UNIVERSITÉ DE STRASBOURG

École doctorale des sciences de la vie et de la santé

Institut de Biologie Moléculaire des Plantes (CNRS)

THÈSE présentée par :

Tina ILC

soutenue le : **18 décembre 2015**

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

Discipline : Aspects moléculaires et cellulaires de la biologie

---

Role of cytochromes P450 in wine aroma

THÈSE dirigée par :

Dr WERCK-REICHHART Danièle Directeur de recherche, CNRS, IBMP

RAPPORTEURS :

Prof BAUDINO Sylvie Professeur, Université de Lyon Saint-Étienne

Prof WÜST Matthias Professeur, Universität Bonn

EXAMINATEUR :

Dr SCHALLER Hubert Directeur de recherche, CNRS, IBMP

Table of Contents

Acknowledgements

I would like to thank Daniele Werck for initiating this project and giving me the great opportunity to do my PhD thesis under her supervision. Thank you for your trust, for the opportunity to bring my own ideas to the project and for your continuous availability for discussion. I also wish to thank Nicolas Navrot for his day-to-day supervision, support and encouragement, as well as all of his help when I came to Strasbourg 3 years ago.

I am grateful to Prof Sylvie Baudino, Prof Matthias Wüst and Dr Hubert Schaller for accepting to participate in my jury.

This work would not be possible without strong collaborations. I am grateful to all the collaborators from INRA Colmar, for teaching me a lot about grape and aroma in particular, for the many useful discussions, as well as the practical help: Philippe Hugueney, Camille Rustenholz, Eric Duchene, Gautier Arista, Gisele Butterlin, Andrea Ilg and Marc Fischer. I want to thank Laurence Miesch and her team for synthesis of standards, as well as for many interesting discussions. I am also grateful to Raquel Tavares from University of Lyon for the collaboration on phylogeny.

I am grateful to all the members of Marie Curie ITN "P4fifty". It was a great pleasure to take part in this network and to get to know so many excellent scientists. In particular, I would like to thank everyone who welcomed me for one of my secondments: Fanny Lambert (Mane S.A.); Sabine Flitsch, Slavomira Husarova (University of Manchester), Bernhard Hauer, Stefan Hammer and Silke Bastian (University of Stuttgart). I also wish to thank my mentor Birger Lindberg Møller for many interesting discussions and useful guidelines. Special thanks also to Margaret Cafferky for administrating the network and being always available for help. I also thank the European Union for funding the "P4fifty" project.

I want to thank the current and former colleagues from the "Cytochrome P450" team: Francois Bernier, Benoit Boachon, Carole Gavira, Thierry Heitz, Lucie Kriegshauser, Agnès Lesot, Rozenn Menard, Fabienne Philippon, Emmanuelle Pineau, Franck Pinot, Hugues Renault, Pascaline Ullman, Gaetan Verdier, and Emilie Widemann. Thank you all for your help, all the discussions, useful tips and challenging questions. My very special thanks go to Annette Alber, Claire Parage, Juliana Iglesias and Zhenhua Liu for always taking time for my scientific as well as "real-life" problems. I also wish to thank everyone from the Metabolomics platform (Dimitri Heintz, Julien Delecolle, Arnaud Agin, Maximilien Nuel, Julie Zumsteg and Raphael Lugan) and Gene expression platform (Malek Alioua) for your help and training.

I want to thank my family for teaching me the importance of education in my mind from an early age. Thanks to my grandma for bringing me up to be so stubborn. I also want to thank

my friends in Slovenia for not letting the distance come between us. Finally, a massive thank you to my big “Strasbourg family” who have supported me throughout this PhD and made Strasbourg my home.

Preface

This work was part of the Marie Curie Initial Training Network "P4fifty", the goal of which was the application of cytochrome P450 enzymes for white biotechnological processes. My principal role in this network was the discovery of P450 enzymes with novel functions. Wine is a rich source of chemically diverse aroma compounds, yet the underlying enzymes are largely unknown. The aim of this work is therefore to identify cytochromes P450 that synthesize aroma in grapes. These cytochromes P450 could be used for industrial production of flavor and fragrance compounds. In addition, grapevine is one of the most important crops in France and worldwide and aroma is of crucial importance for the quality of the final product. The knowledge of aroma biosynthetic pathways could thus also help to improve the quality of grapes and wines. This work was also supported by the funding of collaboration with the team of Dr Philippe Hugueney of the French National Institute for Agricultural Research (INRA) in Colmar by ANR (the French National Agency of Research).

In the first part of my thesis, I review the current knowledge on wine aroma and its biosynthesis in the form of a meta-analysis of 19 published datasets of grape and wine aroma profiles. The objective of this work was to identify aroma compounds that are possibly synthesized in grapes by cytochromes P450. This reviewing and compiling study, presented here as **Chapter 1**: Meta-analysis of grape and wine aroma (to be submitted to Journal of Agricultural and Food Chemistry), revealed an elaborate oxidative metabolism of the monoterpenol linalool, and convinced us to focus our efforts on biotransformations of this compound in grape and wine.

After briefly introducing biosynthesis of terpenoids (**Chapter 2**) and cytochromes P450 (**Chapter 3**), I review in detail recent progress in oxidations of monoterpenols by cytochromes P450 in **Chapter 4**: Cytochrome P450-catalyzed oxidations of monoterpenols in plants (invited publication in Frontiers in Plant Science).

In collaboration with the INRA group, we built an exhaustive and reliable annotation of cytochrome P450 genes in the reference grapevine genome. We then used this annotation to study the expression of these genes using publically available and INRA's newly-generated transcriptomic data (**Chapter 5**: Annotation, classification, genomic organization and expression of the *Vitis vinifera* CYPome, to be submitted to BMC Genomics). This allowed us to identify several gene candidates for biosynthesis of aroma in grapes. In addition, it pointed to other cytochromes P450 of interest, involved for instance in the plant interactions with pathogens, and will provide the argument for future projects.

The second part of the results and main experimental section addresses the formation of an important wine aroma compound, the wine lactone (**Chapter 6**: CYP76F14 catalyzes

biosynthesis of wine lactone precursor from linalool, to be submitted to New Phytologist). I first show that this potent aroma is formed during wine ageing from a grapevine-derived precursor, (*E*)-8-carboxylinalool. I then demonstrate that (*E*)-8-carboxylinalool is synthesized from linalool by the grapevine cytochrome P450 CYP76F14. The CYP76F14 gene is highly expressed in grape berries at late ripening stages, indicating that it is likely the dedicated enzyme for production of (*E*)-8-carboxylinalool in grapes.

In **Chapter 7** (Additional results) I introduce the results that were obtained in other parts of the project and were not included in the preceding chapters. These include a study of grape genes expressed in the early stages of berry development, a study of grapevine cytochromes P450 from other families, and a sensory study showing importance of wine lactone in white wine.

In the final chapter (**Chapter 8**) I discuss the conclusions and perspectives of my work, including perspectives in the wine aroma research and flavor and fragrance biotechnology.

Résumé en français

L'arôme est l'un des principaux critères de qualité du vin. Il est donc très important de comprendre l'origine et l'évolution des arômes du vin dans le but de contrôler et de prédire leur qualité. Bien que de nombreux composés aromatiques apparaissent lors de la fermentation par les levures, les composés aromatiques les plus puissants et les plus caractéristiques des cépages sont formés dans la baie de raisin (*Vitis vinifera*). Un grand nombre de ces composés appartiennent à la famille des monoterpénoïdes, qui sont des molécules isoprèniques à 10 atomes de carbone. Les monoterpénoïdes participent à l'arôme et au parfum de nombreuses plantes.

Les monoterpènes sont synthétisés par des enzymes appartenant à la famille des terpènes synthases. Un certain nombre d'entre elles ont déjà été caractérisées chez la vigne. Une analyse bibliographique systématique a par ailleurs montré qu'une grande partie des monoterpènes trouvés chez la vigne et dans le vin sont des dérivés oxydés du linalool. Le dérivé de linalool dont l'arôme est le plus puissant est la lactone du vin (wine lactone), une molécule qui présente une odeur sucrée de noix de coco. D'après la littérature, la wine lactone se forme par une réaction non-enzymatique à partir du précurseur (*E*)-8-carboxylinalool au cours du vieillissement du vin.

Chez de nombreuses plantes, comme *Arabidopsis thaliana* ou *Catharanthus roseus*, les monoterpènes sont oxydés par les cytochromes P450 et en particulier les membres des familles CYP71 et CYP76. Les cytochromes P450, contrairement aux terpènes synthases, n'ont pas été décrits en détail chez la vigne. L'objectif de ma thèse est d'étudier le rôle des cytochromes P450 dans la biosynthèse des arômes du raisin, et plus particulièrement la formation de la wine lactone et de son précurseur le (*E*)-8-carboxylinalool.

La partie expérimentale principale de ma thèse a consisté à étudier la formation de la wine lactone et de son précurseur au cours de la maturation des raisins et la vinification, au travers de l'identification et de la caractérisation des gènes impliqués dans la biosynthèse de ces composés. Les concentrations de la wine lactone dans les vins sont très faibles et, en conséquence, cette molécule n'est pas détectée par les méthodes habituellement utilisées pour l'analyse des arômes du vin. J'ai donc développé une méthode d'extraction et de détection de la wine lactone par LC-MS/MS qui m'a permis de quantifier ce composé dans tous les échantillons de vins analysés. Par LC-MS/MS j'ai également pu quantifier le (*E*)-8-carboxylinalool et son ester de glucose. L'analyse de ces trois composés dans différents échantillons représentatifs des vins d'Alsace démontre que les quantités de ces trois composés sont fortement corrélées (Figure 1). Alors que les concentrations en ester de glucose de (*E*)-8-carboxylinalool sont élevées dans les vins jeunes (1 ou 2 ans), l'ester est complètement hydrolysé en (*E*)-8-carboxylinalool dans les vins plus vieux. Les

concentrations en wine lactone sont plus élevées dans les vins vieux, de même que dans les vins où la concentration en (*E*)-8-carboxylinalool est élevée. La conversion du (*E*)-8-carboxylinalool en wine lactone suit une réaction du premier ordre avec un taux de conversion très faible de 0,00012 an^{-1}. Cette valeur peut nous permettre d'estimer la concentration en wine lactone d'un vin à partir de la concentration en précurseur et de l'âge du vin.

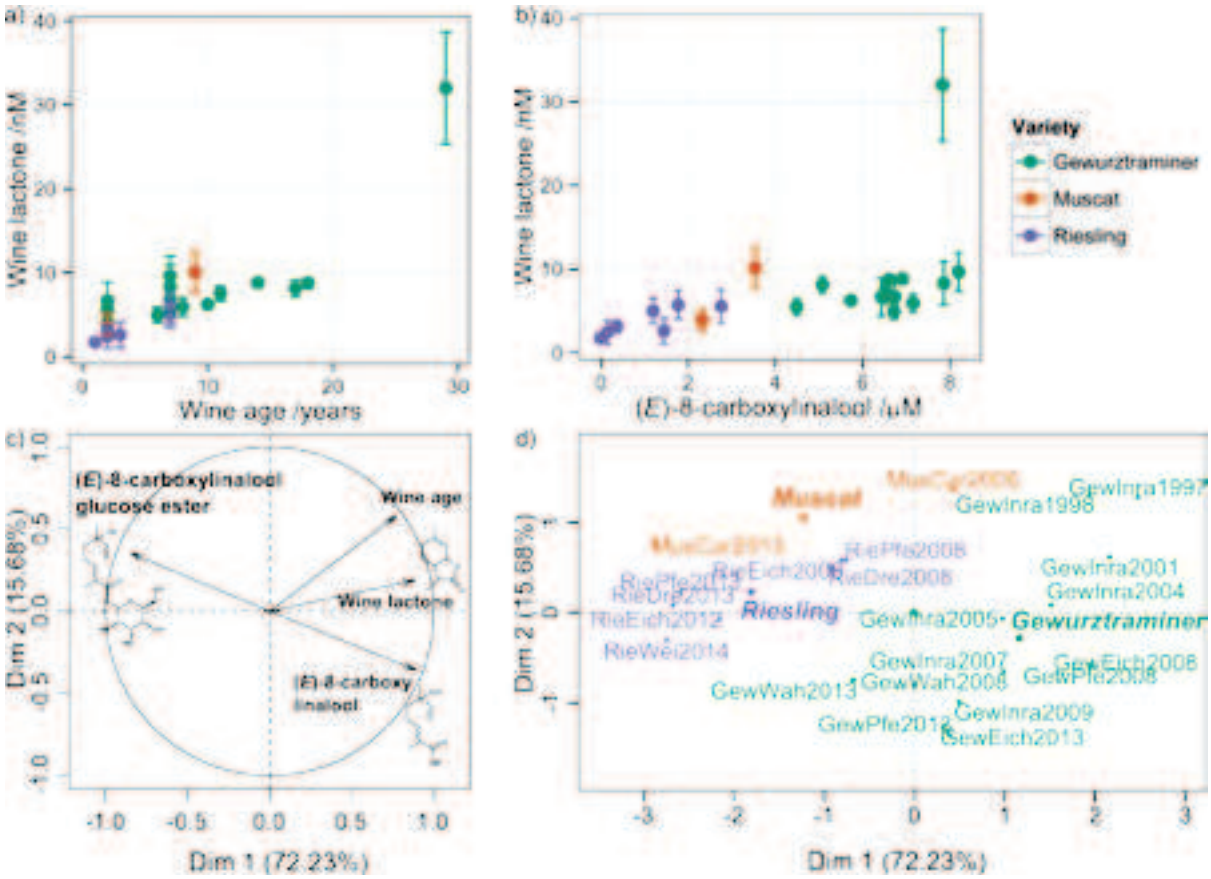

Figure 1. Relations entre l'âge du vin et les concentrations en wine lactone, (*E*)-8-carboxylinalool et (*E*)-8-carboxylinalool glucose ester. Les trois composés ont été quantifiés dans les mêmes vins par LC-MS/MS (N=3). La concentration en wine lactone est élevée dans les vins âgés (a) ainsi que dans les vins présentant une concentration élevée de (*E*)-8-carboxylinalool. L'analyse en composantes principales (ACP) des concentrations de métabolites et de l'âge du vin confirme cette observation : plan factoriel des variables (c) et des individus (d). Mus = Muscat, Rie = Riesling, Gew = Gewurztraminer. Le vin vieux de 29 ans est exclu de l'ACP.

Il a été déjà suggéré que le précurseur de la wine lactone, le (*E*)-8-carboxylinalool, est synthétisé par oxydation enzymatique du linalool. Plus récemment, les cytochromes P450, particulièrement les familles CYP76 et CYP71 ont été décrits comme participant au métabolisme des monoterpénols chez *Arabidopsis thaliana* (Figure 2) et d'autres espèces végétales. Un préalable nécessaire à la caractérisation fonctionnelle d'une famille de gènes est une annotation correcte des gènes. Les annotations automatiques peuvent contenir des erreurs, des artefacts ou encore des omissions. Cette partie de mon travail a eu donc pour objectif une annotation complète et correcte des cytochromes P450 à partir des données de séquençage d'un cultivar de référence de vigne Pinot Noir (*V. vinifera* PN40024). Après

avoir effectué une recherche grâce à l'algorithme BLAST en utilisant des gènes de P450 connus et vérifié la présence des séquences obtenues dans les données génétiques publiées, tous les gènes ont été manuellement vérifiés et classés en gènes, gènes partiels, pseudogènes, et pseudogènes putatifs.

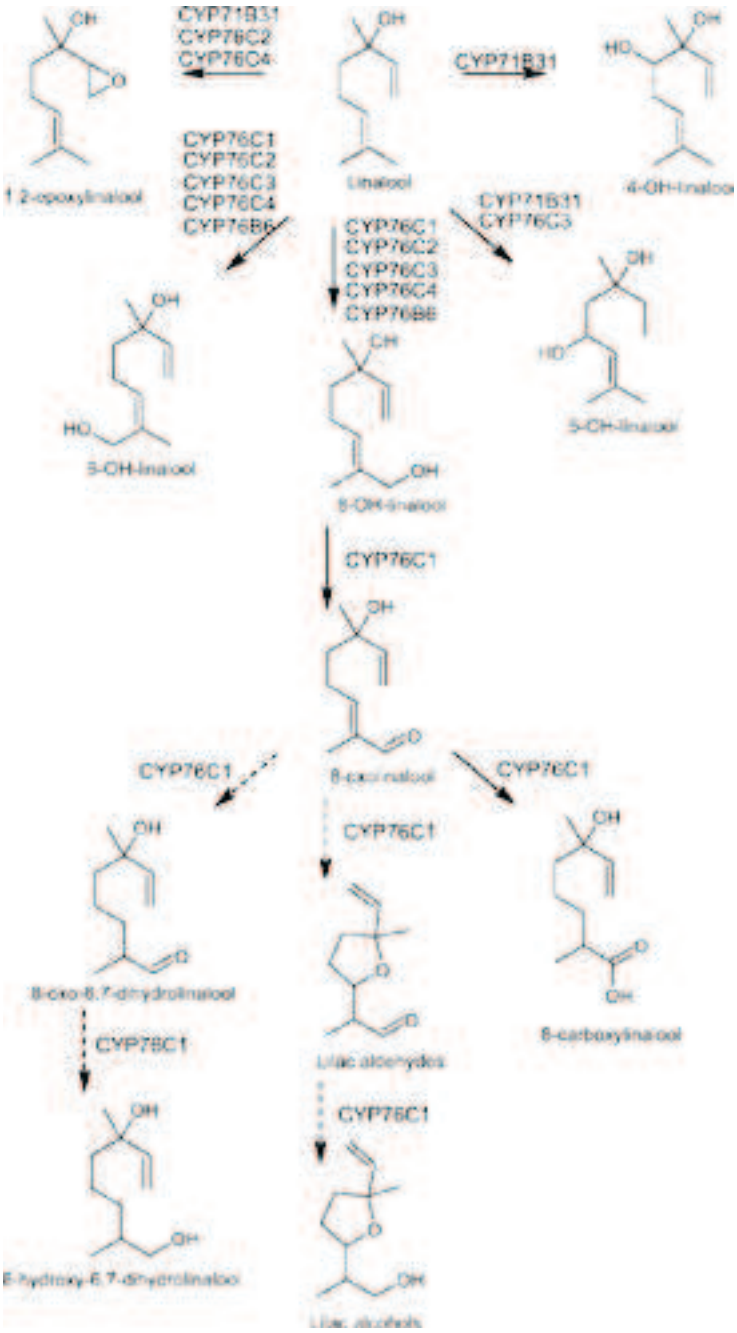

Figure 2. Le métabolisme du linalool par les cytochromes P450 chez *Arabidopsis thaliana*. Les enzymes recombinantes ont été produites dans la levure ou dans les feuilles de *N. benthamiana*. Les flèches en pointillés indiquent les étapes qui pourraient nécessiter d'autres enzymes non décrites.

Cette annotation a révélé 279 gènes complets, 20 gènes partiels, 11 pseudogènes putatifs et 227 pseudogènes de cytochromes P450 chez le cultivar séquencé. Souvent ces gènes forment des « clusters » qui regroupent jusqu'à 35 séquences (Figure 3). Ces clusters regroupent des gènes très similaires qui généralement appartiennent à la même famille de cytochromes

P450 (Figure 4). En conséquence, la plupart des gènes qui font partie d'un cluster ont leur homologue le plus proche au sein du même cluster. Ces clusters ont été probablement formés par duplication en tandem et dans certains cas par duplication segmentale. La famille CYP82, qui est formée de 25 gènes complets, constitue la plus vaste famille de cytochromes P450 chez la vigne.

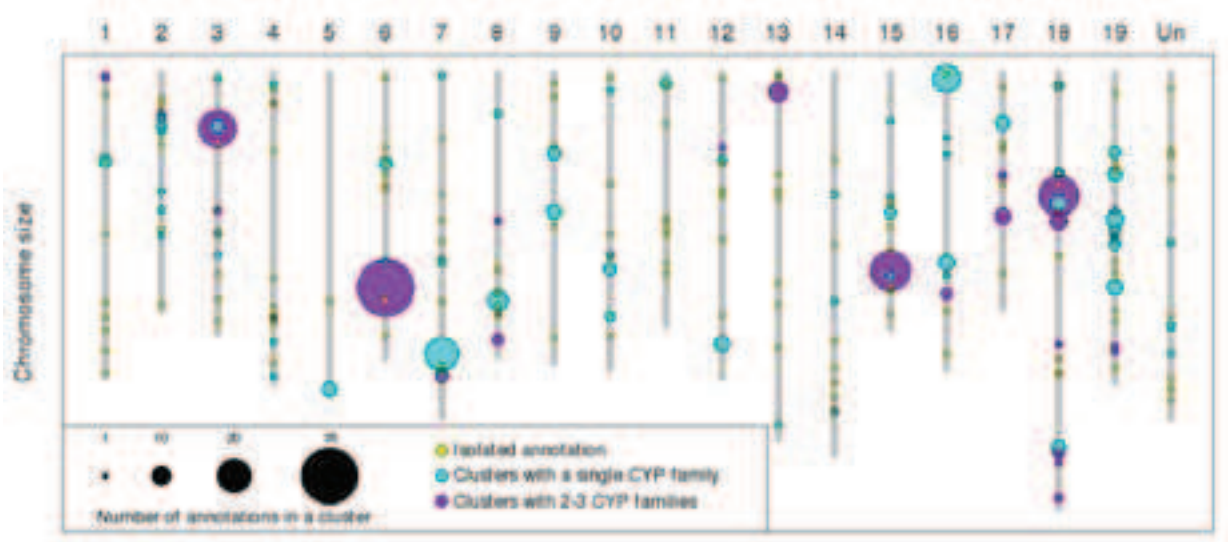

Figure 3. Carte physique des séquences de cytochromes P450 sur les 19 chromosomes de *V. vinifera*. Les cercles jaunes représentent les annotations isolées, les cercles bleus clairs représentent les clusters physiques composés de membres d'une seule famille de P450 et les cercles violets les clusters physiques composés de membres de 2-3 familles de P450. La taille du cercle est proportionnelle au nombre de séquences du cluster. Les numéros 1-19 notent les différents chromosomes et « Un » est le « chromosome inconnu » qui contient des séquences dont l'emplacement chromosomique est inconnu.

Des données de séquençage d'ARN ont ensuite été comparées à cette annotation. Elles ont permis de mettre en évidence les gènes candidats exprimés dans la baie en fin de maturation, ainsi que dans les feuilles infectées par des agents pathogènes comme le mildiou et l'oïdium. Cette analyse d'expression a montré qu'une grande partie des gènes de P450s étaient activés par l'infection par des agents pathogènes et suggère que ce derniers pourraient être impliqués dans la biosynthèse de composés de défense.

De nombreux gènes de cytochromes P450 sont surexprimés dans les baies mûres, parmi eux les gènes de la famille CYP75, qui participent à la pigmentation (biosynthèse des flavonoïdes). Une attention toute particulière a été portée aux différentes familles susceptibles de contribuer aux propriétés aromatiques du raisin et vin, plus particulièrement les familles CYP76 et CYP71 qui ont récemment été décrites comme participant au métabolisme des monoterpénols chez autres espèces végétales (Figure 2). Six gènes candidats fortement exprimés dans les baies mûres ont été sélectionnés : *CYP76F12, CYP76F14, CYP76T21, CYP76Y1, CYP76Y2* et *CYP71AT7*. Le niveau d'expression de ces gènes a ensuite été évalué par qRT-PCR à cinq stades de développement de la baie. L'accumulation de (*E*)-8-carboxylinalool, libre ou conjugué a été quantifiée dans les mêmes échantillons. Les

candidats pour une analyse fonctionnelle ont ensuite été sélectionnés en comparant profils d'expression et l'accumulation du (*E*)-8-carboxylinalool.

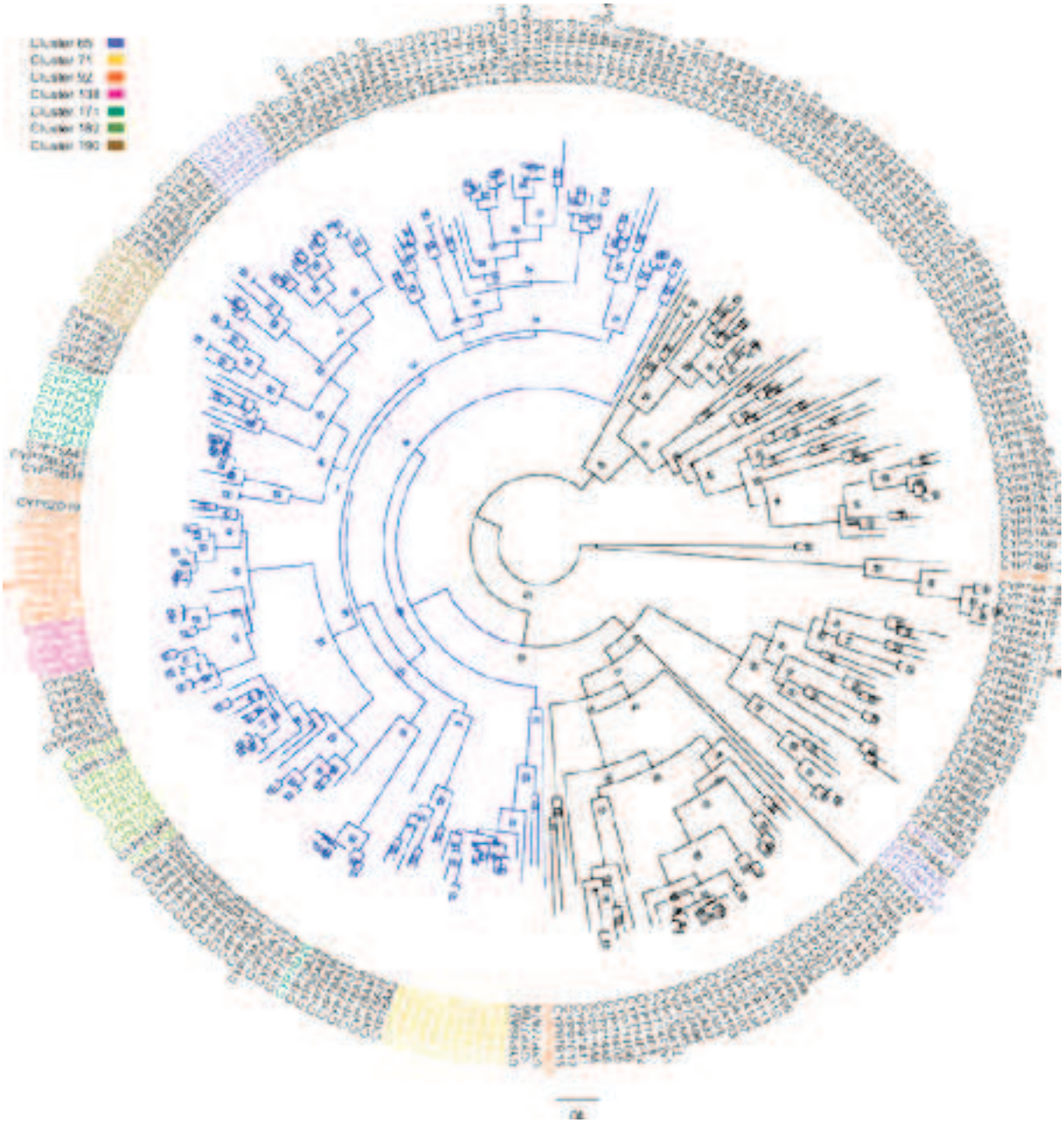

Figure 4. Analyse phylogénétique moléculaire des cytochromes P450 de la vigne. L'alignement des séquences de protéines a été utilisé pour générer un arbre inféré au maximum de vraisemblance. Le clade bleu foncé est le clan 71 qui contient souvent des gènes impliqués dans le métabolisme spécialisé. Les gènes colorés appartiennent aux sept plus grands clusters physiques.

Le (*E*)-8-carboxylinalool s'accumule sous forme conjuguée dans la baie de raisin. La concentration en glucose ester de (*E*)-8-carboxylinalool dans les baies de Gewurztraminer augmente durant la maturation de la baie de raisin. L'analyse de l'expression des gènes dans la baie durant la maturation m'a permis de mettre en évidence 5 gènes de la famille CYP76 et un de la famille CYP71 dont l'expression augmente également durant la maturation. Plus

précisément, l'expression de ces gènes est activée lors des stades tardifs du développement, en particulier au stades mi-mature et mature. Le candidat le plus exprimé est *CYP76F14*.

Les candidats ont été clonés et exprimés dans la souche WAT11 de *Saccharomyces cerevisiae*, qui exprime une P450 réductase d'origine végétale (*A. thaliana*). Les cytochromes P450 sont des enzymes ancrées sur les membranes du réticulum endoplasmique. J'ai donc utilisé les membranes microsomales isolées de ces levures recombinantes pour effectuer les tests d'activité *in vitro*. Les produits d'oxydation du linalool ont été analysés par chromatographie en phase gazeuse et liquide couplée à la spectrométrie de masse (GC-MS et LC-MS/MS). Afin d'évaluer l'activité des candidats *in planta*, ces derniers ont aussi été exprimés transitoirement dans les feuilles de *Nicotiana benthamiana*. Pour compenser la faible production de linalool par cette plante, un gène de linalool synthase de vigne a été co-exprimé pour former le substrat.

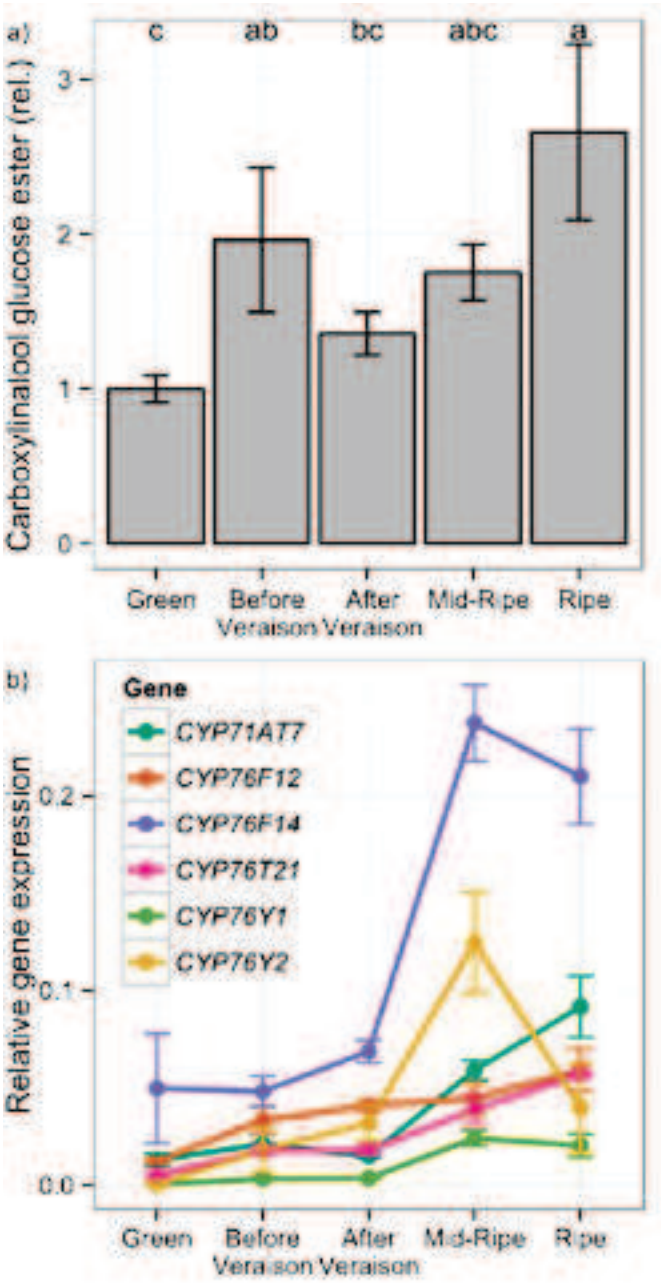

Figure 5. Evolution de la concentration en ester de glucose de (*E*)-8-carboxylinalool et de l'expression des gènes candidats pendant la maturation de baies de raisin. a) Concentration relative (± écart type) du glucose ester de (*E*)-8-carboxylinalool pendant le développement des baies de Gewurztraminer. L'analyse de variance monofactorielle a montré des différences significatives (p=0,0016) entre les stades de développement (« Green » – baies vertes, « Before veraison » – avant véraison, « After veraison » – après véraison, « Mid-ripe » – baies mi-matures, « Ripe » – baies mûres). Les lettres indiquent des différences significatives entre les moyennes calculées par le test de Tukey HSD (α = 0,05). b) Expression des gènes candidats dans les baies de Gewurztraminer, mesurée par RT-PCR quantitative. Les barres d'erreur représentent l'écart type de 3 échantillons. L'expression des gènes candidats est normalisée par rapport à l'expression de 3 gènes constitutifs.

J'ai pu montrer que trois des enzymes candidates exprimées dans la levure, CYP76F12, CYP76F14 et CYP76T21, métabolisent le linalool *in vitro*. Toutes catalysent l'oxydation au niveau du carbone terminal en position (*E*)-8, et l'une d'entre elle en position (*Z*)-8. Cependant, les enzymes diffèrent en ce qui concerne le degré d'oxydation des produits formés : seule l'une des enzymes, CYP76F14, peut catalyser la triple oxydation du linalool en

(*E*)-8-carboxylinalool, le précurseur de la wine lactone (Figure 6). La conversion de (*E*)-8-hydroxylinalool (alcool) en (*E*)-8-oxolinalool (aldéhyde) semble être l'étape cruciale qui fait de CYP76F14 l'enzyme la plus efficace pour la formation de (*E*)-8-carboxylinalool. Le voie de biosynthèse du (*E*)-8-carboxylinalool a aussi été reconstituée *in planta*. J'ai pu montrer que les feuilles de *N. benthamiana* exprimant la linalool synthase et CYP76F14 produisaient du (*E*)-8-carboxylinalool sans accumulation de produits intermédiaires. En revanche, les feuilles exprimant la linalool synthase et CYP76T21 produisent à la fois du (*E*)-8-carboxylinalool et du (*E*)-8-hydroxylinalool, tandis que l'expression de CYP76F12 ne permet de détecter aucune trace de (*E*)-8-carboxylinalool.

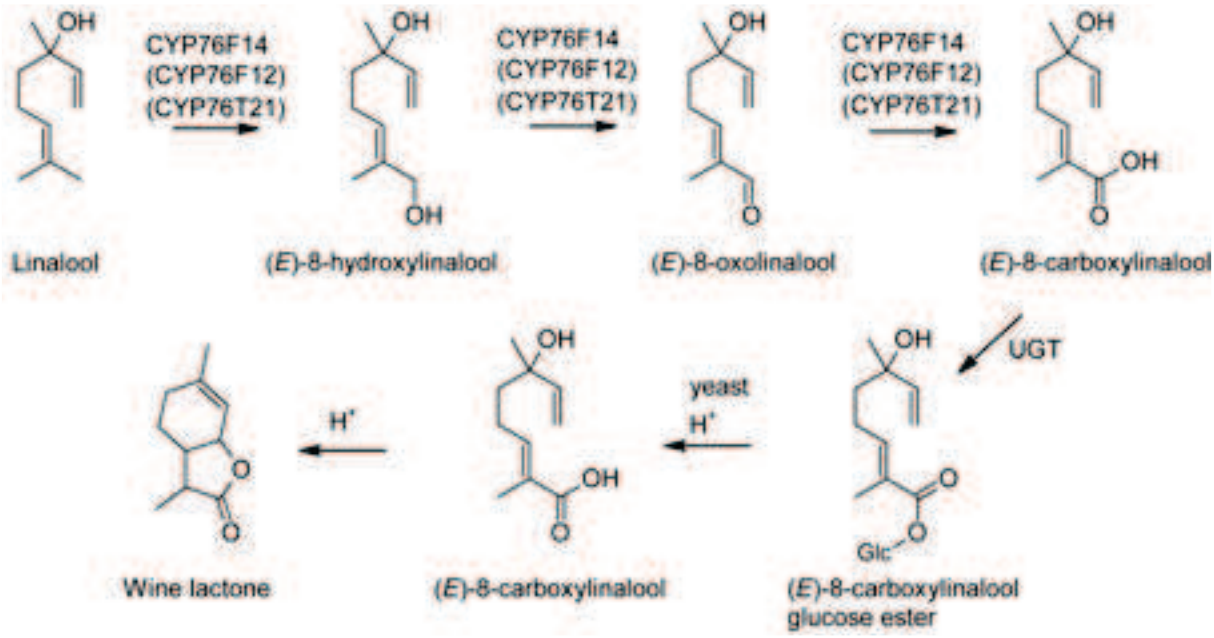

Figure 6. Voie de (bio)synthèse proposée pour la wine lactone. Les quatre premières étapes sont catalysées par les enzymes de la baie de raisin. Les deux dernières étapes ont lieu lors de la vinification et de la maturation du vin.

Les (*E*)-8-hydroxy- et (*E*)-8-carboxylinalool ne sont que deux des dérivés oxygénés de linalool qui peuvent être détectés dans la baie de raisin (Figure 7). Cependant, aucune des autres candidats n'a permis d'obtenir d'autres dérivés, comme par exemple les oxydes de linalool. Quelques études ont montré que certains de ces composés sont synthétisés en début de maturation du raisin. Deux gènes de la famille CYP76, exprimés dans les baies vertes, ont été identifiés et analysés pour mettre en évidence leur possible implication dans la formation de ces molécules. Tous deux catalysent en fait le même type de réaction que les autres candidats exprimés dans les baies mûres : l'oxydation du carbone terminal (position 8). Les autres dérivés de linalool sont donc probablement synthétisés par des membres d'autres familles de cytochromes P450. Quatre enzymes d'autres familles, CYP80E3, CYP81Q19, CYP82D13 et CYP89A41 ont été testées *in vitro*. Pour aucune d'entre elles une métabolisation du linalool ou d'autres monoterpenols n'a pu être détectée. L'une de ces

enzymes, CYP82D13, montre une mutation rare de l'acide aminé le plus conservé chez les cytochromes P450, la cystéine impliquée dans la liaison avec l'hème, qui est remplacée par une tyrosine. Certains mutants artificiels de cystéine en tyrosine ont déjà montré une activité enzymatique, mais si *V. vinifera* CYP82D13 est catalytiquement actif, il sera le premier cas d'un cytochrome P450 naturel qui présente une telle mutation.

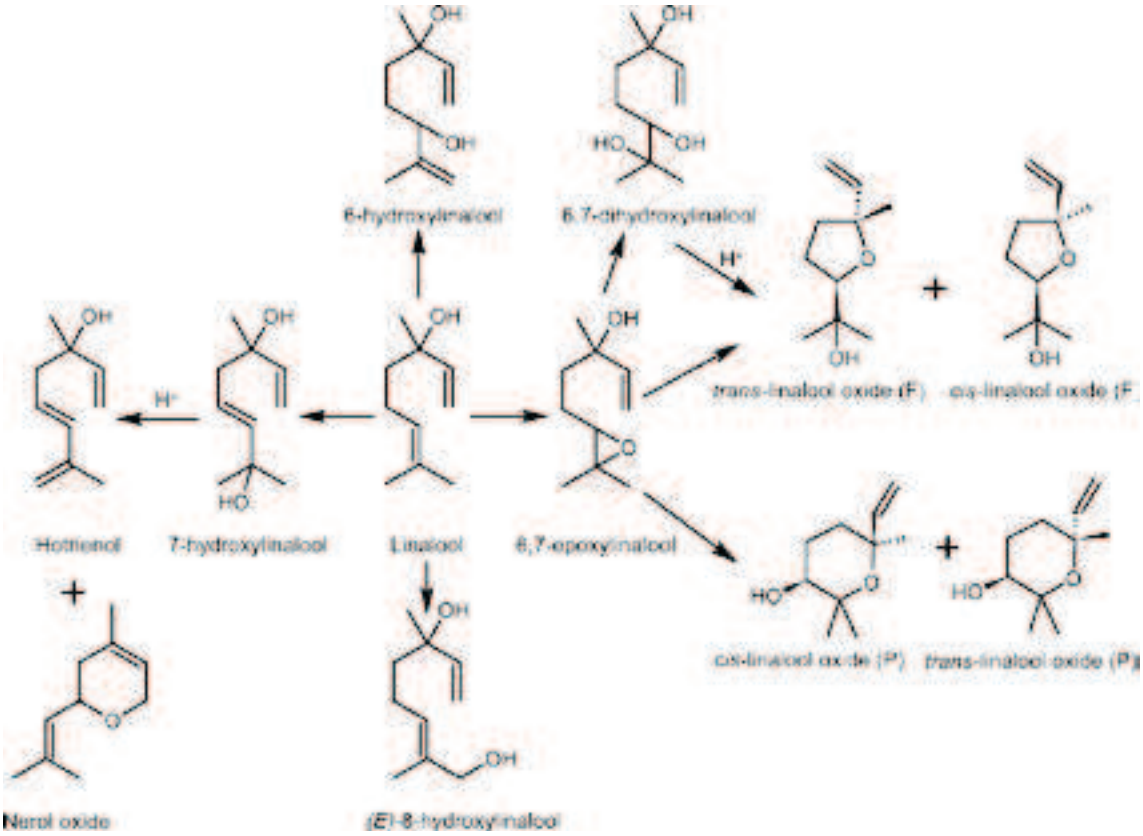

Figure 7. Métabolisme oxydatif du linalool dans les baies de *V. vinifera* cv. Morio Muscat. Le « Feeding » de baies de raisin avec un analogue deutéré de linalool a permis d'identifier de nombreux produits de son métabolisme oxydatif (Luan *et al.* Analy Chim Acta 2006).

En conclusion, les cytochromes P450 jouent un rôle important dans la formation des arômes du vin. L'enzyme CYP76F14 catalyse l'oxydation du linalool en (*E*)-8-carboxylinalool. Cette molécule est ensuite lentement convertie en wine lactone, une molécule odorante puissante, au cours du vieillissement du vin. Ce gène et ses paralogues pourront être utilisés comme marqueurs pour la sélection de nouvelles variétés de vigne présentant un profil aromatique souhaité. De plus, ces enzymes ou les levures recombinantes qui les expriment pourront être utilisées pour produire des arômes par des procédés biotechnologiques. Cependant, la conversion chimique du (*E*)-8-carboxylinalool en wine lactone dans des conditions ambiantes est très lente. Deux solutions sont possibles pour accélérer sa formation: soit réaliser la réaction dans des conditions plus drastiques de température et (ou) de pH, soit identifier des enzymes qui permettent cette cyclisation. Enfin, le (*E*)-8-carboxylinalool

pourrait être utilisé comme un arôme à libération lente dans les cosmétiques ou autres produits.

Cette étude a, d'autre part, démontré l'importance d'une analyse ciblée des composants de l'arôme du vin. Une méta-analyse de l'arôme du vin et du raisin m'a permis de montrer que, même si l'analyse non ciblée permet la quantification d'un grand nombre de composés volatils, de nombreux composés importants, tels que la wine lactone, ne sont pas détectés en raison de leur faible concentration. J'ai, par conséquent, développé une méthode ciblée pour l'analyse de la wine lactone qui m'a permis de mettre en évidence la lente transformation du (*E*)-8-carboxylinalool en wine lactone, ainsi que les différences de concentration de ces composés entre les cépages.

Enfin, l'annotation fine des gènes de cytochrome P450 dans génome de la vigne ouvre la voie à une exploration des autres rôles des cytochromes P450 dans la résistance ou le développement de la vigne et la qualité du vin, par exemple dans la résistance aux pathogènes, la taille ou la pigmentation de la baie. L'analyse de l'expression des gènes a d'ores et déjà mis en évidence plusieurs gènes de cytochromes P450 qui sont activés dans les feuilles infectées par les pathogènes.

Chapter 1
Meta-analysis of grape and wine aroma

Tina Ilc, Danièle Werck-Reichhart, Nicolas Navrot

Institute of Plant Molecular Biology, Centre National de la Recherche Scientifique, University of Strasbourg, Strasbourg, France

Abstract

Wine aroma strongly influences wine quality, yet its composition and evolution during the winemaking process are poorly understood. Volatile compounds that constitute wine aroma are traditionally divided into three classes according to their origin: grape, fermentation and maturation aroma. We challenge this view with meta-analysis and review of grape and wine volatiles and their precursors from 82 profiling experiments. We compiled a list of 141 common grape and wine volatiles and quantitatively compared 43 of them. Our work offers insight into complex relationships between biosynthesis of aroma in grapes and the changes during the winemaking process. With a particular focus on monoterpenes we show that the diversity of these compounds in wines is mainly due to the oxidative metabolism of linalool. Furthermore, we demonstrate that most of the linalool produced in grapes is converted to these oxidized derivatives.

Introduction

Aroma is a crucial determinant of the wine quality, but we are still far from understanding its composition, let alone the influence of genetic factors and environment on its development. Yet we would ultimately want to understand and control wine aroma to ensure consistent production of high quality wines (1). Wine aroma is a complex mixture of volatile organic compounds. These are small, nonpolar molecules that readily enter the gas phase and reach our nasal cavity while we smell or drink a glass of wine. Different volatile organic compounds in wine span a large range of concentrations. While it is tempting to think that the more abundant compounds impact the aroma more than the trace compounds, it is not necessarily the case. We can smell some of the compounds at very low concentrations, whereas others remain undetected even at high levels. The strength of aroma of a particular compound is expressed as an odor detection threshold, which is the lowest concentration perceivable to the human smell. Compounds with low sensory thresholds are often responsible for the characteristic smell of a particular food (2).

Wine aroma compounds differ in their origin and evolution during the winemaking process. Many authors have classified wine aroma compounds into three categories based on their origin: grape (or varietal) aroma, fermentation aroma and ageing aroma, also called the wine bouquet (3–5). However, these three classes are not so clear-cut: ultimately most of the aroma precursors (even the simple ones) originate from grapes and are in some way modified by the fermentation process or ageing.

The grape aroma is synthesized in grape berries by a variety of enzymes, including terpene synthases, *O*-methyl transferases, carotenoid cleavage dioxygenases, cytochromes P450 and probably other not yet characterized enzymes. Genetic variation in aroma biosynthesis genes cause differences in aroma between grapevine varieties: an allelic variant of 1-deoxy-D-xylulose-5-phosphate synthase, a terpenoid biosynthetic gene, causes accumulation of terpenoids in Muscat and Gewurztraminer grapes. An abundance of terpenoids gives these wines a distinct floral aroma (6–8). In another example, differential expression of an *O*-methyl transferase gene results in higher production of methoxypyrazines, compounds evoking the typical capsicum aroma in Sauvignon wines (9). The genetic factors underlying the aroma typicity of all other grapevine varieties—only in Europe over 2000 (10)—remain unexplored.

Grape berries store most of the volatiles they produce as glycosides (11). Possible functions of glycosylation are sequestration, detoxification and decrease of volatility and reactivity (12). Since glycosides are not volatile, they do not directly contribute to wine aroma. They do, however, affect the aroma indirectly: they form a precursor pool from which volatile aglycons

can be released during yeast and malolactic fermentation, during vinification by addition of exogenous glycosidases, during wine ageing by the low pH (13) and, as demonstrated recently, by enzymatic hydrolysis in the mouth, catalyzed by the enzymes in the saliva (14). Aroma glycosides in grapes have either one or two sugar moieties attached to the aglycone (mono and diglycosides, respectively). The first sugar moiety, directly attached to the aglycone, is in all cases glucose. The majority of glycosides (at least in the case of terpenoids) are present in the form of diglycosides, in which a second sugar—arabinose, apiose or rhamnose—is attached to the glucose (13). The diglycosides cannot be hydrolyzed by a β-glucosidase, and require other glycosidase enzymes to release the volatile aglycon (15).

The winemaking process profoundly influences the wine aroma. Hydrolysis of glycosides is not the only effect fermentation has on the aroma composition. Yeast (*Saccharomyces cerevisiae*) produces many components of wine aroma, particularly alcohols and esters, from primary metabolites. Different yeast strains can produce remarkably different aroma profiles (16). After alcoholic fermentation wines are sometimes subjected to malolactic fermentation by *Oenococcus oeni*, the principal role of which is to reduce tartness or acidity of wine by converting malic acid to lactic acid. These bacteria can also alter the composition of aroma by, for example, promoting deglycosylation (17).

The evolution of wine aroma continues after the fermentation process. In addition to above mentioned acid hydrolysis of aroma glycosides during wine ageing, low pH can cause other important chemical changes. Williams *et al.* postulated that some hydroxylated linalool derivatives undergo cyclisation or other rearrangements at low pH (18). In addition, during storage in oak barrels compounds from the wood penetrate the wine and influence its taste and aroma.

Standardized sample preparation (19, 20) and analysis procedure is routinely used for analysis of grape and wine volatiles and their precursors: grape juices or wines are extracted to a non-ionic solid phase, free volatiles are eluted by a non-polar solvent (pentane or a mixture of pentane and dichloromethane), and glycosidically bound volatiles are eluted by a more polar solvent (ethyl acetate or methanol). Free fraction can then be directly analyzed by gas chromatography, whereas the bound fraction is enzymatically hydrolyzed prior to the analysis. Gas chromatography allows for simultaneous analysis of hundreds of volatiles with good resolution of structurally similar molecules. Furthermore, coupling to a mass spectrometer enables reliable identification of compounds by searching mass spectra databases even when analytical standards are not available.

Many research laboratories used this method to investigate how aroma is influenced by grape variety, grape ripening, environment and different winegrowing and winemaking

techniques. While these studies have without doubt answered some of these important questions, a global picture of wine aroma using an ensemble of these valuable data has not been painted yet. We collected, curated and analyzed these data to answer the following questions: what are the components of grape and wine aroma, at which concentration are they present and to what extent are they glycosylated. We compared these parameters between grapes and wines. We then compared these data to the existing knowledge on the origin and evolution of these aroma components with a particular focus on their biosynthesis. This meta-analysis will hopefully add to our understanding of wine aroma composition and development.

Composition of wine aroma

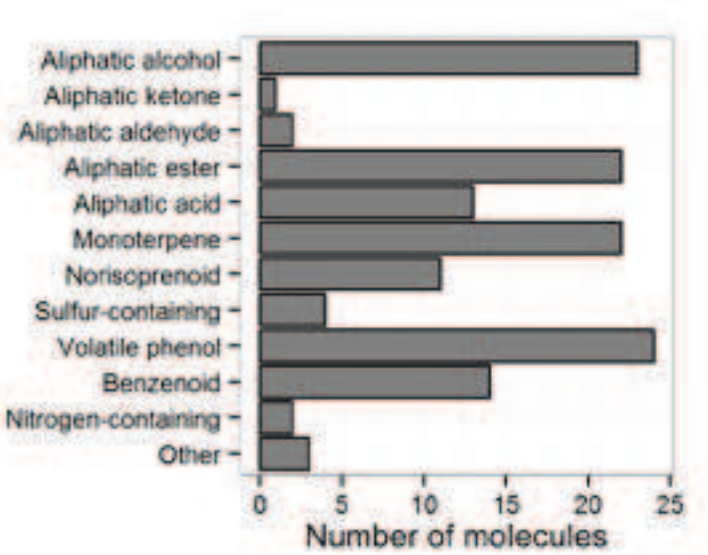

Figure 8. Classes of grape and wine volatile compounds evaluated in this study.

Analysis of 19 publications describing grape or wine aroma revealed 385 different volatile organic compounds. More than half of them were only identified in one or two studies, and were eliminated from further analysis to account for the possibility of incorrect peak annotation. Some of the eliminated compounds might be important varietal compounds, but this study focuses on the similarities, not the differences between the varieties. The remaining 141 validated volatile compounds (Table S1) were assigned to one of the 12 classes based on their chemical structure and biosynthetic origin. Authors of most reviews make a clear distinction between grape and fermentation derived wine aroma (4, 5, 21, 22) and therefore use a different classification of wine aroma components. Classes defined in this study are based on classifications from other authors, while trying to contain all the molecules and to minimize the overlap between classes. The most represented classes (with respect to the number of validated compounds) were aliphatic alcohols, aliphatic esters, monoterpenes and volatile phenols (Figure 1).

Not all volatiles contribute to the aroma equally. Their impact depends on their concentration, as well as aroma intensity. The latter is usually expressed as the odor detection threshold, which is the lowest concentration of a particular compound that can be detected by the human smell. In a recent meta-analysis, Dunkel *et al.* (2) compiled a list of food volatiles occurring at concentrations exceeding their odor detection threshold. Somehow surprisingly, a total of only 220 key food odorants are responsible for the aroma of most of the foods and beverages we consume, and among those fewer than 40 contribute to aroma of an individual food item. Another meta-analysis identified 57 supra-threshold volatiles in wines (all but 2 of these volatiles are also included among the 220 key food odorants). Only 60% (N=35) of those key wine odorants overlap with our set of common wine volatiles, compiled from profiling experiments (Figure 2). The other 40% (N= 22) of key wine odorants are overlooked in volatile profiling experiments because of their low concentrations. In particular, sulfur-containing volatiles and fatty acid lactones appear systematically overlooked (Table S2). Monoterpenes are one of the most represented classes of molecules in our study, but two odor-active monoterpenes, *cis*-rose oxide and wine lactone, are still systematically overlooked in profiling experiments. Conversely, the high-impact aliphatic alcohols, acids and esters, as well as phenols and benzenoids, are apparently present in sufficient concentrations to be easily detectable in profiling experiments. A group of 15 common wine volatiles features on the list of key food odorants, but not the key wine odorants. We labeled those compounds "potential" key odorants (Table S2). The progress in analytical chemistry will likely permit discovery of more key wine odorants in future. Rotundone, the odorant conferring the peppery character to Shiraz wines, was only discovered recently (23) and was not included among the key food odorants nor the key wine odorants.

Although odor detection threshold is commonly used to describe an influence of a particular compound on aroma, this value needs to be used with caution for describing complex mixtures, including wines. Both synergistic and antagonistic effects can occur in odor perception: sometimes a mixture of compounds can be perceived even if all components are at sub-threshold concentrations; conversely, some compounds can mask the perception of other compounds, so they remain undetected at supra-threshold concentrations (24, 25). In addition, odor detection threshold is averaged across the population. Because of the variability in odor receptor genes in human population, each individual has a highly personalized odor perception (26). Finally, threshold values are not known for all the wine volatiles. For these reasons we decided to include all the volatiles, not just the key wine odorants, in our meta-analysis.

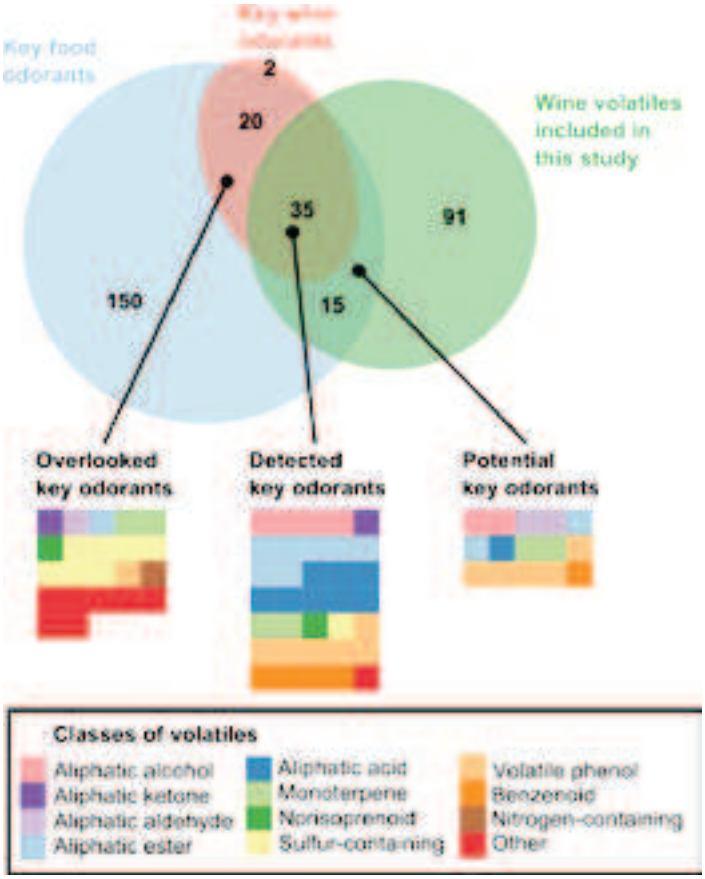

Figure 9. Euler diagram of wine volatiles included in this study and key food and wine odorants, identified in two other meta-analyses. Two other meta-analyses identified volatile compounds with the strongest impact on aroma of food (2) and wine (27). While the overlap between the three studies is sizeable, many key odorants are overlooked in volatile profiling experiments. Each tile represents one volatile molecule and its color corresponds to the molecular class.

Origin and evolution of wine aroma

During the winemaking process the aroma undergoes major changes, in particular deglycosylation of aroma compounds synthesized in grapes and biosynthesis of new compounds. To analyze these changes quantitavely we selected 43 compounds with sufficient number of available data points (>30) across the 82 volatile profiling datasets. For each compound we computed the total concentration (the sum of free and bound concentration) and percentage that is glycosylated (bound/total concentration), and tested for differences between these values in grapes and wines (Figure 3). These differences in concentration and degree of glycosylation are largely characteristic of each class of molecule.

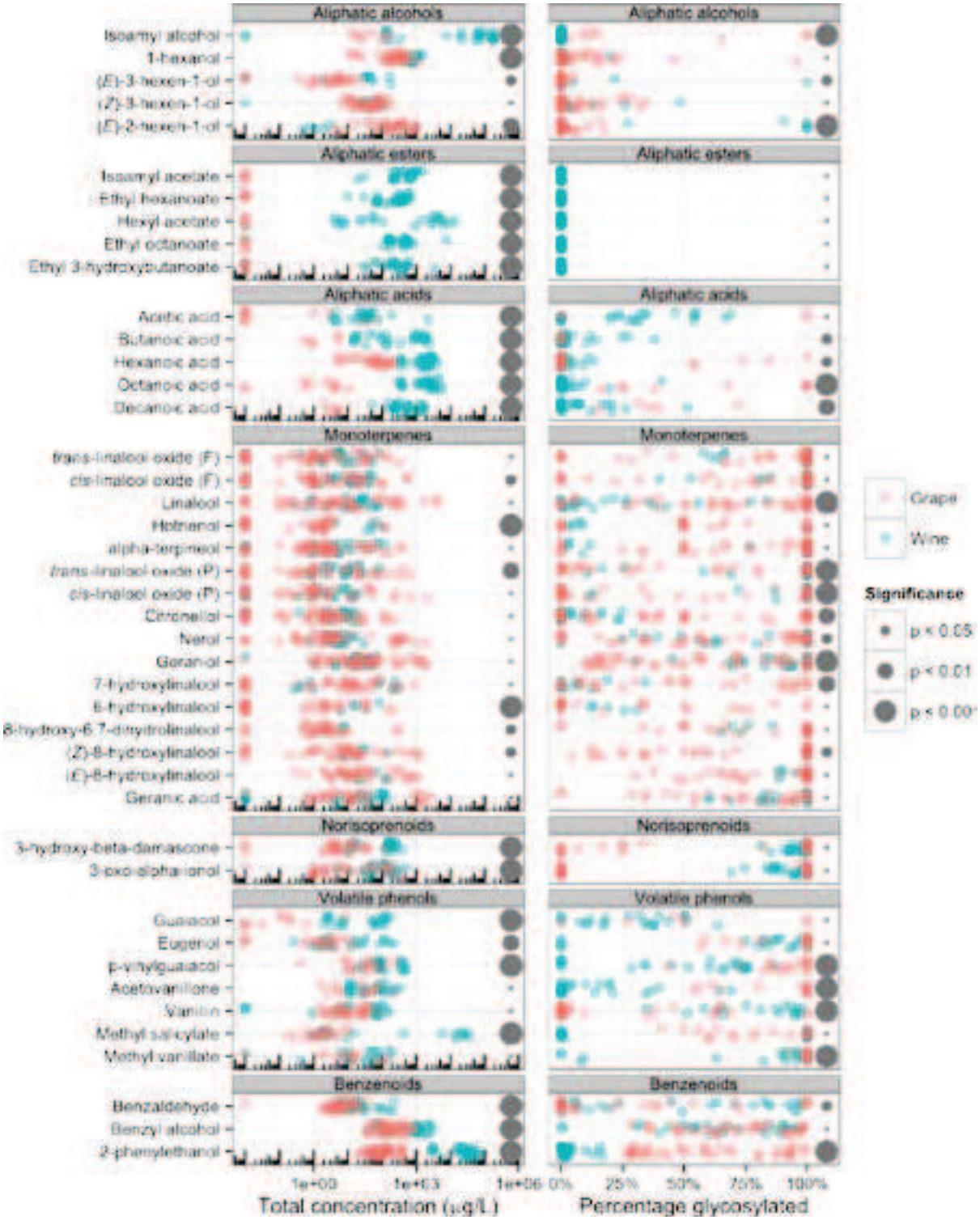

Figure 10. Comparison of the total concentration (left) and degree of glycosylation (right) of selected grape and wine volatile organic compounds. Total concentration is the sum of free and bound concentration and percentage is glycosylated is bound divided by total concentration. Each point represents one grape or wine sample from one of the 19 publications included in the study. Non-detected compounds (concentration zero) were assigned a concentration 0.01 µg L^{-1} to allow their representation on a logarithmic scale. Grey dots on the right indicate significant difference between grapes and wines and their size is proportional to the p-value of the statistical test. Student's t-test was used on log-transformed concentrations and Wilcoxon rank sum test was used on glycosylated fraction.

Aliphatic alcohols

Aliphatic alcohols are a diverse group of compounds that can originate both from grapes and yeast fermentation. C6-alcohols, a common group of plant volatiles with six carbon atoms, smell characteristically "green", reminiscent of leaves and fresh cut grass. They are derived from corresponding C6-aldehydes, also important aroma compounds, by alcohol dehydrogenase enzymes. C6-aldehydes are products of hydroxyperoxide lyase (CYP74) enzymes (28), which were recently characterized in grapevine (29). C6 alcohols can be consumed by the yeast during fermentation (30), which may explain why concentration of (*E*)-2-hexen-1-ol is lower in wines compared to grapes.

Short chain alcohols, also known as fusel alcohols, can add a negative impact to the wine aroma at high concentrations (31). They are formed by the yeast during fermentation from amino acid catabolism (30). A typical example is isoamyl alcohol, the concentration of which is much higher in wines compared to grapes and is also the most abundant compound in this study (Figure 3a).

Aliphatic alcohols also contribute to the aroma as precursors of esters, which are discussed below.

Aliphatic acids

The yeast produce aliphatic acids during fermentation (5) from primary metabolites: fatty acids (C6 or longer), amino acids (short and branched) or sugar (acetic acid). Aroma of volatile fatty acids is generally unpleasant, ranging from sweaty and cheesy to goaty and rancid. Although all aliphatic acids included in this study were detected in both grapes and wines, their concentration was significantly higher in wines (Figure 3a), confirming they are predominantly a fermentation product.

Longer aliphatic acids (C8 and C10) in grapes are glycosylated at least to some extent, which is unexpected because they lack a hydroxyl group to which the sugar moiety is usually attached. Instead, they are probably stored as glucose esters, a less common type of glycoconjugate, where sugar and aglycone are connected via an ester bond. Fatty acid glucose esters have been described in other plants (32), but their role as wine aroma precursors has not yet been extensively studied.

Aliphatic acids can be transformed to more pleasant smelling compounds, such as esters (described below) and lactones. Five fatty acid derived lactones (gamma-nonalactone, gamma-decalactone, gamma-dodecalactone, gamma-(*Z*)-6-dodecenolactone and (*Z*)-oak lactone, also known as whisky lactone) are among the key wine odorants, but are

systematically overlooked in the volatile profiling experiments, presumably because of their low concentration in wines (27).

Aliphatic esters

Esters are a group of volatiles that contribute to fruity notes of wine and other fermented beverages. They are produced during fermentation from alcohol and acyl-CoA by yeast alcohol acyltransferase enzymes. The two major groups of esters in wine are ethyl esters and acetate esters. Their concentrations in grapes are negligible.

The concentration of ethyl esters of medium-chain fatty acids depends on the concentration of the fatty acid precursor (33). Our data confirms this observation: concentrations of ethyl hexanoate and ethyl octanoate are strongly positively correlated to concentrations of their precursors, hexanoic and octanoic acids, respectively (Figure 4). Similar relationship was recently found for C9 and, to a lesser extent, C12 ethyl esters (34). While hexanoic acid is predominantly a fermentation product, non-negligible amounts are present in grapes as well. The biosynthesis of hexanoic acid in grapes could thus influence the concentration of a typical fermentation product, ethyl hexanoate, in wine.

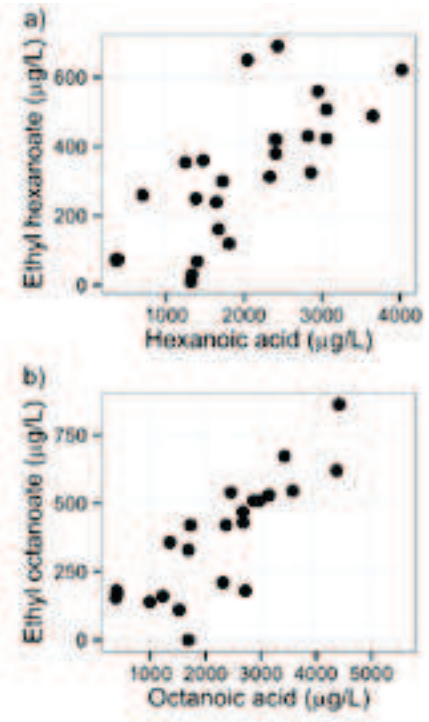

Figure 11. Relationships between the concentrations of ethyl esters and their acid precursors in wines. a) Concentration of ethyl hexanoate is correlated to concentration of hexanoic acid (R=0.780, p-value<0.001) and b) concentration of ethyl butanoate is correlated to concentration of octanoic acid (R=0.830, p-value<0.001).

Conversely, precursor concentration does not determine the concentration of acetate esters. The limiting factor in the production of acetate esters by *S. cerevisiae* is expression of the

alcohol acetyltransferase gene in yeast (35). Indeed, isoamyl acetate and hexyl acetate do not correlate to the concentrations of their precursors (Figure 5).

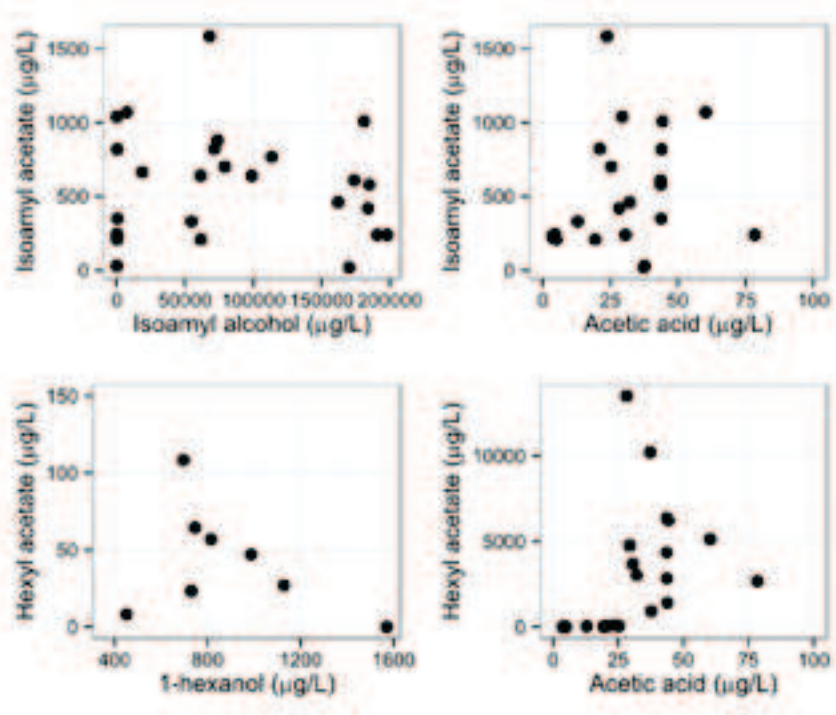

Figure 12. Relationships between the concentrations of acetate esters and their precursors in wines. Concentration of the isoamyl acetate and hexyl acetate is not correlated to the concentration of neither the alcohol nor the acid precursor (α=0.01).

Monoterpenes

Monoterpenes are a large class of plant specialized metabolites. They are built from two isoprenoid units, giving them a backbone of 10 carbon atoms. These compounds give many fruits, flowers, herbs and spice their characteristic aroma. Most of the wine monoterpenes contribute towards floral and citrusy notes. For example, monoterpenols and their derivatives give the characteristic aroma to Muscat (36) and Gewurztraminer (37) wines. High monoterpenol concentration in these varieties results from a mutation in an early terpenoid biosynthesis gene deoxy-D-xylulose synthase (6–8). Our data show that monoterpenols span a large concentration range in grapes and wines (Figure 3a) which points to their role as varietal aroma compounds in some grapevine varieties. Unlike other classes of volatiles described here their concentrations in grapes and wines are similar, as expected for grape-derived compounds. Conversely, the glycosylated percentage of many terpene compounds is lower in wines compared to grapes, suggesting fermentation nonetheless affects the monoterpene content in wine by releasing volatile monoterpenes from their glycosylated precursors.

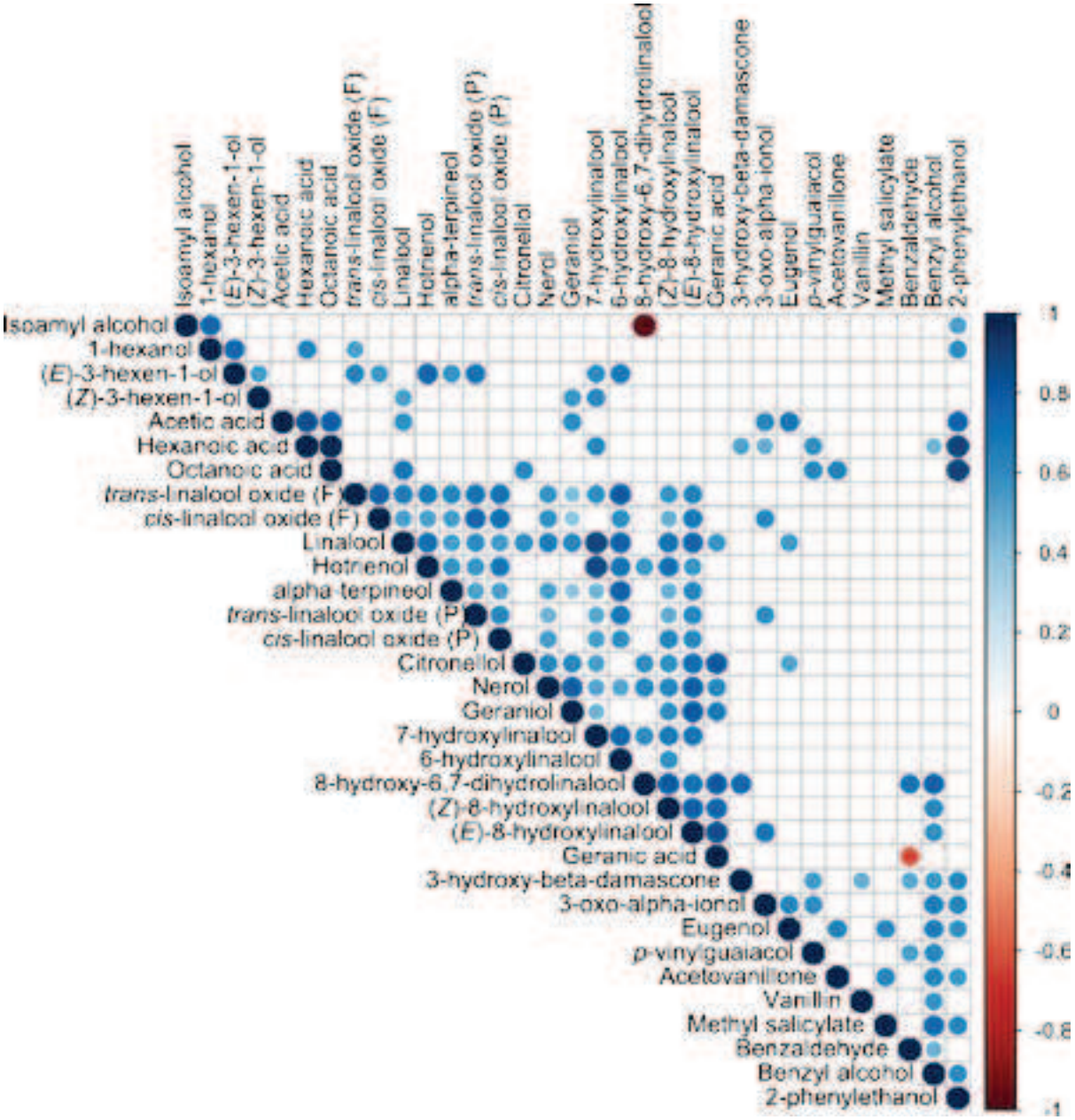

Figure 13. Correlation matrix of selected volatile compounds in wines and grapes. Total concentrations were log-transformed prior to calculation of correlation. Only the compounds with more than 40 data points were included in the calculation. Color and dot size are proportional to the correlation coefficient. Only coefficients with p-value<0.001 are displayed.

Monoterpenes are products of terpene synthase enzymes. The terpene synthase gene family has expanded in grapevine, which underlines the importance of terpenoids in this species (38). More than half of terpene synthase genes were functionally characterized (38) and were found to produce a large number of different mono- and sesquiterpene backbones *in vitro*. Interestingly, this variability is not reflected in the volatile profiles of grapes and wines in our selected studies. Sesquiterpenes do not appear in our dataset of 141 validated volatiles. However, a labeling study revealed a production of 14 sesquiterpenes in grape skins of two different varieties, suggesting the sesquiterpene metabolism in grapes is nonetheless active

(39). The concentrations of sesquiterpenes in grapes and wines are probably too low to be detected in non-targeted profiling experiments, but they do contribute to wine aroma of at least some varieties: rotundone, an oxygenated sesquiterpene, is responsible for the peppery aroma of Shiraz wines (23). Monoterpenes, on the other hand, are one of the largest molecular classes in our study, with 22 different molecules identified in grapes or wines (Figure 1). Interestingly, this large chemical variability among wine monoterpenes does not result from the variability of different backbones synthesized by terpene synthases. Half of the validated monoterpenes in this study are derivatives of the same monoterpene: linalool. Seemingly the variability of monoterpenes in grapes arises from the enzymes that oxygenate linalool at different positions.

Elaborate oxidative linalool metabolism in grapevine was previously demonstrated by feeding experiments (40), (Figure 7). All of the described linalool derivatives, with the exception of nerol oxide, 6,7-epoxylinalool and 6,7-dihydroxylinalool are also present on our list of validated volatiles. Common metabolic origin of monoterpenes, in particular linalool derivatives, is also apparent from the correlation matrix of grape and wine volatiles (Figure 6). Linalool is positively correlated to all the monoterpenes included in the study, which supports its role as a central monoterpene metabolite in grapes. Enzymes catalyzing linalool oxidation in grapes have not yet been identified, but in other plants enzymes from cytochrome P450 superfamily were shown to oxidize monoterpenes (41, 42), including linalool (43, 44). Hydroxylated linalool derivatives were discovered in grapes in early 1980s (11, 45). These compounds have a very weak odor so they are unlikely to contribute to the wine aroma directly, but they may have an indirect contribution. They were found to spontaneously transform to compounds with strong aroma in conditions mimicking wine maturation: in acidic conditions they undergo spontaneous elimination of water and rearrangement to either linear (hotrienol) or cyclic compounds (linalool oxides) (18). Although low concentrations of hotrienol were detected in grapes, its concentration in wines is significantly higher (Figure 3a), which supports the hypothesis of acid-catalyzed formation from 7-hydroxylinalool during winemaking and wine maturation. In addition, concentrations of 7-hydroxylinalool and hotrienol are strongly correlated (Figure 6). We do not find, however, the same amount of evidence for non-enzymatic formation of linalool oxides from 6,7-dihydroxylinalool or 6,7-epoxylinalool. Concentrations of linalool oxides in wines and grapes are comparable. In addition they are glycosylated to a very high degree in grapes, which suggests they are formed in metabolically active grape berries, as demonstrated previously in the feeding study (40). Formation through acid-catalyzed cyclisation during wine maturation is therefore probably of lesser significance.

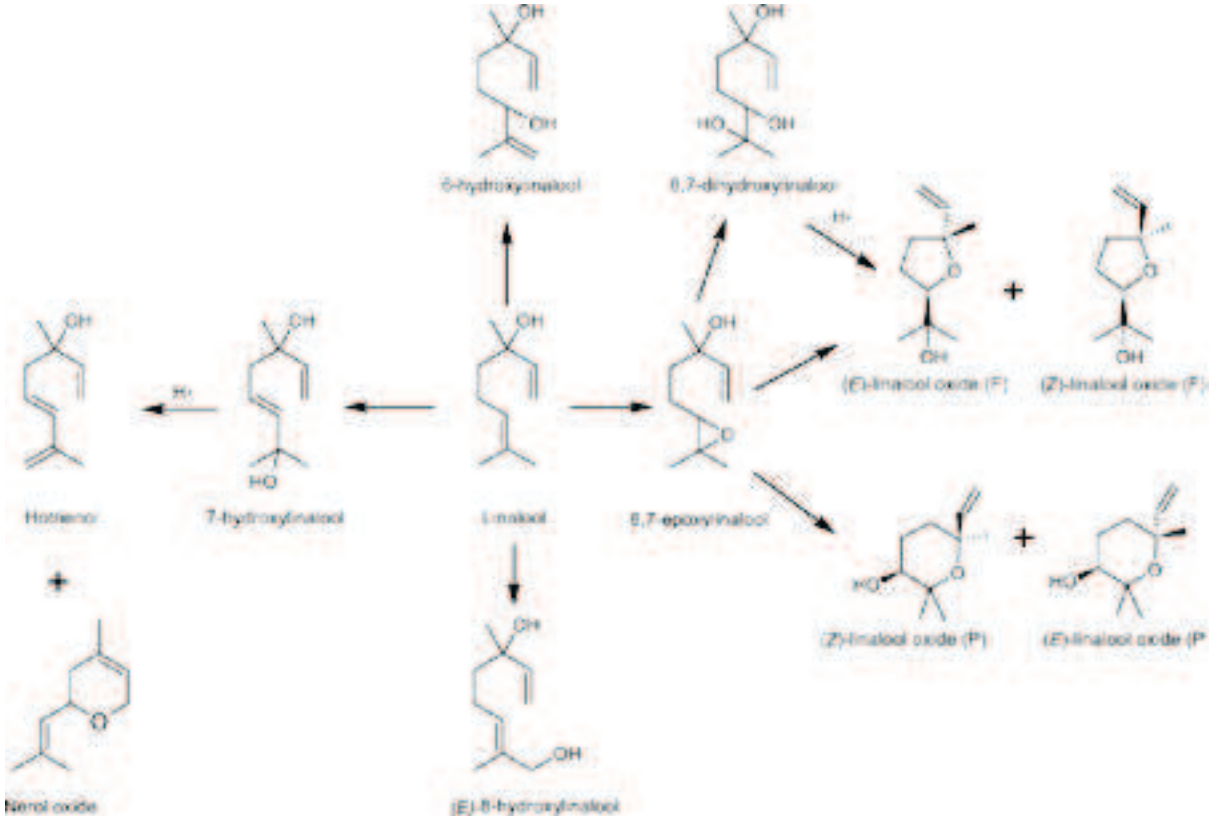

Figure 14. Oxidative metabolism in *Vitis vinifera* cv. Morio Muscat grape berries. Feeding with deuterium-labelled linalool revealed complex linalool metabolism in grape berries. Adapted from (40).

From a single profiling experiment it is difficult to estimate what proportion of the total linalool pool is transformed to oxygenated derivatives, mainly because not all derivatives are quantified in all the experiments. Strong correlation between concentrations of monoterpenes in the studies included in this meta-analysis (Figure 6) allowed us to describe and predict relationships between concentrations of linalool and oxygenated linalool derivatives with a set of linear models (Table S3). In the investigated concentration range (0.001–10 μM) most of linalool is oxygenated (Figure 8). At low concentrations virtually all linalool (97%) is oxygenated and the main linalool derivative is (*E*)-8-hydroxylinalool. At high concentrations oxygenated derivatives represent 52% of the complete linalool pool, and the most abundant derivative is 7-hydroxylinalool. 6-hydroxylinalool and linalool oxides represent a minor part of linalool derivatives. The total fraction of oxygenated linalool derivatives in the linalool pool is underestimated since not all linalool derivatives were included in the calculation.

Monoterpenes in grapes are predominantly glycosylated, although variation between the samples is high (Figure 3). Since only free compounds can impact the wine aroma, this variability could be important for varietal characteristics. Several monoterpenes are glycosylated to a lesser extent in wines as compared to grapes. This can be attributed to

hydrolysis of glycosides by the yeast during fermentation, as well as acid hydrolysis during wine maturation.

In spite of the large number of monoterpenes included in this study, the list presented here is not exhaustive. The lowest limit of quantification in the analyzed set of profiling experiments is around 10^{-7} µg L^{-1} (Figure 3) and concentrations of some monoterpenes, including rose oxide or wine lactone, in wines are below this limit. Nonetheless, because of their low sensory threshold, they still contribute to wine aroma as one of the key wine odorants even at these low concentrations (27). Furthermore, another linalool derivative, (*E*)-8-carboxylinalool, has been detected in grapes and wines, but is not detectable by gas chromatography, hence the absence from the data collected here. This compound also indirectly influences wine aroma as a precursor to wine lactone.

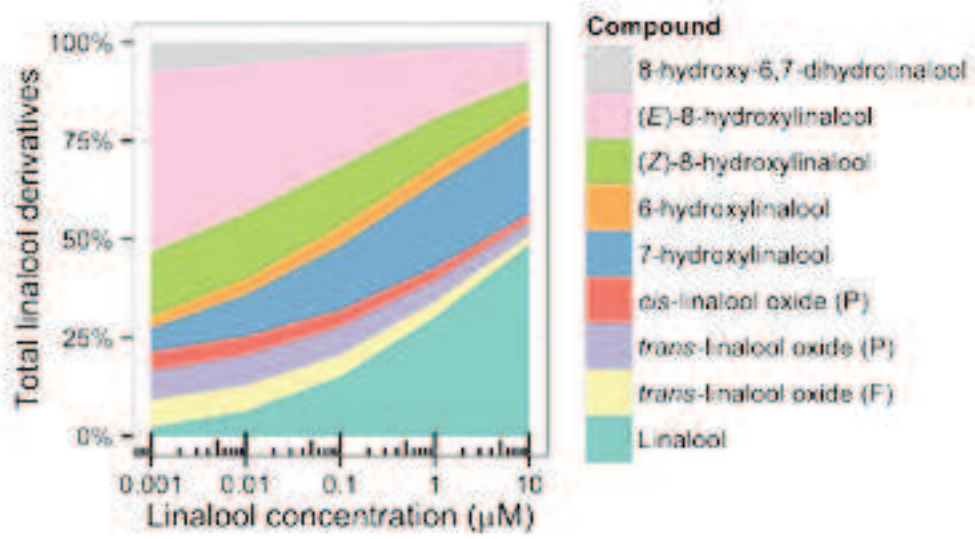

Figure 15. Estimated flux of linalool to different oxygenated derivatives. Relationship between log concentrations of each linalool derivative and linalool was described with a linear regression. These models were then used to estimate the concentration of each derivative in the function of linalool concentration. (cis)-linalool oxide (F) and hotrienol were excluded from the figure because of low R^2 value (**Table S3**).

Norisoprenoids

Norisoprenoids are a group of carotenoid-derived metabolites. Similarly to monoterpenes, their aroma is mostly described as floral or fruity, although some, for example 1,1,6-trimethyl-1,2-dihydronaphthalene (TDN) which is described as petrol or kerosene-like, can have a negative impact on aroma (46).

Norisoprenoids are synthesized by carotenoid cleavage dioxygenases, enzymes that cleave carotenoid substrates at different in-chain positions, yielding products of different sizes. Most of the norisoprenoids in grapes are derived from β-carotene or neoxanthin (Figure 9) (47) and have 13 carbon atoms (C13-norisoprenoids), with the exception of 4-oxoisosporone

(C9). So far only one grapevine carotenoid cleavage dioxygenase has been characterized (48). It was shown to cleave zeaxanthin, a minor grape carotenoid, whereas neoxanthin and β-carotene were not tested as substrates. A hypothetical reductase in the β-damascenone pathway (Figure 9a) has to our knowledge not yet been identified.

The chemical diversity of norisoprenoids in grapes appears to originate from different non-enzymatic reactions, including photooxygenation, thermal degradation or acid hydrolysis (47). Although norisoprenoids are considered grape-derived metabolites, our data show higher concentrations of two norisoprenoids (3-hydroxy-β-damascone and 3-oxo-α-ionol) in wines as compared to the grapes (Figure 3). However, it is unlikely these compounds would form during fermentation or wine maturation, since they are glycosylated to a large extent. It is unknown if non-enzymatic transformations occur also on glycosylated derivatives of these compounds.

