



**HAL**  
open science

# Apport de la génomique à l'exploration de la diversité génétique des vanilliers cultivés et à l'identification des régions impliquées dans la résistance à la fusariose

Félicien Favre

► **To cite this version:**

Félicien Favre. Apport de la génomique à l'exploration de la diversité génétique des vanilliers cultivés et à l'identification des régions impliquées dans la résistance à la fusariose. Sciences agricoles. Université de la Réunion, 2022. Français. NNT : 2022LARE0039 . tel-04049920

**HAL Id: tel-04049920**

**<https://theses.hal.science/tel-04049920v1>**

Submitted on 29 Mar 2023

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

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

# APPORT DE LA GÉNOMIQUE À L'EXPLORATION DE LA DIVERSITÉ GÉNÉTIQUE DES VANILLIERS CULTIVÉS ET À L'IDENTIFICATION DES RÉGIONS IMPLIQUÉES DANS LA RÉSISTANCE À LA FUSARIOSE

**Thèse de doctorat de l'Université de la Réunion**

Ecole doctorale 542, Sciences, Technologie et Santé  
Spécialité du doctorat : Génétique

Thèse préparée dans l'Unité Mixte de Recherche Peuplements végétaux et bioagresseurs  
en milieu tropical, sous la direction de Pascale BESSE, professeure des universités et le co-  
encadrement de Carine CHARRON, docteure

**Thèse soutenue à Saint-Pierre, le 8 décembre 2022, par**

**Félicien FAVRE**

## Composition du Jury

<b>Isabelle FOCK-BASTIDE</b> Professeure des universités, HDR, Université de la Réunion	Présidente
<b>Véronique DECROOCQ</b> Directrice de recherche, HDR, INRAE	Rapporteur
<b>Lorenzo BARCHI</b> Professeur des universités, Université de Turin	Rapporteur
<b>Cyrille SAINTENAC</b> Chargé de recherche, INRAE	Examineur
<b>Pascale BESSE</b> Professeure des universités, HDR, Université de la Réunion	Directrice de thèse
<b>Carine CHARRON</b> Docteure, CIRAD	Co-encadrante de thèse



***« d'un naturel curieux, intelligent, il aime observer la nature et fait preuve de beaucoup d'attention »***

Ferréol Bellier Beaumont à propos de Edmond Albius

*À Théa  
À Mila-Rose*





## REMERCIEMENTS

---

Tout d'abord, je remercie Véronique Decroocq et Lorenzo Barchi d'avoir accepté d'évaluer ce travail de thèse. Je remercie également Isabelle Fock-Bastide et Cyrille Saintenac d'avoir accepté de faire partie du jury de thèse. Je tiens à remercier les membres de mon comité de suivi de thèse Valérie Geffroy, Nathalie Chantret et Yann Pécrix pour leurs conseils avisés et leurs encouragements.

J'ai eu la chance de réaliser ce travail de thèse au Pôle de protection des plantes au sein l'UMR PVBMT, et je remercie le directeur d'unité, Bernard Reynaud, pour son accueil.

Je remercie tout particulièrement Pascale Besse et Carine Charron pour la confiance qu'elles m'ont témoignée en me permettant de poursuivre ce projet de thèse. Un grand merci à Carine pour toutes ces années de travail ensemble, sa disponibilité et pour ses belles qualités humaines.

Je tiens à remercier Cyril Jourda et tous les membres de l'équipe Diversité et utilisation durable des ressources génétiques végétales en milieu tropical de l'UMR PVBMT, et en particulier l'équipe Vanille, Michel Grisoni, Katia Jade, Jean-Bernard Dijoux, Déborah Turpin, Cathucia Andriamihaja et Quentin Piet pour leur accompagnement sur ce projet et le partage de leurs connaissances précieuses sur le vanillier. Je remercie également les stagiaires qui m'ont apporté leur aide pendant ces travaux de thèse, Taj, Julie, Sarah, Manon et Quentin Da Silva à qui je souhaite beaucoup de réussite pour la suite de son travail de thèse dans l'équipe.

Une pensée à toutes les personnes que j'ai croisées pendant ces 3 années, au laboratoire et pendant les pauses. Je vous remercie pour votre gentillesse et vos conseils. Votre aide aura été déterminante dans ce travail. Je tiens à remercier le grand statisticien Fred pour son aide précieuse. Je remercie Magalie, Océane et Chantal pour le soutien administratif de la thèse, et surtout Virginie pour toute l'énergie qu'elle met

dans le soutien aux doctorants.

Un grand merci à tous mes collègues thésards, et notamment mes collègues de bureau Pauline, Cathucia et Quentin, pour les moments de partage et leur soutien moral.

Je dédie cette thèse à mes parents, qui m'ont toujours soutenu et qui m'ont transmis le goût de la nature, de l'aventure et de la découverte sans lequel tout ceci n'aurait pas été possible.

Je remercie mes amis qui me soutiennent depuis tant d'années et qui sont toujours prêts à parcourir les océans pour m'apporter leur bonne humeur légendaire, Florine, Elsa, Camille, Jean-Baptiste, Ana, Aurélien, Léa, Coralie, Pierre-Antoine...

Merci Damien pour la relecture du manuscrit et ton soutien indéfectible pendant ces 3 années.

Un immense merci à toutes les personnes qui ont fait de cette thèse une formidable aventure sur l'île intense, Corentin E, Justine, Marine, François, Joseph, Mickaël, Margot, Emma, Mathilde, Stan, Corentin G, Maria, Alicia, Selim, Lysa, Thomas, Nathan...

Nartrou' sa !

# TABLE DES MATIÈRES

---

<b>LISTE DES FIGURES ET TABLEAUX.....</b>	<b>1</b>
<b>LISTE DES ABRÉVIATIONS ET DES ACRONYMES.....</b>	<b>3</b>
<b>INTRODUCTION GÉNÉRALE   L'ÉTUDE DES PLANTES À L'ÈRE DE LA GÉNOMIQUE .....</b>	<b>5</b>
<b>CHAPITRE 1   SYNTHÈSE BIBLIOGRAPHIQUE.....</b>	<b>9</b>
1. Origine et diversité des vanilliers .....	9
1.1. Le genre <i>Vanilla</i> .....	9
1.2. Les vanilliers cultivés .....	11
1.3. Le génome de <i>V. planifolia</i> .....	12
2. Caractéristiques agronomiques des vanilliers cultivés .....	14
2.1. Culture du vanillier .....	14
2.2. Les maladies majeures des vanilliers .....	14
3. <i>Fusarium oxysporum</i> f.sp. <i>radicis-vanillae</i> .....	17
3.1. <i>Fusarium oxysporum</i> .....	17
3.2. Diversité génétique de Forv .....	17
3.3. Infection et pathogénicité.....	18
4. L'interaction <i>Vanilla</i> -Forv .....	19
4.1. Mécanismes de défense des plantes contre les agents pathogènes .....	19
4.2. Aspect génétique de la résistance des vanilliers à Forv .....	20
4.3. Aspects morphologiques et histologiques de l'interaction vanillier-Forv.....	21
4.4. Aspects moléculaires de l'interaction vanillier-Forv .....	23
5. Problématique et objectifs de la thèse .....	25
<b>CHAPITRE 2   LE GÉNOTYPAGE À HAUT DÉBIT EN APPUI AUX RECHERCHES SUR LES PLANTES TROPICALES COMPLEXES.....</b>	<b>27</b>
1. Préambule .....	27
2. Publication 1 : Genotyping-by-Sequencing Technology in Plant Taxonomy and Phylogeny.....	29
3. Conclusion .....	43
<b>CHAPITRE 3   RÉOLUTION DE LA DIVERSITÉ GÉNÉTIQUE INTRA-SPÉCIFIQUE AU SEIN DES VANILLIERS CULTIVÉS .....</b>	<b>45</b>
1. Préambule .....	45

2. Publication 2 : A genome-wide assessment of the genetic diversity, evolution and relationships with allied species of the clonally propagated crop <i>Vanilla planifolia</i> Jacks. ex Andrews .....	47
3. Conclusion .....	65
<b>CHAPITRE 4   CARTOGRAPHIE GÉNÉTIQUE DE LA RÉSISTANCE A LA FUSARIOSE DES VANILLIERS</b> .....	67
1. Préambule .....	67
2. Publication 3 : First <i>Vanilla planifolia</i> high-density genetic linkage map provides quantitative trait loci for resistance to <i>Fusarium oxysporum</i> .....	69
3. Conclusion .....	99
<b>CHAPITRE 5   EXPLORATION DU GÉNOME DE <i>V. PLANIFOLIA</i> : CRIBLAGE DE GÈNES CANDIDATS POTENTIELLEMENT IMPLIQUÉS DANS LA RÉSISTANCE À LA FUSARIOSE</b> .....	101
1. Introduction .....	101
2. Matériel et méthodes .....	103
2.1. Annotation des gènes <i>LRR</i> chez <i>V. planifolia</i> .....	103
2.2. Identification de gènes candidats potentiellement impliqués dans la résistance à <i>F. oxysporum</i> .....	104
3. Résultats .....	105
3.1. Identification d'un faible nombre de gènes <i>LRR</i> chez <i>V. planifolia</i> .....	105
3.2. Gènes candidats potentiellement impliqués dans la résistance à <i>F. oxysporum</i> ..	109
4. Discussion et perspectives .....	111
<b>CHAPITRE 6   DISCUSSION GÉNÉRALE</b> .....	115
1. Conclusions .....	115
2. Perspectives .....	119
2.1. Expression et diversité des gènes de résistance par des approches omiques.....	119
2.2. Spectre d'action et durabilité de la résistance des vanilliers.....	122
2.3. Vers une culture de vanilliers résiliente dans un contexte de changements environnementaux.....	123
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b> .....	127
<b>DONNÉES SUPPLÉMENTAIRES PUBLIÉES</b> .....	137
1. Chapitre 3 .....	137
2. Chapitre 4 .....	150
<b>ANNEXES</b> .....	171

# LISTE DES FIGURES ET TABLEAUX

---

## INTRODUCTION GENERALE

Figure 1	Contributions de la génomique sur la génétique et la sélection des plantes non-modèles	6
----------	--	---

## CHAPITRE 1 |

Figure 1.1	Phylogénie moléculaire et temps de divergence du genre <i>Vanilla</i>	10
Figure 1.2	Cytogénétique et génomique de <i>V. planifolia</i>	13
Figure 1.3	Symptômes de la pourriture des racines et des tiges sur <i>V. planifolia</i> et <i>V. x tahitensis</i>	16
Figure 1.4	Cycle d'infection de <i>Fusarium oxysporum</i> f.sp. <i>radicis-vanillae</i>	18
Figure 1.5	Schéma illustrant le système immunitaire des plantes	20
Figure 1.6	Différences dans la structure et la composition de la paroi cellulaire de racines <i>in vitro</i> non inoculées	22
Figure 1.7	Voie de biosynthèse des lignines	23

## CHAPITRE 2 |

Figure 1	Diagram of the genomic library construction method based on reducing genome complexity	37
----------	--	----

## CHAPITRE 3 |

Figure 1	Genotyping-By-Sequencing SNP distribution and density on <i>V. planifolia</i> cv Daphna chromosomes	53
Figure 2	Phylogenetic structuration between cultivated vanillas and wild relative species	54
Figure 3	Origin of <i>V. x tahitensis</i> shown by bayesian clustering and comparison of morphological traits	56
Figure 4	Genetic structure of 84 <i>V. planifolia</i> using 1,129 informative SNPs	57
Table 1	Diversity indexes in <i>Vanilla</i> species using 2,004 SNPs identified across all species	55
Table S1	List of 88 <i>Vanilla planifolia</i> accessions used in this study	139
Table S2	List of 7 <i>Vanilla x tahitensis</i> accessions used in this study	142
Table S3	List of 23 <i>Vanilla</i> wild relatives accessions used in this study	143
Table S4	List of 19 interspecific hybrid accessions used in this study	144
Figure S1	GBS loci distribution and density on <i>V. planifolia</i> cv Daphna chromosomes	145
Figure S2	Principal coordinates analysis (PCoA) based on 2,004 SNPs	146
Figure S3	Principal coordinates analysis (PCoA) based on 1,129 <i>Vanilla planifolia</i> SNPs	147
Figure S4	Genetic structure of 84 <i>V. planifolia</i> using 1,129 informative SNPs	148

Figure S5	Morphology of a. <i>V. odorata</i> cf. CR3612 flower and fruits, b. <i>V. insignis</i> cf. CR0087 flower and leaves	149
-----------	---	-----

**CHAPITRE 4 |**

Figure 1	Genetic linkage map of the population obtained from the self-fertilization of <i>V. planifolia</i> CR0040	83
Figure 2	Mapping of the 16 linkage groups (VpLG)	84
Figure 3	Distribution of Fo072 resistance of the progeny of CR0040 accession	86
Figure 4	Candidate genes for significant quantitative trait loci related to Fo072 resistance	88
Table 1	Number of SNP markers identified and aligned in haplotypes A and B of the CR0040 genome assembly	81
Table 2	Distribution of SNP markers among 16 linkage groups (LG) of the mapping population	82
Table 3	Quantitative trait loci (QTL) associated with resistance to Fo072	87
Table S1	DNA samples used for genotyping by sequencing (GBS)	151
Table S2	Accessions tested for the resistance to <i>Fusarium oxysporum</i> f.sp. <i>radicis-vanillae</i> Fo072	154
Table S3	Candidate genes for significant quantitative trait loci for Fo072 resistance in CR0040 genome	157
Figure S1	Genetic linkage map of the population, with 1,804 markers across the linkage groups	164
Figure S2	Phenotyping of CR0040 progenies	166
Figure S3	Quantitative trait loci for AUDPC (black), d.5 (pink), d.10 (purple), d.15 (green) and d.myc (blue)	169

**CHAPITRE 5 |**

Figure 5.1	Schéma de la structure protéique des 3 sous-familles de récepteurs LRR : LRR-RLK, LRR-RLP et NLR	101
Figure 5.2	Localisation des gènes codant pour des gènes LRR sur les 14 chromosomes assemblés	106
Figure 5.3	Densité des gènes codant pour des LRRs identifiés par LRRprofiler	107
Figure 5.4	Localisation et expression (FPKM) par organe des gènes NLR sur le génome de <i>V. planifolia</i> CR0040	108
Figure 5.5	Localisation des gènes de la voie de biosynthèse des phénylpropanoïdes	110
Figure 5.6	Localisation et expression (FPKM) des gènes impliqués dans la voie de biosynthèse des phénylpropanoïdes	112
Tableau 5.1	Nombre de gènes codant pour des LRRs identifiés par LRRprofiler	105

**CHAPITRE 6 |**

Figure 6.1	Bilan des régions d'intérêt identifiées pour la résistance à la fusariose au cours de ma thèse, par les approches combinées de génétique et de génomique	118
------------	--	-----

## LISTE DES ABRÉVIATIONS ET DES ACRONYMES

---

<b>4CL</b>	4-coumarate:CoA ligase
<b>4HBS</b>	4-hydroxybenzaldéhyde synthase
<b>ADN</b>	Acide désoxyribonucléique
<b>AFLP</b>	Amplification fragment length polymorphism
<b>ARN</b>	Acide ribonucléique
<b>BLAST</b>	Basic local alignment search tool
<b>C3'H</b>	4-coumaroyl shikimate/quinate 3'-hydroxylase
<b>C3H</b>	4-coumarate 3-hydroxylase
<b>C4H</b>	Cinnamate 4-hydroxylase
<b>CAD</b>	Cinnamyl alcohol déshydrogenase
<b>CC</b>	Coiled coil domain
<b>CCoAOMT</b>	Caffeoyl CoA 3-O-méthyltransférase
<b>CCR</b>	Cinnamoyl CoA réductase
<b>CIRAD</b>	Centre de coopération internationale en recherche agronomique pour le développement
<b>CMV</b>	<i>Cucumber mosaic virus</i>
<b>COMT</b>	Caffeate/5-hydroxyfêrulate3-Ométhyltransférase
<b>CR</b>	Collection réunion
<b>CRB</b>	Centre de ressources biologiques
<b>CSE</b>	Caffoyl shikimate estérase
<b>CV</b>	Cultivar
<b>CymMV</b>	<i>Cymbidium mosaic virus</i>
<b>DEG</b>	Differentialy expressed gene
<b>EF1<math>\alpha</math></b>	Elongation factor 1 alpha
<b>ETI</b>	Effector-triggered immunity
<b>F5H</b>	Fêrulate 5-hydroxylase/conifêraldéhyde 5-hydroxylase
<b>Forv</b>	<i>Fusarium oxysporum</i> f. sp. <i>radicis-vanillae</i>
<b>FPKM</b>	Fragments per kilobase million
<b>GBS</b>	Génotypage par séquençage
<b>GWAS</b>	Genome wide association studies
<b>H<sup>2</sup></b>	Broad-sense heritability
<b>HCT</b>	4- hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyltransférase



<b>Ho</b>	Hétérozygotie observée
<b>IGS</b>	Ribosomal intergenic spacer
<b>ITS</b>	Internal transcribed spacer
<b>jpi</b>	Jour post-inoculation
<b>LG</b>	Groupe de liaison
<b>LRR</b>	Leucin-rich repeat
<b>LRR-RLK</b>	Leucin-rich repeat receptor-like kinase
<b>LRR-RLP</b>	Leucin-rich repeat receptor-like protein
<b>MAMP</b>	Microbe-associated molecular pattern
<b>NB-ARC</b>	Nuceotide binding-APAF1 (apoptotic protease-activating factor-1), R proteins and CED-4 ( <i>Caenorhabditis elegans</i> death-4 protein)
<b>NBS-LRR (NLR)</b>	Nucleotide binding site leucin-rich repeat
<b>NGS</b>	Next-generation sequencing
<b>OMT</b>	O-methyl transferase
<b>PAL</b>	Phénylalanine ammonia lyase
<b>PAMP</b>	Pathogens-associated molecular pattern
<b>PCoA</b>	Principal coordinate analysis
<b>PPR</b>	Pentatricopeptide-repeat protein
<b>PRR</b>	Pattern recognition receptor
<b>PTAL</b>	Phénylalanine/tyrosine ammonia-lyase
<b>PTI</b>	PAMP-triggered immunity
<b>PVBMT</b>	Peuplements végétaux et bioagresseurs en milieu tropical
<b>QTL</b>	Quantitative trait locus
<b>RAPD</b>	Random amplified polymorphic DNA
<b>RE</b>	Restriction enzyme
<b>RFLP</b>	Restriction fragment length polymorphism
<b>RGA</b>	Resistance gene analog
<b>ROS</b>	Reactive oxygen species
<b>RPW8</b>	Resistance to powdery mildew 8
<b>SNP</b>	Single nucleotide polymorphism
<b>SOOI</b>	Sud-Ouest de l'océan Indien
<b>Spp.</b>	Species
<b>SSR</b>	Simple sequence repeat
<b>TIR</b>	Toll-interleukin receptor
<b>TM</b>	Transmembrane domain
<b>UMR</b>	Unité mixte de recherche

# INTRODUCTION GÉNÉRALE |

## L'ÉTUDE DES PLANTES À L'ÈRE DE LA GÉNOMIQUE

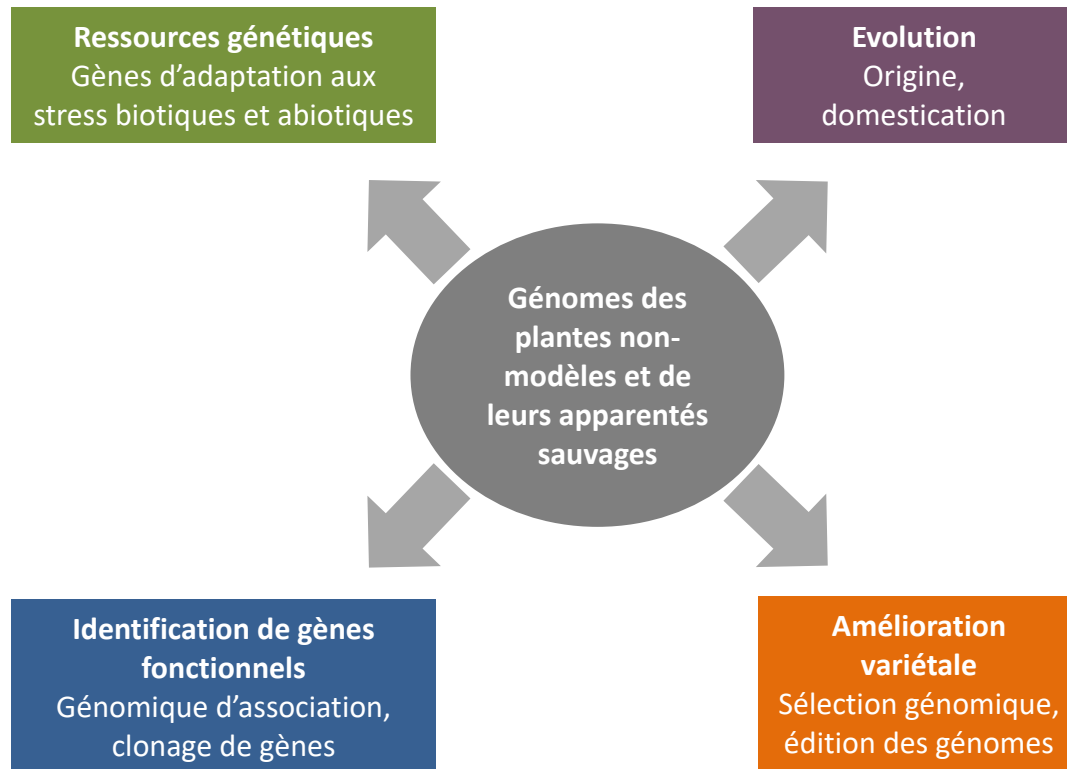
---

Depuis la publication du génome d'*Arabidopsis thaliana* en 2000 (The Arabidopsis Genome Initiative 2000), plus de 400 espèces de plantes à fleur ont été séquencées ([https://www.plabipd.de/plant\\_genomes\\_pa.ep](https://www.plabipd.de/plant_genomes_pa.ep)). Cette avancée spectaculaire dans le séquençage des génomes complets a notamment permis de développer les techniques d'amélioration variétale, telles que les études d'associations génomiques, la sélection génomique et l'édition de gènes. Grâce à la baisse des coûts de séquençage, les plantes non modèles, ou dites « orphelines », sont à leur tour entrées dans cette nouvelle ère de la génomique, permettant d'accélérer les processus de sélection par l'identification de gènes d'intérêt agronomique et d'augmenter nos connaissances sur la diversité et l'évolution de ces espèces.

Des initiatives de séquençage *de novo* ciblant les plantes non modèles ont été entreprises ces dix dernières années. Le séquençage du génome de 101 plantes alimentaires africaines promeut notamment l'amélioration génétique de cultures sous-étudiées, en s'appuyant sur le génome, pour proposer des solutions alimentaires nutritives basées sur la biodiversité en Afrique (Hendre et al. 2019; Jamnadass et al. 2020). Le re-séquencage de génomes entiers, possible quand un génome de référence a déjà été produit, permet également d'obtenir une grande quantité d'informations sur plusieurs milliers d'individus représentatifs de la diversité des espèces étudiées. De nombreuses études de re-séquencage menées sur les céréales et les légumineuses non modèles ont permis de caractériser les processus de domestication mis en jeu et les régions génomiques sous sélection qui sont impliquées, et d'identifier de nouvelles ressources génétiques adaptées aux contraintes biotiques et abiotiques (Ye and Fan 2021).

Les études génomiques des espèces non modèles permettent donc la caractérisation de nouveaux gènes fonctionnels adaptés à des stress environnementaux autres que

ceux identifiés dans les plantes modèles très étudiées, et permettent d'enrichir nos connaissances sur l'évolution des espèces et de favoriser le processus d'amélioration variétale (**Figure 1**).



**Figure 1.** Contributions de la génomique sur la génétique et la sélection des plantes non-modèles (Ye and Fan 2021).

A l'ère de la génomique, les marqueurs moléculaires traditionnels utilisés précédemment, tels que les microsatellites (Simple-Sequence Repeats, SSR), ont été remplacés par les marqueurs de type Single-Nucleotide Polymorphism (SNP) grâce à l'essor des technologies de séquençage. La cartographie de QTL (Quantitative Trait Loci), première étape importante pour l'identification de gènes d'intérêt et la sélection assistée par marqueurs, a aussi bénéficié de ce changement d'échelle dans l'analyse des données. L'apparition d'études GWAS (Genome-wide Association Studies) permet, grâce à la production de plusieurs milliers de SNPs, d'identifier des dizaines de loci associés à des caractères agronomiques chez plusieurs centaines d'individus (Tibbs Cortes et al. 2021; Alseekh et al. 2021). Cependant, la grande majorité des facteurs

moléculaires correspondant aux régions identifiées par ces méthodes reste inexpliquée, et il s'avère donc essentiel de conduire en parallèle des études génétiques et moléculaires pour valider les résultats obtenus par la méthode GWAS.

Les récentes avancées en séquençage et géotypage haut débit fournissent donc l'opportunité de générer rapidement et à moindre coût des ressources génétiques, génomiques, transcriptomiques et métabolomiques pour les organismes non-modèles très peu étudiés. Le vanillier s'insère dans ce contexte d'essor technologique. Très peu de ressources étaient disponibles au début de ma thèse : aucun génome de référence et aucune carte génétique de l'espèce cultivée *Vanilla planifolia* G. Jackson ex Andrews. Ce travail de thèse illustre l'apport de la génomique pour les vanilliers cultivés et leurs apparentés sauvages, en rendant possible l'exploration fine de la variabilité génétique et l'identification de régions impliquées dans un stress biotique majeur causé par le champignon *Fusarium oxysporum*.



# CHAPITRE 1 |

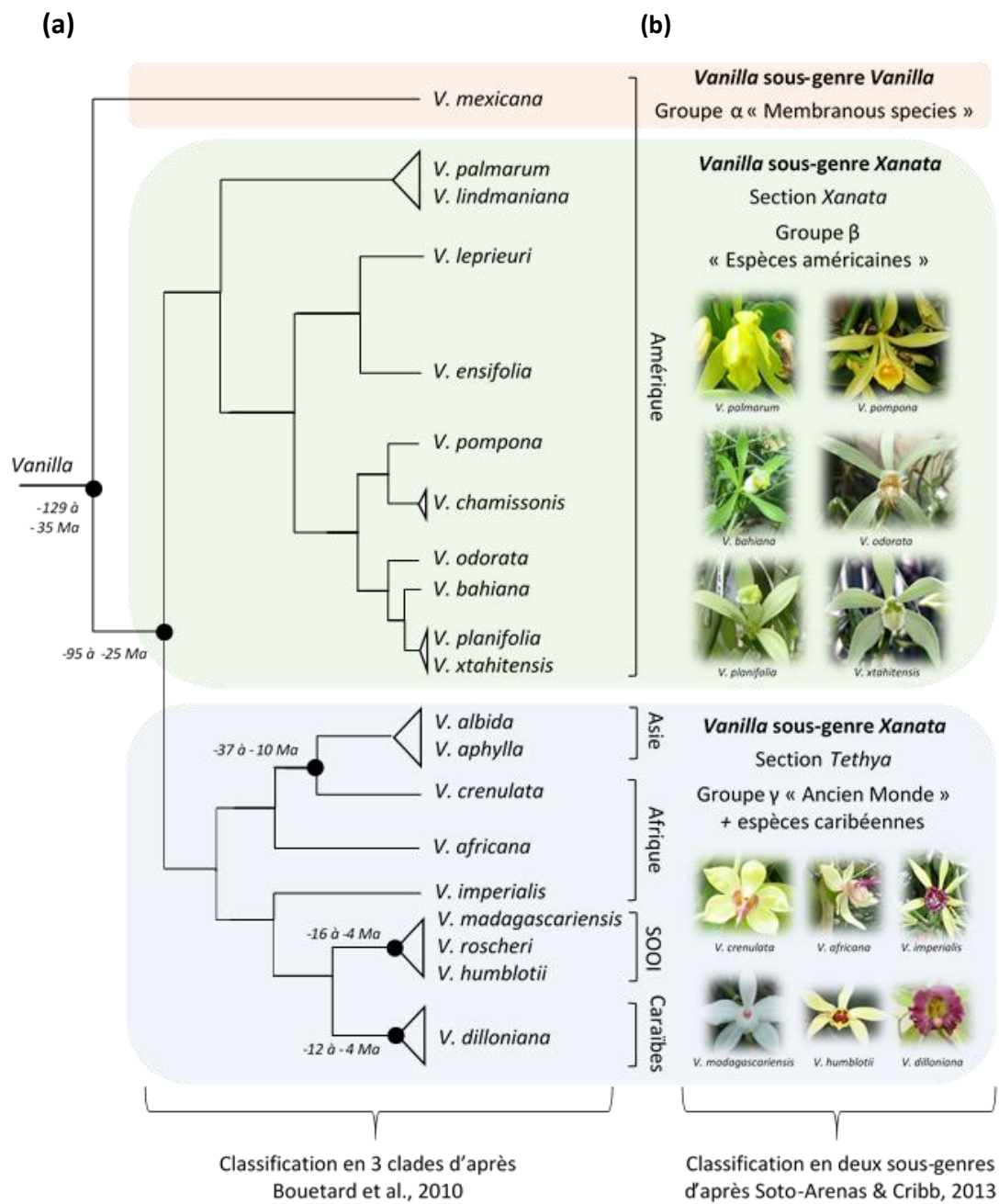
## SYNTHÈSE BIBLIOGRAPHIQUE

---

### 1. Origine et diversité des vanilliers

#### 1.1. Le genre *Vanilla*

Le vanillier est une orchidée d'importance économique majeure dans la région du Sud-Ouest de l'océan Indien (SOOI). Il est cultivé pour ses fruits aromatiques dont est tirée la vanille, une des épices les plus utilisées et appréciées dans le monde, mais aussi une des plus chères. Les vanilliers représentent un groupe d'espèces réparties dans la zone intertropicale (27°N/27°S) appartenant au genre *Vanilla* Plumier ex Miller. Ce genre se place dans la grande famille des *Orchidaceae* et la sous-famille des *Vanilloideae*. Son apparition est estimée à 65 millions d'années (Bouetard et al. 2010). Les premières classifications se basant sur la morphologie des feuilles séparaient les espèces aphyllées (*Aphyllae*) des espèces foliées (*Foliosae*) (Rolfe 1896), elles-mêmes subdivisées en *Papillosae*, *Lamellosae* et *Membranacae* (Portères 1954). L'apport de la biologie moléculaire a cependant confirmé le caractère non monophylétique de ces groupes. Une classification en 3 clades à partir de l'ADN plastidial a été proposée par Bouetard et al. (2010), confirmant l'origine américaine du genre et séparant les vanilliers américains des vanilliers afro-asiatiques (**Figure 1.1**). Deux sous-genres ont été proposés par Soto Arenas et Cribb (2013) à partir de l'étude des séquences nucléaires intergéniques (ITS) de 106 espèces appartenant au genre *Vanilla*, le sous-genre *Vanilla* qui comprend l'espèce ancestrale *V. mexicana* et le sous-genre *Xanata* qui regroupe la section *Xanata* (vanilliers américains) et la section *Tethya* (vanilliers afro-asiatiques) (**Figure 1.1**). Plus de 120 espèces ont été identifiées dans le genre *Vanilla*, parmi lesquelles 18 (Portères 1954) à 35 (Soto Arenas 2003) espèces américaines sont considérées comme aromatiques.



**Figure 1.1.** Phylogénie moléculaire et temps de divergence du genre *Vanilla* estimés à partir des séquences chloroplastiques. **(a)** Classification en 3 clades d'après Bouetard et al. (2010) ; **(b)** classification en deux sous-genres d'après Soto Arenas and Cribb (2013). Ma : million d'années. Les photos des espèces proviennent des bases de données du CRB Vatel (CIRAD, La Réunion).

## 1.2. Les vanilliers cultivés

Les deux espèces aromatiques *V. planifolia* et *Vanilla* × *tahitensis* J. W. Moore sont cultivées pour leur forte teneur en vanilline dans les gousses. *V. planifolia* est majoritairement cultivée dans la région SOOI et représente près de 95% de la production mondiale (Lubinsky et al. 2008). Originaires de Mésoamérique, *V. planifolia* était initialement cultivée sur la côte est du Mexique pour répondre à la demande européenne en épice à partir du milieu du 18<sup>e</sup> siècle. L'espèce *V. planifolia* a ensuite été conservée dans les jardins botaniques européens à partir desquels elle a été introduite en Asie du sud-est et dans la région SOOI. L'absence du pollinisateur naturel dans ces régions a rendu la culture du vanillier difficile jusqu'à la découverte d'une technique de pollinisation manuelle en 1836 par le botaniste belge Charles François Antoine Morren et en 1841 par le jeune esclave réunionnais Edmond Albius. Aujourd'hui, la région SOOI (Madagascar, La Réunion, les Seychelles et l'archipel des Comores) fournit à elle seule 60% de la production mondiale (FAOSTAT 2017; Mazhar 2022).

*V. × tahitensis* est cultivée majoritairement dans les îles du Pacifique Sud. Elle a été introduite à Tahiti depuis les Philippines en 1848 (Constantin and Bois 1915). Plusieurs hypothèses ont été avancées pour expliquer son origine. À partir des caractéristiques morphologiques, Portères (1954) propose une hybridation entre *V. planifolia* et *Vanilla pompona* Schiede. L'utilisation de marqueurs de l'ADN ribosomique (ITS) et de l'ADN plastidial suggère plutôt une origine hybride entre *V. planifolia* comme parent maternel et *Vanilla odorata* C. Presl comme parent paternel (Lubinsky et al. 2008). Deux autres espèces aromatiques sont consommées : *V. pompona*, cultivée marginalement dans certaines îles des Caraïbes et *V. odorata* en Amérique centrale et du sud.

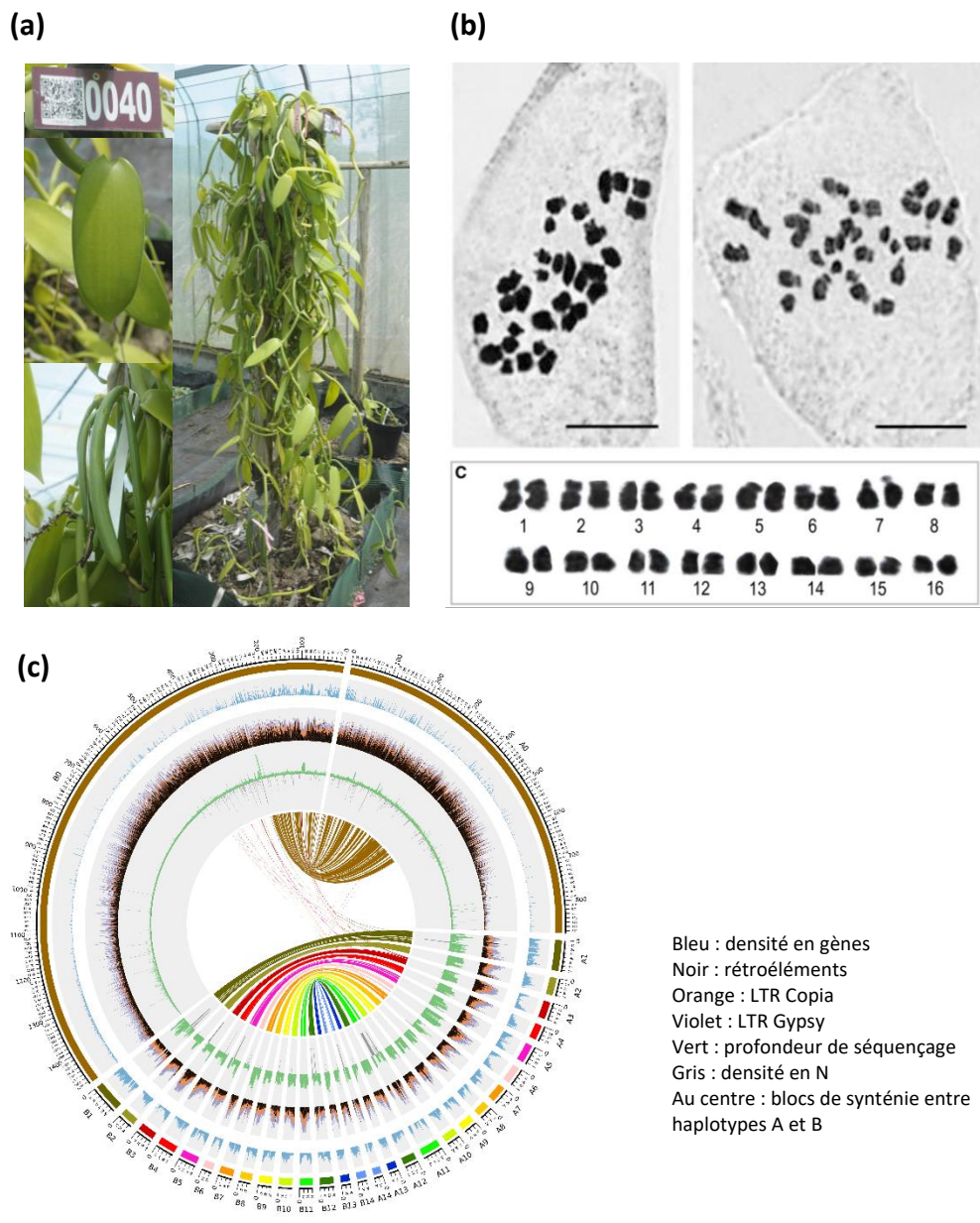
Comme beaucoup d'autres espèces tropicales introduites hors de leur aire d'origine, le vanillier cultivé possède une base génétique très restreinte (Bory et al. 2008b). Ce phénomène est d'autant plus important que les individus cultivés ne sont pour la plupart que des clones obtenus à partir d'une unique bouture ancestrale. Cependant,



au sein du genre *Vanilla*, une grande diversité phénotypique est observée, potentiellement exploitable en amélioration variétale pour les caractères d'intérêt, comme la qualité aromatique des gousses et la résistance aux stress biotiques et abiotiques.

### **1.3. Le génome de *V. planifolia***

La plupart des vanilliers sont diploïdes ( $2n = 32$ ) et seize paires de chromosomes ont été identifiées par des études de cytogénétique chez le cultivar traditionnel réunionnais CR0040 de l'espèce *V. planifolia* (Bory et al. 2008a; Piet et al. 2022). La taille du génome a très récemment été estimée à 4.09 Gb (Piet et al. 2022) (**Figure 1.2**). La diversité phénotypique observée chez certains clones de *V. planifolia*, *V. pompona* et *V. × tahitensis* a été expliquée par l'existence de génomes triploïdes (vanilliers dits « mâles » ou « stériles ») ou tétraploïdes (baptisés « grosse vanille » à La Réunion) (Bory et al. 2008a), montrant l'importance du phénomène de polyploïdisation dans l'évolution du génome de ce genre. Le génome des *Vanilla* spp. est également caractérisé par des événements d'endoréplication partielle, c'est-à-dire la réplication d'ADN de 4 à 64 fois dans certains noyaux (Brown et al. 2017). Ces processus mettent en évidence la complexité du génome et soulèvent de nombreuses questions sur sa dynamique évolutive. Le séquençage du génome du cultivar CR0040, effectué dans l'équipe pendant ma thèse, représente un véritable réservoir de connaissances en appui au programme d'amélioration variétale du vanillier, en particulier en donnant accès aux séquences codantes pour des gènes d'intérêt (Piet et al. 2022).



**Figure 1.2.** Cytogénétique et génomique de *V. planifolia*. **(a)** Plant de vanillier du cultivar CR0040 utilisé pour le comptage des chromosomes, **(b)** métaphases mitotiques et caryotype et **(c)** séquençage du génome complet avec une représentation circulaire des chromosomes assemblés d'après Piet et al. (2022).

## 2. Caractéristiques agronomiques des vanilliers cultivés

### 2.1. Culture du vanillier

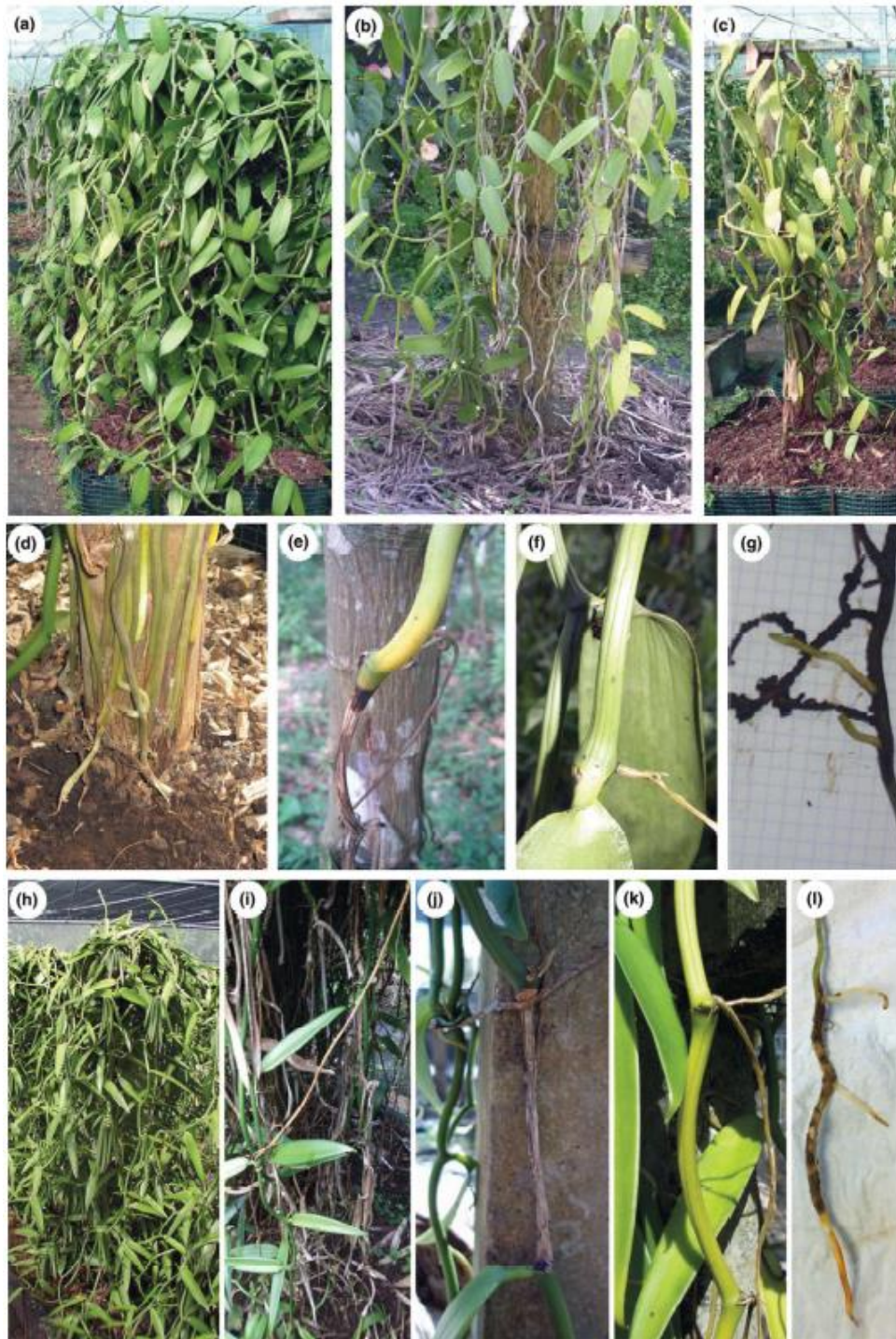
Le vanillier est une liane semi-épiphyte, qui prend racine dans le sol, se développe le long des arbres et peut atteindre jusqu'à 15 mètres de haut. Le vanillier produit deux types de racines : les racines terrestres, longues et ramifiées qui pénètrent le substrat pour absorber l'eau et les sels minéraux et les racines aériennes, généralement courtes, qui permettent l'ancrage des tiges au support et peuvent se transformer en racines terrestres pour améliorer l'absorption. La reproduction des vanilliers est majoritairement végétative (Bory et al. 2008b). L'autofécondation naturelle est impossible car le rostellum empêche le stigmate d'être en contact direct avec les grains de pollen. La fécondation naturelle nécessite l'intervention d'insectes spécifiques, présents uniquement dans les forêts d'Amérique centrale et absents des principales zones de culture (Bouriquet 1954; Gigant et al. 2011). Les modes de culture des vanilliers en sous-bois, en champ ou en ombrières, sont caractérisés par une faible variabilité intra-spécifique, par une forte homogénéité spatiale, et par une densité en plants élevée. Tous ces facteurs participent à la vulnérabilité accrue des vanilliers aux stress environnementaux et aux maladies (Dequaire 1976; Besse et al. 2004; Duval et al. 2006).

### 2.2. Les maladies majeures des vanilliers

La production mondiale de vanille est limitée par des contraintes abiotiques et biotiques, qui aboutissent à d'importantes pertes à l'origine d'une importante fluctuation des prix. La chute prématurée des fruits a récemment été observée et est probablement liée aux changements globaux de températures et de pluviométrie à des saisons décalées (Hernandez-Hernandez, 2011 ; Nascimento et al., 2019). Les vanilliers sont aussi affectés par des virus (essentiellement *Cymbidium mosaic virus* - CymMV et *Cucumber mosaic virus* - CMV) et des cochenilles (*Conchaspis angraeci*) présentes depuis 1997 à La Réunion (Odoux and Grisoni 2010), mais les maladies fongiques

constituent les menaces principales pour la culture du vanillier. Le champignon phytopathogène *Colletotrichum gloeosporioides* est par exemple l'agent responsable de l'antracnose du vanillier, caractérisée par des taches brunes et des lésions sur les feuilles ou les fruits conduisant au flétrissement de la plante (Divakaran et al. 2008). *Phytophthora* spp. provoque la pourriture noire des racines, des tiges, des feuilles et des fruits (Divakaran et al. 2008). Le champignon *Colletotrichum orchidophilum* a quant à lui été observé pour la première fois sur l'île de La Réunion en 2011. La maladie est caractérisée par des taches noires sur les feuilles, les tiges et les fruits (Charron et al. 2018).

La principale maladie des vanilliers est la pourriture des racines et des tiges (**Figure 1.3**), causée par le champignon du sol *Fusarium oxysporum f. sp. radicis-vanillae* (Forv) (Koyyappurath et al. 2015a). Les symptômes ont été observés dans les sols des zones de culture dès le début du 20<sup>ème</sup> siècle, à la Réunion (1901), en Indonésie (1903), à Porto Rico (1918) et aux Seychelles (1921), puis la maladie s'est répandue dans toutes les zones de culture du vanillier. Les symptômes sont principalement caractérisés par la pourriture des racines et des tiges souvent limitée à 2 ou 3 entre-nœuds, limitant l'absorption des nutriments et de l'eau. Des tiges rainurées sont occasionnellement observées témoignant un manque d'eau dans la plante. Une autre caractéristique de cette maladie est la transformation des racines aériennes en racines terrestres pour compenser celles infectées par le champignon. Les méthodes de lutte chimique ou prophylactique se révèlent toutes inefficaces et l'utilisation de plantes résistantes paraît être la meilleure stratégie pour une gestion à long terme de la maladie.



**Figure 1.3.** Symptômes de la pourriture des racines et des tiges sur *V. planifolia* (a-g) et *V. × tahitensis* (h-l). Plantes saines (a,h) ; Plantes malades présentant des chloroses sur les feuilles et sur de nombreuses racines aériennes de vanilliers cultivés en sous-bois (b) et en ombrière (c,i) ; Prolifération de racines aériennes qui présentent des symptômes une fois qu'elles ont atteint le sol (d) ; Pourriture de la tige restreinte à un ou deux entre-nœuds (e,j) ; Tiges rainurées sur plant flétri (f,k) ; Nouvelles racines générées en amont des racines pourries (g,l). D'après Koyyappurath et al. (2015a).



### **3. *Fusarium oxysporum* f.sp. *radicis-vanillae***

#### **3.1. *Fusarium oxysporum***

*Fusarium oxysporum* f.sp. *radicis vanillae* (communément appelé Forv) est un champignon microscopique du sol qui se développe en milieu humide, à une température comprise entre 22 et 37°C (Agrios 1997; Koyyappurath et al. 2015a). Il possède trois types de spores asexués : les microconidies, les macroconidies et les chlamydospores (Nelson 1981). Sur un milieu de culture en laboratoire, il donne un mycélium filamenteux, cotonneux, blanc, saumon, pêche ou violet, en fonction des mycotoxines produites par la souche. Sous la forme de chlamydospores, le champignon est capable de survivre plusieurs dizaines d'années dans le sol et représente donc une menace particulièrement importante pour la culture du vanillier. Très peu d'études concernent les formes spéciales « *radicis* », contrairement aux formes « vasculaires » retrouvées notamment dans les grandes cultures tropicales comme le bananier et le coton, et qui colonisent le xylème. Parmi la centaine de formes spéciales de *F. oxysporum* décrites, seulement six sont de type « *radicis* » : *radicis-betae*, *radicis-capsici*, *radicis-cucumerinum*, *radicis-lupini*, *radicis-lycopersici*, *radicis-vanillae* (Edel-Hermann and Lecomte 2019).

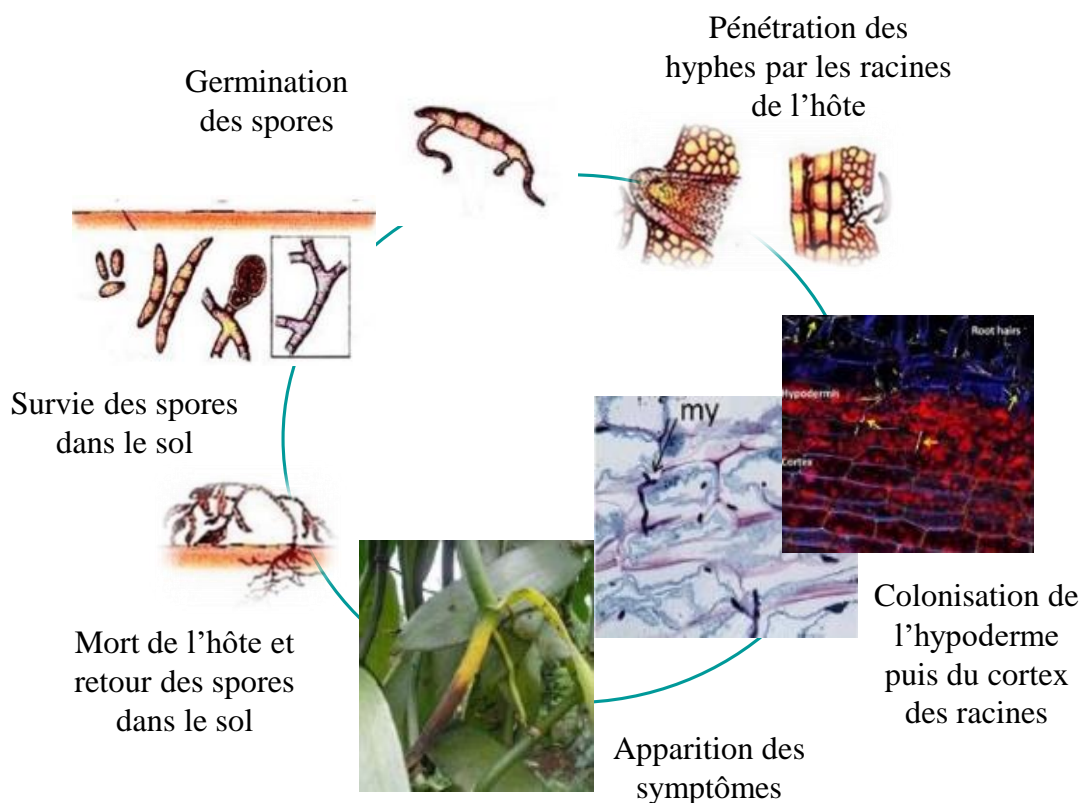
#### **3.2. Diversité génétique de Forv**

La diversité génétique des souches de *Forv* a été étudiée par génotypage moléculaire (séquences *EF1 $\alpha$*  et IGS) dans le cadre des travaux de thèse de S. Koyappurath (2015). Les 377 isolats de *Fusarium* retrouvés dans les vanilleraies de Madagascar, à La Réunion et en Polynésie française présentent une grande diversité, avec l'identification de 20 haploypes *Forv* groupés en 11 haplogroupes. Aucune corrélation n'a été observée entre les haplogroupes et l'origine géographique ou le niveau de pathogénicité des souches. Ces résultats suggèrent donc un schéma d'évolution hautement polyphylétique (Koyyappurath et al. 2015a). Un gradient continu de pathogénicité des différentes souches a également été mis en évidence. La souche

hautement pathogène Fo72 a été utilisée au cours de ma thèse pour les expériences d'infection des racines de vanilliers en milieu contrôlé.

### 3.3. Infection et pathogénicité

Forv est un champignon à mycélium cloisonné, appelé aussi hyphe. Il émet des spores, qui, après germination, pénètrent dans la plante par les poils absorbants afin d'atteindre les systèmes corticaux sans jamais atteindre le système vasculaire, contrairement à la majorité des formes spécialisées. La pénétration est permise tout d'abord par des cellules de passage, puis par une dégradation de la paroi cellulaire de l'hôte (**Figure 1.4** ; Koyyappurath et al. 2015b). Il en résulte un brunissement et une nécrose des racines et des tiges menant la plante jusqu'à la mort (Jarvis and Shoemaker 1978; Perez-Nadales et al. 2014; Koyyappurath et al. 2015a).



**Figure 1.4.** Cycle d'infection de *Fusarium oxysporum* f.sp. *radicis-vanillae*. my : mycélium. Adapté de Perez-Nadales et al. (2014) et Koyyappurath (2015).

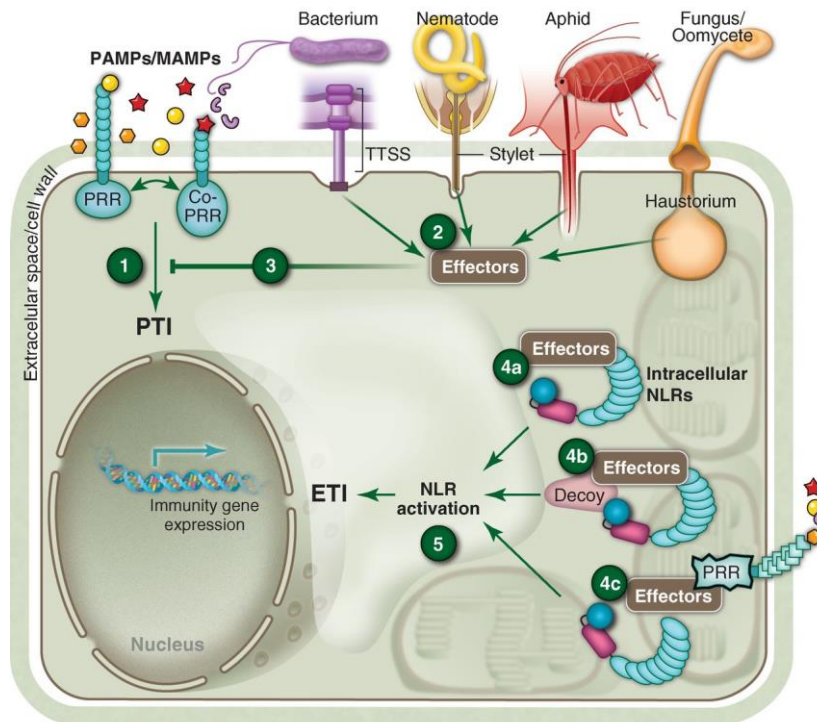
## 4. L'interaction *Vanilla-Forv*

### 4.1. Mécanismes de défense des plantes contre les agents pathogènes

Les plantes ont développé des systèmes pour se protéger des agents pathogènes envahissants. Les barrières physiques, comme la surface composée de cuticules et de cire, permettent de bloquer les agents pathogènes. La libération de composants chimiques comme les phénols, les terpènes et les composés contenant du soufre ou de l'azote, est parfois également impliquée pour dissuader ou éliminer les ennemis envahissants (Keen 1999).

En plus de ces barrières directes, les plantes possèdent un système immunitaire inné qui leur permet d'induire des réponses de défense rapides (Karban and Baldwin 1997) (**Figure 1.5**). Ce système immunitaire permet la mise en place de réactions spécifiques et est basé sur le concept gène pour gène, dans lequel le produit du gène d'avirulence (*avr*) du pathogène est reconnu par le produit du gène de résistance (*R*) de la plante (Jones and Dangl 2006). La réaction hypersensible est une de ces réactions spécifiques qui entraîne l'apoptose des cellules infectées, bloquant la réplication et la propagation de l'agent pathogène (Nicholson and Hammerschmidt 1992; Keen 1999). Les gènes *R* peuvent être divisés en 8 groupes (Gururani et al. 2012). Parmi eux, se trouvent les Leucine-rich repeat (LRR), la famille de gènes *R* majoritaire, constituée de trois sous-familles qui codent pour les LRR-RLK/RLK (domaine LRR + récepteur de type kinase), les LRR-RLP/RLP (domaine LRR + récepteur de type protéine) et les NBS-LRR/NLR (domaine LRR + site de liaison aux nucléotides) (Jones and Dangl 2006). Des gènes *R* dirigés contre *F. oxysporum* ont été trouvés dans des cultures comme la tomate et le concombre (Vakalounakis 1996; Simons et al. 1998; Sela-Buurlage et al. 2001). D'autres études récentes ont mis en évidence une faible densité de gènes *R* chez certaines *Orchidaceae* comparé à d'autres angiospermes (Xue et al. 2020).





**Figure 1.5.** Schéma illustrant le système immunitaire des plantes (Dangl et al. 2013). Les pathogènes expriment des PAMPs (Pathogen-associated molecular patterns) et MAMPs (Microbe-associated molecular patterns) quand ils colonisent les plantes (étoiles, cercles et hexagones colorés). Les plantes les perçoivent *via* des PRRs (Pattern recognition receptors) extracellulaires et initient l'immunité (PTI, PAMP-triggered immunity, étape 1). Les pathogènes produisent des effecteurs de virulence pour bloquer la perception des PAMPs/MAMPs (non montré) (étape 2). Ces effecteurs sont adressés à des localisations subcellulaires spécifiques où ils peuvent supprimer le PTI (étape 3). Les récepteurs intracellulaires de type NLR peuvent reconnaître les effecteurs par 3 voies principales : par interaction directe entre le ligand et le récepteur (étape 4a); en détectant une altération médiée par un effecteur via une protéine leurre qui imite la structure de la cible effectrice, mais qui n'a pas d'autre fonction dans la cellule (étape 4b); ou enfin, en détectant l'altération médiée par l'effeteur d'une cible de virulence de l'hôte, comme le domaine cytosolique d'un PRR (étape 4c). Ces mécanismes conduisent à l'immunité ETI (NLR-dependent effector-triggered immunity).

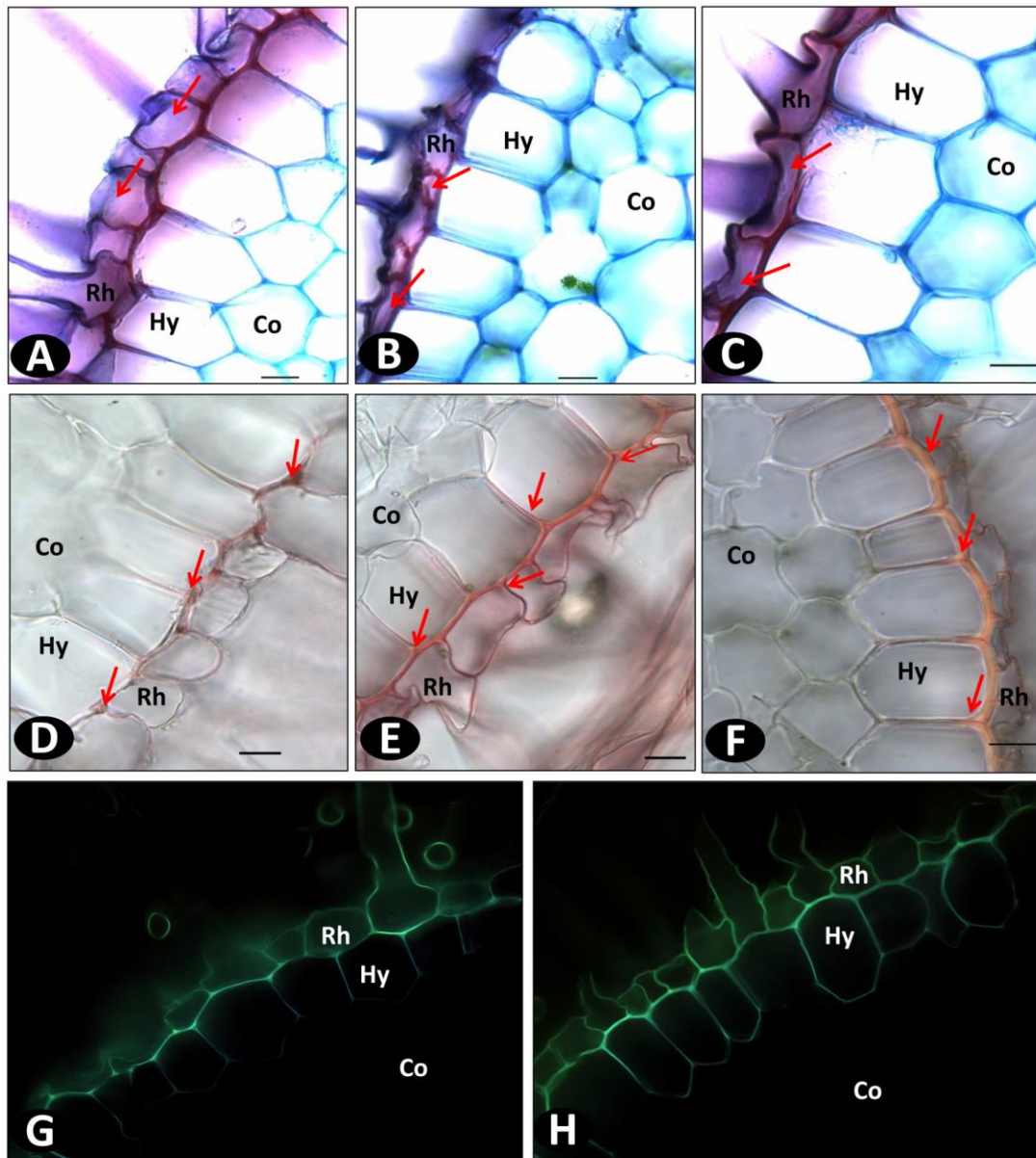
#### 4.2. Aspect génétique de la résistance des vanilliers à Forv

Les phénotypes de résistance vs sensibilité ont été évalués afin d'identifier les degrés de résistance de 18 espèces et 6 hybrides de vanilliers du Centre de Ressources Biologiques (CRB) Vatel (Cirad, La Réunion) vis-à-vis de la souche hautement pathogène Fo072. Parmi les plantes testées, des accessions très résistantes ont été

identifiées, comme CR0020 (*V. planifolia*) et CR0018 (*V. pompona*), ainsi que des accessions très sensibles comme CR0001 ou CR0040 (*V. planifolia*) (Koyyappurath et al. 2015b). De plus, le phénotypage de dizaines d'individus issus de croisements intra-*V. planifolia* a été réalisé (Favre 2019) et a révélé un large spectre de sensibilité à la souche Fo072 de Forv, de très résistants à très sensibles, avec tous les intermédiaires. La répartition des individus dans les différentes classes de sensibilité n'est pas celle attendue dans le cas d'une résistance monogénique, c'est-à-dire ne faisant intervenir qu'un seul gène majeur récessif de résistance. La résistance semble donc plutôt être quantitative polygénique, c'est-à-dire faisant intervenir plusieurs voies, chacune ayant une contribution plus ou moins grande dans le phénotype de résistance global (Koyyappurath et al. 2015b; Declas 2021).

### **4.3. Aspects morphologiques et histologiques de l'interaction vanillier-Forv**

Les processus morphologiques et histologiques au cours de l'infection des racines par Forv ont été décrits chez des accessions sensibles et résistantes du genre *Vanilla* en collection dans le CRB Vatel (Cirad, La Réunion) (Koyyappurath et al. 2015b) (**Figure 1.6**). Chez les accessions sensibles comme CR0001 (*V. planifolia*), la colonisation des hyphes du champignon se produit deux jours post-inoculation (2 jpi) au niveau du rhizoderme et des cellules épidermiques. Le champignon transite ensuite par les cellules de l'hypoderme afin d'atteindre le cortex de la racine à 3 jpi. À 7 jpi, la plasmolyse des cellules corticales commence, et les premiers symptômes apparaissent. Chez les accessions résistantes comme CR0018 (*V. pompona*) et CR0020 (*V. planifolia*), les hyphes restent inféodés aux cellules épidermiques voir hypodermiques qui se transforment même en cellules caliciformes chez *V. pompona*. En effet, la présence d'une paroi hypodermique lignifiée, continue et épaisse bloque la colonisation du cortex par le champignon. Cette caractéristique physique des cellules de l'épiderme et du cortex racinaire joue vraisemblablement un rôle de barrière physique empêchant la pénétration du champignon (Koyyappurath et al. 2015b). Ces observations ouvrent donc de nouvelles perspectives pour l'identification de gènes de résistance à Forv.



**Figure 1.6.** Différences dans la structure et la composition de la paroi cellulaire de racines *in vitro* non-inoculées, observées par microscopie multiphotonique à champ large et différentes méthodes de coloration. **(A-C)** Coupes transversales et coloration FASGA de racines *in vitro*. Les flèches indiquent les cellules du rhizoderme (Rh) qui sont arrondies chez l'accésion sensible CR0001 (*V. planifolia*) **(A)** et compactes chez les accèsions résistantes **(B)** CR0020 (*V. planifolia*) et **(C)** CR0018 (*V. pompona*). **(D-F)** Coupes transversales de racines *in vitro* colorées au phloroglucinol. Les flèches indiquent le dépôt de lignine sur l'hypoderme (Hy) qui est discontinu et fin chez CR0001 **(D)**, et qui est au contraire continu et épais sur les parois longitudinales et radiales des cellules de l'hypoderme chez CR0020 **(E)** et CR0018 **(F)**. **(G, H)** Différences entre CR0001 **(G)** et CR0020 **(H)** dans la composition des cellules du rhizoderme et des cellules de l'hypoderme par coloration à l'hémisulfate de berbérine qui a une affinité avec la subérine et la lignine. Co : Cortex. Barre d'échelle = 20  $\mu$ m. D'après Koyyappurath et al. (2015b).



concentrations significatives en acide p-coumarique et coniférylique et en acide sinapique dans la paroi et les cellules des racines à 7 jpi ont été observées en réponse à l'infection par Forv (Koyyappurath et al. 2015b). Dans cette même étude, la présence de DEGs (Differentially expressed genes) impliqués dans la voie de biosynthèse de la lignine dans les racines a été observée entre des accessions sensibles CR0001 et résistantes CR0020 après inoculation par Forv. Il a ainsi été observé qu'il existait une surexpression des gènes codant pour PAL5 (Phénylalanine Ammoniac Lyase 5) et 4CL (4-Coumarate Coenzyme A ligase) à 1 jpi chez les accessions résistantes. Ces surexpressions arrivent beaucoup plus tardivement chez les accessions sensibles, à 7 jpi. Ces observations pourraient alors suggérer que la résistance serait liée à un aspect temporel de la reconnaissance du champignon par l'hôte (Pilloni 2017). Cette étude a été réalisée sur des intervalles réguliers de 1 jour. Il serait donc intéressant d'étudier ces DEGs de manière précoce après inoculation afin de mieux comprendre l'aspect temporel de la mise en place de la résistance (Morin 2021). La voie de biosynthèse de la lignine repose sur celle des phénylpropanoïdes qui est également la voie impliquée dans la biosynthèse de la vanilline (Dixon et al. 2002; Koyyappurath et al. 2015b). Dans une étude portant sur les gènes impliqués dans la voie de synthèse des phénylpropanoïdes, l'implication des gènes codant pour Vp4HBS (*V. planifolia* 4-hydroxybenzaldéhyde synthase), VpOMT2 (*V. planifolia* O-méthyltransferase 2) et VpOMT3 (*V. planifolia* O-méthyltransferase 3) dans la biosynthèse de la lignine au niveau des racines a été mise en évidence (Fock-Bastide et al. 2014). Outre les études menées sur les gènes impliqués dans les voies de biosynthèse de la lignine, une autre étude portait sur le génome et le transcriptome complets de l'accession Daphna de *V. planifolia*. L'annotation ciblée des gènes analogues de résistance (RGAs, Resistance gene analogs) et la comparaison des séquences des gènes potentiellement impliqués dans la résistance à *Fusarium* chez d'autres espèces de plantes ont été réalisées et ont pu permettre l'identification de 11 gènes putatifs possiblement liés à la résistance à Forv. Parmi eux, 9 présentent des profils d'expression différentiels liés à l'infection, dont VpI02Ag05172.1 (Rao et al. 2014; Hasing et al. 2020).

## 5. Problématique et objectifs de la thèse

Les recherches menées dans l'équipe vanillier de l'UMR PVBMT, où j'ai effectué ce travail de thèse, visent à mieux caractériser la génétique et la génomique des vanilliers cultivés et sauvages, afin d'identifier des régions d'intérêt agronomique, notamment impliquées dans la résistance aux stress biotiques. Ces connaissances s'avèrent indispensables pour mener à bien le programme d'amélioration variétale démarré au début des années 2000 au CRB Vatel du CIRAD de La Réunion, dans le contexte d'une agriculture durable et respectueuse de l'environnement, et ainsi élaborer de nouvelles stratégies de lutte contre les agents pathogènes qui affectent la production de vanille à l'échelle mondiale. L'équipe s'intéresse en particulier aux facteurs génétiques impliqués dans la résistance à la fusariose, maladie retrouvée dans toutes les zones de culture du vanillier, dont le SOOI qui constitue le premier producteur mondial de vanille. Les études menées s'appuient sur la disponibilité en collection de plus de 700 accessions représentatives de la diversité mondiale du genre *Vanilla*, et d'une population issue de l'autofécondation du cultivar traditionnel réunionnais CR0040 qui ségrège pour la résistance à Forv. L'obtention de populations de cartographie est un processus très long chez le vanillier : un pas de temps de 8 à 10 ans est nécessaire entre la première étape de croisement ou d'autopollinisation et la caractérisation complète en champ des nouveaux génotypes qui en sont issus. La disponibilité d'une telle population est donc la clé de la réussite de ce travail de thèse. Une expertise forte en phénotypage précoce de la maladie sur des vanilliers cultivés *in vitro* a également été à la base des travaux initiés au cours de cette thèse. Enfin, très récemment, l'équipe a séquencé plus de 83% du génome du cultivar CR0040, offrant ainsi, au cours de ma thèse, un réservoir de gènes annotés potentiellement impliqués dans des mécanismes de résistance aux pathogènes. Au niveau international, les travaux sur les interactions plantes-*Fusarium* portent essentiellement sur les formes « vasculaires » et non « radicales » de *F. oxysporum*, suggérant la mise en place de mécanismes de résistance très différents. Au début de ma thèse, il n'existait aucune étude génétique sur la résistance à Forv chez le vanillier, et la variabilité génétique au sein des vanilliers cultivés était insuffisamment résolutive.

Dans ce contexte, ce projet de thèse a été initié pour mieux appréhender la diversité intra-spécifique des vanilliers cultivés, dans l'objectif d'identifier de la variabilité au niveau des régions impliquées dans la résistance à la fusariose. La stratégie employée repose sur l'apport de la génomique et l'utilisation de technologies de génotypage à haut débit. Aucun génome de référence et aucune carte génétique n'était disponible au début de ma thèse, et le choix d'utiliser la méthode GBS s'est avéré approprié pour les études génétiques sur cette plante orpheline d'un point de vue des données scientifiques disponibles en 2019.

La suite du manuscrit s'articule autour de quatre chapitres de résultats, dont trois sont rédigés sous forme de publications (deux articles publiés et un soumis). Nous aborderons successivement (i) l'adéquation de la méthodologie GBS avec nos questions de recherche portant sur la diversité et la cartographie génétique dans un contexte de plante tropicale complexe telle que le vanillier, (ii) la résolution de la diversité intra-spécifique des espèces cultivées *V. planifolia* et *V. × tahitensis* grâce à l'obtention d'une grande quantité de SNPs répartis de façon homogène sur ce génome de plus de 4 Gb, (iii) la construction de la première carte génétique haute densité de *V. planifolia* et la mise en évidence de QTL impliqués dans la résistance à la fusariose, et enfin (iv) une analyse, à l'échelle globale du génome, des grandes familles de gènes de résistance connues. Une discussion générale portant sur l'ensemble de mes résultats et les perspectives de ce travail constituent le dernier chapitre.

## CHAPITRE 2 |

# LE GÉNOTYPAGE À HAUT DÉBIT EN APPUI AUX RECHERCHES SUR LES PLANTES TROPICALES COMPLEXES

---

### 1. Préambule

L'étude des variations génétiques des espèces cultivées et des espèces qui leur sont apparentées est cruciale pour la conservation de la biodiversité et l'identification de facteurs génétiques impliqués dans des caractères d'intérêt en appui aux programmes d'amélioration variétale. Les marqueurs moléculaires sont des séquences d'ADN utilisées comme indicateurs pour étudier le polymorphisme chez un organisme. Il existe différents types de marqueurs moléculaires, tels que les marqueurs RFLP (Restriction fragment length polymorphism), RAPD (Random amplified polymorphic DNA), AFLP (Amplification fragment length polymorphism), SSR (Simple sequence repeats) ou les marqueurs SNP (Single nucleotide polymorphism). De précédentes études de diversité génétique dans le genre *Vanilla* ont échoué à identifier la diversité intra-spécifique au sein des espèces cultivées (*i.e.* *V. planifolia* et *V. x tahitensis*) pour lesquelles des variations phénotypiques sont pourtant observées (Besse et al. 2004; Bory et al. 2008b).

L'exploration de la diversité inter- et intraspécifique est aujourd'hui facilitée par l'essor des nouvelles technologies de séquençage (Next generation sequencing, NGS) et le développement des méthodes de génotypage à haut débit pour générer des marqueurs moléculaires. Parmi ces méthodes, le génotypage par séquençage (GBS) est une approche qui permet l'identification de marqueurs de type SNP, simultanément chez un grand nombre d'individus. Cette méthode a pour avantage de pouvoir être utilisée avec ou sans génome de référence chez des espèces avec des génomes complexes comme celui des vanilliers. Ce deuxième chapitre de la thèse a pour objectif de présenter la méthode GBS, avec une description détaillée de la technique et des



développements bioinformatiques pour le traitement des grandes quantités de SNPs générés, et ses nombreuses applications pour l'étude des plantes complexes. Ce travail a été valorisé par une publication en tant que premier auteur dans l'ouvrage « Molecular Plant Taxonomy, Methods and Protocols » en 2021.

