



HAL
open science

**De la rose sauvage à la rose domestiquée :
caractérisation du rôle d'APETALA2L dans la formation
de la fleur double chez le rosier**

Léa François

► **To cite this version:**

Léa François. De la rose sauvage à la rose domestiquée : caractérisation du rôle d'APETALA2L dans la formation de la fleur double chez le rosier. Biologie végétale. Université de Lyon, 2019. Français. NNT : 2019LYSEN029 . tel-03738137

HAL Id: tel-03738137

<https://theses.hal.science/tel-03738137>

Submitted on 25 Jul 2022

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.



Numéro National de Thèse : 2019LYSEN029

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée par

l'École Normale Supérieure de Lyon

École Doctorale N°340

Biologie Moléculaire, Intégrative et Cellulaire

Spécialité de doctorat : Sciences de la Vie

Discipline : Sciences de la Vie

Soutenue publiquement le 16/07/2019, par :

Léa FRANÇOIS

De la rose sauvage à la rose domestiquée : caractérisation du rôle d'*APETALA2L* dans la formation de la fleur double chez le rosier.

Devant le jury composé de :

DAMERVAL Catherine, Directrice de Recherche, CNRS, Rapporteur

JOUANNIC Stefan, Chargé de Recherche, IRD, Rapporteur

ABERLENC Frédérique, Ingénieure de Recherche, IRD, Examinatrice

SCUTT Charlie, Directeur de Recherche, CNRS, Examinateur

BENDAHMANE Mohammed, Directeur de Recherche, INRA, Directeur de thèse

RAYMOND Olivier, Maître de Conférence, Lyon 1, Co-encadrant de thèse

REMERCIEMENTS

Choisir de faire une thèse, et la mener à bien, n'est pas une chose aisée. C'est même une chose impossible à faire seule. En y réfléchissant, c'est comme marcher dans la nuit au bord d'une falaise avec pour seule source lumineuse le ciel étoilé : il faut beaucoup d'étoiles pour s'orienter et garder l'équilibre. Et aujourd'hui, j'ai une infinité d'étoiles à remercier. Evidemment, il y a les plus grandes étoiles, que l'on voit quoi qu'il arrive. Mais dès que nos yeux s'habituent, il est aussi possible de distinguer de bien nombreuses petites étoiles qui, ensemble, éclairent et rassurent tout autant. Et c'est comme ça que je me représente mes trois années de thèse, voire mes sept années à l'ENS.

Mais que trouve-t-on dans ce ciel de thèse étoilé ? Pour commencer, on trouve une voie lactée scientifique.

Pour m'avoir accueillie dans ton équipe et formée, pour m'avoir appris à défendre mes idées et pour m'avoir donné l'opportunité d'enseigner et d'encadrer des étudiants, Mohammed, merci.

Jérémy, étoile de la bioinformatique ! Etoile de la bonne cuisine ! Etoile du bricolage ! J'ai conscience que sans toi, ce manuscrit ne serait pas aussi élégant. Peut-être même qu'il ne serait pas tout court. Pour m'avoir tant appris, pour m'avoir épaulée, orientée et ouverte au code, un merci infini.

Olivier. Merci de m'avoir éclairée de ton soutien, de ta tempérance et de tes discussions et conseils avisés. Il fallait bien ça !

Darcy, moi non plus, je n'aurais pas pu rêver d'une meilleure stagiaire ! Tu as illuminé ma deuxième année de ton enthousiasme et de ta joie de vivre : cette thèse te doit beaucoup !

Au reste de l'équipe aussi, je veux dire merci. Merci Marion, en bonne éclairceuse, tu as su trouver le gène candidat dont chaque thésard rêve. Xiaopeng, thank you for welcoming me so generously. Merci Philippe et Priscilla pour votre aide et vos brillants enseignements de culture *in vitro*. Merci Véronique et Judit pour vos expertises en biologie moléculaire, merci Julie et Léo, pour m'avoir montré en avance ce qu'était une thèse. Merci Claudya pour ta bonne humeur flamboyante et tes excellents gâteaux. Ensemble, vous formez une constellation bien à vous.

J'aimerais remercier aussi les membres du jury, Catherine Damerval, Stefan Jouannic, Frédérique Aberlenc et Charlie Scutt qui vont prendre le temps d'évaluer mon travail et qui, je l'espère en écrivant ces mots, me donneront l'occasion de finir ce périple par un joli feu de joie. Merci aussi aux membres de mon comité de thèse, Christel Carles et Stéphane Vincent, pour vos discussions éclairées et vos encouragements. Merci aux collaboratrices de Bordeaux, Aline Potier et Béatrice Denoyer, pour les transformations génétiques.

Aussi, j'aimerais remercier l'équipe du département de biologie de l'ENS, qui m'a permis de découvrir les dessous de l'enseignement à l'Université. Enseigner a vraiment été une source d'équilibre pour moi, apportant à chaque fois sa lueur d'enthousiasme !

Le RDP est un endroit idéal pour se construire en tant que scientifique. Chaque étoile à son importance. Chaque étoile éclaire, de son expertise scientifique ou humaine, un pan de chaque projet. Et je vous suis reconnaissante, à tous, de toutes vos attentions et suggestions. En particulier, Pierre, la star de la biomol ; Raquel et Joseph, les étoiles montantes de la biochimie ; les brillants Patrice, Alexis et Justin, sans qui nos plantes seraient à la dérive ; Hervé et Isa, sans qui mes milieux auraient été, eux aussi, à la dérive ; Roxane, Stéphanie, Cindy et Brigitte qui gèrent à merveille toute la partie administrative ; Michiel, le roi du développement floral.

Aussi, je tiens à remercier chaleureusement Alexis, Thomas, Marie-Cécile, Fabrice et Carlos pour votre bienveillance, vos conseils et vos éclairages nouveaux. Je suis convaincue que c'est seulement en rassemblant différentes perspectives que l'on peut faire les bons choix et trouver son propre chemin.

Dans mon ciel étoilé, on trouve aussi de nombreux astres qui, bien qu'éloignés de mon sujet scientifique, ont tout autant apporté à la réussite de ce projet par leur présence et leur soutien.

Nicolas, je n'oublierai ni ce matin de L3 où je me suis rendue compte que j'avais enfin trouvé un compère à l'ENS, ni ce – ou plutôt ces – jours de préparation AGREG où j'ai compris que tu étais bien plus que ça : un véritable ami, et brillant qui plus est ! Si ce n'est grâce à toi, je ne sais pas si je pourrais, aujourd'hui, regarder mon avenir d'enseignante aussi sereinement.

Marie, chère Marie, toi aussi tu es une de ces étoiles lumineuses qui indiquent toujours le Nord. Aucun risque de se perdre tant que tu es dans les parages !

Noémie, petite sœur. Je ne suis plus gênée de le dire, tu es mon plus grand modèle, une Grande Ourse flamboyante. Il n'y a pas de mots pour donner la mesure de l'admiration et la reconnaissance que j'ai pour toi.

Raquel, tu es comme un feu follet, à toujours apporter la lumière où tu vas. J'ai de la chance de t'avoir eue à mes côtés dans cette dernière ligne droite.

Enfin, Patrick, ami martien, merci pour ta parka à Roscoff, je serais certainement morte de froid sinon. Mais surtout, merci de ne t'être jamais lassé depuis.

A vous cinq : je vous remercie du fond du cœur. Pour m'avoir toujours montré la voie. Pour m'avoir toujours prouvé, par votre présence parfois discrète, toujours indéfectible, que je n'ai jamais été seule face à aucun de mes choix d'adulte, professionnels et personnels. Et c'est le cœur plein de gratitude, d'attachement et d'émerveillement que je me dis que je peux vous compter parmi mes plus proches amis !

Mais vous n'êtes pas les seuls. Garance, tu as été la co-bureau de rédaction – et bien plus que ça – parfaite. Les poulettes, vous avez été des co-bureaux flamboyantes ! Nathalie, tu as été une tutrice de premier choix, et malgré les difficultés, je ne regrette pas d'avoir suivi tes conseils éclairés car je peux dorénavant regarder en arrière avec fierté et aller de l'avant avec optimisme et enthousiasme.

Sana, Yannick, Patrick et Aurélie, merci pour vos conseils avisés quant à mon orientation future : sans vous, je serais certainement encore pétrifiée devant la multitude d'établissements de banlieue parisienne.

Merci à Adriana, et à toutes les autres étoiles de la danse, pour m'avoir initiée à la danse puis à son enseignement : vous avez été, vous aussi, une source d'équilibre sans pareil.

Merci Maman, Papa, Karin et Christian, savoir que vous assurez toujours mes arrières est une chance sans mesure. Merci aux étoiles filantes qui, si elles lisent ces mots, se reconnaîtront certainement. Vos lumières, même brèves, m'ont fait grandir et m'ont appris à avoir confiance.

Et tout le RDP – la version étendue, dont je ne veux pas faire de liste par peur d'oublier quelqu'un –, pour vos rires, bières, cafés, goûters, tricots, sorties et soutiens en cas de crise. Pour vos petites attentions qui mettent toujours du baume au cœur. Savoir qu'à la fin de la journée, il y a toujours quelqu'un pour décompresser rend tout plus facile, ou parfois, au moins, moins difficile.

Bientôt, je partirai vers de nouveaux horizons mais j'espère ne jamais oublier le ciel étoilé lyonnais et ses nombreux repères !

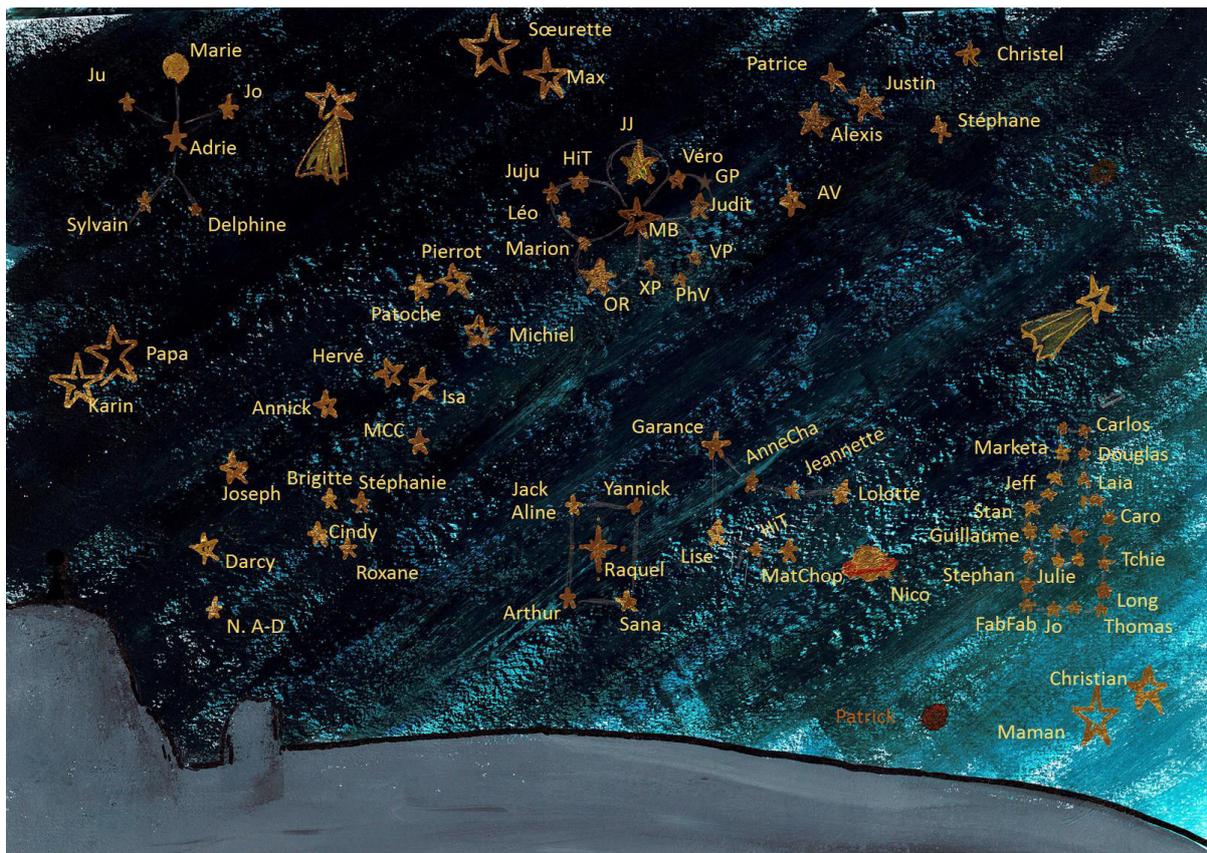


Figure 0 : Aperçu du panorama à l'origine de l'aboutissement de cette thèse, du point de vue d'une doctorante en biologie végétale.

Une voie lactée et différentes constellations peuvent être identifiées. Chaque étoile, et certainement d'autres étoiles non représentées ici, a pu jouer un rôle essentiel et est chaleureusement remerciée !

RÉSUMÉ

Les roses à fleurs doubles attirent sélectionneurs et scientifiques depuis de nombreux siècles. L'analyse des taux de ségrégation et cartes génétiques indique que le passage de la fleur simple à la fleur double est dû à une seule mutation dominante située sur le chromosome 3. Cette mutation conduit à une conversion homéotique d'une partie des étamines en pétales, soulignant la possibilité que cette mutation impacte certains gènes du modèle ABC. Il y a quelques années, notre équipe a démontré que l'augmentation du nombre de pétales chez le rosier était corrélée à une restriction de l'expression de *RcAGAMOUS* (*RcAG*) vers le centre du méristème floral. Cependant, *RcAG* étant porté par le chromosome 5, il ne peut être le déterminant génétique de la fleur double. Il a donc été supposé que la mutation en cause se trouvait dans un gène intervenant en amont de *RcAG*.

Récemment, nous avons séquencé, assemblé et publié le génome de *Rosa chinensis* cv 'Old Blush' un ancêtre des rosiers modernes qui produit déjà des fleurs doubles. L'assemblage, de très bonne qualité, nous a aidé à reconstruire la séquence des deux haplotypes de l'intervalle contenant la mutation liée à la fleur double. Nous avons identifié, parmi les 631 gènes de cet intervalle, un gène *APETALA2-LIKE* (*RcAP2L*) comme candidat plus que prometteur. En effet, il a été découvert que ce gène existait sous la forme de deux allèles, l'un d'entre eux contenant un grand élément transposable, donnant lieu à un allèle tronqué résistant à l'inhibition par miR172, appelé *RcAP2L^{Δ172}*. Sachant que la surexpression d'un variant résistant au miR172 entraîne souvent la formation de pétales supplémentaires chez *A. thaliana*, j'ai démontré que la présence de ce variant corrèle avec le phénotype « fleur double » chez les rosiers d'origine chinoise. Enfin, alors qu'AP2 est capable d'inhiber l'expression d'AG en se liant directement à ses séquences régulatrices chez *A. thaliana*, j'ai confirmé la capacité des protéines codées par les deux allèles de *RcAP2L* à lier les séquences régulatrices de *RcAG*, *in vitro*. Fait intéressant, un mutant spontané d'Old Blush' à fleurs simples porte une mutation dans le domaine de liaison à l'ADN de *RcAP2L^{Δ172}* qui bloque l'interaction entre la protéine et les séquences régulatrices de *RcAG*, suggérant que la non-fonctionnalisation de cet allèle est suffisante pour réverter le phénotype.

À partir de ces résultats, je propose donc un modèle pouvant expliquer la formation de fleurs doubles chez les rosiers chinois et peut-être d'autres Rosaceae, dans lesquelles la protéine *RcAP2L^{Δ172}* peut s'accumuler du fait de sa résistance au miR172 et restreindre davantage l'expression de *RcAG* au centre du méristème floral. Ainsi, la frontière entre les domaines A et C se trouve elle aussi déplacée vers le centre du méristème, ce qui induit la conversion des étamines en pétales.

ABSTRACT

Roses exhibiting double flowers have intrigued both breeders and scientists for decades. Based on segregation ratios and genetic maps, it is known that the switch from simple to double flower is due a single dominant locus on chromosome 3. When present in its mutated form, this locus leads to a homeotic conversion of stamens into petals, suggesting a mechanism involving the ABC genes. A few years ago, our team demonstrated that the increase in petal number correlates with a restriction of *RCAGAMOUS* (*RcAG*) expression domain towards the center of the floral meristem. However, as *RcAG* is located on chromosome 5, the causative mutation was assumed to act as a regulator of this gene.

Recently, we sequenced, assembled and published the double-flowered *Rosa chinensis* cv 'Old Blush' genome sequence with a high-quality assembly that helped us to reconstruct the sequence of the two haplotypes of the interval containing the double flower mutation. Among the 631 genes from this interval, we identified here an *APETALA2-LIKE* (*RcAP2L*) gene as a strong candidate. Indeed, this gene was found to exist as two alleles, with one containing a large transposable element resulting in a truncated, miR172-resistant, variant named *RcAP2L*^{Δ172}. Knowing that the overexpression of a miR172-resistant variant of AP2 leads to the formation of extra petals (and sometimes stamens) in *Arabidopsis*, we investigated the presence of this variant in simple and double flower varieties. The presence of *RcAP2L*^{Δ172} was found to correlate with the double flower phenotype in Chinese roses and was not observed in any of the simple-flowered roses studied. Finally, as AP2 is able to inhibit AG expression by directly binding to its regulatory sequences in *A. thaliana*, I confirmed that both *RcAP2L* proteins are also able to recognize *RcAG* regulatory sequences *in vitro*. Interestingly, we identified a mutation in the *RcAP2L*^{Δ172} DNA binding domain of a spontaneous simple-flowered mutant of 'Old Blush' that prevents the interaction between *RcAP2L*^{Δ172} and *RcAG* regulatory sequences, suggesting that the non-functionalization of this allele is sufficient to the revert the phenotype.

A working model is thus proposed for double flower formation in rose, that could be valid for other Rosaceae, whereby *RcAP2L*^{Δ172} protein may accumulate due to its resistance to miR172 and consequently may repress more *RcAG* towards the center of the floral meristem, leading to the sliding of the A/C border and thus the conversion of stamens into petals.

Table des matières

Chapitre 1 : Introduction	15
I. Qu'est-ce qu'une fleur ?.....	16
1. La fleur, un outil de reproduction.....	16
2. Origine évolutive de la fleur.....	18
II. Le développement floral : du méristème à la différenciation des organes floraux....	19
1. Les méristèmes apicaux caulinares produisent l'ensemble des organes aériens de la plante	19
3. Déterminisme des organes floraux : le modèle ABC chez Arabidopsis et Antirrhinum	20
4. Mécanismes moléculaires et évolution du modèle en modèle ABCE	24
5. Le modèle ABC est-il généralisable à toutes les Angiospermes ?.....	26
6. Effecteurs/cibles des gènes homéotiques	27
7. Différenciation des organes, l'exemple des pétales	28
III. La fleur double, une anomalie florale très appréciée et étudiée.....	30
1. Description des fleurs doubles.....	32
2. Origine développementale des pétales des fleurs doubles.....	32
3. Propagation des plantes à fleur double.....	34
4. Origine des fleurs doubles	35
5. Importance d'AGAMOUS dans la formation des fleurs doubles	36
IV. Le rosier : une plante modèle	38
1. Le rosier appartient aux Rosaceae.....	40
2. Le rosier, une plante domestiquée	42
3. Culture et propagation du rosier	45
4. Disponibilité de nombreuses cartes génétiques.....	45
5. Transformation génétique par Agrobacterium.....	46
6. Disponibilité de données d'expression des gènes	48
V. Le séquençage du génome du rosier, un outil puissant.....	50
1. Le génome d'Old Blush', un génome séquencé de très bonne qualité	50
2. Reconstruction de l'histoire évolutive entre espèces proches du rosier	51
3. Histoire évolutive au sein du genre Rosa : exemple d'Old Blush'	51
4. Reconstruction de l'intervalle associé à la fleur double chez le rosier.....	52
Chapitre II : Objectifs de la thèse	55
I. État de l'art avant le début de la thèse	56
II. Première étape : identification des candidats pour le déterminant de la fleur double	56

III.	Deuxième étape : validation du rôle de ces candidats	57
IV.	Troisième étape : caractérisation du mécanisme d'induction de la fleur double lié à cette mutation.....	57
V.	Quatrième étape : étude de la conservation de ce mécanisme chez d'autres plantes ornementales	58
	Chapitre III : Résultats	59
I.	Identification d'un premier gène candidat pour le contrôle du nombre de pétales chez le rosier : le cas de <i>RcUGO</i>	60
1.	Contexte.....	60
2.	<i>RcUGO</i> , un « small leucine zipper » présent dans l'intervalle	62
3.	Chez <i>A. thaliana</i> , les orthologues de <i>RcUGO</i> interviennent dans le contrôle de l'activité méristématique	64
4.	Le niveau d'expression de <i>RcUGO</i> diffère peu entre fleurs simples et fleurs doubles 66	
5.	Patron d'expression de <i>RcUGO</i> dans les bourgeons floraux	70
6.	La diminution d'expression de <i>RcUGO</i> n'induit pas de modification du nombre de pétales chez 'Old Blush'	72
7.	La surexpression de <i>RcUGO</i> chez 'Old Blush' n'a pas permis de régénérer de plantes transformées génétiquement	75
8.	Conclusions	76
II.	Identification d'un second candidat pour le déterminisme de la fleur double : le cas de <i>RcAP2L</i>	78
1.	Rappel du contexte à propos du gène de la fleur double chez 'Old Blush'	78
2.	<i>AP2L</i> , un candidat pertinent pour la formation de la fleur double	80
3.	Résultats complémentaires et conclusions	92
III.	Caractérisation fonctionnelle de <i>RcAP2L</i>	94
1.	<i>RcAP2L</i> , un facteur de transcription spécifique des jeunes tissus en division	94
2.	<i>RcAP2L</i> , un facteur de transcription pouvant se fixer aux séquences régulatrices de <i>RcAG</i>	98
3.	Une mutation dans le domaine <i>AP2R1</i> de <i>RcAP2</i> ^{Δ172} pourrait induire la réversion du phénotype « fleur double ».....	102
4.	Le blocage de l'expression de <i>RcAP2L</i> ^{Δ172} chez 'Old Blush' induit des retards de développement végétatif.....	106
5.	Surexpression de la version tronquée chez les Rosaceae	108
6.	Surexpression de la version tronquée chez <i>A. thaliana</i>	112
7.	Conclusions	116
	Chapitre IV : Discussion et perspectives	117
I.	Vers une meilleure compréhension du réseau de signalisation déterminant le nombre de pétales chez les rosiers.....	120

1.	Confirmation que RcAP2L ^{Δ172} est nécessaire et suffisant au phénotype « fleur double » chez les rosiers d'origine chinoise.....	120
2.	Identification des partenaires de RcAP2L dans la régulation de l'expression de ses cibles	121
3.	Identification des gènes cibles de RcAP2L.....	122
4.	Les rosiers non chinois à fleurs doubles seraient-ils mutés dans un autre intervenant de la voie RcAP2L ?.....	123
5.	Origine du contrôle polygénique du nombre de pétales au sein des rosiers à fleurs doubles	124
6.	Recherche d'un lien entre RcAP2L et RcUGO	126
II.	Importance horticole de l'identification de la mutation de <i>RcAP2L</i>	126
1.	Vers une accélération du processus de création variétale chez le rosier.....	126
2.	Mutation fleur double dominante chez d'autres espèces ornementales : une modification de la famille euAP2 ?.....	127
III.	Importance fondamentale : conservation des mécanismes de contrôle de la fonction C chez les plantes à fleurs ?.....	127
1.	Deux voies de restriction de la fonction C ont été sélectionnées en fonction des espèces	127
2.	Histoire évolutive de la famille euAP2.....	128
IV.	Quels moteurs de l'évolution phénotypique ?.....	129
1.	Importance des duplications de séquences dans la diversification des fonctions au sein d'une famille de gènes.....	129
2.	Les éléments transposables comme source de variabilité phénotypique importante	131
	Chapitre V : Matériel et Méthodes	133
I.	Culture et transformation des plantes.....	134
1.	Rosiers.....	134
2.	<i>Arabidopsis thaliana</i>	134
3.	Transformation d' <i>Agrobacterium tumefaciens</i>	134
4.	Transformation génétique des rosiers et sélection des transformants.....	136
5.	Transformation génétique d' <i>A. thaliana</i> et sélection des lignées transformées..	136
6.	Transformation génétique des fraisiers.....	138
II.	Biologie moléculaire.....	138
1.	Extraction d'ADN.....	138
2.	Extraction des ARNm et production d'ADNc	138
3.	Clonage des séquences codantes à surexprimer	140
4.	Clonage des séquences codantes pour production des protéines dans <i>E. coli</i>	140
5.	Choix et clonage des microARN artificiels pour diminuer l'expression d'un gène	140

6.	Choix et clonage des ARN guides pour éditer un gène par CRISPR/Cas9.....	142
7.	Vérification de la présence de l'élément transposable par PCR.....	144
III.	Analyses d'expression des gènes	144
1.	RT-PCR quantitatives (qRT-PCR)	144
2.	Hybridations in situ	146
IV.	Analyses des interactions protéine-promoteur	147
1.	Prédictions structurales	147
2.	Synthèse des protéines dans E. coli	147
3.	Purification des protéines.....	147
4.	Synthèse des sondes d'ADN marquées.....	148
5.	Expériences de retards sur gel.....	148
V.	Analyses phénotypiques des lignées.....	150
1.	Comptage du nombre d'organes floraux.....	150
2.	Suivi de la date de floraison d'A. thaliana.....	150
3.	Microscopie Électronique à Balayage	150
VI.	Analyses bioinformatiques.....	150
1.	Génomomes et séquences utilisés.....	150
2.	Analyses des données de transcriptomique	152
3.	Analyses de l'expression allèle-spécifique.....	153
4.	Correction des annotations du génome de rosier	155
VII.	Analyses phylogénétiques	155
1.	Identification des séquences d'intérêt	155
2.	Alignement des séquences	156
3.	Construction des arbres.....	156
	BIBLIOGRAPHIE.....	157
	ANNEXES.....	173
I.	Plasmides.....	175
II.	Article : le génome de <i>Rosa chinensis</i> cv 'Old Blush'	176

LISTE DES FIGURES

- Figure 1** : Organisation d'une fleur sauvage de rosier.
- Figure 2** : Boucle de rétrocontrôle *CLV/WUS* à l'origine du déterminisme des cellules souches dans les méristèmes apicaux caulinaires.
- Figure 3** : Modèle ABC proposé dans les années 1990.
- Figure 4** : Le « quartet model », un modèle moléculaire de l'interaction entre les différentes protéines MADS du modèle ABCE lors du déterminisme des organes floraux.
- Figure 5** : Quelques gènes impliqués dans la croissance et différenciation du pétale lors du développement floral.
- Figure 6** : Exemples de fleurs ornementales étudiées pour l'apparition du caractère « fleur double ».
- Figure 7** : Morphologie d'une fleur double ; exemple de différentes variétés de rosier.
- Figure 8** : Arbre phylogénétique replaçant les différentes espèces modèles utilisées en laboratoire.
- Figure 9** : Arbre phylogénétique des principales Rosaceae à importance économique.
- Figure 10** : Carte génétique et étude de traits d'intérêt chez le rosier : le cas de la fleur double.
- Figure 11** : Résumé des avantages de *R. chinensis* cv 'Old Blush' en tant que modèle d'étude scientifique.
- Figure 12** : Méthode de reconstruction des deux haplotypes de l'intervalle contenant le déterminant de la fleur double.
- Figure 13** : Méthode ayant permis l'identification d'un premier candidat pour le déterminisme génétique de la fleur double.
- Figure 14** : *RcUGO* code pour une petite protéine de la famille des ZPR, ou *small leucine zipper*.
- Figure 15** : Modèle de fonctionnement des ZPR comme inhibiteurs des HD-ZIP de classe III.
- Figure 16** : Comparaison du niveau d'expression de *RcUGO* chez 'Old Blush' et chez 'Single Pink'.
- Figure 17** : Les niveaux d'expression de *RcUGO* chez 'Old Blush' et 'Single Pink' sont proches.
- Figure 18** : Patron d'expression de *RcUGO*.
- Figure 19** : Le domaine d'expression de *RcUGO* recoupe celui de *RcWUS* et *RcAG*.
- Figure 20** : L'inhibition de *RcUGO* chez 'Old Blush' ne modifie pas le nombre d'organes floraux ou la polarité adaxiale-abaxiale des organes aériens.
- Figure 21** : Le rosier modèle 'Old Blush' est hétérozygote pour la mutation à l'origine du phénotype « fleur double ».
- Figure 22** : Un dérivé d'Old Blush' obtenu après culture *in vitro* a perdu l'allèle *RcAP2L*^{Δ172} et produit de nouveau des fleurs simples.
- Figure 23** : *RcAP2L* est exprimé essentiellement dans les jeunes tissus aériens.
- Figure 24** : Comparaison des niveaux d'expression des différents membres de la famille des euAP2.
- Figure 25** : Comparaison des niveaux d'expression des différents membres de la famille des euAP2 dans les jeunes bourgeons floraux.
- Figure 26** : Modèle prédictif de la structure du domaine AP2R1 de *RcAP2L* et de son interaction putative avec l'ADN.
- Figure 27** : *RcAP2L*^{WT} et *RcAP2L*^{Δ172} sont capables d'interagir spécifiquement avec une portion du deuxième intron de *RcAG*.
- Figure 28** : Caractérisation d'un mutant spontané d'Old Blush' produisant des fleurs simples et exprimant une version de *RcAP2L*^{Δ172} avec une mutation dans le domaine de liaison à l'ADN AP2R1.

Figure 29 : La mutation identifiée dans le domaine AP2R1 de 'Single Pink' impacte un acide aminé extrêmement conservé au sein de la famille des euAP2 et réduit la capacité de liaison de la protéine au promoteur de *RcAG*.

Figure 31 : Chez *Rosa chinensis*, l'inhibition de l'expression de *RcAP2L*^{Δ172} par des microARN artificiels semble perturber le développement végétatif.

Figure 32 : La surexpression des différentes versions de *FvAP2L* chez le fraisier n'a pas permis de régénérer de plantes transformées.

Figure 33 : Phénotypes floraux observés chez les plantes surexprimant différents allèles de *RcAP2L*.

Figure 34 : Impact de la surexpression de *RcAP2L* sur la durée de la période végétative de *A. thaliana*.

Figure 35 : Bilan des connaissances produites lors de la thèse.

Figure 36 : Généalogie des rosiers modernes et déterminisme génétique du phénotype « fleur double ».

Figure 37 : Vue d'ensemble du protocole de transformation chez le rosier.

Figure 38 : Construction d'un microARN artificiel.

LISTE DES TABLEAUX

Tableau 1 : Résumé des études sur le lien entre les gènes de classe C et le phénotype « fleur double ».

Tableau 2 : Comparaison des caractéristiques des différentes Rosaceae modèles.

Tableau 3 : Ensemble des variétés de rosiers étudiées.

Tableau 4 : Ensemble des lignées transformées étudiées.

Tableau 5 : Amorces utilisées (partie 1).

Tableau 5 : Amorces utilisées (partie 2).

Tableau 6 : Liste des milieux et tampons utilisés dans cette étude.

LISTE DES ABRÉVIATIONS

ADN : <u>A</u> cide <u>D</u> éoxyribo <u>N</u> ucléique	MBP : « <u>M</u> altose <u>B</u> inding <u>P</u> rotein »
AFLP : « <u>A</u> mplified <u>F</u> ragment <u>L</u> ength <u>P</u> oly- morphism »	MOS : <i>R. moschata</i>
AG : AGAMOUS	MS : milieu <u>M</u> urashige et <u>S</u> koog
ANT : AINTEGUMENTA	MUT : <i>R. chinensis</i> cv ' <u>M</u> utabilis'
AP1/AP2/AP3 : APETALA1/APETALA2/APETALA3	NF-YA : NUCLEAR FACTOR-YA
ARN : <u>A</u> cide <u>R</u> ibo <u>N</u> ucléique	NOIS : <i>R. hybrida</i> cv ' <u>B</u> lush <u>N</u> oisette'
BFL : <i>R. chinensis</i> cv ' <u>B</u> ébé <u>F</u> leuri'	'OB' : <i>R. hybrida</i> cv ' <u>O</u> ld <u>B</u> lush'
BENG : <i>R. chinensis</i> cv ' <u>B</u> engale d'Automne'	PCR : « <u>P</u> olymerase <u>C</u> hain <u>R</u> eaction »
BPE : BIGPETAL	PHYRE2 : « <u>P</u> rotein <u>H</u> omology/ <u>A</u> nalog <u>Y</u> <u>R</u> eco- gnition <u>E</u> ngine 2 »
CaMV : « <u>C</u> auliflower <u>M</u> osaic <u>V</u> irus »	PI : PISTILLATA
CaRG : CC-(A rich)-GG	PLE : PLENA
COMT : <i>R. hybrida</i> cv ' <u>C</u> omtesse de Cayla'	PVP : <u>P</u> oly <u>V</u> inyl <u>P</u> yrrolidone
CRAM : <i>R. hybrida</i> cv ' <u>C</u> ramoisi Supérieur'	QTL : « <u>Q</u> uantitative <u>T</u> rait <u>L</u> ocus »
CDS : « <u>C</u> oding <u>S</u> equence »	RAPD : « <u>R</u> andom <u>A</u> mplified <u>P</u> olymorphic <u>D</u> NA »
ChIP-seq : <u>C</u> hromatine <u>I</u> mmuno <u>P</u> recipitation suivie d'un <u>s</u> équence	REV : REVOLUTA
CLV1/CLV3 : CLAVATA1/CLAVATA3	RME : <i>R. hybrida</i> cv ' <u>R</u> ouge <u>M</u> eilland'
CNA : CORONA	RT-qPCR : « <u>R</u> everse <u>T</u> ranscription and <u>q</u> uantitative <u>P</u> CR »
cv : cultivar	SAN : <i>R. chinensis</i> cv ' <u>S</u> anguinea'
DUC : <i>R. hybrida</i> cv ' <u>D</u> ucher'	SEP1/SEP2/SEP3/SEP4 : SEPALATA1/ SEPALATA 2/ SEPALATA 3/ SEPALATA 4
ERF : « <u>E</u> thylene <u>R</u> esponse <u>F</u> actor »	SHP1/2 : SHATTERPROOF1/ SHATTERPROOF2
FAR : FARINELLI	SIM : « <u>S</u> hoot <u>I</u> nducing <u>M</u> edium »
FPKM : « <u>F</u> ragments <u>P</u> er <u>K</u> ilobase of transcript per <u>M</u> illion mapped reads »	SNP : « <u>S</u> ingle <u>N</u> ucleotide <u>P</u> olymorphism »
FRA : <i>R. hybrida</i> cv 'La <u>F</u> Rance'	'SP' : <i>R. chinensis</i> cv 'Single Pink'
GBD : « <u>G</u> CC-box <u>B</u> inding <u>D</u> omain »	SPO : <i>R. chinensis</i> cv ' <u>S</u> pontanea'
GDR : « <u>G</u> enome <u>D</u> atabase for <u>R</u> osaceae »	SSR : « <u>S</u> imple <u>S</u> equence <u>R</u> epeat »
GEN : <i>R. hybrida</i> cv ' <u>G</u> énéral Schablikine'	START : « <u>S</u> teroidogenic <u>A</u> cute <u>R</u> egulatory protein-related lipid- <u>T</u> ransfer »
GIG : <i>Rosa gigantea</i>	STK : SEEDSTICK
HD-ZIP : <u>H</u> oméo <u>D</u> omaine couplé à un domaine leucine <u>Z</u> ipper	SMZ : SCHLAFMUTZE
HDA19 : HISTONE DÉACÉTYLASE19	SNZ : SNARCHZAPFEN
HERO : <i>R. hybrida</i> cv ' <u>H</u> érodiade'	TAE : tampon <u>T</u> ris, <u>A</u> cétate, <u>E</u> DTA
HUM : <i>R. chinensis</i> cv ' <u>H</u> ume's Blush'	TE : « <u>T</u> ransposable <u>E</u> lement »
IPTG : <u>I</u> so <u>P</u> ropyl- β - <u>D</u> -1- <u>T</u> hyo <u>G</u> alactopiranoside	TOE1/TOE2/TOE3 : TARGET OF EAT1/ TARGET OF EAT2/ TARGET OF EAT3
ISH : « <u>I</u> n <u>S</u> itu <u>H</u> ybridization »	TPL : TOPLESS
IVT : « <u>I</u> n <u>V</u> itro <u>T</u> ranscription »	VIGS : « <u>V</u> irus <u>I</u> nduced <u>G</u> ene <u>S</u> ilencing »
LG : « <u>L</u> inkage <u>G</u> roup »	WMD3 : <u>W</u> eb <u>M</u> icroRNA <u>D</u> esigner 3
LOU : <i>R. hybrida</i> cv ' <u>L</u> ouise d'Arzens'	WUS : WUSCHEL
LTR : « <u>L</u> ong <u>T</u> erminal <u>R</u> epeat »	WIS : <i>R. wichurana</i>
LZ : « <u>L</u> eucine <u>Z</u> ipper »	
MAC : <u>M</u> éristème <u>A</u> pical <u>C</u> aulinaire	
MADS : <u>M</u> CM1, <u>A</u> GAMOUS, <u>D</u> EFICIENS, <u>S</u> RF	

Chapitre 1 : Introduction



Illustrations prouvant l'existence fleurs doubles à travers le monde, parfois dès l'Antiquité.

À gauche: fresque de la maison des Vettii, I^{er} siècle après J.-C., Pompéi.
Source: <https://leg8.fr/monde-romain/rose-rome-antique>

À droite: peinture réalisée par l'artiste Huang Jucai (黄居案) à l'époque de la dynastie Song (X^e siècle après J.-C.), Chine.
Source: <https://ltfc.net/img/5c5fef1063ff454f3bf118c>

Parmi les merveilles de la nature, les fleurs passionnent et fascinent artistes, poètes et « commun des mortels » plus que tout. Elles embellissent le quotidien par leur beauté et leur parfum, caractéristiques qui ont été sélectionnées par l'homme. La sélection des plantes ornementales s'est faite essentiellement sur l'aspect visuel, notamment sur l'augmentation du nombre de pétales par fleur, qui leur donne une apparence plus dense et très attractive pour l'œil. Ce caractère, potentiellement défavorable pour la plante, a été sélectionné, indépendamment et de nombreuses fois, chez des espèces variées provenant de différentes familles et différents genres. Quels sont les mécanismes sélectionnés ? Sont-ils analogues dans les différentes espèces ? Que nous apportent ces informations sur le processus de domestication des plantes ornementales ? C'est dans ce cadre que se place mon travail de thèse qui, pour tenter de répondre à ces questions, utilise le rosier comme modèle.

I. Qu'est-ce qu'une fleur ?

Lorsque nous pensons aux fleurs, nous pensons d'abord à ces choses colorées et voyantes qui décorent nos jardins et maisons. Mais les fleurs sont avant tout des organes extrêmement spécialisés et efficaces, fournissant à la plante qui les porte un avantage reproductif sur les plantes qui en sont dépourvues.

1. La fleur, un outil de reproduction

Les plantes à fleurs, ou Angiospermes, constituent le groupe de plantes le plus représenté sur Terre avec à l'heure actuelle plus de 350 000 espèces ayant colonisé tous les habitats (Joppa et al., 2011). En comparaison, le groupe des Gymnospermes, qui bien que paraphylétique est encore souvent considéré comme le groupe frère des Angiospermes, ne contient qu'environ 1000 espèces (Wang and Ran, 2014). Cette réussite évolutive s'expliquerait en partie par l'apparition de la fleur, structure dédiée à la reproduction sexuée et caractéristique des Angiospermes. De plus, après la fécondation, celle-ci se transforme en fruit, une autre structure spécialisée qui a pour rôle de favoriser la dispersion des graines.

Dans sa structure typique, une fleur est composée de quatre cercles d'organes concentriques, chacun appelé verticille, mis en place séquentiellement au cours du développement floral et ayant chacun une fonction principale précise (Figure 1). De l'extérieur vers l'intérieur d'une fleur sont rencontrés : (1) les sépales, généralement verts et à fonction protectrice du bouton floral, (2) les pétales, souvent colorés et à fonction attractrice des pollinisateurs, (3) les étamines, organes reproducteurs mâles, enfin (4) les carpelles, organes reproducteurs femelles parfois contenus dans un pistil. Les termes calice, corolle, androcée et gynécée font chacun référence à un de ces différents

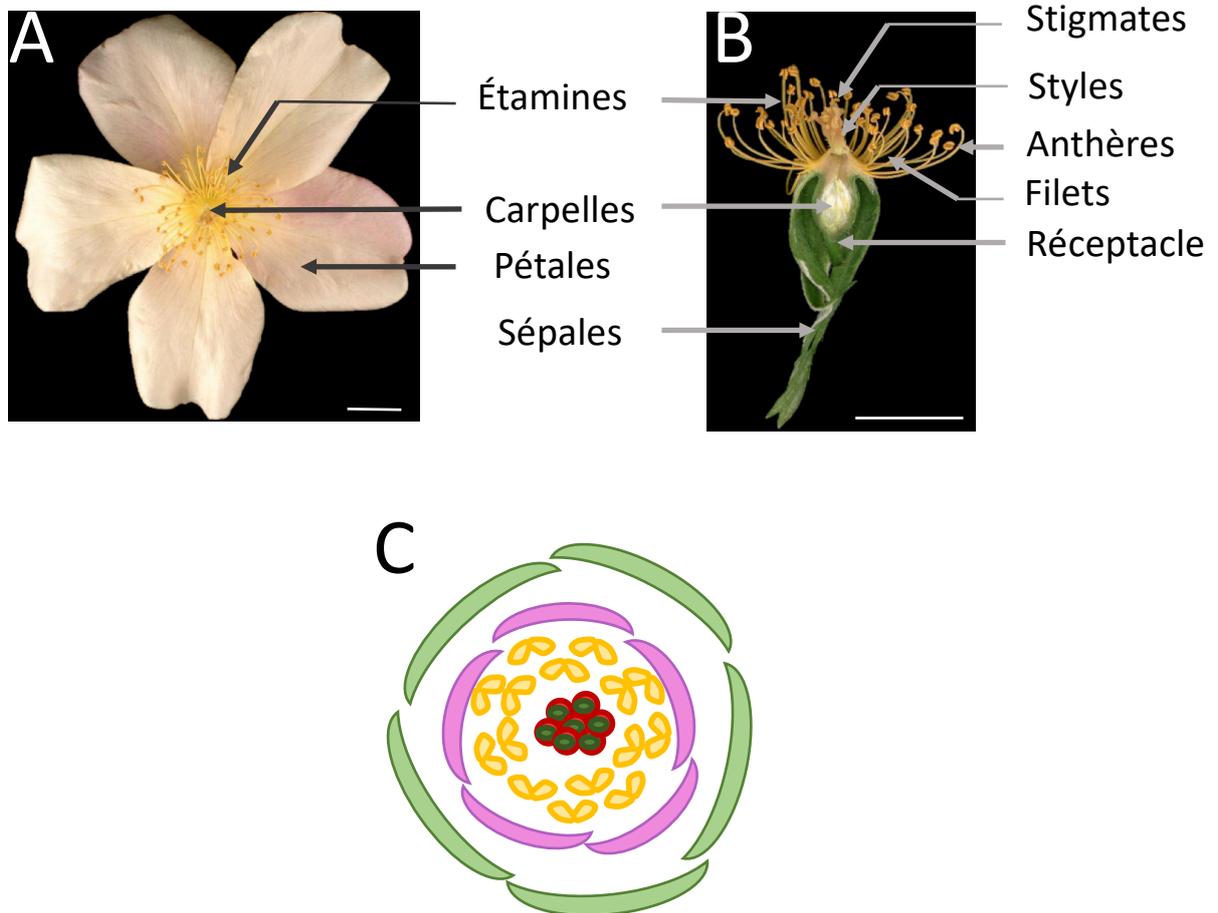


Figure 1 : Organisation d'une fleur sauvage de rosier.

(A) Vue de dessus.

(B) Vue en coupe après retrait des pétales.

(C) Diagramme floral correspondant.

Les rosiers sauvages possèdent un verticille de 5 sépales, un verticille de 5 pétales, un nombre indéterminé d'étamines arrangées en spirale et un nombre indéterminé de carpelles libres insérés dans un réceptacle.

Barre d'échelle : 1 cm

verticilles, de l'extérieur vers l'intérieur respectivement, alors que le terme périanthe fait référence aux deux verticilles extérieurs.

En fonction de la sexualité et du mode de pollinisation de l'espèce, les différents organes floraux peuvent être plus ou moins développés. Ainsi, si la plupart des fleurs sont hermaphrodites (présence simultanée d'étamines et de carpelles), certaines espèces d'Angiospermes ont des fleurs unisexuées, c'est-à-dire portant soit des étamines soit des carpelles. C'est le cas par exemple des fleurs du melon *Cucumis melo* (Martin et al., 2009) ou du kiwi *Actinidia chinensis* (Fraser et al., 2009). Lorsque les fleurs sont pollinisées par le vent (anémophilie), elles portent généralement des étamines produisant une très grande quantité de pollen et des styles plumeux capables de mieux recevoir le pollen en suspension dans l'air. Ces principales adaptations sont souvent accompagnées de la réduction en taille des deux premiers verticilles, dont l'importance fonctionnelle est réduite. Ces fleurs sont donc rarement attractives aux yeux des insectes et des humains et ne sont que peu cultivées pour leurs qualités esthétiques. Ainsi, les plantes de la famille des Poaceae, par exemple, ne produisent pas de fleurs colorées et ont été généralement domestiquées pour leurs qualités nutritives. En revanche, dans le cas de plantes pollinisées par des insectes (entomophilie), les pétales ont pour fonction principale d'attirer les pollinisateurs par leur couleur, leur forme et leur parfum, traits qui attirent aussi particulièrement l'intérêt des jardiniers, sélectionneurs et industries cosmétiques. Alors que ces traits sont donc essentiels à la dissémination de l'espèce, ils peuvent être accompagnés d'autres adaptations utiles à l'attraction des pollinisateurs, comme la production de nectar.

2. Origine évolutive de la fleur

En 1768, Caspar Friedrich Wolff analyse les extrémités des plantes et identifie que tous les organes sont formés dans ces extrémités (Wolff, 1759 : *Theoria Generationis*), que l'on nommera plus tard un méristème. Ainsi, malgré leurs différences de formes et de fonctions, les quatre organes floraux auraient la même origine développementale qu'ils partageraient avec les ébauches foliaires. Peu après, Johann Wolfgang von Goethe proposera la même théorie et envisagera les organes floraux comme des feuilles « métamorphosées » ou « variations autour d'un même thème » (Goethe, 1790 : *The Metamorphosis of plants* ; Coen, 2001). Goethe se fonda notamment sur l'observation de fleurs anormales, et notamment certains rosiers à pétales en très grand nombre et sans organes reproducteurs mâles, pour proposer que chaque organe peut être converti en un autre organe, et qu'ils dérivent tous de feuilles modifiées.

Plusieurs siècles plus tard, les analyses de divers traits floraux au sein de l'ensemble des Angiospermes ont permis la reconstruction théorique du phénotype floral du dernier ancêtre commun des plantes à fleurs, qui existait certainement il y a 250 à 140 millions d'années (Sauquet et al., 2017). Leurs fleurs devaient être hermaphrodites et avec au moins 3 verticilles de 3 organes pétaloïdes chacun

(ou tépales), au moins 3 verticilles de 3 étamines et au moins 5 carpelles libres disposés en spirale. Cette reconstruction suggère que l'ancêtre des Angiospermes possédait certainement des fleurs dont la pollinisation dépendait d'insectes ou autres petits animaux et qu'une réduction du nombre de verticilles a eu lieu dans les débuts de l'évolution des Angiospermes. En effet, les premiers fossiles floraux découverts ont été datés de la fin de l'Aptien (il y a 100 à 120 millions d'années) et correspondent à des fleurs plus simples, avec déjà une réduction du nombre de verticilles au sein de leur périanthe. Cette sélection d'un nombre de verticilles réduit pourrait indiquer que la présence de deux verticilles de tépales, et plus tard sa potentielle évolution en un verticille de sépales et un de pétales, est suffisante, voire avantageuse pour la reproduction et la propagation des individus. Notamment, cela aurait facilité la coévolution entre les pollinisateurs et la structure florale (Woźniak and Sicard, 2018), et par conséquent l'augmentation de la diversité morphologique des fleurs observable aujourd'hui.

II. Le développement floral : du méristème à la différenciation des organes floraux

1. *Les méristèmes apicaux caulinaires produisent l'ensemble des organes aériens de la plante*

Contrairement aux mammifères dont les organes sont tous mis en place durant l'embryogénèse, les plantes sont capables de produire de nouveaux organes tout au long de leur vie. Ceux-ci sont produits par les méristèmes, des tissus spécialisés dans la production continue de tous les tissus différenciés de la plante. Les Méristèmes Apicaux Caulinaires (ou MAC) sont à l'origine de la production des organes végétatifs aériens de la plante. Leur activité est séquentielle : ils produisent à intervalles réguliers des phytomères, qui correspondent à un organe latéral (une feuille pour les méristèmes végétatifs ou un bourgeon floral pour les méristèmes d'inflorescence) et une portion de tige. Les très jeunes bourgeons floraux ou méristèmes floraux produisent, quant à eux, l'ensemble des organes de la fleur.

Des analyses au microscope à transmission puis au microscope confocal ont permis de montrer que les méristèmes sont très structurés. Ainsi, dans les MAC sont trouvées : (1) une zone centrale contenant des cellules indifférenciées et se divisant lentement, (2) une zone périphérique contenant des cellules se divisant plus rapidement et (3) encore plus à l'extérieur du méristème, une zone où sont formés les primordia, c'est-à-dire de petits dômes de cellules en division à l'origine des futures feuilles ou fleurs (Figure 2-A). Le maintien de la structure et de l'activité des méristèmes au cours du temps est permis grâce à la présence en leur centre de cellules souches indifférenciées capables de se diviser tout en maintenant leur nombre. La perte de ces cellules souches induit la perte du méristème ; c'est le cas notamment lors de la perte du facteur de transcription *WUSCHEL* (*WUS*), où un arrêt précoce du

méristème après production de quelques organes et consommation de toutes les cellules souches peut être observé (Figure 2-C ; Laux et al., 1996). Normalement exprimé dans un petit nombre de cellules au centre du méristème, *WUS* est un facteur capable de promouvoir l'identité des cellules souches (Mayer et al., 1998). Au contraire, la mutation du gène *CLAVATA3* (*CLV3*), induit la formation d'un méristème élargi produisant un plus grand nombre d'organes de manière désorganisée (Figure 2-D ; Clark et al., 1995), indiquant que *CLV3* est capable de restreindre le nombre de cellules souches. Ces deux gènes, entre autres (Barton and Poethig, 1993), ont été identifiés comme des facteurs clés de l'homéostasie du méristème. Étant donné que les doubles mutants *clv3/wus* et *clv1/wus* possèdent un phénotype similaire au mutant *wus* (Laux et al., 1996 ; Schoof et al., 2000), il a été proposé qu'ils étaient impliqués dans la même voie de signalisation. Ainsi, le peptide *CLV3*, en se liant à son récepteur *CLAVATA1* (*CLV1*), serait responsable de la restriction de *WUS* vers le centre du méristème (Figure 2-E).

Singulièrement, chaque méristème floral produit l'ensemble des organes contenus dans une – et une seule – fleur puis se termine et arrête de fonctionner. Les verticilles de sépales, pétales, étamines et carpelles sont ainsi séquentiellement produits et *WUS* cesse alors d'être exprimé. Chaque organe est tout d'abord initié sous la forme d'un primordium, qui grandira et se différenciera en un organe adulte selon le programme génétique qu'il exprime.

3. Déterminisme des organes floraux : le modèle ABC chez *Arabidopsis* et *Antirrhinum*

Les gènes à l'origine de la différenciation des primordia en organes floraux spécifiques ont été largement étudiés durant les 30 dernières années. En particulier, au début des années 1990, des expériences de mutagenèse conduites chez *Arabidopsis thaliana* et *Antirrhinum majus* (gueule-de-loup ou muflier), suivies de cribles sur la morphologie florale, ont tout d'abord identifié trois classes de mutants, affectant chacun l'identité des organes de deux verticilles adjacents (Bowman et al., 1991 ; Carpenter and Coen, 1990 ; Coen, 2001).

D'une manière générale, ces mutants sont caractérisés par un maintien de l'organisation florale en verticilles, mais accompagnée d'une conversion de certains organes en d'autres. Ils sont qualifiés de mutants homéotiques, terme tout d'abord introduit par William Bateson en 1894 pour décrire la transformation d'un type d'organes en un autre chez les insectes (Bateson, 1894 ; *Materials for the Study of Variation, Treated with Fatal Regard to Discontinuity in the Origin of Species*). Des phénotypes similaires ont été identifiés chez les deux espèces, bien que celles-ci soient phylogénétiquement assez éloignées. Chez les mutants de classe A, on observe généralement la transformation des sépales en carpelles et des pétales en étamines dans les deux premiers verticilles, avec le pattern correspondant : sépales carpelloïdes (voire carpelles), étamines, étamines, carpelles

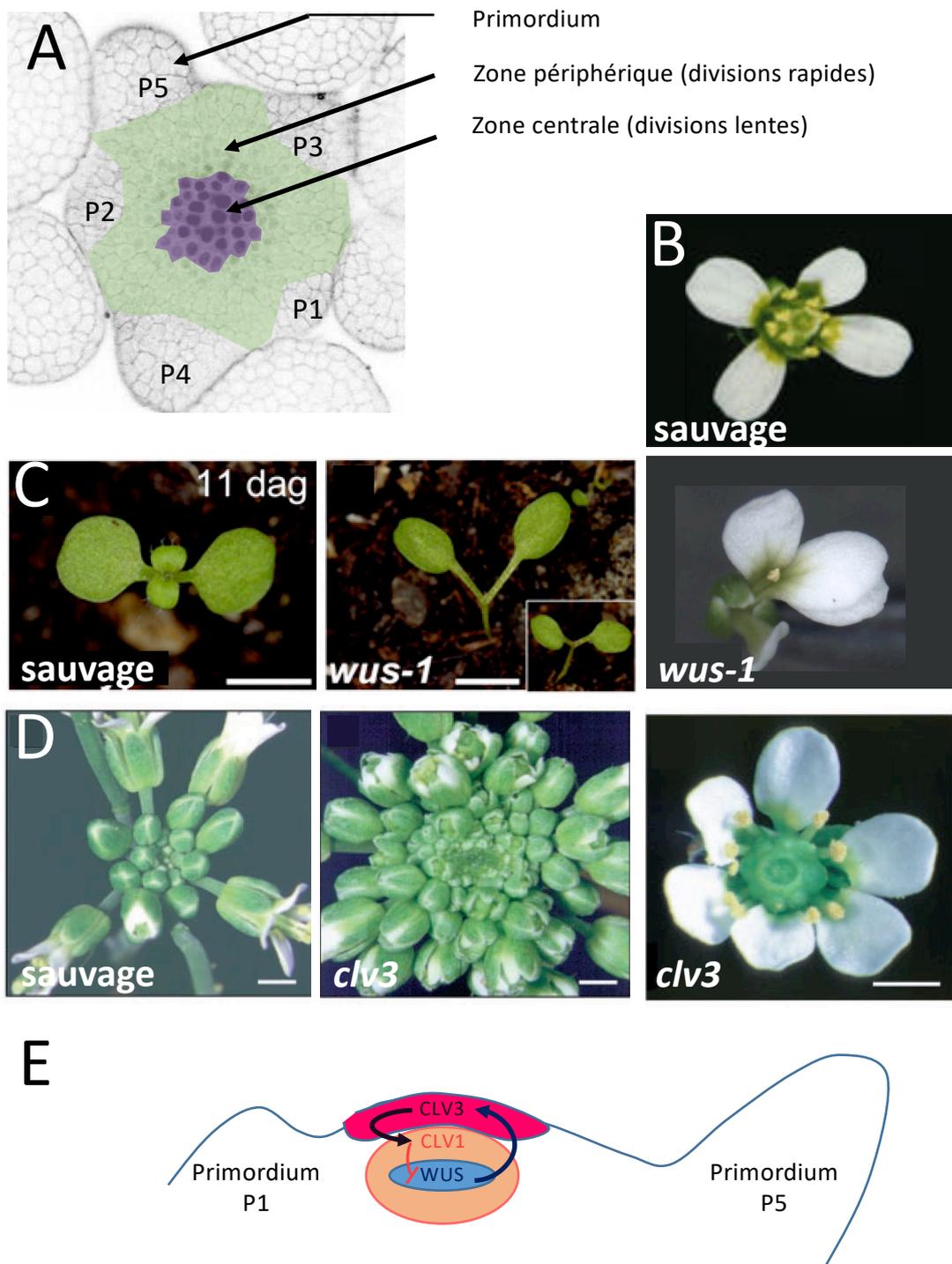


Figure 2 : Boucle de rétrocontrôle *CLV/WUS* à l'origine du déterminisme des cellules souches dans les méristèmes apicaux caulinaires.

- (A) Microscopie confocale d'un méristème d'inflorescence (image de Carlos Galvan-Ampudia). La zone centrale est identifiée grâce au domaine d'expression de *CLV3*. Les primordia en cours de formation sont numérotés du plus jeune au plus ancien, de P1 à P5.
- (B) Fleur d'*Arabidopsis* sauvage montrant 4 pétales et 4 étamines.
- (C) Effet de la mutation du facteur pro-méristématique *WUS* (selon Graf et al., 2010).
- (D) Effet de la mutation de *CLV3*, un inhibiteur de *WUS* (selon Fletcher et al., 1999).
- (E) Schéma résumant les interactions entre *CLV1*, *CLV3* et *WUS* dans la régulation du nombre de cellules souches.

(Figure 3-B ; Bowman et al., 1989 ; Coen and Meyerowitz, 1991). Similairement, les mutants de classe B montrent, dans les verticilles 2 et 3, une conversion des pétales en sépales et des étamines en carpelles et donc le motif floral suivant : sépales, sépales, carpelles, carpelles (Figure 3-C ; Bowman et al., 1989 ; Tröbner et al., 1992).

Enfin, les mutants de classe C sont affectés au niveau des verticilles 3 et 4, avec un remplacement des étamines par des pétales et la production de nouveaux verticilles de sépales et pétales au centre de la fleur (Figure 3-D). Ils sont alors caractérisés par le motif (sépales, pétales, pétales) répété plusieurs fois chez *A. thaliana* où il est même possible d'observer l'apparition de fleurs dans la fleur (Bowman et al., 1989). Chez *A. majus*, le quatrième verticille montre, quant à lui, une nature variable entre sépales, pétales et carpelles et contient de nouveau plusieurs nouveaux verticilles de pétales (Bradley et al., 1993). L'augmentation du nombre total de verticilles observé chez ces deux espèces suggère une persistance anormale de l'activité des cellules souches, aussi appelée perte de détermination du méristème.

Les auteurs ont conjointement proposé un modèle, dit modèle ABC, permettant d'expliquer la spécification des différents verticilles en fonction de la combinaison de l'expression de 3 classes de gènes (Figure 3-A). Ces travaux ont alors amorcé l'identification des gènes correspondant à ces 3 classes. En bref, de l'extérieur vers l'intérieur : (1) dans le premier verticille, les gènes de classe A (*APETALA1* et *APETALA2* chez *A. thaliana*) exprimés seuls déterminent la formation des sépales, (2) dans le deuxième verticille, l'expression simultanée des gènes de classe A et de classe B (*PISTILLATA* et *APETALA3* chez *A. thaliana*), induisent la formation des pétales, (3) dans le troisième verticille, les gènes de classe C (*AGAMOUS* chez *A. thaliana*), associés aux gènes de classe B, spécifient la formation des étamines, (4) dans le quatrième verticille, les gènes de classe C déterminent, seuls, la différenciation des primordia en carpelles. Pour compléter le modèle et correspondre parfaitement aux phénotypes observés, deux autres suppositions sont nécessaires. Tout d'abord, les gènes de classe A et C se répriment mutuellement, ce qui détermine deux territoires d'expression non chevauchants, séparés par une frontière précise entre les deux. En effet, dans les mutants de classe A, la formation d'organes carpelloïdes dans le premier verticille ne peut s'expliquer, selon ce modèle, qu'avec l'extension de l'expression des gènes de classe C à l'ensemble des quatre verticilles et non plus seulement aux verticilles 3 et 4. Et d'autre part, la classe C est aussi responsable de l'arrêt du méristème et limite la production d'organes floraux aux 4 verticilles classiques. Au contraire, dans un mutant *ag*, le méristème fonctionne indéfiniment, ce qui induit la formation de verticilles surnuméraires voire un phénotype de fleur dans la fleur. Il a plus tard été démontré que c'est bien l'expression d'*AGAMOUS* dans le centre du méristème floral qui induit la consommation des cellules souches du méristème ; cela en inhibant l'expression de *WUSCHEL*, le facteur de transcription essentiel au maintien des cellules souches déjà évoqué plus haut dans le cas des méristèmes végétatifs et floraux (Lenhard et al., 2001).

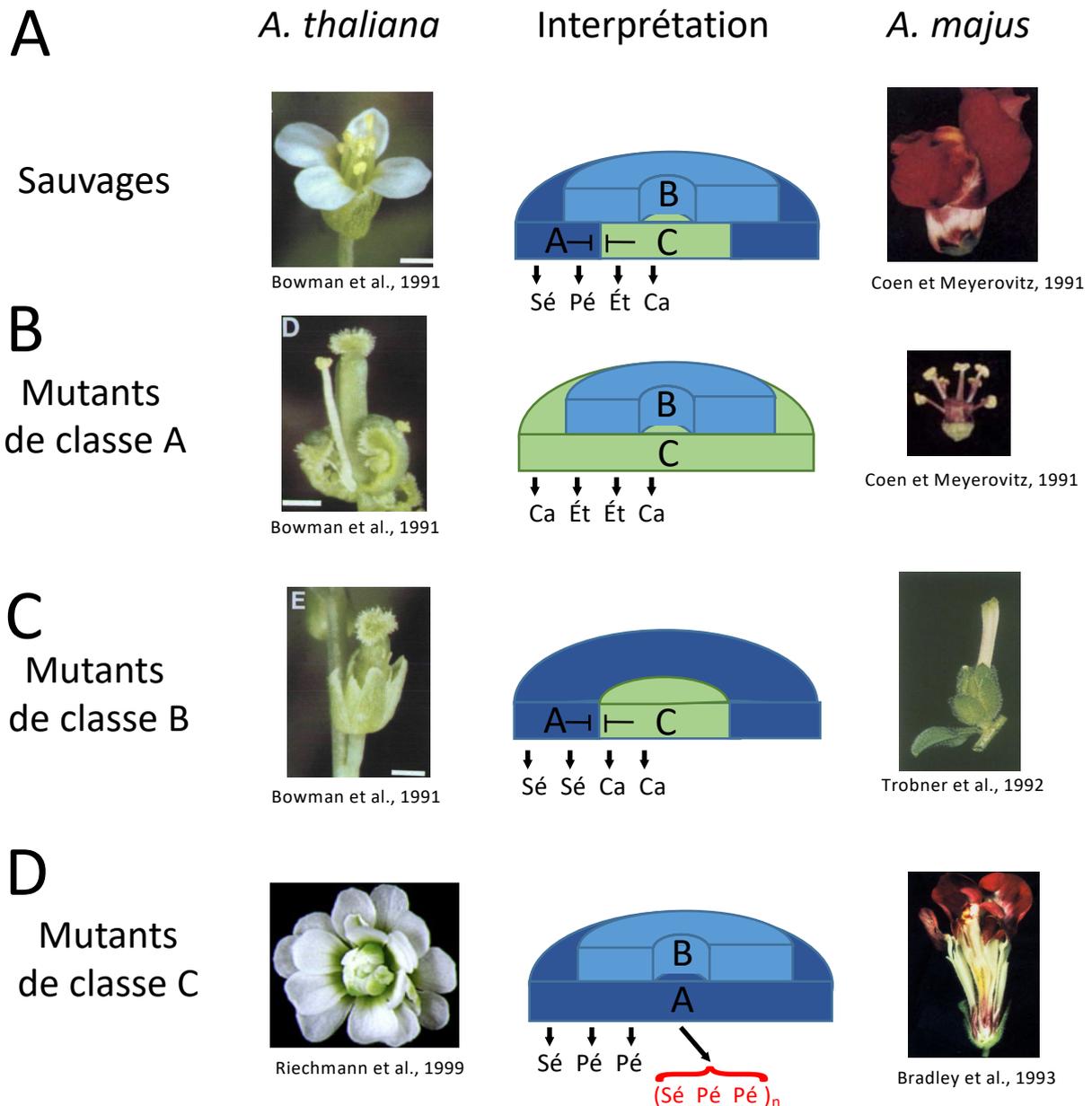


Figure 3 : Modèle ABC proposé dans les années 1990.

Phénotypes et schémas interprétatifs des différents mutants homéotiques du développement floral.

- (A) Fleurs d'*Arabidopsis thaliana* et d'*Antirrhinum majus* sauvages.
- (B) Les mutants de classe A sont caractérisés par la transformation des sépales en carpelles et des pétales en étamines.
- (C) Les mutants de classe B le sont par la transformation des étamines en sépales et des pétales en organes carpelloïdes.
- (D) Les mutants de classe C montrent une transformation des étamines en pétales et la production indéterminée de sépales et pétales à la place des carpelles.

L'observation de ces mutants a amené la communauté scientifique à proposer le modèle ABC : la différenciation en sépales est induite par la présence des gènes de classe A exprimés seuls alors que la différenciation en pétales l'est en présence des gènes de classe A associés aux gènes de classe B. De la même manière, les gènes de classe C, exprimés conjointement aux gènes de classe B permettent la différenciation en étamines alors que seuls, ils induisent la formation des carpelles et la terminaison du méristème.

4. Mécanismes moléculaires et évolution du modèle en modèle ABCE

À l'exception d'*APETALA2*, qui appartient à la famille des facteurs de transcription de type AP2/EREBP, spécifique des plantes (Riechmann and Meyerowitz, 1997), les gènes du modèle ABC appartiennent tous à la très grande famille des facteurs de transcription de type MADS. Représentée chez presque tous les eucaryotes étudiés (Gramzow et al., 2010), le nom de cette famille découle de ses gènes-types identifiés chez différents organismes modèles : *MCM1* chez la levure, *AGAMOUS* chez l'Arabette, *DEFICIENS* chez le muflier, *SRF* chez l'homme). Plus précisément, ils appartiennent à la sous-famille MADS de type II, dont les protéines contiennent, en plus de leur domaine de liaison à l'ADN de type MADS, deux domaines, notés I et K, responsables des interactions protéine-protéine et un domaine C de transactivation (Ng and Yanofsky, 2001). Ils sont ainsi capables de former des complexes se fixant au niveau de séquences d'ADN cibles plus ou moins spécifiques, d'y recruter des cofacteurs capables de moduler le niveau de transcription des gènes cibles et donc de modifier le programme développemental des cellules dans lesquelles ils sont exprimés (Figure 4-D ; Pajoro et al., 2014 ; Wuest et al., 2012).