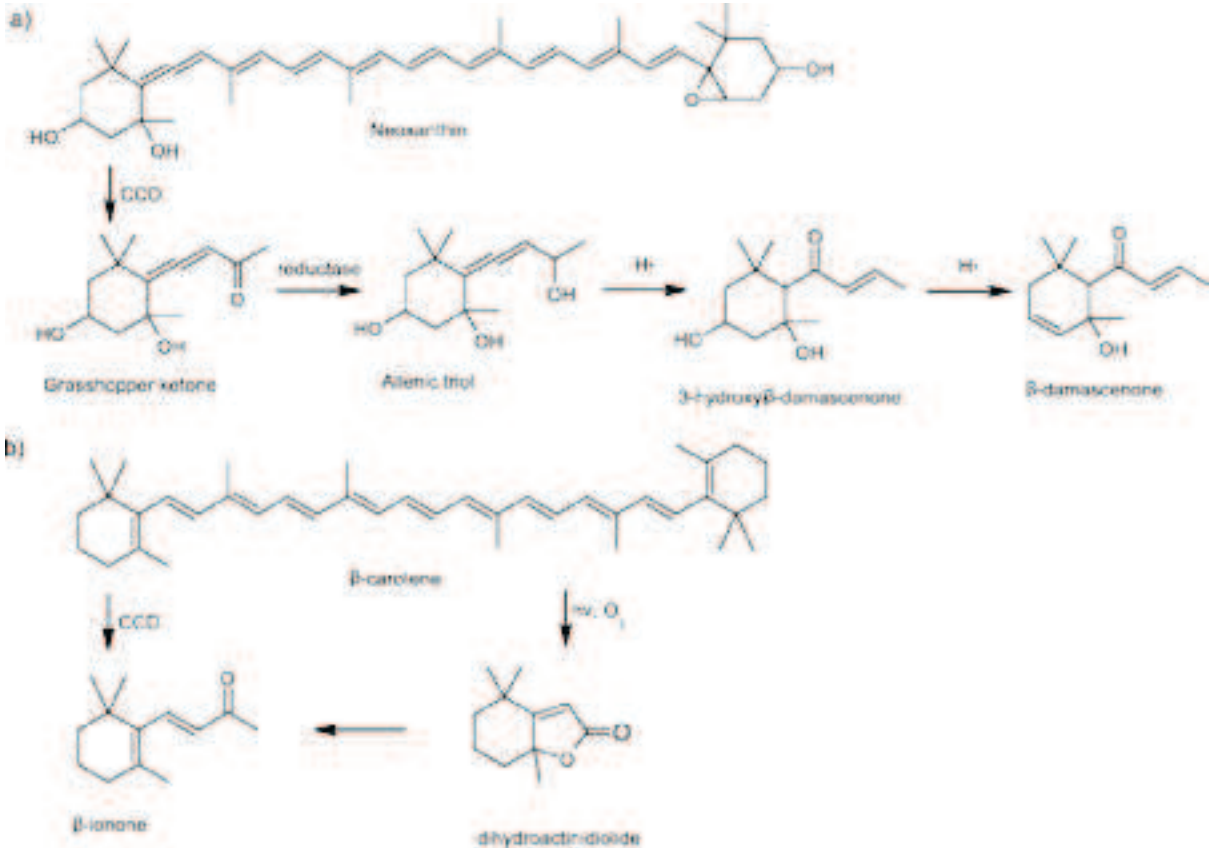

Figure 16. Formation of norisoprenoids from carotenoids. a) Formation of damascenone-type derivatives from neoxanthin. b) Formation of β-ionone from β-carotene via enzymatic or non-enzymatic pathway. CCD – carotenoid cleavage dioxygenase.

Volatile phenols

Volatile phenols are a diverse group of wine volatiles with respect to both their origin and impact on wine aroma. Many of them are common plant volatiles, derived from ferulic acid or related metabolites, and contribute to pleasant spicy aroma notes. Although enzymes

catalyzing their biosynthesis have not yet been characterized in grapes, they have been studied in other plants. Examples include clove aroma eugenol, which is synthesized by an enzyme reducing coniferyl acetate in basil or petunia flowers (49), or vanillin, synthesized from ferulic acid by a hydratase/lyase enzyme in vanilla pods (50). Most of volatile phenols are stored in grapes as glycosides, and can be hydrolyzed during winemaking. It is noteworthy that vanillin in wine can also originate from ageing in oak barrels (51).

Not all volatile phenols are associated with pleasant aroma: some of them, for example guaiacol, are described as smokey, ashy or medicinal and are considered as off flavors in wine. These compounds can originate from fermentation, contamination with spoilage yeast *Brettanomyces* (52) or smoke exposure of grapes, for example from nearby forest fires (53). Concentration of guaiacol is much higher in wines compared to grapes (Figure 3), which confirms fermentation origin of this compound. Interestingly, some authors suggested that glycosylated precursors from grapes represent only a minor source of *p*-vinyl guaiacol in wine (54), but our data suggest this contribution can be substantial.

Benzenoids

Benzenoid compounds in wines are cinnamic acid derivatives with varying side chain lengths and oxidation states. Phenylethanol and phenylacetaldehyde have a weak floral aroma. The high odor detection threshold of phenylethanol is compensated by its high concentration. Although these volatiles can be produced by both plants (55) and yeast (56), their concentration in wines is generally much higher than in grapes, which suggests that contribution from yeast is larger (Figure 3). An exception is phenylacetaldehyde, which is reduced by the yeast to yield 2-phenylethanol.

Conclusions

Analysis of volatile profiles of grapes and wines and their precursors from 20 publications resulted in a list of 141 common grape and wine volatiles (Table S1). A comparison to the previously published list of key wine odorants (27) revealed an overlap of 35 compounds, whereas 22 key wine odorants are overlooked in the volatile profiling experiments we reviewed. In particular, sulfur-containing volatiles and fatty acid-derived lactones are consistently overlooked. This result points to the need of targeted analysis of low-abundance compounds.

To explore the complex origin and evolution of wine aroma, 43 volatiles with sufficient number of data points were analyzed in greater detail. Some compounds occur in similar concentration range in both grapes and wines. These compounds are most likely derived from grapes and their concentration remains constant during fermentation and wine maturation. Concentrations of other compounds differ dramatically between grapes and

wines, because they are either produced or consumed by the yeast during fermentation. Some glycosylated compounds are also hydrolyzed during fermentation, which is reflected as a lower glycosylated fraction in wines compared to grapes.

The boundaries between the roles of grape metabolism, fermentation and ageing in the formation of aroma are blurred because many compounds are affected by all three of these processes. Monoterpenoids, typically considered grape-derived volatiles, accumulate in grapes as glycosides, which can be hydrolysed during the fermentation and can undergo chemical rearrangements during wine maturation. Hotrienol is a monoterpene thought to be formed from 7-hydroxylinalool by a non-enzymatic acid-catalyzed reaction. Although it is present in grapes, its concentration is significantly higher in wines, which is in agreement with its wine maturation origin. Yet its concentration also depends on the concentration of its precursor, 7-hydroxylinalool, which is synthesized in the grapes. What is more, 7-hydroxylinalool is partially de-glycosylated during fermentation, which may also affect hotrienol concentration. To understand and be able to predict wine aroma, all these different factors will need to be taken into consideration.

Monoterpenes are one of the largest groups of volatiles in wines, but their chemical diversity in wines is not due to the different backbones synthesized by a large family of grapevine terpene synthases. Half of the monoterpenes in this study are oxidized derivatives of the same monoterpene—linalool—with main derivatives being 7- and (*E*)-8-hydroxylinalool. We showed that most of linalool produced in grapes is oxygenated. Linalool oxygenases thus not only influence wine aroma by formation of new aroma compounds, but also by depletion of their substrate, one of the key wine odorants.

Although considerable progress in biosynthesis of wine aroma has been made since the sequencing of *Vitis vinifera* genome in 2007 (57), many enzymatic steps leading to aroma compounds remain to be identified. This knowledge is important for the breeding of new grape varieties, and could as well be used to produce ingredients for flavor and fragrance biotechnology.

Material and methods

Data collection

Table 1. Publications selected for the analysis.

Publication	Sample type	Number of samples	Genotypes/ Varieties
(58)	Wine	1	Karaoglan
(59)	Wine	5	Verdejo
(60)	Wine	2	Narince
(61)	Wine	2	Negroamaro
(62)	Wine	5	Lourerio, Alvarinho
(63)	Wine	1	Sangiovese
(64)	Wine	5	Bobal
(65)	Wine	5	Moravia Agria
(66)	Wine	3	Moravia Dulce, Rojal, Tortosi
(67)	Grape	1	Tannat
(68)	Grape	6	F3P30, IASMA ECO3, F3P63, Riesling, Gewürztraminer, Moscato Rosa*
(69)	Grape	2	Muscat of Frontignan
(70)	Grape	5	Mencía, Espadeiro, Caíño Redondo, Pedral, Sousón
(71)	Grape	2	Muscat Hamburg
(72)	Grape	14	Alvarinho, Arinto, Avesso, Azal, Batoca, Lourerio, Trajadura
(73)	Grape	1	Fiano
(74)	Grape	5	Sauvignon Blanc, Chardonnay, Gewürztraminer, Grüner Veltliner, Sangiovese
(75)	Grape	2	Superior Seedless
(76)	Grape	15	Albarinho, Listan, Treixadura, Greanche, Parellada, Tempranillo, Viura, Xarel·lo
(77)	Grape	(3)	Vitis rotundifolia (Muscadine) **

*Non-vinifera genotypes (Nero, Isabella, *Vitis arizonica* and *Vitis cinerea*) were excuded from the quantitative analysis. **Excluded from the quantitative analysis.

We collected published datasets on free and glycosylated grape and wine volatiles. We used search term »grape aroma profiling free glycosylated« on Google Scholar and reviewed citations of included publications. We reviewed 104 publications, 45 of them were excluded because they only contained information on free volatiles, and 8 because they only contained information on glycosylated volatiles. To ensure the compounds were annotated correctly, only publications that contained information of Kovats retention index were included. Further 31 publications were thus excluded, which resulted in 19 selected publications. Publications or samples that described *Vitis* species other than *Vitis vinifera* subsp. *vinifera* were excluded from all quantitative analysis, but used for validation of retention indices.

Data curation

Molecule names and their Kovats retention indices were aggregated in one file and manually checked for duplicates (synonyms), which resulted in 385 unique molecules. They were assigned to one of the 16 classes of volatiles (aliphatic alcohol, ketone, aldehyde, ester or acid, monoterpene, norisoprenoid, sulfur containing, phenol, aromatic alcohol, ketone, aldehyde, ester or acid, nitrogen containing or other). 244 of unique molecules appeared in less than three publications and were excluded from further analysis because of insufficient information. Kovats retention indices of the remaining 141 compounds were compared and in the cases where standard deviation of the retention indices exceeded 50, outliers were excluded (annotation of a compound was considered incorrect in the particular publication and was excluded from the dataset).

We paid particular attention to geometric isomers of some compounds, which elute close to each other and have similar or identical mass spectra, so they are often misannotated. Such pairs of compounds are: pyranic linalool oxides, furanic linalool oxides and 8-hydroxylinalools. In case of pyranic linalool oxides, the two isomers were incorrectly annotated in the first paper reporting their structure (78). This error was identified and corrected later (79), but some authors continued to use the incorrect annotation. All the data were curated so that the first eluted furanoid linalool oxide was annotated as trans isomer, and the second one as cis. Similarly, first eluted pyranoid linalool oxide that elutes first was annotated as trans and the second one as cis (80). Retention indices of (*Z*) and (*E*)-8-hydroxylinalool were first reported in (81) and the author claimed the NMR spectra were in agreement with previously published data. However in a subsequent publication (82), which is also cited in the widely used database of volatiles Pherobase (http://www.pherobase.com/), the retention indices of the two compounds are inverted even if the author received standards from the author of (81). Recent publication (44) confirms the correct elution order is (*Z*)-, followed by (*E*)-8-hydroxylinalool, as reported in the original publication (81).

Statistical analysis

Statistical analysis was performed using the R software version 3.0.2 (83). For quantitative analysis 43 molecules with more than 30 data points were selected from the pool of 141 validated compounds. Concentrations of these compounds in free and glycosidically bound forms were then extracted from the 20 publications. If a compound was only detected in the free fraction, it was assigned a concentration of 0 in the glycosylated fraction, and vice versa. Labels »not detectable«, »not quantifiable« and »trace« were all converted to zero concentration. Concentrations of free and bound form of each molecule were then added to

calculate the »total concentration«. »Fraction glycosylated« was calculated by dividing the concentration of glycosylated compound with the total concentration.

Concentration units were not the same in all studies. All concentrations in wines were given as μg L^{-1}, some concentrations in grapes were given as μg L^{-1} and some as μg kg^{-1}. In addition, grape juice and wine have different, so the concentrations are not directly comparable, but since we compared them on a logarithmic scale we considered these differences negligible and did not adjust for them.

Log-transformed concentrations were compared using a two-sided t-test. Only molecules with 5 or more data points in each category (grape and wine) were tested stastistically. Glycosylated fractions were compared using Mann-Whitney-Wilcoxon u-test. Correlations were calculated on total concentrations (free + glycosylated) for compounds with more than 40 complete observations. Linear models for linalool derivatives were calculated on log total molar concentrations (free + glycosylated) with non-zero values. Linalool derivatives with low R2 value (<0.45) were excluded from the graph (hotrienol, (*cis*)-linalool oxide (pyranoid)). Data, residuals normality and leverage were visually evaluated for each model.

Acknowledgements

We thank Julien Denos for helpful discussions on data analysis.

Authors' contributions

Investigation, T.I.; Formal Analysis, T.I.; Writing – Original Draft, T.I.; Writing – Review & Editing, N.N and D.W-R.; Visualization, T.I.; Supervision, N.N.and D.W-R.; Project administration, D.W-R.; Funding Acquisition, D.W-R.

References

1. Polášková P, Herszage J, Ebeler SE (2008) Wine flavor: chemistry in a glass. *Chem Soc Rev* 37(11):2478.
2. Dunkel A, et al. (2014) Nature's chemical signatures in human olfaction: A foodborne perspective for future biotechnology. *Angew Chemie - Int Ed* 53(28):7124–7143.
3. Rapp A, Mandery H (1986) Wine aroma. *Experientia* 42(8):873–884.
4. Ebeler SE (2001) Analytical chemistry: Unlocking the secrets of wine flavor. *Food Rev Int* 17(1):45–64.
5. Styger G, Prior B, Bauer FF (2011) Wine flavor and aroma. *J Ind Microbiol Biotechnol* 38(9):1145–59.
6. Duchêne E, et al. (2009) Variation of linalool and geraniol content within two pairs of aromatic and non-aromatic grapevine clones. *Aust J Grape Wine Res* 15(2):120–130.
7. Battilana J, et al. (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(4):653–669.
8. Battilana J, et al. (2011) Functional effect of grapevine 1-deoxy-D-xylulose 5-phosphate synthase substitution K284N on Muscat flavour formation. *J Exp Bot* 62(15):5497–508.
9. Guillaumie S, 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(2):604–615.
10. Lacombe T, et al. (2011) Grapevine European catalogue: Towards a comprehensive list. *Vitis - J Grapevine Res* 50(2):65–68.

11. Strauss CR, et al. (1988) Novel monoterpene disaccharide glycosides of Vitis vinifera grapes and wines. *J Agric Food Chem* 36(8):569–573.

12. Hjelmeland AK, Ebeler SE (2014) Glycosidically Bound Volatile Aroma Compounds in Grapes and Wine: A Review. *Am J Enol Vitic* 66(1):1–11.

13. Maicas S, Mateo JJ (2005) Hydrolysis of terpenyl glycosides in grape juice and other fruit juices: a review. *Appl Microbiol Biotechnol* 67(3):322–35.

14. Mayr CM, et al. (2014) Determination of the importance of in-mouth release of volatile phenol glycoconjugates to the flavor of smoke-tainted wines. *J Agric Food Chem* 62(11):2327–2336.

15. Gunata Z, Bitteur S, Brillouet J-M, Bayonove C, Cordonnier R (1988) Sequential enzymic hydrolysis of potentially aromatic glycosides from grape. *Carbohydr Res* 184:139–149.

16. Romano P, Fiore C, Paraggio M, Caruso M, Capece A (2003) Function of yeast species and strains in wine flavour. *Int J Food Microbiol* 86(1-2):169–180.

17. Ugliano M, Moio L (2006) The influence of malolactic fermentation and Oenococcus oeni strain on glycosidic aroma precursors and related volatile compounds of red wine. *J Sci Food Agric* 86:2468–2476.

18. Williams PJ, Straws CR, Wilson B, Strauss CR, Wilson B (1980) Hydroxylated Linalool Derivatives as Precursors of Volatile Monoterpenes of Muscat Grapes. *J Agric Food Chem* 28(4):766–771.

19. Gunata YZ, Bayonove CL, Baumes RL, Cordonnier RE (1985) The aroma of grapes i. extraction and determination of free and glycosidical- ly bound fractions of some grape aroma components. *J Chromatogr* 331:83–90.

20. Voirin SG, et al. (1992) Analytical methods for monoterpene glycosides in grape and wine. I. XAD-2 extraction and gas chromatographic-mass spectrometric determination of synthetic glycosides. *J Chromatogr* 590(2):313–28.

21. Dunlevy JD, Kalua CM, Keyzers RRA, Boss PK (2009) The production of flavour and aroma compounds in grape berries. *Grapevine Molecular Physiology & Biotechnology*, ed Roubelakis-Angelakis KA (Springer Netherlands, Dordrecht), pp 293–340. 2nd edn.

22. Robinson AL, et al. (2014) Origins of grape and wine aroma. Part 1. Chemical components and viticultural impacts. *Am J Enol Vitic* 65(1):1–24.

23. Wood C, et al. (2008) From wine to pepper: Rotundone, an obscure sesquiterpene, is a potent spicy aroma compound. *J Agric Food Chem* 56(10):3738–3744.

24. Ishii A, et al. (2008) Synergy and Masking in Odor Mixtures: An Electrophysiological Study of Orthonasal vs. Retronasal Perception. *Chem Senses* 33(6):553–561.

25. Dalton P, Doolittle N, Nagata H, Breslin P a (2000) The merging of the senses: integration of subthreshold taste and smell. *Nat Neurosci* 3(5):431–432.

26. Mainland JD, et al. (2013) The missense of smell: functional variability in the human odorant receptor repertoire. *Nat Neurosci* 17(1):114–120.

27. Francis IL, Newton JL (2005) Determining wine aroma from compositional data. *Aust J Grape Wine Res* 11(2):114–126.

28. Matsui K (2006) Green leaf volatiles: hydroperoxide lyase pathway of oxylipin metabolism. *Curr Opin Plant Biol* 9(3):274–280.

29. Zhu B-Q, Xu X-Q, Wu Y-W, Duan C-Q, Pan Q-H (2012) Isolation and characterization of two hydroperoxide lyase genes from grape berries. *Mol Biol Rep* 39(7):7443–7455.

30. Mauricio JC, Moreno J, Zea L, Ortega JM, Medina M (1997) The effects of grape must fermentation conditions on volatile alcohols and esters formed by Saccharomyces cerevisiae. *J Sci Food Agric* 75(2):155–160.

31. Ebeler SE, Thorngate JH (2009) Wine chemistry and flavor: Looking into the crystal glass. *J Agric Food Chem* 57(18):8098–8108.

32. Dembitsky VM (2004) Chemistry and biodiversity of the biologically active natural glycosides. *Chem Biodivers* 1(5):673–781.

33. Saerens SMG, et al. (2008) Parameters affecting ethyl ester production by Saccharomyces cerevisiae during fermentation. *Appl Environ Microbiol* 74(2):454–461.

34. Boss P, et al. (2015) Potential grape-derived contributions to volatile ester concentrations in wine. *Molecules* 20(5):7845–7873.

35. Verstrepen KJ, et al. (2003) Expression levels of the yeast alcohol acetyltransferase genes ATF1, Lg-

ATF1, and ATF2 control the formation of a broad range of volatile esters. *Appl Environ Microbiol* 69(9):5228–5237.

36. Ribéreau-Gayon P, Boidron N, Terrier A (1975) Aroma of Muscat grape varieties. *J Agric Food Chem* 23(6):1042–1047.

37. Guth H (1997) Identification of character impact odorants of different white wine varieties. *J Agric Food Chem* 45(8):3022–3026.

38. Martin DM, et al. (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(1):226.

39. May B, Lange BM, 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* 95:135–144.

40. Luan F, Mosandl A, Degenhardt A, Gubesch M, 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(1-2):353–364.

41. Collu G, et al. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett* 508(2):215–220.

42. Haudenschild C, Schalk M, Karp F, Croteau R (2000) Functional expression of regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha spp.) in Escherichia coli and Saccharomyces cerevisiae. *Arch Biochem Biophys* 379(1):127–36.

43. Ginglinger J-F, et al. (2013) Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in Arabidopsis flowers. *Plant Cell* 25(11):4640–57.

44. Hofer R, et al. (2014) Dual function of the cytochrome P450 CYP76 family from Arabidopsis thaliana in the metabolism of monoterpenols and phenylurea herbicides. *Plant Physiol* 166(3):1149–1161.

45. Williams PJ, Strauss CR, Wilson B (1980) New linalool derivatives in Muscat of Alexandria grapes and wines. *Phytochemistry* 19(6):1137–1139.

46. Marais J, Wyk CJ Van, Rapp A (1992) Effect of storage time, temperature and region on the levels of 1,1,6-trimethyl-1,2-dihydronaphthalene and other volatiles, and on quality of Weisser Riesling wines. *South African J Enol Vitic* 13:33–44.

47. Mendes-Pinto MM (2009) Carotenoid breakdown products the—norisoprenoids—in wine aroma. *Arch Biochem Biophys* 483(2):236–245.

48. Mathieu S, Terrier N, Procureur J, Bigey F, 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(420):2721–2731.

49. Koeduka T, et al. (2006) Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. *Proc Natl Acad Sci U S A* 103(26):10128–10133.

50. Gallage NJ, et al. (2014) Vanillin formation from ferulic acid in Vanilla planifolia is catalysed by a single enzyme. *Nat Commun* 5(May):4037.

51. Spillman PJ, Pollnitz AP, Liacopoulos D, Skouroumounis GK, Sefton MA (1997) Accumulation of vanillin during barrel-aging of white, red, and model wines. *J Agric Food Chem* 45(7):2584–2589.

52. Chatonnet P, Dubourdie D, Boidron J-N, Pons M (1992) The origin of ethylphenols in wines. *J Sci Food Agric* 60(2):165–178.

53. Hayasaka Y, et al. (2010) Glycosylation of smoke-derived volatile phenols in grapes as a consequence of grapevine exposure to bushfire smoke. *J Agric Food Chem* 58(20):10989–10998.

54. Chatonnet P, Dubourdieu D, Boidron JN, Lavigne V (1993) Synthesis of volatile phenols by Saccharomyces cerevisiae in wines. *J Sci Food Agric* 62(2):191–202.

55. Tieman D, et al. (2006) Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proc Natl Acad Sci U S A* 103(21):8287–8292.

56. Vuralhan Z, et al. (2005) Physiological characterization of the ARO10-dependent, broad-substrate-specificity 2-oxo acid decarboxylase activity of Saccharomyces cerevisiae. *Appl Environ Microbiol* 71(6):3276–3284.

57. Jaillon O, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161):463–7.

58. Yilmaztekin M, Kocabey N, Hayaloglu AA (2015) Effect of maceration time on free and bound volatiles of red wines from cv. Karaoğlan (Vitis vinifera L.) grapes grown in Arapgir, Turkey. *J Food Sci* 80(3):C556–C563.

59. Sánchez-Palomo E, Alonso-Villegas R, González Viñas MA (2015) Characterisation of free and glycosidically bound aroma compounds of La Mancha Verdejo white wines. *Food Chem* 173:1195–1202.

60. Selli S, et al. (2006) Effect of skin contact on the free and bound aroma compounds of the white wine of Vitis vinifera L. cv Narince. *Food Control* 17:75–82.

61. Toci AT, et al. (2012) Free and bound aroma compounds characterization by GC-MS of Negroamaro wine as affected by soil management. *J Mass Spectrom* 47(9):1104–12.

62. Oliveira JM, Oliveira P, Baumes RL, Maia MO (2008) Volatile and glycosidically bound composition of Loureiro and Alvarinho wines. *Food Sci Technol Int* 14(4):341–353.

63. Becatti E, Genova G, Ranieri A, Tonutti P (2014) Postharvest treatments with ethylene on Vitis vinifera (cv Sangiovese) grapes affect berry metabolism and wine composition. *Food Chem* 159:257–266.

64. Gómez García-Carpintero E, et al. (2011) Aroma characterization of red wines from cv. Bobal grape variety grown in La Mancha region. *Food Res Int* 44(1):61–70.

65. García-Carpintero EG, Sánchez-Palomo E, Gallego M a G, González-Viñas M a. (2011) Volatile and sensory characterization of red wines from cv. Moravia Agria minority grape variety cultivated in La Mancha region over five consecutive vintages. *Food Res Int* 44(5):1549–1560.

66. Gómez García-Carpintero E, Sanchez Palomo E, Gómez Gallego MA, Gonzalez Vinas MA (2012) Free and bound volatile compounds as markers of aromatic typicalness of Moravia Dulce , Rojal and Tortosí red wines. *Food Chem* 131:90–98.

67. Boido E, et al. (2003) Aroma composition of Vitis vinifera cv . Tannat: the typical red wine from Uruguay. *J Agric Food Chem* 51(18):5408–5413.

68. Ghaste M, et al. (2015) Chemical composition of volatile aroma metabolites and their glycosylated precursors that can uniquely differentiate individual grape cultivars. *Food Chem* 188:309–319.

69. Bureau SM, Razungles AJ, Baumes RL (2000) The aroma of Muscat of Frontignan grapes: Effect of the light environment of vine or bunch on volatiles and glycoconjugates. *J Sci Food Agric* 80(14):2012–2020.

70. Canosa P, Oliveira JM, Masa A, Vilanova M (2011) Study of the volatile and glycosidically bound compounds of minority Vitis vinifera red cultivars from NW Spain. *J Inst Brew* 117(3):462–471.

71. Fenoll J, Manso A, Hellin P, Ruiz L, Flores P (2009) Changes in the aromatic composition of the Vitis vinifera grape Muscat Hamburg during ripening. *Food Chem* 114(2):420–428.

72. Genisheva Z, Oliveira JM (2009) Monoterpenic characterization of white cultivars from Vinhos Verdes appellation of origin (North Portugal). *J Inst Brew* 115(4):308–317.

73. Ugliano M, Moio L (2008) Free and hydrolytically released volatile compounds of Vitis vinifera L. cv. Fiano grapes as odour-active constituents of Fiano wine. *Anal Chim Acta* 621:79–85.

74. Vrhovsek U, et al. (2014) Quantitative metabolic profiling of grape, apple and raspberry volatile compounds (VOCs) using a GC/MS/MS method. *J Chromatogr B* 966:132–139.

75. Hellín P, Manso A, Flores P, Fenoll J (2010) Evolution of aroma and phenolic compounds during ripening of "Superior Seedless" grapes. *J Agric Food Chem* 58(10):6334–6340.

76. Lopez-Tammames E, et al. (1997) Potential Aroma in Several Varieties of Spanish Grapes. *J Agric Food Chem* 45:1729–1735.

77. Baek HH, Cadwallader KR (1999) Contribution of Free and Glycosidically Bound Volatile Compounds to the Aroma of Muscadine Grape Juice. *J Food Sci* 64(3):441–444.

78. Felix D, Melera A, Seibl J, Kovats E (1963) Zur Kenntnis ätherischer Öle. 2. Mitteilung. Die Struktur der sogenannten «Linalooloxide». *Helv Chim Acta* 46(5):1513–1536.

79. Kreis P, Dietrich a., Mosandl a. (1996) Elution Order of the Furanoid Linalool Oxides on Common Gas Chromatographic Phases and Modified Cyclodextrin Phases. *J Essent Oil Res* 8(4):339–341.

80. Luan F, Mosandl A, Gubesch M, Wüst M (2006) Enantioselective analysis of monoterpenes in different grape varieties during berry ripening using stir bar sorptive extraction- and solid phase extraction-enantioselective-multidimensional gas chromatography-mass spectrometry. *J Chromatogr A* 1112(1-2):369–74.

81. Winterhalter P, Katzenberger D, Schreier P (1986) 6,7-Epoxylinalool and related oxygenated terpenoids from Carica papaya fruit. *Phytochemistry* 25(6):1347–1350.

82. Chassagne D, Boulanger R, Crouzet J (1999) Enzymatic hydrolysis of edible Passiflora fruit glycosides. *Food Chem* 66:281–288.

83. R Development Core Team R (2011) R: A Language and Environment for Statistical Computing. *R Found Stat Comput* 1(2.11.1):409.

Supplementary information

Table S1. List of grape and wine volatiles reported in at least three studies with mean retention indices.

Compound	RI	Compound	RI
Aliphatic alcohols			
Methanol	879	1-hexanol	1357
1-propanol	1045	(*E*)-3-hexen-1-ol	1371
2-methyl-propanol (isobutanol)	1085	3-ethoxy-1-propanol	1377
1-butanol	1146	(*Z*)-3-hexen-1-ol	1387
2-methyl-1-butanol	1212	(*E*)-2-hexen-1-ol	1408
3-methyl-1-butanol (isoamyl alcohol)	1213	(*Z*)-2-hexen-1-ol	1414
1-pentanol	1249	1-octen-3-ol	1452
3-Methyl-3-buten-1-ol	1252	1-heptanol	1460
2-heptanol	1318	2-ethyl-1-hexanol	1497
4-methyl-1-pentanol	1319	2,3-butanediol	1547
(*Z*)-2-penten-1-ol	1321	1-octanol	1561
3-methyl-1-pentanol	1326		
Aliphatic ketones			
3-hydroxy-2-butanone (acetoin)	1284		
Aliphatic aldehydes			
Hexanal	1095	(*E*)-2-hexenal	1223
Aliphatic esters			
Ethyl acetate	836	Ethyl 3-hydroxybutanoate	1509
Isobutyl acetate	1012	Diethyl malonate	1597
Ethyl butanoate	1055	Methyl decanoate	1629
Ethyl 3-methylbutanoate	1070	Ethyl decanoate	1650
Butyl acetate	1070	Diethyl succinate	1688
Isoamyl acetate	1134	Ethyl methyl succinate	1743
Ethyl hexanoate	1218	Methyl 4-hydroxybutyrate	1783
Hexyl acetate	1285	Ethyl 4-hydroxybutyrate	1822
Ethyl lactate	1336	Ethyl dodecanoate	1852
Ethyl octanoate	1437	Diethyl malate	2057
2-Hydroxy 2-methylpropyl butyrate	1461	Ethyl succinate	2359
Aliphatic acids			
Acetic acid	1440	(*E*)-2-hexenoic acid	1942
Propanoic acid	1524	Octanoic acid	2055
2-methylpropanoic acid (isobutyric acid)	1579	Nonanoic acid	2142
Butanoic acid (butyric acid)	1617	Decanoic acid	2293

Isovaleric acid	1659	Dodecanoic acid	2447
Pentanoic acid	1705	Tetradecanoic acid (myristic acid)	2669
Hexanoic acid	1835		

Monoterpenes

Limonene	1202	Nerol	1811
trans-linalool oxide (F)	1447	Geraniol	1846
cis-linalool oxide (F)	1476	exo-2-hydroxy-1,8-cineole	1851
Linalool	1547	7-hydroxylinalool (diendiol I)	1950
Terpinen-4-ol	1599	7-hydroxy-6,7-dihydrolinalool	1988
Hotrienol	1624	6-hydroxylinalool (diendiol II)	2152
alpha-terpineol	1697	8-hydroxy-6,7-dihydrolinalool	2217
trans-linalool oxide (P)	1727	(*Z*)-8-hydroxylinalool	2274
Citral/ (*E*)-geranial	1743	(*E*)-8-hydroxylinalool	2304
Citronellol	1764	Geranic acid	2317
cis-linalool oxide (P)	1766	*p*-menth-1-ene-7,8-diol	2502

Norisoprenoids

4-Oxo-isophorone	1703	Dehydrovomifoliol	2554
beta-damascenone	1804	3-Oxo-alpha-ionol	2617
3,4-dihydro-3-oxo-actinidol I	2382	3-Hydroxy-7,8-dihydro-beta-ionol	2684
3,4-dihydro-3-oxo-actinidol III	2412	3-Hydroxy-7,8-dehydro-beta-ionol	2732
3,4-Dihydro-3-oxo-actinidol II	2420	Vomifoliol	3148
3-Hydroxy-beta-damascone	2553		

Sulfur-containing volatiles

3(2H)-2-methyldihydro-thiophenone	1505	3-Methylthiopropanoic acid	1757
3-(methylthio)-1-propanol (methionol)	1719	1,2-Benzothiazole	1899

Volatile phenols

Methyl salicylate	1779	4-Methoxyphenylethyl alcohol	2302
Guaiacol (2-methoxyphenol)	1869	4-vinylphenol	2388
Phenol	1982	(*E*)-4-Allylsyringol	2424
4-Ethylguaiacol	2055	Vanillin	2550
4-methyl phenol (p-cresol)	2087	Methyl vanillate	2568
3-methyl phenol (m-cresol)	2119	Acetovanillone (Apocynin)	2654
Eugenol	2181	Ethyl vanillate	2676
4-ethylphenol	2187	3,4-dimethoxyphenol	2756
p-vinylguaiacol	2200	Zingerone	2796
4-Hydroxy-2-methyl acetophenone	2212	Tyrosol	3008
Syringol (2,6-dimethoxyphenol)	2243	Methyl vanillyl eter	3030
Isoeugenol	2302	3,4,5-trimethoxyphenol	3032
Methyl salicylate	1779	4-Methoxyphenylethyl alcohol	2302
Guaiacol (2-methoxyphenol)	1869	4-vinylphenol	2388

Benzenoids

Benzaldehyde	1519	Benzyl alcohol	1883
Phenylacetaldehyde	1639	2-phenylethanol	1913
Acetophenone	1667	Benzenepropanol	2037
Ethylbenzaldehyde	1728	Ethyl cinnamate	2286
Benzyl acetate	1735	Benzoic acid	2391
Ethyl benzene acetate	1782	Benzylacetic acid	2502
2-phenyl ethyl acetate	1809	Cinnamic acid	3045

Nitrogen-containing volatiles

N-ethyl-benzamine	1750	*N*-(2-phenylethyl)-acetamide	2563

Other volatiles

gamma-butyrolactone	1628	2,3-dihydrobenzofuran	2377
Furaneol	2023		

Table S2. Key wine odorants. Included key wine odorants are compounds that were detected in at least three profiling experiments included in this meta-analysis, as well as identified as one of the key wine odorants in the Francis and Newton meta-analysis (27). Potential key wine odorants are compounds that were detected in at least three profiling experiments included in this meta-analysis as well as identified as one of the key food odorants in the Dunkel *et al.* meta-analysis (2). Overlooked key wine odorants are compounds that were not detected in at least three profiling experiments included in this meta-analysis, but were identified as one of the key wine odorants in the Francis and Newton meta-analysis (27).

Class	Included key wine odorants	Potential key wine odorants	Overlooked key wine odorants
Aliphatic alcohols	(*Z*)-3-hexen-1-ol	2-methyl-1-butanol	
	1-hexanol	1-butanol	
	Isobutanol		
	Isoamyl alcohol		
Aliphatic ketones	Acetoin		Butane-2,3-dione
Aliphatic aldehydes		(*E*)-2-hexenal	Acetaldehyde
		Hexanal	
Aliphatic acids	2-methylpropanoic acid	Pentanoic acid	
	Decanoic acid		
	Octanoic acid		
	Propanoic acid		
	Hexanoic acid		
	Acetic acid		
	Butanoic acid		
	Isovaleric acid		
Aliphatic esters	Ethyl acetate	Butyl acetate	ethyl 2- and 3-methylbutanoate
	Ethyl decanoate	Hexyl acetate	
	Ethyl butanoate		
	Ethyl hexanoate		
	Ethyl octanoate		
	Isoamyl acetate		
	Ethyl 3-methylbutanoate		
Monoterpenes	Geraniol	alpha-terpineol	Wine lactone
	Linalool	Limonene	*cis*-rose oxide
Norisoprenoids	beta-damascenone		beta-ionone
Sulfur-containing volatiles	Methionol		Ethyl 2-methylpropanoate
			Dimethyl sulfide
			2-methylfuran-3-thiol
			3-sulfanyl-1-hexanol
			3-sulfanylhexyl acetate
			4-methyl-4-sulfanylpentan-2-one
			Benzenemethanethiol
Volatile phenols	4-Ethylguaiacol	*m*-cresol	
	Eugenol	*p*-cresol	

	Guaiacol	4-vinylphenol	
	p-vinylguaiacol	Isoeugenol	
	Vanillin	Syringol	
	4-ethylphenol	Ethyl phenylacetate	
Benzenoids	Ethyl cinnamate		Ethyl dihydrocinnamate
	Phenethyl acetate		
	Phenylethyl alcohol		
	Phenylethanal		
Nitrogen-containing volatiles			2-methoxy-3-(2-methylpropyl)pyrazine
Others	Furaneol		1,1-diethoxyethane
			Sotolon
			gamma-nonalactone
			gamma-decalactone
			(*Z*)-oak lactone
			gamma-dodecalactone
			gamma-(*Z*)-6-dodecenolactone

Table S3. Parameters of linear regression describing relationships between linalool and each of its oxygenated derivatives (ln[Derivative] = a·ln[Linalool] + b).

Compound	M_w	b	a	R^2	p-value
Linalool	154.25	0.00	1.00		
trans-linalool oxide (F)	170.25	-2.08	0.55	0.56	2.35E-12
trans-linalool oxide (P)	170.25	-1.69	0.59	0.57	2.62E-11
cis-linalool oxide (P)	170.25	-2.28	0.57	0.48	1.48E-08
7-hydroxylinalool	170.25	-0.34	0.82	0.60	8.86E-10
6-hydroxylinalool	170.25	-1.84	0.71	0.67	3.18E-07
8-hydroxy-6,7-dihydrolinalool	172.26	-2.76	0.44	0.45	3.71E-05
(*Z*)-8-hydroxylinalool	170.25	-0.94	0.59	0.68	3.72E-11
(*E*)-8-hydroxylinalool	170.25	-0.57	0.49	0.51	1.11E-07
cis-linalool oxide (F)	170.25	-2.16	0.42	0.34	5.05E-07
Hotrienol	152.23	-2.69	0.50	0.34	7.10E-04

Chapter 2
Biosynthesis of terpenoids in plants

Terpenes (also isoprenoids or terpenoids) are the largest group of plant natural products with an incredible structural and functional diversity (1). Regular terpenes have a backbone composed of a variable number of 5-carbon (5C) isoprene units. Terpenes are grouped by the number of isoprene units: monoterpenes have 2 isoprene units, sesquiterpenes 3, diterpenes 4, until tetraterpenes which have 8 isoprene units (Figure 1). The name hemiterpene is sometimes used for isoprene, a 5-carbon compound which some plants emit in large quantities as protection from abiotic stress (2). Monoterpenes and sesquiterpenes (10C and 15C isoprenoids, respectively), are a large and important class of volatile organic compounds: they are emitted by flowers to attract pollinators, deter herbivores and prevent fungal infections. They are also produced in fruits to attract seed dispersers (3). Diterpenes (20C isoprenoids) are not volatile, but nonetheless have important ecological roles: they serve in plant defense as phytoalexins (antimicrobial compounds) and antiherbivore compounds (4). Sesterterpenes (C25) are present in some plants (5) but their function is not very well documented. Triterpenes (30C isoprenoids) are important for membrane structure, as well as in plant defense (e.g. saponins, which probably act through membrane disruption) (6). The most important function of tetraterpenes (40C, also known as carotenoids) is light harvesting in thylakoid membranes. Natural rubber, an economically important high molecular weight polymer, is produced by polymerization of 5C building blocks (7). In addition to the above mentioned functions, many isoprenoids act as plant hormones, e.g. gibberellins (C20), brassinosteroids (C30), abscisic acid and strigolactones (both C40-derived).

Common precursors of all terpenoids are two 5-carbon isomers dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP). These precursors can be synthesized via two independent and spatially separated pathways: the cytosolic mevalonate (MVA) pathway and the chloroplastic 2-methyl-D-erythrol-4-phosphate (MEP) pathway (8, 9) (Figure 1). The next step in terpene biosynthesis pathway is condensation of the C5 units, a reaction catalyzed by short-chain prenyltransferases. Farnesylpyrophosphate synthase (FPPS), which synthesizes the 15C precursor to sesquiterpenes and triterpenes, is located in the cytosol, whereas geranylpyrophosphate synthase (GPPS) and geranylgeranylpyrophosphate synthase (GGPPS), which synthesize C10 and C20 precursors to monoterpenes and diterpenes, respectively, are located in the chloroplast (3). Squalene, the C30 precursor, and phytoene, the C40 precursor, are synthesized by head-to-head condensation of FPP and GGPP,

respectively. Because of compartmentalization of short-chain prenyltransferases, the mevalonate pathway mostly supplies precursors for sesqui- and triterpenes biosynthesis, whereas mono-, di- and tetraterpenes source their precursors from the MEP pathway. Until recently precursors with all-trans configuration were considered the only common terpene precursors. However, cis-short-chain-prenyltransferases were found to synthesize neryldiphosphate (NPP) and *Z,Z*-farnesyldiphosphate in tomato species (10, 11).

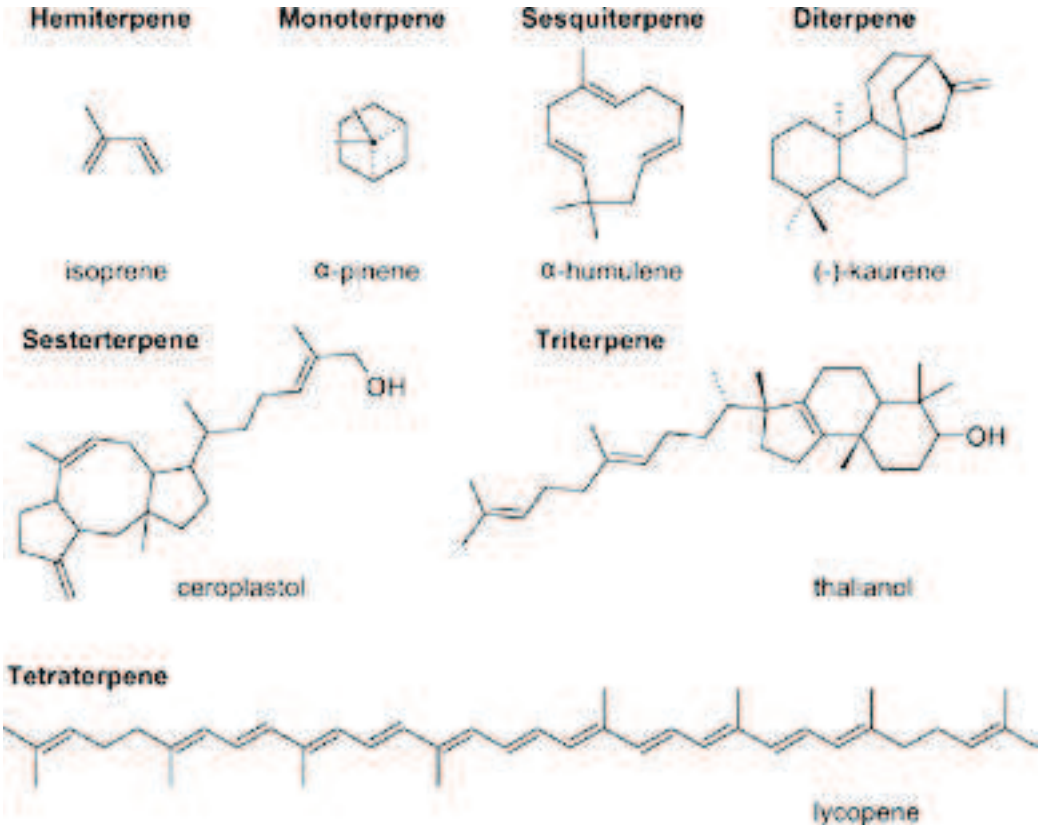

Figure 17. Structural diversity of terpenoids in plants. Examples of different structural types of plant terpenoids.

The final step in terpene biosynthesis (C10-C20) is catalyzed by terpene synthases (TPS). These enzymes use isoprenyl diphosphate substrates and are able to transform them into a remarkable variety of cyclic and acyclic structures. The reaction proceeds via a carbocation intermediate, which can undergo cyclisations or rearrangement reactions in the TPS active site, until the reaction is stopped by proton loss or nucleophile (water) addition (12). In the latter case the resulting product contains a hydroxyl functional group. Biosynthesis of some diterpenes proceeds in two steps: first a class II diTPS synthesizes a bicyclic diphosphate intermediate, which is then used as a substrate by the class I diTPS (4). Recently, TPS-independent biosynthesis of geraniol has been discovered in rose petals: a so-called Nudix

hydrolase (NUDX1) hydrolyses geraniol diphosphate to geraniol phosphate, which is then further hydrolyzed by a phosphatase to yield geraniol (13).

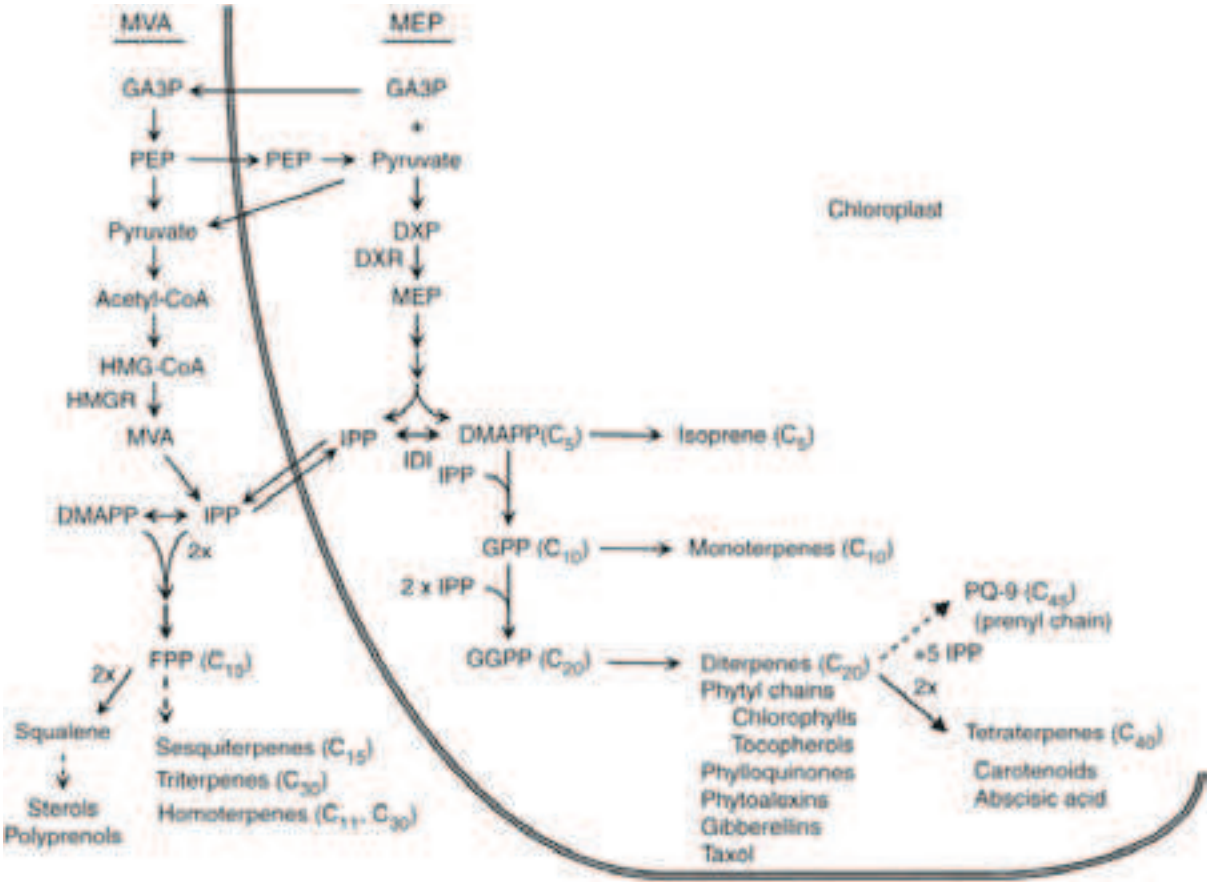

Figure 18. The two biosynthetic pathways leading to isoprenoids: the cytosolic mevalonate pathway (MVA) and the chloroplastic 2-methyl-D-erythrol-4-phosphate (MEP) pathway (2). GA3P – glycerolaldehyde-3-phosphate, PEP – phosphoenolpyruvate, CoA – Coenzyme A, HMG-CoA – 3-hydroxy-3-methylglutaryl-coenzyme A, MVA-mevalonic acid, DMAPP – dimethylallyl pyrophosphate, IPP – isopentenylpyrophosphate, FPP – farnesyl pyrophosphate, DXP - 1-deoxy-D-xylulose 5-phosphate, MEP - 2-methyl-D-erythritol 4-phosphate, GPP – geranyl pyrophosphate, GGPP – geranylgeranyl pyrophosphate, PQ – plastoquinone.

Many terpenes in plants undergo further modifications: about 97% of the known terpenes are oxygenated (14), a large part of which are likely products of cytochromes P450. These enzymes are known to oxidize mono- to tetraterpenes. P450s act on different terpenoids across plant species and their genes often cluster on genomes with those of terpene synthases from the same pathways (15). They usually introduce a hydroxyl functional group, but also carbonyl or carboxyl products are known. Some P450s also catalyze oxidative cleavage of terpenoids (16). Another very common modification of terpenes is glycosylation, catalyzed by family 1 glycosyltransferases, also known as UDP-glycosyltransferases or UGTs. Glycosylation increases water solubility of terpenes and can regulate their homeostasis, storage and transport (17). Other modifications are rarer and often restricted to particular classes of terpenes, but are nonetheless important. For example, oxidative cleavage of

carotenoids by carotenoid cleavage dioxygenases produces apocarotenoids, which include hormones strigolactones and norisoprenoids, an important class of aroma compounds (18).

References

1. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. *Nat Chem Biol* 3(7):408–414.
2. Vickers CE, Gershenzon J, Lerdau MT, Loreto F (2009) A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nat Chem Biol* 5(5):283–291.
3. Dudareva N, Klempien A, Muhlemann JK, Kaplan I (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol* 198(1):16–32.
4. Zerbe P, et al. (2013) Gene Discovery of Modular Diterpene Metabolism in Nonmodel Systems. *Plant Physiol* 162(2):1073–1091.
5. Wang L, Yang B, Lin X-P, Zhou X-F, Liu Y (2013) Sesterterpenoids. *Nat Prod Rep* 30(3):455. Available at: http://xlink.rsc.org/?DOI=c3np20089b.
6. Phillips DR, Rasbery JM, Bartel B, Matsuda SP (2006) Biosynthetic diversity in plant triterpene cyclization. *Curr Opin Plant Biol* 9(3):305–314.
7. Cornish K (2001) Similarities and differences in rubber biochemistry among plant species. *Phytochemistry* 57(7):1123–1134.
8. Lichtenthaler HK, Schwender J, Disch A, Rohmer M (1997) Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. *FEBS Lett* 400(3):271–274.
9. Rohmer M (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat Prod Rep* 16(5):565–574.
10. Schilmiller AL, et al. (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc Natl Acad Sci U S A* 106(26):10865–70.
11. Sallaud C, et al. (2009) A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato Solanum habrochaites. *Plant Cell* 21(1):301–317.
12. Degenhardt J, Köllner TG, Gershenzon J (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70(15-16):1621–37.
13. Magnard J-L, et al. (2015) Biosynthesis of monoterpene scent compounds in roses. *Science (80-)* 349(6243):81–83.
14. Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos Trans R Soc Lond B Biol Sci* 368(1612):20120426.
15. Boutanaev AM, et al. (2015) Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci* 112(1):E81–E88.
16. Lee S, et al. (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. *Proc Natl Acad Sci U S A* 107(49):21205–10.
17. Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J* 66(1):182–93.
18. Auldridge ME, McCarty DR, Klee HJ (2006) Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Curr Opin Plant Biol* 9(3):315–321.

Chapter 3
Cytochromes P450

Cytochromes P450 are a large enzyme family of monooxygenases present in all kingdoms of life. They catalyze the insertion of an oxygen atom to relatively inert positions in organic molecules. Their substrates are structurally very diverse, which also reflect the diversity of their functional roles: detoxification of xenobiotics, including drugs in humans, to biosynthesis of basic and specialized metabolites.

Cytochromes P450 are named after their unusual spectral properties. In the 1950s and 1960s a carbon monoxide-binding proteinaceous pigment was discovered in rat and rabbit liver microsomes (1–3). Like other heme-containing proteins (or cytochromes) this newly-discovered pigment had an intense absorption peak around 400, also called Soret peak or band. Unlike other cytochromes, this peak shifted to 450 nm upon binding of carbon monoxide to the reduced form of the protein, hence the name "P-450". Based on this property, a simple spectrophotometric method, also called CO-difference spectroscopy, is still widely used for characterization of cytochromes P450: the difference in molar extinction coefficient (ε = 91 $mM^{-1}cm^{-1}$) between 450 and 490 nm in the difference spectrum allows for the estimation of cytochrome P450 concentration (4). Enzymatic activity of cytochromes P450 was first demonstrated in 1965 (5).

At the heart of the remarkable P450 catalytic activity is their ability to bind and activate molecular oxygen. Their active site contains a heme prosthetic group; in addition to the four porphyrine nitrogen atoms, the iron ion is coordinated with a cysteine residue of the enzyme via a thiolate bond. In the P450 resting state the sixth ligand of the Fe(III) ion is usually a water molecule Figure 1-**1**). This molecule can be displaced by the substrate, which influences the redox potential of the iron atom and facilitates its reduction to ferrous (III) state (Figure 1-**3**). This reduced intermediate binds molecular oxygen, which is first reduced again and then protonated to form the ferric hydroperoxo intermediate (Figure 1-**5**). A second protonation causes the cleavage of the bond between the two atoms of oxygen and an exit of a water molecule. The resulting radical ferryl oxo intermediate, also named compound I (Figure 1-**6**), abstracts a hydrogen atom from the substrate. The resulting iron-coordinated hydroxyl group rapidly recombines with the substrate radical (Figure 1-**7**). This unique reaction mechanism results in the insertion of an oxygen atom into relatively inert substrates, which are difficult to functionalize by other enzymes. P450s usually catalyze C-hydroxylation reactions, but many other types of P450-catalyzed reactions have been described:

epoxidation, heteroatom oxidation, heteroatom release, group migration, C-C coupling and many more (7, 8).

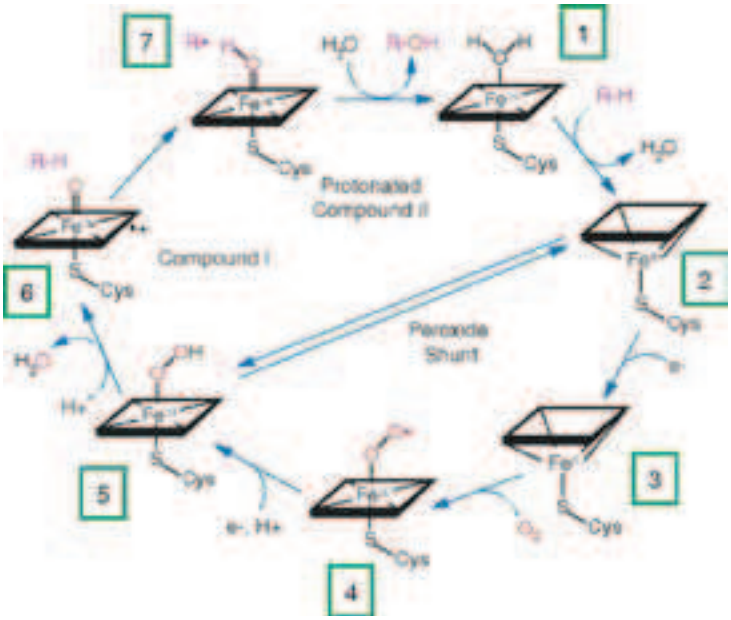

Figure 19. The cytochrome P450 catalytic cycle. A water molecule is bound in the active site in the resting state **1**. Substrate displaces the water molecule **2** which triggers reduction of the Fe ion from (III) to (II) **3**. A molecule of oxygen binds into the active site **4** and is protonated **5** after a second iron reduction. The bond between two atoms of oxygen is cleaved **5** and a water molecule is released. The resulting ferryl oxo radical **6** abstracts a proton from the substrate **7.** The reaction is terminated when the resulting hydroxide combines with the substrate radical (6).

The electrons required for oxygen cleavage are supplied from reduced cofactors nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH). Cytochromes P450 require auxiliary enzymes, which transfer the electrons from the cofactor to P450. In bacteria, the most common auxiliary enzymes are ferredoxin (Fe-S protein) or flavodoxin (a FMN-containing protein), which in turn receive electrons from ferredoxin or flavodoxin reductase, respectively. In eukaryotes the most common electron donor is an enzyme called P450 reductase (also CPR or POR), which contains both flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) prosthetic groups that transport the electrons from the NADPH to the P450. Most eukaryotic P450 systems also differ from bacterial in the fact that both P450 and CPR are anchored to the endoplasmic reticulum (ER) membrane via a N-terminal transmembrane anchor, whereas all components of bacterial P450 systems are soluble and localized in the cytosol. In some cases, plant P450s also require another C-terminally membrane anchored electron donor, cytochrome b_5 to transfer the second electron (9). Finally, CPRs are not specific for their P450 substrate, as higher plants have no more than three reductase genes (usually two) and more than 400 P450 genes (10).

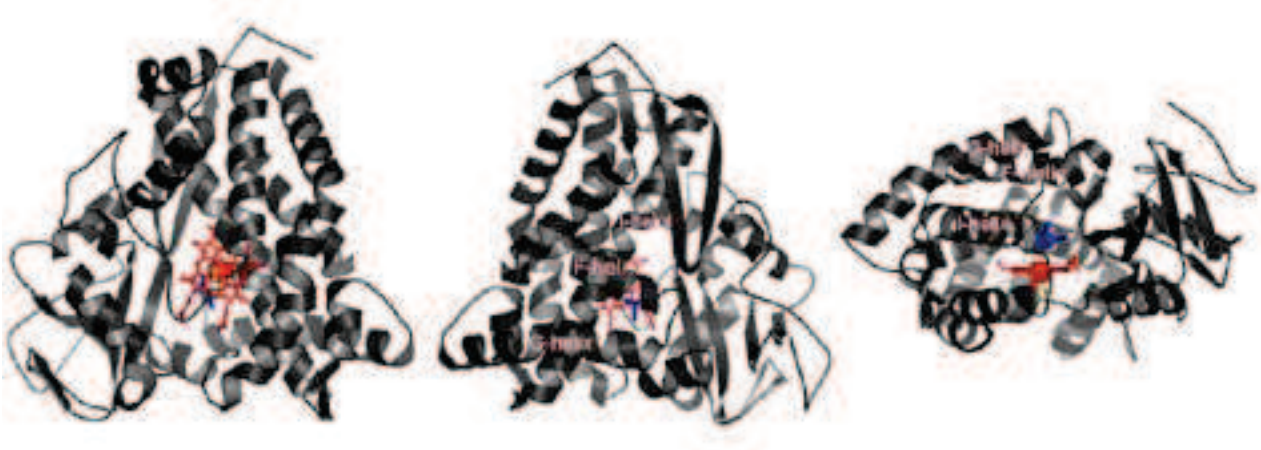

Figure 20. Three dimensional structure of the model cytochrome P450 CYP101A1 (P450cam, 2CPP) from *Pseudomonas putida* with bound substrate (camphor). Heme prosthetic group is displayed in red, cysteine residue participating in thiolate bond with iron in yellow and camphor (substrate) in blue. The protein structure is displayed from three different angles: from the so-called (heme) proximal side (left), the distal side (centre) and a lateral view along the heme axis.

With the advances in sequencing techniques and increase in the number of sequenced genomes, the number of discovered P450 genes has skyrocketed in the last years. The NCBI RefSeq database currently contains almost 50 000 unique cytochrome P450 genes. Although not present in all species (some bacteria, for example *E. coli*, do not have any P450 genes), they are spread in all kingdoms of life, including viruses, which suggests their common origin before accumulation of oxygen in the Earth atmosphere. The most distant P450s can now show less than 20% sequence identity. More closely related P450s are grouped into families which usually share at least 40% amino acid sequence identity; these families are further partitioned into subfamilies with at least 55% amino acid sequence identity. P450s nomenclature is further refined based on phylogenetic analysis. The name of each P450 consists of a prefix CYP, number of the family, letter of the subfamily and number of the isoform.

Despite very low sequence identity between distantly related P450s, numerous 3D structures solved to date show that their structural fold is highly conserved: a two-domain triangular structure with centrally-bound heme prosthetic group and adjacent substrate cavity (Figure 2). The largest domain, also named α-domain, mostly consists of α-helices, whereas the smaller β-domain is rich in β-sheets. α-helices in P450s have been assigned labels from A to L. I-helix (Figure 2) participates in binding of the oxygen molecule, whereas F and G helices modulate the access to the active site. The loop between F and G helices is highly variable and participates in substrate recognition.

References

1. Klingenberg M (2003) Pigments of rat liver microsomes. *Arch Biochem Biophys* 409(1):2–6.
2. Omura T, Sato R (1964) The Carbon Monoxide-biding pigment of Liver Microsomes. *J Biol Chem* 239(7):2370–2378.
3. Omura T, Sato R (1962) A new cytochrome in liver microsomes. *J Biol Chem* 237(April):1375–1376.
4. Omura T, Sato R (1963) Fractional solubilization of haemoproteins and partial purification of carbon monoxide-binding cytochrome from liver microsomes. *Biochim Biophys Acta* 71:224–226.
5. Cooper DY, Levin S, Narasimhulu S, Rosenthal O (1965) Photochemical Action Spectrum of the Terminal Oxidase of Mixed Function Oxidase Systems. *Science* 147(13):400–402.
6. Ener ME, Lee Y-T, Winkler JR, Gray HB, Cheruzel L (2010) Photooxidation of cytochrome P450-BM3. *Proc Natl Acad Sci U S A* 107(44):18783–18786.
7. Mizutani M, Sato F (2011) Unusual P450 reactions in plant secondary metabolism. *Arch Biochem Biophys* 507(1):194–203.
8. Lamb DC, Waterman MR (2013) Unusual properties of the cytochrome P450 superfamily. *Philos Trans R Soc B Biol Sci* 368(1612):20120434–20120434.
9. De Vetten N, et al. (1999) A cytochrome b5 is required for full activity of flavonoid 3', 5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. *Proc Natl Acad Sci U S A* 96(2):778–783.
10. Jensen K, Møller BL (2010) Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry* 71(2-3):132–141.

Chapter 4
Plant monoterpenols: role and oxidative metabolism

Tina Ilc, Claire Parage, Benoît Boachon, Nicolas Navrot and Danièle Werck-Reichhart

Institut de Biologie Moléculaire des Plantes du Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, 67000 Strasbourg, France

Abstract

Plants use monoterpenols as building blocks for production of functionally and structurally diverse molecules, which they use for interactions with other organisms: pollinators, flower visitors, herbivores, fungal or microbial pathogens. For humans, many of these monoterpenol derivatives are economically important because of their pharmaceutical, nutraceutical, flavor or fragrance applications. The biosynthesis of these derivatives is to a large extent catalyzed by the enzymes from the cytochrome P450 superfamily. Here we review the knowledge on monoterpenol metabolism in plants with special focus on recent elucidations of oxidation steps leading to diverse linalool and geraniol derivatives. We evaluate the common points between oxidation pathways of these two monoterpenols, such as involvement of the CYP76 family, and highlight the differences. Finally, we discuss the missing steps and other open questions in the biosynthesis of oxygenated monoterpenol derivatives.

Introduction

Monoterpenols and their derivatives are a group of plant secondary metabolites with important ecological roles, such as flower or fruit scent. In addition they act as precursors to numerous allelochemicals such as iridoids or monoterpenoid indole alkaloids (MIAs). They are thus an economically important raw material in flavor or fragrance, nutraceutical and pharmaceutical industry. To accomplish such a variety of roles, diversification of chemical structures is paramount. Monoterpenols belong to the large family of monoterpenes, which are compounds composed of two isoprenyl units (to form 10-carbon structures). Monoterpenes are products of terpene synthase enzymes, which convert a monoterpenyl diphosphate substrate (usually geranyl or neryl diphosphate, GPP or NPP respectively) into a variety of structures. The reaction starts with ionization of monoterpenyl diphosphate, which yields a carbocation intermediate, prone to rearrangements and cyclisation (Figure 1). The carbocation then follows one of two fates, depending on the catalytic site of the terpene synthase: it either loses a proton and forms a cyclic or an aliphatic hydrocarbon, or is attacked by hydroxyl ion to form a monoterpenol (1). Recently, a terpene synthase-independent synthesis of geraniol was discovered in roses (*Rosa x hybrida*) (2). A cytosolic enzyme Nudix hydrolase (NUDX1), which hydrolyses geranyl diphosphate, is the main geraniol producing enzyme in rose flowers.

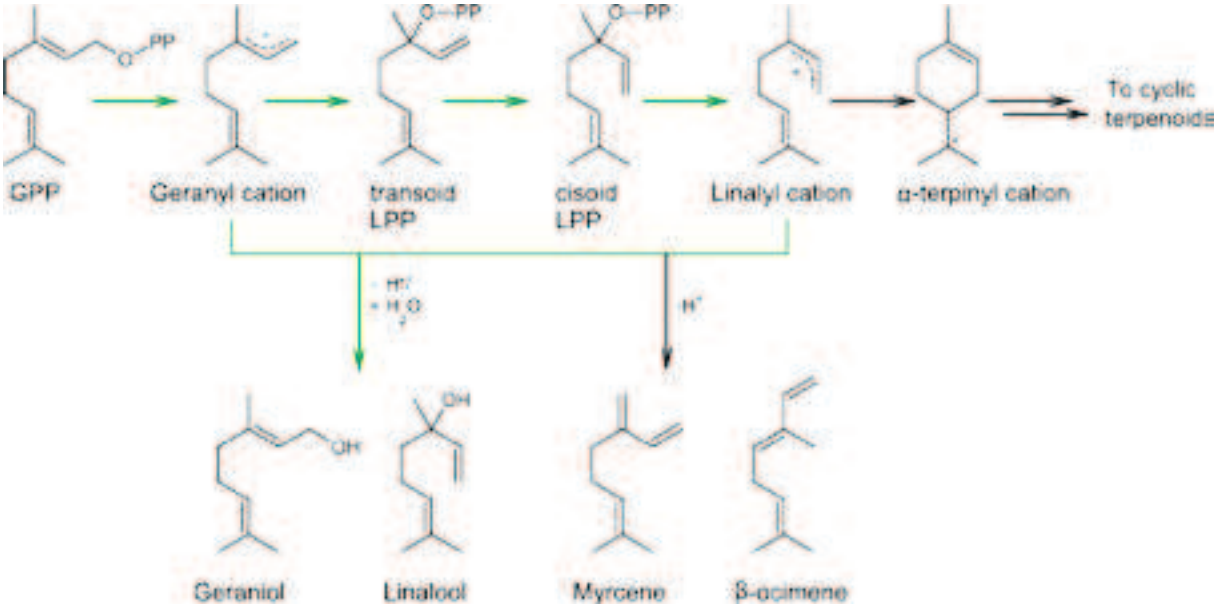

Figure 21. Terpene synthase catalyzed biosynthesis of monoterpenols from geranyl diphosphate (GPP) (adapted from (1)). Terpene synthases cleave the diphosphate group from the substrate, which results in the formation of a carbocation intermediate that can further rearrange. The reaction can be terminated by proton abstraction, which yields hydrocarbon monoterpenes (here illustrated by myrcene and β-ocimene), or by addition of water, which yields monoterpene alcohols, or monoterpenols. The route leading to monoterpenols is highlighted in green. GPP – geranyl diphosphate, LPP – linalyl diphosphate.

The most common monoterpenols in plants are linalool, nerol and geraniol, the latter two being *cis* and *trans* isomers, respectively (Figure 2). Linalool is an optically active compound,

present in the flower scent of both monocotyledonous and dicotyledonous plants (3). Both enantiomers, 3*S*- and 3*R*-linalool, occur naturally at enantiomeric excess ranging from 100% (3*S*) to 97.5% (3*R*) in different plant species (4). Irregular monoterpenols, *i.e.* built from non-head-to-tail isoprenyl condensation, such as chrysanthemol, lavandulol or artemisinol are outside the scope of this review, in which we focus on linalool, geraniol and their oxygenated derivatives.

Monoterpenols differ from other aliphatic monoterpenes in their chemical properties. The presence of an alcohol functional group makes them not only more polar and, therefore, soluble in water (5), but also more chemically reactive. Reports show monoterpenols isomerize to one another in acidic aqueous solvents (6, 7), although these harsh conditions are probably not achieved in biological systems. The alcohol functional group also makes monoterpenols more amenable to secondary transformations, such as oxidation, glycosylation (8-10), esterification (11) or methylation (12). Among these transformations, glycosylation is the most ubiquitous: many plants store monoterpenols as water-soluble glycosides. Plants presumably glycosylate monoterpenols to decrease their toxicity, facilitate their storage and transport between plant organs (13). In this review we focus on the recent advances in understanding monoterpenol oxidative metabolism, in particular mediated by cytochrome P450 oxygenases. These enzymes can greatly diversify the structures of monoterpenols and change their chemical properties. This diversity is reflected in different roles these compounds have in plants, as well as their different economical uses.

Figure 22. Plant non-cyclic monoterpene alcohols, with carbon numbering used in this paper shown for (*R*)-linalool. 8- and 9-linalool oxidized derivatives are also referred to as (*E*)-8- and (*Z*)-8-linalool derivatives, respectively.

Ecological and economic importance of monoterpenol derivatives

Iridoids are geraniol derivatives with important medicinal properties

Only few oxygenated linear geraniol derivatives have been reported in plants. Those include 8-hydroxygeraniol, foliamenthoic acid and carboxygeranic acid, conjugated to hexose, pentose, malonyl or acetyl groups. These compounds were detected upon expression of cytochromes P450 CYP76C4 and CYP76B6 in *N. benthamiana* (14), but their biological relevance and occurrence *in vivo* has not yet been demonstrated.

The best documented oxidized derivatives of geraniol are iridoids. Those are cyclic derivatives, with a fused cyclopentane and pyran ring system bearing either a hydroxyl or a glucosyloxy groups at position C-1 (e.g. 7-deoxy-loganetic and loganic acids, respectively, in Figure 5). In plants, iridoids are mostly found as conjugates (iridoid glycosides or IGs). They were first isolated in the mid-1800s from the root *of Rubia tinctorum* (15), and were named according to their structural similarity and biosynthetic origin to iridodial and iridomyrmecin found in the ants of the genus *Iridomyrmex* (16, 17). Iridoids are found in more than 57 dicot families (16) that belong to Asteridae, such as Apocynaceae, Rubiaceae, Lamiaceae, Loganiaceae, Verbenaceae, Valerianaceae, Gentianaceae or Scrophulariacea. Because of their wide distribution and diversity they can be used as chemotaxonomic markers. Iridoid producing plants have been used in folk medicine for the treatment of many diseases such as parasitoses, inflammation, diabetes and others (18, 19).

The number of plant iridoids is estimated to reach 3000 molecules according to a thorough literature-based inventory performed by Dinda and coworkers (20-23), and, due to their proven or potential pharmacological properties, the number of studies on these molecules is rising quickly. Many iridoids and iridoid-derived compounds, such as monoterpene indole alkaloids (MIAs) are used in modern medicine for their antibacterial, anti-inflammatory, anti-tumoral, chemopreventive and various others activities (24, 25). Among this variety of molecules, vindoline and vincristine, extracted from Madagascar periwinkle (*Catharanthus roseus*) are used directly or as precursor for the synthesis of drugs for treatment of lung cancer and lymphoma. Other extensively used MIAs include compounds as diverse as the antitumoral camptothecin (26), the antihypertensive ajmalicin or the classical antimalarial drug quinine (27) (Figure 3). This review focuses on the role and origin of monoterpenol iridoids in plants. Description of the medicinal properties of plant iridoids and their derivatives will not be detailed in this work, and can be found in a number of dedicated reviews (24, 25).

Figure 3. Examples of iridoids (aucubine and catalpol) and iridoid-derived alkaloids.

Iridoids accumulate in plants as a protection against various herbivores and pathogens. A toxic effect of isolated compounds against vertebrate and invertebrate predators has been reported (16, 28, 29). They have also antibacterial (30-32) and antifungal activities (30, 33-35). The toxic and antimicrobial effect of the IGs is due to activity of their aglycone moeity, which is released through enzymatic or non-enzymatic acidic hydrolysis (36). It has been shown that the antimicrobial activity strictly requires enzymatic hydrolysis to liberate the aglycone (30, 35, 37). In this regard, a single report describes a β-glucosidase from leaves of *Ligustrum obtusifolium* able to convert the secoiridoid glucoside moiety of oleuropein into a glutaraldehyde-like structure that shows strong protein denaturing, protein-crosslinking, and lysine-alkylating activities (38).

It was recently suggested that IGs could also have a role during oxidative stress, such as drought stress. Drought stress increases the accumulation of the IGs catalpol, aucubine, harpagide and harpagoside in roots of the medicinal plant *Scrophularia ningpoensis* (39). Based on the protective activities of catalpol and aucubin observed in animal studies (40), it was suggested that this increase in IGs content may help the plant cell to deal with oxidative stress (41). An increase in indole alkaloids concentrations and particularly of the antioxidant alkaloid ajmalicine in *Catharanthus roseus* plants submitted to drought stress has also been reported (42, 43).

The toxic effect of iridoids towards insects is not general and has to be examined from a co-evolutionary point of view. They can have a clear deterrent effect for non-adapted insects, but can be phagostimulants for adapted ones (44). Feeding generalist insects with iridoid

glycosides induces a reduced growth rate, an increase in larval stages duration and a diminution of survival rates (28, 29, 45). As for the deterrent effect, it is still not clear if it is due to the glycosides or to the release of the algycones via acid-hydrolysis or insect-derived β-glycosidases in the insect midgut (34). Contrary to generalists, adapted insects can feed on iridoids producing plants, with even beneficial effects. Iridoids can act as oviposition stimulants (46, 47) and feeding stimulants (48) for both adults and larvae. The iridoid resistance of specialist insects can be conferred by the ability to absorb and substract them from the gut before hydrolysis can occur (36). The uptaken iridoids can either be degraded, as observed in noctuid and geometrid larvae (49), or sequestrated as observed in many adapted Coleoptera, Homoptera, Hymenoptera or Lepidoptera, thereby conferring them obvious advantages against herbivores, parasites or pathogens (50-52). Only a few iridoids, such as aucubin and catalpol, are sequestrated, suggesting either highly specific transport mechanisms from the insects gut or a differential degradation process between compounds (36). Other insects have been shown to exploit the protective iridoid properties through *de novo* biosynthesis, as observed in some species of *Chrysomeline* (53). *De novo* synthetized iridoids can constitute a part of the defensive chemical arsenal of insects, and they can also participate to a tightly regulated sex pheromone signal, as observed in some species of aphids (54, 55).

Monoterpenol derivatives in food and beverage aroma

Development of gas chromatography in the 1950s (56) triggered interest for the analysis of volatile constituents of food, beverages and aromatic plants. Many of them included putative oxygenated linalool derivatives, but they were always found in complex mixtures so unambiguous identification was impossible. In 1963 a method for synthesis of 6,7-epoxylinalool was reported (57). This epoxide could be further transformed in acid to form furanic or pyranic linalool oxides (structures with 5- or 6-membered ring, respectively, Figure 4) in *cis* or *trans* diastereomers. In addition, each of these linalool oxides exists in two enantiomers, giving a total of 8 stereoisomers. Interestingly, pairs of enantiomers have different sensory properties, depending on whether they formed from (*S*)- or (*R*)-linalool: linalool oxides derived from (*S*)-linalool are described as sweet, floral and creamy, and those derived from (*R*)-linalool as earthy or leafy (59). Availability of synthetic reference standards enabled identification of linalool oxides in green and black tea (58). After this pioneering work on tea aroma, linalool oxides were discovered in many other foodstuffs (see Table 1). Other oxygenated linalool derivatives were later discovered in other plants, such as 6- and 7-hydroxylinalool in camphor tree (*Cinnamomum camphora*) (60). Analysis of glycosylated volatiles allowed for identification of additional monoterpenol derivatives, such as 8- and 9-hydroxylinalool in birch (*Betula alba*) and Japanese quince (*Chaenomeles japonica*) (61).

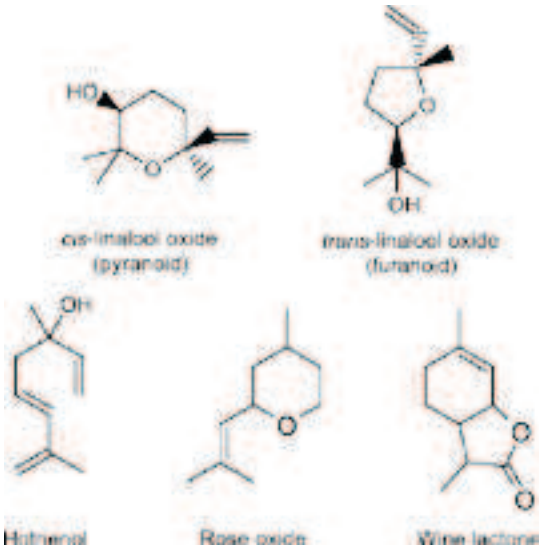

Figure 4. Selected monoterpenol derivatives that occur in food and beverages.

Studies of grape (*Vitis vinifera*) volatiles revealed a rich linalool metabolism. Some of linalool diols in grape juices are unstable in acidic solutions and can spontaneously rearrange to more stable structures by cyclisation or elimination of a water molecule (62). These rearrangements can be accelerated by heating. The products of these rearrangements are more volatile and can have a stronger odor compared to their precursors: for example, hotrienol (Figure 4), which forms from relatively odorless 7-hydroxylinalool upon acidification, has sweet tropical scent. Similarly, a monoterpene acid 8-carboxylinalool was shown to cyclize in acidic medium to form a potent sweet and coconut-like odorant wine lactone (63) (Figure 4). This reaction is thought to occur during wine ageing. Rose oxide is another monoterpenol derivative with a strong rose-like aroma, which was first discovered in rose flowers (64) and later as a constituent of fruits and other foodstuffs (Table 1). Rose oxide differs from other compounds described above in that it is not derived from linalool, but citronellol, a reduced geraniol/ nerol derivative.

Many oxygenated monoterpenol derivatives contribute to characteristic aroma of plants, among them many agricultural crops. Other oxygenated monoterpenols have a weaker odor but can rearrange in acidic medium with help of elevated temperatures. Given their ubiquitous presence, they might act as flavor precursor in foods and beverages during preparation or storage. Similarly, monoterpene glycosides can act as flavor precursors and release their volatile aglycons during food or beverage preparation, storage or fermentation.

Table 1. Linalool derivatives in fruits and other agriculturally important plant species (F: free, B: bound).

Plant	5-hydroxylinalool	6-hydroxylinalool	7-hydroxylinalool	6,7-epoxylinalool	6,7-dihydroxylinalool	8-hydroxylinalool	9-hydroxylinalool	8-carboxylinalool	Wine lactone	*trans*-linalool oxide (5)	*cis*-linalool oxide (5)	*trans* linalool oxide (6)	*cis* -linalool oxide (6)	Anhydrolinalool oxide (5)	Hotrienol	7-hydroxydihydrolinalool	Rose oxide	Ref.
Actinidia arguta (baby kiwi)			B			B	B											(65)
Camellia sinensis (tea)										F B	F B	F B	F B					(66-68)
Carica papaya (papaya)		F	F B	F	F B	F B	F B			F B	F B	F	F		B			(69, 70)
Chichorium endivia (endive)										F	F							(71)
Citrus paradisi (grapefruit)									F									(72)
Coffea arabica (coffee)										F	F			F				(73)
Cymbopogon (Citronelle)																	F	(74)
Ficus carica (fig tree)										F	F	F	F					(75)
Solanum lycopersicum (tomato)										B	B	F			B			(76, 77)
Mangifera indica (Mango)						F B				B	B							(78, 79)
Orange (Citrus sinensis)						F			F	F								(80, 81)
Passiflora (Passionfruit)		B	B			B	B											(82)
Prunus armeniaca (Apricot)		B				B				B	B	B	B			B		(83, 84)
Prunus domestica (Yellow Plum)						B												(83)
Prunus persica (Peach)						B						B	B					(83)
Ribes nigrum (Blackcurrant)										F B							F	(85)
Vitis vinifera (Grapevine)	B	F B	F B		F	B		B		F B	F B	F B	F B		F		F	(62, 86-90)
Zingiber officinale (ginger)						B												(91)

Linalool and its derivatives in plant-insect interactions

Linalool has been extensively investigated for its role in plant-insect interactions, including pollinator attraction (3, 92, 93), defense (94-96), and involvement in multi-trophic interaction by attraction of herbivory predators and parasites (96-98). However, the role of a large number of compounds deriving from linalool oxidative metabolism remains elusive. A few studies showed that pyranoid and furanoid linalool oxides from *Clarkia breweri* (3) and

Daphne mezereum (92), as well as lilac compounds in *Silene latifolia* (99-101) are important olfactory cues for pollinators. For example, the moth *Hadena bicruris* specifically recognizes lilac aldehydes emitted by *Silene latifolia* at night among the many volatiles of the scent bouquet in a nursery pollination system. The moth then lays its eggs in the flower and pollinates it at the same time (99-101). In another example, a link was established between the emission level of lilac aldehydes and attractiveness of the plants from the *Asimitellaria* lineage for different pollinators, namely short- and long-tongued fungus gnat (102). Olfactometer trials demonstrated that lilac aldehydes induced nectaring behavior of the long-tongued fungus gnats but repelled short-tongued fungus gnats. The volatile composition of several *Asimitellaria* species is adapted to their specific pollinator species of fungus gnat, illustrating the scent-mediated speciation of *Asimitellaria* to their pollinators. Moreover, pure lilac aldehydes and lilac alcohols were shown to repel thrips and hoverflies (103), suggesting linalool and its derivatives have a dual role of in attraction of beneficial and repellence of neutral or detrimental insects. *Arabidopsis thaliana* flowers do not need pollinators to reproduce, although out-crossing events have been observed in natural populations since flowers are visited by insects such as solitary bees, hoverflies and thrips (104-106). As a consequence, *A. thaliana* chemical profile may favor protective functions, such as defense against flower visitors and pollen thieves, over attractiveness.

In addition to its role in scent production and plant-insect interactions, linalool oxidative metabolism in plants might serve for linalool detoxification as was observed in insect guts (107, 108), fungus *Botrytis cinerea* (109) and soil fungi (110). Local metabolism of linalool into its furanoid and pyranoid oxides in pistils of *C. breweri* was proposed as a defense mechanism to protect pollen tube from linalool toxicity (3). Oxidation and subsequent glycosylation also increases the solubility of monoterpenols and may favor both their sequestration and transport to other organs. Their presence in phloem might also serve as a line of defense against phloem feeding insect (111).

First insights into geraniol and linalool oxidative metabolism

Battersby and coworkers first demonstrated in labelling experiments that geraniol was cyclized and then transformed to loganin on the pathways leading to MIAs (112). Since then, efforts have been devoted to understanding the steps leading to the common MIA precursor secologanin, as well as other species-specific downstream alkaloids. The Damtoft and Inouye groups paved the way in the 80s (113-117), with extensive precursor feeding experiments in which the patterns of labelled carbon scrambling were studied for many plants and pathways, thus asserting the different iridoid pathways leading from the monoterpenol geraniol (and citronellol in a few cases) to iridodial and alkaloids (reviewed in details in (118)). In particular, they demonstrated that geraniol is the precursor of most of the iridoids, and

established that deoxyloganic and epi-deoxyloganic acids are crucial intermediates in the synthesis of MIAs, each present in specific pathways. The putative pathway leading to these compounds was predicted to involve a succession of oxido-reductions, but until recently the exact sequence and most of the enzymes involved remained unknown.

Active linalool oxidative metabolism in plants was initially observed when feeding linalool to tobacco cell cultures led to the production 8-hydroxylinalool (120). In Clarkia, both linalool and linalool oxides are emitted from flowers, but linalool oxides are likely formed in other specific tissues from an enzymatic linalool oxidation pathway (3, 121). The existence of this oxidative metabolism in plants was then further supported by the detection of an increased production of both volatile and soluble oxygenated linalool derivatives upon heterologous expression of linalool synthases in different plant species including tomato, carnation and Arabidopsis (122-124).

The biosynthetic route and fate of linalool in the plant was more thoroughly investigated using labeled precursors. The first investigations, carried out with lilac flowers using both deuterated and ^{18}O labeled precursors indicated that lilac compounds derived from linalool via a plastidial (MEP) pathway and proceeded with low enantioselectivity, but with high conservation of the (*R*) or (*S*) configuration of the linalool precursor (125, 126). The pathway appeared to sequentially involve 8-hydroxylinalool, 8-oxolinalool, and lilac aldehydes converted into lilac alcohols. Interestingly, labeled lilac compounds were found associated with the plastids only when feeding was performed on intact plant tissues but not isolated plastids. Linalool feeding experiments in kiwi (*Actinidia arguta*) flowers revealed formation of lilac alcohol epoxides (128). (*S*)-linalool in kiwi flower petals is stereospecifically converted to lilac compounds (129).