## **2. Publication 1 : Genotyping-by-Sequencing Technology in Plant Taxonomy and Phylogeny**

### **Auteurs**

Félicien Favre<sup>1</sup>, Cyril Jourda<sup>2</sup>, Pascale Besse<sup>1</sup> and Carine Charron<sup>2</sup>

<sup>1</sup> Université de La Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>2</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France

### **Ouvrage de publication**

Molecular Plant Taxonomy, Methods and Protocols, Second Edition, pages 167-178, Springer, 2021

[https://doi.org/10.1007/978-1-0716-0997-2\\_10](https://doi.org/10.1007/978-1-0716-0997-2_10)

### **Mots clés**

Génotypage par séquençage (GBS), Séquençage nouvelle génération (NGS), Polymorphisme nucléotidique (SNP), Marqueurs moléculaires, Diversité des plantes





## Genotyping-by-Sequencing Technology in Plant Taxonomy and Phylogeny

Félicien Favre, Cyril Jourda, Pascale Besse, and Carine Charron

### Abstract

Genotyping-by-sequencing (GBS) is a method to discover and genotype simultaneous genome-wide high-throughput single nucleotide polymorphisms (SNPs). GBS is based on reducing genome complexity with restriction enzymes. Here we describe a method developed by Elshire et al. for constructing simplified GBS libraries and recent bioinformatic approaches developed to analyze the large volume of polymorphism data generated by this method. GBS approach is suitable for population studies, taxonomic and phylogenetic studies, germplasm characterization, and breeding and trait mapping for a wide range of organisms, including plants with complex genomes.

Key words Genotyping-by-sequencing (GBS), Next-generation sequencing (NGS), High-throughput single nucleotide polymorphism (SNP), Plant diversity markers

---

## 1 Introduction

### 1.1 *Genotyping-by-Sequencing*

Nowadays, the exploration of genetic diversity of plant is enhanced by advanced high-throughput sequencing (HTS) technologies, which provide the opportunity to simultaneously discover a high number of molecular markers at relatively low cost. In particular, genotyping-by-sequencing (GBS) is a method to discover and genotype genome-wide high-throughput single nucleotide polymorphisms (SNPs) in a large number of individuals at the same time. Plant genomes are complex, and sequencing more than one entire genome would be expensive and time-consuming. GBS in any large genome species requires reduction of genome complexity, which can be done by different approaches. The target enrichment approaches can use long-range PCR of specific genes or genomic subsets, molecular inversion probes, or hybridization-based sequence capture methods such as microarrays [1]. However, these methods require invariant primer binding site and remain technically difficult and expensive for assaying many samples at the same time. GBS is associated with restriction-site-reduced

complexity approach (RAD) [2]. The concept is based on acquiring the sequence adjacent to a set of particular restriction enzyme (RE) recognition sites. Large volumes of polymorphism data can be generated by applying massively parallel sequencing and multiplexing RAD with RAD tag libraries. Herein we report a method for constructing GBS libraries based on reducing genome complexity with REs [3]. This approach is simple, quick, extremely specific, and highly reproducible and may reach important regions of the genome that are inaccessible to sequence capture approaches. A large number of methylation-sensitive REs with different size recognition sites can be chosen. Methylation-sensitive REs are not able to cleave methylated cytosine residues; thus they target gene regions and filter out repetitive genomic regions. Thousands of genome-wide markers can be identified with better chance to get something linked to the cause of the polymorphism. GBS focuses on next-generation sequencing (NGS) power to sequence the end of restriction fragments. Advances in NGS throughout the last decade have enabled GBS to be used for high diversity and large genome species. The method is based on a multiplex sequencing strategy that uses an inexpensive barcoding system. Barcodes are included in one of the adapter sequence and located just upstream of the RE cut site in genomic DNA. This procedure generates restriction fragments with appropriate adapter, limits the sample handling, and facilitates the association of fragments to the sample. GBS was applied initially to maize and barley mapping populations but provides results independently of the target species or population and does not require having previous available genomic information. Recent advances in bioinformatics and development of new software programs such as STACKS [4] are able to overcome the lack of reference genome, by using de novo assembly of short sequenced reads. GBS was already used in a large amount of studies in recent years. For instance, an analysis of genetic diversity of European blueberry cultivars by GBS has allowed to better define phylogeny and adaptation of plants to their environment in terms of flowering and fruit ripening [5]. These results should help the preservation of genetic resources and contribute to further breeding programs. GBS is also useful to explore the genetic structure of populations, such as in *Cynara cardunculus*, showing subpopulations within artichokes and cultivated cardoon [6]. Molecular markers identified by GBS are particularly useful for marker-assisted selection (MAS) to enhance genomic selection in plant breeding programs in wheat [7]. GBS has been successfully used in pepper with a wide range of applications. An important amount of informative genome-wide SNPs were identified and enabled to analyze germplasm diversity and population structure as a result of domestication or local adaptation [8, 9]. GBS-generated SNP markers have been also useful in the detection of trait-associated quantitative trait loci (QTLs) for both *Capsicum annuum* and *Capsicum*

*baccatum* and will support genome-wide association mapping studies and marker-assisted selection programs [10, 11]. This approach is also particularly efficient to identify QTLs or genes of interest involved in resistance to plant disease [12], in plant architecture [13], and in plant metabolite content [14]. Here we describe the highly multiplexed system developed by Elshire et al. for constructing GBS libraries for Illumina sequencing. Then we describe the de novo assembly using STACKS and a bioinformatic way to identify SNPs.

### 1.2 Some Limitations to the GBS Method

1. GBS sequencing produces a lot of missing data [15]. This may be partly explained by three main reasons: (a) the lack of the restriction site in particular samples, (b) polymorphism in restriction site, and (c) a low sequence coverage rate. Simulations showed that locus identification was highly reproducible with a sequence coverage somewhere between 20 and 40X [4].
2. This approach gives a random access to genomic regions, because of structure variations and repeated sequences, which are different in each individual.
3. The larger the library is, the more missing data is generated (see Note 1).
4. The most important under the GBS approach is to obtain enough high-quality molecular markers to answer to our questions. The GBS protocol can be modified to be used with new species or different enzymes, mainly to obtain more markers or fewer markers but with a deeper sequence coverage per locus, to increase multiplexing, to avoid more repetitive DNA classes, or for novel applications.

---

## 2 Materials

### 2.1 DNA Extraction and Quantification

1. DNA spin columns-based commercial kit such as DNeasy Plant Mini Kit (Qiagen, Hilden, Germany).
2. Qubit 4 Fluorometer and Qubit assays for DNA quantification (Invitrogen, Carlsbad, CA, USA).
3. *Hind*III or *Eco*RI restriction endonuclease (New England Biolabs, Ipswich, Massachusetts, USA). The nucleic acid recognition sequences where the enzymes cut are, respectively, 5'-A/AGCTT-3' and 5'-G/AATTC-3'.
4. TAE buffer (1X): 0.04 M Tris-acetate and 0.001 M EDTA pH 8.0.
5. Agarose gel 2%.

## 2.2 GBS Library Construction

1. Sequences of double-stranded barcode adapter:  
5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT  
xxxx-3' and  
5'-CWGyyyyAGATCGGAAGAGCGTCGTGTAGGGAAAG  
AGTGT-3',  
where "xxxx" and "yyyy" indicate, respectively, the barcode and barcode complement end sequences.
2. Sequence of double-stranded common adapter:  
5'-CWGAGATCGGAAGAGCGGTTCAGCAGGAATGCCG  
AG-3' and  
5'-CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3'.
3. TE buffer (1X): 10 mM Tris-HCl and 1 mM EDTA-NaOH, pH 8.0.
4. Thermocycler.
5. PicoGreen (Invitrogen, Carlsbad, CA, USA) or similar instrument for quantification of the adapter.
6. PCR 96-well plate.
7. *Ape*KI that recognizes the sequence 5'-G/CWGC-3' (New England Biolabs, Ipswich, Massachusetts, USA) or appropriate RE.
8. NEB Buffer 3 (1X): 50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 1 mM DTT, pH 7.9 at 25 °C (New England Biolabs, Ipswich, Massachusetts).
9. Ligase buffer with ATP and T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA).
10. QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).
11. PCR primer 1: 5'-AATGATACGGCGACCACCGAGATCTCACTCTTTCCCTACACGACGCTCTTCCGATCT-3'.
12. PCR primer 2: 5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCG ATCT-3'.
13. *Taq* Master Mix (1X) (New England Biolabs, Ipswich, Massachusetts, USA).
14. Bio-Rad Experion (Bio-Rad, Hercules, California, USA) or similar instrument.

## 2.3 Illumina Workflow

The protocol should be optimized depending on the sequencer used. Here, we describe the main steps of Illumina sequencing (Illumina Inc., San Diego, California, USA).

## 2.4 Data Analysis Equipment and Softwares

Data must be processed within a high-performance computing cluster. The following softwares and tools are used:

1. FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).
2. Cutadapt [16].
3. Trimmomatic [17], the FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)), or similar tool.
4. GBS barcode splitter (<https://sourceforge.net/projects/gbsbarcode/>) and FASTQ/A Barcode splitter from the FASTX-Toolkit.
5. STACKS [4].
6. South Green bioinformatics platform (<https://www.southgreen.fr>).
7. R package “pegas” [18].
8. STRUCTURE [19].

---

## 3 Methods

### 3.1 DNA Extraction and Quantification

1. Extract high molecular weight DNAs from young leaves using a standard CTAB protocol or using DNA spin column-based commercial kits according the manufacturer’s instructions (see Chapter 3).
2. Quantify genomic DNA by fluorimetric assays, and normalize gDNA concentrations at 50 ng/ $\mu$ L (see Note 2).
3. To test DNA homogeneity, mix 1  $\mu$ L of uncut sample DNA with 4  $\mu$ L of loading dye, load in a 2% agarose gel, and run at 110 V for 2 h. The gel must reveal one clear band for each sample.
4. Test whether DNA extractions are of sufficient quality by enzymatic digestion. The digestion test doesn’t need to be done with methylation-sensitive enzymes that cleave only at unmethylated recognition sites. Cheaper RE that is not methylation-sensitive such as *HindIII* or *EcoRI* should be chosen. Pool some DNAs from the same extraction series to have 500 ng of DNA. Mix 10  $\mu$ L of DNA (50 ng/ $\mu$ L) with 7.3  $\mu$ L of ultrapure water, 2  $\mu$ L of RE 10X, 0.2  $\mu$ L of BSA, and 0.5  $\mu$ L of enzyme. The mix is incubated first at 37 °C for 4 h and at 80 °C for 20 min, loaded in a 2% agarose gel, and run at 110 V for 2 h. The gel must reveal a regular smear without band.

### 3.2 GBS Library Construction

1. Two kinds of adapters are used for constructing GBS libraries, a barcode adapter and a common adapter. Adapters are designed to fit with Illumina sequencing (see Note 3). Dilute



oligonucleotides of each barcode and common adapters separately in TE buffer (50  $\mu$ M each), and anneal them in a thermocycler at 95 °C during 2 min. Then decrease the temperature to 25 °C by 0.1 °C/s and wait 30 min at 25 °C. Hold the adapters at 4 °C.

2. Quantify the adapters with an intercalating dye (PicoGreen (Invitrogen, Carlsbad, CA, USA) or similar instrument) diluted in water to 0.6 ng/ $\mu$ L. Mix then barcode and common adapters together in a 1:1 ratio; plate the mix into a 96-well plate, each well with a different barcode adapter; and dry the plate (Fig. 1, step 1, see Note 4).
3. Add 100 ng of DNA in a volume of 10  $\mu$ L into each well (96 different DNA samples can be studied on one plate). Dry the plate again (Fig. 1, step 1).
4. In each well, digest DNA for 2 h at 75 °C with *Ape*KI, in 20  $\mu$ L volumes containing 1X NEB Buffer 3 and 3.6 U *Ape*KI (Fig. 1, step 2). This step should be optimized depending on the RE that is used (see Note 5).
5. Ligate adapters to the ends of the genomic DNA inserts: add 30  $\mu$ L of a solution containing 1.66X ligase buffer with ATP and T4 ligase into each well. (Fig. 1, step 3).
6. T4 ligase inactivation: incubate samples at 22 °C for 1 h and then heat them to 65° for 30 min.
7. Pool an aliquot of each sample (5  $\mu$ L) into an Eppendorf tube, and apply it to a size exclusion column to remove unreacted adapters. Purify samples using a commercial kit (QIAquick PCR Purification Kit, Qiagen, Hilden, Germany). DNA samples are then eluted in a final volume of 50  $\mu$ L (Fig. 1, step 4).
8. Perform a PCR to amplify the fragment pool in 50  $\mu$ L volumes containing 2  $\mu$ L pooled DNA fragments from step 7, 1X *Taq* Master Mix, and 25 pmol of each PCR primers 1 and 2 (see Note 6). Use the following PCR temperature cycling: 72 °C for 5 min, 98 °C for 30 s; 18 cycles of 98 °C for 30 s, 65 °C for 30 s, and 72 °C for 30 s with a final *Taq* extension step at 72 °C for 5 min (Fig. 1, step 5).
9. Clean up PCR products, and evaluate fragment sizes of the resulting library on a DNA analyzer (Bio-Rad Experion or similar instrument) (Fig. 1, step 6). Libraries without adapter dimer are retained for DNA sequencing (see Note 7).

### 3.3 Illumina Sequencing Workflow

Illumina sequencing being most often outsourced to private companies, we simply propose here a step-by-step workflow rather than a classical wetlab protocol.

1. Perform single-end sequencing or paired-end sequencing of the library in a flow cell channel using the HiSeq 3000/

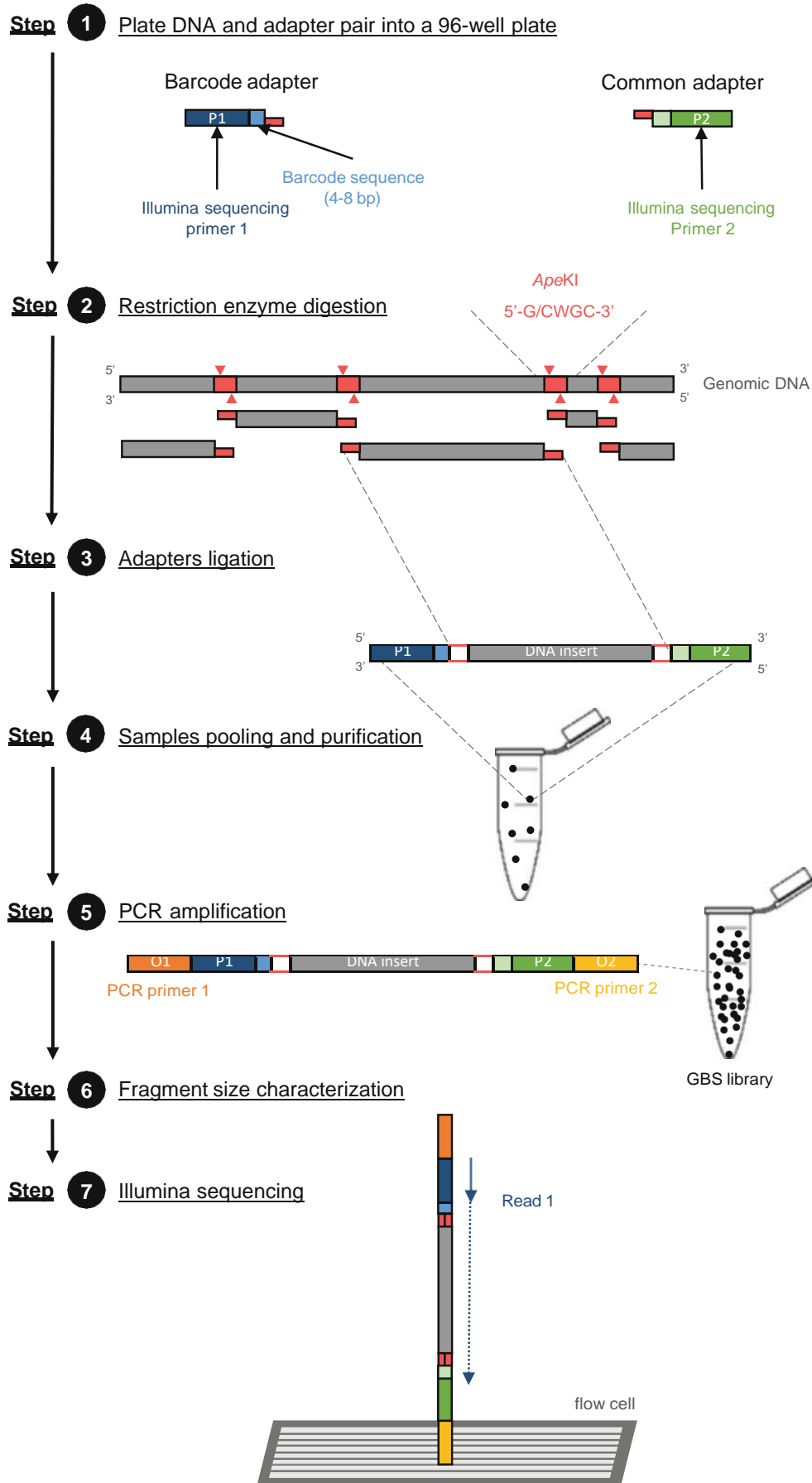


Fig. 1 Diagram of the genomic library construction method based on reducing genome complexity with restriction enzymes (REs) for Illumina sequencing

HiSeq 4000 Systems (Illumina Inc., San Diego, California, USA) (see Note 8).

2. Drop PCR products off at an Illumina flow cell (Fig. 1, step 7).
3. Bridge amplification of DNA fragments in cluster. The amplification is based on solid-phase PCR. Cluster formation amplifies sequencing signal.
  - (a) Bind single-stranded fragments randomly to the inside surface of the flow cell channels.
  - (b) Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.
  - (c) First amplification: the enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.
  - (d) Linearization: denature the double-stranded DNA to leave single-stranded DNA anchored to the substrate.
  - (e) Complete amplification: several million dense clusters of same single-stranded DNA are generated in each channel of the flow cell.
4. Sequencing by synthesis:
  - (a) To initiate the first sequencing cycle, add Illumina sequencing primer P1, DNA polymerase enzyme, and all four nucleotides, each labeled with a different dye, to the flow cell.
  - (b) After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster. Cleave dye and terminating groups and wash.
  - (c) To initiate the next sequencing cycle, add all four labeled nucleotides and enzyme to the flow cell. After laser excitation, collect the image data as previously described. Record the identity of the second base for each cluster.
  - (d) Repeat cycles of sequencing to determine the sequence of bases. This entire process generates millions of reads, representing all the genomic fragments (see Note 9).

### 3.4 GBS Data Processing

#### 3.4.1 Filtering Raw Sequence Data

1. Illumina sequencing produces a large amount of data. Outputs are “fastq” files with four lines per sequence with (a) sequence name; (b) DNA sequence; (c) metadata with sample information such as plant line, location, and year; and (d) quality score of each base of the sequence.
2. View sequences with the quality control tool for high-throughput sequence fastQC.

3. Remove adapter sequences at the end of reads and low-quality extremities. Some tools can be used such as Cutadapt [16] (see Note 10).
4. Remove low-quality reads based both of quality score, read length, low complexity, and N (unsequenced) bases using Trimmomatic [17], the FASTX-Toolkit, or similar tool.
5. Reads are then demultiplexed and assigned to each sample (see Note 11). GBS barcode splitter and FASTQ/A Barcode splitter from the FASTX-Toolkit are tools to split GBS reads by barcode.

#### 3.4.2 Mapping

1. Analyze assigned reads with STACKS to identify all the SNPs [4].
2. Identified SNPs are conserved in a Variant Call Format (VCF) file for analysis [20].
3. Screen SNPs according to several criteria: sequencing depth, missing data, and gene frequency (see Note 12). VCF files derived from SNP calling can be filtered directly by some tools such as SNIPlay which is part of the South Green bioinformatics platform.

#### 3.4.3 Analysis

The identified SNPs can be used in phylogenetic studies. SNIPlay computes on the web series of tools for analyses at a whole-genome scale (general statistics, polyploid analysis, chromosome viewer, SNP density, diversity analysis, association studies, etc.). VCF files can also be analyzed using the R package “pegas” to calculate similarity and construct phylogenetic trees [18]. To analyze the genetic structure of a population and identify groups that are genetically linked, the Bayesian method of the software STRUCTURE can be used [19]. Many other tools are efficient to analyze a VCF files.

---

## 4 Notes

1. The technical option to limit the missing data is to reduce the multiplexing level or sequence the same library several times, and the molecular option is to choose less frequently cutting enzymes.
2. Leaves can be first lyophilized. A quantity of 20–30 mg of lyophilized leaves should be enough for DNA extraction. A great amount of DNA is not needed (50–100 ng/sample). However, quality and quantity of the DNA must be homogeneous. Use preferably fluorimetric method for quantification than spectrophotometric method which could overestimate DNA concentrations. DNA purity is crucial for complete enzyme digestion.

3. The DNA fragment is ligated to a barcode adapter and a common adapter. The barcode adapter contains an Illumina sequencing primer 1 and a barcode which is a 4–8 bp sequence used to identify a sample. The barcode is on the 3<sup>0</sup> end of the top strand of the adapter. The 5<sup>0</sup> end of its bottom strand is terminated by a 3 bp overhang which is complementary to the end genomic DNA fragments generated by the RE. The common adapter is complementary to the other end and is only containing an Illumina sequencing primer 2 end. The adapters are designed for either single-end or paired-end sequencing on the Illumina (Fig. 1, step 1). Barcodes are enzyme specific: they must not recreate the enzyme recognition site to avoid being cut and must have complementary overhangs. Barcodes must be of variable length and different enough from each other to avoid confusion if there is a sequencing error (at least 3 bp differences among barcodes).
4. Up to 96 DNA samples can be processed simultaneously (48/96/384-well plate).
5. Choose methylation-sensitive REs to avoid repetitive regions of genomes and target lower copy regions. Select REs that leave 2 to 3 bp overhangs for efficient adapter ligation to fragments of DNA. *ApeKI* (New England Biolabs) is often used and suitable for maize because it is known to have low recognition sites in maize retrotransposons [3]. *ApeKI* recognizes the sequence 5'-GCWGC-3' (with W is A or T) and leave a 5<sup>0</sup> overhang with 3 bp. *PstI* (New England Biolabs) and *EcoT22I* can also be used and recognize, respectively, the sequences 5'-CTGCA/G-3' and 5'-ATGCA/T-3'. *PstI* was used in artichoke GBS libraries [6].
6. PCR primer 1 is designed to bind, on one hand to 3' strand of barcode adapter and on the other hand to flow cell oligonucleotide 1 for Illumina sequencing. PCR primer 2 binds to 3<sup>0</sup> strand of common adapter and to flow cell oligonucleotide 2. The PCR primers 1 and 2 contain sequences for amplifying restriction fragments with ligated adapters and binding PCR products to oligonucleotides contained in Illumina flow cell (Fig. 1, step 5).
7. Libraries were considered suitable for sequencing if adapter dimers (around 128 bp) are minimal or absent and the majority of other DNA fragments are between 170 and 350 bp. Do the protocol again and adapt and decrease adapter amounts if adapter dimers are present in excess of 0.5%.
8. Two kinds of sequencing can be used for GBS. The single-end sequencing produces reads up to 300 bp and is better for species without reference genome. The paired-end sequencing from both fragment ends generates longer reads from 300 to

500 bp and should be used preferably in species with high-quality reference genome.

9. GBS captures barcode and insert DNA sequence in single read. It ensures that the barcode fits well with its sample because they are physically attached.
10. Different parameters and tools should be tested to remove adapters and low-quality extremities. Reads must be verified each time with fastQC.
11. A table of correspondence between barcode sequences and samples would be very useful for demultiplexing.
12. At least five reads at each locus for each sample are recommended. The minimal gene frequency recommended is 30% to avoid keeping sequencing error. However, gene frequency can be adjusted and reduced depending on the sampling. If the sampling is unbalanced, a low frequency provides access to alleles that are associated with underrepresented individuals in the dataset.

## References

1. Mamanova L, Coffey AJ, Scott CE et al (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–118
2. Miller MR, Dunham JP, Amores A et al (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* 17:240–248
3. Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379
4. Catchen J, Hohenlohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140
5. Campa A, Ferreira JJ (2018) Genetic diversity assessed by genotyping by sequencing (GBS) and for phenological traits in blueberry cultivars. *PLoS One* 13:e0206361
6. Pavan S, Curci PL, Zuluaga DL et al (2018) Genotyping-by-sequencing highlights patterns of genetic structure and domestication in artichoke and cardoon. *PLoS One* 13:e0205988
7. Poland J, Endelman J, Dawson J et al (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5:103–113
8. Taranto F, D'Agostino N, Greco B et al (2016) Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genomics* 17:943
9. Taitano N, Bernau V, Jardón-Barbolla L et al (2019) Genome-wide genotyping of a novel Mexican Chile pepper collection illuminates the history of landrace differentiation after *Capsicum annuum* L. domestication. *Evol Appl* 12:78–92
10. Nimmakayala P, Abburi VL, Saminathan T et al (2016) Genome-wide divergence and linkage disequilibrium analyses for *Capsicum baccatum* revealed by genome-anchored single nucleotide polymorphisms. *Front Plant Sci* 7:1646
11. Nimmakayala P, Abburi VL, Saminathan T et al (2016) Genome-wide diversity and association mapping for Capsaicinoids and fruit weight in *Capsicum annuum* L. *Sci Rep* 6:38081
12. Salgon S, Raynal M, Lebon S et al (2018) Genotyping by sequencing highlights a polygenic resistance to *Ralstonia pseudosolanacearum* in eggplant (*Solanum melongena* L.). *Int J Mol Sci* 19:357
13. Gabay G, Dahan Y, Izhaki Y et al (2018) High-resolution genetic linkage map of European pear (*Pyrus communis*) and QTL fine-mapping of vegetative budbreak time. *BMC Plant Biol* 18:175
14. Gonda I, Ashrafi H, Lyon DA et al (2019) Sequencing-based bin map construction of a tomato mapping population, facilitating high-resolution quantitative trait loci detection. *Plant Genome* 12:1–14
15. Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to

- characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol J* 15:149–161
16. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10
  17. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
  18. Paradis E (2010) Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419–420
  19. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multi-locus genotype data. *Genetics* 155:945–959
  20. Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158

### 3. Conclusion

Le génotypage par séquençage est donc une méthode qui permet l'identification et le génotypage simultané de milliers de polymorphismes nucléotidiques dans tout le génome et est ainsi adaptée aux études de génétique des populations, de taxonomie, de phylogénie, de diversité et vient en appui aux programmes d'amélioration variétale. La méthode a déjà été utilisée chez un grand nombre d'organismes, et notamment chez les plantes avec un génome complexe.

Dans le cadre de ma thèse, cette méthode de génotypage haut débit a donc été choisie pour répondre à deux questions de recherche qui seront développées dans les chapitres 3 et 4 :

- (i) Quelle est la variabilité génétique au sein des vanilliers cultivés et entre individus issus d'une population d'autofécondation d'un cultivar de *V. planifolia* ?
- (ii) Comment exploiter les marqueurs GBS pour générer une carte génétique ultra-dense de *V. planifolia* et identifier des facteurs génétiques liés à la résistance à la fusariose ?





# CHAPITRE 3 |

## RÉSOLUTION DE LA DIVERSITÉ GÉNÉTIQUE INTRA-SPÉCIFIQUE AU SEIN DES VANILLIERS CULTIVÉS

---

### 1. Préambule

Les vanilliers sont un groupe d'espèces complexes caractérisées par un mode de reproduction végétatif, des hybridations intra- et interspécifiques et des événements de polyploïdisation. Ces processus mettent en évidence la complexité du génome et ont un impact sur la diversification du genre *Vanilla*. Parmi la centaine d'espèces de vanilliers, *V. planifolia* et *V. × tahitensis* sont les principales espèces aromatiques cultivées.

L'objectif de ce troisième chapitre est de caractériser la diversité génétique au sein de ces vanilliers cultivés et de leurs espèces apparentées, grâce à la méthode GBS décrite précédemment. Ces travaux ont été valorisés par une publication en tant que premier auteur dans *Genetic Resources and Crop Evolution* en 2022.

L'étude a porté sur 133 accessions de vanilliers conservées au CRB Vatel (Roux-Cuvelier et al. 2021), parmi lesquelles 84 *V. planifolia*, 7 *V. × tahitensis*, 2 *Vanilla bahiana*, 4 *Vanilla cribbiana*, 1 *Vanilla helleri*, 2 *Vanilla insignis*, 8 *V. odorata*, 2 *V. pompona*, 4 *Vanilla sotoarenasii* M. Pignal et 19 hybrides interspécifiques. Les relations entre ces individus ont été étudiées à partir de 2004 SNPs filtrés.



## 2. Publication 2 : A genome-wide assessment of the genetic diversity, evolution and relationships with allied species of the clonally propagated crop *Vanilla planifolia* Jacks. ex Andrews

### Auteurs

Félicien Favre<sup>1</sup>, Cyril Jourda<sup>2</sup>, Michel Grisoni<sup>3</sup>, Quentin Piet<sup>2</sup>, Ronan Rivallan<sup>4,5</sup>, Jean-Bernard Dijoux<sup>2</sup>, Jérémy Hascoat<sup>2</sup>, Sandra Lepers-Andrzejewski<sup>6</sup>, Pascale Besse<sup>1</sup> and Carine Charron<sup>2\*</sup>

<sup>1</sup> Université de La Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>2</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>3</sup> CIRAD, UMR PVBMT, Toamasina 501, Madagascar

<sup>4</sup> CIRAD, UMR AGAP, F-34398 Montpellier, France

<sup>5</sup> AGAP, Université de Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

<sup>6</sup> Etablissement Vanille de Tahiti, 98713 Papeete, Tahiti, Polynésie française, France

### Journal de publication

Genetic Resources and Crop Evolution 69, pages 2125–2139, Springer, 2022

<https://doi.org/10.1007/s10722-022-01362-1>


### Mots clés

Amélioration variétale, Domestication, *Vanilla planifolia*, *Vanilla × tahitensis*





# A genome-wide assessment of the genetic diversity, evolution and relationships with allied species of the clonally propagated crop *Vanilla planifolia* Jacks. ex Andrews

Félicien Favre · Cyril Jourda · Michel Grisoni · Quentin Piet ·  
Ronan Rivallan · Jean-Bernard Dijoux · Jérémy Hascoat ·  
Sandra Lepers-Andrzejewski · Pascale Besse · Carine Charron 

Received: 16 September 2021 / Accepted: 18 February 2022  
© The Author(s) 2022

**Abstract** The *Vanilla* genus is a complex taxonomic group characterized by a vegetative reproduction mode combined with intra- and inter-specific hybridizations, and polyploidy events. These factors strongly impact the diversification of the genus and complicate the delimitation of taxa. Among the hundred *Vanilla* species, *Vanilla planifolia* Jacks. ex Andrews and *Vanilla × tahitensis* J. W. Moore are the main cultivated aromatic species. We applied Genotyping-by-Sequencing to explore the genetic diversity

of these two cultivated vanilla species, seven closely related species and nineteen interspecific hybrids. The inter- and intra-specific relationships of 133 vanilla accessions were examined based on 2004 filtered SNPs. Our results showed a strong genetic structuring between the nine species studied, with wild species showing much lower heterozygosity levels than cultivated ones. Moreover, using Bayesian clustering analyses, the kinship of several hybrids could be verified. We evidenced in particular that *Vanilla sotoarenasii* and *Vanilla odorata* C.Presl may be the parental species of *V. × tahitensis*. The analysis of 1129 SNPs for 84 *V. planifolia* accessions showed a clear genetic

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10722-022-01362-1>.

F. Favre · P. Besse  
Université de La Réunion, UMR PVBMT,  
97410 St Pierre, La Réunion, France  
e-mail: felicien.favre@cirad.fr

P. Besse  
e-mail: pascale.besse@univ-reunion.fr

C. Jourda · Q. Piet · J.-B. Dijoux · J. Hascoat ·  
C. Charron (✉)  
CIRAD, UMR PVBMT, 97410 St Pierre, La Réunion,  
France  
e-mail: carine.charron@cirad.fr

Q. Piet  
e-mail: quentin.piet@cirad.fr

J.-B. Dijoux  
e-mail: jean-bernard.dijoux@cirad.fr

J. Hascoat  
e-mail: jeremy.hascoat@cirad.fr

M. Grisoni  
CIRAD, UMR PVBMT, 501 Toamasina, Madagascar  
e-mail: michel.grisoni@cirad.fr

R. Rivallan  
CIRAD, UMR AGAP, 34398 Montpellier, France  
e-mail: ronan.rivallan@cirad.fr

R. Rivallan  
AGAP, Université de Montpellier, CIRAD, INRAE,  
Institut Agro, Montpellier, France

S. Lepers-Andrzejewski  
Etablissement Vanille de Tahiti, 98713 Papeete, Tahiti,  
French Polynesia  
e-mail: sandra.lepers@vanilledetahiti.pf

demarcation between the vegetatively propagated traditional vanilla cultivars compared to the accessions derived from sexual reproduction, and a higher genetic diversity and lower heterozygosity of the latter ( $H_o = 0.206$ ) compared to the former ( $H_o = 0.362$ ). Our data are consistent with a single-step domestication for *V. planifolia* in accordance with the recent history of its cultivation. It also opens avenues to breed new *V. planifolia* varieties adapted to biotic and abiotic constraints and to reduce mutational load induced by clonal propagation.

**Keywords** Breeding · Domestication ·  
*Vanilla × tahitensis* · *Vanilla planifolia*

## Introduction

The genus *Vanilla* Plumier ex Miller belongs to the Orchidaceae family and is composed of about 120 species, among which 18 (Portères 1954) to 35 (Soto Arenas 2003) are considered to bear aromatic fruits. The main cultivated species is *Vanilla planifolia* Jacks. ex Andrews that contributes to more than 95% of the world's vanilla production. Native to Mesoamerican tropical forests, *V. planifolia* has been vegetatively propagated and cultivated in the Eastern coast of Mexico since the mid eighteenth century in response to the growing demand for vanilla pods in Europe. Vanilla cuttings were subsequently transferred into European botanical gardens, then reached the Indian Ocean region where no natural pollinator was present (Bory et al. 2008b; Lubinsky et al. 2010). The discovery in 1841 of an easy manual pollination technique by Edmond Albius in Reunion Island has led to the fast diffusion of *V. planifolia* into the south west Indian ocean region (Madagascar, Reunion Island, Comoros). *Vanilla × tahitensis* J.W.Moore, is mostly cultivated in French Polynesia and Papua New Guinea. It is supposed to have been introduced to Tahiti Island from the Philippines in 1848 (Constantin and Bois 1915). However, *V. × tahitensis* is no longer found in the wild and its origin has long been debated. Based on morphological characteristics, a hybrid origin between *V. planifolia* and *Vanilla pompona* Schiede (Portères 1954) or *Vanilla odorata* C.Presl (Portères 1954; Soto Arenas 1999) was suggested. A genetic analysis using the nuclear Internal Transcribed Spacer (ITS) and plastid DNA sequences

rather suggested a hybrid origin between *V. planifolia* as the maternal parent and *V. odorata* as the paternal one (Lubinsky et al. 2008). Other aromatic species are cultivated or harvested in the wild at small scales in some localities such as *V. pompona* in the French Caribbean islands and *V. odorata* in Central and South America (Soto Arenas 1999).

Domestication, according to Martínez-Ainsworth and Tenaillon (2016), can be described as a set of consecutive stages that begins with the onset of domestication followed by an increase in the frequency of a set of desirable traits. McKey et al. (2010) highlighted the lack of knowledge on the evolutionary ecology of domesticated plants that are clonally-propagated. Vanilla is no exception and the impact of domestication on the genetics of cultivated *Vanilla* has received little attention. From 1793 to 1875, five introduction events of *V. planifolia* cuttings into Reunion Island were reported, but only one introduction in 1822 by Marchant from Europe is supposed to have been successful and to be at the origin of vanilla cultivation in Reunion Island (Bory et al. 2008b). From a historical perspective and given the very limited number of introductions, “single-step domestication” (i.e. identification of interesting genotype and direct clonal propagation) might be the rule in *V. planifolia*, which would generate a crop that remains close to wild progenitors (Zohary 2004). Low levels of genetic diversity are therefore expected in *V. planifolia* in cultivation areas such as Reunion Island, in accordance with the vegetative mode of multiplication of vanilla vines, and their recent introduction in the Indian Ocean region. Random amplified polymorphic DNA (RAPD) (Besse et al. 2004), amplified fragment length polymorphism (AFLP) (Bory et al. 2008c) and microsatellite (SSR) (Bory et al. 2008a) markers succeeded to discriminate the species and confirmed the genetic uniformity of most *V. planifolia* cultivars in the Indian Ocean and other cultivation areas. AFLP patterns of variation suggested that *V. planifolia* has evolved in introduction areas by the accumulation of point mutations through vegetative multiplication. However, these markers, and even those based on methylation patterns (MSAP) (Gigant et al. 2011), have failed to identify clusters of intraspecific genetic diversity congruent with the phenotypic variations described in cultivation (Bory et al. 2008b, c; Gigant et al. 2011). On the contrary, based on AFLP studies and linkage mapping, varieties described in

*V. × tahitensis* were shown to result mainly from self-pollination or full sib crosses of plants belonging to the most ancient ‘Tahiti’ morphotype, with subsequent heterozygous selection (Lepers-Andrzejewski et al. 2012).

Genotyping-by-sequencing (GBS) is able to generate thousands of Single Nucleotide Polymorphisms (SNPs) markers by applying massively parallel sequencing and multiplexing methods (Elshire et al. 2011). GBS-generated SNP markers are useful to explore the genetic diversity and structure of population in order to better define phylogeny, adaptation of plants to their environment or domestication (Favre et al. 2021). Previous studies validated the efficiency of GBS to characterize *Vanilla* genetic diversity and to identify hybrids (Hu et al. 2019; Alomia et al. 2021). Herein, we developed SNP markers derived from GBS data to study intraspecific diversity and enlighten the evolutionary history of cultivated vanilla in its introduction areas. Our study focused on genetic diversity of *Vanilla* resources conserved in the Biological Resources Centers (BRCs) Vatel and Etablissement Vanille de Tahiti (EVT) (Roux-Cuvelier et al. 2021): cultivated *V. planifolia* and *V. × tahitensis* species, seven wild relatives originating from tropical America (Soto Arenas 2003) and selfed-progenies of *V. planifolia* and interspecific hybrids. This well characterized germplasm constitutes a material of choice to assess the genome-wide genetic diversity, the impact of domestication processes and breeding on genetic diversity levels in cultivated *Vanilla* compared to wild genotypes.

## Materials and methods

### Plant material and DNA extraction

A panel of 137 *Vanilla* spp. accessions was used for GBS sequencing, including (i) the two cultivated species *V. planifolia* (88 accessions) and *V. × tahitensis* (7 accessions), (ii) 7 closely related species *V. odorata* (8 accessions), *Vanilla cribbiana* Soto Arenas (4 accessions), *V. sotoarenasii* (4 accessions), *Vanilla insignis* Ames (2 accessions), *V. pompona* (2 accessions), *Vanilla bahiana* Hoehne (2 accessions), *Vanilla helleri* A.D. Hawkes (1 accession) and (iii) 19 interspecific hybrids (Supporting Information Table S1). All accessions were grown in

shade house or in vitro in Reunion Island and French Polynesia and were selected from a large collection of over 700 accessions conserved in the French BRCs Vatel (CIRAD, Reunion Island) and EVT (Raiatea) (Roux-Cuvelier et al. 2021). Accessions were selected in order to maximize variability for origin, variety and ploidy level and better evaluate the species diversity. Traditional cultivars of *V. planifolia* and *V. × tahitensis* collected in fields were classified as vegetatively (asexually) propagated clones or ‘cuttings’. Accessions obtained by sexual reproduction (selfed-progenies or intra-specific hybrids) were classified as ‘seedlings’. All accessions were clonally propagated by cuttings or micro cuttings in vitro for their conservation in BRCs Vatel and EVT (Supporting information S1 and S2). High molecular weight DNA of each accession was extracted from 25 mg of lyophilized young leaves using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Genomic DNA was quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and normalized at 50 ng/μL. DNA homogeneity and quality was assessed by enzymatic digestion with *HindIII* (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and run in a 2% agarose gel.

### Library preparation and sequencing

Library preparation was performed by the Regional genotyping technology platform (UMR AGAP, CIRAD, Montpellier, France) as described by Elshire et al. (2011). Two kinds of adapters are used for constructing GBS libraries, a common adapter and a sample-specific barcode adapter. Both adapters are designed to fit with Illumina sequencing. Adapters were mixed together in a 1:1 ratio and plated into two 96-well plates, so that each well contained one specific barcode. Extracted DNA was added into the 96-well plates and digested with *PstI* methylation-sensitive restriction enzyme (New England Biolabs, Ipswich, Massachusetts, USA). Adapters were ligated to the ends of the DNA fragments using a T4 ligase (New England Biolabs, Ipswich, Massachusetts, USA). Samples were then pooled together, amplified by Polymerase Chain Reaction (PCR) and purified to remove unreacted adapters. The GBS library was sequenced on Illumina HiSeq3000 sequencer (Illumina Inc., San



Diego, California, USA) with DNA-seq single-read protocol at the GeT-PlaGe platform (INRAE, Toulouse, France).

#### Sequence analysis and SNP calling

Sequence quality was checked with FastQC (Andrews 2010). Low-quality reads, reads with uncalled bases and reads with Illumina adapter sequences were removed using the Cutadapt software (Martin 2011). The remaining reads were assigned to each sample using the GBS barcode splitter tool (<https://sourceforge.net/projects/gbsbarcode/>). Demultiplexed sequences were trimmed to 140 bp to normalize the length between individuals. SNP calling was performed using STACKS de novo pipeline (Catchen et al. 2013) and identified SNPs were converted into Variant Call Format file (VCF; Danecek et al. 2011). Low-quality SNPs were filtered out with vcfr package 1.11.0 (Knaus and Grünwald 2017) from Rstudio (version 3.6.3) (R Development Core Team 2010) and using successive filters: minimum minor allele frequency < 10%, missing data per site > 30% and up to 3 SNPs per locus. SNPs with an allele frequency below 10% were discarded because these very rare variants probably resulted from genotyping errors, while retaining rare alleles that are associated to under-represented samples in the dataset. The *V. planifolia* accessions were used to study structure within such a clonally propagated vanilla, and only polymorphic SNP markers were kept. Loci with unfiltered SNP markers and filtered SNPs were mapped over the *Daphna V. planifolia* chromosomes (Hasing et al. 2020) to check SNP distribution and density. The genotyping data sets were converted into biallelic tables by vcfr package (Knaus and Grünwald 2017), which were used for both phylogenetic relationships and population structure analysis.

#### Genetic diversity and phylogenetic analysis based on SNP markers

The number of effective alleles ( $N_e$ ), Shannon's information index ( $I$ ), observed heterozygosity ( $H_o$ ) and percentage of polymorphic loci ( $P$ ) were calculated using GenAIEx (version 6.502) (Peakall and Smouse 2012). The values were compared across the nine species and interspecific hybrids using the complete genotyping data set, and across different types of *V.*

*planifolia* cultivars. From the complete genotyping dataset, a dissimilarity coefficient was calculated with *DarWIN* software (Perrier and Jacquemoud-Collet 2006) using the simple matching index (Sokal and Michener 1958):  $d_{ij} = 1 - \frac{1}{L} \sum_{l=1}^L \frac{m_l}{\pi}$  where  $d_{ij}$  is the dissimilarity between units  $i$  and  $j$ ;  $L$  the number of loci;  $m_l$  the number of matching alleles for locus  $l$ ; and  $\pi$  the ploidy. Distance trees were constructed from 1000 bootstrap replicates using the Unweighted Neighbor-Joining method (Saitou and Nei 1987). Trees were then converted into *Phylip* file and plotted with *FigTree* software (Rambaut 2006).

#### Population structure analyses

Principal coordinates analyses (PCoA) were performed using the complete genotyping data set and the *V. planifolia* data set with GenAIEx (version 6.502) (Peakall and Smouse 2012). The population structure was analyzed to identify clusters of genetically related individuals using the Bayesian clustering method implemented in *STRUCTURE* (version 2.3.4) (Pritchard et al. 2000). The *STRUCTURE* analysis was first performed between *V. odorata*, *V. planifolia*, *V. pompona*, *V. × tahitensis* and hybrids, and then between all the *V. planifolia* accessions. The admixture model of *STRUCTURE* was chosen on the assumption that each individual had ancestry from one or more of  $K$  genetically distinct sources. Ten independent runs were performed for each  $K$  from  $K = 1$  to  $K = 10$ , with a burn-in period of 10,000 and 100,000 Markov-chain Monte Carlo (MCMC) iterations after burn-in. The best number of  $K$  was chosen with the  $\Delta K$  method (Evanno et al. 2005) by running the *STRUCTURE HARVESTER* (Earl and vonHoldt 2012). For *STRUCTURE* analysis within *V. planifolia*, accessions were assigned to one cluster if their probability of belonging to this cluster is higher than 60%.

## Results

#### Sequencing and SNP calling

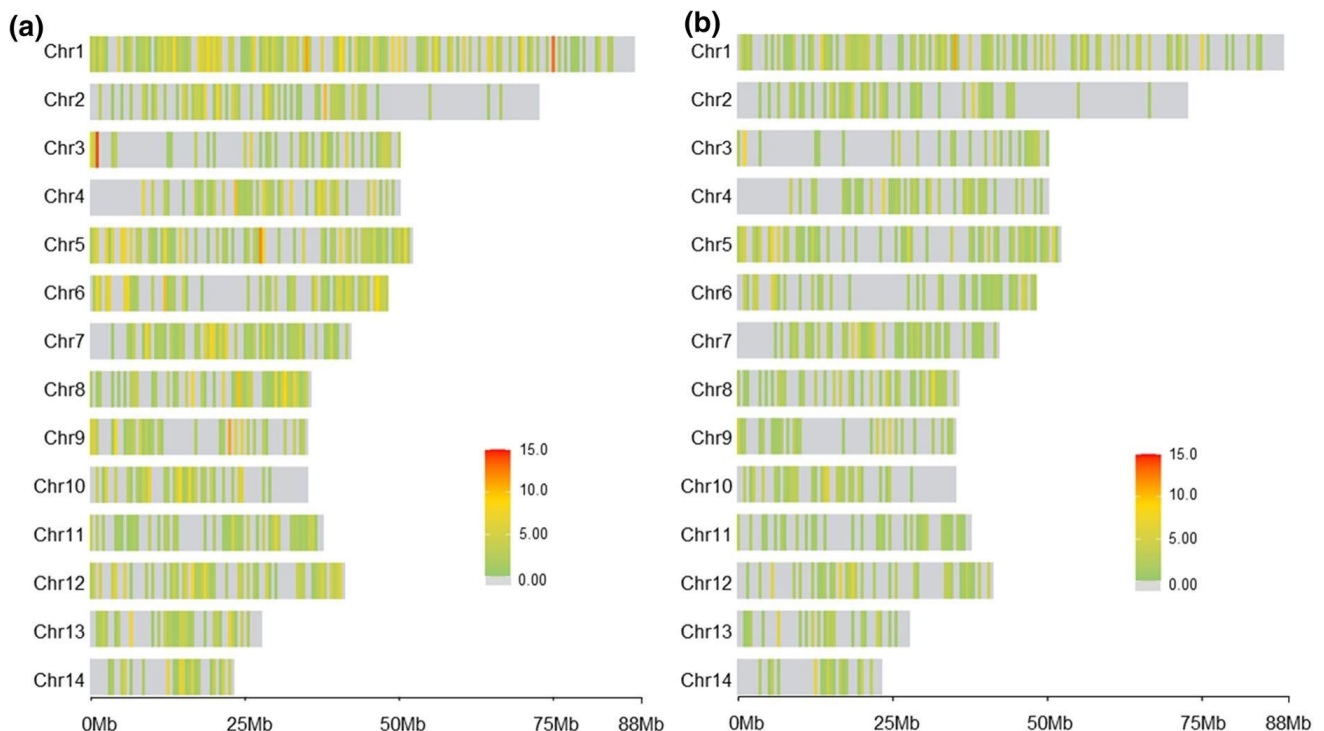
The sequencing of GBS libraries resulted in 1,143,846,935 single-reads of 150 bp for 137 *Vanilla* accessions (Supporting information Tables S1, S2,

S3 and S4). After the cleaning step, 554,565,581 reads of 140 bp were demultiplexed for each individual (51.5% of the reads were discarded). CR0026, CR0151 and CR2564 samples, with a low sequencing depth (0; 6377 and 2657 reads, respectively), were removed from the study. The de novo pipeline reconstructed 828,215 loci including 81,987 identified as biallelic (10.3%) with 225,857 raw SNPs. Among the 80,203 loci with a sequencing depth  $\geq 5$  reads, 32% mapped on the 14 chromosomes of the published *V. planifolia* Daphna cv genome (Hasing et al. 2020) (Supporting information Fig. S1). The 194,625 SNPs distributed on these loci were filtered out based on: minimum minor allele frequency (MAF)  $< 10\%$ , missing data per site  $> 30\%$ , and a maximum of 3 SNPs per locus. The data set obtained consisted of 2040 high-quality filtered SNPs. The accession *V. planifolia* CR0844 was removed due to a missing data rate  $> 45\%$ . The final data set consisted of 133 genotyped vanilla samples using 2004 filtered SNPs, with a mean heterozygosity of 19.4% ( $\pm 16.7\%$ ) and a mean missing data rate of 19.5% ( $\pm 6.3\%$ ). A specific subset of 84 *V. planifolia* individuals using 1129

SNPs was produced from the complete genotyping matrix using the same filters, with a mean heterozygosity of 43.6% ( $\pm 14.9\%$ ) and a mean missing data rate of 15.2% ( $\pm 7.5\%$ ). Among the 2004 and 1129 SNPs, 1916 (95.6%) and 1081 (95.8%) mapped onto the 14 chromosomes of the *V. planifolia* cv Daphna genome. For the 2004 SNP matrix, the density of SNPs per chromosome ranged from one SNP every 527 kb (chromosome 2) to one SNP every 230 kb (chromosome 1), with an average of one SNP every 359 kb (Fig. 1a). For the 1129 SNP matrix, the lowest density of SNPs was detected on chromosome 3 (one SNP every 1024 kb) and the highest on chromosome 1 (one SNP every 429 kb), with an average of one SNP every 641 kb (Fig. 1b).

#### Genetic relationships across species within the genus *Vanilla*

The unweighted Neighbor-Joining (NJ) tree built with the complete genotyping matrix revealed two major groups statistically supported with bootstrap values equal to 1. On one side of the tree, a group



**Fig. 1** Genotyping-By-Sequencing SNP distribution and density on *V. planifolia* cv Daphna chromosomes. **a** From 2004 filtered SNPs dataset. **b** From 1129 filtered SNPs dataset. Horizontal axis displays the chromosome length. The density scale

indicates the number of SNPs within 500 Kb window size. The plot shows the distribution of GBS SNPs across the 14 chromosomes



**Table 1** Diversity indexes in *Vanilla* species using 2004 SNPs identified across all species

Species	N	Ne	(SE)	I	(SE)	Ho	(SE)	P (%)
<i>V. planifolia</i>	84	1.522	(0.011)	0.374	(0.008)	0.287	(0.001)	56.34
Vegetative	38	1.515	(0.011)	0.369	(0.008)	0.362	(0.008)	55.24
Sexual	39	1.498	(0.010)	0.368	(0.007)	0.206	(0.005)	56.24
Unknown	7	1.483	(0.010)	0.358	(0.007)	0.350	(0.008)	54.69
<i>V. x tahitensis</i>	7	1.192	(0.013)	0.205	(0.007)	0.239	(0.009)	30.89
<i>V. bahiana</i>	2	0.782	(0.009)	0.003	(0.001)	0.002	(0.001)	0.45
<i>V. cribbiana</i>	4	0.825	(0.009)	0.012	(0.002)	0.001	(0.000)	2.10
<i>V. helleri</i>	1	0.746	(0.010)	0.001	(0.001)	0.002	(0.001)	0.20
<i>V. insignis</i>	2	0.923	(0.008)	0.031	(0.003)	0.018	(0.002)	5.04
<i>V. odorata</i>	8	1.068	(0.007)	0.110	(0.005)	0.002	(0.001)	22.36
<i>V. pompona</i>	2	0.679	(0.011)	0.004	(0.001)	0.002	(0.001)	0.60
<i>V. sotoarenasii</i>	4	1.056	(0.009)	0.094	(0.005)	0.031	(0.002)	15.57
Interspecific crossing	19	1.664	(0.007)	0.551	(0.004)	0.246	(0.004)	96.56

*N* number of accessions, *Ne* number of effective alleles, *I* Shannon's information index, *Ho* observed heterozygosity, *SE* standard error, *P* percentage of polymorphic SNPs (%). Vegetative, sexual and unknown indicate the reproduction mode of the *V. planifolia* accessions

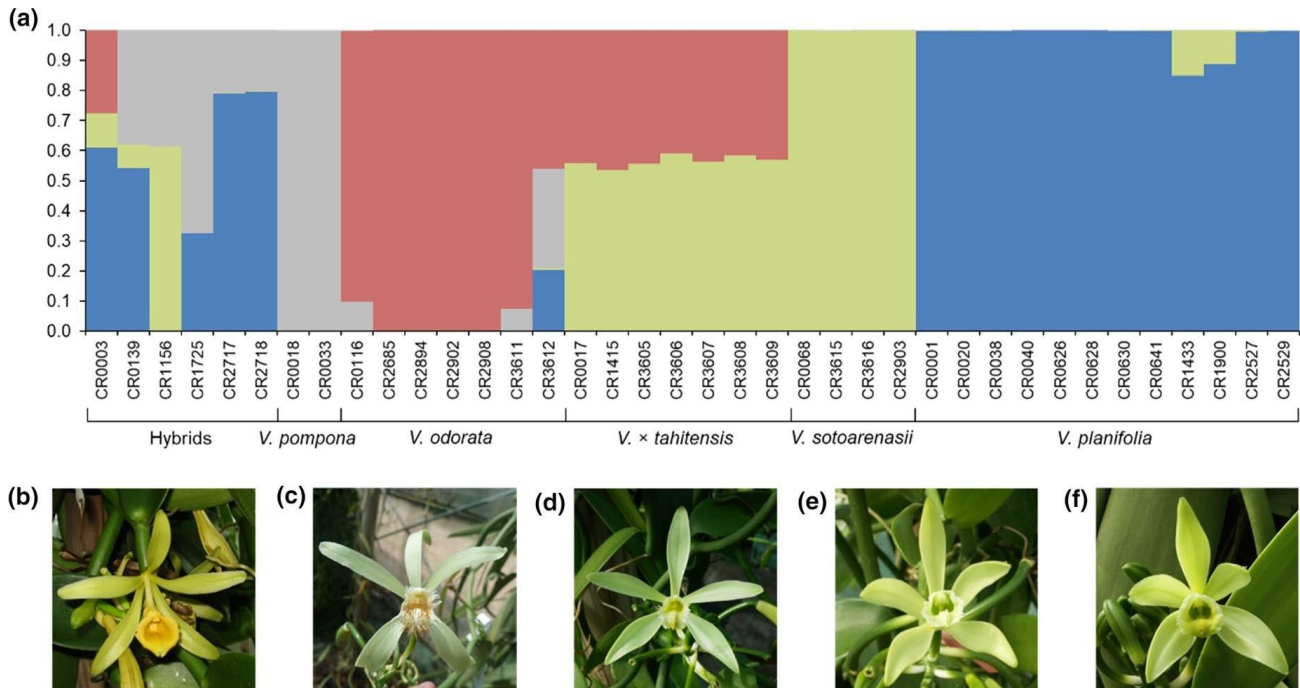
genotyping matrix (Table 1). *Ho* values were of similar range in the cultivated species, *V. planifolia* ( $0.287 \pm 0.001$ ) and *V. x tahitensis* ( $0.239 \pm 0.009$ ), and on average a hundred times higher than the *Ho* values observed in wild relative species (ranging from  $0.001 \pm 0.000$  in *V. cribbiana* to  $0.031 \pm 0.002$  in *V. sotoarenasii*). The mean *Ho* of vegetatively propagated *V. planifolia* ( $0.362 \pm 0.008$ ) was significantly higher than the mean *Ho* in selfed-progenies ( $0.206 \pm 0.005$ ). Hybrids revealed *Ho* levels ( $0.246 \pm 0.004$ ) close to those observed in cultivated vanilla. Shannon's information index (*I*) was highest for interspecific hybrids ( $0.551 \pm 0.551$ ), followed by *V. planifolia* and *V. x tahitensis* ( $0.374 \pm 0.008$  and  $0.205 \pm 0.007$  respectively). Among the wild species, *I* values ranged from  $0.001 \pm 0.001$  in *V. helleri* to  $0.110 \pm 0.005$  in *V. odorata*. The percentage of polymorphic SNPs was higher in the cultivated vanillas and in interspecific hybrids (> 30%), compared to wild relatives (< 23%).

#### Genetic relationships between *V. planifolia* and several of its hybrids

In the light of our GBS-generated SNPs we explored the relationships between several species and derived interspecific hybrids. A Bayesian clustering analysis based on 2004 SNPs (Fig. 3a) was performed on a reduced dataset including six hybrids, two *V. pompona* (Fig. 3b), seven *V. odorata* (Fig. 3c), seven *V. x tahitensis* (Fig. 3d), four *V. sotoarenasii* (Fig. 3e), and a representative subset of 12 *V.*

*planifolia* accessions (Fig. 3f) selected among subgroups identified in Fig. 2. The estimated likelihood was greatest for  $K = 4$ , suggesting the presence of 4 clusters corresponding to the 4 species *V. odorata*, *V. planifolia*, *V. pompona* and *V. sotoarenasii*. For supposed first-generation hybrids, we were expecting a probability of 0.50 for each parental genome assignment. The seven *V. x tahitensis* accessions had in average 56.5% of their SNPs attributed to *V. sotoarenasii* and 43.5% to *V. odorata* (X-squared = 1.69,  $df = 1$ ,  $p$ -value = 0.1936), confirming the interspecific hybrid status of this species. The accession CR1415 (CR0017 x CR0017) had similar proportions in its genome (53.6/46.4%) than its parent *V. x tahitensis* CR0017 (55.9/44.1%). For the *V. pompona* x *V. sotoarenasii* hybrid CR1156, 61.4% of the SNPs were assigned to *V. sotoarenasii* and 38.6% to *V. pompona*. These proportions did not fit the hypothesis of first-generation hybrid (X-squared = 5.20,  $df = 1$ ,  $p$ -value = 0.0226). The SNPs of hybrid CR1725 were assigned to *V. pompona* (67.6%) and *V. planifolia* (32.4%) which was also inconsistent with hypothesis of 0.50/0.50 distribution of parental genomes (X-squared = 12.39,  $df = 1$ ,  $p$ -value = 0.0004). The SNPs of hybrid CR0139 were assigned to *V. planifolia* (54.2%), *V. pompona* (38.0%) and *V. sotoarenasii* (7.80%). Therefore, these three first-generation hybrids showed parental inheritances of SNPs significantly deviating from the hypothetical ratio of 0.5/0.5. For back-cross hybrids, we were expecting 0.75/0.25 distribution of the parental genomes. The hybrid CR2717 (*V. planifolia* x *V. pompona*)





**Fig. 3** Origin of *V. × tahitensis* shown by bayesian clustering and comparison of morphological traits. **a** Population structure of 38 vanilla accessions using 2004 significant SNPs. Colours represent different assigned clusters. The X-axis provides accession and species names and the y-axis provides the proba-

bility of each accession belonging to the assigned cluster. Front view of entire flowers of **b** *V. pompona* CR0018, **c** *V. odorata* CR0116, **d** *V. × tahitensis* CR0017, **e** *V. sotoarenasii* CR0068 and **f** *V. planifolia* CR0040

$\times V. planifolia$ ) had 79.1% of its genome assigned to *V. planifolia* and 20.8% to *V. pompona*, which was consistent with back-cross ratio hypothesis ( $X^2 = 0.93$ ,  $df = 1$ ,  $p\text{-value} = 0.3347$ ). The hybrid CR2718, which is a cutting of CR2717 showed similar results (79.7% assigned to *V. planifolia* and 20.2% to *V. pompona*, and  $X^2 = 1.22$ ,  $df = 1$ ,  $p\text{-value} = 0.2699$ ). The hybrid CR0003, that comes from a *V. planifolia*  $\times$  *V. × tahitensis* cross, revealed ancestry from *V. planifolia* (60.9%), *V. odorata* (27.6%) and *V. sotoarenasii* (11.5%) as expected, but the proportions (60.9/27.6/11.5%) were inconsistent with the 50/25/25% proportions expected for such a cross ( $X^2 = 9.94$ ,  $df = 2$ ,  $p\text{-value} = 0.0070$ ).

#### Intraspecific genetic structuration within *V. planifolia*

The *V. planifolia* accessions derived from sexual reproduction showed some structuration in the NJ tree (Fig. 2) with subgroups supported by bootstrap values equal to 1, whereas no genetic structuration was supported by bootstraps values in vegetatively propagated accessions. Among the accessions with

unknown reproduction mode, CR2093, ‘Colibri’ CR2687, CR2100 and CR2102 branched within cuttings, while CR0510, CR0196 and CR0628 (‘Aiguille’) branched within seedlings. The tetraploids ‘Grosse Vanille’ CR0802 and CR0641 and the triploids ‘Sterile’ CR0645 and CR0630, were grouped together with cuttings without significant structuration between these accessions. The cultivar ‘Petite Mexique’ CR0632, supposed to derive from clonal propagation, was branched with accessions derived from sexual reproduction.

The first and second coordinates of the PCoA of *V. planifolia* accessions explained 13.92% and 8.29% of the genetic variability, respectively (Supporting information Fig. S3a). These values were lower than those observed from the complete dataset PCoA (Supporting information Fig. S2), indicating a low structuration within the *V. planifolia* group. Individuals obtained by vegetative propagation were clustered along the first axis, while individuals derived from sexual reproduction were much more dispersed throughout the plan. No structuration related to geographic origin or the ploidy level of cultivars could

be evidenced by PCoA (Supporting information Fig. S3b and c).

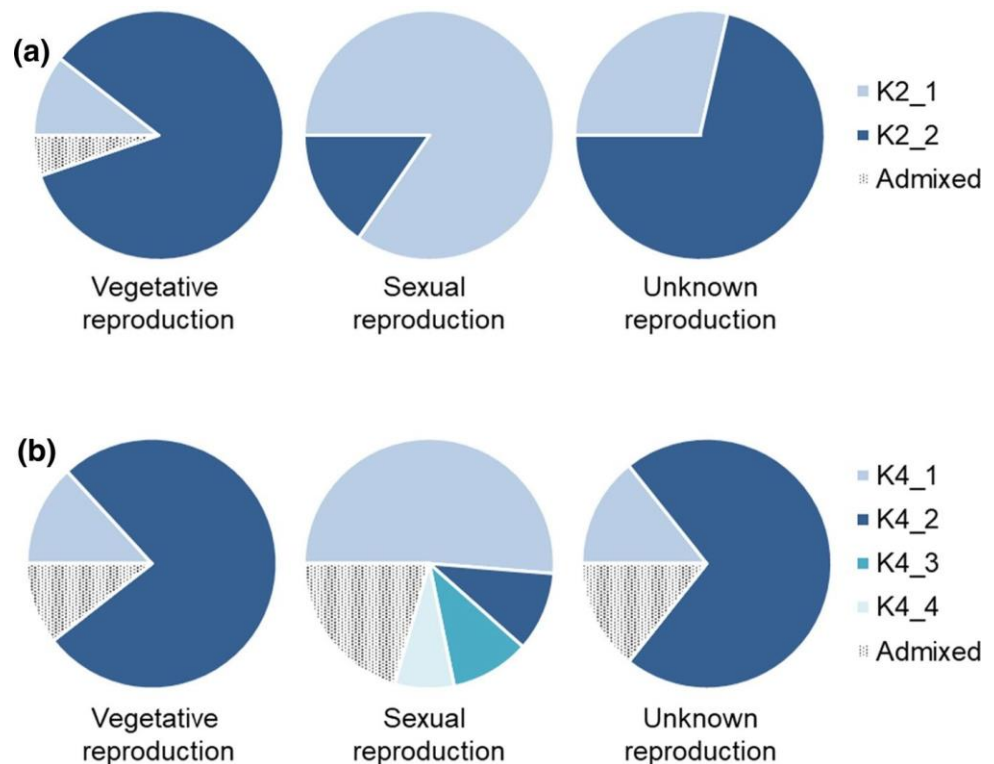
Genetic structure was explored by a Bayesian clustering analysis with 84 *V. planifolia* accessions using the 1129 SNP matrix (Supporting information Fig. S4a). The estimated likelihood was the greatest for  $K = 2$ ,  $K = 3$  and  $K = 4$ , suggesting the presence of 2 to 4 genetic clusters (Supporting information Fig. S4b and c). Accessions were assigned to one specific cluster if their probability of belonging to this cluster was higher than 60%, in order to facilitate interpretation based on a majority rule basis. If not, they were considered admixed. For  $K = 2$ , the majority (84.21%) of the accessions derived from vegetative propagation were assigned to cluster K2\_2, whereas 84.62% of the accessions derived by sexual reproduction were assigned to cluster K2\_1 (Fig. 4a). For  $K = 4$ , 76.31% of the accessions derived from vegetative propagation were assigned to the cluster K4\_2, whereas admixture and a greater diversity of clusters were observed in accessions derived from sexual reproduction (Fig. 4b). Among unknown accessions, five were assigned to the 'vegetative' clusters (K2\_2 and K4\_2), one to the 'sexual' clusters (K2\_1 and K4\_1), and one appeared admixed for  $K = 4$  (Supporting information Fig. S4a).

## Discussion

Our study confirmed that GBS is a powerful genomic tool for the identification of highly informative SNPs for the study of the inter- and intra-specific genetic diversity in the *Vanilla* genus. Here, we applied GBS to explore the genetic relationships and genetic structure of the two cultivated *Vanilla* species, seven closely related species, and 19 interspecific hybrids from the well documented vanilla collections maintained ex situ at BRCs Vatel and Vanille de Tahiti. GBS genotyping yielded 2004 high quality filtered SNPs for the *Vanilla* genus, among those a subset of 1129 SNPs was used for *V. planifolia*. The majority of the SNPs (95%) successfully mapped on the published *V. planifolia* cv Daphna genome (Hasing et al. 2020), except for the extremity of the chromosome 2 not covered with GBS-generated SNPs. This might result from erroneous chromosome 2 assembly in the Daphna genome or a richness in repetitive elements and low-complexity regions in this chromosome part.

Our GBS data supported the current taxonomy for most of the species studied (Bouetard et al. 2010). As expected, *V. planifolia* and wild relative species were strongly separated. Hybrids from crosses between *V. planifolia* and a wild species were in intermediate

**Fig. 4** Genetic structure of 84 *V. planifolia* using 1129 informative SNPs. **a** Proportion of accessions assigned to one of the two clusters determined by STRUCTURE for  $K = 2$ , according to mode of reproduction. **b** Proportion of accessions assigned to one of the four clusters determined by STRUCTURE for  $K = 4$ , according to mode of reproduction. The dotted ground indicates accessions classified as admixed



position on the NJ tree or close to the *V. planifolia* parent. The *V. pompona* group was the most distant from the *V. planifolia* group in agreement with plastid DNA analysis (Bouetard et al. 2010). Accessions attributed to the species *V. odorata* formed a large and structured group, except CR3612 which probably has a hybrid origin. Indeed, morphology of flowers and fruits obtained recently for CR3612 suggested a kinship with a species close to *V. pompona*, while the long and narrow leaves were related to *V. odorata* (Supporting information Fig. S5a). *V. odorata* cf. CR2686 was probably misidentified and should rather be classified within *V. insignis*, since it is genetically close to *V. insignis* CR2688 (Fig. 2) and showed broad leaves and rough stem typical of this species. This would make *V. insignis* a very close but nevertheless distinct group (bootstrap = 1) from *V. odorata*. Accession CR0087 identified as *V. insignis* cf. showed proximity in the NJ tree to the *V. bahiana* group and is very close to its hybrid with *V. bahiana*. The flowers of CR0087 differ from those of *V. bahiana* by their slightly larger size but above all by the presence of very developed orange papillae on the lip of the labellum (Supporting information Fig. S5b). The *V. helleri* accession CR3614, nested within the *V. odorata* group, could be conspecific to this species. The *V. × tahitensis* accessions formed a large and structured group close to, but clearly distinct from its hybrids and from the *V. planifolia* group, contrary to what was observed with plastid DNA analysis (Bouetard et al. 2010).

The *V. sotoarenasii* group was closely related but distinct from *V. planifolia* (bootstrap value = 1), as previously established (Bory et al. 2008c; Bouetard et al. 2010; Azofeifa-Bolaños et al. 2017). *V. sotoarenasii* did not appear like a hybrid in the STRUCTURE analyses, but like a distinct genetic group with specific SNPs. In a commendable effort to limit specific inflation in the genus *Vanilla*, it was recently proposed that *V. sotoarenasii* is conspecific to *V. planifolia* (Karremans et al. 2020). The authors based their proposal on the absence of genetic differences between the two species and argue that the morphological variations of vanilla vines in the Cahuita area (Costa Rica), where large populations of *V. sotoarenasii* have been described (Azofeifa-Bolaños et al. 2017), do not allow to distinguish one species from the other. These two arguments do not hold. First, as indicated previously

(Azofeifa-Bolaños et al. 2017), a small (2 nucleotides) but steady difference separates the ITS sequences of *V. sotoarenasii* from those of *V. planifolia*. The GBS data presented here, based on 2004 SNPs, provided clear evidence of genetic differentiation between *V. planifolia* and *V. sotoarenasii*. Second, the argument that the morphological variability of *V. planifolia* contains that of *V. sotoarenasii* is not supported by any quantitative data. On the contrary, the photograph of numerous fruits produced by the authors (Karremans et al. 2020) to show the wide variability in *V. sotoarenasii* demonstrated that the fruits of *V. sotoarenasii* were always between 9 and 15 cm in length, and indehiscent, which is much less than the 21 cm average length of *V. planifolia* fruits, that are moreover very predominantly dehiscent (Díaz-Bautista et al. 2018). In addition, the variations in shape and color of the fruits are much more certainly due to different stages of maturity, and incomplete natural pollinations, than to phenotypic plasticity that remains to be demonstrated. Similarly, variations in leaf shape and color are common in vanilla plants in relation to biotic and abiotic factors during their growth and have little taxonomic value (Soto Arenas and Cribb 2013). The other objection of these authors that the flower size measurements of *V. sotoarenasii* (Azofeifa-Bolaños et al. 2017) would be biased by the fact that they were made on an accession grown under controlled conditions is not supported by any measurement of variability in wild populations. Unlike vegetative organs, the morphology of reproductive organs is little impacted by environmental conditions and therefore has a better taxonomic value. On the other hand, as mentioned previously (Azofeifa-Bolaños et al. 2017) and documented recently in *V. pompona* (Watteyn et al. 2022) flower size is, among others, an important trait affecting reproductive isolation. In the case of difficult groups such as the genus *Vanilla*, alpha-taxonomy is often not very discriminating or even risky, and an integrative taxonomy approach (Andriamihaja et al. 2022) allows more effectively to dissect the relationships between closely related taxa. In this perspective, the bundle of genetic, morphologic and ecologic arguments clearly plead in favor of the recognition of *V. sotoarenasii* as a valid species which might have recently evolved from *V. planifolia* by geographic and/or reproductive isolation. However,



the hypothesis of a hybrid origin of *V. sotoarenasii* involving *V. planifolia* and a species not included in our study, cannot be completely ruled out.

Bayesian clustering analyses based on GBS-generated SNPs proved to be a powerful tool to assess parental genetic contributions of hybrids, in case of lack of information on a genetic resource, mislabelling or uncontrolled pollination. In this study, we confirmed, or infirmed the taxonomic position and kinship for several species and hybrids. However, some first-generation hybrids did not show the expected 50/50 parental assignment probabilities. Most of the hybrids assessed have *V. pompona* as parental species, which is known to have a much larger genome than *V. planifolia* ( $2C = 8.18$  to  $10.72$  pg) (Bory 2007). The observed deviations could therefore be explained by parental genome complexity. For instance, molecular cytogenetics studies on polyploid sugarcane cultivars, deriving from a few interspecific hybridization events performed a century ago by breeders, highlighted an uneven contribution of each parental genome, with 75–85% of their chromosome originating from one parental species (*S. officinarum*) and 15–25% from the other parent (*S. spontaneum*), with some chromosomes derived from interspecific recombination (Piperidis and D'Hont 2020). A recent GBS study (Alomia et al. 2021) also included some *V. planifolia* x *V. pompona* hybrids, but the parental contributions were unfortunately not quantified, preventing any comparison. Nevertheless, GBS based Bayesian clustering enabled us to undoubtedly identify the parental origin of all the hybrids studied. Our results open important questions about possible genome rearrangement in interspecific crosses between relatively distant species in the genus *Vanilla*, that will need to be addressed further.

Previous assessment of *V. × tahitensis* origin, based on nuclear (ITS, GBS) and plastid (rbcL) loci, suggested a hybrid origin between *V. planifolia* and *V. odorata* (Lubinsky et al. 2008; Hasing et al. 2020; Alomia et al. 2021). Our GBS data confirmed at the genomic level (whole genome scale) the hybrid origin of Tahitian *Vanilla* evidenced in Lubinsky et al. (2008) study, and the proportions of genetic parental contributions (close to 0.50 for all accessions) were compatible with the hypothesis of a first-generation hybrid. However, *V. sotoarenasii* material was missing in Lubinsky et al.'s work. We contributed to clarify the question of the origin of *V. × tahitensis*

by including for the first time well characterized *V. sotoarenasii* accessions and genomic data. According to our analysis, *V. odorata* is indeed one parent, but the second parent is closer to *V. sotoarenasii* than to *V. planifolia* as previously stated. The flower morphology of *V. × tahitensis* (Fig. 3c), with traits that are close to *V. odorata* (Fig. 3b) and to *V. sotoarenasii* (Karremans et al. 2020), supports this observation (Fig. 3d). Alomia et al. (2021) identified, by GBS, *V. planifolia* wild accessions from Belize (referred to as Type 2) as the possible true parent of *V. × tahitensis*. These accessions appeared very close genetically to *V. sotoarenasii* (Alomia et al. 2021). This observation reinforces our hypothesis and we suggest that the Type 2 *V. planifolia* from Belize in Alomia et al. (2021) could possibly be *V. sotoarenasii*. Recent data from Chambers et al. (2021) indeed showed that Type 2 *V. planifolia* and *V. sotoarenasii* cannot be separated genetically in PCA and Structure analyses. A morphological characterization of these Belize accessions could also allow to further verify this hypothesis. Karremans et al. (2020) have also argued that *V. sotoarenasii* is not a species at all, but most likely an introgressed hybrid involving *V. odorata*, *V. × tahitensis*, and/or *V. planifolia*. Our results clearly rather show that *V. sotoarenasii* is a species close to *V. planifolia*, and that *V. × tahitensis* is a hybrid between *V. sotoarenasii* and *V. odorata*. Nevertheless, even if *V. sotoarenasii* was dismissed as a species as argued by Karremans et al. (2020), although we do not agree with this, the maternal origin of *V. × tahitensis* was more precisely addressed by our study and should be searched for in *V. sotoarenasii*-like populations. According to low divergence of ITS sequence data, *V. × tahitensis* appeared to be evolutionarily recent and it was suggested that it resulted from natural or man-mediated pollination in Mesoamerica (Lubinsky et al. 2008). *V. sotoarenasii* is present in southwestern Costa Rica (Azofeifa-Bolaños et al. 2017), and may also be also present in the north of Costa Rica. It was observed in northeastern Colombia (Choco, MG personal observation) and if it is confirmed that it is also present in Belize, this would suggest that the geographic range of *V. sotoarenasii* is much wider in America than the Caribbean area of Costa Rica where it has been reported so far. The other parent of *V. × tahitensis*, *V. odorata*, has a large distribution area covering Central and tropical South America. Thus, the hybrid origin of *V. × tahitensis*



is compatible with the sympatric range of both parents. Historically, the species is said to have been introduced from America to French Polynesia via the Philippines by Amiral Hamelin in 1848 (Correll 1953; Portères 195). The history of migrations and exchanges between the Viceroyalty of New Spain, the Philippines and Pacific islands between the sixteenth and eighteenth centuries (Merrill 1954) is compatible with this hypothesis. Indeed, it can be suggested that the shipment of pods or cuttings could have occurred 300 years ago on board of the Manila Galleons, the first Spanish ships that crossed the Pacific Ocean and introduced many plants from America to Asia and Oceania (Merrill 1954; Lubinsky et al. 2008).

