



HAL
open science

Génétique écologique et génomique des événements de divergence chez les complexes d'espèces en forêt tropicale humide

Alexandra Tinaut

► **To cite this version:**

Alexandra Tinaut. Génétique écologique et génomique des événements de divergence chez les complexes d'espèces en forêt tropicale humide. Génétique des plantes. Université de Guyane, 2015. Français. NNT : 2015YANE0005 . tel-01405400

HAL Id: tel-01405400

<https://theses.hal.science/tel-01405400>

Submitted on 19 Dec 2016

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.



Université de la Guyane

Ecole doctorale Diversités, santé et développement en Amazonie

Thèse Pour obtenir le grade de Docteur de l'Université de Guyane

En Génétique

Par

Alexandra TINAUT

Génétique écologique et génomique des évènements de divergence chez les complexes d'espèces en forêt tropicale humide

Sous la direction d'Ivan SCOTTI

et

Caroline SCOTTI-SAINTAGNE

Soutenue le 17 Décembre 2015, à Kourou

Jury

Giorgio BINELLI	Rapporteur
Philippe LASHERMES	Rapporteur
Christophe DUPLAIS	Examineur

Remerciements

Je tiens en premier lieu à remercier mes directeurs de thèse Ivan SCOTTI et Caroline SCOTTI-SANTAGNE pour cette opportunité qui m'a été donnée. Je voudrais vous remercier pour votre ouverture d'esprit et votre confiance en moi. Cela m'a permis de vraiment faire de cette thèse mon projet. A la limite du raisonnable, j'ai pu partir dans toute les directions que je voulais avec votre soutien.

A Ivan particulièrement, je tiens à te remercier pour tous les cours informels que tu as pu me donner en génétique des populations. Tu m'as formé durant quatre ans, dans la bonne humeur et sans jugement. Ce sont des qualités que je respecte particulièrement.

A Caro, merci pour ta disponibilité, ta sympathie, ta bonne humeur et pour toute l'aide que tu as pu m'apporter dans mes réflexions et dans mes analyses.

Je tiens à présent à remercier les membres du jury de cette thèse, Giorgio BINELLI, Philippe LASHERMES et Christophe DUPLAIS, pour l'intérêt qu'ils portent à ces travaux de thèse. Je vous remercie de bien vouloir juger ces travaux.

Je tiens également à remercier Eric Marcon pour m'avoir accueilli au sein de l'UMR EcoFoG, et pour tous les moyens et bourses qui ont été mis à ma disposition, et qui ont permis le bon déroulement de ma thèse.

Je souhaite maintenant remercier les membres du laboratoire de génétique. Tout d'abord, Eliane et Valérie, je vous remercie pour l'aide, le soutien et les conseils que vous avez pu me donner durant ce séjour que j'ai fait parmi vous et qui vous a quelques fois semblé très long. Henri, je te remercie également pour tous les gâteaux que tu m'as fait partager avec mes collègues. C'était difficile de me lancer pour les faire, mais un réel plaisir de les partager avec vous ! Plus sérieusement, merci pour tes conseils, le partage de tes connaissances et expériences, et merci pour la joie que tu peux mettre au labo. Monsieur Saint-Omer, merci de m'avoir toujours accompagné sur le terrain, assurant ma sécurité et partageant ton humour et ta bonne humeur. Je crois que je t'ai pas mal fait galérer ! Je te remercie pour ta disponibilité et ta gentillesse. Niklas, tu es un nouveau venu dans la « team », mon nouveau chef ! En peu de temps, tu t'es installé et a réussi à contribuer pour beaucoup dans mon projet. Je te remercie pour cet intérêt. Je tiens également à vous remercier, Hadrien et Marina, pour votre bonne humeur, votre simplicité et votre humour.

Mister CAZAL, Jocelyn, je tiens à te remercier pour ta disponibilité pour les sorties de terrains, mais aussi pour ta simplicité et ta joie de vivre ! Steph, je te remercie également pour la majorité de tes conseils avisés que je n'ai pas suivi. Je te remercie pour ton amitié.