Curieusement, la mutation simultanée de ces trois classes de gènes n'empêche pas la formation de fleurs, mais les fleurs produites sont entièrement constituées d'organes semblables à des feuilles carpelloïdes (Bowman et al., 1991). Cela supporte ainsi l'hypothèse de Goethe, formulée plus de 200 ans auparavant et qui supposait que les fleurs sont des feuilles modifiées. En revanche, l'expression ectopique d'*AG*, d'*AP1* ou d'*AP3* et *PI* n'est pas suffisante pour induire la transformation complète des feuilles de rosettes en organes floraux (Krizek and Meyerowitz, 1996 ; Mandel and Yanofsky, 1995 ; Mizukami and Ma, 1992), ce qui indique l'existence probable d'une autre classe de gènes nécessaires à la différenciation en fleur. C'est ainsi que les gènes de classe E (*SEPALATA1/2/3/4*) ont été identifiés et ajoutés au modèle. L'expression de ces gènes de classe E, dont la fonction est redondante, est nécessaire à l'induction de chaque type d'organe et en particulier à celle des pétales, étamines et carpelles. En effet, leur perte de fonction induit la formation de fleurs uniquement composées de sépales, comme cela peut être observé chez le triple mutant *sep1/sep2/sep3* (Figure 4-A). Les protéines de classe E sont capables d'interagir physiquement avec l'ensemble des autres protéines MADS du modèle et sont nécessaires à leur fonctionnalisation (Pelaz et al., 2000 ; Puranik et al., 2014). Ainsi, la formation de complexes tétramériques composés de différentes combinaisons de protéines de classe E et de protéines codées par les trois autres classes de gènes permet l'interaction avec différentes séquences régulatrices cibles spécifiques (Ó'Maoiléidigh et al., 2013 ; Pajoro et al., 2014 ; Puranik et al., 2014 ; Wuest et al., 2012). Cette observation a valu au modèle le nom de « quartet model », permettant d'expliquer la spécification des quatre types d'organes floraux en fonction de la composition des tétramères formés dans chaque verticille (Figure 4-C). Par exemple, l'interaction entre

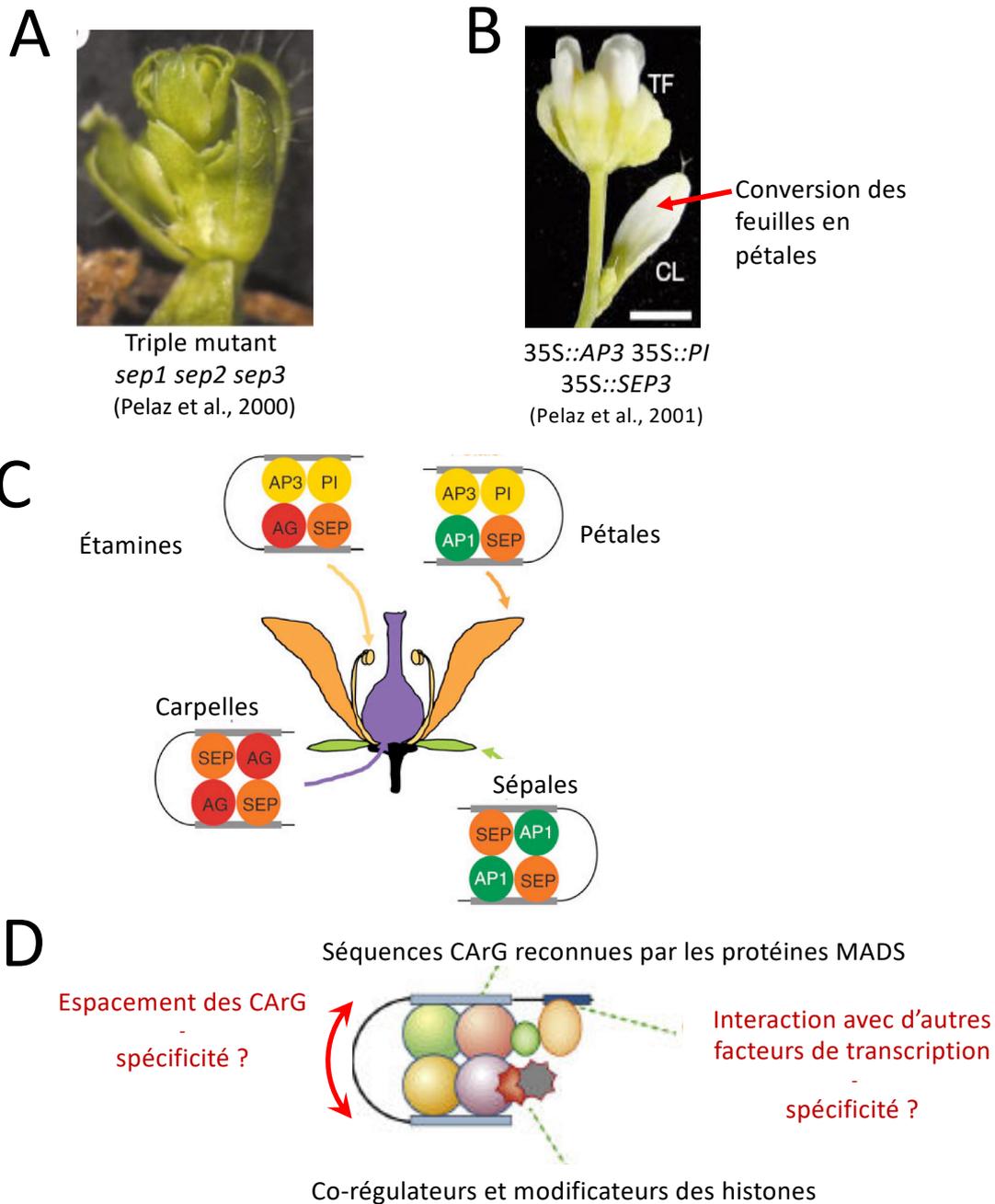


Figure 4 : Le « quartet model », un modèle moléculaire de l'interaction entre les différentes protéines MADS du modèle ABCE lors du déterminisme des organes floraux.

- (A) Phénotype floral d'une *A. thaliana* triple mutante *sep1/sep2/sep3* ; ses fleurs sont seulement composées de sépales.
- (B) Conversion des feuilles en pétales chez une *A. thaliana* surexprimant *SEP3*, *AP3* et *PI*.
- (C) Différentes combinaisons de protéines MADS s'organisent en tétramères spécifiant, en fonction de la composition des tétramères, les séquences-cibles reconnues et donc le type d'organe produit (figure extraite de Soltis et al., 2007).
- (D) Une fois fixé aux séquences régulatrices cibles CARG, le tétramère de MADS peut recruter des corégulateurs qui modifieront le niveau d'expression de ses cibles. Il a été proposé que chaque tétramère tire sa spécificité de l'espacement entre les séquences CARG ou de son interaction avec d'autres facteurs de transcription, entre autre (figure extraite de Yan et al., 2016).

AP1 (classe A), PI (classe B), AP3 (classe B) et SEP3 (classe E) permet la spécification des pétales dans le verticille 2. Leur expression constitutive est suffisante pour transformer les feuilles de rosette d'*Arabidopsis* en pétales (Figure 4-B), alors que la conversion n'est pas vraiment effective sans l'expression de *SEP3* (Pelaz et al., 2001).

5. Le modèle ABC est-il généralisable à toutes les Angiospermes ?

Proposé pour la première fois en 1991 et par la suite étudié en détails chez plusieurs espèces, le modèle ABC fournit à présent un cadre général pour comprendre les mécanismes conduisant à la détermination des organes floraux chez les Angiospermes. Il est globalement bien conservé parmi les plantes à fleurs, avec des variations qui peuvent facilement expliquer les différentes organisations florales observées dans la nature. En particulier, la présence d'un périanthe dont les deux verticilles sont composés de tépales pétaloïdes chez certaines monocotylédones comme la tulipe ou le lys, ainsi que chez de nombreuses Angiospermes basales comme *Amborella*, peut être expliquée par l'expression des gènes de classe B dans les trois premiers verticilles (Soltis et al., 2007). Ainsi, les deux verticilles extérieurs partagent alors le programme développemental des pétales. Compte tenu que le dernier ancêtre commun des Angiospermes devait lui aussi avoir un périanthe indifférencié, il semblerait que la présence de deux verticilles d'apparences distinctes, l'un de sépales et l'autre de pétales, soit liée à une restriction ultérieure des gènes de classe B aux deuxième et troisième verticilles et ainsi à la mise en place du modèle ABC tel qu'il est décrit dans les livres scolaires.

Cependant, si le modèle est conservé, certains de ses acteurs ont pu être substitués dans certaines des espèces étudiées. Le plus frappant a certainement été décrit chez *A. majus* et *P. hybrida*, où l'essentiel de la fonction A n'est pas conférée par des facteurs de transcription de la famille AP2 ou MADS, mais par un microARN (Cartolano et al., 2007). Ainsi, les mutations des orthologues d'*AP1* (*SQUAMOSA* chez *A. majus*) ou d'*AP2* (*LIPLESS1* et *LIPLESS2* chez *A. majus*) n'induisent pas de phénotype de type « mutant de classe A » mais plutôt une perte de l'identité du méristème floral ou la mise en place de pétales mal formés, respectivement (Huijser et al., 1992 ; Keck et al., 2003). Ainsi, les seuls mutants arborant un phénotype de classe A identifiés chez *A. majus* ou *P. hybrida* s'avèrent être soit des mutants de gain de fonction de la fonction C soit des mutants n'exprimant plus le microARN169 responsable de l'inhibition de *NUCLEAR FACTOR-YA (NF-YA)*, un activateur de l'expression des gènes de classe C. Par conséquent, la formation d'un périanthe dépendrait potentiellement plus de l'absence d'expression des gènes de classe C dans les verticilles 1 et 2 que de l'expression de gènes de classe A dans ces mêmes verticilles. Le modèle serait alors plutôt de la forme (A)BC, où les orthologues des gènes de classe A d'*Arabidopsis* sont davantage responsables de la transformation du méristème végétatif en méristème floral que de la détermination des organes eux-mêmes (Soltis et al., 2007) et où différentes voies d'inhibition de la fonction C auraient été sélectionnées en fonction des clades.

De façon plus anecdotique, certains événements de duplication de gènes ou de génomes peuvent également être à l'origine de différences mineures dans le modèle en fonction de l'espèce. En effet, les trajectoires évolutives aléatoires suivies par les différentes copies des gènes résultantes, et leur néofonctionnalisation ou leur sous-fonctionnalisation ultérieure, ont pu simplement conduire à la rétention de la fonction par des paralogues différents en fonction de l'espèce étudiée. Ainsi, la fonction n'est pas forcément portée par l'orthologue exact du gène correspondant chez *A. thaliana*. Cela peut être illustré par l'histoire évolutive de la fonction C dans différentes espèces, telle que décrite par Causier et al. (2005). Par exemple, chez *A. thaliana*, la fonction C est portée par *AGAMOUS* tandis que ses paralogues *SHATTERPROOF1* et *2 (SHP1/2)* ne sont pas impliqués dans le déterminisme des organes floraux à proprement parler mais sont nécessaires au développement des ovules et du fruit ; ils sont parfois appelés gènes de classe D dans le modèle ABCDE du développement floral (Favaro et al., 2003). Inversement, chez *A. majus*, le phénotype mutant de type C, avec conversion homéotique des organes reproducteurs en organes de périanthe et en fleurs indéterminées, n'est observé que lorsque l'orthologue de *SHATTERPROOF* (i.e. *PLENA*) est muté, tandis que l'orthologue d'*AG* (i.e. *FARINELLI*) est responsable de la bonne formation des étamines (Davies et al., 1999).

La disponibilité de séquences de génome pour de plus en plus d'espèces non-modèles permet désormais des études approfondies permettant de mieux comprendre ces échanges de fonctions entre gènes et donc de mieux appréhender le développement des plantes non modèles, en particulier des espèces à fortes importances agricole, ornementale ou économique.

6. Effecteurs/cibles des gènes homéotiques

Dans chacun des verticilles, la formation de tétramères de MADS spécifiques contrôle l'expression de nombreux gènes, permettant la modulation de voies de signalisation et d'induction qui leur sont tout autant spécifiques et ayant pour conséquence l'induction de la formation de sépales pour les complexes composés de protéines de classe A et E, de pétales pour les complexes composés de protéines de classe A, B et E, d'étamines pour les complexes composés de protéines de classe B, C et E et de carpelles pour les complexes composés de protéines de classe C et E. L'ensemble des MADS sont capables de reconnaître des séquences cibles appelées « CARG boxes » dont la séquence consensus correspond à CC[A,T]₆GG ou CC[A,T]₇G (Huang et al., 1993 ; West et al., 1997). Étant donné que les tétramères reconnaissent généralement deux motifs CARG à proximité l'un de l'autre, il a été proposé que leur fixation coopérative sur ces deux sites induise une courbure de l'ADN favorisant l'accès aux corégulateurs (Figure 4-D ; Mendes et al., 2013). Cette courbure a notamment été observée dans le cas de l'interaction entre *SEP3* et *SEEDSTICK (STK)* lors du développement des carpelles (Melzer et al., 2009). Cependant, alors que des analyses d'immunoprécipitation de chromatine (ChiP-seq) révèlent que les différentes protéines MADS possèdent à la fois des cibles spécifiques et des cibles

partagées, les mécanismes assurant la spécificité de ces complexes sont encore en cours d'investigation (Yan et al., 2016). Plusieurs hypothèses, soutenues par des expériences réalisées sur certains de ces complexes et non mutuellement exclusives, ont été proposées. Tout d'abord l'espacement entre les deux motifs pourrait être source de spécificité de chaque complexe puisque des homodimères de protéines SEP1, SEP2 ou SEP3 ont chacun des affinités différentes en fonction de la distance entre les motifs CArG *in vitro* (Figure 4-D ; Jetha et al., 2014). D'autre part, alors que des co-immunoprécipitations utilisant AP1 comme proie ont permis d'identifier des facteurs de transcription appartenant à d'autres familles de gènes, comme AUXIN-RESPONSE FACTOR 2 (ARF2) et SQUAMOSA PROMOTER BINDING PROTEIN LIKE 8 (SPL8 ; Smaczniak et al., 2012) et qu'il a été montré que STK et SHATTERPROOF1/2 interagissent physiquement avec BELLRINGER dans le développement des ovules (Brambilla et al., 2007), il est aussi possible que ces interactions permettent d'augmenter la spécificité des complexes avec leurs gènes cibles (Figure 4-D). Enfin, il est aussi possible que, bien qu'étant capables de reconnaître des séquences régulatrices communes, différents tétramères induisent différents effets sur l'expression du gène cible correspondant en recrutant des co-régulateurs et des remodeleurs de chromatine différents (Figure 4-D ; Smaczniak et al., 2012), ce qui mènerait à des effets opposés. Ainsi, les gènes MADS sont en amont d'un ensemble de voies de signalisation induisant la différenciation en types d'organes : ils ciblent généralement divers facteurs de transcription et intermédiaires des voies de signalisation ou de réponse aux hormones (Ó'Maoiléidigh et al., 2013 ; Pajoro et al., 2014 ; Wuest et al., 2012). Par exemple, Wuest et al. (2012) ont montré que AP3 et PI interagissent et activent les promoteurs de nombreux facteurs de transcription liés au développement des pétales (en particulier *RABBIT EARS* (Takeda et al., 2004) et *UNUSUAL FLOWER* (Durfee et al., 2003)) et des étamines (comme *SPOROCTELESS* ; Liu et al., 2009), tout en reconnaissant les promoteurs et en inhibant l'expression de gènes du développement des carpelles, comme *SHATTERPROOF1/2* (Favaro et al., 2003) et *CRABS-CLAW* (Bowman and Smyth, 1999) entre autres. La modulation de l'expression de ces gènes permettra de modifier de nombreuses voies de signalisation permettant la différenciation des organes des verticilles 2 et 3 en pétales et étamines respectivement.

7. Différenciation des organes, l'exemple des pétales

Une fois les primordia de pétales initiés, la morphogénèse des pétales se fait en plusieurs étapes consécutives, avec tout d'abord une étape de croissance par prolifération cellulaire (croissance mitotique) puis une étape de croissance par élongation (croissance post-mitotique) et une étape de différenciation cellulaire, avec en particulier la formation de cellules coniques sur la portion distale de leur face adaxiale. Le contrôle de chacune de ces étapes est alors essentiel pour obtenir des pétales ayant une taille et une forme bien définies. Il a été montré que certains gènes exprimés dans les pétales en cours de formation, comme *JAGGED*, *AINTEGUMENTA* ou *ARGOS*, promeuvent la prolifération cellulaire

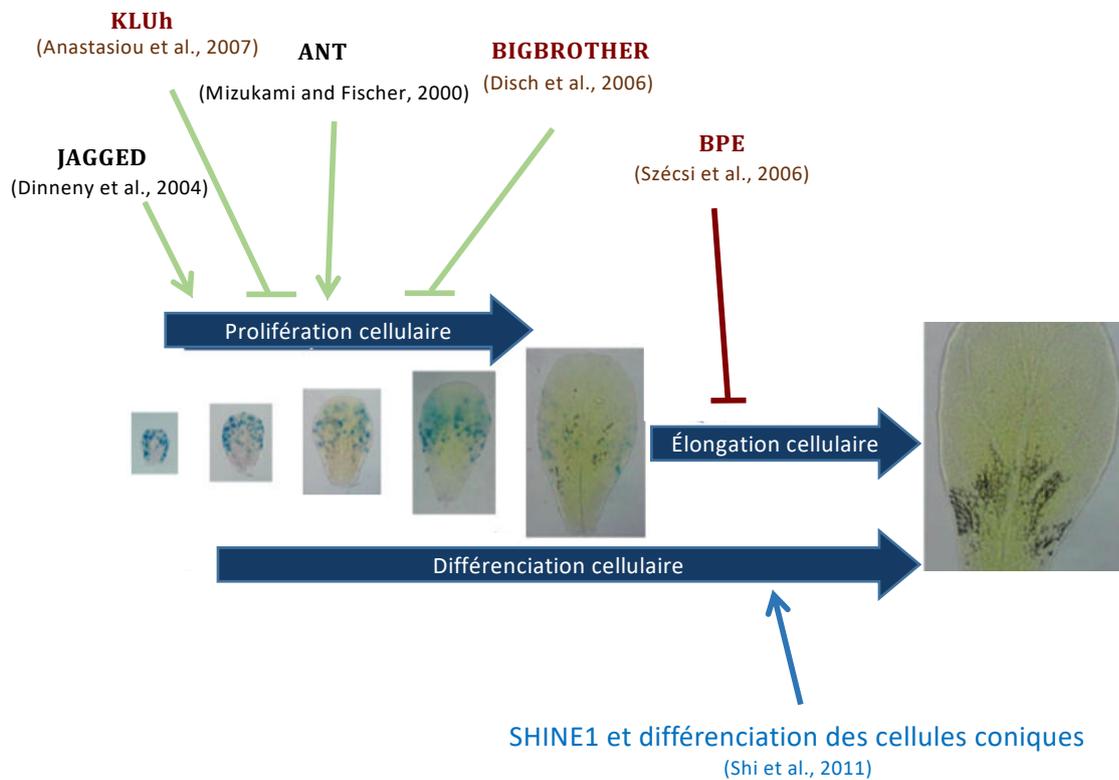


Figure 5 : Quelques gènes impliqués dans la croissance et la différenciation du pétale lors du développement floral.

Une première étape de prolifération cellulaire est contrôlée par la balance entre l'expression de gènes promouvant les divisions cellulaires, comme par exemple *JAGGED* et *AINTEGUMENTA*, et celle de gènes inhibant ces divisions, comme par exemple *KLUH* et *BIGBROTHER* notamment. L'élongation cellulaire est quant à elle inhibée par la présence de *BPE*, ce qui permet d'obtenir la forme finale du pétale. Chez *A. thaliana*, peu de gènes de la différenciation cellulaire ont été identifiés spécifiquement chez le pétale; un des seuls exemples correspond à *SHINE1* qui promeut la synthèse de la cuticule et permet la différenciation des cellules coniques.

Images extraites de Breuninger et Lenhard (2010).

(Figure 5 ; Dinneny et al., 2004 ; Hu et al., 2003 ; Mizukami and Fischer, 2000), alors qu'au contraire *BIGBROTHER* et *KLU* l'inhibent (Figure 5 ; Anastasiou et al., 2007 ; Disch et al., 2006). Ensuite, il semblerait que la transition entre la phase de division cellulaire vers celle de l'expansion cellulaire soit contrôlée par la protéine NAP, dont l'expression serait activée par AP3/PI (Sablowski and Meyerowitz, 1998), puis que *BIGPETALp* (BPEp), un facteur de transcription hélice-boucle-hélice basique préférentiellement exprimé dans les pétales, régulerait finalement la croissance des pétales en limitant l'expansion cellulaire (Figure 5 ; Brioudes et al., 2009 ; Szécsi et al., 2006). En effet, une voie de signalisation initiée par la phytohormone jasmonate régule positivement l'expression de BPEp qui, en interagissant avec un facteur de transcription de la famille des ARF, contrôle l'expansion cellulaire et la croissance post-mitotique du pétale (Brioudes et al., 2009 ; Varaud et al., 2011). Ainsi, ce serait une juste expression de l'ensemble de ces gènes qui permettrait de former des pétales avec une taille homogène au sein de chaque fleur. Ces quelques gènes ne représentent qu'une petite fraction des gènes supposés contrôler la croissance du pétale, sujet à propos duquel les connaissances restent encore limitées. Enfin, la modulation des gènes de différenciation cellulaire permet l'acquisition structurelle et fonctionnelle de cet organe spécialisé dans l'attraction des pollinisateurs. C'est le cas notamment de *SHINE1*, dont l'expression semble activée par AP1 lors du développement des pétales (Pajoro et al., 2014) et induit l'expression des gènes de biosynthèse des crêtes de cuticules présentes uniquement sur les cellules coniques de la face adaxiale des pétales (Shi et al., 2011).

Pour conclure, le développement floral, en permettant la formation de quatre organes avec des fonctions distinctes, met en jeu un ensemble de voies de signalisation distinctes, originales et complexes ; celles-ci gagnent encore à être mieux comprises, que ce soit d'un point de vue moléculaire ou d'un point de vue évolutif, afin de mieux comprendre la variété de structures et de formes que l'on peut rencontrer parmi les fleurs d'aujourd'hui.

III. La fleur double, une anomalie florale très appréciée et étudiée

Si, depuis Linné (1735), la classification des Angiospermes et la définition de leurs familles reposent essentiellement sur la structure de la fleur (nombre d'organes, symétrie...), c'est parce que cet ensemble de caractères est à la fois très stable et conservé dans l'évolution à l'échelle d'une famille et à la fois suffisamment variable pour être un critère de classification entre les familles.

Le strict contrôle du nombre de pétales dans la fleur est certainement lié à son rôle essentiel dans la reproduction de la plante : il faut un nombre de pétales suffisant pour attirer les pollinisateurs tout en laissant suffisamment d'espace pour que les organes reproducteurs se développent puis soient accessibles, et que l'espèce puisse perdurer et potentiellement conquérir de nouveaux espaces. Cependant, parmi les espèces appréciées comme plantes d'ornement dans les jardins, on observe que certaines ont un nombre de pétales considérablement plus élevé que leur équivalent sauvage ;

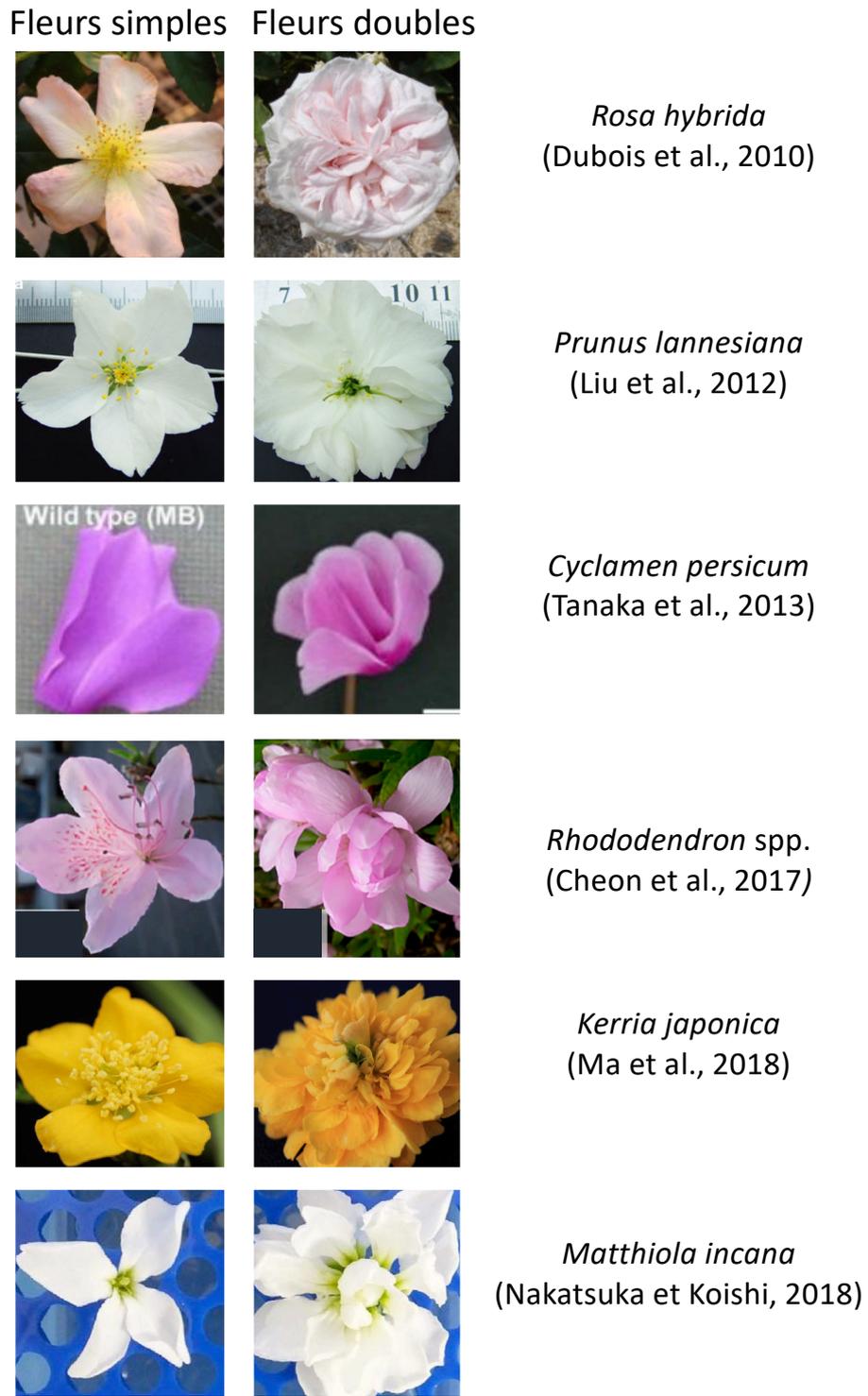


Figure 6 : Exemples de fleurs ornementales étudiées pour l'apparition du caractère « fleur double ».
 Le phénotype est apparu indépendamment dans chaque espèce.
 Les photos sont extraites des articles correspondants, à l'exception de la fleur simple de rosier.

c'est le cas par exemple des variétés commerciales de rosier, prunier, corète du Japon, camélia, ou cyclamen (Figure 6). Ainsi, lors du processus de domestication des plantes ornementales, l'homme les a sélectionnées pour une augmentation importante du nombre de pétales, rendant les fleurs plus attractives à ses yeux : ce caractère est appelé « fleur double ».

1. Description des fleurs doubles

Jean-Jacques Rousseau dans ses *Lettres écrites de la Montagne* (1764) les décrit ainsi : « Ce mot de fleur double ne marque pas dans le nombre des pétales une quelconque duplication, mais une multiplication quelconque. Soit que le nombre des pétales devienne double, triple, quadruple, etc. ». En d'autres termes, alors que les fleurs sauvages, dites « fleurs simples », présentent généralement dans leur deuxième verticille entre 3 et 5 pétales selon le type de fleur (Figures 6 et 7), leurs homologues à « fleur double » peuvent être décrites comme des fleurs avec au moins deux fois plus d'organes pétaloïdes. À ce stade, ces organes sont simplement définis comme des organes plats et colorés, indépendamment de leur position dans la fleur. Les fleurs de rosier sont un exemple bien connu de fleur double : si les fleurs des rosiers sauvages n'offrent que cinq pétales, les roses modernes que nous apprécions dans nos jardins présentent beaucoup plus de pétales, souvent plus de 50 (Figure 7) et parfois jusqu'à 200. De même, les coquelicots sauvages que l'on trouve dans les champs présentent tous quatre pétales, tandis que les variétés à fleurs « doubles » égalaient elles aussi nos jardins. Similairement, les espèces de *Prunus* cultivées dans les vergers comme arbres fruitiers (pour leurs pêches, leurs prunes, leurs abricots, leurs amandes...) présentent toutes des fleurs simples, mais des espèces proches appréciées comme arbres d'ornement produisent des floraisons doubles impressionnantes, comme c'est le cas des cerisiers du Japon (*Prunus japonica*). Ainsi, de nombreuses variétés à fleurs doubles ont été domestiquées au sein de différentes familles, avec potentiellement des mécanismes de formation de ces pétales surnuméraires distincts en fonction des variétés.

2. Origine développementale des pétales des fleurs doubles

L'excès de pétales peut provenir de différentes modifications de l'architecture florale, dont les mieux caractérisées sont décrites dans les paragraphes suivants.

Le mécanisme le plus courant consiste en la conversion homéotique des organes d'un autre verticille en pétales. Le plus souvent ce sont les étamines qui sont transformées en pétales, comme cela est illustré chez *R. chinensis* sur la Figure 7, où une diminution du nombre d'étamines produites est observée en même temps que l'augmentation du nombre de pétales. C'est aussi le cas dans une grande variété de familles telles que les Rosaceae (*Kerria japonica* (Ma et al., 2018) ou *Prunus lannesiana* (Liu et al., 2012)), les Ericaceae (*Rhododendron* spp. (Cheon et al., 2017)), les Solanaceae (*Petunia hybrida* (Krol and Chua, 1993)), les Gentianaceae (*Gentiana scabra* (Nakatsuka et al., 2015)),

R. chinensis cv 'Single Pink' *R. chinensis* cv 'Old Blush' *R. chinensis* cv 'SB'

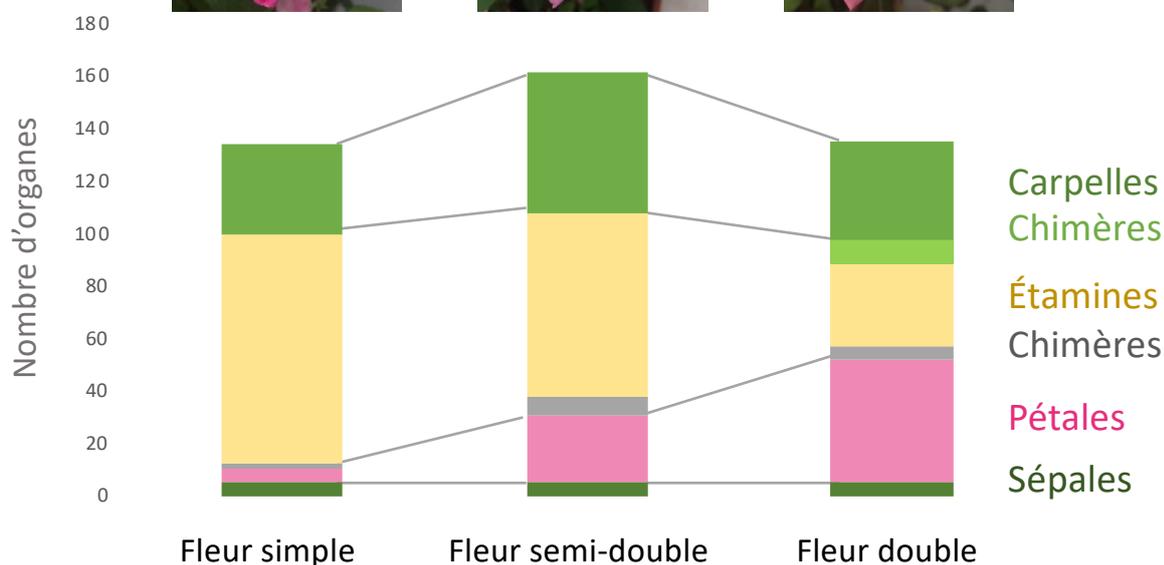


Figure 7 : Morphologie d'une fleur double; exemple de différentes variétés de rosier.

Une fleur de rosier simple contient généralement 5 sépales, 5 pétales, environ 80 étamines et 20 carpelles. Une fleur semi-double contient davantage de pétales, ici autour de 25, et moins d'étamines, autour de 50, qu'une fleur simple. Il y a eu conversion homéotique d'une partie des étamines en pétales. Une fleur très double peut contenir de 50 à plus de 200 pétales. Dans certains, cas non représentés ici, la formation de pétales surnuméraires se fait aussi grâce à une augmentation de l'activité méristématique, et donc du nombre total d'organes. Dans les cas extrêmes, non représentés ici, il est possible d'observer chez certaines roses très doubles, la conversion de l'ensemble des étamines et carpelles en pétales ou en fleurs dans la fleur, ce qui rend la fleur stérile.

les Primulaceae (*Cyclamen persicum* (Tanaka et al., 2013)) ou les Liliaceae (*Lilium longiparum* (Akita et al., 2008)), pour ne citer que les plus étudiées. Dans d'autres cas, en particulier chez les Ranunculaceae et certains hortensias, ce sont les sépales, colorés, qui attirent les pollinisateurs alors que les pétales sont normalement réduits ou absents. Les variétés à « fleur double » présentent alors une conversion homéotique des étamines et des pétales en sépales pétaloïdes (Galimba et al., 2012).

Aussi, la simple augmentation du nombre total d'organes est assez fréquente et peut être combinée avec le mécanisme précédent. Elle peut être due soit à une activité méristématique renforcée (Brand et al., 2000 ; Fletcher et al., 1999) soit à l'apparition de verticille(s) additionnel(s) par rapport à la fleur simple et capable(s) de se différencier en pétales. Dans ce dernier cas, comme chez le lilas par exemple (Dadpour et al., 2011), on parle de néohétérotopie.

Enfin, certains cas trouvent leur explication hors du cadre du développement floral *stricto sensu*, chez des espèces où ce que l'on appelle la « fleur » d'un point de vue ornemental correspond en réalité à un regroupement de fleurs d'un point de vue biologique. Chez les Asteraceae par exemple, de nombreuses petites fleurs sont disposées en capitules et présentent une différenciation stricte : les fleurs périphériques stériles produisent une et une seule ligule, long pétale unique, jouant le rôle de pétale pour l'inflorescence et les fleurs centrales, très étroites, produisent essentiellement des organes reproducteurs. Ensemble, elles reproduisent l'organisation d'une fleur classique. Les variétés à « fleur double » correspondent alors en réalité à des capitules comportant davantage de fleurs à ligules que les variétés à « fleur simple ». Similairement, chez les bougainvillées, 3 bractées en forme de pétales sont disposées en pseudo fleur qui attire les pollinisateurs et contiennent 3 véritables petites fleurs en leur centre. Les bougainvillées à « fleur double » présentent en fait une augmentation du nombre de bractées plutôt qu'une augmentation du nombre de pétales.

3. Propagation des plantes à fleur double

Les formes à « fleur double », bien que très représentées dans les jardins d'ornement, sont généralement absentes du milieu naturel. Dans *De la variation des animaux et des plantes sous l'action de la domestication* (1868), Charles Darwin qualifiait déjà ces modifications de « monstruosité [et potentiellement] de causes de stérilité », notamment en raison de la perte ou de la réduction du nombre d'organes reproducteurs. En effet, l'accroissement du nombre de pétales peut réduire l'accessibilité des organes reproducteurs par les pollinisateurs. La conversion des étamines en pétales peut aussi fortement réduire, voire annuler, la capacité de la plante à produire du pollen. Certaines plantes sélectionnées par l'homme ont même des fleurs sans aucun organe reproducteur : ceci peut notamment être observé chez le lys, où les variétés avec le plus de pétales n'ont même plus de carpelles (Benedito et al., 2004). Rousseau, dans ses *Lettres écrites de la Montagne*, en fait alors une catégorie à part « lorsque les pétales trop multipliés font disparaître les étamines et avorter le germe,

alors la fleur perd le nom de fleur double et prend celui de fleur pleine. On voit par-là que la fleur double est encore dans l'ordre de la nature, mais que la fleur pleine n'y est plus, et n'est qu'un véritable monstre ». Il signifie ici que ces modifications du motif floral causent une réduction drastique de la fitness de la plante ; et les quelques mutants spontanés produits naturellement sont généralement contre-sélectionnés dans la nature à cause de leur capacité réduite à se reproduire sexuellement. Ainsi, si les plantes à fleurs doubles sont sélectionnées comme ornements, elles sont essentiellement tributaires de l'homme pour leur maintien et propagation, souvent par multiplication végétative. Il est intéressant de noter que l'une des seules espèces pour laquelle des fleurs doubles existent à l'état sauvages, la petite pervenche *Vinca minor*, est connue pour se reproduire principalement par multiplication végétative (Wang et al., 2011). Dans ce cas, la fitness de la plante ne dépend que peu de la capacité à produire des graines et donc de la qualité de ses organes reproducteurs, ce qui explique son expansion en milieu naturel.

4. Origine des fleurs doubles

Des variétés à fleurs doubles existent au sein de nombreux genres de plantes à fleurs (*Rosa*, *Dianthus*, *Tulipa*, *Lilium*, *Camellia*, pour n'en citer que quelques-uns). Dans chacun de ces genres, les variétés à fleur double proviennent de variétés ancestrales présentant des fleurs simples, ce qui indique que le caractère « fleur double » a été domestiqué plusieurs fois, de manière strictement indépendante. Il est très probable que ces anomalies soient apparues par des mutations spontanées, éventuellement à plusieurs reprises au fil des siècles, et aient été sélectionnées et maintenues par des jardiniers en raison de leur apparence attrayante. Par exemple, des rosiers à fleurs doubles étaient déjà cultivés dans les jardins de Pompéi, pendant l'Antiquité (Illustrations p 15). Plus tard, les jardiniers et les obtenteurs ont continué à sélectionner des variétés présentant un nombre toujours croissant de pétales. Cette sélection des individus a été faite sur des phénotypes et sans connaissance des gènes à l'origine de ces phénotypes d'intérêt, ouvrant un champ entier à la recherche de ces déterminants.

L'étude d'autres processus de domestication a montré que des changements drastiques d'apparence pouvaient être liés à la modification d'un nombre réduit de gènes maîtres. Ainsi, la mutation de quelques gènes suffit pour transformer la téosinte, ancêtre du maïs très difficilement comestible, en une plante ressemblant beaucoup plus à la source de nourriture majeure à l'échelle mondiale qu'est le maïs (Dong et al., 2019 ; Stitzer and Ross-Ibarra, 2018). Par son apparition relativement fréquente chez différentes espèces, on peut supposer que le phénotype « fleur double » est également contrôlé par un seul ou un très petit nombre de gènes. C'est notamment le cas chez les rosiers chinois, où le passage d'une fleur simple à une fleur double semble ne dépendre que d'un seul locus, avec une mutation dominante (Spiller et al., 2010), mais aussi chez certaines ipomées et gentianes japonaises, où il semblerait que ce soit également un seul locus, avec une mutation récessive

cette fois-ci (Tableau 1 ; Nakatsuka et al., 2015 ; Nitasaka, 2003 ; Tasaki et al., 2017). Cependant, ce n'est pas nécessairement la même mutation pour l'ensemble des espèces à fleur double domestiquées. En effet, comme l'apparition de ce phénotype a été indépendante chez chaque espèce, les mécanismes moléculaires à son origine sont supposés être distincts, bien qu'ayant des conséquences similaires.

5. Importance d'*AGAMOUS* dans la formation des fleurs doubles

Jusqu'à récemment, les mécanismes moléculaires à l'origine du passage d'un phénotype « fleur simple » sauvage à un phénotype « fleur double » restaient inconnus. Cependant, la disponibilité récente de cartes génétiques précises et de nouvelles séquences de génomes complets a permis de mieux étudier les déterminants génétiques de cette surproduction de pétales.

Le phénotype des espèces ornementales à « fleur double » ressemble aux fleurs d'*Arabidopsis* mutantes pour le gène *AGAMOUS* (Bowman et al., 1989), où les étamines sont converties en pétales et où une nouvelle fleur est produite à la place des carpelles. Dès le début des années 2000, Enrico Coen, scientifique essentiel dans la théorisation du modèle ABC, proposa que les fleurs doubles ornementales puissent avoir leur fonction C affectée (Coen, 2001). Dès lors, les homologues d'AG ont été les premiers gènes étudiés pour comprendre le déterminisme des fleurs doubles et de nombreuses études ont alors identifié une non-fonctionnalisation ou une dérégulation d'AG chez celles-ci (Tableau 1). La première démonstration du lien entre la modification d'expression de *RcAG* et la formation de la fleur double chez le rosier a été réalisée dans l'équipe en 2010 (Dubois et al., 2010). Dans cette étude, des expériences d'hybridation *in situ* ont révélé qu'une restriction du domaine d'expression de *RcAG* vers le centre du méristème floral est associée à la formation de fleur possédant plus de pétales. Cette étude a également montré que cette restriction du domaine d'expression de *RcAG* peut être observée chez des rosiers provenant de deux centres de domestication indépendants, indiquant qu'elle a dû être sélectionnée au minimum par deux fois au cours des processus de sélection et amélioration variétales. Les auteurs ont ainsi proposé un modèle dit de « Sliding border » permettant d'expliquer les différentes formes de fleur double chez le rosier par le déplacement de la frontière A/C vers le centre du méristème (Dubois et al., 2010). De même, les orthologues d'AG apparaissent mutés chez certaines variétés à fleurs doubles d'ipomée du Japon (Nitasaka, 2003), de Renonculacées (Galimba et al., 2012) et de gentiane japonaise (Nakatsuka et al., 2015). D'autre part, l'analyse de différents cyclamens indique que *CpAG1*, un des orthologues d'*AGAMOUS* chez le cyclamen, n'est exprimé que dans les méristèmes floraux des fleurs simples et non dans ceux des fleurs doubles (Tanaka et al., 2013). Enfin, un épissage alternatif conduisant à la suppression des motifs C-terminaux impliqués dans l'activation de la transcription et les interactions protéine-protéine chez l'orthologue d'AG n'a été observé que chez les variétés à fleurs doubles de *Prunus lannesiana* et non chez les variétés à fleurs simples (Liu et al., 2012). Étonnamment, la mutation à l'origine de cet épissage

Tableau 1 : Résumé des études sur le lien entre les gènes de classe C et le phénotype « fleur double ».

Espèce	Gène muté	État des gènes de classe C	Transmission	Références
<i>Rosa chinensis</i>	Inconnu	Expression de <i>RcAG</i> restreinte	Dominante	Dubois et al., 2010
<i>Kerria Japonica</i>	<i>KjAG</i>	Insertion d'un élément transposable dans <i>KjAG</i> et formation d'une protéine tronquée	Supposée récessive	Ma et al., 2018
<i>Prunus lannesiana</i>	<i>PrseAG</i>	Saut d'un exon dans <i>PrseAG</i> mais mutation non décrite	N.D. ^d	Liu et al., 2012
<i>Rhododendron</i> spp.	<i>RmAG1</i>	Expression réduite de <i>RmAG1</i> avec parfois des INDELS	N.D. ^d	Cheon et al., 2016
<i>Camellia japonica</i>	Inconnu	Expression réduite de <i>CjAG</i> ^c	N.D. ^d	Sun et al., 2014
<i>Cyclamen persicum</i>	Inconnu	Perte d'expression de <i>CpAG1</i>	N.D. ^d	Tanaka et al., 2013
<i>Ipomoea nil</i>	<i>DP/FAR/AG</i>	Insertion d'un élément transposable dans <i>DP</i> induisant sa perte de fonction	Récessive	Nitasaka et al., 2003
<i>Gentiana scabra</i>	<i>CsAG1</i> ^a	Insertion d'un élément transposable et diminution de l'expression de <i>CsAG1</i>	Récessive	Nakatsuka et al., 2012 Tasaki et al., 2016
<i>Silene dioica</i>	<i>SDM1</i> ^b	Deux INDELS indépendants dans <i>SDM1</i> dans deux variétés distinctes à fleur double	Supposée récessive	Ingle et al., 2015
<i>Thalictrum thalictroides</i>	<i>ThtAG1</i>	Insertion d'un élément transposable dans <i>ThtAG1</i> menant à la délétion du domaine K	Supposée récessive	Galimba et al., 2012
<i>Lilium formolongi</i>	Inconnu	Corrélation entre le niveau d'inhibition de <i>LeLAG1</i> et le phénotype	N.D. ^d	Akita et al., 2011

^a Malgré son nom, *CsAG1* est un orthologue de *PLE*.

^b *SDM1* est un gène de classe C mais l'article ne précise pas s'il appartient à la sous-famille de *PLE* ou d'*AG*.

^c Parmi les camélias, il existe aussi des variétés à fleurs anémones correspondant à un type particulier de fleurs doubles qui montrent une expansion du domaine d'expression de *CjAG* dans les pétales intérieurs, différents des pétales extérieurs classiques.

^d Non Déterminé.

différent n'est pas décrite dans cet article, suggérant que celle-ci n'est pas aisément identifiable.

Le rôle clé des orthologues d'*AGAMOUS* dans la formation des fleurs doubles est confirmé par des expériences de génétique inverse chez quelques plantes ornementales pour lesquelles des protocoles de transformation génétique existent. En effet, l'inhibition artificielle des orthologues d'*AGAMOUS* est suffisante pour augmenter le nombre de pétales de ces espèces, sans pour autant perdre l'organisation générale de la fleur. Chez le cyclamen, l'expression d'un dominant négatif inhibant *CpAG1* a conduit à la conversion des étamines en pétales (Tanaka et al., 2013) et, chez le pommier, l'inhibition de *MdAG* par interférence à ARN (ARNi) a entraîné la production de fleurs très doubles (Klocko et al., 2016). Enfin, une conversion homéotique d'étamines en pétales a également été obtenue en supprimant transitoirement par VIGS (pour « Virus Induced Gene Silencing ») l'expression d'orthologues d'*AG* chez la gentiane japonaise (Nakatsuka et al., 2015), le rosier (Ma et al., 2015), le pétunia (Noor et al., 2014) ainsi qu'une Renonculacée (Galimba et al., 2012). La plupart des phénotypes induits par ces inhibitions ressemblent aux fleurs doubles commerciales, ce qui confirme que la restriction de la fonction C pourrait être l'une des clés de la domestication de la fleur double.

Cependant, une mutation directe d'un orthologue d'*AGAMOUS* n'a pas été détectée chez toutes les espèces pour lesquelles des variétés à fleurs doubles existent. C'est notamment le cas pour le rosier, le cyclamen, le pétunia ou le lys, pour lesquels le déterminant génétique de la double fleur est toujours recherché (Tableau 1). Chez *Rosa chinensis*, plusieurs cartographies génétiques indépendantes, réalisées avec des parents variés, ont identifié un large intervalle contenant la mutation (Crespel et al., 2002 ; Debener, 1999 ; Spiller et al., 2010) à l'origine de l'augmentation du nombre de pétales, mais les orthologues d'*AG* ne se trouvent pas dans cet intervalle. Dans ce cas, le déterminant de la double fleur pourrait agir en amont de *RcAGAMOUS* et en réguler l'expression, contrôlant par conséquent le nombre de pétales. Cela pourrait être également vrai chez les nombreuses espèces pour lesquelles une sous-expression d'*AGAMOUS* a été effectivement observée, sans pour autant avoir détecté une quelconque mutation à proximité de celui-ci qui pourrait expliquer cette sous-expression (Akita et al., 2011 ; Sun et al., 2014 ; Tanaka et al., 2013). Il est alors probable que la mutation à l'origine de la formation de fleurs doubles affecte plutôt un (ou des) gène(s) capable(s) de réguler l'expression d'*AGAMOUS* et que des études supplémentaires sont nécessaires à l'identification de ce (ou ces) gène(s).

IV. Le rosier : une plante modèle

Depuis les années 80, *A. thaliana* s'est avérée une plante facile d'utilisation comme modèle d'étude grâce à son temps de génération court et sa faible demande en surface de culture et a été progressivement adoptée par l'ensemble de la communauté de biologie végétale (Meinke et al., 1998). Cet engouement s'est ensuite accru par effet boule de neige au fur et à mesure que plus de ressources étaient produites : séquence complète du génome (Bevan et al., 2001), banques de

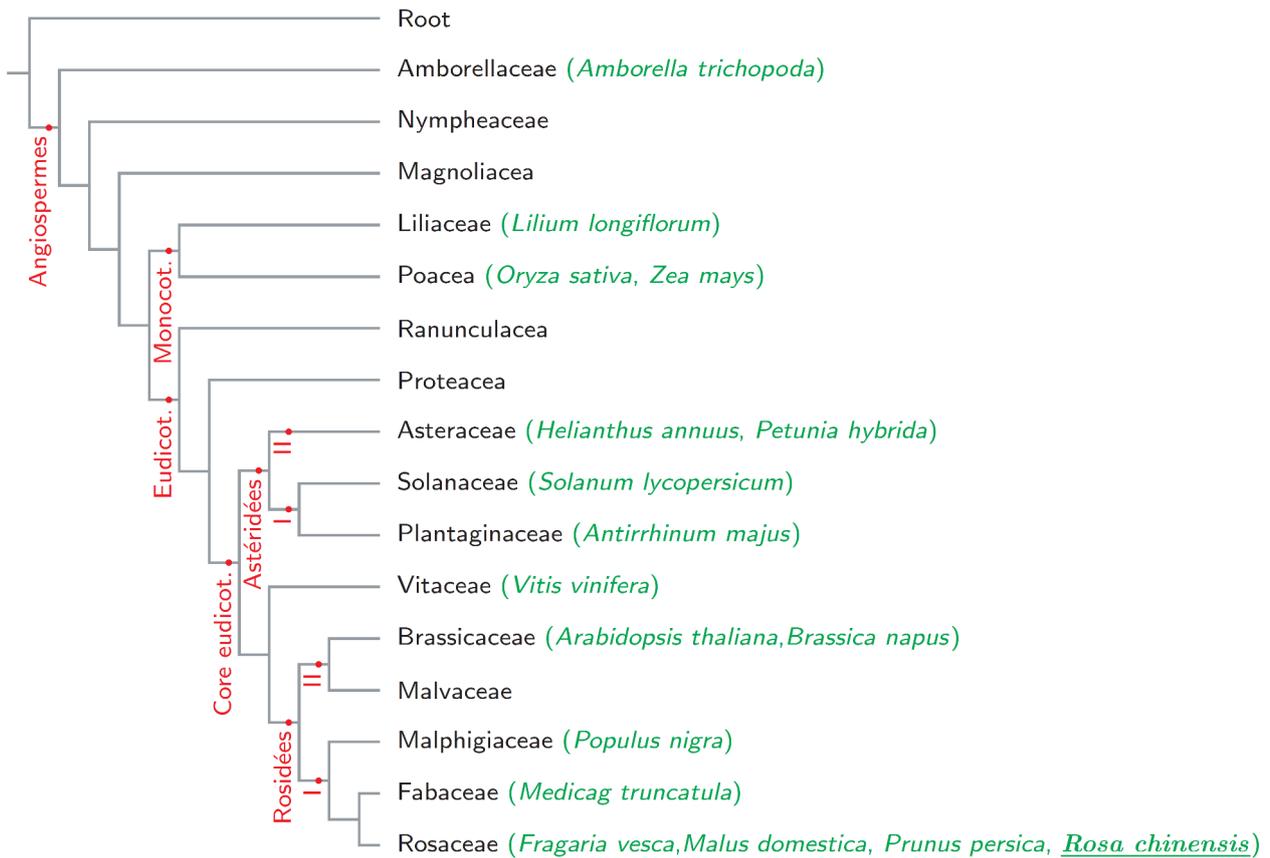


Figure 8 : Arbre phylogénétique remplaçant les différentes espèces modèles utilisées en laboratoire.

Adapté de l'arbre phylogénétique des Angiospermes disponible sur

<http://www.mobot.org/MOBOT/research/APweb/>.

Les principales espèces modèles de chaque famille sont indiquées entre parenthèses. Il est ainsi possible de noter la multiplicité de ces espèces et de leur phénotype.

mutants (Scholl et al., 2000), données d'expression des gènes (Winter et al., 2007), notamment. Seulement, la simple comparaison morphologique entre des espèces aussi variées qu'une arabette des dames (*Arabidopsis thaliana*, Brassicaceae), un coton (*Gossypium* spp., Malvaceae), un pêcher (*Prunus persica*, Rosaceae) ou un maïs (*Zea mays*, Poaceae) montre la nécessité d'utiliser des modèles différents pour mieux comprendre la mise en place de certains caractères, absents chez l'arabette. En particulier, comment analyser les étapes de la mise en place du bois chez une herbacée ? Comment étudier la coloration des pétales chez une plante qui produit des fleurs blanches ? Ou encore comment analyser le développement d'un fruit charnu chez une plante développant des siliques ? Enfin, comment vérifier la conservation d'un mécanisme à l'échelle de plusieurs espèces, comme cela a été le cas par exemple pour le modèle ABC, sans l'utilisation de différents modèles ?

Ainsi, afin d'élargir les connaissances à ces sujets, il est bon de pouvoir utiliser d'autres modèles, provenant de différentes familles, comme par exemple la luzerne (*Medicago truncatula*) pour les Fabaceae, le riz (*Oryza sativa*) et le brachypodium (*Brachypodium distachyon*) pour les Poaceae, la tomate (*Solanum lycopersicum*), le pétunia (*Petunia axillaris*) et le tabac (*Nicotiana benthamiana*) pour les Solanaceae (Figure 8).

On observe ainsi souvent la présence de plusieurs modèles par famille, ce qui est certainement dû à la diversité des traits à étudier mais peut être aussi en partie lié à la concomitance des avancées techniques et des recherches sur ces modèles. Les Rosaceae, entre autres familles, n'échappent pas à la nécessité de mener les recherches sur un petit nombre d'espèces choisies comme modèles, avant de tenter de généraliser ces résultats au reste de la famille. Ainsi, sa position au sein des Rosaceae, les techniques de culture relativement aisées ainsi que les outils génétiques, moléculaires, transcriptomiques et biotechnologiques développés depuis les années 2000 ont permis de faire du rosier une plante modèle de premier choix pour étudier le développement floral (Bendahmane et al., 2013), ce qui sera développé dans les paragraphes suivants.

1. *Le rosier appartient aux Rosaceae*

Les Rosaceae sont une grande famille comprenant près de 3 000 espèces dans le monde, classiquement réparties en une centaine de genres (Thomas, *Petite Flore de France*, 2016). Malgré l'apparente hétérogénéité de cette famille qui contient à la fois des plantes herbacées et des plantes ligneuses, ses membres partagent généralement des caractéristiques florales communes comme la présence d'un réceptacle formé par la concrétion des sépales, pétales et étamines, un nombre important d'étamines et de carpelles, et l'insertion de ces carpelles à l'intérieur du réceptacle. Ses espèces sont réparties en 3 sous-familles, les Rosoideae, les Amygdaloideae et les Dryadoideae, cette dernière étant minoritaire en termes de nombre de genres et d'espèces cultivés (Figure 9 ; Shulaev et al., 2008). Au contraire, Rosoideae et Amygdaloideae contiennent de nombreuses espèces cultivées,

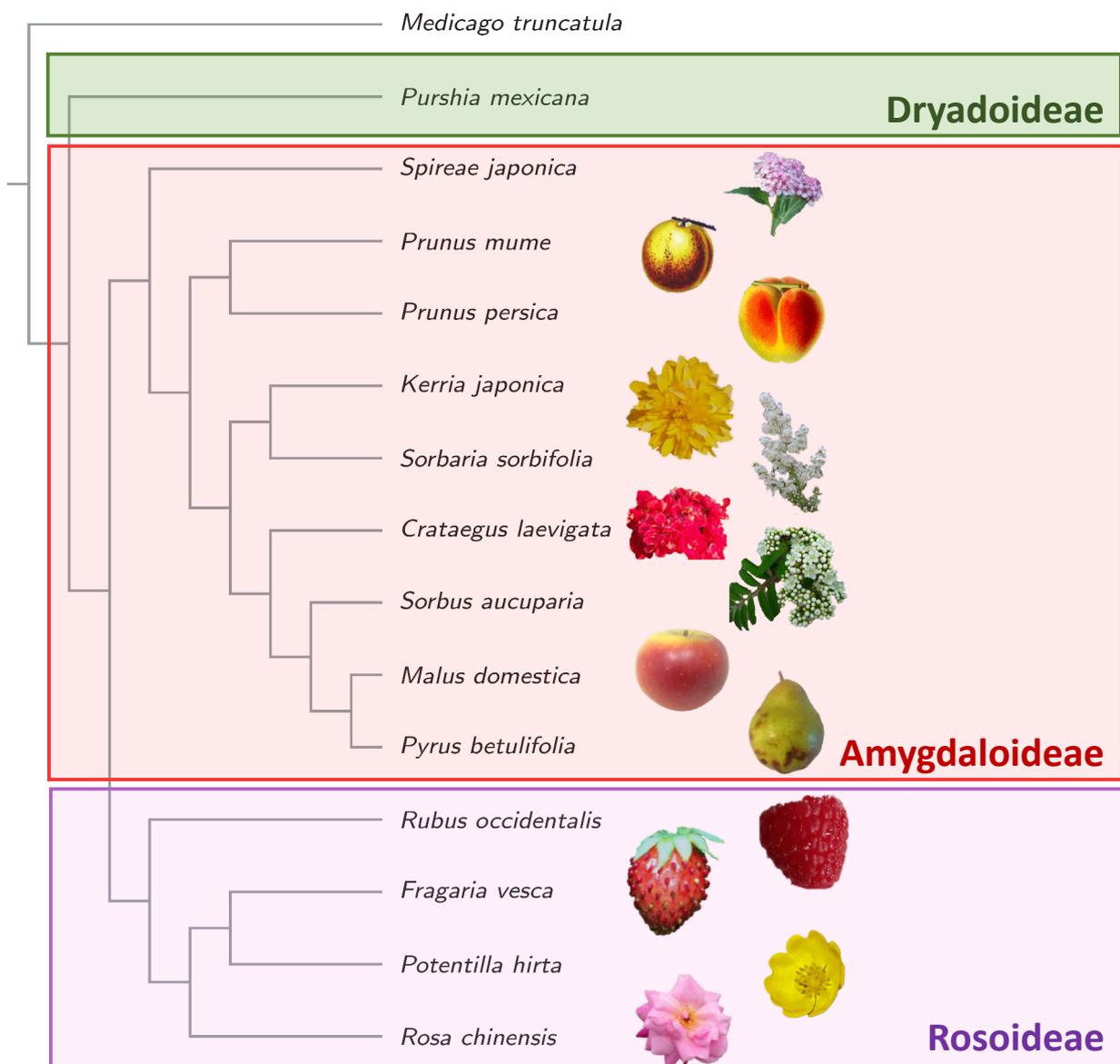


Figure 9 : Arbre phylogénétique des principales Rosaceae à importance économique.

Cet arbre se base sur l'un des derniers articles publiés (Xiang et al., 2017) où les auteurs ont analysé les séquences de plusieurs centaines de gènes nucléaires pour établir la phylogénie des Rosaceae. Le nom de la sous-famille des Amygdaloideae et la position de certaines des espèces au sein de cette sous-famille semblent encore aujourd'hui non consensuels.

Les photos de fleurs ou de fruits à côté de chaque espèce indiquent leur utilisation principale, ornementale ou agricole. Celles qui n'ont pas été prises par mes soins proviennent de la page Wikipedia correspondant à l'espèce citée.

notamment les fraisiers, rosiers et framboisiers au sein des Rosoideae et les pommiers, poiriers, amandiers, pêchers, abricotiers et cerisiers au sein des Amygdaloideae (Figure 9). La plupart de ces espèces sont des arbres cultivés pour leurs fruits et sont difficiles à étudier à cause de la durée de croissance végétative entre chaque génération.

Grâce à leur importance économique, trois principaux modèles ont été établis en recherche, chacun avec leurs avantages et inconvénients (Tableau 2 ; Shulaev et al., 2008). Le pommier, qui appartient aux Amygdaloideae, est certainement le plus étudié. Il possède un génome généralement diploïde à 17 paires de chromosomes (longueur haploïde : 750 Mb ; Tatum et al., 2005), qui a été séquencé d'abord en 2010 sous forme de 1 629 contigs ancrés sur une carte génétique (Velasco et al., 2010), puis en 2017 sous forme de 17 pseudo-chromosomes (Daccord et al., 2017). Des techniques de transformation génétique stable efficaces ont été mises au point, mais la majorité des variétés sont aussi auto-incompatibles et possèdent une longue période juvénile. Le pêcher, qui appartient lui aussi à la sous-famille des Amygdaloideae, est auto-compatible mais difficilement transformable génétiquement et possède un génome composé de 8 paires de chromosomes convenablement séquencées (The International Peach Genome Initiative et al., 2013). Enfin, le fraisier appartient aux Rosoideae et semble, quant à lui, avoir tout pour lui : un petit génome bien séquencé (Shulaev et al., 2011) avec 206 Mb et 7 paires de chromosomes, un temps de génération de seulement 10 à 16 semaines, l'existence de techniques de transformation génétiques efficaces et des espèces modèles et cultivars auto-compatibles. Seulement, contrairement aux deux modèles précédents, il correspond à un des rares genres de la famille des Rosaceae qui n'est pas ligneux, ce qui en fait une exception.

Il est important de noter que ces trois espèces modèles, à haute importance économique, ont été domestiquées et sont cultivées pour leurs fruits et non pour leurs fleurs, suggérant la nécessité d'utiliser un modèle différent pour l'étude du développement floral. Ainsi, le rosier, qui est une plante ornementale, ligneuse et pérenne et a été domestiquée pour des raisons esthétiques et cosmétiques, paraît être le candidat idéal (Tableau 2). D'autre part, sa proximité phylogénétique avec le fraisier (Raymond et al., 2018) pourrait permettre le transfert rapide de connaissances entre les deux genres et le passage d'un modèle à l'autre en cas de nécessité. De la même manière, la mise en commun de l'ensemble des données génomiques à propos des Rosaceae sur le GDR, pour Genome Database for Rosaceae (Jung et al., 2019), donne accès à de nombreuses et diverses ressources permettant d'élargir rapidement l'impact des recherches réalisées au départ sur une espèce.

2. *Le rosier, une plante domestiquée*

À l'heure actuelle, le genre *Rosa* comporte environ 200 espèces, incluant de nombreux polyploïdes. Les espèces actuellement cultivées sont le fruit de deux domestications indépendantes durant l'Antiquité : l'une en Chine (Section Chinenses) et l'autre dans l'espace

Tableau 2: Comparaison des caractéristiques des différentes Rosaceae modèles^a.

	<i>Malus x domestica</i>	<i>Prunus persica</i>	<i>Rosa chinensis</i>	<i>Fragaria vesca</i>	<i>Arabidopsis thaliana</i>
Taille du génome (haploïde)	750 Mb	280 Mb	560 Mb	206 Mb	157 Mb
Nombre de chromosomes	2n = 34	2n = 16	2n = 14	2n = 14	2n = 10
Type	Ligneuse	Ligneuse	Ligneuse	Herbacée	Herbacée
Durée entre 2 générations	4 à 8 ans	3 à 5 ans	< 1 an	10 à 16 semaines	6 à 8 semaines
Plantes par m² au sol	< 1	< 1	Une dizaine	Une centaine	Plusieurs centaines
Type de fruit	Pomme	Drupe	Cynorhodon (non comestible)	Réceptacle charnu portant des akènes	Silique (non comestible)
Reproduction végétative	Boutures	Boutures	Boutures	Stolons	Non
Auto-compatibilité	Dépend des variétés	Oui	Dépend des variétés	Oui	Oui
Transformation	Oui (<i>in vitro</i>)	Faiblement efficace (<i>in vitro</i>)	Oui (<i>in vitro</i>)	Oui (<i>in vitro</i>)	Oui (<i>in planta</i>)
Séquençage du génome	Pseudochromosomes disponibles	Pseudochromosomes disponibles (brouillon)	Pseudochromosomes disponibles	Pseudochromosomes disponibles (brouillon)	Pseudochromosomes disponibles
Cartes génétiques^b	Plus de 2 000 000 marqueurs	Plus de 100 000 marqueurs	Environ 2 500 marqueurs	Environ 96 000 marqueurs	Non utilisé

^a Adapté de Shulaev et al., 2008.

^b D'après <https://www.rosaceae.org/>

périméditerranéen (Sections Cinnamoneae et Galicanae). Durant cette période, des traits spécifiques ont été sélectionnés, comme par exemple la récurrence de la floraison, un parfum particulier, dit « senteur de thé », ou la résistance au froid chez les rosiers européens (Martin et al., 2001). Au 18^e siècle, les missionnaires et commerçants ont introduit l'espèce chinoise diploïde *Rosa chinensis* en Europe et, en particulier, les variétés 'Hume's Blush', 'Park's Yellow' et 'Old Blush' (Raymond, 1999). Leurs croisements avec les espèces locales sont à l'origine de la création des variétés de rosiers connues aujourd'hui. En effet, il est aujourd'hui admis que l'ensemble des rosiers modernes cultivés, ou *Rosa x hybrida*, dérive d'une dizaine d'espèces fondatrices, notamment des *Rosa chinensis*. Ils sont le produit d'une histoire de domestication complexe où les variations de niveau de ploïdie (généralement $4n$, avec des passages par $2n$ et $3n$) côtoient les hybridations interspécifiques et les introgressions en de multiples et indépendants événements. Cela a permis la combinaison de traits propres à chaque espèce ancestrale au sein des variétés modernes.

Ainsi, l'étude de l'histoire de la domestication et des processus de sélection menant aux rosiers modernes, en même temps que celle du déterminisme génétique des traits d'intérêts horticoles, est un sujet particulièrement intéressant pour les chercheurs d'aujourd'hui. La communauté scientifique a alors choisi une variété modèle pour ses études sur le genre *Rosa* : la variété *Rosa chinensis* cv 'Old Blush'. Ce choix s'est fondé sur trois principales raisons. Tout d'abord, celle-ci possède déjà de nombreux traits d'intérêts horticoles, dont le phénotype « fleur double » et la remontée de floraison, qui facilite l'étude du développement floral. D'autre part, contrairement à la majorité des rosiers commerciaux généralement tétraploïdes, 'Old Blush' possède un génome diploïde, ce qui évite une complexité génétique trop importante et facilite donc les études scientifiques. Enfin, elle a été fortement impliquée dans de nombreux croisements à l'origine des cultivars modernes (Bendahmane et al., 2013 ; Martin et al., 2001 ; Raymond, 1999 ; Raymond et al., 2018). Elle a donc très certainement transmis une partie de son génome et de ses caractéristiques phénotypiques à ces cultivars, ce qui signifie que les découvertes réalisées sur cette variété modèle devraient être aussi applicables aux variétés modernes.

En résumé, sa position intermédiaire dans l'histoire de la domestication des rosiers en fait un excellent modèle pour étudier cette histoire et en particulier les traits d'intérêts qui ont été sélectionnés (Raymond et al., 2018) et notamment le caractère « fleur double ». De nouveaux outils ont par conséquent été développés et améliorés par la communauté scientifique, ce qui facilite d'autant plus les analyses qui pourraient ensuite être transférées chez des espèces plus difficiles à étudier. Dans les prochains paragraphes, les outils nouvellement disponibles et les avantages du rosier, et en particulier de la variété *R. chinensis* cv 'Old Blush', comme organisme modèle seront discutés.

3. Culture et propagation du rosier

D'une manière générale, les rosiers sont des plantes faciles à cultiver en laboratoire, que ce soit en chambre de culture ou en serre. Certains rosiers, à l'image d'Old Blush', sont arbustifs et ne nécessitent qu'un faible volume de culture (environ 0,25 m³ par plante).

Le bouturage des rosiers, très aisé chez de nombreuses variétés, permet d'obtenir aisément des clones génétiques facilitant la récolte rapide en quantité suffisante de matériel frais et la répétition des expériences sur différentes plantes de même fond génétique. D'autre part, des fleurs peuvent être obtenues dès la première année après germination ou seulement 3 mois après un bouturage. Ceci, comparé à d'autres Rosaceae ligneuses (notamment le pommier), est un véritable avantage accélérant significativement la réalisation des expériences.

Parmi les variétés de rosiers commerciales, nombreuses sont celles qui fleurissent en continu du printemps jusqu'à l'automne, tant que les conditions extérieures sont favorables. Ce caractère, nommé « remontée de floraison » ou encore « floraison continue », s'oppose au caractère ancestral où la plante ne fleurit qu'une fois par an, sur une courte période, généralement au printemps. Il est largement apprécié des sélectionneurs et jardiniers puisqu'il permet d'embellir les jardins durant plusieurs mois par an mais aussi des chercheurs puisqu'il permet d'échantillonner de jeunes bourgeons tout au long de l'année. En effet, dans le cas des rosiers à floraison unique, l'expérimentateur ne peut récolter des bourgeons que durant une courte durée au printemps et, si la quantité de matériel n'est pas suffisante, il sera nécessaire d'attendre l'année suivante pour en obtenir davantage. Ainsi, la floraison continue d'Old Blush' en chambre de culture est un vrai atout pour l'étude du développement floral chez les rosiers et ouvre potentiellement une généralisation à d'autres Rosaceae ne fleurissant qu'annuellement, comme le cerisier par exemple.

Un des obstacles principaux à l'utilisation du rosier comme plante modèle en laboratoire est certainement la présence chez certaines espèces, y compris 'Old Blush', d'une auto-incompatibilité élevée (Debener and Mattiesch, 1999), limitant fortement les croisements.

4. Disponibilité de nombreuses cartes génétiques

Un sondage mené en 2012-2013 par la *Rose Hybridizer Association* et l'université A&M au Texas a montré qu'à la fois obtenteurs et jardiniers amateurs avaient des attentes élevées concernant plusieurs caractéristiques du rosier telles que la résistance à diverses maladies, le parfum de la fleur et le nombre de fleurs produites (<https://roses.tamu.edu/research/rose-surveys/>). D'autres caractères importants sont la période de floraison, l'absence d'aiguillons, la couleur et la forme de la fleur. Aussi, de nombreux croisements ont été réalisés pour identifier les régions du génome portant les caractères correspondants, et potentiellement les gènes majeurs impliqués (Crespel et al., 2002 ; Debener and Mattiesch, 1999 ; Linde et al., 2006). Pour résumer, une cartographie génétique utilise la fréquence

des recombinaisons méiotiques pour estimer la distance entre des loci. Ces fréquences de recombinaisons sont calculées au sein d'une population issue d'un croisement entre deux individus parents suffisamment éloignés génétiquement pour distinguer l'origine de chaque locus : plus la taille de la population est grande, plus la carte sera précise. Les loci peuvent quant à eux être définis soit par des marqueurs moléculaires (séquence détectée par hybridation ou par PCR, pour « Polymerase Chain Reaction »), soit par leur association avec un trait phénotypique. À partir des distances entre loci, on peut inférer leur ordre sur le chromosome, et dessiner une « carte génétique » des chromosomes. Chez le rosier, plus de six cartes génétiques indépendantes ont été développées depuis les années 90. Elles sont fondées sur l'analyse moléculaire de 50 à 170 descendants en fonction des projets et utilisent des parents aussi variés que *Rosa multiflora*, *Rosa wichurana* ou des variétés de rosiers hybrides. Les marqueurs utilisés, eux aussi variés, ont pu être des marqueurs RAPD (« Random Amplified Polymorphic DNA »), AFLP (« Amplified Fragment Length Polymorphism ») ou SSR (« Simple Sequence Repeat »). La mise en place de ces cartes a été possible en particulier grâce à la facilité de culture de petites populations de rosiers, y compris en plein champ. Cependant, les rosiers nécessitant une plus grande surface de culture qu'une Arabette ou une autre plante herbacée, le nombre de descendants semés et analysés, et donc la précision des cartes génétiques correspondantes, restent généralement plus faibles que pour d'autres plantes modèles.