Genome-wide  $H_o$  levels in wild species were very low ( $H_o = 0.001$  to  $0.031$ ) compared to cultivated species ( $H_o = 0.287$  and  $H_o = 0.239$ ), suggesting frequent inbreeding in the wild. Although vanilla flowers possess a rostellum preventing self-pollination, most species studied are self-compatible and selfing can occur via geitonogamy between different flowers from the same individual (Gigant et al. 2016). Further population genetic studies are needed to precisely assess possible genetic threats such as inbreeding on these species in the wild. The very low  $H_o$  in wild species can also be shown when observing GBS data in Alomia et al. (2021) and Hu et al. (2019), although the authors did not discuss this result. *V. planifolia* in cultivation was shown to have very low levels of diversity (AFLP, RAPD, SSR) (Besse et al. 2004; Bory et al. 2008a, c), and a single introduction in the southwestern Indian ocean area was suggested (Bory et al. 2008c). This pattern, compatible with a single-step domestication, was confirmed for *V. planifolia* using GBS. As clonal propagation is known to increase heterozygosity by the accumulation of point mutations (Balloux et al. 2003), single-step domestication process is always associated with high levels of heterozygosity, as shown for cassava (Elias et al. 2004) and hops (Jakše et al. 2001). Long established clones were indeed highly heterozygous as shown in cassava landraces ( $H_o = 0.50$  to  $0.71$ ) (Pujol et al. 2005). The high genome-wide  $H_o$  levels detected ( $0.362$ ) in cultivated clonal *V. planifolia* varieties worldwide are in accordance with this hypothesis. Most vanilla cultivars indeed result from almost 200 years of intense clonal propagation of the cuttings initially introduced to La Reunion in 1822. The present values are indeed high compared to those reported in *V. planifolia*

wild populations using 15 allozyme loci ( $H_o = 0$  to  $0.078$ ) (Soto Arenas 1999). They are higher than those revealed by 14 SSR markers ( $H_o = 0,154$ ) (Bory et al. 2008a), but close to those estimated from AFLP ( $H_o = 0,295$ ) (Bory et al. 2008c) in a similar set of *V. planifolia* varieties. SSR markers' length evolves more quickly than point mutation, and SSR were shown to reveal higher  $H_o$  than SNPs in population genetics studies (Fischer et al. 2017). However, in our particular case, if the levels of heterozygosity are due to the accumulation of mutations during clonal propagation (rather than from demographic and reproductive history as in natural populations), only SNPs can detect such mutations, since SSR only assess length variations in the number of microsatellite repeats. The  $H_o$  value obtained from dominant AFLP markers was deduced from the proportions of segregating bands in self-progenies (Bory et al. 2008c) and appears slightly underestimated. Our  $H_o$  levels are much higher than those revealed by Hu et al (2019) ( $H_o = 0.0322$  to  $0.0457$ ) but this might be due to their SNP filtration of  $H_o > 0,2$  (Hu et al. 2019) because the same accessions studied in a recent GBS analysis (Alomia et al. 2021) showed much higher  $H_o$  levels, compatible with our results. Interestingly, we therefore demonstrate here, thanks to a well characterized set of accessions from BRC Vatel, that  $H_o$  levels can be used to differentiate wild from cultivated vanillas. This could be very useful for less characterized collections. Indeed, this might indicate that the "hidden diversity" detected in *V. planifolia* by Alomia et al. (2021) using GBS could rather simply correspond to cultivated ( $H_o > 0.20$ , Type 1) compared to wild ( $H_o < 0,03$ , Type 2) accessions. Type 3 accessions with intermediate  $H_o$  values ( $0.08$  to  $0.15$ ) could be cultivars naturalized in the wild, or wild accessions with high natural cloning rate.

Based on AFLP analysis, *V. × tahitensis* was suggested to have a different domestication history than *V. planifolia* (Gigant et al. 2011; Lepers-Andrzejewski et al. 2011, 2012). Although a single introduction origin was also suggested, it has been followed by one or two generations of self-pollination, as shown by the detection of recombination events using AFLP graphical genotypes (Lepers-Andrzejewski et al. 2012). *V. × tahitensis* therefore rather fits the pattern of a single-step domestication (one introduction) followed by subsequent recombination-and-selection cycles (McKey et al. 2010). This hypothesis is supported

by genome-wide  $H_o$  that is lower in *V. × tahitensis* ( $H_o = 0.239$ ) than in *V. planifolia* ( $H_o = 0.362$ ).

Analysis of 1129 SNPs for 84 *V. planifolia* accessions showed a clear demarcation between the vegetatively propagated traditional vanilla cultivars compared to the accessions derived from sexual reproduction. GBS-generated SNPs could be used efficiently to better define the origin of unknown accessions: CR0196, CR2093, CR2100, CR2102 and 'Colibri' CR2687 were assigned to vegetatively propagated accessions, CR510 was assigned to accessions derived from sexual reproduction and CR0628 remains unknown. The cultivar 'Petite Mexique' CR0632 supposed to be a traditional cultivar with clonal propagation was branched and clustered with accessions derived from sexual reproduction.

Selfed-progenies and intra-specific crosses in *V. planifolia* showed an increased level of diversity (Fig. 4, Supporting information Fig. S3a, S4) as previously suggested by AFLP study (Bory et al. 2008c). They also showed a reduced genome-wide heterozygosity ( $H_o = 0.206$ ) as compared to the original parental cuttings group ( $H_o = 0.362$ ), as expected following selfing. Although heterozygosity is a favorable trait (ie hybrid vigor), the heterozygosity born from strict clonal propagation leads to the accumulation of deleterious mutations (McKey et al. 2010). Therefore *V. planifolia* plantations sustainability could be threatened by this mutational load (McKey et al. 2010). To preserve the adaptive potential of *V. planifolia*, the maintenance of mixed clonal/sexual systems is considered the best strategy (McKey et al. 2010). Efforts that have been engaged at BRCs Vatel and EVT to create new varieties from selfing will contribute to increase diversity levels, and also to reduce heterozygosity levels and therefore release mutational load. This has been a very successful strategy to create *Fusarium* resistant plant (Handa) (Grisoni and Dijoux 2017) from the selfing of a *V. planifolia* parent. Our study clarified the origin of *Vanilla* spp. from a large collection and their genetic diversity and structure, and provides new informations for breeding programs that contribute to the enhancement and protection of these materials. Tahitian vanilla, known to be more fragrant and fruity, illustrates the potential of hybridization to select new varieties.

**Acknowledgements** We are grateful to Katia Jade for preparing tissue cultured plant material and to the Plant Protection

Platform (3P, IBISA) for lab facilities and plant resources (BRC Vatel) access. We would like to thank the SouthGreen Bioinformatics Platform (<http://www.southgreen.fr/>) for computational resources and the GeT-PlaGe platform (INRAE, Toulouse, France) for Illumina sequencing. KeyGene N.V. owns patents and patent applications protecting its Sequence Based Genotyping technologies.

**Author contributions** CC, MG and CJ conceived and designed the experiment. JBD produced hybrids and grew the plant material at CRB Vatel. SL provided *V. × tahitensis* from the Etablissement Vanille de Tahiti (EVT) and provided her expertise on the genetics of *V. × tahitensis*. JH extracted DNA, RR produced GBS libraries, CJ performed the SNPs calling and QP the genomic mapping of SNPs. FF and CC analyzed the data and wrote the manuscript. PB, MG and CJ contributed to and edited the manuscript. All authors have reviewed and approved the final manuscript.

**Funding** This research was funded by the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) and the Université de la Réunion. Felicien Favre was supported by a MENRT grant from the French Ministry of Higher Education, Research and Innovation. This work was also funded by the European Regional Development Fund (ERDF), the Conseil Régional de la Réunion, and the Conseil Général de la Réunion. This work was carried out with the financial support of FORDECYT-CONACYT through the Vaniclim Project (Number 297484) "Estrategias para la adaptacion y mitigacion al cambio climatico necesarias para el rescate del cultivo de la vainilla en Mexico.

**Data availability** The raw sequence data have been deposited in the Short Read Archive under NCBI BioProject ID PRJNA756864. Biological samples used in this study have been deposited in the BioSamples Database under the accession identifier SAMN 20,927,883 to 20,928,015, and SAMN20981991 to SAMN20981994.

#### Declarations

**Conflict of interest** The authors declare that they have no conflict of interests.

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

## References

- Alomia YA, Chambers A, Brym M et al (2021) Genotyping-by-sequencing diversity analysis of international Vanilla collections uncovers hidden diversity and enables plant improvement. *Plant Sci*. <https://doi.org/10.1016/j.plantsci.2021.111019>
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Andriamihaja CF, Botomanga A, Misandeau C, et al (2022) Integrative taxonomy and phylogeny of leafless Vanilla orchids from the South West Indian ocean region reveal two new Malagasy species. *J Syst Evol*
- Azofeifa-Bolaños JB, Gigant LR, Nicolás-García M, et al (2017) A new vanilla species from Costa Rica closely related to *V. planifolia* (Orchidaceae). *EJT*. <https://doi.org/10.5852/ejt.2017.284>
- Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164:1635
- Besse P, Silva DD, Bory S et al (2004) RAPD genetic diversity in cultivated vanilla: *Vanilla planifolia*, and relationships with *V. tahitensis* and *V. pompona*. *Plant Sci* 167:379–385. <https://doi.org/10.1016/j.plantsci.2004.04.007>
- Bory S (2007) Diversity of *Vanilla planifolia* in the Indian ocean and its related species : genetics, cytogenetics and epigenetics aspect. Université de La Réunion
- Bory S, Da Silva D, Risterucci A-M et al (2008a) Development of microsatellite markers in cultivated vanilla: polymorphism and transferability to other vanilla species. *Sci Hort* 115:420–425. <https://doi.org/10.1016/j.scienta.2007.10.020>
- Bory S, Grisoni M, Duval M-F, Besse P (2008b) Biodiversity and preservation of vanilla: present state of knowledge. *Genet Resour Crop Evol* 55:551–571. <https://doi.org/10.1007/s10722-007-9260-3>
- Bory S, Lubinsky P, Risterucci A-M et al (2008c) Patterns of introduction and diversification of *Vanilla planifolia* (Orchidaceae) in Reunion Island (Indian Ocean). *Am J Bot* 95:805–815. <https://doi.org/10.3732/ajb.2007332>
- Bouetard A, Lefeuvre P, Gigant R et al (2010) Evidence of transoceanic dispersion of the genus *Vanilla* based on plastid DNA phylogenetic analysis. *Mol Phylogenet Evol* 55:621–630. <https://doi.org/10.1016/j.ympev.2010.01.021>
- Catchen J, Hohenlohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140. <https://doi.org/10.1111/mec.12354>
- Chambers A (2021) Advancing vanilla genomics and plant breeding for the Americas. IV Congreso Internacional de Vainilla <https://www.youtube.com/watch?v=vSElz3cYpJI>, 5:47:00
- Constantin D, Bois J (1915) Sur trois types de vanilles commerciales de TAHITI. *Comptes Rendus De L'académie Des Sciences De Paris* 161:196–202
- Correll DS (1953) Vanilla-its botany, history, cultivation and economic import. *Econ Bot* 7:291–358. <https://doi.org/10.1007/BF02930810>
- Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Díaz-Bautista M, Francisco-Ambrosio G, Espinoza-Pérez J et al (2018) Morphological and phytochemical data of Vanilla species in Mexico. *Data Brief* 20:1730–1738. <https://doi.org/10.1016/j.dib.2018.08.212>
- Earl DA, vonHoldt BM (2012) Structure harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genet Resour* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Elias M, Mühlen GS, McKey D et al (2004) Genetic diversity of traditional South American landraces of Cassava (*Manihot Esculenta* Crantz): an analysis using microsatellites. *Econ Bot* 58:242–256. [https://doi.org/10.1663/0013-0001\(2004\)058\[0242:GDOTSA\]2.0.CO;2](https://doi.org/10.1663/0013-0001(2004)058[0242:GDOTSA]2.0.CO;2)
- Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Favre F, Jourda C, Besse P, Charron C (2021) Genotyping-by-sequencing technology in plant taxonomy and phylogeny. In: Besse P (ed) *Molecular plant taxonomy*. Springer, New York, pp 167–178
- Fischer MC, Rellstab C, Leuzinger M et al (2017) Estimating genomic diversity and population differentiation—an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics* 18:69. <https://doi.org/10.1186/s12864-016-3459-7>
- Gigant R, Bory S, Grisoni M, Besse P (2011) Biodiversity and Evolution in the *Vanilla* Genus. In: *The dynamical processes of biodiversity - case studies of evolution and spatial distribution*, InTech, pp 1–26
- Gigant RL, De Bruyn A, M'sa T et al (2016) Combining pollination ecology and fine-scale spatial genetic structure analysis to unravel the reproductive strategy of an insular threatened orchid. *S Afr J Bot* 105:25–35. <https://doi.org/10.1016/j.sajb.2016.02.205>
- Grisoni M, Dijoux J-B (2017) Vanilla variety named “Handa.” USA patent application US14/999,830
- Hasing T, Tang H, Brym M et al (2020) A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nat Food* 1:811–819. <https://doi.org/10.1038/s43016-020-00197-2>
- Hu Y, Resende MFR, Bombarely A et al (2019) Genom- ics-based diversity analysis of *Vanilla* species using a *Vanilla planifolia* draft genome and genotyping-by-sequencing. *Sci Rep* 9:3416. <https://doi.org/10.1038/s41598-019-40144-1>
- Jakše J, Kindlhofer K, Javornik B (2001) Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. *Génome* 44:773–782. <https://doi.org/10.1139/gen-44-5-773>
- Karremans AP, Chinchilla IF, Rojas-Alvarado G, et al (2020) A reappraisal of Neotropical *Vanilla*. With a note on taxonomic inflation and the importance of alpha taxonomy

- in biological studies. *Lankesteriana*. <https://doi.org/10.15517/lank.v20i3.45203>
- Knaus BJ, Grünwald NJ (2017) vCFR : a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17:44–53. <https://doi.org/10.1111/1755-0998.12549>
- Lepers-Andrzejewski S, Siljak-Yakovlev S, Brown SC et al (2011) Diversity and dynamics of plant genome size: an example of polysomaty from a cytogenetic study of Tahitian vanilla (*Vanilla × tahitensis*, Orchidaceae). *Am J Bot* 98:986–997. <https://doi.org/10.3732/ajb.1000415>
- Lepers-Andrzejewski S, Causse S, Caromel B et al (2012) Genetic linkage map and diversity analysis of Tahitian Vanilla (*Vanilla × tahitensis*, Orchidaceae). *Crop Sci* 52:795–806. <https://doi.org/10.2135/cropsci2010.11.0634>
- Lubinsky P, Cameron KM, Molina MC et al (2008) Neotropical roots of a Polynesian spice: the hybrid origin of Tahitian vanilla, *Vanilla tahitensis* (Orchidaceae). *Am J Bot* 95:1040–1047. <https://doi.org/10.3732/ajb.0800067>
- Lubinsky P, Romero-González GA, Heredia SM, Zabel S (2010) Origins and patterns of *Vanilla* cultivation in tropical America (1500–1900): no support for an independent domestication of *Vanilla* in South America. In: Havkin-Frenkel D, Belanger FC (eds) *Handbook of Vanilla science and technology*. Wiley-Blackwell, Oxford, pp 117–138
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10. <https://doi.org/10.14806/ej.17.1.200>
- Martínez-Ainsworth NE, Tenaillon MI (2016) Superheroes and masterminds of plant domestication. *CR Biol* 339:268–273. <https://doi.org/10.1016/j.crvi.2016.05.005>
- McKey D, Elias M, Pujol B, Duputié A (2010) The evolutionary ecology of clonally propagated domesticated plants: Tansley review. *New Phytol* 186:318–332. <https://doi.org/10.1111/j.1469-8137.2010.03210.x>
- Merrill ED (1954) The botany of Cook's Voyages and it's Unexpected Significance in Relation to Anthropology, Biogeography and History (*Chronica Botanica*). Waltham, Massachusetts
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Perrier X, Jacquemoud-Collet JP (2006) DARwin: dissimilarity analysis and representation for windows. <https://darwin.cirad.fr/index.php>
- Piperidis N, D'Hont A (2020) Sugarcane genome architecture decrypted with chromosome-specific oligo probes. *Plant J* 103:2039–2051. <https://doi.org/10.1111/tpj.14881>
- Portères R (1954) Le genre *Vanilla* et ses espèces. Le vanillier et la vanille dans le monde. Paris, Editions Paul Lechevalier XL VI:94–290
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Pujol B, Mühlen G, Garwood N et al (2005) Evolution under domestication: contrasting functional morphology of seedlings in domesticated cassava and its closest wild relatives. *New Phytol* 166:305–318. <https://doi.org/10.1111/j.1469-8137.2004.01295.x>
- R Development Core Team (2010) A language and environment for statistical computing: reference index. R Foundation for Statistical Computing, Vienna
- Rambaut A (2006) FigTree. <http://tree.bio.ed.ac.uk/software/figtree/>
- Roux-Cuvellier M, Grisoni M, Bellec A et al (2021) Conservation of horticultural genetic resources in France. *Chronica Horticulturae* 61:21–36
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. In: *University of Kansas Science Bulletin*. pp 1409–1438
- Soto Arenas MÁ (1999) Filogeografía y recursos genéticos de las vainillas de México. Instituto Chinoin - Herbario de la Asociación Mexicana de Orquideología AC, Mexico
- Soto Arenas MÁ (2003) *Vanilla*. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN (eds) *Genera orchidacearum: Orchidoideae*. Oxford University Press, USA, p 402
- Soto Arenas MA, Cribb P (2013) A new infrageneric classification and synopsis of the genus *Vanilla* Plum. ex mill. (Orchidaceae: Vanillinae). *Lankesteriana*. <https://doi.org/10.15517/lank.v0i0.12071>
- Watteyn C, Scaccabarozzi D, Muys B et al (2022) Trick or treat? Pollinator attraction in *Vanilla pompona* (Orchidaceae). *Biotropica* 54:268–274. <https://doi.org/10.1111/btp.13034>
- Zohary D (2004) Unconscious selection and the evolution of domesticated plants. *Econ Bot* 58:5–10. [https://doi.org/10.1663/0013-0001\(2004\)058\[0005:USATEO\]2.0.CO;2](https://doi.org/10.1663/0013-0001(2004)058[0005:USATEO]2.0.CO;2)

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





### 3. Conclusion

Dans ce troisième chapitre, les relations de parenté entre les vanilliers cultivés et des espèces qui leur sont proches ont pu être résolues. Nous avons appliqué le génotypage par séquençage pour explorer la diversité génétique de deux espèces de vanilliers cultivés, de sept espèces étroitement apparentées et de dix-neuf hybrides interspécifiques. Les relations inter- et intraspécifiques de 133 accessions de vanilliers ont été examinées sur la base de 2004 SNPs filtrés. Nos résultats ont montré une forte structuration génétique entre les neuf espèces étudiées, les espèces sauvages présentant des niveaux d'hétérozygotie beaucoup plus faibles que ceux des espèces cultivées. De plus, des analyses bayésiennes de classification ont été utilisées pour vérifier la parenté de plusieurs hybrides. Nous avons notamment mis en évidence le fait que *V. sotoarenasii* et *V. odorata* pourraient être les espèces parentales de *V. × tahitensis*. L'analyse de 1129 SNPs pour 84 accessions de *V. planifolia* a montré une démarcation génétique claire entre les cultivars traditionnels de vanilliers à propagation clonale et les accessions issues de la reproduction sexuée. Une diversité génétique plus élevée et un taux d'hétérozygotie plus faible ont été observés chez ces dernières ( $H_o = 0,206$ ) par rapport aux cultivars traditionnels ( $H_o = 0,362$ ). Nos résultats suggèrent une domestication de type « single-step » pour *V. planifolia*, conformément à l'histoire récente de la culture du vanillier.

Cette étude a confirmé l'efficacité des marqueurs SNPs générés par GBS pour identifier des variations génétiques chez une espèce avec une base génétique très restreinte comme *V. planifolia*. Une plus grande diversité génétique a notamment été observée chez les 24 accessions issues de l'autofécondation du cultivar CR0040 de *V. planifolia* en comparaison au cultivar traditionnel multiplié par bouturage. Ces résultats confirment la pertinence de l'utilisation des marqueurs GBS pour générer une carte génétique à partir des individus issus d'une population d'autofécondation, et ainsi mettre en évidence des caractères d'intérêt chez les vanilliers cultivés, comme ceux impliqués dans la résistance à la fusariose.



## CHAPITRE 4 |

# CARTOGRAPHIE GÉNÉTIQUE DE LA RÉSISTANCE A LA FUSARIOSE DES VANILLIERS

---

### 1. Préambule

Le quatrième chapitre de cette thèse est consacré à l'étude génétique de la population issue de l'autofécondation du cultivar traditionnel CR0040 développée dans le cadre du programme d'amélioration variétale du vanillier. Cette population est un jeu de données précieux pour les études génétiques chez *V. planifolia* en donnant notamment accès aux recombinaisons chromosomiques par développement de marqueurs moléculaires dans la descendance. Un grand nombre d'individus est généralement nécessaire mais le faible taux de germination observé (Knudson 1950) et le délai de 3 à 4 ans pour obtenir les premières fleurs compliquent la création d'une telle population chez les vanilliers. Dans ce chapitre, 125 descendants de CR0040 conservés *in vitro* ont été génotypés par la méthode GBS.

Le criblage de populations pour étudier les caractères d'intérêt est souvent limité par le temps de développement des plants et les conditions environnementales. Dans le cadre de la thèse de S. Koyyappurath (2015), une méthode de phénotypage précoce de la résistance à *F. oxysporum* a été développée chez les vanilliers. Cette méthode permet de suivre le développement du champignon et les symptômes sur vitroplants âgés de 3 à 4 mois en conditions contrôlées. Cette méthode présente donc un gain de temps précieux. Les niveaux de résistance évalués par cette méthode sont similaires à ceux observés en ombrières (Koyyappurath 2015). Une caractérisation préliminaire de certaines accessions de la population avait mis en évidence une ségrégation du caractère de résistance à la fusariose des vanilliers (Koyyappurath et al. 2015b), suggérant l'hypothèse d'une résistance polygénique à Forv.



Des cartes génétiques à haute densité de marqueurs obtenus par GBS sont disponibles chez d'autres espèces, et sur lesquelles des QTLs de résistance à *F. oxysporum* ont été localisés (e.g. bananiers, Ahmad et al. 2020). La seule carte de liaison génétique connue dans le genre *Vanilla* a été développée à partir de marqueurs AFLP dans la descendance d'un croisement entre *V. × tahitensis* et *V. pompona* (Lepers-Andrzejewski et al. 2012), mais aucune carte de l'espèce la plus cultivée *V. planifolia* n'a à ce jour été rapportée. La disponibilité de nos nouveaux jeux de données (descendance et génome de CR0040), couplée aux méthodes de génotypage et de phénotypage, a été utilisée dans ce chapitre pour construire la première carte de liaison génétique de *V. planifolia* et identifier des régions génétiques associées à la résistance à Forv. Ces travaux ont été soumis sous forme de publication en tant que premier auteur à Plant Disease en octobre 2022.

## 2. Publication 3 : First *Vanilla planifolia* high-density genetic linkage map provides quantitative trait loci for resistance to *Fusarium oxysporum*

### Auteurs

Félicien Favre<sup>1</sup>, Cyril Jourda<sup>2</sup>, Michel Grisoni<sup>3</sup>, Frédéric Chiroleu<sup>2</sup>, Jean-Bernard Dijoux<sup>2</sup>, Katia Jade<sup>2</sup>, Ronan Rivallan<sup>4,5</sup>, Pascale Besse<sup>1</sup> and Carine Charron<sup>2a</sup>

<sup>1</sup> Université de La Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>2</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>3</sup> CIRAD, UMR PVBMT, Toamasina 501, Madagascar

<sup>4</sup> CIRAD, UMR AGAP, F-34398 Montpellier, France

<sup>5</sup> AGAP, Université de Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

<sup>a</sup> Corresponding author: Carine Charron (carine.charron@cirad.fr)

### Soumission

Soumis à Plant disease le 13/10/2022 - En relecture

### Mots clés

*Vanilla planifolia*, *Fusarium oxysporum* f.sp. *radicis-vanillae*, Génotypage par séquençage (GBS), Carte génétique dense, quantitative trait locus (QTL)



Plant Disease



**First *Vanilla planifolia* high-density genetic linkage map provides quantitative trait loci for resistance to *Fusarium oxysporum***

Journal:	<i>Plant Disease</i>
Manuscript ID	Draft
Manuscript Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	FAVRE, Félicien; Université de la Réunion, UMR PVBMT Jourda, Cyril; CIRAD BIOS, UMR PVBMT grisoni, michel; CIRAD BIOS, CHIROLEU, Frederic; CIRAD BIOS, UMR PVBMT Dijoux, Jean-Bernard; CIRAD BIOS, UMR PVBMT Jade, Katia; CIRAD BIOS, UMR PVBMT Rivallan, Ronan; CIRAD BIOS, UMR AGAP; AGAP, University of Montpellier, CIRAD, INRAE, Institut Agro Besse, Pascale; Université de la Réunion, UMR PVBMT CHARRON, Carine; CIRAD BIOS, UMR PVBMT
Keywords:	<i>Vanilla planifolia</i> G. Jackson ex Andrews, <i>Fusarium oxysporum</i> f.sp. <i>radicis-vanillae</i> , Genotyping-by-Sequencing, Quantitative trait locus, High-density genetic map

SCHOLARONE™  
Manuscripts

## ABSTRACT

*Fusarium oxysporum* f.sp *radicis vanilla* (Forv), the causal agent of Root and Stem Rot disease, is the main pathogen affecting vanilla production. Sources of resistance have been reported in *Vanilla planifolia* G. Jackson ex Andrews, the main cultivated vanilla species. Here we developed the first high-density genetic map in this species with 1,804 GBS-generated single nucleotide polymorphism (SNP) markers using 125 self-progenies of the CR0040 traditional vanilla cultivar. Sixteen linkage groups (LG) were successfully constructed, with a mean of 113 SNPs and an average length of 207 cM per LG. The map had a high density with an average of 5.45 SNP every 10 cM and an average distance of 1.85 cM between adjacent markers. The first 3 LG were aligned against the first assembled chromosome of CR0040 and the other 13 LG were correctly associated with the other 13 assembled chromosomes. The population was challenged with the highly pathogenic Forv strain Fo072 using the root-dip inoculation method. Five traits were mapped and 20 QTLs were associated with resistance to Fo072. Among genes retrieved in the CR0040 physical regions associated with QTLs, genes potentially involved in biotic resistance mechanisms have been highlighted, coding for kinases, E3 ubiquitin ligases, pentatricopeptide repeat-containing proteins, and one leucine-rich repeat receptor underlying the *qFo72\_08.1* QTL. This study should provide useful resources for marker-assisted selection in *V. planifolia*.

## KEYWORDS

*Vanilla planifolia* G. Jackson ex Andrews, *Fusarium oxysporum* f.sp. *radicis-vanillae*, Genotyping-by-Sequencing (GBS), high-density genetic map, quantitative trait locus (QTL).

## INTRODUCTION

The tropical orchid *Vanilla planifolia* G. Jackson ex Andrews is the major species for vanilla production in the world. Originating from Central America, *V. planifolia* was introduced into many areas, including the South-West Indian Ocean region, that has become the main vanilla-producing basin. Diversity studies using various genotyping technologies confirmed low level of genetic diversity in *V. planifolia*, consistent with its vegetative reproduction mode and recent introduction in cultivated areas (Besse et al. 2004; Bory et al. 2008b; Hu et al. 2019; Favre et al. 2022). *V. planifolia* has a diploid genome ( $2n = 32$ ), characterized by polyploidization events and partial genome endoreplication resulting in unbalanced nuclear DNA content in cells (Bory et al. 2008a; Brown et al. 2017; Piet et al. 2022). Cytogenetic studies revealed 16 chromosome pairs and a diploid genome size estimated at 4.09 Gb (Piet et al. 2022). In the two *V. planifolia* genome assemblies recently reported, only 14 chromosome pairs were assembled and diploid genome assembly size varied greatly from 1.5 Gb (Daphna cultivar, (Hasing et al. 2020) to 3.4 Gb (CR0040 cultivar, Piet et al. 2022). Although the latest version of the genome assembly is the most complete (about 83% of the expected genome size), only approximately one third is correctly anchored on 14 chromosome pairs (Piet et al. 2022). The authors demonstrated that the non-endoreplicated fraction of the CR0040 genome was not successfully assembled as it showed low-sequencing depth. Most of this fraction is represented in CR0040 chromosome 0 and was totally overlooked and therefore missing in the Daphna genome assembly.

The soilborne fungus *Fusarium oxysporum* causing Root and Stem Rot (RSR) of vanilla vines is the most hindering disease for the crop. The *F. oxysporum* Schlechtend. species complex includes both pathogenic and nonpathogenic strains (Edel-Hermann and Lecomte 2019). Pathogenic *F. oxysporum* strains cause most often wilt or in some cases root and crown rot in a wide range of economic plant hosts, comprising dicots such as tomato and bean, and monocots such as banana, palms and orchids. *F. oxysporum* causing the vascular wilt penetrates the roots to colonize xylem, whereas *F. oxysporum* causing the rot penetrates the roots, the crown and sometimes the stem

without reaching the vascular system. The term “radicis” is commonly accepted to designate *F. oxysporum* strains causing rot (Edel-Hermann and Lecomte 2019). Among the 106 well-characterized *formae speciales* (f. sp.), only six *F. oxysporum* f. sp. “radicis” have been reported (e.g. *radicis-betae*, *radicis-capsici*, *radicis-cucumerinum*, *radicis-lupini*, *radicis-lycopersici*, *radicis-vanillae*). *F. oxysporum* f. sp. *radicis-vanillae* (Forv) have been described since the beginning of the twentieth century and is found in all vanilla producing countries (Tucker 1927). Symptoms observed are browning and death of underground roots, grooved stem and transformation of the aerial grasping roots into nutritional roots descending to the soil. Stem rot is occasionally observed, but is confined to 2 or 3 internodes. A high diversity of strains with a continuous gradient in pathogenicity was described using *EF1 $\alpha$*  sequence and pathogenicity tests under controlled conditions (Koyyappurath et al. 2015a). The hyphae penetrate the rhizodermis in the hairy root region and disrupts the cortical cells without infection of the vascular system.

Some *Vanilla* species, such as *V. pompona* Schiede or *V. phaeantha* Rchb.f. and their hybrids, have already been reported as resistant to RSR (Knudson 1950; Theis and Jimenez 1957; Delassus 1963). Koyyappurath et al. (2015b) reported for the first time *V. planifolia* accessions with high level of resistance to RSR, but the underlying mechanisms have been so far poorly investigated. A constitutive defense was observed in resistant accessions that show a thickening of lignin on hypodermal cell-walls compared to susceptible plants (Koyyappurath et al. 2015b). Moreover, segregation of resistance to Forv was observed in a population derived from the self-fertilization of the susceptible accession CR0040 conserved in the VATEL Biological Resources Center (BRC, CIRAD, Reunion Island, Roux-Cuvelier et al. 2021), suggesting genetic variations associated to the resistance. Given the persistence of chlamydospores in soil (over 20 years), the pathogen is difficult to eradicate from the field. The use of resistant varieties meets the current need to control disease with environmentally friendly methods. Therefore, understanding the genetic mechanisms underlying plant resistance to Forv and mapping quantitative trait loci (QTLs) related to resistance are crucial for breeding *V. planifolia* varieties resistant to RSR.

Genetic maps and QTL mapping related to *Fusarium oxysporum* vascular wilt resistance have been reported in several species (e.g. watermelon, banana, radish and upland cotton) (Branham et al. 2019; Ahmad et al. 2020; Ma et al. 2021; Zhang et al. 2022), but little is known about the genetic mechanisms involved in resistance to *formae speciales* “*radicis*”. A first vanilla genetic linkage map was developed using molecular markers based on amplified fragment length polymorphism (AFLP) in the progeny of an interspecific cross between *V. ×tahitensis* J. M. Moore and *V. pompona* (Lepers-Andrzejewski et al. 2012). Nevertheless, no study related to the construction of high-density genetic maps and mapping of Forv resistance QTLs has been reported in *V. planifolia* to date. High throughput genotyping, such as genotyping by sequencing (GBS, Elshire et al. 2011), has been used in many crop species to develop a large number of molecular markers for high-density genetic map construction and QTL mapping. GBS method was successfully applied in *Vanilla* species to study genetic diversity by identifying thousands of single nucleotide polymorphisms (SNP) (Hu et al. 2019; Alomia et al. 2021; Favre et al. 2022), but has never been used to map QTLs related to RSR resistance in *V. planifolia*.

We present here the first high-density genetic map reported in *V. planifolia* using GBS in a population derived from the self-fertilization of a traditional vanilla cultivar (CR0040). The population was challenged with a highly pathogenic Forv strain (Fo072) in climatic chamber conditions in order to identify and map QTLs associated with the resistance. The genetic map was anchored to the CR0040 genome assembly to detect candidate genes in genomic regions related to RSR resistance. This study should provide useful resources for marker-assisted selection in *V. planifolia*.

## MATERIALS AND METHODS

### Plant materials and DNA extraction

A population of 128 accessions was produced from the self-fertilization of the susceptible *V. planifolia* accession CR0040 (**Supplementary Table S1**). CR0040 is a



traditional cultivar from Reunion Island and was provided by the VATEL BRC (CIRAD, Reunion Island, Roux-Cuvelier *et al.* 2021). Seeds were collected from the pod, disinfected and plated in MS medium without growth hormone, and maintained in the dark at 25°C for germination. All accessions were clonally propagated by stem cuttings and conserved *in vitro*, under conditions of 12 hours day/12 hours night at 25°C. Young leaves from the 128 accessions and the parent CR0040 were collected and lyophilized for genomic DNA extraction. The total genomic DNA was extracted with the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Genomic DNA was quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and DNA concentrations were normalized at 50 ng/μL. DNA homogeneity and quality was assessed by electrophoresis in a 2% agarose gel.

### **GBS library preparation and sequencing**

Library preparation was performed by the Regional genotyping technology platform (UMR AGAP, CIRAD, Montpellier, France) as described by Elshire *et al.* (2011). DNA was digested with *Pst*I restriction enzyme (New England Biolabs, Ipswich, Massachusetts, USA) and libraries were sequenced on a HiSeq3000 sequencer (Illumina Inc., San Diego, California, USA) with Paired-End DNA sequencing protocol at the GeT-PlaGe platform (INRAE, Toulouse, France).

### **Sequence analysis and SNP calling**

Sequence quality of 150 bp paired raw reads was checked with FastQC (v0.11.7, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads with Illumina adapter sequences and overrepresented sequences were removed using the Cutadapt software (Martin 2011). The remaining reads were assigned to each sample using the GBS barcode splitter tool (<https://sourceforge.net/projects/gbsbarcode/>). Demultiplexed sequences were trimmed to 142 bp to normalize the length between accessions. Samples with less than 400,000 reads were removed from the study and reads from replicates were concatenated. SNP calling was performed using both

STACKS *de novo* and reference-based pipelines (Catchen et al. 2013). Both pipelines can be run for a genetic map. The *de novo* pipeline is run without a reference genome and comprises additional steps to build a catalog of loci and match reads against the catalog. To run the reference-based pipeline, reads were first aligned to the reference genome CR0040 Haplotype A (Piet et al. 2022) using BWA aligner (v0.7.17, Li and Durbin 2009). Identified SNPs with 30% or more missing data and with minor allele frequency below 10% were removed. SNPs were converted into Variant Call Format file (VCF; Danecek *et al.* 2011) and visualized with vcfR package (v1.12.0, Knaus and Grünwald 2017, R Development Core Team 2010). Low-quality SNPs with depth value < 5 were filtered out and the genotyping data were extracted for further analyses. The loci sequences bearing SNPs identified with the *de novo* pipeline were aligned against CR0040 haplotypes A and B using ncbi-blast (v2.12.0+, Camacho et al. 2009). The loci sequences bearing SNPs identified with CR0040 haplotype A as reference were also aligned against CR0040 haplotype B.

### **Genetic map construction and anchoring on the *V. planifolia* genome**

SNP discovered with the *de novo* pipeline and with the reference-based pipeline were both used for genetic linkage map construction using JoinMap<sup>®</sup>5 (Van Ooijen 2018). The similarity between each pair of markers and segregation distortion were computed. For each identical pair, one marker was excluded to speed up the analysis. The remaining markers were grouped with the independence Logarithm of odds (LOD) score of 10. Linkage groups (LG) with 100 or more markers were kept. The ungrouped markers were assigned to strongest cross link (SCL) groups with a threshold for SCL-value of LOD > 3. The markers were then ordered using the regression mapping algorithm with the following parameters in JoinMap: a recombination frequency smaller than 0.50 and LOD larger than 0.50. The genetic distances were computed with the Kosambi function without the third round to discard the poorer fitting markers. The map was visualized with Spidermap (v1.6.12b, Rami 2014). The estimated genome length ( $L$ ) of each LG was calculated as described by Chakravarti et al. (1991) method 4 with  $L = l(m+1)/(m-1)$ , where  $l$  is the observed length of the LG and  $m$  the number of markers in each LG. The estimated genome coverage was calculated with  $c = 1 - e^{-2dn/L}$ ,

where  $d$  is the average marker interval,  $n$  is the number of markers, and  $L$  is the estimated genome length (Fishman et al. 2001). Based on the BLAST information of SNP markers on both CR0040 haplotypes A and B, each LG was assigned to a corresponding chromosome of CR0040, and positions of the SNPs were visualized with RIdeogram R package (v0.2.2, Hao et al. 2020; R Development Core Team 2010).

### ***F. oxysporum* f. sp. *radicis-vanillae* strain and inoculum preparation**

The highly pathogenic *F. oxysporum* f. sp. *radicis-vanillae* (Forv) strain Fo072 was used for inoculation experiments as described by Koyyappurath et al. (2015a). Fo072 was selected because of its stability and high pathogenicity on susceptible vanilla accessions. Fo072 was grown on potato dextrose agar (PDA) plates and stored in the dark at 25°C. The mycelium grown on PDA for 7 days was scraped and put into a minimal liquid medium prepared as follows: in 1L of distilled water, saccharose, 30g;  $\text{KH}_2\text{PO}_4$ , 1 g ;  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0,5 g ; KCl, 0,5 g ;  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.01 g ;  $\text{NaNO}_3$ , 2 g (Correll et al. 1987; Olivain et al. 2006; Koyyappurath et al. 2015a), and incubated in the dark for 7 days at 25°C on a rotary shaker at 125 rpm. The mycelium was removed by filtration and the conidial suspension was adjusted to  $1.10^6$  conidia.mL<sup>-1</sup>. The Forv strain Fo072 used in this study was available under the accession number MIAE01765 in the fungal collection of microorganisms of agro-environmental interest (MIAE, UMR Agroécologie, INRAe, Dijon, France).

### **Phenotyping and statistical analysis of traits**

The root-dip inoculation method was used to test the resistance of the population to Forv (Koyyappurath et al. 2015a). This method consists of dipping the roots of *in vitro* grown plants of 3-4 months in the conidial suspension of Fo072 for 5 min. The inoculated plants were transferred to plastic plots containing sterile coco fiber and bagasse as substrate and incubated in a growth chamber at 25°C and 80% relative humidity with a 12 h - 12 h day/night photoperiod. For each of the 124 CR0040 self-progenies, 2 independent tests were performed with 4 biological replicates if enough

material was available. The accessions CR0001 and CR0020 were used as susceptible and resistant control, respectively. The susceptibility of CR0040 was also tested. Disease on plantlet was rated as described in Koyyappurath (2015a). A disease score scale from 0 to 4 was used and recorded every day during 15 days (0: no symptom, 1: loss of vigor and loss of leaves brilliance, 2: browning of the stem base 3: development of the mycelium, 4: death of the plant). The area under the disease progress curve (AUDPC, Jeger and Viljanen-Rollinson 2001), was calculated using *agricolae* R package (v1.3-5, de Mendiburu 2021; R Development Core Team 2010). The disease score at 5, 10 and 15 days post-inoculation (dpi), d.5, d.10 and d.15 respectively, and the number of days with mycelium growth (d.myc) were also recorded. Broad-sense heritability ( $H^2$ ) was estimated for each trait with the *inti* R package (v0.5.1, Lozano-Isla 2022).

### **QTL mapping for RSR resistance and candidate genes analysis**

QTL mapping was performed on the AUDPC, d.5, d.10, d.15 and d.myc traits using Multiple-QTL Model mapping (MQM) in MapQTL<sup>®</sup>6 software (Van Ooijen 2009). The automatic cofactor selection function was used to select significantly associated markers as cofactors. First, markers every 20 cM were tested at a significance level of  $P < 0.02$ . The linkage groups were tested one after the other, while keeping all identified cofactors, as far as the results were fixed. Then, markers surrounding cofactors were also tested to refine their position. The remaining cofactors were used for detecting QTLs in MQM analysis. The phenotypic variation explained by the QTLs and positive or negative additive effect for the traits were computed in MapQTL<sup>®</sup>6 software. For each trait, the genome-wide LOD threshold was determined by a 10,000-permutations test at significance level of  $P < 0.05$ . The significant interval of each QTL was determined based on its LOD peak location and the surrounding region over the score threshold. QTLs detected were named *qFo072\_“linkage group”.*“QTL number”. Based on the BLAST information of SNP markers on both CR0040 haplotypes A and B, the position of the nearest marker to the peak was mapped against the CR0040 genome assembly to identify genomic regions of interest. All genes annotated within 100 kb on both sides of the marker were explored to identify candidate genes and similarity was searched using BLASTN (<https://blast.ncbi.nlm.nih.gov/>).

## RESULTS

### **A dense anchored genetic map for *Vanilla planifolia***

The sequencing of GBS libraries generated 1,766,927,320 reads. By removing the Illumina adapter sequences and overrepresented sequences, 987,711,332 reads were discarded (55.90% of the initial reads). A total of 554,312,542 reads were correctly demultiplexed (71.14% of the cleaned reads). CR2563, CR2869 and CR2959 accessions with low coverage (186,396 reads, 297,606 reads and 43,490 reads respectively) were removed from the study to avoid genotyping errors. A total of 7,282,844 reads in pairs were assigned to the parent CR0040. The F2 population of 125 accessions yielded a mean of 4,370,488 reads, with a minimum of 483,628 (CR2961) and a maximum of 12,224,658 (CR1063). A mean of 94,6% of reads was correctly aligned in pairs to the haplotype A of the reference genome CR0040 (**Supplementary Table S1**). A total of 160,819 and 77,649 loci were constructed with the *de novo* pipeline and the reference-based pipeline, respectively. After applying the filters, 1,553 SNPs on 759 loci were identified with the *de novo* pipeline and 4,012 SNPs on 1,812 loci were identified with the reference-based pipeline. Among the 1,553 SNPs identified with the *de novo* pipeline, 1,455 SNPs mapped simultaneously on the haplotypes A and B of CR0040 reference genome, 56 SNPs mapped specifically on the haplotype A and 39 SNPs exclusively on the haplotype B (**Table 1**). The density of *de novo* identified SNPs was estimated in the 14 assembled chromosomes of CR0040 with an average of 2.70 and 2.78 SNPs per 1 Mb in haplotypes A and B, respectively. The number of SNPs was lower in the unanchored sequence (chromosomes 0) with 0.04 and 0.07 SNPs per 1 Mb in A0 and B0 respectively. Among the 4,012 SNPs identified using the sequence of the haplotype A as reference, 3,916 SNPs were also found in haplotype B and 96 SNPs were specific to the haplotype A. The density was almost three times higher than *de novo* SNPs in the 14 assembled chromosomes of CR0040 with an average of 7.13 and 7.31 SNPs per 1 Mb in haplotypes A and B, respectively, and low in the chromosomes 0 with 0.04 and 0.17 SNPs per 1 Mb in A0 and B0, respectively.

**Table 1.** Number of SNP markers identified and aligned in (a) haplotype A and (b) haplotype B of the CR0040 genome assembly.**a.**

CR0040 chromosomes		Stacks <i>de novo</i> pipeline SNPs			Stacks reference-based pipeline SNPs		
ID	Size (bp)	Mapped SNPs	Specific-haplotype A SNPs	Density (SNPs per 1 Mb)	Mapped SNPs	Specific-haplotype A SNPs	Density (SNPs per 1 Mb)
A01	73,495,092	302	1	4.11	769	8	10.46
A02	42,591,792	97	0	2.28	269	8	6.32
A03	41,867,667	84	0	2.01	239	0	5.71
A04	36,435,613	100	2	2.74	223	2	6.12
A05	44,518,492	107	0	2.40	338	4	7.59
A06	42,387,411	100	12	2.36	352	46	8.30
A07	39,337,659	125	2	3.18	303	9	7.70
A08	31,377,353	125	1	3.98	297	0	9.47
A09	33,505,165	82	2	2.45	238	8	7.10
A10	32,653,060	76	2	2.33	257	9	7.87
A11	47,220,675	69	0	1.46	221	1	4.68
A12	32,781,290	83	1	2.53	231	0	7.05
A13	22,012,422	80	4	3.63	149	0	6.77
A14	20,090,243	46	0	2.29	94	0	4.68
A0	876,467,686	35	29	0.04	32	1	0.04

**b.**

CR0040 chromosomes		Stacks <i>de novo</i> pipeline SNPs			Stacks reference-based pipeline SNPs		
ID	Size (bp)	Mapped SNPs	Specific-haplotype B SNPs	Density (SNPs per 1 Mb)	Mapped SNPs	Specific-haplotype B SNPs	Density (SNPs per 1 Mb)
B01	68,775,951	292	4	4.25	710	0	10.32
B02	37,534,495	96	0	2.56	257	0	6.85
B03	40,029,660	71	0	1.77	208	0	5.20
B04	44,095,133	90	6	2.04	199	0	4.51
B05	42,224,683	104	5	2.46	332	0	7.86
B06	24,058,676	66	0	2.74	235	0	9.77
B07	33,373,288	125	8	3.75	261	0	7.82
B08	28,944,165	122	0	4.22	297	0	10.26
B09	31,370,254	80	0	2.55	212	0	6.76
B10	32,197,048	71	0	2.21	243	0	7.55
B11	31,724,078	69	0	2.18	214	0	6.75
B12	33,341,981	81	0	2.43	228	0	6.84
B13	21,853,202	79	4	3.62	146	0	6.68
B14	23,577,544	48	1	2.04	122	0	5.17
B0	1,476,298,157	100	11	0.07	252	0	0.17

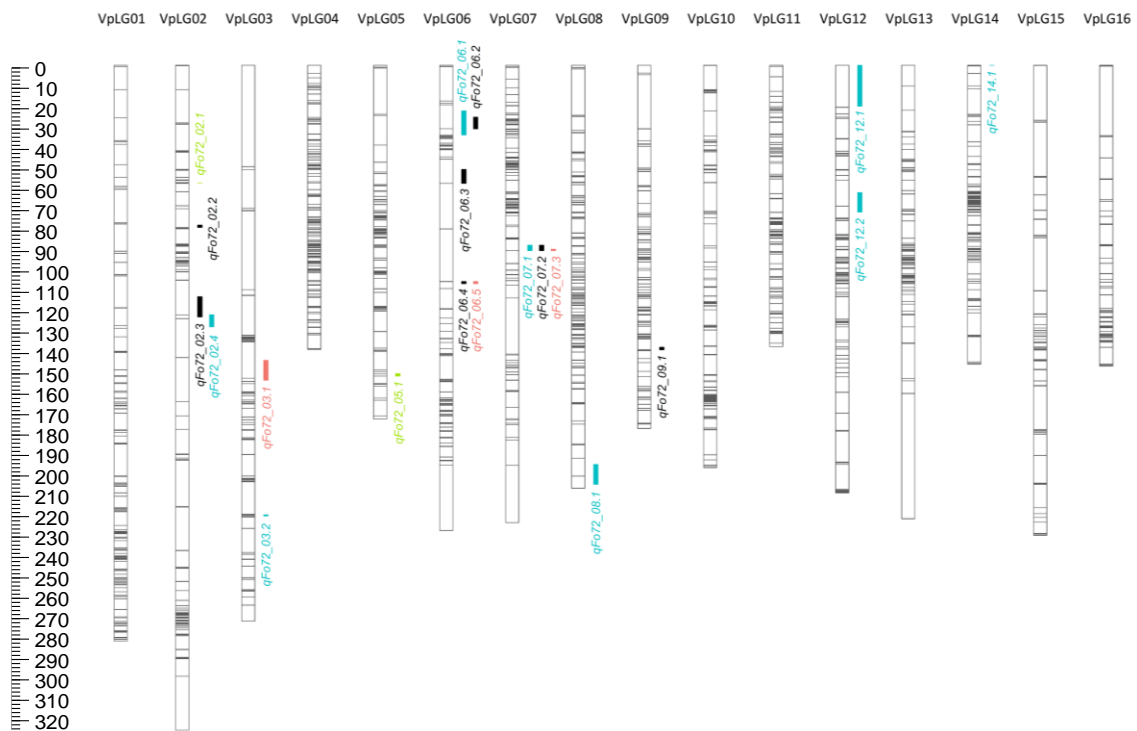
The 5,565 SNPs genotyped *de novo* and with reference-based for the 125 accessions were analyzed together using JoinMap. For each identical pair, one marker was excluded and 4,571 SNPs were used to create groups. At independence LOD of 10, a total of 4,410 SNPs formed 16 linkage groups (LG). Among the 161 ungrouped SNPs, 155 of them were assigned to strongest cross link (SCL) groups with a threshold for SCL-value of LOD > 3. The resulting 16 *V. planifolia* LG (VpLG) had a total of 4,565 SNPs with a mean of 285 SNPs, a minimum of 146 SNPs for VpLG16 and a maximum of 482 for VpLG01 (**Table 2**).

**Table 2.** Distribution of SNP markers among 16 linkage groups (LG) of the mapping population.

LG	Total number of SNPs	Number of SNPs for genetic mapping	Map length (cM)	Average marker interval (cM)	Maximum interval (cM)	Density (SNPs per 10 cM)	Estimated genome length (cM)	Genome coverage (%)
VpLG01	482	137	280.97	2.07	16.33	4.88	285.10	86.27
VpLG02	362	122	324.51	2.68	26.30	3.76	329.88	86.24
VpLG03	156	93	271.33	2.95	49.47	3.43	277.23	86.18
VpLG04	292	150	138.69	0.93	6.79	10.82	140.55	86.29
VpLG05	268	107	172.69	1.63	22.57	6.20	175.95	86.21
VpLG06	276	93	227.01	2.47	31.77	4.10	231.95	86.18
VpLG07	273	124	223.20	1.81	27.94	5.56	226.83	86.25
VpLG08	346	153	206.44	1.36	22.60	7.41	209.16	86.29
VpLG09	373	116	177.24	1.54	26.33	6.54	180.33	86.23
VpLG10	349	120	196.38	1.65	14.03	6.11	199.68	86.24
VpLG11	255	118	137.32	1.17	7.15	8.59	139.67	86.24
VpLG12	281	114	208.75	1.85	20.51	5.46	212.44	86.23
VpLG13	236	98	221.33	2.28	61.10	4.43	225.89	86.19
VpLG14	279	119	145.89	1.24	12.40	8.16	148.36	86.24
VpLG15	191	74	229.43	3.14	26.79	3.23	235.72	86.10
VpLG16	146	66	146.82	2.26	33.93	4.50	151.34	86.06
<b>Average</b>	285.31	112.75	206.75	1.85	25.38	5.45	210.63	86.21
<b>Total</b>	4565	1804	3308.01	-	-	-	3370.07	-

The construction of the genetic map resulted in a total of 1,804 SNPs (467 *de novo* SNPs and 1337 reference-based SNPs), with a mean of 113 SNPs per LG, a minimum of 66 SNPs for VpLG14 and a maximum of 153 SNPs for VpLG08 (**Figure 1, Supplementary Figure S1**). The length of LG ranged from 139 cM (VpLG04) to 325 cM (VpLG02). The map had a high density with an average of 5.45 SNPs every 10 cM and an average

distance of 1.85 cM between adjacent markers. The map was estimated to cover 86.21% of the anchored genome sequences (**Table 2**).

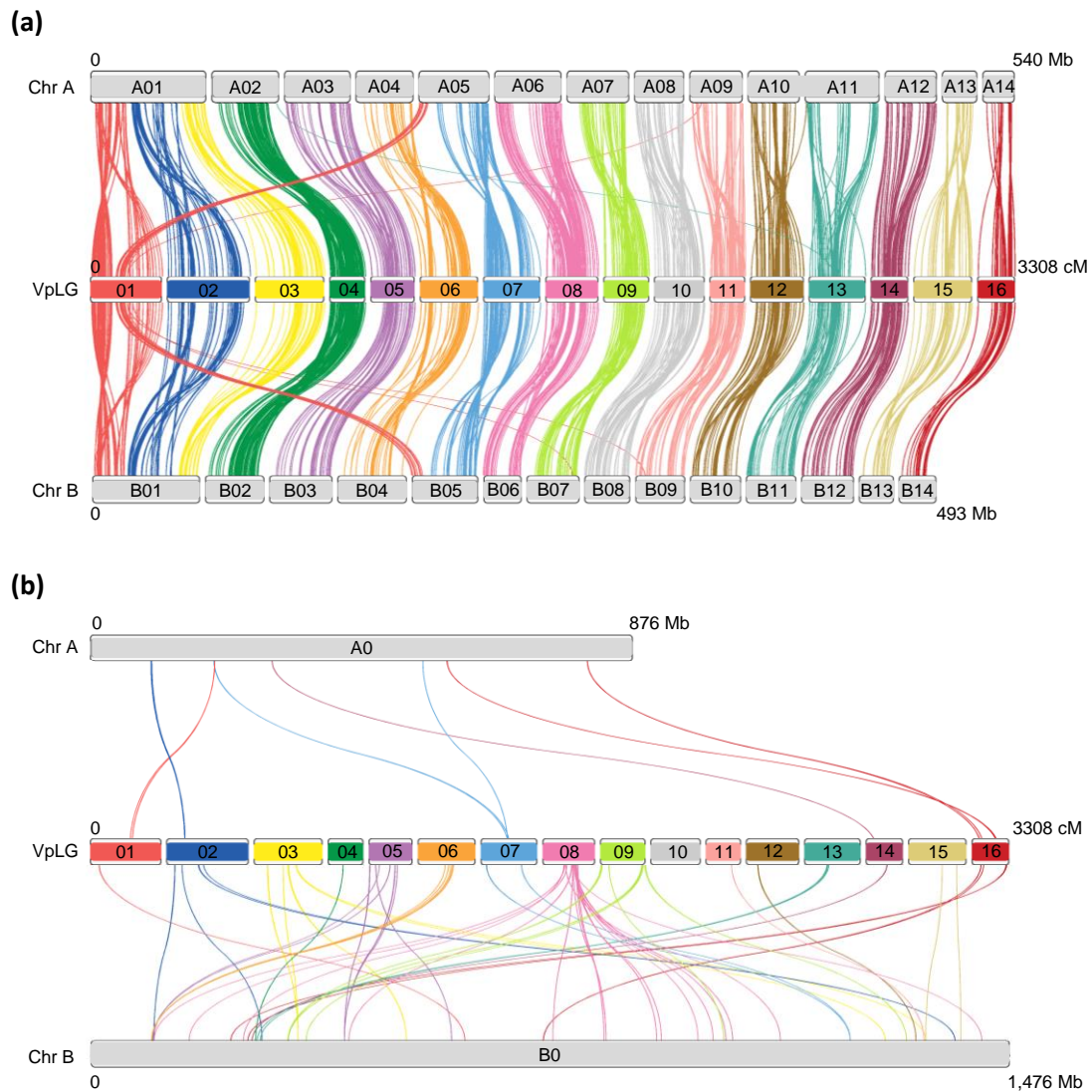


**Figure 1.** Genetic linkage map of the population obtained from the self-fertilization of *Vanilla planifolia* accession CR0040. The 16 linkage groups (LG) were constructed with 1,804 markers, symbolized by horizontal lines. Distances between markers are indicated in cM on the left ruler. Marker names and positions are found in Supplementary Figure S1. Quantitative trait loci (QTLs) related to resistance to Fo072 and their name are indicated on the right of their respective LG. Vertical bars indicate the QTL confidence interval and are colored according to the trait: AUDPC in black, d.5 in pink, d.15 in green and d.myc in blue.

Among the 1,804 SNPs, 1,780 SNPs (98.67 %) and 1,614 SNPs (89.47 %) were aligned against the 14 assembled chromosomes of haplotype A and B, respectively (**Figure 2.a**), 13 SNPs (0.72%) and 137 SNPs (7.59%) were aligned against chromosomes A0 and B0, respectively (**Figure 2.b**). These anchored markers helped to associate each LG to the 14 assembled chromosomes of CR0040. The first chromosome (A01 and B01) was divided into 3 LG (VpLG01, VpLG02 and VpLG03), and the size of each corresponding chromosome part was 22 Mb and 21 Mb for A01.1 and B01.1, 33 Mb and 30 MB for A01.2 and B01.2 and 19 Mb and 18 Mb for A01.3 and B01.3. The linkage groups VpLG04 to VpLG16 matched with the chromosomes A02 to A14 and B02 to B14. A



group of 27 SNPs of VpLG01 was aligned against the extremity of chromosomes A05 and B05, 3 SNPs of VpLG01 against A09 and B09 and another one SNPs of VpLG01 against B07. One marker on the extremity of VpLG12 was aligned against the extremity of A11 and one marker of VpLG13 with the extremity of A02.

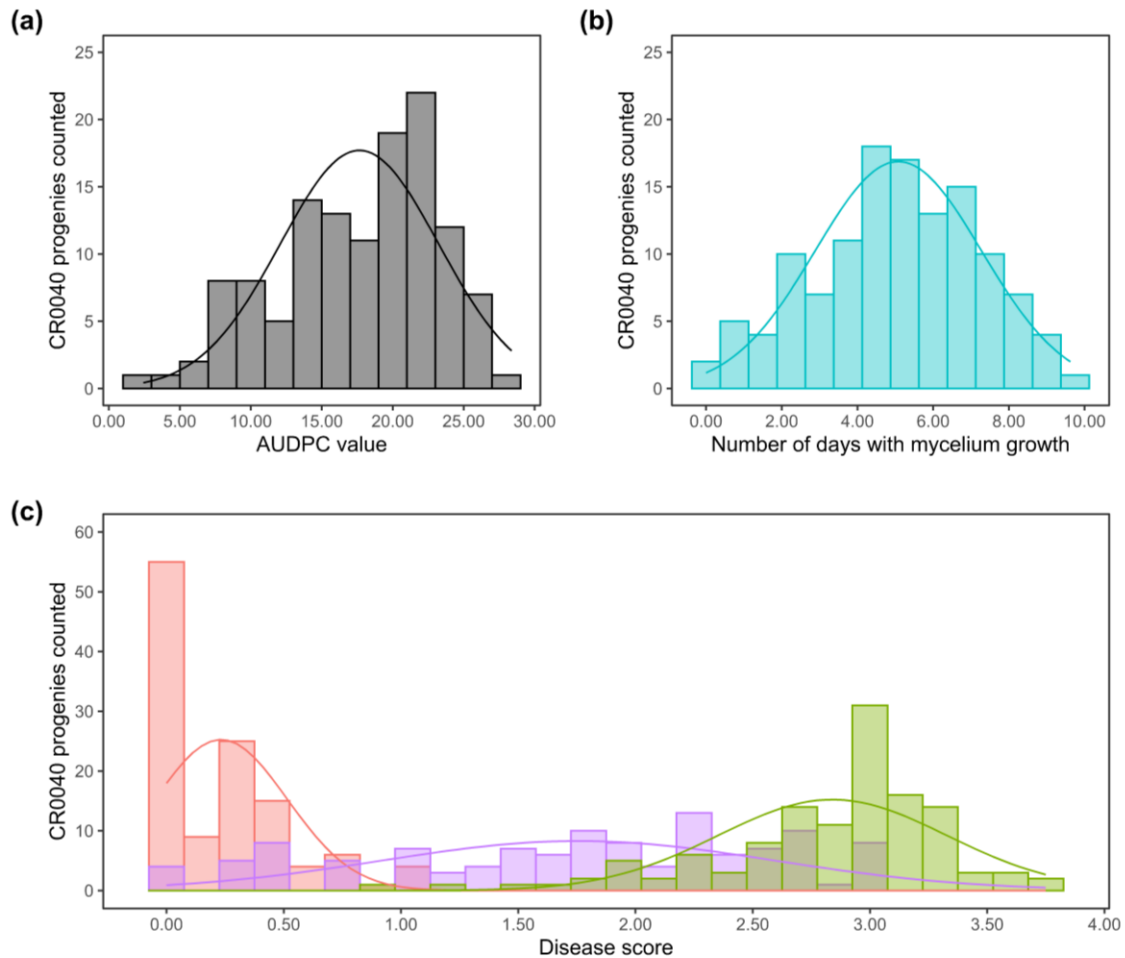


**Figure 2.** Mapping of the 16 linkage groups (VpLG) against **(a)** the 14 chromosomes of the CR0040 genome assembly, haplotypes A (Chr A) and B (Chr B), and **(b)** against the chromosome 0 of haplotypes A (Chr A) and B (Chr B).

## Identification of Forv resistance QTLs

The resistance to the strain Fo072 was recorded from 124 progenies of CR0040 using the root-dip inoculation method. The *V. planifolia* accessions CR0001 and CR0020, known to be susceptible and resistant to Fo072, respectively, were used as controls (Koyyappurath et al. 2015b). The disease scores from 0 to 4 were recorded every day during 15 days on 8 biological replicates for each accession. However, only 7 replicates were tested for 4 accessions and 4 replicates for 10 accessions because of undersize *in vitro* plants (Supplementary Table S2). Based on the disease scores, the AUDPC and the number of days with mycelium growth (d.myc) were calculated (**Supplementary Figure S2**). The mean values ranged from 2.44 (CR2977) to 28.38 (CR2869) for AUDPC and from 0.00 (CR2977) to 9.63 (CR2825) for d.myc.

The distributions of the mean values for AUDPC and d.myc within the population (**Figure 3 a. and b.**) were compared to a normal distribution (Shapiro-Wilk normality test, *P*-value of 0.006 for AUDPC and *P*-value of 0.131 for d.myc). To investigate resistance at different stages of infection, the disease scores at 5 dpi (d.5), 10 dpi (d.10) and 15 dpi (d.15) were recorded and the mean values ranged from 0.00 to 1.13, 0.00 to 3.00 and 0.88 to 3.75 for each trait respectively (Supplementary Table S2 and Figure S2). The distributions of the mean values (Figure 3 c.) were compared to a normal distribution (Shapiro-Wilk normality test *P*-value of 3.547e-11, 2.151e-4 and 8.928e-08 for d.5, d.10 and d.15 respectively). The *P*-values of the Shapiro-Wilk test for normality were lower than the threshold of 0.05, indicating the hypothesis of non-normal distribution of these traits in the population, except for d.myc. The d.myc trait was correlated with the development of the mycelium after inoculation and appeared to be a good variable to investigate the resistance to the fungus growth. The broad-sense heritability values ( $H^2$ ) ranged from 0.74 (d.5) to 0.87 (AUDPC).



**Figure 3.** Distribution of Fo072 resistance of the progeny of CR0040 accession evaluated by **(a)** the AUDPC value during 15 days, **(b)** the number of days with mycelium growth (d.myc), and **(c)** the disease score from 0 to 4 recorded at 5 days post-inoculation (dpi, d.5) in red, at 10 dpi (d.10) in purple and at 15 dpi (d.15) in green. The lines indicate the expected normal distribution for each measured trait.

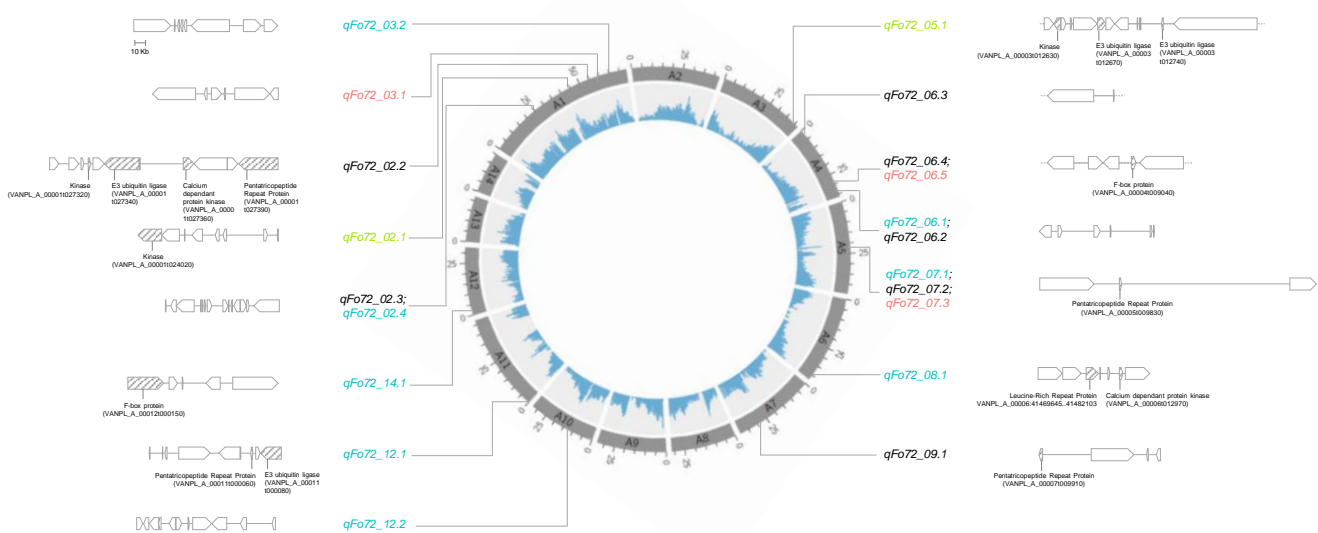
AUDPC, d.5, d.10, d.15 and d.myc traits were analyzed using the Multiple-QTL Model mapping (MQM) with automatic cofactor selection. A total of 13 markers were used as cofactors for AUDPC, 17 markers for d.myc, 9 markers for d.5, 6 markers for d.10 and 8 markers for d.15. A genome-wide LOD threshold of 4.5 was used for the AUDPC, d.10, and d.myc traits. A genome-wide LOD threshold of 4.7 and 4.8 was used for d.5 and d.15, respectively. A total of 20 QTLs for resistance to Fo072 were identified in 9 LG (**Figure 1 and Table 3**). Seven QTLs were detected with the AUDPC in the groups VpLG02 (*qFo72\_02.2*, *qFo72\_02.3*), VpLG06 (*qFo72\_06.2*, *qFo72\_06.3*, *qFo72\_06.4*), VpLG07 (*qFo72\_07.2*), VpLG09 (*qFo72\_09.1*). Eight QTLs were detected with d.myc in the groups VpLG02 (*qFo72\_02.4*), VpLG03 (*qFo72\_03.2*), VpLG06 (*qFo72\_06.1*),

VpLG07 (*qFo72\_07.1*), VpLG08 (*qFo72\_08.1*), VpLG12 (*qFo72\_12.1* and *qFo72\_12.2*) and VpLG14 (*qFo72\_14.1*). Three QTLs were detected with d.5 in VpLG03 (*qFo72\_03.1*), VpLG06 (*qFo72\_06.5*) and VpLG07 (*qFo72\_07.3*). Two QTLs were detected with d.15 in VpLG02 (*qFo72\_02.1*) and VpLG05 (*qFo72\_05.1*). No QTL was identified with d.10 (**Supplementary Figure S3**). The percentage of phenotypic variation explained by these QTLs varied from 5.9% (*qFo72\_14.1*) to 15.5% (*qFo72\_03.1*, Table 3) and the total is of 64% for AUDPC, 66% for d.myc, 36.7% for d.5 and 22.8% for d.15. The QTL *qFo72\_02.3* and *qFo72\_06.2* detected with AUDPC were co-located respectively with *qFo72\_02.4* and *qFo72\_06.1* detected with d.myc in VpLG02 and VpLG06. The three QTLs *qFo72\_07.1*, *qFo72\_07.2* and *qFo72\_07.3*, identified with d.myc, AUDPC and d.5, respectively, co-localized in VpLG07. The intervals varied from 1 to 20 cM and there were 7 main effect QTLs that explained more than 10% of the phenotypic variation (**Table 3**).

**Table 3.** Quantitative trait loci (QTL) associated with resistance to Fo072.

LG	QTL name	Traits	Interval (cM)	Peak		Phenotypic variation explained	Additive	Dominance
				Position (cM)	LOD			
VpLG02	<i>qFo72_02.1</i>	d.15	57.49 - 57.49	57.49	4.84	9.9	-8.85	-7.18
VpLG02	<i>qFo72_02.2</i>	AUDPC	78.01 - 79.23	79.23	4.77	7.1	2.09	2.19
VpLG02	<i>qFo72_02.3</i>	AUDPC	112.97 - 122.83	121.83	8.96	14.4	-1.43	-6.89
VpLG02	<i>qFo72_02.4</i>	d.myc	121.83 - 127.67	123.67	5.57	6.7	0.18	-1.72
VpLG03	<i>qFo72_03.1</i>	d.5	144.00 - 153.72	151.00	7.32	15.5	-0.03	-0.28
VpLG03	<i>qFo72_03.2</i>	d.myc	219.42 - 219.90	219.90	6.49	8.0	-0.79	0.63
VpLG05	<i>qFo72_05.1</i>	d.15	150.47 - 151.67	150.67	6.12	12.9	0.14	-0.33
VpLG06	<i>qFo72_06.1</i>	d.myc	22.27 - 33.98	30.98	9.1	11.8	-1.28	0.20
VpLG06	<i>qFo72_06.2</i>	AUDPC	25.27 - 30.98	28.27	5.46	8.1	-3.82	-1.44
VpLG06	<i>qFo72_06.3</i>	AUDPC	50.84 - 57.71	53.84	5.58	8.1	2.69	-3.53
VpLG06	<i>qFo72_06.4</i>	AUDPC	105.53 - 106.56	105.56	6.7	10.3	-2.48	0.66
VpLG06	<i>qFo72_06.5</i>	d.5	105.53 - 106.56	105.53	5.41	11.1	-0.12	0.06
VpLG07	<i>qFo72_07.1</i>	d.myc	87.79 - 90.50	90.50	6.36	7.8	-0.25	1.34
VpLG07	<i>qFo72_07.2</i>	AUDPC	87.79 - 90.50	89.79	4.91	7.3	-0.22	3.72
VpLG07	<i>qFo72_07.3</i>	d.5	89.79 - 90.50	90.50	4.96	10.1	-0.04	0.17
VpLG08	<i>qFo72_08.1</i>	d.myc	194.76 - 204.55	201.55	7.09	8.7	0.71	-1.42
VpLG09	<i>qFo72_09.1</i>	AUDPC	137.60 - 138.86	138.86	5.76	8.7	2.45	2.66
VpLG12	<i>qFo72_12.1</i>	d.myc	0.00 - 20.00	20.00	6.39	8.3	58.52	-14.18
VpLG12	<i>qFo72_12.2</i>	d.myc	62.08 - 71.74	69.74	7.09	8.8	-1.02	0.56
VpLG14	<i>qFo72_14.1</i>	d.myc	0.00 - 0.00	0.00	4.95	5.9	0.66	-0.52

*V. planifolia* CR0040 genome regions within significant QTL peaks were delimited for candidate genes. QTLs mapped in gene-rich regions, with 16 genes potentially involved in biotic resistance mechanisms (**Figure 4, Supplementary Table S3**). Most notably, one leucine-rich repeat (LRR) containing protein gene, VANPL\_B\_00006t008860, was present under *qFo72\_08.1* for d.myc in chromosome B06. This gene was also annotated in *V. planifolia* Daphna genome in both haplotypes A (Vpl06Ag14341) and B (Vpl06Bg13813.1), and the corresponding gDNA sequence was retrieved in CR0040 chromosome A06, but without associated annotation. A BLASTN analysis of this gene showed a high sequence identity with the predicted *Dendrobium catenatum* LRR protein 1 (LOC110113527) and *Phalaenopsis equestris* LRR receptor kinase BAK1-like (LOC110028109). In addition, four pentatricopeptide repeat-containing protein genes were highlighted in haplotypes A and B, and one supplemental copy on chromosome B03. Four genes encoding E3 ubiquitin ligase and two F-box proteins were identified in both haplotypes A and B. Finally, five QTLs regions revealed genes encoding different type of kinases (kinase domain-containing protein, membrane-associated kinase, calcium-dependent protein kinase, CBL-interacting protein kinase and chloroplastic adenylate kinase) (**Figure 4, Supplementary Table S3**).



**Figure 4.** Candidate genes for significant quantitative trait loci related to Fo072 resistance and position in the CR0040 haplotype A assembly. Names of QTLs are colored according to the trait AUDPC (black), d.myc (light blue), d.5 (red) and d.15 (green). Chromosomes are indicated in grey and gene density every 500 Kb in blue in the circos representation. Boxes indicate genes annotated in the genomic region of the QTLs and hatched boxes indicate putative candidate genes for resistance.

## DISCUSSION

### **The first dense genetic map revealed 16 genetic linkage groups for *V. planifolia***

Here we presented the first high-density genetic linkage map for the main cultivated species *V. planifolia*. The only genetic map for vanilla to date was generated from a cross between *V. xtahitensis* and *V. pompona* using 225 AFLP markers mapped on 18 LGs covering 1,035.85 cM (Lepers-Andrzejewski et al. 2012). The current genetic map contained 1,804 GBS-generated SNP markers distributed across 16 LGs covering 3,308.01 cM. A F2-like mapping population containing 125 CR0040 selfed progenies was used for GBS. Previous GBS-based diversity studies confirmed that CR0040, and more broadly cultivated vanilla species, showed high genome-wide heterozygosity and that selfed progenies of CR0040 showed an increased level of diversity and a reduced heterozygosity as compared to the parent (Favre et al. 2022). QTL mapping resolution is dependent on marker density and mapping population size. The F2 population size was sufficient to construct the genetic map as the current markers were distributed every 1.85 cM on average and a linkage map with interval < 2 cM is commonly considered as high-density map (Slate et al. 2009).

