

Maladie fongique et profils de vulnérabilité : le cas de la rouille des Myrtaceae (Austropuccinia psidii) en Nouvelle-Calédonie

Julia Soewarto

▶ **To cite this version:**

Julia Soewarto. Maladie fongique et profils de vulnérabilité : le cas de la rouille des Myrtaceae (Austropuccinia psidii) en Nouvelle-Calédonie. Phytopathologie et phytopharmacie. Université de la Nouvelle-Calédonie, 2017. Français. NNT : 2017NCAL00010 . tel-02934097

HAL Id: tel-02934097
https://unc.hal.science/tel-02934097

Submitted on 9 Sep 2020

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

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

Doctorat

Discipline : Biologie des populations et écologie

Julia SOEWARTO

Maladie fongique et profils de vulnérabilité : le cas de la rouille des Myrtaceae (*Austropuccinia psidii*) en Nouvelle-Calédonie.

Thèse dirigée par Laurent MAGGIA

Soutenue le 07 décembre 2017

Jury :

Hamid AMIR	Professeur des universités, Université de la Nouvelle-Calédonie	Président
Valérie BURTET-SARRAMEGNA	Maître de conférences, Université de la Nouvelle-Calédonie	Examinateur
Michel DRON	Professeur émérite, Université de Paris Sud – Paris Saclay	Examinateur
Gaël LECELLIER	Maître de conférences, HDR, Université de Versailles	Rapporteur
Jonathan PLETT	Chercheur, Hawkesbury Institute for the Environment, Western Sydney University, Australie	Rapporteur
Fabian CARRICONDE	Chercheur, Institut Agronomique néo-Calédonien	Co-encadrant
Laurent MAGGIA	Chercheur HDR, CIRAD	Directeur de thèse-

À mes parents, Jules et Georgiane Soewarto, qui m'ont enseigné la persévérance, le courage et le respect d'autrui.

Remerciements

Je tiens à remercier en premier lieu mes parents, Jules et Georgiane Soewarto. Merci d'avoir été à mes côtés tout au long de mes études, de m'avoir transmis votre force, de m'avoir soutenue aussi bien moralement que financièrement. Merci d'être des parents extraordinaires, capables de déployer tous les moyens possibles et inimaginables pour me protéger et me permettre de m'épanouir dans cette voie que j'ai choisie de suivre. Vous êtes et serez toujours mes piliers, mes héros au quotidien et mes modèles de réussite en tout point. Cette thèse est vôtre, elle est le fruit de notre travail, de nos sacrifices à tous les trois. Il n'existe pas de mots assez forts pour vous dire à quel point je vous suis reconnaissante et fière d'être votre fille.

Merci à mon directeur de thèse Laurent Maggia pour m'avoir donné l'opportunité de faire une thèse dans mon pays et dans un domaine qui me tient tant à cœur. Vous avez été l'initiative de ce sujet de thèse, de son côté novateur et à nous deux, nous avons parcouru un long chemin. Merci de m'avoir poussée à toujours me mettre en avant, souvent malgré moi et ma timidité légendaire, cela a toujours été payant au final. Merci de m'avoir remise sur les rails quand je m'écartais du chemin, d'avoir pris le temps de réfléchir avec moi quand je me retrouvais coincée, merci d'avoir cru en moi et d'avoir tout mis en œuvre pour trouver les financements de ce projet.

Merci à Stéphanie Sidibe-Bocs et Chantal Hamelin, mes encadrantes en bioinformatique. Merci pour tout le temps que vous m'avez accordé, pour votre investissement, votre gentillesse, votre patience à toute épreuve. Merci de vous être rendus disponibles chaque fois que j'en avais besoin, de m'avoir tant aidée et pour tous nos skypes réguliers. Sans vos précieuses connaissances, je doute que j'aurai pu arriver seule à tous nos résultats. Je suis partie de pratiquement zéro en bioinformatique et avec beaucoup d'appréhension. Il a fallu tout apprendre et vous avez réussi à m'y faire prendre goût. J'espère à l'avenir continuer à m'améliorer dans le domaine. Travailler avec vous a été l'une des plus belles expériences que j'ai vécues dans cette thèse et la plus enrichissante.

Merci à mon co-encadrant Fabian Carriconde pour avoir cru en moi et pour m'avoir boostée sur la fin de la thèse avec des discours ultra-positifs. Merci pour les corrections (in english et en français) et pour les sages conseils sur la thèse et la vie en général.

Je remercie celles qui ont ensoleillé chaque journée de travail. Elles sont mes collègues, mes poupoules, mes IACettes, je ne vous oublierai jamais. Kelly Letellier, ma Kellaii, tu as été là pour moi du début à la fin. Merci de m'avoir soutenue tout ce temps, d'avoir toujours cru en moi, de m'avoir fait rire chaque jour, de m'avoir gâtée en câlins et en bonbons. Anne Houlés, mon Anna Banana merci pour tes sourires, ta joie de vivre, tes bons conseils sur la thèse, la cuisine, la mode, et toutes nos joyeuses aventures. Mes plus beaux souvenirs de thèse, c'est à toi que je les dois. Géraldine Bidau-Ken, ma Gégé merci de m'avoir transmis ton savoir au labo, de m'avoir encouragée du début à la fin, et d'avoir toujours su trouver les paroles sages dans mes moments de fatigue et de rébellion. Sarah Gigante, ma princesse Sarah, merci pour ta douceur, ta gentillesse, ton soutien, tes sourires de chaque jour. Merci d'avoir toujours été là pour m'écouter, me conseiller, me donner du courage et pour tous nos moments passés ensemble. Laure Barrabé, ma Laurette à mes yeux une véritable divinité de la botanique et une source d'inspiration, merci pour toute l'aide que tu m'as apportée dans cette thèse. Mélanie Lelièvre, ma Mélanie jolie, tes bons conseils et ta sagesse ont littéralement changé ma vie…ce n'est pas une erreur, c'est de l'expérience ! Sylvie Russet, ma compatriote de

délires ! Merci pour nos moments de rigolade intense et nos conversations philosophiques du soir.

J'aimerais remercier le GDS-V pour toute leur aide et plus particulièrement Nicolas Hugot pour sa gentillesse, sa disponibilité et ses précieux conseils sur la reconnaissance de la rouille.

Je souhaiterais également exprimer ma gratitude à Carmela Gonzales et Philippe Rinck, propriétaires de la pépinière des Hauts de Farino. Merci à tous les deux de m'avoir laissée effectuer le suivi de la rouille au sein de votre pépinière. Merci d'avoir toujours été à l'écoute et merci également à Valériane pour sa gentillesse et l'entretien des plants au quotidien.

Je remercie chaleureusement le service SIVAP de la DAVAR pour toute leur aide et pour avoir été à l'initiative de la base de données rouille en Nouvelle-Calédonie. Je remercie en particulier Elodie Nakamura et Aurélie Chan pour leur gentillesse et leur aide.

Merci à Clarisse Majorel pour avoir partagé avec moi son précieux savoir sur l'extraction d'ARN et tous ses bons conseils !

Merci à Angélique Berger et Hélène Vignes pour m'avoir encadrée à Montpellier sur la banque d'ADNc et à Pierre Mournet pour tous ses bons conseils et pour avoir pris en charge l'envoi au séquençage.

A l'ensemble de l'équipe Intégration des Données du CIRAD de Montpellier, et en particulier à Alexis Dereeper, Gauthier Sarah, Alix Armero, Bertrand Pitollat, Gaetan Droc, Dominique This… Merci à tous(tes) de m'avoir accueillie dans votre équipe, pour votre gentillesse et pour toute l'aide que vous m'avez apportée durant cette thèse.

Je tiens à remercier Guillaume Martin de l'équipe Structure et Evolution des Génomes pour m'avoir permis d'utiliser ses scripts et pour toute l'aide précieuse qu'il m'a apportée pour les adapter à mes données.

Merci à Luc Baudouin pour ses conseils sur les FST et sur l'analyse des SNPs.

J'adresse également mes remerciements à Frédéric Breton pour ses conseils en phytopathologie.

Un grand merci à toutes les personnes qui m'ont accompagnée sur le terrain et qui m'ont aidée à traquer la rouille, Guillaume Lanuzel, Alexandre Bouarat, Anthony Pain, René Guiglion, Jean-Paul Latai et en particulier l'aide précieuse des déesses de la forêt, les botanistes aussi passionnantes que passionnées, Marion Anquez, Céline Chambrey, et Thais Vasconcelos.

Un grand merci à Vanessa Hequet, véritable gourou de la botanique, pour l'identification des espèces au niveau de l'herbier.

Merci à tous les contributeurs de la base de données de la rouille en Nouvelle-Calédonie, et ceux qui m'ont apportée des échantillons de très loin, en particulier Sophie Tronc pour Ouvéa et Julien Drouin pour Maré.

Merci à la société GLENCORE pour m'avoir aidée financièrement dans ce travail de thèse et en particulier merci à Sylvain Capo pour sa gentillesse et pour avoir cru en ce projet.

Merci à la Province Sud pour le prix d'encouragement à la recherche qu'ils m'ont attribué.

Merci à l'ensemble du personnel de l'Institut Agronomique néo-Calédonien pour leur aide, leur gentillesse et pour m'avoir soutenue tout au long de cette thèse.

Merci à Estelle Bonnet-Vidal qui m'a coachée pour le concours MT180 et auprès de qui j'ai beaucoup appris.

Merci à Armelle Tardivel pour sa gentillesse et son œil d'experte sur la mise en page.

Merci à Laurent Millet, maître jedi incontesté de la galaxie Plateforme du Vivant ! Merci pour ton aide précieuse au labo et pour tes conseils (toujours avisés). Et merci aussi à ta famille, Mylène, Norah et Timothée pour votre amitié, et votre soutien.

Merci à tous(tes) les copains et copines, compatriotes de thèse ou déjà docteurs, pour leurs conseils avisés et les joyeux moments de partage qu'on a passé ensemble. En particulier un grand merci à Julien Demenois, Bryan Vincent, Nicolas Folcher, Marine Toussirot, Simon Gensous, Laura Lagourgue, Patrick Houssard, Elefterios Chalkiadakis, Laurent Bordez.

Merci à Claire Billot et ses enfants pour m'avoir si gentiment accueillie durant toutes mes missions à Montpellier et pour m'avoir gâtée en confiture de fraise faite maison !

A ma grand-mère Aline Clavel, merci pour tout ton amour et ton soutien. Merci de me protéger depuis toujours, de me gâter de ton affection et de tes bons petits plats. Je t'aime par-dessus tout et j'espère te rendre fière.

A mon parrain et poète Jean-Yves Meuric, merci pour ton optimisme légendaire et ton soutien indéfectible. Merci d'avoir toujours été présent dans ma vie et de m'avoir autant encouragée dans mes études.

A l'ensemble de mes proches, famille et amis(es), de Koné à Montpellier, qui m'ont encouragée de près ou de loin, je vous remercie.

Résumé

Les espèces exotiques envahissantes et les maladies infectieuses représentent un problème croissant au niveau mondial et constituent l'une des principales menaces pour la biodiversité. La Nouvelle-Calédonie, archipel situé dans le sud-ouest du Pacifique, est reconnue comme l'un des principaux points chauds de la biodiversité. Sa flore vasculaire native illustre cette singularité biologique remarquable, avec 3250 espèces dont plus de 74% sont endémiques. Dans ce contexte, la gestion et la préservation de cette biodiversité est une question prioritaire.

Austropuccinia psidii est un champignon, considéré comme l'une des principales menaces environnementales pour les plantes appartenant à la famille des Myrtaceae à l'échelle mondiale. Communément connu sous le nom de "rouille des Myrtaceae", ce pathogène a été observé en Nouvelle-Calédonie en 2013. L'établissement de ce phytopathogène exotique est préoccupant compte tenu de l'importance environnementale occupée par les Myrtaceae dans les écosystèmes de l'archipel. L'objectif appliqué de cette thèse est de mettre en place une stratégie de lutte permettant à la fois de contrôler l'expansion du pathogène en milieu naturel, et dans des pépinières productrices de Myrtaceae.

Dans une première partie, nous allons mettre en évidence l'importance de l'impact d'*A. psidii* en termes d'étendue géographie, de gamme hôte et de sévérité des symptômes induits. Ces informations sont primordiales pour cibler spécifiquement les priorités de conservation des espèces et préservation des milieux les plus touchés. Les résultats montrent qu'A. *psidii* impacte un large panel d'espèces de Myrtaceae sur le territoire. Les variations dans la sévérité des symptômes exprimés entre individus d'une même espèce suggèrent la possibilité d'engager une lutte intégrée axée sur la sélection de caractères liés à la résistance à cette maladie.

Dans une seconde partie, nous tenterons d'identifier des marqueurs génétiques permettant de discriminer les individus en fonction du phénotype de résistance qu'ils expriment face à l'infection causée par la rouille. C'est ainsi que via une approche RNA-Seq, combinant à la fois une analyse de différentiel d'expression des gènes et la recherche de variants nucléotidiques directement positionnés sur des parties codantes des génomes de plantes concernées, nous sommes parvenus à identifier des gènes candidats discriminants pour les caractères résistants *vs* sensibles chez des espèces de Myrtaceae locales connues pour leurs caractères patrimonial, économique, ou encore écologique. Parmi ces gènes candidats certains sont impliqués dans les mécanismes de résistance aux pathogènes.

Mots clés : *Austropuccinia psidii*, Myrtaceae, gamme hôte, RNA-Seq, sélection, gènes candidats, résistance, Nouvelle-Calédonie

Abstract

Invasive alien species and emerging infectious diseases represent a growing worldwide problem and constitute one of the major threats to biodiversity. New Caledonia, an archipelago located in the South West Pacific, is recognized as one of the major biodiversity hotspot. The native vascular flora illustrates this outstanding biological uniqueness, with 3250 species of which over 74 % are endemic. In this context, management and preservation of this biodiversity is a priority issue.

Austropuccinia psidii is a fungus considered to be one of the major environmental threats to plants belonging to the family Myrtaceae on a global scale. Commonly known as "Myrtle rust", this pathogen has been observed in New Caledonia in 2013. The establishment of this exotic plant pathogen is of concern given the environmental importance of Myrtaceae in the archipelago ecosystems of Myrtaceae.

The objective of this thesis is to put in place a strategy to control both the expansion of the pathogen in a natural environment and in nurseries cultivating Myrtaceae.

In a first part, we will highlight the importance of the impact of *A. psidii* in terms of geographic extent, host range and severity of symptoms. This information is essential to specifically target the conservation priorities of the species most affected and preserved their habitats. The results show that *A. psidii* impacts a wide range of Myrtaceae species on the territory. Variations in the severity of the symptoms expressed between individuals of the same species suggest the possibility of initiating integrated control strategy based on the selection of traits related to resistance to this disease.

In a second part, we will try to identify genetic markers to discriminate individuals according to the phenotype of resistance they express to the infection caused by rust. By using an RNA-Seq approach, combining both differential gene expression and the search for nucleotide variations positioned in coding parts of the genomes, we were able to identify discriminant candidate genes for resistant vs susecptible in endemic Myrtaceae species known for their patrimonial, economic or ecological role. Among these candidate genes, some are involved in mechanisms of resistance to pathogens.

Keywords: *Austropuccinia psidii*, Myrtaceae, host range, RNA-Seq, selection, candidate gene, resistance, New-Caledonia

Ce travail de thèse a été effectué au sein de l'équipe Sol & Végétation (SolVeg) de l'Institut Agronomique néo-Calédonien (IAC) basée sur le centre IRD (Institut de Recherche pour le Développement) de Nouméa.

Il a bénéficié de :

- l'accès à la Plateforme du Vivant de la Nouvelle-Calédonie (IAC, IFREMER, IPNC, IRD, UNC) ;
- l'appui de l'équipe Intégration de données (ID), UMR AGAP, CIRAD.
- l'appui du Grand plateau technique régional de génotypage, UMR AGAP, CIRAD
- l'appui de l'équipe Dynamique de la diversité, sociétés et environnements (DDSE), UMR AGAP, CIRAD

Ce travail a été financé par :

- le Prix d'Encouragement à la Recherche (PER) de la Province Sud de la Nouvelle-Calédonie pour une durée de trois ans ;
- un complément de bourse délivré par l'IAC d'une durée de 3 ans et 4 mois ;
- des frais de fonctionnement propres à l'IAC et au CIRAD ;
- et d'une convention de partenariat de recherche avec la société GLENCORE.

Avertissement

Le contenu de cette thèse est bilingue. Le Chapitre 1, le Chapitre 4 ainsi que la Discussion générale sont rédigés en français. Les Chapitres 2 et 3 présentent deux articles scientifiques rédigés en anglais, dont l'un a été soumis à Forest Pathology et accepté et l'autre est en cours de finalisation pour soumission à Genomics Data.

Warning

The content of this thesis is bilingual. Chapter 1, Chapter 4 and the general conclusion are written in French. Chapters 2 and 3 present two scientific papers written in English, one has been submitted in Forest Pathology and accepted and the other is in process of finalization for submission to Genomics Data.

Tables des matières

Remerciements……v
Résumé……ix
Abstract……x
Tables des matières……xiv
Liste des figures……xix
Liste des tableaux……xxi
Liste des annexes et des données supplémentaires……xxiv
Abbreviations……xxv

Préambule……1

Introduction générale et objectifs de la thèse……3

Chapitre I : Synthèse bibliographique……6

1.1. Généralités sur les relations plantes hôtes-parasites……7

1.1.1. Le concept de maladie chez les plantes……7

1.1.2. Le triangle d'interaction……9

1.1.3. Les formes de parasitisme chez les végétaux……10

1.1.4. Champignons phytopathogènes……10

1.1.5. Les rouilles……12

1.1.5.1. Taxonomie……12

1.1.5.2. Cycle de vie……14

1.1.5.3. Mode d'infection des rouilles……16

1.1.5.4. Gamme d'hôtes……18

1.2. Le système immunitaire des végétaux……19

1.2.1. Les défenses préformées ou constitutives……19

1.2.2. Les défenses inductibles……21

1.2.2.1. Plant Triggered Immunity (PTI) : la résistance basale ou non-spécifique……23

1.2.2.2. Effector Triggered Susceptibility (ETS) : la désactivation de la résistance basale……25

1.2.2.3. Effector Triggered Immunity (ETI) : La résistance induite spécifique……26

1.2.2.4. Transduction du signal : réponse précoce à la détection de pathogènes……28

1.2.2.5. Les réponses de défense des plantes……28

i. Le renforcement des parois cellulaires……28

i. La réaction d'hypersensibilité……29

ii. Les phytoalexines……31

iii. Les Pathogenesis-Related Proteins 31
1.3. Gestion des maladies de plantes par la sélection 31
1.3.1. Historique de la sélection : de la sélection massale à l'utilisation des marqueurs moléculaires 32
1.3.2. Sélection de caractères de résistance aux pathogènes 33
i. La résistance qualitative 33
ii. La résistance quantitative 34
1.3.3. Stratégie d'utilisation de la résistance en sélection 35
1.3.4. La sélection chez les espèces forestières 35
1.3.5. Le cas des espèces forestières non-modèles 37
1.4. Le pathosystème Myrtaceae-*Austropuccinia psidii* : État des connaissances. 39
1.4.1. Les Myrtaceae 39
1.4.1.1. Taxonomie et caractéristiques 39
1.4.1.2. Origine et distribution 40
1.4.1.3. Importance écologique et économique 42
1.4.1.4. Ressource génomique 42
1.4.2. *Austropuccinia psidii* 43
1.3.2.1. Taxonomie 43
1.3.2.2. Distribution 45
1.3.2.3. Gamme hôte et épidémies 46
1.3.2.4. Cycle de vie d'*A. psidii* 47
1.3.2.5. Description des symptômes 48
1.3.2.6. Les facteurs épidémiologiques 49
1.3.2.7. Diversité d'*A. psidii* 49
i. Variabilité physiologique des urédospores en lien avec la pathogénicité 49
ii. Diversité génétique des populations d'A. psidii 50
1.3.2.8. Ressource génomique 51
1.3.2.9. Modèle d'interaction avec Myrtaceae 51
1.4.3. Gestion de la maladie 52
1.4.3.1. Détection et identification du pathogène 52
1.4.3.2. Moyens de lutte chimique 53
1.4.3.3. Lutte par la sélection de plants résistants à la rouille 53
i. Variabilité inter- et intraspécifique de la résistance à la rouille 54
ii. Gène de résistance 55

Chapitre II : Etat des lieux de la présence d'*Austropuccinia psidii* en Nouvelle-Calédonie : distribution, gamme hôtes et diversité génétique..... 58
2.1. Introduction générale du Chapitre 2 59
2.2. Article: The impact of *Austropuccinia psidii* in New Caledonia, a biodiversity hotspot. 60
2.2.1. Abstract 61
2.2.2. Introduction 62
2.2.3. Materials and methods 64
2.2.3.1. Disease surveys and sampling of *A. psidii* 64
2.2.3.2. DNA extraction, PCR and sequencing 64
2.2.3.3. Microsatellite genotyping 65
2.2.3.4. Myrtle rust symptoms monitoring trial under natural infection 65
2.2.4. Results 68
2.2.4.1. Identity confirmation 68
2.2.4.2. Geographical distribution of A. psidii in New Caledonia 68
2.2.4.3. Host range of *A. psidii* 70
2.2.4.4. Genotyping of A. psidii populations 74
2.2.4.5. Progress of disease severity over time on 35 cultivated Myrtaceae species in a field nursery 75
2.2.5. Discussion 76
2.2.5.1. Arrival of Austropuccinia psidii in New Caledonia 76
2.2.5.2. Spread of A. psidii in natural environments 76
2.2.5.3. Host range expansion 78
2.2.5.4. Potential consequences for biodiversity 78
2.2.5.5. Genetic uniformity of A. psidii isolates 79
2.2.5.6. Variation in disease susceptibility over a range of cultivated Myrtaceae species 80
2.2.6. Conclusion 81
2.2.7. References 82

Chapitre III : Mise en place d'une méthodologie pour la selection de marqueurs de résistance à la rouille chez des espèces de Myrtaceae endémiques à la Nouvelle-Calédonie incluant une appoche RNA-Seq 94
3.1. Introduction générale et objectifs du Chapitre 3 95
3.2. Matériel et Méthodes 95
3.2.1. Choix des espèces à étudier, localités et protocole d'échantillonnage 95
3.2.2. Extraction de l'ARN 97
3.2.3. Construction des banques d'ADNc et séquençage 98
3.2.4. Analyses bioinformatiques des données RNA-Seq 99

3.2.4.1. Contrôle qualité des *reads* bruts 99
3.2.4.2. L'alignement des reads 99
3.2.4.3. L'annotation 100
3.2.4.4. Quantification des transcrits et normalisation 101
3.2.4.5. Différentiel d'expression 102
3.2.4.6. SNP calling : du prétraitement des données à la prédiction des SNPs 102
3.3. Article : A transcriptomic-based approach to identify candidate markers for resistance to myrtle rust (*Austropuccinia psidii*) in three endemic Myrtaceae species from New Caledonia 105
3.3.1. Abstract 106
3.3.2. Introduction 107
3.3.3. Material and methods 109
Plant Material 109
Total RNA extraction 111
cDNA library preparation and sequencing 111
RNA-seq data processing 112
Raw reads cleaning 112
Reference- guided method 112
Aligning to a reference genome from a related species 112
Calling SNPs using E. grandis reference genome 113
De novo methods 113
De novo transcriptome assembly 113
Aligning to the de novo transcriptomes 114
Calling SNPs using de novo transcriptome assemblies 114
Variant filtering 114
Differential gene expression analysis 115
Annotation 115
Data visualization 116
Datasets release 117
3.3.4. Results and discussion 119
3.3.4.1. Raw data summarization and cleaning reads 120
3.3.4.2. Transcriptome assembly 120
3.3.4.3. Mapping cleaned reads to *E. grandis* genome and *de novo* transcriptome 121
3.3.4.4. Differentially Expressed Genes 122
3.3.4.5. SNP Calling results 126
3.3.4.6. Genome browser 127
3.3.5. Conclusion and perspectives 128

3.3.6. Acknowledgements 129
3.3.7. References 130
3.3.8. Supplementary data 133

Chapitre IV : Association entre variation de l'ADN et gènes différentiellement exprimés en réponse à la rouille (*A. psidii*) 138
4.1. Introduction 139
4.2. Materiel et methodes 140
4.2.1. Best Blast Mutual Hits 141
4.2.2. Choix des SNPs discriminants entre phénotypes résistants et sensibles à la rouille 141
4.2.3 Compilation des résultats du SNP calling, du différentiel d'expression et du BBMH 142
4.3. Resultats et discussion 142
4.3.1. Sélection des SNPs discriminants résistants/sensibles à *A. psidii* pour chaque espèce.. 142
4.3.2. Compilation des SNPs discriminants et les gènes différentiellement exprimés identifiés via l'alignement effectué à partir du génome de l'*E. grandis*. 143
4.3.3. Sélection des SNPs discriminants communs entre les trois espèces 144
4.3.4. Sélection des gènes différentiellement exprimés communs aux trois espèces via les alignements effectués sur les transcriptomes *de novo* 148
4.3.5. Compilation des SNPs discriminants et des gènes différentiellement exprimés à partir des alignements sur transcriptome *de novo* pour chaque espèce 149
4.3.6. Illustration d'un SNP discriminant dans un gène différentiellement exprimé du JBrowse *S. longifolium*. 151
4.4. Conclusion 153

Chapitre V : Discussion generale et perspectives 155
5.1. Conséquences écologiques et économiques de la rouille (*A. psidii*) en Nouvelle-Calédonie156
5.2. Une possible part de génétique dans l'expression des réponses à la rouille mise en évidence par un dispositif d'infection naturelle 157
5.3. La construction de transcriptomes *de novo* pour trois espèces de Myrtaceae endémiques à la Nouvelle-Calédonie 158
5.3. Une nouvelle approche méthodologique pour trouver des gènes candidats 158
5.4. Des mécanismes de réponse à la rouille spécifiques à chaque espèce 160
5.5. L'identification de gènes candidats pour deux espèces de Myrtaceae 161

Bibliographie……..……..162
Annexes……182

Liste des figures

INTRODUCTION

Figure 1 : Localisation de la Nouvelle-Calédonie dans le sud-ouest de l'océan Pacifique (A) et distribution des principaux types de végétation (B). Adapté de Isnard et al. (2016).

CHAPITRE 1

Figure 1.1 : Schéma représentant les principales fonctions d'une plante (gauche) pouvant être affectées par des pathogènes (droite) (extrait d'Agrios, G.N. 2005. Plant Pathology, 5th ed.).

Figure 1.2 : Les deux paradigmes de la phytopathologie, A. Le triangle de maladie et B. Le cycle de maladie**.** Extrait d'Agrios, G.N. 2005. Plant Pathology, 5th ed

Figure 1.3 : Espèces de rouilles chez les plantes cultivées.

Figure 1.4 : Représentation schématique du cycle d'une rouille macrocyclique et dioïque *Puccinia graminis* f. sp. *tritici* dont l'hôte principal est le blé (*Triticum aestivum*) et l'hôte alternatif est l'épinette verte (*Berberis vulgaris*).

Figure 1.5 : Processus d'infection d'une rouille

Figure 1.6 : Illustrations des barrières physiques des plantes contre l'entrée des pathogènes.

Figure 1.7 : Schéma du modèle en Zig-Zag illustrant le système immunitaire inné des plantes et l'évolution simultanée des mécanismes de défense des plantes et d'attaque des agents pathogènes. D'après Jones and Dangl, 2006.

Figure 1.8 : Systèmes de perception des MAMPs/DAMPs et des effecteurs dérivés des agents pathogènes chez les plantes. D'après Sanabria, Huang, & Dubery, 2010

Figure 1.9 : Schéma des cascades de signalisation des MAPK (Mitogen-Activated Protein Kinase) activées lors de la reconnaissance des signaux moléculaires associés aux microorganismes MAMP (Microbe-Associated Molecular Pattern), DAMP (Damage-Associated Molecular Pattern) de la PTI, ainsi que les éliciteurs d'avirulence de l'ETI. D'après (Meng and Zhang, 2013).

Figure 1.10 : Schéma illustrant les différents types de SNPs

Figure 1.11 : Répartition mondiale de la famille des Myrtaceae (en noir). Extrait de Grattapaglia et al., 2012.

Figure 1.12 : Phylogénie des Myrtaceae d'après Thornhill et al., 2015.

Figure 1.13 : Positionnement phylogénétique d'*Austropuccinia psidii* (syn. *Puccinia psidii*), récoltés à partir de plusieurs hôtes de Myrtaceae, parmi la famille des Sphaerophragmiaceae. D'après (Beenken, 2017)**.**

Figure 1.14 : Distribution globale d'*Austropuccinia psidii* avec les années de détection et le génotype. D'après Graça et al., 2013 et Granados et al., 2017.

Figure 1.15 : Représentation schématique du cycle de vie d'*Austropuccinia psidii*. D'après (Glen et al., 2007)

Figure 1.16 : Pathogénèse d'*Austropuccinia psidii* chez l'*Eucalyptus grandis.* (Xavier et al., 2001).

CHAPITRE 2

Figure 2.1: Map of myrtle rust reports in New Caledonia according to years.

Figure 2.2: Distribution of myrtle rust reports repartitions in terms of environment type.

Figure 2.3: Myrtle rust symptoms on (a) leaves of *Cloezia artensis var. artensis;* on fruits of (b) *Eugenia gacognei, (*c) *Eugenia stricta,* (d) *Eugenia bullata;* on flowers and buds of (e) *Metrosideros operculata var. francii,* (f) *Cloezia artensis var. artensis,* (g) *Eugenia noumeensis, (*h) *Eugenia excorticata.*

Figure 2.4: Myrtle rust symptoms on leaves of *Syzygium jambos* (a) and *Metrosideros brevistylis* (c); and tree death following several infections on *Syzygium jambos* (b) and *Metrosideros brevistylis* (d).

CHAPITRE 3

Figure 3.1 : Schéma du dogme central de la biologie moléculaire décrivant les mécanismes d'expression des gènes en protéines chez les eucaryotes. © 2010 Nature Education

Figure 3.2 : Illustration des espèces étudiées pour le RNA-Seq.

Figure 3.3 : Résultats issus du bioanalyzer pour l'échantillon Ag19 (*Arillastrum gummiferum*).

Figure 3.4 : Rappel sur la structure d'un gène.

Figure 3.5 : Transcrits de différentes longueurs avec différents niveaux de couverture de reads (gauche), nombre total de reads pour chaque transcrit (milieu) et nombre de reads normalisés par FPKM (droite). Extrait de Garber *et al.*, 2011.

Figure 3.6 : Workflow d'analyse recommandé par GATK.

Figure 3.7: Bioinformatics pipeline showing the different steps involved in RNA-seq analysis until alignement to the two kind of reference. Additional steps required for de novo transcriptome assembly are shown in the red box.

Figure 3.8: Analytic pipeline to identify potential candidate genes using both differentially expressed genes and SNP calling. The dotted arrows will not be treated in this work but in a future

Figure 3.9: Quality control statistics generated by FastQC for individual Syl18 (*S. longifolium*) at different stages of the data cleaning process.

Figure 3.10: Venn diagram showing the differentially expressed genes in *A. gummiferum*, *S. longifolium* and *T. glauca* (FAR and BDS) using alignments with *E. grandis* reference genome.

Figure 3.11: Numbers of SNPs per calling methods on species alignments with *E. grandis* reference genome (A) and with each *de novo* transcriptomes (B).

Figure 3.12: Screenshot of the JBrowse

CHAPITRE 4

Figure 4.1. Illustration d'un SNP discriminant localisé dans le gène SYL_g255820 différentiellement exprimé chez *S. longifolium* à partir de l'alignement sur référence *de novo*.

Liste des tableaux

CHAPITRE 1

Tableau 1.1 : Exemples de maladies causées par des agents pathogènes appartenant à l'ordre des Pucciniales et leurs hôtes respectifs.

Tableau 1.2 : Modèle génétique de la résistance gène-pour-gène (d'après Flor, 1971).

CHAPITRE 2

Table 2.1: List of Myrtaceae species monitored for myrtle rust symptoms under natural infection and disease assessments parameters.

Table 2.2: Host range of *A. psidii* in New Caledonia.

Table 2.3: Genotyping results on 65 *A. psidii* isolates from New-Caledonia.

CHAPITRE 3

Table 3.1: Detailed sampling of three Myrtaceae species for RNA-seq analysis.

Table 3.2: Number of raw and cleaned reads from the three species.

Table 3.3: Statistics of the *de novo* transcriptome assembly for each species using Trinity assembler.

Table 3.4: Alignment statistics indicative of reads aligned to the assembled transcriptome using SamTools flagstat

Table 3.5: Differentially expressed gene resulting from EdgeR.

Table 3.6: List of common differential expressed genes between *A. gummiferum*, *T. glauca* and *S. longifolium* using *E. grandis* reference genome.

Table 3.7: Overlapped and uniques SNPs called using two different calling methods (GATK and in-house script) on *de novo* transcriptomes.

CHAPITRE 4

Tableau 4.1 : Sélection des SNP discriminants sur la base du génotype de la référence.

Tableau 4.2 : Nombre de SNPs discriminants par espèce sur les alignements avec les transcriptomes *de novo*

Tableau 4.3 : Récapitulatif des SNPs discriminants communs entre *A. gummiferum, S. longifolium* et *T. glauca*

Tableau 4.4 : Annotation fonctionnelle des 9 SNPs communs entre les trois espèces et identifiés avec la méthode Unified Genotyper.

Tableau 4.5 : Annotation fonctionnelle des 4 SNPs communs entre les trois espèces et identifiés avec la méthode in-house.

Tableau 4.6 : Comparaisons des gènes différentiellements exprimés à partir des alignements sur les transcriptomes *de novo* et communs aux trois espèces étudiées.

Tableau 4.7 : Liste des gènes différentiellement exprimés communs aux 3 espèces trouvés sur *de novo* avec les SNP disciriminants (méthode Unified Genotyper)

Tableau 4.8 : Liste des gènes DE communs aux 3 espèces trouvés sur denovo avec les SNP disciriminant (méthode In-house)

Tableau 4.9 : Croisement par espèce des SNPs obtenus avec les méthodes Unified Genotyper et In-house.

Liste des annexes et des tableaux supplémentaires

Chapitre II

Supplementary table 2.1. List of DDBJ accession number of *Austropuccinia psidii* isolates from New Caledonia for the ITS, β-tubulin and EF1-α loci.

Supplementary table 2.2 : List of *Austropuccinia psidii* host species in New-Caledonia

Supplementary table 2.3 : Details of *A. psidii* isolates used for microsatellites genotyping

Supplementary table 2.4 : Maximum disease severity scores for each replicate of each species monitored in the field nursery and under natural infection. (-) was set for missing replicates.

Chapitre III

Annexe 3.1 : Du gène à l'ARN

Annexe 3.2 : Structure d'une banque d'ADNc pour le séquençage Illumina en paired-end.

Annexe 3.3 : Rappel de la technologie de séquençage haut-débit d'Illumina.

Annexe 3.4 : Extrait du descriptif d'un *read* de l'individu Ag19 (*Arillastrum gummiferum*) dans un fichier .FASTQ.

Supplementary table 3.1: Detailed of RNA extracts

Supplementary table 3.2: Number of raw and cleaned reads. The individual used for *de novo* transcriptome assembly are shown in yellow.

Supplementary table 3.3: Number of SNP following each filtering step.

Chapitre IV

Annexe 4.1 : Annotation fonctionnelle des 5 gènes différentiellement exprimés en communs entre les 3 espèces sur les alignements *de novo*.

Annexe 4.2 : Annotation fonctionnelle des 29 des SNPs discriminants localisés dans des gènes différentiellement exprimés chez *A. gummiferum* (Unified Genotyper). En jaune sont indiqués les SNPs communs avec la méthode In-house script.

Annexe 4.3 : Annotation fonctionnelle et positions des 3 SNPs discriminants localisés dans des gènes différentiellement exprimés chez *A. gummiferum* (In-house).

Annexe 4.4 : Annotation fonctionnelle des 63 SNPs discriminants localisés dans des gènes différentiellement exprimés chez *S. longifolium* (Unified Genotyper).

Annexe 4.5 : Annotation fonctionnelle et positions des 68 SNPs discriminants localisés dans des gènes différentiellement exprimés chez *S. longifolium* (In-house). En jaune sont indiqués les SNPs communs avec la méthode GATK.

Annexe 4.6 : Annotation fonctionnelle et positions des 68 SNPs discriminants localisés dans des gènes différentiellement exprimés chez *S. longifolium* (In-house). En jaune sont indiqués les SNPs communs avec la méthode GATK.

Abbreviations

ßtub	ß-tubulin gene
ADN	Acide désocyribonucléique
ADNc	ADN complémentaire
ARCAD	Agropolis Resource Center for Crop Conservation, Adaptation and Diversity
ARN	Acide ribonucléique
BBMH	Best blast mutual hit
BDS	Réserve naturelle de Bois du Sud (Nouvelle-Calédonie)
bp	paire de base
DAMP	Damage-Associated Molecular Patterns
DHF	Dense Humid Forest
EF1-α	Translation elongation factor 1- α
ETI	Effector triggered immunity
ETS	Effector triggered susceptibility
FAR	Localité de Farino en Nouvelle-Calédonie (pépinière les Hauts de Farino)
FC	Fold Change
GATK	Genome Analysis Tool Kit
gène Avr	gène d'avirulence
gène R	gène de résistance
HR	réaction d'hypersensibilité
ITS	Internal Transcribed Spacer
LRR	Leucine Rich- Repeats
MAMP	Microbe associated molecular pattern
MAP kinase	Mitogen-activated protein kinase
Mb	Mégabase
NBS	Nucleotide Binding Site
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
PGIP	Polygalacturanase inhibiting proteins
*Ppr*1	Puccinia psidii resistance gene 1
PR	Pathogenesis-Related
PRR	Patern recognition receptor
PTI	Plant triggered immunity
QTL	Quantitative Trait Loci
RLK	Receptor-like Kinase
RLP	Receptor-like Protein
RNA-Seq	Séquençage ARN
ROS	Reactive oxygen species
RPKM	Reads Per Kilobase per Million of reads
SAM	Sélection assistée par marqueurs
SNP	Single-nucleotidepolymorphism
TIR	Toll-interleukine receptor

Préambule

Une histoire de la phytopathologie

Les maladies de plantes ont de tout temps façonné les paysages naturels et agricoles, avec parfois des effets retentissant dans l'histoire de l'Humanité. Citons par exemple les textes de l'Ancien Testament rapportant les ravages de la rouille, de la nielle et des sauterelles sur les cultures de blé, ainsi que les famines qui les ont suivies (Premier livre des rois, « prière pour le peuple », verset 37). Un autre exemple tristement célèbre des conséquences des maladies de plantes cultivées, est la famine de 1846-50 due au mildiou de la pomme de terre (*Phytophtora infestans*). Cette famine emporta avec elle un million de vies humaines en Irlande. Cet épisode catastrophique a profondément marqué l'histoire de l'Irlande et celle des Etats-Unis, véritable terre de refuge pour la vague d'Irlandais qui ont fui la famine et la maladie.

La phytopathologie, l'étude des maladies des plantes, est apparue très tôt au sein des anciennes civilisations humaines. Elle est née du besoin de limiter les impacts des maladies affectant les plantes cultivées. Ecrit il y a environ 3 000 ans en Inde, l'ouvrage *Vrikshayurveda* (science de la vie végétale) fait figure de véritable précurseur dans l'étude des maladies de plantes (Ramachanran, 1984). Cet ouvrage propose une classification des maladies de plantes en deux catégories : les maladies internes (dues aux causes environnementales) et les maladies externes (dues aux insectes et aux vers).

Dans l'Antiquité, Théophraste (372-287 av. J.C.) mettait en cause le climat et la position des astres dans son second ouvrage relatif à la botanique (*De causis plantarum*), pour expliquer l'apparition des maladies de plantes. Alors que les phénomènes naturels, les animaux macroscopiques, et les causes divines étaient mis en cause, il aura fallu attendre l'invention du premier microscope en 1676 pour réaliser qu'un monde invisible à l'œil nu pouvait aussi être incriminé dans les maladies de plantes.

Au 18ème siècle, les origines surnaturelles et la théorie de la génération spontanée des maladies de plantes sont réfutées progressivement par les naturalistes (Tozetti 1767 ; Andanson 1759).

Le 19ème siècle voit l'avènement de la phytopathologie en tant que véritable discipline scientifique. En 1807, Isaac-Bénédict Prévost (1755-1819) décrit le premier cycle parasitaire

de l'histoire des sciences en se basant sur la carie du blé, le champignon *Tilletia foetida* (Canabal, 2011) (Rapilly, 2001). En 1861, la phytopathologie se distingue de la mycologie grâce aux travaux d'Anton de Bary (1831-1888) qui décrivit de nombreux champignons pathogènes de plante. Jusqu'au début du 20ème siècle, la phytopathologie était essentiellement descriptive, s'attelant à classer les agents pathogènes selon leur morphologie leur cycle de vie, les symptômes qu'ils induisent chez les plantes (Crous et al., 2015).

Le milieu du 20ème siècle avec ses prodigieuses avancées techniques, permettent des avancées dans tous les domaines scientifiques. Le premier ouvrage de synthèse exposant les principes de phytopathologie fut rédigé par Ernst Albert Gäumann en 1954. Son concept a donné naissance à la phytopathologie moderne, qui se veut une branche multidisciplinaire intégrant des recherches sur les mécanismes d'infection, les relations hôte-parasite, la physiologie, la génétique des pathogènes, les phénomènes de sensibilité ou de résistance, l'épidémiologie et bien d'autres domaines qui forment l'essence même de la phytopathologie moderne.

A l'aube du 21ème siècle, le développement des techniques de séquençage ADN permet à la phytopathologie de résoudre un grand nombre de problématiques. En 2000, le premier génome de plante (*Arabidopsis thaliana*) a été séquencé (Arabidopsis Genome Initiative, 2000), de même que celui du premier agent phytopathogène (la bactérie gamma *Xylella fastidiosa*) (Klosterman et al., 2016). Appliquées de prime abord à des organismes à petits génomes et des espèces modèles, ces technologies s'adaptent de plus en plus aux organismes ayant des génomes plus complexes.

Les recherches scientifiques ont permis durant ces derniers siècles des avancées significatives dans la connaissance des maladies de plantes et dans le développement de moyens de lutte. Cependant, les défis posés par les agents pathogènes sont toujours importants. Nous allons illustrer dans cette thèse comment l'usage des nouvelles technologies de séquençage peut être utilisé dans le cadre d'une stratégie de lutte contre une pathologie forestière et pour des espèces non-domestiquées.

Introduction générale et objectifs de la thèse

La Nouvelle-Calédonie est un archipel français du Pacifique Sud situé à proximité du Tropique du Capricorne (20–23°S, 164–167°E), à environ 1500 km à l'est de l'Australie et à 2000 km au nord de la Nouvelle-Zélande. Elle est constituée d'une île principale, la Grande-Terre, et de plusieurs îles telles que l'île des Pins au sud, l'archipel des Bélep au nord, les îles Loyauté (Ouvéa, Maré, Lifou et Tiga) à l'est (Figure 1). La superficie totale en terres émergées représente 19 100 km^2.

La Grande-Terre, orientée Nord-Ouest/Sud-Est, s'étend sur 450 km de long et de 40 à 70 km de large. Elle est traversée par une chaîne de montagnes centrale et continue, dont le plus haut sommet culmine à 1628 m (Mont Panié). La Nouvelle-Calédonie est unique à bien des égards et notamment par la richesse et l'originalité de ses formations végétales (Figure 17) qui totalisent environ 3 370 espèces de plantes vasculaires dont 74,4 % sont endémiques (Morat et al., 2012).

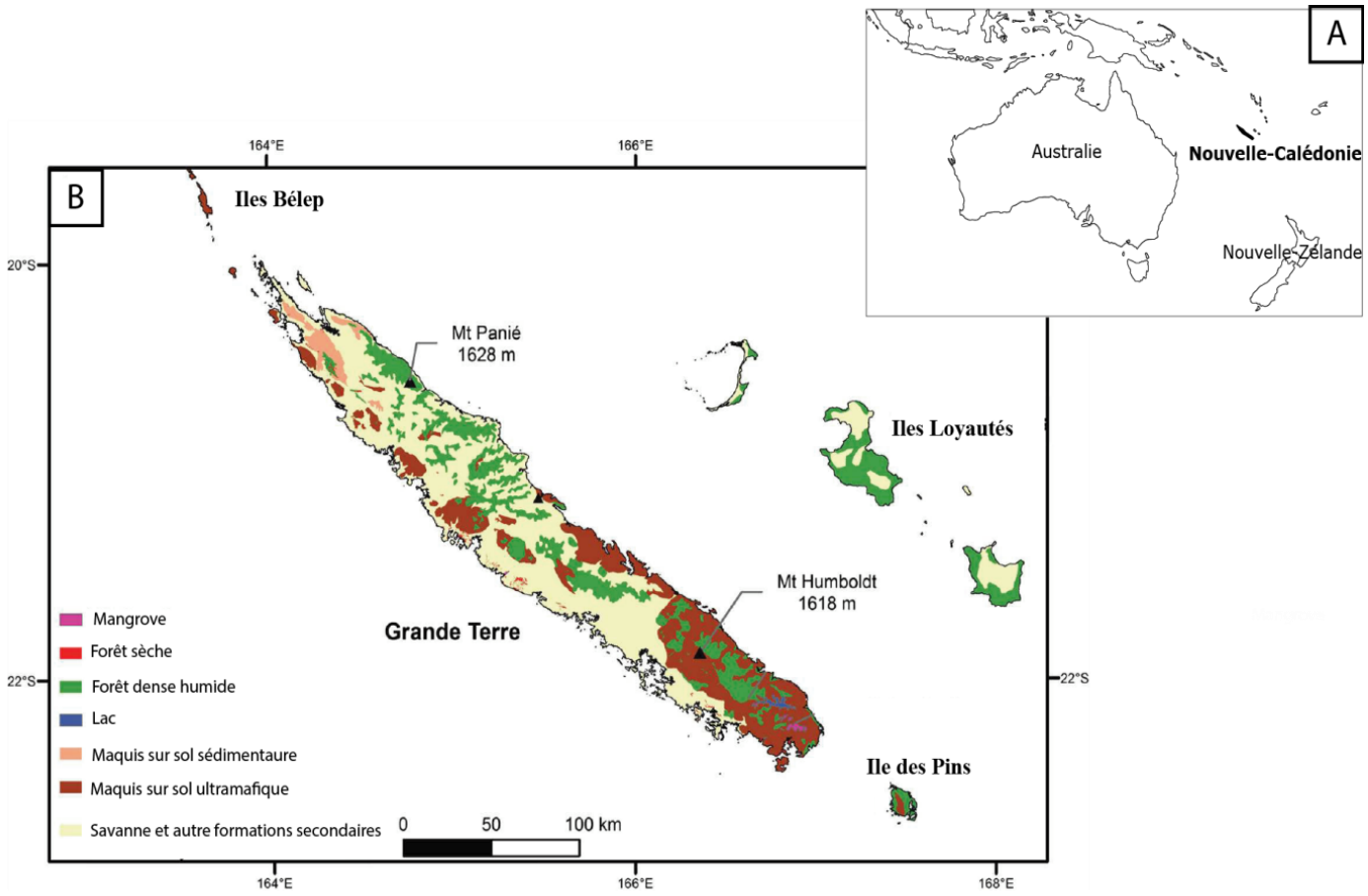


Figure 1 : Localisation de la Nouvelle-Calédonie dans le sud-ouest de l'océan Pacifique (A) et distribution des principaux types de végétation (B). Adapté de Isnard et al. (2016).

En tenant compte de sa surface, la Nouvelle-Calédonie figure au premier rang mondial pour l'endémisme de sa flore (Kier et al., 2009). Cette richesse floristique est attribuée à la combinaison de plusieurs facteurs : l'insularité du territoire, la diversité climatique, géologique et topographique (Isnard et al., 2016). De nombreuses menaces pèsent sur la biodiversité végétale néo-calédonienne, parmi lesquelles figurent : (i) la destruction des habitats par l'exploitation minière, (ii) les feux, (iii) l'urbanisation, et (iv) les invasions biologiques (Wulff et al., 2013). La Nouvelle-Calédonie est considérée comme un "hotspot" de biodiversité mondiale. Selon la définition de Myers et al. (2000), un hotspot est une zone présentant à la fois de fortes concentrations en espèces endémiques et des menaces justifiant la mise en place de mesures prioritaires pour la conservation de la biodiversité terrestre. Selon la base de données globale des espèces invasives (http://www.iucngisd.org/gisd/), 196 espèces invasives ont été recensés dans l'archipel calédonien, en majorité des espèces vertébrées, invertébrées et des espèces végétales. Actuellement peu de connaissances existent sur les dommages occasionnés sur la biodiversité par des espèces invasives de type pathogènes.

La présence d'*A. psidii* a été détectée en Nouvelle-Calédonie au mois d'Avril 2013 au sein de la commune de Farino et sur l'espèce introduite *Syzygium jambos* (Giblin, 2013 ; Nakamura, 2013). De nombreux autres cas de rouille des Myrtaceae ont ensuite été recensés dans le milieu naturel et les pépinières des communes avoisinantes et éloignées. L'établissement de ce phytopathogène exotique est préoccupant compte tenu de l'importance environnementale occupée par les Myrtaceae dans les écosystèmes de l'archipel. En effet, les Myrtaceae constituent la plus grande des familles de plantes à fleurs de Nouvelle-Calédonie (env. 250 espèces). Avec un taux d'endémisme avoisinant les 99%, elles participent de façon importante à la richesse floristique du territoire (Morat, 1993 ; Snow, 2009). Les Myrtaceae sont très présentes dans les écosystèmes néo-calédoniens (en nombre d'espèces, et d'individus) et sont considérées comme structurantes pour la plupart des habitats. A titre d'exemple, sur substrat ultramafique, 180 espèces végétales autochtones sur 250 sont des Myrtaceae (L'Huillier et al., 2010). Elles remplissent de nombreux services écosystémiques, par exemple comme source de nectar et de fruits pour la faune locale (Donovan et al., 2013 ; Fleming et al., 2009). Ces différentes caractéristiques leur valent le statut d'espèces protégées au regard des différents Codes de L'Environnement établis par les Provinces de Nouvelle-Calédonie. Selon les articles 232-2, 232-3 du Code de l'Environnement de la Province Sud de la Nouvelle-Calédonie, la présence d'espèces appartenant à la famille des Myrtaceae (notamment *Arillastrum gummiferum*, *Piliocalyx laurifolius*, *Syzygium* spp., *Eugenia* spp…) dans les forêts denses

humides sempervirentes et les forêts sclérophylles oblige à protéger ces dernières par de la réglementation. L'article 251-1 du Code de l'Environnement de la Province Nord établit également une liste des espèces protégées comprenant de nombreuses Myrtaceae.

D'un point de vue économique, l'arrivée de la rouille sur le territoire néo-calédonien représente une menace pour plusieurs filières dépendantes de l'exploitation des Myrtaceae, en premier lieu celle des pépiniéristes. En effet, les Myrtaceae, par leurs atouts écologiques, comptent parmi les espèces les plus utilisées en revégétalisation des sites dégradés par l'activité minière (Nakamura, 2013). Elles sont également utilisées en tant que plantes ornementales pour les aménagements paysagers publics, ainsi que pour les jardins de particuliers.

Les impacts potentiels de la rouille des Myrtaceae peuvent aussi prendre une dimension socio-culturelle. Certaines espèces de Myrtaceae sont bien connues des néo-calédoniens et constituent une véritable valeur patrimoniale, traditionnelle et parfois même médicinale. Citons par exemple le Chêne-gomme (*Arillastrum gummiferum*), particulièrement apprécié pour ses qualités de bois imputrescible et insensible aux thermites. Il a été jadis utilisé comme bois de construction (poteaux, pont, charpente, quilles de bateaux) (Suprin, 2013). La Fausse bruyère commune (*Sannantha leratii*), à qui on confère des vertus diurétiques et calmantes, a quant à elle été utilisée en infusion. Le Niaouli (*Melaleuca quinquenervia*) est probablement la Myrtaceae la plus connue des néo-calédoniens et l'emblème des paysages de la côte Ouest de la Grande-Terre. Le Niaouli est surtout réputé pour les vertus médicinales de ses feuilles. L'essence qui en est extraite peut être utilisée pour soigner les rhumes et les bronchites. L'écorce de Niaouli est également utilisée dans l'aménagement des toits des cases traditionnelles Kanak. Enfin, les fleurs de Niaouli comptent parmi les plus visitées par les abeilles. Le nectar transformé par les abeilles fournit un miel de qualité et réputé pour les apiculteurs de la Grande Terre.

Compte tenu de l'avancée croissante de la rouille sur le territoire, des comités de pilotages ont été mis en place, sous l'égide du Gouvernement de la Nouvelle-Calédonie. Ils ont réuni plusieurs instances politiques et organismes de recherche, afin d'établir un plan de gestion de la rouille. Les premières actions ont été le déploiement de moyens de mise en quarantaine des pépinières infectées et d'éradication du pathogènes (utilisation de fongicide, destruction des plants infectés, etc.). Ces mesures se sont rapidement révélées inefficaces étant donné l'ampleur des zones infectées en milieu naturel et le coût engendré par les traitements fongicides nécessaires sur le long terme en pépinière. La rouille des Myrtaceae pose aujourd'hui de réelles contraintes à la culture des jeunes plants de Myrtaceae en pépinière et

oblige à un abandon de ces espèces dans les programmes de revégétalisation. En accord avec les conclusions des différents comités de pilotage, il est urgent d'établir un état des lieux des impacts d'*A. psidii* en milieu naturel et de mettre en place une stratégie de lutte adaptée au contexte néo-Calédonien.

Cette thèse s'inscrit dans le cadre de la lutte contre *A. psidii* dans un système insulaire particulièrement vulnérable. Nous avons choisi d'investir nos efforts de recherche dans une stratégie permettant la sélection de plants résistants à la rouille pour des espèces de Myrtaceae impactées par le pathogène. Ces plants cultivés en pépinières seraient majoritairement plantés en milieu naturel dans le cadre d'activités de revégétalisation.
Outres les intérêts écologiques et économiques de cette approche, qui ne nécessite pas l'usage de fongicides, celle-ci permettrait également de renforcer sur le long terme les populations d'hôtes naturelles. Par les mécanismes de la reproduction sexuée et des flux de gènes, nous espérons que les futurs descendants de ces individus, portant les caractères de résistance, puissent constituer une barrière naturelle à la propagation de la rouille.
Afin de réaliser cet objectif pratique, nous avons organisé nos travaux de recherche en deux axes principaux énoncés ci-dessous.

I. Les impacts d'*A. psidii* en Nouvelle-Calédonie (Chapitre 2).

La lutte efficace contre un pathogène, quel qu'il soit, nécessite de connaitre au préalable les dynamiques gouvernant son expansion et le potentiel d'hôtes qu'il peut atteindre dans un nouveau territoire. Ainsi le premier axe de recherche vise à établir un état des lieux de l'établissement d'*A. psidii* en Nouvelle-Calédonie, en termes de distribution géographique et d'impacts de la maladie sur les milieux naturels ainsi que sur l'économie du territoire. Les informations recueillies dans ce cadre permettront de cibler plus spécifiquement les activités de surveillance des populations d'espèces hôtes potentiellement en danger (particulièrement les endémiques et micro-endémiques) dans leurs habitats naturels, et d'élaboration d'une stratégie de lutte sur le long terme. L'ensemble de ces aspects sera traité dans le Chapitre 2 de ce mémoire.

II. Mise au point d'outils pour faciliter la sélection de plants résistants à *A. psidii* (Chapitres 3 et 4)

L'identification de marqueurs moléculaires liés aux mécanismes de résistance à *A. psidii*, est un prérequis essentiel pour la mise en place d'une stratégie de sélection de plants résistants chez les Myrtaceae endémiques de Nouvelle-Calédonie. Alors que les précédents programmes de sélection chez l'Eucalyptus se sont centrés sur l'identification de marqueurs liés à un seul gène majeur de résistance (*Ppr*1) issus de combinaisons de génotypes spécifiques, nous tenterons, dans cette étude, une nouvelle approche afin d'identifier d'autres sources de résistance chez plusieurs espèces de Myrtaceae néo-calédoniennes. L'approche choisie fait appel aux techniques de RNA-Seq et à l'étude globale de transcriptomes de Myrtaceae relevant d'individus différentiés quant aux réponses qu'ils expriment face à la rouille.

Dans un premier temps, nous allons conduire en parallèle deux types d'analyse :

- une analyse d'expression différentielle des gènes entre individus sensibles et individus résistants ;
- une recherche des SNPs, marqueurs des caractères impliqués dans la résistance.

Dans un second temps, les résultats de ces deux analyses seront confrontés afin de sélectionner des marqueurs de gènes exprimés, les plus discriminants, et susceptibles d'être transmis à la descendance.

Dans la suite du mémoire, les résultats de ces travaux seront présentés dans les Chapitres 3 et 4.

Chapitre I : Synthèse bibliographique

Au cours de cette synthèse bibliographique, seront successivement présentés les aspects généraux des maladies de plantes, des mécanismes de résistance des plantes et leur utilisation en sélection, ainsi que des informations concernant plus spécifiquement le pathosystème Myrtaceae-Austropuccinia psidii.

1.1. Généralités sur les relations plantes hôtes-parasites

1.1.1. Le concept de maladie chez les plantes

Les végétaux sont l'épine dorsale de la Vie sur Terre : ils participent à la régulation du climat, purifient l'eau, protègent les sols de l'érosion, et représentent une source de nourriture et d'habitats dont dépendent directement la survie de nombreux animaux et l'Humanité en général (http://www.bgci.org/plantconservationday/whyplantsimportant/). A l'instar des êtres humains et des autres animaux, les végétaux sont sujets aux maladies. Une plante est définie comme "malade" lorsque ses fonctions vitales sont perturbées, altérées ou inhibées, soit par un organisme pathogène (cause biotique), soit par des facteurs environnementaux défavorables (cause abiotique) (Agrios, 2005).

Les maladies d'origine abiotique sont généralement dues à des troubles d'ordre physiologique provoqués par différents facteurs : physiques (températures, lumière, …), édaphiques (pH, salinité, ...) ou chimiques (polluants).

Les maladies d'origine biotique sont causées par des organismes vivants, tels que des champignons, des bactéries, des virus, des oomycètes[1], des nématodes et des insectes. Les changements visibles induits par l'une ou l'autre de ces causes sont les symptômes de la maladie : jaunissement des feuilles, pourriture des fruits, nécrose de l'écorce (chancre), flétrissement, etc. Le diagnostic permet alors d'associer certains symptômes à des maladies spécifiques afin de mieux contrôler les facteurs responsables de ces symptômes. Le type de cellule et de tissus affecté permet de déterminer la fonction physiologique qui sera perturbée (Figure 1.1). Par exemple, un feuillage infecté comme dans le cas des rouilles, des moisissures et des mosaïques, entrave la photosynthèse ; une infection des fruits ou des fleurs interfère sur la reproduction, une infection au niveau des racines peut empêcher l'absorption de l'eau et des nutriments indispensables à la survie de la plante.

[1] Oomycètes : branche de pathogènes similaire aux champignons mais qui est plus apparentée aux algues.

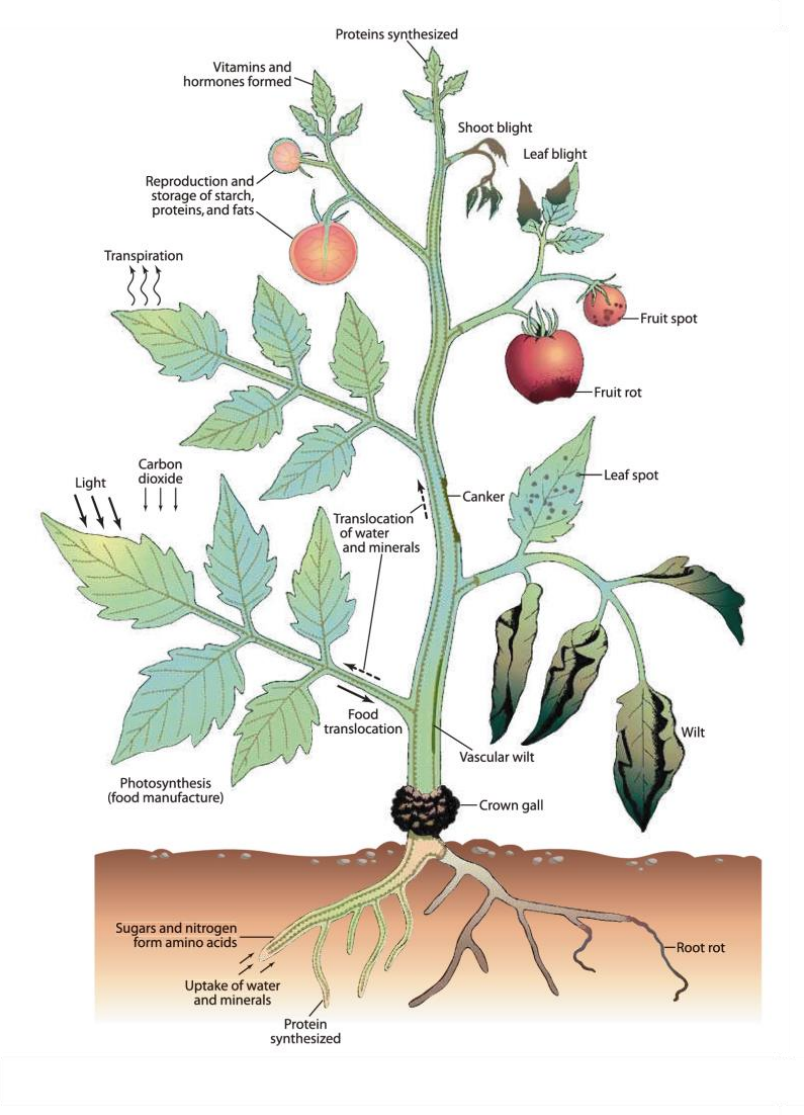


Figure 1.1 : Schéma représentant les principales fonctions d'une plante (gauche) pouvant être affectée par des pathogènes (droite) (extrait d'Agrios, G.N. 2005. Plant Pathology, 5th ed.).

1.1.2. Le triangle d'interaction

Dans le cas des maladies d'origine biotique, le simple contact plante hôte-pathogène peut ne pas être suffisant pour déclencher une maladie. Une troisième composante doit entrer en jeu afin de permettre l'infection : les conditions environnementales (Agrios, 2005). Chacune de ces trois composantes peut considérablement varier, ce qui affecte automatiquement le degré de sévérité de la maladie à l'échelle individuelle dans une population. Par exemple, au sein d'une population d'hôtes la disparité des niveaux d'âge des individus peut favoriser ou non le développement des pathogènes. La taille de la population de pathogènes, peut également jouer un rôle dans l'intensité de la maladie. Enfin, les changements environnementaux peuvent affecter à la fois la croissance des plantes hôtes ainsi que celle des pathogènes. L'interaction entre toutes ces composantes peut se conceptualiser sous forme d'un triangle d'interaction ("triangle de la maladie") (Figure 1.2-A) (Nelson, 1994).

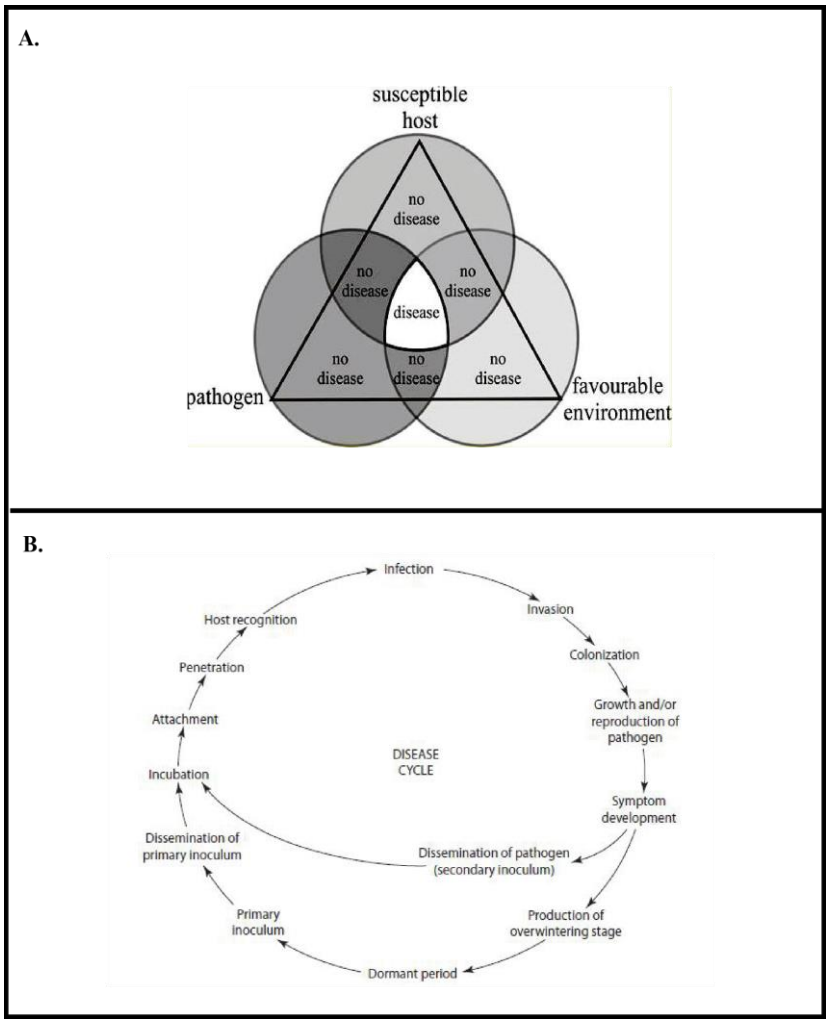


Figure 1.2 : Les deux paradigmes de la phytopathologie, A. Le triangle de maladie et B. Le cycle de maladie**.** Extrait d'Agrios, G.N. 2005. Plant Pathology, 5th ed

1.1.3. Les formes de parasitisme chez les végétaux

Au cours de l'évolution, les microorganismes phytopathogènes ont développé des modes de vie et des stratégies d'infections élaborées pour accéder aux ressources des plantes. Ces modes de vie peuvent être divisés en deux catégories principales : la biotrophie et la nécrotrophie (Rietz and Parker, 2007). Les pathogènes biotrophes (ou obligatoires) ont mis en place des mécanismes afin de détourner les nutriments de la plante hôte pour leur propre alimentation. L' hôte est amputé d'une partie de ses ressources et est maintenu en vie jusqu'à ce que les parasites puissent compléter leur cycle de vie (Mendgen and Hahn, 2002). La survie de tels parasites dépend uniquement de l'hôte et ils ne peuvent être cultivés *in vitro*. Ils concernent des virus, des bactéries (ex : *Pseudomonas*), des champignons (ex : rouille et oidiums), des oomycètes (ex : Albugo), et des nématodes sédentaires.