Je remercie le personnel administratif. Josie, un grand merci pour tous ces conseils administratifs et personnels que tu m'as donné ! Je te remercie pour ces moments où tu m'as administrativement sauvée ! Je remercie également Annie, Catherine, Alain et Georges-Henry pour leur disponibilité.

Je tiens à remercier toutes les personnes que j'ai rencontrées durant ce travail de thèse : chercheurs, techniciens et doctorants. Vous avez participé chacun d'une façon ou d'une autre au bon déroulement de ce projet.

Je vous remercie les « Yaourt nature » pour tous les divertissements, casse-têtes et blagues qu'on a partagés et qu'on partagera. Colette, Marie-florence, Carole, vous avez très souvent allégé ma journée, merci !

Stelly MISSISSIPI !!! Merci pour toutes nos discussions avec ou sans sens, nos réflexions avec ou sans sens !.. Bref, merci !

Karlène, merci pour nos apéros sans fin ! Tu ne te doutes pas de la puissance avec laquelle ils m'ont aidé à accomplir ce projet ! Merci miss !!

Merci à ma famille dans toute sa globalité, tantes, oncles, cousins, cousines, ma grand-mère, merci pour votre soutien. Nous avons une grande famille et une belle famille, et je tiens à elle. Merci !!

Merci mes sœurs pour vos délires et vos paroles quelques fois sensées ! Ludwina, Jennifer, Jane-esther, merci !

Je laissais le meilleur pour la fin... merci Mamounette pour tout ce que tu m'as donné. Je compte bien continuer à te le rendre ! Juste merci d'être ma mère et merci d'être toujours là. C'est sans prétention que je me dis que si derrière chaque grand homme, il y a une femme, il est clair que derrière une femme, il y a une mère !!! Merci maman !!!

Résumé

Connaitre et appréhender les mécanismes de la diversification des espèces est important pour la gestion des écosystèmes, la prévision des impacts des changements climatiques et la compréhension de la biodiversité actuelle et passée. Le but de cette thèse est de comprendre et de déceler les mécanismes génétiques à l'origine de la diversification des espèces en présence de flux de gènes. Cette thèse se focalise sur le modèle biologique *Symphonia globulifera*, qui compte deux écotypes : le *S.globulifera* spécialiste de la *terra firme* et le *S.sp1*, spécialiste des bas-fonds. Ces deux écotypes montrent une faible différenciation génétique, malgré la présence de deux phénotypes différents.

Une première étape a été de mettre en évidence la présence d'une adaptation locale au sein de cette espèce, par le biais d'une expérimentation de jardins de transplantation réciproques, permettant d'expliquer la répartition des écotypes dans leur habitat naturel. Ensuite, dans le but d'identifier les mécanismes sous-jacent à cette adaptation locale, j'ai testé l'hypothèse que la méthylation des gènes pourrait être une marque de l'épigénétique contribuant à la divergence des écotypes, par l'utilisation d'enzyme sensibles à la méthylation dans un protocole de génotypage AFLP. Enfin, un séquençage haut-débit du transcriptome des écotypes en jardins de transplantation réciproques m'a permis de mettre en évidence une expression différentielle des gènes entre les écotypes, qui pourrait expliquer les différences phénotypiques observées entre les écotypes malgré une faible différenciation génétique. Ces travaux de thèse s'appuie, ainsi, sur des données de traits phénotypiques, de génotypage AFLP et de séquençage à haut débit du transcriptome pour montrer la valeur importante de la régulation des gènes dans la divergence des écotypes adaptés localement, et un faible rôle de la méthylation de l'ADN dans l'établissement cette adaptation locale.