We constructed 16 LGs consistent with cytogenetic studies (Bory et al. 2008a; Piet et al. 2022). The map covered the 14 assembled chromosomes of the two available genomes Daphna and CR0040 (Hasing et al. 2020; Piet et al. 2022), but divided chromosome 1 into 3 LGs (VpLG01, VpLG02 and VpLG03). Piet et al. suggested that the missing part of the Daphna genome corresponded mainly to the non-endoreplicated fraction (71.6%) (Brown et al. 2017; Piet et al. 2022), and that the two missing chromosomes might be in the unanchored sequences (chromosome 0), that represented approximately two-third of the CR0040 genome. Here we suggested instead that all the 16 expected chromosomes were sequenced and assembled, but misplaced scaffolds have resulted in assembling the two missing chromosomes together within the first chromosome. As confirmed by our results, high-density genetic maps remain the basis for correct placement of scaffolds on the chromosomes

using genetically ordered markers (Lewin et al. 2009). Our work should help future studies aiming at better anchoring the genome sequence assemblies for *V. planifolia*. CR0040 assembly is the most complete to date and represents 83% of the estimated *V. planifolia* genome. Both *de novo* and reference-based Stacks pipeline were used for SNP calling, aiming to cover the entire genome (Catchen et al. 2013). Almost three times more SNP markers were identified with reference-based pipeline (4,012 SNPs) compared to *de novo* (1,553 SNPs), suggesting that not all loci were reconstructed in the *de novo* catalog. Among the 1,804 SNPs used for genetic map construction, 98.67 % and 89.47 % were correctly aligned to the 14 assembled chromosomes in haplotypes A and B, respectively. Interestingly, only 0.72% and 7.49% SNPs were retrieved in the chromosomes A0 and B0. These chromosomes, which correspond to the non-endoreplicated fraction, were obviously poorly covered in our mapping experiment because the DNA used was extracted from leaves, a tissue known to have higher proportion of endoreplicated fraction than other tissue such as nodes that were used for genome assembly in CR0040 (Piet et al. 2022). Given that only few SNP markers were present in chromosome 0, the map was probably estimated to cover 86.21% of the endoreplicated part of the genome, that comprises 67% of the genes (Piet et al. 2022). To elucidate the anchoring of the chromosomes 0, PCR markers such as simple sequence repeat (SSR) markers could be developed to amplify the non-endoreplicated part of the genome. From the known DNA sequences of chromosome 0, specific primers could be designed to generate specific markers that are expected to be amplified regardless of the tissue dependent low-coverage, which restricts identification of GBS-generated SNPs.

### **Multiple QTLs involved in resistance to *Fusarium oxysporum* f. sp. *radicis-vanillae*.**

*Fusarium oxysporum* is a major threat to vanilla cultivation. To our knowledge, this is the first QTL analysis conducted in a mapping population in *V. planifolia*. Accurate phenotyping is an important part of QTL mapping. We evaluated resistance by measuring symptoms on *in vitro*-grown plants. Koyyappurath et al. observed a

correlation between field and *in vitro* plant evaluation for Forv resistance (Koyyappurath 2015). We identified 20 significant QTLs for 5 traits: AUDPC, the number of days with mycelium growth (d.myc) and the disease score at 5, 10 and 15 dpi (d.5, d.10, d.15). The d.myc trait exhibited a normal distribution and appeared to be a suitable trait for Fo072 resistance mapping ( $H^2 = 0.80$ ). We identified one LRR containing protein gene underlying d.myc *qFo72\_08.1* in chromosome 6. This gene hold promises as candidate gene for Forv resistance. The majority of known plant resistance genes contain NBS and LRR domains (Sekhwal et al. 2015). High sequence identity was identified with the *Phalaenopsis equestris* predicted LRR receptor kinase BAK1-like. Brassinosteroid insensitive 1 associated kinase (BAK1) is a LRR receptor-like kinase (LRR-RLK) that has regulation functions in plant development and PAMP-triggered immunity (PTI) (Chinchilla et al. 2009). It is especially known to interact with the other LRR-RLK flagellin sensing 2 (FLS2) after recognition of the bacterial flagellin (Chinchilla et al. 2007). BAK1, co-receptor of BRI1 in brassinolide-dependent plant development, acts also as regulator of microbial infection-induced cell death (Kemmerling et al. 2007). Arabidopsis *bak1*-mutant exhibited enhanced susceptibility to necrotrophic fungal pathogens *Alternaria brassicicola* or *Botrytis cinerea*. LRR genes were also identified for *Fusarium oxysporum* resistance. The *I* (Immunity) gene conferring resistance against race 1 of *F. oxysporum* f. sp. *lycopersici* (*Fol*) encodes a LRR receptor-like protein (LRR-RLP) (Catanzariti et al. 2017). Like many other LRR-RLP involved in disease resistance, *I* gene action is dependent on LRR-RLK SOBIR1 (suppressor of BAK1-interacting RLK1 1) and SERK3/BAK1 (Somatic embryogenesis receptor kinase 3).

Several other genes, potentially involved in biotic resistance mechanisms, were located in the regions underlying QTLs. In particular, protein kinases play a central role in signaling during microbial recognition and induction of plant defense mechanisms (Romeis 2001). We identified two calcium-dependent protein kinase (CDPK), one of which was located in the same region than the identified LLR containing protein. CDPKs are multifunctional proteins perceiving  $Ca^{2+}$  ion signal and have been reported as one of the earliest responses after biotic and abiotic stresses, conferring for example resistance to fungal pathogens (Romeis et al. 2000; Geng et al. 2013). We also identified 4 pentatricopeptide repeat proteins (PPRs) and 4 E3 ubiquitin ligases.



Encoded by a large gene family, PRRs are better known for their functions in developmental processes, but have also been shown to restrict infection by necrotrophic fungal pathogens in *Arabidopsis thaliana*, by regulating reactive oxygen species homeostasis in mitochondria (Laluk et al. 2011). Some E3 ubiquitin ligase, involved in protein degradation, play also a role in resistance to fungal pathogens (e.g. ALT9 protein in *A. thaliana*) (Guo et al. 2022).

## Conclusions

Historically, *Fusarium* resistance has always been a major concern for vanilla. Breeding programs were set up in Madagascar in the 1960s and then abandoned with the arrival of synthetic vanilla (Delassus 1963). The natural flavor of vanilla is experiencing a revival in recent years and genetic control appears to be the most promising method in the context of global changes. Here, a GBS approach in a CR0040 selfed progenies population has yielded for the first time a high-density genetic map that provided significant QTLs for a critical agronomic trait. This work was based on resistance to the Fo072 strain and the diversity of Forv strains described in vanilla growing soils should be tested to confirm the detected QTLs (Koyyappurath et al. 2015a). Finally, our work showed that vanilla accessions could provide Forv resistance variation for application in breeding programs and generated preliminary information on genes potentially involved in resistance mechanisms.

## REFERENCES

- Ahmad, F., Martawi, N. M., Poerba, Y. S., de Jong, H., Schouten, H., and Kema, G. H. J. 2020. Genetic mapping of *Fusarium* wilt resistance in a wild banana *Musa acuminata* ssp. *malaccensis* accession. *Theor Appl Genet.* 133:3409–3418.
- Alomia, Y. A., Chambers, A., Brym, M., Cibrián-Jaramillo, A., Hasing, T., Hernandez-Hernandez, J., et al. 2021. Genotyping-By-Sequencing diversity analysis of international Vanilla collections uncovers hidden diversity and enables plant improvement. *Plant Science.* :111019.
- Besse, P., Silva, D. D., Bory, S., Grisoni, M., Le Bellec, F., and Duval, M.-F. 2004. RAPD

genetic diversity in cultivated vanilla: *Vanilla planifolia*, and relationships with *V. tahitensis* and *V. pompona*. *Plant Science*. 167:379–385.

Bory, S., Catrice, O., Brown, S., Leitch, I. J., Gigant, R., Chiroleu, F., et al. 2008a. Natural polyploidy in *Vanilla planifolia* (Orchidaceae) ed. Paolo Donini. *Genome*. 51:816–826.

Bory, S., Lubinsky, P., Risterucci, A.-M., Noyer, J.-L., Grisoni, M., Duval, M.-F., et al. 2008b. Patterns of introduction and diversification of *Vanilla planifolia* (Orchidaceae) in Reunion Island (Indian Ocean). *American Journal of Botany*. 95:805–815.

Branham, S. E., Levi, A., and Wechter, W. P. 2019. QTL Mapping Identifies Novel Source of Resistance to Fusarium Wilt Race 1 in *Citrullus amarus*. *Plant Disease*. 103:984–989.

Brown, S. C., Bourge, M., Maunoury, N., Wong, M., Wolfe Bianchi, M., Lepers-Andrzejewski, S., et al. 2017. DNA Remodeling by Strict Partial Endoreplication in Orchids, an Original Process in the Plant Kingdom. *Genome Biology and Evolution*. 9:1051–1071.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.

Catanzariti, A., Do, H. T. T., Bru, P., Sain, M., Thatcher, L. F., Rep, M., et al. 2017. The tomato I gene for Fusarium wilt resistance encodes an atypical leucine-rich repeat receptor-like protein whose function is nevertheless dependent on SOBIR 1 and SERK 3/ BAK 1. *Plant J*. 89:1195–1209.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 22:3124–3140.

Chakravarti, A., Lasher, L. K., and Reefer, J. E. 1991. A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics*. 128:175–182.

Chinchilla, D., Shan, L., He, P., de Vries, S., and Kemmerling, B. 2009. One for all: the receptor-associated kinase BAK1. *Trends in Plant Science*. 14:535–541.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J. D. G., et al. 2007. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature*. 448:497–500.

Correll, J. C., Klittich, C. J. R., and Leslie, J. F. 1987. Nitrate nonutilizing mutants of *Fusarium oxysporum* and their use in vegetative compatibility tests. *The*

- American Phytopathological Society. :1640–1646.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. 2011. The variant call format and VCFtools. *Bioinformatics*. 27:2156–2158.
- Delassus, M. 1963. La lutte contre la fusariose du vanillier par les méthodes génétiques. *L'Agronomie Tropicale. Série 2, Agronomie Générale. Etudes Techniques*. 18:245–246.
- Edel-Hermann, V., and Lecomte, C. 2019. Current Status of *Fusarium oxysporum* Formae Speciales and Races. *Phytopathology*<sup>®</sup>. 109:512–530.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species ed. Laszlo Orban. *PLoS ONE*. 6:e19379.
- Favre, F., Jourda, C., Grisoni, M., Piet, Q., Rivallan, R., Dijoux, J.-B., et al. 2022. A genome-wide assessment of the genetic diversity, evolution and relationships with allied species of the clonally propagated crop *Vanilla planifolia* Jacks. ex Andrews. *Genet Resour Crop Evol.* :1–15.
- Fishman, L., Kelly, A. J., Morgan, E., and Willis, J. H. 2001. A Genetic Map in the *Mimulus guttatus* Species Complex Reveals Transmission Ratio Distortion due to Heterospecific Interactions. *Genetics*. 159:1701–1716.
- Geng, S., Li, A., Tang, L., Yin, L., Wu, L., Lei, C., et al. 2013. TaCPK2-A, a calcium-dependent protein kinase gene that is required for wheat powdery mildew resistance enhances bacterial blight resistance in transgenic rice. *Journal of Experimental Botany*. 64:3125–3136.
- Guo, T., Kong, F., Burton, C., Scaglione, S., Beagles, B., Ray, J., et al. 2022. The E3 Ubiquitin Ligase ATL9 Affects Expression of Defense Related Genes, Cell Death and Callose Deposition in Response to Fungal Infection. *Pathogens*. 11:68.
- Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., et al. 2020. Rideoqram : drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science*. 6:e251.
- Hasing, T., Tang, H., Brym, M., Khazi, F., Huang, T., and Chambers, A. H. 2020. A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nat Food*. 1:811–819.
- Hu, Y., Resende, M. F. R., Bombarely, A., Brym, M., Bassil, E., and Chambers, A. H. 2019. Genomics-based diversity analysis of *Vanilla* species using a *Vanilla planifolia*

- draft genome and Genotyping-By-Sequencing. *Sci Rep.* 9:3416.
- Jeger, M. J., and Viljanen-Rollinson, S. L. H. 2001. The use of the area under the disease-progress curve (AUDPC) to assess quantitative disease resistance in crop cultivars: *Theor Appl Genet.* 102:32–40.
- Kemmerling, B., Schwedt, A., Rodriguez, P., Mazzotta, S., Frank, M., Qamar, S. A., et al. 2007. The BRI1-Associated Kinase 1, BAK1, Has a Brassinolide-Independent Role in Plant Cell-Death Control. *Current Biology.* 17:1116–1122.
- Knaus, B. J., and Grünwald, N. J. 2017. VCFR : a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour.* 17:44–53.
- Knudson, L. 1950. GERMINATION OF SEEDS OF VANILLA. *American Journal of Botany.* 37:241–247.
- Koyyappurath, S., Atuahiva, T., Le Guen, R., Batina, H., Le Squin, S., Gautheron, N., et al. 2015a. *Fusarium oxysporum* f. sp. *radicis-vanillae* is the causal agent of root and stem rot of vanilla. *Plant Pathology.* 65:612–625.
- Koyyappurath, S., Conéjéro, G., Dijoux, J. B., Lapeyre-Montès, F., Jade, K., Chiroleu, F., et al. 2015b. Differential Responses of Vanilla Accessions to Root Rot and Colonization by *Fusarium oxysporum* f. sp. *radicis-vanillae*. *Front. Plant Sci.* 6 Available at: <http://journal.frontiersin.org/Article/10.3389/fpls.2015.01125/abstract> [Accessed August 28, 2020].
- Laluk, K., AbuQamar, S., and Mengiste, T. 2011. The Arabidopsis Mitochondria-Localized Pentatricopeptide Repeat Protein PGN Functions in Defense against Necrotrophic Fungi and Abiotic Stress Tolerance. *Plant Physiology.* 156:2053–2068.
- Lepers-Andrzejewski, S., Causse, S., Caromel, B., Wong, M., and Dron, M. 2012. Genetic Linkage Map and Diversity Analysis of Tahitian Vanilla (*Vanilla × tahitensis*, Orchidaceae). *Crop Science.* 52:795–806.
- Lewin, H. A., Larkin, D. M., Pontius, J., and O'Brien, S. J. 2009. Every genome sequence needs a good map. *Genome Res.* 19:1925–1928.
- Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25:1754–1760.
- Lozano-Isla, F. 2022. inti: Tools and Statistical Procedures in Plant Science. R package version 0.5.7, <https://CRAN.R-project.org/package=inti>.

- Ma, Y., Chhapekar, S. S., Lu, L., Yu, X., Kim, S., Lee, S. M., et al. 2021. QTL mapping for Fusarium wilt resistance based on the whole-genome resequencing and their association with functional genes in *Raphanus sativus*. *Theor Appl Genet.* 134:3925–3940.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* 17:10.
- de Mendiburu, F. 2021. *Agricolae: Statistical Procedures for Agricultural Research.*
- Olivain, C., Humbert, C., Nahalkova, J., Fatehi, J., L'Haridon, F., and Alabouvette, C. 2006. Colonization of Tomato Root by Pathogenic and Nonpathogenic *Fusarium oxysporum* Strains Inoculated Together and Separately into the Soil. *Applied and Environmental Microbiology.* 72:1523–1531.
- Piet, Q., Droc, G., Marande, W., Sarah, G., Bocs, S., Klopp, C., et al. 2022. A chromosome-level, haplotype-phased *Vanilla planifolia* genome highlights the challenge of partial endoreplication for accurate whole-genome assembly. *Plant Communications.* :100330.
- R Development Core Team. 2010. a language and environment for statistical computing: reference index. Vienna: R Foundation for Statistical Computing. Available at: <http://www.polsci.wvu.edu/duval/PS603/Notes/R/fullrefman.pdf> [Accessed July 1, 2020].
- Rami, J. 2014. Spidermap v1.6.12b.
- Romeis, T. 2001. Protein kinases in the plant defence response. *Current Opinion in Plant Biology.* 4:407–414.
- Romeis, T., Piedras, P., and Jones, J. D. G. 2000. Resistance Gene-Dependent Activation of a Calcium-Dependent Protein Kinase in the Plant Defense Response. *Plant Cell.* 12:803–815.
- Roux-Cuvelier, M., Grisoni, M., Bellec, A., Bloquel, E., Charron, C., Delalande, M., et al. 2021. Conservation of horticultural genetic resources in France. *Chronica Horticulturae.* 61:21–36.
- Sekhwal, M., Li, P., Lam, I., Wang, X., Cloutier, S., and You, F. 2015. Disease Resistance Gene Analogs (RGAs) in Plants. *IJMS.* 16:19248–19290.
- Slate, J., Gratten, J., Beraldi, D., Stapley, J., Hale, M., and Pemberton, J. M. 2009. Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica.* 136:97–107.

- Theis, T., and Jimenez, F. A. 1957. A Vanilla hybrid resistant to Fusarium root rot. *Phytopathology*. 47:578–581.
- Tucker, C. M. 1927. Vanilla Root Rot. *Journal of Agricultural Research*. 35:1121–1136.
- Van Ooijen, J. W. 2018. JoinMap®5, Software for the calculation of genetic linkage maps in experimental populations of diploid species. Kyazma B.V., Wageningen, Netherlands.
- Van Ooijen, J. W. 2009. MapQTL®6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma B.V., Wageningen, Netherlands.
- Zhang, J., Abdelraheem, A., Ma, J., Zhu, Y., Dever, J., Wheeler, T. A., et al. 2022. Mapping of dynamic QTLs for resistance to Fusarium wilt (*Fusarium oxysporum* f. sp. *vasinfectum*) race 4 in a backcross inbred line population of Upland cotton. *Mol Genet Genomics*. 297:319–332.

## STATEMENTS AND DECLARATIONS

### Funding

This research was funded by the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) and the University of Reunion Island. Felicien Favre was supported by a MENRT grant from the French Ministry of Higher Education, Research and Innovation. This work was also funded by the European Regional Development Fund (ERDF), the Conseil Régional de la Réunion, and the Conseil Général de la Réunion.

### Conflicts of interest

The authors declare that they have no conflict of interests.

### Availability of data and material

Biological samples used in this study were submitted to the BioSamples Database with the ID SUB12140282. The raw sequence data were submitted to the Sequence Read Archive with NCBI BioProject ID SUB12140289.

### **Authors' contributions**

Félicien Favre, Carine Charron and Pascale Besse conceived and designed the experiment. Jean-Bernard Dijoux produced the population and grew the plant material at CRB Vatel. Katia Jade prepared tissue cultured plant material and helped in inoculation experiments. Frédéric Chiroleu provided help for statistical analysis of the experimental data and Ronan Rivallan produced the GBS libraries. Félicien Favre extracted DNA, performed the SNP calling, conducted the inoculation experiments, analyzed the data and wrote the manuscript. Carine Charron, Pascale Besse, Michel Grisoni and Cyril Jourda contributed to and edited the manuscript. All authors have reviewed and approved the final manuscript.

### **ACKNOWLEDGMENTS**

We are grateful to the Plant Protection Platform (3P, IBISA) for lab facilities and plant resources (BRC Vatel) access. We would like to thank Julie Stanislas for her assistance in inoculation experiments and DNA extraction. We thank the SouthGreen Bioinformatics Platform (<http://www.southgreen.fr/>) for computational resources and the GeT-PlaGe platform (INRAE, Toulouse, France) for Illumina sequencing. KeyGene N.V. owns patents and patent applications protecting its Sequence Based Genotyping technologies.

### 3. Conclusion

Dans ce quatrième chapitre, la première carte de liaison génétique de *V. planifolia* construite à partir de 1804 marqueurs GBS a permis de mettre en évidence 20 QTLs associés à la résistance à Forv. Contrairement à l'étude de diversité du chapitre 3, la première version de la séquence génomique de CR0040 publiée en 2022 a pu être utilisée comme référence pour identifier les marqueurs SNP dans la population et pour ancrer la carte génétique sur le génome. Des erreurs probables d'assemblage ont été mises en évidence par la carte génétique, qui pourra ainsi servir de support pour la deuxième version d'assemblage du génome de CR0040. Très peu de marqueurs ont été identifiés sur les parties non assemblées du génome (chromosomes A0 et B0). Ce résultat était attendu car ces régions correspondent aux régions non endorépliquées pour lesquelles une faible profondeur de séquençage avait été obtenue (Piet et al. 2022). Nous avons montré par cette étude la limite de la méthode GBS chez les vanilliers et la nécessité de combiner nos marqueurs SNP avec des marqueurs PCR, comme des marqueurs microsatellites, pour couvrir l'ensemble du génome, dont les régions très peu répliquées mais présentant potentiellement un contenu génique d'intérêt pour le caractère étudié.

La carte génétique couvre cependant la majeure partie de la région endorépliquée qui contient environ deux tiers des gènes annotés (disponibles sur le Vanilla Génome Hub ; <https://vanilla-genome-hub.cirad.fr/>). Les 20 QTLs identifiés se retrouvent dans des régions riches en gènes. Par comparaison avec les données de la littérature scientifique, certains de ces gènes présentent des caractéristiques de gènes de résistance et ont été retenus comme gènes candidats potentiellement impliqués dans le mécanisme de résistance à la fusariose des vanilliers. Cependant, à ce jour, aucune étude des gènes de résistance n'a été réalisée chez *V. planifolia* à l'échelle du génome complet.





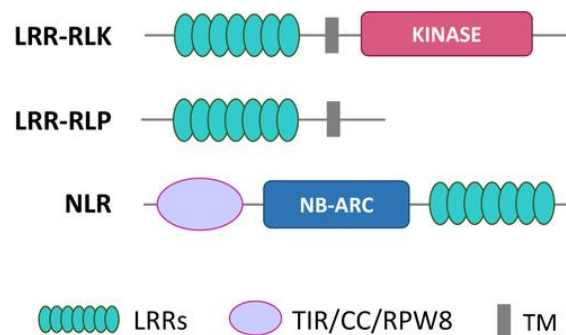
## CHAPITRE 5 |

# EXPLORATION DU GÉNOME DE *V. PLANIFOLIA* : CRIBLAGE DE GÈNES CANDIDATS POTENTIELLEMENT IMPLIQUÉS DANS LA RÉSISTANCE À LA FUSARIOSE

---

## 1. Introduction

Les végétaux ont acquis des mécanismes leur permettant de reconnaître les infections causées par des agents pathogènes et d'y réagir. Les analogues de gènes de résistance (RGAs) sont des gènes de résistance potentiels qui présentent des domaines et des motifs conservés qui jouent des rôles précis dans ce type de mécanisme de résistance actif (Sekhwal et al. 2015). La principale classe de RGAs regroupe les récepteurs qui possèdent un domaine Leucine-rich repeat (LRR). Le domaine LRR contient de 4 à plus de 30 répétitions d'un motif LRR de 24 acides aminés. La famille des LRR comprend 3 sous-familles : LRR receptor-like kinase (LRR-RLK), LRR receptor-like protein, et Nucleotide-binding site LRR (NBS-LRR ou NLR) (**Figure 5.1**).



**Figure 5.1.** Schéma de la structure protéique des 3 sous-familles de récepteurs LRR : LRR-RLK, LRR-RLP et NLR. TM, domaine transmembranaire ; CC, coiled coil domain ; TIR, toll-interleukin receptor ; RPW8, Resistance to powdery mildew 8 ; NB-ARC, Nucleotide binding-APAF1 (apoptotic protease-activating factor-1), R proteins and CED-4 (*Caenorhabditis elegans* death-4 protein) (Gottin et al., 2021).

Les LRR-RLKs et LRR-RLPs sont des récepteurs transmembranaires qui interviennent dans la première réponse immunitaire des plantes appelée PTI (PAMP-triggered immunity). Ils reconnaissent des motifs moléculaires associés aux pathogènes (PAMP). Les NBS-LRRs possèdent un domaine NB-ARC et sont quant à eux des récepteurs intracellulaires qui interviennent dans la deuxième réponse immunitaire ETI (Effector-triggered immunity). Ils reconnaissent de manière spécifique des effecteurs produits par le pathogène (Sekhwal et al. 2015; Gottin et al. 2021). Chez le vanillier, la recherche de ces familles de gènes dans l'annotation de *V. planifolia* Daphna et la comparaison avec les bases de données publiques a mis en évidence 11 gènes *NLR* potentiellement impliqués dans la résistance à Forv (Hasing et al. 2020).

Des méthodes bioinformatiques ont été développées pour identifier les RGAs à partir des données génomiques chez les plantes, sur la base de leurs caractéristiques structurales et leurs domaines et motifs conservés (**Figure 5.1**). L'approche consiste en 4 étapes : (i) une base de données de RGAs plantes est générée, (ii) le génome annoté est aligné sur cette base de données pour identifier les RGAs candidats, (iii) les domaines et motifs de ces RGAs candidats sont détectés et (iv) les RGAs sont triés en classes fonctionnelles selon le motif détecté afin de constituer un jeu de gènes candidats potentiels pour la recherche de gènes de résistance à la fusariose. De plus, des gènes jouant un rôle dans la résistance à *F. oxysporum* ont été identifiés chez d'autres plantes et constituent une base de données précieuse pour la recherche de gènes de résistance chez le vanillier. *F. oxysporum* f. sp. *radicis-lycopersici* (FORL) est l'agent responsable de la pourriture des racines et du collet chez *Solanum lycopersicum*. Le mécanisme de résistance n'a pas été identifié, cependant le locus de résistance *Frl* s'est avéré efficace contre toutes les souches de FORL (Fazio et al. 1999; Devran et al. 2018). *Musa acuminata* Pahang présente une forte résistance à *F. oxysporum* f. sp. *cubense* Foc TR4. Des études transcriptomiques ont révélé que certaines enzymes de la voie métabolique des phénylalanines étaient sur-exprimées durant les dernières étapes de l'infection par Foc TR4, suggérant le rôle de la lignine et des phytotoxines dans la résistance face au pathogène (Zhang et al. 2019).

Les analyses histologiques de racines de vanilliers infectées par une souche pathogène de Forv ont notamment mis en évidence un renforcement pariétal en lignine de l'hypoderme chez les variétés résistantes. Cette caractéristique physique des cellules de l'épiderme et du cortex racinaire joue un rôle de barrière, empêchant la pénétration du champignon (Koyyappurath et al. 2015b). Ces observations soulèvent de nouvelles perspectives sur l'implication de la lignine et des précurseurs de la voie de biosynthèse des phénylpropanoïdes dans le mécanisme de résistance des vanilliers à la fusariose. Les composés phénylpropanoïdes permettent la biosynthèse de plusieurs métabolites induisant les mécanismes de défense des plantes tels que des acides, des flavonoïdes et la lignine (Dixon et al., 2002).

L'objectif de ce chapitre est d'exploiter les données de la littérature et les données de séquençage du génome de CR0040 (Piet et al. 2022) pour rechercher et étudier les caractéristiques structurales des grandes familles de gènes de résistance. J'ai présenté un poster synthétisant ces résultats lors du congrès international « 9th Plant Genomics & Gene Editing » (The Hagues, Netherlands, avril 2022) et ce chapitre a fait l'objet d'un stage de Master 2 (Q. Da Silva) que j'ai encadré en 2022.

## 2. Matériel et méthodes

### 2.1. Annotation des gènes *LRR* chez *V. planifolia*

A partir de l'annotation du génome de CR0040 disponible sur le *Vanilla* Genome Hub (<https://vanilla-genome-hub.cirad.fr/>) (Piet et al. 2022), le pipeline d'annotation des domaines protéiques LRRprofiler (Gottin et al. 2021) a été utilisé pour identifier et classer les gènes *LRR* en 3 sous-familles : *NLR*, *RLK* et *RLP*. LRRprofiler s'appuie sur les outils HMMER (hmmbuild et hmmsearch) pour identifier les séquences LRR et NB-ARC, l'outil iTAK pour identifier les séquences kinase et des bases de données publiques comme SMART pour annoter et classer les domaines. Les sorties graphiques des structures protéiques ont été vérifiées afin d'éliminer les annotations douteuses

(fusion de gènes, gènes tronqués ou splittés). La localisation des gènes *LRR* a été représentée sur les deux haplotypes A et B du génome de CR0040 à l'aide du package Rldeogram (Hao et al. 2020). Les annotations du cultivar Daphna de *V. planifolia*, de 5 autres espèces de la famille des *Orchidaceae* (*Apostasia shenzhenica*, *Dendrobium catenatum*, *Dendrobium chrysotoxum*, *Gastrodia elata* et *Phalaenopsis equestris*), de 4 *Poaceae* (*Miscanthus lutarioriparius*, *Oryza sativa*, *Sorghum bicolor* et *Zea mays*), de *M. acuminata* et de 5 dicotylédones (*S. lycopersicum*, *Coffea arabica*, *Helianthus annuus*, *Theobroma cacao* et *Arabidopsis thaliana*) disponibles dans les bases de données publiques ont été analysées par LLRprofiler pour identifier les gènes *LRR* et comparer avec CR0040. La densité de chaque famille de gènes dans le génome entier a été calculée pour faciliter la comparaison.

## **2.2. Identification de gènes candidats potentiellement impliqués dans la résistance à *F. oxysporum***

Les gènes codant pour des protéines impliquées dans des mécanismes de résistance à *F. oxysporum* ont été recensés à partir de la littérature scientifique. Les séquences orthologues de ces gènes ont été identifiées chez CR0040 par alignement de séquences (Basic local alignment search tool, BLAST) pour constituer un pool de gènes candidats potentiellement impliqués dans la résistance à Forv chez le vanillier. Les gènes de la voie de biosynthèse des phénylpropanoïdes ont été identifiés par (i) recherche des gènes sur l'annotation automatique de CR0040 (Pathways Tools, *Vanilla* Genome Hub, Piet et al. 2022) et par (ii) alignement des séquences (BLAST) de gènes rapportés dans des études précédentes (Rao et al. 2014; Fock-Bastide et al. 2014). La position de ces gènes sur CR0040 est représentée avec Rldeogram (Hao et al. 2020). Les données de transcriptomique obtenues à partir de l'extraction de l'ARN de différents organes de plants CR0040 conservés en ombrière et utilisées pour l'annotation du génome ont été récupérées. Les valeurs de FPKM (Fragments per kilobase million) des bourgeons, boutons, crampons, feuilles, fleurs, fruits matures, ovaires, racines et tiges ont été catégorisées en 5 classes de niveaux d'expression différents : expression négligeable (FPKM < 0.5), expression extrêmement faible (0.5 ≤

FPKM < 3), expression faible ( $3 \leq \text{FPKM} < 50$ ), expression modérée ( $50 \leq \text{FPKM} < 100$ ) et expression forte ( $\text{FPKM} \geq 100$ ).

### 3. Résultats

#### 3.1. Identification d'un faible nombre de gènes *LRR* chez *V. planifolia*

L'annotation des domaines protéiques a révélé 381 gènes *LRR* dans le génome de CR0040, dont 179 se retrouvent sur l'haplotype A (5 *NLRs*, 152 *RLKs* et 22 *RLPs*), et 202 sur l'haplotype B (10 *NLRs*, 165 *RLKs* et 27 *RLPs*) (Tableau 5.1, Figure 5.2).

**Tableau 5.1.** Nombre de gènes codant pour des LRRs identifiés par LRRprofiler à partir des deux annotations de *V. planifolia* (CR0040 et Daphna) et de 15 autres espèces végétales disponibles dans les bases de données.

Clade	Famille	Espèce	BioProject ID	Taille totale du génome (Mb)	NLR <sup>1</sup>	RLK <sup>2</sup>	RLP <sup>3</sup>
Dicot	<i>Solanaceae</i>	<i>Solanum lycopersicum</i>	PRJNA119	828	222	330	103
Dicot	<i>Rubiaceae</i>	<i>Coffea arabica</i>	PRJNA497895	1094	1571	740	306
Dicot	<i>Asteraceae</i>	<i>Helianthus annuus</i>	PRJNA345532	3010	560	551	273
Dicot	<i>Sterculiaceae</i>	<i>Theobroma cacao</i>	PRJEB14326	325	430	295	205
Dicot	<i>Brassicaceae</i>	<i>Arabidopsis thaliana</i>	PRJNA10719	120	91	171	58
Monocot	<i>Musaceae</i>	<i>Musa acuminata</i>	PRJEA82777	472	102	533	135
Monocot	<i>Poaceae</i>	<i>Miscanthus lutarioriparius</i>	PRJEB40463	2075	816	544	265
Monocot	<i>Poaceae</i>	<i>Oryza sativa</i>	PRJDB1747	374	700	472	153
Monocot	<i>Poaceae</i>	<i>Sorghum bicolor</i>	PRJNA13876	709	464	313	86
Monocot	<i>Poaceae</i>	<i>Zea mays</i>	PRJEB32225	2183	254	431	50
Monocot	<i>Orchidaceae</i>	<i>Apostasia shenzhenica</i>	PRJNA310678	349	14	129	27
Monocot	<i>Orchidaceae</i>	<i>Dendrobium catenatum</i>	PRJNA262478	1104	122	271	60
Monocot	<i>Orchidaceae</i>	<i>Dendrobium chrysotoxum</i>	PRJNA664445	1368	78	142	37
Monocot	<i>Orchidaceae</i>	<i>Gastrodia elata</i>	PRJNA632604	1046	3	118	20
Monocot	<i>Orchidaceae</i>	<i>Phalaenopsis equestris</i>	PRJNA192198	1064	42	221	55
Monocot	<i>Orchidaceae</i>	<i>Vanilla planifolia</i> (CR0040 A <sup>4</sup> )	PRJNA753216	1416	5	152	22
Monocot	<i>Orchidaceae</i>	<i>Vanilla planifolia</i> (CR0040 B <sup>5</sup> )	PRJNA754028	1968	10	165	27
Monocot	<i>Orchidaceae</i>	<i>Vanilla planifolia</i> (Daphna A)	PRJNA633886	737	12	156	57
Monocot	<i>Orchidaceae</i>	<i>Vanilla planifolia</i> (Daphna B)	PRJNA668740	744	9	151	49

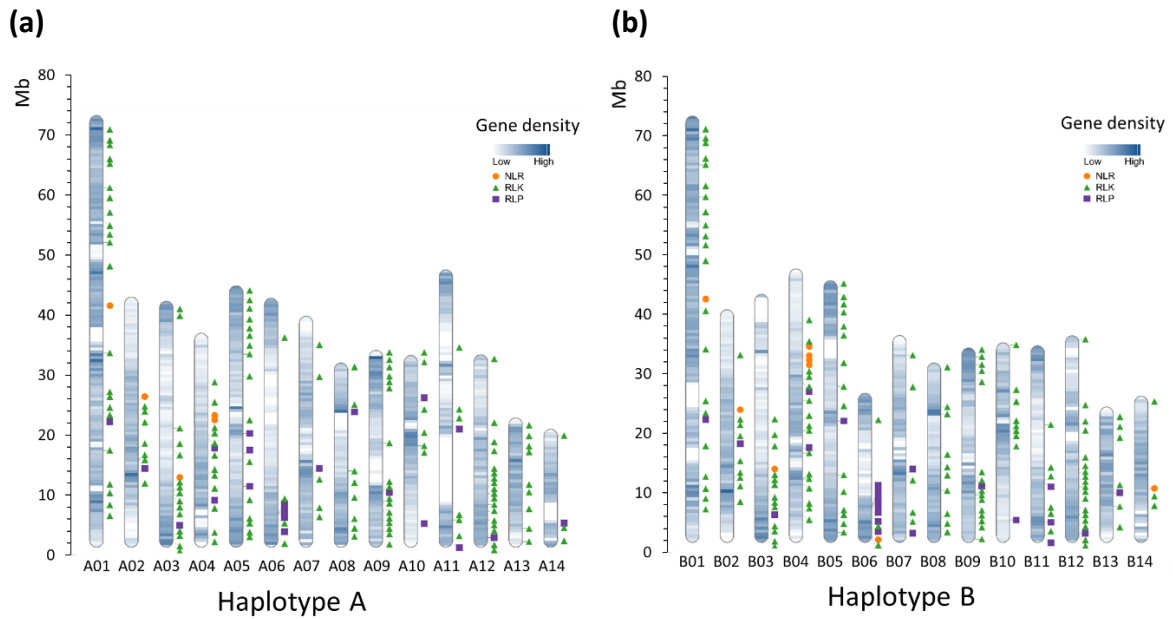
<sup>1</sup>NLR : Nucleotide-binding and leucine-rich repeat immune receptor

<sup>2</sup>RLK : Receptor-like kinase

<sup>3</sup>RLP : Receptor-like protein

<sup>4</sup>A : Haplotype A

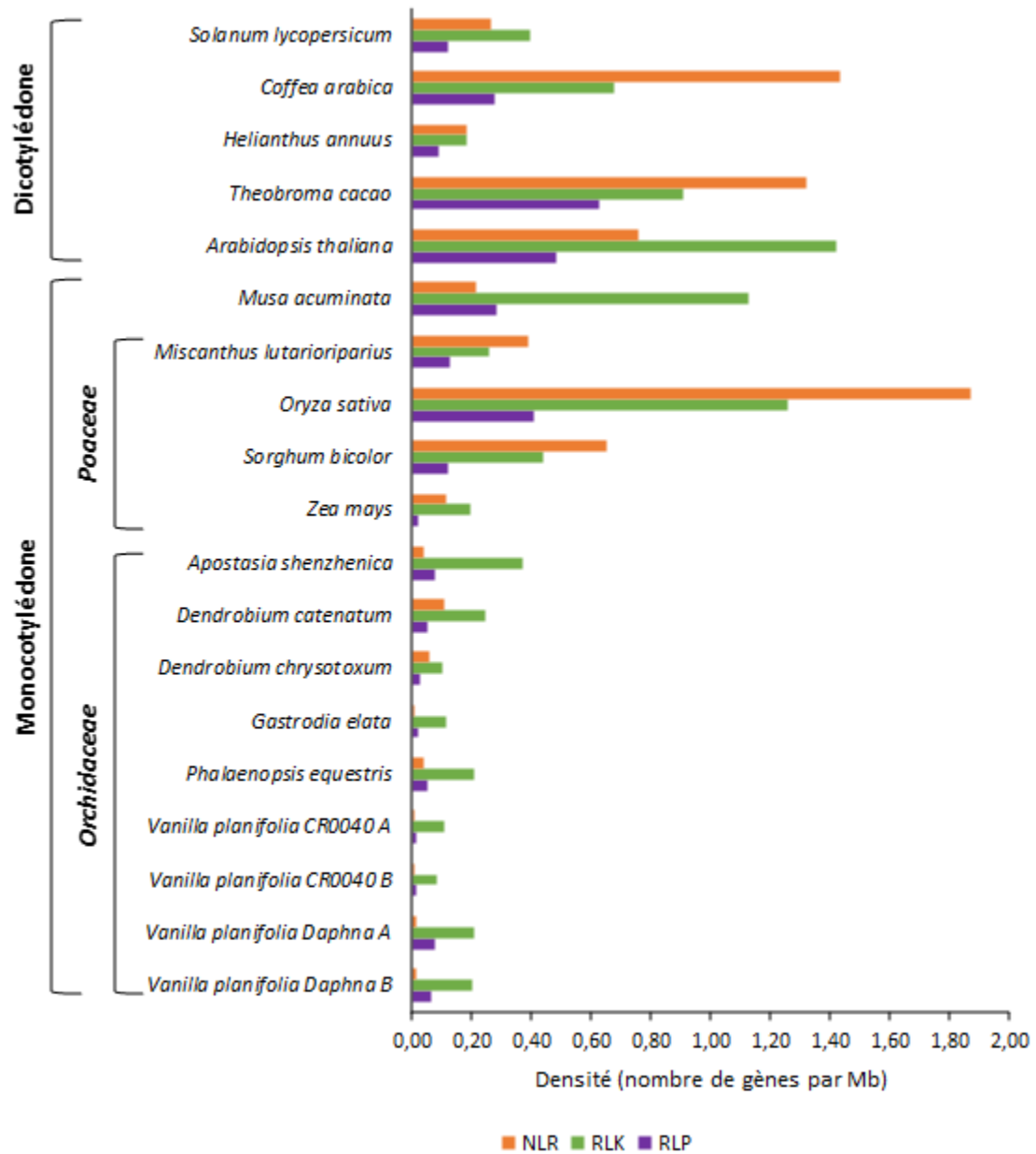
<sup>5</sup>B : Haplotype B



**Figure 5.2.** Localisation des gènes codant pour des gènes *LRR* sur les 14 chromosomes assemblés de (a) l'haplotype A et (b) de l'haplotype B du génome de *V. planifolia* CR0040. Les sous-familles *NLR*, *RLK* et *RLP* sont respectivement symbolisées par un rond orange, un triangle vert et un carré violet à droite des chromosomes. La densité en gènes par fenêtre de 500 Kb est représentée par un dégradé bleu à l'intérieur des chromosomes.

Les gènes *RLK* sont répartis sur tous les chromosomes. Le chromosome B04 présente un cluster de 4 *NLRs*, dont 2 se retrouvent dans la région homologue sur l'haplotype A. Les gènes *LRR* identifiés sur le chromosome 6 se regroupent dans une région située à l'extrémité du chromosome, dans laquelle sont annotés un cluster de gènes *RLP* et un *NLR* sur l'haplotype B (Figure 5.2). Cinq gènes *RLK* ont été identifiés sur le chromosome A0 et un seul *NLR* sur le chromosome B0. L'annotation de *V. planifolia* Daphna présente 225 gènes *LRR* sur l'haplotype A (12 *NLRs*, 156 *RLKs* et 57 *RLPs*) et 209 sur l'haplotype B (9 *NLRs*, 151 *RLKs* et 49 *RLPs*) (Tableau 5.1). Les *NLRs* identifiés chez CR0040 se retrouvent aussi chez Daphna. Cependant 6 *NLRs* de plus ont été annotés chez Daphna comparé à CR0040. Le nombre de *RLKs* identifiés à partir des deux annotations est similaire, alors qu'environ deux fois plus de gènes *RLP* ont été mis en évidence chez Daphna (106 *RLPs* contre 49 *RLPs* chez CR0040).

La densité en gènes *LRR* dans le génome a été calculée chez d'autres orchidées et des espèces plus ou moins éloignées de *V. planifolia* (Figure 5.3).

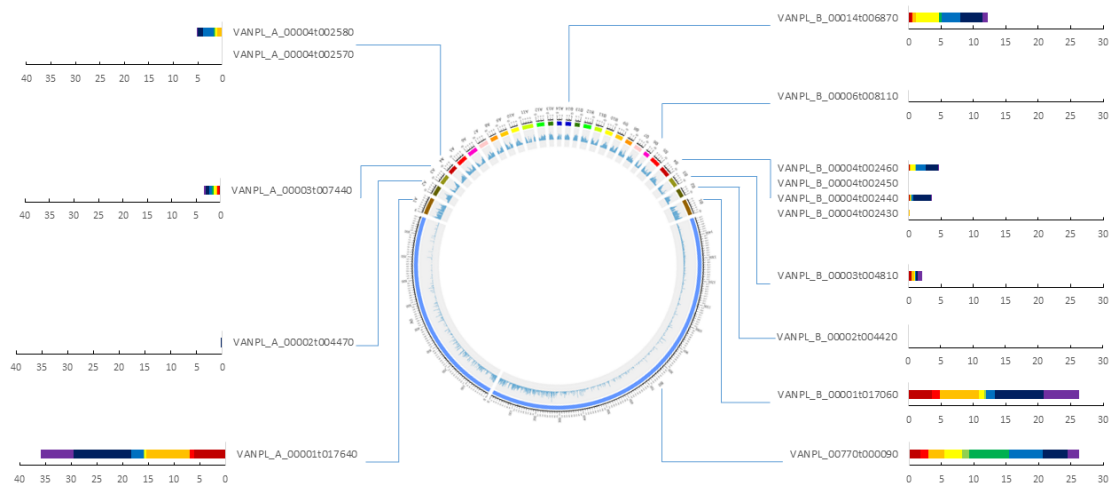


**Figure 5.3.** Densité des gènes codant pour des LRRs identifiés par LRRprofiler chez *V. planifolia* (génomes CR0040 et Daphna) et chez 15 autres espèces végétales. La densité est exprimée en nombre de gènes par Mb. Les sous-familles NLR, RLK et RLP sont respectivement représentées en orange, vert, et violet.

Le nombre de gènes *LRR* et la densité de ces gènes dans le génome sont en moyenne plus faibles chez les différentes espèces de la famille des *Orchidaceae* comparés aux autres espèces. C'est particulièrement le cas des *NLRs*, avec une densité qui varie de 0.003 (*G. elata*) à 0.11 (*D. catenatum*). Chez les autres monocotylédones choisies dans cette étude, la densité en *NLRs* varie de 0.12 (*Z. mays*) à 1.87 (*O. sativa*), et de 0.19 (*H.*



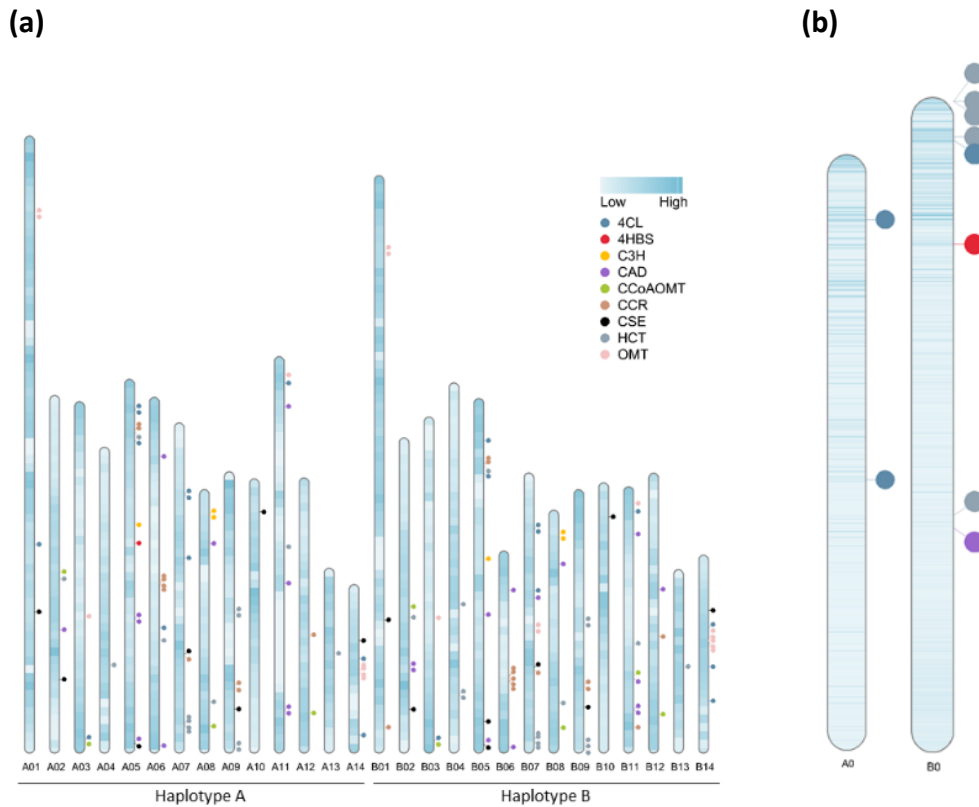
*annuus*) à 1.44 (*C. arabica*) chez les espèces dicotylédones. Les 15 gènes *NLR* identifiés dans le génome de l'accèsion sensible de *V. planifolia* CR0040 présentent des niveaux d'expression différents avec des valeurs de FPKM variant de 0 à 11.25 dans les différents organes (**Figure 5.4**). Les 6 gènes *NLR* dans les chromosomes 2, 4 et 6 présentent des niveaux d'expression négligeables voire nuls avec des valeurs de FPKM de 0 à 0.23 (VANPL\_A\_00002t004470, VANPL\_B\_00002t004420, VANPL\_A\_00004t002570, VANPL\_B\_00004t002430, VANPL\_B\_00004t002450, VANPL\_B\_00006t008110). Les *NLRs* sur le chromosome 1 (VANPL\_A\_00001t017640 et VANPL\_B\_00001t017060) et celui sur le chromosome B0 (VANPL\_00770t000090) présentent des niveaux d'expression relativement élevés avec des valeurs de FPKM variant de 0.1 à 11.25 dans les différents organes. Les niveaux d'expression les plus élevés sont observés au niveau des crampons, des racines, des ovaires et des tiges, et les plus faibles au niveau des boutons et des fleurs.



**Figure 5.4.** Localisation et expression (FPKM) par organe des gènes *NLR* sur le génome de *V. planifolia* CR0040. Les valeurs de FPKM sont indiquées pour les bourgeons (rouge foncé), les boutons (rouge), les crampons (orange), les feuilles (jaunes), les fleurs (vert), les fruits (vert foncé), les ovaires (bleu), les racines (bleu foncé) et les tiges (violet). Les chromosomes et la densité en gènes sont représentés au centre de la figure.

### 3.2. Gènes candidats potentiellement impliqués dans la résistance à *F. oxysporum*.

Par alignement de séquences, un pool de 897 gènes candidats potentiellement impliqués dans la résistance à Forv a été retrouvé dans le génome de CR0040 (**Annexe 1**). Les 15 gènes *NLR* de CR0040 identifiés par LRRprofiler sont ainsi retrouvés à partir des 11 séquences de gènes identifiées par Hasing *et al.* (2020) chez *Daphna* et des séquences d'autres espèces (e.g. *Cucumis melo*, *Brassica rapa*, *Z. mays*, *S. lycopersicum*, *Solanum bulbocastanum*, *M. acuminata*). Des *RLKs* ont aussi été identifiés à partir de séquences de *S. lycopersicum* impliqués dans la résistance à FORL (Fazio *et al.* 1999; Devran *et al.* 2018) et à partir de séquences de *A. thaliana* et *Solanum pennellii*. Enfin, les gènes potentiellement impliqués dans la résistance à *F. oxysporum* f. sp. *cubense* chez *M. acuminata* (Zhang *et al.* 2019) ont mis en évidence chez CR0040 des gènes jouant un rôle dans différents mécanismes de défense contre les champignons pathogènes, comme des chitinases, des protéinases, des lipases et des facteurs de transcription de type WRKY. Les gènes codant pour des enzymes de la voie de biosynthèse des phénylpropanoïdes ont été recherchés et 184 gènes ont été localisés sur le génome de CR0040 (**Figure 5.5**). Neuf familles de gènes sont représentées et se répartissent sur l'ensemble des chromosomes. Les réactions enzymatiques de la voie ont été schématisées avec les gènes associés et leurs valeurs d'expression (FPKM) dans chaque organe de CR0040 (**Figure 5.6**). Douze gènes de la famille des phénylalanine ammonia-lyases (PAL) ont été identifiés avec des valeurs de FPKM variant de 0 à 762.2. Les niveaux d'expression de ces gènes les plus élevés ont été observés dans les crampons, les fleurs et les fruits. Trente gènes ont été identifiés dans la famille des 4-coumarate-CoA ligases (4CL) avec des valeurs de FPKM variant de 0 à 305.5. Parmi eux, 3 gènes présentent des niveaux d'expression élevés dans les crampons, les fleurs et les fruits principalement (VANPL\_A\_00008t010830, VANPL\_B\_00008t010850 et VANPL\_B\_00014t006010).



**Figure 5.5.** Localisation des gènes de la voie de biosynthèse des phénylpropanoïdes **(a)** sur les 14 chromosomes assemblés (haplotypes A et B) et **(b)** sur les chromosomes A0 et B0 (scaffolds non assemblés) du génome de *V. planifolia* CR0040. Chaque famille de gènes est représentée à droite des chromosomes par une couleur différente. La densité en gènes par fenêtre de 500 Kb est représentée par un dégradé bleu à l'intérieur des chromosomes.

Les autres gènes présentent des niveaux d'expression faibles voire nuls. Trente-deux gènes de la famille des hydroxycinnamoyltransferases (HCT) ont été identifiés avec des niveaux d'expressions modérés (valeurs de FPKM variant de 0 à 109.6). Les niveaux les plus forts sont observés dans les boutons et les plus faibles dans les fruits. Le gène VANPL\_B\_00007t012440 présente l'expression la plus élevée de cette famille de gènes, avec des valeurs de FPKM fortes dans les bourgeons et les racines. Les 9 gènes de la famille des caffeoyl-CoA O-methyltransferases (CcoAMT) ont des valeurs de FPKM allant de 0 à 511. Deux gènes présentent de fortes valeurs d'expression (VAMPL\_A\_00008t010420 et VAMPL\_B\_00012t010050), et les valeurs de FPKM les plus élevées sont observées dans les crampons et les fleurs, tandis que les plus faibles sont observées dans les fruits. Les 23 gènes appartenant à la famille des cinnamoyl-CoA reductases (CCR) ont des valeurs de FPKM variant de 0 à 289.6 dans les différents

organes. Pour les deux gènes les plus exprimés (VANPL\_A\_00005t003910 et VANPL\_B\_00005t005050), les fleurs sont les organes dans lesquels le niveau d'expression est le plus élevé. La famille des O-triméthyltransferases (OMT) contient 19 gènes dans le génome de CR0040 avec des valeurs de FPKM variant de 0 à 2977.7. Parmi eux, le gène VANPL\_B\_00007t006130 est le gène le plus exprimé de cette voie de biosynthèse avec notamment une expression de 2977.7 dans les feuilles. La famille des cinnamyl alcohol dehydrogenases (CAD) regroupe 24 gènes qui ont des niveaux d'expression variant de 0 à 541.43 et qui sont globalement plus exprimés dans les racines. Par exemple, le gène VANPL\_B\_00002t009630 est le gène de cette voie le plus exprimé dans les racines. Enfin, une expression différentielle de tous ces gènes, et notamment entre les haplotypes A et B, suggère une régulation transcriptionnelle des gènes homologues.

#### 4. Discussion et perspectives

Dans ce chapitre, des gènes candidats potentiellement impliqués dans la résistance de *V. planifolia* à Forv ont été identifiés en explorant deux mécanismes potentiels :

- (i) La reconnaissance spécifique du pathogène par la plante induite suite à l'infection, avec la sollicitation de gènes de résistance, et donc la recherche de gènes *LRR* sur le génome de *V. planifolia*
- (ii) Le rôle de la voie de biosynthèse des phénylpropanoïdes dans le mécanisme de résistance en se basant sur l'observation d'un épaissement de la paroi de l'hypoderme en lignine chez les accessions résistantes à Forv.

L'annotation des domaines protéiques par LRRprofiler (Gottin et al. 2021) a permis d'identifier 381 gènes *LRR* dans l'annotation de CR0040 et 434 dans celle de Daphna. Hasing *et al.* (2020) ont mis en évidence 599 gènes *LRR* dans l'annotation de Daphna à partir de l'outil RGAugury (Li et al. 2016). Les différences d'annotation peuvent être expliquées par la différence d'outils utilisés. Cependant pour un même outil (LRRprofiler), 6 *NLRs* de plus et deux fois plus de *RLPs* ont été annotés chez Daphna.



Bien que la séquence de CR0040 soit la plus complète avec une estimation de 83% du génome couvert, moins de gènes ont été annotés que chez *Daphna* (Piet et al. 2022) suggérant des erreurs d'annotation automatique. Les *LRRs* sont distribués sur tous les chromosomes de CR0040 et sont majoritairement des *RLKs*. La famille des *RLKs* est une grande famille de gènes intervenant dans plusieurs grande fonctions cellulaires comme le développement et les réponses aux stress biotiques et abiotiques (Li et al. 2018). Des clusters de certains *LRRs* ont été mis en évidence, comme le cluster de *RLPs* sur l'extrémité du chromosome 6. Ces régions présentent un intérêt particulier dans la recherche de régions d'intérêt jouant un rôle dans la résistance.

Le faible nombre de gènes *LRR* identifié chez les vanilliers est cohérent avec celui retrouvé chez les autres *Orchidaceae* de mon étude qui présentent en moyenne 308 *LRRs* par génome alors que la moyenne est de 1122 chez les autres espèces analysées. Les autres monocotylédones (*Poaceae* et *Musaceae*) présentent aussi une densité en *LRRs* plus élevée que celle des orchidées. Ce déficit est particulièrement observé pour les gènes *NLR*. Un nombre de *NLRs* supérieur à 100 est généralement attendu chez les angiospermes (Shao et al. 2016). Cependant, la disparité de l'expression des gènes *NLR* entre les différentes espèces d'angiospermes est importante (Liu et al. 2021) et le nombre peut varier considérablement au sein d'une même famille (Luo et al. 2012; Shao et al. 2014; Zhang et al. 2016; Qian et al. 2017). Le nombre de *NLRs* chez les orchidées de cette étude varie de 3 (*G. elata*) à 122 (*D. catenatum*) et 15 *NLRs* ont été identifiés chez *V. planifolia* CR0040. Le faible nombre de *NLRs* dans la famille des *Orchidaceae* est cohérent avec l'évolution de cette famille de gènes chez les 4 orchidées *P. equestris*, *D. catenatum*, *G. elata* et *A. shenzhenica* (Xue et al. 2020). Les deux espèces *A. shenzhenica* et *V. planifolia* sont les plus proches phylogénétiquement et présentent un nombre similaire de *NLRs* avec respectivement 14 et 15 *NLRs* prédits. Xue et al. (2020) font l'hypothèse de la perte de gènes de résistance chez les orchidées qui pourrait s'expliquer comme une adaptation à une pression de sélection faible exercée par les pathogènes.



## CHAPITRE 6 |

### DISCUSSION GÉNÉRALE

---

#### 1. Conclusions

Les vanilliers, souvent considérés comme des espèces orphelines, ont reçu peu d'attention de la part des généticiens et des phytopathologistes, notamment depuis la production en masse de la vanilline de synthèse et l'arrêt des programmes d'amélioration variétale menés à Madagascar au 20<sup>ème</sup> siècle. Les travaux initiés depuis les années 2000 dans l'équipe « vanille » de l'UMR PVBMT dans laquelle j'ai effectué ce travail de thèse ont apporté des nouvelles connaissances notamment sur la diversité génétique des vanilliers en collection (Besse et al. 2004; Bory et al. 2008b) et sur la résistance à la fusariose, principale maladie affectant la culture des vanilliers. Un premier travail de grande ampleur de description de la diversité des souches de Forv et du mode d'infection (Koyyappurath 2015) a permis l'acquisition de connaissances majeures pour la poursuite d'études sur ce pathogène, qui appartient à un complexe d'espèces responsable de pertes économiques importantes chez d'autres grandes cultures tropicales. De plus, l'identification de sources de résistance chez *V. planifolia* a souligné le besoin de mieux appréhender le fond génétique de ces vanilliers cultivés afin d'identifier les facteurs génétiques impliqués dans le mécanisme de résistance.

L'apport de la génomique dans les travaux de ma thèse, avec le développement de nouvelles méthodes de génotypage à haut débit à l'échelle du génome, couplé au séquençage complet des génomes de *V. planifolia* CR0040 (Piet et al. 2022) et Daphna (Hasing et al. 2020), a permis de résoudre la diversité génétique intra-spécifique au sein des vanilliers cultivés, et de construire la première carte génétique de *V. planifolia* donnant accès à la localisation de régions d'intérêt et de gènes candidats. La séquence du génome de *V. planifolia* n'étant pas disponible au début de la thèse, le génotypage par séquençage est apparu comme la méthode la plus adaptée pour révéler les

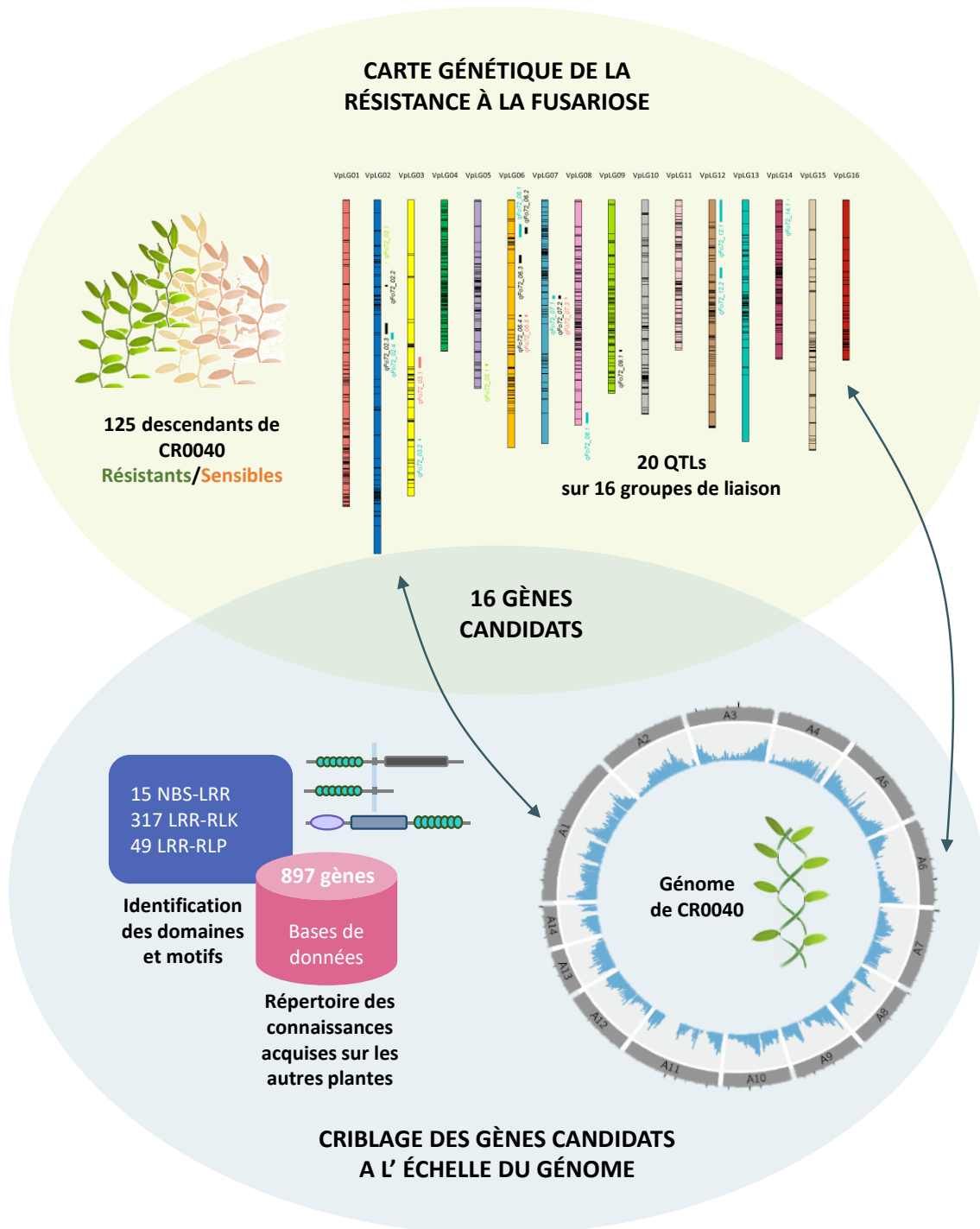


variations inter- et intra-spécifiques, avec respectivement 2004 et 1129 SNPs générés sans génome de référence. L'alignement de ces marqueurs sur l'assemblage de *Daphna* publié par Hasing et al. (2020) a révélé une distribution homogène de ces marqueurs sur le génome. La puissance de l'analyse GBS a permis de confirmer que *V. sotoarensii* est une espèce bien distincte de *V. planifolia*, et a montré pour la première fois que cette espèce sauvage est le parent maternel de l'hybride cultivé *V. x tahitensis*. Cette étude a également montré que le taux d'hétérozygotie est beaucoup plus élevé chez les espèces cultivées que chez les espèces sauvages apparentées, en conséquence d'un taux de mutations ponctuelles plus élevé chez les espèces clonales. Les données suggèrent une domestication de type « single-step » cohérente avec l'histoire d'introduction récente des vanilliers cultivés. Les variations génétiques intra-spécifiques ont pu être résolues grâce à la grande quantité de marqueurs, et notamment au sein d'une population d'autofécondation avec des individus génétiquement proches mais qui présentent un plus grand brassage génétique que chez les cultivars traditionnels. L'utilisation de nouvelles variétés issues d'autofécondation devrait contribuer à augmenter les niveaux de diversité et donc à réduire l'impact des mutations délétères avec une baisse des taux d'hétérozygotie.

La méthode GBS a aussi été privilégiée pour construire la première carte génétique à haute densité chez *V. planifolia*. L'assemblage du génome de CR0040 a été choisi comme référence pour identifier les SNPs pour deux raisons : (i) la séquence est plus complète que celle de *Daphna* avec une couverture du génome estimée à 83% (Piet et al. 2022) et (ii) CR0040 est le parent de la population d'autofécondation utilisée pour la cartographie génétique. Les fragments d'ADN séquencés ont donc été préalablement alignés sur le génome pour identifier les sites polymorphes dans la population. Un total de 4012 SNPs a été identifié par Stacks après l'alignement sur le génome de CR0040. Par comparaison, 1553 SNPs polymorphes ont été obtenus dans cette population par le pipeline *de novo* de Stacks (Catchen et al. 2013) sans génome de référence, ce qui est cohérent avec le nombre de 1129 SNPs utilisés pour étudier la diversité intra-spécifique présentée dans le chapitre 3 de cette thèse. L'utilisation de la séquence du parent de la population comme référence permet en effet de mieux aligner les fragments séquencés et donc de mieux identifier les variants. Le choix a été fait de

combiner ces deux méthodes de SNP calling afin d'augmenter la couverture du génome et accéder aux régions potentiellement absentes de l'assemblage. Pour autant, les marqueurs SNP *de novo* se répartissent bien sur l'ensemble du génome assemblé. Pour construire la carte génétique, les marqueurs similaires ont été retirés de l'étude afin de limiter les temps de calcul. Au final, 1804 marqueurs SNP parmi les plus informatifs et avec le moins de biais ont été gardés pour construire la carte génétique.

Dans cette thèse, deux approches sont abordées pour identifier des régions d'intérêt associées à la résistance à *F. oxysporum* (**Figure 6.1**), (i) une approche génétique à partir de la population d'autofécondation de CR0040 qui ségrège pour la résistance permettant la cartographie de QTLs liés à la résistance et (ii) et une approche globale à l'échelle du génome pour identifier avec *a priori* dans la séquence de CR0040 les structures et domaines conservés de familles de gènes de résistance et rechercher dans les bases de données publiques des gènes impliqués dans des mécanismes de résistance à *F. oxysporum* chez d'autres espèces. La création d'une population qui ségrège pour un caractère est la base des études génétiques. Un grand nombre de lignées fixées après plusieurs générations d'autofécondation d'une population F1 sont généralement utilisées pour identifier les recombinaisons génétiques à partir desquelles des cartes de liaisons génétiques sont développées. Cependant ces processus sont longs pour des plantes comme les vanilliers qui demandent un pas de temps de 8 à 10 ans entre la première étape de croisement et la caractérisation complète en champ des nouveaux génotypes adultes. Dans cette étude, une population de type F2, issue de l'autofécondation de CR0040, avait été générée et était disponible. CR0040 est un cultivar traditionnel et le taux d'hétérozygotie élevé observé dans le chapitre 3 a confirmé l'intérêt d'utiliser cette population qui ségrège pour la résistance à Forv. Le relativement faible nombre d'accessions de la population (125 accessions) est compensé par la grande quantité de SNPs obtenus et 16 groupes de liaison génétique ont pu être reconstruits, en accord avec les 16 chromosomes observés en cytogénétique (Bory et al. 2008a; Piet et al. 2022).



**Figure 6.1.** Bilan des régions d'intérêt identifiées pour la résistance à la fusariose au cours de ma thèse, par les approches combinées de génétique et de génomique.

Une seule souche de Forv (Fo072) a été testée dans le chapitre 4 et les 20 QTLs cartographiés expliquent chacun un pourcentage important de la variance phénotypique. Le gradient de sensibilité/résistance observé entre les différentes accessions de la population et la détection d'un grand nombre de QTLs significatifs confirment que nous étudions une résistance polygénique. Le pool de gènes candidats généré par l'approche globale à l'échelle du génome a mis en lumière différentes fonctions pouvant potentiellement jouer un rôle dans la résistance à Forv. Un gène *LRR-RLP*, sous-jacent au QTL du groupe de liaison 8, a notamment été identifié à l'extrémité du chromosome 6 dans un cluster de *LRR-RLPs*. Aucun gène de la voie de biosynthèse de la lignine ne se retrouve dans les régions génétiques cartographiées. Cependant, les données de transcriptomique de CR0040 infecté par Forv suggèrent une expression différentielle de ces gènes entre les haplotypes A et B.

## 2. Perspectives

### 2.1. Expression et diversité des gènes de résistance par des approches omiques

Ces travaux de thèse ont généré de nouvelles ressources génétiques et génomiques pour l'identification de facteurs génétiques impliqués dans la résistance à Forv. Les résultats obtenus doivent être validés et complétés, notamment en utilisant des approches à des échelles différentes, aujourd'hui rendues possibles par les nouvelles technologies de séquençage. Par exemple, l'expression de l'ensemble des gènes dans le génome peut être explorée grâce au séquençage de l'ARN (RNAseq) chez un individu donné. Les données d'expression dans les différents organes de CR0040 avaient été générées pour aider à l'annotation du génome. Ces données ont été présentées dans le chapitre 5 car CR0040 est sensible à Forv et de la fusariose a été observée sur les racines des plants utilisées pour cette expérience de transcriptomique. Les niveaux d'expression des gènes candidats correspondent donc à une plante sensible en condition d'infection. Cependant, seule une étude des profils transcriptomiques de

racines d'accessions résistantes et sensibles en condition d'inoculation contrôlée, permettrait de révéler l'expression différentielle des gènes sous-jacents. La première étude de transcriptomique *de novo* dans les racines de vanilliers a été réalisée au cours de la thèse de S. Koyyappurath (2015). Cinq gènes présentent une expression différentielle entre l'accession sensible *V. planifolia* CR0001 et les accessions résistantes *V. planifolia* CR0020 et *V. pompona* CR0018, en condition d'infection par la souche Fo072. En particulier, un gène sous-exprimé chez l'accession sensible a été annoté comme codant une protéine Pentatricopeptide repeat (PPR) mitochondriale. Ce résultat est intéressant au regard des 4 gènes codants pour des protéines PPR que nous avons identifiés dans les régions des QTLs. Les protéines PPR sont principalement connues pour leur rôle de régulation dans les processus de développement des plantes, en se liant aux ARNs de façon séquence-spécifique dans les mitochondries et les chloroplastes, mais aussi dans le noyau, afin d'influencer leur expression (Barkan and Small 2014; Manna 2015). Les PPRs, qui représentent une des plus grandes familles de gènes chez les plantes, interviennent également dans la réponse aux stress environnementaux. Des mutants de ces gènes chez *A. thaliana* montrent une plus grande sensibilité aux bactéries et champignons pathogènes (Park et al. 2014). Une autre étude montre l'implication des PPRs dans la régulation des espèces réactives de l'oxygène (ROS) qui interviennent dans la réponse hypersensible suite à l'infection par des champignons pathogènes (Laluk et al. 2011). Nos résultats ouvrent donc des pistes sur la fonction de cette famille de gènes dans la régulation des voies de signalisation en réponse à Forv. Les données transcriptomiques *de novo* préliminaires nécessitent d'être approfondies. S. Koyyappurath (2015) a réalisé ses prélèvements de racines à 5 et 6 jours après l'inoculation, au moment où les premiers symptômes sont observés, en accord avec ce que nous avons observé pendant le phénotypage de la descendance de CR0040. Cependant, une étude transcriptomique *de novo* publiée en 2019 conclut que le principal changement dans l'expression des gènes apparaît à 2 jpi comparé à 10 jpi (Solano-De la Cruz et al. 2019), et suggère que la réponse intervient dans les premières étapes de l'infection. L'étude des profils transcriptomiques dès les premières heures de l'infection apparaît donc nécessaire pour identifier les mécanismes de défense très précoces mis en place par la plante.

L'approche génomique initiée au cours de cette thèse et présentée dans le chapitre 5 doit être complétée. Dans un premier temps, il conviendra d'améliorer l'annotation du génome de CR0040. En effet, la grande majorité des marqueurs SNP couvre les 14 chromosomes assemblés de CR0040 qui contiennent approximativement deux tiers des gènes. Cependant, très peu de SNPs se retrouvent sur les chromosomes 0 sur lesquels un tiers des gènes a été annoté. Piet et al (2022) ont en outre estimé qu'environ 17% du génome n'était pas couvert dans l'assemblage de CR0040 et que des gènes étaient donc potentiellement manquants. De plus, 6 *NLRs* et 2 fois plus de *RLPs* ont été identifiés par l'outil d'annotation LRRprofiler chez *Daphna* comparés à ceux chez CR0040. L'outil se base sur les fichiers protéiques pour rechercher des structures et domaines LRR. Ces résultats suggèrent une annotation automatique incomplète de CR0040. Une annotation manuelle fine de cette famille de gènes aux structures complexes permettrait de mieux comprendre les caractéristiques structurales et évolutives des gènes *LRR*. Peu de *NLRs* ont été identifiés chez *V. planifolia*, en accord avec l'évolution de ces gènes dans la famille des *Orchidaceae* (Xue et al. 2020). Cependant, leur expression à des niveaux faibles dans les tissus complique leur annotation. Un autre outil, appelé NLR-Annotator, s'affranchit des données de transcriptomique en recherchant les structures codant pour les motifs LRR et NB-ARC directement sur la séquence nucléique (Steuernagel et al. 2020). L'utilisation de cet outil permettrait d'identifier des clusters de gènes *NLR* difficilement annotables. Des outils comme RGAugury (Li et al. 2016) permettraient de rechercher d'autres familles de gènes telles que les récepteurs transmembranaires qui possèdent le motif LysM impliqués dans la reconnaissance de la paroi des champignons.

Une analyse de génomique comparative permettrait d'étudier la dynamique d'évolution des gènes de résistance chez le vanillier (voies de biosynthèse de la lignine, *LRR*...) et des gènes *PPR*, en comparant avec les génomes séquencés d'autres monocotylédones disponibles dans les bases de données. Pour identifier les gènes conservés vs spécifiques et comprendre l'histoire évolutive qui a conduit à l'organisation actuelle de ces gènes, de nombreux outils pourraient être utilisés et développés, notamment SynMap et GEvo disponibles sur la plateforme CoGe (<https://genomevolution.org/coge/>) et GreenPhylDB (<http://www.greenphylog.org/cgi->

bin/index.cgi), une ressource Web basée sur la phylogénétique afin de prédire les relations d'homologie entre gènes, qui prend en charge les études évolutives et fonctionnelles pour identifier les gènes candidats ayant une incidence sur les traits agronomiques. De plus, le genome hub (<https://vanilla-genome-hub.cirad.fr/>) regroupe un grand nombre d'outils pour réaliser ces analyses à partir de l'ensemble des données omiques produites par l'équipe. L'arrivée d'un génome de référence pour *V. planifolia* facilite également le re-séquençage d'espèces du genre *Vanilla* représentatives de la diversité. Le re-séquençage de l'espèce résistante *V. pompona* initié par Hasing et al. (2020) a permis de mettre en évidence des gènes potentiellement impliqués par comparaison avec l'annotation de *Daphna*. Nous pouvons également envisager un re-séquençage d'espèces résistantes comme *V. pompona*, afin de comparer avec le génome de CR0040 sensible à la maladie, ou bien un re-séquençage de descendants de CR0040 bien caractérisés dans cette thèse, en choisissant un phénotype extrême pour la résistance et un autre pour la sensibilité. La disponibilité du matériel végétal au sein du CRB Vatel (CIRAD, La Réunion) et l'expertise acquise dans l'analyse des grands jeux de données omiques dans l'équipe ces dernières années pourront permettre de caractériser plus facilement et rapidement les gènes d'intérêt agronomiques.