A l'opposé, les pathogènes nécrotrophes (ou facultatifs) sécrètent des toxines et des enzymes qui vont tuer les cellules hôtes afin d'en extraire les nutriments (Horbach et al., 2011). Ce mode de parasitisme finit irrémédiablement par la mort de l'hôte. Parmi les nécrotrophes se retrouvent des champignons (ex : *Botrytis*, *Pythium*), des oomycètes, des bactéries (*Erwinia*) et des nématodes migrateurs. Entre ces deux extrêmes, une catégorie intermédiaire est celle des hémibiotrophes qui combinent une phase initiale de biotrophe puis une phase nécrotrophe (Horbach et al., 2011). Parmi les pathogènes hémibiotrophes on peut trouver des champignons *(Magnaporthe, Colletotrichum)* et des oomycètes *(Phytophtora).*

1.1.4. Champignons phytopathogènes

A ce jour, sur les plus de 100 000 espèces de champignon connues, plus de 10 000 sont des pathogènes de plantes. Comparativement, une cinquantaine d'espèces de champignons phytopathogènes est connue chez l'Homme et approximativement le même nombre est connu chez les animaux (Agrios, 2005). Les champignons sont la principale cause de maladies chez les plantes et sont responsables d'environ 70 % des maladies des plantes cultivées (Alfen, 2001; Oerke, 2006). Ces maladies dites cryptogamiques sont provoquées par les champignons appartenant aux différentes sous-divisions du règne des Eumycètes (ou champignons vrais) : Ascomycètes, Basidiomycètes, Chytridiomycètes, Zygomycètes et des Deutéromycètes (champignons imparfaits). Les champignons phytopathogènes sont capables d'infecter n'importe quel tissu à n'importe quel stade de croissance de la plante, en suivant un cycle biologique complexe qui peut comporter des stades de reproduction sexuée ou asexuée.

L'infection des plantes par un champignon phytopathogène se déroule selon une série d'évènements successifs conduisant à son développement et son maintien dans la plante hôte. Ce processus est appelé "cycle de la maladie" et inclut le cycle biologique du pathogène ainsi que les effets de la maladie occasionnés sur la plante hôte. La complexité des cycles de la maladie varie selon les espèces, dans le cas des champignons il comprend toujours des étapes obligatoires (inoculation, adhérence, germination, pénétration, colonisation, croissance et reproduction, dissémination, survie et l'absence d'hôte) (Figure 1.2-B).

1.1.5. Les rouilles

1.1.5.1. Taxonomie

REGNE : FUNGI

EMBRANCHEMENT : Basidiomycota

SOUS-EMBRANCHEMENT : Pucciniomycotina

CLASSE : Pucciniomycetes

ORDRE : Pucciniales

La rouille est un terme générique pour caractériser de nombreuses maladies cryptogamiques des végétaux supérieurs (fougères, gymnospermes, et angiospermes). Les responsables de ces maladies sont des champignons appartenant à l'ordre des Pucciniales (Basidiomycètes). Comptant plus de 7 000 espèces réparties dans plus de 166 genres (15 familles), les Pucciniales représentent un tiers des espèces de basidiomycètes connues, et 8% de tous les champignons décrits à ce jour (Aime et al., 2006). Ils constituent le plus grand groupe de champignons phytopathogènes du monde (Tavares et al., 2014), parmi lesquels certains sont responsables d'importants dégâts chez les plantes cultivées (Tableau 1.1, Figure 1.3).

Tableau 1.1 : Exemples de maladies causées par des agents pathogènes appartenant à l'ordre des Pucciniales et leurs hôtes respectifs.

Agent pathogène	**Nom vulgaire**	**Hôte principal**	**Hôte alternatif**	**Type de cycle**	**Références**
Cronartium ribicola	Rouille vésiculeuse du pin blanc	*Pinus strobus*	*Ribes* spp.	macrocyclique	Kinloch, 2003
Hemileia vastatrix	Rouille orangée	*Coffea* spp.	inconnu	macrocyclique	Kolmer *et al*., 2009
Melampsora epitea var. *epitea*	Rouille du saule	*Salix* spp	*Larix decidua*	macrocyclique	Pei, Royle, Hunter & 1996
Phakopsora pachyrhizi	Rouille du soja	*Glycine max*	inconnu	macrocyclique	Goellner *et al*., 2010
Puccinia graminis f. sp. *tritici*	Rouille noire du blé	*Triticum* spp.	*Berberis vulgaris*	macrocyclique	Leonard & Szabo, 2005
Gymnosporangium juniperi virginianae	Rouille du pommier	*Malus* spp.	*Juniperus virginiana*	demicyclique	Zhao *et al*., 2016
Puccinia pittieriana	Rouille de la pomme de terre	*Solanaceae* spp.	aucun	microcyclique	Laundon & Rainbow, 1971

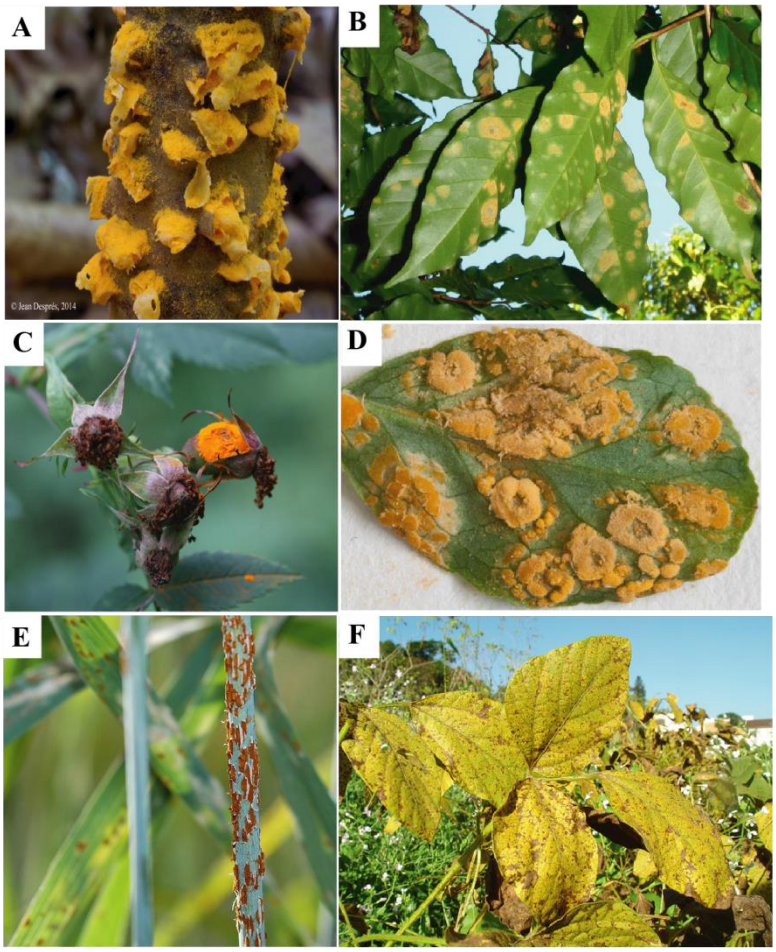


Figure 1.3 : Espèces de rouilles chez les plantes cultivées.

***A** : rouille vésiculeuse du pin blanc (Cronartium ribocola)*
(source: https://www.flickr.com/photos/23521914@N08/14067852639);

***B** : rouille orangée (Hemileia vastatrix)*
(source: http://www.apsnet.org/publications/imageresources/Pages/FI00125.aspx);

***C** : rouille du rosier (Phragmidium mucronatum)*
(source :http://www.naturespot.org.uk/species/rose-rust);

***D** : rouille du saule (Melampsora epitea var. epitea)*
(source: http://www.hampshirefungi.org.uk/progm.php?id=HF1503&ph=664#imageplace);

***E** : rouille noire du blé (Puccinia graminis f. sp. tritici)*
(source : https://en.wikipedia.org/wiki/Stem_rust);

***F** : rouille du soja (Phakopsora pachyrhizi)*
(source: http://www.apsnet.org/publications/imageresources/Pages/IW000048.aspx)

1.1.5.2. Cycle de vie

Tous les Pucciniales sont des parasites obligatoires des plantes (ou biotrophes) dont ils obtiennent des nutriments, sur lesquels ils se reproduisent et complètent leurs cycles de vie. Le cycle de vie des Pucciniales est connu pour être l'un des plus complexes du règne des champignons (Cummins and Hiratsuka 2003). Il peut comporter jusqu'à 5 stades de spores différents : téliospores, basidiospores, pycniospores, aeciospores et urédospores (Figure 1.4) avec des variations dans la composition nucléaire de ces spores (Aime, 2006). Le cycle de vie présente deux phases parasitiques : la phase asexuée dicaryotique (aeciospore, urédospore et téliospore) et la phase sexuée monocaryotique (basidiospore et pycniospores). Les rouilles comportant ces 5 stades de spores ont un cycle qualifié de macrocyclique. D'autres espèces de rouille peuvent compléter leur cycle en se passant de certains types de spores, c'est le cas des rouilles à cycle démicyclique (stade urédospore manquant) ou à cycle microcyclique (stades basidiospore, pycniospore et aeciospores manquant) (*Plant Relationships*, p.74). Les rouilles sont dites autoïques lorsqu'elles accomplissent la totalité de leur cycle sur le même hôte. Cependant, et selon les espèces, l'intervention d'hôtes alternatifs est obligatoire afin de boucler leur cycle biologique. C'est ainsi qu'une rouille, alternant sur au moins 2 espèces d'hôtes distincts, est qualifiée de dioïque. Les hôtes intervenant dans le cycle d'une rouille dioïque ne sont en général pas des espèces apparentées. Les sores et les spores vont du jaune-orangé au noir. Ces couleurs caractéristiques rappellent celles produites par la corrosion des métaux ferreux, et sont à l'origine de la dénomination de ces maladies (rouilles).

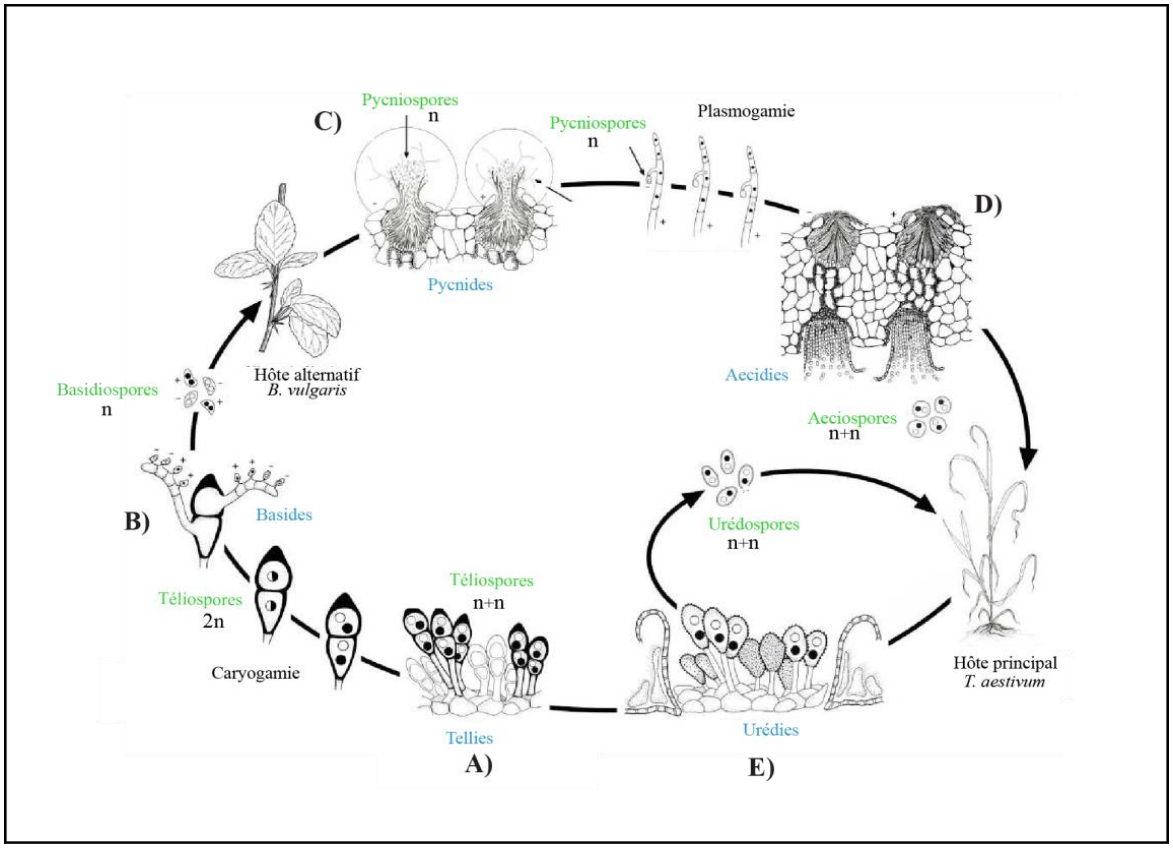


Figure 1.4 : Représentation schématique du cycle d'une rouille macrocyclique et dioïque *Puccinia graminis* f. sp. *tritici* dont l'hôte principal est le blé (*Triticum aestivum*) et l'hôte alternatif est l'épinette verte (*Berberis vulgaris*).
(A) Télies et téliospores, (B) Basides et basidiospores, (C) Pycnides et pycniospores (ou spermogonies et spermaties), (D) Aecidies et aeciospores, (E) Urédies et urédospores. Le niveau de ploidie est indiqué par la lettre « n » (n = haploïde ; n+n=dicaryon ; 2n=diploïde). Adapté de Leonard and Szabo, 2005.

1.1.5.3. Mode d'infection des rouilles

Les champignons biotrophes, tels que les rouilles, dépendent entièrement des cellules vivantes de leur hôte végétal pour leur croissance et leur reproduction (Schulze-Lefert and Panstruga, 2003). Selon la définition de Mendgen and Hahn (2002), les champignons biotrophes obligatoires ont en commun les caractéristiques suivantes :

(i) des structures d'infection hautement développées ;

(ii) une activité de sécrétion limitée, en particulier d'enzymes lytiques ;

(iii) la zone de contact entre les membranes plasmiques du champignon et de la cellule hôte est étroite et riche en glucides et en protéines ;

(iv) la formation d'haustoria, qui sont des hyphes spécialisées pour l'absorption des nutriments et le métabolisme ;

(v) la suppression à long terme des défenses de l'hôte et l'induction de gènes hôtes spécifiques pour l'établissement de la biotrophie.

Les premières étapes du cycle d'infection d'une rouille commencent par l'adhésion d'une spore infectieuse à la surface de la plante hôte (Tucker and Talbot, 2001) (Figure 1.5-a). Dans des conditions climatiques favorables, cette spore germe pour produire un tube germinatif qui évoluera vers une structure spécialisée dans la pénétration de l'hôte : l'appressorium (Leonard and Szabo, 2005). Cet appressorium émet ensuite un hyphe infectieux primaire qui perce la cuticule de la feuille et colonise les cellules adjacentes. Les phases d'infection de la rouille conditionnent le mode de pénétration dans les tissus de la plante hôte. Lors d'une infection de rouille en phase dicaryotique (e.g. urédospores), la pénétration est dite indirecte et s'effectue via les stomates. Néanmoins certaines exceptions de pénétration directe à travers les cellules épidermiques de l'hôte ont été observées chez des rouilles dicaryotiques comme *Austropuccinia psidii* sur *Syzygium jambos* (Hunt, 1968); *Ravenelia humphreyana* sur *Caesalpinia pulcherrima* (Hunt, 1968) ; *Physopella zeae* sur *Zea mays* (Bonde et al., 1982) ; *Phakopsora pachyrhizi* sur *Glycine max* (Bonde, 1976) et *Phakopsora apoda* sur *Pennisetum clandestinum* (Adendorff, 1998).

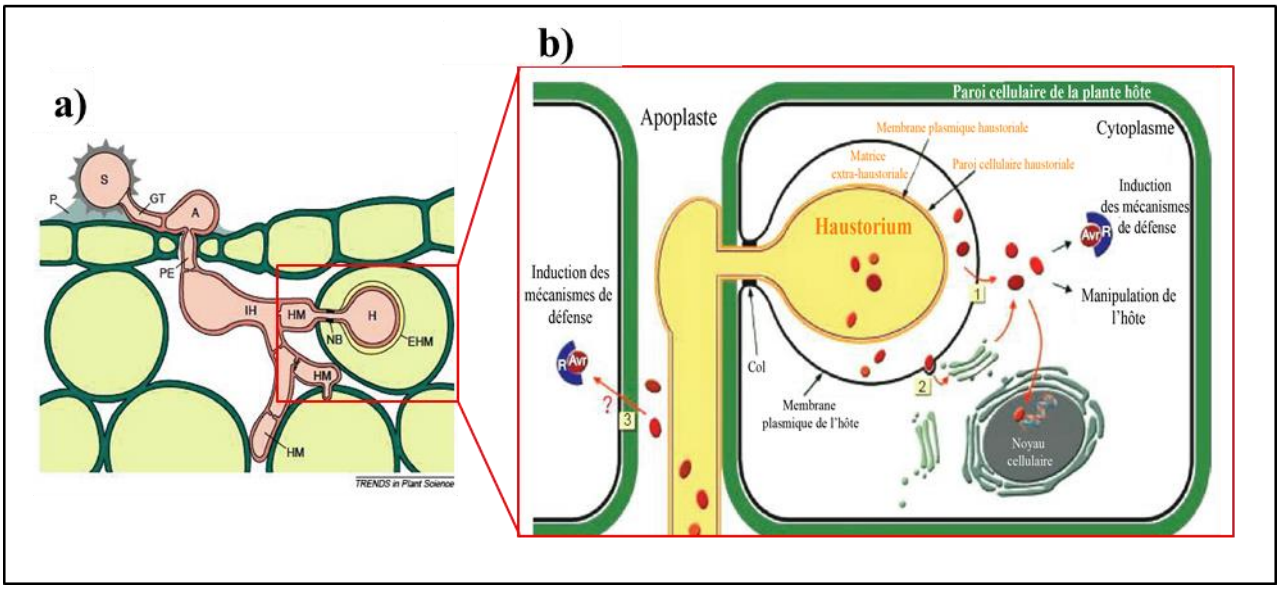


Figure 1.5 : Processus d'infection d'une rouille
a) *Processus d'infection de la rouille de la fève, Uromyces fabae (dycariotique). S : urédospore ; P : surface adhésive de l'hôte ; GT : tube germinatif, A : appressorium ; PE : hyphe de pénétration ; IH : Hyphe d'infection ; HM : cellule mère haustoriale ; H : Haustorium ; EHM : membrane extra-haustoriale ; NM : col. Extrait de Mendgen and Hahn, 2002.*
b) *Schéma d'un haustorium de rouille dans une cellule hôte. Adapté d'après Catanzariti et al., 2007.*

Après pénétration de l'hyphe d'infection, le champignon forme des cellules spécialisées appelées cellules mères d'haustorium (HM, haustorial mother cell) au contact des cellules du mésophylle[2]. Ces cellules sont à l'origine de la formation de l'haustorium (Figure 1.5-b). L'haustorium se développe à l'intérieur de la cellule hôte en invaginant la membrane plasmique de l'hôte (membrane extra-haustoriale). Entre les membranes plasmiques de l'hôte et du pathogène, se forme une interface de communication riche en glucides et en protéines, la matrice extra-haustoriale, (Struck et al., 2004 ; Voegele and Mendgen, 2003). L'haustoria détourne les nutriments de l'hôte pour son propre développement et sécrète des facteurs de virulences appelées effecteurs et impliquées dans la manipulation de l'hôte (Mendgen et al., 2000 ; Voegele and Mendgen, 2003). Ces effecteurs sont des protéines qui ciblent le noyau de la cellule hôte pour y reprogrammer la transcription, dégrader des protéines, détourner les voies métaboliques ou supprimer les voies de défense.

[2] Mésophylle : partie interne de la feuille constituée du parenchyme chlorophyllien et responsable de la photosynthèse.

1.1.5.4. Gamme d'hôtes

La gamme hôte d'une rouille est définie par les espèces végétales qu'elle peut infecter et sur laquelle elle peut compléter son cycle de vie (Schulze-Lefert and Panstruga, 2003). Historiquement, les rouilles ont souvent été définies comme ayant une gamme hôte restreinte et confinée à quelques espèces végétales réparties dans un ou deux genres (*Puccinia hordei, Puccinia sorghi, Puccinia kuehnii*) (Kolmer et al., 2009). Ce haut degré de spécificité dans l'association d'une rouille avec un hôte spécifique a d'ailleurs amené la création d'un niveau taxonomique en dessous de celui d'espèce : forma specialis[3] (f. sp.) (Anikster, 1984). Cependant, de plus en plus d'exemples tendent à prouver que les gammes hôtes chez les rouilles ne restent pas fixes et des expansions spectaculaires de gammes hôtes ont été observés chez certaines espèces : citons par exemple *Puccinia graminis*, connue pour infecter 365 espèces de plantes dans 54 genres, et *Puccinia coronata* qui s'attaque à 290 hôtes appartenant à 72 genres (Bettgenhaeuser et al., 2014).

Les mécanismes conduisant à la diversité de gamme d'hôtes chez les rouilles sont encore méconnus. Les pathogènes évoluent en général en développant des capacités leur permettant d'infecter de nouvelles espèces hôtes qui sont soit proches phylogénétiquement (extension de la gamme hôte), soit distantes de leur hôte de base (″host jump″) (Paul Schulze-Lefert and Panstruga, 2003).

[3] *Forma specialis* : niveau taxonomique en dessous de celui de l'espèce et permettant de classifier les pathogènes (en particulier les champignons).

1.2. Le système immunitaire des végétaux

Malgré leur apparente passivité et leur immobilité, les végétaux sont dotés d'un système immunitaire élaboré, formé d'un éventail de barrières de défense constitutives et inductibles. Leurs mécanismes de défense reposent sur la perception du "non-soi" et le déclenchement d'une cascade d'évènements pour neutraliser et éliminer les pathogènes (Sanabria et al., 2010).

1.2.1. Les défenses préformées ou constitutives

La première ligne de défense des plantes repose sur une résistance préformée dont le rôle est d'empêcher, retarder ou limiter l'invasion par un agent pathogène. Il s'agit en particulier de l'ensemble des barrières physiques de la cellule végétale telle que la cuticule et la paroi cellulaire (Reina-Pinto and Yephremov, 2009) (Figure 1.6).

La cuticule est une structure lipidique composée principalement de cutine et de cires, et recouvrant l'ensemble de la surface des feuilles (Serrano et al., 2014). La paroi pectocellulosique forme quant-à-elle un exosquelette autour de la cellule végétale. Elle contient un réseau dense et complexe de lignine, cellulose, hémicellulose et pectines conférant la rigidité aux cellules (Hématy et al., 2009). À ces barrières physiques s'ajoutent parfois des barrières chimiques faisant intervenir des composés ayant des propriétés antimicrobiennes, des métabolites secondaires : terpénoïdes, phénols et alcaloïdes (Bennett and Wallsgrove, 1994). Ces barrières passives préexistantes sont généralement efficaces contre la plupart des agents pathogènes. Cependant, de nombreux pathogènes produisent des enzymes lytiques visant à dégrader les barrières physiques de la plante et d'accèder aux ressources de la cellule. Parmi ces enzymes se trouvent les exo- ou endopolygalacturonases (dégradant la pectine), les endoglucanases (dégradant la cellulose), les cutinases (dégradant la cuticule), les xylanases (dégradant l'hémicellulose), etc. (Deising, 1992; Juge, 2006 ; Zhang et al., 2014).

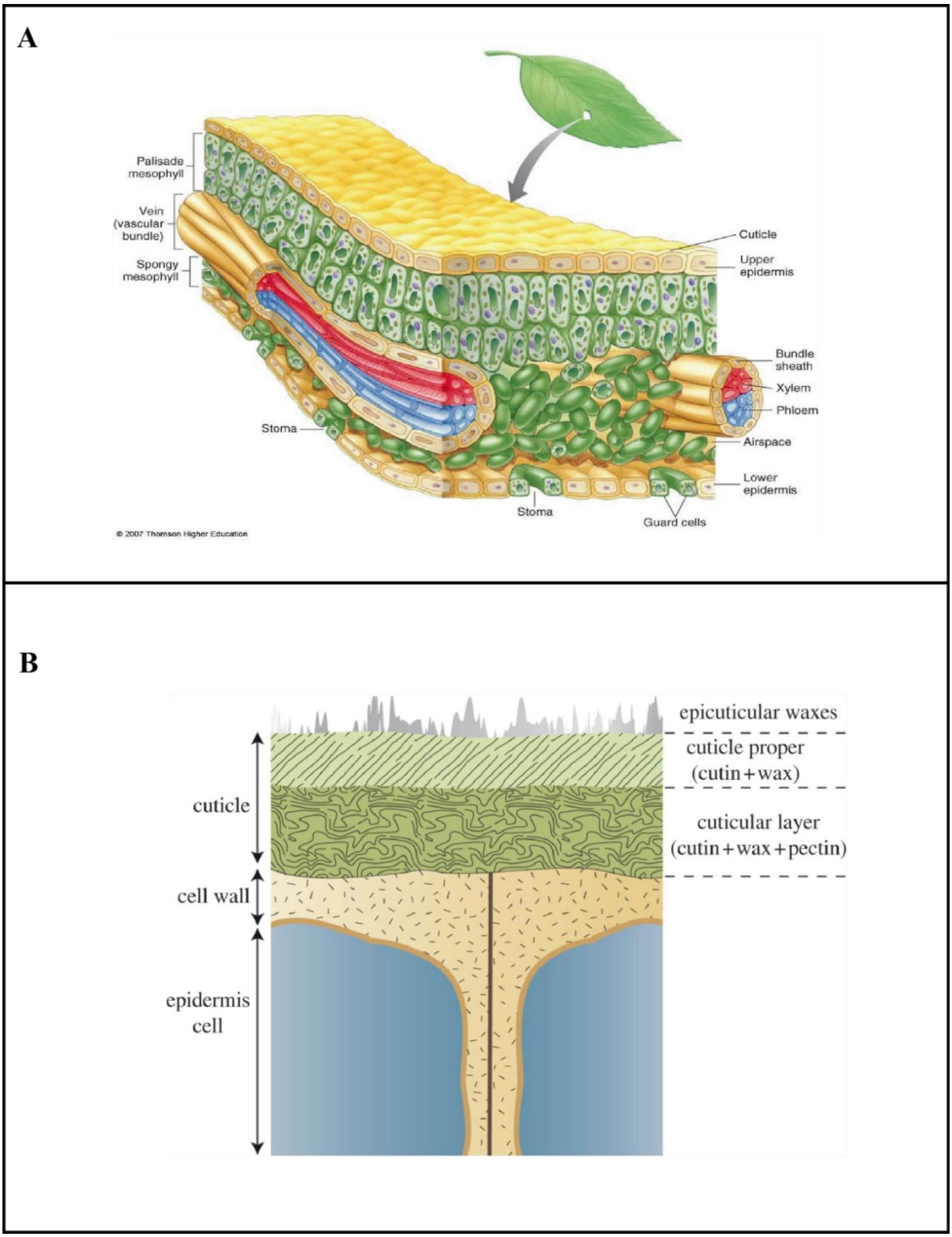


Figure 1.6 : Illustrations des barrières physiques des plantes contre l'entrée des pathogènes.
A : Anatomie d'une feuille de dicotylédone et B : Organisation de la cuticule chez les plantes.

1.2.2. Les défenses inductibles

Même si les défenses préformées permettent de limiter l'entrée dans la plante de la plupart des agents pathogènes, certains parviennent à les contourner. Ils se confrontent alors aux défenses inductibles de la plante, un type de système immunitaire inné à deux ramifications (Jones and Dangl, 2006) :

- La première branche reconnaît et répond à des molécules communes à de nombreuses catégories de microorganismes (éliciteurs généraux), y compris les non-pathogènes : **la résistance basale (PTI).**
- Le deuxième répond soit directement par la reconnaissance de facteurs de virulence spécifiques aux pathogènes, soit indirectement par leurs effets sur l'hôte : **la résistance spécifique (ETI).**

Ces différentes composantes du système de défense des plantes peuvent à leur tour être contournées par les pathogènes. Les différentes issues de l'interaction plante-pathogène peuvent être représentées sous la forme d'un modèle en 5 phases dit en Zig-Zag (Jones and Dangl, 2006) (Figure 1.7). Ce modèle permet de représenter de manière schématique l'évolution simultanée des protéines végétales et des protéines pathogènes, au cours du temps, ainsi que leur conséquence pour la plante : résistance quand les défenses de la plantes sont efficaces ou maladie quand les pathogènes surmontent ces défenses (Jones and Dangl, 2006).
De manière schématique, trois étapes successives sont nécessaires pour aboutir au déclenchement des défenses inductibles : une reconnaissance de la présence de l'agent pathogène par la plante, suivie d'une transduction du signal, qui aboutit à l'activation des gènes de défense (Jones and Dangl, 2006).

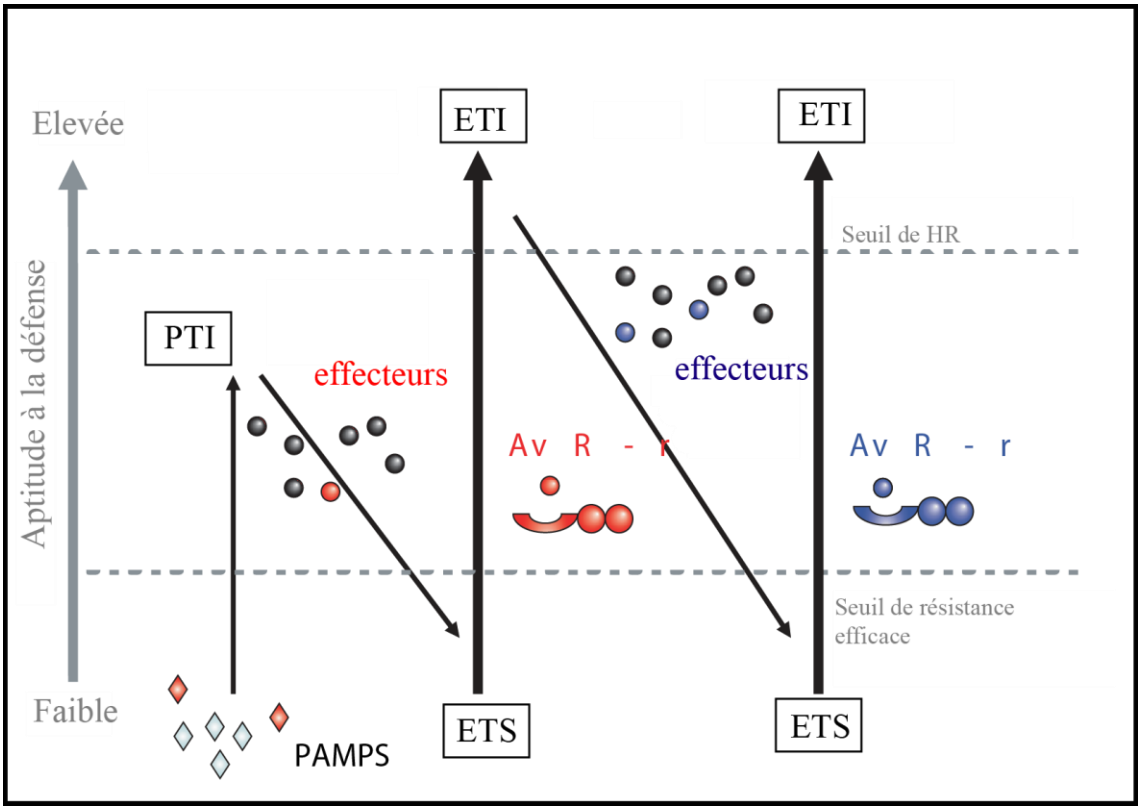


Figure 1.7 : Schéma du modèle en Zig-Zag illustrant le système immunitaire inné des plantes et l'évolution simultanée des mécanismes de défense des plantes et d'attaque des agents pathogènes. D'après Jones and Dangl, 2006.

Phase 1 : *les plantes détectent les MAMPs/PAMPs (Microbial-Pathogen-associated molecular patterns) via les PRRs (Pattern-recognition Receptors) et déclenchent la PTI (Pamp-Triggered Immunity).*

Phase 2 : *les pathogènes sécrètent des effecteurs pour interférer avec la PTI, permettre sa nutrition et sa dispersion. Les plantes sont sensibles (ETS- Effector Triggered Susceptibility).*

Phase 3 : *Un effecteur du pathogène est reconnu par les protéines NB-LRR des plantes, ce qui active l'ETI-Effector-triggered immunity. Cette version amplifiée de la PTI passe souvent un seuil permettant l'induction de la mort cellulaire programmée (HR-Hypersensitive Response).*

Phase 4 : *Un isolat du pathogène, ayant perdu les effecteurs reconnus par les plantes et potentiellement gagné de nouveaux effecteurs, est sélectionné. Cet isolat permet au pathogène de contourner l'ETI.*

Phase 5 : *En réponse, la sélection favorise chez les plantes de nouveaux allèles chez les NB-LRR leur permettant de reconnaitre les effecteurs nouvellement acquis, résultant en un nouveau déclenchement de l'ETI.*

1.2.2.1. Plant Triggered Immunity (PTI) : la résistance basale ou non-spécifique

La PTI est induite lors des phases initiales de l'infection. Ce premier niveau de défense est peu spécifique et peu intense mais permet de contenir de nombreux pathogènes. Elle est déclenchée par des récepteurs membranaires chez l'hôte (PRR – pour "Pattern Recognition Receptors"), capables de reconnaître des signaux moléculaires associés aux microorganismes (MAMPs) ou éliciteurs généraux (Postel and Kemmerling, 2009). La perception de la présence d'agents pathogènes par les PRR déclenche une cascade de signalisation permettant l'induction des mécanismes de défenses : modification du flux ionique (Ca_2^+), cascade de signalisation des MAP kinases[4], renforcement des parois cellulaires, synthèse d'espèces réactives à l'oxygènes (ROS)[5] (Jones and Dangl, 2006). On regroupe également sous ce type de résistance, la résistance non-hôte qui se définit comme la résistance d'une espèce végétale envers tous les variants génétiques d'une espèce pathogène non adaptée pour l'infecter (Lee et al., 2016).

- **Les éliciteurs généraux de la PTI**

La résistance basale de la PTI repose sur la reconnaissance d'éliciteurs généraux, des molécules capables d'activer les mécanismes de défense chez les plantes (Thakur and Sohal, 2013). Ils sont classés en deux types en fonction de leur origine :

- Les éliciteurs exogènes : d'origines microbiennes et appelés MAMP (« Microbial-Associated Molecular Pattern »). Les MAMPs sont des structures ou des composants très conservés chez plusieurs classes de microorganismes, et souvent indispensables à leur survie (Newman et al., 2013). Ils constituent de véritable signaux d'alerte qui ont

[4] **MAP kinases** : protéines présentes chez tous les eucaryotes et qui catalysent la phosphorylation des protéines MAP (Mitogen Activated Protein). Elles ont pour rôle la transduction et l'amplification des signaux extracellulaires (Innes, 2001). Chez les plantes, elles régulent de nombreux processus métaboliques, en particulier l'activation des mécanismes de défense suite à la perception des pathogènes par la PTI ou l'ETI (Nakagami, Pitzschke, & Hirt, 2005).

[5] **Espèces réactives à l'oxygène** : composés chimiques dérivés de l'oxygène comprenant le peroxyde d'hydrogène (H2O2), l'anion superoxyde (O2-) et le radical hydroxyle (OH). En plus de leur effet direct cytotoxiques, les ROS agissent également comme seconds messagers dans diverses voies de signalisation (Stael et al., 2015) incluant : (i) le renforcement de la paroi de la cellule végétale, (ii) l'induction des gènes de défense (PR-pathogenesis-related), (iii) la synthèse de phytoalexines, (iv) l'activation des cascades de MAP kinases, et (v) l'induction de la mort cellulaire de type HR (Réaction d'hypersensibilité) (Hammond-Kosack & Jones, 1997 ; Low & Merida, 1996).

été naturellement sélectionnés au cours de l'évolution par les plantes (Chisholm et al., 2006). Chez les champignons, les MAMPs comprennent les constituants des parois cellulaires, tels que la chitine et l'ergostérol (Nurnberger et al., 2004).

- Les éliciteurs endogènes : aussi appelés DAMPs (« Damage-Associated Molecular Patterns »), molécules d'origine végétale libérées suite aux dommages cellulaires provoqués par l'action d'un microorganisme ou d'un stress environnemental (Choi and Klessig, 2016).

- **<u>Les PRR- Pattern Recognition Receptors</u>**

Chez les végétaux, les éliciteurs généraux sont reconnus par des récepteurs membranaires appelés PRR (« Pattern-Recognition Receptors »). Les PRR sont considérés comme des récepteurs relativement anciens car ils sont retrouvés dans de nombreuses familles végétales. Les PRR font partie de la famille des protéines de résistance (R) composée d'un domaine extracellulaire de type LRR (« Leucin-Rich-Repeats »), un domaine transmembranaire et un domaine cytoplasmique kinase.

Les PRR peuvent être regroupés en trois catégories de récepteurs (Figure 1.8) :

- Les RLK (Receptor-like Kinase) : ils sont ancrés dans la membrane plasmique et comprennent : un domaine extracellulaire (par exemple de type LRR-Leucine Rich Repeats, LysM-Lectine Lysine Motif, WAK-Wall Associated Kinases, etc), un domaine transmembranaire et un domaine kinase cytoplasmique (Shiu & Bleecker, 2001).

- Les RLP (Receptor-like Protein) : ils possèdent un domaine extracellulaire de type LRR et un domaine transmembranaire, mais sont dépourvus de domaine kinase cytoplasmique (Tang, Wang, & Zhou, 2017 ; Tor, Lotze, & Holton, 2009).

- Les PGIP (Polygalacturanase inhibiting proteins) : ils possèdent seulement un domaine extracellulaire LRR (Di Matteo, Bonivento, Tsernoglou, Federici, & Cervone, 2006).

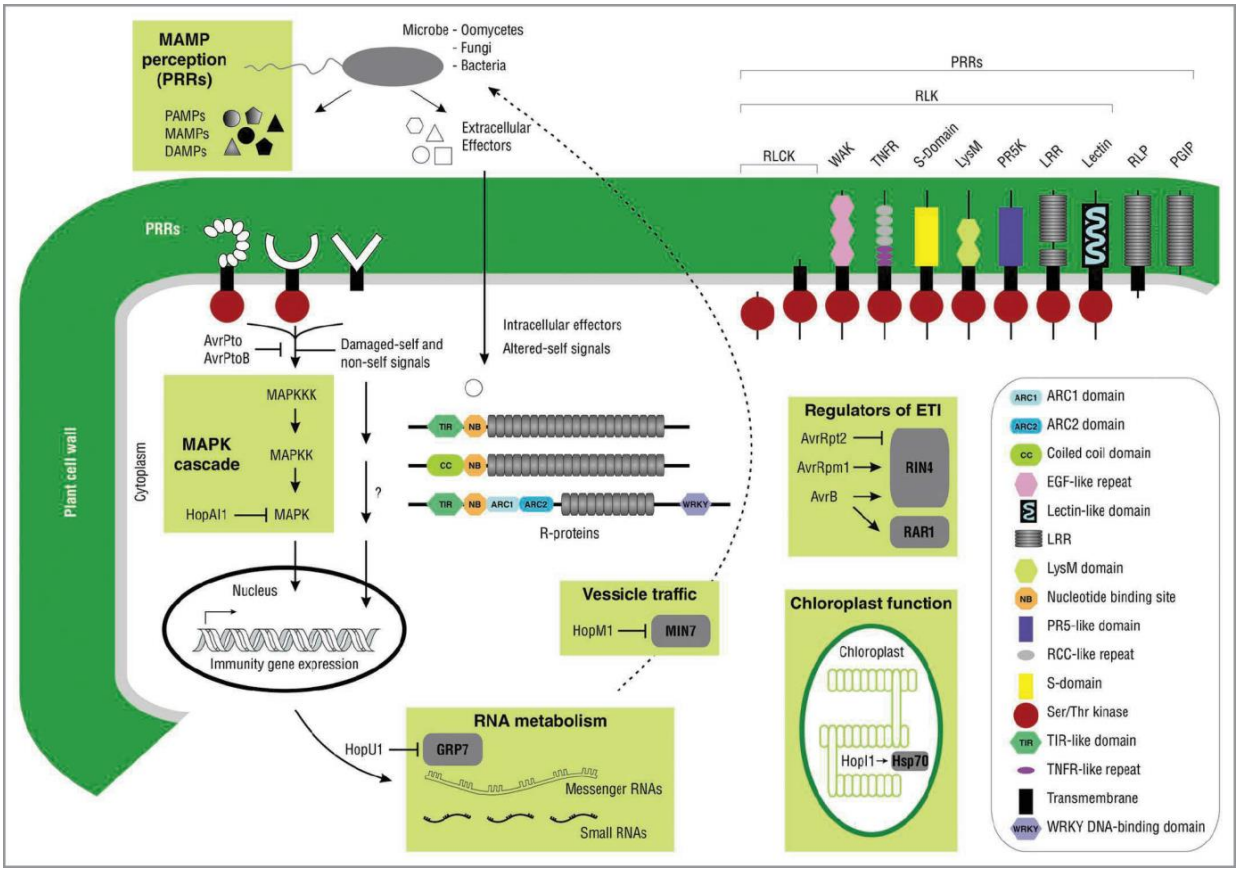


Figure 1.8 : Systèmes de perception des MAMPs/DAMPs et des effecteurs dérivés des agents pathogènes chez les plantes. D'après Sanabria, Huang, & Dubery, 2010.

Abréviations: ***MAMP****, Microbe-Associated Pattern ;* ***PRR****, Pattern recognition receptor ;* ***PGIP****, polygalacturonase inhibitor protein ;* ***RLP****, receptor-like protein ;* ***RLK****, receptor-like kinase ;* ***LRR****, leucine-rich repeats ;* ***LysM****, lysine motif ;* ***PR5K****, pathogenesis related 5-like receptor kinase ;* ***RLCK****, receptor-like cytoplasmic kinase ;* ***S-domain****, self-incompatibility domain ;* ***TNFR****, tumour necrosis factorreceptor ;* ***WAK****, wall associated kinase ;* ***EGF****, epidermal growth factor ;* ***RCC****, regulator of chromosome condensation ;* ***Ser/Thr****, serine/threonine* ***NB****, nucleotide binding ;* ***TIR****, Toll and Interleukin 1 transmembrane receptor ;* ***CC****, coiled-coil ;* ***ARC****, Activity-regulated cytoskeleton-associated protein.*

1.2.2.2. Effector Triggered Susceptibility (ETS) : la désactivation de la résistance basale

Pour contourner la PTI, la plupart des agents pathogènes sont capables de sécréter des effecteurs de virulence (Catanzariti et al., 2007 ; Jones and Dangl, 2006). Les effecteurs de virulence sont de petites protéines (≤ 300 acides aminées), capables de supprimer les réponses de défense de l'hôte et/ou de modifier sa physiologie afin de faciliter l'entrée du pathogène (Duplessis et al., 2011 ; Selin et al., 2016). Cet évènement est aussi connu sous le nom de

ETS-Effector Triggered Susceptibility et aboutira à la sensibilité des plantes hôtes qui seront dans l'incapacité de détecter ces effecteurs de virulences (Bigeard et al., 2015). En d'autres termes, les effecteurs représentent les déterminants clés du pouvoir pathogène afin de contourner les mécanismes de reconnaissance de la PTI. Les effecteurs peuvent être catégorisés en deux groupes suivant la localisation de leurs cibles chez l'hôte (Chaudhari et al., 2014) :

- **Les effecteurs apoplasmiques** demeurent à l'extérieur de la cellule et interagissent avec des cibles extracellulaires ou des récepteurs de surface (Chaudhari et al., 2014). Chez les champignons, ils peuvent être sécrétés par l'appressorium et/ou les hyphes des champignons dans l'espace intercellulaire de la plante hôte.
- **Les effecteurs cytoplasmiques** agissent dans le cytoplasme des cellules et ciblent des protéines intervenant dans les cascades de signalisation du système de défense des plantes. Ces protéines jouent un rôle essentiel dans la virulence des microorganismes en établissant les conditions propices au développement de l'agent pathogène (Houterman et al., 2008).
- **Les effecteurs nucléaires** sont dotés d'un Signal de Localisation Nucléaire (NLS) leur permettant d'être transloqués du cytoplasme vers le noyau de la plante hôte. Une fois sur place, les effecteurs nucléaires agissent comme des facteurs de transcription.

1.2.2.3. Effector Triggered Immunity (ETI) : La résistance induite spécifique

Dans l'étape suivante du modèle de co-évolution en zig-zag, les plantes ont développé des moyens pour reconnaitre certains effecteurs et répondent à l'attaque du pathogène par des mécanismes de défense robustes et rapides (Jones and Dangl, 2006). L'effecteur reconnu par la plante est alors défini comme une protéine avirulente (Avr). Ce deuxième front de résistance est appelé Effector-Triggered Immunity (ETI) et se base sur l'activation de protéines R de type NB-LRR (Nucleotide Binding-Leucine-Rich-Repeat) codées par des gènes de résistance (gènes R) (Tameling and Takken, 2008) (Figure 1.8). Les mécanismes de résistance déclenchés via les protéines R et la perception des effecteurs d'avirulence lors de l'ETI, sont des versions amplifiées de ceux activés lors de la PTI : activation des gènes du système de défense basal contrôlant la production de phytoalexines[6], renforcement des parois

[6] **Phytoalexines** : métabolites secondaires de faible poids moléculaire et possédant une activité anti-microbienne. Elles sont induites lors d'un stress et constituent une part importante du répertoire des mécanismes de défense des plantes (Ahuja et al., 2012).

cellulaires, production de formes réactives à l'oxygène (ROS), modification du flux ionique, cascade de signalisation des MAP kinases, pour aboutir au final à une mort cellulaire localisée au site d'infection (Jones and Dangl, 2006 ; Morel and Dangl, 1997).

- **Les protéines R cytoplasmiques ou NBS-LRR**

Les NBS-LRR (Nucleotide Binding Site-Leucine Rich- Repeats proteins) sont des protéines de résistance cytoplasmiques codées par des gènes de résistance (R) des plantes. La fonction des protéines NB-LRR chez la plante est double : d'une part la reconnaissance du signal lié au produit du gène d'avirulence du pathogène, et d'autre part l'activation de la transduction du signal conduisant à la mise en place des réponses de défense.

Bien que le spectre de résistance assuré par les protéines R soit assez large (résistance aux bactéries, champignons, oomycètes, virus, nématodes, etc.), ces dernières peuvent être classées en seulement cinq grandes catégories sur la base de leurs domaines structuraux (Figure 1.9). La classe la plus représentée concerne les protéines R cytoplasmiques composées d'un domaine C-terminal[7] LRR (leucine-Rich-Repeat) impliqué dans les interactions protéines-protéines, un domaine central de liaison aux nucléotides NBS (Nucleotide Binding-Site) et un domaine variable N-terminal[8] (Figure 1.8) (Padmanabhan et al., 2009). L'agencement et le nombre de répétitions des motifs riches en leucine (LRR) étant à l'origine de la spécificité et de la diversité des gènes de résistance (Parniske et al.1997 ; Dixon et al. 1998 ; Van der Hoorn et al. 2001). Le domaine N-terminal peut contenir un domaine Coiled-coil (CC-NB-LRR) ou un domaine Toll-interleukine-1 Receptor Like (TIR-NB-LRR). C'est d'ailleurs cette extrémité N-terminale qui permet de faire appel aux mécanismes de défense sous-jacents (Feys and Parker, 2000). Les gènes codant pour les NBS-LRR sont inégalement répartis dans les génomes des végétaux, ils peuvent être isolés mais sont la plupart du temps regroupés en cluster[9] (Michelmore and Meyers, 1998 ; Sagi et al., 2017). La distribution des NBS-LRR s'explique par des duplications de gènes en tandem et par la duplication de gènes (individuels ou en petits groupes) dans des sites non homologues (duplication ectopique) (Leister, 2004).

[7] **Le domaine ou extrémité C-terminal** : extrémité d'une protéine caractérisée par la présence d'une fonction carboxyle (-COOH) porté par le dernier acide aminé de la chaîne.
[8] **Le domaine ou extrémité N-terminale** se réfère au début d'une protéine et portant un acide aminé avec une fonction amine libre ($-NH_2$).
[9] Cluster de gènes : ensemble de deux gènes ou plus codant pour un même produit ou des produits similaires.

1.2.2.4. Transduction du signal : réponse précoce à la détection de pathogènes

La reconnaissance spécifique d'un pathogène par la plante déclenche instantanément une cascade de signalisations communes aux mécanismes de la PTI et de l'ETI (Tsuda and Katagiri, 2010). Les signaux acheminent les informations jusqu'au noyau de la cellule végétale pour induire une intense reprogrammation transcriptionnelle[10]et la mise en place de mécanismes de résistance. Cette cascade de signalisations utilise de nombreux seconds messagers impliqués dans un vaste réseau de voies interconnectées impliquées dans des flux ioniques (Bigeard et al., 2015 ; Dodd et al., 2010 ; Du et al., 2009), une cascade de signalisations MAP kinase (Nakagami et al., 2005 ; Romeis et al., 1999), la production d'espèces réactives à l'oxygène (ROS) (Lamb and Dixon, 1997 ; Torres, 2006 ; Van Breusegem, 2006) et des changements hormonaux (Bari and Jones, 2009 ; Jeandet et al., 2013 ; Tsuda and Katagiri, 2010) (Figure 1.9). L'activation des cascades de MAPK est l'un des évènements de signalisation les plus précoces suivant la détection de pathogène par l'une ou l'autre des branches de l'immunité innée des plantes. Une fois activées les MAPK phosphorylent des protéines cibles, dont des facteurs de transcription et des enzymes (Figure 1-9). Les MAPK contrôlent ainsi la synthèse et la signalisation des hormones et gènes de défense, des phytoalexines, la fermeture stomatique, la réaction d'hypersensibilité (HR), etc. Certains effecteurs de virulence des pathogènes sont capables de modifier les composants de la voie de la cascade de signalisation. Ce qui conduit à une déphosphorylation des MAP kinases et une suppression des signaux de défense de la plante (Underwood et al., 2007).

1.2.2.5. Les réponses de défense des plantes

i. Le renforcement des parois cellulaires

La paroi des cellules végétales représente une barrière dynamique contre les attaques de pathogènes (Underwood, 2012). Les mécanismes liés à son renforcement pariétal jouent donc

[10] Reprogrammation transcriptionnelle : phénomène de changements globaux dans l'expression des gènes qui sont généralement initiés par des facteurs de transcription. Au cours de la reprogrammation transcriptionnelle, l'expression de gènes spécifiques est élevée, tandis que d'autres gènes sont réprimés, par rapport à l'état précédent (Sung, 2013).

un rôle majeur dans la résistance des plantes contre leurs agresseurs (Voigt, 2014). Ce renforcement s'effectue par l'incrustation de callose, cellulose, pectine, lignine, de composés phénoliques, et de glycoprotéines structurales (glycoprotéines riches en hydroxyproline-HRGPs). Au final, de nouvelles barrières, appelées papilles, ou appositions pariétales, se forment et contribuent à retarder la progression de l'agent pathogène dans les tissus de la plante. Ces papilles sont également le siège de l'accumulation des composés antimicrobiens (ROS, peroxydases) (Zeyen et al., 2002).

ii. La réaction d'hypersensibilité

La réaction d'hypersensibilité (Hypersensitive Response ou HR) est observée lorsqu'un pathogène est reconnu spécifiquement et précocement par la plante (Tamelling et Takken, 2008). Il s'agit d'une forme de mort cellulaire rapide utilisée comme moyen de défense ultime contre les pathogènes de type biotrophe ou hémibiotrophe (Morel and Dangl, 1997). La plante hôte choisit alors de sacrifier les cellules contaminées afin d'isoler le pathogène et stopper sa propagation (Tameling and Takken, 2008).

La HR se manifeste par le développement de lésions nécrotiques dans lesquelles les cellules s'effondrent, suivi d'une déshydratation localisée et d'un brunissement des cellules affectées (Wright et al., 2000). Elle est activée et régulée par une cascade de signalisation commune à la PTI et l'ETI, faisant intervenir l'accumulation de plusieurs composés (acide salicylique, ROS, monoxyde d'azote), la cascade d'activation des MAPK, les flux d'ions calcium et la forte reprogrammation de l'expression des gènes au sein des cellules végétales (Figure 1.9) (Amirsadeghi et al., 2006).

Simultanément il peut y avoir, dans la nécrose crée par la HR, une accumulation de composés toxiques pour le pathogène telle que la synthèse de protéines PR (Pathogenesis-Related proteins) et des métabolites secondaires antimicrobiens (Mur et al., 2008).

En plus d'agir directement dans la zone infectée, la HR sert de médiateur pour activer les réponses de défense aux cellules adjacentes (LAR-Résistance Locale Acquise) (Kombrink and Schmelzer, 2001) et au niveau systémique (SAR- Résistance systémique acquise) (Strobel et al., 1996).

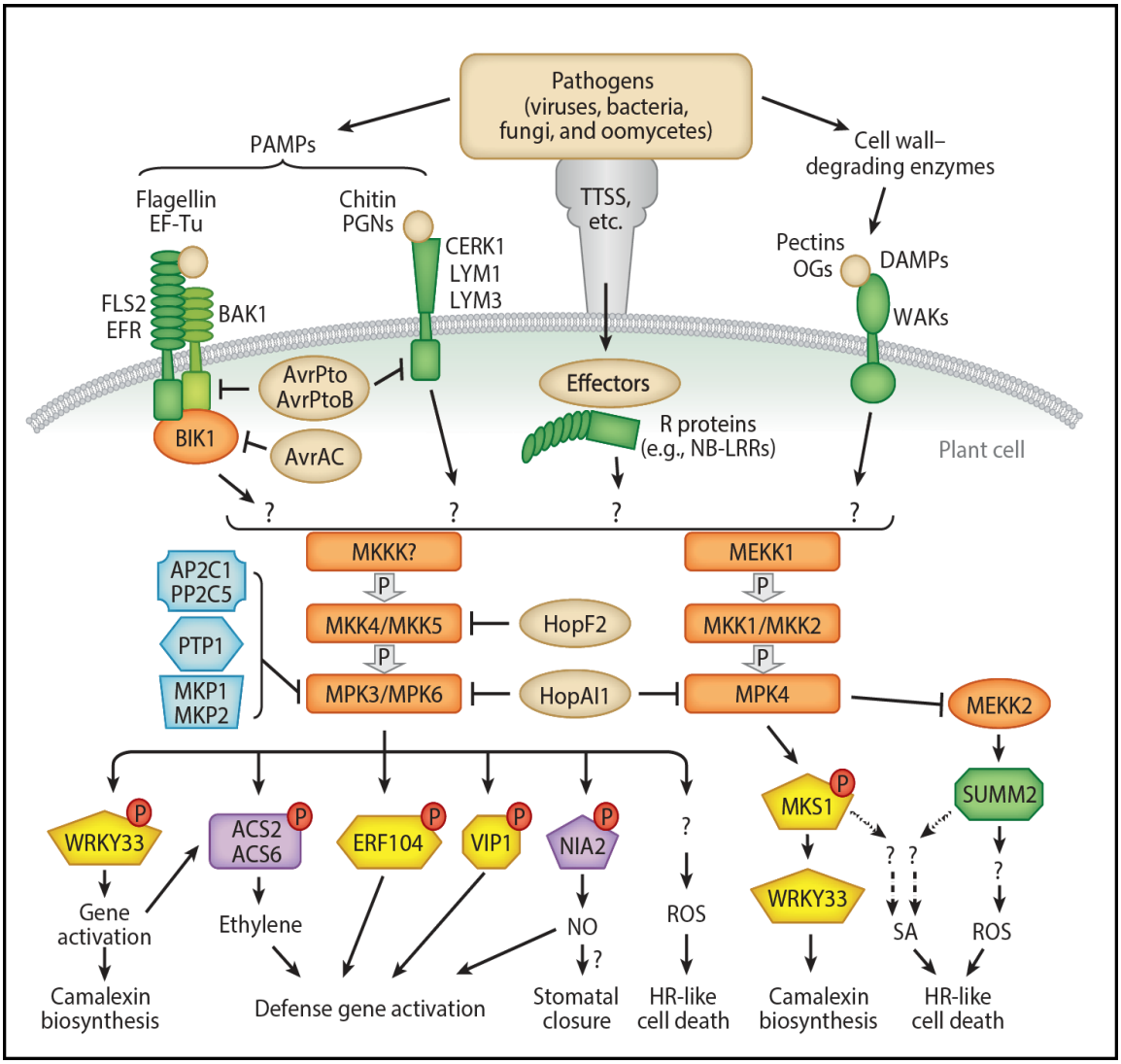


Figure 1.9 : Schéma des cascades de signalisation des MAPK (Mitogen-Activated Protein Kinase) activées lors de la reconnaissance des signaux moléculaires associés aux microorganismes MAMP (Microbe-Associated Molecular Pattern), DAMP (Damage-Associated Molecular Pattern) de la PTI, ainsi que les éliciteurs d'avirulence de l'ETI. D'après (Meng and Zhang, 2013).

Abréviations : NB-LRR (Nucleotide Binding Leucine Rich Repeat) ; R proteins : protéine de résistance ; NO, oxyde nitrique ; OGs : acides oligogalacturoniques; PGN: peptidoglycanes; ROS: espèces réactives à l'oxygène; SA : acide salicylique; TTSS: système de sécrétion de type 3 (utilisé par les bactéries); WAKs (Wall-Associated Kinase):protéine kinases associée à la paroi cellulaire. Les points d'interrogations indiquent les éléments non identifiés de la cascade de signalisation.

iii. Les phytoalexines

Selon la définition de référence énoncée par Paxton, 1981, les phytoalexines sont des composés antimicrobiens de faible poids moléculaires, synthétisés par la plante et s'y accumulant lors d'une interaction avec un agent pathogène. Les phytoalexines sont chimiquement diversifiées, incluant des composés tel que les dérivés phénylpropanoïdes, flavonoïdes, isoflavonoïdes, des sesquiterpènes et des polyketides (Hammerschmidt, 1999). Elles comprennent par exemple, le medicarpin produit par la luzerne (*Medicago sativa*), la rishitine produite à la fois par les tomates et les pommes de terre (Solanaceae) et la camalexine produite par *Arabidopsis thaliana*. La production de phytoalexines est déclenchée par les mécanismes de la reconnaissance spécifique entre les produits des gènes d'avirulence des pathogènes et des gènes R des plantes. Elles s'accumulent rapidement autour du site d'infection et est associée à une réaction hypersensible (Bailey et Ingham 1971 ; Bailey et Deverall, 1971 ; Harper 1987). Leur synthèse peut être activée par plusieurs métabolites secondaires issus des réponses précoces (ROS), principalement par la voie de signalisation de l'acide jasmonique.

iv. Les Pathogenesis-Related Proteins

Les protéines de défense PR (Pathogenesis-Related) possèdent des propriétés antimicrobiennes (toxicité et/ou activité hydrolytique de la paroi cellulaire du pathogène). Elles incluent les : glucanases, chitinases, peroxydases, thaumatin-like protein, défensines, etc. La voie de signalisation principale conduisant à leur synthèse est celle de l'acide salicylique (Sudisha et al., 2012).

1.3. Gestion des maladies de plantes par la sélection

1.3.1. Historique de la sélection : de la sélection massale à l'utilisation des marqueurs moléculaires

Dans la nature, l'évolution résulte de pressions de sélection qui s'exercent sur des populations polymorphes. De même, la création varietale suppose l'existence d'une diversité parmi les plantes cultivées et l'application de pressions sélectives par l'homme. Par un ensemble de processus, l'Homme a longtemps exercé une sélection chez les plantes qu'il cultive afin de développer des cultivars adaptés à ses besoins. La sélection peut avoir divers buts : faciliter la culture, la récolte, améliorer les rendements, favoriser la résistance aux maladies, la qualité nutritionnelle, etc. Les premières traces de sélection dirigée par l'Homme remontent au début de notre civilisation avec le processus de domestication des plantes sauvages. Il y a 10 000 ans, l'Homme se sédentarise, il cultive des plantes pour son alimentation et il choisit de manière empirique les meilleurs individus à ré-ensemencer. L'opération est répétée sur plusieurs générations de plante, transformant ainsi progressivement les espèces sauvages pour aboutir aux premières plantes domistiquées. La domestication des plantes résulte ainsi des choix, plus ou moins conscient, de l'Homme combinés à des phénomènes naturels indépendants (tels que les recombinaisons génétiques et les mutations ponctuelles). La sélection sur la base de caractères observables (phénotype) à l'intérieur d'une population, s'est effectuée durant des milliers d'années et relève de ce qu'on appelle aujourd'hui la sélection massale.

Au cours des deux derniers siècles, la sélection des plantes a évolué en s'appuyant sur les découvertes en génétique. Darwin décrit le principe de l'hybridation et de la sélection naturelle dès la fin du 19ème siècle. A la même époque Mendel définit la relation entre le génotype et le phénotype (loi de l'hérédité). En 1935, Morgan décrit le support physique de l'hérédité (les chromosomes) avec la théorie chromosomique de l'hérédité. Dès lors les progrès en génétique se sont rapidement enchainés, la notion de gène s'est progressivement construite comme étant l'unité de base de l'hérédité prédéterminant un caractère phénotypique. A la fin du 20ème siècle (1985-1990), la découverte des marqueurs moléculaires permet d'accéder directement aux gènes gouvernant l'expression du caractère souhaité. Les marqueurs moléculaires sont des séquences d'ADN aisément détectables sur le génome et permettant l'identification et l'étiquetage de gènes d'intérêt d'une génération à l'autre (Semagn et al., 2006). Ils sont indépendants de l'organe mesuré, du stade de développement et de l'environnement (Moose and Mumm, 2008). Leur utilisation dans la sélection des plantes (sélection assistée par marqueurs ou SAM) facilite la construction de génotypes cumulant les

gènes favorables à l'origine des caractères d'intérêts (Foolad and Sharma, 2005). La sélection ne tient plus seulement compte de la diversité des caractères observables (phénotypes), mais également de la diversité génétique qui y est liée.

1.3.2. Sélection de caractères de résistance aux pathogènes

En agronomie, la sélection pour la résistance a pour but de développer des cultivars, qui montrent peu/pas de réduction de leur rendement, quand ils sont exposés à des pathogènes. Selon plusieurs études, l'utilisation de la résistance en gestion des maladies de plantes s'avère être une technologie respectueuse de l'environnement et peu couteuse vis-à-vis de l'usage de produits phytosanitaire (Brown, 2015 ; Cuartero et al., 1999 ; Poland et al., 2009 ; Wallwork, 2009). La notion de résistance englobe tous les caractères héritables d'une plante lui permettant d'éviter, tolérer et résister aux attaques de pathogènes, dans des conditions qui causent des symptômes sévères sur les individus sensibles de la même espèce (Cuartero et al., 1999). La résistance aux pathogènes peut être classifiée en deux catégories : (i) la résistance complète ou qualitative et (ii) la résistance partielle ou quantitative (Poland et al., 2009).

i. La résistance qualitative

La résistance qualitative (résistance verticale ou spécifique) est gouvernée par l'expression de gènes majeurs[11] de résistance, hérités selon un déterminisme mendélien simple. Ce type de résistance se caractérise par une absence de symptômes, causée par l'arrêt ou un dérèglement majeur du cycle infectieux du pathogène qui ne produit alors plus de descendants. La résistance qualitative est associée à l'hypothèse de la résistance gène-pour-gène énoncée par H.H. Flor en 1955. En se basant sur le pathosystème de la rouille du lin (*Linum usitatissimum -Melampsora lini*), Flor (1955) postule que l'hérédité du caractère de résistance de l'hôte et de la pathogénicité du pathogène sont contrôlées par des paires de gènes dominants[12] et complémentaires : un gène de résistance chez la plante *(R)* et un gène d'avirulence (*Avr*) (Tableau 1.2). L'interaction entre le gène d'avirulence d'un isolat spécifique du pathogène (*Avr*) avec le gène de résistance (*R*) d'un génotype de l'hôte végétal, conduit à la résistance qualitative. De nombreux gènes de résistance majeurs ont été identifiés à partir d'espèces modèles ou cultivées. Par exemple chez le café dans le pathosystème *Coffea arabica* L. -

[11] Gène majeur ou « gène à effet majeur » : qualifie un nombre limité de gènes impliqués dans l'expression d'un caractère qualitatif.
[12] Gène dominant : allèle s'exprimant à l'état hétérozygote

Hemileia vastatrix Berk. & Br. ; chez le maïs dans le pathosystème *Zea mays* L. - *Puccinia sorghi* Schw., chez le blé dans le pathosystème *Triticum aestivum* (L.) Thell.- *Puccinia recondita*, etc (Vale et al., 2001). Bien qu'efficace contre des races spécifiques de pathogènes, la résistance qualitative n'est cependant pas durable dans le temps. Selon le modèle de co-évolution en zig-zag, les pathogènes parviennent à détourner la résistance conférée par des gènes R majeurs en déployant de nouveaux effecteurs de virulence (Jones and Dangl, 2006). Ce type de résistance ne fait sens que lorsqu'elle est opérée contre des populations de pathogènes de taille restreinte et dans des conditions environnementales moins propices à leur développement.

Tableau 1.2 : Modèle génétique de la résistance gène-pour-gène (d'après Flor, 1971)

*La présence de l'allèle R du gène de résistance chez la plante, et de l'allèle Avr correspondant chez l'agent pathogène déclenche la résistance spécifique. On parle alors d'**interaction incompatible**. Dans toutes les autres situations, l'**interaction** est dite **compatible** et la maladie se développe.*

Plante-hôte / Agent Pathogène	*R*	*r*
Avr	Résistance	Maladie
avr	Maladie	Maladie

ii. La résistance quantitative

La résistance quantitative (résistance horizontale ou résistance partielle) est régie par une multitude de gènes mineurs [13] dont les effets sont cumulatifs (Schiff et al., 2001). Elle s'exprime par un ralentissement de la progression et/ou une réduction des symptômes de la maladie (Poland et al., 2009). Contrairement à la résistance qualitative caractérisée par une réponse en "tout ou rien", la résistance quantitative se définit comme un continuum de réponse qui va, selon le génotype hôte, d'un haut niveau de résistance à la sensibilité complète. Ce type de résistance confère l'avantage d'être, en pratique, plus durable que la résistance qualitative (Mundt, 2014). L'adaptation de l'agent pathogène est théoriquement

[13] Gènes mineurs ou plygènes : groupe de gènes qui, lorsqu'ils sont exprimés ensemble, produisent un phénotype particulier.

plus complexe quand plusieurs gènes de résistance agissent par rapport à un seul gène. La résistance quantitative est approchée par l'analyse de QTLs (Quantitative Trait Loci), des portions de chromosomes où sont localisés un ou plusieurs gènes mineurs contribuant significativement à l'expression de la résistance (Lannou, 2012).