Luan and coworkers investigated the diversity of metabolism of linalool in grape berries by feeding labeled linalool analogues (127). Linalool was shown to be the precursor of linalool oxides, 8-hydroxylinalool, 7-hydroxylinalool, 6-hydroxylinalool, hotrienol and nerol oxide, as well as their glycoconjugates. Hydroxylation at the position 7 is stereoselective, which strongly suggests that it results from an enzymatic and not a photooxydation reaction. They also proved that linalool oxides are preferentially formed via a 6,7-epoxylinalool intermediate, and also to a minor extent through 6,7-dihydroxylinalool, thus most likely via an enzymatic reaction. They showed in addition that the total oxygenation activity was stronger at the beginning of the ripening period and that stereospecific formation of furanoid and pyranoid linalool oxides occurred at different stages of berry maturation. Interestingly, the same group also investigated metabolism of deuterated geraniol in grape, but hydroxylated geraniol derivatives could not be detected (119). Some geraniol was, however, converted to rose oxide, cyclic ether derived from oxidative metabolism of citronellol, a

reduced derivative of geraniol. It thus appears that geraniol is mostly oxidized at the terminal positions and subsequently turned into iridodial and epi-iridodial, and not much is known about other natural single-hydroxylated derivatives. On the contrary, oxidation of linalool occurs at different positions in the carbon chain forming a wide diversity of hydroxylated compounds.

Recent advances and role of cytochromes P450 in geraniol and linalool metabolism

Geraniol metabolism in *Catharanthus roseus*

Early in the 1970s, geraniol and nerol were shown to be almost exclusively oxidized at the position 8 by a monooxygenase activity in *Catharanthus roseus* (*Vinca rosea* at that time) extract fraction to form 8-hydroxygeraniol (also sometimes referred as 10-hydroxygeraniol) (130, 131). The first characterized plant P450 showing the 2,3- or 6,7-epoxidation of geraniol and nerol *in vitro*, CYP71A1, was isolated from the avocado fruit (132), and a homolog CYP71 from catmint was found to catalyze the 8-hydroxylation of geraniol and nerol (133), although in both cases no further experiments confirmed the physiological functions of the enzymes *in planta*. Later, investigations focused on the first oxidative step in the pathway leading from geraniol to MIAs led to the characterization of CYP76B6 as a geraniol hydroxylase in *C. roseus,* and this activity was found related to the accumulation of alkaloids in Apocynaceae (134). Recently, a more thorough functional analysis of CYP76B6 catalytic properties demonstrated that this P450 catalyzed two successive regio-specific oxidations at the C8 position of geraniol to form the derived aldehyde (14). This activity is likely widespread in the plant kingdom, as similar oxidized geraniol derivatives have been reported in tobacco (a plant that does not produce geraniol) upon expression of geraniol synthases (14, 135). The strong induction of the *CYP76B6* gene expression by the hormone methyl-jasmonate, as well as its co-expression with the other MIA genes in the phloem-associated parenchyma confirmed that it was the best candidate for catalyzing the geraniol oxidation step toward iridoids and MIAs in *C. roseus* (136). Similarly, in the iridoid producing plant *Swertia mussotii*, the methyl jasmonate-induced CYP76B10 was found to catalyze the 8-hydroxylation of geraniol (137). Other members of the CYP76 family were subsequently found to catalyze oxidation of geraniol or nerol such as CYP76C4 from *A. thaliana*, catalyzing both the geraniol to 8- and 9-hydroxylations, and CYP76B1 from *Helianthus tuberosus*, catalyzing nerol hydroxylation (14). In the same study, other CYP76 members from *A. thaliana* were shown to metabolize other monoterpenols. This suggested that other members of the CYP76 family might be involved in monoterpenol metabolism.

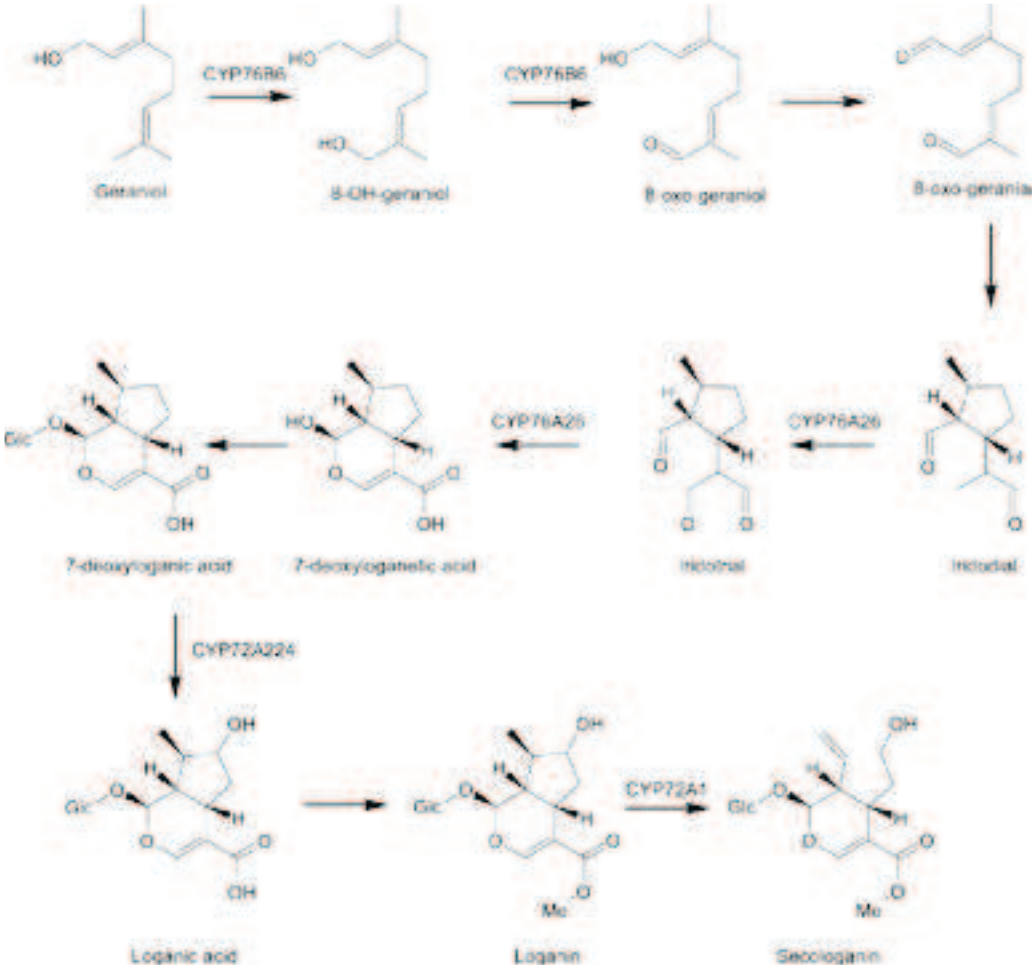

Figure 5. Cytochromes P450 involved in secologanin pathway in *Catharanthus roseus*. Glc: (β-D)-glucopyranosyl moiety.

This was confirmed when deep-sequencing and proteomic approaches led to the elucidation of the complete secologanin pathway in *C.* (Figure 5). These studies led to the characterization of CYP76A26 from *C. roseus* as a nepetalactol (iridodial) oxygenase, catalyzing three consecutive oxygenation steps and to the formation of 7-deoxyloganic acid (136, 138). Interestingly, this enzyme was also found capable to hydroxylate monoterpenols (such as nerol, citronellol), although with low efficiency. CYP76A26, however, did not hydroxylate geraniol, possibly preventing competition with CYP76B6. The same or similar strategies led the characterization of the complete sequence of enzymes forming the secologanin pathway (Figure 5). Those include an oxidoreductase converting 8-oxogeraniol to 8-oxogeranial, a recently characterized iridoid synthase belonging to the progesterone 5-beta reductase family, a glycosyltransferase, and another cytochrome P450, CYP72A224 (136, 139-141). The final steps of methylation and oxidative ring opening and formation of secologanin by loganic acid methyltransferase (LAMT) and secologanin synthase (SLS1 or CYP72A1), respectively, had been described previously (142-144). Recently, a second

functional *C. roseus* SLS2 was characterized and both isoforms were shown to further catalyze the oxidation of secologanin to secoxyloganin *in vitro* (145).

It is interesting to note that both SLS and CYP72A224 share the unusual property of metabolizing larger and very hydrophilic glycosylated substrates. CYP72A224 was confirmed to be inactive on the aglycone (136) (Figure 5), which was not tested for SLS1 and SLS2. The same sequence of two reactions, *i.e.* glycosylation then hydroxylation, was also reported for the related geniposide pathway in *Gardenia jasminoides* (146). It is thus likely that CYP72s derive from a CYP ancestor that was already able to accommodate substrates larger than monoterpenols. This hypothesis is supported by the hydroxylation of triterpenic derivatives by CYP72s in different species such as Medicago (147) or the quinidine hydroxylation by CYP72A8 from *A. thaliana* (148).

Other cytochromes P450 have been identified in biosynthetic pathways from geraniol to MIAs. Twenty years ago, CYP71D12 was characterized in *Catharanthus roseus* as a tabersonine hydroxylase catalyzing the first commited step to vindoline (149), and recently, CYP71D1v2, was shown to act just downstream in the pathway (150). More putative P450-catalyzed steps in MIA pathways remain to be characterized.

Linalool metabolism in *Arabidopsis thaliana* flowers

The first lead to linalool oxygenases emerged from *in silico* analysis of gene co-expression. Two P450 enzymes CYP76C3 and CYB71B31 were tightly co-expressed with two terpene synthases TPS10 and TPS14, ((*R*)- and (*S*)-linalool synthases, respectively) in *A. thaliana* flowers (151). Both P450s were expressed exclusively in the upper segment of the anther filaments and nectaries (and weakly in petals) upon anthesis. Both P450s could metabolize both linalool enantiomers in yeast and after transient expression in *Nicotiana benthamina*. CYP71B31 formed 1,2-epoxylinalool and a mixture of different diastereoisomers of 4- and 5-hydroxylinalool and CYP76C3 8- and 9-hydroxylinalool and a different mixture of 4- and 5-hydroxylinalool diastereoisomers (Figure 6). However, altered expression of CYP76C3 and CYP71B31 in *A. thaliana* mutants had only a minor impact on overall floral linalool oxidative metabolism, which prevented the identification of their end-chain endogenous products, as well as suggested the contribution of other enzymes in the floral linalool metabolism. The other members of the CYP76 family in *A. thaliana* were more extensively investigated. At first their enzyme activity was tested *in vitro* in parallel with CYP76B6 from *C. roseus* and CYP76B1 from *Helianthus tuberosus* (152). All these enzymes metabolized monoterpenols, but among those showing a clear expression in yeast only CYP76C1, CYP76C2, CYP76C4 and CYP76B6 metabolized linalool into 8-hydroxylinalool as major product and 8-hydroxylinalool as minor product. In addition, CYP76C2 and CYP76C4 were shown to produce 1,2-epoxylinalool (Figure 6). This suggested a widespread monoterpenol oxidation

capacity in the CYP76 family, and especially a linalool oxidase role of the CYP76C subfamily in Brassicaceae (152).

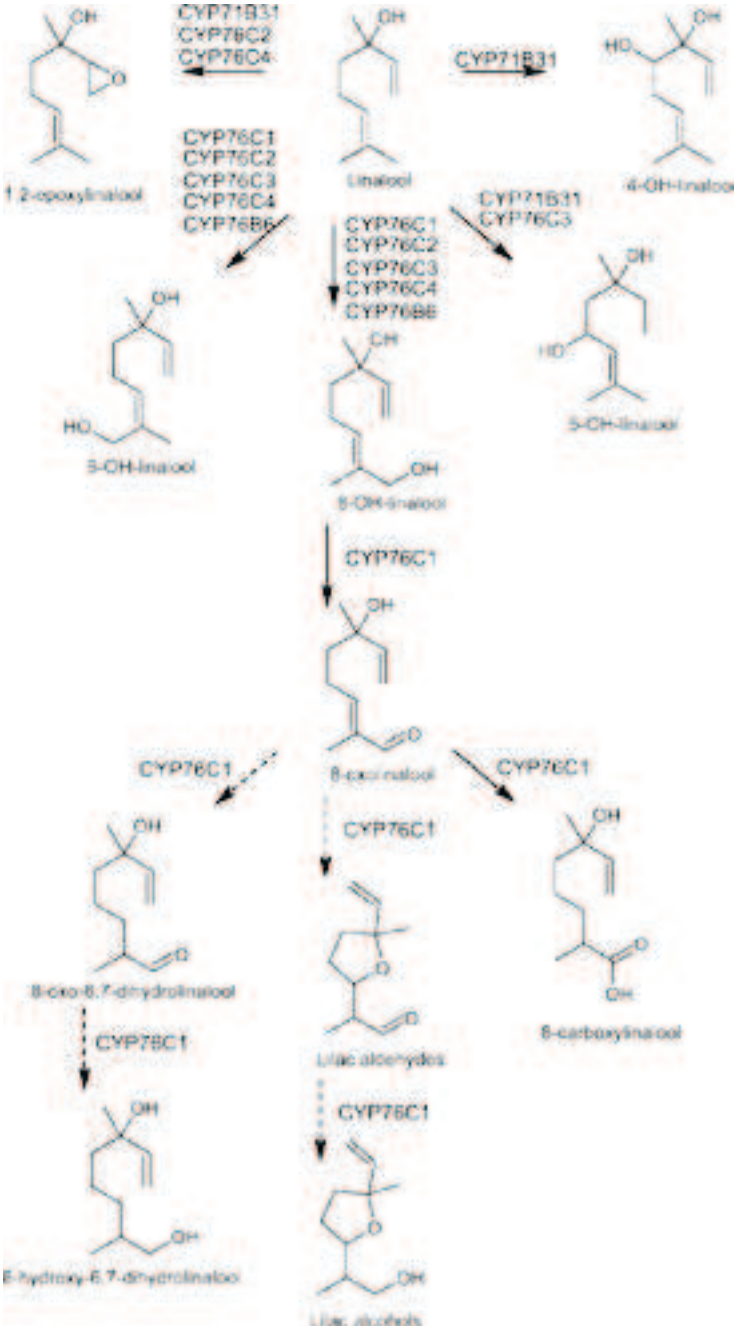

Figure 6. Linalool metabolism by *A. thaliana* cytochromes P450. Recombinant enzymes were produced either in yeast or *N. benthamiana* leaves. Dotted arrows indicate reactions which may require other enzymes.

CYP76C1 was by far the most efficient linalool-converting enzyme *in vitro*, appearing as a prime linalool oxidase candidate in *A. thaliana*. Its enzyme activity and role was thus extensively investigated *in vitro* and in the plant (103). Like *CYP71B31* and *CYP76C3*, *CYP76C1* was co-expressed with *TPS10* and *TPS14* upon flower anthesis, but $CYP76C1_{promoter}$*:GUS* transformants revealed a more widespread expression in flower organs, including anthers, stigma and petals. *In vitro* studies, transient expression in *N.*

benthamiana, as well as targeted metabolic profiling in Arabidopsis mutants altered in *CYP76C1* expression all confirmed that CYP76C1 catalyzed sequential oxidation of the terminal carbon in linalool, forming successively 8-hydroxylinalool, 8-oxolinalool, and 8-carboxylinalool (Figure 6). Furthermore, 8-oxolinalool was converted by CYP76C1 into the lilac aldehydes and lilac alcohols in the *in vitro* assays (Figure 6). All products, except for 8-oxolinalool, were detected as volatile or conjugated compounds in *A. thaliana* flowers and their amount was dependent on *CYP76C1* expression. Among the *A. thaliana* enzymes assayed in vitro, CYP76C1 was the only one to catalyze the oxidation of 8-hydroxylinalool to 8-oxolinalool (B. Boachon, R. Höfer, J.F. Ginglinger and C. Gavira unpublished data). Together with the high linalool oxidase activity of CYP76C1, this most likely explains the minor quantitative impact of CYP76C3 and CYP71B31, as well as of their other flower-expressed paralogues CYP76C2 and CYP71B38 on floral linalool metabolism (unpublished data). The specific reactions catalyzed by CYP71B31 and CYP76C3, as well as their tissue-specific expression, however suggest the existence of a complex linalool metabolism in *A. thaliana* flowers for the production of specific cues or protection compounds.

CYP76C1 mutants offered an opportunity to investigate the role of linalool derivatives in ecological interactions with the flowers of *A. thaliana* (103). The depletion of several volatile and soluble linalool derivatives rendered the *cyp76c1* inactivation mutant flowers more attractive and sensitive to several antagonist insects such as the pollen thieves thrips (*Frankliniella occidentalis*) and florivores, including a generalist herbivore *Plutella xylostella* and the Brassicaceae specialist, *Phaedon cochleariae* and *Spodoptera littoralis*. Correspondingly, the beetles *Phaedon cochleariae* were shown to prefer feeding on control cabbage leaves over leaves treated with 8-hydroxylinalool or 8-carboxylinalool. Since endogenous linalool oxides are most often stored as glycosides, they could function as toxic compounds released from wounded tissues or in the insect guts upon herbivory attack. In addition to soluble compounds, CYP76C1 produced the volatile lilac aldehydes and lilac alcohols while decreasing the emission of their precursor linalool. Since thrips and *A. thaliana* pollinators hoverflies (*Episyrphus balteus*) are attracted by linalool and repelled by both lilac aldehydes and lilac alcohols, CYP76C1 appears to set the balance between an attractive display and defense by consuming linalool and producing repellant lilac compounds. Accordingly, *CYP76C1* is a pseudogene in the genome of the obligate outcrossing relative *Arabidopsis lyrata* (152). Altogether, this suggests that CYP76C1, responsible for the synthesis of most oxygenated linalool derivatives in *A. thaliana*, may increase flower fitness through defense against floral antagonists.

It is noteworthy that additional linalool derivatives, such as 8-hydroxy-6,7-dihydrolinalool and 8-oxo-6,7-dihydrolinalool, were detected in *A. thaliana* flowers. Their concentrations

were barely altered by the silencing of CYP76C1, CYP76C3 or CYP71B31 (103), which suggests other enzymes are involved in unknown branched pathways of *A. thaliana* linalool metabolism. In addition, none of the P450 investigated was found able to catalyze the conversion of linalool into 6,7-epoxylinalool, the proposed precursor of furanoid and pyranoid linalool oxides (70).

Finally, the subcellular localization of CYP76C1, CYP76C3 and CYP71B31 was investigated in parallel with localization of TP10 and TPS14 (103, 151). In all cases the two linalool synthases were detected in vesicular structures associated with the plastids whereas the P450s appeared associated with plastid-wrapping endoplasmic reticulum sheets. This most likely explains the initial observations of Kreck and coworkers (125), who found that labeled lilac compounds associated with the plastids only when feeding was performed on intact plant tissues, but not isolated plastids.

Perspectives and open questions

Recent work started to reveal the core oxidative metabolism of monoterpenols in higher plants. Considering their widespread presence and their economic importance, they are first rank targets in pathway discovery studies. The advances on linalool and geraniol oxidative pathways occurred in parallel, revealing some similarities but also differences, in part resulting from the intrinsic properties of the starting compounds. Metabolism of both linalool and geraniol share common initial oxidation steps. The most common oxidation of both substrates occurs on the 8-carbon, and commonly involves several consecutive oxygenations via alcohol and aldehyde to acid as the final product. Both the alcohol and the acid are usually accumulated in plants as glycosides. Conjugation and branching pathways indicate that each intermediate is released from the enzyme. Polar derivatives, such as carboxylic acids and glycosides were long overlooked due to analytical procedures essentially restricted to GC. This fact points to the importance of using complementary analytical approaches to get the full appraisal of a metabolic pathway.

In all the known monoterpenol oxidative pathways the initial oxidation step is catalyzed by a cytochrome P450 from CYP76 family. These enzymes belong to different subfamilies in different plant taxa and share quite low sequence identities (<55%). The branching point of the oxidation cascade is often, but not always, the aldehyde intermediate. All the intermediates of the oxidation cascade can be glycosylated, although the aldehyde intermediate is usually not accumulating in plants. The balance between the glycotransferase activity and the efficiency of the oxidation cascade may control the flux through the oxidation steps and consequently the distribution among the intermediates. In Asteridae, the first elongation of the geraniol pathway led to emergence of the large families of iridoids and seco-

iridoids, followed by the second elongation to MIAs. This pathway is now elucidated up to the formation of the seco-iridoid and strictosidine structures. In many linalool producing plants, oxidation and cyclization leads to the production of cyclic compounds, often stored in the plant as conjugates. It is currently unknown whether cyclic linalool derivatives are incorporated in more complex structures, equivalent to MIAs.

It is currently unclear whether cytochromes P450 alone catalyze the formation of cyclic linalool derivatives, such as linalool oxides or lilac aldehydes, or if additional enzymes are needed to cyclize activated precursors generated by P450s. A human xenobiotic metabolizing P450 CYP2D6 can efficiently convert linalool into cyclic linalool oxides (153). In plants, however, oxidation/cyclisation activity of P450s has not yet been confirmed. *A. thaliana* CYP76C1 can form lilac aldehydes from 8-oxolinalool *in vitro* with low efficiency, but its overexpression in *N. benthamiana* was not sufficient to generate detectable amounts of lilac compounds. This could be either due to competing *N. benthamiana* enzymes, but may as well signal requirement for additional cyclization enzyme(s), which remain to be identified. In addition, none of the cytochromes P450 characterized so far has been able to produce linalool oxides, in spite of their widespread occurrence in plant kingdom.

A common feature of CYP76 enzymes is their low regio- and stereoselectivity. The enzymes can usually oxidize a whole subset of monoterpenols, the most striking example being CYP76B6 from *C. roseus* that catalyzes the double oxidation of all linear monoterpenols (geraniol, linalool, lavandulol, citronellol) with a very high efficiency (14). Other examples are CYP76C1 and CYP76C3, both which metabolize (*R*)- and (*S*)-linalool with no significant enantioselectivity (103, 151). The low selectivity of CYP76C1 is also illustrated by its diverse product profile, including 8-carboxylinalool, lilac aldehydes or lilac alcohols. In addition, CYP76B1, CYP76C1, CYP76C2, and CYP76C4 were all found to metabolize several phenylurea herbicides. Their herbicide-metabolizing capacity was sufficient to confer an increased herbicide tolerance to plant transformants (14, 154). This raises the question of promiscuity of these enzymes with regard to other plant metabolites and xenobiotics. The CYP76Ms of rice (*Oryza sativa*) evolved as multifunctional enzymes dedicated to the biosynthesis of labdane-related diterpenoid antifungal phytoalexins (155, 156). Structural studies are needed to understand different enzyme specificity and their apparent promiscuity.

Monoterpenol oxidative metabolism usually controls the formation of volatile or soluble oxygenated derivatives, in particular in flowers, fruits and young leaves, which require intensive protection against herbivores and other antagonists or predators. Together with glycosyl transferases, they also control the level of emission of free monoterpenols. There are now several examples where this activity sets the balance between attractiveness for pollinators and repellence of antagonists. In flowers (157) and fruits of *Carica papaya* (70,

158, 159), as well as flowers of *Clarkia breweri* (3) and *Daphne mezereum* (92) linalool and linalool oxides appeared to play a role in insect attractions. The need for maintaining a very high flexibility of this system (possibly in association with speciation or complex interactions, including tri-trophic plant-insect interactions) could be responsible for the high versatility of this pathway. An alternative hypothesis is that monoterpenol metabolism evolved as a detoxification pathway since similar pathways are reported in insects and microorganisms (3), which could provide another explanation for its high versatility. While the ecological role of linalool and its derivatives is already well established, the influence of iridoids and MIAs on plant adaptation and resistance to antagonists and microorganisms remains largely unexplored, in spite of the intense interest in pharmaceutical applications of these compounds.

The recent description of the core linalool and geraniol oxidative metabolism in *A. thaliana* and *C. roseus* paves the way to the discovery and engineering of the monoterpenol-derived pathways of plants of agronomic and economic interest, not only for the production of metabolites of interest but also for a direct plant protection effect. The reconstruction of these pathways in microorganisms for aroma, cosmetic and pharmaceutical industry is also in development. Among the most appealing candidates are fruit, wine and tea aroma, as well as the large diversity of iridoids and alkaloids with documented therapeutic applications.

Authors' contributions

Writing – first draft: Introduction, Monoterpenol derivatives in food and beverages, First insights into geraniol and linalool oxidative metabolism (TI); Geraniol derivatives and iridoids, First insights into geraniol and linalool oxidative metabolism (CP); Role of linalool and its derivatives in plant-insect interactions, Linalool metabolism (BB); Geraniol metabolism (NN), Perspectives and open questions (DWR). Writing – review and editing (NN and DWR).

References

1. Degenhardt J, Kollner TG, & Gershenzon J (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70(15-16):1621-1637.
2. Magnard JL, *et al.* (2015) Plant volatiles. Biosynthesis of monoterpene scent compounds in roses. *Science* 349(6243):81-83.
3. Raguso RA & Pichersky E (1999) New Perspectives in Pollination Biology: Floral Fragrances. A day in the life of a linalool molecule: Chemical communication in a plant-pollinator system. Part 1: Linalool biosynthesis in flowering plants. *Plant Species Biol.* 14(2):95-120.
4. Aprotosoaie AC, Hăncianu M, Costache I-I, & Miron A (2014) Linalool: a review on a key odorant molecule with valuable biological properties. *Flavour and Fragrance Journal* 29(4):193-219.
5. Weidenhamer JD, Macias FA, Fischer NH, & Williamson GB (1993) Just how insoluble are monoterpenes? *Journal of chemical ecology* 19(8):1799-1807.
6. Baxter RL, Laurie WA, & Mchale D (1978) Transformations of monoterpenoids in aqueous acids: The reactions of linalool. geraniol, nerol and their acetates in aqueous citric acid. *Tetrahedron* 34(14):2195-

2199.

7. Cori O, Chayet L, Perez LM, Bunton CA, & Hachey D (1986) Rearrangement of linalool, geraniol, nerol and their derivatives. *The Journal of Organic Chemistry* 51(8):1310-1316.

8. Sheth K, Ramstad E, & Wolinsky J (1961) The structure of loganin. *Tetrahedron Lett.* 2(12):394-397.

9. Inouye H, Arai T, Miyoshi Y, & Yaoi Y (1963) Über die struktur des monotropeins. *Tetrahedron Lett.* 4(16):1031-1038.

10. Francis M & Allcock C (1969) Geraniol β-D-glucoside; occurrence and synthesis in rose flowers. *Phytochemistry* 8(8):1339-1347.

11. Steltenkamp RJ & Casazza WT (1967) Composition of the essential oil of lavandin. *J. Agric. Food Chem.* 15(6):1063-1069.

12. Froehlich O, Duque C, & Schreier P (1989) Volatile constituents of curuba (Passiflora mollissima) fruit. *J. Agric. Food Chem.* 37(2):421-425.

13. Vasserot Y, Arnaud A, & Galzy P (1995) Monoterpenol glycosides in plants and their biotechnological transformation. *Acta biotechnologica* 15(1):77-95.

14. Höfer R, *et al.* (2013) Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early (seco)iridoid pathway. *Metabolic engineering*.

15. Schunck E (1848) Untersuchung des Krapps. *Justus Liebigs Ann. Chem.* 66(2):174-213.

16. Bowers MD (1991) Iridoid glycosides. *Academic Press, Orlando, FL*:297-325.

17. Cavill G, Ford D, & Locksley H (1956) The chemistry of ants. I. Terpenoid constituents of some Australian Iridomyrmex species. *Australian Journal of Chemistry* 9(2):288-293.

18. Chan-Bacab MJ & Peña-Rodríguez LM (2001) Plant natural products with leishmanicidal activity. *Nat. Prod. Rep.* 18(6):674-688.

19. Viljoen A, Mncwangi N, & Vermaak I (2012) Anti-Inflammatory Iridoids of Botanical Origin. *Curr. Med. Chem.* 19(14):2104-2127.

20. Dinda B, Debnath S, & Harigaya Y (2007) Naturally occurring iridoids. A review, part 1. *Chem. Pharm. Bull. (Tokyo)* 55(2):159-222.

21. Dinda B, Debnath S, & Harigaya Y (2007) Naturally occurring secoiridoids and bioactivity of naturally occurring iridoids and secoiridoids. A review, part 2. *Chem. Pharm. Bull. (Tokyo)* 55(5):689-728.

22. Dinda B, Chowdhury DR, & Mohanta BC (2009) Naturally Occurring Iridoids, Secoiridoids and Their Bioactivity. An Updated Review, Part 3. *Chem. Pharm. Bull. (Tokyo)* 57(8):765-796.

23. Dinda B, Debnath S, & Banik R (2011) Naturally Occurring Iridoids and Secoiridoids. An Updated Review, Part 4. *Chem. Pharm. Bull. (Tokyo)* 59(7):803-833.

24. Ghisalberti E (1998) Biological and pharmacological activity of naturally occurring iridoids and secoiridoids. *Phytomedicine* 5(2):147-163.

25. Dinda B & Debnath S (2013) Monoterpenes: iridoids. *Natural Products*, (Springer), pp 3009-3067.

26. Wall ME, *et al.* (1966) Plant antitumor agents. I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from camptotheca acuminata1, 2. *J. Am. Chem. Soc.* 88(16):3888-3890.

27. Block E (1999) [Book Review: Alkaloids: Biochemistry, Ecology, and Medicinal Applications Margaret F. Roberts, Michael Wink]. *Q. Rev. Biol.* 74(2):256.

28. Bowers MD & Puttick GM (1988) Response of generalist and specialist insects to qualitative allelochemical variation. *Journal of chemical ecology* 14(1):319-334.

29. Puttick GM & Bowers MD (1988) Effect of qualitative and quantitative variation in allelochemicals on a generalist insect: Iridoid glycosides and the southern armyworm. *Journal of chemical ecology* 14(1):335-351.

30. Davini E, Javarone C, Trogolo C, Aureli P, & Pasolini B (1986) The quantitative isolation and antimicrobial activity of the aglycone of aucubin. *Phytochemistry* 25(10):2420-2422.

31. Ishiguro K, Yamaki M, & Takagi S (1982) [Studies on the iridoid related compounds. I. On the antimicrobial activity of aucubigenin and certain iridoid aglycones]. *Yakugaku zasshi: Journal of the Pharmaceutical Society of Japan* 102(8):755-759.

32. Rombouts J & Links J (1956) The chemical nature of the antibacterial substance present inAucuba japonica thunbg. *Experientia* 12(2):78-80.

33. Biere A, Marak HB, & van Damme JM (2004) Plant chemical defense against herbivores and pathogens: generalized defense or trade-offs? *Oecologia* 140(3):430-441.

34. Marak HB, Biere A, & Van Damme JM (2002) Two herbivore-deterrent iridoid glycosides reduce the in-vitro growth of a specialist but not of a generalist pathogenic fungus of Plantago lanceolata L. *Chemoecology* 12(4):185-192.

35. Sluis W, Nat J, & Labadie R (1983) Thin-layer chromatographic bioassay of iridoid and secoiridoid glucosides with a fungitoxic aglucone moiety using beta-glucosidase and the fungus Penicillium expansum as a test organism. *J. Chromatogr.* 259:522-526.

36. Dobler S, Petschenka G, & Pankoke H (2011) Coping with toxic plant compounds–the insect's perspective on iridoid glycosides and cardenolides. *Phytochemistry* 72(13):1593-1604.

37. Baden C & Dobler S (2009) Potential benefits of iridoid glycoside sequestration in Longitarsus melanocephalus (Coleoptera, Chrysomelidae). *Basic and Applied Ecology* 10(1):27-33.

38. Konno K, Hirayama C, Yasui H, & Nakamura M (1999) Enzymatic activation of oleuropein: a protein crosslinker used as a chemical defense in the privet tree. *Proceedings of the National Academy of Sciences* 96(16):9159-9164.

39. Wang D, Du F, Liu H, & Liang Z (2010) Drought stress increases iridoid glycosides biosynthesis in the roots of Scrophularia ningpoensis seedlings. *J Med Plants Res* 4(24):2691-2699.

40. Jin L, Xue H-Y, Jin L-J, Li S-Y, & Xu Y-P (2008) Antioxidant and pancreas-protective effect of aucubin on rats with streptozotocin-induced diabetes. *Eur. J. Pharmacol.* 582(1):162-167.

41. Mao Y-R, Jiang L, Duan Y-L, An L-J, & Jiang B (2007) Efficacy of catalpol as protectant against oxidative stress and mitochondrial dysfunction on rotenone-induced toxicity in mice brain. *Environ. Toxicol. Pharmacol.* 23(3):314-318.

42. Jaleel CA, *et al.* (2008) Antioxidant potential and indole alkaloid profile variations with water deficits along different parts of two varieties of Catharanthus roseus. *Colloids and Surfaces B: Biointerfaces* 62(2):312-318.

43. Jaleel CA, *et al.* (2008) Water deficit stress effects on reactive oxygen metabolism in Catharanthus roseus; impacts on ajmalicine accumulation. *Colloids and Surfaces B: Biointerfaces* 62(1):105-111.

44. Dobler S (2001) Evolutionary aspects of defense by recycled plant compounds in herbivorous insects. *Basic and Applied Ecology* 2(1):15-26.

45. Bowers MD & Puttick GM (1989) Iridoid glycosides and insect feeding preferences: gypsy moths (Lymantria dispar, Lymantriidae) and buckeyes (Junonia coenia, Nymphalidae). *Ecological Entomology* 14(3):247-256.

46. Pereyra PC & Bowers MD (1988) Iridoid glycosides as oviposition stimulants for the buckeye butterfly, Junonia coenia (Nymphalidae). *Journal of chemical ecology* 14(3):917-928.

47. Prudic KL, Oliver JC, & Bowers MD (2005) Soil nutrient effects on oviposition preference, larval performance, and chemical defense of a specialist insect herbivore. *Oecologia* 143(4):578-587.

48. Bowers MD (1983) The role of iridoid glycosides in host-plant specificity of checkerspot butterflies. *Journal of chemical ecology* 9(4):475-493.

49. Boros CA, Stermitz FR, & McFarland N (1991) Processing of iridoid glycoside antirrinoside fromMaurandya antirrhiniflora (Scrophulariaceae) byMeris paradoxa (Geometridae) andLepipolys species (Noctuidae). *Journal of chemical ecology* 17(6):1123-1133.

50. Gardner DR & Stermitz FR (1988) Host plant utilization and iridoid glycoside sequestration byEuphydryas anicia (Lepidoptera: Nymphalidae). *Journal of chemical ecology* 14(12):2147-2168.

51. Rimpler H (1991) Sequestration of iridoids by insects. (Oxford: Oxford Science Publications), pp 314-330.

52. Bowers MD & Stamp NE (1997) Fate of Host-Plant Iridoid Glycosides in Lepidopteran Larvae of Nymphalidae and Arcthdae. *Journal of chemical ecology* 23(12):2955-2965.

53. Kunert M, *et al.* (2008) De novo biosynthesis versus sequestration: a network of transport systems supports in iridoid producing leaf beetle larvae both modes of defense. *Insect Biochem. Mol. Biol.* 38(10):895-904.

54. Dawson G, *et al.* (1990) Aphid semiochemicals—a review, and recent advances on the sex pheromone. *Journal of chemical ecology* 16(11):3019-3030.

55. Stewart-Jones A, *et al.* (2007) Structure, ratios and patterns of release in the sex pheromone of an aphid, Dysaphis plantaginea. *J. Exp. Biol.* 210(Pt 24):4335-4344.

56. James A & Martin A (1952) Gas-liquid partition chromatography: the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid. *Biochem. J.* 50(5):679.

57. Felix D, Melera A, Seibl J, & Kovats ES (1963) Zur Kenntnis ätherischer Öle. 2. Mitteilung. Die Struktur der sogenannten «Linalooloxide». *Helv. Chim. Acta* 46(5):1513-1536.

58. Yamanishi T, Sato H, & Ohmura A (1964) Linalool-Epoxides in the Essential Oils from Freshly Plucked Tea-Leaves and Black Tea. *Agric. Biol. Chem.* 28(9):653-655.

59. Wang D, Ando K, Morita K, Kubota K, & Kobayashi A (1994) Optical isomers of linalool and linalool oxides in tea aroma. *Biosci. Biotechnol. Biochem.* 58(11):2050-2053.

60. Takaoka D & Hiroi M (1976) Two acyclic monoterpene diols from Cinnamomum camphorha. *Phytochemistry* 15(2):330.

61. Tschesche R, Ciper F, & Breitmaier E (1977) Monoterpen-Glucoside aus den Blättern von Betula alba und den Früchten von Chaenomeles japonica. *Chem. Ber.* 110(9):3111-3117.

62. Williams PJ, Strauss CR, & Wilson B (1980) Hydroxylated linalool derivatives as precursors of volatile monoterpenes of Muscat grapes. *J. Agric. Food Chem.* 28(4):766-771.

63. Bonnländer B, Baderschneider B, Messerer M, & Winterhalter P (1998) Isolation of two novel terpenoid glucose esters from Riesling wine. *J. Agric. Food Chem.* 46(4):1474-1478.

64. Seidel C, *et al.* (1961) Zur Kenntnis des Rosenöls. 2. Mitteilung. Die Konstitution des Oxyds C10H18O aus bulgarischem Rosenöl. *Helv. Chim. Acta* 44(2):598-606.

65. Garcia CV, Quek S-Y, Stevenson RJ, & Winz RA (2011) Characterization of the bound volatile extract from baby kiwi (Actinidia arguta). *J. Agric. Food Chem.* 59(15):8358-8365.

66. Bondarovich H, *et al.* (1967) VOLATILES IN TEA Some Aspects of the Chemistry of Tea. A Contribution to the Knowledge of the Volatile Constituents. *J. Agric. Food Chem.* 15(1):36-47.

67. Moon J-H, Watanabe N, Ijima Y, Yagi A, & Sakata K (1996) cis-and trans-Linalool 3, 7-oxides and methyl salicylate glycosides and (Z)-3-hexenyl β-D-glucopyranoside as aroma precursors from tea leaves for oolong tea. *Biosci. Biotechnol. Biochem.* 60(11):1815-1819.

68. Wang D, Yoshimura T, Kubota K, & Kobayashi A (2000) Analysis of glycosidically bound aroma precursors in tea leaves. 1. Qualitative and quantitative analyses of glycosides with aglycons as aroma compounds. *J. Agric. Food Chem.* 48(11):5411-5418.

69. Flath RA & Forrey RR (1977) Volatile components of papaya (Carica papaya L., Solo variety). *J. Agric. Food Chem.* 25(1):103-109.

70. Winterhalter P, Katzenberger D, & Schreier P (1986) 6, 7-Epoxy-linalool and related oxygenated terpenoids from Carica papaya fruit. *Phytochemistry* 25(6):1347-1350.

71. Goetz-Schmidt EM & Schreier P (1986) Neutral volatiles from blended endive (Cichorium endivia, L.). *J. Agric. Food Chem.* 34(2):212-215.

72. Buettner A & Schieberle P (1999) Characterization of the most odor-active volatiles in fresh, hand-squeezed juice of grapefruit (Citrus paradisi Macfayden). *J. Agric. Food Chem.* 47(12):5189-5193.

73. Gautschi F, Winter M, Flament I, Willhalm B, & Stoll M (1967) New developments in coffee aroma research. *J. Agric. Food Chem* 15:15-23.

74. Kreis P & Mosandl A (1994) Chiral compounds of essential oils. Part XVII. Simultaneous stereoanalysis of Cymbopogon oil constituents. *Flavour and fragrance journal* 9(5):257-260.

75. Gibernau M, Buser HR, Frey JE, & Hossaert-McKey M (1997) Volatile compounds from extracts of figs of Ficus carica. *Phytochemistry* 46(2):241-244.

76. Buttery R, Seifert R, Guadagni D, & Ling L (1971) Characterization of additional volatile components of tomato. *J. Agric. Food Chem.* 19(3):524-529.

77. Buttery RG, Teranishi R, Ling LC, & Turnbaugh JG (1990) Quantitative and sensory studies on tomato paste volatiles. *J. Agric. Food Chem.* 38(1):336-340.

78. Ollé D, *et al.* (1998) Comparison of Free and Glycosidically Linked Volatile Components from Polyembryonic and Monoembryonic Mango (Mangifera i ndica L.) Cultivars. *J. Agric. Food Chem.* 46(3):1094-1100.

79. Sakho M, Chassagne D, & Crouzet J (1997) African mango glycosidically bound volatile compounds. *J. Agric. Food Chem.* 45(3):883-888.

80. Selli S, Cabaroglu T, & Canbas A (2004) Volatile flavour components of orange juice obtained from the cv. Kozan of Turkey. *Journal of Food Composition and Analysis* 17(6):789-796.

81. Hinterholzer A & Schieberle P (1998) Identification of the most odour-active volatiles in fresh, hand-extracted juice of Valencia late oranges by odour dilution techniques. *Flavour and fragrance journal* 13(1):49-55.

82. Chassagne D, Boulanger R, & Crouzet J (1999) Enzymatic hydrolysis of edible Passiflora fruit glycosides. *Food chemistry* 66(3):281-288.

83. Krammer G, Winterhalter P, Schwab M, & Schreier P (1991) Glycosidically bound aroma compounds in the fruits of Prunus species: apricot (P. armeniaca, L.), peach (P. persica, L.), yellow plum (P. domestica, L. ssp. Syriaca). *J. Agric. Food Chem.* 39(4):778-781.

84. Salles C, Jallageas JC, Fournier F, Tabet JC, & Crouzet JC (1991) Apricot glycosidically bound volatile components. *J. Agric. Food Chem.* 39(11):1979-1983.

85. Varming C, Andersen ML, & Poll L (2006) Volatile monoterpenes in black currant (Ribes nigrum L.) juice: effects of heating and enzymatic treatment by β-glucosidase. *J. Agric. Food Chem.* 54(6):2298-2302.

86. Williams P, Strauss C, & Wilson B (1981) Classification of the monoterpenoid composition of Muscat grapes. *American Journal of Enology and Viticulture* 32(3):230-235.

87. Wilson B, Strauss CR, & Williams PJ (1984) Changes in free and glycosidically bound monoterpenes in developing Muscat grapes. *J. Agric. Food Chem.* 32(4):919-924.

88. Strauss CR, Wilson B, & Williams PJ (1988) Novel monoterpene diols and diol glycosides in Vitis vinifera grapes. *J. Agric. Food Chem.* 36(3):569-573.

89. Winterhalter P, Messerer M, & Bonnländer B (1997) Isolation of the glucose ester of (E)-2, 6-dimethyl-6-hydroxyocta-2, 7-dienoic acid from Riesling wine. *Vitis* 36(1):55-56.

90. Guth H (1997) Identification of character impact odorants of different white wine varieties. *J. Agric. Food Chem.* 45(8):3022-3026.

91. Wu P, Kuo MC, & Ho CT (1990) Glycosidically bound aroma compounds in ginger (Zingiber officinale Roscoe). *J. Agric. Food Chem.* 38(7):1553-1555.

92. Borg-Karlson A-K, Unelius CR, Valterová I, & Nilsson LA (1996) Floral fragrance chemistry in the early flowering shrub Daphne mezereum. *Phytochemistry* 41(6):1477-1483.

93. Reisenman CE, Riffell JA, Bernays EA, & Hildebrand JG (2010) Antagonistic effects of floral scent in an insect–plant interaction. *Proceedings of the Royal Society of London B: Biological Sciences* 277(1692):2371-2379.

94. Junker RR, Gershenzon J, & Unsicker SB (2011) Floral odor bouquet loses its ant repellent properties after inhibition of terpene biosynthesis. *Journal of chemical ecology* 37(12):1323-1331.

95. McCallum EJ, *et al.* (2011) Increased plant volatile production affects oviposition, but not larval development, in the moth Helicoverpa armigera. *The Journal of experimental biology* 214(21):3672-3677.

96. Xiao Y, *et al.* (2012) Specific herbivore-induced volatiles defend plants and determine insect community composition in the field. *Ecology letters* 15(10):1130-1139.

97. Loughrin JH, Potter DA, & Hamilton-Kemp TR (1995) Volatile compounds induced by herbivory act as aggregation kairomones for the Japanese beetle (Popillia japonica Newman). *Journal of chemical ecology* 21(10):1457-1467.

98. Turlings T, *et al.* (1995) How caterpillar-damaged plants protect themselves by attracting parasitic wasps. *Proceedings of the National Academy of Sciences* 92(10):4169-4174.

99. Dötterl S, *et al.* (2006) Linalool and lilac aldehyde/alcohol in flower scents: Electrophysiological detection of lilac aldehyde stereoisomers by a moth. *J. Chromatogr. A* 1113(1):231-238.

100. Dötterl S, *et al.* (2006) Nursery pollination by a moth in Silene latifolia: the role of odours in eliciting antennal and behavioural responses. *New Phytologist* 169(4):707-718.

101. Dötterl S, Burkhardt D, Jürgens A, & Mosandl A (2007) Stereoisomeric pattern of lilac aldehyde in Silene latifolia, a plant involved in a nursery pollination system. *Phytochemistry* 68(4):499-504.

102. Okamoto T, Okuyama Y, Goto R, Tokoro M, & Kato M (2015) Parallel chemical switches underlying pollinator isolation in Asian Mitella. *J. Evol. Biol.* 28(3):590-600.

103. Boachon B, et al. (2015) CYP76C1 (Cytochrome P450)-Mediated Linalool Metabolism and the Formation of Volatile and Soluble Linalool Oxides in Arabidopsis Flowers: A Strategy for Defense against Floral Antagonists. *Plant Cell* 27(10):2972–90.

104. Jones M (1971) population genetics of Arabidopsis thaliana. I. The breeding system. *Heredity (Edinb.)*.

105. Snape J & Lawrence M (1971) Breeding system of Arabidopsis thaliana. *Heredity (Edinb.)*.

106. Hoffmann M, *et al.* (2003) Flower visitors in a natural population of Arabidopsis thaliana. *Plant Biol.* 5(5):491-494.

107. Yu S (1987) Microsomal oxidation of allelochemicals in generalist (Spodoptera frugiperda) and semispecialist (Anticarsia gemmatalis) insect. *Journal of chemical ecology* 13(3):423-436.

108. Southwell IA, Maddox CD, & Zalucki MP (1995) Metabolism of 1, 8-cineole in tea tree (Melaleuca alternifolia andM. linariifolia) by pyrgo beetle (Paropsisterna tigrina). *Journal of chemical ecology* 21(4):439-453.

109. Bock G, Benda I, & Schreier P (1986) Metabolism of linalool by Botrytis cinerea. *ACS Symposium series-American Chemical Society (USA)*.

110. Demyttenaere JC & Willemen HM (1998) Biotransformation of linalool to furanoid and pyranoid linalool oxides by Aspergillus niger. *Phytochemistry* 47(6):1029-1036.

111. Gowan E, Lewis BA, & Turgeon R (1995) Phloem transport of antirrhinoside, an iridoid glycoside, inAsarina scandens (Scrophulariaceae). *Journal of chemical ecology* 21(11):1781-1788.

112. Battersby A, Brown R, Knight J, Martin J, & Plunkett A (1966) Biosynthesis of the indole alkaloids from a monoterpene. *Chemical Communications (London)* (12):346-347.

113. Damtoft S, Jensen SR & Nielsen BJ (1983) The biosynthesis of iridoid glucosides from 8-epi-deoxyloganic acid. *Biochem. Soc. Trans.* 11(5):593-594.

114. Damtoft S, Rosendal S, & Nielsen BJ (1981) 13 C and 1 H NMR spectroscopy as a tool in the configurational analysis of iridoid glucosides. *Phytochemistry* 20(12):2717-2732.

115. Uesato S, Matsuda S, Iida A, Inouye H, & Zenk MH (1984) Intermediacy of 10-Hydroxygeraniol, 10-Hydroxynerol and Iridodial in the Biosynthesis of Ajmaline and Vomilenine in Rauwolfia-Serpentina Suspension-Cultures. *Chem. Pharm. Bull. (Tokyo)* 32(9):3764-3767.

116. Uesato S, Matsuda S, & Inouye H (1984) Mechanism for Iridane Skeleton Formation from Acyclic Monoterpenes in the Biosynthesis of Secologanin and Vindoline in Catharanthus-Roseus and Lonicera-Morrowii. *Chem. Pharm. Bull. (Tokyo)* 32(4):1671-1674.

117. Inouye H & Uesato S (1986) Biosynthesis of iridoids and secoiridoids. *Fortschritte der Chemie organischer Naturstoffe/Progress in the Chemistry of Organic Natural Products*, (Springer), pp 169-236.

118. Jensen S (1991) Plant iridoids, their biosynthesis and distribution in angiosperms. *Ecological chemistry and biochemistry of plant terpenoids. Clarendon Press, Oxford* 133(145):399.

119. Luan F, Mosandl A, Münch A, & Wüst M (2005) Metabolism of geraniol in grape berry mesocarp of Vitis vinifera L. cv. Scheurebe: demonstration of stereoselective reduction, E/Z-isomerization, oxidation and glycosylation. *Phytochemistry* 66(3):295-303.

120. Hirata T, Aoki T, Hirano Y, Ito T, & Suga T (1981) The biotransformation of foreign substrates by tissue cultures. I. The hydroxylation of linalool and its related compounds with the suspension cells of Nicotiana tabacum. *Bull. Chem. Soc. Jpn.* 54(11):3527-3529.

121. Pichersky E, Raguso RA, Lewinsohn E, & Croteau R (1994) Floral scent production in Clarkia (Onagraceae)(I. Localization and developmental modulation of monoterpene emission and linalool synthase activity). *Plant Physiol.* 106(4):1533-1540.

122. Lewinsohn E, *et al.* (2001) Enhanced levels of the aroma and flavor compound S-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. *Plant Physiol.* 127(3):1256-1265.

123. Lavy M, *et al.* (2002) Linalool and linalool oxide production in transgenic carnation flowers expressing the Clarkia breweri linalool synthase gene. *Molecular Breeding* 9(2):103-111.

124. Aharoni A, *et al.* (2003) Terpenoid metabolism in wild-type and transgenic Arabidopsis plants. *The Plant Cell* 15(12):2866-2884.

125. Kreck M, Püschel S, Wüst M, & Mosandl A (2003) Biogenetic studies in Syringa vulgaris L.: synthesis and bioconversion of deuterium-labeled precursors into lilac aldehydes and lilac alcohols. *J. Agric. Food Chem.* 51(2):463-469.

126. Burkhardt D & Mosandl A (2003) Biogenetic studies in Syringa vulgaris L.: bioconversion of (18)O(2H)-labeled precursors into lilac aldehydes and lilac alcohols. *J. Agric. Food Chem.* 51(25):7391-7395.

127. Luan F, Mosandl A, Degenhardt A, Gubesch M, & 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(1):353-364.

128. Matich AJ, Bunn BJ, Hunt MB, & Rowan DD (2006) Lilac alcohol epoxide: A linalool derivative in Actinidia arguta flowers. *Phytochemistry* 67(8):759-763.

129. Matich A, Comeskey D, Bunn B, Hunt M, & Rowan D (2011) Biosynthesis and enantioselectivity in the production of the lilac compounds in Actinidia arguta flowers. *Phytochemistry* 72(7):579-586.

130. Meehan TD & Coscia CJ (1973) Hydroxylation of geraniol and nerol by a monooxygenase from Vinca rosea. *Biochem. Biophys. Res. Commun.* 53(4):1043-1048.

131. Madyastha K, Meehan TD, & Coscia CJ (1976) Characterization of a cytochrome P-450 dependent monoterpene hydroxylase from the higher plant Vinca rosea. *Biochemistry* 15(5):1097-1102.

132. Hallahan DL, *et al.* (1992) Interactions of avocado (Persea americana) cytochrome P-450 with monoterpenoids. *Plant Physiol.* 98(4):1290-1297.

133. Hallahan DL, *et al.* (1994) Cytochrome P-450-catalysed monoterpenoid oxidation in catmint (Nepeta racemosa) and avocado (Persea americana); evidence for related enzymes with different activities. *Biochimica et Biophysica Acta (BBA)-General Subjects* 1201(1):94-100.

134. Collu G, *et al.* (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* 508(2):215-220.

135. Dong L, *et al.* (2013) Characterization of two geraniol synthases from Valeriana officinalis and Lippia dulcis: similar activity but difference in subcellular localization. *Metabolic Engineering*.

136. Miettinen K, *et al.* (2014) The seco-iridoid pathway from Catharanthus roseus. *Nature communications* 5:3606.

137. Wang JF, *et al.* (2010) Cloning and Functional Analysis of Geraniol 10-Hydroxylase, a Cytochrome P450 from Swertia mussotii Franch. *Bioscience Biotechnology and Biochemistry* 74(8):1583-1590.

138. Salim V, Wiens B, Masada-Atsumi S, Yu F, & De Luca V (2014) 7-deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in Catharanthus roseus iridoid biosynthesis. *Phytochemistry* 101:23-31.

139. Asada K, *et al.* (2013) A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell* 25(10):4123-4134.

140. Salim V, Yu F, Altarejos J, & De Luca V (2013) Virus-induced gene silencing identifies Catharanthus roseus 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J.* 76(5):754-765.

141. Geu-Flores F, *et al.* (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* 492(7427):138-142.

142. Madyastha KM, Guarnaccia R, Baxter C, & Coscia CJ (1973) S-Adenosyl-L-methionine: loganic acid methyltransferase. A carboxyl-alkylating enzyme from Vinca rosea. *J. Biol. Chem.* 248(7):2497-2501.

143. Irmler S, *et al.* (2000) Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450CYP72A1 as secologanin synthase. *Plant J.* 24(6):797-804.

144. Murata J, Roepke J, Gordon H, & De Luca V (2008) The leaf epidermome of Catharanthus roseus reveals its biochemical specialization. *Plant Cell* 20(3):524-542.

145. de Bernonville TD, *et al.* (2015) Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced de novo assembly of a Catharanthus roseus transcriptome. *BMC Genomics* 16(1):619.

146. Nagatoshi M, Terasaka K, Nagatsu A, & Mizukami H (2011) Iridoid-specific glucosyltransferase from Gardenia jasminoides. *J. Biol. Chem.* 286(37):32866-32874.

147. Biazzi E, *et al.* (2015) CYP72A67 Catalyzes a Key Oxidative Step in Medicago truncatula Hemolytic Saponin Biosynthesis. *Molecular plant* 8(10):1493-1506.

148. Olry A, Schneider-Belhaddad F, Heintz D, & Werck-Reichhart D (2007) A medium-throughput screening assay to determine catalytic activities of oxygen-consuming enzymes: a new tool for functional characterization of cytochrome P450 and other oxygenases. *Plant J.* 51(2):331-340.

149. St-Pierre B & De Luca V (1995) A cytochrome P-450 monooxygenase catalyzes the first step in the conversion of tabersonine to vindoline in Catharanthus roseus. *Plant Physiol.* 109(1):131-139.

150. Qu Y, *et al.* (2015) Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 112(19):6224-6229.

151. Ginglinger JF, *et al.* (2013) Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in Arabidopsis flowers. *Plant Cell* 25(11):4640-4657.

152. Hofer R, *et al.* (2014) Dual function of the cytochrome P450 CYP76 family from Arabidopsis thaliana in

the metabolism of monoterpenols and phenylurea herbicides. *Plant Physiol.* 166(3):1149-1161.

153. Meesters R, Duisken M, & Hollender J (2007) Study on the cytochrome P450-mediated oxidative metabolism of the terpene alcohol linalool: indication of biological epoxidation. *Xenobiotica* 37(6):604-617.

154. Didierjean L, *et al.* (2002) Engineering herbicide metabolism in tobacco and Arabidopsis with CYP76B1, a cytochrome P450 enzyme from Jerusalem artichoke. *Plant Physiol.* 130(1):179-189.

155. Wang Q, *et al.* (2012) Characterization of CYP76M5–8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *J. Biol. Chem.* 287(9):6159-6168.

156. Yisheng W, Qiang W, Matthew LH, & Reuben JP (2013) Picking sides: distinct roles for CYP76M6 and CYP76M8 in rice oryzalexin biosynthesis. *Biochem. J.* 454(2):209-216.

157. Knudsen JT & Tollsten L (1993) Trends in floral scent chemistry in pollination syndromes: floral scent composition in moth-pollinated taxa. *Botanical Journal of the Linnean Society* 113(3):263-284.

158. Schreier P & Winterhalter P (1986) Precursors of Papaya (Carica papaya, L.) Fruit Volatiles. 317:85-98.

159. Flath RA, Light DM, Jang EB, Mon TR, & John JO (1990) Headspace examination of volatile emissions from ripening papaya (Carica papaya L., Solo Variety). *J. Agric. Food Chem.* 38(4):1060-1063.

Chapter 5

Annotation, classification, genomic organization and expression of the *Vitis vinifera* CYPome

Tina Ilc[1], Gautier Arista[2], Raquel Tavares[3], Nicolas Navrot[1], Frédéric Choulet[4], Etienne Paux[4], Marc Fischer[2], Philippe Hugueney[2], Danièle Werck-Reichhart[1], Camille Rustenholz[2]

[1] Institute of Plant Molecular Biology, Centre National de la Recherche Scientifique, Université de Strasbourg, Strasbourg, France

[2] Laboratoire Métabolisme Secondaire de la Vigne, Institut National de la Recherche Agronomique, Université de Strasbourg, Colmar, France

[3] Laboratoire de Biométrie et Biologie Evolutive, Centre National de la Recherche Scientifique, Université de Lyon 1, Lyon, France

[4] Laboratoire Structure et Evolution du Génome du Blé, Institut National de la Recherche Agronomique, Université Blaise Pascal, Clermont-Ferrand, France

Abstract

Cytochromes P450 are enzymes that control a wide range of functions in plants, from hormonal signaling, biosynthesis of structural polymers, to defense or communication with other organisms. The manual annotation of cytochrome P450 genes in the genome of *Vitis vinifera* PN40024 revealed 579 P450 sequences, including 279 complete genes. Most of the P450 sequences in grapevine genome are organized in physical clusters, resulting from tandem or segmental duplications. Although most of these clusters are small, some of the P450 families, such as CYP76 and CYP82, underwent multiple duplications and formed large clusters of homologous sequences. Analysis of gene expression revealed highly specific expression patterns, which are often shared within the genes in large physical clusters. Some of these genes are induced upon biotic stress, which points to their role in plant defense, whereas others are specifically activated during grape berry ripening and might be responsible for the production of berry-specific metabolites, such as aroma compounds. Our comprehensive gene annotation and expression analysis provide groundwork for further functional characterization of this important gene family in grapevine.

Background

Grapevine (*Vitis vinifera L.*) is one of the oldest (1) and economically the most important (2) fruit crops in the world. The majority of grapes produced worldwide are used in winemaking. Modern cultivated grapevine has been shaped by thousands of years of selection for traits such as berry size, sugar content or skin color (3), but today viticulture is facing new challenges. In addition to pathogen pressure, it has to deal with climate change (4–6), and shift of consumer preference towards higher quality wines with a lower environmental impact (7, 8). Traditional breeding is extremely difficult to apply in grapevine because of its long lifecycle, reduced fitness of progeny and complexity of quality traits (9). Sequencing of the grapevine genome in 2007 (10) and advances in the 'omics' techniques (11) set the stage for more efficient breeding solutions. The next crucial step towards improved grapevine varieties is the identification of genes underlying important traits, such as interactions with pathogens, fruit development and quality.

Many developmental as well as ecological functions in plants are controlled by cytochrome P450 oxygenases (12, 13). These enzymes catalyze regio- and stereospecific insertion of an oxygen atom into small, hydrophobic substrates that range from terpenoids and fatty acids to amino acids and their derivatives, such as phenolic compounds. In the model plant *Arabidopsis thaliana* they control processes as diverse as plant growth and branching (14, 15), flower (16) and fruit development (17), formation of lignin and surface biopolymers (18, 19), emission of volatiles (20) or plant-pathogen and plant-insect interactions (21, 22). This makes cytochromes P450 attractive targets for crop improvement.

Cytochromes P450 in plants evolved into many distinct families, which are usually defined as genes with 40% or higher protein sequence identity. Within one P450 family the biochemical function is often conserved across the plant kingdom. For example, enzymes from the CYP97 family are involved in carotenoid hydroxylation, CYP79s in the *N*-hydroxylation of amino acid to aldoximes, CYP75s in the hydroxylation of flavonoids, and CYP704s in fatty acid hydroxylation to form the precursors to structural polymers sporopollenin and cutin (23). Members of other families, however, have divergent functions: some members of CYP72 family are involved in iridoid biosynthesis, whereas others oxidize triterpene substrates (23). These differences stem from different evolutionary pressures on genes with different functions. Families with essential functions, such as hormone metabolism or synthesis of biopolymers, are usually maintained at low copy number and high purifying selection, whereas families with adaptive functions expanded or "bloomed" in certain taxa (24). A well-documented example is a bloom of CYP76M subfamily in rice (*Oryza sativa*), which consists of 11 genes and 2 pseudogenes. At least 4 members of this subfamily are involved in the biosynthesis of diterpenoid antifungal compounds (25, 26). They are clustered close together

in the genome, which is another common feature of recently duplicated P450s and probably results from sequential tandem duplications (24). Interestingly, CYP76 members from other plants, for example *Arabidopsis thaliana* or *Catharanthus roseus*, have a different biochemical function, namely oxidation of monoterpenols or their iridoid derivatives (27, 28). Recently expanded P450 families might therefore have interesting ecological functions, but those are more difficult to predict compared to functions of conserved P450 families. In addition, function of many P450 families is still unknown or poorly understood.

Previous annotation of P450s has highlighted some potentially interesting gene families in the highly heterozygous *V. vinifera* cv. Pinot Noir genome (29–31). In this work we performed the first complete manual annotation of P450s in the nearly homozygous *V. vinifera* reference genome PN40024 (10). We discuss the structural organization of the genes with particular focus on gene clusters. We evaluate phylogenetic relationships between those genes to identify recently expanded gene families likely linked to adaptive traits or domestication. Finally, we investigate spatio-temporal gene expression patterns, with particular focus on berry development and pathogen response to identify P450s with potential roles in these important physiological processes. This work will support further functional characterization of cytochrome P450 genes in grapevine.

Results

Gene annotation, classification and phylogeny

A similarity search of the *V. vinifera* PN40024 genome with known P450 sequences revealed 579 putative P450 sequences. We manually curated the sequences obtained with a gene prediction algorithm, and validated the annotation with grapevine ESTs and RNAseq reads (see Material and methods). We distributed them into four categories: genes, partial genes, putative pseudogenes and pseudogenes. This led to the identification of 279 full-length genes, which is fewer than 315 genes reported for the heterozygous Pinot Noir genome on the Cytochrome P450 homepage (http://drnelson.uthsc.edu/CytochromeP450.html), and suggests that some sequences previously annotated as different genes are probably allelic variants. The number of cytochromes P450 in grapevine is comparable to their number in other plants (e.g. 273 in *Arabidopsis thaliana* and *Solanum lycopersicum*, 309 in *Oryza sativa*). 20 sequences were annotated as partial genes, lacking a segment of the sequence due to gaps in the genome assembly. 11 putative pseudogenes only contain one nonsense mutation or frame shift, which could originate from sequencing errors or be genuine and still be functional genes in some varieties. Finally, the 269 pseudogenes are fragments, either containing multiple stop codons or frameshift mutations, or sequences not aligning to the whole length of homologous P450 genes.

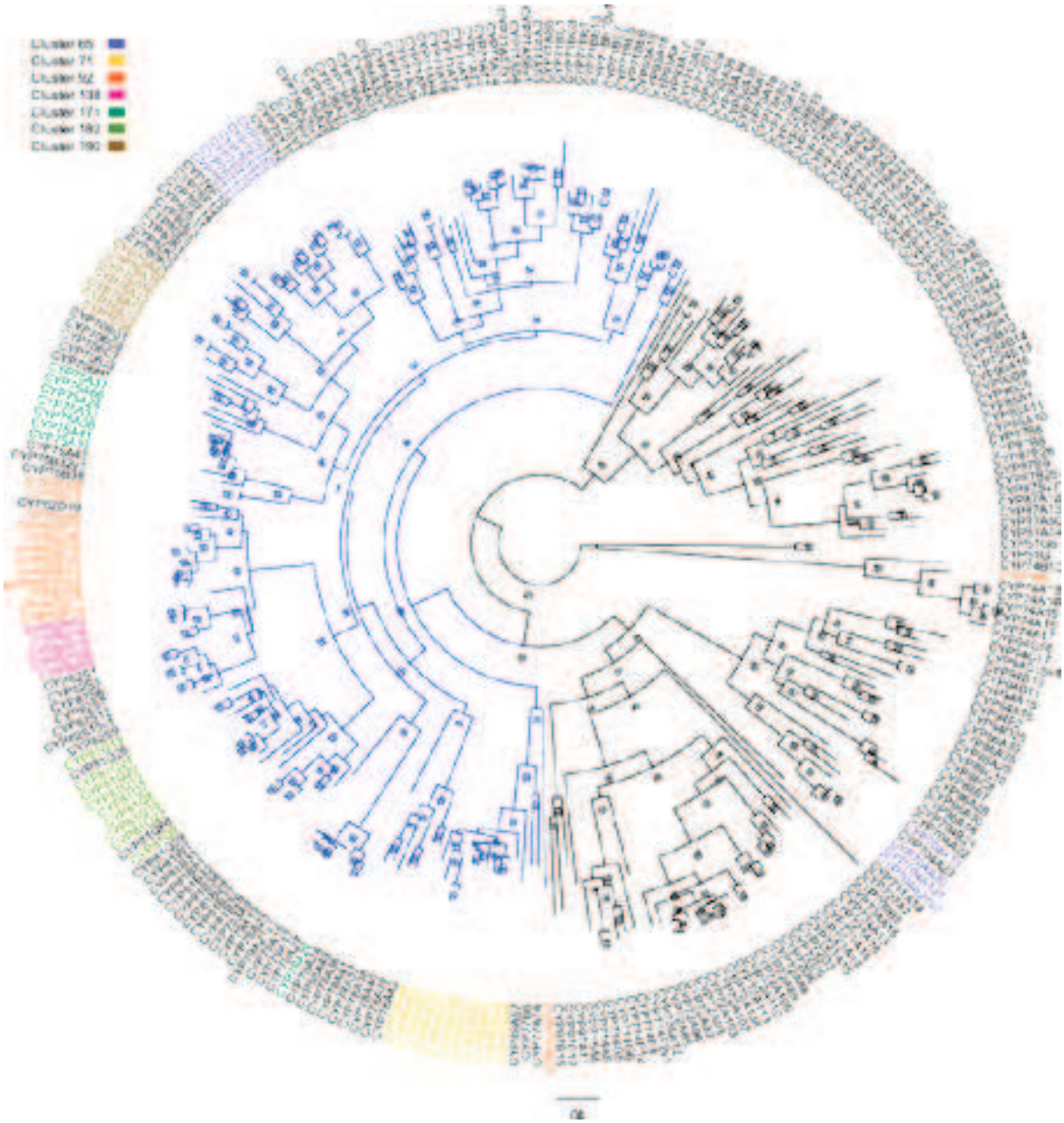

Figure 23. Molecular phylogenetic analysis of grapevine cytochrome P450. The alignment of full length cytochrome P450 protein sequences was used to generate a maximum likelihood tree. The dark blue clade is the clan 71, which often contains genes involved in specialized metabolism. The highlighted genes belong to the seven largest physical clusters.

The cytochrome P450 families usually have conserved or similar functions across species. Grapevine P450s can be assigned to 48 families based on sequence identities A phylogenetic analysis of either the full-length sequences or of a subset of conserved P450 sites confirmed this classification for most of the families. One exception is the CYP90B gene, which is clustered with CYP720 and CYP724 as previously observed (32). Other exceptions are the families CYP76 and CYP80, which form a monophyletic group (Figure 1, see Material and methods). We thus investigated the phylogeny of these two families in the broader context of selected angiosperm species (Figure S1). CYP80 clearly groups with CYP76 sequences, but

the phylogenetical relations of the clades CYP80, CYP76A/G and the rest of CYP76 sequences (labeled core CYP76) remain uncertain. Within the CYP76A/G clade, a eudicot duplication gave rise to the two subfamilies CYP76A and CYP76G. Within the large "core CYP76" clade the uncertain position of both the monocot and *Amborella trichopoda* CYP76s could be due to a problem of long-branch attraction. A specific core eudicot duplication gave rise to CYP76F/B/X on one side and CYP76T/C/E on other side. These tree topologies were obtained both with the full-length alignment and the partial alignment of conserved sites. Although species-specific "blooms" appeared in the whole CYP76/80 family, they are particularly abundant in the "core CYP76" clade. Different subfamilies "expanded" in different species.

Comparison of P450 family sizes between species (Figure S2) allowed us to identify families that potentially expanded in grapevine and might have a role in the production of species-specific specialized metabolites: an expansion of the CYP75 family, involved in anthocyanin biosynthesis, is already well documented (33), whereas the function of CYP82, the largest P450 family in grapevine with 25 members, is currently unknown in this species. Other families that are larger in grapevine than in most other species are: CYP76, CYP79, CYP80, CYP81, CYP87, CYP89 and CYP716.

Structural organization of the P450s in the PN40024 genome

The 579 cytochrome P450 sequences are distributed on all the 19 chromosomes. Some chromosomes, namely 18, 19 and 6, carry a high number of P450s, whereas others, for example chromosome 5, carry very few (Figure S3). 24 P450 sequences (7 genes, 6 partial genes, 11 pseudogenes) are located on the "Unknown" chromosome which is composed of scaffolds that could not be anchored on any of the 19 chromosomes. Since the genome is not completely homozygous (estimated homozygosity is 93% (10)), the "Unknown" chromosome may also contain eventual allelic variants of genes that are placed on the 19 chromosomes.

We further investigated the distribution of cytochrome P450 sequences in clusters or groups in close physical proximity (separated by less than 200 kb and 8 non-P450 genes (34, 35)). Our results show that P450 sequences are organized in clusters and not randomly distributed in the genome (bootstrap test, p-value < 0.0001). A large majority of cytochrome P450 sequences (452 or 78%) are part of one of the 85 clusters and only 22% (127 P450 sequences) are isolated in the grape genome. The largest number of clusters (40%) are only composed of 2 P450 sequences, whereas the largest cluster counts 35 P450 sequences. On average, there are 5 P450s per cluster and the median is 3 P450s per cluster (Figure S4). The clusters are not enriched neither in complete genes nor pseudogenes, compared to isolated annotations (data not shown). Some chromosomes, such as 16 and 18, are enriched in clustered P450s,

whereas others, such as chromosomes 4 and 11, are enriched in isolated P450 (Figure 2, Figure S3).

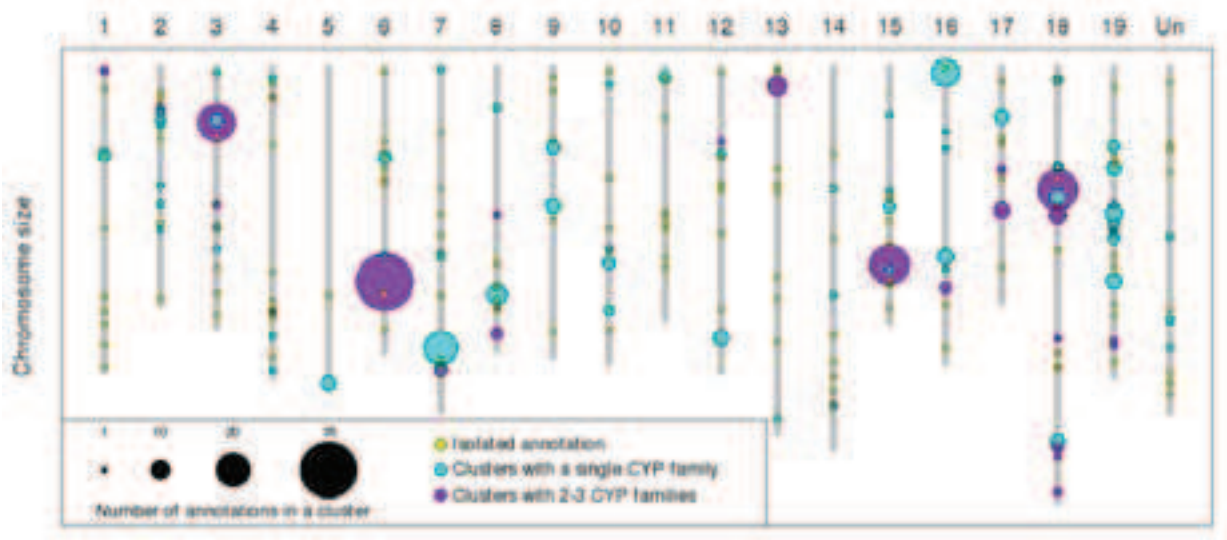

Figure 24. Physical map of cytochrome P450 sequences on the 19 *V. vinifera* chromosomes. Yellow circles represent isolated annotations, light blue circles represent physical clusters composed of members of only one P450 family and the purple circles represent physical clusters composed of members of 2–3 P450 families. The circle size is proportional to the number of sequences in the cluster. The numbers 1–19 are chromosome numbers and "Un" is "Unknown chromosome" which contains sequences with unknown chromosome location.

Cytochrome P450 families group genes with higher sequence similarity (≥40% protein sequence identity) and often a similar function. A majority of physical clusters are composed of members of only one P450 family (63 clusters, 74%) and the remaining clusters are composed of up to 3 P450 families. The 4 largest clusters are composed of several P450 families, whereas the clusters with single P450 families are smaller (Figure 2, Figure S4). Most of the largest P450 families (CYP82, CYP71, CYP81, CYP76, CYP72 and others) are organized in clusters (Table S1).

Clustering by P450 family already indicates that more similar P450 sequences cluster in closer physical proximity. But many P450 families are dispersed among several clusters. We thus wished to explore whether the closest paralogs belong to the same or different clusters (Figure 3). The majority of clustered P450 genes (86%) have their closest paralog (the best BLAST hit) in the same cluster. The second and third closest paralogs (second and third best BLAST hit) are in the same cluster for 58% and 49% of the clustered P450 genes. The sequence similarities within the same cluster are thus higher than between clusters.

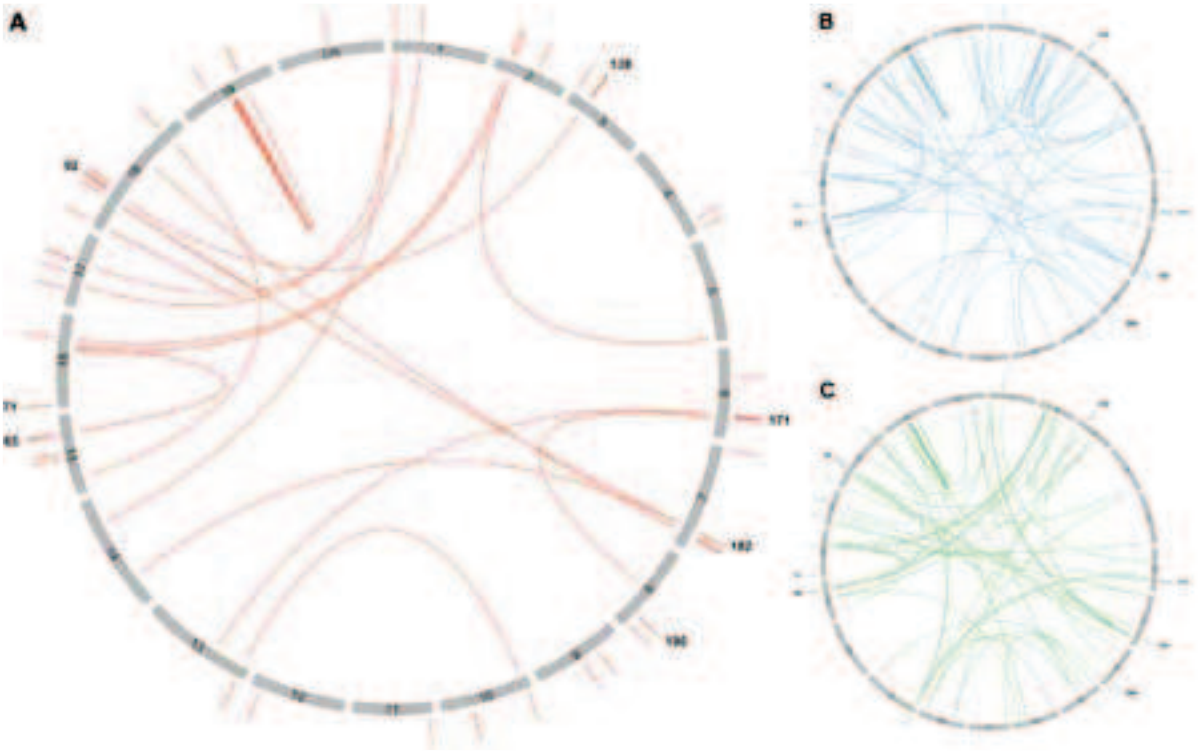

Figure 25. Similarity of the P450 genes between and within clusters. For each circle, the grey bars correspond to the 19 grape chromosomes and the "Unknown chromosome". The lines connect complete P450 genes according to their similarity. The lines outside the circles show the similarity between genes of the same cluster, whereas the lines in the circle connect similar genes of different clusters. Only P450 genes that from clusters composed of at least two complete genes are illustrated here. The seven largest clusters are labeled with numbers corresponding to Table 1. The lines are connecting the genes corresponding to the best BLAST hit (A), second best hit (B) or third best blast hit (C).

Large P450 clusters in *V. vinifera* genome formed *via* different mechanisms

To investigate the mechanisms underlying the formation of large physical clusters of cytochrome P450 genes, we further analyzed the sequence similarity within clusters, taking into account not only the coding P450 sequences, but also the surrounding non-coding-sequences. This allowed us to infer the mechanism of cluster formation. We focused on the 7 largest physical clusters, which comprise of 11 to 35 P450 sequences (Table 1). Together, these 7 clusters contain 23% of all P450 genes, and a similar fraction of total P450 sequences. Most of the sequences in these clusters are part of "clan 71", which is a large clade of plant cytochromes P450 often involved in the biosynthesis of specialized metabolites (Figure 1).

Analysis of similarity blocks within these clusters showed they differ remarkably in their structures (Figure S5). One of the largest physical clusters, cluster 65, is characterized by low similarities, both among the P450 sequences and surrounding non-coding regions. The similarity blocks of two other large physical clusters, **71** and **171**, are restricted to P450 sequences and do not extend to the intergenic regions. Single gene duplications were thus probably the main mechanism of formation of these two clusters. The similarity blocks of physical clusters **138** and **182** extend to the non-coding regions around the cytochromes

P450 annotations. This suggests the duplication events leading to formation of these clusters happened relatively recently. High similarity between the non-coding regions, which include the promoter regions, should result in similar expression profiles. Cluster **138** has the highest fraction (73%) of pseudogenes of all the seven large clusters. In physical clusters **92** and **190**, the similarity blocks extend over even longer regions that include 3–4 cytochrome P450 sequences and their intergenic regions (Figure 4). In addition, the type of annotation (gene or pseudogene) was also maintained in the same order between duplicated blocks. This suggests these two clusters formed through very recent segmental duplications.

Table 2. Description of the seven largest physical P450 clusters in the *V. vinifera* genome. Label – sequential number of each cluster in the genome; Chr – chromosome number; Location – chromosome coordinates; Total seq. – number of P450 sequences in each cluster, including complete and partial genes, putative pseudogenes and pseudogenes with their family distribution; Complete genes – number of complete P450 genes in the cluster; Expressed sequences – number of expressed P450 sequences in the cluster, co-expression – expression pattern of the cluster ("-" signifies low co-expression within the cluster); Organization – description of structural organization and mechanism of formation of each cluster.

Label	Chr	Location	Total seq.	Expressed seq.	Complete genes	Co-expression	Organization
65	15	15572751.. 15909327	20 CYP76 4 CYP704	20	10	Flowers	Low similarity among members
71	16	401789.. 596606	16 CYP89	14	11	All leaves	Single gene duplications
92	18	9625486.. 9912876	22 CYP82 1 CYP74 1 CYP704	21	14	Leaves and ripe berries	Duplicated blocks with co-expression; some single gene duplications
138	3	4387722.. 4512089	22 CYP82	18	5	Constitutive with weak upregulation in young berries	Small duplicated blocks, a few are co-expressed, single gene duplications
171	6	16790972.. 17446396	21 CYP75 14 CYP79	29	8	-	Single gene duplications
182	7	22260680.. 22372250	20 CYP81	17	9	-	Small duplicated blocks, a few are co-expressed, single gene duplications
190	8	18038159.. 18121816	11 CYP76	10	7	Flowers	Duplicated blocks with co-expression; some single gene duplication

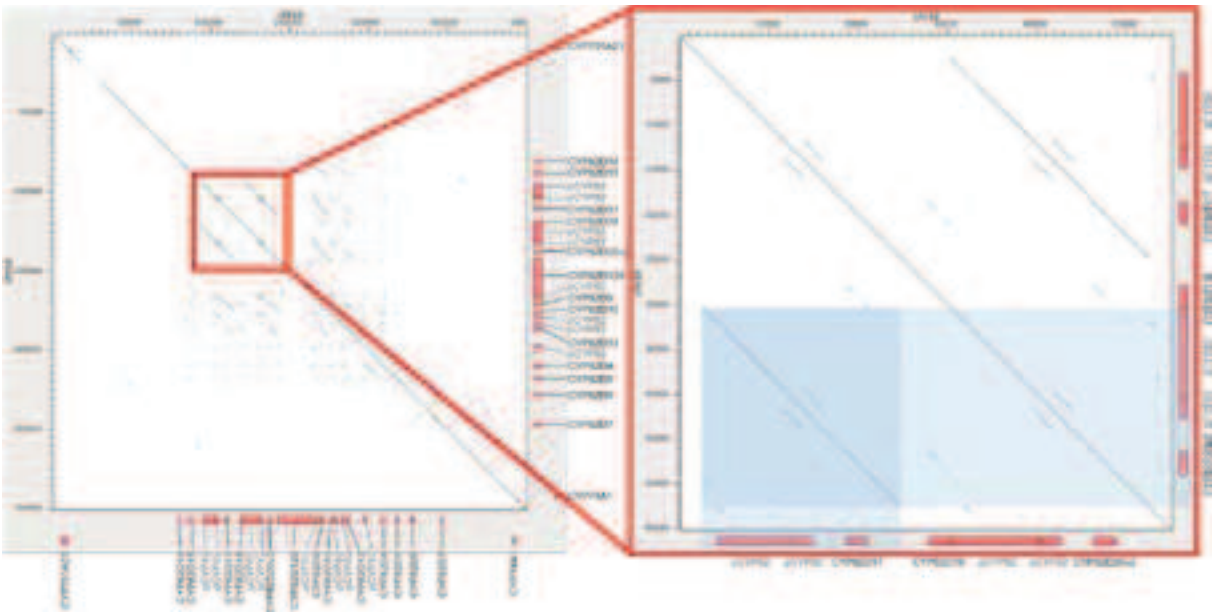

Figure 26. Dot matrix of segmental duplications in the physical cluster 92. Physical cluster 92 is located on chromosome 18 and comprises 22 CYP82 sequences, one CYP74 sequence and one CYP704 sequence. The dots and the black lines represent the sequence similarities in cluster **92** compared to itself. The red rectangles on the sides of the graph represent cytochrome P450 sequences. Complete genes are labeled with their name and pseudogenes are labeled with "p" and the P450 family. A) The similarities for the whole cluster **92**. B) A zoom of the red squared region which contains two 20-kbp-sequence blocks with very high similarity. Analysis of gene expression showed that CYP82D17 and CYP82D20v2 are co-expressed (expression cluster **O**, expression in leaves), and so are the first and the third pseudogene in the enlarged segment (expression cluster **J**, expression in ripe berries).

Expression profiles of grapevine P450s

To identify P450 genes with potential roles in pathogen resistance or biosynthesis of berry metabolites we analyzed the expression of the 579 P450 sequences. Pseudogenes were included in the analysis of expression to account for sequences that might be functional in other varieties, as well as for recently pseudogenized sequences that may still be expressed to some extent. We used 73 RNA-Seq datasets (Table S2) which describe gene expression in different tissues (flowers, berries, leaves) and different stages of berry development, and pathogen infection. We grouped the 73 experiments in 5 categories: flowers, green berries, ripe berries, leaves (control) and leaves under biotic stress. The latter category includes leaves infected with the powdery mildew pathogen *Erysiphe necator* and the downy mildew pathogen *Plasmopara viticola*.

To enable a meaningful comparison of gene expression between different experiments we calculated fragments per kilobase of transcript per million mapped reads (FPKM) for each P450 sequence in the 5 categories of experiments. We grouped the expression levels into 4 classes (no expression, low, average or high expression). The majority of P450 sequences (494 or 85%) were expressed in at least one experiment. Of the remaining 85 non-expressed P450 sequences, only 4 were complete genes. Expression of complete P450 genes (mean FPKM = 11, median FPKM = 0.6) was higher compared to pseudogenes and putative

pseudogenes (mean FPKM = 1.6, median FPKM = 0) or partial genes (mean FPKM = 1.7, median = 0.1). In each of the 5 categories, an average of 12% of the genes were not expressed, 54% had low, 20% average and 14% high expression. Interestingly, in leaves exposed to biotic stress, the fraction of non-expressed genes drops from 28% to only 9%, whereas the fraction of the highly expressed genes increases from 15 to 21%. This indicates a major shift in expression caused by biotic stress. Mean expression per category for all P450 sequencing is available in the **Appendix I**.

