



**HAL**  
open science

# Analyse génétique et fonctionnelle de la biosynthèse des arômes des raisins

Thuy-Thanh Truong

► **To cite this version:**

Thuy-Thanh Truong. Analyse génétique et fonctionnelle de la biosynthèse des arômes des raisins. Génétique des plantes. Université de Strasbourg, 2020. Français. NNT: 2020STRAJ003. tel-04212523

**HAL Id: tel-04212523**

**<https://theses.hal.science/tel-04212523v1>**

Submitted on 20 Sep 2023

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

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

*École doctorale des Sciences de la Vie et de la Santé*

**INRA Grand Est Colmar**

Santé de la Vigne et Qualité du Vin – UMR 1131

# THÈSE

présentée par :

**Thuy-Thanh TRUONG**

soutenue le : **30 janvier 2020**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline : Sciences de la Vie

Spécialité : Aspects moléculaires et cellulaires de la biologie

## Analyse génétique et fonctionnelle de la biosynthèse des arômes des raisins

**THÈSE dirigée par :**

**M. HUGUENEY Philippe**

Directeur de Recherche INRA, Université de Strasbourg

**RAPPORTEURS :**

**Mme ADRIAN Marielle**

**M. FOUCHER Fabrice**

Professeur, Université de Bourgogne Franche-Comté

Directeur de Recherche INRA, IRHS Angers

**AUTRES MEMBRES DU JURY :**

**Mme BAILLIEUL Fabienne**

**M. DUCHÊNE Éric**

**Mme HEMMERLIN Andréa**

Professeur, Université de Reims Champagne Ardenne

Ingénieur de Recherche INRA, Université de Strasbourg

Chargée de Recherche CNRS, IBMP, Université de Strasbourg

Thèse financée par :



*Qu'est-ce que vous regardez ? C'est la carte routière ? – Non ! C'est la carte des vins.*

*C'est pour éviter les bouchons !*

*Raymond Devos*



# Remerciements

Ces trois années de thèse ont été si intenses que je ne les ai pas vu passer et il est déjà temps de remercier les gens qui m'ont accompagné pendant cette aventure.

Je tiens tout d'abord à remercier Marielle Adrian, Fabrice Foucher, Fabienne Baillieul, Andréa Hemmerlin, qui ont accepté d'évaluer mon travail.

Je remercie également pour leur participation à mon comité de thèse, Philippe Darriet et Joseph Schacherer, vos conseils et remarques m'ont été d'une aide précieuse.

Cette thèse n'aurait pas été réalisable sans le soutien de la Région Grand Est Alsace et à l'INRA. Je voudrais donc les remercier pour m'avoir offert cette opportunité. J'aimerais aussi remercier le DAAD et l'INRA pour avoir financé mon séjour de 3 mois à Munich au Helmholtz Zentrum.

J'aimerais remercier Philippe Huguency, mon directeur de thèse et Eric Duchêne mon co-encadrant, sans qui je n'aurais pu réaliser cette thèse. Merci pour votre encadrement, votre écoute, vos précieux conseils et vos encouragements. Ce fut vraiment un plaisir de pouvoir travailler avec vous.

Ce projet de thèse n'aurait pas pu être réalisé sans collaboration scientifique. Je remercie chaleureusement mes collègues de l'équipe de Métabolisme Secondaire de la Vigne, et de l'équipe Génétique et Amélioration de la Vigne. Merci à Nathalie Jaegli, Vincent Dumas et Gisèle Butterlin d'avoir été à mon écoute et pour votre bonne humeur quoi qu'il arrive ! Un grand merci à Raymonde Baltenweck et Patricia Claudel pour tout ce que vous m'avez appris sur les instruments analytiques. Je vous remercie aussi pour votre patience et vos réponses à mes innombrables questions. Merci à toi Raymonde, d'avoir eu la patience de m'aider à déterminer la structure d'un si grand nombre de molécules, sans toi j'aurais vraiment été perdue ! Merci à Camille Rustenholtz, Amandine Velt et Sophie Blanc pour les analyses bioinformatiques qui m'ont permis d'avancer pendant ma thèse. Merci à Camille pour ton écoute et ton soutien (Coldplay Power !). Merci à toi Amandine d'être venue à mon secours tant de fois lorsque R me mettait des bâtons dans les roues. Merci à Lauriane Renault pour ton aide dans les manips mais surtout pour ton amitié et ton écoute tout au long de cette thèse. Les

## Remerciements

mugs cakes vont me manquer et tes petites douceurs aussi (vive le beurre salé) ! Je remercie l'ensemble de mes collègues pour m'avoir si bien accueillie. L'horoscope du café, les blagues de Jef et surtout le tarot vont me manquer !

Merci à Philippe Schmitt-Kopplin d'avoir accepté de m'accueillir dans son équipe pendant trois mois pour les analyses FT-ICR. Merci à Basem Kanawati, Jenny Ulm et Marianna Lucio pour votre aide sur les analyses par FT-ICR. Ce séjour a vraiment été enrichissant tant professionnellement qu'humainement et m'a conforté sur mon envie d'avoir une expérience à l'international.

Je tiens à témoigner toute mon amitié envers les thésardes, thésards et post-doc de l'INRA de Colmar que j'ai eu la chance de côtoyer : merci à Isabelle, Lise, Antoine, Michaël, mes colocataires de bureau. Merci à toi Lise plus particulièrement pour m'avoir fourni en chips et/ou café durant les périodes difficiles.

Enfin, je remercie évidemment ma famille et mes amis, sans qui je n'aurais pu tenir sur la durée. Merci à toi Brice pour ton amour et pour les soirées raclettes improvisées pour me remonter le moral ! Merci à mes parents pour leur soutien durant toutes ces années et surtout d'avoir cru en moi. Merci à vous Danièle, qui avait toujours été là pour moi.

Je conclus ici en remerciant une nouvelle fois l'ensemble du personnel de l'INRA de Colmar, en vous souhaitant une très bonne continuation.

# Table des matières

Remerciements .....	3
Table des matières .....	5
Liste des abréviations .....	9
Liste des figures .....	11
Liste des tableaux .....	15
Introduction .....	19
1. Contexte général.....	21
2. La composition des raisins .....	23
2.1. Sucrosité et acidité .....	23
2.2. Métabolisme secondaire.....	24
3. Les déterminants génétiques et moléculaires de la biosynthèse des arômes des raisins	25
3.1. Les terpènes .....	27
3.2. Les sesquiterpènes des raisins.....	36
3.3. Les norisoprénoïdes .....	39
3.4. Analyse génétique de la biosynthèse des méthoxypyrazines dans les raisins .....	42
3.5. Les produits de dégradation des lipides .....	46
3.6. Les thiols volatils .....	48
4. Objectifs de la thèse .....	50
Matériels et méthodes.....	53
1. Population d'étude RixGw et échantillons analysés pendant la thèse .....	55
1.1. Matériel végétal .....	57
1.2. Extraction des échantillons en vue des analyses métabolomiques .....	57
2. Analyses métabolomiques de la population de croisement RixGw .....	59
2.1. Description des outils analytiques utilisés pendant la thèse .....	59



2.2. Traitement des données issus des analyses métabolomiques .....	63
3. Génotypage de la population de croisement RixGw .....	67
3.1. Génotypage par marqueurs microsatellites .....	67
3.2. Génotypage par puce Illumina 18k .....	69
4. Recherche de « métabolite Quantitative Trait Loci » (mQTL) et recherche des gènes candidats associés.....	73
5. Identification des gènes candidat liés aux mQTL .....	74
5.1. Génomes de référence et données de transcriptomique .....	75
5.2. Fonctions putatives des protéines correspondantes aux gènes annotés dans les intervalles de confiance des QTL.....	75
6. Clonage et caractérisation fonctionnelle de la CAMT .....	76
6.1. Clonage du gène par la technique Gateway .....	76
6.2. Expression chez <i>Nicotiana benthamiana</i> .....	76
6.3. Test enzymatique in vitro avec la protéine CAMT recombinante exprimée dans <i>E.coli</i> 77	
Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS.....	79
1. Introduction .....	81
2. Article.....	82
3. Discussion .....	149
Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées .....	151
1. Introduction .....	153
2. Résultats .....	155
2.1. Analyse non ciblée du déterminisme génétique du métabolome des baies par UHPLC-MS.....	155
2.2. Étude des précurseurs d'arômes présent dans les baies par UHPLC-MS ESI (+)176	
2.3. Apport de l'analyse métabolomique des feuilles dans la compréhension du métabolome de la baie.....	203

3. Discussion .....	212
3.1. Diversité des molécules détectées dans les baies de la population RixGw et déterminisme génétique associé .....	212
3.2. Les glycosides : des précurseurs d'arômes potentiels abondants dans les baies .	213
3.3. Les extraits de feuilles, un modèle intéressant à étudier.....	215
3.4. L'identification précise des molécules, une tâche difficile.....	216
3.5. La métabolomiques non ciblée en UHPLC-MS pour l'analyse du déterminisme génétique de la biosynthèse d'un grand nombre de métabolites .....	216
Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL .....	219
1. Introduction .....	221
2. Article 2.....	222
3. Discussion .....	249
Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats .....	251
1. Introduction .....	253
2. Identification des gènes candidats liés aux QTL d'intérêt et caractérisation fonctionnelle	254
2.1. Gènes candidats sur le chromosome 10 liés au linalol et à ses dérivés .....	254
2.2. Gènes candidats sur le chromosome 2 liés au contenu en vanilline dans les baies	256
3. Discussion .....	261
Conclusion et perspectives .....	265
1. Des outils de spectrométrie de masse complémentaires au service de l'analyse du métabolome de la vigne.....	267
2. Le métabolome des baies de raisins : une source de précurseurs d'arômes des vins..	270
3. Les feuilles : un outil d'étude du métabolome de la vigne à exploiter.....	271
4. Une approche innovante : l'analyse en réseaux du métabolome à la lumière des QTL	272

## Table des matières

5. L'identification des gènes candidats et leur caractérisation fonctionnelle, une étape limitante dans la compréhension du déterminisme génétique du métabolome de la vigne	273
6. Des marqueurs génétiques comme outils pour une application directe dans la création variétale .....	274
Annexe 1 : QTL GC-MS apparaissant uniquement sur un an d'analyse. ....	275
Références bibliographiques .....	281

## Liste des abréviations

**ADN** : Acide désoxyribonucléique

**a.m.u** : Unité de masse atomique

**ATS**:  $\alpha$ -terpénoïde Synthase

**CAMT** : Caffeic Acid O-Methyltransferase

**Chr** : Chromosome

**CCD** : Carotenoid Cleavage Dioxygenase

**cM** : Centimorgan

**CYP**: Cytochrome P450

**Da** : Dalton

**DXS** : 1-déoxy-D-xylulose 5-phosphate synthase

**ESI** : Electrospray Ionization

**FPKM** : Fragments Per Kilobase Million

**FT-ICR** : Fourier Transform - Ion Cyclotron Resonance

**GBS** : Genotyping by Sequencing

**GC-MS** : Gas Chromatography – Mass Spectrometry

**Gw** : Gewürztraminer

**GPP** : Geranylpyrophosphate

**IBMP** : 3-isobutyl-2-méthoxy-pyrazine

**IPTG** : Isopropyl  $\beta$ -D-1-thiogalactopyranoside

**LIS** : Linalol Synthase

**LOD** : Logarithm of Odds

**m/z** : Rapport masse sur charge

**MEP** : Voie plastidiale du méthylérythritol phosphate

## Liste des abréviations

**Min** : Minute

**mQTL** : Metabolite Quantitative Trait Locus

**MVA** : Voie cytosolique du mévalonate

**Nb** : Nombre

**OMT** : O-Methyltransferase

**Pb** : Paire de base

**pH** : Potentiel hydrogène

**PCR** : Reaction de Polymerisation en chaine

**PN40024** : *Vitis vinifera* variété Pinot Noir 40024

**Pos** : Position

**Ppm** : Partie Par Million

**QTL** : Quantitative Trait Locus

**Ri** : Riesling

**RixGw** : Population issu d'un croisement entre un génotype Riesling 49 et un Gewurztraminer 643

**RNA-seq** : Ribonucleic Acid Sequencing

**RMN** : Résonance Magnétique Nucléaire

**Rpm** : Revolution par minute

**SNP** : Single Nucleotid Polymorphism

**SPE** : Solid Phase Extraction

**TDN** : 1,1,6-trimethyl-1,2-dihydronaphtalene

**TPS** : Terpenol Synthase

**UHPLC-MS** : Ultra-High Pressure Liquid Chromatography – Mass Spectrometry

**UV** : Ultraviolet

## Liste des figures

**Figure 1 :** Voies de biosynthèse des terpènes chez les plantes (d'après Champagne and Boutry, 2016).

**Figure 2 :** Biosynthèse du 1-déoxy-D-xylulose 5-phosphate (DXP) catalysée par l'enzyme DXS.

**Figure 3 :** Comparaison des séquences protéiques des allèles VvDXS (allèle neutre), VvDXSK284N des Muscats et VvDXSR306C du Gewurztraminer (vue partielle des protéines d'après Hugueney et al., 2012).

**Figure 4 :** Introduction des mutations de la DXS de vigne, associées au caractère aromatique dans la protéine DXS d'E. coli (d'après Hugueney et al., 2012).

**Figure 5 :** Modèle de la formation de la wine lactone au cours du vieillissement du vin, à partir de précurseurs accumulés dans les raisins (d'après Ilc et al., 2017).

**Figure 6 :** Modèle moléculaire de l'allèle TPS24  $\alpha$ -guaiène synthase (VvGuaS) de la Syrah.

**Figure 7 :** Voie de biosynthèse de la rotundone chez la Syrah.

**Figure 8 :** Activité de clivage du  $\beta$ -carotène par l'enzyme CCD1 de pétunia (d'après Simkin et al., 2004).

**Figure 9 :** Voies de biosynthèse putatives de l'IPMP et l'IBMP, les principales méthoxypyrazines des raisins (d'après Dunlevy et al., 2010).

**Figure 10 :** Corrélation entre le caractère « poivron vert » perçu à la dégustation et la concentration en IBMP dans les vins (d'après Roujou de Boubée et al., 2000).

**Figure 11 :** Analyse de l'expression des gènes VvOMT dans les baies des cépages Carménère (fort producteur de méthoxypyrazines) et Petit Verdot (très faible producteur de méthoxypyrazines) (d'après Guillaumie et al., 2013).

**Figure 12 :** Accumulation de méthoxypyrazines dans le milieu de culture, après expression des ADNc de différents allèles des gènes VvOMT1, VvOMT2, VvOMT3 et VvOMT4 chez E. coli (d'après Guillaumie et al., 2013).

**Figure 13 :** Représentation schématique de la voie des lipoxygénases (d'après Niinemets et al., 2013).

**Figure 14 :** Les principaux thiols volatils impliqués dans les arômes des vins.

**Figure 15 :** Biosynthèse des principaux thiols volatils des vins à partir de leurs précurseurs présents dans les raisins (Roland et al., 2011).

**Figure 16 :** Population d'étude plantée à Bergheim (68) (Source : Eric Duchêne).

**Figure 17 :** Photo de la GC-MS (source : Patricia Claudel).

**Figure 18** : Photo de l'UHPLC-MS.

**Figure 19** : FT-ICR au Helmholtz Zentrum à Munich.

**Figure 20** : Exemple d'un ion ( $m/z = 465.233$ ) aligné entre FT-ICR et UHPLC-MS ESI (+).

**Figure 21** : Carte génétique SSR. Nombre de marqueurs : 157.

**Figure 22**: Carte génétique SNP du Riesling (SNP\_Ri). Nombre de marqueurs : 953.

**Figure 23** : Carte génétique SNP du Gewurztraminer (SNP\_GW). Nombre de marqueurs : 806.

**Figure 24** : Recherche de QTL métaboliques (mQTL).

**Figure 25**: Stratégie de sélection des gènes candidats associés à un QTL.

**Figure 26** : Diagramme de Venn représentant le nombre d'ions détectés par UHPLC-MS ESI (+) dans les baies prélevées en 2014 et en 2015.

**Figure 27** : Aires des 100 pics les plus intenses et  $m/z$  des ions associés dans les analyses de baies de Ri et de Gw par UHPLC-MS ESI (+).

**Figure 28** : Représentation des ions détectés par UHPLC-MS ESI (+) en moyenne dans les baies de Ri, Gw et le maximum détecté dans la population RixGw en 2014 et en 2015.

**Figure 29** : Détection de QTL à l'échelle du génome entier avec les analyses métabolomiques par UHPLC-MS ESI (+) de baies ( $pval < 0.05$ ).

**Figure 30** : Détection de QTL à l'échelle du génome entier avec les analyses métabolomiques par UHPLC-MS ESI (-) de baies ( $pval < 0.05$ ).

**Figure 31** : QTL détectés avec les données UHPLC-MS ESI (+) de baies sur le chromosome 5 avec la carte SNP\_Gw ( $pval \leq 0.05$ ).

**Figure 32** : QTL détectés avec les données UHPLC-MS ESI (+) de baies sur le chromosome 6 avec la carte SNP\_Gw ( $pval \leq 0.05$ ).

**Figure 33** : QTL détectés avec les données UHPLC-MS ESI (+) de baies sur le chromosome 12 avec la carte SNP\_Ri ( $pval \leq 0.05$ ).

**Figure 34** : Courant d'ions des trois fragments caractéristiques de monoterpénols recherchés dans une analyse par UHPLC-MS d'extrait SPE de baies d'un génotype de la population (0074E) en 2014.

**Figure 35** : Sucres pouvant être liés aux terpénols (Hjelmeland and Ebeler, 2015).

**Figure 36** : Spectre de masse ESI (+) au temps de rétention 4.2min dans l'échantillon ayant la plus forte aire de pic pour le  $m/z$  153.127 au temps de rétention 4.3 min en 2015.

**Figure 37 :** Proposition de structure de la molécule reliée au fragment de monoterpénol 153-2, l'apiosyl-glucoside de linalol oxyde.

**Figure 38 :** Histogrammes de la distribution de l'aire de pic de l'ion avec un m/z de 482.260 au temps de rétention 4.2 min détecté dans les baies de la descendance RixGw par UHPLC-MS ESI (+).

**Figure 39 :** Effet des QTL de la carte génétique SSR sur l'aire de pic de la molécule d'apiosyl-glucoside en 2015 dans la population Ri(AB) x Gw(CD).

**Figure 40 :** Effet du QTL 12@19.9 de la carte génétique SNP\_Ri sur l'aire de pic de la molécule d'apiosyl-glucoside de linalol oxyde en 2015.

**Figure 41 :** Courant d'ions des fragments caractéristiques de norisoprénoides d'une analyse par UHPLC-MS d'extrait SPE de baie d'un génotype de la population (0074E) en 2014.

**Figure 42 :** Spectre de masse au temps de rétention 2.3 min dans l'échantillon du génotype 218E en 2015 (échantillon ayant l'aire de pic maximale de la population pour M207T147).

**Figure 43 :** Proposition de structure de la molécule liée au fragment de norisoprénoides 207-2, le glucoside de vomifoliol.

**Figure 44 :** Histogrammes de la distribution de l'aire de pic de l'ion avec un m/z 387.201 au temps de rétention 2.3 min détecté dans les baies de la descendance RixGw par UHPLC-MS ESI (+).

**Figure 45 :** Effet du QTL 1@56 de la carte génétique SNP\_Ri sur l'aire de pic de l'ion avec un m/z de 387.201 au temps de rétention 2.3 min lié au fragment de norisoprénoides 207-2 en 2014

**Figure 46 :** Spectre de masse au temps de rétention 6.7 min, correspondant à la molécule liée au fragment 209-7.

**Figure 47 :** Proposition de structure de la molécule liée au fragment 209-7, le malonyl-glucoside de 3-oxo- $\alpha$ -ionol.

**Figure 48 :** Histogrammes de la distribution de l'aire de pic de l'ion 457.207 au temps de rétention 6.7 minutes détecté dans les baies de la descendance RixGw par UHPLC-MS ESI (+).

**Figure 49 :** Effet du QTL 12@19.9 de la carte génétique SNP\_Ri sur l'aire de pic de la molécule de malonyl-glucoside de norisoprénoides en 2015.

**Figure 50 :** Diagramme de Venn représentant le nombre d'ions détectés par UHPLC-MS ESI (+) dans les feuilles et dans les baies prélevées en 2015.

**Figure 51 :** QTL détectés avec les données UHPLC-MS de manière non ciblée dans les feuilles prélevées en 2015 ( $p_{\text{val}} \leq 0.05$ ).



**Figure 52 :** LOD score des QTL des ions en fonction des aires de pics détectés par UHPLC-MS ESI (+) pour les ions ayant un m/z communs dans les baies et dans les feuilles 2015 ( $pval \leq 0.05$ ).

**Figure 53 :** QTL détectés sur le chromosome 5 avec la carte SNP\_Gw et les données non ciblées par UHPLC-MS ESI (+) de baies et de feuilles en 2015 ( $pval \leq 0.05$ ).

**Figure 54 :** QTL détectés sur le chromosome 3 avec la carte SNP\_Ri et les données non ciblées par UHPLC-MS ESI (+) de baies et de feuilles en 2015 ( $pval \leq 0.05$ ).

**Figure 55 :** Expression des TPS situées sur le scaffold 271 du chromosome 10 dans les baies de Ri et de Gw prélevées en 2016.

**Figure 56 :** Expression de la CAMT dans les baies de Ri et de Gw prélevées en 2016.

**Figure 57 :** Alignement des séquences en acides aminés des 3 isoformes de protéines de la CAMT présentes chez les Ri et le Gw.

**Figure 58 :** SDS-page des protéines de fusion 6His-NusA avec les 3 isoformes de CAMT.

**Figure 59 :** Observation de l'expression de la GFP à la lumière UV, après transformation transitoire de *Nicotiana benthamiana* par le vecteur pBIN-GFP.

Figure 60 : Voie de biosynthèse de la vanilline chez la vanille (*V. planifolia*) (D'après Gallage et al., 2014).

**Figure 61 :** Stratégie de validation de l'impact des QTL sur la qualité des vins par microvinification de couples de QTL.

## Liste des tableaux

**Tableau 1 :** Exemple de molécules aromatiques présentes dans les vins et issues des raisins, soit directement, soit libérée à partir d'un précurseur.

**Tableau 2 :** Liste des principaux monoterpénols détectés dans les baies.

**Tableau 3 :** Classement des cépages en fonction de la teneur des jus ou des vins en monoterpénols libres (D'après Mateo et Jiménez ,2000).

**Tableau 4 :** Structure des norisoprénoïdes les plus courants dans les vins.

**Tableau 5 :** Fragments caractéristiques des monoterpénols recherchés dans les analyses de baies 2014 et 2015 par UHPLC-MS ESI (+).

**Tableau 6 :** Fragments caractéristiques des norisoprénoïdes recherchés dans les analyses de baies de 2014 et 2015 par UHPLC-MS ESI (+).

**Tableau 7 :** Récapitulatif de la carte génétique SSR.

**Tableau 8 :** Récapitulatif de la carte génétique SNP\_Ri.

**Tableau 9 :** Récapitulatif de la carte génétique SNP\_GW.

**Tableau 10 :** Substrats testés pour les tests enzymatiques in vitro.

**Tableau 11 :** Nombre d'ions retenus par XCMS dans les échantillons de baies analysés par UHPLC-MS.

**Tableau 12 :** Top 20 des ions les plus intenses détectés dans les baies par UHPLC-MS ESI (+) dans la population RixGw en 2014.

**Tableau 13 :** Top 20 des ions les plus intenses détectés dans les baies par UHPLC-MS ESI (+) dans la population RixGw en 2015.

**Tableau 14 :** Liste des ions retrouvés à la fois le top 20 des molécules détectées en 2014 et en 2015 par UHPLC-MS ESI (+).

**Tableau 15 :** Nombre de QTL détectés et nombre d'ions correspondants détectés avec les analyses de baies par UHPLC-MS en fonction des cartes génétiques ( $pval \leq 0.05$ ).

**Tableau 16 :** QTL liés aux 12 ions les plus abondants communs en 2014 et 2015, détectés dans les baies par UHPLC-MS ESI (+) ( $pval \leq 0.05$ ).

**Tableau 17 :** Liste des ions correspondant à des monoterpénols pour lesquels un QTL significatif ( $pval \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

**Tableau 18 :** Liste des ions correspondant à des sesquiterpènes pour lesquels un QTL significatif ( $pval \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

**Tableau 19 :** Liste des ions correspondant à des diterpènes pour lesquels un QTL significatif ( $pval \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

**Tableau 20 :** Liste des ions correspondant à des norisoprénoides pour lesquels un QTL significatif ( $pval \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

**Tableau 21 :** Liste des ions correspondant à des flavonoïdes pour lesquels un QTL significatif ( $pval \leq 0.05$ ) est détecté sur le chromosome 6 avec la carte SNP\_Gw pour les deux années d'analyses.

**Tableau 22 :** Liste des ions correspondant à des norisoprénoides, monoterpénols et sesquiterpène pour lesquels un QTL significatif ( $pval \leq 0.05$ ) est détecté sur le chromosome 12 avec la carte SNP\_Ri pour les deux années d'analyses.

**Tableau 23 :** Nombre de fragments de monoterpénols détectés dans les analyse de baies par UHPLC-MS ESI (+).

**Tableau 24 :** Monoterpénols libres ou glycosidés identifiés dans les extraits de baies analysés par UHPLC-MS ESI (+).

**Tableau 25 :** Récapitulatif de la quantification des fragments de monoterpénols détectés dans les baies de la population RixGw. Les fragments ayant des aires de pics significativement différentes entre le Ri et le Gw sur les deux années sont en gras.

**Tableau 26 :** QTL détectés pour les fragments de monoterpénols détectés dans les analyses par UHPLC-MS ESI (+) reproductibles sur les deux années d'analyses de baies ( $pval \leq 0.05$ ).

**Tableau 27 :** Corrélation de la quantification des fragments avec la quantification du fragment 153-2 en 2015 (M153T253).

**Tableau 28 :** Récapitulatif de la quantification du m/z 482.260 au temps de rétention 4.2 min sur les deux années d'analyse par UHPLC-MS ESI (+) de baies du Ri, du Gw et de la population de croisement RixGw.

**Tableau 29 :** QTL détectés pour l'ion avec un m/z 482.260 au temps de rétention 4.2 min dans les baies RixGw ( $pval \leq 0.05$ ).

**Tableau 30 :** Modèle génétique expliquant le contenu en apiosyl-glucoside de linalol oxyde en 2015 avec la carte génétique SSR.

**Tableau 31 :** Modèle génétique expliquant le contenu en apiosyl-glucoside de linalol oxyde en 2015 avec la carte génétique SNP\_Ri.

**Tableau 32 :** Nombre de fragments de norisoprénoïdes détectés dans les analyse de baies par UHPLC-MS ESI(+).

**Tableau 33 :** Norisoprénoïdes identifiés à partir de fragments caractéristiques de norisoprénoïdes dans les extraits de baies analysés par UHPLC-MS ESI (+) en 2014 et 2015.

**Tableau 34 :** Récapitulatif quantification des fragments de norisoprénoïdes.

**Tableau 35 :** QTL détectés pour les fragments de norisoprénoïdes détectés dans les analyses par UHPLC-MS ESI (+) reproductibles sur les deux années d'analyses ( $p_{val} \leq 0.05$ ).

**Tableau 36 :** Corrélations entre les abondances de M207T147 en 2015 avec les ions identifiés comme appartenant au norisoprénoïde à identifier.

**Tableau 37 :** Récapitulatif de la quantification de l'ion de  $m/z = 387.201$  au temps de rétention 2.3 min sur les deux années d'analyses de baies par UHPLC-MS ESI (+).

**Tableau 38 :** QTL détectés pour l'ion 387.201 au temps de rétention 2.3 min lié au fragment de norisoprénoïde 207-2 ( $p_{val} \leq 0.05$ )

**Tableau 39 :** Modèle expliquant le contenu en pour l'ion 387.201 au temps de rétention 2.3 min lié au fragment de norisoprénoïde 207-2 avec la carte génétique SNP\_Ri sur le chromosome 1 à la position 56cM.

**Tableau 40 :** Matrice de corrélation de M209T399 en 2015 avec les ions identifiés comme appartenant à la molécule liée au fragment 209-7.

**Tableau 41 :** Récapitulatif de la quantification de l'ion de  $m/z 457.207$  au temps de rétention 6.6 min.

**Tableau 42 :** QTL détectés pour l'ion 457.207 au temps de rétention 6.6 min lié au fragment de norisoprénoïde 209-7 ( $p_{val} \leq 0.05$ ).

**Tableau 43 :** Modèle expliquant le contenu en malonyl-glucoside de norisoprénoïde en 2015 avec la carte génétique SNP\_Ri.

**Tableau 44 :** Nombre d'ions retenus par XCMS dans les échantillons de baies et de feuilles prélevés en 2015 analysés par UHPLC-MS.

**Tableau 45 :** Top 20 des ions majeurs dans les feuilles détectés par UHPLC-MS ESI (+).

**Tableau 46 :** Nombre de QTL par carte génétique, détectés dans les feuilles et dans les baies en 2015 ( $p_{val} \leq 0.05$ ).

**Tableau 47 :** Liste des ions  $[M + H]^+$  ayant un QTL sur le chr5 avec la carte SNP Gw ( $p_{val} \leq 0.05$ ) à la fois dans les baies et dans les feuilles en 2015 et analysées par UHPLC-MS ESI (+).

## Liste des tableaux

**Tableau 48 :** Ions identifiés dans les feuilles liés au QTL chromosome 3 Carte SNP\_Ri ( $pval \leq 0.05$ ).

**Tableau 49 :** QTL identifiés sur le chromosome 10 liés au linalol et à ses dérivés détectés dans les baies de la population RixGw par GC-MS et UHPLC-MS ESI (+).

**Tableau 50 :** Terpénols synthases (TPS) annotées sur le scaffold 271 du chromosome 10.

**Tableau 51 :** QTL lié au contenu en vanilline libre des baies analysé par GC-MS sur le chromosome 2.

**Tableau 52 :** Gènes candidats liés au QTL de la vanilline sur le chromosome 2.

**Tableau 53 :** SNP détectés dans les exons du gène de la CAMT.

**Tableau 54 :** Substrats potentiels testés pour étudier l'activité de la CAMT.

**Tableau 55 :** Gènes codant pour des linalol synthases ayant été caractérisées.

# **Introduction**



### 1. Contexte général

La culture de la vigne occupait dans le monde 7 449 000 ha en 2018 (OIV, 2019). La production de raisins, avec 74,3 Mt est l'une des principales productions de fruits dans le monde, derrière celle de pastèques (118,4 Mt), de bananes (113,9 Mt) et de pommes (83,1 Mt) (FAOSTAT, données 2017). Sur le plan de la valeur ajoutée en revanche, elle occupe la première place (44 265 millions \$) devant la production de pomme (37 778 millions \$) (FAOSTAT, valeurs 2016). Ces raisins sont consommés directement (36 % de la production), utilisés en raisins secs (3 % de la production) ou pour la fabrication de jus, mais l'essentiel est transformé en vins. La production mondiale de vins était de 292 millions d'hl en 2018, en stagnation sur les dix années précédentes. L'Italie occupe régulièrement la place de premier producteur mondial, en compétition avec la France et l'Espagne. Avec 792 543 ha exploités en 2018, la vigne destinée à la production de vins occupe environ 3 % de la agricole utile mais le vin contribue pour 19,3 % à la valeur de la production agricole en 2018 (INSEE, 2019). La filière viti-vinicole fournirait plus de 250 000 emplois directs (FranceAgrimer, 2019), elle est le premier secteur agroalimentaire exportateur Français, et le deuxième excédent de la balance commerciale de la France derrière l'industrie aéronautique (FEVS, 2013). La filière viticole française contribue significativement à la valeur de la production agricole française et à la richesse de certains territoires, mais elle doit répondre à plusieurs défis.

Les effets des changements climatiques se font sentir sur toute la planète et les activités agricoles, y compris la viticulture, y sont particulièrement sensibles. L'adaptation de la culture de la vigne et de la production de vins aux conditions climatiques du futur est à l'heure actuelle une des préoccupations majeures du monde viti-vinicole. L'augmentation des teneurs en CO<sub>2</sub> de l'air, des températures, et/ou une modification du régime des pluies, peuvent en effet significativement modifier le profil des vins produits, rendre la culture de la vigne impossible mais aussi offrir de nouvelles possibilités. Mieux comprendre les effets du changement climatique sur la composition des raisins pour mieux anticiper les évolutions possibles et proposer des adaptations est une des priorités actuelles des recherches scientifiques sur la vigne.

Réduire l'utilisation de produits phytosanitaires est également un objectif majeur des politiques agricoles en France, toutes productions confondues. La vigne est une des cultures les plus consommatrices de fongicides en France car elle est très sensible à deux maladies fongiques venues du continent Américain : le mildiou (*Plasmopara viticola*) et l'oïdium (*Erysiphe necator*). Des gènes de résistances à ces pathogènes sont présents dans des espèces du genre *Vitis* présentes sur le continent Américain mais aussi dans *Vitis amurensis*, une espèce



asiatique. De nombreux programmes de création variétale, dès le début du XX<sup>ème</sup> siècle, ont cherché à combiner dans les mêmes plantes des caractéristiques de résistance aux maladies fongiques aux aptitudes œnologiques de la vigne cultivée en Europe, *Vitis vinifera*. La faiblesse des surfaces cultivées en Europe sur lesquelles sont plantées des variétés résistantes aux maladies fongiques pour la production de vins témoigne de la difficulté de l'exercice. La composition des raisins de *Vitis vinifera* est très diverse, toujours complexe, mais ne présente jamais de défauts intrinsèques dans les variétés sélectionnées au cours des siècles, ce qui n'est pas le cas des espèces « sauvages », qui n'ont jamais réellement été domestiquées. Retrouver dans des hybrides interspécifiques les qualités œnologiques de *Vitis vinifera* est donc un travail de longue haleine mais possible, comme en témoigne l'inscription par l'INRA de Colmar en janvier 2018 au catalogue français des variétés de quatre variétés de vigne présentant des résistances au mildiou et à l'oïdium et aptes à la production de vins commerciaux. Les programmes de création variétale actuels visent à proposer aux grands ensembles viticoles français des variétés de vigne résistantes aux maladies fongiques mais capables de produire des vins correspondants à la typicité locale attendue.

La sélection de génotypes possédant des loci de résistances utilise des marqueurs moléculaires issus de programmes de recherches antérieurs (Merdinoglu et al., 2018). Disposer de marqueurs moléculaires permettant d'identifier dès le stade « plantule » des plantes dont les raisins auraient des caractéristiques précises conduirait soit à diminuer les effectifs de génotypes à tester au vignoble, soit, à effectif au vignoble constant, élargir les possibilités de sélection sur d'autres critères. À la fois pour appuyer la création variétale et pour anticiper les effets du changement climatique, il est nécessaire de mieux connaître les déterminants génétiques et moléculaires de la composition des raisins.

## 2. La composition des raisins

### 2.1. Sucrosité et acidité

On peut distinguer quatre grands types de vins produits directement à partir des raisins: les vins effervescents, les vins blancs secs, les vins blancs liquoreux et les vins rouges. D'autres types de boissons, comme le Cognac ou les vins doux naturels résultent de transformations ultérieures à partir de vins de base.

La teneur en alcool et l'acidité sont des paramètres essentiels communs à tous les vins. L'alcool des vins est issu de la transformation des sucres présents dans les raisins par des levures, essentiellement représentées par *Saccharomyces cerevisiae*. Le rendement de transformation des levures est variable mais une norme fixe le chiffre de 16,83 g.l<sup>-1</sup> de sucres dans un moût pour obtenir 1 % v/v d'alcool dans un vin (Ribéreau-Gayon et al., 2006).

Les sucres des raisins sont, pour l'espèce *Vitis vinifera*, pour moitié du fructose et pour moitié du glucose. Glucose et fructose sont produits par hydrolyse par des invertases du saccharose importé du phloème (Davies et al., 2012). Ces hexoses, stockés dans les vacuoles de la pulpe des baies, s'accumulent rapidement à partir de la véraison. Les concentrations en sucres mesurées à la récolte dépendent des variétés, des conditions climatiques de l'année et du rapport fruit/feuille des plants de vigne: pour une surface foliaire donnée, les concentrations seront d'autant plus faibles que la quantité de raisins sera élevée (Jackson and Lombard, 1993). Pour donner des ordres de grandeur, la teneur en sucres recherchée à la récolte pour des raisins destinés à l'élaboration de vins effervescents sera de l'ordre de 170 g.l<sup>-1</sup>. Pour un vin liquoreux, cette teneur pourra atteindre 400 g.l<sup>-1</sup>.

L'acidité est un paramètre sensoriel essentiel: trop élevée les vins vont être jugés verts et agressifs, trop faible, les vins vont apparaître plats et mous. L'acidité participe également à la stabilité microbiologique des vins et à leur potentiel de vieillissement.

L'acidité des vins peut être caractérisée par une mesure de pH ou de l'acidité de titration. pH et acidité de titration dépendent des concentrations en acides organiques et des teneurs en cations des vins. La gamme de pH observés va de pH 3,1 pour un vin blanc sec de type Riesling à pH 3,9 pour un vin liquoreux de Gewurztraminer. L'acidité sera d'autant plus élevée que les teneurs en acides organiques seront fortes et que les teneurs en cations seront basses. Les principaux acides organiques des raisins sont l'acide malique et l'acide tartrique, qui représentent à eux seuls plus de 90 % des acides des raisins (Ruffner, 1982). En fonction du

type de vin recherché, l'acide malique sera ou non transformé en acide lactique au cours d'une fermentation malolactique réalisée par la bactérie *Oenococcus oeni*.

Les quantités d'acide tartrique par baie sont considérées comme constantes à partir de la véraison (DeBolt et al., 2008) alors que l'acide malique est dégradé au cours de la maturation, d'autant plus rapidement que les températures sont élevées (Sweetman et al., 2014).

Pour les deux acides les concentrations baissent au cours de la maturation par un phénomène de dilution consécutif à l'augmentation du volume des baies.

Le potassium est le principal cation modulant le pH des raisins (Ribéreau-Gayon et al., 2006). En neutralisant les acides organiques, il conduit à une augmentation du pH. Les concentrations mesurées dans vins ou les raisins à la récolte sont de l'ordre de 1 g.l<sup>-1</sup> (Ribéreau-Gayon et al., 2006) mais peuvent dépasser 2 g.l<sup>-1</sup> (Kodur et al. 2013).

### 2.2. Métabolisme secondaire

Les métabolites secondaires des raisins jouent un rôle central dans les qualités organoleptiques des vins car ils sont à la base de leur couleur, de leurs arômes et de leur structure. Les anthocyanes sont des composés phénoliques synthétisés, sauf pour quelques variétés, seulement dans les pellicules des raisins. Ces molécules très diverses sont à l'origine de la coloration rouge des raisins (Castellarin et al. 2014). Les polymères de flavan-3-ols, ou tannins condensés, sont également des composés phénoliques surtout présents dans les pépins, ils contribuent beaucoup aux sensations en bouche procurées par les vins (Castellarin et al. 2014). En excès, ils procurent une sensation d'astringence.

Les arômes des vins sont dus à des molécules volatiles, soit déjà présentes en l'état dans les raisins, soit issues de la transformation de précurseurs lors des processus de fermentation, soit produites par les levures, soit apparaissant au cours du vieillissement. Des composés terpéniques (monoterpénols en C10) sont à l'origine d'arômes floraux de rose, muguet ou tilleul. Ils sont présents dans les raisins en l'état sous forme libre ou liés à des glucosides. Les formes liées ne sont pas volatiles mais sont hydrolysées au cours du vieillissement du vin, ce qui libère des molécules volatiles. Les C13-norisoprénoïdes sont dérivés de la dégradation des caroténoïdes, qui sont des tétraterpènes (C40). On trouve dans cette famille, parmi d'autres, une molécule aux arômes de violette (la  $\beta$ -ionone), mais aussi une molécule comme le 1,1,6-triméthyl-1,2-dihydronaphtalène (TDN) dont l'odeur rappelle celle du kérosène.

Les méthoxypyrazines sont des molécules hétérocycliques azotées également présentes dans les raisins. Elles ont en général un impact négatif sur la perception aromatique des vins, avec des arômes herbacés de poivron ou de petit pois, mais peuvent participer à la typicité des vins de Sauvignon Blanc par exemple.

Les thiols volatils sont des composés caractérisés par la présence d'un groupement –SH. Ces molécules ont des seuils de perception très bas et sont présentes dans de nombreux fruits. Les descripteurs leur correspondant vont être « pamplemousse », « zeste de citron », « fruit de la passion », « bourgeon de cassis ». Ces molécules apparaissent seulement au cours de la fermentation alcoolique. En effet, elles sont présentes dans les raisins sous la forme de précurseurs S-conjugués.

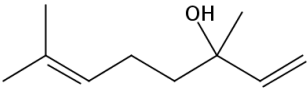
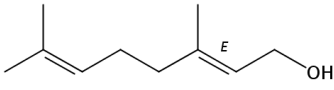
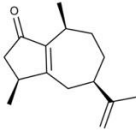
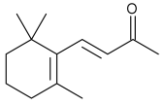
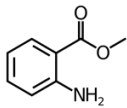
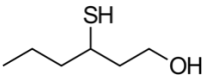
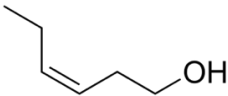
Ce bref panorama vise simplement à illustrer l'extrême diversité des métabolites secondaires de la vigne, à l'origine de la richesse des sensations organoleptiques qu'ils procurent dans les vins. Le profil de métabolites secondaires trouvés dans les vins dépend très largement de la variété qui a servi à son élaboration mais le niveau de connaissance des déterminants génétiques à l'origine de ces différences est très variable selon la famille chimique considérée.

### 3. Les déterminants génétiques et moléculaires de la biosynthèse des arômes des raisins

Les arômes constituent un caractère déterminant de la qualité des vins et jouent un rôle majeur dans les impressions ressenties lors de la dégustation. Les arômes du vin sont dus à un mélange complexe de molécules volatiles, dont certaines sont directement présentes dans les raisins et d'autres sont formées au cours de la fermentation puis du vieillissement du vin. Dans la suite de cette introduction, nous nous focaliserons sur les arômes et leurs précurseurs synthétisés dans les raisins et au déterminisme génétique de leur biosynthèse. Parmi les molécules aromatiques majeures issues des baies de raisin, on peut citer les monoterpènes, les sesquiterpènes, les norisoprénoïdes, les pyrazines, les thiols volatils, les dérivés des produits de dégradation des lipoxygénases. Chaque composé aromatique est caractérisé par un seuil de détection olfactive, correspondant à la concentration minimale (dans l'eau ou dans le vin) perceptible par l'odorat humain. Selon les composés, le seuil de détection olfactive peut varier de plusieurs ordres de grandeur, de quelques ppb à plusieurs dizaines de ppm. Par conséquent, certains composés peuvent avoir des impacts aromatiques majeurs, malgré leur présence en très

faibles concentrations dans les raisins et les vins. Quelques exemples illustrant les différentes familles de molécules aromatiques des vins sont présentés dans le Tableau 1.

Dans ce chapitre, j'ai choisi de me focaliser sur les composés aromatiques des raisins, dont la biosynthèse a été étudiée par des approches de génétique ou de génomique fonctionnelle. Le déterminisme de leur biosynthèse chez la vigne ainsi que les gènes impliqués seront détaillés.

Famille chimique	Composé	Structure chimique	Descripteurs	Seuil de détection olfactive (eau ou vin)
Monoterpène	Linalol		floral	5 ppm
Monoterpène	Géraniol		rose	1 ppm
Sesquiterpène	rotundone		poivré	16 ppb
Norisoprénoïde	$\beta$ -ionone		violette	0,5 ppm
Pyrazine	IBMP		poivron, pois	5 ppb
Thiol	3-mercaptohexanol		pamplemousse, agrumes	60 ppb
Produit de dégradation des lipides	cis-hexenol		herbe, foin	5 ppm

**Tableau 1 : Exemple de molécules aromatiques présentes dans les vins et issues des raisins, soit directement, soit libérée à partir d'un précurseur** (Source : <https://www.vignevin-occitanie.com/fiches-pratiques/descripteurs-sensoriels-et-references-chimiques/>).

### 3.1. Les terpènes

Les terpènes, ou isoprénoïdes, sont des molécules issues de la polymérisation d'unités isoprène ( $C_5H_8$ ) sous la forme d'isopentényl diphosphate (IPP), le précurseur universel de la synthèse des terpènes. Selon le nombre d'isoprènes polymérisés, on obtient des monoterpènes (2 unités,  $C_{10}$ ), sesquiterpènes (3 unités,  $C_{15}$ ), diterpènes (4 unités,  $C_{20}$ ), et plus. Chez les plantes, les terpènes sont produits via la voie cytosolique du mévalonate (MVA Pathway), dite également voie de l'HMG-CoA réductase, et par la voie plastidiale du méthylérythritol phosphate (MEP pathway) (Figure 1). Ces deux voies métaboliques donnent naissance à une extraordinaire diversité de molécules impliquées dans tous les aspects de la vie des plantes: constituants des membranes (stéroïls), photorécepteurs (caroténoïdes, chaîne latérale de la chlorophylle), transporteurs d'électrons (plastoquinone), hormones végétales (gibbérellines, acide abscissique), molécules de défenses (phytoalexines terpéniques), polymères (latex), composés volatils organiques (isoprène, terpènes volatiles des plantes aromatiques et des parfums floraux). Dans la suite de cette introduction, nous nous focaliserons sur les monoterpènes et les sesquiterpènes, car les molécules de plus haut poids moléculaire sont peu volatiles et ne participent pas directement aux arômes et parfums (Croteau and Karp, 1994). Nous évoquerons également les norisoprénoïdes issus de la dégradation des caroténoïdes, qui constituent des arômes importants des vins.

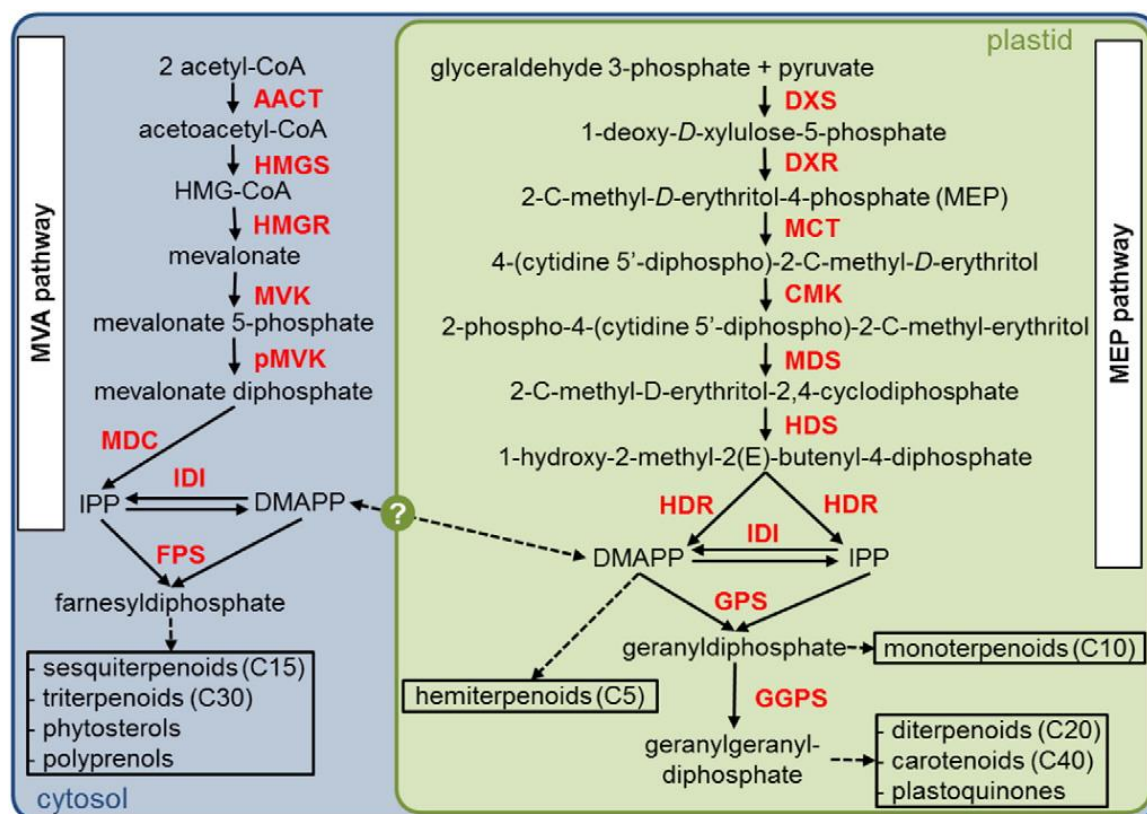


Figure 1 : Voies de biosynthèse des terpènes chez les plantes (d'après Champagne and Boutry, 2016).

MVA Pathway : voie du mévalonate, MEP pathway : voie du méthylérythritol phosphate, AACT, acetoacetyl-coenzyme A thiolase; HMGS, 3-hydroxy-3-méthylglutaryl (HMG)-CoA synthase; HMGR, HMG-CoA réductase; MVK, mevalonate kinase; pMVK, phosphomevalonate kinase; MDC, mevalonate diphosphate décarboxylase; DXS, 1-deoxyxylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate réductoisomérase; MCT, 2-C-méthyl-D-erythritol 4-phosphate cytidyltransferase; CMK, 4-diphosphocytidyl-2-C-méthyl-D-erythritol kinase; MDS, 2C-méthyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-méthylbut-2-enyl diphosphate synthase; HDR, 4-hydroxy-3-méthylbut-2-enyl diphosphate réductase; IPP, isopentényl diphosphate; DMAPP, diméthylallyl diphosphate; IDI, IPP/ DMAPP isomérase; GPS, geranyl diphosphate synthase; GGPS, geranylgeranyl diphosphate synthase; FPS, farnesyl diphosphate synthase.

### 3.1.1. Les monoterpènes des raisins

Les monoterpènes sont des molécules à 10 atomes de carbone dérivées du précurseur géranyl diphosphate (Tableau 2). Chez la vigne, ils sont très majoritairement présents sous la forme d'alcools (monoterpénols) souvent associées à des parfums floraux ou d'agrumes, si bien que nous nous focaliserons essentiellement sur ces alcools dans ce paragraphe. Les monoterpénols sont accumulés en grandes quantités dans les baies des cépages muscatés, tels que les variétés de la famille des Muscats et le Gewurztraminer (Mateo and Jiménez, 2000). Plus de 70 monoterpènes différents ont été identifiés dans les raisins et le vin, dont une douzaine de dérivés du linalol. Chez la plupart des cépages, les monoterpènes sont synthétisés au début du développement des baies, puis les niveaux baissent en dessous du seuil de détection après la

véraison, si bien qu'ils ne contribuent pas significativement à l'arôme des baies mûres (Kalua and Boss, 2009, 2010; Robinson et al., 2014). Les baies des cépages muscatés représentent une exception majeure, car leurs arômes floraux caractéristiques sont dus à une accumulation massive de monoterpénols, les plus importants étant le linalol, le géraniol, le nérol, l' $\alpha$ -terpinéol et le hotriéol (Tableau 2).

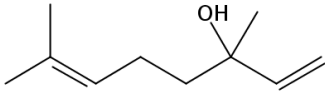
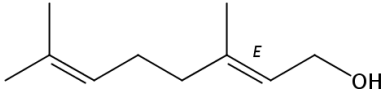
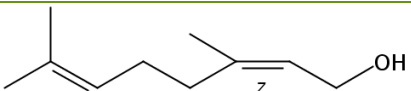
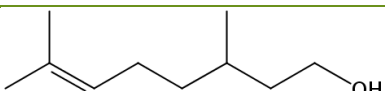
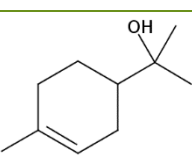

Monoterpénol	Structure chimique	Descripteurs	Seuil de détection olfactive (eau ou vin)
Linalool		floral	5 ppm
Géraniol		rose	1 ppm
Nérol		rose	4 ppm
Citronellol		citronelle	2
$\alpha$ -terpinéol		floral	20 ppm
Hotriéol		tilleuil	

Tableau 2 : Liste des principaux monoterpénols détectés dans les baies.

Les cépages peuvent être divisés en trois groupes généraux en fonction de la concentration totale en monoterpénols libres dans les jus ou les vins, ainsi que par leur profil monoterpénique (Tableau 3). On distingue ainsi les variétés neutres, qui ont de très faibles concentrations, telle que le Chardonnay, les cépages « aromatiques non muscatés » avec 1-4 mg.l<sup>-1</sup> de monoterpénols, tel que le Riesling et les cépages « aromatiques muscatés » comme le Muscat blanc, le Muscat d'Alexandrie ou le Gewurztraminer avec jusqu'à 6 mg.l<sup>-1</sup> de



monoterpénols libres (Mateo and Jiménez, 2000). Dans l'ensemble, les monoterpénols sont principalement concentrés dans la peau des raisins (pellicule) avec une distribution différentielle selon le composé. Le géraniol et le nérol sont beaucoup plus abondants dans la peau que dans la chair, tandis que le linalol est plus généralement réparti dans le fruit (Gunata et al., 1985).

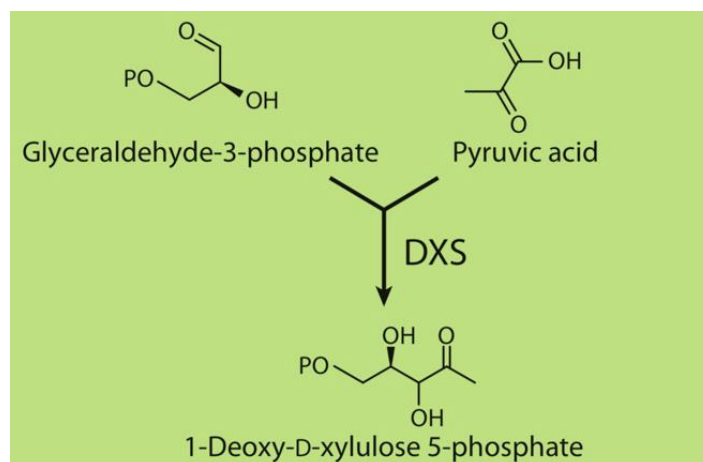
Cépages aromatiques muscatés 4 mg.l <sup>-1</sup> < Monoterpénols < 6 mg.l <sup>-1</sup>	Cépages aromatiques non muscatés 1 mg.l <sup>-1</sup> < Monoterpénols < 4 mg.l <sup>-1</sup>	Cépages neutres Monoterpénols < 1 mg.l <sup>-1</sup>
Gewurztraminer	Müller-Thurgau	Cabernet sauvignon
Muscat d'Alexandrie	Riesling	Chardonnay
Muscat de Frontignan	Scheurebe	Chasselas
Muscat de Hambourg	Sylvaner	Merlot
Muscat Ottonel	Traminer (Savagnin rose)	Syrah

**Tableau 3 : Classement des cépages en fonction de la teneur des jus ou des vins en monoterpénols libres, D'après Mateo et Jiménez, 2000).**

### *3.1.2. Identification d'allèles du gène codant pour la 1-déoxy-D-xylulose 5-phosphate synthase (DXS) responsable du caractère aromatique de certains cépages*

L'analyse du déterminisme génétique de la biosynthèse des monoterpènes des raisins a fait l'objet de nombreux travaux. Une première analyse génétique d'un croisement entre un cépage neutre et le Muscat de Hambourg a permis d'identifier un locus sur le groupe de liaison (LG) 5 associé à l'accumulation de linalol, nérol et géraniol (Doligez et al., 2006). Ce QTL associé à des niveaux élevés de monoterpénols et localisé sur le LG 5 a été confirmé indépendamment par deux équipes de recherche dans une population Muscat × cépage neutre (Battilana et al., 2009) et deux populations issues d'autofécondations de cépages aromatiques (Muscat Ottonel et Gewurztraminer) (Duchêne et al., 2009). L'analyse des gènes présents dans les intervalles de confiance de ces QTL a permis de proposer que le gène codant pour l'enzyme 1-déoxy-D-xylulose 5-phosphate synthase (DXS) pourrait être responsable de l'accumulation massive de monoterpènes dans les cépages aromatiques (Battilana et al., 2009; Duchêne et al., 2009a). La DXS est la première enzyme de la voie du MEP (Figure 1). Elle catalyse la formation du 1-déoxy-D-xylulose 5-phosphate (DXP) à partir du glycéraldéhyde-3-phosphate et du pyruvate (Figure 2). Cette enzyme est connue pour jouer un rôle majeur dans la régulation

de la biosynthèse des terpènes d'origine plastidiale chez les plantes, en raison du caractère limitant de son activité au sein de cette voie de métabolique (Estévez et al., 2001; Mandel et al., 1996; Wright et al., 2014)

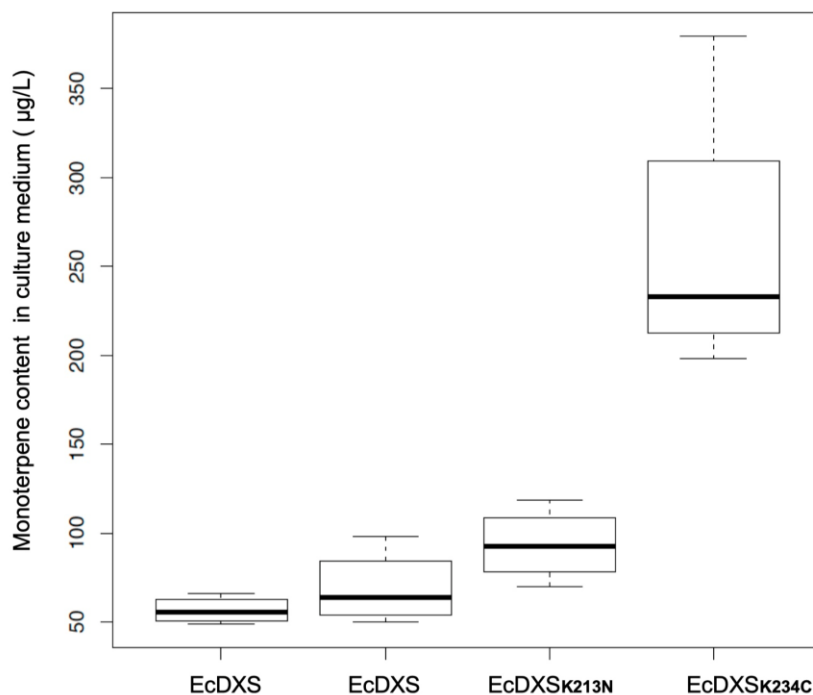


**Figure 2 : Biosynthèse du 1-déoxy-D-xylulose 5-phosphate (DXP) catalysée par l'enzyme DXS.**

Une étude de génétique d'association à grande échelle a été réalisées sur un panel de 148 cépages muscatés, aromatiques ou neutres. L'analyse de la séquence du gène *VvDXS* a révélé 95 polymorphismes nucléotidiques (single-nucleotide polymorphism, SNP), dont quatre aboutissent à des changement d'acides aminés qui modifient potentiellement la fonction de la protéine (Emanuelli et al., 2010). Parmi ces SNP, le plus fréquemment lié au caractère aromatique est le SNP G/T qui entraîne le remplacement de la lysine (K) en position 284 par de l'asparagine (N) (K284N). Au sein du panel, 68 des 69 variétés de Muscat analysées possèdent l'allèle *VvDXS*<sub>K284N</sub> à l'état hétérozygote. En revanche, le Gewurztraminer, également très aromatique, ne présente pas l'allèle *VvDXS*<sub>K284N</sub>, mais possède un allèle *VvDXS*<sub>R306C</sub> à l'état hétérozygote, dans lequel un SNP C/T entraîne le remplacement de l'arginine en position 306 par une cystéine (R306C) (Figure 3).



**Figure 3 : Comparaison des séquences protéiques des allèles VvDXS (allèle neutre), VvDXSK284N des Muscats et VvDXSR306C du Gewurztraminer (vue partielle des protéines d'après Huguenev et al., 2012).**



**Figure 4 : Introduction des mutations de la DXS de vigne, associées au caractère aromatique dans la protéine DXS d'*E. coli* (d'après Huguenev et al., 2012).**

La protéine originale (EcDXS) et les protéines DXS mutantes EcDXSK<sub>213N</sub> (correspondant à VvDXSK<sub>284N</sub>) et EcDXSK<sub>234C</sub> (correspondant à VvDXSR<sub>306C</sub>) ont été exprimées chez *E. coli* et les monoterpènes libres accumulés dans le milieu de culture ont été quantifiés.

La caractérisation des différents SNP étroitement liés au caractère aromatique a permis le développement de marqueurs génétiques utilisables pour sélectionner efficacement ce caractère dans des programmes de création variétale (Emanuelli et al., 2014). Néanmoins, plusieurs équipes de recherches ont poursuivi l'étude des allèles VvDXSK<sub>284N</sub> et VvDXSR<sub>306C</sub>, dans le but de mieux comprendre les bases moléculaires du caractère aromatique chez la vigne.

Afin d'analyser l'impact de la mutation K284N sur l'activité de la DXS, les allèles VvDXSK<sub>284</sub> (allèle neutre) et VvDXSK<sub>284N</sub> ont été exprimés dans *Escherichia coli*. L'étude des paramètres enzymatiques des protéines recombinantes a montré que la protéine correspondant à l'allèle VvDXSK<sub>284N</sub> possédait une efficacité catalytique (K<sub>cat</sub>) deux fois supérieure à celle de la protéine VvDXSK<sub>284</sub>, sans modification significative de l'affinité pour les substrats. De plus, la transformation des deux allèles dans des tabacs transgéniques a permis de montrer que l'expression de l'allèle VvDXSK<sub>284N</sub> aboutit à une accumulation de dérivés monoterpénoïdes dans les feuilles de tabac cinq à dix fois supérieure à celle obtenue après expression de l'allèle

neutre *VvDXS*<sub>K284</sub> (Battilana et al., 2011). Enfin, l'expression de l'allèle *VvDXS*<sub>K284N</sub> dans la variété de vigne naine « Microvine » entraîne une accumulation de monoterpénols dans les baies similaires à celle d'une variété muscatée (Dalla Costa et al., 2018). Le caractère aromatique des cépages de la famille des Muscats semble donc lié à l'efficacité catalytique supérieure de l'allèle *VvDXS*<sub>K284N</sub>. Ce surcroît d'activité de l'enzyme, limitante de la voie du MEP, aboutit à une accumulation massive de monoterpénols responsables des arômes floraux caractéristiques des raisins et des vins des Muscats.

Par ailleurs, la caractérisation d'allèles *VvDXS* responsables de l'accumulation de monoterpènes dans les cépages aromatiques a permis de développer une stratégie d'optimisation de la synthèse de terpènes dans les systèmes procaryotiques. En effet, les procaryotes utilisent également la voie du MEP pour la biosynthèse des terpènes et c'est d'ailleurs chez *E. coli* et d'autres bactéries que cette voie métabolique a été découverte dans les années 1990 (Rohmer et al., 1993). L'introduction dans la protéine DXS d'*E. coli* (*EcDXS*) des mutations K284N caractéristique du Muscat et R306C caractéristique du Gewurztraminer aux positions correspondantes a permis d'obtenir les variants *EcDXS*<sub>K213N</sub> et *EcDXS*<sub>R306C</sub>. L'expression de ces variants chez *E. coli* permet d'augmenter très significativement la synthèse des terpènes chez cette bactérie, en particulier en ce qui concerne *EcDXS*<sub>R306C</sub> (Figure 4, Hugueney et al., 2012). Cette stratégie d'optimisation de l'enzyme DXS procaryotique pour l'obtention de terpènes d'intérêt par voie biotechnologique est actuellement utilisée par la société Déinove pour la production de molécules terpéniques à haute valeur ajoutée.

### *3.1.3. Caractérisation de terpène synthases (TPS) candidates pour la synthèse des monoterpènes des raisins*

Le séquençage du génome de la vigne homozygote PN40024, dérivée du Pinot Noir par de multiples cycles d'autofécondation, a révélé la présence d'une vaste famille d'une centaine de gènes codant pour des terpène synthases (TPS) putatives, ce qui représente le double du nombre de gènes TPS identifiés chez *Arabidopsis* (Jaillon et al., 2007). En effet, parmi les plantes dont le génome a été séquençé, la vigne possède la seconde famille TPS la plus nombreuse, juste derrière l'Eucalyptus (Myburg et al., 2014). Une analyse détaillée de la famille TPS de la vigne PN40024 a permis d'identifier 69 gènes potentiellement fonctionnels, les gènes restants étant constitué de pseudogènes (Martin et al., 2010). Dix-sept gènes TPS ont été caractérisés comme codant pour des monoterpène synthases fonctionnelles, après expression

des protéines correspondantes chez *E. coli*. Après incubation en présence de GPP, plupart des monoterpène synthases caractérisées catalysent la synthèse de mélanges plus ou moins complexes de monoterpènes, le nom de la TPS correspondant au produit majoritaire. Par exemple, la pinène synthase VvTPS44 produit 50% de (+)- $\alpha$ -pinène, mais également du (+)-limonène (25%) et une dizaine d'autres monoterpènes. Néanmoins, quelques monoterpène synthases synthétisent un produit unique, comme c'est le cas des linalol synthases VvTPS54 et VvTPS56 ou bien de la géraniol synthase VvTPS52. De plus, les enzymes VvTPS54 et VvTPS56 qui produisent du linalol en présence de GPP sont également capables de catalyser la synthèse de nérolidol en présence de FPP.

Avant même la disponibilité de la séquence du génome de la vigne, un ADNc codant pour une TPS exprimée dans les raisins de Gewurztraminer a été cloné et l'activité de l'enzyme correspondante a été caractérisée comme  $\alpha$ -terpinéol synthase, après expression de la protéine chez *E. coli* et analyse de l'activité *in vitro* (Martin and Bohlmann, 2004). Plus récemment, une étude de la diversité du gène codant pour l' $\alpha$ -terpinéol synthase dans un panel d'une soixantaine de cépages a révélé que deux SNP dans des régions codantes du étaient corrélés à des concentrations plus élevées d' $\alpha$ -terpinéol chez les cépages correspondants, mais l'impact fonctionnel potentiel de ces SNP n'ait pas été étudié.

La biosynthèse de géraniol et de linalol qui s'accumulent massivement dans les cépages aromatiques a fait l'objet de nombreux travaux. Des analyses transcriptomiques ont montré que l'accumulation de monoterpénols dans les baies était corrélée à l'expression de la linalol synthase VvTPS56 dans les baies de Gewurztraminer (Martin et al., 2012) ou de la géraniol synthase VvTPS52 dans les baies de la variété de Muscat « Moscato bianco » (Matarese et al., 2013). Par ailleurs, l'analyse génétique de la biosynthèse des monoterpénols a identifié des QTL associés à l'accumulation de linalol sur le LG 10 (Battilana et al., 2009; Duchêne et al., 2017). La richesse particulière en linalol des baies de Muscat serait donc due à la combinaison d'allèles favorable de la DXS (LG5) et d'une linalol synthase putative (LG10) (Duchêne et al., 2009a). Cependant, l'assemblage de la séquence du génome de référence de la vigne PN40024 est problématique dans certaines régions du chromosome 10 et cette linalol synthase putative n'a pas encore été identifiée. En résumé, même si des corrélations ont été établies entre l'expression de certains gènes codant pour des TPS et l'accumulation de monoterpènes dans les baies, et malgré la caractérisation fonctionnelle de plusieurs monoterpène synthases *in vitro* ou en système hétérologue, il n'existe actuellement pas de preuve génétique formelle de

l'implication d'un gène TPS particulier dans la variabilité génétique de la biosynthèse de monoterpènes des raisins.

### *3.1.4. Caractérisation d'une enzyme à cytochrome P450 impliquée dans la biosynthèse de la wine lactone, un arôme caractéristique des vins âgés*

Les raisins et les vins contiennent de nombreux dérivés oxydés du linalol (Ilc et al., 2016). Parmi les produits issus de l'oxydation du linalol, la wine lactone est en général peu abondante dans les vins ( $100 \mu\text{g.l}^{-1}$ ), mais elle joue un rôle important dans l'arôme des vins de Gewurztraminer, en raison de son seuil de détection olfactive très bas ( $10 \text{ ng.l}^{-1}$ ) (Guth, 1997). La wine lactone est une monoterpène lactone bicyclique qui possède un arôme agréable qualifié de sucré et boisé, rappelant la noix de coco. La wine lactone provient de la cyclisation en milieu acide du (E)-8-carboxylinalol, qui provient lui-même de l'oxydation du linalol via le (E)-8-hydroxylinalol (Figure 5) (Luan et al., 2006). Il a été montré que des enzymes à cytochrome P450 appartenant à la famille CYP76 étaient impliqués dans le métabolisme oxydatif du linalol dans les fleurs d'*Arabidopsis* (Ginglinger et al., 2013), ce qui en fait de bonnes candidates pour l'oxydation du linalol chez la vigne. Une analyse du transcriptome des raisins de Gewurztraminer a montré que cinq gènes codant pour des cytochromes P450 appartenant à la famille CYP76 étaient exprimés au cours de la maturation, l'expression du gène *CYP76F14* étant la plus forte. L'analyse fonctionnelle de ces gènes candidats a montré que seule l'enzyme CYP76F14 était capable d'oxyder efficacement le linalol en (E)-8-hydroxylinalol, puis en (E)-8-carboxylinalol (Ilc et al., 2017).

Parallèlement, l'analyse génétique de l'accumulation du (E)-8-carboxylinalol dans une population issue du croisement Riesling x Gewurztraminer a montré que cette dernière est associée à trois QTL majeurs: l'un contenant *VvDXS*, un QTL co-localisant avec plusieurs terpènes synthases localisées sur le chromosome 10 et probablement responsables de la synthèse de linalol, et enfin un QTL sur le chromosome 2 dont l'intervalle de confiance contient le gène *CYP76F14*, ce qui confirme son rôle majeur dans la synthèse du (E)-8-carboxylinalol (Ilc et al., 2017). La richesse particulière des vins de Gewurztraminer en wine lactone s'explique sans doute par leur caractère aromatique lié à l'accumulation massive de monoterpènes comme le géraniol et le linalol, ainsi que par la présence d'allèles favorables du gène codant pour CYP76F14. Cette combinaison de caractères permet l'accumulation de glucoside de (E)-8-carboxylinalol dans les raisins, probablement formé grâce à l'activité d'une UDP-glucosyltransférase non identifiée. La wine lactone formée au cours du vieillissement du

vin provient alors de la cyclisation lente en milieu acide du (E)-8-carboxylinalol libéré à partir de son glucoside pendant la fermentation.

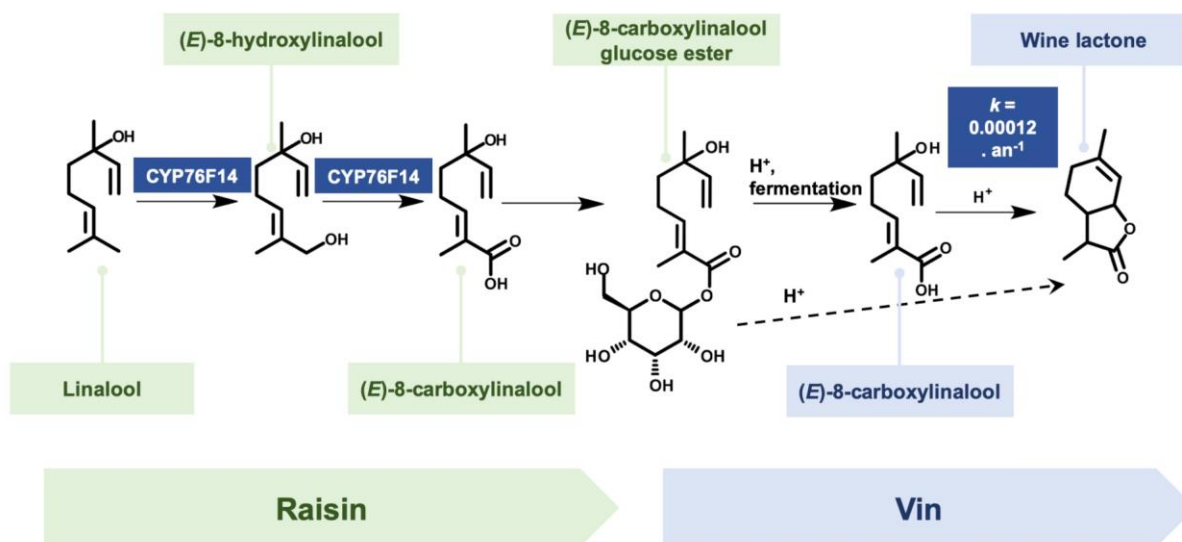


Figure 5 : Modèle de la formation de la wine lactone au cours du vieillissement du vin, à partir de précurseurs accumulés dans les raisins (d'après Ilc *et al.*, 2017).

D'après le modèle proposé par Ilc *et al.* (2017), la wine lactone provient de la cyclisation lente en milieu acide du (E)-8-carboxylinalol au cours du vieillissement du vin. Le (E)-8-carboxylinalol provient lui-même de l'hydrolyse du glucose ester correspondant pendant la fermentation. Le (E)-8-carboxylinalol-glucose ester est formé à partir du linalol, après des réactions d'oxydation catalysées par l'enzyme CYP76F14 et une glycosylation catalysée par une glucosyltransférase non caractérisée.

## 3.2. Les sesquiterpènes des raisins

### 3.2.1. Origine des sesquiterpènes

Les sesquiterpènes sont issus de l'activité de sesquiterpène synthases sur le précurseur farnésyl diphosphate (FPP), qui est synthétisé à partir d'une molécule de DMAPP et de deux molécules d'IPP, majoritairement issues de la voie cytosolique du mévalonate (Figure 1). Néanmoins, dans les raisins, du DMAPP et IPP dérivés de la voie du MEP sont également incorporés aux sesquiterpènes (May *et al.*, 2013). Une cinquantaine de dérivés de

sesquiterpènes ont été identifiés dans les raisins, principalement des hydrocarbures et quelques cétones, oxydes et alcools. Leur impact sur les arômes des raisins et des vins est cependant moins bien connu que celui des monoterpènes (Petronilho et al., 2014). En revanche, les sesquiterpènes sont des terpènes majeurs émis par les fleurs de vigne, en particulier au niveau des anthères pendant le développement du pollen, probablement comme composés de défense contre les herbivores. Chez le Cabernet Sauvignon, la valencène synthase exprimée dans les anthères est responsable de la biosynthèse du (+)-valencène et du (-)-7-epi- $\alpha$ -selinène, qui sont les sesquiterpènes majeurs des fleurs (Martin et al., 2009).

L'analyse du génome de référence de la vigne PN40024 a révélé une famille d'une trentaine de sesquiterpène synthases putatives, dont 13 ont été caractérisées après expression des protéines correspondantes dans des souches d'*E. coli* accumulant du FPP (Martin et al., 2010). Cependant, la participation de ces gènes à la biosynthèse de sesquiterpènes des raisins n'a pas été démontrée. L'essentiel des travaux sur les sesquiterpènes des raisins s'est concentré sur la rotundone, qui contribue aux arômes caractéristiques des vins issus de la Syrah.

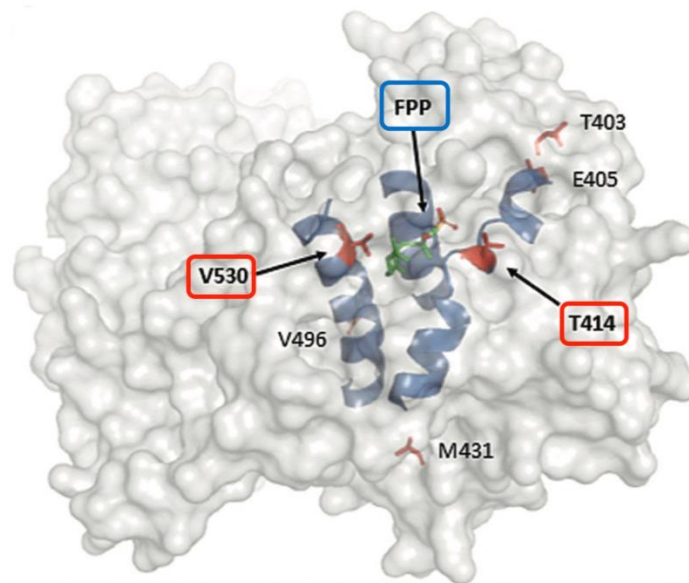
### 3.2.2. Caractérisation de la voie de biosynthèse de la rotundone

La rotundone est une cétone sesquiterpénique qui a été identifiée relativement récemment comme responsable des arômes poivrés caractéristiques de certains cépages, dont la Syrah est le plus connu (Wood et al., 2008). La rotundone possède un arôme puissant et un seuil de détection olfactif extrêmement bas de 16 ng.l<sup>-1</sup> dans le vin rouge. Il semble néanmoins qu'environ 20 % de la population soit anosmique pour la rotundone, c'est-à-dire incapable de percevoir cet arôme (Wood et al., 2008).

Des études *in vitro* ont montré que la rotundone pouvait être obtenue par oxydation du sesquiterpène  $\alpha$ -guaiène, qui pourrait constituer le précurseur de la rotundone *in vivo* (Huang et al., 2015). Cependant, la caractérisation fonctionnelle de sesquiterpène synthases de vigne n'a pas permis d'identifier d'enzyme catalysant la synthèse de quantités significatives d' $\alpha$ -guaiène. Seule l'enzyme VvTPS24 sélinène synthase de PN40024 (VvSelS) est capable de former de l' $\alpha$ -guaiène en très faibles quantités (3,5% des sesquiterpènes produits), en plus de son produit majeur, le sélinène (Martin et al., 2010). Le clonage d'un allèle du gène *VvTPS24* de la Syrah a montré qu'il se différencie de l'allèle de PN40024 par six acides aminés (Drew et al., 2016). L'expression transitoire de l'allèle *VvTPS24* de Syrah dans le tabac *Nicotiana benthamiana* a révélé que la protéine correspondante possédait une activité  $\alpha$ -guaiène synthase très



significative, puisque ce sesquiterpène représentait 45% des produits de l'enzyme, appelée par conséquent VvGuaS. Par des expériences de mutagenèse dirigée complétées par une approche de modélisation moléculaire, Drew et al., (2016) ont également montré que l'activité particulière de la protéine VvGuaS était principalement due à deux acides aminés qui la différenciaient de la sélénène synthase de PN40024. Ces deux acides aminés sont situés au voisinage du site actif de l'enzyme et leur position permet de proposer qu'ils interagissent avec le substrat FPP (Figure 6).

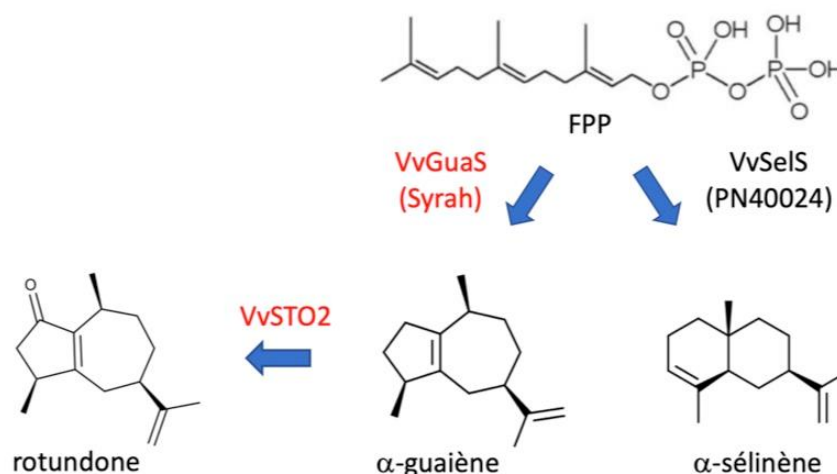


**Figure 6 : Modèle moléculaire de l'allèle TPS24  $\alpha$ -guaiène synthase (VvGuaS) de la Syrah.**

Les six acides aminés (T403, E405, T414, M431, V496, V530) qui différencient l'allèle TPS24 VvGuaS de l'allèle TPS24 VvSelS de PN40024 sont indiqués. Les acides aminés T414 et V530, situés au voisinage du site actif et responsables de l'activité  $\alpha$ -guaiène synthase sont encadrés, de même que le substrat FPP.

La transformation de l' $\alpha$ -guaiène en rotundone nécessite une réaction d'oxydation qui peut être réalisée *in vitro* en présence d'oxygène (Huang et al., 2015). Néanmoins, des enzymes à cytochrome P450 de la famille CYP71D possèdent des activités sesquiterpène oxydase produisant des alcools sesquiterpéniques comme le capsidiol, qui est une phytoalexine des Solanacées (Takahashi et al., 2005). L'analyse du génome de référence de la vigne PN40024 a permis d'identifier un petit groupe de six gènes de la famille CYP71B codant pour des sesquiterpène oxydases (STO) putatives et nommés VvSTO1 à VvSTO6. Les ADNc des gènes VvSTO2, VvSTO4 and VvSTO6 ont été clonés grâce à des ARN préparés à partir de baies de Syrah. L'expression des protéines correspondantes chez la levure a permis de montrer que

l'enzyme VvSTO2 était capable d'oxyder l' $\alpha$ -guaiène en rotundone (Takase et al., 2016). Le gène *VvSTO2* est plus fortement exprimé dans les baies de Syrah que dans les baies de Merlot, qui est un cépage pauvre en rotundone. L'ensemble de ces travaux a permis de proposer une voie de biosynthèse de la rotundone chez la Syrah, qui dépend de la combinaison d'un allèle spécifique de *TPS24*, *VvGuaS*, qui possède une activité  $\alpha$ -guaiène synthase, et de l'expression importante du gène *VvSTO2* codant pour une enzyme à cytochrome P450 capable d'oxyder l' $\alpha$ -guaiène en rotundone (Figure 7).



**Figure 7 : Voie de biosynthèse de la rotundone chez la Syrah.**

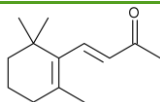
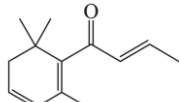
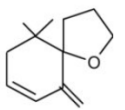
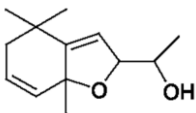
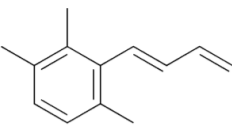
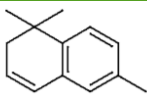
Les produits majeurs des allèles *TPS24 VvGuaS* de la Syrah et *VvSelS* de PN40024 sont indiqués. Les enzymes majeures de la synthèse de la rotundone chez la Syrah sont indiquées en rouge.

### 3.3. Les norisoprénoïdes

Les norisoprénoïdes sont issus du clivage oxydatif des caroténoïdes et servent de précurseurs à deux phytohormones importantes, l'acide abscissique (ABA) et les strigolactones, qui sont des régulateurs essentiels du développement des plantes et des réponses aux stress (Lewinsohn et al., 2005). Les norisoprénoïdes C13 sont généralement plus abondants dans les plantes, mais il existe également des composés de 9 à 11 atomes de carbone (Winterhalter and Rouseff, 2001). Le mécanisme général de formation des norisoprénoïdes commence par le clivage d'un caroténoïde par une dioxygénase de la famille des « carotenoid cleavage dioxygenases » (CCD) (Harrison and Bugg, 2014), suivi de la transformation enzymatique ou non enzymatique des produits de clivage pouvant donner naissance à des précurseurs d'arômes non volatils, puis à des composés aromatiques volatils. Les norisoprénoïdes et leurs précurseurs sont sujets à un grand nombre de réactions possibles, ce qui rend l'étude de leur biosynthèse très complexe (Dunlevy et al., 2009). Les norisoprénoïdes

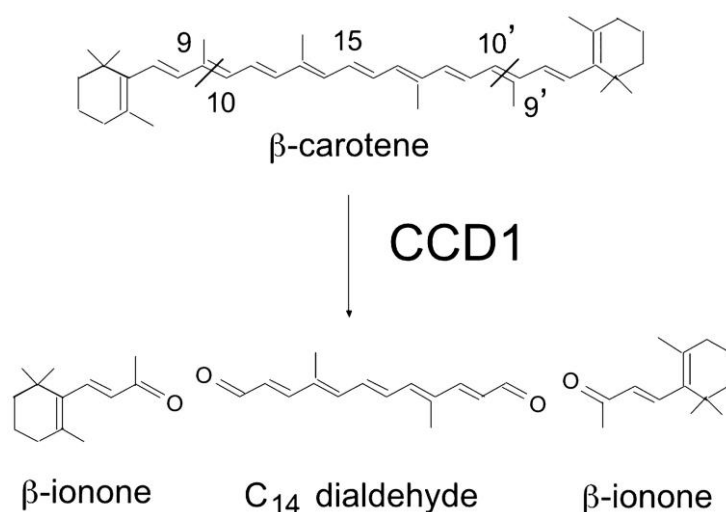
participent aux arômes de nombreuses plantes, fleurs ou fruits, notamment la rose, la tomate, la pêche, le thé, le safran, et la pastèque. Un grand nombre de norisoprénoïdes ont été identifiés dans les baies de raisin et les vins, mais seuls quelques-uns ont un impact olfactif considéré comme significatif (Tableau 4). Comme dans le cas des terpénols, les norisoprénoïdes sont souvent glycosylés par des activités UDP-glycosyltransférases (Schwab et al., 2015). Le pool d'arômes glycosylés peut être libéré sous forme volatile par hydrolyse acide lors de la vinification ou du vieillissement du vin (Loscos et al., 2009).

La  $\beta$ -ionone est un composé à l'arôme de fleur de violette que l'on trouve en quantités supérieures au seuil de détection ( $90 \text{ ng.l}^{-1}$ ) dans les vins issus de nombreux cépages. De même, la  $\beta$ -damascénone est très commune dans les vins, cependant son importance réelle pour l'arôme des vins est sujette à débat (Pineau et al., 2007).

Norisoprénoïde	Structure chimique	Description olfactive	Seuil de détection olfactive (eau ou vin)
$\beta$ -ionone		violette	0.5 ppm
$\beta$ -damascénone		rose	2 ppm
Vitispirane		camphre	8 ppm
Actinidol		camphre	
4- (2,3,6-triméthylphényl) buta-1,3-diène (TPB)		floral	
1,1,6-triméthyl-1,2-dihydronaphtalène (TDN)		kérosène	2 ppm

**Tableau 4 : Structure des norisoprénoïdes les plus courants dans les vins.**

L'analyse des activités potentielles de clivage des molécules de caroténoïdes par les neuf CCD identifiées chez *Arabidopsis* a montré que seuls les CCD1 et CCD4 coupent les caroténoïdes au niveau des doubles liaisons permettant de générer des norisoprénoïdes en C13 et un dialdéhyde en C14 (Figure 8). Cependant, contrairement à la protéine CCD4 localisée dans les plastides, CCD1 semble localisée dans le cytosol, où les caroténoïdes sont beaucoup moins abondants (Auldridge et al., 2006). La caractérisation fonctionnelle d'enzymes CCD4 chez différentes espèces végétales (le crocus, la pomme, la pêche) a montré que toutes ces enzymes sont capables de cliver le  $\beta$ -carotène pour donner la  $\beta$ -ionone (Ahrazem et al., 2010; Auldridge et al., 2006; Brandi et al., 2011; Huang et al., 2009).