Abstract

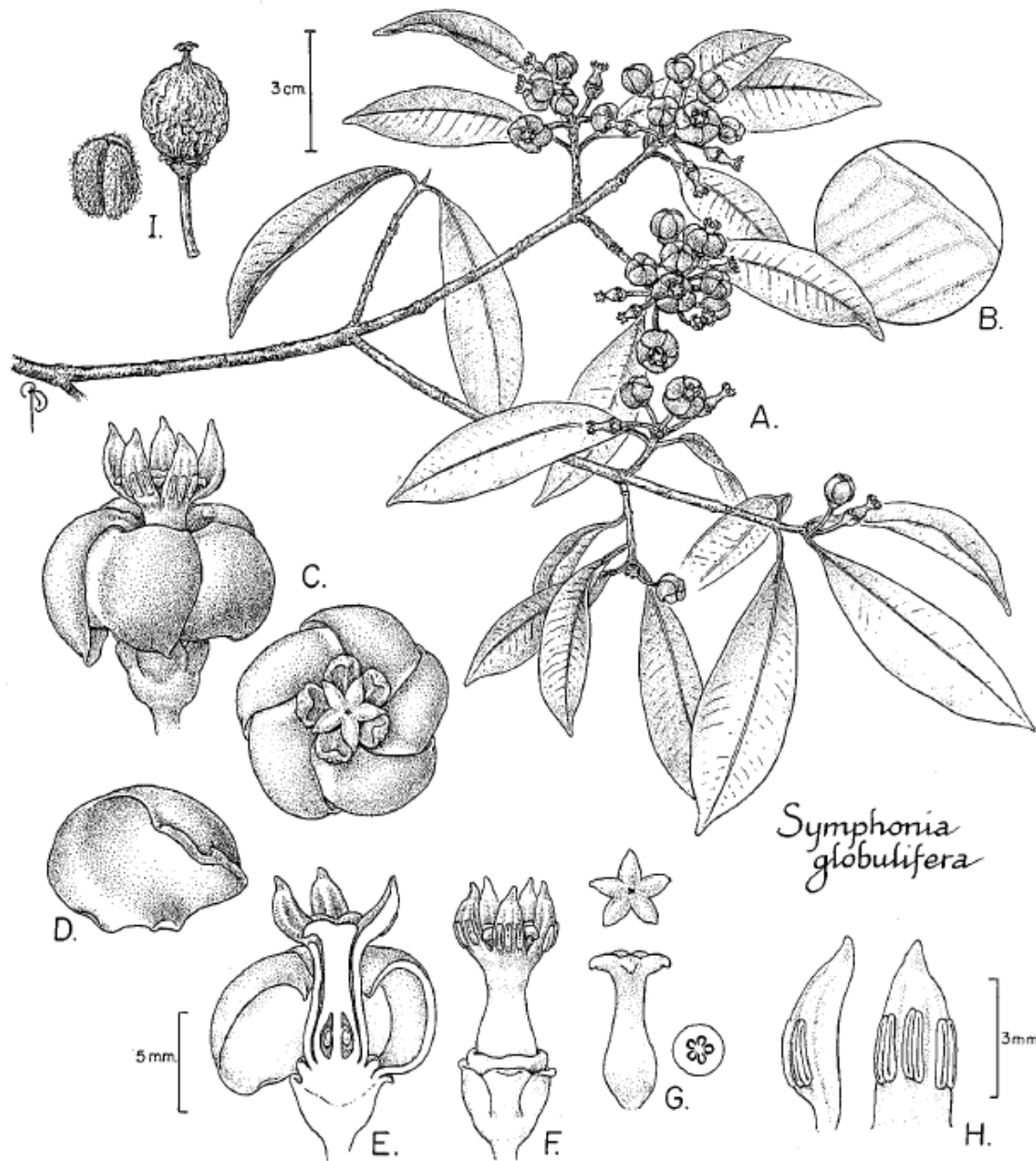
Understanding of the mechanisms driving diversification of species is a significant way to improve the management of ecosystems, predict the impacts of climate change and understand the actual and past biodiversity level. The aim of this thesis is to understand and comprehend genetic mechanisms behind the diversification of species in the presence of gene flow. This thesis is focused on the biological model *Symphonia globulifera*, which presents two ecotypes: the *S.globulifera*, specialist of seasonally flooded lowlands and S.sp1, specialist of *terra firme*. These two ecotypes show low genetic differentiation, despite the presence of two apparent phenotypes.