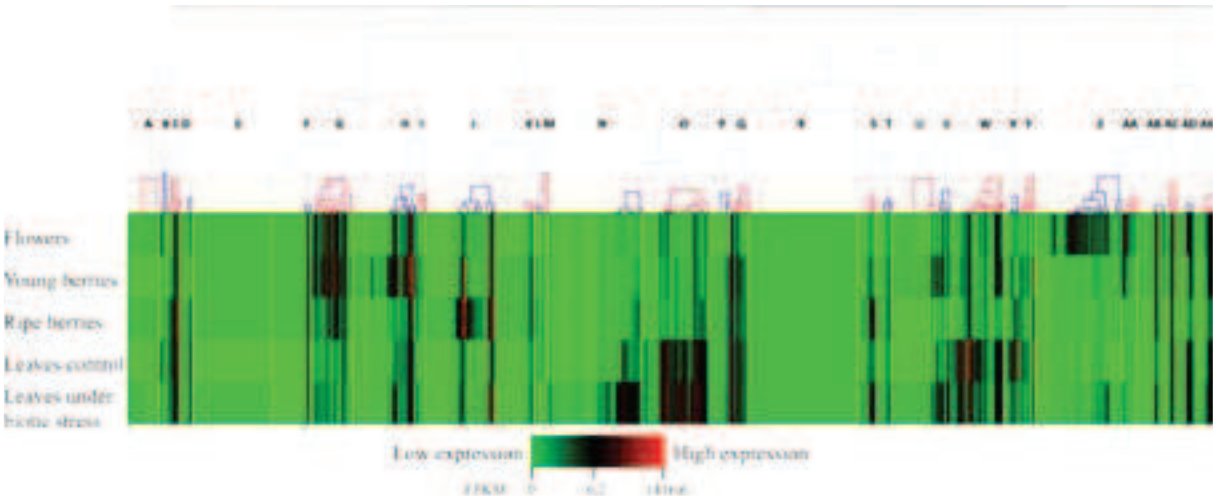

Figure 27. Heat map and dendrogram of the P450 sequences, clustered according to their expression profile. The expression levels were averaged over the experiments classified in one of the five experimental categories: flowers, young berries, ripe berries, control leaves and leaves under biotic stress. The dendrogram on the top of the figure shows 31 co-expression clusters, which include 494 expressed cytochrome P450 sequences.

We analyzed the expression patterns by clustering the expression profiles of the 494 expressed cytochrome P450 sequences. A Pearsons' correlation coefficient cut-off of 0.795 resulted in 31 expression clusters, shown in Figure 5. Five expression clusters (**I**, **J**, **N**, **X** and **Z**) were up-regulated in one of the five experimental categories. These five clusters contained 127 P450s, showing that a quarter of the expressed P450s were specifically up-regulated rather than constitutively expressed in the investigated conditions. The largest expression cluster was **R**, which consisted of 50 sequences with low basal level of expression in all investigated tissues and conditions. These may be sequences with constitutive expression, or with higher expression in conditions or organs not included in this study. Other large expression clusters were cluster **N** with 44 sequences specifically expressed in leaves undergoing biotic stress, cluster **Z** with 40 sequences specifically expressed in flowers and cluster **E** with 37 sequences expressed to a low level in reproductive organs (berries and flowers) as well as leaves under biotic stress.

The analysis of gene expression highlighted some of the P450s with potential role in biotic stress response. From the 106 sequences specifically expressed in leaves (clusters **N**, **O**, **W**

and **X**), 44 were upregulated and 7 were downregulated upon biotic stress, confirming an expression shift upon pathogen infection. The P450 family with the highest number of upregulated genes was CYP82 (7 sequences, 4 complete genes), followed by CYP81 and CYP71 with 6 sequences each (1 and 3 complete genes, respectively). CYP87B25 is the most induced gene upon biotic stress and two other CYP87 genes were among the 7 most upregulated sequences (FPKM>20).

Another major shift in the P450 expression pattern occurs during the grape berry ripening. From 81 sequences specifically expressed in berries (clusters **F**, **I**, **J**, **K** and **P**), only 4 were upregulated in green berries (cluster **I**), compared to 32 in ripe berries (cluster **J**). Interestingly, the most represented families cluster **J** were also CYP82 with 5 sequences (1 complete gene) and CYP71 with 5 sequences (3 complete genes). This cluster in addition featured the P450 gene with the highest expression in all experiments: CYP78A41 with 1417 FPKM.

We more thoroughly investigated co-expression of cytochrome P450 sequences within the 7 largest physical clusters (Table 1). In the previous section we identified 4 large clusters with high similarity, not only among the coding, but also the non-coding regions. These non-coding regions presumably include promoter sequences, so the genes in these clusters are expected to have the same expression pattern. Indeed, the 20 kbp duplicated block within cluster **92** (Figure 4b) retained the same expression profile after the segmental duplication. The duplicated segment consists of three P450 sequences: two pseudogenes and one gene. The first pseudogene in both blocks retained a very low level of expression in ripe berries, the second pseudogene in both blocks was not expressed at any experimental conditions, whereas the two complete genes (CYP82D17 and CYP82D20v2) were co-expressed in leaves and slightly induced upon pathogen infection. 7 out of 24 sequences in this cluster shared this same expression pattern (expression clusters **N** and **O**), whereas 4 other P450 sequences in the same cluster were up-regulated in ripe berries (cluster **J**). Interestingly, cluster **138** is also composed of CYP82 sequences, but these sequences were preferentially expressed in young berries. Another large physical cluster with coordinated expression pattern was cluster **71**, which comprises 11 CYP89 genes and 5 pseudogenes. Four of these genes were specifically expressed in leaves (cluster **O**). Three CYP76 sequences in the physical cluster **190**, on the other hand, were co-expressed in flowers.

Discussion

We produced a reliable and validated manual annotation of cytochromes P450 in the genome of the nearly homozygous grapevine (*V. vinifera)* accession PN40024 (10). Cytochrome P450 superfamily in *Vitis vinifera* contains both very similar and very divergent genes (sequence

identity ranges from 10% to almost 100%), and often form clusters in very close physical proximity, which makes it challenging for automated annotation algorithms. Manual curation is therefore necessary to produce a reliable annotation, suitable for demanding downstream applications such as phylogenetic or gene expression analysis. Grapevine P450s have been previously manually annotated (Cytochrome P450 homepage, http://drnelson.uthsc.edu/vitis.htm) in the highly heterozygous genome of Pinot noir cultivar (29). Our annotation represents an improvement over the existing dataset for several reasons. The assembly of PN40024 genome is of better quality compared to the Pinot noir genome: it contains fewer gaps and a higher fraction of anchored contigs. The homozygosity of the genome not only enabled a better quality of the assembly, but also assured that most of the annotated sequences are individual loci and not allelic variants. This can partially explain a lower number of cytochrome P450 genes in our annotation—279—compared to the 315 genes reported on the Cytochrome P450 homepage. Additionally, the annotation on the Cytochrome P450 homepage classifies the sequences in only two categories, genes and pseudogenes, whereas we employed a more stringent classification into genes, partial genes, putative pseudogenes and pseudogenes. Lastly, we report the exact genomic coordinates of the P450 sequences, which facilitate comparison to annotations of other genes, and provide insights into structural organization of the grapevine CYPome.

Several gene families involved in the biosynthesis of specialized metabolites, such as terpene synthase genes (TPS) and stilbene synthase genes (STS), have expanded in grapevine genome compared to other species (10, 36, 37). Although the total number of cytochrome P450 genes in grapevine is comparable to other species, individual P450 families experienced similar expansions. These expended families, similarly to TPS and STS families, form large physical clusters of more than 10 homologous sequences. One of such families is CYP75, which together with CYP79 family members forms the largest physical cluster of 35 P450 sequences on chromosome 6. Expansion of CYP75 genes in grapevine was previously documented, but the presence of another P450 family, CYP79, in the same cluster was not reported (33). Clustered genes with low or no homology sometimes participate in the same biosynthetic pathway (38, 39), but this is unlikely in the case of CYP75 and CYP79, since both families have well established roles in different biosynthetic pathways: CYP79 genes code for amino acid *N*-hydroxylases (23), whereas CYP75A genes code for flavonoid 3',5'-hydroxylases (40), crucial enzymes in the biosynthesis of blue anthocyanins in the grape skin (33, 41). We can, however, not exclude the recruitment of some of these genes in other pathways. Interestingly, the sequencing of the genome of the grapevine cultivar Tannat, characterized by its very deep color, revealed an even higher number of CYP75 genes compared to the PN40024 accession (42). Copy number of genes in a cluster can therefore vary between cultivars and influence

varietal characteristics. Other expanded P450 families in the grapevine genome that form large clusters are CYP82, CYP76, CYP81 and CYP89.

Analysis of gene expression across several tissues and conditions provides a first hint to the putative P450 functions in grapevine. Pathogen infection causes a major shift in the P450 expression, inducing members from families CYP71, CYP81, CYP82 and CYP87. Their homologs in other species have been shown to participate in biosynthesis of highly specialized defense compounds (Table S1). Interestingly, CYP736A25v1, which was shown to be upregulated upon infection with the Pierce disease pathogen *Xylella fastidiosa* (43), is also upregulated upon infection with powdery mildew and downy mildew pathogens. Other genes from the CYP736 family are not affected by biotic stress. Another large shift in expression occurs in developing grape berries. The most upregulated P450 families in the ripe-berry expression cluster are CYP82 and CYP71 (the two largest P450 families in grapevine). These P450s are likely to participate in the biosynthesis of defense compounds or compounds important for the organoleptic properties of wine (aroma, colour, taste, mouthfeel). The most up-regulated P450 gene in ripe berries and the P450 gene with the overall highest expressions is CYP78A41. A member of the same P450 family in tomato (*S. lycopersicum*) was selected during domestication to increase fruit size (44). High expression of CYP78A41 in grape berries points to a similar event in grapevine domestication.

The phylogenetic and structural data suggest that some P450 families underwent multiple tandem or segmental duplications, which resulted in large physical clusters of homologous sequences. Most of these P450 families are involved in biosynthesis of highly specialized metabolites in other plant species. These genes are often expressed in specific conditions and tissues, such as leaves upon pathogen infection. Our work thus lays the ground for discovery of interesting novel P450 functions in grapevine.

Material and methods

Gene annotation

We annotated the cytochromes P450 using the 12x version of the assembly of the *Vitis vinifera* cv PN40024 genome (10, 45). Four publically available datasets of cytochromes P450 were used to perform similarity searches in the PN40024 genome. 947 protein sequences of grape P450s were downloaded from the NCBI Protein database (http://www.ncbi.nlm.nih.gov/protein, Feb 2014). Three datasets were downloaded from David Nelson's website (http://drnelson.uthsc.edu/CytochromeP450.html, Feb 2014), which stores manually curated annotations of cytochromes P450 for many species: 702 P450 protein sequences of *Vitis vinifera* cv Pinot Noir clone ENTAV115 (28, http://drnelson.uthsc.edu/vitis.htm); 416 P450 protein sequences of *Vitis vinifera* cv

PN40024 from the 8x assembly version of the genome (10, http://drnelson.uthsc.edu/Vitis.additionalP450s.htm); and 288 P450 protein sequences of *Arabidopsis thaliana* (35, http://drnelson.uthsc.edu/Arabidopsis.Blast.file.html). The four datasets were masked for repeat sequences using the online tool "Repeat Masking" from Censor (http://www.girinst.org/censor/index.php).

The four masked datasets were used to perform four independent TBLASTN analyses (47) against the PN40024 12x sequence with an e-value cutoff of 1e-3. The TBLASTN outputs were parsed using a homemade script. The hits from the three grape datasets were kept if they were at least 50 amino acids long with at least 70% sequence identity. The hits from the *Arabidopsis* dataset were kept if they were at least 50 amino acids long with an identity percentage of at least 50%. The software Exonerate (version 2.2.0, build October 2008, 37) was used to predict gene structures using the protein2genome parameter and the same cutoff of sequence identity as above. A homemade script was used to reformat the output files from exonerate into files in the gff format. These gff files were imported to the Artemis genome browser (49) to perform the manual curation of the structures suggested by Exonerate. The parsed hits identified through TBLASTN were used to improve or to complete the Exonerate annotations. Every annotation starting with a start codon, ending with a stop codon and with correct exon-intron borders (GT-AG or sometimes GC-AG) was considered as a complete "gene". Every annotation showing the previously described gene structure but with a single point mutation creating a frameshift, a premature stop codon or a wrong exon-intron border was considered as a "putative pseudogene" also marked "pseudogene?" because it may result from a mistake in the genome assembly. Every annotation interrupted by a gap in the genomic sequence or including one was considered as a "partial" annotation. All the other annotations with wrong gene structure but showing a significant similarity level with a cytochrome P450 from one of the four datasets were annotated as "pseudogenes". The genome annotation V1 stored in Grape Genome Database hosted at CRIBI (39; http://genomes.cribi.unipd.it/DATA/GFF/V1.phase.gff3) and a set of expertized and functional grape cytochromes P450 were used to guide the manual curation.

To validate the gene structure, two transcript datasets were used. First, the *Vitis vinifera* unigene set build #15 from the NCBI database was downloaded (ftp://ftp.ncbi.nih.gov/repository/UniGene/Vitis_vinifera/Vvi.seq.uniq.gz). The 32,193 unigenes were mapped on the PN40024 12x sequence using GMAP version 2013-11-27 (51) using the default parameters except for the format parameter which was set to "gff3_match_cdna". The second transcript dataset was locally assembled using six RNA-Seq experiments ((52), SRR519450, SRR519456, SRR520380 and SRR520385; (53), all four samples; (54), SRR493740- SRR493746; (42), SRR866544, SRR866570, SRR866571 and

SRR866576; (55), SRR522472, SRR522477 and SRR522478; and eight unpublished RNA-Seq datasets acquired by INRA. The software Tophat2 v2.0.11 (56) was used to map the RNA-Seq reads against the PN40024 12x sequence using the following parameters: -p 5 -N 5 --read-edit-dist 5. The software Cufflinks v2.2.1 (57) was used to assemble the transcripts from all the RNA-Seq experiments. First the cufflinks command was used with the -p 5 parameter and then the cuffmerge command with the -p 15 parameter and using the fasta file of the PN40024 12x sequence for the -s parameter. This assembly led to 32,219 transcripts and to a gtf file showing their mapped location in the PN40024 12x sequence. The two transcript datasets were formatted in gff format compatible with the Artemis Browser so that the predicted gene structures of the cytochromes P450 could be compared with the transcripts and edited if needed.

The command maskFastaFromBed v2.19.1 from the bedtools package (58) was used to mask the regions of the PN40024 12x sequence where we annotated cytochrome P450 exons after having reformatted the gff file of the annotations into a bed file. We performed TBLASTN analyses of the four grape cytochrome P450 datasets against the masked PN40024 12x sequence and parsing analyses using the same parameters and cutoffs than previously described. This step allowed to identify the region of the grape genome for which a cytochrome P450 similarity was missed during the manual curation.

To validate the set of complete genes of cytochromes P450 that we annotated, a BLAST against non-redundant sequence database (NR) was performed and only the genes for which the best hit was a cytochrome P450 were kept. For the pseudogenes, a BLASTX was performed against the set of complete P450 genes that we annotated and we kept only the ones that aligned over at least 30% of the query length with the percentage identity of 50%.

The cytochrome P450 annotations were transferred to the improved version of the PN40024 12x assembly when it was released (PN40024 12X.2; https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences) using a homemade script. The presence of physical clusters of cytochrome P450s in the grape genome was tested based on the following definition of a cluster. Two consecutive P450 annotations are part of a cluster if they are separated by 200kb and 8 non-P450 genes at the most (34, 35). The two annotations also have to be located on the same scaffold which guaranties a precise estimation of the intergenic distances. A bootstrap test was performed to check whether the cytochromes P450 were more clustered than what is randomly expected. A homemade script was developed with R version 3.0.2 (59). Ten thousand sampling without replacement of 579 (number of P450 annotations) or 279 features (number of complete P450 genes) were performed on the genome annotation V1 stored in Grape Genome Database hosted at CRIBI counting 29,971 features. The percentage of features organized in clusters

was computed using the same protocol as for cytochromes P450. The p-value was calculated by counting each time a percentage equal of greater than the percentage of P450 in clusters divided by 10000 (number of iterations).

Sequence similarity within and between clusters (Figure 3) was analyzed by performing a BLASTP search of translated complete P450 genes against themselves. Only the genes that aligned over at least 70% of the query length with the percentage identity of 40% were kept. The Circos software (60) was used to draw the figure. Clusters that contained less than two complete genes were excluded from this analysis (i.e. clusters that contained partial genes, pseudogenes and putative pseudogenes with less than 2 complete genes).

The dotter software version 4.23 (61) was used to draw the sequence similarity graphs of the cluster 190 with its fasta sequence and annotations in a gff format as an input.

Sequence classification

Cytochrome P450 genes, partial genes and putative pseudogenes were aligned to the P450 sequences from the heterozygous Pinot Noir genome, retrieved from the cytochrome P450 homepage (http://drnelson.uthsc.edu/CytochromeP450.html). In the case of protein sequence identity above 95%, the original name was kept. New sequences were assigned a family based on the best hit among already named grapevine P450s. 22 sequences were given a new CYP name, and genes previously annotated as members of CYP81V subfamily were re-classified to CYP81Q subfamily.

Phylogeny

Sequences from non-Vitis species were retrieved from the cytochrome P450 homepage (http://drnelson.uthsc.edu/CytochromeP450.html). Pseudogenes and incomplete genes were excluded from the analysis. 279 *Vitis vinifera* CYP (Figure 1) and 191 CYP76, 80 and 706 protein sequences from *Aquilegia caerulea, Nelumbo nucifera, Mimulus guttatus, Solanum lycopersicum, Amborella trichopoda, Oryza sativa, Brachypodium distachyon, Arabidopsis thaliana, Medicago trunculata, Populus trichocarpa* and *Vitis vinifera* (Figure S1) were aligned with MUSCLE (62) implemented in Seaview (63, 64). Conserved sites were selected in the alignment using Gblocks (65) using the less stringent option parameters. Maximum likelihood phylogenies were obtained from the full-length alignments and from the subset of more conserved sites alignments (all Vitis CYP: 166 sites and 11 species CYP alignment: 278 sites) using RAxML (v 8.2.4) (66) via the CIPRES Science Gateway (67) and PhyML (implemented in Seaview v 4.5.4) (68). Bootstrap values are shown on the nodes of the Vitis all CYP phylogeny. Nodes with bootstrap values below 60 were manually suppressed from the 11 species CYP phylogeny and are shown as trifurcations (unsolved topologies). The trees

were visualized and colored using Figtree (http://tree.bio.ed.ac.uk/software/figtree). The species cladogram in (Figure S1) was inferred from the APGIII system (69).

Gene expression

We retrieved raw grape RNA-Seq data from NCBI SRA public database (http://www.ncbi.nlm.nih.gov/sra). 73 sequence files generated in the framework of 7 different experiments (42, 52, 53, 55, 70, 71) and eight unpublished RNA-Seq datasets acquired by INRA were used. The data were formatted in the fastq format using the fastq-dump command from the SRA Toolkit package version 2.3.4 (http://www.ncbi.nlm.nih.gov/books/NBK158900).

Alignments of these reads against the PN40024 12x sequence were then performed using GSNAP version 2013-11-27 (72) with the following parameters: -B 4, -N 1, -n 3, --nofails and the quality protocol according to the experiment. These files were parsed to keep the best, unique and paired (if paired-end reads) alignments using a homemade script.

The number of fragments aligned on each annotation from the genome annotation V1 stored in Grape Genome Database hosted at CRIBI and the cytochromes P450 was counted using the command htseq-count from the HTSeq framework version 0.6.0 (73) with the following parameters: -m intersection-nonempty and -s no. Using a homemade script, FPKMs (Fragments Per Kilo base of exon per Million fragments mapped) were calculated for every annotation.

Using all non-zero FPKM values, the 33th and 66th quantiles were calculated to assign the expression values to one of the four levels of expression chosen: no, low, average and high expression. The experiments were grouped into five categories regarding the conditions under which the samples were obtained. These categories are: flowers, young berries, ripe berries, leaves (control) and leaves under biotic stress. An average expression per category was then calculated for each gene and assigned to one of the four levels of expression regarding its value: no expression if the average was zero, low expression between zero and the 33% quantile, average expression between 33% and 66% quantile and high expression for averages higher than the 66% quantile.

The average expression values for each P450 annotation were used to perform a clustering analysis using HCE version 3.5 (74) with a complete linkage method and a Pearson's correlation as distance measure. The cut-off to define the clusters was set at a Pearson's correlation coefficient of 0.795.

Acknowledgements

We thank David R Nelson for assigning names to newly discovered sequences and updating the names for previously known sequences; and Adrian Arellano Davin for editing the CYP76 phylogenetic tree.

Accession Numbers

Cytochrome P450 annotation will be submitted to public databases.

Authors' contributions

TI contributed to manual curation and phylogenetic analysis, performed an expert check of the manual curation and gene classification, and wrote the manuscript. GA contributed to manual curation, performed the analysis of gene expression and wrote the manuscript. RT performed the phylogenetic analysis. NN contributed to manual curation and wrote the manuscript. FC and EP participated in the study design. MF contributed to manual curation. PH and DWR acquired funding, contributed to study design and supervision and edited the manuscript. CR contributed to study design, performed the analysis of structural organization, coordinated the work and wrote the manuscript.

References

1. Zohary D, Spiegel-Roy P (1972) Beginnings of Fruit Growing in the Old World. *Science (80-)* 187:319–327.
2. Food and Agriculture Organisation fo the United Nations. Food and Agricultural commodities production / Commodities by region (2015) Available at: http://faostat3.fao.org/browse/rankings/commodities_by_regions/E.
3. Myles S, et al. (2011) Genetic structure and domestication history of the grape. *Proc Natl Acad Sci U S A* 108(9):3530–3535.
4. Duchêne E, Schneider C (2005) Grapevine and climatic changes: a glance at the situation in Alsace. *Agron Sustain Dev* 25(1):93–99.
5. Duchêne E, Huard F, Dumas V, Schneider C, Merdinoglu D (2010) The challenge of adapting grapevine varieties to climate change. *Clim Res* 41(3):193–204.
6. Hannah L, et al. (2013) Climate change, wine, and conservation. *Proc Natl Acad Sci U S A* 110(17):6907–12.
7. Bisson LF, Waterhouse AL, Ebeler SE, Walker MA, Lapsley JT (2002) The present and future of the international wine industry. *Nature* 418(6898):696–699.
8. Borneman AR, Schmidt SA, Pretorius IS (2013) At the cutting-edge of grape and wine biotechnology. *Trends Genet* 29(4):263–271.
9. Gray DJ, Li ZT, Dhekney SA (2014) Precision breeding of grapevine (Vitis vinifera L.) for improved traits. *Plant Sci* 228:3–10.
10. Jaillon O, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161):463–7.
11. Langridge P, Fleury D (2011) Making the most of "omics" for crop breeding. *Trends Biotechnol* 29(1):33–40.
12. Nelson D, Werck-Reichhart D (2011) A P450-centric view of plant evolution. *Plant J* 66(1):194–211.
13. Bak S, et al. (2011) Cytochromes P450. *Arabidopsis Book* 9(9):e0144.
14. Helliwell CA, et al. (1998) Cloning of the Arabidopsis ent-kaurene oxidase gene GA3. *Proc Natl Acad Sci*

95(15):9019–9024.

15. Booker J, et al. (2005) MAX1 Encodes a Cytochrome P450 Family Member that Acts Downstream of MAX3/4 to Produce a Carotenoid-Derived Branch-Inhibiting Hormone. *Dev Cell* 8(3):443–449.

16. Anastasiou E, et al. (2007) Control of Plant Organ Size by KLUH/CYP78A5-Dependent Intercellular Signaling. *Dev Cell* 13(6):843–856.

17. Ito T, Meyerowitz EM (2000) Overexpression of a gene encoding a cytochrome P450, CYP78A9, induces large and seedless fruit in arabidopsis. *Plant Cell* 12(9):1541–50.

18. Ehlting J, Hamberger B, Million-Rousseau R, Werck-Reichhart D (2006) Cytochromes P450 in phenolic metabolism. *Phytochem Rev* 5(2-3):239–270.

19. Wellesen K, et al. (2001) Functional analysis of the LACERATA gene of Arabidopsis provides evidence for different roles of fatty acid ω-hydroxylation in development. *Proc Natl Acad Sci U S A* 98(17):9694–9699.

20. Lee S, et al. (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. *Proc Natl Acad Sci U S A* 107(49):21205–10.

21. Nafisi M, et al. (2007) Arabidopsis cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. *Plant Cell* 19(6):2039–52.

22. Hansen CH (2001) Cytochrome P450 CYP79F1 from Arabidopsis Catalyzes the Conversion of Dihomomethionine and Trihomomethionine to the Corresponding Aldoximes in the Biosynthesis of Aliphatic Glucosinolates. *J Biol Chem* 276(14):11078–11085.

23. Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos Trans R Soc Lond B Biol Sci* 368(1612):20120426.

24. Feyereisen R (2011) Arthropod CYPomes illustrate the tempo and mode in P450 evolution. *Biochim Biophys Acta - Proteins Proteomics* 1814(1):19–28.

25. Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ (2009) CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21(10):3315–25.

26. Wang Q, et al. (2012) Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *J Biol Chem* 287(9):6159–68.

27. Hofer R, et al. (2014) Dual function of the cytochrome P450 CYP76 family from Arabidopsis thaliana in the metabolism of monoterpenols and phenylurea herbicides. *Plant Physiol* 166(3):1149–1161.

28. Miettinen K, et al. (2014) The seco-iridoid pathway from Catharanthus roseus. *Nat Commun* 5:3606.

29. Velasco R, et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* 2(12):e1326.

30. Nelson DR (2009) The cytochrome P450 homepage. *Hum Genomics* 4(1):59–65.

31. Nelson DR, Ming R, Alam M, Schuler MA (2008) Comparison of cytochrome P450 genes from six plant genomes. *Trop Plant Biol* 1(3-4):216–235.

32. Nelson DR, Schuler M a, Paquette SM, Werck-Reichhart D, Bak S (2004) Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiol* 135(2):756–772.

33. Falginella L, et al. (2010) Expansion and subfunctionalisation of flavonoid 3',5'-hydroxylases in the grapevine lineage. *BMC Genomics* 11(1):562.

34. Richly E, Kurth J, Leister D (2002) Mode of amplification and reorganization of resistance genes during recent Arabidopsis thaliana evolution. *Mol Biol Evol* 19(1):76–84.

35. Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Genet Genomics* 280(3):187–198.

36. Martin DM, et al. (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(1):226.

37. Parage C, et al. (2012) Structural, functional, and evolutionary analysis of the unusually large stilbene synthase gene family in grapevine. *Plant Physiol* 160(3):1407–1419.

38. Takos AM, Rook F (2012) Why biosynthetic genes for chemical defense compounds cluster. *Trends Plant Sci* 17(7):383–388.

39. Nützmann H-W, Osbourn A (2014) Gene clustering in plant specialized metabolism. *Curr Opin Biotechnol* 26:91–99.

40. Ayabe S, Akashi T (2006) Cytochrome P450s in flavonoid metabolism. *Phytochem Rev* 5(2-3):271–282.

41. Falginella L, Di Gaspero G, Castellarin SD (2012) Expression of flavonoid genes in the red grape berry of "Alicante Bouschet" varies with the histological distribution of anthocyanins and their chemical composition. *Planta* 236(4):1037–51.

42. Da Silva C, et al. (2013) The High Polyphenol Content of Grapevine Cultivar Tannat Berries Is Conferred Primarily by Genes That Are Not Shared with the Reference Genome. *Plant Cell* 25(12):4777–4788.

43. Cheng DW, et al. (2010) Transcriptional regulation of the grape cytochrome P450 monooxygenase gene CYP736B expression in response to Xylella fastidiosa infection. *BMC Plant Biol* 10:135.

44. Chakrabarti M, et al. (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proc Natl Acad Sci U S A* 110(42):17125–30.

45. Grimplet J, et al. (2012) Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. *BMC Res Notes* 5:213.

46. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408(14):796–814.

47. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.

48. Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6(1):31.

49. Rutherford K, et al. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16(10):944–945.

50. Vitulo N, et al. (2014) A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biol* 14(1):99.

51. Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21(9):1859–75.

52. Perazzolli M, et al. (2012) Downy mildew resistance induced by Trichoderma harzianum T39 in susceptible grapevines partially mimics transcriptional changes of resistant genotypes. *BMC Genomics* 13(1):660.

53. Sweetman C, Wong DC, Ford CM, Drew DP (2012) Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. *BMC Genomics* 13(1):691.

54. Vannozzi A, Dry IB, Fasoli M, Zenoni S, Lucchin M (2012) Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses. *BMC Plant Biol* 12(1):1.

55. Venturini L, et al. (2013) De novo transcriptome characterization of Vitis vinifera cv. Corvina unveils varietal diversity. *BMC Genomics* 14(1):41.

56. Kim D, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.

57. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–5.

58. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–2.

59. R Development Core Team R (2011) R: A Language and Environment for Statistical Computing. *R Found Stat Comput* 1(2.11.1):409.

60. Krzywinski M, et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9):1639–45.

61. Sonnhammer ELL, Durbin R (1995) A Dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167:1–10.

62. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.

63. Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Bioinformatics* 12(6):543–548.

64. Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* 27(2):221–224.

65. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17(4):540–552.

66. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

67. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)* (IEEE), pp 1–8.

68. Guindon S, et al. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* 59(3):307–321.

69. The angiosperm phylogeny group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 161(2):105–121.

70. Jones L, et al. (2014) Adaptive genomic structural variation in the grape powdery mildew pathogen, Erysiphe necator. *BMC Genomics* 15(1):1081.

71. Ramos MJ, et al. (2014) Flower development and sex specification in wild grapevine. *BMC Genomics* 15(1):1095.

72. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7):873–881.

73. Anders S, Pyl PT, Huber W (2014) HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.

74. Seo J, Gordish-Dressman H, Hoffman EP (2006) An interactive power analysis tool for microarray hypothesis testing and generation. *Bioinformatics* 22(7):808–14.

75. Beaudoin G a W, Facchini PJ (2013) Isolation and characterization of a cDNA encoding (S)-cis-N-methylstylopine 14-hydroxylase from opium poppy, a key enzyme in sanguinarine biosynthesis. *Biochem Biophys Res Commun* 431(3):597–603.

76. Dang T-TT, Facchini PJ (2014) CYP82Y1 Is N-Methylcanadine 1-Hydroxylase, a Key Noscapine Biosynthetic Enzyme in Opium Poppy. *J Biol Chem* 289(4):2013–2026.

77. Farrow SC, Hagel JM, Beaudoin G a W, Burns DC, Facchini PJ (2015) Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat Chem Biol* 11(9):728–732.

78. Haudenschild C, Schalk M, Karp F, Croteau R (2000) Functional expression of regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha spp.) in Escherichia coli and Saccharomyces cerevisiae. *Arch Biochem Biophys* 379(1):127–36.

79. Jørgensen K, et al. (2011) Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: isolation, biochemical characterization, and expression pattern of CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. *Plant Physiol* 155(1):282–92.

80. Larbat R, et al. (2009) Isolation and functional characterization of CYP71AJ4 encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. *J Biol Chem* 284(8):4776–85.

81. Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS (2006) Artemisia annua L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett* 580(5):1411–6.

82. Pfalz M, Vogel H, Kroymann J (2009) The Gene Controlling the Indole Glucosinolate Modifier1 Quantitative Trait Locus Alters Indole Glucosinolate Structures and Aphid Resistance in Arabidopsis. *Plant Cell* 21(3):985–999.

83. Ono E, et al. (2006) Formation of two methylenedioxy bridges by a Sesamum CYP81Q protein yielding a furofuran lignan, (+)-sesamin. *Proc Natl Acad Sci* 103(26):10116–10121.

84. Boachon B, et al. (2015) CYP76C1 (Cytochrome P450)-Mediated Linalool Metabolism and the Formation of Volatile and Soluble Linalool Oxides in Arabidopsis Flowers: A Strategy for Defense against Floral Antagonists. *Plant Cell* 27(10):2972–90.

85. Collu G, et al. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett* 508(2):215–220.

86. Diaz-Chavez ML, et al. (2013) Biosynthesis of Sandalwood Oil: Santalum album CYP76F Cytochromes P450 Produce Santalols and Bergamotol. *PLoS One* 8(9):e75053.

87. Guo J, et al. (2013) CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proc Natl Acad Sci*:1–6.

88. Hatlestad GJ, et al. (2012) The beet R locus encodes a new cytochrome P450 required for red betalain production. *Nat Genet* 44(7):816–20.

89. Irmler S, et al. (2000) Indole alkaloid biosynthesis in Catharanthus roseus: New enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J* 24(6):797–804.

90. Seki H, et al. (2011) Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell* 23(11):4112–23.

91. Biazzi E, et al. (2015) CYP72A67 catalyses a key oxidative step in Medicago truncatula hemolytic saponin biosynthesis. *Mol Plant* 8(10):1493–1506.

92. Mikkelsen MD, Hansen CH, Wittstock U, Halkier B a (2000) Cytochrome P450 CYP79B2 from Arabidopsis catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. *J Biol Chem* 275(43):33712–7.

93. Wittstock U (2000) Cytochrome P450 CYP79A2 from Arabidopsis thaliana L. Catalyzes the Conversion of L-Phenylalanine to Phenylacetaldoxime in the Biosynthesis of Benzylglucosinolate. *J Biol Chem* 275(19):14659–14666.

94. Andersen MD (2000) Cytochromes P-450 from Cassava (Manihot esculenta Crantz) Catalyzing the First Steps in the Biosynthesis of the Cyanogenic Glucosides Linamarin and Lotaustralin. *J Biol Chem* 275(3):1966–1975.

95. Christ B, et al. (2013) Cytochrome P450 CYP89A9 Is Involved in the Formation of Major Chlorophyll Catabolites during Leaf Senescence in Arabidopsis. *Plant Cell* 25(5):1868–1880.

96. Carelli M, et al. (2011) Medicago truncatula CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell* 23(8):3070–81.

97. Moses T, et al. (2014) Unravelling the Triterpenoid Saponin Biosynthesis of the African Shrub Maesa lanceolata. *Mol Plant* (October 2015).

98. Luo P, Wang YH, Wang GD, Essenberg M, Chen XY (2001) Molecular cloning and functional identification of (+)-delta-cadinene-8-hydroxylase, a cytochrome P450 mono-oxygenase (CYP706B1) of cotton sesquiterpene biosynthesis. *Plant J* 28(1):95–104.

99. Magome H, et al. (2013) CYP714B1 and CYP714B2 encode gibberellin 13-oxidases that reduce gibberellin activity in rice. *Proc Natl Acad Sci U S A* 110(5):1947–52.

100. Kraus PF, Kutchan TM (1995) Molecular cloning and heterologous expression of a cDNA encoding berbamunine synthase, a C--O phenol-coupling cytochrome P450 from the higher plant Berberis stolonifera. *Proc Natl Acad Sci U S A* 92(6):2071–2075.

101. Pauli HH, Kutchan TM (1998) Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3'-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. *Plant J* 13(6):793–801.

102. Greer S, et al. (2007) The cytochrome P450 enzyme CYP96A15 is the midchain alkane hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of Arabidopsis. *Plant Physiol* 145(3):653–667.

103. Laudert D, Pfannschmidt U, Lottspeich F, Holländer-Czytko H, Weiler EW (1996) Cloning, molecular and functional characterization of Arabidopsis thaliana allene oxide synthase (CYP 74), the first enzyme of the octadecanoid pathway to jasmonates. *Plant Mol Biol* 31(2):323–335.

104. Park JH, et al. (2002) A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in Arabidopsis due to a block in jasmonic acid biosynthesis. *Plant J* 31(1):1–12.

105. Hughes RK, De Domenico S, Santino A (2009) Plant cytochrome CYP74 family: Biochemical features, endocellular localisation, activation mechanism in plant defence and improvements for industrial applications. *ChemBioChem* 10(7):1122–1133.

Supplemental information

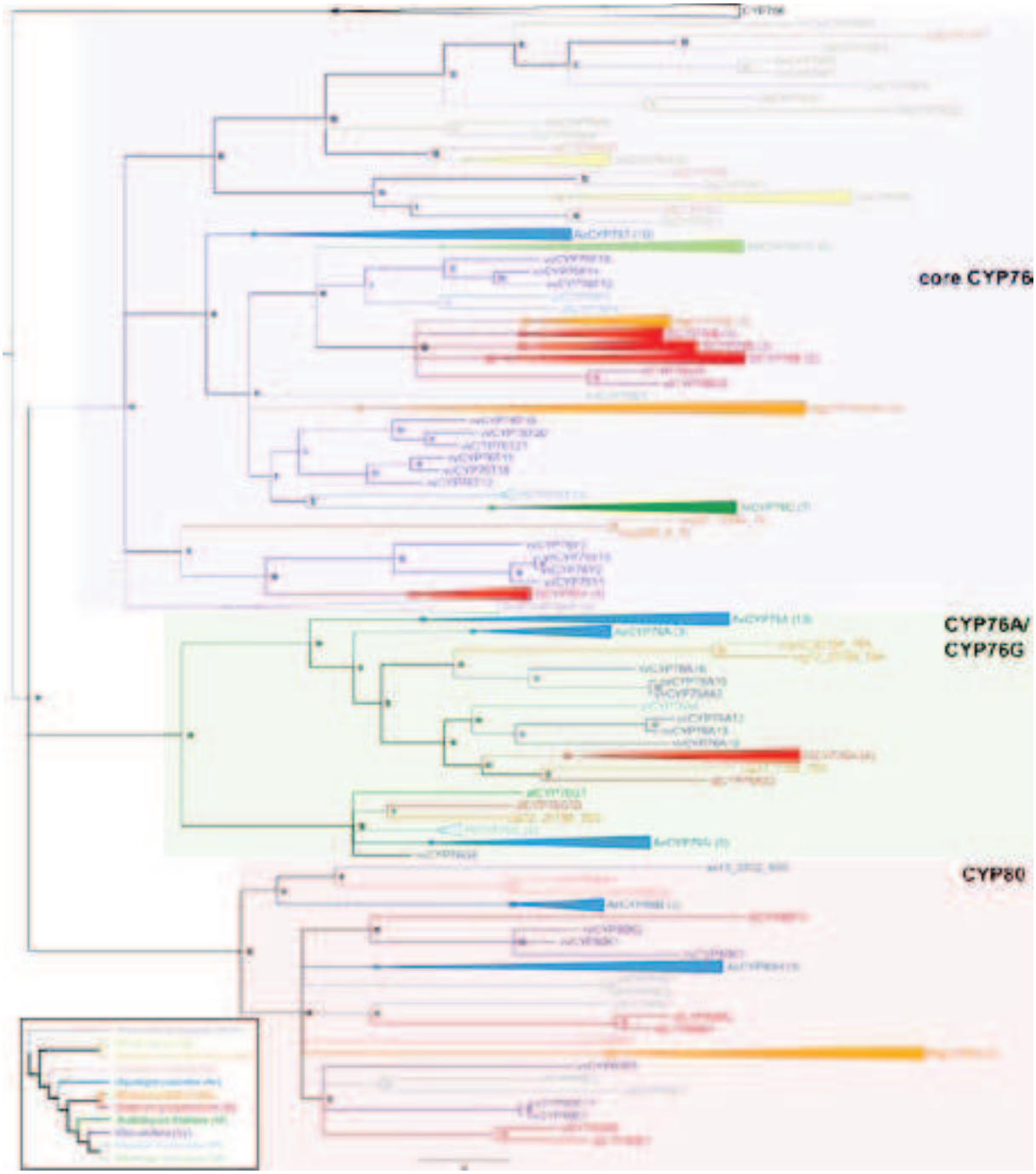

Figure S1. Phylogeny of CYP80 and CYP76 in angiosperms. Maximum likelihood tree of full length CYP76 and CYP80 protein sequences from a selection of angiosperms, rooted with CYP706 from all the included species. Nodes with bootstrap values below 60 are collapsed to trifurcations. Species specific clades with more than two members (except *V. vinifera*) are collapsed to triangles. The label of the triangle gives the subfamily and the number of members contained in the clade.

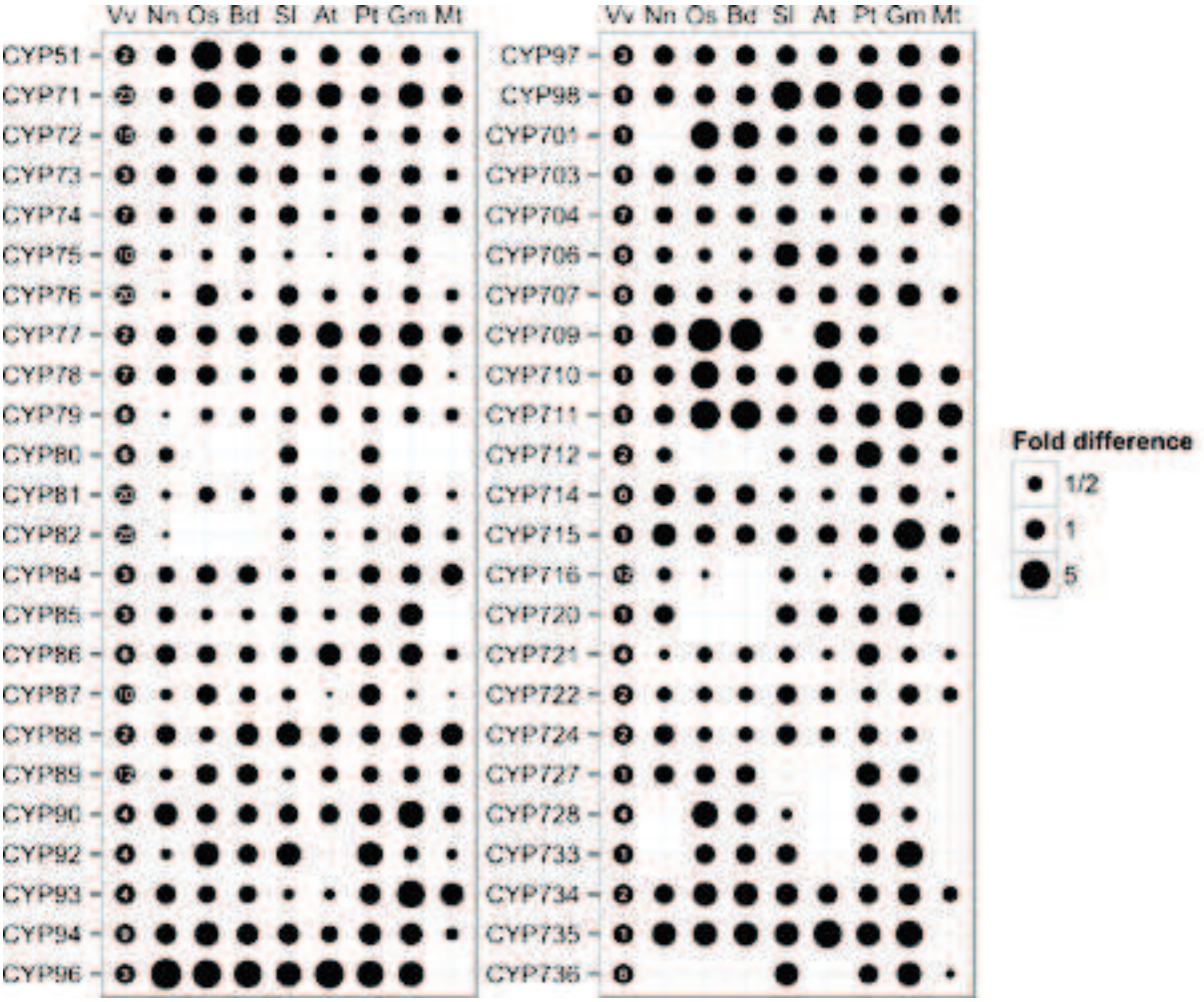

Figure S2. Comparison of the number of P450 genes per family between species. Dot size is proportional to the relative family size (number of genes per family) in a given species compared to *Vitis vinifera* (Vv = *Vitis vinifera*, Nn = *Nelumbo nucifera*, Os = Oryza sativa, Bd = *Brachypodium distachyon*, Sl = *Solanum lycopersicum*, At = *Arabidopsis thaliana*, Pt = *Populus trichocarpa*, Gm = *Glycine max*, Mt = *Medicago truncatula*). The numbers in the first column are the absolute family sizes (numbers of genes per family) in *Vitis vinifera*. The number of genes per family was retrieved from the cytochrome P450 homepage. Pseudogenes and families not present in *V. vinifera* (CYP83, CYP99, CYP702, CYP705, CYP708, CYP718 and CYP729) were excluded from the count.

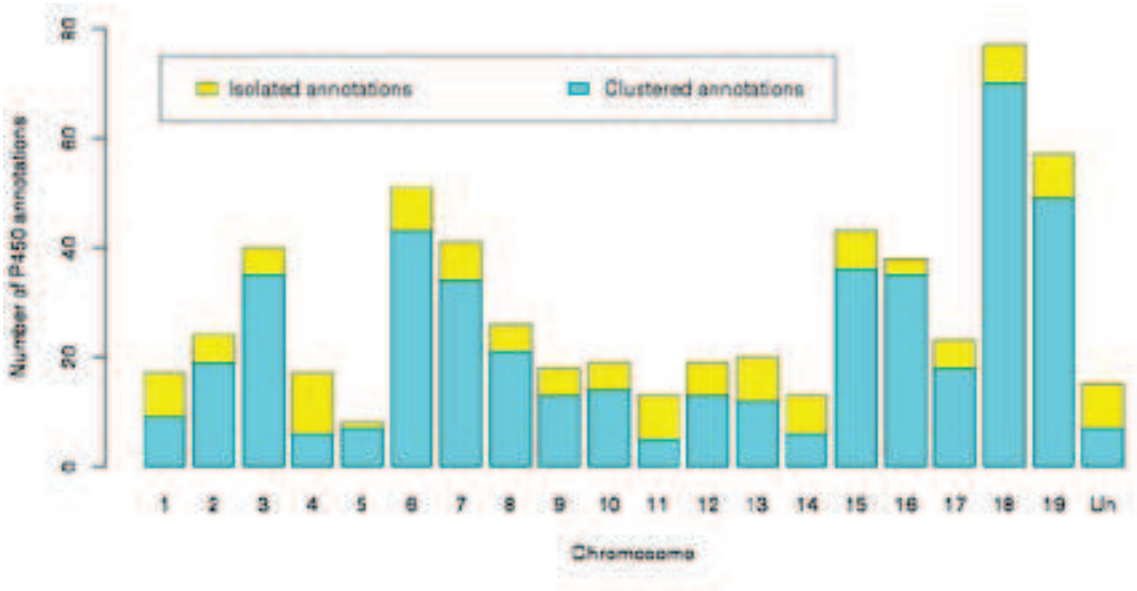

Figure S3. Distribution of the *V. vinifera* P450s per chromosome. The blue bar corresponds to clustered annotations and the yellow bar to the isolated annotations. The "Unknown chromosome" is labeled as "Un".

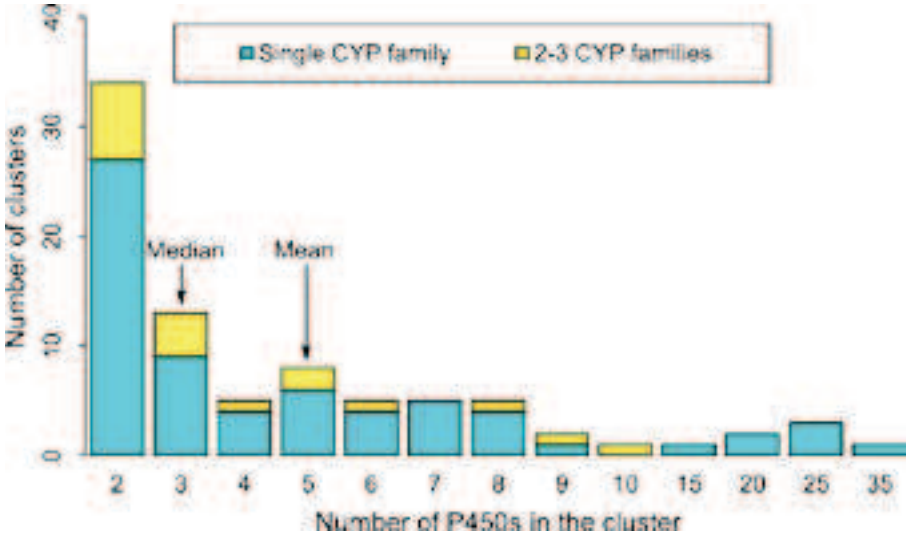

Figure S4. Distribution of the P450 sequences per physical cluster. Median and average values are labeled with arrows. The clusters composed of a single P450 family are represented in blue and those composed of 2 or 3 P450 families in orange.

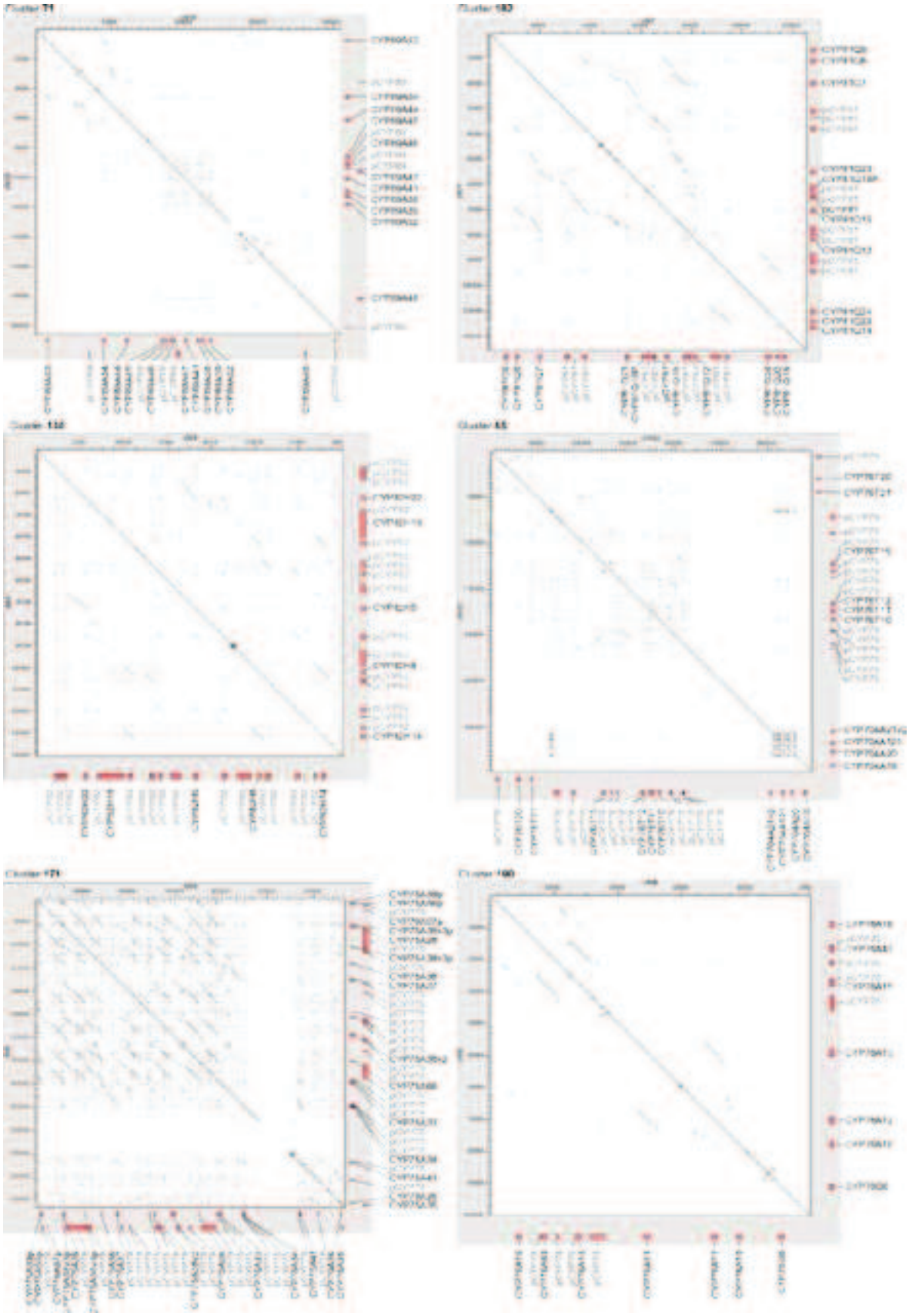

Figure S5. Dot matrix plots of the largest physical clusters. The dots and the black lines represent the sequence similarities in cluster 92 compared to itself. The red rectangles on the sides of the graph represent cytochrome P450 sequences. Complete genes are labeled with their name and pseudogenes are labeled with "p" and the P450 family.

Table S4. List of P450 families with majority of its members grouped in physical clusters.

Family	Family size (total sequences)	Number of clustered members	Examples of functions in other organisms
CYP82	69	66	Biosynthesis of homoterpenes in *A. thaliana* (20), opioids in *Papaver somniferum* (75–77).
CYP71	51	45	Biosynthesis of monoterpenoids in mint species *Mentha x piperita* and *Mentha x spicata* (78), cyanogenic glucosides in cassava (*Manihot esculenta*) (79), furanocoumarins in several species (80), artemisinin in *Artemisia annua* (81), flavonoids in soybean (*Glycine max*). (40)
CYP81	50	42	Biosynthesis of indole glucosinolates in *A. thaliana* (82), isoflavonoid phytoalexins in *Medicago truncatula, G. echinata* and*, Lotus japonicus* (40), sesamin in *Sesamum spp*. (83)
CYP76	42	42	Biosynthesis of monoterpene volatiles in *A. thaliana* (84), monoterpene indole alkaloids in *Catharanthus roseus* (28, 85), sesquiterpene volatiles in sandalwood (*Santalum album*) (86), phytoalexins in rice (*Oryza sativa*) (25, 26), tanshinones in Chinese sage (*Salvia miltiorrhiza*)(87), pigment betalain in beetroot (*Beta vulgaris*) (88).Metabolism of xenobiotics in *A. thaliana* (27).
CYP72	36	33	Biosynthesis of monoterpene indole alkaloids in *C. roseus* (28, 89), glycyrrhizin in licorice (*Glycyrrhiza*) (90), saponins in *M. trucantula* (91).
CYP79	26	25	Biosynthesis of glucosinolates in *A. thaliana* (22, 92, 93), cyanogenic glucosides in cassava (*M. esculenta*) (94).
CYP89	25	21	Chlorophyll degradation in *A. thaliana* (95).
CYP75	24	23	Biosynthesis of flavonoids in *Petunia x hybrida*, *A. thaliana, Gentiana triflora, C. roseus, etc*. (40)
CYP716	23	12	Biosynthesis of saponins in *M. trucantula* (96) and *Maesa lanceolata* (97).
CYP706	21	19	Biosyntheis of sesquiterpenoids in cotton (*Gossypium arboreum*) (98).
CYP87	20	15	Biosynthesis of saponins in *Maesa lanceolata* (97).
CYP714	16	13	Degradation of hormones (gibberelin) in rice (*O. sativa*) (99).
CYP736	13	11	Unknown. Pathogen response in grapevine *V. vinifera* (43).
CYP728	11	9	Unknown.
CYP80	10	9	Alkaloid biosynthesis in barberry *(Berberis stolonifera)* (100) and California poppy (*Eschscholzia californica*) (101).
CYP96	9	9	Biosynthesis of cuticular wax in *A. thaliana* (102).
CYP721	8	8	Unknown.
CYP74	7	7	Biosynthesis of hormones (jasmonates) and C6 volatiles in *A. thaliana* and other plants (103–105).
CYP92	7	5	Unknown.
CYP93	7	6	Biosynthesis of flavonoids in soybean (*G. max*), *Glycyrrhiza echinata*, *Gerbera hybrid*, *Antirrhinum majus*, *Torrenia hybrid*, etc. (40)
CYP712	6	5	Unknown.

Table S5. Description of RNA-Seq experiments used for analysis of gene expression.

Genotype	Tissue	Conditions	Reference
Carignan	Leaves	Infection with powdery mildew pathogen *Erysiphe nectator*	(70)
Pinot noir	Leaves	Infection with downy mildew pathogen *Plasmopara viticola*	(52)
Touriga Nacional	Flowers	Development	(71)
Tannat	Berries	Development	(42)
Shiraz	Berries	Development	(53)
Corvina	Berries	Development	(55)
4 cultivars	Berries	Development	Unpublished

Chapter 6
CYP76F14 catalyzes biosynthesis of wine lactone precursor from linalool

Tina Ilc[1], Laurence Miesch[2], Florian Lauvoisard[2], Lucie Kriegshauser[1], Andrea Ilg[3], Eric Duchêne[3], Philippe Hugueney[3], Danièle Werck-Reichhart[1], Nicolas Navrot[1]

[1] Institute of Plant Molecular Biology, Centre National de la Recherche Scientifique, University of Strasbourg, Strasbourg, France

[2] Laboratoire de Chimie Organique Synthétique, Centre National de la Recherche Scientifique University of Strasbourg, Strasbourg, France

[3] Laboratoire Métabolisme Secondaire de la Vigne, Institut National de la Recherche Agronomique, University of Strasbourg, Colmar, France

Abstract

Monoterpenes constitute an important part of wine aroma, notably the floral notes of some wine varieties. The monoterpene in wine with the lowest sensory threshold is the wine lactone, a bicyclic lactone that can form *via* acid-catalyzed cyclization of (*E*)-8-carboxylinalool (also known as menthiafolic acid) during wine maturation. In a sample of 23 white wines, we show that (*E*)-8-carboxylinalool content correlates with the content in wine lactone and we estimate the kinetic constant for the formation of wine lactone from (*E*)-8-carboxylinalool. The concentration of this precursor was higher in Gewurztraminer wines compared to other two varieties studied. We explored the biosynthetic origin of (*E*)-8-carboxylinalool in grapes and identified a set of candidate cytochrome P450s likely to contribute to the conversion of linalool into (*E*)-8-carboxylinalool based on their gene expression during grape berry maturation. One of them, CYP76F14, which is highly expressed in the late stages of berry ripening, can efficiently oxidize linalool to (*E*)-8-carboxylinalool *in vitro* as well as *in planta*.

Introduction

Grapevine (*Vitis vinifera L.*) is one of economically the most important crops worldwide with most of the 70 million tons of yearly grape harvest used for winemaking (1). The quality of wine to a large extent depends on its aroma, but its chemical complexity makes it a difficult subject to study, let alone predict: it consists of several hundred different volatile compounds at concentrations spanning several orders of magnitude. The advances in analytical chemistry in the second half of the 20th century allowed scientists to study this complex chemical mixture in more detail. It gradually transpired that characteristic wine aroma does not depend as much on the most abundant compounds, but rather on low-concentration compounds with very strong odor. One is isobutyl methoxypyrazine, responsible for characteristic bell pepper scent of Cabernet Sauvignon and Sauvignon Blanc wines, where it reaches concentrations of about 30 pg/L (about an order of magnitude higher than its odor threshold) (2). Another example is rotundone, a sesquiterpene which conveys the peppery aroma to Shiraz wines at concentrations as low as 20 ng/L (3). Wine lactone, a bicyclic monoterpene lactone, is another trace wine aroma compound with a very low odor threshold (10 ng/L) (4). Pure wine lactone has a sweet, woody and coconut-like aroma (5) and was shown to contribute to the aroma of Gewurztraminer wines (4, 6). It might also contribute to the aroma profile of fresh orange (7) and grapefruit (8) juice.

As opposed to rotundone or isobutyl methoxypyrazine, wine lactone is not synthetized in grapes, but predominantly during wine maturation (9). It forms in a slow, non-enzymatic, acid-catalyzed cyclisation from an odorless precursor, (*E*)-8-carboxylinalool or menthiafolic acid (10–12). It has therefore been suggested that wine lactone concentration is increasing with wine aging (13), but the data on both (*E*)-8-carboxylinalool and wine lactone concentrations in wine are scarce. Concentrations of wine lactone in wine are very low, which causes analytical difficulties (14), whereas (*E*)-8-carboxylinalool is often overlooked because gas chromatography is traditionally used for analysis of aroma compounds, although it is not be appropriate for the analysis of more polar molecules. The analysis is additionally impeded by the fact that (*E*)-8-carboxylinalool is found both in grapes and wines in the form of a glucose ester, with the sugar moiety attached to the carboxyl functional group (10, 15, 16). Monoterpene glycosides and glucose esters were suggested to act as a precursor pool for the formation of aroma compounds during wine fermentation and maturation, as they are hydrolyzed chemically (at low pH) or enzymatically (17).

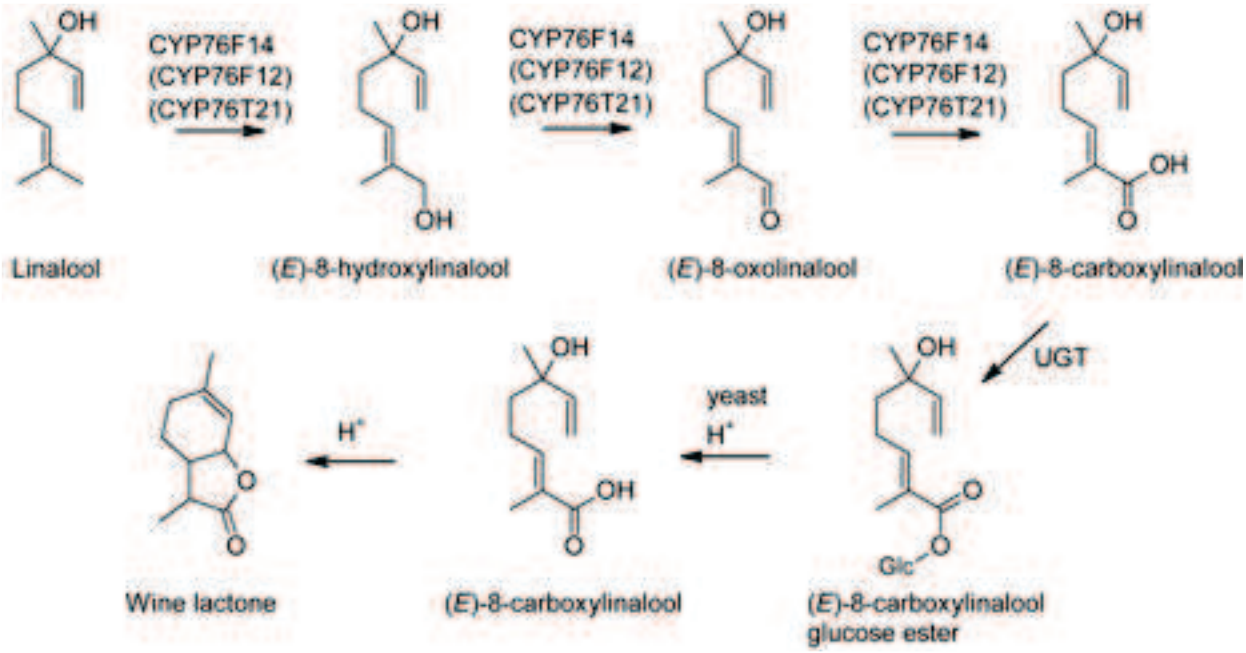

Figure 28. The proposed pathway for wine lactone formation.

Luan *et al.* (18) proposed that (*E*)-8-carboxylinalool in grapes is formed from the monoterpene linalool, another important aroma compound, in a multi-step enzymatic reaction: linalool is first oxidized at C8 to yield (*E*)-8-hydroxylinalool, also a common grape metabolite (19), which is then further oxidized at the same carbon atom to (*E*)-8-carboxylinalool (Figure 1), followed by glycosylation to form the glucose ester.

Formation of linalool by grapevine terpene synthases (TPS) has already been studied (20), but the enzymes catalyzing downstream oxidation to (*E*)-8-carboxylinalool remained unknown. In many plants monoterpenes are oxidized by cytochromes P450 (21), heme containing enzymes that use molecular oxygen to oxidize small, non-polar substrates, usually with remarkable regio- and stereospecificity. P450s from families CYP76 and CYP71 often oxidize monoterpene alcohols, such as linalool (22–24) or geraniol (25). CYP76C1, for example, converts linalool into (*E*)-8-carboxylinalool in *Arabidopis thaliana* flowers (24). Enzymes from CYP76 and CYP71 families thus emerged as likely candidates for (*E*)-8-carboxylinalool synthesis in grapes.

The aim of this work was to elucidate the pathway of wine lactone formation in wines, from the early enzymatic steps during grape berry development to the non-enzymatic steps during wine maturation and aging. Using liquid chromatography coupled to mass spectrometry, we first establish the relationship between the concentrations of wine lactone and its precursors in a selection of wines of different ages. We show that the expression of a subset of P450 genes is induced in the later stages of berry ripening and demonstrate that, among these, only CYP76F14, which is the most highly expressed CYP76 in ripe berry, efficiently catalyzes the

conversion of linalool to (*E*)-8-carboxylinalool both *in vitro* and *in planta*. We therefore suggest that CYP76F14 is responsible for the synthesis of (*E*)-8-carboxylinalool, the wine lactone precursor in grapes.

Results

Wine lactone concentration in wine is determined by wine age and precursor concentration

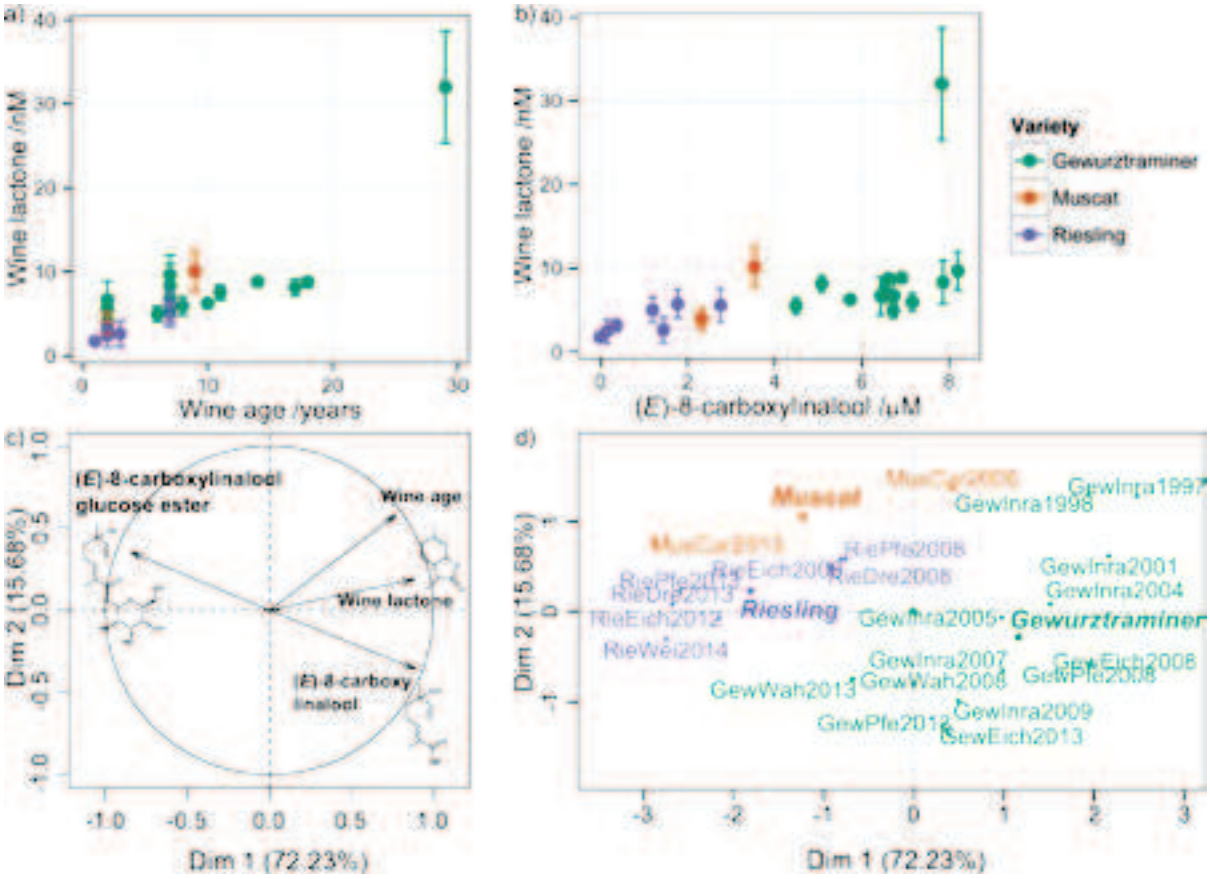

Figure 29. Relationships between wine age and concentrations of wine lactone, (*E*)-8-carboxylinalool and (*E*)-8-carboxylinalool glucose ester. All three compounds were quantified in the same wines by LC-MS/MS (N=3). Concentration of wine lactone is higher in older wines (a) and wines with higher concentration of its precursor, (*E*)-8-carboxylinalool (b). Principal component analysis of metabolite concentrations and the age of wine confirms this observation: (c) variables factor map and (d) individuals factor map (Mus = Muscat, Rie = Riesling, Gew = Gewurztraminer). The 29-year-old wine was excluded from PCA analysis.

Low abundance of wine lactone makes it a difficult analytical subject with only a few successful attempts of wine lactone quantification in wines reported to date (6, 14). We first attempted analysis of wine lactone in concentrated wine extracts by gas chromatography coupled to mass spectrometry (GC-MS), the conventional method for analysis of volatile compound. However, the limit of detection was too high to allow for the analysis of wine lactone at the low concentrations expected in wines. We therefore developed a liquid chromatography–tandem mass spectrometry (LC-MS/MS) method, which enabled the

quantification of wine lactone in all 23 investigated wines (Figure S1). We also used LC-MS/MS to analyze (*E*)-8-carboxylinalool by direct injection of wines without extraction (Figure S2). With this method we also detected a conjugated form of (*E*)-8-carboxylinalool, putatively identified as (*E*)-8-carboxylinalool glucose ester.

To determine the factors controlling wine lactone concentrations in wine, we analyzed wines of different ages and varieties for their content of wine lactone and its two precursors, (*E*)-8-carboxylinalool and (*E*)-8-carboxylinalool glucose ester. This analysis showed that concentration of wine lactone was increasing with wine age (Figure 2a), as well as with concentration of (*E*)-8-carboxylinalool (Figure 2b). The concentration of (*E*)-8-carboxylinalool was strongly dependent on wine variety, and much higher in Gewurztraminer than in Riesling.

Concentrations of wine lactone were generally about 1000 times lower compared to (*E*)-8-carboxylinalool. Only a small proportion of (*E*)-8-carboxylinalool was thus converted to wine lactone, which was in good agreement with the relatively stable concentration of (*E*)-8-carboxylinalool over time. Although the concentrations of wine lactone were low (in the nM range), they were all above the sensory threshold of wine lactone, which is 60 pM in model wine (4).

Principal component analysis was performed on wine age and concentrations of (*E*)-8-carboxylinalool, (*E*)-8-carboxylinalool glucose ester and wine lactone to further explore the relationships between these factors (Figure 2c–d). This analysis revealed a strong negative correlation between the concentrations of (*E*)-8-carboxylinalool glucose ester and (*E*)-8-carboxylinalool, supporting a decrease in (*E*)-8-carboxylinalool glucose ester by acid hydrolysis yielding free (*E*)-8-carboxylinalool (Figure S3). In Gewurztraminer wines (*E*)-8-carboxylinalool glucose ester was only detected in 1-year-old wines, but not in 6-year- and older wines. In the wines without (*E*)-8-carboxylinalool glucose ester, the concentration of (*E*)-8-carboxylinalool was independent of age. Wine lactone concentration correlated with wine age and concentration of (*E*)-8-carboxylinalool, supporting its slow formation from (*E*)-8-carboxylinalool.

We used the first order reaction rate law to model conversion of (*E*)-8-carboxylinalool to wine lactone. Using the approximation of a constant concentration of (*E*)-8-carboxylinalool, we determined a rate constant $k = (0.00012 \pm 0.00002)$ year^{-1}. Because the kinetic constant was so low, we could use a linear approximation for the reaction rate, with [wine lactone] = k · [(*E*)-8-carboxylinalool] · wine age. As evident from the plot of observed vs. calculated values, the deviations from predicted values were higher in young wines (Figure S4). A

possible explanation is formation of wine lactone from (*E*)-8-carboxylinalool glucose ester, or in grape berries or by the yeast during fermentation.

Metabolic profiling pointed to differences between varieties. Gewurztraminer wines contained more (*E*)-8-carboxylinalool than Riesling or Muscat, and, as a result, they were also enriched in wine lactone. To explore the underlying causes of this differential accumulation we investigated the biosynthesis of (*E*)-8-carboxylinalool.

Expression of candidate P450 genes parallels increasing (*E*)-8-carboxylinalool concentration during grape berry ripening

The concentrations of many grape volatiles, including oxygenated monoterpene derivatives, are reported to be higher in the ripe compared to green berries (26, 27). To establish (*E*)-8-carboxylinalool accumulation profile we thus analyzed its content in developing Gewurztraminer berries at five stages from green to ripe berry. Free (*E*)-8-carboxylinalool was not detected at any developmental stage. We could, however, detect it in a conjugated form, putatively identified as (*E*)-8-carboxylinalool glucose ester. Its concentration increased during ripening of Gewurztraminer berries (Figure 3a).

We postulated that the expression of genes involved in the biosynthesis of (*E*)-8-carboxylinalool glucose ester would parallel its accumulation in ripening grape berries. In the first step, we compared gene expression in Gewurztraminer berries at two developmental stages (green and mid-ripe berry) by RNA sequencing, with main focus on members of the CYP71 and CYP76 families recently shown to metabolize linalool in *A. thaliana* (22–24). Annotation of cytochromes P450 in *V. vinifera* PN40024 genome revealed an expanded CYP76 family with 20 full-length coding sequences (**Chapter 5**). Grapevine CYP71 family, on the other hand, is of comparable size or smaller than CYP71 families in other plant genomes and comprises of 23 predicted genes (28). RNA sequencing of Gewurztraminer berries highlighted five ripening induced CYP76 (*CYP76F12*, *CYP76F14*, *CYP76T21*, *CYP76Y1* and *CYP76Y2*) and one CYP71 (*CYP71AT7*) gene candidate (Figure S5). Analysis of expression of the CYP76 family in the Corvina cultivar (29) confirmed induced expression of gene candidates in ripe berries (Figure S6). Most of these genes are expressed to comparable levels in leaves, buds and roots.

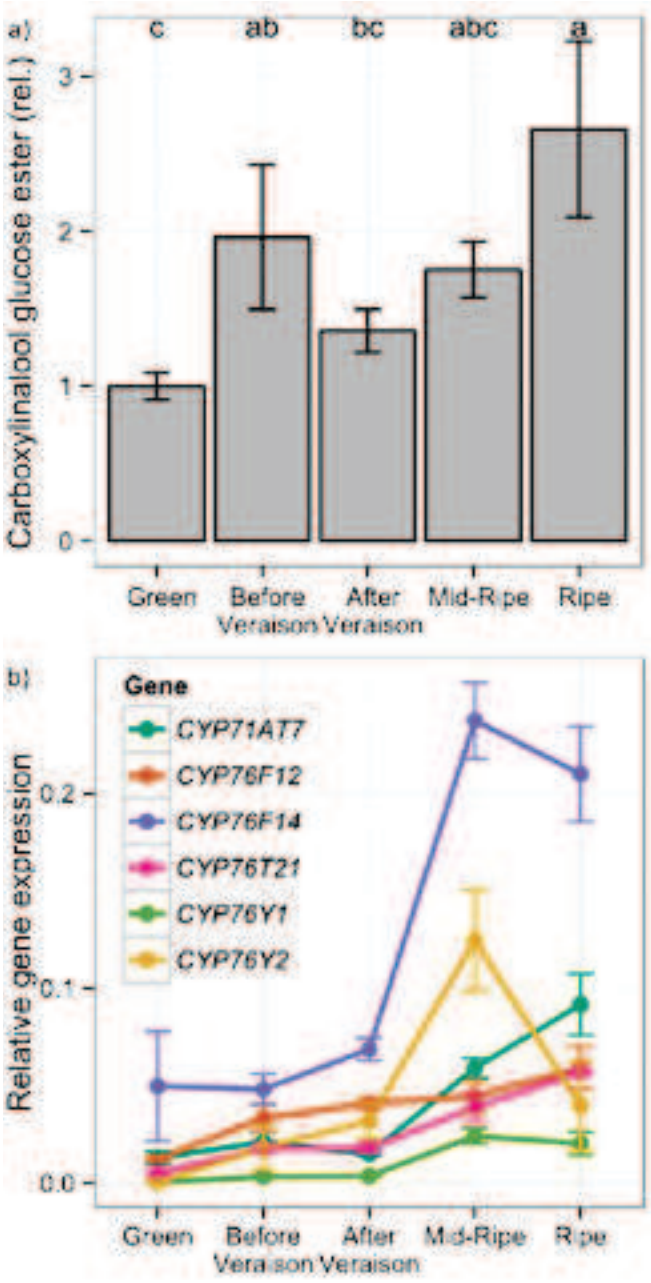

Figure 30. Comparative increase in (*E*)-8-carboxylinalool glucose ester content and candidate P450 genes expression during grape berry development. a) Relative concentration (±SD) of (*E*)-8-carboxylinalool glucose ester in developing Gewurztraminer berries. One-way ANOVA revealed significant differences between developmental stages (p=0.0016). Letters indicate significant differences among the means calculated with Tukey HSD test (α = 0.05). b) Expression of candidate genes in Gewurztraminer grape berries as determined by qPCR. Error bars represent standard deviation of 3 technical and 3 biological replicates. Gene expression was normalized to the expression of three constitutively expressed genes (30).

The expression level of the candidates was investigated in more detail at 5 developmental stages (green, before veraison, after veraison, mid-ripe and ripe) of Gewurztraminer berries by qRT-PCR. Expression of all candidate genes increased during the ripening (Figure 3). However, only expression of *CYP71AT7*, *CYP76F12* and *CYP76T21* kept increasing until the ripe stage. Expression of *CYP76Y1* and *CYP76F14* peaked in mid-ripe berries and remained constant in the ripe berries. Expression of *CYP76Y2* peaked in mid-ripe berries as well, but

dropped in ripe berries. *CYP76F14* reached the highest expression level of all candidates in the late stages of Gewurztraminer berry ripening.

Expression of several CYP76 and one CYP71 genes therefore parallels the accumulation of (*E*)-8-carboxylinalool throughout the grape ripening. Those were considered as strong candidates for (*E*)-8-carboxylinalool biosynthesis.

Several candidate enzymes metabolize linalool *in vitro*

To evaluate the catalytic activity of the candidate P450s, we cloned their cDNA from *V. vinifera* cv. Muscat Ottonel and expressed them in yeast together with the P450 reductase 1 from *A. thaliana*. CO-differential spectroscopy (31) confirmed the expression of all of them, except CYP76Y1 (Figure S7). As previously observed for other members of CYP76 family (23), the yeast expression level of grape CYP76 enzymes was relatively low.

We evaluated the enzyme activities by incubating the recombinant yeast microsomal fractions with linalool, (*E*)-8-hydroxylinalool and (*E*)-8-oxolinalool. Three out of five candidate enzymes (CYP76F14, CYP76F12 and CYP76T21) metabolized linalool (Figure 4a) and (*E*)-8-hydroxylinalool (Figure 4b), CYP76F14 was the most active on both substrates. Conversion of (*E*)-8-oxolinalool proved difficult to evaluate due to a competing reduction of this substrate, probably catalyzed by a microsomal yeast enzyme (24).

Gas chromatography of the reaction products (Figure S7) revealed that CYP76F12, CYP76F14 and CYP76T21 converted linalool to (*E*)-8-hydroxylinalool. Two additional products of the three enzymes (Figure S8) were identified as downstream products of (*E*)-8-oxolinalool ((*E*)-8-hydroxy-6,7-dihydrolinalool and (*E*)-8-oxo-6,7-dihydrolinalool), presumably generated by an endogenous membrane bound yeast enzyme (24). (*E*)-8-oxolinalool was sometimes detected as an intermediate product, but was in most cases reduced to these two side products, which were also found in the negative control, reaction with microsomes from yeast transformed with an empty vector. CYP76F12 differed from the other two enzymes in that it produced both (*E*) and (*Z*) isomers of 8-hydroxylinalool. No major product could be detected for CYP71AT7, despite the consumption of low amounts of substrate.

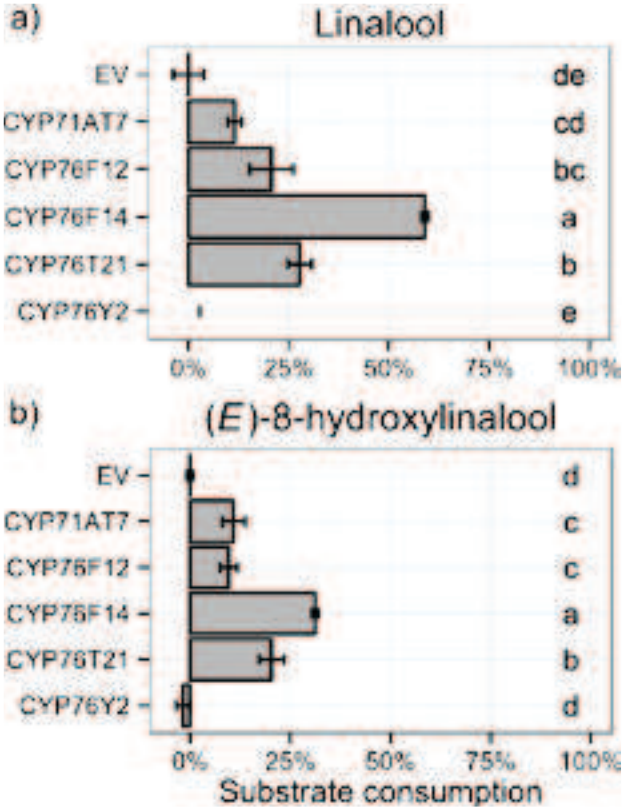

Figure 31. Conversion of linalool and *(E)*-8-hydroxylinalool by grapevine cytochromes P450 expressed in yeast. Microsomal membranes prepared from yeast expressing the candidate P450s were incubated with substrates at 100 µM concentration for 1h in the presence of NADPH. Reaction mixtures were extracted and analyzed by GC-FID. Conversion (± SD, N=3) is calculated relative to the residual concentration of substrate in negative control (microsomal fraction from yeast transformed with an empty vector - EV) incubated under the same conditions. One-way analysis of variance (ANOVA) revealed significant differences ($P < 0.001$) for both substrates. Letters on the right axes indicate significant differences between means calculated with Tukey HSD test ($\alpha = 0.05$).

The reaction products were also analyzed by LC-MS/MS. This analysis confirmed the products already identified by GC and in addition allowed for the detection of the final product, (*E*)-8-carboxylinalool, which is not readily detectable by GC (13). It showed that only one of the candidate P450s, CYP76F14, catalyzed the conversion of linalool to (*E*)-8-carboxylinalool as the main product (Figure 5). Similarly, when the next molecule in the pathway, (*E*)-8-hydroxylinalool, was used as a substrate, only CYP76F14 could oxidize it to (*E*)-8-carboxylinalool. However, when (*E*)-8-oxolinalool was used as a substrate, three enzymes, CYP76F14, CYP76F12 and CYP76T21, oxidized it to (*E*)-8-carboxylinalool (Figure S11). CYP76F14 is thus the only candidate that catalyzes the whole cascade of oxidations leading to the conversion of linalool into (*E*)-8-carboxylinalool, mainly because of its ability to efficiently oxidize (*E*)-8-hydroxylinalool to (*E*)-8-oxolinalool.

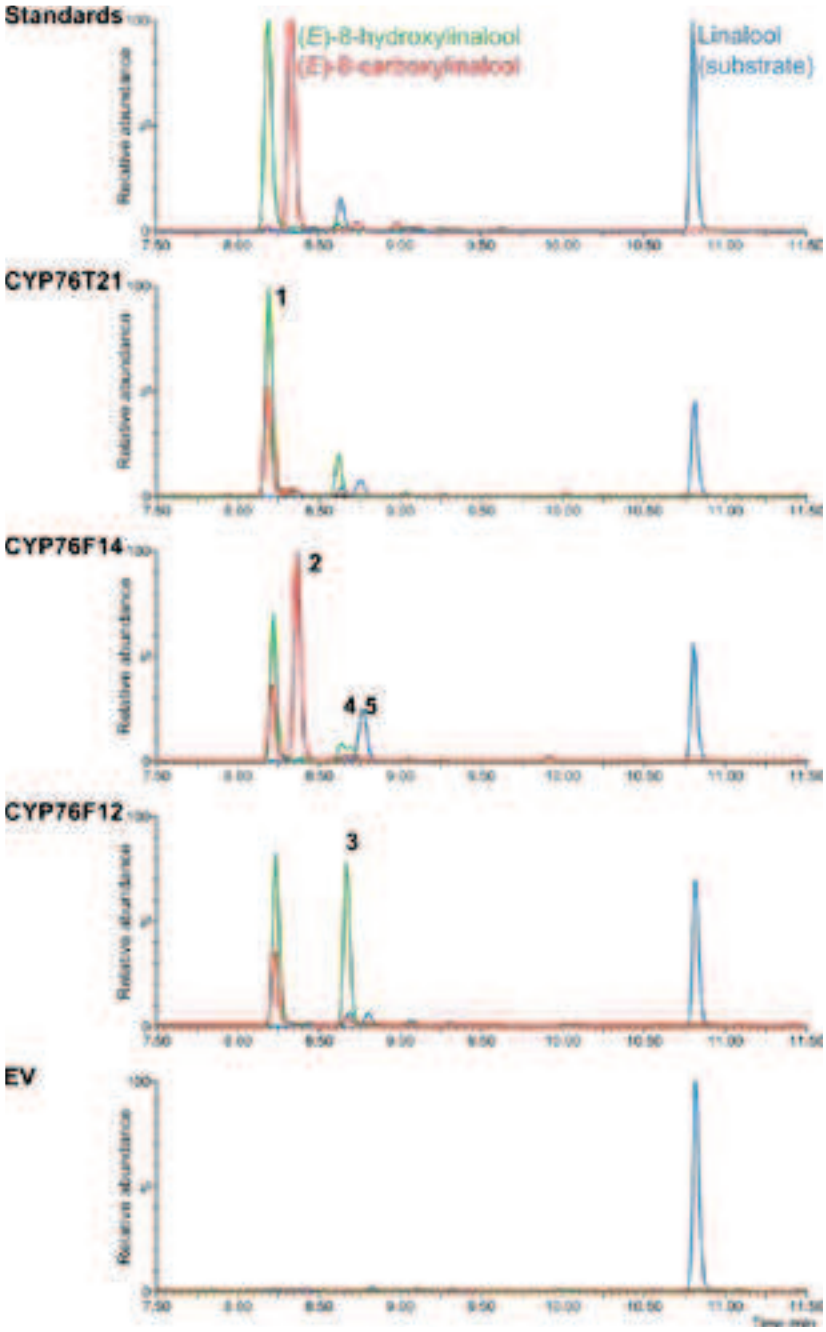

Figure 32. Identification of the products of linalool metabolism by relevant candidate grape P450s expressed in yeast. LC-MS/MS reaction profiles of three enzyme candidates in multiple-reaction-monitoring (MRM) mode. Incubations are performed in the same conditions as for Figure 4. Each color represents an MRM channel for specific detection of target molecules: blue for linalool (137>80.7), green for (*E*)-8-hydroxylinalool (135>106.8), and red for (E)-8-carboxylinalool (167.2>92.8). Mass spectra of identified (Figure S9) and putative products (Figure S10) are available in supplementary information. Negative control is a reaction with microsomes from yeast transformed with an empty vector. **1**–(*E*)-8-hydroxylinalool, **2**–(*E*)-8-carboxylinalool, **3**–(*Z*)-8-hydroxylinalool, **4**–(*E*)-8-oxo-6,7-dihydrolinalool, **5**–(*E*)-8-hydroxy-6,7-dihydrolinalool. **4** and **5** are side products of yeast metabolism of (*E*)-8-oxolinalool.