**Figure 8 : Activité de clivage du  $\beta$ -carotène par l'enzyme CCD1 de pétunia (d'après Simkin et al., 2004).**

Chez la vigne, la première caractérisation fonctionnelle d'une CCD concerne le gène *VvCCD1* de la Syrah (Mathieu et al., 2005). L'expression de *VvCCD1* dans des souches d'*E. coli* accumulant des caroténoïdes a montré que la protéine recombinante catalyse le clivage de la zéaxanthine et de la lutéine pour former de la 3-hydroxy- $\beta$ -ionone, mais n'a généré aucun produit de clivage avec le  $\beta$ -carotène (Mathieu et al., 2005, 2007). Plus récemment, la caractérisation fonctionnelle des gènes *VvCCD1*, *VvCCD4a* et *VvCCD4b* isolés du cépage Pinotage a montré que les protéines correspondantes étaient toutes capables de catalyser le clivage des doubles liaisons en position 9,10 et 9',10' (Figure 8) de divers substrats de type caroténoïdes, *VvCCD1* étant le seul CCD à produire la  $\beta$ -ionone à partir de  $\beta$ -carotène (Dunlevy et al., 2009). Les profils d'expression de *VvCCD1*, *VvCCD4a* et *VvCCD4b* apparaissent compatibles avec un rôle dans la biosynthèse des norisoprénoïdes. Bien que l'analyse de l'expression du gène *VvCCD1* et la caractérisation fonctionnelle de la protéine correspondante

*in vitro* soient en faveur d'un rôle de VvCCD1 dans le clivage des caroténoïdes des raisins, les expériences de transformation génétiques ne sont pas concluantes. En effet, la surexpression de VvCCD1 chez le cépage Sultana n'a pas révélé de modifications des profils de caroténoïdes ni de norisoprénoïdes chez les plants de vigne transformés, ce qui ne permet pas de conclure quant à la participation réelle de VvCCD1 à la biosynthèse des norisoprénoïdes chez la vigne (Lashbrooke et al., 2013).

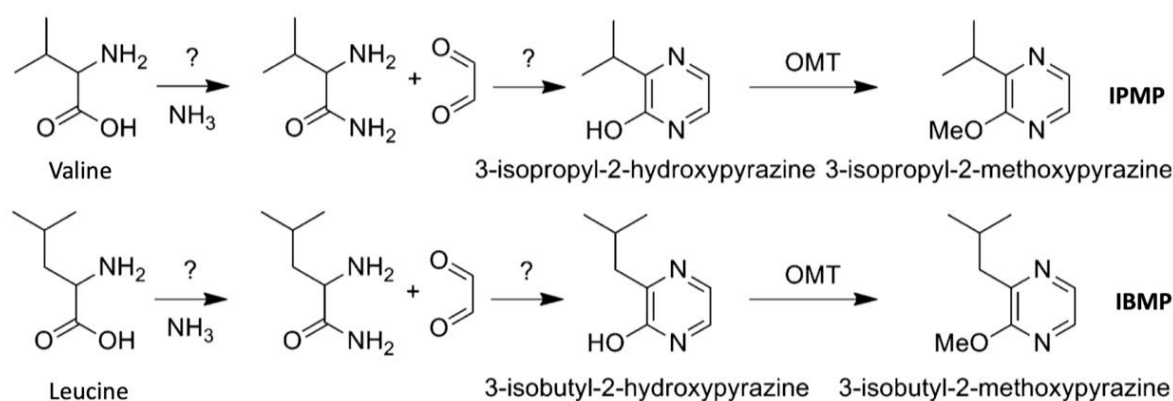
Contrairement à la  $\beta$ -ionone et à la  $\beta$ -damascénone, qui sont présentes dans un grand nombre de vins, seuls quelques cépages produisent des vins contenant des quantités significatives de TDN. Le plus connu d'entre eux est le Riesling, dont les vins peuvent contenir des quantités importantes de TDN (200  $\mu\text{g}\cdot\text{l}^{-1}$ ). Le TDN est responsable des arômes pétrolés caractéristiques des vins de Riesling âgés, arômes parfois perçus comme des défauts s'ils sont trop prononcés (Sacks et al., 2012; Winterhalter, 1991). Le TDN est produit par l'hydrolyse lente de précurseurs glycosylés lors de la fermentation et du vieillissement du vin (Strauss et al., 1987). Bien que le TDN provienne des caroténoïdes qui sont eux même issus de la voie du MEP, la biosynthèse de TDN ne semble pas affectée chez les cépages aromatiques, dont le métabolisme du MEP est particulièrement actif. En effet, les Muscats et le Gewurztraminer ne produisent pas de quantités significatives de TDN. Paradoxalement, le Riesling, caractérisé par sa richesse en TDN, contient des quantités de terpènes moyennement élevées (Tableau 3) et ne produit pas de quantités significativement plus élevées de  $\beta$ -damascénone ou de  $\beta$ -ionone que les autres cépages (Sefton et al., 2011). Ceci suggère l'existence de mécanismes génétiques responsables la formation de TDN spécifiques au Riesling, indépendamment du contenu global en terpènes de ce cépage. Malgré l'importance significative des norisoprénoïdes dans les arômes de nombreux vins, il n'existe pas de preuve génétique de l'implication d'un gène particulier de la vigne dans leur biosynthèse, même si des expériences ont montré que certaines enzymes des raisins pouvaient constituer des candidats potentiels.

### **3.4. Analyse génétique de la biosynthèse des méthoxypyrazines dans les raisins**

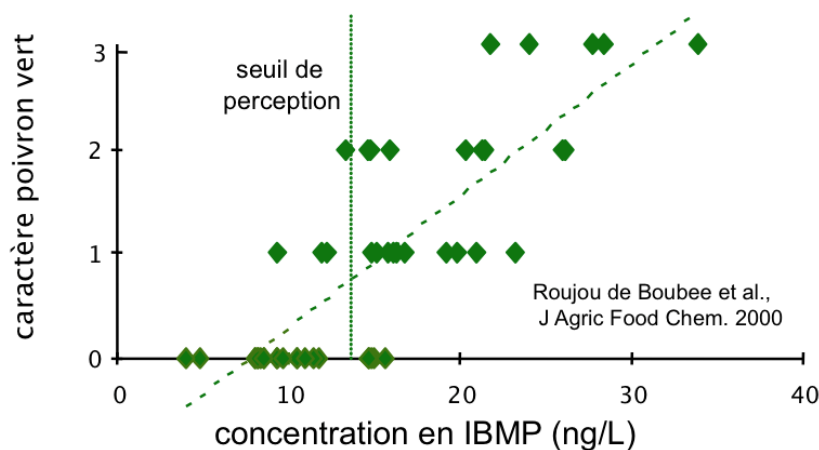
Les pyrazines (1,4-diazines) sont des composés hétérocycliques contenant de l'azote qui sont largement répandus dans la nature, tant dans le règne animal que végétal (Roujou de Boubée et al. 2000). Les pyrazines produites chez les plantes sont en général méthoxylées pour donner des méthoxypyrazines volatiles. Les méthoxypyrazines sont des molécules puissamment odorantes (seuils de détection de 2 à 16  $\mu\text{g}\cdot\text{L}^{-1}$  dans le vin) qui se caractérisent par

des arômes qualifiés d'herbacés, qui rappellent le poivron ou le pois (Heymann et al., 1986; Lei et al., 2018). Les méthoxypyrazines sont présentes à l'état volatil libre dans les raisins et s'accumulent pendant la phase de croissance herbacée de la baie, pour décliner pendant la phase de maturation (Dunlevy et al., 2010, 2013; Guillaumie et al., 2013). Les mécanismes de ce déclin ne sont pas connus et pourraient être dus à un effet de dilution résultant de l'expansion des baies ou à des réactions de déméthylation pour donner des hydroxypyrazines non volatiles (Guillaumie et al., 2013). Les méthoxypyrazines les plus abondantes dans les raisins sont la 3-isopropyl-2-méthoxypyrazine (IPMP), la 3-sec-butyl-2-méthoxypyrazine (SBMP) et la 3-isobutyl-2-méthoxypyrazine (IBMP) (Figure 9), l'IBMP étant considérée comme celle ayant le plus d'impact sur l'arôme des vins. L'IBMP est accumulée en quantités notables dans les raisins des cépages Cabernet Sauvignon, Sauvignon Blanc, Cabernet Franc, Merlot et Carménère. En revanche, les cépages Pinot Noir, Chardonnay et Petit Verdot produisent très peu de méthoxypyrazines (Lei et al., 2018; Sala et al., 2004). Bien qu'elles puissent être perçues négativement lorsqu'elles sont présentes en excès dans le vin, les méthoxypyrazines sont recherchées en petites quantités, en particulier dans certains vins blancs (Robinson et al., 2014).

Contrairement à de nombreuses molécules aromatiques dont les descripteurs sont très variables en fonction des dégustateurs, l'IBMP est associée de manière extrêmement reproductible à des arômes de poivron vert immédiatement reconnaissable par les dégustateurs (Figure 10) (Roujou de Boubée et al., 2000).



**Figure 9 : Voies de biosynthèse putatives de l'IPMP et l'IBMP, les principales méthoxypyrazines des raisins (d'après Dunlevy *et al.*, 2010).**

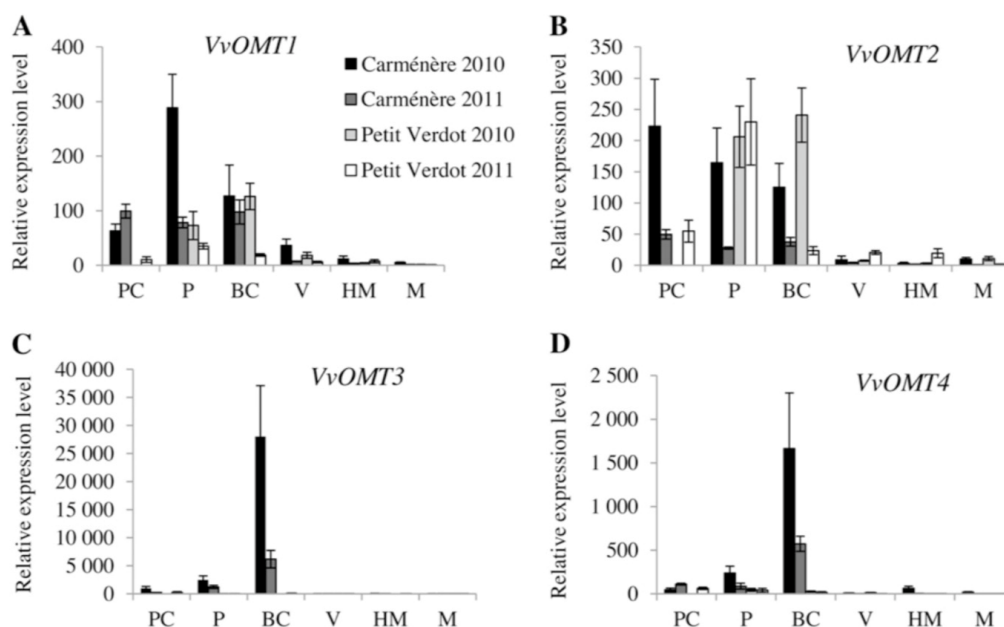


**Figure 10 : Corrélation entre le caractère « poivron vert » perçu à la dégustation et la concentration en IBMP dans les vins (d'après Roujou de Boubée *et al.*, 2000).**

La voie de biosynthèse des méthoxypyrazines IPMP et IBMP n'est pas totalement élucidée, mais le modèle actuel propose qu'elles dérivent respectivement des acides aminés valines et leucine. La dernière étape de la voie de biosynthèse implique une étape de méthylation qui rend les méthoxypyrazines très volatiles et qui est catalysée par une enzyme de type O-méthyltransférase (OMT). Grâce à une approche classique de biochimie des protéines, une OMT capable de méthyler les hydroxypyrazines a été purifiée à partir de tiges de Cabernet Sauvignon, mais le gène correspondant n'avait pu être caractérisé (Hashizume *et al.*, 2001). Une stratégie de caractérisation de gènes candidats a abouti au clonage de deux gènes, *VvOMT1* et *VvOMT2*, codant pour des OMT capables de catalyser la méthylation de la 3-isobutyl-2-hydroxypyrazine (IBHP) en IBMP (Dunlevy *et al.*, 2010). Cependant, ces enzymes possédaient des efficacités et des affinités relativement faibles pour le substrat IBHP ( $K_m$  de l'ordre de 500  $\mu\text{M}$ ), ce qui paraissait peu compatible avec une activité efficace sur un substrat très peu abondant comme l'IBHP. Par ailleurs, *VvOMT1* et *VvOMT2* montraient une activité de méthylation 100 fois plus élevée sur des flavonoïdes comme la quercétine que sur l'IBHP. Plus récemment, deux analyses génétiques de l'accumulation des méthoxypyrazines dans les raisins ont été menées indépendamment, la première sur une population issue d'un croisement entre le Cabernet Sauvignon et *Vitis riparia* cv. Gloire à Montpellier (Guillaumie *et al.*, 2013) et la seconde issue d'un croisement entre le Cabernet Sauvignon et une vigne naine dérivée du Pinot Meunier (Dunlevy *et al.*, 2013). Ces deux études ont permis d'identifier un QTL majeur associé à l'accumulation de méthoxypyrazines sur le chromosome 3. L'intervalle de confiance de ce QTL contenait des gènes codant pour deux OMT, *VvOMT3* et *VvOMT4*, très similaires mais différentes de *VvOMT1* et *VvOMT2*.

## Introduction

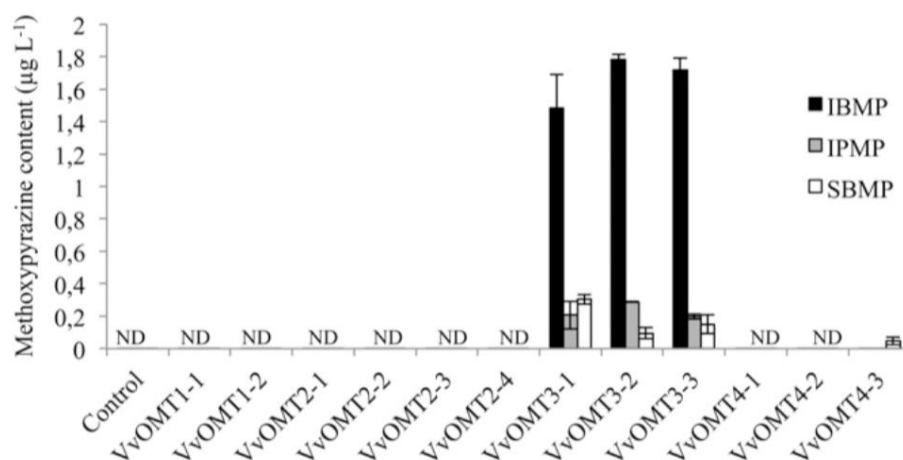
L'analyse de l'activité des protéines VvOMT3 et VvOMT4 en présence de différents substrats a permis de montrer que, contrairement à VvOMT1 et VvOMT2, VvOMT3 possédait des caractéristiques enzymatiques compatibles avec la biosynthèse des méthoxypyrazines, c'est à dire une efficacité et une affinité élevée pour le substrat IBHP. De plus, l'analyse de l'expression des gènes VvOMT dans les baies des cépages fort producteurs ou faible producteurs de méthoxypyrazines a montré que l'expression du gène VvOMT3 et, dans une moindre mesure celle du gène VvOMT4, étaient très bien corrélées avec l'accumulation des méthoxypyrazines dans les raisins (Figure 11) (Dunlevy et al., 2013; Guillaumie et al., 2013). Enfin, l'expression des ADNc de différents allèles des gènes VvOMT1, VvOMT2, VvOMT3 et VvOMT4 chez *E. coli* a abouti, de manière inattendue, à l'accumulation de quantités importantes de méthoxypyrazines dans le milieu de culture des souches exprimant les allèles VvOMT3 (Figure 12) (Guillaumie et al., 2013). L'accumulation d'IBMP dans les cultures d'*E. coli* exprimant les allèles VvOMT3, les rendant immédiatement reconnaissables « au nez » grâce à leur forte odeur de poivron, renforce l'hypothèse du rôle prépondérant de VvOMT3 dans la synthèse des méthoxypyrazines et suggère que le substrat IBHP s'accumule chez *E. coli*, mais n'est normalement pas métabolisé en IBMP faute d'activité OMT appropriée.



**Figure 11 : Analyse de l'expression des gènes VvOMT dans les baies des cépages Carménère (fort producteur de méthoxypyrazines) et Petit Verdot (très faible producteur de méthoxypyrazines) (d'après Guillaumie *et al.*, 2013).**



Stade de développement des baies : PC, Peppercorn size; P, pea size; BC, bunch closure; V, véraison; HM, half mature; M, mature.



**Figure 12 : Accumulation de méthoxy-pyrazines dans le milieu de culture, après expression des ADNc de différents allèles des gènes *VvOMT1*, *VvOMT2*, *VvOMT3* et *VvOMT4* chez *E. coli* (d'après Guillaumie et al., 2013).**

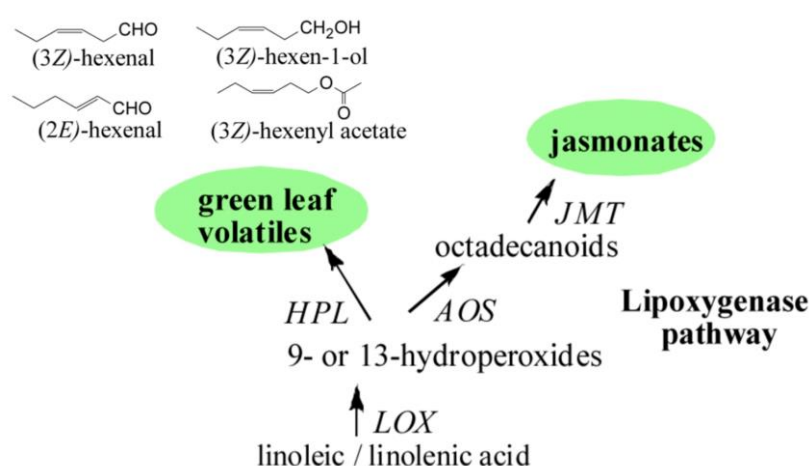
L'ensemble de ces travaux a permis de montrer que le gène *VvOMT3* joue un rôle déterminant dans la biosynthèse des méthoxy-pyrazines, et que les cépages dont les raisins sont pauvres en méthoxy-pyrazines sont caractérisés par une expression très faible de ce gène. La comparaison des régions du génome voisine des gènes *VvOMT3* et *VvOMT4* chez le Cabernet Sauvignon et le Pinot a montré des différences significatives. En particulier, la présence d'un transposon ou d'une délétion au voisinage des allèles *VvOMT3* et *VvOMT4* du Pinot Noir pourrait expliquer la faible expression de ces gènes *OMT* et la faible accumulation de méthoxy-pyrazines chez ce cépage (Dunlevy et al., 2013).

### 3.5. Les produits de dégradation des lipides

Les produits de dégradation des lipides qui participent aux arômes des raisins sont principalement les aldéhydes et les alcools en C6 et C9, tels que les dérivés de l'hexanal, de l'hexanol ou de l'hexenol. Ces composés sont regroupés sous l'appellation de « green leaf volatiles » (GLVs) (Figure 13). Les GLVs sont largement répandus chez les plantes et interviennent dans les réactions aux blessures et dans les défenses contre les pathogènes (Schwab et al., 2008). Les GLVs représentent un pourcentage important des molécules volatiles totales émises par des raisins pressés, allant de 40 à 97% en fonction des cépages (Yang et al., 2009). Les GLVs sont souvent décrits comme ayant des arômes de type « vert » ou herbacés ,

qui peuvent contribuer négativement à l'arôme du vin s'ils sont présents en très grande quantité (Dennis et al., 2012; Kotseridis and Baumes, 2000). Cependant, ils représentent également une source de substrats nécessaires à la formation d'esters souhaitables lors de la fermentation (Dennis et al., 2012). Les aldéhydes en C6 peuvent également être impliqués dans la production de thiols volatils.

Les GLVs sont produits à partir d'acides gras libres comme l'acide linoléique et l'acide  $\alpha$ -linoléique par l'action successive d'enzymes lipoxygénases (LOX) et hydroperoxyde lyases (HPL) (Figure 13). Les LOX végétales peuvent être classées en deux groupes en fonction de leur site d'oxygénation au niveau du C9 (9-LOX) ou du C13 (13-LOX) de l'acide gras. Les 13-LOX peuvent également être classées selon la présence (type II) ou l'absence (type I) d'un peptide de transit plastidial (Liavonchanka and Feussner, 2006; Schwab et al., 2008).



**Figure 13 : Représentation schématique de la voie des lipoxygénases (d'après Niinemets et al., 2013).** Quelques exemples de green leaf volatiles (GLVs) sont indiqués. LOX: lipoxygénase, HPL : hydroperoxyde lyase, AOS : allène oxyde synthase, JMT : jasmonic acid carboxyl methyl transferase.

Le génome de référence PN40024 contient une douzaine de gènes codant pour des LOXs putatives (Podolyan et al., 2010). Deux 13-LOX de type II (VvLOXA et VvLOXO) et deux 9-LOX (VvLOXC et VvLOXD), sont exprimées à différents stade du développement des baies de Sauvignon blanc. Les protéines recombinantes VvLOXA et VvLOXO catalysent l'oxydation des acides gras insaturés *in vitro* pour produire des hydroperoxydes (Podolyan et al., 2010). Les hydroperoxydes subissent ensuite un clivage par des hydroperoxyde lyases (HPL) pour former des aldéhydes en C6 et C9. Les HPL sont des enzymes à cytochrome P450 classées en trois groupes en fonction de leur spécificité de substrat. Les 9-HPL et les 13-HPL

sont respectivement spécifiques des 9-hydroperoxydes et des 13-hydroperoxydes, tandis que les 9/13-HPL sont capables de catalyser le clivage des deux types d'hydroperoxydes (Schwab et al., 2008).

Les ADNc des gènes de deux HPL, *VvHPL1* et *VvHPL2*, ont été clonés à partir d'ARN de baies de Cabernet Sauvignon. La protéine VvHPL1 catalyse le clivage des 13-hydroperoxydes pour produire des aldéhydes en C6. La protéine VvHPL2 clive les 9-hydroperoxydes pour produire des aldéhydes en C9, mais également les 13-hydroperoxydes, bien que son activité soit très inférieure à celle de VvHPL1 (Zhu et al., 2012). Le clonage de gènes LOX et HPL chez la vigne a montré que les enzymes correspondantes possédaient des activités *in vitro* compatibles avec leur implication dans la biosynthèse des GLVs responsables de certains arômes des raisins. Néanmoins, il n'existe pas de preuve génétique de l'implication d'un gène particulier dans ces voies de biosynthèse à l'heure actuelle.

### 3.6. Les thiols volatils

Les composés soufrés volatils participent aux arômes de nombreuses plantes comme le fruit de la passion, le cassis, l'ail et les asperges. (Dubourdieu and Tominaga, 2009; Dunlevy et al., 2009; Robinson et al., 2014). La 4-mercapto-4-méthylpentan-2-one (4MMP) a été le premier thiol volatil identifié comme contribuant de manière importante à l'arôme de Sauvignon blanc (Darriet et al., 1995). La 4MMP est une molécule extrêmement odorante (seuil de perception : 0,8 ng.l<sup>-1</sup>) qui possède une odeur marquée de buis et de cassis. Le 3-mercaptohexan-1-ol (3MH) et son ester d'acétate (3MHA), ainsi que le 4-mercapto-4-méthylpentan-2-ol (4MMPOH), ont été identifiés peu après (Tominaga et al., 1996, 1998) (Figure 14). Le 3MH (seuil de perception : 60 ng.l<sup>-1</sup>) et son ester 3MHA (seuil de perception : 4 ng.l<sup>-1</sup>) contribuent aux notes de pamplemousse et de fruit de la passion caractéristiques des vins de Sauvignon blanc. Même s'ils sont particulièrement importants dans les vins de Sauvignon, les thiols volatils sont également des contributeurs majeurs aux arômes typiques des vins issus d'autres cépages tels que les Muscats, le Gewurztraminer, le Riesling, le Chenin blanc et le Sémillon (Robinson et al., 2014; Tominaga et al., 2000).

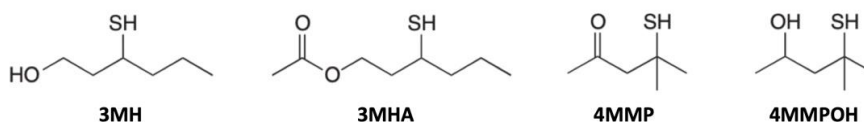


Figure 14 : Les principaux thiols volatils impliqués dans les arômes des vins.

Contrairement à la majorité des composés volatils importants pour l'arôme des raisins, les thiols volatils 3MH, 3MHA, 4MMP et 4MMPOH n'existent pas sous forme libre ni sous forme de conjugués glycosylés dans les raisins. La source exacte des précurseurs de thiols et des thiols dans les vins est encore controversée, néanmoins des conjugués de cystéine (Tominaga et al., 1998) et de glutathion ont été isolés dans les baies.

Trois voies de biosynthèse sont communément acceptées pour expliquer la libération de 4MMP et de 3MH dans les vins (Figure 15) (Roland et al., 2011). La première voie implique des précurseurs cystéinylés, initialement identifiés dans les cépages Sauvignon blanc, puis dans d'autres cépages comme les Petit et Gros Manseng, le Riesling et le Gewurztraminer. Ces conjugués de cystéine sont clivés au cours de la fermentation, par des activités de type « cystéine conjugate  $\beta$ -lyase » des levures (Tominaga et al., 1998). Une seconde voie impliquerait des précurseurs glutathionylés. Le précurseur du 3MH glutathionylé (3MH-glut) semble résulter de la conjugaison du glutathion avec le (E)-2-hexéanal par l'action d'activités glutathion S-transférases (GST). L'expression des gènes *GST* peut être induite en condition de stress parallèlement à la voie des lipoxygénases, ce qui lierait la production des précurseurs de thiols à la voie des lipoxygénases et aux GLVs (Kobayashi et al., 2011; Thibon et al., 2011). Une troisième voie de biogénèse impliquerait des composés insaturés en C6 issus de la dégradation des lipides, tels que le (E) -2-hexéanal, qui subissent une addition de soufre pendant la fermentation alcoolique (Harsch et al., 2013). Le donneur de soufre n'a pas encore été identifié, mais il pourrait s'agir de  $H_2S$ , de la cystéine, du glutathion ou d'autres molécules du moût ayant une fonction SH thiol libre. Enfin, la biosynthèse de 3MHA repose totalement sur la fermentation, car 3MHA est produit à partir de 3MH par l'action de l'alcool acétyltransférase ATF1 de levure (Swiegers and Pretorius, 2007).

Malgré l'existence de nombreux résultats plaçant en faveur d'un rôle majeur des précurseurs cystéinylés et glutathionylés dans la richesse des vins en thiols volatils, certains travaux ont montré de faibles corrélations entre les concentrations de ces précurseurs dans les jus de raisin et la présence de 3MH et de 4MMP dans les vins (Patel et al., 2010; Pinu et al., 2012; Pinu et al., 2014), ce qui explique que l'origine exacte des thiols volatils des vins soit

encore débattue. Néanmoins, de nombreux cépages connus pour la richesse en thiols de leurs vins partagent des relations de parenté (Dubourdieu and Tominaga, 2009; Myles et al., 2011), ce qui suggère l'existence d'un déterminisme génétique de la teneur en thiols des vins, déterminisme dont les acteurs sont actuellement inconnus.

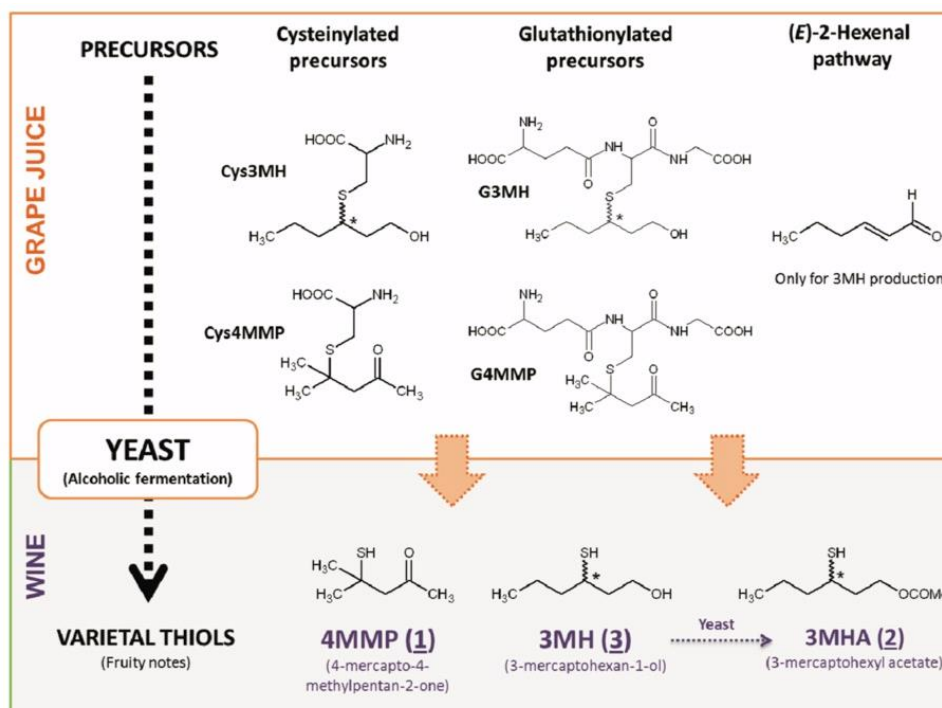


Figure 15 : Biosynthèse des principaux thiols volatils des vins à partir de leurs précurseurs présents dans les raisins (Roland et al., 2011).

#### 4. Objectifs de la thèse

Malgré l'importance des molécules aromatiques des raisins dans la qualité et la typicité des vins, les déterminants de leur biosynthèse sont encore très mal connus, à l'exception des quelques gènes présentés précédemment. En particulier, peu de marqueurs génétiques sont actuellement utilisables dans les programmes de création variétale, pour la sélection de variétés ayant des profils aromatiques bien définis. L'objectif de ma thèse est d'identifier les régions génomiques et les gènes clés gouvernant la biosynthèse de certains des arômes et des précurseurs d'arômes de la baie de raisin. A court et moyen terme, une meilleure connaissance des bases génétiques de la biosynthèse des arômes de raisins et l'identification de marqueurs associés permettra d'optimiser les profils aromatiques dans les programmes de sélection de nouvelles variétés de vigne.

Dans le cadre de ma thèse, j'ai principalement utilisé une population d'étude constituée de génotypes de vigne issus d'un croisement entre les cépages Riesling (Ri) et Gewurztraminer (Gw), variétés qui ont été choisies en raison de leurs caractéristiques très différentes (phénologie, arômes, acidité, couleur des baies). Pour identifier des gènes clés de la biosynthèse de composés aromatiques des raisins, la stratégie utilisée pendant ma thèse comporte différentes étapes :

- L'analyse des métabolites secondaires associés aux arômes présents dans les baies de raisin au sein de la population RixGw
- La recherche de Quantitative Trait Loci (QTL) liés aux teneurs en métabolites secondaires associés aux arômes et précurseurs d'arômes dans la descendance du croisement RixGw, c'est à dire des régions du génome responsables des variations quantitatives du contenu en composés aromatiques considérés au sein de la population,
- L'analyse précise des régions génomiques associées aux QTL pour ces métabolites, dans le but d'identifier des gènes candidats potentiellement impliqués dans leur synthèse,
- La caractérisation fonctionnelle de certains gènes candidats identifiés.

Mon travail s'articule ainsi en 4 chapitres :

Le chapitre 1 porte sur l'analyse génétique de la biosynthèse des molécules potentiellement volatiles des raisins quantifiés par chromatographie en phase gazeuse couplée à de la spectrométrie de masse (GC-MS).

Le chapitre 2 s'attache à l'analyse génétique de la biosynthèse des molécules du métabolisme secondaire de la vigne quantifiée par chromatographie en phase liquide à haute pression couplée à de la spectrométrie de masse (UHPLC-MS).

Le chapitre 3 porte sur Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par spectrométrie de masse à résonance cyclotronique ionique (FT-ICR) et UHPLC-MS à la lumière des QTL

Le chapitre 4 se concentre sur les gènes liés aux loci des QTL trouvés dans les 3 premiers chapitres, en essayant de trouver des explications biologiques concernant les allèles du Ri et du Gw qui pourraient avoir une influence sur la différence de contenu en arômes ou en précurseurs d'arômes entre le Ri et le Gw ainsi que parmi les génotypes de la descendance du croisement RixGw.



## **Matériels et méthodes**





### 1. Population d'étude RixGw et échantillons analysés pendant la thèse

La population d'étude de 383 individus (RixGw) est issue d'un croisement entre deux cultivars : le Riesling (Ri) clone 49 et le Gewurztraminer (Gw) clone 643. Les cépages Ri et Gw ont été choisis en raison de leurs caractéristiques très différentes (phénologie, arômes, acidité et couleur des baies, etc...). Chaque génotype a été greffé sur le porte-greffe Couderc 161-49 et trois plants par génotype ont été plantés sur un vignoble expérimental à Bergheim (48°210 N, 7°340 E) en 2006 (Figure 16).

Les analyses ont porté sur 117 individus de la population et sur les parents sur 3 années de prélèvements de baies (2014, 2015 et 2018) et une année de prélèvement de feuilles (2015). L'étude de cette population a déjà permis l'analyse génétique de nombreux caractères d'intérêts agronomiques, avec entre autres le déterminisme génétique des teneurs en géraniol et en linalol (Duchêne *et al.*, 2009). Cette population parfaitement caractérisée constitue un outil extrêmement performant pour la génomique fonctionnelle des caractères d'intérêts agronomiques chez la vigne. Cet outil a permis la caractérisation d'un gène majeur de la biosynthèse des arômes terpéniques qui est déjà utilisé par l'industrie pour des applications biotechnologiques (Hugueney *et al.*, 2012, brevet international). L'étude de cette population a déjà permis l'analyse génétique de caractères d'intérêts agronomique tels que la phénologie, l'acidité, les contenus en terpénols (Duchêne *et al.*, 2012a, Duchêne *et al.*, 2012b, Hugueney *et al.*, 2012, Duchêne *et al.*, 2014, Ilc *et al.*, 2017).



Riesling (Ri)

Gewurztraminer (Gw)

X



Population d'étude  
(RixGw)

Figure 16 : Population d'étude plantée à Bergheim (68) (Source : Eric Duchêne).

### 1.1. Matériel végétal

#### *1.1.1. Prélèvement des baies et des jus*

Pour les 117 génotypes du croisement RixGw et leurs parents, 250 ml de baies ont été collectés pendant leur maturation sur 3 années : 2014, 2015 et 2018. La date de véraison, qui correspond au début du processus de maturation des raisins, a été déterminée pour chaque génotype, à partir d'observations régulières, comme la date à laquelle 50% des baies étaient molles (entrant dans le processus de maturation). La date de prélèvement a été déterminée pour chaque génotype comme la date où la somme des températures de 350 °C-jours (calculée avec les températures maximales, température de base = 10 °C) a été atteinte après la véraison. Cette procédure a été validée précédemment pour évaluer les différences génotypiques quant à la capacité d'accumulation de sucre (Duchêne et al., 2012a) et de l'acidité dans les baies. Soixante grammes de baies ont été congelées à -20 °C directement après les prélèvements de 2014 et 2015 et pour les jus issus de baies prélevées en 2018, 120 µl d'une solution de Na<sub>2</sub>SO<sub>3</sub> à 10 g.l<sup>-1</sup> ont été ajoutés aux jus avant congélation à -80 °C.

#### *1.1.2. Prélèvement des feuilles*

Des feuilles de 117 génotypes ainsi que de Ri et de Gw ont été prélevées le 21 mai 2015 afin d'étudier le métabolisme secondaire des feuilles de vigne. La 5<sup>ème</sup> feuille en partant de l'apex a été prélevée puis des disques foliaires de 20 mm ont été découpés et congelés pour les analyses métabolomiques par UHPLC-MS.

### 1.2. Extraction des échantillons en vue des analyses métabolomiques

L'extraction des métabolites secondaires des échantillons de baies et de jus a été réalisée par Extraction sur Phase Solide (SPE). Cette technique d'extraction a pour avantage d'éliminer les molécules polaires du métabolisme primaire tels que les sucres, mais aussi de concentrer les métabolites secondaires qui sont présents à de faibles concentrations.

#### *1.2.1. Extraction des baies par SPE pour des analyses par GC-MS et UHPLC-MS*

Quarante g de baies congelées ont été épépinées et broyées sous azote liquide. La poudre a été mise en suspension dans 4 ml d'une solution de Na<sub>2</sub>SO<sub>3</sub> à 10 g.l<sup>-1</sup> et 20 µl d'une solution

de 3-octanol à  $1 \text{ g.l}^{-1}$  a été ajouté comme étalon interne pour les analyses de la fraction libre par GC-MS). Après centrifugation (25 min, 16 500 g à  $4 \text{ }^{\circ}\text{C}$ ), le surnageant a été passé à travers un préfiltre en fibre de verre et un filtre en verre et  $150 \mu\text{l}$  d'une solution d'apigénine à  $100 \text{ mg.l}^{-1}$  ont été ajoutés comme étalon externe pour les analyses par UHPLC-MS. Le filtrat a été passé à travers une cartouche d'extraction C18 en phase solide (SPE) de 1 g liée à la silice (Bond Elut Jr., technologies Agilent), préalablement rincée avec 5 ml d'éthanol et 15 ml d'eau ultrapure, à un débit de  $1 \text{ goutte.sec}^{-1}$ . Après lavage avec 15 ml d'eau ultrapure, les composés d'intérêt ont été élués avec 4 ml de dichlorométhane pour obtenir la fraction libre qui est analysée par GC-MS.

Une seconde élution SPE a été réalisée avec 4 ml d'éthanol pour obtenir la fraction glycosylée.  $200 \mu\text{L}$  de cette 2<sup>ème</sup> élution ont été réservées pour l'analyse par UHPLC-MS. Les 4.8 ml restants ont été dilués dans 40 ml de tampon citrate/phosphate (pH 4,5) et la solution résultante a été incubée avec 50 mg d'enzyme glycolytique AR2000 (Gist-Brocades, Seclin, France) pendant une nuit à  $37.5 \text{ }^{\circ}\text{C}$  pour libérer les molécules volatiles liées à des sucres (Tamborra et al., 2004). La solution a ensuite été passée à nouveau sur cartouche SPE C18 et élué avec 4 ml de dichlorométhane. Pour les analyses par GC-MS des fractions libres et liées, les échantillons ont été déshydratés dans des pipettes Pasteur remplies avec 0.5 g de sulfite de sodium anhydre et concentrés à  $500 \mu\text{l}$  sous un léger flux d'azote. Vingt  $\mu\text{l}$  d'une solution à  $1 \text{ g.l}^{-1}$  de m-crésol ont été ajoutés à chaque échantillon comme étalon externe.

### *1.2.2. Extraction des jus pour des analyses par UHPLC-MS et FT-ICR*

Douze ml de jus ont été décongelés à température ambiante pendant 1 h puis agités pour homogénéisation et centrifugés à 8000 rcf pendant 10 min. 5 ml du surnageant a été prélevé et  $2 \mu\text{L}$  d'une solution d'apigénine à  $100 \text{ mg.l}^{-1}$  a été ajouté. Le surnageant a été passé à travers une cartouche SPE C18 de 200 mg (Bond Elut Jr., technologies Agilent). Les éluions ont été réalisées par centrifugation à 450 rcf pendant 4 min. La colonne avait été préalablement rincée avec 4 ml d'éthanol et 8 ml d'eau ultrapure. Cinq ml de surnageant ont été déposés sur la cartouche SPE. Après lavage avec 4 ml d'eau ultrapure, puis  $100 \mu\text{l}$  d'éthanol, les composés d'intérêt ont été élués avec  $500 \mu\text{l}$  d'éthanol.

Pour les analyses par FT-ICR,  $20 \mu\text{l}$  d'extrait SPE ont été dilués dans 1 mL de méthanol avant analyse. Pour les analyses par UHPLC-MS,  $10 \mu\text{l}$  d'isobutanol ont été ajoutés à  $200 \mu\text{L}$  d'extrait SPE et l'ensemble a été concentré au Speedvac pendant 45 min à  $30 \text{ }^{\circ}\text{C}$ . Quatre-vingt-

dix  $\mu\text{l}$  d'éthanol absolu ont ensuite été ajoutés aux 10  $\mu\text{l}$  d'extrait concentré pour obtenir un volume final de 100  $\mu\text{l}$ .

### *1.2.3. Extraction des feuilles pour des analyses par UHPLC-MS*

Chaque disque foliaire a été pesé avant séchage pendant une nuit à 95 °C. Chaque disque a été broyé avec des billes en métal puis 60  $\mu\text{l}$  de solvant d'extraction (MeOH + diosmétine (2  $\mu\text{g}/\text{ml}$ )) par mg de poids frais de feuille a été ajouté. Une incubation à 65 °C pendant 30 min a été réalisée puis une centrifugation 4 °C 10 min à 12000 rpm.

## **2. Analyses métabolomiques de la population de croisement RixGw**

Durant ma thèse, j'ai pu obtenir des données analytiques sur des extraits SPE de baies, de jus et de feuilles par GC-MS, UHPLC-MS et FT-ICR, outils que je vais détailler ci-dessous.

### **2.1. Description des outils analytiques utilisés pendant la thèse**

#### *2.1.1. Etude du « volatilome » des baies par GC-MS*

L'analyse des composés volatiles a été réalisée à l'aide d'un chromatographe en phase gazeuse GC6890N (Agilent) équipé d'un injecteur Split/Splitless et couplé à un spectromètre de masse MDS5975 (Agilent) (Figure 17). La séparation de 1  $\mu\text{l}$  d'extrait a été réalisée sur une colonne DB-Wax (30 m x 0,32 mm, épaisseur de film 0,5  $\mu\text{m}$  ; JW Scientific) en utilisant l'hélium comme gaz vecteur à un débit de colonne de 1.5  $\text{ml}\cdot\text{min}^{-1}$ . Les conditions de séparation étaient les suivantes : injecteur à 250°C en mode Pulsed Splitless (25 psi pendant 0.5 min) ; température du four passant de 45° C à 82 °C avec un gradient de 20 °C.min<sup>-1</sup>, plateau de 1 min puis jusqu'à 235 °C avec un gradient de 3 °C.min<sup>-1</sup>, plateau à 235 °C maintenu pendant 10 min. La ligne de transfert a été maintenue à 270 °C. Le spectromètre de masse a été utilisé en mode impact électronique à 70 eV. Les températures de la source et de l'analyseur quadripolaire étaient respectivement de 230 et 150 °C. Les ions ont été enregistrés en mode scan sur une plage de m/z allant de 30 à 300 unité de masse atomique (a.m.u). Le logiciel ChemStation (G1701DA, Rév. D.03.00) a été utilisé pour le contrôle des instruments et le traitement des données.



Figure 17 : Photo de la GC-MS (source : Patricia Claudel).

### 2.1.2. Etude du métabolisme secondaire de la vigne par UHPLC-MS

L'analyse des composés solubles dans les différents échantillons a été réalisée par l'utilisation d'une chromatographie liquide à haute pression couplé à un spectromètre de masse (UHPLC-MS) (Figure 18). Le système est décrit dans les paragraphes suivants.

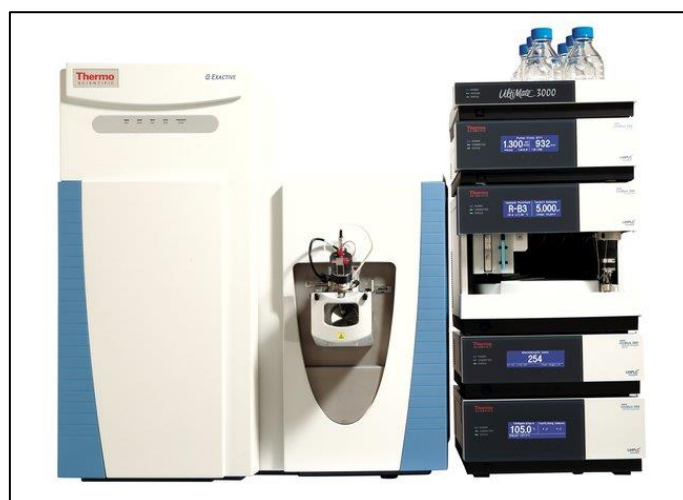


Figure 18 : Photo de l'UHPLC-MS.

#### a. Séparation physico-chimique par chromatographie liquide

Les analyses ont été réalisées sur un système de chromatographie liquide à ultra-haute performance (UHPLC) Dionex Ultimate 3000, équipé d'une colonne Nucleodur C18 HTec (Macherey-Nagel, 50 x 2 mm diamètre interne, particules de 1,8  $\mu\text{m}$  de diamètre) maintenue à

30 °C. Les solvants utilisés étaient les suivants : acétonitrile/acide formique (0,1 %, v/v ; éluant A) et eau/acide formique (0,1 %, v/v ; éluant B) ; le débit a été maintenu à 0,4 mL.min<sup>-1</sup>. Le programme de séparation utilisé est le suivant : 0–4 min, 80–70% de B; 4-5 min, 70-50% de B; 5 à 6,5 min, 50% de B; 6,5 à 8,5 min, 50 à 0% de B; 8,5 à 10 min 0% B. Le volume d'échantillon injecté était de 1 µl.

La séparation par UHPLC permet de diminuer la complexité du mélange de molécule arrivant au spectromètre de masse et ainsi diminuer l'effet de suppression d'ion lorsque le mélange est trop riche ou concentré par rapport à une injection directe (sans séparation). On obtient ainsi une meilleure sensibilité pour une meilleure quantification plus précise. Cela permet aussi de séparer les isomères de molécules ayant les mêmes m/z et donc les mêmes formules brutes mais n'ayant pas la même structure. Les composés les plus polaires vont sortir en premier, du fait de leur faible affinité avec la colonne C18 alors que les composés apolaires sortiront à la fin du fait de leur forte affinité avec la colonne C18. Le gradient de solvant lui, permet de changer l'affinité des molécules sur la colonne (phase stationnaire) et de séparer encore plus les molécules qui auraient un temps de rétention proche. L'acide formique à 0,1% permet d'aider à l'ionisation des molécules pour la partie analyse par spectromètre de masse.

### b. Analyse par le spectromètre de masse

Le système de chromatographie en phase liquide est couplé à un spectromètre de masse Exactive Orbitrap (Thermo Fisher Scientific) équipé d'une source d'ionisation par électrospray (ESI). Le spectromètre de masse permet une analyse qualitative et quantitative de tous les métabolites présents dans un échantillon complexe. Les métabolites provenant de l'UHPLC sont ionisés par une source d'ionisation électrospray (ESI : electro-spray ionisation source), soit en mode positif : ESI (+) et soit en mode négatif : ESI (-). Le flux d'ions est dirigé par deux gaz d'azote : le flux de gaz enrobant la source d'ionisation, et le gaz auxiliaire qui permet de focaliser le spray au niveau du point d'entrée du spectromètre de masse. Ces gaz ont été maintenu à 40 et 5 unités arbitraires, respectivement. Les ions peuvent ensuite être sélectionnés selon les rapports m/z au niveau de l'octapole et du flatapole (Transfer Multipole), ce qui permet de cibler une gamme particulière de masses. Pour nos analyses, les spectres ont été acquis dans la plage de masse m/z de 95 à 1200 a.m.u., avec une résolution de 50 000 à m/z 200 a.m.u., avec une température du capillaire de transfert d'ions à 300 °C et un voltage à 3400 V. La cellule C-Trap permet de rediriger le flux d'ions soit vers la cellule HCD qui permet de fragmenter les ions (utilisation pour la détermination de structure), soit vers la cellule Orbitrap.



La cellule Orbitrap est une électrode creuse dans laquelle est placée une électrode en fuseau, cette cellule permet de piéger les ions par une force électrostatique. Les ions qui pénètrent dans la cellule Orbitrap entament un mouvement oscillato-rotatoire autour de l'électrode. Ce mouvement est détecté et analysée par la cellule Orbitrap et permet une mesure de  $m/z$  très précise. Le système est étalonné en ESI (+) et ESI (-) en utilisant des solutions de calibration externe selon les instructions du constructeur. De plus, le système est étalonné en interne en ESI (+) en utilisant du phtalate de dibutyle comme lockmass ( $m/z=279,1591$ ). Les analyses ont ainsi une précision en masse inférieure à 1 ppm en ESI (+) et inférieure à 10 ppm en ESI (-).

$$\text{Précision en masse (ppm)} = \frac{m/z \text{ observée} - m/z \text{ attendue}}{m/z \text{ attendue}} \times 10^6$$

**Équation 1 : Précision en masse (ppm).**

### *2.1.3. Profilage métabolique par FT-ICR de jus provenant de baies de la population d'étude*

Les spectres de masse à haute résolution ont été acquis sur un spectromètre de masse à transformée de Fourier par résonance cyclotron ionique Bruker (FTICR-MS) (Bruker Daltonics GmbH, Bremen, Allemagne) équipé d'un aimant supraconducteur de 12 Tesla (Magnex Scientific Inc., Yarnton, GB) et une source APOLOII ESI (Bruker Daltonics GmbH, Bremen, Allemagne) en mode d'ionisation négative et positive (Figure 19). Les spectres ont été étalonnés de manière externe sur des solution d'arginine ( $10 \text{ mg.l}^{-1}$ ). Un étalonnage interne supplémentaire a été effectué pour chaque échantillon en utilisant des molécules omniprésentes trouvées dans les échantillons, atteignant des précisions de masse inférieures à 0,1 ppm en mesure de routine. Les spectres ont été acquis sur une plage de masse  $m/z$  de 122 à 1000 a.m.u. Deux cent balayages ont été accumulés pour chaque échantillon dans chaque mode d'ionisation.



Figure 19 : FT-ICR au Helmholtz Zentrum à Munich.

## 2.2. Traitement des données issus des analyses métabolomiques

Pour chaque donnée de semi-quantification des molécules issues des analyses métabolomique, l'héritabilité ( $h^2$ ) a été calculée en tant que rapport entre la variance génétique et la somme de la variance génétique et de la variance résiduelle (Équation 2). La variance résiduelle a été obtenue à l'aide d'une analyse de variance avec les réplicats des deux parents dans chaque jeu de données.

$$\text{Héritabilité } (h^2) = \text{Variance génétique} / (\text{Variance génétique} + \text{Variance résiduelle})$$

Équation 2 : Calcul de l'héritabilité ( $h^2$ )

### 2.2.1. Traitement des données obtenues par GC-MS

Les fichiers correspondants aux spectres obtenus par GC-MS ont été convertis en .AIA par le logiciel Agilent ChemStation. Ils ont ensuite été alignés année par année avec l'outil en ligne Workflow4Metabolomics (Giacomoni et al., 2015).

Les paramètres suivants ont été utilisés:  $fwhm = 5$ ,  $rtdiff = 0.1$ ,  $minfeat = 5$ ,  $simthreshold = 0.7$ ,  $minclassfraction = 0.1$ ,  $minclasssize = 3$ . Une liste de composés quantifiés mais non identifiés est obtenue suite à l'alignement des spectres. L'identification des molécules volatiles détectées a été faite en comparant le spectre de masse de la molécule dans l'échantillon où la quantification est la plus intense, à ceux d'étalons authentiques et de bibliothèques de spectres. L'Institut national américain des normes et de la technologie (NIST-05a) et les bibliothèques de spectres de masse Wiley Registry 7e édition ont été utilisés pour l'identification. Les seuils

d'appariement des spectres étaient d'au moins 85% et l'indice de rétention était pris en compte pour l'identification du composé. Seules les molécules détectées au cours des deux années d'analyse d'arôme de raisin ont été conservées pour analyse ultérieure et détection QTL.

$$\text{Indice de rétention de Kovats} = 100 \times \left[ n + \frac{\log(t(\text{inconnu}) - \log(t(n)))}{\log(t(N) - \log(t(n)))} \right]$$

n, le nombre d'atomes de carbone dans le plus petit *n*-alcane ;

N, le nombre d'atomes de carbone dans le plus grand *n*-alcane ;

t, le temps de rétention

**Équation 3 : Calcul de l'indice de rétention (RI)**

### *2.2.2. Traitement des données obtenues par UHPLC-MS pour une approche non ciblée*

Les fichiers de données brutes issues des analyses UHPLC-MS ont été convertis en fichier de type .mzXML via le logiciel MSconvert (<https://bio.tools/msconvert>). Le logiciel XCMS (Smith et al., 2006) est ensuite utilisé afin de quantifier des ions de manière non ciblée.

Pour l'ensemble des analyses, les paramètres suivant ont été utilisés dans XCMS: method="centWave", ppm=2, noise=30000, mzdiff=0.001, prefilter=c(5,15000), snthresh=6, peakwidth=c(6,35).

Il faut noter qu'un plus grand nombre de molécules sont ionisées en ESI (+) qu'en ESI (-) et que l'ESI (+) a la particularité de générer plus de fragments moléculaires, c'est pourquoi le nombre d'ions détectés en ESI (+) est toujours supérieur à celui en ESI (-).

En complément de ces analyses séparées des différents jeux de données, un alignement entre les données UHPLC-MS de baies prélevées en 2014 et en 2015 a été tenté mais n'a pas été concluant, du fait de la différence entre les temps de rétention observés sur les analyses entre les deux années (différence de temps de rétention de 0.2 min pour le standard interne). La comparaison des deux années a été faite sur les alignements séparés des deux jeux de données puis par recouplement de m/z entre des gammes de temps de rétention ou encore par comparaison directe des m/z détectés. La trop grande différence de temps de rétentions entre les données de feuilles et de baies dû à des périodes d'injection trop éloignés des échantillons

de feuilles et de baies prélevés en 2015 n'a pas permis de pouvoir aligner les données analytiques des baies 2015 avec celles des feuilles 2014. Dans ce cas, ces données ont été comparées en utilisant seulement la masse  $m/z$  des ions, sans tenir compte du temps de rétention. En effet, la masse des ions  $m/z$  mesurée par l'Orbitrap est à la fois très précise et très stable, et ne varie pas, même si les analyses sont effectuées à plusieurs années d'intervalle.

### *2.2.3. Traitement des données obtenues par UHPLC-MS pour une approche ciblée par famille de molécules*

Pour l'analyse ciblée, j'ai surtout utilisé les analyses réalisées en ESI (+) car ces analyses ont une meilleure précision en masse (< 1 ppm) et permettent d'avoir une fragmentation nous permettant d'identifier plus facilement la structure des molécules et d'obtenir des fragments caractéristiques de certaines familles de molécules. L'utilisation de standard, quand cela est possible, a été réalisée afin de confirmer l'identification des molécules par le temps de rétention et par son  $m/z$ . La recherche ciblée de molécules appartenant à des molécules de la famille des monoterpénols et de la famille des norisoprénoïdes a été réalisée. Pour cela, des fragments caractéristiques de ces familles de molécules ont été recherchés et sont détaillés ci-dessous.

#### *c. Recherche ciblée de molécules appartenant à la famille des monoterpénols*

L'arôme typique des Gw est en partie lié au monoterpénols présents en grande quantité. Pour la recherche ciblée de monoterpénols dans les analyses par UHPLC-MS ESI (+), trois fragments caractéristiques des monoterpénols ont été recherchés (Tableau 5).

<b>m/z</b>	<b>Fragment terpénol</b>
<b>137.133</b>	$[\text{C}_{10}\text{H}_{18}\text{O} - \text{H}_2\text{O} + \text{H}]^+$
<b>153.127</b>	$[\text{C}_{10}\text{H}_{18}\text{O}_2 - \text{H}_2\text{O} + \text{H}]^+$
<b>185.117</b>	$[\text{C}_{10}\text{H}_{16}\text{O}_3 + \text{H}]^+$

**Tableau 5 : Fragments caractéristiques des monoterpénols recherchés dans les analyses de baies 2014 et 2015 par UHPLC-MS ESI (+).**

Ces 3 fragments ne sont pas les seuls pouvant être utilisés pour identifier les monoterpénols, cependant je me suis concentrée sur ces 3  $m/z$  pendant ma thèse car ils sont les plus détectés dans les données UHPLC-MS.

#### d. Recherche ciblée de molécules appartenant à la famille des norisoprénoïdes

L'arôme typique de vins âgés de Riesling est dû à un norisoprénoïde : le TDN. C'est pourquoi nous avons cherché à étudier des molécules de cette famille. La recherche de norisoprénoïdes dans les échantillons analysés en LC-MS a consisté à rechercher un fragment caractéristique d'un norisoprénoïde. Par exemple, un norisoprénoïde de formule brute  $C_{13}H_{20}O$  a un  $m/z$  attendu de 193.1586 en mode positif en UHPLC-MS. Seules les molécules détectées sur les 2 années et pour lesquelles un QTL a été détecté sont détaillées dans ce manuscrit.

$m/z$	Fragment norisoprénoïde
<b>193.1586</b>	$[C_{13}H_{22}O_2 - H_2O + H]^+$
<b>207.138</b>	$[C_{13}H_{20}O_3 - H_2O + H]^+$
<b>209.154</b>	$[C_{13}H_{22}O_3 - H_2O + H]^+$

**Tableau 6 : Fragments caractéristiques des norisoprénoïdes recherchés dans les analyses de baies de 2014 et 2015 par UHPLC-MS ESI (+).**

#### 2.2.4. Traitement des données obtenues par FT-ICR

L'alignement des spectres de masses obtenus par FT-ICR de tous les génotypes a été réalisé grâce à un script R, avec un seuil d'erreur maximum toléré de 1 ppm. Un filtrage des  $m/z$  a été réalisé sur les paramètres suivants :

- Rapport Signal/Bruit > 4
- Ion apparaissant dans au moins dans 10% des échantillons.
- Elimination de l'effet « wiggles » autour des pics
- Filtrage des isotopes

L'annotation des ions a été réalisée grâce au logiciel NetCalc développé par Mariana Lucio (logiciel non publié).

#### 2.2.5. Alignement des données UHPLC-MS sur les données FT-ICR

L'alignement des données UHPLC-MS sur les données FT-ICR obtenus sur des jus de la population RixGw a été réalisé pour les analyses réalisées en ESI (+) et en ESI (-). Pour les

analyses par ESI (-), les données UHPLC-MS ESI (-) ont été préalablement recalibrées manuellement pour avoir une précision en masse proche de 1 ppm.

L'alignement des deux types de données a été effectué à l'aide d'un script R personnalisé, tenant en compte des éventuels adduits possibles ( $\text{Cl}^-$  pour ESI (-),  $\text{Na}^+$  et  $\text{NH}_4^+$  pour ESI (+)). Pour chaque  $m/z$  provenant des données UHPLC-MS et FTICR-MS, l'erreur de masse exacte a été calculée. Une correspondance a été retenue lorsque l'erreur était inférieure à 1 ppm.

L'alignement des deux types de données nous ainsi a permis d'obtenir (Figure 20):

- Une précision en masse  $< 0.1$  ppm nous permettant d'avoir une annotation du  $m/z$  par une formule brute,
- La séparation des isomères et la fragmentation de la molécule pour l'élucidation de la structure.

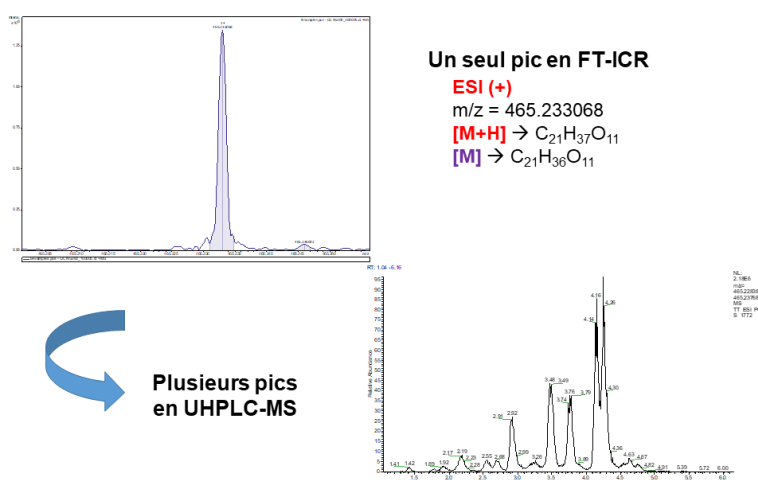


Figure 20 : Exemple d'un ion ( $m/z = 465.233$ ) aligné entre FT-ICR et UHPLC-MS ESI (+).

### 3. Génotypage de la population de croisement RixGw

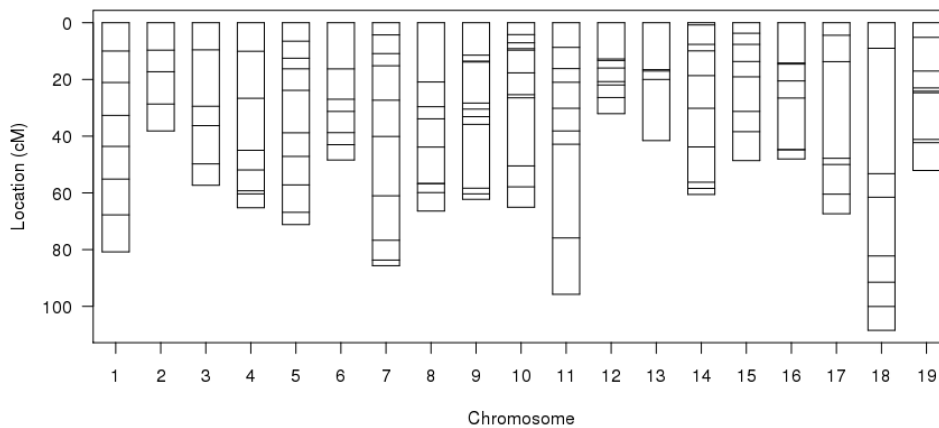
#### 3.1. Génotypage par marqueurs microsatellites

Les 117 individus de la population RixGw ont été précédemment génotypés à l'aide de marqueurs génétiques de type microsatellites, appelés aussi « Simple Sequence Repeat » (SSR). Un génotypage par marqueur SSR avait été réalisé avant ma thèse avec 153 marqueurs SSR (Duchêne et al., 2012b). Les séquences des amorces SSR utilisées dans cette étude sont disponibles dans la base de données NCBI UniSTS.

Quatre marqueurs génétiques ont été ajoutés à la carte initiale dont 2 marqueurs génétiques importants :

- **VVsc271** est un marqueur SSR sur le scaffold 271 du génome de référence PN40024 12x (<http://www.genoscope.cns.fr/cgi-bin/ggb/vg/s/12/12/gbrowse/vitis>). Les amorces utilisées pour amplifier cette région étaient VVsc271f (5'-CTTTCATGACAGATTGAGATTTACAACAAGA-3') et VVsc271r (5'-TAGCTCTTATGACAACCTTGAAAATGGAGAG-3').
- le gène **DXS1** (GSVIVT00017832001) dont le génotypage a été réalisé par «High Resolution Melting Assay» pour cartographier un polymorphisme de nucléotide simple [A / C] (SNP) en position 4789 dans la séquence du gène Gewurztraminer *DXS1* (Duchêne et al., 2009).

La carte consensus finale a été construite avec 157 marqueurs génétiques et couvre 1195 cM le long des 19 chromosomes du génome de la vigne (Figure 21).



**Figure 21 : Carte génétique SSR. Nombre de marqueurs : 157.**

Les caractéristiques du génotypage SSR selon les chromosomes sont résumées dans le Tableau 7. L'espace moyen entre marqueurs varie entre 4.6 à 15.5 cM et l'espace maximal entre marqueurs varie entre 12.8 et 44.2 cM selon les chromosomes.

Chromosome	Nombre de marqueurs génétiques	Taille (cM)	Espacement entre marqueurs moyen (cM)	Espacement entre marqueurs max (cM)
Chr 1	8	80.8	11.5	13
Chr 2	5	38.2	9.5	11.4
Chr 3	6	57.3	11.5	19.9
Chr 4	8	65.2	9.3	18.3
Chr 5	10	71.2	7.9	15
Chr 6	7	48.4	8.1	16.3
Chr 7	10	85.7	9.5	20.9
Chr 8	9	66.4	8.3	20.9
Chr 9	11	62.3	6.2	22.5
Chr 10	11	65.1	6.5	24
Chr 11	9	95.8	12	33.1
Chr 12	8	32.1	4.6	12.8
Chr 13	5	41.6	10.4	21.6
Chr 14	10	60.6	6.7	13.6
Chr 15	8	48.6	6.9	12.2
Chr 16	8	48	6.9	18.2
Chr 17	7	67.3	11.2	34.1
Chr 18	8	108.5	15.5	44.2
Chr 19	9	52.1	6.5	16.5
<b>Total</b>	157	1195	8.7	44.2

Tableau 7 : Récapitulatif de la carte génétique SSR.

La carte génétique SSR est peu dense, mais permet d'avoir les informations sur l'effet combiné des allèles provenant des deux parents.

### 3.2. Génotypage par puce Illumina 18k

Descartes génétiques parentales ont été construites avec les informations de Single Nucleotid Polymorphism (SNP) après le génotypage de la population par BeadChip Illumina Grape20K Infinium®, qui permet de révéler les SNP à 18071 positions dans le génome de la vigne (Le Paslier et al., 2013). Ce travail avait réalisé avant ma thèse. Pour résumer, sur 18071 SNP caractérisés pour la descendance et pour les génotypes parentaux, 8720 ne se sont pas homozygotes ou n'étaient pas exploitables. Les marqueurs avec plus de 30% de données manquantes, avec une ségrégation non mendélienne (test du Khi Deux,  $p < 0,05$ ) ou ne correspondant pas au génotype des parents ont été écartés. La phase des marqueurs SNP restants a été extraite à l'aide de JoinMap® 3.0 (Van Ooijen & Voorrips, 2001). Les cartes parentales



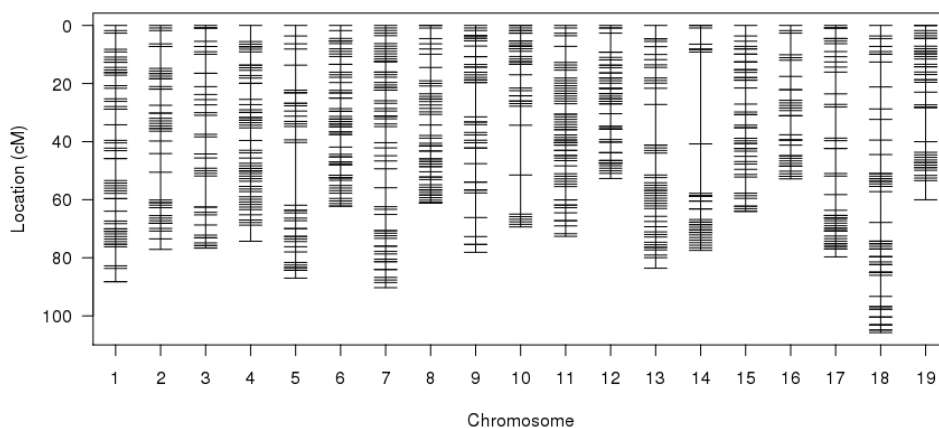
ont été construites en tenant compte uniquement des ségrégations parentales, comme dans deux rétrocroisements, et les marqueurs ont été codés en deux sous-ensembles correspondant à chaque parent avec un format “AA, AB” pour R / qtl (Broman et al., 2003). Les marqueurs en double ont été éliminés.

Le package R / qtl a été utilisé pour former les groupes de liaison, pour calculer les distances génétiques et pour vérifier les ordres des marqueurs. Les marqueurs avec une séparation <h<sub>k</sub>xh<sub>k</sub>> ont été supprimés avant le processus de calcul de la carte.

Les cartes SNP (Figure 22 et Figure 23) sont beaucoup plus denses que la carte SSR mais elles ne contiennent que les informations de ségrégation d’un des parents.

### 3.2.1. Carte génétique SNP\_Ri

Une carte SNP\_Ri avec 953 marqueurs SNP spécifiques du Ri a été construite avec les données de génotypage par puce Illumina (Figure 22). Cette carte couvre 1413cM sur les 19 chromosomes. L’espacement moyen entre marqueurs varie entre 0.9 à 2.2 cM et l’espacement maximal entre marqueurs varie entre 4.6 et 31.6 cM selon les chromosomes (Tableau 8).



**Figure 22:** Carte génétique SNP du Riesling (SNP\_Ri). Nombre de marqueurs : 953.

Chromosome	Nombre de marqueurs génétiques	Taille (cM)	Espacement entre marqueurs moyen (cM)	Espacement entre marqueurs max (cM)
Chr 1	55	88.3	1.6	7.6
Chr 2	44	77.1	1.8	9.6
Chr 3	38	76.7	2.1	10.7
Chr 4	58	74.3	1.3	5.6
Chr 5	40	87	2.2	21.6
Chr 6	64	62.2	1	4.4
Chr 7	62	90.3	1.5	6.5
Chr 8	55	61	1.1	4.6
Chr 9	51	78.1	1.6	11.8
Chr 10	45	69.4	1.6	17.2
Chr 11	63	72.6	1.2	5.6
Chr 12	57	52.7	0.9	6.6
Chr 13	47	83.6	1.8	14
Chr 14	36	77.4	2.2	31.6
Chr 15	49	64	1.3	5.6
Chr 16	34	52.8	1.6	7.4
Chr 17	50	79.7	1.6	10.7
Chr 18	60	105.8	1.8	10.6
Chr 19	45	60	1.4	11.8
<b>Total</b>	953	1413.1	1.5	31.6

Tableau 8 : Récapitulatif de la carte génétique SNP\_Ri.

### 3.2.2. Carte génétique SNP\_Gw

Une carte SNP\_Gw avec 809 marqueurs SNP spécifiques du Gw a été construite avec les données de génotypage par puce Illumina (Figure 23). Cette carte génétique couvre 1165 cM sur les 19 chromosomes. L'espacement moyen entre marqueurs varie entre 1.1 à 2 cM et l'espacement maximal entre marqueurs varie entre 4.6 et 21.6 cM selon les chromosomes

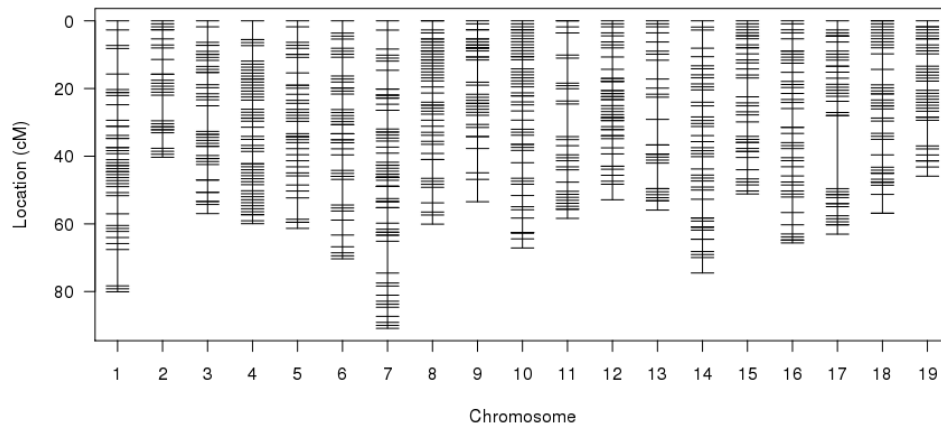


Figure 23 : Carte génétique SNP du Gewurztraminer (SNP\_GW). Nombre de marqueurs : 806.

Chromosome	Nombre de marqueurs génétiques	Taille (cM)	Espacement entre marqueurs moyen (cM)	Espacement entre marqueurs max (cM)
Chr 1	44	80.1	1.9	10.7
Chr 2	28	40.3	1.5	7.6
Chr 3	42	57	1.4	7.6
Chr 4	55	60	1.1	5.6
Chr 5	39	61.3	1.6	6.3
Chr 6	45	70.3	1.6	7.6
Chr 7	64	90.9	1.4	9.4
Chr 8	50	60.1	1.2	5.6
Chr 9	36	53.5	1.5	7.2
Chr 10	51	67.1	1.3	4.6
Chr 11	30	58.4	2	9.6
Chr 12	47	52.9	1.1	4.6
Chr 13	32	55.9	1.8	7.6
Chr 14	45	74.5	1.7	5.6
Chr 15	39	51.1	1.3	5.6
Chr 16	41	65.6	1.6	5.6
Chr 17	39	63	1.7	21.6
Chr 18	43	56.8	1.4	5.6
Chr 19	36	45.9	1.3	7.6
<b>Total</b>	<b>806</b>	<b>1164.8</b>	<b>1.5</b>	<b>21.6</b>

Tableau 9 : Récapitulatif de la carte génétique SNP\_GW.

#### 4. Recherche de « métabolite Quantitative Trait Loci » (mQTL) et recherche des gènes candidats associés

Les données issues d'analyses métabolomiques ciblées et non ciblées (GC-MS, UHPLC-MS à haute résolution Orbitrap et FT-ICR) de baies, de jus et de feuilles de 117 génotypes de la population ainsi que de données de génotypage (SSR et puce Illumina 18k) ont été utilisées pour la recherche de mQTL afin d'identifier des régions génomiques responsables de la variation des concentrations de molécules d'intérêt du métabolisme secondaire.

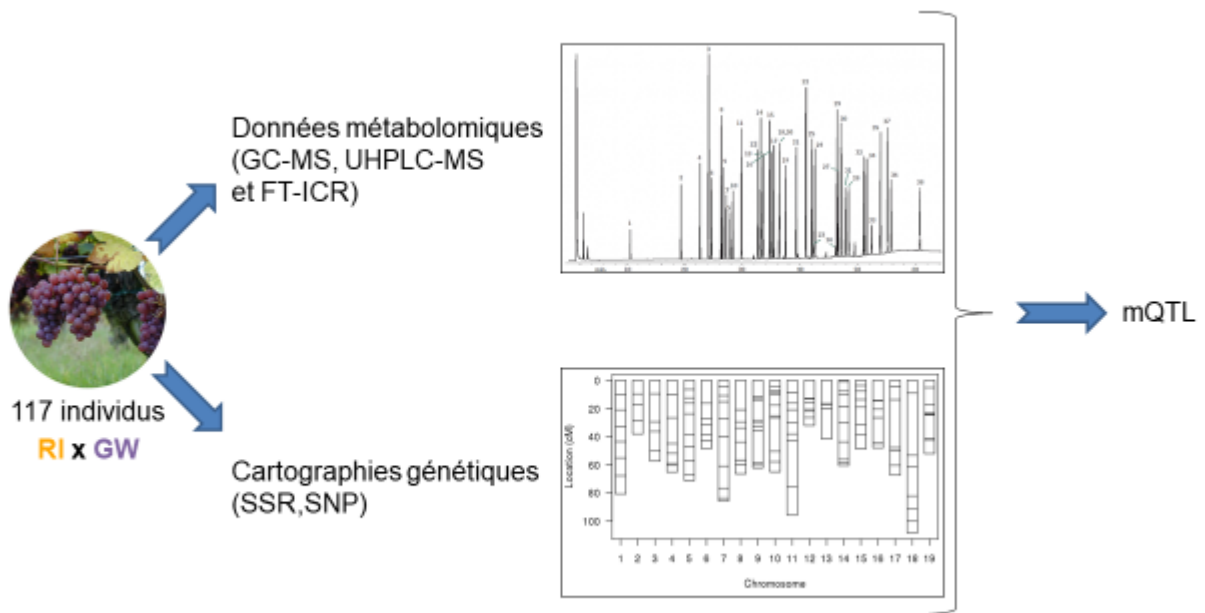


Figure 24 : Recherche de QTL métaboliques (mQTL).

La détection des QTL a été effectuée avec le package R/qtl du logiciel R. La vraisemblance de la présence d'un QTL a été évaluée par une valeur de LOD score. La probabilité (p-value) que cette valeur soit due au hasard a été déterminée en réalisant 1000 permutations aléatoires du jeu de données.

Les modèles de QTL ont été construits pas à pas après le raffinement de la position des QTL (*refineqtl*), la recherche de QTL supplémentaires (*addqtl*) et la recherche d'interactions entre QTL (*addint*). Le score LOD et le pourcentage de variance expliqué par un QTL dans un modèle QTL ont été évalués par analyse de variance en utilisant des sommes de carrés de type III (*fitqtl*). Les intervalles de confiance des QTL ont été calculés en intervalles de confiance

bayésiens (bayesian) avec une probabilité de couverture de 0,95. Comme la position des marqueurs SNP est incluse dans les noms des marqueurs, l'option «expand to marker = true» permettait un accès direct aux positions physiques des intervalles de confiance dans l'ensemble 12xV1 du génome de référence PN40024.

### 5. Identification des gènes candidat liés aux mQTL

Après l'analyse du QTL, un tri des QTL a été réalisé en fonction de 2 critères :

- L'impact potentiel de la molécule associée au QTL pour les arômes des vins,
- La vraisemblance de l'existence de ce QTL (LOD score, p-value),

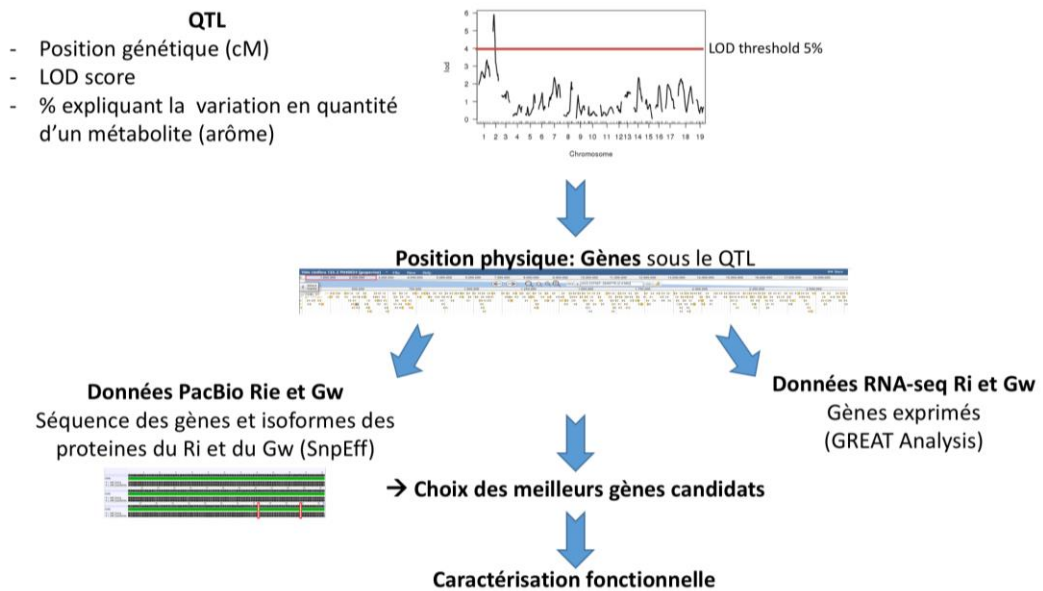


Figure 25: Stratégie de sélection des gènes candidats associés à un QTL.

Les QTL paraissant les plus intéressants ont été explorés afin d'identifier des gènes pouvant expliquer la variation en un ou plusieurs composés d'intérêt. Pour cela, différents outils ont été utilisés :

- L'utilisation de différentes données génomiques de référence et de données de transcriptomique des gènes situés sur l'intervalle de confiance
- La recherche des fonctions putatives des protéines codées par les gènes grâce à la recherche dans des banques de données de protéines

### 5.1. Génomes de référence et données de transcriptomique

#### 5.1.1. PN40024

PN40024 est un génotype issu de neuf autofécondations d'un clone de Pinot noir, homozygote à 93%, dont le génome a été séquencé en 2007 (Jaillon et al., 2007). Il a permis d'établir le génome de référence de la vigne. Nous sommes à la 3<sup>ième</sup> version de l'assemblage et à la 3<sup>ième</sup> version de son annotation du génome (Canaguier et al., 2017). La taille du génome est de 475 Mb dont environ 40% de séquences répétées, 19 chromosomes et 30 000 gènes annotés (Jaillon et al., 2007). Ce génome a été le modèle utilisé pour toutes les analyses bio-informatiques. J'ai utilisé la première version de l'annotation (CRIBI V1) du génome de référence pour mes analyses ([https://urgi.versailles.inra.fr/gb2/gbrowse/vitis\\_12x\\_pub/](https://urgi.versailles.inra.fr/gb2/gbrowse/vitis_12x_pub/)).

En complément au génome de référence PN40024, nous disposons des séquences des génomes complets des parents Ri et Gw, récemment séquencés et assemblés au laboratoire après un séquençage profond (150 x) grâce à la technologie Pacific Biosciences (PacBio) (Rhoads and Au, 2015). L'analyse des formes alléliques des gènes d'intérêt chez les parents Ri et Gw sont réalisées dans leur génome respectif.

Nous disposons également de données RNA-seq de baies de Ri et de Gw à différents stades de maturité : baies vertes, baies verte au cours de la véraison, baies molles au cours de la véraison et baies à mi-maturité. Ces données de transcriptomique nous ont permis d'analyser l'expression des gènes candidats dans les baies.

### 5.2. Fonctions putatives des protéines correspondantes aux gènes annotés dans les intervalles de confiance des QTL

Pour chaque QTL retenu, j'ai recensé les gènes présents dans l'intervalle de confiance du QTL à partir de l'annotation CRIBI V1 du génome de référence PN40024 (12xV2). J'ai ensuite réalisé un « pblast » en série des protéines correspondantes contre la base de données Swissprot afin d'obtenir des fonctions putatives les plus actualisées possible grâce à un script Perl (ncbiblast\_soaplite.pl) (Lopez et al., 2014). Les 5 meilleurs hits du pblast de chaque protéine ont été retenus afin de voir si la fonction putative d'une protéine parmi les protéines codées par les gènes compris dans l'intervalle de confiance des QTL pourrait participer à la voie de biosynthèse des composés d'intérêt.

Grâce aux données PacBio des génomes du Riesling et du Gewurztraminer, j'ai recherché s'il y avait des SNP différentiels dans la séquence des gènes candidats et j'ai vérifié si ceux-ci étaient exprimés ou non grâce aux données RNA-seq des baies de Riesling et Gewurztraminer. J'ai ainsi sélectionné les meilleurs gènes candidats pour un QTL sur la base de leur fonction putative et de leur profil d'expression, en vue de leur caractérisation fonctionnelle.

### 6. Clonage et caractérisation fonctionnelle de la CAMT

Pour le moment seule la caractérisation fonctionnelle d'un gène codant pour une Caffeic Acid O-Méthyl Transférase (CAMT) putative a été initiée. En effet, celui-ci constitue un gène candidat intéressant à caractériser car il se trouve dans l'intervalle de confiance d'un QTL de teneur en vanilline dans les baies de raisin.

#### 6.1. Clonage du gène par la technique Gateway

Le gène CAMT a été amplifié par PCR à partir d'ADNc issu de baies de Ri et de Gw. Un clonage pJet a ensuite été réalisé afin d'insérer l'amplicon d'ADNc de CAMT dans un vecteur pJet qui va permettre d'amplifier de manière efficace l'ADNc de la CAMT. A partir de la construction pJet, un clonage gateway a été réalisé afin de cloner l'amplicon dans deux types de vecteurs d'expression différents : pHNGWA, vecteur d'expression chez *Escherichia coli* (Busso et al., 2005) et pMDC32, vecteur d'expression chez les plantes via *Agrobacterium tumefaciens* (Curtis and Grossniklaus, 2003).

#### 6.2. Expression chez *Nicotiana benthamiana*

Le plasmide pMDC32 contenant le gène CAMT a été introduit dans *Agrobacterium tumefaciens*, puis la souche transformée a été infiltrée dans des plants de *Nicotiana benthamiana* (Voinnet, 2001). Pour ce faire, une colonie isolée d'agrobactéries transformées a été mise en culture dans 15 ml de milieu LB, contenant les antibiotiques adéquats, puis incubée une nuit, à 28 °C sous agitation. Le lendemain, la culture a été centrifugée 10 min à 5 000 rpm, puis le culot a été repris dans 40 ml de tampon d'infiltration (MgCl<sub>2</sub> 10 µM, acétosyringone 100 µM), avant d'être centrifugé 10 min à 5 000 rpm. Le culot a ensuite été repris dans un volume final choisi pour obtenir une DO<sub>600nm</sub> de 0,4 par construction infiltrée. Une souche d'*A. tumefaciens* contenant un plasmide permettant l'expression de la Green Fluorescent Protein (GFP) a été utilisée comme témoin. Les métabolites des feuilles agroinfiltrées ont ensuite été

extraits et analysés par UHPLC-MS. Des composés accumulés différentiellement en présence de la GFP et de la CAMT constituent des substrats ou des produits potentiels de cette enzyme.

### 6.3. Test enzymatique *in vitro* avec la protéine CAMT recombinante exprimée dans *E.coli*

Le plasmide pHNGWA contenant le gène CAMT a été introduit dans *E. coli* et la production de protéine recombinante a été induite par l'IPTG. Une colonie isolée d'*E. coli* BL21 DE3 transformée a été mise en pré-culture de 2 ml dans un milieu LB sélectif à 37 °C une nuit. Le lendemain, cette culture a servi à ensemencer 25 ml de milieu LB supplémenté des antibiotiques adéquats, et remise en culture à 37 °C jusqu'à l'obtention d'une DO<sub>600nm</sub> comprise entre 0,4 et 0,6. La synthèse de protéines recombinantes a été induite par 0,5 mM d'IPTG (Isopropyl-1-thio-β-D-galactopyranoside), après quoi la culture a été placée à 20 °C sous agitation, une nuit. La culture a été récoltée le lendemain pour une purification ou des tests enzymatiques. Les tests enzymatiques ont été réalisés avec les extraits bruts d'*E.coli* et l'enzyme CAMT recombinante purifiée, en présence de substrats potentiels à des concentrations à 1mM. Les produits de réaction ont été analysés par UHPLC-MS.

Molécule	Formule chimique
3,5-dihydroxyanisole (DHA)	C7H8O3
3-phenylphenol	C12H10O
4-phenylphenol	C12H10O
Caffeic acid	C9H8O4
Gallic acid monohydrate	C7H6O5
Ethyl gallate	C9H10O5
2-isobutyl-3-hydroxypyrazine	C8H12N2O
Esculetin 6,7-dihydroxycoumarin	C9H6O4
Resveratrol	C14H12O3
Olivetol	C11H16O2
Naringenin	C15H12O5
(+/-)-Catechin	C15H14O6
Kaempferol	C15H10O6
Quercetin dihydrate	C15H10O7
Myricetin	C15H10O8
Coniferyl alcool	C10H12O3
3-hydroxybenzaldehyde	C7H6O2
4-hydroxybenzaldehyde	C7H6O2
3,4-dihydroxybenzaldehyde	C7H6O3
3-Caffeoylquinic acid	C16H18O9
4-O-Caffeoylquinic acid	C16H18O9
5-O-Caffeoylquinic acid	C16H18O9

Tableau 10 : Substrats testés pour les tests enzymatiques *in vitro*.





**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

# Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

## 1. Introduction

Les arômes des raisins et des vins sont dus à des molécules directement volatiles, ou bien à des molécules présentes sous la forme de précurseurs liés à des sucres (glycosides) qui sont hydrolysés pendant la fermentation puis le vieillissement des vins. Des analyses des molécules volatiles présentes dans les raisins des individus de la population RixGw ont été réalisées au laboratoire par la technique de chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS). Ces analyses ont été réalisées sur des baies prélevées en 2014 et 2015, et concernent à la fois les composés volatils libres et glycosides libérés après hydrolyse enzymatique par Rapidase® AR2000. L'identification des molécules détectées est rendue possible par l'injection de standards, mais aussi grâce aux nombreuses banques de spectres de masses disponibles pour la GC-MS.

Grâce à une analyse conjointe de ces données analytiques des molécules volatiles et liées à des sucres par GC-MS, de données génétiques ainsi que de données génomiques, j'ai réalisé une analyse génétique de la biosynthèse des molécules volatiles des raisins. L'ensemble de ces résultats fait l'objet d'un article actuellement en fin de rédaction et qui sera soumis prochainement pour publication (Truong TT et al., Genetic analysis of volatile compounds biosynthesis reveals genomic hot spots determining aroma profiles of grape berries).

Pour ces travaux, les prélèvements des baies ont été réalisés par Nathalie Jaegli et Vincent Dumas, l'extraction des composés volatils par SPE a été réalisée par Nathalie Jaegli et Ariane Nansenet, le génotypage a été réalisé par Gisèle Butterlin, les injections par GC-MS ont été réalisées par Patricia Claudel et j'ai réalisé le traitement des données GC-MS : traitement de données par W4M, identification des molécules, recherche des QTL et des gènes candidats.