## 2.2. Spectre d'action et durabilité de la résistance des vanilliers

Seule la souche Fo072 hautement pathogène a été utilisée pour étudier l'interaction vanillier/Forv dans cette thèse. Cependant, une cinquantaine d'isolats de Forv a été identifiée dans les sols réunionnais, présentant un schéma d'évolution hautement polyphylétique et un gradient continu de pathogénicité (Koyyappurath et al. 2015a). L'étude du spectre d'action de la résistance vis-à-vis d'une large diversité de souches de Forv permettrait de mieux comprendre la spécificité/généricité des mécanismes moléculaires mis en place suite à l'infection. La disponibilité de la carte génétique, de la population d'autofécondation de CR0040 et des souches Forv en collection constitue une base solide pour cartographier des marqueurs génétiques liés aux facteurs de résistance à ces souches et déterminer la spécificité de la résistance. Au cours de ma

thèse, 19 accessions de la population ont été testées pour la résistance à trois souches réunionnaises, hautement pathogènes mais de groupes phylogénétiques différents (Fo072, Fo294 et Fo254) (résultats non montrés ; Declas, 2021). Trois de ces accessions présentent un profil de résistance élevé aux 3 souches, suggérant l'hypothèse d'une résistance généraliste. Des avancées notoires ont été obtenues en génétique/génomique vanillier, mais très peu de connaissances sont disponibles sur l'épidémiologie et les facteurs de pathogénicité du champignon, connaissances pourtant indispensables pour élaborer une stratégie de gestion durable de la maladie.

Les résistances polygéniques contrôlées par plusieurs gènes, chacun contribuant à une part plus ou moins importante dans le phénotype de résistance, sont considérées comme plus durables et donc, d'une certaine façon, plus intéressantes que les résistances conférées par un seul gène majeur (Pilet-Nayel et al. 2017). En effet, il est plus difficile pour le pathogène de contourner plusieurs gènes au lieu d'un seul, et les pathogènes qui réussissent à contourner les QTLs ne gagneraient qu'un avantage marginal (Poland et al. 2009). Cependant, l'amélioration variétale durable basée sur la résistance polygénique peut s'avérer difficile à mettre en œuvre, avec la nécessité de pyramider les différents QTLs identifiés, et plus particulièrement ceux qui sont impliqués dans différents mode d'action, pour augmenter le spectre de la résistance. Cette stratégie a été notamment proposée pour la résistance durable du blé contre *Fusarium pseudograminearum* (Bovill et al. 2010) et *Fusarium graminearum* (Hu et al. 2022). Sur vanillier, aucun programme de pyramidage de gènes, ni pour la résistance aux maladies, ni pour des qualités agronomiques (rendement, floraison...) n'a à ce jour été publié.

### **2.3. Vers une culture de vanilliers résiliente dans un contexte de changements environnementaux**

Longtemps considérée comme une plante « orpheline », le vanillier connaît un regain d'intérêt, d'une part pour la qualité aromatique de ses fruits très recherchée pour



proposer une vanille naturelle et non plus de synthèse, et d'autre part pour la mise en lumière récente de cette plante atypique d'un point de vue scientifique. L'arrivée de nouvelles connaissances génétiques et génomiques, et une meilleure caractérisation et conservation des ressources biologiques, placent *V. planifolia* dans une nouvelle ère en devenant une espèce modèle pour le genre *Vanilla* et la famille des *Orchidaceae*. Ces récentes avancées nous permettent également d'envisager la mise en place de programmes d'amélioration variétale pour des caractéristiques d'intérêt comme la résistance à la fusariose. Le séquençage du génome de *V. planifolia* et maintenant la première cartographie génétique de la résistance donnent accès aux séquences de gènes d'intérêt. Des marqueurs moléculaires « breeder-friendly » pourraient être développés dans le cadre de la sélection assistée par marqueurs.

Les résultats de la thèse confirment le potentiel de la résistance dans la gestion de la maladie. Cependant, pour augmenter l'efficacité de la lutte génétique, il est important de continuer les recherches sur le pathogène. En effet, la culture du vanillier en monoculture favorise le risque d'épidémies massives par les populations de *Fusarium* présentes dans les sols (Xiong et al. 2015). L'intérêt de préserver la culture traditionnelle en sous-bois ou sur support naturel s'avère être une stratégie complémentaire à la lutte génétique pour une gestion durable de la maladie. Dans un contexte de changement climatique important, avec des projections d'augmentation des températures et du nombre de jours consécutifs sans pluie dans les zones de production de vanille à Madagascar notamment (source ©EcoAct, confidentiel), il s'avère d'autant plus essentiel de mieux caractériser la diversité génétique naturelle pour faire face aux nouvelles contraintes abiotiques et possiblement biotiques. Des vanilliers aphylls sauvages endémiques présents dans des zones plus arides à Madagascar, aux Comores, à Mayotte, en Tanzanie et aux Seychelles (Andriamihaja et al., 2020) représentent un matériel de choix pour rechercher des caractères d'intérêt, et ainsi utiliser ces connaissances pour la sélection de vanilliers aromatiques mieux adaptés aux changements environnementaux.

D'autre part, l'utilisation de stratégies de biocontrôle, utilisant notamment des microorganismes antagonistes, représente une voie encore peu explorée pour

contribuer à la diminution de l'impact de la fusariose. Récemment, des solutions de biocontrôle contre la fusariose vasculaire chez les céréales, basées sur l'utilisation de micro-organismes endophytes comme des bactéries du genre *Bacillus* et des champignons filamenteux, ont été proposées par des sociétés privées (<https://www.cultivar.fr/sinformer/des-espoirs-de-biocontrole-pour-lutter-contre-la-fusariose-des-epis>), mais peu d'études scientifiques expliquent les mécanismes impliqués et l'efficacité de ces méthodes. Par ailleurs, il a été montré chez la tomate que l'utilisation de souches non-pathogènes de *F. oxysporum* pouvaient réduire l'impact de maladies causées par des champignons vasculaires tels que *Verticillium dahliae* et des souches de *Fusarium* pathogènes (de Lamo and Takken 2020). Le développement de telles stratégies chez le vanillier via le criblage d'agents de biocontrôle en conditions contrôlées confinées au laboratoire puis au champ pourraient être combinées à une évaluation des pratiques agroécologiques dans les plantations de vanilliers (impact du compost, de la densité de plantation, des mélanges variétaux dans les parcelles...). Des expérimentations à grande échelle dans les lieux de production de la zone SOOI pourraient être réalisées sur des géotypes résistants à la fusariose, dont les individus issus de l'autofécondation de CR0040 caractérisés au cours de ma thèse.



## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- Agrios GN (1997) Plant pathology, 4th ed. Academic Press, San Diego
- Ahmad F, Martawi NM, Poerba YS, et al (2020) Genetic mapping of Fusarium wilt resistance in a wild banana *Musa acuminata* ssp. *malaccensis* accession. *Theor Appl Genet* 133:3409–3418. <https://doi.org/10.1007/s00122-020-03677-y>
- Alomia YA, Chambers A, Brym M, et al (2021) Genotyping-By-Sequencing diversity analysis of international *Vanilla* collections uncovers hidden diversity and enables plant improvement. *Plant Science* 111019. <https://doi.org/10.1016/j.plantsci.2021.111019>
- Alseekh S, Kostova D, Bulut M, Fernie AR (2021) Genome-wide association studies: assessing trait characteristics in model and crop plants. *Cell Mol Life Sci* 78:5743–5754. <https://doi.org/10.1007/s00018-021-03868-w>
- Andriamihaja C, Ramarosandratana V, Grisoni M, Jeannoda V, Besse P (2020). The leafless *Vanilla* species-complex from the South-West Indian Ocean Region: A taxonomic puzzle and a model for orchid evolution and conservation research. *Diversity*, 12:443, 25 p. <https://doi.org/10.3390/d12120443>
- Barkan A, Small I (2014) Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol* 65:415–442. <https://doi.org/10.1146/annurev-arplant-050213-040159>
- Barros J, Escamilla-Trevino L, Song L, et al (2019) 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nat Commun* 10:1994. <https://doi.org/10.1038/s41467-019-10082-7>
- Besse P, Silva DD, Bory S, et al (2004) RAPD genetic diversity in cultivated vanilla: *Vanilla planifolia*, and relationships with *V. tahitensis* and *V. pompona*. *Plant Science* 167:379–385. <https://doi.org/10.1016/j.plantsci.2004.04.007>
- Bory S, Catrice O, Brown S, et al (2008a) Natural polyploidy in *Vanilla planifolia* (Orchidaceae). *Genome* 51:816–826. <https://doi.org/10.1139/G08-068>
- Bory S, Lubinsky P, Risterucci A-M, et al (2008b) Patterns of introduction and diversification of *Vanilla planifolia* (Orchidaceae) in Reunion Island (Indian Ocean). *American Journal of Botany* 95:805–815. <https://doi.org/10.3732/ajb.2007332>
- Bouetard A, Lefeuvre P, Gigant R, et al (2010) Evidence of transoceanic dispersion of the genus *Vanilla* based on plastid DNA phylogenetic analysis. *Molecular Phylogenetics and Evolution* 55:621–630. <https://doi.org/10.1016/j.ympev.2010.01.021>

- Bouriquet G (1954) Le vanillier et la vanille dans le monde, Lechevalier
- Bovill WD, Horne M, Herde D, et al (2010) Pyramiding QTL increases seedling resistance to crown rot (*Fusarium pseudograminearum*) of wheat (*Triticum aestivum*). *Theor Appl Genet* 121:127–136. <https://doi.org/10.1007/s00122-010-1296-7>
- Branham SE, Levi A, Wechter WP (2019) QTL Mapping Identifies Novel Source of Resistance to *Fusarium* Wilt Race 1 in *Citrullus amarus*. *Plant Disease* 103:984–989. <https://doi.org/10.1094/PDIS-09-18-1677-RE>
- Brown SC, Bourge M, Maunoury N, et al (2017) DNA Remodeling by Strict Partial Endoreplication in Orchids, an Original Process in the Plant Kingdom. *Genome Biology and Evolution* 9:1051–1071. <https://doi.org/10.1093/gbe/evx063>
- Camacho C, Coulouris G, Avagyan V, et al (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
- Catanzariti A, Do HTT, Bru P, et al (2017) The tomato I gene for *Fusarium* wilt resistance encodes an atypical leucine-rich repeat receptor-like protein whose function is nevertheless dependent on SOBIR 1 and SERK 3/ BAK 1. *Plant J* 89:1195–1209. <https://doi.org/10.1111/tpj.13458>
- Catchen J, Hohenlohe PA, Bassham S, et al (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140. <https://doi.org/10.1111/mec.12354>
- Chakravarti A, Lasher LK, Reefer JE (1991) A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128:175–182. <https://doi.org/10.1093/genetics/128.1.175>
- Charron C, Hubert J, Minatchy J, et al (2018) Characterization of *Colletotrichum orchidophilum*, the agent of black spot disease of vanilla. *J Phytopathol* 166:525–531. <https://doi.org/10.1111/jph.12714>
- Chinchilla D, Shan L, He P, et al (2009) One for all: the receptor-associated kinase BAK1. *Trends in Plant Science* 14:535–541. <https://doi.org/10.1016/j.tplants.2009.08.002>
- Chinchilla D, Zipfel C, Robatzek S, et al (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* 448:497–500. <https://doi.org/10.1038/nature05999>
- Constantin D, Bois J (1915) Sur trois types de vanilles commerciales de TAHITI. *Comptes Rendus de l'Académie des sciences de Paris* 161:196–202
- Correll JC, Klittich CJR, Leslie JF (1987) Nitrate nonutilizing mutants of *Fusarium oxysporum* and their use in vegetative compatibility tests. *The American Phytopathological Society* 1640–1646

- Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Dangl JL, Horvath DM, Staskawicz BJ (2013) Pivoting the Plant Immune System from Dissection to Deployment. *Science* 341:746–751. <https://doi.org/10.1126/science.1236011>
- de Lamo FJ, Takken FLW (2020) Biocontrol by *Fusarium oxysporum* Using Endophyte-Mediated Resistance. *Front Plant Sci* 11:37. <https://doi.org/10.3389/fpls.2020.00037>
- de Mendiburu F (2021) *Agricolae: Statistical Procedures for Agricultural Research*
- Declas S (2021) Evaluation du spectre de résistance de vanilliers issus d’une population d’autofécondation à différentes souches du champignon *Fusarium oxysporum*
- Delassus M (1963) La lutte contre la fusariose du vanillier par les méthodes génétiques. *L’Agronomie Tropicale Série 2, Agronomie Générale Etudes Techniques* 18:245–246
- Dequaire J (1976) L’amélioration du vanillier à Madagascar. *Journal d’agriculture traditionnelle et de botanique appliquée* 23:139–158. <https://doi.org/10.3406/jatba.1976.3246>
- Devran Z, Kahveci E, Hong Y, et al (2018) Identifying molecular markers suitable for Frl selection in tomato breeding. *Theor Appl Genet* 131:2099–2105. <https://doi.org/10.1007/s00122-018-3136-0>
- Divakaran M, Pillai G, Nirmal Babu K, Peter K (2008) Isolation and fusion of protoplasts in *Vanilla* species. *Current Science* 94:115–120
- Dixon RA, Achnine L, Kota P, et al (2002) The phenylpropanoid pathway and plant defence—a genomics perspective: Phenylpropanoids and plant defence. *Molecular Plant Pathology* 3:371–390. <https://doi.org/10.1046/j.1364-3703.2002.00131.x>
- Duval M-F, Bory S, Andrzejewski S, et al (2006) Diversité génétique des vanilliers dans leurs zones de dispersion secondaire. In: *Les ressources génétiques : des ressources partagées : 6ème Colloque national BRG, La Rochelle, 2-4 octobre 2006*. <https://agritrop.cirad.fr/534974/>. Accessed 11 Jan 2022
- Edel-Hermann V, Lecomte C (2019) Current Status of *Fusarium oxysporum* Formae Speciales and Races. *Phytopathology*® 109:512–530. <https://doi.org/10.1094/PHYTO-08-18-0320-RVW>
- Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- FAOSTAT (2017) Production quantities of vanilla by country.

<http://www.fao.org/faostat/en/?#data/QC/visualize>.

- Favre F (2019) Exploitation de données génotypage par séquençage (GBS) pour l'analyse de la diversité génétique de *Vanilla* spp. et la recherche de marqueurs moléculaires associés à la résistance à la fusariose. Université Paris-Saclay. 66 pages
- Favre F, Jourda C, Grisoni M, et al (2022) A genome-wide assessment of the genetic diversity, evolution and relationships with allied species of the clonally propagated crop *Vanilla planifolia* Jacks. ex Andrews. *Genet Resour Crop Evol* 1–15. <https://doi.org/10.1007/s10722-022-01362-1>
- Fazio G, Stevens MR, Scott JW (1999) Identification of RAPD markers linked to fusarium crown and root rot resistance (Frl) in tomato. *Euphytica* 105:205–210. <https://doi.org/10.1023/A:1003497719705>
- Fishman L, Kelly AJ, Morgan E, Willis JH (2001) A Genetic Map in the *Mimulus guttatus* Species Complex Reveals Transmission Ratio Distortion due to Heterospecific Interactions. *Genetics* 159:1701–1716. <https://doi.org/10.1093/genetics/159.4.1701>
- Fock-Bastide I, Palama TL, Bory S, et al (2014) Expression profiles of key phenylpropanoid genes during *Vanilla planifolia* pod development reveal a positive correlation between PAL gene expression and vanillin biosynthesis. *Plant Physiology and Biochemistry* 74:304–314. <https://doi.org/10.1016/j.plaphy.2013.11.026>
- Geng S, Li A, Tang L, et al (2013) TaCPK2-A, a calcium-dependent protein kinase gene that is required for wheat powdery mildew resistance enhances bacterial blight resistance in transgenic rice. *Journal of Experimental Botany* 64:3125–3136. <https://doi.org/10.1093/jxb/ert146>
- Gigant R, Bory S, Grisoni M, Besse P (2011) Biodiversity and Evolution in the *Vanilla* Genus. In: *The dynamical processes of biodiversity - case studies of evolution and spatial distribution*, InTech. pp 1–26
- Gottin C, Dievart A, Summo M, et al (2021) A new comprehensive annotation of leucine-rich repeat-containing receptors in rice. *Plant J* tpj.15456. <https://doi.org/10.1111/tpj.15456>
- Guo T, Kong F, Burton C, et al (2022) The E3 Ubiquitin Ligase ATL9 Affects Expression of Defense Related Genes, Cell Death and Callose Deposition in Response to Fungal Infection. *Pathogens* 11:68. <https://doi.org/10.3390/pathogens11010068>
- Gururani MA, Venkatesh J, Upadhyaya CP, et al (2012) Plant disease resistance genes: Current status and future directions. *Physiological and Molecular Plant Pathology* 78:51–65. <https://doi.org/10.1016/j.pmpp.2012.01.002>

- Hao Z, Lv D, Ge Y, et al (2020) RIdiogram : drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science* 6:e251. <https://doi.org/10.7717/peerj-cs.251>
- Hasing T, Tang H, Brym M, et al (2020) A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nat Food* 1:811–819. <https://doi.org/10.1038/s43016-020-00197-2>
- Hendre PS, Muthemba S, Kariba R, et al (2019) African Orphan Crops Consortium (AOCC): status of developing genomic resources for African orphan crops. *Planta* 250:989–1003. <https://doi.org/10.1007/s00425-019-03156-9>
- Hu W, Fu L, Gao D, et al (2022) Pyramiding Three QTL Improved Fusarium Head Blight Resistance in Wheat (*Triticum Aestivum* L.). In Review
- Hu Y, Resende MFR, Bombarely A, et al (2019) Genomics-based diversity analysis of *Vanilla* species using a *Vanilla planifolia* draft genome and Genotyping-By-Sequencing. *Sci Rep* 9:3416. <https://doi.org/10.1038/s41598-019-40144-1>
- Jamnadass R, Mumm RH, Hale I, et al (2020) Enhancing African orphan crops with genomics. *Nat Genet* 52:356–360. <https://doi.org/10.1038/s41588-020-0601-x>
- Jarvis WR, Shoemaker RA (1978) Taxonomic status of *Fusarium oxysporum* causing foot and root rot of tomato
- Jeger MJ, Viljanen-Rollinson SLH (2001) The use of the area under the disease-progress curve (AUDPC) to assess quantitative disease resistance in crop cultivars: *Theor Appl Genet* 102:32–40. <https://doi.org/10.1007/s001220051615>
- Jones JDG, Dangl JL (2006) The plant immune system. *Nature* 444:323–329. <https://doi.org/10.1038/nature05286>
- Karban R, Baldwin IT (1997) *Induced Responses to Herbivory*. University of Chicago Press
- Keen N (1999) Mechanisms of Pest Resistance in Plants. In: *Workshop on Ecological Effects of Pest Resistance Genes in Managed Ecosystem*, Bethesda
- Kemmerling B, Schwedt A, Rodriguez P, et al (2007) The BRI1-Associated Kinase 1, BAK1, Has a Brassinolide-Independent Role in Plant Cell-Death Control. *Current Biology* 17:1116–1122. <https://doi.org/10.1016/j.cub.2007.05.046>
- Knaus BJ, Grünwald NJ (2017) VCFR : a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17:44–53. <https://doi.org/10.1111/1755-0998.12549>
- Knudson L (1950) GERMINATION OF SEEDS OF VANILLA. *American Journal of Botany* 37:241–247. <https://doi.org/10.1002/j.1537-2197.1950.tb12189.x>
- Koyyappurath S (2015) Histological and molecular approaches for resistance to



- Fusarium oxysporum* f.sp. *radicis-vanillae*, causal agent of root and stem rot in *Vanilla* spp. (Orchidaceae). Université de la Réunion
- Koyyappurath S, Atuahiva T, Le Guen R, et al (2015a) *Fusarium oxysporum* f. sp. *radicis-vanillae* is the causal agent of root and stem rot of vanilla. *Plant Pathol* 65:612–625. <https://doi.org/10.1111/ppa.12445>
- Koyyappurath S, Conéjéro G, Dijoux JB, et al (2015b) Differential Responses of *Vanilla* Accessions to Root Rot and Colonization by *Fusarium oxysporum* f. sp. *radicis-vanillae*. *Front Plant Sci* 6:. <https://doi.org/10.3389/fpls.2015.01125>
- Laluk K, AbuQamar S, Mengiste T (2011) The Arabidopsis Mitochondria-Localized Pentatricopeptide Repeat Protein PGN Functions in Defense against Necrotrophic Fungi and Abiotic Stress Tolerance. *Plant Physiology* 156:2053–2068. <https://doi.org/10.1104/pp.111.177501>
- Lepers-Andrzejewski S, Causse S, Caromel B, et al (2012) Genetic Linkage Map and Diversity Analysis of Tahitian *Vanilla* (*Vanilla × tahitensis*, Orchidaceae). *Crop Science* 52:795–806. <https://doi.org/10.2135/cropsci2010.11.0634>
- Lewin HA, Larkin DM, Pontius J, O'Brien SJ (2009) Every genome sequence needs a good map. *Genome Res* 19:1925–1928. <https://doi.org/10.1101/gr.094557.109>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li P, Quan X, Jia G, et al (2016) RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* 17:852. <https://doi.org/10.1186/s12864-016-3197-x>
- Li X, Salman A, Guo C, et al (2018) Identification and Characterization of LRR-RLK Family Genes in Potato Reveal Their Involvement in Peptide Signaling of Cell Fate Decisions and Biotic/Abiotic Stress Responses. *Cells* 7:120. <https://doi.org/10.3390/cells7090120>
- Liu Y, Zeng Z, Zhang Y-M, et al (2021) An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. *Molecular Plant* 14:2015–2031. <https://doi.org/10.1016/j.molp.2021.08.001>
- Lozano-Isla F (2022) inti: Tools and Statistical Procedures in Plant Science. R package version 0.5.7, <https://CRAN.R-project.org/package=inti>.
- Lubinsky P, Cameron KM, Molina MC, et al (2008) Neotropical roots of a Polynesian spice: the hybrid origin of Tahitian vanilla, *Vanilla tahitensis* (Orchidaceae). *American Journal of Botany* 95:1040–1047. <https://doi.org/10.3732/ajb.0800067>

- Luo S, Zhang Y, Hu Q, et al (2012) Dynamic Nucleotide-Binding Site and Leucine-Rich Repeat-Encoding Genes in the Grass Family. *Plant Physiology* 159:197–210. <https://doi.org/10.1104/pp.111.192062>
- Ma Y, Chhapekar SS, Lu L, et al (2021) QTL mapping for Fusarium wilt resistance based on the whole-genome resequencing and their association with functional genes in *Raphanus sativus*. *Theor Appl Genet* 134:3925–3940. <https://doi.org/10.1007/s00122-021-03937-5>
- Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* 113:93–99. <https://doi.org/10.1016/j.biochi.2015.04.004>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 17:10. <https://doi.org/10.14806/ej.17.1.200>
- Mazhar MS (2022) Vanilla production in Australia: Feasibility report. AgriFutures Australia
- Morin M (2021) Annotation et analyse de l'expression différentielle de gènes candidats impliqués dans la résistance des vanilliers au *Fusarium oxysporum*
- Mt S-D la C, J A-G, J G-J, et al (2019) Functional categorization of de novo transcriptome assembly of *Vanilla planifolia* Jacks. potentially points to a translational regulation during early stages of infection by *Fusarium oxysporum* f. sp. *vanillae*. *BMC genomics* 20:. <https://doi.org/10.1186/s12864-019-6229-5>
- Nelson PE (1981) Life Cycle and Epidemiology of *Fusarium oxysporum*. In: *Fungal Wilt Diseases of Plants*. Elsevier, pp 51–80
- Nicholson RL, Hammerschmidt R (1992) Phenolic Compounds and Their Role in Disease Resistance. *Annual Review of Phytopathology* 30:369–389. <https://doi.org/10.1146/annurev.py.30.090192.002101>
- Odoux E, Grisoni M (eds) (2010) *Vanilla*, 0 edn. CRC Press
- Olivain C, Humbert C, Nahalkova J, et al (2006) Colonization of Tomato Root by Pathogenic and Nonpathogenic *Fusarium oxysporum* Strains Inoculated Together and Separately into the Soil. *Applied and Environmental Microbiology* 72:1523–1531. <https://doi.org/10.1128/AEM.72.2.1523-1531.2006>
- Park YJ, Lee HJ, Kwak KJ, et al (2014) MicroRNA400-Guided Cleavage of Pentatricopeptide Repeat Protein mRNAs Renders *Arabidopsis thaliana* More Susceptible to Pathogenic Bacteria and Fungi. *Plant and Cell Physiology* 55:1660–1668. <https://doi.org/10.1093/pcp/pcu096>
- Perez-Nadales E, Almeida Nogueira MF, Baldin C, et al (2014) Fungal model systems and the elucidation of pathogenicity determinants. *Fungal Genetics and Biology* 70:42–67. <https://doi.org/10.1016/j.fgb.2014.06.011>

- Piet Q, Droc G, Marande W, et al (2022) A chromosome-level, haplotype-phased *Vanilla planifolia* genome highlights the challenge of partial endoreplication for accurate whole-genome assembly. *Plant Communications* 100330. <https://doi.org/10.1016/j.xplc.2022.100330>
- Pilet-Nayel M-L, Moury B, Caffier V, et al (2017) Quantitative Resistance to Plant Pathogens in Pyramiding Strategies for Durable Crop Protection. *Front Plant Sci* 8:1838. <https://doi.org/10.3389/fpls.2017.01838>
- Pilloni R (2017) Analyse transcriptionnelle de gènes potentiellement impliqués dans la résistance de *Vanilla planifolia* à For
- Poland JA, Balint-Kurti PJ, Wisser RJ, et al (2009) Shades of gray: the world of quantitative disease resistance. *Trends in Plant Science* 14:21–29. <https://doi.org/10.1016/j.tplants.2008.10.006>
- Portères R (1954) Le genre *Vanilla* et ses espèces. Le vanillier et la vanille dans le monde. Paris, Editions Paul Lechevalier XL VI:94–290
- Qian L-H, Zhou G-C, Sun X-Q, et al (2017) Distinct Patterns of Gene Gain and Loss: Diverse Evolutionary Modes of NBS-Encoding Genes in Three Solanaceae Crop Species. *G3 Genes|Genomes|Genetics* 7:1577–1585. <https://doi.org/10.1534/g3.117.040485>
- R Development Core Team (2010) a language and environment for statistical computing: reference index. R Foundation for Statistical Computing, Vienna
- Rami J (2014) Spidermap v1.6.12b
- Rao X, Krom N, Tang Y, et al (2014) A deep transcriptomic analysis of pod development in the vanilla orchid (*Vanilla planifolia*). *BMC Genomics* 15:964. <https://doi.org/10.1186/1471-2164-15-964>
- Rolfe R (1896) A revision of the genus *Vanilla*. *Journal of the Linnean Society of London, Botany* 32:439–478
- Romeis T (2001) Protein kinases in the plant defence response. *Current Opinion in Plant Biology* 4:407–414. [https://doi.org/10.1016/S1369-5266\(00\)00193-X](https://doi.org/10.1016/S1369-5266(00)00193-X)
- Romeis T, Piedras P, Jones JDG (2000) Resistance Gene-Dependent Activation of a Calcium-Dependent Protein Kinase in the Plant Defense Response. *Plant Cell* 12:803–815. <https://doi.org/10.1105/tpc.12.5.803>
- Roux-Cuvelier M, Grisoni M, Bellec A, et al (2021) Conservation of horticultural genetic resources in France. *Chronica Horticulturae* 61:21–36
- Sekhwai M, Li P, Lam I, et al (2015) Disease Resistance Gene Analogs (RGAs) in Plants. *IJMS* 16:19248–19290. <https://doi.org/10.3390/ijms160819248>
- Sela-Buurlage M, Budai-Hadrian O, Pan Q, et al (2001) Genome-wide dissection of

- Fusarium resistance in tomato reveals multiple complex loci. *Mol Gen Genomics* 265:1104–1111. <https://doi.org/10.1007/s004380100509>
- Shao Z-Q, Xue J-Y, Wu P, et al (2016) Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant Physiology* 170:2095–2109. <https://doi.org/10.1104/pp.15.01487>
- Shao Z-Q, Zhang Y-M, Hang Y-Y, et al (2014) Long-Term Evolution of Nucleotide-Binding Site-Leucine-Rich Repeat Genes: Understanding Gained from and beyond the Legume Family. *Plant Physiology* 166:217–234. <https://doi.org/10.1104/pp.114.243626>
- Simons G, Groenendijk J, Wijbrandi J, et al (1998) Dissection of the Fusarium I2 Gene Cluster in Tomato Reveals Six Homologs and One Active Gene Copy. *The Plant Cell* 10:1055–1068. <https://doi.org/10.1105/tpc.10.6.1055>
- Slate J, Gratten J, Beraldi D, et al (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* 136:97–107. <https://doi.org/10.1007/s10709-008-9317-z>
- Soto Arenas MÁ (2003) Vanilla. In: Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN (eds) *Genera orchidacearum: Orchidoideae*. Oxford University Press, USA, p 402
- Soto Arenas MA, Cribb P (2013) A new infrageneric classification and synopsis of the genus *Vanilla* Plum. ex mill. (Orchidaceae: Vanillinae). *Lankesteriana*. <https://doi.org/10.15517/lank.v0i0.12071>
- Steuernagel B, Witek K, Krattinger SG, et al (2020) The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol* 183:468–482. <https://doi.org/10.1104/pp.19.01273>
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815. <https://doi.org/10.1038/35048692>
- Theis T, Jimenez FA (1957) A Vanilla hybrid resistant to Fusarium root rot. *Phytopathology* 47:578–581
- Tibbs Cortes L, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. *Plant Genome* 14:. <https://doi.org/10.1002/tpg2.20077>
- Tucker CM (1927) Vanilla Root Rot. *Journal of Agricultural Research* 35:1121–1136
- Vakalounakis DJ (1996) Allelism of the Fcu-1 and Foc genes conferring resistance to fusarium wilt in cucumber. *Eur J Plant Pathol* 102:855–858. <https://doi.org/10.1007/BF01877055>
- Van Ooijen JW (2018) JoinMap®5, Software for the calculation of genetic linkage maps

- in experimental populations of diploid species. Kyazma B.V., Wageningen, Netherlands
- Van Ooijen JW (2009) MapQTL®6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma B.V., Wageningen, Netherlands
- Xiong W, Zhao Q, Zhao J, et al (2015) Different Continuous Cropping Spans Significantly Affect Microbial Community Membership and Structure in a Vanilla-Grown Soil as Revealed by Deep Pyrosequencing. *Microb Ecol* 70:209–218. <https://doi.org/10.1007/s00248-014-0516-0>
- Xue J-Y, Zhao T, Liu Y, et al (2020) Genome- Wide Analysis of the Nucleotide Binding Site Leucine-Rich Repeat Genes of Four Orchids Revealed Extremely Low Numbers of Disease Resistance Genes. *Front Genet* 10:1286. <https://doi.org/10.3389/fgene.2019.01286>
- Ye C-Y, Fan L (2021) Orphan Crops and their Wild Relatives in the Genomic Era. *Molecular Plant* 14:27–39. <https://doi.org/10.1016/j.molp.2020.12.013>
- Zhang J, Abdelraheem A, Ma J, et al (2022) Mapping of dynamic QTLs for resistance to Fusarium wilt (*Fusarium oxysporum* f. sp. *vasinfectum*) race 4 in a backcross inbred line population of Upland cotton. *Mol Genet Genomics* 297:319–332. <https://doi.org/10.1007/s00438-021-01846-2>
- Zhang L, Cenci A, Rouard M, et al (2019) Transcriptomic analysis of resistant and susceptible banana corms in response to infection by *Fusarium oxysporum* f. sp. *cubense* tropical race 4. *Sci Rep* 9:8199. <https://doi.org/10.1038/s41598-019-44637-x>
- Zhang Y-M, Shao Z-Q, Wang Q, et al (2016) Uncovering the dynamic evolution of nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes in Brassicaceae: Evolution of NBS-LRR genes in Brassicaceae. *J Integr Plant Biol* 58:165–177. <https://doi.org/10.1111/jipb.12365>

# DONNÉES SUPPLÉMENTAIRES PUBLIÉES

---

## 1. Chapitre 3

**Supporting Information Table S1.** List of 88 *Vanilla planifolia* accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

**Supporting Information Table S2.** List of 7 *Vanilla* × *tahitensis* accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

**Supporting Information Table S3.** List of 23 *Vanilla* wild relatives accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

**Supporting Information Table S4.** List of 19 interspecific hybrids accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

**Supporting information Fig. S1 GBS loci distribution and density on *V. planifolia* cv Daphna chromosomes** from 194,625 unfiltered SNP dataset. Horizontal axis displays the chromosome length. The density scale indicates the number of SNPs within 500 Kb window size. The plot shows the distribution of GBS SNPs across the 14 chromosomes.

**Supporting Information Fig. S2 Principal coordinates analysis (PCoA) based on 2,004 SNPs.** Accessions are color-coded by species.

**Supporting Information Fig. S3 Principal coordinates analysis (PCoA) based on 1,129 *Vanilla planifolia* SNPs.** Accessions are color-coded by **a.** type of reproduction, **b.** growing area of cultivars, **c.** ploidy level.

**Supporting Information Fig. S4 Genetic structure of 84 *V. planifolia* using 1,129**

**informative SNPs. a.** Genetic clustering as determined by STRUCTURE for  $K = 2$  et  $K = 4$ . The X-axis provides accessions and the y-axis provides the probability of each accession belonging to a cluster. Accessions are assigned to one cluster if their probability of belonging to this cluster is higher than 60%. The remaining accessions are classified as admixed. The dotted ground indicates accessions with admixture. **b.** DeltaK graph shows the best K as determined by STRUCTURE HARVESTER. **c.** Mean likelihood  $L(K)$  and standard deviation (SD) per K value from STRUCTURE HARVESTER.

**Supporting Information Fig. S5 Morphology of a.** *V. odorata* cf. CR3612 flower and fruits, **b.** *V. insignis* cf. CR0087 flower and leaves.

**Supporting Information Table S1.** List of 88 *Vanilla planifolia* accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

Identifier	Material type	Reproduction mode	Parents (female x male)	Variety name	Place of origin	Ploidy level <sup>(1)</sup>	AFLP group <sup>(2)</sup>	Missing data (%) <sup>(1)</sup>	Heterozygous SNPs (%) <sup>(1)</sup>	Structure cluster <sup>(3)</sup>
CR0001	Traditional cultivar	Vegetative propagation	-	bleue	Reunion Island, France	2C	AFLP1	15.68	71.85	cluster 4
CR0010	Traditional cultivar	Vegetative propagation	-	-	Costa Rica	2C	AFLP1	17.01	82.28	cluster 4
CR0019	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	nd	AFLP2	11.43	28.00	cluster 4
CR0020	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	2C	AFLP2	28.43	60.89	cluster 3
CR0021	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	2C	AFLP2	11.69	53.56	cluster 2
CR0022	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	2C	AFLP2	8.50	30.88	cluster 3
CR0023	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	2C	AFLP2	10.27	42.74	cluster 2
CR0024	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	2C	AFLP2	7.62	42.95	cluster 3
CR0025	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	2C	AFLP2	10.63	24.28	cluster 4
CR0026	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	nd	AFLP2	nd	nd	nd
CR0027	Plant breeding material	Selfed-progeny	Unknown	-	BRC Vatel, CIRAD, France	nd	AFLP2	19.93	15.82	admixed
CR0034	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	12.05	78.95	cluster 3
CR0035	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	18.42	61.24	admixed
CR0036	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	2C	AFLP1	9.74	80.37	admixed
CR0038	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	2C	AFLP1	11.60	86.97	cluster 3
CR0040	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	11.25	77.94	admixed
CR0041	Traditional cultivar	Vegetative propagation	-	-	Madagascar	nd	AFLP1	14.53	82.07	admixed
CR0151	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	nd	nd	nd
CR0196	Improved cultivar	Unknown	-	-	Reunion Island, France	nd	AFLP1	28.79	88.43	admixed
CR0510	Improved cultivar	Unknown	-	-	Reunion Island, France	nd	nd	12.13	25.10	cluster 4
CR0622	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	21.43	62.46	cluster 3
CR0623	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	12.75	74.82	admixed
CR0626	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	13.37	68.51	cluster 4
CR0628	Traditional cultivar	Unknown	-	Aiguille	Reunion Island, France	2C	AFLP2	6.02	46.75	cluster 3
CR0630	Traditional cultivar	Vegetative propagation	-	Sterile	Reunion Island, France	3C	AFLP1	18.25	54.06	cluster 4
CR0631	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	2C	AFLP1	22.41	57.42	cluster 4
CR0632	Traditional cultivar	Vegetative propagation	-	Petite Mexique	Reunion Island, France	2C	AFLP2	8.50	46.56	cluster 3
CR0640	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	24.62	42.19	cluster 4



DONNÉES SUPPLÉMENTAIRES PUBLIÉES |

CR0641	Traditional cultivar	Vegetative propagation	-	Grosse vanille	Reunion Island, France	4C	AFLP1	17.63	71.40	cluster 4
CR0642	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	2C	AFLP1	9.30	79.49	cluster 4
CR0645	Traditional cultivar	Vegetative propagation	-	Sterile	Reunion Island, France	3C	AFLP1	12.05	52.87	cluster 4
CR0654	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	8.86	79.98	cluster 3
CR0656	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	2C	AFLP1	33.13	39.47	cluster 4
CR0659	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	10.01	69.69	cluster 4
CR0661	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	26.04	44.19	cluster 4
CR0673	Traditional cultivar	Vegetative propagation	-	-	Samoa	2C	AFLP1	13.82	78.42	admixed
CR0674	Traditional cultivar	Vegetative propagation	-	-	Fiji	nd	AFLP1	10.54	76.24	admixed
CR0676	Traditional cultivar	Vegetative propagation	-	-	Fiji	nd	AFLP1	25.95	54.67	admixed
CR0684	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	11.96	64.49	cluster 3
CR0695	Traditional cultivar	Vegetative propagation	-	-	Guadeloupe, France	nd	AFLP1	17.18	62.57	cluster 3
CR0717	Traditional cultivar	Vegetative propagation	-	-	Reunion Island, France	nd	AFLP1	36.49	40.45	cluster 4
CR0763	Traditional cultivar	Vegetative propagation	-	-	India	nd	AFLP1	14.61	70.12	cluster 4
CR0802	Traditional cultivar	Vegetative propagation	-	Grosse vanille	Reunion Island, France	4C	nd	13.46	62.23	admixed
CR0844	Traditional cultivar	Vegetative propagation	-	Grosse vanille	Mayotte Island, France	nd	AFLP1	nd	nd	nd
CR0885	Traditional cultivar	Vegetative propagation	-	-	Comoros	nd	nd	15.50	57.55	cluster 4
CR0925	Plant breeding material	Selfed-progeny	CR0041 x CR0041	-	BRC Vatel, CIRAD, France	nd	nd	8.68	29.87	cluster 3
CR0974	Plant breeding material	Selfed-progeny	CR0066 x CR0066	-	BRC Vatel, CIRAD, France	nd	nd	9.57	35.85	cluster 3
CR1004	Plant breeding material	Selfed-progeny	CR0051 x CR0051	-	BRC Vatel, CIRAD, France	nd	nd	18.51	28.70	cluster 4
CR1048	Traditional cultivar	Vegetative propagation	-	-	Mauritius	nd	nd	11.51	71.07	cluster 3
CR1068	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	6.20	23.32	cluster 3
CR1076	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	6.73	31.62	cluster 4
CR1097	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	5.93	25.99	cluster 2
CR1098	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	13.46	28.66	cluster 4
CR1123	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	14.53	35.03	cluster 4
CR1126	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	4.78	19.81	cluster 3
CR1129	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	8.24	29.34	cluster 3
CR1140	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	7.62	42.47	cluster 3
CR1157	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	21.26	39.60	admixed
CR1161	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	9.21	47.02	cluster 3
CR1426	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	7.00	34.48	cluster 3

DONNÉES SUPPLÉMENTAIRES PUBLIÉES |

CR1433	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	3.37	22.91	admixed
CR1434	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	10.19	17.26	cluster 4
CR1701	Plant breeding material	Selfed-progeny	CR0041 x CR0041	-	BRC Vatel, CIRAD, France	nd	nd	9.30	37.11	cluster 3
CR1712	Plant breeding material	Selfed-progeny	CR0136 x CR0136	-	BRC Vatel, CIRAD, France	nd	nd	12.13	37.40	cluster 4
CR1900	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	9.30	38.18	cluster 3
CR1901	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	8.33	37.29	cluster 3
CR1917	Plant breeding material	Selfed-progeny	CR0041 x CR0041	-	BRC Vatel, CIRAD, France	nd	nd	11.34	53.55	cluster 1
CR1972	Plant breeding material	Intra-specific hybrid	CR0802 x CR0631	-	BRC Vatel, CIRAD, France	3C	nd	16.30	63.92	cluster 4
CR2030	Plant breeding material	Selfed-progeny	CR0802 x CR0802	-	BRC Vatel, CIRAD, France	4C	nd	32.60	64.91	cluster 4
CR2093	Improved cultivar	Unknown	-	-	Reunion Island, France	nd	nd	12.31	73.13	cluster 4
CR2100	Improved cultivar	Unknown	-	-	Reunion Island, France	nd	nd	15.94	61.85	admixed
CR2102	Improved cultivar	Unknown	-	-	Reunion Island, France	nd	nd	12.93	78.43	cluster 4
CR2526	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	9.39	39.98	cluster 4
CR2527	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	9.74	36.51	cluster 4
CR2528	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	11.60	33.47	cluster 1
CR2529	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	16.12	32.63	cluster 3
CR2560	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	27.10	36.57	cluster 1
CR2561	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	23.65	59.98	cluster 1
CR2562	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	16.83	42.92	cluster 3
CR2564	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	nd	nd	nd
CR2567	Plant breeding material	Selfed-progeny	CR0040 x CR0040	-	BRC Vatel, CIRAD, France	nd	nd	26.66	42.87	admixed
CR2687	Traditional cultivar	Unknown	-	Colibri	Mexico	nd	nd	15.59	70.30	cluster 4
CR2699	Traditional cultivar	Vegetative propagation	-	Maryse M1	Reunion Island, France	nd	nd	30.29	51.46	admixed
CR2844	Traditional cultivar	Vegetative propagation	-	Maryse M2	Reunion Island, France	nd	nd	22.14	63.59	cluster 4
CR3601	Traditional cultivar	Vegetative propagation	-	-	Mauritius	nd	nd	39.77	39.71	cluster 4
CR3602	Traditional cultivar	Vegetative propagation	-	-	Mauritius	nd	nd	22.41	58.90	cluster 3
CR3603	Traditional cultivar	Vegetative propagation	-	-	Mauritius	nd	nd	15.68	65.65	cluster 4
CR3604	Traditional cultivar	Vegetative propagation	-	-	Mauritius	nd	nd	18.16	60.28	cluster 4

(1) From Bory (2008); *nd*: no data.

(2) From Bory et al. (2008); *nd*: no data.

(3) This study, from dataset with 1,129 SNPs and 84 samples. *Missing data*: percentage of absent SNP data per accession; *Heterozygous SNPs*: percentage of heterozygous SNPs per accession; *nd*: no data.

**Supporting Information Table S2.** List of 7 *Vanilla × tahitensis* accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

Identifier	Material type	Reproduction mode	Parents (female x male)	Variety name	Place of origin	Missing data (%) <sup>(1)</sup>	Heterozygous SNPs (%) <sup>(1)</sup>
CR0017	Traditional cultivar	Vegetative propagation	-	Haapape	French Polynesia	29.69	32.79
CR1415	Plant breeding material	Selfed-progeny	CR0017 x CR0017	-	BRC Vatel, CIRAD, France	23.45	16.69
CR3605	Traditional cultivar	Vegetative propagation	-	Tahiti H5-27	French Polynesia	33.88	31.47
CR3606	Traditional cultivar	Vegetative propagation	-	Haapape OFE2	French Polynesia	28.74	28.78
CR3607	Traditional cultivar	Vegetative propagation	-	T4-49	French Polynesia	32.19	31.57
CR3608	Traditional cultivar	Vegetative propagation	-	H5-66	French Polynesia	30.69	30.02
CR3609	Traditional cultivar	Vegetative propagation	-	S70	French Polynesia	38.17	27.68

<sup>(1)</sup> This study, from dataset with 2,004 SNPs and 133 samples. *Missing data*: percentage of absent SNP data per accession; *Heterozygous SNPs*: percentage of heterozygous SNPs per accession.

**Supporting Information Table S3.** List of 23 *Vanilla* wild relatives accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

Identifier	Species	Place of origin	Missing data (%) <sup>(1)</sup>	Heterozygous SNPs (%) <sup>(1)</sup>
CR0072	<i>V. bahiana</i>	Brazil	28.59	0.28
CR2074	<i>V. bahiana</i>	BRC Vatel, CIRAD, France	25.15	0.40
CR0109	<i>V. cribbiana</i>	French Guyana	34.93	0.38
CR2897	<i>V. cribbiana</i>	Ecuador	40.47	0.17
CR2899	<i>V. cribbiana</i>	Ecuador	39.62	0.25
CR2901	<i>V. cribbiana</i>	Private collection, France	29.99	0.00
CR3614	<i>V. helleri</i>	Costa Rica	25.60	0.27
CR2688	<i>V. insignis</i> cf.	Mexico	26.00	0.00
CR0087	<i>V. insignis</i> cf.	MNHN, France	34.08	5.30
CR2685	<i>V. odorata</i>	Mexico	34.28	0.08
CR2902	<i>V. odorata</i>	Private collection, France	38.62	0.00
CR2908	<i>V. odorata</i>	Private collection, France	40.92	0.76
CR3611	<i>V. odorata</i>	Brazil	26.65	0.07
CR3612	<i>V. odorata</i>	Brazil	26.75	0.07
CR0116	<i>V. odorata</i>	French Guyana	33.48	0.83
CR2686	<i>V. odorata</i>	Mexico	26.00	0.20
CR2894	<i>V. odorata</i>	Ecuador	23.85	0.13
CR0018	<i>V. pompona</i>	French Polynesia	38.52	0.24
CR0033	<i>V. pompona</i>	Reunion Island, France	40.92	0.51
CR0068	<i>V. sotoarenasii</i>	Costa Rica	14.02	5.11
CR3615	<i>V. sotoarenasii</i>	Costa Rica	13.37	4.49
CR3616	<i>V. sotoarenasii</i>	Costa Rica	23.90	4.00
CR2903	<i>V. sotoarenasii</i> cf.	Private collection, France	14.17	0.12

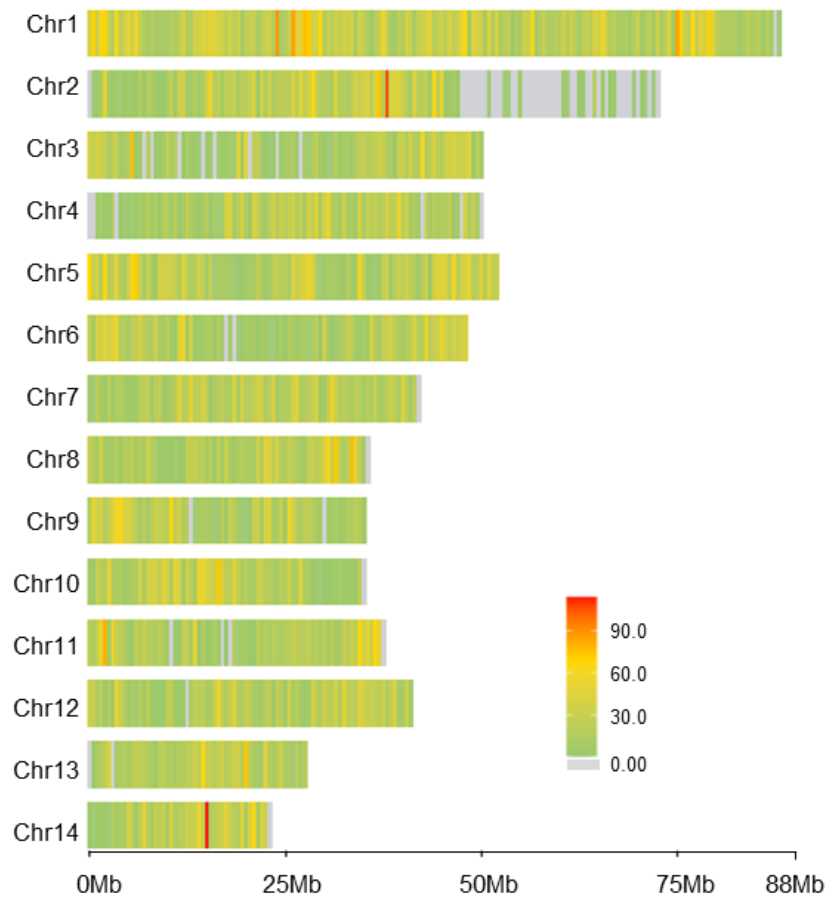
<sup>(1)</sup> This study, from dataset with 2,004 SNPs and 133 samples. *Missing data*: percentage of absent SNP data per accession; *Heterozygous SNPs*: percentage of heterozygous SNPs per accession.

**Supporting Information Table S4.** List of 19 interspecific hybrids accessions used in this study, sampled from the *ex situ* collection of BRC Vatel (Cirad, Reunion Island, France).

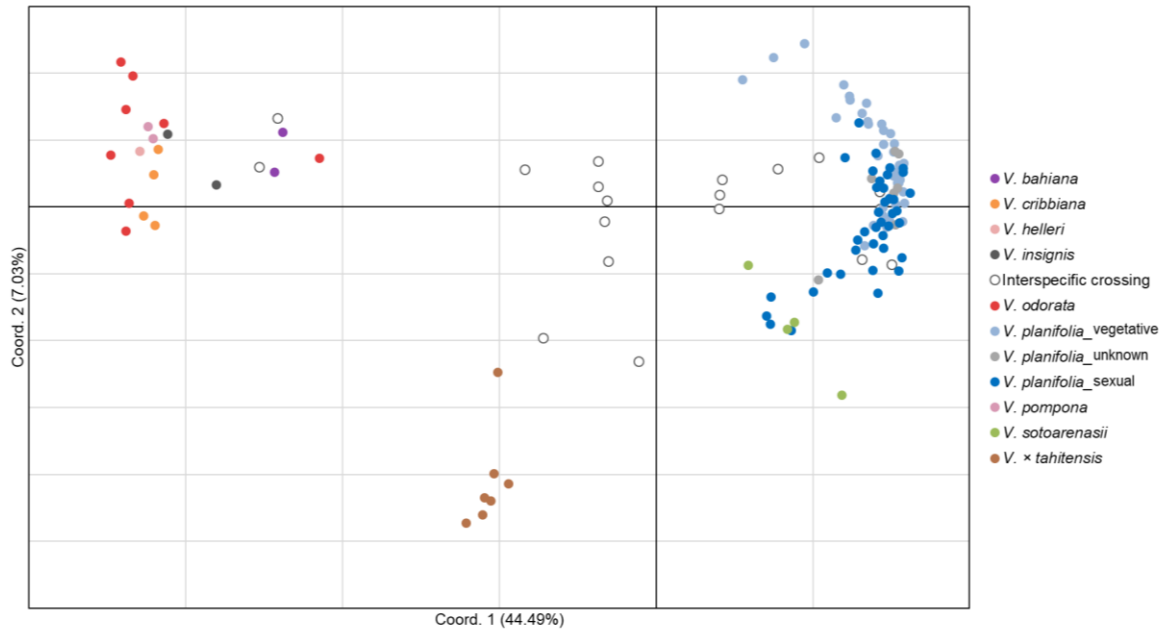
Identifier	Parents (female x male)	Place of origin	Missing data (%) <sup>(1)</sup>	Heterozygous SNPs (%) <sup>(1)</sup>
CR0003	<i>V. planifolia</i> x <i>V. × tahitensis</i> *	FOFIFA, Madagascar	32.88	60.07
CR0139	<i>V. pompona</i> x <i>V. planifolia</i> *	FOFIFA, Madagascar	41.92	39.00
CR0166	<i>V. planifolia</i> x <i>V. phaeantha</i>	BRC Vatel, CIRAD, France	28.69	23.16
CR0757	[( <i>V. planifolia</i> x <i>V. pompona</i> ) x <i>V. planifolia</i> ] cf x [( <i>V. planifolia</i> x <i>V. pompona</i> ) x <i>V. planifolia</i> ]*	Madagascar	8.63	16.11
CR1156	<i>V. sotoarenasii</i> x <i>V. pompona</i>	BRC Vatel, CIRAD, France	27.99	18.99
CR1531	( <i>V. planifolia</i> x <i>V. pompona</i> ) x ( <i>V. planifolia</i> x <i>V. × tahitensis</i> )*	FOFIFA, Madagascar	7.88	24.65
CR1540	<i>V. planifolia</i> x ( <i>V. planifolia</i> x <i>V. pompona</i> )*	FOFIFA, Madagascar	8.23	16.37
CR1725	<i>V. planifolia</i> x <i>V. pompona</i>	BRC Vatel, CIRAD, France	32.49	19.22
CR1773	<i>V. planifolia</i> x ( <i>V. planifolia</i> x <i>V. phaeantha</i> )	BRC Vatel, CIRAD, France	21.16	25.70
CR1810	<i>V. planifolia</i> x <i>V. bahiana</i>	BRC Vatel, CIRAD, France	26.35	0.34
CR1936	( <i>V. planifolia</i> x <i>V. phaeantha</i> ) x <i>V. planifolia</i>	BRC Vatel, CIRAD, France	15.92	38.46
CR1943	<i>V. planifolia</i> x ( <i>V. planifolia</i> x <i>V. phaeantha</i> )	BRC Vatel, CIRAD, France	20.51	27.87
CR1947	<i>V. insignis</i> x <i>V. bahiana</i>	BRC Vatel, CIRAD, France	31.99	2.86
CR2354	( <i>V. planifolia</i> x <i>V. phaeantha</i> ) x ( <i>V. planifolia</i> x <i>V. phaeantha</i> )	BRC Vatel, CIRAD, France	21.61	13.49
CR2711	<i>V. chamissonis</i> x <i>V. planifolia</i>	BRC Vatel, CIRAD, France	26.70	23.76
CR2715	<i>V. planifolia</i> x <i>V. chamissonis</i>	BRC Vatel, CIRAD, France	27.50	24.02
CR2717	( <i>V. planifolia</i> x <i>V. pompona</i> ) x <i>V. planifolia</i>	Costa Rica	28.04	44.31
CR2718	( <i>V. planifolia</i> x <i>V. pompona</i> ) x <i>V. planifolia</i>	Costa Rica	31.39	43.05
CR2731	<i>V. planifolia</i> X <i>V. madagasacariensis</i> cf.	BRC Vatel, CIRAD, France	13.97	28.65

<sup>(1)</sup> This study, from dataset with 2,004 SNPs and 133 samples. *Missing data*: percentage of absent SNP data per accession; *Heterozygous SNPs*: percentage of heterozygous SNPs per accession.

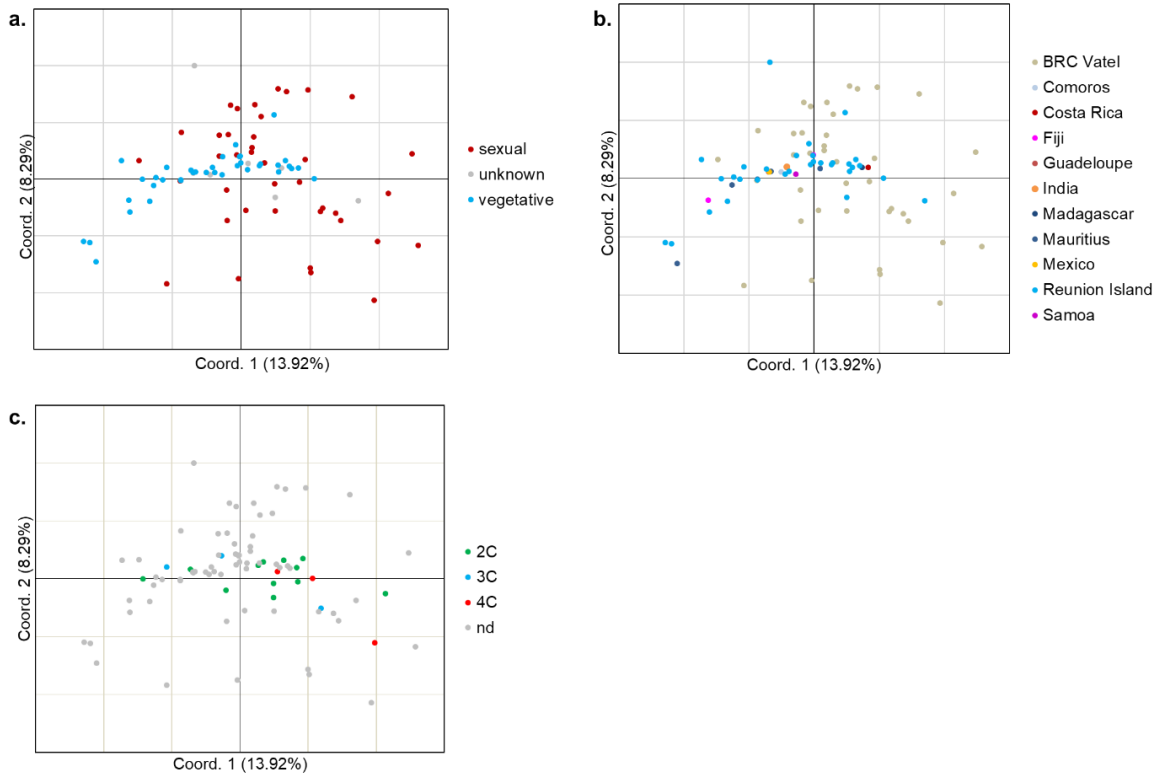
\* uncertainty about the parental cross used



**Supporting information Fig. S1 GBS loci distribution and density on *V. planifolia* cv *Daphna* chromosomes** from 194,625 unfiltered SNP dataset. Horizontal axis displays the chromosome length. The density scale indicates the number of SNPs within 500 Kb window size. The plot shows the distribution of GBS SNPs across the 14 chromosomes.

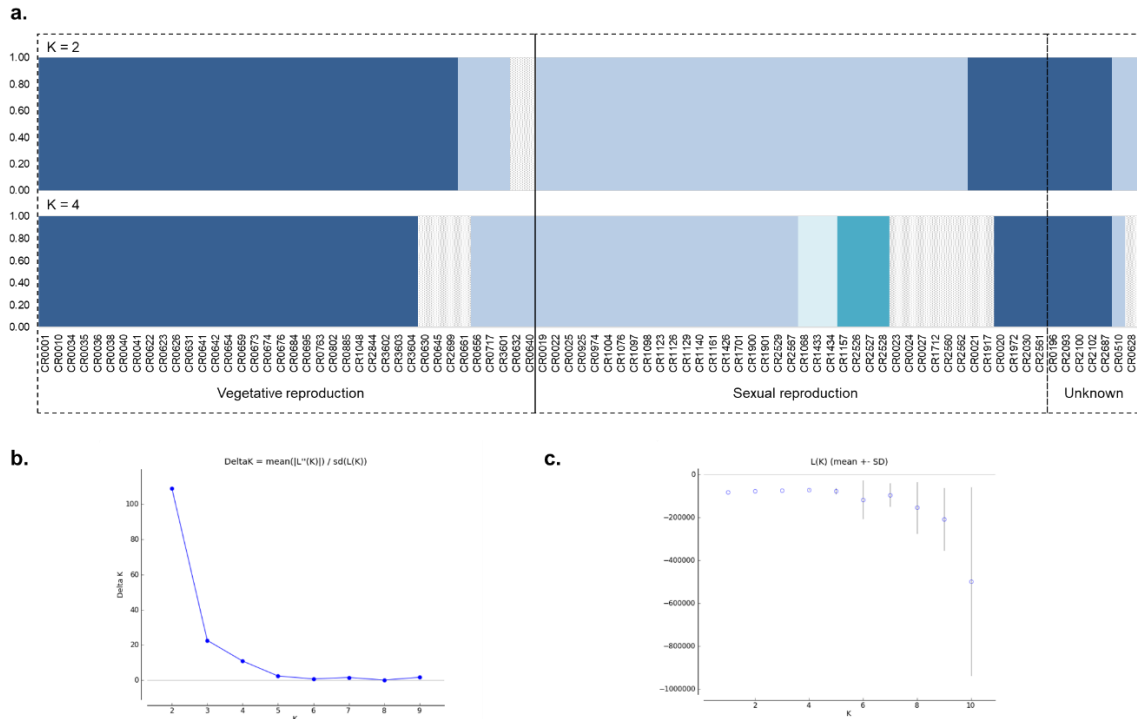


**Supporting Information Fig. S2 Principal coordinates analysis (PCoA) based on 2,004 SNPs.** Accessions are color-coded by species.



**Supporting Information Fig. S3 Principal coordinates analysis (PCoA) based on 1,129 *Vanilla planifolia* SNPs.** Accessions are color-coded by **a.** type of reproduction, **b.** growing area of cultivars, **c.** ploidy level.





**Supporting Information Fig. S4 Genetic structure of 84 *V. planifolia* using 1,129 informative SNPs. a.** Genetic clustering as determined by STRUCTURE for K = 2 et K = 4. The X-axis provides accessions and the y-axis provides the probability of each accession belonging to a cluster. Accessions are assigned to one cluster if their probability of belonging to this cluster is higher than 60%. The remaining accessions are classified as admixed. The dotted ground indicates accessions with admixture. **b.** DeltaK graph shows the best K as determined by STRUCTURE HARVESTER. **c.** Mean likelihood L(K) and standard deviation (SD) per K value from STRUCTURE HARVESTER.



**Supporting Information Fig. S5 Morphology of a. *V. odorata* cf. CR3612 flower and fruits, b. *V. insignis* cf. CR0087 flower and leaves.**

## 2. Chapitre 4

**Supplementary Table S1.** DNA samples used for genotyping by sequencing (GBS). Their associated barcode sequences and the number of reads after cleaning, demultiplexing and mapping steps are indicated.

**Supplementary Table S2.** Accessions tested for the resistance to *Fusarium oxysporum* f.sp. *radicis-vanillae* Fo072. The mean and sd values for the five traits recorded (AUDPC, d.myc, d.5, d.10 and d.15) are indicated.

**Supplementary Table S3.** Candidate genes for significant quantitative trait loci for Fo072 resistance in CR0040 genome annotation.

**Supplementary Figure S1.** Genetic linkage map of the population, with 1,804 markers across the linkage groups **(a)** VpLG01 to VpLG08 and **(b)** VpLG09 to VpLG16. Genetic distances (cM) are indicated on the left and markers names on the right of the map bars. The colors indicate *de novo* (red) and reference-based (black) markers.

**Supplementary Figure S2.** Phenotyping of CR0040 progenies. The x-axis indicates the name of the accessions and the y-axis **(a)** the AUDPC values, **(b)** the number of days with mycelium growth (d.myc) and the disease score at **(c)** 5 days post-inoculation (dpi, d.5), **(d)** 10 dpi (d.10) and **(e)** 15 dpi (d.15). Each colored point indicates a replicate value. Mean values and standard error of the mean are represented respectively by the black points and error bars.

**Supplementary Figure S3.** Quantitative trait loci for AUDPC (black), d.5 (pink), d.10 (purple), d.15 (green) and d.myc (blue) on the linkage groups VpLG01 to VpLG16 **(a-p)**. The dotted lines indicate the LOD threshold of 4.7 for d.5 (pink), 4.8 for d.15 (green) and 4.5 for AUDPC, d.10 and d.myc (black). LOD, logarithm of odds.

**Supplementary Table S1.** DNA samples used for genotyping by sequencing (GBS). Their associated barcode sequences and the number of reads after cleaning, demultiplexing and mapping steps are indicated.

Accession	Population	Barcodes used for each replicate	Reads in pairs in sequencing	Forward reads	Reverse reads	Reads mapped properly in pairs
CR0040	parent	GAAGA ; AGAATCCA ; AGTCG	7,282,844	3,641,422	3,641,422	6,866,344
CR1059	progeny	CAGAT ; GCCAA	5,999,392	2,999,696	2,999,696	5,664,408
CR1062	progeny	AATGCA	4,037,932	2,018,966	2,018,966	3,807,534
CR1063	progeny	GATACG ; AATAGACA	12,224,658	6,112,329	6,112,329	11,567,214
CR1068	progeny	GGATCG	1,569,690	784,845	784,845	1,490,994
CR1070	progeny	AGAATCA ; ACGCTCA	4,581,274	2,290,637	2,290,637	4,337,440
CR1074	progeny	GACATCA	5,693,680	2,846,840	2,846,840	5,385,560
CR1076	progeny	CGCCTCA	7,924,606	3,962,303	3,962,303	7,498,540
CR1077	progeny	AACCTGCA	9,958,236	4,979,118	4,979,118	9,386,124
CR1096	progeny	AACACATG ; CCGTCA	4,465,926	2,232,963	2,232,963	4,229,214
CR1097	progeny	CCAG ; CGATT	4,911,246	2,455,623	2,455,623	4,660,600
CR1098	progeny	CCACG	2,357,960	1,178,980	1,178,980	2,236,984
CR1103	progeny	ACGAG ; GGTTACCA	6,289,856	3,144,928	3,144,928	5,957,430
CR1107	progeny	GAGCG	1,906,266	953,133	953,133	1,806,130
CR1123	progeny	ATAGG	1,454,596	727,298	727,298	1,370,374
CR1125	progeny	TGAACA ; AGGTCCA	4,340,152	2,170,076	2,170,076	4,086,666
CR1126	progeny	CTCCGA	4,543,906	2,271,953	2,271,953	4,291,752
CR1127	progeny	CTGAAG	905,524	452,762	452,762	862,430
CR1129	progeny	CACTGCA	1,011,652	505,826	505,826	957,318
CR1411	progeny	ACATGCA	2,585,936	1,292,968	1,292,968	2,424,610
CR1421	progeny	AATTCGA ; CCACCTAG	4,875,818	2,437,909	2,437,909	4,601,150
CR1424	progeny	CCATGCCA	1,712,396	856,198	856,198	1,617,464
CR1426	progeny	CCACACGT	3,203,472	1,601,736	1,601,736	3,013,470
CR1433	progeny	TTGA	1,965,030	982,515	982,515	1,872,582
CR1434	progeny	TATAA ; TGGCACA	3,267,194	1,633,597	1,633,597	3,084,356
CR1899	progeny	TTCTA	3,191,094	1,595,547	1,595,547	3,010,376
CR1900	progeny	TCGCT ; AATGT	3,659,912	1,829,956	1,829,956	3,456,154
CR1901	progeny	GACTCA	4,932,962	2,466,481	2,466,481	4,669,394
CR1903	progeny	TCAGCG ; AACAGCTA	3,468,482	1,734,241	1,734,241	3,289,248
CR1904	progeny	TATCGA	1,617,062	808,531	808,531	1,522,674
CR2524	progeny	TCTGCCA ; CCGTACA	5,534,694	2,767,347	2,767,347	5,227,734
CR2526	progeny	CGTGACA	3,576,470	1,788,235	1,788,235	3,389,036
CR2528	progeny	TCAAGACA	521,380	260,690	260,690	493,996
CR2529	progeny	TTGGCGCA	1,995,306	997,653	997,653	1,874,408
CR2560	progeny	TTGTCTAA ; ACCTCG	2,231,306	1,115,653	1,115,653	2,104,944
CR2563	progeny	-	0	0	0	0
CR2566	progeny	GAGCG	2,580,842	1,290,421	1,290,421	2,447,290
CR2568	progeny	AGTCG ; GGTGAGCT	5,158,752	2,579,376	2,579,376	4,861,350

CR2569	progeny	CGATT	6,202,088	3,101,044	3,101,044	5,869,040
CR2577	progeny	ATGACA ; GAATACA	6,298,146	3,149,073	3,149,073	5,965,296
CR2820	progeny	AGCTAA	1,842,652	921,326	921,326	1,743,166
CR2821	progeny	GCCGTA ; TCGCAA	5,757,116	2,878,558	2,878,558	5,461,858
CR2822	progeny	TGGCACA	679,576	339,788	339,788	649,366
CR2825	progeny	ATGCCGA	2,361,974	1,180,987	1,180,987	2,222,198
CR2826	progeny	GTGGACCA ; TGCG	3,549,452	1,774,726	1,774,726	3,356,826
CR2828	progeny	GGTTACCA	4,513,334	2,256,667	2,256,667	4,277,864
CR2829	progeny	GGTGAGCT	2,782,070	1,391,035	1,391,035	2,622,252
CR2830	progeny	ATTG	2,224,340	1,112,170	1,112,170	2,096,982
CR2832	progeny	ACATA ; ATGCCGA	4,045,006	2,022,503	2,022,503	3,833,790
CR2833	progeny	GCCAA	3,393,516	1,696,758	1,696,758	3,218,482
CR2834	progeny	AATGT ; CTAGCA	4,198,958	2,099,479	2,099,479	3,971,042
CR2835	progeny	CTAGCA	3,323,976	1,661,988	1,661,988	3,131,636
CR2837	progeny	CGGT	514,652	257,326	257,326	486,456
CR2839	progeny	AGGTCCA	1,433,116	716,558	716,558	1,349,568
CR2840	progeny	GAATACA ; CACTCGCA	4,530,086	2,265,043	2,265,043	4,298,096
CR2842	progeny	GATACAA	5,366,964	2,683,482	2,683,482	5,064,324
CR2864	progeny	AATAGACA	9,275,804	4,637,902	4,637,902	8,775,880
CR2865	progeny	AACAGCTA ; ATACCG	3,886,974	1,943,487	1,943,487	3,677,744
CR2866	progeny	CGGT ; AGCTAA	3,161,616	1,580,808	1,580,808	2,994,232
CR2867	progeny	CTCAG	1,818,278	909,139	909,139	1,721,692
CR2868	progeny	TAATG ; ATTG	9,568,290	4,784,145	4,784,145	9,072,452
CR2869	progeny	-	0	0	0	0
CR2870	progeny	TCGCAA ; GATACAA	8,083,140	4,041,570	4,041,570	7,666,532
CR2875	progeny	GCTAGA	4,432,090	2,216,045	2,216,045	4,207,082
CR2876	progeny	CCTCGCA ; GCTAGA	10,294,156	5,147,078	5,147,078	9,752,482
CR2877	progeny	ACGCTCA	1,915,186	957,593	957,593	1,814,122
CR2878	progeny	CCGTACA	4,886,304	2,443,152	2,443,152	4,611,400
CR2879	progeny	CACTCGCA ; TCTGG	3,012,550	1,506,275	1,506,275	2,843,818
CR2881	progeny	CCGATACA	6,624,390	3,312,195	3,312,195	6,190,726
CR2884	progeny	CCACCTAG ; GTGACCA	6,831,376	3,415,688	3,415,688	6,447,564
CR2948	progeny	TGCG	647,300	323,650	323,650	610,768
CR2949	progeny	TCTGG ; GTGGACCA	6,768,684	3,384,342	3,384,342	6,406,810
CR2951	progeny	GTTC A	7,340,726	3,670,363	3,670,363	6,949,936
CR2952	progeny	CCGTCA ; CAGTGG	3,500,314	1,750,157	1,750,157	3,301,582
CR2953	progeny	ACCTCG	1,347,666	673,833	673,833	1,278,054
CR2955	progeny	ATACCG ; TAATG	5,130,836	2,565,418	2,565,418	4,847,182
CR2956	progeny	TTAGACA	1,812,482	906,241	906,241	1,716,478
CR2957	progeny	CTCGCCA ; CCGATACA	4,643,766	2,321,883	2,321,883	4,390,612
CR2959	progeny	-	0	0	0	0
CR2961	progeny	TCGGAACA	483,628	241,814	241,814	452,390
CR2962	progeny	TTACAGCA	5,344,120	2,672,060	2,672,060	5,038,072
CR2963	progeny	TTGTTACA ; TTAGACA	3,142,430	1,571,215	1,571,215	2,957,378
CR2964	progeny	GTAT ; GCCGTA	4,114,734	2,057,367	2,057,367	3,903,584
CR2965	progeny	CGGCA	1,046,650	523,325	523,325	985,992

CR2966	progeny	GGCGG ; ACATA	3,861,552	1,930,776	1,930,776	3,631,364
CR2967	progeny	GTTCCA	1,339,512	669,756	669,756	1,265,496
CR2968	progeny	CGTCAA ; AGAATCCA	5,500,060	2,750,030	2,750,030	5,171,234
CR2970	progeny	TACGAA	3,315,154	1,657,577	1,657,577	3,140,070
CR2971	progeny	GTGACCA ; CCTCGCA	7,970,626	3,985,313	3,985,313	7,552,810
CR2973	progeny	TGTAGCA	1,358,250	679,125	679,125	1,283,360
CR2975	progeny	GTCGGAA	3,302,390	1,651,195	1,651,195	3,133,366
CR2976	progeny	GGTCGACA ; GTTCA	5,155,440	2,577,720	2,577,720	4,903,944
CR2977	progeny	AACT ; CTCGCCA	8,431,368	4,215,684	4,215,684	7,921,496
CR2980	progeny	AACCA	2,924,144	1,462,072	1,462,072	2,733,966
CR2981	progeny	GAAGA ; TCGGAACA	11,302,714	5,651,357	5,651,357	10,640,910
CR2982	progeny	CAGAT	2,784,260	1,392,130	1,392,130	2,649,838
CR2983	progeny	AATGCA	6,154,458	3,077,229	3,077,229	5,809,824
CR2985	progeny	GATACG	8,361,194	4,180,597	4,180,597	7,924,168
CR2988	progeny	GGATCG ; GTAT	3,553,888	1,776,944	1,776,944	3,380,832
CR3771	progeny	AGAATCA	3,706,344	1,853,172	1,853,172	3,514,796
CR3772	progeny	GACATCA	4,190,304	2,095,152	2,095,152	3,977,424
CR3773	progeny	CGCCTCA ; TACGAA	7,994,740	3,997,370	3,997,370	7,553,698
CR3774	progeny	AACCTGCA	6,142,584	3,071,292	3,071,292	5,805,928
CR3775	progeny	AACACATG ; ATGACA ; GTCGGAA	6,369,250	3,184,625	3,184,625	6,024,412
CR3776	progeny	CCAG	6,377,168	3,188,584	3,188,584	6,010,770
CR3777	progeny	CCACG ; CTCAG	5,662,192	2,831,096	2,831,096	5,357,534
CR3778	progeny	ACGAG	5,714,140	2,857,070	2,857,070	5,429,256
CR3779	progeny	ATAGG	3,160,874	1,580,437	1,580,437	2,999,584
CR3780	progeny	TGAACA ; TTGTTACA	8,180,214	4,090,107	4,090,107	7,744,474
CR3781	progeny	CTCCGA	2,206,156	1,103,078	1,103,078	2,113,808
CR3783	progeny	CTGAAG	851,722	425,861	425,861	816,142
CR3784	progeny	CACTGCA ; GGCGG	4,749,714	2,374,857	2,374,857	4,487,632
CR3785	progeny	ACATGCA ; CGTCAA	7,781,384	3,890,692	3,890,692	7,391,220
CR3786	progeny	AATTCGA	2,775,932	1,387,966	1,387,966	2,618,736
CR3787	progeny	CCATGCCA ; TGTAGCA	4,682,658	2,341,329	2,341,329	4,440,738
CR3788	progeny	CCACACGT	5,167,646	2,583,823	2,583,823	4,888,192
CR3789	progeny	TTGA	6,982,592	3,491,296	3,491,296	6,616,590
CR3790	progeny	TATAA	1,099,276	549,638	549,638	1,044,006
CR3791	progeny	TTCTA ; TTACAGCA	11,241,350	5,620,675	5,620,675	10,662,486
CR3792	progeny	TCGCT	2,297,588	1,148,794	1,148,794	2,170,556
CR3793	progeny	GACTCA	4,228,192	2,114,096	2,114,096	4,000,812
CR3794	progeny	TCAGCG	1,589,052	794,526	794,526	1,505,634
CR3796	progeny	TATCGA ; CGGCA	3,650,326	1,825,163	1,825,163	3,459,656
CR3797	progeny	TCTGCCA ; GTTCCA	4,847,374	2,423,687	2,423,687	4,581,080
CR3798	progeny	CGTGACA	2,072,298	1,036,149	1,036,149	1,969,428
CR3799	progeny	TCAAGACA	2,855,322	1,427,661	1,427,661	2,706,642
CR3800	progeny	TTGGCGCA	7,831,192	3,915,596	3,915,596	7,419,562
CR3801	progeny	TTGTCTAA ; GGTCGACA	8,276,818	4,138,409	4,138,409	7,824,370
CR3802	progeny	GGTA	3,014,078	1,507,039	1,507,039	2,861,758

**Supplementary Table S2.** Accessions tested for the resistance to *Fusarium oxysporum* f.sp. *radicis-vanillae* Fo072. The mean and sd values for the five traits recorded (AUDPC, d.myc, d.5, d.10 and d.15) are indicated.