1.3.3. Stratégie d'utilisation de la résistance en sélection

Historiquement, la gestion des maladies de plantes de culture s'est surtout fondée sur la résistance qualitative et l'utilisation de gènes majeurs de résistance (R). Une méthode qui s'est révélée au départ efficace. Toutefois, le déploiement à grande échelle des gènes majeurs R finissait par aboutir à un contournement de la résistance qualitative. Ce contournement de la résistance résultait d'une adaptation du pathogène (nouveaux pathotypes plus virulents), facilitée par l'homogénéité du paysage variétal. Actuellement, les tendances en sélection visent à augmenter la durabilité de la résistance par l'utilisation de la résistance quantitative. Il est ainsi largement admis que le pyramidage de gènes est l'approche la plus durable d'utilisation de la résistance en sélection (REX Consortium, 2016). La faible probabilité de contourner simultanément chaque gène employé dans la résistance, permet alors de cultiver les espèces dans des conditions favorables au développement de la maladie. Cette méthode a permis de contrôler l'un des pathogènes les plus dévastateur du blé (la rouille du blé, *Puccinia graminis* f. sp. *tritici*) à l'échelle mondiale pendant plus d'un demi-siècle (Mundt, 2014). Jusqu'à récemment, la restriction majeure à l'utilisation du pyramidage de gènes était le temps de génération d'une variété combinant l'ensemble des gènes sélectionnés. De nouvelles approches existent désormais afin de réduire ce temps, elles incluent entre autre, la sélection assistée par marqueurs moléculaire (SAM) et l'utilisation des nouvelles techniques de séquençage haut-débit (Kolmer, 1996 ; Foolad et al., 2002 ; Singh et al., 2005 ; Liu et al., 2004 ; Richardson et al., 2006 ; Asea et al., 2009 ; Moloney et al., 2009).

1.3.4. La sélection chez les espèces forestières

Les espèces forestières sont définies comme des espèces végétales ligneuses, pérennes, pouvant se situer à différents stade de domestication[14] (Neale, 2007). Le développement des techniques de sélection applicables aux espèces forestières cultivées est en constante

[14] Domestication : production de génotypes ou de variétés arborant des caractères phénotypiques désirables pour l'Homme.

expansion pour les besoins de nombreux secteurs économiques (la production de bois, de papier, de biocarburant ou bio-matériel).

La sélection de génotypes résistants représente la forme de lutte la plus économique pour réduire les impact des maladies sur les arbres forestiers (Sobrosa and Martins–Corder, 2001). Malgré un large engouement de la sélection assistée par marqueurs appliquée aux espèces d'intérêt agricole et horticole, les espèces forestières sont longtemps restées en marge de cette technologie. En cause, le temps de génération particulièrement long de ces dernières (temps de germination des graines, floraison, etc.) (Boerjan, 2005). Le potentiel de la SAM chez les arbres forestiers est également limité par les effets de la recombinaison génétique entrainant une diminution de la liaison entre le marqueur et le caractère d'intérêt à chaque génération (Boerjan, 2005). Il en résulte des faux-positifs, pour lesquels la détection des marqueurs originellement sélectionnés n'est plus liée au locus d'intérêt.

L'avènement des nouvelles techniques de séquençage d'ADN (NGS) (notamment le haut débit et la baisse des coûts) ont permis de surmonter les limites de la sélection chez les espèces forestières (Deschamps et al., 2012 ; Studholme et al., 2011). Ces dernières permettent notamment d'accéder rapidement à l'abondante source de variation génétique des organismes que représentent les SNP (Single Nucleotide Polymorphism) (Harfouche et al., 2012). Un SNP caractérise la différence d'une base nucléotidique entre deux séquences d'ADN. Ce type de polymorphisme est très fréquent au sein des génomes complexes d'eucaryotes, dont font parties les plantes (en moyenne 1 SNP tous les 100-300 pb) (Maughan et al 2010). Les SNPs peuvent être classés selon les substitutions de nucléotides qui les caractérisent, soit des transitions (C / T ou G / A), soit des transversions (C / G, A / T, C / A ou T / G). Une grande partie des SNPs se situent dans les régions non codantes du génome (introns et régions intergéniques) et peuvent, dans certains cas, intervenir dans la régulation des gènes (épissage, liaison aux facteurs de transcription ou séquence d'ARN non-codant). D'autres SNPs peuvent être présents à l'intérieur des séquences codantes (exon) et possiblement induire un changement d'acide aminé dans la protéine produite (mutation non-synonyme) (Brodie et al., 2016; Wright, 2005) (Figure 10). Les avantages principaux de l'utilisation des SNPs dans la SAM des espèces forestières, sont la haute résolution des associations marqueur-caractères (SNPs situés dans les séquences codantes) et leur transfert entre les familles ou les espèces. L'utilisation des SNPs pour la sélection de plants résistants a déjà été entreprise chez plusieurs espèces forestières, pour lesquelles les ressources génomiques étaient disponibles. Il s'agit entre autres des Peupliers (*Populus*) (Carletti et al.,

2016 ; La Mantia et al., 2013), des Pins (*Pinus)* (Liu et al., 2016 ; Quesada et al., 2010) et des *Eucalyptus* (Barbosa-da-Silva et al., 2005).

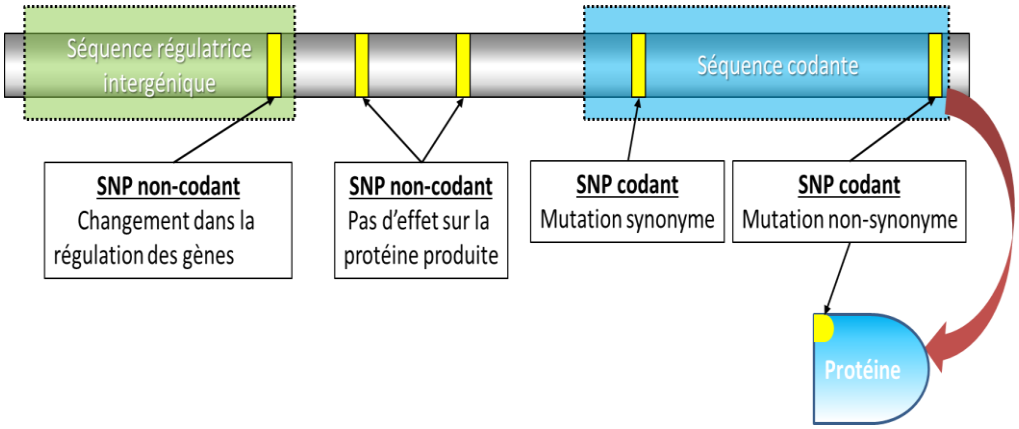


Figure 1.10 : Schéma illustrant les différents types de SNPs

1.3.5. Le cas des espèces forestières non-modèles

Les pipelines d'analyses[15] pour identifier les SNPs utilisables en sélection sont actuellement bien développés pour les espèces modèles, mais dépendent de la disponibilité d'un génome de référence annoté (localisation et fonction des gènes). Ils ne peuvent donc pas être appliqués aux espèces non-modèles pour lesquelles les ressources génomiques sont limitées, voir même inexistantes (Seeb et al., 2011). Or une grande partie des espèces forestières ne sont pas domestiquées et il n'existe pas dans la litérature de génome de référence pour elles. Les espèces forestières constituent ~ 82% de la biomasse continentale et abritent plus de 50% de la biodiversité terrestre (Neale and Kremer, 2011). Elles fournissent plusieurs services écologiques pour le bien-être de l'Humanité (séquestration de carbone, régulation du climat, etc.). Ainsi, une préoccupation globale sur l'état de santé des populations d'arbres à l'état naturel émerge dans la société et s'intensifie face aux menaces grandissantes que représentent les maladies biotiques, dans un contexte de changements climatiques avéré.

Dans ce cadre, le séquençage d'ARN (ou RNA-Seq) représente une alternative pour l'identification de SNPs dans les régions transcrites chez des espèces d'intérêts écologiques peu étudiées (Lopez-Maestre et al., 2016). Le RNA-Seq est une technique de séquençage

[15] Pipeline d'analyse : chaîne de traitement des données bio-informatiques formée par un ensemble d'outils d'analyse.

utilisant la technologie des NGS pour étudier le transcriptome, c'est-à-dire l'ensemble des gènes transcrits au même moment dans un tissu et à un moment particulier (Strickler et al., 2012). Les objectifs du RNA-Seq comprennent l'identification de l'ensemble des transcrits dans un échantillon biologique et l'estimation de leur abondance dans des conditions physiologiques particulières ou à des stades de développement spécifique (Singh et al., 2012). En mettant l'accent sur le séquençage des parties codantes plutôt que le génome entier, le RNA-Seq constitue une forme de réduction de la complexité des génomes (Everett et al., 2011 ; Wang et al., 2009). Même si l'absence d'un génome de référence ne constitue plus une limite pour l'étude des espèces non-modèles, les pipelines d'analyses sont quant à eux très hétérogènes et difficilement transposables d'une expérimentation à l'autre. Afin de répondre au challenge de gestion des données RNA-Seq, le développement d'outils de bio-informatiques et de bio-analyses adaptés sont devenus une prérogative de la biologie moderne (Mardis, 2016 ; Schmidt and Hildebrandt, 2017).

1.4. Le pathosystème Myrtaceae-*Austropuccinia psidii* : État des connaissances.

1.4.1. Les Myrtaceae

1.4.1.1. Taxonomie et caractéristiques

REGNE : Plantae
CLASSE : Magnoliopsida
SOUS-CLASSE : Rosidae
ORDRE : Myrtales
FAMILLE : Myrtaceae

La famille des Myrtaceae est considérée comme la huitième plus grandes familles d'angiospermes du monde, comprenant 142 genres et 5500 espèces (Grattapaglia et al., 2012 ; Wilson, 2010). Ce sont majoritairement des arbustes et des arbres prédominants dans l'hémisphère Sud, principalement en Amérique du Sud, en Asie du Sud-Est, en Afrique, en Inde, en Nouvelle-Calédonie et dans de nombreuses îles du Pacifique jusqu'à Hawaii (Thornhill et al., 2015) (Figure 1.11).

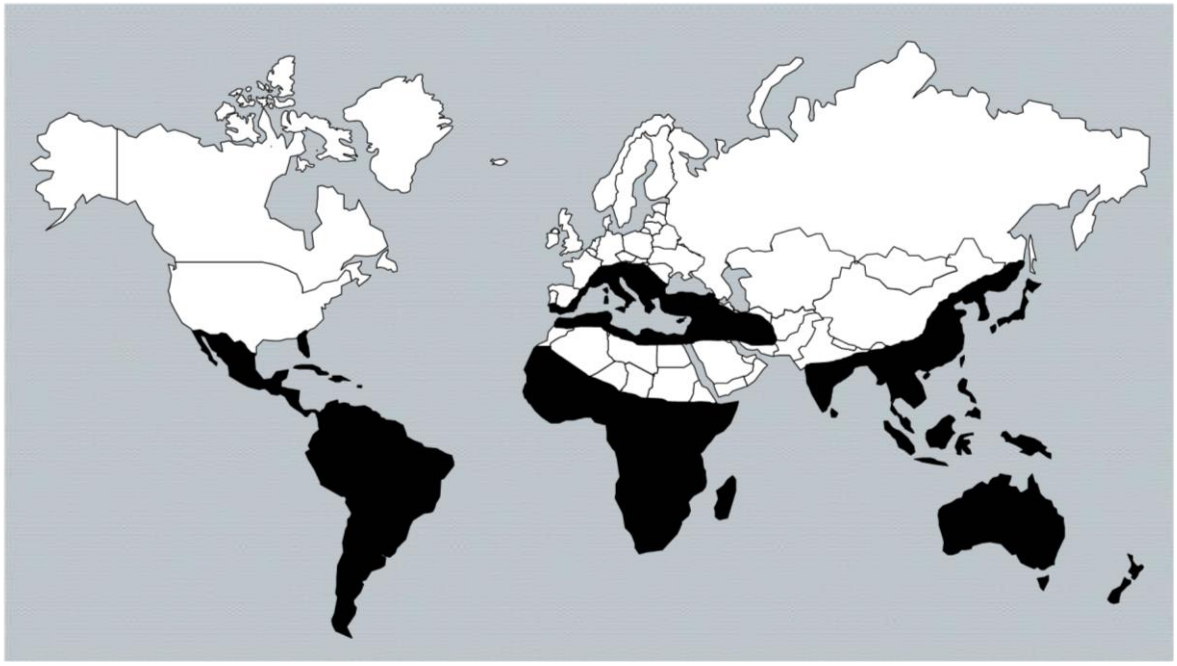


Figure 1.11 : Répartition mondiale de la famille des Myrtaceae (en noir).
Extrait de Grattapaglia et al., 2012.

La Myrtaceae se distingue généralement par une combinaison de caractéristiques, incluant la présence de glandes schizogènes sur les feuilles persistantes, des fleurs disposées en multiples de quatre ou cinq avec de nombreuses étamines, un phloème situé de part et d'autre du xylème (non pas à l'intérieur comme dans la plupart des autre plantes), et la présence de « fosses » sur les vaisseaux du xylème (Wilson et al., 2001).

La famille a historiquement été divisée en trois sous-familles, principalement sur la bases des caractéristiques morphologiques des fruits : les Myrteae (fruits charnus), les Leptospermeae (fruits capsulaires), et les Chamelaucieae (fruits secs) (Candolle, 1828). Plus tard, cette classification a été réduite à deux groupes : les Myrtoïdeae, aux fruits charnus et aux feuilles opposées, et les Leptospermoideae possédant des fruits capsulaires et des feuilles alternes (Niedenzu 1893). Cette classification a été contestée pour la première fois par Johnson et Briggs (1979) sur la base d'analyse utilisant des caractères morphologiques. Enfin, l'utilisation de marqueurs moléculaires en phylogénie a permis une nouvelle classification intrafamiliale qui reconnait désormais deux nouvelles sous-familles : les Myrtoideae et les Psiloxydeae (Wilson et al., 2005). Par ailleurs, les études moléculaires récentes indiquent que plusieurs genres anciennement décrits et acceptés parmi les Myrtaceae ne sont pas monophylétiques et nécessitent une révision (Lam et al. 2002 ; Steane et al. 2002 ; Harrington and Gadek 2004 ; Sytsma et al. 2004 ; Lucas et al. 2005 ; van der Merwe et al. 2005).

1.4.1.2. Origine et distribution

La classification intrafamiliale des Myrtaceae, et l'origine de leur distribution sont complexes et sujettes à débat depuis de nombreuses années. De nombreuses études en biogéographie basées sur des données morphologiques (Johnson and Briggs, 1981 ; Raven and Axelrod, 1974) et phylogénie moléculaires (Rutschmann et al.,2007 ; Sytsma et al., 2004) suggèrent que les Myrtaceae ont une origine gondwanienne. De récents travaux montrent qu'elles sont apparues au cours du Crétacé, avant le morcellement du Gondwana (Figure 1.12) (Biffin et al., 2010 ; Thornhill et al., 2015). Une grande partie de la distribution moderne des taxons de Myrtaceae dans le monde peut être expliqué par des évènements tels que la dispersion longue-distance et la vicariance. À noter que les origines de la distribution restreinte de certains genres monotypiques sont encore inconnues (par exemple l'apparition du genre *Arillastrum* en Nouvelle-Calédonie).

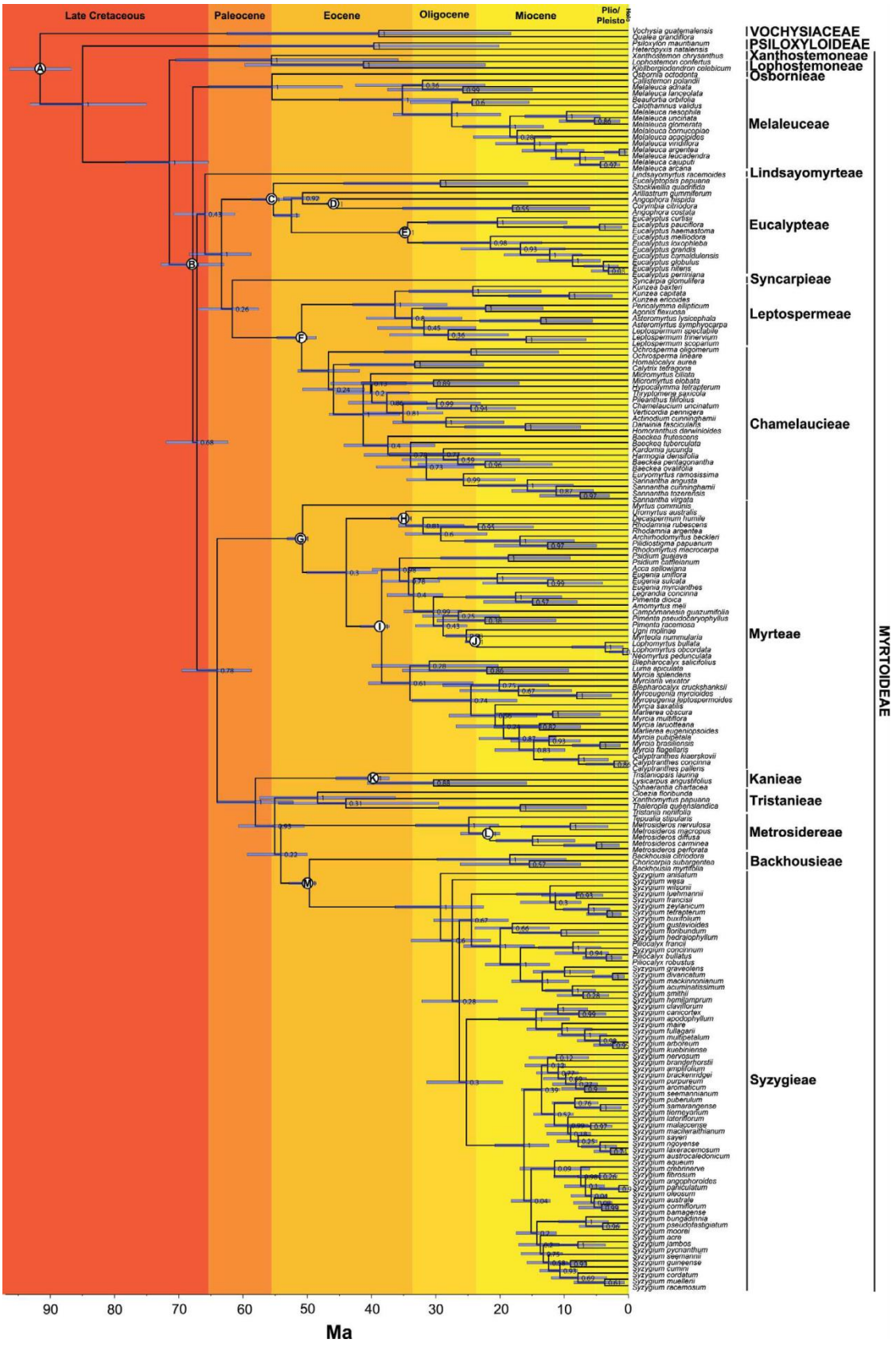


Figure 1.12 : Phylogénie des Myrtaceae d'après Thornhill et al., 2015.

1.4.1.3. Importance écologique et économique

La famille des Myrtaceae comprend plusieurs genres d'importance écologique et économique dans le monde entier. Les produits naturels fournis par cette famille comprennent le bois pour l'industrie papetière et la construction, la production d'énergie (*Eucalyptus* spp), les épices (clous de girofle : *Syzygium aromaticum*, quatre-épices : *Pimenta dioica*), les fruits charnus pour l'industrie agroalimentaire (*Psidium*-goyaves, *Eugenia*-Pitanga) et les plantes ornementales (Reynertson et al., 2008). L'aspect remarquable dans cette famille tient à sa richesse en huile essentielle, tanins, composés phénoliques et terpènes, qui en fait l'une des familles de plantes les plus recherchées pour l'industrie cosmétique et thérapeutique (Keszei *et al*., 2008 ; Padovan et al., 2014 ; Asgar *et al*., 2008).

1.4.1.4. Ressource génomique

L'exploration des génomes des Myrtaceae est due en grande partie aux besoins de recherche exprimés par l'industrie forestière. L'accent a été mis sur les *Eucalyptus*, principales espèces forestières plantées à travers le monde pour leur croissance rapide, leur capacité d'adaptation et leur utilité dans de nombreux secteurs économiques (bois de construction, industrie papetière, etc.). Le genre *Eucalyptus* représentent un groupe diversifié d'arbres, comptant plus de 700 espèces réparties dans sept sous-genres polytypiques et six sous-genres monotypiques (Brooker, 2000). La plupart des espèces d'*Eucalyptus* appartiennent au sous-genre Symphyomyrtus, qui comprend des espèces d'importance économique telles que *E. grandis*, *E. urophylla* et *E. globulus* (Doughty, 2000 ; Grattapaglia et al., 2012).

Le génome de l'*Eucalyptus grandis* constitue le génome de référence pour les *Eucalyptus*. C'est le premier génome séquencé pour la famille de Myrtaceae, sa taille est estimée à 640 Mb (Myburg et al., 2014). Sur plus de 36 000 protéines prédites, 34% résultent de duplication en tandem. Les duplications en tandem impliquent souvent les gènes de réponses au stress, leur rôle est primordial dans les dynamiques adaptatives des espèces (Yu et al., 2015). *E. grandis* montre également une forte diversité de gènes impliqués dans les métabolites secondaire (par exemple les terpènes), c'est l'une des plus importantes détectée à ce jour chez les plantes. Ces métabolites secondaires agissent notamment comme des moyens de défenses chimiques contre un éventail de pathogènes et d'herbivores dans le milieu naturel. Beaucoup de ces métabolites entrent également dans la composition d'huiles d'intérêt pour la pharmacologie et la cosmétique.

Récemment, un autre génome de Myrtaceae a été séquencé, c'est le cas de l'espèce endémique d'Hawaii *Metrosideros polymorpha* (Izuno et al., 2016). La taille de son génome a été estimée à 304 Mb avec plus de 30 000 protéines prédites.

1.4.2. *Austropuccinia psidii*

Austropuccinia psidii (syn. *Puccinia psidii*, Sphaerophragmiaceae, Pucciniales) est une rouille qui a été découverte et décrite par G. Winter en 1884 sur *Psidium guajava L.* (goyavier) au Brésil (Beenken, 2017 ; Glen et al., 2007). Ce champignon est connu pour être l'une des principales menaces pesant sur les plantes appartenant à la famille des Myrtaceae à l'échelle mondiale. Sa gamme d'hôte a été recensée à 460 espèces hôtes couvrant 73 genres de Myrtaceae (Carnegie et al., 2016a ; Giblin and Carnegie, 2014 ; Roux et al., 2016). La maladie causée par ce pathogène est plus communément connue sous les noms de rouille du goyavier (*Psidium guajava*), rouille de l'*Eucalyptus*, ou encore rouille des Myrtaceae.

1.3.2.1. Taxonomie

Depuis sa découverte, la position taxonomique d'*A. psidii* est sujette à débat. *A. psidii* a été initialement nommé *Puccinia psidii* et placé dans le genre Puccinia (Basidiomycota, Pucciniales, Pucciniaceae) par Winter (1884). Ce dernier s'était alors basé sur une description morphologique du stade téléomorphe (reproduction sexuée) du champignon. Le stade anamophe (asexué) du champignon quant à lui, ayant été alors décrit comme étant une espèce différente (*Uredo rangelii*) (Simpson et al., 2006). L'identification d'*A. psidii*, sur la seule base de critères morphologiques, a ainsi conduit à de nombreuses confusions avec pas moins de 27 synonymes du même pathogène (Glen et al., 2007). L'utilisation de moyens moléculaires en taxonomie a permis de préciser la classification de ce champignon parmi les Pucciniales (rouilles). En 2008, une phylogénie basée sur le gène de la *ß-tubuline* exclue totalement *A. psidii* de la famille Pucciniaceae (van der Merwe et al., 2008). En 2016, *A. psidii* est placée dans la famille des Sphaerophragmiaceae grâce à une phylogénie basée sur les régions LSU (large sous-unité) et SSU (petite sous-unité) de l'ADN ribosomale et la sous-unité 3 du cytochrome c oxydase (*cox3*) de l'ADN mitochondrial (McTaggart et al., 2016). Enfin en 2017, le nouveau genre *Austropuccinia* est créé par (Beenken, 2017) en se basant sur l'épitype[16] de Machado *et al.* 2015 (Figure 1.13).

[16] Selon la définition officielle du code international de nomenclature en botanique (Art. 9.18), un **épitype** désigne un spécimen (ou illustration) choisi comme modèle d'interprétation lorsque l'holotype, le lectotype, le

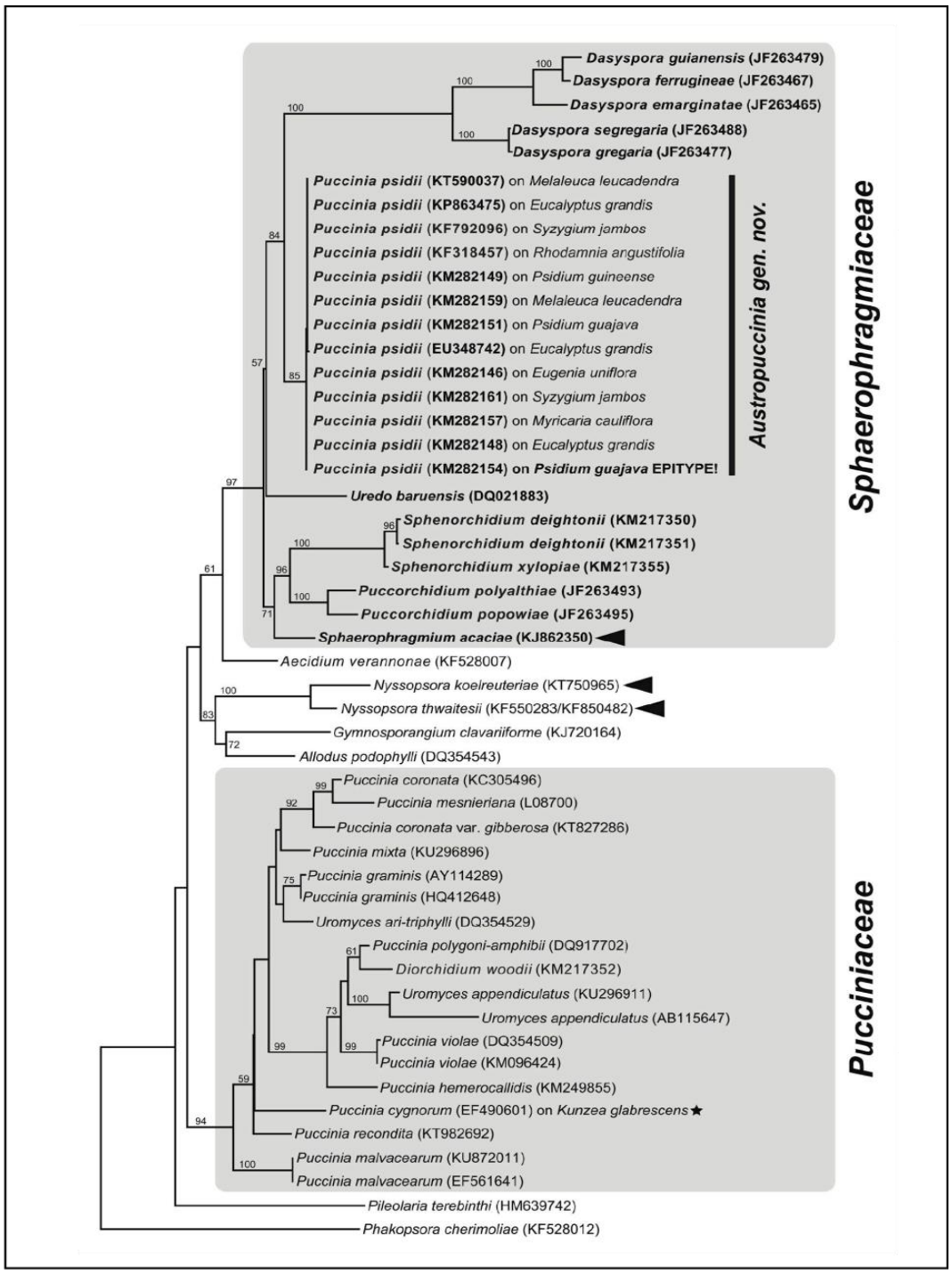


Figure 1.13 : Positionnement phylogénétique d'*Austropuccinia psidii* (syn. *Puccinia psidii*), récoltés à partir de plusieurs hôtes de Myrtaceae, parmi la famille des Sphaerophragmiaceae. D'après (Beenken, 2017)**.**

néotype ou l'ensemble du matériel original, associé à un nom validement publié, est manifestement ambigu et ne peut être identifié de manière probante en vue de l'application précise d'un nom de taxon. Lorsqu'un épitype est désigné, le type auquel il fait référence (holotype, lectotype ou néotype) doit être explicitement cité.

1.3.2.2. Distribution

A. psidii est un pathogène natif d'Amérique du Sud et Centrale. On le retrouve de l'Argentine à Cuba, en passant par l'Uruguay, le Paraguay, le Brésil, la Colombie, le Venezuela, l'Equateur, le Mexique, le Salvador, le Guatemala, la Jamaïque, Porto Rico et la République Dominicaine (Coutinho et al., 1998 ; Rodas et al., 2015). Le pathogène est resté cantonné dans sa région d'origine jusqu'à la fin des années 1970 puisqu'alors détecté pour la première fois au sud de la Floride (Marlatt and Kimbrough, 1979). Par la suite, l'aire de répartition d'*A. psidii* s'est progressivement étendu à la Californie (Mellano, 2006), à Hawaii (Uchida et al., 2006), au Japon (Kawanishi et al., 2009), à la Chine (Zhuang, J-Y. & Wei, S-X.), à l'Australie (y compris la Tasmanie) (Carnegie et al., 2010 ; Potts et al., 2016), l'Afrique du Sud (Roux et al., 2013), la Nouvelle-Calédonie (Giblin, 2013), l'Indonésie (Sumatra) (McTaggart et al., 2015), Singapore (du Plessis et al., 2017) et dernièrement à la Nouvelle-Zélande (Department of Conservation, May 2017) (Figure 1.14). La dispersion des spores par le vent constitue le principal moyen de propagation des agents pathogènes. Toutefois, les effets combinés du changement climatique et de l'intensification des activités humaines (en particulier le commerce international et l'urbanisation) ont été aussi reconnues comme causes principales d'introduction ou migration des espèces invasives et d'émergence des maladies (Brasier and Buck, 2001 ; Desprez-Loustau et al., 2016).

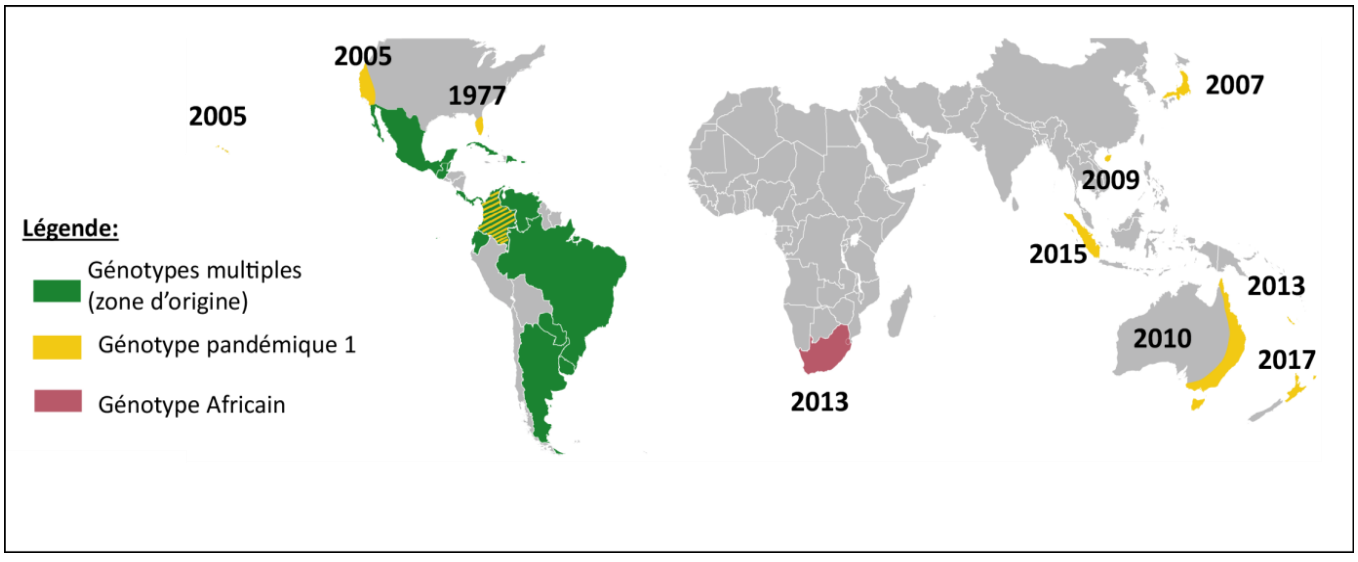


Figure 1.14 : Distribution globale d'*Austropuccinia psidii* avec les années de détection et le génotype. D'après Graça et al., 2013 et Granados et al., 2017.

1.3.2.3. Gamme hôte et épidémies

Région d'origine d'*A. psidii* :

Il existe relativement peu d'études concernant l'ampleur de la gamme hôte d'*A. psidii* en Amérique du Sud et Centrale. On sait cependant que les dommages et la sévérité des symptômes sont relativement moindres sur la flore indigène que sur les espèces introduites, à l'exception d'épidémies observées dans les plantations de goyaviers (*Psidium guajava*) au Brésil (Ribeiro and Pommer, 2004) et de *Pimenta dioica* (quatre-épices) en Jamaïque (Maclachlan, 1938). Les activités économiques liées à la culture de ces deux espèces de Myrtaceae ont été mises en péril à de nombreuses reprises du fait des épidémies récurrentes et particulièrement virulentes d'*A. psidii* (Ferrari et al., 1997 ; Maclachlan, 1938).

Les *Eucalyptus* ont été introduits au Brésil au début du 20ème siècle pour le développement de la filière du bois en foresterie. Des clones hybrides d'*E. grandis*, sélectionnés pour leur qualité de bois, leur rapidité de croissance et leur rendement, ont été massivement plantés. Dès la seconde moitié du 20ème siècle, le Brésil est devenu l'un des principaux producteurs de pulpe d'*Eucalyptus* au monde pour l'industrie papetière (Droulers and Andrade, 2000). La première épidémie de rouille due à *A. psidii* sur ces *Eucalyptus* introduits a eu lieu en 1973 dans la province d'Espirito Santo, entrainant au passage des pertes économiques considérables dans les pépinières et les plantations (Coutinho et al., 1998). En général, la rouille entraine une réduction de 25-35% de la hauteur et du diamètre des arbres sensibles et une perte de 41% du volume de bois (Rodas et al., 2015). Dès lors, la maladie, communément appelée rouille de l'*Eucalyptus*, est devenue une contrainte majeure pour la culture des *Eucalyptus* en Amérique du Sud.

En dehors de la région d'origine d'*A.psidii* :

L'expansion de l'aire de répartition d'*A. psidii* en dehors de sa région d'origine, s'est accompagnée d'un accroissement de sa gamme hôte. En Floride, le pathogène a d'abord été observé sur des espèces introduites (*Pimenta dioica* et *Melaleuca quinquinervia*), puis sur des espèces natives (*Myrcianthes fragrans*) (Glen et al., 2007). A Hawaii, la rouille a été détectée pour la première fois sur l'île d'Oahu au niveau des semis en serre de *Metrosideros polymorpha* ('Ōhi'a), une espèce endémique et commune de l'archipel (Uchida et al., 2006). La maladie s'est ensuite très rapidement répandue à l'ensemble des îles hawaiiennes en

élargissant son spectre d'hôtes de 33 nouvelles espèces (~16,5% des espèces de Myrtaceae connues à Hawaii) (Silva et al., 2014). En Afrique du Sud, le pathogène a été détecté pour la première fois sur une espèce ornementale non-native (*Myrtus communis*), puis sur d'autres espèces natives (Roux et al., 2016). L'Australie est à ce jour, le pays le plus impacté par la rouille en terme de gamme d'hôtes avec plus de 230 espèces attaquées en milieu naturel et 115 autres sensibles après inoculations contrôlées (Carnegie and Lidbetter, 2011 ; Giblin and Carnegie, 2014 ; Morin et al., 2012 ; Pegg et al., 2014a).

1.3.2.4. Cycle de vie d'*A. psidii*

A. psidii se comporte comme une rouille macrocyclique et autoïque (Coutinho et al., 1998 ; McTaggart et al., 2017). Tous les différents stades de spores sont produits sur le même hôte (aeciospores, urédospores, téliospores et basidiospores) à l'exception des pycniospores (Coutinho et al., 1998 ; Figueiredo, 2001 ; Glen et al., 2007) (Figure 1.15). En condition naturelle, les urédospores d'*A. psidii* sont produits en très grandes quantités et sont plus souvent observables que les téliospores, les basiodiospores et les aeciospores[17]. En règle générale chez les rouilles, cette phase d'urédospores mitotiques se répète de manière récurrente et permet au pathogène d'accroitre exponentiellement sa population. Les masses d'urédospores ainsi produites (reproduction asexuée), sont disséminées par le vent et transportées vers d'autres hôtes pour initier un nouveau cycle d'infection (Tessmann and Dianese, 2002).

Dans certaines circonstances, la production d'urédospores du pathogène change au profit de celle de téliospores, puis de basidiospores par reproduction sexuée. Cet évènement constitue une stratégie de survie en réponse à des conditions environnementales défavorables (paroi plus résistante des téliospores). Ce mode de reproduction permet des recombinaisons génétiques à l'origine de l'émergence de nouvelles formes de virulence qui renforcent les capacités adaptatives du pathogène.

Certaines étapes du cycle de vie d'*A. psidii* sont encore méconnues, tels que la formation et le rôle des basidiospores, la caryogamie, la méiose et l'ensemble des processus nucléaires capables de générer la variabilité génétique essentielle à l'émergence de nouvelles souches (Machado, 2015).

[17] Les **aecidies** et aeciospores sont morphologiquement identiques aux urédies et urédospores (Figueiredo, 2001).

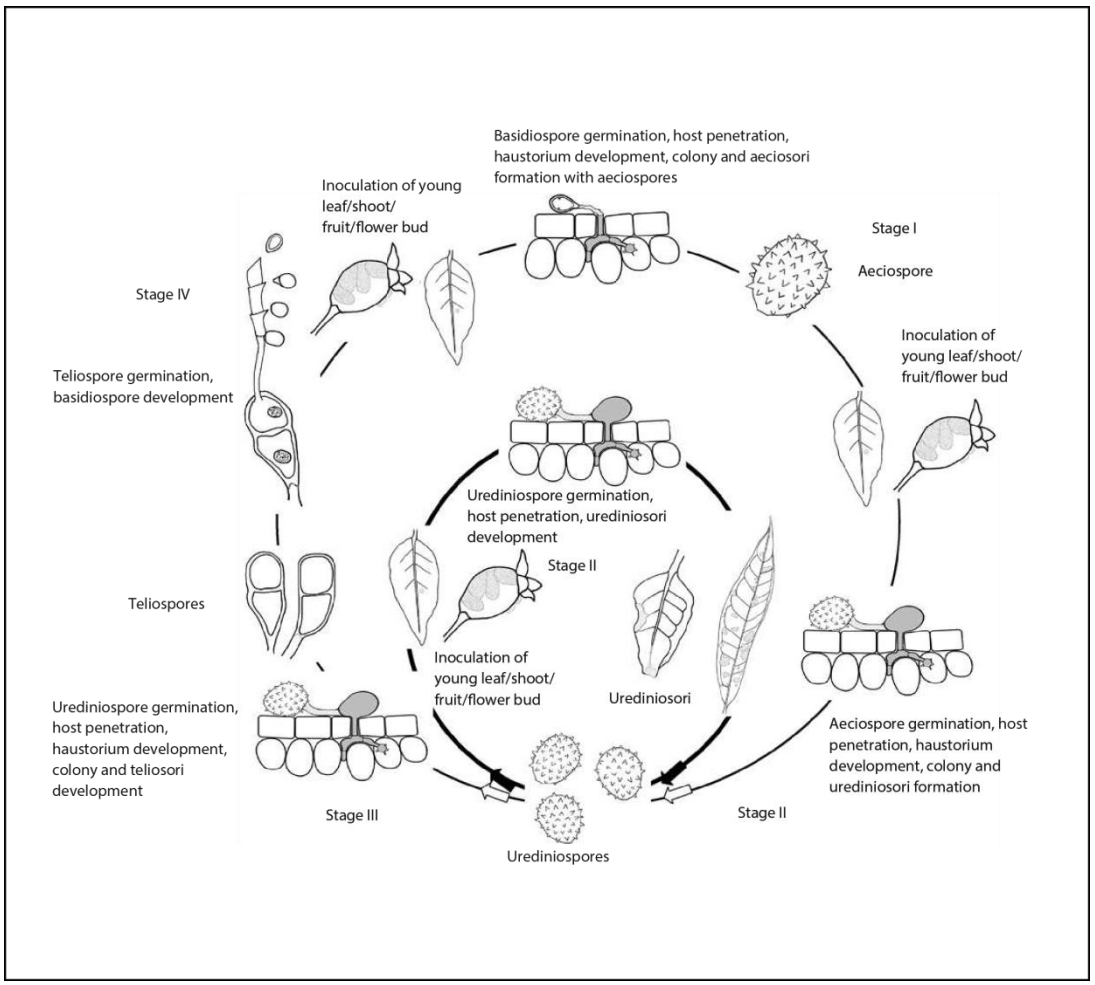


Figure 1.15 : Représentation schématique du cycle de vie d'*Austropuccinia psidii*.
D'après (Glen et al., 2007)

1.3.2.5. Description des symptômes

De nombreuses études décrivent les symptômes d'*A. psidii* sur divers hôtes. Elles s'accordent à dire que le champignon attaque préférentiellement les individus juvéniles et les organes en croissance de l'appareil aérien (feuille, bourgeons, fleurs, fruits, tige) (Coutinho et al., 1998 ; Tommerup et al., 2003). Les premiers symptômes visibles sont caractérisés par des lésions (évoluant du violet au gris) recouvertes de masses d'urédospores jaune-orangées d'aspect poudreux (et plus rarement des téliospores jaune-brun) (Figure 1.16) (Glen et al., 2007). Cette sporulation massive amoindrit les capacités photosynthétiques de la plante, mettant à mal les fonctions vitales de la plante pour sa croissance et sa survie (Alves et al., 2011). Dans les

stades les plus avancés de la maladie, l'infection peut conduire à une distorsion des organes infectés, une défoliation, un dépérissement des organes infectés, allant même jusqu'à la mort de l'hôte (Glen et al., 2007).

Les urédospores apparaissent le plus souvent d'abord sous la surface abaxial des feuilles, puis tendent à envahir l'ensemble de l'organe infecté. Ils peuvent être de dimension variable (19-27 x 15-26 µm), avec des formes ellipsoïdale à ovoïdes (Pegg et al., 2014a). La paroi des urédospores, d'une épaisseur d'environ 1.5-2.5 µm, est échinulée[18] et peut contenir ou non une tonsure. Les téliospores (22–42×15–23 µm) sont bruns foncés, cylindriques à ellipsoïdales avec un sommet arrondi et une surface lisse. Ils sont formés par deux cellules séparées et sont pédicellés (Carnegie, 2014 ; Pegg et al., 2014a).

1.3.2.6. Les facteurs épidémiologiques

Plusieurs études ont montré que la germination des spores d'*A. psidii* et le succès de l'infection dépendent des paramètres environnementaux (température, humidité foliaire, intensité lumineuse, photopériode) (Ruiz et al., 1989). Les conditions optimales pour la germination des spores incluent une forte humidité relative pendant plus de 8h à l'obscurité et des températures comprises entre 15-25°C pendant 5 à 7 jours (Alfenas et al., 2003a ; Ruiz et al., 1989). La sporulation peut être observée entre 10 et 12 jours après l'inoculation. En période hivernale, une période de latence de quatre à cinq semaines peut intervenir entre l'inoculation et la sporulation (Carnegie and Lidbetter, 2011). L'absence de tissus juvéniles et d'humidité foliaire, ainsi que des températures en dehors de la gamme de tolérance (10-30°C) inhibent complètement l'infection (Zauza et al., 2010a).

1.3.2.7. Diversité d'*A. psidii*

i. Variabilité physiologique des urédospores en lien avec la pathogénicité

La gamme d'hôtes très étendue d'*A. psidii* chez les Myrtaceae montre que ce pathogène, à l'image de nombreuses rouilles chez les céréales, possède des niveaux taxonomiques complexes intégrant la notion de race (syn. souche ; biotype, haplotype multilocus) (Tan et al., 2014). Ces races ont la capacité d'infecter et de causer la maladie soit sur un hôte, soit des groupes d'hôtes particuliers. Plusieurs expériences d'inoculations croisées d'urédospores d'*A.*

[18] Echinulée = recouvert de petites épines.

psidii sur des espèces de Myrtaceae différentes, ont ainsi montré l'existence d'une variabilité physiologique au sein des populations d'*A. psidii* d'Amérique du sud (Aparecido et al., 2003 ; Coelho et al.; Maclachlan, 1938). Cette variabilité physiologique des urédospores a permis d'identifier plusieurs biotypes de rouille, chacun infectant des combinaisons d'hôtes différentes (par exemple les urédospores n'infectant que les *Psidium guajava,* ceux n'infectant que les *Eucalyptus grandis* et *Psidium guajava* et *E. grandis* et enfin ceux n'infectant que les *Syzygium jambos* (Coelho, Alfenas, & Ferreira, 2001). Rayachhetry et al. (2001) ont aussi montré que deux isolats différents d'*A. psidii* prélevés sur *Melaleuca quinquenervia* et *Pimenta dioica* ne peuvent pas infecter *Syzygium jambos*. Cette variabilité physiologique des urédospores peut s'expliquer par un cortège de protéines émis par les spores qui diffère selon l'hôte infecté. Dans leur étude, Quecine et al. (2016) ont montré que les urédospores d'*A. psidii* infectant spécifiquement *Psidium guajava* possèdent de nombreuses protéines en lien avec la virulence fongique (malate déshydrogénase, sous-unités de protéosomes, les énolases, etc.) ; alors que pour les urédospores prélevées sur *Eucalyptus*, les protéines observées sont impliquées dans la biogenèse et la translocation des protéines. Ces différences montrent que les urédospores, et leur réservoir de protéines spécifiques, jouent un rôle clé dans l'établissement de la maladie et dépendent des structures physiques ou chimiques de l'espèce hôte infectée.

ii. Diversité génétique des populations d'A. psidii

La diversité génétique d'*A. psidii* a été étudiée dans de nombreux travaux à l'aide de marqueurs microsatellites (Simple Sequence Repeats) (Graça et al., 2011a ; Granados et al., 2017 ; Machado et al., 2015b ; Roux et al., 2016 ; Stewart et al., 2017 ; Zhong et al., 2008, 2011).

C'est ainsi que dans sa région native du Brésil, onze génotypes différents d'*A. psidii* ont pu être détectés par Zhong et al. (2011). Une autre étude a montré que la structure des populations de rouille brésilienne était également fortement influencée par les espèces hôtes (Graça et al., 2011a, 2013). Ainsi ils rapportent l'existence de cinq génotypes du pathogène spécifiquement associés avec certaines espèces hôtes. L'ensemble de ces résultats soutient l'hypothèse de l'existence d'espèces cryptiques au sein des populations d'*A. psidii* au Brésil.

D'autres études se sont intéressées à la diversité génétique du pathogène en dehors de sa zone d'origine et ont mis à jour l'existence de deux génotypes différents de ceux trouvés au Brésil. Le premier génotype est dit "pandémique" et se retrouve en Californie, à Hawaii, en Australie, en Chine, en Indonésie et en Nouvelle-Calédonie (Graça et al., 2011a ; Machado et

al., 2015b ; McTaggart et al., 2016 ; Ross-Davis et al., 2014). Le deuxième génotype diffère du génotype pandémique et ne se retrouve qu'en Afrique du Sud (Roux et al., 2016). L'origine de ces deux génotypes agissant en dehors de la région native de la rouille est inconnue. Néanmoins en 2017, le premier cas de génotype pandémique a été retrouvé en Colombie sur les espèces *Corymbia citrioda*, *Eucalyptus spp.* et *Syzygium jambos*, posant de nouvelles interrogations quant aux dynamiques spatiales de ce génotype pandémique (Granados et al., 2017).

1.3.2.8. Ressource génomique

Le génome d'*A. psidii* a été séquencé en 2014, sa taille est estimée entre 103 et 145 Mb (Tan et al., 2014). La taille de ce génome est comparable à celle d'autres rouilles (100-120 Mb pour *Puccinia tritica* et 110 Mb pour *P. striiformis*) (Tan et al., 2014). Une grande partie de ce génome est composé d'éléments transposables[19] (TE) (27%), connus pour générer la variabilité génétique nécessaire à l'adaptation des pathogènes (Spanu et al., 2010 ; Tan et al., 2014). Le nombre de protéines codées par le génome est estimé à plus de 19 000, avec une forte proportion de protéines associées à des fonctions moléculaires de liaison à l'ADN, d'intégration d'ADN ou de réplication d'ADN ARN-dépendante.

1.3.2.9. Modèle d'interaction avec Myrtaceae

Comme vu précédemment, *A. psidii* pénètre préférentiellement ses hôtes de manière directe, en perforant la paroi externe des cellules épidermiques. Le processus d'infection d'*A. psidii* a été étudié chez des génotypes sensibles et résistants de clones d'*Eucalyptus grandis* (Xavier et al., 2001). Cette étude a démontré que les phases préliminaires d'infection d'*A. psidii* se déroulaient de manière identique pour les deux génotypes : germination des urédospores, formation de l'appressorium et pénétration de l'hôte. Les premières différences histopathologiques sont observées après la phase de pénétration. La formation des hyphes primaires d'infection et des haustoria est retardée de 6h chez les génotypes résistants comparés aux génotypes sensibles.

[19] Elément transposable (transposons) = séquence d'ADN capable de se déplacer de manière autonome dans le génome par le biais du mécanisme de la transposition.

Chez les hôtes résistants, une réaction d'hypersensibilité autour des cellules colonisées par les haustoria se produit 48 h après l'infection, mettant fin à la propagation du pathogène dans la feuille (Figure 1.16).

Chez les hôtes sensibles, les hyphes continuent de se ramifier et de coloniser les cellules du mésophylle. Les premiers symptômes macroscopiques de la maladie apparaissent entre 3-5 jours après infection (hypertrophie des tissus colonisés). La sporulation commence à s'observer à partir de 6 jours après infection, et aboutira au bout de 12 jours à la formation d'un nouveau pool d'urédospores infectieux.

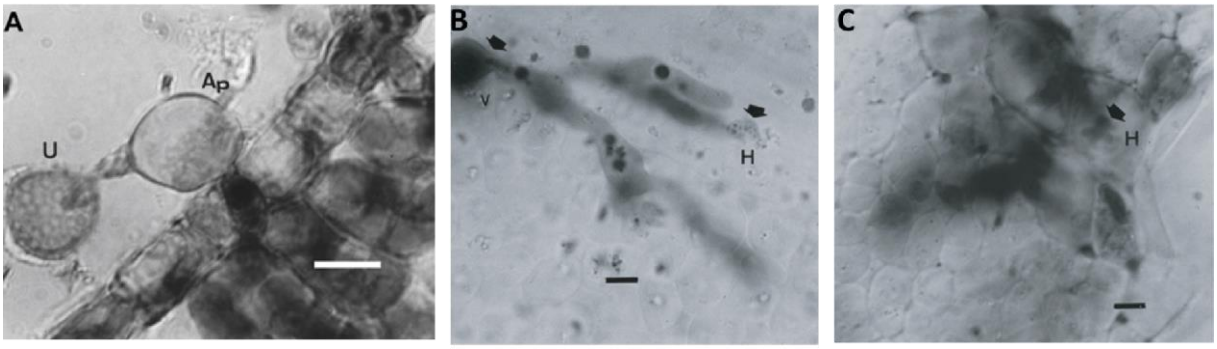


Figure 1.16 : Pathogénèse d'*Austropuccinia psidii* chez l'*Eucalyptus grandis.* (Xavier et al., 2001).
***A** : Pénétration directe des cellules épidermiques, Ap=Appressorium et U= urédospore ;*
***B** : Développement de A. psidii chez un génotype sensible et **C** : chez un génotype résistant, V= vésicule et H= haustorium.*
La coloration du cytoplasme (Bleu d'aniline+ bleu de trypan) montre la réaction d'hypersensibilité après la formation de l''haustorium chez le génotype résistant.
Echelle = 10 µm

1.4.3. Gestion de la maladie

1.4.3.1. Détection et identification du pathogène

Face à la menace globale que représente *A. psidii*, plusieurs moyens de prévention et de lutte ont été mis en place dans les diverses régions riches en Myrtaceae. Les mesures de quarantaines constituent le moyen de prévention le plus efficace face à l'introduction de pathogènes exogènes ou de nouvelles souches de pathogènes déjà introduits. La détection rapide et précise de la présence d'*A. psidii* dans une nouvelle aire de distribution est une étape cruciale dans la protection des zones à haut risque favorables à l'établissement du pathogène. Cette détection peut se faire grâce à des méthodes moléculaires. La PCR-nichée est la

première méthode mise au point pour la détection de la rouille (Langrell et al., 2008). Cette méthode se base sur l'emploi de deux paires d'amorces amplifiant spécifiquement la région ITS (Internal-transcribed spacer) de l'ADN nucléaire ribosomique d'*A. psidii*. Elle permet notamment de révéler la présence du pathogène dans divers tissus de plantes (feuilles, fruits, tige, etc.), y compris dans les échantillons asymptomatiques. Une seconde méthode de détection de la rouille se base sur la technique de la PCR en temps réel (Baskarathevan *et al.,* 2016). Ces deux méthodes constituent un moyen rapide pour détecter la présence du pathogène dans un grand nombre d'échantillons.

L'identification formelle et définitive du pathogène peut être effectuée par séquençage de l'ADN extrait de ces spores. Les séquences amplifiées de régions génomiques définies sont comparées à l'épitype et aux autres séquences disponibles d'*A. psidii* sur GenBank (Kawanishi et al., 2009 ; Machado, 2015). L'approche par barcoding des régions ITS de l'ADNr est communément utilisée pour cette identification. D'autres régions peuvent également être utilisées, telle que les séquences des gènes *β-tubulin*, *elongation factor* 1α et de la région LSU de l'ADNr.

1.4.3.2. Moyens de lutte chimique

L'application de produits phytosanitaires selon des protocoles établis, est un moyen efficace pour lutter contre la rouille au sein d'espaces restreints et confinés. Le succès de l'utilisation des fongicides a ainsi contribué à sauvegarder la filière bois d'*Eucalyptus* en Amérique du Sud. Pour la culture des goyaviers, l'application quotidienne de fongicides a également permis de limiter les dommages causés par la maladie à des niveaux soutenables pour la filière (Ribeiro and Pommer, 2004). Les substances actives utilisées dans les fongicides, comme moyen préventif et/ou curatif contre la rouille, comprennent : l'azoxystrobine, l'oxychlorure de cuivre, le mancozeb, l'oxycarboxine, le propiconazole, le triadimenol et la triforine (Glen et al., 2007). Une rotation de ces produits et une application régulière doivent être effectuées afin de garantir leur efficacité, ce qui accroit les coûts de production (Ribeiro and Pommer, 2004). Les bénéfices de l'application de fongicides doivent être évalués en fonction des coûts économiques et écologiques qu'ils occasionnent.

1.4.3.3. Lutte par la sélection de plants résistants à la rouille

Selon Glen et al., 2007, la sélection de plants résistants à la rouille représente un moyen viable économiquement pour contrôler les dommages occasionnés par *A. psidii,* une fois que le

pathogène est établi. C'est le moyen de lutte préférentiel contre la rouille appliquée en Amérique du Sud. Il permet de limiter l'application des fongicides (Old, 2002). Au Brésil, des cultivars de goyaviers ont été sélectionnés pour leur capacité à résister à la rouille (Ribeiro and Pommer, 2004), et la sélection de clones hybrides résistants à la rouille chez les espèces cultivées d'*Eucalyptus* (notamment *E. grandis*) est devenue une priorité de recherche (Xavier et al., 2001).

i. Variabilité inter- et intraspécifique de la résistance à la rouille

Le prérequis essentiel à tout programme de sélection est l'existence de variations phénotypiques liées à de la diversité génétique. Ainsi de nombreux travaux se sont attachés à étudier les niveaux de vulnérabilités de plusieurs hôtes de Myrtaceae face à l'attaque d'*A. psidii*, à la fois dans le milieu naturel, et en conditions contrôlées. Les premiers suivis relatifs à l'étude de la résistance à la rouille ont été menées prioritairement sur les espèces cultivées d'*Eucalyptus* au Brésil et ont conclu à l'existence d'une variabilité interspécifique de la résistance (entre les espèces) et entre les provenances pour les mêmes espèces (Dianese, 1984). Cette variabilité interspécifique de la réponse à la maladie a été classifiée en cinq catégories sur la base du nombre de plants infectés par taxon, et du nombre de pustules par plant (hautement sensible, sensible, modérément sensible, résistant, totalement résistant).

Dans leur étude Zauza et al. (2010b), ont évalué les niveaux de résistance à la rouille de plusieurs espèces de *Corymbia*, *Eucalyptus* et *Melaleuca* provenant de localités différentes (Australie, Indonésie, Papouasie-Nouvelle-Guinée, Thailande et Vietnam) et inoculés en conditions contrôlées avec une souche d'*A. psidii* brésilienne (UFV-02-Race 1). Douze jours après l'inoculation, des variations dans l'expression phénotypique de la résistance à la rouille ont été observées entre ces espèces (variabilité interspécifique), de même qu'entre les individus d'une même espèce (variabilité intraspécifique) indépendamment de la provenance. Ces deux niveaux de variabilité ont également pu être confirmés par Morin et al. (2012) ainsi que Pegg et al. (2014b) sur des Myrtaceae australiennes. Ils ont aussi démontré l'absence de lien entre le développement des symptômes et les relations phylogénétiques entre les taxons (Morin et al., 2012).

Selon Tobias et al. (2015), la variabilité intraspécifique est une preuve de la capacité d'*A. psidii* à contourner la résistance médiée par les MAMP. De même que l'existence de mécanismes de résistance médiés par les gènes R chez les individus résistants (Thumma et al., 2013) et le déclenchement rapide de la HR suite à la reconnaissance du pathogène (Xavier et al., 2001).

ii. Gène de résistance

En se basant sur la variabilité des réponses du pathosystème *E. grandis* - *A. psidii* et la méthode d'analyse de ségrégants regroupés (Bulked segregant analysis[20]), Junghans et al. (2003) ont démontré que le contrôle des mécanismes de résistance à la rouille était lié à l'action d'un gène majeur R modulée par des gènes mineurs. Les plantes possédant le gène R et la combinaison favorable de gènes mineurs exprimeront ainsi un phénotype immunisé contre la rouille. Ce gène R a été dénommé *Ppr*1 (*Puccinia psidii resistance gene*-1) et cartographié avec six marqueurs RAPD. *Ppr*1 est ainsi devenu le premier gène de résistance à une maladie identifiée chez l'*Eucalyptus*, le processus de résistance qui y est lié se rapproche du modèle de la résistance qualitative retrouvé chez le Pin (Kinloch et al., 1999) et le Peuplier (Lefèvre et al., 1998) en réponse à des pathogènes avec lesquels ils n'ont pas co-évolué (Junghans et al., 2003). Plusieurs autres études ont confirmé l'existence de ce gène, son rôle majeur dans la résistance à la rouille pour l'*Eucalyptus* et ont également affiné sa position sur le génome de l'hôte. En se basant sur des marqueurs microsatellites, *Ppr*1 a été positionné sur le groupe de liaison 3 de la carte génétique consensus de référence de l'*Eucalyptus* (Brondani et al., 2006 ; Lima et al., 2011 ; Mamani et al., 2010 ; Rosado et al., 2010). Cependant, ces marqueurs liés au locus *Ppr1* ne peuvent être utilisés que dans le cas de croisements spécifiques. Les populations d'arbres forestiers contiennent généralement plusieurs familles non apparentées et il est peu probable que ces marqueurs puissent être utilisé dans d'autres familles que celles de référence, ou d'autres espèces, en raison de la rupture de liaison entre les marqueurs et *Ppr*1 (Mamani et al., 2010). En 2011, l'approche de lutte contre la rouille de l'*Eucalyptus* basée sur la seule identification de *Ppr*1 s'effondre du fait de l'apparition d'une nouvelle souche d'*A. psidii* plus virulente (Graça et al., 2011b). Ce contournement rapide de la résistance (en moins de 10 ans) illustre l'urgence d'améliorer les connaissances sur les bases moléculaires de la résistance à *A. psidii* au-delà de l'*Eucalyptus*, et de les étendre à d'autres espèces de Myrtaceae afin de trouver de nouvelles sources de résistance.

Thumma et al. (2013) recommande de se baser sur des études d'association à l'échelle populationnelle pour identifier des marqueurs directement liés au locus de résistance (SNPs) et utilisables pour des génomes différents. C'est ainsi qu'ils ont mis en évidence 13 SNPs situés sur plusieurs chromosomes de l'*Eucalyptus*, au niveau de loci codant pour une

[20] **Bulked segregant analysis** : méthode permettant l'identification rapide de marqueurs dans des régions spécifiques du génome. La méthode compare deux groupes d'ADN provenant d'individus au phénotype contrastés et issu d'un même croisement. Les deux groupes sont analysés pour identifier des marqueurs permettant de les distinguer (en général des RAPD pour Restriction Fragment Length Polymorphism) (Michelmore et al., 1991).

Glutathione S-transferase, des protéines NB-LRR et des facteurs de transcription (MYB, bZIP et WYRKY).

Chapitre II : Etat des lieux de la présence d'*Austropuccinia psidii* en Nouvelle-Calédonie : distribution, gamme hôtes et diversité génétique.

Ce chapitre fait l'objet d'un article scientifique soumis dans la revue Forest Pathology et accepté avec modifications mineures.

2.1. Introduction générale du Chapitre 2

Les écosystèmes insulaires abritent une large proportion d'espèces endémiques, rares, menacées et particulièrement vulnérables aux impacts d'envahisseurs biologiques (Bisconti et al., 2013 ; Keitt et al., 2011). En effet, bien que représentant seulement 5% de la surface en terres émergées, les îles regroupent environ un quart des espèces de plantes vasculaires connues (Kier et al., 2009). Comme vu précédemment dans la synthèse bibliographique, la rouille *A. psidii* est considérée comme la principale menace pour les plantes appartenant à la famille des Myrtaceae. La zone tropicale de l'Asie-Pacifique abrite une riche diversité biologique de Myrtaceae (Loope, 2010), augmentant de fait la probabilité pour *A. psidii* de trouver des espèces apparentées à ses hôtes d'origine. L'historique des invasions d'*A. psidii* a de part-ailleurs montré les capacités de ce pathogène à établir de nouvelles interactions hôte-pathogène dans une gamme vaste de conditions climatiques. Dans leurs travaux de modélisation de niches climatiques Kriticos et al. (2013) identifiaient de nombreuses îles du Pacifique Sud, dont la Nouvelle-Calédonie, comme particulièrement favorables à l'établissement d' *A. psidii*. Ces prédictions se sont révélées exactes puisqu'*A. psidii* a été détectée la même année sur le territoire néo-calédonien, au sein de la commune de Farino. Peu de temps après cette détection, de nombreux témoignages attestent d'une distribution bien plus vaste à l'échelle du territoire.

Le but de ce travail qui a débuté en 2014, est d'établir un premier état des lieux de la dispersion d'*A. psidii* en Nouvelle-Calédonie. Il s'agit tout particulièrement d'identifier :

- la distribution géographique de la rouille en Nouvelle-Calédonie ainsi que sa gamme d'espèces hôtes ;
- le niveau de diversité génétique d'*A. psidii* sur le territoire ;
- les différents degrés de virulence de la maladie en fonction des hôtes concernés dans des conditions semi-contrôlées.

L'ensemble des résultats acquis formera une base de données qui permettra de cibler plus spécifiquement les milieux et les espèces à risque.

2.2. Article: The impact of *Austropuccinia psidii* in New Caledonia, a biodiversity hotspot.

Soewarto Julia[1], Carriconde Fabian[1], Hugot Nicolas[2], Bocs Stéphanie[3,4], Hamelin Chantal[3,4], Maggia Laurent[1,3,5]

[1] *Institut Agronomique néo-Calédonien (IAC), "Equipe Sol & Végétation" (SolVeg), BP 18239, 98800 Nouméa, New Caledonia*

[2] *Chambre d'Agriculture Nouvelle-Calédonie, Groupement de Défense Sanitaire Végétal (GDS-V), 98845, Nouméa, Nouvelle-Calédonie*

[3] *AGAP, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.*

[4] *CIRAD, UMR AGAP, F-34398 Montpellier, France.*

[5] *CIRAD, UMR AGAP, F-98800 Nouméa, Nouvelle-Calédonie, France.*

*Corresponding authors:

J. Soewarto (jsoewarto@hotmail.com)

L. Maggia (laurent.maggia@cirad.fr)

2.2.1. Abstract

Myrtle rust, caused by the fungus *Austropuccinia psidii*, infects a wide range of host species within the Myrtaceae family worldwide. Since its first report in 2013 from New Caledonia, the extent of the host range, geographical distribution and genetic diversity of *A. psidii* in this territory have been unknown. This study reports 67 new host species distributed in 14 different Myrtaceae genera, including five new genera. The pathogen was found in various types of plant communities where Myrtaceae species are the dominant or codominant species, as well as in several nurseries. It is now considered a significant threat to the biodiversity of ecosystems and the Myrtaceae-related economy. A three-year disease monitoring trial with 35 cultivated Myrtaceae species showed variations in disease severity at the species and individual scales. This suggests that some level of genetic resistance to the rust may possibly exist among the endemic Myrtaceae of New Caledonia. A population genetic investigation revealed that only one single genotype of *A. psidii* occurs across its entire host range in New Caledonia. Therefore, it is essential to carry out population genetic surveillance to identify the appearance of mutations or the introduction of potentially more virulent genotypes of *A. psidii*. The outcomes from this study will assist with the ongoing management of the disease in New Caledonia.

Keywords: *Austropuccinia psidii*, myrtle rust, New Caledonia, host range, genetic diversity

2.2.2. Introduction

In natural ecosystems, reports of exotic plant pathogens have increased over the past few years and are one of the leading causes of the decline in dominant tree species, resulting in permanent changes in the species composition of ecosystems (Desprez-Loustau et al., 2016; Ellison et al., 2005; Ennos, 2015; Orwig, 2002). Causal agent of the disease myrtle rust is *Austropuccinia psidii* (Sphaerophragmiaceae, Pucciniales) has become a major threat to members of the Myrtaceae plant family worldwide, including several dominant woody species and key commercial species(Beenken, 2017; Glen et al., 2007; Morin et al., 2012). This pathogen is considered to be unusual among the rusts (Pucciniales) in having a very broad host range, which includes at least 460 host species from 73 genera across the Myrtaceae plant family (Carnegie et al., 2016b; Giblin and Carnegie, 2014).