## **2. Article**

### **Genetic analysis of volatile compounds biosynthesis reveals genomic hot spots determining aroma profiles of grape berries**

**Truong T.T., Claudel P, Jaegli N, Baltenweck R, Butterlin G, Dumas V,  
Rustenholz C, Merdinoglu D, Hugueney P, Duchêne E.**

UMR1131 SVQV, University of Strasbourg - INRA, 68000 Colmar, France

## **Summary**

Grape berry aromas are crucial for wine quality and contribute significantly to the characteristic sensory profiles of wines produced using a given variety. However, the genetic determinism of grape berry aromas as well as their biosynthetic pathways are still largely unknown. To better understand the biosynthesis of grape berry aromas, berry volatile molecules were studied over 2 growing seasons in a progeny from a bi-parental cross between Riesling and Gewurztraminer. Each offspring was genotyped by simple sequence repeats markers and by hybridization on a 20k Grapevine Illumina® SNP chip. Segregations for the abundance of 96 molecules (free volatile or glycosides) were observed in the progeny. Quantitative Trait Loci (QTL) for the content of monoterpenols and their hydroxylated derivatives were detected, as well as QTL for other isoprenoids such as hemiterpenols, norisoprenoids and one sesquiterpene. In this study, we identify genomic hotspots on chromosome 5 and 10, with major impacts on 29 and 14 molecules respectively. We confirm a two-step model where a QTL on chromosome 5, linked to a deoxy-xylulose phosphate synthase gene (DXS), drives the total amount of isoprenoid synthesis and a QTL on chromosome 10, colocalizing with linalool synthase genes, allows the synthesis of linalool and its derivatives. Additional QTL for potential aroma compounds such as phenylpropanoids, fatty alcohols and lipoxygenase pathway products were detected.

Key words: Grapevine, QTL, berry aroma

# Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

## Introduction

Grapevine is one of the most widely cultivated fruit crop in the world and grapes are used for fresh consumption, grape juice, raisins, as well as for wine production. Viticulture and wine industry are currently facing two challenges: a high societal demand for reducing the use of pesticides, and a modification of environmental conditions due to climate change. Breeding new varieties is a promising way to improve the resistance to fungal diseases (Merdinoglu *et al.*, 2018) and to cumulate agronomic traits that can help in adapting to climate change, such as better water use efficiency or adapted phenology (Duchêne, 2016). Molecular markers are routinely used to screen progenies for resistance to downy or powdery mildew (Eibach *et al.*, 2007; Merdinoglu *et al.*, 2018) but efficient breeding of varieties adapted to local specifications requires molecular markers related to berry composition. Sugars, organic acids and potassium concentrations in grapevine berries are key determinant of alcohol content and acidity in wines but the pleasant experience of drinking a glass of wine would not exist without secondary metabolites that determine colour, structure and aroma profiles.

Aromas are volatile organic compounds belonging to several chemical families that may be already present in grape berries, *e.g.* terpenols, norisoprenoids or methoxypyrazines, but they also can originate from hydrolysis of precursors during wine making, as exemplified by volatile thiols (Roland *et al.*, 2011). In addition, wine aromas may be chemically released during wine ageing, as demonstrated for 1,1,6-trimethyl-1,2-dihydronaphthalene (TDN) (Daniel *et al.*, 2009) or wine lactone (Ilc *et al.*, 2017).

The large diversity of aroma profiles in wines produced all over the world depends on climate and terroir variability but the first choice made by winegrowers is the grapevine variety to be planted. Genetic differences within the *Vitis vinifera* species, both between varieties and clones, are widely reported for the terpenol content of the berries (Black *et al.*, 2015; Duchêne *et al.*, 2009a; Duchêne *et al.*, 2009b; Mateo and Jiménez, 2000), for methoxypyrazines

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

(Guillaumie *et al.*, 2013) or for sulfur compounds (Roland *et al.*, 2011). However, the knowledge about the genes involved and the molecular mechanisms responsible for differences of aroma potential between grapevine cultivars is still fragmentary (see Lin *et al.* (2019) for a comprehensive review). The genetic bases for cultivar-specific aroma profiles are completely elucidated in only two cases, namely the overall level of terpenol accumulation, and the synthesis of rotundone, a molecule responsible for the peppery aroma of wines made from Shiraz.

1-deoxy-D-xylulose 5-phosphate (DOXP) is the precursor of geranyl diphosphate (GPP) and geranyl geranyl diphosphate (GGPP), the substrates used by terpene synthases (VvTPS) to produce monoterpenes and sesquiterpenes. In aromatic genotypes, a mutation of a single base in the gene coding for the 1-deoxy-D-xylulose 5-phosphate synthase (*DXSI*) enables a higher synthesis of DOXP, and subsequently of GPP or GGPP (Battilana *et al.*, 2011; Dalla Costa *et al.*, 2018). The consequences of allelic variations in the *DXSI* gene were first demonstrated on the concentrations of monoterpenols such as geraniol, nerol, linalool, citronellol and  $\alpha$ -terpineol (Battilana *et al.*, 2009; Duchêne *et al.*, 2009a; Emanuelli *et al.*, 2010) and were further extended to linalool derivatives (Ilc *et al.*, 2017).

Rotundone is synthesized from  $\alpha$ -guaiene following a hydroxylation reaction catalyzed by the cytochrome P450 enzyme CYP71BE5 (Takase *et al.*, 2016). However, efficient  $\alpha$ -guaiene synthesis requires a specific allele of the TPS24 sesquiterpene synthase, that differs from the TPS24 protein sequence in the PN40024 genome by six amino acids. Site-directed mutagenesis confirmed the role of two mutations in the active site of the enzyme (Drew *et al.*, 2016).

Knowing the precise molecular mechanisms underlying genetic variation in aroma synthesis is however not necessary for breeding purposes. Quantitative trait loci (QTLs) detection can provide breeders with molecular markers surrounding the genomic regions of

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

interest. QTLs for aromas or aromas precursors were already published for the biosynthesis of monoterpenols. The main ones are located on chromosome 10 for linalool (Battilana *et al.*, 2009; Duchêne *et al.*, 2009a), chromosome 13 for  $\alpha$ -terpineol (Duchêne *et al.*, 2009a) or on chromosome 2 for carboxylinalool (Ilc *et al.*, 2017). The other well documented example is the biosynthesis of methoxypyrazines, which is controlled by five loci, including a locus on chromosome 3 influencing the expression of the *VvOMT3* gene encoding a pyrazine-specific O-methyltransferase (Guillaumie *et al.*, 2013).

The challenge in our context is to better understand the genetic determinism of traits specific of Riesling or Gewurztraminer in order to breed new grapevine varieties with desired organoleptic properties. Gewurztraminer wines are well known for their highly aromatic flavor, presumably linked to high amounts of terpenols (Duchêne *et al.*, 2009b; Furdikova *et al.*, 2017), volatiles thiols (Tominaga *et al.*, 2000), but also volatile phenols such as vinylguaiacol (Furdikova *et al.*, 2017; Versini, 1985). Riesling wines also contain terpenes (Friedel *et al.*, 2016; Luo *et al.*, 2019) but TDN, with its kerosene-like flavor, is the molecule that best characterizes these wines (Sacks *et al.*, 2012). More recently, Schüttler *et al.* (2015) demonstrated the contribution of volatile thiols to the typicality of Riesling wines.

Untargeted analytical strategies, together with modern data analysis procedures allow to access to the quantification and identification of an unprecedented number of molecules. In order to better understand the genetic determinism of grapevine aroma biosynthesis, we performed gas chromatography-mass spectrometry (GC-MS) untargeted analysis of berries collected in the progeny from a bi-parental cross between Riesling (Ri) and Gewurztraminer (Gw), over two growing seasons. Genotypic data of the individuals, obtained from simple sequence repeats (SSR) markers and single nucleotide polymorphisms (SNP) were used for genetic map construction and QTL detection. We show here that allelic variations at a small number of loci exert major impacts on grape aroma profile and we provide new hypotheses



## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

explaining variations of the concentrations of potentially important volatile molecules in grapevine berries.

### **Material and Methods**

#### **Plant material**

A grapevine population consisting of the progeny from a cross between *V. vinifera* cv Riesling (Ri) clone 49 (female) and *V. vinifera* cv Gewurztraminer (Gw) clone 643 (male) (Ri x Gw) was studied. 383 genotypes from the Ri x Gw cross were grafted onto the Couderc 161-49 rootstock and three plants per genotype were planted in an experimental vineyard in Bergheim (48°210 N, 7°340 E) in 2006. The veraison dates (onset of the ripening process) were calculated for each individual plot, after successive scorings, as the dates when 50% of the berries were soft. Three replicates of the parent varieties Ri and Gw (3 plants per replicate) at different locations in the experimental plot were included in the sampling protocol.

#### **Berry sampling and extraction of potentially volatile compounds (PVCs) by solid phase extraction (SPE)**

Seventy grams of maturing berries were collected from 117 genotypes of the progeny and three replicates for each parent variety during the 2014 and 2015 growing seasons. The sampling date was determined for each genotype as the date when a cumulative heat sum of 350 degree-days (calculated with maximum temperatures, base temperature of 10 °C) after veraison was reached. This procedure was validated previously for assessing genotypic differences for sugar accumulation capacity (Duchêne *et al.*, 2012) and berry acidity (Duchêne *et al.*, 2014). Some of the berries were used for sugar and acid analysis and a sample was frozen at -20°C before GC-MS analysis.

Forty grams of berries were deseeded and crushed under liquid nitrogen. The powder was suspended in a 1 g l<sup>-1</sup> Na<sub>2</sub>SO<sub>3</sub> solution. 20 µl of a 1 g l<sup>-1</sup> 3-octanol solution in ethanol was

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

added as external standard for the free fraction. After centrifugation (25 min, 16 500 g at 4°C), the supernatant was passed through a glass fibre prefilter and a glass filter. The filtrate was passed through a 1 g C18 silica-bonded nonpolar solid phase extraction (SPE) cartridge (Bond Elut Jr., Agilent technologies), previously rinsed with 5 ml of ethanol and 15 ml of ultrapure water, at a rate of *ca.* 1 drop s<sup>-1</sup>. After washing with 15 ml of ultrapure water, compounds of interest were eluted with 4 ml of dichloromethane to release the free fraction. A second elution was performed on the same cartridge with 4 ml of ethanol to obtain the glycosylated fraction. This extract was diluted in 40 ml of a citrate/phosphate buffer (pH 4.5) and the resulting solution was incubated with 50 mg of AR2000 glycolytic enzyme (Gist-Brocades, Seclin, France) overnight at 37.5 °C to release the glycosidically bound molecules (Tamborra et al. 2004). After addition of 20 µl of a 1 g l<sup>-1</sup> solution of 4-nonanol, the released non-polar molecules were separated again by SPE on a C18 column and eluted with 4 ml of dichloromethane after rinsing with water.

The two extracts from the same original sample (bound and free fractions) were dried in Pasteur pipettes filled with 0.5 g of anhydrous sodium sulphite and concentrated to 500 µl under a gentle nitrogen flux. Twenty µl of a 1g l<sup>-1</sup> solution of m-cresol was added to each sample as internal control. Samples were stored at -20 °C prior to GC-MS analysis

### **GC-MS profiling, molecule identification and semi-quantification**

The analysis of the volatile compounds was carried out using an Agilent GC6890N gas chromatograph equipped with a split/splitless injector and coupled to an Agilent 5975 mass spectrometer. The separation of 1 µl of extract was carried out on a DB-Wax column (30 m \* 0.32 mm, film thickness 0.5 mm, JW Scientific) using helium as carrier gas at a column flow rate of 1.5 ml min<sup>-1</sup>. Separation was operated with the following parameters : injector at 250 ° C in pulsed splitless mode (25 psi for 0.5min), oven temperature rising from 45 ° C to 82 ° C

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

at a rate of 20 ° C min<sup>-1</sup> (held for 1 min) then up to 235 ° C with a rate of 3 ° C min<sup>-1</sup> and held at the final temperature for 10 min. The transfer line was maintained at 270 ° C.

The mass spectrometer was used in electronic impact mode at 70 eV and positive ions were recorded with a scan range from m/z 30 to m/z 400. The temperatures of the source and the quadrupole analyser were respectively 230 °C and 150 °C. The ChemStation software (G1701DA, Rev. D.03.00) was used for instrument control and data acquisition.

GC-MS raw data files from Ri, Gw and RixGw berries extracts were pre-processed with the *metaMS.runGC* tool available on the “Workflow4Metabolomics” online platform (W4M, workflow4metabolomics.org) (Giacomoni *et al.*, 2014) with the following parameters : fwhm = 5, rtdiff=0.1, minfeat=5, simthreshold=0.7, minclassfraction=0.1, minclasssize=3. Annual datasets were preprocessed separately, but, for each year, chromatograms for free and bound extracts were aligned together.

The identities of the detected molecules were determined by comparing their mass spectra with those of authentic standards and NIST14 mass spectral library. Match thresholds of spectra were at least 85% and linear retention indices (LRI) were considered for compound identification.

Quantitation of 3-octanol, 4-nonanol, m-cresol, linalool,  $\alpha$ -terpineol, citronellol, nerol and geraniol were was achieved by using calibration curves with known amounts of authentic standards (Sigma-Aldrich) and normalization to the sample weight and peak area of the internal standard

### **Preliminary statistical analyses**

Analysis of variance using the tree replicates of Ri and Gw berries were performed with the “lm” procedure in the R “stats” package (R Core Team, 2019) for all the quantified molecules. The broad sense heritabilities of the traits within the progeny were calculated as

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

$\sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$ .  $\sigma_e^2$  was extracted from the anovas with the parent varieties.  $\sigma_g^2$  was calculated as the difference between the total variability observed in the progeny,  $\sigma_t^2$ , and  $\sigma_e^2$ .

### Genetic markers and QTL detection

For the detection of quantitative trait loci (QTL), three genetic maps were used (details in Duchêne *et al.* (2019)): a consensus map with 157 simple sequence repeats markers (SSR) and a SNP for the *DXSI* gene, that covers 1195.0 cM along the 19 chromosomes of the grapevine genome, and two parental maps, *i.e.* a Riesling map and a Gewurztraminer map with respectively 953 and 806 single nucleotide polymorphism (SNP) markers. As previously described (Ilc *et al.*, 2017), two genetic markers related to aroma biosynthesis were added to the initial maps. VVsc271 is a simple sequence repeat (SSR) on scaffold 271 in the PN40024 reference genome (<http://www.genoscope.cns.fr/cgi-bin/ggb/vitis/12X/gbrowse/vitis>). Primers used to amplify this region were

VVsc271f : 5' -CTTTCATGACAGATTGAGATTTACAACAAG-3' and

VVsc271r : 5' -TAGCTCTTATGACAACCTTGAAAATGGAG CC-3'.

Scaffold 271 was chosen because several terpene synthases were annotated in this region (Martin *et al.*, 2010). The *DXSI* gene (GSVIVT00017832001) was mapped by 'High Resolution Melting Assay' (Wittwer *et al.*, 2003) for a [A/C] Single Nucleotide Polymorphism (SNP) at position 4789 in the GwDXS1 gene sequence (Duchêne *et al.*, 2009a).

Only molecules that were detected in both years of grape aroma PVCs analysis were kept for QTL detection. QTL detection was performed on the three genetic maps with the R/qtl software (Broman *et al.*, 2003) using the multiple imputation method ('draws' = 128) and the one-dimension scan command 'scanone'. LOD significances were ensured with permutation tests (1000 permutations). Confidence intervals were calculated as Bayesian credible intervals (bayesint) with a probability of coverage of 0.95. The physical positions of the confidence

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

intervals were assessed according to the physical positions of the genetic markers in the genome.

### **Search for candidate genes**

An exploration of the possible candidate genes for the main QTL was performed using the smallest confidence interval for each QTL, *i.e.* the region that was common to all the datasets. The physical positions of these intervals were determined in the 12X.v2 assembly of the PN40024 reference genome (Canaguier et al., 2014) using the coordinates of the SNPs of the Illumina 18k chip. The positions of the genes in the confidence intervals were directly extracted from the genome browser ([https://urgi.versailles.inra.fr/jbrowse/gmod\\_jbrowse](https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse)). The protein sequences were submitted to UniProtKB/Swiss-Prot and the results of the five best hits were compiled in one table per confidence interval.

## **Results**

### **Untargeted analysis of PVCs in berries of Ri, Gw and genotypes from the RixGw cross population**

GC-MS analyses were performed for two consecutive years on free and bound fraction of berries extracts. In our study, each aglycone detected in the free fraction is considered as one volatile molecule.

Untargeted analysis was first performed using the W4M workflow with raw GC-MS data from samples collected in 2014. The molecules detected after this step were searched again in 2015 GC-MS data. Out of seventy volatile compounds identified and quantified in 2014, fifty-three were detected and quantified again in 2015, and were selected for further genetic analyses (Table 1). Several chemical families were represented: fatty alcohols, lipoxigenase pathway products, phenylpropanoids, norisoprenoids, sesquiterpenes, terpenols, hemiterpenols, and one aldehyde ((E,E)-2,4-heptadienal). Among these compounds, forty-three volatile

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

molecules were detected both in free and bound forms. Three volatile molecules were detected only in the free fraction: vanillin (#44), benzeneacetaldehyde (#13) and (E,E)-2,4-heptadienal (#9) , while seven volatile molecules were detected only following hydrolysis of glycosydes: 1-nonanol (#14), (Z)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#16), lilac alcohol formate (#27), ethyl cinnamate (#29), 8-hydroxycarvotanacetone (#37), 3-hydroxy- $\beta$ -damascone (#43) and dihydro-3-oxo- $\beta$ -ionol (#50).

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

Peak#	Compound	RI	RT (min)	CAS	Identification	Fraction
1	(E)-2-Hexenal	1234	5.4	6728-26-3	A	F+H
2	Isoprenol	1258	5.9	763-32-6	A	F+H
3	Prenol	1332	7.6	556-82-1	A	F+H
4	1-Hexanol	1365	8.4	111-27-3	A	F+H
5	(Z)-3-Hexen-1-ol	1396	9.3	928-96-1	A	F+H
6	(E)-2-Hexen-1-ol,	1419	10.0	928-95-0	A	F+H
7	(E)-Linalool oxide (furanoid)	1456	11.2	11063-78-8	A	F+H
8	(Z)-Linalool oxide (furanoid)	1483	12.1	34995-77-2	A	F+H
9	(E,E)-2,4-Heptadienal	1503	12.8	05/03/4313	A	F
10	Linalool	1560	14.8	78-70-6	A	F+H
11	1-Octanol	1571	15.2	111-87-5	A	F+H
12	Hotrienol	1622	17.1	53834-70-1	A	F+H
13	Benzeneacetaldehyde	1649	18.0	122-78-1	A	F
14	1-Nonanol	1673	19.1	143-08-8	A	H
15	$\alpha$ -Terpineol	1707	20.4	98-55-5	A	F+H
16	(Z)-2,6-Dimethyl-3,7-octadiene-2,6-diol	1713	20.7	13741-21-4	B	H
17	(E)-Linalool oxide (pyranoid)	1748	22.0	10448-31-4	A	F+H
18	(Z)-Linalool oxide (pyranoid)	1771	23.0	14009-71-3	A	F+H
19	Citronellol	1777	23.1	106-22-9	A	F+H
20	Lilac alcohol C	1793	23.8	33081-36-6	A	F+H
21	Nerol	1809	24.4	106-25-2	A	F+H
22	Lilac alcohol D	1835	25.3	33081-37-7	A	F+H
23	Geraniol	1857	26.2	106-24-1	A	F+H
24	2-Hydroxy-1,8-cineole	1866	26.5	60761-00-4	B	F+H
25	Benzyl alcohol	1880	27.0	100-51-6	A	F+H
26	Phenylethyl Alcohol	1914	28.3	60-12-8	A	F+H
27	Lilac alcohol formate	1951	29.6	115-99-1	B	H
28	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol	1955	29.9	51276-34-7	B	F+H
29	Ethyl cinnamate	2127	35.8	103-36-6	A	H
30	2,6-Dimethyl-1,7-octadiene-3,6-diol	2135	36.2	51276-33-6	B	F+H
31	Eugenol	2166	37.1	97-53-0	A	F+H
32	4-Vinylguaiaicol	2190	38.1	7786-61-0	A	F+H
33	3,7-Dimethyl-1,7-octanediol	2213	38.8	107-74-4	B	F+H
34	2,6-Dimethyl-1,7-octadien-3-ol	2215	38.9	22460-59-9	B	F+H
35	(Z)-8-Hydroxylinalool	2278	40.9	103619-06-3	B	F+H
36	(E)-8-Hydroxylinalool	2318	42.2	75991-61-6	A	F+H
37	8-Hydroxycarvotanacetone	2329	42.5	7712-46-1	B	H
38	Terpene (n.i.)	2336	42.8	/	C	F+H
39	Geranic acid	2381	43.3	459-80-3	A	F+H
40	Terpene (n.i.)	2401	44.8	/	C	F+H
41	Terpene (n.i.)	2469	46.8	/	C	F+H
42	Isopregol	2508	48.1	7786-67-6	B	F+H
43	3-Hydroxy- $\beta$ -damascone	2536	48.8	102488-09-5	B	H
44	Vanillin	2559	49.3	121-33-5	B	F
45	Methyl vanillate	2600	50.6	3943-74-6	B	F+H
46	Acetovanillone	2630	51.5	498-02-2	A	F+H
47	(Z)-8-Hydroxygeraniol	2633	51.7	026488-97-1	B	F+H
48	Sesquiterpene (n.i.)	2647	52.2	/	C	F+H
49	Dihydro- $\beta$ -ionone	2658	52.5	17283-81-7	A	F+H
50	Dihydro-3-oxo- $\beta$ -ionol	2666	52.7	27185-79-1	B	H
51	3-Oxo-7,8-dihydro- $\alpha$ -ionol	2692	53.4	36151-02-7	B	F+H
52	Zingerone	2764	55.8	122-48-5	B	F+H
53	Homovanillin alcohol	2807	57.3	2380-78-1	B	F+H

**Table 1:** Volatile components identified by GC-MS in berries collected in the RixGw progeny

Peak numbers, compound names, retention indices (RI), retention times (RT), and CAS numbers are indicated. RI were determined relative to n-alkanes (C8–C28). Identification methods were as follows: (A) identities confirmed by comparison of mass spectrum and retention time with those of authentic standards, (B) identities tentatively assigned by comparison with the NIST14 mass spectral library or literature, (C) unidentified compound. The compounds were detected either in the free fraction (F), the hydrolyzed fraction (H), or in both fractions (F+H)

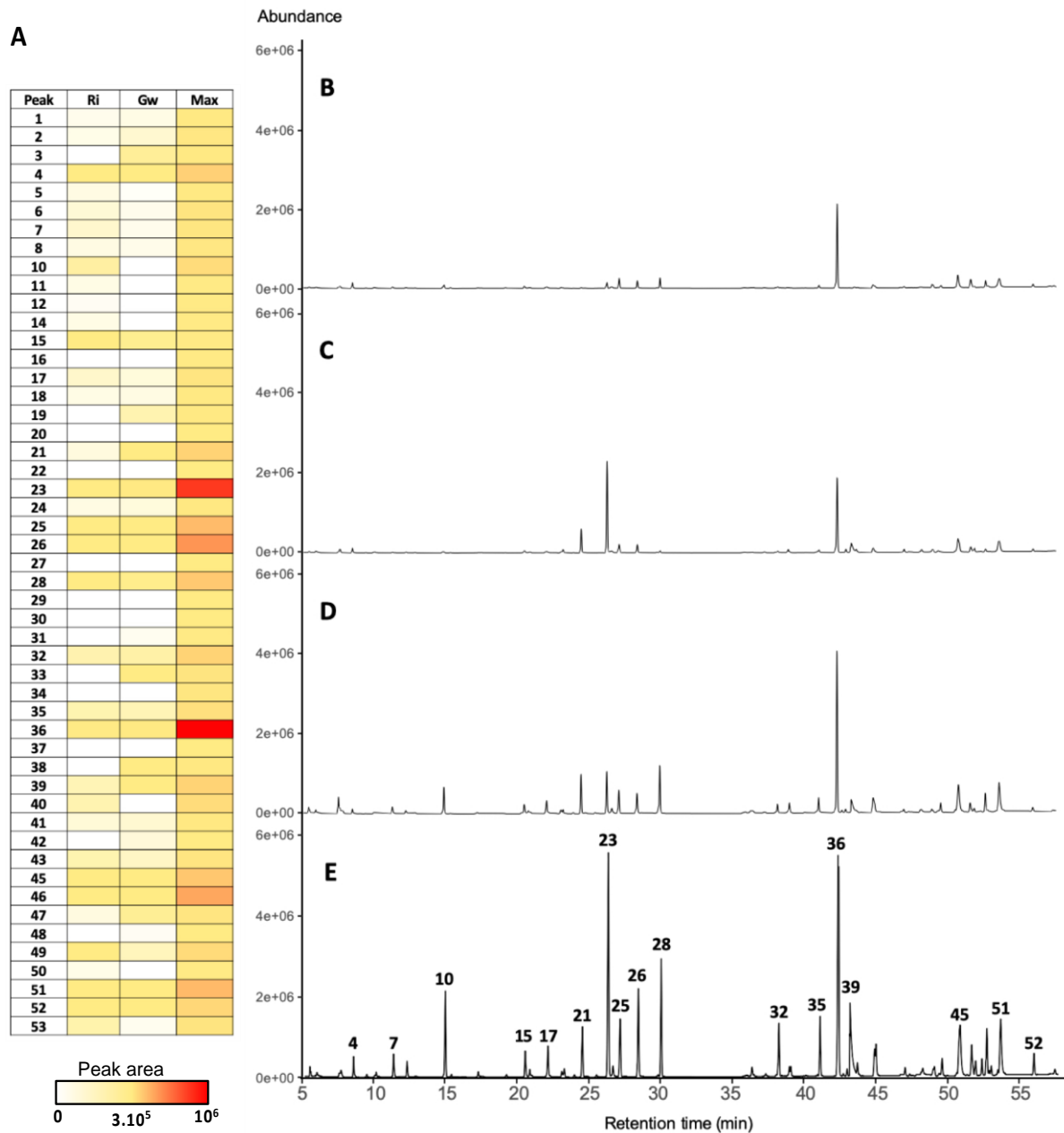
## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

Out of the fifty-three volatile molecules, thirty-two molecules were differentially accumulated between Ri and Gw (with  $p \leq 0.05$ , Supplementary Table S1). For example, fatty alcohols were almost absent in Gw and norisoprenoids such as dihydro- $\beta$ -ionone (#49), 3-hydroxy- $\beta$ -damascone (#43) and dihydro-3-oxo- $\beta$ -ionol (#50) were found in higher amounts in Ri. Conversely, a number of monoterpenes such as geraniol (#23), nerol (#21), geranic acid (#39) and citronellol (#19) were found in much higher amounts in Gw than in Ri. However, amounts of linalool (#10) and some of its derivatives (#7, #8) were higher in Ri than in Gw. Ri contained slightly more phenylpropanoids (#44, #53) and there were no clear differences for lipoxygenase pathway products (#4, #5, #6).

The amounts of all the quantified molecules segregated in the progeny, with heritabilities ranging from 0.24 to 0.99 (Supplementary Table S1). Crossing Ri and Gw created a large genetic variability of PVCs profiles. Indeed, some molecules could not be quantified in the parent varieties but were detected in high amounts in some genotypes of the progeny (#29, #30, #34, lilac alcohols #20, #22, #27, #37) (Fig. 1A). Comparison of GC-MS profiles showed clear transgressive segregations, with molecules found in off spring in much higher amounts than in the parent lines. Some genotypes, such as 74E (Fig. 1D), had clearly a more complex profile of PVCs than its parents (Fig. 1B, Fig. 1C). Among genotypes in the RixGw progeny, increased profile complexity and enhanced PVCs accumulation affected all chemical families, including isoprenoids (#10, #15, #17, #21, #23, #36), phenylpropanoids (#25, #26, #32, #45, #52) or C6-compounds (#4). (Fig. 1E),



# Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS



**Fig. 1:** Diversity of PVCs profiles in the RixGw progeny

(A) Heat map comparing mean peak areas (n=3) of identified compounds (hydrolyzed fraction) in Ri and Gw, with the highest value of all the genotypes of the RixGw progeny (Max). Peak numbers correspond to compounds listed in Table 1.

GC-MS profiles of berry PVCs (hydrolyzed fraction) from Ri (B), Gw (C), genotype 74E (D) and synthetic profile with the highest value at each retention time of all the genotypes of the RixGw progeny (E).

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

### QTL analysis of PVCs accumulation in the RixGw population

QTL detection was performed with three genetic maps (SSR consensus, SNPs Gw and SNPs Ri) on four datasets: for the free and the hydrolyzed fractions, and for two vintages, 2014 and 2015. A LOD threshold corresponding to a genome-wide p-value of 0.05 was determined for each molecule in every dataset after performing 1000 permutations. Only QTLs with a LOD score above this threshold are presented in this manuscript. In total, 283 QTLs were detected: 57 with the Ri parental map, 106 with the Gw parental map and 120 with the consensus map. QTL were detected for each of the 53 quantified volatile molecules, at least for one year, either in the free or in the glycosylated fraction. 109 QTL were detected both in 2014 and 2015 for 51 molecules (Supplementary Table S2).

#### *QTLs for phenylpropanoids*

Even if Riesling wines are not characterized by a high content in phenolic compounds, QTLs for homovanillin alcohol, zingerone and 4-vinylguaiacol were detected with the Ri map for both vintages on chromosomes 3, 12 and 13, respectively. QTLs for acetovanillone, benzyl alcohol, ethyl cinnamate and phenylethyl alcohol were found with the Gw map, on chromosomes 6, 7 and 8.

#### *QTLs for lipoxygenase pathway products*

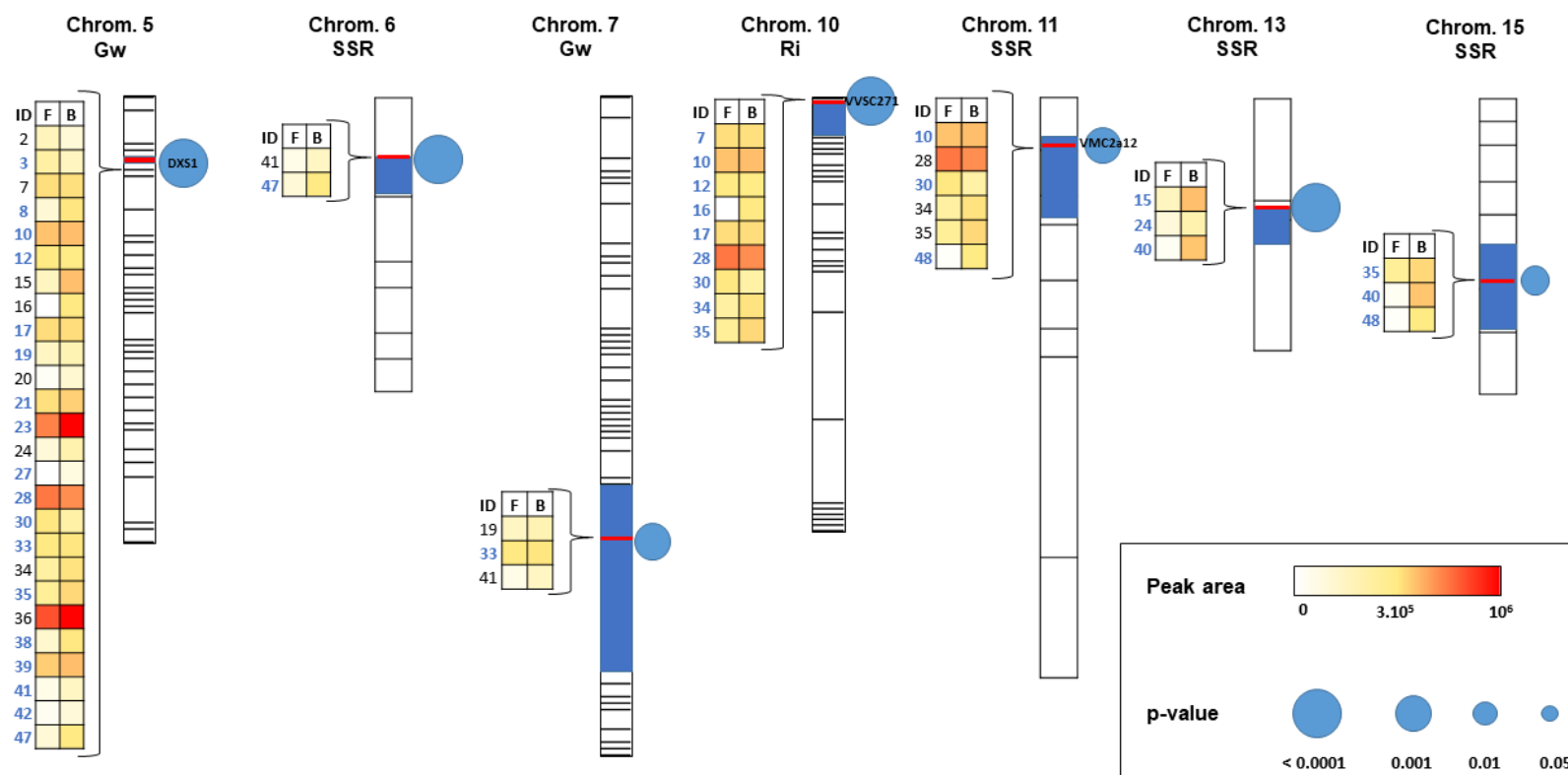
The fatty acid-derived alcohols 1-octanol and 1-nonanol were only detected bound to glycosides and their concentrations depended on a QTL on chromosome 1 in the Ri map and a QTL on chromosome 18 in the Gw map. The concentrations of (E)-2-hexen-1-ol, both free and glycosylated, were related to genetic variations on chromosome 7, whereas concentrations of free (Z)-3-hexen-1-ol were related to genetic variations on chromosome 13, in both cases only in the Gw map.

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

### *Isoprenoids: monoterpenols, norisoprenoids and sesquiterpenes*

Major QTLs on chromosomes 5, 10 and 11 impacted the concentrations of large numbers of PVCs. These loci had reproducible impacts on 29, 14 and 3 molecules, respectively (Supplementary Table S2, Fig. 2.). On chromosome 5, allelic variations in the Gw genome impacted a large range of molecules of the isoprenoid family, including hemiterpenols and the dihydro- $\beta$ -ionone norisoprenoid. This is consistent with the role of *DXS1* as the first gene of the plastidial isoprenoid biosynthetic pathway. On chromosome 10, a QTL was associated to the level of linalool biosynthesis, impacting a number of linalool derivatives, and hotrienol. A QTL on chromosome 11 had a similar effect. Additional QTLs on chromosome 4, 6, 7, 13 and 15 had effects on a lower number of compounds. The QTL on chromosome 13 impacted three terpenols but predominantly  $\alpha$ -terpineol concentration. Two QTL for an unidentified sesquiterpene (#48) were found both in 2014 and 2015 on chromosomes 11 and 15, in the same genomic region as QTLs for monoterpenols on chromosome 15.

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS



**Fig. 2:** Characteristics of QTL with a major impact on volatile terpenes concentrations.

The red line shows the position where the LOD score was the highest. The blue areas represent the confidence intervals. Impacted terpene compounds are indicated with their peak number (Table 1), those in blue are those with a reproducible QTL on both years of analysis. Relative abundance of the terpene compounds is given as mean peak area of free (F) and bound forms (B) in 2015, according to the color scale bar.

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

### **Candidate genes**

An exploration of the possible candidate genes for the main QTL was performed using the smallest confidence interval for each QTL, *i.e.* the region that was common to all the datasets. These intervals contained between 84 and more than 600 genes, depending on the power of the QTLs to explain the variations of amounts in the progeny. Several genes that may participate in the biosynthesis pathways of the quantified PVCs were selected in these intervals, based on their putative function (Supplementary Table S3), and their relevance will be discussed below.

### **Discussion**

#### **The RixGw population exhibits an increased PVCs diversity in comparison to parental lines**

This study is up to date the more complete one for the number of quantified PVCs in a grapevine segregating progeny or genetic association population. Previous works on grapevine berry secondary metabolism only focused on terpenols (Battilana *et al.*, 2009; Duchêne *et al.*, 2009a; Yang *et al.*, 2017b ) or on non-volatile molecules such as polyphenols (Fournier-Level *et al.*, 2009; Huang *et al.*, 2012). Berries from the genotypes of the RixGw progeny contained a large diversity of molecules from different classes, whose amounts segregated in the progeny: terpenols, sesquiterpenes, norisoprenoids, phenylpropanoids, fatty acids and C6-compounds. Most of these molecules were detected in higher amounts in the progeny than in parental lines, showing the effect of combination of two genetic backgrounds (Fig. 1). Some molecules were even not detected in the parents berries but only in the progeny. A classical definition of heterosis is that when crossing two lines, of maize for example, F1 hybrids have greater biomass and yield potential than the two parents. The molecular basis of the phenomenon is still unclear

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

(Swanson-Wagner *et al.*, 2006) but one of the first working hypothesis was that F1 hybrids can cumulate complementary alleles from each parent. The definition of heterosis does not strictly apply in our case because our parent varieties are not homozygous lines. However, the result is similar, with genotypes from the progeny accumulating higher amounts of aroma compounds than the parents. This unexpected result shows that sexual reproduction can give birth to completely new varieties, combining for example the acidity and low sugar content of Ri berries, the pink colour of Gw berries, and innovative aromatic profiles. There is currently a high societal demand for new grapevine varieties, which are resistant to fungal diseases and require less fungicide protection. Current breeding programs constitute therefore an opportunity to dramatically expand the grapevine phenotypic space and, in particular, to explore unknown aromatic horizons.

### **The *DXS1* gene: a key gene responsible for major difference in isoprenoid content between Gw and Ri and within the RixGw progeny**

A major QTL on chromosome 5 was reported in many studies on the genetic variability of isoprenoid content in several genetic backgrounds (Battilana *et al.*, 2009; Doligez *et al.*, 2006; Duchêne *et al.*, 2009a; Yang *et al.*, 2017d) and a molecular marker for breeding purposes was already developed (Emanuelli *et al.*, 2014). This QTL has been associated to *DXS*, which is the first enzyme in the biosynthesis of isopentenyl-pyrophosphate (IPP), the skeleton of all the isoprenoids molecules. A specific allelic *DXS1* allele from Muscato Bianco has been shown to possess enhanced enzymatic activity (Battilana *et al.*, 2011; Dalla Costa *et al.*, 2018). The association of a specific *DXS1* form (*DXS1+* allele) from Gw with high isoprenoid synthesis was detected in a self-progeny of Gw, but also in a genetic association study (Dalla Costa *et al.*, 2018; Emanuelli *et al.*, 2010). A R306C mutation in the *DXS1* protein in Gw is responsible for the enhanced enzymatic activity in this variety (Huguency *et al.*, 2012). We confirm here

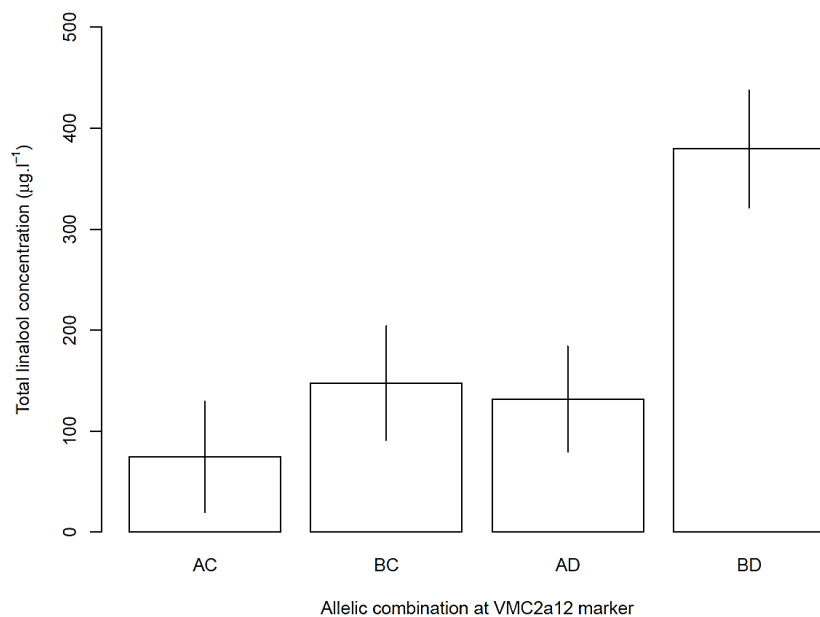
## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

the architecture proposed for the level of terpenol synthesis in a Muscat Ottonel self-progeny (Duchêne *et al.*, 2009a): a favorable allele (*DXSI+*) at the *DXSI* locus on chromosome 5 is necessary for a global high level of isoprenoid synthesis but without a favorable allele on chromosome 10 (*LIS+*), the synthesis of linalool is low, and only geraniol is produced (Fig. 3B). Conversely, combination of (*DXSI+*) from Gw with (*LIS+*) from Ri results in the accumulation of high amounts of linalool in berries, reminiscent of aromatic profiles usually found in Muscat varieties. Therefore, a cross between Ri and Gw can give rise to new genotypes with unexpected Muscat aroma profiles, showing that it possible to breed ‘Muscat-like’ varieties without any direct contribution of a Muscat variety.

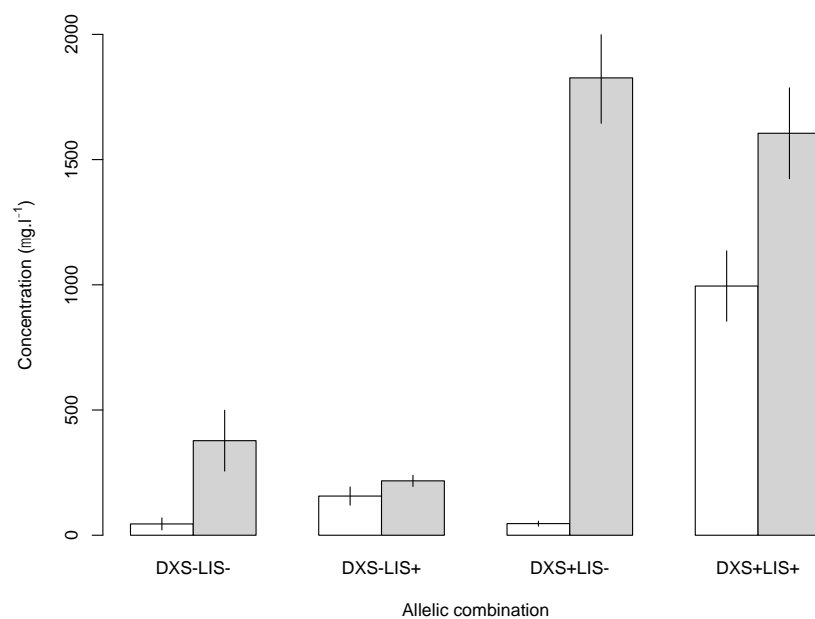
In addition to its impact on monoterpenols, we show here that the presence of the *DXSI+* allele in a genotype can increase the concentrations of at least 28 molecules in the berries (Fig. 2). This is a clear example showing that a very simple genetic variation of one nucleotide can generate specific complex aroma profiles in grape berries and consequently so-called “varietal” aromas in the corresponding wines. This result adds some consistency to the concept of “varietal bouquet”.

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

**A**



**B**



**Fig. 3:** Combined effect of QTLs on chromosomes 5, 10 and 11 on terpenol profiles in the RixGw progeny

**A.** Concentrations of linalool (sum of free and bound linalool) in the RixGW progeny berries in 2015. AC, BC, AD, BD are allelic combination at VMC2a12 (11@8.7) from Ri (A,B) x Gw (C,D)

**B.** Concentrations of geraniol (grey) and linalool (white) (sum of free and bound compounds) in the RixGW progeny berries in 2014. DXS -/+ is the genotype at the DXS1 locus, on chromosome 5, in the Gw genome; LIS -/+ is the genotype at VVSC271 on chromosome 10 in the Ri genome.



## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

### Linalool synthases genes (LIS) are strong candidates associated to a major QTL on chromosome 10

A major QTL for linalool amounts was already reported on chromosome 10 in segregating populations where one of the parent belonged to the Muscat family (Battilana *et al.*, 2009; Duchêne *et al.*, 2009a) but also in a germplasm collection (Dalla Costa *et al.*, 2018). Here we show that the concentrations of eight other linalool derivatives are impacted by this locus, which add complexity to the aroma profiles. Linalool synthases (LIS) genes are the most likely candidate genes controlling the biosynthesis of linalool. Out of 33 potential LIS genes identified in the reference genome, functional activity of the corresponding proteins was demonstrated for only six of them (Martin *et al.*, 2010). However, the structure of the grapevine genomic regions containing LIS genes is unresolved in the reference genome and we can demonstrate, using a self-progeny of PN40024, that some of these regions are heterozygous (Supplementary Table 4). This result can explain why scaffold 271, where 3 LIS sequences were annotated (Martin *et al.*, 2010), is still in the 'Unknown' chromosome in the 12x.2 version of the reference genome ([https://urgi.versailles.inra.fr/jbrowse/gmod\\_jbrowse](https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse)) (Supplementary Table S3).

Nevertheless, we designed the SSR marker VVSC271 on scaffold 271 (Ilc *et al.*, 2017) and we showed that VVSC271 collocates with the QTL for linalool on chromosome 10 and is close to VrZAG64, a marker previously associated with the variations of the linalool content in berries (Duchêne *et al.*, 2009a). Four putative LIS genes, VvTPS55, VvTPS65 and VvTPS56, and VviTPS144 are annotated on scaffold 271. Two alleles of the VvTPS56 gene from Pinot Noir and one allele from Cabernet Sauvignon were shown encode functional LIS enzymes (Martin *et al.*, 2010). In addition, the LIS activity of a TPS56 allele from Ri was confirmed (Zhu *et al.*, 2014), indicating that the genomic region linked to VVSC271 contains a functional LIS allele, which is a very strong candidate explaining variations of linalool biosynthesis.

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

VVSC271 can be already considered as a reliable marker for breeding purposes but future work will be needed to determine the precise genomic positions of *LIS* genes on chromosome 10, together with the detailed description of the allelic variations necessary to obtain high linalool amounts. Finally, whether Ri and varieties of the Muscat family share common alleles, or common molecular mechanisms controlling linalool synthesis remains to be further investigated as well.

### **New QTLs for linalool and linalool oxides on chromosomes 11 and 15: a role for cytochromes P450?**

A QTL impacting linalool content was previously described on chromosome 15 in a Muscat Ottonel self-progeny (Duchêne *et al.*, 2009a), but no such QTL was stable in the present study. However, QTLs for the linalool oxide (Z)-8-hydroxylinalool (#35) and for an unidentified sesquiterpene (#40) were detected on chromosome 15. Cytochrome P450 enzymes belonging to the CYP76 family have been shown to be involved in the oxidative metabolism of linalool in *Arabidopsis* flowers (Ginglinger *et al.*, 2013) and in grapes (Ilc *et al.*, 2016). Cytochrome P450 genes are organized in clusters in the grapevine genome (Ilc *et al.*, 2018) and exploring the confidence interval for the QTL for (Z)-8-hydroxylinalool provided 21 relevant P450 candidate genes, including genes in the CYP76 family (Supplementary Table S3).

A QTL on chromosome 11 impacted the amounts of both linalool (#10) and the linalool oxide 2,6-dimethyl-1-7-octadiene-3,6-diol (#28). Similarly, a cytochrome P450 was also found to colocalize with this QTL. Higher linalool concentrations were observed only in genotypes that combine two recessive alleles from the parents (Fig. 3A). As a P450 enzyme using linalool as a substrate would lower its concentration, a likely scenario would be that higher linalool concentrations could result from the combination in the same genotype of two P450 alleles, from Ri and Gw, with reduced enzymatic activities. This ‘aroma trap’ effect, would be in line

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

with observations that glycoside transferases, by sequestering PVCs, can reduce the pool of free aromas (Yauk *et al.*, 2014).

### **An $\alpha$ -terpineol synthase gene collocates with a QTL for $\alpha$ -terpineol and 2-hydroxy-1,8-cineole**

A QTL on chromosome 13 explained up to 35% of the variations of  $\alpha$ -terpineol and 31% of 2-hydroxy-1,8-cineole contents. A QTL colocalizing with an  $\alpha$ -terpineol synthase (ATS) gene characterized previously (GenBank accessions AY572896 and AY572987, Martin and Bohmann (2004) was already described in a self-progeny of Muscat Ottonel (Duchêne *et al.*, 2009a) and was confirmed by a genetic association study (Yang *et al.*, 2017c). This ATS is also able to synthesize 1-8 cineole (Martin and Bohmann, 2004), the precursor of 2-hydroxy-1,8-cineole. Moreover, a parallel evolution between the level of expression of an ATS gene and  $\alpha$ -terpineol concentration in grapevine berries was also observed (Matarese *et al.*, 2014). The ATS gene is found in the confidence interval for  $\alpha$ -terpineol and 2-hydroxy-1,8-cineole synthesis on chromosome 13, but together with twelve other TPS genes with similarity to ATS (Supplementary Table S3). A detailed analysis of the haplotypes of the parent varieties at this locus will be necessary before considering functional analysis studies.

### **Lipoxygenase pathway products and fatty alcohols**

The degradation of lipids such as linoleic acid and  $\alpha$ -linolenic acid by the lipoxygenase pathway produces C6 or C9 compounds (Ameye *et al.*, 2018). C6 compounds appearing during the crushing of the grapes (Podolyan *et al.*, 2010) are often associated with “green leaf” fragrances (Ameye *et al.*, 2018; Green *et al.*, 2011). In wines, they are indicators of a lack of maturity of the grapes at harvest (Ribéreau-Gayon *et al.*, 2006), but can contribute, to some

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

extent to the typicality of the aroma profiles of the wines (Green *et al.*, 2011). More importantly, Harsch *et al.* (2013) demonstrated that (E)-2-hexen-1-ol is a precursor of 3-sulfanyl hexanol, a volatile thiol of major interest both in Gw and Ri wines (Schüttler *et al.*, 2015; Tominaga *et al.*, 2000). We have identified QTLs in the Gw map for (E) 2-hexen-1-ol on chromosome 7 and (Z)-3-hexen-1-ol on chromosome 13. (E) 2-hexen-1-ol can result from the degradation of membrane lipids after the successive actions of lipases, lipoxygenases (LOX), hydroperoxide lyases (HPL), (3Z):(2E)-enal isomerases and alcohol dehydrogenases (ADH) (Ameys *et al.*, 2018). Hydroperoxide lyases (HPL) are members of the cytochrome P450 CYP74 family (Feussner and Wasternack, 2002). (Z)-3-hexen-1-ol can be synthesized from (E)-2-hexen-1-ol by a (3Z):(2E)-enal isomerase (Lin *et al.*, 2019).

None of the *Vitis vinifera* lipoxygenase genes identified by Podolyan *et al.* (2010) or CYP74s annotated by Ilc *et al.* (2018) were found in the confidence intervals of the QTLs on chromosome 7 and 13. However, three esterase/lipase genes colocalizing with the QTL on chromosome 13 (Supplementary Table S3) could participate to the genetic variations for (E)-2-hexen-1-ol concentrations.

Although the relationships between (E)-2-hexen-1-ol concentrations in grapes and concentrations of volatile thiols is not strongly established, the QTLs detected for C6-compounds open new perspectives for further investigations on the genetic determinism of the presence of volatile thiols in wines.

The fatty alcohols 1-nonanol and 1-octanol can be found in grapes (Bahena-Garrido *et al.*, 2019) but also in wines (Crandles *et al.*, 2015; Soares *et al.*, 2015). In table grapes, 1-nonanol and 1-octanol may contribute to “rose-orange” and “jasmine-lemon” flavours, respectively (Wu *et al.*, 2016). Although our local experience of sensory tests of the pure molecules is positive, their contributions to wine aroma profiles is uncertain. They likely result from the degradation of

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

lipids, and lipases coding genes, as for C6 compounds, and relevant candidate genes could be found on the confidence intervals on chromosomes 1 and 18 (Supplementary Table S3).

### QTLs for phenylpropanoids

Phenylpropanoids such as 4-vinylguaiacol (2-methoxy-4-vinylphenol) and 4-vinylphenol participate in the spicy aromas of beers (Langos and Granvogl, 2016) and 4-vinylguaiacol was also proposed to be responsible for the spicy aroma of Gw wines (Grando *et al.*, 1993; Versini, 1985). However, at high concentrations, a volatile phenol such as 4-ethylphenol is associated with unpleasant flavours resulting from wine contamination by *Brettanomyces/Dekkera spp* (Loureiro and Malfeito-Ferreira, 2003). Although 4-vinylphenol and 4-vinylguaiacol can be found in grapes (Cabrita *et al.*, 2006), the likely precursors of vinylphenol and vinylguaiacol in wines are coumaric acid and ferulic acid, respectively, that undergo decarboxylation by specific yeast strains of the Pof+ (phenolic off-flavour) type (Grando *et al.*, 1993; Langos and Granvogl, 2016). GC is not the appropriate method to detect and quantify cinnamic acids but QTLs for 4-vinylguaiacol, acetovanillone, benzyl alcohol, cinnamic acid ethyl ester, homovanillyl alcohol, phenylethyl alcohol and zingerone were detected on six chromosomes in the RixGw progeny. These compounds are found in grapes and wines (Ilc *et al.*, 2016; Schueuermann *et al.*, 2016) and can play a role in aroma profiles: zingerone is one of the main compounds of ginger aroma (Menon *et al.*, 2007), cinnamic acid ethyl ester can be associated to a “cinnamon” descriptor, phenylethyl alcohol to a “rose” descriptor in wines (Rutan *et al.*, 2014) and acetovanillone to “vanilla” (Caldeira *et al.*, 2016). Even though these phenolic compounds likely have a positive impact on grape flavours, their precise contribution to the aroma profiles of wines after the fermentation process is to be demonstrated.

Confidence intervals for the two major QTLs, on chromosomes 6 and 12, include glucosyltransferase genes, which is in accordance with the fact that cinnamic acid ethyl ester

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

and zingerone were present only as glycosylated forms in grape extracts. Interestingly, QTLs were detected for vanillin and some of its derivatives. Combined genetic and functional analyses may help deciphering the vanillin biosynthesis pathway, which is still a matter of debate (Gallage *et al.*, 2014; Yang *et al.*, 2017a).

### **Conclusion**

The untargeted analysis of PVCs in the RixGw progeny allowed to quantify 96 molecules, either free or bound to a glycoside. A lot of them were terpenols, norisoprenoids or sesquiterpenes but segregations for phenylpropanoids and lipoxygenase pathway products were also observed. Reproducible QTLs were detected for most of them, including QTLs for phenylpropanoids and lipoxygenase pathway products, which were not reported before. The observation that polymorphism at a small number of genomic hotspots can impact a large number of molecules supports the concept of ‘varietal aroma’. Indeed, the characteristic ‘bouquet’ linked to a variety is more likely the consequence of the perception of dozens of molecules than of a single one. The transgressive segregations observed shows that sexual reproduction can generate new and unexpected aroma profiles, which open new perspectives for expanding the world of grape aroma. A striking result is that, the combination of a specific DXS allele from Gw and a specific genomic region that contains linalool synthase genes on chromosome 10 from Ri results in genotypes with novel Muscat-like aroma profiles. A two-step model for a high level of linalool synthesis was confirmed and the precise molecular tools available for tracking the favorable alleles open new opportunities for breeding programs. The next step is to complete the GC-MS analyses by liquid chromatography – mass spectrometry analyses (LC-MS) and to produce wines with subset of genotypes to confirm the sensory interest of some of the new QTLs. At last, this study provides new candidate genes for functional genomic studies. A more precise knowledge on the molecular mechanisms of aroma

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

biosynthesis can help in anticipating the consequences of higher temperatures or water stress during the ripening period, and consequently to prepare the adaptation of viticulture to climate change.

### References

- Ameje M, Allmann S, Verwaeren J, Smagghe G, Haesaert G, Schuurink RC, Audenaert K.** 2018. Green leaf volatile production by plants: a meta-analysis. *New Phytologist* **220**, 666-683.
- Bahena-Garrido SM, Ohama T, Suehiro Y, Hata Y, Isogai A, Iwashita K, Goto-Yamamoto N, Koyama K.** 2019. The potential aroma and flavor compounds in *Vitis* sp. cv. Koshu and *V. vinifera* L. cv. Chardonnay under different environmental conditions. *Journal of the Science of Food and Agriculture* **99**, 1926-1937.
- Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Grando MS.** 2009. The 1-deoxy-D-xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theoretical and Applied Genetics* **118**, 653-669.
- Battilana J, Emanuelli F, Gambino G, Gribaudo I, Gasperi F, Boss PK, Grando MS.** 2011. Functional effect of grapevine 1-deoxy-D-xylulose 5-phosphate synthase substitution K284N on Muscat flavour formation. *Journal of Experimental Botany* **62**, 5497-5508.
- Black CA, Parker M, Siebert TE, Capone DL, Francis IL.** 2015. Terpenoids and their role in wine flavour: recent advances. *Australian Journal of Grape and Wine Research* **21**, 582-600.
- Cabrita MJ, Freitas AMC, Laureano O, Stefano RD.** 2006. Glycosidic aroma compounds of some Portuguese grape cultivars. *Journal of the Science of Food and Agriculture* **86**, 922-931.
- Caldeira I, Santos R, Ricardo-da-Silva JM, Anjos O, Mira H, Belchior AP, Canas S.** 2016. Kinetics of odorant compounds in wine brandies aged in different systems. *Food Chemistry* **211**, 937-946.
- Crandles M, Reynolds AG, Khairallah R, Bowen A.** 2015. The effect of yeast strain on odor active compounds in Riesling and Vidal blanc icewines. *Lwt-Food Science and Technology* **64**, 243-258.
- Dalla Costa L, Emanuelli F, Trenti M, Moreno-Sanz P, Lorenzi S, Coller E, Moser S, Slaghenauffi D, Cestaro A, Larcher R, Gribaudo I, Costantini L, Malnoy M, Grando MS.** 2018. Induction of Terpene Biosynthesis in Berries of Microvine Transformed with *VvDXS1* Alleles. *Frontiers in plant science* **8**.
- Daniel MA, Capone DL, Sefton MA, Else GM.** 2009. Riesling acetal is a precursor to 1,1,6-trimethyl-1,2-dihydronaphthalene (TDN) in wine. *Australian Journal of Grape and Wine Research* **15**, 93-96.
- Doligez A, Audiot E, Baumes R, This P.** 2006. QTLs for muscat flavor and monoterpene odorant content in grapevine (*Vitis vinifera* L.). *Molecular Breeding* **18**, 109-125.
- Drew DP, Andersen TB, Sweetman C, Moller BL, Ford C, Simonsen HT.** 2016. Two key polymorphisms in a newly discovered allele of the *Vitis vinifera* TPS24 gene are responsible for the production of the rotundone precursor alpha-guaiene. *Journal of Experimental Botany* **67**, 799-808.
- Duchêne E.** 2016. How can grapevine genetics contribute to the adaptation to climate change? *Oeno-one* **50**, 113-124.

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

- Duchêne E, Butterlin G, Claudel P, Dumas V, Jaegli N, Merdinoglu D.** 2009a. A grapevine (*Vitis vinifera* L.) deoxy-d-xylulose synthase gene colocalizes with a major quantitative trait loci for terpenol content. *Theoretical and Applied Genetics* **118**, 541-552.
- Duchêne E, Dumas V, Butterlin G, Jaegli N, Rustenholz C, Chauveau A, Bérard A, Le Paslier M-C, Gaillard I, Merdinoglu D.** 2019. Genetic variations of acidity in grape berries are controlled by the interplay between organic acids and potassium. *Theoretical and Applied Genetics*.
- Duchêne E, Dumas V, Jaegli N, Merdinoglu D.** 2012. Deciphering the ability of different grapevine genotypes to accumulate sugar in berries. *Australian Journal of Grape and Wine Research* **18**, 319-328.
- Duchêne E, Dumas V, Jaegli N, Merdinoglu D.** 2014. Genetic variability of descriptors for grapevine berry acidity in Riesling, Gewürztraminer and their progeny. *Australian Journal of Grape and Wine Research* **20**, 91-99.
- Duchêne E, Legras JL, Karst F, Merdinoglu D, Claudel P, Jaegli N, Pelsy F.** 2009b. Variation of linalool and geraniol content within two pairs of aromatic and non aromatic grapevine clones. *Australian Journal of Grape and Wine Research* **15**, 120-130.
- Eibach R, Zyprian E, Welter L, Töpfer R.** 2007. The use of molecular markers for pyramiding resistance genes in grapevine breeding. *Vitis* **46**, 120-124.
- Emanuelli F, Battilana J, Costantini L, Le Cunff L, Boursiquot JM, This P, Grando MS.** 2010. A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biology* **10**, 241.
- Emanuelli F, Sordo M, Lorenzi S, Battilana J, Grando MS.** 2014. Development of user-friendly functional molecular markers for gene conferring muscat flavor in grapevine. *Molecular Breeding* **33**, 235-241.
- Feussner I, Wasternack C.** 2002. The lipoxygenase pathway. *Annual Review of Plant Biology* **53**, 275-297.
- Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet JM, Cheynier V, This P.** 2009. Quantitative Genetic Bases of Anthocyanin Variation in Grape (*Vitis vinifera* L. ssp *sativa*) Berry: A Quantitative Trait Locus to Quantitative Trait Nucleotide Integrated Study. *Genetics* **183**, 1127-1139.
- Friedel M, Frotscher J, Nitsch M, Hofmann M, Bogs J, Stoll M, Dietrich H.** 2016. Light promotes expression of monoterpene and flavonol metabolic genes and enhances flavour of winegrape berries (*Vitis vinifera* L. cv. Riesling). *Australian Journal of Grape and Wine Research*.
- Furdikova K, Bajnociova L, Malik F, Spanik I.** 2017. Investigation of volatile profile of varietal Gewürztraminer wines using two-dimensional gas chromatography. *Journal of Food and Nutrition Research* **56**, 73-85.
- Gallage NJ, Hansen EH, Kannagara R, Olsen CE, Motawia MS, Jorgensen K, Holme I, Hebelstrup K, Grisoni M, Moller BL.** 2014. Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nature Communications* **5**, 4037.
- Giacomoni F, Le Corguillé G, Monsoor M, Landi M, Pericard P, Pétéra M, Duperier C, Tremblay-Franco M, Martin J-F, Jacob D, Goulitquer S, Thévenot EA, Caron C.** 2014. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* **31**, 1493-1495.
- Ginglinger JF, Boachon B, Hofer R, Paetz C, Kollner TG, Miesch L, Lugan R, Baltenweck R, Mutterer J, Ullmann P, Beran F, Claudel P, Verstappen F, Fischer MJC, Karst F, Bouwmeester H, Miesch M, Schneider B, Gershenzon J, Ehrling J, Werck-Reichhart D.** 2013. Gene



## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in Arabidopsis flowers. *The Plant Cell* **25**, 4640-4657.

**Grando MS, Versini G, Nicolini G, Mattivi F.** 1993. Selective use of wine yeast strains having different volatile phenol production. *Vitis* **32**, 43-50.

**Green JA, Parr WV, Breitmeyer J, Valentin D, Sherlock R.** 2011. Sensory and chemical characterisation of Sauvignon blanc wine: Influence of source of origin. *Food Research International* **44**, 2788-2797.

**Guillaumie S, Ilg A, Rety S, Brette M, Trossat-Magnin C, Decroocq S, Leon C, Keime C, Ye T, Baltenweck-Guyot R, Claudel P, Bordenave L, Vanbrabant S, Duchene E, Delrot S, Darriet P, Huguene P, Gomes E.** 2013. Genetic analysis of the biosynthesis of 2-methoxy-3-isobutylpyrazine, a major grape-derived aroma compound impacting wine quality. *Plant Physiology* **162**, 604-615.

**Harsch MJ, Benkwitz F, Frost A, Colonna-Ceccaldi B, Gardner RC, Salmon JM.** 2013. New precursor of 3-mercaptohexan-1-ol in grape juice: thiol-forming potential and kinetics during early stages of must fermentation. *Journal of Agricultural and Food Chemistry* **61**, 3703-3713.

**Huang YF, Doligez A, Fournier-Level A, Le Cunff L, Bertrand Y, Canaguier A, Morel C, Miralles V, Veran F, Souquet JM, Cheynier V, Terrier N, This P.** 2012. Dissecting genetic architecture of grape proanthocyanidin composition through quantitative trait locus mapping. *BMC Plant Biology* **12**, 30.

**Huguene P, Duchêne E, Merdinoglu D.** 2012. 1-deoxy-D-xylulose 5-phosphate synthase alleles responsible for enhanced terpene biosynthesis In: *GénoPlante-Valor*, ed, Vol. WO 2012/052171 A1. France, 73.

**Ilc T, Arista G, Tavares R, Navrot N, Duchene E, Velt A, Choulet F, Paux E, Fischer M, Nelson DR, Huguene P, Werck-Reichhart D, Rustenholz C.** 2018. Annotation, classification, genomic organization and expression of the *Vitis vinifera* CYPome. *PLoS One* **13**, e0199902.

**Ilc T, Halter D, Miesch F, Lauvoisard F, Kriegshauser L, Ilg A, Baltenweck R, Huguene P, Werck-Reichhart D, Duchêne E, Navrot N.** 2017. A grapevine cytochrome P450 generates the precursor of wine lactone, a key odorant in wine. *New Phytologist* **213**, 264-274.

**Ilc T, Werck-Reichhart D, Navrot N.** 2016. Meta-Analysis of the Core Aroma Components of Grape and Wine Aroma. *Frontiers in plant science* **7**, 1472.

**Langos D, Granvogl M.** 2016. Studies on the Simultaneous Formation of Aroma-Active and Toxicologically Relevant Vinyl Aromatics from Free Phenolic Acids during Wheat Beer Brewing. *Journal of Agricultural and Food Chemistry* **64**, 2325-2332.

**Lin J, Massonnet M, Cantu D.** 2019. The genetic basis of grape and wine aroma. *Horticulture Research* **6**.

**Loureiro V, Malfeito-Ferreira M.** 2003. Spoilage yeasts in the wine industry. *International Journal of Food Microbiology* **86**, 23-50.

**Luo JQ, Brotchie J, Pang M, Marriott PJ, Howell K, Zhang PZ.** 2019. Free terpene evolution during the berry maturation of five *Vitis vinifera* L. cultivars. *Food Chemistry* **299**, 13.

**Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, Lund ST, Bohlmann J.** 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biology* **10**, 226.

**Martin DM, Bohlmann J.** 2004. Identification of *Vitis vinifera* (-)-alpha-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry* **65**, 1223-1229.

## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

- Matarese F, Cuzzola A, Scalabrelli G, D'Onofrio C.** 2014. Expression of terpene synthase genes associated with the formation of volatiles in different organs of *Vitis vinifera*. *Phytochemistry* **105**, 12-24.
- Mateo JJ, Jiménez M.** 2000. Monoterpenes in grape juice and wines. *Journal of Chromatography A* **881**, 555-567.
- Menon AN, Padmakumari KP, Kutty BS, Sumathikutty MA, Sreekumar MM, Arumugham C.** 2007. Effects of processing on the flavor compounds of Indian fresh ginger (*Zingiber officinale* Rosc.). *Journal of Essential Oil Research* **19**, 105-109.
- Merdinoglu D, Schneider C, Prado E, Wiedemann-Merdinoglu S, Mestre P.** 2018. Breeding for durable resistance to downy and powdery mildew in grapevine. *OENO One* **52**, 203-209.
- Podolyan A, White J, Jordan B, Winefield C.** 2010. Identification of the lipooxygenase gene family from *Vitis vinifera* and biochemical characterisation of two 13-lipoxygenases expressed in grape berries of Sauvignon Blanc. *Functional Plant Biology* **37**.
- R Core Team.** 2019. R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing.
- Ribéreau-Gayon P, Dubourdieu D, Donèche B, Lonvaud A.** 2006. *Handbook of Enology Volume 1. The Microbiology of Wine and Vinifications. 2nd Edition*: John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,.
- Roland A, Schneider R, Razungles A, Cavelier F.** 2011. Varietal thiols in wine: discovery, analysis and applications. *Chemical Reviews* **111**, 7355-7376.
- Rutan T, Herbst-Johnstone M, Pineau B, Kilmartin PA.** 2014. Characterization of the Aroma of Central Otago Pinot noir Wines Using Sensory Reconstitution Studies. *American Journal of Enology and Viticulture* **65**, 424-434.
- Sacks GL, Gates MJ, Ferry FX, Lavin EH, Kurtz AJ, Acree TE.** 2012. Sensory Threshold of 1,1,6-Trimethyl-1,2-dihydronaphthalene (TDN) and Concentrations in Young Riesling and Non-Riesling Wines. *Journal of Agricultural and Food Chemistry* **60**, 2998-3004.
- Schueuermann C, Khakimov B, Engelsen SB, Bremer P, Silcock P.** 2016. GC-MS Metabolite Profiling of Extreme Southern Pinot noir Wines: Effects of Vintage, Barrel Maturation, and Fermentation Dominate over Vineyard Site and Clone Selection. *Journal of Agricultural and Food Chemistry* **64**, 2342-2351.
- Schüttler A, Friedel M, Jung R, Rauhut D, Darriet P.** 2015. Characterizing aromatic typicality of Riesling wines: merging volatile compositional and sensory aspects. *Food Research International* **69**, 26-37.
- Soares RD, Welke JE, Nicolli KP, Zanús M, Caramão EB, Manfroí V, Zini CA.** 2015. Monitoring the evolution of volatile compounds using gas chromatography during the stages of production of Moscatel sparkling wine. *Food Chemistry* **183**, 291-304.
- Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS.** 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F-1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6805-6810.
- Takase H, Sasaki K, Shinmori H, Shinohara A, Mochizuki C, Kobayashi H, Ikoma G, Saito H, Matsuo H, Suzuki S, Takata R.** 2016. Cytochrome P450 CYP71BE5 in grapevine (*Vitis vinifera*) catalyzes the formation of the spicy aroma compound (-)-rotundone. *Journal of Experimental Botany* **67**, 787-798.
- Tominaga T, Baltenweck-Guyot R, Peyrot des Gachons C, Dubourdieu D.** 2000. Contribution of volatile thiols to the aromas of white wines made from several *Vitis vinifera* grape varieties. *American Journal of Enology and Viticulture* **51**, 178-181.

## **Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

- Versini G.** 1985. [On the aroma of 'Traminer aromatico' or 'Gewuerztraminer' wine.]. *Vignevine* **12**, 57-65.
- Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ.** 2003. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical Chemistry* **49**, 853-860.
- Wu Y, Duan S, Zhao L, Gao Z, Luo M, Song S, Xu W, Zhang C, Ma C, Wang S.** 2016. Aroma characterization based on aromatic series analysis in table grapes. *Scientific Reports* **6**, 31116.
- Yang H, Barros-Rios J, Kourteva G, Rao X, Chen F, Shen H, Liu C, Podstolski A, Belanger F, Havkin-Frenkel D, Dixon RA.** 2017a. A re-evaluation of the final step of vanillin biosynthesis in the orchid *Vanilla planifolia*. *Phytochemistry* **139**, 33-46.
- Yang X, Guo Y, Zhu J, Niu Z, Shi G, Liu Z, Li K, Guo X.** 2017b. Genetic Diversity and Association Study of Aromatics in Grapevine. *Journal of the American Society for Horticultural Science* **142**, 225-231.
- Yang XX, Guo YS, Zhu JC, Ma N, Sun T, Liu ZD, Li K, Guo XW.** 2017c. Associations between the alpha-terpineol synthase gene and alpha-terpineol content in different grapevine varieties. *Biotechnology & Biotechnological Equipment* **31**, 1100-1105.
- Yang XX, Guo YS, Zhu JC, Shi GL, Niu ZZ, Liu ZD, Li K, Guo XW.** 2017d. Associations between the 1-deoxy-D-xylulose-5-phosphate synthase gene and aroma in different grapevine varieties. *Genes & Genomics* **39**, 1059-1067.
- Yauk YK, Ged C, Wang MY, Matich AJ, Tessarotto L, Cooney JM, Chervin C, Atkinson RG.** 2014. Manipulation of flavour and aroma compound sequestration and release using a glycosyltransferase with specificity for terpene alcohols. *Plant Journal* **80**, 317-330.
- Zhu BQ, Cai JA, Wang ZQ, Xu XQ, Duan CQ, Pan QH.** 2014. Identification of a Plastid-Localized Bifunctional Nerolidol/Linalool Synthase in Relation to Linalool Biosynthesis in Young Grape Berries. *International journal of molecular sciences* **15**, 21992-22010.

## SUPPLEMENTARY DATA

**Supplementary Table S1:** Intensities and heritabilities of the means for molecules identified in mid-ripening berries of Riesling, Gewurztraminer (n=3) and their progeny (RixGw, 117 genotypes). Significance of the difference between the two parents are indicated as follows:  
\*  $P \leq 0.05$ , \*\*  $P \leq 0.01$ , \*\*\*  $P \leq 0.001$ .

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Fatty alcohol	1-Nonanol (#14)	2014	Bound	3033 ***	0	0	2058	8421	0.76
Fatty alcohol	1-Nonanol (#14)	2015	Bound	3607 ***	0	0	4286	256142	0.99
Fatty alcohol	1-Octanol (#11)	2014	Bound	2900 **	440	0	2281	6253	0.54
Fatty alcohol	1-Octanol (#11)	2015	Bound	3531 ***	0	0	5911	351070	0.99
Fatty alcohol	1-Octanol (#11)	2014	Free	1765 ***	0	0	539	3077	0.97
Fatty alcohol	1-Octanol (#11)	2015	Free	475	0	0	62	2094	NA
Heptadienal	2,4-Heptadienal, (E,E)- (#9)	2014	Free	7638	8063	0	4848	13601	0.26
Heptadienal	2,4-Heptadienal, (E,E)- (#9)	2015	Free	3188	3421	0	5425	27573	0.91
Hemiterpenol	Isoprenol (#2)	2014	Bound	1099	3052	0	2451	6402	0.24
Hemiterpenol	Isoprenol (#2)	2015	Bound	3232	6579 ***	0	15314	910565	0.99
Hemiterpenol	Isoprenol (#2)	2014	Free	6522	7582	3982	10010	21015	0.56
Hemiterpenol	Isoprenol (#2)	2015	Free	5677	9422 **	3591	8574	16454	0.71
Hemiterpenol	Prenol (#3)	2014	Bound	0	0	0	753	11140	NA
Hemiterpenol	Prenol (#3)	2015	Bound	0	14609 ***	0	8435	515384	0.99
Hemiterpenol	Prenol (#3)	2014	Free	0	5825	0	4652	26466	0.46
Hemiterpenol	Prenol (#3)	2015	Free	0	16249 ***	0	2947	22244	0.87

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Norisoprenoid	1.α,β-Dihydro-β-ionone (#49)	2014	Bound	0	0	0	154	2197	NA
Norisoprenoid	1.α,β-Dihydro-β-ionone (#49)	2015	Bound	19951 *	9227	0	58578	3470582	0.99
Norisoprenoid	1.α,β-Dihydro-β-ionone (#49)	2015	Free	0	0	0	248	4505	NA
Norisoprenoid	3-Hydroxy-β-damascone (#43)	2014	Bound	9067 **	4992	4383	8371	16876	0.49
Norisoprenoid	3-Hydroxy-β-damascone (#43)	2015	Bound	10818	8261	0	24607	1468154	0.99
Norisoprenoid	3-Oxo-7,8-dihydro-α-ionol (#51)	2014	Bound	5643	4631	3855	8696	19926	0.68
Norisoprenoid	3-Oxo-7,8-dihydro-α-ionol (#51)	2015	Bound	36299	42753	0	176567	10447733	0.98
Norisoprenoid	3-Oxo-7,8-dihydro-α-ionol (#51)	2015	Free	49213	45405	12426	63251	330054	0.5
Norisoprenoid	Dihydro-3-oxo-β-ionol (#50)	2014	Bound	1957 ***	41	0	2036	4453	0.55
Norisoprenoid	Dihydro-3-oxo-β-ionol (#50)	2015	Bound	3298 ***	0	0	5440	323203	0.99
Sesquiterpene	n.i. sesquiterpene (#48)	2014	Bound	0	1947 ***	0	1533	21339	0.91
Sesquiterpene	n.i. sesquiterpene (#48)	2015	Bound	0	1714 ***	0	3165	187732	0.98
Sesquiterpene	n.i. sesquiterpene (#48)	2014	Free	0	0	0	13	1536	NA
Sesquiterpene	n.i. sesquiterpene (#48)	2015	Free	0	0	0	8	926	NA
Monoterpenol	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#28)	2014	Bound	25949	26251	0	61787	354516	0.74
Monoterpenol	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#28)	2015	Bound	62830 **	16057	0	118148	7029993	0.98
Monoterpenol	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#28)	2014	Free	75033	46729	0	50585	565537	0.64
Monoterpenol	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#28)	2015	Free	46159 ***	0	0	41413	613386	0.96
Monoterpenol	n.i. terpene (#41)	2014	Bound	6465	3919	1326	4834	13452	0.52
Monoterpenol	n.i. terpene (#41)	2015	Bound	5610	6374	0	13154	787515	0.99
Monoterpenol	n.i. terpene (#41)	2014	Free	3625 **	2027	0	2221	6283	0.69
Monoterpenol	n.i. terpene (#41)	2015	Free	2194	1461	0	2512	8022	0.72
Monoterpenol	(E) 8-Hydroxylinalool (#36)	2014	Bound	19612	46331	0	111950	1245553	0.66
Monoterpenol	(E) 8-Hydroxylinalool (#36)	2015	Bound	267412	482739 *	0	825960	49443857	0.98
Monoterpenol	(E) 8-Hydroxylinalool (#36)	2014	Free	9113	16055	0	23258	355882	0.71

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Monoterpenol	(E) 8-Hydroxylinalool (#36)	2015	Free	11477	51175	0	209356	1521334	0.86
Monoterpenol	(Z)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#16)	2014	Bound	3183	1829	0	4414	29096	0.68
Monoterpenol	(Z)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#16)	2015	Bound	0	0	0	5764	337184	NA
Monoterpenol	(Z) 8-Hydroxygeraniol (#47)	2014	Bound	1214	8975 ***	0	8264	40505	0.83
Monoterpenol	(Z) 8-Hydroxygeraniol (#47)	2015	Bound	4356	14808 ***	0	21638	1294598	0.98
Monoterpenol	(Z) 8-Hydroxygeraniol (#47)	2014	Free	0	2436	0	1040	8690	0.37
Monoterpenol	(Z) 8-Hydroxygeraniol (#47)	2015	Free	0	2099	0	978	11737	0.42
Monoterpenol	(Z) 8-Hydroxylinalool (#35)	2014	Bound	8507	11011	0	24052	107288	0.88
Monoterpenol	(Z) 8-Hydroxylinalool (#35)	2015	Bound	10832	10142	0	45667	2702994	0.99
Monoterpenol	(Z) 8-Hydroxylinalool (#35)	2014	Free	0	0	0	1783	21604	NA
Monoterpenol	(Z) 8-Hydroxylinalool (#35)	2015	Free	0	0	0	1112	24963	NA
Monoterpenol	1,7-Octadien-3-ol, 2,6-dimethyl- (#34)	2014	Bound	0	5132	0	14754	90609	0.68
Monoterpenol	1,7-Octadien-3-ol, 2,6-dimethyl- (#34)	2015	Bound	0	0	0	20139	1178119	NA
Monoterpenol	1,7-Octadien-3-ol, 2,6-dimethyl- (#34)	2014	Free	0	0	0	2163	26100	NA
Monoterpenol	1,7-Octadien-3-ol, 2,6-dimethyl- (#34)	2015	Free	0	0	0	688	22688	NA
Monoterpenol	1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	2014	Bound	0	0	0	1108	14453	NA
Monoterpenol	1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	2015	Bound	0	0	0	3031	177300	NA
Monoterpenol	1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	2014	Free	0	902	0	3493	54879	0.87
Monoterpenol	1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	2015	Free	751	0	0	4297	68868	0.91
Monoterpenol	1,7-Octanediol, 3,7-dimethyl- (#33)	2014	Bound	0	9872	0	7648	41559	0.43
Monoterpenol	1,7-Octanediol, 3,7-dimethyl- (#33)	2015	Bound	0	28335 ***	0	27226	1635244	0.98
Monoterpenol	1,7-Octanediol, 3,7-dimethyl- (#33)	2014	Free	0	11629 ***	0	2705	32074	0.71
Monoterpenol	1,7-Octanediol, 3,7-dimethyl- (#33)	2015	Free	0	7610	0	4525	47830	0.51
Monoterpenol	2-Hydroxy-1,8-cineole (#24)	2014	Bound	2421	862	0	4061	19546	0.66
Monoterpenol	2-Hydroxy-1,8-cineole (#24)	2015	Bound	4069	4979	0	13889	826053	0.99

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Monoterpenol	2-Hydroxy-1,8-cineole (#24)	2014	Free	0	0	0	89	4013	NA
Monoterpenol	2-Hydroxy-1,8-cineole (#24)	2015	Free	6768	4058	0	6318	17109	0.39
Monoterpenol	n.i. terpene (#40)	2014	Bound	7017 ***	0	0	19824	173377	0.96
Monoterpenol	n.i. terpene (#40)	2015	Bound	10934 ***	0	0	54420	3199949	0.99
Monoterpenol	n.i. terpene (#40)	2014	Free	0	0	0	118	5489	NA
Monoterpenol	8-Hydroxycarvotanacetone (#37)	2014	Bound	0	0	0	2105	20299	NA
Monoterpenol	8-Hydroxycarvotanacetone (#37)	2015	Bound	0	0	0	1648	96387	NA
Monoterpenol	$\alpha$ -Terpineol (#15)	2014	Bound	10505	6125	0	17174	124265	0.81
Monoterpenol	$\alpha$ -Terpineol (#15)	2015	Bound	20289	15268	769	43968	358725	0.94
Monoterpenol	$\alpha$ -Terpineol (#15)	2014	Free	0	361	0	691	5489	0.64
Monoterpenol	$\alpha$ -Terpineol (#15)	2015	Free	3468	4083	1275	6141	26612	0.91
Monoterpenol	Citronellol (#19)	2014	Bound	1407	9114 ***	0	3832	14769	0.78
Monoterpenol	Citronellol (#19)	2015	Bound	257	11117 ***	0	9897	596035	0.99
Monoterpenol	Citronellol (#19)	2014	Free	907	8868 ***	0	3109	17143	0.54
Monoterpenol	Citronellol (#19)	2015	Free	0	4729 ***	0	1970	10317	0.88
Monoterpenol	n.i. terpene (#38)	2014	Bound	0	9501 ***	0	4199	56026	0.93
Monoterpenol	n.i. terpene (#38)	2015	Bound	0	18726 ***	0	13417	812976	0.99
Monoterpenol	n.i. terpene (#38)	2014	Free	0	5208	0	1046	14276	NA
Monoterpenol	n.i. terpene (#38)	2015	Free	0	3991 ***	0	1492	15801	0.73
Monoterpenol	Geranic acid (#39)	2014	Bound	798	1698	0	11375	218776	0.96
Monoterpenol	Geranic acid (#39)	2015	Bound	10071	22107 ***	0	89225	5267957	1
Monoterpenol	Geranic acid (#39)	2014	Free	0	82020 **	0	32400	170832	0.58
Monoterpenol	Geranic acid (#39)	2015	Free	7340	46209 ***	0	42437	307398	0.97
Monoterpenol	Geraniol (#23)	2014	Bound	40530	291087 ***	3870	210823	1355706	0.94
Monoterpenol	Geraniol (#23)	2015	Bound	39510	602719 ***	0	630667	37857344	1

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Monoterpenol	Geraniol (#23)	2014	Free	32244	218826 ***	3646	143374	716025	0.83
Monoterpenol	Geraniol (#23)	2015	Free	19582	184891 ***	0	103084	539953	0.84
Monoterpenol	Hotrienol (#12)	2014	Bound	0	1641	0	2227	19626	0.6
Monoterpenol	Hotrienol (#12)	2015	Bound	1779	0	0	4555	269151	0.91
Monoterpenol	Hotrienol (#12)	2014	Free	7179	8258	0	8578	43214	0.43
Monoterpenol	Hotrienol (#12)	2015	Free	4215	0	0	1887	31199	0.57
Monoterpenol	Isopregol (#42)	2014	Bound	0	2968 ***	0	677	7266	0.62
Monoterpenol	Isopregol (#42)	2015	Bound	0	5538 ***	0	2398	148578	0.98
Monoterpenol	Isopregol (#42)	2014	Free	0	1154	0	62	2538	NA
Monoterpenol	Isopregol (#42)	2015	Free	0	1129	0	120	5077	NA
Monoterpenol	Lilac alcohol C (#20)	2014	Bound	0	0	0	772	11455	NA
Monoterpenol	Lilac alcohol C (#20)	2015	Bound	0	0	0	1991	116452	NA
Monoterpenol	Lilac alcohol C (#20)	2014	Free	0	0	0	152	4716	NA
Monoterpenol	Lilac alcohol D (#22)	2014	Bound	0	0	0	527	6929	NA
Monoterpenol	Lilac alcohol D (#22)	2015	Bound	0	0	0	744	43515	NA
Monoterpenol	Lilac alcohol D (#22)	2014	Free	0	0	0	149	4030	NA
Monoterpenol	Lilac alcohol formate (#27)	2014	Bound	0	0	0	332	4329	NA
Monoterpenol	Lilac alcohol formate (#27)	2015	Bound	0	0	0	1362	79655	NA
Monoterpenol	Linalool (#10)	2014	Bound	6663	10837	0	30699	224191	0.8
Monoterpenol	Linalool (#10)	2015	Bound	12952 *	0	0	53194	3131269	0.99
Monoterpenol	Linalool (#10)	2014	Free	5464	8276	0	24821	241755	0.83
Monoterpenol	Linalool (#10)	2015	Free	6213 ***	0	0	12831	215855	1
Monoterpenol	Linalool oxide (pyranoid), cis- (#18)	2014	Bound	2784	1700	0	1714	8064	NA
Monoterpenol	Linalool oxide (pyranoid), cis- (#18)	2015	Bound	3605	3959	0	8012	480019	0.99
Monoterpenol	Linalool oxide (pyranoid), cis- (#18)	2014	Free	6449	6664	3785	6704	11703	0.28



**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Monoterpenol	Linalool oxide (pyranoid), cis- (#18)	2015	Free	3815	4532	0	4411	8340	NA
Monoterpenol	Linalool oxide (pyranoid), trans- (#17)	2014	Bound	6364	8269	0	14204	78365	0.8
Monoterpenol	Linalool oxide (pyranoid), trans- (#17)	2015	Bound	7310 **	5081	0	23946	1419404	1
Monoterpenol	Linalool oxide (pyranoid), trans- (#17)	2014	Free	12647	8246	0	23131	143300	0.84
Monoterpenol	Linalool oxide (pyranoid), trans- (#17)	2015	Free	13218 **	2565	0	11099	132004	0.89
Monoterpenol	Linalool oxide, (furanoid), cis- (#8)	2014	Bound	4047	3835	0	9452	56789	0.92
Monoterpenol	Linalool oxide, (furanoid), cis- (#8)	2015	Bound	3905 ***	2990	0	15608	923435	1
Monoterpenol	Linalool oxide, (furanoid), cis- (#8)	2014	Free	0	618	0	1818	19812	0.76
Monoterpenol	Linalool oxide, (furanoid), cis- (#8)	2015	Free	0	0	0	1009	15860	NA
Monoterpenol	Linalool oxide, (furanoid), trans- (#7)	2014	Bound	5925	7808	0	15240	83997	0.78
Monoterpenol	Linalool oxide, (furanoid), trans- (#7)	2015	Bound	6906 *	2715	0	20105	1190547	0.99
Monoterpenol	Linalool oxide, (furanoid), trans- (#7)	2014	Free	0	0	0	4208	95931	NA
Monoterpenol	Linalool oxide, (furanoid), trans- (#7)	2015	Free	0	0	0	1942	88187	NA
Monoterpenol	Nerol (#21)	2014	Bound	4195	72128 ***	0	29479	133553	0.77
Monoterpenol	Nerol (#21)	2015	Bound	4552	104909 ***	0	89615	5406684	0.99
Monoterpenol	Nerol (#21)	2014	Free	3767	33403 ***	0	15614	79977	0.82
Monoterpenol	Nerol (#21)	2015	Free	0	23117 ***	0	10730	131127	0.78
Lipoxygenase pathway products	1-Hexanol (#4)	2014	Bound	25099	16059	7681	27598	76076	0.44
Lipoxygenase pathway products	1-Hexanol (#4)	2015	Bound	35348	28877	0	97893	5823107	0.99
Lipoxygenase pathway products	1-Hexanol (#4)	2014	Free	57101	48686	10717	45586	102133	0.45
Lipoxygenase pathway products	1-Hexanol (#4)	2015	Free	44862 **	23002	8776	39481	108706	0.68
Lipoxygenase pathway products	2-Hexen-1-ol, (E)- (#6)	2014	Bound	5150	1915	0	8008	22215	0.45
Lipoxygenase pathway products	2-Hexen-1-ol, (E)- (#6)	2015	Bound	5752	2799	0	25505	1504872	0.96
Lipoxygenase pathway products	2-Hexen-1-ol, (E)- (#6)	2014	Free	52451	36664	10147	44847	137763	0.54
Lipoxygenase pathway products	2-Hexen-1-ol, (E)- (#6)	2015	Free	51855 ***	18527	8313	38572	134726	0.74