Quatre de ces cartes (Debener and Mattiesch, 1999 ; Hibrand-Saint Oyant et al., 2007 ; Linde et al., 2006) ont été unifiées par Spiller et al. (2010). La carte consensus ainsi obtenue (Figure 10) porte 597 marqueurs génétiques et est dorénavant une référence pour la recherche des déterminants génétiques à l'origine des caractères phénotypiques cités précédemment, et notamment du phénotype « fleur double » dont le déterminant génétique a été localisé sur une portion du « Linkage Group 3 » ou LG3. Cela a ainsi permis d'affirmer que *RcAG*, bien qu'impliqué dans la formation de la fleur double, ne pouvait pas être le déterminant de ce phénotype étant donné qu'il n'est pas sur le même chromosome (Dubois et al., 2010). Ainsi, chez le rosier, la génétique directe, qui consiste en la recherche du gène induisant un phénotype observé, est donc possible et adaptée à sa longue histoire de domestication (Figure 11).

5. Transformation génétique par *Agrobacterium*

Depuis les premières transformations stables de *Nicotiana benthamiana* en 1984 (Block et al., 1984) et d'*Arabidopsis thaliana* en 1986 (Lloyd et al., 1986), cette méthode de génétique inverse est régulièrement utilisée en laboratoire sur les espèces modèles et en cours de développement ou d'optimisation pour de nombreuses autres espèces. Cette approche, qui permet d'insérer un fragment d'ADN d'intérêt dans le génome de l'organisme étudié, est maintenant indispensable à l'analyse de la fonction des gènes, en permettant l'application d'une démarche scientifique complète. Ainsi, en

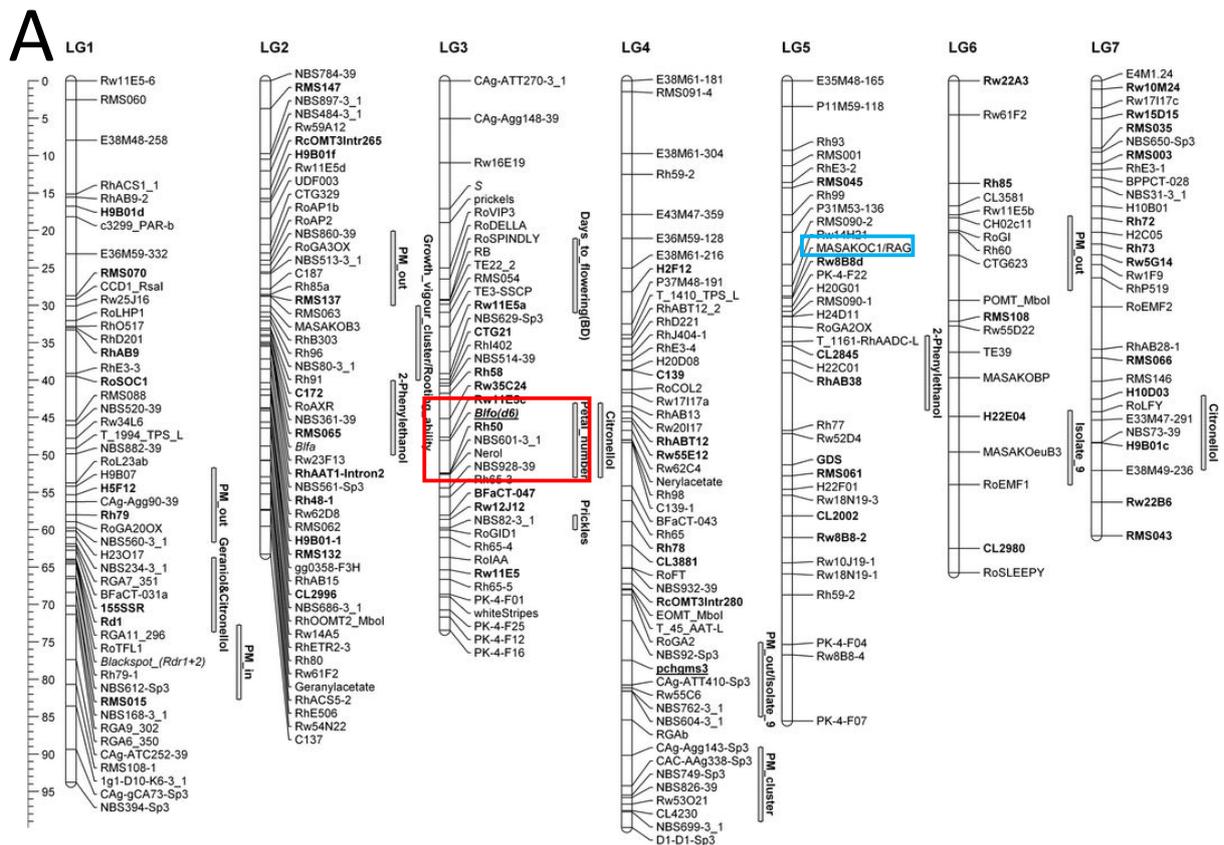


Figure 10 : Carte génétique et étude de traits d'intérêt chez le rosier : le cas de la fleur double.

- (A) Carte génétique consensus intégrant les marqueurs positionnés à partir de quatre populations distinctes. Les marqueurs en gras correspondent à ceux qui ont permis de lier les quatre différentes cartes entre elles. La nouvelle carte porte 597 marqueurs et montre les intervalles génétiques contenant les déterminants de traits d'intérêt comme le phénotype « fleur double », dont le déterminant a été localisé sur le chromosome 3 (intervalle en rouge), l'absence d'aiguillons, la résistance au mildiou, la production de certaines molécules de parfum ou la remontée de floraison. Extraite de Spiller et al., 2010.
- (B) Hybridation *in situ* extraite de Dubois et al. (2010), montrant une corrélation entre le nombre de pétales par fleur et la restriction du domaine d'expression du gène de classe C *RcAG*, localisé sur le chromosome 5 (encadré en bleu sur la carte). Les auteurs ont ainsi proposé que le déterminant de la fleur double chez *Rosa* devait intervenir en amont de *RcAG*.

inhibant un gène, on peut observer l'impact de sa perte de fonction sur le phénotype, puis en réinsérant une copie, on peut valider la restauration du phénotype et ainsi confirmer la fonction du gène étudié.

À l'heure actuelle, de nombreuses espèces de Rosaceae sont génétiquement transformables par application d'Agrobactéries sur des tissus placés en culture *in vitro* (Aldwinckle et al., 2009), y compris le rosier. Dans notre laboratoire, un protocole optimisé de transformation génétique d'Old Blush' est disponible et en cours de publication (Vergne et al., 2009 ; Vergne et al., non publié ; développé dans la section Matériel et Méthodes). Il permet un taux d'efficacité proche de 10 %, ce qui assure normalement de toujours obtenir plusieurs événements de transformation au cours d'une expérience (Figure 11). Diverses constructions peuvent être utilisées, permettant notamment la surexpression constitutive de gènes d'intérêt, de microARN artificiels ou d'ARN interférents pour diminuer l'expression endogène du gène cible ou de protéines fluorescentes sous le promoteur du gène étudié pour étudier son patron d'expression. D'autre part, un protocole d'édition de génome par CRISPR/Cas9 a été mis au point par Jeanne Loue-Manifel, Jérémy Just et Philippe Vergne (IE, IR et IR dans l'équipe respectivement). Ainsi, le rosier, est dorénavant une plante modèle à la pointe de la recherche et à jour au niveau des dernières avancées techniques.

6. Disponibilité de données d'expression des gènes

Comme pour beaucoup d'organismes, les premières données de séquençage produites sur le rosier ont été transcriptomiques, ce qui permet de concentrer les efforts de séquençage sur les gènes. Mon équipe d'accueil a été parmi les pionniers dans ce domaine en développant un atlas transcriptomique du rosier (Dubois et al., 2012). Pour cela, un séquençage 454 d'ADNc obtenus à partir d'un mélange de 13 différents tissus de *Rosa chinensis* cv 'Old Blush' a permis d'obtenir l'assemblage de 80 714 transcrits, aussi appelés clusters unigenes. En parallèle, un séquençage Illumina réalisé sur ces 13 tissus séparément indique le niveau d'expression de chaque gène dans chacun des tissus. À présent qu'une séquence génomique est disponible (Raymond et al., 2018), les clusters unigenes ne sont plus utilisés comme référence, mais les données de séquençage Illumina servent toujours pour l'exploration du niveau d'expression des gènes dans les différents organes du rosier.

D'autre part, deux ensembles de données de RNA-seq comparant le développement d'une fleur simple et d'une fleur double sont également disponibles dans l'équipe (données non publiées). Ils ont été réalisés sur de très jeunes bourgeons floraux, aux stades où apparaissent les primordia des sépales, pétales ou étamines. Le premier compare 'Old Blush' et la variété commerciale *Rosa chinensis* 'Single Pink', décrite comme un sport de celui-ci, c'est-à-dire un variant lié à une mutation somatique apparue spontanément et propagé végétativement (Phillips, 1988 ; Soule, 2009). À l'exception de ses fleurs simples, son apparence est identique à 'Old Blush', ce qui peut se décrire comme une réversion

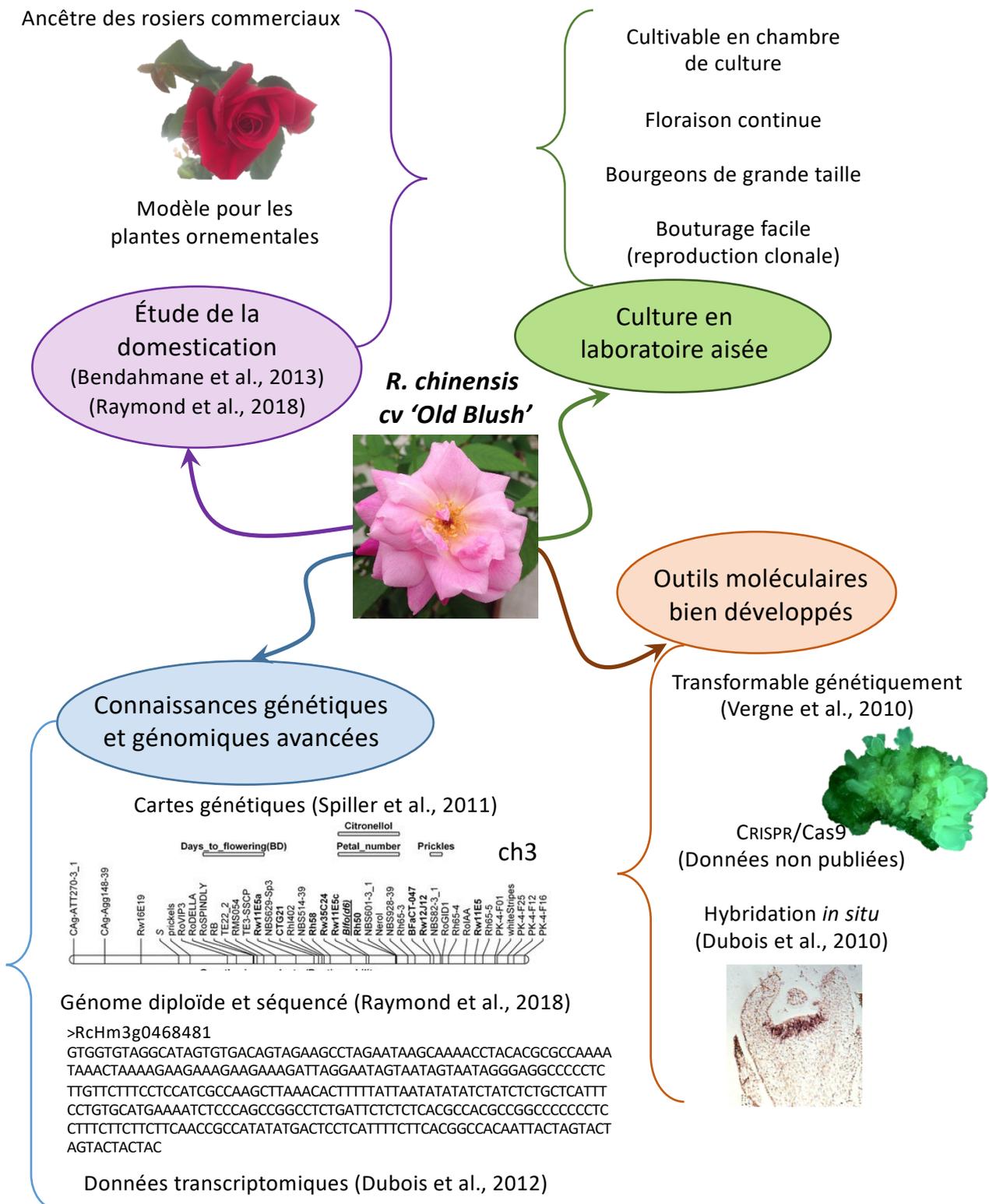


Figure 11 : Résumé des avantages de *R. chinensis* cv 'Old Blush' en tant que modèle d'étude scientifique.

Son génome diploïde et sa position clé lors de la domestication et de l'amélioration variétale du rosier font de 'Old Blush' une variété adaptée pour l'étude des mécanismes de la domestication et de l'origine génétique des traits d'intérêt. Pour répondre à ces questions, de nombreux outils sont désormais disponibles : un génome séquencé et assemblé, des cartes génétiques, des données transcriptomiques, ainsi que des protocoles d'hybridation *in situ*, de transformation génétique et d'édition de génome.

vers le phénotype sauvage. Le second compare 'Old Blush' et un autre mutant spontané à fleur simple obtenu fortuitement dans l'équipe après passage en culture *in vitro*. Ces deux jeux de données sont une source importante d'informations sur le développement floral et en particulier celui des pétales.

Par conséquent, l'ensemble des outils culturels, génétiques et moléculaires développés chez 'Old Blush', ainsi que son rôle important dans l'histoire de la domestication du rosier et de l'obtention des rosiers modernes font de cette variété un bon modèle d'étude pour les traits phénotypiques liés à la domestication, comme cela est résumé sur la Figure 11.

V. Le séquençage du génome du rosier, un outil puissant

Une autre avancée technologique essentielle pour la recherche fondamentale est l'obtention de la séquence qui représente la presque totalité du génome d'Old Blush', ce qui est assez rare chez les plantes ornementales. Elle permet notamment d'accéder aux séquences complètes des gènes, y compris leurs régions régulatrices, ainsi qu'aux séquences répétées du génome. Ce puissant outil, issu d'un projet coordonné par mon équipe d'accueil, a été publié très récemment.

1. Le génome d'Old Blush', un génome séquencé de très bonne qualité

Le génome du rosier est un relativement petit génome de 560 Mb à l'état haploïde (Raymond et al., 2018 ; Yokoya et al., 2000). Pourtant, étant donné les hybridations interspécifiques suivies de multiplication des variétés par bouturage majoritairement, nombre de variétés de rosiers, y compris 'Old Blush', présentent une forte hétérozygotie, c'est-à-dire une grande différence de séquence entre les deux chromosomes de chaque paire. Cela a longtemps limité la qualité de l'assemblage du génome complet de 'Old Blush'.

Le séquençage et l'assemblage de ce génome hétérozygote ont été réalisés en 2012-2014 dans mon équipe. Environ 400x de reads Illumina paired-end avec des tailles d'insert variant de 168 pb à 11 000 pb ont été produits. Leur assemblage a abouti à la reconstruction de 15 938 scaffolds (Raymond et al., 2018), pour une longueur totale de 882 Mb et plus de la moitié des séquences contenues dans des contigs de taille supérieure à 227 kb (N50 = 227kb). Sur cette séquence, 75 573 gènes ou allèles ont été annotés. La comparaison avec les données transcriptomiques précédemment décrites et les protéines prédites sur les génomes de fraisier et pommier a permis d'estimer que, pour environ les deux tiers des gènes, les deux allèles avaient été assemblés séparément. Donc bien que fragmenté, cet assemblage donne accès au génome hétérozygote complet du rosier. Une stratégie parallèle a été développée pour augmenter la continuité de l'assemblage. Dans d'autres espèces hétérozygotes, telles que la vigne (Zharkikh et al., 2008), un long processus d'autofécondation commencé dans les années 50 a permis d'obtenir une lignée quasi-homozygote, ce qui a facilité l'assemblage de pseudo-chromosomes (Jaillon et al., 2007). Dans le cas du rosier, l'auto-fécondation semblait vraiment longue

et hasardeuse. Afin de surmonter ce blocage, l'équipe a développé une méthode originale visant à produire une plantule homozygote dérivée de tissu haploïde (microspores) cultivé *in vitro* après doublement chromosomique (Vergne et al., 2018). Ce génome a ensuite été séquencé par une méthode de séquençage « long read » sur la plateforme PacBio II (Eid et al., 2009). L'utilisation d'un haploïde doublé, la profondeur de couverture du séquençage de 80x et les méthodes d'assemblage innovantes ont permis d'obtenir un assemblage d'excellente qualité, avec un N50 de 24 Mb, ce qui permet l'analyse de longs intervalles continus. Les 82 contigs produits ont ensuite été ancrés sur les 25 695 marqueurs génétiques de la carte publiée par Bourke et al. (2017) pour les ordonner sur les 7 groupes de liaison du rosier (Figure 12-B). Cela a permis de valider l'excellente colinéarité des contigs et de la carte génétique, et de reconstruire les 7 pseudo-chromosomes. L'assemblage d'excellente qualité, ainsi que les analyses moléculaires qui ont été basées dessus et auxquelles j'ai participé, ont permis de publier ces résultats dans *Nature Genetics* en 2018. Encore aujourd'hui, il correspond à un des assemblages les plus continus existant chez les plantes (Belser et al., 2018), ce qui ouvre de nombreuses possibilités quant aux recherches sur le rosier voire sur d'autres espèces proches.

2. Reconstruction de l'histoire évolutive entre espèces proches du rosier

L'existence de données de séquençage de qualité chez le rosier, mais aussi chez le pêcher, le pommier, le fraisier et, dans une moindre mesure chez le framboisier, permet de comparer les séquences et les caryotypes de ces espèces. À partir de ces comparaisons, il est alors possible de reconstruire l'évolution des génomes et des espèces de Rosaceae depuis leur dernier ancêtre commun. Ainsi, il a été mis en évidence que celui-ci devait posséder 9 protochromosomes et environ 9 000 gènes. L'ancêtre des Rosoideae, qui comptent le genre *Rosa* et ses proches voisins *Fragaria* et *Rubus*, possédait quant à lui 8 protochromosomes résultant d'une fission et de deux fusions de chromosomes (Raymond et al., 2018). De façon similaire, l'ancêtre des *Prunus* possédait certainement lui aussi 8 protochromosomes issus de 2 fissions et 4 fusions différentes. En revanche, l'ancêtre des *Malus* et *Pyrus* possédait probablement 17 protochromosomes issus d'une duplication complète du génome suivie de réarrangements correspondant à 11 fissions et 12 fusions (Raymond et al., 2018).

Au vu du petit nombre de réarrangements, il semble donc possible de comparer ces génomes sur des blocs de synténie de plusieurs dizaines de mégabases, promouvant ainsi les recherches translationnelles et le transfert de connaissances entre Rosaceae.

3. Histoire évolutive au sein du genre *Rosa* : exemple d'« Old Blush »

Les roses sauvages de la section Chinenses, comme *R. gigantea* ou *R. chinensis*, sont caractérisés par des fleurs simples, des stipules entières et des styles courts et libres. La section Synstylae est, quant à elle, caractérisée par des styles rassemblés dans une colonne mince (d'où le nom de la section, « syn » signifiant ensemble), plusieurs fleurs par hampe florale (appelée cyme) et des

stipules pectinées ou dentées. *Rosa chinensis* cv 'Old Blush' présente des inflorescences de type cymes, des styles libres mais saillants et des stipules dentées. Ces traits morphologiques pourraient donc indiquer une origine hybride d'Old Blush', en relation étroite avec les sections Chineses et Synstylae.

L'analyse de la distribution et de la structuration de la diversité des SNP (pour « Single Nucleotide Polymorphism ») dans les données de reséquençage de différentes variétés ancestrales de rosier provenant de ces deux sections, alignées sur le génome haploïde d'Old Blush', a permis d'éclaircir en partie les origines de ce dernier. Tout d'abord, elle indique que la diversité est structurée le long de certaines régions chromosomiques selon des patrons intermédiaires entre ceux des espèces Synstylae et Chineses ancestrales, ce qui confirme l'hypothèse d'une origine hybride d'Old Blush' (Raymond et al., 2018). D'autre part, parmi les 2 variétés reséquencées de la section Chineses, le génome d'Old Blush' semble extrêmement proche du génome de *R. gigantea*, avec environ 28 % de son génome haploïde très similaire à celui de *R. gigantea* (moins d'un variant par kb), ce qui indique que cette fleur simple sauvage (ou une de ses très proches voisines) est très certainement un ancêtre de notre rosier modèle (Raymond et al., 2018). En revanche, parmi les 3 génotypes reséquencés de la section Synstylae, aucune proximité aussi nette n'a été identifiée, indiquant que l'ancêtre d'Old Blush' au sein des Synstylae reste à découvrir (Raymond et al., 2018).

4. Reconstruction de l'intervalle associé à la fleur double chez le rosier

La séquence génomique obtenue à partir de la lignée homozygote dérivée d'Old Blush' a permis d'obtenir la séquence presque complète d'un et d'un seul des haplotypes de chaque chromosome du génome d'Old Blush'. L'obtention d'un assemblage de très haute qualité est un événement déterminant pour l'étude des traits d'intérêt chez les espèces à importance horticole ou agronomique car elle permet d'obtenir la séquence complète correspondant aux intervalles génétiques identifiés comme contenant les déterminants génétiques de ces traits (Figure 12). Ainsi, jusqu'au séquençage du génome homozygote d'Old Blush', il avait été impossible d'obtenir le contenu complet de l'intervalle génétique contenant la mutation à l'origine du phénotype « fleur double » chez le rosier. En effet, seuls des fragments de cet intervalle avaient pu être identifiés, comme cela est par ailleurs toujours le cas chez d'autres espèces, comme *Petunia hybrida* et *Dianthus chinensis*, dont les séquences génomiques sont encore trop fragmentées pour avoir accès à l'ensemble des gènes contenus dans les intervalles génétiques contenant la mutation « fleur double ». Cependant, étant donné que la recherche des déterminants génétiques est généralement fondée sur une cartographie génétique comparant une population ségrégant pour le caractère d'intérêt, il apparaît nécessaire de pouvoir accéder à chacun des deux haplotypes afin de connaître l'état de chacun des allèles du ou des gène(s) considéré(s). Enfin, dans le cadre de mon projet, reconstruire le second haplotype correspondant à l'intervalle du chromosome 3 où se trouve le locus déterminant la formation de fleurs

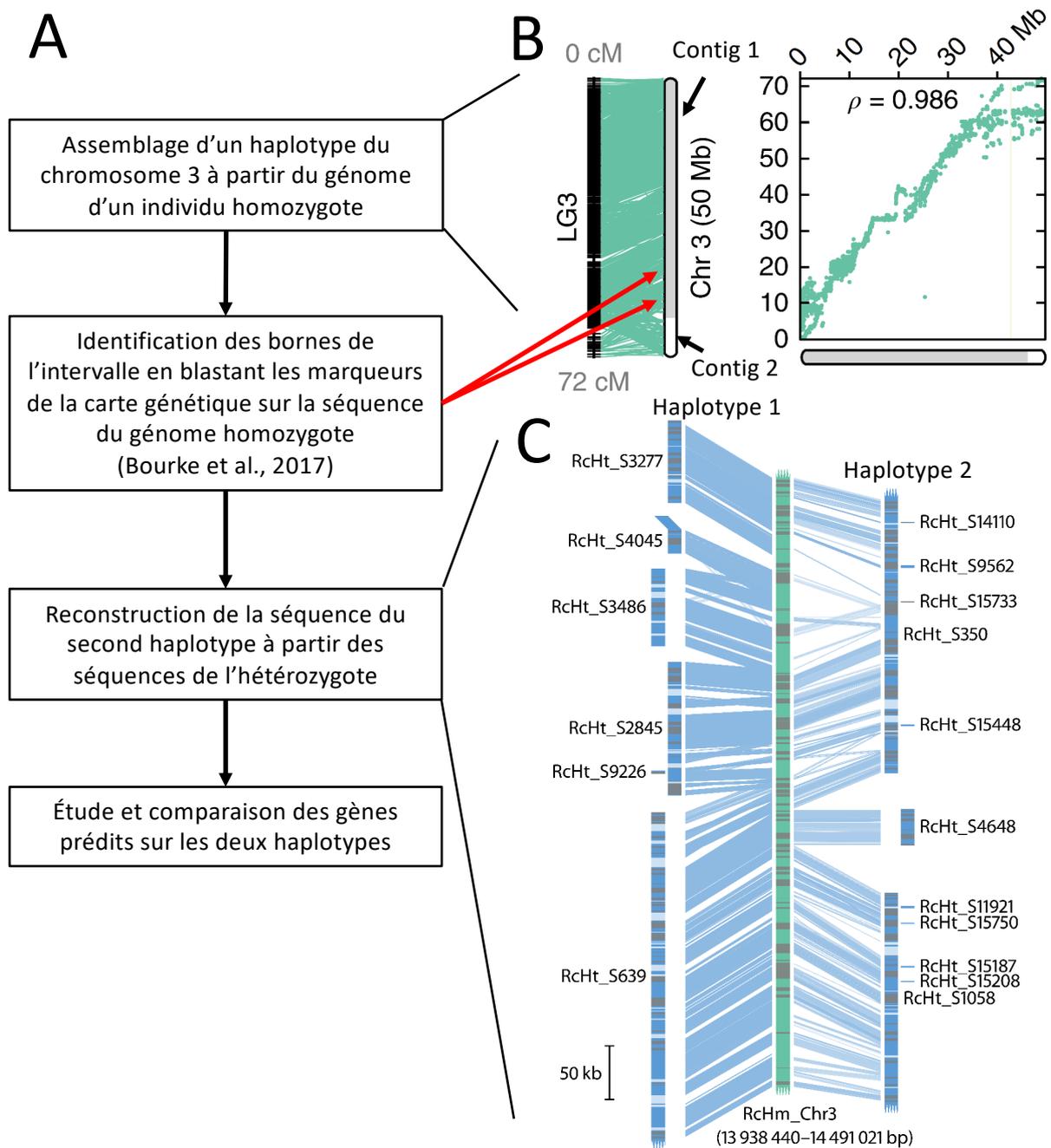


Figure 12 : Méthode de reconstruction des deux haplotypes de l'intervalle contenant le déterminant de la fleur double.

- (A) Résumé des différentes étapes de la reconstruction des deux haplotypes de l'intervalle. Mon projet de thèse débute à partir de la dernière étape.
- (B) Représentation du chromosome 3 avec ses deux contigs ancrés grâce à des marqueurs génétiques (panneau de gauche) et la correspondance entre la position (en centimorgans) des marqueurs sur la carte génétique et leur position physique le long de la séquence du génome (panneau de droite). Extraits de Raymond et al. (2018).
- (C) Identification du second haplotype par BLAST de la séquence homozygote sur les scaffolds du génome hétérozygote : exemple d'une portion de l'intervalle. Au centre, est représentée une portion du pseudo-chromosome du génome homozygote (en vert). À gauche, se trouvent les scaffolds hétérozygotes (en bleu) correspondant au même haplotype. Leur pourcentage de similarité avec le génome homozygote est plus élevé que ne l'est celui des scaffolds du second haplotype (à droite, en bleu également).

simples ou doubles était d'autant plus essentiel que la mutation induisant le phénotype « fleur double » est dominante. Ainsi, à partir des séquences produites par séquençage Illumina du génome complet et hétérozygote d'«Old Blush», et grâce à la structure du premier haplotype, il a été possible d'identifier les deux allèles de chacun des 631 gènes présents dans l'intervalle et de les comparer (Figure 12). Ce travail, qui a été commencé par Annick Dubois (chargée de recherche dans l'équipe) avant mon arrivée puis mené à bout par Marion Verdenaud (ingénieure d'étude dans l'équipe) au début de ma thèse, est à la base de mon projet qui donc a pu se fonder sur des connaissances solides quant au contenu de l'intervalle contenant la mutation induisant la formation de la fleur double.

Chapitre II : Objectifs de la thèse

Alors que les roses à fleurs doubles sont si fascinantes pour l'œil grâce à leurs très nombreux pétales qui leur donnent une apparence plus dense, la nature de la mutation causant le passage d'une fleur simple à une fleur double restait un mystère pour les sélectionneurs et scientifiques depuis des siècles. Les travaux récents des membres, actuels et passés, de mon équipe d'accueil ont permis de mettre en place l'ensemble des outils indispensables à l'amélioration des connaissances fondamentales sur le rosier (Bendahmane et al., 2013 ; Dubois et al., 2010, 2011, 2012 ; Vergne et al., 2009), tout en commençant à lever le voile du mystère de la fleur double il y a quelques années (Dubois et al., 2010). C'est dans la continuation de ces recherches que s'est inscrit mon travail de thèse.

I. État de l'art avant le début de la thèse

Pour résumer les informations disponibles, à mon arrivée, à propos du déterminisme de la fleur double chez le rosier :

- (1) Le passage d'une fleur simple à une fleur double dépend d'un déterminant unique et dominant (Debener, 1999). Ce déterminant est présent à l'état hétérozygote chez 'Old Blush' car la moitié des descendants du croisement de cette variété avec une variété à fleurs simples sont à fleurs doubles (Shuppert, 2009).
- (2) Les cartes génétiques, combinées à l'assemblage des données de séquençage du génome d'Old Blush', localisent ce déterminant dans un intervalle de 6 207 462 pb sur le chromosome 3, composé de 631 gènes et appelé Df.
- (3) *RcAGAMOUS*, qui est quant à lui localisé sur le chromosome 5, semble avoir une expression plus restreinte dans les fleurs doubles que dans les fleurs simples (Figure 10 ; Dubois et al., 2010), suggérant un mouvement de la frontière entre les gènes de classe A et les gènes de classe C vers le centre du méristème floral.

D'un point de vue global, ma thèse visait donc à identifier quel(s) mécanisme(s) a(ont) été sélectionné(s) par l'homme afin de produire des rosiers à fleurs doubles. Elle peut être divisée en plusieurs étapes permettant de répondre à cette question.

II. Première étape : identification des candidats pour le déterminant de la fleur double

À partir de l'intervalle assemblé, plusieurs stratégies complémentaires ont alors été envisagées pour identifier quelques candidats. Pour commencer, étant donnée la restriction du territoire d'expression de *RcAG* vers le centre du méristème floral dans le cas de la formation de fleurs doubles, il est possible de s'intéresser aux gènes de l'intervalle dont un homologue est connu pour inhiber *AGAMOUS* chez d'autres espèces modèles, en particulier *Arabidopsis thaliana*. D'autre part, alors que les candidats recherchés doivent être présents à l'état hétérozygote chez 'Old Blush', une première

comparaison locale des séquences des deux haplotypes (Figure 12) a montré que les éléments transposables représentaient une part importante des différences. Il est donc important de ne pas se limiter à la recherche de polymorphismes courts et de pouvoir aussi étudier les polymorphismes liés aux mouvements d'éléments transposables. Ensuite, dans le cas où la mutation causale serait difficile à identifier au niveau génomique, par exemple s'il s'agit d'une modification épigénétique ou d'une mutation dans des séquences régulatrices éloignées de la séquence codante, une autre stratégie possible consiste à chercher dans cet intervalle des gènes qui sont différentiellement exprimés entre une fleur simple et une fleur double. Ces derniers pourraient alors être soit de bons candidats pour porter le déterminant de la fleur double, soit au minimum des gènes dont l'expression est dérégulée par la surproduction de pétales et donc aider à identifier des conséquences de la mutation recherchée.

III. Deuxième étape : validation du rôle de ces candidats

Une fois les candidats sélectionnés, il sera important de confirmer leur implication dans le développement floral et en particulier dans le contrôle du nombre de pétales ; le but ultime étant de vérifier quelle mutation est suffisante pour obtenir le phénotype « fleur double ». Pour cela, il est intéressant de réaliser des transformations génétiques stables chez différentes variétés de rosiers et d'en analyser les conséquences phénotypiques. En particulier, il est important de supprimer ou sous-exprimer le candidat chez 'Old Blush' pour voir si on induit la formation d'une fleur simple et, réciproquement, d'exprimer le candidat chez une fleur simple et vérifier si on induit la production de fleurs doubles. Enfin, il pourrait être envisagé de vérifier si cette mutation est retrouvée chez l'ensemble des rosiers à fleurs doubles ou si des mutations différentes portant sur le même gène ont pu être sélectionnées lors d'événements de sélection du phénotype « fleur double » potentiellement indépendants.

IV. Troisième étape : caractérisation du mécanisme d'induction de la fleur double lié à cette mutation

Afin de mieux documenter le mécanisme par lequel une fleur double est mise en place, il sera important de s'intéresser à la fonction de la protéine à l'origine du phénotype. Pour cela, les informations disponibles dans la littérature ou l'analyse de mutants, en particulier chez *Arabidopsis*, pourront servir de base de recherche pour déterminer la fonction de la protéine et les voies de signalisation impactées par la mutation.

D'autre part, étant donné le lien entre le niveau d'expression d'*AGAMOUS* et la formation de la fleur double chez de nombreuses espèces et variétés (Galimba et al., 2012; Liu et al., 2012; Ma et al., 2018; Nakatsuka et al., 2015; Nitasaka, 2003), y compris chez le rosier (Dubois et al., 2010), il sera intéressant d'étudier le lien entre ce gène homéotique et les candidats sélectionnés. Pour commencer,

il serait envisageable de vérifier si les candidats interagissent directement avec *RcAGAMOUS* ou par l'intermédiaire d'autres intervenants. En effet, l'étude du mécanisme par lequel le déterminant de la fleur double réprime *RcAGAMOUS* pourrait aussi indiquer d'autres gènes impliqués dans la régulation du nombre de pétales et donc de potentiels autres candidats pour l'induction de la fleur double chez d'autres variétés.

V. Quatrième étape : étude de la conservation de ce mécanisme chez d'autres plantes ornementales

Pour finir, il serait intéressant de vérifier si des mutations similaires pourraient être impliquées dans la formation de fleurs doubles chez d'autres espèces proches, d'abord parmi les Rosaceae, puis chez des espèces plus éloignées. Des transformations génétiques hétérologues des gènes candidats pourront être envisagées chez *F. vesca*, *A. thaliana* et/ou *P. hybrida*. D'autre part, l'analyse de l'état des gènes candidats chez des variétés à fleurs doubles provenant d'autres genres pourrait aussi permettre de vérifier si la voie de signalisation et peut-être la mutation de ces candidats sont conservées à travers différents genres, familles ou clades.

Chapitre III : Résultats

I. Identification d'un premier gène candidat pour le contrôle du nombre de pétales chez le rosier : le cas de *RcUGO*

1. Contexte

À mon arrivée dans l'équipe, une première recherche de gènes candidats avait déjà été effectuée. Vu l'existence de plusieurs révertants somatiques et spontanés produisant de nouveau des fleurs simples, une hypothèse était que la mutation causale du passage d'une fleur simple à une fleur double pouvait être l'insertion/excision d'un élément transposable dans ou à proximité du gène déterminant la formation de fleurs simples ou doubles. En effet, l'insertion d'un élément transposable est parfois réversible ce qui serait cohérent avec la présence de révertants, comme la variété à fleurs simples *Rosa chinensis* cv 'Single Pink' qui dérive de la variété à fleurs doubles *R. chinensis* cv 'Old Blush'. Ces insertions peuvent alors modifier la fonctionnalité d'un gène de multiples façons, notamment en induisant un arrêt prématuré de la séquence codante ou en modifiant le niveau d'expression de celui-ci. Il avait donc été décidé de comparer les gènes exprimés lors du développement floral de ces deux variétés de *R. chinensis* afin de trouver des candidats impliqués dans le contrôle du nombre de pétales chez le rosier. L'idée sous-jacente du RNA-seq était que, bien que la majorité des gènes différentiellement exprimés entre ces deux variétés correspondront à des effecteurs du développement des pétales, ceux contenus dans l'intervalle Df représenteront de bons candidats pour le déterminisme de la fleur double. Il est cependant important de noter que cette approche ne concerne que les cas où la mutation liée à la fleur double provoquerait un défaut d'expression (*silencing* d'un élément transposable produisant l'extinction du gène à proximité, ou insertion dans un promoteur par exemple). En cas d'échec de l'approche par analyse différentielle de l'expression des gènes de l'intervalle, le RNA-seq donnera tout de même accès au polymorphisme dans les séquences des gènes exprimés, menant pourquoi pas à une seconde approche pour identifier des candidats pour le déterminisme des fleurs doubles. La méthode consistait donc à identifier des gènes localisés dans l'intervalle génétique identifié comme contenant la mutation Df et différentiellement exprimés entre les jeunes bourgeons floraux de la variété modèle à fleurs doubles 'Old Blush' et ceux de son mutant somatique à fleurs simples 'Single Pink' (Figure 13). Pour cela, des méristèmes floraux au stade soit de formation des sépales (stade 1), soit de formation des pétales (stade 2), soit de formation des étamines (stade 3) ont été collectés et utilisés pour une extraction des ARN totaux, puis leur séquençage (RNA-seq). Après identification d'une centaine de transcrits contenus dans l'intervalle (qui à l'époque n'était pas encore totalement assemblé), mon équipe d'accueil avait pu identifier un candidat très prometteur puisque exprimé presque 3 fois plus fortement chez 'Old Blush' que chez 'Single Pink'. Celui-ci, nommé *RcUGO* pour « *Upregulated Gene in 'Old Blush'* » (*RcHt_S6248.1* dans le génome hétérozygote,

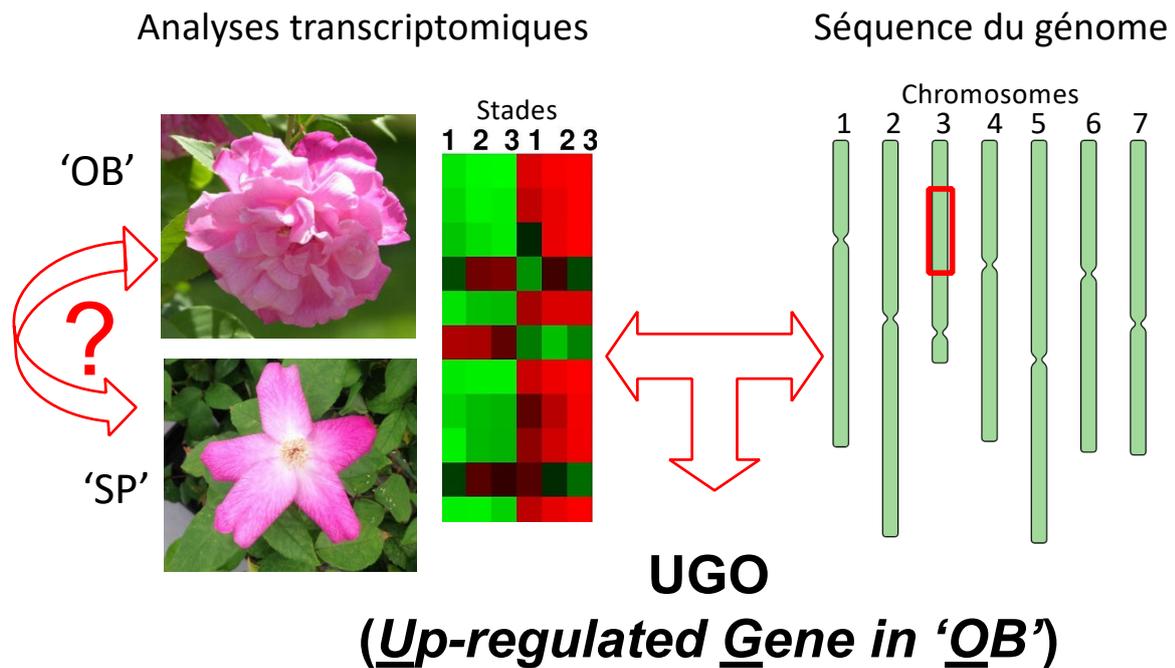


Figure 13 : Méthode ayant permis l'identification d'un premier candidat pour le déterminisme génétique de la fleur double.

Recherche de gènes différentiellement exprimés entre une fleur simple et une fleur double et localisés dans l'intervalle contenant le déterminant 'fleur double'.

Les données transcriptomiques ont été obtenues à partir d'ARN extraits de bourgeons aux stades précoces du développement floral de la variété à fleur double *Rosa chinensis* cv 'Old Blush' ('OB') et de son mutant spontané *R. chinensis* cv 'Single Pink' ('SP').

Stade 1 : initiation des sépales ; Stade 2 : initiation des pétales ; Stade 3 : initiation des étamines.

RcHm_v2.0_Chr3g0472971 dans le génome homozygote), était le candidat présent dans l'intervalle avec la plus forte surexpression chez 'Old Blush'.

2. *RcUGO*, un « *small leucine zipper* » présent dans l'intervalle

À partir de cette observation, nous avons décidé de mieux caractériser ce gène en recherchant en particulier en quoi il pourrait influencer le développement floral. C'est à ce moment-là qu'a commencé mon travail.

Les prédictions de gènes réalisées sur la séquence du génome du rosier indiquent que *RcUGO* est un gène de taille relativement petite (1721 pb), situé sur le chromosome 3 (entre les positions 18 954 240 pb et 18 955 961 pb). Il est composé de deux exons séparés par un intron et donne lieu à la production d'un ARN messager de 850 pb. La protéine codée par cet ARNm contient 82 acides aminés, ce qui en fait une protéine de petite taille moléculaire. Une première analyse par BLASTP (Altschul et al., 1990 ; <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) a permis de la classer dans la famille des ZPR (ou LITTLE ZIPPERs en anglais) qui contient quatre représentants chez *A. thaliana* (Wenkel et al., 2007). Ceux-ci codent pour de petites protéines de 67 à 105 acides aminés avec un unique domaine leucine-zipper (LZ) très conservé et typique de la famille (Wenkel et al., 2007). Les domaines leucine-zipper sont caractérisés par un enchaînement de plusieurs motifs de sept résidus (heptades) contenant chacun une leucine, ou parfois un autre acide aminé hydrophobe, en position 4 de l'heptade. Les leucines appartenant à deux domaines leucine zipper sont capables d'interagir entre elles et permettent la dimérisation des protéines qui les portent.

La recherche de domaines protéiques par des logiciels comme Pfam (Finn et al., 2016) ou PROSITE (Sigrist et al., 2013 ; <https://prosite.expasy.org/>) n'est pas parvenue à détecter de domaine caractéristique chez *RcUGO*. En revanche, après identification par BLASTP des membres de la famille ZPR chez les espèces *F. vesca* (Shulaev et al., 2011), *M. truncatula* (Tang et al., 2014), *P. hybrida* (Bombarely et al., 2016) et *P. persica* (The International Peach Genome Initiative et al., 2013), l'alignement de leurs séquences protéiques avec celle de *RcUGO* révèle un domaine fortement conservé, riche en acides aminés hydrophobes, correspondant bien au domaine à leucine zipper typique de la famille (Figure 14-B). À partir de cet alignement multiple, un arbre phylogénétique a été construit et permet de distinguer deux sous-familles de ZPR : la première, qui contient ZPR1, ZPR2 et *RcHm_v2.0_Chr1g0350821*, est caractérisée par la présence d'une séquence en N-terminal, en amont du domaine leucine zipper, d'une trentaine d'acides aminés alors que la seconde, qui contient ZPR3 et ZPR4 et *RcUGO* (*RcHm_v2.0_Chr3g0472971*), est caractérisée par l'absence de ces acides aminés (Figure 14-B ; Floyd et al., 2014). La présence de deux orthologues chez *A. thaliana* pour chaque ZPR de rosier, éventuellement due à la récente duplication complète de génome ayant eu lieu spécifiquement chez les Brassicaceae (Blanc et al., 2003), suggère une potentielle redondance

fonctionnelle chez *A. thaliana*, confirmée par l'absence de phénotype visible chez les simples mutants, au contraire des doubles mutants (Kim et al., 2008).

3. Chez *A. thaliana*, les orthologues de *RcUGO* interviennent dans le contrôle de l'activité méristématique

De manière intéressante pour notre étude des mécanismes d'augmentation du nombre de pétales chez le rosier, il semblerait que les ZPR soient impliqués dans la régulation et la structuration des méristèmes caulinaires d'*A. thaliana*. En effet, chez *A. thaliana*, le double mutant *zpr3/zpr4* présente une perte de structure des méristèmes, une phyllotaxie altérée et parfois une production d'organes surnuméraires due à l'apparition de méristèmes ectopiques (Kim et al., 2008). Les analyses moléculaires indiquent que *WUS* est légèrement surexprimé et que *CLV3* l'est très fortement, avec une expansion de son domaine d'expression à l'ensemble de la tunica, correspondant aux couches externes L1 et L2 du méristème (Kim et al., 2008). Au contraire, la surexpression de *ZPR3* induit non seulement une abaxialisation des feuilles mais aussi, dans le cas des phénotypes les plus sévères, un arrêt du méristème apical caulinaire et une réduction du nombre de méristèmes secondaires (Brandt et al., 2013 ; Floyd et al., 2014 ; Kim et al., 2008). Par conséquent, les ZPR semblent impliqués dans l'organisation et le fonctionnement du méristème ce qui pourrait être cohérent avec l'augmentation de l'activité méristématique observée chez les rosiers à fleurs très doubles et, bien que ces mutants ne produisent pas de pétales surnuméraires chez *A. thaliana*, on peut imaginer que ce soit le cas chez d'autres espèces, comme *R. chinensis*.

L'étude de la structure de ces petites protéines, ainsi que de leur origine phylogénétique (Figure 14), permet de proposer un mécanisme moléculaire pour leur fonction. L'analyse des séquences des domaines leucine zipper des ZPR a en effet montré que ceux-ci sont très proches du domaine leucine zipper des facteurs de transcription à homéodomaine couplé à un domaine leucine zipper de classe III (ou HD-ZIP de classe III, Floyd et al., 2014). En particulier, la présence d'asparagine en première position des heptades, qui est prédite comme source de reconnaissance et de dimérisation des domaines ayant la même structure (Brandt et al., 2013), est partagée par ces deux familles, suggérant à la fois une origine commune et pourquoi pas une capacité d'hétérodimérisation. La première hypothèse a été supportée par des analyses phylogénétiques comparant les domaines leucine zipper de l'ensemble des HD-ZIP de différentes espèces d'Embryophytes, qui ont révélé que la famille des ZPR provient de la duplication d'un HD-ZIP de classe III suivie d'une dégénérescence partielle des domaines adjacents (Floyd et al., 2014). La deuxième hypothèse a été corroborée par Kim et al. (2008), qui ont montré que *ZPR3* reconnaît et interagit avec l'ensemble des HD-ZIP III. En parallèle, Wenkel et al. (2007) ont montré que, réciproquement, l'HD-ZIP de classe III *REVOLUTA* (*REV*) est capable d'interagir *in vitro* avec chacun des quatre ZPR présents chez *A. thaliana*. Quand un ZPR interagit avec

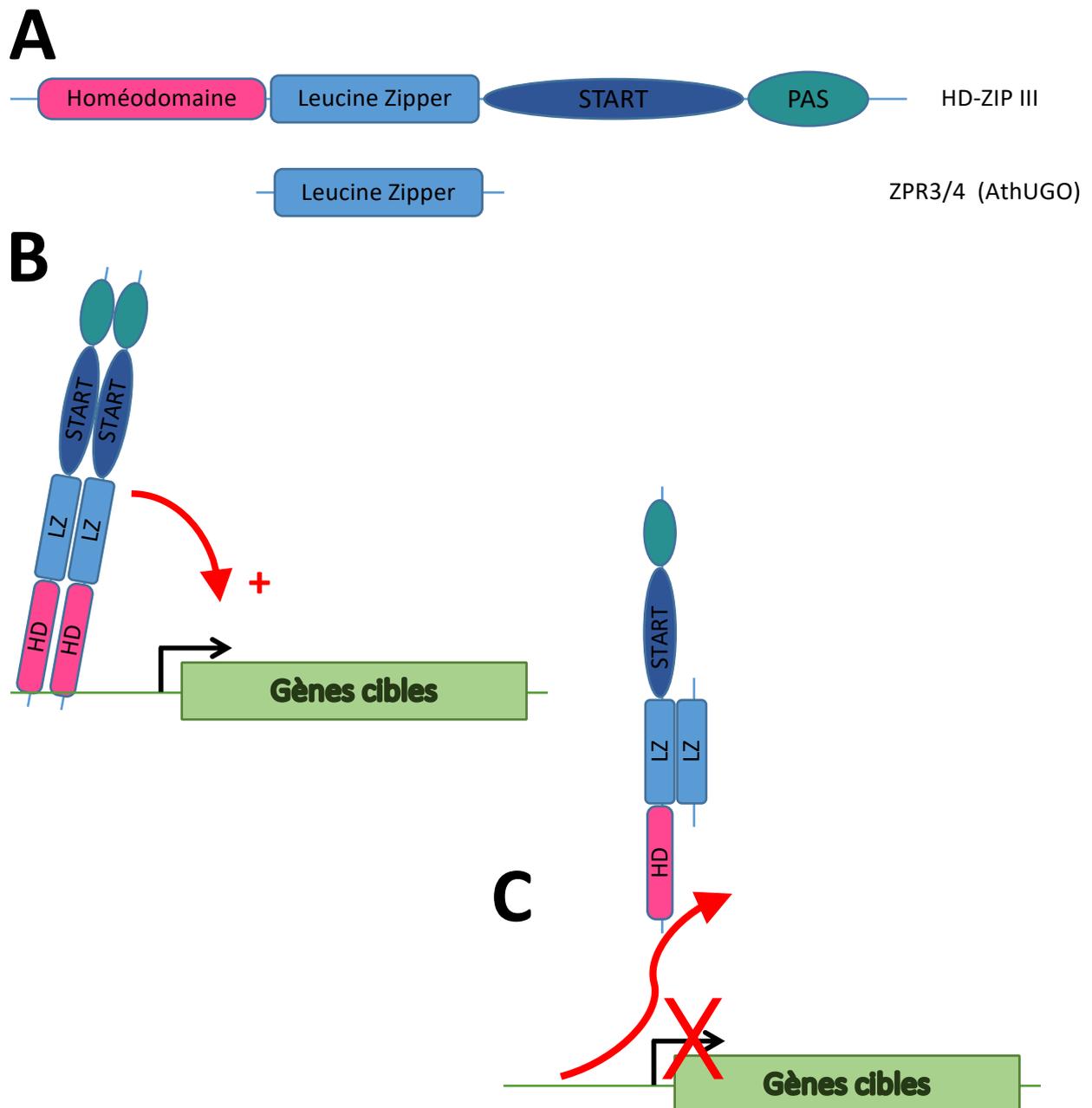


Figure 15 : Modèle de fonctionnement des ZPR comme inhibiteurs des HD-ZIP de classe III.

- (A) Structure des HD-ZIP de classe III et des ZPR. L'homéodomaine est responsable des interactions avec l'ADN, le domaine leucine zipper permet la dimérisation des protéines, le domaine START est potentiellement capable de lier de petites molécules hydrophobes telles que les stéroïdes et le domaine PAS, ou MEKHLA, pourrait servir de senseur ou d'aide à la dimérisation.
- (B) Mécanisme moléculaire de régulation du niveau d'expression des gènes cibles de HD-ZIP III. Une fois dimérisés grâce à leur domaine leucine zipper, les HD-ZIP III sont capables de reconnaître leur séquence d'ADN cible et de moduler l'expression des gènes adjacents.
- (C) Étant donné que les ZPR ne présentent pas d'homéodomaine, la dimérisation entre un HD-ZIP III et un ZPR ne permet plus l'interaction avec l'ADN; ils jouent alors le rôle de dominants négatifs vis-à-vis des HD-ZIP III.

REV, celui-ci n'est alors plus capable de lier ses séquences d'ADN cibles *in vitro* (Wenkel et al., 2007). Ainsi, alors que les HD-ZIP sont actifs sous forme de dimère, la dimérisation hétérologue entre un HD-ZIP III et un ZPR, dont l'homéodomaine a été perdu, empêche l'interaction entre l'HD-ZIP III et ses séquences d'ADN cibles et rend le complexe inactif (Figure 15-C). Les ZPR sont ainsi parfois qualifiés soit de dominants négatifs soit de microprotéines en référence aux microARN qui correspondent à des portions d'ARN complémentaires de leur cible et qui bloquent eux aussi la fonction des ARN avec lesquels ils interagissent (Brandt et al., 2013).

Les HD-ZIP III, au nombre de cinq à la fois chez *A. thaliana* (PHABULOSA, REVOLUTA, PHAVOLUTA, CORONA et ATHB8) et chez *R. chinensis*, sont caractérisés par la présence en N-terminal d'un homéodomaine adjacent à un domaine de dimérisation leucine zipper (Schena and Davis, 1992), puis d'un domaine START (pour STeroidogenic Acute Regulatory protein-related lipid-Transfer ; Ponting and Aravind, 1999) prédit pour lier des stéroïdes ou autres molécules hydrophobes, et d'un domaine MEKHLA, riche en méthionines (M), glutamates (E), lysines (K), histidines (H), leucines (L) et alanines (A), potentiellement capable de lier de petits métabolites et de faciliter la dimérisation (Mukherjee and Bürglin, 2006). Ces protéines jouent de nombreux rôles lors du développement d'*A. thaliana*, allant de la polarisation des organes végétatifs (McConnell et al., 2001) et des tissus conducteurs (Emery et al., 2003) au maintien de l'organisation du méristème (Green et al., 2005). En particulier, la perte de fonction de *CORONA* (*CNA*) produit de plus grands méristèmes (Green et al., 2005), alors que celle de *REV* mène à la formation de plantes avec moins de méristèmes secondaires et parfois une diminution du nombre d'organes par fleurs (Otsuga et al., 2001). Ces observations, cohérentes avec les phénotypes observés lors de la surexpression d'un *ZPR* (Wenkel et al., 2007), suggèrent que la modification de la balance entre *ZPR* et HD-ZIP est impliquée dans le contrôle de l'activité méristématique et pourrait donc l'être aussi dans la formation de fleurs doubles chez le rosier. Ainsi, notre intérêt pour *RcUGO* en a été d'autant plus renforcé.

4. Le niveau d'expression de *RcUGO* diffère peu entre fleurs simples et fleurs doubles

Afin de valider la différence d'expression de *RcUGO* observée en RNA-seq entre la variété 'Old Blush' à fleurs doubles et de son mutant spontané à fleurs simples, nous avons quantifié l'expression par reverse transcription et PCR quantitative (RT-qPCR) sur des ARNm extraits de jeunes bourgeons triés selon leur stade de développement, comme pour le RNA-seq : stade 1 correspondant à la formation des sépales, stade 2 à celle des pétales et stade 3 à celle des étamines. Les premières expériences, réalisées sur des échantillons prélevés par Xiao-Peng Fu (post-doctorante dans l'équipe de 2014 à 2015), ont permis d'obtenir les résultats présentés en Figure 16-A,C, où *RcUGO* est exprimé jusqu'à 40 fois plus chez 'Old Blush' que chez 'Single Pink'. Cependant, l'utilisation d'échantillons

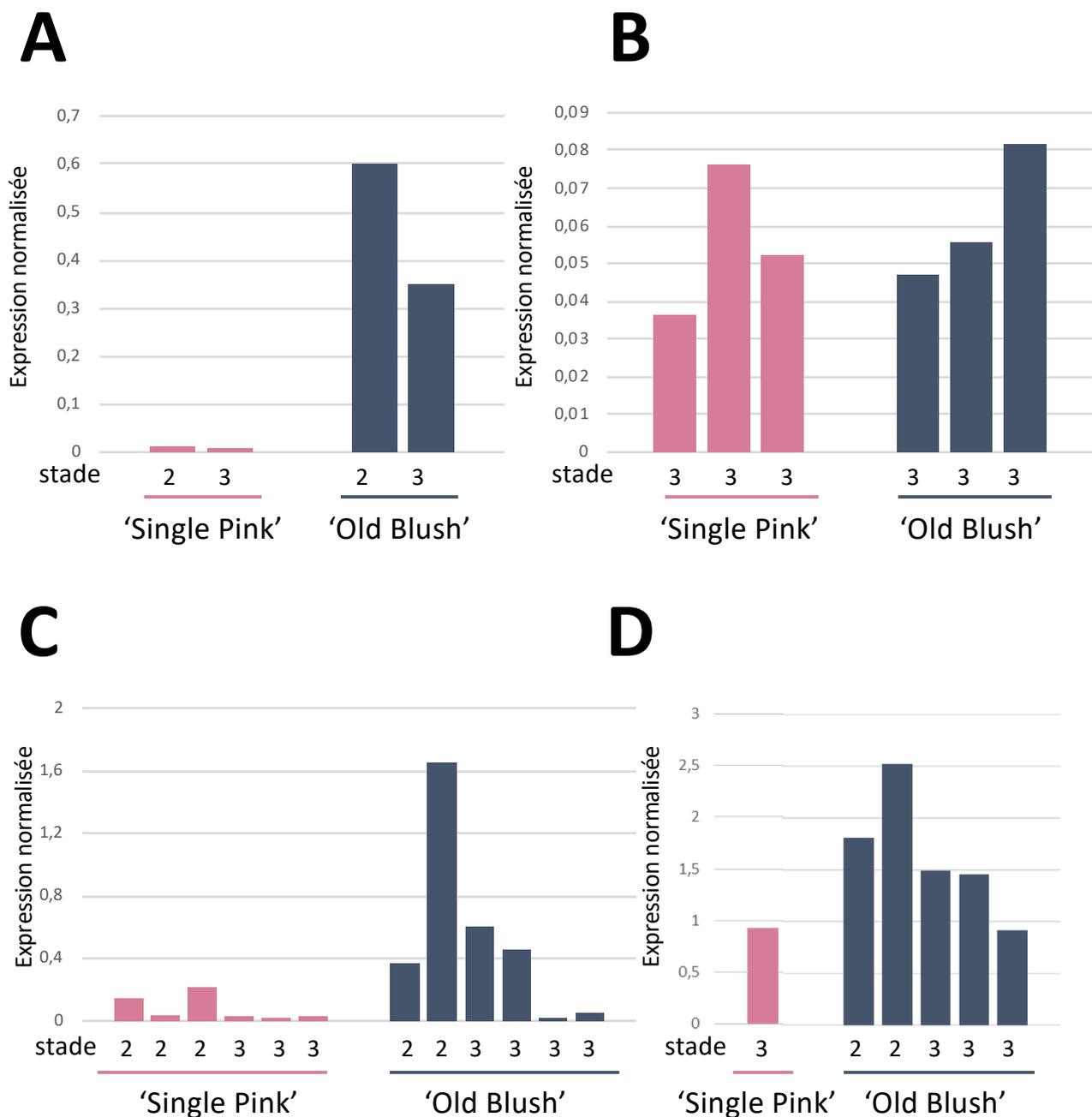


Figure 16 : Comparaison du niveau d'expression de *RcUGO* chez 'Old Blush' et chez 'Single Pink'.

RT-qPCR réalisées sur des ADNc obtenus à partir d'ARN extraits de jeunes bourgeons floraux à différents stades.

(A) et (C): ARNm extraits par Xiao-Peng Fu, RT-qPCR réalisées par Léa François.

(B) et (D): Extraction d'ARNm et RT-qPCR réalisées par Léa François. Les niveaux d'expression ont été normalisés sur les gènes de ménage *RcTCTP*, *RcGAPDH* et *RcTUB* pour les expériences (B), (C) et (D) et grâce au seul gène de ménage *RcTCTP* pour l'expérience (A). La différences de niveaux d'expression observés pour *RcUGO* chez 'Single Pink' entre les deux expérimentatrices ne permet pas de conclure quant au niveau d'expression de *RcUGO* chez cette variété.

Stade 2 : initiation des pétales ; stade 3 : initiation des étamines.

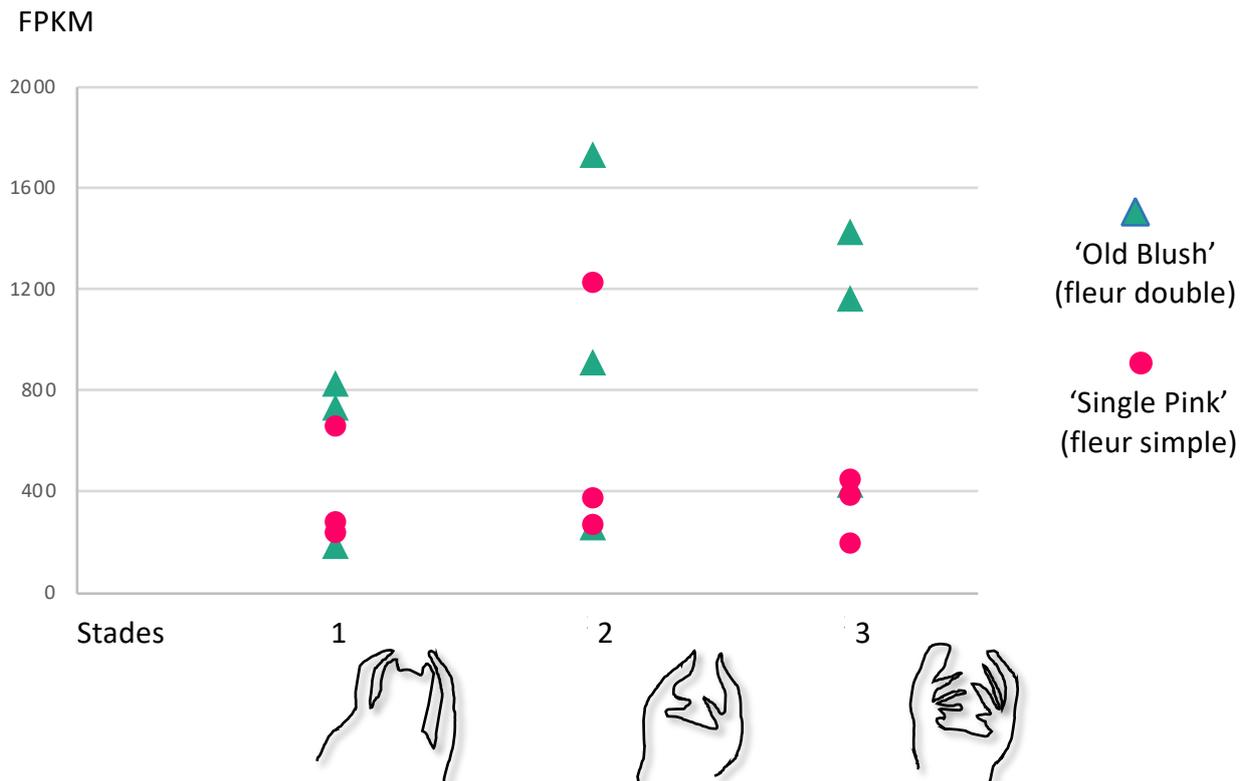


Figure 17 : Les niveaux d'expression de *RcUGO* chez 'Old Blush' et 'Single Pink' sont proches.

Niveaux d'expression de *RcUGO* déterminés à partir des données de RNA-seq obtenues lors des stades précoces du développement floral de la variété à fleurs doubles 'OB' et de la variété à fleurs simples 'SP'.

Les réplicats biologiques sont présentés séparément.

Stade 1 : initiation des sépales; stade 2 : initiation des pétales; stade 3 : initiation des étamines.

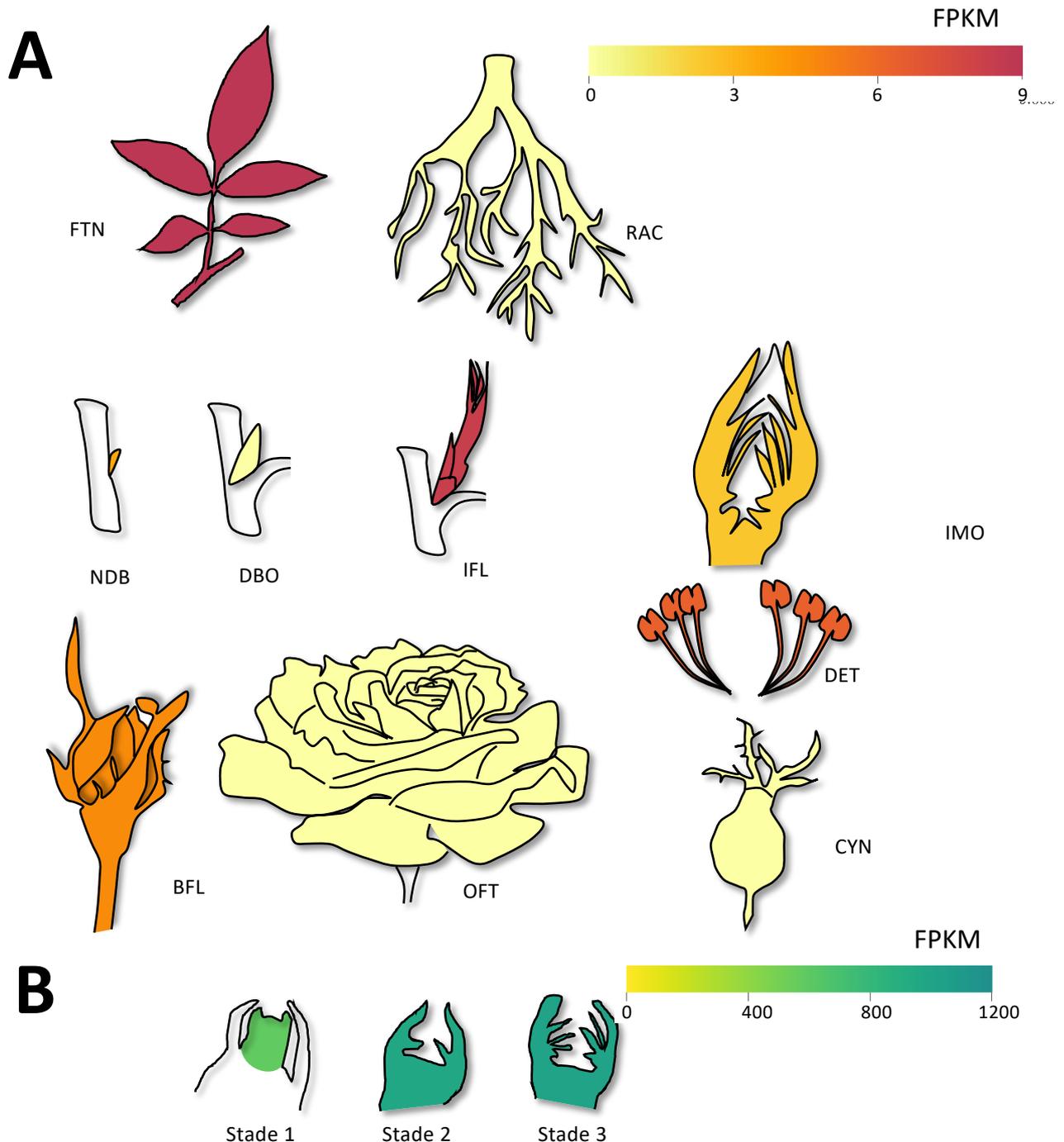


Figure 18 : Patron d'expression de *RcUGO*.

(A) RNA-seq réalisé sur différents organes de *R. chinensis* cv 'Old Blush' et décrit dans Dubois et al. (2012). FTN : jeunes feuilles; RAC : jeunes racines; NDB : bourgeons axillaires dormants; DBO: bourgeons axillaires actifs; IFL : bourgeons floraux lors de la transition vers un méristème floral; IMO : méristèmes floraux et développement précoce des organes floraux; BFL : bouton floral fermé; DET : étamines aux stades de microsporogénèse et microgamétogénèse; OFT : fleurs épanouies; SEN : fleurs sénescents; CYN : cynorhodons.

(B) RNA-seq réalisé sur de très jeunes bourgeons en séparant les stades où se forment respectivement les primordia de sépales (stade 1), les primordia de pétales (stade 2) et les primordia d'étamines (stade 3).