A. psidii is native to South and Central America and was first described in 1884 on *Psidium guajava* in Brazil (Coutinho et al., 1998). Symptoms of infection caused by *A. psidii* are characterized by powdery yellow pustules that can cover actively growing leaves and shoots as well as fruits and flower buds (Glen et al., 2007). Infection by *A. psidii* can lead to leaf distortion, branch defoliation, shoot-tip dieback and plant death (Carnegie et al., 2016a). The pathogen has caused severe outbreaks on densely cultivated and economically important hosts from South and Central America, such as *Pimenta dioica* (allspice) in Jamaica and *Psidium guajava* in Brazil (Maclachlan, 1938; Ribeiro and Pommer, 2004). It was also found on introduced eucalypts in Brazil in the 1930s and resulted in serious losses to this economically important forest resource (Alfenas et al., 2003b; Coutinho et al., 1998). During the last decade, *A. psidii* has invaded many climatically favorable regions around the world where Myrtaceae occur: Florida (Rayachhetry et al., 1997), California (Mellano, 2006), Hawaii (Uchida et al., 2006), Japan (Kawanishi et al., 2009) , China (Zhuang, J-Y. & Wei, S-X.), Australia (Carnegie et al., 2010; Potts et al., 2016), South Africa (Roux et al., 2013), New Caledonia (Giblin, 2013), Indonesia (McTaggart et al., 2015), Singapore (du Plessis et al., 2017) and New Zealand (Department of Conservation, May 2017).

The population structure of *A. psidii* has been investigated around the world using microsatellite markers originally designed by Zhong et al. (2008). Brazil, one of the putative centers of *A. psidii* diversity, is home to at least five major genotypes that were specifically associated with particular host species (Graça et al., 2013).

A number of recent studies showed that at least two other genotypes of unknown origin have spread globally and are able to infect both native and exotics hosts. A unique genotype of *A. psidii* was reported in South Africa (Roux et al., 2016), and another different genotype

considered to be “pandemic” was reported in areas of the Asia-Pacific region, including Australia, California, China, Hawaii, Indonesia, Japan and New Caledonia (Machado et al., 2015b; Ross-Davis et al., 2014; Stewart et al., 2017). In 2017, this pandemic genotype was found for the first time in Colombia, which raised new questions about its origin from a native host from Central or South America or from an external introduction (Granados et al., 2017).

Since the spread of the pandemic *A. psidii* genotype in Asia-Pacific areas, there has been increasing concern about the deleterious effect this pathogen could have on non-native island ecosystems where Myrtaceae occur. Species from such ecosystems are believed to be highly susceptible to this rust because they lack co-evolutionary recognition processes or defense mechanisms (Dodds and Thrall, 2009; Morin et al., 2012).

New Caledonia is a French archipelago located in the Southwest Pacific. It comprises a main island, called “Grande Terre”, and small islands, including the “Loyauté Islands”. New Caledonia is renowned for its extraordinary botanical richness, which is ranked third in terms of world endemism (Gomez et al., 2015; Morat, 1993). The flora of New Caledonia counts approximately 257 species of Myrtaceae, of which 99% are endemic, which makes Myrtaceae the most species-rich angiosperm family in this territory (Grandcolas et al., 2008; Kier et al., 2009; Morat et al., 2012; Myers et al., 2000). In terms of abundance, Myrtaceae is one of the families that dominate or co-dominate the different types of ecosystems that occur on the archipelago, especially in rainforests, sclerophyllous forests, savannas and maquis shrublands (Bouchet et al., 1995; Read et al., 2000). As a global biodiversity hotspot, New Caledonia has set priorities for the conservation of biodiversity in the face of the various threats posed by mines, fires and invasive species. In this particular context, the spread of *A. psidii* is of a great concern given the combination of the richness of endemic Myrtaceae species, the abundance of potential hosts and the tropical climatic conditions that predominate in New Caledonia. In addition to the unpredictable effects of this rust on biodiversity and natural environments, it is important to consider the patrimonial, cultural and economic importance of Myrtaceae to New Caledonia. Fruit consumption, traditional medicine, landscaping and construction are some of the common uses of Myrtaceae species and a number of endemic Myrtaceae species have also been extensively used as pioneer species in revegetation programs, contributing to the restoration of ecosystems degraded by mining activities.

In this context and after three years of detection of *A. psidii* in New Caledonia, the present work aims to (i) determine the geographical spread and host range of *A. psidii* on the archipelago, (ii) characterize the genetic variability within *A. psidii* across a range of different

plant hosts and habitats, (iii) evaluate the disease severity on several cultivated Myrtaceae species under natural infections.

2.2.3. Materials and methods

2.2.3.1. Disease surveys and sampling of *A. psidii*

An epidemiological surveillance program for myrtle rust disease in New Caledonia was conducted from 2013 to 2015 by the governmental organization of the Department of Veterinary, Food and Rural Affairs (DAVAR). This program included public and industry communications to raise awareness of myrtle rust. Host species showing symptoms of myrtle rust were reported by nursery staff, researchers and members of the public over the entire territory. These reports were centralized in a public database after being verified from photos. We supplemented this myrtle rust database with additional reports, especially from natural habitats. A sample from each infected specimen was collected during these additional surveys were, pressed and dried prior to examination by a botanist to confirm the host species.
Isolates of *A. psidii* were collected from different host species from across New Caledonia, this included natural field sites and nurseries. Leaves, shoots, fruits and/or flowers were taken from the infected plants and stored with silica beads prior to DNA extraction.

2.2.3.2. DNA extraction, PCR and sequencing

Fungal (single pustule) and host plant genomic DNA was extracted from infected leaves using the Qiagen Plant DNeasy Mini Kit (Qiagen Inc., Valencia, CA) following the manufacturer's instructions. The internal transcribed spacer region of the nuclear ribosomal DNA (ITS), a portion of the ß-tubulin gene (ßtub) and a portion of the translation elongation factor 1- α gene (EF1-α) were used for DNA sequence-based identification of *A. psidii* ((Machado et al., 2015a)). The primers used for these loci were: ITS-1F (Gardes and Bruns, 1993) and Psi6 (Langrell et al., 2008), Ppsi-ßtubF and Ppsi-ßtubR (Machado et al., 2015a), and PPEFF and PPEFR (Machado et al., 2015a), respectively. Reaction conditions (in 25 µL final volume) and PCR amplification were undertaken as described by Machado et al., (2015a). The PCR products were purified on Sephadex® G-50 columns in multiscreen filter plates (Millipore). The purified products were then amplified in both forward and reverse directions using the same primers as for the PCR and the ABI BigDye Terminator v3.1 cycle sequencing kit

(Applied Biosystems). Sequencing was then performed on the ABI 3730xl automated DNA sequencer at Plateforme du Vivant de Nouvelle-Calédonie (PFV-NC).

The assembly and editing of the sequences using both forward and reverse sequences were conducted using Sequencher v.4.10.1 (Gene Codes Inc., Ann Arbor, MI, USA). Sequences from each locus were aligned with MEGA6 (Tamura et al., 2013) and submitted to a BLASTn search on the NCBI database (http://blast.ncbi.nlm.nih.gov).

2.2.3.3. Microsatellite genotyping

Fifteen microsatellite loci that previously showed diversity in populations of *A. psidii* were tested (Zhong et al., 2008): PpSSR010, PpSSR012, PpSSR014, PpSSR018, PpSSR022, PpSSR078, PpSSR80, PpSSR087, PpSSR102, PpSSR136, PpSSR146, PpSSR161, PpSSR176, PpSSR195, and PpSSR208. For each primer pair, a fluorescent-labeled oligonucleotide was added to the 5' end of the forward primer (6-FAM™, VIC®, NED™, PET™). The reaction conditions (10 µL total volume) involved the following contents: 500 µM of dNTPs (5-Prime), 0.5 U of DNA Taq polymerase (GoTaq® Flexi DNA Polymerase M829, Promega™), 1 X reaction buffer (Promega™), 2 to 3 mM of $MgCl_2$ (Promega™) depending on the locus, 0.2 µM of each primer (forward and reverse), sterile water and 3 µL of genomic DNA (at a standard concentration of 20 ng μL^{-1}). PCR amplifications were performed using the following program: 95°C for 5 min, then 35 cycles of 94°C for 15 s, 49-58°C (depending on the locus) for 15 s, and 72°C for 45 s, followed by a final extension at 72°C for 5 min. Fragment analysis was conducted on an ABI 3730xl automated sequencer (Plateforme du Vivant de Nouvelle-Calédonie (PFV-NC)). Allele sizes were determined using the Genescan™ 500 LIZ™ dye size standard (Applied Biosystems) and analyzed with GeneMapper™ software v. 3.5 (Applied Biosystems).

2.2.3.4. Myrtle rust symptoms monitoring trial under natural infection

A disease monitoring trial was conducted between DAVAR, the Groupement de Défense Sanitaire du Végétal and the Institut Agronomique néo-Calédonien (IAC) from October 3, 2013, to September 4, 2016. The objective was to examine the progress of disease symptoms over time on several cultivated species of Myrtaceae to *A. psidii* and under natural infection.

The trial was conducted in a private field nursery (165.772024; -21.663800) at the Farino location, where *A. psidii* has been present since March 2013.

The trial included 35 cultivated Myrtaceae species that are mostly endemic to New Caledonia and represented 9 tribes of the Myrtoideae subfamily (Thornhill et al., 2015; Vasconcelos et al., 2017) (Table 2.1). Healthy plants of each species were sourced from different commercial nurseries at a juvenile stage (less than one year) and free of myrtle rust symptoms. Plants were derived from seed or from cuttings, depending on the species and the nursery of origin. Six individuals per species were used, except when there was insufficient nursery stock available. Individual plants were repotted in 5 L pots and watered twice a day using an automatic drip watering system. The field nursery soil was covered with plastic tarp, and no fungicides or phytosanitary products were applied against the rust to avoid bias in terms of the species response to the disease. The field nursery is not located in a confined area, which allows natural infection processes by surrounding infected trees in the natural field. The field nursery is surrounded within a radius of 100 meters by a population of *Syzygium jambos* infected by myrtle rust. The presence of these trees near the field nursery is sufficient to induce infection on the juvenile Myrtaceae plants. However, as a natural inoculum, the amount of spores at the time of infection on each plant is not controlled. However, we assume that a three years monitoring is sufficient to observe rust symptoms, at least once, on all susceptible individuals of the experiment. We assessed every two weeks the progress of myrtle rust symptoms by counting the number of diseased individuals from each species and scoring the disease severity on each individual plant. The disease severity was evaluated on the youngest leaves of each individual and according to the disease index for *A. psidii* on *Melaleuca quinquenervia* (http://www.canc.nc/fileadmin/documents/canc/internet/a_telecharger/GDSV/Doc_techniques_GDSV/Synth%C3%A8se_finale_test_de_s%C3%A9v%C3%A9rit%C3%A9_rouille_08_10_2014.pdf). Briefly, this disease index includes 6 scores of severity based on the extent of *A. psidii* infection on the leaves: 0 (immunity, no evident symptom of the disease); 1 (minor infection with small sized lesions covering less than 10% of the total new leaf); 2 (minor to moderate infection covering less than 25% infection of the leaf); 3 (moderate infection on 25-50% of the leaf surface and possibly extending on the stem); 4 (severe infection on 50-100% of the leaf surface with possibly leaf distortion); 5 (severe infection; leaf, shoot and blackening and abscise); 6 (severe infection; dead shoots).

Table 2.1: List of Myrtaceae species monitored for myrtle rust symptoms under natural infection and disease assessments parameters.

Species	Plant source	Number of replicates	Status [a]	Maximal disease severity score	Disease incidence (%)	Number of death[b]	
Melaleuca vinimalis var. *captain cook*	*seeds*	6	I	0	0	0	**Resistant**
Pleurocalyptus pancheri	*seeds*	4	G	0	0	0	
Psidium guajava	*seeds*	6	I	0	0	0	
Syzygium aggregatum	*seeds*	6	E	0	0	0	
Syzygium macranthum	*seeds*	6	E	0	0	0	
Syzygium malaccense	*seeds*	6	A	0	0	0	
Syzygium sp. La Confiance	*seeds*	6	E	0	0	0	
Syzygium australe	*cuttings*	6	I	0	0	0	
Sannantha pinifolia	*cuttings*	4	E	0	0	0	
Xanthostemon myrtifolius	*cuttings*	6	E	0	0	0	
Sannantha leratii	*cuttings*	6	E	1	17	0	**Susceptible**
Eugenia balansae	*seeds*	6	E	1	33	0	
Syzygium sp. Karikouié	*seeds*	6	E	1	33	0	
Syzygium ngoyense	*seeds*	6	E	1	33	0	
Sannantha procera	*seeds*	6	E	1	50	0	
Syzygium fructescens	*seeds*	6	E	1	67	0	
Syzygium acre	*seeds*	5	E	2	80	0	
Syzygium longifolium	*seeds*	6	E	3	33	0	
Tristaniopsis guillainii	*seeds*	6	E	3	50	0	
Syzygium baudouini	*seeds*	6	E	3	67	0	
Syzygium wagapense	*seeds*	6	E	3	100	0	
Syzygium eugenioides	*seeds*	6	E	3	100	0	
Metrosideros operculata var. *francii*	*cuttings*	6	E	4	100	0	
Metrosideros operculata var. *operculata*	*cuttings*	5	E	4	100	0	
Metrosideros punctata	*seeds*	6	E	4	83	0	
Metrosideros laurifolia var. *laurifolia*	*seeds*	3	E	4	100	0	
Arillastrum gummiferum	*seeds*	6	G	5	33	0	
Xanthostemon laurinus	*seeds*	6	E	5	50	0	
Xanthostemon aurantiacus	*seeds*	6	E	5	50	0	
Tristaniopsis glauca	*seeds*	6	E	5	100	1	
Eugenia noumeensis	*seeds*	6	E	5	100	1	
Cloezia artensis var. *artensis*	*seeds*	6	G	5	100	0	
Myrtastrum rufopunctatum	*cuttings*	5	G	6	100	4	
Melaleuca quinquenervia	*seeds*	6	A	6	100	1	
Eugenia sp. Maa	*seeds*	6	E	6	100	4	

[a] E: endemic species; A: indigenous species; G: endemic genera; I: introduced species (source: FLORICAL vers. 22.IV.2016)

[b] Number of death following several cycle of infection of *A. psidii*.

2.2.4. Results

2.2.4.1. Identity confirmation

We sequenced 8 isolates collected from 4 host species and 6 distant locations in New Caledonia. All the sequences were identical for the three loci (ITS, β-tubulin and EF1-α). Sequence accession numbers are given in Supplementary table 2.1. The BLASTn results from the NCBI database suggested that the rust strain was most closely related to *Austropuccinia psidii*. We compared the New Caledonian sequences to the reference sequences from the *A. psidii* epitype, which are KM282154 (ITS), KM282123 (β tubulin) and KM282143 (EF1-α) (Machado et al., 2015a). The ITS sequences of *A. psidii* from New Caledonia were 99% identical to KM282154 (e.g., 523 out of 525 base pairs for LC228282); the β-tubulin sequences were 99% identical to KM282123 (e.g., 744 out of 745 base pairs for LC228295); and the EF1-α sequences were 99% identical to KM282143 (e.g., 558 out of 561 base pairs for LC228298).

2.2.4.2. Geographical distribution of A. psidii in New Caledonia

The myrtle rust database contained 373 disease reports from 115 localities (nurseries, gardens, private properties and natural field sites) (Figure 2.1). It should be noted that the number of occurrences per year does not reflect the progression of the disease in the territory but rather a declining sampling effort. Myrtle rust was frequently found in nurseries, representing 26% of the total number of disease reports. At least 29 commercial nurseries reported the presence of myrtle rust, with serious damage to juvenile plants. The disease has also spread widely in natural field sites, including six types of ecosystems: maquis of low and medium altitudes on ultramafic soils, savannas and other secondary formations on non-ultramafic soils, dense humid forests of low and medium altitudes on sedimentary volcanic soils, dense humid forests of low and medium altitudes on ultramafic soils, dense humid forests on calcareous soils, and dry forests (Figure 2.2).

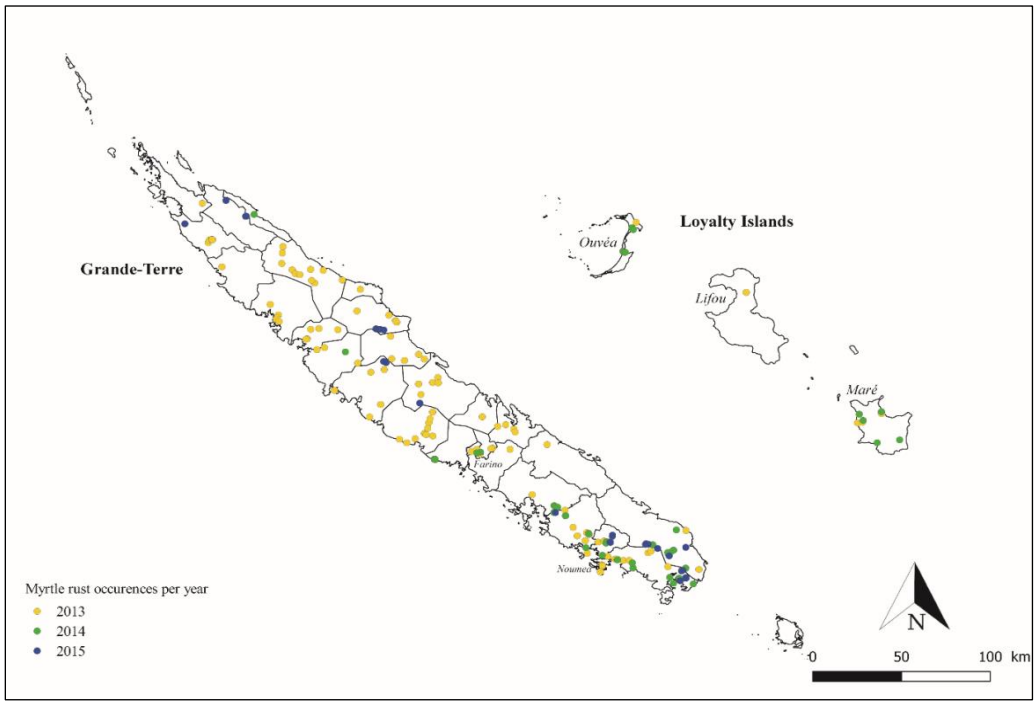


Figure 2.1 : Map of myrtle rust reports in New Caledonia according to years (2013: 147 reports; 2014: 136 reports and 2015: 89 reports).

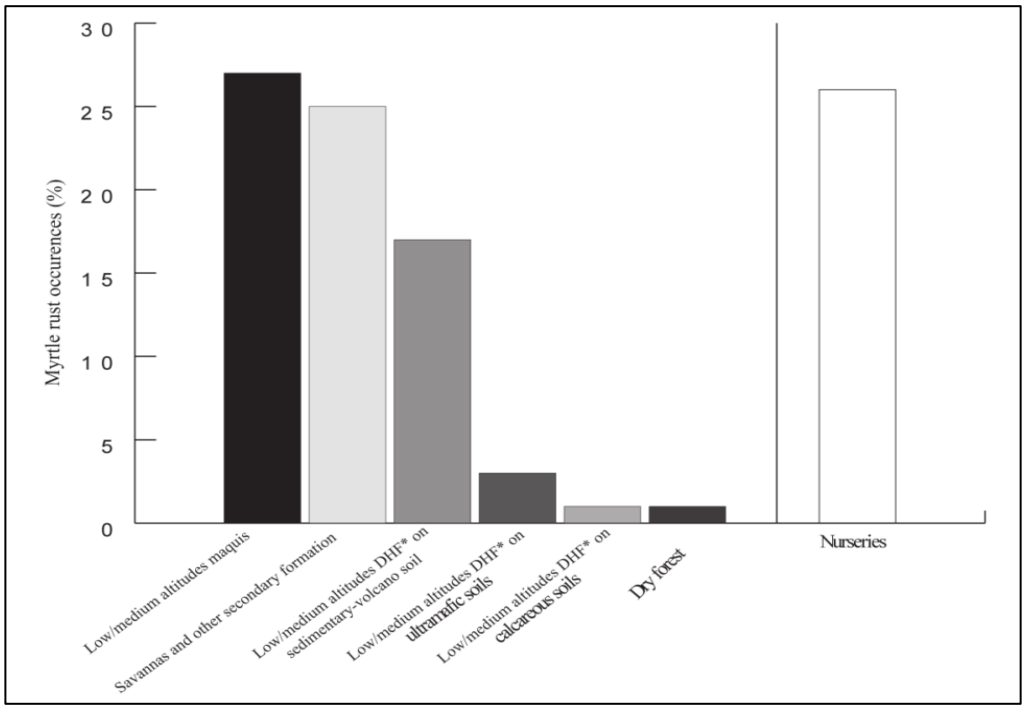


Figure 2.2 : Distribution of myrtle rust reports repartitions in terms of environment type: DHF: Dense Humid Forest.

The highest disease incidence (in terms of number of disease occurrences) was reported in maquis of low and medium altitudes, with 27% of the total number of disease records. Savannas and other secondary formations together account for 25% of the total number of disease reports. Dense and humid forests of low and medium altitudes on sedimentary volcanic soils represent 17% of the disease reports, followed by 3% of disease reports from the same kind of formation on ultramafic substrate and 1% on calcareous soil. The remaining 1% of the disease reports includes dry forests and dense humid forests of low to medium altitudes on ultramafic soils and calcareous soil.

2.2.4.3. Host range of *A. psidii*

We identified 70 plant species infected by *A. psidii*. These 70 species belong to 15 different Myrtaceae genera (Table 2.2, Supplementary table 2.2). Among them, three species were already recorded as hosts of *A. psidii* in locations other than New Caledonia (*Melaleuca quinquenervia, Syzygium jambos* and *Syzygium cumini*), and 67 correspond to new records (Supplementary table 2.2). The present work expands the host range of *A. psidii* to five new genera: *Cloezia, Stereocaryum, Arillastrum, Sannantha* and *Myrtastrum*. We also recorded myrtle rust symptoms on 18 Myrtaceae samples (*Myrtaceae* spp, Table 2.2) that remain unidentified because of the lack of reproductive structures that could have allowed botanical identification.

Nine rare and endangered species were found to be infected with myrtle rust: *Eugenia balansae, E. daenikeri, E. excorticata, E. munzingeri, E. noumeensis, Cloezia artensis* var. *riparia, Metrosideros punctata, Tristaniopsis polyandra* and *T. reticulata*. *A. psidii* infection was also recorded on the flower buds, flowers and fruits of seven species: *Eugenia bullata, E. excorticata, E. gacognei, E. paludosa, E. stricta, Metrosideros operculata* var. *francii* and *Cloezia artensis* var. *artensis* (Figure 2.3).

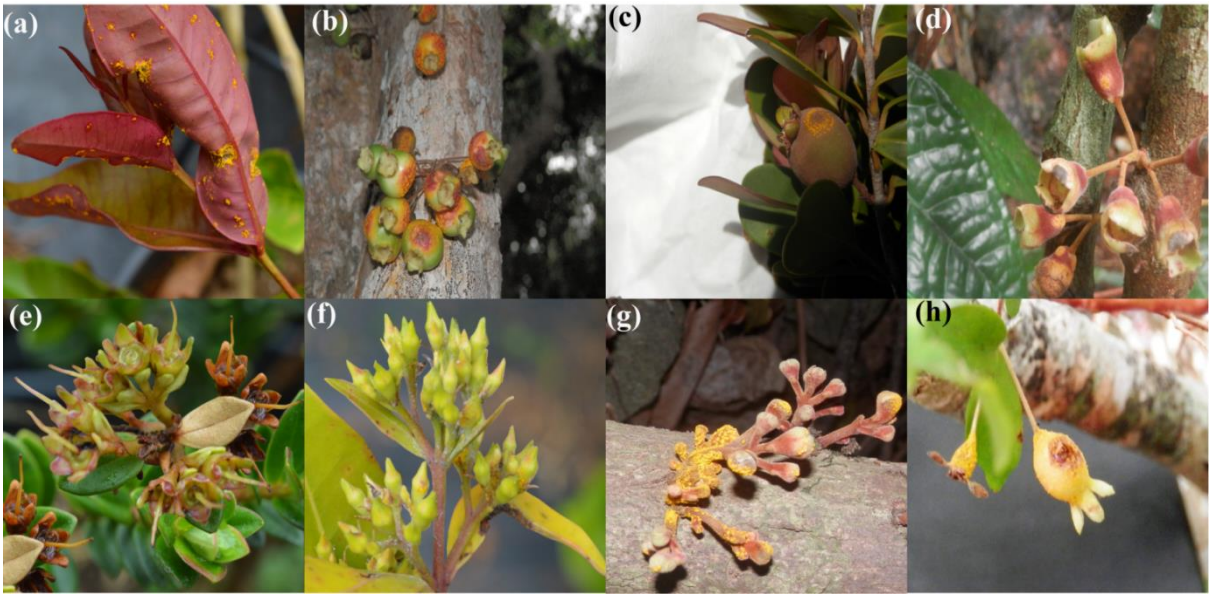


Figure 2.3 : Myrtle rust symptoms on (a) leaves of *Cloezia artensis var. artensis;* on fruits of (b) *Eugenia gacognei*, (c) *Eugenia stricta*, (d) *Eugenia bullata;* on flowers and buds of (e) *Metrosideros operculata var. francii*, (f) *Cloezia artensis var. artensis*, (g) *Eugenia noumeensis*, (h) *Eugenia excorticata.*

Table 2.2 : Host range of *A. psidii* in New Caledonia

	Number of disease occurences					
Host Genus	**Natural field**	**Nurseries**	**Total**	**Number of host species**	**Number of host species with flowers/fruits infection**	**Number of rare and endangered host species**
Eugenia	86	12	98	19	5	5
Syzygium	99	24	123	17	0	0
Cloezia	20	9	29	4	1	1
Metrosideros	11	18	29	9	1	1
Tristaniopsis	8	7	15	5	0	2
Melaleuca	10	4	14	1	0	0
Uromyrtus	7	2	9	3	0	0
Xanthostemon	3	3	6	4	0	0
Stereocaryum	5	0	5	1	0	0
Gossia	3	0	3	2	0	0
Arillastrum	0	2	2	1	0	0
Sannantha	0	2	2	2	0	0
Archirhodomyrtus	0	1	1	1	0	0
Myrtastrum	0	1	1	1	0	0
Myrtaceae .spp	29	7	36	18	0	0
Total	**281**	**92**	**373**	**88**	**7**	**9**

Most host species were observed to be susceptible to *A. psidii* both in nurseries and their natural environment. However, for some specific Myrtaceae genera, susceptible host species were detected in nurseries but were not observed with rust symptoms in the field (*Arillastrum*, *Sannantha*, *Archirhodomyrtus* and *Myrtastrum*) (Table 2.2). Cases of death following myrtle rust infection mainly occurred among juvenile plants in nurseries as well as in adult trees of *Syzygium jambos* (exotic) and *Metrosideros brevistylis* (endemic) in natural habitats (Figure 2.4).



Figure 2.4. Myrtle rust symptoms on leaves of *Syzygium jambos* (a) and *Metrosideros brevistylis* (c); and tree death following several infections on *Syzygium jambos* (b) and *Metrosideros brevistylis* (d).

2.2.4.4. Genotyping of A. psidii populations

Among the 15 markers tested, 8 were successfully amplified: PpSSR010, PpSSR014, PpSSR018, PpSSR022, PpSSR136, PpSSR161, PpSSR195, and PpSSR208. Genotyping was performed for 70 infected samples, representing 24 different host species from 20 locations across the New Caledonian archipelago. All the samples shared the same alleles at the 8 loci, indicating that they belong to a single multilocus genotype (Table 2.3; Supplementary table 2.3).

Table 2.3 : Genotyping results on 65 *A. psidii* isolates from New-Caledonia

Markers from Zhong et al., 2008	**Repeated pattern from Zhong et al., 2008**	**Allele size (bp) observed in this study**	
		allele 1	**allele 2**
PpSSR010	C_{12} $(CA)_9$	141	148
PpSSR014	$(AG)_{13}$	212	216
PpSSR018	$(AG)_{10}$	169	171
PpSSR022	$(AG)_{18}$	156	168
PpSSR136	$(AG)_{14}$	133	144
PpSSR161	$(AG)_{25}$	280	293
PpSSR195	$(AG)_{18}$	116	126
PpSSR208	$(AG)_{12}$	84	84

2.2.4.5. Progress of disease severity over time on 35 cultivated Myrtaceae species in a field nursery

We followed the progress of myrtle rust symptoms on 35 cultivated Myrtaceae species and under natural infection. Despite a trial which does not allow control of inoculation, we observed that, over a relatively long observation period (3 years), many species developed symptoms of myrtle rust. Of the 35 Myrtaceae species screened for myrtle rust symptoms, 7 were determined to be resistant to *A. psidii*, as they did not develop any sign of infection during the whole observation period (Table 2.1).

Development of myrtle rust symptoms was observed on 21 species, which could be considered as susceptible to *A. psidii*. Symptoms monitoring, every two weeks, showed that rust carried out several cycles of infection on the same individuals. Recent sporulation events were observed by the presence of fresh yellow-orange pustules, turning grey in color with age. The emergence of new yellow-orange pustules on the same individuals a few weeks later proves the beginning of a new cycle of infection. A new infection cycle often coincided with the emergence of new leaves. The disease severity score represents the proportion of host tissue covered by symptoms and reflects the extent of damage caused by myrtle rust attack. Minor to moderate infection (maximum disease score of 1, 2 and 3) has been recorded in 12 species. These species showed variations in disease incidence (percentage of diseased plants) from 17 to 100%. Highest myrtle rust disease severity (score of 4; 5 and 6) has been recorded in 13 species (Table 2.1). For these species, the disease incidence varied from 33 to 100%. We also recorded 11 individual plant deaths following several infection cycles.

2.2.5. Discussion

This study is the first statement of the impacts of myrtle rust in New Caledonia. Our research reveals that one genotype of *A. psidii* is occurring in New Caledonia since at least 2013. This genotype was found to infect 70 Myrtaceae species, 67 of which are new host reports for *A. psidii*. Ecological integrity could be compromised in the natural communities where Myrtaceae are the dominant or co-dominant species. It is also likely to affect commercial activities related to Myrtaceae culture.

2.2.5.1. Arrival of Austropuccinia psidii in New Caledonia

Myrtle rust was first detected in New Caledonia in March 2013 on the introduced species *Syzygium jambos* in Farino, South Province. Following its initial detection, several other host species were reported in commercial nurseries, gardens and the natural environment. The evidence of a potentially widespread epidemic strongly supports the hypothesis that the pathogen was established in the natural environment for much longer. The experiences of Australia and Hawaii demonstrate that once the pathogen becomes established in non-confined areas, control and eradication are unlikely to be achieved (Carnegie and Cooper, 2011; Loope and Uchida, 2012). Local nurseries were particularly affected as the establishment of *A. psidii* in the territory due to the optimal conditions for pathogen infection and dispersal (e.g., a high concentration of Myrtaceae in the juvenile stage, regular light and watering), and had a substantial loss of stock. A best practices management plan against myrtle rust was promoted by the Chambre d'Agriculture de Nouvelle-Calédonie to control the disease in confined spaces and maintain nursery activity. However, increased financial costs resulted from the ongoing efforts to control the disease with fungicides and the repetitive disease outbreaks of *A. psidii*. Myrtle rust now represents a serious constraint to the cultivation and propagation of Myrtaceae plants in nurseries.

2.2.5.2. Spread of A. psidii in natural environments

In its native range of Central and South America, *A. psidii* is not usually severe on the native vegetation (Glen et al., 2007). Rust epidemics have mainly been restricted to plantations of host species grown on a large scale, such as guavas, allspice and exotic *Eucalyptus* (Ribeiro and Pommer, 2004). The first example of this rust disease in the natural environment was

reported in Hawaii, where it spread to all major islands of the archipelago (Uchida et al., 2006). In Australia, the disease has caused significant impact in a range of natural ecosystems, including coastal heaths, coastal and riparian wetlands, sand island ecosystems, and littoral, montane, subtropical and tropical rainforests (Carnegie et al., 2016a; Pegg et al., 2014a). It is likely that the same scenario could apply to other island ecosystems containing a high proportion of potential Myrtaceae host species.

Our study showed that many natural plant communities of New Caledonia are suitable for epidemics of *A. psidii* and include many genera and species of susceptible hosts. Some plant communities appear to be more impacted than others in terms of number of host species and observed occurrences of rust disease. We assumed that open field sites dominated by potential Myrtaceae host species are more exposed to wind-dispersed airborne *A. psidii* urediniospores and constitute good conductors for the epidemic spread of disease. These characteristics are found in low-to-medium altitude maquis, savannas and other secondary formations and would explain the higher occurrences of the disease. The incidence of myrtle rust was observed to be lower in closed-type field sites, such as dense humid forest. It is most likely that a high diversity of non-host tree species may contribute to reducing the spread of infectious spores through physical interference in such habitats. We also observed that the different types of dense humid forests occurring in New Caledonia are not impacted to the same degree by myrtle rust. Species from dense humid forests on sedimentary volcanic soils appeared to be more susceptible to myrtle rust than those occurring on ultramafic soils. Great dissimilarities between rainforests on ultramafic soils and those on non-ultramafic soils in terms of floristic composition were previously shown by (Ibanez et al., 2014). We hypothesized that these dissimilarities could partly explain why more susceptible species are found on volcano-sedimentary soils. It is possible that other ecological characteristics that distinguish dense humid forests on these two types of substrate can influence the susceptibility of species, such as specific mycorrhizal associations conferring resistance to biotic stress (Selosse et al., 2004; Smith and Read, 2008). Dense humid forests on calcareous soil are mainly distributed on the Loyalty Islands and Ile des Pins (Morat et al., 2001). In terms of species richness, the Myrtaceae plant family is poorly represented on this type of substrate, which could explain the low representation of myrtle rust occurrences. Relatively few occurrences of myrtle rust were reported in dry forest (sclerophyll forest) because of the fragmentation and scarcity of this type of formation. Nonetheless, the presence of the pathogen in dry forest is of serious concern given the risk of extinction among some species. In addition to its endemism rate (55%) and high proportion of threatened species, dry forest is one of the most endangered

vegetation types worldwide (Gillespie and Jaffré, 2003; Jaffré et al., 2009). From a global point of view, our observations show that the disease incidence in natural environments depends on plant community structure and Myrtaceae species richness and occurrences. It is imperative to initiate monitoring of plant communities impacted by myrtle rust to fully understand the long-term impact of this invasive pathogen. While this study already shows the diversity of plant communities impacted by rust, many plant formations were not investigated in this study due to their inaccessibility. Further efforts should be undertaken to determine whether myrtle rust can occur in other types of formations, for example, dense humid forests at higher altitudes (> 900 m).

2.2.5.3. Host range expansion

We identified 67 new hosts of *A. psidii* under conditions of natural infection, ranking New Caledonia second after Australia in terms of susceptible species. However, the entire host range may not be fully realized in this study for several reasons. On one hand, all factors required for disease development may not be available to some host species in the field (e.g., absence of actively growing shoots or juvenile-stage plants, environmental conditions not conductive to infection or insufficient rust inocula). This could explain why some genera were only found to be susceptible to myrtle rust in nurseries and not under natural field conditions (*Arillastrum*, *Piliocalyx*, *Sannantha*, *Archirhodomyrtus*, *Myrtastrum*). Further investigations should be carried out to verify that these genera are resistant in the natural environment at all times. On the other hand, there are a number of infected samples for which the host genus and species could not be determined. Molecular identification of these samples by DNA barcoding (Hebert et al., 2003) could be beneficial by identifying the host species and would undoubtedly result in an expansion of the *A. psidii* host range list. This approach has already been initiated in New Zealand with the construction of a Myrtaceae DNA barcoding database that would enable the rapid identification of large numbers of host species (Buys et al., 2016).

2.2.5.4. Potential consequences for biodiversity

To date, a majority of the *A. psidii* host species reported in New Caledonia have been impacted at the level of their expanding foliage. Over time, repeated rust infections will likely result in a decrease in photosynthetic capacity, the depletion of foliage regrowth, gradual crown loss, and reductions in reproductive ability and the survivorship of the most susceptible individuals (Booth et al., 2000; Jarosz and Davelos, 1995). After at least three years of *A.*

psidii presence in New Caledonia, records of adult tree deaths in the natural environment are rare and relate mainly to two species: the endemic *Metrosideros brevistylis* and the introduced *Syzygium jambos*. In this situation, it is difficult to grasp over the short term the consequences of the spread of *A. psidii* on the structure and functioning of ecosystems. Such an extent of *A. psidii* attacks has already been observed in native ecosystems on two susceptible key rainforest species in Australia (Carnegie et al., 2016a). One of them, *Rhodomyrtus psidioides*, now suffers from rapid population decline (Carnegie et al., 2016a). The risk that myrtle rust poses to endemic New Caledonian host species becomes more apparent when it attacks the reproductive system. At least four *Eugenia* and one *Metrosideros* species have also been reported as having myrtle rust symptoms on flowers and/or fruits, often leading to other secondary infections and organ necrosis. The natural regeneration of these species could be jeopardized by repeated *A. psidii* infections over time. Some other plant species are at risk of disappearing altogether from their natural ecosystems because of *A. psidii* infections, especially microendemic species that are already rare and endangered.

2.2.5.5. Genetic uniformity of A. psidii isolates

Understanding the structure of a pathogen population is important for forecasting its evolutionary potential and may prove useful for optimizing disease control management (McDonald, 1997; Palumbi, 2001). Many cross inoculation, spore morphology, and molecular studies using microsatellites markers have shown that the *A. psidii* complex is composed of several “biotypes” or races with physiological specialization (Glen et al., 2007; Graça et al., 2011c). (Machado et al., 2015b) previously showed that the same *A. psidii* genotype occurred in Australia, China, Hawaii and New Caledonia. The same study by (Machado et al., 2015b) used 6 microsatellite loci and 4 samples of *A. psidii* from 2 species and 2 locations in New Caledonia. Our work was carried out with a larger sampling effort, including 70 samples, 24 host species and 20 locations scattered across the island. The genotyping of 8 microsatellite loci confirmed that isolates of *A. psidii* from New Caledonia represent one single genotype of the pathogen. Like in Australia (Machado et al., 2015b), *A. psidii* is thought to clonally reproduce throughout New Caledonia(Machado et al., 2015b). Howerver, the potential for genetic changes in genomic regions resulting in the emergence of new pathotypes is known to exist in *A. psidii* populations. (Machado et al., 2015b) showed the emergence of a few non-persistent mutants of the dominant *A. psidii* genotype only 5 years after its first report in Australia. In Brazil, a new and more aggressive race of *A. psidii* emerged to overcome the selected rust-resistance gene *Ppr*-1 in *Eucalyptus grandis* (Graça et al., 2011b). It is important

to maintain the surveillance of *A. psidii* population diversity in New Caledonia by regular sampling and the use of more polymorphic microsatellite markers to identify any pathogenicity shifts resulting from mutations or the introduction of new genotypes.

2.2.5.6. Variation in disease susceptibility over a range of cultivated Myrtaceae species

We followed the progress of myrtle rust symptoms on 35 cultivated Myrtaceae species and under natural infection process. More than half of these expressed symptoms of myrtle rust disease, at least once during the total duration of the monitoring. Among the susceptible species, we included four species originated from cuttings and representing one single repeated plant genotype. In this specific case we could not generalize the observations made from this study and we would rather assess the disease severity on other genetically different individuals before concluding on their true level of susceptibility.

The observed responses of the species susceptible to *A. psidii*, varied on several scales. Primarily at the scale of the species, as the disease severity assessment for each species varies from minor to severe infection, and even death in some cases. Secondly, variations in the disease incidence indicate that not all the replicates of the same species have been affected by rust. This response variability on an individual scale could be explained by several factors including, inoculum quantity, microclimatic conditions within each plant (Tommerup et al., 2003), or differences in the expression of resistance mechanisms depending on intrinsic genetic factors. Such variability in responses toward *A. psidii* have also been observed among and within a range of indigenous Myrtaceae species from Australia (Pegg et al., 2014a, 2014b; Zauza et al., 2010b) and in *Eucalyptus* spp. plantations in Brazil (De Carvalho et al., 1998; Dianese, 1984; Tommerup et al., 2003) and indicates the existence of some level of genetic resistance against myrtle rust (Morin et al., 2012).

At present and among all existing methods of control against A. psidii, the selection of resistant genotypes has been one of the main disease control measures *(Alves et al., 2011).* The genetic basis of resistance toward A. psidii in eucalypts was shown to be complex through the combined action of major- and minor-effect genes *(Junghans et al., 2003; Mamani et al., 2010; Santos et al., 2014).* A greater understanding of the genetic basis of resistance mechanisms toward A. psidii infection across a range of New Caledonian host species showing response variability may provide interesting insights into the drivers of resistance.

2.2.6. Conclusion

This study is the first report of the geographic distribution and host range of *Austropuccinia psidii* in New Caledonia. The pathogen has been reported from a wide range of plant communities, emphasizing the ability of this pathogen to adapt to different conditions. This work expands the host range of *A. psidii* to include 67 new hosts and 5 new genera. Spatial modeling of the distribution of the Myrtaceae plant family in New Caledonia could be undertaken to identify priority conservation areas where a large diversity of host species may be impacted.

An important result of this study was that all the isolates collected across the archipelago represented one single genotype, possibly resulting from one introduction event. Therefore, it is essential to monitor the emergence of potentially more virulent strains by mutation or the incursion of other genotypes. Resistance to rust has also been identified in many host species under field nursery conditions, which demonstrates their potential use in the search for candidate resistance genes and breeding programs.

Acknowledgements

The authors acknowledge the scientific and technical assistance of Nicolas Hugot from "Groupement de défense sanitaire végétal (GDS-V)", Aurélie Chan and Elodie Nakamura from "Service d'Inspection Vétérinaire, Alimentaire et Phytosanitaire (SIVAP)". We wish to particularly thank the owners of the field nursery "Les Hauts de Farino", Philippe Rinck, Carmela Gonzales and Benjamin Rinck for their help and support during this study. We are grateful to those who participated to the rust surveys, and especially Julien Drouin, Marion Anquez, Guillaume Lanuzel, Anthony Pain, Alexandre Bouarat, René Guiglion, Jean-Paul Lataï, Laure Barrabé and Marine Toussirot from the Institut Agronomique néo Calédonien, Sophie Tronc from the Chambre d'Agriculture de Nouvelle-Calédonie, Nathalie Kaquea and Aurélie Virtel from Société Montagnat. We thank Céline Chambrey, Vanessa Hecquet from Institut de Recherche pour le Développement and Thais Vasconcelos from Kew Garden for their help during Myrtaceae species identification. We also would like to thank Kelly Letellier for her help during the sequencing process. Funding for the work was provided by GLENCORE society, Institut Agronomique néo-Calédonien, South Province of New Caledonia and Centre de coopération international en recherche agronomique pour le développement (CIRAD).

2.2.7. References

Alfenas, A. C., Zauza, E. a. V. and Assis, T. F. (2003). First record of *Puccinia psidii* on *Eucalyptus globulu*s and *E. viminalis* in Brazil. Australasian Plant Pathology, 32, 325–326.

Alves, A. A., Rosado, C. C. G., Faria, D. A., Guimarães, L. M. da S., Lau, D., Brommonschenkel, S. H., Grattapaglia, D. and Alfenas, A. C. (2011). Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of *Eucalyptus* to *Puccinia psidii* rust infection. Euphytica, 183, 27–38.

Beenken, L. (2017). *Austropuccinia*: A new genus name for the myrtle rust *Puccinia psidii* placed within the redefined family Sphaerophragmiaceae (Pucciniales). Phytotaxa, 297, 53-61.

Booth, T. H., Old, K. M. and Jovanovic, T. (2000). A preliminary assessment of high risk areas for *Puccinia psidii* (*Eucalyptus* rust) in the Neotropics and Australia. Agriculture, Ecosystems & Environment 82, 295–301.

Bouchet, P., Jaffre, T. and Veillon, J.-M. (1995). Plant extinction in New Caledonia: protection of sclerophyll forests urgently needed. Biodiversity and Conservation 4, 415–428.

Buys, M. H., Flint, H. J., Miller, E. M., Yao, H., Caird, A. R. and Ganley, R. J. (2016). Preparing for the invasion: efficacy of DNA barcoding to discern the host range of myrtle rust (*Puccinia psidii*) among species of Myrtaceae. Forestry, 89, 263–270.

Carnegie, A. J. and Cooper, K. (2011). Emergency response to the incursion of an exotic myrtaceous rust in Australia. Australasian Plant Pathology, 40, 346–359.

Carnegie, A. J., Lidbetter, J. R., Walker, J., Horwood, M. A., Tesoriero, L., Glen, M. and Priest, M. J. (2010). *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. Australasian Plant Pathology, 39, 463–466.

Carnegie, A. J., Kathuria, A., Pegg, G. S., Entwistle, P., Nagel, M. and Giblin, F. R. (2016a). Impact of the invasive rust *Puccinia psidii* (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biological Invasions, 18, 127-144.

Carnegie, A. J., Kathuria, A., Pegg, G. S., Entwistle, P., Nagel, M. and Giblin, F. R. (2016b). Impact of the invasive rust *Puccinia psidii* (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biological Invasions, 18, 127–144.

Coutinho, T. A., Wingfield, M. J., Alfenas, A. C. and Crous, P. W. (1998). *Eucalyptus* Rust: A disease with the potential for serious international implications. Plant Disease, 82, 819–825.

De Carvalho, A. D., Alfenas, A. C., Maffia, L. A. and Do Carmo, M. G. F. (1998). Resistance of *Eucalyptus* species, progenies and provenances to *Puccinia psidii* winter. Pesquisa agropecuaria brasileira, 33, 139–147.

Desprez-Loustau, M.-L., Aguayo, J., Dutech, C., Hayden, K. J., Husson, C., Jakushkin, B., Marçais, B., Piou, D., Robin, C. and Vacher, C. (2016). An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. Annals of Forest Science, 73, 45–67.

Dianese, J. C. (1984). Response of *Eucalyptus* species to field infection by *Puccinia psidii*. Plant Disease, 68, 314.

Dodds, P. and Thrall, P. (2009). Goldacre paper: Recognition events and host–pathogen co-evolution in gene-for-gene resistance to flax rust. Functional Plant Biology, 36, 395.

du Plessis, E., McTaggart, A. R., Granados, G., Wingfield, M., Roux, J., Pegg, G. S., Makinson, J., Purcell, M. and Ali, M. (2017). First report of myrtle rust caused by *Austropuccinia psidii* on *Rhodomyrtus tomentosa* (Myrtaceae) from Singapore. Plant Disease.

Ellison, A. M., Bank, M. S., Clinton, B. D., Colburn, E. A., Elliott, K., Ford, C. R., Foster, D. R., Kloeppel, B. D., Knoepp, J. D., Lovett, G. M., Mohan, J., Orwig, D. A., Rodenhouse, N. L., Sobczak, W. V., Stinson, K. A., Stone, J. K., Swan, C. M., Thompson, J., Von Holle, B. and

Webster, J. R. (2005). Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Frontiers in Ecology and the Environment, 3, 479–486.

Ennos, R. A. (2015). Resilience of forests to pathogens: an evolutionary ecology perspective. Forestry, 88, 41–52.

Gardes, M. and Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. Molecular ecology, 2, 113–118.

Giblin, F. R. (2013). Myrtle rust report: New Caledonia. University of the Sunshine Coast, Maroochydore, Queensland, Australia.

Giblin, F. R. and Carnegie, A. J. (2014). Puccinia psidii (Myrtle Rust) – Australian host list. Version current at 24 Sept. 2014. http://www.anpc.asn.au/resources/Myrtle_Rust.html.

Gillespie, T. W. and Jaffré, T. (2003). Tropical dry forests in New Caledonia. Biodiversity & Conservation, 12, 1687–1697.

Glen, M., Alfenas, A. C., Zauza, E. a. V., Wingfield, M. J. and Mohammed, C. (2007). *Puccinia psidii*: a threat to the Australian environment and economy —a review. Australasian Plant Pathology, 36, 1–16.

Gomez, C., Mangeas, M., Curt, T., Ibanez, T., Munzinger, J., Dumas, P., Jérémy, A., Despinoy, M. and Hély, C. (2015). Wildfire risk for main vegetation units in a biodiversity hotspot: modeling approach in New Caledonia, South Pacific. Ecology and Evolution, 5, 377–390.

Graça, R. N., Alfenas, A. C., Ross-Davis, A. L., Klopfenstein, N., Kim, M.-S., Peever, T. L., Cannon, P. G., Uchida, J. Y., Kadooka, C. Y. and Hauff, R. D. (2011a). Multilocus genotypes indicate differentiation among *Puccinia psidii* populations from South America and Hawaii. In: Proceedings of the 58th Annual Western International Forest Disease Work Conference; 2010 October 4–8; Valemount, BC: 131–134 Flagstaff, AZ: US Forest Service, AZ Zone Forest Health.

Graça, R. N., Aun, C. P., Guimarães, L. M. S., Rodrigues, B. V. A., Zauza, E. A. V. and Alfenas, A. C. (2011b). A new race of *Puccinia psidii* defeats rust resistance in eucalypt. Australasian Plant Pathology, 40, 442–447.

Graça, R. N., Ross-Davis, A. L., Klopfenstein, N. B., Kim, M.-S., Peever, T. L., Cannon, P. G., Aun, C. P., Mizubuti, E. S. G. and Alfenas, A. C. (2013). Rust disease of eucalypts, caused by *Puccinia psidii*, did not originate via host jump from guava in Brazil. Molecular Ecology, 22, 6033–6047.

Granados, G. M., McTaggart, A. R., Barnes, I., Rodas, C. A., Roux, J. and Wingfield, M. J. (2017). The pandemic biotype of *Austropuccinia psidii* discovered in South America. Australasian Plant Pathology, 40, 442-447.

Grandcolas, P., Murienne, J., Robillard, T., Desutter-Grandcolas, L., Jourdan, H., Guilbert, E. and Deharveng, L. (2008). New Caledonia: a very old Darwinian island? Philosophical Transactions of the Royal Society B: Biological Sciences 363, 3309–3317.

Hebert, P. D. N., Cywinska, A., Ball, S. L. and deWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences, 270, 313–321.

Ibanez, T., Munzinger, J., Dagostini, G., Hequet, V., Rigault, F., Jaffré, T. and Birnbaum, P. (2014). Structural and floristic diversity of mixed tropical rain forest in New Caledonia: new data from the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN). Applied Vegetation Science, 17, 386–397.

Jaffré, T., Rigault, F., Dagostini, G., Tinel-Fambart, J., Wulff, A. and Munzinger, J. (2009). Input of the different vegetation units to the richness and endemicity of the New Caledonian flora. Flora, 162, 54–9.

Jarosz, A. M. and Davelos, A. L. (1995). Effects of disease in wild plant populations and the evolution of pathogen aggressiveness. New Phytologist, 129, 371–387.

Junghans, D. T., Alfenas, A. C., Brommonschenkel, S. H., Oda, S., Mello, E. J. and Grattapaglia, D. (2003). Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. Theoretical and Applied Genetics, 108, 175–180.

Kawanishi, T., Uematsu, S., Kakishima, M., Kagiwada, S., Hamamoto, H., Horie, H. and Namba, S. (2009). First report of rust disease on ohia and the causal fungus, *Puccinia psidii*, in Japan. Journal of General Plant Pathology, 75, 428–431.

Kier, G., Kreft, H., Lee, T. M., Jetz, W., Ibisch, P. L., Nowicki, C., Mutke, J. and Barthlott, W. (2009). A global assessment of endemism and species richness across island and mainland regions. Proceedings of the National Academy of Sciences, 106, 9322–9327.

Langrell, S. R. H., Glen, M. and Alfenas, A. C. (2008). Molecular diagnosis of *Puccinia psidii* (guava rust) – a quarantine threat to Australian eucalypt and Myrtaceae biodiversity. Plant Pathology, 57, 687–701.

Loope, L. L. and Uchida, J. Y. (2012). The challenge of retarding erosion of island biodiversity through phytosanitary measures: An update on the case of *Puccinia psidii* in Hawai‘i. Pacific Science 66, 127–139.

Machado, P., Glen, M., Pereira, O. L., Silva, A. A. and Alfenas, A. C. (2015a). Epitypification of *Puccinia psidii*, causal agent of guava rust. Tropical Plant Pathology, 40, 5–12.

Machado, P., Alfenas, A. C., Alfenas, R. F., Mohammed, C. L. and Glen, M. (2015b). Microsatellite analysis indicates that *Puccinia psidii* in Australia is mutating but not recombining. Australasian Plant Pathology, 44, 455–462.

Maclachlan, J. D. (1938). A rust of the Pimento tree in Jamaica, B.W.I. Phytopathology, 28, 157-170.

Mamani, E. M. C., Bueno, N. W., Faria, D. A., Guimarães, L. M. S., Lau, D., Alfenas, A. C. and Grattapaglia, D. (2010). Positioning of the major locus for *Puccinia psidii* rust resistance (Ppr1) on the *Eucalyptus* reference map and its validation across unrelated pedigrees. Tree Genetics & Genomes, 6, 953–962.

McDonald, B. A. (1997). The Population genetics of fungi: Tools and techniques. Phytopathology, 87, 448–453.

McTaggart, A. R., Roux, J., Granados, G. M., Gafur, A., Tarrigan, M., Santhakumar, P. and Wingfield, M. J. (2015). Rust (*Puccinia psidii*) recorded in Indonesia poses a threat to forests and forestry in South-East Asia. Australasian Plant Pathology, 45, 83-89.

Mellano, V. (2006). Rust on myrtle found in San Diego County. Healthy Garden-Healthy Home 3.

Morat, P. (1993). Our Knowledge of the Flora of New Caledonia: Endemism and Diversity in Relation to Vegetation Types and Substrates. Biodiversity Letters 1, 72.

Morat, P., Jaffré, T. and Veillon, J.-M. (2001). The flora of New Caledonia’s calcareous substrates. Adansonia 23, 109–127.

Morat, P., Jaffré, T., Tronchet, F., Munzinger, J., Pillon, Y., Veillon, J.-M., Chalopin, M., Birnbaum, P., Rigault, F., Dagostini, G., Tinel, J. and Lowry, P. P. (2012). Le référentiel taxonomique Florical et les caractéristiques de la flore vasculaire indigène de la Nouvelle-Calédonie. Adansonia 34, 179–221.

Morin, L., Aveyard, R., Lidbetter, J. R. and Wilson, P. G. (2012). Investigating the host-range of the rust fungus *Puccinia psidii* sensu lato across tribes of the family Myrtaceae present in Australia. PLoS One, 7(4), e35434.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B. da and Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403, 853–858.

Orwig, D. A. (2002). Ecosystem to regional impacts of introduced pests and pathogens: historical context, questions and issues. Journal of Biogeography, 29, 1471–1474.

Palumbi, S. R. (2001). Humans as the world's greatest evolutionary force. Science, 293, 1786–1790.

Pegg, G. S., Giblin, F. R., McTaggart, A. R., Guymer, G. P., Taylor, H., Ireland, K. B., Shivas, R. G. and Perry, S. (2014a). *Puccinia psidii* in Queensland, Australia: disease symptoms, distribution and impact. Plant Pathology, 63, 1005–1021.

Pegg, G. S., Brawner, J. T. and Lee, D. J. (2014b). Screening *Corymbia* populations for resistance to *Puccinia psidii*. Plant Pathology, 63, 425–436.

Potts, B. M., Sandhu, K. S., Wardlaw, T., Freeman, J., Li, H., Tilyard, P. and Park, R. F. (2016). Evolutionary history shapes the susceptibility of an island tree flora to an exotic pathogen. Forest Ecology and Management, 368, 183–193.

Rayachhetry, M. B., Elliott, M. L. and Van, T. K. (1997). Natural epiphytotic of the rust *Puccinia psidii* on *Melaleuca quinquenervia* in Florida. Plant Disease, 81, 831–831.

Read, J., Jaffré, T., Godrie, E., Hope, G. S. and Veillon, J.-M. (2000). Structural and floristic characteristics of some monodominant and adjacent mixed rainforests in New Caledonia. Journal of Biogeography, 27, 233–250.

Ribeiro, I. J. A. and Pommer, C. V. (2004). Breeding guava (*Psidium guajava*) for resistance to rust caused by *Puccinia psidii*. Acta Horticulturae, 75–78.

Ross-Davis, A. L., Graca, R. N., Alfenas, A. C., Peever, T. L., Hanna, J. W., Uchida, J. Y., Hauff, R. D., Kadooka, C. Y., Kim, M.-S., Cannon, P. G. and others (2014). Tracking the distribution of *Puccinia psidii* genotypes that cause rust disease on diverse myrtaceous trees and shrubs. In: Chadwick K Comp Proceedings of the 61st Annual Western International Forest Disease Work Conference; October 6–11, Waterton Lakes National Park, Alberta, Canada.

Roux, J., Greyling, I., Coutinho, T. A., Verleur, M. and Wingfield, M. J. (2013). The myrtle rust pathogen, *Puccinia psidii*, discovered in Africa. IMA Fungus, 4, 155–159.

Roux, J., Granados, G. M., Shuey, L., Barnes, I., Wingfield, M. J. and McTaggart, A. R. (2016). A unique genotype of the rust pathogen, *Puccinia psidii*, on Myrtaceae in South Africa. Australasian Plant Pathology, 45, 645–652.

Santos, M. R., Guimares, L. M. da S., Resende, M. D. V. de, Rosse, L. N., Zamprogno, K. C. and Alfenas, A. C. (2014). Resistance of *Eucalyptus pellita* to rust (*Puccinia psidii*). Crop Breeding and Applied Biotechnology, 14, 244–250.

Selosse, M.-A., Baudoin, E. and Vandenkoornhuyse, P. (2004). Symbiotic microorganisms, a key for ecological success and protection of plants. Comptes Rendus Biologies, 327, 639–648.

Smith, S. E. and Read, D. J. (2008). Mycorrhizal symbiosis (third edition). Academic, New York.

Stewart, J. E., Ross-Davis, A. L., Graça, R. N., Alfenas, A. C., Peever, T. L., Hanna, J. W., Uchida, J. Y., Hauff, R. D., Kadooka, C. Y., Kim, M.-S., Cannon, P. G., Namba, S., Simeto, S., Pérez, C. A., Rayamajhi, M. B., Lodge, D. J., Arguedas, M., Medel-Ortiz, R., López-Ramirez, M. A., Tennant, P., Glen, M., Machado, P. S., McTaggart, A. R., Carnegie, A. J. and Klopfenstein, N. B. (2017). Genetic diversity of the myrtle rust pathogen (*Austropuccinia psidii*) in the Americas and Hawaii: Global implications for invasive threat assessments. Forest Pathology, e12378. https://doi.org/10.1111/efp.12378

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution, 30, 2725–2729.

Thornhill, A. H., Ho, S. Y. W., Külheim, C. and Crisp, M. D. (2015). Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Molecular Phylogenetics and Evolution, 93, 29–43.

Tommerup, I. C., Alfenas, A. C., Old, K. M. and others (2003). Guava rust in Brazil–a threat to *Eucalyptus* and other Myrtaceae. New Zealand Journal of Forestry Science, 33, 420–428.

Uchida, J., Zhong, S. and Killgore, E. (2006). First report of a rust disease on Ohia caused by *Puccinia psidii* in Hawaii. Plant Disease, 90, 524–524.

Vasconcelos, T. N. C., Proença, C. E. B., Ahmad, B., Aguilar, D. S., Aguilar, R., Amorim, B. S., Campbell, K., Costa, I. R., De-Carvalho, P. S., Faria, J. E. Q., Giaretta, A., Kooij, P. W., Lima, D. F., Mazine, F. F., Peguero, B., Prenner, G., Santos, M. F., Soewarto, J., Wingler, A. and Lucas, E. J. (2017). Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae. Molecular Phylogenetics and Evolution, 109, 113–137.

Zauza, E. A. V., Alfenas, A. C., Old, K., Couto, M. M. F., Graça, R. N. and Maffia, L. A. (2010). Myrtaceae species resistance to rust caused by *Puccinia psidii*. Australasian Plant Pathology, 39, 406–411.

Zhong, S., Yang, B. and Alfenas, A. C. (2008). Permanent genetic resources: Development of microsatellite markers for the guava rust fungus, *Puccinia psidii*. Molecular Ecology Resources, 8, 348–350.

Zhuang, J-Y. & Wei, S-X. Additional materials for the rust flora of Hainan Province, China.

Supplementary table 2.1. List of DDBJ accession number of *Austropuccinia psidii* isolates from New Caledonia for the ITS, β-tubulin and EF1-α loci.