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Lipoxygenase pathway products	2-Hexenal, (E)- (#1)	2014	Bound	15029	25953 *	3444	16023	44953	0.5
Lipoxygenase pathway products	2-Hexenal, (E)- (#1)	2015	Bound	2906	3502	0	10418	619045	0.96
Lipoxygenase pathway products	2-Hexenal, (E)- (#1)	2014	Free	49577	134100 *	5148	100084	357690	0.7
Lipoxygenase pathway products	2-Hexenal, (E)- (#1)	2015	Free	24207	24128	2958	61536	335384	0.92
Lipoxygenase pathway products	3-Hexen-1-ol, (Z)- (#5)	2014	Bound	1026	0	0	1317	7833	0.35
Lipoxygenase pathway products	3-Hexen-1-ol, (Z)- (#5)	2015	Bound	3979	1132	0	9044	536715	0.97
Lipoxygenase pathway products	3-Hexen-1-ol, (Z)- (#5)	2014	Free	5176	5550	0	8029	33326	0.7
Lipoxygenase pathway products	3-Hexen-1-ol, (Z)- (#5)	2015	Free	9340	4484	0	9375	42287	0.51
Phenylpropanoid	4-Vinylguaiacol (#32)	2014	Bound	9960	7567	0	13848	79963	0.78
Phenylpropanoid	4-Vinylguaiacol (#32)	2015	Bound	11165	12321	0	89552	5274028	1
Phenylpropanoid	4-Vinylguaiacol (#32)	2014	Free	2440	6129	0	4137	37515	0.47
Phenylpropanoid	4-Vinylguaiacol (#32)	2015	Free	0	0	0	3728	26691	NA
Phenylpropanoid	Acetovanillone (#46)	2014	Bound	73334	115878	0	79200	461426	0.38
Phenylpropanoid	Acetovanillone (#46)	2015	Bound	24176	18183	0	247908	14566171	1
Phenylpropanoid	Acetovanillone (#46)	2014	Free	2356	0	0	3725	41824	0.77
Phenylpropanoid	Acetovanillone (#46)	2015	Free	0	0	0	3962	33528	NA
Phenylpropanoid	Benzeneacetaldehyde (#13)	2014	Free	8840	0	0	3896	56456	0.24
Phenylpropanoid	Benzeneacetaldehyde (#13)	2015	Free	0	0	0	1666	22890	NA
Phenylpropanoid	Benzyl alcohol (#25)	2014	Bound	40036	32083	26729	77738	326579	0.86
Phenylpropanoid	Benzyl alcohol (#25)	2015	Bound	52069 **	43163	0	172905	10257773	1
Phenylpropanoid	Benzyl alcohol (#25)	2014	Free	61752	31678	21489	110529	400206	0.73
Phenylpropanoid	Benzyl alcohol (#25)	2015	Free	44035	42465	12963	70583	308896	0.8
Phenylpropanoid	Cinnamic acid, ethyl ester (#29)	2014	Bound	0	0	0	371	5702	NA
Phenylpropanoid	Cinnamic acid, ethyl ester (#29)	2015	Bound	0	0	0	1095	64033	NA
Phenylpropanoid	Eugenol (#31)	2014	Bound	0	0	0	829	6822	NA

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Family	Molécule	Year	Free / Bound	Parents		RixGw progeny			
				Ri	Gw	Min.	Mean	Max	Heritability
Phenylpropanoid	Eugenol (#31)	2015	Bound	0	2109 ***	0	4476	265013	0.99
Phenylpropanoid	Eugenol (#31)	2014	Free	1134	150	0	265	2718	NA
Phenylpropanoid	Eugenol (#31)	2015	Free	0	0	0	308	6679	NA
Phenylpropanoid	Homovanillin alcohol (#53)	2014	Bound	16106 ***	0	0	12756	77423	0.8
Phenylpropanoid	Homovanillin alcohol (#53)	2015	Bound	11656 *	2335	0	25499	1512680	0.98
Phenylpropanoid	Homovanillin alcohol (#53)	2014	Free	7196	5856	0	8658	39072	0.88
Phenylpropanoid	Homovanillin alcohol (#53)	2015	Free	4476	8302	0	7356	51557	0.78
Phenylpropanoid	Methyl vanillate (#45)	2014	Bound	71090 *	44724	27332	86226	225293	0.71
Phenylpropanoid	Methyl vanillate (#45)	2015	Bound	51708	33541	0	128822	7663984	0.98
Phenylpropanoid	Methyl vanillate (#45)	2014	Free	7423	4890	0	10749	50017	0.78
Phenylpropanoid	Methyl vanillate (#45)	2015	Free	0	0	0	2774	20188	NA
Phenylpropanoid	Phenylethyl Alcohol (#26)	2014	Bound	55737	56261	37389	116712	563221	0.91
Phenylpropanoid	Phenylethyl Alcohol (#26)	2015	Bound	68144	72386	0	307383	18192702	1
Phenylpropanoid	Phenylethyl Alcohol (#26)	2014	Free	77847	54694	29338	139715	473914	0.74
Phenylpropanoid	Phenylethyl Alcohol (#26)	2015	Free	53736	50268	18522	91958	365345	0.75
Phenylpropanoid	Vanillin (#44)	2014	Free	25527 ***	7064	2683	19035	48590	0.74
Phenylpropanoid	Vanillin (#44)	2015	Free	26285 **	6338	4930	19793	72361	0.6
Phenylpropanoid	Zingerone (#52)	2014	Bound	14120	7856	0	17689	66768	0.6
Phenylpropanoid	Zingerone (#52)	2015	Bound	24234	22339	0	72455	4308506	0.99
Phenylpropanoid	Zingerone (#52)	2014	Free	1474	1275	0	3540	43055	0.62
Phenylpropanoid	Zingerone (#52)	2015	Free	825	641	0	3843	27715	0.78

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

**Supplementary Table S2:** QTLs for berry PVCs significant at  $P \leq 0.05$  in 2014 and 2015

Chr	Family	Molécule (peak #)	Free / Bound	Genetic Map	LOD range	%Var. range	Confidence interval (cM)
Chr1	Fatty alcohol	1-Nonanol (#14)	Bound	Ri	6.2-9.8	21.7-32.1	27.9-40.5
				SSR	7.5-10.9	13.2-25.2	40-46
		1-Octanol (#11)	Bound	Ri	4.5-5.4	16.5-19.1	25.3-53.4
				SSR	4.9-6.6	8.1-15.6	26-55.1
Chr3	Phenylpropanoid	Homovanillin alcohol (#53)	Free	Ri	3-3.7	7.8-13.5	9.9-27.3
				SSR	4.5-5	14.2-18.1	0-16
Chr4	Terpenol	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#28)	Bound	Ri	2.5-2.8	5.5-5.8	18.1-74.3
Chr5	Hemiterpenol	Prenol (#3)	Free	Gw	4.5-9.4	16.3-31.2	1.8-15.4
				SSR	7.2-9.9	24.6-32.2	0-22
			Bound	Gw	3.1-8.5	11.7-22.7	1.8-15.4
	Terpenol	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#28)	Bound	Gw	2.8-4.5	10.5-16.4	9.9-21.7
				Bound	Gw	7.4-8.2	19.6-24.4
	n.i. terpene (#41)		SSR		8.2-8.2	27.3-27.7	0-8

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Chr	Family	Molecule (peak #)	Free / Bound	Genetic Map	LOD range	%Var. range	Confidence interval (cM)
		(Z) 8-Hydroxylinalool (#35)	Bound	Gw	4.7-6.1	13-18.8	9.9-19
				SSR	5.2-7.5	18.9-21.2	4-10
		(Z) 8-Hydroxygeraniol (#47)	Free	Gw	6.8-6.9	23.7-23.9	1.8-10.8
				SSR	7.3-7.5	21.8-25	0-8
			Bound	Gw	10.7-15.5	34.1-46	7.3-10.8
				SSR	11.6-17.2	32.8-45.3	6-8
		1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	Free	Gw	3.3-4.1	12.3-12.8	9.9-21.7
			Bound	Gw	3.4-4.4	12.7-13.8	9.9-19
				SSR	3.9-4.2	11.3-18.8	2-12
		1,7-Octanediol, 3,7-dimethyl- (#33)	Free	Gw	4-7.9	13.7-26.7	9.9-19
				SSR	4.5-8.1	16.3-27.3	4-16.3
			Bound	Gw	11.6-22	36.6-58.2	10.8-15.4
				SSR	12.1-25	38.1-62.9	4-10
		Citronellol (#19)	Free	Gw	12.1-15.3	39.8-45.3	9.9-15.4
SSR	12.8-16.6			39.5-48	4-10		

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Chr	Family	Molecule (peak #)	Free / Bound	Genetic Map	LOD range	%Var. range	Confidence interval (cM)
			Bound	Gw	19.2-20.3	52.9-55.3	9.9-15.4
				SSR	19.9-21	49.7-54.4	4-10
		n.i. terpene (#38)	Free	Gw	5.2-8.4	18.3-21.9	0-10.8
				SSR	5.2-9.2	18.4-24.5	0-6.5
			Bound	Gw	8.9-18.2	29.7-45.2	9.9-15.4
				SSR	9.3-21.2	30.8-56.8	6-10
		Geranic acid (#39)	Free	Gw	8.6-14.5	28.9-43.4	9.9-15.4
				SSR	10.8-15.8	34.6-46.2	4-8
		Geraniol (#23)	Free	Gw	14.8-23.4	43.7-60.2	9.9-10.8
				SSR	15.2-23.7	45-60.6	4-8
			Bound	Gw	15.1-24.2	44-61.7	9.9-10.8
				SSR	15.7-24.4	46-62	8-10
		Hotrienol (#12)	Bound	Gw	2.9-4	10.7-14.6	0-21.7
		Isopregol (#42)	Bound	Gw	3.2-7.2	12-20.9	9.9-15.4
SSR	6-9.3			21.1-30.9	4-10		

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Chr	Family	Molécule (peak #)	Free / Bound	Genetic Map	LOD range	%Var. range	Confidence interval (cM)
		Lilac alcohol formate (#27)	Bound	Gw	2.7-2.8	8.4-10.1	1.8-33.3
				SSR	4.2-5.3	13-15	0-12
		Linalool (#10)	Free	Gw	3.8-4.7	12.1-17	9.9-19
				SSR	4-4.9	12.6-16	2-12
			Bound	Gw	4.2-4.3	13.4-15.5	9.9-21.7
				SSR	4.7-5	19.3-20.3	2-10
		Linalool oxide (pyranoid), trans- (#17)	Free	Gw	2.7-2.8	10-10.4	0-21.7
			Bound	Gw	3.3-5.3	déc-19	9.9-21.7
		Linalool oxide, (furanoid), cis- (#8)	Free	Gw	2.5-4.3	9.5-15.4	0-24.4
			Bound	Gw	8.7-8.8	28.8-29.1	1.8-15.4
				SSR	9.8-10.3	33.1-41.5	0-18
		Nerol (#21)	Free	Gw	10.4-11.2	32.6-35.6	7.3-15.4
				SSR	12.1-12.2	37.8-38	4-10
			Bound	Gw	13.8-18.8	41.7-52.5	9.9-15.4
SSR	14.9-18.6			44.1-52.2	4-10		

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

<b>Chr</b>	<b>Family</b>	<b>Molécule (peak #)</b>	<b>Free / Bound</b>	<b>Genetic Map</b>	<b>LOD range</b>	<b>%Var. range</b>	<b>Confidence interval (cM)</b>
<b>Chr6</b>	Terpenol	(Z) 8-Hydroxygeraniol (#47)	Bound	SSR	6.1-6.7	16-18.2	12-18
	Phenylpropanoid	Benzyl alcohol (#25)	Bound	Gw	4.2-6	11.6-21	21.6-54.4
				SSR	5.4-7.5	16.7-22.1	24-38
			Bound	Gw	2.7-3.3	10.1-12.1	10.8-33.4
				SSR	6.4-6.4	22.2-22.5	14-30
<b>Chr7</b>	Terpenol	1,7-Octanediol, 3,7-dimethyl- (#33)	Free	Gw	3-3.4	10.7-11.7	55.2-81.1
	Lipoxygenase pathway products	2-Hexen-1-ol, (E)- (#6)	Free	Gw	3-6.1	11.2-21.3	55.2-74.5
			Bound	Gw	4.4-5.4	16.5-19.3	55.2-74.5
	Phenylpropanoid	Acetovanillone (#46)	Free	Gw	2.9-3.6	10.7-13.1	74.5-90.9
<b>Chr8</b>	Phenylpropanoid	Phenylethyl Alcohol (#26)	Free	Gw	3.4-3.4	12.4-12.5	41-60.1
<b>Chr10</b>	Terpenol	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#28)	Free	Ri	2.6-2.9	9.6-10.6	0-12.5
			Bound	Ri	4.3-4.9	11.8-12.3	0-10.7
				SSR	5.6-6.2	19.3-19.7	0-12
		(Z) 8-Hydroxylinalool (#35)	Bound	Ri	3.7-4.8	13.7-17.2	0-10.7
				SSR	5.2-5.6	11.5-12.9	0-10



**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Chr	Family	Molecule (peak #)	Free / Bound	Genetic Map	LOD range	%Var. range	Confidence interval (cM)
		(Z)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#16)	Bound	Ri	2.8-5.9	10.6-14	1.8-10.7
		1,7-Octadien-3-ol, 2,6-dimethyl- (#34)	Bound	Ri	4.4-6.2	15.9-21.5	0.9-8
				SSR	4.9-7.2	12.9-17.8	0-14
		1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	Free	Ri	3.3-3.5	12.2-12.7	0-10.7
				SSR	4.3-5.7	8.9-12.6	0-6
			Bound	Ri	2.5-3.1	9.5-11.3	0-12.5
		Hotrienol (#12)	Bound	Ri	2.6-2.8	6.5-10.6	0-12.5
		Linalool (#10)	Free	Ri	3-4.7	11.1-17	0-10.7
				SSR	4.3-6.4	13.5-20.6	0-9.1
			Bound	Ri	4.6-6.6	16.7-19.6	0-8.9
				SSR	8.4-8.6	16.7-16.8	0-6
		Linalool oxide (pyranoid), trans- (#17)	Free	Ri	4.4-5	13.2-17.8	2.7-10.7
				SSR	5.2-5.3	18.4-18.8	2-14
			Bound	Ri	3.6-4.1	13.2-15	0-8.9
				SSR	5.5-5.8	17.2-20.3	0-10

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

<b>Chr</b>	<b>Family</b>	<b>Molecule (peak #)</b>	<b>Free / Bound</b>	<b>Genetic Map</b>	<b>LOD range</b>	<b>%Var. range</b>	<b>Confidence interval (cM)</b>
		Linalool oxide, (furanoid), trans- (#7)	Bound	Ri	4.6-4.6	12.6-16.5	0-10.7
				SSR	5.5-5.9	19.6-20.8	0-10
<b>Chr11</b>	Sesquiterpene	n.i. sesquiterpene (#48)	Bound	Ri	2.8-5.1	10.7-18.1	30.6-54.6
	Terpenol	1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	Free	SSR	4-4.5	7.7-8.6	2-92
		Linalool (#10)	Bound	SSR	4.2-6.1	9.1-14.6	8.7-22
<b>Chr12</b>	Phenylpropanoid	Zingerone (#52)	Bound	Ri	3.5-7.6	17.5-23.2	25.2-40.2
				SSR	4.7-8.4	10.2-28.3	18-26.4
<b>Chr13</b>	Terpenol	2-Hydroxy-1,8-cineole (#24)	Bound	SSR	4.3-9.8	15.7-31	18-40
		n.i. terpene (#40)	Bound	Ri	3.6-4.6	8.7-12.8	43.9-45.1
			Bound	SSR	6.2-6.9	14.3-18.1	14-36
	$\alpha$ -Terpineol (#15)	Bound	SSR	8.1-10.8	13.4-34.7	20-26	
	Lipoxygenase pathway products	3-Hexen-1-ol, (Z)- (#5)	Free	Gw	2.7-3	10.2-10.7	11.6-49.6
	Phenylpropanoid	4-Vinylguaiacol (#32)	Bound	Ri	2.7-3.5	10-12.8	41.2-65.7
<b>Chr15</b>	Sesquiterpene	n.i. sesquiterpene (#48)	Bound	Gw	2.4-4.9	7.7-14.8	16-46.7
	Terpenol	(Z) 8-Hydroxylinalool (#35)	Bound	Gw	3.5-5.6	9.9-15.9	27.8-47.6

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

<b>Chr</b>	<b>Family</b>	<b>Molecule (peak #)</b>	<b>Free / Bound</b>	<b>Genetic Map</b>	<b>LOD range</b>	<b>%Var. range</b>	<b>Confidence interval (cM)</b>
				SSR	4.4-5.4	7.8-7.9	26-44
		n.i. terpene (#40)	Bound	Ri	3.8-4.2	10.8-13.3	21.5-38.8
				SSR	4.2-5.4	9.7-11.4	19.1-40
<b>Chr18</b>	Fatty alcohol	1-Nonanol (#14)	Bound	Gw	4.2-4.5	16-16.2	0-14.3
				SSR	7.8-7.9	16.6-16.9	0-28.2
		1-Octanol (#11)	Bound	SSR	5.9-7.8	14.5-19.5	1.8-28.2

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

**Supplementary Table S3:** Selection of candidate genes potentially associated to PVCs QTLs.

Candidate genes were selected in genomic regions associated to PVCs QTLs. (coordinates in the PN40024 12X.v2 reference genome), based on their putative function. Gene ID in italics correspond to terpenol synthases (TPS) annotated by Martin *et al.* (2010).

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
Fatty alcohol	1	6298545	7518054	109	VIT_01s0137g00700	chr1:7230367.. 7232691	GDL76_ARATH GDSL esterase/lipase	2.00E-135	52.17
					VIT_01s0137g00710	chr1:7234024.. 7236188	GDL53_ARATH GDSL esterase/lipase	2.00E-176	62.67
					VIT_01s0137g00720	chr1:7237322.. 7240382	GDL53_ARATH GDSL esterase/lipase	0	63.16
					VIT_01s0137g00730	chr1:7240959.. 7245528	GDL53_ARATH GDSL esterase/lipase	0	66.14

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

<b>Molécule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i></b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
	18	2507904	4015808	138	VIT_18s0001g02760	chr18:2910928. .2914749	EST_HEVBR Esterase	2.00E-179	65.24
					VIT_18s0001g02770	chr18:2914955. .2920663	EST_HEVBR Esterase	2.00E-178	65.24
					VIT_18s0001g02780	chr18:2921245. .2922708	EST_HEVBR Esterase	1.00E-73	60.45
Hemiterpenol	5	1191543	5194837	396	VIT_05s0020g02130	chr5:3851143.. 3856263	DXS_CAPAN Probable 1-deoxy-D-xylulose-5-phosphate synthase, chloroplastic	0	86.77
Sesquiterpene	11	6 029 650	11 387 070	245	VIT_11s0206g00020	chr11:7348840. .7350264	IDI1_CAMAC Isopentenyl-diphosphate Delta-isomerase I	3.00E-64	67.55
	15	12545262	17665532	420	VIT_15s0046g00330	chr15:1734370 4..17345642	C70A2_ARATH Cytochrome P450 703A2	0	76.15
					VIT_15s0048g02900	chr15:1700538 4..17007131	C78A3_SOYBN Cytochrome P450 78A3	0	69.32
					VIT_15s0048g01970	chr15:1608893 1..16091721	THAH_ARATH Cytochrome P450 708A2	9.00E-31	56.32

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID  CRIBI V1 /  <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
					VIT_15s0048g01960	chr15:1608596 4..16088384	C87A3_ORYSJ Cytochrome P450 87A3	2.00E-118	54.52
					VIT_15s0048g01950	chr15:1607023 7..16070890	C87A3_ORYSJ Cytochrome P450 87A3	1.00E-36	45.39
					VIT_15s0048g01700	chr15:1590668 4..15909439	C04C1_PINTA Cytochrome P450 704C1	2.00E-170	49.1
					VIT_15s0048g01690	chr15:1589212 5..15894930	C04C1_PINTA Cytochrome P450 704C1	0	51.07
					VIT_15s0048g01680	chr15:1588245 6..15885257	C04C1_PINTA Cytochrome P450 704C1	4.00E-179	50.49
					VIT_15s0048g01670	chr15:1586880 4..15872017	C04C1_PINTA Cytochrome P450 704C1	3.00E-179	50.49
					VIT_15s0048g01610	chr15:1577535 9..15776766	C76C4_ARATH Cytochrome P450 76C4	3.00E-69	69.74
					VIT_15s0048g01600	chr15:1574159 3..15744038	CYT24_CATRO Cytochrome P450 76T24	0	55.09

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID  CRIBI V1 /  <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
					VIT_15s0048g01590	chr15:1573037 1..15740256	CYT24_CATRO Cytochrome P450 76T24	0	55.4
					VIT_15s0048g01570	chr15:1570353 0..15707230	C76C4_ARATH Cytochrome P450 76C4	2.00E-77	69.33
					VIT_15s0048g01550	chr15:1569712 4..15697786	C76C1_ARATH Cytochrome P450 76C1	4.00E-80	65.03
					VIT_15s0048g01540	chr15:1569658 9..15697021	CYT24_CATRO Cytochrome P450 76T24	6.00E-15	61.7
					VIT_15s0048g01530	chr15:1568863 3..15691321	CYT24_CATRO Cytochrome P450 76T24	0	54.05
					VIT_15s0048g01500	chr15:1563576 9..15636949	CYT24_CATRO Cytochrome P450 76T24	2.00E-100	52.5
					VIT_15s0048g01490	chr15:1560985 7..15612332	C76B6_CATRO Geraniol 8-hydroxylase	0.00E+00	55.82
					VIT_15s0048g01480	chr15:1559799 1..15598545	CYT24_CATRO Cytochrome P450 76T24	1.00E-81	62.84

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
					VIT_15s0048g01470	chr15:1559560 7..15596584	CYT24_CATRO Cytochrome P450 76T24	6.00E-91	46.71
					VIT_15s0048g01440	chr15:1557314 2..15574026	CYT24_CATRO Cytochrome P450 76T24	1.00E-61	47.83
					VIT_15s0048g01620	chr15:1578114 6..15781403	CYT24_CATRO Cytochrome P450 76T24	2.00E-16	67.31
Terpenol	4	3320724	23410170	1104	/	/	/	/	/
	5	3851143	4758001	87	VIT_05s0020g02130	chr5:3851143.. 3856263	DXS_CAPAN Probable 1-deoxy-D- xylulose-5-phosphate synthase, chloroplastic	0	86.77
	6	5054535	6362346	140	VIT_06s0004g04330	chr6:5308774.. 5311089	8HGO_CATRO 8-hydroxygeraniol oxidoreductase	2.00E-176	60.89
					VIT_06s0004g04320	chr6:5298004.. 5302347	8HGO_CATRO 8-hydroxygeraniol oxidoreductase	1.00E-157	56.61
					VIT_06s0004g04310	chr6:5282932.. 5294571	8HGO_CATRO 8-hydroxygeraniol oxidoreductase	4.00E-149	58.72



**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID  CRIBI V1 /  <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
	7	19002549	24445899		VIT_07s0129g00700	chr7:22261529. .22262324	CYQ32_CATRO Cytochrome P450 81Q32	1.00E-100	59.73
VIT_07s0129g00710					chr7:22264819. .22266546	CYQ32_CATRO Cytochrome P450 81Q32	0	60.28	
VIT_07s0129g00730					chr7:22273787. .22275794	CYQ32_CATRO Cytochrome P450 81Q32	0	59.8	
VIT_07s0129g00760					chr7:22291738. .22293436	CYQ32_CATRO Cytochrome P450 81Q32	0	60.04	
VIT_07s0129g00770					chr7:22309401. .22310597	CYQ32_CATRO Cytochrome P450 81Q32	3.00E-111	66.08	
VIT_07s0129g00800					chr7:22323990. .22325858	CYQ32_CATRO Cytochrome P450 81Q32	0	60.64	
VIT_07s0129g00810					chr7:22331762. .22333283	CYQ32_CATRO Cytochrome P450 81Q32	1.00E-152	60.47	
VIT_07s0129g00820					chr7:22335270. .22337246	CYQ32_CATRO Cytochrome P450 81Q32	0	57.2	

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID  CRIBI V1 /  <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
					VIT_07s0129g00830	chr7:22344121. .22349743	CYQ32_CATRO Cytochrome P450 81Q32	6.00E-112	66.09
					VIT_07s0129g00850	chr7:22364071. .22366337	CYQ32_CATRO Cytochrome P450 81Q32	5.00E-163	57.86
					VIT_07s0129g00860	chr7:22368125. .22372421	CYQ32_CATRO Cytochrome P450 81Q32	0	60.24
					VIT_07s0129g00750	chr7:22285114. .22291186	CYQ32_CATRO Cytochrome P450 81Q32	2.00E-70	59.55
					VIT_07s0031g01370	chr7:23782514. .23783995	C7A12_PANGI Cytochrome P450 CYP736A12	3.00E-136	52.82
					VIT_07s0031g01380	chr7:23788875. .23790710	C7A12_PANGI Cytochrome P450 CYP736A12	0	57.35
					VIT_07s0031g01390	chr7:23791882. .23793696	C7A12_PANGI Cytochrome P450 CYP736A12	0	57.35
					VIT_07s0031g01570	chr7:23993103. .23994887	C78A7_ARATH Cytochrome P450 78A7	0	69.48
	10	1473028	2922216	151	/	/	/	0	99.31

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
		Scaffold 271		4	<i>VviTPS55</i>	chrUkn:844363 8..8446031	<i>Linalool synthase</i>	9.00E-108	68.58
					<i>VviTPS144</i>	chrUkn:850588 8..8506873	<i>Linalool synthase</i>	/	/
					<i>VviTPS65</i>	chrUkn:851074 5..8513032	<i>Linalool synthase</i>	0	64.45
					<i>VviTPS56</i>	chrUkn:852611 1..8528398	<i>Linalool synthase</i>	0	58.97
	11	1 542 611	4 958 166	358	VIT_11s0016g04810	chr11:4101704. .4104591	C734A_ARATH Cytochrome P450 734A1	0	75,95
	13	7637509 20272268		479	<i>VviTPS127</i>	chr13:1191516 7..11915523	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	/	/
					<i>VviTPS122</i>	chr13:1976592 4..19768001	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	0	95.49
<i>VviTPS39</i>					chr13:1977135 0..19773899	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	0	93.39	

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID  CRIBI V1 /  <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
					<i>VviTPS121</i>	chr13:1978647 5..19793544	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	/	/
					<i>VviTPS120</i>	chr13:1979918 8..19806801	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	/	/
					<i>VviTPS119</i>	chr13:1982447 1..19826083	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	0	91.38
					<i>VviTPS118</i>	chr13:1984963 1..19851342	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	/	/
					<i>VviTPS117</i>	chr13:1988547 1..19887380	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	/	/
					<i>VviTPS37</i>	chr13:1989838 3..19900842	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	0	88.85
					<i>VviTPS116</i>	chr13:1992023 6..19922669	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	0	92
					<i>VviTPS106</i>	chr13:1996522 6..19967492	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	3E-115	88.6

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID  CRIBI V1 /  <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
					<i>VviTPS115</i>	chr13:2000496 5..20005438	ATESY_VITVI (-)- $\alpha$ -terpineol synthase	/	/
	15	14797253	17332248	246	VIT_15s0048g02900	chr15:1700538 4..17007131	C78A3_SOYBN Cytochrome P450 78A3	0	69.32
VIT_15s0048g01970					chr15:1608893 1..16091721	THAH_ARATH Cytochrome P450 708A2	9.00E-31	56.32	
VIT_15s0048g01960					chr15:1608596 4..16088384	C87A3_ORYSJ Cytochrome P450 87A3	2.00E-118	54.52	
VIT_15s0048g01950					chr15:1607023 7..16070890	C87A3_ORYSJ Cytochrome P450 87A3	1.00E-36	45.39	
VIT_15s0048g01700					chr15:1590668 4..15909439	C04C1_PINTA Cytochrome P450 704C1	2.00E-170	49.1	
VIT_15s0048g01690					chr15:1589212 5..15894930	C04C1_PINTA Cytochrome P450 704C1	0	51.07	
VIT_15s0048g01680					chr15:1588245 6..15885257	C04C1_PINTA Cytochrome P450 704C1	4.00E-179	50.49	

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

<b>Molécule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID  CRIBI V1 /  <i>Martin et al., 2010</i></b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
					VIT_15s0048g01670	chr15:1586880 4..15872017	C04C1_PINTA Cytochrome P450 704C1	3.00E-179	50.49
					VIT_15s0048g01610	chr15:1577535 9..15776766	C76C4_ARATH Cytochrome P450 76C4	3.00E-69	69.74
					VIT_15s0048g01600	chr15:1574159 3..15744038	CYT24_CATRO Cytochrome P450 76T24	0	55.09
					VIT_15s0048g01590	chr15:1573037 1..15740256	CYT24_CATRO Cytochrome P450 76T24	0	55.4
					VIT_15s0048g01570	chr15:1570353 0..15707230	C76C4_ARATH Cytochrome P450 76C4	2.00E-77	69.33
					VIT_15s0048g01550	chr15:1569712 4..15697786	C76C1_ARATH Cytochrome P450 76C1	4.00E-80	65.03
					VIT_15s0048g01540	chr15:1569658 9..15697021	CYT24_CATRO Cytochrome P450 76T24	6.00E-15	61.7
					VIT_15s0048g01530	chr15:1568863 3..15691321	CYT24_CATRO Cytochrome P450 76T24	0	54.05

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

<b>Molécule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i></b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
					VIT_15s0048g01500	chr15:1563576 9..15636949	CYT24_CATRO Cytochrome P450 76T24	2.00E-100	52.5
					VIT_15s0048g01490	chr15:1560985 7..15612332	C76B6_CATRO Geraniol 8-hydroxylase	0	55.82
					VIT_15s0048g01480	chr15:1559799 1..15598545	CYT24_CATRO Cytochrome P450 76T24	1.00E-81	62.84
					VIT_15s0048g01470	chr15:1559560 7..15596584	CYT24_CATRO Cytochrome P450 76T24	6.00E-91	46.71
					VIT_15s0048g01440	chr15:1557314 2..15574026	CYT24_CATRO Cytochrome P450 76T24	1.00E-61	47.83
					VIT_15s0048g01620	chr15:1578114 6..15781403	CYT24_CATRO Cytochrome P450 76T24	2.00E-16	67.31
Lipoxygenase pathway products	7	19002549	21473110	242	VIT_07s0191g00050	chr7:20619161. .20620800	PER17_ARATH Peroxidase 17	0	76.16
	13	3906495	13827815	477	VIT_13s0106g00320	chr13:1000195 0..10003856	GDL71_ARATH GDSL esterase/lipase	4.00E-88	46.69

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
					VIT_13s0106g00350	chr13:1007515 5..10077313	GDL71_ARATH GDSL esterase/lipase	2.00E-106	50.96
					VIT_13s0106g00440	chr13:1017439 7..10178420	GDL71_ARATH GDSL esterase/lipase	3.00E-117	48.7
Phenylpropanoid	3	2415247	3773168	172	VIT_03s0038g04530	chr3:3324275.. 3325170	BAMT_ANTMA Benzoate carboxyl methyltransferase	7.00E-56	46.05
	6	5890683	12579867	393	VIT_06s0004g08150	chr6:8882451.. 8885910	TCMO_ZINVI Trans-cinnamate 4- monooxygenase	0	87.05
					VIT_06s0004g07310	chr6:8036795.. 8038646	U87A1_ARATH UDP-glycosyltransferase 87A1	5.00E-165	52.61
					VIT_06s0004g07280	chr6:8023357.. 8024290	U87A1_ARATH UDP-glycosyltransferase 87A1	1.00E-55	53.53
					VIT_06s0004g07270	chr6:8022216.. 8023356	U87A1_ARATH UDP-glycosyltransferase 87A1	2.00E-65	56.35



**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

<b>Molecule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID CRIBI V1 / Martin et al., 2010</b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
					VIT_06s0004g07250	chr6:7994693.. 7996798	U87A1_ARATH UDP-glycosyltransferase 87A1	6.00E-168	54.28
					VIT_06s0004g07240	chr6:7986280.. 7988465	U87A1_ARATH UDP-glycosyltransferase 87A1	2.00E-158	50.56
					VIT_06s0004g07230	chr6:7983729.. 7986100	U87A2_ARATH UDP-glycosyltransferase 87A2	2.00E-128	52.05
					VIT_06s0004g07220	chr6:7980168.. 7981988	U87A2_ARATH UDP-glycosyltransferase 87A2	4.00E-98	56.13
					VIT_06s0004g07300	chr6:8034332.. 8034776	U87A1_ARATH UDP-glycosyltransferase 87A1	6.00E-24	44.72
	7	24397840	25737398	85	VIT_07s0031g02630	chr7:25286923.. .25292785	AL3H1_ARATH Aldehyde dehydrogenase family 3 member H1	0	68.19
	8	18120696	22491293	482	VIT_08s0007g07530	chr8:21025975.. .21026898	ADRC3_ARATH NADPH-dependent aldehyde reductase-like protein, chloroplastic	6.00E-128	68.85

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

<b>Molécule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID CRIBI V1 / Martin et al., 2010</b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
					VIT_08s0007g07520	chr8:21012902. .21014341	ADRC3_ARATH NADPH-dependent aldehyde reductase-like protein, chloroplastic	6.00E-118	64.73
	12	15687643	19629932	205	VIT_12s0034g01120	chr12:1858256 1..18584204	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	2.00E-173	54.64
VIT_12s0034g00160					chr12:1718988 0..17191367	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	0	57	
VIT_12s0034g00150					chr12:1718039 1..17181645	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	2.00E-30	75	
VIT_12s0034g00140					chr12:1715103 6..17157548	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	0	56.85	
VIT_12s0034g00130					chr12:1710122 2..17102809	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	0	56.76	
VIT_12s0034g00080					chr12:1702874 8..17030350	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	2.00E-175	55.88	
VIT_12s0034g00060					chr12:1700816 7..17009798	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	8.00E-179	56.11	

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

<b>Molecule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i></b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
					VIT_12s0034g00040	chr12:1698368 0..16985327	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	1.00E-173	55.26
					VIT_12s0034g00030	chr12:1695822 7..16960109	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	5.00E-159	60.26
					VIT_12s0055g00020	chr12:1688802 5..16889618	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	2.00E-179	56.52
					VIT_12s0055g00030	chr12:1686837 1..16871175	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	0	57.08
					VIT_12s0055g00050	chr12:1684611 9..16847905	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	8.00E-177	56.25
					VIT_12s0055g00060	chr12:1683307 7..16833955	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	6.00E-46	49.74
					VIT_12s0055g00070	chr12:1682738 6..16828287	U7A15_MALDO UDP-glycosyltransferase 71A15	5.00E-119	58.72
					VIT_12s0055g00160	chr12:1674129 4..16743067	U7A15_MALDO UDP-glycosyltransferase 71A15	1.00E-159	54.94

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

<b>Molécule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i></b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
					VIT_12s0055g00170	chr12:1672871 5..16729497	UFOG6_MANES Anthocyanidin 3-O-glucosyltransferase 6 (Fragment)	1.00E-99	59.11
					VIT_12s0055g00180	chr12:1672497 5..16726517	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	4.00E-176	55.72
					VIT_12s0055g00200	chr12:1669325 3..16696124	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	7.00E-176	55.81
					VIT_12s0055g00250	chr12:1660146 5..16603103	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	3.00E-178	55.49
					VIT_12s0055g00280	chr12:1656088 2..16562261	U7A15_MALDO UDP-glycosyltransferase 71A15	2.00E-162	51.73
					VIT_12s0055g00290	chr12:1652116 6..16543422	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	3.00E-168	54.17
					VIT_12s0055g00310	chr12:1649439 8..16495945	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	8.00E-175	54.94
					VIT_12s0055g00320	chr12:1648444 8..16485962	UFOG6_FRAAN UDP-glucose flavonoid 3-O-glucosyltransferase 6	1.00E-174	54.85

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS**

Molecule family	Chr	Lower limit (bp)	Upper limit (bp)	Number of genes	Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i>	Position (bp)	Best hit on Uniprot/Swissprot	e-value	% identity
	13	4768465	18211439	522	VIT_13s0101g00260	chr13:1204715 0..12047308	CCR1_ARATH Cinnamoyl-CoA reductase 1	0.048	41.46
VIT_13s0019g05260					chr13:7173843. .7176208	4CLL5_ARATH 4-coumarate--CoA ligase- like 5	4.00E-150	44.79	
VIT_13s0047g00990					chr13:1790363 2..17918950	CCR1_ARATH Cinnamoyl-CoA reductase 1	4.00E-92	51.88	
VIT_13s0047g00940					chr13:1777219 8..17783457	CCR1_ARATH Cinnamoyl-CoA reductase 1	8.00E-97	52.02	
VIT_13s0047g00770					chr13:1746798 6..17481341	CCR1_ARATH Cinnamoyl-CoA reductase 1	2.00E-19	46.6	
VIT_13s0047g00760					chr13:1745817 3..17462104	CCR1_ARATH Cinnamoyl-CoA reductase 1	3.00E-18	30.27	
VIT_13s0047g00550					chr13:1702268 3..17025108	CCR1_ARATH Cinnamoyl-CoA reductase 1	3.00E-76	44.38	
VIT_13s0047g00540					chr13:1701118 8..17022465	CCR1_ARATH Cinnamoyl-CoA reductase 1	6.00E-94	50.78	

**Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw  
détectées par GC-MS**

<b>Molecule family</b>	<b>Chr</b>	<b>Lower limit (bp)</b>	<b>Upper limit (bp)</b>	<b>Number of genes</b>	<b>Candidate gene ID CRIBI V1 / <i>Martin et al., 2010</i></b>	<b>Position (bp)</b>	<b>Best hit on Uniprot/Swissprot</b>	<b>e-value</b>	<b>% identity</b>
					VIT_13s0074g00220	chr13:7768507. .7768905	COMT1_CATRO Caffeic acid 3-O- methyltransferase	1.00E-41	62
		5890759	12835683	281	VIT_13s0101g00260	chr13:1204715 0..12047308	CCR1_ARATH Cinnamoyl-CoA reductase 1	0.048	41.46
					VIT_13s0101g00070	chr13:1168124 9..11691265	BCRA_THAAR Benzoyl-CoA reductase subunit A	2.2	43.33

Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

**Supplementary Table S4:** Genotyping of individuals from a self-progeny of the *Vitis vinifera* PN40024 line with PCR primers designed on scaffold 271 in the reference genome.

Forward primer: 5'-CCTTGTGCTGTATCAGCAACACA-3'

Reverse primer : 5'-TTCATTCAATCTTTGTGCGTGTTGA-3'

Genotype	locus	Allele 1	Allele 2
3974	lis-sc271	382	420
4952	lis-sc271		420
4953	lis-sc271	382	420
4954	lis-sc271	382	420
4955	lis-sc271		420
4956	lis-sc271	382	420
4957	lis-sc271	382	420
4958	lis-sc271	382	420
4959	lis-sc271		420
4960	lis-sc271	382	420
4961	lis-sc271	382	420
4962	lis-sc271	382	420
4963	lis-sc271		420
4964	lis-sc271		420
4965	lis-sc271	382	420
4966	lis-sc271	382	420
4967	lis-sc271		420
4968	lis-sc271		420
4969	lis-sc271		420
4970	lis-sc271		420
4971	lis-sc271		420
4972	lis-sc271	382	420
4973	lis-sc271		420
4974	lis-sc271	382	420
4975	lis-sc271		420
4976	lis-sc271	382	420
4977	lis-sc271	382	420
4978	lis-sc271		420
4979	lis-sc271	382	420
4980	lis-sc271		420
4981	lis-sc271		420
4982	lis-sc271		420
4983	lis-sc271	382	420
4984	lis-sc271		420
4985	lis-sc271	382	420
4986	lis-sc271	382	420
4987	lis-sc271		420
4988	lis-sc271	382	420
4989	lis-sc271	382	420
4990	lis-sc271		420
4991	lis-sc271	382	420
4992	lis-sc271		420

### 3. Discussion

Une première analyse globale des données obtenues en GC-MS m'a permis d'observer la ségrégation parmi les descendants de profils d'accumulation de 53 molécules aromatiques d'intérêt dans la population d'étude dans les fractions libres et liées des extraits SPE de baies. Au total, 96 molécules libres et liées ont été quantifiées sur deux années de prélèvements de baies dans la descendance RixGw. Ces premiers résultats m'ont tout d'abord permis d'appréhender la diversité de molécules présents dans la population, d'évaluer leur distribution au sein de la population RixGw et de mettre en évidence les molécules présentes en quantités différentes entre le Ri et le Gw. Les molécules détectées appartiennent à différentes familles chimiques : des isoprénoïdes (hémiterpénols, monoterpénols, norisoprénoïdes, sesquiterpènes), des alcools à longue chaîne, des dérivés de la voie des lipoxygénases et des phénylpropanoïdes. La famille de molécule la plus représentée dans les analyses est la famille des monoterpénols avec 28 composés sur les 53 composés d'intérêt. Cette analyse a révélé l'apparition de molécules dans les baies de la population RixGw qui n'avaient été détectées ni dans les baies de Ri, ni dans les baies de Gw. Cette observation souligne le fait qu'à partir de deux cépages, des individus ayant un profil aromatique très différent des parents ont été créés grâce à la combinaison d'allèles du Ri et du Gw.

Une analyse QTL avec 3 cartes génétiques différentes a ensuite été réalisée à partir de ces données de quantification de molécules volatiles ou potentiellement volatiles présents dans les baies. Cette analyse a relevé 283 QTL impactant la biosynthèse de ces molécules dont 109 QTL reproductible sur les deux années de prélèvements pour 51 molécules différentes. Ces analyses m'ont permis de confirmer les QTL précédemment identifiés et liés au contenu en cinq terpénols majeurs (linalol,  $\alpha$ -terpinéol, citronellol, nérol et géraniol) situés sur les chromosomes 5, 10 et 13 (Duchêne et al., 2009). Cependant, en utilisant de nouvelles cartes génétiques plus denses (Illumina 18k), j'ai pu obtenir des intervalles de confiance des QTL plus petits, me permettant une localisation plus précise des gènes candidats associés.

Cette analyse génétique portant sur l'analyse des molécules volatiles détectés de manière non ciblée par GC-MS a permis de mettre en évidence l'impact majeur d'un petit nombre de régions du génome sur le profil aromatique global des raisins de la population RixGw. Le QTL localisé sur le chromosome 5 de la carte génétique SNP\_Gw, qui est lié au gène de la DXS, est un « hotspot » génétique majeur de notre étude. En effet, il impacte non seulement le contenu en cinq terpénols identifiés précédemment (Duchêne et al., 2009), mais également un grand



## Chapitre 1 Déterminisme génétique de la biosynthèse de molécules potentiellement volatiles dans les raisins de la population RixGw détectées par GC-MS

nombre d'autres dérivés hydroxylés ou oxydés des terpénols et hémiterpénols, participant à la richesse aromatique des individus de la population RixGw. Un autre hotspot localisé sur le chromosome 10 de la carte génétique SNP\_Ri, lié à gènes codant pour des linalol synthases, impacte plus spécifiquement la biosynthèse de composés dérivés du linalol. D'autres QTL ont été détectés pour les isoprénoïdes sur les chromosomes 6, 7, 8, 11, 13 et 15. Des QTL liés aux contenus en alcools à longue chaîne ont été trouvés sur les chromosomes 1 et 18. Des QTL liés aux molécules de la famille des phénylpropanoïdes ont été détectés sur les chromosomes 3, 6, 7, 8, 12 et 13. Des QTL ont été détectés pour les molécules dérivées des voies de la lipoxigénase sur les chromosomes 7 et 13. Aucun QTL reproductible n'a été détecté pour les molécules de la famille des norisoprénoïdes.

A partir des positions génétiques des intervalles de confiance des QTL, puis de leurs positions physiques, des gènes candidats ont été identifiés sur la base du génome de référence de la vigne, PN40024. Des cytochromes P450 ont notamment été identifiés comme pouvant être impliqués dans la biosynthèse de terpénols oxydés.

Certains des QTL n'ont été mis en évidence qu'une seule des deux années de prélèvements (Annexe 1). Un QTL sur le chromosome 2 lié au contenu en vanilline des baies a notamment été identifié et présenté de manière plus approfondie dans le chapitre 4.

L'étude du déterminisme génétique de la biosynthèse des molécules aromatique a été poursuivie dans le chapitre 2 avec l'étude des précurseurs glycosylés par UHPLC-MS.

**Chapitre 2 Caractérisation du métabolome des  
raisins par UHPLC-MS et analyses génétiques  
associées**



## **1. Introduction**

L'analyse par GC-MS des extraits de baies de raisins de la population RixGw a permis d'avoir une vision globale des molécules directement volatiles ou libérées lors de l'hydrolyse par l'activité glycosidase de l'enzyme Rapidase® AR2000. Cependant, l'intégralité des précurseurs d'arômes n'est pas détectée par GC-MS. En effet, l'hydrolyse enzymatique par l'AR2000 n'est pas parfaite, les précurseurs complexes (diglycosylés, par exemple) étant très mal hydrolysés (Hjelmeland and Ebeler, 2015). L'autre désavantage de cette approche est l'absence d'information sur le sucre lié à l'aglycone. Enfin, certains composés, bien que volatils, sont mal détectés dans nos conditions d'analyses GC-MS. Cette analyse a donc été complétée par une analyse non ciblée du métabolisme secondaire des baies par chromatographie liquide couplée à de la spectrométrie de masse (UHPLC-MS) sur les mêmes échantillons de baies prélevées en 2014 et 2015, afin d'avoir une vision plus exhaustive des molécules présentes dans les baies.

Une partie des résultats de cette analyse a déjà été exploitée lors d'une précédente thèse réalisée par Tina Ilc (Ilc, 2015). Ces travaux ont porté sur l'analyse ciblée de la voie de biosynthèse des 8-OH-linalol et carboxylinalol (précurseurs de la wine lactone dans les vins) détectés dans les baies prélevées en 2014 avec la carte génétique de type SSR (Ilc et al., 2017). Des QTL avaient été détectés sur les chromosomes 5, 10 et 2 pour ces deux composés.

Dans le cadre de ma thèse, j'ai complété ces travaux par une analyse globale du métabolome des baies de manière non ciblée sur les deux années de prélèvement 2014 et 2015.

Le logiciel XCMS (Smith et al., 2006) a été utilisé afin de quantifier des ions de manière non ciblée.

Schématiquement, XCMS réalise les opérations suivantes à partir des données brutes issues d'un ensemble d'analyses en UHPLC-MS:

- Détection des ions présents dans l'ensemble des analyses (ions présents dans au moins 5% des échantillons au-delà d'un seuil minimal fixé) et quantification de chaque ion par intégration de la surface du pic correspondant,
- Alignement de l'ensemble des chromatogrammes, pour compenser d'éventuels décalages des temps de rétention et obtenir un temps de rétention unique pour chaque ion dans l'ensemble des analyses,

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

- Attribution à chaque ion d'un identifiant unique  $M_{xxx}T_{yyy}$ , où xxx désigne la masse de l'ion en Da et yyy désigne le temps de rétention en secondes, c'est à dire le temps écoulé à partir du début de l'analyse,
- Production d'une matrice de résultats où chaque analyse UHPLC-MS correspond à une ligne et chaque ion détecté correspond à une colonne. La quantification de chaque ion est ainsi obtenue pour chaque analyse.

XCMS est utilisé couramment pour traiter plusieurs centaines d'analyses comportant plusieurs milliers d'ions, ce qui génère des matrices de données de taille importante.

Ces analyses m'ont permis de caractériser la grande diversité de molécules pouvant être détectées dans les baies par UHPLC-MS. J'ai identifié les molécules les plus abondamment détectées dans la population RixGw et les molécules présentes de manière différentielle entre le Ri et le Gw.

Une analyse génétique a été réalisée en utilisant les données de métabolomiques non ciblée avec 3 cartes génétiques, ceci me permettant d'identifier des hotspots génétiques gouvernant la biosynthèse de nombreux composés présents dans les baies. Dans le cadre de l'étude des arômes et des précurseurs d'arômes, je me suis concentrée sur l'étude de la diversité et du déterminisme génétique des composés de type monoterpénols et norisoprénoïdes dans les baies.

Ces travaux réalisés sur les baies ont aussi été complétés par une analyse du métabolome des feuilles de la population RixGw au cours l'année 2015. Celui-ci a été comparé au métabolome des baies prélevées la même année afin de déterminer s'il était possible de mettre en évidence les relations entre le métabolome des baies prélevées à mi-maturité et le métabolome de feuilles prélevées au début printemps.

Pour ces travaux, les prélèvements ont été réalisés par Nathalie Jaegli et Vincent Dumas, l'extraction des métabolites secondaires des échantillons de baies a été réalisée par Nathalie Jaegli et celles des feuilles par David Halter, le génotypage a été réalisé par Gisèle Butterlin. Les analyses par UHPLC-MS ont été réalisées par Raymonde Baltenweck et David Halter. A partir des données analytiques brutes et j'ai réalisé le traitement des données à l'aide de la suite logicielle XCMS, j'ai effectué les recherches de QTL et j'ai analysé les régions génomiques associées pour la recherche de gènes candidats. J'ai enfin réalisé l'identification d'une partie des molécules d'intérêt, avec l'aide de Raymonde Baltenweck.

## 2. Résultats

### 2.1. Analyse non ciblée du déterminisme génétique du métabolome des baies par UHPLC-MS

L'analyse non ciblée du déterminisme génétique du métabolome des baies par UHPLC-MS a été réalisée en combinant deux types de données : des données de phénotypage par UHPLC-MS non ciblé et des données de génotypage avec les cartes génétiques SSR, SNP\_Ri et SNP\_Gw.

#### 2.1.1. Analyse non ciblée du métabolome des baies de la population par UHPLC-MS

La diversité du métabolisme secondaire détecté dans les baies de la population RixGw et de Ri et de Gw a tout d'abord été appréhendée par une analyse non ciblée par UHPLC-MS ESI (+) et ESI (-) sur deux années de prélèvements de baies de la population RixGw. L'alignement des ensembles d'analyses obtenues par UHPLC-MS a permis de détecter des milliers d'ions dans chaque ensemble (Tableau 11). Dans chaque jeu de données analytiques traitées simultanément, les ions sont désignés par un identifiant unique MxxxTyyy attribué par XCMS, où xxx désigne la masse de l'ion en Da et yyy désigne le temps de rétention en secondes, c'est à dire le temps écoulé à partir du début de l'analyse.

UHPLC-MS	Année	Nombre d'ions détectés dans les extraits de baies
ESI (+)	2014	4654
	2015	8152
ESI (-)	2014	3380
	2015	4108

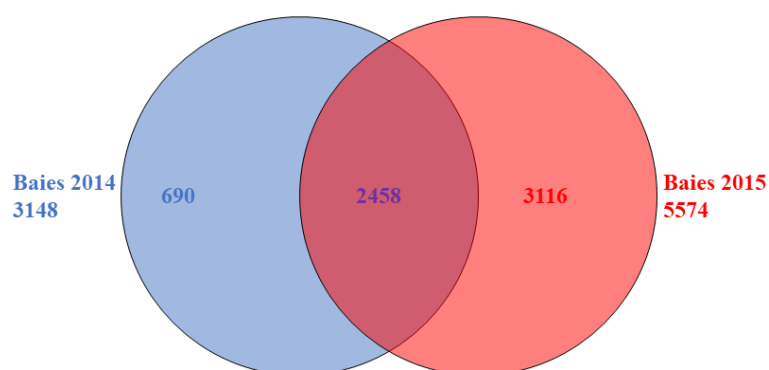
Tableau 11 : Nombre d'ions retenus par XCMS dans les échantillons de baies analysés par UHPLC-MS.

Le nombre de molécules différentes détectées par les analyses par UHPLC-MS est difficile à déterminer. En effet, il ne correspond pas au nombre d'ions puisque ce dernier dépend de la fragmentation et du nombre d'adduits. Le nombre de molécules détectées se rapproche cependant plus du nombre d'ions détecté en ESI (-) qu'en ESI (+).

Le nombre d'ions dans les analyses par UHPLC-MS ESI (+) en 2015 est bien supérieur à celui en 2014.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Afin d'identifier les ions communs et différentiels entre les deux années d'analyses, j'ai comparé les ions de même  $m/z$  sans prendre en compte leurs temps de rétention (Figure 26). Sur les 3148 ions détectés en 2014 et 5574 en 2015, 2458 ions ont été détectés à la fois sur les deux années d'analyses non ciblées avec les paramètres d'alignement utilisés.



**Figure 26 : Diagramme de Venn représentant le nombre d'ions détectés par UHPLC-MS ESI (+) dans les baies prélevées en 2014 et en 2015.**

L'observation des 100 ions majeurs (c'est-à-dire correspondant aux aires des pics moléculaires les plus élevées) dans les analyses par UHPLC-MS ESI (+) des deux années d'analyses du Ri et du Gw donne un indice sur la raison probable de la grande différence en nombre d'ions détectés entre les deux années (Figure 27). De nombreux ions présentent des aires de pics plus élevées en 2015 qu'en 2014 et un nombre plus important d'ions en 2015 passent le seuil d'aire de pic à 30 000, imposé lors de l'alignement non ciblé par XCMS pour la sélection des ions. Nous pouvons ainsi voir que déjà dans les analyses réalisées sur les baies des parents, des différences significatives sont déjà observables entre les 2 années de prélèvements. Par exemple, l'aire de pic maximale observée chez le Gw est de  $2.5 \times 10^8$  en 2014 au temps de rétention 3.9 minutes pour un  $m/z$  de 465.103 alors qu'il est de  $4 \times 10^8$  en 2015 pour le même.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

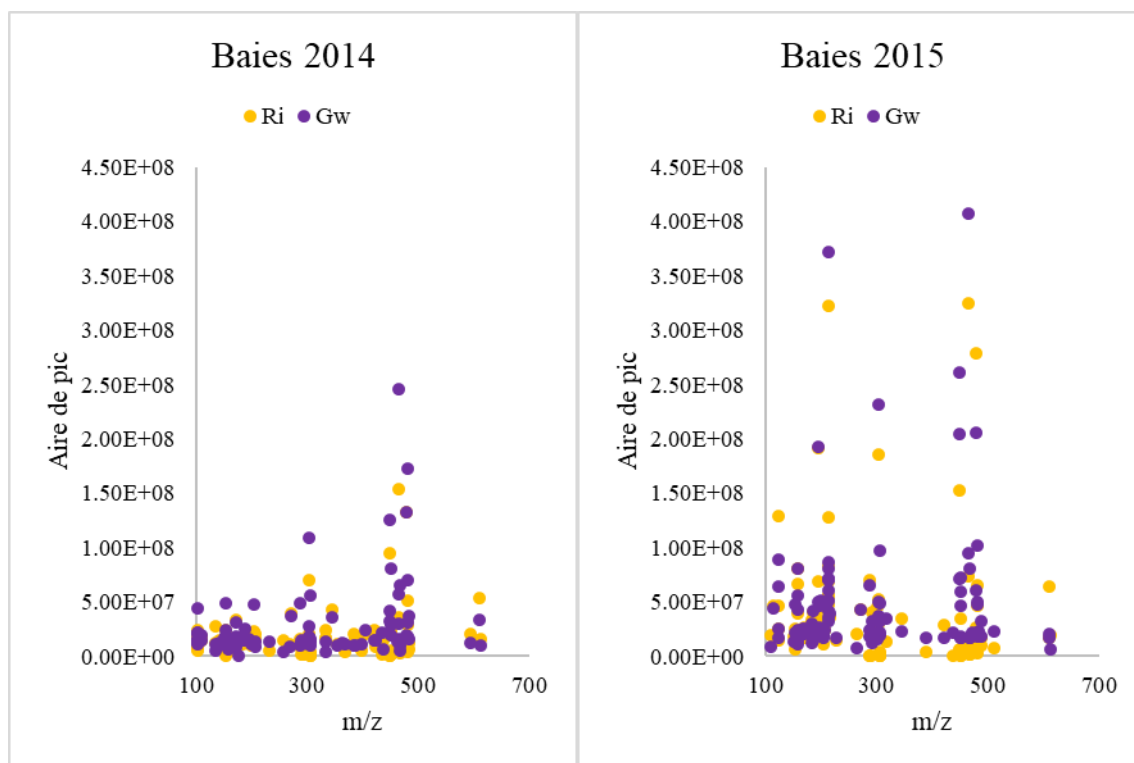


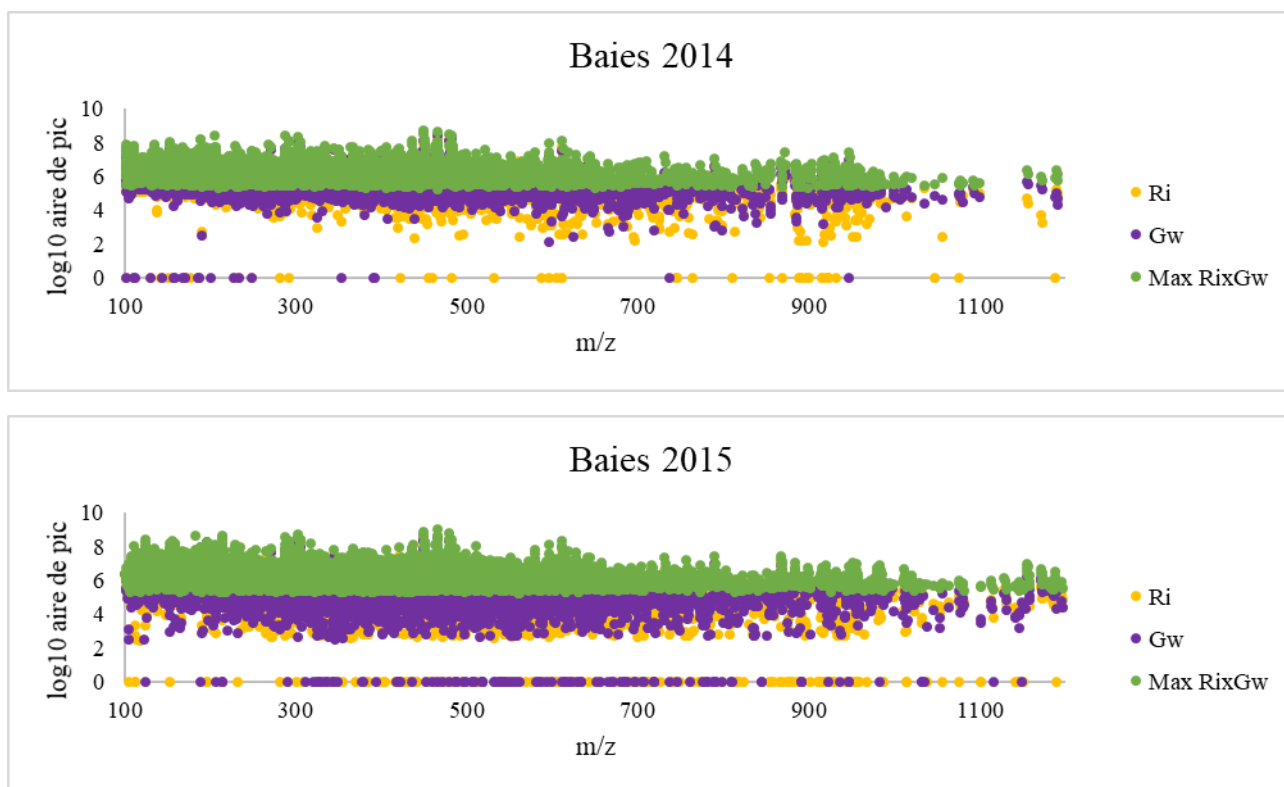
Figure 27 : Aires des 100 pics les plus intenses et m/z des ions associés dans les analyses de baies de Ri et de Gw par UHPLC-MS ESI (+).

Cette différence interannuelle est aussi observable dans les analyses de la population RixGw (Figure 28). Par exemple avec les analyses par UHPLC-MS ESI (+), une aire de pic maximale de  $5.4 \times 10^8$  est observée pour l'ion de m/z 449.108 au temps de rétention 3.9 min en 2014 alors qu'en 2015 une aire de pic maximale est observée pour l'ion de m/z 465.103 au temps de rétention 4 min avec une aire de pic maximale de  $1 \times 10^9$ .

On observe aussi une augmentation de la diversité de molécules présentes dans les descendants RixGw par rapport aux parents. En effet certains ions sont détectés dans la population mais sont absents dans les échantillons de Ri et/ou Gw (Figure 28).



## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées



**Figure 28 :** Représentation des ions détectés par UHPLC-MS ESI (+) en moyenne dans les baies de Ri, Gw et le maximum détecté dans la population RixGw en 2014 et en 2015.

Une étude plus approfondie des molécules les plus abondamment détectées dans les analyses de baies par UHPLC-MS ESI (+) a ensuite été réalisée. Pour cela, les 20 ions avec les aires de pics les plus élevées en 2014 et en 2015 ont été étudiés (Tableau 12 et Tableau 13).

Parmi les 20 ions majeurs détectés en 2014, 12 ions correspondent à des flavonoïdes, 6 ions correspondent à des terpénols, 1 ion à un acide aminé et 1 ion n'a pas pu être identifié. En 2015, 17 ions correspondant à des flavonoïdes ont été détectés, 2 à des terpénols et 1 ion n'a pas pu être identifié. Les molécules les plus détectées dans nos analyses sont des molécules de la famille des flavonoïdes puis les monoterpénols, notamment des glycosides de monoterpénols.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Ion	RT (min)	m/z	Ion pseudo-moléculaire	Standard / structure proposée	Famille	Moyenne Ri	Moyenne Gw	Min RixGw	Moyenne RixGw	Max RixGw	h <sup>2</sup>
M449T72	1.2	449.108	[C21H20O11 + H] <sup>+</sup>	<i>Glucoside de flavonoïde</i>	Flavonoïde	2.69E+05	4.18E+07 ***	2.80E+04	5.72E+07	5.36E+08	0.96
M465T234	3.9	465.103	[C21H20O12 + H] <sup>+</sup>	<b>Quercétine-glucoside</b>	Flavonoïde	1.54E+08	2.46E+08	4.26E+07	2.58E+08	4.60E+08	0.05
M449T282	4.7	449.108	[C21H20O11 + H] <sup>+</sup>	<b>Kaempférol 3-O-glucoside</b>	Flavonoïde	9.52E+07	1.26E+08	2.43E+07	1.60E+08	4.32E+08	NA
M479T263	4.4	479.082	[C21H18O13 + H] <sup>+</sup>	<b>Quercétine 3-O-glucuronide</b>	Flavonoïde	1.32E+08	1.32E+08	5.07E+07	1.81E+08	3.22E+08	NA
M449T100	1.7	449.108	[C21H20O11 + H] <sup>+</sup>	<i>Glucoside de flavonoïde</i>	Flavonoïde	2.72E+05	3.17E+07 ***	5.30E+04	3.62E+07	3.04E+08	0.95
M205T100	1.7	205.097	[C11H12O2N2 + H] <sup>+</sup>	<b>Tryptophane</b>	Acide aminé	2.32E+07	4.79E+07	4.57E+06	5.60E+07	2.76E+08	0.76
M482T243	4.1	482.26	[C21H36O11 + NH4] <sup>+</sup>	<i>Apiosyl glucoside de linalol oxyde</i>	Monoterpénol	5.13E+07	1.73E+08 ***	7.22E+06	8.93E+07	2.64E+08	0.89
M287T268	4.5	287.055	[C15H10O6 + H] <sup>+</sup>	<b>Fragment kaempférol</b>	Flavonoïde	1.52E+07	4.88E+07	3.67E+06	5.41E+07	2.46E+08	0.09
M303T234	3.9	303.050	[C15H10O7 + H] <sup>+</sup>	<b>Fragment quercétine</b>	Flavonoïde	6.95E+07	1.09E+08	1.82E+07	1.14E+08	2.17E+08	0.12
M188T100	1.7	188.071	[C11H9O2N + H] <sup>+</sup>	<b>Fragment tryptophane</b>	Acide aminé	1.30E+07	2.53E+07	2.59E+06	3.07E+07	1.54E+08	0.76
M451T259	4.3	451.123	[C21H22O11 + H] <sup>+</sup>	<b>Astilbine</b>	Flavonoïde	5.23E+06	8.02E+07 ***	1.76E+06	4.72E+07	1.53E+08	0.93
M611T202	3.4	611.161	[C27H30O16 + H] <sup>+</sup>	<b>Quercétine-rutinoside</b>	Flavonoïde	5.31E+07	3.31E+07	1.31E+07	6.23E+07	1.41E+08	0.19
M436T361	6	436.218	[C19H30O10 + H] <sup>+</sup>	Malonyl glucoside de terpenol oxydé	Monoterpénol	1.42E+07	5.87E+06 ***	2.78E+05	1.66E+07	1.31E+08	0.97
M291T132	2.2	291.086	[C15H14O6 + H] <sup>+</sup>	<b>Catéchine</b>	Flavonoïde	6.43E+06	9.20E+06	8.66E+04	3.37E+07	1.22E+08	0.89
M153T243	4.1	153.127	[C10H18O2 - H2O + H] <sup>+</sup>	<i>Fragment de l'apiosyl glucoside de linalol oxyde</i>	Monoterpénol	2.43E+07	4.83E+07 ***	4.42E+06	4.03E+07	1.17E+08	0.90
M482T197	3.3	482.26	[C21H36O11 + NH4] <sup>+</sup>	<i>Apiosyl glucoside de linalol oxyde</i>	Monoterpénol	2.91E+07	6.99E+07 **	4.56E+06	3.96E+07	1.15E+08	0.80
M466T404	6.7	466.265	[C21H36O10 + H] <sup>+</sup>	<i>Apiosyl-glucoside de monoterpenol</i>	Monoterpénol	1.02E+07	6.48E+07 ***	1.25E+06	2.25E+07	1.14E+08	0.79
M465T243	4.1	465.233	[C21H36O11 + H] <sup>+</sup>	<i>Apiosyl glucoside de linalol oxyde</i>	Monoterpénol	1.35E+07	2.99E+07 **	2.76E+06	2.75E+07	1.07E+08	0.90
M305T259	4.3	305.066	[C15H12O7 + H] <sup>+</sup>	<b>Fragment astilbine</b>	Flavonoïde	3.61E+06	5.63E+07 ***	1.06E+06	3.29E+07	1.04E+08	0.93
M291T154	2.6	291.086	[C15H14O6 + H] <sup>+</sup>	<b>Epicatechine</b>	Flavonoïde	2.05E+06	2.91E+06	3.02E+04	1.59E+07	1.02E+08	0.91

**Tableau 12 : Top 20 des ions les plus intenses détectés dans les baies par UHPLC-MS ESI (+) dans la population RixGw en 2014.** Les ions sont triés en fonction de l'aire de pic maximale dans la population RixGw. Des identifications putatives sont proposées sur la base de l'analyse détaillée des spectres de masse. Les identifications confirmées par des standards sont indiquées en gras.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Ion	RT (min)	m/z	Ion	Standard / structure proposée	Famille	Moyenne Ri	Moyenne Gw	Min RixGw	Moyenne RixGw	Max RixGw	h <sup>2</sup>
M465T239	4	465.103	[C21H20O12 + H] <sup>+</sup>	<b>Quercétine-glucoside</b>	Flavonoïde	3.25E+08	4.07E+08	0.00E+00	5.06E+08	1.01E+09	0.56
M449T74	1.2	449.108	[C21H20O11 + H] <sup>+</sup>	<i>Glucoside de flavonoïde</i>	Flavonoïde	6.01E+06	2.61E+08 ***	3.77E+03	1.14E+08	7.56E+08	0.91
M479T273	4.6	479.082	[C21H18O13 + H] <sup>+</sup>	<b>Quercétine 3-O-glucuronide</b>	Flavonoïde	2.79E+08	2.06E+08	0.00E+00	4.03E+08	6.90E+08	0.43
M303T238	4	303.05	[C15H10O7 + H] <sup>+</sup>	<b>Fragment quercétine-glucoside</b>	Flavonoïde	1.86E+08	2.31E+08	3.29E+03	2.86E+08	5.79E+08	0.58
M449T288	4.8	449.108	[C21H20O11 + H] <sup>+</sup>	<b>Kaempférol 3-O-glucoside</b>	Flavonoïde	1.52E+08	7.16E+07	6.46E+03	2.14E+08	5.11E+08	0.38
M449T104	1.7	449.108	[C21H20O11 + H] <sup>+</sup>	<i>Glucoside de flavonoïde</i>	Flavonoïde	4.80E+06	2.04E+08 ***	2.96E+03	6.95E+07	4.25E+08	0.91
M287T274	4.6	287.055	[C15H10O6 + H] <sup>+</sup>	<i>Luteolin-7 glucoside?</i>	Flavonoïde	6.96E+07	6.51E+07	4.42E+03	9.48E+07	4.06E+08	0.15
M303T270	4.5	303.05	[C15H10O7 + H] <sup>+</sup>	<b>Fragment quercétine 3-O-glucuronide</b>	Flavonoïde	5.28E+07	5.02E+07	1.40E+03	8.18E+07	3.30E+08	0.61
M482T253	4.2	482.26	[C21H36O11 + NH4] <sup>+</sup>	<i>Apiosyl glucoside de linalol oxyde</i>	Terpénol	4.68E+07	1.02E+08 *	0.00E+00	8.01E+07	2.81E+08	0.79
M305T265	4.4	305.066	[C15H12O7 + H] <sup>+</sup>	<b>Fragment astilbine</b>	Flavonoïde	3.50E+06	9.70E+07 ***	0.00E+00	8.07E+07	2.61E+08	0.97
M479T297	5	479.119	[C22H22O12 + H] <sup>+</sup>	<b>Isorhamnetin-3-O-glucoside</b>	Flavonoïde	2.60E+07	6.09E+07 *	0.00E+00	7.28E+07	2.57E+08	0.79
M611T206	3.4	611.161	[C27H30O16 + H] <sup>+</sup>	<b>Quercétine-rutinoside</b>	Flavonoïde	6.36E+07	2.09E+07	1.11E+03	8.91E+07	2.36E+08	0.62
M291T137	2.3	291.086	[C15H14O6 + H] <sup>+</sup>	<b>Catéchine</b>	Flavonoïde	4.11E+07	3.17E+07	5.90E+02	7.03E+07	2.28E+08	0.79
M451T265	4.4	451.124	[C21H22O11 + H] <sup>+</sup>	<b>Astilbine</b>	Flavonoïde	2.58E+06	7.22E+07 ***	0.00E+00	5.97E+07	1.86E+08	0.96
M153T253	4.2	153.127	[C10H18O2 - H2O + H] <sup>+</sup>	<i>Apiosyl glucoside de linalol oxyde</i>	Terpénol	2.48E+07	4.77E+07 **	1.08E+03	4.11E+07	1.59E+08	0.86
M317T297	5	317.066	[C16H12O7 + H] <sup>+</sup>	<b>Fragment isorhamnetin</b>	Flavonoïde	1.38E+07	3.51E+07 *	3.07E+03	4.15E+07	1.48E+08	0.79
M289T327	5.5	289.071	[C15H12O6 + H] <sup>+</sup>	Flavonoïde	Flavonoïde	7.80E+05	1.99E+07 *	0.00E+00	2.94E+07	1.33E+08	0.84
M463T316	5.3	463.087	[C21H18O12 + H] <sup>+</sup>	<b>Kaempférol glucuronide</b>	Flavonoïde	1.47E+07	1.17E+07	1.87E+03	3.45E+07	1.33E+08	0.68
M595T255	4.3	595.166	[C27H30O15 + H] <sup>+</sup>	<b>Kaempférol-rutinoside</b>	Flavonoïde	1.58E+07	5.17E+06	1.32E+03	3.24E+07	1.28E+08	0.75
M437T128	2.1	437.144	[C21H24O10 + H] <sup>+</sup>	<i>Glucoside de flavonoïde</i>	Flavonoïde	1.41E+05	2.16E+07 ***	3.78E+03	9.07E+06	1.22E+08	0.89

**Tableau 13 : Top 20 des ions les plus intenses détectés dans les baies par UHPLC-MS ESI (+) dans la population RixGw en 2015.** Les ions sont triés en fonction de l'aire de pic maximale dans la population RixGw. Des identifications putatives sont proposées sur la base de l'analyse détaillée des spectres de masse. Les identifications confirmées par des standards sont indiquées en gras.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Treize molécules communes entre les deux années d'analyses sont retrouvées dans le top 20 des deux années d'analyses (Tableau 14). Seules deux familles de molécules sont présentes parmi ces molécules communes: les flavonoïdes avec 11 ion correspondants, et les monoterpénols avec 2 ion correspondants. Les molécules communes avec les ID 1, 5, 6, 11 et 12 ont des aires de pics significativement plus élevées dans les baies du Gw que du Ri sur les deux années d'analyse.

Molécule commune	Ion 2014 – Ion 2015	RT (min)	m/z	Ion pseudo-moléculaire	Standart / Proposition structure	Famille	Anova Ri vs. Gw
1	M449T72-449T74	1.2	449.108	[C <sub>21</sub> H <sub>20</sub> O <sub>11</sub> + H] <sup>+</sup>	Glucoside de flavonoïde	Flavonoïde	Gw ***
2	M465T234-465T239	3.9	465.103	[C <sub>21</sub> H <sub>20</sub> O <sub>12</sub> + H] <sup>+</sup>	Quercétine-glucoside	Flavonoïde	/
3	M449T282-M449T288	4.7	449.108	[C <sub>21</sub> H <sub>20</sub> O <sub>11</sub> + H] <sup>+</sup>	Kaempférol 3-O-glucoside	Flavonoïde	/
4	M479T263-M479T273	4.4	479.082	[C <sub>21</sub> H <sub>18</sub> O <sub>13</sub> + H] <sup>+</sup>	Quercétine 3-O-glucuronide	Flavonoïde	/
5	M449T100-M449T104	1.7	449.108	[C <sub>21</sub> H <sub>20</sub> O <sub>11</sub> + H] <sup>+</sup>	Glucoside de flavonoïde	Flavonoïde	Gw ***
6	M482T243-M482T253	4.1	482.26	[C <sub>21</sub> H <sub>36</sub> O <sub>11</sub> + NH <sub>4</sub> ] <sup>+</sup>	Apiosyl glucoside de linalol oxyde	Terpénol	Gw *
7	M287T268-M287T274	4.5	287.055	[C <sub>15</sub> H <sub>10</sub> O <sub>6</sub> + H] <sup>+</sup>	Fragment kaempférol	Flavonoïde	/
8	M303T234-M303T238	3.9	303.050	[C <sub>15</sub> H <sub>10</sub> O <sub>7</sub> + H] <sup>+</sup>	Fragment Quercétine	Flavonoïde	/
9	M451T259-M451T265	4.3	451.123	[C <sub>21</sub> H <sub>22</sub> O <sub>11</sub> + H] <sup>+</sup>	Astilbine	Flavonoïde	Gw ***
10	M611T202-M611T206	3.4	611.161	[C <sub>27</sub> H <sub>30</sub> O <sub>16</sub> + H] <sup>+</sup>	Quercétine-rutinoside	Flavonoïde	/
11	M291T132-M291T137	2.2	291.086	[C <sub>15</sub> H <sub>14</sub> O <sub>6</sub> + H] <sup>+</sup>	Catéchine	Flavonoïde	/
12	M153T243-M153T253	4.1	153.127	[C <sub>10</sub> H <sub>18</sub> O <sub>2</sub> - H <sub>2</sub> O + H] <sup>+</sup>	Fragment d'apiosyl glucoside de linalol oxyde	Terpénol	Gw **
13	M305T259-M305T265	4.3	305.066	[C <sub>15</sub> H <sub>12</sub> O <sub>7</sub> + H] <sup>+</sup>	Fragment d'astilbine	Flavonoïde	Gw ***

Tableau 14 : Liste des ions retrouvés à la fois le top 20 des molécules détectées en 2014 et en 2015 par UHPLC-MS ESI (+). Les p-value des tests anova montrant une différence significative entre Ri et Gw sont signalées par des astérisques : \*\*\* ≤ 0.001, \*\* ≤ 0.01, \* ≤ 0.05.

### 2.1.2. Analyse génétique non ciblée des baies de la population RixGw

A partir des données non ciblées par UHPLC-MS des baies et de 3 cartes génétiques, une analyse QTL a été réalisée (Tableau 15). Des milliers de QTL ont été localisés sur l'ensemble des 19 chromosomes. Le nombre de QTL détecté en 2015 est supérieur à celui observé en 2014, surtout en ESI (+). Ceci peut être expliqué par le nombre plus importants d'ions détectés en 2015 qu'en 2014 avec des aires de pics plus élevées. Les LOD scores sont plus élevés dans les analyses ESI (-). Ceci peut être expliqué par le fait du moindre effet de

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

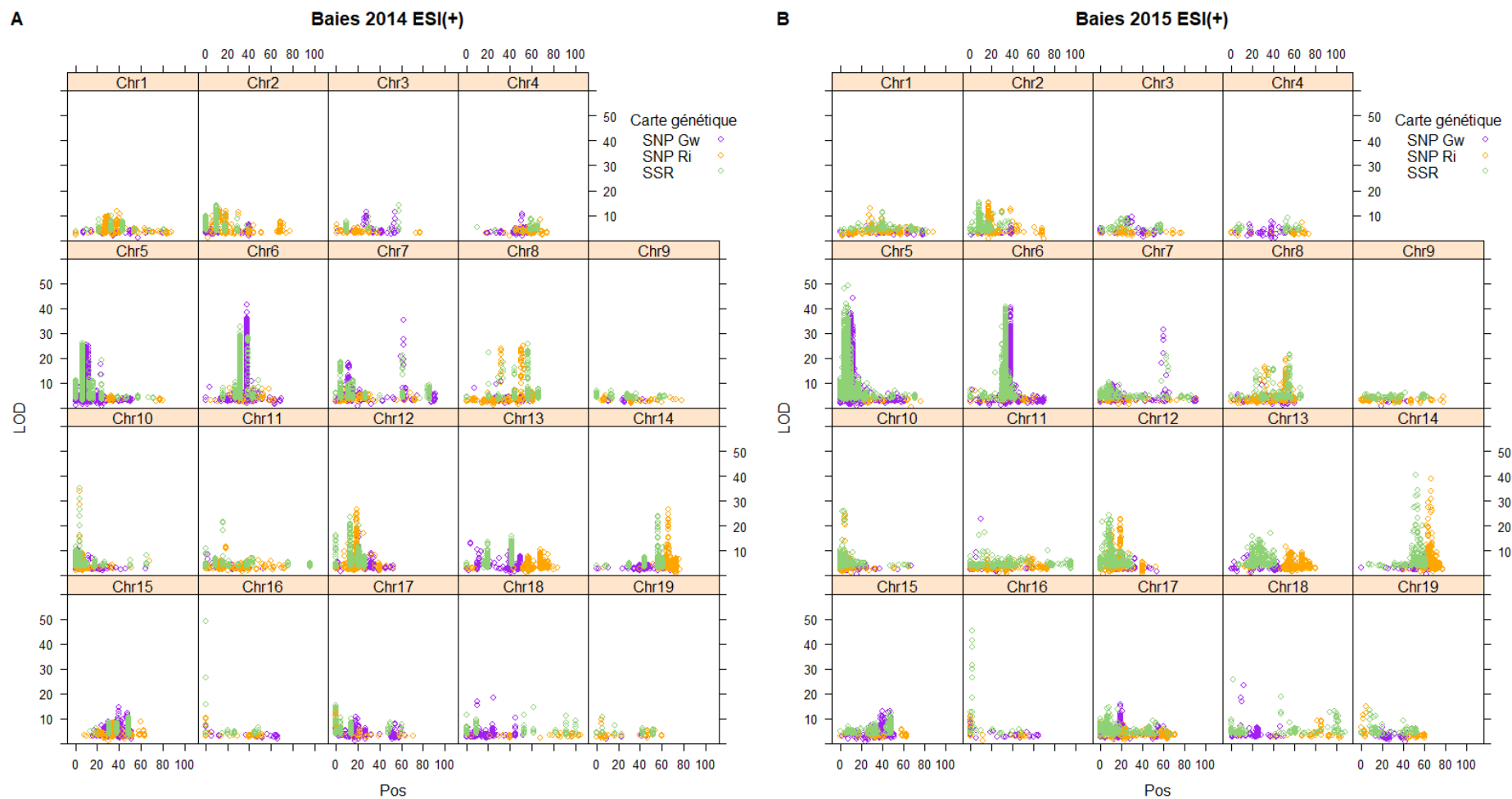
suppression d'ion dans les analyses par rapport aux analyses réalisées par ESI (+) du fait du moins grand nombre d'ions détectés et donc d'une quantification probablement plus fiable. Selon les cartes génétiques, le nombre de QTL est assez variable avec une constante : nombre QTL SSR > QTL Gw > QTL Ri.

Carte génétique	UHPLC-MS Baies 2014				UHPLC-MS Baies 2015			
	ESI (-)		ESI (+)		ESI (-)		ESI (+)	
	Nb QTL	Nb d'ions	Nb QTL	Nb de pics	Nb QTL	Nb d'ions	Nb QTL	Nb de pics
<b>SSR</b>	2949	2286	3992	3092	3310	2637	5633	4598
<b>SNP_RI</b>	1470	1171	2269	1760	1441	1216	2667	2264
<b>SNP_GW</b>	2533	2141	3269	2810	2913	2477	4712	4195

**Tableau 15 : Nombre de QTL détectés et nombre d'ions correspondants détectés avec les analyses de baies par UHPLC-MS en fonction des cartes génétiques (pval ≤ 0.05).**

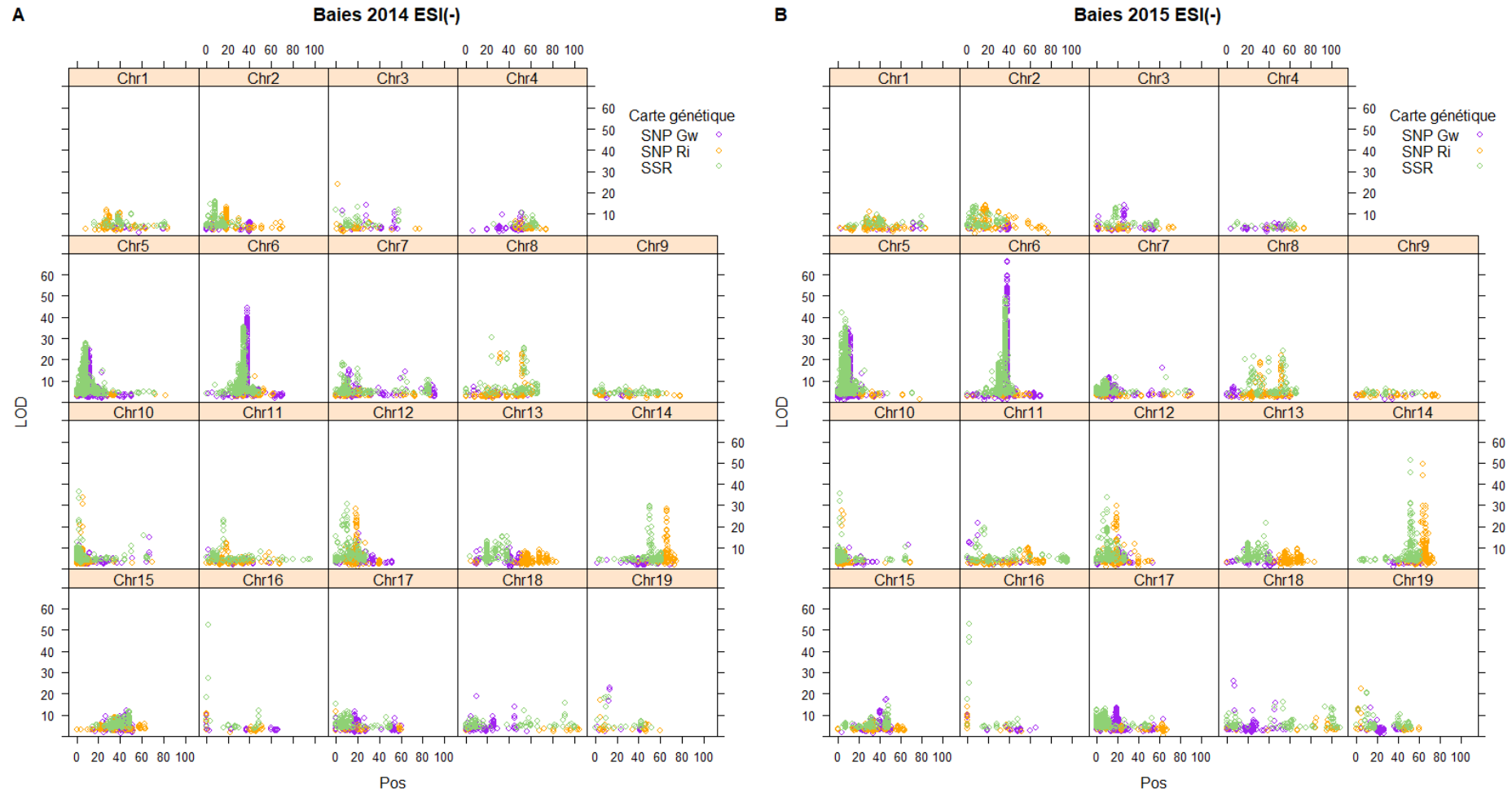
La Figure 29 et la Figure 30 représentent l'ensemble des QTL détectés avec les 3 cartes génétiques et les données de phénotypage du métabolome des baies prélevées en 2014 et en 2015. L'analyse QTL non ciblée à l'échelle du génome entier révèle des « **hotspots** » **génétiques**, se traduisant par une co-localisation de nombreux QTL possédant des LOD scores élevés. Ces régions du génome exercent donc des impacts majeurs sur grand nombre de composés du métabolome des baies. Ces hotspots sont présents notamment sur les chromosomes 5, 6, 12 et 14. Ces QTL sont reproductibles à la fois sur les deux années d'analyses mais aussi sur les analyses en ESI (+) et ESI (-).