N.B. Les deux expériences ayant été réalisées indépendamment, les niveaux d'expression en (A) et (B) ne sont pas comparables, ce qui justifie le choix de couleurs différentes pour les deux figures.

fraîchement récoltés par moi-même (Figure 16-B,D) n'a pas confirmé ces résultats puisqu'aucune différence du niveau d'expression n'a pu être observée entre les deux variétés. Il est possible de formuler quelques hypothèses quant à ce résultat surprenant : tout d'abord, le niveau d'expression de *RcUGO* pourrait être très labile et pourrait varier rapidement lors du développement floral. Ensuite, il faut savoir qu'aux stades 1 et 2 les bourgeons sont uniquement disséquables sous la loupe binoculaire, et que pour disposer de suffisamment de matériel biologique, jusqu'à 40 méristèmes doivent être disséqués et poolés pour chaque extraction d'ARN. Ainsi, alors que le développement du bourgeon se fait de façon continue, la délimitation de stades est un peu arbitraire et il peut y avoir des erreurs ou un biais lié à l'expérimentateur. Comme les mesures sont au final des valeurs moyennes sur plusieurs méristèmes, la notion de variabilité entre méristèmes individuels est perdue. Enfin, la conservation des échantillons durant plusieurs années à -20°C pourrait avoir induit une dégradation des ARNm, en particulier des petits ARNm comme celui de *RcUGO*, et le travail sur du matériel fraîchement récolté (Figure 16-B,C) serait plus fiable.

À ce stade du projet, l'analyse du RNA-seq était encore préliminaire car, notamment, le génome était encore en cours d'annotation. Une fois l'annotation des gènes terminée et validée, Jérémy Just (ingénieur de recherche dans l'équipe) et moi-même avons repris tout le traitement des données avant d'aller plus loin. Il s'est alors avéré que lors de la première analyse, une étape de normalisation selon la taille des bibliothèques avait été omise : les bibliothèques d'Old Blush' comptant par hasard davantage de reads que celles de 'Single Pink', un biais avait été introduit. Après notre nouveau traitement des données de séquençage, les résultats présentés en Figure 17 ont été obtenus. On peut y observer une légère augmentation du niveau d'expression entre les stades les plus avancés d'Old Blush' et de 'Single Pink', ce qui est cohérent avec la tendance observée dans les RT-qPCR mais aussi, confirmer la forte variabilité des résultats observés par RT-qPCR au sein des réplicats du RNA-seq. Ainsi, alors que les analyses statistiques ne trouvent pas de différence significative entre ces variétés, il est difficile de confirmer la différence d'expression de *RcUGO* entre 'Old Blush' et 'Single Pink'.

5. Patron d'expression de *RcUGO* dans les bourgeons floraux

Les données de RNA-seq (Dubois et al., 2010) indiquent que *RcUGO* est exprimé dans une grande variété de tissus, en particulier dans les organes aériens comme les feuilles, les bourgeons végétatifs et les bourgeons floraux (Figure 18). Afin de mieux caractériser son patron d'expression dans les bourgeons floraux, des hybridations *in situ* (ISH) ont été réalisées sur de jeunes bourgeons d'Old Blush' et 'Single Pink' à différents stades de développement (Figure 19-A), avec l'aide de XiaoPeng Fu (ancienne post-doctorante). Ainsi, il s'avère que le patron d'expression de *RcUGO* est similaire chez les deux variétés, et que l'expression est potentiellement plus faible chez 'Single Pink' bien que les hybridations *in situ* ne soient pas des expériences quantitatives. En premier lieu, *RcUGO*

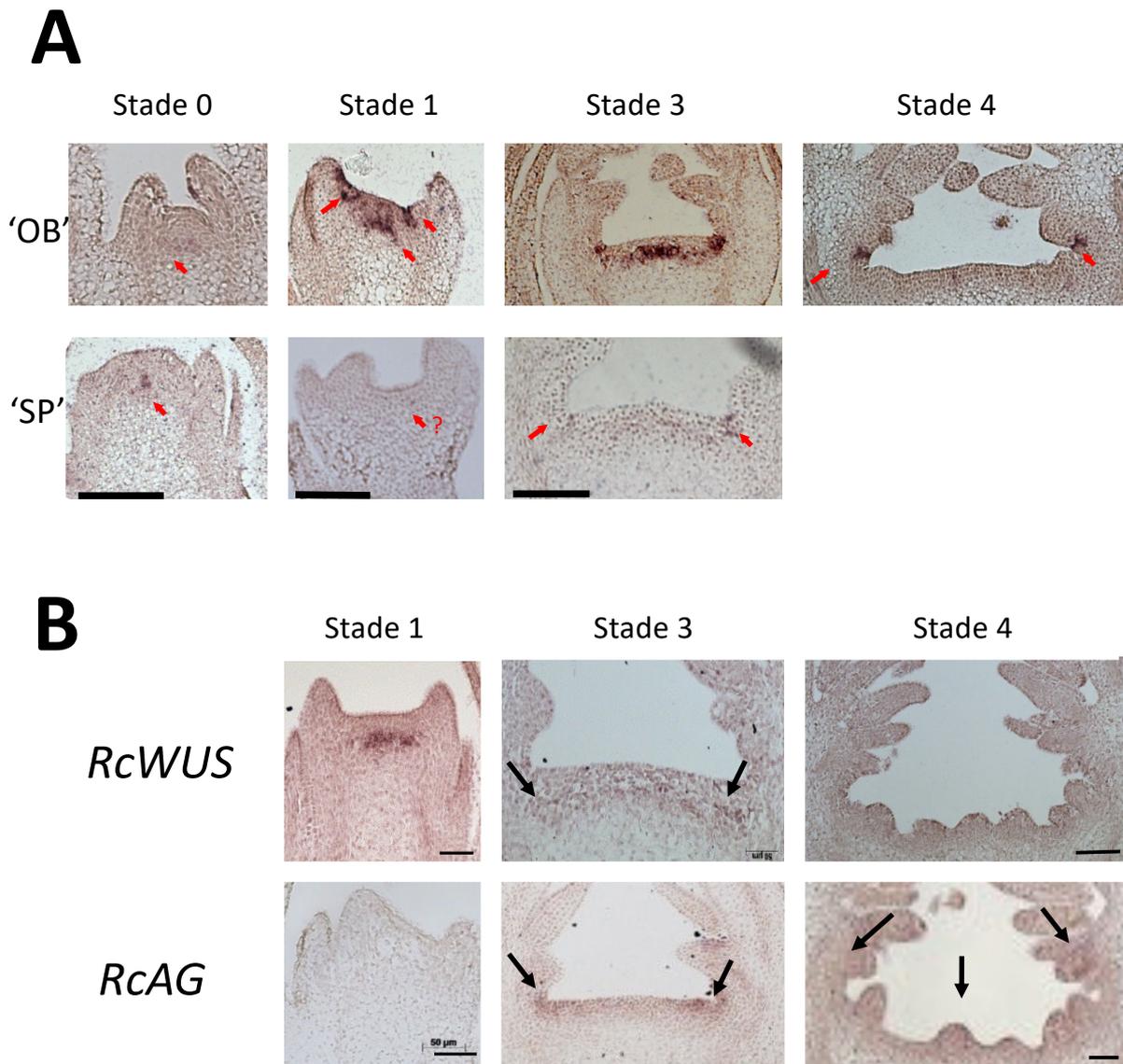


Figure 19 : Le domaine d'expression de *RcUGO* recoupe celui de *RcWUS* et *RcAG*.

Hybridation *in situ* sur des coupes transversales de méristèmes floraux à différents stades de développement.

(A) Sonde dirigée contre *RcUGO* sur des méristèmes provenant des variétés *Rosa chinensis* cv 'Old Blush' (fleurs doubles) et *R. chinensis* cv 'Single Pink' (fleurs simples).

(B) Méristèmes provenant de 'Old Blush', marqués avec des sondes dirigées contre *RcWUS* et *RcAG*, deux gènes-clés de l'organisation florale et potentiellement de la formation de fleurs doubles. Ces images ont été réalisées par ou avec l'aide de Xiaopeng Fu (ancienne post-doctorante de l'équipe).

Stade 0 : initiation du méristème floral; stade 1 : initiation des sépales; stade 2 : initiation des pétales;

stade 3 : initiation des étamines; stade 4 : initiation des carpelles.

Echelle : 50 μ m.

est exprimé au centre du méristème floral, dans quelques cellules centrales de la zone quiescente avant de s'étaler vers la périphérie de la zone méristématique. Dès le stade 2, au moment de l'initiation des primordia de pétales, *RcUGO* commence aussi à s'exprimer dans les cellules des couches les plus superficielles, mais à la périphérie du méristème, légèrement à la base des primordia en cours d'initiation. Ce deuxième domaine d'expression, en forme d'anneau organisé en périphérie du méristème, pourrait correspondre à la frontière entre le méristème à proprement parler et les organes en cours de développement. L'ensemble du domaine d'expression de *RcUGO* semble proche de celui de *REV* chez *A. thaliana* (Emery et al., 2003 ; Otsuga et al., 2001), en particulier au centre du méristème, ce qui est cohérent avec l'activation de l'expression des ZPR par les HD-ZIP III (Wenkel et al., 2007).

D'autre part, il a été remarqué qu'il recouvrait aussi les domaines d'expression de certains gènes clés de l'organisation florale chez *A. thaliana*, en particulier ceux des orthologues de *WUS* et *AG*. En effet, *RcWUS* est exprimé dès le stade 0, au centre du méristème floral en cours d'initiation, et se maintient jusqu'au stade 3 où son expression commence à diminuer (Figure 19-B). Il s'exprime tout d'abord dans les quelques cellules quiescentes très centrales, au niveau des couches cellulaires 4 à 6, puis s'étend en même temps que le centre du méristème s'élargit et s'aplatit (stades 2 et 3). Il semblerait donc que le territoire d'expression de ce facteur de transcription nécessaire au maintien des cellules souches, et dont le patron d'expression spatial et temporel est à l'origine du contrôle du nombre d'organes produits par chaque méristème (Brand et al., 2000 ; Ikeda et al., 2009 ; Laux et al., 1996), soit inclus dans celui de *RcUGO*, suggérant un lien (direct ou indirect) entre les deux gènes. D'autre part, *RcAG* commence à être exprimé au stade 3 au niveau de l'anneau à la périphérie du méristème où se trouve aussi *RcUGO*. Il s'étend ensuite au centre du méristème où il inhibe probablement l'expression de *RcWUS* directement (Liu et al., 2011) ou indirectement en activant la transcription de *KNUCKLES (KNU)*, un répresseur *WUS* chez *A. thaliana* (Sun et al., 2009). Ensemble, ces résultats pourraient suggérer que *RcUGO* est potentiellement un nouvel acteur dans ce réseau d'interaction et pourquoi pas dans la détermination du nombre de pétales chez le rosier. Cependant, nous ne savons toujours pas comment *RcUGO* pourrait être impliqué dans cette interaction, soulignant la nécessité d'études fonctionnelles.

6. La diminution d'expression de *RcUGO* n'induit pas de modification du nombre de pétales chez 'Old Blush'

Afin de vérifier le rôle de *RcUGO* dans le développement floral, il était important d'essayer de modifier le niveau d'expression de ce gène afin de regarder si le phénotype était altéré. Étant donné qu'il était envisagé que c'était la surexpression de *RcUGO* qui induisait le passage d'un phénotype fleur simple vers un phénotype fleur double, nous avons tout d'abord décidé d'inhiber l'expression de *RcUGO* chez la variété modèle à fleurs doubles 'Old Blush' afin de voir si cela induisait une réversion

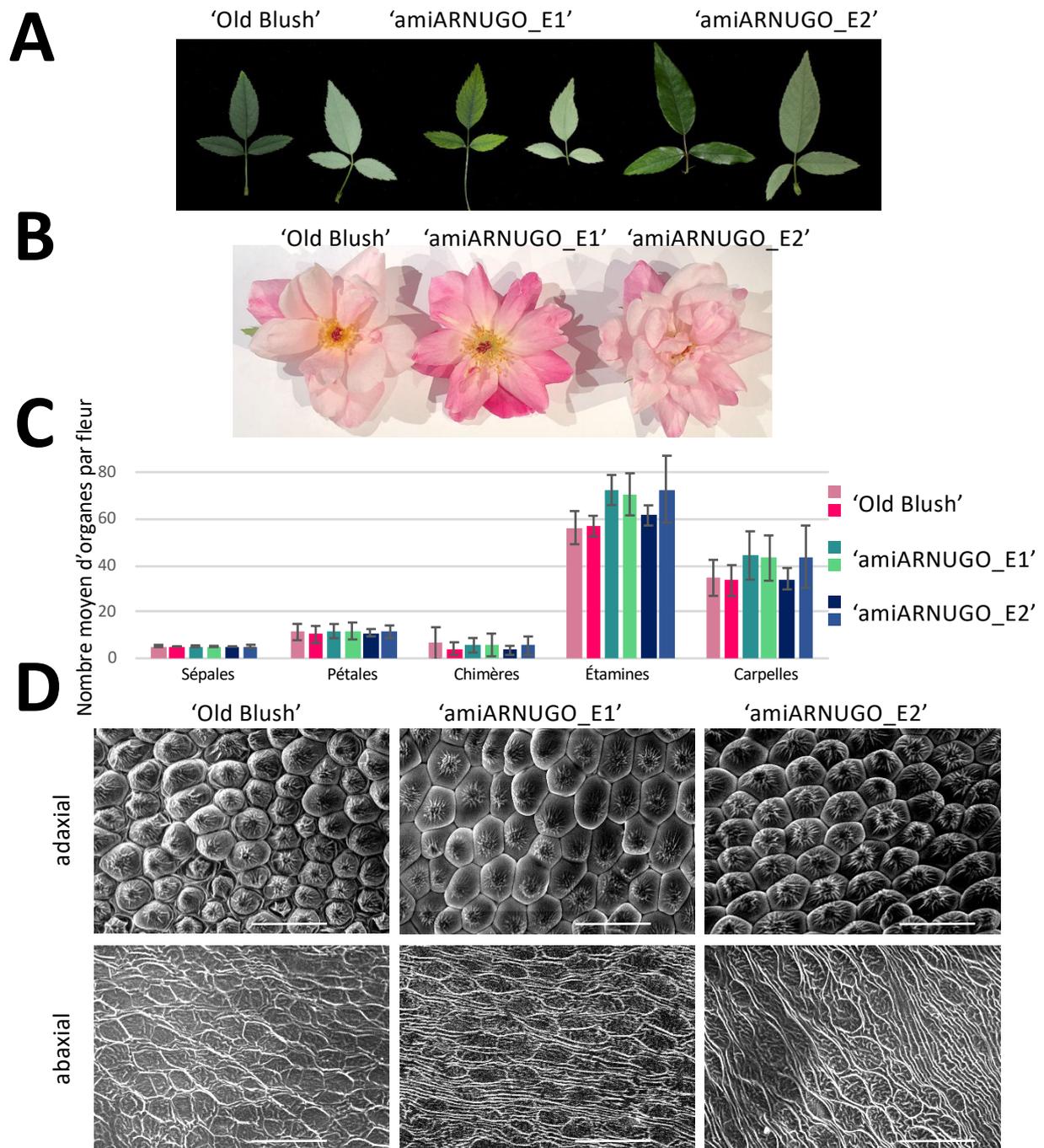


Figure 20 : L'inhibition de *RcUGO* chez 'Old Blush' ne modifie pas le nombre d'organes floraux ou la polarité adaxiale-abaxiale des organes aériens.

(A) et (B) Photographies des deux faces de feuilles (A) et de fleur (B) sauvage ou exprimant un microARN dirigé contre *RcUGO*.

(C) Nombre d'organes par fleur sauvage ou exprimant un microARN dirigé contre *RcUGO* (analyse de deux plantes par génotype).

(D) Image au microscope électronique à balayage des deux faces d'un pétale de fleurs sauvage ou exprimant un microARN dirigé contre *RcUGO*. Aucune différence n'a été identifiée entre les pétales sauvages et mutants : les cellules coniques sont bien visibles sur la face adaxiale alors qu'elles ne sont pas sur la face abaxiale.

(barres d'échelle: 30 μ m)

du phénotype. Pour cela, la technique des microARN artificiels a été utilisée. Les microARN ont d'abord été étudiés comme un phénomène naturel de régulation de l'expression des gènes *in planta* (Llave et al., 2002) : ils correspondent à des petits ARN formant une structure tige-boucle qui est ensuite clivée par un complexe enzymatique pour libérer un ARN simple brin de 21 nucléotides capable de reconnaître une séquence complémentaire dans le transcrit d'un gène et d'en induire la dégradation. Ce mécanisme a été détourné comme outil de laboratoire en changeant artificiellement la séquence de 21 nucléotides, pour cibler un gène d'intérêt (Ossowski et al., 2008 ; Schwab et al., 2006), sans avoir à cribler des milliers de lignées pour obtenir une perte de fonction du gène d'intérêt. Cette méthode permet aussi d'inhiber des gènes sans pour autant les rendre complètement non fonctionnels, ce qui peut être un avantage dans le cas des gènes dont la perte de fonction complète est létale.

Un protocole permettant le choix des séquences des microARN artificiels ciblant les gènes d'intérêt grâce à l'interface web WMD3 (Schwab et al., 2006) ainsi que leur insertion dans le backbone du miR319 caractérisé chez *A. thaliana* était déjà établi dans l'équipe, comme décrit dans la section Matériel et Méthodes (Figure 38). En revanche, il était à l'époque impossible de déterminer si les microARN choisis pouvaient aussi influencer l'expression d'autres gènes de rosier possédant une séquence cible similaire, aussi appelés off-targets. Pour remédier à cela, j'ai installé et paramétré une instance locale du programme WMD3 sur une machine virtuelle accessible à toute l'équipe et permettant la recherche des off-targets sur le génome du rosier. Étant donné les nombreux mois de travail liés à la transformation du rosier, ceci est une étape importante pour réduire les risques liés à l'utilisation de microARN inhibant plusieurs gènes et donc pour lesquels l'analyse phénotypique des rosiers transformés génétiquement risque d'être biaisée par les off-targets. Pour cette étude, deux constructions spécifiques de *RcUGO* ont été placées sous le promoteur fort du virus de la mosaïque du chou-fleur 35S (Odell et al., 1988) et introduites chez 'Old Blush' par transformation génétique médiée par *A. tumefaciens*. Le premier a permis de régénérer deux événements de transformation indépendants, dont le nombre d'organes floraux a été compté et comparé à celui de plants d'Old Blush' ayant été cultivés *in vitro* et régénérés en parallèle des transformants. Ainsi, il semblerait que tous les rosiers, transformés ou non, produisent en moyenne 5 sépales, une dizaine de pétales, une moyenne de 5 chimères étamines/pétales, une soixantaine d'étamines et autour de 35 carpelles par fleur. Par conséquent, l'utilisation de microARN artificiels pour inhiber l'expression de *RcUGO* ne paraît pas modifier le nombre de pétales par fleur.

Étant donné le rôle important des ZPR dans la différenciation dorso-ventrale des organes chez *A. thaliana*, les deux faces des feuilles et pétales ont été étudiées afin de vérifier que l'asymétrie dorso-ventrale de ces organes était maintenue chez les rosiers mutants. Une observation macroscopique permet déjà de supposer que c'est le cas, étant donné la forte différenciation entre les faces adaxiales, plus sombres, plus riches en cuticule et abaxiales, plus claires, des feuilles. Similairement, la forme des

pétales semble aussi indiquer une structuration normale de ceux-ci. Cela a été confirmé au microscope électronique à balayage, où la présence de cellules coniques a pu être identifiée seulement sur la face adaxiale.

Ainsi, il semblerait que l'inhibition de *RcUGO* ne soit pas suffisante pour voir une modification phénotypique chez le rosier, suggérant peut-être une redondance fonctionnelle avec le deuxième ZPR présent chez le rosier, *RcHm_v2.0_Chr1g0350821*. Ainsi, il serait intéressant de répéter cette expérience en ciblant les deux gènes simultanément ou, au contraire, de surexprimer *RcUGO* chez un rosier à fleur simple.

7. La surexpression de RcUGO chez 'Old Blush' n'a pas permis de régénérer de plantes transformées génétiquement

Il serait en effet intéressant de vérifier si la surexpression de *RcUGO* chez une fleur simple est suffisante pour induire le passage d'une fleur simple à une fleur double. Malheureusement, à l'époque de cette étude, aucun cal embryogène n'avait pu être établi en culture *in vitro* pour une variété à fleurs simples, étape indispensable à la transformation génétique. Afin d'obtenir des résultats complémentaires à l'expérience de diminution de l'expression de *RcUGO* par microRNA chez 'Old Blush', nous avons choisi d'aussi surexprimer *RcUGO* chez 'Old Blush', afin de mieux comprendre sa fonction et notamment de voir s'il inhibait réellement la fonction des HD-ZIP de classe III. Les fleurs d''Old Blush', bien que doubles, n'ont qu'environ 20 pétales, contre 80 et plus pour d'autres variétés, ce qui laisserait potentiellement encore la possibilité de voir une augmentation du nombre d'organes. Pour cela, la séquence codante de *RcUGO* a été clonée et placée sous un promoteur 35S et introduite dans le génome d''Old Blush' par transformation génétique médiée par *A. tumefaciens*. Suivant le protocole habituel développé dans l'équipe, plus d'une centaine de cals ont été transformés, ce qui permet usuellement d'obtenir au moins une dizaine d'événements de transformation indépendants. Cependant, lors de cette transformation, de nombreuses proliférations *ex planta* d'*A. tumefaciens* ont contaminé les boîtes et entraîné la perte de nombreux événements. Les quatre événements restants semblent résistants aux efforts pour les régénérer sous forme de plantules, et cela malgré de nombreux essais d'induction de la formation de feuilles. On observe en particulier la présence de cals blancs translucides au lieu des cals verts habituels. Plusieurs hypothèses pourraient expliquer cette observation. Tout d'abord, les cals ont pu être trop stressés et sont maintenant incapables de remettre en place un programme génétique cohérent permettant de produire des organes complets ; la couleur anormale des cals supporterait cette hypothèse. Ou bien, les cals sélectionnés étant très certainement ceux qui expriment le plus fortement le transgène, il est possible que la surexpression de *RcUGO* soit toxique pour les plantes et ralentisse fortement leur croissance. Dans l'hypothèse où *RcUGO* se dimériserait avec les HD-ZIP III et les inactiverait, on pourrait attendre à ce que ceux-ci ne puissent plus

jouer leur rôle, décrit chez *Arabidopsis*, dans la polarisation des organes végétatifs (McConnell et al., 2001) et des tissus conducteurs (Emery et al., 2003), d'où l'impossibilité de différencier des feuilles. Cette hypothèse est notamment cohérente avec le rôle connu des ZPR dans l'organisation des méristèmes chez *A. thaliana* (Green et al., 2005) et le fait que, chez le *Ginkgo biloba*, la surexpression de l'unique GbZPR trouvé induit un retard de croissance allant parfois jusqu'à la terminaison du méristème (Floyd et al., 2014).

8. Conclusions

Cette étude a fourni de nouvelles informations sur *RcUGO*, un gène de la famille des ZPR. En particulier, son patron d'expression dans les méristèmes floraux chevauche celui de *RcAG* et *RcWUS*, ce qui suggère un rôle dans le fonctionnement du méristème et notamment dans le contrôle de l'activité méristématique. En effet, *WUS* est un facteur de transcription promouvant le maintien des cellules souches et l'activité du méristème floral et dont l'expression est en partie contrôlée par *AG*. Ainsi, la perte de fonction d'*AG* induit un prolongement de l'expression de *WUS* (Lenhard et al., 2001). *AG* semble pouvoir inhiber *WUS* directement et indirectement : indirectement en activant l'expression de *KNUCKLES*, un autre inhibiteur de *WUS*, et directement en reconnaissant des séquences régulatrices dans le promoteur de *WUS* et en recrutant les complexes inhibiteurs Polycomb qui méthylent les histones 3 en position K27 (H3K27). Cela compacte l'ADN à proximité et réduit l'accès du promoteur aux protéines permettant la transcription (Liu et al., 2011). Ainsi, le patron d'expression de *RcUGO* pourrait suggérer qu'il est impliqué, d'une façon encore inconnue, dans cette boucle de régulation. D'autre part, ce patron d'expression original de *RcUGO* dans le méristème floral au stade de la détermination de l'identité des organes floraux pourrait faire de son promoteur un outil particulièrement intéressant pour de futurs projets.

Cependant, à l'heure actuelle, il est encore difficile de vérifier la fonction jouée par *RcUGO* chez le rosier. En effet, alors que l'inhibition de l'expression de *RcUGO* ne semble pas induire de phénotype visible, des études fonctionnelles plus approfondies auraient été nécessaires. En particulier, la réalisation d'une perte de fonction complète par la méthode CRISPR/Cas9, une surexpression de *RcUGO* sous un promoteur spécifique des méristèmes ou des complémentations du double *zrp3/zpr4* chez *A. thaliana* auraient pu nous permettre d'y voir plus clair. Cependant, alors qu'un deuxième candidat pour le déterminisme de la fleur double a été identifié et supporté par davantage de données préliminaires, il a été décidé de concentrer mes recherches sur ce dernier, comme cela est développé dans les deux sections suivantes.

II. Identification d'un second candidat pour le déterminisme de la fleur double : le cas de *RcAP2L*

1. Rappel du contexte à propos du gène de la fleur double chez 'Old Blush'

Comme développé durant l'introduction, il a été montré que, chez le rosier, le passage d'un phénotype « fleur simple » à un phénotype « fleur double » est déterminé par une mutation unique localisée sur le bras long du chromosome 3 (Figure 10). Fait important, d'après plusieurs études génétiques indépendantes, il semblerait que ce soit la même région qui soit mutée chez de nombreux rosiers cultivés (Crespel et al., 2002 ; Debener, 1999 ; Dugo et al., 2005 ; Schupert, 2005 ; Spiller et al., 2010).

En 1999, l'analyse d'une descendance d'un croisement entre un rosier à fleur simple et un rosier à fleur double avait montré qu'une moitié de descendants produisaient des « fleurs doubles » et l'autre moitié arboraient des « fleurs simples » (Debener, 1999) indiquant qu'une mutation dominante est à l'origine de l'apparition de la fleur double chez le rosier. Ultérieurement, d'autres études génétiques ont montré que le croisement entre 'Old Blush' et une variété ancestrale à « fleur simple » telle que *Rosa wichurana* produit également une moitié de descendants à « fleur double » et une autre moitié à « fleur simple » et que le rétrocroisement d''Old Blush' avec un de ses descendants à « fleur simple » conduit au même ratio (Figure 21 ; Schupert, 2005). Il en a été conclu qu''Old Blush' était hétérozygote pour la mutation. Ainsi, nous avons pu chercher, après reconstruction de l'intervalle contenant le déterminant génétique à l'origine du phénotype « fleur double », des gènes présents à l'état hétérozygote dans le génome d''Old Blush'. Pour ce faire, nous nous sommes intéressés en priorité aux mutations facilement identifiables, notamment l'insertion d'éléments transposables, qui sont connus pour être une véritable force motrice pour l'évolution génétique et phénotypique des organismes (Kidwell and Lisch, 1997 ; Nitasaka, 2003 ; Wessler, 1988).

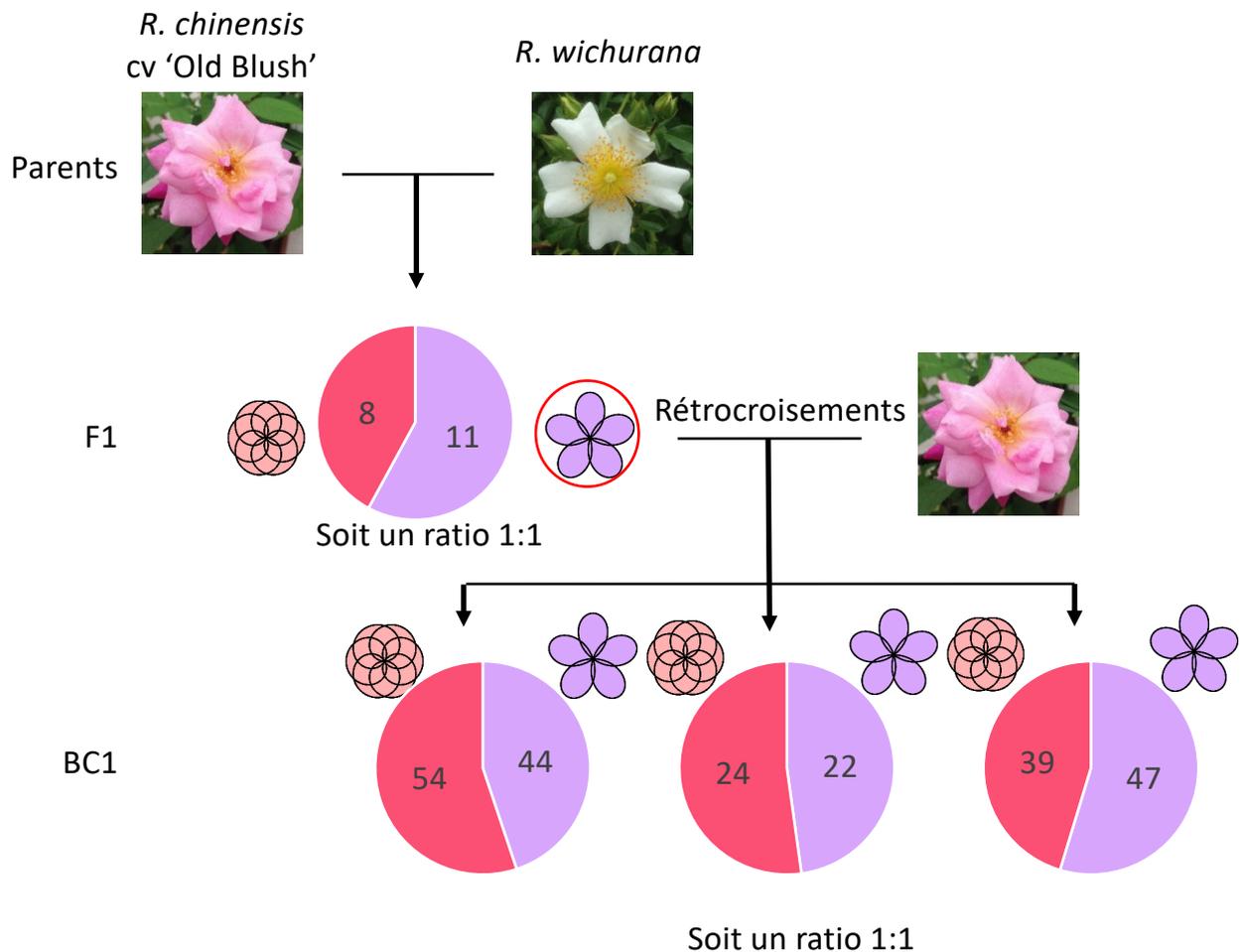


Figure 21: Le rosier modèle 'Old Blush' est hétérozygote pour la mutation à l'origine du phénotype « fleur double ».

Résumé de l'étude génétique réalisée par D. A. Shupert durant sa thèse, *Inheritance of flower, stem, leaf, and disease traits in three diploid interspecific rose populations* (2005).

Le croisement entre *Rosa chinensis* 'Old Blush' et la fleur simple *R. wichurana* produit des plantes à fleurs doubles et des plantes à fleurs simples en proportions cohérentes avec un rapport 1/2 versus 1/2. Lorsque les fleurs simples issues de ce croisement sont à leur tour croisées avec 'Old Blush', le rapport 1/2 versus 1/2 est conservé.

2. *AP2L*, un candidat pertinent pour la formation de la fleur double

Résumé de l'article : Alors que les roses sauvages n'ont que cinq pétales, la nature de la mutation causant le passage de cette fleur simple à une fleur double possédant plus de 20 pétales, restait inconnue. Jusqu'à présent, l'analyse des taux de ségrégation et cartes génétiques indiquait que le trait « fleur double » était dû à une seule mutation, dominante, hétérozygote chez 'Old Blush', située sur le chromosome 3 (Schupert, 2005 ; Spiller et al., 2010) et que cette mutation était à l'origine d'une restriction du territoire d'expression du gène *RcAGAMOUS* (*RcAG*) vers le centre du méristème floral (Dubois et al., 2010). À partir du génome de *Rosa chinensis* cv 'Old Blush' nouvellement séquencé, assemblé et publié par l'équipe (Raymond et al., 2018, *Nature Genetics*), il a été possible de reconstruire la séquence des deux haplotypes sur l'intervalle génétique contenant le locus « fleur double ». Parmi les 631 gènes de cet intervalle, un gène de la famille des euAP2 (*APETALA2-LIKE* ou *RcAP2L*) est apparu comme candidat répondant à tous les critères. En effet, il a été observé que ce gène existait sous la forme de deux allèles différents, l'un de type sauvage et l'autre contenant un grand élément transposable. L'étude de différentes variétés de rosiers à fleurs simples et à fleurs doubles a ensuite permis d'identifier une corrélation entre cette mutation et le phénotype « fleur double » chez les rosiers d'origine chinoise. Enfin, l'analyse de données de RNA-seq a montré que la transcription de cette forme mutée produit un ARN messager tronqué, résistant à l'inhibition normale par miR172. Alors que chez *A. thaliana*, AP2 est connu pour inhiber l'expression d'AG et que la surexpression d'un variant résistant au miR172 entraîne souvent la formation de pétales supplémentaires (Chen, 2004), nous avons pu proposer un modèle permettant d'expliquer la formation des fleurs doubles chez le rosier. Selon ce modèle, la protéine *RcAP2L* ^{Δ 172} s'accumulerait du fait de la résistance de son ARNm au miR172 et restreindrait davantage l'expression de *RcAG* au centre du méristème floral. Ainsi, la frontière entre les domaines A et C se trouverait elle aussi déplacée vers le centre du méristème, ce qui induirait la conversion des étamines en pétales.

SCIENTIFIC REPORTS

OPEN A miR172 target-deficient AP2-like gene correlates with the double flower phenotype in roses

Léa François¹, Marion Verdenaud^{1,2}, Xiaopeng Fu^{1,3}, Darcy Ruleman¹, Annick Dubois¹, Michiel Vandebussche¹, Abdelhafid Bendahmane², Olivier Raymond¹, Jérémy Just¹ & Mohammed Bendahmane¹

Received: 29 May 2018

Accepted: 6 August 2018

Published online: 27 August 2018

One of the well-known floral abnormalities in flowering plants is the double-flower phenotype, which corresponds to flowers that develop extra petals, sometimes even containing entire flowers within flowers. Because of their highly priced ornamental value, spontaneous double-flower variants have been found and selected for in a wide range of ornamental species. Previously, double flower formation in roses was associated with a restriction of *AGAMOUS* expression domain toward the centre of the meristem, leading to extra petals. Here, we characterized the genomic region containing the mutation associated with the switch from simple to double flowers in the rose. An *APETALA2*-like gene (*RcAP2L*), a member of the Target Of EAT-type (TOE-type) subfamily, lies within this interval. In the double flower rose, two alleles of *RcAP2L* are present, one of which harbours a transposable element inserted into intron 8. This insertion leads to the creation of a miR172 resistant *RcAP2L* variant. Analyses of the presence of this variant in a set of simple and double flower roses demonstrate a correlation between the presence of this allele and the double flower phenotype. These data suggest a role of this miR172 resistant *RcAP2L* variant in regulating *RcAGAMOUS* expression and double flower formation in *Rosa* sp.

Roses are widely used as garden ornamental plants and cut flowers worldwide. A number of their agricultural and decorative traits specify their commercial value¹ and have been selected during domestication. Examples of these important traits are recurrent flowering, double flowers, petal colour and fragrance². Double flower refers to a characteristic of modern roses giving blooms with an increased number of petals that can vary from 10 to more than 200 petals per flower, whereas wild-type simple flowers are composed of 5 petals. This characteristic is tightly associated with flower development and organ identity patterning, as it results from homeotic conversion of stamens into petals³. However, the underlying molecular mechanisms are not fully understood in roses, or in other non-model species.

In the past three decades, most of the genetic and molecular networks controlling floral development have been extensively studied in model species such as *Arabidopsis thaliana* and *Antirrhinum majus*. These studies led to the establishment of the ABC model of flower development^{4,5}. In this model, the combinatorial actions of four classes of homeotic genes (A, B, C and E) determine flower organ identity. Briefly, from the outer to the inner whorl of the floral meristem, the A-class genes (*APETALA1*, *API*; *APETALA2*, *AP2*) alone determine sepal formation; the A-class genes together with the B-class genes (*PISTILLATA*, *PI*; *APETALA3*, *AP3*) determine petal fate, the C-class gene (*AGAMOUS*, *AG*) associated with the B-class genes specify stamen formation, and finally the C-class gene determines carpel fate. E-class genes are necessary for all floral organ identity. A-class genes have also an antagonistic role toward the expression of the C-class gene *AG*, and vice versa. This leads to the expression of the A-class genes in the sepal and petal whorls and of *AG* in the stamen and carpel whorls. In *Arabidopsis*, *AG* loss-of-function leads to over-accumulation of A-class genes in the third whorl and homeotic conversion of stamens into petals⁶. Similarly, over-accumulation of AP2 protein leads to a reduced expression of *AG* in the third whorl and a similar homeotic conversion of stamens into petals⁷. This conceptual framework for floral organ

¹Laboratoire Reproduction et Développement des Plantes, Univ Lyon, ENS de Lyon, UCB Lyon 1, CNRS, INRA, F-69364, Lyon, France. ²Institute of Plant Sciences Paris-Saclay IPS2, CNRS, INRA, University Paris-Sud, University of Evry, University Paris-Diderot, Sorbonne Paris-Cite, University of Paris-Saclay, F-91405, Orsay, France. ³Key Lab of Horticultural Plant Biology, College of Horticulture & Forestry Sciences, Huazhong Agricultural University, Wuhan, China. Correspondence and requests for materials should be addressed to M.B. (email: mohammed.bendahmane@ens-lyon.fr)

identity patterning is broadly valid for flowering plant species that have been studied^{8,9}. However, during evolution, some genes underwent duplication and neo- or sub-functionalization, leading to small differences in their regulatory interactions. For example, the canonical C-function, performed by *AG* in *Arabidopsis*, is carried out by *PLENA* in *Antirrhinum*, that is orthologous to the *Arabidopsis* *SHATTERPROOF* genes (*SHP*)¹⁰. In *Petunia*, the restriction of the C-class gene expression needs mainly the actions of the microRNA *BLIND*, but involves a gene from the euAP2 family, *PhBEN*^{11,12}. This diversity of the canonical ABCE functions, together with the absence of comprehensive genome data giving access to all members of each gene family, hampered the identification of the key genes determining floral organ identity in non-model species, such as in rose. Recently, efforts have been made to identify canonical rose A-, B-, C- and E-class gene orthologues, but we are still far from understanding their exact role in rose floral phenotype patterning^{3,13–19}.

Previously, we demonstrated that a downregulation and a restricted expression domain of the rose orthologue of *AGAMOUS* (*RcAG*) correlates with an increase in petal number in domesticated roses³. This was later confirmed by transient *RcAG* downregulation using Virus Induced Gene Silencing²⁰. Similar associations between *AG* expression and double flower formation were shown in other species such as Ranunculids, Cyclamen, Japanese gentian and *Prunus*^{21–24}. Yet, the molecular mechanism by which the restriction of the expression of *RcAG* occurs remains unknown. Indeed, the rose *RcAG* gene does not co-segregate with the major locus (*Df*) located on Linkage Group 3 that has been shown to control the switch from the simple flower to the double flower phenotype^{25,26}. In roses, a yet unknown gene located in the *Df* locus and acting upstream of *RcAG* must be the determinant for double flower formation³.

In order to identify the genetic determinant of the double flower phenotype, we localized and analysed the sequence of the double flower interval using the recent high-quality *Rosa chinensis* ‘Old Blush’ genome assemblies²⁷. The first corresponds to the homozygous rose assembly^{27,28} consisting of seven assembled pseudomolecules and representing a haplotype of the rose genome. The second assembly corresponds to the heterozygous *Rosa chinensis* ‘Old Blush’ consisting of 15,937 scaffolds, and provides access to the two haplotypes of the genome and information on alleles. Among the candidate genes in the interval, we identified a gene belonging to the euAP2 family, of which certain members are known to repress *AG* expression in many species^{6,12,29,30}. We show that in double flower roses this gene is present as two different alleles, one of which harbours a transposable element insertion that is never found in simple flower roses. This insertion leads to a truncated *RcAP2L* version that lacks the miR172 binding site, meaning it is no longer negatively regulated by this microRNA. The data provide a basis for a mechanism by which double flowers are formed and open new perspectives to dissect in detail the underlying molecular and biochemical mechanisms in roses and likely in other species.

Results

Localization of the genetic interval associated with the double flower phenotype. In roses, the double flower phenotype is associated with a dominant mutation in the yet unknown *Df* (*DOUBLE FLOWER*) locus. This locus was previously shown to map on LG3^{25,26}. We used the high-quality genome assembly (*RcHm*)^{27,28} to identify flanking markers that define the mapping interval containing *Df*. Flanking markers were retrieved from the previously reported genetic maps^{26,31,32} and mapped on the rose genome sequence²⁷ and those that had unique match allowed to mark out an interval of 6.2 Mb on Chromosome 3 at coordinates 13,535,933 to 19,743,495 (Fig. 1a). Genes within this interval were then retrieved using the gene annotation of the reference rose genome^{27,28} (Supplementary Table 1). The assembled interval on Chromosome 3 contained 631 annotated genes (Supplementary Table 1). Alleles for each gene were then retrieved using the genome assembly of the heterozygous genome (*RcHt*)²⁷. Previous studies showed that a modified expression pattern of *RcAG* was associated with double flower formation in rose³. *RcAG* maps on Chromosome 5 of the rose genome, thus corroborating previous data indicating that *RcAG* is not the *Df* gene³. Among the 631 annotated genes that lie within the double flower mapping interval, no gene showed similarities to *RcAG* gene. These data suggest that the gene responsible for double flower formation could be an upstream regulator of *RcAG*.

A mutant allele of an AP2-like gene lies within Double Flower interval. To narrow down the number of *Df* gene candidates, we searched within the assembled double flower interval for genes that share homologies with those known to regulate *AG* expression in *Arabidopsis* and that are present at heterozygous state in the double flower rose *R. chinensis* ‘Old Blush’. Indeed, previous genetic segregation analyses involving ‘Old Blush’ or other rose cultivars as parents showed that the double flower trait is controlled by a dominant allele at heterozygous state^{25,33–35}. Interestingly, one candidate gene had high sequence similarity to *APETALA2* (*AP2*). In *Arabidopsis*, *AP2* was shown to negatively regulate the expression of *AG* in the sepal and petal whorls, restricting its expression to the stamen and carpel whorls^{6,36}. The identified rose *AP2*-like gene (*RcAP2L*, *RcHm3g0468481*; Fig. 1c) contains 10 exons and 9 introns, and encodes for a 460 amino-acid protein. Analysis of the predicted *RcAP2L* protein showed the presence of two AP2 DNA-binding domains, indicating a similar structure to the *Arabidopsis* A-class gene *AP2*³⁷. Additionally, a miR172 binding site and two EAR motifs (Ethylene-responsive element binding factor-associated amphiphilic repression) were also found in this gene. These three features are characteristic of the euAP2 family members^{38,39}.

BLASTP of the *Arabidopsis* *AP2* protein on rose and strawberry predicted protein sequences identified four potential members of the euAP2 family in each of the species (Fig. 2). Protein sequence alignments and phylogenetic analyses using the AP2 domains of euAP2 genes from *Petunia hybrida*, *Solanum lycopersicum*, *Arabidopsis thaliana*, *Capsella rubella*, *Medicago truncatula*, *Vitis vinifera* and *Prunus persica* showed that each of the four rose predicted proteins groups with a single and unique strawberry and *Prunus* predicted protein, supporting their orthologous relationship and the quality of the tree (Fig. 2). Bootstrap values highly support the presence of a single rose member of the AP2-type subfamily (*RcHm2g0106221*). The remaining three rose euAP2 members, including *RcAP2L*, likely belong to the Target Of EAT-type (TOE-type) subfamily (Fig. 2). The rose TOE-type

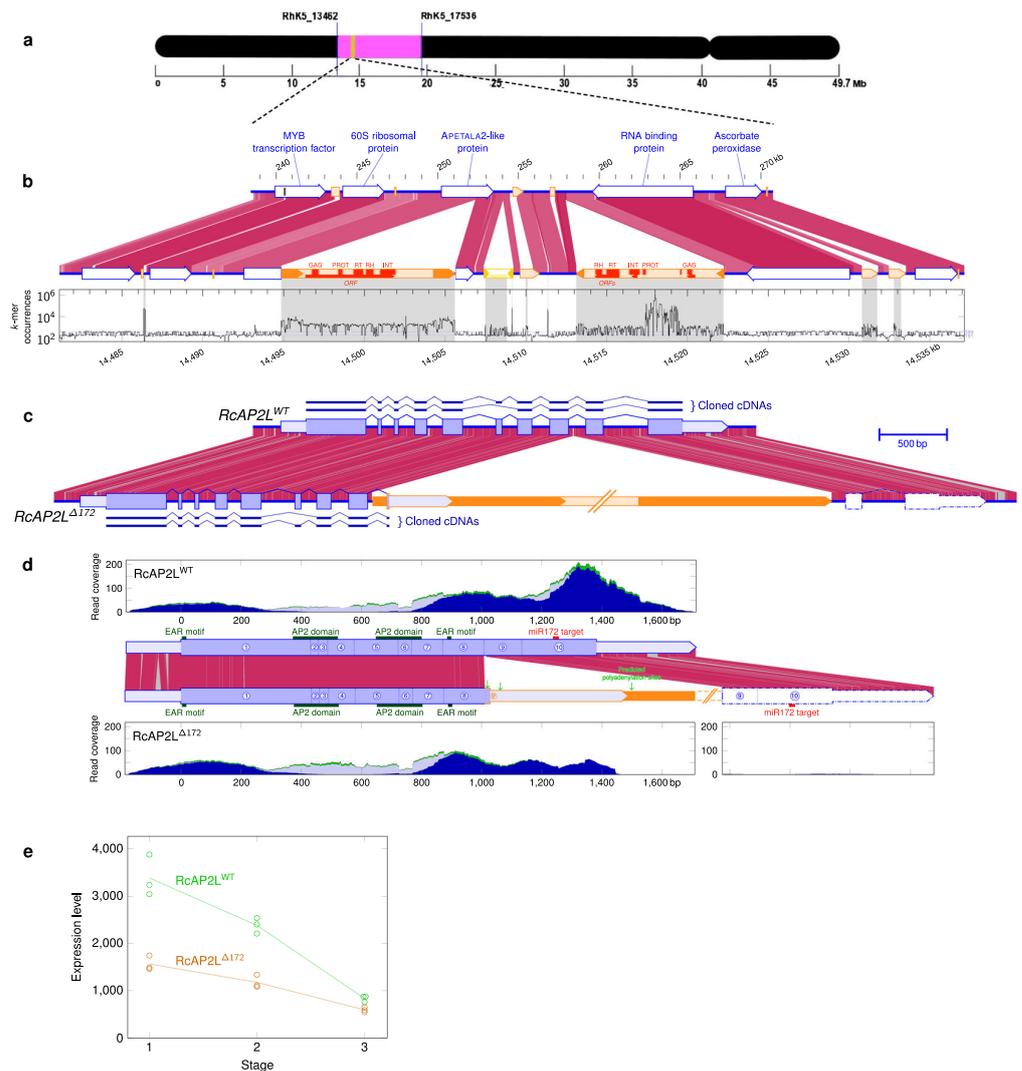


Figure 1. *RcAP2L* lies within the double flower interval in the rose. (a) Graphical representation of chromosome 3 showing the localization of the double flower interval and its flanking markers from the genetic map. (b) Detailed comparison of the two haplotypes containing *RcAP2L* and neighbouring sequences. The upper haplotype contains the wild-type allele of *RcAP2L*. The lower haplotype contains *RcAP2L*^{Δ172} with TE insertion. Blue and orange box-arrows indicate annotated genes and transposable elements, respectively. Sequence similarities between the two haplotypes are shown with purple shades. The darker is the purple colour intensity, the highest is the Blastn identity (ranging between 100 to 90%). The plot shows *k*-mer occurrences along the region: corresponds to the number of times each word of 47 bp in length is found in the 375 Gb set of reads used to assemble the heterozygous genome²⁷. High occurrence counts indicate repeated sequences (grey shading). (c) Intron/exon structure of the two alleles of *RcAP2L*, confirmed by cDNA sequencing. Exons are shown as blue boxes. Light blue indicate 5' and 3' UTR. cDNA sequences confirmed by RT-PCR are shown. The transposable element inserted in *RcAP2L*^{Δ172} is shown in orange. Purple lines show identical sequence and grey lines indicate single nucleotide polymorphisms and short INDELS (overall identity: 93,4%). (d) Structure of mature mRNAs from *RcAP2L* alleles. The three area plots show RNA-seq read coverage along mRNAs. Dark blue: reads matching specifically one of the two alleles. Light blue: reads that could come from any of the two *RcAP2L* alleles. Green: reads also matching at other loci in rose genome. Expression data confirm the structure of the two alleles, including the correct splicing of exon 9' of *RcAP2L*^{Δ172} originating from the inserted TE, and the appearance of a new polyadenylation site at position 1,500 bp. The region of *RcAP2L*^{Δ172} homologous to exons 9 and 10 exhibits a negligible expression level (highest value ≤ 2). (e) Expression of wild-type allele *RcAP2L*^{WT} (green) and *RcAP2L*^{Δ172} allele (orange) early on in flower formation. Stage 1: sepal initiation, stage 2: petal initiation, stage 3: stamen initiation. Data were extracted from RNA-seq analyses. Y-axis is labelled in FPKM.

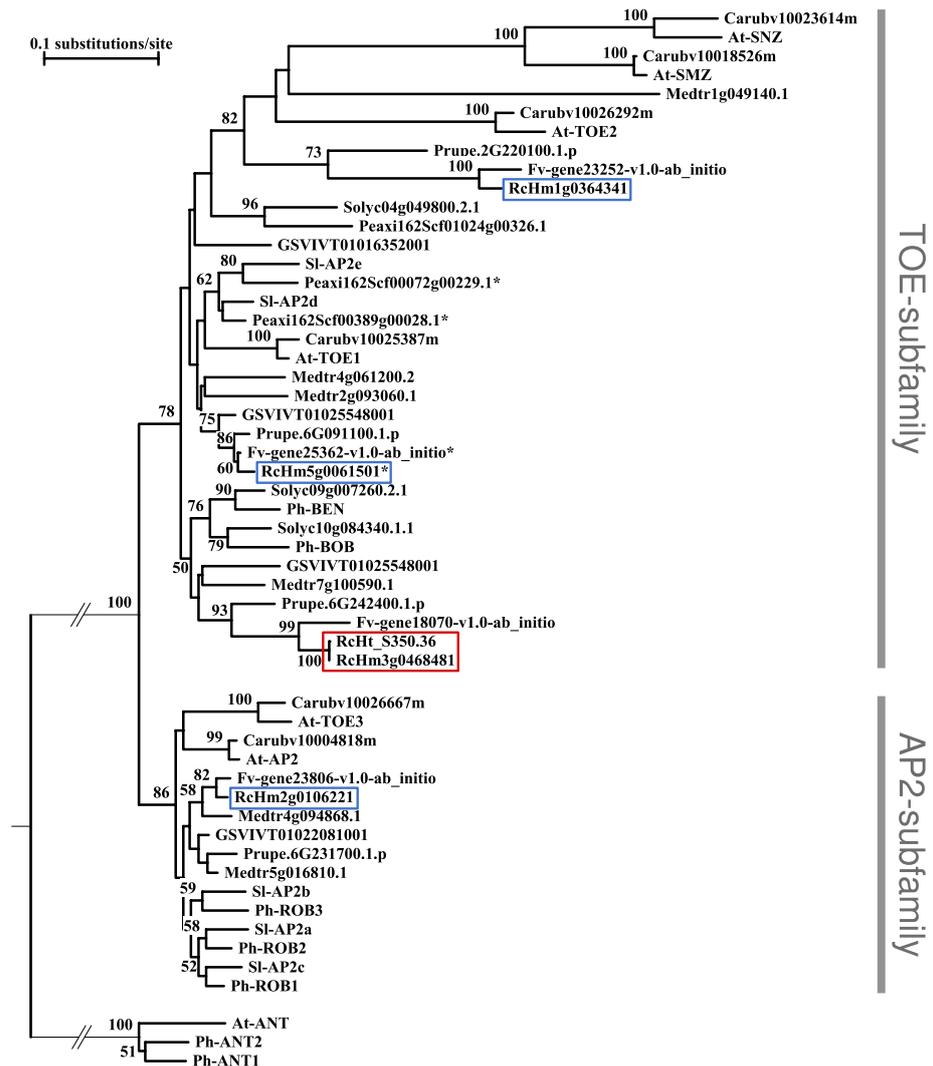


Figure 2. *RcAP2L* belongs to the euAP2 family and the TOE-subfamily. Neighbor-Joining tree based on the aligned AP2 DNA binding domains of the euAP2 members from *Rosa chinensis* (RcHm and RcHt)²⁷, *Fragaria vesca* (Fv), *Petunia hybrida* and *P. axillaris* (Ph and Peaxil, respectively), *Solanum lycopersicum* (Solyc or Sl), *Arabidopsis thaliana* (At), *Capsella rubella* (Carub), *Medicago truncatula* (Medtr), *Vitis vinifera* (GSVIVT) and *Prunus persica* (Prupe). The tree was rooted with *AINTEGUMENTA* (ANT) and tested with 2,000 bootstraps. EuAP2 members clearly group together with a bootstrap of 100. Among this family, the AP2- and TOE-subfamilies are also well separated (bootstrap of 86 and 78 respectively). Each distinct gene of close species generally groups together confirming the reliability of the tree. *RcAP2L* is part of the TOE-subfamily (observed for both the truncated *RcHm3g0468481* and the wild-type *RcHt_S350.36* version).

subfamily contains a single homolog for *AtTOE1* (*RcHm5g0061501*) and a gene (*RcHm1g0364341*) that groups in a branch with *Arabidopsis* *TOE2*, *SMZ* (*SCHLAFMUTZE*) and *SNZ* (*SCHNARCHZAPPEN*). The higher divergence of this last branch from the rest of the tree is likely due to the presence of a non-functional second AP2 DNA-binding domain, that could have accumulated more mutations and putatively acquired a new function^{12,39}.

Phylogenetic analyses, using AP2 domains, revealed no direct orthologue of *RcAP2L* (*RcHm3g0468481*) in *Arabidopsis* genome. Interestingly, *RcAP2L* appears to group with the *Petunia* *PhBEN* and *PhBOB* gene. *PhBEN* was reported to repress the expression of the C-function genes in the perianth, and together with *PhBOB*, it is required for organ growth in the second whorl¹².

Gene sequence analyses, using the assembled heterozygous genome of ‘Old Blush’, revealed that in the double flower of *R. chinensis* ‘Old Blush’, *RcAP2L* is present as two different alleles. The first allele, located on scaffold *RcHt_S350*²⁷, corresponds to the wild-type sequence of *RcAP2L* (*RcAP2L*^{WT}). A second allele, located on two assembled scaffolds (*RcHt_S3277* and *RcHt_S1251*), contains an additional sequence of 10,790 bp inserted in its 8th intron (genome coordinates 14,494,849 – 14,505,638; Fig. 1b).

Genotype	Flower multiplicity	Ploidy level	Overlapping reads per 10 ⁸ reads		
			WT	5' of TE	3' of TE
<i>R. chinensis</i> 'Old Blush'	Double flower	2x	5.7	7.2	7.4
<i>R. chinensis</i> homozygous genome	NA	2x	0	22.2	16.7
<i>R. odorata</i> 'Hume's Blush'	Double flower	2x	8.4	9.2	6.9
<i>R. x hybrida</i> 'La France'	Double flower	3x	5.1	5.1	8.0
<i>R. chinensis</i> 'Sanguinea'	Simple flower	2x	22.5	0	0
<i>R. chinensis</i> 'Spontanea'	Simple flower	2x	13.2	0	0
<i>R. wichurana</i>	Simple flower	2x	8.6	0	0

Table 1. Identification of *RcAP2L* alleles present in 7 resequenced genotypes. Number of 100 bp genomic reads overlapping intra-gene (denoting the presence of a wild-type *RcAP2L* allele) or gene-TE (mutated allele, *RcAP2L*^{Δ172}) junctions, for genotypes with double or simple flowers. The read counts were normalized according to the read library size, and expressed as reads per 100 million reads. Ploidy level: According to Raymond *et al.*²⁷.

The inserted sequence is repeated in the rose genome and corresponds to a transposable element (TE) belonging to the Gypsy LTR retrotransposon family (Fig. 1b). Sequence alignment also showed that the two LTRs of the inserted TE are 100% identical on their whole length, indicating a recent insertion⁴⁰. The inserted TE contains an open reading frame of 5,535 pb and DANTE software predicted the presence of at least 5 retroviral sequences coding for the structural protein GAG, a protease, a reverse transcriptase, a H-Ribonuclease and an integrase. We found 17 complete copies from this TE family in the rose genome, and 66 solo-LTRs, making it moderately repeated.

The TE insertion in *RcAP2L* creates a new splicing acceptor site that is predicted to lead to a fusion of the 8th exon of *RcAP2L* with a sequence from the 5' LTR from the TE. This new splicing creates a premature STOP codon and the loss of the 9th and the 10th exons, which causes the formation of a truncated protein composed of 342 amino acids, and the loss of the miR172 binding site (Fig. 1d). This allele was consequently named *RcAP2L*^{Δ172}.

We mapped RNA-seq reads on the predicted transcripts to validate the mRNA structures; as a few SNPs and INDELS exist between the sequences of the two alleles (Fig. 1d), we were able to distinguish reads coming from each. The RNA-seq coverage drastically decreased at one of the predicted polyadenylation sites of *RcAP2L*^{Δ172} identified by PASPA software⁴¹, indicating that the corresponding mRNA existed and was properly spliced, and thus must be stable. The RNA-seq mapping also showed that exons 9 and 10 of the mutated allele, located after the TE insertion, are not expressed, indicating that the mRNA from this allele no longer have a miR172 binding site. Sequencing of cDNA prepared from RNA extracted from *R. chinensis* 'Old Blush' confirmed the predicted intron/exon structures but also indicated a potential alternative splicing with the loss of the 6th exon (Fig. 1c).

Expression analyses showed that both alleles *RcAP2L*^{WT} and *RcAP2L*^{Δ172} are expressed during flower formation (Fig. 1e). The expression of both alleles is high in flower primordia at stages 1 and 2 (sepal and petal initiation, respectively), and their expression starts to significantly decrease at stage 3 (stamen initiation), thus consistent with a role in perianth formation.

The presence of the *RcAP2L*^{Δ172} allele correlates with double flower formation in Chinese and modern roses. To further address the correlation between the presence of the *RcAP2L*^{Δ172} allele and double flower formation, we investigated its presence in the available genomic data from five other rose genotypes that exhibit either double flowers (*R. odorata* 'Hume's Blush', *R. x hybrida* 'La France') or simple flowers (*R. chinensis* 'Sanguinea', *R. chinensis* 'Spontanea' and *R. wichurana*)²⁷. The insertion of the TE in intron 8 of *RcAP2L* was investigated by the presence of reads overlapping both 5' and 3' junctions, while its absence was confirmed by reads overlapping the intact position on the wild type *RcAP2L* gene (Table 1). For example, a mean of 6.9 reads per 10⁸ reads and 9.2 reads per 10⁸ reads were shown to overlap respectively the 3' and 5' TE junctions in *R. odorata* 'Hume's Blush' and 8.4 reads per 10⁸ reads were overlapping the wild type position of the gene (Table 1), indicating that this genotype had one of each allele. Conversely, *R. wichurana* had no read overlapping the TE junctions and 8.6 reads per 10⁸ reads overlapping the intact position, indicating that this genotype only has the wild type allele at homozygous state. This analysis indicated that all double flower roses of the panel harbour both the wild-type *RcAP2L*^{WT} allele and the truncated *RcAP2L*^{Δ172} allele. Conversely, simple flower roses harbour only the *RcAP2L*^{WT} allele and never the *RcAP2L*^{Δ172}. Together, these data corroborate the observation in 'Old Blush' and show the existence of a correlation between the presence of *RcAP2L*^{Δ172} and double flower formation.

To further confirm our hypothesis, we investigated the presence of *RcAP2L*^{Δ172} in a set of modern rose cultivars (Supplementary Table S2). The presence or absence of *RcAP2L*^{Δ172} was investigated by PCR amplification of the TE insertion junctions using DNA extracted from 6 rose plants exhibiting simple flowers and from 13 rose plants exhibiting double flowers (Supplementary Table 2; Fig. 3). DNA fragments overlapping both left and right borders of the transposon were detected in all these double flower roses (Fig. 3a), while no similar DNA fragment could be detected in the simple flower roses (Fig. 3b). Our data show a correlation between the double flower phenotype and the presence of the transposable element insertion, thus providing another argument in favour of the role of *RcAP2L*^{Δ172} during double flower formation. It should be noted that none of the analysed double flower cultivars were homozygous for the truncated allele and all had wild type and truncated alleles.

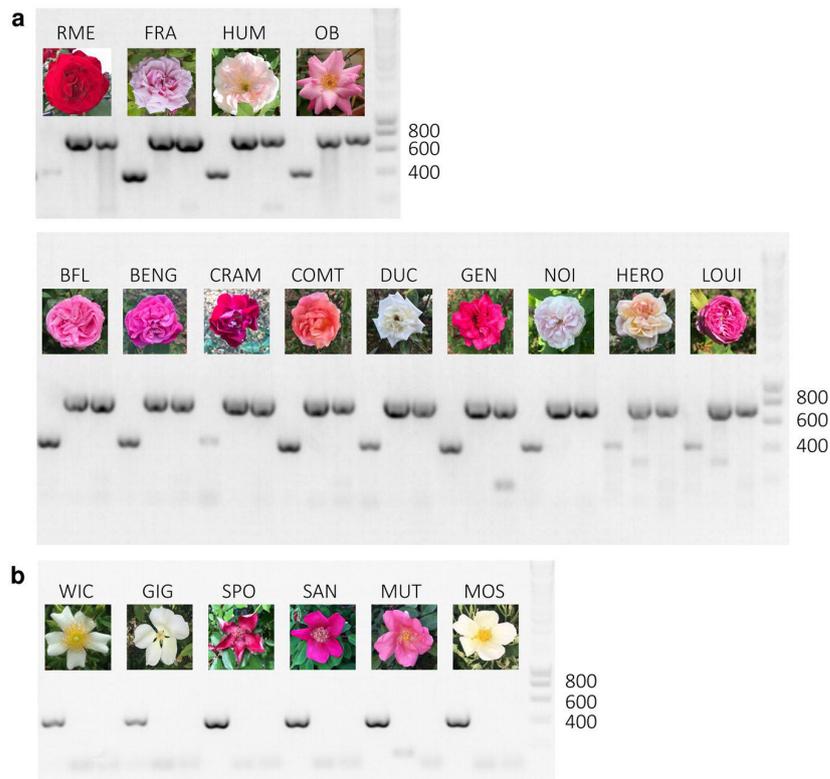


Figure 3. The transposable element insertion is only observed in the analysed double flower roses. PCR to detect the TE were performed on genomic DNA from different rose cultivars exhibiting simple (**b**) or double flowers deriving from *R. chinensis* (**a**). The lower band (419 bp) corresponds to the combination of primers that amplifies the wild type allele, while the two higher ones (770 bp and 754 bp) correspond to amplification of the left and right borders of the transposable element insertion, respectively.

Discussion

In roses, the formation of double flowers is associated with a shift of the A/C boundary and a restriction of *RcAG* expression domain toward the centre of the meristem which in turn leads to a conversion of stamens into petals³. Genetic mapping identified the major locus *Df* as involved in the control of rose double flower formation²⁵. However, *RcAG* does not lie within the double flower interval suggesting that a yet unknown upstream regulator of *RcAG* must be the determinant of double flower formation.

In this study, we identified a gene of the euAP2 family, *RcAP2L*, that localizes within the double flower interval. We identified a truncated allele version of *RcAP2L* (*RcAP2L*^{Δ172}) whose presence correlates with double flower formation in Chinese roses and modern roses, such as *R. hybrida* ‘La France’, that have Chinese rose cultivars as ancestors. Such an allelic form is absent in all simple flower roses. A TE insertion in the 8th intron of *RcAP2L* leads to the loss of the miR172 target site. We demonstrate that the position of the TE insertion is conserved among different double flower rose varieties that have a parent from the Chinese section. These data indicate that this allele must have been inherited from a single common Chinese ancestor and spread among its double flower modern descendants due to the positive human selection during rose domestication.

Phylogenetic analysis indicates that *RcAP2L* is a member of the TOE-subfamily. In *Arabidopsis*, most studies focused on *AP2* and only a few reports addressed the role of other euAP2 family members such as *TOE1*, *TOE2* and *TOE3*. *Arabidopsis AP2* is known to restrict *AG* expression to the third and fourth whorls (stamens and carpels, respectively). Knockout of *AP2* results in ectopic expression of *AG* in the sepal and petal whorls, which is associated with a conversion of sepals to carpeloid structures and loss of petals⁶. Recently, ChIP-qPCR experiments in *Arabidopsis* showed that both *AP2* and *TOE3* bind to *AG* second intron (containing transcription cis-regulatory elements) to decrease its expression level^{29,42}. These published data support our findings on *RcAP2L* as a pertinent candidate for double flower determination, likely by regulating the expression pattern of *RcAG*.

In *Arabidopsis*, *TOE1* overexpression induces late flowering while its loss-of-function leads to early flowering with no apparent flower phenotype variations⁴³. Conversely to *Arabidopsis*, in some species certain TOE-subfamily genes have a flower patterning function. In *Petunia*, *PhBEN* was shown to inhibit the C-function in the perianth primordium, thus consistent with a function similar to that of the *Arabidopsis AP2*. It is clear that *TOE* genes have evolved to perform different functions in different species, and in roses, their misexpression is likely at the origin of the appearance of the double flower abnormality.

EuAP2 family members are characterized by the presence of a miR172 binding site that is important for their post-transcriptional regulation. It has been reported that overexpression of miR172 induces early flowering as well as floral defects similar to the ones observed in *ap2* mutants⁴³. During flower formation, miR172 is highly expressed in whorls 3 and 4 and targets AP2 transcripts to prevent its expression in the centre of the meristem, where the stamen- and carpel-identity gene *AG* is expressed³⁰. However, in *Arabidopsis* it has been reported that when AP2 lacks the miR172 binding site (as for the *RcAP2L^{Δ172}*), its expression is maintained in the centre of the meristem, which leads to continuous downregulation of *AG* expression^{30,44}. Such downregulation of *AG* expression results into the formation of flowers with an increased number of petals or stamens and a loss of floral determinacy^{7,44}, a phenotype resembling that of *ag* loss-of-function mutant⁴⁵. Similarly, *AG* expression is reduced when a miR172-resistant *TOE3* is expressed in *Arabidopsis* flowers²⁹, indicating that other members of the euAP2 family can also have antagonistic role on the expression of *AG*.

In a recent study, Han *et al.*⁴⁶, reported that the down-regulation of the rose AP2 orthologue leads to the reduction of petal number⁴⁶. However, the *RcAP2* gene (corresponding to *RcHm2g0106221²⁷*) studied by Han *et al.*⁴⁶, is located on chromosome 2 and not on chromosome 3, where the interval containing the double flower mutation lies.

In double flower roses, the *RcAP2L^{Δ172}* truncated allele lost its miR172 binding site but still contained both AP2 DNA binding domains and EAR domains (Ethylene-responsive element binding factor-associated amphiphilic repression), thus consistent with an *RcAP2L* gain of function hypothesis and the dominant character of the *Df* gene. Indeed, *Arabidopsis* AP2 is known to interact with TOPLESS via its EAR domain to recruit the histone deacetylase HDA19 to its DNA binding sites including the *AG* second intron⁴⁷. It has been demonstrated that a fusion between the AP2 DNA binding domains and TOPLESS, with the addition of an artificial miR172 binding site, is sufficient to complement the *Arabidopsis ap2* phenotype, indicating that the main floral function of AP2 is established via its TOPLESS interaction and recruitment to DNA target sites. Further investigation also revealed that TOE1, TOE2 and TOE3 interact with TOPLESS⁴⁸, indicating a potentially conserved mechanism among the whole euAP2 family. Therefore, it is likely that in roses, the expression of the miR172 resistant allele (*RcAP2L^{Δ172}*) is responsible for the observed restricted expression of *RcAG* toward the centre of the flower, which in turn leads to the formation of flowers with an increased number of petals.

Our data, taken together with published data in *Arabidopsis* and other plants, suggest the following model. In simple flower roses, *RcAP2L* and *RcAP2* proteins are produced only in the first two whorls, where they can inhibit *RcAG* transcription. This induces sepal and petal organ identity determination and development in whorls 1 and 2, respectively. In the third and fourth whorls, the accumulation of miR172 interferes with euAP2 mRNA accumulation, which in turn results in the expression of *RcAG* that mediates stamen and carpel organ identity determination and development (Fig. 4). However, in a double flower such as ‘Old Blush’, the presence of the miR172 insensitive variant *RcAP2L^{Δ172}* leads to a prolonged accumulation of *RcAP2L* protein toward the centre of the meristem and consequently prolonged downregulation of *RcAG* expression (Fig. 4). This causes a restriction of *RcAG* expression toward the centre of the flower, as described previously³. As a consequence, the homeotic conversion of stamens into petals leads to the formation of a double flower.

Our model is supported by previous work on kiwifruit, where a downregulation of miR172 and an up-regulation of AP2 were observed in flower buds from the ‘Pukekohe dwarf’ kiwifruit double flower cultivars, but not in the ‘Hayward’ and ‘Chieftain’ simple flower cultivars⁴⁹.