C. roseus and *A. thaliana* members of the CYP76 family were recently shown to metabolize a broad set of monoterpenols (23, 32, 33). We therefore also tested the candidate enzymes for their activity with geraniol, another major monoterpenol in grapes (**Chapter 1**). All candidate enzymes converted at least small amounts of geraniol. CYP76T21 was the most

active on geraniol (Figure S12) and the main reaction product detected by GC analysis was (*E*)-8-hydroxygeraniol (Figure S13).

The functional analysis of the selected P450 candidates expressed in yeast thus indicates that, whereas several of them can potentially contribute to linalool or geraniol oxidation in maturing grape berries, CYP76F14 is the most active on linalool and its oxidation products, and the only enzyme that can efficiently catalyze the whole cascade of oxidation reactions from linalool to (*E*)-8-carboxylinalool.

Reconstruction of the (*E*)-8-carboxylinalool pathway in *Nicotiana benthamiana* leaves

We used Agrobacterium-mediated transient transformation of *N. benthamiana* leaves to investigate the activity of the selected P450 enzymes *in planta*. In order to provide the P450 enzymes with linalool substrate, a linalool synthase was first cloned from grapevine. cDNA sequences of putative LIS expressed in Gewurztraminer berries were reconstituted based on RNAseq data and on homology with LIS identified in the grapevine reference genome (20). A full-length cDNA corresponding to a putative TPS56 was then amplified from RNAs prepared from mid-ripe Gewurztraminer berries. Transient expression this cDNA in leaves of *N. benthamiana* resulted in the biosynthesis of significant amounts of linalool (data not shown), demonstrating that this cDNA encoded an active linalool synthase.

To test if the simultaneous expression of a grapevine linalool synthase and the candidate linalool oxidases were sufficient to generate significant amounts of (*E*)-8-carboxylinalool in a plant cell environment, we transiently co-expressed LIS together with one of the three best candidate P450 genes in leaves of *N. benthamiana*. Linalool and its derivatives were analyzed in methanol leaf extracts by LC-MS/MS (Figure 6). All the targeted metabolites—linalool, (*E*)-8-hydroxylinalool and (*E*)-8-carboxylinalool—were detected in both free and glycosylated forms. Neither free nor glycosylated (*E*)-8-oxolinalool could be detected.

Both CYP76F14 and CYP76T21 depleted all the linalool produced by LIS, but only CYP76F14 completely converted it to (*E*)-8-carboxylinalool without accumulation of intermediary product, (*E*)-8-hydroxylinalool. The products of candidate enzymes in plant tissues therefore corroborate their *in vitro* activity, in particular the (*E*)-8-carboxylinalool synthase activity of CYP76F14.

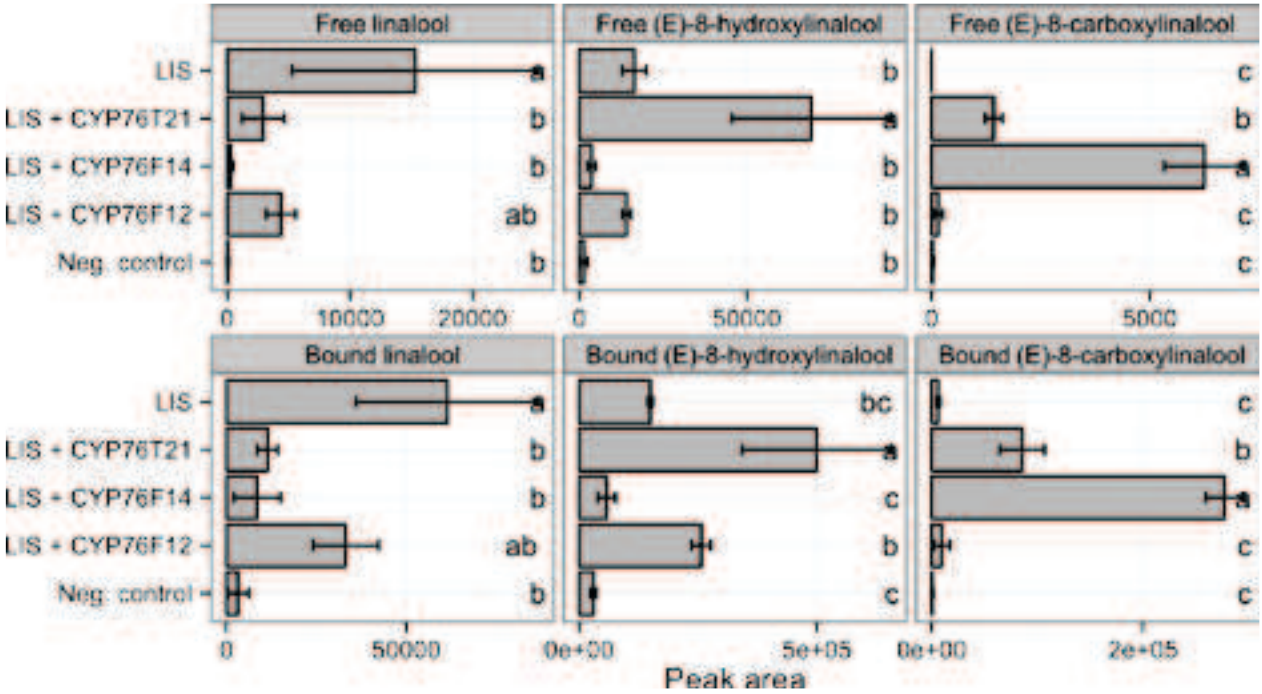

Figure 33. Quantification of oxidized linalool metabolites in *N. benthamiana* leaves expressing grapevine linalool synthase and candidate P450 genes. *Agrobacterium tumefaciens* containing candidate P450 genes was infiltrated to *N. benthamiana* leaves Expression of green fluorescent protein (GFP) was used as a negative control. Amounts of metabolites are expressed as means of peak areas in the corresponding MRM channel (± standard deviation, N=3). Free and bound forms of the same metabolite were detected with the same channel so that peak areas are roughly comparable. One-way analysis of variance (ANOVA) revealed significant differences between groups ($P < 0.05$ for linalool, $P < 0.01$ for malonylhexosyl linalool and $P < 0.001$ for other four metabolites). Letters on the right axes indicate significant differences between means calculated with Tukey HSD test ($\alpha = 0.05$).

Gene polymorphisms between grapevine varieties

Due to the limited availability of plant material we used two different *V. vinifera* varieties in our experiments, Gewurztraminer (Gw) for the analysis of gene expression and (*E*)-8-carboxylinalool content, and Muscat Ottonel (Mo) for functional gene characterization. To explore the nucleotide polymorphisms among the two varieties, we isolated and sequenced the candidate genes from both of them and compared them to the PN40024 reference genome (**Appendix II**). *CYP76F14* was the gene with the most polymorphisms. Sequences in Gewurztraminer and in PN40024 were identical, but differed from the Muscat Ottonel sequence (12 SNP, 6 of which are nonsynonymous). The latter variant was previously named *CYP76F2* (34), but it is most likely an allelic variant of *CYP76F14* and not a distinct locus.

To test the effect of this polymorphism on CYP76F14 enzymatic activity, the Gewurztraminer and Muscat Ottonel enzyme activities were compared. Interestingly, we did not find any difference in product profiles, neither *in vitro* (Figure S15) nor *in planta* (Figure S16). The amino acid differences observed between the two variants thus do not seem to alter the biochemical activity of the two enzymes.

Discussion

Wine lactone forms in wine from (*E*)-8-carboxylinalool and its glucose ester

To gain better insight into the formation of wine lactone in wine, we measured the content of (*E*)-8-carboxylinalool, (*E*)-8-carboxylinalool glucose ester and wine lactone in 23 white wine samples. Concentrations of (*E*)-8-carboxylinalool in wines were in micromolar range, comparable to the concentrations of other linalool derivatives (**Chapter 1**). Concentrations of wine lactone, on the other hand, were much lower—in the nanomolar range—but higher than those reported previously (4, 9), which could be due to the fact that we profiled older wines. Wine lactone was shown to form from (*E*)-8-carboxylinalool via a slow non-enzymatic cyclisation (11). Wedler *et al.* (35) recently challenged this unusual mechanism and based of a quantum chemistry approach suggested an alternative process involving oxidation of cyclic terpenes such as limonene or α-terpineol. We demonstrate here that the concentration of wine lactone correlates with both wine age and (*E*)-8-carboxylinalool concentration, which is consistent with the formation from (*E*)-8-carboxylinalool. In addition, oxygenated limonene or α-terpineol derivatives have to our knowledge never been detected in wines. This does not, however, exclude the possibility that wine lactone may result from limonene oxidation in plant tissues where limonene is the main terpene constituent, such as *Citrus* fruits. The formation of wine lactone from (*E*)-8-carboxylinalool in wine follows first-order reaction kinetics with an extremely low reaction rate constant $k = 0.00012$ year^{-1}. This slow reaction rate is in agreement with the high activation energy required for the unlikely hydride shift proposed in the reaction mechanism (11). The concentration of wine lactone can thus be estimated from the wine age and the easily measurable (*E*)-8-carboxylinalool concentration, circumventing the tedious sample preparation for analysis of wine lactone.

All (*E*)-8-carboxylinalool is present as a glucose ester in grape berries. In young wines it is present in both free and conjugated forms, suggesting some of (*E*)-8-carboxylinalool might be released by the yeast during fermentation (14). In older wines the content of (*E*)-8-carboxylinalool glucose ester is lower whereas the content of free (*E*)-8-carboxylinalool is higher.

We found large differences in wine lactone contents between different wines, stemming from differences in wine age and (*E*)-8-carboxylinalool content. Of the three wine varieties analyzed, the highest concentrations were measured in Gewurztraminer wines, which supports the role of wine lactone as Gewurztraminer varietal aroma compound (4). Our data also support the hypothesis that this is due to the higher concentration of the wine lactone precursor, (*E*)-8-carboxylinalool, in Gewurztraminer wines. In addition, preliminary data shows that the content of (*E*)-8-carboxylinalool glucose ester is higher in berries of

Gewurztraminer compared to the wild grapevine *Vitis vinifera* subsp. *sylvestris* (Figure S17). This could suggest that wine lactone is an aromatic trait selected during grapevine domestication.

The biosynthesis of (*E*)-8-carboxylinalool in grapevine

The wine lactone precursor, (*E*)-8-carboxylinalool, accumulates in ripening grape berries. We chose a set of candidate genes potentially involved in the biosynthesis of (*E*)-8-carboxylinalool in grape based on their preferential expression in ripening berries. Together, the activities of recombinant enzymes and the pathway reconstruction in *N. benthamiana* demonstrate that a terpene synthase (LIS) and a cytochrome P450 (CYP76F14) are sufficient to efficiently produce (*E*)-8-carboxylinalool from common plant precursors. We cannot, however, exclude the involvement of other redundant enzymes in the grape berries. Grapevine genome contains 69 terpene synthase genes, of which at least 7 have linalool synthase activity in vitro (20). Within the set of investigated P450 candidates expressed in ripening grape berries, two other enzymes, CYP76F12 and CYP76T21, could catalyze some of the oxidation steps that lead to the conversion of linalool into (*E*)-8-carboxylinalool, and may thus also participate in the production of production of the wine lactone precursor (Figure 1).

Linalool and its first oxygenated product, (*E*)-8-hydroxylinalool, accumulate in grapes as glycosylated derivatives (19), formation of which requires the activity of UDP glycosyltransferases (UGTs). UGTs compete with P450 enzymes for available linalool and (*E*)-8-hydroxylinalool product. They can therefore prevent further oxidation and the formation of the final product. It is noteworthy that *N. benthamiana* leaves expressing CYP76F14 accumulated only the final product, (*E*)-8-carboxylinalool, and no glycosylated linalool or (*E*)-8-hydroxylinalool. CYP76F14 is thus efficient enough to outcompete UGTs and is efficiently coupled with LIS for the channeling of linalool. This was not the case for the other P450 candidates. *N. bentamiana* leaves co-expressing CYP76T21 or CYP76F12 accumulated significant amounts of glycosylated linalool or (*E*)-8-hydroxylinalool, and less (*E*)-8-carboxylinalool compared to CYP76F14. The high efficiency of CYP76F14 in (*E*)-8-carboxylinalool biosynthesis, together with its high expression level in ripening berries, thus support the role of CYP76F14 as the main (*E*)-8-carboxylinalool synthase in grapes. It is also noteworthy that CYP76F14 and the other monoterpenol oxidases identified in this work may influence wine aroma by depleting powerful odorants such as linalool and geraniol from the terpenoid pool. CYP76F14 and other berry-expressed P450s could thus be used as markers for the selection of grapevine varieties producing desired amounts of monoterpenols and the wine lactone precursor.

Wine lactone has 3 chiral centers, resulting in 8 possible stereoisomers. Only one of them, the (3*S*,3a*S*,7a*R*) form, is a powerful odorant and was until recently the only stereoisomer

detected in wines (5). Interestingly, it was shown to be the cyclisation product (3*R*, 6*E*)-8-carboxylinalool, whereas the main linalool enantiomer in wine is (3*S*)-linalool (18). Recent work confirmed that the predominant configuration of (*E*)-8-carboxylinalool is also (3*S*), although the ratio between enantiomers varied between samples (14). Consistent with presence of both (*E*)-8-carboxylinalool enantiomers the same investigation also revealed the presence of two enantiomers of wine lactone, the major one being the less fragrant (3*R*,3a*R*,7a*S*) enantiomer. CYP76F14 is not stereoselective and equally metabolizes both linalool enantiomers to produce both (*R*) and (*S*)-carboxylinalool (Figure S18).

(*E*)-8-hydroxylinalool is an intermediate in the biosynthesis of (*E*)-8-carboxylinalool and a product of three cytochromes P450 investigated in this study. These enzymes, CYP76F12, CYP76F14 and CYP76T21 thus likely synthesize (*E*)-8-hydroxylinalool in grapes, where this compound accumulates in glycosylated form (36). Strauss *et al.* showed that (*E*)-8-hydroxylinalool can act as a precursor to dill ether, another potent odorant, in acidic solutions (36). Despite non-negligible amounts of its precursor, dill ether has (to our knowledge) never been detected neither in grapes nor in wines. We could confirm that (*E*)-8-hydroxylinalool cyclizes to dill ether in acidic buffer, but we were also unable to detect dill ether in any of the wines we analyzed. This could be related to poor ionisation, low concentration, slow formation, or the combination of these factors. A more sensitive targeted analytical approach could give us more information on concentrations of this compound in wines and reveal new roles of CYP76 enzymes in wine aroma.

Other oxygenated linalool derivatives in grapes are 6-hydroxylinalool, 7-hydroxylinalool, (*Z*)-8-hydroxylinalool, furanic and pyranic linalool oxides. None of the CYP76 enzymes included in this study could synthesize these compounds, with the exception of CYP76F12, which could produce (*Z*)-8-hydroxylinalool. While our findings confirm the role of the CYP76 family in oxidation of monoterpenols (22, 23), they also hint at other P450 families potentially involved in oxidations at non-terminal positions.

We showed that (*E*)-8-carboxylinalool accumulates in grape berries in the form of glucose ester. (*E*)-8-carboxylinalool was not included as a substrate in the recently reported characterization of grapevine glycosyltransferases (37, 38), so the enzyme catalyzing this reaction remains to be discovered. Profiling of wines showed that hydrolysis of this glucose ester occurs both during fermentation and during wine maturation. The role of yeast in this process remains to be explored, but could be an important factor affecting the wine lactone formation and more generally wine aroma.

Material and methods

Synthesis of monoterpenoids

(*E*)-8-hydroxylinalool, (*E*)-8-oxolinalool and (*E*)-8-carboxylinalool were synthesized and purified as described previously (22, 24). Enantiomerically pure wine lactone was synthesized according to the procedure of Chavan et al. (39) with the following modifications: hydroboration was accomplished following the procedure of Bode and Carreira (40) and separation of the diastereomers was carried out after the dehydrohalogenation step, contrary to recommendation in (39). Wine lactone displayed identical spectroscopic data to those reported previously.

Wine samples

Wines were purchased from Paul Ginglinger (Eguisheim, France) or produced by the INRA Colmar. One sample (Riesling Weingut 2014) was purchased at the supermarket. The INRA Gewurztraminer wines were made from grapes collected from the same vineyard localized in the "Grand Cru Osterberg" area in Ribeauvillé (France) in 1986, 1997, 1998, 2001, 2004, 2005, 2007 and 2009. All wines were stored in the INRA cellar, which is maintained at the constant temperature. For the analysis of (*E*)-8-carboxylinalool and (*E*)-8-carboxylinalool glucose ester, wines were injected directly to LC-MS/MS without being extracted or concentrated. For the analysis of wine lactone, 40 mL of wine was spiked with internal standard ((*E*)-8-oxolinalool), then extracted with 10 mL pentane: ethyl acetate (1:1 vol.). The extract was then evaporated to dryness and the residue resuspended in 200 µL of methanol prior to LC-MS/MS analysis.

Liquid chromatography coupled to tandem mass spectrometry

Liquid chromatography-mass spectrometry analyses were performed on a Waters Quattro Premier XE mass spectrometer equipped with an electrospray ionization source and coupled to an Acquity UPLC system (Waters USA). Chromatographic separation was achieved using an Acquity UPLC bridged ethyl hybrid C18 column (100 · 3 · 2.1 mm, 1.7 µm; Waters) and precolumn. The mobile phase consisted in (A) water and (B) methanol, both containing 0.1% formic acid. The solvent gradient was as follows: 95% A 0–2 min, linear gradient to 0% A 2–12 min, 0% A 12–14 min, linear gradient to 95% A 14–15 min, 95% A 15–17 min. The flow rate was 0.350 mL min^{-1}. The column was heated to 35°C. Injection volume was 5 µL for multiple reaction monitoring (MRM) runs and 3 µL for non-targeted (scan) runs. Nitrogen was used as the drying and nebulizing gas. The nebulizer gas flow was set to 50 L/h, and the desolvation gas flow to 900 L/h. The interface temperature was 400°C and the source temperature 135°C. The capillary voltage was set to 3.4 kV; for the scan runs the cone voltage was set to 25 V and the ionization was in positive or negative mode. Low mass and high mass

resolution was 15 for both mass analyzers, ion energies 1 and 2 were 0.5 V, entrance and exit potential were 50 V, and detector (multiplier) gain was 650 V. Parent and daughter ions used for MRM were: 137>80.7 for linalool, 135>106.8 for (*E*)-8-hydroxylinalool, 151.2>92.8 for (*E*)-8-oxolinalool, 167.2>92.8 for (*E*)-8-carboxylinalool and 167.2>93 for wine lactone. Glycosylated derivatives were putatively identified on the basis of their mass spectra (Table S3). Principal component analysis was performed in R with the package FactoMineR.

Selection of candidate genes

Cytochrome P450 genes were annotated in the PN40024 genome as described in **Chapter 5**. RNAseq data from Riesling and Gewurztraminer grape berries from two time points (before and after veraison) were aligned to these gene models. For further investigations, we chose the genes with increased expression in the later time point (after veraison).

Total RNA isolation from grape berries and qPCR

Grape berries were collected from the Gewurztraminer 643 clone from the ampelographic collection of INRA in Colmar, France, at five different developmental stages: green (4 July 2014), before and after veraison (8 August 2014), mid-ripe (8 September 2014) and ripe (29 September 2014). Berries were flash frozen in liquid nitrogen and stored at -80°C. They were crushed with a hammer to allow for the removal of seeds. Then they were ground using Qiagen Teflon grinding jar set and TissueLyser for 30 s at 30 s^{-1}.

Total RNA was isolated from 1 g of berry powder following the protocol in (30) with the following modifications: 7 mL of extraction buffer was used per g of berry powder and after LiCl precipitation the samples were treated as described above. The concentration of RNA, 260/280 and 260/230 ratios was measured by Nano-Drop 2000 spectrophotometer (Thermo Scientific) and integrity of RNA was confirmed by gel electrophoresis. cDNA was synthesized from 1 µg total RNA using $(dT)_{23}$ and Superscript® III Reverse Transcriptase (Life technologies) following manufacturer's instructions. The resulting cDNA was diluted 5x and used as a template in a quantitative PCR (qPCR) experiment.

Primers for qPCR were designed using the Primer3 plus website (Table S 2). Specificity of the primers was confirmed *in silico* using the Primer Blast website (NCBI) and by measuring the melting curve of the products. The reaction mixture contained 1x SYBR® Green Master Mix, 250 nM of forward and reverse primers and 50x diluted cDNA from reverse transcription reaction. The qPCR experiment was performed on Roche LightCycler® 480. The three reference genes, EF1-α (XM_002284928.3), UBQ_L40 (XM_002273532.2) and MDH (m) (XM_002278676.3), were selected from 6 constitutively expressed grapevine genes (30) using the GeNorm algorithm (41). Amplification efficiency were calculated for each set of

primers using LinRegPCR (42). Relative gene expression was calculated as described in (43). Each point was obtained from 3 technical and 3 biological replicates.

Amplification of coding sequences, construction of vectors and recombinant expression

RNA was extracted from grape skins of *V. vinifera* cv. Muscat Ottonel following the protocol in (30), except that PVPP in extraction buffer was replaced with PVP. After LiCl precipitation and ethanol wash, the nucleic acids were digested using RQ1 RNase-Free DNase (Promega, Cat.# M6101). The reaction mixture was diluted to 250 µL and extracted with equal volume of phenol: chloroform : isoamyl alcohol (25:24:1 vol.). The aqueous phase was extracted twice with an equal volume of chloroform. RNA was precipitated with ethanol and resuspended in 30 µL of MilliQ water. cDNA was synthesized using Superscript® III Reverse Transcriptase (Life technologies) following manufacturer's instructions, and was then used as a template for PCR amplification of the target sequences.

CYP76F14 and CYP76F12 PCR products were cloned into the pYeDP60 yeast expression vector using restriction enzymes BamHI and KpnI. The N-terminal segment of CYP76F14 was recoded taking into account yeast codon preference using a 150 nt forward primer. The PCR product was first inserted into pGem®-T Easy vector (Promega) before being transferred to pYeDP60 using the same restriction enzymes. CYP71AT7, CYP76T21 and CYP76Y2 were inserted into the yeast expression vector pYeDP60u2 using the USER™ cloning method (44). Several colonies obtained for each gene were verified by sequencing. Muscat Ottonel was used for sequence amplification and some sequence divergence with the reference genome (PN40024) was expected. In the cases where sequence divergences were consistent in several colonies, they were considered as relevant and the corresponding amplicons were expressed in yeast. The list of primers used for sequence amplification is provided in Table S 2. The GwCYP76F14 coding sequnce was synthesized de novo by Genecust with codon optimization for expression in yeast. CYP76F12, CYP76F14 and CYP76T21 were cloned into the pCAMBIA2300 vector with the USER™ cloning method. Linalool synthase is a Gewurztraminer allele of *Vitis vinifera* TPS56 (HM807392 PN55M1) and was cloned into vector pMDC32.

Heterologous expression in yeast

The *Saccharomyces cerevisiae* strain WAT11 was transformed with pYeDP60:CYP or "empty" pYeDP60 plasmid as described in (45). Yeast cultures were grown as described for the high density procedure in (46) with the following modifications. A 200 mL culture volume was inoculated with 15 mL of the overnight culture, and the gene expression was induced by the addition of 5 vol. % of 200 g L^{-1} galactose. The microsomal membranes were isolated as described in (47). The expression of P450 was evaluated by CO differential

spectroscopy of reduced microsomes (31) (Figure S7). The intensity of the peak at 450 nm was too low to allow for reliable quantification of enzymes.

P450 enzyme assays

In vitro P450 enzyme assays were conducted in 20 mM Na^+/K^+ phosphate buffer, pH 7.4, in the presence of 730 µM reduced NADPH, 100 µM of substrate and 10 % vol. of microsomal membrane suspension. Reactions were conducted at 27°C for 1 hour under agitation and were stopped by the addition of solvent (ethyl acetate for samples to be analyzed by GC or methanol for samples to be analyzed by LC), vortexed for 30 seconds and centrifuged to pellet the membranes and proteins. Ethyl acetate extracts were filtered over sodium sulfate to remove residual water. All assays were performed in triplicates.

Transient expression in *N. benthamiana*

N. benthamiana plants were grown under 16 h day (24°C) / 8 h night (20°C) conditions. The *Agrobacterium tumefaciens* strain LBA4404 was transformed with the pCambia2300 or pCambia3300 plasmids by electroporation. Liquid cultures of transformants (100 mL LB, 50 µM rifampicin, 10 µM gentamicin, 50 µM kanamycin) were grown overnight at 28°C (180 rpm agitation). The next day they were harvested by 10 min centrifugation at 3000 g and washed twice with tap water. Agrobacteria transformed with the p19 gene of tomato bushy stunt virus (48), linalool synthase and P450 genes were then mixed with tap water in the same tube in 1:1:1 ratios, so that the OD_{600} of each of the strains corresponded to 0.4. The bacteria were then infiltrated into the abaxial epidermis of 20 day-old *N. benthamiana* leaves with a 1 mL plastic syringe (without needle). For negative controls, strains for expression of either P450 or LIS were replaced with a GFP expression strain.

Leaves tissues were collected 4 days post-infiltration, frozen in liquid nitrogen, and ground with a metal bead (7 or 10 mm diameter) in 12 mL plastic tubes. Leaves expressing GFP were also viewed under UV light to confirm successful agroinfiltration and gene expression. Leaf powder was stored at -80°C and extracted with methanol (1:4 m/v) by vortexing. Extracts were then stored overnight at -20°C and centrifuged just prior to analysis. The supernatant was transferred into a glass vials for LC-MS/MS analysis.

Gas chromatography analysis

Ethyl acetate extracts were analyzed on a Varian 3900 gas chromatograph (Agilent Technologies) with flame ionization detector, equipped with a DB5 column (Agilent Technologies). Injector temperature was 250°C and injection was splitless. Temperature gradient was: 0.5 min at 50°C, 10°C/min to 320°C, and 5 min at 320°C. Temperature of the FID detector was 280°C.

Acknowledgements

We thank Gisèle Butterlin (INRA, Colmar) for help with grape berry sampling and RNA preparation and Grégory Lemarquis (INRA, Colmar) for providing Gewurztraminer wine samples.

Accession Numbers

Accession numbers to be added after cytochrome P450 annotation is submitted to public databases

Authors' contributions

Conceptualization, T.I., N.N., E.D., P.H. and D.W-R.; Investigation, T.I., L.K., A.I.; Chemical synthesis : L.M. and F.L. ; Formal Analysis, T.I.; Writing – Original Draft, T.I., N.N., E.D., P.H. and D.W-R.; Writing – Review & Editing, N.N., P.H. and D.W-R.; Visualization, T.I.; Supervision, N.N., E.D. P.H. and D.W-R.; Project administration, P.H. and D.W-R.; Funding Acquisition, D.W-R. and P.H.

References

1. International Organisation of Vine and Wine (2015) Available at: www.oiv.int/oiv/info/enstatoivextracts.
2. Harris RLN, Lacey MJ, Brown W V, Allen MS (1987) Determination of 2-methoxy-3-alkylpyrazines in wine by gas chromatography/mass spectrometry. *Vitis* 26:201–207.
3. Wood C, et al. (2008) From wine to pepper: rotundone, an obscure sesquiterpene, is a potent spicy aroma compound. *J Agric Food Chem* 56(10):3738–44.
4. Guth H (1997) Identification of character impact odorants of different white wine varieties. *J Agric Food Chem* 45(8):3022–3026.
5. Guth H (1996) Determination of the configuration of wine lactone. *Helv Chim Acta* 79:1559–1571.
6. Guth H (1997) Quantitation and Sensory Studies of Character Impact Odorants of Different White Wine Varieties. *J Agric Food Chem* 43:35–37.
7. Hinterholzer A, Schieberle P (1998) Identification of the most odour-active volatiles in fresh, hand-extracted juice of Valencia late oranges by odour dilution techniques. *Flavour Fragr J* 13(1):49–55.
8. Buettner A, Schieberle P (1999) Characterization of the most odor-active volatiles in fresh, hand-squeezed juice of grapefruit (Citrus paradisi Macfayden). *J Agric Food Chem* 47(12):5189–5193.
9. Guth H (1998) *Chemistry of Wine Flavor* eds Waterhouse AL, Ebeler SE.
10. Bonnländer B, Baderschneider B, Messerer M, Winterhalter P (1998) Isolation of two novel terpenoid glucose esters from Riesling wine. *J Agric Food Chem* 8561(642):1474–1478.
11. Luan F, Degenhardt A, Mosandl A, Wüst M (2006) Mechanism of wine lactone formation: demonstration of stereoselective cyclization and 1,3-hydride shift. *J Agric Food Chem* 54(26):10245–52.
12. Giaccio J, et al. (2011) The formation of wine lactone from grape-derived secondary metabolites. *J Agric Food Chem* 59(2):660–664.
13. Giaccio J, et al. (2011) The formation of wine lactone from grape-derived secondary metabolites. *J Agric Food Chem* 59(2):660–4.
14. Giaccio J, Curtin CD, Sefton MA, Taylor DK (2015) Relationship between menthiafolic acid and wine lactone in wine. *J Agric Food Chem* 63(37):8241–8246.
15. D'Ambrosio M, Harghel P, Guantieri V (2013) Isolation of intact glycosidic aroma precursors from grape juice by hydrophilic interaction liquid chromatography. *Aust J Grape Wine Res* 19(2):189–192.
16. Winterhalter P, Messerer M, Bonnländer B (1997) Isolation of the glucose ester of (E) -2,6-dimethyl-6-hydroxyocta-2,7-dienoic acid from Riesling wine. *Vitis* 36(1):55–56.

17. Williams PJ, Straws CR, Wilson B, Strauss CR, Wilson B (1980) Hydroxylated Linalool Derivatives as Precursors of Volatile Monoterpenes of Muscat Grapes. *J Agric Food Chem* 28(4):766–771.

18. Luan F, Hampel D, Mosandl A, Wüst M (2004) Enantioselective analysis of free and glycosidically bound monoterpene polyols in Vitis vinifera L. cvs. Morio Muscat and Muscat Ottonel: Evidence for an oxidative monoterpene metabolism in grapes. *J Agric Food Chem* 52(7):2036–2041.

19. Strauss CR, et al. (1988) Novel monoterpene disaccharide glycosides of Vitis vinifera grapes and wines. *J Agric Food Chem* 36(8):569–573.

20. Martin DM, et al. (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(1):226.

21. Dudareva N, Pichersky E, Gershenzon J (2004) Biochemistry of plant volatiles. *Plant Physiol* 135(4):1893–1902.

22. Ginglinger J-F, et al. (2013) Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in Arabidopsis flowers. *Plant Cell* 25(11):4640–57.

23. Hofer R, et al. (2014) Dual function of the cytochrome P450 CYP76 family from Arabidopsis thaliana in the metabolism of monoterpenols and phenylurea herbicides. *Plant Physiol* 166(3):1149–1161.

24. Boachon B, et al. (2015) CYP76C1 (Cytochrome P450)-Mediated Linalool Metabolism and the Formation of Volatile and Soluble Linalool Oxides in Arabidopsis Flowers: A Strategy for Defense against Floral Antagonists. *Plant Cell* 27(10):2972–90.

25. Collu G, et al. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett* 508(2):215–220.

26. Wilson B, Strauss CR, Williams PJ (1984) Changes in free and glycosidically bound monoterpenes in developing Muscat grapes. *J Agric Food Chem* 32(4):919–924.

27. Martin DM, Chiang A, Lund ST, 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(3):919–929.

28. Nelson DR, Ming R, Alam M, Schuler MA (2008) Comparison of cytochrome P450 genes from six plant genomes. *Trop Plant Biol* 1(3-4):216–235.

29. Fasoli M, et al. (2012) The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* 24(9):3489–3505.

30. Reid KE, Olsson N, Schlosser J, Peng F, Lund ST (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biol* 11:1–11.

31. Omura T, Sato R (1964) The carbon monoxide-biding pigment of liver microsomes. *J Biol Chem* 239(7):2370–2378.

32. Höfer R, et al. (2013) Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco)iridoid pathway. *Metab Eng* 20:221–232.

33. Miettinen K, et al. (2014) The seco-iridoid pathway from Catharanthus roseus. *Nat Commun* 5:3606.

34. Nelson DR (2009) The cytochrome P450 homepage. *Hum Genomics* 4(1):59–65.

35. Wedler H, Pemberton R, Tantillo D (2015) Carbocations and the complex flavor and bouquet of wine: mechanistic aspects of terpene biosynthesis in wine grapes. *Molecules* 20(6):10781–10792.

36. Strauss CR, Wilson B, Williams PJ (1988) Novel monoterpene diols and diol glycosides in Vitis vinifera grapes. *J Agric Food Chem* 36(3):569–573.

37. Bönisch F, et al. (2014) Activity-Based Profiling of a Physiologic Aglycone Library Reveals Sugar Acceptor Promiscuity of Family 1 UDP-Glucosyltransferases from Grape. *Plant Physiol* 166(1):23–39.

38. Bönisch F, et al. (2014) A UDP-glucose:monoterpenol glucosyltransferase adds to the chemical diversity of the grapevine metabolome. *Plant Physiol* 165(2):561–581.

39. Chavan SP, Kharul RK, Sharma AK, Chavan SP (2001) An efficient and simple synthesis of (-)-wine lactone. *Tetrahedron Asymmetry* 12(21):2985–2988.

40. Bode JW, Carreira EM (2009) Stereoselective synthesis of erythronolide A via nitrile oxide cycloadditions and related studies. *J Org Chem* 74(22):8695–8712.

41. Vandesompele J, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric

averaging of multiple internal control genes. *Genome Biol* 3(7):1–12.

42. Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 339(1):62–66.

43. Hilliou F, Tran T (2013) RqPCRAnalysis: Analysis of Quantitative Real-time PCR Data. *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms* (SciTePress - Science and and Technology Publications), pp 202–211.

44. Nour-Eldin HH, Hansen BG, Norholm MHH, Jensen JK, Halkier BA (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res* 34(18):e122–e122.

45. Gietz D, St Jean A, Woods R a, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* 20(6):1425.

46. Pompon D, Louerat B, Bronine A, Urban P (1996) Yeast expression of animal and plant P450s in optimized redox environments. *Methods Enzymol* 272(Ii):51–64.

47. Gavira C, et al. (2013) Challenges and pitfalls of P450-dependent (+)-valencene bioconversion by Saccharomyces cerevisiae. *Metab Eng* 18:25–35.

48. Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* 33(5):949–956.

49. Winter D, et al. (2007) An "electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2(8):e718.

Supplementary information

Table S6. Primers used for cloning.

Gene	Forward primer	Reverse primer
CYP71AT7u	GGCTTAAUATGATGATTTTGCTTCTTATCCTTTTAGCTCTC	GGTTTAAUTTAATGACTAGGGATCCTCGCCAAAAG
CYP76F2	ATAGGATCCATGGAGCTTCTATCTTGTCTTCTTTGCTTTTTAGCAGCCTGGACATCGATTTACATCATGTTCTCAGCCAGAAGGGGAAGGAAACATGCGGCTCATAAATTACCACCAGGACCAGTACCCTTACCAATAATAGGAAGCCTCTTAAACCTG	GGTACCTCAAACCCGTACAGGTAGAGCTTG
CYP76F2u	GGCTTAAUATGGAGTTGTTGAGTTGTCTGCTG	GGTTTAAUTCAAACCCGTACAGGTAGAGCTTG
CYP76F12	ATAGGATCCATGGAGATGTTGAGCTGTCTGCTG	TAGGTACCTTACTTCGGCACACAAGAGCAACAAC
CYP76F12u	GGCTTAAUATGGAGATGTTGAGCTGTCTGCTG	GGTTTAAUTTACTTCGGCACACAAGAGCAACAAC
CYP76T21u	GGCTTAAUATGGATTACACCCCACTTGTTCTTC	GGTTTAAUTCATGGTTTGGTGGGAACAGC
CYP76Y2u	GGCTTAAUATGGAACTTAACACCTTCCTCTTGC	GGTTTAAUTTAATTTTCTTCGATAAATGGAATTGCAGAAAG

Table S7. Primers used for qPCR experiments.

Gene	Forward primer	Reverse primer
EF1-α (m)	CTGCTTGCTTTCACCCTTGG	TCATCGTACCTTGCCTTGGAG
UBQ-L40	GGTTCTGAGGCTTCGTGGTGG	TGCAGTTGACAGCACGTGGG
MDH (m)	GTTCATACAATGTTCCAGCAGGGC	CCTGTGCGGTCAAGTCCAACTTC
CYP71AT7	ACCAGAGCCTGAGAAAGAAGACATC	AGCTGCGCTTGTATCTGTCCC
CYP76F12	AGATACTCTTCTCAACATCAGCGAGG	TGCCATTGCCCATTCCAAGGTG
CYP76F14	AGCTAGCAGTGATGTGTTAGACGTTC	GTTGTGTCAGTCCCCGCAGC
CYP76T21	CCGTTAAAAGCTTCTTCTCCGG	GAAATGGTGTCTGTTCCCGC
CYP76Y1	GAGCCTCCAATCCAGTGACC	TGGTAGATGTATCACTTCCAGC
CYP76Y2	ACCATTTGGAGCAGGTAGGC	TGGTCAGTGGAGTAGTTCCT

Table S8. Putative peak annotation in *N. bentamiana* extracts.

Name	RT /min	Authentic standard	Mr	Detected m/z - positive mode												Detected m/z - negative mode		
				$[M-H]^-$	$[M+Na]^+$	$[M+K]^+$	$[M+NH_4]^+$	$[M+H-H_2O]^+$	$[M+H-2H_2O]^+$	$[M+H-CO]^+$	$[M+H-C_5H_8O_6]^+$	$[M+H-C_6H_{10}O_5-H_2O]^+$	$[M+H-C_6H_{10}O_5-CO]^+$	$[M+H-C_6H_{10}O_5-CO-H_2O]^+$	$[M+H-C_6H_{10}O_5-C_2H_4O_2]^+$	$[M-H]^-$	$[M-H-C_6H_{10}O_5-H_2O]^-$	$[M-H-HCOOH]^-$
Linalool	10.7	YES	154					137										
Malonyl hexosyl linalool	9.8	NO	402		425	441	420								317			
(*E*)-8-hydroxylinalool	8.0	YES	170		193	209		153	135	107								
Hexosyl hydroxylinalool	7.3	NO	332	333	355	371	349	315	297		153	135						
(*E*)-8-carboxylinalool	8.2	YES	186					167								183		
Carboxylinalool glucose ester	6.9	NO	348	347	369	385	364	329	311		167	149	139	121			183	393

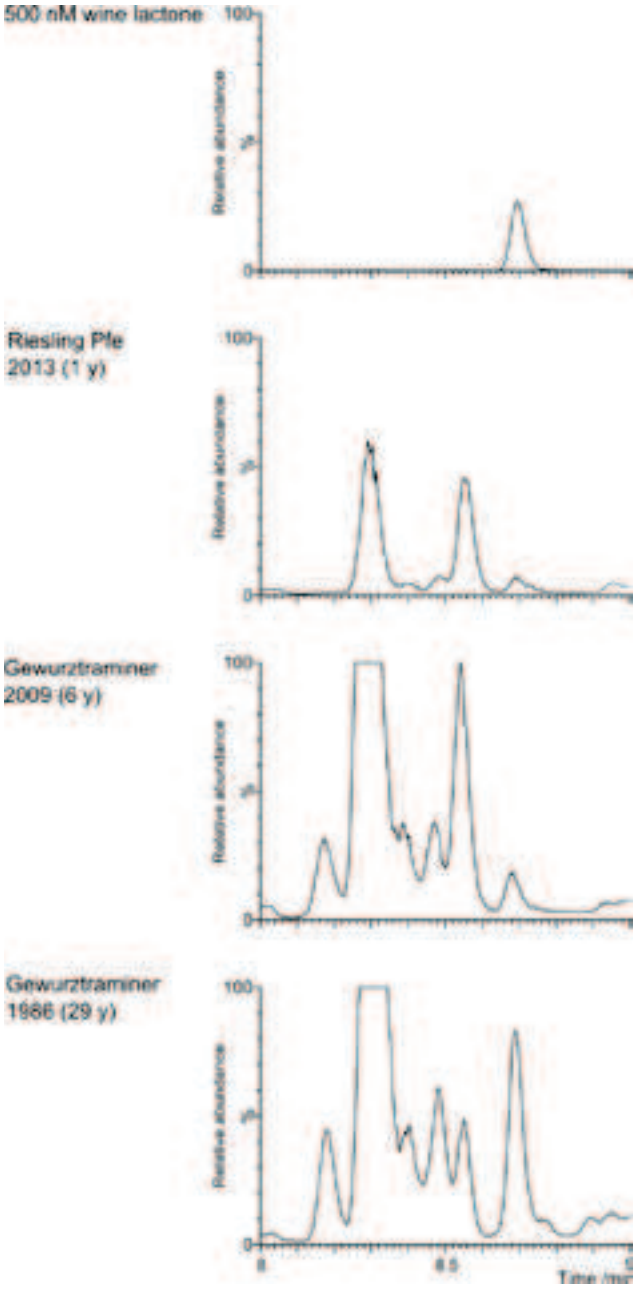

Figure S6. Targeted LC-MS/MS quantification of wine lactone in wine extracts. Specific MS/MS transition (167.3 >92.8 m/z) was used for wine lactone detection.

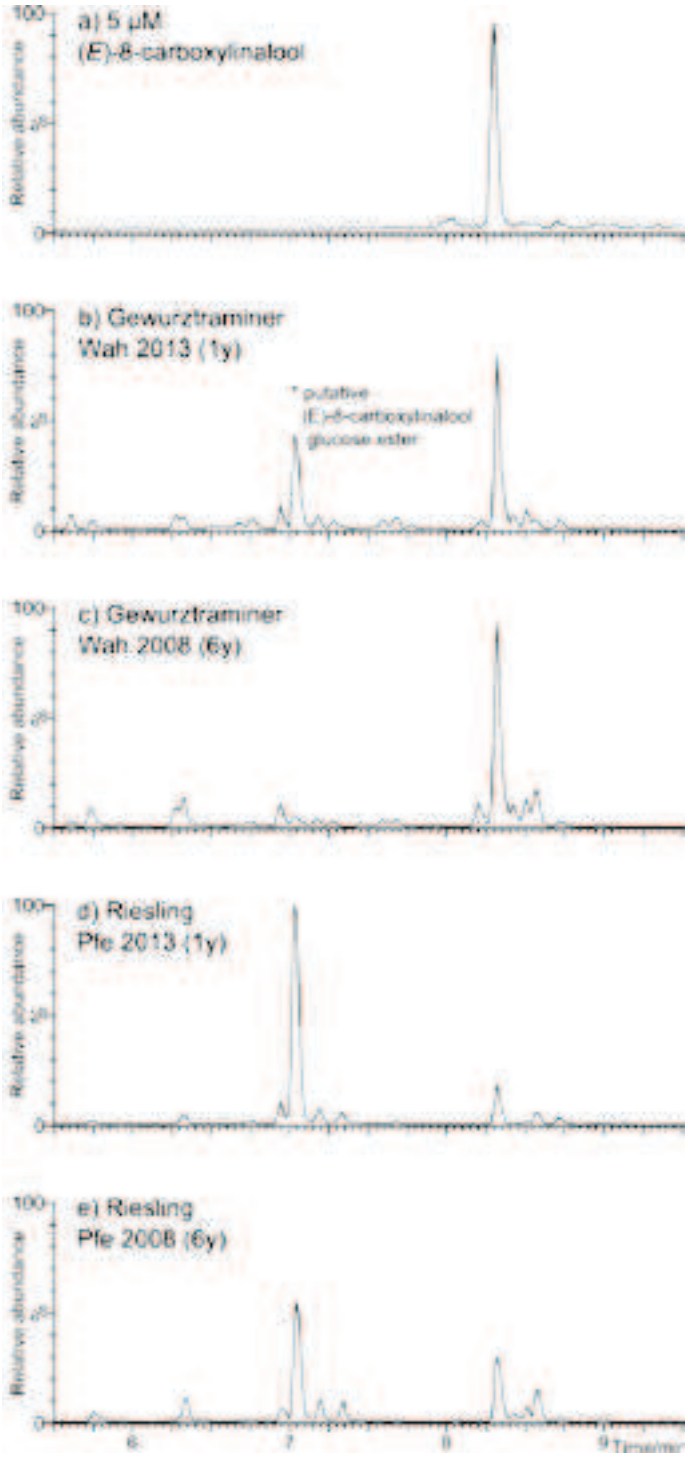

Figure S7. LC-MS/MS detection of (*E*)-8-carboxylinalool and its glucose ester in wines. Specific MS/MS transition was used for detection of (E)-8-carboxylinalool (167.3 >93 m/z). Peak at 7 min was putatively identified as (*E*)-8-carboxylinalool glucose ester on the basis of its mass spectrum (Table S3).

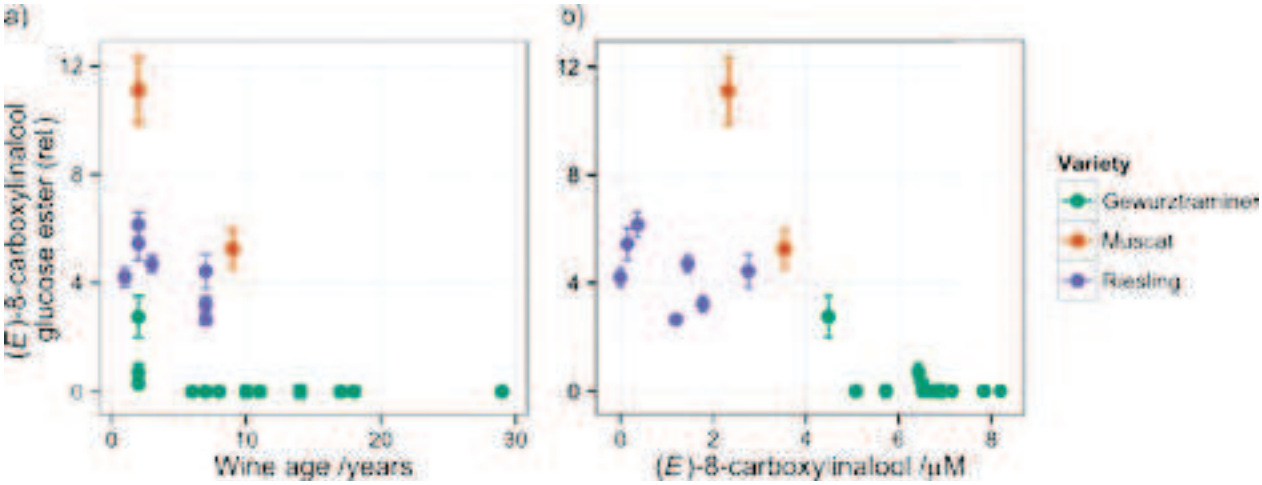

Figure S8. Relationship between (*E*)-8-carboxylinalool glucose ester and wine age (a) and free (*E*)-8-carboxylinalool (b) in 23 wine samples. Both compounds were quantified by LC-MS/MS and are expressed as means (±SD, N=3). Due to unavailability of (*E*)-8-carboxylinalool glucose ester standard the concentration of this compound is expressed in (*E*)-8-carboxylinalool equivalents.

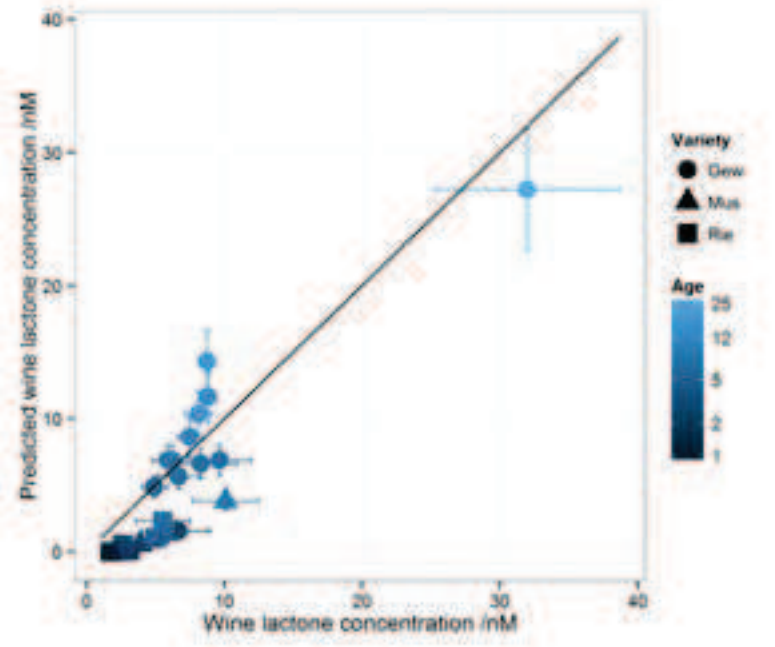

Figure S9. Predicted vs. real concentrations of wine lactone. Wine lactone concentrations were measured by LC-MS/MS and are expressed as means of technical replicates (±SD, N=3). Observed values were fitted to the function [wine lactone] = (1 - exp(-*k**age))*[(E)-8-carboxylinalool], using SD^{-2} as weights. Value of kinetic constant is k = (0.00012±0.00002) $year^{-1}$ (p-value<0.001).

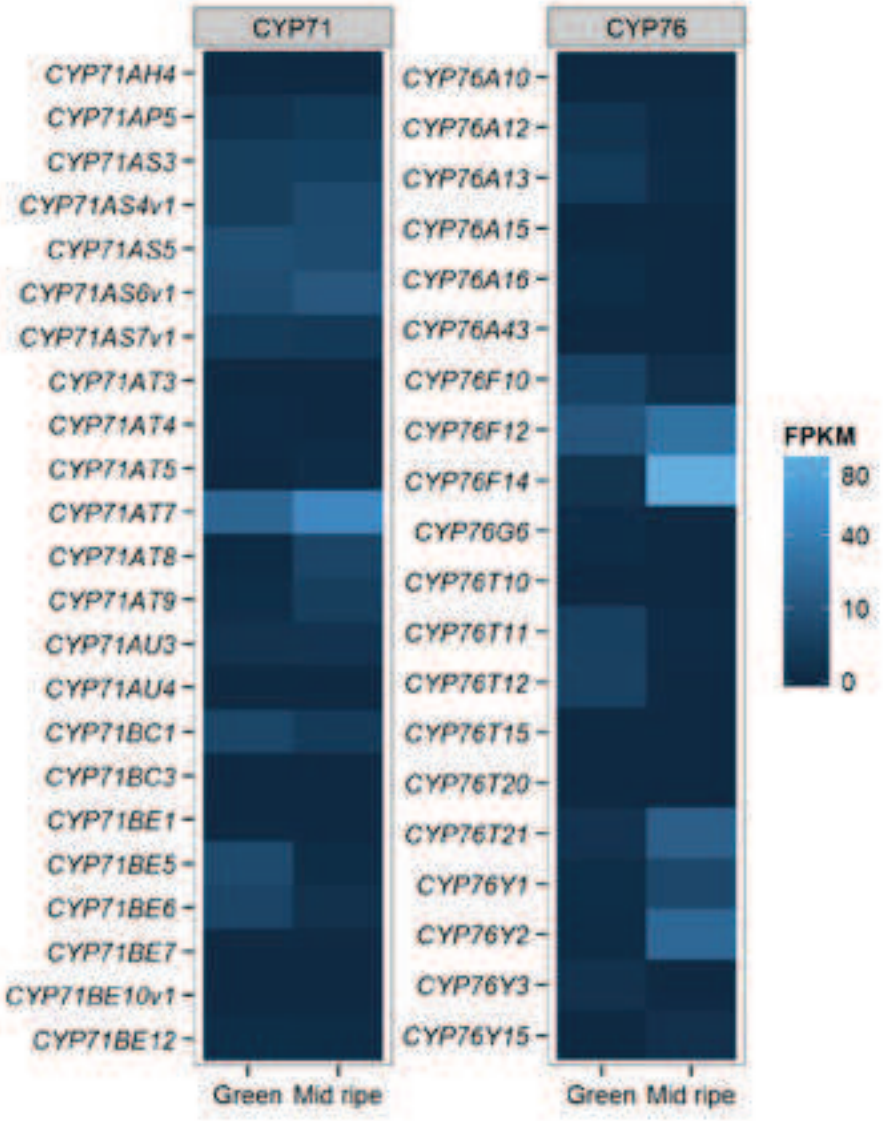

Figure S10. Expression of CYP71 (left panel) and CYP76 (right panel) genes at two different ripeness stages of Gewurztraminer berries. Gene expression is given in fragments per kilobase of transcript per million mapped reads (FPKM).

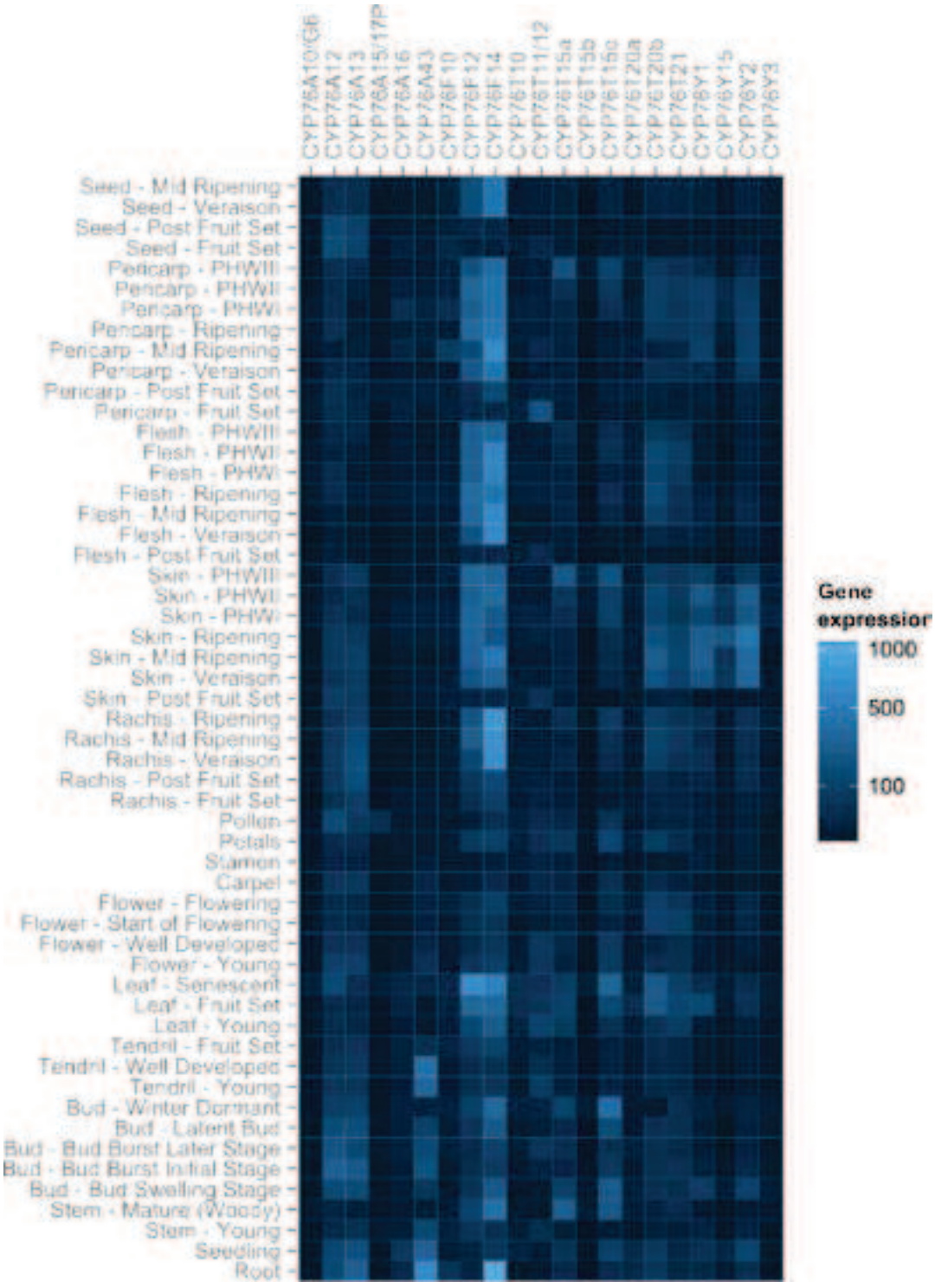

Figure S11. Expression of CYP76 genes in different organs and developmental stages of *Vitis vinifera* cv. *Corvina*. Normalized "grapevine expression atlas" data (29) was retrieved from Grape eFP Browser (49). Because of errors in gene annotation used for this dataset some pairs of adjacent genes were annotated as one gene.

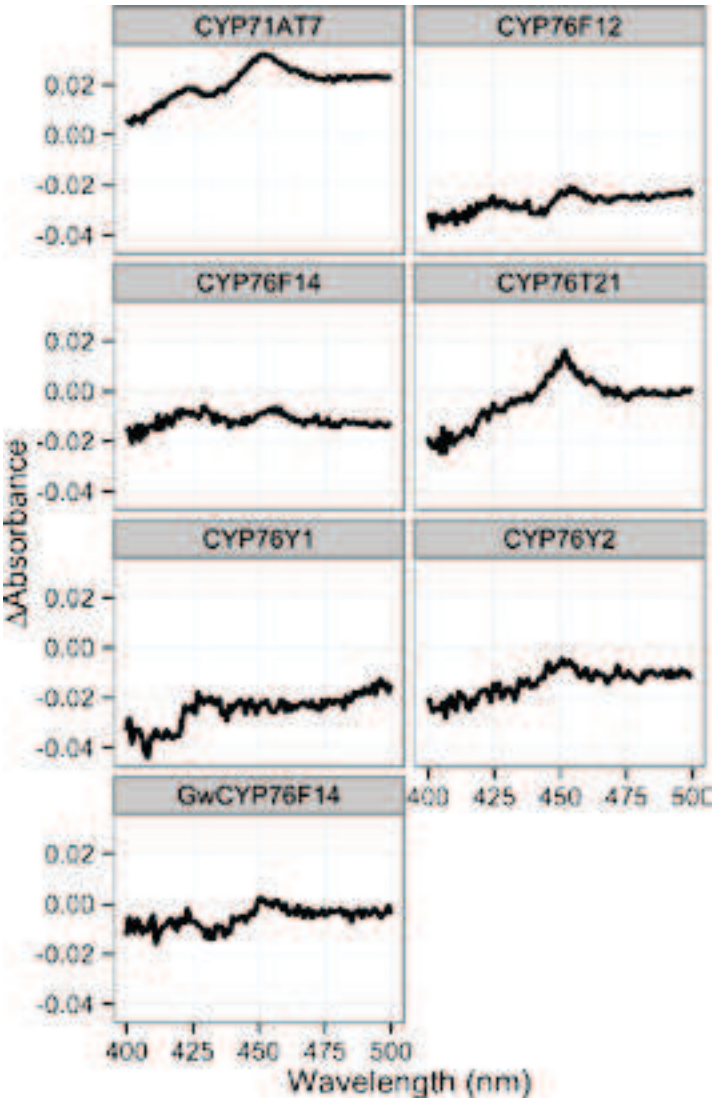

Figure S12. CO difference spectra of reduced microsomal membranes isolated from yeast expressing the P450 candidates. A peak at 450 nm indicates expression of cytochrome P450 (31). GwCYP76F14 is the variant of CYP76F14 cloned from Gewurztraminer cDNA.

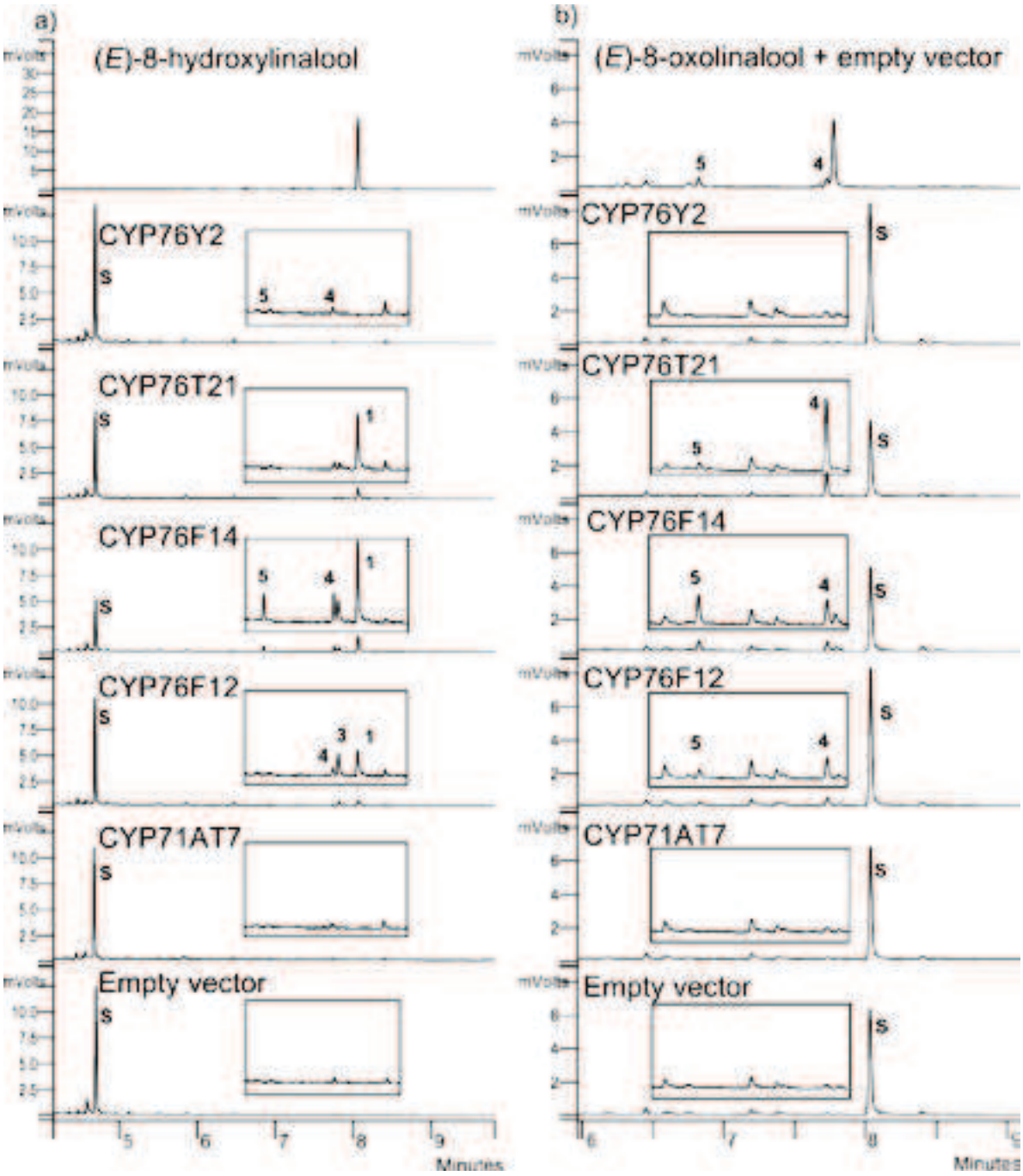

Figure S13. GC-FID profiles of the linalool (a) and (*E*)-8-hydroxylinalool (b) conversion products by the yeast-expressed grape P450s. The reactions were carried out using recombinant yeast microsomal membranes. Microsomal membranes from yeast transformed with an empty vector were used as a negative control. The upper panels show authentic standards of the main reaction products: (*E*)-8-hydroxylinalool (left panel) and (*E*)-8-oxolinalool (right panel). Peak annotation: **1**–(*E*)-8-hydroxylinalool, **3**–(*Z*)-8-hydroxylinalool, **4**–(*E*)-8-oxo-6,7-dihydrolinalool, **5**–(*E*)-8-hydroxy-6,7-dihydrolinalool. **4** and **5** are side products of yeast metabolism of (*E*)-8-oxolinalool. 3 and (*E*)-8-oxolinalool coelute so the annotation is supported by LC-MS/MS profiling (**Figure 5**).

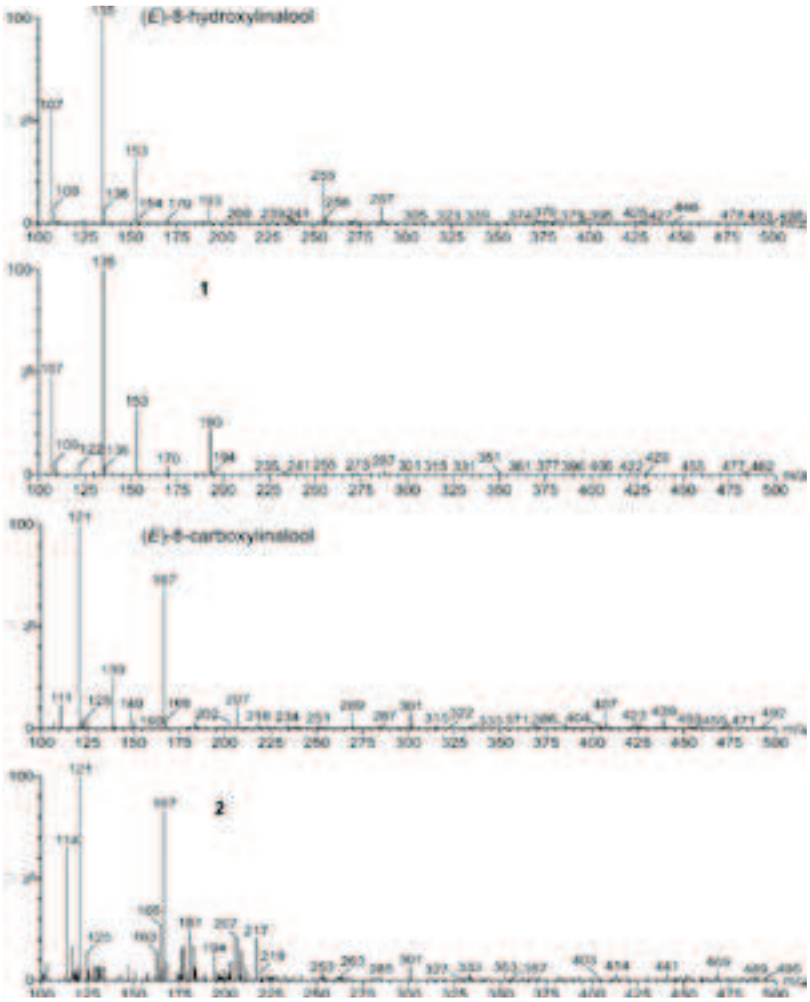

Figure S14. Mass spectra (ESI-MS) of the products identified by comparison with authentic standards.. Peak numbers refer to those in Figure 5.

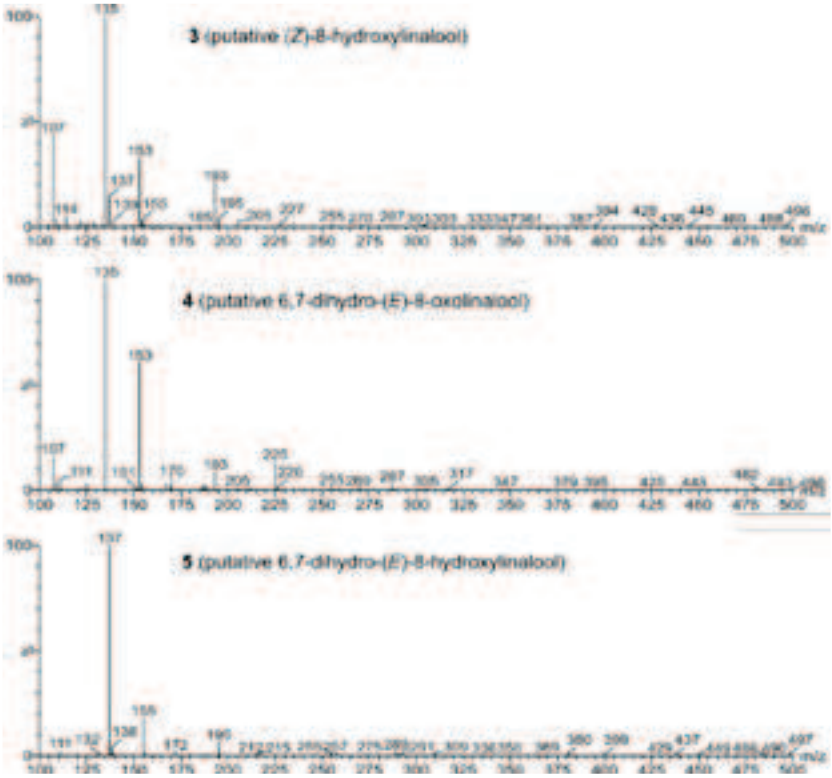

Figure S15. Mass spectra (ESI-MS) of the products putatively identified based on MS fragmentation. Peak numbers refer to those in Figure 5. **3** was identified on the basis of the similarity of its mass spectrum to that of **1**. **4** and **5** were identified in (24).

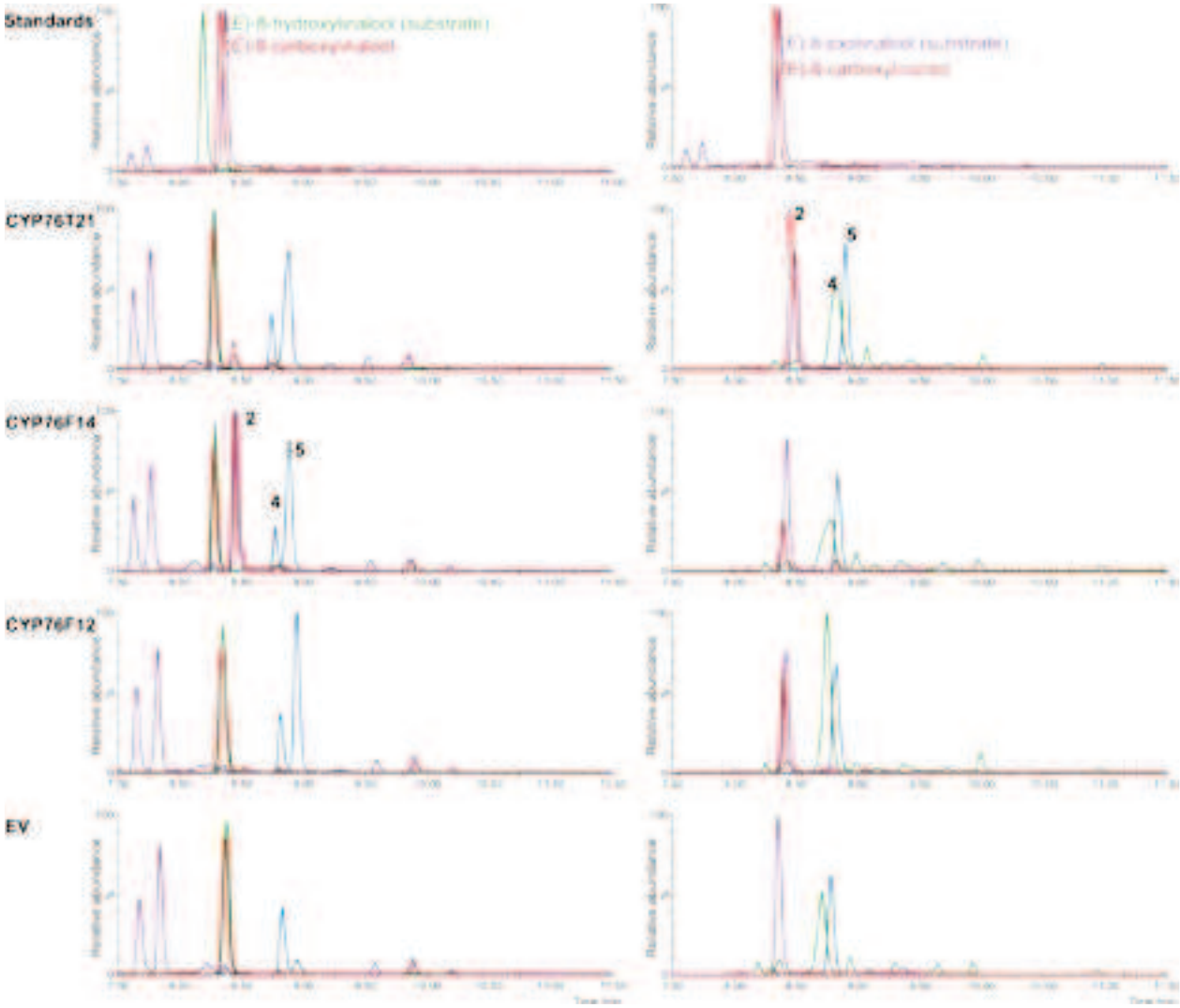

Figure S16. UPLC-MS analysis of the products of (*E*)-8-hydroxylinalool (left) and (*E*)-8-oxolinalool (right) metabolism by the yeast-expressed candidate P450scatalyzed the whole cascade of linalool oxidations. LC-MS/MS profiles of the products of three candidates enzymes in multiple-reaction-monitoring (MRM) mode. Each color represents an MRM channel for specific detection of target molecules: blue–linalool (137>80.7), green–(*E*)-8-hydroxylinalool (135>106.8), purple–(*E*)-8-oxolinalool (151.2>92.8) and red–(*E*)-8-carboxylinalool (167.2>92.8). Peak numbers refer to those in Figure 5.

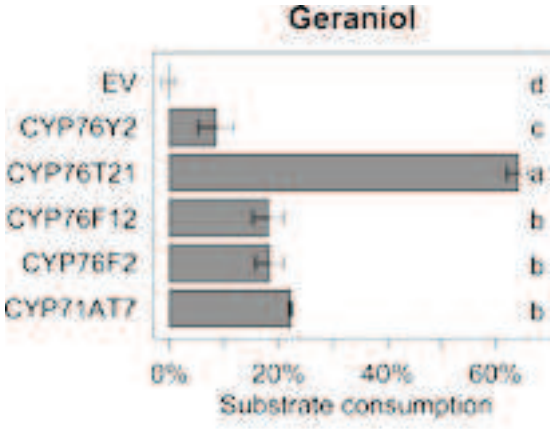

Figure S17. Conversion of geraniol by grapevine cytochromes P450. The experimental procedure and the evaluation of activity are the same as those used in Figure 4.

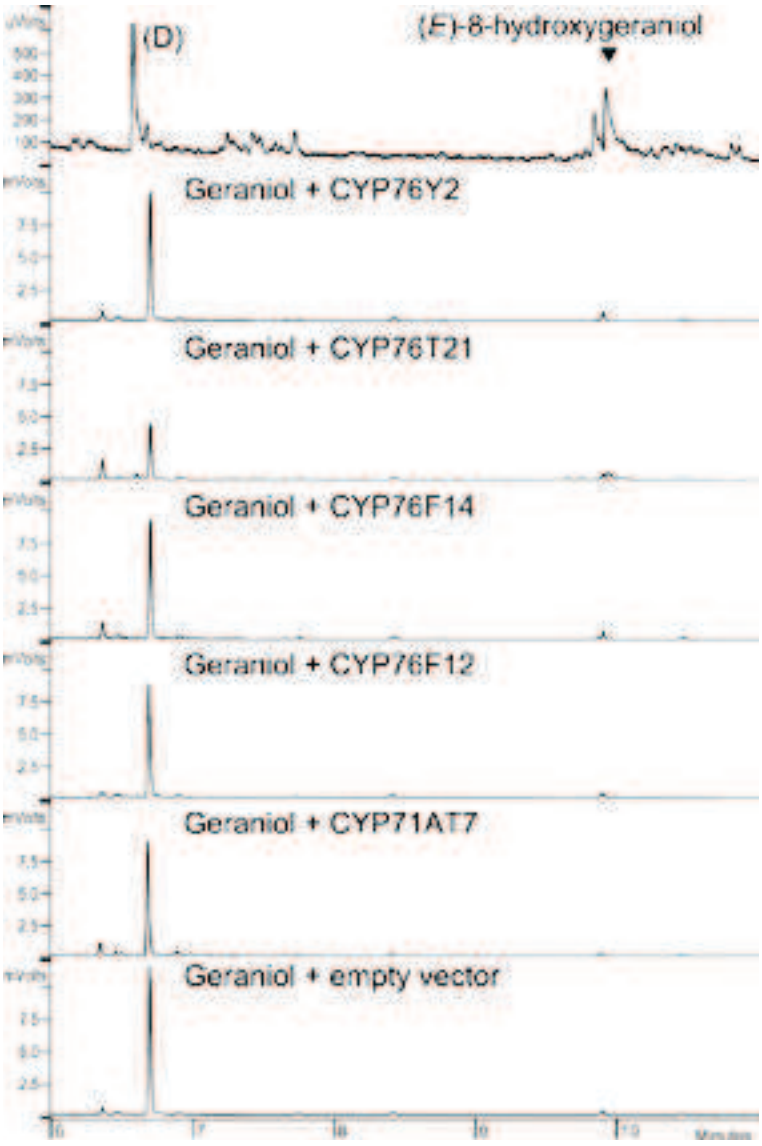

Figure S18. GC-FID analysis of the products of geraniol conversion by the yeast-expressed candidate enzymes. Reaction conditions are those used in Figure S11. D – a degradation product of (*E*)-8-hydroxygeraniol.

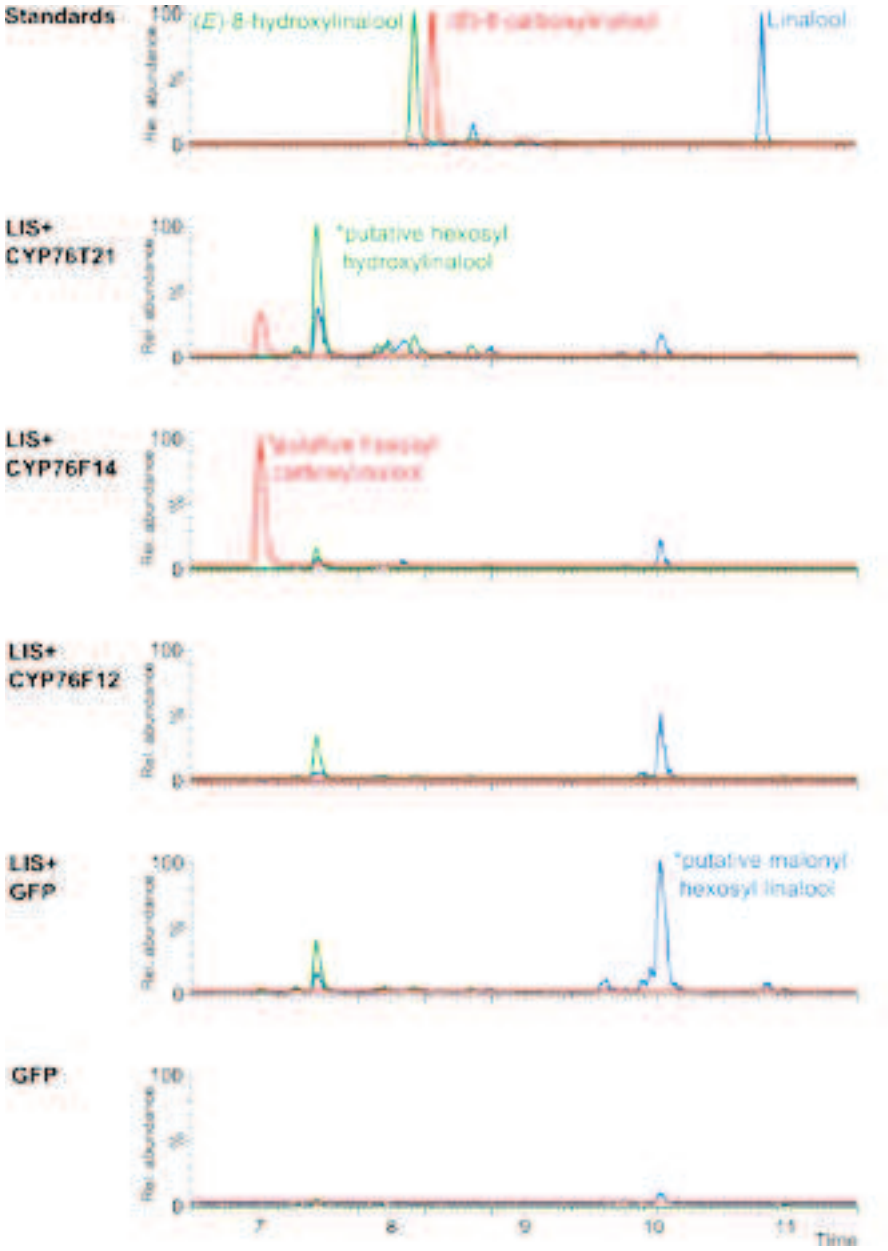

Figure S19. UPLC-MS-MS targeted analysis of the *N. benthamiana* leaves expressing the grapevine LIS and one of P450 candidates. Plant transformed with a GFP expression vector are used as a negative control. Each color represents an MRM channel for specific detection of target molecules: blue–linalool (137>80.7), green–(*E*)-8-hydroxylinalool (135>106.8) and red–(*E*)-8-carboxylinalool (167.2>92.8). Conjugated products were putatively identified on the basis of their mass spectra (Table S3).

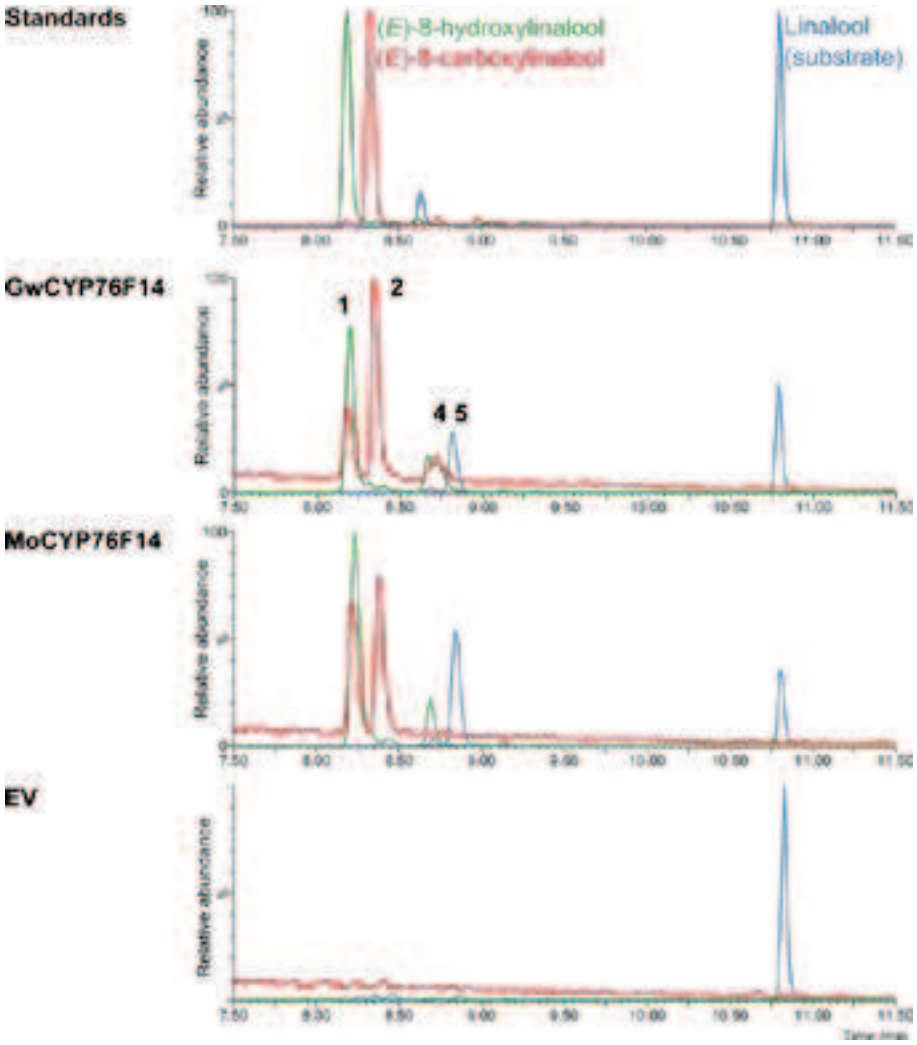

Figure S20. Comparison of the *in vitro* activities of Gewurztraminer and Muscat Ottonel CYP76F14. UPLC-MS/MS analysis of the reaction products of three enzyme candidates in multiple-reaction-monitoring (MRM) mode. Each color represents one MRM channel for the specific detection of target molecules: blue–linalool (137>80.7), green–(*E*)-8-hydroxylinalool (135>106.8) and red–(*E*)-8-carboxylinalool (167.2>92.8). Peaks 4 and 5 were putatively identified on the basis of their mass spectra (Figure S 9).