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées



**Figure 29 : Détection de QTL à l'échelle du génome entier avec les analyses métabolomiques par UHPLC-MS ESI (+) de baies ( $p_{val} < 0.05$ ). (A) 2014 (B) 2015** Chaque point correspond à un ion placé en fonction des caractéristiques du QTL associé : sa position en cM sur chaque chromosome en abscisse et le LOD du QTL en ordonnée.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées



**Figure 30 : Détection de QTL à l'échelle du génome entier avec les analyses métabolomiques par UHPLC-MS ESI (-) de baies ( $p_{\text{val}} < 0.05$ ). (A) 2014 et (B) 2015** Chaque point correspond à un ion placé en fonction des caractéristiques du QTL associé : sa position en cM sur chaque chromosome en abscisse et le LOD du QTL en ordonnée.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### a. QTL impactants les molécules communes les plus abondamment détectées dans les baies

Parmi les 12 ions communs des top 20 les mieux détectés en ESI (+) en 2014 et en 2015 (Tableau 14), seuls 4 ions ont des QTL reproductibles sur les 2 années d'analyses sur les chromosomes 2, 5, 6 et 12. Ceci suggère que les hotspots génétiques impactent l'accumulation de certaines des molécules majeures détectées par UHPLC-MS.

Chromosome	Carte	Id	Molécule	Pos (cM)	LOD	pval
<b>Chr2</b>	SNP_Gw	<b>5</b>	<i>Glucoside de flavonoïde</i>	39.44-39.44	2.84-4.52	0-0.04
<b>Chr5</b>	SNP_Gw	<b>6</b>	<i>Apiosyl glucoside de linalol oxyde</i>	10-10.81	21.97-25.69	0
		<b>12</b>	<i>Fragment d'apiosyl glucoside de linalol oxyde</i>		24.08-25.72	0
	SSR	<b>6</b>	<i>Apiosyl glucoside de linalol oxyde</i>	6.5-8	23.26-28.29	0
		<b>12</b>	<i>Fragment d'apiosyl glucoside de linalol oxyde</i>		24.91-28.38	0
<b>Chr6</b>	SNP_Gw	<b>9</b>	<b>Astilbine</b>	38	31.79-36.39	0
		13	<b>Fragment astilbine</b>		29.59-31.25	0
	SSR	<b>9</b>	<b>Astilbine</b>	31.27-34	32.69	0
		<b>13</b>	<b>Fragment astilbine</b>		26.03-25.54-32.71	0
<b>Chr12</b>	SNP_Ri	<b>6</b>	<i>Apiosyl glucoside de linalol oxyde</i>	11.88-19.9	3.12-3.82	0.01-0.04
		12	<i>Fragment d'apiosyl glucoside de linalol oxyde</i>	19.9-19.9	3.27-3.51	0.01-0.05

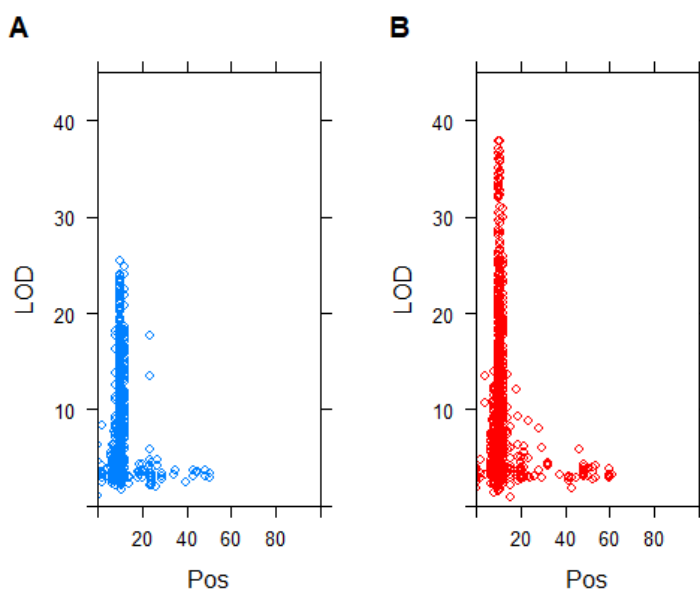
Tableau 16 : QTL liés aux 12 ions les plus abondants communs en 2014 et 2015, détectés dans les baies par UHPLC-MS ESI (+) (pval ≤ 0.05).



## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### b. Analyse détaillée du « hotspot » situé sur le chromosome 5, carte SNP\_Gw

La combinaison d'analyses métabolomiques et génétiques a relevé un « hotspot » génétique sur le chromosome 5, responsable de la variation d'un très grand nombre de molécules, représentées par 751 et 1128 ions respectivement en 2014 et en 2015.



**Figure 31 : QTL détectés avec les données UHPLC-MS ESI (+) de baies sur le chromosome 5 avec la carte SNP\_Gw ( $p_{\text{val}} \leq 0.05$ ). (A) : Baies 2014, (B) : Baies 2015.**

Parmi les ions reliés à ce QTL, j'ai pu identifier 351 ions pouvant correspondre à des molécules ou des fragments de molécules de type isoprénoïde. J'ai pu relier la plupart des m/z à des molécules de différentes familles d'isoprénoïdes : monoterpénols, norisoprénoïdes, sesquiterpènes et diterpènes (Tableau 17,

Tableau 18, Tableau 19 et Tableau 20).

La plupart des ions correspondent à des molécules qui appartiennent à la famille de monoterpénols (Tableau 17) : 108 ions correspondent à des m/z de monoterpénols. Grâce au nombre de carbones de la molécule, il est possible d'en déduire le type de glycoside. On trouve ainsi :

- 39 ions correspondant à des monoterpénols monoglycosylés (molécules avec 16 carbones),
- 20 ions correspondant à des monoterpénols glucoside-pentosides (molécules avec 21 carbones),
- 26 ions correspondant à des monoterpénols glucoside-hexosides (molécules en C22)
- 2 ions correspondant à des monoterpénols triglycosylés (molécules en C27).

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Parmi ces ions, l'ion de m/z 482.26 correspondant à [C<sub>21</sub>H<sub>36</sub>O<sub>11</sub> + NH<sub>4</sub>]<sup>+</sup> a déjà été identifié précédemment parmi les molécules présentes dans le top 20 des baies sur les deux années d'analyses pour la molécule identifiée au temps de rétention 4.2 comme le glucoside apiosyl du linalol oxyde. Il a un LOD score maximal de 36.22 en 2015, LOD score le plus élevé sur le chromosome 5 avec la carte génétique SNP\_Gw parmi les monoterpénols identifiés. Trois pics avec un m/z de 482.26 avec différents temps de rétention ont un QTL sur le chromosome 5 avec la carte génétique SNP\_GW. Ainsi, 3 monoterpénols ayant comme formule brute C<sub>21</sub>H<sub>36</sub>O<sub>11</sub> ont un QTL au même locus.

Ion pseudo-moléculaire monoterpénol	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD max	Nb de pics	LOD max	Nb de pics
[C <sub>10</sub> H <sub>14</sub> + H] <sup>+</sup>	135.1169	6.37-19.04	20.29	7	22.75	5
[C <sub>10</sub> H <sub>14</sub> O + H] <sup>+</sup>	151.1118	8.14-15.4	18	3	21.52	5
[C <sub>10</sub> H <sub>14</sub> O <sub>2</sub> + H] <sup>+</sup>	167.1067	10-10.81	7.66	3	20.56	3
[C <sub>10</sub> H <sub>14</sub> O <sub>4</sub> + H] <sup>+</sup>	199.0966	10-14	7.92	4	12.17	4
[C <sub>10</sub> H <sub>16</sub> O <sub>3</sub> - H <sub>2</sub> O + H] <sup>+</sup>	169.1224	8.14-12	16	4	28.69	6
[C <sub>10</sub> H <sub>16</sub> O <sub>3</sub> + H] <sup>+</sup>	185.1173	7.25-9.92	13.03	4	22.11	3
[C <sub>10</sub> H <sub>16</sub> O <sub>6</sub> + H] <sup>+</sup>	233.102	10	3.5	1	4.47	1
[C <sub>10</sub> H <sub>16</sub> O <sub>8</sub> + H] <sup>+</sup>	265.0919	9.92-10.81	6.97	2	15.63	2
[C <sub>10</sub> H <sub>18</sub> O - H <sub>2</sub> O + H] <sup>+</sup>	137.1326	9.92-21.71	17.85	7	27.84	5
[C <sub>10</sub> H <sub>18</sub> O + H] <sup>+</sup>	155.1431	9.92-15.4	11.42	2	15.19	5
[C <sub>10</sub> H <sub>18</sub> O <sub>2</sub> - H <sub>2</sub> O + H] <sup>+</sup>	153.1275	9.92-10.81	24.08	6	28.43	6
[C <sub>10</sub> H <sub>18</sub> O <sub>2</sub> + H] <sup>+</sup>	171.138	7.25-10.81	23.23	2	16.19	3
[C <sub>10</sub> H <sub>18</sub> O <sub>3</sub> + H] <sup>+</sup>	187.133	8.14-12	10.37	4	15.53	6
[C <sub>10</sub> H <sub>18</sub> O <sub>4</sub> + H] <sup>+</sup>	203.1278	7.25-10.81	6.49	1	10.86	2
[C <sub>10</sub> H <sub>18</sub> O <sub>5</sub> + H] <sup>+</sup>	219.1228	7.25	4.51	1	3.68	1
[C <sub>10</sub> H <sub>18</sub> O <sub>7</sub> + H] <sup>+</sup>	251.1126	12	4.17	1	3.58	1
[C <sub>10</sub> H <sub>20</sub> O + H] <sup>+</sup>	157.1588	9.92-10.81	17.31	3	26.92	2
[C <sub>10</sub> H <sub>20</sub> O <sub>2</sub> + H] <sup>+</sup>	173.1537	9.92-10.81	11.22	3	22.54	3
[C <sub>10</sub> H <sub>20</sub> O <sub>3</sub> + H] <sup>+</sup>	189.1486	7.25-9.92	7.18	2	3.65	2
[C <sub>10</sub> H <sub>20</sub> O <sub>4</sub> + H] <sup>+</sup>	205.1435	21.71	11.64	1	5.11	1
[C <sub>10</sub> H <sub>20</sub> O <sub>6</sub> + H] <sup>+</sup>	237.1333	10	11.28	1	5.14	1
[C <sub>16</sub> H <sub>24</sub> O <sub>6</sub> + H] <sup>+</sup>	313.1646	10.81	17.9	2	35.87	2
[C <sub>16</sub> H <sub>24</sub> O <sub>7</sub> + H] <sup>+</sup>	329.1595	10.81-12	16.11	4	21.76	3
[C <sub>16</sub> H <sub>26</sub> O <sub>13</sub> + H] <sup>+</sup>	427.1447	9.92-10.81	15.87	4	21.39	3
[C <sub>16</sub> H <sub>26</sub> O <sub>4</sub> + H] <sup>+</sup>	283.1904	10.81	4.79	1	5.94	1
[C <sub>16</sub> H <sub>26</sub> O <sub>5</sub> + H] <sup>+</sup>	299.1853	10-10.81	17.47	3	19.95	4
[C <sub>16</sub> H <sub>26</sub> O <sub>6</sub> + H] <sup>+</sup>	315.1802	9.92-12	14.64	2	28.5	2
[C <sub>16</sub> H <sub>26</sub> O <sub>6</sub> + NH <sub>4</sub> ] <sup>+</sup>	332.2068	12	2.59	1	2.93	1
[C <sub>16</sub> H <sub>26</sub> O <sub>7</sub> + H] <sup>+</sup>	331.1753	8.14-12	16.65	4	35.63	3
[C <sub>16</sub> H <sub>26</sub> O <sub>8</sub> + H] <sup>+</sup>	347.1701	10.81	15.89	3	26.7	2
[C <sub>16</sub> H <sub>26</sub> O <sub>8</sub> + NH <sub>4</sub> ] <sup>+</sup>	364.1966	12	3.15	1	5.91	1
[C <sub>16</sub> H <sub>28</sub> O <sub>10</sub> + H] <sup>+</sup>	381.1756	8.14	5.11	1	5.06	1
[C <sub>16</sub> H <sub>28</sub> O <sub>10</sub> + NH <sub>4</sub> ] <sup>+</sup>	398.2021	8.14	10.97	1	9.98	1
[C <sub>16</sub> H <sub>28</sub> O <sub>11</sub> + H] <sup>+</sup>	397.1706	9.92	18.63	1	11.13	1
[C <sub>16</sub> H <sub>28</sub> O <sub>11</sub> + Na] <sup>+</sup>	419.1525	9.92	9.2	1	10.01	1
[C <sub>16</sub> H <sub>28</sub> O <sub>11</sub> + NH <sub>4</sub> ] <sup>+</sup>	414.1971	9.92-46	19.06	1	14.39	2
[C <sub>16</sub> H <sub>28</sub> O <sub>14</sub> + H] <sup>+</sup>	445.1556	10.81	7.96	2	22.74	1
[C <sub>16</sub> H <sub>28</sub> O <sub>5</sub> + H] <sup>+</sup>	301.201	10.81-12	6.69	2	14.07	2
[C <sub>16</sub> H <sub>28</sub> O <sub>6</sub> + H] <sup>+</sup>	317.1959	9.92-10	21.82	4	23.18	3

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Ion pseudo-moléculaire monoterpénol	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD max	Nb de pics	LOD max	Nb de pics
[C16H28O6 + NH4] <sup>+</sup>	334.2224	9.92-10.81	8.19	2	12.98	2
[C16H28O7 + H] <sup>+</sup>	333.1908	9.92-10.81	22.73	3	32.29	3
[C16H28O7 + Na] <sup>+</sup>	355.1727	6.37-12	4.14	1	5.71	2
[C16H28O7 + NH4] <sup>+</sup>	350.2173	9.92-10	6.75	2	7.64	3
[C16H28O8 + H] <sup>+</sup>	349.1857	9.92	14.6	3	16.15	2
[C16H28O8 + Na] <sup>+</sup>	371.1677	8.14-10.81	3.85	1	5.44	1
[C16H28O8 + NH4] <sup>+</sup>	366.2123	9.92-10.81	11.94	3	16.89	5
[C16H28O9 + H] <sup>+</sup>	365.1807	10.81	15.34	1	29.87	1
[C16H28O9 + NH4] <sup>+</sup>	382.2072	10.81	12.18	2	24.14	1
[C16H30O10 + H] <sup>+</sup>	383.1912	7.25	3.44	2	3.16	1
[C16H30O11 + H] <sup>+</sup>	399.1862	10	15.92	1	9.79	1
[C16H30O11 + Na] <sup>+</sup>	421.1682	9.92	15.57	1	11.05	1
[C16H30O6 + H] <sup>+</sup>	319.2116	10	18.47	1	24.08	1
[C16H30O7 + H] <sup>+</sup>	335.2064	8.14-10.81	15.83	1	32.28	3
[C16H30O8 + H] <sup>+</sup>	351.2013	9.92-10.81	17.74	2	16.68	3
[C16H30O8 + Na] <sup>+</sup>	373.1833	9.92-10.81	15.14	1	10.92	1
[C16H30O8 + NH4] <sup>+</sup>	368.228	9.92-10	15.45	2	9.16	2
[C16H30O9 + H] <sup>+</sup>	367.1963	8.14-9.92	10.73	2	12.21	1
[C16H30O9 + Na] <sup>+</sup>	389.1783	1.78-10.81	10.28	1	9.11	1
[C16H30O9 + NH4] <sup>+</sup>	384.2229	10.81	11.44	1	18.55	1
[C16H32O7 + H] <sup>+</sup>	337.2221	8.14-12	24.89	1	29.88	3
[C21H36O10 + H] <sup>+</sup>	449.2383	9.92	11.88	3	13.79	2
[C21H36O10 + Na] <sup>+</sup>	471.2201	8.14-12	16.68	2	30.86	3
[C21H36O10 + NH4] <sup>+</sup>	466.2648	8.14-12	16.16	3	12.43	3
[C21H36O11 + H] <sup>+</sup>	465.2332	9.92-10.81	21.42	2	30.46	4
[C21H36O11 + Na] <sup>+</sup>	487.215	9.92-10.81	23.26	3	24.26	4
[C21H36O11 + NH4] <sup>+</sup>	482.2597	9.92-10.81	21.97	3	36.22	3
[C21H36O12 + H] <sup>+</sup>	481.2282	9.92-10.81	16.59	3	19.84	3
[C21H36O12 + Na] <sup>+</sup>	503.21	6.37-10.81	9.59	2	15.04	3
[C21H36O12 + NH4] <sup>+</sup>	498.2541	10.81	3.4	1	12.55	2
[C21H36O13 + H] <sup>+</sup>	497.223	10.81	13.19	1	17.32	1
[C21H38O10 + H] <sup>+</sup>	451.2541	10	16	1	16.69	1
[C21H38O10 + NH4] <sup>+</sup>	468.2804	10-10.81	18.04	1	21.72	2
[C21H38O11 + H] <sup>+</sup>	467.2489	9.92-10.81	15.29	2	29.39	2
[C21H38O11 + Na] <sup>+</sup>	489.2308	9.92-10.81	11.69	1	16.21	1
[C21H38O11 + NH4] <sup>+</sup>	484.2754	9.92	9.44	1	11.75	1
[C21H38O12 + H] <sup>+</sup>	483.2438	9.92-10.81	18.35	2	24.48	2
[C21H38O12 + Na] <sup>+</sup>	505.2256	9.92-10.81	12.56	1	11.51	2
[C21H38O12 + NH4] <sup>+</sup>	500.2703	9.92-12	11.05	2	21.63	2
[C21H38O13 + H] <sup>+</sup>	499.2386	8.14	8.05	1	12.35	1
[C21H38O13 + NH4] <sup>+</sup>	516.2652	10.81	9.53	1	18.74	1
[C22H34O12 + H] <sup>+</sup>	491.2125	10-10.81	4.57	2	16.34	2
[C22H34O14 + H] <sup>+</sup>	523.2024	10.81	7.56	1	13.34	1
[C22H34O14 + NH4] <sup>+</sup>	540.2293	10.81	6.86	1	8.31	1
[C22H36O10 + H] <sup>+</sup>	461.2382	12	7.97	1	18.33	1
[C22H36O11 + H] <sup>+</sup>	477.2331	10-10.81	8.89	2	20.56	2
[C22H36O11 + Na] <sup>+</sup>	499.2151	10.81	9.54	1	15.64	1
[C22H36O11 + NH4] <sup>+</sup>	494.2597	9.92-12	10.3	2	17	2
[C22H36O12 + H] <sup>+</sup>	493.2281	10-12	7.78	2	14.1	2
[C22H36O12 + NH4] <sup>+</sup>	510.2546	9.92-10.81	11.78	4	20.55	2
[C22H36O13 + H] <sup>+</sup>	509.2231	9.92-10.81	12.14	2	17.34	2
[C22H36O13 + Na] <sup>+</sup>	531.205	9.92-10.81	11.33	1	9.41	2
[C22H36O9 - H <sub>2</sub> O + H] <sup>+</sup>	445.2434	10.81	9.96	1	34.57	1

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Ion pseudo-moléculaire monoterpénol	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD max	Nb de pics	LOD max	Nb de pics
[C22H38O10 + H] <sup>+</sup>	463.2538	9.92-10.81	11.82	3	13.03	2
[C22H38O10 + Na] <sup>+</sup>	485.236	9.92-12	5.03	1	12.43	1
[C22H38O10 + NH <sub>4</sub> ] <sup>+</sup>	480.2804	9.92-12	14.52	1	25.4	1
[C22H38O11 + H] <sup>+</sup>	479.2491	12	9.92	1	16.57	3
[C22H38O11 + NH <sub>4</sub> ] <sup>+</sup>	496.2754	9.92-12	12.88	4	25.03	2
[C22H38O12 + H] <sup>+</sup>	495.2437	12	10.57	1	6.25	1
[C22H38O12 + NH <sub>4</sub> ] <sup>+</sup>	512.2703	9.92-10.81	12.87	2	12.62	3
[C22H38O13 + H] <sup>+</sup>	511.2389	10.81	11.54	2	18.12	2
[C22H38O13 + NH <sub>4</sub> ] <sup>+</sup>	528.2652	10.81	10.92	2	17.14	2
[C22H38O19 + H] <sup>+</sup>	607.2081	9.92	9.39	1	15.62	1
[C22H40O11 + H] <sup>+</sup>	481.2645	10.81	13.28	2	14.39	1
[C22H40O11 + NH <sub>4</sub> ] <sup>+</sup>	498.2911	9.92-10.81	7.99	1	14.23	1
[C22H40O12 + H] <sup>+</sup>	497.2594	10	5.68	1	8.34	1
[C22H42O12 + H] <sup>+</sup>	499.275	10.81	10.19	1	16.04	1
[C27H46O15 + NH <sub>4</sub> ] <sup>+</sup>	628.3177	10.81	10.87	1	20.51	1
[C27H46O16 + H] <sup>+</sup>	627.2858	9.92-10.81	6.68	3	10.02	3

**Tableau 17** : Liste des ions correspondant à des monoterpénols pour lesquels un QTL significatif ( $p_{\text{val}} \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

Cinquante-sept ions ayant un QTL sur le chromosome 5 avec la carte génétique SNP\_Gw ont pu être reliés à des sesquiterpènes, dont 24 ions correspondant à des sesquiterpènes glycosylés (

Tableau 18). L'ion de m/z 485.1995 correspondant à l'ion pseudo-moléculaire [C21H34O11 + Na]<sup>+</sup> a un LOD score de 37.87, LOD score le plus élevé parmi les sesquiterpènes.

Ion pseudo-moléculaire sesquiterpène	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD max	Nb de pics	LOD max	Nb de pics
[C15H16O2 + H] <sup>+</sup>	229.1224	48.52	2.7	1	3.39	1
[C15H18O2 + H] <sup>+</sup>	231.138	9.92-24.38	11.22	1	12.57	5
[C15H20 + H] <sup>+</sup>	201.1639	12	2.8	1	7.3	1
[C15H20O + H] <sup>+</sup>	217.1588	9.92	9.83	2	9.54	2
[C15H20O2 + H] <sup>+</sup>	233.1536	9.92-19.92	16.57	1	13.12	4
[C15H20O3 + H] <sup>+</sup>	249.1485	9.92-19.04	9.97	6	7.93	6
[C15H20O4 + H] <sup>+</sup>	265.1435	8.14-48.52	3.71	1	5.07	2
[C15H20O5 + H] <sup>+</sup>	281.1384	12	5.01	1	12.58	1
[C15H22 + H] <sup>+</sup>	203.1795	8.14-10.81	4.76	1	10.88	3
[C15H22O + H] <sup>+</sup>	219.1744	7.25-10	4.84	1	8.77	5
[C15H22O2 + H] <sup>+</sup>	235.1693	9.92	9.36	4	9.82	5
[C15H22O3 + H] <sup>+</sup>	251.1642	1.78-21.71	14.63	5	18.24	7
[C15H22O4 + H] <sup>+</sup>	267.1591	9.92-10.81	15.77	4	23.28	6
[C15H22O5 + H] <sup>+</sup>	283.154	9.92-10.81	13.51	3	37.1	5
[C15H22O6 + H] <sup>+</sup>	299.149	12-50.3	4.88	1	12.48	2
[C15H24 + H] <sup>+</sup>	205.1952	8.14-9.92	4.95	1	8.03	2
[C15H24O + H] <sup>+</sup>	221.1901	6.37-9.92	9.8	3	8.24	2
[C15H24O2 + H] <sup>+</sup>	237.185	8.14-10.81	8.96	2	10.89	7

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Ion pseudo-moléculaire sesquiterpène	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD max	Nb de pics	LOD max	Nb de pics
[C15H24O3 + H] <sup>+</sup>	253.1798	9.92-10.81	9.16	3	11.78	5
[C15H24O4 + H] <sup>+</sup>	269.1748	9.92-10.81	16.32	3	26.08	4
[C15H24O5 + H] <sup>+</sup>	285.1697	6.37-14	22.74	5	33.57	6
[C15H24O6 + H] <sup>+</sup>	301.1646	10.81-12	11.39	2	23.46	4
[C15H24O7 + H] <sup>+</sup>	317.1595	12	13.89	2	9.94	1
[C15H26O2 + H] <sup>+</sup>	239.2006	9.92-10.81	9.9	1	11.47	2
[C15H26O3 + H] <sup>+</sup>	255.1955	8.14-10	6.77	2	6.48	3
[C15H26O4 + H] <sup>+</sup>	271.1904	9.92-10.81	6.67	2	9.13	3
[C15H26O4 + Na] <sup>+</sup>	293.1722	9.92	5.99	1	9.5	2
[C15H26O5 + H] <sup>+</sup>	287.1853	9.92-12	24	4	27.09	6
[C15H26O6 + H] <sup>+</sup>	303.1803	9.92	23.84	3	33.24	2
[C15H26O7 + H] <sup>+</sup>	319.1752	10.81	10.96	2	27.42	1
[C15H26O8 + H] <sup>+</sup>	335.1701	12	12.57	1	10.06	1
[C15H28O5 + H] <sup>+</sup>	289.201	9.92-10	17.75	2	29.66	2
[C15H28O8 + H] <sup>+</sup>	337.1857	8.14	5.95	1	9.25	1
[C21H32O10 + H] <sup>+</sup>	445.2069	9.92-10.81	18.15	1	26.47	2
[C21H32O11 + H] <sup>+</sup>	461.2019	10.81	7.6	1	19.96	1
[C21H32O7 + H] <sup>+</sup>	397.2222	9.92	8.19	1	5.78	1
[C21H32O9 + H] <sup>+</sup>	429.2121	6.37-9.92	18.19	2	17.67	3
[C21H34O10 + H] <sup>+</sup>	447.2225	9.92-10	10.51	2	22.25	2
[C21H34O10 + Na] <sup>+</sup>	469.2045	10	12.93	1	11.33	1
[C21H34O10 + NH4] <sup>+</sup>	464.2491	10.81-12	8.52	1	10.12	4
[C21H34O11 + H] <sup>+</sup>	463.2176	9.92-12	15.78	3	32.21	3
[C21H34O11 + Na] <sup>+</sup>	485.1995	10.81-12	18.58	1	37.87	2
[C21H34O11 + NH4] <sup>+</sup>	480.2439	9.92-12	15.93	3	36.01	3
[C21H34O12 + H] <sup>+</sup>	479.2126	10.81	4.95	1	13.99	2
[C21H34O12 + Na] <sup>+</sup>	501.1943	10.81	4.45	1	9.04	1
[C21H34O12 + NH4] <sup>+</sup>	496.2389	10.81	5.03	1	12.96	1
[C21H34O8 + H] <sup>+</sup>	415.2327	10-10.81	11.41	2	8.88	2
[C21H34O9 + H] <sup>+</sup>	431.2275	9.92	12.34	1	23.11	1
[C21H34O9 + Na] <sup>+</sup>	454.2174	9.92-10.81	11.37	1	14.99	2
[C21H34O9 + NH4] <sup>+</sup>	448.2541	10	6.34	1	10.32	1
[C21H36O7 + H] <sup>+</sup>	401.2534	10.81	6.25	1	6.73	1
[C21H36O8 + H] <sup>+</sup>	417.2483	8.14-10.81	4.65	1	11.78	2
[C21H36O9 + H] <sup>+</sup>	433.2433	10-10.81	7.71	2	8.88	3
[C21H36O9 + NH4] <sup>+</sup>	450.2698	9.92	7.48	1	10.27	1
[C21H38O8 + H] <sup>+</sup>	419.264	10.81	7.08	1	6.92	1
[C21H40O10 + H] <sup>+</sup>	453.2695	10.81	5.92	1	9.81	1
[C21H40O12 + H] <sup>+</sup>	485.2595	10.81	13.3	1	14.32	1

**Tableau 18** : Liste des ions correspondant à des sesquiterpènes pour lesquels un QTL significatif ( $p_{val} \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

Quarante-six ions ayant un QTL sur le chromosome 5 avec la carte génétique SNP\_Gw ont pu être reliés à des diterpènes, dont 21 ions correspondant à des diterpènes glycosylés (Tableau 19). L'ion de m/z 598.307 correspondant à l'ion pseudo-moléculaire [C26H44O14 + NH4]<sup>+</sup> a un LOD score de 34.81 en 2015, LOD score le plus élevé parmi les diterpènes.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Ion pseudo-moléculaire diterpène	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD max	Nb de pics	LOD max	Nb de pics
[C20H26O + H] <sup>+</sup>	283.2057	12	6.18	1	10.3	1
[C20H26O2 + H] <sup>+</sup>	299.2006	10.81-12	11.42	2	25.18	3
[C20H28O + H] <sup>+</sup>	285.2213	9.92	7.63	2	14.31	2
[C20H28O2 + H] <sup>+</sup>	301.2162	9.92-12	7.27	4	12.9	4
[C20H28O3 + H] <sup>+</sup>	317.2112	10.81	9.19	2	20.37	2
[C20H30O + H] <sup>+</sup>	287.2369	12	7.6	1	11.35	1
[C20H30O2 + H] <sup>+</sup>	303.2318	10.81-12	11.2	3	23.36	2
[C20H30O3 + H] <sup>+</sup>	319.2268	10.81	9.1	3	13.65	3
[C20H30O4 + H] <sup>+</sup>	335.2217	10.81	13.98	3	20.54	4
[C20H30O5 + H] <sup>+</sup>	351.2166	10.81	12.83	1	17.19	1
[C20H32 + H] <sup>+</sup>	273.2577	10	15.39	1	16.78	1
[C20H32O2 + H] <sup>+</sup>	305.2475	9.92-10.81	5.16	1	10.98	2
[C20H32O3 + H] <sup>+</sup>	321.2424	10.81-12	13.93	3	19.39	3
[C20H32O4 + H] <sup>+</sup>	337.2374	10-12	9.56	3	14.88	7
[C20H32O5 + H] <sup>+</sup>	353.2322	12	11.43	3	11.34	1
[C20H34O3 + H] <sup>+</sup>	323.2581	12	7.96	1	7.17	1
[C20H34O4 + H] <sup>+</sup>	339.253	10.81-12	11.64	3	14.53	2
[C20H34O5 + H] <sup>+</sup>	355.2479	10.81-12	11.61	3	15.12	3
[C20H34O5 + Na] <sup>+</sup>	377.2298	10.81	5.88	1	5.34	1
[C20H34O6 + H] <sup>+</sup>	371.2428	10.81-12	14.56	1	19.04	2
[C20H36O4 + H] <sup>+</sup>	341.2687	9.92	4.85	1	5.68	1
[C20H36O5 + H] <sup>+</sup>	357.2635	10.81-12	9.8	1	11.45	2
[C20H36O6 + H] <sup>+</sup>	373.2585	9.92-12	9.59	3	12.87	3
[C20H38O5 + H] <sup>+</sup>	359.2792	12	6.28	1	7.59	2
[C20H38O6 + H] <sup>+</sup>	375.2742	10	4.75	1	4.42	1
[C26H38O8 + H] <sup>+</sup>	479.2637	10	6.38	1	10.94	1
[C26H40O9 + H] <sup>+</sup>	497.2749	12	11.6	1	9.12	1
[C26H42O10 + H] <sup>+</sup>	515.2852	8.14-10.81	11.53	2	16.58	2
[C26H42O10 + NH4] <sup>+</sup>	532.3117	9.92-10.81	12.41	2	12.74	2
[C26H42O11 + NH4] <sup>+</sup>	548.3066	9.92-10.81	12.25	2	12.48	2
[C26H42O9 + H] <sup>+</sup>	499.2903	10-10.81	7.69	2	14.32	2
[C26H44O10 + H] <sup>+</sup>	517.3009	10-10.81	10.48	3	15.73	2
[C26H44O10 + Na] <sup>+</sup>	539.2829	10-10.81	5.64	1	13.53	2
[C26H44O10 + NH4] <sup>+</sup>	534.3274	9.92	7.02	2	6.58	1
[C26H44O11 + H] <sup>+</sup>	533.2959	4-8.14	18.26	2	13.48	2
[C26H44O11 + NH4] <sup>+</sup>	550.3224	9.92-10.81	7.44	2	13.34	2
[C26H44O14 + NH4] <sup>+</sup>	598.307	9.92-10.81	18.16	2	34.81	2
[C26H44O15 + NH4] <sup>+</sup>	614.3019	10.81-12	3.67	1	14.93	2
[C26H44O8 + H] <sup>+</sup>	485.3111	9.92	7.07	1	6.82	1
[C26H44O9 + H] <sup>+</sup>	501.3059	10.81-12	5.64	1	19.25	2
[C26H46O10 + NH4] <sup>+</sup>	536.3432	12	7.86	1	13.63	1
[C26H46O11 + NH4] <sup>+</sup>	552.3384	10-12	7.92	1	8.39	2
[C26H46O12 + H] <sup>+</sup>	551.3064	8.14-9.92	13.43	2	6.58	2
[C26H46O12 + NH4] <sup>+</sup>	568.333	8.14-9.92	7.11	1	7.64	2
[C26H46O9 + H] <sup>+</sup>	503.3217	12	2.19	1	9.89	1
[C26H48O10 + H] <sup>+</sup>	521.3324	9.92	3.71	2	4.75	2

Tableau 19 : Liste des ions correspondant à des diterpènes pour lesquels un QTL significatif ( $p_{\text{val}} \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

Neuf ions correspondant à des fragments spécifiques de norisoprénoides ont un QTL sur le chromosome 5 avec la carte génétique SNP\_Gw, notamment pour l'ion de m/z 193.159

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

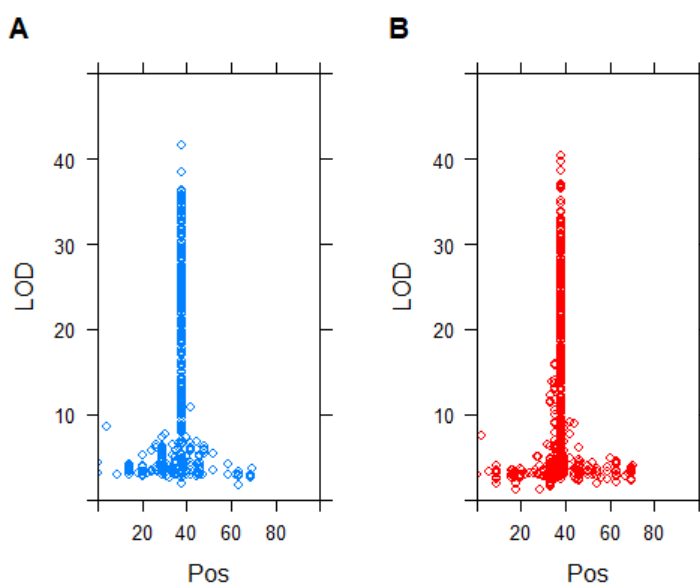
avec un LOD max de 14.22 (Tableau 20). Parmi ces m/z, on retrouve 5 ions correspondant à des norisoprénoïdes glucoside-pentoside (C<sub>24</sub>H<sub>40</sub>O<sub>11</sub>).

Ion pseudo-moléculaire norisoprénoïde	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD max	Nb de pics	LOD max	Nb de pics
[C <sub>13</sub> H <sub>18</sub> + H] <sup>+</sup>	175.1482	8.14-14	4.16	2	9.21	4
[C <sub>13</sub> H <sub>18</sub> O + H] <sup>+</sup>	191.1432	8.14-59.52	3.57	1	3.1	2
[C <sub>13</sub> H <sub>20</sub> + H] <sup>+</sup>	177.1639	10.81-14	3.06	1	7.79	3
[C <sub>13</sub> H <sub>20</sub> O + H] <sup>+</sup>	193.1589	10.81-14	14.22	2	7.42	2
[C <sub>24</sub> H <sub>34</sub> O <sub>12</sub> + NH <sub>4</sub> ] <sup>+</sup>	532.239	9.92	4.24	1	4.78	1
[C <sub>24</sub> H <sub>38</sub> O <sub>13</sub> + NH <sub>4</sub> ] <sup>+</sup>	552.2653	10	2.79	1	12.88	1
[C <sub>24</sub> H <sub>38</sub> O <sub>14</sub> + NH <sub>4</sub> ] <sup>+</sup>	568.2601	10.81	7.39	1	8.68	2
[C <sub>24</sub> H <sub>40</sub> O <sub>11</sub> + NH <sub>4</sub> ] <sup>+</sup>	522.2911	4-10.81	4.38	1	10.76	2
[C <sub>24</sub> H <sub>40</sub> O <sub>14</sub> + NH <sub>4</sub> ] <sup>+</sup>	570.2756	9.92	8.48	1	8.25	1

**Tableau 20** : Liste des ions correspondant à des norisoprénoïdes pour lesquels un QTL significatif ( $p_{\text{val}} \leq 0.05$ ) est détecté sur le chromosome 5 avec la carte SNP\_Gw pour les deux années d'analyses.

### c. Analyse détaillée du « hotspot » situé sur le chromosome 6, carte SNP\_Gw

Un hotspot a été détecté sur le chromosome 6 avec la carte SNP\_Gw à un locus à 40 cM. En 2014 et 2015, 408 et 647 ions différents sont reliés à ce locus.



**Figure 32** : QTL détectés avec les données UHPLC-MS ESI (+) de baies sur le chromosome 6 avec la carte SNP\_Gw ( $p_{\text{val}} \leq 0.05$ ). (A) : Baies 2014, (B) : Baies 2015.

Parmi ces ions ayant un QTL sur le chromosome 6 de la carte SNP\_Gw, 62 m/z ont pu être annotés et sont retrouvés sur les deux années de prélèvement (Tableau 21). Des ions correspondant à des molécules de la famille des flavonoïdes ont été identifiés avec notamment

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

un ion correspondant à une molécule d'astilbine précédemment identifiée qui a un m/z de 451.123.

Ion pseudo-moléculaire	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD Max	Nb de pics	LOD Max	Nb de pics
[C13H10O4 + H] <sup>+</sup>	231.0654	38	10.49	1	14.07	1
[C13H12O6 + NH4] <sup>+</sup>	282.0973	35.15-63.31	3.16	1	2.79	1
[C13H12O7 + H] <sup>+</sup>	281.0656	37.82-38	33.31	2	31.11	2
[C13H14O2 + H] <sup>+</sup>	203.1067	28-39.61	3.33	1	5.02	1
[C13H14O8 + H] <sup>+</sup>	299.0761	37.82-38	29.81	2	32.08	2
[C13H16O7 + H] <sup>+</sup>	285.0969	38	4.42	1	6.17	1
[C14H10O4 + H] <sup>+</sup>	243.0652	37.82-38	25.63	1	6.85	1
[C14H10O5 + H] <sup>+</sup>	259.0602	37.82-42	27.44	1	31.54	3
[C15H10O5 + H] <sup>+</sup>	271.0601	37.82-38	31.71	2	21.76	3
[C15H10O6 + H] <sup>+</sup>	287.055	17.2-39.61	25.81	4	10.66	6
[C15H10O7 + H] <sup>+</sup>	303.0499	37.82-38	29.23	1	29.52	1
[C15H12O6 + H] <sup>+</sup>	289.0706	33.36-44	36.14	3	35.02	5
[C15H12O7 + H] <sup>+</sup>	305.0656	37.82-52	31.25	4	29.59	4
[C15H20O5 + H] <sup>+</sup>	281.1384	38-45.08	15.7	1	3.59	1
[C16H14O4 + H] <sup>+</sup>	271.0964	35.15-38	24.66	1	6.22	1
[C16H14O7 + H] <sup>+</sup>	319.0813	33.36-63.31	14.14	3	15.66	4
[C17H14O7 + H] <sup>+</sup>	331.0811	37.82-38	33.95	1	28.27	1
[C17H14O8 + H] <sup>+</sup>	347.0762	35.15-38	27.28	1	21.3	2
[C17H18O9 + NH4] <sup>+</sup>	384.1289	37.82-42	4.33	1	3.81	1
[C18H26O10 + H] <sup>+</sup>	403.1601	35.15-50	4.48	1	10.19	2
[C18H30O12 + H] <sup>+</sup>	439.1816	38	13.38	1	13	1
[C20H18O10 + H] <sup>+</sup>	419.0972	35.15-37.82	15.01	1	13.64	1
[C20H18O11 + H] <sup>+</sup>	435.0923	37.82-38	24.27	1	22.05	1
[C20H20O10 + H] <sup>+</sup>	421.1129	38	32.7	1	30.21	1
[C20H20O11 + H] <sup>+</sup>	437.1079	38	22.78	1	27.23	1
[C20H20O11 + Na] <sup>+</sup>	459.0899	37.82-38	19.76	1	18.67	1
[C21H18O8 + H] <sup>+</sup>	399.1074	37.82-38	32.57	1	27.51	1
[C21H18O9 + H] <sup>+</sup>	415.1025	38	27.24	1	29.71	1
[C21H20O10 + H] <sup>+</sup>	433.113	37.82-60	29.16	2	29.22	4
[C21H20O9 + H] <sup>+</sup>	417.118	37.82-38	35.65	1	30.42	1
[C21H22O10 + H] <sup>+</sup>	435.1287	37.82-38	35.89	2	36.71	2
[C21H22O10 + Na] <sup>+</sup>	457.1104	37.82-38	35.66	1	27.45	1
[C21H22O10 + NH4] <sup>+</sup>	452.155	37.82-38	31.75	1	27.89	2
[C21H22O11 + H] <sup>+</sup>	451.1235	35.15-38	34.06	2	32.69	4
[C21H22O11 + Na] <sup>+</sup>	473.1055	37.82-38	35.35	2	29.27	2



## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

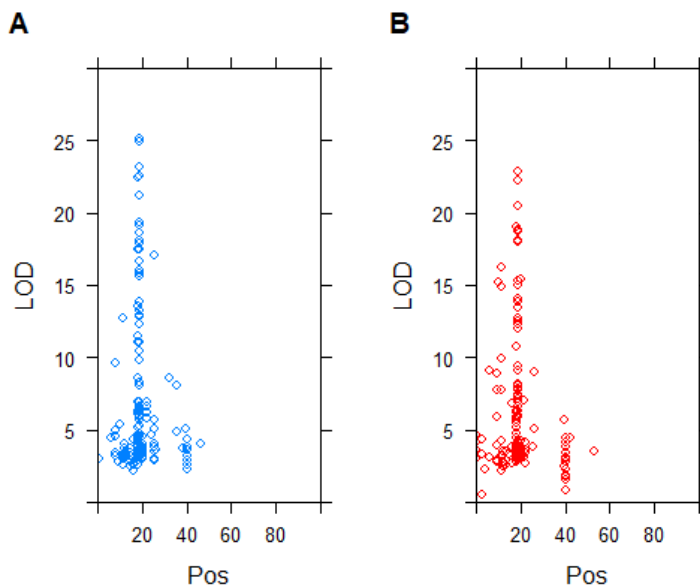
Ion pseudo-moléculaire	m/z	Position génétique (cM)	Baies 2014		Baies 2015	
			LOD Max	Nb de pics	LOD Max	Nb de pics
[C21H22O11 + NH4] <sup>+</sup>	468.1502	38	33.21	2	31.17	1
[C21H22O12 + H] <sup>+</sup>	467.1185	37.82-38	23.36	1	27.67	2
[C21H22O12 + Na] <sup>+</sup>	489.1005	37.82-38	23.39	1	21.57	1
[C21H22O12 + NH4] <sup>+</sup>	484.1451	38	20.43	1	24.29	1
[C21H22O9 + H] <sup>+</sup>	419.1337	38	5.43	1	5.19	1
[C21H24O10 + H] <sup>+</sup>	437.1443	37.82-39.61	6.71	1	8.42	2
[C22H20O13 + H] <sup>+</sup>	493.0977	38	3.32	1	3.44	1
[C22H22O11 + H] <sup>+</sup>	463.1235	35.15-37.82	9.59	1	14	1
[C22H24O11 + H] <sup>+</sup>	465.1392	38	14.14	1	16.01	1
[C23H16O10 + NH4] <sup>+</sup>	470.1083	42-45.97	5.94	1	3.27	1
[C25H19O8 + Na] <sup>+</sup>	470.0972	37.82-38	33.96	1	25.03	1
[C26H26O13 + NH4] <sup>+</sup>	564.1713	45.97-46	5.19	1	3.08	1
[C26H28O10 + H] <sup>+</sup>	501.1754	34-37.82	4.25	1	3.13	1
[C26H28O11 + H] <sup>+</sup>	517.1706	38	4.39	1	2.76	1
[C26H32O12 + H] <sup>+</sup>	537.1967	33.36-37.82	3.42	1	4.32	1
[C27H28O13 + H] <sup>+</sup>	561.1604	38	18.23	1	16.31	1
[C27H28O17 + H] <sup>+</sup>	625.1401	37.82-38	20.48	1	8.43	1
[C27H32O14 + H] <sup>+</sup>	581.1865	37.82-38	22.55	1	17.4	1
[C27H32O15 + H] <sup>+</sup>	597.1817	38	20.78	1	22.01	1
[C27H32O16 + H] <sup>+</sup>	613.1765	35.15-37.82	17.64	2	10.59	2
[C27H32O16 + NH4] <sup>+</sup>	630.2029	34-37.82	18.39	1	16.68	2
[C27H32O17 + H] <sup>+</sup>	629.1713	38	17.09	1	14.66	1
[C29H26O15 + NH4] <sup>+</sup>	632.1609	42-45.97	6.62	1	3.06	1
[C29H26O16 + NH4] <sup>+</sup>	648.1559	39.61-44.2	5.91	1	4.29	1
[C30H18O10 + H] <sup>+</sup>	539.098	37.82-38	36.25	1	13.99	1
[C33H42O19 + H] <sup>+</sup>	743.2394	37.82-38	14.99	1	17.08	1
[C39H52O23 + H] <sup>+</sup>	889.2973	38	10.56	1	12.81	1

**Tableau 21** : Liste des ions correspondant à des flavonoïdes pour lesquels un QTL significatif ( $p_{\text{val}} \leq 0.05$ ) est détecté sur le chromosome 6 avec la carte SNP\_Gw pour les deux années d'analyses.

### d. Analyse détaillée du « hotspot » situé sur le chromosome 12, carte SNP\_Ri

En 2014 et 2015, 280 et 296 ions respectivement ont un QTL sur le chromosome 12 avec la carte SNP\_RI (Figure 33). On observe deux QTL à des locus à 20 et 40 cM sur les deux années d'analyses

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées



**Figure 33 : QTL détectés avec les données UHPLC-MS ESI (+) de baies sur le chromosome 12 avec la carte SNP\_Ri ( $p_{val} \leq 0.05$ ). (A) : Baies 2014, (B) : Baies 2015.**

Dans le cadre de ma thèse, je me suis concentrée sur le locus situé à 20 cM car j'y ai trouvé des ions spécifiques à des molécules de types norisoprénoïdes, monoterpénols et sesquiterpène (Tableau 22).

Ion	m/z	Famille	Position génétique (cM)	Baies 2014		Baies 2015	
				LOD max	Nb de pics	LOD max	Nb de pics
[C13H20 + H <sup>+</sup> ]	177.164	Norisoprénoïde	22-39.28	6.58	1	5.68	1
[C13H18O + H <sup>+</sup> ]	191.1432	Norisoprénoïde	18.13-19.01	6.25	1	3.69	1
[C13H18O2 + H <sup>+</sup> ]	207.1381	Norisoprénoïde	19.01	8.05	1	8.04	1
[C13H20O2 + H <sup>+</sup> ]	209.1537	Norisoprénoïde	19.01	17.48	1	12.54	1
[C24H38O14 + NH <sub>4</sub> <sup>+</sup> ]	568.2601	Norisoprénoïde	19.9	3.67	1	3.01	2
[C10H14 + H <sup>+</sup> ]	135.1169	Monoterpénol	18.13-42	4.45	1	4.41	1
[C10H18O2 + H <sub>2</sub> O + H <sup>+</sup> ]	153.1274	Monoterpénol	18.13-19.9	5.13	2	5.69	2
[C10H20O3 + H <sup>+</sup> ]	189.1486	Monoterpénol	19.01-11	18.59	1	7.7	1
[C16H28O7 + H <sup>+</sup> ]	333.1908	Monoterpénol	11.88-19.9	3.18	1	3.08	1
[C16H30O7 + NH <sub>4</sub> <sup>+</sup> ]	352.233	Monoterpénol	14.55-18.13	3.96	1	3.34	1
[C19H30O9 + H <sup>+</sup> ]	403.1964	Monoterpénol	19.01	3.54	1	4.13	1
[C22H36O10 + H <sup>+</sup> ]	461.2382	Monoterpénol	22-21.68	6.94	1	7.01	1
[C21H34O10 + NH <sub>4</sub> <sup>+</sup> ]	464.2491	Sesquiterpène	19.01-40.16	6.95	1	3.44	1
[C21H36O11 + H <sup>+</sup> ]	465.2332	Monoterpénol	11.88-19.9	3.62	2	3.5	1
[C21H36O11 + NH <sub>4</sub> <sup>+</sup> ]	482.26	Monoterpénol	11.88-19.9	3.12	1	3.82	1

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Tableau 22 : Liste des ions correspondant à des norisoprénoides, monoterpénols et sesquiterpène pour lesquels un QTL significatif ( $p_{val} \leq 0.05$ ) est détecté sur le chromosome 12 avec la carte SNP\_Ri pour les deux années d'analyses.

### 2.2. Étude des précurseurs d'arômes présent dans les baies par UHPLC-MS ESI (+)

Parmi les milliers d'ions détectés, certains correspondent à des molécules qui appartiennent à des précurseurs d'arômes. Durant ma thèse, j'ai décidé de me concentrer sur les monoterpénols et sur les norisoprénoides, molécules importantes participant à la typicité des vins de Ri et de Gw.

#### 2.2.1. Analyse ciblée des monoterpénols dans les baies

Les monoterpénols participent fortement à la typicité des vins de Gw et représentent un pool important de précurseurs d'arômes dans les baies pouvant être libérés lors de la vinification, c'est pourquoi j'ai voulu étudier leur diversité et leur déterminisme génétique dans les baies de la population RixGw.

##### a. Recherche des ions issus des monoterpénols dans les baies

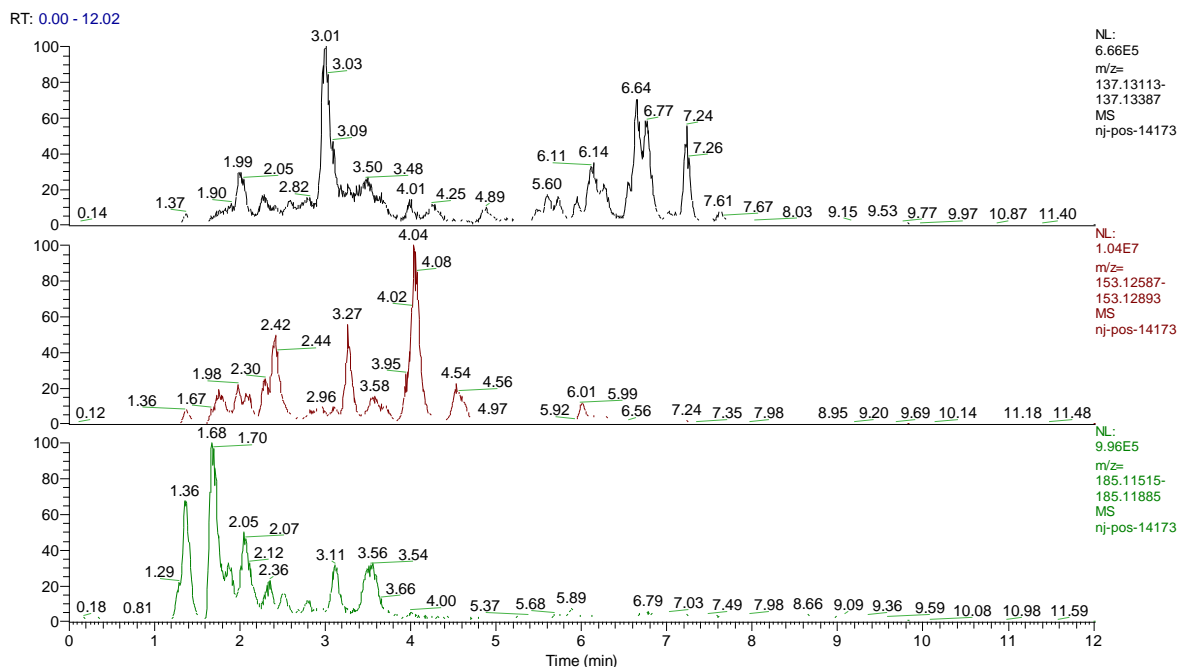
Afin d'analyser les monoterpénols présents dans les extraits de baies analysées par UHPLC-MS ESI (+), j'ai tout d'abord identifié des fragments d'ions ayant des m/z caractéristiques des monoterpénols. Cette recherche n'est pas exhaustive mais correspond aux fragments de monoterpénols les plus fréquemment trouvés dans les analyses de baies de la population RixGw. Pour chaque année d'analyse de baies, les ions ayant des m/z de 137.133, 153.127 et 185.117 ont été recherchés. Pour chaque ion recherché, plusieurs ions ayant des temps de rétention différents sont détectés (Tableau 23).

m/z	Type de monoterpénol	Fragment monoterpénol	Nombre de pics en 2014	Nombre de pics en 2015
137.133	Monoterpénol	$[C_{10}H_{18}O - H_2O + H]^+$	9	8
153.127	Monoterpénol oxydé ou hydroxylé	$[C_{10}H_{18}O_2 - H_2O + H]^+$	6	7
185.117	Carboxy-monoterpénol	$[C_{10}H_{16}O_3 + H]^+$	5	3

Tableau 23 : Nombre de fragments de monoterpénols détectés dans les analyse de baies par UHPLC-MS ESI (+).

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Un exemple de courants d'ions observés pour les fragments de monoterpénols recherchés dans une analyse d'un descendant de la population RixGw est représenté dans la Figure 34. Pour le courant d'un m/z donné, on observe bien plusieurs pics à des temps de rétention différents avec des aires de pics différentes.



**Figure 34 : Courant d'ions des trois fragments caractéristiques de monoterpénols recherchés dans une analyse par UHPLC-MS d'extrait SPE de baies d'un génotype de la population (0074E) en 2014.**

Parmi ces fragments de monoterpénols identifiés dans chaque jeu de données UHPLC-MS ESI (+) de baies, je n'ai gardé que les fragments caractéristiques des monoterpénols qui ont été retrouvés sur les deux années d'analyses avec des temps de rétention comparables.

Au total, 15 monoterpénols différents ont été suivi de manière plus précise pour la proposition de structure, leur quantification et les analyses QTL (Tableau 24).

**Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées**

Fragment monoterpénol	Proposition de structure / Identification par un standard	Ion pseudo-moléculaire	Temps de rétention (min)	
			2014	2015
137-1	Apiosyl-glucoside de terpénol	[C <sub>21</sub> H <sub>38</sub> O <sub>11</sub> + H] <sup>+</sup> = 467.249 [C <sub>21</sub> H <sub>38</sub> O <sub>11</sub> + NH <sub>4</sub> ] <sup>+</sup> = 484.275	2.0	2.1
137-2	Apiosyl-glucoside de terpénol	[C <sub>21</sub> H <sub>38</sub> O <sub>11</sub> + H] <sup>+</sup> = 467.249 [C <sub>21</sub> H <sub>38</sub> O <sub>11</sub> + NH <sub>4</sub> ] <sup>+</sup> = 484.275	3.0	3.2
137-3	Glucoside de terpénol	[C <sub>16</sub> H <sub>28</sub> O <sub>6</sub> + H] <sup>+</sup> = 317.196	4.0	4.2
137-4	Terpènediol libre	[C <sub>10</sub> H <sub>18</sub> O - H <sub>2</sub> O + H] <sup>+</sup> = 137.133	4.9	5.2
137-5	Terpènediol libre	[C <sub>10</sub> H <sub>18</sub> O - H <sub>2</sub> O + H] <sup>+</sup> = 137.133	5.7	6.2
137-6	Rutinoside de terpénol	[C <sub>22</sub> H <sub>38</sub> O <sub>10</sub> + H] <sup>+</sup> = 463.254 [C <sub>22</sub> H <sub>38</sub> O <sub>10</sub> + NH <sub>4</sub> ] <sup>+</sup> = 480.280	6.7	6.8
137-7	Malonyl-glucoside de terpénol	[C <sub>19</sub> H <sub>30</sub> O <sub>9</sub> + H] <sup>+</sup> = 403.192 [C <sub>19</sub> H <sub>30</sub> O <sub>9</sub> + NH <sub>4</sub> ] <sup>+</sup> = 420.223	7.2	7.3
153-1	Apiosyl-glucoside de linalol oxyde	[C <sub>21</sub> H <sub>36</sub> O <sub>11</sub> + H] <sup>+</sup> = 465.233 [C <sub>21</sub> H <sub>36</sub> O <sub>11</sub> + NH <sub>4</sub> ] <sup>+</sup> = 482.260	3.3	3.4
153-2	Apiosyl-glucoside de linalol oxyde	[C <sub>21</sub> H <sub>36</sub> O <sub>11</sub> + H] <sup>+</sup> = 465.233 [C <sub>21</sub> H <sub>36</sub> O <sub>11</sub> + NH <sub>4</sub> ] <sup>+</sup> = 482.260	4.1	4.2
153-3	Glucoside de terpénol oxydé	[C <sub>16</sub> H <sub>28</sub> O <sub>7</sub> + H] <sup>+</sup> = 333.191 [C <sub>16</sub> H <sub>28</sub> O <sub>7</sub> + NH <sub>4</sub> ] <sup>+</sup> = 350.217	4.6	4.7
153-4	Malonyl-glucoside de terpénol oxydé	[C <sub>19</sub> H <sub>30</sub> O <sub>10</sub> + H] <sup>+</sup> = 419.191 [C <sub>19</sub> H <sub>30</sub> O <sub>10</sub> + NH <sub>4</sub> ] <sup>+</sup> = 436.218	6.0	6.4
153-5	<b>Oxyde furanique de linalool</b>	[C <sub>10</sub> H <sub>18</sub> O <sub>2</sub> - H <sub>2</sub> O + H] <sup>+</sup>	7.2	7.3
185-1	Glucoside de carboxylinalool	[C <sub>21</sub> H <sub>36</sub> O <sub>13</sub> + H] <sup>+</sup> = 497.223 [C <sub>21</sub> H <sub>36</sub> O <sub>13</sub> + NH <sub>4</sub> ] <sup>+</sup> = 514.25	1.7	1.7
185-2	Apiosyl-glucoside de carboxylinalool	C <sub>21</sub> H <sub>34</sub> O <sub>12</sub> + H] <sup>+</sup> = 479.21230 C <sub>21</sub> H <sub>34</sub> O <sub>12</sub> + NH <sub>4</sub> ] <sup>+</sup> = 496.239	2.3	2.4
185-3	Glucoside carboxylinalool	[C <sub>16</sub> H <sub>26</sub> O <sub>8</sub> + H] <sup>+</sup> = 347.170 [C <sub>16</sub> H <sub>26</sub> O <sub>8</sub> + NH <sub>4</sub> ] <sup>+</sup> = 364.197	3.1	3.3

**Tableau 24 : Monoterpénols libres ou glycosidés identifiés dans les extraits de baies analysés par UHPLC-MS ESI (+).**

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Concernant les aglycones identifiés, 7 ont un m/z de 137.133, 5 ont un m/z de 153.127, 3 ont un m/z de 185.117.

D'après Hjelmeland and Ebeler, 2015, les sucres pouvant être liés aux molécules de type monoterpénols sont au nombre de 5 (Figure 35) :

- Glucoside simple
- Malonyl-glucoside (glucose lié à un groupement malonyl)
- Arabino-glucoside (glucose lié à un arabinose)
- Apiosyl-glucoside (glucose lié à un apiose)
- Rutinoside (glucose lié à un rhamnose)

Les glycones arabino-glucoside et apiosyl-glucoside ont exactement la même masse et ne peuvent être différenciés que par une analyse précise de leur structure par RMN. Les aglycones de type apiosyl-glucosides sont plus fréquents et abondants que les arabinoglucosides, c'est pourquoi dans ma thèse, lorsqu'un hexose-pentose est identifié, la proposition est en premier lieu l'apiosyl-glucoside comme glycone.

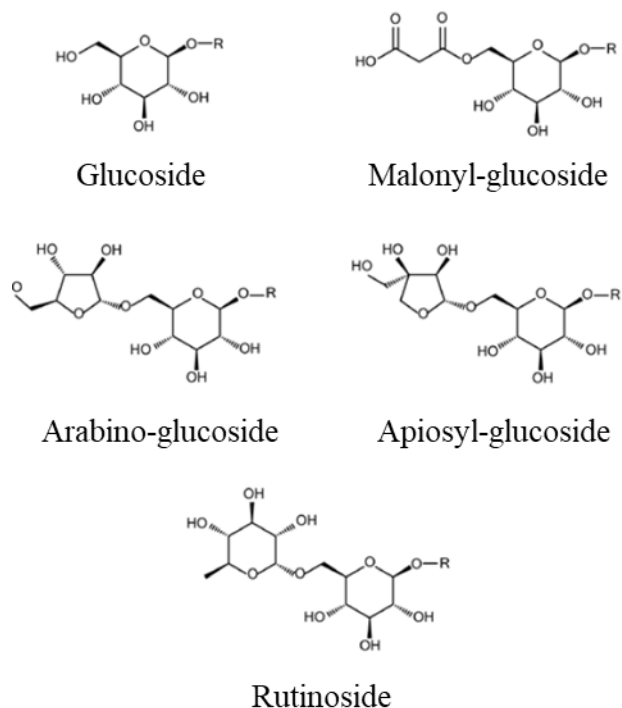


Figure 35 : Sucres pouvant être liés aux terpénols (Hjelmeland and Ebeler, 2015).

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Parmi ces 15 monoterpénols, 3 monoterpénols ont été identifiés comme étant sous forme libre (137-4, 137-5 et 153-5), les autres étant liés à un ou plusieurs sucres.

Parmi les glycones, j'ai pu identifier :

- 4 monoterpénols lié à un glucose,
- 5 monoterpénols lié à un glucose-pentoses (apiose ou arabinose),
- 1 monoterpénol lié à un glucose-rhamnose (rutinoside),
- 2 monoterpénols liés à un groupement malonyl-glucoside.

Seul l'oxyde furanique de linalol au temps de rétention 7.2 min a pu être confirmé par l'injection d'un standard correspondant. Les autres molécules n'ont pas pu être confirmées car il n'existe pas de standard commercial de monoterpénols glycosylés à ce jour.

L'identification des monoterpénols présents dans les extraits de baies a été poursuivie par la quantification des fragments de monoterpénols dans la population RixGw et dans le Ri et dans le Gw grâce aux données issues de l'alignement des analyses par UHPLC-MS ESI (+) des baies de la population RixGw en 2014 et 2015 (Tableau 25).

Cette analyse m'a permis d'identifier les fragments de monoterpénols les plus abondamment détectés dans les baies. L'aire de pic des fragments de monoterpénols est significativement différentielle entre le Gw pour le Ri sur deux années d'analyses pour 3 fragments : 137-6, 153-2 et 185-1. Les fragments de monoterpénols les plus abondamment détectés dans les analyses sont les fragments correspondants à des monoterpénols oxydés ou hydroxylés qui ont un m/z de 153.127 (153-1 à 153-5).

**Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées**

Fragment	Année	Moyenne Ri	Moyenne Gw	Minimum RixGw	Moyenne RixGw	Maximum RixGw	h <sup>2</sup>
137-1	2014	9.74E+05	8.00E+05	2.56E+05	1.19E+06	6.76E+06	0.85
	2015	9.83E+05	8.74E+05	4.48E+03	1.60E+06	7.73E+06	0.79
137-2	2014	3.78E+06	4.97E+06	5.42E+05	5.36E+06	2.25E+07	0.70
	2015	3.48E+06	6.79E+06	3.98E+03	6.17E+06	2.34E+07	0.71
137-3	2014	6.50E+05	9.34E+05	3.48E+05	8.66E+05	2.91E+06	0.71
	2015	6.48E+05	1.43E+06 *	1.78E+03	1.55E+06	4.11E+06	0.57
137-4	2014	8.83E+05 **	4.62E+05	3.05E+05	7.50E+05	3.02E+06	0.79
	2015	1.13E+06	8.89E+05	3.70E+03	1.39E+06	2.99E+06	0.66
137-5	2014	1.22E+06	2.14E+06 **	3.12E+05	2.41E+06	1.77E+07	0.93
	2015	9.97E+05	1.33E+06	2.22E+03	1.68E+06	5.88E+06	0.69
<b>137-6</b>	2014	<b>1.25E+06</b>	<b>6.36E+06 **</b>	<b>4.61E+05</b>	<b>3.71E+06</b>	<b>1.39E+07</b>	<b>0.65</b>
	2015	<b>2.54E+06</b>	<b>1.19E+07 **</b>	<b>4.90E+03</b>	<b>7.41E+06</b>	<b>3.55E+07</b>	<b>0.73</b>
137-7	2014	4.59E+05	9.61E+05 *	2.20E+05	1.05E+06	4.90E+06	0.75
	2015	4.27E+05	1.90E+06 ***	7.31E+03	1.42E+06	5.87E+06	0.86
153-1	2014	1.12E+07	1.84E+07	2.36E+06	1.98E+07	7.80E+07	0.78
	2015	6.59E+06	2.15E+07 *	6.72E+03	1.70E+07	7.23E+07	0.59
<b>153-2</b>	2014	<b>2.43E+07</b>	<b>4.83E+07 ***</b>	<b>4.42E+06</b>	<b>4.03E+07</b>	<b>1.17E+08</b>	<b>0.90</b>
	2015	<b>2.48E+07</b>	<b>4.77E+07 ***</b>	<b>1.08E+03</b>	<b>4.11E+07</b>	<b>1.59E+08</b>	<b>0.86</b>
153-3	2014	6.12E+06	8.70E+06	7.03E+05	1.58E+07	8.57E+07	0.68
	2015	5.41E+06	2.98E+06	3.15E+03	1.05E+07	7.78E+07	0.85
153-4	2014	8.84E+06 ***	3.86E+06	4.00E+05	1.23E+07	8.47E+07	0.98
	2015	6.41E+06	3.93E+06	8.00E+00	1.01E+07	7.12E+07	0.91
153-5	2014	5.09E+05	5.28E+05	4.37E+05	6.02E+05	1.52E+06	0.08
	2015	3.91E+05	9.43E+05 ***	6.81E+03	8.52E+05	1.71E+07	0.96
<b>185-1</b>	2014	<b>1.84E+06</b>	<b>2.80E+06 **</b>	<b>5.75E+05</b>	<b>2.51E+06</b>	<b>1.13E+07</b>	<b>0.89</b>
	2015	<b>1.16E+06</b>	<b>2.49E+06 **</b>	<b>7.54E+02</b>	<b>2.60E+06</b>	<b>8.46E+06</b>	<b>0.89</b>
185-2	2014	9.35E+05	1.53E+06	4.53E+05	1.42E+06	7.89E+06	0.65
	2015	6.06E+05	1.24E+06	7.39E+03	1.61E+06	6.10E+06	0.71
185-3	2014	1.08E+06	1.05E+06	3.87E+05	1.22E+06	3.42E+06	0.79
	2015	5.07E+05	5.08E+05	5.52E+03	9.13E+05	5.44E+06	0.86

**Tableau 25 : Récapitulatif de la quantification des fragments de monoterpénols détectés dans les baies de la population RixGw.** Les fragments ayant des aires de pics significativement différentes entre le Ri et le Gw sur les deux années sont en gras.



## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### b. Analyse génétique des fragments de monoterpénols dans les baies

Une recherche QTL a été effectuée sur les fragments spécifiques des monoterpénols avec les 2 années d'analyses de baies par UHPLC-MS ESI (+). Seuls les QTL reproductibles sur les 2 années d'analyses sont présentés ici (Tableau 26).

Chr	Carte génétique	Position génétique (cM)	Fragment monoterpénol	LOD	pval
5	SNP Gw	6.37-10.81	137-7	6.31-20.27	0
		7.25-8.14	185-2	3.98-4.77	0
		9.92	137-2	13.06-24.58	0
		9.92	137-3	7.1-8.38	0
		9.92-10.81	137-6	16.79-27.84	0
		9.92	153-3	4.99-6.8	0
		9.92-10.81	153-4	5.41-6.51	0
		9.92	185-1	13.03-22.11	0
		9.92	185-3	6.93-10.32	0
		10	153-1	15.87-25.22	0
	10-10.81	153-2	24.08-25.72	0	
	SSR	0-6.5	137-7	6.91-21.85	0
		2-6.5	185-2	4.29-4.36	0.01-0.03
		4-6.5	137-2	13.46-26.38	0
		4-6.5	185-1	13.63-23.29	0
		4-6.5	185-3	7.02-11.3	0
		6-6.5	153-3	5.31-7.11	0
		6.5	137-3	8.26-8.72	0
		6.5-8	137-6	17.35-30.13	0
		6.5	153-1	17.6-26.48	0
6.5-8		153-2	24.91-28.38	0	
6.5-18	153-4	7.35-9.62	0		
10	SNP Ri	2.65-5.36	153-3	3.87-4.25	0-0.01
		6.24-7.13	185-3	2.56-3.25	0.03-0.05
	SSR	0	153-3	5.32-5.77	0
		0	185-3	4.18-5.1	0.01-0.02
12	SNP Ri	18.13	153-4	5.13-5.69	0
		19.9	153-2	3.27-3.51	0.01-0.05
	SSR	12-26.37	153-4	5.94-5.95	0
13	SNP Ri	52.39-57.71	137-1	2.99-3.97	0-0.02
	SSR	19.99-20	137-1	4.08-7.4	0-0.03

Tableau 26 : QTL détectés pour les fragments de monoterpénols détectés dans les analyses par UHPLC-MS ESI (+) reproductibles sur les deux années d'analyses de baies (pval  $\leq$  0.05).

Des QTL reproductibles ont été détectés sur les chromosomes 5, 10, 12 et 13. La quasi-totalité des fragments sont liés au QTL sur le chromosome 5 relatif à la DXS. Deux fragments ont un QTL sur le chromosome 10. On peut supposer que ces composés sont des dérivés du linalol, comme l'a montré l'analyse des données GC-MS. Le fragment 137-1 a un QTL sur le

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

chromosome 13. On peut supposer que ce composé soit un dérivé d' $\alpha$ -terpinéol. Deux fragments ont un QTL sur le chromosome 12 avec les cartes SNP Ri et SSR.

### c. Analyse détaillée du monoterpénol majeur lié au fragment 153-2

L'analyse des fragments de monoterpénols a permis de relever le fragment 153-2 comme fragment le plus abondant dans les analyses par UHPLC-MS ESI (+) de baies dans la population RixGw ainsi que dans les analyses de baies de Ri et de Gw. Deux QTL, sur les chromosomes 5 et 12, ont été mis en évidence pour ce fragment. C'est pourquoi, j'ai voulu étudier de manière plus précise la molécule reliée à ce fragment avec notamment une proposition de structure et en vérifiant la cohérence des données entre le fragment de monoterpénol et l'ion pseudo-moléculaire correspondant.

#### ❖ Proposition de structure

L'étude de la structure de la molécule reliée au 153-2 a été basée sur le spectre de masse au temps de rétention 4.2 minutes dans l'échantillon ayant la plus forte aire de pic (Figure 36).

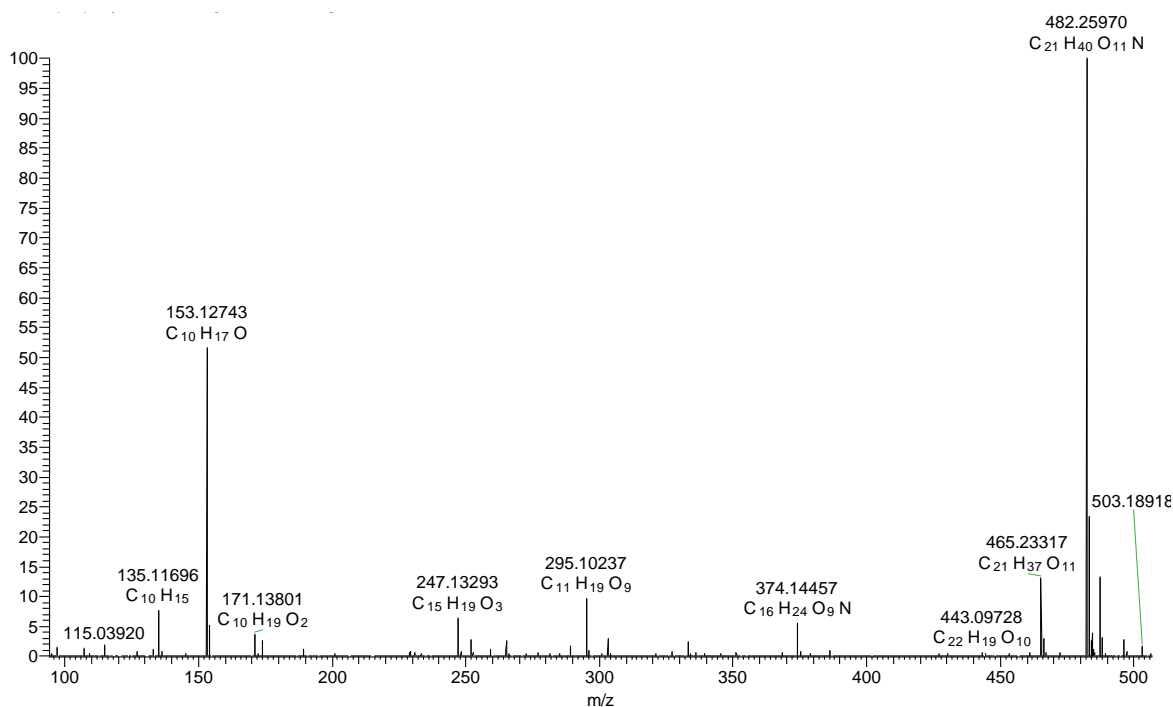


Figure 36 : Spectre de masse ESI (+) au temps de rétention 4.2min dans l'échantillon ayant la plus forte aire de pic pour le m/z 153.127 au temps de rétention 4.3 min en 2015.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Sur ce spectre de masse, on peut observer que l'ion pseudo-moléculaire le plus intense a un m/z de 482.260, correspondant à  $[C_{21}H_{36}O_{11} + NH_4]^+$ . On observe aussi l'ion avec un m/z de 465.233 correspondant au  $[C_{21}H_{36}O_{11} + H]^+$ . Un fragment de monoterpénoïde oxydé glycosylé avec un m/z de 333.191 correspond à  $[C_{16}H_{28}O_7 + H]^+$  et un fragment d'apiosylglucose avec un m/z de 295.102.

Cette observation est confirmée par les corrélations entre les données de quantification de ces fragments dans toutes les analyses de baies réalisées par UHPLC-MS (Tableau 27). La quantification des ions détectés au temps de rétention 4.3 min sont corrélées avec un R supérieure à 0.97 pour les ions avec un m/z de 482.260 et 295.102.

Ion	m/z	Ion pseudo-moléculaire	Corrélation avec M153T253
M153T253	153.127	$[C_{10}H_{18}O_2 - H_2O + H]^+$	1.00
M295T253	295.102	$[C_{11}H_{18}O_9 + H]^+$	0.97
M482T253	482.260	$[C_{21}H_{36}O_{11} + NH_4]^+$	0.97
M277T253	277.092	$[C_{11}H_{18}O_9 - H_2O + H]^+$	0.95
M465T253	465.233	$[C_{21}H_{36}O_{11} + H]^+$	0.93
M259T253	259.081	$[C_{11}H_{18}O_9 - 2 H_2O + H]^+$	0.92
M135T253	135.117	$[C_{10}H_{18}O_2 - 2 H_2O + H]^+$	0.91
M333T253	333.191	$[C_{16}H_{28}O_7 + H]^+$	0.86
M171T253	171.138	$[C_{10}H_{18}O_2 + H]^+$	0.84
M487T253	487.215	$[C_{21}H_{36}O_{11} + Na]^+$	0.79

Tableau 27 : Corrélation de la quantification des fragments avec la quantification du fragment 153-2 en 2015 (M153T253).

Les différents fragments nous indiquent que la molécule est un monoterpénoïde oxydé ou hydroxylé lié à un apiosyl glucoside. Une identification probable de cette molécule est l'apiosyl glucoside de linalol (Figure 37). En effet cette molécule est présente de manière abondante dans les vins de Gw où elle a déjà été identifiée et caractérisée (Baltenweck-Guyot et al., 2000).

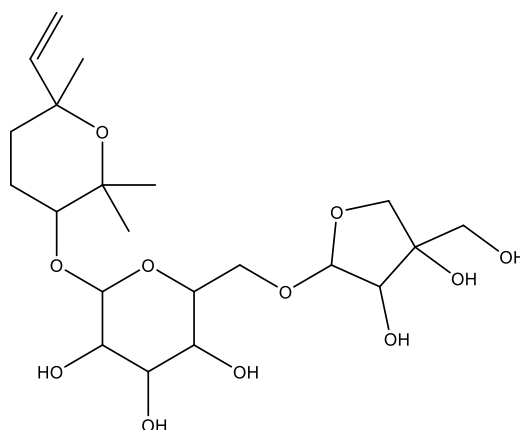


Figure 37 : Proposition de structure de la molécule reliée au fragment de monoterpénoïde 153-2, l'apiosyl-glucoside de linalol oxydé.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### ❖ Quantification

Pour confirmer les observations réalisées à partir de la quantification du fragment 153-2, j'ai étudié l'ion avec un  $m/z$  482.260 au temps de rétention 4.2 min sur les deux années d'analyses de baies (Tableau 28).

La quantification de l'apiosyl glucoside de linalol oxyde en prenant l'ion avec un  $m/z$  482.260 au temps de rétention 4.2 min permet de confirmer la différence d'abondance observée avec le fragment 153-2, plus élevée chez le Gw que chez le Ri sur les deux années d'analyses.

Année	Moyenne Ri	Moyenne Gw	Min RixGw	Moyenne RixGw	Max RixGw	$h^2$
2014	5.13E+07	1.73E+08 *	7.22E+06	8.93E+07	2.64E+08	0.89
2015	4.68E+07	1.02E+08 *	0.00E+00	8.01E+07	2.81E+08	0.79

Tableau 28 : Récapitulatif de la quantification du  $m/z$  482.260 au temps de rétention 4.2 min sur les deux années d'analyse par UHPLC-MS ESI (+) de baies du Ri, du Gw et de la population de croisement RixGw.

L'observation de la ségrégation dans la population de la distribution de la molécule (Figure 38) permet de voir que ce n'est pas une distribution normale : 50 à 60% des descendants présentent des aire de pics faibles voire nulles.

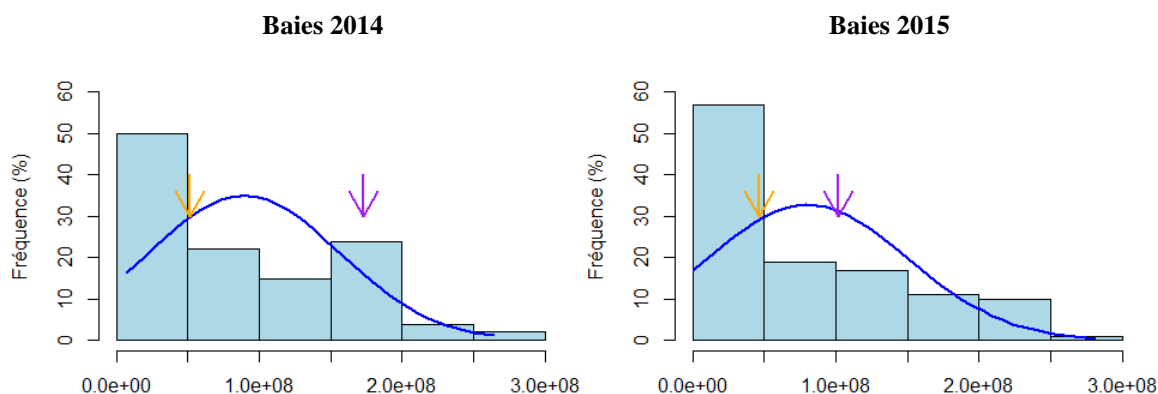


Figure 38 : Histogrammes de la distribution de l'aire de pic de l'ion avec un  $m/z$  de 482.260 au temps de rétention 4.2 min détecté dans les baies de la descendance RixGw par UHPLC-MS ESI (+). Les flèches en orange indiquent les moyennes observées dans le Ri, celles en violet indique celles observées dans le Gw.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### ❖ Analyse génétique

Une analyse QTL a été réalisée avec la quantification réalisée précédemment dans la population de croisement RixGw (Tableau 29).

Année	Carte génétique	Ion	LOD 5%	LOD 20%	Marqueur génétique	Chr	Pos (cM)	LOD	pval
2014	SNP_Gw	M482T243	3.18	2.41	c5.loc10	5	10	21.97	0
2015	SNP_Gw	M482T253	3.08	2.27	chr5b4758001	5	10.81	25.69	0
2014	SSR	M482T243	3.88	3.25	DXS1	5	6.5	23.26	0
2015	SSR	M482T253	4.05	3.35	c5.loc8	5	8	28.29	0
2014	SNP_Ri	M482T243	3.04	2.37	chr12b6485036	12	11.88	3.12	0.04
2015	SNP_Ri	M482T253	3.05	2.35	chr12b8519447	12	19.9	3.82	0.01

**Tableau 29 : QTL détectés pour l'ion avec un m/z 482.260 au temps de rétention 4.2 min dans les baies RixGw (pval ≤ 0.05).**

Deux QTL ont été identifiés de manière non ciblée, l'un sur le chromosome 5 avec la carte SNP\_Gw, colocalisant avec la DXS, et sur le chromosome 12 avec la carte SNP\_Ri. Ces deux QTL sont reproductible sur les deux années de prélèvements (Tableau 29).

- *Modèle avec la carte génétique SSR*

Avec la carte SSR, l'étude génétique de l'ion M482T253 de 2015 m'a permis d'identifier un modèle additif avec 3 QTL ( $y \sim Q1 + Q2 + Q3$ ) expliquant 81% de la variance du contenu en apiosyl-glucoside d'oxyde de linalol dans la population RixGw avec un LOD de 42 (Tableau 30) :

- Un QTL situé sur le chromosome 5 lié à la DXS expliquant 63.5 % de la variance,
- 2 autres QTL obtenus grâce à la commande `addqt1`,
- Un QTL sur le chromosomes 10 et lié au scaffold 271, expliquant 6 % de la variance,
- Un QTL sur le chromosome 12 expliquant 7.7 % de la variance.

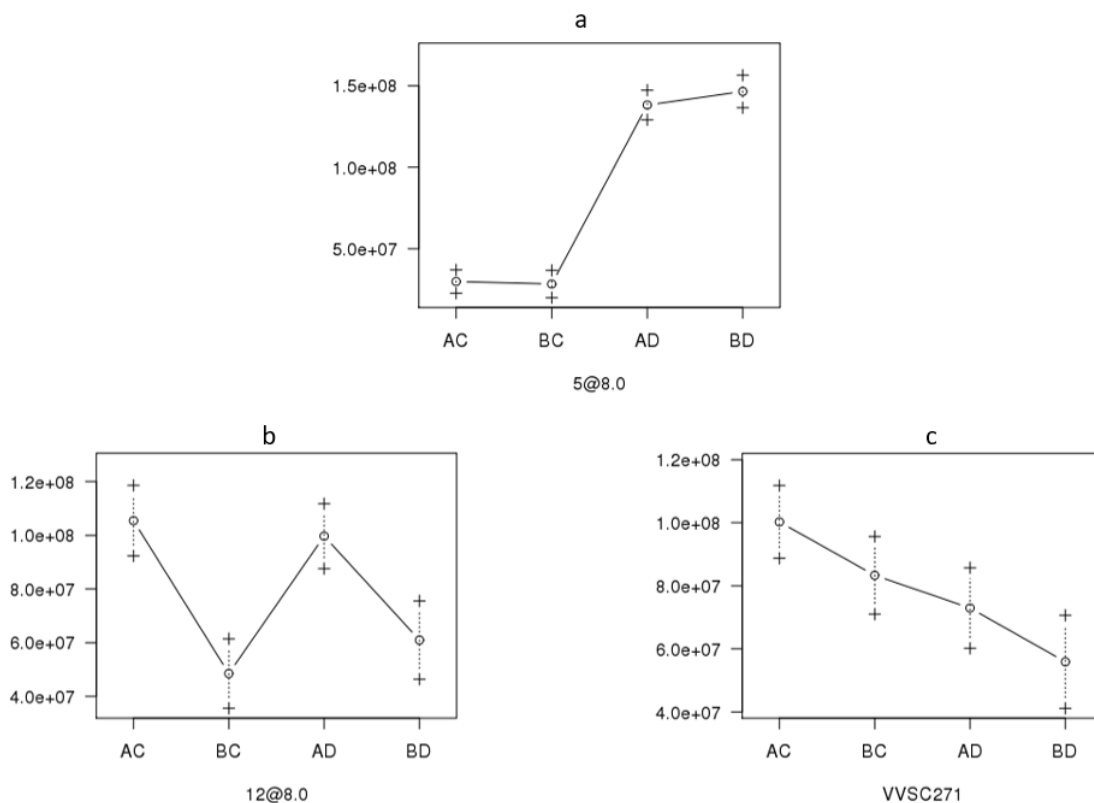
En 2014, le même modèle a été observée avec des QTL à 5@8, 10@4.2 et 12@22 avec respectivement 64%, 7.4% et 5 % de variance expliquée.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Source de variation	Degrés de liberté	Somme des carrés des écarts	Carrés moyens	LOD	% variance expliquée	pval ( $\chi^2$ )	pval (F)
Modèle	9	4.47E+17	4.97E+16	41.7	81.4	0 ***	0 ***
5@8.0	3	3.49E+17		36.8	63.5	0 ***	< 2E-16 ***
10@4.2	3	3.30E+16		6.9	6	0 ***	1.9E-06 ***
12@8.0	3	4.24E+16		8.6	7.7	0 ***	6.5E-08 ***
Erreur	104	1.02E+17	9.81E+14				
Totale	113	5.49E+17					

**Tableau 30 : Modèle génétique expliquant le contenu en apiosyl-glucoside de linalol oxyde en 2015 avec la carte génétique SSR.**

Les effets en 2015 des différents QTL sont illustrés dans la Figure 39.



**Figure 39 : Effet des QTL de la carte génétique SSR sur l'aire de pic de la molécule d'apiosyl-glucoside en 2015 dans la population Ri(AB) x Gw(CD). (a) QTL 5@8, (b) QTL 12@8, (c) QTL 10@4.2.**

- Pour le QTL 5@8, la présence de l'allèle D du Gw (AD et BD) dans la population RixGw multiplie par 3 le contenu en apiosyl-glucoside de linalol oxyde.
- Pour le QTL 12@8, c'est la présence de l'allèle A du Ri (AC et AD) qui multiplie par 2 le contenu en apiosyl-glucoside de linalol oxyde par rapport aux deux autres combinaisons alléliques.
- Pour le QTL 10@4 situé sur le scaffold 271, on voit clairement un effet additif des allèles C du Gw et A du Ri. C'est l'association des allèles A et C qui donne les

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

contenus en apiosyl-glucoside de linalol oxyde les plus élevées dans la population RixGw.

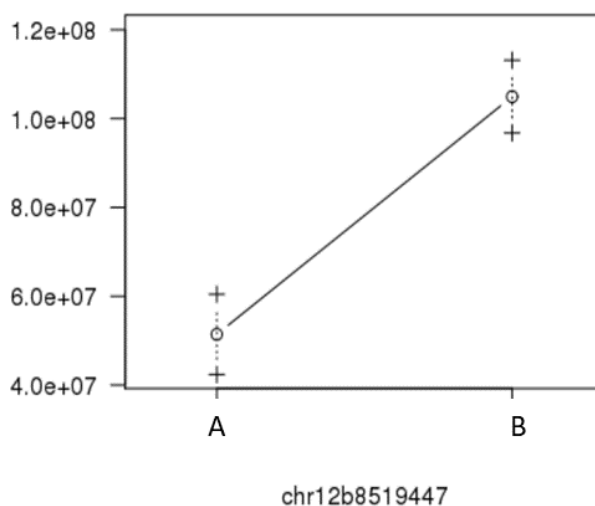
- *Modèle avec la carte SNP\_Ri*

Avec la carte SNP\_Ri, l'étude génétique de l'ion M482T253 de 2015 m'a permis d'identifier un modèle avec un seul QTL situé sur le chromosome 12 expliquant 14.3% de la variance du contenu en apiosyl-glucoside d'oxyde de linalol dans la population RixGw avec un LOD de 3.8. Aucun QTL additionnel n'a été trouvé. En 2014, le même modèle a été observée avec des QTL à 12@11.9 avec 11.5 % de variance expliquée.