Our data taken together with that in the literature, strengthen the conclusion that misregulation of the miR172/AP2 loop is likely the cause of the double flower phenotype in many species. To address more in depth the molecular mechanisms that link the miR172 insensitive allele of *RcAP2L* to the double flower formation, a future experiment would be to overexpress it in simple flower roses.

The fact that many double flower roses still develop carpels suggests that the accumulation of the miR172 insensitive variant *RcAP2L^{Δ172}* affects *RcAG* expression only in the third whorl, but not in the fourth whorl. This indicates that the rose may have evolved differently from *Arabidopsis* where miR172 insensitive variant of AP2 affect both whorls 3 and 4. It will be interesting to address the molecular mechanisms of such difference and whether such mechanism is applicable to other species with double flowers.

Material and Methods

Plant material. Double flower rose cultivars *Rosa chinensis* ‘Old Blush’, *R. odorata* ‘Hume’s Blush’, *R. x hybrida* ‘La France’, *R. x hybrida* ‘Rouge Meilland’, *R. x hybrida* ‘Bébé Fleuri’, *R. x hybrida* ‘Bengale d’Automne’, *R. x hybrida* ‘Cramoisi Supérieur’, *R. x hybrida* ‘Comtesse de Cayla’, *R. x hybrida* ‘Ducher’, *R. x hybrida* ‘General Shablikine’, *R. x hybrida* ‘Blush Noisette’, *R. x hybrida* ‘Herodiade’ and *R. x hybrida* ‘Louise d’Arzens’, and simple flower cultivars *R. chinensis* ‘Spontanea’, *R. chinensis* ‘Sanguinea’, *R. chinensis* ‘Mutabilis’, *R. wichurana*, *R. gigantea* and *R. moschata* were field grown at the Lyon-Botanical-Garden and/or in environmentally controlled greenhouse conditions at the Ecole Normale Supérieure of Lyon with 16 h/8 h day/night periods and 25 °C/18 °C day/night temperatures.

Staging rose flower development. Flower development stages were distinguished and dissected under a binocular microscope as previously defined by Dubois *et al.*¹⁴. Stages 1 to 3 correspond to development stages when sepal, petal and stamen primordia arise, respectively. During stage 4, carpels are produced in the centre of the meristem, which will then sink below at stage 5.

DNA extraction and genotyping PCR. Young leaves or axillary buds were collected and ground in PVP and homogenization buffer (Tris pH8 15 mM, EDTA 2 mM, NaCl 20 mM, KCl 20 mM, β-mercaptoethanol 0.1%, Triton 0.5%). DNA extraction was performed using the DNeasy kit (Qiagen).

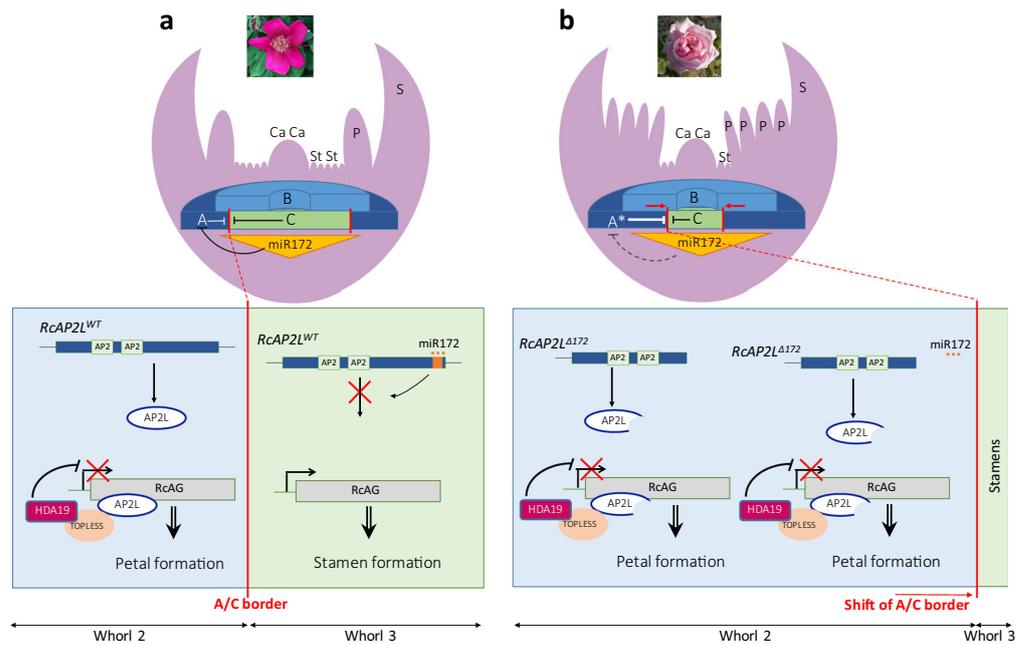


Figure 4. Model showing how a miR172-resistant euAP2 could lead to double flower formation. **(a)** In wild-type flowers, euAP2 are expressed in the first and second whorls where they can inhibit *RcAG* expression likely by recruiting cofactors and histone modifiers, such as *TOPLESS* and *HDA19*. Sepals and petals are consequently formed. In the 3rd whorl, miR172 is expressed and inhibits euAP2 proteins production, releasing the inhibition of *RcAG*. *RcAG* will then determine stamens and carpels identity and formation. **(b)** In double flowers, the truncated version *RcAP2L* Δ 172 mRNA (following TE insertion) is insensitive to miR172 inhibition. *RcAP2L* Δ 172 expression is maintained in the meristem and down-regulates *RcAG* expression. This leads to the sliding of the A/C border toward the centre of the meristem and the formation of more petals, producing the so-called double-flowers. The wild-type *RcAP2L* (not shown on the Figure) is degraded. Ca, carpels; St, stamens; P, petals; S, sepals. A, A-class genes (*RcAP1*, euAP2); A*, A-class genes including a miR172-resistant euAP2 such as *RcAP2L* Δ 172; B, B-class genes; C, C-class gene (*RcAG*).

DNA fragments were amplified using the GoTaq Polymerase Chain Reaction according to the manufacturer's recommendation (Promega). An initial denaturing step was carried at 95 °C for 5 min. Fifteen cycles of touch-down PCR were then performed 95 °C for 30 s, 65 °C (with a decrease of 1 °C per cycle) for 30 s, 72 °C for 1 min 30 s. This was followed by 30 cycles of standard PCR with the following cycle 95 °C for 30 s, 50 °C for 30 s and 72 °C for 1 min 30 s. A final elongation step was performed for 10 min at 72 °C.

RNA purification and cDNA sequencing. Total RNA was prepared from floral meristems at different developmental stages (1 to 3) using the Spectrum plant total RNA kit (Sigma) and TURBO DNA-freeTM AM 1907 (Ambion), mainly as previously described³. Contaminating DNA was removed using the DNA-freeTM kit following the manufacturer's recommendations (Ambion). One microgram of total RNA was then used in a reverse transcription assay. cDNAs were PCR amplified, cloned and sequenced using primers designed to specifically target *RcAP2L* or *RcAP2L* Δ 172 (Supplementary Table 3).

Characterisation of the double flower interval. The high quality rose genome from *Rosa chinensis* 'Old Blush' was recently published²⁷ in the form of two complementary assemblies. The first one was obtained from PacBio long read sequencing using a homozygous rose material derived from the heterozygous *Rosa chinensis* 'Old Blush'^{27,28} and consists of 7 assembled pseudomolecules representing a haplotype of the rose genome. The genome of the heterozygous *Rosa chinensis* 'Old Blush' (Illumina sequencing) consists of 15,937 scaffolds and provides access to the two haplotypes of the genome.

Flanking markers of double flower interval³¹ were mapped on the *Rosa chinensis* homozygous reference genome²⁷ using the following parameters: $evalue < 10^{-6}$, $lengthHSP > 40$, $percentage\ identity > 97\%$. Markers that had unique match were kept and used to define the corresponding physical region on the rose genome sequence. Genes within this interval were analyzed using Blast and Pfam web interface^{50,51}.

Analysis of the presence of the TE element in *RcAP2L* was also performed using the available genome sequences of rose cultivars²⁷. Single reads from resequenced genomes of the different rose cultivars²⁷ were trimmed using cutadapt⁵² and custom Perl scripts. They were cut to an homogeneous length of 100 bp and aligned on the reference rose genome using bwa software⁵³ allowing up to two mismatches on the whole length of the read

(end-to-end alignment). Reads overlapping genomic positions of interest over at least 15 bp on each side were counted. Read counts were normalized on the library size for each genotype. psRNATarget webserver interface was used to detect miR targets⁵⁴.

Haplotype identification and comparison. Sequence analysis was performed using the high-quality genome assembly of homozygous *R. chinensis* ‘Old Blush’²⁷. The two distinct haplotypes within the double flower interval were retrieved from the heterozygous genome assembly²⁷. Blastn⁵⁵ and gene synteny were used to confirm alleles sequences. The water program from EMBOSS suite⁵⁶ was used to obtain optimal end-to-end alignments between allelic regions and identify polymorphisms.

Gene sequence analysis. Gene models were recovered from the rose reference genome sequence annotation²⁷. Splicing site predictions and untranslated region (UTR) boundaries were manually adjusted based on cloned cDNA sequences and RNA-seq data. Putative functions for genes flanking *RcAP2L* were inferred from *Arabidopsis* best blast hit. Unknown protein domains were identified using InterProScan software version 5.27.-66.0⁵⁷, Pfam database version 31.0⁵⁰ and manual annotation. miR172 putative binding sites were predicted using a local instance of WMD3 software (Ossowski Stephan, Fitz Joffrey, Schwab Rebecca, Riester Markus and Weigel Detlef, personal communication).

Transposable element annotation. To identify repeated regions, the genomic sequence of *RcAP2L* neighbourhood was cut into 47 bp overlapping *k*-mers, and the number of occurrences of each *k*-mer was counted in the 375 Gb-dataset of genomic reads used to assemble the *Rosa chinensis* heterozygous rose genome sequence²⁷. These occurrence counts were plotted along the sequence (Fig. 1b) and compared to the mean occurrence counts for homozygous and heterozygous regions. Automatic transposable element (TE) annotations from the rose genome²⁷ were used as a starting point, and manually curated.

The boundaries of the two long terminal repeats (LTRs) were accurately identified using a graphical dotplot program⁵⁸ and LTR sequences were compared using bl2seq alignment⁵⁵. Open reading frames were predicted in the TE internal region using Pfam software⁵⁰ and protein domains were annotated by similarity search using DANTE (<http://repeatexplorer.org/>).

Using the LTR sequence as a Blastn query (parameters: M = 6 N = -7 Q = 8 R = 8; e-value $\leq 10^{-20}$, match length ≥ 600 bp) and the internal part sequence as tBlastx query (parameters: BLOSUM80 Q = 9 R = 3; e-value $\leq 10^{-15}$, overall coverage of query ≥ 800 bp), we looked for LTR pairs flanking a putative internal part, to detect complete copies of TEs from the same family. Using more stringent criteria (e-value $\leq 10^{-80}$ and match length ≥ 900 bp), we also identified solo-LTRs from the same family⁵⁹.

Expression analysis. Paired-end RNA-seq data from young flower buds at stage 1, 2 and 3 were previously described²⁷.

Pairs of reads putatively originating from *RcAP2L*^{WT} and *RcAP2L* ^{Δ 172} alleles were selected using Tophat version 2.1.1⁶⁰ with relaxed parameters (“-read-realign-edit-dist 0-b2-very-sensitive -max-intron-length 25000” and insert size and insert size SD estimated beforehand on the whole predicted transcriptome for each library). These read pairs were remapped on the whole genome with Tophat allowing up to 5 multimatches and secondary alignments. Based on the number of matches and the alignment scores, read pairs were sorted into four categories: (i) specific to *RcAP2L*^{WT}, (ii) specific to *RcAP2L* ^{Δ 172}, (iii) coming indiscriminately from *RcAP2L*^{WT} or *RcAP2L* ^{Δ 172}, and (iv) coming indiscriminately from *RcAP2L*^{WT}, *RcAP2L* ^{Δ 172} or other loci in the genome (hereafter called non-specific read pairs). Read pairs that Tophat could match on the extracted genomic sequences of *RcAP2L* alleles, but not on the whole genome, were put in category (iv). This case is expected for reads originating from repeated sequences. The read pairs from each category were mapped on the predicted transcripts using Bowtie2 version 2.3.4.1⁶¹. Coverage at each position was computed using Samtools version 1.5⁶². Normalization was done using the library sizes (custom Perl scripts), before adding up the coverage values from all libraries. Reads from categories (iii) and (iv) were spread between the two alleles according to the ratio (i)/(ii), computed on a sliding window of width 241 bp. After ensuring that sequence polymorphism between *RcAP2L*^{WT} and *RcAP2L* ^{Δ 172} transcripts was sufficient to estimate independently their expression level, we used Tophat version 2.1.1⁶⁰ on the annotated *Rosa chinensis* heterozygous genome²⁷, with corrected annotations for *RcAP2L* alleles, and we normalized read counts using DESeq. 2 version 1.2.0⁶³.

Phylogenetic analysis. EuAP2 family members were identified by using the *Arabidopsis thaliana* AP2 protein as a Blast query against rose and strawberry predicted proteins²⁷. Sequences were aligned using ClustalW⁶⁴ and BioEdit software⁶⁵. Where applicable, gene annotation was corrected manually.

Neighbor-Joining tree based on the aligned AP2 DNA binding domains of the euAP2 members from *Rosa chinensis* (RcHm and RcHt)²⁷, *Fragaria vesca* (Fv), *Petunia*¹², *Solanum lycopersicum*, *Arabidopsis thaliana*, *Capsella rubella*, *Medicago truncatula*, *Vitis vinifera* and *Prunus persica*. Sequences were downloaded from the Phytozome website (<https://phytozome.jgi.doe.gov/pz/portal.html>). The aligned regions containing the two AP2 domains (Supplemental Data File 1) were selected for phylogenetic analysis. Neighbor-Joining tree was computed with Treecon software⁶⁶ using the following parameters (1) Distance estimation options: Tajima and Nei⁶⁷; Distance calculations; insertions and deletion not taken into account; Alignment positions: all; Bootstrap analysis: yes, 2000 samples. (2) Infer tree topology options: Neighbor-joining; Bootstrap analysis: yes. (3) Root unrooted trees options: single sequence (forced); bootstrap analysis: yes. Tree was rooted using the *Arabidopsis* ANT protein.

References

- Cairns, T. *Modern roses XI, The word Encyclopedia of roses*, (Academic Press, San Diego, California, 2003).
- Bendahmane, M., Dubois, A., Raymond, O. & Bris, M. L. Genetics and genomics of flower initiation and development in roses. *J Exp Bot* **64**, 847–857 (2013).
- Dubois, A. *et al.* Tinkering with the C-function: a molecular frame for the selection of double flowers in cultivated roses. *PLoS One* **5**, e9288 (2010).
- Irish, V. F. & Litt, A. Flower development and evolution: gene duplication, diversification and redeployment. *Current Opinion in Genetics & Development* **15**, 454–460 (2005).
- Krizek, B. A. & Fletcher, J. C. Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet* **6**, 688–98 (2005).
- Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genetic interactions among floral homeotic genes of Arabidopsis. *Development* **112**, 1–20 (1991).
- Chen, X. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* **303**, 2022–5 (2004).
- Irish, V. The ABC model of floral development. *Curr Biol* **27**, R887–R890 (2017).
- Theissen, G. & Melzer, R. Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann Bot* **100**, 603–19 (2007).
- Bradley, D., Carpenter, R., Sommer, H., Hartley, N. & Coen, E. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. *Cell* **72**, 85–95 (1993).
- Cartolano, M. *et al.* A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. *Nat Genet* **39**, 901–5 (2007).
- Morel, P. *et al.* Divergence of the Floral A-Function between an Asterid and a Rosid Species. *Plant Cell* **29**, 1605–1621 (2017).
- Dubois, A. *et al.* Transcriptome database resource and gene expression atlas for the rose. *BMC Genomics* **13**, 638–648 (2012).
- Dubois, A. *et al.* Genomic approach to study floral development genes in *Rosa* sp. *PLoS One* **6**, e28455 (2011).
- Hibino, Y., Kitahara, K., Hirai, S. & Matsumoto, S. Structural and functional analysis of rose class B MADS-box genes 'MASAKO BP, euB3, and B3: Paleo-type AP3 homologue 'MASAKO B3' association with petal development. *Plant Science* **170**, 778–785 (2006).
- Kitahara, K., Hibino, Y., Aida, R. & Matsumoto, S. Ectopic expression of the rose AGAMOUS-like MADS-box genes 'MASAKO C1 and D1' causes similar homeotic transformation of sepal and petal in Arabidopsis and sepal in *Torenia*. *Plant Science* **166**, 1245–1252 (2004).
- Kitahara, K., Hirai, S., Fukui, H. & Matsumoto, S. Rose MADS-box genes 'MASAKO BP and B3' homologous to class B floral identity genes. *Plant Science* **161**, 549–557 (2001).
- Kitahara, K. & Matsumoto, S. Rose MADS-box genes 'MASAKO C1 and D1' homologous to class C floral identity genes. *Plant Science* **151**, 121–134 (2000).
- Mibus, H., Heckl, D. & Serek, M. Cloning and Characterization of Three APETALA1/FRUITFULL-like Genes in Different Flower Types of *Rosa × hybrida* L. *J Plant Growth Regul* **30**, 272 (2011).
- Ma, N. *et al.* Low temperature-induced DNA hypermethylation attenuates expression of RhAG, an AGAMOUS homolog, and increases petal number in rose (*Rosa hybrida*). *BMC Plant Biol* **15**, 237 (2015).
- Galimba, K. D. *et al.* Loss of deeply conserved C-class floral homeotic gene function and C- and E-class protein interaction in a double-flowered ranunculid mutant. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2267–E2275 (2012).
- Liu, Z., Zhang, D., Liu, D., Li, F. & Lu, H. Exon skipping of AGAMOUS homolog PrseAG in developing double flowers of *Prunus lannesiana* (Rosaceae). *Plant Cell Reports* **32**, 227–237 (2012).
- Nakatsuka, T. *et al.* Isolation and characterization of the C-class MADS-box gene involved in the formation of double flowers in Japanese gentian. *BMC plant biology* **15**, 182 (2015).
- Tanaka, Y. *et al.* Multi-petal cyclamen flowers produced by AGAMOUS chimeric repressor expression. *Scientific Reports* **3**, 2641 (2013).
- Debener, T. & Mattiesch, L. Construction of a genetic linkage map for roses using RAPD and AFLP markers. *Theoretical and Applied Genetics* **99**, 891–899 (1999).
- Spiller, M. *et al.* Towards a unified genetic map for diploid roses. *Theor Appl Genet* **122**, 489–500 (2011).
- Raymond, O. *et al.* The *Rosa* genome provides new insights into the domestication of modern roses. *Nature Genetics* **50**, 772–777 (2018).
- Vergne, P. *et al.* Production of homozygous rose line derived from heterozygous genotype. (2018).
- Jung, J.-H., Lee, S., Yun, J., Lee, M. & Park, C.-M. The miR172 target TOE3 represses AGAMOUS expression during Arabidopsis floral patterning. *Plant Science* **215–216**, 29–38 (2014).
- Wollmann, H., Mica, E., Todesco, M., Long, J. A. & Weigel, D. On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. *Development* **137**, 3633–42 (2010).
- Bourke, P. M. *et al.* Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *Plant J* **90**, 330–343 (2017).
- Koning-Boucoiran, C. F. *et al.* Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68 k Axiom SNP array for rose (*Rosa* L.). *Front Plant Sci* **6**, 249 (2015).
- Shupert, D. A., Byrne, D. H. & Brent Pemberton, H. Inheritance of Flower Traits, Leaflet Number and Prickles in Roses. 751 edn 331–335 (International Society for Horticultural Science (ISHS), Leuven, Belgium, 2007).
- Debener, T. Genetics: Inheritance of characteristics. in *Encyclopedia of rose science*. (eds Roberts, A.V., Debener, T. & Gudim, S.) 286–292 (Elsevier, Oxford, 2003).
- Crespel, L. *et al.* Mapping of qualitative and quantitative phenotypic traits in *Rosa* using AFLP markers. *Theor Appl Genet* **105**, 1207–1214 (2002).
- Drews, G. N., Bowman, J. L. & Meyerowitz, E. M. Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell* **65**, 991–1002 (1991).
- Jofuku, K. D., den Boer, B. G., Van Montagu, M. & Okamoto, J. K. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *The Plant Cell* **6**, 1211–1225 (1994).
- Kim, S., Soltis, P. S., Wall, K. & Soltis, D. E. Phylogeny and domain evolution in the APETALA2-like gene family. *Mol Biol Evol* **23**, 107–20 (2006).
- Wang, P. *et al.* Expansion and Functional Divergence of AP2 Group Genes in Spermatophytes Determined by Molecular Evolution and Arabidopsis Mutant Analysis. *Frontiers in Plant Science* **7**(2016).
- SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. The paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**, 43–45 (1998).
- Ji, G. *et al.* PASPA: a web server for mRNA poly(A) site predictions in plants and algae. *Bioinformatics* **31**, 1671–3 (2015).
- Yant, L. *et al.* Orchestration of the Floral Transition and Floral Development in Arabidopsis by the Bifunctional Transcription Factor APETALA2[W][OA]. *The Plant Cell* **22**, 2156–2170 (2010).
- Aukerman, M. J. & Sakai, H. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**, 2730–41 (2003).
- Zhao, L., Kim, Y., Dinh, T. T. & Chen, X. miR172 regulates stem cell fate and defines the inner boundary of APETALA3 and PSTILLATA expression domain in Arabidopsis floral meristems. *Plant J* **51**, 840–9 (2007).
- Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genes directing flower development in Arabidopsis. *Plant Cell* **1**, 37–52 (1989).

46. Han, Y. *et al.* An APETALA2 Homolog, RcAP2, Regulates the Number of Rose Petals Derived From Stamens and Response to Temperature Fluctuations. *Front Plant Sci* **9**, 481 (2018).
47. Krogan, N. T., Hogan, K. & Long, J. A. APETALA2 negatively regulates multiple floral organ identity genes in Arabidopsis by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19. *Development* **139**, 4180–4190 (2012).
48. Causier, B., Ashworth, M., Guo, W. & Davies, B. The TOPLESS Interactome: A Framework for Gene Repression in Arabidopsis1[W] [OA]. *Plant Physiology* **158**, 423–438 (2012).
49. Varkonyi-Gasic, E., Lough, R. H., Moss, S. M. A., Wu, R. & Hellens, R. P. Kiwifruit floral gene APETALA2 is alternatively spliced and accumulates in aberrant indeterminate flowers in the absence of miR172. *Plant Molecular Biology* **78**, 417–429 (2012).
50. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**, D279–D285 (2016).
51. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **45**, D12–D17 (2017).
52. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
53. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–95 (2010).
54. Dai, X. & Zhao, P. X. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* **39**, W155–159 (2011).
55. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–10 (1990).
56. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276–7 (2000).
57. Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic acids research* **45**, D190–D199 (2017).
58. Sonnhammer, E. L. & Durbin, R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**, GC1–10 (1995).
59. Vitte, C. & Panaud, O. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol Biol Evol* **20**, 528–40 (2003).
60. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–9 (2012).
62. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9 (2009).
63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
64. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–80 (1994).
65. Hall, T. A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
66. Van de Peer, Y. & De Wachter, R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer applications in the biosciences: CABIOS* **10**, 569–570 (1994).
67. Tajima, F. & Nei, M. Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* **1**, 269–85 (1984).

Acknowledgements

We thank Judit Szécsi (ENS de Lyon) and Philippe Vergne (ENS de Lyon) for useful discussions and for critical reading of the manuscript. We thank Alexis Lacroix, Patrice Bolland and Justin Berger (ENS de Lyon), and the “Lyon Botanical Garden-France” for providing plant material. We gratefully acknowledge support from the Pôle Scientifique de Modélisation Numérique of the ENS de Lyon for the computing resources. We thank Loïs Taulelle (PSMN) for his help with computing clusters and Emmanuel Quemener (Centre Blaise Pascal, ENS de Lyon) for setting up *ad-hoc* servers when needed. This work was supported by funds from the French National Institute of Agronomic Research (INRA), from the Ecole Normale Supérieure-Lyon-France, and from French National Research Agency program DODO (ANR-16CE20-0024-03).

Author Contributions

L.F., D.R., X.F., J.J., O.R. performed experiments. M.B., L.F., J.J., O.R., A.B. analysed data. M.Ve., A.D., J.J., L.F. performed bioinformatics analyses. M.Va., L.F. performed phylogenetic analyses. M.B. designed and supervised the study. M.B., L.F., J.J., O.R. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30918-4>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

3. Résultats complémentaires et conclusions

Depuis la publication de cet article, de nouvelles corrélations ont été identifiées et permettent d'appuyer d'autant plus le potentiel rôle de *RcAP2L*^{Δ172} dans la détermination du phénotype « fleur double ».

Tout d'abord, un mutant d'« Old Blush », ci-après référencé en tant que 'T49', produisant des fleurs simples a été obtenu par mon équipe (P. Vergne et coll.) après culture *in vitro* et transformation par *A. tumefaciens* (Figure 22-A). Ce phénotype n'ayant été obtenu qu'une fois sur la quinzaine d'événements de transformation obtenus lors de cette expérience, il a été supposé que le phénotype n'était pas lié à l'insert en lui-même. De manière surprenante, il n'a pas été possible d'amplifier par PCR les jonctions entre le TE et *RcAP2L* dans le génome de cette lignée 'T49' (Figure 22-B). Ces jonctions sont caractéristiques de l'allèle *RcAP2L*^{Δ172} qui est normalement présent dans le génome d'« Old Blush » et responsable du phénotype « fleur double » selon notre hypothèse (François et al., 2018). Cette observation suggérait une élimination de l'allèle *RcAP2L*^{Δ172} lors de la culture *in vitro* d'un cal d'« Old Blush », ce qui aurait permis l'apparition du phénotype « fleur simple » chez 'T49'. De plus, le mapping allèle-spécifique de données de RNA-seq comparant les stades précoces du développement de bourgeons floraux chez 'Old Blush' et 'T49', réalisé par Jérémie Just (Ingénieur de Recherche dans l'équipe), confirme que l'allèle *RcAP2L*^{Δ172} n'est pas exprimé chez 'T49' (Figure 22-C), prolongeant la corrélation entre la présence de cet allèle et le phénotype « fleur double ». Une étude à l'échelle du génome de l'expression spécifique des deux haplotypes a permis de montrer qu'une grande région de cet haplotype du chromosome 3 n'était plus exprimée chez cet individu (Figure 22-D), suggérant une potentielle délétion d'au moins 28,8 mégabases, incluant l'allèle *RcAP2L*^{Δ172}, qui serait à l'origine de la réversion du phénotype.

D'autre part, des études chez le pêcher *Prunus persica* (L.) Batsch ont aussi montré qu'une mutation, dominante et localisée sur une portion du chromosome 6, pouvait aussi être à l'origine de l'induction du phénotype « fleur double » chez cette espèce (Pascal et al., 2017). Récemment, il a été démontré que l'orthologue de *RcAP2L* se trouvait dans cet intervalle génétique et qu'une délétion du site d'inhibition par miR172 dans celui-ci corrélait aussi avec le phénotype « fleur double » chez le pêcher (Gattolin et al., 2018).

L'ensemble de ces informations concorde avec l'hypothèse formulée dans notre article, proposant que la résistance au miR172 de l'allèle *RcAP2L*^{Δ172} chez le rosier, mais certainement aussi chez d'autres Rosaceae, soit à l'origine d'une prolongation de l'activité de la protéine *RcAP2L*, qui serait alors capable d'inhiber *RcAG* plus longtemps et d'accroître la production de pétales, aux dépens de celle des étamines.

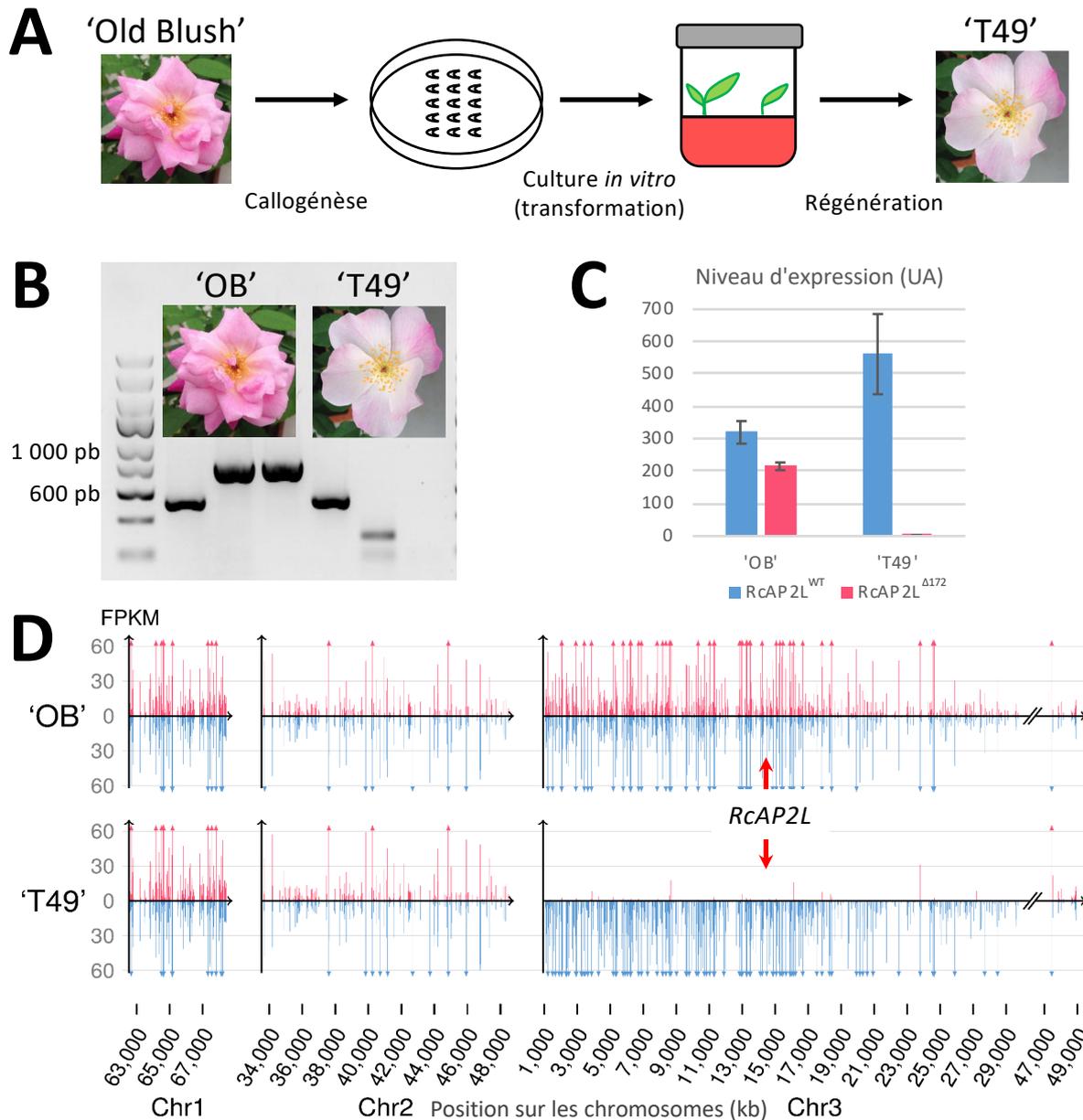


Figure 22 : Un variant d'Old Blush' obtenu après culture *in vitro* a perdu l'allèle $RcAP2L^{\Delta 172}$ et produit de nouveau des fleurs simples.

- (A) Obtention inattendue d'un mutant d'Old Blush' à fleurs simples après culture *in vitro* et transformation. De manière surprenante, aucun insert n'a pu être détecté dans le génome de 'T49'.
- (B) PCR amplifiant la jonction sauvage (416 pb) et les deux jonctions entre $RcAP2L^{\Delta 172}$ et l'élément transposable (770 et 754 pb) chez la fleur double 'Old Blush' et son mutant à fleur simple 'T49'. Ce dernier semble avoir perdu l'allèle $RcAP2L^{\Delta 172}$.
- (C) Niveaux d'expression de $RcAP2L^{\Delta 172}$ et $RcAP2L^{WT}$ déterminés à partir des données transcriptomiques obtenues lors des stades précoces du développement floral d'OB' et 'T49'.
- (D) Expression des deux allèles de chaque gène, en FPKM, sur les portions du génome dont les deux haplotypes sont complètement assemblés (représentés en rose et bleu; 2889 gènes), dans des boutons floraux d'Old Blush' et 'T49'. On observe chez 'T49' une perte d'expression de l'haplotype du chromosome 3 contenant $RcAP2L^{\Delta 172}$ sur une portion d'au moins 28,8 Mb (1349 gènes). Les autres paires de chromosomes, y compris les chromosomes 4 à 7 non représentés sur la figure, ont leurs deux haplotypes exprimés.

III. Caractérisation fonctionnelle de RcAP2L

1. *RcAP2L, un facteur de transcription spécifique des jeunes tissus en division*

L'étude des données transcriptomiques produites dans l'équipe (Dubois et al., 2010) permet d'identifier le profil d'expression global de *RcAP2L* (Figure 23-A). Ainsi, il semble spécifiquement exprimé dans les bourgeons végétatifs et floraux, des tissus méristématiques en cours de division, ainsi que dans les anthères matures. Un deuxième RNA-seq réalisé spécifiquement sur les bourgeons floraux indique que son expression est forte lors de la formation des verticilles de sépales (stade 1) et de pétales (stade 2), avant de commencer à diminuer au stade 3, lors de la mise en place des étamines (Figure 23-B), ce qui est cohérent avec un rôle dans le déterminisme des organes du périanthe. En comparaison, les autres membres de la famille euAP2 ont un profil d'expression plus étendu, puisqu'ils semblent aussi exprimés dans les racines, les feuilles et les fleurs jeunes, matures voire sénescents (Figures 24 et 25). Les orthologues d'*AP2* et de *TOE1* (*RcHm2g0106221* et *RcHm5g0061501*, respectivement) ont des niveaux d'expression plutôt homogènes au sein de l'ensemble des organes de la plante, avec *RcAP2* plus fortement exprimé que *RcTOE1* en moyenne. En revanche, le niveau d'expression de *RcHm1g0364341*, plus proche gène de *TOE2*, *SNZ* et *SMZ* (Figure 2 de l'article François et al., 2018), montre une bien plus grande variabilité puisque celui-ci semble fortement exprimé dans les feuilles et les racines mais pas dans les fleurs matures et les cynorhodons. Ainsi, bien que ces quatre gènes possèdent chacun un motif de régulation par miR172, leurs profils d'expression sont assez différents, indiquant que la régulation de leur expression ne repose pas uniquement sur l'activité du microARN. En particulier, il est probable que le promoteur de *RcAP2L* soit plus spécifique que celui des autres membres de la famille et qu'il ne permette l'expression de *RcAP2L* que dans les méristèmes et jeunes tissus floraux. Au contraire, les promoteurs de *RcAP2* et *RcTOE1* pourraient promouvoir l'expression de ces gènes dans l'ensemble des organes de la plante.

Étonnamment, l'orthologue d'*AP2* reste fortement exprimé au stade 3 (Figure 25), lors de la mise en place des primordia d'étamines, ce qui semble en contradiction avec son rôle supposé dans l'inhibition de *RcAG* et de la formation des étamines (Han et al., 2018). Cette observation est cependant cohérente avec certaines études réalisées chez *A. thaliana*, identifiant que miR172 pourrait essentiellement bloquer la traduction de ses cibles plutôt que d'induire la dégradation des ARNm (Aukerman and Sakai, 2003). Ainsi, bien que les ARNm de *RcAP2* et *RcAP2L* soient détectés dans les méristèmes floraux de stade 3, il est possible que peu ou pas de protéine soit produite à ce stade.

À l'heure actuelle, aucune hybridation *in situ* n'a pour l'instant permis de confirmer le profil d'expression de *RcAP2L*, et ce certainement à cause d'un problème lié aux sondes utilisées puisqu'un témoin positif, dont le niveau d'expression en RNA-seq dans les méristèmes floraux est similaire à celui mesuré pour *RcAP2L*, permet d'obtenir un signal net lors de ces expériences. Il serait donc nécessaire de continuer l'optimisation de ce protocole afin d'obtenir les différents profils d'expression des

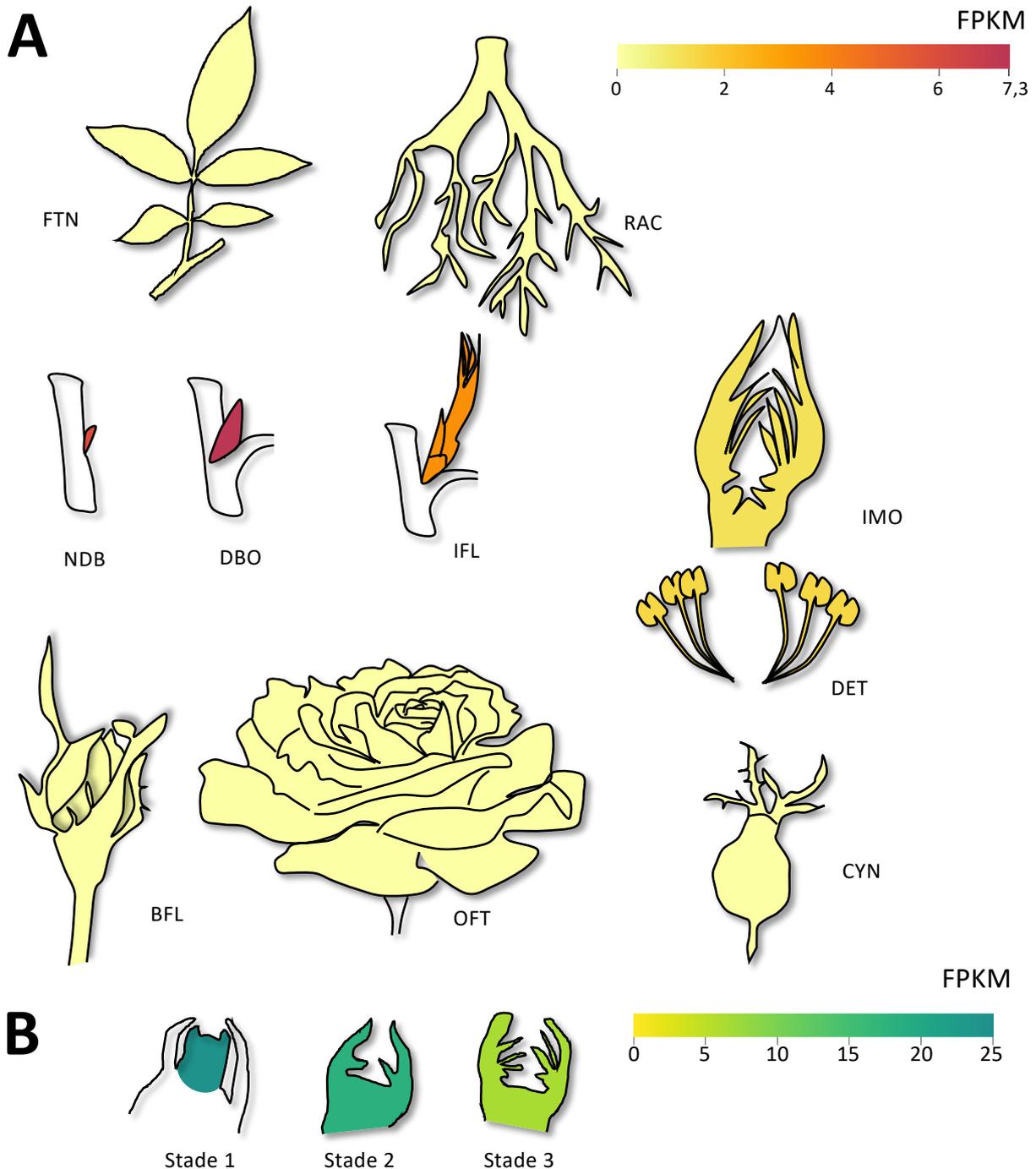


Figure 23 : *RcAP2L* est exprimé essentiellement dans les jeunes tissus aériens.

- (A) RNA-seq réalisé sur différents organes de *R. chinensis* cv 'Old Blush' et décrit dans Dubois et al. (2012). FTN : jeunes feuilles; RAC : jeunes racines; NDB : bourgeons axillaires dormants; DBO : bourgeons axillaires actifs; IFL : bourgeons floraux lors de la transition vers un méristème floral; IMO : méristèmes floraux et développement précoce des organes floraux; BFL : bouton floral fermé; DET : étamines aux stades de microsporogénèse et microgamétogénèse; OFT : fleurs épanouies; SEN : fleurs sénescents; CYN : cynorhodons.
- (B) RNA-seq réalisé sur de très jeunes bourgeons en séparant les stades où se forment respectivement les primordia de sépales (stade 1), les primordia de pétales (stade 2) et les primordia d'étamines (stade 3).

N.B. Les deux expériences ayant été réalisées indépendamment, les niveaux d'expression en (A) et (B) ne sont pas comparables, ce qui justifie le choix de couleurs différentes pour les deux figures.

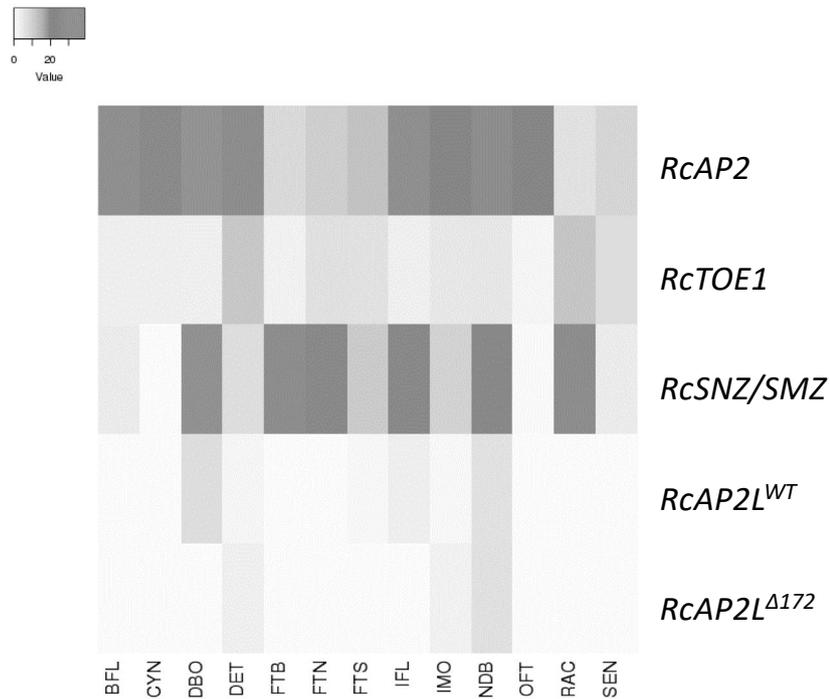


Figure 24 : Comparaison des niveaux d'expression des différents membres de la famille des euAP2.

Heatmaps réalisées à partir des données de RNA-seq réalisé sur différents organes de *R. chinensis* cv 'Old Blush' et décrit dans Dubois et al., 2012. FTN : jeunes feuilles; RAC : jeunes racines; NDB : bourgeons axillaires dormants; DBO: bourgeons axillaires actifs; IFL : bourgeons floraux lors de la transition vers un méristème floral; IMO : méristèmes floraux et développement précoce des organes floraux; BFL : bouton floral fermé; DET : étamines aux stades de microsporogénèse et microgamétogénèse; OFT : fleurs épanouies; SEN : fleurs sénescents; CYN : cynorhodons.

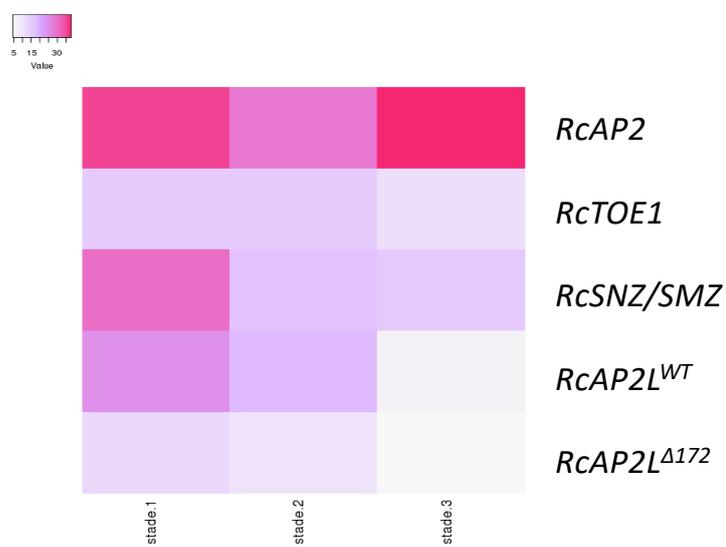


Figure 25 : Comparaison des niveaux d'expression des différents membres de la famille des euAP2 dans les jeunes bourgeons floraux.

Heatmaps réalisées à partir des données de RNA-seq réalisé sur de très jeunes bourgeons en séparant les stades où se forment respectivement les primordia de sépales (stade 1), les primordia de pétales (stade 2) et les primordia d'étamines (stade 3).

membres de la famille des euAP2, et éventuellement de celui de miR172, pour lequel un protocole spécifique serait nécessaire (Wollmann et al., 2010).

2. *RcAP2L, un facteur de transcription pouvant se fixer aux séquences régulatrices de RcAG*

Les membres de la famille AP2/ERF sont caractérisés par la présence d'au moins un domaine de liaison à l'ADN de type AP2 et sont divisées en deux principales sous-familles : les membres de la sous-famille ERF (pour « Ethylene Response Factor ») ne contiennent qu'un seul domaine de liaison à l'ADN alors que ceux de la sous-famille AP2 en contiennent deux (Kim et al., 2006). Bien qu'AP2 soit un facteur de transcription particulièrement important pour le développement floral, aucune structure cristallographique de ses domaines de liaison à l'ADN n'existe à ce jour. En revanche, elle a été déterminée pour deux facteurs de transcription de la sous-famille ERF (*ERF1* et *ERF96*) (Allen et al., 1998 ; Chen et al., 2018). Ce domaine AP2 d'environ 65 acides aminés, aussi appelé GBD pour « GCC-Binding Domain », est organisé en trois brins β antiparallèles, suivis d'une hélice α riche en acides aminés hydrophobes (Allen et al., 1998). Les brins β peuvent s'insérer dans le grand sillon de l'ADN et contiennent des acides aminés capables d'interagir avec l'ADN, soit avec des bases azotées spécifiques soit avec le squelette phosphoribosique (Figure 26-B), alors que l'hélice α est certainement responsable de la structuration et de la compaction du domaine complet (Allen et al., 1998).

L'alignement et la comparaison des séquences protéiques du premier domaine AP2 de RcAP2L et celui de ERF96 permet d'identifier 29 acides aminés conservés entre les deux protéines (Figure 26-A). En particulier, il est possible d'observer une plus grande conservation au niveau des structures secondaires du brin β_3 et de l'hélice α , indiquant que l'organisation du domaine pourrait être suffisamment similaire entre ces deux protéines pour prédire la structure du domaine de RcAP2L à l'aide de celle d'ERF96. Et effectivement, la modélisation de la structure de ce domaine avec le logiciel PHYRE2 (pour Protein Homology/AnalogY Recognition Engine 2 ; Kelley et al., 2015), a permis la prédiction structurale du domaine AP2R1 (Figure 26-B), en se fondant sur la comparaison de la séquence de celui-ci avec ERF96 (présenté sur la figure) ou ERF1 (non représenté mais similaire). Il indique une confiance de 99,9% sur l'ensemble de la structure prédite pour le domaine, ce qui permet d'utiliser cette prédiction pour la suite des analyses.

Alors que près de la moitié des acides aminés capables d'interagir avec le squelette phosphoribosique sont conservés entre les deux protéines (Figure 26-A), seulement un des acides aminés responsables de l'interaction avec les bases elles-mêmes est identique dans les deux protéines (Figure 26-A). Il est donc fort probable que les séquences d'ADN reconnues par les deux protéines soient différentes, ce qui est cohérent avec le fait que les deux sous-familles de ces facteurs de transcription sont impliquées dans des processus biologiques différents et régulent des cibles

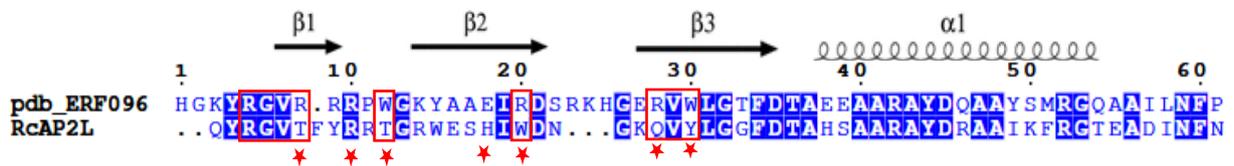
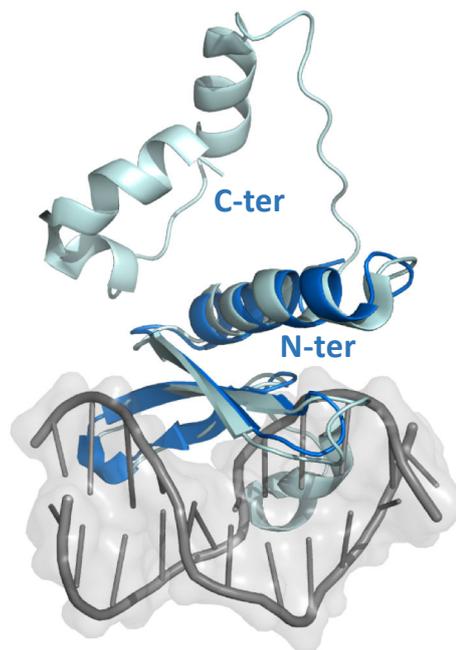
A**B**

Figure 26 : Modèle prédictif de la structure du domaine AP2R1 de RcaAP2L et de son interaction putative avec l'ADN.

- (A) Alignement du domaine AP2R1 de RcaAP2L avec le domaine AP2 de la protéine ERF96. Les acides aminés prédits pour interagir avec le squelette phosphocarboné de l'ADN sont encadrés en rouge alors que ceux prédits pour interagir avec les bases azotées sont marqués d'une étoile (Allen et al., 1998; Chen et al., 2018).
- (B) Superposition de la structure 3D de ERF96 (en bleu clair; Chen et al., 2018) et de la prédiction structurale d'AP2R1 (en bleu foncé) modélisée par PHYRE2 (confiance de 99,9%) à partir de la structure de ERF96.

différentes (Kim et al., 2006). Ainsi, alors que les ERF sont connus pour lier des séquences riches en GC (Allen et al., 1998), il a été montré que le domaine AP2R2 d'AP2 lie une séquence riche en AT (Dinh et al., 2012) ce qui confirme cette idée.

Afin de confirmer l'hypothèse énoncée dans la section précédente et proposant que la résistance au miR172 de *RcAP2L*^{Δ172} est à l'origine de la restriction du domaine d'expression de *RcAG* chez les fleurs doubles, j'ai voulu vérifier si *RcAP2L* était effectivement capable de lier le promoteur de *RcAG*, comme c'est le cas pour les protéines AP2 et TOE3 chez *A. thaliana* (Dinh et al., 2012 ; Jung et al., 2014 ; Yant et al., 2010). Pour cela, une analyse sans a priori semblait adaptée pour quatre principales raisons :

(1) Les séquences régulatrices de l'intron 2 d'AG et *RcAG* sont globalement trop divergentes pour être alignées et comparées entre elles, avec des tailles de 3027 pb et 3491 pb pour AG et *RcAG* respectivement. En particulier, les deux fragments de l'intron 2 d'AG liés par AP2 *in vivo*, nommés fragment AB (159 pb) et fragment II (230 pb), ne sont pas aisément identifiables dans l'intron de *RcAG*.

(2) La séquence nucléotidique cible identifiée par Dinh et al. en 2012 n'a été étudiée qu'à partir du domaine de liaison à l'ADN AP2R2 et non à partir de la protéine entière. Lors de leurs analyses, aucune spécificité dans la séquence d'ADN liée n'a pu être identifiée en utilisant la protéine entière d'AP2.

(3) Cette séquence ne fait que 6 nucléotides (TTTGTT), et est donc trop fréquente dans un génome. En effet, celle-ci est en moyenne trouvée 12 fois rien qu'à proximité du codon d'initiation (+/- 3kb) de chaque gène du génome.

(4) *RcAP2L* appartenant à une sous-famille des euAP2 différentes de celle d'AP2 et de TOE3, ses séquences cibles pourraient être substantiellement différentes.

Par conséquent, j'ai décidé de réaliser des expériences de retard sur gel en présence des protéines codées par les deux allèles de *RcAP2L* et de différents fragments du second intron de *RcAG* (*RcHm5g0012431*), arbitrairement découpé en amplicons chevauchants de 500 pb (Figure 27-A). Ces derniers ont été amplifiés et marqués grâce à des amorces liées au fluorophore IRD dye 700, permettant une révélation du signal grâce au système d'imagerie Odyssey CLx (LI-COR). En parallèle, les protéines sauvage et tronquée ont été fusionnées à un tag Histidine et à la Maltose Binding Protein (MBP), puis produites en système *E. coli* et purifiées sur colonnes d'affinité Nickel-Sepharose grâce au tag Histidine. Elles seront désormais désignées comme *RcAP2L*^{WT}-MBP et *RcAP2L*^{Δ172}-MBP respectivement. Les deux protéines ont ensuite été mises en présence de chacun des fragments du promoteur de *RcAG* et d'ADN non compétitif de saumon à 60 μg·mL⁻¹ permettant de s'assurer de la spécificité de la fixation. Après une incubation d'une heure, les mélanges ont été migrés sur gel d'agarose à 2% permettant alors de visualiser une potentielle interaction physique entre la protéine et le fragment d'intérêt par retard de migration en comparaison à la migration du fragment

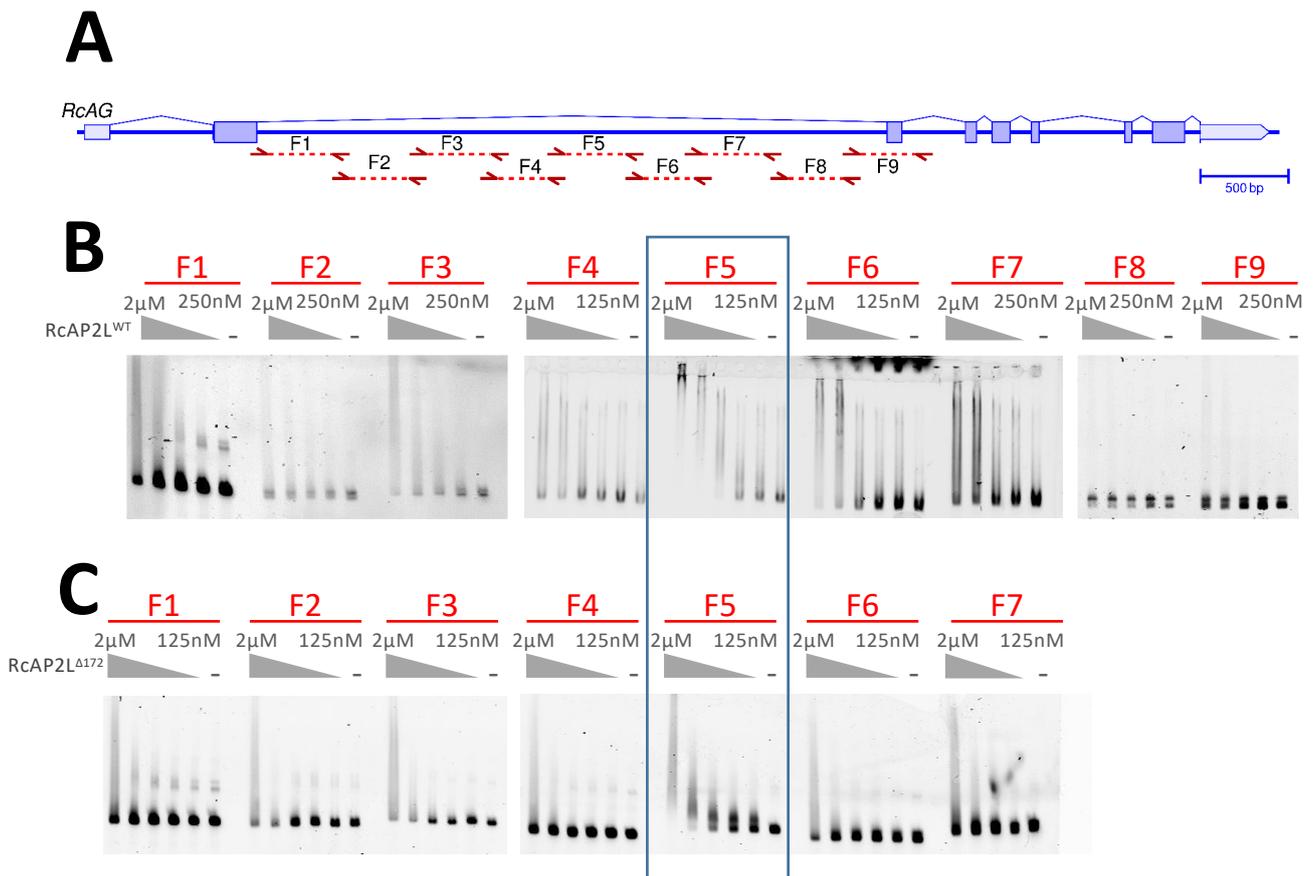


Figure 27 : RcAP2L^{WT} et RcAP2L^{Δ172} sont capables d'interagir spécifiquement avec une portion du deuxième intron de RcAG.

- (A) Structure du gène *RcAG* et position des fragments amplifiés pour les expériences de retard sur gel. L'ensemble de l'intron a été découpée en neuf fragments de 500 pb.
- (B) Retard sur gel comparant la capacité d'interaction entre la protéine RcAP2L^{WT}, fusionnée aux tags Histidine et MBP, et les différents fragments du promoteur de *RcAG*.
- (C) Retard sur gel comparant la capacité d'interaction entre la protéine RcAP2L^{Δ172}, fusionnée aux tags Histidine et MBP, et les différents fragments du promoteur de *RcAG*.

libre, incubé en l'absence de protéine. Dans le cas d'une interaction, l'augmentation progressive de la concentration en protéine, variant de 125 nM à 2 µM, permet de fixer de plus en plus d'ADN libre et ainsi de faire disparaître la bande correspondant à la sonde libre. Les premiers résultats, présentés sur la Figure 27-B, révèlent ainsi qu'à l'exception du fragment 5 et peut-être du fragment 6, seul un faible signal de retard peut être identifié aux plus fortes concentrations, ce qui correspond certainement à une interaction aspécifique. En revanche, dans le cas du fragment 5, la bande correspondant à la sonde libre commence à disparaître dès la plus faible concentration et est totalement consommée à partir de 0,5 µM et 1 µM pour la protéine sauvage et la protéine tronquée, respectivement. Celui-ci est localisé entre les positions 1846 pb et 2376 pb à partir du codon d'initiation de *RcAG* (Figure 27-A). Le fragment 6 présente quant à lui un profil intermédiaire. Ces expériences devront bien sûr être répétées, mais elles indiquent, par comparaison aux autres fragments qui peuvent servir de témoins négatifs, que RcAP2L, aussi bien sous sa forme sauvage que sous sa forme mutée RcAP2L^{Δ172} (Figure 27-B et C), a une plus grande affinité pour le fragment 5 de l'intron de *RcAG* que pour les autres séquences utilisées, confirmant que RcAP2L pourrait reconnaître spécifiquement des séquences régulatrices de *RcAG*. Ainsi, ces premiers résultats permettent désormais d'orienter la recherche du site exact reconnu par cette protéine. En effet, il serait intéressant de continuer à découper le fragment 5 en fragments plus petits afin de mieux caractériser la séquence liée et mieux la comparer à celles d'*A. thaliana*. Alors que l'alignement des séquences complètes de l'intron d'*AG* et de *RcAG* est difficile du fait de la forte variation entre ces séquences, l'alignement MUSCLE du seul fragment 5 avec l'intron de *AG* indique qu'il pourrait contenir la région II, une des régions liées par AP2 chez *A. thaliana* (Bombliès et al., 1999 ; Deyholos and Sieburth, 2000 ; Dinh et al., 2012 ; Sieburth and Meyerowitz, 1997). Enfin, pour aller plus loin, il serait intéressant d'envisager de vérifier cette interaction *in vivo* grâce à la fusion des différents variants de RcAP2L avec la protéine trans-activatrice VP16 et l'utilisation d'un promoteur de *RcAG* incluant son second intron et fusionné à une protéine fluorescente.

Pour conclure, bien que ces expériences doivent être répétées et approfondies, les premiers résultats obtenus semblent en accord avec le modèle que nous avons proposé dans l'article publié en 2018 et où nous proposons que RcAP2L^{Δ172} pourrait inhiber l'expression de *RcAG* en se liant à ses séquences régulatrices (François et al., 2018).

3. Une mutation dans le domaine AP2R1 de RcAP2^{Δ172} pourrait induire la réversion du phénotype « fleur double »

La variété *Rosa chinensis* cv 'Single Pink' a été décrite comme un mutant somatique spontané, ou sport, d'« Old Blush » arborant des fleurs simples (Figure 28-A ; Phillips, 1988 ; Soule, 2009). Or l'élément transposable inséré dans RcAP2L^{Δ172} et proposé comme à l'origine de la production de fleurs

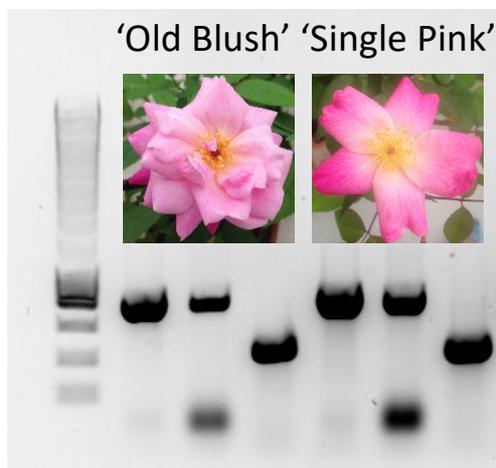
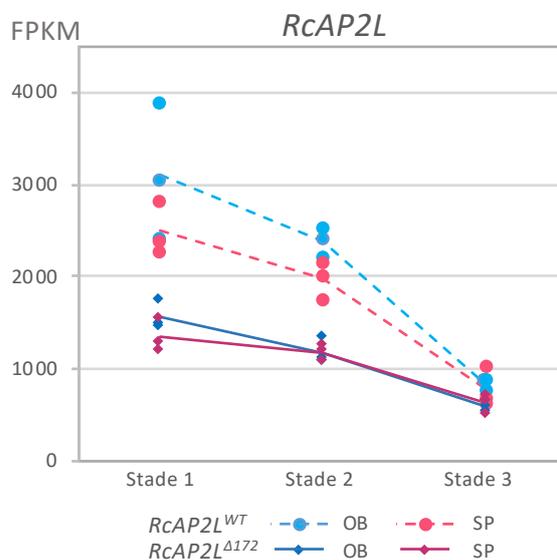
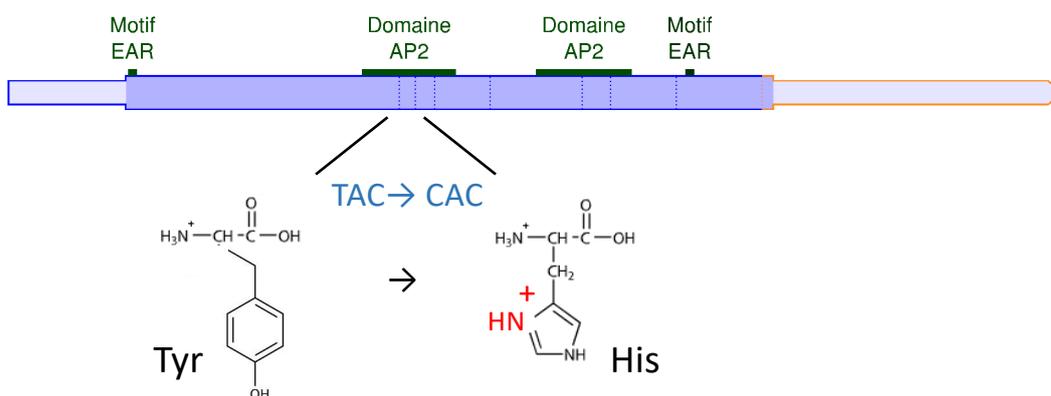
A**B****C**

Figure 28 : Caractérisation d'un mutant spontané d'Old Blush' produisant des fleurs simples et exprimant une version de $RcAP2L^{\Delta 172}$ avec une mutation dans le domaine de liaison à l'ADN AP2R1.

- (A) PCR amplifiant la jonction sauvage (416 pb) et les deux jonctions entre $RcAP2L^{\Delta 172}$ et l'élément transposable (770 et 754 pb) chez la fleur double 'Old Blush' et son mutant spontané à fleur simple 'Single Pink'.
- (B) Niveaux d'expression de $RcAP2L^{\Delta 172}$ et $RcAP2L^{WT}$ déterminés à partir des données transcriptomiques utilisées dans François et al. (2018) et obtenues lors des stades précoces du développement floral de 'OB' et 'SP'.
- (C) Identification d'une mutation ponctuelle dans le premier domaine de liaison à l'ADN de $RcAP2L^{\Delta 172}$ de 'SP'. Cette mutation induit la substitution d'une tyrosine par une histidine, chargée positivement au pH cellulaire.

doubles chez 'Old Blush' est un rétrotransposon de type Ty3/Gypsy (élément à ARN, dit de classe I) qui se propage par un mécanisme de copier-coller dans le génome et ne s'excise généralement pas (Wessler, 2006). Il est donc surprenant de voir une restauration du phénotype « fleur simple » dans un mutant somatique d'Old Blush dans le génome duquel il est encore possible d'identifier par PCR la présence de l'élément transposable inséré dans *RcAP2L*^{Δ172} (Figure 28-A). De plus, l'ARNm correspondant à *RcAP2L*^{Δ172} a pu être détecté par RT-PCR et par RNA-seq réalisés sur des ARNm extraits de jeunes bourgeons floraux (Figure 28-B). Cela confirme l'expression de l'allèle résistant à miR172, comme chez 'Old Blush', ce qui d'après notre modèle devrait conduire à la production de fleurs doubles et semble donc à première vue incohérent avec le phénotype « fleur simple » observé. Cependant, l'alignement des reads provenant du RNA-seq a permis d'identifier une mutation dans le domaine de liaison à l'ADN AP2R1 de *RcAP2L*^{Δ172}, qui a ensuite été confirmée par clonage et séquençage de l'ADNc correspondant. Cette mutation induit le remplacement d'une tyrosine par une histidine à la position 152 dans la protéine (Figure 28-C), et donc dans le brin β3 de la structure tridimensionnelle (Figure 29-A). L'alignement des domaines AP2R1 de l'ensemble des membres de la famille euAP2 provenant de *Petunia hybrida*, *Arabidopsis thaliana*, *Rosa chinensis*, *Prunus persica* et *Medicago truncatula* indique que cette mutation est à une position extrêmement conservée à l'échelle de cette famille (Figure 29-A). Ainsi, cette mutation pourrait être à l'origine de la réversion du phénotype « fleur double » vers un phénotype sauvage « fleur simple », en causant la non-fonctionnalisation de cet allèle.