A first part of this thesis was to test the presence of local adaptation of this using reciprocal transplant experiment gardens, allowing the understanding of the ecotypes distributions in their natural habitats. Then, this local adaptation in the presence of gene flow, directed me to the regulation of gene methylation in order to see the role this brand of epigenetics can have in the divergence of the ecotypes. In a third part of the thesis, new generation sequencing of the transcriptome ecotypes in reciprocal gardens transplantations allowed me to show the evidence of gene regulation to differentiate the ecotypes. This work thesis is based on phenotype records data, AFLP genotyping and high-throughput sequencing of the transcriptome, in order to show the important value of gene regulation in the divergence of the locally adapted ecotypes, and a weak role of DNA methylation in the establishment of local adaptation.

INTRODUCTION

Contexte	14
Le Symphonia globulifera.....	14
Description de l'espèce et utilisation.....	14
Le S.sp1. et les différentes populations du Symphonia globulifera	16
La biodiversité dans les forêts tropicales.....	16
Hypothèse du berceau et théorie de la reine rouge.....	18
Hypothèse du musée	18
Hypothèse du refuge	18
La Guyane et son environnement hétérogène.....	19
Micro-habitat : bas-fonds et terra firme	19
Variations régionales.....	19
.....	21
<i>Origine, maintien et Évolution de la diversité génétique et phénotypique</i>	22
Origine de la diversité génétique.....	22
Le polymorphisme	22
Les mutations	22
La migration.....	23
La dérive génétique.....	23
La sélection	24
Interaction Génotype-Environnement	27
La plasticité phénotypique	27
La plasticité phénotypique adaptative	27
Plasticité phénotypique adaptative et expression des gènes.....	27
<i>Détection de la divergence et outils moléculaires</i>	29
Divergence et fréquences alléliques.....	29
Les marqueurs moléculaires.....	29
Fréquences alléliques et test de sélection	29
Expressions différentielles et techniques	31
Observation de l'expression différentielle	31
Objectifs.....	31
Le RNAseq.....	31
Techniques de détection et tests statistiques	31
<i>Epigénétique et expression des gènes</i>	34
La méthylation de l'ADN.....	34
Epigénétique et évolution.....	34
<i>Hypothèses et expérimentations</i>	35
Objectifs et hypothèses.....	35
Guyane française : 3 sites d'étude.....	36
Les sites de Laussat et de Régina	36
Le site de Paracou	36

***Bibliographie*..... 37**



George E. et al., 1998

Figure 1 . *Symphonia globulifera*

A. Tige avec ses feuilles et ses inflorescences. B. Détail des nervures des feuilles. C. Vue latérale (en haut) et vue du dessus de la fleur. D. le pétale. E. coupe transversale de la fleur avec le tube staminal. F. fleur sans pétale avec le tube staminal, les anthères et les lobes du stigmate. G. Le pistil, vue apicale du stigmate (en haut) et coupe médiane de l'ovaire. H. Vue latérale (gauche) et vue abaxiale du lobes du tube staminal avec des anthères adnées. I. le fruit (à droite) et la graine (à gauche).

CONTEXTE

LE SYMPHONIA GLOBULIFERA

DESCRIPTION DE L'ESPECE ET UTILISATION

De la famille des Clusiaceae (Guttiferae), le *Symphonia globulifera* (Figure 1.) est appelé « manil marécage » en Guyane, « palétuvier jaune » aux Antilles et « anani » au Brésil ; il peut mesurer jusqu'à 40m de haut et 90cm de diamètre. Le genre *Symphonia* est originaire de Madagascar où on peut rencontrer 16 espèces différentes[1]. Selon les études de datation, à l'aide de fossile de *Pachydermites diderexi*, le *Symphonia globulifera* est apparu au Nigeria au milieu de l'Eocène (~45 Ma) [2], et s'est dispersé par voie marine dans les Néotropiques[3], il a 10 millions d'années, durant le Pliocène[4] (Figure 2.). Présent en Amérique centrale, dans les Caraïbes et en Amérique du Sud, le *Symphonia globulifera* est la seule espèce du genre *Symphonia* à évoluer hors de Madagascar.