Source de variation	Degrés de liberté	Somme des carrés des écarts	Carrés moyens	LOD	% variance expliquée	pval ( $\chi^2$ )	pval (F)
12@19.9	1	7.9e+16	4.97E+16	3.8	14.3	2.7e-05 ***	3.3e-05 ***
Erreur	112	4.7E+17	4.20E+15				
Totale	113	5.5E+17					

**Tableau 31 : Modèle génétique expliquant le contenu en apiosyl-glucoside de linalol oxyde en 2015 avec la carte génétique SNP\_Ri.**

Le QTL sur le chromosome 12 de la carte Ri est dû à l'allèle B au marqueur chr12b8519447 dans la population, qui ont un contenu en apiosyl-glucoside de linalol oxyde 2 fois plus élevé que pour l'allèle A.



**Figure 40 : Effet du QTL 12@19.9 de la carte génétique SNP\_Ri sur l'aire de pic de la molécule d'apiosyl-glucoside de linalol oxyde en 2015. (A) : allèle A du Ri (B) : allèle B du Ri.**

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### ❖ Conclusion sur l'apiosyl-glucoside de linalol oxyde

L'apiosyl-glucoside de linalol oxyde est le monoterpénol le plus abondamment retrouvé dans les baies de Ri, de Gw et de la population RixGw. Cette molécule est présente de manière abondante dans les vins de Gw où elle a déjà été identifiée et caractérisée (Baltenweck-Guyot et al., 2000). Des aires de pics dans les baies de Gw sont significativement plus élevées que celles de Ri dans nos analyses. L'analyse génétique du contenu en apiosyl-glucoside dans les baies de la population RixGw montre un déterminisme lié à 3 locus génétique : un locus important lié au Gw sur le chromosome 5 et deux loci liés au Ri sur les chromosomes 10 et 12. L'apiosyl-glucoside de linalol oxyde pourrait être hydrolysé et libéré lors de la vinification pour donner du linalol oxyde, molécule ayant des caractéristiques organoleptiques intéressantes : les descriptifs olfactifs le décrivant sont de type floral, vert, lavande (TheGoodScentCompagny).

### 2.2.2. Analyse ciblée des norisoprénoides dans les baies

Les norisoprénoides sont issus de la dégradation de molécules de types caroténoïdes et participent à la typicité des vins de Ri. En effet, l'arôme « kérosène » typique de vieux vins de Ri est liée à la présence de TDN, molécule qui n'est pas présente dans les baies. Le précurseur du TDN dans les baies n'a pas encore été identifié. C'est pourquoi j'ai voulu étudier la diversité des norisoprénoides détectés dans les baies, leur déterminisme génétique et de trouver de potentiels précurseurs de TDN.

#### a. Recherche des ions issus des norisoprénoides dans les baies

Pour analyser les norisoprénoides présents dans les extraits de baies analysées par UHPLC-MS ESI (+), j'ai tout d'abord identifié des ions ayant des m/z caractéristiques des norisoprénoides dans les baies prélevées en 2014 et 2015 (Tableau 32). Cette recherche n'est pas exhaustive mais correspond aux fragments de norisoprénoides les plus fréquemment trouvés dans les analyses. Pour chaque m/z recherché, on retrouve plusieurs ions ayant des temps de rétentions différents.

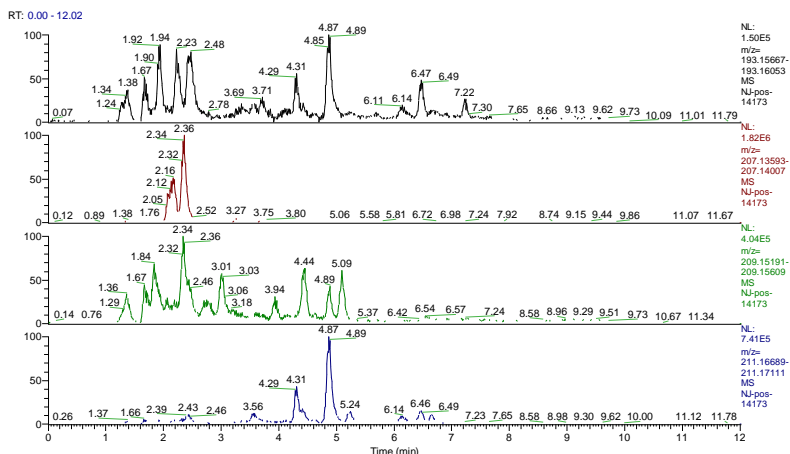
m/z	Fragment norisoprénoides	Nombre d' ions en 2014	Nombre d'ions en 2015
193.159	$[C_{13}H_{22}O_2 - H_2O + H]^+$	7	8
207.138	$[C_{13}H_{20}O_3 - H_2O + H]^+$	3	4
209.154	$[C_{13}H_{22}O_3 - H_2O + H]^+$	7	7

Tableau 32 : Nombre de fragments de norisoprénoides détectés dans les analyse de baies par UHPLC-MS ESI(+).



## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Un exemple de courants d'ions observés pour les fragments de norisoprénoides recherchés dans une analyse d'un descendant de la population RixGw est représenté dans la Figure 41. Pour un courant d'un m/z donné, on observe bien plusieurs pics à des temps de rétention différents avec des aires de pics différentes.



**Figure 41 : Courant d'ions des fragments caractéristiques de norisoprénoides d'une analyse par UHPLC-MS d'extrait SPE de baie d'un génotype de la population (0074E) en 2014.**

Des propositions de structures ont pu être faite à partir des spectres de masses au temps de rétention correspondant au fragment de norisoprénoides (Tableau 33) ainsi que les corrélations entre les différents ions à un temps de rétention donné. Ainsi 17 norisoprénoides, pour la plupart présents sous forme glycosylée ont été identifiés.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Fragment norisoprénoïde	Proposition de structure	Ion pseudo-moléculaire	Temps de rétention	
			2014	2015
193-1	Rutinoside de norisoprénoïde	$[C_{25}H_{42}O_{12} + H]^+ = 535.275$	1.4	1.4
193-2	Glucoside de norisoprénoïde	$[C_{19}H_{32}O_7 + H]^+ = 373.222$ $[C_{19}H_{32}O_7 + NH_4]^+ = 390.2486$	2.2	2.6
193-3	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_8 + H]^+ = 387.201$	3.7	3.7
193-4	Apiosyl-glucoside de norisoprénoïde	$[C_{19}H_{32}O_7 + H]^+ = 373.222$ $[C_{24}H_{40}O_{11} + H]^+ = 505.264$	4.3	4.5
193-5	Glucoside de norisoprénoïde	$[C_{19}H_{32}O_7 + H]^+ = 373.222$ $[C_{19}H_{32}O_7 + NH_4]^+ = 390.2486$	4.9	5.0
193-6	Norisoprénoïde libre	$M_{193}T_{382} [ + H]^+$	6.5	6.4
193-7	Norisoprénoïde libre	$M_{193}T_{439} [ + H]^+$	7.2	7.3
207-1	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_{10} + NH_4]^+$	1.4	1.4
207-2	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_8 + H]^+ = 387.201$	2.3	2.5
207-3	Malonyl-glucoside de norisoprénoïde	$[C_{22}H_{32}O_{11} + H]^+ = 473.201$ $[C_{22}H_{32}O_{11} + NH_4]^+ = 490.228$	3.5	3.8
209-1	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_7 + H]^+ = 371.206$	1.7	1.7
209-2	Glucoside de norisoprénoïde	$[C_{19}H_{32}O_8 + H]^+ = 389.217$	2.3	2.5
209-3	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_7 + H]^+ = 371.206$	3.0	3.1
209-4	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_7 + H]^+ = 371.207$	3.9	4.1
209-5	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_7 + H]^+ = 371.207$	4.4	4.6
209-6	Glucoside de norisoprénoïde	$[C_{19}H_{30}O_7 + NH_4]^+ = 388.233$	5.1	5.3
209-7	Malonyl-glucoside de norisoprénoïde	$[C_{22}H_{32}O_{10} + NH_4]^+ = 474.234$ $[C_{22}H_{32}O_{10} + H]^+ = 457.207$	6.6	6.7

**Tableau 33 : Norisoprénoïdes identifiés à partir de fragments caractéristiques de norisoprénoïdes dans les extraits de baies analysés par UHPLC-MS ESI (+) en 2014 et 2015.**

## **Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées**

Pour aller plus loin, une étude de la ségrégation de la quantification des différents fragments de norisoprénoïdes détectés dans les extraits de baies de 2014 et 2015 a permis d'identifier les norisoprénoïdes les plus abondants et les plus intéressants à étudier (Tableau 34). Ainsi, le fragment 207-2 est le fragment de norisoprénoïde le plus intensément détecté dans les baies à la fois en 2014 et 2015 et est détecté à la fois dans les baies de Ri et de Gw. Le fragment 193-7 est l'unique fragment de norisoprénoïde ayant des aires de pics significativement plus élevées dans le Gw que dans le Ri. Les fragments 209-2, 209-3 et 209-7 sont significativement plus intenses dans les échantillons de baies de Ri que de Gw. De ce fait, les molécules liées à ces 3 fragments sont de potentiels candidats comme précurseurs du TDN.

**Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées**

Fragment	Aonée	Moyenne Ri	Moyenne Gw	Minimum RixGw	Moyenne RixGw	Maximum RixGw	h <sup>2</sup>
193-1	2014	2.41E+05	2.39E+05	1.15E+05	3.26E+05	8.81E+05	0.86
	2015	3.41E+05	3.22E+05	1.40E+03	5.40E+05	1.44E+06	0.77
193-2	2014	2.76E+06	9.86E+05	4.67E+05	1.94E+06	6.27E+06	0.12
	2015	3.02E+06 **	9.29E+05	4.05E+03	4.54E+06	1.14E+07	0.78
193-3	2014	4.03E+05 **	2.35E+05	1.58E+05	3.76E+05	9.97E+05	0.76
	2015	4.50E+05	3.33E+05	4.39E+03	6.87E+05	9.37E+06	0.9
193-4	2014	3.90E+05 *	2.21E+05	1.61E+05	3.68E+05	9.38E+05	0.63
	2015	7.57E+05	5.36E+05	2.10E+02	1.04E+06	2.92E+06	0.65
193-5	2014	1.07E+06 *	7.74E+05	2.56E+05	1.12E+06	2.42E+06	0.76
	2015	1.34E+06	1.19E+06	4.13E+03	2.30E+06	5.76E+06	0.74
193-6	2014	2.31E+05 **	1.38E+05	6.96E+04	2.29E+05	6.45E+05	0.84
	2015	2.87E+05	3.22E+05	1.19E+03	4.97E+05	2.85E+06	0.76
<b>193-7</b>	<b>2014</b>	<b>7.06E+04</b>	<b>1.35E+05 **</b>	<b>5.84E+04</b>	<b>1.36E+05</b>	<b>4.33E+05</b>	<b>0.86</b>
	<b>2015</b>	<b>1.00E+05</b>	<b>2.28E+05 ***</b>	<b>6.75E+03</b>	<b>2.42E+05</b>	<b>3.95E+06</b>	<b>0.99</b>
207-1	2014	4.30E+05 **	2.94E+05	1.75E+05	4.70E+05	1.44E+06	0.81
	2015	6.50E+05	5.61E+05	6.94E+03	1.02E+06	2.64E+06	0.76
207-2	2014	1.52E+07	1.30E+07	6.23E+06	1.64E+07	3.28E+07	0.32
	2015	2.40E+07	2.36E+07	2.42E+03	3.37E+07	7.58E+07	0.29
207-3	2014	8.99E+05 ***	3.77E+05	3.60E+05	9.85E+05	3.53E+06	0.89
	2015	2.66E+06	6.72E+05	2.40E+03	2.05E+06	7.91E+06	0.4
209-1	2014	1.43E+06	1.38E+06	8.33E+05	2.13E+06	4.90E+06	0.63
	2015	1.66E+06	1.65E+06	3.05E+02	2.61E+06	7.24E+06	0.77
<b>209-2</b>	2014	<b>3.79E+06 ***</b>	<b>1.18E+06</b>	<b>1.47E+06</b>	<b>3.06E+06</b>	<b>7.23E+06</b>	<b>0.86</b>
	2015	<b>4.51E+06 *</b>	<b>1.86E+06</b>	<b>7.29E+02</b>	<b>5.74E+06</b>	<b>1.66E+07</b>	<b>0.66</b>
<b>209-3</b>	2014	<b>2.46E+06 **</b>	<b>6.50E+05</b>	<b>5.39E+05</b>	<b>1.83E+06</b>	<b>4.38E+06</b>	<b>0.51</b>
	2015	<b>5.07E+06 ***</b>	<b>8.62E+05</b>	<b>5.12E+03</b>	<b>4.27E+06</b>	<b>1.06E+07</b>	<b>0.69</b>
209-4	2014	1.32E+06 **	6.68E+05	5.21E+05	1.17E+06	2.87E+06	0.77
	2015	3.41E+06	2.01E+06	1.25E+03	3.31E+06	1.09E+07	0.49
209-5	2014	3.10E+06	1.91E+06	8.60E+05	2.57E+06	5.66E+06	0.28
	2015	6.65E+06 ***	3.67E+06	1.31E+03	5.80E+06	1.76E+07	0.89
209-6	2014	3.64E+06 ***	1.79E+06	1.04E+06	2.44E+06	5.63E+06	0.76
	2015	2.44E+06	1.64E+06	5.50E+02	3.50E+06	1.25E+07	0.74
<b>209-7</b>	<b>2014</b>	<b>9.56E+05 ***</b>	<b>2.43E+05</b>	<b>1.15E+05</b>	<b>3.80E+05</b>	<b>1.20E+06</b>	<b>0.69</b>
	<b>2015</b>	<b>1.09E+06 ***</b>	<b>4.28E+05</b>	<b>2.32E+03</b>	<b>9.41E+05</b>	<b>3.10E+06</b>	<b>0.86</b>

**Tableau 34 : Récapitulatif quantification des fragments de norisoprénoides.** Les fragments ayant des aires de pics significativement différentes entre le Ri et le Gw sur les deux années sont en gras.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### b. Analyse génétique de l'accumulation des norisoprénoïdes dans les baies

A partir des quantifications des fragments de norisoprénoïdes décrits précédemment, une recherche QTL a été réalisée avec les 3 cartes génétiques (Tableau 35). Seuls les QTL reproductibles sur les deux années d'analyses sont présentés ici. Des QTL reproductibles ont été trouvés sur les chromosomes 1, 5, 11, 12 et 15. Nous pouvons noter que les chromosomes 5 et 12 font partie des hotspots QTL qui ont été identifiés précédemment avec l'analyse non ciblée.

Chr	Carte génétique	Position génétique maximale	Fragment norisoprénoïde	LOD	pval
1	SNP_Ri	28.82	193-5	4.1-8.5	0-0.01
		27.94-38	209-5	4.5-6.8	0
5	Gw	10.81-14	193-6	5.1-7.4	0
	SSR	6.5-12	193-6	5.2-8	0-0.01
11	SNP_Ri	18.99-25.18	209-5	3.2-3.4	0.01-0.04
12	SNP_Ri	19.01	207-3	8-8.1	0
			209-7	12.5-17.5	0
	SSR	12.77-21.95	207-3	8.7-8.8	0
			209-7	14.8-18.3	0
15	SNP_Gw	40-40.38	209-1	3.9-4.6	0-0.01
		40.38	209-3	6.3-7	0
	SSR	48.61	209-1	4.2-4.3	0.03
		48.61	209-3	5.5-6.8	0

**Tableau 35 : QTL détectés pour les fragments de norisoprénoïdes détectés dans les analyses par UHPLC-MS ESI (+) reproductibles sur les deux années d'analyses (pval ≤ 0.05).**

A partir de ces données de quantification et d'analyse génétique, je me suis concentrée sur 2 fragments de norisoprénoïdes : le fragment 207-2 du fait de sa forte abondance dans les analyses de baies de Ri et de Gw et le fragment 207-3 du fait de la présence d'un QTL avec la carte SNP\_Ri sur le chromosome 12.

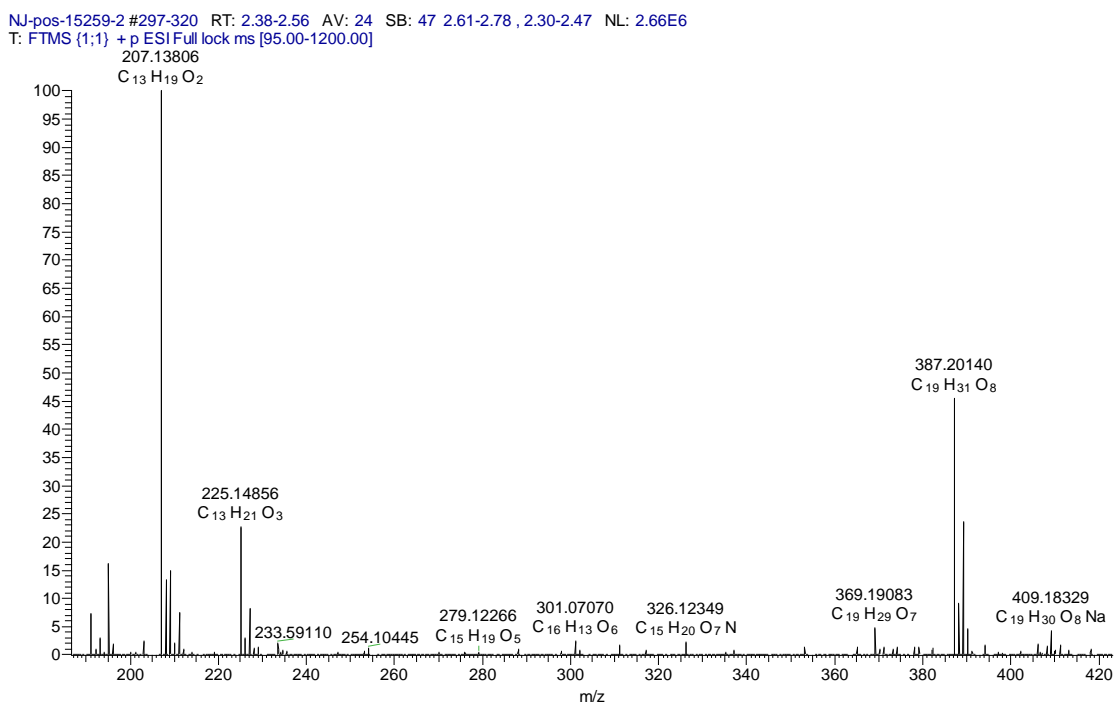
### c. Analyse détaillée du norisoprénoïde lié au fragment 207-2

La recherche de norisoprénoïde grâce à un fragment spécifique a révélé que le fragment 207-2 est le fragment de norisoprénoïde le plus intensément retrouvé dans les extraits de baies de Ri, de Gw et de la population RixGw, c'est pourquoi une analyse plus détaillée du norisoprénoïde lié au fragment 207-2 a été réalisée.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### ❖ Proposition de structure

Le spectre de masse dans l'échantillon le plus intense au temps de rétention 2.4 min est représenté dans la Figure 42. De fragments caractéristiques sont présents, avec notamment des ions avec des  $m/z$  189.128, 207.138 et 225.149 qui sont caractéristiques des norisoprénoïdes. Les ions pseudo-moléculaires confirmant la présence d'un glucose sur l'aglycone de norisoprénoïde sont les ions avec les  $m/z$  387.201, 404.228 et 409.183.



**Figure 42 :** Spectre de masse au temps de rétention 2.3 min dans l'échantillon du génotype 218E en 2015 (échantillon ayant l'aire de pic maximale de la population pour M207T147).

Cette observation est confirmée par les corrélations entre les données de quantification de ces fragments dans toutes les analyses de baies réalisées par UHPLC-MS (Tableau 36). La quantification des ions détectés au temps de rétention 2.3 min sont corrélées avec un R supérieure à 0.8 et 0.7 pour les ions avec un  $m/z$  de 387.201 et 369.191.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

ion	m/z	Ion pseudo-moléculaire	Corrélation avec M207T147
M207T147	207.138	[C <sub>13</sub> H <sub>20</sub> O <sub>3</sub> - H <sub>2</sub> O + H] <sup>+</sup>	1
M387T148	387.201	[C <sub>19</sub> H <sub>30</sub> O <sub>8</sub> + H] <sup>+</sup>	0.81
M369T148	369.191	[C <sub>19</sub> H <sub>28</sub> O <sub>7</sub> - H <sub>2</sub> O + H] <sup>+</sup>	0.78
M225T147	225.149	[C <sub>13</sub> H <sub>20</sub> O <sub>3</sub> + H] <sup>+</sup>	0.71
M404T137	404.228	[C <sub>19</sub> H <sub>30</sub> O <sub>8</sub> + NH <sub>4</sub> ] <sup>+</sup>	0.52
M189T149	189.128	[C <sub>13</sub> H <sub>20</sub> O <sub>3</sub> - 2 H <sub>2</sub> O + H] <sup>+</sup>	0.25
M409T138	409.183	[C <sub>19</sub> H <sub>30</sub> O <sub>8</sub> + Na] <sup>+</sup>	0.23

Tableau 36 : Corrélations entre les abondances de M207T147 en 2015 avec les ions identifiés comme appartenant au norisoprénoïde à identifier.

Une proposition de structure serait le glucoside de vomifoliol, un glucoside de norisoprénoïde qui a déjà été caractérisé dans les feuilles de Riesling (Skouroumounis and Winterhalter, 1994) mais aussi dans les vins de Gw (Baltenweck-Guyot et al., 1996).

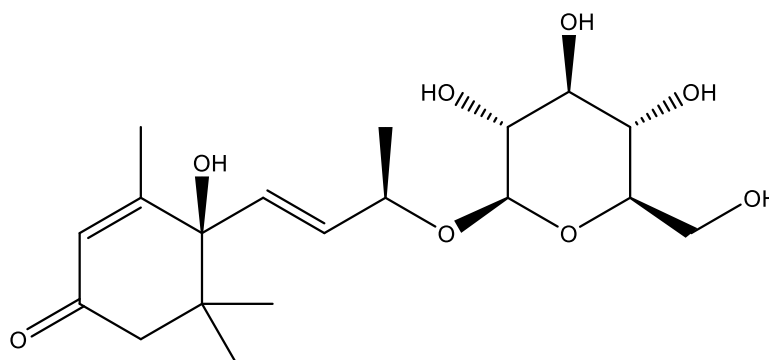


Figure 43 : Proposition de structure de la molécule liée au fragment de norisoprénoïde 207-2, le glucoside de vomifoliol.

### ❖ Quantification

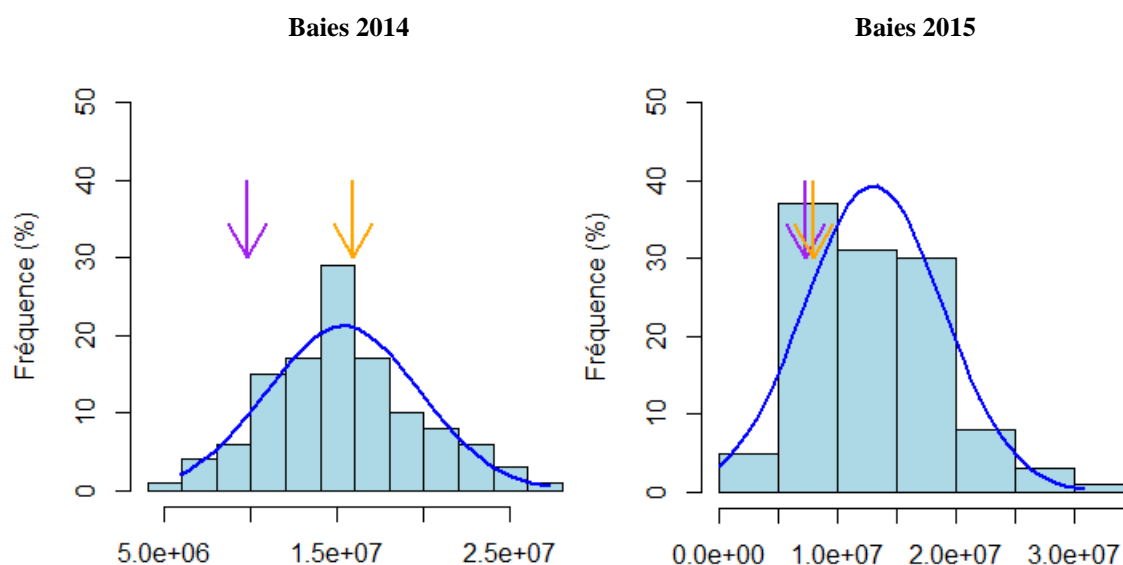
La quantification de l'ion 387.201 relié au fragment 207-2 ne relève pas de différence significative reproductible sur les deux années de prélèvement de baies (Tableau 37)

Année	Ion	Moyenne Ri	Moyenne Gw	Minimum RixGw	Moyenne RixGw	Maximum RixGw	h <sup>2</sup>
2014	M387T140	1.59E+07 *	9.80E+06	5.85E+06	1.53E+07	2.73E+07	0.58
2015	M387T148	8.06E+06	7.37E+06	0	1.31E+07	3.08E+07	0.64

Tableau 37 : Récapitulatif de la quantification de l'ion de m/z = 387.201 au temps de rétention 2.3 min sur les deux années d'analyses de baies par UHPLC-MS ESI (+).

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

L'observation de la ségrégation dans la population de la distribution de la molécule (Figure 44) permet de voir que c'est une distribution normale centrée autour de l'aire de pic à  $1.5 \times 10^7$ .



**Figure 44 :** Histogrammes de la distribution de l'aire de pic de l'ion avec un m/z 387.201 au temps de rétention 2.3 min détecté dans les baies de la descendance RixGw par UHPLC-MS ESI (+). Les flèches en orange indiquent les moyennes observées dans le Ri, celles en violet indique celles observées dans le Gw.

### 2.2.2.1.1. Analyse génétique

Un QTL a été détecté sur le chromosome 1 en 2014 pour ce norisoprénoïde mais n'a pas été retrouvé à nouveau dans les analyses des baies en 2015 (Tableau 38). Nous pouvons cependant noter qu'un QTL sur le chromosome 1 avait été détecté en 2014 pour la molécule 3-Oxo-7,8-dihydro- $\alpha$ -ionol en GC-MS dans la fraction liée dont l'intervalle de confiance était situé 21.6 et 45.8 cM avec un LOD de 4.9 (**Annexe 1**). Ce QTL a aussi détecté pour les fragments de norisoprénoïdes 193-5 et 209-5 sur les deux années d'analyse (Tableau 35).

Année	Carte	Ion	LOD 5%	LOD 20%	Marqueur	Chr	Pos	LOD	pval
2014	SNP_Ri	M387T140	2.95	2.33	chr1b10849423	1	56.05	4.35	0

**Tableau 38 :** QTL détectés pour l'ion 387.201 au temps de rétention 2.3 min lié au fragment de norisoprénoïde 207-2 (pval  $\leq 0.05$ )



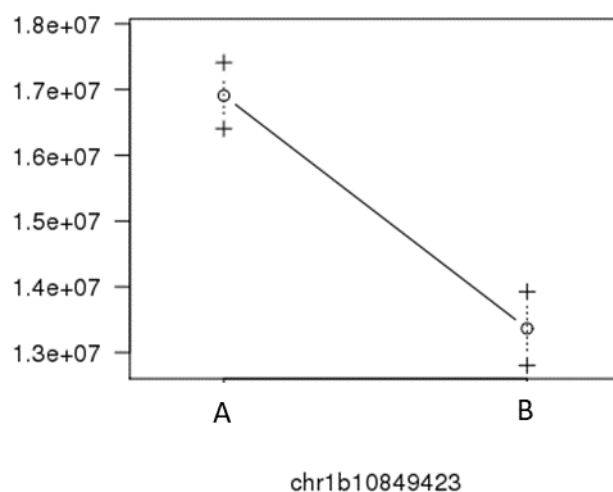
## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Un modèle lié au marqueur génétique sur le chromosome 1 à la position 56 cM expliquant 16% de la variance a été trouvé (Tableau 39).

Source de variation	Degrés de liberté	Somme des écarts des carrés	Carrés moyens	LOD	% variance expliquée	pval ( $\chi^2$ )	pval (F)
1@56.05	1	3.57e+14	3.57e+14	4.38	15.82	7.12e-06 ***	8.91e-06 ***
Error	115	1.90e+15	1.65e+13				
Total	115	2.26e+15					

**Tableau 39** : Modèle expliquant le contenu en pour l'ion 387.201 au temps de rétention 2.3 min lié au fragment de norisoprénoïde 207-2 avec la carte génétique SNP\_Ri sur le chromosome 1 à la position 56cM.

Les individus ayant l'allèle A du Ri au locus chr1b20849423 ont des aires de pics significativement plus élevées que ceux ayant l'allèle B (Figure 45).



**Figure 45** : Effet du QTL 1@56 de la carte génétique SNP\_Ri sur l'aire de pic de l'ion avec un m/z de 387.201 au temps de rétention 2.3 min lié au fragment de norisoprénoïde 207-2 en 2014 (A) : allèle A du Ri, (B) : allèle B du Ri.

### ❖ Conclusion sur le glucoside de vomifoliol

Le glucoside de vomifoliol est le norisoprénoïde le plus abondamment détecté dans les extraits de baies de Ri, de Gw et de la population RixGw. Ce composé a déjà été caractérisé dans les feuilles de Riesling (Skouroumounis and Winterhalter, 1994) mais aussi dans les vins de Gw (Baltenweck-Guyot et al., 1996). Son contenu n'est pas différentiel entre les baies de Ri et de Gw et aucun QTL reproductible présent sur les deux années d'analyse n'a été identifié pour cette molécule. Malgré le fait que ce composé n'est pas différentiel entre le Ri et le Gw, son étude reste intéressante du fait de sa forte teneur dans les baies. Lors de la vinification, il

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

pourrait subir une hydrolyse pour donner du vomifoliol, molécule présente dans de nombreux fruits exotiques tels que la mangue (Sakho et al., 1997), la nectarine (Aubert et al., 2003) ou encore le fruit de la passion (Chassagne et al., 1999) et ayant des arômes de type fruité (FoodDB).

- Analyse détaillée du norisoprénoïde lié au fragment 209-7

Parmi les fragments de norisoprénoïdes identifiés, le fragment 209-7 est un fragment qui paraît intéressant à étudier car son aire de pic est significativement plus élevée dans le Ri que dans le Gw (Tableau 34) et qu'un QTL pour ce fragment a été mis en évidence sur le chromosome 12 avec les cartes génétiques SSR et SNP\_Ri.

- Proposition de structure

L'analyse du spectre de masse au temps de rétention 6.7 min de l'échantillon 0002E montre une co-élution de plusieurs molécules ionisées au même temps de rétention (

Figure 46), avec un fragment spécifique de monoterpénol avec un m/z de 135.117 correspondant à l'ion  $[C_{10}H_{16}O - H_2O + H^+]$  par exemple. L'ion de m/z 457.107 étant le plus intense.

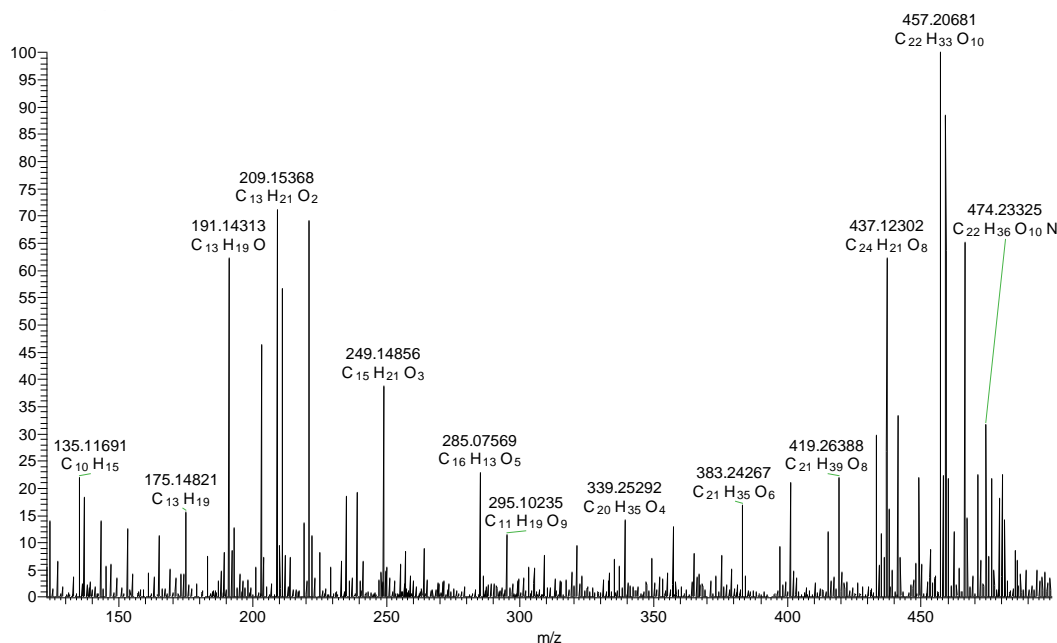


Figure 46 : Spectre de masse au temps de rétention 6.7 min, correspondant à la molécule liée au fragment 209-7.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

L'analyse des corrélations par rapport au fragment de m/z 209.154 au temps de rétention permet d'identifier les fragments correspondants au norisoprénoïde à identifier (Tableau 40). Ainsi le fragment 209-7 semble lié à des ions correspondants à des adduits de la formule brute C<sub>22</sub>H<sub>32</sub>O<sub>10</sub>.

ion	m/z	Ion pseudo-moléculaire	Corrélation par rapport au M209T399
<b>M209T399</b>	209.154	[C <sub>13</sub> H <sub>20</sub> O <sub>2</sub> + H] <sup>+</sup>	1
<b>M457T397</b>	457.207	[C <sub>22</sub> H <sub>32</sub> O <sub>10</sub> + H] <sup>+</sup>	0.77
<b>M474T400</b>	474.233	[C <sub>22</sub> H <sub>32</sub> O <sub>10</sub> + NH <sub>4</sub> ] <sup>+</sup>	0.77
<b>M479T395</b>	479.188	[C <sub>22</sub> H <sub>32</sub> O <sub>10</sub> + Na] <sup>+</sup>	0.69
<b>M439T389</b>	439.1963	[C <sub>22</sub> H <sub>32</sub> O <sub>10</sub> - H <sub>2</sub> O + H] <sup>+</sup>	0.66
<b>M191T393</b>	191.143	[C <sub>13</sub> H <sub>20</sub> O <sub>2</sub> - H <sub>2</sub> O + H] <sup>+</sup>	0.56

Tableau 40 : Matrice de corrélation de M209T399 en 2015 avec les ions identifiés comme appartenant à la molécule liée au fragment 209-7.

La molécule est un norisoprénoïde lié à un malonyl-glucoside. Une structure possible de cette molécule est le malonyl-glucoside de 3-oxo- $\alpha$ -ionol (Figure 47).

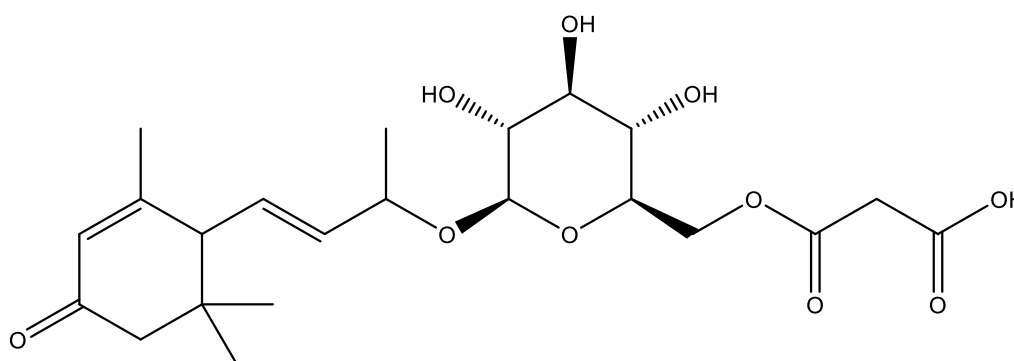


Figure 47 : Proposition de structure de la molécule liée au fragment 209-7, le malonyl-glucoside de 3-oxo- $\alpha$ -ionol.

### o Quantification

L'ion pseudo-moléculaire ayant un m/z de 457.207 a été utilisé pour le suivi de la quantification de la molécule (Tableau 41). Les analyses révèlent que les extraits de baies de Riesling ont des aires de pics significativement plus élevées que ceux de Gewurztraminer pour l'ion pseudo-moléculaire de m/z 457.207 au temps de rétention 6.6 min. On observe des aires de pics 5 fois plus importantes dans les analyses de Ri que dans les analyses de Gw. En 2015, le maximum d'aire de pic observé dans la population est 2 fois supérieur à celui observé en 2014, révélant un effet millésime en 2015 sur ce norisoprénoïde.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Année	Ion	Moyenne Ri	Moyenne Gw	Minimum RixGw	Moyenne RixGw	Maximum RixGw	h <sup>2</sup>
2014	M457T391	1.41E+06 ***	2.47E+05	6.03E+04	5.51E+05	1.86E+06	0.69
2015	M457T397	1.50E+06 **	2.79E+05	4.82E+03	1.02E+06	3.86E+06	0.67

Tableau 41 : Récapitulatif de la quantification de l'ion de m/z 457.207 au temps de rétention 6.6 min. Les différences significatives d'aire de pic entre Ri et Gw sont mises en évidence par \*\* pval ≤ 0.01, \*\*\* pval ≤ 0.001.

La distribution du contenu en ce norisoprénoïde n'est pas de type normal, 70% des individus de la population RixGw ont des aires de pics inférieurs à  $1.10^6$ .

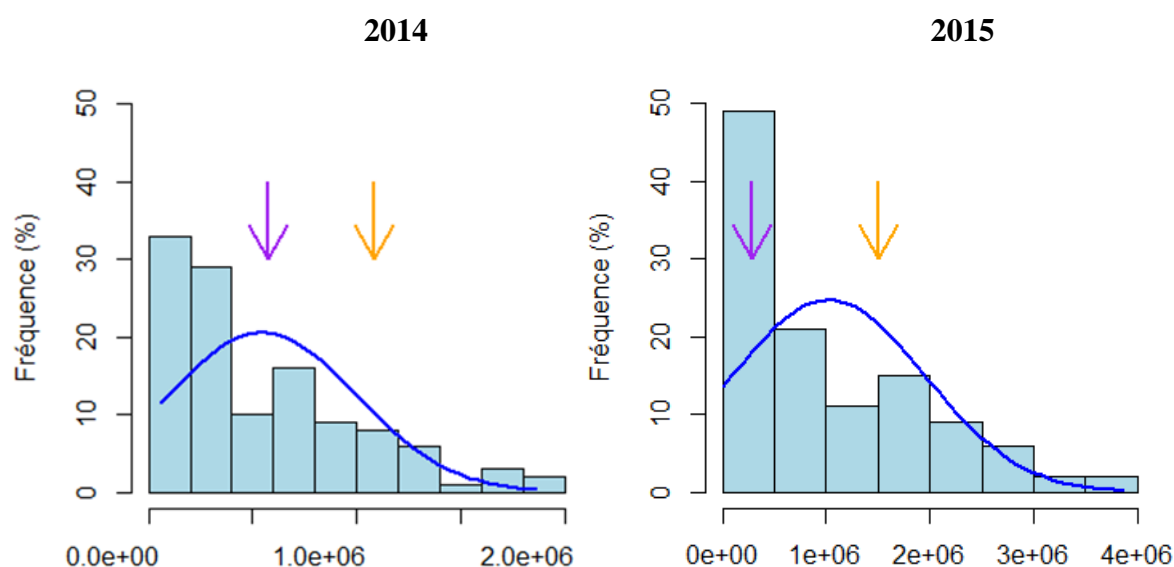


Figure 48 : Histogrammes de la distribution de l'aire de pic de l'ion 457.207 au temps de rétention 6.7 minutes détecté dans les baies de la descendance RixGw par UHPLC-MS ESI (+). La flèche en orange indique la moyenne observée dans les analyses de Ri, celle en violet indique celle observée dans le Gw.

### ○ Analyse génétique

La recherche de QTL pour le malonyl-glucoside de norisoprénoïde au temps de rétention 6.6 min dans les baies de la population RixGw a été réalisé avec les 3 cartes génétiques. Elle révèle un QTL sur le chromosome 12 avec les cartes génétiques SSR et SNP\_Ri, QTL qui est reproductible sur les deux années d'analyses (Tableau 42).

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Année	Carte génétique	Ion	LOD 5%	LOD 20%	Marqueur	Chr	Pos (cM)	LOD	pval
2014	SNP_Ri	M457T391	3.02	2.36	chr12b8358979	12	19.01	17.94	0
2014	SSR	M457T391	4.02	3.18	VMC1g3.2	12	13.3	16.93	0
2015	SNP_Ri	M457T397	3.08	2.37	chr12b8358979	12	19.01	18.1	0
2015	SSR	M457T397	4.03	3.29	c12.loc8	12	8	17.9	0

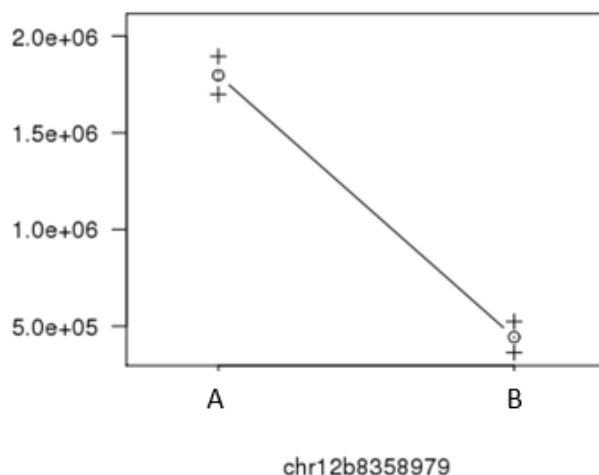
**Tableau 42 :** QTL détectés pour l'ion 457.207 au temps de rétention 6.6 min lié au fragment de norisoprénoïde 209-7 (pval  $\leq$  0.05).

La molécule ayant un QTL à la fois avec les cartes génétiques SNP\_Ri et SSR sur le chromosome 12, je n'ai fait un modèle qu'avec la carte génétique SNP\_Ri car le LOD score y est plus élevé. Un modèle expliquant 51.9 % de la variance du contenu en ce malonyl-glucoside a pu être fait à partir des analyses QTL avec la carte SNP\_RI (Tableau 43).

Source de variation	Degrés de liberté	Somme des carrés des écarts	Carrés moyens	LOD	% variance expliquée	pval ( $\chi^2$ )	pval (F)
12@19.9	1	5.02E+13	5.02e+13	18.1	51.87	0 ***	0 ***
Erreur	112	4.65e+11					
Totale	113	9.67e+13					

**Tableau 43 :** Modèle expliquant le contenu en malonyl-glucoside de norisoprénoïde en 2015 avec la carte génétique SNP\_Ri.

L'analyse de l'effet des allèles du Ri sur le contenu en malonyl-glucoside de norisoprénoïde dans la population RixGw montre que les individus ayant l'allèle A au marqueur chr12b8358979 en ont 4 fois plus que ceux qui ont l'allèle B (Figure 40).



**Figure 49 :** Effet du QTL 12@19.9 de la carte génétique SNP\_Ri sur l'aire de pic de la molécule de malonyl-glucoside de norisoprénoïde en 2015. (A) : allèle A du Ri, (B) : allèle B du Ri.

❖ Conclusion sur le malonyl-glucoside d'oxo- $\alpha$ -ionol

Le malonyl-glucoside d'oxo- $\alpha$ -ionol est présent de manière différentielle dans les baies de Ri que dans les baies de Gw. L'analyse génétique du contenu en malonyl-glucoside d'oxo- $\alpha$ -ionol révèle un déterminisme génétique lié à un locus génétique du Ri sur le chromosome 12. Ce composé représente donc un candidat comme précurseur de la molécule de TDN dans les vins de Ri mais pourrait aussi être un précurseur d'une molécule intéressante : le 3-oxo- $\alpha$ -ionol, molécule ayant des arômes de type épice (TheGoodScentCompagny).

**2.3. Apport de l'analyse métabolomique des feuilles dans la compréhension du métabolome de la baie**

Après avoir étudié le métabolisme secondaire des baies et leur déterminisme génétique dans la population RixGw, une étude du métabolome des feuilles a été réalisée sur des feuilles prélevées en 2015 afin de le comparer avec celui des baies prélevées la même année.

*2.3.1. Analyse non ciblée de la diversité du métabolome des baies et des feuilles*

Afin d'avoir une vision globale de la diversité du métabolisme secondaire détecté dans les baies et les feuilles, une première analyse non ciblée a été réalisée. Des milliers d'ions avec des m/z et des temps de rétentions différents ont pu être détectés. Leur nombre est quatre fois plus élevé dans les baies que dans les feuilles (Tableau 44).

UHPLC-MS	Nombre d'ions détectés dans les extraits de baies 2015	Nombre d'ions détectés dans les extraits de feuilles 2015
ESI (+)	8152	1396
ESI (-)	4108	819

Tableau 44 : Nombre d'ions retenus par XCMS dans les échantillons de baies et de feuilles prélevés en 2015 analysés par UHPLC-MS.

Une comparaison des ions détectés par UHPLC-MS ESI (+) dans les baies et les feuilles prélevées en 2015 de la population d'étude a été réalisé afin de voir quelles molécules sont détectables dans les deux types d'échantillons (Figure 50).

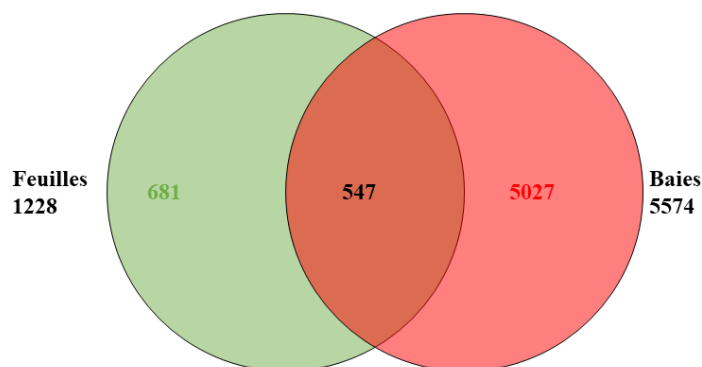


Figure 50 : Diagramme de Venn représentant le nombre d'ions détectés par UHPLC-MS ESI (+) dans les feuilles et dans les baies prélevées en 2015.

Le contenu en ions détectés dans les baies et dans les feuilles est très différent. Parmi les ions détectés à la fois dans les feuilles et les baies prélevées la même année, seul 547 ions sont retrouvés sur les 1228 ions des feuilles et sur les 5574 ions des baies. Ils représentent moins de la moitié des ions détectés dans les feuilles. Cette différence peut être expliquée par le type d'échantillon mais aussi par une extraction qui est différente entre les deux types d'échantillons : les baies subissent une extraction par SPE afin d'éliminer les composés les plus polaires tels que les sucres et les acides qui, par leur trop forte présence, gênerait les analyses. Cette étape élimine les composés polaires tels que les acides féruliques, cinnamiques et coumariques. Cette étape n'est pas nécessaire pour des échantillons de feuilles, ce qui fait que l'on peut détecter des molécules qui n'ont pas pu être détectées dans les extraits de baies. Ceci est observable facilement lorsque l'on se concentre sur les 20 ions les plus intenses dans les extraits de feuilles (Tableau 45). Parmi ces molécules, cinq ne sont pas présentes du tout dans les analyses réalisées sur les baies (acides gras et sucres), d'autres sont détectées mais ne sont pas présentes dans le top 20 des molécules détectées le plus dans les baies (acides aminés, acide chlorogénique). Parmi les molécules présentes dans le top 20 des molécules détectées dans les baies et les feuilles, on retrouve les molécules de type flavonoïdes et une molécule de type monoterpénol.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Ion	RT (min)	m/z	ion	Standard / Proposition de structure	Famille	Présent dans les baies ?	Moyenne Ri	Moyenne Gw	Min RixGw	Moyenne RixGw	Max RixGw	h <sup>2</sup>
M479T259	4.3	479.082	[C21H18O13 + H+]	<b>Quercétine - Glucuronide</b>	Flavonoïde	OUI	2.81E+08	2.89E+08	1.59E+08	2.55E+08	3.66E+08	0.34
M147T67_1	1.1	147.076	[C5H10N2O3 + H+]	<b>Glutamine</b>	Acide aminé	OUI	2.77E+07	2.92E+07	1.91E+06	3.97E+07	1.55E+08	0.06
M465T235	3.9	465.103	[C21H20O12 + H+]	Quercétine glucoside	Flavonoïde	OUI	2.51E+07	1.71E+07	8.94E+06	2.88E+07	8.45E+07	0.69
M148T67_1	1.1	148.06	[C5H9NO4 + H+]	<b>Acide glutamique</b>	Acide aminé	OUI	4.39E+07	4.13E+07	1.70E+07	4.27E+07	7.58E+07	0.33
M611T202	3.4	611.161	[C27H30O16 + H+]	<b>Quercétine rutinoside</b>	Flavonoïde	OUI	2.45E+07 *	1.16E+07	5.53E+06	2.25E+07	7.20E+07	0.66
M463T306	5.1	463.087	[C21H18O12 + H+]	<b>Kaempférol glucuronide</b>	Flavonoïde	OUI	1.83E+07	2.73E+07	5.74E+06	2.06E+07	5.91E+07	0.38
M303T259	4.3	303.05	[C15H10O7 + H+]	<b>Fragment de Quercétine Glucoronide</b>	Flavonoïde	OUI	4.09E+07	4.22E+07	2.54E+07	3.78E+07	5.42E+07	0.1
M116T69_2	1.2	116.071	[C5H9NO2 + H+]	<b>Proline</b>	Acide aminé	OUI	1.96E+07	1.55E+07	3.91E+06	1.96E+07	4.44E+07	0.05
M365T72	1.2	365.105	[C12H22O11 + Na+]	Dissaccharide	sucre	OUI	2.87E+07	2.51E+07	1.60E+07	3.13E+07	4.26E+07	0.2
M291T132	2.2	291.086	[C15H14O6 + H+]	<b>Catéchine</b>	Flavonoïde	OUI	1.01E+07	2.50E+07 **	5.84E+06	1.70E+07	4.17E+07	0.66
M163T106	1.8	163.039	[C9H6O3 + H+]	<b>Fragment de Caffeoyl quinic acid</b>	Acide chlorogénique	OUI	1.82E+07 ***	1.41E+06	4.73E+05	1.23E+07	4.00E+07	0.87
M355T105	1.8	355.102	[C16H18O9 + H+]	<b>Caffeoyl quinic acid</b>	Acide chlorogénique	OUI	1.54E+07 ***	2.18E+04	2.62E+04	9.80E+06	3.53E+07	0.89
M325T70	1.2	325.113	[C12H20O10 + H+]	Dissaccharide	sucre	OUI	2.40E+07	2.17E+07	1.20E+07	2.52E+07	3.47E+07	0.55
M490T67_1	1.1	490.177	[C17H28O15 + NH4+]	Dérivé sucre	sucre	NON	1.70E+07	1.63E+07	6.74E+06	1.67E+07	3.26E+07	0.27
M138T69	1.2	138.055	[C7H7NO2 + H+]	Acide Anthranilique	Phénol	NON	1.20E+07	1.49E+07	1.95E+06	1.39E+07	3.17E+07	0.04
M607T556	9.3	607.292	[C36H39O5N4 + H+]	Methylpheophorbide	Produit de dégradation de la chlorophylle	NON	8.27E+06	8.48E+06	5.90E+04	7.19E+06	3.12E+07	0.19
M613T565	9.4	613.483	[C39H64O5 + H+]	Diglycérade	Diglyceride	NON	9.25E+06	1.00E+07	1.64E+02	1.11E+07	3.06E+07	0.46
M360T71	1.2	360.15	[C12H22O11 + NH4+]	Dissaccharide	sucre	OUI	1.73E+07	1.54E+07	1.34E+07	1.91E+07	2.63E+07	0.49
M153T280	4.7	153.127	[C10H16O + H+]	Fragment de glucoside d'oxyde de monoterpénol	Monoterpénol	OUI	1.42E+06	6.47E+06	1.48E+05	3.01E+06	2.48E+07	0.46
M297T428	7.1	297.279	[C19H36O2 + H+]	Méthyl ester de l'acide oléique	Acide gras	NON	7.79E+06	5.37E+06	7.91E+04	6.93E+06	2.41E+07	0.48

**Tableau 45 : Top 20 des ions majeurs dans les feuilles détectés par UHPLC-MS ESI (+).**



## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### 2.3.2. QTL responsables de la variation du contenu en métabolites secondaires des feuilles

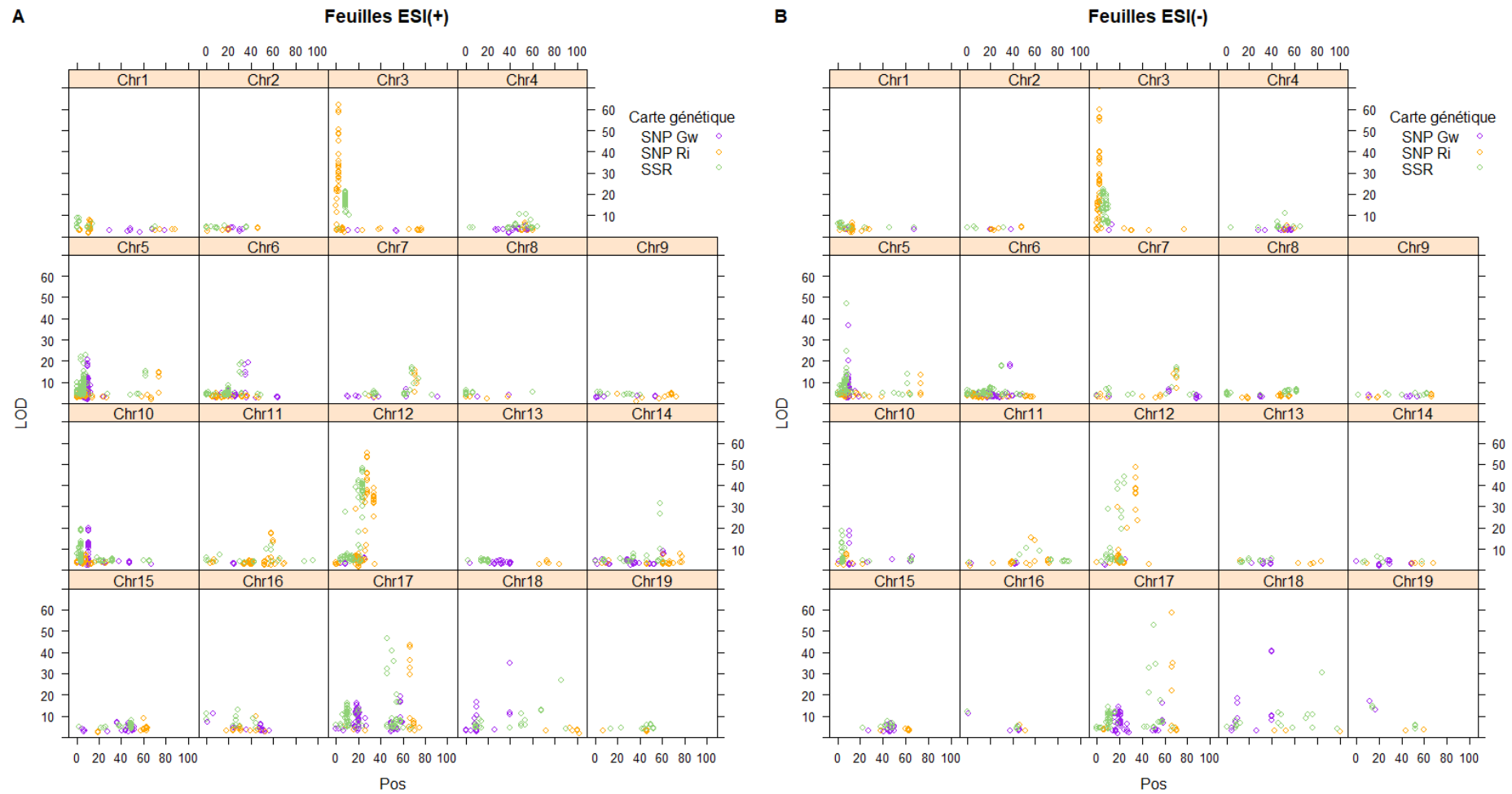
A partir des données non ciblées UHPLC-MS des extraits de feuilles, une recherche QTL a été réalisée avec les 3 cartes génétiques SSR, SNP\_Ri et SNP\_GW (Tableau 46) . Le nombre de QTL détectés dans les feuilles est bien inférieur à celui observé dans les baies prélevées la même année. Ceci est cohérent avec la différence en nombre d'ions détectés dans les feuilles par rapport à celui détecté dans les baies (Tableau 44) .

Carte génétique	UHPLC-MS Baies		UHPLC-MS Feuilles	
	ESI (-)	ESI (+)	ESI (-)	ESI (+)
<b>SSR</b>	3310	2620	489	706
<b>SNP_RI</b>	1441	1354	341	468
<b>SNP_GW</b>	2913	2116	360	560

Tableau 46 : Nombre de QTL par carte génétique, détectés dans les feuilles et dans les baies en 2015 (pval  $\leq 0.05$ ).

L'observation des LOD score et des locus des QTL détectés dans les feuilles permet d'identifier des hotspots génétiques reliés aux contenus en métabolites secondaires des feuilles (Figure 51). Des QTL sur les chromosomes 3 et 17 sont propres aux analyses d'échantillons de feuilles. Des QTLs communs aux baies et aux feuilles sont détectés sur les chromosomes 5, 6 et 12 (Figure 29 (B) et Figure 51). Cependant, les LOD score sont plus faibles dans les échantillons de feuilles ceux des baies.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées



**Figure 51 : QTL détectés avec les données UHPLC-MS de manière non ciblée dans les feuilles prélevées en 2015 ( $p_{val} \leq 0.05$ ). (A) ESI (+) (B) ESI (-). Chaque point correspond à un ion placé en fonction des caractéristiques du QTL associé : sa position en cM sur chaque chromosome en abscisse et le LOD du QTL en ordonnée.**

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

Les LOD score dans les analyses des feuilles sont globalement plus faibles par rapport à ceux retrouvés dans les analyses de baies. Ceci peut être expliqué par l'observation des aires de pics des ions ayant un QTL à la fois dans les baies et dans les feuilles (Figure 52). Plus les aires de pics sont élevées, plus les LOD scores sont élevés pour la plupart des QTL observés. Les aires de pics sont pour la plupart plus élevées dans les baies que dans les feuilles, même si la comparaison directe est délicate en raison de la nature différente des extraits.

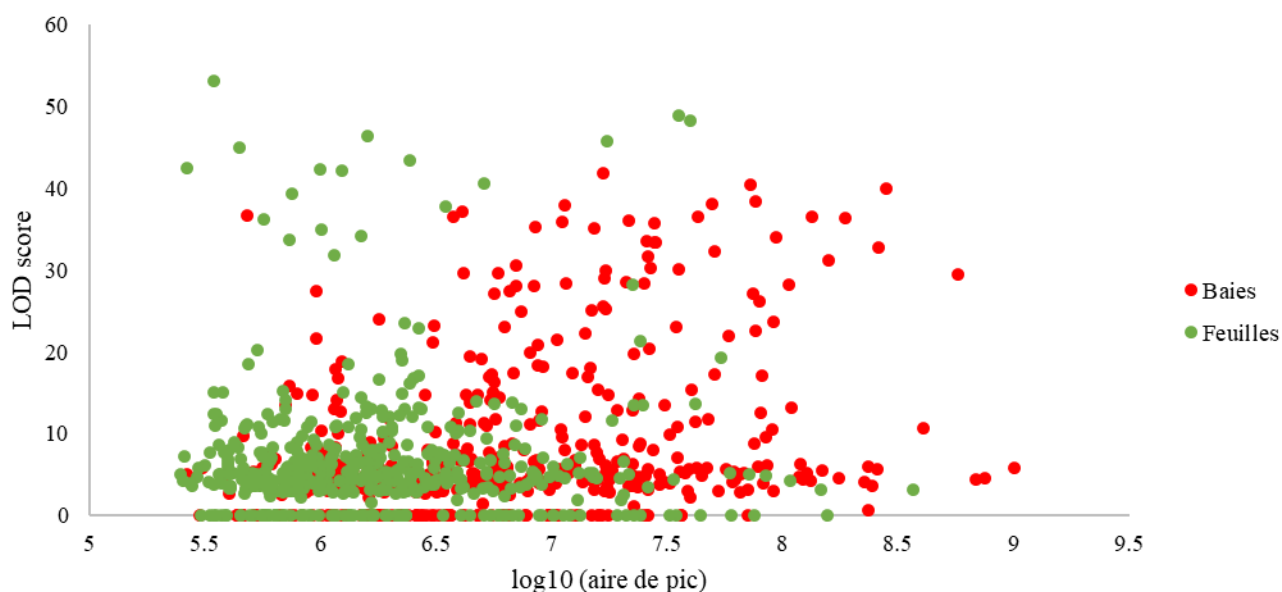
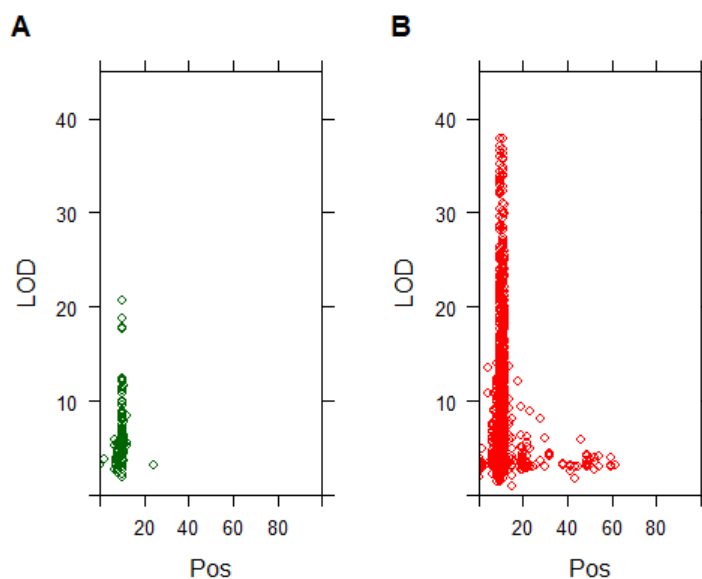


Figure 52 : LOD score des QTL des ions en fonction des aires de pics détectés par UHPLC-MS ESI (+) pour les ions ayant un m/z communs dans les baies et dans les feuilles 2015 ( $p_{val} \leq 0.05$ ).

### a. Analyse détaillée du « hotspot » situé sur le chromosome 5, carte SNP\_Gw

Le hotspot retrouvé avec les analyses UHPLC-MS de baies sur le chromosome 5 avec la carte génétique SNP\_GW est retrouvé avec les données du métabolome des feuilles. Le QTL est situé à 10 cM et colocalisent avec le gène de la DXS comme pour les baies. Avec les analyses par UHPLC-MS ESI (+), 330 et 2309 ions pour les feuilles et les baies respectivement ont un QTL sur le chromosome 5 avec des LOD maximums de 22.85 pour les analyses des feuilles et 37.87 pour les analyses de baies.

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées



**Figure 53 : QTL détectés sur le chromosome 5 avec la carte SNP\_Gw et les données non ciblées par UHPLC-MS ESI (+) de baies et de feuilles en 2015 ( $p_{\text{val}} \leq 0.05$ ). (A) : Feuilles 2015 , (B) : Baies 2015.**

Parmi ces ions, 43 ions ayant un QTL à la fois dans les feuilles et dans les baies ont pu être reliés à des ions pseudo-moléculaires correspondant à des isoprénoïdes : 36 m/z correspondent à des monoterpénols, 5 à des sesquiterpènes, 1 à un diterpène et 1 à un norisoprénoïde.

**Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées**

m/z	Ion pseudo-moléculaire	Famille	Max quanti m/z baies	Max quanti m/z feuilles	QTL baies		QTL feuilles	
					Nb de pic	Max LOD	Nb de pic	Max LOD
135.117	[C10H14 + H] <sup>+</sup>	Monoterpénol	7.96E+07	5.65E+06	5	22.75	4	6.01
151.112	[C10H14O + H] <sup>+</sup>	Monoterpénol	1.06E+07	1.53E+06	5	21.52	2	5.4
137.133	[C10H18O – H2O + H] <sup>+</sup>	Monoterpénol	3.55E+07	3.38E+06	5	27.84	2	5.85
153.127	[C10H18O2 – H2O + H] <sup>+</sup>	Monoterpénol	1.59E+08	2.48E+07	6	28.43	5	12.42
169.122	[C10H16O2 + H] <sup>+</sup>	Monoterpénol	1.72E+07	9.20E+05	6	28.69	3	4.65
171.138	[C10H18O2 + H] <sup>+</sup>	Monoterpénol	1.43E+07	1.26E+06	3	16.19	1	2.67
187.133	[C10H18O3 + H] <sup>+</sup>	Monoterpénol	8.15E+07	8.10E+05	6	15.53	1	5.26
269.175	[C15H24O4 + H] <sup>+</sup>	Sesquiterpène	7.04E+06	4.51E+05	4	26.08	1	4.75
285.17	[C15H24O5 + H] <sup>+</sup>	Sesquiterpène	1.53E+07	1.86E+06	6	33.57	1	8.7
303.18	[C15H26O6 + H] <sup>+</sup>	Sesquiterpène	8.49E+06	1.48E+06	2	33.24	1	8.17
297.17	[C16H24O5 + H] <sup>+</sup>	Monoterpénol	9.16E+06	6.74E+05	2	15.66	1	5.72
315.18	[C16H26O6 + H] <sup>+</sup>	Monoterpénol	2.11E+07	1.87E+06	2	28.5	2	11.48
331.175	[C16H26O7 + H] <sup>+</sup>	Monoterpénol	4.10E+06	2.39E+06	3	35.63	2	5.65
333.191	[C16H28O7 + H] <sup>+</sup>	Monoterpénol	2.79E+07	3.85E+06	3	32.29	3	9.48
350.217	[C16H28O7 + NH4] <sup>+</sup>	Monoterpénol	2.24E+07	6.91E+06	3	7.64	3	6.38
355.173	[C16H28O7 + Na] <sup>+</sup>	Monoterpénol	2.44E+07	2.28E+07	2	5.71	3	7.43
366.212	[C16H28O8 + NH4] <sup>+</sup>	Monoterpénol	8.72E+06	9.56E+05	5	16.89	1	5.33
371.168	[C16H28O8 + Na] <sup>+</sup>	Monoterpénol	6.28E+06	3.04E+06	1	5.44	2	6.74
373.183	[C16H30O8 + Na] <sup>+</sup>	Monoterpénol	1.40E+07	3.98E+06	1	10.92	1	9.52
367.196	[C16H30O9 + H] <sup>+</sup>	Monoterpénol	8.07E+07	1.59E+06	1	12.21	1	5.64
389.178	[C16H30O9 + Na] <sup>+</sup>	Monoterpénol	4.20E+07	3.12E+06	1	9.11	1	3.74
436.218	[C19H30O10 + NH4] <sup>+</sup>	Monoterpénol	8.44E+07	1.14E+06	3	5.96	1	7.12
441.173	[C19H30O10 + Na] <sup>+</sup>	Monoterpénol	3.24E+07	1.11E+06	3	6.12	1	6.62
390.249	[C19H32O7 + NH4] <sup>+</sup>	Norisoprénoïde	8.96E+05	7.27E+05	1	3.13	1	2.7
301.216	[C20H28O2 + H] <sup>+</sup>	Diterpène	2.39E+07	6.85E+05	4	12.9	1	11.54
480.244	[C21H34O11 + NH4] <sup>+</sup>	Sesquiterpène	4.99E+07	2.03E+06	3	36.01	1	6.14
485.199	[C21H34O11 + Na] <sup>+</sup>	Sesquiterpène	1.67E+07	1.32E+06	2	37.87	1	6.29
466.265	[C21H36O10 + NH4] <sup>+</sup>	Monoterpénol	1.09E+08	1.01E+07	3	12.43	1	4.78
471.22	[C21H36O10 + Na] <sup>+</sup>	Monoterpénol	2.58E+07	6.24E+06	3	30.86	1	4.96
465.233	[C21H36O11 + H] <sup>+</sup>	Monoterpénol	5.07E+07	3.68E+06	4	30.46	2	9.85
482.26	[C21H36O11 + NH4] <sup>+</sup>	Monoterpénol	2.81E+08	9.03E+06	3	36.22	2	10.76
487.551	[C21H36O11 + Na] <sup>+</sup>	Monoterpénol	7.42E+07	7.00E+06	4	24.26	2	9.5
503.21	[C21H36O12 + Na] <sup>+</sup>	Monoterpénol	5.37E+06	3.84E+05	3	15.04	1	5.84
489.231	[C21H38O11 + Na] <sup>+</sup>	Monoterpénol	8.04E+06	9.76E+05	1	16.21	1	3.05
494.26	[C22H36O11 + NH4] <sup>+</sup>	Monoterpénol	2.27E+07	1.37E+06	2	17	1	4.72
499.215	[C22H36O11 + Na] <sup>+</sup>	Monoterpénol	6.82E+06	1.42E+06	1	15.64	1	4.48
510.255	[C22H36O12 + NH4] <sup>+</sup>	Monoterpénol	3.08E+06	3.83E+06	2	20.55	1	5.89
515.21	[C22H36O12 + Na] <sup>+</sup>	Monoterpénol	1.23E+06	3.44E+06	2	16.66	1	5.9
480.28	[C22H38O10 + NH4] <sup>+</sup>	Monoterpénol	2.70E+07	2.04E+07	1	25.4	1	4.49
485.236	[C22H38O10 + Na] <sup>+</sup>	Monoterpénol	4.41E+06	9.13E+06	1	12.43	1	4.92
496.275	[C22H38O11 + NH4] <sup>+</sup>	Monoterpénol	2.52E+07	8.39E+05	2	25.03	1	5.23
498.291	[C22H40O11 + NH4] <sup>+</sup>	Monoterpénol	5.58E+06	8.45E+05	1	14.23	1	4.26
510.291	[C23H40O11 + NH4] <sup>+</sup>	Monoterpénol	1.06E+08	7.89E+05	3	26.42	1	6.66

**Tableau 47 : Liste des ions [M + H]<sup>+</sup> ayant un QTL sur le chr5 avec la carte SNP Gw (pval ≤ 0.05) à la fois dans les baies et dans les feuilles en 2015 et analysées par UHPLC-MS ESI (+).**

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

### b. Analyse détaillée du « hotspot » situé sur le chromosome 3, carte SNP\_Ri

Un QTL a été identifié sur le chromosome 3 à partir des données de génotypage SNP Ri et les données de phénotypage des feuilles prélevées en 2015. Ce QTL n'impacte pas les métabolites dans les baies de la population RixGw mais est détecté pour 53 ions différents dans les feuilles (Figure 54).

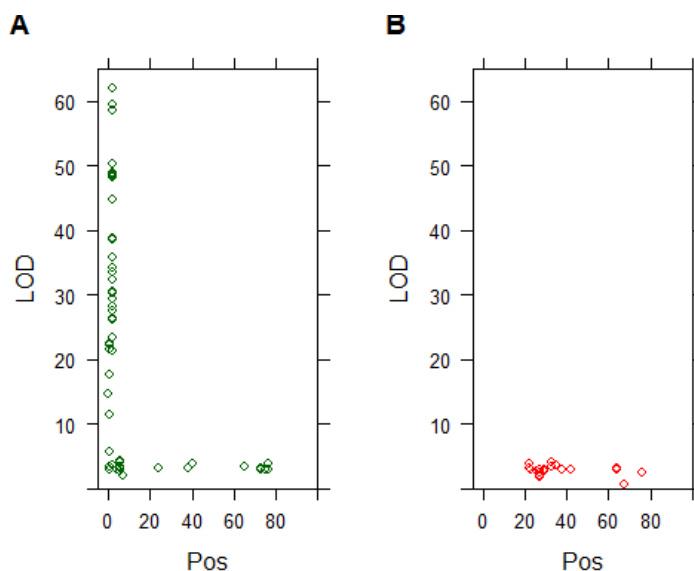


Figure 54 : QTL détectés sur le chromosome 3 avec la carte SNP\_Ri et les données non ciblées par UHPLC-MS ESI (+) de baies et de feuilles en 2015 ( $p_{val} \leq 0.05$ ). (A) : Feuilles 2015 (B) : Baies 2015.

Vingt et une formules brutes ont pu être identifiées, appartenant à la famille d'acides chlorogéniques ou encore des flavonoïdes (Tableau 48). Parmi les ions rattachés à ce QTL, l'injection de standard a permis de confirmer l'identification de deux molécules : le 5-caffeoylquinic acid et le coumaroyl quinic acid.

Ion	RT	LOD 5%	LOD 20%	Chr	Pos (cM)	LOD	pval	m/z	Ion pseudo-moléculaire	Identification / Proposition de structure
M517T96	1.6	2.81	2.25	3	0.87	22.5	0.00	517.155	[C22H28O14 + H] <sup>+</sup>	Glucoside de caffeoyl quinic acid
M353T108	1.8	3.17	2.42	3	2	38.64	0.00	353.087	[C16H16O9 + H] <sup>+</sup>	Chlorogenoquinone
M355T105	1.8	3.13	2.45	3	2	48.84	0.00	355.102	[C16H18O9 + H] <sup>+</sup>	5-caffeoylquinic acid
M377T106	1.8	3.11	2.45	3	2	62.06	0.00	377.084	[C16H18O9 + Na] <sup>+</sup>	5-caffeoylquinic acid
M372T108	1.8	3.01	2.34	3	2	44.91	0.00	372.129	[C16H18O9 + NH4] <sup>+</sup>	5-caffeoylquinic acid
M393T106	1.8	3.05	2.35	3	2	26.23	0.00	393.058	[C19H14O8 + Na] <sup>+</sup>	
M615T105	1.8	3.02	2.37	3	2	34.13	0.00	615.132	[C27H28O15 + Na] <sup>+</sup>	Flavonoïde
M661T106	1.8	2.95	2.27	3	0.87	17.78	0.00	661.176	[C31H32O16 + H] <sup>+</sup>	
M731T106	1.8	3.05	2.38	3	2	27.64	0.00	731.179	[C32H36O18 + Na] <sup>+</sup>	Flavonoïde
M163T106	1.8	3.06	2.39	3	2	48.18	0.00	163.039	[C9H6O3 + H] <sup>+</sup>	
M147T127	2.1	2.88	2.27	3	2	21.27	0.00	147.044	[C9H6O2 + H] <sup>+</sup>	
M230T129	2.2	2.80	2.26	3	0.87	21.54	0.00	230.057	[C13H10O4 + ] <sup>+</sup>	
M339T129	2.2	2.93	2.27	3	2	28.17	0.00	339.107	[C16H18O8 + H] <sup>+</sup>	Coumaroyl quinic acid
M361T129	2.2	2.99	2.30	3	2	29.44	0.00	361.089	[C16H18O8 + Na] <sup>+</sup>	Coumaroyl quinic acid
M629T132	2.2	2.70	2.11	3	0	14.64	0.00	629.187	[C31H32O14 + H] <sup>+</sup>	
M356T135	2.3	2.82	2.20	3	2	26.36	0.00	356.134	[C16H18O8 + NH4] <sup>+</sup>	Coumaroyl quinic acid

Tableau 48 : Ions identifiés dans les feuilles liés au QTL chromosome 3 Carte SNP\_Ri ( $p_{val} \leq 0.05$ ).

### 3. Discussion

#### 3.1. Diversité des molécules détectées dans les baies de la population RixGw et déterminisme génétique associé

L'analyse non ciblée par UHPLC-MS m'a d'abord permis d'identifier les familles de molécules les plus abondamment détectées dans les baies par cette technique. Les flavonoïdes sont les molécules les plus représentées parmi les molécules les plus abondamment détectées (par exemple l'astilbine), puis les monoterpénols. Ces composés majeurs sont présents sous forme mono ou pluri-glycosylés. Les analyses réalisées par UHPLC-MS par rapport aux analyses par GC-MS ont l'avantage de donner une information sur le glycone lié à l'aglycone.

La quantification de ces ions dans les baies de la population de croisement RixGw, de Ri et de Gw a permis d'identifier les ions majeurs présents de manière différentielle entre le Ri et le Gw mais aussi des molécules présentes dans les analyses de la population RixGw mais qui sont absentes dans les analyses des parents. C'est le cas par exemple de l'astilbine et de l'apiosyl-glucoside de linalol oxyde qui sont présents de manière plus importante dans les baies de Gw que dans les baies de Ri. Parmi ces molécules se trouvent des dérivés glycosylés de molécules volatiles qui peuvent ainsi constituer des précurseurs d'arômes potentiels.

L'analyse QTL réalisée avec les données métabolomiques non ciblées et 3 cartes génétiques m'a permis de déterminer les régions génomiques gouvernant des variations dans l'accumulation d'un grand nombre de molécules présentes dans les baies de raisins, les 2 années d'analyse me permettant de confirmer la robustesse du lien entre une région génomique et le contenu en un métabolite secondaire spécifique. J'ai identifié des hotspots génétiques sur les chromosomes 5, 6 et 12, ces hotspots gouvernant chacun la biosynthèse d'un grand nombre de molécules

Un hotspot crucial sur le chromosome 5 a été détecté avec la carte génétique du Gw. Ce hotspot génétique est le plus important car il impacte le plus grand nombre de précurseurs d'arômes dans la population d'étude. Celui-ci colocalise avec le gène de la DXS, un gène qui a déjà été caractérisé chez la vigne (Battilana et al., 2009; Dalla Costa et al., 2018; Duchêne et al., 2009; Huguency et al., 2012). Ce gène, situé dans la voie de biosynthèse plastidiale MEP, est responsable de la biosynthèse du précurseur d'un très grand nombre de molécules de type isoprénoïdes, l'IPP. Parmi les isoprénoïdes, il impacte surtout la biosynthèse de composés de type monoterpénols mais des sesquiterpènes, des diterpènes ainsi que des norisoprénoïdes sont

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

aussi impactés par ce QTL. L'impact de ce gène sur la biosynthèse de monoterpénols, de diterpènes ainsi que de norisoprénoïdes n'est pas étonnant puisque ce gène se trouve au début de la voie de biosynthèse plastidiale de ces molécules. L'impact de ce gène sur les molécules de type sesquiterpènes est une plus étonnant puisque leur voie de biosynthèse est cytosolique (voie MVA) et que la DXS ne participe pas à cette voie de biosynthèse. Ces résultats sont en faveur d'un transport de l'IPP du plaste vers le cytosol lorsque celui-ci est présent en forte concentration dans le plaste, IPP qui serait ensuite utilisé pour la biosynthèse de sesquiterpènes en complément de l'IPP produit par la voie du MVA. Un tel transport a déjà été observé chez d'autres plantes (Laule et al., 2003), mais le ou les transporteurs impliqués n'ont pas été caractérisés à ce jour (Pick and Weber, 2014). L'impact global du locus DXS sur le métabolisme des raisins est analysé plus en détail dans le chapitre 3.

Un hotspot sur le chromosome 12 lié à des molécules de type norisoprénoïdes et monoterpénols a été détecté avec la carte SNP\_Ri. Ce QTL a un impact surtout sur les molécules de type norisoprénoïdes où l'on détecte des LOD scores élevés. Ce QTL semble lié à un gène codant pour une glucosyltransférase, gène candidat qui ne sera pas détaillé dans ce manuscrit.

Un hotspot sur le chromosome 6 lié à des molécules de la famille des flavonoïdes a été détecté avec la carte génétique SNP\_Gw. L'effet de ce QTL est détecté pour de nombreuses molécules avec notamment un effet sur l'astilbine, une molécule retrouvée de manière abondante dans les baies.

### 3.2. Les glycosides : des précurseurs d'arômes potentiels abondants dans les baies

Dans le cadre de l'étude des précurseurs d'arômes et des arômes des baies, j'ai recherché de manière ciblée à identifier et à quantifier les molécules de la famille des monoterpénols et des norisoprénoïdes. Ceci a été possible grâce à l'utilisation de leurs fragmentations caractéristiques. L'analyse des monoterpénols dans les baies a permis d'identifier 15 monoterpénols, la plupart d'entre eux étant liés à des sucres.

L'analyse génétique des monoterpénols par UHPLC-MS a permis de confirmer les QTL déjà identifiés pour les monoterpénols par GC-MS dans le chapitre 1. Ainsi, pour les monoterpénols, on retrouve les QTL sur les chromosomes 5, 10, 13 qui ont déjà été trouvés auparavant avec les analyses par GC-MS. Un nouveau QTL sur le chromosome 12 impacte 2



## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

monoterpénols, dont l'apiosyl-glucoside de linalol oxyde qui est retrouvé le plus abondamment dans les baies de la population RixGw.

Parmi les monoterpénols identifiés, leurs quantifications dans les baies de la population RixGw et de ses parents a montré que l'apiosyl-glucoside de linalol oxyde est présent en très forte quantité dans les baies de certains individus de la population RixGw mais aussi qu'il y en a significativement plus dans le Gw que dans le Ri. L'abondance de cette molécule dans les baies de Gw soulève la question de la raison de la très faible concentration de linalol dans les vins de Gw par rapport aux vins de Ri (Mateo and Jiménez, 2000). Ceci soulève l'hypothèse d'une biosynthèse du linalol dans les baies de Gw qui serait oxydé puis glycosylé pour donner l'apiosyl-glucoside de linalol oxyde. Ce composé avait déjà été auparavant isolé et caractérisé à partir de vins de Gw dont il est un composé présent de manière importante (Baltenweck-Guyot et al., 2000).

L'analyse des norisoprénoïdes dans les baies avait pour but d'identifier de potentiels précurseurs du TDN, une molécule spécifique de l'arôme de type kérosène caractéristiques des vins de Ri âgés. Aucun QTL reproductible sur les deux années de prélèvements n'avait été détectés avec les analyses par GC-MS. Des QTL reproductibles pour les norisoprénoïdes détectés par UHPLC-MS ESI (+) sur les deux années de prélèvement ont été détectés pour les norisoprénoïdes sur les chromosomes 1, 12, 15.

Un premier glycoside de norisoprénoïde a été détecté de manière abondante à la fois chez le Ri et le Gw ainsi que dans leur descendance RixGw. Une identification putative de cette molécule est le glycoside de vomifoliol. Ce composé avait déjà été identifié dans les feuilles de Riesling (Skouroumounis and Winterhalter, 1994) et des vins de Gw (Baltenweck-Guyot et al., 1996). Le contenu en ce norisoprénoïde n'est pas significativement différent entre le Ri et le Gw, ce qui en fait un mauvais candidat comme précurseur du TDN dans les vins de Ri mais représente un précurseur de norisoprénoïde majeur pouvant être libéré pendant la vinification.

Un deuxième norisoprénoïde a été identifié comme étant différentiel dans les analyses de baies de Ri et de Gw de manière reproductible sur deux années d'analyse. La nature possible de cette molécule est un malonyl-glucoside de 3-oxo- $\alpha$ -ionol. Ce composé est impacté par un QTL sur le chromosome 12 avec la carte génétique SNP\_Ri et est un potentiel candidat comme précurseur de TDN. Pour voir si ce locus est responsable de la concentration en TDN des vins de Ri, nous pourrions sélectionner les génotypes en fonction de leur génotypage au locus du QTL et en étudier les vins correspondants. Si ce locus gouverne la biosynthèse de précurseurs

## Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées

du TDN dans le Ri, une différence significative en quantité de TDN devrait être retrouvée entre les vins issus des deux groupes de génotypes, cette différence apparaissant cependant seulement après plusieurs années de garde.

### 3.3. Les extraits de feuilles, un modèle intéressant à étudier

En complément des analyses métabolomiques de baies de raisins, des analyses par UHPLC-MS ont été faites sur des feuilles prélevées en 2015. Ces analyses ont pour but d'étudier si les feuilles, faciles à échantillonner, pouvaient être utilisées comme modèle d'étude pour prévoir certaines caractéristiques des baies, qui ne peuvent être récoltées qu'une fois par an.

L'extraction réalisée sur les baies et celle sur les feuilles est différente, ceci pouvant expliquer les différences de nombres d'ions détectés dans les baies et les feuilles. En effet, l'extraction directe au méthanol des échantillons de feuilles est plus simple et rapide que l'extraction des baies qui comporte une étape de SPE. Il n'y a pas d'étape SPE pour les échantillons de feuilles, étape qui élimine les composés polaires tels que les sucres et les acides et qui permet de concentrer les échantillons. On observe ainsi un nombre d'ions bien inférieur dans les feuilles que dans les baies, avec des molécules communes d'intérêt, mais aussi des molécules qui ne sont pas ou peu détectées dans les extraits SPE de baies. En effet, l'étude des composés polaires dans les baies est rendue difficile par la présence en grande quantité de composés tels que les sucres et les acides organiques et n'est pas possible avec les analyses des extraits SPE.

L'analyse du déterminisme génétique du métabolome des feuilles a mis en évidence des hotspots similaires aux analyses génétiques des baies sur les chromosomes 5, 6 et 12. Des QTL spécifiques des feuilles sont trouvés sur le chromosome 3. L'analyse des feuilles que j'ai réalisée pendant ma thèse constitue un premier test et devrait être optimisée pour aller plus loin. Une étape de concentration en plus des extraits de feuilles (par évaporation du solvant au Speedvac par exemple) pourrait être testée pour tenter de détecter davantage de molécules.

À terme, les molécules répondant de la même manière dans les baies et dans les feuilles avec des QTL similaires pourront être utilisées comme marqueur de qualité des baies. Ainsi il serait possible d'analyser les feuilles d'une population de croisement afin prédire précocement certains paramètres de la qualité des baies avant même leur apparition. En effet, l'analyse des

## **Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées**

feuilles peut être réalisée la première année lors de la mise en place de la population alors que les baies ne sont disponibles que 3 ans après la mise en place du dispositif expérimental en champ. Avec notre population de croisement RixGw, les molécules pouvant être utilisées comme marqueur de qualité seraient les molécules de type flavonoïdes et terpénols. Ceci sera à confirmer avec une autre année d'analyse de feuilles et de baies de la population RixGw.

### **3.4. L'identification précise des molécules, une tâche difficile**

L'analyse du métabolome par UHPLC-MS permet d'avoir une vision très complète du métabolome des baies et des feuilles. La très bonne précision en masse en ESI (+) de moins d'1 ppm ainsi que la fragmentation des molécules permet d'avoir des indices sur la structure des molécules, cependant elle ne permet pas à elle seule de déterminer la structure exacte. Avec cette technique, l'identification précise ne peut se faire que par l'injection d'un standard de la molécule à identifier dans les mêmes conditions que nos analyses. Les molécules les plus détectées dans nos analyses de baies et de feuilles par UHPLC-MS correspondaient à des molécules glycosylées, pour lesquelles il existe très peu de standards commerciaux. En l'absence de standards, la seule manière de pouvoir obtenir la structure de ce type de molécules serait de les isoler et de les analyser par Résonance Magnétique Nucléaire (RMN). Ceci ne serait cependant faisable que pour les molécules présentes en grande quantité dans les échantillons.

### **3.5. La métabolomique non ciblée en UHPLC-MS pour l'analyse du déterminisme génétique de la biosynthèse d'un grand nombre de métabolites**

L'analyse globale du métabolome des baies de raisins ainsi que de feuilles par UHPLC-MS des baies prélevées en 2014 et 2015 m'a permis de caractériser la grande diversité de molécules présentes dans échantillons. Grâce à ces analyses, des milliers d'ions ont pu être détectés et quantifiés dans les baies de Ri, Gw et RixGw. L'étude des molécules de manière non ciblée par UHPLC-MS permet d'ouvrir l'analyse aux molécules précurseurs d'arômes, mais aussi à d'autres molécules de type flavonoïdes qui n'étaient pas détectées par les analyses par GC-MS du fait de leur faible volatilité (Chapitre 1). L'une des difficultés à gérer avec ce type de données est de diminuer la complexité des jeux de données. Dans mon cas, j'ai utilisé les m/z comme point de référence entre les différents jeux de données obtenus par UHPLC-MS ESI (+).