Accessio n	No. of replicates	AUDPC		d.myc		d.5		d.10		d.15	
		Mea n	SD	Mea n	SD	Mea n	SD	Mea n	SD	Mea n	SD
CR0001 <sup>a</sup>	21	20.50	7.25	6.00	3.15	0.29	0.47	2.26	0.87	3.10	0.70
CR0020 <sup>b</sup>	26	5.79	6.20	1.81	2.58	0.00	0.00	0.38	0.77	1.50	1.33
CR0040 <sup>c</sup>	29	21.40	6.82	6.52	2.89	0.26	0.44	2.17	0.96	3.17	0.47
CR1059	8	10.63	3.09	3.75	1.39	0.00	0.00	0.50	0.53	3.00	0.00
CR1062	8	23.44	3.23	7.13	0.83	0.50	0.53	2.25	0.46	3.63	0.52
CR1063	4	14.13	4.31	4.75	0.96	0.00	0.00	1.00	1.15	3.25	0.50
CR1068	8	8.88	4.10	1.75	2.43	0.00	0.00	0.38	0.52	2.50	0.53
CR1070	8	8.19	4.37	2.50	2.51	0.00	0.00	0.25	0.46	2.63	1.06
CR1074	8	22.94	3.05	8.88	1.13	0.75	0.46	2.50	0.58	3.00	0.00
CR1076	8	22.19	3.26	6.88	1.13	0.63	0.52	2.25	0.46	3.13	0.35
CR1077	8	22.81	2.94	6.88	1.81	0.75	0.46	2.38	0.52	3.13	0.35
CR1096	8	22.31	6.57	7.00	0.76	0.50	0.53	2.25	0.89	3.38	0.52
CR1097	8	18.75	6.38	5.75	4.83	0.50	0.53	-	-	2.63	0.52
CR1098	8	22.06	4.76	6.63	1.85	0.75	0.50	2.38	0.92	3.13	0.35
CR1103	8	20.50	7.66	5.75	3.49	0.00	0.00	2.13	1.25	3.00	0.53
CR1107	0	-	-	-	-	-	-	-	-	-	-
CR1123	8	5.63	6.39	2.13	2.95	0.00	0.00	0.50	0.53	1.25	1.49
CR1125	8	18.13	5.82	4.25	3.54	0.00	0.00	2.38	0.74	2.75	1.04
CR1126	4	19.63	6.28	4.75	3.30	-	-	2.25	0.96	2.75	0.50
CR1127	8	7.50	3.86	1.00	1.51	0.00	0.00	0.50	0.53	2.00	0.93
CR1129	8	16.56	3.49	4.38	2.20	0.00	0.00	1.88	0.64	2.88	0.35
CR1411	8	15.31	8.35	5.00	3.12	0.00	0.00	1.63	1.30	2.88	0.99
CR1421	8	11.13	6.49	3.25	3.81	0.13	0.35	0.00	0.00	2.25	0.89
CR1424	8	24.00	3.38	8.13	1.13	0.00	0.00	2.75	0.46	3.00	0.00
CR1426	8	23.06	2.64	7.00	0.76	0.50	0.58	2.75	0.46	3.13	0.35
CR1433	8	16.56	4.93	5.13	2.42	0.25	0.50	2.13	0.83	2.88	0.35
CR1434	8	19.44	4.04	5.75	2.60	0.13	0.35	2.00	0.82	2.88	0.35
CR1899	8	15.50	7.25	4.13	2.95	0.00	0.00	1.50	1.20	2.75	0.71
CR1900	8	19.44	6.96	5.50	3.07	0.50	0.53	1.75	1.04	3.13	0.35
CR1901	8	19.38	6.01	6.63	2.07	0.38	0.52	1.75	0.89	3.00	0.00
CR1903	8	25.75	3.28	8.75	0.71	0.50	0.58	3.00	0.00	3.00	0.00
CR1904	8	19.06	8.30	4.38	3.62	0.00	0.00	2.25	1.04	2.88	1.13
CR2524	8	23.13	3.58	6.50	2.27	0.50	0.53	2.63	0.52	3.00	0.00
CR2526	8	24.91	9.37	5.50	4.69	0.63	0.74	2.25	1.26	3.19	0.75
CR2528	8	26.31	3.96	8.50	1.07	0.50	0.58	3.00	0.00	3.63	0.52
CR2529	8	14.13	7.86	3.38	3.50	0.00	0.00	1.50	1.20	2.75	1.04
CR2560	8	19.88	7.98	5.75	3.49	0.50	0.53	1.88	1.25	3.00	0.00
CR2563	8	25.38	2.18	8.25	1.04	0.50	0.58	3.00	0.00	3.25	0.46
CR2566	7	21.64	6.65	7.43	1.72	0.71	0.76	2.29	0.76	3.29	0.49

CR2568	8	14.44	5.17	4.50	2.20	0.00	0.00	1.38	1.06	2.88	0.35
CR2569	4	26.63	3.64	6.50	0.58	1.00	0.00	2.50	0.58	3.75	0.50
CR2577	8	8.94	5.66	1.38	2.45	0.31	0.46	1.00	0.93	2.00	0.93
CR2820	8	19.81	6.64	5.25	4.46	0.75	0.46	2.25	0.89	2.50	0.76
CR2821	8	21.50	3.81	7.13	0.83	0.00	0.00	2.75	0.89	3.25	0.46
CR2822	8	23.31	5.84	6.25	1.28	0.38	0.52	2.38	0.74	3.75	0.46
CR2825	8	25.75	4.27	9.63	0.52	0.50	0.53	3.00	0.00	3.50	0.53
CR2826	8	13.69	6.86	3.38	2.83	0.13	0.23	1.63	0.92	2.63	0.74
CR2828	4	22.00	0.58	7.75	0.50	-	-	3.00	0.00	3.00	0.00
CR2829	8	5.88	6.29	0.50	0.76	0.00	0.00	0.50	0.76	1.88	1.13
CR2830	7	13.43	3.83	1.00	2.65	0.29	0.49	1.71	0.76	2.33	1.03
CR2832	8	14.69	6.57	3.75	2.76	0.25	0.46	1.38	0.92	2.75	0.89
CR2833	8	24.38	1.96	8.50	0.76	1.00	0.00	2.75	0.46	3.00	0.00
CR2834	4	19.50	2.71	7.75	1.26	0.00	0.00	2.00	0.82	3.00	0.00
CR2835	8	19.00	4.60	7.00	1.07	0.00	0.00	2.25	0.89	3.00	0.00
CR2837	4	22.75	3.77	7.50	1.00	-	-	2.75	0.50	3.00	0.00
CR2839	8	8.38	8.61	3.13	3.36	0.00	0.00	1.00	1.20	1.75	1.39
CR2840	4	15.88	5.66	3.50	4.04	0.00	0.00	2.00	1.15	2.75	0.96
CR2842	8	23.25	2.45	7.13	0.83	0.25	0.50	2.75	0.46	3.25	0.46
CR2864	8	17.00	8.66	5.38	3.38	0.00	0.00	1.50	1.20	3.00	0.53
CR2865	8	10.88	5.07	2.50	2.20	0.13	0.35	0.38	0.74	2.75	1.04
CR2866	8	19.50	7.22	5.75	2.49	0.75	0.71	1.63	0.92	3.13	0.35
CR2867	8	20.19	7.03	5.50	2.62	0.31	0.59	2.00	0.93	3.25	0.71
CR2868	4	11.63	8.64	2.25	2.87	0.25	0.50	0.75	0.96	2.25	0.96
CR2869	8	28.38	2.60	8.50	1.07	1.13	0.35	-	-	3.25	0.46
CR2870	8	22.88	3.20	6.38	1.30	0.88	0.35	2.25	0.71	3.00	0.00
CR2875	8	21.81	8.24	5.75	3.73	0.50	0.53	1.50	0.58	3.13	0.64
CR2876	8	16.63	4.82	4.63	2.62	0.13	0.35	0.75	0.50	3.00	0.00
CR2877	8	21.94	4.27	7.63	1.30	0.13	0.35	-	-	3.13	0.35
CR2878	8	17.25	5.04	4.75	2.71	0.00	0.00	1.75	0.96	3.00	0.53
CR2879	8	14.25	7.37	3.75	3.45	0.13	0.35	0.50	0.58	2.75	0.46
CR2881	8	10.19	6.85	5.25	4.65	0.00	0.00	0.00	0.00	2.63	0.74
CR2884	8	20.00	4.38	7.88	2.59	0.25	0.46	2.00	1.15	3.00	0.00
CR2948	7	17.50	8.79	4.86	2.73	0.57	0.53	1.71	1.25	2.71	0.49
CR2949	8	22.19	5.62	7.13	3.09	0.38	0.52	-	-	2.88	0.35
CR2951	8	7.50	5.72	2.25	2.49	0.00	0.00	0.75	0.96	2.00	1.31
CR2952	8	26.38	2.55	7.63	1.30	0.50	0.58	2.75	0.50	3.25	0.46
CR2953	8	16.25	3.92	4.25	1.98	0.00	0.00	1.75	0.71	3.00	0.00
CR2955	8	14.50	9.73	4.75	3.33	0.00	0.00	2.25	0.50	2.25	1.39
CR2956	8	15.81	13.27	4.63	4.10	0.00	0.00	2.75	0.50	2.13	1.25
CR2957	8	19.63	3.52	5.88	0.35	0.50	0.53	2.00	0.00	3.00	0.00
CR2959	0	-	-	-	-	-	-	-	-	-	-
CR2961	8	11.50	6.38	1.88	2.95	0.00	0.00	1.13	0.99	2.38	0.92
CR2962	8	18.44	6.49	5.13	3.60	0.00	0.00	1.50	0.58	3.13	0.35
CR2963	8	19.69	3.06	4.63	3.85	0.63	0.52	-	-	2.63	0.52
CR2964	8	10.94	10.32	3.88	4.16	0.00	0.00	0.25	0.50	1.63	1.51



CR2965	8	11.50	7.91	2.38	3.16	0.38	0.52	1.00	0.82	2.00	1.20
CR2966	8	17.88	5.45	5.13	2.70	0.31	0.46	2.00	1.15	2.88	0.35
CR2967	8	13.31	5.37	1.50	2.14	0.25	0.46	1.25	0.50	2.13	0.99
CR2968	8	19.56	7.22	4.63	3.38	0.00	0.00	2.13	0.99	3.13	0.83
CR2970	8	10.69	3.29	2.38	2.00	0.00	0.00	0.50	0.53	2.88	0.35
CR2971	8	7.25	3.82	0.63	1.77	0.00	0.00	0.50	0.53	2.00	0.93
CR2973	8	20.94	6.50	5.50	3.25	0.00	0.00	2.25	0.89	2.88	0.35
CR2975	7	24.14	3.72	7.29	3.55	1.00	0.00	2.33	0.58	3.43	0.53
CR2976	8	10.13	1.94	1.13	1.55	0.25	0.38	0.75	0.46	2.25	0.71
CR2977	8	2.44	3.92	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.64
CR2980	8	13.50	5.16	4.38	3.07	0.00	0.00	1.00	0.00	2.75	0.46
CR2981	8	21.63	2.95	5.63	1.30	0.25	0.46	2.50	0.53	3.00	0.00
CR2982	8	15.25	3.45	3.75	0.71	0.00	0.00	1.63	0.92	3.00	0.00
CR2983	8	21.69	4.46	6.75	1.91	0.25	0.50	2.25	0.50	3.13	0.35
CR2985	8	13.19	4.83	2.38	2.00	0.00	0.00	1.25	0.71	2.63	0.52
CR2988	8	17.38	5.51	5.25	1.28	0.25	0.50	1.75	0.89	3.25	0.46
CR3771	8	18.06	3.53	4.63	1.85	0.00	0.00	1.75	0.89	3.38	0.52
CR3772	8	8.75	5.81	2.00	2.14	0.00	0.00	0.75	0.89	2.25	0.89
CR3773	8	21.06	4.62	6.25	2.71	0.25	0.50	2.50	0.58	3.13	0.64
CR3774	8	21.50	3.21	5.50	1.20	0.25	0.46	2.50	0.53	3.00	0.00
CR3775	8	22.25	3.28	5.38	1.51	0.00	0.00	2.38	0.52	3.50	0.53
CR3776	0	-	-	-	-	-	-	-	-	-	-
CR3777	8	14.25	4.77	3.63	2.88	0.13	0.35	1.50	0.58	2.75	0.46
CR3778	8	21.00	5.73	5.88	3.87	0.25	0.46	-	-	2.75	0.46
CR3779	8	21.31	3.62	5.25	1.58	0.00	0.00	2.50	0.76	3.63	0.52
CR3780	0	-	-	-	-	-	-	-	-	-	-
CR3781	8	16.50	3.74	4.50	2.00	0.00	0.00	1.75	0.71	3.00	0.00
CR3783	8	4.88	3.81	0.25	0.71	0.06	0.18	0.50	0.53	1.50	0.76
CR3784	8	13.13	2.57	3.75	2.49	0.00	0.00	1.00	0.00	2.75	0.46
CR3785	8	22.13	5.82	5.50	2.07	0.00	0.00	2.13	0.99	3.25	0.46
CR3786	8	12.13	5.26	3.63	2.77	0.00	0.00	0.25	0.50	2.75	0.46
CR3787	8	10.81	7.23	2.88	2.80	0.00	0.00	1.38	1.06	2.38	0.92
CR3788	8	17.13	5.68	4.63	2.67	0.00	0.00	1.75	1.16	3.00	0.00
CR3789	8	10.50	7.44	3.00	3.74	0.00	0.00	0.00	0.00	2.25	1.04
CR3790	8	16.31	5.15	2.75	1.39	0.25	0.46	1.50	0.53	3.00	0.53
CR3791	8	20.44	5.63	7.25	1.04	0.00	0.00	3.00	0.00	2.88	0.35
CR3792	4	22.75	4.27	7.00	0.82	-	-	2.75	0.50	3.00	0.00
CR3793	8	18.81	4.17	5.88	0.83	0.00	0.00	2.00	0.76	3.13	0.35
CR3794	8	21.50	2.88	8.38	1.51	0.38	0.52	-	-	3.13	0.35
CR3796	8	15.13	4.80	2.13	1.25	0.25	0.46	1.25	0.71	3.00	0.53
CR3797	8	26.63	2.95	8.75	0.89	0.50	0.58	3.00	0.00	3.00	0.00
CR3798	8	23.50	4.77	8.50	2.45	0.25	0.46	-	-	3.00	0.00
CR3799	8	21.00	5.09	8.75	1.16	0.00	0.00	3.00	0.00	3.25	0.46
CR3800	4	24.75	3.30	7.75	1.71	-	-	2.75	0.50	3.00	0.00
CR3801	8	24.13	2.43	6.38	1.77	0.00	0.00	2.88	0.35	3.25	0.46
CR3802	8	14.31	5.89	4.00	2.27	0.13	0.35	1.38	1.06	3.13	0.35

SD standard deviation, <sup>a</sup> susceptible control, <sup>b</sup> resistant control and <sup>c</sup> parent

**Supplementary Table S3. Candidate genes for significant quantitative trait loci for Fo072 resistance in CR0040 genome annotation.**

QTL name	Trait	Haplotype A			Haplotype B		
		Gene name	Gene location	Gene annotation	Gene name	Gene location	Gene description
<i>qFo72_02.4</i>	d.myc	VANPL_A_00001t014380	A01:26899445..26900200	putative Peptidyl-prolyl cis-trans isomerase CYP37, chloroplastic	VANPL_B_00001t013840	B01:24622242..24649994	Peptidyl-prolyl cis-trans isomerase CYP37, chloroplastic
<i>qFo72_02.3</i>	AUDPC	VANPL_A_00001t014390	A01:26904734..26910330	putative cell wall protein	-	-	-
		VANPL_A_00001t014400	A01:26910333..26928256	vacuolar protein sorting-associated protein 22 homolog 1	VANPL_B_00001t013850	B01:24656605..24671814	vacuolar protein sorting-associated protein 22 homolog 1
		VANPL_A_00001t014410	A01:26933474..26935395	conserved hypothetical protein	VANPL_B_00001t013870	B01:24678297..24679323	conserved hypothetical protein
		VANPL_A_00001t014420	A01:26936070..26937023	conserved hypothetical protein	VANPL_B_00001t013880	B01:24679754..24680748	conserved hypothetical protein
		VANPL_A_00001t014430	A01:26938456..26939794	putative protein FAR1-RELATED SEQUENCE	VANPL_B_00001t013890	B01:24680984..24685544	putative protein FAR-RED IMPAIRED RESPONSE 1
		VANPL_A_00001t014440	A01:26941416..26946131	putative Glutamine amidotransferase type-2 domain-containing protein	-	-	-
		VANPL_A_00001t014450	A01:26957245..26962078	putative asparagine synthetase domain-containing protein 1 isoform v <sub>7</sub>	VANPL_B_00001t013900	B01:24685546..24706131	asparagine synthetase domain-containing protein 1 isoform X1
		VANPL_A_00001t014460	A01:26962516..26963359	putative Glutathione synthetase, chloroplastic	VANPL_B_00001t013910	B01:24706163..24708188	putative Glutathione synthetase, chloroplastic
		VANPL_A_00001t014470	A01:26965525..26966868	putative Glutathione synthetase, chloroplastic	VANPL_B_00001t013920	B01:24709243..24710944	conserved hypothetical protein
		VANPL_A_00001t014480	A01:26968029..26974329	putative Glutathione synthetase, chloroplastic	VANPL_B_00001t013930	B01:24711417..24718135	putative glutathione synthetase, chloroplastic [Phalaenopsis equestris]
		VANPL_A_00001t014490	A01:26974981..26980807	Calcium-transporting ATPase, endoplasmic reticulum-type	VANPL_B_00001t013940	B01:24718877..24724572	Calcium-transporting ATPase, endoplasmic reticulum-type
		VANPL_A_00001t014500	A01:26981185..26983665	DUF569 domain-containing protein	VANPL_B_00001t013950	B01:24725018..24727468	DUF569 domain-containing protein
		VANPL_A_00001t014510	A01:26988271..27014016	putative elongation factor TypA-like SVR3, chloroplastic	VANPL_B_00001t013960	B01:24732503..24758132	Translation factor GUF1 like, chloroplastic
<i>qFo72_02.1</i>	d.15	VANPL_A_00001t024020	A01:45820964..45844946	Protein kinase domain-containing protein	VANPL_B_00001t024350	B01:42781113..42805056	Protein kinase domain-containing protein

VANPL_A_00001t024030	A01:45844948..45862532	putative protein YIP4a	VANPL_B_00001t024360	B01:42817489..42822141	Protein YIP4a
VANPL_A_00001t024040	A01:45867834..45868469	putative Egg cell-secreted protein 1.3	VANPL_B_00001t024370	B01:42829398..42829823	Egg cell-secreted protein 1.3
VANPL_A_00001t024050	A01:45875285..45886036	Tubby-like F-box protein 8	VANPL_B_00001t024380	B01:42836824..42847310	Tubby-like F-box protein 8
VANPL_A_00001t024060	A01:45898652..45902645	WAT1-related protein	VANPL_B_00001t024390	B01:42859658..42863632	WAT1-related protein
VANPL_A_00001t024070	A01:45906498..45909997	conserved hypothetical protein	-	-	-
VANPL_A_00001t024080	A01:45947458..45951152	Actin-7 isoform B	VANPL_B_00001t024400	B01:42907845..42911319	Actin
VANPL_A_00001t024090	A01:45961097..45961994	protein MKS1-like [Phalaeopsis equestris]	VANPL_B_00001t024410	B01:42921666..42923560	protein MKS1-like [Phalaeopsis equestris]
VANPL_A_00001t027290	A01:54233558..54244691	50S ribosomal protein L1, chloroplastic	VANPL_B_00001t026080	B01:51095998..51105189	50S ribosomal protein L1, chloroplastic
VANPL_A_00001t027300	A01:54254756..54264784	conserved hypothetical protein	VANPL_B_00001t026090	B01:51115227..51125133	protein RETICULATA-RELATED 4, chloroplastic [Phalaeopsis equestris]
VANPL_A_00001t027310	A01:54267073..54269659	Cation/H(+) antiporter 15	VANPL_B_00001t026100	B01:51127525..51130111	Cation/H(+) antiporter 15
VANPL_A_00001t027320	A01:54274312..54276480	putative membrane-associated kinase regulator 1	VANPL_B_00001t026110	B01:51134766..51136868	putative membrane-associated kinase regulator 1
VANPL_A_00001t027330	A01:54278921..5429061	trafficking protein particle complex subunit 4 [Phalaeopsis equestris]	VANPL_B_00001t026120	B01:51139356..51140409	Trafficking protein particle complex subunit
VANPL_A_00001t027340	A01:54290637..54325391	E3 ubiquitin-protein ligase KEG	VANPL_B_00001t026130	B01:51150841..51186588	E3 ubiquitin-protein ligase KEG
VANPL_A_00001t027350	A01:54325393..54326505	conserved hypothetical protein	VANPL_B_00001t026140	B01:51203073..51203708	putative RNase H domain-containing protein
VANPL_A_00001t027360	A01:54369775..54380251	CDPK-related kinase 1	VANPL_B_00001t026160	B01:51236187..51246352	CDPK-related kinase 1
VANPL_A_00001t027370	A01:54380446..54413859	conserved hypothetical protein	VANPL_B_00001t026170	B01:51266789..51278849	conserved hypothetical protein
VANPL_A_00001t027380	A01:54414167..54424429	DNA topoisomerase 6 subunit A	VANPL_B_00001t026180	B01:51279180..51284758	Thioredoxin-like protein AAED1, chloroplastic
VANPL_A_00001t027390	A01:54425720..54465201	pentatricopeptide repeat-containing protein At5g02830, chloroplastic isoform X2 [Phalaeopsis equestris]	VANPL_B_00001t026200	B01:51290922..51328247	DNA topoisomerase 6 subunit A
VANPL_A_00001t029570	A01:60110295..60153611	Phosphoenolpyruvate phosphatase	VANPL_B_00001t027290	B01:56255461..56298530	putative Phosphoenolpyruvate phosphatase

VANPL_A_00001t029580	A01:60162625..60165286	Monogalactosyldiacylglycerol synthase 2, chloroplastic	VANPL_B_00001t027300	B01:56307647..56310415	Monogalactosyldiacylglycerol synthase		
VANPL_A_00001t029590	A01:60169561..60179936	Protein S-acyltransferase 24	VANPL_B_00001t027310	B01:56314533..56324858	Protein S-acyltransferase 24		
VANPL_A_00001t029600	A01:60182407..60183503	40S ribosomal protein S17	VANPL_B_00001t027320	B01:56327177..56328549	putative 40S ribosomal protein S17		
VANPL_A_00001t029610	A01:60193145..60228083	Threonyl-tRNA synthetase	VANPL_B_00001t027330	B01:56333248..56369355	Threonine-tRNA ligase, mitochondrial 1		
VANPL_A_00001t029620	A01:60228591..60236833	40S ribosomal protein S11 [Phalaenopsis equestris]	VANPL_B_00001t027340	B01:56369748..56375671	40S ribosomal protein S11		
-	-	-	VANPL_B_00001t027360	B01:56411187..56448484	Adenylate kinase 5, chloroplastic		
<hr/>							
<i>qFo72_03.2</i>	<i>d.myc</i>	VANPL_A_00001t032150	A01:65555503..65592615	Cytochrome P450 97B2, chloroplastic	VANPL_B_00001t029670	B01:60646805..60690498	Cytochrome P450 97B2, chloroplastic
		VANPL_A_00001t032160	A01:65596259..65597070	VQ domain-containing protein	VANPL_B_00001t029680	B01:60694031..60694889	VQ domain-containing protein
		VANPL_A_00001t032170	A01:65599661..65601476	Polygalacturonase QRT3	VANPL_B_00001t029690	B01:60698170..60700037	Polygalacturonase QRT3
		VANPL_A_00001t032180	A01:65602800..65604811	Polygalacturonase QRT3	VANPL_B_00001t029700	B01:60701406..60703427	Polygalacturonase QRT3
		VANPL_A_00001t032190	A01:65605921..65608584	Polygalacturonase QRT3	VANPL_B_00001t029710	B01:60704514..60707170	Polygalacturonase QRT3
		VANPL_A_00001t032200	A01:65613329..65651609	Ent-cassa-1,2,15-diene synthase	VANPL_B_00001t029720	B01:60711949..60737232	Ent-kaur-16-ene synthase, chloroplastic
		-	-	-	VANPL_B_00001t029730	B01:60737234..60750339	conserved hypothetical protein
		VANPL_A_00001t032210	A01:656665690..65683501	Protein GID8	VANPL_B_00001t029740	B01:60766908..60784615	Protein GID8
		VANPL_A_00001t032220	A01:65686382..65700071	Secretory carrier-associated membrane protein 3	VANPL_B_00001t029750	B01:60787570..60801017	Secretory carrier-associated membrane protein 3
<hr/>							
<i>qFo72_05.1</i>	<i>d.15</i>	VANPL_A_00003t012620	A03:40745878..40747362	CBL-interacting protein kinase 19	VANPL_B_00003t010200	B03:39008973..39010919	CBL-interacting protein kinase 19
		VANPL_A_00003t012630	A03:40747371..40756918	Protein kinase domain-containing protein	VANPL_B_00003t010210	B03:39011388..39018336	Protein kinase domain-containing protein
		VANPL_A_00003t012640	A03:40757150..40763923	Aspartate aminotransferase, cytoplasmic	VANPL_B_00003t010220	B03:39018347..39025290	Aspartate aminotransferase
		VANPL_A_00003t012650	A03:40764254..40769273	TOM1-like protein 9	VANPL_B_00003t010230	B03:39025579..39030564	TOM1-like protein 9
		VANPL_A_00003t012660	A03:40773524..40774933	putative F-box protein SKIP23	VANPL_B_00003t010240	B03:39034322..39035731	putative DUF295 domain-containing protein

	VANPL_A_00003t012670	A03:40776673..40800689	E3 ubiquitin ligase BIG BROTHER	VANPL_B_00003t010250	B03:39037470..39061386	E3 ubiquitin ligase BIG BROTHER	
	VANPL_A_00003t012680	A03:40801027..40809216	5'-phosphoribosylglycinamide transfermlyase	VANPL_B_00003t010260	B03:39061736..39069922	Phosphoribosylglycinamide formyltransferase, chloroplatic	
	VANPL_A_00003t012690	A03:40809218..40818022	putative L-gulonolactone oxidase 3	VANPL_B_00003t010270	B03:39069924..39076257	Pentatricopeptide repeat-containing protein	
	VANPL_A_00003t012700	A03:40818308..40831778	conserved hypothetical protein	VANPL_B_00003t010280	B03:39076259..39078704	L-gulonolactone oxidase 3	
	VANPL_A_00003t012710	A03:40840706..40841145	transmembrane protein 234 homolog [Phalaenopsis equestris]	VANPL_B_00003t010290	B03:39078714..39088423	conserved hypothetical protein	
	VANPL_A_00003t012720	A03:40843201..40843611	conserved hypothetical protein	VANPL_B_00003t010300	B03:39088425..39099177	transmembrane protein 234 homolog [Phalaenopsis equestris]	
	VANPL_A_00003t012730	A03:40843973..40844625	hypothetical protein	VANPL_B_00003t010310	B03:39101098..39101460	conserved hypothetical protein	
	VANPL_A_00003t012740	A03:40865246..40867031	E3 ubiquitin-protein ligase ATL42-like isoform X1 [Phalaenopsis equestris]	VANPL_B_00003t010320	B03:39102009..39102401	conserved hypothetical protein	
	VANPL_A_00003t012750	A03:40877495..40961268	Two pore calcium channel protein 1	VANPL_B_00003t010330	B03:39114788..39116565	E3 ubiquitin-protein ligase ATL42-like isoform X1 [Phalaenopsis equestris]	
<i>qFo72_06.4</i>	AUDPC	VANPL_A_00004t009010	A04:28125223..28151781	Polyadenylate-binding protein 2	VANPL_00259t000490	B0:97082949..97109527	Polyadenylate-binding protein 2
<i>qFo72_06.5</i>	d.5	VANPL_A_00004t009020	A04:28166489..28180930	putative Cytochrome c oxidase subunit 6b-3	VANPL_00259t000500	B0:97119953..97135214	putative Cytochrome c oxidase subunit 6b-3
		VANPL_A_00004t009030	A04:28181577..28197221	conserved hypothetical protein	VANPL_00259t000510	B0:97135970..97151343	conserved hypothetical protein
		VANPL_A_00004t009040	A04:28209877..28214399	F-box protein FBX14	VANPL_00259t000520	B0:97164135..97168892	F-box protein FBX14
		VANPL_A_00004t009050	A04:28218108..28261941	Protein ALTERED SEED GERMINATION 2	VANPL_00259t000530	B0:97172430..97216424	Protein ALTERED SEED GERMINATION 2
<i>qFo72_06.1</i>	d.myc	VANPL_A_00004t011320	A04:34479949..34491944	Protein argonaute 10	VANPL_B_00004t010870	B04:33594642..33606662	Protein argonaute 10
<i>qFo72_06.2</i>	AUDPC	VANPL_A_00004t011330	A04:34498767..34502725	Beta-carotene isomerase D27, chloroplatic	VANPL_B_00004t010880	B04:33613592..33617576	Beta-carotene isomerase D27, chloroplatic
		VANPL_A_00004t011340	A04:34534732..34541212	conserved hypothetical protein	VANPL_B_00004t010890	B04:33645535..33652009	putative BHLH domain-containing protein
		VANPL_A_00004t011350	A04:34551016..34552135	putative Ion channel POLLUX-like 2	VANPL_B_00004t010900	B04:33652305..336597099	putative ion channel POLLUX-like 2 isoform X1 [Phalaenopsis equestris]

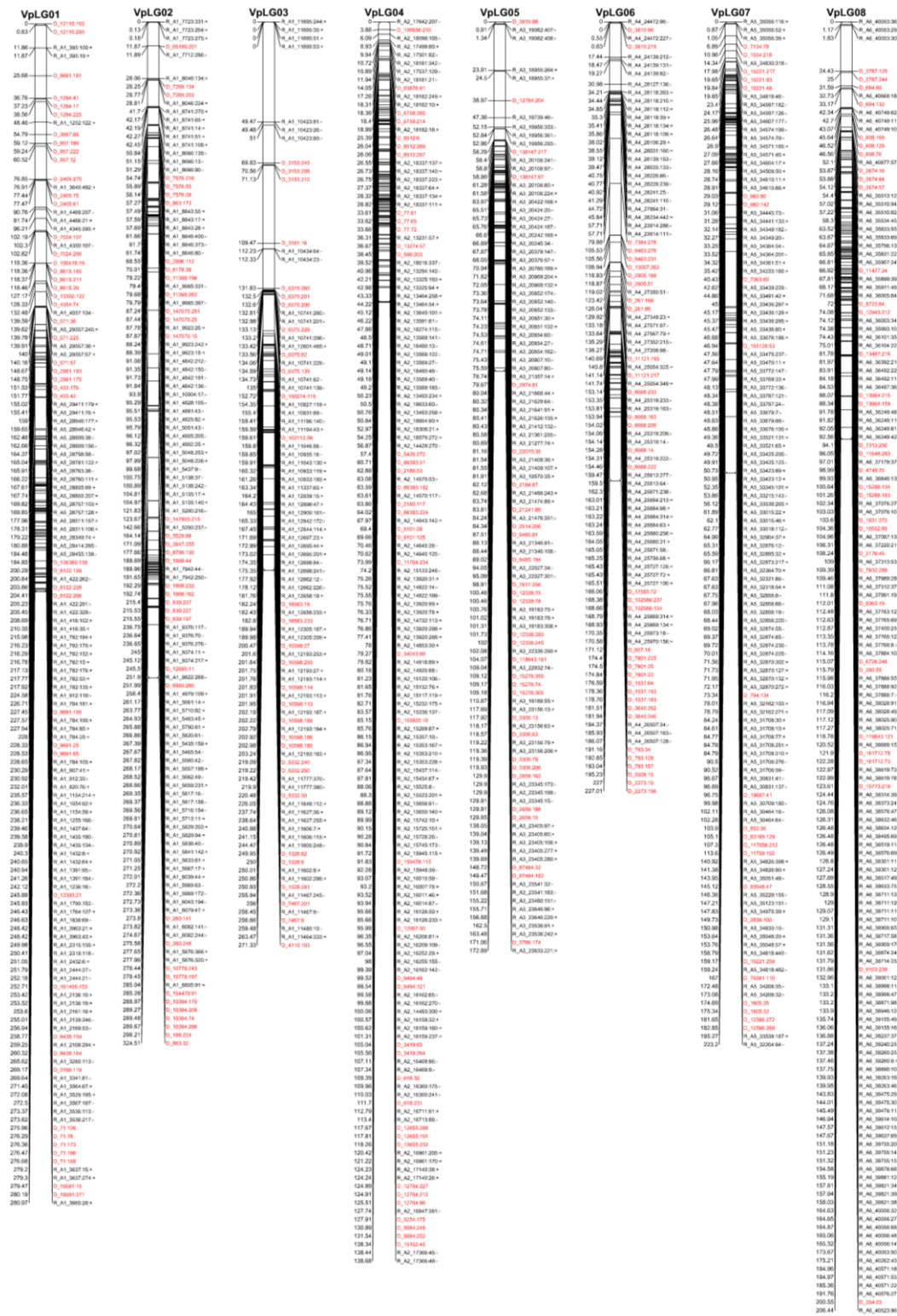
	VANPL_A_00004t011360	A04:34591263..34593213	putative ion channel POLLUX-like 1	VANPL_B_00004t010910	B04:33697101..33699637	putative ion channel POLLUX-like 1
	VANPL_A_00004t011370	A04:34594912..34595463	putative ion channel POLLUX-like 2	VANPL_B_00004t000640	B04:4626094..4675250	ubiquitin-like-specific protease 1D isoform X2 [Phalaenopsis equestris]
<i>qFo72_06.3</i>	AUDPC	VANPL_A_00004t000590	A04:4644613..4691004	VANPL_B_00004t000650	B04:4688844..4713841	RING-box protein 1a [Phalaenopsis equestris]
<i>qFo72_07.1</i>	d.myc	VANPL_A_00005t009820	A05:23172306..23229115	VANPL_B_00005t008130	B05:20573287..20632915	DNA mismatch repair protein PMS1
<i>qFo72_07.2</i>	AUDPC	VANPL_A_00005t009830	A05:23255454..23257835	VANPL_B_00005t008140	B05:20659334..20661751	Pentatricopeptide repeat-containing protein
<i>qFo72_07.3</i>	d.5	VANPL_A_00005t009840	A05:234333485..23461027	VANPL_B_00005t008150	B05:20794482..20821642	double-stranded RNA-binding protein 8-like [Phalaenopsis equestris]
<i>qFo72_08.1</i>	d.myc	VANPL_A_00006t012930	A06:41421413..41445641	VANPL_B_00006t008840	B06:23272603..23295856	Alpha-L-arabinofuranosidase 1
		VANPL_A_00006t012940	A06:41446061..41464702	VANPL_B_00006t008850	B06:23295858..23303736	Aquaporin SIP2-1
		Vp106Ag14341.1	A06:41469645..41482103	VANPL_B_00006t008860	B06:23306136..23322480	Daphna_Dendrobium catenatum leucine-rich repeat protein 1 (LOC110110749)
		VANPL_A_00006t012950	A06:41483492..41484365	VANPL_B_00006t008870	B06:23323347..23324091	conserved hypothetical protein
		VANPL_A_00006t012960	A06:41491560..41493275	VANPL_B_00006t008880	B06:23332065..23333553	Alcohol dehydrogenase 9
		VANPL_A_00006t012970	A06:41502994..41506563	VANPL_B_00006t008890	B06:23345428..23349005	calcium-dependent protein kinase 2-like [Phalaenopsis equestris]
		VANPL_A_00006t012980	A06:41509052..41533517	VANPL_B_00006t008900	B06:23362763..23376134	Transcription factor HBP-1b(c38)
<i>qFo72_09.1</i>	AUDPC	VANPL_A_00007t009910	A07:28560787..28564793	VANPL_B_00007t009100	B07:24257689..24260654	pentatricopeptide repeat-containing protein At2g29760, chloroplastic-like [Phalaenopsis equestris]
		VANPL_A_00007t009930	A07:28615624..28659819	VANPL_B_00007t009110	B07:24305499..24352568	protein CHROMATIN REMODELING 4 [Phalaenopsis equestris]
		VANPL_A_00007t009940	A07:28673353..28675131	VANPL_B_00007t009120	B07:24384521..24386334	globulin-1 S allele [Phalaenopsis equestris]
		VANPL_A_00007t009950	A07:28683425..28687815	VANPL_B_00007t009130	B07:24394733..24399161	Type I inositol polyphosphate 5-phosphatase 4

<i>qFo72_12.2</i>	d.myc	VANPL_A_00010t005640	A10:13245170..13251408	putative Glucan endo-1,3-beta-glucosidase 14	VANPL_B_00010t005530	B10:13449644..13455857	Glucan endo-1,3-beta-glucosidase 14		
		VANPL_A_00010t005650	A10:13252434..13257774	Phospholipase D alpha 1	VANPL_B_00010t005540	B10:13457725..13462126	Phospholipase D alpha 1		
		VANPL_A_00010t005660	A10:13258404..13268037	conserved hypothetical protein	VANPL_B_00010t005550	B10:13462699..13475371	conserved hypothetical protein		
		VANPL_A_00010t005670	A10:13268039..13270512	putative Eugenol synthase 2	VANPL_B_00010t005590	B10:13506921..13508477	Phenylcoumaran benzylic ether reductase Betv6		
		VANPL_A_00010t005680	A10:13279055..13285645	putative Phenylcoumaran benzylic ether reductase Betv6	VANPL_B_00010t005610	B10:13513992..13519386	hypothetical protein		
		VANPL_A_00010t005690	A10:13286312..13292181	hypothetical protein	VANPL_B_00010t005620	B10:13525787..13526986	Transcription factor RADIALIS		
		VANPL_A_00010t005700	A10:13298612..13299838	Transcription factor RADIALIS	VANPL_B_00010t005630	B10:13533178..13552747	ATPase family AAA domain-containing protein 3-like [Phalaenopsis equestris]		
		VANPL_A_00010t005710	A10:13303637..13323258	ATPase family AAA domain-containing protein 3-like	VANPL_B_00010t005640	B10:13552971..13568823	DNA-3-methyladenine glycosylase II		
		VANPL_A_00010t005720	A10:13323510..13339228	DNA-3-methyladenine glycosylase II	VANPL_B_00010t005650	B10:13585969..13589573	Protein KINESIN LIGHT CHAIN-RELATED 2		
		VANPL_A_00010t005730	A10:13353186..13359791	Protein KINESIN LIGHT CHAIN-RELATED 2	VANPL_B_00010t005690	B10:13620588..13624478	Peptidyl-prolyl cis-trans isomerase CYP38, chloroalastic		
		VANPL_A_00010t005770	A10:13386469..13389965	Peptidyl-prolyl cis-trans isomerase CYP38, chloroalastic	Vpl11Ag20841.1	B11:61967..62747	hypothetical protein		
		<i>qFo72_12.1</i>	d.myc	Vpl11Ag20841.1	A11:281702..282481	hypothetical protein	VANPL_B_00011t000010	B11:93359..93718	putative WEB family protein At2g38370
				VANPL_A_00011t000010	A11:295539..295898	putative WEB family protein At2g38370	VANPL_B_00011t000030	B11:96085..98171	conserved hypothetical protein
				VANPL_A_00011t000020	A11:298265..300344	putative Nsp1_C domain-containing protein	VANPL_B_00011t000040	B11:113346..150656	F-actin-capping protein subunit beta
VANPL_A_00011t000030	A11:312659..345281			putative F-actin-capping protein subunit beta	VANPL_B_00011t000050	B11:155585..178200	protein LHY-like isoform X3 [Phalaenopsis equestris]		
VANPL_A_00011t000040	A11:354509..376721			protein LHY-like isoform X3 [Phalaenopsis equestris]	VANPL_B_00011t000060	B11:188135..190402	Pentatricopeptide repeat-containing protein At5g15340, mitochondrial		
VANPL_A_00011t000050	A11:376723..377368			hypothetical protein	VANPL_B_00011t000070	B11:193768..198226	NAC domain-containing protein 74		
VANPL_A_00011t000060	A11:387520..389556			Pentatricopeptide repeat-containing protein At5g15340, mitochondrial	VANPL_B_00011t000080	B11:198691..220348	E3 ubiquitin-protein ligase MARCH6		
VANPL_A_00011t000070	A11:392843..397514			NAC domain-containing protein 74	VANPL_B_00012t000150	B12:357119..361590	F-box protein		
VANPL_A_00011t000080	A11:397768..419678			E3 ubiquitin-protein ligase MARCH6	VANPL_B_00012t000160	B12:365574..407961	Oligopeptidase B		
VANPL_A_00012t000150	A12:440699..476024			F-box protein	VANPL_B_00012t000170	-	-		
VANPL_A_00012t000160	A12:481190..489637			putative Peptidase S9A/B/C	-	-	-		
VANPL_A_00012t000170	A12:494208..495029			conserved hypothetical protein	-	-	-		

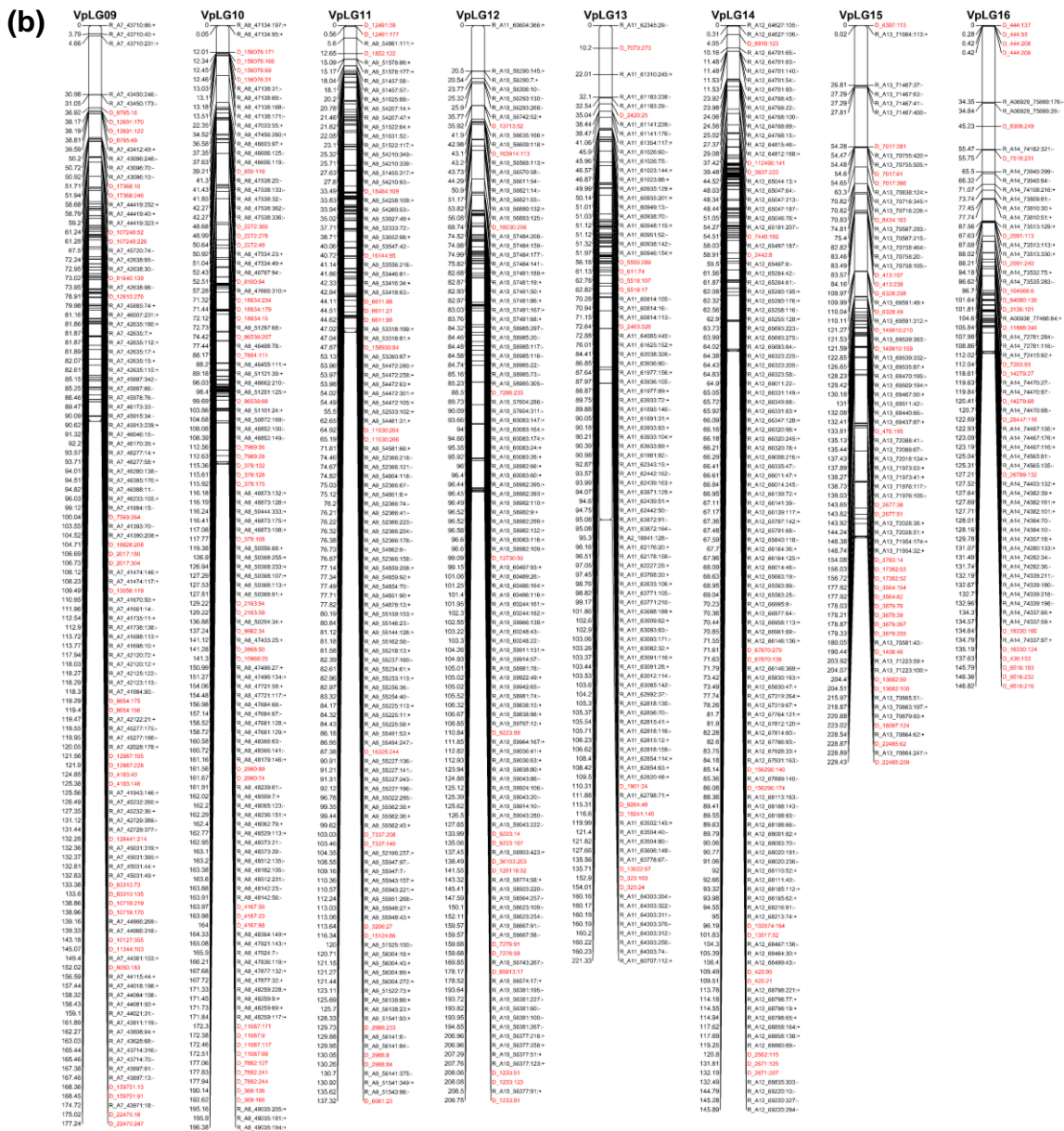
VANPL_A_00012t000180	A12:517436..531601	ABC transporter I family member 19	VANPL_B_00012t000170	B12:408752..409094	ABC transporter I family member 19
VANPL_A_00012t000190	A12:543781..588422	Plastid division protein CDP1, chloroplast	VANPL_B_00012t000180	B12:420925..465161	Plastid division protein CDP1, chloroplast



(a)

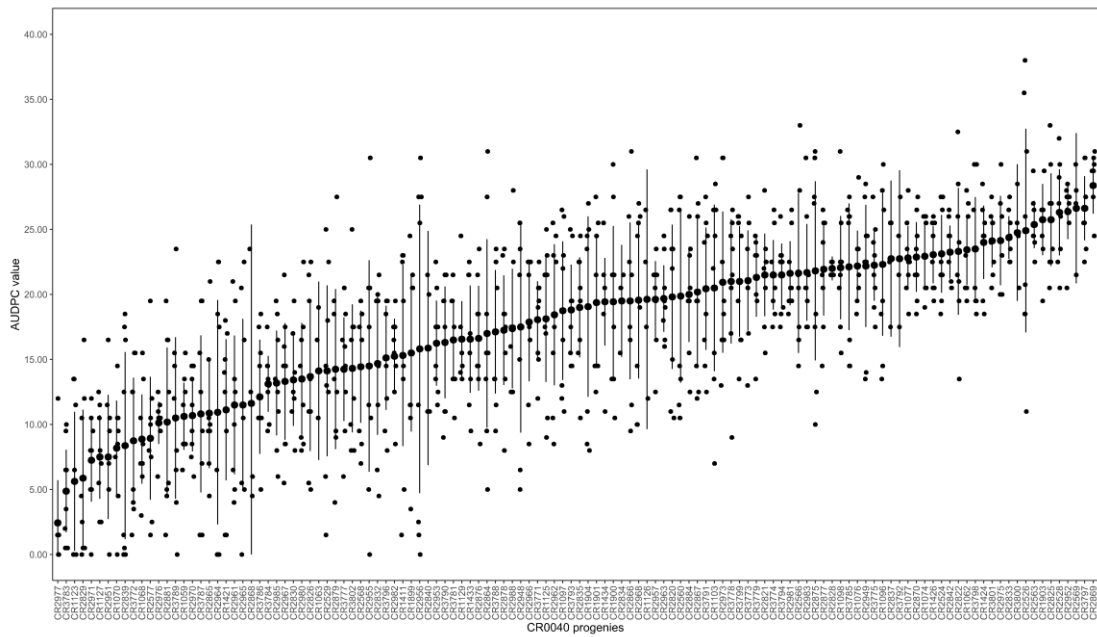


**Supplementary Figure S1.** Genetic linkage map of the population, with 1,804 markers across the linkage groups (a) VpLG01 to VpLG08 and (b) VpLG09 to VpLG16. Genetic distances (cM) are indicated on the left and markers names on the right of the map bars. The colors indicate *de novo* (red) and reference-based (black) markers.

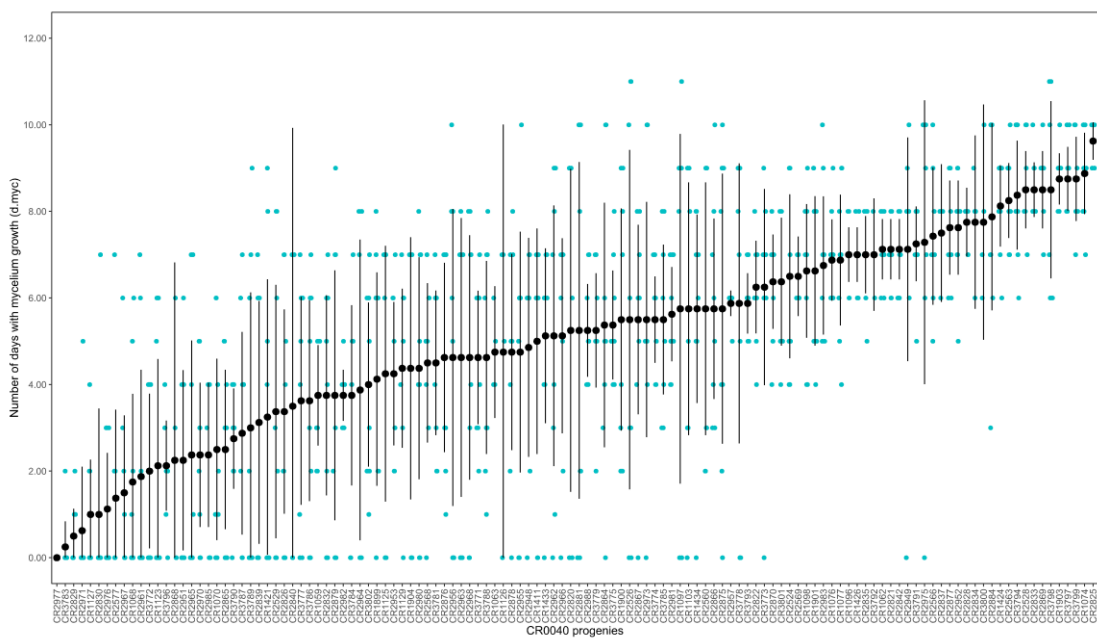


**Supplementary Figure S1 (continued).** Genetic linkage map of the population, with 1,804 markers across the linkage groups (a) VpLG01 to VpLG08 and (b) VpLG09 to VpLG16. Genetic distances (cM) are indicated on the left and markers names on the right of the map bars. The colors indicate *de novo* (red) and reference-based (black) markers.

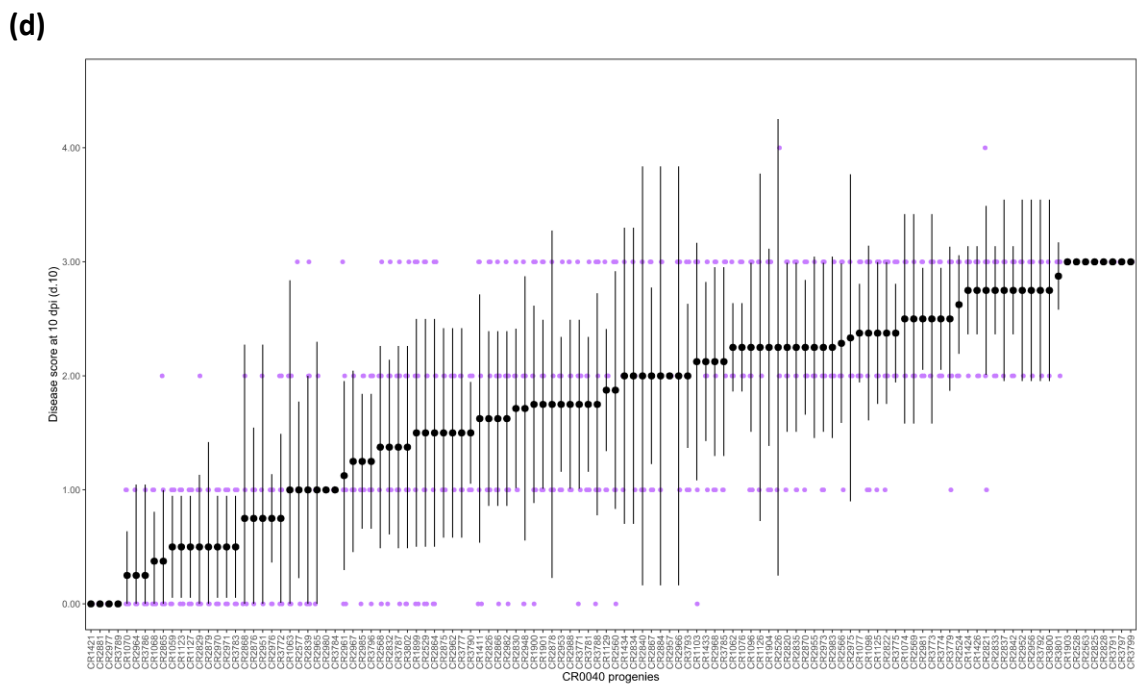
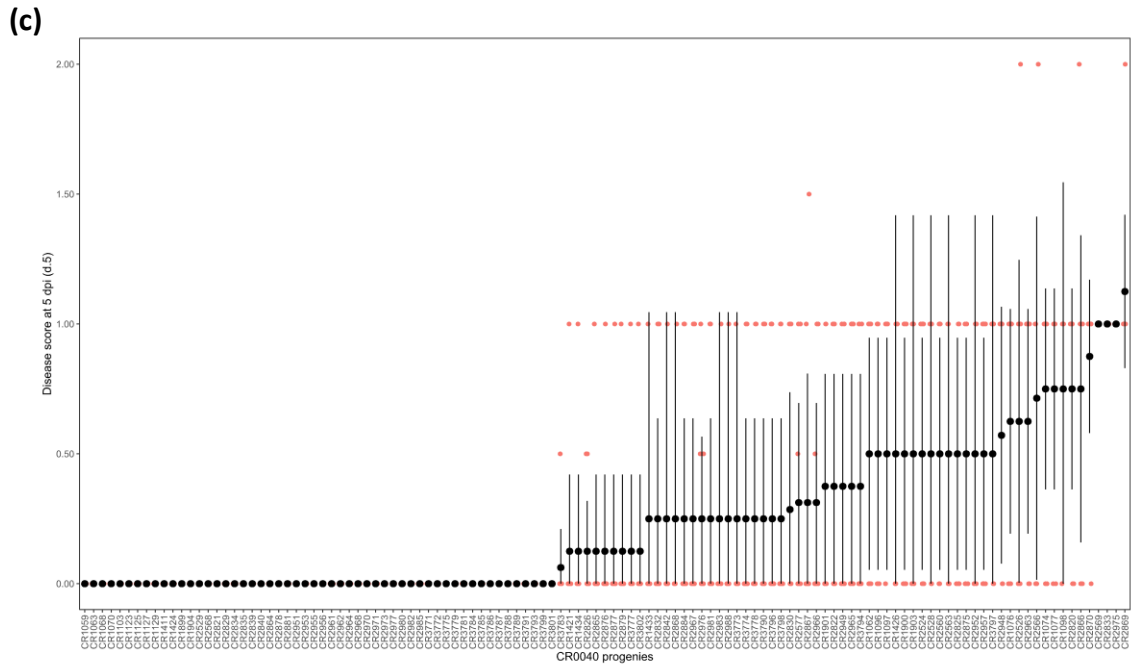
(a)



(b)

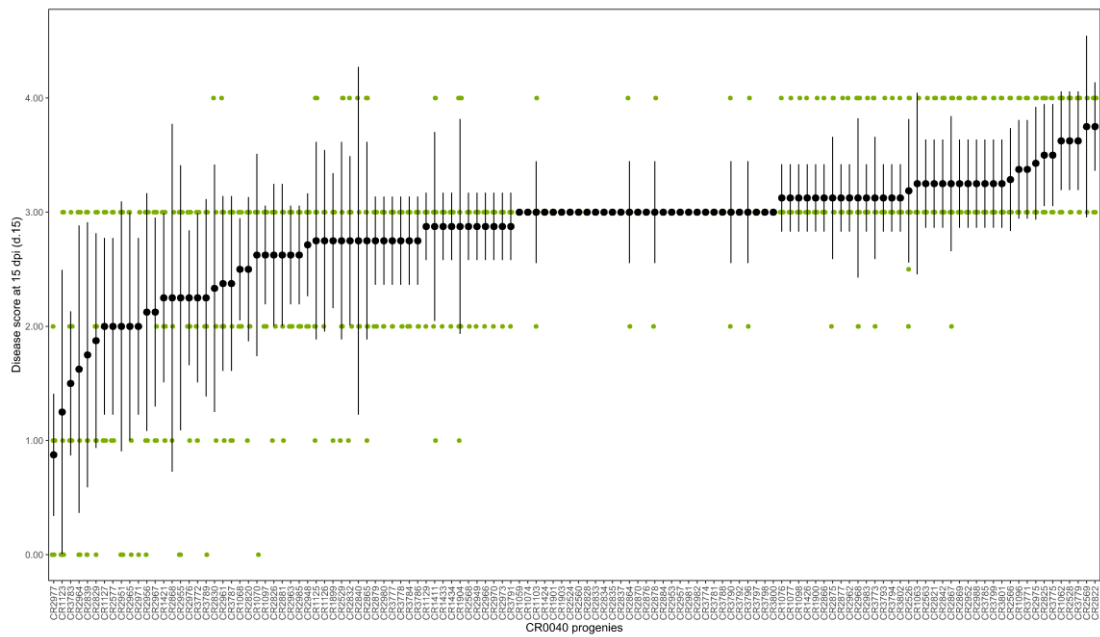


**Supplementary Figure S2.** Phenotyping of CR0040 progenies. The x-axis indicates the name of the accessions and the y-axis **(a)** the AUDPC values, **(b)** the number of days with mycelium growth (d.myc) and the disease score at **(c)** 5 days post-inoculation (dpi, d.5), **(d)** 10 dpi (d.10) and **(e)** 15 dpi (d.15). Each colored point indicates a replicate value. Mean values and standard error of the mean are represented respectively by the black points and error bars.



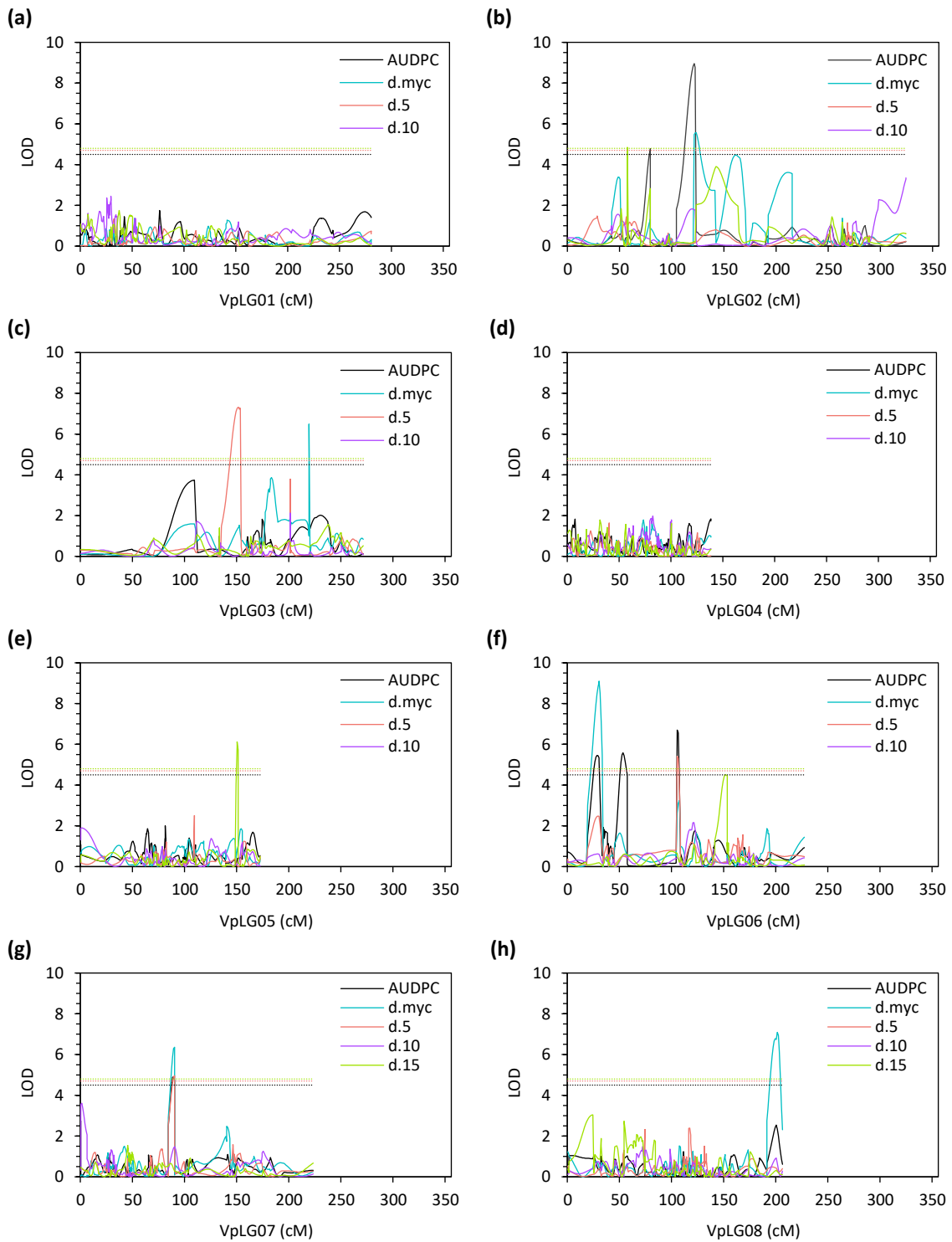
**Supplementary Figure S2 (continued).** Phenotyping of CR0040 progenies. The x-axis indicates the name of the accessions and the y-axis **(a)** the AUDPC values, **(b)** the number of days with mycelium growth (d.myc) and the disease score at **(c)** 5 days post-inoculation (dpi, d.5), **(d)** 10 dpi (d.10) and **(e)** 15 dpi (d.15). Each colored point indicates a replicate value. Mean values and standard error of the mean are represented respectively by the black points and error bars.

(e)

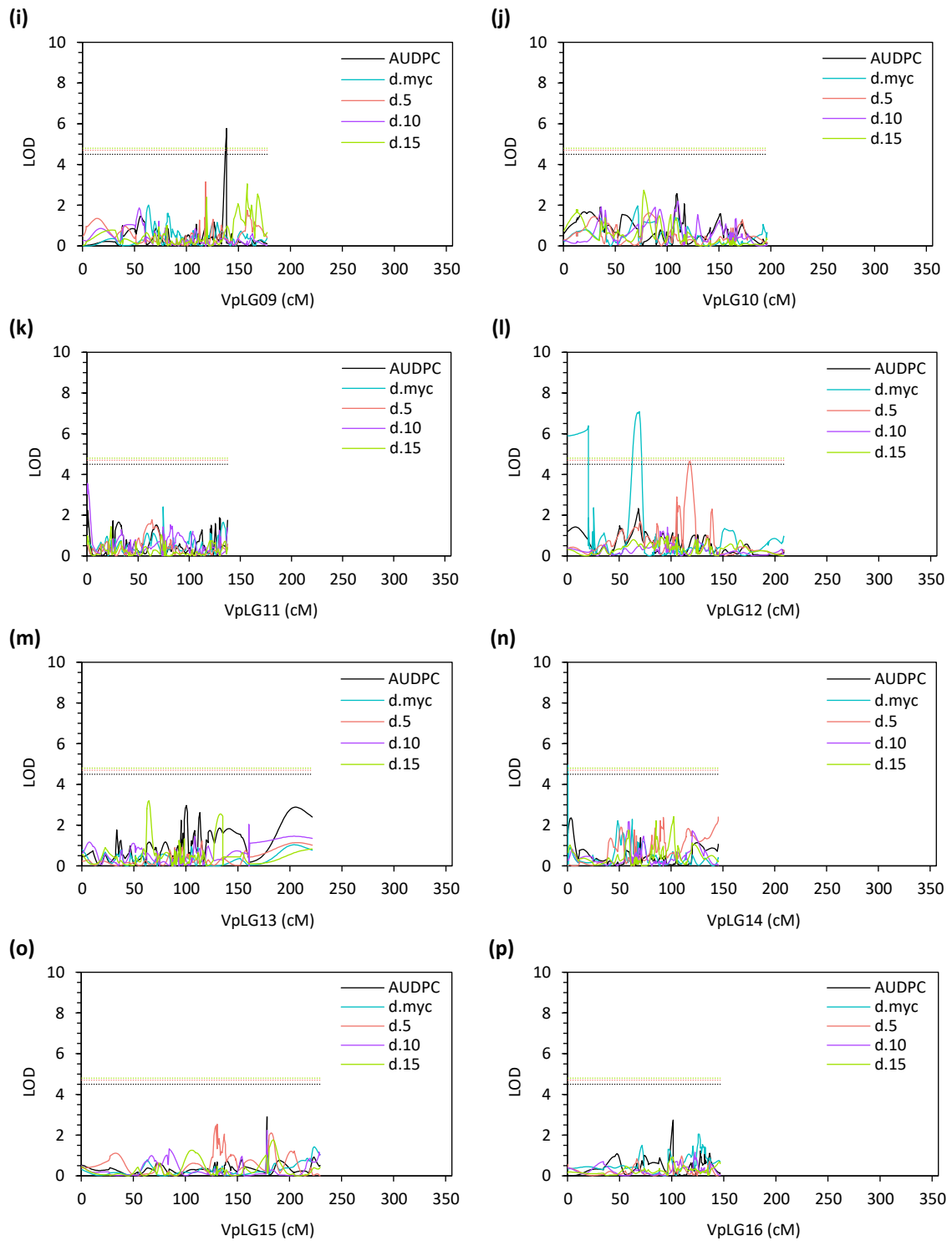


**Supplementary Figure S2 (continued).** Phenotyping of CR0040 progenies. The x-axis indicates the name of the accessions and the y-axis **(a)** the AUDPC values, **(b)** the number of days with mycelium growth (d.myc) and the disease score at **(c)** 5 days post-inoculation (dpi, d.5), **(d)** 10 dpi (d.10) and **(e)** 15 dpi (d.15). Each colored point indicates a replicate value. Mean values and standard error of the mean are represented respectively by the black points and error bars.





**Supplementary Figure S3.** Quantitative trait loci for AUDPC (black), d.5 (pink), d.10 (purple), d.15 (green) and d.myc (blue) on the linkage groups VpLG01 to VpLG16 (a-p). The dotted lines indicate the LOD threshold of 4.7 for d.5 (pink), 4.8 for d.15 (green) and 4.5 for AUDPC, d.10 and d.myc (black). LOD, logarithm of odds.



**Supplementary Figure S3 (continued).** Quantitative trait loci for AUDPC (black), d.5 (pink), d.10 (purple), d.15 (green) and d.myc (blue) on the linkage groups VpLG01 to VpLG16 (**a-p**). The dotted lines indicate the LOD threshold of 4.7 for d.5 (pink), 4.8 for d.15 (green) and 4.5 for AUDPC, d.10 and d.myc (black). LOD, logarithm of odds.