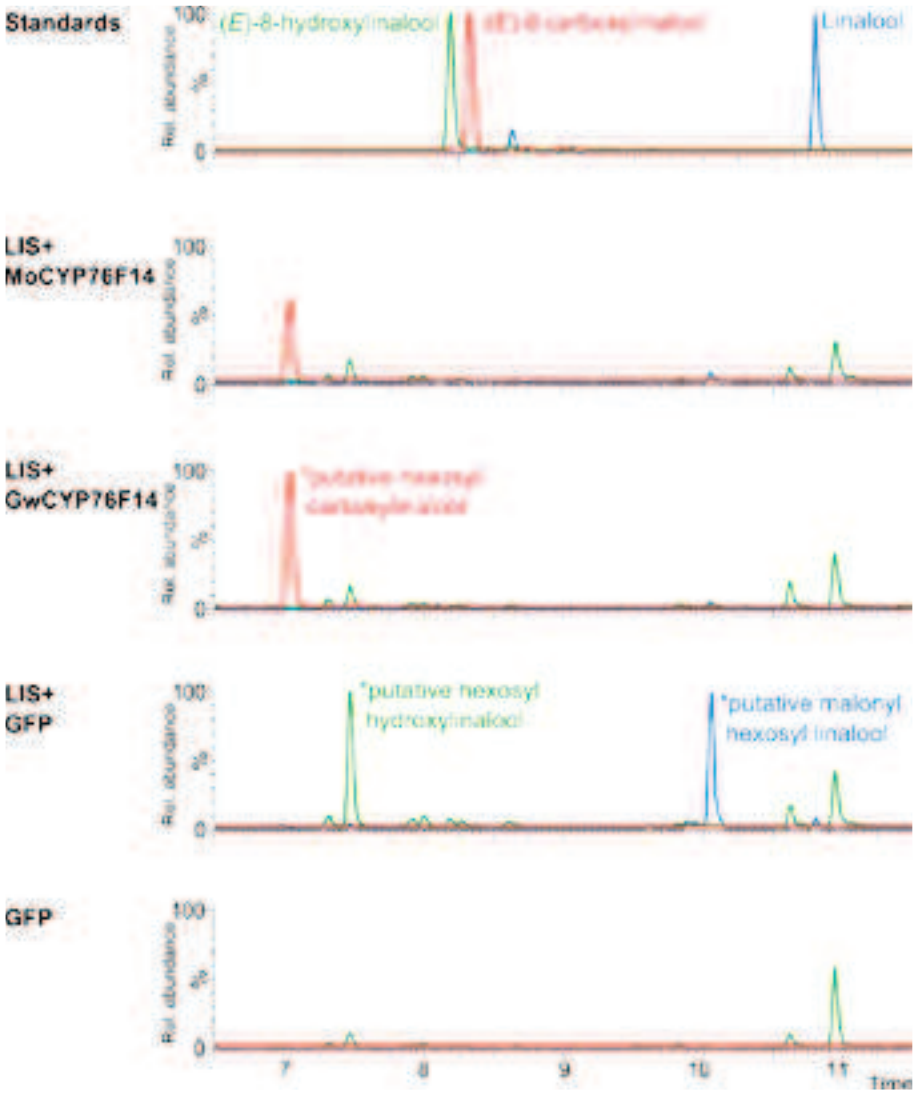

Figure S21. Comparison of the activities of Gewurztraminer and Muscat Ottonel CYP76F14 in *N. benthamiana*. Each color represents one MRM channel for specific detection of target molecules: blue–linalool (137>80.7), green–(*E*)-8-hydroxylinalool (135>106.8) and red–(*E*)-8-carboxylinalool (167.2>92.8). Conjugated products were putatively identified on the basis of their mass spectra (Table S3).

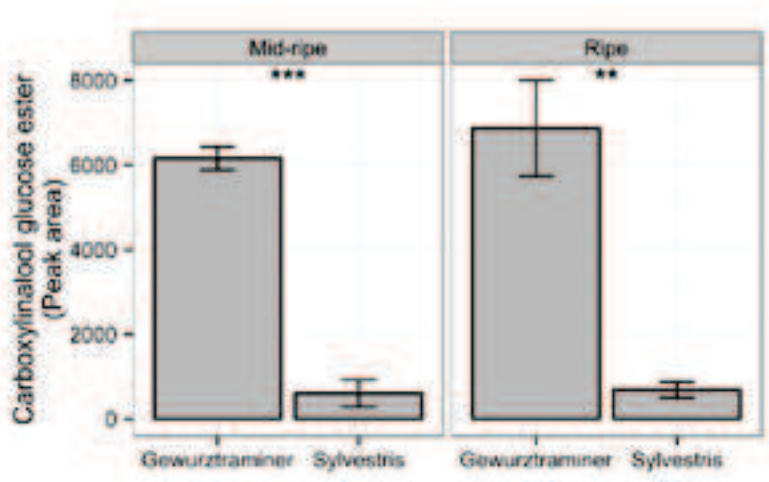

Figure S22. Comparison of (*E*)-8-carboxylinalool glucose ester concentration in berries of wild (*Vitis vinifera* subsp. *sylvestris*) and cultivated grapevine (*Vitis vinifera* subsp. *sylvestris* cv. Gewurztraminer). (*E*)-8-carboxylinalool glucose ester was quantified in methanol extracts of grape berries via LC-MS/MS and is expressed as a mean (±SD, N=3). Data were normalized relative to fresh weight. Student's t-test revealed significant differences between the two subspecies (p-value<0.001 ***, <0.001 **).

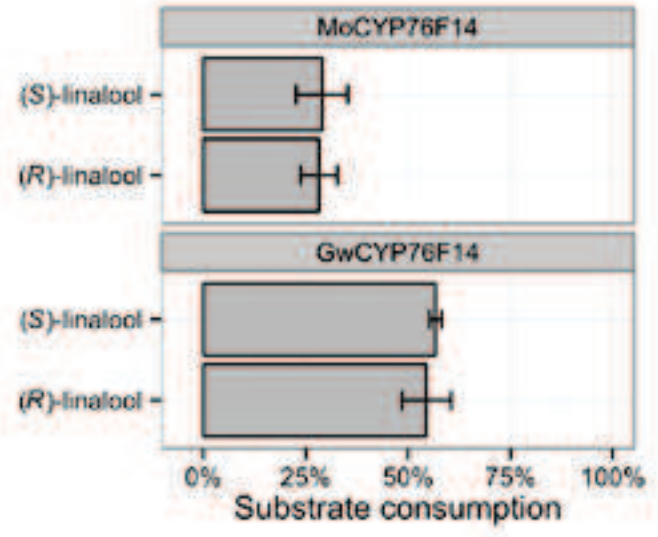

Figure S23. Conversion of two linalool enantiomers by CYP76F14 expressed in yeast. (Mo= Muscat Ottonel and Gw = Gewurztraminer). Sample preparation is identical to those used to generate data in Figure 4. Conversion (± SD, N=3) is calculated relative to the residual concentration of substrate in negative control (microsomal fraction from yeast transformed with an empty vector - EV) incubated under the same conditions. Student's t-test showed no significant difference between the two substrates at α=0.05 for neither of the enzyme variants.

Chapter 7
Additional results

Characterization of gene candidates using *Nicotiana benthamiana* leaf discs

Some plant cytochrome P450 genes express poorly in *S. cerevisiae* which impedes their functional characterization. In particular, this is the case for many members of the CYP76 family (1, 2). For functional characterization of such genes transient expression in *N. benthamiana* leaves represents an attractive alternative (2, 3). When the enzyme substrate is not present in the leaves endogenously, it can be vacuum infiltrated into the leaf discs expressing the gene to be characterized. With this method we tested three candidate P450s (CYP76Y2, CYP76T21 and CYP76F14) with four grape monoterpenols: linalool, geraniol, nerol and citronellol.

We expressed the candidate P450s in *N. benthamiana* leaves as described in the **Chapter 6**, except that they were not co-expressed with a terpene synthase. Instead, we cut discs from the leaves expressing the P450s, vacuum infiltrated them with a buffer solution containing the substrate (20 mM K^+/Na^+ phosphate buffer pH 7.4, 400 µM substrate). Ten discs of 1.4 cm diameter were incubated in 10 mL substrate solution for 4 h. The buffer solution was then collected and extracted using solid phase extraction cartridges (Oasis HLB 3cc with 60 mg sorbent). Ethyl acetate eluent was concentrated under argon and analyzed using GC-FID (for quantification) and GC-MS (for product identification). GC-FID conditions were identical to those used in **Chapter 6.** GC-MS was performed on PerkinElmer Clarus 680/680T system fitted with a HP5ms column (30 m, 0.25 mm, 0.25 µm; Agilent technologies). Temperature program was identical to the one used on GC-FID. Injector temperature was 250°C, ionization energy was 70 eV and m/z was collected in the 50-300 range.

Both CYP76T21 and CYP76F14 metabolized linalool, and the main product was in both cases (*E*)-8-hydroxylinalool (Figure 1a, Figure 2a). A second product of CYP76F14 was identified as a downstream product, (*E*)-8-oxolinalool (Figure 1a, Figure 2a). Although the non-volatile fraction was not analyzed in this experiment, it is informative to compare this experiment to the co-expression of linalool synthase and P450s in *N. benthamiana* leaves (**Chapter 6**). In the latter case, CYP76F14 converted all the linalool to (*E*)-8-carboxylinalool without accumulation of intermediary products. But when substrate was fed to the leaf discs expressing CYP76F14, two intermediary products, (*E*)-8-hydroxy- and (*E*)-8-oxolinalool, were accumulating. A possible reason for this difference is a high substrate concentration

used in the leaf disc experiment. If the affinity for linalool is higher compared to the downstream products, the enzyme is saturated with the first substrate and unable to perform the subsequent oxidation steps.

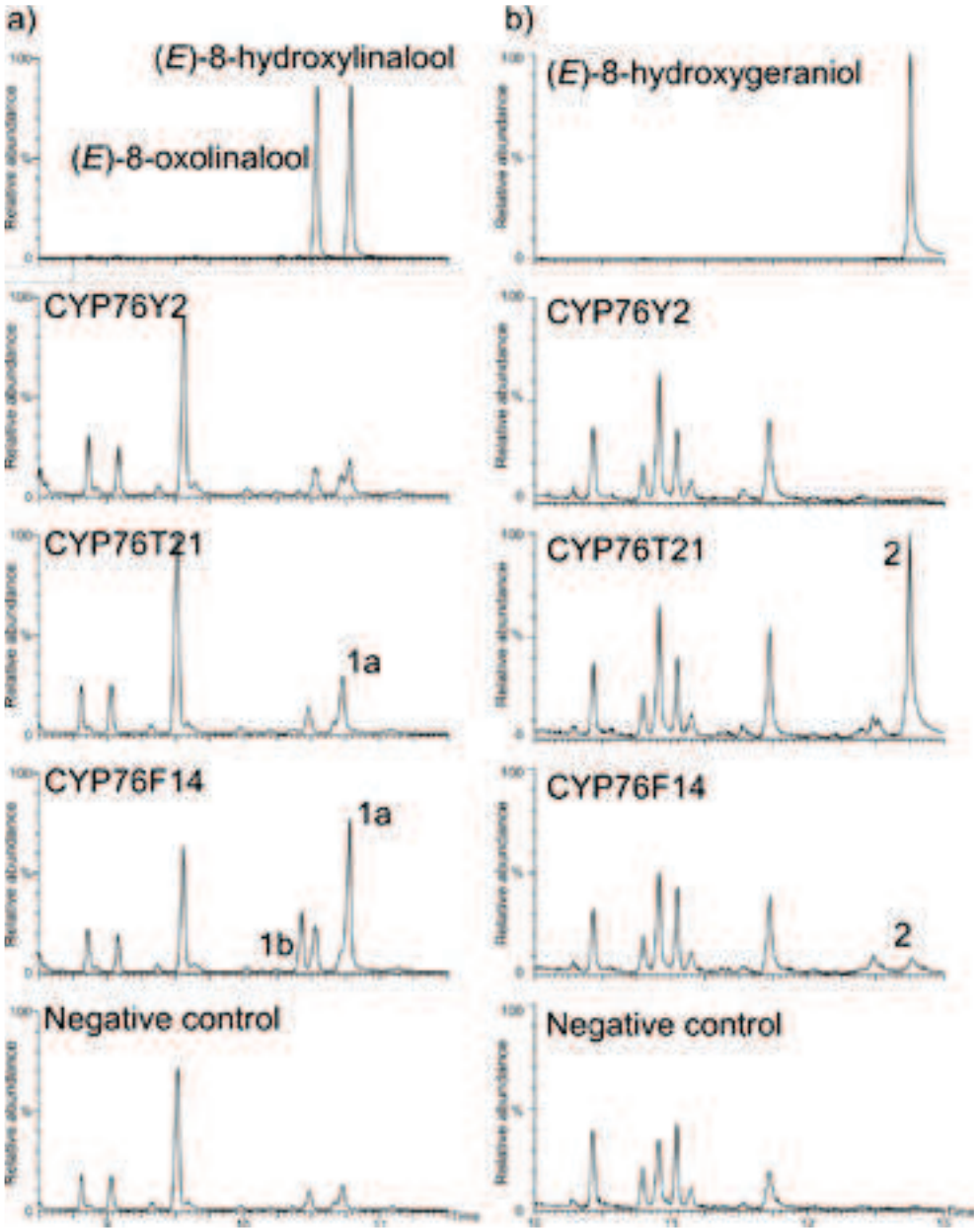

Figure 34. GC-MS profiles of *N. benthamiana* leaf discs expressing grapevine P450s after vacuum infiltration of substrates linalool (a) and geraniol (b) compared to authentic standards. *N. benthamiana* leaves were infiltrated with *Agrobacterium tumefaciens* bearing P450 genes. GFP was used as a negative control. Leaf discs were incubated in a substrate solution for 4 h. Buffer containing substrate and products was extracted using solid phase extraction cartridges and analyzed using GC-MS. Chromatograms show relative abundance of selected m/z 71 (a) and 68 (b). **1a** was identified as (*E*)-8-hydroxylinalool, **1b** as (*E*)-8-oxolinalool and **2** as (*E*)-8-hydroxygeraniol. Corresponding mass spectra are displayed in **Figure 2**.

The same two enzymes, CYP76F14 and CYP76T21, also metabolized geraniol in the leaf disc experiment. The main product was in both cases identified as (*E*)-8-hydroxygeraniol (Figure 1b, Figure 2b). CYP76T21 produced more of the (*E*)-8-hydroxygeraniol, whereas CYP76F14 was the most efficient at linalool oxidation. These results are in agreement with *in vitro*

experiments, where CYP76F14 was the most efficient at linalool oxidation and CYP76T21 was the most efficient enzyme at geraniol oxidation.

Leaf discs expressing CYP76F14 and CYP76T21 also metabolized the two other monoterpenols assayed: nerol and citronellol. The products of both enzymes were the same and were putatively identified as (*E*)-8-hydroxynerol and (*E*)-8-hydroxycitronellol (Figure 3, Figure 4). Comparison of relative amounts of products produced by different enzymes (Figure 5) revealed that while both enzymes, CYP76F14 and CYP76T21, produced comparable amounts of (*E*)-8-hydroxynerol and (*E*)-8-hydroxycitronellol, they differed in their efficiency to hydroxylate the other two substrates: CYP76F14 produced more (*E*)-8-hydroxylinalool compared to CYP76T21, whereas CYP76T21 produced more (*E*)-8-hydroxygeraniol.

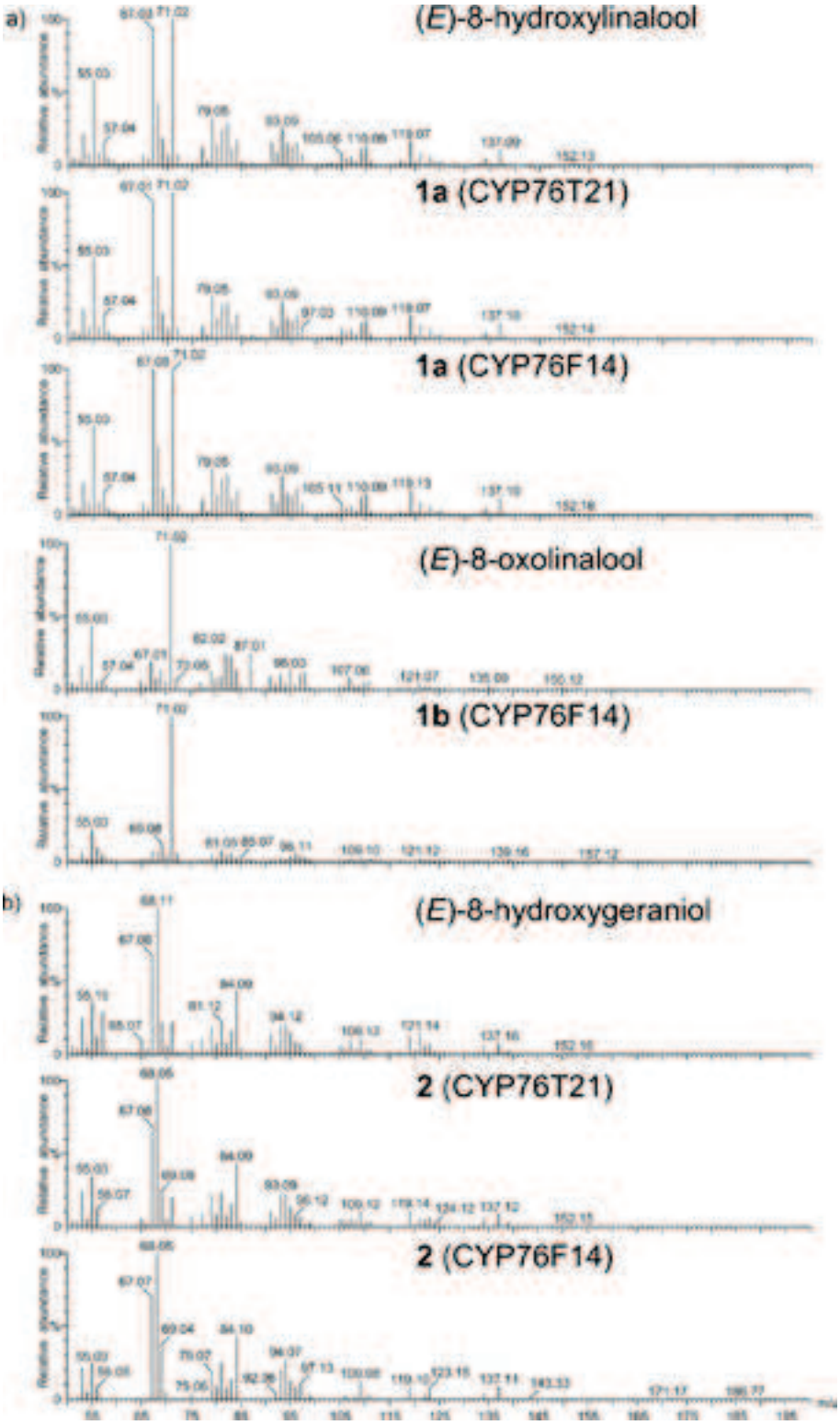

Figure 35. Mass spectra of CYP76F14 and CYP76T21 products formed from linalool (a) and geraniol (b) in *N. benthamiana* leaf discs (Figure 1).

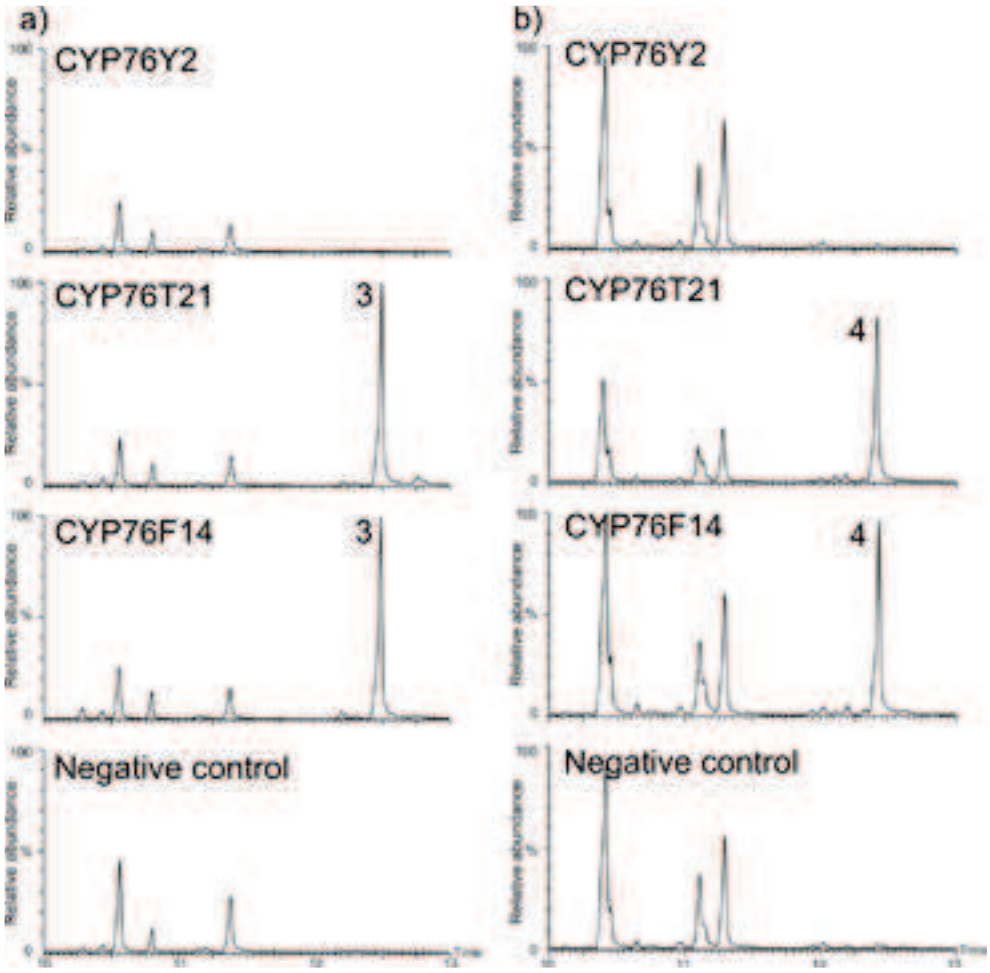

Figure 36. GC-MS profiles of *N. benthamiana* leaf discs expressing grapevine P450s after vacuum infiltration of substrates nerol (a) and citronellol (b). Samples were prepared as described in the legend of **Figure 1**. Chromatograms show the relative abundance of selected m/z 68 **(c)** and 55 **(d)**. Reaction products were tentatively identified on the basis of their mass spectra: **3** as (*E*)-8-hydroxynerol and **4** as (*E*)-8-hydroxycitronellol (**Figure 4**).

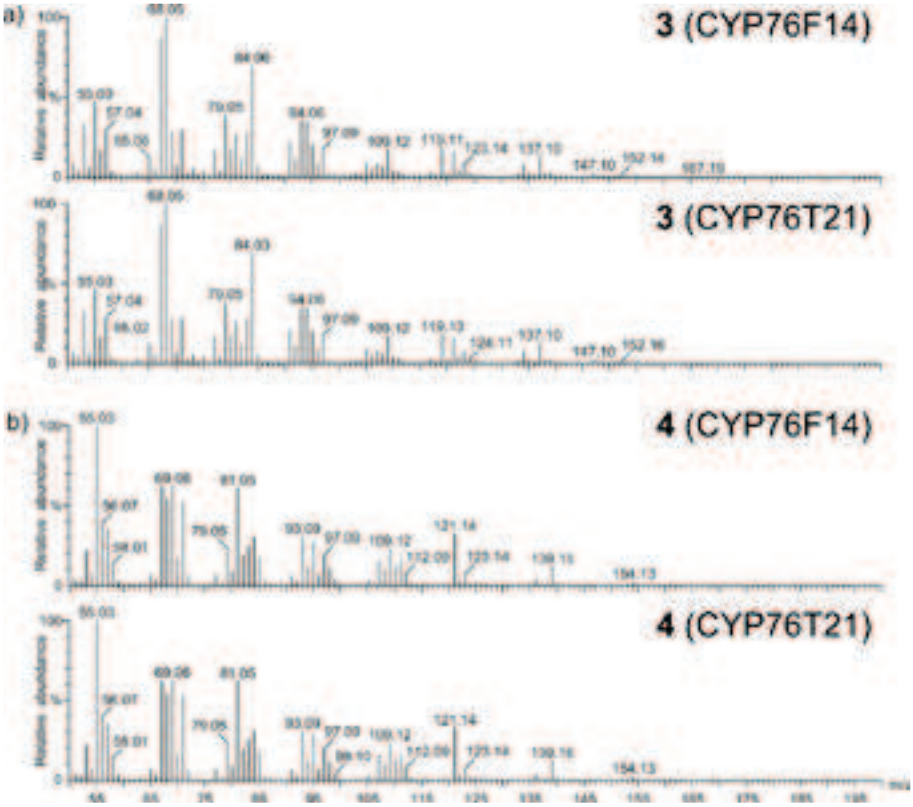

Figure 37. Mass spectra of CYP76F14 and CYP76T21 products formed from nerol (a) and citronellol (b) by grapevine in *N. benthamiana* leaf discs (Figure 3). Product formed from nerol **(a)** was tentatively identified as (*E*)-8-hydroxynerol and the product formed from citronellol **(b)** was tentatively identified as (*E*)-8-hydroxycitronellol based on comparison of mass spectra to those published in (1).

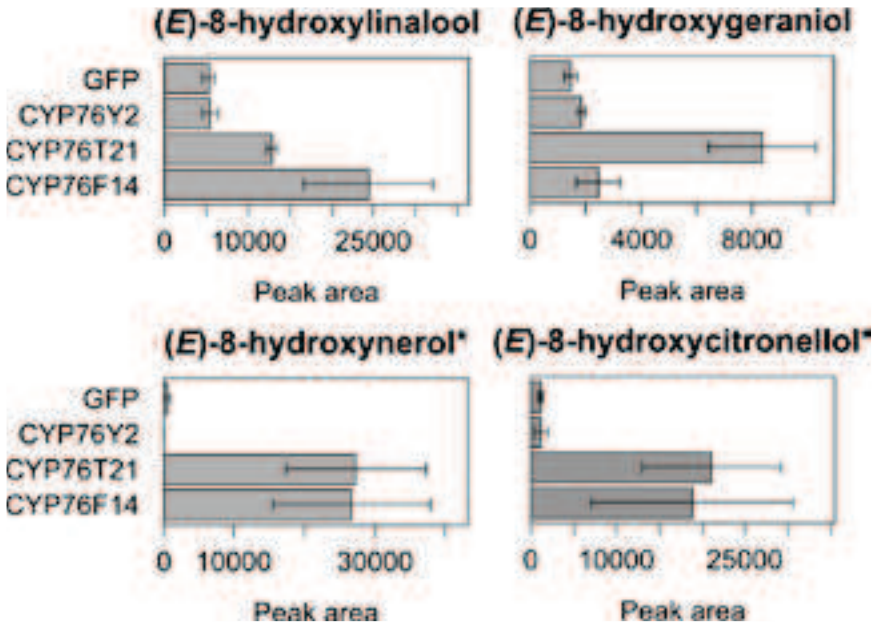

Figure 38. Relative amounts of the main products of grapevine P450s in *N. benthamiana* leaf discs. The leaf discs were incubated in 400 µM solution of substrate for 4 h. The products labeled with an asterisk ((*E*)-hydroxynerol and (*E*)-8-hydroxycitronellol) were putatively identified on the basis of their mass spectrum (Figure 4).

Enantioselectivity of grapevine linalool synthase

The linalool synthase used in the *N. benthamiana* coexpression experiment had not been previously characterized in terms of its enantioselectivity. To determine whether it produces (*R*)- or (*S*)-linalool, we transiently expressed it in *N. benthamiana* leaves as described in the **Chapter 6**. We then collected headspace volatiles with 30 mg Porapak Type Q 50-80 Mesh cartridges (Waters). After elution with hexane, the samples were run on the chiral column Agilent CycloSil B.

Comparison to authentic standards of (*S*)- and (*R*)-linalool showed that VvLIS produces exclusively (*S*)-linalool (Figure 6), which is the predominant form of linalool found in grapes.

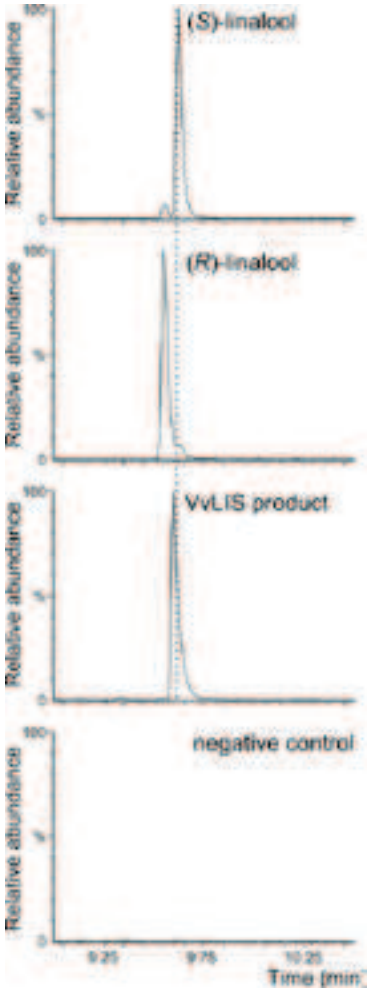

Figure 39. Determination of the configuration of the linalool produced by the grapevine linalool synthase used in our *N. benthamiana* experiments. GC-MS chromatograms of the headspace volatiles collected from *N. benthamiana* leaves transiently expressing the grapevine linalool synthase (VvLIS). Negative control is *N. benthamiana* leaves infiltrated with *A. tumefaciens* bearing the *GFP* gene.

Kinetics of linalool oxidation by CYP76F14

Kinetics of cytochrome P450 reactions can usually be modeled with the Michaelis-Menten equation (4). To evaluate the efficiency of the CYP76F14 enzyme we monitored its reaction kinetics *in vitro*. Reaction mixtures containing 0.009 µM of the CYP76F14 enzyme and 0.5 mM NADPH were prepared in 20 mM K^+/Na^+ phosphate buffer. Enzyme concentration was estimated from its differential CO-spectrum. For the lowest substrate concentration the amount of enzyme was halved to avoid excessive substrate consumption. Reaction was started with addition of (*S*)-linalool at varying concentrations. Reaction mixtures were incubated with agitation at 27°C for 30 min in closed glass test tubes to avoid sample evaporation. Reaction was stopped with the addition of 500 µL ethyl acetate and immediately extracted by vortexing. Ethyl acetate extracts were dried over Na_2SO_4 and injected to GC-FID. Experimental data were fitted to Michaelis-Menten equation using the R software. Velocity was divided by enzyme concentration because the latter was not constant for all substrate concentrations. The kinetic parameters of linalool oxidation were $k_{cat} = (1.11 \pm 0.6)$ s^{-1} and $K_m = (22 \pm 3)$ µM. The K_m is comparable to those obtained for the CYP76 enzymes from *A. thaliana* (5), whereas the k_{cat} is lower. However, the value of k_{cat} has to be considered with care due to the inaccurate determination of enzyme concentration, resulting from the low peak intensity in the CO-difference spectrum. Furthermore, because CYP76F14 catalyses three sequential oxidations of linalool to (*E*)-8-carboxylinalool we monitored substrate depletion, and not product formation. This compromised the reliability of results because it forced us to work at relatively long reaction times to reduce the error in substrate consumption, whereas the Michaelis-Menten kinetic model best describes the initial reaction rate when substrate concentration remains approximately constant ($[S] \approx [S]_0$). In addition enzyme products compete with linalool for binding to the enzyme.

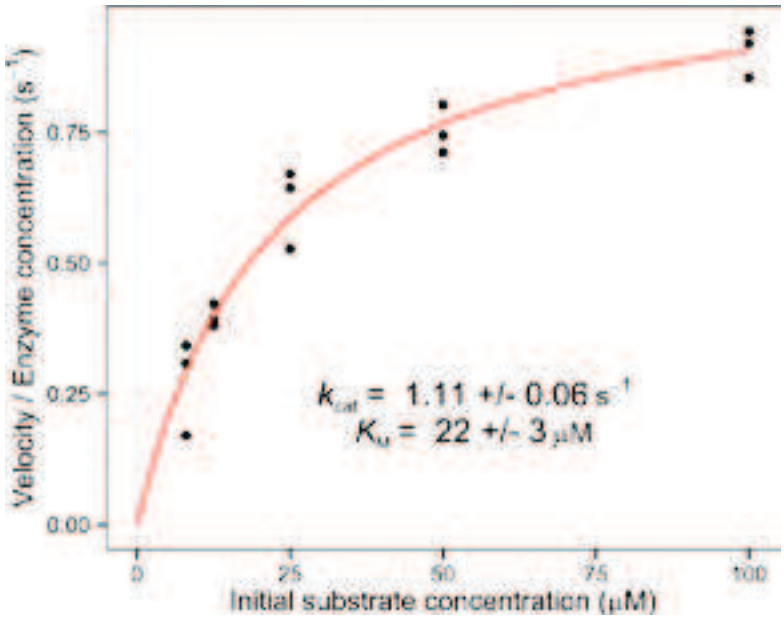

Figure 40. Kinetics of *in vitro* (*S*)-linalool conversion by CYP76F14. The reaction was stopped after 30 min with addition of ethyl acetate. Samples were extracted and substrate conversion was monitored by GC-FID.

Functional characterization of grapevine CYP76 genes expressed in green berry

In grape berries, linalool undergoes many different oxidation reactions, which produce a number of hydroxylated and cyclic linalool derivatives (6). Like other aroma compounds in grapes, their concentration is generally reported to increase during grape berry ripening (7, 8). Some of the linalool oxides, however, are shown to peak in green berries and to decrease during the berry ripening (9). In addition, our previous work showed that although the concentration of (*E*)-8-carboxylinalool, the wine lactone precursor, is increasing during the grape ripening, it is already present in green berries. We therefore decided to explore cytochromes P450 that metabolize linalool in green grape berries.

Comparison of RNA sequencing on cDNA from green and mid-ripe Gewurztraminer berries highlighted two *CYP76* gene candidates for linalool metabolism in green berries: *CYP76F10* and *CYP76T12* (**Chapter 6**, Figure S5). These two genes were cloned from Gewurztraminer green berry cDNA (Table 1). Sequences of several clones revealed that one of the genes, CYP76T12, was expressed as two splicing variants: *CYP76T12v1* with two introns and *CYP76T12v2* with one intron (Figure 8). Other CYP76 genes only have one intron and are homologous to *CYP76T12v2* in the 93-base pair region of the first intron of *CYP76T12v1*. The additional intron in *CYP76T12v1* does not disrupt the open reading frame of the sequence.

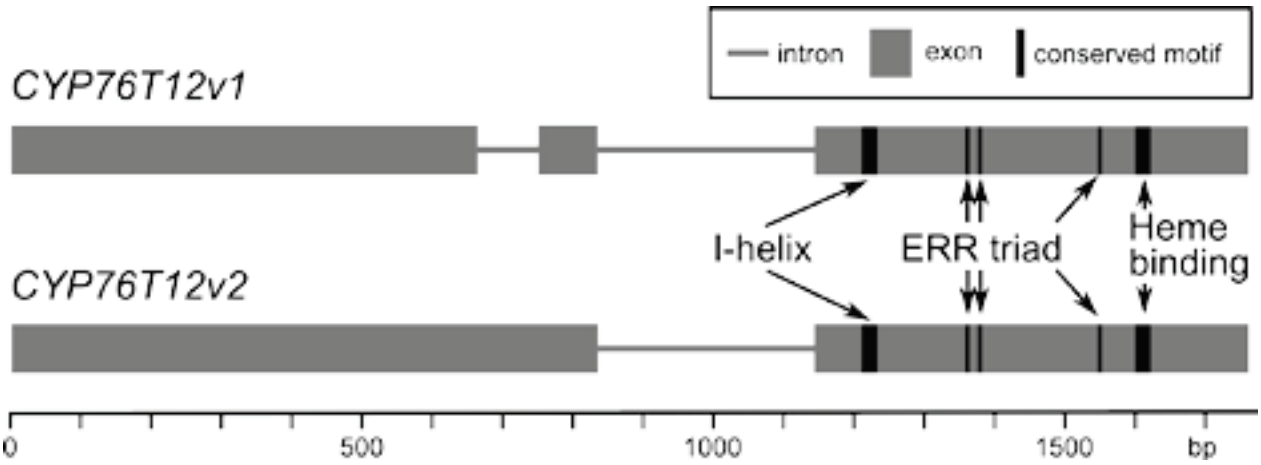

Figure 41. Structure of the alternative splicing variants of the *CYP76T12* gene. *CYP76T12* was cloned from Gewurztraminer green berry cDNA. Two alternative splicing variants were present among the sequenced clones. Conserved P450 motifs (I-helix motif, the catalytic ERR triad and the heme binding motif) are highlighted.

We investigated the expression of the two genes in developing Gewurztraminer berries using qRT-PCR (Figure 9). Two sets of primers were designed for specific detection of the two alternative splicing variants of *CYP76T12* (Table 1). Limited options for primer design resulted in suboptimal primer design with, as a consequence, a low signal also in the negative control (without cDNA). However, this signal was lower compared to samples with cDNA.

All three investigated sequences showed the highest expression in green berries, which is consistent with the RNAseq experiment. The two splicing variants of *CYP76T12* had a nearly identical expression pattern. However, the expression level of these genes was much lower compared to previously investigated genes expressed in ripe berries (**Chapter 6**, Figure 3). Expression of *CYP76F14*, the most highly expressed *CYP76* gene in ripe berries, exceeded the expression of *CYP76F10* and *CYP76T12* even in the green berries.

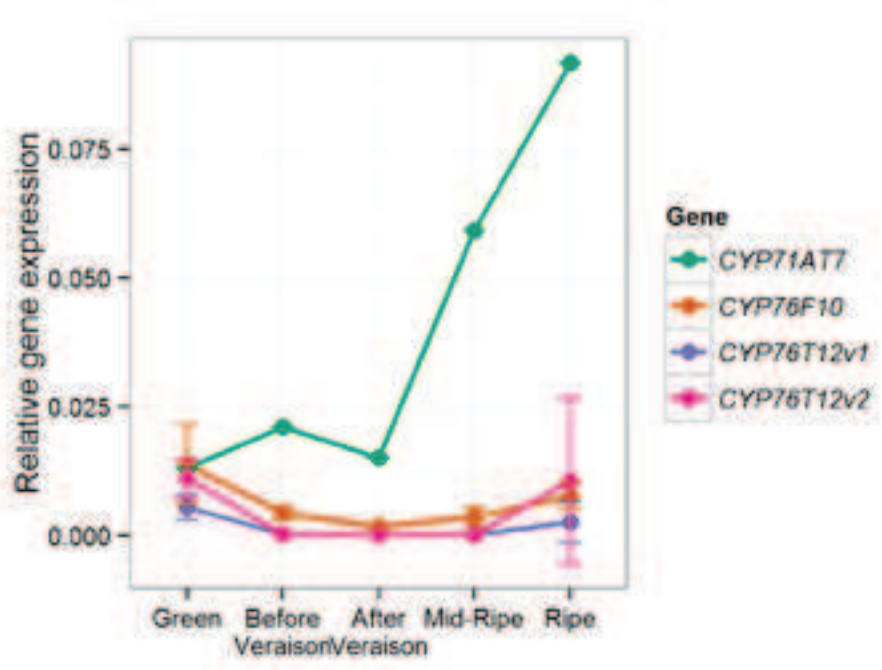

Figure 42. Relative expression of *CYP76F10* and *CYP76T12* during berry development. Error bars represent standard deviation of three technical and three biological replicates. *CYP71AT7* is added for comparison of relative gene expression levels to genes discussed in the previous chapter.

We expressed CYP76F10 and both variants CYP76T12 sequences in yeast (*S. cerevisiae* WAT11). CO-differential spectroscopy of reduced CYP76F10 microsome preparation (Figure 10) revealed weak ΔA peak at 450 nm, similar to that observed with the other two members of this subfamily previously characterized (CYP76F12 and CYP76F14). CYP76T12v2 had a more intense peak at 450 nm, but we could not detect a peak at 450 nm in the microsomal preparation of CYP76T12v1. Splicing of 93 nucleotides, which corresponds to 31 amino acids, thus probably disrupts the protein structure and renders the enzyme non-functional.

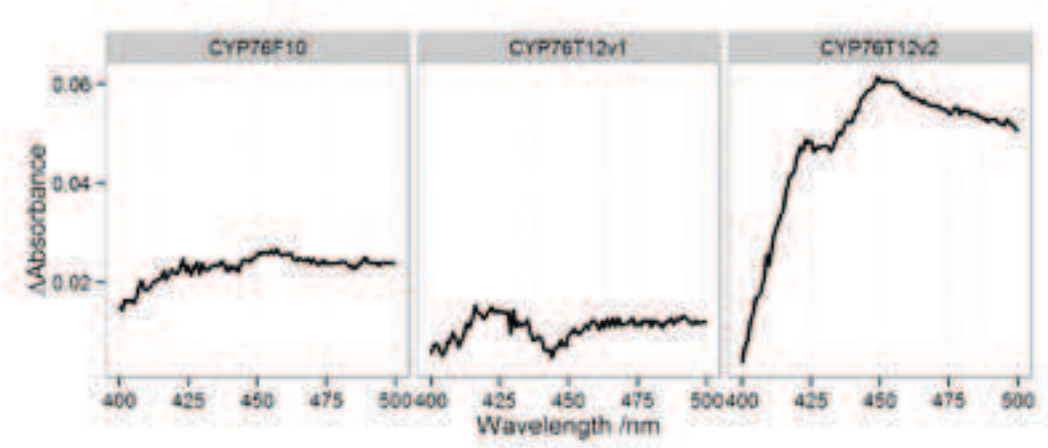

Figure 43. CO-difference spectra of reduced microsomes of yeast expressing CYP76F10 and the two splicing variants of CYP76T12.

We tested the recombinant enzymes with linalool and (*E*)-8-hydroxylinalool. CYP76F10 and CYP76T12v2, metabolized both of the substrates (Figure 11). GC-FID analysis revealed that both of these enzymes can hydroxylate linalool at the position (*E*)-8 and further metabolize (*E*)-8-hydroxylinalool. CYP76T12v1 did not show any enzymatic activity, which confirms the deleterious effect of the splicing of additional intron.

LC-MS/MS analysis of the reaction product showed that both enzymes produce (*E*)-8-carboxylinalool from linalool (Figure 12a). In addition, CYP76T12v2 produced compounds putatively identified as (*Z*)-8-hydroxylinalool **3** and (*Z*)-8-carboxylinalool **6**. Because analytical standards of these products were unavailable, they could not be unambiguously identified. Their putative identification is supported by the fact that they are not produced from (*E*)-8-hydroxylinalool (Figure 12b) or (*E*)-8-oxolinalool (Figure 12c). Upon incubation with these two substrates only (*E*)-8-carboxylinalool was produced in addition to the two side products, probably produced by an endogenous yeast reductase (**4** and **5**).

CYP76F10 and CYP76T12 were transiently expressed in *N. benthamiana* leaves with grapevine linalool synthase (LIS) to evaluate their activity *in planta*. Leaves were extracted with methanol and analyzed by UPLC-MS/MS. Both enzymes produced (*E*)-8-carboxylinalool, which accumulated mostly in glycosylated form. (*Z*)-8-carboxylinalool, one of the putative products of CYP76T12 *in vitro*, could not be detected in extracts of *N. benthamiana* leaves expressing this gene. Neither could putative (*Z*)-8-carboxylinalool be detected in any of the wine samples analyzed in **Chapter 6**.

Like other linalool-metabolizing enzymes from the CYP76 family presented in this work, CYP76F10 and CYP76T12 catalyze the oxidation of the terminal (*E*)- and (*Z*)-8-carbon atoms of the linalool molecule. However, due to their low expression they are unlikely to significantly contribute to the production of (*E*)-8-carboxylinalool in grapes.

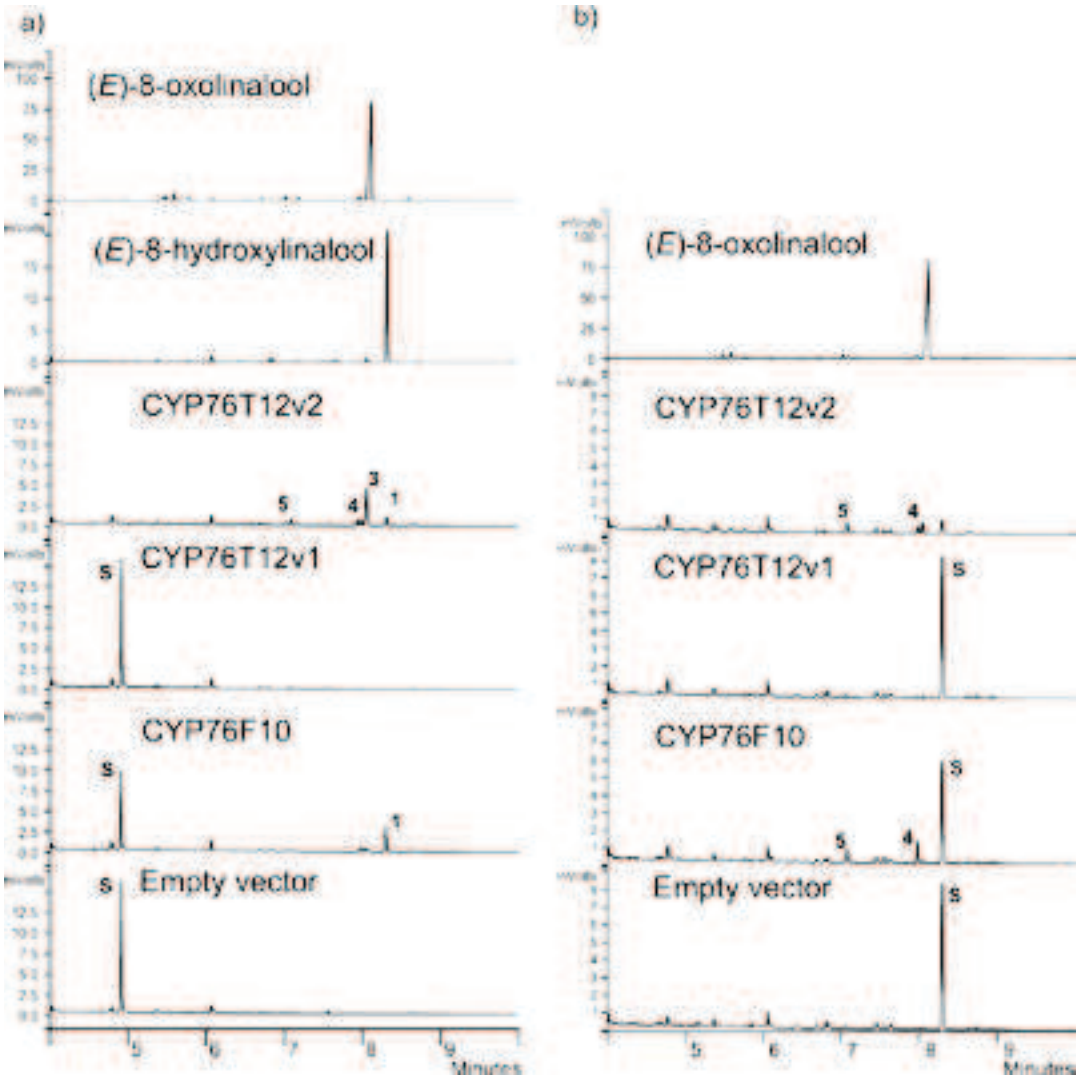

Figure 44. GC-FID identification of the products of linalool (a) and (*E*)-8-hydroxylinalool (b) *in vitro* conversion by the recombinant CYP76F10 and CYP76T12 enzymes Reactions were carried out using recombinant yeast microsomal membranes. Microsomal membranes from yeast transformed with an empty vector were used as a negative control. Top lanes are authentic standards of main reaction products. **1**–(*E*)-8-hydroxylinalool, **3**–(*Z*)-8-hydroxylinalool, **4**–(*E*)-8-oxo-6,7-dihydrolinalool, **5**–(*E*)-8-hydroxy-6,7-dihydrolinalool. **4** and **5** are side products of yeast metabolism of (*E*)-8-oxolinalool.

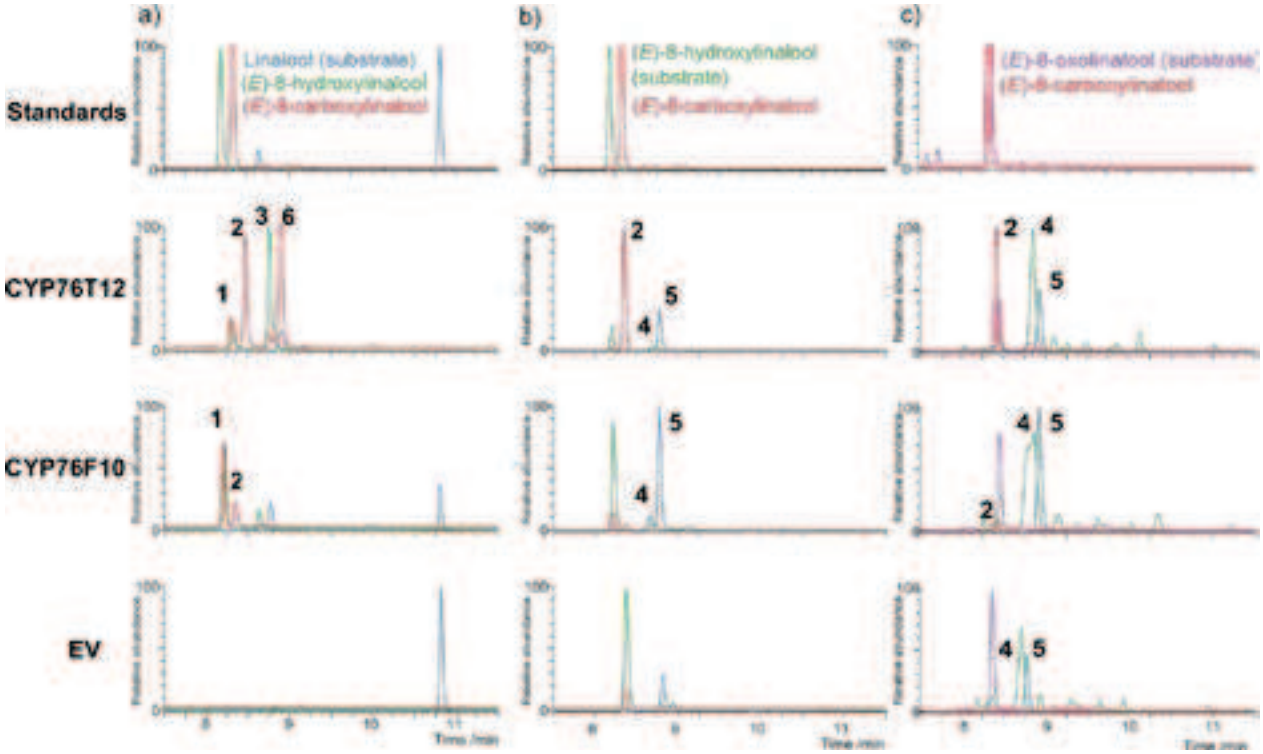

Figure 45. LC-MS/MS identification of the products of the *in vitro* conversion of linalool and its derivatives by the recombinant CYP76F10 and CYP76T21 enzymes. Microsomal preparations of yeast expressing each enzyme and a negative control (microsomes of yeast transformed with an empty vector) were incubated with the following substrates: linalool (a), (*E*)-8-hydroxylinalool (b) and (*E*)-8-oxolinalool (c). Each color represents an MRM (multiple-reaction-monitoring) channel for specific detection of target molecules: blue–linalool (137>80.7), green–(*E*)-8-hydroxylinalool (135>106.8), purple–(*E*)-8-oxolinalool (151.2>92.8) and red–(*E*)-8-carboxylnalool (167.2>92.8). **1**–(*E*)-8-hydroxylinalool, **2**–(*E*)-8-carboxylinalool, **3**–(*Z*)-8-hydroxylinalool, **4**–(*E*)-8-oxo-6,7-dihydrolinalool, **5**–(*E*)-8-hydroxy-6,7-dihydrolinalool, **6**–(*Z*)-8-carboxylinalool. **4** and **5** are side products of yeast metabolism of (*E*)-8-oxolinalool.

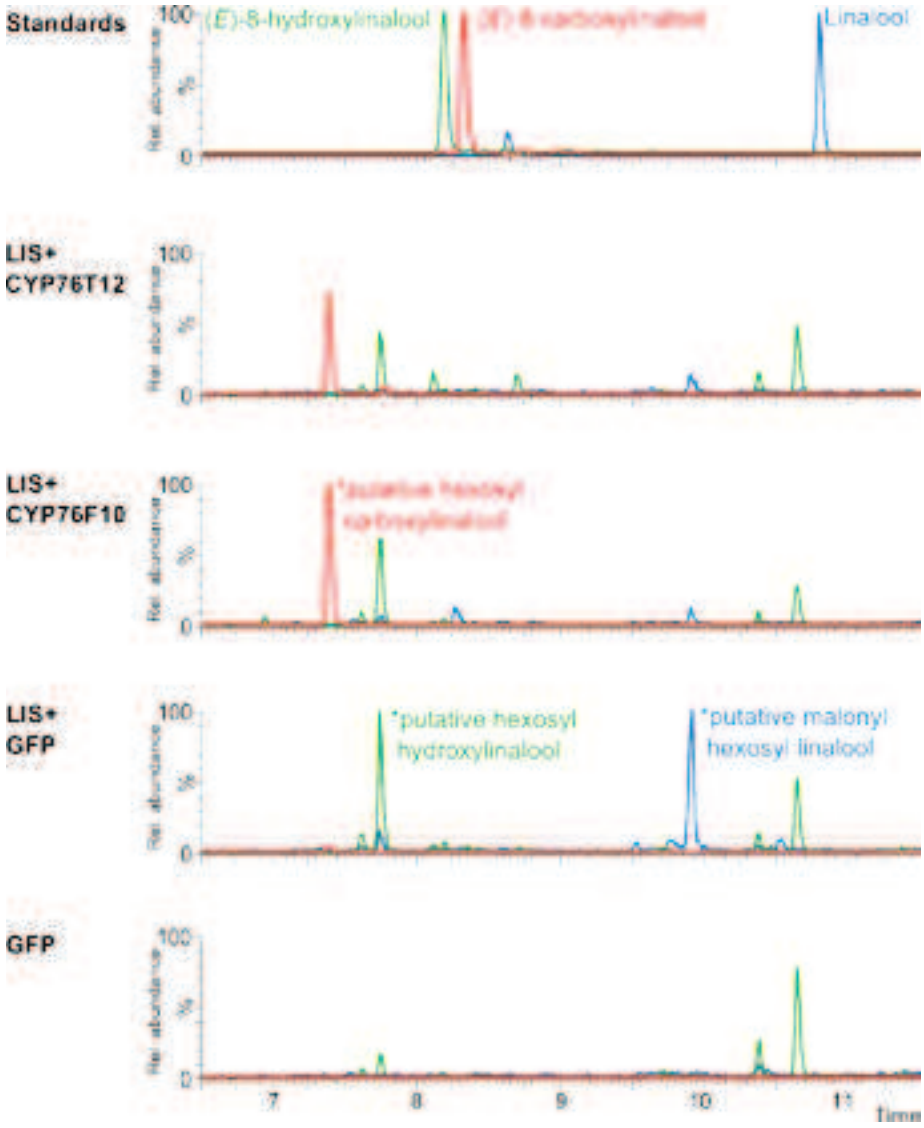

Figure 46. LC-MS/MS profiles of *N. benthamiana* leaves expressing linalool synthase (LIS) alone or together with CYP76T12 or CYP76F10. Leaves were transfected with a combination LIS and a P450 gene or GFP as a negative control. 4 days after infection leaves were collected, frozen in liquid nitrogen and extracted with methanol. Each color represents an MRM channel for specific detection of target molecules: blue–linalool (137>80.7), green–(*E*)-8-hydroxylinalool (135>106.8) and red–(*E*)-8-carboxylinalool (167.2>92.8). Most of linalool derivatives were present as conjugated forms, which were putatively identified on the basis of their mass spectra. Putatively identified conjugated products are labeled with an asterisk.

Table 3. Primers used for cloning and qRT-PCR of the "green berry" P450 genes. Restriction sites and USER™ cloning sites are underlined.

Gene	Forward primer	Reverse primer
CYP76F10	ATAGGATCCATGGACTTGATGAG CTATTTGCTATGTC	TAGGTACCTCATACATGGATAGG GAGAACACG
CYP76T12	ATAGGATCCATGGATTACATAAC ATTTTTGCTTTTGC	TAGGTACCTTAAACCTTGATGGG AATAGCC
CYP76F10u	GGCTTAAUATGGACTTGATGAG CTATTTGCTATGTC	GGTTTAAUTCATACATGGATAGG GAGAACACG
CYP76T12u	GGCTTAAUGAAGTAGACAATGG ATTACATAAC	GGTTTAAUTTAAACCTTGATGGG AATAGCC
CYP76F10 (qPCR)	AGTGCAGGGTTGGATGGCTAGC	TCGCAGTTGTATCAGTCCCCGC
CYP76T12v1 (qPCR)	TGATCCACAAGCAAGCAAGGATG	AGTCGTGTCAGTTCCCGCAAC
CYP76T12v2 (qPCR)	TGCCCGGGGTGTGATGGAAG	TGTTCCACCGTATACCTTGTGGA

Oxidation of olefinic monoterpenes by CYP71AT7

Whereas several *A. thaliana* CYP76 enzymes have been described as monoterpenol oxidases, only one of the functionally characterized CYP71 exhibited this activity. However, some CYP71 enzymes from different plants catalyze hydroxylations of olefinic (hydrocarbon) monoterpenes. For example, CYP71D13 and CYP71D15 oxidize limonene in peppermint (10), and CYP71AR1 oxidizes α-pinene in strawberry (11). This motivated us to test the activity of CYP71AT7, the most highly expressed CYP71 in ripe berries, with olefinic monoterpenes.

We tested five monoterpenes: α-pinene, *p*-cymene, limonene, β-ocimene and γ-terpinene (Figure 14). Reactions were carried out in 500 µL reaction volume in closed glass vials in 20 mM Na^+/K^+ phosphate buffer (pH 7.4) containing 200 µM of substrate and 5 % vol. of the microsomal preparation. After 30 min at 27°C, the reaction was stopped and extracted with a mixture of pentane and ethyl acetate (1:1 vol.), and the extract analyzed using GC-MS.

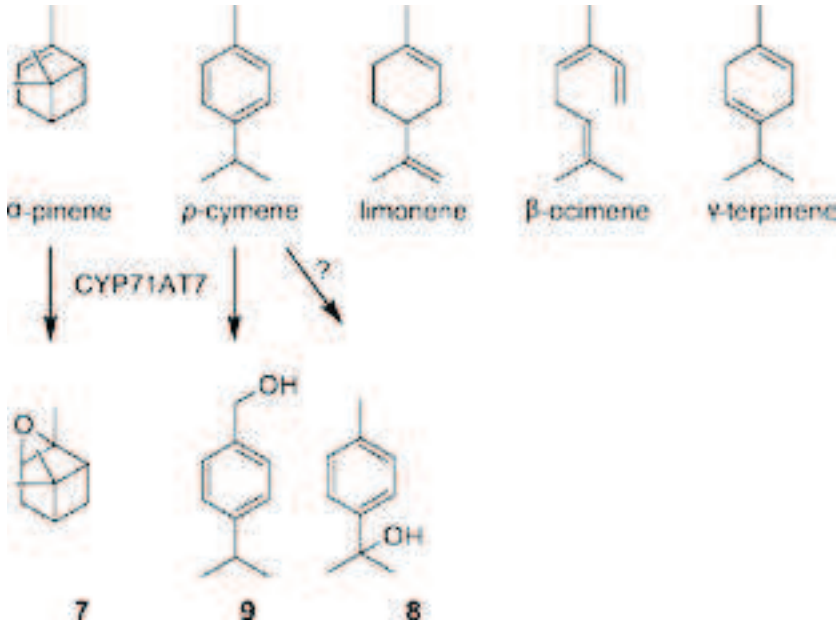

Figure 47. Monoterpene oxidation catalyzed by CYP71AT7. Five olefinic monoterpenes (top row) were tested as substrates of the yeast-expressed CYP71AT7. Two of them were oxygenated as shown.

Although the conversion was relatively low, oxygenated products could be detected for two of the substrates (Figure 15). Because of unavailability of analytical standards, the products were only tentatively identified on the basis of the best match in the NIST mass spectral library search (Figure 16). CYP71AT7 catalyzed the epoxidation of α-pinene at the double bond to yield 2,3-epoxypinane (**7**) and the hydroxylation of *p*-cymene at the methyl group to yield 4-isopropylbenzyl alcohol or cuminol (**9**). The solution of *p*-cymene already contained some p-cymen-8-ol (**8**), but its concentration increased after incubation with CYP71AT7.

All of the tested monoterpenes except *p*-cymene were reported as *in vitro* products of grapevine terpene synthases (12), but to our knowledge neither of them has been detected in

grapes or wines. Neither has any of the three products been detected in grapes or wines. In addition, the conversion was relatively low (Figure 15), so we concluded that the observed catalytic activity was unlikely relevant *in vivo*.

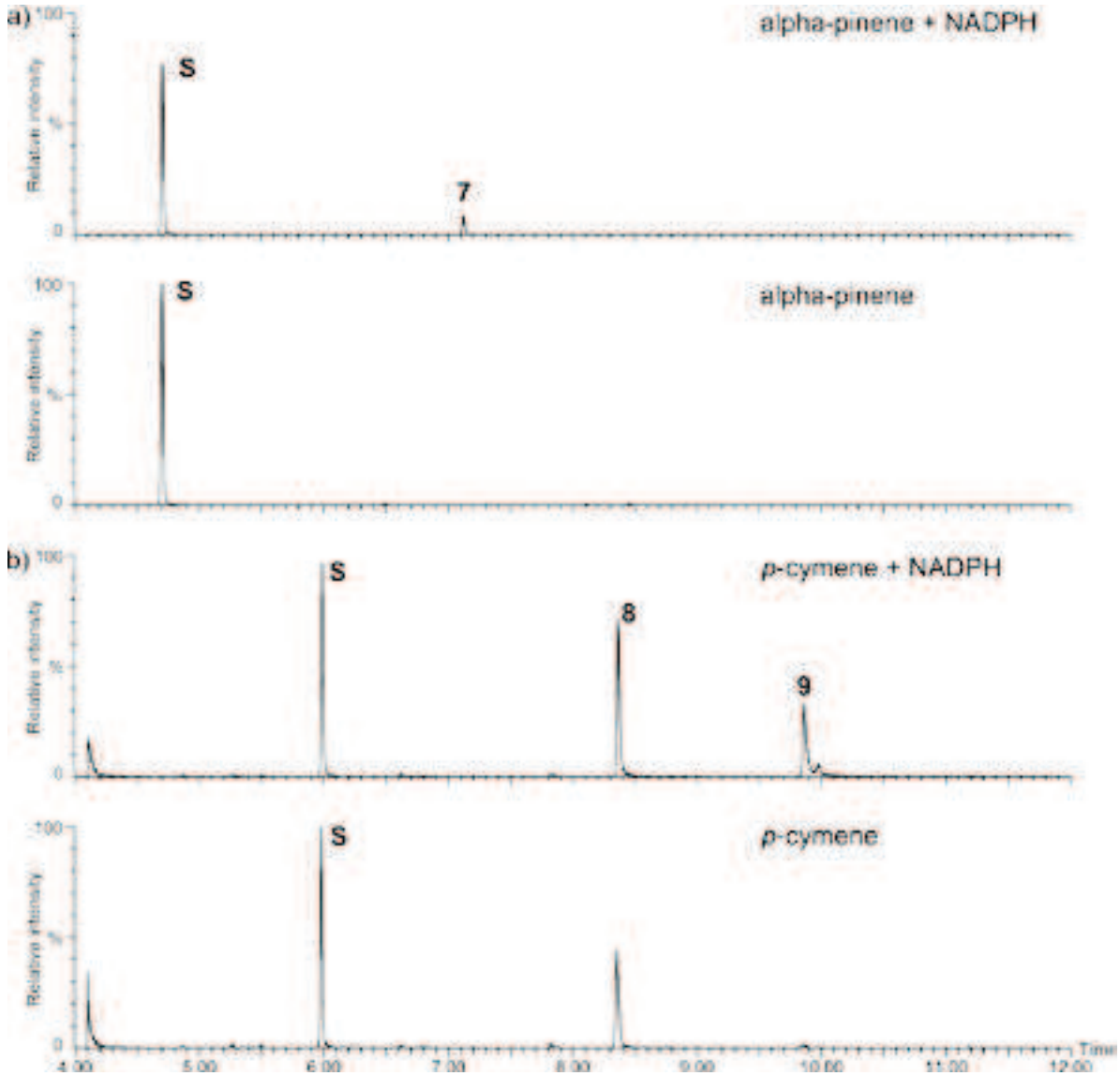

Figure 48. GC-MS profiles of the products of the *in vitro* conversion of α-pinene (a) and *p*-cymene by yeast-expressed CYP71AT7 (b). Reaction in the absence of NADPH was used as a negative control. Chromatograms show the relative abundance of selected m/z 93+67 **(a)** and 135 **(b)**. Products were tentatively identified on the basis of their mass spectra (**Figure 16**). **S** – substrate, **7** – 2,3-epoxypinane, **8** – *p*-cymen-8-ol , **9** – cuminol.

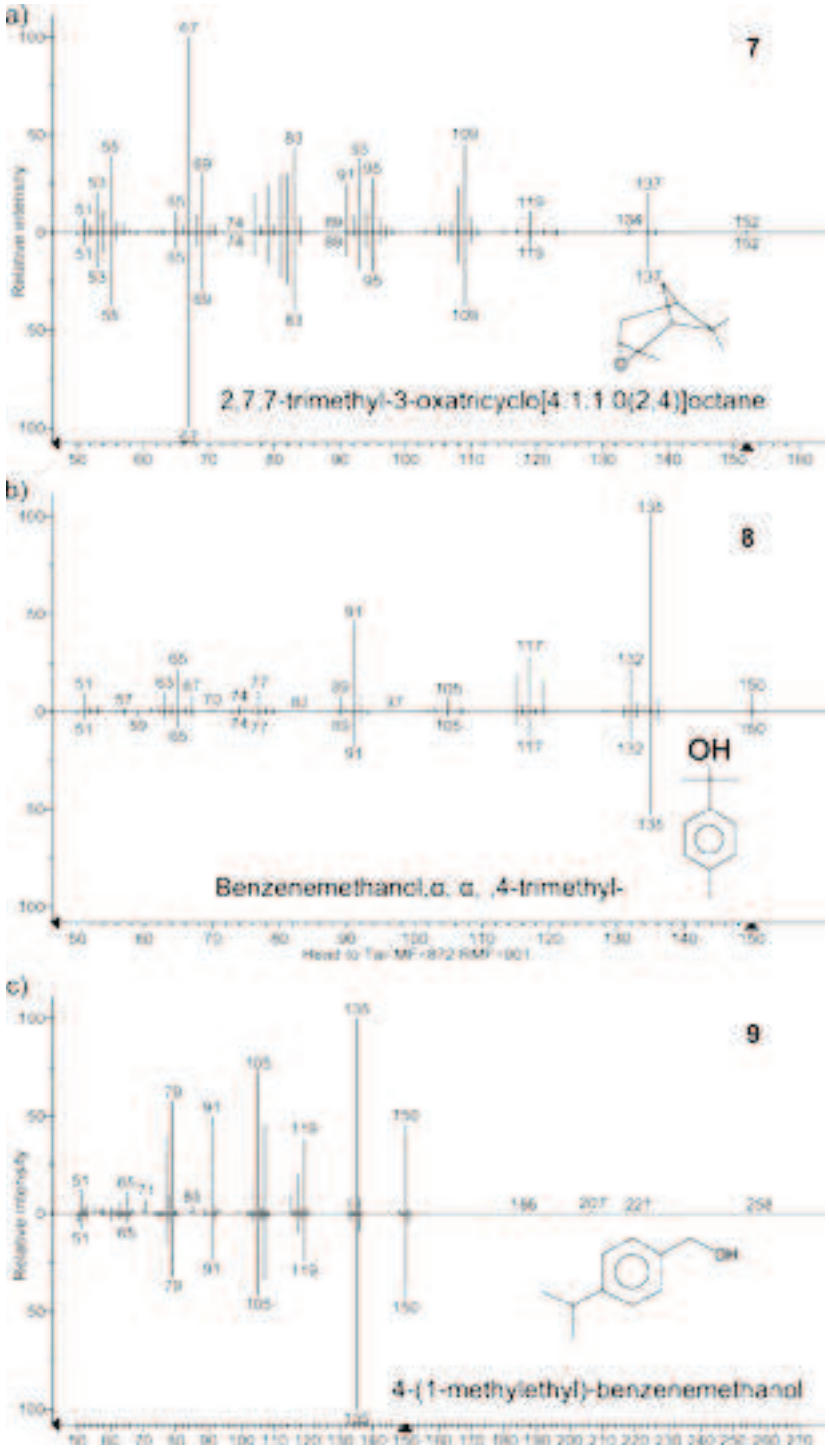

Figure 49. Identification of the conversion products of α-pinene (a) and *p*-cymene (b) conversion by CYP71AT7. Mass spectra of the product peaks were searched against the NIST Mass spectral library database. This enabled putative identification of all three peaks: **7** – 2,3-epoxypinane, **8** – *p*-cymen-8-ol and **9** – cuminol. The top spectrum of each panel corresponds to the product peak and the bottom one to the spectrum from the database.

Role of other CYP families in grape aroma biosynthesis

In **Chapter 6** and this chapter we showed that grapevine CYP76 enzymes oxidize linalool at positions (*E*)-8 and (*Z*)-8. However, none of these enzymes oxidized linalool at in-chain positions, despite numerous other oxygenated linalool derivatives commonly found in grapes. This led us to consider that other P450 families might be involved in linalool oxidations in grapes. RNA sequencing experiment revealed other P450 candidates with increased expression in mid-ripe Gewurztraminer berries (Figure 17). Some of the genes featuring in this list belong to P450 families with conserved functions and were not further evaluated. For example, CYP90 are involved in metabolism of brassinosteroids, CYP97 metabolize carotenoids and CYP74 catalyze rearrangement of fatty acid hydroxyperoxides (13).

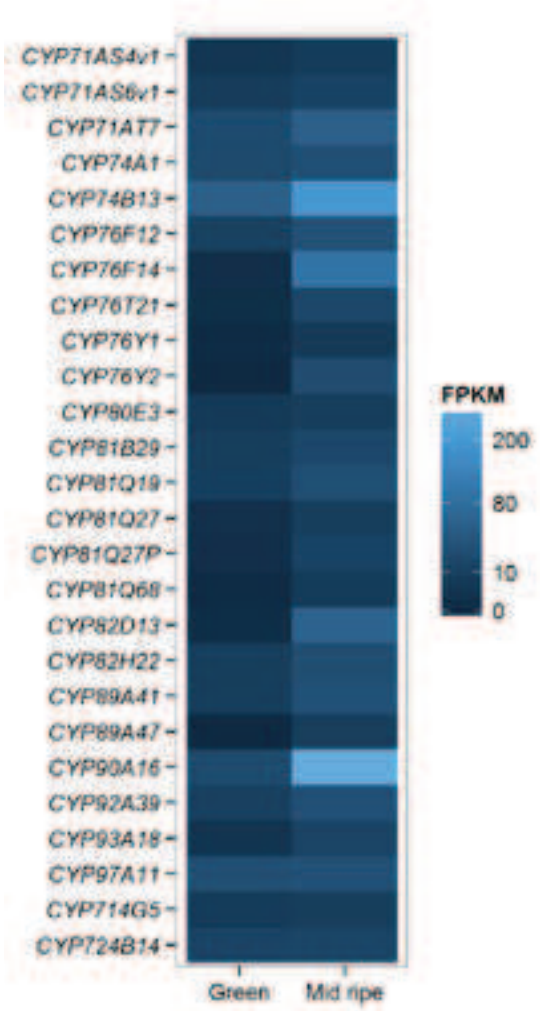

Figure 50. Cytochrome P450 genes induced during ripening of Gewurztraminer berries. Gene expression of cytochrome P450 genes was evaluated with RNAseq on green and mid ripe Gewurztraminer grape berries. Heatmap displays cytochrome P450 genes with higher expression in mid ripe berries compared to green berries and with expression exceeding 5 fragments per kilobase of transcript per million mapped reads (FPKM).

Genes from the CYP82 family appeared possible candidates for monoterpenol oxidation. Members of this family are involved in terpenol metabolism in *A. thaliana*. They synthetize

homoterpenes (shortened terpenoids) by an unusual oxidative cleavage of nerolidol and geranylgeraniol (14). In addition, this family is larger in grapevine compared to other plant species, which could indicate expansion and subfunctionalisation. Two *CYP82* genes are induced during the berry ripening: *CYP82D13* and *CYP82H22*. We attempted to clone both of them from Muscat Ottonel ripe berry skin cDNA, but only succeeded at amplifying *CYP82D13*. The expression of *CYP82H22* might be too low (Figure 17).

We amplified the *CYP82D13* sequence from different varieties: Muscat Ottonel, Gewurztraminer and Riesling, as well as wild grapevine *V. sylvestris* subsp. *sylvestris*. Surprisingly, all of the sequences, except one of the putative alleles of *V. sylvestris* subsp. *sylvestris* had a mutation that changed the most conserved residue in the P450 sequences, the heme-binding cysteine, into a tyrosine residue (Figure 18). Until recently the cysteine residue was considered the only conserved amino acid in all P450 sequences, but more recently P450 genes with variations at this position have been discovered, although it is still unclear whether they have an enzymatic activity or not (15).

Figure 51. Alignment of translated CYP82D13 sequences from different *V. vinifera* varieties. CYP82D13 was cloned from berry skin cDNA of different grapevine varieties and wild grapevine *V. sylvestris* subsp. *sylvestris*. The heme-binding cysteine residue is labeled with an arrow.

We expressed the Muscat Ottonel *CYP82D13* gene in yeast. The reduced microsomes of yeast expressing *CYP82D13* lacked the typical absorbance peak at 450 nm. A minor peak at 420 nm was however observed (Figure 19), indicating that at least a minor proportion of the protein bound the heme in the active site. Nevertheless, no enzyme activity was observed when this enzyme was tested with monoterpenols.

Other families considered for monoterpenol oxidations were CYP80, CYP81 and CYP89. Members of the CYP80 family were shown to catalyze phenol coupling reactions in alkaloid biosynthesis in other plants, but we decided to review this family because of its phylogenetic proximity with CYP76 (**Chapter 5**). The family CYP81 had the most representatives in the list of ripening-induced cytochrome P450s (along with CYP76), we therefore decided to test it for involvement in aroma formation. Members of CYP81 catalyze different reactions in specialized metabolism of other species, such as isoflavonoid, glucosinolate or lignan biosynthesis (16). Finally, two CYP89 members are induced in ripe berries. So far the biochemical function of this family is unknown.

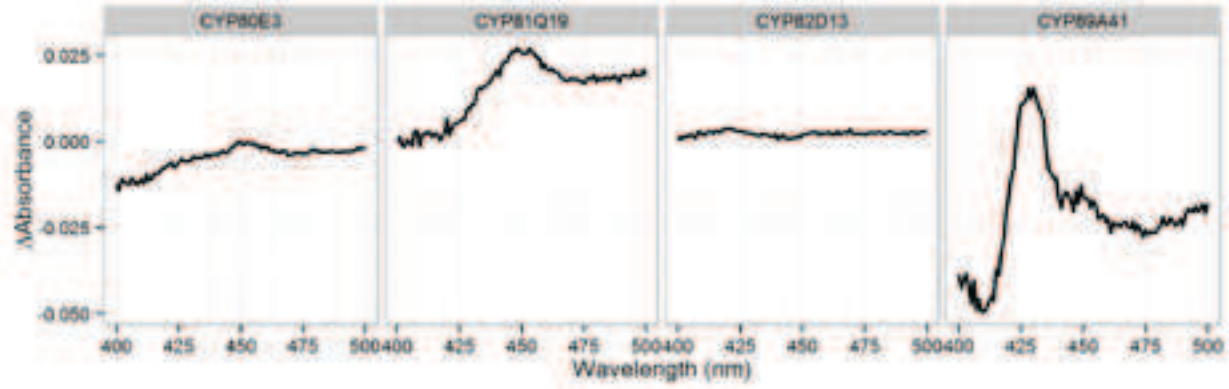

Figure 52. CO-difference spectra of reduced yeast microsomes expressing the selected cytochrome P450 candidates from CYP80, CYP81, CYP82 and CYP89 families.

We chose to investigate the most highly expressed member of each of these families: CYP80E3, CYP81Q19 (previously named CYP81V13) and CYP89A41. Their coding sequences were amplified and expressed in yeast as described in **Chapter 6**. The list of primers used for amplification is available in Table 2. All three P450s were expressed in yeast at least to a minor extend, as indicated by a peak at 450 nm in the CO-difference spectrum of recombinant yeast microsomes (Figure 19). However, none of the enzymes showed any monoterpenol oxidation activity, which suggests these P450s are involved in other ripening related processes.

Table 4. Primers used for amplification of the CYP80, CYP81, CYP82 and CYP98 candidates.
Restriction sites are underlined.

Gene	Forward primer	Reverse primer
CYP80E3	TAGGTACCATGGTAGCCATGTTAGCAGAAG GAAC	TAGGTACCTCATTTTCGAACTTTAGGAATGAT GAG
CYP81Q19	ATAGGATCCATGGAAGCTAGATGGCTATACT CATC	TAGGTACCTCAGACACTGTTCAGAATTTTGGA TG
CYP82D13	GGATCCATGGATTTCCTTCTCCAATGCCTAA AC	GGTACCCTACTTATAAAGATTATAAGAAGCTAA GCGTGGG
CYP89A41	ATAGGATCCATGGAGATATGGGTCTTCTTCT TCATC	TAGGTACCTTATTTCAACCTCGGAGATAAGTG GG

Effect of wine lactone on sensory properties of wine

Wine lactone was reported to have a very low odor detection threshold, which was determined at 60 pM in model wine (17). Model wine only contains water and ethanol, and is as such a very simple mixture, as opposed to wines, which contain hundreds of different volatiles. We wished to investigate if wine lactone has an impact on wine aroma at concentrations detected in old Gewurztraminer wines (10 nM).

We tested the null hypothesis that addition of 10 nM wine lactone does not affect the sensory properties of wine. We used the methodology of the triangle test (18), in which panelists were presented with three samples, of which two are identical and one different. The panelists were then asked to identify the different samples. In our experiment all three samples were the wine that contained the lowest amount of wine lactone (<1 year old Riesling from the producer Emile Weingut, which contained 1.74 nM wine lactone), except that one of the samples was spiked with 10 nM wine lactone.

Our study included 48 volunteers from Institut de Biologie Moléculaire des Plantes (21 females and 27 males) between the age of 21 and 60 (mean age 37, standard deviation 12). They were presented with three coded glasses containing 10 mL wine, arranged in a triangular form, and ballot papers, which instructed them to circle the code that represented the different sample.

The majority of panelists (27) chose the correct answer; the two false answers got 9 and 12 answers. The exact binomial test was calculated with the sensR package (19) in R software. The results showed that the d-prime parameter, which describes the sensory difference, is greater than 0 (p-value < 0.001). The d-prime parameter was estimated to 1.8, which indicates a medium to large sensory difference. We therefore could conclude that wine lactone impacts the complex wine aroma at concentrations above 10 nM.

Bibliography

1. Hofer R, et al. (2014) Dual function of the cytochrome P450 CYP76 family from Arabidopsis thaliana in the metabolism of monoterpenols and phenylurea herbicides. *Plant Physiol* 166(3):1149–1161.
2. Höfer R, et al. (2013) Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco)iridoid pathway. *Metab Eng* 20:221–232.
3. Ginglinger J-F, et al. (2013) Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in Arabidopsis flowers. *Plant Cell* 25(11):4640–57.
4. Shou M, et al. (2012) Enzyme Kinetics of Cytochrome P450-Mediated Reactions. *Curr Drug Metab* 2(1):17–36.
5. Boachon B, et al. (2015) CYP76C1 (Cytochrome P450)-Mediated Linalool Metabolism and the Formation of Volatile and Soluble Linalool Oxides in Arabidopsis Flowers: A Strategy for Defense against Floral Antagonists. *Plant Cell* 27(10):2972–90.
6. Luan F, Mosandl A, Degenhardt A, Gubesch M, 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(1-2):353–364.
7. Wilson B, Strauss CR, Williams PJ (1984) Changes in free and glycosidically bound monoterpenes in developing Muscat grapes. *J Agric Food Chem* 32(4):919–924.
8. Fenoll J, Manso A, Hellin P, Ruiz L, Flores P (2009) Changes in the aromatic composition of the Vitis vinifera grape Muscat Hamburg during ripening. *Food Chem* 114(2):420–428.
9. Luan F, Mosandl A, Gubesch M, Wüst M (2006) Enantioselective analysis of monoterpenes in different grape varieties during berry ripening using stir bar sorptive extraction- and solid phase extraction-enantioselective-multidimensional gas chromatography-mass spectrometry. *J Chromatogr A* 1112(1-2):369–74.
10. Haudenschild C, Schalk M, Karp F, Croteau R (2000) Functional expression of regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha spp.) in Escherichia coli and Saccharomyces cerevisiae. *Arch Biochem Biophys* 379(1):127–36.
11. Aharoni A, et al. (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* 16(11):3110–3131.
12. Martin DM, et al. (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(1):226.
13. Bak S, et al. (2011) Cytochromes P450. *Arabidopsis Book* 9(9):e0144.
14. Lee S, et al. (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. *Proc Natl Acad Sci U S A* 107(49):21205–10.
15. Sezutsu H, Le Goff G, Feyereisen R (2013) Origins of P450 diversity. *Philos Trans R Soc Lond B Biol Sci* 368(1612):20120428.
16. Nelson D, Werck-Reichhart D (2011) A P450-centric view of plant evolution. *Plant J* 66(1):194–211.
17. Guth H (1997) Identification of character impact odorants of different white wine varieties. *J Agric Food Chem* 45(8):3022–3026.
18. Naes T, Brockhoff PB, Tomic O (2010) Discrimination and Similarity Testing. *Statistics for Sensory and Consumer Science*, eds Naes T, Brockhoff PB, Tomic O (J. Wiley & Sons), pp 79–94.
19. Christensen R, Brockhoff P (2015) sensR---An R-package for sensory discrimination.

Chapter 8
Conclusions and perspectives

Chemical composition of wine aroma

Although economical importance of wine makes it one of the best researched beverages, we are still struggling to understand its chemical composition. The quality of wine largely depends on its aroma, but its complexity makes it a difficult subject to study: wine aroma comprises a large number of chemically diverse compounds at very different concentrations. We analyzed data from published grape and wine profiling experiments and compiled a list of 141 common grape and wine volatiles. However, many potent odorants in wine consistently slip under the radar of untargeted profiling because of their low concentrations. In terms of aroma, abundance of a compound is rarely correlated to its importance and for some compounds a human nose is a better detector than sophisticated analytical equipment.

An example of low-abundance/high-impact aroma compound is the wine lactone: a monoterpene lactone with a sweet and coconut-like aroma, shown to be an important determinant of aroma of some wines. This compound was not detected in any of the profiling experiments in our meta-analysis. In fact, in almost 20 years between its discovery in wine and this study, it has only been quantified in three wine samples (1, 2). Besides its low concentration (nM range), its relatively high polarity (theoretical $\log P_{ow} = 2$) decreases the efficiency of extraction to organic solvents, so the only successful attempts of wine lactone quantification involved a laborious continuous extraction procedure. We developed an unconventional analytical method, which exploits the unusually high sensitivity for wine lactone on LC-ESI-MS/MS, sufficient to leverage for low extraction efficiency. Our workflow thus allowed for quantification of wine lactone in 23 samples. The differences in concentration of wine lactone between samples were over 10-fold (from 2 to 32 nM) and always exceeded the odor detection threshold (60 pM in model wine) in all the analyzed wines. From the three analyzed wine varieties, Gewurztraminer had the highest wine lactone content.