## **Chapitre 2 Caractérisation du métabolome des raisins par UHPLC-MS et analyses génétiques associées**

L'analyse du métabolisme secondaire de manière non ciblée par UHPLC-MS est une approche toute récente. L'analyse non ciblée du métabolisme secondaire a été réalisée sur des plantes tels que le riz (Navarro-Reig et al., 2017) , l'orge (Ghosson et al., 2018), le sorgho (Mareya et al., 2019) ou encore la fraise (Antunes et al., 2019). Nos travaux sont les premiers à analyser de manière non ciblée le métabolome de baies de raisins et de feuilles de vigne par UHPLC-MS, et à utiliser ces données pour la recherche de QTL à l'échelle du génome chez les plantes. Cette approche innovante permet d'identifier des régions génomiques impactant un grand nombre de métabolites différents. Des milliers de QTL ont été identifiés avec les 3 cartes génétiques utilisés, avec une certaine redondance entre les QTL identifiés avec la carte génétique SSR et les cartes génétiques parentales. Cependant, l'effet additif des allèles des parents ainsi que les effets additifs de plusieurs QTL sur une même molécule ne peut être étudié de manière exhaustive que grâce à la carte consensus SSR.



**Chapitre 3 Analyse en réseaux du métabolome de la  
vigne par la combinaison d'analyses par FT-ICR et  
UHPLC-MS à la lumière des QTL**

**Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses  
par FT-ICR et UHPLC-MS à la lumière des QTL**

## Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

### 1. Introduction

Malgré ses avantages, la technique UHPLC-MS souffre de l'imprécision relative des mesures de masse qui complique l'identification des molécules. En effet, plus la précision en masse est bonne, plus il est facile d'attribuer une formule brute précise aux molécules et il est possible d'automatiser l'annotation des m/z. Pour contourner ce problème, j'ai eu l'opportunité d'utiliser la technique de spectrométrie de masse à résonance cyclotronique ionique (FT-ICR) permettant d'obtenir les mesures de masse les plus précises à l'heure actuelle, afin d'obtenir des identifications plus fiables des ions et de pouvoir les annoter. Ces analyses ont été réalisées lors d'un séjour de 3 mois que j'ai effectué dans l'équipe du Professeur Philippe Schmitt-Kopplin au Helmholtz Zentrum de Munich (Allemagne).

Des analyses par UHPLC-MS et FT-ICR ont été réalisées de manière sur les mêmes extraits de jus de la population RixGw prélevées en 2018. Ces analyses m'ont permis de quantifier des milliers de molécules détectées par les deux instruments. L'analyse QTL réalisée à partir de ces deux types de données m'a permis de valider les résultats obtenus grâce à chacune des deux technologies et de combiner les avantages des deux types d'instruments, pour une exploitation optimale des données métabolomiques.

Cette analyse génétique très complète du métabolisme secondaire de la baie de raisin et l'ensemble de ces résultats fait l'objet d'un article actuellement en préparation (Truong TT et al., Genetic-driven molecular network analysis of grape berry metabolism). Ces résultats sont originaux, car la combinaison d'analyses UHPLC-MS et FT-ICR n'avait jamais été utilisée auparavant dans le cadre d'analyses génétiques.

Pour ces travaux, les prélèvements ont été réalisés par Nathalie Jaegli et Vincent Dumas, l'extraction des échantillons a été réalisé par Sonia Hommay et Camille Banholzer, le génotypage a été réalisé par Gisèle Butterlin, les injections par UHPLC-MS ont été réalisées par Raymonde Baltenweck et Camille Banholzer. J'ai réalisé les analyses par FT-ICR avec l'aide de Basem Kanawati dans l'équipe BioGeoChemistry de Philippe Schmitt-Kopplin lors d'un séjour de 3 mois au Helmholtz Zentrum à Munich. J'ai réalisé le traitement des données conjointe des données FT-ICR et UHPLC-MS avec l'aide de Jenny Ulm et de Raymonde Baltenweck ainsi que la recherche des QTL. L'annotation des ions étude des réseaux métabolomiques a été réalisé avec le logiciel Netcalc développé par Marianna Lucio dans l'équipe du professeur Philippe Schmitt-Kopplin.



2. Article 2

**Genetic-driven molecular network analysis of grape berry  
metabolism**

Truong TT, Baltenweck R, Jaegli N, Butterlin G, Dumas V, Claudel P, Maia-Grondard A, Uhl J, Kanawati B, Lucio M, Rustenholz C, Merdinoglu D, Schmitt-Kopplin P, Duchêne E, Hugueney P.

**Abstract**

Grape berry metabolism is crucial for wine quality and contributes to cultivar-specific aromatic signatures, which greatly add to the enjoyment of wine tasting. Using a nontargeted approach by ultrahigh-resolution mass spectrometry, we analyzed the berries in a grapevine cross population, in order to evaluate the impact of sexual reproduction on grape berry metabolome. These metabolomic analyses revealed a considerably enlarged chemical landscape in the progeny in comparison to parental lines, showing that hybridization may generate individuals with greatly enhanced chemical diversity in berries. Furthermore, genome-wide QTL analysis showed that specific genetic variations in a limited number of genomic regions impact the biosynthesis of hundreds of molecules, with major consequences on grape aroma profiles. Finally, we show that the combination of untargeted metabolomics, genetics analyses and molecular networking constitutes a powerful approach to decipher complex metabolic pathways in plants.

Key words: Grapevine, QTL, FTICR-MS, UHPLC-MS, molecular network, terpenoid metabolism

## **Introduction**

Grape (*Vitis vinifera*) is one of the earliest domesticated fruit crops (McGovern et al., 1996) and, since Antiquity, it has been widely cultivated for its fruit and for winemaking. Many starch or sugar-rich plant derived products may be fermented, giving rise to a wide variety of alcoholic beverages. However, among those, wine has a special status, being closely linked to human history and still being today one of the most widely prized beverages. The worldwide success of wines and grape-derived spirits is most probably due to their extraordinary aromatic diversity and complexity. The composition of wine is determined by a complex interplay between environmental factors, genetic factors and viticultural practices (Roullier-Gall et al., 2014a). The terroir, which describes the typicality of wine in relation to its geographical origin, is considered to greatly impact wine quality (van Leeuwen et al., 2004; Roullier-Gall et al., 2014b). Nevertheless, the winemaker's primary choice of planting a given grapevine variety is largely influenced by desired grape characteristics, many of which are determined by specific genetic factors. Grapevine has retained high levels of genetic diversity since its domestication (Myles et al., 2011) and the current diversity of grape cultivars used in the wine industry offers a wide panel of berry characteristics, such as size, color, flavors, and aromas. This diversity of cultivars has been created through natural sexual crossing or breeding and through natural mutations, which may affect clones of the same cultivar (This et al., 2006; Pelsy et al., 2010). In recent years, this diversity has been exploited to decipher the genetic determinism of different grape compositional traits, which are known to greatly impact grape and wine quality. Indeed, genetic analyses of grape composition using different mapping population has led to the identification of quantitative trait loci (QTLs) associated to the contents in both primary and secondary metabolites, including sugars, organic acids, phenolics and aroma compounds (Poni et al., 2018; Lin et al., 2019). Grapevine secondary metabolites, such as anthocyanins, tannins and aroma compounds, are major determinants of grape and wine sensory characteristics (Polášková et al., 2008; Ali et al., 2010). The richness of grape specialized metabolism and its potential for giving rise to an exceptional diversity of flavor compounds contribute greatly to the aromatic complexity of wine. Analysis of the grapevine genome has shown a remarkable expansion of several gene families linked to secondary metabolism, compared to other plants (Jaillon et al., 2007). A striking example of gene family expansion is the terpene synthase (TPS) family that generates aromatic molecules constituting major contributors to wine aromas (Schwab and Wüst, 2015). Grapevine exhibits the second largest TPS family of all plant species for which a genome sequence is available (Martin et al., 2010). Interestingly, the expansion of

### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

the TPS family in grapevine has been accompanied by a parallel blooming of some cytochrome P450 subfamilies, in particular those shown to be involved in the oxidation of monoterpenols (Ilc et al., 2018). In some cases, combined genetic and functional genomic analyses have allowed the identification and characterization of genes associated to QTLs impacting the biosynthesis of distinctive grape secondary metabolites, such as anthocyanins (Fournier-Level et al., 2009), proanthocyanidins (Huang et al., 2012), monoterpenols (Battilana et al., 2009; Duchêne et al., 2009) or methoxypyrazines (Dunlevy et al., 2013; Guillaumie et al., 2013). However, the current understanding of the genetic bases of grape composition is far from complete and few genetic markers related to grape and wine organoleptic quality traits are available, although such markers are much needed. In particular, there is currently a considerable interest in breeding new grapevine varieties with durable resistance to major diseases affecting viticulture, by exploiting natural resistance sources from wild *Vitis* species (Feechan et al., 2013). Indeed, disease-resistant grapevine varieties offer a unique opportunity to satisfy the strong societal demand for the production of high-quality wines, while lowering the environmental impacts of viticulture (Merdinoglu et al., 2018). However, efficient breeding of qualitative grapevine varieties, whose characteristics are adapted local specifications, require molecular markers related to berry composition, that allow the combination of multiple factors to reach the expected phenotype. In order to better understand the genetic determinism of secondary metabolism, we studied a progeny from a bi-parental cross between Riesling (Ri) and Gewurztraminer (Gw). Ri and Gw are major grapevine cultivars of the Rhine Valley, which strongly differ for numerous oenologically relevant traits. This Riesling x Gewurztraminer (RixGw) progeny has successfully been used for analyzing the genetic determinism of phenology (Duchêne et al., 2012a), sugar accumulation (Duchêne et al., 2012b), berry acidity (Duchêne et al., 2014) and grape aroma biosynthesis (Ilc et al., 2017). As aroma-related metabolism greatly differs in Ri and Gw berries, an analytical approach allowing a wide characterization of berry metabolome was used.

Combinations of non-targeted metabolomic approaches using complementary analytical strategies have already demonstrated their unique potential for the metabolite coverage, validation of data and identification of compounds of interest (Forcisi et al., 2013; Forcisi et al., 2015). In particular, Fourier transform ion cyclotron resonance-mass spectrometry (FTICR-MS) has proven to be a powerful technique for studying wine, mostly employed in direct infusion experiments to benefit from its ultra-high resolving power and mass accuracy (Liger-Belair et al., 2009; Roullier-Gall et al., 2014a; Roullier-Gall et al., 2014c). However, direct

### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

infusion techniques do not allow to discriminate between isomeric and isobaric compounds and are sensitive to ion suppression phenomena (Forcisi et al., 2013). On the other hand, ultra-high-pressure liquid chromatography coupled to mass spectrometry (UHPLC-MS) takes advantage of high-performance chromatographic separation and fragmentation-derived structural information, although mass spectrometric accuracy does not match that of FTICR-MS. Therefore, combined analysis with FTICR-MS and UHPLC-MS allows optimal exploitation of the best of both techniques, to offer a unique power for the analysis of complex metabolomes (Forcisi et al., 2015). In this study, we have applied for the first time a FTICR-MS-based metabolomic profiling strategy to obtain a non-targeted genome-wide QTL analysis of grape berry metabolome. Although this large-scale non-targeted approach allowed the detection of several thousands of metabolic QTLs spanning the entire grapevine genome, we show that specific genetic variations in a small number of genomic regions impact the biosynthesis of hundreds of molecules, with major consequences on grape aroma profiles. Furthermore, taking advantage of the power of combined FTICR-MS and UHPLC-MS metabolomics, we analyzed the metabolic network linked to a major QTL to decipher the pathway leading to major terpenoids in grape berries.

## **Results and Discussion**

### **FTICR-MS profiling reveals a wide metabolic diversity in a grapevine progeny**

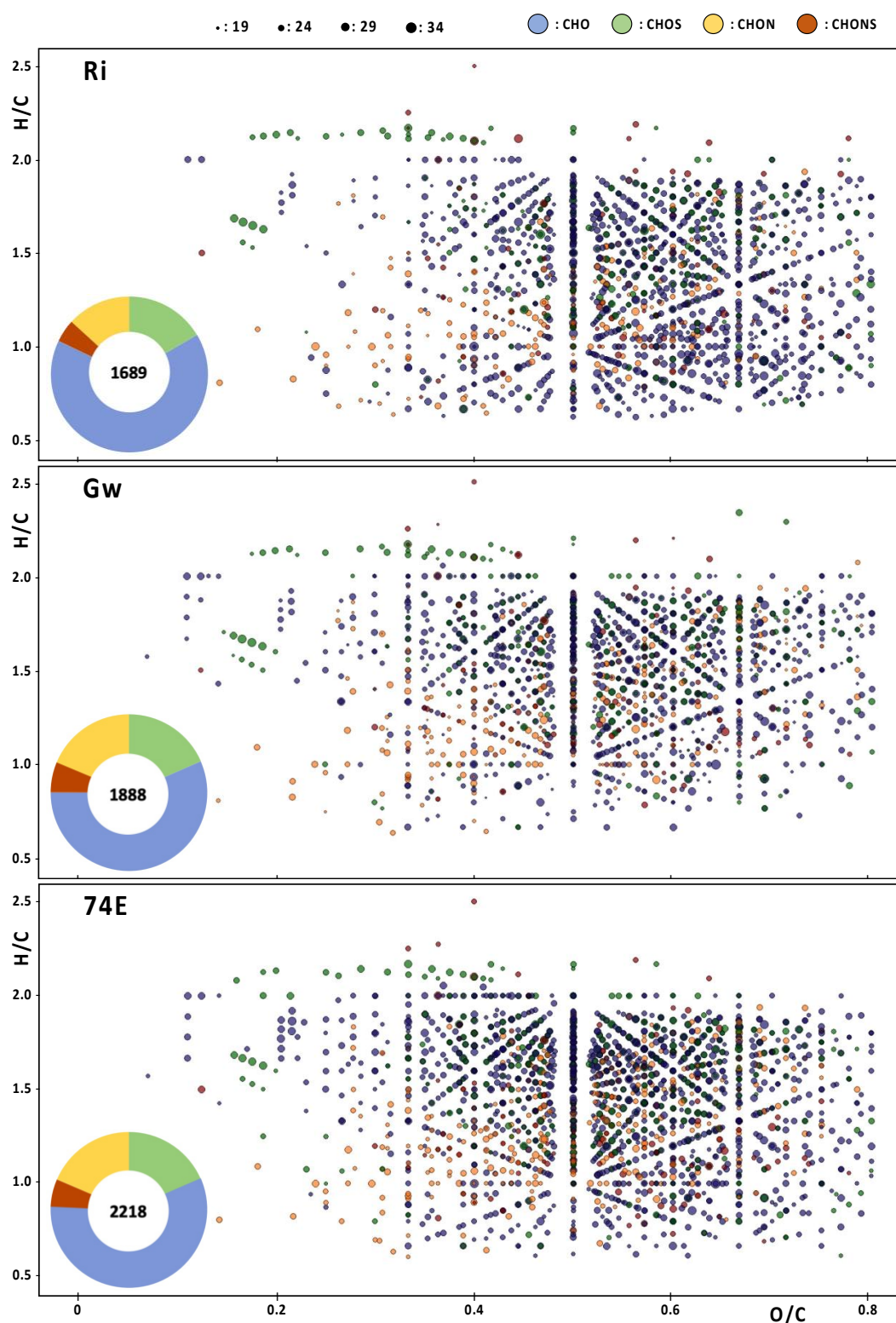
FTICR-MS was used both in the negative and the positive electrospray ionization (ESI) mode for the non-targeted analysis of the low-molecular weight metabolites present in grape berry extracts of all genotypes from the RixGw progeny. These analyses were focused on semi-polar compounds, the most polar ones being excluded by the solid phase extraction (SPE) step used for the preparation of berry extracts. SPE was used to eliminate the large amounts of sugars and organic acids accumulated in mature grape berries, which would have interfered with the characterization of other classes of metabolites. Nevertheless, semi-polar compounds from grapes include large numbers of specialized metabolites, which contribute greatly to the organoleptic properties and aromatic complexity of wine.

Within the mass range explored (92-1,000 m/z), FTICR-MS spectra exhibited several thousand peaks, which corresponded to metabolites that can be ionized under the selected electrospray conditions. Totals of 13,149 and 37,720 raw mass values were obtained in ESI (-) and ESI (+),

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

respectively. Only masses occurring in a minimum of 10% of all samples were used for the calculation of elemental formulas with the NetCalc software (Tziotis et al., 2011). Two datasets were thus obtained, containing respectively 3,251 masses in ESI (-) and 9,047 masses in ESI (+), assigned to an elementary composition. This information was used to construct van Krevelen (VK) diagrams representing the chemical landscape associated to selected grape berry extracts, by distributing each elemental composition onto 2 axes according to its H/C and O/C atomic ratios (**Fig. 1**). As examples, the chemical composition and the VK diagrams highlighted a wide chemical diversity in berry extracts from the parental genotypes Ri and Gw, as well as in the 74E hybrid. Each sample contained about 2,000 elements exhibiting similar distribution of elementary compositions, with a significant contribution of CHO formulas (**Fig. 1**). However, VK diagram highlighted a more complex metabolic profile in berries of the 74E hybrid, than in the parental lines.

## Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

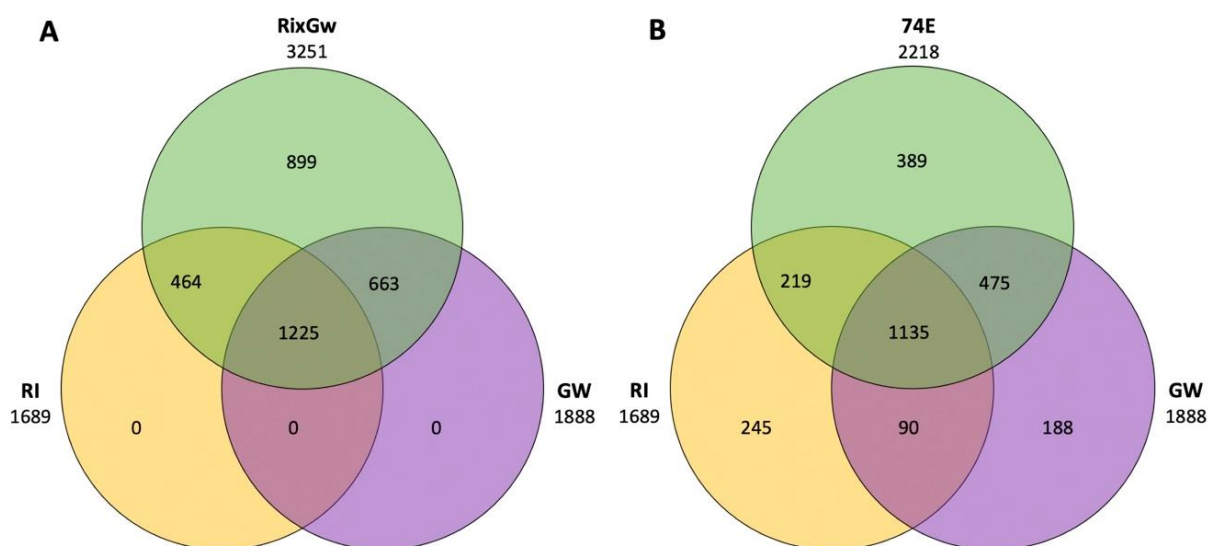


**Fig. 1. Overview of grape berry metabolome represented as van Krevelen diagrams**

ESI (-) FTICR-MS data of berry extracts from Ri, Gw and from the 74E genotype of the RixGw progeny were assigned to elementary compositions and used to construct van Krevelen diagrams presenting H/C vs O/C atomic ratios. Elemental compositions were attributed as CHO (blue), CHOS (green), CHON (yellow) and CHONS (red). Point size indicate log<sub>2</sub> mass peak intensity for each compound. For each genotype, a pie chart indicates the total number of attributed compounds, as well as their distribution among elemental formula classes.

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

In order to visualize the metabolic diversity in the RixGw progeny, Venn diagrams were used to compare the berry metabolome of parental lines to those of the entire population or to that of a single sibling (**Fig. 2**). Nearly half of the metabolites contained in the berries of Ri and Gw were specific to only one of the parental lines, highlighting substantial differences in berry compositions. The RixGw progeny captured the whole parental metabolic diversity, since all elementary compositions present in Ri or Gw were also present in the progeny (**Fig. 2A**). Furthermore, nearly one third of the elementary compositions present in the progeny were not found in parental lines, indicating that genetic mixing gave rise to large diversity of novel metabolites. Comparing the berry metabolome of the 74E hybrid genotype to that of its parents revealed a remarkable metabolic complexity. Berry extracts of the 74E hybrid contained more elementary compositions (2,218) than the Ri and Gw parents (1,689 and 1,888, respectively), including a large number of 74E-specific compounds (389), which were not detected in the berries of the parental genotypes (**Fig. 2B**). Interestingly, berries of the 74E hybrid alone harbored a significant proportion (over 40%) of the metabolic novelty generated in the whole RixGw progeny. Cultivated grapevine varieties have been shown to harbor very high levels of genetic diversity, comparable to that of maize (Lijavetzky et al., 2007; Myles et al., 2011). This tremendous genetic diversity bears the potential to generate wide phenotypic variations in a single cross between grapevine varieties. This potential may be used to create new varieties with greatly enhanced metabolic complexity, possibly leading to novel oenological properties.



**Fig. 2. FTICR-MS analysis of grape metabolic diversity in the RixGw progeny**

The distribution of elementary compositions in berry extracts from Ri, Gw and from all genotypes of the RixGw progeny are shown in Venn diagrams. **A.** Metabolic diversity in the whole RixGw progeny compared to parental lines **B.** Metabolic diversity in the 74E genotype compared to parental lines

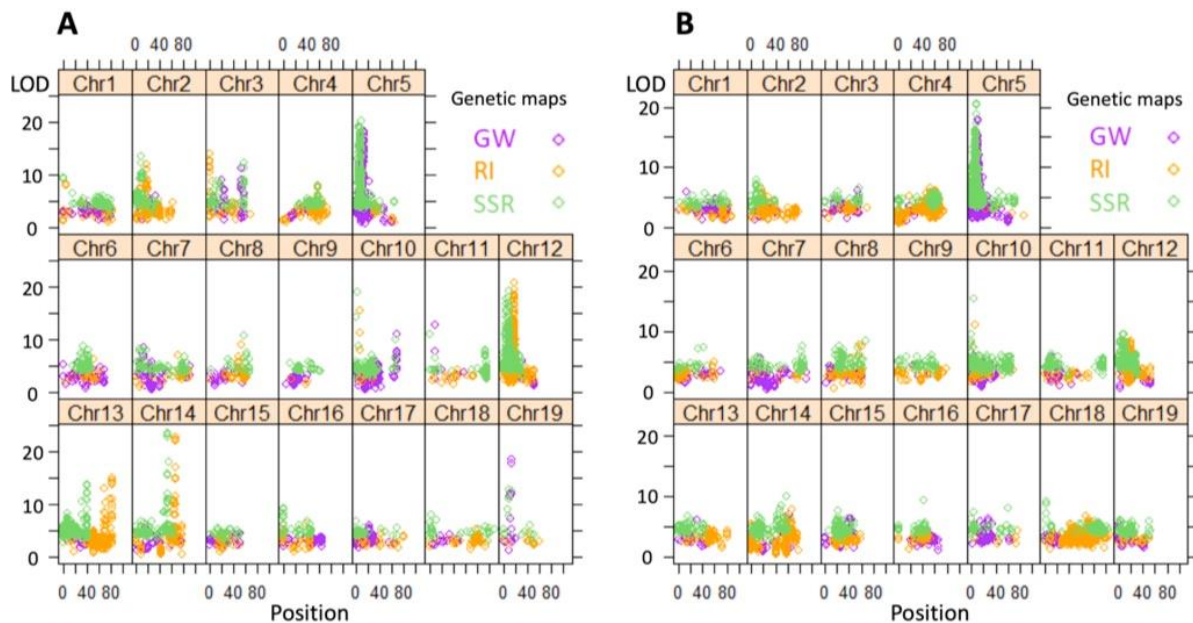
### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

#### *Genetic bases of metabolic diversity in the RixGw progeny*

High density genotyping of the RixGw progeny has allowed targeted genetic analysis of several grape metabolic traits, including the accumulation of sugars, organic acids and of the precursor of wine lactone (Duchêne et al., 2012b; Duchêne et al., 2013; Ilc et al., 2017). Taking advantage of the high resolving power and mass accuracy of FTICR-MS, both ESI (-) and ESI (+) analytical datasets were used for genome-wide mapping of quantitative trait loci (QTLs). Metabolic QTL (mQTL) analysis using ESI (-) and ESI (+) data led, respectively, to the detection of 2942 and 6037 significant mQTLs ( $p < 0.05$ ) impacting the accumulation of a broad diversity of metabolites. mQTLs were distributed all over the 19 chromosomes of grapevine, with LOD scores reaching 20 (**Fig. 3**). Interestingly, polymorphism at a small number of genomic hotspots resulted in high-LOD scores mQTLs, which impacted large number of molecules. This is especially true for mQTLs mapped on chromosomes 5 and 12 with ESI (-) data (**Fig. 3A**), the mQTL on chromosome 5 being detected with ESI (+) data as well (**Fig. 3B**). Most mQTLs detected with the simple sequence repeat (SSR) and Gw maps on chr5 converged to one genetic locus, which colocalized with the gene coding for the enzyme 1-deoxy-D-xylulose-5-phosphate synthase 1 (*DXS1*) (**Fig. S1**). Specific *DXS1* alleles were previously associated with the high terpene contents in berries from aromatic grapevine varieties such as Muscats and Gewurztraminer (Battilana et al., 2009; Duchêne et al., 2009; Emanuelli et al., 2010; Huguency et al., 2012). However, the enhanced enzymatic activity of these particular *DXS1* alleles were shown to influence a small number of volatile monoterpenols such as geraniol, linalool, nerol, citronellol and  $\alpha$ -terpineol (Battilana et al., 2009; Duchêne et al., 2009; Battilana et al., 2011). Untargeted FTICR-MS metabolomic analyses revealed here that the enhanced activity of the enzyme encoded by the Gw-specific *DXS1* allele was likely to impact the accumulation of hundreds of metabolites, resulting in major changes in berry terpenoid profiles. Indeed, most of the QTLs associated to chromosome 5 co-localized with the *DXS1* locus (**Fig. S1A**), which was linked to 1113 mQTLs ( $p < 0.05$ ), potentially affecting the accumulation of a similar number of metabolites (**Fig. S1B**).



### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL



**Fig. 3. Genome-wide QTL analysis of grape metabolism in the RixGw progeny**

Relative quantifications of elementary compositions in berry extracts from all genotypes of the RixGw progeny were used to map metabolic QTLs. Each point corresponds to a mQTL significantly associated ( $p < 0.05$ ) to the amount of an elementary composition in berry extracts, using the SSR, Ri or Gw genetic maps as indicated. mQTL positions on chromosomes are shown on the horizontal axis and the LOD score of each QTL is indicated on the vertical axis. **A.** mQTL analysis using ESI (-) FTICR-MS data. **B.** mQTL analysis using ESI (+) FTICR-MS data.

In order to better characterize the extent of these changes affecting grape terpenoid metabolism, FTICR-MS analyses were complemented by UHPLC-MS analyses of the same samples. A simple comparison of extracted ion chromatograms (EIC) focused on major terpenoids such as geraniol and linalool-derived compounds showed that a genotype with favorable *DXS1* allele such as 74E exhibited a much more complex terpenoid profile than a non-aromatic genotype (**Fig. S2**). *DXS* is the first enzyme in the methylerythritol phosphate (MEP) pathway for terpenoid biosynthesis in plants. This enzyme is known to play a major role in the regulation of terpene biosynthesis in plastids, due to its limiting activity within this metabolic pathway (Estévez et al., 2001; Wright et al., 2014). The massive impact of highly active *DXS* alleles may offer unique insight into terpenoid metabolism in grape. Indeed, we reasoned that a large proportion of the compounds mapping to the QTL associated to the *DXS1* gene should derive from the grape berry terpenoid pathway, whose organization may be deduced from the chemical relationships existing among this group of compounds.

*Grape terpenoids as a proof-of-concept for genetic-driven pathway analysis*

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

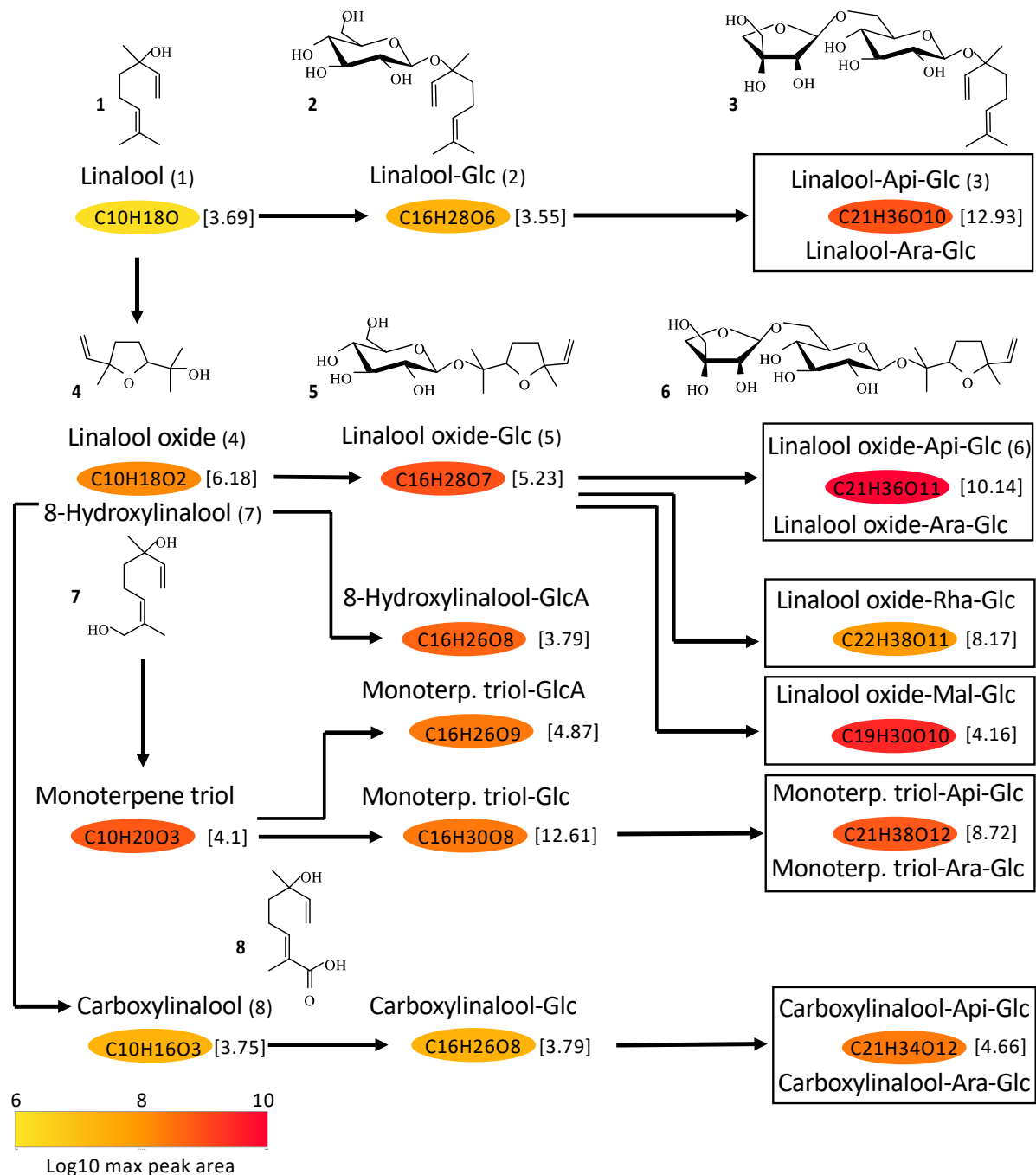
Selection of CHO elemental compositions associated with the QTL co-localizing with the *DXSI* gene resulted in a set of 129 compounds with putative formula, most of which could be matched with compounds detected with UHPLC-MS (**Table S1**). Matching of FTICR-MS and UHPLC-MS analytical data allowed to combine the high mass accuracy of FTICR-MS with fragmentation-derived structural information from UHPLC-MS (Forcisi et al., 2015). In order to decipher the chemical relationships existing among these 129 compounds, network analysis was performed as previously described (Tziotis et al., 2011), with nodes representing metabolite candidates and their associated elementary composition and edges representing chemical reactions (**Fig. S3**). Most metabolites were distributed in two main networks, starting respectively with C10 or C20 compounds. Most of the major compounds giving large peak areas in both FTICR-MS and UHPLC-MS analyses, belonged to the sub-network derived from C10 precursors (**Fig. S3**). Based on this molecular sub-network, detailed analysis of UHPLC-MS data and comparison to published literature, we constructed the possible pathway leading to major linalool derivatives of grape berries, with proposed molecular identifications of both intermediates and end products (**Fig. 4**). Oxidized derivatives of linalool form a complex family of plant metabolites, some of them participating in the aroma of flowers and fruits (Ilc et al., 2016). Hydroxylated linalool derivatives such as 8-hydroxylinalool ( $C_{10}H_{18}O_2$ ) and monoterpene triols ( $C_{10}H_{20}O_3$ ) have been proposed as aroma precursors in Muscat grapes (Williams et al., 1980) and some glycosides have been characterized (Strauss et al., 1988; Marinos et al., 1994). Carboxylinalool ( $C_{10}H_{16}O_3$ ) and its glycosides ( $C_{16}H_{26}O_8$  and  $C_{21}H_{34}O_{12}$ ) are the precursors of wine lactone, which contributes to the aroma of aged Gewurztraminer wines (Ilc et al., 2017). Finally, furanic linalool oxide glycosides have been characterized in Gewurztraminer wines (Baltenweck et al., 2000), with molecular formula identical to those of the major compounds linalool oxide-Glc ( $C_{16}H_{28}O_7$ ) and linalool oxide-Api-Glc ( $C_{21}H_{36}O_{11}$ ) found in the berries of the RixGw progeny (**Fig. 4**). This proof-of-concept experiment showed that QTL-driven metabolic network analysis could be used to reconstruct a pathway leading to major linalool-derived compounds found in grape and wine.

In this work, we have used ultrahigh-resolution mass spectrometry to characterize the metabolome of grape berries in a grapevine cross population. As specific genetic variations in a small number of genomic regions may impact the biosynthesis of hundreds of molecules, hybridization can considerably enlarge the chemical landscape of grape berries and open new perspectives for winemaking. Furthermore, we have used a combination of untargeted metabolomics, genetic analyses and molecular networking for pathway elucidation. Indeed,

### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

mass spectrometry-based molecular networking has proven very useful for the dereplication of complex biological samples, which still remains a challenging task (Yang et al., 2013). Here, we have added a genetic dimension to this approach, to develop a QTL-driven molecular networking strategy, which constitutes a powerful approach to decipher complex metabolic pathways.

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL



**Fig. 4. QTL-driven reconstruction of the biosynthetic pathway leading to major linalool derivatives of grape berries.**

This pathway was constructed with elemental formula impacted by QTLs co-localizing with the *DXS1* locus, based on chemical network relationships (Fig. S3) and detailed analysis of UHPLC-MS fragmentation data. Abbreviations for glycosides are the following: glucoside: Glc, glucuronide: GlcA, apiosyl-glucoside: Api-Glc, arabinosyl-glucoside: Ara-Glc, malonyl-glucoside: Mal-Glc, rhamnosyl-glucoside: Rha-Glc. Molecular structures of some intermediates are presented. For each elemental formula, the maximum FTICR-MS peak area detected among all genotypes of the RixGw progeny is indicated according to the color scale and the LOD score of the corresponding QTL on the GW map is indicated between brackets [ ].

## **Materials and Methods**

### **Plant material**

The grapevine population used in this work is a progeny from a cross between *V. vinifera* cv Riesling (Ri) clone 49 (female) and *V. vinifera* cv Gewurztraminer (Gw) clone 643 (male) (RixGw). Siblings from the RixGw cross were grafted onto the Couderc 161-49 rootstock and three plants per genotype were planted in an experimental vineyard in Bergheim (48°210 N, 7°340 E) in 2006. The veraison dates (onset of the ripening process) were calculated for each individual plot, after successive scorings, as the dates when 50% of the berries were soft (entering the ripening process).

### **Extraction of grape berry metabolites**

Berries were collected from 103 genotypes at mid-ripening in 2018. The sampling date was determined for each genotype as previously, as the date when a cumulative heat sum of 350 degree-days (calculated with maximum temperatures, base temperature = 10°C) after veraison was reached (Duchêne et al., 2012b; Duchêne et al., 2014). Berry juice was extracted using a Robot Coupe C80 juice extractor (Vincennes, France), and 12 mL of juice of each genotype was stored at -80°C after addition of 120 µL of a Na<sub>2</sub>SO<sub>3</sub> solution (1 g.L<sup>-1</sup>). Thawed berry juices were centrifugated (25 min, 16500 g at 4°C) and 5 mL of supernatant was used for the solid phase extraction (SPE), after addition of apigenin (0.25 mg.L<sup>-1</sup>), as an external standard. The supernatants were passed through a 200 mg C18 silica-bonded nonpolar SPE cartridge (Bond Elut Jr., Agilent Technologies), rinsed previously with 5 mL of methanol and 15 mL of ultrapure water. After washing with 4 mL of ultrapure water, compounds of interest were eluted with 500 µL of ethanol.

### **FTICR-MS metabolic profiling**

For FTICR analysis, 20 µL of SPE extracts were diluted in 1 mL of methanol. High-resolution mass spectra were acquired on a Bruker solariX Ion Cyclotron Resonance Fourier Transform Mass Spectrometer (FTICR-MS) (Bruker Daltonics GmbH, Bremen, Germany) equipped with a 12 Tesla superconducting magnet (Magnex Scientific Inc., Yarnton, GB) and a Apollo II ESI source (Bruker Daltonics GmbH, Bremen, Germany) operated in negative or positive ionization modes (ESI). Spectra were externally calibrated on clusters of arginine (10

### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

mg.L<sup>-1</sup> in methanol). Further internal calibration was performed for each sample by using ubiquitous molecules found in the samples, reaching mass accuracies lower than 0.1 ppm in routine day-to-day measurement (Gougeon et al., 2009; Roullier-Gall et al., 2014a; Roullier-Gall et al., 2014b). Spectra were acquired with a time domain of 4 mega words over a mass range of 92 to 1,000 *m/z* in ESI (-) and 122 to 1,000 in ESI (+). 200 scans were accumulated for each sample in both ESI modes. FTICR mass spectra were exported to peak lists with a cut-off signal-to noise ratio (S/N) of 4. Peak alignment was performed with maximum error thresholds of 1 ppm and filtered for masses occurring in a minimum of 10 % of all samples.

#### **UHPLC-MS profiling and preprocessing**

For UHPLC-MS, SPE extracts were concentrated two times using a vacuum concentrator (Concentrator Plus, Eppendorf, Hamburg, Germany) prior analysis using a Dionex Ultimate 3000 UHPLC system (Thermo Fisher Scientific, San Jose, CA, USA) equipped with a diode array detector (DAD). The chromatographic separation was performed on a Nucleodur HTec column (150 x 2 mm, 1.8 µm particle size; Macherey-Nagel, Düren, Germany) maintained at 30°C. The mobile phase consisted of acetonitrile/formic acid (0.1%, v/v) (eluant A) and water/formic acid (0.1%, v/v) (eluant B) at a flow rate of 0.25 mL.min<sup>-1</sup>. Mass spectrometry-grade acetonitrile and formic acid were supplied by Thermo Fisher. The gradient elution program was as follows: 0–4 min, 80–70% B; 4–5 min, 70–50% B; 5–6.5 min, 50% B; 6.5–8.5 min 50–0% B; 8.5–10 min 0% B. The sample volume injected was 1 µL. The liquid chromatography system was coupled to an Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with an electrospray ionization source operating in positive and negative modes. Parameters were set at 360°C for ion transfer capillary temperature and 3400 V for needle voltages. Nebulization with nitrogen sheath gas and auxiliary gas were maintained at 40 and 5 arbitrary units, respectively. The spectra were acquired within the *m/z* mass range of 95–1200 atomic mass units (a.m.u.), using a resolution of 50 000 at *m/z* 200 a.m.u. The system was calibrated internally in positive mode using dibutyl phthalate as lock mass (*m/z* 279.1591), giving a mass accuracy below 1 ppm. In negative mode, the system was calibrated externally using a negative ion calibration solution (Pierce, Thermo Fisher Scientific), giving a mass accuracy below 5 ppm. The instruments were controlled using the Xcalibur software and data was processed using the XCMS software package (Smith et al., 2006). Raw data were converted to the mzXML format using MSconvert before analysis. Settings of the xcmsSet

### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

function of XCMS were as follows: method = "centWave", ppm = 2, noise = 50,000, mzdiff = 0.001, prefilter = c(5,15,000), snthresh = 6, peakwidth = c(6,35). Peaks were aligned using the obiwrap function using the followings settings of the function group.density: bw = 10, mzwid = 0.0025. Ion identifiers were generated automatically by XCMS as MxxxTyyy, where xxx is the m/z and yyy the retention time in seconds.

#### **Metabolite annotation and networks**

FTICR-MS exact masses were subjected to the Netcalc software tool to obtain chemical formulas (Tziotis et al., 2011). The formulas were validated by setting sensible chemical constraints (n rule; O/C ratio  $\leq 1$ ; H/C ratio  $\leq 2nC_2$ ; element counts: C  $\leq 100$ , H  $\leq 200$ , O  $\leq 80$ , N  $\leq 3$ , S  $\leq 3$ , and P  $\leq 1$ ). In order to facilitate the interpretation of elemental formulas attributed to m/z values, formulas were next represented using two-dimensional van Krevelen (VK) diagrams. These diagrams display the hydrogen/carbon (H/C) versus oxygen/carbon (O/C) ratios of annotated elemental formulas and provide a commonly used qualitative description of the molecular complexity of biological samples (Tziotis et al., 2011; Roullier-Gall et al., 2014b). The m/z distribution according to their elemental compositions (CHO, CHOS, CHON, CHONS, CHOP, CHONP, CHONSP) was also represented. Network analysis was performed from the complete FTICR sample set data as previously described (Tziotis et al., 2011). Nodes represent metabolite candidates and their associated elementary composition and edges represent chemical reactions. Matching of UHPLC-MS and FTICR-MS data was performed using a custom R in-house script, taking possible adducts into account (Forcisi et al., 2015).

#### **Genetic markers and QTL analysis**

QTL detection was performed on a consensus genetic map used previously for the detection of QTLs for developmental stages (Duchêne et al., 2012a) and for (E)-8-carboxylinalool content (Ilc et al., 2017). Two genetic markers were added to the initial map. VVsc271 is a simple sequence repeat (SSR) on scaffold 271 in the PN40024 129 reference genome (<http://www.genoscope.cns.fr/cgi-bin/ggb/vitis/12X/gbrowse/vitis>). Primers used to amplify this region were VVsc271f : 5'-CTTTCATGACAGATTGAGATTTACAACAAG-3' and VVsc271r : 5'-TAGCTCTTATGACAACCTTGAAAATGGAGCC-3'. Scaffold 271 was chosen because several terpene synthases were annotated in this region (Martin et al., 2010). The DXS1 gene (GSVIVT00017832001) was mapped by 'High Resolution Melting Assay'

### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

(Wittwer et al., 2003). The final map was built with 157 simple sequence repeats markers (SSR) and a SNP for the DXS1 gene and covers 1131.1 cM along the 19 chromosomes of the grapevine genome. SNP Illumina 18k chip were used to obtain two SNP parental maps: a Riesling map and a Gewurztraminer for the detection of quantitative trait loci (QTL) (Duchêne et al., 2019).

QTL detection was performed on the three genetic maps with the R/qtl software (Broman et al., 2003) using the multiple imputation method ('draws' = 128) and the one-dimension scan command 'scanone'. LOD significances were ensured with permutation tests (1000 permutations). QTL models were constructed step-by-step after the refinement of the QTL position (refineqtl), the search for supplementary QTLs (addqtl). The LOD score and the percentage of variance explained by a QTL in a QTL model were assessed with analysis of variance using type III sums of squares (fitqtl). Confidence intervals were calculated as Bayesian credible intervals (bayesesint) with a probability of coverage of 0.95. The physical positions of our confidence intervals were assessed according to the physical positions of the genetic markers in the genome.



## Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

### References

- Ali K, Maltese F, Choi YH, Verpoorte R** (2010) Metabolic constituents of grapevine and grape-derived products. *Phytochem Rev* **9**: 357–378
- Baltenweck R, Trendel J, Albrecht P, Schaeffer A** (2000) Glycosides and phenylpropanoid glycerol in *Vitis vinifera* Cv. Gewurztraminer wine. *J Agric Food Chem* **48**: 6178–6182
- Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Stella Grando M** (2009) The 1-deoxy-D: -xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theor Appl Genet* **118**: 653–669
- Broman KW, Wu H, Sen S, Churchill GA** (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890
- Duchêne E, Butterlin G, Claudel P, Dumas V, Jaegli N, Merdinoglu D** (2009) A grapevine (*Vitis vinifera* L.) deoxy-D: -xylulose synthase gene colocalizes with a major quantitative trait loci for terpenol content. *Theor Appl Genet* **118**: 541–552
- Duchêne E, Butterlin G, Dumas V, Merdinoglu D** (2012a) Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages. *Theor Appl Genet* **124**: 623–635
- Duchêne E, Dumas V, Butterlin G, Jaegli N, Rustenholz C, Chauveau A, Bérard A, Le Paslier M-C, Gaillard I, Merdinoglu D** (2019) Genetic variations of acidity in grape berries are controlled by the interplay between organic acids and potassium. *Theor Appl Genet*, *in press*
- Duchêne E, Dumas V, Jaegli N, Merdinoglu D** (2012b) Deciphering the ability of different grapevine genotypes to accumulate sugar in berries. *Aust J Grape Wine Res* **18**: 319–328
- Duchêne E, Dumas V, Jaegli N, Merdinoglu D** (2014) Genetic variability of descriptors for grapevine berry acidity in Riesling, Gewürztraminer and their progeny. *Aust J Grape Wine Res* **20**: 91–99
- Dunlevy JD, Dennis EG, Soole KL, Perkins MV, Davies C, Boss PK** (2013) A methyltransferase essential for the methoxy-pyrazine-derived flavour of wine. *Plant J*. doi: 10.1111/tj.12224
- Emanuelli F, Battilana J, Costantini L, Le Cunff L, Boursiquot J-M, This P, Grando MS** (2010) A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol* **10**: 1–17
- Estévez JM, Cantero A, Reindl A, Reichler S, León P** (2001) 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J Biol Chem* **276**: 22901–22909
- Feechan A, Anderson C, Torregrosa L, Jermakow A, Mestre P, Wiedemann-Merdinoglu S, Merdinoglu D, Walker AR, Cadle-Davidson L, Reisch B, et al** (2013) Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia rotundifolia* identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. *Plant J* **76**: 661–674
- Forcisi S, Moritz F, Kanawati B, Tziotis D, Lehmann R, Schmitt-Kopplin P** (2013) Liquid chromatography-mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. *J Chromatogr A* **1292**: 51–65
- Forcisi S, Moritz F, Lucio M, Lehmann R, Stefan N, Schmitt-Kopplin P** (2015) Solutions for low and high accuracy mass spectrometric data matching: a data-driven annotation strategy in nontargeted metabolomics. *Anal Chem* **87**: 8917–8924
- Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet J-M, Cheynier V, This P** (2009) Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp. *sativa*) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. *Genetics* **183**: 1127–1139

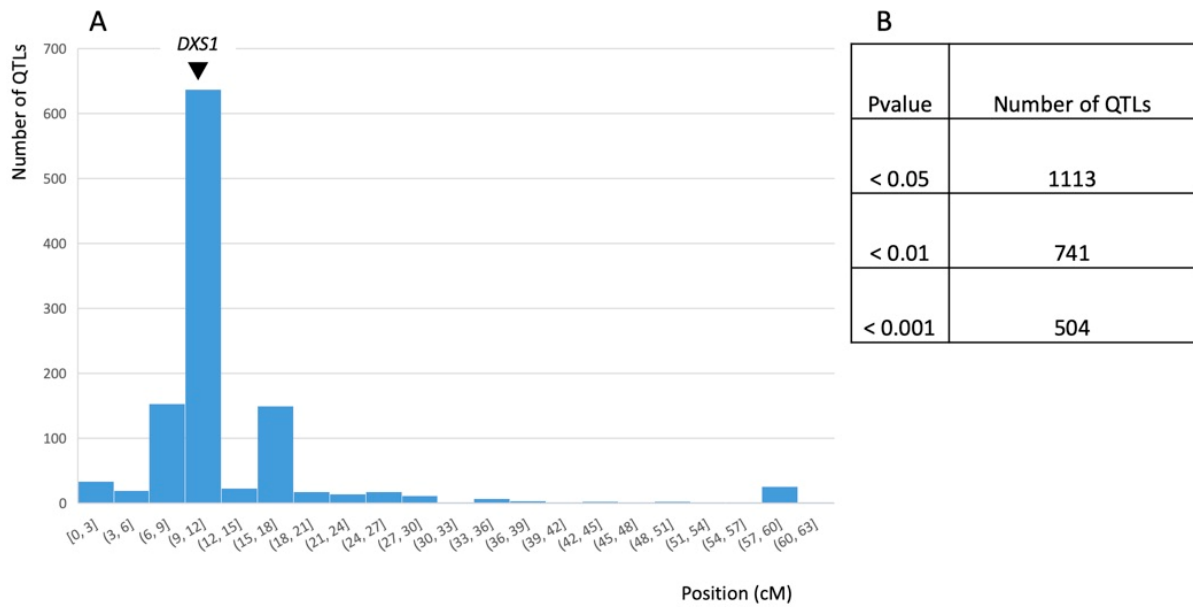
### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

- Gougeon RD, Lucio M, Frommberger M, Peyron D, Chassagne D, Alexandre H, Feuillat F, Voilley A, Cayot P, Gebefügi I, et al** (2009) The chemodiversity of wines can reveal a metabo-geography expression of cooperage oak wood. *Proc Natl Acad Sci USA* **106**: 9174–9179
- Guillaumie S, Ilg A, Réty S, Brette M, Trossat-Magnin C, Decroocq S, Léon C, Keime C, Ye T, Baltenweck R, et al** (2013) Genetic analysis of the biosynthesis of 2-methoxy-3-isobutylpyrazine, a major grape-derived aroma compound impacting wine quality. *Plant Physiol* **162**: 604–615
- Huang Y-F, Doligez A, Fournier-Level A, Le Cunff L, Bertrand Y, Canaguier A, Morel C, Miralles V, Veran F, Souquet J-M, et al** (2012) Dissecting genetic architecture of grape proanthocyanidin composition through quantitative trait locus mapping. *BMC Plant Biol* **12**: 30–19
- Hugueney P, Duchêne E, Merdinoglu D** (2012) International patent WO 2012\_052171.
- Ilc T, Arista G, Tavares R, Navrot N, Duchêne E, Velt A, Choulet F, Paux E, Fischer M, Nelson DR, et al** (2018) Annotation, classification, genomic organization and expression of the *Vitis vinifera* CYPome. *PLoS ONE* **13**: e0199902
- Ilc T, Halter D, Miesch L, Lauvoisard F, Kriegshauser L, Ilg A, Baltenweck R, Hugueney P, Werck-Reichhart D, Duchêne E, et al** (2017) A grapevine cytochrome P450 generates the precursor of wine lactone, a key odorant in wine. *New Phytol* **213**: 264–274
- Ilc T, Parage C, Boachon B, Navrot N, Werck-Reichhart D** (2016) Monoterpenol Oxidative Metabolism: Role in Plant Adaptation and Potential Applications. *Front Plant Sci* **7**: 509
- 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
- Liger-Belair G, Cilindre C, Gougeon RD, Lucio M, Gebefügi I, Jeandet P, Schmitt-Kopplin P** (2009) Unraveling different chemical fingerprints between a champagne wine and its aerosols. *Proc Natl Acad Sci USA* **106**: 16545–16549
- Lijavetzky D, Cabezas JA, Ibáñez A, Rodríguez V, Martínez-Zapater JM** (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* **8**: 424
- Lin J, Massonnet M, Cantu D** (2019) The genetic basis of grape and wine aroma. *Horticulture Research* 1–24
- Marinos VA, Tate ME, Williams PJ** (1994) Protocol for FAB MS/MS characterization of terpene disaccharides of wine. *J Agric Food Chem* **42**: 2486–2492
- Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, Lund ST, Bohlmann J** (2010) Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol* **10**: 226
- McGovern PE, Glusker DL, Exner LJ, Voigt MM** (1996) Neolithic resinated wine. *Nature* **381**: 480–481
- Merdinoglu D, Schneider C, Prado E, Wiedemann-Merdinoglu S, Mestre P** (2018) Breeding for durable resistance to downy and powdery mildew in grapevine. *OENO One* **52**: 203–209
- Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia J-M, Ware D, et al** (2011) Genetic structure and domestication history of the grape. *Proc Natl Acad Sci USA* **108**: 3530–3535
- Pelsy F, Hocquigny S, Moncada X, Barbeau G, Forget D, Hinrichsen P, Merdinoglu D** (2010) An extensive study of the genetic diversity within seven French wine grape variety collections. *Theor Appl Genet* **120**: 1219–1231

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

- Polášková P, Herszage J, Ebeler SE** (2008) Wine flavor: chemistry in a glass. *Chem Soc Rev* **37**: 2478–2489
- Poni S, Gatti M, Palliotti A, Dai Z, Duchêne E, Truong T-T, Ferrara G, Matarrese AMS, Gallotta A, Bellincontro A, et al** (2018) Grapevine quality\_ A multiple choice issue. **234**: 445–462
- Roullier-Gall C, Boutegrabet L, Gougeon RD, Schmitt-Kopplin P** (2014a) A grape and wine chemodiversity comparison of different appellations in Burgundy: vintage vs terroir effects. *Food Chem* **152**: 100–107
- Roullier-Gall C, Lucio M, Noret L, Schmitt-Kopplin P, Gougeon RD** (2014b) How subtle is the "terroir" effect? Chemistry-related signatures of two "climats de Bourgogne". *PLoS ONE* **9**: e97615
- Roullier-Gall C, Witting M, Gougeon RD, Schmitt-Kopplin P** (2014c) High precision mass measurements for wine metabolomics. *Front Chem* **2**: 102
- Schwab W, Wüst M** (2015) Understanding the Constitutive and Induced Biosynthesis of Mono- and Sesquiterpenes in Grapes (*Vitis vinifera*): A Key to Unlocking the Biochemical Secrets of Unique Grape Aroma Profiles. *J Agric Food Chem* **63**: 10591–10603
- Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G** (2006) XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal Chem* **78**: 779–787
- Strauss CR, Wilson B, Williams PJ** (1988) Novel monoterpene diols and diol glycosides in *Vitis vinifera* grapes. *J Agric Food Chem* **36**: 569–573
- This P, Lacombe T, Thomas MR** (2006) Historical origins and genetic diversity of wine grapes. *Trends Genet* **22**: 511–519
- Tziotis D, Hertkorn N, Schmitt-Kopplin P** (2011) Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur J Mass Spectrom (Chichester)* **17**: 415–421
- van Leeuwen C, Friant P, Choné X, Tregoat O, Koundouras S, Dubourdieu D** (2004) Influence of Climate, Soil, and Cultivar on Terroir. *Am J Enology Vitic* **55**: 207–217
- Williams PJ, Strauss CR, Wilson B** (1980) Hydroxylated linalool derivatives as precursors of volatile monoterpenes of muscat grapes. *J Agric Food Chem* **28**: 766–771
- Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ** (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem* **49**: 853–860
- Wright LP, Rohwer JM, Ghirardo A, Hammerbacher A, Ortiz-Alcaide M, Raguschke B, Schnitzler J-P, Gershenzon J, Phillips MA** (2014) Deoxyxylulose 5-Phosphate Synthase Controls Flux through the Methylerythritol 4-Phosphate Pathway in *Arabidopsis*. *Plant Physiol* **165**: 1488–1504
- Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, de Felicio R, Fenner A, et al** (2013) Molecular networking as a dereplication strategy. *J Nat Prod* **76**: 1686–1699

Supporting information



**Fig. S1. Distribution of QTLs on chromosome 5**

**A.** Distribution of the QTLs on chromosome 5 ( $p < 0.05$ ), with the GW map. The position of the *DXS1* gene (9.9 cM) is indicated. **B.** Number of significant QTLs associated to chromosome 5 as a function of the P value.

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

ID_FT	m/z	Netcalc formula	Max peak area	Peak number_LC	QTL LOD
22351	501.3058307	C26H44O9	46077344	0	15.39
23145	513.2694358	C26H40O10	45853200	0	15.31
24324	531.2801093	C26H42O11	378144480	0	14.35
23401	517.3008249	C26H44O10	372240544	2	13.62
18672	449.2381414	C21H36O10	2212981248	3	12.93
23544	519.3164323	C26H46O10	160101360	1	12.62
12068	351.2013719	C16H30O8	700371072	2	12.61
24612	535.3114049	C26H46O11	426904832	2	12.27
14653	387.2377293	C20H34O7	105190128	0	12.06
24469	533.2957555	C26H44O11	865602048	0	12.04
13317	369.2271347	C20H32O6	60268640	0	11.73
13485	371.2427876	C20H34O6	132604656	1	11.52
22200	499.2902237	C26H42O9	41446800	2	11.5
14785	389.2533845	C20H36O7	196879872	0	11.31
23269	515.2851204	C26H42O10	311684448	1	10.53
25518	549.2906383	C26H44O12	364913216	1	10.53
24186	529.26446	C26H40O11	24192318	0	10.26
25658	551.3063271	C26H46O12	558114112	0	10.15
19719	465.2330551	C21H36O11	23102722048	4	10.14
12360	355.2478927	C20H34O5	112840224	2	9.97
15864	405.2482914	C20H36O8	60188756	0	9.92
18926	453.2846614	C25H40O7	16810264	0	9.82
12210	353.2322795	C20H32O5	161467248	1	9.45
21155	485.2592871	C21H40O12	189970352	1	8.81
20990	483.2435823	C21H38O12	1272427392	1	8.72
22029	497.2229998	C21H36O13	595054144	1	8.67
14499	385.2221322	C20H32O7	34981532	0	8.44
22999	511.238564	C22H38O13	332129984	2	8.29
20704	479.2487075	C22H38O11	273132416	2	8.17
19862	467.2487565	C21H38O11	1232572928	3	8.11
22045	497.2745675	C26H40O9	45002948	2	8.06
22496	503.3215345	C26H46O9	37726452	0	7.87
10029	319.226735	C20H30O3	63740584	3	7.8
22867	509.2228984	C22H36O13	658760512	1	7.61
19586	463.2173347	C21H34O11	1259309312	3	7.61
752	153.1273968	C10H16O	158054736	8	7.51
8668	299.2005889	C20H26O2	44449036	2	7.43
21889	495.2588349	C26H38O9	22678738	0	7.38

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

12506	357.2635068	C20H36O5	45994192	1	7.35
17715	433.2432075	C21H36O9	2087667712	2	7.24
12076	351.2166309	C20H30O5	68372656	2	7.23
9881	317.2111473	C20H28O3	86529952	2	7.23
9727	315.1955179	C20H26O3	27301012	1	7.17
26662	567.3012887	C26H46O13	81312704	0	7.11
20012	469.2643313	C21H40O11	128718408	1	6.94
10891	333.2060373	C20H28O4	51015532	3	6.77
9575	313.1128904	C11H20O10	52364736	2	6.75
8803	301.216215	C20H28O2	31539176	2	6.72
13666	373.2584495	C20H36O6	163348096	1	6.69
20707	479.2639726	C26H38O8	29032494	1	6.58
13177	367.2115371	C20H30O6	28613848	0	6.43
11945	349.2009914	C20H28O5	25229662	1	6.38
10747	331.190384	C20H26O4	14783564	0	6.26
11179	337.2373142	C20H32O4	107754808	3	6.25
11044	335.2216912	C20H30O4	165163600	8	6.25
20846	481.2280533	C21H36O12	1782558592	3	6.19
1309	171.137956	C10H18O2	446455232	5	6.18
15315	397.1704964	C16H28O11	490926816	1	6.16
8939	303.2319092	C20H30O2	21207024	1	6.15
14632	387.1649784	C18H26O9	41812288	5	6.04
15483	399.1861059	C16H30O11	61297592	1	6.03
4932	247.1176216	C11H18O6	13165014	1	6.0
7827	287.1852828	C15H26O5	48431644	6	5.99
7418	281.1899991	C20H24O	7607771	0	5.95
21734	493.2431808	C26H36O9	7704776	0	5.9
8521	297.1849048	C20H24O2	8751672	1	5.82
20860	481.2796088	C26H40O8	17052618	0	5.62
13759	375.1286326	C16H22O10	17302054	1	5.6
24049	527.2334393	C22H38O14	157810816	0	5.54
10146	321.2424347	C20H32O3	29949362	1	5.52
1060	163.0753545	C10H10O2	17423382	7	5.36
7566	283.2056366	C20H26O	6893483	0	5.33
10885	333.1907622	C16H28O7	2171522048	5	5.3
9599	313.1798123	C20H24O3	7828042	1	5.14
11511	343.1023829	C15H18O9	43850012	1	5.11
7684	285.2213258	C20H28O	10571397	0	5.0
23002	511.2537918	C26H38O10	11200744	0	4.92

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

13028	365.1806516	C16H28O9	532613728	1	4.87
20582	477.2482771	C26H36O8	7005323	0	4.74
20693	479.2122759	C21H34O12	683649920	2	4.66
15731	403.1596594	C18H26O10	92480336	1	4.62
7830	287.2005744	C19H26O2	5245220	0	4.61
14486	385.1703744	C15H28O11	1013921280	1	4.6
11427	341.2686418	C20H36O4	17446572	1	4.55
11310	339.2529629	C20H34O4	43653144	2	4.54
10500	327.107403	C15H18O8	677299520	2	4.54
8224	293.122994	C12H20O8	65358048	1	4.53
11038	335.2064011	C16H30O7	56114644	1	4.42
7962	289.216232	C19H28O2	5260175	0	4.38
21885	495.2435796	C22H38O12	23427066	0	4.38
6471	267.1438482	C11H22O7	186930960	1	4.29
6497	267.2107217	C20H26	3242473	0	4.26
8760	301.0918098	C13H16O8	14893882	0	4.26
16764	419.1911547	C19H30O10	7988525056	4	4.16
10262	323.2581145	C20H34O3	28670958	1	4.14
10349	325.0918473	C15H16O8	34344688	0	4.1
1898	189.148532	C10H20O3	1825182848	2	4.1
21003	483.2800425	C22H42O11	13237950	0	4.09
10716	331.102342	C14H18O9	120183120	2	4.06
3464	223.1176289	C9H18O6	37058004	1	4.02
3813	229.1070656	C11H16O5	35719980	1	3.94
17911	437.1442382	C21H24O10	35848252	3	3.93
11645	345.118052	C15H20O9	18923084	0	3.91
13030	365.1958593	C20H28O6	9007537	0	3.83
19463	461.2533591	C26H36O7	5885748	0	3.83
11793	347.1700146	C16H26O8	1279191424	2	3.79
5031	249.1121337	C14H16O4	37794132	5	3.78
10489	327.0710966	C14H14O9	601540352	0	3.78
18797	451.2537842	C21H38O10	11817545	1	3.76
1737	185.1172201	C10H16O3	67912392	3	3.75
12607	359.1336493	C16H22O9	20153868	1	3.72
821	155.14304	C10H18O	5386143	5	3.69
12061	351.1802156	C19H26O6	6413037	1	3.67
7107	277.0918202	C11H16O8	48969352	4	3.66
15834	405.1543791	C21H24O8	19692576	0	3.65
5277	253.1070587	C13H16O5	12823588	1	3.64

### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

6972	275.0913781	C15H14O5	41992788	3	3.59
10259	323.2216913	C19H30O4	5435785	0	3.58
14056	379.1962902	C17H30O9	8024154	0	3.58
15222	395.2640033	C19H38O8	5234418	0	3.58
9875	317.1958555	C16H28O6	75014568	5	3.55
9210	307.2267756	C19H30O3	5642588	0	3.54
21006	483.2952809	C26H42O8	12064608	0	3.52
2984	213.1637716	C16H20	4166710	0	3.49
12869	363.1286707	C15H22O10	11801369	0	3.45
7960	289.2009529	C15H28O5	26178168	2	3.41
9317	309.0969129	C15H16O7	39153772	2	3.4
18894	453.1604409	C18H28O13	16691685	1	3.38
8340	295.1023877	C11H18O9	490503136	6	3.31
8621	299.0914016	C17H14O5	38847152	1	3.28
2953	213.0757531	C10H12O5	74458048	2	3.27
22181	499.2385165	C21H38O13	243900144	1	3.26
9093	305.2475354	C20H32O2	20566644	1	3.23
6729	271.1176195	C13H18O6	26960408	2	3.22
2291	199.0600971	C9H10O5	14038339	2	3.21
14764	389.1806413	C18H28O9	23603400	1	3.14
6751	271.1903737	C15H26O4	56515240	4	3.13
15847	405.1755046	C18H28O10	22382464	3	3.12
18765	451.1446707	C18H26O13	15591674	1	3.11
2591	205.1434486	C10H20O4	736507392	1	3.09
5304	253.1798382	C15H24O3	83021920	3	3.08
16371	413.2016976	C17H32O11	16130700	2	3.06
7376	281.1019621	C14H16O6	18346130	1	2.98
6752	271.2056652	C19H26O	6186064	0	2.96
6840	273.0757316	C15H12O5	689855616	3	2.96
1102	165.0546253	C9H8O3	116599456	4	2.92
16225	411.1133205	C15H22O13	7614295	0	2.83
2742	209.0444461	C10H8O5	11263934	1	2.81

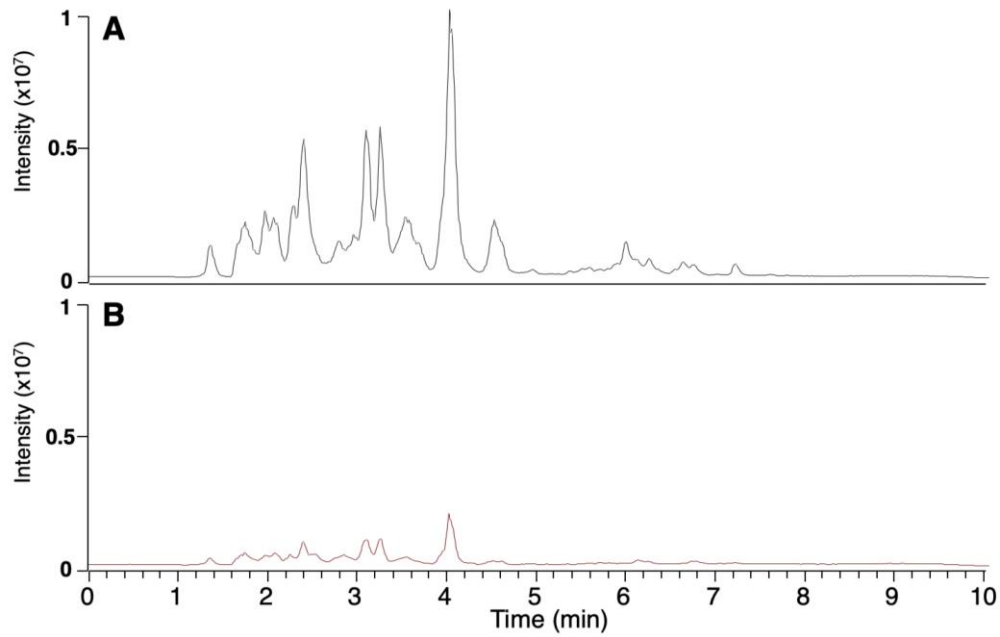
**Table S1. List of elemental compositions associated to significant QTLs ( $p < 0.05$ ) co-localizing with the *DXSI* locus in the RixGw progeny, with the Gw map.**

The  $m/z$  of detected ions and the corresponding neutral elemental compositions are indicated, as well as the LOD score of the associated QTL and the maximum peak area detected among all genotypes of the RixGw progeny. Number of mass peaks detected for the same  $m/z$  in UHPLC-MS analyses are indicated.



### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

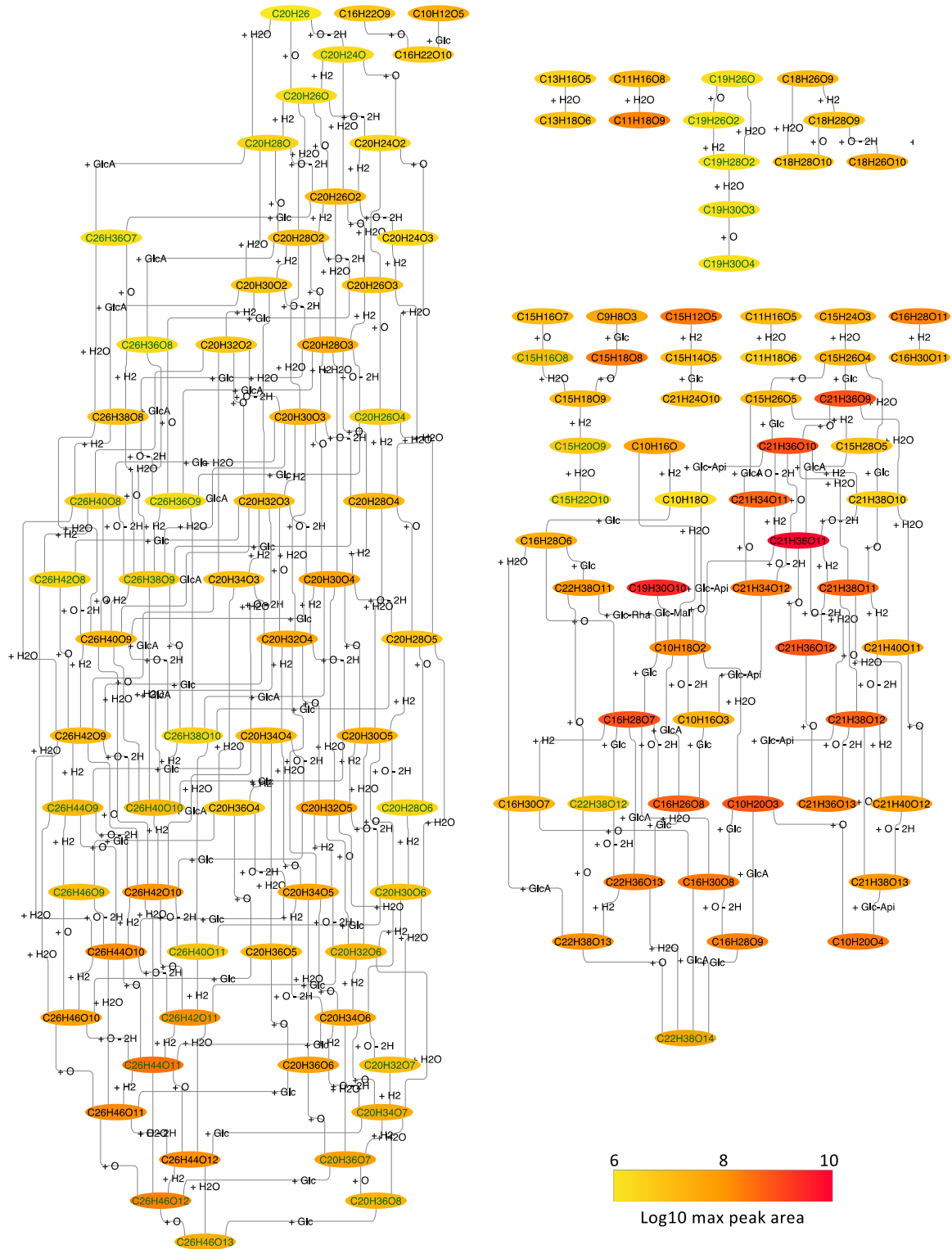
### Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL



**Fig. S2. Major impact of the *DXS1* locus on terpenoid profile in grape berries**

UHPLC-MS analysis of grape berry extracts presenting the sum of the extracted ion chromatograms (EIC) for the characteristic ions of major terpenoid compounds ( $m/z \pm 2$  ppm) 137.1325 + 153.1274 + 167.1066, corresponding respectively to  $[C_{10}H_{16} + H]^+$ ,  $[C_{10}H_{16}O + H]^+$  and  $[C_{10}H_{14}O_2 + H]^+$ . **A.** Typical profile obtained with berry extracts from an aromatic genotype (74E) harboring a favorable *DXS1* allele from Gw; **B.** Typical profile obtained with berry extracts from a non-aromatic genotype (216E). For easier comparison, both profiles are presented at the same intensity scale ( $1 \cdot 10^7$  arbitrary units).

# Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL



**Fig. S3. Molecular network of CHO elemental formulas impacted by the *DXS1* locus**  
 Elemental formula indicated black are detected both in FTICR-MS and UHPLC-MS. Elemental formula indicated green are detected in FTICR-MS. The chemical relationships between the elemental formula are the following: addition of 2 hydrogen atoms (+H2), addition of oxygen (+O), addition of water (+H2O), addition of oxygen and loss of 2 hydrogen atoms (+O-2H), addition of glucose (+Glc), addition of glucuronic acid (+GlcA), addition of glucose-apiose (+Glc-Api). For each elemental formula, the maximum peak area detected among all genotypes of the RixGw progeny is indicated according to the color scale.

## Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL

### 3. Discussion

L'analyse conjointe d'extraits de jus issus de la population de croisement RixGw par UHPLC-MS et par FT-ICR a permis d'allier les avantages des deux instruments : par UHPLC une séparation des isomères ainsi qu'une fragmentation des molécules, permettant ainsi d'avoir des informations sur leurs structures, et par FT-ICR une précision en masse inférieure à 0,1ppm, permettant une détermination précise de la formule brute, ainsi qu'une haute résolution supérieure à 100 000, permettant de distinguer des m/z proches. Ces données ont permis d'appréhender de manière la plus précise possible le métabolisme secondaire des baies, avec la quantification et l'identification de milliers d'ions mais surtout la mise en réseau des molécules entre elles. Les annotations précises des ions détectés en FT-ICR par le logiciel Netcalc m'ont permis d'obtenir de manière automatisée des formules brutes ainsi que les liens chimiques qui pourraient relier les molécules entre elles.

La combinaison de toutes ces données métabolomiques de haute résolution et de génotypage à haut débit a permis de confirmer les QTL que j'ai pu identifier précédemment dans le chapitre 2. L'un des hotspots les plus importants se situe sur le chromosome 5 et est relié au QTL de la DXS. 129 ions correspondant à des molécules de la famille des monoterpénols, des sesquiterpènes et des diterpènes.

La combinaison des données de type QTL avec les données de liens chimiques entre molécules permet d'établir un nouveau concept : l'étude des réseaux entre des molécules ayant des voies de biosynthèse communes et des liens chimiques entre elles. Ceci permet d'avoir une vision novatrice du fonctionnement de la biosynthèse de composés d'intérêt et mais aussi des conséquences des variations génétiques au niveau de l'ADN sur le paysage métabolique des raisins.

### **Chapitre 3 Analyse en réseaux du métabolome de la vigne par la combinaison d'analyses par FT-ICR et UHPLC-MS à la lumière des QTL**

**Chapitre 4 Identification et caractérisation  
fonctionnelle de gènes candidats**



### 1. Introduction

Les analyses QTL réalisées sur les jeux de données de type GC-MS, UHPLC-MS et FT-ICR m'ont permis d'identifier des régions génomiques gouvernant la biosynthèse d'un très grand nombre de molécules du métabolisme secondaire de la vigne. La recherche des gènes candidats liés à ces QTL et pouvant expliquer la variabilité en métabolites secondaires, dont notamment les arômes et les précurseurs d'arômes, est la suite logique de ces analyses. Le passage de la position génétique (cM) à la position physique (pb) sur le génome de référence PN40024 se fait grâce à l'information de position physique des marqueurs génétiques utilisés pour le génotypage. Tous les QTL n'ont pas pu être exploités du fait de leur grand nombre. Le choix du QTL à étudier de manière plus approfondie a été fait en fonction de plusieurs critères :

- L'impact potentiel de la molécule associée au QTL pour les arômes des vins,
- La vraisemblance de l'existence de ce QTL (LOD score, p-value).

Après l'analyse QTL, une identification des gènes présents sur l'intervalle de confiance du QTL à partir de l'annotation des gènes du génome de référence PN40024 a été réalisée. Les gènes candidats ont été choisis sur la base de leur fonction putative et de leur profil d'expression dans les baies de Ri et Gw, analysé grâce aux données transcriptomiques de RNA-seq. Des gènes candidats pouvant participer à la biosynthèse de monoterpénols ont été identifiés dans le Chapitre 1.

Ce chapitre détaillera des gènes candidats identifiés pour deux QTL :

- Un QTL sur le chromosome 10, sur la carte SSR, relié à la synthèse de linalol et de ses dérivés,
- Un QTL sur le chromosome 2 lié à la synthèse de vanilline libre quantifiée en GC-MS. Malgré le fait que ce QTL n'ait été retrouvé qu'en 2014, la recherche d'un gène candidat pour ce QTL a été réalisée car la voie de biosynthèse de la vanilline chez les plantes reste controversée, même chez *Vanilla planifolia* où elle est beaucoup étudiée (Gallage et al., 2014; Yang et al., 2017). Ce gène candidat a été caractérisé fonctionnellement chez *Nicotiana benthamiana* et *in vitro*.

La recherche de gènes candidats a été possible grâce aux données bioinformatiques de séquençage génomiques et RNA-seq du Ri et du Gw, dont les analyses ont été réalisées par Camille Rustenholz, Amandine Velt et Sophie Blanc. J'ai réalisé moi-même la caractérisation de la CAMT avec l'aide de Lauriane Renault.



## 2. Identification des gènes candidats liés aux QTL d'intérêt et caractérisation fonctionnelle

### 2.1. Gènes candidats sur le chromosome 10 liés au linalol et à ses dérivés

Les analyses QTL par GC-MS et UHPLC-MS ont révélé un QTL lié aux variations en contenu de différents dérivés de linalol situé sur le scaffold 271 du chromosome 10 (Tableau 49, Chapitre 1).

Chromosome	Analyse	Molécule	Fraction	Carte génétique	LOD	%Var.	Intervalle de confiance (cM)
Chr10	GC-MS	(E)-2,6-Dimethyl-3,7-octadiene-2,6-diol	Libre	SNP_Ri	2.6-2.9	9.6-10.6	0-12.5
			Liée	SNP_Ri	4.3-4.9	11.8-12.3	0-10.7
		SSR		5.6-6.2	19.3-19.7	0-12	
		(Z) 8-Hydroxylinalol	Liée	SNP_Ri	3.7-4.8	13.7-17.2	0-10.7
				SSR	5.2-5.6	11.5-12.9	0-10
		(Z)-2,6-Dimethyl-3,7-octadiene-2,6-diol	Liée	SNP_Ri	2.8-5.9	10.6-14	1.8-10.7
		1,7-Octadien-3-ol, 2,6-dimethyl-	Liée	SNP_Ri	4.4-6.2	15.9-21.5	0.9-8
				SSR	4.9-7.2	12.9-17.8	0-14
		1,7-Octadiene-3,6-diol, 2,6-dimethyl-	Libre	SNP_Ri	3.3-3.5	12.2-12.7	0-10.7
			SSR	4.3-5.7	8.9-12.6	0-6	
		Liée	SNP_Ri	2.5-3.1	9.5-11.3	0-12.5	
			SSR	2.6-2.8	6.5-10.6	0-12.5	
		Hotrienol	Liée	SNP_Ri	2.6-2.8	6.5-10.6	0-12.5
				SSR	2.6-2.8	6.5-10.6	0-12.5
		Linalol	Libre	SNP_Ri	3-4.7	11.1-17	0-10.7
				SSR	4.3-6.4	13.5-20.6	0-9.1
			Liée	SNP_Ri	4.6-6.6	16.7-19.6	0-8.9
				SSR	8.4-8.6	16.7-16.8	0-6
	Linalol oxyde (pyranoid), trans-	Libre	SNP_Ri	4.4-5	13.2-17.8	2.7-10.7	
			SSR	5.2-5.3	18.4-18.8	févr-14	
Liée		SNP_Ri	3.6-4.1	13.2-15	0-8.9		
		SSR	5.5-5.8	17.2-20.3	0-10		
Linalol oxyde, (furanoid), trans-	Liée	SNP_Ri	4.6-4.6	12.6-16.5	0-10.7		
		SSR	5.5-5.9	19.6-20.8	0-10		
UHPLC-MS		Apiosylglucoside de linalol oxyde	/	SSR	6.3-6.9	6-7.4	4-10

Tableau 49 : QTL identifiés sur le chromosome 10 liés au linalol et à ses dérivés détectés dans les baies de la population RixGw par GC-MS et UHPLC-MS ESI (+).

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats

### 2.1.1. Identification de gènes candidats sur les intervalles de confiance

Deux intervalles de confiance différents avaient été identifiés dans le chapitre 1 :

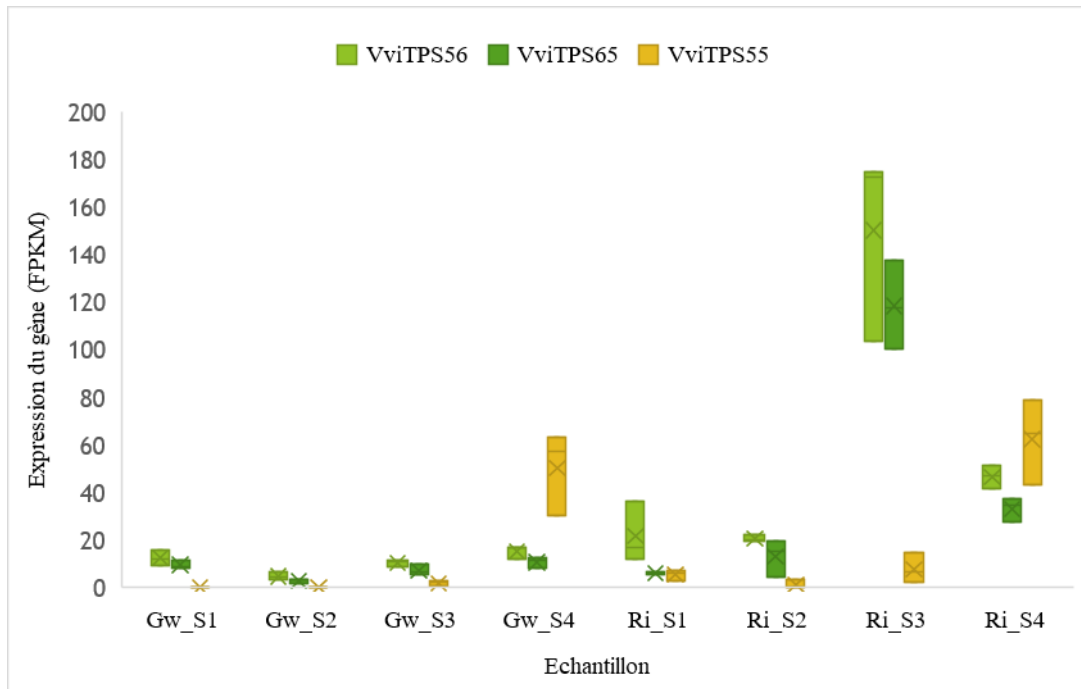
- Un premier intervalle de confiance avec la carte SNP\_Ri, où aucun gène candidat n'avait pu être identifié,
- Un deuxième intervalle de confiance avec la carte SSR incluant le marqueur génétique VVsc271, lié au scaffold 271. Dans cette intervalle quatre gènes candidats codant pour des TPS ont été identifiés par Martin et al. (2010). Ces gènes peuvent coder pour une linalol synthase qui pourrait être responsable de la biosynthèse de linalol dans les baies de la population RixGw (Tableau 50). Parmi ces TPS, la TPS56 a déjà été caractérisée dans les cépages de type Cabernet Sauvignon, Pinot Noir (Martin et al., 2010) et le Riesling (Zhu et al., 2014).

Intervalle de confiance	Annotation TPS d'après Martin et al., 2010	Position physique du gène (pb)	Fonction d'après Martin et al., 2010
Chr10 : Scaffold 271	<i>VviTPS55</i>	chrUkn:8443638..8446031	<i>Linalol synthase</i>
	<i>VviTPS144</i>	chrUkn:8505888..8506873	<i>Linalol synthase</i>
	<i>VviTPS65</i>	chrUkn:8510745..8513032	<i>Linalol synthase</i>
	<i>VviTPS56</i>	chrUkn:8526111..8528398	<i>Linalol synthase</i>

Tableau 50 : Terpénols synthases (TPS) annotées sur le scaffold 271 du chromosome 10.

### 2.1.2. Analyse de l'expression génique par RNA-seq

Parmi les 4 TPS situées sur le scaffold 271 identifiées précédemment, la TPS144 n'est pas du tout exprimée dans les baies de Ri et de Gw. Des différences significatives d'expressions sont observables entre les baies de Ri et de Gw pour les trois autres TPS. Les TPS56 et TPS65 sont les plus exprimées dans les baies de Ri, notamment dans les baies molles au cours de la véraison (Figure 55). Le plus faible contenu en linalol dans le Gw pourrait être expliqué par une différence d'activité de la TPS55.



**Figure 55 :** Expression des TPS situées sur le scaffold 271 du chromosome 10 dans les baies de Ri et de Gw prélevées en 2016. S1 : Baies vertes, S2 : Baies vertes au cours de la véraison, S3 : Baies molles au cours de la véraison, S4 : Baies à mi-maturité.

### 2.1.3. Analyse des allèles des transcrits grâce aux données RNA-seq

L'observation des reads RNAseq alignés sur PN40024 de la TPS 55, TPS 56 et TPS 65 montre que le Ri et le Gw sont hétérozygotes pour ces 3 gènes. Des SNP différentiels entre Ri et Gw sont détectés dans les exons des 3 gènes candidats mais aussi au niveau de la séquence d'adressage de l'ARNm.

Ces différences nucléotidiques dans les transcrits peuvent être responsables de la production d'isoformes de protéines pouvant avoir des activités enzymatiques différentes. Différents allèles correspondant aux TPS 55, 56 et 65 ont été clonés et leur caractérisation fonctionnelle est actuellement en cours.

## 2.2. Gènes candidats sur le chromosome 2 liés au contenu en vanilline dans les baies

Un QTL lié au contenu en vanilline libre des baies de la population RixGw prélevées en 2014 a été détecté avec les données GC-MS avec la carte génétique SSR sur le chromosome 2 à la position 6 cM. Il a un LOD score de 5.1 (p-value de 0.009) et explique 18 % de la variance observée dans la population RixGw (Tableau 51).

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats

Molécule	Carte génétique	Année	LOD	pval	% Variance	Intervalle de confiance (cM)	Intervalle de confiance (pb)
Vanilline	Gw	2014	3.6	0.006	4.3	0-40.3	236 840 - 6 301 453
	SSR	2014	5.1	0.009	18.1	0-12	2 161 587 - 4 496 744

Tableau 51 : QTL lié au contenu en vanilline libre des baies analysé par GC-MS sur le chromosome 2.

### 2.2.1. Identification de gènes candidats sur les intervalles de confiance

Dans l'intervalle de confiance associé au QTL, les protéines liées à 244 gènes annotés sur PN400274 ont été comparées aux séquences disponibles dans Swissprot/Uniprot. Deux gènes candidats codant pour des caffeic acid O-méthyltransférases (CAMT) paraissent intéressants à étudier, compte tenu de ce que l'on sait de la voie de synthèse de la vanilline (Tableau 52).