RGa LRR Profiler	VANPL_000501000920	RLK	LRR receptor-like serine/threonine-protein kinase SIK1	A	0	VANPL_A_00000:96213666..96219751	
RGa LRR Profiler	VANPL_000501000960	RLK	Protein kinase domain-containing protein	A	0	VANPL_A_00000:96240922..96254209	
RGa Fusa bananier FORL	VANPL_000501001010	POD	Peroxidase 4	B	0	VANPL_00050:2489155..2513111	
RGa Fusa bananier	VANPL_000531001410	ASP	Serine/threonine-protein kinase BSK2	B	0	VANPL_B_00000:275640116..275657379	
RGa Fusa bananier	VANPL_000531001940	NLR	Protein ASPARTIC PROTEASE IN GUARD CELL 2	A	0	VANPL_A_00000:3271061..3273556	
RGa LRR Profiler	VANPL_000941000070	RLK	Transcription factor MYB106	A	0	VANPL_A_00000:197900144..197901697	
RGa LRR Profiler	VANPL_000941000110	RLK	Tyrosine-sulfated glycopeptide receptor 1	A	0	VANPL_A_00000:197998802...198002262	
RGa LRR Profiler	VANPL_000941000810	RLK	Protein kinase domain-containing protein	A	0	VANPL_A_00000:199594310..199597413	
RGa LRR Profiler	VANPL_001111000310	RLK	Protein kinase domain-containing protein	B	0	VANPL_B_00000:97876715..97879693	
RGa Fusa bananier	VANPL_001111000350	POD	Peroxidase 4	B	0	VANPL_00111:553915..555463	
RGa LRR Profiler	VANPL_001451000290	RLP	putative leucine-rich repeat receptor-like protein kinase At5g49770	B	0	VANPL_B_00000:312411811..312413486	
Phénylpropanoïde	VANPL_001701000450	4-HBS ?	Vanillin synthase	B	0	VANPL_B_00000:330742707..330745551	
RGa Fusa bananier	VANPL_001811000120	NLR	MYB-like transcription factor 4	B	0	VANPL_00181:407115..408112	
RGa LRR Profiler	VANPL_001981000010	RLP	putative leucine-rich repeat receptor-like protein kinase At5g49770	B	0	VANPL_B_00000:219676492..219678165	
RGa LRR Profiler	VANPL_002081000030	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1	A	0	VANPL_A_00000:311935603..311939148	
RGa LRR Profiler	VANPL_003481000440	RLK	Leucine-rich repeat receptor protein kinase HPCA1	B	0	VANPL_B_00000:407004240..407009211	
Phénylpropanoïde	VANPL_005131000070	4CL	Oxalate--Coa ligase	A	0	VANPL_A_00000:478281286..478283355	
RGa Fusa autres espèces / RGA Daphna / RGA LRR Profiler / RGA Fusa bananier	VANPL_007701000090	RLK	Disease resistance protein RGA3	B	0	VANPL_B_00000:224259030..224262364	
RGa LRR Profiler	VANPL_009461000010	GLIP	putative protein kinase domain-containing protein	B	0	VANPL_B_00000:119688487..119692088	
RGa Fusa bananier	VANPL_013011000130	HCT	GDSL esterase/lipase At5g5050 [Phalaenopsis equestris]	B	0	VANPL_01301:144015..145781	
Phénylpropanoïde	VANPL_016371000010	HCT	putative Shikimate O-hydroxycinnamoyltransferase	B	0	VANPL_B_00000:8621023..8622673	
Phénylpropanoïde	VANPL_016371000050	HCT	putative Shikimate O-hydroxycinnamoyltransferase	B	0	VANPL_B_00000:8686624..8704499	
Phénylpropanoïde	VANPL_016371000070	HCT	putative Shikimate O-hydroxycinnamoyltransferase	B	0	VANPL_B_00000:8746388..8748047	
Phénylpropanoïde	VANPL_049091000020	HCT	omega-hydroxypalmitate O-feruloyl transferase [Phalaenopsis equestris]	B	0	VANPL_B_00000:942432680..942434270	
RGa LRR Profiler	VANPL_051361000010	RLK	receptor protein kinase-like protein ZAR1 [Phalaenopsis equestris]	B	0	VANPL_B_00000:955834104..955836227	
Phénylpropanoïde	VANPL_053841000040	CAD	PKS_ER domain-containing protein	B	0	VANPL_B_00000:970767055..970781799	
RGa LRR Profiler	VANPL_055621000020	RLK	conserved hypothetical protein	B	0	VANPL_B_00000:980567416..980572522	
RGa LRR Profiler	VANPL_073531000050	RLK	receptor-like protein kinase HSL1 [Phalaenopsis equestris]	B	0	VANPL_B_00000:1071794653..1071798951	
Phénylpropanoïde	VANPL_122771000010	4CL	putative 4-coumarate--CoA ligase-like 5 [Phalaenopsis equestris]	B	0	VANPL_B_00000:95962013..95981903	
RGa LRR Profiler	VANPL_129021000010	RLK	putative receptor-like protein kinase At5g18500	B	0	VANPL_B_00000:1275983967..1275997442	
RGa Fusa bananier	VANPL_142081000020	NLR	Transcription factor MYB106	B	0	VANPL_B_00000:156441921..156443378	
RGa Fusa bananier	VANPL_171031000020	WRKY	WRKY transcription factor 71-like	B	0	VANPL_17103:16046..18793	
RGa LRR Profiler	VANPL_186071000010	RLK	putative receptor-like protein kinase At5g18500	B	0	VANPL_B_00000:1426632401..1426664215	
RGa LRR Profiler	VANPL_198171000010	RLK	inactive receptor-like serine/threonine-protein kinase At2g40270 [Phalaenopsis equestris]	B	0	VANPL_B_00000:1450664463..1450674810	
RGa LRR Profiler	VANPL_200641000010	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase RGI4	B	0	VANPL_B_00000:1455107552..1455110924	
RGa LRR Profiler	VANPL_201341000010	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase RGI4	B	0	VANPL_B_00000:1456263080..1456266452	
RGa LRR Profiler	VANPL_210201000010	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase RGI4	B	0	VANPL_B_00000:1470133086..1470136458	

RGA Fusa bananier	VANPL_A_00001t000010	NLR	conserved hypothetical protein	A	1	VANPL_A_00001:7706..19291	
RGA LRR Profiler	VANPL_A_00001t002280	RLK	putative LRR receptor-like serine/threonine-protein kinase At2g24230	A	1	VANPL_A_00001:3164937..3168144	
RGA Fusa bananier	VANPL_A_00001t002540	POD	Peroxidase 72	A	1	VANPL_A_00001:3441224..3443414	
RGA LRR Profiler	VANPL_A_00001t003160	RLK	Leucine-rich repeat receptor protein kinase MSP1	A	1	VANPL_A_00001:4199965..4209478	
RGA LRR Profiler	VANPL_A_00001t003210	RLK	Pollen receptor-like kinase 4	A	1	VANPL_A_00001:4257517..4260124	
RGA Fusa bananier	VANPL_A_00001t003960	ASP	aspartyl protease family protein 2-like [Phalaenopsis equestris]	A	1	VANPL_A_00001:5369469..5371661	
RGA LRR Profiler	VANPL_A_00001t004930	RLK	LRR receptor-like serine/threonine-protein kinase RGS5	A	1	VANPL_A_00001:7405868..7410121	
RGA LRR Profiler	VANPL_A_00001t005280	RLK	putative inactive receptor kinase At5g10020	A	1	VANPL_A_00001:8204106..8213217	
RGA Fusa bananier	VANPL_A_00001t005320	Cytochrome p450	Flavonoid 3',5'-hydroxylase	A	1	VANPL_A_00001:8309367..8311338	
RGA Fusa autres espèces	VANPL_A_00001t005470	Cytochrome p450	Wall-associated receptor kinase-like 20	A	1	VANPL_A_00001:8799012..8810281	
RGA Fusa bananier	VANPL_A_00001t005500	WRKY	WRKY transcription factor WRKY24-like isoform X2 [Phalaenopsis equestris]	A	1	VANPL_A_00001:8925128..8927783	
Phénylpropanoïde	VANPL_A_00001t005660	OMT	Tricetin 3',4',5'-O-trimethyltransferase	A	1	VANPL_A_00001:9232452..9233762	
Phénylpropanoïde	VANPL_A_00001t005670	OMT	Tricetin 3',4',5'-O-trimethyltransferase	A	1	VANPL_A_00001:9244799..9246135	
RGA LRR Profiler	VANPL_A_00001t007170	RLK	LRR receptor-like serine/threonine-protein kinase RPK2	A	1	VANPL_A_00001:12299608..12310192	
RP Soiano-DelaCruz	VANPL_A_00001t007630	RPS18A	40S ribosomal protein S18	A	1	VANPL_A_00001:13020556..13024763	
RP Soiano-DelaCruz	VANPL_A_00001t007960	RPS23A	40S ribosomal protein S23	A	1	VANPL_A_00001:13597980..13604746	
RP Soiano-DelaCruz	VANPL_A_00001t008110	RPS5	40S ribosomal protein S5	A	1	VANPL_A_00001:13859471..13863250	
RGA LRR Profiler	VANPL_A_00001t008190	RLK	inactive receptor-like serine/threonine-protein kinase At2g40270 [Phalaenopsis equestris]	A	1	VANPL_A_00001:13990318..13995673	
RGA Fusa bananier	VANPL_A_00001t008680	NLR	Myb-related protein 308	A	1	VANPL_A_00001:14706139..14706839	
RGA Fusa bananier	VANPL_A_00001t008690	NLR	myb-related protein Zn38-like [Phalaenopsis equestris]	A	1	VANPL_A_00001:14709985..14711014	
RGA Fusa bananier	VANPL_A_00001t008960	NLR	myb-related protein P-like [Phalaenopsis equestris]	A	1	VANPL_A_00001:15167931..15169485	
RGA Fusa bananier	VANPL_A_00001t009530	WRKY	AP2-like ethylene-responsive transcription factor AIL7 [Phalaenopsis equestris]	A	1	VANPL_A_00001:16114309..16116896	
RGA LRR Profiler	VANPL_A_00001t010080	RLK	Pollen receptor-like kinase 1	A	1	VANPL_A_00001:17255658..17258256	
RGA LRR Profiler	VANPL_A_00001t010520	RLK	Protein NSP-INTERACTING KINASE 3	A	1	VANPL_A_00001:17979210..17987694	
FORL	VANPL_A_00001t011540	FORL	L-type lectin-domain containing receptor kinase IV.2	A	1	VANPL_A_00001:20084986..20087250	
RGA LRR Profiler	VANPL_A_00001t011630	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g63430	A	1	VANPL_A_00001:20226036..20255595	
RGA Fusa bananier	VANPL_A_00001t011780	NLR	protein ODORANT1-like [Phalaenopsis equestris]	A	1	VANPL_A_00001:20449406..20450500	
FORL	VANPL_A_00001t012370	FORL	Serine/threonine-protein kinase BSK1	A	1	VANPL_A_00001:21502531..21521231	
RGA Fusa bananier	VANPL_A_00001t012410	NLR	Myb-related protein 308	A	1	VANPL_A_00001:21611296..21612594	
RGA LRR Profiler	VANPL_A_00001t012430	RLK	Leucine-rich repeat receptor-like protein kinase PEPRI1	A	1	VANPL_A_00001:21626422..21630272	
FORL	VANPL_A_00001t012890	FORL	L-type lectin-domain containing receptor kinase IV.1	A	1	VANPL_A_00001:24838038..24840426	
RGA LRR Profiler	VANPL_A_00001t013360	RLK	Protein kinase domain-containing protein	A	1	VANPL_A_00001:25627458..25631832	
RP Soiano-DelaCruz	VANPL_A_00001t014170	RPL10A	60S ribosomal protein L10	A	1	VANPL_A_00001:26647246..26654028	
RGA Fusa bananier	VANPL_A_00001t015460	WRKY	AP2-like ethylene-responsive transcription factor AIL7 [Phalaenopsis equestris]	A	1	VANPL_A_00001:28455993..28459371	
RGA Fusa bananier	VANPL_A_00001t016260	NLR	Myb-related protein 308	A	1	VANPL_A_00001:298883024..29884673	
RGA LRR Profiler / RGA Daphnia / RGA autres espèces / RGA	VANPL_A_00001t017640	NLR	Disease resistance protein RGA3	A	1	VANPL_A_00001:32377332..32380790	

RGa Fusa bananier	VANPL_A_000011018730	NLR	Transcription factor MYB92	A	1	VANPL_A_00001:34243922..34256867	
RP Solano-DelaCruz	VANPL_A_000011019630	RPL9B	60S ribosomal protein L9 [Phalaenopsis equestris]	A	1	VANPL_A_00001:35708981..35714163	
RGa LRR Profiler	VANPL_A_000011020470	RLK	Protein kinase domain-containing protein	A	1	VANPL_A_00001:40432049..40437303	
RP Solano-DelaCruz	VANPL_A_000011021150	RPL27A	putative 60S ribosomal protein L27-3	A	1	VANPL_A_00001:41315842..41330465	
FORL	VANPL_A_000011022510	GLIP	L-type lectin-domain containing receptor kinase VIII.2	A	1	VANPL_A_00001:43112236..43114636	
RGa Fusa bananier	VANPL_A_000011023660	POD	GDSL esterase/lipase A12g23540	B	1	VANPL_A_00001:45034803..45036683	
RGa Fusa bananier	VANPL_A_000011024680	POD	Peroxidase 4	A	1	VANPL_A_00001:47703357..47706886	
RGa LRR Profiler	VANPL_A_000011024730	RLK	Protein kinase domain-containing protein	A	1	VANPL_A_00001:47805255..47809387	
RGa LRR Profiler	VANPL_A_000011024770	RLK	LRR receptor-like serine/threonine-protein kinase SIK1	A	1	VANPL_A_00001:47845779..47851874	
RGa Fusa bananier	VANPL_A_000011024940	NLR	protein ODORANT1-like [Phalaenopsis equestris]	A	1	VANPL_A_00001:48470864..48472325	
Phénylpropanoïde	VANPL_A_000011025020	4CL	4-coumarate-CoA ligase-like 6	A	1	VANPL_A_00001:48627491..48633431	
RGa Fusa bananier	VANPL_A_000011025080	CHIT	Acidic endochitinase	A	1	VANPL_A_00001:48761634..48762839	
RGa Fusa bananier	VANPL_A_000011025100	CHIT	Acidic endochitinase	A	1	VANPL_A_00001:48774929..48776006	
RGa Fusa bananier	VANPL_A_000011025110	CHIT	Acidic endochitinase	A	1	VANPL_A_00001:48792603..48793632	
FORL	VANPL_A_000011025590	FORL	L-type lectin-domain containing receptor kinase SIT1	A	1	VANPL_A_00001:50155786..50157819	
RGa LRR Profiler	VANPL_A_000011025690	RLK	serine/threonine-protein kinase RIPK-like [Phalaenopsis equestris]	A	1	VANPL_A_00001:50446868..50471233	
RGa LRR Profiler	VANPL_A_000011025720	RLK	LRR receptor-like serine/threonine-protein kinase FEI 1	A	1	VANPL_A_00001:50558943..50619762	
RGa LRR Profiler	VANPL_A_000011025850	RIP	LRR receptor-like serine/threonine-protein kinase GSO2 [Phalaenopsis equestris]	A	1	VANPL_A_00001:50968147..50971666	
RGa Fusa bananier	VANPL_A_000011026270	GLIP	GDSL esterase/lipase A14g16230-like [Phalaenopsis equestris]	A	1	VANPL_A_00001:51845973..51848011	
FORL	VANPL_A_000011026350	FORL	Serine/threonine-protein kinase BSK1	A	1	VANPL_A_00001:52052296..52091274	
RGa Fusa bananier	VANPL_A_000011027540	POD	Peroxidase 4	A	1	VANPL_A_00001:54733141..54734879	
Phénylpropanoïde	VANPL_A_000011028090	CSE	Caffeoylshikimate esterase	A	1	VANPL_A_00001:56658535..56672134	
RGa LRR Profiler	VANPL_A_000011028280	RLK	putative LRR receptor-like serine/threonine-protein kinase HSL2	A	1	VANPL_A_00001:57006609..57008786	
RGa Fusa bananier	VANPL_A_000011030830	POD	Peroxidase 4	A	1	VANPL_A_00001:62748304..62749593	
RGa Fusa bananier	VANPL_A_000011030840	POD	Peroxidase 4	A	1	VANPL_A_00001:62751002..62752423	
RGa LRR Profiler	VANPL_A_000011030850	RLK	Protein NSP-INTERACTING KINASE 2	A	1	VANPL_A_00001:62756441..62762555	
RGa LRR Profiler	VANPL_A_000011031160	RLK	LRR receptor kinase SERK2	A	1	VANPL_A_00001:64246719..64267990	
RGa Fusa bananier	VANPL_A_000011031760	GLIP	Cytochrome P450 703A2	A	1	VANPL_A_00001:65016707..65018411	
RGa Fusa bananier	VANPL_A_000011032740	NLR	Transcription factor MYB93	A	1	VANPL_A_00001:66238197..66239477	
RGa LRR Profiler	VANPL_A_000011033150	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 7	A	1	VANPL_A_00001:67009852..67033659	
RGa LRR Profiler	VANPL_A_000011033390	RLK	LRR receptor-like serine/threonine-protein kinase ERECTA	A	1	VANPL_A_00001:673631357..67369943	
RGa Fusa bananier	VANPL_A_000011034120	NLR	Myb-related protein 308	A	1	VANPL_A_00001:69051966..69053643	
RGa Fusa bananier	VANPL_A_000011034480	GLIP	GDSL esterase/lipase A14g71250	A	1	VANPL_A_00001:69831004..69832650	
RGa Fusa bananier	VANPL_A_000011034890	WRKY	AP2-1	A	1	VANPL_A_00001:71051706..71054254	
RGa Fusa bananier	VANPL_A_000021000350	CHIT	Endochitinase EP3	A	2	VANPL_A_00002:3126393..3127680	
RGa Fusa bananier	VANPL_A_000021000360	CHIT	Endochitinase EP3	A	2	VANPL_A_00002:3304742..3305742	
RGa Fusa bananier	VANPL_A_000021000370	CHIT	Endochitinase EP3	A	2	VANPL_A_00002:3332653..3334004	
RGa Fusa bananier	VANPL_A_000021000390	CHIT	Endochitinase EP3	A	2	VANPL_A_00002:3362770..3364040	
RGa Fusa bananier	VANPL_A_000021000520	WRKY	putative WRKY transcription factor 4	A	2	VANPL_A_00002:4553811..4556060	
RGa Fusa bananier	VANPL_A_000021001010	GLIP	GDSL esterase/lipase A12g04570	A	2	VANPL_A_00002:7581621..7583959	

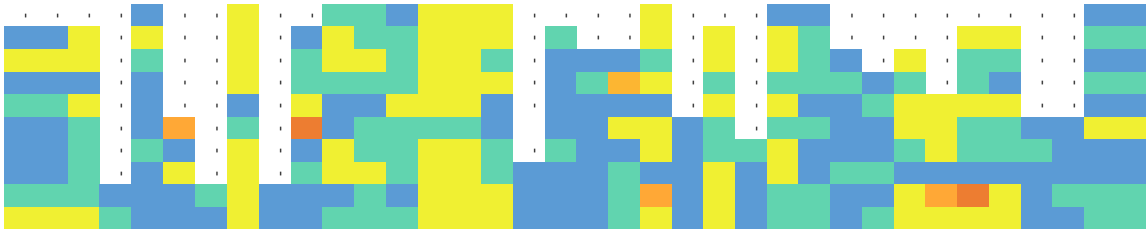
RGa Daphnia (Hasing) / RGa Daphnia (Hasing) / RGa LRR Profiler	LRR	VANPL_A_000021004460	conserved hypothetical protein	A	2	VANPL_A_00002:17670144...17691295	
	LRR	VANPL_A_000021004470	Disease resistance protein RPS5	A	2	VANPL_A_00002:17704670...17718867	
RGa LRR Profiler	RLK	VANPL_A_000021004510	Protein kinase domain-containing protein	A	2	VANPL_A_00002:17870955...17874676	
RGa Fusa autres espèces		VANPL_A_000021004640	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	A	2	VANPL_A_00002:18112801...18115672	
RGa Fusa autres espèces		VANPL_A_000021004660	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	A	2	VANPL_A_00002:18155877...18158520	
RGa LRR Profiler	RLK	VANPL_A_000021005670	Protein NSP-INTERACTING KINASE 1	A	2	VANPL_A_00002:20200890...20213812	
RGa LRR Profiler	RLK	VANPL_A_000021005810	Leucine-rich repeat receptor-like protein kinase TDR	A	2	VANPL_A_00002:20561718...20565356	
RGa Fusa bananier	NLR	VANPL_A_000021005820	MYB-like transcription factor ODO1	A	2	VANPL_A_00002:20585856...20587216	
Phénylpropanoïde	CcoAOMT	VANPL_A_000021005960	Norbelladine O-methyltransferase	A	2	VANPL_A_00002:20968023...21003777	
Phénylpropanoïde	HCT	VANPL_A_000021006310	omega-hydroxyacetyltransferase O-fenulovyl transférase [Phalaenopsis equestris]	A	2	VANPL_A_00002:21827410...21829271	
RGa Fusa bananier	NLR	VANPL_A_000021006390	transcription factor MYB39-like isoform X1 [Phalaenopsis equestris]	A	2	VANPL_A_00002:21953855...21958350	
RGa LRR Profiler	RLK	VANPL_A_000021008280	receptor protein kinase TMK1-like [Phalaenopsis equestris]	A	2	VANPL_A_00002:25705834...25709699	
RP Solano-DelaCruz	RPS6B	VANPL_A_000021008750	40S ribosomal protein S6	A	2	VANPL_A_00002:26352354...26358110	
RGa LRR Profiler	RLK	VANPL_A_000021009380	Non-specific serine/threonine protein kinase	A	2	VANPL_A_00002:27628050...27643900	
RGa LRR Profiler	RLK	VANPL_A_000021009390	Non-specific serine/threonine protein kinase	A	2	VANPL_A_00002:27650906...27676772	
RGa LRR Profiler	RLP	VANPL_A_000021009410	putative LRR receptor-like serine/threonine-protein kinase At1g56130	A	2	VANPL_A_00002:27690316...27712899	
Phénylpropanoïde	CAD	VANPL_A_000021009540	Alcohol dehydrogenase superfamily, zinc-type	A	2	VANPL_A_00002:27883904...27893860	
RGa Fusa bananier	WRKY	VANPL_A_000021010220	putative AP2-like ethylene-responsive transcription factor ANT [Phalaenopsis equestris]	A	2	VANPL_A_00002:29565686...29601385	
RGa Fusa bananier	NLR	VANPL_A_000021011050	Protein ODORANT1	A	2	VANPL_A_00002:30549127...30550375	
RGa Fusa bananier	ASP	VANPL_A_000021011300	outactive protein ASPARTIC PROTEASE IN GIJARD CELL 1	A	2	VANPL_A_00002:30903241...30928908	
RGa LRR Profiler	RLK	VANPL_A_000021011870	Receptor-like kinase TMK4	A	2	VANPL_A_00002:31678858...31684083	
Phénylpropanoïde	CSE	VANPL_A_000021012770	caffeoylshikimate esterase [Phalaenopsis equestris]	A	2	VANPL_A_00002:33794585...33817042	
RGa Fusa bananier	NLR	VANPL_A_000031000030	conserved hypothetical protein	A	3	VANPL_A_00003:108518...129365	
FORI	RLK	VANPL_A_000031000180	L-type lectin domain containing receptor kinase S.4	A	3	VANPL_A_00003:379547...382097	
RGa LRR Profiler	RLK	VANPL_A_000031000920	Leucine-rich repeat receptor protein kinase HPCA1	A	3	VANPL_A_00003:1298895...1303887	
RGa LRR Profiler	RLK	VANPL_A_000031001650	putative inactive leucine-rich repeat receptor-like protein kinase IMK2	A	3	VANPL_A_00003:2477319...2480152	
RGa Fusa bananier	POD	VANPL_A_000031001680	Cationic peroxidase 1	A	3	VANPL_A_00003:2518363...2520108	
RGa Fusa bananier	POD	VANPL_A_000031001690	Peroxidase 4	A	3	VANPL_A_00003:2525087...2526634	
RGa Fusa bananier	GLIP	VANPL_A_000031002740	GDSL esterase/lipase At2g23540	A	3	VANPL_A_00003:4674966...4676522	
RP Solano-DelaCruz	RPL17A	VANPL_A_000031003170	60S ribosomal protein L17	A	3	VANPL_A_00003:5753793...5762559	
RGa Fusa bananier	NLR	VANPL_A_000031005140	MYB-related protein 308	A	3	VANPL_A_00003:17759551...17760854	
RGa LRR Profiler	RLK	VANPL_A_000031005680	Phytosulfokine receptor 1	A	3	VANPL_A_00003:21608214...21613429	
RGa Fusa bananier	NLR	VANPL_A_000031005820	MYB-like transcription factor 4	A	3	VANPL_A_00003:22665211...22666720	
FORI	NLR	VANPL_A_000031006170	Glutathione S-transferase U1R	A	3	VANPL_A_00003:24841094...24842195	
RGa LRR Profiler	RLK	VANPL_A_000031006220	putative LRR receptor-like serine/threonine-protein kinase At1g12460	A	3	VANPL_A_00003:25018860...25022385	
RGa LRR Profiler	RLK	VANPL_A_000031006260	Protein kinase domain-containing protein	A	3	VANPL_A_00003:25420023...25424032	
Phénylpropanoïde	OMT	VANPL_A_000031006290	Tricetin 3',4',5'-O-trimethyltransferase	A	3	VANPL_A_00003:25562281...25566430	
RGa Fusa bananier	Cytochrome p450	VANPL_A_000031006690	Cytochrome P450 84A1	A	3	VANPL_A_00003:28484800...28487541	



RGa Fusa bananier	VANPL_A_00004t006150	Cytochrome p450	Flavonoid 3'-monooxygenase CYP75B137	A	4	VANPL_A_00004:2275887..22758400	
RGa Fusa bananier	VANPL_A_00004t006500	GLIP	GDSL esterase/lipase A15g33370	A	4	VANPL_A_00004:23312506..23314174	
RGa LRR Profiler	VANPL_A_00004t006800	RLK	Phytosulfokinase receptor 1	A	4	VANPL_A_00004:23956542..23960880	
RGa LRR Profiler	VANPL_A_00004t007000	RLK	Non-specific serine/threonine protein kinase	A	4	VANPL_A_00004:24182001..24221651	
RGa LRR Profiler	VANPL_A_00004t007840	RLK	Inactive LRR receptor-like serine/threonine-protein kinase BIR2	A	4	VANPL_A_00004:25619219..25621984	
Phénylpropanoïde	VANPL_A_00004t007980	HCT	omega-hydroxypalmitate O-fenuloyl transférase-like [Phalaenopsis equestris]	A	4	VANPL_A_00004:25940450..25942409	
RGa Fusa bananier	VANPL_A_00004t008000	GLIP	GDSL esterase/lipase A14g71691	A	4	VANPL_A_00004:25975626..25979106	
RGa LRR Profiler	VANPL_A_00004t009200	RLP	receptor-like protein kinase At3g21340 isoform X1 [Phalaenopsis equestris]	A	4	VANPL_A_00004:28423529..28430895	
RGa Fusa bananier	VANPL_A_00004t009450	PPO	Polyphenol oxidase, chloroplasmic	A	4	VANPL_A_00004:28816055..28818938	
RGa LRR Profiler	VANPL_A_00004t010080	RLK	LRR receptor-like serine/threonine-protein kinase FLS2	A	4	VANPL_A_00004:29796778..29801521	
RGa LRR Profiler	VANPL_A_00004t011430	RLK	putative inactive receptor kinase RLK902	A	4	VANPL_A_00004:34680560..34683423	
RGa LRR Profiler	VANPL_A_00004t011480	RLK	putative inactive receptor kinase At1g48480	A	4	VANPL_A_00004:34739769..34743863	
RGa Fusa bananier	VANPL_A_00004t011560	NLR	protein ODORANT1-like [Phalaenopsis equestris]	A	4	VANPL_A_00004:34866803..34867894	
RGa Fusa bananier	VANPL_A_00005t000440	GLIP	GDSL esterase/lipase A15g55050-like [Phalaenopsis equestris]	A	5	VANPL_A_00005:556577..558226	
RGa LRR Profiler	VANPL_A_00005t000620	RLK	LRR receptor-like serine/threonine-protein kinase GSO1	A	5	VANPL_A_00005:801003..805446	
RGa LRR Profiler	VANPL_A_00005t001670	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 6	A	5	VANPL_A_00005:2447663..2457765	
RGa Fusa bananier	VANPL_A_00005t001700	WRKY	AP2-like ethylene-responsive transcription factor At1g16060	A	5	VANPL_A_00005:2525628..2528232	
RGa Fusa bananier	VANPL_A_00005t001730	ASP	aspartyl protease family protein A15g10770 isoform X1 [Phalaenopsis equestris]	A	5	VANPL_A_00005:2566979..2569571	
Phénylpropanoïde	VANPL_A_00005t002420	4CL	putative 4-coumarate-CoA ligase 2	A	5	VANPL_A_00005:3572724..3573509	
Phénylpropanoïde	VANPL_A_00005t002430	4CL	4-coumarate-CoA ligase 1	A	5	VANPL_A_00005:3573782..3580946	
RGa LRR Profiler	VANPL_A_00005t002630	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 8	A	5	VANPL_A_00005:3863161..3877648	
RGa Fusa bananier	VANPL_A_00005t002970	NLR	Transcription factor MYB61	A	5	VANPL_A_00005:4323642..4345291	
RGa Fusa bananier	VANPL_A_00005t003660	WRKY	AP2-like ethylene-responsive transcription factor ANT	A	5	VANPL_A_00005:5715824..5718775	
Phénylpropanoïde	VANPL_A_00005t003910	CCR	Cinnamoyl-CoA reductase 1	A	5	VANPL_A_00005:6113858..6115617	
Phénylpropanoïde	VANPL_A_00005t003950	CCR	putative Cinnamoyl-CoA reductase 1	A	5	VANPL_A_00005:6140146..6140555	
Phénylpropanoïde	VANPL_A_00005t003990	HCT	Hydroxycinnamoyltransferase	A	5	VANPL_A_00005:6214915..6220030	
RGa Fusa bananier	VANPL_A_00005t004030	WRKY	AP2-like ethylene-responsive transcription factor AIL5	A	5	VANPL_A_00005:6306992..6309748	
RGa LRR Profiler	VANPL_A_00005t004120	RLK	Receptor-like protein kinase	A	5	VANPL_A_00005:6461955..6466597	
Phénylpropanoïde	VANPL_A_00005t004140	RLK	Receptor-like protein kinase	A	5	VANPL_A_00005:6480561..6485292	
Phénylpropanoïde	VANPL_A_00005t004580	4CL	4-coumarate-CoA ligase-like 9	A	5	VANPL_A_00005:7224150..7233683	
RGa LRR Profiler	VANPL_A_00005t005200	RLK	Protein kinase domain-containing protein	A	5	VANPL_A_00005:8485841..8488929	
RGa LRR Profiler	VANPL_A_00005t005910	RLK	Tyrosine-sulfated glycopeptide receptor 1	A	5	VANPL_A_00005:10179059..10183177	
RGa Fusa bananier	VANPL_A_00005t005950	NLR	Transcription factor MYB106	A	5	VANPL_A_00005:10291404..10292948	
RGa LRR Profiler	VANPL_A_00005t006550	RLK	putative Pollen receptor-like kinase 3	A	5	VANPL_A_00005:11655174..11657193	
RGa LRR Profiler	VANPL_A_00005t007730	RLK	LRR receptor kinase BAK1	A	5	VANPL_A_00005:15395808..15448691	
Phénylpropanoïde	VANPL_A_00005t008220	C3H	p-coumarate 3-hydroxylase	A	5	VANPL_A_00005:17315446..17319092	
RGa Fusa bananier	VANPL_A_00005t008620	Cytochrome p450	Flavonoid 3'-monooxygenase CYP75B137	A	5	VANPL_A_00005:18606898..18608991	
Phénylpropanoïde	VANPL_A_00005t008840	4-HBS ?	Vanillin synthase	A	5	VANPL_A_00005:19507210..19510118	
RGa LRR Profiler	VANPL_A_00005t009740	RLK	Pollen receptor-like kinase 3	A	5	VANPL_A_00005:22860142..22862286	
RGa LRR Profiler	VANPL_A_00005t010850	RLP	Receptor-like protein 51	A	5	VANPL_A_00005:25135045..25136628	

Phénylpropanoïde	VANPL_A_000051011870	CAD	cinnamyl alcohol dehydrogenase 2 [Phalaenopsis equestris]	A	5	VANPL_A_00005:28450252..28453070	
Phénylpropanoïde	VANPL_A_000051011880	CAD	cinnamyl alcohol dehydrogenase 2 [Phalaenopsis equestris]	A	5	VANPL_A_00005:28476520..28479822	
RGA LRR Profiler	VANPL_A_000051011910	RLP	putative leucine-rich repeat extensin-like protein 6 [Phalaenopsis equestris]	A	5	VANPL_A_00005:28733192..28735057	
RGA LRR Profiler	VANPL_A_000051011970	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 5	A	5	VANPL_A_00005:29205859..29226017	
RGA Fusa bananier	VANPL_A_000051012390	WRKY	AP2-like ethylene-responsive transcription factor BBW2, partial [Phalaenopsis equestris]	A	5	VANPL_A_00005:30932116..30935426	
RGA LRR Profiler	VANPL_A_000051013340	RLP	putative Receptor-like protein 4	A	5	VANPL_A_00005:34163849..34164948	
RGA LRR Profiler	VANPL_A_000051014140	RLK	Receptor-like kinase TMK3	A	5	VANPL_A_00005:364441632..36445818	
RGA Fusa bananier	VANPL_A_000051014540	POD	Peroxidase 4	A	5	VANPL_A_00005:37274595..37275947	
RGA LRR Profiler	VANPL_A_000051016300	RLK	Protein kinase domain-containing protein	A	5	VANPL_A_00005:40369156..40374637	
RGA LRR Profiler	VANPL_A_000051016330	RLK	Receptor-like protein kinase 7	A	5	VANPL_A_00005:40409208..40415727	
RGA LRR Profiler	VANPL_A_000051016780	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 3	A	5	VANPL_A_00005:41158853..41168941	
RGA Fusa bananier	VANPL_A_000051016970	NLR	Transcription factor MYB102	A	5	VANPL_A_00005:41406695..41408247	
RGA LRR Profiler	VANPL_A_000051017920	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g06840	A	5	VANPL_A_00005:42730966..42764780	
Phénylpropanoïde	VANPL_A_000051018250	CAD	Mannitol dehydrogenase	A	5	VANPL_A_00005:43195163..43196830	
Phénylpropanoïde	VANPL_A_000051018340	CSE	caffeoylshikimate esterase-like [Phalaenopsis equestris]	A	5	VANPL_A_00005:433331269..43333236	
RGA Fusa autres espèces	VANPL_A_000051018510		wall-associated receptor kinase 3-like [Phalaenopsis equestris]	A	5	VANPL_A_00005:43564384..43566685	
FORL	VANPL_A_000061000160		Receptor-like serine/threonine-protein kinase	A	6	VANPL_A_00006:3115116..316009	
RGA Fusa bananier	VANPL_A_000061000480	NLR	Myb-related protein 308	A	6	VANPL_A_00006:861238..868953	
RGA Fusa bananier	VANPL_A_000061002110	NLR	Myb-related protein 308	A	6	VANPL_A_00006:3279025..3286586	
RGA Fusa bananier	VANPL_A_000061002310	GLIP	GDSL esterase/lipase At2g40250	A	6	VANPL_A_00006:3629739..3631571	
RP Solano-DelaCruz	VANPL_A_000061002420	RPL9B	60S ribosomal protein L9-like [Phalaenopsis equestris]	A	6	VANPL_A_00006:3798324..3807151	
RGA Fusa bananier	VANPL_A_000061003030	NLR	Transcription factor MYB106	A	6	VANPL_A_00006:4954750..4956151	
RGA Fusa bananier	VANPL_A_000061003090	GLIP	GDSL esterase/lipase At1g74460	A	6	VANPL_A_00006:5085850..5087892	
RGA LRR Profiler	VANPL_A_000061003790	RLK	Receptor protein kinase TMK1	A	6	VANPL_A_00006:6679695..6694471	
Phénylpropanoïde	VANPL_A_000061003920	CAD	putative mannitol dehydrogenase	A	6	VANPL_A_00006:7045312..7058770	
Phénylpropanoïde	VANPL_A_000061005770	CCR	Phenylacetaldehyde reductase	A	6	VANPL_A_00006:22027097..22037348	
Phénylpropanoïde	VANPL_A_000061005780	CCR	Phenylacetaldehyde reductase	A	6	VANPL_A_00006:22053317..22065775	
Phénylpropanoïde	VANPL_A_000061005790	CCR	Phenylacetaldehyde reductase	A	6	VANPL_A_00006:22083982..22101226	
Phénylpropanoïde	VANPL_A_000061005810	CCR	Phenylacetaldehyde reductase	A	6	VANPL_A_00006:22153580..22165453	
Phénylpropanoïde	VANPL_A_000061006660	4CL	4-coumarate--CoA ligase-like 9	A	6	VANPL_A_00006:27443969..27475670	
Phénylpropanoïde	VANPL_A_000061006940	HCT	omega-hydroxypalmitate O-feruloyl transferase [Phalaenopsis equestris]	A	6	VANPL_A_00006:28955953..28957379	
RGA Fusa bananier	VANPL_A_000061008360	PR1	Pathogenesis-related protein 1	A	6	VANPL_A_00006:33352259..33353122	
RGA LRR Profiler	VANPL_A_000061008780	RLK	conserved hypothetical protein	A	6	VANPL_A_00006:34093226..34109509	
RGA Fusa bananier	VANPL_A_000061009180	POD	Cationic peroxidase 1	A	6	VANPL_A_00006:35046680..35048104	
RGA LRR Profiler	VANPL_A_000061009910	RLP	putative Receptor-like protein 53	A	6	VANPL_A_00006:36540223..36541108	
RGA LRR Profiler	VANPL_A_000061009920	RLP	putative Receptor-like protein 46	A	6	VANPL_A_00006:36542182..36543014	
RGA LRR Profiler	VANPL_A_000061009970	RLP	putative Receptor-like protein 46	A	6	VANPL_A_00006:36595954..36597318	
RGA LRR Profiler	VANPL_A_000061009990	RLP	putative Leucine-rich repeat-containing protein	A	6	VANPL_A_00006:36608523..36614587	
RGA LRR Profiler	VANPL_A_000061010420	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase At1g17230	A	6	VANPL_A_00006:37525502..37529530	
RP Solano-DelaCruz	VANPL_A_000061011290	RPL10A	60S ribosomal protein L10	A	6	VANPL_A_00006:38970561..38971561	

RGa Fusa bananier	VANPL_A_00006:01.1310	Cytochrome p450	Premnapirodienne oxygénase	A 6	VANPL_A_00006:39004362..39006285	
RGa Fusa bananier	VANPL_A_00006:01.1350	Cytochrome p450	Premnapirodienne oxygénase	A 6	VANPL_A_00006:39034866..39036799	
RGa LRR Profiler	VANPL_A_00006:01.1770	RLP	Receptor-like protein 4	A 6	VANPL_A_00006:39737964..39766781	
RGa Daphna (Hasing)	VANPL_A_00006:01.2170	LRR	glucan endo-1,3-beta-D-glucosidase-like [Phalaenopsis equestris]	A 6	VANPL_A_00006:40299529..40306310	
RGa LRR Profiler	VANPL_A_00006:01.2650	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g06840	A 6	VANPL_A_00006:40954589..41010113	
Phénylpropanoïde	VANPL_A_00006:01.2960	CAD	Alcohol dehydrogénase 9	A 6	VANPL_A_00006:41491560..41493275	
RGa Fusa autres espèces	VANPL_A_00006:01.3170		Wall-associated receptor Kinase 5	A 6	VANPL_A_00006:41693064	
FORL	VANPL_A_00006:01.3250		L-type lectin-domain containing receptor kinase SIT2	A 6	VANPL_A_00006:41799711..41802601	
RGa LRR Profiler	VANPL_A_00007:000570	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g12460	A 7	VANPL_A_00007:4866609..4869943	
RGa Fusa bananier	VANPL_A_00007:000740	NLR	Transcription repressor MYB5	A 7	VANPL_A_00007:5435847..5437741	
RGa Fusa bananier	VANPL_A_00007:001210	GLIP	GDSL esterase/lipase At2g23540	A 7	VANPL_A_00007:7450088..7451980	
RGa Fusa bananier	VANPL_A_00007:001390	NLR	Myb-related protein Zn38	A 7	VANPL_A_00007:8105770..8106881	
RGa Fusa bananier	VANPL_A_00007:001490	GLIP	putative GDSL esterase/lipase At5g37690	A 7	VANPL_A_00007:8433820..8442813	
Phénylpropanoïde	VANPL_A_00007:001510	4CL	putative 4-coumarate--CoA ligase-like 4	A 7	VANPL_A_00007:8490804..8491861	
Phénylpropanoïde	VANPL_A_00007:001520	4CL	4-coumarate--CoA ligase-like 4	A 7	VANPL_A_00007:8539582..8542614	
RGa LRR Profiler	VANPL_A_00007:002170	RLK	LRR receptor-like serine/threonine-protein kinase GSO2	A 7	VANPL_A_00007:10327481..10332330	
FORL	VANPL_A_00007:003780		Serine/threonine-protein kinase BSK7	A 7	VANPL_A_00007:13782458..13810247	
RGa Fusa bananier	VANPL_A_00007:004070	NLR	transcription factor MYB399-like isoform X2 [Phalaenopsis equestris]	A 7	VANPL_A_00007:14372411..14377273	
Phénylpropanoïde	VANPL_A_00007:004860	4CL	4-coumarate--CoA ligase-like 5	A 7	VANPL_A_00007:16089412..16091275	
FORL	VANPL_A_00007:005320		Glutathione S-transferase U18	A 7	VANPL_A_00007:16880638..16881742	
RGa LRR Profiler	VANPL_A_00007:008730	RLP	DNA damage-repair/tolerance protein DRT100	A 7	VANPL_A_00007:26683417..26685659	
RGa Fusa bananier	VANPL_A_00007:008950	GLIP	GDSL esterase/lipase 7	A 7	VANPL_A_00007:26974676..26977136	
RGa LRR Profiler	VANPL_A_00007:009000	RLK	Protein NSP-INTERACTING KINASE 1	A 7	VANPL_A_00007:27036468..27040779	
FORL	VANPL_A_00007:009100		Serine/threonine-protein kinase BSK5	A 7	VANPL_A_00007:27140252..27144500	
Phénylpropanoïde	VANPL_A_00007:009390	CSE	caffeoylshikimate esterase-like [Phalaenopsis equestris]	A 7	VANPL_A_00007:27561924..27566135	
Phénylpropanoïde	VANPL_A_00007:009540	CCR	Cinnamoyl-CoA reductase 1	A 7	VANPL_A_00007:27776823..27783158	
RGa LRR Profiler	VANPL_A_00007:01.2040	RLK	Leudne-rich repeat receptor-like serine/threonine-protein kinase	A 7	VANPL_A_00007:33378372..33382048	
RGa LRR Profiler	VANPL_A_00007:01.2050	RLK	Leudne-rich repeat receptor-like serine/threonine-protein kinase	A 7	VANPL_A_00007:33430181..33433872	
Phénylpropanoïde	VANPL_A_00007:01.3150	HCT	putative Shikimate O-hydroxycinnamoyltransferase	A 7	VANPL_A_00007:35797080..35798457	
Phénylpropanoïde	VANPL_A_00007:01.3180	HCT	putative Shikimate O-hydroxycinnamoyltransferase	A 7	VANPL_A_00007:35856158..35857535	
Phénylpropanoïde	VANPL_A_00007:01.3230	HCT	putative Shikimate O-hydroxycinnamoyltransferase	A 7	VANPL_A_00007:35921552..35922862	
Phénylpropanoïde	VANPL_A_00007:01.3290	HCT	putative Shikimate O-hydroxycinnamoyltransferase	A 7	VANPL_A_00007:36002506..36003918	
RGa LRR Profiler	VANPL_A_00008:000250	RLK	Receptor protein-tyrosine kinase CEPR2	A 8	VANPL_A_00008:667724..674043	
RGa Fusa bananier	VANPL_A_00008:000780	WRKY	putative AP2-like ethylene-responsive transcription factor All1	A 8	VANPL_A_00008:2578899..2588823	
Phénylpropanoïde / RGa Fusa bananier	VANPL_A_00008:000880	C3H	Trimethyltridecatetraene synthase	A 8	VANPL_A_00008:2898217..2900095	
Phénylpropanoïde / RGa Fusa bananier	VANPL_A_00008:000890	C3H	Trimethyltridecatetraene synthase	A 8	VANPL_A_00008:2913532..2915293	





Phénylpropanoïde	VANPL_A_00008:002330	CAD	Alcohol dehydrogenase 9	A	8	VANPL_A_00008:6401904..6403488	
RGA LRR Profiler	VANPL_A_00008:002780	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 7	A	8	VANPL_A_00008:7110429..7131427	
RGA LRR Profiler	VANPL_A_00008:003760	RLP	putative inactive receptor kinase At5g10020	A	8	VANPL_A_00008:8312611..8324114	
RGA Fusa bananier	VANPL_A_00008:003960	PPO	Polyphenol oxidase, chloroplastic	A	8	VANPL_A_00008:11674396..11676108	
RGA Fusa bananier	VANPL_A_00008:004000	PPO	Polyphenol oxidase, chloroplastic	A	8	VANPL_A_00008:11888026..11890414	
RGA Fusa bananier	VANPL_A_00008:004020	PPO	Polyphenol oxidase, chloroplastic	A	8	VANPL_A_00008:11909126..11910859	
RGA Fusa bananier	VANPL_A_00008:005500	NLR	Transcription factor MYB16	A	8	VANPL_A_00008:18016932..18018167	
RGA LRR Profiler	VANPL_A_00008:005570	RLK	Receptor protein kinase TMK1	A	8	VANPL_A_00008:18401514..18406483	
FORL	VANPL_A_00008:005880	RLK	Glutathione transferase GST 23	A	8	VANPL_A_00008:19556647..19557716	
RGA LRR Profiler	VANPL_A_00008:006220	RLK	LRR receptor-kinase BAK1	A	8	VANPL_A_00008:20357600..20378518	
RGA LRR Profiler	VANPL_A_00008:007620	RLK	Protein kinase domain-containing protein	A	8	VANPL_A_00008:22907568..22909508	
RGA Fusa bananier	VANPL_A_00008:008740	WRKY	AP2-like ethylene-responsive transcription factor At4g16060	A	8	VANPL_A_00008:25176093..25178605	
Phénylpropanoïde	VANPL_A_00008:008770	HCT	Hydroxycinnamoyltransferase	A	8	VANPL_A_00008:25273303..25275949	
RGA Fusa bananier	VANPL_A_00008:009380	NLR	myb-related protein P-like [Phalaenopsis equestris]	A	8	VANPL_A_00008:26491511..26493388	
RGA LRR Profiler	VANPL_A_00008:009810	RLK	Protein kinase domain-containing protein	A	8	VANPL_A_00008:27294028..27298505	
RGA LRR Profiler	VANPL_A_00008:010390	RLK	putative inactive receptor kinase At5g67200	A	8	VANPL_A_00008:28116620..28136650	
Phénylpropanoïde	VANPL_A_00008:010420	CcoAOMT	Caffeoyl-CoA O-methyltransferase 3	A	8	VANPL_A_00008:28170298..28171733	
RGA LRR Profiler	VANPL_A_00008:010930	RLK	Protein NSP-INTERACTING KINASE 1	A	8	VANPL_A_00008:28777447..28780864	
FORL	VANPL_A_00008:011780	RLK	L-type lectin-domain containing receptor kinase IX.1	A	8	VANPL_A_00008:30222877..30225174	
RGA Fusa bananier	VANPL_A_00008:011920	GLIP	GDSL esterase/lipase At5g5050-like [Phalaenopsis equestris]	A	8	VANPL_A_00008:30495443..30497703	
RGA LRR Profiler	VANPL_A_00009:000190	RLK	putative inactive leucine-rich repeat receptor-like protein kinase At3g03770	A	9	VANPL_A_00009:1108630..1113462	
RGA LRR Profiler	VANPL_A_00009:001210	RLK	LRR receptor-like serine/threonine-protein kinase FEI 1	A	9	VANPL_A_00009:2324305..2333128	
RGA LRR Profiler	VANPL_A_00009:001500	RLK	receptor protein kinase-like protein ZAR1 [Phalaenopsis equestris]	A	9	VANPL_A_00009:2842567..2844740	
RP Solano-DelaCruz	VANPL_A_00009:001930	RPL24B	60S ribosomal protein L24-like [Phalaenopsis equestris]	A	9	VANPL_A_00009:3456223..3474403	
RGA LRR Profiler	VANPL_A_00009:002030	RLK	putative inactive receptor kinase At5g58300	A	9	VANPL_A_00009:3650842..3658474	
RGA Fusa bananier	VANPL_A_00009:002190	WRKY	WRKY transcription factor WRKY24-like isoform X2 [Phalaenopsis equestris]	A	9	VANPL_A_00009:3780580..3782654	
RGA LRR Profiler	VANPL_A_00009:003280	RLK	Leucine-rich repeat receptor protein kinase HPCA1	A	9	VANPL_A_00009:5332030..5338787	
RGA LRR Profiler	VANPL_A_00009:003500	RLK	inactive receptor-like serine/threonine-protein kinase At2g40270 [Phalaenopsis equestris]	A	9	VANPL_A_00009:5663976..5669556	
RGA Fusa bananier	VANPL_A_00009:004250	Cytochrome p450	Flavonoid 3'-monooxygenase CYP75B137	A	9	VANPL_A_00009:6923856..6925718	
RP Solano-DelaCruz	VANPL_A_00009:004790	RP52C	40S ribosomal protein S2-4	A	9	VANPL_A_00009:8104913..8109052	
RGA Fusa bananier	VANPL_A_00009:006750	ASP	Aspartic proteinase nepenthesin-2	A	9	VANPL_A_00009:13513884..13515518	
RGA LRR Profiler	VANPL_A_00009:007250	RLK	Nodulation receptor kinase	A	9	VANPL_A_00009:15753718..15773807	
Phénylpropanoïde	VANPL_A_00009:007620	HCT	putrescine hydroxycinnamoyltransferase 1-like	A	9	VANPL_A_00009:16713113..16714306	
Phénylpropanoïde	VANPL_A_00009:007630	HCT	Putrescine hydroxycinnamoyltransferase 1	A	9	VANPL_A_00009:16720214..16722030	
RGA LRR Profiler	VANPL_A_00009:008870	RLK	LRR receptor-like serine/threonine-protein kinase GHR1	A	9	VANPL_A_00009:23888431..23893464	
RGA LRR Profiler	VANPL_A_00009:008970	RLK	Receptor-like protein kinase 7	A	9	VANPL_A_00009:23998873..24002356	
Phénylpropanoïde	VANPL_A_00009:010090	CCR	Tetrateptide alpha-pyrone reductase 1	A	9	VANPL_A_00009:25508839..25510211	
Phénylpropanoïde	VANPL_A_00009:010150	CCR	Cinnamoyl-CoA reductase 1	A	9	VANPL_A_00009:25604937..25611358	
RGA LRR Profiler	VANPL_A_00009:010340	RLP	putative protein MALE DISCOVERER 2	A	9	VANPL_A_00009:25963777..25964312	
RGA LRR Profiler	VANPL_A_00009:010350	RLK	Receptor kinase-like protein Xa21	A	9	VANPL_A_00009:25994238..25998126	

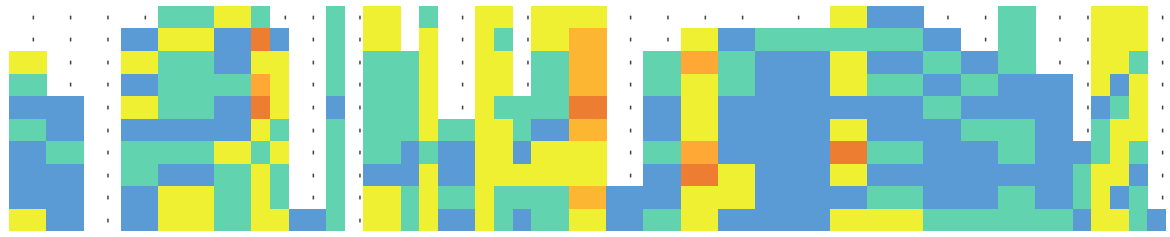
RGa LRR Profiler	VANPL_A_000091010360	RLK	Inactive leucine-rich repeat receptor-like protein kinase	A	9	VANPL_A_00009:26027050..26030693	
RGa Fusa bananier	VANPL_A_000091010440	ASP	Aspartyl protease family protein AT5g10770	A	9	VANPL_A_00009:26258934..26260671	
RGa Fusa bananier	VANPL_A_000091010500	NLR	transcription factor MYB80 [Phalaenopsis equestris]	A	9	VANPL_A_00009:26387833..26388837	
RGa LRR Profiler	VANPL_A_000091010590	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase RGI4	A	9	VANPL_A_00009:26480799..26496763	
RGa Fusa bananier	VANPL_A_000091010700	WRKY	ethylene-responsive transcription factor WR1-like [Phalaenopsis equestris]	A	9	VANPL_A_00009:26713126..26714692	
RP Solano-DelaCruz	VANPL_A_000091010800	RPL24B	60S ribosomal protein L24	A	9	VANPL_A_00009:26994076..27000969	
RGa LRR Profiler	VANPL_A_000091010840	RLK	Leucine-rich repeat receptor protein kinase HPCA1	A	9	VANPL_A_00009:27061059..27065948	
Phénylpropanoïde	VANPL_A_000091011260	CSE	caffeoylshikimate esterase-like [Phalaenopsis equestris]	A	9	VANPL_A_00009:28252790..28254743	
RGa LRR Profiler	VANPL_A_000091011520	RLK	protein MALE DISCOVERER 2-like [Phalaenopsis equestris]	A	9	VANPL_A_00009:28808770..28815660	
RGa LRR Profiler	VANPL_A_000091011760	RLK	LRR receptor-like serine/threonine-protein kinase ER2	A	9	VANPL_A_00009:29330087..29337290	
RGa LRR Profiler	VANPL_A_000091012150	RLK	Protein kinase domain-containing protein	A	9	VANPL_A_00009:30453143..30493221	
FORL	VANPL_A_000091012640		Thaumatin-like protein	A	9	VANPL_A_00009:31674852..31675858	
FORL	VANPL_A_000091012750		Serine/threonine-protein kinase ATG1b	A	9	VANPL_A_00009:31805417..31835180	
Phénylpropanoïde	VANPL_A_000091013040	HCT	Tryptamine hydroxycinnamoyltransferase 1	A	9	VANPL_A_00009:32694393..32695744	
Phénylpropanoïde	VANPL_A_000091013050	HCT	Tryptamine hydroxycinnamoyltransferase 2	A	9	VANPL_A_00009:32695747..32707048	
RGa LRR Profiler	VANPL_A_000091013180	RLK	LRR receptor kinase SERK2	A	9	VANPL_A_00009:32966339..32990358	
RGa LRR Profiler	VANPL_A_000101000080	RLK	Receptor protein-tyrosine kinase CEPR2	A	10	VANPL_A_00010:239007..258741	
RGa LRR Profiler	VANPL_A_000101000100	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1	A	10	VANPL_A_00010:387785..392213	
Phénylpropanoïde	VANPL_A_000101001320	CSE	caffeoylshikimate esterase-like [Phalaenopsis equestris]	A	10	VANPL_A_00010:3936083..3938017	
RGa LRR Profiler	VANPL_A_000101003190	RLP	putative Leucine-rich repeat receptor-like serine/threonine-protein kinase RGI4	A	10	VANPL_A_00010:7988465..7990976	
RGa LRR Profiler	VANPL_A_000101003420	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g06840	A	10	VANPL_A_00010:8425430..8482858	
RGa Fusa bananier	VANPL_A_000101003490	POD	Cationic peroxidase 1	A	10	VANPL_A_00010:8664419..8665912	
RGa Fusa bananier	VANPL_A_000101003590	WRKY	WRKY transcription factor WRKY24-like isoform X1 [Phalaenopsis equestris]	A	10	VANPL_A_00010:8834505..8836300	
RP Solano-DelaCruz	VANPL_A_000101003880	RPS23A	putative 40S ribosomal protein S23	A	10	VANPL_A_00010:9431448..9450995	
FORL	VANPL_A_000101004340		Thaumatin-like protein	A	10	VANPL_A_00010:10212225..10213048	
FORL	VANPL_A_000101004350		L-type lectin-domain containing receptor kinase S.4	A	10	VANPL_A_00010:10219224..10221855	
RGa Fusa bananier	VANPL_A_000101004880	GLIP	GDSL esterase/lipase 7	A	10	VANPL_A_00010:10928913..10936344	
RGa LRR Profiler	VANPL_A_000101006190	RLK	calmodulin-binding receptor kinase CaMRLK	A	10	VANPL_A_00010:13931962..13934889	
RGa LRR Profiler	VANPL_A_000101007380	RLK	Receptor-like protein kinase HSL1	A	10	VANPL_A_00010:15397653..15401248	
RGa LRR Profiler	VANPL_A_000101007640	RLK	Receptor protein kinase-like protein ZAR1	A	10	VANPL_A_00010:15797174..15800953	
RP Solano-DelaCruz	VANPL_A_000101008780	RPS12A	40S ribosomal protein S12	A	10	VANPL_A_00010:17933607..17942377	
RP Solano-DelaCruz	VANPL_A_000101009490	RPS12A	40S ribosomal protein S12-like [Phalaenopsis equestris]	A	10	VANPL_A_00010:19486096..19492713	
RGa LRR Profiler	VANPL_A_0001011910	RIP	Receptor-like protein 44	A	10	VANPL_A_00010:28616004..28617889	
RGa Fusa bananier	VANPL_A_000101012010	POD	Lignin-forming anionic peroxidase	A	10	VANPL_A_00010:29131066..29132264	
RGa Fusa bananier	VANPL_A_000101012030	POD	Lignin-forming anionic peroxidase	A	10	VANPL_A_00010:29254505..29255655	
RGa Fusa bananier	VANPL_A_000101012040	POD	Lignin-forming anionic peroxidase	A	10	VANPL_A_00010:29306034..29307152	
Phénylpropanoïde	VANPL_A_000111001320	OMT	Caffeic acid 3-O-methyltransferase 1	1	11	VANPL_A_00011:2186421..2188082	
Phénylpropanoïde	VANPL_A_000111001990	4CL	Oxalate--CoA ligase	A	11	VANPL_A_00011:3153920..3155898	
Phénylpropanoïde	VANPL_A_000111003310	CAD	Cinnamyl alcohol dehydrogenase 3	A	11	VANPL_A_00011:5938222..5940429	
RGa Fusa bananier	VANPL_A_000111003880	NLR	transcription factor WER-like [Phalaenopsis equestris]	A	11	VANPL_A_00011:7379849..7380982	

RGa LRR Profiler	VANPL_A_00011:1004450	RLK	Leucine-rich repeat receptor protein kinase HPCA1	A	11	VANPL_A_00011:13207780..13212353
Phénylpropanoïde	VANPL_A_00011:1006410	HCT	omega-hydroxypalmitate O-feruloyl transférase	A	11	VANPL_A_00011:22650646..22652263
RGa LRR Profiler	VANPL_A_00011:1006600	RLK	putative inactive receptor kinase RLK902	A	11	VANPL_A_00011:23743458..23745826
RGa LRR Profiler	VANPL_A_00011:1007410	RLK	Protein kinase domain-containing protein	A	11	VANPL_A_00011:25996626..26021212
RGa LRR Profiler	VANPL_A_00011:1007510	RLP	Receptor-like protein CLAVATA2	A	11	VANPL_A_00011:26332104..26335084
Phénylpropanoïde	VANPL_A_00011:1007870	CAD	Alcohol dehydrogenase 9	A	11	VANPL_A_00011:26975115..26976570
RP Solano-DelaCruz	VANPL_A_00011:1008150	RPS5	40S ribosomal protein S5	A	11	VANPL_A_00011:27540979..27548704
RGa LRR Profiler	VANPL_A_00011:1009370	RLK	Receptor-like protein kinase HSL1	A	11	VANPL_A_00011:41650228..41657064
Phénylpropanoïde	VANPL_A_00011:1009530	CAD	Alcohol dehydrogenase 9	A	11	VANPL_A_00011:42070873..42072352
Phénylpropanoïde	VANPL_A_00011:1009550	CAD	Alcohol dehydrogenase 9	A	11	VANPL_A_00011:42085044..42086742
RGa LRR Profiler	VANPL_A_00011:1009710	RLK	Protein kinase domain-containing protein	A	11	VANPL_A_00011:42526020..42574588
RGa Fusa autres espèces	VANPL_A_00011:1009890		G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	A	11	VANPL_A_00011:42873353..42876241
RGa Fusa bananier	VANPL_A_00011:1011720	WRKY	putative WRKY transcription factor 4	A	11	VANPL_A_00011:45956734..45976183
RGa LRR Profiler	VANPL_A_00011:1011790	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 5	A	11	VANPL_A_00011:46045256..46067213
RGa LRR Profiler	VANPL_A_00011:1012060	RLP	Leucine-rich repeat receptor-like serine/threonine-protein kinase	A	11	VANPL_A_00011:46525371..46545299
RGa LRR Profiler	VANPL_A_00012:1000280	RLK	Kinase-like protein TMKL1	A	12	VANPL_A_00012:719831..722756
RGa Fusa bananier	VANPL_A_00012:1000940	NLR	MYB-like transcription factor 4	A	12	VANPL_A_00012:3164857..3165919
RGa LRR Profiler	VANPL_A_00012:1002910	RLK	Inactive leucine-rich repeat receptor-like serine/threonine-protein kinase At1g60630	A	12	VANPL_A_00012:11579663..11582581
RGa LRR Profiler	VANPL_A_00012:1004520	RLK	MDS1-interacting receptor like kinase 1	A	12	VANPL_A_00012:15616407..15619976
RGa LRR Profiler	VANPL_A_00012:1004530	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase At1g17230	A	12	VANPL_A_00012:15634254..15639968
RP Solano-DelaCruz	VANPL_A_00012:1005740	RPL13B	60S ribosomal protein L13-2	A	12	VANPL_A_00012:18229521..18234049
Phénylpropanoïde	VANPL_A_00012:1005860	CCR	cinnamoyl-CoA reductase 2-like [Phalaenopsis equestris]	A	12	VANPL_A_00012:18679106..18684813
RGa LRR Profiler	VANPL_A_00012:1006180	RLK	receptor protein kinase-like protein ZAR1 [Phalaenopsis equestris]	A	12	VANPL_A_00012:19361744..19364271
RP Solano-DelaCruz	VANPL_A_00012:1006730	RPL10A	60S ribosomal protein L10-1	A	12	VANPL_A_00012:20835147..20836565
RGa LRR Profiler	VANPL_A_00012:1006740	RLK	putative inactive receptor kinase At5g58300	A	12	VANPL_A_00012:20887529..20899725
RGa LRR Profiler	VANPL_A_00012:1006860	RLK	Leucine-rich repeat receptor-like protein kinase PXC2	A	12	VANPL_A_00012:21315985..21319869
RGa LRR Profiler	VANPL_A_00012:1007290	RLK	Systemin receptor SR160	A	12	VANPL_A_00012:22556419..22560345
RGa LRR Profiler	VANPL_A_00012:1007630	RLK	Non-specific serine/threonine protein kinase	A	12	VANPL_A_00012:23343030..23367172
RGa LRR Profiler	VANPL_A_00012:1007700	RLK	Non-specific serine/threonine protein kinase	A	12	VANPL_A_00012:23587019..23594870
RGa LRR Profiler	VANPL_A_00012:1007710	RLK	Non-specific serine/threonine protein kinase	A	12	VANPL_A_00012:236632942..23645943
RGa LRR Profiler	VANPL_A_00012:1008570	RLK	Leucine-rich repeat receptor-like protein kinase PXC1	A	12	VANPL_A_00012:25521142..25523296
RGa LRR Profiler	VANPL_A_00012:1009290	RLK	Protein kinase domain-containing protein	A	12	VANPL_A_00012:26704676..26707607
RGa Fusa bananier	VANPL_A_00012:1009320	WRKY	WRKY transcription factor 28-like	A	12	VANPL_A_00012:26756613..26758504
Phénylpropanoïde	VANPL_A_00012:1010040	CCoAOMT	Norbelladine 4'-O-methyltransferase	A	12	VANPL_A_00012:27994997..27996368
RGa LRR Profiler	VANPL_A_00012:1010110	RLK	Protein kinase domain	A	12	VANPL_A_00012:28103845..28107774
RGa Fusa bananier	VANPL_A_00012:1011130	NLR	MYB-related protein 315	A	12	VANPL_A_00012:29746999..29748325
RGa LRR Profiler	VANPL_A_00012:1011170	RLK	putative inactive receptor kinase At1g48480	A	12	VANPL_A_00012:29782165..29785349
RGa LRR Profiler	VANPL_A_00012:1011350	RLK	Mitogen-activated protein kinase kinase YODA	A	12	VANPL_A_00012:30051273..30066627
RGa LRR Profiler	VANPL_A_00012:1011950	RLP	Leucine-rich repeat protein 2	A	12	VANPL_A_00012:31142756..31159855
RGa Fusa bananier	VANPL_A_00012:1012130		Flavonoid 3'-monooxygenase CYP75B137	A	12	VANPL_A_00012:31414801..31417035
			Cytochrome p450			



RGA LRR Profiler	VANPL_B_000011002720	RLK	Leucine-rich repeat receptor protein kinase MSP1	B	1	VANPL_B_00001:3601738..3611734	
RGA LRR Profiler	VANPL_B_000011004600	RLK	LRR receptor-like serine/threonine-protein kinase RGS5	B	1	VANPL_B_00001:6829704..6833449	
RGA LRR Profiler	VANPL_B_000011005110	RLK	putative inactive receptor kinase At5g10020	B	1	VANPL_B_00001:7882311..7891150	
RGA Fusa bananier	VANPL_B_000011005150	Cytochrome p450	Flavonoid 3',5'-hydroxylase	B	1	VANPL_B_00001:7965856..7967922	
RGA Fusa autres espèces	VANPL_B_000011005310	OMT	Wall-associated receptor kinase-like 20	B	1	VANPL_B_00001:8484475..8495236	
Phénylpropanoïde	VANPL_B_000011005530	OMT	Tricetin 3',4',5'-O-trimethyltransferase	B	1	VANPL_B_00001:8910279..8911464	
Phénylpropanoïde	VANPL_B_000011005540	OMT	Tricetin 3',4',5'-O-trimethyltransferase	B	1	VANPL_B_00001:8918763..8920132	
RGA LRR Profiler	VANPL_B_000011006670	RLK	Protein kinase domain-containing protein	B	1	VANPL_B_00001:11283969..11286769	
RP Solano-DelaCruz	VANPL_B_000011007250	RPS18A	40S ribosomal protein S18	B	1	VANPL_B_00001:12053642..12058026	
RP Solano-DelaCruz	VANPL_B_000011007560	RPS23A	putative 40S ribosomal protein S23	B	1	VANPL_B_00001:12671203..12676995	
RP Solano-DelaCruz	VANPL_B_000011007710	RPS5	40S ribosomal protein S5	B	1	VANPL_B_00001:12910715..12913097	
RGA LRR Profiler	VANPL_B_000011007810	RLK	inactive receptor-like serine/threonine-protein kinase At2g40270 [Phalaenopsis equestris]	B	1	VANPL_B_00001:13045007..13050426	
RGA Fusa bananier	VANPL_B_000011008210	NLR	Myb-related protein 308	B	1	VANPL_B_00001:13639731..13640389	
RGA Fusa bananier	VANPL_B_000011008220	NLR	myb-related protein Zn38-like [Phalaenopsis equestris]	B	1	VANPL_B_00001:13643557..13644538	
RGA Fusa bananier	VANPL_B_000011008490	NLR	myb-related protein P-like [Phalaenopsis equestris]	B	1	VANPL_B_00001:14090221..14091866	
RGA Fusa bananier	VANPL_B_000011009020	WRKY	AP2-like ethylene-responsive transcription factor AIL7 [Phalaenopsis equestris]	B	1	VANPL_B_00001:15046497..15049509	
RGA LRR Profiler	VANPL_B_000011009620	RLK	Pollen receptor-like kinase 5	B	1	VANPL_B_00001:16205510..16208080	
RGA LRR Profiler	VANPL_B_000011010020	RLK	putative ATP-dependent DNA helicase MER3	B	1	VANPL_B_00001:16891894..16940779	
FORL	VANPL_B_000011011040	RLK	L-type lectin-domain containing receptor kinase IV.1-like [Phalaenopsis equestris]	B	1	VANPL_B_00001:19229867..19239862	
RGA LRR Profiler	VANPL_B_000011011130	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g63430	B	1	VANPL_B_00001:19377629..19404464	
RGA Fusa bananier	VANPL_B_000011011330	NLR	conserved hypothetical protein	B	1	VANPL_B_00001:19659945..19664376	
FORL	VANPL_B_000011011900	NLR	Serine/threonine-protein kinase BSK1	B	1	VANPL_B_00001:20671714..20689904	
RGA Fusa bananier	VANPL_B_000011011940	NLR	Myb-related protein 308	B	1	VANPL_B_00001:20780074..20781128	
RGA LRR Profiler	VANPL_B_000011011960	RLK	Leucine-rich repeat receptor-like protein kinase PEPRI1	B	1	VANPL_B_00001:20796942..20800557	
RGA LRR Profiler	VANPL_B_000011012830	RLK	Protein kinase domain-containing protein	B	1	VANPL_B_00001:23342687..23347066	
RP Solano-DelaCruz	VANPL_B_000011013690	RPL10A	60S ribosomal protein L10	B	1	VANPL_B_00001:24408852..24410022	
RGA Fusa bananier	VANPL_B_000011014920	WRKY	AP2-like ethylene-responsive transcription factor AIL7 [Phalaenopsis equestris]	B	1	VANPL_B_00001:26221491..26224954	
RGA Fusa bananier	VANPL_B_000011015730	NLR	Myb-related protein 308	B	1	VANPL_B_00001:27658669..27660316	
RGA Fusa autres espèces / RGA Daphnia / RGA LRR Profiler / RGA	VANPL_B_000011017060		Disease resistance protein RGA3	B	1	VANPL_B_00001:30227099..30230631	
RGA LRR Profiler	VANPL_B_000011017260	RLK	pollen receptor-like kinase 3 [Phalaenopsis equestris]	B	1	VANPL_B_00001:30698580..30700797	
RGA Fusa bananier	VANPL_B_000011018190	NLR	Transcription factor MYB92	B	1	VANPL_B_00001:32332929..32356692	
RP Solano-DelaCruz	VANPL_B_000011019120	RPL9B	putative 60S ribosomal protein L9	B	1	VANPL_B_00001:33872550..33877533	
FORL	VANPL_B_000011020240	POD	Serine/threonine-protein kinase BSK2	B	1	VANPL_B_00001:35772863..35789797	
RGA Fusa bananier	VANPL_B_000011020600	POD	Peroxidase 4	B	1	VANPL_B_00001:36357839..36358959	
RGA Fusa bananier	VANPL_B_000011020610	POD	Cationic peroxidase 1	B	1	VANPL_B_00001:36370270..36371506	
RGA Fusa bananier	VANPL_B_000011020620	POD	Peroxidase 4	B	1	VANPL_B_00001:36374454..36375971	
RGA LRR Profiler	VANPL_B_000011021270	RLK	Protein kinase domain-containing protein	B	1	VANPL_B_00001:37614549..37619987	

RP Solano-DelaCruz	RPL27A	VANPL_B_000011021950	60S ribosomal protein L27-3	B 1	VANPL_B_00001:38489610..38493594
FORL	GLIP	VANPL_B_000011022960	putative L-type lectin-domain containing receptor kinase S.7	B 1	VANPL_B_00001:39980641..39993993
RGA Fusa bananier	GLIP	VANPL_B_000011023980	GDSL esterase/lipase A1g223540	B 1	VANPL_B_00001:41945890..41947853
RGA LRR Profiler	RLK	VANPL_B_000011024560	serine/threonine-protein kinase RPK-like [Phalaenopsis equestris]	B 1	VANPL_B_00001:47380988..47399252
RGA LRR Profiler	RLK	VANPL_B_000011024590	LRR receptor-like serine/threonine-protein kinase FEI1	B 1	VANPL_B_00001:47488555..47549407
RGA LRR Profiler	RLP	VANPL_B_000011024710	LRR receptor-like serine/threonine-protein kinase GS02 [Phalaenopsis equestris]	B 1	VANPL_B_00001:47903977..47907447
RGA Fusa bananier	GLIP	VANPL_B_000011025130	GDSL esterase/lipase A1g416230-like [Phalaenopsis equestris]	B 1	VANPL_B_00001:48748204..48750232
FORL	POD	VANPL_B_000011025200	Serine/threonine-protein kinase BSK1	B 1	VANPL_B_00001:48924673..48963480
RGA Fusa bananier	CSE	VANPL_B_000011026360	Peroxidase 4	B 1	VANPL_B_00001:51588568..51590290
Phénylpropanoïde	RLK	VANPL_B_000011026880	Caffeoylshikimate esterase	B 1	VANPL_B_00001:52887626..52901099
RGA LRR Profiler	RLK	VANPL_B_000011027080	putative LRR receptor-like serine/threonine-protein kinase HSL2	B 1	VANPL_B_00001:53229194..53232180
RGA Fusa bananier	POD	VANPL_B_000011028490	Peroxidase 4	B 1	VANPL_B_00001:58868461..58869571
RGA Fusa bananier	POD	VANPL_B_000011028500	Peroxidase 4	B 1	VANPL_B_00001:58871213..58876542
RGA LRR Profiler	RLK	VANPL_B_000011028510	Protein NSP-INTERACTING KINASE 2	B 1	VANPL_B_00001:58876780..58882657
RGA LRR Profiler	RLK	VANPL_B_000011028720	LRR receptor kinase SERK2	B 1	VANPL_B_00001:59400118..59421528
RGA Fusa bananier	Cytochrome p450	VANPL_B_000011029260	Cytochrome P450 703A2	B 1	VANPL_B_00001:60069519..60071144
RGA Fusa bananier	NLR	VANPL_B_000011030310	Transcription factor MYB93	B 1	VANPL_B_00001:61349796..61351088
RGA LRR Profiler	RLK	VANPL_B_000011030760	Protein STRUBBELIG-RECEPTOR FAMILY 7	B 1	VANPL_B_00001:62417136..62441629
RGA LRR Profiler	RLK	VANPL_B_000011030990	LRR receptor-like serine/threonine-protein kinase ERECTA	B 1	VANPL_B_00001:62716564..62725185
RGA Fusa bananier	NLR	VANPL_B_000011031750	MYB-related protein 308	B 1	VANPL_B_00001:64462524..64464192
RGA Fusa bananier	GLIP	VANPL_B_000011032070	GDSL esterase/lipase A1g71250	B 1	VANPL_B_00001:65290595..65293530
Phénylpropanoïde	CCR	VANPL_B_000011032220	cinnamoyl-CoA reductase 2 [Phalaenopsis equestris]	B 1	VANPL_B_00001:65700555..65704577
RGA Fusa bananier	WRKY	VANPL_B_000011032780	conserved hypothetical protein	B 1	VANPL_B_00001:68102503..68104954
RGA Fusa bananier	CHIT	VANPL_B_000021000290	Endochitinase EP3	B 2	VANPL_B_00002:2279530..2280862
RGA Fusa bananier	CHIT	VANPL_B_000021000310	Chitinase 5	B 2	VANPL_B_00002:2439230..2440491
RGA Fusa bananier	WRKY	VANPL_B_000021000450	putative WRKY transcription factor 4	B 2	VANPL_B_00002:3672755..3679781
RGA Fusa bananier	GLIP	VANPL_B_000021001000	GDSL esterase/lipase A1g604570	B 2	VANPL_B_00002:6806581..6808834
RGA LRR Profiler	RLK	VANPL_B_000021001080	Receptor protein kinase-like protein A1g434220	B 2	VANPL_B_00002:7301418..7304669
RGA Daphna (Hasing) / RGA Fusa bananier	LRR	VANPL_B_000021004420	Disease resistance protein RPS5	B 2	VANPL_B_00002:16843918..16846814
RGA LRR Profiler	RLK	VANPL_B_000021004480	Protein kinase domain-containing protein	B 2	VANPL_B_00002:16978152..16981875
RGA Fusa autres espèces	RLK	VANPL_B_000021004600	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	B 2	VANPL_B_00002:17211571..17214347
RGA Fusa autres espèces	RLK	VANPL_B_000021004610	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	B 4	VANPL_B_00002:17254994..17257887
RGA LRR Profiler	RLK	VANPL_B_000021005640	Protein NSP-INTERACTING KINASE 1	B 2	VANPL_B_00002:19304667..19317589
RGA LRR Profiler	RLK	VANPL_B_000021005790	Leucine-rich repeat receptor-like protein kinase TDR	B 2	VANPL_B_00002:19673117..19676503
RGA Fusa bananier	NLR	VANPL_B_000021005800	MYB-like transcription factor ODO1	B 2	VANPL_B_00002:19692532..19693935
Phénylpropanoïde	CCoAOMT	VANPL_B_000021005920	Norbelladine O-methyltransferase	B 2	VANPL_B_00002:20071931..20111672
RGA Fusa bananier	NLR	VANPL_B_000021006430	transcription factor MYB39-like isoform X1 [Phalaenopsis equestris]	B 2	VANPL_B_00002:21246029..21250872
Phénylpropanoïde	HCT	VANPL_B_000021006500	omega-hydroxypalmitate O-feruloyl transferase [Phalaenopsis equestris]	B 2	VANPL_B_00002:21368661..21369287



putative leucine-rich repeat receptor-like protein kinase At1g35710	RLP	VANPL_B_000021006640	B	2	VANPL_B_00002:21650277...21652712
receptor protein kinase TMK1-like [Phalaenopsis equestris]	RLK	VANPL_B_000021008470	B	2	VANPL_B_00002:25104940...25108778
putative 40S ribosomal protein S6	RP56B	VANPL_B_000021008910	B	2	VANPL_B_00002:25748476...25772837
Non-specific serine/threonine protein kinase	RLK	VANPL_B_000021009510	B	2	VANPL_B_00002:26993244...27010382
Non-specific serine/threonine protein kinase	RLK	VANPL_B_000021009520	B	2	VANPL_B_00002:27019131...27047269
Non-specific serine/threonine protein kinase	RLK	VANPL_B_000021009530	B	2	VANPL_B_00002:27062107...27093621
Alcohol dehydrogenase 9	CAD	VANPL_B_000021009630	B	2	VANPL_B_00002:27263624...27265229
Alcohol dehydrogenase 9	CAD	VANPL_B_000021009640	B	2	VANPL_B_00002:27269882...27271301
AP2-like ethylene-responsive transcription factor ANT	WRKY	VANPL_B_000021010290	B	2	VANPL_B_00002:28840233...28863890
Protein ODORANT1	NLR	VANPL_B_000021011200	B	2	VANPL_B_00002:29837645...29838962
putative protein ASPARTIC PROTEASE IN GUJARO CELL 1	ASP	VANPL_B_000021011430	B	2	VANPL_B_00002:30164023...30190916
Receptor-like kinase TMK4	RLK	VANPL_B_000021011970	B	2	VANPL_B_00002:30955076...30960366
caffeoylshikimate esterase [Phalaenopsis equestris]	CSE	VANPL_B_000021012660	B	2	VANPL_B_00002:32315130...32337605
L-type lectin-domain containing receptor kinase S.4	GLIP	VANPL_B_000031000170	B	3	VANPL_B_00003:343928...346523
putative GDSL esterase/lipase At2g23540	RPL17A	VANPL_B_000031000290	B	3	VANPL_B_00003:4611390...4620888
Pectinase M28 domain-containing protein	NLR	VANPL_B_000031000720	B	3	VANPL_B_00003:5707846...5779716
Mvb-related protein 308	RLK	VANPL_B_000031002650	B	3	VANPL_B_00003:16226299...16227599
Phytosulfakinase receptor 1	NLR	VANPL_B_000031003160	B	3	VANPL_B_00003:20146781...20151641
MYB-like transcription factor 4	NLR	VANPL_B_000031003320	B	3	VANPL_B_00003:21213862...21215294
Glutathione S-transferase U17	FORL	VANPL_B_000031003660	B	3	VANPL_B_00003:23204878...23205994
putative LRR receptor-like serine/threonine-protein kinase At1g12460	RLK	VANPL_B_000031003710	B	3	VANPL_B_00003:23377190...23380619
Protein kinase domain-containing protein	OMT	VANPL_B_000031003750	B	3	VANPL_B_00003:23774164...23777926
Tricetin 3',4',5'-O-trimethyltransferase	Cytochrome p450	VANPL_B_000031003770	B	3	VANPL_B_00003:23908204...23910728
Cytochrome P450 84A1	Cytochrome p450	VANPL_B_000031004120	B	3	VANPL_B_00003:26640611...26643365
Disease resistance protein RGA2-like [Phalaenopsis equestris]	ASP	VANPL_B_000031004810	B	3	VANPL_B_00003:29182212...29189718
aspartyl protease family protein At5g10770-like [Phalaenopsis equestris]	RLK	VANPL_B_000031004840	B	3	VANPL_B_00003:29230322...29232931
Leucine-rich repeat receptor-like protein kinase PXC2	RLK	VANPL_B_000031005320	B	3	VANPL_B_00003:30531186...30534907
putative inactive receptor kinase At5g58300	RLK	VANPL_B_000031005410	B	3	VANPL_B_00003:30713853...30722518
Inactive leucine-rich repeat receptor-like protein kinase IMK2	RLK	VANPL_B_000031005640	B	3	VANPL_B_00003:31316913...31320038
inactive receptor-like serine/threonine-protein kinase At2g40270 [Phalaenopsis equestris]	RLK	VANPL_B_000031005690	B	3	VANPL_B_00003:31429547...31432454
Leucine-rich repeat receptor protein kinase HPCA1	RLK	VANPL_B_000031005790	B	3	VANPL_B_00003:31738889...31769733
receptor-like protein kinase HSL1 [Phalaenopsis equestris]	RIP	VANPL_B_000031006680	B	3	VANPL_B_00003:33604201...33607264
putative Leucine-rich repeat protein 2	RLK	VANPL_B_000031007810	B	3	VANPL_B_00003:35557626...35566674
Inactive LRR receptor-like serine/threonine-protein kinase BIR2	4CL	VANPL_B_000031009600	B	3	VANPL_B_00003:38115768...38117706
4-coumarate-CoA lyase-like 7	CCoAOMT	VANPL_B_000031009800	B	3	VANPL_B_00003:38594202...38597410
Norbelladine O-methyltransferase	CCoAOMT	VANPL_B_000031009850	B	3	VANPL_B_00003:38630251...38633403