Dynamic changes in wine aroma

During the winemaking process wine aroma undergoes complex chemical and biochemical changes. Although wine aroma is commonly divided into grape-derived, fermentation-derived and aging-derived, our meta-analysis and review show that the same compounds can be affected by all three of these processes. For example, many aroma compounds are produced in the grapes as non-volatile glycosides, and released as volatile aglycone upon

fermentation, enzymatic hydrolysis, acid hydrolysis (during wine aging) or by enzymatic hydrolysis in human mouth. Some volatile compounds can undergo secondary transformations, such as acid-catalyzed rearrangements, to yield compounds with altered aromatic properties.

Development of wine lactone follows both of these transformations. Soon after discovery of wine lactone in wines, its putative precursor, (*E*)-8-carboxylinalool, was isolated from wines and shown to yield wine lactone in an acid-catalyzed chemical rearrangement (3). This reaction was studied in the laboratory under somewhat harsh conditions (simultaneous distillation-extraction). At realistic conditions, however, the reaction advanced extremely slowly, it was thus proposed that wine lactone forms from (*E*)-8-carboxylinalool during wine aging (4). Our LC-MS/MS approach allowed for the quantification of both wine lactone and (*E*)-8-carboxylinalool in the same wine samples. Our results confirmed the slow formation of wine lactone from (*E*)-8-carboxylinalool during wine aging with an estimated kinetic constant $k = (0.00012 \pm 0.00002)$ year^{-1}, meaning that every year of wine aging only roughly one out of 10000 (*E*)-8-carboxylinalool molecules will transform into wine lactone. In spite of this extremely slow rate of formation, wine lactone reaches concentrations above the odor detection threshold even in young wines, due to the high concentration of the precursor (*E*)-8-carboxylinalool.

Analysis of grape berry extracts showed that (*E*)-8-carboxylinalool is present in grapes exclusively in glycosylated form, previously characterized as a glucose ester (3, 5). Wines, however, contain both free and conjugated (*E*)-8-carboxylinalool, pointing to partial hydrolysis of the glucose ester during the fermentation. Additionally, the content of the glucose ester was higher in younger compared to older wines, suggesting acid hydrolysis of the glucose ester takes place during wine aging. We could thus show that the formation of wine lactone depends on the three processes: biosynthesis in grapes, fermentation and aging. Hydrolysis of (*E*)-8-carboxylinalool glucose ester during fermentation and wine maturation needs to be studied further to understand why it seemingly occurs faster in some wines (Gewurztraminer) compared to the others (Riesling and Muscat).

Biosynthesis of wine aroma

The long history of grapevine domestication and clonal selection resulted in cultivars with desired organoleptic properties, but at a cost of reduced fitness of the plants. This is reflected in a high pesticide use in viticulture. Genes that influence wine quality are therefore important for the creation of new pathogen resistant varieties with desired organoleptic properties. Information about aroma biosynthetic genes can be applied in the process of marker-assisted breeding, as well as in genetic engineering of new varieties. We identified

and characterized the most likely candidate for the biosynthesis of (*E*)-8-carboxylinalool, the wine lactone precursor in grapes. CYP76F14 is a grapevine cytochrome P450, which is highly expressed in ripe grape berries. We showed both *in vitro* and in *N. benthamiana* that CYP76F14 synthesizes (*E*)-8-carboxylinalool from linalool. Preliminary genetic data, obtained at the INRA of Colmar, indicate an overlap between the *CYP76F14* locus and the quantitative trait locus (QTL) for (*E*)-8-carboxylinalool content (data not shown). This result consolidates the role of CYP76F14 as the main (*E*)-8-carboxylinalool synthase in grapevine.

As for many aroma compounds, chirality is detrimental of the perceived aroma characteristic of wine lactone. Of all eight wine lactone diastereomers, the one initially detected in wine (3*S*, 3*aS*, 7*aR*) has the lowest odor detection threshold (6). This isomer is derived from (3*R, 6E*)-8-carboxylinalool, whereas the predominant configuration of linalool in grapes was found to be 3*S* (7). Only recently both enantiomers of (*E*)-8-carboxylinalool—(3*S)* and (3*R)*—were found in wines, along with the both resulting enantiomers of wine lactone—(3*S*, 3*aS*, 7*aR*) and (3*S*, 3*aS*, 7*aR*). We showed CYP76F14 can metabolize both (3*S)*- and (3*R)*-linalool to corresponding carboxylic acid without clear stereoselectivity. The ratio of (3*S, 6E*)- to (3*R, 6E*)-8-carboxylinalool, and therefore the amount of the fragrant enantiomer of wine lactone, thus likely depend on the amount of (3*S)*- and (3*R)*-linalool produced by the terpene synthases. In the functional analysis of the linalool synthase gene family in grapevine, only one isoform was found to produce (3*R)*-linalool *in vitro*, as opposed to several 3*S*-linalool producing terpene synthases (8). The expression of the (3*R)*-linalool synthase might therefore be crucial for the production of the fragrant enantiomer of wine lactone.

The manual annotation and expression analysis of the cytochrome P450 genes present in the grapevine genome was instrumental to identify candidates for (*E*)-8-carboxylinalool biosynthesis. It also highlighted P450 genes that might be involved in biosynthesis of other organoleptic compounds. Our meta-analysis showed the presence of 6 other oxygenated linalool derivatives in grapes (6-hydroxylinalool, 7-hydroxylinalool and four linalool oxide isomers), yet none of the enzymes we tested was able to produce any of these compounds. This indicates possible involvement of other P450 families, not previously implicated in monoterpenol metabolism.

The genetic background of the differences between grapevine cultivars, including differences in aroma, is still largely unexplored. The reference grapevine genome belongs to a highly inbred descendant of Pinot noir cultivar and therefore significantly differs, both genotypically and phenotypically, from cultivated varieties. In the reference genome we identified 11 putative cytochrome P450 pseudogenes, which only contain one nonsense or frameshift mutation. Those might be functional genes in cultivated varieties and explain some of the varietal differences. Another possible source of varietal differences is copy number variation.

Genome and transcriptome sequencing of the grapevine variety Tannat showed that many of the genes contributing to its high polyphenol content, including the genes coding for flavonoid hydroxylases from the CYP75 family, are not present in the PN40024 genome. We attempted to amplify putative *CYP76* varietal genes from Muscat Ottonel cDNA with degenerated primers designed on the most conserved *CYP76* regions, but failed to amplify any novel genes (data not shown). This result does, however, not exclude the presence of other *CYP76* genes in Muscat Ottonel or other genotypes. The dropping price of sequencing will without doubt soon result in genome and transcriptome data of other grapevine varieties. This will enable identification of varietal genes, involved in the biosynthesis of aroma and other compounds.

Plant cytochromes P450 are often involved in biosynthesis of defense compounds. Our gene expression analysis highlighted grapevine P450 genes that are upregulated in response to pathogen infection. Studies in other species have shown that many biosynthetic genes have "bloomed" and produced large gene families involved in the biosynthesis of species-specific antimicrobial and other defense compounds. The largest P450 family in grapevine is CYP82, which underwent two separate blooms in two different subfamilies. One of them, CYP82D, is preferentially expressed in leaves upon pathogen infection, whereas the other CYP82H, is expressed in green berries. We attempted to functionally characterize CYP82D13, which is expressed in ripe berries, but analysis of sequences from different grapevine varieties revealed that in most of them this gene bears a mutation in the codon coding for the most conserved amino acid in cytochromes P450, the heme-binding cysteine. This might either signify a gene inactivation or a dramatic change in catalytic activity.

Perspectives in flavor and fragrance biotechnology

Over the last years the market demand for natural flavors has been increasing, but the flavor molecules concentrations in the source materials are often too low to allow for a sustainable production. This has boosted the research in flavor functional genomics, since the identification of biosynthetic genes allows for the production of flavor compounds in bioreactor. Our study adds five more enzymes to the repertory of known monoterpenol-oxidizing cytochromes P450, reviewed in the **Chapter 4**. CYP76F14 produces the precursor of the flavor compound wine lactone, (*E*)-8-carboxylinalool. However, the chemical conversion of (*E*)-8-carboxylinalool to wine lactone is likely too slow for industrial use even at harsher conditions. A synthetic biology approach could be used to complement the grapevine biosynthetic pathway with a hypothetical (*E*)-8-carboxylinalool cyclizing enzyme and allow for the industrial production of wine lactone. We screened a library of bacterial squalene/hopene cyclase mutants, which enable cyclization of non-phosphorylated terpenes, such as geraniol, with several linalool derivatives (9). None of these mutants was able to

cyclize (*E*)-8-hydroxylinalool, (*E*)-8-oxolinalool or (*E*)-8-carboxylinalool (data not shown). Surprisingly, before wine lactone was detected in wine, it had been discovered as a metabolite in koala urine (10). Koalas feed on eucalyptus trees, which are rich in (*E*)-8-carboxylinalool conjugates (11). Koala liver might thus contain an enzyme capable of cyclizing (*E*)-8-carboxylinalool to wine lactone.

Finally, volatile precursors may be used as a smart material for slow and controlled release of flavor or aroma. For example, glycosylated aroma compounds have been proposed for applications in cosmetics, where the skin microbiome can catalyze the release of the volatile aglycone (12). Furthermore, the release of aroma from glycosylated precursors can be controlled by addition of glycosidase enzymes. Similar applications could be developed for *in situ* conversion non-glycosylated aroma precursors, such as (*E*)-8-carboxylinalool. Although (*E*)-8-carboxylinalool conversion to wine lactone in wines is very slow, it can be controlled by the amount of the precursor, pH, temperature or a presence of catalyst. A patent describes a use of a different wine lactone precursor in a hydrogen tank coating for leak detection (13). The precursor is covalently bound to a resin, and, upon contact with hydrogen, the precursor is released and converted to wine lactone. Wine lactone is suitable for these types of applications because of its low odor detection threshold.

References

1. Guth H (1997) Quantitation and Sensory Studies of Character Impact Odorants of Different White Wine Varieties. *J Agric Food Chem* 43:35–37.
2. Giaccio J, Curtin CD, Sefton MA, Taylor DK (2015) Relationship between menthiafolic acid and wine lactone in wine. *J Agric Food Chem* 63(37):8241–8246.
3. Bonnländer B, Baderschneider B, Messerer M, Winterhalter P (1998) Isolation of two novel terpenoid glucose esters from Riesling wine. *J Agric Food Chem* 8561(642):1474–1478.
4. Giaccio J, et al. (2011) The formation of wine lactone from grape-derived secondary metabolites. *J Agric Food Chem* 59(2):660–4.
5. Winterhalter P, Messerer M, Bonnländer B (1997) Isolation of the glucose ester of (E) -2,6-dimethyl-6-hydroxyocta-2,7-dienoic acid from Riesling wine. *Vitis* 36(1):55–56.
6. Guth H (1996) Determination of the configuration of wine lactone. *Helv Chim Acta* 79:1559–1571.
7. Luan F, Degenhardt A, Mosandl A, Wüst M (2006) Mechanism of wine lactone formation: demonstration of stereoselective cyclization and 1,3-hydride shift. *J Agric Food Chem* 54(26):10245–52.
8. Martin DM, et al. (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(1):226.
9. Hammer SC, Marjanovic A, Dominicus JM, Nestl BM, Hauer B (2014) Squalene hopene cyclases are protonases for stereoselective Brønsted acid catalysis. *Nat Chem Biol* 11(2):121–126.
10. Southwell IA (1975) Essential oil metabolism in the koala iii novel urinary monoterpenoid lactones. *Tetrahedron Lett* 16(24):1885–1888.
11. Goodger JQD, Woodrow IE (2011) α,β-Unsaturated monoterpene acid glucose esters: Structural diversity, bioactivities and functional roles. *Phytochemistry* 72(18):2259–2266.
12. Schwab W, Fischer TC, Giri A, Wüst M (2015) Potential applications of glucosyltransferases in terpene glucoside production: impacts on the use of aroma and fragrance. *Appl Microbiol Biotechnol* 99(1):165–

174.

13. Thorn DL, Jonietz KK, Boncella JM (2011) Coating for leak detection and method.

Appendix I

Expression of *Vitis vinifera* cytochrome P450 sequences

Gene expression data was retrieved from publically available RNA-Seq datasets and 8 unpublished datasets (**Chapter 5**). Datasets were divided into 5 categories and average gene expression (FPKM) was calculated for each category. Expression clusters correspond to the heatmap in **Chapter 5**, Figure 5. Abbreviations: "Chr" – chromosome, "Or" – orientation, "par" – partial gene, "ps" – pseudogene, "ps?" – putative pseudogene.

Name	Chr	Start	End	Or	Type	Flowers	Young berries	Ripe berries	Leaves (contr.)	Leaves (biotic stress)	Expr. cluster
CYP51G	13	18366040	18374499	+	gene	41.40	59.54	26.58	25.47	18.96	H
CYP51G6	1	23836524	23839703	-	gene	82.95	129.10	48.98	5.98	10.58	G
CYP71AH4	17	11510534	11512151	+	gene	0.00	0.00	0.00	0.00	0.00	AE
CYP71AP5	17	11481497	11483149	+	gene	19.80	14.24	7.75	13.58	19.86	AE
CYP71AS3	1	7047898	7050179	-	gene	0.57	1.07	2.12	1.24	5.40	N
CYP71AS4v1	1	7040703	7042665	-	gene	1.60	3.49	8.98	9.15	8.36	C
CYP71AS5	1	7027729	7030098	-	gene	1.03	2.70	3.69	0.76	4.31	T
CYP71AS6v1	1	7022625	7024610	-	gene	0.84	3.58	12.17	15.55	9.68	C
CYP71AS7v1	1	7005787	7007375	-	gene	8.73	8.60	2.34	16.61	14.02	W
CYP71AT3	18	11832431	11833989	-	gene	0.00	0.03	0.01	0.00	0.00	K
CYP71AT4	18	11801940	11803508	-	gene	0.06	0.01	0.01	0.00	0.00	F
CYP71AT5P	18	11794593	11796161	-	gene	0.20	0.01	0.01	0.03	0.02	R
CYP71AT7	18	11779385	11781003	-	gene	3.49	15.61	69.93	7.44	13.03	J
CYP71AT8	18	11774368	11775969	-	gene	0.08	0.02	1.59	1.56	9.67	N
CYP71AT9	18	11767728	11769317	-	gene	0.26	0.15	4.92	0.06	4.56	S
CYP71AU3	17	11416516	11418486	+	gene	0.05	0.45	3.06	0.46	0.34	J
CYP71AU4	17	11311172	11313097	-	gene	0.00	0.01	0.00	0.05	0.03	R
CYP71BC1	6	7205213	7206911	+	gene	5.41	8.22	1.77	8.37	17.14	V
CYP71BC3	6	7192216	7193925	+	gene	8.87	0.04	0.01	0.00	5.16	Z
CYP71BE1	8	3263145	3264737	+	gene	0.02	0.03	3.25	0.00	0.07	J
CYP71BE10v1	10	15601736	15603474	+	gene	0.02	0.01	0.02	0.00	0.01	E
CYP71BE13	10	15489011	15490811	+	gene	102.85	36.46	1.69	0.82	5.48	AA
CYP71BE5	19	8086568	8088361	-	gene	0.18	2.65	0.52	4.03	4.15	Q
CYP71BE6	19	8065226	8067199	-	gene	0.16	2.72	0.81	0.01	0.31	H
CYP71BE7	19	8052648	8054350	-	gene	0.02	0.00	0.00	0.00	0.01	E
CYP72A103	19	12496523	12499936	-	gene	6.05	67.53	36.69	135.68	82.78	Y
CYP72A105	19	13603264	13607320	+	gene	0.32	3.75	3.08	38.40	36.26	O
CYP72A107	19	12924169	12928065	+	gene	0.00	0.09	0.11	12.15	14.80	O
CYP72A108	19	13551182	13555209	+	gene	0.00	0.18	0.11	0.66	0.81	R
CYP72A110	19	13198362	13202257	+	gene	0.00	0.05	0.06	2.59	2.03	O
CYP72A113	19	12608034	12612007	+	gene	0.00	0.02	0.01	7.36	4.62	W
CYP72A85	19	11236194	11240749	+	gene	0.00	0.02	0.01	0.05	0.02	P
CYP72A88	19	17128817	17131872	-	gene	0.00	0.00	0.00	0.06	0.04	O
CYP72A89	19	17094959	17098064	-	gene	0.00	0.00	0.00	0.24	0.28	Y
CYP72A92	19	17018343	17027900	+	gene	10.61	35.49	27.96	61.06	78.33	Q
CYP72A93	19	16963761	16968293	+	gene	1.00	17.71	9.68	19.29	20.79	Q
CYP72A96	19	12274776	12280141	-	gene	0.00	0.00	0.00	0.00	0.00	outlier
CYP72D3	Un	13520477	13522989	+	gene	8.37	9.72	4.09	70.62	42.90	W
CYP72D5	Un	13525184	13527672	+	gene	13.97	14.40	2.53	27.92	15.77	W
CYP72D6	4	1479669	1482394	+	gene	1.11	1.17	1.23	1.66	2.88	N
CYP74A1	18	9911314	9912876	-	gene	17.20	26.51	63.56	12.84	31.77	J
CYP74A13	3	4246162	4247613	+	gene	1.55	0.02	0.03	0.00	0.14	Z
CYP74A14	3	4255743	4257194	+	gene	10.49	0.08	0.01	0.00	0.02	Z
CYP74A15	3	4264484	4265980	+	gene	0.17	0.07	0.02	0.00	0.01	E
CYP74A16	3	4276107	4277603	+	gene	0.27	0.70	0.21	0.54	0.62	R
CYP74A17	3	4282491	4283954	+	gene	0.23	0.05	1.76	0.00	0.06	J

CYP74B13	12	5973055	5975632	-	gene	171.30	217.82	457.56	99.18	286.90	J
CYP75A28	6	16843289	16845218	+	gene	0.21	4.11	6.98	4.25	2.95	L
CYP75A33	6	17182819	17184734	+	gene	0.00	0.02	0.01	0.09	0.03	P
CYP75A35	6	17444319	17446396	-	gene	0.00	0.91	0.04	0.00	0.00	K
CYP75A36	6	16923866	16925797	+	gene	0.15	0.09	0.16	0.86	0.33	R
CYP75A37P	6	16958423	16961597	+	gene	0.00	0.50	1.53	0.12	0.12	L
CYP75A41	6	17353534	17355859	+	gene	0.00	0.00	0.01	0.00	0.21	E
CYP75A43	8	19086144	19087784	+	gene	10.56	6.12	1.28	1.77	0.61	AB
CYP75A66	6	17133934	17135874	+	gene	1.29	0.60	1.66	1.07	0.27	B
CYP75B32v1	17	8143553	8145765	+	gene	105.07	171.05	62.25	66.02	50.43	H
CYP75B38	17	8167015	8169119	+	gene	10.77	13.07	5.86	3.13	2.82	G
CYP76A10	8	18106975	18108596	-	gene	0.00	0.00	0.01	0.00	0.01	A
CYP76A12	8	18099069	18101173	-	gene	0.74	0.21	0.12	0.30	0.11	R
CYP76A13	8	18077891	18079993	-	gene	3.16	0.73	0.52	0.89	0.68	Z
CYP76A15	8	18056637	18058400	-	gene	1.51	0.08	0.00	0.00	0.00	AB
CYP76A16	8	18038159	18039776	-	gene	0.80	0.33	0.11	0.05	0.01	R
CYP76A43	8	18045974	18047737	-	gene	8.25	0.26	0.04	0.00	0.03	Z
CYP76F10	2	9407260	9410219	+	gene	3.85	4.58	2.28	0.08	1.60	G
CYP76F12	2	4192069	4193694	-	gene	8.52	8.80	29.57	41.53	22.09	C
CYP76F14	2	4211828	4213481	+	gene	5.61	17.58	102.13	42.08	130.77	S
CYP76G6	8	18120170	18121816	-	gene	0.13	0.01	0.01	0.02	0.00	R
CYP76T10	15	15741636	15744038	-	gene	0.45	0.29	0.04	0.96	0.39	R
CYP76T11	15	15737556	15740218	-	gene	5.72	5.21	0.56	6.86	11.96	V
CYP76T12	15	15732111	15733934	-	gene	6.23	37.92	2.01	9.16	8.69	H
CYP76T15	15	15688774	15691309	-	gene	0.73	0.07	0.25	6.86	14.83	O
CYP76T20	15	15595607	15598545	+	gene	0.03	0.04	0.00	0.00	0.00	F
CYP76T21	15	15609892	15612071	+	gene	6.49	1.98	14.36	32.66	7.32	X
CYP76Y1	3	10833717	10835596	+	gene	0.28	0.70	6.81	21.97	34.71	O
CYP76Y15	2	10945511	10947340	-	gene	1.05	1.70	0.48	0.00	0.23	G
CYP76Y2	2	10897022	10898851	-	gene	2.52	3.73	13.52	0.66	0.56	J
CYP76Y3	2	12936358	12938671	-	gene	1.29	0.89	1.26	0.46	0.37	A
CYP77A14	13	8130065	8131612	-	gene	113.83	94.28	25.29	7.49	20.02	AB
CYP77B6	10	364438	365964	+	gene	67.93	1.75	0.25	0.15	0.24	Z
CYP78A36	1	381267	382921	-	gene	0.78	0.97	1.21	6.11	1.63	X
CYP78A37	1	22106024	22109360	+	gene	1.66	4.31	0.31	0.88	3.12	U
CYP78A38	2	2044344	2046005	-	gene	5.37	1.22	8.48	10.98	9.18	D
CYP78A39	7	23993183	23994886	-	gene	35.19	0.15	0.47	0.00	0.01	Z
CYP78A40	15	17005434	17007131	+	gene	8.47	0.09	0.07	1.29	1.73	Z
CYP78A41	17	5732157	5734202	-	gene	59.96	366.44	1417	10.97	7.86	J
CYP78A42	18	21535124	21537036	+	gene	0.01	0.04	0.00	0.00	0.01	U
CYP79A15	13	1541967	1543728	-	gene	0.00	0.00	0.00	0.01	0.04	AD
CYP79A17	13	1527055	1528816	-	gene	0.00	0.00	0.00	0.00	0.00	outlier
CYP79A19	13	1519993	1521754	-	gene	0.00	0.00	0.00	0.00	0.00	outlier
CYP79A21v1	13	1467798	1469559	-	gene	0.00	0.00	0.00	0.00	0.00	AE
CYP79A23	13	1456751	1458509	-	gene	0.01	0.02	0.00	0.00	0.00	F
CYP79A24	13	9944010	9945774	+	gene	0.10	0.00	0.25	0.00	0.03	A
CYP79A26	6	17440343	17442119	+	gene	0.23	0.02	0.35	0.00	0.11	E
CYP79A27	6	14924997	14926722	-	gene	0.10	0.00	0.01	0.00	0.01	E
CYP80E11	2	3326398	3328772	-	gene	1.99	13.88	12.33	11.76	18.99	P
CYP80E3	2	3336426	3338880	-	gene	0.84	8.01	9.09	15.06	13.92	Q
CYP80E5	2	3646097	3647904	-	gene	0.05	1.52	0.77	0.00	0.06	G
CYP80K1	16	17550032	17551625	+	gene	1.21	0.00	0.00	0.00	0.00	AB
CYP80K2	16	17555607	17557181	+	gene	0.00	0.00	0.00	0.00	0.00	Z
CYP80K3	16	17539829	17541415	+	gene	33.78	1.37	0.03	0.00	0.03	Z
CYP81B26	18	7959549	7961518	-	gene	0.20	0.32	0.09	0.82	4.42	N
CYP81B27	18	7956240	7958165	-	gene	8.84	37.44	27.67	12.58	48.61	T
CYP81B29	18	7943100	7953002	-	gene	11.70	13.11	8.72	0.23	1.40	G
CYP81B31	9	6420606	6422261	+	gene	0.03	0.00	0.00	0.02	0.02	D

CYP81B32	9	6390839	6392566	-	gene	0.02	0.00	0.01	0.00	0.08	A
CYP81B34	9	6372932	6374662	-	gene	0.16	0.01	0.00	0.00	0.03	U
CYP81B81	9	6425981	6427661	+	gene	0.02	0.00	0.00	0.31	0.08	AD
CYP81Q12	7	22335270	22337246	+	gene	0.27	0.02	0.01	0.00	0.00	F
CYP81Q15	7	22324027	22325725	+	gene	0.04	0.10	0.38	0.00	0.17	E
CYP81Q19	7	22370472	22372250	+	gene	7.05	16.99	24.07	8.82	10.86	K
CYP81Q20	7	22368071	22369733	+	gene	0.42	0.10	0.24	3.29	2.33	W
CYP81Q23	7	22291738	22293436	+	gene	0.11	0.08	0.16	0.06	0.12	R
CYP81Q24	7	22364071	22366219	+	gene	0.02	0.01	0.00	0.00	0.00	U
CYP81Q26	14	6899413	6901418	-	gene	0.00	0.00	0.01	0.00	0.01	E
CYP81Q29	18	8053061	8054991	+	gene	0.02	0.01	0.23	0.09	0.39	R
CYP81Q68	Un	20111880	20113598	-	gene	0.80	1.92	3.59	14.90	48.12	O
CYP81Q7	7	22273978	22275621	+	gene	23.19	59.64	18.84	11.75	16.43	G
CYP81Q8	7	22264821	22266514	+	gene	0.06	0.29	0.01	0.00	0.01	E
CYP81Q9	7	22260680	22262324	+	gene	0.02	0.10	0.01	0.00	0.00	F
CYP81W1	18	8045718	8048970	+	gene	22.13	30.11	21.11	84.29	125.68	O
CYP82D10	18	9794763	9797055	+	gene	0.03	0.02	1.15	0.02	1.02	S
CYP82D13	18	9804817	9807375	+	gene	0.04	1.05	32.14	3.44	4.34	J
CYP82D14	18	9699734	9701502	+	gene	0.11	0.00	0.00	0.03	0.01	AD
CYP82D15	18	9706002	9708470	+	gene	0.00	0.00	0.07	0.00	0.00	S
CYP82D17	18	9728437	9730964	+	gene	0.12	1.89	1.45	55.48	113.42	O
CYP82D18	18	9737764	9740222	+	gene	0.00	0.00	0.13	0.00	0.00	F
CYP82D19	7	18198115	18200236	-	gene	0.07	0.04	0.63	0.42	6.45	N
CYP82D20v2	18	9756250	9758781	+	gene	0.56	1.23	1.27	80.52	142.72	O
CYP82D4	18	9827989	9830254	+	gene	0.01	0.00	0.00	0.00	0.00	A
CYP82D5	18	9836516	9838612	+	gene	0.11	0.05	0.01	0.00	0.05	E
CYP82D6	18	9846632	9848417	+	gene	0.28	1.82	2.08	2.87	19.66	N
CYP82D7	18	9865562	9867293	+	gene	1.61	4.25	3.41	51.22	148.20	O
CYP82D9	18	9789032	9791275	+	gene	0.19	0.26	0.30	0.49	5.98	N
CYP82H14a	3	4510174	4512089	-	gene	1.40	35.82	4.27	1.14	3.03	H
CYP82H18	3	4421629	4423866	-	gene	0.65	0.72	0.08	0.15	0.09	R
CYP82H22	3	4401302	4403199	-	gene	22.94	44.74	13.59	3.29	1.01	G
CYP82H5	3	4464581	4466474	-	gene	3.69	7.91	2.25	3.90	1.86	H
CYP82H8	3	4480778	4482691	-	gene	0.79	2.83	0.12	0.42	0.52	H
CYP82S1	12	21576153	21578134	-	gene	0.61	0.03	0.01	0.02	0.13	R
CYP82S12	9	10967030	10971388	-	gene	0.08	0.06	0.01	0.00	0.01	E
CYP82S15v1	9	11038532	11043816	-	gene	0.27	0.21	0.03	0.01	0.41	R
CYP82S3	12	21514206	21516333	-	gene	0.12	0.08	0.01	0.03	2.20	N
CYP82S6	4	24105688	24107881	+	gene	0.92	0.70	0.70	14.89	18.68	O
CYP82S7	4	24114649	24116860	+	gene	0.12	0.04	0.03	0.03	0.18	R
CYP82S9	Un	22322297	22324147	+	gene	0.17	0.05	0.02	0.00	0.12	E
CYP84A30	3	510098	511819	+	gene	0.05	0.19	0.25	0.00	0.00	F
CYP84A31	4	19859411	19861121	+	gene	2.44	3.54	7.70	21.00	41.18	O
CYP84A32v1	3	457065	458789	+	gene	0.25	1.63	2.02	0.00	0.00	K
CYP85A1	14	23435565	23438398	-	gene	7.47	99.05	36.10	3.97	19.19	I
CYP85A10	1	284363	287452	-	gene	1.69	0.32	0.02	0.00	0.02	Z
CYP85A11P	1	263082	281163	-	gene	0.00	0.01	0.00	0.00	0.00	K
CYP86A28	15	19173918	19175594	+	gene	9.92	2.20	0.95	1.08	2.11	Z
CYP86A29	2	2641688	2643325	+	gene	60.51	68.08	30.35	8.14	28.66	G
CYP86A30	6	7005672	7007231	+	gene	0.38	1.13	0.95	0.08	0.05	H
CYP86B7	1	1748732	1751009	+	gene	8.17	1.50	2.26	2.41	4.68	Z
CYP86C10	1	6936328	6937908	+	gene	4.64	17.22	2.09	10.02	38.01	V
CYP86C8	3	17994179	17995744	-	gene	0.24	0.00	0.00	0.00	0.00	A
CYP87A12	16	17599422	17601831	+	gene	1.36	0.01	0.00	0.00	0.00	AB
CYP87A13	16	17607035	17610511	+	gene	0.29	0.01	0.00	0.00	0.00	E
CYP87A14	2	3403487	3405967	+	gene	1.39	0.10	0.11	0.09	0.13	Z
CYP87A50	15	16079897	16088218	+	gene	1.20	0.17	0.09	2.97	20.07	N
CYP87B25	18	1054808	1057685	-	gene	2.70	0.10	0.15	11.22	63.53	N

CYP87B26	18	1007964	1010738	+	gene	53.91	0.00	0.00	0.02	0.03	Z
CYP87B27	18	1039533	1042444	-	gene	0.08	0.04	0.15	0.15	34.15	N
CYP87B6	4	21366769	21368843	-	gene	0.00	0.03	0.00	0.00	0.00	H
CYP87B8	4	21374128	21376321	-	gene	0.00	1.54	0.28	0.54	1.19	P
CYP87B9	18	1084914	1087799	-	gene	0.35	0.14	0.05	22.74	85.42	O
CYP88A23	15	9799164	9801559	-	gene	0.00	1.83	0.01	0.00	0.00	I
CYP88A24	15	9763285	9768505	-	gene	9.30	3.25	0.79	0.52	0.60	AA
CYP89A38	16	503395	504960	+	gene	6.41	6.28	6.29	18.26	24.92	O
CYP89A39	16	506303	507868	+	gene	15.69	7.22	5.29	108.58	108.71	O
CYP89A41	16	494863	496464	+	gene	13.58	17.96	16.49	64.28	47.75	W
CYP89A44	16	455290	456840	+	gene	0.13	0.07	0.00	0.04	2.98	N
CYP89A45	16	479387	480916	+	gene	0.02	0.31	0.17	0.05	0.02	R
CYP89A46	16	484984	486636	+	gene	1.03	0.18	0.26	0.24	0.45	Z
CYP89A47	16	490073	491623	-	gene	1.06	0.03	9.20	1.92	0.49	J
CYP89A48	16	575469	577025	-	gene	8.13	6.82	4.00	13.13	22.31	O
CYP89A52	16	511725	513281	+	gene	0.67	0.44	0.99	2.03	2.07	O
CYP89A53	16	401789	403348	+	gene	0.87	1.60	1.39	0.03	0.68	K
CYP89A54	16	439624	441177	+	gene	0.31	5.25	0.85	0.16	0.05	H
CYP89A58	19	23346184	23347689	-	gene	0.02	0.07	0.06	0.01	0.01	R
CYP90A16	13	360424	363684	+	gene	43.45	38.58	143.22	96.62	72.86	C
CYP90B12	4	18533795	18537529	+	gene	10.26	15.52	4.69	5.30	22.02	V
CYP90C5	4	19556221	19560597	+	gene	10.57	36.83	4.18	5.59	10.61	H
CYP90D8	9	1887505	1891994	-	gene	4.42	3.15	0.52	3.76	2.72	AC
CYP92A34	11	833917	835545	+	gene	0.98	0.23	0.01	0.28	0.07	R
CYP92A36	11	856265	857896	+	gene	65.24	116.99	6.32	30.40	28.33	H
CYP92A38	11	865734	867363	+	gene	0.84	0.05	0.02	0.08	0.03	R
CYP92A39	9	829435	831954	+	gene	22.42	20.18	21.22	249.92	175.08	W
CYP93A16	8	21172463	21174755	-	gene	0.34	0.22	0.13	0.31	0.59	R
CYP93A17	8	21179266	21180990	-	gene	0.06	25.14	0.65	0.96	0.94	H
CYP93A18	8	21184526	21186221	-	gene	2.20	1.63	8.77	0.00	0.98	J
CYP93A9	8	21190543	21192196	-	gene	0.02	0.01	0.05	0.18	0.09	R
CYP94A15	7	196513	198030	-	gene	0.07	0.29	1.21	9.08	5.49	W
CYP94A16	7	185744	187264	-	gene	4.71	13.23	3.00	27.02	19.71	W
CYP94B10	17	1192477	1194015	+	gene	0.47	9.81	0.42	0.00	0.05	H
CYP94B11	14	25721500	25723110	+	gene	1.23	0.86	7.02	0.16	0.81	J
CYP94C15v1	13	9347916	9349436	-	gene	1.18	0.35	0.72	0.02	0.29	Z
CYP94C16	6	20926941	20928455	-	gene	1.30	4.70	1.39	0.79	0.57	G
CYP94C17	8	17605079	17606608	-	gene	3.68	1.03	0.14	1.15	1.21	Z
CYP94D25	8	11777899	11779401	+	gene	14.41	13.95	3.14	41.75	21.90	W
CYP94F2	6	18030826	18032412	+	gene	22.90	11.38	5.73	10.13	29.59	AE
CYP96A18	7	24088359	24089867	-	gene	9.71	15.44	10.94	18.20	21.72	Q
CYP96A28	7	24108258	24109781	-	gene	12.45	59.70	4.53	2.92	13.68	H
CYP96A30	7	24096876	24098399	-	gene	20.23	0.02	0.01	0.05	0.02	Z
CYP97A11	4	16300756	16323693	+	gene	26.44	31.10	30.32	13.04	16.96	F
CYP97B15	3	19697017	19725026	-	gene	22.06	13.95	7.06	26.07	28.17	AD
CYP97C12	8	18452646	18457877	-	gene	23.39	15.36	5.02	27.80	28.52	AD
CYP98A43	8	11733570	11736439	-	gene	72.13	70.30	17.13	69.98	25.10	AC
CYP701A23	18	9625486	9628671	-	gene	10.32	12.82	3.93	1.97	4.45	G
CYP703A9	15	17343706	17345623	+	gene	0.01	0.03	0.01	0.01	0.01	R
CYP704A121	15	15882456	15885296	-	gene	3.78	0.85	0.80	0.02	1.13	Z
CYP704A19	15	15906789	15909327	-	gene	15.57	9.97	6.75	163.20	206.09	O
CYP704A20	15	15892125	15894969	-	gene	1.50	0.69	0.78	0.00	0.56	Z
CYP704A21v2	15	15869209	15872056	-	gene	6.21	1.93	1.48	2.42	8.53	AE
CYP704A22	16	16087902	16090810	+	gene	1.59	0.78	1.56	3.24	2.36	W
CYP704A23	16	16103541	16105639	+	gene	0.00	0.06	0.17	0.01	0.05	R
CYP704B21	1	12831204	12833528	-	gene	6.74	0.08	0.25	0.00	0.04	Z
CYP706C7	2	4648241	4650489	-	gene	0.82	0.03	0.23	11.75	6.67	W

CYP706C8	5	25119017	25121047	+	gene	0.02	0.00	0.01	0.00	0.00	A
CYP706G1	5	25126474	25128301	+	gene	9.61	1.45	1.60	1.24	2.74	Z
CYP706H1	16	14890580	14893448	+	gene	1.82	4.10	0.25	48.56	28.45	W
CYP706J3	16	15070282	15072491	-	gene	0.00	0.00	0.00	0.00	0.00	H
CYP707A38	3	5493454	5496019	-	gene	13.49	1.16	0.20	11.22	12.19	AD
CYP707A39	7	23186058	23189004	+	gene	0.01	0.16	0.12	0.00	0.01	E
CYP707A40	18	8847866	8850489	+	gene	7.19	21.94	3.10	36.08	8.15	Y
CYP707A41	2	18361159	18363573	-	gene	1.75	2.05	2.12	15.33	7.65	W
CYP707A42	6	5973687	5977447	-	gene	0.25	0.02	0.05	1.04	0.07	X
CYP709B5	7	5228639	5231228	-	gene	2.77	0.25	0.06	0.00	22.41	N
CYP710A18	10	8866057	8867586	-	gene	7.67	2.47	0.52	1.06	1.30	AA
CYP711A14	4	916173	919492	-	gene	0.67	2.82	0.49	1.95	5.35	V
CYP712A12	12	6865075	6866746	+	gene	0.02	0.00	0.01	0.00	0.00	A
CYP712D1	13	27872873	27875625	-	gene	9.00	12.31	3.80	3.13	6.37	G
CYP714A8	13	98215	101623	+	gene	23.56	62.01	9.98	7.72	4.24	G
CYP714E12	15	11091518	11095999	+	gene	9.78	17.08	8.85	0.14	0.22	G
CYP714E8	15	11039737	11043505	+	gene	3.46	4.30	0.82	0.00	0.21	G
CYP714F2v1	10	19305012	19307177	+	gene	0.17	0.37	0.00	0.07	0.08	R
CYP714G5	3	12677869	12679890	+	gene	1.86	3.94	4.03	23.40	26.08	O
CYP714G6	18	33740007	33742019	+	gene	0.13	0.25	0.10	13.91	15.81	O
CYP715A6	16	18863490	18866569	+	gene	0.03	0.18	1.76	0.00	0.02	J
CYP716A15	11	11819497	11821972	+	gene	4.62	153.79	27.41	137.11	73.07	Y
CYP716A17	11	11497296	11499344	+	gene	0.46	3.11	0.34	1.20	0.59	H
CYP716A18	18	22742900	22746514	-	gene	6.63	1.15	0.09	0.14	0.10	AA
CYP716A19	4	2470501	2472554	+	gene	0.04	0.00	0.00	0.00	0.11	U
CYP716A20	11	12916608	12918493	+	gene	0.14	2.22	0.04	0.00	0.00	G
CYP716C4	10	14697235	14698833	-	gene	0.00	0.01	0.00	0.01	0.02	Y
CYP716E12	18	29734314	29736327	-	gene	0.00	0.40	1.30	0.00	0.00	K
CYP716E13	18	29727686	29729919	-	gene	0.00	0.60	0.02	0.00	0.00	K
CYP716E16	18	30150782	30152498	+	gene	0.31	0.02	0.01	0.00	0.00	E
CYP716E17	18	29831533	29833367	+	gene	0.01	1.01	0.14	0.00	0.00	G
CYP716E20	15	3854845	3856658	-	gene	0.11	0.36	0.00	0.00	0.00	F
CYP716E21	18	29838175	29839989	-	gene	2.64	0.15	0.10	0.00	0.00	AA
CYP720A1	16	140769	143178	-	gene	0.73	0.05	0.00	0.07	0.07	R
CYP721A13	18	10381889	10384207	-	gene	0.64	0.25	0.15	0.54	0.88	R
CYP721A15	18	10370740	10373021	-	gene	0.29	3.91	0.43	0.20	0.55	H
CYP721A16	18	10363578	10366849	-	gene	10.25	5.06	1.18	2.84	5.11	Z
CYP721A18	18	10346531	10348852	-	gene	1.64	1.32	0.69	7.72	32.38	N
CYP722A1	18	10968173	10971522	-	gene	1.93	0.18	0.02	1.19	0.55	B
CYP722C1 ortholog	4	19254663	19258281	-	gene	0.49	1.93	1.07	0.08	10.20	N
CYP724B13	12	9778144	9780799	-	gene	9.64	0.06	0.03	0.03	0.18	Z
CYP724B14	14	26743473	26745706	+	gene	6.13	3.87	4.56	17.89	12.79	W
CYP727A7	13	21804250	21822377	-	gene	14.31	10.11	9.53	32.40	28.52	W
CYP728B15	10	1408056	1410051	+	gene	0.06	0.04	0.00	0.00	0.69	U
CYP728B6	19	6307747	6309528	-	gene	0.80	0.08	0.63	0.11	0.10	R
CYP728B8	10	1413013	1415051	+	gene	0.09	0.07	0.00	0.38	1.28	V
CYP728G1	13	1550154	1552089	-	gene	0.00	0.00	0.00	0.00	0.00	outlier
CYP73A78	11	15123526	15125254	+	gene	0.10	0.31	1.38	0.00	0.24	J
CYP73A81	11	12482399	12484118	-	gene	2.03	3.11	16.22	0.51	5.08	J
CYP73A82	6	8882615	8885897	-	gene	373.60	404.67	187.23	144.55	108.97	G
CYP733A1	18	14491507	14493912	-	gene	0.38	0.34	1.11	0.29	0.52	J
CYP734A13	11	4101725	4104591	+	gene	4.58	11.46	2.04	0.45	8.94	U
CYP734A15	9	5785961	5788706	+	gene	4.35	7.72	1.54	0.67	2.16	G
CYP735A12	14	21269651	21271982	-	gene	1.04	0.65	0.56	0.05	0.79	Z
CYP736A17	17	4024600	4027523	-	gene	9.65	0.61	0.22	9.33	2.48	B
CYP736A20	17	3986773	3989015	-	gene	0.04	0.05	0.00	0.01	2.45	N
CYP736A21	17	3978890	3980556	-	gene	0.65	0.11	0.01	25.48	2.06	X

CYP736A22	17	3960581	3963005	-	gene	4.95	9.22	1.33	4.24	4.90	H
CYP736A25v1	7	23789074	23790629	+	gene	1.08	0.55	0.78	7.96	50.54	N
CYP736A26	7	23792052	23793692	+	gene	0.00	0.00	0.00	0.18	0.25	D
CYP706C42P	Un	25955860	25956180	-	par	0.01	0.00	0.00	2.76	0.98	W
CYP706C8-de2b	2	4641194	4641698	-	par	0.00	0.00	0.00	0.00	0.00	outlier
CYP706G14	5	25132926	25144134	+	par	0.17	0.49	0.97	0.86	4.70	N
CYP706G2	Un	24349799	24351490	-	par	1.79	2.65	4.04	5.70	21.85	N
CYP706J1	16	15129842	15132139	-	par	0.00	0.00	0.00	0.00	0.00	outlier
CYP72A109	19	18868716	18872610	+	par	0.00	0.24	0.11	13.50	9.41	W
CYP72A111P	19	13176940	13178921	+	par	0.00	0.04	0.00	1.20	0.84	Y
CYP72A99	19	11317773	11332874	-	par	0.23	0.49	0.40	0.02	0.07	R
CYP75A34	6	17233375	17235650	+	par	0.01	0.03	0.06	0.00	0.17	E
CYP75A38v3 fragment	6	16839638	16841240	+	par	0.11	1.68	2.80	4.72	3.84	Q
CYP75A38v3 fragment	6	16888832	16889203	+	par	0.00	0.08	0.08	0.00	0.02	T
CYP75A39Pv1	6	16790972	16791875	+	par	0.21	0.53	0.76	0.70	1.00	R
CYP75A65P	6	16795214	16796903	+	par	1.11	3.81	4.97	4.63	0.95	M
CYP79A97P	6	16808866	16809255	-	par	0.00	0.02	0.00	0.00	0.13	U
CYP81Q27	2	5697654	5698052	+	par	1.83	2.52	5.97	4.81	10.00	N
CYP81Q27P	Un	19293144	19295125	-	par	2.73	3.42	8.30	6.95	13.01	N
CYP82H29P	3	4408518	4419784	-	par	0.00	0.00	0.00	0.00	0.00	E
CYP82S10	7	15002340	15002747	+	par	0.00	0.00	0.02	0.05	0.00	C
CYP82S8	7	15032655	15033104	+	par	0.09	0.19	0.08	0.00	0.10	E
CYP96A27	14	18153311	18154738	+	par	1.83	0.01	0.00	0.00	0.01	Z
CYP706J5	16	15042801	15045000	-	ps?	0.00	0.00	0.00	0.00	0.00	outlier
CYP71BE12	10	15636913	15638660	+	ps?	2.30	0.03	0.01	0.00	0.00	Z
CYP714F7P	10	19415029	19417130	-	ps?	0.65	0.19	0.02	0.00	0.09	E
CYP716E22	18	30173601	30175611	-	ps?	4.60	0.00	0.00	0.01	0.00	AA
CYP72A95P	19	16870273	16875654	+	ps?	0.00	0.00	0.00	0.00	0.01	E
CYP728B7P	19	6327712	6329567	-	ps?	1.30	0.26	0.94	6.29	7.52	O
CYP75A38v2	6	17075239	17083168	+	ps?	0.67	0.46	0.56	0.09	0.08	R
CYP81Q18P	7	22308571	22310389	+	ps?	1.02	1.63	0.35	0.04	0.22	AB
CYP82D126P	18	9762307	9784672	+	ps?	0.01	0.02	0.18	0.07	0.72	R
CYP82S5Pv1	12	21496352	21499253	-	ps?	0.00	0.00	0.00	0.00	0.00	Z
CYP96A25	7	24083515	24085024	-	ps?	1.69	1.47	1.76	1.73	3.59	N
CYP51	9	21009310	21009468	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP71	1	6998483	7003367	-	ps	0.31	0.15	0.04	0.03	0.06	R
CYP71	10	12802550	12825281	+	ps	0.67	1.37	9.13	0.00	0.00	J
CYP71	10	15484875	15485357	+	ps	0.10	0.01	0.00	0.00	0.00	AB
CYP71	10	15557846	15564545	+	ps	0.47	0.01	0.00	0.00	0.00	AB
CYP71	10	15617435	15625455	+	ps	0.18	0.01	0.00	0.00	0.01	U
CYP71	15	10505471	10505620	+	ps	0.00	0.00	0.00	0.00	0.24	N
CYP71	17	11333952	11335880	+	ps	0.44	0.90	0.72	135.26	72.81	W
CYP71	17	11461406	11463348	+	ps	0.00	0.00	0.00	0.00	0.00	N
CYP71	17	11501663	11503471	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP71	18	11761743	11763899	-	ps	0.02	0.00	0.01	0.00	0.03	A
CYP71	18	11784360	11786092	-	ps	0.36	0.19	0.03	2.64	2.31	O
CYP71	18	23686363	23688848	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP71	19	8035825	8043860	-	ps	0.62	0.22	0.03	0.00	0.01	E
CYP71	19	8070325	8070579	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP71	19	8099578	8104810	-	ps	2.12	2.31	0.79	2.07	2.34	AC
CYP71	19	8102748	8103653	-	ps	0.07	0.09	0.09	0.00	0.08	E
CYP71	19	15915999	15916199	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP71	19	21762696	21762887	+	ps	1.28	1.57	1.00	6.01	13.26	O
CYP71	19	22085890	22086081	+	ps	2.03	1.46	0.86	5.67	7.53	O
CYP71	4	23027748	23028889	-	ps	0.01	0.01	0.04	0.00	0.02	E

CYP71	5	18194882	18195031	-	ps	0.00	0.00	0.07	0.00	0.00	J
CYP71	6	7181152	7181748	+	ps	5.13	0.12	0.01	0.00	0.22	Z
CYP71	6	7188737	7188943	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP71	6	7200813	7201016	+	ps	0.00	0.00	0.00	0.00	0.01	N
CYP71	6	7203558	7203687	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP71	8	3227289	3228820	-	ps	0.01	0.00	0.00	0.00	0.22	U
CYP71	8	3294600	3296493	+	ps	0.16	0.10	0.04	0.48	1.20	N
CYP72	11	1037048	1039454	-	ps	1.41	2.61	2.10	0.17	0.38	F
CYP72	19	11363419	11365229	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP72	19	11491860	11505082	-	ps	0.01	0.06	0.03	0.00	0.28	E
CYP72	19	11542515	11553113	-	ps	0.00	0.00	0.00	0.00	0.00	J
CYP72	19	11749525	11749755	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP72	19	11750774	11751052	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP72	19	11929910	11936584	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP72	19	12089424	12090654	-	ps	0.00	0.01	0.00	0.00	0.01	U
CYP72	19	12471722	12472487	-	ps	0.01	0.09	0.10	0.00	0.01	E
CYP72	19	12659626	12662564	+	ps	0.00	0.01	0.00	1.81	1.52	Q
CYP72	19	12895095	12895531	-	ps	0.00	0.00	0.00	0.00	0.02	N
CYP72	19	13514979	13515415	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP72	19	13642480	13645186	+	ps	0.52	3.33	2.36	4.08	4.81	Q
CYP72	19	13653034	13653818	-	ps	0.00	0.02	0.00	0.00	0.00	H
CYP72	19	16900351	16911180	+	ps	0.02	0.02	0.01	0.00	0.01	E
CYP72	19	17149886	17150165	-	ps	0.00	0.00	0.00	0.40	0.28	O
CYP72	19	17155186	17157835	-	ps	0.00	0.00	0.00	0.01	0.01	O
CYP73	7	11748277	11748525	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP73	8	19258073	19258288	-	ps	0.35	0.53	0.14	0.42	0.23	R
CYP75	6	17029031	17029192	+	ps	0.00	0.12	0.24	0.16	0.07	P
CYP75	6	17039852	17041162	+	ps	0.99	0.74	1.93	2.25	0.48	X
CYP75	6	17173264	17175176	+	ps	0.42	2.53	8.31	0.77	0.28	J
CYP75	6	17178251	17180108	+	ps	0.00	0.00	0.01	0.00	0.01	S
CYP75	6	17228910	17229132	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP75	6	17230884	17230930	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP75	6	17231742	17232638	+	ps	0.00	0.01	0.01	0.00	0.05	T
CYP75	6	17351383	17352663	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP76	15	15572751	15574418	+	ps	0.13	0.01	0.05	0.04	0.01	R
CYP76	15	15635790	15641299	-	ps	7.78	0.01	0.01	0.00	0.00	Z
CYP76	15	15655039	15657376	-	ps	38.90	1.21	0.06	0.07	19.80	Z
CYP76	15	15687544	15687705	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP76	15	15696741	15697813	-	ps	0.06	0.03	0.00	0.06	0.08	AC
CYP76	15	15703653	15704147	-	ps	0.00	1.09	0.04	0.33	1.60	P
CYP76	15	15729977	15730267	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP76	15	15730383	15730532	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP76	15	15748389	15750981	-	ps	0.03	0.01	0.00	0.00	0.01	U
CYP76	15	15761436	15761968	-	ps	0.64	0.49	0.01	0.00	0.08	E
CYP76	15	15762606	15762773	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP76	15	15775408	15775722	-	ps	0.00	0.06	0.01	0.08	0.09	P
CYP76	15	15776163	15776714	-	ps	0.62	0.12	0.00	0.00	0.06	E
CYP76	15	15781236	15781415	-	ps	0.39	0.02	0.00	0.00	0.00	AB
CYP76	2	4206572	4208527	+	ps	1.68	0.28	3.47	0.04	0.76	J
CYP76	2	4215729	4217381	+	ps	2.28	4.74	20.44	1.27	13.03	S
CYP76	2	9431620	9433136	+	ps	0.00	0.01	0.03	0.00	0.00	T
CYP76	2	12929477	12931790	-	ps	0.00	0.00	0.01	0.02	0.00	M
CYP76	8	18045443	18045610	-	ps	0.19	0.10	0.05	0.00	0.00	F
CYP76	8	18050466	18051495	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP76	8	18056106	18056276	-	ps	0.08	0.00	0.00	0.00	0.00	Z
CYP76	8	18061137	18065782	-	ps	0.00	0.00	0.00	0.00	0.00	N
CYP78	1	19495236	19495586	-	ps	0.01	0.07	0.17	0.00	0.01	E
CYP78	15	12242009	12242440	+	ps	0.00	0.01	0.04	0.00	0.00	K

CYP78	16	6461201	6461380	+	ps	0.00	2.84	0.60	5.65	1.09	Y
CYP78	16	6495720	6495899	+	ps	0.04	16.71	2.58	111.33	15.34	X
CYP78	17	15926210	15926458	-	ps	0.00	0.03	0.07	0.00	0.00	K
CYP78	19	10947850	10948997	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP78	19	10961752	10962585	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP78	Un	8425555	8425839	-	ps	0.15	0.08	0.19	0.00	0.00	F
CYP78	Un	25122006	25122299	-	ps	0.34	0.45	1.58	0.16	0.03	J
CYP79	13	1462698	1463898	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP79	13	1509614	1510117	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP79	13	1522766	1523331	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP79	13	1533302	1533805	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP79	6	14924703	14924780	+	ps	0.05	0.00	0.00	0.00	0.00	Z
CYP79	6	16797881	16798285	-	ps	0.10	0.29	0.15	10.18	10.71	O
CYP79	6	16846244	16888762	-	ps	0.05	0.07	0.06	0.71	0.66	R
CYP79	6	16889828	16902090	-	ps	0.07	0.10	0.10	5.49	3.16	W
CYP79	6	16968377	16968810	-	ps	0.00	0.00	0.00	0.00	0.03	N
CYP79	6	16985870	16986142	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP79	6	16989296	16989980	-	ps	0.00	0.00	0.00	0.00	0.01	N
CYP79	6	17042627	17054757	-	ps	0.01	0.01	0.04	0.73	0.29	R
CYP79	6	17058028	17058494	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP79	6	17085430	17092527	-	ps	0.00	0.02	0.00	0.00	0.02	U
CYP79	6	17112372	17116456	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP79	6	17140982	17173186	-	ps	0.00	0.00	0.00	0.10	0.05	AD
CYP79	6	17391147	17392925	+	ps	0.07	0.12	22.66	0.00	0.09	J
CYP80	12	5975633	5975782	-	ps	1.19	0.42	1.92	0.14	0.19	A
CYP80	2	3650500	3652330	-	ps	0.00	0.07	0.02	0.00	0.00	K
CYP80	2	3653947	3654366	-	ps	0.00	0.02	0.00	0.00	0.00	H
CYP80	4	6224782	6224931	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	13	16689102	16689299	+	ps	0.00	0.02	0.01	0.00	0.00	K
CYP81	16	5141828	5141980	+	ps	0.02	0.00	0.06	0.00	0.00	A
CYP81	16	5206567	5206719	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	17	11135422	11135634	-	ps	0.00	0.00	0.01	0.10	0.07	C
CYP81	18	7953688	7954908	-	ps	0.04	0.03	0.12	0.06	0.02	R
CYP81	18	8051010	8052099	+	ps	0.05	0.07	0.11	0.17	0.07	R
CYP81	18	30379654	30379854	-	ps	0.02	0.32	0.25	0.13	0.04	R
CYP81	18	33775420	33775623	-	ps	0.04	0.05	0.04	0.00	0.07	E
CYP81	19	1686328	1686519	-	ps	0.00	0.00	0.00	0.00	0.21	N
CYP81	3	11027883	11028035	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	4	2337448	2337678	+	ps	0.00	0.00	0.00	0.00	0.20	N
CYP81	6	8098711	8099391	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	7	22284641	22286688	+	ps	0.00	0.01	0.01	0.00	0.00	K
CYP81	7	22288447	22288605	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	7	22288605	22288766	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	7	22315074	22318353	+	ps	0.00	0.00	0.00	0.00	0.00	J
CYP81	7	22319145	22320449	+	ps	0.02	0.01	0.02	0.00	0.00	F
CYP81	7	22319885	22320076	+	ps	0.00	0.05	0.00	0.00	0.00	H
CYP81	7	22331331	22333217	+	ps	0.56	1.04	0.21	0.10	0.17	H
CYP81	7	22333219	22334430	+	ps	0.00	0.01	0.00	0.00	0.00	H
CYP81	7	22342161	22346238	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	7	22348142	22349740	+	ps	0.01	0.00	0.00	0.02	0.01	D
CYP81	9	6367010	6367372	-	ps	0.00	0.11	0.15	0.00	0.00	K
CYP81	9	6381065	6381507	-	ps	0.00	0.23	0.13	0.00	0.00	K
CYP81	9	12160854	12161012	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP81	Un	20111356	20111517	-	ps	0.03	0.08	0.12	0.68	5.88	N
CYP81	Un	20111602	20111760	-	ps	0.12	0.02	0.04	3.26	9.48	O
CYP82	1	18256205	18271611	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	12	21517775	21517978	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	12	21517978	21518142	-	ps	0.00	0.00	0.00	0.00	0.00	outlier

CYP82	12	21518331	21518549	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	12	21523571	21523930	-	ps	0.01	0.00	0.00	0.00	0.00	Z
CYP82	12	21572975	21574919	-	ps	0.01	0.00	0.00	0.00	0.04	AE
CYP82	18	9714071	9724784	+	ps	0.00	0.00	0.00	0.00	0.00	J
CYP82	18	9722237	9722536	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	18	9740228	9752656	+	ps	0.00	0.00	0.00	0.00	0.00	J
CYP82	18	9750107	9750406	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	18	9785814	9786200	+	ps	0.00	0.00	0.00	0.00	0.07	N
CYP82	18	9787198	9787389	+	ps	0.00	0.00	0.00	0.00	0.06	N
CYP82	18	9798051	9799754	+	ps	0.00	0.00	0.01	0.00	0.00	J
CYP82	18	9802541	9803172	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	18	9816450	9818702	+	ps	0.00	0.00	0.00	0.00	0.00	Z
CYP82	18	24024667	24024906	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	18	30963952	30964101	-	ps	0.00	0.02	0.03	0.00	0.00	K
CYP82	3	4387722	4389809	-	ps	0.29	0.25	0.08	0.02	0.05	R
CYP82	3	4390430	4390660	-	ps	0.00	0.02	0.03	0.00	0.00	K
CYP82	3	4391087	4393737	-	ps	0.00	0.01	0.02	0.00	0.00	T
CYP82	3	4407592	4407870	-	ps	0.02	0.00	0.00	1.17	0.77	W
CYP82	3	4431524	4431883	-	ps	0.00	0.02	0.04	0.00	0.00	K
CYP82	3	4432393	4434102	-	ps	0.00	0.00	0.00	0.00	0.00	Z
CYP82	3	4435647	4437210	-	ps	0.00	0.07	0.02	0.00	0.00	F
CYP82	3	4441106	4445712	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	3	4451674	4453568	-	ps	0.45	1.04	0.15	0.33	0.27	H
CYP82	3	4471514	4478285	-	ps	0.00	0.00	0.00	0.00	0.00	N
CYP82	3	4484494	4484661	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	3	4485190	4486900	-	ps	0.00	0.00	0.00	0.00	0.00	T
CYP82	3	4498176	4500076	-	ps	0.00	0.09	0.02	0.03	0.00	M
CYP82	3	4502155	4502392	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	3	4507057	4507766	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	7	15036736	15037353	+	ps	0.12	0.23	0.02	0.00	0.13	E
CYP82	9	10900808	10901203	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	9	10976494	10976945	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	9	11018408	11019025	-	ps	0.00	0.00	0.01	0.00	0.00	J
CYP82	9	11085979	11086383	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP82	9	11098021	11101612	-	ps	0.03	0.06	0.00	0.00	0.00	AB
CYP82	Un	22319624	22319995	+	ps	0.09	0.00	0.01	0.00	0.02	A
CYP84	4	19437461	19437679	-	ps	0.00	0.00	0.02	0.00	0.02	S
CYP86	3	12943438	12943662	-	ps	0.03	0.06	0.14	0.09	0.04	R
CYP87	10	20947075	20947224	+	ps	0.00	0.00	0.00	0.00	0.02	N
CYP87	12	7235277	7235432	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP87	15	16070237	16071923	+	ps	0.03	0.00	0.00	0.00	0.13	AE
CYP87	15	16091360	16091926	+	ps	0.12	0.03	0.00	0.00	0.01	U
CYP87	18	10150 47	10150 41	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP87	19	15010817	15010966	-	ps	0.00	0.00	0.03	0.00	0.00	J
CYP87	3	4466475	4466660	-	ps	0.08	0.15	0.10	0.00	0.06	E
CYP87	3	14418227	14419952	+	ps	0.01	0.01	0.00	0.00	0.05	U
CYP87	3	14419142	14419393	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP87	7	13346228	13346377	-	ps	0.00	0.11	0.05	0.39	0.26	P
CYP89	10	14384050	14385260	+	ps	0.00	0.00	0.00	0.00	0.00	J
CYP89	10	14410296	14413552	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP89	14	9641971	9643690	-	ps	0.00	0.00	0.00	0.00	0.01	U
CYP89	14	9683532	9685074	-	ps	0.00	0.01	0.00	0.00	0.02	U
CYP89	16	430419	430754	-	ps	0.13	0.00	0.00	0.39	0.13	AD
CYP89	16	482163	482720	+	ps	0.01	0.00	0.00	0.07	0.00	D
CYP89	16	487714	487875	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP89	16	488950	489258	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP89	16	595695	596606	+	ps	0.00	0.01	0.02	0.00	0.18	E
CYP89	17	8166177	8166368	+	ps	0.00	0.00	0.01	0.00	0.00	J

CYP89	19	7698087	7699668	+	ps	0.35	0.07	0.17	5.60	3.19	W
CYP89	Un	1269962	1270159	-	ps	0.86	2.20	1.93	1.81	1.57	I
CYP89	Un	6571254	6571532	-	ps	0.00	0.00	0.00	0.00	0.02	N
CYP90	8	21136300	21136464	-	ps	1.19	2.63	24.84	6.25	17.96	S
CYP92	1	20437600	20440651	-	ps	0.38	0.19	0.30	0.33	0.22	R
CYP92	11	849767	850875	+	ps	0.62	0.44	0.15	0.15	0.16	R
CYP92	11	861031	862541	+	ps	1.30	1.18	0.24	0.11	0.02	AB
CYP93	12	9294958	9295107	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP93	19	21678232	21678731	+	ps	0.63	0.57	0.88	1.72	2.93	O
CYP93	19	22251878	22278125	+	ps	0.14	0.26	0.28	0.63	1.10	N
CYP94	11	15928617	15928790	-	ps	32.90	47.22	31.77	61.04	68.25	Q
CYP96	14	18141203	18142713	+	ps	4.70	0.02	0.00	0.00	0.02	Z
CYP96	14	26932881	26942356	+	ps	0.94	0.97	0.52	1.18	1.66	O
CYP96	14	26948964	26949173	+	ps	0.00	0.03	0.17	0.10	0.07	P
CYP96	7	24100997	24102541	-	ps	0.22	0.01	0.00	0.06	0.01	R
CYP98	2	12422555	12422704	+	ps	0.14	0.12	0.03	0.00	0.08	E
CYP706	16	14997999	14998590	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP706	16	15060342	15060974	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP706	16	15085511	15086083	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP706	16	15152141	15152386	-	ps	0.14	0.03	0.12	0.00	0.00	F
CYP706	18	11784192	11784356	-	ps	0.00	0.00	0.00	0.00	0.04	N
CYP706	2	4646292	4646459	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP706	5	25109265	25112558	+	ps	0.28	0.33	0.20	0.07	0.10	R
CYP706	5	25128847	25130640	+	ps	0.04	0.01	0.00	0.00	0.01	E
CYP706	5	25138783	25139019	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP706	5	25141028	25141509	+	ps	0.02	0.00	0.01	0.00	0.04	A
CYP707	12	516043	517552	+	ps	0.01	0.01	0.02	0.00	0.00	E
CYP711	14	13645354	13645542	-	ps	0.00	0.00	0.00	0.20	0.00	X
CYP711	4	914904	915965	-	ps	0.09	0.06	0.01	0.10	0.23	R
CYP712	12	6873001	6885757	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP712	12	6884613	6884867	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP712	13	27865564	27868383	-	ps	0.05	0.05	0.15	0.04	0.04	R
CYP712	15	14191180	14191344	+	ps	0.00	0.12	0.00	0.00	0.02	U
CYP714	10	19326909	19327353	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP714	10	19448799	19449173	+	ps	0.09	0.00	0.02	0.00	0.00	A
CYP714	15	11118508	11122879	+	ps	3.73	1.15	0.34	0.00	2.01	Z
CYP714	15	11128769	11129348	+	ps	0.01	0.02	0.00	0.00	0.00	AB
CYP714	15	11135862	11137145	+	ps	0.01	0.00	0.01	0.00	0.00	F
CYP714	18	33751092	33751664	-	ps	0.07	0.00	0.00	0.00	0.08	AE
CYP714	19	7218428	7219436	+	ps	0.10	0.02	0.04	0.00	0.09	E
CYP714	3	12684548	12684718	-	ps	0.00	0.00	0.00	0.00	0.02	N
CYP714	6	9241542	9249703	-	ps	0.00	0.01	0.00	0.00	0.03	U
CYP716	15	3877587	3884324	+	ps	0.00	0.01	0.00	0.00	0.00	H
CYP716	16	22337676	22337825	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP716	18	29715995	29721439	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP716	18	29756697	29758528	+	ps	0.00	2.32	0.28	0.00	0.00	I
CYP716	18	29782679	29783278	-	ps	0.56	0.01	0.01	0.00	0.00	F
CYP716	18	30184390	30184686	+	ps	0.00	0.00	0.02	0.00	0.00	J
CYP716	19	19765514	19767375	+	ps	0.69	0.13	0.16	0.04	0.00	R
CYP716	6	391297	393052	-	ps	0.05	0.12	0.08	0.05	0.20	R
CYP716	7	9648701	9649402	+	ps	0.10	0.00	0.00	0.00	0.00	Z
CYP716	8	15649279	15649428	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP721	18	10356448	10358765	-	ps	0.92	9.97	0.84	2.15	14.13	V
CYP721	18	10368994	10369234	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP721	18	10376295	10381065	-	ps	0.01	0.03	0.00	0.00	0.00	AB
CYP721	18	10385705	10391303	-	ps	0.11	0.03	0.43	0.04	0.12	R
CYP724	12	13211586	13211735	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP724	17	6175797	6175955	+	ps	164.85	287.83	169.79	208.05	162.68	H

CYP724	3	15992473	15992622	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP728	15	10216724	10216873	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP728	18	21551452	21551601	+	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP728	19	6322164	6322676	-	ps	0.37	0.16	0.19	0.96	1.70	N
CYP728	19	6323028	6323270	-	ps	0.38	0.08	0.19	0.44	1.54	N
CYP728	19	6323471	6323647	-	ps	0.61	0.16	0.42	0.60	1.51	N
CYP728	7	14674801	14674950	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP734	12	19245533	19245688	+	ps	0.00	0.04	0.00	0.00	0.00	H
CYP734	17	8856272	8856421	-	ps	1.60	0.25	0.48	2.48	1.20	W
CYP736	14	24626807	24626977	-	ps	0.02	0.00	0.03	0.00	0.10	A
CYP736	17	3941935	3942509	-	ps	0.01	0.17	0.15	0.00	0.00	F
CYP736	17	3949030	3951005	-	ps	3.39	5.77	0.83	2.31	0.98	H
CYP736	17	3994682	4010408	-	ps	3.23	1.34	0.92	1.40	0.46	AA
CYP736	18	31019885	31020046	-	ps	0.00	0.00	0.00	0.00	0.00	outlier
CYP736	7	23782439	23783987	+	ps	1.51	0.22	0.19	4.13	2.61	W
CYP736	Un	6176251	6176457	+	ps	0.00	0.04	0.00	0.00	0.01	U

Appendix II

DNA and protein sequence alignments of gene candidates for (*E*)-8-carboxylinalool biosynthesis