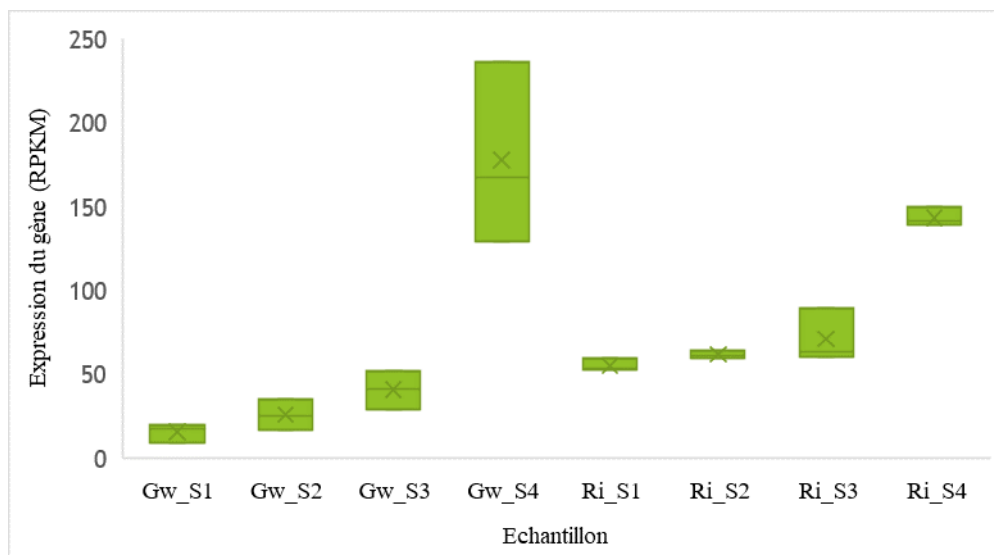
Gène	Position physique	Fonction
<i>VIT_02s0025g02920</i>	chr2:2308584..2310417	Caffeic acid 3-O-methyltransferase
<i>VIT_02s0025g02930</i>	chr2:2314343..2314782	Caffeic acid 3-O-methyltransferase

Tableau 52 : Gènes candidats liés au QTL de la vanilline sur le chromosome 2.

### 2.2.2. Analyse de l'expression génique par RNA-seq des gènes candidats

Les données RNA-seq montrent que le gène *VIT\_02s0025g02930* n'est pas exprimé dans les baies et que le gène *VIT\_02s0025g02920* est exprimé dans les baies de Riesling et de Gewurztraminer (Figure 56). Le gène *VIT\_02s0025g02920* (ou CAMT) est surtout exprimé dans les baies à mi-maturité. L'expression du gène *VIT\_02s0025g02930* ne semble pas être différentielle entre le Ri et le Gw. La différence observée en teneur en vanilline entre les baies de Ri et de Gw ne semble donc pas liée à une différence d'expression du gène *VIT\_02s0025g02920*,

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats



**Figure 56 : Expression de la CAMT dans les baies de Ri et de Gw prélevées en 2016.** S1 : Baies vertes, S2 : Baies vertes au cours de la véraison, S3 : Baies molles au cours de la véraison, S4 : Baies à mi-maturité.

### 2.2.3. Analyse des allèles du gène *VIT\_02s0025g02920* et des isoformes de la protéine associée

Les analyses de séquences nucléotidiques du gène *VIT\_02s0025g02920* (ou *CAMT*) ont révélé que le Ri et le Gw sont hétérozygotes pour ce gène. Le Ri et le Gw ont tous les deux un SNP commun à la position 2310073 sur le chromosome 2 et des SNP différentiels aux positions 2308997 pour le Gw et 2309074 pour le Ri (Tableau 53). Ces deux SNP différentiels sont intéressants car ce sont des SNP modifiant la séquence en acides aminés des protéines. Ces différences en acides aminés pourraient être responsables d'une modification d'activité enzymatique de la protéine.

Chr	SNP	Position	Référence (PN40024)	Alternatif	Localisation	Effet	Changement d'acide aminé
chr2	Gw	2308997	C	A	exon	missense	Gly → Cys
chr2	Ri	2309074	G	T	exon	missense	Pro → His
chr2	Ri	2310073	C	T	exon	missense	Arg → His
chr2	Gw	2310073	C	T	exon	missense	Arg → His

**Tableau 53 : SNP détectés dans les exons du gène de la CAMT.**

A partir des séquences des génomes du Riesling et du Gewurztraminer, j'ai pu déterminer que ces 2 cépages ont chacun 2 allèles différents de *CAMT*, dont un est commun

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats

aux deux cépages (correspondant à l'allèle présent dans le génome de référence PN40024) et un allèle différentiel. Ces 3 allèles codent ainsi pour 3 isoformes de protéines CAMT différentes qui ont un poids moléculaire de 40.3 kDa (Figure 57).

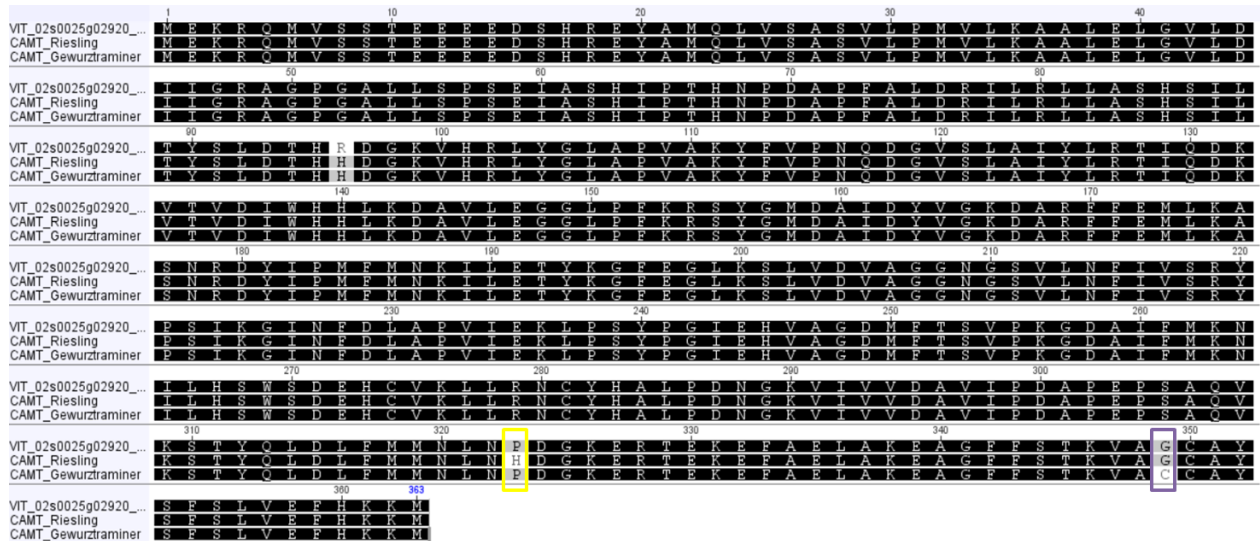


Figure 57 : Alignement des séquences en acides aminés des 3 isoformes de protéines de la CAMT présentes chez les Ri et le Gw.

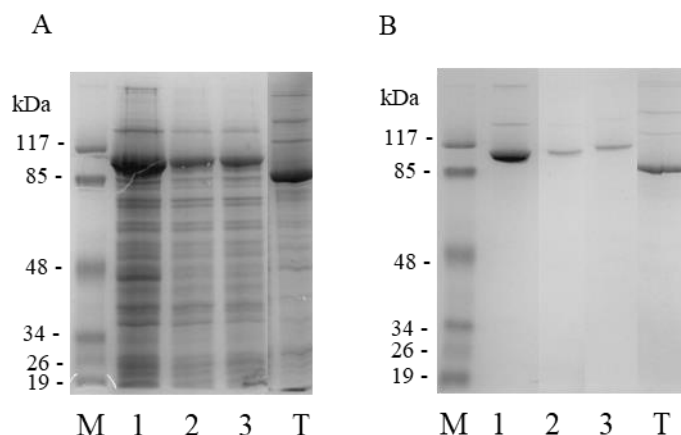
### 2.2.4. Caractérisation fonctionnelle de la CAMT

Afin de déterminer si le QTL sur le chromosome 2 de la vanilline est lié à une différence d'activité enzymatique entre les différents isoformes de protéine de la CAMT, le gène de la CAMT a été cloné et des expériences de caractérisation fonctionnelle ont été réalisées *in vitro* et *in planta*.

#### a. Expression hétérologue dans *E. coli*

Des expériences d'expression hétérologue de la protéine CAMT dans *E. coli* ont été réalisées afin de produire les 3 isoformes de protéines de CAMT *in vitro*. La production de la protéine CAMT a été induite dans *E. coli* puis la protéine a été purifiée via une colonne d'affinité Ni-TED grâce au tag contenant six histidines (6His). La protéine de fusion 6His-NusA-CAMT avec un poids moléculaire de 97.6 kDa (40.3 kDa+57.6 kDa) a été ainsi purifiée (Figure 58).

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats



**Figure 58 : SDS-page des protéines de fusion 6His-NusA avec les 3 isoformes de CAMT.**

(A) Extrait Brut (B) OMT purifiés sur Ni-TED 1: allèle Gw, 2: allèle Ri CAMT, 3: allèle commun PN40024 CAMT, T: Témoin positif AOMT, M : Marqueur de poids moléculaire.

Des tests enzymatiques ont ensuite été réalisés sur les extraits bruts d'*E. coli* exprimant les différentes isoformes de CAMT avec des mélanges de substrats potentiels. Sur les 22 substrats potentiels testés (Tableau 54), aucun n'a révélé l'activité enzymatique des différentes isoformes de la protéine CAMT.

Molécule	Formule chimique
3,5-dihydroxyanisole (DHA)	C7H8O3
3-phenylphenol	C12H10O
4-phenylphenol	C12H10O
Caffeic acid	C9H8O4
Gallic acid monohydrate	C7H6O5
Ethyl gallate	C9H10O5
2-isobutyl-3-hydroxypyrazine	C8H12N2O
Esculetin 6,7-dihydroxycoumarin	C9H6O4
Resveratrol	C14H12O3
Olivetol	C11H16O2
Naringenin	C15H12O5
(+/-)-Catechin	C15H14O6
Kaempferol	C15H10O6
Quercetin dihydrate	C15H10O7
Myricetin	C15H10O8
Coniferyl alcool	C10H12O3
3-hydroxybenzaldehyde	C7H6O2
4-hydroxybenzaldehyde	C7H6O2
3,4-dihydroxybenzaldehyde	C7H6O3
3-Caffeoylquinic acid	C16H18O9
4-O-Caffeoylquinic acid	C16H18O9
5-O-Caffeoylquinic acid	C16H18O9

**Tableau 54 : Substrats potentiels testés pour étudier l'activité de la CAMT.**

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats

### b. Expression transitoire dans *Nicotiana benthamiana*

Afin de tester l'activité enzymatique potentielle *in planta*, les 3 allèles du Ri et du Gw ont été clonés et exprimés chez *Nicotiana Benthamiana* par agroinfiltration avec *Agrobacterium tumefaciens* contenant un plasmide pMDC32-CAMT. Un contrôle négatif a été réalisé avec une transformation avec le vecteur pBIN-GFP permettant d'exprimer la GFP (Figure 59).



**Figure 59 :** Observation de l'expression de la GFP à la lumière UV, après transformation transitoire de *Nicotiana benthamiana* par le vecteur pBIN-GFP.

Aucune activité enzymatique différentielle n'a été détectée dans les feuilles de *Nicotiana benthamiana* exprimant les différents allèles de CAMT par rapport au contrôle négatif GFP.

### 3. Discussion

L'identification de gènes candidats et leurs caractérisations fonctionnelles sont les dernières étapes pour comprendre le déterminisme génétique de la biosynthèse de composés d'intérêts. De nombreux gènes candidats pouvant être reliés à des QTL ont ainsi pu être identifier (Chapitre 1). Pour étudier si ces gènes sont impliqués dans le déterminisme génétique de la biosynthèse de composés d'intérêts, des données génomiques et RNA-seq ont été utilisées afin de déterminer s'il y a des différences d'expressions ou des différences alléliques entre le Ri et le Gw pour les gènes candidats.

L'étude du déterminisme du contenu en linalol et de ses dérivés dans les baies de la population RixGw a révélé une région génomique sur le scaffold 271 comme étant responsable de la variation en linalol dans la population RixGw. Ce QTL avait déjà été identifié dans différentes études (Battilana et al., 2009; Duchêne et al., 2009, 2017; Emanuelli et al., 2010). Six TPS ont été caractérisées comme ayant une fonction de linalol synthase, cependant seule la TPS56 est située sur un intervalle de confiance du QTL expliquant la variation en contenu en linalol sur le chromosome 10 (Tableau 55).



## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats

Gène	Fonction	CRIBI ID	Références
<i>VviTPS54</i> <i>VviLinNer1</i>	(3S)-Linalol/(E)-Nerolidol synthase	VIT_00s0385g00020	Martin et al.,2010; Matarese et al.,2013
<i>VviTPS56</i> <i>VviLinNer2</i>	(3S)-Linalol/(E)-Nerolidol synthase	VIT_00s0271g00060	Martin et al.,2010; Zhu et al.,2014
<i>VviTPS57</i> <i>VviLNGI1</i>	Linalol/(E)-Nerolidol/ (E,E)-Geranyl linalol synthase	VIT_00s0372g00040	Martin et al.,2010
<i>VviTPS58</i> <i>VviLNGI3</i>	Linalol/(E)-Nerolidol/ (E,E)-Geranyl linalool synthase	VIT_00s0372g00070	Martin et al.,2010
<i>VviTPS61</i> <i>VviLNGI4</i>	Linalol/(E)-Nerolidol/ (E,E)-Geranyl linalol synthase	VIT_00s0372g00070	Martin et al.,2010
<i>VviTPS63</i> <i>VviLNGI2</i>	Linalol/(E)-Nerolidol/ (E,E)-Geranyl linalol synthase	VIT_00s0847g00020	Martin et al.,2010

**Tableau 55 : Gènes codant pour des linalol synthases ayant été caractérisés.**

Parmi les TPS présents sur le scaffold 271, trois gènes candidats ont été identifiés comme pouvant expliquer la différence en linalol entre les baies de Ri et de Gw. La TPS55 est la TPS la plus exprimée chez le Gw et présente des SNP différentiels. Cette TPS pourrait être responsable de la biosynthèse de linalol dans les baies de Gw. Les TPS 56 et 65 sont les TPS les plus exprimées dans le Ri, surtout dans les baies au moment de la véraison et présentent des SNP différentiels. Ainsi, la différence de concentration en linalol dans les baies de Ri et de Gw pourrait être expliquée par :

- Des différences d'expression des gènes codant pour les TPS56 et TPS65,
- Des différences liées aux SNP présents au niveau des séquences d'adressage de l'ARNm, la protéine étant plastidiale,
- Des différences d'activité enzymatique liées aux SNP présents dans les exons des gènes.

Les hypothèses concernant une différence d'adressage ainsi qu'une différence d'activité enzymatique des TPS doivent être testée par une caractérisation fonctionnelle des différents gènes candidats. Cette caractérisation fonctionnelle est en cours.

La voie de biosynthèse de la vanilline est très controversée (Gallage et al., 2014, 2018; Yang et al., 2017), c'est pourquoi il paraissait intéressant de l'étudier même si la contribution des baies dans le contenu des vins en vanilline est généralement plus faible que celle provenant de l'élevage du vin dans les fûts de chêne. Néanmoins, la vanilline des raisins peut être importante pour l'arôme des vins blancs qui ne sont pas élevés en fûts. Chez la vanille (*Vanilla planifolia*), la vanilline serait synthétisée par une vanilline synthase (VAN) à partir d'acide

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats

férulique (Figure 60) (Gallage et al., 2014). Néanmoins, cette voie de synthèse est controversée, les résultats précédents n'ayant pas pu être reproduits par d'autres équipes (Yang et al., 2017).

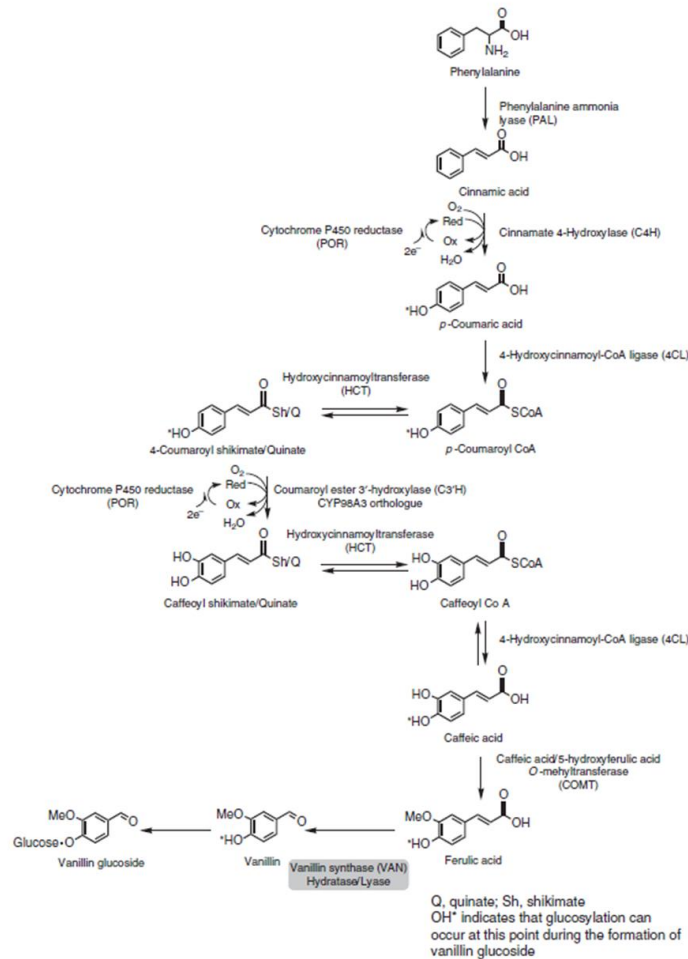


Figure 60 : Voie de biosynthèse de la vanilline chez la vanille (*V. planifolia*) (D'après Gallage et al., 2014).

L'étude du déterminisme génétique de la biosynthèse de vanilline dans les baies a révélé un QTL sur le chromosome 2 de la carte SSR sur une année d'analyse par GC-MS. La recherche d'un gène candidat a été poursuivie malgré la non reproductibilité du QTL. Un gène candidat a été identifié sur l'intervalle de confiance du QTL : une caffeic acid O methyltransférase (CAMT) putative. En effet, ce gène pourrait être impliqué dans la biosynthèse d'acide férulique à partir d'acide cafféique. Ce gène est exprimé dans les baies de Ri et de Gw de manière non différentielle. Chez le Ri et le Gw, ce gène est présent sous deux formes alléliques différentes, chacune ayant un allèle différentiel avec un SNP modifiant un acide aminé de la protéine et un autre allèle commun correspondant à l'allèle de PN40024.

## Chapitre 4 Identification et caractérisation fonctionnelle de gènes candidats

Les 3 allèles du gène *CAMT* ont été clonés et les isoformes de protéines caractérisées fonctionnellement *in vitro* et *in planta*. Les caractérisations fonctionnelles n'ont cependant pas été concluante : aucune activité enzymatique n'a pu être identifiée avec les différents substrats testés *in vitro* ni *in planta*. L'absence d'activité enzymatique pourrait être expliquée par le fait que le substrat de la protéine n'a pas été testé *in vitro* ou que ce substrat n'est pas présent dans les feuilles de *Nicotiana benthamiana*. Néanmoins, différents substrats potentiels liés à la voie de biosynthèse de la vanilline ayant été testés sans succès, il est possible que le gène *CAMT* ne soit pas le bon le gène candidat lié au QTL détecté.

Durant ma thèse, l'identification des gènes candidats et leur caractérisation fonctionnelle a été la partie la plus limitante. En effet, j'y ai rencontré des difficultés de différentes natures. Tout d'abord, la taille des intervalles de confiance des QTL pouvait être très grande et ceux-ci pouvaient contenir des centaines de gènes. Parmi ces gènes, certains font partie de familles multigéniques souvent constituées de gènes très similaires dupliqués en tandem comme les TPS ou encore les P450. Il est alors difficile de déterminer quel gène est vraiment impliqué dans la biosynthèse du composé d'intérêt. Une autre limite de l'approche est que l'identification de la fonction des protéines se fait par comparaison avec des bases de données de protéines. Si la voie de biosynthèse étudiée n'est pas connue et que le gène candidat code pour une protéine encore inconnue ou non annotée, elle ne pourra pas être identifiée ni caractérisée. L'utilisation du génome de PN40024 comme génome de référence fut aussi une limite dans mes analyses. En effet, le génome de PN40024 est incomplet car des régions hétérozygotes subsistent et ne sont pas bien séquencées et/ou alignées. De nombreux gènes se situent sur le chromosome « Unknown », chromosome regroupant les scaffolds qui n'ont pas pu être alignés sur le génome de référence. C'est le cas par exemple du scaffold 271 où se situent les gènes codant pour des linalol synthases. Pour pallier à cela, l'étude des gènes candidats devrait se faire directement sur les génomes des parents. Les génomes du Ri et du Gw, dont les séquences ont été obtenues et assemblées *de novo* au laboratoire, sont actuellement en cours d'analyse et pourront être utilisés à terme comme génomes de référence pour l'étude de la population RixGw.

## **Conclusion et perspectives**



### 1. Des outils de spectrométrie de masse complémentaires au service de l'analyse du métabolome de la vigne

Durant ma thèse, j'ai réalisé une étude non ciblée du métabolisme secondaire de la vigne grâce à des instruments de séparation physico-chimique des molécules couplés à des spectromètres de masse. Chaque outil a ses avantages et ses inconvénients et leur utilisation conjointe permet d'avoir une vision globale des molécules du métabolisme de la vigne (Tableau 56). Pour les analyses par GC-MS et UHPLC-MS, une donnée quantitative est liée à un ion caractérisé par un rapport  $m/z$ , et à un temps de rétention.

La GC-MS est un outil analytique robuste, permettant de séparer, d'identifier et de quantifier de manière précise des molécules volatiles que sont les arômes. L'identification des molécules grâce à leur spectre de masse ainsi qu'à leur indice de rétention permet d'avoir une information fiable sur la structure de la molécule. Cependant l'outil est limité aux molécules volatiles et non thermolabile et il n'est pas possible d'analyser directement les précurseurs d'arômes qui ne sont pas volatils. Pour pallier à cela, des analyses par UHPLC-MS ont été réalisées.

L'UHPLC-MS est un outil analytique performant, permettant de quantifier un grand nombre de molécules. La fragmentation des molécules en ESI (+) donne des informations structurales sur les molécules étudiées. La bonne précision des analyses réalisées par UHPLC-MS ESI (+) m'a permis d'annoter de nombreux ions avec des formules brutes, cependant cette étape, manuelle et limitante, ne permet pas d'établir des liens entre les molécules. Pour pallier à cela, des analyses ont été réalisées par FT-ICR.

Le FT-ICR est un outil de spectrométrie de masse à haute résolution et haute précision en masse, tant en ESI (+) qu'en ESI (-). Cette résolution des pics et cette précision en masse est cruciale pour pouvoir annoter les ions avec des formules brutes de manière automatique et par conséquent de construire des réseaux de relations chimiques entre molécules.

L'alignement des analyses conjointes réalisées par UHPLC-MS et FT-ICR sur les mêmes échantillons a permis d'associer les avantages des deux instruments.

Pour la plupart des molécules, les quantifications réalisées sont en réalité des semi-quantifications par l'utilisation des aires de pics. Cette semi-quantification est suffisante pour pouvoir comparer des analyses réalisées sur des échantillons de même type extraits de la même manière. Pour pouvoir obtenir la concentration exacte des molécules, il faudrait avoir à

disposition les standards des molécules, ce qui n'est pas possible pour la plupart des molécules de type précurseur d'arôme.

Les propositions de structure des molécules, déduites de l'analyse des données obtenues par UHPLC-MS et par FT-ICR sont pour la plupart putatives. Pour aller plus loin en l'absence de standard approprié, il faudrait pouvoir isoler les molécules qui paraissent intéressantes afin de les analyser par RMN.

L'analyse non ciblée du métabolome des baies grâce à ces trois outils analytiques a permis de quantifier des milliers de molécules. Cette analyse a révélé la richesse des profils métaboliques de la population de croisement RixGw, profils métabolomiques très riches avec l'apparition de molécules qui n'étaient pas présentes chez leurs parents.

L'intégralité des jeux de données n'a pas été exploitée de manière exhaustive. En particulier, du fait de leur moins bonne précision en masse, les analyses réalisées par UHPLC-MS ESI (-) n'ont pas été pleinement valorisées et des molécules détectées uniquement dans ce mode mériteraient un travail d'investigation complémentaire. Ces énormes jeux de données représentent une importante source d'information sur le métabolome de la vigne qui n'a pas fini de révéler tous ses secrets.

## Conclusion et perspectives

Caractéristiques	GC-MS	UHPLC-MS	FT-ICR
<b>Types de molécules détectées</b>	Uniquement molécules volatiles et non thermolabiles	Molécules solubles	Molécules solubles
<b>Précision en masse</b>	1 a.m.u	< 1 ppm en ESI (+) < 10 ppm en ESI (-)	< 0.2 ppm en ESI (+) et (-)
<b>Résolution des pics</b>	1000	25 000 – 100 000	> 100 000 → permet de différencier de manière précise des ions de m/z très proches
<b>Fragmentation</b>	Oui → Informations de structure	Oui pour ESI (+) → Informations de structure	Peu de fragmentation → Jeu de données plus simple à interpréter → Pas d'information de structure
<b>Séparation</b>	<b>Séparation par la température et par la colonne de chromatographie</b> Séparation des isomères possible	<b>Séparation par la colonne de chromatographie C18 et par le gradient de solvant</b> Séparation des isomères possible	<b>Injection par infusion</b> Pas de séparation des isomères Tout est injecté dans le spectromètre de masse en même temps
<b>Quantification</b>	Peu d'effet de suppression d'ion → Bonne quantification	Peu d'effet de suppression d'ion → Bonne quantification	Effet de suppression d'ion → Moins bonne quantification
<b>Identification</b>	- Spectre de masse - Indice de rétention	- m/z - Fragmentation - Temps de rétention	m/z
<b>Banque de données de spectre de masse</b>	Oui	Oui	Non

Tableau 56 : Avantages et inconvénients des outils utilisés pour les analyses métabolomiques.



### 2. Le métabolome des baies de raisins : une source de précurseurs d'arômes des vins

L'exploitation de grands jeux de données issus de différents outils analytiques très performants (GC-MS, UHPLC-MS et FT-ICR) a permis d'appréhender de manière la plus exhaustive possible les molécules pouvant participer de manière directe (GC-MS) ou indirecte (UHPLC-MS et FT-ICR) aux arômes des vins.

Les analyses par GC-MS permettent de quantifier les molécules directement volatiles participant directement aux arômes des vins. Elles permettent aussi de quantifier les molécules glycosylées après hydrolyse enzymatique sans pour autant avoir une information sur le sucre lié à la molécule. Cependant, l'hydrolyse enzymatique est incomplète et ne permet pas de visualiser la totalité du contenu en précurseurs d'arômes. Ce problème est compensé par l'utilisation de l'UHPLC-MS et la FT-ICR pour l'étude des composés solubles. La fragmentation des analyses réalisées par UHPLC-MS ESI (+) m'a permis d'identifier des molécules précurseurs d'arômes de type monoterpénols et norisoprénoïdes grâce à des fragments spécifiques. Ainsi, j'ai pu identifier et quantifier 15 monoterpénols et 17 norisoprénoïdes dans les baies. Ces molécules étaient présentes pour la plupart sous la forme glycosylées. L'apiosyl-glucoside de linalol oxyde et le glucoside de vomifoliol sont les molécules les plus abondamment détectées dans nos analyses.

L'étude du devenir des précurseurs d'arôme tels que les glucosides de terpénols lors de la vinification pourrait aider à identifier les précurseurs d'arômes présents dans les baies qui sont importants et qui participeraient de manière directe aux arômes des vins. C'est le cas par exemple du glucoside de carboxylinalol qui serait transformé en wine lactone par hydrolyse lors de la vinification (Ilc et al., 2017). L'étude du devenir de l'apiosyl-glucoside de linalol oxyde me paraît ainsi cruciale. Sous sa forme glycosylée, l'apiosyl-glucoside de linalol oxyde n'est pas détectable par nos récepteurs olfactifs. Seule une libération du sucre permettrait de libérer une molécule perceptible lors de la dégustation d'un vin. Pour savoir si cette molécule est hydrolysée pendant la vinification, des études de cinétique d'un vin de Gw sur plusieurs années seraient intéressantes. Si c'est le cas, cette molécule pourrait participer à la typicité des vins de Gw.

Le précurseur du TDN, molécule typique des arômes des vins de Ri reste encore à être identifié. Le malonyl-glucoside d' $\alpha$ -ionol, un norisoprénoïde présent préférentiellement dans les baies de Ri, pourrait constituer un précurseur du TDN. Cette hypothèse reste néanmoins à

tester par une étude du devenir de précurseurs d'arômes potentiels. Des expériences d'ajout de malonyl-glucoside d' $\alpha$ -ionol dans des vins permettraient de tester l'impact de ce composé sur l'accumulation de TDN au cours du vieillissement.

Une meilleure connaissance du métabolome des baies de Riesling et de Gewurztraminer permettra, à terme, de mieux comprendre la typicité des vins qui en sont issus.

### 3. Les feuilles : un outil d'étude du métabolome de la vigne à exploiter

L'étude du déterminisme génétique de la biosynthèse des métabolites secondaires à la fois dans les baies et dans les feuilles prélevées la même année montre qu'il était possible d'utiliser les feuilles comme outil d'étude d'un certain nombre de métabolites que l'on trouve à la fois dans les feuilles et dans les baies. C'est le cas par exemple des monoterpénols que l'on retrouve à la fois dans les baies et dans les feuilles. La transposition des résultats obtenus avec les analyses des feuilles sur les baies est à faire cependant avec précaution. Les feuilles présentent un nombre bien inférieur de molécules et ne contiennent pas toute la diversité des molécules présentes dans les baies.

Cependant, les feuilles ont l'avantage de contenir peu de composés très polaires tels que les sucres et les acides, permettant une extraction très simple sans passer par une étape de SPE. Les analyses réalisées pourraient cependant être optimisées avec une étape de concentration des extraits méthanoliques de feuille par exemple. Les feuilles représentent un matériel intéressant à utiliser car elles sont disponibles dès la première année de croissance des plantules et présentes toute l'année lorsque la plante est en serre alors que les baies ne sont disponibles qu'après trois ans de croissance et uniquement durant l'été. Les feuilles pourraient ainsi être utilisées pour l'analyse précoce de certaines caractéristiques des raisins, lors de l'étude d'une population de croisement dont les baies ne sont pas encore disponibles.

### 4. Une approche innovante : l'analyse en réseaux du métabolome à la lumière des QTL

La combinaison des données issues des analyses métabolomiques avec des données de cartographies génétique m'a permis de détecter des milliers de QTL sur tout le génome de la vigne. Ces milliers de QTL ont pu être détectés grâce à la richesse de l'information génétique constituée de milliers de SNPs, combinée avec l'information de la quantification de milliers d'ions. Une partie significative des milliers de QTL co-localisent cependant en un nombre restreint de régions génomiques que nous avons désigné sous le nom de « hotspots ».

Parmi ces hotspots, des régions génomiques gouvernant les concentrations de nombreux précurseurs d'arômes ont été identifiées :

- Un hotspot majeur du Gw sur le chromosome 5 impacte de nombreuses molécules de la famille des isoprénoïdes : monoterpénols, norisoprénoïdes, sesquiterpènes, et diterpènes.
- Un hotspot du Ri sur le chromosome 12 impacte de nombreuses molécules de la famille des norisoprénoïdes et des monoterpénols.

D'autres hotspots génétiques non liés à des arômes ont été découverts, dont notamment un hotspot lié à la biosynthèse de flavonoïdes sur le chromosome 6 et un hotspot sur le chromosome 3 lié à la biosynthèse de dérivés d'acides chlorogéniques dans les feuilles.

L'étude des liens chimiques et la mise en réseau des molécules impactées par un même locus permet d'identifier les relations qu'il peut y avoir entre elles. Il est possible par exemple de relier des ions correspondants à des molécules différenciellement oxydées ou encore à des molécules différenciellement glycosylées. Cette approche a été validée par l'analyse des conséquences des variations génétiques sur le chromosome 5 lié à la DXS sur des ensembles de molécules en combinant des analyses par FT-ICR avec des analyses par UHPLC-MS. Cette approche pourra être appliquée à d'autres hotspots comme celui sur le chromosome 12 impactant la biosynthèse des norisoprénoïdes et des monoterpénols.

Cette approche d'étude en réseau des molécules impactées par un même QTL permettra à terme d'avoir de meilleures connaissances sur les voies de synthèse et équilibres biochimiques par l'analyse de réseaux de molécules impactées par une même région génomique. Elle permettra aussi d'avoir de nouvelles connaissances sur la notion d'arôme variétal et de « bouquet » afin de pouvoir anticiper l'impact de la création variétale sur les arômes des vins.

### 5. L'identification des gènes candidats et leur caractérisation fonctionnelle, une étape limitante dans la compréhension du déterminisme génétique du métabolome de la vigne

L'analyse détaillée des gènes liés aux QTL permet potentiellement d'identifier les gènes responsables de la variation en contenu en arômes et en précurseurs. Ce type d'analyse est rendu possible par la combinaison de différentes informations génomiques et transcriptomiques disponibles pendant ma thèse (séquence génomique de PN40024, données RNAseq du Ri et du Gw). A partir des gènes se situant sur l'intervalle de confiance des QTL, un tri est réalisé en fonction des fonctions putatives des protéines codées par ces gènes, de leur expression dans les baies et enfin de la présence d'allèles différents chez le Ri et le Gw.

A partir du QTL localisé sur le chromosome 10 et lié au contenu en linalol, trois TPS (TPS55, TPS56 et TPS65) ont été identifiées comme pouvant participer à la biosynthèse de linalol dans les baies de Ri et de Gw. Ces trois gènes sont exprimés dans les baies et sont présents sous différentes formes alléliques chez le Ri et le Gw. La caractérisation fonctionnelle des protéines correspondantes permettra d'étudier si elles possèdent des activités linalol synthase et si les différents allèles de Ri et de Gw codent pour des isoformes ayant des caractéristiques enzymatiques différentes.

Un gène candidat potentiellement impliqué dans la biosynthèse de vanilline a été identifié : la caffeic acid-O-Methyltransférase. Ce gène a été cloné en vue de sa caractérisation fonctionnelle, mais aucune activité enzymatique n'a pu être détectée, malgré le test d'un nombre important de substrats potentiels. Cet exemple illustre la difficulté à identifier le gène responsable du caractère étudié, même quand un gène candidat semble prometteur en raison de sa fonction putative.

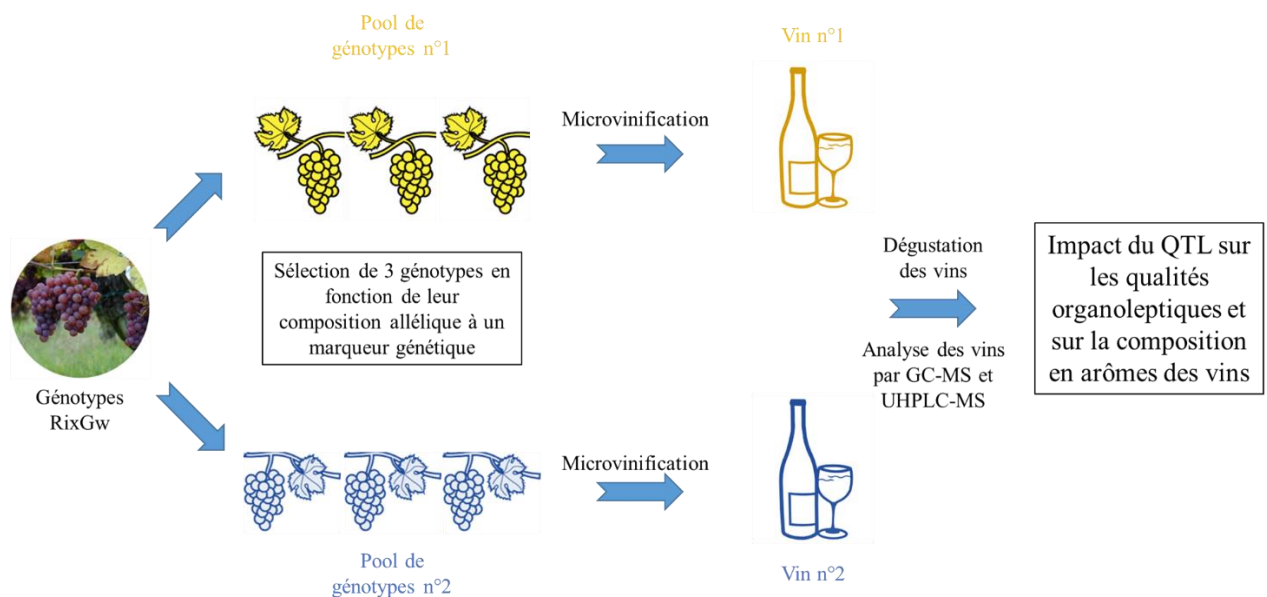
L'étape d'identification et de caractérisation des gènes candidats est l'étape la plus limitante du fait du grand nombre de gènes pouvant être présent sur un intervalle de confiance. La taille des intervalles de confiance des QTL reste encore grande malgré la densité des marqueurs génétiques. Pour pallier à cela, des cartes génétiques plus denses de type Genotyping By Sequencing ont été réalisées sur davantage de génotypes (252) afin de faire des recherches QTL plus précises. Afin d'augmenter la puissance de la recherche QTL, le génotypage et l'analyse du métabolome d'autres génotypes de la population RixGw serait possible.

Enfin, l'utilisation de PN40024 comme génome de référence comporte des avantages, mais aussi des inconvénients. En effet, la présence de régions génomiques hétérozygotes et/ou non

assemblée complique la recherche de gènes candidats. Pour pallier à cela, les génomes de Ri et de Gw sont en cours d'assemblage et d'annotation afin de pouvoir les utiliser directement comme génomes de référence pour les études génétiques réalisées grâce à la population RixGw.

### 6. Des marqueurs génétiques comme outils pour une application directe dans la création variétale

Mon travail de thèse a permis la découverte de nombreux marqueurs génétiques liés aux molécules aromatiques ainsi qu'à leurs précurseurs. La validation de l'utilisation de marqueurs moléculaires que j'ai pu mettre en évidence est actuellement réalisée par l'analyse chimique et gustative de vins produits à partir de génotypes de la population RixGw choisis en fonction de marqueurs génétiques associés à des QTL d'intérêt (Figure 61). Cette approche permet de vérifier directement l'impact d'un QTL sur la composition et les qualités organoleptiques du produit final qu'est le vin.



**Figure 61 : Stratégie de validation de l'impact des QTL sur la qualité des vins par microvinification de couples de QTL.**

Ces marqueurs génétiques pourront être utilisés pour la sélection variétale, dans le cadre par exemple des croisements entre vignes américaines et vignes européennes, afin d'obtenir des vins ayant des caractéristiques organoleptiques spécifiques.

## Annexe 1 : QTL GC-MS apparaissant uniquement sur un an d'analyse.

Chrom.	Famille	Molécule	Libre - Lié	Carte génétique	Année	LOD	% Variance	Intervalle de confiance (cM)
Chr1	Norisoprenoïde	3-Oxo-7,8-dihydro- $\alpha$ -ionol (#51)	Lié	Ri	2014	4.9	17.7	21.6-45.8
	Phénylpropanoïde	Vanillin (#44)	Libre	Gw	2014	3.5	1.5	0-80.1
Chr2	Norisoprenoïde	Dihydro-3-oxo- $\beta$ -ionol (#50)	Lié	Ri	2015	3.2	11.8	32-77.1
	Phénylpropanoïde	Eugenol (#31)	Libre	Ri	2015	2.8	10.3	17.5-62.8
			Lié	Ri	2015	6.6	22.9	35.5-77.1
		Vanillin (#44)	Libre	SSR	2015	7.1	24.5	28-38.2
				Gw	2014	3.6	4.3	0-40.3
	SSR	2014	5.1	18.1	0-12			
Chr3	Norisoprenoïde	3-Hydroxy- $\beta$ -damascone (#43)	Lié	SSR	2014	4.7	17.0	40-56
	Sesquiterpène	n.i. Sesquiterpène (#48)	Lié	Gw	2015	2.0	6.1	25.1-57
	MonoMonoterpénol	Hotrienol (#12)	Lié	Ri	2015	2.4	6.9	0.9-37.5
		Linalool oxide (pyranoid), trans- (#17)	Libre	Ri	2015	2.4	6.5	0.9-37.5
	Phénylpropanoïde	Benzyl alcohol (#25)	Lié	Gw	2015	4.0	10.5	15.2-57
				SSR	2015	4.9	15.7	40-57.3
		Homovanillin alcohol (#53)	Lié	SSR	2015	6.9	23.8	0-16
				Ri	2014	3.1	8.9	9.9-50.1
Methyl vanillate (#45)	Lié	Gw	2015	3.3	12.3	0-9		
Chr4	Monoterpénol	(Z)-2,6-Dimethyl-3,7-octadiene-2,6-diol (#16)	Lié	Ri	2014	3.0	4.6	13.7-74.3
		Linalool oxide, (furanoid), trans- (#7)	Lié	Ri	2015	2.6	6	17.2-74.3
	Phénylpropanoïde	Benzeneacetaldehyde (#13)	Libre	Ri	2014	2.6	10.0	17.2-68.8
		Homovanillin alcohol (#53)	Lié	Gw	2014	2.9	10.5	23.4-60
		Methyl vanillate (#45)	Lié	Gw	2014	5.1	18.4	48.4-57.3
				SSR	2014	5.3	18.6	62-65.2
		Phenylethyl Alcohol (#26)	Lié	Gw	2015	3.3	12.4	31.4-48.4
				SSR	2015	4.0	14.7	42-59.3
Chr5	Fatty alcohol	1-Octanol (#11)	Libre	Gw	2014	5.8	20.3	1.8-15.4
				SSR	2014	7.1	24.3	0-10
	Hemiterpénol	isoprenol (#2)	Lié	Gw	2015	3.8	14.1	0-21.7
				SSR	2015	4.3	15.5	0-12.5
	Norisoprenoïde	Prenol (#3)	Lié	SSR	2015	8.9	23.9	0-10
				Gw	2015	4.0	14.6	0-19

**Annexe 1 : QTL GC-MS apparaissant uniquement sur un an d'analyse.**

Chrom.	Famille	Molécule	Libre - Lié	Carte génétique	Année	LOD	% Variance	Intervalle de confiance (cM)
		1.α,β-Dihydro-β-ionone (#49)		SSR	2015	4.7	17.0	0-14
	Monoterpénoïde	(E) 8-Hydroxylinalool (#36)	Libre	Gw	2015	4.6	16.7	1.8-15.4
				SSR	2015	4.4	15.8	0-10
			Lié	Gw	2015	9.7	32.1	8.1-15.4
				SSR	2015	10.1	33.1	2-10
		(E)-2,6-Dimethyl-3,7-octadiène-2,6-diol (#28)	Libre	Gw	2015	2.6	9.7	9.9-27.1
			Lié	SSR	2015	5.1	17.2	2-12
		(Z) 8-Hydroxylinalool (#35)	Libre	Gw	2014	4.4	15.8	9.9-19
				SSR	2014	4.3	19.3	2-12
		(Z)-2,6-Dimethyl-3,7-octadiène-2,6-diol (#16)	Lié	Gw	2015	3.7	13.6	9.9-24.4
				SSR	2015	4.1	15.0	2-20
		1,7-Octadiène-3-ol, 2,6-diméthyl- (#34)	Lié	Gw	2015	3.7	13.8	0-21.7
		2-Hydroxy-1,8-cineole (#24)	Lié	Gw	2015	2.7	10.1	0-29.7
		8-Hydroxycarvotanacétone (#37)	Lié	Gw	2014	3.9	14.2	1.8-19
		α-Terpénoïde (#15)	Libre	Gw	2014	9.0	29.8	9.9-15.4
				SSR	2014	9.2	30.4	0-12
			Lié	Gw	2015	3.2	11.8	0-21.7
				SSR	2015	3.9	5.9	0-18
		Geranic acid (#39)	Lié	Gw	2015	9.0	30.0	9.9-19
	SSR			2015	9.3	30.9	4-16	
Lilac alcohol C (#20)	Lié	Gw	2015	3.5	13.1	0-21.7		
Linalool oxide (pyranoid), trans- (#17)	Lié	SSR	2015	6.4	21.2	4-18		
Linalool oxide, (furanoid), cis- (#8)	Libre	SSR	2014	4.7	17.0	0-32		
Linalool oxide, (furanoid), trans- (#7)	Lié	Gw	2015	3.9	14.5	7.3-23.5		
		SSR	2015	4.4	17.9	2-12.5		
Produit dérivés de lipoxygénase		2-Hexen-1-ol, (E)- (#6)	Libre	SSR	2014	5.1	18.1	20-32
Chr6	Monoterpénoïde	n.i. terpène (#41)	Libre	Gw	2015	5.0	18.0	17.2-35.1
				SSR	2015	5.8	20.5	12-32
		(Z) 8-Hydroxygeraniol (#47)	Libre	Gw	2014	3.0	10.3	5.4-26.2
				SSR	2014	6.6	19.0	10-20
				Lié	Gw	2014	3.1	10.0
	1,7-Octadiène-3-ol, 2,6-diméthyl- (#34)	Lié	Gw	2014	3.0	10.3	4.5-44.2	
			SSR	2014	4.2	7.3	0-44	
Produit dérivés de lipoxygénase		2-Hexen-1-ol, (E)- (#6)	Libre	Ri	2014	3.3	12.1	1.8-29.5
Phénylpropanoïde		Benzyl alcohol (#25)	Libre	Gw	2014	3.0	11.3	10.8-54.4
Chr7	Heptadiénal	2,4-Heptadiénal, (E,E)- (#9)	Libre	Gw	2014	3.3	12.1	0-42.7
	Monoterpénoïde	n.i. terpène (#41)	Libre	Gw	2014	4.0	14.5	53.4-78.4
		Citronellol (#19)	Libre	Gw	2014	3.0	13.4	62.5-81.1

**Annexe 1 : QTL GC-MS apparaissant uniquement sur un an d'analyse.**

Chrom.	Famille	Molécule	Libre - Lié	Carte génétique	Année	LOD	% Variance	Intervalle de confiance (cM)
	Produit dérivés de lipoxygénase	2-Hexen-1-ol, (E)- (#6)	Lié	SSR	2015	4.4	12.5	6-70
	Phénylpropanoïde	Acetovanillone (#46)	Libre	SSR	2014	5.7	20.1	80-85.7
<b>Chr8</b>	Hemiterpénoïde	Isoprenol (#2)	Lié	SSR	2014	4.2	10.3	40-62
	Produit dérivés de lipoxygénase	1-Hexanol (#4)	Lié	SSR	2014	4.9	10.2	26-56.7
<b>Chr9</b>	Norisoprenoïde	1.α,β-Dihydro-β-ionone (#49)	Libre	SSR	2015	4.6	16.6	24-62.3
	Phénylpropanoïde	Benzèneacétaldéhyde (#13)	Libre	Gw	2015	2.6	9.8	0-30.6
		Benzyl alcohol (#25)	Lié	Ri	2014	3.9	14.3	36.8-72.7
<b>Chr10</b>	Monoterpénoïde	(E)-2,6-Diméthyl-3,7-octadiène-2,6-diol (#28)	Libre	SSR	2015	4.0	14.5	0-12
		(Z) 8-Hydroxylinalool (#35)	Libre	Ri	2014	3.4	12.5	0-11.6
			SSR	2014	4.8	9.0	0-24	
		(Z)-2,6-Diméthyl-3,7-octadiène-2,6-diol (#16)	Lié	SSR	2014	6.8	23.5	0-9.7
		1,7-Octadiène-3,6-diol, 2,6-diméthyl- (#30)	Lié	SSR	2015	4.7	14.0	0-6
	Hotrienol (#12)	Lié	SSR	2015	4.3	15.5	0-4	
	Phénylpropanoïde	Benzyl alcohol (#25)	Lié	SSR	2014	5.7	16.2	6-28
		Phényléthyl Alcohol (#26)	Libre	Ri	2015	2.9	10.8	2.7-34.3
SSR			2015	4.0	14.7	4-20		
<b>Chr11</b>	Sesquiterpène	n.i. Sesquiterpène (#48)	Lié	SSR	2014	5.8	14.4	26-46
	Monoterpénoïde	(E)-2,6-Diméthyl-3,7-octadiène-2,6-diol (#28)	Libre	SSR	2014	4.6	16.5	4-14
		(Z) 8-Hydroxylinalool (#35)	Libre	SSR	2014	4.2	11.1	2-14
			Lié	Gw	2014	3.1	8.3	1.8-36.9
			SSR	2014	5.5	8.1	0-20	
		1,7-Octadiène-3-ol, 2,6-diméthyl- (#34)	Lié	Gw	2014	3.3	8.5	3.6-35.1
		SSR	2014	4.5	6.2	0-90		
		1,7-Octadiène-3,6-diol, 2,6-diméthyl- (#30)	Lié	SSR	2014	5.2	21.4	4-12
		Citronellol (#19)	Lié	SSR	2015	4.5	6.0	14-52
	Linalool (#10)	Lié	Gw	2014	2.9	8.6	3.6-41.4	
		Ri	2014	3.0	7.3	0-41.2		
	Linalool oxide, (furanoid), cis- (#8)	Lié	Ri	2014	3.2	11.8	7.2-43	
Phénylpropanoïde	4-Vinylguaiacol (#32)	Lié	SSR	2015	4.3	15.7	92-95.8	
	Phényléthyl Alcohol (#26)	Lié	Ri	2014	2.9	10.6	19-48.4	
		SSR	2014	3.5	13.4	14-48		
Zingerone (#52)	Lié	SSR	2015	4.8	10.7	0-95.8		
<b>Chr12</b>	Norisoprenoïde	3-Hydroxy-β-damascone (#43)	Lié	Ri	2015	3.6	13.4	14.6-40.2
			SSR	2015	4.1	15.1	6-28	



**Annexe 1 : QTL GC-MS apparaissant uniquement sur un an d'analyse.**

Chrom.	Famille	Molécule	Libre - Lié	Carte génétique	Année	LOD	% Variance	Intervalle de confiance (cM)
	Monoterpénol	Linalool oxide, (furanoid), cis- (#8)	Lié	Ri	2015	2.7	10.3	13.7-46.5
	Produit dérivés de lipoxygénase	2-Hexen-1-ol, (E)- (#6)	Lié	Ri	2015	4.5	16.3	43.8-52.7
				SSR	2015	4.8	14.1	10-32.1
		3-Hexen-1-ol, (Z)- (#5)	Libre	Ri	2014	4.3	16.0	27-50.9
	Phénylpropanoïde	Benzeneacetaldehyde (#13)	Libre	Ri	2015	2.6	9.7	27-52.7
Gw				2014	2.9	11.0	4.4-39.3	
Chr13	Monoterpénol	2-Hydroxy-1,8-cineole (#24)	Lié	Gw	2014	2.9	10.8	36.7-55.9
				Ri	2014	7.6	25.8	55.1-74.6
		n.i terpene (#40)	Lié	Gw	2014	3.2	9.3	21.7-55.9
		α-Terpineol (#15)	Lié	Gw	2014	3.8	14.0	36.7-50.5
				Ri	2014	6.5	22.7	51.5-71.1
		Geranic acid (#39)	Lié	Gw	2014	2.0	7.7	0-49.6
	Ri			2014	1.9	7.0	13.6-83.6	
	Lilac alcohol C (#20)	Libre	Gw	2014	1.4	5.5	0-36.7	
	Lilac alcohol D (#22)	Libre	Gw	2014	1.6	6	0-36.7	
	Produit dérivés de lipoxygénase	3-Hexen-1-ol, (Z)- (#5)		SSR	2015	4.7	16.8	6-26
	Phénylpropanoïde	Acetovanillone (#46)	Libre	SSR	2015	5.6	19.7	10-24
				Ri	2015	3.1	11.4	43.1-67.5
SSR				2015	4.1	14.8	8-30	
	Zingerone (#52)	Libre	Ri	2015	2.9	14.2	43.9-65.7	
Chr14	Fatty alcohol	1-Nonanol (#14)	Lié	Gw	2014	3.2	11.5	20.4-45.6
				Gw	2014	3.2	11.6	19.5-52.7
	Monoterpénol	Isopregol (#42)	Libre	Ri	2015	1.2	4.6	0.9-77.4
				Gw	2015	3.3	7.8	38.4-64.5
	Produit dérivés de lipoxygénase	1-Hexanol (#4)	Lié	Gw	2014	3.4	13.7	37.5-64.5
				Ri	2015	3.3	12.3	9.2-74.8
SSR				2015	4.1	15.2	30-56	
Chr15	Norisoprenoïde	3-Hydroxy-β-damascone (#43)	Lié	Gw	2014	3.0	11.6	12.4-40.4
				Gw	2015	4.3	15.7	38.6-51.1
		Dihydro-3-oxo-β-ionol (#50)	Lié	SSR	2015	4.1	11.0	14-48.6
	Sesquiterpène	n.i. Sesquiterpène (#48)		SSR	2014	5.2	12.6	26-46
	Monoterpénol	1,7-Octadiene-3,6-diol, 2,6-dimethyl- (#30)	Libre	Gw	2015	2.5	7.5	16-51.1
				Lié	Gw	2014	3.0	8.9
		Lilac alcohol formate (#27)	Lié	Gw	2015	2.8	8.4	25.1-51.1
				SSR	2015	5.0	11.8	36-48.6
		Linalool (#10)	Libre	Gw	2015	2.2	6.5	5.3-51.1
	Lié			Gw	2014	2.9	7	9.8-46.7
Phénylpropanoïde	Homovanillin alcohol (#53)	Lié	Gw	2015	3.5	13.1	25.1-46.7	
Chr16	Norisoprenoïde	3-Oxo-7,8-dihydro-α-ionol (#51)	Lié	SSR	2015	4.2	15.2	22-32

**Annexe 1 : QTL GC-MS apparaissant uniquement sur un an d'analyse.**

Chrom.	Famille	Molécule	Libre - Lié	Carte génétique	Année	LOD	% Variance	Intervalle de confiance (cM)
	Monoterpénoïde	n.i. terpene (#41)	Lié	Gw	2015	3.4	7.0	0-36.9
		n.i. terpene (#38)	Libre	Gw	2015	2.8	4.3	9.9-65.6
			Lié	SSR	2015	4.1	7.3	20-42
			Nerol (#21)	Libre	Gw	2015	3.1	2.9
	Phénylpropanoïde	Homovanillin alcool (#53)	Libre	Ri	2014	4.3	12.4	31.1-52.8
				SSR	2014	6.0	20.9	34-48
			Lié	Ri	2014	3.9	11.4	41.2-52.8
				SSR	2014	4.7	15.6	32-46
<b>Chr17</b>	Monoterpénoïde	Linalool oxide (pyranoid), cis- (#18)	Lié	Ri	2015	3.3	12.3	0.9-23.6
				SSR	2015	4.3	15.8	0-22
<b>Chr18</b>	Produit dérivés de lipoxygénase	1-Octanol (#11)	Lié	Gw	2014	3.6	12.9	0-14.3
	Hemiterpénoïde	Isoprenol (#2)	Lié	SSR	2014	7.5	21.0	4-16
	Norisoprenoïde	Dihydro-3-oxo- $\beta$ -ionol (#50)	Lié	SSR	2015	4.1	11.1	80-104
	Sesquiterpène	n.i. Sesquiterpène (#48)	Lié	Gw	2014	2.7	7.6	0-56.8
	Monoterpénoïde	n.i. terpene (#41)	Lié	Gw	2014	3.8	10.6	0.9-14.3
			Lié	Gw	2014	2.9	9.4	0-45
			Lié	Ri	2015	2.4	7.9	21.2-77
	Produit dérivés de lipoxygénase	1-Hexanol (#4)	Libre	Gw	2014	4.6	16.6	35-48.6
				SSR	2014	4.5	16.3	86-108
			Lié	Gw	2014	3.0	12.3	0-46.8
				SSR	2014	6.2	15.1	0-30
	2-Hexen-1-ol, (E)- (#6)	Lié	SSR	2014	4.2	15.1	8-108.5	
Phénylpropanoïde	Phenylethyl Alcohol (#26)	Libre	SSR	2015	4.2	14.7	0-14	
<b>Chr19</b>	Fatty alcohol	1-Nonanol (#14)	Lié	SSR	2015	4.9	13.5	36-42
		1-Octanol (#11)	Lié	SSR	2015	4.8	14.5	36-42
	Hemiterpénoïde	Prenol (#3)	Lié	Gw	2015	4.4	10.2	0-26.7
				SSR	2015	5.0	10.3	0-28
	Monoterpénoïde	$\alpha$ -Terpineol (#15)	Lié	SSR	2015	4.0	4.2	2-20
	Phénylpropanoïde	Zingerone (#52)	Libre	Ri	2014	2.3	8.7	0-28.2
			Lié	Ri	2014	3.1	9.7	1.8-28.2

**Annexe 1 : QTL GC-MS apparaissant uniquement sur un an d'analyse.**

## Références bibliographiques

- Ahrazem, O., Trapero, A., Gómez, M.D., Rubio-Moraga, A., and Gómez-Gómez, L. (2010). Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family: A deeper study in *Crocus sativus* and its allies. *Genomics* 96, 239–250.
- Auldridge, M.E., McCarty, D.R., and Klee, H.J. (2006). Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Curr. Opin. Plant Biol.* 9, 315–321.
- Battilana, J., Costantini, L., Emanuelli, F., Sevini, F., Segala, C., Moser, S., Velasco, R., Versini, G., and Grando, M.S. (2009). The 1-deoxy-d-xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theor. Appl. Genet.* 118, 653–669.
- Battilana, J., Emanuelli, F., Gambino, G., Gribaudo, I., Gasperi, F., Boss, P.K., and Grando, M.S. (2011). Functional effect of grapevine 1-deoxy-D-xylulose 5-phosphate synthase substitution K284N on Muscat flavour formation. *J. Exp. Bot.* 62, 5497–5508.
- Brandi, F., Bar, E., Mourgues, F., Horváth, G., Turcsi, E., Giuliano, G., Liverani, A., Tartarini, S., Lewinsohn, E., and Rosati, C. (2011). Study of “Redhaven” peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC Plant Biol.* 11, 24.
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisine, N., Mohellibi, N., Guichard, C., Rombauts, S., Le Clainche, I., et al. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* 14, 56–62.
- Champagne, A., and Boutry, M. (2016). Proteomics of terpenoid biosynthesis and secretion in trichomes of higher plant species. *Biochim. Biophys. Acta* 1864, 1039–1049.
- Croteau, R., and Karp, F. (1994). Origin of Natural Odorants. In *Perfumes: Art, Science and Technology*, P.M. Müller, and D. Lamparsky, eds. (Dordrecht: Springer Netherlands), pp. 101–126.
- Dalla Costa, L., Emanuelli, F., Trenti, M., Moreno-Sanz, P., Lorenzi, S., Coller, E., Moser, S., Slaghenaufi, D., Cestaro, A., Larcher, R., et al. (2018). Induction of Terpene Biosynthesis in Berries of Microvine Transformed with VvDXS1 Alleles. *Front. Plant Sci.* 8.
- Darriet, P., Tominaga, T., Lavigne, V., Boidron, J.-N., and Dubourdieu, D. (1995). Identification of a powerful aromatic component of *Vitis vinifera* L. var. sauvignon wines: 4-mercapto-4-methylpentan-2-one. *Flavour Fragr. J.* 10, 385–392.
- Davies, C., Boss, P.K., Geros, H., Lecourieux, F., and Delrot, S. (2012). Source/Sink Relationships and Molecular Biology of Sugar Accumulation in Grape Berries. In *The Biochemistry of the Grape Berry*, H. Gerós, M. Manuela Chaves, and S. Delrot, eds. (BENTHAM SCIENCE PUBLISHERS), pp. 44–66.
- DeBolt, S., Ristic, R., Iland, P.G., and Ford, C.M. (2008). Altered Light Interception Reduces Grape Berry Weight and Modulates Organic Acid Biosynthesis During Development. *HortScience* 43, 957–961.

## Références bibliographiques

- Dennis, E.G., Keyzers, R.A., Kalua, C.M., Maffei, S.M., Nicholson, E.L., and Boss, P.K. (2012). Grape Contribution to Wine Aroma: Production of Hexyl Acetate, Octyl Acetate, and Benzyl Acetate during Yeast Fermentation Is Dependent upon Precursors in the Must. *J. Agric. Food Chem.* *60*, 2638–2646.
- Doligez, A., Audiot, E., Baumes, R., and This, P. (2006). QTLs for muscat flavor and monoterpenic odorant content in grapevine (*Vitis vinifera* L.). *Mol. Breed.* *18*, 109–125.
- Drew, D.P., Andersen, T.B., Sweetman, C., Møller, B.L., Ford, C., and Simonsen, H.T. (2016). Two key polymorphisms in a newly discovered allele of the *Vitis vinifera* TPS24 gene are responsible for the production of the rotundone precursor  $\alpha$ -guaiene. *J. Exp. Bot.* *67*, 799–808.
- Dubourdieu, D., and Tominaga, T. (2009). Polyfunctional Thiol Compounds. In *Wine Chemistry and Biochemistry*, M.V. Moreno-Arribas, and M.C. Polo, eds. (New York, NY: Springer New York), pp. 275–293.
- Duchêne, E., Butterlin, G., Claudel, P., Dumas, V., Jaegli, N., and Merdinoglu, D. (2009a). A grapevine (*Vitis vinifera* L.) deoxy-d-xylulose synthase gene colocalizes with a major quantitative trait loci for terpenol content. *Theor. Appl. Genet.* *118*, 541–552.
- Duchêne, E., Butterlin, G., Claudel, P., Dumas, V., Jaegli, N., and Merdinoglu, D. (2009b). A grapevine (*Vitis vinifera* L.) deoxy-d-xylulose synthase gene colocalizes with a major quantitative trait loci for terpenol content. *Theor. Appl. Genet.* *118*, 541–552.
- Duchêne, E., Butterlin, G., Dumas, V., and Merdinoglu, D. (2012). Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages. *Theor. Appl. Genet.* *124*, 623–635.
- Duchêne, E., Butterlin, G., Claudel, P., Dumas, V., Jaegli, N., Huguency, P., Arnold, G., and Merdinoglu, D. (2017). Genetic determinism of the “Muscat” flavour in grapevine (*Vitis vinifera* L.) cultivars. In *Ix International Symposium on Grapevine Physiology and Biotechnology*, M. Pinto, ed. (Leuven 1: Int Soc Horticultural Science), pp. 87–91.
- Dunlevy, J.D., Kalua, C.M., Keyzers, R.A., and Boss, P.K. (2009). The Production of Flavour & Aroma Compounds in Grape Berries. In *Grapevine Molecular Physiology & Biotechnology*, K.A. Roubelakis-Angelakis, ed. (Dordrecht: Springer Netherlands), pp. 293–340.
- Dunlevy, J.D., Soole, K.L., Perkins, M.V., Dennis, E.G., Keyzers, R.A., Kalua, C.M., and Boss, P.K. (2010). Two O-methyltransferases involved in the biosynthesis of methoxypyrazines: grape-derived aroma compounds important to wine flavour. *Plant Mol. Biol.* *74*, 77–89.
- Dunlevy, J.D., Dennis, E.G., Soole, K.L., Perkins, M.V., Davies, C., and Boss, P.K. (2013). A methyltransferase essential for the methoxypyrazine-derived flavour of wine. *Plant J.* *75*, 606–617.
- Emanuelli, F., Battilana, J., Costantini, L., Le Cunff, L., Boursiquot, J.-M., This, P., and Grando, M.S. (2010). A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol.* *10*, 241.
- Emanuelli, F., Sordo, M., Lorenzi, S., Battilana, J., and Grando, M.S. (2014). Development of user-friendly functional molecular markers for VvDXS gene conferring muscat flavor in grapevine. *Mol. Breed.* *33*, 235–241.

- Estévez, J.M., Cantero, A., Reindl, A., Reichler, S., and León, P. (2001). 1-Deoxy-d-xylulose-5-phosphate Synthase, a Limiting Enzyme for Plastidic Isoprenoid Biosynthesis in Plants. *J. Biol. Chem.* *276*, 22901–22909.
- Gallage, N.J., Hansen, E.H., Kannangara, R., Olsen, C.E., Motawia, M.S., Jørgensen, K., Holme, I., Hebelstrup, K., Grisoni, M., and Møller, B.L. (2014). Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nat. Commun.* *5*, 1–14.
- Giacomoni, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., Duperier, C., Tremblay-Franco, M., Martin, J.-F., Jacob, D., et al. (2015). Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* *31*, 1493–1495.
- Ginglinger, J.-F., Boachon, B., Höfer, R., Paetz, C., Köllner, T.G., Miesch, L., Lugan, R., Baltenweck, R., Mutterer, J., Ullmann, P., et al. (2013). Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in *Arabidopsis* flowers. *Plant Cell* *25*, 4640–4657.
- Guillaumie, S., Ilg, A., Réty, S., Brette, M., Trossat-Magnin, C., Decroocq, S., Léon, C., Keime, C., Ye, T., Baltenweck-Guyot, R., et al. (2013). Genetic Analysis of the Biosynthesis of 2-Methoxy-3-Isobutylpyrazine, a Major Grape-Derived Aroma Compound Impacting Wine Quality. *Plant Physiol.* *162*, 604.
- Gunata, Y.Z., Bayonove, C.L., Baumes, R.L., and Cordonnier, R.E. (1985). The aroma of grapes I. Extraction and determination of free and glycosidically bound fractions of some grape aroma components. *J. Chromatogr. A* *331*, 83–90.
- Guth, H. (1997). Identification of Character Impact Odorants of Different White Wine Varieties. *J. Agric. Food Chem.* *45*, 3022–3026.
- Harrison, P.J., and Bugg, T.D.H. (2014). Enzymology of the carotenoid cleavage dioxygenases: reaction mechanisms, inhibition and biochemical roles. *Arch. Biochem. Biophys.* *544*, 105–111.
- Harsch, M.J., Benkwitz, F., Frost, A., Colonna-Ceccaldi, B., Gardner, R.C., and Salmon, J.-M. (2013). New Precursor of 3-Mercaptohexan-1-ol in Grape Juice: Thiol-Forming Potential and Kinetics during Early Stages of Must Fermentation. *J. Agric. Food Chem.* *61*, 3703–3713.
- Hashizume, K., Tozawa, K., Hiraga, Y., and Aramaki, I. (2001). Purification and characterization of a O-methyltransferase capable of methylating 2-hydroxy-3-alkylpyrazine from *Vitis vinifera* L. (cv. Cabernet Sauvignon). *Biosci. Biotechnol. Biochem.* *65*, 2213–2219.
- Heymann, H., Noble, A.C., and Boulton, R.B. (1986). Analysis of methoxypyrazines in wines. 1. Development of a quantitative procedure. *J. Agric. Food Chem.* *34*, 268–271.
- Huang, A.-C., Sefton, M.A., Sumby, C.J., Tiekink, E.R.T., and Taylor, D.K. (2015). Mechanistic studies on the autoxidation of  $\alpha$ -guaiene: structural diversity of the sesquiterpenoid downstream products. *J. Nat. Prod.* *78*, 131–145.
- Huang, F.-C., Molnár, P., and Schwab, W. (2009). Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. *J. Exp. Bot.* *60*, 3011–3022.

- Hugueney, P., Duchene, E., and Merdinoglu, D. (2012). Brevet international Génoplatte-Valor WO2012/052171 1-deoxy-D-xylulose 5-phosphate synthase alleles responsible for enhanced terpene biosynthesis.
- Ilc, T., Werck-Reichhart, D., and Navrot, N. (2016). Meta-Analysis of the Core Aroma Components of Grape and Wine Aroma. *Front. Plant Sci.* 7.
- Ilc, T., Halter, D., Miesch, L., Lauvoisard, F., Kriegshauser, L., Ilg, A., Baltenweck, R., Hugueney, P., Werck-Reichhart, D., Duchêne, E., et al. (2017). A grapevine cytochrome P450 generates the precursor of wine lactone, a key odorant in wine. *New Phytol.* 213, 264–274.
- Ilc, T., Arista, G., Tavares, R., Navrot, N., Duchêne, E., Velt, A., Choulet, F., Paux, E., Fischer, M., Nelson, D.R., et al. (2018). Annotation, classification, genomic organization and expression of the *Vitis vinifera* CYPome. *PLOS ONE* 13, e0199902.
- INSEE (2019). Industrie agroalimentaire – Tableaux de l'économie française.
- Jackson, D.I., and Lombard, P.B. (1993). Environmental and Management Practices Affecting Grape Composition and Wine Quality - A Review. *Am. J. Enol. Vitic.* 44, 409–430.
- 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.
- Kalua, C.M., and Boss, P.K. (2009). Evolution of Volatile Compounds during the Development of Cabernet Sauvignon Grapes (*Vitis vinifera* L.). *J. Agric. Food Chem.* 57, 3818–3830.
- Kalua, C.M., and Boss, P.K. (2010). Comparison of major volatile compounds from Riesling and Cabernet Sauvignon grapes (*Vitis vinifera* L.) from fruitset to harvest. *Aust. J. Grape Wine Res.* 16, 337–348.
- Kobayashi, H., Takase, H., Suzuki, Y., Tanzawa, F., Takata, R., Fujita, K., Kohno, M., Mochizuki, M., Suzuki, S., and Konno, T. (2011). Environmental stress enhances biosynthesis of flavor precursors, S-3-(hexan-1-ol)-glutathione and S-3-(hexan-1-ol)-L-cysteine, in grapevine through glutathione S-transferase activation. *J. Exp. Bot.* 62, 1325–1336.
- Kotseridis, Y., and Baumes, R. (2000). Identification of Impact Odorants in Bordeaux Red Grape Juice, in the Commercial Yeast Used for Its Fermentation, and in the Produced Wine. *J. Agric. Food Chem.* 48, 400–406.
- Lashbrooke, J.G., Young, P.R., Dockrall, S.J., Vasanth, K., and Vivier, M.A. (2013). Functional characterisation of three members of the *Vitis vinifera* L. carotenoid cleavage dioxygenase gene family. *BMC Plant Biol.* 13, 156.
- Le Paslier, M.-C., Choisne, N., Scalabrin, S., Bacilieri, R., Berard, A., Bounon, R., Boursiquot, J.-M., Bras, M., Brunel, D., Chauveau, A., et al. (2013). The GrapeReSeq 18k *Vitis* genotyping chip. In IX International Symposium on Grapevine Physiology and Biotechnology, (April 21-26 2013, La Serena, Chile), p.
- Lei, Y., Xie, S., Guan, X., Song, C., Zhang, Z., and Meng, J. (2018). Methoxypyrazines biosynthesis and metabolism in grape: A review. *Food Chem.* 245, 1141–1147.

- Lewinsohn, E., Sitrit, Y., Bar, E., Azulay, Y., Ibdah, M., Meir, A., Yosef, E., Zamir, D., and Tadmor, Y. (2005). Not just colors—carotenoid degradation as a link between pigmentation and aroma in tomato and watermelon fruit. *Trends Food Sci. Technol.* *16*, 407–415.
- Liavonchanka, A., and Feussner, I. (2006). Lipoxygenases: Occurrence, functions and catalysis. *J. Plant Physiol.* *163*, 348–357.
- Loscos, N., Hernández-Orte, P., Cacho, J., and Ferreira, V. (2009). Fate of grape flavor precursors during storage on yeast lees. *J. Agric. Food Chem.* *57*, 5468–5479.
- Luan, F., Mosandl, A., Degenhardt, A., Gubesch, M., and Wüst, M. (2006). Metabolism of linalool and substrate analogs in grape berry mesocarp of *Vitis vinifera* L. cv. Morio Muscat: demonstration of stereoselective oxygenation and glycosylation. *Anal. Chim. Acta* *563*, 353–364.
- Mandel, M.A., Feldmann, K.A., Herrera-Estrella, L., Rocha-Sosa, M., and León, P. (1996). CLA1, a novel gene required for chloroplast development, is highly conserved in evolution. *Plant J.* *9*, 649–658.
- Martin, D.M., and Bohlmann, J. (2004). Identification of *Vitis vinifera* (-)-alpha-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry* *65*, 1223–1229.
- Martin, D.M., Toub, O., Chiang, A., Lo, B.C., Ohse, S., Lund, S.T., and Bohlmann, J. (2009). The bouquet of grapevine (*Vitis vinifera* L. cv. Cabernet Sauvignon) flowers arises from the biosynthesis of sesquiterpene volatiles in pollen grains. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 7245–7250.
- Martin, D.M., Aubourg, S., Schouwey, M.B., Daviet, L., Schalk, M., Toub, O., Lund, S.T., and Bohlmann, J. (2010). Functional Annotation, Genome Organization and Phylogeny of the Grapevine (*Vitis vinifera*) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays. *BMC Plant Biol.* *10*, 226.
- Martin, D.M., Chiang, A., Lund, S.T., and Bohlmann, J. (2012). Biosynthesis of wine aroma: transcript profiles of hydroxymethylbutenyl diphosphate reductase, geranyl diphosphate synthase, and linalool/nerolidol synthase parallel monoterpenol glycoside accumulation in Gewürztraminer grapes. *Planta* *236*, 919–929.
- Matarese, F., Scalabrelli, G., and D’Onofrio, C. (2013). Analysis of the expression of terpene synthase genes in relation to aroma content in two aromatic *Vitis vinifera* varieties. *Funct. Plant Biol.* *40*, 552–565.
- Mateo, J.J., and Jiménez, M. (2000). Monoterpenes in grape juice and wines. *J. Chromatogr. A* *881*, 557–567.
- Mathieu, S., Terrier, N., Procureur, J., Bigey, F., and Günata, Z. (2005). A carotenoid cleavage dioxygenase from *Vitis vinifera* L.: functional characterization and expression during grape berry development in relation to C13-norisoprenoid accumulation. *J. Exp. Bot.* *56*, 2721–2731.
- Mathieu, S., Bigey, F., Procureur, J., Terrier, N., and Günata, Z. (2007). Production of a recombinant carotenoid cleavage dioxygenase from grape and enzyme assay in water-miscible organic solvents. *Biotechnol. Lett.* *29*, 837–841.



- May, B., Lange, B.M., and Wüst, M. (2013). Biosynthesis of sesquiterpenes in grape berry exocarp of *Vitis vinifera* L.: evidence for a transport of farnesyl diphosphate precursors from plastids to the cytosol. *Phytochemistry* 95.
- Merdinoglu, D., Schneider, C., Prado, E., Wiedemann-Merdinoglu, S., and Mestre, P. (2018). Breeding for durable resistance to downy and powdery mildew in grapevine. *OENO One* 52, 203–209.
- Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., et al. (2014). The genome of *Eucalyptus grandis*. *Nature* 510, 356–362.
- Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., et al. (2011). Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3530–3535.
- Niinemets, U., Kännaste, A., and Copolovici, L. (2013). Quantitative patterns between plant volatile emissions induced by biotic stresses and the degree of damage. *Front. Plant Sci.* 4, 262.
- OIV (2019). Note de conjecture mondiale - Situation du secteur en 2018.
- Patel, P., Herbst-Johnstone, M., Lee, S.A., Gardner, R.C., Weaver, R., Nicolau, L., and Kilmartin, P.A. (2010). Influence of Juice Pressing Conditions on Polyphenols, Antioxidants, and Varietal Aroma of Sauvignon blanc Microferments. *J. Agric. Food Chem.* 58, 7280–7288.
- Petronilho, S., Coimbra, M.A., and Rocha, S.M. (2014). A critical review on extraction techniques and gas chromatography based determination of grapevine derived sesquiterpenes. *Anal. Chim. Acta* 846, 8–35.
- Pineau, B., Barbe, J.-C., Van Leeuwen, C., and Dubourdieu, D. (2007). Which Impact for  $\beta$ -Damascenone on Red Wines Aroma? *J. Agric. Food Chem.* 55, 4103–4108.
- Pinu, F.R., Jouanneau, S., Nicolau, L., Gardner, R.C., and Villas-Boas, S.G. (2012). Concentrations of the Volatile Thiol 3-Mercaptohexanol in Sauvignon blanc Wines: No Correlation with Juice Precursors. *Am. J. Enol. Vitic.* 63, 407–412.
- Pinu, F.R., Edwards, P.J.B., Jouanneau, S., Kilmartin, P.A., Gardner, R.C., and Villas-Boas, S.G. (2014). Sauvignon blanc metabolomics: grape juice metabolites affecting the development of varietal thiols and other aroma compounds in wines. *Metabolomics* 10, 556–573.
- Podolyan, A., White, J., Jordan, B., and Winefield, C. (2010). Identification of the lipoxygenase gene family from *Vitis vinifera* and biochemical characterisation of two 13-lipoxygenases expressed in grape berries of Sauvignon Blanc. *Funct. Plant Biol.* 37, 767–784.
- Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* 13, 278–289.
- Ribéreau-Gayon, P., Dubourdieu, D., Donèche, B., and Lonvaud, A. (2006). *Handbook of enology* (Chichester, England: John Wiley).

## Références bibliographiques

- Robinson, A.L., Boss, P.K., Solomon, P.S., Trengove, R.D., Heymann, H., and Ebeler, S.E. (2014). Origins of Grape and Wine Aroma. Part 1. Chemical Components and Viticultural Impacts. *Am. J. Enol. Vitic.* *65*, 1–24.
- Rohmer, M., Knani, M., Simonin, P., Sutter, B., and Sahn, H. (1993). Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J.* *295*, 517–524.
- Roland, A., Schneider, R., Razungles, A., and Cavelier, F. (2011). Varietal Thiols in Wine: Discovery, Analysis and Applications. *Chem. Rev.* *111*, 7355–7376.
- Roujou de Boubée, D., Van Leeuwen, C., and Dubourdieu, D. (2000). Organoleptic impact of 2-methoxy-3-isobutylpyrazine on red bordeaux and loire wines. Effect of environmental conditions on concentrations in grapes during ripening. *J. Agric. Food Chem.* *48*, 4830–4834.
- Ruffner, H.P. (1982). Metabolism of tartaric and malic acids in *Vitis*: A review. *Vitis* *21*.
- Ryona, I., Leclerc, S., and Sacks, G.L. (2010). Correlation of 3-Isobutyl-2-methoxypyrazine to 3-Isobutyl-2-hydroxypyrazine during Maturation of Bell Pepper (*Capsicum annuum*) and Wine Grapes (*Vitis vinifera*). *J. Agric. Food Chem.* *58*, 9723–9730.
- Sacks, G.L., Gates, M.J., Ferry, F.X., Lavin, E.H., Kurtz, A.J., and Acree, T.E. (2012). Sensory Threshold of 1,1,6-Trimethyl-1,2-dihydronaphthalene (TDN) and Concentrations in Young Riesling and Non-Riesling Wines. *J. Agric. Food Chem.* *60*, 2998–3004.
- Sala, C., Busto, O., Guasch, J., and Zamora, F. (2004). Influence of vine training and sunlight exposure on the 3-alkyl-2-methoxypyrazines content in musts and wines from the *Vitis vinifera* variety cabernet sauvignon. *J. Agric. Food Chem.* *52*, 3492–3497.
- Schwab, W., Davidovich-Rikanati, R., and Lewinsohn, E. (2008). Biosynthesis of plant-derived flavor compounds. *Plant J.* *54*, 712–732.
- Schwab, W., Fischer, T.C., Giri, A., and Wüst, M. (2015). Potential applications of glucosyltransferases in terpene glucoside production: impacts on the use of aroma and fragrance. *Appl. Microbiol. Biotechnol.* *99*, 165–174.
- Sefton, M.A., Skouroumounis, G.K., Elsey, G.M., and Taylor, D.K. (2011). Occurrence, Sensory Impact, Formation, and Fate of Damascenone in Grapes, Wines, and Other Foods and Beverages. *J. Agric. Food Chem.* *59*, 9717–9746.
- Simkin, A.J., Underwood, B.A., Auldridge, M., Loucas, H.M., Shibuya, K., Schmelz, E., Clark, D.G., and Klee, H.J. (2004). Circadian Regulation of the PhCCD1 Carotenoid Cleavage Dioxygenase Controls Emission of  $\beta$ -Ionone, a Fragrance Volatile of *Petunia* Flowers. *Plant Physiol.* *136*, 3504–3514.
- Strauss, C.R., Wilson, B., Anderson, R., and Williams, P.J. (1987). Development of Precursors of C13 Nor-Isoprenoid Flavorants in Riesling Grapes. *Am. J. Enol. Vitic.* *38*, 23–27.
- Sweetman, C., Sadras, V.O., Hancock, R.D., Soole, K.L., and Ford, C.M. (2014). Metabolic effects of elevated temperature on organic acid degradation in ripening *Vitis vinifera* fruit. *J. Exp. Bot.* *65*, 5975–5988.

- Swiegers, J.H., and Pretorius, I.S. (2007). Modulation of volatile sulfur compounds by wine yeast. *Appl. Microbiol. Biotechnol.* *74*, 954–960.
- Takahashi, S., Zhao, Y., O’Maille, P.E., Greenhagen, B.T., Noel, J.P., Coates, R.M., and Chappell, J. (2005). Kinetic and molecular analysis of 5-epiaristolochene 1,3-dihydroxylase, a cytochrome P450 enzyme catalyzing successive hydroxylations of sesquiterpenes. *J. Biol. Chem.* *280*, 3686–3696.
- Takase, H., Sasaki, K., Shinmori, H., Shinohara, A., Mochizuki, C., Kobayashi, H., Ikoma, G., Saito, H., Matsuo, H., Suzuki, S., et al. (2016). Cytochrome P450 CYP71BE5 in grapevine (*Vitis vinifera*) catalyzes the formation of the spicy aroma compound (–)-rotundone. *J. Exp. Bot.* *67*, 787–798.
- Tamborra, P., Martino, N., and Esti, M. (2004). Laboratory tests on glycosidase preparations in wine. *Anal. Chim. Acta* *513*, 299–303.
- Thibon, C., Cluzet, S., Mérillon, J.M., Darriet, P., and Dubourdieu, D. (2011). 3-Sulfanylhexanol Precursor Biogenesis in Grapevine Cells: The Stimulating Effect of *Botrytis cinerea*. *J. Agric. Food Chem.* *59*, 1344–1351.
- Tominaga, T., Darriet, P., and Dubourdieu, D. (1996). Identification de l’acétate de 3-mercaptohexanol, composé à forte odeur de buis, intervenant dans l’arôme des vins de Sauvignon. *Vitis* *35(4)*, 207–210.
- Tominaga, T., Furrer, A., Henry, R., and Dubourdieu, D. (1998). Identification of new volatile thiols in the aroma of *Vitis vinifera* L. var. Sauvignon blanc wines. *Flavour Fragr. J.* *13*, 159–162.
- Tominaga, T., Baltenweck-Guyot, R., Gachons, C.P.D., and Dubourdieu, D. (2000). Contribution of Volatile Thiols to the Aromas of White Wines Made From Several *Vitis vinifera* Grape Varieties. *Am. J. Enol. Vitic.* *51*, 178–181.
- Winterhalter, P. (1991). 1,1,6-Trimethyl-1,2-dihydronaphthalene (TDN) formation in wine. 1. Studies on the hydrolysis of 2,6,10,10-tetramethyl-1-oxaspiro[4.5]dec-6-ene-2,8-diol rationalizing the origin of TDN and related C13 norisoprenoids in Riesling wine. *J. Agric. Food Chem.* *39*, 1825–1829.
- Winterhalter, P., and Rouseff, R. (2001). Carotenoid-Derived Aroma Compounds: An Introduction. In *Carotenoid-Derived Aroma Compounds*, P. Winterhalter, and R.L. Rouseff, eds. (Washington, DC: American Chemical Society), pp. 1–17.
- Wood, C., Siebert, T.E., Parker, M., Capone, D.L., Elsey, G.M., Pollnitz, A.P., Eggers, M., Meier, M., Vössing, T., Widder, S., et al. (2008). From Wine to Pepper: Rotundone, an Obscure Sesquiterpene, Is a Potent Spicy Aroma Compound. *J. Agric. Food Chem.* *56*, 3738–3744.
- Wright, L.P., Rohwer, J.M., Ghirardo, A., Hammerbacher, A., Ortiz-Alcaide, M., Raguschke, B., Schnitzler, J.-P., Gershenzon, J., and Phillips, M.A. (2014). Deoxyxylulose 5-Phosphate Synthase Controls Flux through the Methylerythritol 4-Phosphate Pathway in *Arabidopsis*. *Plant Physiol.* *165*, 1488–1504.
- Yang, C., Wang, Y., Liang, Z., Fan, P., Wu, B., Yang, L., Wang, Y., and Li, S. (2009). Volatiles of grape berries evaluated at the germplasm level by headspace-SPME with GC-MS.

## Références bibliographiques

Yang, H., Barros-Rios, J., Kourteva, G., Rao, X., Chen, F., Shen, H., Liu, C., Podstolski, A., Belanger, F., Havkin-Frenkel, D., et al. (2017). A re-evaluation of the final step of vanillin biosynthesis in the orchid *Vanilla planifolia*. *Phytochemistry* 139, 33–46.

Zhu, B.-Q., Xu, X.-Q., Wu, Y.-W., Duan, C.-Q., and Pan, Q.-H. (2012). Isolation and characterization of two hydroperoxide lyase genes from grape berries. *Mol. Biol. Rep.* 39, 7443–7455.



# Thuy Thanh TRUONG

## Analyse génétique et fonctionnelle de la biosynthèse des arômes des raisins

### Résumé

Malgré l'importance des molécules aromatiques des raisins dans la qualité et la typicité des vins, les déterminants de leur biosynthèse sont encore mal connus. L'objectif de ma thèse est d'identifier les régions génomiques et les gènes clés de la biosynthèse des arômes de la baie de raisin. Pour cela, j'ai combiné différentes approches de génétique, de métabolomique ciblée et non ciblée et de génomique fonctionnelle. Différents outils analytiques (GC-MS, UHPLC-MS et FT-ICR) m'ont permis de quantifier plusieurs milliers de molécules pouvant participer de manière directe ou indirecte aux arômes des baies et des vins. Des loci à effets quantitatifs (Quantitative trait loci, QTL) associés à l'accumulation d'un grand nombre de ces métabolites dans les raisins ont été détectés dans la descendance d'un croisement entre Riesling et Gewurztraminer. Des gènes candidats liés à des QTL ont été identifiés et j'ai initié la caractérisation fonctionnelle de certains d'entre eux. À terme, une meilleure connaissance des marqueurs génétiques liés aux arômes des raisins permettra d'augmenter l'efficacité des programmes de création de nouvelles variétés de vigne.

Mot-clés : Vigne, raisins, arômes, métabolisme secondaire, Riesling, Gewurztraminer, QTL

### Summary

Despite the importance of grape aroma compounds in the quality and typicity of wines, the determinants of their biosynthesis are still poorly known. The goal of my thesis is to identify genomic regions and key genes for the biosynthesis of grape berry aromas. For this, I combined different approaches including genetics, targeted and non-targeted metabolomics and functional genomics. Different analytical tools (GC-MS, UHPLC-MS and FT-ICR) allowed me to quantify several thousand molecules that can participate directly or indirectly in the aroma of berries and wines. Quantitative trait loci (QTLs) associated with the accumulation of a large number of these metabolites in grapes were detected in a progeny from a cross between Riesling and Gewurztraminer. Candidate genes linked to QTLs have been identified and I have initiated the functional characterization of some of them. Ultimately, a better knowledge of the genetic markers related to grape aroma characteristics will increase the efficiency of the breeding programs for new grapevine varieties.

Key-words : Vine, Grapes, Flavors, Secondary Metabolism, Riesling, Gewurztraminer, QTL