D'autre part, selon la prédiction structurale déterminée au paragraphe précédent (Figure 26-B), cette tyrosine semble positionnée au contact du grand sillon de l'ADN, à une distance de 4,2 Å d'une base azotée (Figure 29-B), indiquant qu'elle pourrait être impliquée dans la reconnaissance par *RcAP2L* de ses séquences d'ADN cibles. Le remplacement de cette tyrosine par une histidine, en ajoutant une charge positive à cette position, pourrait alors modifier, voire annuler, l'affinité de *RcAP2L* pour ses séquences cibles. Cette proposition a d'autant plus de sens que chez ERF96 tout comme chez ANT, une mutation de l'acide aminé à cette même position, empêche la liaison de ces facteurs à leurs séquences cibles respectives (Chen et al., 2018a ; Krizek, 2003). Pour valider cette hypothèse, j'ai donc répété les expériences de retard sur gel avec la protéine tronquée de 'Single Pink' et le fragment 5 du promoteur de *RcAG*, identifié précédemment comme étant reconnu par *RcAP2L* chez le rosier (Figure 29-C). Ainsi, même si la protéine *RcAP2L*^{Δ172} produite chez 'Single Pink' lie bien ce fragment, c'est avec moins d'affinité que celle d'Old Blush : en effet, à tout autre paramètre constant, pour une même concentration, il reste toujours une bien plus grande quantité de sondes libres en présence de la protéine mutée qu'en présence de la protéine *RcAP2L*^{Δ172} d'Old Blush (Figure 29-C). Par conséquent, il est possible qu'*in vivo* cette mutation dans la protéine *RcAP2L*^{Δ172} de 'Single Pink' ne permette plus

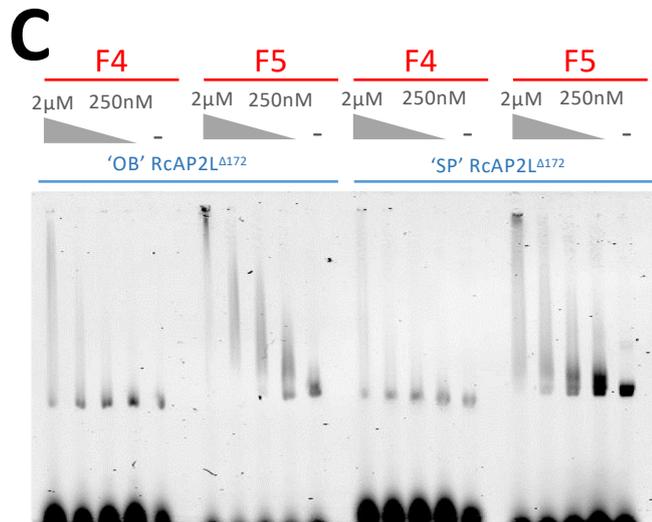
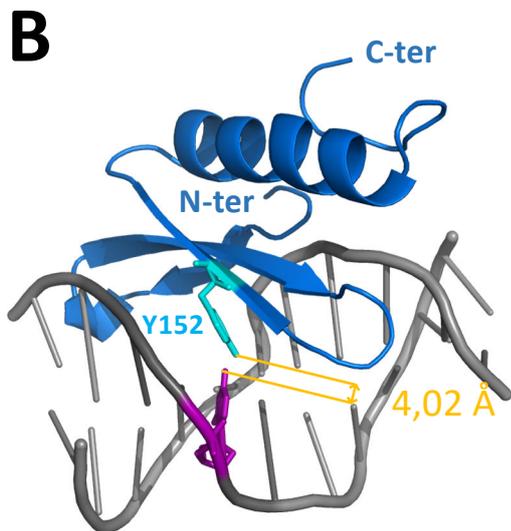
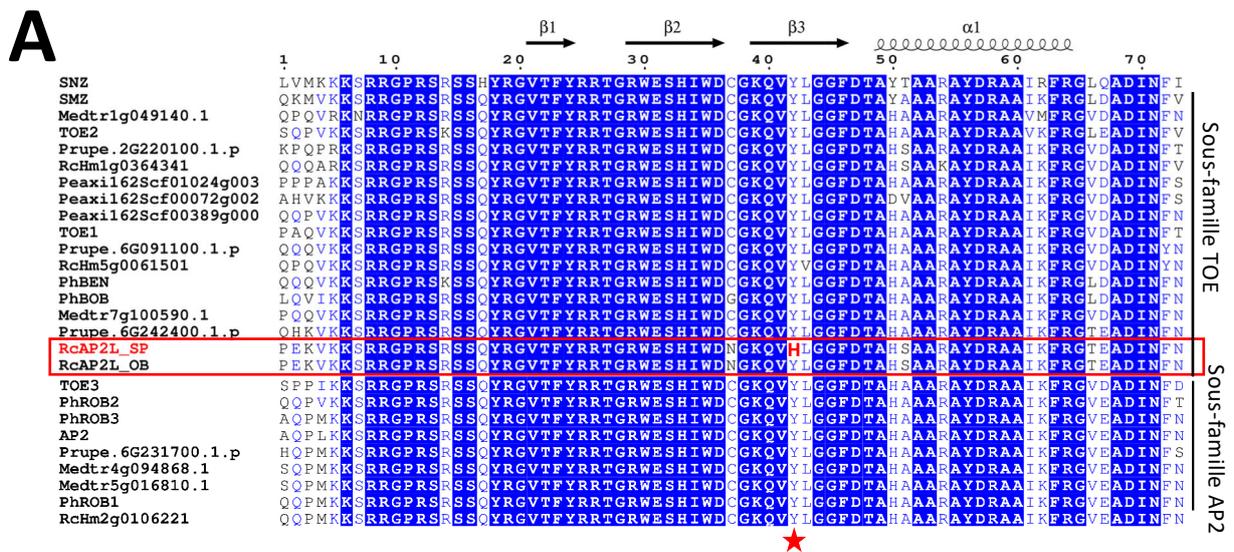


Figure 29 : La mutation identifiée dans le domaine AP2R1 de 'Single Pink' impacte un acide aminé extrêmement conservé au sein de la famille des euAP2 et réduit la capacité de liaison de la protéine au promoteur de *RcAG*.

- (A) Alignement protéique du domaine AP2R1 des euAP2 de *R. chinensis*, *P. persica*, *P. hybrida*, *M. truncatula* et *A. thaliana*. L'étoile rouge indique la position de la tyrosine mutée en histidine, identifiée chez la variété 'Single Pink'. Cette tyrosine est localisée dans le brin β3 interagissant avec l'ADN et est conservée dans toutes les espèces étudiées, à l'exception du mutant 'Single Pink'.
- (B) Prédiction de la position de la tyrosine mutée chez 'SP' selon le modèle structural du domaine AP2R1 réalisée dans la figure 26-B. Celle-ci semble se trouver en contact direct avec l'ADN.
- (C) Retard sur gel comparant les capacités de liaison des protéines RcaP2L⁴¹⁷² d'Old Blush' et de 'Single Pink' aux fragments 4 et 5 du promoteur de *RcAG*. Pour augmenter la solubilité des protéines et faciliter leur purification, celles-ci ont été fusionnées aux tags Histidine et MBP.

la reconnaissance des séquences régulatrices de *RcAG*, qui, nous l'avons vérifié à partir de données de séquençage génomique, sont intactes chez cette variété.

L'identification d'une corrélation secondaire entre la réversion du phénotype « fleur double » vers un phénotype « fleur simple » et la perte de fonction de l'allèle *RcAP2L*^{Δ172} est un argument fort en faveur du modèle déjà proposé dans l'article de 2018. À partir de ces observations, il est alors possible d'étendre ce modèle et de proposer le scénario suivant (Figure 30) : lors du processus de domestication des rosiers chinois, un phénotype de type « fleur double » est apparu suite à l'insertion d'un élément transposable dans *RcAP2L*, le laissant toujours fonctionnel mais le rendant insensible à l'inhibition par miR172 ; l'expression prolongée de cet allèle, et par conséquent l'inhibition prolongée de *RcAG* dans le centre du méristème floral, est alors à l'origine de l'augmentation du nombre de pétales dans cette variété originale. Le phénotype a ensuite été maintenu et introduit dans d'autres variétés de rosiers, ce qui a permis l'obtention des variétés modernes actuelles à « fleur double ». Plus tard, la réversion du phénotype vers un phénotype sauvage à « fleur simple », comme c'est le cas pour les variétés 'Single Pink' et 'T49', peut être induite par la non-fonctionnalisation ou l'absence d'expression de cet allèle résistant au miR172.

4. Le blocage de l'expression de *RcAP2L*^{Δ172} chez 'Old Blush' induit des retards de développement végétatif

Pour aller plus loin dans la confirmation du rôle de *RcAP2L*^{Δ172} dans le passage d'un phénotype « fleur simple » à un phénotype « fleur double », il était important de vérifier si *RcAP2L*^{Δ172} est nécessaire et suffisant pour obtenir des fleurs doubles. Pour cela, une stratégie se fondant sur l'utilisation de microARN artificiels pour cibler et inhiber l'expression *RcAP2L*^{Δ172} a été mise en œuvre durant ma thèse (Figure 38).

Un protocole permettant le choix des séquences des microARN artificiels ciblant les gènes d'intérêt grâce à l'interface web WMD3 (Schwab et al., 2006) ainsi que leur insertion dans le backbone du miR319 caractérisé chez *A. thaliana* était déjà établi dans l'équipe, comme décrit dans la section Matériel et Méthodes (Figure 38). D'autre part, l'installation de l'instance locale du programme WMD3 sur une machine virtuelle hébergée au Centre Blaise Pascal (ENS Lyon), et son adaptation pour tourner sur le génome du rosier décrite dans le *Chapitre III Résultats, section I. Identification d'un premier candidat pour le contrôle du nombre de pétales, partie 6. Diminution du niveau d'expression de RcUGO*, permet dorénavant une vérification de l'absence des off-targets chez le rosier et donc une plus grande précision dans le dessin des guides. Pour cette étude, deux constructions ont ainsi été réalisées et introduites chez 'Old Blush' par transformation médiée par *A. tumefaciens*. La première (amiARN1) est spécifique de *RcAP2L*^{Δ172} alors que la seconde (amiARN2) cible les deux allèles de *RcAP2L* (Figure 31-A). Ces transformations, réalisées en février 2018, commencent tout juste à permettre la régénération

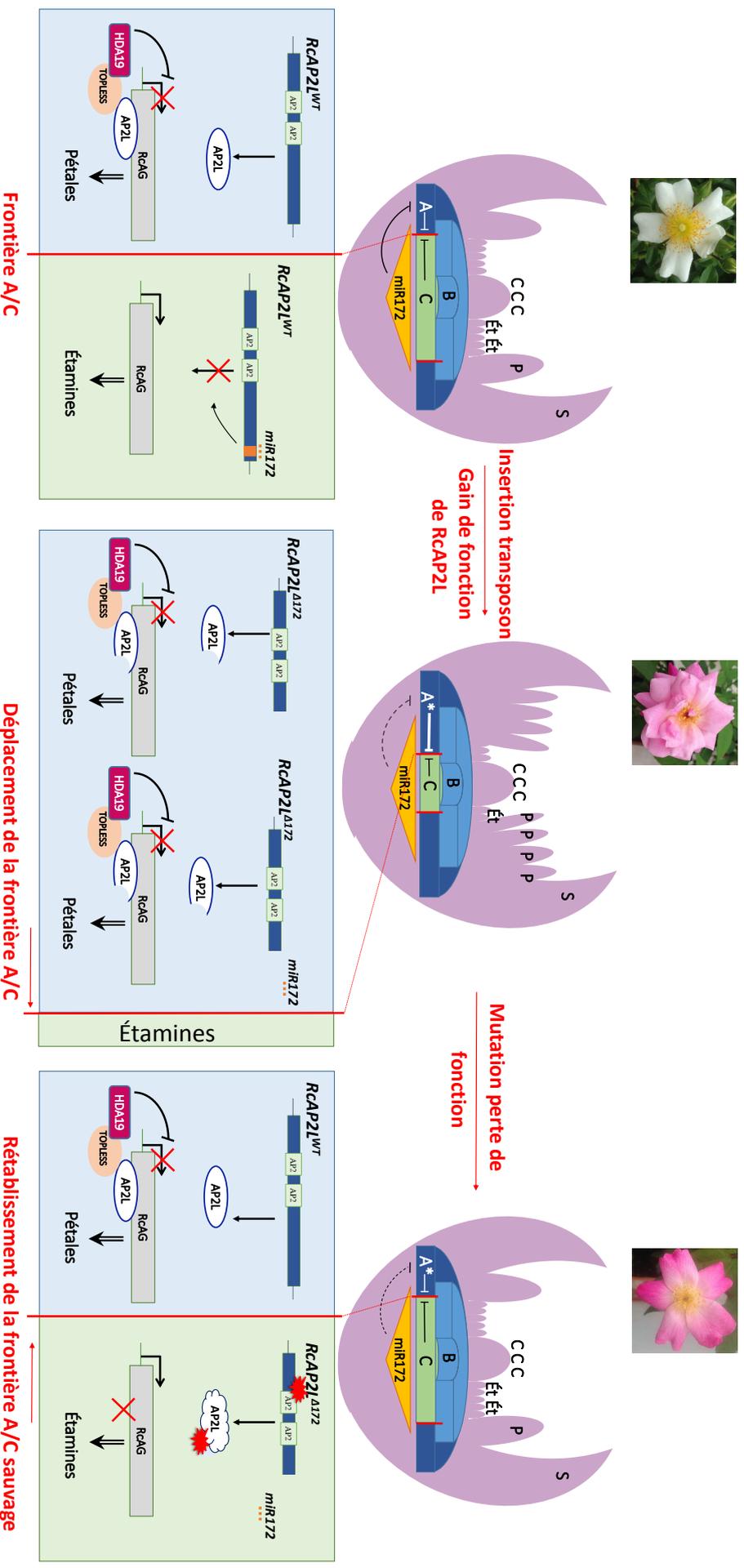


Figure 30 : Rôle de *RcAP2L* dans le contrôle du nombre de pétales.

Modifié à partir de François et al. (2018).

Chez les plantes ancestrales à fleurs simples, *RcAP2L*^{WT} est exprimé seulement dans les verticilles de sépales et pétales où la protéine peut inhiber l'expression de *RcAG* en se liant à ses séquences régulatrices. Dans les verticilles d'étamines et carpelles, *RcAP2L*^{WT} est inhibé par miR172 : *RcAG* peut alors être exprimé et induire la différenciation des primordia en organes sexuels.

Chez 'Old Blush' et ses descendants, l'insertion d'un élément transposable dans *RcAP2L*^{Δ172} rend l'ARNm de cet allèle insensible à l'inhibition par miR172 : il est donc exprimé plus longtemps et restreint d'autant plus l'expression de *RcAG* vers le centre du méristème. Cela induit le déplacement de la frontière A/C vers le centre du méristème et donc la formation de pétales surnuméraires.

Chez le mutant 'Single Pink', une mutation dans le domaine de liaison à l'ADN de *RcAP2L*^{Δ172} empêche la protéine de se fixer aux séquences régulatrices de *RcAG* et la rend non-fonctionnelle. Malgré sa présence prolongée, *RcAP2L*^{Δ172} n'a pas d'effet et seule *RcAP2L*^{WT} peut inhiber *RcAG*, ce qui rétablit une frontière A/C de type sauvage et provoque le retour à un phénotype « fleur simple ».

de plantes exprimant les microARN et les résultats obtenus restent donc préliminaires. Pour l'instant, ces plantes s'acclimatent difficilement à la culture sur terreau et montrent des retards de développement forts comparé à une plante sauvage transférée de culture *in vitro* au même moment (Figure 31-B). Leurs feuilles sont petites et mal formées et aucune fleur n'a encore été produite. Ces effets pléiotropiques, alors que *RcAP2L* n'est en théorie exprimé que dans les méristèmes et anthères, suggèrent soit un rôle essentiel de *RcAP2L* y compris dans le développement végétatif, soit que les produits de dégradation du transcrite de *RcAP2L* dans ces organes pourraient être exportés dans d'autres organes et cibler à leur tour d'autres membres de la famille AP2 partageant des séquences nucléotidiques similaires (Meng et al., 2011). Malheureusement, l'absence de fleurs produites jusqu'à ne permet pas de statuer sur le rôle de *RcAP2L* dans le déterminisme du phénotype « fleur double » pour l'instant.

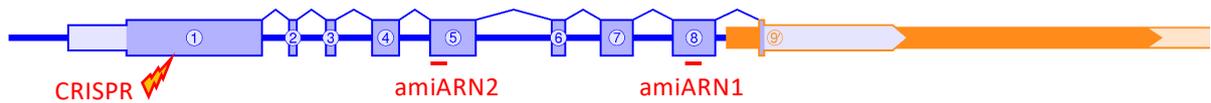
Pour contourner cette difficulté, l'utilisation de la technologie CRISPR/Cas9 (Mali et al., 2013), qui n'a en théorie pas de répercussion sur les autres gènes de la famille, semble une alternative adaptée. Cependant, comme la méthode de mutagenèse par CRISPR en est encore à ses premiers balbutiements chez le rosier, la transformation correspondante a nécessité une durée plus longue de préparation et a été lancée plus tardivement, avec un guide ciblant l'exon 1 de *RcAP2L*^{Δ172} (Figure 31-A). D'ici quelques semaines, il sera possible de vérifier l'édition de *RcAP2L*^{Δ172} dans les cals transformés, avant d'induire la régénération des plantes dont l'édition génomique pourrait rendre la protéine *RcAP2L*^{Δ172} non fonctionnelle. Ensuite, les plantes adultes, qui sont attendues pour le mois d'octobre 2019 au plus tôt, pourront être phénotypées afin d'apporter ou non la validation de nos hypothèses.

5. Surexpression de la version tronquée chez les Rosaceae

Afin de poursuivre les confirmations fonctionnelles, il est aussi nécessaire de confirmer que la seule expression de *RcAP2L*^{Δ172} dans un génotype produisant normalement des fleurs simples est suffisante pour induire la formation de fleurs doubles. Pour cela, la variété commerciale *Rosa pimpinellifolia* cv 'Pimpinelle', seule variété transformable et à fleur simple dans les chambres de culture de l'ENS de Lyon, a été utilisée pour surexprimer sous le promoteur 35S du virus de la mosaïque du chou-fleur (ou CaMV, pour « Cauliflower Mosaic Virus ») les différents allèles de *RcAP2L* identifiés lors de cette étude, soit *RcAP2L*^{WT}, *RcAP2L*^{Δ172} et la version tronquée identifiée chez 'Single Pink' *RcAP2L*_{Sp}^{Δ172}. Bien qu'une première tentative de transformation n'ait pas permis de régénérer de plantes à cause d'un taux de mortalité élevé chez les cals, la seconde a permis d'obtenir récemment de jeunes plantules transformées. Ainsi, une première très jeune plante surexprimant *RcAP2L*^{Δ172} vient d'être obtenue et semble croître correctement. À l'heure actuelle, celle-ci ne fait encore qu'une dizaine de centimètres et devrait fleurir dans les mois qui arrivent. Si une augmentation du nombre de pétales par fleur était observée, ce serait une confirmation directe de l'importance de l'insertion de l'élément

A

RcAP2L^{Δ172}



B



'OB'

RcAP2L_amiARN1

Figure 31 : Chez *Rosa chinensis*, l'inhibition de l'expression de *RcAP2L*^{Δ172} par des microARN artificiels semble perturber le développement végétatif.

(A) Illustration des deux méthodes mises en œuvre pour inhiber l'expression de *RcAP2L*^{Δ172}. Le microARN1 cible l'exon 8 de l'ARNm de l'allèle *RcAP2L*^{Δ172} spécifiquement, alors que le microARN2 cible l'exon 5 des deux allèles de *RcAP2L*. Une édition de génome par la technique CRISPR/Cas9 utilisant un guide spécifique de l'exon 1 de *RcAP2L*^{Δ172} a également été initiée.

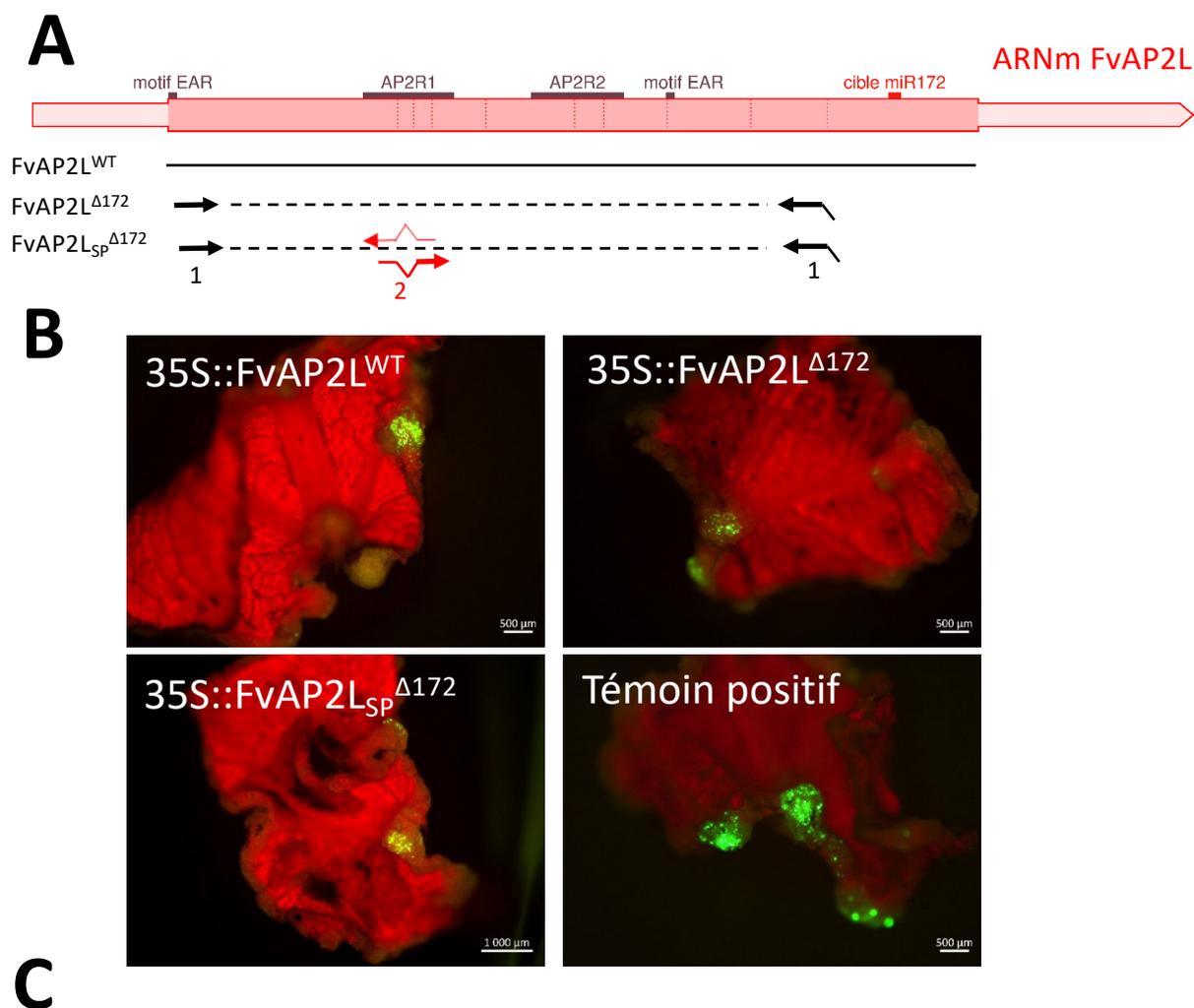
(B) Comparaison du phénotype d'un plant d'"Old Blush" sauvage et d'un plant d'"Old Blush" exprimant le microARN1 trois mois après régénération de culture *in vitro*. Les deux autres plantules génétiquement modifiées sont mortes lors de l'acclimatation. À l'heure actuelle, aucune plantule exprimant le microARN2 n'a été suffisamment forte pour être acclimatée. La forte altération du développement, visible sur cette photo, ne permet pas à la plante de produire des bourgeons floraux; il nous est donc impossible de statuer sur le phénotype floral induit par l'inhibition de *RcAP2L*^{Δ172}.

Les plantes mutagénisées par la méthode CRISPR/Cas9 sont, quant à elles, encore sous forme de cals et en cours de régénération.

transposable dans *RcAP2L*^{Δ172} comme déterminant génétique du phénotype « fleur double ». Dans le but d'aussi confirmer l'impact de la mutation dans le domaine AP2R1 de *RcAP2L*^{Δ172} identifiée chez 'Single Pink', cette construction a aussi été introduite chez 'Pimprenelle' ; les plantes sont en cours de génération et sont attendues plus tardivement (octobre 2019). D'après nos hypothèses, aucune modification du nombre de pétales n'est attendue, par rapport aux cinq pétales habituellement produits par cette variété en laboratoire.

Pour aller plus loin dans l'étude du déterminisme de la formation de la fleur double, nous avons souhaité vérifier si la fonction d'*AP2L* était conservée au sein des Rosaceae, et en particulier chez une espèce très proche du genre *Rosa*, comme le fraisier *Fragaria vesca* (Figure 32). Lors de l'analyse du génome d'Old Blush' en vue de la publication de sa séquence, la récente publication du génome du framboisier, *Rubus occidentalis* (VanBuren et al., 2016), nous a amenés à nous poser des questions sur les relations phylogénétiques précises entre les trois genres très proches *Rosa*, *Fragaria* et *Rubus*. J'ai alors participé à la résolution de leurs liens de parenté, jusque-là encore incertains (Longhi et al., 2014) ou supportés par peu de données (Xiang et al., 2017). Pour cela, 748 séquences de gènes, identifiés comme uniques dans chacun des génomes grâce au logiciel BUSCO (Simão et al., 2015), ont permis de comparer la distance moléculaire entre ces espèces. Les résultats, publiés dans *Nature Genetics* (Raymond et al., 2018) et présentés en annexe, indiquent une distance en moyenne similaire entre chacune des séquences du génome de *R. chinensis* et de celles de *Fragaria vesca* (Shulaev et al., 2008) ou de *Rubus occidentalis* (VanBuren et al., 2016), confirmant que la divergence entre ces trois genres a été quasiment simultanée, ce qui peut être qualifié de radiation évolutive. L'alignement avec MUSCLE (Edgar, 2004) de chacune de ces séquences indépendamment a ensuite permis de réaliser 748 arbres phylogénétiques indépendants révélant que plus de 68% des arbres à très forte confiance (bootstrap > 0,996) supportent l'hypothèse d'une plus forte proximité entre les genres *Rosa* et *Fragaria*. En particulier, la concaténation de l'ensemble de ces séquences, et donc la réalisation d'un arbre phylogénétique intégrant l'ensemble des données de ces 748 gènes, permet d'obtenir un arbre consensus aussi cohérent avec cette hypothèse.

Par ailleurs, l'analyse phylogénétique des membres de la famille euAP2, présentée dans François et al. (2018), a permis d'identifier *FvH4_6g15180* (ou gene18070-v1.0 pour la première version de l'annotation du génome de *F. vesca*) comme l'orthologue de *RcAP2L*. Il sera dorénavant nommé *FvAP2L*. Sa séquence codante, présente à l'état sauvage et donc référée comme *FvAP2L*^{WT}, a été synthétisée par l'entreprise IDT (Integrated DNA Technologies, <https://eu.idtdna.com/>) et clonée dans le vecteur d'entrée pENTR-D/TOPO. Afin de produire les constructions mimant l'impact de l'insertion de l'élément transposable chez 'Old Blush' et 'Single Pink', une version tronquée de la séquence codante (*FvAP2L*^{Δ172}) a été générée à partir de la séquence précédente en utilisant une amorce reverse se terminant par 18 nucléotides additionnels mimant ceux de *RcAP2L*^{Δ172} provenant



Vecteur	Nombre d'explants	Nombre d'explants transformés	Taux de transformation à 10 jours	Survie à 3 mois
Témoin positif	42	28	66.6 %	30%
35S::FvAP2L ^{WT}	104	73	70.1 %	0 %
35S::FvAP2L ^{Δ172}	121	77	63.6 %	0 %
35S::FvAP2L _{SP} ^{Δ172}	96	92	95.8 %	0 %

Figure 32 : La surexpression des différentes versions de *FvAP2L* chez le fraisier n'a pas permis de régénérer de plantes transformées.

- (A) Représentation des 3 constructions générées à partir de la séquence codante de *FvH4_6g15180*, l'orthologue de *RcAP2L* chez le fraisier.
- (B) Signal GFP obtenu à la loupe à épifluorescence à partir de cals de fraisier transformés 10 jours auparavant. Afin de s'assurer de l'efficacité des transformations, chaque construction utilisée possède, en addition du gène d'intérêt, la séquence de la GFP sous le promoteur ubiquitaire proID. Chaque amas de cellules fluorescentes correspond à un événement de transformation génétique indépendant.
- (C) Nombre d'explants transformés identifiés et taux de survie pour chaque construction. Ces transformations, à l'exception de la production des constructions, réalisée par mes soins, ont été réalisées par Aline Potier (INRA de Bordeaux, équipe « Organogénèse du fruit et endoreduplication »). Malheureusement, aucun explant n'a survécu suffisamment longtemps pour permettre la régénération de plantes transformées.

de la fusion de l'exon 8 avec une portion du LTR de l'élément transposable (Tableau 5 ; Figure 32-A). D'autre part, une mutagenèse dirigée permettant d'introduire la mutation identifiée chez 'Single Pink' a été réalisée à partir du plasmide précédent. Ces trois constructions ont ensuite été mises sous le contrôle du promoteur ubiquitaire p35S, dans le plasmide pK7WG2D (Annexe 1), et transformées dans *A. tumefaciens*. Les transformations de cals de fraisier ont ensuite été réalisées en collaboration avec l'INRA de Bordeaux, et plus précisément avec Béatrice Denoyers et Aline Potier (Équipe « Organogenèse du fruit et endoreduplication »), selon le protocole décrit dans la section Matériel et Méthodes. Bien que les vérifications de l'efficacité de transformation aient été positives à 10 jours après transformation (Figure 32-B), aucune plante n'a malheureusement pu être régénérée à partir de ces transformations et ce malgré les deux réplicats indépendants réalisés à un mois d'intervalle (Figure 32-C). Des difficultés similaires avaient été rencontrées lors de la réalisation des transformations équivalentes chez la variété *R. pimpinellifolia* cv. 'Pimprenelle', avec la perte de tous les événements de transformations dans les trois premiers mois de la première expérience, mais une deuxième transformation chez 'Pimprenelle' a cependant été plus fructueuse. Le taux de mort prématurée des cals transformés, inhabituel chez ces deux espèces, pourrait être dû à une certaine toxicité cellulaire d'AP2L, quand celui-ci est exprimé en trop grande quantité.

6. Surexpression de la version tronquée chez *A. thaliana*

Afin de contourner les difficultés rencontrées chez *F. vesca* et le délai important entre le début et la fin d'une transformation chez 'Old Blush', nous avons aussi transformé génétiquement l'écotype Col0 d'*A. thaliana*, avec les constructions déjà utilisées chez 'Pimprenelle'. Cela devait nous permettre d'analyser la conservation de la fonction de *RcAP2L* chez *A. thaliana*. D'après les fonctions des membres de la famille euAP2 assurées chez cette Brassicaceae (Aukerman and Sakai, 2003 ; Chen, 2004 ; Jung et al., 2014), j'ai particulièrement prêté attention à l'impact de ces surexpressions sur la durée de la période végétative avant floraison et le phénotype floral.

Après sélection des transformants sur milieu MS supplémenté en kanamycine, une observation rapide semblait indiquer que les premières fleurs de certains transformants surexprimant *RcAP2L^{WT}* ou *RcAP2L^{Δ172}* avaient plus d'organes qu'une plante sauvage (Figure 33). En effet, certaines fleurs possédaient jusqu'à 7 pétales, une dizaine d'étamines ou encore parfois deux pistils. Cette augmentation du nombre d'organes a aussi été observée chez les plantes surexprimant la version mutée de *RcAP2L_{Sp}^{Δ172}* normalement exprimée chez 'Single Pink', mais le phénotype paraissait moins fréquent et moins fort que pour les deux autres constructions. D'autre part, il s'est avéré que les plantes surexprimant *RcAP2L^{Δ172}* ont fleuri plus tardivement que les plantes sauvages et que les plantes surexprimant la version mutée de *RcAP2L_{Sp}^{Δ172}* (Figure 34-A). L'ensemble de ces observations, obtenues lors la première génération de plantes transformées (T1), semblait cohérent avec le modèle



Figure 33 : Phénotypes floraux observés chez les plantes surexprimant différents allèles de *RcAP2L*.

Effets de la surexpression des CDS de *RcAP2L*^{WT} et *RcAP2L*^{Δ172} lors des premières et deuxièmes générations. L'augmentation du nombre total d'organes (pétales, étamines et carpelles) n'a pas pu être confirmée à la troisième génération, indiquant soit un artéfact soit une diminution de l'expression du transgène au cours des générations (« *silencing* »).

N.B. L'augmentation du nombre d'organes a aussi pu être observée chez certaines plantes surexprimant la version mutée de *RcAP2L*^{Δ172} clonée chez 'Single Pink', mais certainement moins fréquemment que pour les autres transformants. Elle n'a pas été observée chez les plantes sauvages (Col0) cultivées dans les mêmes conditions.

proposé chez le rosier et le rôle des membres de la famille euAP2 dans le déterminisme des pétales, la promotion de l'activité méristématique ou l'inhibition de la mise à fleur chez *A. thaliana* (Aukerman and Sakai, 2003 ; Jung et al., 2014 ; Wollmann et al., 2010). Cependant, après sélection des insertions uniques en T2, ces phénotypes ont commencé à être plus subtils : certaines fleurs possédaient encore des organes surnuméraires mais aucun retard de floraison n'était observable. Encore plus surprenant, après sélection des lignées homozygotes pour l'insert, les troisièmes générations de plantes (T3) semblaient complètement sauvages (Figure 34-B) et ce y compris après un deuxième semis indépendant.

À la lumière de ces résultats, il est malheureusement à déplorer que les phénotypes des T1 et T2 n'aient pas été monitorés plus précisément et notamment qu'ils n'aient pas été quantifiés. En particulier, le nombre de transformants indépendants concernés par l'augmentation du nombre d'organes et l'intensité du délai de floraison, ainsi que la comparaison entre la version sauvage et la version tronquée de *RcAP2L*, sont des informations cruciales qui restent manquantes. Seulement, il est généralement préférable et attendu de mesurer ces données sur des lignées homozygotes pour une insertion unique (T3), et non sur une population hétérogène, avec parfois plusieurs insertions, telle que celle directement issue de la transformation (T1), d'où l'absence de quantification pour ces générations.

Plusieurs hypothèses pourraient expliquer les observations décrites ci-dessus. La première serait que les phénotypes soient artefactuels, liés aux conditions de culture, en particulier pour la durée de la période végétative. En effet, le délai de floraison observé chez les plantes surexprimant *RcAP2L*^{Δ172} pourrait être lié à un stress plus important de ces plantes, comme par exemple des champignons sur les boîtes de culture ou un défaut d'arrosage. Un nouveau semis et phénotypage pourront invalider cette hypothèse, le cas échéant. La deuxième pourrait être celle d'une inhibition progressive de l'expression du transgène par la plante au cours des générations ; ceci est aussi appelé « gene silencing » en anglais. En effet, il arrive que l'utilisation d'un promoteur 35S puisse induire ce type de réaction de défense (Daxinger et al., 2008 ; Okumura et al., 2016 ; Weinhold et al., 2013). Bien que de rares données concernant le phénotype induit par la fusion transcriptionnelle du promoteur 35S avec la séquence d'*AP2* aient déjà été publiées concernant *A. thaliana* (Chen, 2004), il reste possible que l'expression de la construction introduite ici ait été inhibée. Pour étayer cette hypothèse, une extraction d'ARNm à partir de jeunes feuilles des différentes lignées en T2 et T3 a été réalisée, suivie d'une RT-PCR pour révéler la présence ou non de l'ARNm codé par le transgène. Bien que ces données soient encore préliminaires, il semblerait que le transgène ne soit pas exprimé de manière stable et homogène dans toutes les plantes puisque, indépendamment de la génération, certaines plantes l'expriment alors que d'autres non (Figure 34-C). Cela pourrait donc expliquer pourquoi le phénotype est aussi variable en fonction des plantes et serait à confirmer en réalisant de nouvelles

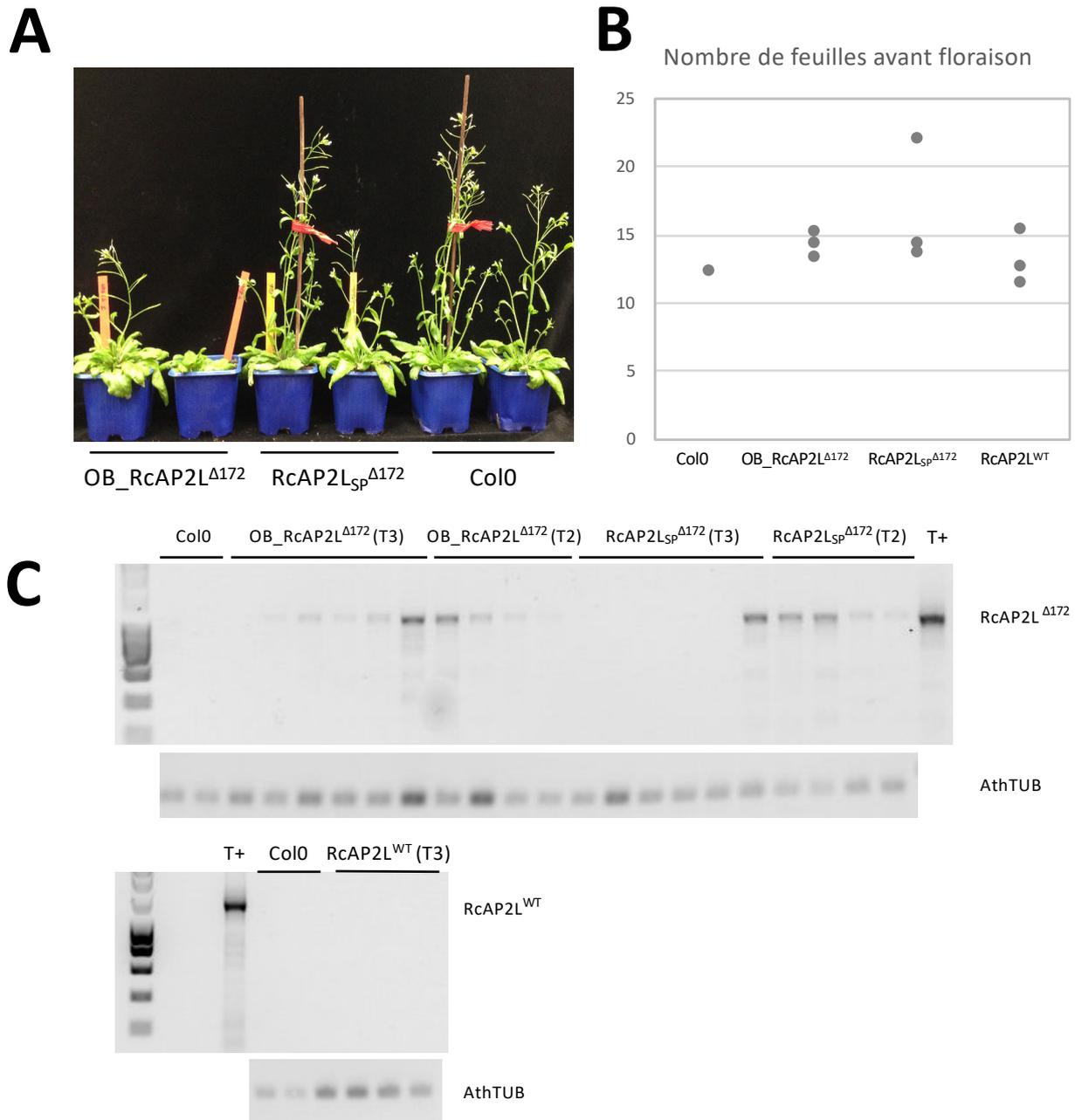


Figure 34 : Impact de la surexpression de *RcAP2L* sur la durée de la période végétative de *A. thaliana*.

- (A) Comparaison de la croissance et de la phase végétative entre Col0 et la première génération de transformants surexprimant le CDS de *RcAP2L*^{Δ172} cloné chez 'Old Blush' ou sa version mutée clonée chez 'Single Pink'. Pour chaque construction, les plantes les plus avancées et les moins avancées parmi les 9 étudiées sont montrées sur cette photo.
- (B) Comparaison du nombre de feuilles de rosette produites avant mise à fleur entre les plantes sauvages (Col0), et les différents transformants après sélection de trois lignées indépendantes par construction, homozygotes et à une seule insertion (3^e génération). Chaque point correspond à la moyenne mesurée sur 8 individus.
- (C) Vérification de l'expression des transgènes par RT-PCR sur des ARNm extraits de jeunes feuilles d'*A. thaliana* transformées. Les témoins positifs (T+) ont été réalisés à partir d'ARNm de jeunes bourgeons de rosier.

Le retard de floraison, observé à la première génération chez les plantes surexprimant OB_RcAP2L^{Δ172}, n'est pas confirmé à la troisième génération, indiquant soit un artéfact soit une diminution de l'expression du transgène au cours des générations (« *silencing* »).

transformations, RT-PCR quantitatives et phénotypages sur des T1 afin de corrélérer l'intensité des phénotypes et le niveau d'expression des transgènes.

7. Conclusions

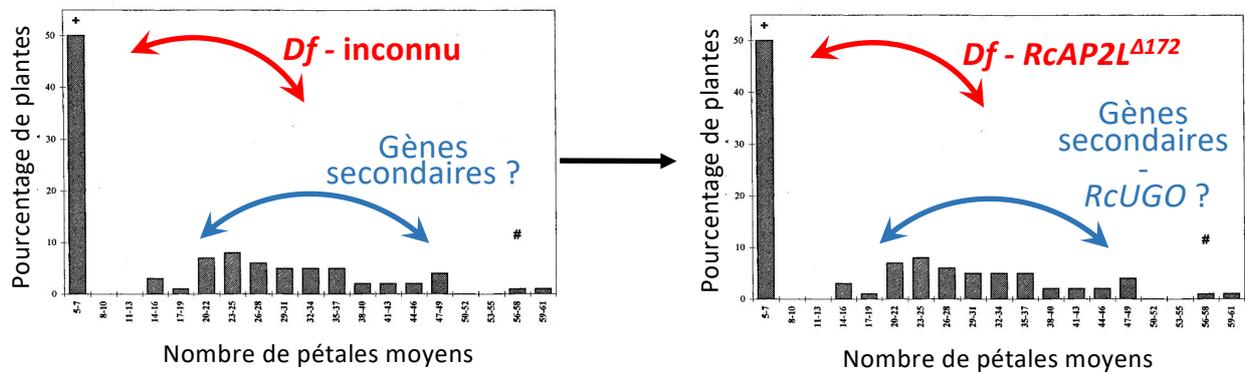
Dans le chapitre précédent et comme nous l'avons publié dans *Scientific Reports* (François et al., 2018), nous avons proposé que la présence de l'allèle muté *RcAP2L*^{Δ172} était à l'origine de la formation de la fleur double chez le rosier. Selon notre hypothèse, *RcAP2L* serait capable d'inhiber l'expression de *RcAG* et, dans le cas des fleurs doubles, son expression prolongée induite par la délétion du site d'inhibition par miR172 induirait une prolongation de cette inhibition et un déplacement de la frontière A/C vers le centre du méristème. Dans ce chapitre, j'ai confirmé que *RcAP2L* interagissait spécifiquement avec une portion du deuxième intron de *RcAG in vitro*, intron connu pour contenir les séquences régulatrices de l'expression d'*AG* chez *A. thaliana*. J'ai aussi confirmé que, bien que tronquée, *RcAP2L*^{Δ172} était aussi capable de reconnaître ce fragment d'intron. Enfin, nous avons identifié un mutant spontané d'Old Blush' à fleurs simples portant une mutation dans le domaine de liaison à l'ADN de *RcAP2L*^{Δ172} qui diminue l'affinité entre la protéine et les séquences régulatrices de *RcAG*. Cette observation, qui suggère que la non-fonctionnalisation de cet allèle est suffisante pour réverter le phénotype, est une première confirmation fonctionnelle. Des transformations génétiques de *R. chinensis*, *R. pimpinellifolia*, *F. vesca* et *A. thaliana* ont été initiées afin de confirmer ces analyses et vérifier si ce mécanisme d'inhibition d'un euAP2 de la sous-famille TOE pourrait être conservé chez différentes espèces. Malheureusement, il s'est avéré que manipuler le niveau d'expression de *RcAP2L* n'est pas chose aisée et de nouvelles analyses seront nécessaires.

Chapitre IV : Discussion et perspectives

Il y a quelques années, notre équipe a démontré que l'augmentation du nombre de pétales chez le rosier était corrélée à une restriction du territoire d'expression du gène *RcAGAMOUS* (*RcAG*) vers le centre du méristème floral (Dubois et al., 2010). Alors que les cartes génétiques localisent la mutation induisant le passage d'une fleur simple à une fleur double sur le chromosome 3, *RcAG* est porté par le chromosome 5, il ne peut donc pas être le déterminant de la fleur double. Nous avons donc cherché des candidats, localisés dans l'intervalle génétique contenant la mutation « fleur double » et intervenant en amont dans la régulation de l'expression de *RcAG*.

Plus récemment, nous avons séquencé, assemblé et publié le génome de *Rosa chinensis* cv 'Old Blush' (Raymond et al., 2018, *Nature Genetics*), un ancêtre des rosiers modernes qui produit déjà des fleurs doubles. Après ce travail auquel j'ai participé, l'assemblage de pseudomolécules continues à l'échelle des chromosomes nous a permis de reconstruire la séquence des deux haplotypes sur l'intervalle génétique contenant le locus « fleur double ». Durant ma thèse, nous avons choisi, parmi les 631 gènes de cet intervalle, deux candidats qui nous ont semblé particulièrement pertinents. J'ai ensuite concentré mon travail sur un gène de la famille des euAP2 (*APETALA2-LIKE* ou *RcAP2L*) qui était localisé dans l'intervalle. Ce gène a attiré notre attention car il existait sous la forme de deux allèles différents chez 'Old Blush', l'un de type sauvage et l'autre contenant l'insertion d'un grand élément transposable. L'analyse de données de RNA-seq a montré que la transcription de cette forme mutée produit un ARN messager tronqué, résistant à l'inhibition normalement assurée par miR172. Sachant que, chez *A. thaliana*, AP2 est connu pour inhiber l'expression d'AG (Dinh et al., 2012 ; Drews et al., 1991), j'ai pu confirmer que les deux allèles de *RcAP2L* sont eux aussi capables de lier spécifiquement une portion du second intron de *RcAG* connu pour contenir les séquences régulatrices de son expression au sein du méristème (Dinh et al., 2012 ; Sieburth and Meyerowitz, 1997), témoignant que *RcAP2L* pourrait inhiber l'expression de *RcAG*. D'autre part, alors que la surexpression d'un variant résistant au miR172 entraîne souvent la formation de pétales supplémentaires (Chen, 2004), j'ai établi que la présence de l'allèle muté (désormais appelé *RcAP2L*^{Δ172}) corrélait avec le phénotype de « fleur double » chez tous les rosiers d'origine chinoise étudiés. Une exception a cependant été identifiée : un mutant spontané d'« Old Blush » à fleurs simples, 'Single Pink', porte une mutation dans le domaine de liaison à l'ADN de *RcAP2L*^{Δ172} qui diminue l'affinité de la protéine pour la séquence régulatrice de *RcAG* identifiée. Cette observation suggère que la non-fonctionnalisation de cet allèle est suffisante pour réverter le phénotype observé, et constitue une nouvelle indication de l'implication de ce gène dans le déterminisme de la fleur double. Selon le modèle que nous proposons, la protéine *RcAP2L*^{Δ172} s'accumulerait du fait de la résistance de son ARNm au miR172 et restreindrait davantage l'expression de *RcAG* vers le centre du méristème floral. Ainsi, la frontière entre les territoires A et C se trouverait elle aussi déplacée vers le centre du méristème, ce qui induirait la conversion des étamines en pétales.

A



B

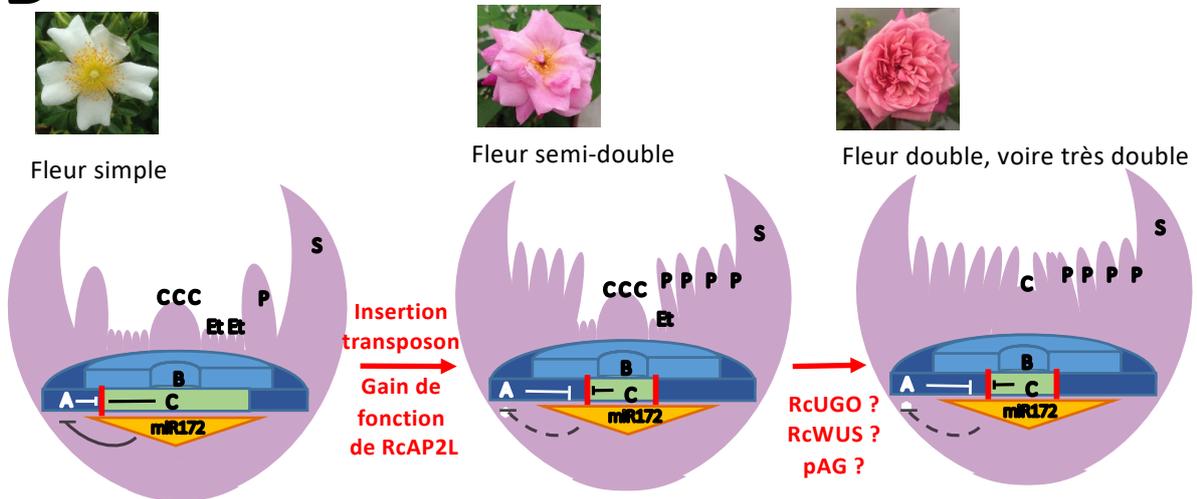


Figure 35 : Bilan des connaissances produites lors de la thèse.

- (A) Nombre moyen de pétales par fleur chez les descendants d'un croisement entre un rosier à « fleur simple » et un rosier à « fleur double » (Debener, 1999). La moitié des descendants produit des fleurs à 5 pétales, l'autre moitié produit des fleurs à plus de 12 pétales mais dont le nombre varie grandement en fonction de l'individu. Le passage d'un rosier produisant des fleurs à 5 pétales vers un rosier produisant des fleurs à plus de 10 pétales est sous le contrôle d'un seul locus chez les rosiers étudiés. Grâce aux travaux présentés dans ce manuscrit, il apparaît dorénavant évident que l'insertion d'un élément transposable dans *RcAP2L* est à l'origine de cette modification de phénotype. La grande variation du nombre de pétales par individu à « fleurs doubles » souligne aussi la présence de gènes pouvant avoir un effet additif sur l'allèle *RcAP2L^{A172}*.
- (B) Proposition de candidats futurs pour le contrôle du nombre de pétales au sein des rosiers à « fleurs doubles ».

Le second gène étudié, *RcUGO*, possède un patron d'expression intéressant étant donné qu'il chevauche ceux de *RcWUS* et *RcAG* et que leurs orthologues chez *A. thaliana* ont un rôle dans l'organisation des méristèmes floraux, suggérant que celui-ci pourrait intervenir dans le contrôle de l'activité méristématique et, pourquoi pas, dans le contrôle du nombre de pétales au sein des fleurs doubles.

Si l'ensemble de ces résultats permet de proposer un modèle pouvant expliquer la formation des fleurs doubles chez le rosier, des validations fonctionnelles supplémentaires sont encore nécessaires pour confirmer définitivement ce modèle et l'étendre à d'autres Rosaceae et pour mieux comprendre l'évolution de la lignée des gènes euAP2 chez les Angiospermes et le rôle de *RcUGO* dans l'organisation du méristème floral. Dans les paragraphes suivants, des expériences permettant d'approfondir les connaissances acquises jusque-là seront discutées, en même temps que la portée des résultats obtenus durant ma thèse. Toutefois, les points déjà abordés dans la discussion du papier François et al. (2018) ne seront pas approfondis de nouveau ici.

I. Vers une meilleure compréhension du réseau de signalisation déterminant le nombre de pétales chez les rosiers

1. Confirmation que *RcAP2L*^{Δ172} est nécessaire et suffisant au phénotype « fleur double » chez les rosiers d'origine chinoise

Afin de confirmer que l'expression de l'allèle muté *RcAP2L*^{Δ172} est nécessaire et suffisant pour produire des fleurs doubles, il est important de montrer que l'insertion de cet allèle dans des rosiers à fleurs simples induit la formation de fleurs doubles et, qu'au contraire, la suppression de cet allèle du fond génétique d'un rosier à fleurs doubles induit la réversion de celui-ci vers un phénotype à fleurs simples. L'identification d'une mutation dans l'allèle *RcAP2L*^{Δ172} chez un mutant somatique de *R. chinensis* cv 'Old Blush' produisant de nouveau des fleurs simples, est un argument fort en faveur de la nécessité de la présence de cet allèle pour produire des fleurs doubles. En effet, cette mutation, localisée dans le domaine de liaison à l'ADN de la protéine *RcAP2L*^{Δ172}, réduit en conditions expérimentales l'affinité de cet allèle pour les séquences régulatrices de *RcAG*, suggérant que c'est sa non-fonctionnalisation qui induit la réversion du phénotype « fleur double » vers un phénotype « fleur simple ». Seulement, bien que cette observation soit très convaincante ce n'est pas une preuve irréfutable étant donné que ce n'est pas la seule différence génomique présente entre 'Single Pink' et 'Old Blush', dont la divergence a commencé il y a au moins une centaine d'années (Phillips, 1988 ; Soule, 2009). La confirmation définitive du modèle implique la transformation génétique d'"Old Blush" pour inhiber *RcAP2L*^{Δ172}. Étant donné les problèmes de croissance observés lors de la première manipulation du niveau d'expression de *RcAP2L*^{Δ172} par microARN artificiels exprimés sous le promoteur ubiquitaire fort 35S, une première possibilité serait de répéter l'insertion du même microARN

artificiel sous un promoteur spécifique des méristèmes floraux ; cela devrait limiter l'impact de celui-ci aux primordia floraux et ne plus perturber la croissance et la floraison des plantes. Idéalement, il serait intéressant d'utiliser le propre promoteur de *RcAP2L* mais, jusqu'à présent, les tentatives de clonage de celui-ci n'ont pas été fructueuses. Une autre serait de mutagéniser l'allèle *RcAP2L*^{Δ172} d'Old Blush' par la méthode CRISPR/Cas9 (Mali et al., 2013) ; cette solution a déjà été mise en œuvre, les plantes sont en cours de régénération et sont attendues pour le début de l'hiver 2019. D'autre part, la génération de rosiers sauvages initialement à « fleurs simples » surexprimant les allèles *RcAP2L*^{WT}, *RcAP2L*_{OB}^{Δ172} et *RcAP2L*_{SP}^{Δ172} devrait aussi permettre de savoir si la présence d'un allèle *RcAP2L*^{Δ172} fonctionnel est suffisante pour induire la formation de fleurs doubles ; ces plantes sont attendues pour l'automne 2019. Selon le modèle, il est attendu que seules les plantes surexprimant *RcAP2L*^{Δ172}, et peut-être les plantes surexprimant trop fortement *RcAP2L*^{WT} pour que l'ensemble des ARNm soient inhibés par miR172, verront leur nombre de pétales augmenté.

Enfin, des constructions plus exotiques auraient aussi pu être envisagées : en particulier il aurait été possible de réintroduire une cible de miR172 à l'allèle *RcAP2L*^{Δ172} ou de muter spécifiquement le site de fixation de miR172 de l'allèle *RcAP2L*^{WT} afin de vérifier que le phénotype est bien lié à la résistance au microARN et non à la perte de la portion en C-terminal de la protéine ou à un autre effet secondaire de l'insertion de l'élément transposable.

2. Identification des partenaires de *RcAP2L* dans la régulation de l'expression de ses cibles

Il serait intéressant de mieux caractériser la ou les voies de signalisation impliquée(s) avec *RcAP2L* dans le contrôle du nombre de pétales.

Pour commencer, il serait pertinent d'identifier quelles sont les protéines, et notamment les cofacteurs, nécessaires au fonctionnement de *RcAP2L*^{Δ172}. Pour cela, deux approches pourraient être envisagées. La première est une approche avec *a priori* ; elle pourrait se fonder sur la littérature existante chez *A. thaliana* pour choisir des candidats probables pour jouer le rôle de cofacteurs de *RcAP2L*. Notamment, il a été montré qu'AP2 pouvait, grâce à son domaine EAR positionné à proximité de son domaine de liaison à l'ADN, interagir physiquement avec TOPLESS (Causier et al., 2012 ; Krogan et al., 2012) et le recruter à proximité de ses gènes cibles (Krogan et al., 2012). Ensemble, ils recrutent l'histone déacétylase HDA19, capable d'effacer notamment les marques H4K16Ac, réduisant ainsi l'accessibilité de l'ADN aux polymérase et inhibant l'expression des gènes à proximité (Krogan et al., 2012). Le rôle de ces protéines dans la restriction de l'expression d'AG aux verticilles 3 et 4 est confirmé par l'observation d'une diminution du nombre total d'organes floraux et de la présence de sépales carpelloïdes chez le quintuple mutant de la redondante famille TOPLESS *tpl-2 tpr1 tpr2 tpr3* et *tpr4* ou chez le mutant *hda19* (Krogan et al., 2012). Fait particulièrement intéressant, la complémentation du

mutant *ap2* par l'insertion d'une fusion de TPL avec le domaine de liaison à l'ADN d'AP2 seulement, indique que, chez *A. thaliana*, AP2 fonctionne essentiellement en recrutant TPL à proximité de ses cibles pour inhiber leur expression. Ainsi, il pourrait être intéressant de vérifier si cette interaction est conservée chez le rosier et si la perte de fonction de *RcTPL* pourrait être une source de réversion de phénotype « fleur double ». Pour cela, des expériences de *Virus Induced Gene Silencing* (VIGS) (Unver and Budak, 2009) pourraient être envisagées afin d'étudier l'effet de la diminution de l'expression d'un ou plusieurs membres de la famille *TPL* ou de *RcAP2L* sur la structure florale. Il serait ensuite possible d'analyser les niveaux d'acétylation des histones à proximité du gène *RcAG*, en même temps que son niveau d'expression par qRT-PCR selon ces différentes conditions d'inhibition. La comparaison entre le niveau d'acétylation dans des bourgeons aux stades de formation des sépales et pétales avec celui des bourgeons où *RcTPL* et/ou de *RcAP2L* sont inhibés permettrait de confirmer l'interaction entre les deux protéines et leur effet sur l'expression de *RcAG*.

La deuxième approche, sans *a priori*, pourrait se baser sur des expériences de co-immunoprécipitation couplées à une spectroscopie de masse, méthode de choix pour identifier toutes les protéines capables d'interagir avec une protéine d'intérêt, et dans ce cas *RcAP2L* (Köcher and Superti-Furga, 2007). Cette expérience serait dorénavant facilitée par l'existence de protéines *RcAP2L* marquées par un tag Histidine couplé ou non à un tag MBP, que j'ai générées pour les expériences de retard sur gel. Elles permettront en effet l'utilisation d'anticorps commerciaux, manipulés quotidiennement au laboratoire, sans passer par la production d'anticorps spécifiques à *RcAP2L*. Ces analyses pourraient faciliter l'identification de l'ensemble des interactants de la voie de signalisation liée à *RcAP2L* et, pourquoi pas, mettre en évidence de nouveaux candidats potentiels dans la régulation du nombre de pétales au sein des rosiers à fleurs doubles ou bien dans la domestication des fleurs doubles chez d'autres espèces ornementales. Cependant, la possibilité ou non que ceux-ci aient joué un rôle dans la domestication des fleurs doubles dépend aussi de la nature de la protéine et de ses fonctions. En effet, plus un gène est ubiquitaire, ou plus il possède des effets pléiotropiques, moins il est vraisemblable que celui-ci soit un candidat plausible car sa modification aurait induit trop d'effets délétères pour la plante. Ainsi, il est probable que, malgré le rôle essentiel de TPL dans l'action d'AP2 chez *A. thaliana*, les membres de la famille de *TOPLESS* n'aient pas été ciblés lors de la sélection des plantes ornementales à fleurs doubles car leur modification induirait certainement des modifications phénotypiques secondaires potentiellement inadéquates.

3. Identification des gènes cibles de *RcAP2L*

D'autre part, il serait intéressant de vérifier si l'effet de *RcAP2L*^{*Δ172*} sur l'augmentation du nombre de pétales se fait uniquement via son inhibition de l'expression de *RcAG* ou si d'autres cibles de *RcAP2L* sont tout aussi importantes. Pour commencer, il serait pertinent de mieux identifier la

séquence d'ADN liée par RcAP2L dans l'intron de *RcAG* en répétant les expériences de retard sur gel sur des portions plus courtes du fragment 5. La connaissance exacte de la séquence ciblée aurait deux conséquences pour le futur. Tout d'abord, cela permettrait de tenter de muter cette séquence par la méthode CRISPR/Cas9 et ainsi voir si cela induit la réversion du phénotype fleur double vers un phénotype fleur simple. Alors que la méthode de mutagenèse par CRISPR/Cas9 en est encore à ses premiers pas chez le rosier, l'utiliser sur les séquences régulatrices de *RcAG* serait une façon élégante et relativement aisée de confirmer son applicabilité. En effet, il suffirait d'un allèle de *RcAG* insensible à la régulation de RcAP2L pour contrer l'effet du gain de fonction induit par l'insertion de l'élément transposable dans *RcAP2L*^{Δ172} sur le domaine d'expression de *RcAG*. Si une réversion du phénotype était observée, cela confirmerait que contrôler l'expression de *RcAG* est vraiment la clé de la formation des fleurs doubles chez le rosier, comme c'est le cas chez de nombreuses autres espèces appartenant à d'autres familles (Galimba et al., 2012 ; Ingle and Gilmartin, 2015 ; Liu et al., 2012 ; Ma et al., 2018 ; Nitasaka, 2003).

Cependant, si la réversion n'était pas complète, cela signifierait que d'autres gènes cibles de *RcAP2L* sont aussi impliqués. Alors, il serait nécessaire de rechercher d'autres gènes régulés par *RcAP2L*, par exemple grâce à des analyses de ChIP-seq (pour Chromatine ImmunoPrecipitation suivie d'un séquençage), qui permettrait de séquencer l'ensemble des séquences d'ADN auxquelles RcAP2L peut se fixer et ainsi avoir accès à l'ensemble des cibles potentielles de celle-ci (Chen et al., 2018b). Les candidats obtenus à partir de cette analyse sans *a priori* pourraient ensuite être confirmés en recherchant les séquences cibles de RcAP2L identifiées dans l'intron de *RcAG* ou en répétant les expériences de retard sur gel afin de regarder si RcAP2L interagit véritablement avec leurs séquences régulatrices. En se fondant sur les gènes cibles déjà identifiées chez *A. thaliana* et *P. hybrida*, on attendrait par exemple à retrouver les orthologues d'*AP3* et de *SEP3* (Krogan et al., 2012 ; Morel et al., 2017).

L'identification de l'ensemble des voies de signalisation impliquées en aval de *RcAP2L* permettrait ensuite d'orienter les recherches pour répondre aux nombreuses questions encore en suspens à l'heure actuelle et développées dans les paragraphes suivants.

4. *Les rosiers non chinois à fleurs doubles seraient-ils mutés dans un autre intervenant de la voie RcAP2L ?*

L'histoire de la domestication des rosiers modernes possède deux volets parallèles. En effet, les rosiers sauvages ont été domestiqués par deux fois, de façon indépendante : une fois en Asie et une en Europe (Bendahmane et al., 2013). Les multiples croisements et changements de niveau de ploïdie rendent la reconstruction de l'histoire des rosiers modernes difficile. La généalogie la plus précise disponible aujourd'hui a été publiée dans Bendahmane et al. (2013) et est présentée en

Figure 36. Ainsi, le phénotype « fleur double » a été sélectionné deux fois, indépendamment, avant que les deux lignées ne soient mises en contact au XVIII^e siècle, suggérant que les mécanismes pourraient être différents. Dans les deux cas, une restriction de l'expression de *RcAG* a été observée (Dubois et al., 2010), indiquant que les deux mutations, bien qu'indépendantes, doivent avoir un effet similaire sur *RcAG*.

Les travaux effectués dans cette thèse ont majoritairement concerné les rosiers modernes d'origine chinoise, dont la caractérisation génétique et moléculaire du mécanisme à l'origine de la formation des fleurs doubles a commencé par les cartographies présentées en introduction et indiquant que la mutation était dominante (Debener, 1999 ; Dugo et al., 2005 ; Spiller et al., 2010). Cependant, des données préliminaires et non publiées semblent indiquer que ce n'est pas le cas pour les rosiers d'origine péri-méditerranéenne dont la domestication et la sélection du phénotype « fleur double » a été faite indépendamment (Bendahmane et al., 2013 ; Dubois et al., 2010). En effet, l'allèle *RcAP2L*^{Δ172} est absent du génome des rosiers péri-méditerranéens à fleurs doubles analysés, comme par exemple *R. damascena* et *R. gallica* cv 'officinalis' (Figure 36). Cela est d'autre part cohérent avec le mécanisme potentiellement co-dominant identifié (Olivier Raymond, communication personnelle, données non publiées). Alors que la position de la mutation n'a pas encore été étudiée, un mécanisme différent doit probablement être en jeu, peut-être en lien avec les autres intervenants de la voie de signalisation de *RcAP2L* (section précédente) ou bien directement au niveau de *RcAG* ou de ses séquences régulatrices. On pourrait par exemple imaginer qu'une des séquences activatrices, c'est-à-dire recrutant un facteur de transcription capable d'induire l'expression de *RcAG*, comme LEAFY, APETALA1 ou NF-Y (*NUCLEAR FACTOR-Y*) (Hong et al., 2003), ait été rendue non fonctionnelle. Ainsi, bien que l'activateur soit présent, celui-ci ne pourrait plus se lier au promoteur de *RcAG* et induire son expression, induisant un délai dans l'activation de *RcAG* et donc un déplacement de la frontière A/C vers le centre du méristème selon un effet dose cohérent avec la semi-dominance du phénotype.

5. Origine du contrôle polygénique du nombre de pétales au sein des rosiers à fleurs doubles

Alors que le passage d'une fleur simple à une fleur double est sous le contrôle d'un unique gène chez les rosiers chinois, les fleurs doubles issues d'un croisement entre un rosier à fleur simple et d'un rosier à fleur double présentent un nombre de pétales par fleur variant d'une dizaine à plusieurs dizaines (Debener, 1999), suggérant l'intervention de nombreux autres gènes secondaires qui pourraient moduler l'effet de la mutation de *RcAP2L* (Figure 35). Jusqu'à présent, étant donné que le déterminant de la fleur double lui-même n'était pas encore identifié, les QTL (pour « Quantitative Trait Loci ») associés à la variance du nombre de pétales au sein des fleurs semi-doubles à doubles n'ont été que très peu étudiés. Ainsi, une des rares tentatives pour localiser ces QTL avait donné lieu à la

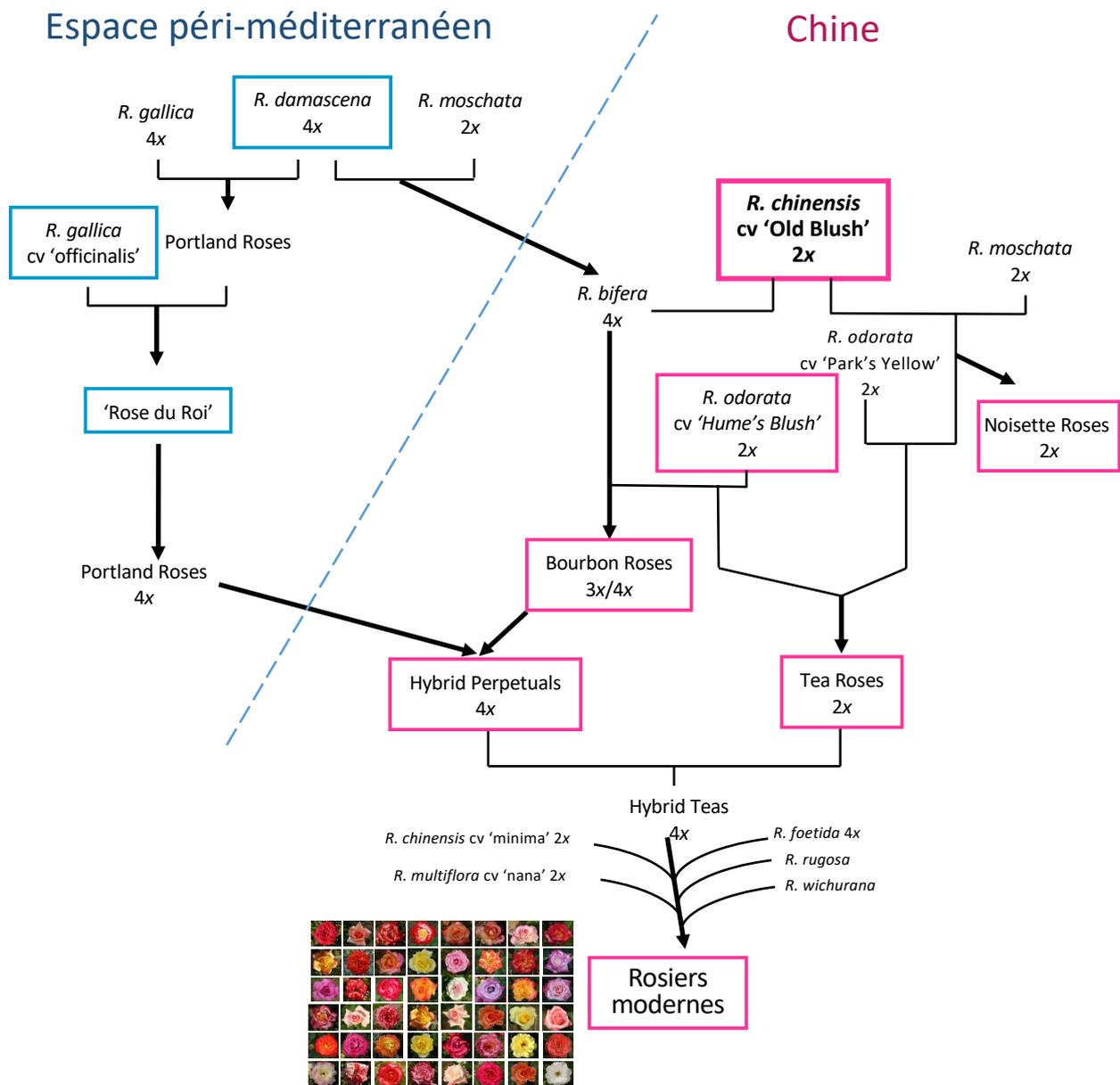


Figure 36 : Généalogie des rosiers modernes et déterminisme génétique du phénotype « fleur double ». Généalogie théorique des rosiers modernes selon l'origine de leurs ancêtres, adaptée de Bendahmane et al. (2013). À droite, les rosiers d'origine chinoise (sections Chineses et Synstylae) et leurs descendants. À gauche, les rosiers domestiqués dans l'espace péri-méditerranéen (section Gallicanae). Les premières indications de mise en contact des différentes variétés des deux bassins de domestication datent du XVIII^e siècle, suggérant une absence de flux génétiques jusqu'à cette époque et donc deux sélections indépendantes du phénotype « fleur double ».