			Accession number		
***A. psidii* isolates**	**Host species**	**Locality, Province**	**ITS**	**β-tubulin**	**EF1-α**
PpNC1	*Cloezia artensis var.artensis*	Païta, South Province	LC228282	LC228290	LC228298
PpNC2	*Eugenia gacognei*	Moindou, South Province	LC228283	LC228291	-
PpNC3	*Eugenia noumeensis*	Ouégoa, North Province	LC228284	LC228292	-
PpNC4	*Syzygium jambos*	Farino, South Province	LC228285	LC228293	LC228299
PpNC5	*Syzygium jambos*	Maré, Isles Province	LC228286	LC228294	LC228300
PpNC6	*Syzygium jambos*	Maré, Isles Province	LC228287	LC228295	-
PpNC7	*Syzygium jambos*	Mont Dore, South Province	LC228288	LC228296	-
PpNC8	*Syzygium jambos*	Farino, South Province	LC228289	LC228297	LC228301

Supplementary table 2.2 : List of *Austropuccinia psidii* host species in New-Caledonia

	Number of reported myrtle rust occurrences			
Host species	**Natural field**	**Nurseries**	**Total**	**New host record**
Archirhodomyrtus sp. *Koné*	0	1	1	yes
Arillastrum gummiferum	0	2	2	yes
Cloezia artensis var. *artensis*	17	8	25	yes
Cloezia artensis var. *riparia*	3	0	3	yes
Cloezia floribunda	2	0	2	yes
Cloezia sp. Koné	0	1	1	yes
Eugenia balansae	0	1	1	yes
Eugenia brongniartiana	1	0	1	yes
Eugenia bullata	12	1	13	yes
Eugenia daenikeri	1	0	1	yes
Eugenia excorticata	6	0	6	yes
Eugenia gacognei	39	0	39	yes
Eugenia hurlimannii	3	0	3	yes
Eugenia mouensis	1	0	1	yes
Eugenia munzingeri	1	0	1	yes
Eugenia noumeensis	5	4	9	yes
Eugenia ovigera	1	0	1	yes
Eugenia paludosa	1	0	1	yes
Eugenia sp. *Baaba*	1	0	1	yes
Eugenia sp. *Dumbéa*	1	1	2	yes
Eugenia sp. *Gohapin*	1	0	1	yes
Eugenia sp. *Maa*	0	3	3	yes
Eugenia sp. *Ouate*	1	0	1	yes
Eugenia sp. *Tiawe*	1	0	1	yes
Eugenia stricta	18	2	20	yes
Gossia alaternoides	1	0	1	yes
Gossia sp. *Pic du Grand Kaori*	2	0	2	yes
Melaleuca quinquenervia	10	4	14	no
Metrosideros brevistylis	6	0	6	yes
Metrosideros elegans	1	1	2	yes
Metrosideros laurifolia	4	9	13	yes
Metrosideros nitida	0	1	1	yes
Metrosideros operculata	0	1	1	yes
Metrosideros operculata var. *francii*	0	2	2	yes
Metrosideros operculata var. *operculata*	0	1	1	yes
Metrosideros punctata	0	2	2	yes
Metrosideros sp. *Koné*	0	1	1	yes
Myrtastrum rufopunctatum	0	1	1	yes
Syzygium baudouinii	0	1	1	yes
Syzygium eugenioides	0	1	1	yes
Sannantha leratii	0	1	1	yes

Supplementary table 2.2 continued from previous page

	Number of reported myrtle rust occurrences			
Species	**Natural field**	**Nurseries**	**Total**	**New host record**
Sannantha procera	0	1	1	yes
Stereocaryum neocaledonicum	5	0	5	yes
Syzygium acre	0	1	1	yes
Syzygium brongniartii	1	0	1	yes
Syzygium cumini	2	0	2	no
Syzygium frutescens	0	1	1	yes
Syzygium jambos	86	4	90	no
Syzygium longifolium	3	7	10	yes
Syzygium macranthum	0	1	1	yes
Syzygium mouanum	1	0	1	yes
Syzygium ngoyense	1	3	4	yes
Syzygium pancheri	0	1	1	yes
Syzygium sp. *Dumbéa*	1	0	1	yes
Syzygium sp. *Karikouié*	0	1	1	yes
Syzygium sp. *Moindou*	4	0	4	yes
Syzygium sp. *Mont Mou*	0	1	1	yes
Syzygium wagapense	0	2	2	yes
Tristaniopsis calobuxus	0	1	1	yes
Tristaniopsis glauca	5	4	9	yes
Tristaniopsis guillainii	0	1	1	yes
Tristaniopsis polyandra	3	0	3	yes
Tristaniopsis reticulata	0	1	1	yes
Uromyrtus emarginata	3	0	3	yes
Uromyrtus sp. Katepahie	1	0	1	yes
Uromyrtus sp. Koné	0	2	2	yes
Xanthostemon aurantiacus	1	2	3	yes
Xanthostemon laurinus	0	1	1	yes
Xanthostemon sp. *Parc Forestier*	1	0	1	yes
Xanthostemon vieillardii	1	0	1	yes
Total	**281**	**92**	**373**	**67**

Supplementary table 2.3 : Details of *A. psidii* isolates used for microsatellites genotyping

	***A. psidii* isolate**	**Host species**	**Collection date**	**Type of sampling**	**Locality**	**Longitude**	**Latitude**
1	F4_3	*Cloezia artensis* var. *artensis*	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
2	F4_6	*Cloezia artensis* var. *artensis*	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
3	PpNC1	*Cloezia artensis* var. *artensis*	11/09/2014	nursery	Païta, South Province	166.382602	-22.093598
4	JS38	*Eugenia bullata*	02/09/2015	natural field	Yaté, South Province	166.88287	-22.34114
5	P13	*Eugenia bullata*	17/09/2014	natural field	Mont Dore, South Province	166.88287	-22.34114
6	P14	*Eugenia bullata*	17/09/2014	natural field	Mont Dore, South Province	166.88287	-22.34114
7	P15	*Eugenia bullata*	17/09/2014	natural field	Mont Dore, South Province	166.88287	-22.34114
8	P17	*Eugenia bullata*	17/09/2014	natural field	Mont Dore, South Province	166.88287	-22.34114
9	SL2	*Eugenia bullata*	16/09/2014	nursery	Mont Dore, South Province	166.540299	-22.231296
10	Md1	*Eugenia excorticata*	13/03/2014	natural field	Moindou, South Province	165.542113	-21.696027
11	Md13	*Eugenia excorticata*	13/03/2014	natural field	Moindou, South Province	165.541298	-21.696610
12	Md2	*Eugenia excorticata*	13/03/2014	natural field	Moindou, South Province	165.542043	-21.696061
13	Md25	*Eugenia excorticata*	13/03/2014	natural field	Moindou, South Province	165.545151	-21.698724
14	JS7	*Eugenia gacognei*	31/08/2015	natural field	Dumbéa, South Province	166.499944	-22.1360278
15	JS9	*Eugenia gacognei*	31/08/2015	natural field	Dumbéa, South Province	166.499944	-22.1360278
16	Ma10	*Eugenia gacognei*	10/01/2014	natural field	Maré, Isles Province	167.950047	-21.598583
17	Ma11	*Eugenia gacognei*	10/01/2014	natural field	Maré, Isles Province	167.950047	-21.598583
18	Ma4	*Eugenia gacognei*	10/01/2014	natural field	Maré, Isles Province	167.97121	-21.433454
19	Ma8	*Eugenia gacognei*	10/01/2014	natural field	Maré, Isles Province	167.872084	-21.480707
20	Md12	*Eugenia gacognei*	13/03/2014	natural field	Moindou, South Province	165.541314	-21.696620
21	Md20	*Eugenia gacognei*	13/03/2014	natural field	Moindou, South Province	165.544948	-21.698146
22	Md28	*Eugenia gacognei*	13/03/2014	natural field	Moindou, South Province	165.545624	-21.698888
23	Md5	*Eugenia gacognei*	13/03/2014	natural field	Moindou, South Province	165.541890	-21.696060
24	Md7	*Eugenia gacognei*	13/03/2014	natural field	Moindou, South Province	165.541598	-21.696393
25	Md9	*Eugenia gacognei*	13/03/2014	natural field	Moindou, South Province	165.541510	-21.696525
26	PpNC2	*Eugenia gacognei*	13/03/2014	natural field	Moindou, South Province	165.545373	-21.698518
27	Cp25	*Eugenia hurlimannii*	29/04/2014	natural field	Yaté, South Province	166.8456720	-22.1786114
28	Cp26	*Eugenia hurlimannii*	29/04/2014	natural field	Yaté, South Province	166.8392775	-22.1825791
29	Cp27	*Eugenia hurlimannii*	29/04/2014	natural field	Yaté, South Province	166.8392775	-22.1825791
30	JS100	*Eugenia mouensis*	08/09/2015	natural field	Ponérihouen, North Province	165.26888465	-21.00918716
31	F6_2	*Eugenia noumeensis*	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
32	F6_5	*Eugenia noumeensis*	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
33	JS118	*Eugenia noumeensis*	09/09/2015	natural field	Ouégoa, North Province	164.42191	-20.31466
34	PpNC3	*Eugenia noumeensis*	09/09/2015	natural field	Ouégoa, North Province	164.42191	-20.31466
35	JS122	*Eugenia noumeensis*	09/09/2015	natural field	Ouégoa, North Province	164.42191	-20.31466
36	JS120	*Eugenia paludosa*	09/09/2015	natural field	Ouégoa, North Province	164.42191	-20.31466
37	F7_2	*Eugenia* sp. Maa	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
38	JS13	*Eugenia* sp. Dumbéa	31/08/2015	natural field	Dumbéa, South Province	166.50075	-22.1388333
39	Cp1	*Eugenia stricta*	28/04/2014	natural field	Yaté, South Province	166.8175480	-22.2009598
40	JS36	*Eugenia stricta*	02/09/2015	natural field	Yaté, South Province	166.896284	-22.285214
41	Vil4	*Eugenia stricta*	11/09/2014	nursery	Païta, South Province	166.382602	-22.093598
42	JS28	*Gossia* sp.Pic du Grand Kaori	01/09/2015	natural field	Yaté, South Province	166.896284	-22.285214

Supplementary table 2.3 continued from previous page

	A. psidii isolate	Taxon host	Collection date	Type of sampling	Locality	Longitude	Latitude
43	F8_1	*Melaleuca quinquenervia*	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
44	F8_2	*Melaleuca quinquenervia*	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
45	JS123	*Melaleuca quinquenervia*	10/09/2015	natural field	Ouégoa, North Province	164.42192318	-20.31465152
46	Vil2	*Melaleuca quinquenervia*	11/09/2014	nursery	Païta, South Province	166.382602	-22.093598
47	JS152	*Metrosideros brevistylis*	10/09/2015	natural field	Ouégoa, North Province	164.52945436	-20.3996182
48	JS154	*Metrosideros brevistylis*	10/09/2015	natural field	Ouégoa, North Province	164.53046044	-20.39924151
49	JS67	*Metrosideros brevistylis*	07/09/2015	natural field	Poya, North Province	165.267820	-21.176370
50	JS73	*Metrosideros brevistylis*	07/09/2015	natural field	Poya, North Province	165.280944	-21.1824168
51	SL3	*Metrosideros laurifolia*	16/09/2014	nursery	Mont Dore, South Province	166.540299	-22.231296
52	JS32	*Myrtaceae* sp. 1 Forêt Nord	02/09/2015	natural field	Yaté, South Province	166.896284	-22.285214
53	JS40	*Myrtaceae* sp. Prony	02/09/2015	natural field	Yaté, South Province	166.88287	-22.34114
54	JS50	*Myrtaceae sp.* 2 Forêt Nord	03/09/2015	natural field	Yaté, South Province	166.91437	-22.3232
55	JS51	*Myrtaceae sp.* 2 Forêt Nord	03/09/2015	natural field	Yaté, South Province	166.91437	-22.3232
56	F11	*Myrtastrum rufopunctatum*	23/09/2014	nursery	Farino, South Province	165.772024	-21.663800
57	PpNC4	*Syzygium jambos*	25/03/2015	nursery	Farino, South Province	165.772024	-21.663800
58	PpNC8	*Syzygium jambos*	08/04/2015	natural field	Farino, South Province	165.773834	-21.658162
59	PpNC6	*Syzygium jambos*	30/01/2014	natural field	Maré, Isles Province	167.8509	-21.447738
60	PpNC5	*Syzygium jambos*	30/01/2014	natural field	Maré, Isles Province	167.8509	-21.447738
61	Ma3	*Syzygium jambos*	10/01/2014	natural field	Maré, Isles Province	168.072087	-21.581587
62	PpNC7	*Syzygium jambos*	16/09/2014	nursery	Mont Dore, South Province	166.540299	-22.231296
63	JS54	*Syzygium mouanum*	05/09/2015	natural field	Yaté, South Province	166.82204	-22.20901
64	Md10	*Syzygium* sp. Moindou	13/03/2014	natural field	Moindou, South Province	165.541510	-21.696525
65	Md15	*Syzygium* sp. Moindou	13/03/2014	natural field	Moindou, South Province	165.541203	-21.696717
66	Md26	*Syzygium* sp. Moindou	13/03/2014	natural field	Moindou, South Province	165.545204	-21.698777
67	Md6	*Syzygium* sp. Moindou	13/03/2014	natural field	Moindou, South Province	165.541598	-21.696393
68	Tont13	*Stereocaryum neocaledonicum*	15/04/2014	natural field	Tontouta, South Province	166.2565474	-21.9971353
69	Tont6	*Stereocaryum neocaledonicum*	15/04/2014	natural field	Tontouta, South Province	166.2551455	-21.9976416
70	JS18	*Uromyrtus emarginata*	01/09/2015	natural field	Yaté, South Province	166.890166666666	-22.2885278

Supplementary table 2.4 : Maximum disease severity scores for each replicate of each species monitored in the field nursery and under natural infection. (-) was set for missing replicates.

	Plant Replicates					
Species	**Plant 1**	**Plant 2**	**Plant 3**	**Plant 4**	**Plant 5**	**Plant 6**
Melaleuca vinimalis var. *captain cook*	0	0	0	0	0	0
Pleurocalyptus pancheri	0	0	0	0	-	-
Psidium guajava	0	0	0	0	0	0
Syzygium aggregatum	0	0	0	0	0	0
Syzygium macranthum	0	0	0	0	0	0
Syzygium malaccense	0	0	0	0	0	0
Syzygium sp. La Confiance	0	0	0	0	0	0
Syzygium australe	0	0	0	0	0	0
Sannantha pinifolia	0	0	0	0	-	-
Xanthostemon myrtifolius	0	0	0	0	0	0
Sannantha leratii	0	0	0	0	0	1
Eugenia balansae	0	0	1	0	1	0
Syzygium sp. Karikouié	0	0	0	1	0	1
Syzygium ngoyense	1	0	0	1	0	0
Sannantha procera	0	1	0	0	1	1
Syzygium fructescens	0	1	1	0	1	1
Syzygium acre	1	0	2	1	1	-
Syzygium longifolium	3	0	0	0	0	3
Tristaniopsis guillainii	0	2	0	2	0	3
Syzygium baudouini	2	1	1	0	3	0
Syzygium wagapense	1	1	2	2	3	2
Syzygium eugenioides	3	2	2	3	3	1
Metrosideros operculata var. *francii*	3	4	2	2	3	3
Metrosideros operculata var. *operculata*	4	2	3	2	3	-
Metrosideros punctata	1	4	3	3	3	0
Metrosideros laurifolia var. *laurifolia*	2	3	4	-	-	-
Arillastrum gummiferum	0	0	0	0	2	5
Xanthostemon laurinus	5	0	0	1	3	0
Xanthostemon aurantiacus	0	0	1	0	5	3
Tristaniopsis glauca	5	5	1	5	5	4
Eugenia noumeensis	5	5	5	5	5	5
Cloezia artensis var. *artensis*	5	5	5	5	5	5
Myrtastrum rufopunctatum	4	6	6	6	6	-
Melaleuca quinquenervia	3	6	1	1	2	2
Eugenia sp. Maa	5	5	5	6	6	5

Chapitre III : Mise en place d'une méthodologie pour la selection de marqueurs de résistance à la rouille chez des espèces de Myrtaceae endémiques à la Nouvelle-Calédonie incluant une appoche RNA-Seq.

Ce chapitre fait l'objet d'un datapaper à soumettre dans la revue Genomics Data.

3.1. Introduction générale et objectifs du Chapitre 3

Les observations menées au cours de l'étude du Chapitre 2 de cette thèse, ont montré une variabilité à l'échelle individuelle dans la réponse à la rouille chez plusieurs espèces d'hôtes. Cela suggère l'existence, et la mise en œuvre, de mécanismes de défense chez les Myrtaceae néo-calédoniennes, lors d'une infection par *A. psidii*. Afin d'infirmer ou confirmer cette hypothèse, nous nous attacherons dans ce troisième Chapitre à définir si les variations observées dans les réponses phénotypiques à la rouille des Myrtaceae néo-calédoniennes admettent une base génétique. Pour se faire, nous soumettrons des individus aux phénotypes contrastés (résistant ou sensible à la rouille) à deux types d'études:

- d'une part, une analyse d'expression différentielle des gènes, afin d'identifier des gènes admettant des expressions différentiées aux regards des deux types de phénotypes retenus ;
- et d'autre part, une recherche de variants génomiques, identifiables par l'observation de mutations nucléotidiques (ou SNPs)s, caractéristiques de chacun des deux types de réponses à la rouille.

Pour mener à bien ces études, sur des espèces de Myrtaceae non-modèles, nous utiliserons une approche RNA-Seq (Annexe 3.1). Ce Chapitre fait l'objet d'un article de format "datapaper" intitulé "A transcriptomic-based approach to identify candidate markers for resistance to myrtle rust (*Austropuccinia psidii*) in three endemic Myrtaceae species from New Caledonia".

3.2. Matériel et Méthodes

3.2.1. Choix des espèces à étudier, localités et protocole d'échantillonnage

Pour comparer les variations de transcrits en réponse à *A. psidii*, nous avons choisi trois espèces de Myrtaceae hôtes d'intérêt pour la revégétalisation ou le secteur ornemental : *Arillastrum gummiferum*, *Syzygium longifolium* et *Tristaniopsis glauca*. Ces trois espèces ont été cultivées à un stade juvénile dans une pépinière naturellement infectée par *A. psidii* sur la commune de Farino. Nous avons empiriquement classé les individus selon les phénotypes résistants ou sensibles tels que décrit dans le Chapitre 2. Les trois espèces sont illustrées dans la Figure 3.2.

Nous avons échantillonné les feuilles de vingt-quatre individus répartis parmi les trois espèces préalablement citées : 6 individus (3 résistants ; 3 sensibles) pour *A. gummiferum*, 7 individus (5 résistants; 2 sensibles) pour *S. longifolium* et 5 individus (1 résistant, 4 sensibles) pour *T. glauca*. Les feuilles récoltées ont été immédiatement placées dans l'azote liquide puis conservés à -80°C pour la suite des expérimentations.



Figure 3.2 : Illustration des espèces étudiées pour le RNA-Seq.
Les images de gauche sont issues du site http://endemia.nc/ et représentent les espèces adultes dans le milieu naturel. Les images de droites ont été prises au sein de la pépinière de Farino sur des plants juvéniles. Légende : a-d : *Syzygium longifolium* ; e-h : *Tristaniopsis glauca* ; i-m : *Arillastrum gummiferum*.

En plus des individus récoltés en pépinière, nous avons également récolté les feuilles de 6 individus en milieu naturel pour *T. glauca*. La localité où ont été récoltés ces individus se situe dans la réserve naturelle Bois du Sud, au Sud de la Nouvelle-Calédonie. Une visite préliminaire a été effectuée avant l'échantillonnage afin de vérifier les phénotypes exprimés des individus présents dans cette population. Le jour de l'échantillonnage nous avons également classé empiriquement ces individus comme résistants s'ils ne présentaient pas de symptômes à la rouille (2) et sensibles s'ils étaient infectés (4). Ces échantillons, prélevés *in*

situ, seront traités à part des *T. glauca* issus de la pépinière de Farino dans les analyses subséquentes.

3.2.2. Extraction de l'ARN

La première étape dans l'expérimentation du RNA-Seq est d'isoler l'ARN des échantillons récoltés. Les Myrtaceae sont connues pour contenir de fort taux de métabolites secondaires pouvant potentiellement interférer ou limiter l'étape d'extraction de l'ARN. Le choix de la méthode d'extraction pour obtenir l'ARN de haute qualité requis pour le RNA-Seq est donc primordial. Nous avons choisi d'utilisé un protocole au CTAB (cetyltrimethylammonium bromide) décrit par Salzman et al., 1999 et adapté à ce type d'échantillon. L'extraction de l'ARN total de chaque individu est suivie d'une étape de purification pour enlever toute trace de contamination ADN. La qualité des ARN extraits est déterminée à l'aide d'un bioanalyseur 2100 (Agilent Technologies) (Figure 3.3).

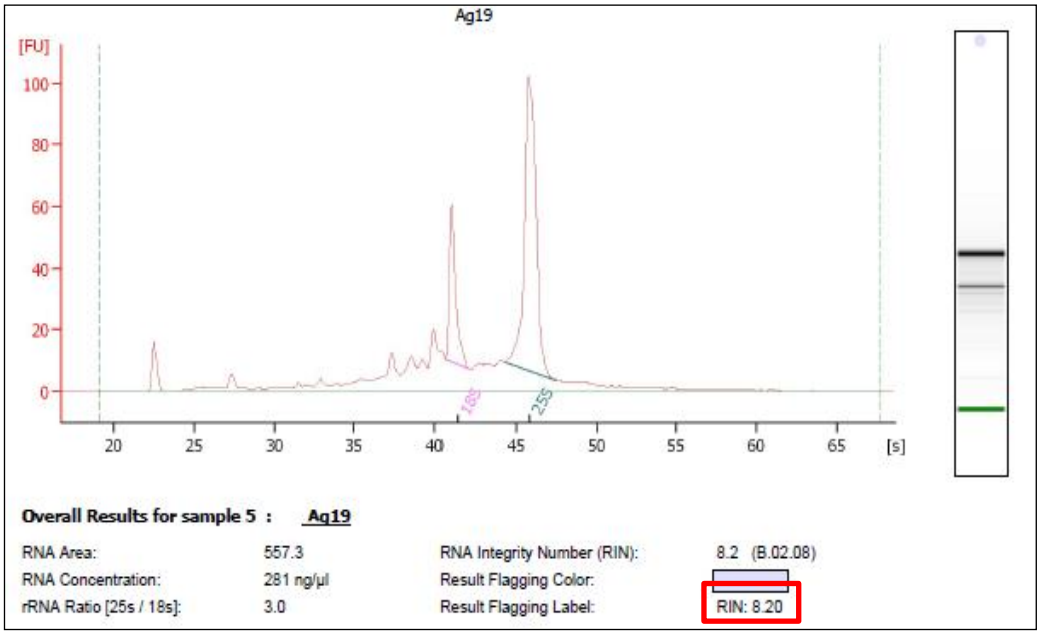


Figure 3.3 : Résultats issus du bioanalyzer pour l'échantillon Ag19 (*Arillastrum gummiferum*). L'ARN est considéré de haute qualité lorsque le ratio des bandes 28S / 18S est d'au moins 2 et le RIN (numéro d'intégrité) doit être au moins de 6.

3.2.3. Construction des banques d'ADNc et séquençage

En général la préparation de l'ARN pour le séquençage en NGS comprend les étapes suivantes : (i) la fragmentation des séquences cibles, (ii) la conversion des cibles en ADN complémentaire (banque d'ADNc), (iii) la fixation d'adaptateurs aux extrémités des fragments cibles et (iv) la quantification des banques d'ADNc finales pour le séquençage (Head et al., 2014).

Sont effectuées successivement à l'aide de du kit "TruSeq ® RNA Sample Preparation" :

- Une étape de préparation de l'ARNm qui inclu leur isolement et purification sur des billes magnétiques oligo-(dT)[21] à partir des extraits d'ARN totaux obtenus, et leur découpage chimique en petits fragments comme l'exige la technologie de séquençage à haut débit utilisée ;
- Une étape de transcription inverse (RT) des fragments d'ARNm (simple brin, fragile) en ADNc (double brin, physiquement plus stables) et amplification par PCR des copies d'ADNc obtenues, opération globale de RT-PCR, dans le but d'obtenir une banque d'ADNc représentative du pool des ARNm d'origine pour chaque échantillon ;
- Une étape de préparation au séquençage qui inclu une ligation aux extrémités de chaque brin d'ADNc d'adaptateurs[22] Illumina TruSeq™, pour leur fixation sur la *flowcell* [23] et pour leur identification individuelle (Barcode). La structure d'une banque d'ADNc est représentée dans l'Annexe 3.2.

Chaque banque d'ADNc est ensuite quantifiée par qPCR à l'aide du kit "KAPA library quantification" (Kapa Biosystems Ltd, SA) puis ajustée à 20 nM. Toutes les banques sont rassemblées dans un même tube (pool de banque d'ADNc) et envoyées pour séquençage multiplexé[24] en paired-end à la plateforme Genotoul de Toulouse (Illumina HiSeq^{TM} 2500, Annexe 3.3). A l'issu de ce séquençage, chaque fragment d'ADNc aboutit à la formation de deux séquences courtes de 150 pb appelées *"reads"*.

[21] Billes magnétiques oligo-(dT) : matrice d'affinité permettant l'isolement de l'ARNm par hybdridation avec la queue poly (A).
[22] Adaptateurs : petites séquences d'ADN double brin.
[23] Flowcell : surface de verre utilisée par la technologie Illumina, et recouverte d'adaptateurs complémentaires à ceux liés aux extrémités de l'ADNc à séquencer.
[24] Multiplexage : séquençage simultané de plusieurs échantillons.

3.2.4. Analyses bioinformatiques des données RNA-Seq

Les étapes majeures du traitement des données issues du RNA-Seq, telles que décrites par (Conesa et al., 2016), sont les suivantes : (i) contrôler la qualité des séquences brutes issues du RNA-Seq, (ii) aligner les *reads* contre une référence, (iii) quantifier les transcrits et (iv) analyser l'expression différentielle des gènes. Nous avons entrepris l'ensemble de ces 4 étapes en ligne de commande[25]. Les protocoles d'analyses (scripts) utilisés ont été adaptés à partir de ceux mis en place par le projet européen ARCAD (Agropolis Resource Center for Crop Conservation, Adaptation and Diversity) (Sarah et al., 2016). Les fichiers étant très volumineux et les programmes d'analyses nécessitant des dizaines de GB de, nous avons traité l'ensemble des données à distance via le cluster[26] de calcul du CIRAD de Montpellier.

3.2.4.1. Contrôle qualité des *reads* bruts

Les séquences nucléiques de chaque reads sont répertoriées dans un fichier au format texte **FASTQ**. Ce fichier contient également des informations sur la qualité du séquençage (score de qualité[27]) comme décrit dans l'Annexe 3.4. Contrôler la qualité des reads bruts implique d'effecuer des analyses de la qualité des séquences, du taux en GC[28], des k-mers[29] surreprésentés, des reads dupliquées, afin de détecter des erreurs de séquençage, des artefacts de PCR ou encore des contaminations.

3.2.4.2. L'alignement des reads

L'ordre de positionnement des *reads* les uns par rapport aux autres est déterminé grâce à l'alignement de la séquence obtenue à une référence. Dans l'idéal cette référence est l'ensemble des séquences d'ADN type de l'espèce étudiée (génome) disponible dans les bases de données internationales (format fna). Les programmes d'alignement transforment les fichiers de reads bruts (FASTQ) en fichier binaire de format BAM (Binary Alignement Map) ou texte de format SAM (Sequence Alignment Map). Les fichiers .SAM et .BAM contiennent l'information de position de chaque read sur le génome. Ils sont volumineux et ne peuvent pas

[25] Ligne de commande : permet d'exécuter une tâche dans le cadre d'un système d'exploitation informatique (dans notre cas UNIX).
[26] Cluster de calcul : équipement informatique (serveurs de calcul, de gestion, de stockage) mis à la disposition par l'UMR AGAP et accessible à distance.
[27] Un score de qualité (Phred) est une prédiction de la probabilité d'une erreur dans le séquençage. Dans notre étude le score qualité est au format Phred+33. https://fr.wikipedia.org/wiki/FASTQ
[28] Taux en GC exprime le pourcentage de liaisons des bases nucléotidiques Guanine-Cytosine dans la molécule d'ADN.
[29] K-mers : répétition de courts motifs d'ADN en tandem.

être visualisés. Il est nécessaire de les manipuler à l'aide d'outils spécifiques tel que samtools (Li et al., 2009).

Dans cette expérimentation nous ne disposions pas de base de génome de référence. Nous avons choisi une double approche afin de pallier à ce problème.

Par une approche de génomique comparative, nous utiliserons le génome disponible de référence de haute qualité d'une espèce sœur, l'*Eucalyptus grandis* (Myburg et al., 2014). Le génome de l'*E. grandis* servira de guide pour l'alignement des reads, en supposant la conservation de l'ordre des gènes (synténie) entre les espèces de Myrtaceae (Ekblom and Galindo, 2011).

Par une approche *de novo*, il s'agit là de reconstruire les transcriptomes de références pour chacune des espèces étudiées. Pour ce faire, et par espèce, l'assemblage de *novo* pour chaque transcrit d'ARNm, à partir des millions de reads disponibles au niveau des données d'un échantillon choisi pour la qualité de ses séquences, est conduit. Les outils bioinformatiques actuellement utilisés pour ce type d'opération recourent au chevauchement de séquences afin de construire des contigs les plus longs possibles.

3.2.4.3. L'annotation

L'étape d'annotation permet de faire le lien entre les données RNA-Seq et les informations biologiques des espèces étudiées (Conesa et al., 2016). L'annotation d'une séquence génomique peut être abordée à deux niveaux d'analyse. Le premier niveau, aussi appelé annotation structurale, consiste à identifier les éléments du génome (gènes, introns, exons, promoteurs, etc.) tels que représentés sur la Figure 3.4. Le deuxième niveau, ou annotation fonctionnelle, cherche à assigner une fonction biologique à chacun des gènes hypothétiques (Médigue et al., 2002). Les informations concernant l'annotation d'un génome de référence sont stockées dans un fichier au format gff3.

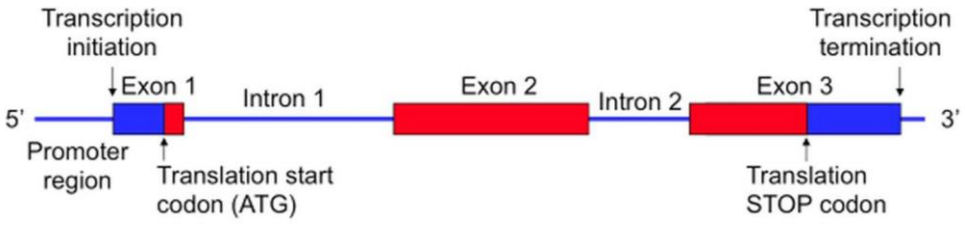


Figure 3.4 : Rappel sur la structure d'un gène

Dans notre étude, nous disposons déjà de l'annotation du génome de l'*E. grandis* (V1.1 annotation, www.phytozome.net) pour les alignements effectués à l'aide de ce même génome de référence. Pour l'alignement *de novo*, l'annotation des transcrits peut se faire par génomique comparative avec les bases de données protéiques internationales telles que Swiss-Prot ou RefSeq.

3.2.4.4. Quantification des transcrits et normalisation

L'un des objectifs les plus récurrents dans les études de RNA-seq est l'identification des gènes différentiellement exprimées (DE) entre des groupes d'individus d'une même espèce exprimant des phénotypes contrastés. L'étape préalable à la comparaison des transcrits est la normalisation des données et la quantification du nombre de reads par gène (Conesa et al., 2016). En effet, il est largement reconnu que le RNA-Seq induit des biais techniques dans le nombre de reads produits pour chaque individu (effet de la taille des banques, de la longueur des transcrits). Par exemple, pour des niveaux d'expression similaires, les transcrits longs auront plus de fragments que les transcrits courts (Trapnell et al., 2012).

Les méthodes les plus utilisées pour corriger ces biais sont la normalisation par la mesure des RPKM (Reads Per Kilobase per Million of reads) (Mortazavi *et al.*, 2008) ou des FPKM (Fragment Per Kilobase per Million of reads) (Conesa et al., 2016) (Figure 3.5). Le fichier produit en sortie de cette étape d'estimation d'abondance des reads par région génomique est un tableau de comptage.

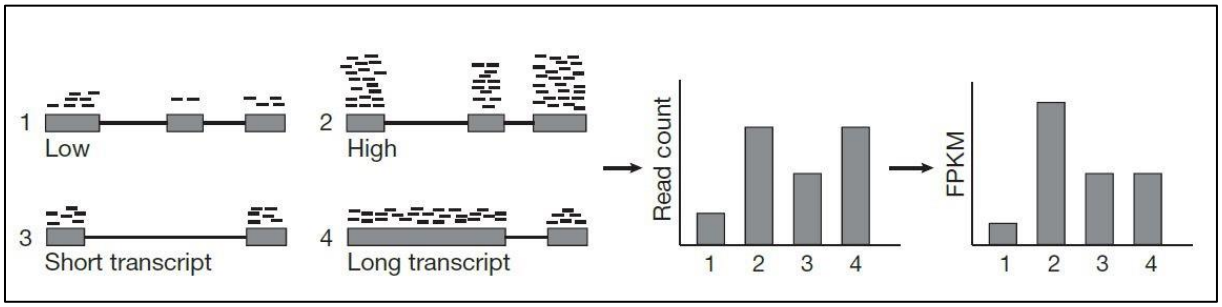


Figure 3.5 : Transcrits de différentes longueurs avec différents niveaux de couverture de reads (gauche), nombre total de reads pour chaque transcrit (milieu) et nombre de reads normalisés par FPKM (droite). Extrait de Garber *et al.*, 2011.

3.2.4.5. Différentiel d'expression

Une fois quantifiés et normalisés, les échantillons peuvent être comparés par des méthodes statistiques afin de déterminer quels sont les gènes différentiellement exprimés (DE) entre plusieurs conditions. Le package EdgeR de Bioconductor (Open Source Software for bioinformatics) qui réalise des tests statistiques à partir des comptages pour déterminer l'expression différentielle des gènes applicables au RNASeq a été utilisé dans le cadre de ce travail. Les tests sont basés sur la loi binomiale négative. Pour chaque gène et chaque test (comparaison entre les groupes étudiés) une p-value est calculée qui est comparée à une valeur seuil que l'on choisit. Cependant dans le cas de l'analyse différentielle, des tests indépendants sont réalisés sur plusieurs milliers de gènes simultanément. Cela tend à augmenter considérablement le nombre de faux-positif[30] (Storey, 2003) c'est-à-dire de gènes déclarés DE alors qu'en réalité ils ne le sont pas. Il faut donc corriger les p-values pour tenir compte de ce biais. La méthode de correction utilisée dans ce travail est la FDR (False Discovery Rate) (Benjamini and Hochberg, 1995). Les p-values corrigées (FDR) sont ensuite comparées à une valeur seuil choisie. Elles doivent être inférieure à cette valeur pour que le gène soit déclaré DE. Bien qu'il n'existe pas véritablement de convention sur les valeurs seuils de la FDR, il est recommandé d'utiliser les paramètres de la FDR à 10, 5 ou 1%.

Afin de filtrer sur les niveaux d'expression des gènes DE le rapport logarithmique entre deux conditions, aussi appelé ratio du log de "Fold Change" ($FC = \frac{expression\ condition\ 1}{expression\ condition\ 2}$) est calculé (Cui and Churchill, 2003). Le seuil du logFC a été fixé à 1 ce qui signifie qu'il y a au moins un rapport de 2 entre l'expression du gène dans la conditon 1 par rapport à la condition 2. Le logFC est négatif pour un gène sous-exprimé et positif pour un gène surexprimé.

3.2.4.6. SNP calling : du prétraitement des données à la prédiction des SNPs

Un SNP (Single Nucleotide Polymorphism) est une mutation ponctuelle ne modifiant qu'un nucléotide entre les séquences génomiques de deux individus d'une même espèce. La recherche de variants SNPs se base sur la comparaison des reads alignés avec une référence (génome de référence ou transcriptome *de novo*) (Olson et al., 2015).

Préalablement au SNP calling, il est nécessaire de nettoyer les fichiers d'alignement (au format BAM ou SAM) afin de diminuer les risques d'erreurs dans l'identification des SNPs.

[30] Faux-positif désigne dans cette étude les gènes déclarés différentiellement exprimé alors qu'ils ne le sont pas.

Ces risques d'erreurs sont inhérents aux méthodes d'amplification de l'ADN et de séquençage. Les bonnes pratiques de GATK (Genome Analysis Tool Kit) recommande l'utilisation d'outils (Figure 3.6) pour effectuer le prétraitement des fichiers d'alignement des séquences en vue d'effectuer du SNP calling (McKenna et al., 2010). Nous détaillerons ici les erreurs les plus connues ainsi que les outils bioinformatiques mis en place pour les corriger :

(1) L'une des erreurs les plus communes dans l'analyse de SNP calling est la présence de « duplicats », autrement dit des artefacts créés au moment de la PCR, à l'étape de préparation des banques d'ADNc. Ces duplicats se présentent sous la forme de paires de *reads* alignés sur la même localisation génomique du même brin d'ADN (Alkan et al., 2011). Les duplicats augmentent artificiellement le nombre de SNP et conduisent à une identification erronée (DePristo et al., 2011). Afin de pallier au problème des duplicats, GATK recommande l'utilisatio de l'outil MarkDuplicate, permettant des les identifier et de les retirer des séquences.

(2) Les INDELS (Insertion/Délétions de Nucléotides) représentent une large source d'erreur qui modifie la quantité de nucléotide et la qualité de la séquence. Un INDEL se définit soit par l'ajout d'un ou de plusieurs nucléotides à une séquence donnée (insertion), soit parla suppression d'un ou plusieurs nucléotides (délétion). Ces variations structurales sont prises en compte dans le workflow recommandé par GATK avec l'outil « Realignment around indels ». Cet outil améliore la qualité des séquences en effectuant un alignement optimal au niveau des INDELS.

(3) Les erreurs de lecture par le séquenceur lui-même peuvent générer des erreurs sur l'identification des SNPs. Pour pallier aux erreurs de lecture, il est conseillé d'effectuer une étape de recalibration des scores de qualité de chaque base nucléotidique. Cela peut être entrepris par l'outil « Base Recalibration » de GATK (DePristo et al., 2011).

Suite à ce pré-traitement des données, l'étape du SNP-calling peut être effectuée. Dans son workflow d'analyse, GATK recommande l'utilisation de deux types d'outils pour détecter les SNPs: Unified Genotyper et Haplotype Caller. Les algorythmes de ces deux outils détectent les variations de bases par rapport aux séquences d'un individu de référence et attribuent un génotype par individu à chaque site de variation. Le fichier qui résulte de l'analyse de SNP calling est un format VCF (variant calling format).

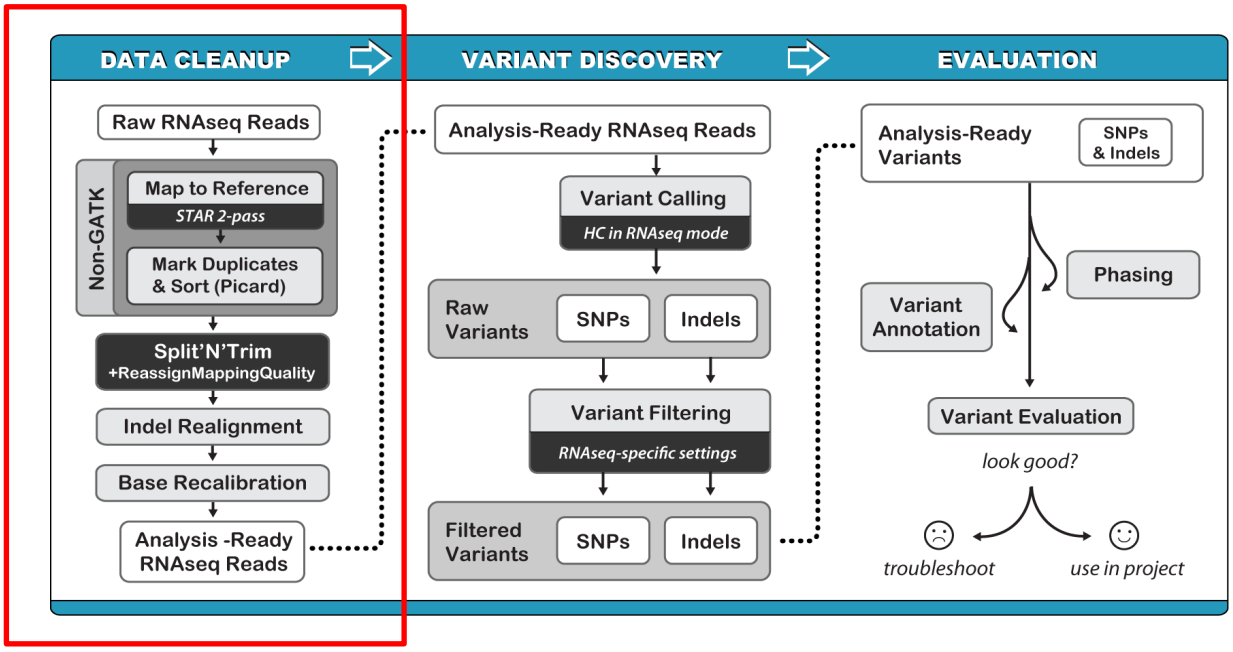


Figure 3.6 : Workflow d'analyses recommandé par GATK.

L'encadré rouge représente les étapes de pré-traitement des fichiers d'alignement de séquences. Source : GATK Best Practices.

3.3. Article : A transcriptomic-based approach to identify candidate markers for resistance to myrtle rust (*Austropuccinia psidii*) in three endemic Myrtaceae species from New Caledonia.

Authors list:

Soewarto Julia[1*], Hamelin Chantal[2,4*], Bocs Stéphanie[2,4, 6], Mournet Pierre[2,6], Vignes Hélène[2,6], Berger Angélique[2,6], Armero Alix[2,6], Martin Guillaume[2,6], Dereeper Alexis[3,4], Sarah Gautier[2,4,6], Carriconde Fabian[1], Maggia Laurent[1,5, 6]

[1] *Institut Agronomique néo-Calédonien (IAC), 98848 Noumea, New Caledonia*

[2] *CIRAD, UMR AGAP, F-34398 Montpellier, France*

[3]*IRD, UMR IPME (IRD-UM2-Cirad) 911, avenue Agropolis, BP 64501 34394 Montpellier Cedex 5, France.*

[4] *South Green Bioinformatics Platform, Montpellier, France*

[5] *CIRAD, UMR AGAP, 98800 Noumea, New Caledonia*

[6]*AGAP, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France*

**The authors have equal contribution to the manuscript*

3.3.1. Abstract

The myrtle rust disease, caused by the fungus *Austropuccinia psidii*, infects a wide range of host species within the Myrtaceae family worldwide. Since its first report in 2013 in New Caledonia, it was found on various types of native environments where Myrtaceae are the dominant or codominant species, as well as in several nurseries. It is now considered as a significant threat to ecosystems biodiversity and Myrtaceae-related economy. Finding predictive molecular markers for resistance against Myrtle rust may help to identify resistant plants and provide a means of maintaining Myrtaceae production in nursery, especially for ecosystems restoration purposes. Next generation sequencing and analysis tools may provide new research opportunity for work in non-model organism, such as the endemic Myrtaceae. We described here a pre-breeding methodology to find disease resistance markers in three endemic Myrtaceae, by combining RNA-sequencing, differential gene expression (DGE) and variant calling. DGE and variant calling were conducted in parallel using the closest and only available sequenced Myrtaceae genome (*Eucalyptus grandis*) and a *de novo* assembly transcriptome from each species.

Keywords: Myrtle rust, Myrtaceae, RNA-Seq, Differential gene expression, SNP calling, *de novo* transcriptome, functional annotation

3.3.2. Introduction

Myrtaceae is a large flowering plant family that comprises around 142 genera, 5500 species of shrubs or trees, and occurring mainly in pan-tropical regions with centers of diversity in South America, Southeast Asia and Australia (Thornhill et al., 2015; Wilson, 2010). It comprises several genera of outstanding ecological and economic relevance worldwide. Valuable products provided by this family include timber, paper pulp and energy production (Eucalyptus spp), spices (clove: *Syzygium aromaticum*, allspice: *Pimenta dioica*), fleshy fruits (*Psidium*-guavas; *Eugenia*-pitanga) and ornamentals (Reynertson et al., 2008). A remarkable aspect in this family is the richness of essential oil, tannins, phenolic and terpenes compounds making it one of the most valuable plant families for cosmetic and therapeutic industry (Keszei et al., 2008; Padovan et al., 2014). Many pest and pathogens are known to infect Myrtaceae (Simpson et al., 2006). Among these, the rust disease caused by the rust fungi *Austropuccinia psidii* (syn. myrtle rust) is one of the major threat for Myrtaceae worldwide (Beenken, 2017; Glen et al., 2007). Myrtle rust is a foliar disease which attack actively growing leaves, shoot tips, young stems, causing distortions and dieback following the emergence of urediniospores (Tommerup et al., 2003). Inflorescence of several species, including their fruits, can also be impacted (Glen et al., 2007). Reduced fertility, depletion of resource for growth, dieback, crown loss, and finally plant death are the main impacts resulting from repeated infections of *A. psidii* (Carnegie et al., 2016b). The host-range of this fungus is unprecedented for a biotrophic pathogen, reaching at least 460 species Myrtaceae hosts species worldwide (Carnegie et al., 2016b; Giblin and Carnegie, 2014).

The French archipelago New-Caledonia is home of approximately 257 Myrtaceae species, of which 99% are endemic (Morat et al., 2012; Snow, 2009). Myrtaceae is one of the major families that dominate or co-dominate the different types of vegetation communities that occur in the archipelago, especially in rainforests, sclerophyllous forest, savannas and maquis shrubland (Bouchet et al., 1995; Ibanez et al., 2014; Perry and Enright, 2002; Read et al., 2000). Because of their ecologically importance and abundance, Myrtaceae are often chosen as pioneer species in revegetation programs to mitigate deforestation of mining land (L'Huillier et al., 2010). In addition to many valuable economic resources cited above, some Myrtaceae species have a patrimonial value for Caledonian people (folk-medecine use, timber for traditional construction, and iconic species from Caledonian landscape). *A. psidii* was first detected in New Caledonia in March 2013. Since then, the pathogen thrives on the territory by invading all types of Myrtaceae-colonized habitats and infecting at least 67 endemic species

(Soewarto et al., 2017 accepted). In the neo-caledonian context where biodiversity conservation is a priority, myrtle rust represents a threat for the health and stability of Myrtaceae-dominated vegetation communities and for various economic activities that rely on their cultivation.

The main control/eradication strategies known against *A. psidii* are chemical treatments with the use of fungicides and host resistance (Glen et al., 2007). Breeding for host resistance is regarded as the most economically viable, and environmentally-friendly, approach to control *A. psidii* where it has established (Glen et al., 2007). Disease control through selection for resistant to *A. psidii* has long been carried out in *Eucalyptus* species in Brazil (Dianese, 1984; Xavier et al., 2001). By molecular approaches, a major resistance gene to *A. psidii,* designated *Ppr*1 (*Puccinia psidii resistance gene*-1), was identified in *Eucalyptus grandis* (Junghans et al., 2003). This locus has been successfully positioned on linkage group 3 (~chromosome 3) of the Eucalyptus genome (Mamani et al., 2010). Identification of several markers tighly linked to this disease-resistant gene has been performed to facilitate the marker-assisted selection of resistant *E. grandis* genotype (Junghans et al., 2003; Laia et al., 2015; Mamani et al., 2010).

In light of these previous researches, we aim to identify molecular variations in genes that confer resistance to *A. psidii* in new Caledonian endemic Myrtaceae. Using molecular markers linked to these genes in a marker-assisted based selection may help to identify resistant genotype and provide a means of maintaining Myrtaceae production in nursery, especially for ecosystems restoration purposes in New Caledonia.

Although the pathosystem *A. psidii*-Eucalyptus has been well studied, knowledge about resistance mechanisms in other Myrtaceae species is far less known in particular because of the lack of genetic resource. Indeed, *Eucalyptus grandis* is the only representative of Myrtaceae family whose genome has been sequenced and annotated (Myburg et al., 2014). A recent alternative for genetic analysis in non-model species without a reference-genome is RNA sequencing, also called RNA-seq (Harper, 2016). Transcriptome-base methods are cost-effective strategy for gene discovery, development of molecular markers and large scale analyses of genetic variation associated with trait variation across a diversity panel (Wang et al., 2009).

In this paper we described a new pre-breeding methodology to find predictive rust resistance markers in three endemic Myrtaceae species from New-Caledonia that previously showed variation in disease response to myrtle rust: *Arillastrum gummiferum, Tristaniopsis glauca*

and *Syzygium longifolium* (Soewarto et al., 2017 accepted). We extracted RNA from resistant and susceptible individuals for each species against *A. psidii*. We generated RNA-seq datasets from these two contrasting phenotypes. In order to obtain a better understanding of the genes involved in resistance or susceptibility to *A. psidii* and provide a reliable analysis of the transcriptomic landscape of gene regulation in host Myrtaceae species, we compared two approaches to treat the data: (1) we took advantage of the recently published and annotated genome of a related species, *E. grandis*, and performed differential expression (DE) and SNP calling; (2) we also mapped out dataset to a reference transcriptome assembly obtained for each studies species and performed differential expression and SNP calling.

3.3.3. Material and methods

Plant Material

Three Myrtaceae species of interest were chosen for their panel of contrasting responses to myrtle rust: *Arillastrum gummiferum*, *Syzygium longifolium* and *Tristaniopsis glauca*.

These species were grown in an open-field nursery where myrtle rust disease was naturally present. These three species were harvested at the same time in April the 28th, 2015.

For *T. glauca*, a population of contrasting individuals regarding the response to myrtle rust was sampled in May, the 1st, 2015 in natural environment (Bois du Sud, protected area from South Province of New Caledonia). *T. glauca* samples from natural field "Bois du Sud" were treated separately from those from the nursery. In the subsequent analysis we will use *A. gummiferum*-FAR, *S. longifolium*-FAR, *T. glauca-FAR* for individuals from the nursery, and *T. glauca*-BDS for individuals from natural field.

Individual free of symptoms (considered as resistant) and infected ones (considered as susceptible) were sampled at a juvenile stage for the most recent leaves. The sampling details are provided in Table 3.1. All samples were frozen at the time of collection and then stored at −80°C until total RNA was extracted.

Table 3.1: Detailed sampling of three Myrtaceae species for RNA-seq analysis

Sample name	ID sampling	Biological material	Species	Sampling site	Life stage	Phenotype toward myrtle rust infection
Sample 1	Ag19	leaf	*Arillastrum gummiferum*	nursery	juvenile	susceptible
Sample 2	Ag28	leaf	*Arillastrum gummiferum*	nursery	juvenile	susceptible
Sample 3	Ag2	leaf	*Arillastrum gummiferum*	nursery	juvenile	resistant
Sample 4	Ag3	leaf	*Arillastrum gummiferum*	nursery	juvenile	resistant
Sample 5	Ag4	leaf	*Arillastrum gummiferum*	nursery	juvenile	resistant
Sample 6	Ag6	leaf	*Arillastrum gummiferum*	nursery	juvenile	susceptible
Sample 7	Syl10	leaf	*Syzygium longifolium*	nursery	juvenile	resistant
Sample 8	Syl13	leaf	*Syzygium longifolium*	nursery	juvenile	susceptible
Sample 9	Syl15	leaf	*Syzygium longifolium*	nursery	juvenile	susceptible
Sample 10	Syl18	leaf	*Syzygium longifolium*	nursery	juvenile	resistant
Sample 11	Syl2	leaf	*Syzygium longifolium*	nursery	juvenile	resistant
Sample 12	Syl4	leaf	*Syzygium longifolium*	nursery	juvenile	resistant
Sample 13	Syl7	leaf	*Syzygium longifolium*	nursery	juvenile	resistant
Sample 14	Tg2	leaf	*Tristaniopsis glauca*	nursery	juvenile	susceptible
Sample 15	Tg3	leaf	*Tristaniopsis glauca*	nursery	juvenile	resistant
Sample 16	Tg4	leaf	*Tristaniopsis glauca*	nursery	juvenile	susceptible
Sample 17	Tg5	leaf	*Tristaniopsis glauca*	nursery	juvenile	susceptible
Sample 18	Tg6	leaf	*Tristaniopsis glauca*	nursery	juvenile	susceptible
Sample 19	V1	leaf	*Tristaniopsis glauca*	natural field	juvenile	susceptible
Sample 20	V2	leaf	*Tristaniopsis glauca*	natural field	juvenile	susceptible
Sample 21	V3	leaf	*Tristaniopsis glauca*	natural field	juvenile	resistant
Sample 22	V4	leaf	*Tristaniopsis glauca*	natural field	juvenile	resistant
Sample 23	V6	leaf	*Tristaniopsis glauca*	natural field	juvenile	susceptible
Sample 24	V7	leaf	Tristaniopsis glauca	natural field	juvenile	susceptible

Total RNA extraction

Total RNA was extracted using 3-10 g of fresh material from the cetyltrimethylammonium bromide-based protocol (CTAB) as described by Salzman et al., 1999. Briefly, 3-10 g of frozen leaves from each individual was ground in liquid nitrogen using a mortar and pestle. Then 3-10 mL of pre-heated extraction buffer (2% CTAB, 2% polyvinylpyrrolidone, PVP-40 (2% w/v), 2% β-mercaptoethanol, 100 mM Tris-HCl, 25 mM EDTA and 2 M NaCl) was added to the grinded samples, and incubated at 65°C for 30 min. An equal volume of mixture of chloroform:isoamyl alcohol (24:1) was added and mixed immediately for 2 min using a vortex mixer. The samples were then centrifuged at 10,000 g for 10 min. The upper aqueous phase was transferred to new tubes and 1/3 volume of 10 M LiCl was added. The samples were mixed and stored at 4°C overnight. The samples are then centrifuged at 18,000 g for 20 min. The supernatant was removed and the pellet is washed with 75% ethanol and air-dried. The pellet was suspended in 30 µl of RNase-free water and 70 µl of SSTE buffer (1M NaCl, SDS (0.5% w/v), 10 mM Tris-HCl, 1 mM EDTA). An equal volume of acid:phenol:chloroform:isoamyl alcohol (25:24:1) is added to each sample, and vortexed. The samples were then centrifuged at 12,000 g for 10 min and the upper aqueous phase was transferred to new tubes with 2 volumes of cooled ethanol 100% and 1/10 volumes of NaAc (pH5.2). The samples are then mixed and incubated at -20°C for 2 hrs. The samples were centrifuged at 18,000 g for 20 min. The supernatant was removed and the pellet was washed three times with 75% ethanol before being air-dried. Finally, the pellet is resuspended in 30 µl of RNase-free water.

The extraction is followed by removal of DNA with the TURBO DNA-*free*™ kit (Ambion) according to the manufacturer's instructions. RNA quantity and quality control was performed using a 2100 Bioanalyzer (Agilent Technologies).

cDNA library preparation and sequencing

Paired-end Illumina mRNA libraries were generated using the TruSeq RNA-Seq Sample Prep kit according to the manufacturer's protocol (Illumina Inc., San Diego, CA, USA). Briefly, Poly-A containing mRNA molecules were isolated using poly-T oligo-attached magnetic beads. The purified mRNA was then chemically fragmented. Reverse transcription of first- and second-strand of cDNA are performed and were followed by end repair. Single 'A'

nucleotide is added to the 3' ends of each fragment before ligation of adapters. The purified cDNA templates were enriched by PCR to form libraries of 300 pb. Each indexed cDNA library was verified and quantified using a 2100 Bioanalyzer (Agilent Technologies). The final libraries were then quantified by qPCR with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA) and normalized to 20 nM before being pooled. Sequencing was conducted on a single lane of a flow cell on Illumina HiSeq™ 2500 (Genotoul platform, INRA) as paired-ends reads of length 150 bp.

RNA-seq data processing

RNA-seq data cleaning processing was followed according the ARCAD (Agropolis Resource Center for Crop Conservation, Adaptation and Diversity) workflow analysis (http://arcad-bioinformatics.southgreen.fr/) in command line as described in Figure 3.7. Scripts are available on the SouthGreen GitHub repository (https://github.com/SouthGreenPlatform/arcad-hts).

Raw reads cleaning

Raw reads were processed with Cutadapt (Martin, 2011) using the script arcad_hts_1_cutadapt_in_chain.pl and the TruSeq index sequences corresponding to the samples. We also used Cutadapt to improve reads quality by trimming the start and the end of each reads. We then filtered the reads on the basis of their mean quality score, keeping those with a mean quality higher than 30 using arcad_hts_2_Filter_Fastq_On_Mean_Quality.pl. Reads with length inferior to 35 bp were discarded. Each processing steps were followed by a FastQC quality control of reads (arcad_hts_0_fastqc_in_chains.pl). Thereafter, single reads (i.e. those for which the mate pair was discarded in the previous steps) were separated from paired-reads using arcad_hts_3_synchronized_paired_fastq_end.pl.

Reference- guided method

Aligning to a reference genome from a related species

The reference genome assembly and annotation of *E. grandis* were downloaded from NCBI RefSeq portal (https://www.ncbi.nlm.nih.gov/assembly/GCF_000612305.1/).
Reads were mapped to reference genome *Eucalyptus grandis* (EUCGR) using HISAT 2.04 (Kim et al., 2015). The aligned data were passed to StringTie (Pertea et al., 2016) for

transcript assembly. A reference genome annotation file in GFF3 format (intron/exon positions) was provided to guide the transcripts assembly. The resulting binary files (.bam) were then filtered on quality using SAMtools view (-q0 –F4) to remove unmapped and multimapped reads (Li et al., 2009). Finally, the reads were sorted using SAMtools sort. Mapping statistics were verified using SAMtools flagstat (Li et al., 2009).

Calling SNPs using E. grandis reference genome

Variant calling algorithms compare mapped reads to the reference genome and identify potential variants. We followed the best practices recommended by the GATK (Genome Analysis Toolkit) pipeline (McKenna et al., 2010). Consistent with GATK's recommendations, mapped reads against *E. grandis* reference genome have been submitted to cleaning process before SNP calling step. We used Picard tools to remove PCR duplicates (MarkDuplicates) and reorder reads to match the contigs according to the reference genome (ReorderSam). Then, we used the GATK tool SplitNCigarReads (with parameters -RMQF 255 -RMQT 60 -U ALLOW_N_CIGAR_READS), which splits reads into exon segment and hard-clip any sequences overhanging into the intronic regions. The reads were realigned around INDELS and recalibrate base quality values using RealignerTargetCreator and IndelRealigner GATK tools. Once the reads were pre-processed with Picard and GATK tools, variant calling was undertaken by two programs: Haplotype Caller (GATK) and an in-house script (Martin et al., *in prep.*, Baurens et al., *in prep.*). Contrary to GATK, which use statistics based on population genetics (which is not the case here), the in-house program only count the number of reads supporting each bases (A, T, G, C) at each covered sites for each accession. Based on this base count for each accession, a genotype was emitted based on a binomial test. The analysis pipeline we used is illustrated in Figure 3.8.

De novo methods

De novo transcriptome assembly

In order to identify pathways of responses, potentially specific to each species, and overcome divergence between the targeted species and the reference, we constructed a *de novo* assembly for each species.

De novo transcriptome assembly for the three species was carried out using Trinity software

with 50 GB of memory (Inchworm, Chrysalis, and Butterfly modules) (Grabherr et al., 2011). The individual with the more reads was chosen as the representative of the species for each *de novo* transcriptome. Briefly, overlapping k-mers are extracted from the cleaned paired-reads. Inchworm module assembled the reads into contigs. Next, Chrysalis module ranked Inchworm contigs into clusters and constructed complete de Bruijn graphs for each cluster. Finally, Butterfly module processed the individual graphs in parallel to reconstruct transcript sequences in a manner that reflects the original cDNA molecules. To avoid redundant transcripts, we kept the longest isoform for each "trinity gene". Assembly statistics (N50, contig length, GC content, etc) were computed by TrinityStats.pl embedded in Trinity.

Aligning to the de novo transcriptomes

We used the Burrows-Wheeler alignment tool (BWA-MEM) to map the cleaned reads from the 24 samples to the corresponding *de novo* transcriptome assembles of each species (Li, 2013). The read aligner, BWA-MEM, aligns each mate of a paired-end read at the same time and produces SAM/BAM files containing the alignments. Samtools view from the SAMtools 1.3 package (http://www.htslib.org/doc/samtools-1.1.html) was used to sort and index the BAM files by coordinate and remove duplicates (parameters -q0 -F4).

Calling SNPs using de novo transcriptome assemblies

We processed the aligned reads to a cleaning process on Galaxy platform (Afgan et al., 2016). Based upon established GATK best practices, we used Realigner targer creator and IndelRealigner programs (LOD=5) to fix the misalignments due to the mapping process and perform a local realignment near the INDELS (DePristo et al., 2011; Van der Auwera et al., 2002). Variant calling was launched both with Unified Genotyper (GATK), using a minimum phred-scaled confidence threshold of 30 (McKenna et al., 2010), and an in-house scripts (Martin et al., *in prep.*, Baurens et al., *in prep.*).

Variant filtering

Because SNP calling implied to detect any difference between the reads and the reference (*E. grandis* reference genome or de novo transcriptome), the initial output is large and must be filtered. Several criteria were used to exclude the less reliable SNPs from the dataset.

Firstly, we applied an in-house pre-filter script (VcfPreFilter.1.0.py) on .vcf files resulting from the in-house script SNP calling, in order to compare them with those resulting from the GATK methods. This pre-filter included the following parameters: a minimal site coverage (5 reads), a maximal site coverage (1000 reads) per SNP, a minimum allele frequency (0.01) and a minimum allele coverage (2 reads per allele).
Secondly, we applied other thresholds to all the vcf output regardless of the SNP calling used. We used the program VCF-filter, which required: a minimum coverage (10), a maximal coverage (1000), a minimum frequency (0.05), a minimum allele coverage (3) to keep a genotype; and no missing data were allowed to keep a variant site.
Thirdly, we used an in-house script to remove monomorphic and multiallelic variants.
And finally, as recommended by GATK, we used the VariantFiltration program to remove clustered SNP: in a window size of 10 pb we considered that 3 SNPS constitute a cluster.
The final SNP set comprised the de novo identified SNPs that had passed all filters.
We compared the overlap among SNP positions obtained for each calling methods using vcf-compare from VCFtools (Danecek et al., 2011)

Differential gene expression analysis

The edgeR (Bioconductor package) was used to identify Differentially Expressed Genes (DEGs) on the three species aligned with the two types of reference (*E. grandis* genome and *de novo* transcriptome) (Robinson et al., 2010).
Because longer transcripts have higher read counts (at the same expression level), a common method for within-library normalization is to divide the summarized counts by the length of the gene. We used the RPKM (reads per kilobase per million mapped reads) to minimize the influence of variation in gene length and total number of reads (Mortazavi et al., 2008). RPKM values were log2 transformed for gene quantification. Differential expression was selected based on fold change ≥ 2 ($logFC \geq 1$ and $logFC \leq 1$) and false discovery rate (FDR)-adjusted *p* value threshold of 0.1. An FDR of 0.1 implies that we are willing to accept that 10% of the differentially expressed genes are false positives.

Annotation

We ran FrameDP V1.2.2 software with default parameters (Gouzy et al., 2009, for the prediction of coding regions in the unigenes using the *E. grandis* protein database from

Universal Protein Resource (UniProt-Swissprot). The polypeptides sequences obtained were submitted to the Basic Local Alignment Search Tool (BLAST) to search against homologs in the UnitProt (Swiss-Prot and Trembl) databases.

Data visualization

To make easier the analysis of the many different data types that were generated during this study (*de novo* transcriptomes assemblies, alignments, differential expression analysis, SNP calling and annotation), we used the JBrowse tool (Skinner et al., 2009). JBrowse is a web genome browser from GMOD project (Generic Model Organism Database project). It allows visualize on each tracks, the values of different data types for the genes along a reference sequence (*E. grandis* genome or *de novo* transcriptomes). A genome browser (Kent et al., 2002; Stalker et al., 2004; Stein et al., 2002) visually conveys the spatial relationships between different pieces of genomic data, helping users form hypotheses about their functional relationships. To build the jbrowse we rely on Skinner and Holmes, 2002 workflow and we imported the following files:

- The reference sequence files (*de novo* transcriptomes or *E. grandis* reference genome) in FASTA format
- The alignments files in the BAM format.
- The variants calling files (SNPs) in the VCF format
- The annotations files in the GFF3 format
- The EdgeR output files containing the RPKM (expression level) of each locus for the samples in the TSV format.

To load the following data we used the following perl scripts provided by the JBrowse developers: prepare-refseqs.pl for fasta files and flatfile-to-json.pl for .gff3 and .tsv files. To load the others data types we had to modify the two files tracks.conf and trackList.json where we described in a pre-defined format how the tracks will be displayed and indicated the names of to the corresponding files. We also managed to display a track showing for the SNPs the corresponding genotypes of the samples of the study. These genotypes are displayed in three different colors: cyan for heterozygotes, grey for homozygotes for the reference allele and deep blue for the homozygotes of the alternative allele. To achieve this we used the MVV (multivariantviewerjbrowsepluging) (https://github.com/elsiklab/multivariantviewer). To load

the MVV track data as the name of the sample, its category (sensible or resistant) and the color we wanted for each category we used the trackList.json file.

Datasets release

Raw RNA-seq data in FASTQ format will be deposited at the NCBI Sequence Read Archivedatabase (SRA). The transcriptome assembly, reads alignment, de novo transcriptome annotations and differential expression files will be deposited at DDBJ/EMBL/GenBank.

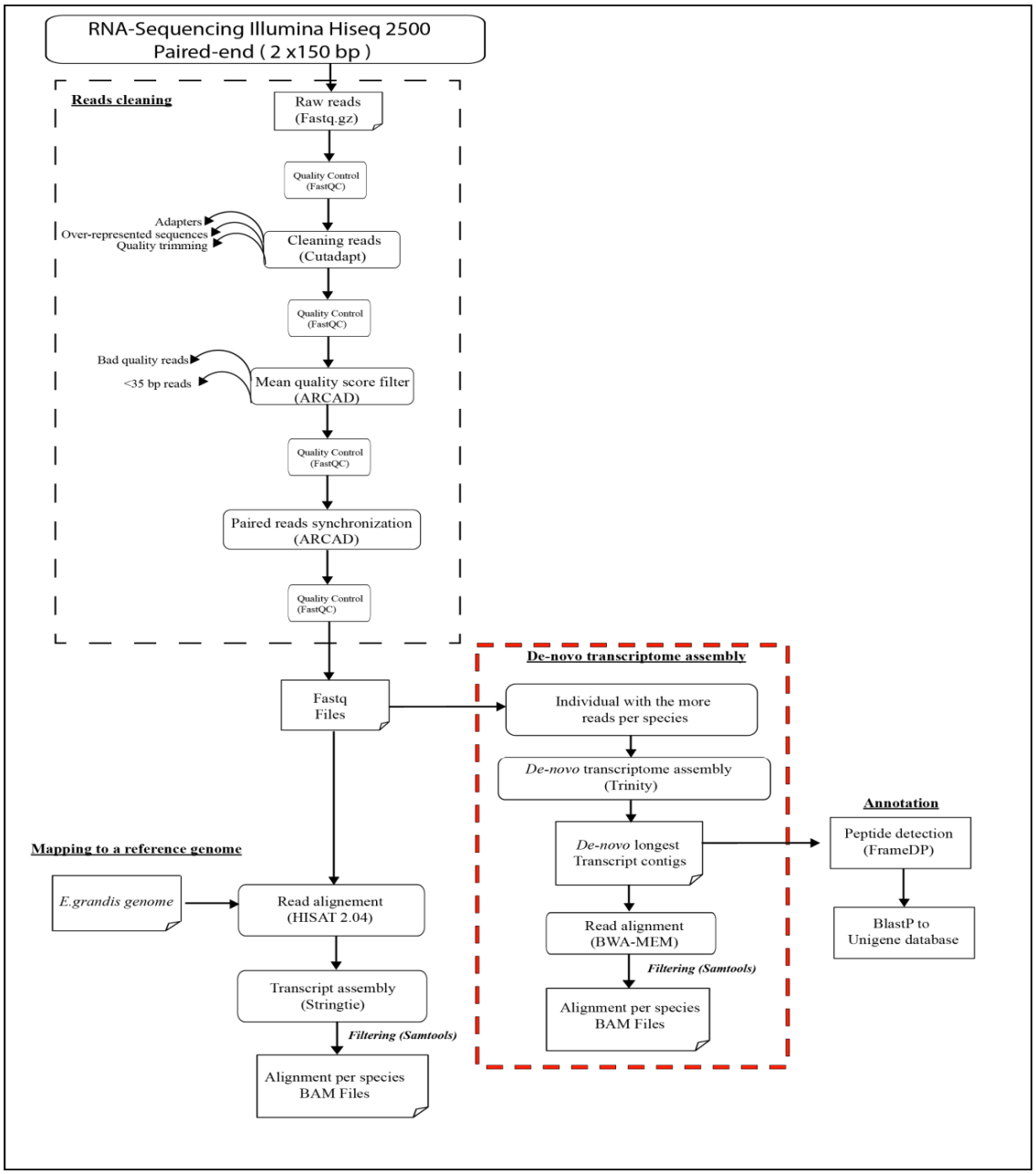


Figure 3.7: Bioinformatics pipeline showing the different steps involved in RNA-seq analysis until alignement to the two kind of reference. Additional steps required for de novo transcriptome assembly are shown in the red box.