Gene candidates for (*E*)-8-carboxylinalool biosynthesis (*CYP71AT7, CYP76F12, CYP76F14, CYP76T21, CYP76Y1* and *CYP76Y2*) were amplified from Gewurztraminer (Gw) and Muscat Ottonel (Mo) grape skin cDNA. Their sequences were aligned to the sequence from the reference grapevine genome PN40024 (Pn) using ClustalW. Differences between the varieties are highlighted with light gray.

```
PnCYP71AT7  ATGATGATTTTGCTTCTTATCCTTTTAGCTCTCCCTCTCTTTCTCTTGTTTCTTCTTCGGAATCGAAGAA 70
MoCYP71AT7  ATGATGATTTTGCTTCTTATCCTTTTAGCTGTCCCTCTCTTTCTCTTGTTTCTTCTTCGGAATCGAAGAA 70
GwCYP71AT7  ATGATGATTTTGCTTCTTATCCTTTTAGCTCTCCCTCTCTTTCTCTTGTTTCTTCTTCGGAATCGAAGAA 70

PnCYP71AT7  GAACGCCCCTCCCACCAGGTCCTCCAGGGCTTCCCCTGATTGGAAACTTGCTCCAGCTTGATAAATCAGC 140
MoCYP71AT7  GAACGCCCCTCCCACCAGGTCCTCCAGGGCTTCCCCTGATTGGAAACTTGCTCCAGCTTGATAAATCAGC 140
GwCYP71AT7  GAACGCCCCTCCCACCAGGTCCTCCAGGGCTTCCCCTGATTGGAAACTTGCTCCAGCTTGATAAATCAGC 140

PnCYP71AT7  TCCTCATATCTACCTATGGCGACTTTCTAAGCAATATGGCCCCCTCATGATCTTGCGCCTTGGTTTTGTG 210
MoCYP71AT7  TCCTCATATCTACCTATGGCGACTTTCTAAGCAATACGGCCCCCTCATGATCTTGCGCCTTGGTTTTGTG 210
GwCYP71AT7  TCCTCATATCTACCTATGGCGACTTTCTAAGCAATATGGCCCCCTCATGATCTTGCGCCTTGGTTTTGTG 210

PnCYP71AT7  CCAACCCTAGTGGTTTCTTCAGCAAGAATGGCAAAAGAGGTCATGAAAACACATGATCTTGAATTTTCTG 280
MoCYP71AT7  CCAACCCTAGTGGTTTCTTCAGCAAGAATGGCAAAAGAGGTCATGAAAACACATGATCTTGAATTTTCTG 280
GwCYP71AT7  CCAACCCTAGTGGTTTCTTCAGCAAGAATGGCAAAAGAGGTCATGAAAACACATGATCTTGAATTTTCTG 280

PnCYP71AT7  GTAGGCCTTCCTTGCTTGGTCTGCGGAAGCTATCCTACAATGGCCTGGATGTGGCCTTTTCACCATATAA 350
MoCYP71AT7  GTAGGCCTTCCTTGCTTGGTCTGCAGAAGCTATCCTACAATGGCCTGGATGTGGCCTTTTCACCATATAA 350
GwCYP71AT7  GTAGGCCTTCCTTGCTTGGTCTGCGGAAGCTATCCTACAATGGCCTGGATGTGGCCTTTTCACCATATAA 350

PnCYP71AT7  TGATTATTGGAGAGAAATGAGGAAGATTTGTGTTCTCCATCTTTTCAACTCTAAAAGGGCGCAATCCTTT 420
MoCYP71AT7  TGATTATTGGAGAGAAATGAGGAAGATTTGTGTTCTCCATCTTTTCAACTCTAAAAGGGCGCAATCCTTT 420
GwCYP71AT7  TGATTATTGGAGAGAAATGAGGAAGATTTGTGTTCTCCATCTTTTCAACTCTAAAAGGGCGCAATCCTTT 420

PnCYP71AT7  CGTCCCATTCGGGAAGATGAGGTTTTGGAAATGATTAAAAAAATTTCCCAGTTTGCTTCTGCTTCCAAGC 490
MoCYP71AT7  CGTCCCATTCGGGAAGATGAGGTTTTGGAAATGATTAAAAAAATTTCCCAGTTTGCTTCTGCTTCCAAGC 490
GwCYP71AT7  CGTCCCATTCGGGAAGATGAGGTTTTGGAAATGATTAAAAAAATTTCCCAGTTTGCTTCTGCTTCCAAGC 490

PnCYP71AT7  TCACTAACTTAAGTGAGATCTTAATTTCTCTCACAAGCACCATAATCTGCAGGGTTGCTTTCAGTAAGAG 560
MoCYP71AT7  TCACTAACTTAAGTGAGATCTTAATTTCTCTCACAAGCACCATAATCTGCAGGGTTGCTTTCAGTAAGAG 560
GwCYP71AT7  TCACTAACTTAAGTGAGATCTTAATTTCTCTCACAAGCACCATAATCTGCAGGGTTGCTTTCAGTAAGAG 560

PnCYP71AT7  GTACGATGATGAAGGATATGAAAGAAGCAGATTTCAGAAACTCGTAGGTGAAGGTCAGGCTGTGGTGGGA 630
MoCYP71AT7  GTACGATGATGAAGGATATGAAAGAAACAGATTTCAGAAACTCGTAGGTGAAGGTCAGGCTGTGGTGGGA 630
GwCYP71AT7  GTACGATGATGAAGGATATGAAAGAAGCAGATTTCAGAAACTCGTAGGTGAAGGTCAGGCTGTGGTGGGA 630

PnCYP71AT7  GGCTTCTATTTCTCGGATTATTTTCCTTTGATGGGCTGGGTTGATAAACTCACAGGGATGATTGCTCTGG 700
MoCYP71AT7  GGCTTCTATTTCTCGGATCATTTTCCTTTGATGGGCTGGGTTGATAAACTCACAGGGATGATTGCTCTGG 700
GwCYP71AT7  GGCTTCTATTTCTCGGATTATTTTCCTTTGATGGGCTGGGTTGATAAACTCACAGGGATGATTGCTCTGG 700

PnCYP71AT7  CCGATAAAAATTTCAAGGAATTTGATTTGTTTTATCAAGAAATCATTGATGAGCACCTGGATCCAAATAG 770
MoCYP71AT7  CCGATAAAAATTTCAAGGAATTTGATTTGTTTTATCAAGAAATCATTGATGAGCACCTGGATCCAAATAG 770
GwCYP71AT7  CCGATAAAAATTTCAAGGAATTTGATTTGTTTTATCAAGAAATCATTGATGAGCACCTGGATCCAAATAG 770

PnCYP71AT7  ACCAGAGCCTGAGAAAGAAGACATCACTGACGTCTTACTTAAACTGCAGAAGAATCGTCTGTTTACAATT 840
MoCYP71AT7  ACCAGAGCCTGAGAAAGAAGACATCACTGACGTCTTACTTAAACTGCAGAAGAATCGTCTGTTTACAATT 840
GwCYP71AT7  ACCAGAGCCTGAGAAAGAAGACATCACTGACGTCTTACTTAAACTGCAGAAGAATCGTCTGTTTACAATT 840

PnCYP71AT7  GATCTCACTTTTGATCACATTAAAGCAGTTCTCATGAACATATTTCTCGCAGGGACAGATACAAGCGCAG 910
MoCYP71AT7  GATCTCACTTTTGATCACATTAAAGCAGTTCTCATGAACATATTTCTCGCAGGGACAGATACAAGCGCAG 910
GwCYP71AT7  GATCTCACTTTTGATCACATTAAAGCAGTTCTCATGAACATATTTCTCGCAGGGACAGATACAAGCGCAG 910

PnCYP71AT7  CTACCTTGGTTTGGGCAATGACAATGCTCATGAAGAATCCTAGAACGATGACGAAAGCACAAGAGGAACT 980
MoCYP71AT7  CTACCTTGGTTTGGGCAATGACAATGCTCATGAAGAATCCTAGAACGATGACGAAAGCACAAGAGGAACT 980
GwCYP71AT7  CTACCTTGGTTTGGGCAATGACAATGCTCATGAAGAATCCTAGAACGATGACGAAAGCACAAGAGGAACT 980

PnCYP71AT7  CAGAAATTTAATCGGAAAGAAAGGTTTTGTAGATGAAGATGATCTTCAGAAGCTTCCGTATCTGAAAGCA 1050
MoCYP71AT7  CAGAAATTTAATCGGAAAGAAAGGTTTTGTAGATGAAGATGATCTTCAGAAGCTTCCGTATCTGAAAGCA 1050
GwCYP71AT7  CAgAAATTTAATCGGAAAGAAAGGTTTTGTAGATGAAGATGATCTTCAGAAGCTTCCGTATCTGAAAGCA 1050

PnCYP71AT7  ATTGTGAAGGAGACGATGAGATTGCACCCTGCAAGCCCATTGCTAGTCCCGAGAGAAACACTTGAGAAGT 1120
MoCYP71AT7  ATTGTGAAGGAGACGATGAGATTGCACCCTGCAAGCCCATTGCTAGTCCCGAGAGAAACACTTGAGAAGT 1120
GwCYP71AT7  ATTGTGAAGGAGACGATGAGATTGCACCCTGCAAGCCCATTGCTAGTCCCGAGAGAAACACTTGAGAAGT 1120

PnCYP71AT7  GTGTGATAGATGGGTATGAGATACCACCCAAAACCCTAGTGTATGTGAATGCATGGGCAATCGGAAGAGA 1190
MoCYP71AT7  GTGTGATAGATGGGTATGAGATACCACCCAAAACCCTAGTGTATGTGAATGCATGGGCAATCGGAAGAGA 1190
GwCYP71AT7  GTGTGATAGATGGGTATGAGATACCACCCAAAACCCTAGTGTATGTGAATGCATGGGCAATCGGAAGAGA 1190

PnCYP71AT7  TCCTGAGTCCTGGGAAAACCCAGAAGAGTTCATGCCTGAGAGATTCTTGGGAACTTCCATAGACTTCAAA 1260
MoCYP71AT7  TCCTGAGTCCTGGGAAAACCCAGAAGAGTTCATGCCTGAGAGATTCTTGGGAACTTCCATAGACTTCAAA 1260
GwCYP71AT7  TCCTGAGTCCTGGGAAAACCCAGAAGAGTTCATGCCTGAGAGATTCTTGGGAACTTCCATAGACTTCAAA 1260

PnCYP71AT7  GGACAGGATTACCAACTCATACCCTTCGGAGGAGGACGAAGAATTTGCCCAGGCTTAAATCTAGGAGCTG 1330
MoCYP71AT7  GGACAGGATTACCAACTCATACCCTTCGGAGGAGGACGAAGAATTTGICCAGGCTTAAATCTAGGAGCTG 1330
GwCYP71AT7  GGACAGGATTACCAACTCATACCCTTCGGAGGAGGACGAAGAATTTGCCCAGGCTTAAATCTAGGAGCTG 1330
```

```
PnCYP71AT7  CGATGGTGGAGCTGACACTAGCTAATCTTCTTTACTCGTTTGACTGGGAAATGCCCGCTGGAATGAACAA 1400
MoCYP71AT7  CGGTGGTGGAGCTGACACTAGCTAATCTTCTTTACTCGTTTGACTGGGAAATGCCCGCTGGAATGAACAA 1400
GwCYP71AT7  CGATGGTGGAGCTGACACTAGCTAATCTTCTTTACTCGTTTGACTGGGAAATGCCCGCTGGAATGAACAA 1400

PnCYP71AT7  GGAAGACATAGACATCGACGTCAAACCTGGTATTACAATGCACAAGAAGAATGCTCTTTGCCTTTTGGCG 1470
MoCYP71AT7  GGAAGACATAGACATCGACGTCAAACCTGGTATTACAATGCACAAGAAGAATGCTCTTTGCCTTTTGGCG 1470
GwCYP71AT7  GGAAGACATAGACATCGACGTCAAACCTGGTATTACAATGCACAAGAAGAATGCTCTTTGCCTTTTGGCG 1470

PnCYP71AT7  AGGATCCCTAGTCATTAA 1488
MoCYP71AT7  AGGATCCCTAGTCATTAA 1488
GwCYP71AT7  AGGATCCCTAGTCATTAA 1488
```

```
PnCYP71AT7  MMILLLILLALPLFLLFLLRNRRRTPLPPGPPGLPLIGNLLQLDKSAPHIYLWRLSKQYGPLMILRLGFV 70
MoCYP71AT7  MMILLLILLAVPLFLLFLLRNRRRTPLPPGPPGLPLIGNLLQLDKSAPHIYLWRLSKQYGPLMILRLGFV 70
GwCYP71AT7  MMILLLILLALPLFLLFLLRNRRRTPLPPGPPGLPLIGNLLQLDKSAPHIYLWRLSKQYGPLMILRLGFV 70

PnCYP71AT7  PTLVVSSARMAKEVMKTHDLEFSGRPSLLGLRKLSYNGLDVAFSPYNDYWREMRKICVLHLFNSKRAQSF 140
MoCYP71AT7  PTLVVSSARMAKEVMKTHDLEFSGRPSLLGLQKLSYNGLDVAFSPYNDYWREMRKICVLHLFNSKRAQSF 140
GwCYP71AT7  PTLVVSSARMAKEVMKTHDLEFSGRPSLLGLRKLSYNGLDVAFSPYNDYWREMRKICVLHLFNSKRAQSF 140

PnCYP71AT7  RPIREDEVLEMIKKISQFASASKLTNLSEILISLTSTIICRVAFSKRYDDEGYERSRFQKLVGEGQAVVG 210
MoCYP71AT7  RPIREDEVLEMIKKISQFASASKLTNLSEILISLTSTIICRVAFSKRYDDEGYERNRFQKLVGEGQAVVG 210
GwCYP71AT7  RPIREDEVLEMIKKISQFASASKLTNLSEILISLTSTIICRVAFSKRYDDEGYERSRFQKLVGEGQAVVG 210

PnCYP71AT7  GFYFSDYFPLMGWVDKLTGMIALADKNFKEFDLFYQEIIDEHLDPNRPEPEKEDITDVLLKLQKNRLFTI 280
MoCYP71AT7  GFYFSDHFPLMGWVDKLTGMIALADKNFKEFDLFYQEIIDEHLDPNRPEPEKEDITDVLLKLQKNRLFTI 280
GwCYP71AT7  GFYFSDYFPLMGWVDKLTGMIALADKNFKEFDLFYQEIIDEHLDPNRPEPEKEDITDVLLKLQKNRLFTI 280

PnCYP71AT7  DLTFDHIKAVLMNIFLAGTDTSAATLVWAMTMLMKNPRTMTKAQEELRNLIGKKGFVDEDDLQKLPYLKA 350
MoCYP71AT7  DLTFDHIKAVLMNIFLAGTDTSAATLVWAMTMLMKNPRTMTKAQEELRNLIGKKGFVDEDDLQKLPYLKA 350
GwCYP71AT7  DLTFDHIKAVLMNIFLAGTDTSAATLVWAMTMLMKNPRTMTKAQEELRNLIGKKGFVDEDDLQKLPYLKA 350

PnCYP71AT7  IVKETMRLHPASPLLVPRETLEKCVIDGYEIPPKTLVYVNAWAIGRDPESWENPEEFMPERFLGTSIDFK 420
MoCYP71AT7  IVKETMRLHPASPLLVPRETLEKCVIDGYEIPPKTLVYVNAWAIGRDPESWENPEEFMPERFLGTSIDFK 420
GwCYP71AT7  IVKETMRLHPASPLLVPRETLEKCVIDGYEIPPKTLVYVNAWAIGRDPESWENPEEFMPERFLGTSIDFK 420

PnCYP71AT7  GQDYQLIPFGGGRRICPGLNLGAAMVELTLANLLYSFDWEMPAGMNKEDIDIDVKPGITMHKKNALCLLA 490
MoCYP71AT7  GQDYQLIPFGGGRRICPGLNLGAAVVELTLANLLYSFDWEMPAGMNKEDIDIDVKPGITMHKKNALCLLA 490
GwCYP71AT7  GQDYQLIPFGGGRRICPGLNLGAAMVELTLANLLYSFDWEMPAGMNKEDIDIDVKPGITMHKKNALCLLA 490

PnCYP71AT7  RIPSH* 496
MoCYP71AT7  RIPSH* 496
GwCYP71AT7  RIPSH* 496
```

```
PnCYP76F12  ATGGAGATGTTGAGCTGTCTGCTGTGTTTTCTCGTCGCTTGGACTTCAATTTACATCATGTTTTCAGTTA 70
MoCYP76F12  ATGGAGATGTTGAGCTGTCTGCTGTGTTTTCTCGTCGCTTGGACTTCAATTTACATCATGTTTTCAGTTA 70
GwCYP76F12  ATGGAGATGTTGAGCTGTCTGCTGTGTTTTCTCGTCGCTTGGACTTCAATTTACATCATGTTTTCAGTTA 70

PnCYP76F12  GAAGAGGAAGCCAACATACTGCTTACAAACTTCCTCCGGGACCAGTTCCACTTCCCATAATAGGAAACCT 140
MoCYP76F12  GAAGAGGAAGCCAACATACTGCTTACAAACTTCCTCCGGGACCAGTTCCACTTCCCATAATAGGAAACCT 140
GwCYP76F12  GAAGAGGAAGCCAACATACTGCTTACAAACTTCCTCCGGGACCAGTTCCACTTCCCATAATAGGAAACCT 140

PnCYP76F12  CTTAAACCTGGGTAACAGACCCCATGAGTCCCTAGCCGAACTTGCAAAAACTTACGGCCCAATAATGACT 210
MoCYP76F12  CTTAAACCTGGGTAACAGACCCCATGAGTCCCTAGCCGAACTTGCAAAAACTTACGGCCCAATAATGACT 210
GwCYP76F12  CTTAAACCTGGGTAACAGACCCCATGAGTCCCTAGCCGAACTTGCAAAAACTTACGGCCCAATAATGACT 210

PnCYP76F12  CTTAAACTTGGCTATGTAACCACAATAGTCATCTCTTCTGCACCCATGGCCAAAGAGGTCCTCCAAAAGC 280
MoCYP76F12  CTTAAACTTGGCTATGTAACCACAATAGTCATCTCTTCTGCACCCATGGCCAAAGAGGTCCTCCAAAAGC 280
GwCYP76F12  CTTAAACTTGGCTATGTAACCACAATAGTCATCTCTTCTGCACCCATGGCCAAAGAGGTCCTCCAAAAGC 280

PnCYP76F12  AGGACCTCTCCTTCTGCAACCGCTTCGTCCCTGATGCCATCCGAGCCACAAACCACAACCAACTGTCAAT 350
MoCYP76F12  AGGACCTCTCCTTCTGCAACCGCTTCGTCCCTGATGCCATCCGAGCCACAAACCACAACCAACTGTCAAT 350
GwCYP76F12  AGGACCTCTCCTTCTGCAACCGCTTCGTCCCTGATGCCATCCGAGCCACAAACCACAACCAACTGTCAAT 350

PnCYP76F12  GGCCTGGATGCCTGTTTCAACAACTTGGAGAGTCCTTCGAAAGATATGCAATTCGCACTTATTCACAACA 420
MoCYP76F12  GGCCTGGATGCCTGTTTCAACAACTTGGAGAGTCCTTCGAAAGATATGCAATTCGCACTTATTCACAACA 420
GwCYP76F12  GGCCTGGATGCCTGTTTCAACAACTTGGAGAGTCCTTCGAAAGATATGCAATTCGCACTTATTCACAATA 420

PnCYP76F12  CAAAAACTAGACTCTAACACCCATCTCCGCCACCACACAAAGTGCAAGAACTTCTTGCGAAAGTTGAAGAGA 490
MoCYP76F12  CAAAAACTAGACTCTAACACCCATCTCCGCCACCACACAAAGTGCAAGAACTTCTTGCGAAAGTTGAAGAGA 490
GwCYP76F12  CAAAAACTAGACTCTAACACCCATCTCCGCCACCACACAAAGTGCAAGAACTTCTTGCGAAAGTTGAAGAGA 490

PnCYP76F12  GCCGCCAAGCCGGTGATGCTGTGTATATAGGCCGTGAAGCTTTTAGAACTAGTCTCAACTTGTTATCTAA 560
MoCYP76F12  GCCGCCAAGCCGGTGATGCTGTGTATATAGGCCGTGAAGCTTTTAGAACTAGTCTCAACTTGTTATCTAA 560
GwCYP76F12  GCCGCCAAGCCGGTGATGCTGTGTATATAGGCCGTGAAGCTTTTAGAACTAGTCTCAACTTGTTATCTAA 560

PnCYP76F12  CACCATATTTTCTGTGGATCTTGTGGATCCAATTTCTGAGACTGTACTAGAGTTTCAGGAGTTGGTGCGT 630
MoCYP76F12  CACCATATTTTCTGTGGATCTTGTGGATCCAATTTCTGAGACTGTACTAGAGTTTCAGGAGTTGGTGCGT 630
GwCYP76F12  CACCATATTTTCTGTGGATCTTGTGGATCCAATTTCTGAGACTGTACTAGAGTTTCAGGAGTTGGTGCGT 630

PnCYP76F12  TGTATAATAGAGGAAATTGAGAGACCCAACTTGGTAGATTATTTTCCGGTGCTCAGAAAGATTGATCCAC 700
MoCYP76F12  TGTATAATAGAGGAAATTGAGAGACCCAACTTGGTAGATTATTTTCCGGTGCTCAGAAAGATTGATCCAC 700
GwCYP76F12  TGTATAATAGAGGAAATTGAGAGACCCAACTTGGTAGATTATTTTCCGGTGCTCAGAAAGATTGATCCAC 700

PnCYP76F12  AAGGTATAAGGCGTCGTTTGACAATTTATTTTGGTAAGATGATTGGGATATTTGATAGAATGATCAAGCA 770
MoCYP76F12  AAGGTATAAGGCGTCGTTTGACAATTTATTTTGGTAAGATGATTGGGATATTTGATAGAATGATCAAGCA 770
GwCYP76F12  AAGGTATAAGGCGTCGTTTGACAATTTATTTTGGTAAGATGATTGGGATATTTGATAGAATGATCAAGCA 770

PnCYP76F12  ACGGTTACAGTTAAGAAAAATGCAAGGTTCAATAGCTACCAGTGACGTGTTAGATACTCTTCTCAACATC 840
MoCYP76F12  ACGGTTGCAGTTAAGAAAAATGCAAGGTTCAATAGCTACCAGTGACGTGTTAGATACTTTTCTCAACATC 840
GwCYP76F12  ACGGTTACAGTTAAGAAAAATGCAAGGTTCAATAGCTACCAGTGACGTGTTAGATACTCTTCTCAACATC 840

PnCYP76F12  AGCGAGGATAACAGCAATGAGATTGAAAGAAATCATATGGAACATTTGTTATTGGACTTATTTGTTGCGG 910
MoCYP76F12  AGCGAGGATAACAGCAATGAGATTGAAAGAAATCATATGGAACATTTGTTATTGGACTTATTTGTTGCGG 910
GwCYP76F12  AGCGAGGATAACAGCAATGAGATTGAAAGAAATCATATGGAACATTTGTTATTGGACTTATTTGTTGCGG 910

PnCYP76F12  GGACTGACACAACTTCGAGCACCTTGGAATGGGCAATGGCAGAGCTGCTACACAACCCTGAAAAACTTTT 980
MoCYP76F12  GGACTGACACAACTTCGAGCACCTTGGAATGGGCAATGGCAGAGCTGCTACACAACCCTGAAAAACTTTT 980
GwCYP76F12  GGACTGACACAACTTCGAGCACCTTGGAATGGGCAATGGCAGAGCTGCTACACAACCCTGAAAAACTTTT 980

PnCYP76F12  AAAAGCCCGAGTGGAGCTCCTGCAAACCATCGGCAAAGACAAACAGGTAAAAGAATCAGACATCACTCGA 1050
MoCYP76F12  AAAAGCCCAAGTGGAGCTCCTGCAAACCATCGGCAAAGACAAACAGGTAAAAGAATCAGACATCACTCGA 1050
GwCYP76F12  AAAAGCCCGAGTGGAGCTCCTGCAAACCATCGGCAAAGACAAACAGGTAAAAGAATCAGACATCACTCGA 1050

PnCYP76F12  CTCCCTTTCTTGCAAGCGGTCGTGAAAGAAACTTTCCGATTGCACCCAGTAGTTCCATTTTTAATCCCAC 1120
MoCYP76F12  CTCCCTTTCTTGCAAGCGGTCGTGAAAGAAACTTTCCGATTGCACCCAGTAGTTCCATTTTTAATCCCAC 1120
GwCYP76F12  CTCCCTTTCTTGCAAGCGGTCGTGAAAGAAACTTTCCGATTGCACCCAGTAGTTCCATTTTTAATCCCAC 1120

PnCYP76F12  ACAGAGTTGAAGAGGACACAGATATAGATGGGCTTACAGTCCCAAAGAACGCACAGGTGCTGGTGAACGC 1190
MoCYP76F12  ACAGAGTTGAAGAGGACACAGATATAGATGGGCTTACAGTCCCAAAGAACGCACAGGTGCTGGTGAACGC 1190
GwCYP76F12  ACAGAGTTGAAGAGGACACAGATATAGATGGGCTTACAGTCCCAAAGAACGCACAGGTGCTGGTGAACGC 1190

PnCYP76F12  ATGGGCTATAGGTCGAGACCCGAACATATGGGAGAACCCCAACTCCTTTGTACCCGAGAGATTCTTGGAG 1260
MoCYP76F12  ATGGGCTATAGGTCGAGACCCGAACATATGGGAGAACCCCAACTCCTTTGTACCCGAGAGATTCTTGGAG 1260
GwCYP76F12  ATGGGCTATAGGTCGAGACCCGAACATATGGGAGAACCCCAACTCCTTTGTACCCGAGAGATTCTTGGAG 1260

PnCYP76F12  TTGGACATGGACGTGAAGGGCCAGAATTTTGAGTTGATTCCGTTTGGTGCTGGCAGGAGAATCTGTCCTG 1330
MoCYP76F12  TTGGACATGGACGTGAAGGGCCAGAATTTTGAGTTGATTCCGTTTGGTGCTGGCAGGAGAATCTGTCCTG 1330
GwCYP76F12  TTGGACATGGACGTGAAGGGCCAGAATTTTGAGTTGATTCCGTTTGGTGCTGGCAGGAGAATCTGTCCTG 1330
```

```
PnCYP76F12  GGTTGCCATTGGCAACCCGGATGGTTCACTTGATGCTGGCTTCACTCATTCACTCCTGTGATTGGAAACT 1400
MoCYP76F12  GGTTGCCATTGGCAACCCGGATGGTTCACTTGATGCTGGCCTCACTCATTCACTCCTGTGATTGGAAACT 1400
GwCYP76F12  GGTTGCCATTGGCAACCCGGATGGTTCACTTGATGCTGGCTTCACTCATTCACTCCTGTGATTGGAAACT 1400

PnCYP76F12  TGAAGATGGGATGACACCGGAAAACATGAACATGGAAGACAGGTTTGGCATTACCTTACAGAAGGCTCAG 1470
MoCYP76F12  TGAAGATGGGATAACACCGGAAAACATGAACATGGAAGACAGGTTTGGCATTACCTTACAGAAGGCTCAG 1470
GwCYP76F12  TGAAGATGGGATGACACCGGAAAACATGAACATGGAAGACAGGTTTGGCATTACCTTACAGAAGGCTCAG 1470

PnCYP76F12  CCCCTGAAAGCTATACCGATACGTGTGTGA 1500
MoCYP76F12  CCCCTGAAAGCTATACCGATACGTGTGTGA 1500
GwCYP76F12  CCCCTGAAAGCTATACCGATACGTGTGTGA 1500
```

```
PnCYP76F12  MEMLSCLLCFLVAWTSIYIMFSVRRGSQHTAYKLPPGPVPLPIIGNLLNLGNRPHESLAELAKTYGPIMT 70
MoCYP76F12  MEMLSCLLCFLVAWTSIYIMFSVRRGSQHTAYKLPPGPVPLPIIGNLLNLGNRPHESLAELAKTYGPIMT 70
GwCYP76F12  MEMLSCLLCFLVAWTSIYIMFSVRRGSQHTAYKLPPGPVPLPIIGNLLNLGNRPHESLAELAKTYGPIMT 70

PnCYP76F12  LKLGYVTTIVISSAPMAKEVLQKQDLSFCNRFVPDAIRATNHNQLSMAWMPVSTTWRVLRKICNSHLFTT 140
MoCYP76F12  LKLGYVTTIVISSAPMAKEVLQKQDLSFCNRFVPDAIRATNHNQLSMAWMPVSTTWRVLRKICNSHLFTT 140
GwCYP76F12  LKLGYVTTIVISSAPMAKEVLQKQDLSFCNRFVPDAIRATNHNQLSMAWMPVSTTWRVLRKICNSHLFTI 140

PnCYP76F12  QKLDSNTHLRHHKVQELLAKVEESRQAGDAVYIGREAFRTSLNLLSNTIFSVDLVDPISETVLEFQELVR 210
MoCYP76F12  QKLDSNTHLRHHKVQELLAKVEESRQAGDAVYIGREAFRTSLNLLSNTIFSVDLVDPISETVLEFQELVR 210
GwCYP76F12  QKLDSNTHLRHHKVQELLAKVEESRQAGDAVYIGREAFRTSLNLLSNTIFSVDLVDPISETVLEFQELVR 210

PnCYP76F12  CIIEEIERPNLVDYFPVLRKIDPQGIRRRLTIYFGKMIGIFDRMIKQRLQLRKMQGSIATSDVLDTLLNI 280
MoCYP76F12  CIIEEIERPNLVDYFPVLRKIDPQGIRRRLTIYFGKMIGIFDRMIKQRLQLRKMQGSIATSDVLDTFLNI 280
GwCYP76F12  CIIEEIERPNLVDYFPVLRKIDPQGIRRRLTIYFGKMIGIFDRMIKQRLQLRKMQGSIATSDVLDTLLNI 280

PnCYP76F12  SEDNSNEIERNHMEHLLLDLFVAGTDTTSSTLEWAMAELLHNPEKLLKARVELLQTIGKDKQVKESDITR 350
MoCYP76F12  SEDNSNEIERNHMEHLLLDLFVAGTDTTSSTLEWAMAELLHNPEKLLKAQVELLQTIGKDKQVKESDITR 350
GwCYP76F12  SEDNSNEIERNHMEHLLLDLFVAGTDTTSSTLEWAMAELLHNPEKLLKARVELLQTIGKDKQVKESDITR 350

PnCYP76F12  LPFLQAVVKETFRLHPVVPFLIPHRVEEDTDIDGLTVPKNAQVLVNAWAIGRDPNIWENPNSFVPERFLE 420
MoCYP76F12  LPFLQAVVKETFRLHPVVPFLIPHRVEEDTDIDGLTVPKNAQVLVNAWAIGRDPNIWENPNSFVPERFLE 420
GwCYP76F12  LPFLQAVVKETFRLHPVVPFLIPHRVEEDTDIDGLTVPKNAQVLVNAWAIGRDPNIWENPNSFVPERFLE 420

PnCYP76F12  LDMDVKGQNFELIPFGAGRRICPGLPLATRMVHLMLASLIHSCDWKLEDGMTPENMNMEDRFGITLQKAQ 490
MoCYP76F12  LDMDVKGQNFELIPFGAGRRICPGLPLATRMVHLMLASLIHSCDWKLEDGITPENMNMEDRFGITLQKAQ 490
GwCYP76F12  LDMDVKGQNFELIPFGAGRRICPGLPLATRMVHLMLASLIHSCDWKLEDGMTPENMNMEDRFGITLQKAQ 490

PnCYP76F12  PLKAIPIRV* 500
MoCYP76F12  PLKAIPIRV* 500
GwCYP76F12  PLKAIPIRV* 500
```

```
PnCYP76F14  ATGGAGTTGTTGAGTTGTCTGCTGTGTTTTCTTGCTGCTTGGACTTCAATTTACATCATGTTTTCAGCCA 70
MoCYP76F14  ATGGAGTTGTTGAGTTGTCTGCTGTGTTTTCTTGCTGCTTGGACTTCAATTTACATCATGTTTTCAGCCA 70
GwCYP76F14  ATGGAGTTGTTGAGTTGTCTGCTGTGTTTTCTTGCTGCTTGGACTTCAATTTACATCATGTTTTCAGCCA 70

PnCYP76F14  GAAGAGGAAGAAAGCATGCTGCTCACAAACTTCCTCCAGGACCAGTTCCGCTTCCCATAATAGGAAGCCT 140
MoCYP76F14  GAAGAGGAAGAAAGCATGCTGCTCACAAACTTCCTCCAGGACCAGTTCCGCTTCCCATAATAGGAAGCCT 140
GwCYP76F14  GAAGAGGAAGAAAGCATGCTGCTCACAAACTTCCTCCAGGACCAGTTCCGCTTCCCATAATAGGAAGCCT 140

PnCYP76F14  CTTAAACCTGGGTAACAGACCCCATGAGTCCCTCGCCAATCTTGCAAAAACTTACGGCCCAATTATGACT 210
MoCYP76F14  CTTAAACCTGGGTAACAGACCCCATGAGTCCCTCGCCAATCTTGCAAAAACTTACGGCCCAATTATGACT 210
GwCYP76F14  CTTAAACCTGGGTAACAGACCCCATGAGTCCCTCGCCAATCTTGCAAAAACTTACGGCCCAATTATGACT 210

PnCYP76F14  CTCAAACTTGGCTATGTAACCACAATAGTCATCTCTTCTGCCCCCATGGCCAAAGAAGTCCTCCAAAAGC 280
MoCYP76F14  CTCAAACTTGGCTATGTAACCACAATAGTCATCTCTTCTGCCCCCATGGCCAAAGAAGTCCTCCAAAAGC 280
GwCYP76F14  CTCAAACTTGGCTATGTAACCACAATAGTCATCTCTTCTGCCCCCATGGCCAAAGAAGTCCTCCAAAAGC 280

PnCYP76F14  AGGATCTCTCCTTCTGCAACCGATCCATCCCTGATGCCATCCGAGCCGCAAAACACAACCAACTGTCGAT 350
MoCYP76F14  AGGATCTCTCCTTCTGCAACCGATCAATCCCTGATGCCATCCGAGCCGCAAAACACAACCAACTGTCGAT 350
GwCYP76F14  AGGATCTCTCCTTCTGCAACCGATCCATCCCTGATGCCATCCGAGCCGCAAAACACAACCAACTGTCGAT 350

PnCYP76F14  GGCCTGGCTACCTGTTTCAACAACTTGGAGAGCCCTTCGAAGGACATGCAATTCGCACTTATTCACTCCT 420
MoCYP76F14  GGCCTGGATACCTGTTTCAACAACTTGGAGAGCCCTTCGAAGGACATGCAATTCGCACTTATTCACTTCT 420
GwCYP76F14  GGCCTGGCTACCTGTTTCAACAACTTGGAGAGCCCTTCGAAGGACATGCAATTCGCACTTATTCACTCCT 420

PnCYP76F14  CAAAAACTAGACTCTAACACCCATCTCCGCCACCAGAAAGTGCAAGAGCTTCTTGCGAATGTTGAACAGA 490
MoCYP76F14  CAAAAACTAGACTCTAACACCCATCTCCGCCACCAGAAAGTGCAAGAGCTTCTTGCGAATGTTGAACAGA 490
GwCYP76F14  CAAAAACTAGACTCTAACACCCATCTCCGCCACCAGAAAGTGCAAGAGCTTCTTGCGAATGTTGAACAGA 490

PnCYP76F14  GCTGCCAAGCTGGTGGTCCTGTAGATATAGGCCAAGAAGCTTTTAGAACTAGTCTCAACTTGTTATCCAA 560
MoCYP76F14  GCTGCCAAGCTGGTGGTCCTGTAGATATAGGCCAAGAAGCTTTTAGAACTAGTCTCAACTTGTTATCCAA 560
GwCYP76F14  GCTGCCAAGCTGGTGGTCCTGTAGATATAGGCCAAGAAGCTTTTAGAACTAGTCTCAACTTGTTATCCAA 560

PnCYP76F14  CACCATATTTTCTGTGGATCTTGTTGATCCAATTTCTGAGACTGCACAAGAGTTTAAAGAGTTGGTGCGT 630
MoCYP76F14  CACCATATTTTCTGTGGATCTTGTTGATCCAATTTCTGAGACTGCACAAGAGTTTAAAGAGTTGGTGCGT 630
GwCYP76F14  CACCATATTTTCTGTGGATCTTGTTGATCCAATTTCTGAGACTGCACAAGAGTTTAAAGAGTTGGTGCGT 630

PnCYP76F14  GGTGTGATGGAGGAAGCTGGGAAACCCAACTTGGTAGATTATTTTCCGGTGCTCAGACGGATTGATCCAC 700
MoCYP76F14  GGTGTGATGGAGGAAGCTGGGAAACCCAACTTGGTTGATTATTTTCCGGTGCTCAGACAGATTGATCCAC 700
GwCYP76F14  GGTGTGATGGAGGAAGCTGGGAAACCCAACTTGGTAGATTATTTTCCGGTGCTCAGACGGATTGATCCAC 700

PnCYP76F14  AAAGTATAAGGCGTCGTTTGACAATTTATTTTGGAAGGATGATTGAGATCTTTGATAGAATGATCAAGCA 770
MoCYP76F14  AAGGTATAAGGCGTCGTTTGACAATTTATTTTGGAAGGATGATTGAGATCTTTGATAGAATGATCAAGCA 770
GwCYP76F14  AAAGTATAAGGCGTCGTTTGACAATTTATTTTGGAAGGATGATTGAGATCTTTGATAGAATGATCAAGCA 770

PnCYP76F14  ACGGTTACAGCTAAGAAAAAATCAAGGTTCAATAGCTAGCAGTGATGTGTTAGACGTTCTTCTCAACATC 840
MoCYP76F14  ACGGTTACAGCTAAGAAAAATTCAAGGTTCAATAGCTAGCAGTGATGTGTTAGACGTTCTTCTCAACATC 840
GwCYP76F14  ACGGTTACAGCTAAGAAAAAATCAAGGTTCAATAGCTAGCAGTGATGTGTTAGACGTTCTTCTCAACATC 840

PnCYP76F14  AGTGAAGATAACAGCAGTGAGATTGAAAGAAGTCATATGGAACATTTGTTATTGGACTTATTTGCTGCGG 910
MoCYP76F14  AGTGAAGATAACAGCAATGAGATTGAAAGAAGTCATATGGAACATTTGTTATTGGACTTATTTGCTGCGG 910
GwCYP76F14  AGTGAAGATAACAGCAGTGAGATTGAAAGAAGTCATATGGAACATTTGTTATTGGACTTATTTGCTGCGG 910

PnCYP76F14  GGACTGACACAACTTCGAGCACATTGGAATGGGCAATGGCAGAGCTGCTACACAACCCTGAAACACTTTT 980
MoCYP76F14  GGACTGACACAACTTCGAGCACATTGGAATGGGCAATGGCAGAGCTGCTACACAACCCTGAAACACTTTT 980
GwCYP76F14  GGACTGACACAACTTCGAGCACATTGGAATGGGCAATGGCAGAGCTGCTACACAACCCTGAAACACTTTT 980

PnCYP76F14  GAAAGCCCGAATGGAACTCCTGCAAACCATCGGCCAAGACAAACAGGTAAAAGAATCAGACATCAGTCGA 1050
MoCYP76F14  GAAAGCCCGAATGGAACTCCTGCAAACCATCGGCCAAGACAAACAGGTAAAAGAATCAGACATCAGTCGA 1050
GwCYP76F14  GAAAGCCCGAATGGAACTCCTGCAAACCATCGGCCAAGACAAACAGGTAAAAGAATCAGACATCAGTCGA 1050

PnCYP76F14  CTCCCATACTTGCAAGCAGTGGTGAAAGAGACCTTCCGATTGCACCCAGCAGTTCCATTTTTACTCCCAC 1120
MoCYP76F14  CTCCCATACTTGCAAGCAGTGGTGAAAGAGACCTTCCGATTGCACCCAGCAGTTCCATTTTTACTCCCAC 1120
GwCYP76F14  CTCCCATACTTGCAAGCAGTGGTGAAAGAGACCTTCCGATTGCACCCAGCAGTTCCATTTTTACTCCCAC 1120
PnCYP76F14  GCAGAGTTGAAGGGGACGCAGATATAGATGGGTTCGCAGTCCCAAAGAACGCACAGGTGCTGGTGAACGC 1190
MoCYP76F14  GCAGAGTTGAAGGGGACGCGGATATAGATGGGTTCGCAGTCCCAAAGAACGCACAGGTGCTGGTGAACGC 1190
GwCYP76F14  GCAGAGTTGAAGGGGACGCAGATATAGATGGGTTCGCAGTCCCAAAGAACGCACAGGTGCTGGTGAACGC 1190

PnCYP76F14  ATGGGCTATAGGTCGAGACCCGAACACATGGGAGAACCCCAACTCATTTGTGCCAGAGAGGTTCTTGGGG 1260
MoCYP76F14  ATGGGCTATAGGTCGAGACCCGAACACATGGGAGAACCCCAACTCATTTGTGCCAGAGAGGTTCTTGGGG 1260
GwCYP76F14  ATGGGCTATAGGTCGAGACCCGAACACATGGGAGAACCCCAACTCATTTGTGCCAGAGAGGTTCTTGGGG 1260

PnCYP76F14  TTGGACATGGACGTGAAGGGCCAGAATTTTGAGCTGATTCCGTTTGGTGCTGGCAGGAGAATCTGTCCTG 1330
MoCYP76F14  TTGGACATGGACGTGAAGGGCCAGAATTTTGAGCTGATTCCGTTTGGTGCTGGCAGGAGAATCTGTCCTG 1330
GwCYP76F14  TTGGACATGGACGTGAAGGGCCAGAATTTTGAGCTGATTCCGTTTGGTGCTGGCAGGAGAATCTGTCCTG 1330
```

```
PnCYP76F14  GGCTGCCATTGGCAATCCGGATGGTTCACTTGATGCTGGCCTCGCTCATTCACTCCTATGATTGGAAACT 1400
MoCYP76F14  GGTTGCCATTGGCAATCCGGATGGTTCACTTGATGCTGGCCTCGCTCATTCACTCCTATGATTGGAAACT 1400
GwCYP76F14  GGCTGCCATTGGCAATCCGGATGGTTCACTTGATGCTGGCCTCGCTCATTCACTCCTATGATTGGAAACT 1400

PnCYP76F14  TGAAGATGGGGTGACACCGGAAAACATGAACATGGAAGAAAGATATGGCATTAGTTTACAAAAGGCTCAG 1470
MoCYP76F14  TGAAGATGGGGTGACACCGGAGAACATGAACATGGAAGAAAGATATGGTATTAGTTTACAAAAGGCTCAG 1470
GwCYP76F14  TGAAGATGGGGTGACACCGGAAAACATGAACATGGAAGAAAGATATGGCATTAGTTTACAAAAGGCTCAG 1470

PnCYP76F14  CCCCTGCAAGCTCTACCTGTACGGGTTTGA 1500
MoCYP76F14  CCCCTGCAAGCTCTACCTGTACGGGTTTGA 1500
GwCYP76F14  CCCCTGCAAGCTCTACCTGTACGGGTTTGA 1500
```

```
PnCYP76F14  MELLSCLLCFLAAWTSIYIMFSARRGRKHAAHKLPPGPVPLPIIGSLLNLGNRPHESLANLAKTYGPIMT 70
MoCYP76F14  MELLSCLLCFLAAWTSIYIMFSARRGRKHAAHKLPPGPVPLPIIGSLLNLGNRPHESLANLAKTYGPIMT 70
GwCYP76F14  MELLSCLLCFLAAWTSIYIMFSARRGRKHAAHKLPPGPVPLPIIGSLLNLGNRPHESLANLAKTYGPIMT 70

PnCYP76F14  LKLGYVTTIVISSAPMAKEVLQKQDLSFCNRSIPDAIRAAKHNQLSMAWLPVSTTWRALRRTCNSHLFTP 140
MoCYP76F14  LKLGYVTTIVISSAPMAKEVLQKQDLSFCNRSIPDAIRAAKHNQLSMAWIPVSTTWRALRRTCNSHLFTS 140
GwCYP76F14  LKLGYVTTIVISSAPMAKEVLQKQDLSFCNRSIPDAIRAAKHNQLSMAWLPVSTTWRALRRTCNSHLFTP 140

PnCYP76F14  QKLDSNTHLRHQKVQELLANVEQSCQAGGPVDIGQEAFRTSLNLLSNTIFSVDLVDPISETAQEFKELVR 210
MoCYP76F14  QKLDSNTHLRHQKVQELLANVEQSCQAGGPVDIGQEAFRTSLNLLSNTIFSVDLVDPISETAQEFKELVR 210
GwCYP76F14  QKLDSNTHLRHQKVQELLANVEQSCQAGGPVDIGQEAFRTSLNLLSNTIFSVDLVDPISETAQEFKELVR 210

PnCYP76F14  GVMEEAGKPNLVDYFPVLRRIDPQSIRRRLTIYFGRMIEIFDRMIKQRLQLRKNQGSIASSDVLDVLLNI 280
MoCYP76F14  GVMEEAGKPNLVDYFPVLRQIDPQGIRRRLTIYFGRMIEIFDRMIKQRLQLRKIQGSIASSDVLDVLLNI 280
GwCYP76F14  GVMEEAGKPNLVDYFPVLRRIDPQSIRRRLTIYFGRMIEIFDRMIKQRLQLRKNQGSIASSDVLDVLLNI 280

PnCYP76F14  SEDNSSEIERSHMEHLLLDLFAAGTDTTSSTLEWAMAELLHNPETLLKARMELLQTIGQDKQVKESDISR 350
MoCYP76F14  SEDNSNEIERSHMEHLLLDLFAAGTDTTSSTLEWAMAELLHNPETLLKARMELLQTIGQDKQVKESDISR 350
GwCYP76F14  SEDNSSEIERSHMEHLLLDLFAAGTDTTSSTLEWAMAELLHNPETLLKARMELLQTIGQDKQVKESDISR 350

PnCYP76F14  LPYLQAVVKETFRLHPAVPFLLPRRVEGDADIDGFAVPKNAQVLVNAWAIGRDPNTWENPNSFVPERFLG 420
MoCYP76F14  LPYLQAVVKETFRLHPAVPFLLPRRVEGDADIDGFAVPKNAQVLVNAWAIGRDPNTWENPNSFVPERFLG 420
GwCYP76F14  LPYLQAVVKETFRLHPAVPFLLPRRVEGDADIDGFAVPKNAQVLVNAWAIGRDPNTWENPNSFVPERFLG 420

PnCYP76F14  LDMDVKGQNFELIPFGAGRRICPGLPLAIRMVHLMLASLIHSYDWKLEDGVTPENMNMEERYGISLQKAQ 490
MoCYP76F14  LDMDVKGQNFELIPFGAGRRICPGLPLAIRMVHLMLASLIHSYDWKLEDGVTPENMNMEERYGISLQKAQ 490
GwCYP76F14  LDMDVKGQNFELIPFGAGRRICPGLPLAIRMVHLMLASLIHSYDWKLEDGVTPENMNMEERYGISLQKAQ 490

PnCYP76F14  PLQALPVRV* 500
MoCYP76F14  PLQALPVRV* 500
GwCYP76F14  PLQALPVRV* 500
```

```
PnCYP76T21  ATGGATTACACCCCACTTGTTCTTCTACTTCTTCTTCCTTGCTTTGTCTGGTTATGCTTCCATTTCCTCA 70
MoCYP76T21  ATGGATTACACCCCACTTGTTCTTCTACTTCTTCTTCCTTGCTTTGTCTGGTTATGCTTCCATTTCCTCA 70
GwCYP76T21  ATGGATTACACCCCACTTGTTCTTCTACTTCTTCTTCCTTGCTTTGTCTGGTTATGCTTCCATTTCCTCA 70

PnCYP76T21  TCCTTGGCTCCACCCATCGGAAATCCTTCCAAGCCAGGCTTCCGCCGGGCCCCCGCCCTCTACCCATCAT 140
MoCYP76T21  TCCTTGGCTCCACCCATCGGAAATCCTTCCAAGCCAGGCTTCCGCCGGGCCCCCGCCCTCTACCCATCAT 140
GwCYP76T21  TCCTTGGCTCCACCCATCGGAAATCCTTCCAAGCCAGGCTTCCGCCGGGCCCCCGCCCTCTACCCATCAT 140

PnCYP76T21  CGGAAACCTCCTTGAACTCGGCGATAAACCCCACCAATCATTCACAACTCTTTCCAAAACGTATGGCCCT 210
MoCYP76T21  CGGAAACCTCCTTGAACTCGGCGATAAACCCCACCAATCACTCACAACTCTTTCCAAAACGTATGGCCCT 210
GwCYP76T21  CGGAAACCTCCTTGAACTCGGCGATAAACCCCACCAATCACTCACAACTCTTTCCAAAACGTATGGCCCT 210

PnCYP76T21  CTGATGTCTCTCAAGCTAGGAAGCACCACCACCATAGTCATTTCCTCGCCCAAAACAGCCCAAGAAGTAC 280
MoCYP76T21  CTGATGTCTCTCAAGCTAGGAAGCACCACCACCATAGTCATTTCCTCGCCCAAAACAGCCCAAGAAGTAC 280
GwCYP76T21  CTGATGTCTCTCAAGCTAGGAAGCACCACCACCATAGTCATTTCCTCGCCCAAAACAGCCCAAGAAGTAC 280

PnCYP76T21  TAAACAAAAAAGACCAAGCCTTCGCCAGCAGGACAGTTCTCAATGCCATCCAAATCCAAGACCATCACAA 350
MoCYP76T21  TAAACAAAAAAGACCAAGCCTTCGCCAGCAGGACAGTTCTCAATGCCATCCAAATCCAAGACCATCACAA 350
GwCYP76T21  TAAACAAAAAAGACCAAGCCTTCGCCAGCAGGACAGTTCTCAATGCCATCCAAATCCAAGACCATCACAA 350

PnCYP76T21  GTTTTCAATGGTCTTTTTACCCGCGTCTGCTCATTGGCGCAACCTCAGGAAGATTTGCAGCATGCAAATA 420
MoCYP76T21  GTTTTCAATGGTCTTTTTACCCGCGTCTGCTCATTGGCGCAACCTCAGGAAGATTTGCAGCATGCAAATA 420
GwCYP76T21  GTTTTCAATGGTCTTTTTACCCGCGTCTGCTCATTGGCGCAACCTCAGGAAGATTTGCAGCATGCAAATA 420

PnCYP76T21  TTTTCCCCGCAACGTGTTGAGGCCAGCCAAGACCTGCGTCGAAAAGTTGTGCAACAACTTCTAGAGCATG 490
MoCYP76T21  TTTTCCCCGCAACGTGTTGAGGCCAGCCAAGACCTGCGTCGAAAAGTTGTGCAACAACTTCTAGAGCATG 490
GwCYP76T21  TTTTCCCCGCAACGTGTTGAGGCCAGCCAAGACCTGCGTCGAAAAGTTGTGCAACAACTTCTAGAGCATG 490

PnCYP76T21  CCCGTGAAAGTTGCAACAGCGGTCGGGCAGTCGATGTTGGCAGAGCAGCCTTCACAACCACCCTCAATTT 560
MoCYP76T21  CCCGTGAAAGTTGCAACAGCGGTCGGGCAGTCGATGTTGGCAGAGCAGCCTTCACAACCACCCTCAATTT 560
GwCYP76T21  CCCGTGAAAGTTGCAACAGCGGTCGGGCAGTCGATGTTGGCAGAGCAGCCTTCACAACCACCCTCAATTT 560

PnCYP76T21  GTTATCGAACACTTTTTTCTCTGTTGATTTGGCTCACTATGATTCCAATTTGTCACAAGAGTTCAAGGAC 630
MoCYP76T21  GTTATCGAACACTTTTTTCTCTGTTGATTTGGCTCACTATGATTCCAATTTGTCACAAGAGTTCAAGGAC 630
GwCYP76T21  GTTATCGAACACTTTTTTCTCTGTTGATTTGGCTCACTATGATTCCAATTTGTCACAAGAGTTCAAGGAC 630

PnCYP76T21  CTTATATGGAGTATAATGGTAGAAGCTGGAAAGCCTAATCTTGCAGACTTCTTTCCGGGCCTCAGATTGG 700
MoCYP76T21  CTTATATGGAGTATAATGGTAGAAGCTGGAAAGCCTAATCTTGCAGACTTCTTTCCGGGCCTCAGATTGG 700
GwCYP76T21  CTTATATGGAGTATAATGGTAGAAGCTGGAAAGCCTAATCTTGCAGACTTCTTTCCGGGCCTCAGATTGG 700

PnCYP76T21  TTGATCCACAAGGAATACAGAAAAGGATGACGGTTTATTTTAATAAACTGTTAGATGTTTTTGACGGTTT 770
MoCYP76T21  TTGATCCACAAGGAATACAGAAAAGGATGACGGTTTATTTTAATAAACTGTTAGATGTTTTTGACGGTTT 770
GwCYP76T21  TTGATCCACAAGGAATACAGAAAAGGATGACGGTTTATTTTAATAAACTGTTAGATGTTTTTGACGGTTT 770

PnCYP76T21  TATCAATCAAAGGTTACCGTTAAAAGCTTCTTCTCCGGACAACGATGTACTAGATGCCCTCCTCAATCTC 840
MoCYP76T21  TATCAATCAAAGGTTACCGTTAAAAGCTTCTTCTCCGGACAACGATGTACTAGATGCCCTCCTCAATCTC 840
GwCYP76T21  TATCAATCAAAGGTTACCGTTAAAAGCTTCTTCTCCGGACAACGATGTACTAGATGCCCTCCTCAATCTC 840

PnCYP76T21  AACAAACAACATGACCATGAGTTGAGCTCCAACGATATCAGACATTTGCTTACTGACCTATTCTCTGCGG 910
MoCYP76T21  AACAAACAACATGACCATGAGTTGAGCTCCAACGATATCAGACATTTGCTTACTGACTTATTCTCTGCGG 910
GwCYP76T21  AACAAACAACATGACCATGAGTTGAGCTCCAACGATATCAGACATTTGCTTACTGACTTATTCTCTGCGG 910

PnCYP76T21  GAACAGACACCATTTCGAGCACGATAGAGTGGGCAATGGCTGAGTTATTAAACAACCCTAAAGCGATGGC 980
MoCYP76T21  GAACAGACACCATTTCGAGCACGATAGAGTGGGCAATGGCTGAGTTATTAAACAACCCTAAAGCGATGGC 980
GwCYP76T21  GAACAGACACCATTTCGAGCACGATAGAGTGGGCAATGGCTGAGTTATTAAACAACCCTAAAGCGATGGC 980

PnCYP76T21  GAAAGCTCAAGATGAACTCAGTCAAGTAGTGGGCAAAGACAGGATAGTTGAAGAATCAGACGTCACAAAG 1050
MoCYP76T21  GAAAGCTCAAGATGAACTCAGTCAAGTAGTGGGCAAAGACAGGATAGTTGAAGAATCAGACGTCACAAAG 1050
GwCYP76T21  GAAAGCTCAAGATGAACTCAGTCAAGTAGTGGGCAAAGACAGGATAGTTGAAGAATCAGACGTCACAAAG 1050

PnCYP76T21  CTCCCTTATTTACAGGCAGTAGTAAAAGAAACCTTCAGGTTGCACCCACCCGCCCCATTCCTGGTTCCTA 1120
MoCYP76T21  CTCCCTTATTTACAGGCAGTAGTAAAAGAAACCTTCAGGTTGCACCCACCCGCCCCATTCCTGGTTCCTA 1120
GwCYP76T21  CTCCCTTATTTACAGGCAGTAGTAAAAGAAACCTTCAGGTTGCACCCACCCGCCCCATTCCTGGTTCCTA 1120

PnCYP76T21  GAAAGGCCGAAATGGACTCAGAAATATTAGGTTATGCGGTGCCCAAAAATGCACAAGTACTTGTCAATGT 1190
MoCYP76T21  GAAAGGCCGAAATGGACTCAGAAATATTAGGTTATGCGGTGCCCAAAAATGCACAAGTACTTGTCAATGT 1190
GwCYP76T21  GAAAGGCCGAAATGGACTCAGAAATATTAGGTTATGCGGTGCCCAAAAATGCACAAGTACTTGTCAATGT 1190

PnCYP76T21  GTGGGCTATTGGCAGAGATTCGCGGACGTGGTCGAACCCGAATTCGTTTGTGCCTGAAAGGTTTTTAGAG 1260
MoCYP76T21  GTGGGCTATTGGCAGAGATTCGCGGACGTGGTCGAATCCGAATTCATTTGTGCCTGAAAGGTTTTTAGAG 1260
GwCYP76T21  GTGGGCTATTGGCAGAGATTCGCGGACGTGGTCGAATCCGAATTCATTTGTGCCTGAAAGGTTTTTAGAG 1260

PnCYP76T21  TGCCAAATTGATGTCAAGGGCCGAGATTTCCAACTCATTCCCTTCGGCGCTGGAAGAAGAATCTGTCCTG 1330
MoCYP76T21  TGCCAAATTGATGTCAAGGGCCGAGATTTCCAACTCATTCCCTTCGGCGCTGGAAGAAGAATCTGTCCTG 1330
GwCYP76T21  TGCCAAATTGATGTCAAGGGCCGAGATTTCCAACTCATTCCCTTCGGCGCTGGAAGAAGAATCTGTCCTG 1330
```

```
PnCYP76T21  GATTGCTGTTAGGCCATAGGATGGTGCACTTGATGTTGGCTTCTCTTCTTCACTCCTTTGATTGGAAGCT 1400
MoCYP76T21  GATTGCTGTTAGGCCATAGGATGGTGCACTTGATGTTGGCTTCTCTTCTTCACTCCTTTGATTGGAAGCT 1400
GwCYP76T21  GATTGCTGTTAGGCCATAGGATGGTGCACTTGATGTTGGCTTCTCTTCTTCACTCCTTTGATTGGAAGCT 1400

PnCYP76T21  TGAAGATAGTATGAGACCGGAAGACATGGACATGAGTGAGAAGTTTGGATTTACATTACGAAAAGCCCAA 1470
MoCYP76T21  TGAAGATAGTATGAGACCGGAAGACATGGACATGAGTGAGAAGTTTGGATTTACATTACGAAAAGCCCAA 1470
GwCYP76T21  TGAAGATAGTATGAGACCGGAAGACATGGACATGAGTGAGAAGTTTGGATTTACATTACGAAAAGCCCAA 1470

PnCYP76T21  CCTCTCCGGGCTGTTCCCACCAAACCATGA 1500
MoCYP76T21  CCTCTCCGGGCTGTTCCCACCAAACCATGA 1500
GwCYP76T21  CCTCTCCGGGCTGTTCCCACCAAACCATGA 1500




PnCYP76T21  MDYTPLVLLLLLPCFVWLCFHFLILGSTHRKSFQARLPPGPRPLPIIGNLLELGDKPHQSFTTLSKTYGP 70
MoCYP76T21  MDYTPLVLLLLLPCFVWLCFHFLILGSTHRKSFQARLPPGPRPLPIIGNLLELGDKPHQSLTTLSKTYGP 70
GwCYP76T21  MDYTPLVLLLLLPCFVWLCFHFLILGSTHRKSFQARLPPGPRPLPIIGNLLELGDKPHQSLTTLSKTYGP 70

PnCYP76T21  LMSLKLGSTTTIVISSPKTAQEVLNKKDQAFASRTVLNAIQIQDHHKFSMVFLPASAHWRNLRKICSMQI 140
MoCYP76T21  LMSLKLGSTTTIVISSPKTAQEVLNKKDQAFASRTVLNAIQIQDHHKFSMVFLPASAHWRNLRKICSMQI 140
GwCYP76T21  LMSLKLGSTTTIVISSPKTAQEVLNKKDQAFASRTVLNAIQIQDHHKFSMVFLPASAHWRNLRKICSMQI 140

PnCYP76T21  FSPQRVEASQDLRRKVVQQLLEHARESCNSGRAVDVGRAAFTTTLNLLSNTFFSVDLAHYDSNLSQEFKD 210
MoCYP76T21  FSPQRVEASQDLRRKVVQQLLEHARESCNSGRAVDVGRAAFTTTLNLLSNTFFSVDLAHYDSNLSQEFKD 210
GwCYP76T21  FSPQRVEASQDLRRKVVQQLLEHARESCNSGRAVDVGRAAFTTTLNLLSNTFFSVDLAHYDSNLSQEFKD 210

PnCYP76T21  LIWSIMVEAGKPNLADFFPGLRLVDPQGIQKRMTVYFNKLLDVFDGFINQRLPLKASSPDNDVLDALLNL 280
MoCYP76T21  LIWSIMVEAGKPNLADFFPGLRLVDPQGIQKRMTVYFNKLLDVFDGFINQRLPLKASSPDNDVLDALLNL 280
GwCYP76T21  LIWSIMVEAGKPNLADFFPGLRLVDPQGIQKRMTVYFNKLLDVFDGFINQRLPLKASSPDNDVLDALLNL 280

PnCYP76T21  NKQHDHELSSNDIRHLLTDLFSAGTDTISSTIEWAMAELLNNPKAMAKAQDELSQVVGKDRIVEESDVTK 350
MoCYP76T21  NKQHDHELSSNDIRHLLTDLFSAGTDTISSTIEWAMAELLNNPKAMAKAQDELSQVVGKDRIVEESDVTK 350
GwCYP76T21  NKQHDHELSSNDIRHLLTDLFSAGTDTISSTIEWAMAELLNNPKAMAKAQDELSQVVGKDRIVEESDVTK 350

PnCYP76T21  LPYLQAVVKETFRLHPPAPFLVPRKAEMDSEILGYAVPKNAQVLVNVWAIGRDSRTWSNPNSFVPERFLE 420
MoCYP76T21  LPYLQAVVKETFRLHPPAPFLVPRKAEMDSEILGYAVPKNAQVLVNVWAIGRDSRTWSNPNSFVPERFLE 420
GwCYP76T21  LPYLQAVVKETFRLHPPAPFLVPRKAEMDSEILGYAVPKNAQVLVNVWAIGRDSRTWSNPNSFVPERFLE 420

PnCYP76T21  CQIDVKGRDFQLIPFGAGRRICPGLLLGHRMVHLMLASLLHSFDWKLEDSMRPEDMDMSEKFGFTLRKAQ 490
MoCYP76T21  CQIDVKGRDFQLIPFGAGRRICPGLLLGHRMVHLMLASLLHSFDWKLEDSMRPEDMDMSEKFGFTLRKAQ 490
GwCYP76T21  CQIDVKGRDFQLIPFGAGRRICPGLLLGHRMVHLMLASLLHSFDWKLEDSMRPEDMDMSEKFGFTLRKAQ 490

PnCYP76T21  PLRAVPTKP* 500
MoCYP76T21  PLRAVPTKP* 500
GwCYP76T21  PLRAVPTKP* 500
```

```
PnCYP76Y1    ATGGAACTTAACTCCTTCCTCTTGCTTTGCATGCCACTGGTCCTGTGCCTCTTCTTCCTTCAATTTCTGC 70
MoCYP76Y1    ATGGAACTTAACTCCTTCCTCTTGCTTTGCATGCCACTGGTCCTGTGCCTCTTCTTCCTTCAATTTCTGC 70
GwCYP76Y1v1  ATGGAACTTAACTCCTTCCTCTTGCTTTGCATGCCACTGGTCCTGTGCCTCTTCTTCCTTCAATTTCTGC 70
GwCYP76Y1v2  ATGGAACTTAACTCCTTCCTCTTGCTTTGCATGCCACTGGTCCTGTGCCTCTTCTTCCTTCAATTTCTGC 70

PnCYP76Y1    GCCCATCATCCCACGCTACCAAGCTACCGCCTGGCCCGACTGGCCTCCCCATTTTGGGCTCCCTACTCCA 140
MoCYP76Y1    GCCCATCATCCCACGCTACCAAGCTACCGCCTGGCCCGACTGGCCTCCCCATTTTGGGCTCCCTACTCCA 140
GwCYP76Y1v1  GCCCATCATCCCACGCTACCAAGCTACCGCCTGGCCCGACTGGCCTCCCCATTTTGGGCTCCCTACTCCA 140
GwCYP76Y1v2  GCCCATCATCCCACGCTACCAAGCTACCGCCTGGCCCGACTGGCCTCCCCATTTTGGGCTCCCTACTCCA 140

PnCYP76Y1    AATTGGCAAACTTCCTCATCACTCACTTGCAAGATTGGCGAAAATCCACGGTCCTCTCATCACCCTCCGC 210
MoCYP76Y1    AATTGGCAAACTTCCTCATCACTCACTTGCAAGATTGGCGAAAATCCACGGTCCTCTCATCACCCTCCGC 210
GwCYP76Y1v1  AATTGGCAAACTTCCTCATCACTCACTTGCAAGATTGGCGAAAATCCACGGTCCTCTCATCACCCTCCGC 210
GwCYP76Y1v2  AATTGGCAAACTTCCTCATCACTCACTTGCAAGATTGGCGAAAATCCACGGTCCTCTCATCACCCTCCGC 210

PnCYP76Y1    CTTGGGTCCATCACCACCGTCGTCGCCTCCTCTCCCCAGACAGCCAAACTAATTCTCCAAACACATGGCC 280
MoCYP76Y1    CTTGGGTCCATCACCACCGTCGTCGCCTCCTCTCCCCAGACAGCCAAACTAATTTTCCAAACACATGGCC 280
GwCYP76Y1v1  CTTGGGTCCATCACCACCGTCGTCGCCTCCTCTCCCCAGACAGCCAAACTAATTCTCCAAACACATGGCC 280
GwCYP76Y1v2  CTTGGGTCCATCACCACCGTCGTCGCCTCCTCTCCCCAGACAGCCAAACTAATTCTCCAAACACATGGCC 280

PnCYP76Y1    AGAATTTCCTGGACCGTCCTGTTCCCGAGGCCATCGACAGTCCTCAAGGCACAATCGCATGGACTCCTGT 350
MoCYP76Y1    AGAACTTCCTGGACCGTCCTGTTCCCGAGGCCGTCGACAGTCCTCAAGGCACAATCGCATGGACTCCTGT 350
GwCYP76Y1v1  AGAACTTCCTGGACCGTCCTGTTCCCGAGGCCATCGACAGTCCTCAAGGCACAATCGCATGGACTCCTGT 350
GwCYP76Y1v2  AGAACTTCCTGGACCGTCCTGTTCCCGAGGCCATCGACAGTCCTCAAGGCACAATCGCATGGACTCCTGT 350

PnCYP76Y1    GGACCATGTATGGCGCAGCCGCCGCCGTGTTTGCAACAACCACTTGTTCACATCCCAGAGCCTGGACTCA 420
MoCYP76Y1    GGACCATGTATGGCGCAGCCGCCGCCGTGTTTGCAACAACCACTTGTTCACATCCCAGAGCCTGGACTCA 420
GwCYP76Y1v1  GGACCATGTATGGCGCAGCCGCCGCCGTGTTTGCAACAACCACTTGTTCACATCCCAGAGCCTGGACTCA 420
GwCYP76Y1v2  GGACCATGTATGGCGCAGCCGCCGCCGTGTTTGCAACAACCACTTGTTCACATCCCAGAGCCTGGACTCA 420

PnCYP76Y1    CTCCAACACCTTCGATACAAAAAGGTGGAACAACTTCTCCAACATATCCGTAAGCATTGTGTTTCCGGTA 490
MoCYP76Y1    CTCCAACACCTTCGATACAAAAAGGTGGAACAACTTCTCCAACATATCCGTAAGCATTGTGTTTCCGGTA 490
GwCYP76Y1v1  CTCCAACACCTTCGATACAAAAAGGTGGAACAACTTCTCCAACATATCCGTAAGCATTGTGTTTCCGGTA 490
GwCYP76Y1v2  CTCCAACACCTTCGATACAAAAAGGTGGAACAACTTCTCCAACATATCCGTAAGCATTGTGTTTCCGGTA 490

PnCYP76Y1    CACCAGTGGATATCGGCCTACTCGCCTCTGCCACCAACTTGAACGTGCTATCAAACGCCATTTTCTCTGT 560
MoCYP76Y1    CACCAGTGGATATCGGCCTACTCGCCTCTGCCACCAACTTGAACGTGCTATCAAACGCCATTTTCTCTGT 560
GwCYP76Y1v1  CACCAGTGGATATCGGCCTACTCGCCTCTGCCACCAACTTGAACGTGCTATCAAACGCCATTTTCTCTGT 560
GwCYP76Y1v2  CACCAGTGGATATCGGCCTACTCGCCTCTGCCACCAACTTGAACGTGCTATCAAACGCCATTTTCTCTGT 560

PnCYP76Y1    TGACCTTGTTGATCCAGGATTTGAGTCGGCTCAGGATTTCAGGGATCTGGTGTGGGGAATCATGGAGGGT 630
MoCYP76Y1    TGACCTTGTTGATCCAGGATTTGAGTCGGCTCAGGATTTCAGGGATCTGGTGTGGGGAATCATGGAGGGT 630
GwCYP76Y1v1  TGACCTTGTTGATCCAGGATTTGAGTCGGCTCAGGATTTCAGGGATCTGGTGTGGGGAATCATGGAGGGT 630
GwCYP76Y1v2  TGACCTTGTTGATCCAGGATTTGAGTCGGCTCAGGATTTCAGGGATCTGGTGTGGGGAATCATGGAGGGT 630

PnCYP76Y1    GCTGGCAAGTTTAATATTTCAGATTATTTTCCCATGTTTCGAAGGTTCGATTTGCTAGGTGTGAAGCGTG 700
MoCYP76Y1    GCTGGCAAGTTTAATATTTCAGATTATTTTCCCATGTTTCGAAGGTTCGATTTGCTAGGTGTGAAGCGTG 700
GwCYP76Y1v1  GCTGGCAAGTTTAATATTTCAGATTATTTTCCCATGTTTCGAAGGTTCGATTTGCTAGGTGTGAAGCGTG 700
GwCYP76Y1v2  GCTGGCAAGTTTAATATTTCAGATTATTTTCCCATGTTTCGAAGGTTCGATTTGCTAGGTGTGAAGCGTG 700

PnCYP76Y1    ACACTTTTTCATCTTATAGAAGGTTTTATGAAATAGTTGGTGATATAATCAAAAGCCGTATCAAGTGTAG 770
MoCYP76Y1    ACACTTTTTCATCTTATAGAAGGTTTTATGAAATAGTTGGTGATATAATCAAAAGCCGTATCAAGTGTAG 770
GwCYP76Y1v1  ACACTTTTTCATCTTATAGAAGGTTTTATGAAATAGTTGGTGATATAATCAAAAGCCGTATCAAGTGTAG 770
GwCYP76Y1v2  ACACTTTTTCATCTTATAGAAGGTTTTATGAAATAGTTGGTGATATAGTCAAAAGCCGTATCAAGTGTAG 770

PnCYP76Y1    AGCCTCCAATCCAGTGACCAGGAATGACGATTTCTTGGATGTGATTCTCGATCAGTGCCAAGAAGATGGT 840
MoCYP76Y1    AGCCTCCAATCCGGTGAGCAGGAATGAAGACTTCTTGGATGTGATTCTTGATCAGTGCCAAGAAGATGGC 840
GwCYP76Y1v1  AGCCTCCAATCCAGTGACCAGGAATGACGATTTCTTGGATGTGATTCTCGATCAGTGCCAAGAAGATGGC 840
GwCYP76Y1v2  AGCCTCCAATCCGGTGAGCAGGAATGACGACTTCTTGGATGTGATTCTTGATCAGTGCCAAGAAGATGTC 840



PnCYP76Y1   TCTTTATTCGATTCTGAAAATATCCAGGTTTTGATTGTGGAATTGTTTTATGCTGGAAGTGATACATCTA 910
MoCYP76Y1   TCTTATTTCGATTCTGAAAATATCCAGGTTTTGATTGTGGAACTGTTTTATGCTGGAAGTGATACATCTA 910
GwCYP76Y1v1 TCTTTATTCGATTCTGAAAATATCCAGGTTTTGATTGTGGAATTGTTTTATGCTGGAAGTGATACATCTA 910
GwCYP76Y1v2 TCTTATTTCGATTCTGAAAATATCCAGGTTTTGATTGTGGAACTGTTTTATGCTGGAAGTGATACATCTA 910

PnCYP76Y1   CCATAACAACTGAATGGGCAATGACTGAATTTCTTCGAAATCCAGGGGTGATGCAAAAGGTTCGGCAGGA 980
MoCYP76Y1   CCATAACAACTGAATGGGCAATGACTGAATTTCTTCGAAATCCAGGGGTGATGCAAAAGGTTCGGCAGGA 980
GwCYP76Y1v1 CCATAACAACTGAATGGGCAATGACTGAATTTCTTCGAAATCCAGGGGTGATGCAAAAGGTTCGGCAGGA 980
GwCYP76Y1v2 CCATAACAACTGAATGGGCAATGACTGAATTTCTTCGAAATCCAGGGGTGATGCAAAAGGTTCGGCAGGA 980

PnCYP76Y1   ACTTAGCGAAGTAATCGGGGCAGGTCAAATGGTTAGAGAATCAGATATGGATCGACTTCCATATTTTCAA 1050
MoCYP76Y1   ACTTAGCGAAGTAATCGGGGCAGGTCAAATGGTTAGAGAATCAGATATGGATCGACTTCCATATTTTCAA 1050
GwCYP76Y1v1 ACTTAGCGAAGTAATCGGGGCAGGTCAAATGGTTAGAGAATCAGATATGGATCGACTTCCATATTTTCAA 1050
GwCYP76Y1v2 ACTTAGCGACGTAATCGGGGCAGGTCAAATGGTTAGAGAAGCAGATATGGATCGACTTCCATATTTTCAA 1050
```

```
PnCYP76Y1   GCTGTTGTGAAAGAGACACTCAGACTCCATCCGGCTGGGCCCCTTCTGTTACCTTTTAAAGCAAAGAATG 1120
MoCYP76Y1   GCTGTTGTGAAAGAGACACTCAGACTCCATCCGGCTGGGCCCCTTCTGTTACCTTTTAAAGCAAAGAATG 1120
GwCYP76Y1v1 GCTGTTGTGAAAGAGACACTCAGACTCCATCCGGCTGGGCCCCTTCTGTTACCTTTTAAAGCAAAGAATG 1120
GwCYP76Y1v2 GCTGTTGTGAAAGAGACACTCAGACTCCATCCTGCTGGGCCCCTTCTGTTACCTTTTAAAGCAAAGAATG 1120

PnCYP76Y1   ATGTGGAATTATCCGGTTTCACCATACCCAGTAACAGTCATGTCCTTGTGAATATGTGGGCTATTGCAAG 1190
MoCYP76Y1   ATGTGGAATTATCCGGTTTCACCATACCCAGTAACAGTCATGTCCTTGTGAATATGTGGGCTATTGCAAG 1190
GwCYP76Y1v1 ATGTGGAATTATCCGGTTTCACCATACCCAGTAACAGTCATGTCCTTGTGAATATGTGGGCTATTGCAAG 1190
GwCYP76Y1v2 ATGTGGAATTATCCGGTTTCACCATACCCAGTAACAGTCATGTCCTTGTGAATATGTGGGCTATTGCAAG 1190

PnCYP76Y1   AGATCCAAGTTATTGGGAGGATCCTTTATCCTTCCTTCCTGAAAGATTCTTGGGCTCTAAGATAGATTAT 1260
MoCYP76Y1   AGATCCAAGTTATTGGGAGGATCCTTTATCCTTCCTTCCTGAAAGATTCTTGGGCTCTAAGATAGATTAT 1260
GwCYP76Y1v1 AGATCCAAGTTATTGGGAGGATCCTTTATCCTTCCTTCCTGAAAGATTCTTGGGCTCTAAGATAGATTAT 1260
GwCYP76Y1v2 AGATCCAAGTTATTGGGAGGATCCTTTATCCTTCCTTCCTGAAAGATTCTTGGGCTCTAAGATAGATTAT 1260

PnCYP76Y1   AGAGGCCAAGATTTTGAATATATACCATTTGGAGCTGGTAGGCGAATCTGCCCAGGCATGCCTCTTGCCG 1330
MoCYP76Y1   AGAGGCCAAGATTTTGAATATATACCATTTGGAGCTGGTAGGCGAATCTGCCCAGGCATGCCTCTTGCCG 1330
GwCYP76Y1v1 AGAGGCCAAGATTTTGAATATATACCATTTGGAGCTGGTAGGCGAATCTGCCCAGGCATGCCTCTTGCCG 1330
GwCYP76Y1v2 AGAGGCCAAGATTTTGAATATATACCATTTGGAGCTGGTAGGCGAATCTGCCCAGGCATGCCTCTTGCCG 1330

PnCYP76Y1   TCAGAATGGTTCAACTAGTGTTAGCTTCCATTATCCACTCCTTCAACTGGAAGCTTCCTGAAGGAACAAC 1400
MoCYP76Y1   TCAGAATGGTTCAACTAGTGTTAGCTTCCATTATCCACTCCTTCAACTGGAAGCTTCCTGAAGGAACAAC 1400
GwCYP76Y1v1 TCAGAATGGTTCAACTAGTGTTAGCTTCCATTATCCACTCCTTCAACTGGAAGCTTCCTGAAGGAACAAC 1400
GwCYP76Y1v2 TCAGAATGGTTCAACTAGTGTTAGCTTCCATTATCCACTCCTTCAACTGGAAGCTTCCTGAAGGAACAAC 1400

PnCYP76Y1   CCCACTGACCATTGACATGCAAGAACATTGCGGAGCTACCTTGAAGAAGGCCATTCCTCTTTCTGCCATT 1470
MoCYP76Y1   CCCACTGACCATTGACATGCAAGAACATTGCGGAGCTACCTTGAAGAAGGCCATTCCTCTTTCTGCCATT 1470
GwCYP76Y1v1 CCCACTGACCATTGACATGCAAGAACATTGCGGAGCTACCTTGAAGAAGGCCATTCCTCTTTCTGCCATT 1470
GwCYP76Y1v2 CCCACTGACCATTGACATGCAAGAACATTGCGGAGCTACCTTGAAGAAGGCCATTCCTCTTTCTGCCATT 1470

PnCYP76Y1   CCATTTATAGAAGAAAATTAA 1491
MoCYP76Y1   CCATTTATAGAAGAAAATTAA 1491
GwCYP76Y1v1 CCATTTATAGAAGAAAATTAA 1491
GwCYP76Y1v2 CCATTTATAGAAGAAAATTAA 1491
```

```
PnCYP76Y1    MELNSFLLLCMPLVLCLFFLQFLRPSSHATKLPPGPTGLPILGSLLQIGKLPHHSLARLAKIHGPLITLR 70
MoCYP76Y1    MELNSFLLLCMPLVLCLFFLQFLRPSSHATKLPPGPTGLPILGSLLQIGKLPHHSLARLAKIHGPLITLR 70
GwCYP76Y1v1  MELNSFLLLCMPLVLCLFFLQFLRPSSHATKLPPGPTGLPILGSLLQIGKLPHHSLARLAKIHGPLITLR 70
GwCYP76Y1v2  MELNSFLLLCMPLVLCLFFLQFLRPSSHATKLPPGPTGLPILGSLLQIGKLPHHSLARLAKIHGPLITLR 70

PnCYP76Y1    LGSITTVVASSPQTAKLILQTHGQNFLDRPVPEAIDSPQGTIAWTPVDHVWRSRRRVCNNHLFTSQSLDS 140
MoCYP76Y1    LGSITTVVASSPQTAKLIFQTHGQNFLDRPVPEAVDSPQGTIAWTPVDHVWRSRRRVCNNHLFTSQSLDS 140
GwCYP76Y1v1  LGSITTVVASSPQTAKLILQTHGQNFLDRPVPEAIDSPQGTIAWTPVDHVWRSRRRVCNNHLFTSQSLDS 140
GwCYP76Y1v2  LGSITTVVASSPQTAKLILQTHGQNFLDRPVPEAIDSPQGTIAWTPVDHVWRSRRRVCNNHLFTSQSLDS 140

PnCYP76Y1    LQHLRYKKVEQLLQHIRKHCVSGTPVDIGLLASATNLNVLSNAIFSVDLVDPGFESAQDFRDLVWGIMEG 210
MoCYP76Y1    LQHLRYKKVEQLLQHIRKHCVSGTPVDIGLLASATNLNVLSNAIFSVDLVDPGFESAQDFRDLVWGIMEG 210
GwCYP76Y1v1  LQHLRYKKVEQLLQHIRKHCVSGTPVDIGLLASATNLNVLSNAIFSVDLVDPGFESAQDFRDLVWGIMEG 210
GwCYP76Y1v2  LQHLRYKKVEQLLQHIRKHCVSGTPVDIGLLASATNLNVLSNAIFSVDLVDPGFESAQDFRDLVWGIMEG 210

PnCYP76Y1    AGKFNISDYFPMFRRFDLLGVKRDTFSSYRRFYEIVGDIIKSRIKCRASNPVTRNDDFLDVILDQCQEDG 280
MoCYP76Y1    AGKFNISDYFPMFRRFDLLGVKRDTFSSYRRFYEIVGDIIKSRIKCRASNPVSRNEDFLDVILDQCQEDG 280
GwCYP76Y1v1  AGKFNISDYFPMFRRFDLLGVKRDTFSSYRRFYEIVGDIIKSRIKCRASNPVTRNDDFLDVILDQCQEDG 280
GwCYP76Y1v2  AGKFNISDYFPMFRRFDLLGVKRDTFSSYRRFYEIVGDIVKSRIKCRASNPVSRNDDFLDVILDQCQEDV 280

PnCYP76Y1    SLFDSENIQVLIVELFYAGSDTSTITTEWAMTEFLRNPGVMQKVRQELSEVIGAGQMVRESDMDRLPYFQ 350
MoCYP76Y1    SYFDSENIQVLIVELFYAGSDTSTITTEWAMTEFLRNPGVMQKVRQELSEVIGAGQMVRESDMDRLPYFQ 350
GwCYP76Y1v1  SLFDSENIQVLIVELFYAGSDTSTITTEWAMTEFLRNPGVMQKVRQELSEVIGAGQMVRESDMDRLPYFQ 350
GwCYP76Y1v2  SYFDSENIQVLIVELFYAGSDTSTITTEWAMTEFLRNPGVMQKVRQELSDVIGAGQMVREADMDRLPYFQ 350

PnCYP76Y1    AVVKETLRLHPAGPLLLPFKAKNDVELSGFTIPSNSHVLVNMWAIARDPSYWEDPLSFLPERFLGSKIDY 420
MoCYP76Y1    AVVKETLRLHPAGPLLLPFKAKNDVELSGFTIPSNSHVLVNMWAIARDPSYWEDPLSFLPERFLGSKIDY 420
GwCYP76Y1v1  AVVKETLRLHPAGPLLLPFKAKNDVELSGFTIPSNSHVLVNMWAIARDPSYWEDPLSFLPERFLGSKIDY 420
GwCYP76Y1v2  AVVKETLRLHPAGPLLLPFKAKNDVELCGFTIPSNSHVLVNMWAIARDPSYWEDPLSFLPERFLGSKIDY 420

PnCYP76Y1    RGQDFEYIPFGAGRRICPGMPLAVRMVQLVLASIIHSFNWKLPEGTTPLTIDMQEHCGATLKKAIPLSAI 490
MoCYP76Y1    RGQDFEYIPFGAGRRICPGMPLAVRMVQLVLASIIHSFNWKLPEGTTPLTIDMQEHCGATLKKAIPLSAI 490
GwCYP76Y1v1  RGQDFEYIPFGAGRRICPGMPLAVRMVQLVLASIIHSFNWKLPEGTTPLTIDMQEHCGATLKKAIPLSAI 490
GwCYP76Y1v2  RGQDFEYIPFGAGRRICPGMPLAVRMVQLVLASIIHSFNWKLPEGTTPLTIDMQEHCGATLKKAIPLSAI 490

PnCYP76Y1    PFIEEN* 497
MoCYP76Y1    PFIEEN* 497
GwCYP76Y1v1  PFIEEN* 497
GwCYP76Y1v2  PFIEEN* 497
```

```
PnCYP76Y2  ATGGAACTTAACACCTTCCTCTTGCTTTGCATGCCACTCATCCTGTGCTTTTTCCTCCTCCAATTTCTGC 70
MoCYP76Y2  ATGGAACTTAACACCTTCCTCTTGCTTTGCATGCCACTCATCCTGTGCTTTTTCCTCCTCCAATTTCTGC 70
GwCYP76Y2  ATGGAACTTAACACCTTCCTCTTGCTTTGCATGCCACTCATCCTGTGCTTTTTCCTCCTCCAATTTCTGC 70

PnCYP76Y2  GCCCATCATCCCACGCTACCAAGCTACCGCCTGGCCCAACTGGCCTCCCTATTTTGGGCTCCCTACTGGA 140
MoCYP76Y2  GCCCATCATCCCACGCTACCAAGCTACCGCCTGGCCCAACTGGCCTCCCTATTTTGGGCTCCCTACTGGA 140
GwCYP76Y2  GCCCATCATCCCACGCTACCAAGCTACCGCCTGGCCCAACTGGCCTCCCTATTTTGGGCTCCCTACTGGA 140

PnCYP76Y2  AATTGGCAAACTTCCTCATCGCTCACTTGCAAGATTGGCCAAAATCCACGGCCCTCTCATCACTCTCCGC 210
MoCYP76Y2  AATTGGCAAACTTCCTCATCGCTCACTTGCAAGATTGGCCAAAATCCACGGCCCTCTCATCACTCTCCGC 210
GwCYP76Y2  AATTGGCAAACTTCCTCATCGCTCACTTGCAAGATTGGCCAAAATCCACGGCCCTCTCATCACTCTCCGC 210

PnCYP76Y2  CTTGGCTCCATCACCACCGTCGTCGCCTCCTCTCCCCAGACAGCCAAACTAATTCTCCAAACACATGGCC 280
MoCYP76Y2  CTTGGCTCCATCACCACCGTCGTCGCCTCCTCTCCCCAGACAGCCAAACTAATTCTCCAAACACATGGCC 280
GwCYP76Y2  CTTGGCTCCATCACCACCGTCGTCGCCTCCTCTCCCCAGACAGCCAAACTAATTCTCCAAACACATGGCC 280

PnCYP76Y2  AGAACTTCCTGGACCGTCCTGCTCCCGAGGCCCTCGACAGTCCTCAAGGCACAATCGGATGGATTCCTGC 350
MoCYP76Y2  AGAACTTCCTGGACCGTCCTGCTCCCGAGGCCCTCGACAGTCCTCAAGGCACAATCGGATGGATTCCTGC 350
GwCYP76Y2  AGAACTTCCTGGACCGTCCTGCTCCCGAGGCCCTCGACAGTCCTCAAGGCACAATCGGATGGATTCCTGC 350

PnCYP76Y2  GGACCATGTATGGCGCAGCCGCCGCCGTGTTTGCATCAACCACTTGTTCACATCCCAGAGCCTGGACTCA 420
MoCYP76Y2  GGACCATGTATGGCGCAGCCGCCGCCGTGTTTGCATCAACCACTTGTTCACATCCCAGAGCCTGGACTCA 420
GwCYP76Y2  GGACCATGTATGGCGCAGCCGCCGCCGTGTTTGCATCAACCACTTGTTCACATCCCAGAGCCTGGACTCA 420

PnCYP76Y2  CTCCAACACCTTCGATACAAAAAGGTGGAACAACTTCTCCAACATATCCGTAAGCACTGTGTTTCCGGTA 490
MoCYP76Y2  CTCCAACACCTTCGATACAAAAAGGTGGAACAACTTCTCCAACATATCCGTAAGCACTGTGTTTCCGGTA 490
GwCYP76Y2  CTCCAACACCTTCGATACAAAAAGGTGGAACAACTTCTCCAACATATCCGTAAGCACTGTGTTTCCGGTA 490

PnCYP76Y2  CACCGGTGGATATCGGCCTACTCACCTCTGCCATCAACTTGAACGTGCTTTCAAACGCCATTTTCTCTGT 560
MoCYP76Y2  CACCGGTGGATATCGGCCTACTCACATCTGCCATCAACTTGAACGTGCTTTCAAACGCCATTTTCTCTGT 560
GwCYP76Y2  CACCGGTGGATATCGGCCTACTCACATCTGCCATCAACTTGAACGTGCTTTCAAACGCCATTTTCTCTGT 560

PnCYP76Y2  TGACCTTGTTGATCCAGGATTTGAGTCGGCTCAGGATTTCAGGGATCAGGTGTGGGGAATCATGGAGGGT 630
MoCYP76Y2  TGACCTTGTTGATCCAGGATTTGAGTCGGCTCAGGATTTCAGGGATCAGGTGTGGGGAATCATGGAGGGT 630
GwCYP76Y2  TGACCTTGTTGATCCAGGATTTGAGTCGGCTCAGGATTTCAGGGATCAGGTGTGGGGAATCATGGAGGGT 630

PnCYP76Y2  GCTGGCAAGTTTAATATTTCAGATTATTTTCCCATGTTTCGAAGGTTCGATTTGCTAGGTGTGAAGCGCG 700
MoCYP76Y2  GCTGGCAAGTTTAATATTTCAGATTATTTTCCCATGTTTCGAAGGTTCGATTTGCTAGGTGTGAAGCGCG 700
GwCYP76Y2  GCTGGCAAGTTTAATATTTCAGATTATTTTCCCATGTTTCGAAGGTTCGATTTGCTAGGTGTGAAGCGCG 700

PnCYP76Y2  ACACTTTCTCATGTTATAAACGGCTTTATGAAATAGTTGGTGGTATAATCAAAAGCCGTATCAAGTGTAG 770
MoCYP76Y2  ACACTTTCTCATGTTATAAACGGCTTTATGAAATAGTTGGTGGTATAATCAAAAGCCGTATCAAGTGTAG 770
GwCYP76Y2  ACACTTTCTCATGTTATAAACGGCTTTATGAAATAGTTGGTGGTATAATCAAAAGCCGTATCAAGTGTAG 770

PnCYP76Y2  AGCCTCCAATCCGATGAGCAGGAATGACGACTTCTTGGATGTGATTCTTGATCAGTGCCAAGAAGATGGC 840
MoCYP76Y2  AGCCTCCAATCCGATGAGCAGGAATGACGACTTCTTGGATGTGATTCTTGATCAGTGCCAAGAAGATGGC 840
GwCYP76Y2  AGCCTCCAATCCGATGAGCAGGAATGACGACTTCTTGGATGTGATTCTTGATCAGTGCCAAGAAGATGGC 840

PnCYP76Y2  TCTGTTTTCAATTCTGATAATATCCAGGTTTTGATTGTGGAACTATTTTATGCTGGGAGTGATACATCTA 910
MoCYP76Y2  TCTGTTTTCAATTCTGATAATATCCAGGTTTTGATTGTGGAACTATTTTATGCTGGGAGTGATACATCTA 910
GwCYP76Y2  TCTGTTTTCAATTCTGATAATATCCAGGTTTTGATTGTGGAACTATTTTATGCTGGGAGTGATACATCTA 910

PnCYP76Y2  CCATAACAACTGAATGGGCAATGACTGAACTTCTTCGAAATCCTCGGCTGATGCAAAAGGTTCGGCAAGA 980
MoCYP76Y2  CCATAACAACTGAATGGGCAATGACTGAACTTCTTCGAAATCCTCGGCTGATGCAAAAGGTTCGGCAAGA 980
GwCYP76Y2  CCATAACAACTGAATGGGCAATGACTGAACTTCTTCGAAATCCTCGGCTGATGCAAAAGGTTCGGCAAGA 980

PnCYP76Y2  ACTTAGCGAAGTAATCGGGGCAGGTCAAATGGTTAGAGAATCAGATATGGACCGACTTCCATATTTTCAA 1050
MoCYP76Y2  ACTTAGCGAAGTAATCGGGGCAGGTCAAATGGTTAGAGAATCAGATATGGACCGACTTCCATATTTTCAA 1050
GwCYP76Y2  ACTTAGCGAAGTAATCGGGGCAGGTCAAATGGTTAGAGAATCAGATATGGACCGACTTCCATATTTTCAA 1050

PnCYP76Y2  GCTGTTGTGAAAGAGACGCTTAGACTCCATCCGGCTGGGCCCCTTCTGTTACCTTTTAAAGCAAAGAATG 1120
MoCYP76Y2  GCTGTTGTGAAAGAGACGCTTAGACTCCATCCGGCTGGGCCCCTTCTGTTACCTTTTAAAGCAAAGAATG 1120
GwCYP76Y2  GCTGTTGTGAAAGAGACGCTTAGACTCCATCCGGCTGGGCCCCTTCTGTTACCTTTTAAAGCAAAGAATG 1120
PnCYP76Y2  ATGTGGAATTATGCGGTTTCACCATACCCAGTAACAGTCATGTCCTTGTCAATATGTGGGCTATTGCAAG 1190
MoCYP76Y2  ATGTGGAATTATGCGGTTTCACCATACCCAGTAACAGTCATGTCCTTGTCAATATGTGGGCTATTGCAAG 1190
GwCYP76Y2  ATGTGGAATTATGCGGTTTCACCATACCCAGTAACAGTCATGTCCTTGTCAATATGTGGGCTATTGCAAG 1190

PnCYP76Y2  AGATCCAGGTTATTGGGAGGATCCTTCATCCTTCCTTCCCGAAAGATTCTTGGGCTCTAAGATAGATTAT 1260
MoCYP76Y2  AGATCCAGGTTATTGGGAGGATCCTTCATCCTTCCTTCCCGAAAGATTCTTGGGCTCTAAGATAGATTAT 1260
GwCYP76Y2  AGATCCAGGTTATTGGGAGGATCCTTCATCCTTCCTTCCCGAAAGATTCTTGGGCTCTAAGATAGATTAT 1260

PnCYP76Y2  AGAGGCCAAGATTACGAGTATATACCATTTGGAGCAGGTAGGCGAATCTGCCCAGGCATACCTCTTGCCA 1330
MoCYP76Y2  AGAGGCCAAGATTACGAGTATATACCATTTGGAGCAGGTAGGCGAATCTGCCCAGGCATACCTCTTGCCA 1330
GwCYP76Y2  AGAGGCCAAGATTACGAGTATATACCATTTGGAGCAGGTAGGCGAATCTGCCCAGGCATACCTCTTGCCA 1330
```

```
PnCYP76Y2  TCAGAATGGTTCAACTAGTGTTAGCTTCCATTATCCACTCCTTTAACTGGAAGCTTCCTGAAGGAACTAC 1400
MoCYP76Y2  TCAGAATGGTTCAACTAGTGTTAGCTTCCATTATCCACTCCTTTAACTGGAAGCTTCCTGAAGGAACTAC 1400
GwCYP76Y2  TCAGAATGGTTCAACTAGTGTTAGCTTCCATTATCCACTCCTTTAACTGGAAGCTTCCTGAAGGAACTAC 1400

PnCYP76Y2  TCCACTGACCATTGACATGCAAGAACAATGCGGAGCTACCTTGAAGAAGGCCATTCCTCTTTCTGCAATT 1470
MoCYP76Y2  TCCACTGACCATTGACATGCAAGAACAATGCGGAGCTACCTTGAAGAAGGCCATTCCTCTTTCTGCAATT 1470
GwCYP76Y2  TCCACTGACCATTGACATGCAAGAACAATGCGGAGCTACCTTGAAGAAGGCCATTCCTCTTTCTGCAATT 1470

PnCYP76Y2  CCATTTATCGAAGAAAATTAA 1491
MoCYP76Y2  CCATTTATCGAAGAAAATTAA 1491
GwCYP76Y2  CCATTTATCGAAGAAAATTAA 1491
```

```
PnCYP76Y2  MELNTFLLLCMPLILCFFLLQFLRPSSHATKLPPGPTGLPILGSLLEIGKLPHRSLARLAKIHGPLITLR 70
MoCYP76Y2  MELNTFLLLCMPLILCFFLLQFLRPSSHATKLPPGPTGLPILGSLLEIGKLPHRSLARLAKIHGPLITLR 70
GwCYP76Y2  MELNTFLLLCMPLILCFFLLQFLRPSSHATKLPPGPTGLPILGSLLEIGKLPHRSLARLAKIHGPLITLR 70

PnCYP76Y2  LGSITTVVASSPQTAKLILQTHGQNFLDRPAPEALDSPQGTIGWIPADHVWRSRRRVCINHLFTSQSLDS 140
MoCYP76Y2  LGSITTVVASSPQTAKLILQTHGQNFLDRPAPEALDSPQGTIGWIPADHVWRSRRRVCINHLFTSQSLDS 140
GwCYP76Y2  LGSITTVVASSPQTAKLILQTHGQNFLDRPAPEALDSPQGTIGWIPADHVWRSRRRVCINHLFTSQSLDS 140

PnCYP76Y2  LQHLRYKKVEQLLQHIRKHCVSGTPVDIGLLTSAINLNVLSNAIFSVDLVDPGFESAQDFRDQVWGIMEG 210
MoCYP76Y2  LQHLRYKKVEQLLQHIRKHCVSGTPVDIGLLTSAINLNVLSNAIFSVDLVDPGFESAQDFRDQVWGIMEG 210
GwCYP76Y2  LQHLRYKKVEQLLQHIRKHCVSGTPVDIGLLTSAINLNVLSNAIFSVDLVDPGFESAQDFRDQVWGIMEG 210

PnCYP76Y2  AGKFNISDYFPMFRRFDLLGVKRDTFSCYKRLYEIVGGIIKSRIKCRASNPMSRNDDFLDVILDQCQEDG 280
MoCYP76Y2  AGKFNISDYFPMFRRFDLLGVKRDTFSCYKRLYEIVGGIIKSRIKCRASNPMSRNDDFLDVILDQCQEDG 280
GwCYP76Y2  AGKFNISDYFPMFRRFDLLGVKRDTFSCYKRLYEIVGGIIKSRIKCRASNPMSRNDDFLDVILDQCQEDG 280

PnCYP76Y2  SVFNSDNIQVLIVELFYAGSDTSTITTEWAMTELLRNPRLMQKVRQELSEVIGAGQMVRESDMDRLPYFQ 350
MoCYP76Y2  SVFNSDNIQVLIVELFYAGSDTSTITTEWAMTELLRNPRLMQKVRQELSEVIGAGQMVRESDMDRLPYFQ 350
GwCYP76Y2  SVFNSDNIQVLIVELFYAGSDTSTITTEWAMTELLRNPRLMQKVRQELSEVIGAGQMVRESDMDRLPYFQ 350

PnCYP76Y2  AVVKETLRLHPAGPLLLPFKAKNDVELCGFTIPSNSHVLVNMWAIARDPGYWEDPSSFLPERFLGSKIDY 420
MoCYP76Y2  AVVKETLRLHPAGPLLLPFKAKNDVELCGFTIPSNSHVLVNMWAIARDPGYWEDPSSFLPERFLGSKIDY 420
GwCYP76Y2  AVVKETLRLHPAGPLLLPFKAKNDVELCGFTIPSNSHVLVNMWAIARDPGYWEDPSSFLPERFLGSKIDY 420

PnCYP76Y2  RGQDYEYIPFGAGRRICPGIPLAIRMVQLVLASIIHSFNWKLPEGTTPLTIDMQEQCGATLKKAIPLSAI 490
MoCYP76Y2  RGQDYEYIPFGAGRRICPGIPLAIRMVQLVLASIIHSFNWKLPEGTTPLTIDMQEQCGATLKKAIPLSAI 490
GwCYP76Y2  RGQDYEYIPFGAGRRICPGIPLAIRMVQLVLASIIHSFNWKLPEGTTPLTIDMQEQCGATLKKAIPLSAI 490

PnCYP76Y2  PFIEEN* 497
MoCYP76Y2  PFIEEN* 497
GwCYP76Y2  PFIEEN* 497
```

Le vin tient une grande place tant culturelle et qu'économique en Europe. Alors que la qualité d'un vin dépend directement de ses arômes, la composition et la formation de ces derniers est encore mal connue. Une méta-analyse des données bibliographiques nous a permis de montrer que les arômes du raisin et du vin sont riches en monoterpènes, en particulier en dérivés du linalool. Les récentes avancées dans la compréhension du métabolisme oxydatif du linalool et du géraniol ont ensuite été résumées dans une revue bibliographique. L'annotation détaillée de la superfamille des cytochromes P450 dans le génome de la vigne nous a permis d'étudier sa structure génétique, sa phylogénie et son expression, mais aussi d'identifier des gènes dont l'expression est activée dans le grain à maturité, lors de la synthèse de nombreux composés aromatiques. La lactone du vin est la molécule dont le seuil de détection olfactive est le plus bas, ce qui en fait un composant essentiel de l'arôme du vin. Nous avons pu démontrer que cette lactone se forme au cours du vieillissement du vin par une réaction lente et non-enzymatique à partir du 8-carboxylinalool. L'accumulation de ce dernier dans la baie est concomitante à l'expression de plusieurs P450s, dont *CYP76F14* est le plus fortement exprimé. L'activité des candidats a été évaluée *in vitro* et *in planta*. Trois enzymes catalysent des étapes d'oxydation conduisant du linalool au (*E*)-8-carboxylinalool, mais seul CYP76F14 catalyse efficacement la formation de l'acide. Tant par son activité catalytique que son profil d'expression, CYP76F14 apparaît donc comme le responsable le plus probable de la formation du précurseur de la lactone du vin. Ce travail établit une base solide pour l'étude des autres P450s impliqués dans la formation des arômes du vin ainsi que dans d'autres processus biologiques tels que le développement du fruit ou les interactions plante-pathogène chez la vigne.

Mots-clés : **arômes, monoterpenol, Vitis vinifera, cytochrome P450, annotation**.

Wine has an important cultural, as well as economical value in Europe. Our appreciation of wine quality is profoundly shaped by its aroma, but aroma composition and formation are still poorly understood. With a meta-analysis of grape and wine aroma, we show that both are rich in monoterpenes, in particular oxygenated linalool derivatives. We then review the recent advances in the oxidative metabolism of the two major plant monoterpenols, linalool and geraniol. A thorough annotation of the P450 superfamily in grapevine, revealed its genomic organization, phylogeny and expression. Specifically, we identified genes showing an activated expression in the ripe grape berry, the stage during which the biosynthesis of many aroma compounds takes place. Among the known oxygenated monoterpenols in grapevine, wine lactone has the lowest odor detection threshold and therefore the largest potential impact on wine aroma. We demonstrated that wine lactone is formed during wine ageing via a slow non-enzymatic reaction from the precursor (E)-8-carboxylinalool. We showed that the accumulation of this precursor in grape berries parallels the expression of several cytochrome P450 genes, among which CYP76F14 has the highest expression. We functionally characterized of the candidate enzymes in vitro and in the leaves of *Nicotiana benthamiana*. While three of them catalyzed some of the oxidative steps from linalool to (E)-8-carboxylinalool, only CYP76F14 efficiently catalyzed the whole pathway. Taken together, CYP76F14 catalytic activity and expression pattern indicate that it is a prime candidate for the formation of the wine lactone precursor in grape berries. This work lays ground for the elucidation of novel roles of P450s in the biosynthesis of aroma in grapevine, as well as other physiological processes, such as fruit development or plant-pathogen interactions.

Keywords: **aroma, monoterpenol, Vitis vinifera, cytochrome P450, gene annotation.**