RGa LRR Profiler	VANPL_B_000031010110	RLK	Receptor-like protein kinase HSL1	B 3	VANPL_B_00003:38947973..38951470	
RGa LRR Profiler	VANPL_B_000031010370	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM3	B 3	VANPL_B_00003:39249608..39253532	
RGa Fusa bananier	VANPL_B_000031010470	WRKY	WRKY transcription factor 71-like	B 3	VANPL_B_00003:39344287..39346037	
RGa LRR Profiler	VANPL_B_000031010910	RLK	LRR receptor-like serine/threonine-protein kinase RPK2 [Phalaenopsis equestris]	B 3	VANPL_B_00003:39863326..39866697	
RGa LRR Profiler	VANPL_B_000041001110	RLK	Receptor protein kinase CLAVATA1	B 4	VANPL_B_00004:8183677..8187595	
RGa Fusa bananier	VANPL_B_000041001930	GLIP	GDSL esterase/lipase LIT1	B 4	VANPL_B_00004:12150617..12156094	
RGa LRR Profiler	VANPL_B_000041001980	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase At1g17230	B 4	VANPL_B_00004:12380632..12386314	
RGa Fusa autres espèces / Daphna / RGa LRR Profiler	VANPL_B_000041002430		conserved hypothetical protein	B 4	VANPL_B_00004:13934237..13937667	
RGa Fusa autres espèces / Daphna / RGa LRR Profiler / RGa Fusa bananier	VANPL_B_000041002440		Disease resistance protein RGA3	B 4	VANPL_B_00004:13964465..13968170	
RGa Fusa autres espèces / Daphna / RGa LRR Profiler	VANPL_B_000041002450		Disease resistance protein RGA3	B 4	VANPL_B_00004:13994942..13998211	
RGa Fusa autres espèces / Daphna / RGa LRR Profiler / RGa Fusa bananier	VANPL_B_000041002460		Disease resistance protein RGA3	B 4	VANPL_B_00004:14025113..14028592	
RGa Fusa bananier	VANPL_B_000041003160	NLR	Transcription factor MYB15	B 4	VANPL_B_00004:16129866..16131124	
RGa LRR Profiler	VANPL_B_000041003980	RLK	Brassinosteroid LRR receptor kinase BRI1	B 4	VANPL_B_00004:17932121..17936467	
RGa LRR Profiler	VANPL_B_000041004020	RLK	Receptor kinase-like protein Xa21	B 4	VANPL_B_00004:18058463..18061778	
RGa LRR Profiler	VANPL_B_000041004060	RLK	receptor-like protein kinase HSL1 [Phalaenopsis equestris]	B 4	VANPL_B_00004:18239172..18242753	
RGa LRR Profiler	VANPL_B_000041004100	RLP	putative Receptor-like protein kinase HSL1	B 4	VANPL_B_00004:18344357..18346992	
RGa Fusa autres espèces	VANPL_B_000041004270		G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	B 4	VANPL_B_00004:18752608..18755245	
RGa Fusa autres espèces / FORL	VANPL_B_000041004290		G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	B 4	VANPL_B_00004:18790235..18793050	
RGa LRR Profiler	VANPL_B_000041005400	RLK	Protein NSP-INTERACTING KINASE 1	B 4	VANPL_B_00004:21223279..21227594	
RGa Fusa bananier	VANPL_B_000041005580	NLR	protein ODORANT1-like [Phalaenopsis equestris]	B 4	VANPL_B_00004:21574253..21575330	
RGa Fusa bananier	VANPL_B_000041006350	Cytochrome p450	Flavonoid 3'-monooxygenase CYP75B137	B 4	VANPL_B_00004:23147593..23149950	
RGa Fusa bananier	VANPL_B_000041006650	GLIP	GDSL esterase/lipase At5g33370	B 4	VANPL_B_00004:23700274..23701900	
RGa LRR Profiler	VANPL_B_000041006950	RLK	Phytosulfokine receptor 1	B 4	VANPL_B_00004:24351047..24355110	
RGa LRR Profiler	VANPL_B_000041007150	RLK	Non-specific serine/threonine protein kinase	B 4	VANPL_B_00004:24577240..24616676	
RGa LRR Profiler	VANPL_B_000041008050	RLK	Inactive LRR receptor-like serine/threonine-protein kinase BIR2	B 4	VANPL_B_00004:26040631..26043412	
Phénylpropanoide	VANPL_B_000041008180	HCT	omega-hydroxypalmitate O-feruloyl transferase-like [Phalaenopsis equestris]	B 4	VANPL_B_00004:26351259..26356719	
RGa Fusa bananier	VANPL_B_000041008200	GLIP	GDSL esterase/lipase At1g71691	B 4	VANPL_B_00004:26383835..26387630	
RGa Fusa bananier	VANPL_B_000041008720	PPO	Polyphenol oxidase, chloroplastic	B 4	VANPL_B_00004:28418907..28421273	



RGA LRR Profiler	VANPL_B_00004t008950	RLP	receptor-like protein kinase At3g21340 isoform X1 [Phalaenopsis equestris]	B	4	VANPL_B_00004:28840609..28846495	
RGA LRR Profiler	VANPL_B_00004t009460	RLK	LRR receptor-like serine/threonine-protein kinase FL52	B	4	VANPL_B_00004:29684501..29689227	
RGA LRR Profiler	VANPL_B_00004t010970	RLK	putative inactive receptor kinase RLK902	B	4	VANPL_B_00004:33786160..33788955	
RGA LRR Profiler	VANPL_B_00004t011060	RLK	putative inactive receptor kinase At1g48480	B	4	VANPL_B_00004:33849048..33853047	
RGA Fusa bananier	VANPL_B_00004t011180	NLR	protein ODORANT1 [Phalaenopsis equestris]	B	4	VANPL_B_00004:33986310..33987425	
RGA LRR Profiler	VANPL_B_00004t011560	RLK	Receptor protein kinase-like protein ZAR1	B	4	VANPL_B_00004:34650200..34655476	
Phénylpropanoïde	VANPL_B_00004t012560	HCT	Omega-hydroxypalmitate O-feruloyl transferase	B	4	VANPL_B_00004:37102499..37112972	
Phénylpropanoïde	VANPL_B_00004t012570	HCT	Omega-hydroxypalmitate O-feruloyl transferase	B	4	VANPL_B_00004:37116974..37121349	
RGA LRR Profiler	VANPL_B_00004t012810	RLK	putative MID51-interacting receptor like kinase 2	B	4	VANPL_B_00004:37786886..37789096	
RGA LRR Profiler	VANPL_B_00004t013320	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g67720	B	4	VANPL_B_00004:39287421..39306872	
RGA Fusa bananier	VANPL_B_00004t013430	Cytochrome p450	Cytochrome P450 71AP13	B	4	VANPL_B_00004:39529592..39531451	
RGA Fusa bananier	VANPL_B_00004t013440	Cytochrome p450	Cytochrome P450 71AP13	B	4	VANPL_B_00004:39548889..39550754	
RGA Fusa bananier	VANPL_B_00004t013450	Cytochrome p450	Cytochrome P450 71AP13	B	4	VANPL_B_00004:39559437..39561397	
RGA LRR Profiler	VANPL_B_00004t013530	RLK	Protein kinase domain-containing protein	B	4	VANPL_B_00004:39718930..39750689	
RGA Fusa bananier	VANPL_B_00005t000390	GLIP	GDSL esterase/lipase At5g55050-like [Phalaenopsis equestris]	B	5	VANPL_B_00005:600082..601576	
RGA LRR Profiler	VANPL_B_00005t000900	RLK	LRR receptor-like serine/threonine-protein kinase GSO1	B	5	VANPL_B_00005:1211690..1216073	
RGA LRR Profiler	VANPL_B_00005t001320	RLK	Leucine-rich repeat receptor-like tyrosine-protein kinase PXC3	B	5	VANPL_B_00005:1914377..1917017	
RGA LRR Profiler	VANPL_B_00005t002680	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 6	B	5	VANPL_B_00005:3797086..3807201	
RGA Fusa bananier	VANPL_B_00005t002720	WRKY	AP2-like ethylene-responsive transcription factor At1g16060	B	5	VANPL_B_00005:3865241..3867446	
RGA Fusa bananier	VANPL_B_00005t002740	ASP	aspartyl protease family protein At5g10770 isoform X1 [Phalaenopsis equestris]	B	5	VANPL_B_00005:3906755..3908978	
Phénylpropanoïde	VANPL_B_00005t003450	4CL	4-coumarate-CoA ligase 1	B	5	VANPL_B_00005:4975243..4982494	
RGA LRR Profiler	VANPL_B_00005t003570	RLK	putative protein STRUBBELIG-RECEPTOR FAMILY 8	B	5	VANPL_B_00005:5122350..5149617	
RGA Fusa bananier	VANPL_B_00005t004060	NLR	Transcription factor MYB61	B	5	VANPL_B_00005:5864877..5886662	
RGA Fusa bananier	VANPL_B_00005t004780	WRKY	AP2-like ethylene-responsive transcription factor ANT	B	5	VANPL_B_00005:7422488..7425615	
Phénylpropanoïde	VANPL_B_00005t005030	CCR	Cinnamoyl-CoA reductase 1	B	5	VANPL_B_00005:7825291..7839128	
Phénylpropanoïde	VANPL_B_00005t005050	CCR	Cinnamoyl-CoA reductase 1	B	5	VANPL_B_00005:7859545..7861197	
Phénylpropanoïde	VANPL_B_00005t005090	HCT	Hydroxycinnamoyltransferase	B	5	VANPL_B_00005:7922594..7927826	
RGA Fusa bananier	VANPL_B_00005t005140	WRKY	AP2-like ethylene-responsive transcription factor AIL5	B	5	VANPL_B_00005:8013980..8016330	
RGA LRR Profiler	VANPL_B_00005t005230	RLK	Receptor-like protein kinase	B	5	VANPL_B_00005:8107324..8114869	
RGA Fusa autres espèces / RGA Fusa bananier	VANPL_B_00005t005250	4CL	Receptor-like protein kinase	B	5	VANPL_B_00005:8119277..8124027	
Phénylpropanoïde	VANPL_B_00005t005660	4CL	4-coumarate-CoA ligase-like 9	B	5	VANPL_B_00005:8859302..8861318	
RGA LRR Profiler	VANPL_B_00005t006140	RLK	Pollen receptor-like kinase 3	B	5	VANPL_B_00005:13239504..13241765	
RGA LRR Profiler	VANPL_B_00005t007280	RLK	LRR receptor kinase BAK1	B	5	VANPL_B_00005:17061018..17116097	
Phénylpropanoïde / RGA Fusa Fusa bananier	VANPL_B_00005t007760	C3H	p-coumarate 3-hydroxylase	B	5	VANPL_B_00005:19079107..19082991	
RGA LRR Profiler	VANPL_B_00005t008030	RLK	Pollen receptor-like kinase 3	B	5	VANPL_B_00005:20195909..20198998	
RGA LRR Profiler	VANPL_B_00005t009240	RLP	Receptor-like protein 51	B	5	VANPL_B_00005:22632752..22634062	
Phénylpropanoïde	VANPL_B_00005t010080	CAD	cinnamyl alcohol dehydrogenase 2 [Phalaenopsis equestris]	B	5	VANPL_B_00005:25704980..25708094	
RGA Fusa bananier	VANPL_B_00005t010800	WRKY	AP2-like ethylene-responsive transcription factor BBM2, partial [Phalaenopsis equestris]	B	5	VANPL_B_00005:28589664..28592836	
RGA LRR Profiler	VANPL_B_00005t011560	RLK	Pollen receptor-like kinase 1	B	5	VANPL_B_00005:31345279..31347452	



RGa Fusa bananier	Cytochrome p450	VANPL_B_000061007180	Vaniladiene synthase	B	6	VANPL_B_00006:2092344..20924173	
RGa Fusa bananier	Cytochrome p450	VANPL_B_000061007190	Premnaspirodiene oxygenase	B	6	VANPL_B_00006:20942871..20944522	
RGa Fusa bananier	Cytochrome p450	VANPL_B_000061007200	Premnaspirodiene oxygenase	B	6	VANPL_B_00006:20972811..20974710	
FORL		VANPL_B_000061007210	G-type lectin 5-receptor-like serine/threonine-protein kinase At2g19130	B	6	VANPL_B_00006:20975464..20988263	
RGa LRR Profiler	RLP	VANPL_B_000061007670	Receptor-like protein 4	B	6	VANPL_B_00006:21604253..21641822	
FORL		VANPL_B_000061007780	L-type/lectin-domain containing receptor kinase IX.1	B	6	VANPL_B_00006:21803082..21805424	
RGa Fusa autres espèces / RGA Daphna / RGA LRR Profiler		VANPL_B_000061008110	Disease resistance protein RGA3	B	6	VANPL_B_00006:22174480..22178492	
RGa LRR Profiler	RLK	VANPL_B_000061008570	putative LRR receptor-like serine/threonine-protein kinase At1g06840	B	6	VANPL_B_00006:22827589..22881730	
RGa LRR Profiler	RLP	VANPL_B_000061008860	Leucine-rich repeat protein 1	B	6	VANPL_B_00006:23306136..23322480	
Phénylpropanoïde	CAD	VANPL_B_000061008880	Alcohol dehydrogenase 9	B	6	VANPL_B_00006:233332065..23333553	
RGa Fusa autres espèces		VANPL_B_000061009050	Wall-associated receptor kinase 2	B	6	VANPL_B_00006:23535893..23538734	
FORL		VANPL_B_000061009130	L-type lectin-domain containing receptor kinase SIT2	B	6	VANPL_B_00006:23624227..23627013	
RGa LRR Profiler	RLK	VANPL_B_000071000440	putative LRR receptor-like serine/threonine-protein kinase At1g12460	B	7	VANPL_B_00007:3181994..3185177	
RGa Fusa bananier	NLR	VANPL_B_000071000600	Transcription repressor MYB5	B	7	VANPL_B_00007:3749827..3750873	
RGa Fusa bananier	GLIP	VANPL_B_000071001050	GDSL esterase/lipase At2g23540	B	7	VANPL_B_00007:5565401..5567218	
RGa Fusa bananier	NLR	VANPL_B_000071001250	Myb-related protein Zm38	B	7	VANPL_B_00007:6207477..6208185	
RGa Fusa bananier	GLIP	VANPL_B_000071001370	GDSL esterase/lipase At4g16230-like [Phalaenopsis equestris]	B	7	VANPL_B_00007:6509891..6511842	
Phénylpropanoïde	4CL	VANPL_B_000071001390	4-coumarate--CoA ligase-like 4	B	7	VANPL_B_00007:6564097..6566868	
Phénylpropanoïde	4CL	VANPL_B_000071001400	4-coumarate--CoA ligase-like 4	B	7	VANPL_B_00007:6605060..6607089	
RGa LRR Profiler	RLK	VANPL_B_000071002050	LRR receptor-like serine/threonine-protein kinase GSO2	B	7	VANPL_B_00007:8306630..8312291	
RGa Fusa bananier	NLR	VANPL_B_000071002410	myb-related protein Zm1-like [Phalaenopsis equestris]	B	7	VANPL_B_00007:9379863..9381085	
FORL		VANPL_B_000071003560	Serine/threonine-protein kinase BSK7	B	7	VANPL_B_00007:11593825..11622005	
Phénylpropanoïde	4CL	VANPL_B_000071004640	4-coumarate--CoA ligase-like 5	B	7	VANPL_B_00007:14004273..14006368	
FORL		VANPL_B_000071005030	Glutathione S-transferase U18	B	7	VANPL_B_00007:14776022..14777117	
Phénylpropanoïde	CAD	VANPL_B_000071005070	PKS_ER domain-containing protein	B	7	VANPL_B_00007:14854017..14868792	
Phénylpropanoïde	OMT	VANPL_B_000071006130	Caffeic acid O-methyltransferase	B	7	VANPL_B_00007:18468106..18469948	
Phénylpropanoïde	OMT	VANPL_B_000071006140	Tricetin 3',4',5'-O-trimethyltransferase	B	7	VANPL_B_00007:18483376..18493548	
RGa LRR Profiler	RLP	VANPL_B_000071007890	DNA damage-repair/tolerance protein DRT100	B	7	VANPL_B_00007:22258770..22261130	
RGa Fusa bananier	GLIP	VANPL_B_000071008110	GDSL esterase/lipase 7	B	7	VANPL_B_00007:22570446..22572768	
RGa LRR Profiler	RLK	VANPL_B_000071008160	Protein NSP-INTERACTING KINASE 1	B	7	VANPL_B_00007:22630097..22634428	
FORL		VANPL_B_000071008260	putative Serine/threonine-protein kinase BSK5	B	7	VANPL_B_00007:22741677..22746933	
Phénylpropanoïde	CSE	VANPL_B_000071008530	caffeoylshikimate esterase-like [Phalaenopsis equestris]	B	7	VANPL_B_00007:23187764..23191989	
Phénylpropanoïde	CCR	VANPL_B_000071008670	Cinnamoyl-CoA reductase 1	B	7	VANPL_B_00007:23417671..23424287	
RGa LRR Profiler	RLK	VANPL_B_000071011190	Leucine-rich repeat receptor-like serine/threonine-protein kinase	B	7	VANPL_B_00007:29251522..29255095	
RGa LRR Profiler	RLK	VANPL_B_000071011210	Leucine-rich repeat receptor-like serine/threonine-protein kinase	B	7	VANPL_B_00007:29298327..29301772	
Phénylpropanoïde	HCT	VANPL_B_000071012390	putative Shikimate O-hydroxycinnamoyltransferase	B	7	VANPL_B_00007:31767388..31769271	
Phénylpropanoïde	HCT	VANPL_B_000071012430	putative Shikimate O-hydroxycinnamoyltransferase	B	7	VANPL_B_00007:31792630..31794294	

Phénylpropanoïde	VANPL_B_00007012440	HCT	putative Shikimate O-hydroxycinnamoyltransferase	B 7	VANPL_B_00007:31824735..31826438	
RGA LRR Profiler	VANPL_B_00007012500	RLP	conserved hypothetical protein	B 7	VANPL_B_00007:31905341..31907323	
Phénylpropanoïde	VANPL_B_00007012520	HCT	putative Shikimate O-hydroxycinnamoyltransferase	B 7	VANPL_B_00007:31923448..31924013	
RGA LRR Profiler	VANPL_B_00008000240	RLK	Receptor protein-tyrosine kinase CEP2A2	B 8	VANPL_B_00008:656841..662746	
RGA Fusa bananier	VANPL_B_00008000780	WRKY	putative AP2-like ethylene-responsive transcription factor ALL1	B 8	VANPL_B_00008:2629590..2639394	
Phénylpropanoïde / RGA Fusa bananier	VANPL_B_00008000870	C3H	Trimethyltridecatetraene synthase	B 8	VANPL_B_00008:2991219..2993204	
Phénylpropanoïde / RGA Fusa bananier	VANPL_B_00008000880	C3H	Trimethyltridecatetraene synthase	B 8	VANPL_B_00008:3006680..3008373	
Phénylpropanoïde	VANPL_B_00008002250	CAD	Alcohol dehydrogenase 9	B 8	VANPL_B_00008:6418656..6420756	
RGA LRR Profiler	VANPL_B_00008002640	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 7	B 8	VANPL_B_00008:7084124..7103002	
RGA LRR Profiler	VANPL_B_00008003690	RLK	putative inactive receptor kinase At5g10020	B 8	VANPL_B_00008:8247221..8252981	
RGA Fusa bananier	VANPL_B_00008003940	PPO	Polyphenol oxidase, chloroplastic	B 8	VANPL_B_00008:9284884..9289115	
RGA Fusa bananier	VANPL_B_00008003980	PPO	Polyphenol oxidase, chloroplastic	B 8	VANPL_B_00008:9499086..9501175	
RGA Fusa bananier	VANPL_B_00008003990	PPO	Polyphenol oxidase chloroplastic	B 8	VANPL_B_00008:9519267..9520996	
RGA LRR Profiler	VANPL_B_00008005440	RLK	Serine/threonine-protein kinase BRI1-like 2	B 8	VANPL_B_00008:15466785..15471165	
RGA Fusa bananier	VANPL_B_00008005480	NLR	Transcription factor MYB16	B 8	VANPL_B_00008:15720232..15721467	
RGA LRR Profiler	VANPL_B_00008005560	RLK	Receptor protein kinase TMK1	B 8	VANPL_B_00008:16084281..16088046	
FORL	VANPL_B_00008005880	RLK	Glutathione transferase GST 23	B 8	VANPL_B_00008:17205514..17206348	
RGA LRR Profiler	VANPL_B_00008006260	RLK	LRR receptor kinase BAK1	B 8	VANPL_B_00008:18088294..18109632	
RGA LRR Profiler	VANPL_B_00008007660	RLK	Protein kinase domain-containing protein	B 8	VANPL_B_00008:20596065..20598309	
RGA Fusa bananier	VANPL_B_00008008780	WRKY	AP2-like ethylene-responsive transcription factor At1g16060	B 8	VANPL_B_00008:22873416..22875863	
Phénylpropanoïde	VANPL_B_00008008810	HCT	Hydroxycinnamoyltransferase	B 8	VANPL_B_00008:22969623..22972512	
RGA Fusa bananier	VANPL_B_00008009420	NLR	myb-related protein P-like [Phalaenopsis equestris]	B 8	VANPL_B_00008:24227115..24229396	
RGA LRR Profiler	VANPL_B_00008009860	RLK	Protein kinase domain-containing protein	B 8	VANPL_B_00008:25026673..25031286	
RGA LRR Profiler	VANPL_B_00008010410	RLK	putative inactive receptor kinase At5g67200	B 8	VANPL_B_00008:25877248..25895529	
Phénylpropanoïde	VANPL_B_00008010440	CCoAMT	Caffeoyl-CoA O-methyltransferase	B 8	VANPL_B_00008:25930547..25931961	
RGA LRR Profiler	VANPL_B_00008010930	RLK	Protein NSP-INTERACTING KINASE 1	B 8	VANPL_B_00008:26536715..26540356	
FORL	VANPL_B_00008011760	GLIP	L-type lectin-domain containing receptor kinase IX.1	B 8	VANPL_B_00008:28000223..28008179	
RGA Fusa bananier	VANPL_B_00008011900	RLK	GDSL esterase/lipase At5g55050 [Phalaenopsis equestris]	B 8	VANPL_B_00008:28293233..28294867	
RGA LRR Profiler	VANPL_B_00009000180	RLK	putative inactive leucine-rich repeat receptor-like protein kinase At3g03770	B 9	VANPL_B_00009:279339..284289	
FORL	VANPL_B_00009000390	RLK	putative Glutathione S-transferase U24	B 9	VANPL_B_00009:471789..477607	
RGA LRR Profiler	VANPL_B_00009001130	RLK	LRR receptor-like serine/threonine-protein kinase FEI 1	B 9	VANPL_B_00009:1417876..1427163	
RP Sobano-DelaCruz	VANPL_B_00009001890	RPL24B	60S ribosomal protein L24-like [Phalaenopsis equestris]	B 9	VANPL_B_00009:2513707..2531862	
RGA LRR Profiler	VANPL_B_00009001980	RLK	putative inactive receptor kinase At5g58300	B 9	VANPL_B_00009:2703489..2710720	
RGA Fusa bananier	VANPL_B_00009002120	WRKY	WRKY transcription factor WRKY24-like isoform X2 [Phalaenopsis equestris]	B 9	VANPL_B_00009:2821472..2823510	
RGA Fusa bananier	VANPL_B_00009002910	ASP	aspartyl protease family protein At5g10770-like [Phalaenopsis equestris]	B 9	VANPL_B_00009:3927877..3931278	
RGA LRR Profiler	VANPL_B_00009003250	RLK	Leucine-rich repeat receptor protein kinase HPCA1	B 9	VANPL_B_00009:4419673..4425724	
RGA LRR Profiler	VANPL_B_00009003460	RLK	putative protein MALE DISCOVER 2	B 9	VANPL_B_00009:4732494..4755096	

RGA Fusa bananier	VANPL_B_000091004260	Cytochrome p450	Flavonoid 3'-monooxygenase CYP75B137	B	9	VANPL_B_00009:5985365..5987241	
RGA Fusa bananier	VANPL_B_000091004270	Cytochrome p450	Flavonoid 3'-monooxygenase CYP75B137	B	9	VANPL_B_00009:6009712..6011839	
RP Sobano-DeliaCruz	VANPL_B_000091004850	RP52C	40S ribosomal protein S2-4	B	9	VANPL_B_00009:7189732..7194070	
RGA Fusa bananier	VANPL_B_000091006400	ASP	Aspartic proteinase nepenthesin-2	B	9	VANPL_B_00009:12629399..12630965	
Phénylpropanoïde	VANPL_B_000091007250	HCT	Putrescine hydroxycinnamoyltransferase 1	B	9	VANPL_B_00009:15737555..15738919	
Phénylpropanoïde	VANPL_B_000091007270	HCT	Putrescine hydroxycinnamoyltransferase 1	B	9	VANPL_B_00009:15777139..15778503	
RGA LRR Profiler	VANPL_B_000091009170	RLK	LRR receptor-like serine/threonine-protein kinase GHR1	B	9	VANPL_B_00009:21450853..21455725	
RGA LRR Profiler	VANPL_B_000091009270	RLK	Receptor-like protein kinase 7	B	9	VANPL_B_00009:21583500..21586886	
Phénylpropanoïde	VANPL_B_000091010400	CCR	Tetraketide alpha-pyrone reductase 1	B	9	VANPL_B_00009:23262238..23263610	
Phénylpropanoïde	VANPL_B_000091010460	CCR	Cinnamoyl-CoA reductase 1	B	9	VANPL_B_00009:23352030..23357901	
RGA LRR Profiler	VANPL_B_000091010630	RLP	conserved hypothetical protein	B	9	VANPL_B_00009:23714893..23715456	
RGA LRR Profiler	VANPL_B_000091010640	RLK	receptor kinase-like protein Xa2.1 [Phalaenopsis equestris]	B	9	VANPL_B_00009:23739603..23743823	
RGA LRR Profiler	VANPL_B_000091010650	RLK	Inactive leucine-rich repeat receptor-like protein kinase	B	9	VANPL_B_00009:23758267..23761980	
RGA Fusa bananier	VANPL_B_000091010720	ASP	Aspartyl protease family protein At5g10770	B	9	VANPL_B_00009:23991040..23993333	
RGA Fusa bananier	VANPL_B_000091010780	NLR	transcription factor MYB80 [Phalaenopsis equestris]	B	9	VANPL_B_00009:24109339..24110315	
RGA LRR Profiler	VANPL_B_000091010870	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase RGI4	B	9	VANPL_B_00009:24210692..24214567	
RGA Fusa bananier	VANPL_B_000091010980	WRKY	ethylene-responsive transcription factor WR11-like [Phalaenopsis equestris]	B	9	VANPL_B_00009:24429225..24430916	
RP Sobano-DeliaCruz	VANPL_B_000091011060	RP124B	60S ribosomal protein L24	B	9	VANPL_B_00009:24696349..24703854	
RGA LRR Profiler	VANPL_B_000091011100	RLK	Leucine-rich repeat receptor protein kinase HPCA1	B	9	VANPL_B_00009:24759484..24764390	
Phénylpropanoïde	VANPL_B_000091011540	CSE	caffeoylshikimate esterase-like [Phalaenopsis equestris]	B	9	VANPL_B_00009:25901200..25905805	
RGA LRR Profiler	VANPL_B_000091011760	RLK	protein MALE DISCOVERER 2-like [Phalaenopsis equestris]	B	9	VANPL_B_00009:26455596..26463098	
RGA LRR Profiler	VANPL_B_000091012000	RLK	LRR receptor-like serine/threonine-protein kinase ER2	B	9	VANPL_B_00009:26993100..27000263	
RGA LRR Profiler	VANPL_B_000091012400	RLK	Protein kinase domain-containing protein	B	9	VANPL_B_00009:28195952..28233285	
FORL	VANPL_B_000091012850		Thaumatin-like protein	B	9	VANPL_B_00009:29518347..29519310	
FORL	VANPL_B_000091012970		Serine/threonine-protein kinase ATG1b	B	9	VANPL_B_00009:29644152..29670138	
Phénylpropanoïde	VANPL_B_000091013250	HCT	Tryptamine hydroxycinnamoyltransferase 2	B	9	VANPL_B_00009:30537941..30539456	
Phénylpropanoïde	VANPL_B_000091013260	HCT	Tryptamine hydroxycinnamoyltransferase 1	B	9	VANPL_B_00009:30549035..30550582	
Phénylpropanoïde	VANPL_B_000091013270	HCT	Tryptamine hydroxycinnamoyltransferase 2	B	9	VANPL_B_00009:30559446..30561098	
RGA LRR Profiler	VANPL_B_000091013390	RLK	LRR receptor kinase SEK2	B	9	VANPL_B_00009:30820508..30844382	
RGA LRR Profiler	VANPL_B_000101000070	RLK	Receptor protein-tyrosine kinase CEPR2	B	10	VANPL_B_00010:343192..354272	
Phénylpropanoïde	VANPL_B_000101001270	CSE	caffeoylshikimate esterase-like [Phalaenopsis equestris]	B	10	VANPL_B_00010:4032430..4049279	
RGA LRR Profiler	VANPL_B_000101003270	RLK	Leucine-rich repeat receptor-like serine/threonine-protein kinase RGI4	B	10	VANPL_B_00010:8263173..8266881	
RGA LRR Profiler	VANPL_B_000101003510	RLK	putative LRR receptor-like serine/threonine-protein kinase AT1g06840	B	10	VANPL_B_00010:8724300..8786217	
RGA Fusa bananier	VANPL_B_000101003590	POD	Cationic peroxidase 1	B	10	VANPL_B_00010:9107682..9109110	
RGA Fusa bananier	VANPL_B_000101003690	WRKY	WRKY transcription factor WRKY24-like isoform X1 [Phalaenopsis equestris]	B	10	VANPL_B_00010:9322625..9324465	
RP Sobano-DeliaCruz	VANPL_B_000101004000	RP523A	40S ribosomal protein S23	B	10	VANPL_B_00010:9938213..9947644	
FORL	VANPL_B_000101004460		Thaumatin-like protein	B	10	VANPL_B_00010:11139678..11140717	
FORL	VANPL_B_000101004470		L-type lectin-domain containing receptor kinase S.4	B	10	VANPL_B_00010:11146790..11149481	
RGA Fusa bananier	VANPL_B_000101004980	GLIP	GDSL esterase/lipase 7	B	10	VANPL_B_00010:11896778..11901196	
RGA LRR Profiler	VANPL_B_000101005350	RLK	leucine-rich repeat receptor-like protein kinase PXC1 [Phalaenopsis equestris]	B	10	VANPL_B_00010:13253551..13256296	

RGA Fusa autres espèces	VANPL_B_000101006050		G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	B	10	VANPL_B_00010:14131610..14134152	
RGA LRR Profiler	VANPL_B_000101006080	RLK	calmodulin-binding receptor kinase CaMRLK	B	10	VANPL_B_00010:14176211..14179142	
RGA LRR Profiler	VANPL_B_000101006230	RLK	Protein kinase domain-containing protein	B	10	VANPL_B_00010:14320952..14324121	
RGA LRR Profiler	VANPL_B_000101006620	RLK	Receptor-like protein kinase HSL1	B	10	VANPL_B_00010:14846984..14850515	
RGA LRR Profiler	VANPL_B_000101006940	RLK	Receptor protein kinase-like protein ZAR1	B	10	VANPL_B_00010:15266798..15270392	
RP Solano-DeIaCruz	VANPL_B_000101008110	RP51ZA	40S ribosomal protein S12	B	10	VANPL_B_00010:17393424..17402112	
RP Solano-DeIaCruz	VANPL_B_000101008860	RP51ZA	40S ribosomal protein S12	B	10	VANPL_B_00010:18986903..18992550	
RGA LRR Profiler	VANPL_B_000101011220	RLP	Receptor-like protein 44	B	10	VANPL_B_00010:28584543..28586566	
RGA Fusa bananier	VANPL_B_000101011320	POD	Lignin-forming anionic peroxidase	B	10	VANPL_B_00010:29037867..29039145	
RGA Fusa bananier	VANPL_B_000101011330	POD	Lignin-forming anionic peroxidase	B	10	VANPL_B_00010:29094532..29095807	
RGA Fusa bananier	VANPL_B_000101011340	POD	Lignin-forming anionic peroxidase	B	10	VANPL_B_00010:29146213..29147359	
RGA Fusa bananier	VANPL_B_000101011350	POD	Lignin-forming anionic peroxidase	B	10	VANPL_B_00010:29204871..29206017	
Phénylpropanoïde	VANPL_B_000111001370	OMT	Caffeic acid 3-O-methyltransferase 1	B	11	VANPL_B_00011:1958191..1959700	
Phénylpropanoïde	VANPL_B_000111002020	4CL	Oxalate-CoA ligase	B	11	VANPL_B_00011:2928051..2930122	
Phénylpropanoïde	VANPL_B_000111003360	CAD	Cinnamyl alcohol dehydrogenase 3	B	11	VANPL_B_00011:5626211..5627933	
RGA Fusa bananier	VANPL_B_000111003890	NLR	transcription factor WER-like [Phalaenopsis equestris]	B	11	VANPL_B_00011:6979754..6980889	
RGA Fusa bananier	VANPL_B_000111004730	WRKY	ethylene-responsive transcription factor WR1-like [Phalaenopsis equestris]	B	11	VANPL_B_00011:10902380..10904662	
RGA LRR Profiler	VANPL_B_000111005130	RLK	Leucine-rich repeat receptor protein kinase HPCA1	B	11	VANPL_B_00011:12761825..12767523	
Phénylpropanoïde	VANPL_B_000111005780	HCT	omega-hydroxypalmitate O-feruloyl transferase	B	11	VANPL_B_00011:18647821..18649539	
RGA LRR Profiler	VANPL_B_000111005970	RLK	putative inactive receptor kinase At1g48480	B	11	VANPL_B_00011:19612653..19614955	
RGA LRR Profiler	VANPL_B_000111006750	RLK	Protein kinase domain-containing protein	B	11	VANPL_B_00011:21743580..21773493	
RGA LRR Profiler	VANPL_B_000111006860	RLP	Receptor-like protein CLAVATA2	B	11	VANPL_B_00011:22042683..22045509	
Phénylpropanoïde	VANPL_B_000111007100	CCoAMT	Norbaldine O-methyltransferase	B	11	VANPL_B_00011:22491245..22491919	
Phénylpropanoïde	VANPL_B_000111007280	CAD	Alcohol dehydrogenase 9	B	11	VANPL_B_00011:22812068..22813523	
RP Solano-DeIaCruz	VANPL_B_000111007570	RP55	40S ribosomal protein S5	B	11	VANPL_B_00011:23443019..234440415	
RGA LRR Profiler	VANPL_B_000111008620	RLK	Receptor-like protein kinase HSL1	B	11	VANPL_B_00011:26052581..26056209	
Phénylpropanoïde	VANPL_B_000111008790	CAD	Alcohol dehydrogenase 9	B	11	VANPL_B_00011:26486089..26487560	
Phénylpropanoïde	VANPL_B_000111008870	CAD	Alcohol dehydrogenase 9	B	11	VANPL_B_00011:26525333..26526804	
RGA LRR Profiler	VANPL_B_000111009010	RLK	Protein kinase domain-containing protein	B	11	VANPL_B_00011:27044749..27053526	
RGA Fusa autres espèces	VANPL_B_000111009190		G-type lectin S-receptor-like serine/threonine-protein kinase SD2-2 [Phalaenopsis equestris]	B	11	VANPL_B_00011:27361418..27364267	
RGA LRR Profiler	VANPL_B_000111010030	RLP	piriformospora indica-insensitive protein 2-like [Phalaenopsis equestris]	B	11	VANPL_B_00011:28472268..28475031	
Phénylpropanoïde	VANPL_B_000111010200	CCR	Tetraketide alpha-pyrone reductase 2	B	11	VANPL_B_00011:28652308..28653817	
RGA Fusa bananier	VANPL_B_000111011220	WRKY	putative WRKY transcription factor 4	B	11	VANPL_B_00011:30537745..30557205	
RGA LRR Profiler	VANPL_B_000111011280	RLK	Protein STRUBBELIG-RECEPTOR FAMILY 5	B	11	VANPL_B_00011:30623722..30647142	
RGA LRR Profiler	VANPL_B_000111011570	RLP	Leucine-rich repeat receptor-like serine/threonine-protein kinase	B	11	VANPL_B_00011:31057050..31074652	
RGA LRR Profiler	VANPL_B_000121000260	RLK	Kinase-like protein TWK11	B	12	VANPL_B_00012:595001..597850	
RGA LRR Profiler	VANPL_B_000121002960	RLK	Inactive leucine-rich repeat receptor-like serine/threonine-protein kinase At1g60630	B	12	VANPL_B_00012:11198727..11201490	
Phénylpropanoïde	VANPL_B_000121003970	CAD	short-chain dehydrogenase TIC 32, chloroplastic	B	12	VANPL_B_00012:13830795..13833243	
RGA LRR Profiler	VANPL_B_000121004500	RLK	MDIS1-interacting receptor-like kinase 1	B	12	VANPL_B_00012:14524680..14528282	

RGa LRR Profiler	VANPL_B_00012i004510	RLK	Leudne-rich repeat receptor-like serine/threonine-protein kinase At1g17230	B 12	VANPL_B_00012:14542545..14548316	
RP Solano-DelaCruz	VANPL_B_00012i005690	RPL13B	60S ribosomal protein L13-2	B 12	VANPL_B_00012:18993140..18997750	
Phénylpropanoïde	VANPL_B_00012i005810	CCR	cinnamoyl-CoA reductase 2-like [Phalaenopsis equestris]	B 12	VANPL_B_00012:19452211..19457905	
RGa LRR Profiler	VANPL_B_00012i006160	RLK	receptor protein kinase-like protein ZAR1 [Phalaenopsis equestris]	B 12	VANPL_B_00012:20320421..20322933	
RGa LRR Profiler	VANPL_B_00012i006730	RLK	putative inactive receptor kinase At5g58300	B 12	VANPL_B_00012:21720072..21732088	
RGa LRR Profiler	VANPL_B_00012i006840	RLK	Leucine-rich repeat receptor-like protein kinase PXC2	B 12	VANPL_B_00012:22138444..22142148	
RGa LRR Profiler	VANPL_B_00012i007290	RLK	Systemin receptor SR160	B 12	VANPL_B_00012:23367745..23371571	
RGa LRR Profiler	VANPL_B_00012i007610	RLK	Non-specific serine/threonine protein kinase	B 12	VANPL_B_00012:24129673..24153740	
RGa LRR Profiler	VANPL_B_00012i007680	RLK	Non-specific serine/threonine protein kinase	B 12	VANPL_B_00012:24358716..24374404	
RGa LRR Profiler	VANPL_B_00012i007690	RLK	Non-specific serine/threonine protein kinase	B 12	VANPL_B_00012:24400922..24413242	
RGa LRR Profiler	VANPL_B_00012i008570	RLK	Leucine-rich repeat receptor-like protein kinase PXC1	B 12	VANPL_B_00012:26261590..26263741	
RGa LRR Profiler	VANPL_B_00012i009260	RLK	Protein kinase domain-containing protein	B 12	VANPL_B_00012:27440660..27443478	
RGa Fusa bananier	VANPL_B_00012i009280	WRKY	WRKY domain-containing protein	B 12	VANPL_B_00012:27492711..27494645	
Phénylpropanoïde	VANPL_B_00012i010050	COAO/MIT	Norbeldaline 4'-O-methyltransferase	B 12	VANPL_B_00012:28715430..28716631	
RGa LRR Profiler	VANPL_B_00012i010120	RLK	putative hypothetical protein	B 12	VANPL_B_00012:28816765..28820684	
RGa Fusa bananier	VANPL_B_00012i011170	NLR	Myb-related protein 315	B 12	VANPL_B_00012:30480120..30481151	
RGa LRR Profiler	VANPL_B_00012i011210	RLK	putative inactive receptor kinase At1g48480	B 12	VANPL_B_00012:30513409..30516513	
RGa LRR Profiler	VANPL_B_00012i011370	RLK	Mitogen-activated protein kinase kinase kinase YODA	B 12	VANPL_B_00012:30779918..30792956	
RGa LRR Profiler	VANPL_B_00012i012010	RLP	Somatic embryogenesis receptor kinase 1	B 12	VANPL_B_00012:31894875..31901524	
RGa Fusa bananier	VANPL_B_00012i012170	Cytochrome p450	Flavonoid 3'-monoxygenase CYP75B137	B 12	VANPL_B_00012:32131013..32133193	
RGa Fusa bananier	VANPL_B_00012i012180	Cytochrome p450	Flavonoid 3'-monoxygenase CYP75B137	B 12	VANPL_B_00012:32141001..32143112	
RGa LRR Profiler	VANPL_B_00012i012200	RLK	LRR receptor-like serine/threonine-protein kinase RPK2	B 12	VANPL_B_00012:32183981..32187868	
RGa LRR Profiler	VANPL_B_00012i012360	RLK	putative LRR receptor-like serine/threonine-protein kinase At2g16250	B 12	VANPL_B_00012:32428095..32430517	
RGa LRR Profiler	VANPL_B_00013i000140	RLK	Brassinosteroid LRR receptor kinase BRL1	B 13	VANPL_B_00013:15525111..15573961	
RP Solano-DelaCruz	VANPL_B_00013i000820	RPL27A	60S ribosomal protein L27-3-like [Phalaenopsis equestris]	B 13	VANPL_B_00013:2737216..2738136	
RGa LRR Profiler	VANPL_B_00013i001060	RLK	Receptor-like protein kinase HSL1	B 13	VANPL_B_00013:3177131..3180613	
RGa Fusa bananier	VANPL_B_00013i001500	WRKY	WRKY transcription factor 22-like	B 13	VANPL_B_00013:4059897..4061187	
RGa LRR Profiler	VANPL_B_00013i002010	RLK	Protein kinase domain-containing protein	B 13	VANPL_B_00013:4940450..4942932	
RGa Fusa bananier	VANPL_B_00013i004570	GLIP	GDSL esterase/lipase At5g37690	B 13	VANPL_B_00013:9685982..9687805	
Phénylpropanoïde	VANPL_B_00013i005580	HCT	<b>Omega-hydroxypalmitate O-feruloyl transferase</b>	B 13	VANPL_B_00013:11538828..11541113	
RGa LRR Profiler	VANPL_B_00013i006110	RLK	putative inactive receptor kinase At2g26730	B 13	VANPL_B_00013:12584658..12587054	
RGa LRR Profiler	VANPL_B_00013i006950	RLP	LRRT_2 domain-containing protein	B 13	VANPL_B_00013:13869786..13871609	
RGa LRR Profiler	VANPL_B_00013i008210	RLK	putative inactive receptor kinase At1g48480	B 13	VANPL_B_00013:16073959..16079055	
RGa Fusa bananier	VANPL_B_00013i008280	NLR	Protein ODORANT1	B 13	VANPL_B_00013:16258126..16259133	
RGa LRR Profiler	VANPL_B_00013i009460	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g74360	B 13	VANPL_B_00013:19385881..19389661	
RGa Fusa bananier	VANPL_B_00013i009570	Cytochrome p450	Flavonoid 3'-monoxygenase CYP75B137	B 13	VANPL_B_00013:19642291..19645160	
RGa Fusa bananier	VANPL_B_00013i009580	Cytochrome p450	Flavonoid 3'-monoxygenase CYP75B137	B 13	VANPL_B_00013:19669550..19671681	
RGa LRR Profiler	VANPL_B_00014i000120	RLK	putative LRR receptor-like serine/threonine-protein kinase IRK	B 14	VANPL_B_00014:824752..827584	
RGa Fusa bananier	VANPL_B_00014i001060	NLR	myb-related protein 308-like isoform X1 [Phalaenopsis equestris]	B 14	VANPL_B_00014:4156358..4158400	
RGa Fusa bananier	VANPL_B_00014i002000	PR1	Pathogenesis-related protein 1	B 14	VANPL_B_00014:6407772..6420411	

RGA Fusa bananier	VANPL_B_00014t002020	PR1	Pathogenesis-related protein 1	B	14	VANPL_B_00014:6443131..6444033	
Phénylpropanoïde	VANPL_B_00014t002100	CSE	caffeoylshikimate esterase-like [Phalaenopsis equestris]	B	14	VANPL_B_00014:6572858..6574797	
FORL	VANPL_B_00014t002230		G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	B	14	VANPL_B_00014:6866164..6869023	
RP Solano-DelaCruz	VANPL_B_00014t002640	RPL17A	60S ribosomal protein L17	B	14	VANPL_B_00014:7595690..7597375	
Phénylpropanoïde	VANPL_B_00014t003170	4CL	4-coumarate--CoA ligase-like 8	B	14	VANPL_B_00014:8628130..8634537	
Phénylpropanoïde	VANPL_B_00014t004070	OMT	Tricetin 3',4',5'-O-trimethyltransferase	B	14	VANPL_B_00014:10135593..10136969	
Phénylpropanoïde	VANPL_B_00014t004080	OMT	Tricetin 3',4',5'-O-trimethyltransferase	B	14	VANPL_B_00014:10144575..10146132	
Phénylpropanoïde	VANPL_B_00014t004090	OMT	Caffeic acid 3-O-methyltransferase	B	14	VANPL_B_00014:10149812..10151260	
Phénylpropanoïde	VANPL_B_00014t004110	OMT3	O-methyltransferase-3	B	14	VANPL_B_00014:10176989..10178593	
Phénylpropanoïde	VANPL_B_00014t004140	OMT2	O-methyltransferase-2	B	14	VANPL_B_00014:10194642..10197648	
RGA Fusa autres epibacter	VANPL_B_00014t004210		Callose synthase 10	B	14	VANPL_B_00014:10262896..10357192	
RGA Fusa bananier	VANPL_B_00014t004360	NLR	putative Myb-related protein Myb4	B	14	VANPL_B_00014:10608842..10611324	
Phénylpropanoïde	VANPL_B_00014t006010	4CL	Oxalate--CoA ligase	B	14	VANPL_B_00014:13289685..13292890	
RGA Fusa autres espèces / RGA Daphnia / RGA LRR Profiler / RGA Fusa bananier	VANPL_B_00014t006870		Disease resistance protein RGA3	B		VANPL_B_00014:14854308..14857698	
RGA LRR Profiler	VANPL_B_00014t007500	RLK	LRR receptor-like serine/threonine-protein kinase RGI1	B	14	VANPL_B_00014:16111911..16115824	
Phénylpropanoïde	VANPL_B_00014t008060	4CL	4-coumarate--CoA ligase-like 7	B	14	VANPL_B_00014:17356367..17361999	
RGA LRR Profiler	VANPL_B_00014t008180	RLK	putative LRR receptor-like serine/threonine-protein kinase At1g63430	B	14	VANPL_B_00014:17632958..17644779	





**Annexe 2.** Résumé de la présentation au congrès international Vanilla 2019, Moshi, Tanzanie (novembre 2019).

### **Advancements in Breeding Fusarium-Resistant Vanilla Varieties Using Next Generation Sequencing**

Félicien Favre<sup>1</sup>, Carine Charron<sup>2</sup>, Michel Grisoni<sup>2</sup>, Cyril Jourda<sup>2</sup> and Pascale Besse<sup>1</sup>

<sup>1</sup>University of La Réunion, UMR PVBMT, La Réunion

<sup>2</sup>Cirad, UMR PVBMT, La Réunion

**Keywords:** *Vanilla planifolia*, *Fusarium oxysporum* f. sp. *radicis-vanillae*, Resistance, Next-Generation Sequencing (NGS), Genetic diversity

#### **Abstract**

Demand for natural vanilla flavor is increasing, but the cultivated *Vanilla planifolia* faces critical challenges in response to biotic and abiotic stresses. In particular, Root and Stem Rot (RSR) disease caused by *Fusarium oxysporum* f. sp. *radicis-vanillae* (Forv) is the most damaging disease of vanilla. What can we do to reduce the damage? Prophylactic measures and use of chemicals or biocontrol agents are not effective in controlling RSR in the field. In this context, the use of resistant varieties is the best option to reduce Fusarium impact. Low genetic diversity has been detected within *V. planifolia* in accordance with the vegetative mode of dispersion in vanilla cultivation. However, among *V. planifolia* cultivars conserved in the Biological Resources Center Vatel at CIRAD La Réunion, several showed Forv resistant in lab tests as well as in the field. The exploration and deciphering of vanilla genomes using Next Generation Sequencing (NGS) provides a real reservoir of knowledge in support of the vanilla breeding programs. In particular, it opens access to genes involved in RSR plant resistance, allowing the development of molecular markers genetically linked to resistance.

**Annexe 3.** Résumé et poster présenté au 23<sup>rd</sup> World Orchid Conference, Taichu, Taiwan (avril 2021).

### **Genome-wide Assessment of the Genetic Diversity in *Vanilla***

Félicien Favre<sup>1</sup>, Cyril Jourda<sup>2</sup>, Michel Grisoni<sup>2</sup>, Katia Jade<sup>2</sup>, Jean-Bernard Dijoux<sup>2</sup>, Jérémy Hascoat<sup>2</sup>, Sandra Lepers<sup>3</sup>, Pascale Besse<sup>1</sup> and Carine Charron<sup>2</sup>

<sup>1</sup> University of La Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>2</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>3</sup> Etablissement Vanille de Tahiti, 98713 Papeete, Tahiti, Polynésie française, France

**Keywords:** Genetic diversity, Genotyping-by-Sequencing, *Vanilla planifolia*, *Vanilla* hybrids

### **Abstract**

*Vanilla* belongs to a complex taxonomic group characterized in particular by a vegetative reproduction mode combined with interspecific hybridizations in sympatric areas and polyploidy events. These mechanisms have a strong impact on the diversification of the *Vanilla* and complicate the delimitation of taxa. All of these genomic processes shaped the *Vanilla* genus diversity into more than 120 species distributed over three continents. However, due to its major clonal propagation, the genetic basis of cultivated vanillas (namely *V. planifolia* and *V. xtahitensis*) is extremely narrow. Genomics tools, such as Genotyping-by-Sequencing (GBS), are the key to explore the genetic diversity within the complex genus *Vanilla*. This powerful high-throughput genotyping method was applied on 133 vanilla accessions available at the Biological Resource Center Vatel (CIRAD, Reunion Island). Based on the 2,004 filtered SNPs obtained, inter and intra-specific relationships of *V. planifolia* and wild relative species were precisely analyzed with phylogenetic and Bayesian clustering analyses. Our results (i) showed a strong phylogenetic structuring between cultivated species and wild relative species, (ii) specified parental species at the origin of hybrids and (iii) allowed detecting intra-*V. planifolia* genetic diversity. Individuals resulting from sexual multiplication (selfing-pollination or intraspecific cross) exhibited greater genetic diversity than individuals resulting from vegetative reproduction. Our work highlights the strong potential of self-fertilization in cultivated vanilla to increase genetic diversity, and in particular, to breed new vanilla varieties adapted to major biotic and abiotic constraints.



# Genome-Wide Assessment of the Genetic Diversity in Vanilla

Félien Favre<sup>\*1</sup>, Cyril Jourda<sup>2</sup>, Michel Grisoni<sup>2</sup>, Katia Jade<sup>2</sup>, Jean-Bernard Dijoux<sup>2</sup>, Jérémy Hascoat<sup>2</sup>, Sandra Lepers<sup>3</sup>, Pascale Besse<sup>1</sup> and Carine Charron<sup>2</sup>

<sup>1</sup> University of La Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>2</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>3</sup> Etablissement Vanille de Tahiti, 98713 Papeete, Tahiti, Polynésie française, France

## Vanilla: a complex taxonomic group

- Vegetative reproduction mode
- Interspecific hybridizations in sympatric areas
- Polyploidy events

Strong impact on the diversification of *Vanilla* and complicated delimitation of taxa<sup>1,2,3,10</sup>

*Vanilla* genus diversity shaped into more than 120 species distributed over three continents<sup>8</sup>

- Two cultivated species: *Vanilla planifolia* and *Vanilla tahitensis*
- Low level of genetic diversity in *V. planifolia*<sup>2</sup>

## Objective

Exploitation of genomic tools to better understand phylogenetic relationships and impact of domestication in the *Vanilla* genus

## Genotyping by Sequencing (GBS) strategy to identify a large number of Single Nucleotide Polymorphisms (SNPs) in complex plants<sup>6</sup>

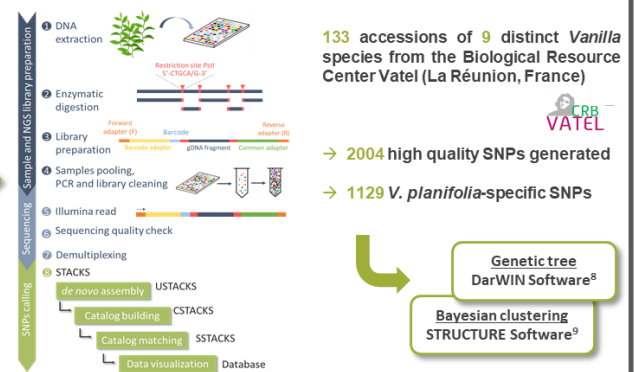


Fig. 1. Principle of GBS<sup>6</sup> and SNPs search strategy with the STACKS<sup>4</sup> pipeline

## GBS refines existing genetic relationships between cultivated vanillas and wild relative species

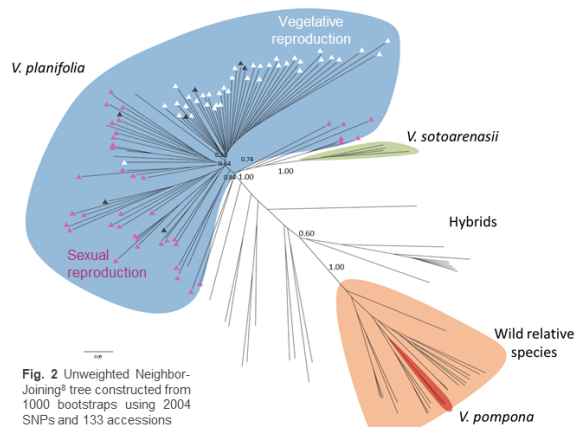


Fig. 2 Unweighted Neighbor-Joining<sup>9</sup> tree constructed from 1000 bootstraps using 2004 SNPs and 133 accessions

- High bootstraps values (>0.75) between the *V. planifolia* group and wild relatives (*V. bahiana*, *V. cribrana*, *V. helleri*, *V. insignis*, *V. odorata* and *V. pompona*)
- Hybrids in intermediate position
  - Strong genetic structuration between cultivated and wild relative
- Low bootstraps values (<0.50) between cultivated *V. planifolia* clones

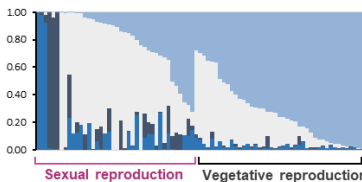


Fig. 3 Genetic structure<sup>9</sup> of 84 *V. planifolia* using 1129 SNPs

- Higher level of genetic diversity between *V. planifolia* obtained by sexual reproduction than those obtained by vegetative propagation
- Diversity generated by sexual reproduction

## GBS specifies the parental species at the origin of hybrids

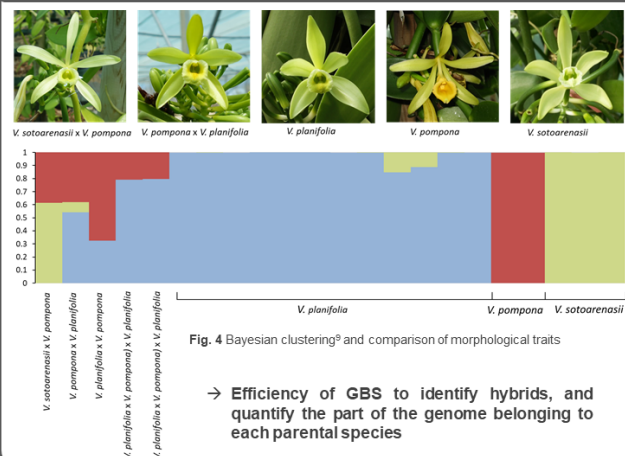


Fig. 4 Bayesian clustering<sup>9</sup> and comparison of morphological traits

→ Efficiency of GBS to identify hybrids, and quantify the part of the genome belonging to each parental species

- GBS: a powerful tool for detecting intra-*V. planifolia* genetic variability and identifying the parents of inter-specific hybrids

- Strong potential of self-fertilization or interspecific cross in cultivated vanilla to generate genetic diversification, an hopeful result for breeding program

- Besse P., Silva D.D., Bory S., et al. (2004). RAPD genetic diversity in cultivated vanilla: *Vanilla planifolia*, and relationships with *V. tahitensis* and *V. pompona*. *Plant Science* 167: 379–385.
- Bory S., Lubinsky P., Risterucci A.M., et al. (2008). Patterns of introduction and diversification of *Vanilla planifolia* (Orchidaceae) in Reunion Island (Indian Ocean). *American Journal of Botany* 95: 805–815.
- Bouetard A., Lefeuvre P., Gigant R., et al. (2010). Evidence of transoceanic dispersion of the genus *Vanilla* based on plastid DNA phylogenetic analysis. *Molecular Phylogenetics and Evolution* 55: 621–630.
- Catchen J., Hohenlohe P.A., Bassham S., et al. (2013). Stacks: an analysis tool set for population genomics. *Mol Ecol* 22: 3124–3140.
- Elshire R.J., Glaubitz J.C., Sun Q., et al. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6: e19379.
- Favre F., Jourda C., Besse P., Charron C. (2021). Genotyping-by-Sequencing Technology in Plant Taxonomy and Phylogeny. In: Besse P. (eds) *Molecular Plant Taxonomy, Methods in Molecular Biology*, 2222: 167–178.
- Lubinsky P., Cameron K.M., Molina M.C., et al. (2008). Neotropical roots of a Polynesian spice: the hybrid origin of Tahitian vanilla, *Vanilla tahitensis* (Orchidaceae). *American Journal of Botany* 95: 1040–1047.
- Perrier X., Flou A., Bonnad F. (2003). Data analysis methods. In: Hamon, P., Seguin, M., Perrier, X., Glaszmann, J. C. Ed., *Genetic diversity of cultivated tropical plants*. Entfield, Science Publishers 43–76.
- Pritchard J.K., Stephens M., Donnelly P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Soto Arenas M.A. and Cribb P. (2010). A new infrageneric classification and synopsis of the genus *Vanilla* Plum. ex mill. (Orchidaceae: *Vanillinae*). *Linnean Society* 9: 335–398.



This research was supported by the European Union: European Regional Development Fund, the Conseil Régional de La Réunion and the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD).

**Annexe 4.** Résumé et poster présenté au 9<sup>th</sup> Plant Genomics & Gene Editing congress: Europe, The Hague, Netherlands (avril 2022).

### **Identification and Annotation of Resistance Gene Analogs in *Vanilla planifolia***

Félicien Favre<sup>1</sup>, Quentin Da Silva<sup>2</sup>, Manon Morin<sup>2</sup>, Quentin Piet<sup>1</sup>, Cyril Jourda<sup>2</sup>, Michel Grisoni<sup>2</sup>, Pascale Besse<sup>1</sup> & Carine Charron<sup>2</sup>

<sup>1</sup> University of La Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France

<sup>2</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France

**Keywords:** *Vanilla planifolia*, *Fusarium oxysporum* f. sp. *radicis-vanillae*, Resistance Gene Analogs (RGA), Leucin-Rich-Repeat (LRR)

#### **Abstract**

Vanilla cultivation is limited by numerous diseases, including the Root and Stem Rot Disease caused by the soil fungus *Fusarium oxysporum* f. sp. *radicis-vanillae* (Forv). Obtaining crops with high levels of disease resistance is a major challenge for breeders. The genome of the main cultivated *Vanilla planifolia* is now available and provides the sequences of genes involved in resistance mechanisms. Well-documented bioinformatics tools were used for identification and annotation of Resistance Gene Analogs (RGA) in *V. planifolia* genome CR0040. Plant resistance genes are mainly encoded by three large Leucin-Rich Repeat (LRR)-containing receptor subfamilies: the LRR-receptor-Like Kinase (RLK), LRR-Receptor-Like Protein (RLP) and Nucleotide-binding LRR Receptor (NLR) subfamilies. The genome-wide analysis of the NLR genes of *V. planifolia* revealed extremely low numbers of disease resistance genes, and these results were consistent with the low number of such genes known in the Orchidaceae family. Genome-wide RGA analysis is a crucial step towards identification of genes involved in Forv resistance and provides new information for breeding programs.

# Identification and Annotation of Resistance Gene Analogs in *Vanilla planifolia*

## The complex Leucin-Rich Repeat (LRR)-containing receptor family

Félicien Favre<sup>1</sup>, Quentin Da Silva<sup>2</sup>, Manon Morin<sup>2</sup>, Quentin Piet<sup>3</sup>, Cyril Jourda<sup>2</sup>, Michel Grisoni<sup>2</sup>, Pascale Besse<sup>2</sup> & Carine Charron<sup>2</sup>

<sup>1</sup> University of La Réunion, UMR PVBMT, F-97410 St Pierre, La Réunion, France  
<sup>2</sup> CIRAD, UMR PVBMT, F-97410 St Pierre, La Réunion, France



UNIVERSITÉ DE LA RÉUNION

### 1. Introduction

- Vanilla cultivation is limited by numerous diseases, including the Root and Stem Rot Disease caused by *Fusarium oxysporum* f. sp. *radicis-vanillae* (Forv)
- Sources of resistance to Forv are observed in *Vanilla planifolia*, the main cultivated species. However, no molecular factor involved in resistance mechanisms has been identified

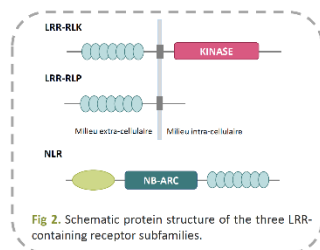


Fig 2. Schematic protein structure of the three LRR-containing receptor subfamilies.

➤ *V. planifolia*, a complex genome (Fig. 1)

➤ Plant resistance genes are mainly encoded by 3 large LRR-containing receptor subfamilies (Fig. 2)

Sekhwai et al., 2015

**Objectives : Identification and annotation of LRR-containing receptor subfamilies in the *V. planifolia* complex genome**

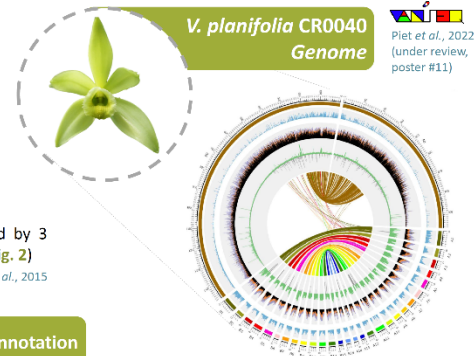


Fig 1. Circos representation of the *V. planifolia* genome showing the synteny between the 14 assembled chromosomes of both haplotype A and B

### 2. Method

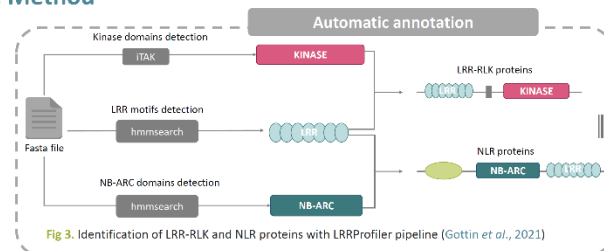


Fig 3. Identification of LRR-RLK and NLR proteins with LRRProfiler pipeline (Gottin et al., 2021)

- Manual checking of structures and domains of candidate genes
- Comparison of suspect gene sequences with proteomic data of other species for manual annotation
- Location of the genes in the genome of *V. planifolia* CR0040

### 3. A low number of NLR genes in *V. planifolia*

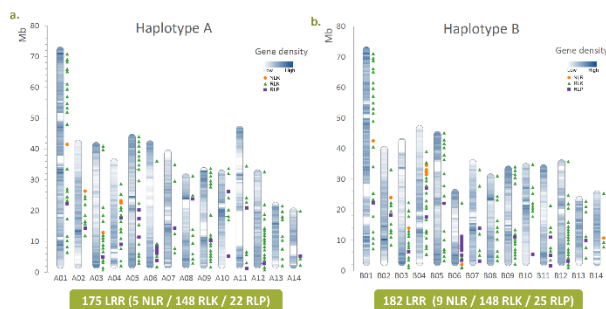


Fig 4. Location of LRR-RLK and NLR proteins within the *V. planifolia* complete genome a. Haplotype A and b. Haplotype B. Chromosomes are filled with gene density from Eugene annotation

- Genome-wide analysis of the disease resistance genes of *V. planifolia* revealed extremely low numbers of NLR compared with other crops.
- These results were consistent with the low number of NLR genes known in the Orchidaceae family (Fig. 5)

Xue et al., 2020

### 4. Perspectives

- Further analyses, in particular in a second version of the genome, are needed to identify other resistance gene analogs in *V. planifolia*.
- Transcriptomic and genetic data from a population segregating for the resistance are needed to identified molecular factors involved in the Forv resistance.

- 357 LRR-containing receptors were distributed all over the 14 chromosomes (Fig. 4)
- 15 NLR were annotated and one NLR cluster was identified in chromosome 4 in both haplotypes

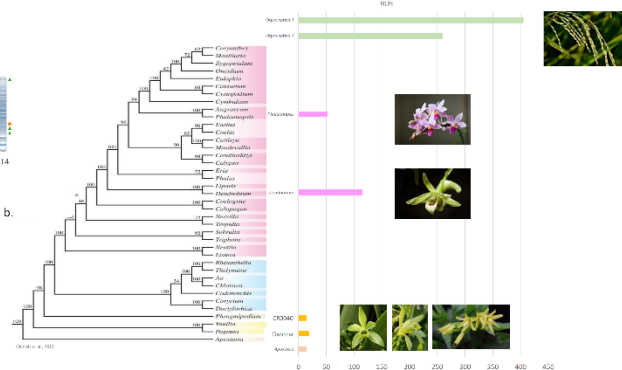


Fig 5. Plastome phylogeny for Orchidaceae, based on an ML analysis of 75 genes from the plastid genome of 39 orchid species (Givnish et al., 2015) and number of NLR identified in sequenced genomes

#### References

- Koyyappurath S, et al. (2016) Differential responses of Vanilla accessions to Root Rot and colonization by *Fusarium oxysporum* f. sp. *radicis-vanilla*, *Frontiers in Plant Science*, 6
- Piet Q, et al. (2022, under review) A chromosome-level, haplotype phased genome of *Vanilla planifolia* highlights that partial endoreplication challenges accurate whole genome assembly, *Plant Communications*.
- Sekhwai MK, et al. (2015) Disease Resistance Gene Analogs (RGAs) in plants, *International Journal of Molecular Sciences*, 16: 19248-19290
- Gottin C, et al. (2021) A New Comprehensive Annotation of Leucin-Rich Repeat Containing Receptors in Rice, *The Plant Journal*, 15456
- Xue JY, et al. (2020) Genome-wide analysis of the Nucleotide Binding Site Leucin-Rich Repeat genes of four Orchids revealed extremely low numbers of disease resistance genes, *Frontiers in Genetics*, 10:1286

This research was supported by the European Union: European Regional Development Fund, the Conseil Régional de La Réunion and the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD).





## Contribution of genomics to the exploration of the genetic diversity of cultivated vanilla and to the identification of regions involved in resistance to *Fusarium oxysporum*.

### Abstract

Cultivated for its aromatic beans, vanilla is an orchid that plays an important economic role in the southwestern Indian Ocean region. Like many other tropical species introduced outside their native area, the most cultivated species *Vanilla planifolia* G. Jackson has a narrow genetic base. Vanilla cultivation is limited by numerous diseases, including the root and stem rot disease caused by the soil fungus *Fusarium oxysporum* f. sp. *radicis-vanillae* (Forv). Chemical and prophylactic control methods have been unsuccessfully used and the study of resistance sources found in the natural diversity of the *Vanilla* genus seems to be the best strategy for a sustainable management of the disease.

Genetic studies in many crop species have been facilitated by the new generation sequencing technologies and the development of high-throughput molecular markers. Genotyping by sequencing (GBS) is a cost-effective method to discover a large number of single nucleotide polymorphism (SNP) molecular markers, in several individuals simultaneously, in species with a complex genome such as vanilla. GBS was applied in this study to the two cultivated species *V. planifolia* and *Vanilla × tahitensis* J.W. Moore, 7 closely related species and 19 interspecific hybrids. Based on 2,004 SNPs, the relationships between the different species were refined. The hybrid origin of *V. × tahitensis*, from *Vanilla odorata* C. Presl and a species genetically close to *Vanilla sotoarenasii* M. Pignal, was clarified by studying the genetic structure. A higher heterozygosity level was observed in the cultivated *V. planifolia* accessions compared to the wild species, suggesting a "single-step domestication" in accordance with their recent introduction and their clonal multiplication mode. Greater genetic diversity was observed in individuals derived from crosses, such as in individuals from the population obtained by self-fertilization of the traditional cultivar CR0040. GBS analysis of 125 individuals from this population segregating for Forv resistance identified 1,804 SNPs used for the construction of the first high-density genetic map in *V. planifolia*. The markers were ordered into 16 linkage groups, in accordance with the 16 chromosomes detected by previous cytogenetic studies. Their alignment against the 14 assembled chromosomes of the sequenced CR0040 genome suggested the assembly of 3 linkage groups on chromosome 1. Population resistance to the highly pathogenic strain Fo072 was assessed by an *in vitro* method. The 20 quantitative trait loci (QTLs) located on the genetic map explained, for each of the traits studied, from 23 to 66% of the total phenotypic variability, confirming the polygenic resistance to Forv. Genetic regions associated with resistance QTLs were investigated in the CR0040 genome annotation. These gene-rich chromosome regions included candidate genes potentially involved in biotic resistance mechanisms and encoding kinases, E3 ubiquitin ligases and pentatricopeptide repeat-containing proteins. Bioinformatics analysis of the sequencing data, focusing on the study of resistance gene analogs that have conserved domains and motifs, such as leucine-rich repeat (LRR) motif, indicated a low representation of this gene family in *V. planifolia*. A gene encoding a LRR receptor-like protein was found in a region underlying a QTL and appeared to be a key candidate for resistance to Forv.

Obtaining crops with high level of disease resistance is a major challenge for breeders and these results should contribute to a better understanding of the molecular mechanisms induced in response to infection by *F. oxysporum* f. sp. *radicis*, that cause severe losses in other economically important species.

**Keywords:** *Vanilla planifolia*, *Fusarium oxysporum* f. sp. *radicis-vanillae*, GBS, genetic diversity, QTL mapping, resistance gene



## Apport de la génomique à l'exploration de la diversité génétique des vanilliers cultivés et à l'identification des régions impliquées dans la résistance à la fusariose.

Cultivé pour ses fruits aromatiques, le vanillier est une orchidée d'importance économique majeure dans la région Sud-Ouest de l'océan Indien. Comme beaucoup d'autres espèces tropicales introduites hors de leur aire d'origine, l'espèce cultivée *Vanilla planifolia* G. Jackson possède une base génétique restreinte, la rendant vulnérable aux maladies. La pourriture des racines et des tiges, causée par le champignon *Fusarium oxysporum* f.sp. *radicis-vanillae* (Forv), est une maladie majeure chez les vanilliers. Les méthodes de lutte chimique et prophylactiques se sont montrées inefficaces et l'utilisation de sources de résistance présentes dans la diversité naturelle du genre *Vanilla* apparaît comme la meilleure stratégie pour une gestion durable de la maladie.

Les études génétiques chez les plantes ont été facilitées par les nouvelles technologies de séquençage et le développement de marqueurs moléculaires haut-débit. Le génotypage par séquençage (GBS) permet de générer rapidement un grand nombre de marqueurs moléculaires de type single nucleotide polymorphism (SNP) simultanément chez plusieurs individus avec des génomes complexes, comme celui du vanillier. La méthode GBS a été appliquée dans cette étude chez les deux espèces cultivées *V. planifolia* et *Vanilla × tahitensis* J.W. Moore, 7 espèces proches et 19 hybrides interspécifiques. Les 2004 SNPs identifiés ont permis d'affiner les relations de parenté entre les espèces. L'origine hybride de *V. × tahitensis*, entre *Vanilla odorata* C. Presl et une espèce proche de *Vanilla sotoarenasii* M. Pignal, a pu être précisée par l'étude de la structure génétique. Un taux d'hétérozygotie plus élevé a été observé chez les *V. planifolia* cultivées comparées aux espèces sauvages, suggérant une domestication « single step » en accord avec leur introduction récente et leur mode de multiplication clonale. Une plus grande diversité génétique a été observée chez les individus issus de croisements, comme la population obtenue par l'autofécondation du cultivar traditionnel CR0040. L'analyse GBS des 125 individus de cette population en ségrégation pour la résistance à Forv a permis de construire à l'aide de 1804 SNPs la première carte génétique haute-densité chez *V. planifolia*. Les marqueurs se regroupent en 16 groupes de liaison, conformément aux 16 chromosomes identifiés par cytogénétique. Leur alignement sur les 14 chromosomes du génome de CR0040 séquencé suggère l'assemblage de 3 groupes de liaison sur le chromosome 1. La résistance de la population à la souche hautement pathogène Fo072 a été évaluée par méthode *in vitro*. Les 20 quantitative trait loci (QTLs) localisés sur la carte génétique expliquent, pour chacun des caractères étudiés, entre 23 et 66 % de la variabilité phénotypique totale, montrant le caractère polygénique de la résistance à Forv. Les régions génétiques associées aux QTLs de résistance ont été recherchées dans le génome de CR0040. Ces régions sont riches en gènes, dont des gènes candidats potentiellement impliqués dans des mécanismes de résistance biotiques, codant pour des kinases, des ubiquitines ligases E3 et des protéines à motifs pentatricopeptide repeat. L'analyse bio-informatique des données de séquençage, en ciblant l'étude sur les analogues de gènes de résistance qui présentent des domaines et des motifs conservés, comme le domaine leucine-rich repeat (LRR), a révélé une faible représentation de cette famille de gènes chez *V. planifolia*. Un gène codant une protéine LRR receptor-like se retrouve dans une région sous-jacente à un QTL et apparaît comme un candidat clé pour la résistance à Forv.

L'étude de la diversité génétique et des facteurs impliqués dans la résistance à Forv appuie les programmes d'amélioration variétale et permet de mettre en évidence des mécanismes moléculaires mis en place en réponse à l'infection par des *F. oxysporum* de forme *radicis*, responsables de maladies majeures chez d'autres espèces d'importance économique.

**Mots clés :** *Vanilla planifolia*, *Fusarium oxysporum* f. sp. *radicis-vanillae*, GBS, diversité génétique, QTL, gène de résistance