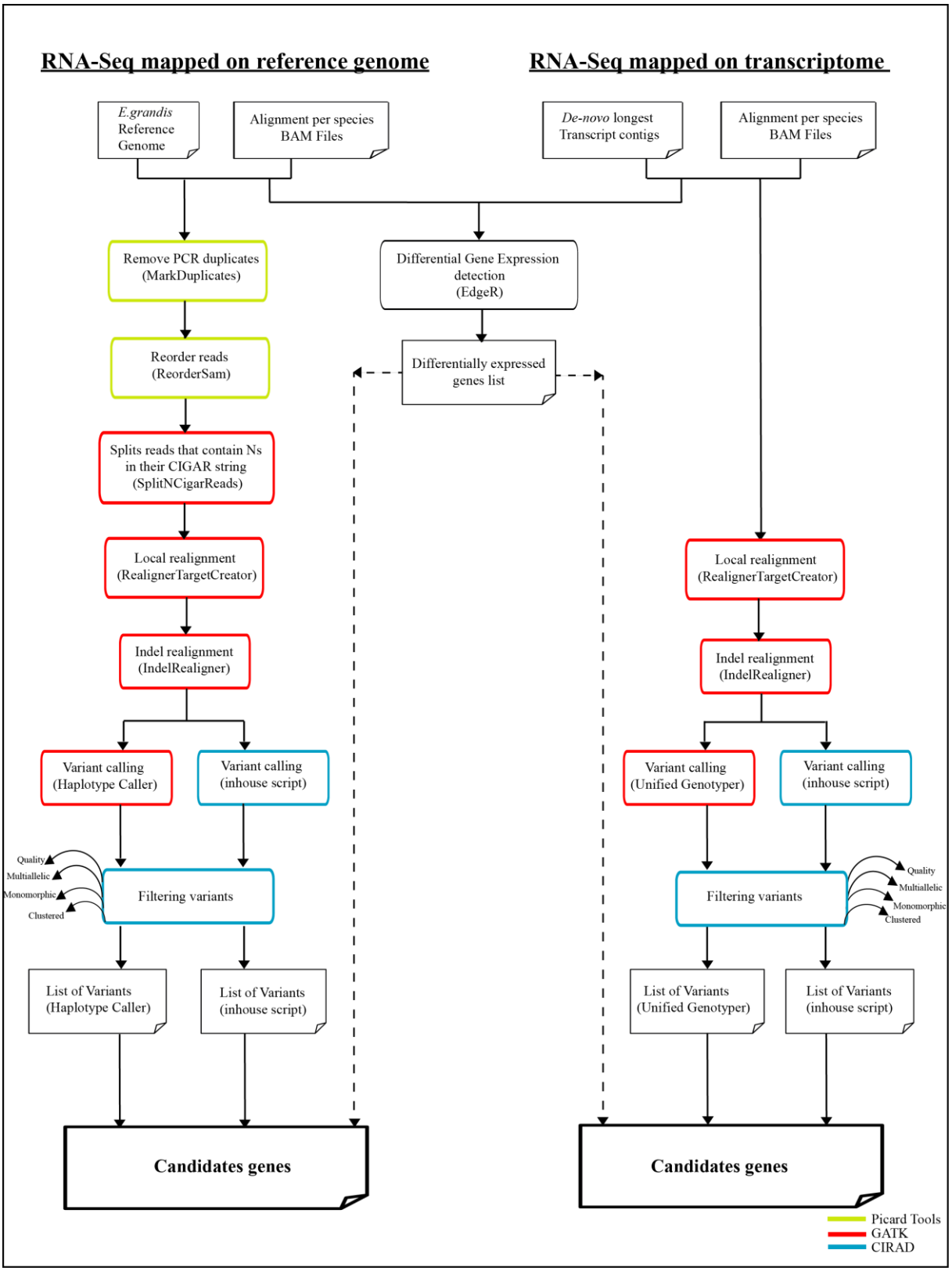
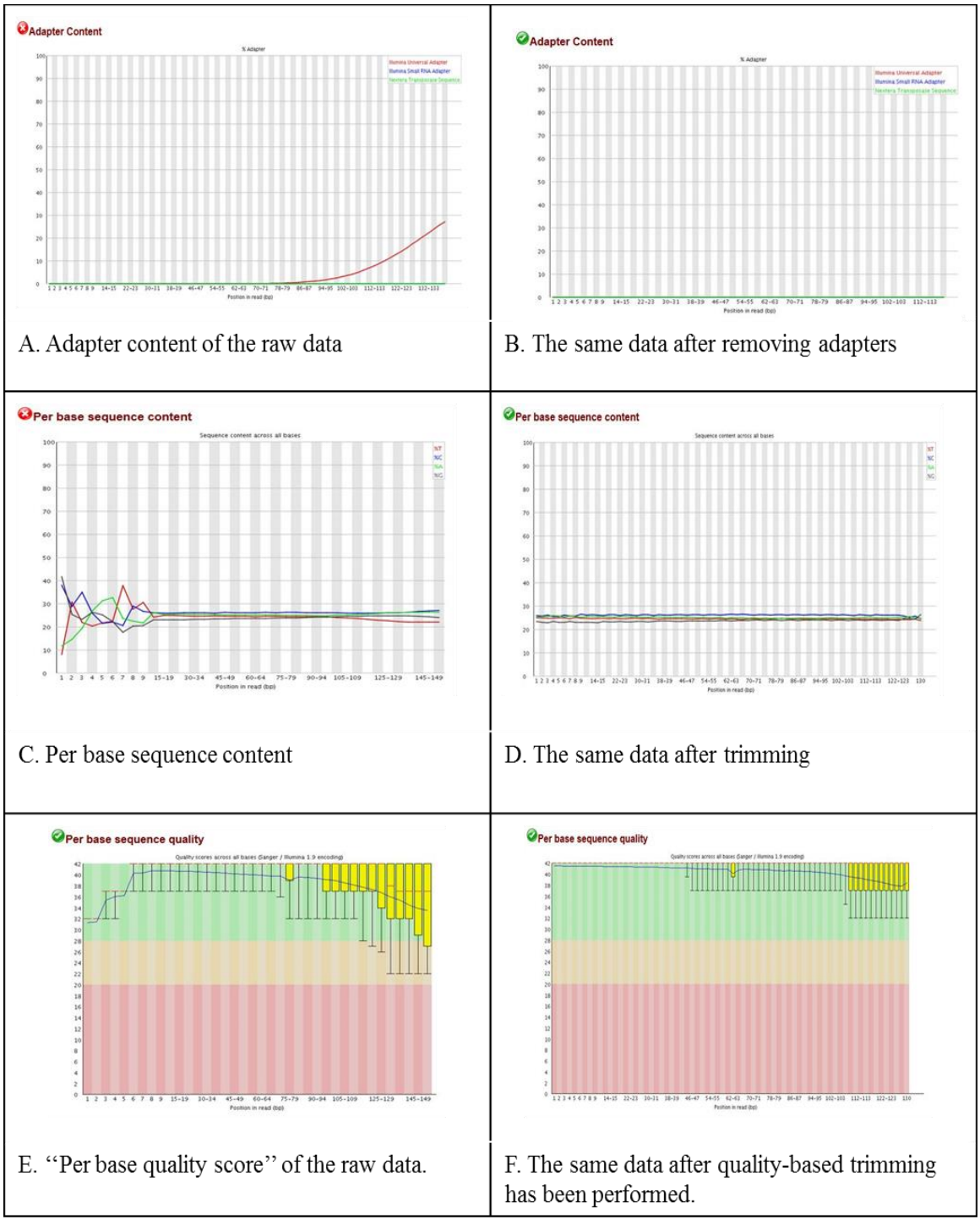


Figure 3.8: Analytic pipeline to identify potential candidate genes using both differentially expressed genes and SNP calling. The dotted arrows will not be treated in this work but in a future publication.

3.3.4. Results and discussion



A. Adapter content of the raw data

B. The same data after removing adapters

C. Per base sequence content

D. The same data after trimming

E. ‘‘Per base quality score’’ of the raw data.

F. The same data after quality-based trimming has been performed.

Figure 3.9: Quality control statistics generated by FastQC for individual Syl18 (*S. longifolium*) at different stages of the data cleaning process.

In this study, we constructed normalized cDNA libraries from resistant and susceptible individuals to *A. psidii* (myrtle rust) from three Myrtaceae species: *Tristaniopsis glauca, Arillastrum gummiferum* and *Syzygium longifolium*. The quality of our samples was suitable for the Illumina requirement sequencing. Details of RNA samples were supplied in Supplementary table 3.1.

3.3.4.1. Raw data summarization and cleaning reads

RNA sequencing with Illumina HiSeq2500 in paired-end, for the three species, produced a total of 668,165,595 raw reads. Of those, 96% were kept after the pre-processed cleaning stage (644,707,222). We improved the reads quality by trimming the first 15 bp and last 5 bp of each reads l and removing Illumina adapters as illustrated in Figure 3.9. A summary of the RNA-seq raw and clean data is presented in Table 3.2 and detailed in Supplementary table 3.2.

Table 3.2 : Number of raw and cleaned reads from the three species.

	A. gummiferum	*S. longifolium*	*T. glauca*-FAR	*T. glauca*-BDS
Number of libraries/individuals	6	7	5	6
Length of raw reads (bp)	150	150	150	150
Total number of raw reads	176,074,893	200,293,564	137,602,172	154,194,966
Total number of clean reads	169,892,296	193,48, 676	132,410,558	148,922,692
Length of cleaned reads (bp)	130	130	130	130

3.3.4.2. Transcriptome assembly

For each species, we selected the individual with the more reads and we used it to assembly a *de novo* transcriptome (Table 3.3; Supplementary table 3.2). Transcriptome assembly for *A. gummiferum* resulted in 117 839 putative transcripts with an average contig length of 501 bp and N_{50} of 1378 base pairs. Transcriptome assembly for *S. longifolium* yielded a total number of 89 782 putative transcripts with an average contig length of 530 bp in length and N_{50} of 1406 base pairs. Transcriptome assembly for *T. glauca-FAR* resulted in 108 823 putative transcripts, with an average contig length of 547 bp and N_{50} of 1396 base pairs. And finally, de novo transcriptome assembly for *T. glauca-BDS* resulted in 74 684 putative transcripts with an avegarge contig length of 525 bp in length and N_{50} of 1315 base pairs. However, Trinity *de novo* assembly resulted in larger number of transcripts than expected number of

genes, likely because of alternative splicing. To avoid redundant transcripts, we kept the longest isoform for each "gene" identified by TRINITY (unigene). Overall size of filtered de novo assembly yielded 57 Mb for *A. gummiferum*, 47 Mb for *S. longifolium*, 58 Mb for *T. glauca*-FAR and 39 Mb for *T. glauca*-BDS. Number of unigenes in each speacies were comparatively higher than in *E. grandis* genome (43,432) (Myburg et al., 2014).

Table 3.3 : Statistics of the *de novo* transcriptome assembly for each species using Trinity assembler.

	Species	***A. gummiferum***	***S. longifolium***	***T.glauca*-FAR**	***T.glauca*-BDS**
	Individual reference	**Ag3**	**Syl10**	**Tg4**	**V6**
Counts of transcripts	Total number of trinity genes (unigene)	84,919	64,716	76,982	53,527
	Total number of trinity transcripts	117,839	89,780	108,823	74,684
	Percent GC	45.86	46.37	44.45	46.21
Stats based on all transcript contigs	Contig N50	1,378	1,406	1,396	1,315
	Median contig length (bp)	501	530	547	525
	Average contig length (bp)	843.45	867.27	876.12	839.11
	Total assembled bases	99,391,026	77,863,296	95,341,718	62,667,790
Stats based on only LONGEST ISOFORM per 'GENE'	Contig N50	1,021	1,219	1,263	1,199
	Median contig length (bp)	386	402	421	415
	Average contig length (bp)	672.69	727.09	755.46	734.46
	Total number of assembled bases	57,124,398	47,054,345	58,156,816	39,313,402

3.3.4.3. Mapping cleaned reads to *E. grandis* genome and *de novo* transcriptome

We have chosen to compare the alignments rate using the genome of a related species (*Eucalyptus grandis)* and a *de novo* assembled transcriptome from each species. Mapping results are available in Table 3.4. High quality raw reads from each of the 24 libraries were mapped to *E. grandis* genome with HISAT 2.04, generating mapping rates of approximately 68% for *A. gummiferum*, *S. longigolium*, *T. glauca*- FAR and *T. glauca*-BDS. And around 5% of the total reads mapped correspond to singletons. By using *de novo* transcriptome assembles with BWA-MEM mapper, we improve the mapping rate to reach 96% of transcripts mapped and properly paired. Concomitantly, we reduced the percent of singletons to 0.1%.

Table 3.4 : Alignment statistics indicative of reads aligned to the assembled transcriptome using SamTools flagstat.

		***E. grandis* ref. genome**			***De novo* transcriptome**		
Species	Biological replicate	Total reads mapped	Properly paired (%)*	Singletons mapped (%)**	Total reads mapped	Properly paired (%)*	Singletons mapped (%)**
A. gummiferum	Ag19	22 195 213	70	4,8	19 953 353	95	0,1
	Ag28	23 611 351	71	4,8	20 537 263	96	0,1
	Ag2	25 655 914	68	4,5	23 462 050	96	0,1
	Ag3	27 318 150	68	4,6	24 139 896	96	0,1
	Ag4	18 576 164	69	4,5	17 399 264	97	0,1
	Ag6	21 277 541	74	4,2	17 259 913	96	0,1
S. longifolium	Syl10	26 685 822	68	6,1	25 218 899	96	0,1
	Syl13	16 676 981	65	5,0	16 426 865	97	0,1
	Syl15	17 382 267	63	5,2	15 522 277	97	0,1
	Syl18	19 263 137	69	5,4	18 980 838	96	0,1
	Syl2	23 370 932	67	5,3	22 382 429	96	0,1
	Syl4	23 806 097	66	5,7	21 148 273	95	0,1
	Syl7	19 456 046	66	5,1	19 209 000	97	0,1
T. glauca-FAR	Tg2	18 287 633	71	5,5	16 225 569	96	0,2
	Tg3	20 985 018	70	5,5	19 345 206	97	0,2
	Tg4	19 284 476	70	6,9	19 094 992	95	0,3
	Tg5	19 370 817	65	4,2	17 549 513	97	0,1
T. glauca-BDS	Tg6	17 264 359	67	4,9	16 126 662	96	0,1
	V1	12 259 586	58	3,6	12 446 883	99	0,0
	V2	20 957 190	70	4,6	18 089 903	98	0,1
	V3	18 561 244	66	4,1	17 126 183	98	0,1
	V4	17 177 833	66	5,1	17 252 577	96	0,1
	V6	22 656 615	66	4,9	20 231 181	97	0,0
	V7	19 387 135	70	5,7	17 984 129	97	0,1

*: Number of proper pairs in proportion to the total reads mapped reads.

**: Number of reads where one from a pair in proportion to the total mapped reads.

3.3.4.4. Differentially Expressed Genes

The total number of identified DEGs along with their up-/down-regulation are summarized in Table 3.5 and displayed as plotsmear in Supplementary Figure 3.2. In edgeR, dispersion was

estimated on a common dispersion basis for all species except for *T. glauca*-FAR mapped on *de novo* transcriptome for which no DE gene was found with this method. In this particular case, we used the tagwise dispersion. The differential expression analysis of genes was analyzed in susceptible individual compared to resistant ones. When using the *E. grandis* reference genome we showed that around 10% of the total expressed genes are differentially expressed in *A. gummiferum* and *S. longifolium*. Only 1-2% of total expressed genes are differentially expressed in *T. glauca*. When using the *de novo* transcriptome of each species, proportion of DE genes compared to other expressed genes decrease for all species, except for *A. gummiferum* (10%). We observed 7% of DE genes in *S. longifolium*, 1% in *T. glauca*-BDS and 0.2% in *T-glauca*-FAR.

Table 3.5 : Differentially expressed gene resulting from EdgeR.

Reference	Species	Total number of genes	DE genes (common dispersion)	over express genes (LogFC≥1)	under express genes (LogFC≤-1)
E. grandis genome	*A. gummiferum*	43 432	4727	3919	808
	S. longifolium	43 432	3971	2797	1174
	T. glauca-FAR	43 432	359	171	188
	T. galuca-BDS	43 432	742	617	125
de novo transcriptome	*A. gummiferum*	84 919	8849	2293	6556
	S. longifolium	64 717	4590	2809	1781
	T. glauca-FAR	41 905	84*	20	64
	T. galuca-BDS	53 527	512	352	160

*Tagwise dispersion

As gene expression differences exist among the two groups of individuals (resistant/susceptible), it should be expected that biological replicate of the same condition will cluster together. We used a multidimensional scaling (MDS) plot to see a spatial configuration of how similar or dissimilar the resistant and susceptible individuals are (Supplementary figure 1). We observed a clustering of individuals with the same phenotype for *A. gummiferum* and *T. glauca-FAR*. While for *S. longifolium* the individual Syl10 initially considered resistant grouped with susceptible individuals. On *T. glauca*-BDS, the susceptible individuals have very similar expression patterns, while the resistant ones are dispersed.

A cross comparison of differentially expressed genes (DEGs) is presented as Venn diagram illustrating the overlapped DEGs (Figure 3.10). We found 31 over-expressed genes in

susceptible individual that overlapped between the three species *A. gummiferum*, *S. longifolium* and *T. glauca* (FAR/BDS) (Table 3.6). No under-expressed gene was found in common between the three species, suggesting that resistance process toward *A. psidii* is potentially specific to each species.

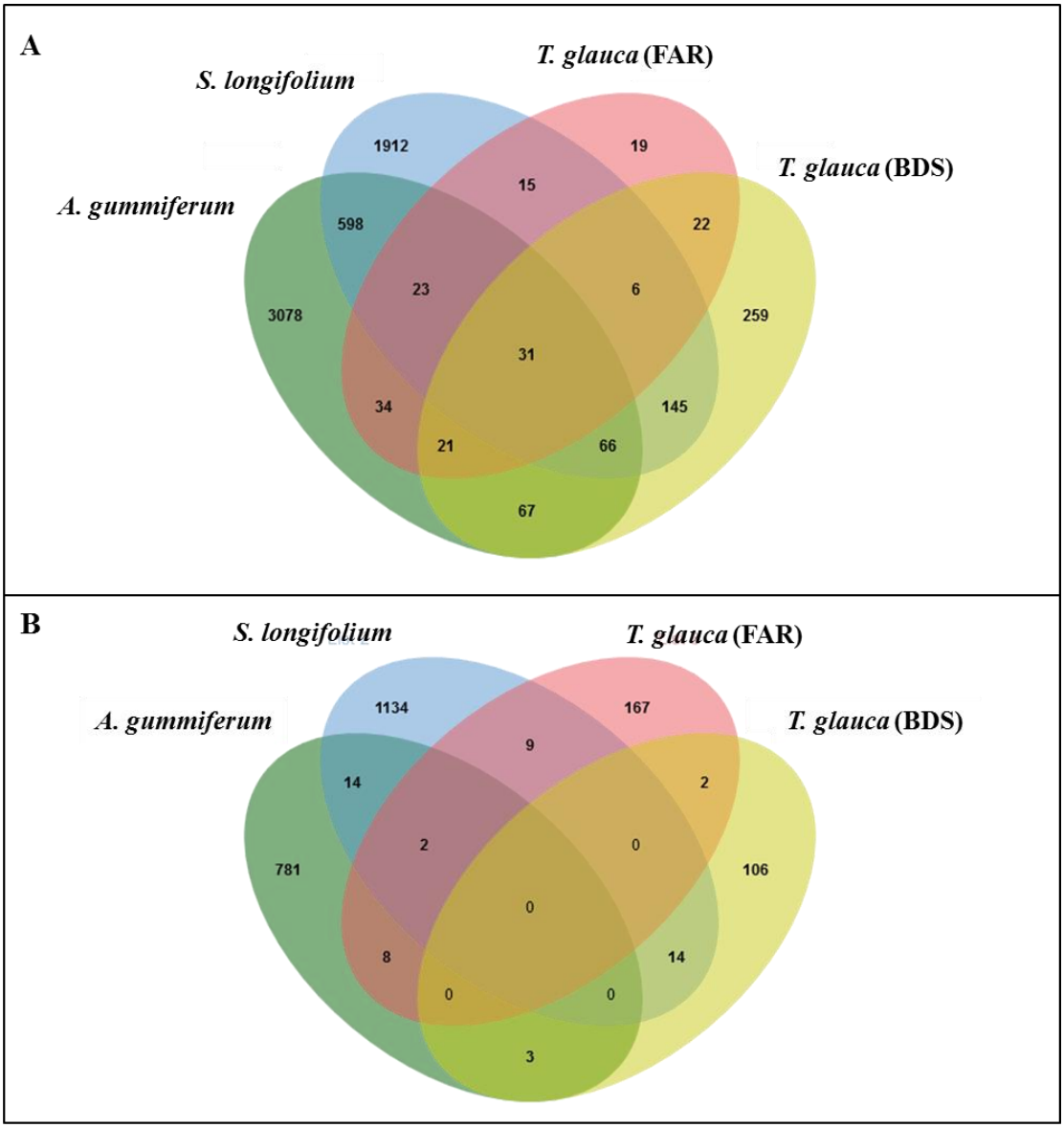


Figure 3.10: Venn diagram showing the differentially expressed genes in *A. gummiferum*, *S. longifolium* and *T. glauca* (FAR and BDS) using alignments with *E. grandis* reference genome. A: over-expressed genes in susceptible individuals and B: under-expressed genes in susceptible individuals.

Among the 31 over-expressed genes in susceptibles, we found several genes potentially involved in disease response processes (Table 6), such as a mitogen-activated protein kinase

kinase 9-like (LOC104425852) and various chitinase coding genes (LOC104456215, LOC104456221, LOC104456217, LOC104456219, LOC104456223, LOC104456220). Mitogen-activated protein kinase kinase are known for beeing essential component of the MAP kinase signal transduction pathway that regulated stress response to biotic factors (Innes, 2001; Tsuda and Katagiri, 2010). Chitinases are enzymes that hydrolyze the polymer chitin from many chitinolytic biotic aggressor (fungi, bacteria, viruses, viroids) (Punja and Zhang, 1993). They are considered as Pathogeneis-Related proteins playing crucial role in resistance against pathogens (Hammond-Kosack et al., 2007).

Table 3.6: List of common differential expressed genes between *A. gummiferum*, *T. glauca* and *S. longifolium* using *E. grandis* reference genome.

Gene name	Description	Scaffold	Position	
			Begin	End
LOC104430480	cationic peroxidase 1-like	scaffold 1	10988229	10991371
LOC104447583	1-aminocyclopropane-1-carboxylate oxidase homolog 4-like	scaffold 1	37081804	37083288
LOC104430395	cationic peroxidase 1-like	scaffold 1	10916608	10919639
LOC104450568	lichenase	scaffold 1	4937091	4938832
LOC104437048	germin-like protein subfamily 1 member 13	scaffold 3	29816958	29817912
LOC104442643	polyphenol oxidase%2C chloroplastic-like	scaffold 4	11999480	12001939
LOC104441046	polyphenol oxidase%2C chloroplastic-like	scaffold 4	11894466	11897452
LOC104443132	uncharacterized	scaffold 4	32944403	32946072
LOC104441265	germin-like protein subfamily 1 member 13	scaffold 4	19180316	19181238
LOC104443409	1-aminocyclopropane-1-carboxylate synthase-like	scaffold 5	896325	899493
LOC104456215	chitinase 6-like	scaffold 8	4487298	4488395
LOC104456221	chitinase 6-like	scaffold 8	4546083	4547171
LOC104457265	peroxidase 10-like	scaffold 8	25674984	25676740
LOC104415211	major allergen Pru ar 1-like	scaffold 8	58078042	58078973
LOC104415342	germin-like protein subfamily 1 member 13	scaffold 8	59956406	59957531
LOC104415205	major allergen Pru ar 1-like	scaffold 8	58026813	58027698
LOC104415213	major allergen Pru ar 1-like	scaffold 8	58084821	58085687
LOC104456217	endochitinase PR4-like	scaffold 8	4496392	4497507
LOC104456219	endochitinase PR4-like	scaffold 8	4522275	4523440
LOC104416148	uncharacterized	scaffold 8	4427514	4431668
LOC104456223	endochitinase PR4-like	scaffold 8	4553763	4554816
LOC104415341	germin-like protein subfamily 1 member 13	scaffold 8	59940992	59942099
LOC104456220	endochitinase PR4-like	scaffold 8	4533169	4534322
LOC104426676	basic 7S globulin-like	scaffold 11	42720151	42721601
LOC104425852	mitogen-activated protein kinase kinase 9-like	scaffold 11	29992106	29993402
LOC104427666		scaffold 11	30401717	30402386
LOC104425880	miraculin-like	scaffold 11	30418678	30419730
LOC104431040	cinnamoyl-CoA reductase 1-like	scaffold 1321	1144	3482

3.3.4.5. SNP Calling results

We performed SNP calling using the two different pipelines (GATK and in-house script) and two references type (*E. grandis* reference genome and *de novo* transcriptomes).

We succeed to apply all the filters on SNPs originating from alignments on the *de novo* transcriptomes. Only a part of the filters has been applied on the SNPs originating from alignment on the *E. grandis* reference genome. Details of the numbers of raw SNPs called and SNP counts after each filtering steps are listed in Supplementary table 3.3.

While we cannot compare SNPs between these two type of reference, we analyzed the total number of variants separately. We displayed on the Figure 3.11 (A) the total counts of SNPs on each species according the alignments on *E. grandis* reference genome. At this stage of the analysis, we can see that the in-house methods generated more SNPs than the GATK recommended method (Haplotype Caller).

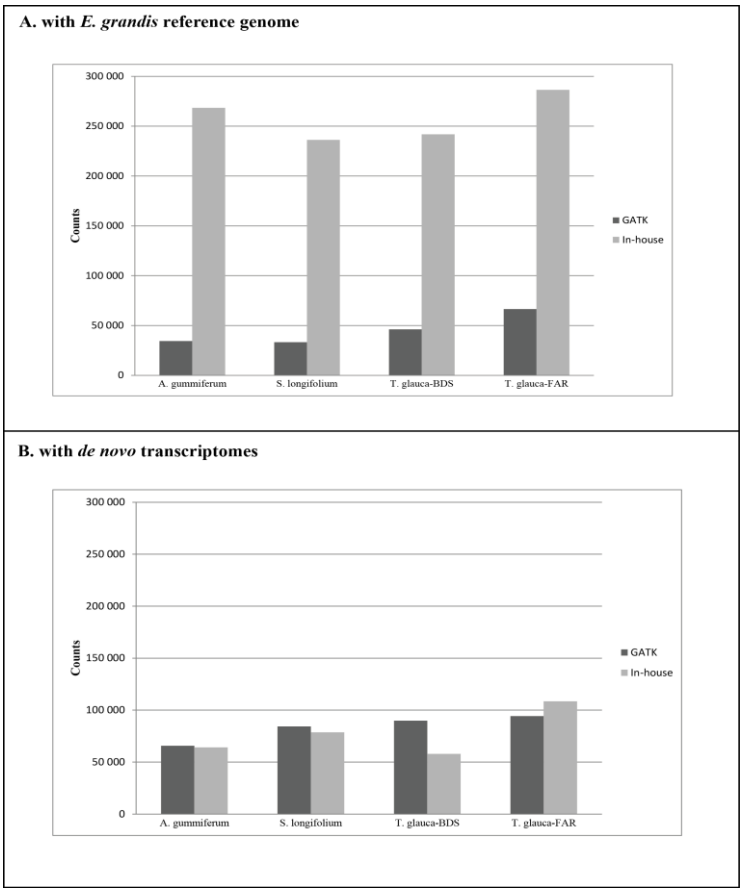


Figure 3.11: Numbers of SNPs per calling methods on species alignments with *E. grandis* reference genome (A) and with each *de novo* transcriptomes (B). *The proportion of SNP calls with GATK methods (black) and with In-house script (grey) are displayed*

We displayed on the Figure 3-11 (B) the total counts of SNPs on each species according the alignments on each *de novo* transcriptomes. After applying all the required filters we found similar counts of SNPs using GATK pipelines than in-house script in *A. gummiferum*, *S. longifolium* and *T. glauca*-BDS. And we only found more SNPs using the in-house script on the T. glauca-FAR.

We checked for the overlapped SNP positions obtained for each calling methods on *de novo* alignments (Table 3-7). We found around $^2/_3$ of the SNPs found with Unified Genotyper (GATK) and in-house methods sharing identical positions on *A. gummiferum*, *S. longifolium* and *T. glauca*-BDS. On *T. glauca*-FAR we found around 83% of shared SNPs positions between Unified Genotyper and in-house script.

Table 3.7: Overlapped and uniques SNPs called using two different calling methods (GATK and in-house script) on *de novo* transcriptomes.

Species	**Methods**	**filtered SNP counts**	**Unique SNP per method**	**Shared SNP positions between methods**
A. gummiferum	GATK (Unified Genotyper)	65,623	22,489	43,134
	In-house script	64,098	20,964	
S. longifolium	GATK (Unified Genotyper)	84,242	28,337	55,905
	In-house script	78,612	22,707	
T. glauca-BDS	GATK (Unified Genotyper)	89,791	34,549	55,242
	In-house script	57,835	2,593	
T. glauca-FAR	GATK (Unified Genotyper)	94,274	15,599	78,675
	In-house script	108,495	29,820	

3.3.4.6. Genome browser

We built two different JBrowses with the data from species *Syzygium longifolium*, taking first the *E. grandis* reference genome as a reference for alignment, then the de novo *S. longifolium* transcriptome. From now using the JBrowse, we can see at once for each individual, if a gene is differentially expressed, containing single nucleotide variations and the functional annotation of its gene products. We can also see the gene expression level (RPKM) for the samples.

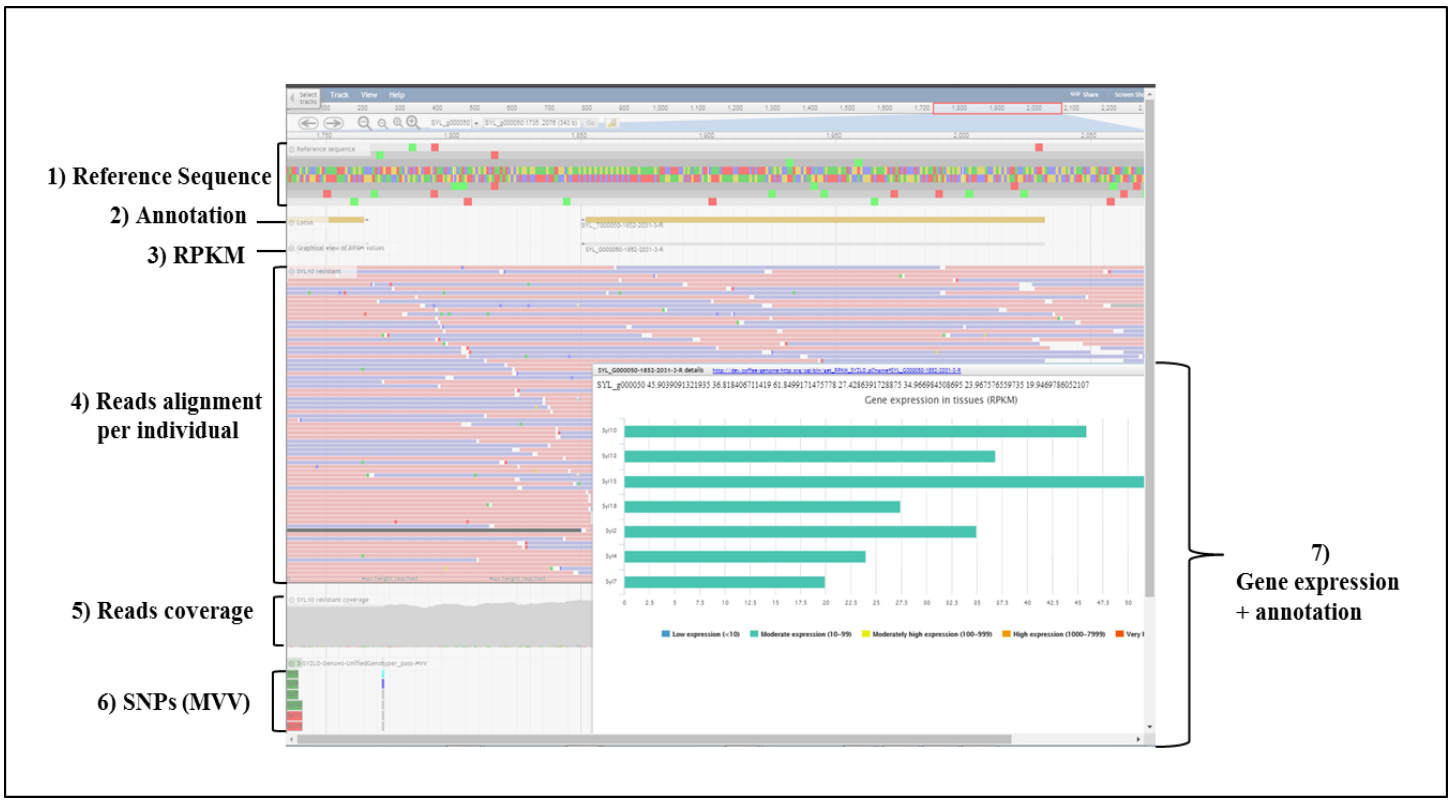


Figure 3.12: Screenshot of the JBrowse showing from top to bottom:
1) Reference Sequence track with each nucleotide base represented in different colors (S. longifolium- de novo transcriptome), 2) Annotation track, 3) RPKM track from edgeR, 4) Alignments track (light red for alignments on the forward strand and light blue for alignment on the reverse strand), 5) Reads coverage track, 6) SNPs per individual (cyan for heterozygotes, grey for homozygotes for the reference allele and deep blue for the homozygotes of the alternative allele) and MVV track (green for resistant individuals and red for susceptible) and 7) a popup with the expression of the gene displayed on the annotation track

3.3.5. Conclusion and perspectives

Transcriptome analysis is an important tool for characterization and understanding of the molecular basis of phenotypic variation in biology, including diseases (Soneson and Delorenzi, 2013). Although large genome size in non-domesticated forest tree renders complete genome sequencing unfeasible in many species, transcriptome sequencing represents a cost-effective alternative to obtain a large amount of genome-wide data.

In this study, we used both *de novo* transcriptome and a related species reference genome to perfom preliminary comparative analysis of RNA-Seq data originated from three Myrtaceae host species to the *A. psidii* rust pathogen. We have successfully obtained the first reference transcriptomes for *A. gummiferum*, *S. longifolium* and *T. glauca*, and therefore provide a valuable resource for comparative genomics, population genetics and evolutionary biology.
We evaluated the differential expression of genes on both types of alignment. We compared overlapped DE genes between species using the reference genome of *E. grandis*.
Differential expression shows that some of the genes overexpressed in susceptible individuals are in common among the three species studied. Overexpressed genes in resistant individuals have also been detected but are not common between species, we assume that rust-resistance mechanisms are species-specific. We also demonstrated here that large SNP panels can be discovered from both using *E. grandis* reference genome and *de novo* transcriptome.
We made four SNP call sets with two mapping references (*E. grandis* reference genome and *de novo* transcriptome) and two variants callers (GATK and in-house).
The RNAseq methodology based on several biological replicates allowed a genome wide analysis of the expression patterns that distinguish resistant and susceptible phenotypes among the three endemic species. Our results provide new valuable genetic resources for further investigations of myrtle rust interactions and resistance-related pathways.
The principal perspective will be to join the results obtained from differential expression and SNP calling obtained to identify congruent list of gene markers potentially involved in *A. psidii* resistance.

3.3.6. Acknowledgements

The authors acknowledge the Plateforme du Vivant (Nouméa, New Caledonia) and the Grand plateau technique regional de génotypage (UMR AGAP, CIRAD, Montpellier, France) for their facilities accesses. We thank the SouthGreen Bioinformatics Platform – UMR AGAP - CIRAD (http://southgreen.cirad.fr) for providing us with computational resources.
Funding for the work was provided by GLENCORE society, IAC and CIRAD.

3.3.7. References

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44, W3–W10.

Beenken, L. (2017). Austropuccinia: a new genus name for the myrtle rust Puccinia psidii placed within the redefined family Sphaerophragmiaceae (Pucciniales). Phytotaxa 297, 53.

Bouchet, P., Jaffre, T., and Veillon, J.-M. (1995). Plant extinction in New Caledonia: protection of sclerophyll forests urgently needed. Biodivers. Conserv. 4, 415–428.

Carnegie, A.J., Kathuria, A., Pegg, G.S., Entwistle, P., Nagel, M., and Giblin, F.R. (2016). Impact of the invasive rust Puccinia psidii (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biol. Invasions 1–18.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Dianese, J.C. (1984). Response of Eucalyptus species to field infection by Puccinia psidii. Plant Dis. 68, 314.

Giblin, F.R., and Carnegie, A.J. (2014). Puccinia psidii (Myrtle Rust) – Australian host list. Version current at 24 Sept. 2014. http://www.anpc.asn.au/resources/Myrtle_Rust.html.

Glen, M., Alfenas, A.C., Zauza, E. a. V., Wingfield, M.J., and Mohammed, C. (2007). Puccinia psidii: a threat to the Australian environment and economy —a review. Australas. Plant Pathol. 36, 1–16.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Harper, A.L. (2016). A Transcriptomics Approach to the Development of Predictive Molecular Markers for Tolerance to Tree Diseases: Ash Dieback. (Plant and Animal Genome), p.

Ibanez, T., Munzinger, J., Dagostini, G., Hequet, V., Rigault, F., Jaffré, T., and Birnbaum, P. (2014). Structural and floristic diversity of mixed tropical rain forest in New Caledonia: new data from the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN). Appl. Veg. Sci. 17, 386–397.

Junghans, D.T., Alfenas, A.C., Brommonschenkel, S.H., Oda, S., Mello, E.J., and Grattapaglia, D. (2003). Resistance to rust (Puccinia psidii Winter) in Eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers. Theor. Appl. Genet. 108, 175–180.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. 12, 996–1006.

Keszei, A., Brubaker, C.L., and Foley, W.J. (2008). A molecular perspective on terpene variation in Australian Myrtaceae. Aust. J. Bot. 56, 197.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360.

Laia, M.L., Alfenas, A.C., Brommonschenkel, S.H., Oda, S., de Melo, E.J., de Araújo Silva, I.M., Gonçalves, J.F., and Marques, A. (2015). Cloning and characterization of gene-resistant analogs (RGAs) involved in rust (Puccinia psidii) resistance in Eucalyptus grandis. J. For. Res. 26, 711–718.

L'Huillier, L., Jaffré, T., and Wulff, A. (2010). Mines et environnement en Nouvelle-Calédonie : les milieux sur substrats ultramafiques et leur restauration (Éditions IAC).

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 25, 2078–2079.

Mamani, E.M.C., Bueno, N.W., Faria, D.A., Guimarães, L.M.S., Lau, D., Alfenas, A.C., and Grattapaglia, D. (2010). Positioning of the major locus for Puccinia psidii rust resistance (Ppr1) on the Eucalyptus reference map and its validation across unrelated pedigrees. Tree Genet. Genomes 6, 953–962.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Morat, P., Jaffré, T., Tronchet, F., Munzinger, J., Pillon, Y., Veillon, J.-M., Chalopin, M., Birnbaum, P., Rigault, F., Dagostini, G., et al. (2012). Le référentiel taxonomique Florical et les caractéristiques de la flore vasculaire indigène de la Nouvelle-Calédonie. Adansonia 34, 179–221.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., et al. (2014). The genome of Eucalyptus grandis. Nature 510, 356–362.

Padovan, A., Keszei, A., Külheim, C., and Foley, W.J. (2014). The evolution of foliar terpene diversity in Myrtaceae. Phytochem. Rev. 13, 695–716.

Perry, G.L.W., and Enright, N.J. (2002). Humans, fire and landscape pattern: understanding a maquis-forest complex, Mont Do, New Caledonia, using a spatial "state-and-transition" model. J. Biogeogr. 29, 1143–1158.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11, 1650–1667.

Read, J., Jaffré, T., Godrie, E., Hope, G.S., and Veillon, J.-M. (2000). Structural and floristic characteristics of some monodominant and adjacent mixed rainforests in New Caledonia. J. Biogeogr. 27, 233–250.

Reynertson, K.A., Yang, H., Jiang, B., Basile, M.J., and Kennelly, E.J. (2008). Quantitative analysis of antiradical phenolic constituents from fourteen edible Myrtaceae fruits. Food Chem. 109, 883–890.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Salzman, R.A., Fujita, T., Zhu-Salzman, K., Hasegawa, P.M., and Bressan, R.A. (1999). An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. Plant Mol. Biol. Report. 17, 11–17.

Simpson, J.A., Thomas, K., and Grgurinovic, C.A. (2006). Uredinales species pathogenic on species of Myrtaceae. Australas. Plant Pathol. 35, 549–562.

Skinner, M.E., and Holmes, I.H. (2002). Setting Up the JBrowse Genome Browser. In Current Protocols in Bioinformatics, (John Wiley & Sons, Inc.), p.

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., and Holmes, I.H. (2009). JBrowse: A next-generation genome browser. Genome Res. 19, 1630–1638.

Snow, N. (2009). Kanakomyrtus (Myrtaceae): A New Endemic Genus from New Caledonia with Linear Stigma Lobes and Baccate Fruits. Syst. Bot.

Soewarto, J., Carriconde, F., Hugot, N., Bocs, S., Hamelin, C., and Maggia, L. (2017). The impact of Austropuccinia psidii in New Caledonia, a biodiversity hotspot. For. Pathol. in press.

Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14, 91.

Stalker, J., Gibbins, B., Meidl, P., Smith, J., Spooner, W., Hotz, H.-R., and Cox, A.V. (2004). The Ensembl Web Site: Mechanics of a Genome Browser. Genome Res. 14, 951–955.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., et al. (2002). The Generic Genome Browser: A Building Block for a Model Organism System Database. Genome Res. 12, 1599–1610.

Thornhill, A.H., Ho, S.Y.W., Külheim, C., and Crisp, M.D. (2015). Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Mol. Phylogenet. Evol. 93, 29–43.

Tommerup, I.C., Alfenas, A.C., Old, K.M., and others (2003). Guava rust in Brazil–a threat to Eucalyptus and other Myrtaceae. N. Z. J. For. Sci. 33, 420–428.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63.

Wilson, P.G. (2010). Myrtaceae. In Flowering Plants. Eudicots, (Springer, Berlin, Heidelberg), pp. 212–271.

Xavier, A.A., Alfenas, A.C., Matsuoka, K., and Hodges, C.S. (2001). Infection of resistant and susceptible Eucalyptus grandis genotypes by urediniospores of Puccinia psidii. Australas. Plant Pathol. 30, 277–281.

3.3.8. Supplementary data

Supplementary Table 3.1: Detailed of RNA extracts

Sample name	ID sampling	Biological material	Organism	BioAnalyzeur RNA concentration (ng/µl)	rRNA Ratio [25s/18s]	RIN
Sample 1	Ag19	leaf	*Arillastrum gummiferum*	281	3	8,2
Sample 2	Ag28	leaf	*Arillastrum gummiferum*	98	3,4	8,2
Sample 3	Ag2	leaf	*Arillastrum gummiferum*	353	2,3	6,9
Sample 4	Ag3	leaf	*Arillastrum gummiferum*	473	3,1	7,4
Sample 5	Ag4	leaf	*Arillastrum gummiferum*	531	2,9	7,4
Sample 6	Ag6	leaf	*Arillastrum gummiferum*	843	1,5	7,8
Sample 7	Syl10	leaf	*Syzygium longifolium*	414	2,3	7,7
Sample 8	Syl13	leaf	*Syzygium longifolium*	1248	3,4	7,9
Sample 9	Syl15	leaf	*Syzygium longifolium*	499	4,3	8
Sample 10	Syl18	leaf	*Syzygium longifolium*	602	1,8	7,5
Sample 11	Syl2	leaf	*Syzygium longifolium*	790	2,1	8,2
Sample 12	Syl4	leaf	*Syzygium longifolium*	667	1,8	8
Sample 13	Syl7	leaf	*Syzygium longifolium*	760	2,8	7,7
Sample 14	Tg2	leaf	*Tristaniopsis glauca* (FAR)	215	3,6	7,5
Sample 15	Tg3	leaf	*Tristaniopsis glauca* (FAR)	197	2,6	7,4
Sample 16	Tg4	leaf	*Tristaniopsis glauca* (FAR)	117	3,2	8,1
Sample 17	Tg5	leaf	*Tristaniopsis glauca* (FAR)	249	2,9	7,8
Sample 18	Tg6	leaf	*Tristaniopsis glauca* (FAR)	309	3,3	8
Sample 19	V1	leaf	*Tristaniopsis glauca* (BDS)	114	3,4	8,4
Sample 20	V2	leaf	*Tristaniopsis glauca* (BDS)	135	3,2	7,9
Sample 21	V3	leaf	*Tristaniopsis glauca* (BDS)	249	2,9	8,3
Sample 22	V4	leaf	*Tristaniopsis glauca* (BDS)	236	3,9	8,2
Sample 23	V6	leaf	*Tristaniopsis glauca* (BDS)	347	2,3	8,3
Sample 24	V7	leaf	*Tristaniopsis glauca* (BDS)	108	1,8	7,9

Supplementary table 3.2: Number of raw and cleaned reads. The individual used for *de novo* transcriptome assembly are shown in yellow.

			Rawreads		Cutadapt		Quality filter		Synchronized paired		
Species	libraries	pair	Number of read	Total reads	Number of read	Total reads	Number of read	Total reads	Number of read in paired sequences	Number of read in single sequence	Total reads
A.,gummiferum	Ag19	R1.fastq.gz	13,999,356	176,074,893	13,979,390	176,550,782	13,826,858	172,686,969	13,487,790	420,309	172,686,969
		R2.fastq.gz	13,999,356		13,963,517		13,569,031		13,487,790		
	Ag28	R1.fastq.gz	15,117,850		15,106,309		14,952,929		14,614,463	424,045	
		R2.fastq.gz	15,117,850		15,090,520		14,700,042		14,614,463		
	Ag2	R1.fastq.gz	16,737,435		16,720,235		16,077,560		15,717,141	756,610	
		R2.fastq.gz	16,737,435		16,703,178		16,113,332		15,717,141		
	Ag3	R1.fastq.gz	17,559,520		17,536,161		17,368,035		16,974,642	485,658	
		R2.fastq.gz	16,737,435		17,518,848		17,066,907		16,974,642		
	Ag4	R1.fastq.gz	11,841,969		11,832,007		11,715,154		11,447,310	330,994	
		R2.fastq.gz	11,841,969		11,823,208		11,510,460		11,447,310		
	Ag6	R1.fastq.gz	13,192,359		13,146,912		13,001,753		12,704,802	377,057	
		R2.fastq.gz	13,192,359		13,130,497		12,784,908		12,704,802		
S.,longifolium	Syl10	R1.fastq.gz	18,410,200	200,293,564	18,399,611	199,924,373	18,214,298	196,307,059	17,786,867	531,480	196,307,059
		R2.fastq.gz	18,410,200		18,380,383		17,890,916		17,786,867		
	Syl13	R1.fastq.gz	11,906,944		11,892,642		11,784,325		11,584,279	267,325	
		R2.fastq.gz	11,906,944		11,880,199		11,651,558		11,584,279		
	Syl15	R1.fastq.gz	12,185,308		12,169,050		12,038,801		11,790,027	321,993	
		R2.fastq.gz	12,185,308		12,155,770		11,863,246		11,790,027		
	Syl18	R1.fastq.gz	13,361,746		13,348,180		13,210,653		12,827,489	448,322	
		R2.fastq.gz	13,361,746		13,328,387		12,892,647		12,827,489		
	Syl2	R1.fastq.gz	15,528,002		15,511,653		15,357,301		15,028,873	417,797	
		R2.fastq.gz	15,528,002		15,496,078		15,118,242		15,028,873		
	Syl4	R1.fastq.gz	15,452,459		15,412,531		15,242,629		14,834,105	497,267	
		R2.fastq.gz	15,452,459		15,391,995		14,922,848		14,834,105		
	Syl7	R1.fastq.gz	13,302,123		13,286,296		13,152,135		12,889,198	341,199	
		R2.fastq.gz	13,302,123		13,271,598		12,967,460		12,889,198		

Supplementary table 3.2 continued from previous page

			Rawreads		Cutadapt		Quality filter		Synchronized paired		
Species	libraries	pair	Number of read	Total reads	Number of read	Total reads	Number of read	Total reads	Number of read in paired sequences	Number of read in single sequence	Total reads
T.glauca-FAR	Tg2	R1.fastq.gz	12,628,209	137,602,172	12,619,729	137,325,570	12,506,099	134,446,070	12,256,035	317,804	134,446,070
		R2.fastq.gz	12,628,209		12,607,318		12,323,775		12,256,035		
	Tg3	R1.fastq.gz	14,403,325		14,391,609		14,257,034		13,943,792	386,364	
		R2.fastq.gz	14,403,325		14,380,994		14,016,914		13,943,792		
	Tg4	R1.fastq.gz	16,207,008		16,174,551		15,618,141		15,124,929	784,505	
		R2.fastq.gz	16,207,008		16,135,866		15,416,222		15,124,929		
	Tg5	R1.fastq.gz	13,874,232		13,841,109		13,716,610		13,528,269	267,298	
		R2.fastq.gz	13,874,232		13,830,270		13,607,226		13,528,269		
	Tg6	R1.fastq.gz	11,688,312		11,676,931		11,570,381		11,352,254	279,541	
		R2.fastq.gz	11,688,312		11,667,193		11,413,668		11,352,254		
T.glauca-BDS	V1	R1.fastq.gz	8,858,469	154,194,966	8,817,505	153,676,220	8,741,346	150,931,462	8,643,046	147,634	150,931,462
		R2.fastq.gz	8,858,469		8,807,090		8,692,380		8,643,046		
	V2	R1.fastq.gz	13,962,685		13,942,810		13,817,346		13,589,262	301,974	
		R2.fastq.gz	13,962,685		13,931,957		13,663,152		13,589,262		
	V3	R1.fastq.gz	12,086,290		12,073,243		11,960,434		11,772,532	256,341	
		R2.fastq.gz	12,086,290		12,062,913		11,840,971		11,772,532		
	V4	R1.fastq.gz	11,998,827		11,936,661		11,528,882		11,209,706	526,788	
		R2.fastq.gz	11,998,827		11,870,792		11,417,318		11,209,706		
	V6	R1.fastq.gz	15,648,452		15,626,992		15,480,924		15,213,148	356,039	
		R2.fastq.gz	15,648,452		15,609,495		15,301,411		15,213,148		
	V7	R1.fastq.gz	14,542,760		14,511,762		14,367,592		14,033,652	419,994	
		R2.fastq.gz	14,542,760		14,485,000		14,119,706		14,033,652		

Supplementary table 3.3: Number of SNP following each filtering step.

	Reference type	Species	raw	prefilter	VCF filter	Monomorph filter	GATK filter
GATK (Haplotype Caller)	*E. grandis* genome	*A. gummiferum*	5,641,707		411,927	44,983	34,378
		S. longifolium	2,550,459		342,695	270,689	181,967
		T. glauca-BDS	3,076,802		280,610	58,222	46,155
		T. glauca-FAR	3,076,802		382,358	85,268	66,465
Inhouse script	*E. grandis* genome	*A. gummiferum*	10,768,345	2,627,198	825,906	396,279	268,221
		S. longifolium	8,860,083	2,246,428	723,591	351,999	236,197
		T. glauca-BDS	11,671,338	2,866,631	708,324	447,700	241,774
		T. glauca-FAR	11,671,338	2,866,631	877,238	523,495	286,384
GATK (Unified Genotyper)	de novo transcriptome	*A. gummiferum*	648,301		199,892	95,741	65,623
		S. longifolium	602,223		192,069	111,489	84,242
		T. glauca-BDS	935,693		257,152	122,951	89,791
		T. glauca-FAR	783,330		252,744	129,592	94,274
Inhouse script	de novo transcriptome	*A. gummiferum*	10,501,327	2,140,954	1,451,164	82,280	64,098
		S. longifolium	10,210,671	2,141,428	1,514,849	91,416	78,612
		T. glauca-BDS	9,436,592	2,100,408	1,409,764	68,608	57,835
		T. glauca-FAR	12,433,821	2,643,534	1,650,084	133,857	108,495

Chapitre IV : Association entre variation de l'ADN et gènes différentiellement exprimés en réponse à la rouille (*A. psidii*)

4.1. Introduction

Le phénotype résultant d'un stress biotique se définit par un ensemble de réactions moléculaires, modulées à la fois par les facteurs environnementaux et le fond génétique préexistant. Lors d'une interaction entre une plante hôte et un pathogène, un certain nombre de gènes sont impliqué, comme décrit dans le Chapitre 1. De nombreuses études ont utilisé le différentiel d'expression de gènes pour identifier les gènes spécifiquement induits par les individus résistants à la maladie. Les changements de niveau d'expression des gènes en réponse aux pathogènes, impliquent le plus souvent: le burst oxydatif, le transport moléculaire, les signaux de transduction, le développement et le maintien de la cellule, la biosynthèse d'acide jasmonique, éthylène, d'auxine et les métabolites secondaires (McGregor et al., 2009; Rao et al., 2013; Wang et al., 2017; Xue et al., 2015; Yang et al., 2009).

Par ailleurs, il a été montré qu'une grande partie des allèles des gènes, responsables du phénotype de résistance à un pathogène, ont un caractère dominant (Crute and Pink, 1996; Li et al., 2001). Les changements de bases nucléotidiques au niveau de ces gènes peuvent alors conduire à la résistance ou à la sensibilité à une maladie. Par exemple, chez l'espèce modèle *Arabidopsis thaliana*, une mutation dans le gène codant pour une pectate lyase confère la résistance à l'agent de l'oïdium *Golovinomyces cichoracearum* (Vogel et al., 2002). Chez la même espèce, une mutation ponctuelle d'un nucléotide dans un gène codant une protéine de résistance (TIR-NBS-LRR) abolit complètement la résistance à la bactérie *Pseudomonas syringae* et à l'oïdium *Peronospora parasitica* (Zhang, 2003).

L'utilisation de marqueurs polymorphes, tels que les SNPs, permet l'exploration des paysages génomiques individuels et l'identification des mutations liées à un phénotype particulier. L'avancée des technologies NGS permet la génération d'un grand jeu de données SNPs, mais paradoxalement l'établissement du lien entre les SNP et les phénotypes reste une étape critique dans l'analyse. La plupart des approches pour identifier des SNPs jouant un rôle « marqueur de phénotype », s'appuient sur des tests statistiques d'association (p-value ; F_{ST}[31]) avec les caractères phénotypiques mesurés (Saccone et al., 2010). Ces approches sont connu sous le terme de GWAS (Genome-wide association studdy) (Zeng et al., 2015). Ces tests statistiques requièrent à la fois l'existence d'un polymorphisme et un échantillonnage conséquent (Cantor et al., 2010).

[31] F_{ST} : aussi connu sous le nom d'indice de fixation, il reflète la variation dans la fréquence des allèles parmi des groupes d'une population.

Dans le Chapitre 3 de cette thèse, nous avions identifié des SNPs chez trois espèces de Myrtaceae hôtes à la maladie de la rouille (*Austropuccinia psidii*) reflétant la variation de chaque individu par rapport à une référence donnée (le génome de référence de l'*E. grandis* ou les transcriptome *de novo*). Nous nous sommes appuyés sur un plan expérimental contenant un faible nombre de répétitions biologiques (n < 8) exprimant des phénotypes contrastés face à la maladie (résistant *vs* sensibles). C'est pourquoi il est proposé, dans le cadre de cette étude, une autre approche cantonnée à un pool restreint de SNPs par espèce. Ces SNPs seront choisis pour leur capacité à caractériser chacun de manière formelle le phénotype résistant, ou sensible, sur l'ensemble des échantillons étudiés. Ce qui n'empêchera pas, à l'avenir, en augmentant le nombre d'échantillons par espèce, de conduire les tests statistiques habituels en élargissant alors l'analyse à l'ensemble des SNPs identifiés.

Dans ce quatrième Chapitre de thèse, nous testerons une nouvelle approche pour identifier des gènes candidats de résistance à la rouille, par une analyse conjointe du différentiel d'expression des gènes et des génotypes SNPs. Il s'agira ici de consolider les SNPs identifiés dans Chapitre 3 via leur confrontation avec des régions génomiques reconnues comme admettant des expressions différentiées aux regards des deux types de phénotypes étudiés, supposées donc jouer un rôle dans la réponse au pathogène. Cette confrontation se fera par espèce entre différentes espèces, via la comparaison de gènes orthologues entre *A. gummiferum, S. longifolium, T. glauca et E. grandis*.

4.2. Materiel et methodes

La comparaison des gènes entre espèces fait appel à la notion d'homologie, d'orthologie et paralogie. L'homologie désigne une origine ancestrale commune. En génétique, on définit par le terme « orthologues » les gènes possédant une origine commune (homologie) mais ayant été séparés par un évènement de spéciation (Kuzniar et al., 2008 ; Sonnhammer and Koonin, 2002). De même on définit par le terme « paralogues » des gènes homologues ayant été séparés par un évènement de duplication au sein d'une même espèce (Fitch, 2000 ; Sonnhammer and Koonin, 2002). L'identification des gènes homologues entre deux espèces se base sur les algorithmes d'alignements de séquences (nucléotidiques ou peptidiques) tel que le BLAST (Altschul et al., 1990). L'une des méthodes les plus couramment utilisé pour la recherche d'homologie entre espèces est le BBMH (Best Blast Mutual Hits) que nous utilisons dans ce travail.

4.2.1. Best Blast Mutual Hits

L'identification de groupes de gènes homologues est particulièrement utile pour la génomique comparative, l'annotation des génomes, la compréhension de l'évolution des gènes/protéines (Li et al., 2003). Les séquences de gènes sont comparées entre deux espèces et les meilleurs alignements réciproques constituent des gènes ortologues (Zheng et al., 2005).

Nous avons ici recherché des gènes orthologues putatifs entre chacune des trois espèces de Myrtaceae, et avec l'espèce de référence *Eucalyptus grandis*, en utilisant la stratégie BBMH. On considère dans cette étude qu'une similitude significative (avec un pourcentage d'identité supérieur à 90%) est signe d'homologie.

Afin de lancer la procédure de BBMH, nous utilisons les fichiers d'annotations fonctionnelles des transcriptomes *de novo* des trois espèces de Myrtaceae étudiées et issues du programme FrameDP (Chapitre 3). Chaque transcriptome a été traité individuellement avec le programme CD-HIT (Fu et al., 2012) afin d'identifier les clusters de gènes paralogues. Nous avons considéré qu'un recouvrement de séquence d'au moins 70% chez une même espèce formera un cluster de gènes paralogues. Le programme BBMH (blast_cluster.pl) a été lancé en ligne de commande deux-à-deux entre les annotations fonctionnelles des espèces de Myrtaceae étudiées (par exemple *A. gummiferum-T. glauca* ; puis *T-glauca-S. longifolium*, puis *S. longifolium-A. gummiferum*) et avec l'annotation du génome de référence de l'*E. grandis* (par exemple *A. gummiferum – E. grandis*, puis *S. longifolium – E. grandis*, et *T. glauca - E. grandis*). Les paramètres d'analyses du BBMH comprennent une e-valeur de 1e-10. L'ensemble des résultats des différentes combinaisons de BBMH a été regroupé dans un seul et même tableau à l'aide d'un script personnalisé (map_myrtaceae4_locus.pl).

4.2.2. Choix des SNPs discriminants entre phénotypes résistants et sensibles à la rouille

Nous utilisons ici les résultats issus des différentes méthodes SNP calling testées dans le Chapitre 3 (méthode GATK : Unified Genotyper et méthode personnalisée : In-house) sur la base des références transcriptomes *de novo* de chaque espèce de Myrtaceae étudiée. Afin de maximiser les chances de trouver des SNPs communs entre espèces, nous avons choisis de travailler sur les SNPs ayant passés l'ensemble des filtres de qualité décrit dans le Chapitre 3, à l'exception du filtre GATK sensé déjà retirer les SNPs regroupés en clusters.

Pour chaque espèce, nous avons sélectionné les SNPs discriminants sur la base du génotype de la référence utilisée tel que décrit dans le Tableau 4.1. L'ensemble des combinaisons

d'allèles spécifiques à l'un, ou l'autre, des phénotypes de réponse à la rouille est considéré comme discriminant.

Tableau 4.1 : Sélection des SNP discriminants sur la base du génotype de la référence

	Individu de référence	Individu 1	Individu 2	
Phénotype	Résistant	Résistant	Sensible	
Génotype SNP position 1	G/G	G/G	G/A	SNP discriminant
Génotype SNP position 2	T/T	T/C	T/C	SNP non-discriminant

4.2.3 Compilation des résultats du SNP calling, du différentiel d'expression et du BBMH

L'ensemble des résultats générés comprenant, le tableau général des BBMH, les différentes listes de SNPs discriminants et le différentiel d'expression, ont été regroupés en utilisant un programme personnalisé (map_myrtaceae4_locus.pl). Les données utilisées pour le différentiel d'expression sont celles issues du Chapitre 3 pour les deux méthodes d'alignement des individus, avec le génome de référence de l'*E. grandis* et avec les transcriptomes *de novo*. Nous avons extrait de cette compilation les différents résultats qui suivent afin de répondre aux différentes interrogations posées dans ce chapitre.

4.3. Resultats et discussion

Nous traitons ici des résultats obtenus en comparant les espèces *A. gummiferum, S. longifolium* et *T. glauca* (BDS) pour lesquelles nous avons pu générer les différents fichiers BBMH. L'espèce *T. glauca* (FAR) ne sera donc pas traitée dans le cadre de ce mémoire.

4.3.1. Sélection des SNPs discriminants résistants/sensibles à *A. psidii* pour chaque espèce.

Dans cette étude, nous avons effectué SNP calling à l'aide de deux méthodes : la méthode recommandée par GATK (Unified Genotyper) et une méthode mise au point par le CIRAD (in-house script de Martin et al., *in prep.*, Baurens et al., *in prep*).

Nous avons filtré les SNPs de chaque individu alignés sur la référence transcriptome *de novo* pour chaque espèce afin d'identifier les marqueurs discriminants entre les phénotypes résistant et sensibles (Tableau 4.2). En moyenne, le nombre total de SNPs discriminants identifiés pour chaque espèce est de 2%. Sur l'espèce *A. gummiferum*, nous avons trouvé 1001 SNPs discriminants avec la méthode In-house et 1224 SNPs discriminants avec la méthode Unified Genotyper ; ce qui représente respectivement 1,2 et 1,3% du nombre total de SNPs trouvés pour chaque méthode. Sur l'espèce *S. longifolium*, nous avons trouvé 2027 SNPs discriminants avec la méthode In-house et 2536 SNPs discriminants avec la méthode Unified Genotyper ; ce qui représente respectivement 2,6 et 3 % du nombre total de SNPs trouvés pour chaque méthode. Sur l'espèce *T. glauca*, nous avons trouvé 1475 SNPs discriminants avec la méthode In-house et 2449 SNPs discriminants avec la méthode Unified Genotyper ; ce qui représente approximativement 2 % du nombre total de SNPs trouvés pour chaque méthode.

Tableau 4.2 : Nombre de SNPs discriminants par espèce sur les alignements avec les transcriptomes *de novo*

			SNPs discriminants	
Espèces	Méthode de SNP calling	Nombre de SNP filtrés	Nombre total	Proportion
Arillastrum gummiferum	In-house script	82 280	1001	1,2
	Unified Genotyper	95 741	1224	1,3
Syzygium longifolium	In-house script	79 486	2027	2,6
	Unified Genotyper	84 242	2536	3,0
Tristaniopsis glauca-BDS	In-house script	68 608	1475	2,1
	Unified Genotyper	122 951	2449	2,0

4.3.2. Compilation des SNPs discriminants et les gènes différentiellement exprimés identifiés via l'alignement effectué à partir du génome de l'*E. grandis*.

Dans le Chapitre 3, nous avons démontré l'existence de 31 gènes communs surexprimés chez les individus sensibles des trois espèces de Myrtaceae étudiées. Pour rappel, ces 31 gènes sont issus de l'étude du différentiel d'expression menés à partir des alignements des reads des trois espèces de Myrtaceae sur le génome de l'*E. grandis*. Le croisement de ces 31 gènes avec les

SNPs discriminants définis pour les trois espèces s'est révélé nul. Néanmoins, il est possible que des variations nucléotides puissent exister au niveau des promoteurs de ces gènes et pourraient ainsi expliquer le différentiel d'expression observé. En effet, il est reconnu que les variations dans les régions promotrices des gènes peuvent affecter l'expression des gènes et le métabolisme (Smirnova and Kochetov, 2015). Par exemple, la comparaison des séquences du gène *Xa23* entre deux variétés de riz (O. *sativa* ssp. *indica*) a montré que le caractère de résistance ou de sensibilité à la bactérie *Xanthomonas oryzae* était modulé par des variations nucléotidiques au niveau du promoteur de ce gène (Cui et al., 2017). Il est aussi connu que chez les plantes, l'activation des mécanismes de résistance peut être induite par la fixation de la protéine d'avirulence du pathogène sur la région promotrice des gènes de résistance (Römer et al., 2009 ; Smirnova and Kochetov, 2015).

4.3.3. Sélection des SNPs discriminants communs entre les trois espèces

La confrontation des listes des SNPs discriminants des trois espèces étudiées est résumée dans le Tableau 4.3 Cette confrontation interspécifique nous a permis d'identifier 9 SNPs communs avec la méthode Unified Genotyper.

Parmi eux, 7 SNPs sont situés dans des gènes exprimant des fonctions connues dans les mécanismes d'interaction hôte-pathogène.

Par exemple, un SNP est situé dans le gène codants pour une « L-type lectin-domain containing receptor kinase IV (LecRK) » chez les trois espèces de Myrtaceae (ARIGU_g779590 ; SYL_g613210 ; TRIGL_g199130). Le rôle des gènes LecRK dans la résistance à *Phytophthora infestans* (oomycète) a été démontré chez *Arabidopsis thaliana* (Bouwmeester et al., 2011).

Un SNP est situé dans un gène codant pour un « Receptor-like protein kinase FERONIA » (ARIGU_g410430 ; SYL_g608330 ; TRIGL_g529070) dont le rôle est connu pour être impliqué dans les voies de signalisation du Ca^{2+}, intervenant notamment dans la réponse immunitaire contre les agents pathogène (Li et al., 2016).

Un SNP est situé dans un gène codant une « Phosphatidylinositol-3-phosphatase myotubularin-1 » (ARIGU_g748290 ; SYL_g320690 ; TRIGL_g086400). Cette protéine possède un rôle dans l'induction des ROS (espèce réactive à l'oxygène) par la voie de l'acide abscissique et est responsable de la fermeture des stomates en réponse aux stress biotiques et abiotiques (Park et al., 2003).

Plusieurs SNPs discriminants sont également situés dans trois gènes codant des « TMV resistance protein N-like ». Les « TMV resistance protein » des protéines de résistance de la famille des TIR-NBS-LRR conférant la résistance au virus de la mosaïque du tabac (Tobacco Mosaic Virus) (Dinesh-Kumar et al., 2000).
Enfin, un SNP est situé dans un gène codant pour une protéine putative LRR receptor-like serine/threonine-protein kinase dont le rôle est connue dans les réactions de défense des plantes contre les pathogènes (Afzal et al., 2008).

Tableau 4.3 : Récapitulatif des SNPs discriminants communs entre *A. gummiferum*, *S. longifolium* et *T. glauca*

Nombre de SNP commun entre les trois espèces

Espèces	Méthode de SNP calling	Par méthode	Commun aux deux méthodes
A. gummiferum / *S. longifolium* / *T. glauca* (BDS)	Unified genotyper	9	1
	In-house script	2	

Avec la méthode de SNP-calling In-house script, seuls 2 SNPs discriminants communs aux trois espèces ont été identifiés, dont un est commun avec la méthode GATK (Tableau 4.3, 4.4 et 4.5). Ce SNP est situé dans un gène codant respectivement pour une protéine MOS2 chez l'*E. grandis*, une protéine photosystème II CP74 hypothétique chez *A. gummiferum*, une protéine TMV resistance N-like chez *S. longifolium* et une protéine hypothétique MOS2 chez *T. glauca*-BDS. Les fonctions et positions de chacun de ces SNPs sont résumées dans les Tableaux 4.3 et 4.4.

Tableau 4.4 : Annotation fonctionnelle des 9 SNPs communs entre les trois espèces et identifiés avec la méthode Unified Genotyper.