Les variétés encadrées en rose correspondent aux variétés à fleurs doubles testées et possédant l'allèle *RcAP2L*^{Δ172} alors que les variétés encadrées en bleu correspondent aux variétés à fleurs doubles testées mais ne possédant pas l'allèle muté. Alors que la corrélation entre *RcAP2L*^{Δ172} et le phénotype « fleur double » a été mise en évidence chez les rosiers d'origine chinoise, ce n'est pas le cas chez les rosiers d'origine péri-méditerranéenne. Un mécanisme différent a certainement été sélectionné ; celui-ci est très peu caractérisé à ce jour.

proposition de la présence d'un autre gène impactant lui aussi le nombre de pétales à proximité du locus déterminant le passage d'une fleur simple à une fleur double (Hibrand-Saint Oyant et al., 2007) mais aucune recherche n'a depuis pu confirmer ces résultats. La découverte de *RcAP2L* comme facteur déclencheur du passage d'une fleur simple à une fleur double permet dorénavant de chercher des candidats pour le contrôle polygénique du nombre de pétales au sein des variétés à fleur double, en commençant par les cibles de *RcAP2L* qui pourraient amplifier le phénotype observé.

6. Recherche d'un lien entre *RcAP2L* et *RcUGO*

À partir des nouvelles données générées par l'étude encore en cours des mécanismes de fonctionnement de *RcAP2L*, il serait intéressant de rechercher un éventuel lien entre *RcAP2L* et *RcUGO* dans le contrôle du nombre de pétales. Ainsi, peut-être que le niveau d'expression de *RcUGO* est partie intégrante de ce contrôle multigénique du nombre de pétales au sein des fleurs doubles. Étant donné qu'ils appartiennent à des familles de protéines complètement différentes qui ne sont pas connues pour interagir entre elles chez *A. thaliana*, il est peu probable que ces deux protéines interagissent physiquement mais peut-être interviennent-elles dans la même voie de signalisation. En particulier, d'après la littérature, ces deux protéines semblent toutes les deux contrôler l'activité méristématique et très certainement le niveau d'expression de *WUS* (Brandt et al., 2013 ; Würschum et al., 2006). Appartiennent-elles à la même voie de signalisation ? *RcUGO* est-il sous le contrôle de *RcAP2L*, comme pourrait le suggérer sa plus forte expression chez le rosier à fleurs doubles 'Old Blush' ? Plus précisément, y aurait-il des séquences cibles de *RcAP2L* dans le promoteur de *RcUGO* ? Et, bien que *RcUGO* ne soit très certainement pas à l'origine de la formation de la « fleur double » chez le rosier, pourrait-il intervenir dans le contrôle polygénique du nombre de pétales chez les rosiers à « fleur double », notamment en promouvant l'activité méristématique ? En effet, la position de *RcUGO* sur le chromosome 3, semble correspondre à l'intervalle identifié comme contenant le principal QTL contrôlant le nombre de pétales chez les fleurs doubles (Hibrand-Saint Oyant et al., 2007), ouvrant la possibilité que *RcUGO* pourrait correspondre à celui-ci (Figure 35).

II. Importance horticole de l'identification de la mutation de *RcAP2L*

1. Vers une accélération du processus de création variétale chez le rosier

L'identification du déterminant de la fleur double est une véritable avancée à la fois technologique et fondamentale. Sur le plan horticole, il sera dorénavant aisé de déterminer si de jeunes plantules issues d'un croisement produiront des fleurs simples ou doubles, permettant ainsi un gain de temps, d'espace et d'eau important pour les sélectionneurs. En effet, une simple extraction d'ADN sur des plantules de quelques semaines, suivi d'une PCR de génotypage, comme présentées dans François et al. (2018), seront suffisantes pour déterminer si l'allèle *RcAP2L*^{A172} est présent, et donc

prédire le phénotype floral avant même que les plantes ne soient capables de fleurir. Dans le cas où d'autres espèces porteraient le même type de mutation, comme cela semble le cas chez les pêcheurs ornementaux (Gattolin et al., 2018), cette méthode pourrait être aussi transférée chez ces espèces.

2. *Mutation fleur double dominante chez d'autres espèces ornementales : une modification de la famille euAP2 ?*

D'autre part, l'identification du rôle d'un membre de la famille euAP2 résistant à miR172 dans la domestication des fleurs doubles chez au moins deux espèces de Rosaceae, le rosier et le pêcheur (François et al., 2018 ; Gattolin et al., 2018) ouvre la voie à des recherches similaires chez d'autres espèces ornementales pour lesquelles le déterminisme de la fleur double est encore inconnu et, en particulier, celles pour lesquelles la mutation est dominante. Alors qu'aucune mutation dominante induisant un phénotype fleur double n'a jamais été identifiée chez *A. thaliana*, le phénotype fleur double est sous le contrôle d'une unique mutation dominante chez *Petunia hybrida* (Krol and Chua, 1993) et la Fabaceae *Cercis canadensis* (Yokoya et al., 2000) notamment. Un séquençage des régions génomiques et des ADNc des orthologues de *RcAP2L* chez ces espèces pourrait apporter de nouvelles informations quant à leur état chez ces espèces.

III. **Importance fondamentale : conservation des mécanismes de contrôle de la fonction C chez les plantes à fleurs ?**

1. *Deux voies de restriction de la fonction C ont été sélectionnées en fonction des espèces*

Deux grandes voies de signalisation responsables du contrôle du domaine d'expression d'AG ou de ses orthologues ont été décrites chez les plantes à fleurs. La première, déjà discutée longuement ici et dans François et al. (2018), dépend de la balance entre l'expression de membres de la famille AP2 et de leur inhibiteur miR172. La seconde, caractérisée chez *P. hybrida* et *Antirrhinum majus*, dépend aussi de la balance entre l'expression d'un gène et d'un microARN capable d'inhiber son expression : c'est la boucle NF-YA – miR169 (Cartolano et al., 2007). En plus de ne pas appartenir à la même famille de gènes, la différence majeure entre ces deux voies est que NF-YA est un régulateur positif de l'expression d'AG et que c'est l'inhibition de celui-ci dans les verticilles 1 et 2 par miR169 qui bloque l'expression des orthologues d'AG et permet donc la formation d'un périante (Cartolano et al., 2007). Chez ces deux dernières espèces, la famille des euAP2 ne joue qu'un rôle anecdotique dans la restriction du patron d'expression des gènes de classe C et dans le déterminisme des sépales et des pétales alors que chez *A. thaliana*, la voie NF-YA n'est pas impliquée dans le déterminisme de l'organisation florale. Ainsi, il semblerait que l'établissement d'un périante dépende majoritairement de la restriction des gènes de classe C et non forcément de l'expression de gènes de classe A comme

le proposait le modèle ABC originel. On parle alors de modèle (A)BC, où ce serait l'absence de l'expression des gènes de classe C qui permettrait la formation des sépales et pétales.

2. Histoire évolutive de la famille euAP2

Sur le plan fondamental, l'existence d'un membre de la sous-famille TOE capable d'inhiber la fonction C suggère que, contrairement à ce qui était identifié chez *A. thaliana*, cette fonction n'est pas spécifique de la sous-famille AP2. Ainsi, cela rejoint les observations réalisées chez *P. hybrida* où les gènes *PhBEN* et *PhBOB* sont eux aussi capables de restreindre la fonction C. *A. thaliana* et *R. chinensis* appartiennent aux Rosidées, alors que *P. hybrida* appartient aux Astéridées, soulignant que la présence d'un TOE capable d'inhiber la fonction C n'est pas limitée à une famille restreinte d'espèces. Alors que l'étude de l'arbre phylogénétique présenté dans François et al. (2018) a montré que *RcAP2L*, *PhBEN* et *PhBOB* étaient potentiellement des orthologues dérivant d'un ancêtre commun, aucun orthologue n'a pu être identifié chez *A. thaliana*, ou *Capsella rubella*, alors qu'il existe un orthologue potentiel chez *Medicago truncatula* (une autre Rosidée). Ainsi, d'un point de vue parcimonieux, il semble envisageable que l'ancêtre commun aux Rosidées et Astéridées possédait un membre de la sous-famille TOE capable d'inhiber la fonction C et que celui-ci ait été perdu durant l'évolution qui a mené à ces deux Brassicaceae. Afin de confirmer cette hypothèse, il serait intéressant d'analyser les membres de la famille euAP2 dans un plus vaste panel d'espèces afin de vérifier s'il est possible de trouver un orthologue à *RcAP2L* et *PhBEN/PhBOB* chez une majorité de ces espèces et d'augmenter le bootstrap associé à cette observation, qui était seulement de 50 avec ce jeu de données.

D'autre part, contrairement à ce qui est observé chez *A. thaliana* où la restriction de la fonction C est assurée uniquement par les membres de la sous-famille AP2 des euAP2, chez *R. chinensis* et *P. hybrida* à la fois *RcAP2L* et *PhBOB/PhBEN*, bien qu'appartenant à la sous-famille TOE, semblent capables d'inhiber la fonction C. Pourtant, un ou plusieurs vrai(s) orthologue(s) d'AP2 est conservé chez ces espèces, ce qui amène à s'interroger sur leur fonction. Chez *P. hybrida*, leur mutation induit la formation de sépales chimériques et colorés, phénotype similaire à celui observé chez *Nicotiana benthamiana* lors de la surexpression de miR172, et donc de l'inhibition des membres de la famille euAP2. La transformation, même partielle, des sépales en organes pétaloïdes, suggère une expansion du domaine d'expression des gènes de classe B au premier verticille. Ainsi, la perte de fonction des membres de la sous-famille AP2 n'induit pas un phénotype de type perte de fonction A similaire à celui d'*A. thaliana* (Morel et al., 2017), mais il semblerait que, tout comme chez *N. benthamiana*, les orthologues d'AP2 aient plutôt conservé un rôle dans la restriction des gènes de la fonction B. Cependant, la surexpression de *PhAP2a* permettrait tout de même de compléter la mutation *ap2* chez *A. thaliana* (Maes et al., 2001), suggérant que *PhAP2a* pourrait avoir conservé

toutes les fonctions présentes chez *A. thaliana* et que, potentiellement ce soient les séquences régulatrices de ses cibles qui aient évolué chez *P. hybrida*.

Chez *Rosa*, peu d'études ont été réalisées sur la famille euAP2 et donc sur la sous-famille AP2. En particulier, les études fonctionnelles par transformation génétique étant longues et laborieuses, aucune donnée n'est encore disponible. Cependant, des inhibitions transitoires de l'expression de *RcAP2* par *Virus Induced Gene Silencing* (VIGS) ont été réalisées (Han et al., 2018) et semblent indiquer que *RcAP2* pourrait effectivement induire une conversion homéotique des pétales en étamines, mais des analyses plus détaillées semblent nécessaires pour confirmer cette observation. Ainsi, excepté cette observation et celle de l'inhibition de l'expression de *RcAG* par VIGS (Ma et al., 2018), les seules analyses fonctionnelles des gènes du développement floral chez le rosier ont été réalisées en système hétérologue, et notamment chez *A. thaliana* (Kitahara and Matsumoto, 2000 ; Kitahara et al., 2001). Placée sous un promoteur 35S, la surexpression de *RcAP2* semble perturber légèrement le développement floral, avec la production de quelques chimères pétales/étamines. À la vue des données de cet article, et conjointement avec les observations réalisées ici lors de la surexpression *RcA2PL* sous un promoteur 35S chez *A. thaliana*, il semblerait que le promoteur 35S ne soit pas l'outil idéal pour une étude en système hétérologue, soulignant la nécessité d'utiliser des promoteurs plus spécifiques du méristème et moins sensibles au silencing lors des transformations futures.

Pour conclure d'une manière globale, l'étude fonctionnelle d'espèces non modèles, ou bien en voie de s'établir en tant qu'espèces modèles, apporte une dimension évolutive au travail sur les familles de gènes et permet de mieux caractériser les grandes voies de signalisation, y compris au cours des processus de domestication.

IV. Quels moteurs de l'évolution phénotypique ?

À partir des gènes étudiés dans ce projet de thèse, il est possible d'entrevoir et souligner plusieurs mécanismes à l'origine de l'apparition de nouvelles voies de régulation, menant elles-mêmes à de nouveaux phénotypes. Ces sources de variation sont particulièrement intéressantes d'un point de vue évolutif étant donné qu'elles sont soumises à la sélection, naturelle ou humaine, et peuvent amener à l'apparition de nouvelles espèces ou variétés.

1. Importance des duplications de séquences dans la diversification des fonctions au sein d'une famille de gènes

Les duplications de gènes ou de génomes représentent une force majeure de l'évolution en levant une partie de la pression de sélection sur la fonction initiale du gène ancestral et en facilitant le partage des fonctions entre les deux gènes fils (subfonctionnalisation) ou l'apparition de nouvelles fonctions (néofonctionnalisation) (Lynch and Conery, 2000 ; Ohno, 1970 ; Otto and Yong, 2002). L'apparition d'un gène *ab initio* à partir d'une séquence intergénique ne contenant aucune information

est extrêmement rare car, afin d'obtenir un gène fonctionnel, une séquence codante, sans codon stop prématuré et codant pour une séquence protéique capable de se structurer et d'accomplir une fonction donnée, ainsi que des séquences régulatrices permettant l'expression du gène sont nécessaires (Kaessmann, 2010). À l'opposé, lorsque les protéines possèdent différents domaines se partageant différentes sous-fonctions modulables, le réarrangement de ces domaines après duplication, ou leur combinaison avec des domaines provenant d'autres protéines, sont plus fréquents et peuvent être à l'origine d'une variation de fonction, et éventuellement d'une variation du phénotype. Dans le cas d'une duplication de gènes ou de génome, la fonction ancestrale peut être maintenue par une des copies, tandis que l'autre peut évoluer librement, sans conséquence délétère sur l'organisme. Cela peut être illustré par l'histoire évolutive du premier ancêtre MADS des gènes de classe C, tel que développée par Causier et al. (2005). En effet, après la duplication de cet ancêtre, la fonction C a tout d'abord été retenue par des paralogues différents chez *Arabidopsis thaliana* (*AGAMOUS*) et chez *Antirrhinum majus* (*AmPLENA*). D'autre part, le deuxième paralogue a acquis des fonctions différentes chez ces deux espèces. En effet, *SHATTERPROOF 1* et *2*, les orthologues d'*AmPLENA*, sont nécessaires au développement des ovules et du fruit d'*A. thaliana* (Favaro et al., 2003) tandis que l'orthologue d'*AG*, *AmFARINELLI*, est responsable de la formation appropriée des étamines (Davies et al., 1999).

De manière intéressante, même lorsque la séquence du gène dupliqué commence à dégénérer (pseudogénéisation), une nouvelle fonction peut apparaître. Cette notion est illustrée par la structure de *RcUGO* et de ses orthologues chez *A. thaliana*, qui, bien que dérivant d'une duplication de gène suivie d'une perte de la majorité des domaines fonctionnels du gène ancestral, conserve un rôle essentiel dans le contrôle de l'homéostasie du méristème (Wenkel et al., 2007). Il en a résulté la production d'un dominant négatif ayant la fonction inverse de l'HD-ZIP III ancestral puisqu'il inhibe ses paralogues possédant encore l'ensemble des domaines fonctionnels. Similairement, il est proposé que les microARN dérivent au départ d'une duplication de gène inversée en tandem, évoluant tout d'abord en petit ARN de type siARN avant parfois d'être réduits jusqu'à former des microARN par perte progressive de la majorité de la séquence (Allen et al., 2004 ; Cuperus et al., 2011). Ainsi, les deux gènes identifiés dans cette étude soulignent l'implication des duplications de gènes et génomes dans l'acquisition de réseaux de régulation post-transcriptionnelle dans le développement des plantes, soit par des microARN soit par des microprotéines, aussi nommées *Small Regulatory Proteins* en anglais (SRP).

2. Les éléments transposables comme source de variabilité phénotypique importante

Les éléments transposables, grâce à leur forte mobilité et leur capacité à s'insérer aléatoirement dans le génome, sont eux aussi une source prolifique de variants et de diversité phénotypique permettant notamment la sélection humaine (Kidwell and Lisch, 1997 ; Lisch, 2013). Chez le melon (*Cucumis melo*) par exemple, la production de fleurs unisexuées femelles résulte de l'insertion d'un élément transposable à proximité du gène *CmWIP1* induisant une hyperméthylation de son promoteur et ainsi l'inhibition de l'expression de ce facteur capable de promouvoir la formation des étamines et d'inhiber celle des carpelles (Martin et al., 2009). Dans le cas de l'induction de la formation de fleurs doubles, pas moins de trois insertions indépendantes dans des orthologues du gène *AG* ont été identifiées comme étant à l'origine de la formation de fleurs doubles chez *Ipomoea nil* (Nitasaka, 2003), *Thalictrum thalictroides* (Galimba et al., 2012) et *Gentiana scabra* (Tasaki et al., 2017). Dans chacun de ces cas, ces insertions rendent les orthologues d'AGAMOUS non fonctionnels alors que dans le cas de *RcAP2L*⁴¹⁷² on observe au contraire un gain de fonction lié à une expansion de son domaine d'activité. De manière similaire, il a été montré que l'insertion de l'élément transposable Hopscotch très en amont du gène *TEOSINTE BRANCHED1 (TB1)* induit son expression ectopique dans les bourgeons axillaires du maïs. Il inhibe alors l'activité des méristèmes axillaires et est à l'origine de la réduction du nombre de branches latérales par plante permettant le passage d'un port arbustif de type téosinte à un port à forte dominance apicale de type maïs (Studer et al., 2011), étape clé de la domestication du maïs. L'ensemble de ces exemples montre à quel point les éléments transposables peuvent être une source importante de variabilité, fournissant une diversité essentielle avec certaines mutations potentiellement sélectionnées dans la nature ou lors de la domestication et à l'origine de nouvelles espèces ou variétés.

Pour conclure, l'ensemble de ces études et hypothèses n'aurait jamais été possible sans des séquences génomiques de haute qualité, produites par l'équipe dans le cas du rosier (Raymond et al., 2018). La disponibilité de ces séquences, pour des espèces non-modèles à l'origine, permet de mieux comprendre l'évolution à différentes échelles, de l'histoire évolutive des familles de gènes jusqu'à une meilleure compréhension des mécanismes de création et de sélection variétales, comme ce fut le cas par exemple lors du décryptage de l'origine hybride du rosier moderne *R. hybrida* cv 'La France'. Dorénavant, seul un petit pas sépare encore ces séquences de la caractérisation des mécanismes moléculaires à l'origine de chacun des traits d'intérêt cultural chez les espèces dont le génome est séquencé, ouvrant un immense éventail de possibilités pour les chercheurs à venir.

Chapitre V : Matériel et Méthodes

I. Culture et transformation des plantes

1. Rosiers

Les rosiers de cette étude (Tableau 3) ont été cultivés au Jardin botanique de Lyon ou dans des chambres de culture dédiées aux rosiers de l'École Normale Supérieure de Lyon, avec une photopériode de 16h de jour et 8h de nuit et des températures de 23°C le jour et de 18°C la nuit, et une hygrométrie de 60%.

Dans le cas des transformations génétiques, après culture *in vitro*, les jeunes rosiers (Tableau 4) ont d'abord été transférés sur terreau et acclimatés dans un incubateur réservé à cet usage dans les mêmes conditions climatiques qu'en chambre de culture (ci-dessus), mais avec une hygrométrie contrôlée à 85% d'humidité, pendant environ 40 jours, jusqu'à ce que les plantes soient suffisamment résistantes pour être cultivées en chambre de culture.

2. *Arabidopsis thaliana*

Les plantes ont été cultivées en chambre de culture avec une intensité lumineuse de 70 $\mu\text{moles}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ de photons et une photopériode de 16h/8h (jour/nuit) et des températures de 25°C/18°C (jour/nuit). Les *Arabidopsis* mutantes ont été obtenues par transformation médiée par *Agrobacterium tumefaciens* (Tableau 4).

Après récolte manuelle, les graines ont été séchées à 37°C pendant 3 à 7 jours, incubées une nuit à -20°C afin de tuer les éventuels œufs de pucerons puis stérilisées au dichlore gazeux pendant environ 4h (mélange de 4mL d'acide chlorhydrique 1 M avec 100 mL d'hypochlorite de sodium à 13% de chlore actif, dans une boîte hermétique de 5 L ; Lindsey et al., 2017). Elles ont ensuite été stratifiées durant 2 jours à l'obscurité et à 4°C pour homogénéiser la levée, puis semées et repiquées une semaine après.

3. Transformation d'*Agrobacterium tumefaciens*

Dans cette étude, deux souches d'*A. tumefaciens* électrocompétentes ont été utilisées : les EHA105 pour les rosiers et les C58PMP90 pour *A. thaliana*. Elles ont été transformées selon la même méthode : 50 μL de bactéries décongelées sur glace ont été mis en contact avec 100 à 200 ng d'ADN plasmidique avant d'être transférées dans des cuves d'électroporation Eurogentec préalablement refroidies, et électroporées (2 kV, 335 Ω , 15 μF , 5 s) avec un appareil Micropulser™ (BioRad). Après addition rapide de 400 μL de milieu LB, les bactéries ont été incubées pendant 3h à 27°C et étalées sur boîte de Petri contenant les antibiotiques adaptés à la sélection plasmidique et à celle de la souche d'*A. tumefaciens* (rifampicine 50 $\mu\text{g}\cdot\text{L}^{-1}$ et gentamicine 20 $\mu\text{g}\cdot\text{L}^{-1}$, ci-après Rif₅₀+Genta₂₀, pour les C58PMP90, Rif₁₀ pour les EHA105).

Tableau 3 : Ensemble des variétés de rosiers étudiées.

Variété	Phénotype floral	Culture	Ploidie supposée	Attribution
<i>R. chinensis</i> cv 'Old Blush'	Fleur double	Chambres de culture ENS Lyon	Diploïde	Chinenses
<i>R. x hybrida</i> cv 'Louise d'Arzens'	Fleur double	Roseraie de la Tête d'Or	N.D.	Noisette
<i>R. x hybrida</i> cv 'Hérodiade'	Fleur double	Roseraie de la Tête d'Or	N.D.	Noisette
<i>R. x hybrida</i> cv 'Blush Noisette'	Fleur double	Roseraie de la Tête d'Or	N.D.	Noisette
<i>R. x hybrida</i> cv 'Ducher'	Fleur double	Roseraie de la Tête d'Or	N.D.	Thé
<i>R. x hybrida</i> cv 'Lady Hilington'	Fleur double	Roseraie de la Tête d'Or	N.D.	Thé
<i>R. x hybrida</i> cv 'Louise Odier'	Fleur double	Roseraie de la Tête d'Or	N.D.	Bourbon
<i>R. x hybrida</i> cv 'Souvenir de la Malmaison'	Fleur double	Roseraie de la Tête d'Or	Triploïde	Bourbon
<i>R. x hybrida</i> cv 'General Schablikine'	Fleur double	Roseraie de la Tête d'Or	N.D.	Chinenses
<i>R. x hybrida</i> cv 'Comtesse de Cayla'	Fleur double	Roseraie de la Tête d'Or	N.D.	Chinenses
<i>R. x hybrida</i> cv 'Cramoisi Supérieur'	Fleur double	Roseraie de la Tête d'Or	N.D.	Chinenses
<i>R. x hybrida</i> cv 'Bengale d'Automne'	Fleur double	Roseraie de la Tête d'Or	N.D.	Chinenses
<i>R. x hybrida</i> cv 'Bébé Fleur'	Fleur double	Roseraie de la Tête d'Or	N.D.	Chinenses
<i>R. chinensis</i> cv 'Hume's Blush'	Fleur double	Chambres de culture ENS Lyon	Diploïde	Chinenses
<i>R. x hybrida</i> cv 'Rouge Meillard'	Fleur double	Chambres de culture ENS Lyon	N.D.	Hybride moderne
<i>R. x hybrida</i> cv 'La France'	Fleur double	Roseraie de la Tête d'Or	Triploïde	Hybride moderne
<i>R. damascena</i> cv 'Kalanzik'	Fleur double	Roseraie de la Tête d'Or	Tetraploïde	Gallicanae
<i>R. gallica</i> cv 'Officinalis'	Fleur double	Roseraie de la Tête d'Or	Tetraploïde	Gallicanae
<i>R. x 'Portland'</i>	Fleur double	Roseraie de la Tête d'Or	N.D.	Portland
<i>R. moschata</i>	Fleur simple	Roseraie de la Tête d'Or	Diploïde	Synstylae
<i>R. wichurana</i>	Fleur simple	Chambres de culture ENS Lyon	Diploïde	Synstylae
<i>R. chinensis</i> cv 'Mutabilis'	Fleur simple	Chambres de culture ENS Lyon	Diploïde	Chinenses
<i>R. chinensis</i> cv 'Sanguinea'	Fleur simple	Chambres de culture ENS Lyon	Diploïde	Chinenses
<i>R. chinensis</i> cv 'Spontanea'	Fleur simple	Chambres de culture ENS Lyon	Diploïde	Chinenses
<i>R. gigantea</i>	Fleur simple	Roseraie de la Tête d'Or	Diploïde	Chinenses
<i>R. chinensis</i> cv 'Single Pink'	Fleur simple	Chambres de culture ENS Lyon	Diploïde	Chinenses
<i>R. chinensis</i> cv 'T49'	Fleur simple	Chambres de culture ENS Lyon	Diploïde	Chinenses
<i>R. pimpinellifolia</i> cv 'Pimprenelle'	Fleur simple en laboratoire	Chambres de culture ENS Lyon	Tétraploïde	Hybride moderne

N.D. : non déterminé.

4. Transformation génétique des rosiers et sélection des transformants

Les cals embryogènes provenant des variétés *R. chinensis* 'Old Blush' (Vergne et al., 2009) et *Rosa pimpinellifolia* cv. 'Pimprenelle' (Vergne et al., données non publiées) ont été transformés génétiquement selon le protocole mis en place dans l'équipe d'accueil et décrit dans Vergne et al. (2010), avec les modifications suivantes :

(1) Les *A. tumefaciens* EHA105 ont d'abord été pré-cultivées 24h à 27°C en milieu AB liquide avant d'être culottées, resuspendues et incubées 24h à température ambiante dans 50 mL de milieu inducteur des gènes de virulence VIR-IND (Tableau 6). Une fois resuspendues dans du milieu MSØ (MS basal salts, pH 5.8), elles ont été mises en co-culture avec des embryons préalablement blessés par agitation avec du sable de Fontainebleau pendant deux fois 1min 30sec.

(2) Le milieu T (Tableau 6) a été utilisé pour l'étape de co-culture entre les embryons et les *A. tumefaciens* ainsi que lors de la semaine suivant l'arrêt de la transformation, avec addition de céfotaxine à 500 µg·mL⁻¹ final et carbénicilline à 50 µg·mL⁻¹ final pour cette étape.

L'ensemble des étapes de la transformation génétique à proprement parler, ainsi que de la régénération des plantes est résumé dans la Figure 37. La sélection des cals transformés génétiquement se fait lors des repiquages successifs sur les milieux EMM, SIM et Multiplication contenant de la kanamycine à 60 µg·L⁻¹ (EMM) puis 30 µg·L⁻¹ (SIM et Multiplication). Cependant, la surexpression de la GFP permet aussi de vérifier l'efficacité de la transformation génétique, notamment en cas de mauvaise santé des embryons.

5. Transformation génétique d'*A. thaliana* et sélection des lignées transformées

Les transformations génétiques d'*A. thaliana* ont été réalisées selon la méthode du « floral dipping », décrite par Clough and Bent (1998) et qui consiste à tremper de jeunes inflorescences dans une solution d'*A. tumefaciens* contenant 5% de saccharose, 10 mM de MgCl₂ et 0,025% de Silwet L-77 (Agridine). Après 24h d'incubation dans un sac en plastique, les plantes transformées génétiquement ont été remises en chambre de culture et leurs graines collectées après séchage sur pied.

Afin de sélectionner les graines génétiquement transformées, celles-ci ont été cultivées sur milieu MS (Murashige and Skoog, 1962) additionné de kanamycine à 50 µg·mL⁻¹ (kana₅₀) pendant une semaine en jours continus à 20°C. Seules les plantes résistantes ont été conservées et auto-pollinisées afin d'identifier les lignées T1 portant une insertion unique. Pour cela, elles ont été semées sur milieu MS + kana₅₀ pour analyse de la ségrégation : les lignées ségrégant selon les lois mendéliennes avec 3/4 des descendants résistants et 1/4 de descendants sensibles étant ensuite considérées comme portant une seule insertion du transgène dans leur génome. Une autopollinisation de ces T1 a ensuite permis d'obtenir des T2 homozygotes pour l'insertion, ce qui a été vérifié par analyse des descendants T3 qui

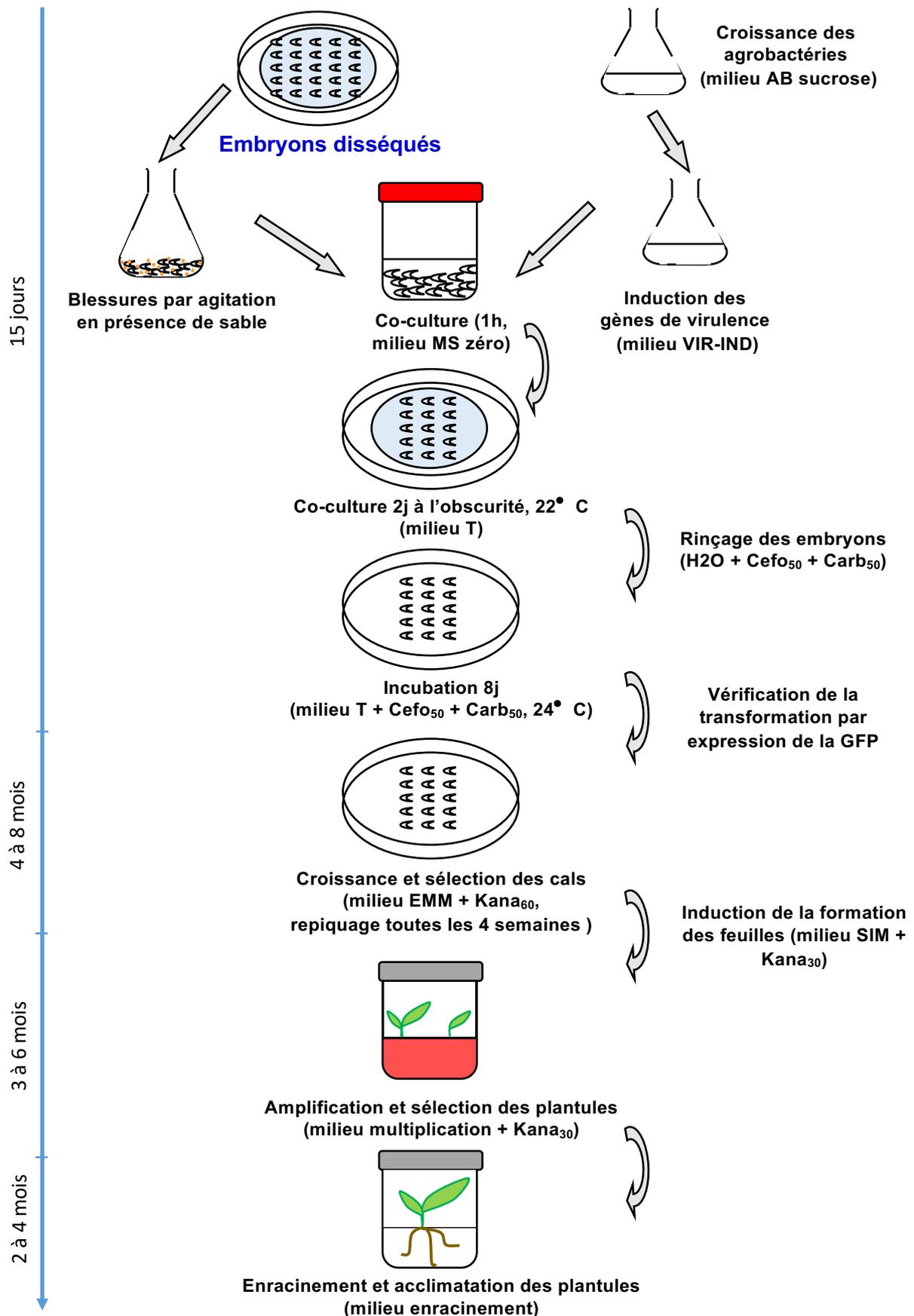


Figure 37 : Vue d'ensemble du protocole de transformation chez le rosier.

Les différentes étapes de la transformation *R. chinensis* cv 'Old Blush' est décrite dans la section Matériel et Méthodes - Cultures de plantes -Transformation des rosiers et sélection des transformants.

devaient être résistants à 100 %. L'ensemble des expériences présentées dans cette thèse a été effectué sur des T3 homozygotes ne possédant qu'une seule insertion (Tableau 4), sauf mention contraire.

6. Transformation génétique des fraisiers

Les transformations génétiques de *Fragaria vesca* ont été réalisées en collaboration avec Béatrice Denoyes et Aline Potier de l'équipe « Organogenèse du fruit et endoreduplication » (UMR 1332 - INRA de Bordeaux), principalement selon le protocole de Oosumi et al. (2006). La différence majeure comparée au protocole de transformation génétique réalisé sur 'Old Blush' correspond au fait que de jeunes feuilles sont utilisées comme matériel de départ pour le contact avec les *A. tumefaciens*, ce qui permet une régénération plus rapide des explants.

II. Biologie moléculaire

1. Extraction d'ADN

Les extractions d'ADN ont été réalisées avec le kit DNeasy (Qiagen) sur de jeunes feuilles ou bourgeons axillaires collectés et broyés dans du polyvinylpyrrolidone (PVP40) et un tampon d'homogénéisation (Tableau 6).

2. Extraction des ARNm et production d'ADNc

Les méristèmes floraux à différents stades de développement ont été identifiés et disséqués sous une loupe binoculaire. Les stades 1, 2, 3 et 4 correspondent respectivement aux moments où les sépales, pétales, étamines et carpelles sont initiés alors que le stade 5 correspond au moment où le centre du méristème commence à se creuser pour donner sa forme au réceptacle (Dubois et al., 2010). Tout de suite après avoir été disséqués, les méristèmes ont été plongés dans l'azote liquide et rassemblés en fonction de leur stade de développement. Ce matériel a ensuite été broyé au pilon dans du tampon de lyse additionné de 1% de β -mercaptoéthanol et de PVP40, dans l'azote liquide, et les ARN totaux purifiés à l'aide du kit NucleoSpin RNA Plant (Macherey-Nagel, Düren, Allemagne), selon les recommandations du fabricant. Un à 5 microgrammes d'acides nucléiques ont ensuite été traités avec la DNase TURBO DNA-free™ AM 1907 (Ambion, Austin Texas, États-Unis). Finalement, la qualité de l'extraction a été vérifiée par migration de 1 μ L de produit sur gel d'agarose à 1% dans du TAE 1x.

Entre 0,5 à 1 μ g d'ARN totaux résultant de ce traitement ont par la suite été utilisés pour la transcription inverse avec l'enzyme RevertAid Reverse Transcriptase (ThermoFisher Scientific) et des amorces oligo(dT) (T11VN, Tableau 5), selon le protocole fourni avec l'enzyme.

Tableau 4 : Ensemble des lignées transformées étudiées.

Lignées	Espèce transformée	Fond génétique	Transgène	Sélection	Description de la lignée
<i>p35S::RcAP2L^{WT}</i>	<i>Arabidopsis thaliana</i>	Col0	<i>p35S::RcAP2L^{WT}</i>	Resistance kanamycine + GFP	Surexpression de la version sauvage de la séquence codante de <i>RcAP2L</i>
<i>p35S::RcAP2L^{Δ172}</i>	<i>Arabidopsis thaliana</i>	Col0	<i>p35S::RcAP2L^{Δ172}</i>	Resistance kanamycine + GFP	Surexpression de la séquence codante de <i>RcAP2L^{Δ172}</i> de 'Old Blush'
<i>p35S::RcAP2L^{spΔ172}</i>	<i>Arabidopsis thaliana</i>	Col0	<i>p35S::RcAP2L^{spΔ172}</i>	Resistance kanamycine + GFP	Surexpression de la séquence codante de <i>RcAP2L^{Δ172}</i> de 'Single Pink'
<i>CRISPR-RcAP2L^{Δ172}</i>	<i>Rosa chinensis</i>	'Old Blush' (fleur double)	CAS9_sg2	Resistance kanamycine + GFP	Perte de fonction de <i>RcAP2L^{Δ172}</i>
<i>AmiARN1_RcAP2L^{Δ172}</i>	<i>Rosa chinensis</i>	'Old Blush' (fleur double)	<i>p35S::AmiARN1_RcAP2L^{Δ172}</i>	Resistance kanamycine + GFP	Surexpression d'un microARN artificiel ciblant <i>RcAP2L^{Δ172}</i>
<i>AmiARN2_RcAP2L^{Δ172}</i>	<i>Rosa chinensis</i>	'Old Blush' (fleur double)	<i>p35S::AmiARN2_RcAP2L^{Δ172}</i>	Resistance kanamycine + GFP	Surexpression d'un microARN artificiel ciblant <i>RcAP2L^{Δ172}</i> et <i>RcAP2L^{WT}</i>
<i>p35S::RcAP2L^{WT}</i>	<i>Rosa pimpinellifolia</i>	'Pimprenelle' (fleur simple)	<i>p35S::RcAP2L^{WT}</i>	Resistance kanamycine + GFP	Surexpression de la version sauvage de la séquence codante de <i>RcAP2L</i>
<i>p35S::RcAP2L^{Δ172}</i>	<i>Rosa pimpinellifolia</i>	'Pimprenelle' (fleur simple)	<i>p35S::RcAP2L^{Δ172}</i>	Resistance kanamycine + GFP	Surexpression de la séquence codante de <i>RcAP2L^{Δ172}</i> de 'Old Blush'
<i>p35S::RcAP2L^{spΔ172}</i>	<i>Rosa pimpinellifolia</i>	'Pimprenelle' (fleur simple)	<i>p35S::RcAP2L^{spΔ172}</i>	Resistance kanamycine + GFP	Surexpression de la séquence codante de <i>RcAP2L^{Δ172}</i> de 'Single Pink'
<i>AmiARN1_RcUGO</i>	<i>Rosa chinensis</i>	'Old Blush' (fleur double)	<i>p35S::AmiARN1_RcUGO</i>	Resistance kanamycine + GFP	Surexpression d'un microARN artificiel ciblant <i>RcUGO</i>
<i>AmiARN2_RcUGO</i>	<i>Rosa chinensis</i>	'Old Blush' (fleur double)	<i>p35S::AmiARN2_RcUGO</i>	Resistance kanamycine + GFP	Surexpression d'un microARN artificiel ciblant <i>RcUGO</i>

3. Clonage des séquences codantes à surexprimer

Afin de surexprimer les différents allèles d'intérêt, les ADNc correspondants ont été amplifiés à partir d'ARN extraits de *R. chinensis* 'Old Blush' avec la polymérase haute-fidélité Phusion (ThermoFisher) et en suivant les recommandations du fabricant. Les amorces utilisées (Tableau 5) correspondent au début et à la fin de la séquence codante (CDS). Après avoir confirmé la taille attendue par électrophorèse, les produits PCR ont été purifiés avec le kit Nucleospin Gene and PCR Clean-up (Macherey-Nagel). L'ajout, en 5' des amorces sens, d'une séquence CACC a permis le clonage directionnel de ces séquences codantes dans le plasmide d'entrée pENTR-D-TOPO (carte présentée sur l'annexe 1) puis leur transformation dans des bactéries DH5 alpha thermocompétentes. La ligation des inserts a ensuite été vérifiée par PCR sur colonies (GoTaq, Promega, amorces M13_Forward et M13_Reverse) puis séquençage Sanger (Société GATC, <https://www.eurofinsgenomics.eu/>). Ces inserts ont été transférés dans le plasmide de destination PK7GW2D (Annexe 1) par recombinaison Gateway (kit LR Clonase™ II Plus Enzyme Mix, Invitrogen) selon les recommandations du fabricant suivie d'une transformation dans *E. coli*. Ce plasmide de destination est adapté à la transformation chez le rosier car il contient un gène de résistance à la kanamycine et exprime la GFP sous un promoteur ubiquitaire, deux éléments essentiels à la sélection des cals transformés.

4. Clonage des séquences codantes pour production des protéines dans *E. coli*

Les inserts des plasmides d'entrée contenant les CDS d'intérêt ont été transférés dans le plasmide de destination pHM-GWA (Annexe 1) par recombinaison Gateway (kit LR Clonase™ II Plus Enzyme Mix, Invitrogen) selon les recommandations du fabricant. Ces vecteurs, conférant une résistance à l'ampicilline, ont permis de fusionner les séquences d'intérêt à un tag Histidine suivi d'un tag MBP, pour « Maltose Binding Protein » qui permet d'augmenter la solubilité et la stabilité des protéines.

5. Choix et clonage des microARN artificiels pour diminuer l'expression d'un gène

Afin de diminuer l'expression endogène des gènes d'intérêt chez le rosier, une stratégie utilisant des microARN artificiels a été adoptée (Figure 38). Tout d'abord, pour dessiner les séquences de 21 nucléotides ciblant spécifiquement les gènes candidats, le programme WMD3 (<http://wmd3.weigelworld.org/cgi-bin/webapp.cgi> ; Ossowski Stephan, Fitz Joffrey, Schwab Rebecca, Riester Markus and Weigel Detlef, communication personnelle) a été installé sur une machine virtuelle et adapté pour fonctionner sur le génome du rosier. Après avoir choisi deux séquences spécifiques par gène, le programme a de nouveau été utilisé pour déterminer la séquence complémentaire correspondante et en incluant un mésappariement en position centrale. Ces deux séquences ont

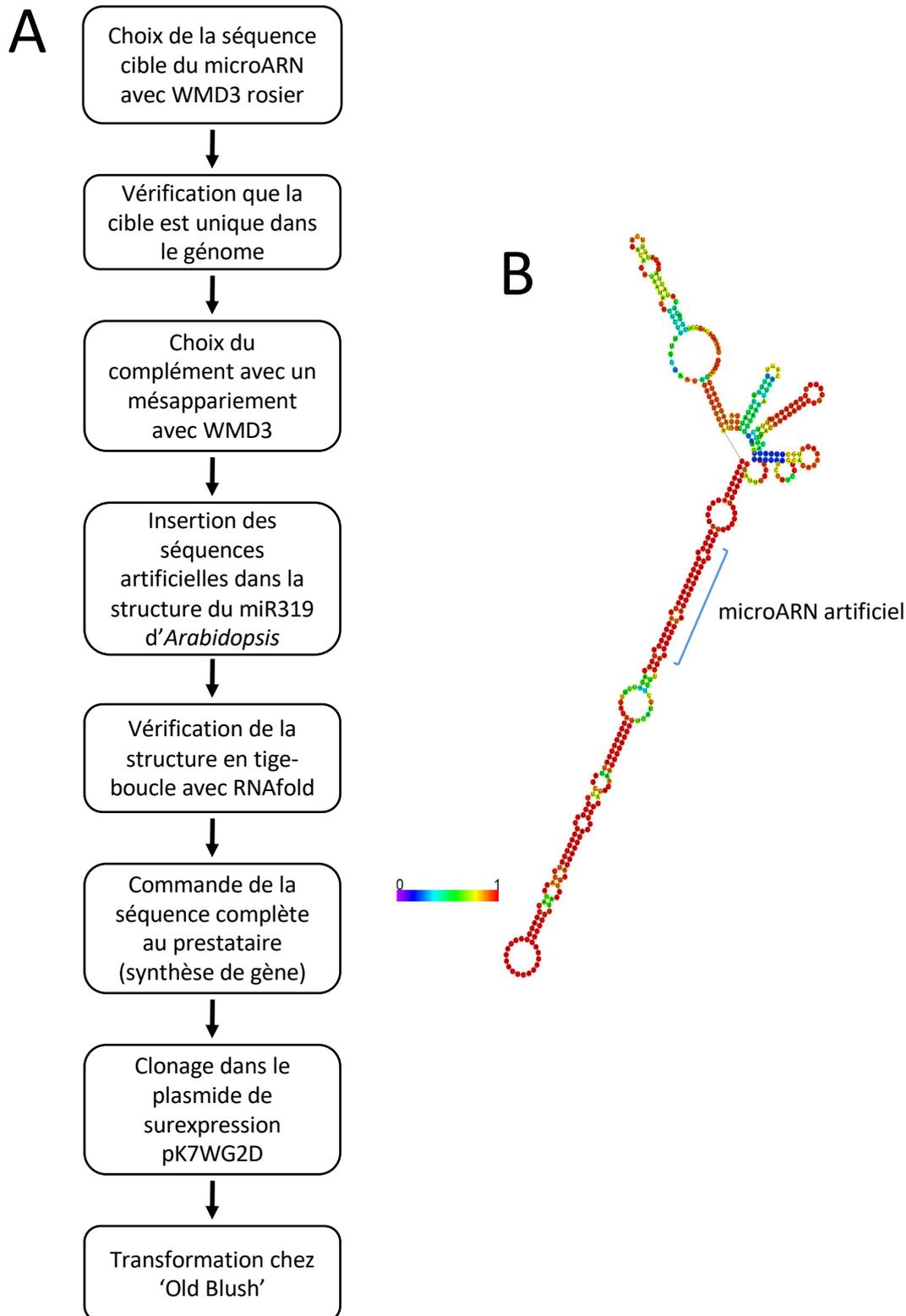


Figure 38 : Construction d'un microARN artificiel.

(A) Vue d'ensemble des différentes étapes nécessaires à l'expression d'un microARN artificiel chez le rosier. WMD3 est un programme disponible sur <http://wmd3.weigelworld.org/> et manuellement adaptable à tous les génomes. RNAfold est un programme en ligne (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) permettant de prédire la structure secondaire des ARN.

(B) Structure secondaire du miR319 d'*Arabidopsis*, modifié pour contenir le microARN d'intérêt.

ensuite été insérées dans la séquence du microARN 319 *d'A. thaliana* à la place des séquences endogènes (Schwab et al., 2006). La complémentarité des séquences et la bonne conformation du microARN ont ensuite été vérifiées avec le programme en ligne RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>).

La séquence résultante a ensuite été synthétisée par la société IDT (Integrated DNA Technologies, <https://eu.idtdna.com/>) avec l'addition en 5' d'une séquence CACC pour faciliter le clonage directionnel dans le pENTR-D-TOPO (Annexe 1). Une recombinaison Gateway (kit LR Clonase™ II Plus Enzyme Mix, Invitrogen) dans le plasmide PK7WG2D (Annexe 1) permettant la surexpression de ces microARN sous le promoteur 35S a finalement été réalisée.

6. Choix et clonage des ARN guides pour éditer un gène par CRISPR/Cas9

Les séquences des ARN guides pour éditer les gènes d'intérêt par CRISPR/Cas9 ont été choisies à partir de l'analyse automatisée des séquences du génome réalisée par Jérémy Just (ingénieur de recherche au sein de mon équipe) par des scripts Perl utilisant Jellyfish v2.2.0 (Marçais and Kingsford, 2011). L'ensemble des guides ayant une cible unique dans le génome ont été identifiés selon les règles suivantes : (1) séquences de 23 nucléotides se terminant en GG, (2) sans N lié à des indécisions de séquençage du génome, (3) sans répétition consécutive de 4 T ou plus, (4) dont les 15 dernières bases apparaissent au plus 10 fois dans le génome, (5) pour lesquels il n'y a pas de variant des 15 dernières bases se terminant par AG. À partir du fichier contenant ces guides potentiels, ceux situés dans les exons les plus en 5' du gène ont été favorisés.

Le protospacer, qui correspond à la séquence du guide sans le NGG en 3', a été commandé en tant qu'amorce à la société IDT (<https://eu.idtdna.com/>) avec l'ajout en 5' de celle-ci de la séquence ATTG. De la même manière, son brin complémentaire a été commandé avec en 5' la séquence AAAC. Ces deux amorces, chacune concentrée à 2 μ M, ont ensuite été chauffées à 95°C pendant 10 min puis hybridées à température ambiante. Les deux séquences ajoutées en 5' des amorces ont permis de créer des bouts collants facilitant le clonage du fragment dans le plasmide d'entrée pEn-Chimera (Annexe 1). La ligation a été réalisée avec la T4 DNA ligase (New England BioLabs) selon les recommandations du fabricant. L'insertion du fragment d'ADN a ensuite été vérifiée par PCR sur colonies et séquençage Sanger (<https://www.eurofinngenomics.eu/>). Pour finir, l'insert a été transféré par recombinaison Gateway (kit LR Clonase™ II Plus Enzyme Mix, Invitrogen) dans un plasmide de destination pDE-CAS9 (Fauser et al., 2014 ; Annexe 1) modifié qui contient lui aussi un gène de résistance à la kanamycine et un gène codant pour la GFP, ainsi que les gènes codant pour l'activité Cas9.

Tableau 5 : Amorces utilisées. (partie 1)

Nom de l'amorce	Séquence	Usage
B1Fseq	CAAGTTTGTACAAAAAGCAGG	Séquencage après LR simple
B2Rseq	CCACTTTGTACAAGAAAGC	
AP2L-TE_Jt1_F	GAAGCTGCAAGGTCTTTACGTTAAT	Amplification jonction <i>RcAP2L</i> /TE en 5'
AP2L-TE_Jt1_R	GAGACTCTCCACTTTGTTTCCTTG	
AP2L_TE_Jt2_F	GTGGTGAGAAGTAGTTCCTAGCTAG	Amplification jonction <i>RcAP2L</i> /TE en 3'
AP2L_TE_Jt2_R	AGACCAAATAGTAGATCCTCACACG	
AP2L_JtWT_F	ATTGCTATGTCATGTCTTCACAACC	Amplification jonction sauvage de <i>RcAP2L</i>
AP2L_JtWT_R	TATGGATGGAAAGAAGTTGGAGTT	
AP2L_all_START_F	CACCATGCTAGATCTCAACCTCAA	Clonage directionnel des CDS de <i>RcAP2L^{WT}</i> et <i>RcAP2L^{Δ172}</i>
APL2_D_STOP_R	TTATTTGCTTCGGAAAACCC	
AP2L_WT_STOP_R	TCAGTCCTGCAATAGAAAT	
AP2D_sg2_F	ATTGTGATGATCGGCGTCGTCGTA	Guide CRISPR contre <i>RcAP2L^{Δ172}</i>
AP2D_sg2_R	AAACTACGACGACGCCGATCATCA	
Fra_AP2L_vers_SP_1	AATGGGAAGCAGGTGCACCTGGGTGGGTTGA	Mutagenèse dirigée pour introduire la mutation de type 'Single Pink' chez le fraisier
Fra_AP2L_vers_SP_2	TCAAACCCACCAAGTGCACCTGCTCCCAT	
Fra_AP2L_start_all_F	CACCATGCTAGATCTCAACCTCAA	Génération d'une version FvAP2L ^{Δ172} similaire à celle d'Old Blush'
Fra_AP2LD_R	TTATTTGCTTCGGAAAACCTTGATCTCATGTGAATAGGC	
AG_reg_Fragment1_F	TAACCGTGGCCGCCTCTATGAG	Amplification du promoteur de <i>RcAG</i> par fragment de 500 pb pour expérience de retard sur gel
AG_reg_Fragment1_R	GCCTGCATACATTCCAAAACCCAC	
AG_reg_Fragment2_F	GGTTTGGTAGAAGCATAGATCTGGTG	
AG_reg_Fragment2_R	CTCCTGTGTTGAACCTTCCAGGTG	
AG_reg_Fragment3_F	GGGTTTTCTTGAACCTGTGTTAG	
AG_reg_Fragment3_R	GTTGAGAGCTGTGAGCTCATTG	
AG_reg_Fragment4_F	GCTGCCTATCTTCTTCTCCTCAC	
AG_reg_Fragment4_R	CCCTAAACCTTATCCTATTACCCACC	
AG_reg_Fragment5_F	GATGTTGTCATCAGAGTTTCAGACAG	
AG_reg_Fragment5_R	CACATATATCTCAATGTGCCAAGGAAG	
AG_reg_Fragment6_F	GAAGCTGTAATCTGGTTGTGCAC	
AG_reg_Fragment6_R	GAAGTTTTCCATGGTGAAGTAAAGGG	
AG_reg_Fragment7_F	GAATGAAGGTCTACTCTAGTTTACCTC	
AG_reg_Fragment7_R	CAAAGCAAAGGGTTGGACCTATC	
AG_reg_7bis_F	AGAAGCTTGATATATAGTATTAGGG	
AG_reg_7bis_R	TGAGATAATAGAAAAGACAGAACCC	
AG_reg_8_F	CGAAGTACTGCTATTCCTCTTTTC	
AG_reg_8_R	TATTGCTAGTTCACACTACCCAG	

7. Vérification de la présence de l'élément transposable par PCR

Les fragments d'ADN génomique spécifiques de l'allèle sauvage *RcAP2L^{WT}* ou des bordures de la jonction entre l'allèle tronqué *RcAP2L^{Δ172}* et l'élément transposable (Tableau 5) ont été amplifiés par PCR avec l'ADN polymérase GoTaq (Promega), conformément aux recommandations du fabricant. Une étape de dénaturation initiale a été réalisée à 95°C pendant 5 min. Afin d'augmenter la spécificité de la réaction, quinze cycles de PCR en gradient ont ensuite été effectués avec les étapes suivantes : 95°C pendant 30 secondes, 65°C (avec une diminution de 1°C par cycle) pendant 30 secondes et 72°C pendant 1 minute et 30 secondes. Cela a été suivi de 30 cycles standard avec le cycle suivant : 95°C pendant 30 secondes, 50°C pendant 30 secondes et 72°C pendant 1 minute et 30 secondes. Une étape finale d'élongation de 10 min à 72°C a permis de terminer la réaction.

III. Analyses d'expression des gènes

1. RT-PCR quantitatives (qRT-PCR)

Deux couples d'amorces différents par gène d'intérêt ont été conçus avec le programme Primer3 v2.4.0 (Untergasser et al., 2012) utilisé en ligne de commande avec les paramètres suivants : (1) taille de l'amplicon entre 80 et 250 paires de base, (2) pourcentage de GC compris entre 40 et 60%, (3) différence de température d'hybridation entre les 2 amorces ne pouvant pas excéder 3°C, (4) amorces sens et antisens de part et d'autre d'un intron quand cela était possible. La spécificité de l'amplification a été contrôlée sur les prédictions d'ARNm du génome de rosier et avec le logiciel de prédiction de PCR *in silico* ePCR v2.3.12-1 (Rotmistrovsky et al., 2004) et les paramètres suivants : « -t4 -d40-450 -w4 -f3 -n2 -g1 -p+ », où -t permet de choisir le format de sortie des résultats ; -p+ active le filtrage des hits redondants ; -d indique la taille de l'amplicon, -w identifie la longueur des mots exactement identiques recherchés par l'heuristique, -f le nombre de mots exacts non contigus et -n et -g correspondent respectivement au nombre de mésappariements et de gaps tolérés.

Cette vérification a ensuite été réalisée par PCR réelle sur les ADNc obtenus à partir des jeunes bourgeons floraux et migration sur gel d'agarose 1%. Enfin, l'efficacité de l'amplification de chaque couple d'amorces a été mesurée par qRT-PCR à partir des produits d'amplification par PCR dilués en série.

Les PCR quantitatives ont été effectuées avec le kit FastStart DNA Master SYBR Green (Roche) et l'appareil Step-OnePlus Real-Time PCR (Applied Biosystems, USA). Des triplicats techniques ont été réalisés. Les niveaux d'expression ont été exprimés en unités arbitraires et calculés avec la méthode de quantification relative suivante $eff_A^{-CtA} / (eff_{TCTP}^{-CtTCTP} \times eff_{GAPDH}^{-CtGAPDH} \times eff_{TUB}^{-CtTUB})^{-1/3}$ où le Ct, pour 'Cycle Threshold', correspond au nombre de cycles de PCR nécessaires pour atteindre un seuil de fluorescence établi par l'appareil et eff correspond à l'efficacité d'amplification spécifique à chaque couple d'amorces conçu précédemment. Les niveaux d'expression ont été normalisés avec les gènes

Tableau 5 : Amorces utilisées. (partie 2)

Nom de l'amorce	Séquence	Usage
RhAG HIS for	AGGGAACTTGACTTGCACAACAATAACC	Hybridation <i>in situ</i> de <i>RcAG</i>
RhAG HIS rev T7	TGTAATACGACTCACTATAGGGCGTAATGCCAGTACTAACTGTGGGAGAGGTT	
littleZIP_ISH1_Fwd	CCCTCACTTCTTCACCAACACA	Hybridation <i>in situ</i> de <i>RcUGO</i>
T7_littleZIP1_R	TAATACGACTCACTATAGGGACACCACTGCTACTTGTCTCT	
RhTCTP-F2	GATGCTGATGAGGGTGTGTA	qPCR et mesure d'expression de <i>RcTCTP</i>
RhTCTP-R2	CTTGTTGCTCCCTCAATGT	
GAPDH_F	GGTCAAGGTCATTGCTTGGT	qPCR et mesure d'expression de <i>RcGAPDH</i>
GAPDH_R	GGATCGATCACATCGACAGA	
TUB_F	ATTGAGCGTCCCACCTACAC	qPCR et mesure d'expression de <i>RcTUB</i>
TUB_R	AGCATGAAATGGATCCTTGG	
littleZIP_qPCR1_Fwd	GATTGCCTCTGCTCTCTCCG	qPCR et mesure d'expression de <i>RcUGO</i>
littleZIP_qPCR1_Rev	GCCTTCTCCTCAGCCTCTC	
5'_IZIP_F	CACCATGGAAAGGCTAAACTCGAAG	Clonage directionnel de la CDS de <i>RcUGO</i>
3'_IZIP_avecTGA_R	TCAGTTGCTGGAAGTGAAGG	
B1Fseq	CAAGTTTGTACAAAAAGCAGG	Séquençage après LR simple (clonage)
B2Rseq	CCACTTTGTACAAGAAAGC	

RcHt_S1070.3 (RcGAPDH), *RcHt_S1963.26 (RcTUB)*, *RcHt_S1460.7 (RcTCTP)* (Dubois et al., 2010 ; Szécsi et al., 2006), après vérification que celui-ci était exprimé de façon stable dans l'ensemble des échantillons.

2. Hybridations *in situ*

L'hybridation *in situ* a été réalisée principalement comme décrit dans Dubois et al. (2010). Les amorces pour amplifier les sondes spécifiques des ARNm d'intérêt ont été dessinées de sorte que l'amplicon mesure 300 à 700 paires de base. Leur spécificité a été vérifiée en utilisant le programme ePCR (Rotmistrovsky et al., 2004) (paramètres : -t4 -d40-450 -w4 -f3 -n2 -g1 -p+) sur les ARNm prédits à partir du génome du rosier publié par Raymond et al. (2018), et la spécificité de la sonde l'a été en utilisant un blast nucléotidique. L'ajout en 5' de l'amorce antisens portant la séquence correspondant au promoteur T7 (TAATACGACTCACTATAGG) permettra ensuite la transcription *in vitro* (Tableau 5).

Afin d'amplifier le fragment d'intérêt, des PCR ont été réalisées sur les ADNc d'Old Blush' puis vérifiées sur gel d'agarose 1% et purifiées avec le kit Macherey Nagel NucleoSpin Gel and PCR Clean Up. La transcription *in vitro* (IVT) d'ARN marqués à la digoxigénine a été réalisée sur la nuit à 37°C avec 1 µg de produit PCR, dans un volume total de 20 µL et selon le protocole du fabricant du kit DIG RNA Labelling (Roche). Après vérification de l'efficacité de la transcription par migration sur gel de 2 µL de produit d'IVT, les ARN ont ensuite été purifiés dans de l'éthanol à 70% additionné de 0,1 volume d'acétate de sodium à 3 mol·L⁻¹. Après précipitation pendant 30 min à -20°C, les ARN ont été centrifugés à 13 000 g pendant 30 min et à 4°C, lavés deux fois à l'éthanol 70%, séchés et dissouts dans 100 µL d'eau ultrapure. L'hydrolyse des sondes dans du tampon carbonate 1X a ensuite été réalisée à 60°C pendant une durée permettant d'obtenir des tailles de sondes d'environ 170 paires de base puis arrêtée avec 10 µL d'acide acétique à 10%. Après une nouvelle purification, cette fois en présence de 380 µg·mL⁻¹ d'ARNt, les sondes ont été resuspendues dans 40 µL d'eau ultrapure et conservées à -20°C.

De jeunes boutons floraux fraîchement collectés ont été immédiatement placés dans une solution de fixation contenant 50% d'éthanol, 5% d'acide acétique et 4% de paraformaldéhyde. Les échantillons ont été infiltrés sous vide deux fois pendant 1 heure avant d'être transférés dans une solution d'éthanol à 70% et incubés pendant 12 heures à 4°C. Deux lavages dans de l'éthanol à 90% ont été effectués avant stockage des échantillons dans du méthanol à -20°C pour être ultérieurement inclus dans de la paraffine. Pour cela, ils ont d'abord été déshydratés lors de lavages successifs à l'éthanol, à l'éthanol additionné d'éosine puis à l'Histo-Clear (Fisher Scientific) avant d'être inclus dans du Paraplast X-TRA (Sigma-Aldrich). Des coupes de 8 micromètres ont été réalisées avec le microtome Microm HM355S (Thermo Scientific).

Les échantillons ont ensuite été débarrassés de la paraffine dans de l'Histo-Clear et progressivement réhydratés avant d'être incubés 15 min dans de l'acide chlorhydrique 0,8%, 10 min

dans du citrate de sodium salin 2X pour la perméabilisation cellulaire, dans du PBS avec $10 \mu\text{g}\cdot\text{mL}^{-1}$ de protéinase K Fungal (Invitrogen) pour éliminer les protéines liées à l'ARN. Entre chaque traitement, des lavages à l'eau ultrapure ont été effectués. Après une étape de fixation de 10 minutes dans du PBS contenant 3,7% de formaldéhyde, les échantillons ont été à nouveau déshydratés et incubés sur la nuit à 50°C dans $80 \mu\text{L}$ de tampon d'hybridation (DIG Easy Hyb™, Roche) contenant $3 \mu\text{L}$ de sondes marquées. Le jour suivant, les sections ont été lavées 2 fois 1h30 dans un tampon avec 50% de formamide et du citrate de sodium salin 2X. Un traitement avec du TBS 1X additionné d'un agent de blocage (Blocking reagent, Roche) a ensuite été effectué pour réduire le bruit de fond, suivi d'un lavage de 30 minutes dans une solution de TBT (Tableau 6). Une incubation de 90 minutes avec un anticorps anti-DIG 1:1000 dans du TBT a ensuite été réalisée avant la révélation du signal en présence de NBT / BCIP pendant une nuit.

IV. Analyses des interactions protéine-promoteur

1. Prédictions structurales

Les alignements de séquences protéiques ont été réalisés avec les outils en ligne MULTALIN (<http://multalin.toulouse.inra.fr/multalin/>; Corpet, 1988; paramètres par défaut) et ESPript (<http://esript.ibcp.fr/ESPript/ESPript/>; Robert and Gouet, 2014). La modélisation de la structure des protéines a été réalisée avec l'outil en ligne PHYRE2 (Kelley et al., 2015) et en utilisant comme modèle la structure cristallographique de ERF96 identifiée dans (Chen et al., 2018a). La visualisation de la structure protéique 3D et l'alignement des structures 3D ont été réalisés avec le logiciel PyMOL v2.3 (<https://pymol.org/>).

2. Synthèse des protéines dans *E. coli*

Toutes les protéines recombinantes ont été exprimées dans la souche *E. coli* BL21 DE3. Les bactéries ont été cultivées dans du milieu LB liquide (Tableau 6) jusqu'à une densité optique à 600 nm entre 0,6 et 0,9. L'expression des protéines d'intérêt a alors été induite avec l'isopropyl- β -D-1-thyogalactopiranoside (IPTG) à une concentration finale de $400 \mu\text{M}$, pour une production de protéines à 18°C sur la nuit.

3. Purification des protéines

Les cultures de bactéries ont ensuite été centrifugées 30 min à $3000 g$ et les culots de bactéries résultants ont été resuspendus dans du tampon de lyse (Tableau 6), avant sonication. Une centrifugation à $13\ 000 g$ a permis de récupérer les fractions solubles qui ont ensuite été passées à travers des résines de Ni-sépharose, préalablement équilibrées avec le tampon de lyse (Tableau 6). Ensuite, les protéines marquées avec un tag MBP-His ont été éluées avec de l'imidazole 300 mM après une étape de lavage précédente avec de l'imidazole 30 mM .

Afin de vérifier l'efficacité de la purification, les différentes fractions résultant de l'expression et de la purification des protéines (fractions non induites, induites, totales, solubles, de lavage et d'éluion) ont été analysées par électrophorèse sur gel de polyacrylamide SDS-page afin de révéler la présence ou non de la protéine d'intérêt dans chaque fraction. Les protéines ont été dénaturées en les mélangeant avec un tampon de dénaturation (Tableau 6) durant 5 min à 95°C. Les protéines dénaturées ont été chargées sur des gels de polyacrylamide SDS-PAGE à 10% conjointement avec un marqueur de poids moléculaire (Precision Plus Standar Dual Color, Bio-Rad). La migration des protéines a été réalisée dans du tampon d'électrophorèse (Tableau 6) jusqu'à ce que le bleu de bromophénol quitte le gel. Les protéines ont été révélées dans les gels en les colorant avec une solution de bleu de Coomassie (Instant Blue Coomassie™, SigmaAldrich), puis en les décolorant dans de l'eau.

Enfin, la concentration en protéine a été quantifiée par mesure d'absorbance à 280 nm avec le Nanodrop 2000 (ThermoFisher). Les concentrations ont été calculées par la loi de Beer-Lambert, selon la formule $A = \epsilon \cdot l \cdot C$; où l est la distance optique et ϵ est le coefficient d'extinction molaire calculé pour chaque protéine à l'aide de l'outil ProtParam (ExPASy ; <https://web.expasy.org/protparam/>).

4. Synthèse des sondes d'ADN marquées

Afin de visualiser les ADN après interaction avec les protéines, des amorces marquées au fluorophore IRD dye 700 (Tableau 5) ont été commandées chez IDT (Integrated DNA Technologies, <https://eu.idtdna.com/>) et utilisées pour amplifier les fragments d'ADN génomique par PCR avec la polymérase PrimeStar, conformément aux recommandations du fabricant (Ambion). Les produits PCR ont ensuite été migrés sur gel d'agarose à 1% afin d'estimer la concentration en ADN et rapporter les échantillons les plus concentrés à des concentrations similaires à celles des moins concentrés.

5. Expériences de retards sur gel

Aussi appelées EMSA (pour « Electrophoretic Mobility Shift Assay »), elles ont été réalisées sur des gels d'agarose à 2% préparés avec du tampon d'électrophorèse TAE 0,5X (Tableau 6). Les gels ont été équilibrés dans du tampon TAE 0,5X à 90V pendant 90 minutes à 4°C. Les mélanges protéine-ADN contenant d'ADN compétitif de poisson à une concentration finale 45 $\mu\text{g}\cdot\text{mL}^{-1}$ et d'ADN cible marqué ont été incubés 1h à 4°C dans le tampon d'interaction (Tableau 6). Ils ont ensuite été déposés sur gel pour une migration d'une heure à 90 V et à 4°C. Les retards de migration ont ensuite été visualisés grâce au système d'imagerie Odyssey CLx (LI-COR).

Tableau 6 : Liste des milieux et tampons utilisés dans cette étude.

Nom du milieu	Composition	Utilisation	Référence
Milieu LB (Lysogeny Broth)	10 g·L ⁻¹ Tryptone (Sigma-Aldrich), 5 g·L ⁻¹ Yeast Extract (Sigma-Aldrich), 10 g·L ⁻¹ NaCl (Duchefa)	Culture <i>E. coli</i>	Bertani, 1959
Milieu AB liquide	Solution 'AB salt' 1X, 0,375% saccharose, 7,5 µg·mL ⁻¹ rifampicine et 75 µg·mL ⁻¹ spectinomycine	Culture <i>A. tumefaciens</i>	Gelvin et al., 2006
Milieu VIR-IND	Solution 'AB salt' 1X, 2 mM NAH ₂ PO ₄ , 50 mM MES, 0,5% glucose (v/v) et 0,1 mM d'acétosyringone préparée extemporanément	Induction <i>A. tumefaciens</i>	Gelvin et al., 2006
Milieu T	4,3 g·L ⁻¹ de MS basal salt and vitamins (Duchefa), 30 g·L ⁻¹ saccharose, 1,34 µM NAA, 3,42 µM zéatine, 3,33 µM BAP et 0,58 µM GA ₃ , pH 5,8 puis 7,2 g·L ⁻¹ agar (Merck)	Transformation rosiers	Communication personnelle de Philippe Vergne
EMM (Embryogenic callus Maintenance Medium)	4,3 g·L ⁻¹ MS basal salt and vitamins (Duchefa), 2,4 g·L ⁻¹ Gelrite™ ou 6,5 g·L ⁻¹ VitroAgar (Kalys) et 60 µg·mL ⁻¹ kanamycine	Transformation rosiers	Dohm et al., 2001
SIM (Shoot Inducing Medium)	4,3 g·L ⁻¹ MS basal salt and vitamins (Duchefa), 30 g·L ⁻¹ saccharose, 0,05 µM AIB, 8,9 µM BAP, pH 5,8 puis 7,2 g·L ⁻¹ Agar (Merck) et 30 µg·mL ⁻¹ kanamycine	Transformation rosiers	Vergne et al., 2010
Milieu de multiplication	4,3 g·L ⁻¹ MS salts and vitamins (Duchefa), 0,1 g·L ⁻¹ Fe-EDDHA, 30 g·L ⁻¹ saccharose, 0,02 µM NAA, 0,29 µM GA ₃ , 2,22 µM BAP, pH 5,8 puis 6,5 g·L ⁻¹ VitroAgar (Kalys) et 30 µg·mL ⁻¹ kanamycine	Transformation rosiers	Vergne et al., 2010
Tampon d'homogénéisation	Tris HCl pH 8 15 mM, EDTA 2 mM, NaCl 20 mM, KCl 20 mM, β-mercaptoéthanol 0,1%, Triton 0,5%	Extraction ADN	François et al., 2018
TBS (Tris Buffer Saline)	Tris HCl pH 7,5 100 mM, NaCl 400 mM	Hybridation <i>in situ</i>	Javelle et al., 2011
TBT (TBS BSA Triton)	TBS 1X, Triton à 0,3%, BSA à 1%	Hybridation <i>in situ</i>	Javelle et al., 2011
Tampon de dénaturation	Tris-HCl 50 mM, pH 6,8; glycérol à 10%; SDS à 1%; bleu de bromophénol 0,0025%; dithiothréitol à 0,4%	Migration des protéines	Stigliani et al., 2018
Tampon d'électrophorèse	Tris-HCl 25 mM; glycine 192 mM; SDS 0,1% p/v	Migration des protéines	Stigliani et al., 2018
Tampon de lyse	Tris-HCl pH8 20 mM; NaCl 500 mM; TCEP 1mM; antiprotéase (cOmplete™ ULTRA Tablets, Roche)	Extraction des protéines	Stigliani et al., 2018
Tampon d'interaction	Tris-HCl pH 7,5 400mM; KCl 200 mM; KCl 100 mM; glycérol 10%; Triton 0,1%	Retard sur gel	Stigliani et al., 2018

V. Analyses phénotypiques des lignées

1. Comptage du nombre d'organes floraux

Les différents organes floraux qui ont été comptés sont les sépales, les pétales, les étamines (organes mâles) et les carpelles (organes femelles) correspondants à des organes insérés à l'intérieur du réceptacle et dont le style et le stigmate dépassent à l'extérieur de ce dernier (Figure 1-B). Il arrive parfois que certains organes aient une identité intermédiaire entre deux organes différents, ils ont alors été appelés chimères. Tous les types de chimères ont pu être observés mais seules les chimères entre pétales et étamines ont été vus de manière récurrente et en quantité significative pour cette étude.

2. Suivi de la date de floraison d'*A. thaliana*

Afin d'analyser la date de floraison des plants d'*A. thaliana*, deux paramètres généralement utilisés ont été pris en compte : la date d'ouverture de la première fleur et le nombre de feuilles principales mises en place avant que le méristème végétatif ne se transforme en méristème floral. Ce dernier est utilisé comme indicateur de la date de floraison depuis les années 90 (Devlin et al., 1996).

3. Microscopie Électronique à Balayage

Les échantillons fraîchement collectés ont été imagés avec l'appareil HS-300 (Hirox) à une température de -20°C et un voltage de 15 kV.

VI. Analyses bioinformatiques

Ce travail a été effectué en partie grâce aux ressources du PSMN (Pôle Scientifique de Modélisation Numérique) et du CBP (Centre Blaise Pascal) de l'ENS de Lyon et avec l'aide de Jérémie Just (IR dans l'équipe) et Marion Verdnaud (IE dans l'équipe).

1. Génomes et séquences utilisés

Les analyses présentées dans ce manuscrit ont fait appel à l'utilisation de nombreux résultats de séquençage obtenus à partir de matériel biologique varié et en utilisant des méthodes de séquençage variées. Le contenu est résumé ci-dessous.

- (1) Le génome du rosier diploïde modèle *R. chinensis* cv 'Old Blush' hétérozygote est disponible dans l'équipe depuis 2014 sous la forme de 15 937 scaffolds, pour une longueur totale de 882,7 megabases (pour une taille de génome haploïde estimée à 560 Mb). Ces séquences ont été obtenues en séquençant, par la méthode Illumina, l'ADN génomique de *R. chinensis* cv 'Old Blush'. Il est appelé dans ce manuscrit le « génome hétérozygote » et est dorénavant disponible en ligne sous le numéro d'accès BioProject PRJEB24406 (Raymond et al., 2018). À mon arrivée, il n'était pas possible de plus assembler entre eux

ces scaffolds à cause de la forte hétérozygotie du génome et la présence de nombreuses séquences répétées. Au cours des différentes analyses effectuées depuis l'obtention de cet assemblage, y compris celles de mon travail de thèse, nous avons estimé qu'environ les 2/3 du génome ont été assemblés sous forme de deux haplotypes séparés, et le tiers restant sous forme d'un consensus entre les deux haplotypes.

- (2) En 2016, un deuxième type de séquençage a été réalisé sur un haploïde doublé dérivant d'Old Blush' (Vergne et al., 2018). Ce matériel haploïde a été produit dans l'équipe à partir de microspores extraites d'anthères d'Old Blush' et régénéré sous forme de calcs embryogènes. Il est possible de produire des jeunes tiges à partir de ces calcs, en culture *in vitro*, mais pas de régénérer des plantes entières. L'ADN de ce matériel a été utilisé pour séquencer un haplotype du génome d'Old Blush', avec la méthode de séquençage de reads longs PacBio. Ceux-ci ont permis l'assemblage de cet haplotype en 7 pseudo-molécules correspondant aux 7 chromosomes présents chez le rosier. Ce génome, appelé « génome homozygote » d'Old Blush' dans cette étude, a été déposé après publication (Raymond et al., 2018) dans GenBank sous le numéro d'accèsion PDCK00000000.
- (3) Les données de reséquençage génomique de différentes variétés de rosiers (en particulier *R. chinensis* cv 'Single Pink', *R. odorata* cv 'Hume's Blush', *R. x hybrida* 'La France', *R. chinensis* cv 'Sanguinea', *R. chinensis* cv 'Spontanea' and *R. wichurana* ; Raymond et al., 2018), sous forme de paires de reads, ont été utilisées durant ma thèse pour étudier la présence du TE dans d'autres génotypes que celui d'Old Blush'. Elles sont quant à elles disponibles dans la « Sequence Read Archive », sous le numéro d'accèsion SRP119986.
- (4) Des données de RNA-seq comparant le niveau d'expression des gènes dans de jeunes feuilles (FTN), de jeunes racines (RAC), de bourgeons axillaires, aussi appelé méristèmes végétatifs, dormants (NDB), de bourgeons axillaires actifs (DBO), de bourgeons floraux lors de la transition vers un méristème floral (IFL), de méristèmes floraux lors du développement précoce des organes floraux (IMO), de boutons floraux fermés (BFL), d'étamines aux stades de microsporogénèse et microgamétogénèse (DET), de fleurs épanouies (OFT), de fleurs sénescents (SEN) et de cynorrhodons (CYN), comme décrits dans Dubois et al. (2012). Chacune de ces bibliothèques ne contient que 400 000 à 1,2 millions de reads de 32 pb, mais c'est encore actuellement la ressource transcriptomique la plus riche en termes de variété de stades et d'organes existant pour le rosier.
- (5) Des données de RNA-seq comparant les stades de bourgeons floraux correspondant à l'initiation des primordia des sépales (stade 1), des pétales (stade 2) et des étamines (stade 3) chez 'Old Blush' et son mutant spontané à fleurs simples 'Single Pink'. Le

séquençage a été fait en paired-ends Illumina 2x100 pb orientés, en triplicats et chaque librairie contient 30 à 40 millions de reads.

- (6) Des données de RNA-seq comparant des bourgeons floraux d'Old Blush' à ceux de 'T49', mutant obtenu fortuitement parmi la régénération d'explants d'Old Blush' passés en culture *in vitro*, en 2012. Le séquençage a été fait en single end 101 pb, en triplicats et chaque librairie contient 20 à 30 millions de reads.

Il est important de noter qu'à chacun des génomes hétérozygote et homozygote sont associées des annotations du génome, disponibles sous la forme de gènes prédits, ARN prédits (ARNm mais aussi ARNr et ARNc), protéines prédites et éléments transposables prédits (Raymond et al., 2018). Ils sont eux aussi disponibles en ligne et ont été essentiels à l'analyse de l'intervalle contenant la mutation entraînant la formation de la fleur double.

2. Analyses des données de transcriptomique

L'ensemble de ce travail, à l'exception de l'étape 3, a été réalisé avec Jérémie Just (IR dans l'équipe).

0/ Contrôle qualité et nettoyage des reads.

La validation de la qualité globale des séquences brutes de RNA-seq et l'identification des adaptateurs de séquençage ont été réalisés avec FastQC v0.11.3 à 0.11.7 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Les adaptateurs ont été détectés et supprimés des séquences avec Cutadapt v1.16 (Martin, 2011 ; paramètres : « --error-rate=0.1 --times=3 --overlap=5 --minimum-length=40 --zero-cap --max-n=2 ») et les régions ayant une qualité Phred (Ewing and Green, 1998) inférieure à 28 sur une fenêtre glissante de 4 pb ont été coupées au moyen de scripts maison utilisant BioPerl v1.7.2 (Stajich et al., 2002). Les reads de moins de 25 pb à l'issue du nettoyage, et les reads non pairés (dans le cas de séquençage paired-end) ont été supprimés.

1/ Outils et paramètres de mapping.

Les reads ont été mappés sur les différents génomes avec TopHat2 v2.1.1 (Kim et al., 2013) et Bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012) en utilisant les annotations de gènes déjà obtenues, avec les paramètres suivants : « --read-realign-edit-dist 0 --min-anchor-length 5 --min-intron-length 9 --max-intron-length 20000 --max-multihits 50 --coverage-search --microexon-search --min-segment-intron 8 --max-segment-intron 50000 -b2-very-sensitive -prefilter-multihits -G <GFF3 file> ». Les fichiers BAM obtenus ont été triés avec Picard v2.4.1 (<http://broadinstitute.github.io/picard/>), puis le comptage du nombre de reads par gène a été fait avec HTSeq-count v0.11.1 (Anders et al., 2015).

2/ Normalisation des librairies, analyse différentielle.

La normalisation et le calcul des FPKM (pour « Fragments Per Kilobase of transcript per Million mapped reads ») ont été faits avec Cufflinks v2.2.1 (Trapnell et al., 2012), avec les paramètres « --frag-

bias-correct --multi-read-correct --min-isoform-fraction 0.05 --max-multiread-fraction 1 --max-mle-iterations 15000 --min-intron-length 20 --max-intron-length 50000 --overhang-tolerance 8 ». Les analyses différentielles entre échantillons ont été faites à partir des comptages bruts, avec le package R DESeq2 (Love et al., 2014). Pour limiter les effets des tests multiples, un filtrage *a priori* a été fait, et seuls les gènes ayant au moins une moyenne d'un read par échantillon ont été considérés.

3/ Comparaison des données d'expression sur les bourgeons entre différents projets.

Afin de vérifier si les deux expériences de RNA-seq (Dubois et al., 2012, et 'Old Blush' vs 'Single Pink') étaient comparables, il était nécessaire d'étudier la corrélation entre les niveaux d'expression de gènes de ménage dans les échantillons communs, les bourgeons floraux au stade 3 du développement floral et dans les méristèmes floraux. Pour cette comparaison, tous les reads ont été remappés en single ends (pour les librairies paired-ends, seul le read forward a été utilisé). Ensuite, les orthologues, chez *R. chinensis*, des gènes de ménages caractérisés par Gu et al. (2014) chez *A. thaliana* ont été identifiés par BLAST (Altschul et al., 1990), soit *RcHt_S1070.3* (*RcGAPDH*), *RcHt_S1621.18* (*RcACT*), *RcHt_S1460.7* (*RcTCTP*), *RcHt_S313.10* et *RcHt_S1849.8* (*RcEF1 α*), *RcHt8199.1* (*RcH3*), *RcHt_S1963.26* (*RcTUB*), *RcHt_S1883.1* (*RcUBC*) et *RcHt_S1682.11* (*RcEIF5*). Bien qu'une tendance fût visible, le coefficient de Pearson n'était que de 0,6113, indiquant qu'il n'était pas possible de convertir les mesures du RNA-seq réalisé par Dubois et al. (2012) en mesures du même ordre de grandeur que celui réalisé sur les bourgeons floraux. Par conséquent, il était plus prudent d'utiliser des échelles séparées lors de l'analyse et la présentation des données des différents RNA-seq.

Les données d'expression obtenues ont été représentées sous une forme similaire à celle de l'eFP Browser d'*A. thaliana* (Winter et al., 2007). À chaque mesure d'expression en FPKM a été associée une couleur selon les échelles du package R viridis v0.5.1 (première moitié des échelles de couleur Viridis et Inferno). La représentation graphique des données d'expression de plusieurs gènes d'une même famille a été réalisée sous forme de « heatmap » en utilisant le programme du site web <http://www.heatmapper.ca/expression/>, avec les options suivantes : (1) sans normalisation indépendante pour chaque gène (« no scaling ») et (2) sans regroupement des gènes selon leur patron d'expression (« no clustering method »).

3. Analyses de l'expression allèle-spécifique

J'avais observé dans les résultats de Cufflinks, utilisé avec l'option « --multi-read-correct » sur le génome hétérozygote (paragraphe ci-dessus), des scaffolds entiers du chromosome 3 apparemment non exprimés. À partir de cette constatation, Jérémy Just (IR dans l'équipe) a analysé l'expression allèle-spécifique à l'échelle du génome.

1/ Identification des régions pour lesquelles on a les deux haplotypes

Les séquences des chromosomes des deux génomes homozygotes extraits d'Old Blush' (Raymond et al., 2018) ont été comparées avec LAST (Kielbasa et al., 2011), après entraînement de l'algorithme sur les deux chromosomes 1. Le pourcentage de similarité entre les deux chromosomes a été calculé à partir de ces alignements sur une fenêtre glissante de 10 kb et représenté graphiquement. Lorsque les deux assemblages représentent le même haplotype, ce pourcentage est proche de 100% ; lorsque deux haplotypes distincts ont été séquencés, il est très variable, compris entre 90 et 99%. Les positions des régions correspondant à ce dernier cas ont été relevées manuellement.

2/ Identification des groupes d'allèles par similarité.

Les transcrits prédits sur les deux assemblages ont été comparés avec BLASTN entre eux et avec les transcriptomes prédits de *Fragaria vesca*, *Rubus occidentalis*, *Malus x domestica*, *Prunus domestica*, *P. avium* et *P. persica* (paramètres: « M=1 N=-3 Q=3 R=3 W=7 E=1e-30 wordmask=dust B=200 V=200 hspsepSmax=1000 hspsepQmax=1000 links pingpong sump »). Pour chaque transcrit de rosier, seuls ses hits sur d'autres transcrits de rosier meilleurs que tous ses hits sur les autres espèces ont été considérés. Ces hits filtrés ont été utilisés pour trouver les meilleurs hits réciproques de chaque transcrit.

3/ Filtrage des paires d'allèles par colinéarité.

Les alignements des chromosomes complets obtenus avec LAST ont été utilisés pour filtrer les allèles en position synténique à partir des candidats obtenus ci-dessus. Autour de chaque gène, la synténie a été modélisée sous forme d'une relation affine entre coordonnées, à partir des alignements sur une région de 2 Mb. Si les coordonnées des deux allèles potentiels prédites par cette relation étaient distantes de moins de 400 kb, ils ont été considérés comme allèles réels (à la fois par similarité de séquence et position synténique). Ainsi, 23 199 paires potentielles d'allèles ont été obtenues sur l'ensemble du génome, dont 3 526 sont situées dans des régions pour lesquelles on a deux haplotypes distincts.

4/ Mapping allèle-spécifique des reads.

Pour déterminer pour chaque read s'il pouvait être attribué spécifiquement à un allèle, les données de RNA-seq ont été mappées sur l'union des deux génomes homozygotes (Raymond et al., 2018) et du génome hétérozygote (Raymond et al., 2018) avec HISAT2 v2.1.0 et les paramètres « --min-intronlen 20 --max-intronlen 50000 -k 30 --no-discordant --no-mixed --no-unal --no-softclip --non-deterministic » et en utilisant les sites d'épissages prédits dans l'annotation des génomes. Ce mapping non déterministe a été itéré 100 fois, et le résultat de l'ensemble des itérations a permis de classer les reads en « spécifiques d'un haplotype », « communs aux deux haplotypes » ou « non spécifiques » (pour les reads ayant plus de deux meilleurs matches possibles).

Le mapping des reads a été filtré en utilisant cette information, pour produire des fichiers BAM distincts pour chaque haplotype et chaque catégorie. Ces fichiers ont été traités avec Cufflinks

(Trapnell et al., 2012), comme décrit précédemment, pour estimer l'expression de chaque allèle en termes de FPKM.

4. Correction des annotations du génome de rosier

Ce travail a été réalisé avec l'aide de Jérémie Just (IR dans l'équipe). Les modèles de gènes ont été extraits de l'annotation des séquences des génomes homozygotes et hétérozygotes présentés ci-dessus, après identification par comparaison avec les séquences protéiques d'*A. thaliana* (TAIR 10, <http://www.arabidopsis.org/>; Reiser et al., 2017). Quand cela s'est avéré nécessaire, les séquences ont été comparées entre elles et les annotations manuellement modifiées. En particulier, (1) les prédictions de site d'épissage et les limites des régions non traduites (UTR) ont été ajustées manuellement en fonction des séquences d'ADNc clonées et alignées avec EST2GENOME (EMBOSS v6.6.0.0; Mott, 1997; Rice et al., 2000) ainsi que des données transcriptomiques alignées avec TopHat2 v2.1.1 (Kim et al., 2013) sans utiliser les annotations déjà obtenues, et visualisées avec Artemis v17.0.1 (Rutherford et al., 2000). (2) les domaines protéiques ont été identifiés à l'aide du logiciel InterProScan version 5.27.-66.057 (Quevillon et al., 2005), de la base de données Pfam version 31.050 (Finn et al., 2016) et/ou d'une annotation manuelle. (3) Les sites de liaison supposés à miR172 ont été prédits à l'aide d'une instance locale du logiciel WMD3 (Ossowski Stephan, Fitz Joffrey, Schwab Rebecca, Riester Markus et Weigel Detlef, communication personnelle). (4) Les annotations automatiques d'éléments transposables du génome d'Old Blush', réalisées avec REPET (Flutre et al., 2011) ont été confirmées en découpant les séquences génomiques d'intérêt en *k*-mers chevauchants de 47 pb et en comptant leur nombre d'occurrences dans l'ensemble des 375 Gb de reads génomiques utilisés pour assembler le génome hétérozygote, avec Jellyfish v2.2.3 (Marçais and Kingsford, 2011). Quand cela s'est avéré pertinent, les limites des LTR (pour Long Terminal Repeats) ont été identifiées avec précision à l'aide du programme DOTTER (Sonnhammer and Durbin, 1995) et les séquences des LTR ont été comparées entre elles à l'aide du programme d'alignement bl2seq (Tatusova and Madden, 1999). Les cadres de lectures ont été prédits dans la région interne des TE à l'aide du logiciel Fgenesh (Solovyev et al., 2006) et les domaines protéiques ont été quant à eux annotés par recherche de similarité au moyen de DANTE (<http://repeatexplorer.org/>; Polpitiya et al., 2008).

VII. Analyses phylogénétiques

1. Identification des séquences d'intérêt

Les séquences des génomes, ARNm et protéines prédits des espèces *Fragaria vesca*, *Petunia*, *Solanum lycopersicum*, *Arabidopsis thaliana*, *Capsella rubella*, *Medicago truncatula*, *Vitis vinifera* and *Prunus persica* ont été téléchargées depuis le site web Phytozome (Goodstein et al., 2012; <https://phytozome.jgi.doe.gov/pz/portal.html>) et utilisées pour les analyses phylogénétiques. Les

proches parents des protéines d'intérêt ont été identifiés par BLAST protéique à partir de protéines entières et/ou de domaines conservés en fonction des besoins.

2. Alignement des séquences

Les séquences ont été alignées en utilisant les programmes ClustalW v2.1 (Gouy et al., 2010) et BioEdit (Tajima and Nei, 1984) ou Seaview v4.6.4 (Gouy et al., 2010). Quand cela a été nécessaire, l'alignement a été manuellement corrigé, avec l'aide de Michiel Vandenbussche (Équipe « Évolution et développement de la fleur », laboratoire Reproduction et Développement des Plantes). En particulier, la position de certains motifs conservés au sein de la famille des euAP2 a été vérifiée. En cas d'absence de certains de ces motifs, la prédiction de la structure intron/exon a été contrôlée et corrigée si nécessaire.

3. Construction des arbres

Les arbres phylogénétiques ont été construits avec la méthode Neighbor-Joining (Saitou and Nei, 1987) et les logiciels Treecon (Van de Peer and De Wachter, 1994) et/ou Seaview v4.6.4 (Gouy et al., 2010). Les paramètres suivants y ont été utilisés : (1) Estimation de distance : Méthode = Tajima et Nei (Tajima and Nei, 1984) ; Calculs de distance = insertions et délétions non prises en compte ; Positions d'alignement : tous ; Analyse par bootstrap: oui avec 2000 répétitions. (2) Topologie d'arborescence : Méthode = Neighbor-Joining ; Analyse Bootstrap = oui. (3) Racines : outgroup = séquence unique et forcée.

BIBLIOGRAPHIE

- Akita, Y., Horikawa, Y., and Kanno, A. (2008). Comparative analysis of floral MADS-box genes between wild-type and a putative homeotic mutant in lily. *J. Hortic. Sci. Biotechnol.* *83*, 453–461.
- Akita, Y., Nakada, M., and Kanno, A. (2011). Effect of the expression level of an AGAMOUS-like gene on the petaloidy of stamens in the double-flowered lily, 'Elodie.' *Sci. Hortic.* *128*, 48–53.
- Aldwinckle, H. s, bullet, and Malnoy, M. (2009). Plant Regeneration and Transformation in the Rosaceae. *Transgenic Plant J* *3*.
- Allen, E., Xie, Z., Gustafson, A.M., Sung, G.-H., Spatafora, J.W., and Carrington, J.C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* *36*, 1282–1290.
- Allen, M.D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998). A novel mode of DNA recognition by a β -sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* *17*, 5484–5496.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403–410.
- Anastasiou, E., Kenz, S., Gerstung, M., MacLean, D., Timmer, J., Fleck, C., and Lenhard, M. (2007). Control of Plant Organ Size by KLUH/CYP78A5-Dependent Intercellular Signaling. *Dev. Cell* *13*, 843–856.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* *15*, 2730–2741.
- Barton, M.K., and Poethig, R.S. (1993). Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and in the shoot meristemless mutant. *Development* *119*, 823–831.
- Bateson, W. (1894). *Materials for the Study of Variation, Treated with Fatial Regard to Discontinuity in the Origin of Species.*
- Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C., Genete, M., Berrabah, W., Chèvre, A.-M., Delourme, R., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* *4*, 879.
- Bendahmane, M., Dubois, A., Raymond, O., and Bris, M.L. (2013). Genetics and genomics of flower initiation and development in roses. *J. Exp. Bot.* *64*, 847–857.
- Benedito, V.A., Angenent, G.C., van Tuyl, J.M., and Krens, F.A. (2004). *Lilium longiflorum* and molecular floral development: the ABCDE model. *Acta Hortic.* *83*–89.
- Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* *62*, 293–300.

- Bevan, M., Mayer, K., White, O., Eisen, J.A., Preuss, D., Bureau, T., Salzberg, S.L., and Mewes, H.W. (2001). Sequence and analysis of the Arabidopsis genome. *Curr. Opin. Plant Biol.* *4*, 105–110.
- Blanc, G., Hokamp, K., and Wolfe, K.H. (2003). A Recent Polyploidy Superimposed on Older Large-Scale Duplications in the Arabidopsis Genome. *Genome Res.* *13*, 137–144.
- Block, M.D., Herrera-Estrella, L., Montagu, M.V., Schell, J., and Zambryski, P. (1984). Expression of foreign genes in regenerated plants and in their progeny. *EMBO J.* *3*, 1681–1689.
- Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S., Blied, M., Boersma, M.R., Borghi, L., Bruggmann, R., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* *2*, 16074.
- Bomblies, K., Dagenais, N., and Weigel, D. (1999). Redundant enhancers mediate transcriptional repression of AGAMOUS by APETALA2. *Dev. Biol.* *216*, 260–264.
- Bourke, P.M., Arens, P., Voorrips, R.E., Esselink, G.D., Koning-Boucoiran, C.F.S., Van't Westende, W.P.C., Santos Leonardo, T., Wissink, P., Zheng, C., van Geest, G., et al. (2017). Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *Plant J. Cell Mol. Biol.* *90*, 330–343.
- Bowman, J.L., and Smyth, D.R. (1999). CRABS CLAW, a gene that regulates carpel and nectary development in Arabidopsis, encodes a novel protein with zinc finger and helix-loop-helix domains. *Dev. Camb. Engl.* *126*, 2387–2396.
- Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in Arabidopsis. *Plant Cell* *1*, 37–52.
- Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. *Development* *112*, 1–20.
- Bradley, D., Carpenter, R., Sommer, H., Hartley, N., and Coen, E. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of *Antirrhinum*. *Cell* *72*, 85–95.
- Brambilla, V., Battaglia, R., Colombo, M., Masiero, S., Bencivenga, S., Kater, M.M., and Colombo, L. (2007). Genetic and Molecular Interactions between BELL1 and MADS Box Factors Support Ovule Development in Arabidopsis. *Plant Cell* *19*, 2544–2556.
- Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M., and Simon, R. (2000). Dependence of Stem Cell Fate in Arabidopsis on a Feedback Loop Regulated by CLV3 Activity. *Science* *289*, 617–619.
- Brandt, R., Xie, Y., Musielak, T., Graeff, M., Stierhof, Y.-D., Huang, H., Liu, C.-M., and Wenkel, S. (2013). Control of stem cell homeostasis via interlocking microRNA and microProtein feedback loops. *Mech. Dev.* *130*, 25–33.
- Breuninger, H., and Lenhard, M. (2010). Chapter Seven - Control of Tissue and Organ Growth in Plants. In *Current Topics in Developmental Biology*, M.C.P. Timmermans, ed. (Academic Press), pp. 185–220.
- Brioudes, F., Joly, C., Szécsi, J., Varaud, E., Leroux, J., Bellvert, F., Bertrand, C., and Bendahmane, M. (2009). Jasmonate controls late development stages of petal growth in Arabidopsis thaliana. *Plant J. Cell Mol. Biol.* *60*, 1070–1080.

- Carpenter, R., and Coen, E.S. (1990). Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. *Genes Dev.* *4*, 1483–1493.
- Cartolano, M., Castillo, R., Efremova, N., Kuckenbergh, M., Zethof, J., Gerats, T., Schwarz-Sommer, Z., and Vandenbussche, M. (2007). A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. *Nat. Genet.* *39*, 901–905.
- Causier, B., Castillo, R., Zhou, J., Ingram, R., Xue, Y., Schwarz-Sommer, Z., and Davies, B. (2005). Evolution in Action: Following Function in Duplicated Floral Homeotic Genes. *Curr. Biol.* *15*, 1508–1512.
- Causier, B., Ashworth, M., Guo, W., and Davies, B. (2012). The TOPLESS Interactome: A Framework for Gene Repression in *Arabidopsis*1[W][OA]. *Plant Physiol.* *158*, 423–438.
- Chen, X. (2004). A MicroRNA as a Translational Repressor of APETALA2 in *Arabidopsis* Flower Development. *Science* *303*, 2022–2025.
- Chen, C.-Y., Lin, P.-H., Chen, K.-H., and Cheng, Y.-S. (2018a). Structural insights into DNA sequence recognition by *Arabidopsis* ETHYLENE RESPONSE FACTOR 96. *BioRxiv* 302299.
- Chen, X., Bhaduria, V., and Ma, B. (2018b). ChIP-Seq: A Powerful Tool for Studying Protein-DNA Interactions in Plants. *Curr. Issues Mol. Biol.* *27*, 171–180.
- Cheon, K.-S., Nakatsuka, A., Tasaki, K., and Kobayashi, N. (2017). Floral Morphology and MADS Gene Expression in Double-flowered Japanese Evergreen Azalea. *Hortic. J. advpub*.
- Clark, S.E., Running, M.P., and Meyerowitz, E.M. (1995). CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. *Development* *121*, 2057–2067.
- Coen, E. (2001). Goethe and the ABC model of flower development. *C. R. Acad. Sci. III* *324*, 523–530.
- Coen, E.S., and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* *353*, 31–37.
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* *16*, 10881–10890.
- Crespel, L., Chirollet, M., Durel, C., Zhang, D., Meynet, J., and Gudin, S. (2002). Mapping of qualitative and quantitative phenotypic traits in *Rosa* using AFLP markers. *Theor. Appl. Genet.* *105*, 1207–1214.
- Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011). Evolution and Functional Diversification of MIRNA Genes. *Plant Cell* *23*, 431–442.
- 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.
- Dadpour, M.R., Naghiloo, S., Peighambaroust, S.H., Panahirad, S., Aliakbari, M., and Movafeghi, A. (2011). Comparison of floral ontogeny in wild-type and double-flowered phenotypes of *Syringa vulgaris* L. (Oleaceae). *Sci. Hortic.* *127*, 535–541.

- Davies, B., Motte, P., Keck, E., Saedler, H., Sommer, H., and Schwarz-Sommer, Z. (1999). PLENA and FARINELLI: redundancy and regulatory interactions between two Antirrhinum MADS-box factors controlling flower development. *EMBO J.* *18*, 4023–4034.
- Daxinger, L., Hunter, B., Sheikh, M., Jauvion, V., Gascioli, V., Vaucheret, H., Matzke, M., and Furner, I. (2008). Unexpected silencing effects from T-DNA tags in Arabidopsis. *Trends Plant Sci.* *13*, 4–6.
- Debener, T. (1999). Genetic analysis of horticulturally important morphological and physiological characters in diploid roses. *Gartenbauwissenschaft Ger.*
- Debener, T., and Mattiesch, L. (1999). Construction of a genetic linkage map for roses using RAPD and AFLP markers. *Theor. Appl. Genet.* *99*, 891–899.
- Devlin, P.F., Halliday, K.J., Harberd, N.P., and Whitelam, G.C. (1996). The rosette habit of Arabidopsis thaliana is dependent upon phytochrome action: novel phytochromes control internode elongation and flowering time. *Plant J. Cell Mol. Biol.* *10*, 1127–1134.
- Deyholos, M.K., and Sieburth, L.E. (2000). Separable Whorl-Specific Expression and Negative Regulation by Enhancer Elements within the AGAMOUS Second Intron. *Plant Cell* *12*, 1799–1810.
- Dinh, T.T., Girke, T., Liu, X., Yant, L., Schmid, M., and Chen, X. (2012). The floral homeotic protein APETALA2 recognizes and acts through an AT-rich sequence element. *Development* *139*, 1978–1986.
- Dinneny, J.R., Yadegari, R., Fischer, R.L., Yanofsky, M.F., and Weigel, D. (2004). The role of JAGGED in shaping lateral organs. *Dev. Camb. Engl.* *131*, 1101–1110.
- Disch, S., Anastasiou, E., Sharma, V.K., Laux, T., Fletcher, J.C., and Lenhard, M. (2006). The E3 ubiquitin ligase BIG BROTHER controls arabidopsis organ size in a dosage-dependent manner. *Curr. Biol. CB* *16*, 272–279.
- Dong, Z., Alexander, M., and Chuck, G. (2019). Understanding Grass Domestication through Maize Mutants. *Trends Genet.* *35*, 118–128.
- Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991). Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell* *65*, 991–1002.
- Dubois, A., Raymond, O., Maene, M., Baudino, S., Langlade, N.B., Boltz, V., Vergne, P., and Bendahmane, M. (2010). Tinkering with the C-Function: A Molecular Frame for the Selection of Double Flowers in Cultivated Roses. *PLoS ONE* *5*.
- Dubois, A., Remay, A., Raymond, O., Balzergue, S., Chauvet, A., Maene, M., Pécrix, Y., Yang, S.-H., Jeauffre, J., Thouroude, T., et al. (2011). Genomic Approach to Study Floral Development Genes in Rosa sp. *PLoS ONE* *6*.
- Dubois, A., Carrere, S., Raymond, O., Pouvreau, B., Cottret, L., Rocchia, A., Onesto, J.-P., Sakr, S., Atanassova, R., Baudino, S., et al. (2012). Transcriptome database resource and gene expression atlas for the rose. *BMC Genomics* *13*, 638.
- Dugo, M.L., Satovic, Z., Millán, T., Cubero, J.I., Rubiales, D., Cabrera, A., and Torres, A.M. (2005). Genetic mapping of QTLs controlling horticultural traits in diploid roses. *Theor. Appl. Genet.* *111*, 511–520.

- Durfee, T., Roe, J.L., Sessions, R.A., Inouye, C., Serikawa, K., Feldmann, K.A., Weigel, D., and Zambryski, P.C. (2003). The F-box-containing protein UFO and AGAMOUS participate in antagonistic pathways governing early petal development in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 8571–8576.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* *32*, 1792–1797.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* *323*, 133–138.
- Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr. Biol. CB* *13*, 1768–1774.
- Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* *8*, 186–194.
- Fausser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *Plant J. Cell Mol. Biol.* *79*, 348–359.
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M., and Colombo, L. (2003). MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *Plant Cell* *15*, 2603–2611.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* *44*, D279–285.
- Fletcher, J.C., Brand, U., Running, M.P., Simon, R., and Meyerowitz, E.M. (1999). Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science* *283*, 1911–1914.
- Floyd, S.K., Ryan, J.G., Conway, S.J., Brenner, E., Burris, K.P., Burris, J.N., Chen, T., Edger, P.P., Graham, S.W., Leebens-Mack, J.H., et al. (2014). Origin of a novel regulatory module by duplication and degeneration of an ancient plant transcription factor. *Mol. Phylogenet. Evol.* *81*, 159–173.
- Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PLoS One* *6*, e16526.
- François, L., Verdenaud, M., Fu, X., Ruleman, D., Dubois, A., Vandenbussche, M., Bendahmane, A., Raymond, O., Just, J., and Bendahmane, M. (2018). A miR172 target-deficient AP2-like gene correlates with the double flower phenotype in roses. *Sci. Rep.* *8*, 12912.
- Fraser, L.G., Tsang, G.K., Datson, P.M., De Silva, H.N., Harvey, C.F., Gill, G.P., Crowhurst, R.N., and McNeilage, M.A. (2009). A gene-rich linkage map in the dioecious species *Actinidia chinensis* (kiwifruit) reveals putative X/Y sex-determining chromosomes. *BMC Genomics* *10*, 102.
- Galimba, K.D., Tolkin, T.R., Sullivan, A.M., Melzer, R., Theißen, G., and Di Stilio, V.S. (2012). Loss of deeply conserved C-class floral homeotic gene function and C- and E-class protein interaction in a double-flowered ranunculid mutant. *Proc. Natl. Acad. Sci. U. S. A.* *109*, E2267–E2275.
- Gattolin, S., Cirilli, M., Pacheco, I., Ciacciulli, A., Linge, C. da S., Mauroux, J.-B., Lambert, P., Cammarata, E., Bassi, D., Pascal, T., et al. Deletion of the miR172 target site in a TOE-type gene is a strong candidate variant for dominant double-flower trait in Rosaceae. *Plant J.* *0*.

- Gelvin, S.B. (2006). *Agrobacterium* virulence gene induction. *Methods Mol. Biol.* Clifton NJ 343, 77–84.
- Goethe, J.W. (1790). *The metamorphosis of plants.*
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186.
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224.
- Graf, P., Dolzblasz, A., Würschum, T., Lenhard, M., Pfreundt, U., and Laux, T. (2010). MGOUN1 Encodes an Arabidopsis Type IB DNA Topoisomerase Required in Stem Cell Regulation and to Maintain Developmentally Regulated Gene Silencing. *Plant Cell* 22, 716–728.
- Gramzow, L., Ritz, M.S., and Theissen, G. (2010). On the origin of MADS-domain transcription factors. *Trends Genet. TIG* 26, 149–153.
- Green, K.A., Prigge, M.J., Katzman, R.B., and Clark, S.E. (2005). CORONA, a member of the class III homeodomain leucine zipper gene family in Arabidopsis, regulates stem cell specification and organogenesis. *Plant Cell* 17, 691–704.
- Gu, C.-S., Liu, L., Xu, C., Zhao, Y., Zhu, X., and Huang, S.-Z. (2014). Reference Gene Selection for Quantitative Real-Time RT-PCR Normalization in *Iris. lactea* var. *chinensis* Roots under Cadmium, Lead, and Salt Stress Conditions. *Sci. World J.* 2014.
- Han, Y., Tang, A., Wan, H., Zhang, T., Cheng, T., Wang, J., Yang, W., Pan, H., and Zhang, Q. (2018). An APETALA2 Homolog, RcAP2, Regulates the Number of Rose Petals Derived From Stamens and Response to Temperature Fluctuations. *Front. Plant Sci.* 9.
- Hibrand-Saint Oyant, L., Crespel, L., Rajapakse, S., Zhang, L., and Foucher, F. (2007). Genetic linkage maps of rose constructed with new microsatellite markers and locating QTL controlling flowering traits. *Tree Genet. Genomes* 4, 11.
- Hong, R.L., Hamaguchi, L., Busch, M.A., and Weigel, D. (2003). Regulatory Elements of the Floral Homeotic Gene AGAMOUS Identified by Phylogenetic Footprinting and Shadowing. *Plant Cell* 15, 1296–1309.
- Hu, Y., Xie, Q., and Chua, N.-H. (2003). The Arabidopsis auxin-inducible gene ARGOS controls lateral organ size. *Plant Cell* 15, 1951–1961.
- Huang, H., Mizukami, Y., Hu, Y., and Ma, H. (1993). Isolation and characterization of the binding sequences for the product of the Arabidopsis floral homeotic gene AGAMOUS. *Nucleic Acids Res.* 21, 4769–4776.
- Huijser, P., Klein, J., Lönnig, W.E., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. *EMBO J.* 11, 1239–1249.
- Ikeda, M., Mitsuda, N., and Ohme-Takagi, M. (2009). Arabidopsis WUSCHEL Is a Bifunctional Transcription Factor That Acts as a Repressor in Stem Cell Regulation and as an Activator in Floral Patterning. *Plant Cell* 21, 3493–3505.

- Ingle, E.K.S., and Gilmartin, P.M. (2015). Molecular characterisation of four double-flowered mutants of *Silene dioica* representing four centuries of variation. *J. Exp. Bot.* *66*, 3297–3307.
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* *449*, 463–467.
- Javelle, M., Marco, C.F., and Timmermans, M. (2011). In Situ Hybridization for the Precise Localization of Transcripts in Plants. *J. Vis. Exp. JoVE*.
- Jetha, K., Theißen, G., and Melzer, R. (2014). Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res.* *42*, 10927–10942.
- Joppa, L.N., Roberts, D.L., and Pimm, S.L. (2011). How many species of flowering plants are there? *Proc. R. Soc. B Biol. Sci.* *278*, 554–559.
- Jung, J.-H., Lee, S., Yun, J., Lee, M., and Park, C.-M. (2014). The miR172 target TOE3 represses AGAMOUS expression during Arabidopsis floral patterning. *Plant Sci.* *215*, 29–38.
- Jung, S., Lee, T., Cheng, C.-H., Buble, K., Zheng, P., Yu, J., Humann, J., Ficklin, S.P., 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*, D1137–D1145.
- Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res.* *20*, 1313–1326.
- Keck, E., McSteen, P., Carpenter, R., and Coen, E. (2003). Separation of genetic functions controlling organ identity in flowers. *EMBO J.* *22*, 1058–1066.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* *10*, 845–858.
- Kidwell, M.G., and Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 7704–7711.
- Kiełbasa, S.M., Wan, R., Sato, K., Horton, P., and Frith, M.C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Res.* *21*, 487–493.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* *14*, R36.
- Kim, S., Soltis, P.S., Wall, K., and Soltis, D.E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Mol. Biol. Evol.* *23*, 107–120.
- Kim, Y.-S., Kim, S.-G., Lee, M., Lee, I., Park, H.-Y., Seo, P.J., Jung, J.-H., Kwon, E.-J., Suh, S.W., Paek, K.-H., et al. (2008). HD-ZIP III Activity Is Modulated by Competitive Inhibitors via a Feedback Loop in Arabidopsis Shoot Apical Meristem Development. *Plant Cell* *20*, 920–933.
- Kitahara, K., and Matsumoto, S. (2000). Rose MADS-box genes 'MASAKO C1 and D1' homologous to class C floral identity genes. *Plant Sci.* *151*, 121–134.

- Kitahara, K., Hirai, S., Fukui, H., and Matsumoto, S. (2001). Rose MADS-box genes 'MASAKO BP and B3' homologous to class B floral identity genes. *Plant Sci.* *161*, 549–557.
- Klocko, A.L., Borejsza-Wysocka, E., Brunner, A.M., Shevchenko, O., Aldwinckle, H., and Strauss, S.H. (2016). Transgenic Suppression of AGAMOUS Genes in Apple Reduces Fertility and Increases Floral Attractiveness. *PLoS ONE* *11*.
- Köcher, T., and Superti-Furga, G. (2007). Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* *4*, 807–815.
- Krizek, B.A. (2003). AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain. *Nucleic Acids Res.* *31*, 1859–1868.
- Krizek, B.A., and Meyerowitz, E.M. (1996). The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Dev. Camb. Engl.* *122*, 11–22.
- Krogan, N.T., Hogan, K., and Long, J.A. (2012). APETALA2 negatively regulates multiple floral organ identity genes in Arabidopsis by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19. *Development* *139*, 4180–4190.
- Krol, A.R.V.D., and Chua, N.H. (1993). Flower Development in Petunia. *Plant Cell* *5*, 1195–1203.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* *40*, D1202–D1210.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.
- Laux, T., Mayer, K.F., Berger, J., and Jürgens, G. (1996). The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. *Development* *122*, 87–96.
- Lenhard, M., Bohnert, A., Jürgens, G., and Laux, T. (2001). Termination of Stem Cell Maintenance in Arabidopsis Floral Meristems by Interactions between WUSCHEL and AGAMOUS. *Cell* *105*, 805–814.
- Linde, M., Hattendorf, A., Kaufmann, H., and Debener, T. (2006). Powdery mildew resistance in roses: QTL mapping in different environments using selective genotyping. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* *113*, 1081–1092.
- Lindsey, B.E., Rivero, L., Calhoun, C.S., Grotewold, E., and Brkljacic, J. (2017). Standardized Method for High-throughput Sterilization of Arabidopsis Seeds. *J. Vis. Exp. JoVE*.
- Lisch, D. (2013). How important are transposons for plant evolution? *Nat. Rev. Genet.* *14*, 49–61.
- Liu, X., Huang, J., Parameswaran, S., Ito, T., Seubert, B., Auer, M., Rymaszewski, A., Jia, G., Owen, H.A., and Zhao, D. (2009). The SPOROXYTELESS/NOZZLE Gene Is Involved in Controlling Stamen Identity in Arabidopsis. *Plant Physiol.* *151*, 1401–1411.
- Liu, X., Kim, Y.J., Müller, R., Yumul, R.E., Liu, C., Pan, Y., Cao, X., Goodrich, J., and Chen, X. (2011). AGAMOUS Terminates Floral Stem Cell Maintenance in Arabidopsis by Directly Repressing WUSCHEL through Recruitment of Polycomb Group Proteins[W]. *Plant Cell* *23*, 3654–3670.

- Liu, Z., Zhang, D., Liu, D., Li, F., and Lu, H. (2012). Exon skipping of AGAMOUS homolog PrseAG in developing double flowers of *Prunus lannesiana* (Rosaceae). *Plant Cell Rep.* *32*, 227–237.
- Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and Silencing-Associated Small RNAs in Plants. *Plant Cell* *14*, 1605–1619.
- Lloyd, A.M., Barnason, A.R., Rogers, S.G., Byrne, M.C., Fraley, R.T., and Horsch, R.B. (1986). Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science* *234*, 464–466.
- Longhi, S., Giongo, L., Buti, M., Surbanovski, N., Viola, R., Velasco, R., Ward, J.A., and Sargent, D.J. (2014). Molecular genetics and genomics of the Rosoideae: state of the art and future perspectives. *Hortic. Res.* *1*, 1.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* *290*, 1151–1155.
- Ma, J., Shen, X., Liu, Z., Zhang, D., Liu, W., Liang, H., Wang, Y., He, Z., and Chen, F. (2018). Isolation and Characterization of AGAMOUS-Like Genes Associated With Double-Flower Morphogenesis in *Kerria japonica* (Rosaceae). *Front. Plant Sci.* *9*, 959.
- Ma, N., Chen, W., Fan, T., Tian, Y., Zhang, S., Zeng, D., and Li, Y. (2015). Low temperature-induced DNA hypermethylation attenuates expression of RhAG, an AGAMOUS homolog, and increases petal number in rose (*Rosa hybrida*). *BMC Plant Biol.* *15*.
- Maes, T., Van de Steene, N., Zethof, J., Karimi, M., D’Hauw, M., Mares, G., Van Montagu, M., and Gerats, T. (2001). *Petunia* Ap2-like genes and their role in flower and seed development. *Plant Cell* *13*, 229–244.
- Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. *Nat. Methods* *10*, 957–963.
- Mandel, M.A., and Yanofsky, M.F. (1995). A gene triggering flower formation in *Arabidopsis*. *Nature* *377*, 522–524.
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinforma. Oxf. Engl.* *27*, 764–770.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* *17*, 10–12.
- Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., Pitrat, M., Dogimont, C., and Bendahmane, A. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature* *461*, 1135–1138.
- Martin, M., Piola, F., Chessel, D., Jay, M., and Heizmann, P. (2001). The domestication process of the Modern Rose: genetic structure and allelic composition of the rose complex. *Theor. Appl. Genet.* *102*, 398–404.
- Mayer, K.F.X., Schoof, H., Haecker, A., Lenhard, M., Jürgens, G., and Laux, T. (1998). Role of WUSCHEL in Regulating Stem Cell Fate in the *Arabidopsis* Shoot Meristem. *Cell* *95*, 805–815.

- McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* 411, 709–713.
- Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D., and Koornneef, M. (1998). *Arabidopsis thaliana*: A Model Plant for Genome Analysis. *Science* 282, 662–682.
- Melzer, R., Verelst, W., and Theissen, G. (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. *Nucleic Acids Res.* 37, 144–157.
- Mendes, M.A., Guerra, R.F., Berns, M.C., Manzo, C., Masiero, S., Finzi, L., Kater, M.M., and Colombo, L. (2013). MADS Domain Transcription Factors Mediate Short-Range DNA Looping That Is Essential for Target Gene Expression in *Arabidopsis*[W]. *Plant Cell* 25, 2560–2572.
- Meng, Y., Shao, C., Wang, H., and Chen, M. (2011). The Regulatory Activities of Plant MicroRNAs: A More Dynamic Perspective1. *Plant Physiol.* 157, 1583–1595.
- Mizukami, Y., and Fischer, R.L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 942–947.
- Mizukami, Y., and Ma, H. (1992). Ectopic expression of the floral homeotic gene AGAMOUS in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* 71, 119–131.
- Morel, P., Heijmans, K., Rozier, F., Zethof, J., Chamot, S., Bento, S.R., Vialette-Guiraud, A., Chambrier, P., Trehin, C., and Vandenbussche, M. (2017). Divergence of the Floral A-Function between an Asterid and a Rosid Species. *Plant Cell* 29, 1605–1621.
- Mott, R. (1997). EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* CABIOS 13, 477–478.
- Mukherjee, K., and Bürglin, T.R. (2006). MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins. *Plant Physiol.* 140, 1142–1150.
- Murashige, T., and Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant.* 15, 473–497.
- Nakatsuka, T., and Koishi, K. (2018). Molecular characterization of a double-flower mutation in *Matthiola incana*. *Plant Sci.* 268, 39–46.
- Nakatsuka, T., Saito, M., Yamada, E., Fujita, K., Yamagishi, N., Yoshikawa, N., and Nishihara, M. (2015). Isolation and characterization of the C-class MADS-box gene involved in the formation of double flowers in Japanese gentian. *BMC Plant Biol.* 15, 182.
- Ng, M., and Yanofsky, M.F. (2001). Function and evolution of the plant MADS-box gene family. *Nat. Rev. Genet.* 2, 186–195.
- Nitasaka, E. (2003). Insertion of an En/Spm-related transposable element into a floral homeotic gene DUPLICATED causes a double flower phenotype in the Japanese morning glory. *Plant J.* 36, 522–531.
- Noor, S.H., Ushijima, K., Murata, A., Yoshida, K., Tanabe, M., Tanigawa, T., Kubo, Y., and Nakano, R. (2014). Double flower formation induced by silencing of C-class MADS-box genes and its variation among *petunia* Cultivars. *Sci. Hortic.* 178, 1–7.

- Odell, J.T., Knowlton, S., Lin, W., and Mauvais, C.J. (1988). Properties of an isolated transcription stimulating sequence derived from the cauliflower mosaic virus 35S promoter. *Plant Mol. Biol.* *10*, 263–272.
- Ohno, S. (1970). *Evolution by Gene Duplication* (Berlin Heidelberg: Springer-Verlag).
- Okumura, A., Shimada, A., Yamasaki, S., Horino, T., Iwata, Y., Koizumi, N., Nishihara, M., and Mishiba, K. (2016). CaMV-35S promoter sequence-specific DNA methylation in lettuce. *Plant Cell Rep.* *35*, 43–51.
- Ó'Maoiléidigh, D.S., Wuest, S.E., Rae, L., Raganelli, A., Ryan, P.T., Kwaśniewska, K., Das, P., Lohan, A.J., Loftus, B., Graciet, E., et al. (2013). Control of Reproductive Floral Organ Identity Specification in *Arabidopsis* by the C Function Regulator AGAMOUS. *Plant Cell* *25*, 2482–2503.
- Oosumi, T., Gruszewski, H.A., Blischak, L.A., Baxter, A.J., Wadl, P.A., Shuman, J.L., Veilleux, R.E., and Shulaev, V. (2006). High-efficiency transformation of the diploid strawberry (*Fragaria vesca*) for functional genomics. *Planta* *223*, 1219–1230.
- Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* *53*, 674–690.
- Otsuga, D., DeGuzman, B., Prigge, M.J., Drews, G.N., and Clark, S.E. (2001). REVOLUTA regulates meristem initiation at lateral positions. *Plant J. Cell Mol. Biol.* *25*, 223–236.
- Otto, S.P., and Yong, P. (2002). The evolution of gene duplicates. *Adv. Genet.* *46*, 451–483.
- Pajoro, A., Madrigal, P., Muiño, J.M., Matus, J.T., Jin, J., Mecchia, M.A., Debernardi, J.M., Palatnik, J.F., Balazadeh, S., Arif, M., et al. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol.* *15*, R41.
- Pascal, T., Aberlenc, R., Confolent, C., Hoerter, M., Lecerf, E., Tuéro, C., and Lambert, P. (2017). Mapping of new resistance (Vr2, Rm1) and ornamental (Di2, pl) Mendelian trait loci in peach. *Euphytica* *213*, 132.
- Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* *405*, 200–203.
- Pelaz, S., Tapia-López, R., Alvarez-Buylla, E.R., and Yanofsky, M.F. (2001). Conversion of leaves into petals in *Arabidopsis*. *Curr. Biol. CB* *11*, 182–184.
- Phillips, R. (1988). *Roses* (Random House).
- Polpitiya, A.D., Qian, W.-J., Jaitly, N., Petyuk, V.A., Adkins, J.N., Camp, D.G., Anderson, G.A., and Smith, R.D. (2008). DANTE: a statistical tool for quantitative analysis of -omics data. *Bioinforma. Oxf. Engl.* *24*, 1556–1558.
- Ponting, C.P., and Aravind, L. (1999). START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem. Sci.* *24*, 130–132.
- Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., Marcellin, R., Melzer, R., Brown, E., Hart, D., et al. (2014). Structural Basis for the Oligomerization of the MADS Domain Transcription Factor SEPALLATA3 in *Arabidopsis*[W]. *Plant Cell* *26*, 3603–3615.

- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* *33*, W116–W120.
- Raymond, O. (1999). Domestication et sélection dirigée chez le rosier : analyse historique via les phénotypes morphologique, chimique et biochimique.
- Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., Vergne, P., Moja, S., Choisne, N., Pont, C., et al. (2018). The Rosa genome provides new insights into the domestication of modern roses. *Nat. Genet.* *50*, 772–777.
- Reiser, L., Subramaniam, S., Li, D., and Huala, E. (2017). Using the Arabidopsis Information Resource (TAIR) to Find Information About Arabidopsis Genes. *Curr. Protoc. Bioinforma.* *60*, 1.11.1-1.11.45.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet. TIG* *16*, 276–277.
- Riechmann, J.L., and Meyerowitz, E.M. (1997). MADS domain proteins in plant development. *Biol. Chem.* *378*, 1079–1101.
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* *42*, W320–W324.
- Rotmistrovsky, K., Jang, W., and Schuler, G.D. (2004). A web server for performing electronic PCR. *Nucleic Acids Res.* *32*, W108–W112.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinforma. Oxf. Engl.* *16*, 944–945.
- Sablowski, R.W.M., and Meyerowitz, E.M. (1998). A Homolog of NO APICAL MERISTEM Is an Immediate Target of the Floral Homeotic Genes APETALA3/PISTILLATA. *Cell* *92*, 93–103.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* *4*, 406–425.
- Sauquet, H., Balthazar, M. von, Magallón, S., Doyle, J.A., Endress, P.K., Bailes, E.J., Morais, E.B. de, Bull-Hereñu, K., Carrive, L., Chartier, M., et al. (2017). The ancestral flower of angiosperms and its early diversification. *Nat. Commun.* *8*, ncomms16047.
- Schena, M., and Davis, R.W. (1992). HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily. *Proc. Natl. Acad. Sci. U. S. A.* *89*, 3894–3898.
- Scholl, R.L., May, S.T., and Ware, D.H. (2000). Seed and Molecular Resources for Arabidopsis. *Plant Physiol.* *124*, 1477–1480.
- Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F.X., Jürgens, G., and Laux, T. (2000). The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. *Cell* *100*, 635–644.
- Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. (2006). Highly Specific Gene Silencing by Artificial MicroRNAs in Arabidopsis. *Plant Cell* *18*, 1121–1133.

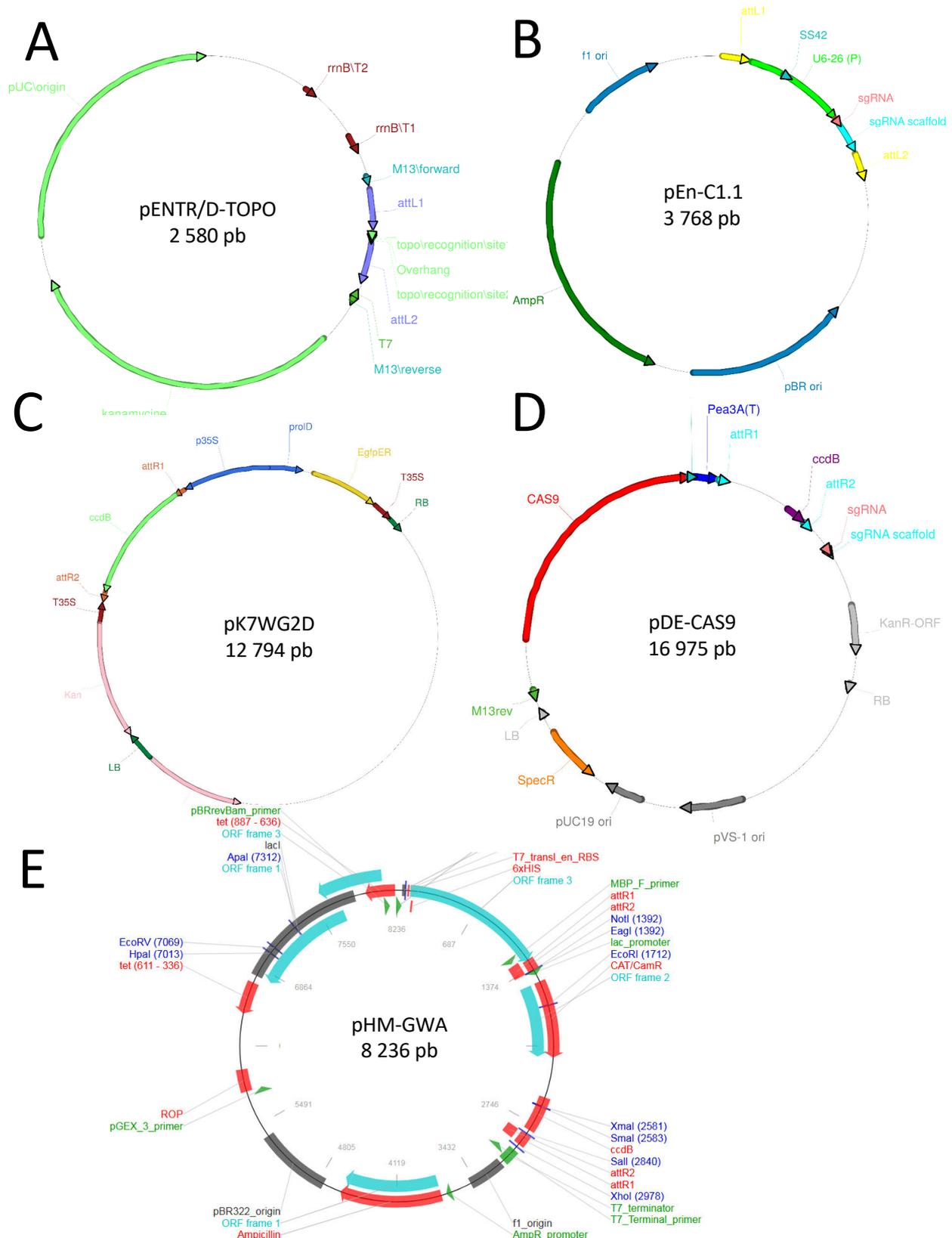
- Shi, J.X., Malitsky, S., De Oliveira, S., Branigan, C., Franke, R.B., Schreiber, L., and Aharoni, A. (2011). SHINE transcription factors act redundantly to pattern the archetypal surface of Arabidopsis flower organs. *PLoS Genet.* *7*, e1001388.
- Shulaev, V., Korban, S.S., Sosinski, B., Abbott, A.G., Aldwinckle, H.S., Folta, K.M., Iezzoni, A., Main, D., Arús, P., Dandekar, A.M., et al. (2008). Multiple Models for Rosaceae Genomics. *Plant Physiol.* *147*, 985–1003.
- Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., et al. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* *43*, 109–116.
- Shupert, D.A. (2005). Inheritance of flower, stem, leaf, and disease traits in three diploid interspecific rose populations.
- Sieburth, L.E., and Meyerowitz, E.M. (1997). Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically. *Plant Cell* *9*, 355–365.
- Sigrist, C.J.A., de Castro, E., Cerutti, L., Cucho, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* *41*, D344-347.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinforma. Oxf. Engl.* *31*, 3210–3212.
- Smaczniak, C., Immink, R.G.H., Muiño, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D. (Peter), Liu, S., Westphal, A.H., Boeren, S., et al. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc. Natl. Acad. Sci.* *109*, 1560–1565.
- Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* *7 Suppl 1*, S10.1-12.
- Soltis, D.E., Chanderbali, A.S., Kim, S., Buzgo, M., and Soltis, P.S. (2007). The ABC Model and its Applicability to Basal Angiosperms. *Ann. Bot.* *100*, 155–163.
- Sonnhammer, E.L., and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* *167*, GC1-10.
- Soule, valerie ann (2009). Analysis of genetic diversity and relationships in the china rose group.
- Spiller, M., Linde, M., Oyant, L.H.-S., Tsai, C.-J., Byrne, D.H., Smulders, M.J.M., Foucher, F., and Debener, T. (2010). Towards a unified genetic map for diploid roses. *Theor. Appl. Genet.* *122*, 489–500.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G.R., Korf, I., Lapp, H., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* *12*, 1611–1618.
- Stigliani, A., Martin-Arevalillo, R., Lucas, J., Bessy, A., Vinos-Poyo, T., Mironova, V., Vernoux, T., Dumas, R., and Parcy, F. (2018). Capturing Auxin Response Factors Syntax Using DNA Binding Models. *Mol. Plant.*
- Stitzer, M.C., and Ross-Ibarra, J. (2018). Maize domestication and gene interaction. *New Phytol.* *220*, 395–408.

- Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* *43*, 1160–1163.
- Sun, B., Xu, Y., Ng, K.-H., and Ito, T. (2009). A timing mechanism for stem cell maintenance and differentiation in the Arabidopsis floral meristem. *Genes Dev.* *23*, 1791–1804.
- Sun, Y., Fan, Z., Li, X., Liu, Z., Li, J., and Yin, H. (2014). Distinct double flower varieties in *Camellia japonica* exhibit both expansion and contraction of C-class gene expression. *BMC Plant Biol.* *14*.
- Szécsi, J., Joly, C., Bordji, K., Varaud, E., Cock, J.M., Dumas, C., and Bendahmane, M. (2006). BIGPETALp, a bHLH transcription factor is involved in the control of Arabidopsis petal size. *EMBO J.* *25*, 3912–3920.
- Tajima, F., and Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* *1*, 269–285.
- Takeda, S., Matsumoto, N., and Okada, K. (2004). RABBIT EARS, encoding a SUPERMAN-like zinc finger protein, regulates petal development in Arabidopsis thaliana. *Dev. Camb. Engl.* *131*, 425–434.
- Tanaka, Y., Oshima, Y., Yamamura, T., Sugiyama, M., Mitsuda, N., Ohtsubo, N., Ohme-Takagi, M., and Terakawa, T. (2013). Multi-petal cyclamen flowers produced by AGAMOUS chimeric repressor expression. *Sci. Rep.* *3*, 2641.
- Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K.L., Yandell, M., Gundlach, H., et al. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* *15*, 312.
- Tasaki, K., Higuchi, A., Fujita, K., Watanabe, A., Sasaki, N., Fujiwara, K., Abe, H., Naito, Z., Takahashi, R., Hikage, T., et al. (2017). Development of molecular markers for breeding of double flowers in Japanese gentian. *Mol. Breed.* *37*, 33.
- Tatum, T.C., Stepanovic, S., Biradar, D.P., Rayburn, A.L., and Korban, S.S. (2005). Variation in nuclear DNA content in *Malus* species and cultivated apples. *Genome* *48*, 924–930.
- Tatusova, T.A., and Madden, T.L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* *174*, 247–250.
- The International Peach Genome Initiative, Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J., 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.
- Thomas, R. (2016). *Petite Flore de France*.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* *7*, 562–578.
- Tröbner, W., Ramirez, L., Motte, P., Hue, I., Huijser, P., Lönning, W.E., Saedler, H., Sommer, H., and Schwarz-Sommer, Z. (1992). GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. *EMBO J.* *11*, 4693–4704.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* *40*, e115.

- Unver, T., and Budak, H. (2009). Virus-Induced Gene Silencing, a Post Transcriptional Gene Silencing Method. *Int. J. Plant Genomics* 2009.
- Van de Peer, Y., and De Wachter, R. (1994). TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Appl. Biosci.* CABIOS 10, 569–570.
- VanBuren, R., Bryant, D., Bushakra, J.M., Vining, K.J., Edger, P.P., Rowley, E.R., Priest, H.D., Michael, T.P., Lyons, E., Filichkin, S.A., et al. (2016). The genome of black raspberry (*Rubus occidentalis*). *Plant J. Cell Mol. Biol.* 87, 535–547.
- Varaud, E., Brioudes, F., Szécsi, J., Leroux, J., Brown, S., Perrot-Rechenmann, C., and Bendahmane, M. (2011). AUXIN RESPONSE FACTOR8 Regulates Arabidopsis Petal Growth by Interacting with the bHLH Transcription Factor BIGPETALp[C][W]. *Plant Cell* 23, 973–983.
- Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., et al. (2010). The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat. Genet.* 42, 833–839.
- Vergne, P., Maene, M., Gabant, G., Chauvet, A., Debener, T., and Bendahmane, M. (2009). Somatic embryogenesis and transformation of the diploid *Rosa chinensis* cv Old Blush. *Plant Cell Tissue Organ Cult.* PCTOC 100, 73.
- Vergne, P., Just, J., Szécsi, J., Dubois, A., Raymond, O., and Bendahmane, M. (2018). Production of homozygous rose line derived from heterozygous genotype.
- Wang, X.-Q., and Ran, J.-H. (2014). Evolution and biogeography of gymnosperms. *Mol. Phylogenet. Evol.* 75, 24–40.
- Wang, Y.-Q., Melzer, R., and Theißen, G. (2011). A double-flowered variety of lesser periwinkle (*Vinca minor* fl. pl.) that has persisted in the wild for more than 160 years. *Ann. Bot.* 107, 1445–1452.
- Weinhold, A., Kallenbach, M., and Baldwin, I.T. (2013). Progressive 35S promoter methylation increases rapidly during vegetative development in transgenic *Nicotiana attenuata* plants. *BMC Plant Biol.* 13, 99.
- Wenkel, S., Emery, J., Hou, B.-H., Evans, M.M.S., and Barton, M.K. (2007). A Feedback Regulatory Module Formed by LITTLE ZIPPER and HD-ZIPIII Genes. *Plant Cell* 19, 3379–3390.
- Wessler, S.R. (1988). Phenotypic diversity mediated by the maize transposable elements Ac and Spm. *Science* 242, 399–405.
- Wessler, S.R. (2006). Transposable elements and the evolution of eukaryotic genomes. *Proc. Natl. Acad. Sci.* 103, 17600–17601.
- West, A.G., Shore, P., and Sharrocks, A.D. (1997). DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. *Mol. Cell. Biol.* 17, 2876–2887.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS ONE* 2.
- Wolff, C.F. (1759). *Theoria Generationis*.

- Wollmann, H., Mica, E., Todesco, M., Long, J.A., and Weigel, D. (2010). On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. *Dev. Camb. Engl.* *137*, 3633–3642.
- Woźniak, N.J., and Sicard, A. (2018). Evolvability of flower geometry: Convergence in pollinator-driven morphological evolution of flowers. *Semin. Cell Dev. Biol.* *79*, 3–15.
- Wuest, S.E., O'Maoileidigh, D.S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., Lohan, A.J., Loftus, B., Graciet, E., and Wellmer, F. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proc. Natl. Acad. Sci.* *109*, 13452–13457.
- Würschum, T., Groß-Hardt, R., and Laux, T. (2006). APETALA2 Regulates the Stem Cell Niche in the Arabidopsis Shoot Meristem. *Plant Cell* *18*, 295–307.
- Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., Chen, H., Xiang, J., and 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*, 262–281.
- Yan, W., Chen, D., and Kaufmann, K. (2016). Molecular mechanisms of floral organ specification by MADS domain proteins. *Curr. Opin. Plant Biol.* *29*, 154–162.
- Yant, L., Mathieu, J., Dinh, T.T., Ott, F., Lanz, C., Wollmann, H., Chen, X., and Schmid, M. (2010). Orchestration of the Floral Transition and Floral Development in Arabidopsis by the Bifunctional Transcription Factor APETALA2[W][OA]. *Plant Cell* *22*, 2156–2170.
- Yokoya, K., Roberts, A.V., Mottley, J., Lewis, R., and Brandham, P.E. (2000). Nuclear DNA Amounts in Roses. *Ann. Bot.* *85*, 557–561.
- Zharkikh, A., Troggio, M., Pruss, D., Cestaro, A., Eldrdge, G., Pindo, M., Mitchell, J.T., Vezzulli, S., Bhatnagar, S., Fontana, P., et al. (2008). Sequencing and assembly of highly heterozygous genome of *Vitis vinifera* L. cv Pinot Noir: Problems and solutions. *J. Biotechnol.* *136*, 38–43.

ANNEXES



Annexe 1 : Cartes des plasmides utilisés pour les clonages.

(A) et (B) plasmides utilisés comme vecteurs d'entrée pour les surexpressions et synthèses de protéines (A) et pour les éditions de gène par CRISPR/Cas9 (B). Plasmide de destination utilisé pour les surexpressions (C), pour les éditions de gène par CRISPR/Cas9 (D) et pour les synthèses de protéines dans *E. coli* (E).

Les cartes ont été obtenues du site web : <https://www.addgene.org/vector-database/>