(Le locus surligné en jaune est commun avec la méthode In-house)

Annotation *E. grandis*		***A. gummiferum***			***S. longifolium***			***T. glauca(BDS)***		
Eg_LOC	Eg_product	Ag_locus_tag	Ag_product	SNP-Position	Sl_locus_tag	Sl_product	SNP-Position	Tg_locus_tag	Tg_product	SNP-Position
LOC104418125	beta-amyrin synthase	ARIGU_g793950:536:2626:2:+	beta-amyrin synthase [Eucalyptus grandis]	516	SYL_g613950:187:2472:1:+	Beta-amyrin synthase	1643	TRIGL_g066660:1:2491:2:+	beta-amyrin synthase [Eucalyptus grandis]	2731
LOC104418518	L-type lectin-domain containing receptor kinase IV.1	ARIGU_g779590:403:2406:1:+	L-type lectin-domain containing receptor kinase IV.1-like [Eucalyptus grandis]	816	SYL_g613210:224:2233:2:+	L-type lectin-domain containing receptor kinase IV.1-like [Eucalyptus grandis]	176	TRIGL_g199130:235:2226:1:+	L-type lectin-domain containing receptor kinase IV.1	2150
LOC104421609	receptor-like protein kinase FERONIA	ARIGU_g410430:191:2911:2:+	Receptor-like protein kinase FERONIA	1679	SYL_g608330:162:2876:3:+	Receptor-like protein kinase FERONIA,Conserved hypothetical protein	1851	TRIGL_g529070:132:1860:3:+	receptor-like protein kinase FERONIA [Eucalyptus grandis]	1442
LOC104421745	phosphatidylinositol-3-phosphatase myotubularin-1	ARIGU_g748290:1:2658:1:+	phosphatidylinositol-3-phosphatase myotubularin-1 [Eucalyptus grandis]	969	SYL_g320690:381:2087:3:+	Phosphatidylinositol-3-phosphatase myotubularin-1	605	TRIGL_g086400:70:2631:1:+	phosphatidylinositol-3-phosphatase myotubularin-1 [Eucalyptus grandis]	1169
LOC104443066	TMV resistance protein N-like	ARIGU_g002450:173:3640:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1990	SYL_g615500:374:2320:2:+	leucine-rich repeat protein soc-2-like [Eucalyptus grandis],TMV resistance protein N-like [Eucalyptus grandis]	759	TRIGL_g531360:45:1850:3:+	TMV resistance protein N-like [Eucalyptus grandis],transforming growth factor-beta receptor-associated protein 1 homolog [Eucalyptus grandis]	1452
LOC104444214	protein MOS2	ARIGU_g076200:358:1746:1:+	Conserved hypothetical protein,Photosystem II CP47 reaction center protein	2487	SYL_g113150:72:1481:3:+	TMV resistance protein N-like [Eucalyptus grandis],Peptide chain release factor PrfB2, chloroplastic	810	TRIGL_g132870:178:1551:1:+	Putative Protein Ycf2 B,protein MOS2 [Eucalyptus grandis]	1773
LOC104445424	unknown_product	ARIGU_g005350:1:1707:1:+	TMV resistance protein N-like isoform X3 [Eucalyptus grandis],leucine-rich repeat protein soc-2-like [Eucalyptus grandis]	3110	SYL_g616660:1150:4383:1:+	TMV resistance protein N-like isoform X2 [Eucalyptus grandis],TMV resistance protein N-like [Eucalyptus grandis]	4716	TRIGL_g532500:213:3878:3:+	TMV resistance protein N-like isoform X2 [Eucalyptus grandis],BAHD acyltransferase At5g47980-like [Eucalyptus grandis]	1321

Suite du Tableau 4.4 : Annotation fonctionnelle des 9 SNPs communs entre les trois espèces et identifiés avec la méthode Unified Genotyper. (Le locus surligné en jaune est commun avec la méthode In-house

Annotation *E. grandis*		***A. gummiferum***			***S. longifolium***			***T. glauca (BDS)***		
Eg_LOC	Eg_product	Ag_locus_tag	Ag_product	SNP-Position	Sl_locus_tag	Sl_product	SNP-Position	Tg_locus_tag	Tg_product	SNP-Position
LOC104448576	probable LRR receptor-like serine/threonine-protein kinase At1g53430	ARIGU_g317660:248:3283:2:+	Putative LRR receptor-like serine/threonine-protein kinase At1g53440	181	SYL_g613230:1:2857:2:+	Putative LRR receptor-like serine/threonine-protein kinase At1g53440	194	TRIGL_g530450:1:3085:2:+	Putative LRR receptor-like serine/threonine-protein kinase At1g53430	100
LOC104452833	uncharacterized LOC104452833	ARIGU_g412560:1:1255:1:+	Conserved hypothetical protein	1128	SYL_g614030:1322:3141:2:+	Conserved hypothetical protein	185	TRIGL_g332820:9:1152:3:+	Conserved hypothetical protein	421

Tableau 4.5 : Annotation fonctionnelle des 4 SNPs communs entre les trois espèces et identifiés avec la méthode in-house. (Le locus surligné en jaune est commun avec la méthode GATK)

Annotation *E. grandis*		***A. gummiferum***			***S. longifolium***			***T. glauca* (BDS)**		
Eg_LOC	Eg_product	Ag_locus_tag	Ag_product	SNP-Position	Sl_locus_tag	Sl_product	SNP-Position	Tg_locus_tag	Tg_product	SNP-Position
LOC104428150	TMV resistance protein N-like	ARIGU_g220620:1:3706:2:+	TMV resistance protein N-like [Eucalyptus grandis],TMV resistance protein N-like [Eucalyptus grandis]	3936	SYL_g067400:113:2219:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1567	TRIGL_g195490:87:3644:3:+	TMV resistance protein N-like [Eucalyptus grandis]	862
LOC104444214	protein MOS2	ARIGU_g076200:358:1746:1:+	Conserved hypothetical protein,Photosystem II CP47 reaction center protein	2487	SYL_g113150:72:1481:3:+	TMV resistance protein N-like [Eucalyptus grandis],Peptide chain release factor PrfB2, chloroplastic	810	TRIGL_g132870:178:1551:1:+	Putative Protein Ycf2 B,protein MOS2 [Eucalyptus grandis]	1773

4.3.4. Sélection des gènes différentiellement exprimés communs aux trois espèces via les alignements effectués sur les transcriptomes *de novo*

Nous avons comparé les gènes différentiellements exprimés à partir des alignements sur les transcriptomes *de novo* de chaque espèce (Tableau 4.5). Cette comparaison a permis d'identifier 5 gènes surexprimés chez les individus sensibles des trois espèces. Parmi ces 5 gènes, un en particulier code pour une « Thaumatin-like protein (TLP) » chez *A. gummiferum* et une « *Pathogenesis-related protein R major form* » chez *T. glauca*-BDS (Annexe 4.1). Les TLP sont reconnus pour leur rôle dans les mécanismes de défense des plantes contre les pathogènes (Liu et al., 2010). Comparativement, nous avions trouvé 66 gènes surexprimés chez les sensibles entre ces trois mêmes espèces dans le Chapitre 3, et 31 gènes surexprimés lorsque les individus de *T. glauca* –FAR étaient ajoutés à l'analyse. Nous supposons alors que l'alignement contre une espèce différente, bien qu'étant informatif au vu de la qualité de l'annotation du génome de l'*E. grandis*, engendre des faux-positifs au niveau de l'analyse du différentiel d'expression.

En comparant les espèces deux-à-deux, nous identifions : 17 gènes surexprimés chez les sensibles entre *A. gummiferum* et *T. glauca* (BDS) ; 23 gènes surexprimés chez les sensibles entre *S. longifolium* et *T. glauca* (BDS) ; 135 gènes différentiellement exprimés (129 surexprimés et 6 sous-exprimés chez les sensibles) entre *A. gummiferum* et *S. longifolium*. Contrairement à ce que l'on aurait pu s'attendre de par les liens phylogénétiques qui unissent ces espèces (Thornhill *et al.*, 2015), rapprochant plus *A. gummiferum* de *T. glauca* par rapport à *S. longifolium*, les résultats obtenus suggèrent une hiérarchie différente dans les liens d'affinités entre ces trois espèces. Il apparaît en effet à la lecture du Tableau 4.6 qu'*A. gummiferum* partage plus de gènes DE avec *S. longifolium* isolant fortement *T. glauca* d'eux. Les traits de résistance et/ou de sensibilité à la rouille, de fait, semblent dissociés de la typologie du phylum caractérisant le groupe de ces espèces. Il est également possible que ce résultat soit dû à des biais lors de l'extraction de l'ARN des feuilles de *Tristaniopsis*. En effet, bien que de bonnes qualités, les concentrations en extraits d'ARN obtenus sur l'espèce *T. glauca* sont inférieures à celles des deux autres espèces.

Tableau 4.6 : Comparaisons des gènes différentiellements exprimés à partir des alignements sur les transcriptomes *de novo* et communs aux trois espèces étudiées.

	Nombre de gènes DE communs		
Espèce	Total	Surexprimés*	Sous-exprimés*
A. gummiferum - *T. glauca* (BDS) - *S. longifolium*	5	5	0
A. gummiferum - *T. glauca* (BDS)	17	17	0
S. longifolium - *T. glauca* (BDS)	23	23	0
A. gummiferum- *S. longifolium*	135	129	6

**sur- et sous-exprimés chez les individus sensibles par rapport aux individus résistants à la rouille.*

4.3.5. Compilation des SNPs discriminants et des gènes différentiellement exprimés à partir des alignements sur transcriptome *de novo* pour chaque espèce

Nous avons confronté la liste des SNPs discriminants (résistants/sensibles) avec celles des gènes différentiellement exprimés, ou non, espèce par espèce. Pour la méthode Unified Genotyper, les résultats acquis (Tableau 4.7) mettent en évidence 6 loci différentiellement exprimés admettant un total de 29 SNPs identifiés pour l'espèce *A. gummiferum*, en plus des 1195 SNPs identifiés sur 400 locus non différentiellement exprimés. De même que 22 loci différentiellement exprimés ont été identifiés chez *S. longifolium* avec 63 SNPs, en plus des 2473 SNPs identifiés sur 1480 locus non différentiellement exprimés, et enfin 1 seul locus différentiellement exprimé à 1 SNP chez *T. glauca*, en plus des 2448 SNPs identifiés sur 1387 locus non différentiellement exprimés.

Tableau 4.7 : Liste des gènes différentiellement exprimés communs aux 3 espèces trouvés sur *de novo* avec les SNP disciriminants (méthode Unified Genotyper)

	Nombre de SNP situés dans des gènes différentiellement exprimés						Nombre de SNP-non différentiellement exprimés		
Espèce	Total SNP	Sous-exprimés	Surexprimés	Total de locus	Sous-exprimés	Surexprimés	Total SNP	Total de locus	TOTAL SNP discriminants
A. gummiferum	29	12	17	12	3	3	1195	400	1224
S. longifolium	63	3	60	22	3	19	2473	1480	2536
T. glauca (BDS)	1	0	1	1	0	1	2448	1387	2449

Pour la méthode In-house, un total de 2 loci différentiellement exprimés, contenant 3 SNPs ont été trouvé sur l'espèce *A. gummiferum*, en plus des 998 SNPs identifiés sur 572 locus non différentiellement exprimés. De même que 19 loci différentiellement exprimés admettant un total de 63 SNPs ont été identifiés chez *S. longifolium*, en plus des 1959 SNPs identifiés sur 1196 locus non différentiellement exprimés. Aucun locus différentiellement exprimé et contenant un SNP discriminants n'a été trouvé pour *T. glauca* avec cette méthode (Tableau 4.8).

Les annotations fonctionnelles de ces loci différentiellement exprimés contenant des SNP discriminants sont fournies dans les Annexes 4.2, 4.3, 4.4, 4.5 et 4.6 pour chaque espèce.

Tableau 4.8 : Liste des gènes DE communs aux 3 espèces trouvés sur denovo avec les SNP disciriminant (méthode In-house)

	Nombre de SNP situés dans des gènes différentiellement exprimés						Nombre de SNP-non différentiellement exprimés		
Espèce	Total SNP	Sous-exprimés	Surexprimés	Total de locus	Sous-exprimés	Surexprimés	Total SNP	Total de locus	TOTAL SNP discriminants
A. gummiferum	**3**	0	3	2	0	2	998	572	1001
S. longifolium	68	3	65	19	3	16	1959	1196	2027
T. glauca (BDS)	0	0	0	0	0	0	1475	834	1475

Nous avons croisé pour chaque espèce, les positions des gènes différentiellement exprimés et contenant des SNP discriminant obtenus à la fois avec la méthode Unified Genotyper et In-house (Tableau 4.9). Il en ressort que 3 SNPs sont communs aux deux méthodes pour l'espèce *A. gummiferum* sur 2 gènes DE (Annexes 2 et 3). De même, nous trouvons 49 SNPs communs pour *S. longifolium* sur 19 gènes DE (Annexe 4 et 5). Parmi ces 19 gènes, 5 gènes possèdent des fonctions liés à la résistance : SYL_g617630 « TMV resistance protein N-like » (Christie et al., 2016; Dinesh-Kumar et al., 2000), SYL_g103460 (pleiotropic drug resistance protein 2-like) (Stukkens et al., 2005), SYL_g283480 (MLO-like protein 12) (Piffanelli et al., 2002), SYL_g361300 (Wall-associated receptor kinase-like 14) (Kohorn and Kohorn, 2012) et SYL_g312850 (Dirigent protein 2) (Li et al., 2017).

Pour l'espèce *T. glauca* aucun SNP discriminant en commun ne peut être trouvé entre les deux méthodes. Cependant nous remarquons que 1154 SNPs discriminants sont partagés entre les deux méthodes pour cette espèce au niveau des gènes non différentiellement exprimés.

Tableau 4.9 : Croisement par espèce des SNPs obtenus avec les méthodes Unified Genotyper et In-house. DE = Différentiellement exprimé.

	SNPs discriminants situés dans des gènes DE et commun entre les 2 méthodes (GATK/Inhouse)			SNPs discriminant communs aux 2 méthodes (GATK/inhouse) situés dehors des gènes DE
Espèces	Surexprimé	Sous-exprimé	TOTAL	TOTAL
A. gummiferum	3	0	3	636
S. longifolium	3	46	49	1570
T. glauca (BDS)	0	0	0	1154

4.3.6. Illustration d'un SNP discriminant dans un gène différentiellement exprimé du JBrowse S. longifolium.

Le navigateur web du transcriptome *de novo* de *S. longifolium* permet de visualiser le contig SYL_g255820 qui contient une région codant une Tetraketide alpha-pyrone reductase (Figure 4.1). On observe que certains SNPs discriminent les individus sensibles des individus résistants : sur cet exemple les individus sensibles (Syl 13 et Syl 15 en rouge) sont hétérozygotes (SNP cyan) tandis que les individus résistants sont homozygotes (SNP gris) à la référence (Syl10). D'autre part ce gène est sous exprimé chez les individus sensibles par rapport aux résistants à l'exception de Syl10. Le JBrowse de *S. longifolium* est pour l'instant herbégé sur le serveur du CIRAD. Il devrait être configuré très prochainement afin de permettre l'accès aux autres utilisateurs extérieurs.

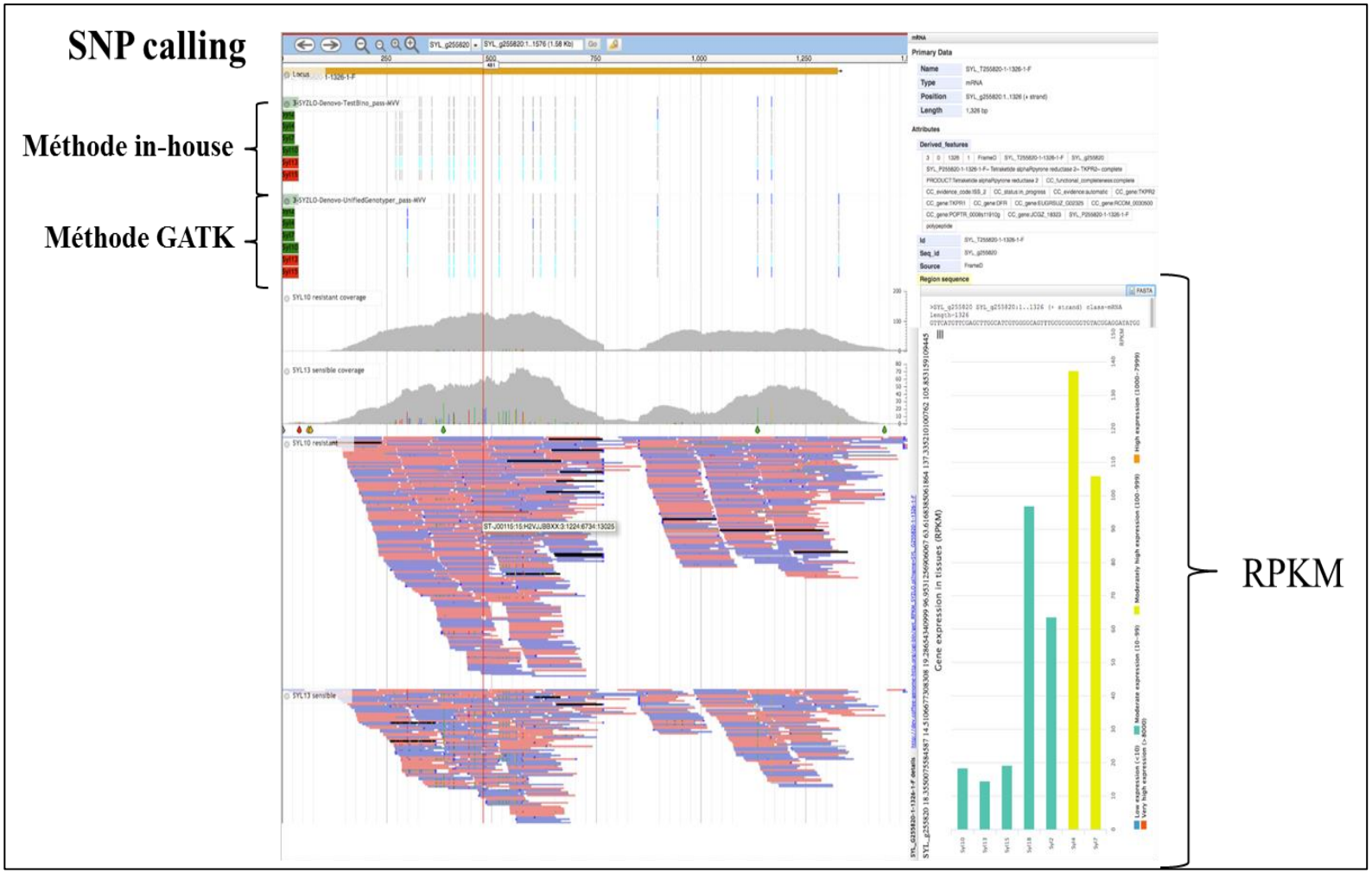


Figure 4.12 : Illustration d'un SNP discriminant localisé dans le gène SYL_g255820 différentiellement exprimé chez *S. longifolium* à partir de l'alignement sur référence *de novo*.

Les individus résistants sont représentés en vert et les sensibles sont en rouge (en haut à gauche pour le SNP calling).

Les niveaux de RPKM sont illustrés en plusieurs couleurs : jaune pour les hauts niveaux de RPKM et vert pour des valeurs moyennes de RPKM.

4.4. Conclusion

La recherche de gènes marqueurs de la résistance à la rouille des Myrtaceae (*Austropuccinia psidii*) est devenue une prérogative de recherche pour la préservation des espèces menacées par le pathogène. Jusqu'à présent, l'effort de recherche s'est surtout concentré sur les espèces d'*Eucalyptus,* et ont démontré que les réponses de résistance à la rouille sont médiées par des gènes de résistance (R) (Mamani et al., 2010 ; Thumma et al., 2013 ; Tobias et al., 2015).
Dans leur étude, (Tobias et al., 2015) émettent l'hypothèse de la persistance de certains gènes de résistance ancestraux conservés chez les Myrtaceae modernes et liés à l'historique des interactions entre une espèce apparentée et *Austropuccinia psidii* en Amérique du Sud. Compte tenu du fait que les gènes de résistance sont conservés entre les génomes des espèces proches, nous nous attendions dans cette étude à identifier des marqueurs discriminants pour le phénotype de réponse à la rouille commun à plusieurs espèces de Myrtaceae. Or, l'analyse du différentiel d'expression et la recherche de SNPs rendent compte de la complexité des réponses induites lors de l'interaction entre *A. psidii* et les Myrtaceae néo-calédoniennes.

Recherche de gènes candidats communs aux trois espèces

L'analyse de variants nucléotidiques a permis d'identifier un seul SNP concordant avec les deux méthodes de SNP calling utilisées et commun aux trois espèces. L'analyse du différentiel d'expression menée par une approche transcriptome *de novo* a permis l'identification de 5 gènes candidats, surexprimés chez les sensibles et communs aux trois espèces étudiées. Le faible nombre de gènes candidats communs aux trois espèces de Myrtceae étudiées peut être expliqué par plusieurs hypothèses. D'une part, il est probable que les mécanismes intervenant dans la résistance à la rouille chez les Myrtaceae néo-calédoniennes soient spécifiques à chaque espèce. D'autre part, l'infection par *A. psidii* n'ayant pas été contrôlée, il est possible que nous ayons prélevé les échantillons à des stades d'évolution de la maladie différents.

Recherche de gènes candidats spécifiques

En se plaçant à l'échelle spécifique, nous avons combiné la recherche de SNP discriminants avec celle de gènes différentiellement exprimés entre les individus résistants et les individus sensibles. Nous sommes ainsi parvenus à identifier 2 gènes candidats pour *A. gummiferu*m et 19 gènes candidats pour *S. longifolium*. Parmi les 19 gènes candidats isolés sur *S. longifolium,* de nombreux gènes possèdent une fonction impliquée dans la résistance au pathogène et ceux-là pourraient potentiellement être utilisés dans un programme de sélection assistée par marqueurs moléculaires.

Chapitre V : Discussion generale et perspectives

5.1. Conséquences écologiques et économiques de la rouille (*A. psidii*) en Nouvelle-Calédonie

Ces travaux nous ont permis d'en apprendre plus sur les dynamiques d'expansion d'*A. psidii* en Nouvelle-Calédonie. Après au moins trois années de présence, la gamme hôte d'*A. psidii* s'élève à 67 espèces de Myrtaceae endémiques, représentant environ 27% de l'ensemble des Myrtaceae décrites à ce jour sur le territoire. En termes d'étendue de la gamme hôte à *A. psidii*, la Nouvelle-Calédonie se place au second rang après l'Australie.

Cette gamme hôte est par ailleurs en constante expansion. De plus, les difficultés liées à l'identification même des espèces de Myrtaceae lorsqu'elles sont au stade juvénile, abîmées ou exemptes de fleurs et de fruits, ne permet pas d'avoir de chiffres fiables quant au nombre réel d'espèces concernées. Des solutions existent afin de contourner ces problèmes, comme par exemple l'identification par le barcoding moléculaire, une technique de taxonomie permettant de caractériser génétiquement des individus et de les classer au sein d'une base de données (Hebert et al., 2003).

Notre étude sur le niveau de diversité des populations d'*A. psidii* en Nouvelle-Calédonie a révélé qu'un seul génotype de cette rouille est présent sur le territoire. La rouille se propage donc localement sur l'ensemble de la gamme hôte de manière clonale, à l'instar du génotype pandémique infectant l'Australie, et Hawaii.

En termes d'espace conquis, la présence du pathogène a été relevée sur l'ensemble de la Grande-Terre, du Nord au Sud, ainsi que sur les trois Iles Loyautés. Les Bélep et l'Ile des Pins ont été, jusqu'à présent, considérés comme indemnes de cette maladie. Cependant cela pourrait également être dû à une surveillance moindre que celle exercée au sein des autres localités de Nouvelle-Calédonie.

De nombreuses pépinières cultivant les Myrtaceae ont été sévèrement touchées dès la première année d'explosion de la rouille. Suite à cela, elles ont été amenées à complètement ou partiellement réduire la production des Myrtaceae devenue trop contraignante. En milieu naturel, nous avons relevé la présence d'*A. psidii* dans différents types d'habitats : les maquis de basses et moyenne altitudes, les savanes et autre formations secondaires, les forêts denses humides de basses et moyennes altitudes, et la forêt sèche. Etant donné l'hétérogénéité des milieux néo-calédoniens et la difficulté d'accès pour certains d'entre eux, de nombreuses formations végétales n'ont pas pu être investiguées au cours de cette étude, en particulier les milieux en haute altitude. En effet, bien que se développant de manière optimale en climat tropical, la rouille peut également se développer en haute altitude dans une gamme de

températures plus froides comme c'est le cas à Hawaii, et également en Tasmanie (Kriticos et al., 2013; Potts et al., 2016), suggérant que les milieux de haute altitude en Nouvelle-Calédonie (maquis et forêts denses humides) pourraient également être susceptibles d'héberger le pathogène. De plus amples investigations en haute altitude sont nécessaires afin d'améliorer les connaissances sur la distribution de la maladie et la tolérance de gamme de température supportée par *A. psidii* en Nouvelle-Calédonie.

Une perspective de recherche axée sur de la modélisation spatiale pourrait également être développée en combinant diverses données de climatologie, de distribution spatiale des Myrtaceae en Nouvelle-Calédonie et de la gamme hôte identifiée dans ces travaux. Les résultats de cette recherche viendraient en appui à la définition des zones de conservation prioritaires à mettre en place.

5.2. Une possible part de génétique dans l'expression des réponses à la rouille mise en évidence par un dispositif d'infection naturelle

L'utilisation d'innoculations contrôlées ou de suivi d'infection à l'état naturel présentent toutes deux des avantages et des inconvénenients. Les inoculations contrôlées constituent la méthode la plus utilisée pour classer les différents types de réactions face à des souches spécifiques de pathogènes (Avellaneda et al., 2015 ; Challagulla et al., 2015 ; Shin et al., 2014). Elles permettent de réduire les variations environnementales afin de maximiser les chances d'observation de phénotypes contrastés. Cependant cette méthode ne permet pas de capturer l'ensemble des composantes de la résistance, tels que les mécanismes passifs de résistance (Laine et al., 2011).

Dans le cas du pathosystème Myrtaceae-*A. psidii*, les différents degrés de sensibilité et de résistance ont été évalués par inoculation contrôlée dans des chambres de culture, sur plants entiers et sur des boutures (Morin et al., 2012). Cette méthode présente l'avantage de permettre un optimum de température pour l'infection. Dans leur étude, Carnegie and Lidbetter (2011) ont testé à la fois l'inoculation contrôlée en chambre de culture et l'infection naturelle. Leur expérimentation, n'exposant les plants que 48 h à la source naturelle d'inoculum de rouille, n'a pas été suffisante pour permettre l'infection.

Dans notre étude, nous avons testé les degrés de sensibilité à la rouille de 35 espèces de Myrtaceae au stade juvénile et au sein d'un dispositif installé au sein d'une pépinière à Farino. Le dispositif, à ciel-ouvert, ne permettait pas de réaliser d'inoculations contrôlées. L'infection des plants par la rouille s'est effectuée de manière naturelle, à partir des *Syzygium jambos*

infectés situés à proximité du dispositif de suivi. Face au génotype unique d'*A. psidii* présent sur le territoire, une variabilité dans les réponses à la rouille a été observée chez les espèces hôtes. Au sein d'un milieu semi-contrôlé, cette variabilité de réponse s'exprime à plusieurs échelles : entre les espèces d'une part, et entre les individus d'une même espèce d'autre part. Ces différences se sont conservées sur une période de 3 ans de suivi. Nous confirmons donc dans cette étude qu'il est possible d'observer des variations dans les niveaux d'incidence et de sévérité à la rouille dans le cadre d'infection naturelle, sans inoculation contrôlée.

5.3. La construction de transcriptomes *de novo* pour trois espèces de Myrtaceae endémiques à la Nouvelle-Calédonie

Comprendre les bases moléculaires de la résistance aux agents pathogènes des plantes est un préalable indispensable dans l'élaboration d'une stratégie de lutte. Dans cette étude nous avons spécifiquement étudié le pathosystème *Austropuccinia psidii*-Myrtaceae, sur trois espèces de Myrtaceae hôtes à la rouille présentant un potentiel de résistance génétique : *Arillastrum gummiefrum*, *Syzygium longifolium*, *Tristaniopsis glauca*.

Nous avons ainsi reconstruit et annoté les transcriptomes *de novo* pour chacune des espèces étudiées afin de nous permettre de mener les différents types d'analyses subséquentes. Il s'agit des premiers transcriptomes de Myrtaceae, endémiques à la Nouvelle-Calédonie, réalisés à ce jour. Ces ressources pourraient être valorisées, en dehors du cadre de la rouille des Myrtaceae, pour bénéficier à la génomique comparative chez des espèces forestières non-modèles (notamment grâce aux Jbrowses). Par ailleurs, au-delà des résultats acquis pour l'annotation fonctionnelle de ces transcriptomes, il serait intéressant d'approfondir l'analyse des fonctions exprimées, en le comparant par exemple à celles d'autres organismes via le programme Gene Ontology (Ashburner et al., 2000).

5.4. Une nouvelle approche méthodologique pour trouver des gènes candidats

La recherche de marqueurs de résistance à la rouille *A. psidii* a été entreprise par différents moyens dans d'autres études et essentiellement sur le genre *Eucalyptus*. Par exemple, Moon et al. (2007) ont utilisé une approche SAGE (Serial analysis of gene expression) pour identifier des gènes différentiellement exprimés entre des individus sensibles et des individus résistants d'*E. grandis* face à l'infection par *A. psidii*. Ils ont pu identifier 40 sequences tags exprimées

préférentiellement par les individus sensibles et regroupant des fonctions de sénéscence des feuilles, de réponses au stress, de détoxification et l'absence d'induction de mécanismes de défense efficace. Ils ont également montré l'expression de 70 sequences tags préférentiellement exprimées chez les *E. grandis* résistants codant pour la restructuration du cytosquelette, le transport des vésicules, la biosynthèse de lignine et de cellulose.

Thumma et al. (2013) a mis en évidence 15 SNPs liés à la résistance à la rouille, par une approche Bulked segregant analysis au sein de populations naturelles d'*Eucalyptus (E. nitens ; E. globulus*, et *E. grandis)*. Ces SNPs étaient situés majoritairement dans des facteurs de transcription bZIP impliqués dans la réponse au stress ; et dans des gènes de résistance NBS-LRR situés dans la même région que le gène de résistance à la rouille *Ppr1*.

Dans une autre étude, Christie et al. (2016) ont identifié 1487 gènes *NBS-LRR* par une recherche analytique du domaine NBS sur l'annotation de l'*E. grandis.* Dans la même étude de Christie et al. (2016), une analyse RNA-Seq a été menée sur des échantillons d'*E. grandis* post-inoculés et infectés avec le champignon pathogène *Chrysoporthe austroafricana*. L'analyse du différentiel d'expression entre ces deux traitements a permis d'identifier 7 transcrits surexprimés chez les plantes résistantes à *C. austroafricana*.

Enfin, une analyse QTL menée par Butler et al. (2016) sur l'*E. globulus*, a permis d'identifier 4 QTLs liés à la présence, ou non, de symptômes et 2 autres QTLs liés à la présence, ou l'abensce, de réaction hypersensible.

Notre étude est la première à s'interesser aux variations de réponses au pathogène *A. psidii* au niveau transcriptomique sur trois espèces de Myrtaceae non-modèles. Ne disposant pas de référence de base comme c'est le cas chez l'Eucalyptus, nous avons choisi l'approche RNA-Seq afin d'accéder uniquement aux parties codantes des génomes étudiés.

Nous avons choisi de combiner les résultats obtenus séparemment par deux approches (différentiel d'expression et analyse SNP) afin d'identifier des gènes marqueurs de la résistance à *A. psidii*. En s'intéressant uniquement aux parties codantes des génomes étudiés, nous espérons pouvoir identifier rapidement des outils moléculaires applicables dans un programme de sélection de plants résistants pour les filières économiques dépendantes de la culture des Myrtaceae. Bien que de plus en plus d'études sur le RNA-Seq aient été publiées ces dernières années, il n'y a pas de consensus commun sur la méthodologie à appliquer pour ce type de travaux (Honaas et al., 2016). Beaucoup de difficultés dans l'adaptation des scripts en ligne de commande et la gestion des big data générés par les différents programmes d'analyse ont dû être levés pour atteindre les résultats obtenus.

5.5. Des mécanismes de réponse à la rouille spécifiques à chaque espèce

La comparaison des patterns d'expression des individus résistants semble montrer l'existence de mécanismes de résistance sépcifiques à chaque espèce, et donc non-partagés entre les espèces. Au regard de ces résultats, plusieurs hypothèses peuvent être formulées quant aux mécanismes de résistance à la rouille :

- Chez les individus résistants, nous supposons l'existence d'une résistance préformée (ou constitutive) empêchant la pénétration du pathogène à l'intérieur de la feuille et entrainant de fait, l'absence d'observation de symptômes de la maladie et la surexpression de gènes codant pour des protéines de défense. Cette résistance pré-formée serait également spécifique à chaque espèce. Il est également possible que cette observation puisse résulter d'un biais au moment de l'échantillonnage. En effet en l'absence d'inoculation contrôlée, il est impossible de savoir à quel stade de l'infection ont été récoltés les échantillons. Le pattern d'expression alors observé chez les individus résistants pourrait ainsi être incomplet.
- Chez les individus sensibles à la rouille, il y a reconnaissance de la présence du pathogène par l'activation de gènes impliqués dans les mécanismes de défense (comme les enzymes chitinases). Malgré ces premières réactions cellulaires, les individus sensibles expriment des symptômes visibles propres à la rouille. Nous pouvons alors supposer que le pathogène *A. psidii* a développé des moyens de contournement des moyens de défense chez les individus sensibles. Par ailleurs, plusieurs degrés de sensibilité au sein d'une même espèce face à la rouille ont été démontrés dans de précédentes études (Morin et al., 2012 ; Pegg et al., 2014a ; Zauza et al., 2010b). A partir de leur expérience avec des inoculations contrôlée de rouille, Pegg et al. (2014a) ont identifié au moins quatres niveaux de sensibilité qu'ils définissent par : relativement tolérant, modéremment sensible, hautement sensible et extrêment sensible. L'existence d'individus tolérants à modéremment sensibles dans le pool d'individus que nous avons considéré comme sensible, pourrait expliquer l'expression de gènes liés à la défense de l'organisme attaqué. Dans notre expérimentation, en l'absence d'inoculation contrôlée, il est difficile de conclure sur l'existence d'une telle hétérogénité des niveaux de sensibilité. Il pourrait alors être recommandé dans le cadre d'une inoculation contrôlé des espèces étudiées, d'identifier plus précisément les niveaux de sensibilité exprimés par chaque individu et d'effectuer pour chacun une étude du différentiel d'expression.

5.6. L'identification de gènes candidats pour deux espèces de Myrtaceae

En combinant l'approche d'analyse différentielle et le SNP calling à l'échelle spécifique, nous sommes parvenus à mettre en évidence 2 gènes candidats sur *A. gummiferum* et 19 gènes candidats sur *S. longifolium*. C'est aussi sans compter sur les nombreux SNPs discriminants identifiés sur des locus non différentiellement exprimés (non-DE) et qui pourraient être en réalité de faux non-DE si observés trop tôt ou trop tard par rapport à l'infection. Nous avons ainsi démontré la faisabilité de notre approche afin d'identifier, dans le cadre du temps impartis d'une thèse, des marqueurs candidats pour la sélection chez des espèces forestières pour lesquelles aucune donnée moléculaire n'était disponible dans la littérature. Il est important de prendre des précautions pour valider les résultats obtenus dans le cadre de cette étude. En effet, nous sommes partis d'une variabilité observée sur des espèces « naturelles » (au fond génétique diversifié par rapport à des espèces cultivées), en condition d'infection naturelle (biais des variations environnementales et la temporalité de l'infection, asynchronisme du développement des feuilles entre les individus). L'ensemble de ces variations pourrait être à l'origine de biais dans l'observation des données.

Ainsi**,** en priorité dans la suite de ce travail, nous voudrions vérifier si les changements de bases dans les gènes candidats affectent la conformité des protéines produites. Dans cet optique, nous utiliserons le programme SnpEff (Cingolani et al., 2012), spécialisé dans l'annotation et la prédiction de leur effet sur les régions codantes (mutation synonyme ou non synonyme).

En perspective de ce travail, nous envisageons également de valider ces gènes sur un plus grand nombre d'individus (par séquençage des régions concernées) et de vérifier le lien entre le génotype et les phénotypes de réponse à la rouille.

Il pourrait également être envisagé de suivre l'expression et la localisation de ces gènes candidats dans les tissus d'individus sensibles, et d'individus résistants, par marquage à la GFP (Green Fluorescent Protein) afin de mieux géolocaliser les phénomènes en cours. Si ces marqueurs sont validés au niveau moléculaire, nous pourrons alors envisager de tester leur héritabilité sur plusieurs générations de plants afin de consolider leur utilisation dans un programme de sélection.

Bibliographie

A

Adendorff, R. (1998). Infection of Kikuyu grass (*Pennisetum clandestinum*) by the rust fungus *Phakopsora apoda*. Thesis.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44, W3–W10.

Afzal, A.J., Wood, A.J., and Lightfoot, D.A. (2008). Plant Receptor-Like Serine Threonine Kinases: Roles in signaling and plant defense. Mol. Plant. Microbe Interact. 21, 507–517.

Agrios, G.N. (2005). chapter one - introduction. In Plant Pathology (Fifth Edition), (San Diego: Academic Press), pp. 3–75.

Ahuja, I., Kissen, R., and Bones, A.M. (2012). Phytoalexins in defense against pathogens. Trends Plant Sci. 17, 73–90.

Aime, M.C., Matheny, P.B., Henk, D.A., Frieders, E.M., Nilsson, R.H., Piepenbring, M., McLaughlin, D.J., Szabo, L.J., Begerow, D., Sampaio, J.P., et al. (2006). An overview of the higher level classification of Pucciniomycotina based on combined analyses of nuclear large and small subunit rDNA sequences. Mycologia 98, 896–905.

Alfen, N.K.V. (2001). Fungal Pathogens of Plants. In Encyclopedia of Life Sciences, John Wiley & Sons, Ltd, ed. (Chichester, UK: John Wiley & Sons, Ltd), p.

Alfenas, A.C., Zauza, E. a. V., and Assis, T.F. (2003a). First record of *Puccinia psidii* on *Eucalyptus globulus* and *E. viminalis* in Brazil. Australas. Plant Pathol. 32, 325–326.

Alfenas, A.C., Zauza, E. a. V., and Assis, T.F. (2003b). First record of *Puccinia psidii* on *Eucalyptus globulus* and *E. viminalis* in Brazil. Australas. Plant Pathol. 32, 325–326.

Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12, 363–376.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Alves, A.A., Guimarães, L.M. da S., Chaves, A.R. de M., DaMatta, F.M., and Alfenas, A.C. (2011). Leaf gas exchange and chlorophyll a fluorescence of *Eucalyptus urophylla* in response to *Puccinia psidii* infection. Acta Physiol. Plant. 33, 1831–1839.

Amirsadeghi, S., Robson, C.A., McDonald, A.E., and Vanlerberghe, G.C. (2006). Changes in plant mitochondrial electron transport alter cellular levels of reactive oxygen species and susceptibility to cell death signaling molecules. Plant Cell Physiol. 47, 1509–1519.

Aparecido, C.C., Figueiredo, M.B., and Furtado, E.L. (2003). Grupos de variabilidade fisiologica em populaçôes de *Puccinia psidii*. Summa Phytopatológica 29, 234–238.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature genetics 25, 25–29.

Avellaneda, M. C., Hoy, J. W. and Pontif, M. J. (2015). Screening for resistance to sugarcane brown rust with controlled-conditions inoculation. Plant Disease 99, 1633–1639.

B

Barbosa-da-Silva, A., Wanderley-Nogueira, A.C., Silva, R.R., Berlarmino, L.C., Soares-Cavalcanti, N.M., and Benko-Iseppon, A.M. (2005). In silico survey of resistance (R) genes in *Eucalyptus* transcriptome. Genet. Mol. Biol. 28, 562–574.

Bari, R., and Jones, J.D.G. (2009). Role of plant hormones in plant defence responses. Plant Mol. Biol. 69, 473–488.

Beenken, L. (2017). Austropuccinia: a new genus name for the myrtle rust *Puccinia psidii* placed within the redefined family Sphaerophragmiaceae (Pucciniales). Phytotaxa 297, 53.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

Bennett, R.N., and Wallsgrove, R.M. (1994). Secondary metabolites in plant defence mechanisms. New Phytol. 127, 617–633.

Bettgenhaeuser, J., Gilbert, B., Ayliffe, M., and Moscou, M.J. (2014). Nonhost resistance to rust pathogens – a continuation of continua. Front. Plant Sci. 5.

Biffin, E., Lucas, E.J., Craven, L.A., Ribeiro da Costa, I., Harrington, M.G., and Crisp, M.D. (2010). Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. Ann. Bot. 106, 79–93.

Bigeard, J., Colcombet, J., and Hirt, H. (2015). Signaling mechanisms in Pattern-Triggered Immunity (PTI). Mol. Plant 8, 521–539.

Bisconti, R., Canestrelli, D., Salvi, D., and Nascetti, G. (2013). A geographic mosaic of evolutionary lineages within the insular endemic newt Euproctus montanus. Mol. Ecol. 22, 143–156.

Boerjan, W. (2005). Biotechnology and the domestication of forest trees. Curr. Opin. Biotechnol. 16, 159–166.

Bonde, M.R. (1976). Histology of the suscept-pathogen relationship between glycine max and *Phakopsora pachyrhizi*, the cause of soybean rust. Phytopathology 66, 1290.

Bonde, M.R., Bromfield, K.R., and Melching, J.S. (1982). Morphological development of *Physopella zeae* on corn [*Zea mays*, maize]. Phytopathol. USA.

Booth, T.H., Old, K.M., and Jovanovic, T. (2000). A preliminary assessment of high risk areas for *Puccinia psidii* (Eucalyptus rust) in the Neotropics and Australia. Agric. Ecosyst. Environ. 82, 295–301.

Bouchet, P., Jaffre, T., and Veillon, J.-M. (1995). Plant extinction in New Caledonia: protection of sclerophyll forests urgently needed. Biodivers. Conserv. 4, 415–428.

Bouwmeester, K., Sain, M. de, Weide, R., Gouget, A., Klamer, S., Canut, H., and Govers, F. (2011). The Lectin Receptor Kinase LecRK-I.9 is a novel phytophthora resistance component and a potential host target for a rxlr effector. PLOS Pathog. 7, e1001327.

Brasier, C.M., and Buck, K.W. (2001). Rapid evolutionary changes in a globally invading fungal pathogen (Dutch Elm Disease). Biol. Invasions 3, 223–233.

Brodie, A., Azaria, J.R., and Ofran, Y. (2016). How far from the SNP may the causative genes be? Nucleic Acids Res. 44, 6046–6054.

Brondani, R.P.V., Williams, E.R., Brondani, C., and Grattapaglia, D. (2006). A microsatellite-based consensus linkage map for species of Eucalyptus and a novel set of 230 microsatellite markers for the genus. BMC Plant Biol. 6, 20.

Brooker, M.I.H. (2000). A new classification of the genus Eucalyptus L'Hér. (Myrtaceae). Aust. Syst. Bot. 13, 79–148.

Brown, J.K.M. (2015). Durable resistance of crops to disease: A Darwinian Perspective. Annu. Rev. Phytopathol. 53, 513–539.

Butler, J. B., Freeman, J. S., Vaillancourt, R. E., Potts, B. M., Glen, M., Lee, D. J. and Pegg, G. S. (2016). Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*. Tree Genetics & Genomes

Buys, M.H., Flint, H.J., Miller, E.M., Yao, H., Caird, A.R., and Ganley, R.J. (2016). Preparing for the invasion: efficacy of DNA barcoding to discern the host range of myrtle rust (*Puccinia psidii*) among species of Myrtaceae. Forestry 89, 263–270.

C

Canabal, M. (2011). De la découverte scientifique à l'occultation du savoir : Isaac-Bénédict Prévost et la carie du blé (1798-1807), Résumé, Summary. Rev. Hist. Sci. Tome 63, 501–527.

Cantor, R.M., Lange, K., and Sinsheimer, J.S. (2010). Prioritizing GWAS Results: A review of statistical methods and recommendations for their application. Am. J. Hum. Genet. 86, 6–22.

Carletti, G., Carra, A., Allegro, G., Vietto, L., Desiderio, F., Bagnaresi, P., Gianinetti, A., Cattivelli, L., Valè, G., and Nervo, G. (2016). QTLs for Woolly Poplar Aphid (*Phloeomyzus passerinii* L.) Resistance Detected in an inter-specific *Populus deltoides* x *P. nigra* Mapping Population. PLoS ONE 11.

Carnegie, A.J. (2014). First report of *Puccinia psidii* (Myrtle Rust) in Eucalyptus plantations in Australia. Plant Dis. 99, 161–161.

Carnegie, A.J., and Cooper, K. (2011). Emergency response to the incursion of an exotic myrtaceous rust in Australia. Australas. Plant Pathol. 40, 346–359.

Carnegie, A.J., and Lidbetter, J.R. (2011). Rapidly expanding host range for Puccinia psidii sensu lato in Australia. Australas. Plant Pathol. 41, 13–29.

Carnegie, A.J., Lidbetter, J.R., Walker, J., Horwood, M.A., Tesoriero, L., Glen, M., and Priest, M.J. (2010). *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. Australas. Plant Pathol. 39, 463–466.

Carnegie, A.J., Kathuria, A., Pegg, G.S., Entwistle, P., Nagel, M., and Giblin, F.R. (2016a). Impact of the invasive rust *Puccinia psidii* (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biol. Invasions 18, 127–144.

Carnegie, A.J., Kathuria, A., Pegg, G.S., Entwistle, P., Nagel, M., and Giblin, F.R. (2016b). Impact of the invasive rust *Puccinia psidii* (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biol. Invasions 1–18.

Catanzariti, A.-M., Dodds, P.N., and Ellis, J.G. (2007). Avirulence proteins from haustoria-forming pathogens: Avirulence proteins from haustoria-forming pathogens. FEMS Microbiol. Lett. 269, 181–188.

Catherine Aime, M. (2006). Toward resolving family-level relationships in rust fungi (Uredinales). Mycoscience 47, 112–122.

Challagulla, V., Bhattarai, S. and Midmore, D. J. (2015). In-vitro vs in-vivo Inoculation: Screening for Resistance of Australian Rice Genotypes Against Blast Fungus. Rice Science 22, 132–137.

Chaudhari, P., Ahmed, B., Joly, D.L., and Germain, H. (2014). Effector biology during biotrophic invasion of plant cells. Virulence 5, 703–709.

Chisholm, S.T., Coaker, G., Day, B., and Staskawicz, B.J. (2006). Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. Cell 124, 803–814.

Choi, H.W., and Klessig, D.F. (2016). DAMPs, MAMPs, and PAMPs in plant innate immunity. BMC Plant Biol. 16.

Christie, N., Tobias, P.A., Naidoo, S., and Külheim, C. (2016). The *Eucalyptus grandis* NBS-LRR Gene Family: Physical Clustering and Expression Hotspots. Front. Plant Sci. 6.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80–92. doi:10.4161/fly.19695.

Coelho, L., Alfenas, A.C., and Ferreira, F.A. Variabilidade fisiológica de *Puccinia psidii* — ferrugem do eucalipto. Summa Phytopathol. 27, 295–300.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., et al. (2016). A survey of best practices for RNA-seq data analysis. Genome Biol. 17.

Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with rna-seq.

Coutinho, T.A., Wingfield, M.J., Alfenas, A.C., and Crous, P.W. (1998). Eucalyptus rust: A disease with the potential for serious international implications. Plant Dis. 82, 819–825.

Crous, P.W., Hawksworth, D.L., and Wingfield, M.J. (2015). Identifying and naming plant-pathogenic fungi: past, present, and future. Annu. Rev. Phytopathol. 53, 247–267.

Crute, I., and Pink, D. (1996). Genetics and utilization of pathogen resistance in Plants. Plant Cell 8, 1747–1755.

Cuartero, J., Laterrot, H., and Lenteren, J.C. van (1999). Host- plant resistance to pathogens and arthropod pests. in integrated pest and disease management in greenhouse crops, (Springer, Dordrecht), pp. 124–138.

Cui, X., and Churchill, G.A. (2003). Statistical tests for differential expression in cDNA microarray experiments. Genome Biol. 4, 210.

Cui, H., Wang, C., Qin, T., Xu, F., Tang, Y., Gao, Y., and Zhao, K. (2017). Promoter variants of Xa23 alleles affect bacterial blight resistance and evolutionary pattern. PloS One 12, e0185925.

D

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

De Carvalho, A.D., Alfenas, A.C., Maffia, L.A., and Do Carmo, M.G.F. (1998). Resistance of Eucalyptus species, progenies and provenances to *Puccinia psidii* winter. Pesqui. Agropecu. Bras. 33, 139–147.

Deising, H. (1992). Adhesion pad formation and the involvement of cutinase and esterases in the attachment of uredospores to the host cuticle. Plant cell online 4, 1101–1111.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

Deschamps, S., Llaca, V., and May, G.D. (2012). Genotyping-by-Sequencing in Plants. Biology 1, 460–483.

Desprez-Loustau, M.-L., Aguayo, J., Dutech, C., Hayden, K.J., Husson, C., Jakushkin, B., Marçais, B., Piou, D., Robin, C., and Vacher, C. (2016). An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. Ann. For. Sci. 73, 45–67.

Dianese, J.C. (1984). Response of *Eucalyptus species* to field infection by *Puccinia psidii*. Plant Dis. 68, 314.

Dinesh-Kumar, S.P., Tham, W.-H., and Baker, B.J. (2000). Structure–function analysis of the tobacco mosaic virus resistance gene N. Proc. Natl. Acad. Sci. U. S. A. 97, 14789–14794.

Dodd, A.N., Kudla, J., and Sanders, D. (2010). The Language of Calcium Signaling. Annu. Rev. Plant Biol. 61, 593–620.

Dodds, P., and Thrall, P. (2009). Goldacre paper : Recognition events and host–pathogen co-evolution in gene-for-gene resistance to flax rust. Funct. Plant Biol. 36, 395.

Donovan, B.J., Munzinger, J., Pauly, A., and McPherson, G. (2013). Flower-visiting records of the native bees of New Caledonia. Ann. Mo. Bot. Gard. 99, 19–43.

Doughty, R.W. (2000). The eucalyptus: a natural and commercial history of the gum tree (Baltimore, Md: Johns Hopkins University Press).

Droulers, M., and Andrade, C. de (2000). Papier, nouveau secteur d'exportation brésilien. Inf. Géographique 64, 262–264.

Du, L., Ali, G.S., Simons, K.A., Hou, J., Yang, T., Reddy, A.S.N., and Poovaiah, B.W. (2009). Ca2+/calmodulin regulates salicylic-acid-mediated plant immunity. Nature 457, 1154–1158.

Duplessis, S., Cuomo, C.A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L., et al. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. 108, 9166–9171.

E

Ekblom, R., and Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107, 1–15.

Ellison, A.M., Bank, M.S., Clinton, B.D., Colburn, E.A., Elliott, K., Ford, C.R., Foster, D.R., Kloeppel, B.D., Knoepp, J.D., Lovett, G.M., et al. (2005). Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Front. Ecol. Environ. 3, 479–486.

Ennos, R.A. (2015). Resilience of forests to pathogens: an evolutionary ecology perspective. Forestry 88, 41–52.

Everett, M.V., Grau, E.D., and Seeb, J.E. (2011). Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. Mol. Ecol. Resour. 11, 93–108.

F

Ferrari, J.T., Nogueira, E.M. d., and dos Santos, A.J.T. (1997). Control of rust (*Puccinia psidii*) in guava (*Psidium guajava*). Acta Hortic. 55–58.

Feys, B.J., and Parker, J.E. (2000). Interplay of signaling pathways in plant disease resistance. Trends Genet. 16, 449–455.

Figueiredo, M.B. (2001). Life cycle and ecology of *Puccinia psidii*.

Fitch, W.M. (2000). Homology a personal view on some of the problems. Trends Genet. TIG 16, 227–231.

Fleming, T.H., Geiselman, C., and Kress, W.J. (2009). The evolution of bat pollination: a phylogenetic perspective. Ann. Bot. 104, 1017–1043.

Flor, H.H. (1955). Host-parasite interactions in flax rust-its genetics and other implications.

Flor, H.H. (1971). Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. 9, 275–296.

Foolad, M.R., and Sharma, A. (2005). Molecular markers as selection tools in tomato breeding. Acta Hortic. 225–240.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150.

G

Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. Nat. Methods 8, 469–477.

Gardes, M., and Bruns, T.D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. Mol. Ecol. 2, 113–118.

Giblin, F.R. (2013). Myrtle rust report : New Caledonia. University of the Sunshine Coast, Maroochydore, Queensland, Australia.

Giblin, F.R., and Carnegie, A.J. (2014). Puccinia psidii (Myrtle Rust) – Australian host list. Version current at 24 Sept. 2014. http://www.anpc.asn.au/resources/Myrtle_Rust.html.

Gillespie, T.W., and Jaffré, T. (2003). Tropical dry forests in New Caledonia. Biodivers. Conserv. 12, 1687–1697.

Glen, M., Alfenas, A.C., Zauza, E. a. V., Wingfield, M.J., and Mohammed, C. (2007). Puccinia psidii: a threat to the Australian environment and economy —a review. Australas. Plant Pathol. 36, 1–16.

Gomez, C., Mangeas, M., Curt, T., Ibanez, T., Munzinger, J., Dumas, P., Jérémy, A., Despinoy, M., and Hély, C. (2015). Wildfire risk for main vegetation units in a biodiversity hotspot: modeling approach in New Caledonia, South Pacific. Ecol. Evol. 5, 377–390.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Graça, R.N., Alfenas, A.C., Ross-Davis, A.L., Klopfenstein, N., Kim, M.-S., Peever, T.L., Cannon, P.G., Uchida, J.Y., Kadooka, C.Y., and Hauff, R.D. (2011a). Multilocus genotypes indicate differentiation among *Puccinia psidii* populations from South America and Hawaii.

Graça, R.N., Aun, C.P., Guimarães, L.M.S., Rodrigues, B.V.A., Zauza, E.A.V., and Alfenas, A.C. (2011b). A new race of *Puccinia psidii* defeats rust resistance in eucalypt. Australas. Plant Pathol. 40, 442–447.

Graça, R.N., Alfenas, A.C., Ross-Davis, A.L., Klopfenstein, N., Kim, M.-S., Peever, T.L., Cannon, P.G., Uchida, J.Y., Kadooka, C.Y., and Hauff, R.D. (2011c). Multilocus genotypes indicate differentiation among *Puccinia psidii* populations from South America and Hawaii. In: Proceedings of the 58th Annual Western International Forest Disease Work Conference; 2010 October 4–8; Valemount, BC: 131–134 Flagstaff, AZ: US Forest Service, AZ Zone Forest Health.

Graça, R.N., Ross-Davis, A.L., Klopfenstein, N.B., Kim, M.-S., Peever, T.L., Cannon, P.G., Aun, C.P., Mizubuti, E.S.G., and Alfenas, A.C. (2013). Rust disease of eucalypts, caused by *Puccinia psidii* , did not originate via host jump from guava in Brazil. Mol. Ecol. 22, 6033–6047.

Granados, G.M., McTaggart, A.R., Barnes, I., Rodas, C.A., Roux, J., and Wingfield, M.J. (2017). The pandemic biotype of *Austropuccinia psidii* discovered in South America. Australas. Plant Pathol.

Grandcolas, P., Murienne, J., Robillard, T., Desutter-Grandcolas, L., Jourdan, H., Guilbert, E., and Deharveng, L. (2008). New Caledonia: a very old Darwinian island? Philos. Trans. R. Soc. B Biol. Sci. 363, 3309–3317.

Grattapaglia, D., Vaillancourt, R.E., Shepherd, M., Thumma, B.R., Foley, W., Külheim, C., Potts, B.M., and Myburg, A.A. (2012). Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. Tree Genet. Genomes 8, 463–508.

H

Hammond-Kosack, K.E., Kanyuka, K., Hammond-Kosack, K.E., and Kanyuka, K. (2007). Resistance Genes (R Genes) in Plants, Resistance Genes (R Genes) in Plants. In eLS, eLS, (John Wiley & Sons, Ltd, John Wiley & Sons, Ltd), p.

Harfouche, A., Meilan, R., Kirst, M., Morgante, M., Boerjan, W., Sabatti, M., and Scarascia Mugnozza, G. (2012). Accelerating the domestication of forest trees in a changing world. Trends Plant Sci. 17, 64–72.

Harper, A.L. (2016). A Transcriptomics Approach to the Development of Predictive Molecular Markers for Tolerance to Tree Diseases: Ash Dieback. (Plant and Animal Genome), p.

Harrison, P.W., Wright, A.E., and Mank, J.E. (2012). The evolution of gene expression and the transcriptome–phenotype relationship. Semin. Cell Dev. Biol. 23, 222–229.

Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., and Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. BioTechniques 56, 61–passim.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003). Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321.

Hématy, K., Cherk, C., and Somerville, S. (2009). Host–pathogen warfare at the plant cell wall. Curr. Opin. Plant Biol. 12, 406–413.

Horbach, R., Navarro-Quesada, A.R., Knogge, W., and Deising, H.B. (2011). When and how to kill a plant cell: Infection strategies of plant pathogenic fungi. J. Plant Physiol. 168, 51–62.

Houterman, P.M., Cornelissen, B.J.C., and Rep, M. (2008). Suppression of plant resistance gene-based immunity by a fungal effector. PLoS Pathog. 4, e1000061.

Hunt, P. (1968). Cuticular penetration by germinating uredospores. Trans. Br. Mycol. Soc. 103–112.

I

Ibanez, T., Munzinger, J., Dagostini, G., Hequet, V., Rigault, F., Jaffré, T., and Birnbaum, P. (2014). Structural and floristic diversity of mixed tropical rain forest in New Caledonia: new data from the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN). Appl. Veg. Sci. 17, 386–397.

Innes, R.W. (2001). Mapping out the roles of MAP kinases in plant defense. Trends Plant Sci. 6, 392–394.

Isnard, S., L'huillier, L., Rigault, F., and Jaffré, T. (2016). How did the ultramafic soils shape the flora of the New Caledonian hotspot? Plant Soil 403, 53–76.

Izuno, A., Hatakeyama, M., Nishiyama, T., Tamaki, I., Shimizu-Inatsugi, R., Sasaki, R., Shimizu, K.K., and Isagi, Y. (2016). Genome sequencing of *Metrosideros polymorpha* (Myrtaceae), a dominant species in various habitats in the Hawaiian Islands with remarkable phenotypic variations. J. Plant Res. 129, 727–736.

J

Jaffré, T., Rigault, F., Dagostini, G., Tinel-Fambart, J., Wulff, A., and Munzinger, J. (2009). Input of the different vegetation units to the richness and endemicity of the New Caledonian flora. Flora 162, 54–9.

Jarosz, A.M., and Davelos, A.L. (1995). Effects of disease in wild plant populations and the evolution of pathogen aggressiveness. New Phytol. 129, 371–387.

Jeandet, P., Clément, C., Courot, E., and Cordelier, S. (2013). Modulation of phytoalexin biosynthesis in engineered plants for disease resistance. Int. J. Mol. Sci. 14, 14136–14170.

Jones, J.D., and Dangl, J.L. (2006). The plant immune system. Nature 444, 323–329.

Juge, N. (2006). Plant protein inhibitors of cell wall degrading enzymes. Trends Plant Sci. 11, 359–367.

Junghans, D.T., Alfenas, A.C., Brommonschenkel, S.H., Oda, S., Mello, E.J., and Grattapaglia, D. (2003). Resistance to rust (Puccinia psidii Winter) in Eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers. Theor. Appl. Genet. 108, 175–180.

K

Kawanishi, T., Uematsu, S., Kakishima, M., Kagiwada, S., Hamamoto, H., Horie, H., and Namba, S. (2009). First report of rust disease on ohia and the causal fungus, Puccinia psidii, in Japan. J. Gen. Plant Pathol. JGPP 75, 428–431.

Keitt, B., Campbell, K., Saunders, A., Clout, M., Wang, Y., Heinz, R., Newton, K., and Tershy, B. (2011). The global islands invasive vertebrate eradication database: a tool to improve and facilitate restoration of island ecosystems. Isl. Invasives Erad. Manag. IUCN Gland Switz. 74–77.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. 12, 996–1006.

Keszei, A., Brubaker, C.L., and Foley, W.J. (2008). A molecular perspective on terpene variation in Australian Myrtaceae. Aust. J. Bot. 56, 197.

Kier, G., Kreft, H., Lee, T.M., Jetz, W., Ibisch, P.L., Nowicki, C., Mutke, J., and Barthlott, W. (2009). A global assessment of endemism and species richness across island and mainland regions. Proc. Natl. Acad. Sci. 106, 9322–9327.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360.

Kinloch, B.B., Sniezko, R.A., Barnes, G.D., and Greathouse, T.E. (1999). A major gene for resistance to white pine blister rust in Western white pine from the Western cascade range. Phytopathology 89, 861–867.

Klosterman, S.J., Rollins, J.R., Sudarshana, M.R., and Vinatzer, B.A. (2016). Disease Management in the Genomics Era—Summaries of Focus Issue Papers. Phytopathology 106, 1068–1070.

Kohorn, B.D., and Kohorn, S.L. (2012). The cell wall-associated kinases, WAKs, as pectin receptors. Front. Plant Sci. 3.

Kolmer, J.A., Ordonez, M.E., and Groth, J.V. (2009). The Rust Fungi. In Encyclopedia of Life Sciences, John Wiley & Sons, Ltd, ed. (Chichester, UK: John Wiley & Sons, Ltd), p.

Kombrink, E., and Schmelzer, E. (2001). The hypersensitive response and its role in local and systemic disease resistance. Eur. J. Plant Pathol. 107, 69–78.

Kriticos, D.J., Morin, L., Leriche, A., Anderson, R.C., and Caley, P. (2013). Combining a climatic niche model of an invasive fungus with its host species distributions to identify risks to Natural Assets: Puccinia psidii Sensu Lato in Australia: e64479. PLoS One 8.

Kuzniar, A., van Ham, R.C.H.J., Pongor, S., and Leunissen, J.A.M. (2008). The quest for orthologs: finding the corresponding gene across genomes. Trends Genet. 24, 539–551.

L

La Mantia, J., Klápště, J., El-Kassaby, Y.A., Azam, S., Guy, R.D., Douglas, C.J., Mansfield, S.D., and Hamelin, R. (2013). Association analysis identifies Melampsora ×columbiana poplar leaf rust resistance SNPs. PloS One 8, e78423.

Laia, M.L., Alfenas, A.C., Brommonschenkel, S.H., Oda, S., de Melo, E.J., de Araújo Silva, I.M., Gonçalves, J.F., and Marques, A. (2015). Cloning and characterization of gene-resistant analogs (RGAs) involved in rust (Puccinia psidii) resistance in *Eucalyptus grandis*. J. For. Res. 26, 711–718.

Laine, A.-L., Burdon, J. J., Dodds, P. N. and Thrall, P. H. (2011). Spatial variation in disease resistance: from molecules to metapopulations: Spatial variation in disease resistance. Journal of Ecology 99, 96–112.

Lamb, C., and Dixon, R.A. (1997). The oxidative burst in plant disease resistance. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48, 251–275.

Langrell, S.R.H., Glen, M., and Alfenas, A.C. (2008). Molecular diagnosis of *Puccinia psidii* (guava rust) – a quarantine threat to Australian eucalypt and Myrtaceae biodiversity. Plant Pathol. 57, 687–701.

Lannou, C. (2012). Variation and selection of quantitative traits in plant pathogens. Annu. Rev. Phytopathol. 50, 319–338.

Lee, S., Whitaker, V.M., and Hutton, S.F. (2016). Mini Review: Potential Applications of Non-host Resistance for Crop Improvement. Front. Plant Sci. 7.

Lefèvre, F., Goué-Mourier, M.C., Faivre-Rampant, P., and Villar, M. (1998). A single gene cluster controls incompatibility and partial resistance to various melampsora Larici-Populina races in hybrid poplars. Phytopathology 88, 156–163.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet. 20, 116–122.

Leonard, K.J., and Szabo, L.J. (2005). Stem rust of small grains and grasses caused by *Puccinia graminis*. Mol. Plant Pathol. 6, 99–111.

L'Huillier, L., Jaffré, T., and Wulff, A. (2010). Mines et environnement en Nouvelle-Calédonie: les milieux sur substrats ultramafiques et leur restauration (Éditions IAC).

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li, C., Wu, H.-M., and Cheung, A.Y. (2016). FERONIA and Her Pals: Functions and Mechanisms. Plant Physiol. 171, 2379–2392.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 25, 2078–2079.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. 13, 2178–2189.

Li, N., Zhao, M., Liu, T., Dong, L., Cheng, Q., Wu, J., Wang, L., Chen, X., Zhang, C., Lu, W., et al. (2017). A Novel Soybean Dirigent Gene GmDIR22 Contributes to Promotion of Lignan Biosynthesis and Enhances Resistance to Phytophthora sojae. Front. Plant Sci. 8.

Li, Z.K., Sanchez, A., Angeles, E., Singh, S., Domingo, J., Huang, N., and Khush, G.S. (2001). Are the dominant and recessive plant disease resistance genes similar? A case study of rice R genes and *Xanthomonas oryzae* pv. *oryzae* races. Genetics 159, 757–765.

Lima, B.M., Teixeira, J.E., Gazaffi, R., Garcia, A.A., Grattapaglia, D., Valle, R.K., and Camargo, L.E. (2011). Identification of a novel QTL contributing to rust resistance in Eucalyptus. BMC Proc. 5, P32.

Liu, J.-J., Sturrock, R., and Ekramoddoullah, A.K.M. (2010). The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Rep. 29, 419–436.

Liu, J.-J., Schoettle, A.W., Sniezko, R.A., Sturrock, R.N., Zamany, A., Williams, H., Ha, A., Chan, D., Danchok, B., Savin, D.P., et al. (2016). Genetic mapping of Pinus flexilis major gene (Cr4) for resistance to white pine blister rust using transcriptome-based SNP genotyping. BMC Genomics 17.

Loope, L. (2010). A summary of information on the rust Puccinia psidii Winter (guava rust) with emphasis on means to prevent introduction of additional strains to Hawaii (U.S. Geological Survey).

Loope, L.L., and Uchida, J.Y. (2012). The challenge of retarding erosion of island biodiversity through phytosanitary measures: An update on the case of Puccinia psidii in Hawai‘i. Pac. Sci. 66, 127–139.

Lopez-Maestre, H., Brinza, L., Marchet, C., Kielbassa, J., Bastien, S., Boutigny, M., Monnin, D., Filali, A.E., Carareto, C.M., Vieira, C., et al. (2016). SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. Nucleic Acids Res. 44, e148–e148.

M

Machado, P. da S. (2015). Epitypification, cytology, genetic and physiological variability, and genomic analysis of Puccinia psidii.

Machado, P., Glen, M., Pereira, O.L., Silva, A.A., and Alfenas, A.C. (2015a). Epitypification of *Puccinia psidii*, causal agent of guava rust. Trop. Plant Pathol. 40, 5–12.

Machado, P., Alfenas, A.C., Alfenas, R.F., Mohammed, C.L., and Glen, M. (2015b). Microsatellite analysis indicates that Puccinia psidii in Australia is mutating but not recombining. Australas. Plant Pathol. 44, 455–462.

Maclachlan, J.D. (1938). A rust of the Pimento tree in Jamaica, B.W.I. Phytopathology 28.

Mamani, E.M.C., Bueno, N.W., Faria, D.A., Guimarães, L.M.S., Lau, D., Alfenas, A.C., and Grattapaglia, D. (2010). Positioning of the major locus for *Puccinia psidii* rust resistance (*Ppr1*) on the Eucalyptus reference map and its validation across unrelated pedigrees. Tree Genet. Genomes 6, 953–962.

Mardis, E.R. (2016). The challenges of big data. Dis. Model. Mech. 9, 483–485.

Marlatt, R.B., and Kimbrough, J.W. (1979). *Puccinia psidii* on *Pimenta dioica* in south Florida. Plant Dis. Report. 63, 510–512.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10.

McDonald, B.A. (1997). The Population genetics of fungi: tools and techniques. Phytopathology 87, 448–453.

McDonald, B.A., Linde, and C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40, 349–379.

McGregor, C.E., Miano, D.W., LaBonte, D.R., Hoy, M., Clark, C.A., and Rosa, G.J.M. (2009). Differential gene expression of resistant and susceptible sweetpotato plants after infection with the causal agents of sweet potato virus disease. J. Am. Soc. Hortic. Sci. 134, 658–666.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A map reduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

McTaggart, A.R., Roux, J., Granados, G.M., Gafur, A., Tarrigan, M., Santhakumar, P., and Wingfield, M.J. (2015). Rust (*Puccinia psidii*) recorded in Indonesia poses a threat to forests and forestry in South-East Asia. Australas. Plant Pathol.

McTaggart, A.R., Roux, J., Granados, G.M., Gafur, A., Tarrigan, M., Santhakumar, P., and Wingfield, M.J. (2016). Rust (Puccinia psidii) recorded in Indonesia poses a threat to forests and forestry in South-East Asia. Australas. Plant Pathol. 45, 83–89.

McTaggart, A.R., Shuey, L.S., Granados, G.M., du Plessis, E., Fraser, S., Barnes, I., Naidoo, S., Wingfield, M.J., and Roux, J. (2017). Evidence that *Austropuccinia psidii* may complete its sexual life cycle on Myrtaceae. Plant Pathol.

Médigue, C., Bocs, S., Labarre, L., Mathé, C., and Vallenet, D. (2002). L'annotation in silico des séquences génomiques - Bio-informatique (1). médecine/sciences 18, 237–250.

Mellano, V. (2006). Rust on myrtle found in San Diego County. Healthy Gard.-Healthy Home 3.

Mendgen, K., and Hahn, M. (2002). Plant infection and the establishment of fungal biotrophy. Trends Plant Sci. 7, 352–356.

Mendgen, K., Struck, C., Voegele, R.T., and Hahn, M. (2000). Biotrophy and rust haustoria. Physiol. Mol. Plant Pathol. 56, 141–145.

Meng, X., and Zhang, S. (2013). MAPK Cascades in plant disease resistance signaling. Annu. Rev. Phytopathol. 51, 245–266.

van der Merwe, M.M., Walker, J., Ericson, L., and Burdon, J.J. (2008). Coevolution with higher taxonomic host groups within the Puccinia/Uromyces rust lineage obscured by host jumps. Mycol. Res. 112, 1387–1408.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8, 1113–1130.

Michelmore, R.W., Paran, I., and Kesseli, R.V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. 88, 9828–9832.

Moon, D. H., Salvatierra, G. R., Caldas, D. G. G., Gallo de Carvalho, M. C. C., Carneiro, R. T., Franceschini, L. M., Oda, S. and Labate, C. A. (2007). Comparison of the expression profiles of susceptible and resistant Eucalyptus grandis exposed to Puccinia psidii Winter using SAGE. Functional Plant Biology 34, 1010–1018.

Moose, S.P., and Mumm, R.H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol. 147, 969–977.

Morat, P. (1993). Our Knowledge of the Flora of New Caledonia: Endemism and Diversity in Relation to Vegetation Types and Substrates. Biodivers. Lett. 1, 72.

Morat, P., Jaffré, T., and Veillon, J.-M. (2001). The flora of New Caledonia's calcareous substrates. Adansonia 23, 109–127.

Morat, P., Jaffré, T., Tronchet, F., Munzinger, J., Pillon, Y., Veillon, J.-M., Chalopin, M., Birnbaum, P., Rigault, F., Dagostini, G., et al. (2012). Le référentiel taxonomique Florical et les caractéristiques de la flore vasculaire indigène de la Nouvelle-Calédonie. Adansonia 34, 179–221.

Morel, J.-B., and Dangl, J.L. (1997). The hypersensitive response and the induction of cell death in plants. Cell Death Differ. 4, 671–683.

Morin, L., Aveyard, R., Lidbetter, J.R., and Wilson, P.G. (2012). Investigating the host-range of the rust fungus *Puccinia psidii* sensu lato across tribes of the family Myrtaceae Present in Australia: e35434. PLoS One 7.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628.

Mundt, C.C. (2014). Durable resistance: A key to sustainable management of pathogens and pests. Infect. Genet. Evol. 27, 446–455.

Mur, L.A.J., Kenton, P., Lloyd, A.J., Ougham, H., and Prats, E. (2008). The hypersensitive response; the centenary is upon us but how much do we know? J. Exp. Bot. 59, 501–520.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., et al. (2014). The genome of *Eucalyptus grandis.* Nature 510, 356–362.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B. da, and Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature 403, 853–858.

N

Nakagami, H., Pitzschke, A., and Hirt, H. (2005). Emerging MAP kinase pathways in plant stress signalling. Trends Plant Sci. 10, 339–346.

Nakamura, E. (2013). Myrtle Rust in New-Caledonia. Personal Communication at the Meeting Thirteen of The Myrtle Rust Transition to Management Group. Australia. May 28, 2013.

Neale, D.B. (2007). Genomics to tree breeding and forest health. Curr. Opin. Genet. Dev. 17, 539–544.

Neale, D.B., and Kremer, A. (2011). Forest tree genomics: growing resources and applications. Nat Rev Genet 12, 111–122.

Nelson, E.B. (1994). The disease triangle and the disease cycle. Turf Grass Trends USA.

Newman, M.-A., Sundelin, T., Nielsen, J.T., and Erbs, G. (2013). MAMP (microbe-associated molecular pattern) triggered immunity in plants. Front. Plant Sci. 4.

Nurnberger, T., Brunner, F., Kemmerling, B., and Piater, L. (2004). Innate immunity in plants and animals: striking similarities and obvious differences. Immunol. Rev. 198, 249–266.

O

Oerke, E.-C. (2006). Crop losses to pests. J. Agric. Sci. 144, 31.

Old, K. (2002). Assessment of eucalypt rust as a pathogen of *Eucalyptus* spp. and other Myrtaceae, and development of sensitive methods for its detection in germplasm in Australia (FST/1996/206). Adopt. ACIAR Proj. Outputs Stud. Proj. Complet. In 2003.

Olson, N.D., Lund, S.P., Colman, R.E., Foster, J.T., Sahl, J.W., Schupp, J.M., Keim, P., Morrow, J.B., Salit, M.L., and Zook, J.M. (2015). Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front. Genet. 6.

Orwig, D.A. (2002). Ecosystem to regional impacts of introduced pests and pathogens: historical context, questions and issues. J. Biogeogr. 29, 1471–1474.

P

Padmanabhan, M., Cournoyer, P., and Dinesh-Kumar, S.P. (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. Cell. Microbiol. 11, 191–198.

Padovan, A., Keszei, A., Külheim, C., and Foley, W.J. (2014). The evolution of foliar terpene diversity in Myrtaceae. Phytochem. Rev. 13, 695–716.

Palumbi, S.R. (2001). Humans as the World's Greatest Evolutionary Force. Science 293, 1786–1790.

Park, K.-Y., Jung, J.-Y., Park, J., Hwang, J.-U., Kim, Y.-W., Hwang, I., and Lee, Y. (2003). A Role for Phosphatidylinositol 3-Phosphate in Abscisic Acid-Induced Reactive Oxygen Species Generation in Guard Cells. Plant Physiol. 132, 92–98.

Paul Schulze-Lefert, and, and Panstruga, R. (2003). Establishment of biotrophy by parasitic fungi and reprogramming of host cells for disease resistance. Annual Review of Phytopathology 41, 641–667.

Pegg, G.S., Giblin, F.R., McTaggart, A.R., Guymer, G.P., Taylor, H., Ireland, K.B., Shivas, R.G., and Perry, S. (2014a). *Puccinia psidii* in Queensland, Australia: disease symptoms, distribution and impact. Plant Pathol. 63, 1005–1021.

Pegg, G.S., Brawner, J.T., and Lee, D.J. (2014b). Screening Corymbia populations for resistance to Puccinia psidii. Plant Pathol. 63, 425–436.

Perry, G.L.W., and Enright, N.J. (2002). Humans, fire and landscape pattern: understanding a maquis-forest complex, Mont Do, New Caledonia, using a spatial "state-and-transition" model. J. Biogeogr. 29, 1143–1158.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11, 1650–1667.

Piffanelli, P., Zhou, F., Casais, C., Orme, J., Jarosch, B., Schaffrath, U., Collins, N.C., Panstruga, R., and Schulze-Lefert, P. (2002). The Barley mlo modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. Plant Physiol. 129, 1076–1085.

du Plessis, E., McTaggart, A.R., Granados, G., Wingfield, M., Roux, J., Pegg, G.S., Makinson, J., Purcell, M., and Ali, M. (2017). First report of myrtle rust caused by *Austropuccinia psidii* on *Rhodomyrtus tomentosa* (Myrtaceae) from Singapore. Plant Dis.

Poland, J.A., Balint-Kurti, P.J., Wisser, R.J., Pratt, R.C., and Nelson, R.J. (2009). Shades of gray: the world of quantitative disease resistance. Trends Plant Sci. 14, 21–29.

Postel, S., and Kemmerling, B. (2009). Plant systems for recognition of pathogen-associated molecular patterns. Semin. Cell Dev. Biol. 20, 1025–1031.

Potts, B.M., Sandhu, K.S., Wardlaw, T., Freeman, J., Li, H., Tilyard, P., and Park, R.F. (2016). Evolutionary history shapes the susceptibility of an island tree flora to an exotic pathogen. For. Ecol. Manag. 368, 183–193.

Punja, Z.K., and Zhang, Y.-Y. (1993). Plant chitinases and their roles in resistance to fungal diseases. J. Nematol. 25, 526–540.

Q

Quecine, M.C., Leite, T.F., Bini, A.P., Regiani, T., Franceschini, L.M., Budzinski, I.G.F., Marques, F.G., Labate, M.T.V., Guidetti-Gonzalez, S., Moon, D.H., et al. (2016). Label-Free Quantitative proteomic analysis of *Puccinia psidii* uredospores reveals differences of fungal populations infecting Eucalyptus and Guava. PLoS ONE 11, e0145343.

Quesada, T., Gopal, V., Cumbie, W.P., Eckert, A.J., Wegrzyn, J.L., Neale, D.B., Goldfarb, B., Huber, D.A., Casella, G., and Davis, J.M. (2010). Association mapping of quantitative disease resistance in a natural population of Loblolly Pine (*Pinus taeda* L.). Genetics 186, 677–686.

R

Ramachanran, C.K. (1984). Vrikshayurveda (Arboreal Medicine in Ancient India). Anc. Sci. Life 4, 110.

Rao, J., Liu, D., Zhang, N., He, H., Ge, F., and Chen, C. (2013). Identification of genes differentially expressed in a resistant reaction to *Fusarium Oxysporum* in Lilium Regale by SSH. IERI Procedia 5, 95–101.

Rayachhetry, M.B., Elliott, M.L., and Van, T.K. (1997). Natural epiphytotic of the rust *Puccinia psidii* on *Melaleuca quinquenervia* in Florida. Plant Dis. 81, 831–831.

Rayachhetry, M.B., Van, T.K., Center, T.D., and Elliott, M.L. (2001). Host Range of Puccinia psidii, a Potential Biological Control Agent of Melaleuca quinquenervia in Florida. Biol. Control 22, 38–45.

Read, J., Jaffré, T., Godrie, E., Hope, G.S., and Veillon, J.-M. (2000). Structural and floristic characteristics of some monodominant and adjacent mixed rainforests in New Caledonia. J. Biogeogr. 27, 233–250.

Reina-Pinto, J.J., and Yephremov, A. (2009). Surface lipids and plant defenses. Plant Physiol. Biochem. 47, 540–549.

REX Consortium (2016). Combining selective pressures to enhance the durability of disease resistance genes. Front. Plant Sci. 7.

Reynertson, K.A., Yang, H., Jiang, B., Basile, M.J., and Kennelly, E.J. (2008). Quantitative analysis of antiradical phenolic constituents from fourteen edible Myrtaceae fruits. Food Chem. 109, 883–890.

Ribeiro, I.J.A., and Pommer, C.V. (2004). Breeding guava (*Psidium guajava*) for resistance to rust caused by Puccinia psidii. Acta Hortic. 75–78.

Rietz, S., and Parker, J.E. (2007). Plant Disease and Defence. In Encyclopedia of Life Sciences, John Wiley & Sons, Ltd, ed. (Chichester: John Wiley & Sons, Ltd), p.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rodas, C.A., Roux, J., Maier, W., Granados, G.M., Bolaños, M.D., McTaggart, A.R., and Wingfield, M.J. (2015). First report of *Puccinia psidii* on *Corymbia citriodora* and Eucalyptus in Colombia. For. Pathol. 45, 534–536.

Romeis, T., Piedras, P., Zhang, S., Klessig, D.F., Hirt, H., and Jones, J.D.G. (1999). Rapid Avr9- and Cf-9–Dependent Activation of MAP Kinases in Tobacco Cell Cultures and Leaves: Convergence of Resistance Gene, Elicitor, Wound, and Salicylate Responses. Plant Cell Online 11, 273–287.

Römer, P., Recht, S., and Lahaye, T. (2009). A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proc. Natl. Acad. Sci. U. S. A. 106, 20526–20531.

Rosado, T.B., Tomaz, R.S., Junior, R., Fernandes, M., Rosado, A.M., Guimarães, L.M. da S., Araújo, E.F. de, Alfenas, A.C., and Cruz, C.D. (2010). Detection of QTL associated with rust resistance using IBD-based methodologies in exogamic Eucalyptus spp. populations. Crop Breed. Appl. Biotechnol. 10, 321–328.

Ross-Davis, A.L., Graca, R.N., Alfenas, A.C., Peever, T.L., Hanna, J.W., Uchida, J.Y., Hauff, R.D., Kadooka, C.Y., Kim, M.-S., Cannon, P.G., et al. (2014). Tracking the distribution of *Puccinia psidii* genotypes that cause rust disease on diverse myrtaceous trees and shrubs.

Roux, J., Greyling, I., Coutinho, T.A., Verleur, M., and Wingfield, M.J. (2013). The Myrtle rust pathogen, *Puccinia psidii*, discovered in Africa. IMA Fungus 4, 155–159.

Roux, J., Granados, G.M., Shuey, L., Barnes, I., Wingfield, M.J., and McTaggart, A.R. (2016). A unique genotype of the rust pathogen, *Puccinia psidii*, on Myrtaceae in South Africa. Australas. Plant Pathol. 45, 645–652.

Ruiz, R.A.R., Alfenas, A.C., Ferreira, F.A., and Vale, F.X.R. (1989). Influência da temperatura, do tempo de molhamento foliar, fotoperíodo e intensidade de luz sobre a infecção de *Puccinia psidii* em eucalipto. Fitopatol Bras 14, 55–61.

S

Saccone, S.F., Bolze, R., Thomas, P., Quan, J., Mehta, G., Deelman, E., Tischfield, J.A., and Rice, J.P. (2010). SPOT: a web-based tool for using biological databases to prioritize SNPs after a genome-wide association study. Nucleic Acids Res. 38, W201–W209.

Sagi, M.S., Deokar, A.A., and Tar'an, B. (2017). Genetic analysis of NBS-LRR Gene Family in Chickpea and their expression profiles in response to ascochyta blight infection. Front. Plant Sci. 8.

Salzman, R.A., Fujita, T., Zhu-Salzman, K., Hasegawa, P.M., and Bressan, R.A. (1999). An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. Plant Mol. Biol. Report. 17, 11–17.

Sanabria, N.M., Huang, J.-C., and Dubery, I.A. (2010). Self/non-self perception in plants in innate immunity and defense. Self/Nonself 1, 40–54.

Santos, M.R., Guimares, L.M. da S., Resende, M.D.V. de, Rosse, L.N., Zamprogno, K.C., and Alfenas, A.C. (2014). Resistance of *Eucalyptus pellita* to rust (*Puccinia psidii*). Crop Breed. Appl. Biotechnol. 14, 244–250.

Sarah, G., Homa, F., Pointet, S., Contreras, S., Sabot, F., Nabholz, B., Santoni, S., Sauné, L., Ardisson, M., Chantret, N., et al. (2016). A large set of 26 new reference transcriptomes dedicated to comparative population genomics in crops and wild relatives. Mol. Ecol. Resour.

Schiff, C.L., Wilson, I.W., Somerville, S.C. (2001). Polygenic powdery mildew disease resistance in *Arabidopsis thaliana*: quantitative trait analysis of the accession Warschau-1. Plant Pathol. 50, 690–701.

Schmidt, B., and Hildebrandt, A. (2017). Next-generation sequencing: big data meets high performance computing. Drug Discov. Today 22, 712–717.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L.W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol. Ecol. Resour. 11, 1–8.

Selin, C., de Kievit, T.R., Belmonte, M.F., and Fernando, W.G.D. (2016). Elucidating the role of effectors in Plant-Fungal Interactions: Progress and Challenges. Front. Microbiol. 7.

Selosse, M.-A., Baudoin, E., and Vandenkoornhuyse, P. (2004). Symbiotic microorganisms, a key for ecological success and protection of plants. C. R. Biol. 327, 639–648.

Semagn, K., Bjørnstad, Å., and Ndjiondjop, M.N. (2006). An overview of molecular marker methods for plants. Afr. J. Biotechnol. 5.

Serrano, M., Coluccia, F., Torres, M., L'Haridon, F., and Métraux, J.-P. (2014). The cuticle and plant defense to pathogens. Front. Plant Sci. 5.

Shin, S., Kim, K.-H., Kang, C.-S., Cho, K.-M., Park, C. S., Okagaki, R. and Park, J.-C. (2014). A Simple Method for the Assessment of Fusarium Head Blight Resistance in Korean Wheat Seedlings Inoculated with Fusarium graminearum. The Plant Pathology Journal 30, 25–32.

Silva, A.C. da, Andrade, P.M.T. de, Alfenas, A.C., Graça, R.N., Cannon, P., Hauff, R., Ferreira, D.C., and Mori, S. (2014). Virulence and Impact of Brazilian Strains of *Puccinia psidii* on Hawaiian ʻŌhiʻa (*Metrosideros polymorpha*). Pac. Sci. 68, 47–56.

Simpson, J.A., Thomas, K., and Grgurinovic, C.A. (2006). Uredinales species pathogenic on species of Myrtaceae. Australas. Plant Pathol. 35, 549–562.

Singh, D., Singh, P.K., Chaudhary, S., Mehla, K., and Kumar, S. (2012). Exome Sequencing and Advances in Crop Improvement. Adv. Genet. 79, 87–121.

Skinner, M.E., and Holmes, I.H. (2002). Setting Up the JBrowse Genome Browser. In Current Protocols in Bioinformatics, (John Wiley & Sons, Inc.), p.

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., and Holmes, I.H. (2009). JBrowse: A next-generation genome browser. Genome Res. 19, 1630–1638.

Smirnova, O.G., and Kochetov, A.V. (2015). Promoters of plant genes responsive to pathogen invasion. Russ. J. Genet. Appl. Res. 5, 254–261.

Smith, S.E., and Read, D.J. (2008). Mycorrhizal symbiosis (third edition). Academic, New York.

Snow, N. (2009). Kanakomyrtus (Myrtaceae): A New Endemic Genus from New Caledonia with linear stigma lobes and baccate fruits. Syst. Bot.

Sobrosa, R.C., and Martins–Corder, M.P. (2001). Genetic and Biotechnological Breeding for Disease Resistance in Forest Species: A Review. Cropp Breed. Appl. Biotechnol. 1, 385–398.

Soewarto, J., Carriconde, F., Hugot, N., Bocs, S., Hamelin, C., and Maggia, L. (2017). The impact of *Austropuccinia psidii* in New Caledonia, a biodiversity hotspot. For. Pathol. in press.

Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14, 91.

Sonnhammer, E.L.L., and Koonin, E.V. (2002). Orthology, paralogy and proposed classification for paralog subtypes. Trends Genet. TIG 18, 619–620.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stuber, K., Loren van Themaat, E.V., Brown, J.K.M., Butcher, S.A., Gurr, S.J., et al. (2010). Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. Science 330, 1543–1546.

Stalker, J., Gibbins, B., Meidl, P., Smith, J., Spooner, W., Hotz, H.-R., and Cox, A.V. (2004). The Ensembl Web Site: Mechanics of a Genome Browser. Genome Res. 14, 951–955.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., et al. (2002). The Generic Genome Browser: A building block for a model organism system database. Genome Res. 12, 1599–1610.

Stewart, J.E., Ross-Davis, A.L., Graça, R.N., Alfenas, A.C., Peever, T.L., Hanna, J.W., Uchida, J.Y., Hauff, R.D., Kadooka, C.Y., Kim, M.-S., et al. (2017). Genetic diversity of the myrtle rust pathogen (*Austropuccinia psidii*) in the Americas and Hawaii: Global implications for invasive threat assessments. For. Pathol.

Storey, J.D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. Ann. Stat. 31, 2013–2035.

Strickler, S.R., Bombarely, A., and Mueller, L.A. (2012). Designing a transcriptome next-generation sequencing project for a nonmodel plant species. Am. J. Bot. 99, 257–266.

Strobel, N.E., Ji, C., Gopalan, S., Kuc, J.A., and He, S.Y. (1996). Induction of systemic acquired resistance in cucumber by Pseudomonas syringae pv. syringae 61 HrpZPss protein. Plant J. 9, 431–439.

Struck, C., Mueller, E., Martin, H., and Lohaus, G. (2004). The Uromyces fabae UfAAT3 gene encodes a general amino acid permease that prefers uptake of in planta scarce amino acids. Mol. Plant Pathol. 5, 183–189.

Studholme, D.J., Glover, R.H., and Boonham, N. (2011). Application of high-throughput DNA sequencing in phytopathology. Annu. Rev. Phytopathol. 49, 87–105.

Stukkens, Y., Bultreys, A., Grec, S., Trombik, T., Vanham, D., and Boutry, M. (2005). NpPDR1, a Pleiotropic Drug Resistance-Type ATP-Binding Cassette Transporter from *Nicotiana plumbaginifolia*, plays a major role in plant pathogen defense. Plant Physiol. 139, 341–352.

Sudisha, J., Sharathchandra, R.G., Amruthesh, K.N., Kumar, A., Shetty, H.S. (2012). Pathogenesis Related Proteins in Plant Defense Response, in: Plant Defence: Biological Control, Progress in Biological Control. Springer, Dordrecht, pp. 379–403.

Sung, M.-H. (2013). Transcriptional reprogramming, in: Encyclopedia of systems biology. Springer, New York, NY, pp. 2258–2259.

Suprin, B. (2013). Mille et une plantes en Nouvelle-Calédonie écrit par Bernard Suprin - Photosynthèse - Faune et flore - Pacific Book'In - Librairie en ligne Nouméa.

T

Tameling, W.I.L., and Takken, F.L.W. (2008). Resistance proteins: scouts of the plant innate immune system. Eur. J. Plant Pathol. 121, 243–255.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Tan, M.-K., Collins, D., Chen, Z., Englezou, A., and Wilkins, M.R. (2014). A brief overview of the size and composition of the myrtle rust genome and its taxonomic status. Mycology 5, 52–63.

Tavares, S., Ramos, A.P., Pires, A.S., Azinheira, H.G., Caldeirinha, P., Link, T., Abranches, R., Silva, M. do C., Voegele, R.T., Loureiro, J., et al. (2014). Genome size analyses of Pucciniales reveal the largest fungal genomes. Front. Plant Sci. 5.

Tessmann, D.J., and Dianese, J.C. (2002). Hentriacontane: a leaf hydrocarbon from *Syzygium jambos* with stimulatory effects on the germination of urediniospores of Puccinia psidii. Fitopatol. Bras. 27, 538–542.

Thakur, M., and Sohal, B.S. (2013). Role of elicitors in inducing resistance in plants against pathogen infection: A Review.

Thornhill, A.H., Ho, S.Y.W., Külheim, C., and Crisp, M.D. (2015). Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Mol. Phylogenet. Evol. 93, 29–43.

Thumma, B.R., Pegg, G.S., Warburton, P., Macdonell, P., Yang, X., and Southerton, S. (2013). Molecular tagging of rust resistance genes in eucalypts. CSIRO Plant Ind.

Tobias, P.A., Guest, D.I., Külheim, C., Hsieh, J.-F., and Park, R.F. (2015). A curious case of resistance to a new encounter pathogen: Myrtle rust in Australia. Mol. Plant Pathol. n/a-n/a.

Tommerup, I.C., Alfenas, A.C., Old, K.M., and others (2003). Guava rust in Brazil–a threat to Eucalyptus and other Myrtaceae. N. Z. J. For. Sci. 33, 420–428.

Torres, M.A. (2006). Reactive Oxygen Species Signaling in Response to Pathogens. PLANT Physiol. 141, 373–378.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578.

Tsuda, K., and Katagiri, F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. Curr. Opin. Plant Biol. 13, 459–465.

Tucker, S.L., and Talbot, N.J. (2001). Surface attachment and pre-penetration stage development by plant pathogenic fungi. Annu. Rev. Phytopathol. 39, 385–417.

U

Uchida, J., Zhong, S., and Killgore, E. (2006). First report of a rust disease on Ohia caused by *Puccinia psidii* in Hawaii. Plant Dis. 90, 524–524.

Underwood, W., Zhang, S., He, S.Y. (2007) The *Pseudomonas syringae* type III effector tyrosine phosphatase HopAO1 suppresses innate immunity in *Arabidopsis thaliana*. Plant J. Cell Mol. Biol. 52, 658–672.

Underwood, W. (2012). The plant cell wall: a dynamic barrier against pathogen invasion. Front. Plant Sci. 3, 85.

V

Vale, F.X.R.D., Parlevliet, J.E., Zambolim, L. (2001). Concepts in plant disease resistance. Fitopatol. Bras. 26, 577–589.

Van Breusegem, F. (2006). Reactive Oxygen Species in Plant Cell Death. PLANT Physiol. 141, 384–390.

Vasconcelos, T.N.C., Proença, C.E.B., Ahmad, B., Aguilar, D.S., Aguilar, R., Amorim, B.S., Campbell, K., Costa, I.R., De-Carvalho, P.S., Faria, J.E.Q., et al. (2017). Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae. Mol. Phylogenet. Evol. 109, 113–137.

Voegele, R.T., and Mendgen, K. (2003). Rust haustoria: nutrient uptake and beyond. New Phytol. 159, 93–100.

Vogel, J.P., Raab, T.K., Schiff, C., and Somerville, S.C. (2002). PMR6, a pectate lyase-like gene required for powdery mildew susceptibility in Arabidopsis. Plant Cell 14, 2095–2106.

Voigt, C.A. (2014). Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. Front. Plant Sci. 5.

W

Wallwork, H. (2009). The use of Host Plant Resistance in Disease Control. In Disease Control in Crops, essor D. Walters, ed. (Wiley-Blackwell), pp. 122–141.

Wang, F., Zhang, F., Chen, M., Liu, Z., Zhang, Z., Fu, J., and Ma, Y. (2017). Comparative transcriptomics reveals differential gene expression related to colletotrichum gloeosporioides resistance in the octoploid strawberry. Front. Plant Sci. 8.

Wang, Y., Ghaffari, N., Johnson, C.D., Braga-Neto, U.M., Wang, H., Chen, R., and Zhou, H. (2011). Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. BMC Bioinformatics 12 Suppl 10, S5.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63.

Wilson, P.G. (2010). Myrtaceae. In Flowering Plants. Eudicots, (Springer, Berlin, Heidelberg), pp. 212–271.

Wilson, P.G., O'Brien, M.M., Gadek, P.A., and Quinn, C.J. (2001). Myrtaceae revisited: a reassessment of infrafamilial groups. Am. J. Bot. 88, 2013–2025.

Wilson, P.G., O'Brien, M.M., Heslewood, M.M., and Quinn, C.J. (2005). Relationships within Myrtaceae sensu lato based on a matK phylogeny. Plant Syst. Evol. 251, 3–19.

Wright, A.F. (2005). Genetic Variation: Polymorphisms and Mutations. In Encyclopedia of Life Sciences, John Wiley & Sons, Ltd, ed. (Chichester: John Wiley & Sons, Ltd), p.

Wright, K.M., Duncan, G.H., Pradel, K.S., Carr, F., Wood, S., Oparka, K.J., and Cruz, S.S. (2000). Analysis of the N Gene Hypersensitive Response Induced by a Fluorescently Tagged Tobacco Mosaic Virus. Plant Physiol. 123, 1375.

Wulff, A.S., Hollingsworth, P.M., Ahrends, A., Jaffré, T., Veillon, J.-M., L'Huillier, L., and Fogliani, B. (2013). Conservation priorities in a biodiversity hotspot: analysis of narrow endemic plant species in new caledonia: e73371. PLoS One 8.

X

Xavier, A.A., Alfenas, A.C., Matsuoka, K., and Hodges, C.S. (2001). Infection of resistant and susceptible Eucalyptus grandis genotypes by urediniospores of *Puccinia psidii*. Australas. Plant Pathol. 30, 277–281.

Xue, R., Wu, J., Zhu, Z., Wang, L., Wang, X., Wang, S., and Blair, M.W. (2015). Differentially expressed genes in resistant and susceptible common bean (*Phaseolus vulgaris* L.) Genotypes in Response to *Fusarium oxysporum f. sp. phaseoli*. PLoS ONE 10.

Y

Yang, S.-M., Dowler, W.M., and Johnson, D.R. (2009). Comparison of methods for selecting fungi pathogenic to leafy spurge.

Yu, J., Ke, T., Tehrim, S., Sun, F., Liao, B., and Hua, W. (2015). PTGBase: an integrated database to study tandem duplicated genes in plants. Database 2015, bav017-bav017.

Z

Zauza, E.A.V., Couto, M.M.F., Lana, V.M., Maffia, L.A., and Alfenas, A.C. (2010a). Vertical spread of *Puccinia psidii* urediniospores and development of eucalyptus rust at different heights. Australas. Plant Pathol. 39, 141.

Zauza, E.A.V., Alfenas, A.C., Old, K., Couto, M.M.F., Graça, R.N., and Maffia, L.A. (2010b). Myrtaceae species resistance to rust caused by Puccinia psidii. Australas. Plant Pathol. 39, 406–411.

Zeng, P., Zhao, Y., Qian, C., Zhang, L., Zhang, R., Gou, J., Liu, J., Liu, L., and Chen, F. (2015). Statistical analysis for genome-wide association study. J. Biomed. Res. 29, 285–297.

Zeyen, R.J., Carver, T.L.W., and Lyngkjær, M.F. (2002). Epidermal cell papillae. In the Powdery Mildews. A comprehensive treatise, (APS Press), pp. 107–125.

Zhang, Y. (2003). A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1. Plant cell online 15, 2636–2646.

Zhang, L., Kars, I., Essenstam, B., Liebrand, T.W.H., Wagemakers, L., Elberse, J., Tagkalaki, P., Tjoitang, D., van den Ackerveken, G., and van Kan, J.A.L. (2014). Fungal Endopolygalacturonases are recognized as microbe-associated molecular patterns by the Arabidopsis Receptor-Like Protein responsiveness to botrytis polygalacturonases1. PLANT Physiol. 164, 352–364.

Zheng, X.H., Lu, F., Wang, Z.-Y., Zhong, F., Hoover, J., and Mural, R. (2005). Using shared genomic synteny and shared protein functions to enhance the identification of orthologous gene pairs. Bioinforma. Oxf. Engl. 21, 703–710.

Zhong, S., Yang, B., and Alfenas, A.C. (2008). Permanent genetic resources: Development of microsatellite markers for the guava rust fungus, Puccinia psidii. Mol. Ecol. Resour. 8, 348–350.

Zhong, S., Yang, B., and Puri, K.D. (2011). Characterization of *Puccinia psidii* isolates in Hawaii using microsatellite DNA markers. J. Gen. Plant Pathol. JGPP 77, 178–181.

Zhuang, J-Y. & Wei, S-X. Additional materials for the rust flora of Hainan Province, China. Plant Relationships.

Annexes

Annexe 3.1. Du gène à l'ARN

Les changements dans l'expression des gènes sont à la base de l'évolution adaptative de nombreux traits de vie (Harrison et al., 2012). Caractériser les gènes associés aux variations observées entre différents états biologiques, est essentiel dans la compréhension des mécanismes liés à l'expression d'un phénotype d'intérêt tel que la résistance aux maladies. Cette caractérisation peut être effectuée par le biais de l'étude du **transcriptome**. Le transcriptome représente l'ensemble des ARN (Acide Ribonucléique) issus de la transcription de l'ADN à un stade de développement donné ou encore dans des conditions physiologiques particulières (Wang et al., 2011). L'information pour la synthèse des protéines déterminant le phénotype est porté par un type d'ARN particulier, l'ARN messager (ARNm/mRNA) (Figure Annexe 3.1).

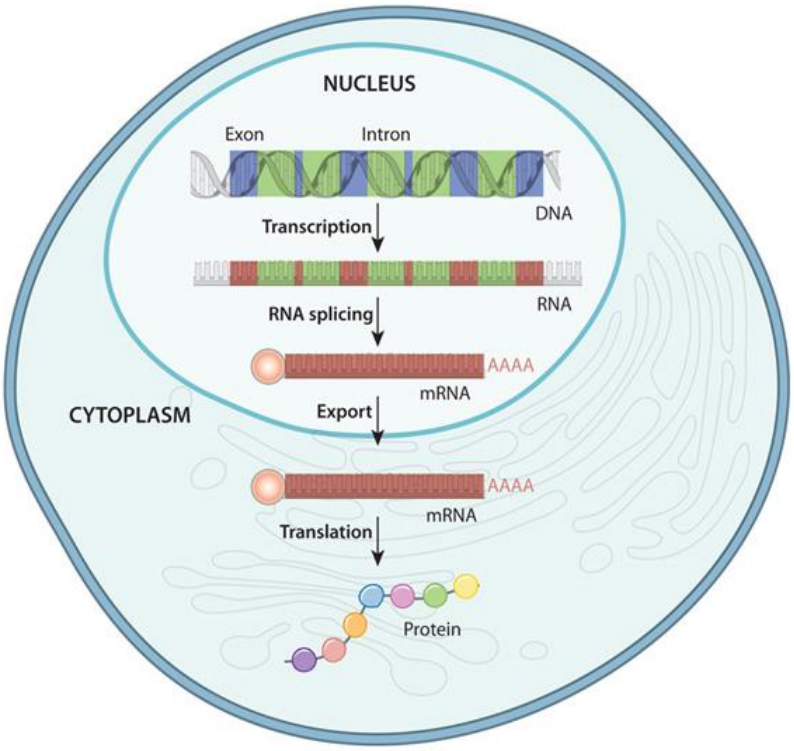
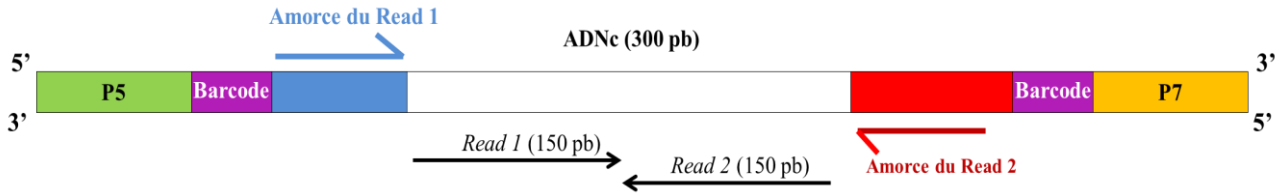


Figure Annexe 1 : Schéma du dogme central de la biologie moléculaire décrivant les mécanismes d'expression des gènes en protéines chez les eucaryotes. © 2010 Nature Education

Le transcriptome est obtenu par "séquençage de l'ARN (ou RNA-SEQ)", réalisé à l'aide de des nouvelles techniques de séquençage haut-débit (NGS) (Wang et al., 2009). Le RNA-Seq

apporte à la fois des informations sur les niveaux d'expression des gènes transcrits, la structure de ces gènes et l'ensemble des isoformes produits (par épissage alternatif ou variants alléliques) (Costa et al., 2010). L'analyse de l'ensemble de ces attributs spécifiques au RNA-Seq constitue l'un des challenges les plus complexes du domaine des NGS (Costa et al., 2010).

Annexe 3.2 : Structure d'une banque d'ADNc pour le séquençage Illumina en paired-end.



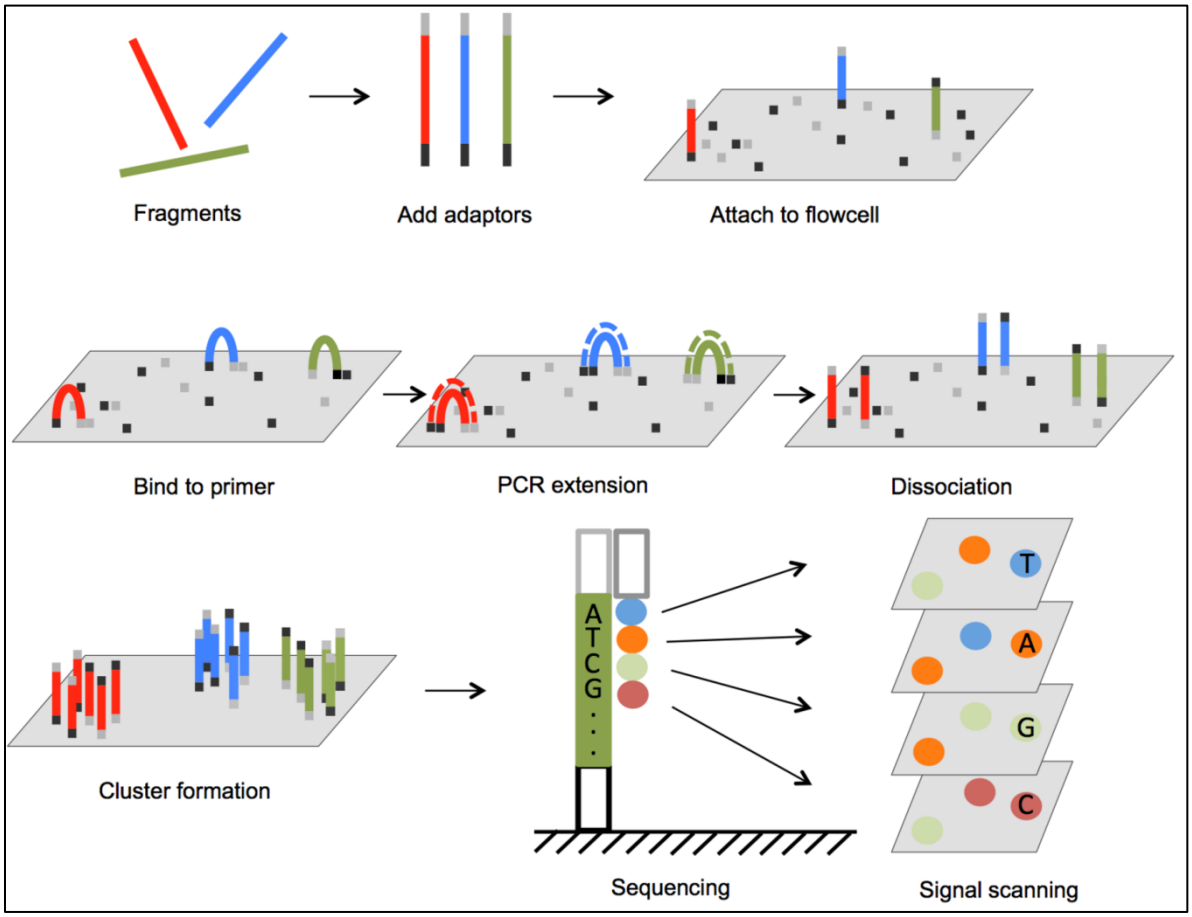
L'adaptateur (P5+amorce du read 1) se lie à l'extrémité 5' du fragment d'ADNc. L'adaptateur (P7+barcode+amorce du read 2) se lie à l'extrémité 3' du fragment d'ADNc.

P5 et P7 servent à lier les fragments d'ADN à la flowcell d'Illumina au moment du séquençage.

Le barcode est un identifiant individuel permettant d'identifier et combiner plusieurs échantillons dans une même expérience de séquençage.

Les amorces Read 1 et Read 2 permettent d'amorcer le séquençage aux deux extrémités des brins d'ADNc.

Annexe 3.3 : Rappel de la technologie de séquençage haut-débit d'Illumina.



(1) Les adaptateurs sont attachés aux extrémités des fragments de séquence.

(2) Les fragments se lient de manière aléatoire sur la flowcell chargée avec des amorces complémentaires aux adaptateurs.

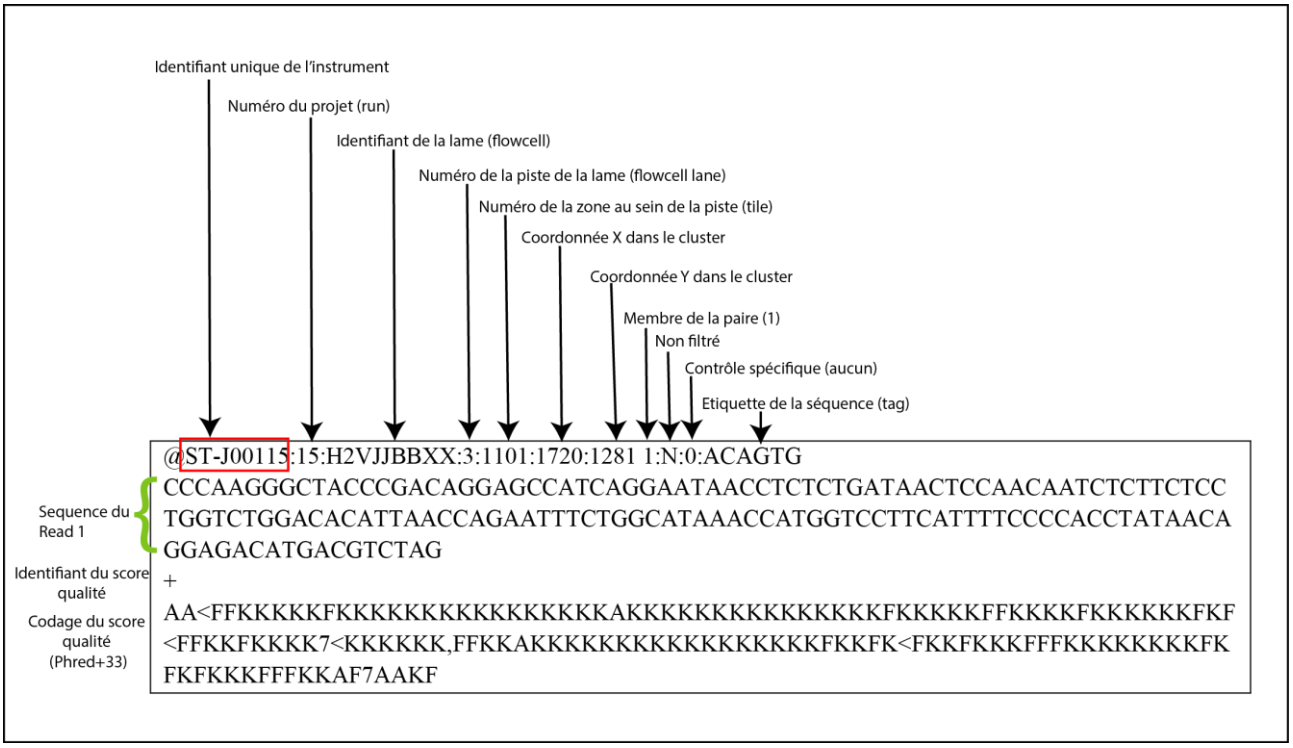
Les réactions "PCR par pont" amplifient chaque fragment attaché à la flow cell pour produire des clusters (regroupement d'ADN clonal).

(3) Une seule base est incorporée à chaque cycle PCR par l'utilisation de nucléotides modifiés ("reversible terminators") et marqués par fluorescence (une couleur par base).

Un cycle individuel de séquençage comprend : (1) l'ajout d'une base, (2) l'acquisition d'image, (3) le clivage chimique du groupement "terminator" et le marquage fluorescent pour le cycle suivant.

Source : http://www.3402bioinformaticsgroup.com/service/

Annexe 3.4. Extrait du descriptif d'un *read* de l'individu Ag19 (*Arillastrum gummiferum*) dans un fichier .FASTQ.



Annexe 4.1 : Annotation fonctionnelle des 5 gènes différentiellement exprimés en communs entre les 3 espèces sur les alignements *de novo*.

Annotation chez l'*E. grandis*		**Différentiel d'expression à partir de l'alignement sur *E. grandis***			***A. gummiferum***			***S. longifolium***			***T. glauca -BDS***		
Eg_LOC	Eg_product	Ag_Eg_logFC	Sl_Eg_logFC	Tg_Eg_logFC	Ag_locus_tag	Ag_product	Ag_logFC	Sl_locus_tag	Sl_product	Sl_logFC	Tg_locus_tag	Tg_product	Tg_logFC
LOC104418805	hemoglobin-2	6,724108284	1,907799412	6,23495902	ARIGU_g752320:94:561:1:+	hemoglobin-2 [Eucalyptus grandis]	6,588285881	SYL_g411340:1:266:2:+	Non-symbiotic hemoglobin 2	3,727009565	TRIGL_g414740:96:563:3:+	hemoglobin-2 [Eucalyptus grandis]	5,883300843
LOC104441780	thaumatin-like protein 1	0	3,557229001	7,618862524	ARIGU_g213530:63:624:3:+	Thaumatin-like protein	4,906914569	SYL_g244860:22:747:1:+	Osmotin	3,688050897	TRIGL_g016370:1:323:3:+	Pathogenesis-related protein R major form	8,311220112
LOC104443781	protein P21	3,601867093	2,226231194	0	ARIGU_g534730:1:713:3:+	protein P21-like [Eucalyptus grandis]	4,232370443	SYL_g103630:1:733:2:+	protein P21-like [Eucalyptus grandis]	1,993706788	TRIGL_g276200:1:314:3:+	Thaumatin-like protein	4,626381092
LOC104445702	hypothetical protein	0	4,047892165	4,957694796	ARIGU_g537930:114:338:3:+	Maturase K	2,706783683	SYL_g192000:250:474:1:+	DNA mismatch repair protein MLH3	4,662607241	TRIGL_g312310:1:230:3:+	Conserved hypothetical protein	5,6393607
LOC104445702	hypothetical protein	0	4,047892165	4,957694796	ARIGU_g537930:114:338:3:+	Maturase K	2,706783683	SYL_g405070:121:345:1:+	Pentatricopeptide repeat-containing protein At4g19191, mitochondrial	3,057471016	TRIGL_g312310:1:230:3:+	Conserved hypothetical protein	5,6393607

Annexe 4.2 : Annotation fonctionnelle des 29 des SNPs discriminants localisés dans des gènes différentiellement exprimés chez *A. gummiferum* (Unified Genotyper). En jaune sont indiqués les SNPs communs avec la méthode In-house script.

Eg_LOC	Eg_product	Ag_Eg_logFC	Ag_locus_tag	Ag_product	Ag_logFC
EUCGR_None	EUCGR_None	0	ARIGU_g036290:1:253:2:+	Ferric reduction oxidase 6,Ferric reduction oxidase 6	-2,8026732
EUCGR_None	EUCGR_None	0	ARIGU_g036290:525:1469:3:+	Ferric reduction oxidase 6,Ferric reduction oxidase 6	-2,8026732
LOC104419257	carbonic anhydrase 2	-3,858885683	ARIGU_g178790:199:1215:1:+	carbonic anhydrase 2-like isoform X1 [Eucalyptus grandis]	-3,099087
LOC104419257	carbonic anhydrase 2	-3,858885683	ARIGU_g178790:199:1215:1:+	carbonic anhydrase 2-like isoform X1 [Eucalyptus grandis]	-3,099087
LOC104424789	isoprene synthase, chloroplastic	0	ARIGU_g628280:1:786:1:+	Isoprene synthase, chloroplastic	-3,7130163
LOC104424789	isoprene synthase, chloroplastic	0	ARIGU_g628280:1:786:1:+	Isoprene synthase, chloroplastic	-3,7130163
LOC104424822	isoprene synthase, chloroplastic	-4,816088701	ARIGU_g628280:1:786:1:+	Isoprene synthase, chloroplastic	-3,7130163
LOC104427201	probable terpene synthase 12	-3,312419001	ARIGU_g628280:1:786:1:+	Isoprene synthase, chloroplastic	-3,7130163
LOC104452881	protein DMR6-LIKE OXYGENASE 2	0	ARIGU_g409320:1:1044:1:+	Putative Protein DMR6-LIKE OXYGENASE 2	-4,2579723
LOC104452898	protein DMR6-LIKE OXYGENASE 2	-3,722500529	ARIGU_g409320:1:1044:1:+	Putative Protein DMR6-LIKE OXYGENASE 2	-4,2579723

LOC104452898	protein DMR6-LIKE OXYGENASE 2	-3,722500529	ARIGU_g409320:1:1044:1:+	Putative Protein DMR6-LIKE OXYGENASE 2	-4,2579723
LOC104452899	protein DMR6-LIKE OXYGENASE 2	-4,091614406	ARIGU_g409320:1:1044:1:+	Putative Protein DMR6-LIKE OXYGENASE 2	-4,2579723
EUCGR_None	EUCGR_None	0	ARIGU_g138750:1:736:2:+	Extensin-3	5,13855279
EUCGR_None	EUCGR_None	0	ARIGU_g333660:921:1082:3:+	Cation/H(+) antiporter 7,(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial	3,57866975
EUCGR_None	EUCGR_None	0	ARIGU_g517120:41:445:2:+	Tubulin beta-3 chain	2,73028992
LOC104415855	histone H3.2	3,770873402	ARIGU_g655070:1:416:3:+	Histone H3.2	4,26552582
LOC104419004	short-chain dehydrogenase reductase 3b	2,803324817	ARIGU_g333660:1:910:2:+	Cation/H(+) antiporter 7,(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial	3,57866975
LOC104419004	short-chain dehydrogenase reductase 3b	2,803324817	ARIGU_g333660:1:910:2:+	Cation/H(+) antiporter 7,(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial	3,57866975
LOC104419023	short-chain dehydrogenase reductase 3b	2,988436911	ARIGU_g333660:1:910:2:+	Cation/H(+) antiporter 7,(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial	3,57866975
LOC104419040	(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial	3,50181235	ARIGU_g333660:1:910:2:+	Cation/H(+) antiporter 7,(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial	3,57866975
LOC104419267	2-methylene-furan-3-one reductase	2,738601977	ARIGU_g703920:1:1081:2:+	2-methylene-furan-3-one reductase	3,05327886
LOC104419268	2-methylene-furan-3-one reductase	2,734437034	ARIGU_g703920:1:1081:2:+	2-methylene-furan-3-one reductase	3,05327886

LOC104436372	pectinesterase-like	3,574701001	ARIGU_g124210:1:1432:2:+	Pectinesterase 3	3,65728359
LOC104436372	pectinesterase-like	3,574701001	ARIGU_g124210:1:1432:2:+	Pectinesterase 3	3,65728359
LOC104436373	pectinesterase-like	3,735241395	ARIGU_g124210:1:1432:2:+	Pectinesterase 3	3,65728359
LOC104439065	anthocyanidin 3-O-glucosyltransferase 7	2,505487508	ARIGU_g231350:1:1550:3:+	anthocyanidin 3-O-glucosyltransferase 7-like [Eucalyptus grandis]	2,84116646
LOC104451753	anthocyanidin 3-O-glucosyltransferase 7	3,142882133	ARIGU_g231350:1:1550:3:+	anthocyanidin 3-O-glucosyltransferase 7-like [Eucalyptus grandis]	2,84116646
LOC104451753	anthocyanidin 3-O-glucosyltransferase 7	3,142882133	ARIGU_g231350:1:1550:3:+	anthocyanidin 3-O-glucosyltransferase 7-like [Eucalyptus grandis]	2,84116646
LOC104451782	anthocyanidin 3-O-glucosyltransferase 7	2,795907991	ARIGU_g231350:1:1550:3:+	anthocyanidin 3-O-glucosyltransferase 7-like [Eucalyptus grandis]	2,84116646

Annexe 4.3 : Annotation fonctionnelle et positions des 3 SNPs discriminants localisés dans des gènes différentiellement exprimés chez *A. gummiferum* (In-house)

Annotation *E. grandis*			*A. gummiferum*			
Eg_LOC	Eg_product	Ag_Eg_logFC	Ag_locus_tag	Ag_product	Ag_logFC	Position-SNP
EUCGR_None	EUCGR_None	0	ARIGU_g517120:41:445:2:+	Tubulin beta-3 chain	2,73028992	471
LOC104419267	2-methylene-furan-3-one reductase	2,738601977	ARIGU_g703920:1:1081:2:+	2-methylene-furan-3-one reductase	3,05327886	703
LOC104419268	2-methylene-furan-3-one reductase	2,734437034	ARIGU_g703920:1:1081:2:+	2-methylene-furan-3-one reductase	3,05327886	703

Annexe 4.4 : Annotation fonctionnelle des 63 SNPs discriminants localisés dans des gènes différentiellement exprimés chez *S. longifolium* (Unified Genotyper)

Eg_LOC	Eg_product	Sl_Eg_logFC	Sl_locus_tag	Sl_product	Sl_logFC
EUCGR_None	EUCGR_None	0	SYL_g417690:1:448:3:+	(+)-neomenthol dehydrogenase	-2,7509299
LOC104449408	alpha-L-fucosidase 1	-2,6180868	SYL_g269220:1:1657:2:+	Alpha-L-fucosidase 1	-2,3468894
LOC104453800	tetraketide alpha-pyrone reductase 2	-2,4166828	SYL_g255820:1:1326:1:+	Tetraketide alpha-pyrone reductase 2	-2,2593768
EUCGR_None	EUCGR_None	0	SYL_g604400:342:518:3:+	Serine/threonine-protein kinase TOR,Fructan 1-exohydrolase	1,87659306
EUCGR_None	EUCGR_None	0	SYL_g103460:49:207:3:-	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558
EUCGR_None	EUCGR_None	0	SYL_g167100:1:247:2:+	Conserved hypothetical protein	1,90642246
EUCGR_None	EUCGR_None	0	SYL_g546780:1:155:3:+	glutathione transferase GST 23-like [Eucalyptus grandis]	1,84373618
EUCGR_None	EUCGR_None	0	SYL_g587120:1:370:1:+	Conserved hypothetical protein	2,25427494
EUCGR_None	EUCGR_None	0	SYL_g604400:1:227:3:+	Serine/threonine-protein kinase TOR,Fructan 1-exohydrolase	1,87659306
EUCGR_None	EUCGR_None	0	SYL_g604400:342:518:3:+	Serine/threonine-protein kinase TOR,Fructan 1-exohydrolase	1,87659306
EUCGR_None	EUCGR_None	0	SYL_g606800	57.9515060640393	1,83011006
EUCGR_None	EUCGR_None	0	SYL_g606830	50.9025085156755	1,799905
LOC104415224	chitinase-like protein 2	1,96355428	SYL_g466010:1:995:3:+	Chitinase-like protein 2	2,01758005
LOC104418455	peroxidase 73	1,94658519	SYL_g376210:144:1208:3:+	Peroxidase 73	2,13880874
LOC104421258	uncharacterized protein At4g22758	1,91414233	SYL_g587120:1:370:1:+	Conserved hypothetical protein	2,25427494
LOC104421273	pleiotropic drug resistance protein 2	2,42914914	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558
LOC104421273	pleiotropic drug resistance protein 2	2,42914914	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558
LOC104421274	pleiotropic drug resistance protein 2	2,52583746	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558
LOC104421275	pleiotropic drug resistance protein 2	2,42670259	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558
LOC104421778	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	0	SYL_g361300:1:920:3:+	Wall-associated receptor kinase-like 14	1,86189319
LOC104421778	LEAF RUST 10 DISEASE-RESISTANCE	0	SYL_g361300:1:920:3:+	Wall-associated receptor kinase-like 14	1,86189319

	LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2				
LOC104421801	MLO-like protein 2	2,13670511	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916
LOC104421801	MLO-like protein 2	2,13670511	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916
LOC104421801	MLO-like protein 2	2,13670511	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916
LOC104422736	oxidoreductase-like domain-containing protein 1	0	SYL_g431040:1:490:2:+	oxidoreductase-like domain-containing protein 1 [Eucalyptus grandis]	1,79273219
LOC104422779	alpha-mannosidase	0	SYL_g550100:58:3090:1:+	lysosomal alpha-mannosidase-like [Eucalyptus grandis]	1,9469498
LOC104423985	alpha-mannosidase	1,97410773	SYL_g550100:58:3090:1:+	lysosomal alpha-mannosidase-like [Eucalyptus grandis]	1,9469498
LOC104423985	alpha-mannosidase	1,97410773	SYL_g550100:58:3090:1:+	lysosomal alpha-mannosidase-like [Eucalyptus grandis]	1,9469498
LOC104425384	glutathione transferase GST 23	2,1614737	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852
LOC104425384	glutathione transferase GST 23	2,1614737	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852
LOC104425384	glutathione transferase GST 23	2,1614737	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852
LOC104425385	glutathione transferase GST 23	2,29717633	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852
LOC104425386	glutathione transferase GST 23	2,57046472	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852
LOC104425388	glutathione transferase GST 23	2,44307067	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852
LOC104428749	dirigent protein 22	1,83219427	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766
LOC104428749	dirigent protein 22	1,83219427	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766
LOC104433195	dnaJ homolog subfamily B member 3	0	SYL_g297010:1:981:1:+	Chaperone protein dnaJ 1, mitochondrial	1,73343439
LOC104433195	dnaJ homolog subfamily B member 3	0	SYL_g297010:1:981:1:+	Chaperone protein dnaJ 1, mitochondrial	1,73343439
LOC104437477	dirigent protein 22	1,68058869	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766
LOC104437530	dirigent protein 22-like	1,85650885	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766
LOC104443781	protein P21	2,22623119	SYL_g103630:1:733:2:+	protein P21-like [Eucalyptus grandis]	1,99370679
LOC104443783	protein P21	1,94918949	SYL_g103630:1:733:2:+	protein P21-like [Eucalyptus grandis]	1,99370679
LOC104443850	heavy metal-associated isoprenylated plant protein 3	2,35244837	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443850	heavy metal-associated isoprenylated plant protein 3	2,35244837	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443850	heavy metal-associated isoprenylated plant protein 3	2,35244837	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443852	heavy metal-associated isoprenylated plant	2,05498103	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782

	protein 3				
LOC104443852	heavy metal-associated isoprenylated plant protein 3	2,05498103	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443852	heavy metal-associated isoprenylated plant protein 3	2,05498103	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443853	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443855	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443856	heavy metal-associated isoprenylated plant protein 3	2,43868125	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443856	heavy metal-associated isoprenylated plant protein 3	2,43868125	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443921	uncharacterized LOC104443921	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443922	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443923	hypothetical protein	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443924	hypothetical protein	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104443924	hypothetical protein	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104446036	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104446073	heavy metal-associated isoprenylated plant protein 3-like	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104446074	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782
LOC104446141	dirigent protein 22	0	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766
LOC104450341	7-deoxyloganetin glucosyltransferase	2,10908661	SYL_g641430:1:1516:2:+	7-deoxyloganetin glucosyltransferase	2,19120046
LOC104455835	MLO-like protein 2	0	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916

Annexe 4.5 : Annotation fonctionnelle et positions des 68 SNPs discriminants localisés dans des gènes différentiellement exprimés chez *S. longifolium* (In-house). En jaune sont indiqués les SNPs communs avec la méthode GATK.

Annotation *E. grandis* ***S. longifolium***

Eg_LOC	Eg_product	Sl_Eg_logFC	Sl_locus_tag	Sl_product	Sl_logFC	SNP-position
EUCGR_None	EUCGR_None	0	SYL_g417690:1:448:3:+	(+)-neomenthol dehydrogenase	-2,7509299	418
LOC104449408	alpha-L-fucosidase 1	-2,6180868	SYL_g269220:1:1657:2:+	Alpha-L-fucosidase 1	-2,3468894	1434
LOC104453800	tetraketide alpha-pyrone reductase 2	-2,4166828	SYL_g255820:1:1326:1:+	Tetraketide alpha-pyrone reductase 2	-2,2593768	1329
EUCGR_None	EUCGR_None	0	SYL_g617630:95:907:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1,7185668	1069
EUCGR_None	EUCGR_None	0	SYL_g617630:95:907:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1,7185668	1069
EUCGR_None	EUCGR_None	0	SYL_g109760:80:774:2:+	Conserved hypothetical protein	1,78462656	609
EUCGR_None	EUCGR_None	0	SYL_g109760:80:774:2:+	Conserved hypothetical protein	1,78462656	609
EUCGR_None	EUCGR_None	0	SYL_g103460:49:207:3:-	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558	3149
EUCGR_None	EUCGR_None	0	SYL_g167100:1:247:2:+	Conserved hypothetical protein	1,90642246	424
EUCGR_None	EUCGR_None	0	SYL_g455170	21.0754598658091	2,20710709	280
EUCGR_None	EUCGR_None	0	SYL_g538880:578:692:2:+	TSL-kinase interacting protein 1	1,74772861	313
LOC104421273	pleiotropic drug resistance protein 2	2,42914914	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558	3149
LOC104421273	pleiotropic drug resistance protein 2	2,42914914	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558	3149
LOC104421274	pleiotropic drug resistance protein 2	2,52583746	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558	3149
LOC104421275	pleiotropic drug resistance protein 2	2,42670259	SYL_g103460:558:4925:3:+	OSJNBa0081G05.10 protein,pleiotropic drug resistance protein 2-like [Eucalyptus grandis]	2,43505558	3149
LOC104421778	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	0	SYL_g361300:1:920:3:+	Wall-associated receptor kinase-like 14	1,86189319	567
LOC104421778	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.2	0	SYL_g361300:1:920:3:+	Wall-associated receptor kinase-like 14	1,86189319	567
LOC104421801	MLO-like protein 2	2,13670511	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916	393
LOC104421801	MLO-like protein 2	2,13670511	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916	393
LOC104421801	MLO-like protein 2	2,13670511	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916	393

LOC104422736	oxidoreductase-like domain-containing protein 1	0	SYL_g431040:1:490:2:+	oxidoreductase-like domain-containing protein 1 [Eucalyptus grandis]	1,79273219	475
LOC104422779	alpha-mannosidase	0	SYL_g550100:58:3090:1:+	lysosomal alpha-mannosidase-like [Eucalyptus grandis]	1,9469498	1354
LOC104423985	alpha-mannosidase	1,97410773	SYL_g550100:58:3090:1:+	lysosomal alpha-mannosidase-like [Eucalyptus grandis]	1,9469498	1354
LOC104423985	alpha-mannosidase	1,97410773	SYL_g550100:58:3090:1:+	lysosomal alpha-mannosidase-like [Eucalyptus grandis]	1,9469498	1354
LOC104425384	glutathione transferase GST 23	2,1614737	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852	889
LOC104425384	glutathione transferase GST 23	2,1614737	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852	889
LOC104425384	glutathione transferase GST 23	2,1614737	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852	889
LOC104425385	glutathione transferase GST 23	2,29717633	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852	889
LOC104425386	glutathione transferase GST 23	2,57046472	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852	889
LOC104425388	glutathione transferase GST 23	2,44307067	SYL_g546790:59:748:2:+	Glutathione transferase GST 23	2,30229852	889
LOC104426303	uncharacterized LOC104426303	0	SYL_g617630:95:907:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1,7185668	1069
LOC104426303	uncharacterized LOC104426303	0	SYL_g617630:95:907:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1,7185668	1069
LOC104428749	dirigent protein 22	1,83219427	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766	550
LOC104428749	dirigent protein 22	1,83219427	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766	550
LOC104434964	unknown_product	0	SYL_g617630:95:907:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1,7185668	1069
LOC104434964	unknown_product	0	SYL_g617630:95:907:2:+	TMV resistance protein N-like [Eucalyptus grandis]	1,7185668	1069
LOC104437477	dirigent protein 22	1,68058869	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766	550
LOC104437530	dirigent protein 22-like	1,85650885	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766	550
LOC104443850	heavy metal-associated isoprenylated plant protein 3	2,35244837	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443850	heavy metal-associated isoprenylated plant protein 3	2,35244837	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443850	heavy metal-associated isoprenylated plant protein 3	2,35244837	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443852	heavy metal-associated isoprenylated plant protein 3	2,05498103	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443852	heavy metal-associated isoprenylated plant protein 3	2,05498103	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443852	heavy metal-associated isoprenylated plant protein 3	2,05498103	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443853	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443855	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443856	heavy metal-associated isoprenylated plant protein 3	2,43868125	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696

LOC104443856	heavy metal-associated isoprenylated plant protein 3	2,43868125	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443921	uncharacterized LOC104443921	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443922	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443923	hypothetical protein	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443924	hypothetical protein	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104443924	hypothetical protein	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104445248	MACPF domain-containing protein NSL1	1,85540191	SYL_g539150:254:2071:2:+	MACPF domain-containing protein NSL1	1,95213104	403
LOC104446036	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104446073	heavy metal-associated isoprenylated plant protein 3-like	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104446074	heavy metal-associated isoprenylated plant protein 3	0	SYL_g390740:257:628:2:+	Conserved hypothetical protein	1,71351782	696
LOC104446081	uncharacterized LOC104446081	1,76849706	SYL_g109760:80:774:2:+	Conserved hypothetical protein	1,78462656	609
LOC104446081	uncharacterized LOC104446081	1,76849706	SYL_g109760:80:774:2:+	Conserved hypothetical protein	1,78462656	609
LOC104446082	uncharacterized LOC104446082	0	SYL_g109760:80:774:2:+	Conserved hypothetical protein	1,78462656	609
LOC104446083	uncharacterized LOC104446083	0	SYL_g109760:80:774:2:+	Conserved hypothetical protein	1,78462656	609
LOC104446084	uncharacterized LOC104446084	2,96820005	SYL_g109760:80:774:2:+	Conserved hypothetical protein	1,78462656	609
LOC104446141	dirigent protein 22	0	SYL_g312850:1:650:3:+	Dirigent protein 2	2,1235766	550
LOC104450341	7-deoxyloganetin glucosyltransferase	2,10908661	SYL_g641430:1:1516:2:+	7-deoxyloganetin glucosyltransferase	2,19120046	510
LOC104455835	MLO-like protein 2	0	SYL_g283480:1:1901:3:+	MLO-like protein 12	2,21098916	393
LOC104456745	protein SENESCENCE-ASSOCIATED GENE 21, mitochondrial	1,98719617	SYL_g447370:1:466:2:+	Indole-3-acetic acid-induced protein ARG2	2,28009025	532
LOC104456745	protein SENESCENCE-ASSOCIATED GENE 21, mitochondrial	1,98719617	SYL_g447370:1:466:2:+	Indole-3-acetic acid-induced protein ARG2	2,28009025	532
LOC104456746	protein SENESCENCE-ASSOCIATED GENE 21, mitochondrial	2,18682422	SYL_g447370:1:466:2:+	Indole-3-acetic acid-induced protein ARG2	2,28009025	532

Annexe 4.6 : Annotation fonctionnelle du gène différentiellement exprimé et contenant un SNP discriminant chez *T. glauca*-BDS (Unified Genotyper)

Eg_LOC	Eg_product	Tg_Eg_logFC	Tg_locus_tag	Tg_product	Tg_logFC
EUCGR_None	EUCGR_None	0	TRIGL_g068700:1:561:1:+	Major allergen Pru ar 1	5,34161705