Cette espèce est majoritairement allogame (90%)[2]. Généralement rencontrée en bas-fonds, elle possède des racines aériennes permettant des échanges gazeux, les pneumatophores. Le *S. globulifera* est caractérisé par ses fleurs globuleuses et rouges (1-1.2cm), des fruits peu charnus possédant une à deux graines globuleuses ou ovoïdes (3-4 cm) (Figure 1.), des feuilles opposées et enfin par la présence de latex jaune qui s'oxyde et devient foncé quand il est en contact avec l'air. Bien que souvent visité par les chauves-souris et les singes[5], la pollinisation du *Symphonia globulifera* se fait par les colibris en Guyane, et par les lépidoptères au Costa Rica[6].

Au Brésil et au Suriname, le *Symphonia globulifera* est utilisé pour la production de contre-plaqué et pour les propriétés médicinales du latex[7], qui s'avère posséder des activités antipaludiques, antimicrobiennes, antileishmaniales et antioxydantes[8]

Box 1. L'espèce biologique et l'écotype

Tiré du latin « species », signifiant « sorte, catégorie », le mot espèce est aujourd'hui dans le monde scientifique un concept donnant un ancrage à la description de la diversité des organismes présents sur terre. La définition de l'espèce a très longtemps été débattue par les naturalistes qui avaient besoin d'un concept qui puisse se généraliser à tous les organismes. Ceci a donné lieu à diverses définitions de l'« espèce » et de la formation des espèces, la spéciation. La première définition de l'espèce s'est basée sur la morphologie, ayant pour principe que des populations proches morphologiquement appartiennent à la même espèce. C'est d'ailleurs de ce même principe que sont nés les domaines de la taxonomie et de la systématique, permettant à des naturalistes comme Carl Von Linné de découvrir, de classer et de nommer des espèces végétales.

L'espèce biologique

La notion d'espèce biologique a été définie par Ernest Mayr en 1940[9] comme étant un ensemble de populations évoluant dans les mêmes sites géographiques et écologiques et qui, partout où elles sont en contact, s'hybrident ou sont potentiellement capables de le faire, dans le cas où il y aurait une barrière écologique ou géographique. Ainsi, l'espèce est définie par la présence de flux de gènes entre les populations et des individus de la même espèce pouvant se reproduire, donnant lieu à des descendance fertiles ayant des caractéristiques morphologiques et génétiques communes.

L'écotype

En 1922, le botaniste Göte Wilhelm Turesson utilise pour la première fois le mot écotype[10], et le définit comme étant une unité écologique résultant de la réponse du génotype d'une écoespèce à un habitat particulier, faisant de l'écotype la sous-unité écologique de l'écoespèce. L'écoespèce représentait alors l'espèce selon la classification de Linné. Ainsi, l'écotype est une sous-unité de l'espèce évoluant dans un habitat particulier.

LE S.SP1. ET LES DIFFERENTES POPULATIONS DU SYMPHONIA GLOBULIFERA

En Guyane et au Suriname, le *Symphonia globulifera* montre un écotype différent du *S.globulifera*, le *S.sp1*, qui est généralement rencontré en *terra firme*[11].

Aussi, il a déjà été montré un décalage de floraison entre ces deux écotypes[12] malgré une faible différenciation génétique[2] ($F_{ST}= 0.03$, Casalis, *non publié*). Au Costa Rica, le *Symphonia globulifera* est un arbre de 15m de haut que l'on rencontre dans les montagnes, et ne montre qu'un seul morphotype[6] ; pourtant, en Afrique et en Amérique du Sud, le *Symphonia globulifera* est décrit comme un arbre de 40 à 50 mètre de haut qui présente ces deux écotypes, spécialistes des bas-fonds et de la *terra firme*[2, 12, 13].

La densité des populations d'arbres de *Symphonia globulifera*, d'un diamètre de plus de 10cm, est variable en Amérique du sud. Les populations ont une forte densité de 104 arbres par hectare au Guyana[14], contre 10.9 arbres par hectare en Guyane[2] et 0.4 arbres par hectare au Panama (Center for Tropical Forest Sciences, recensement de 2005). Il semblerait qu'un seul ou plusieurs événements de dispersions marines aient permis l'établissement du *S. globulifera* dans les Néotropiques, avec des populations de l'Amérique centrale très différenciées par rapport aux populations d'Amérique du sud ($F_{ST}=0.14$)[4]. Enfin, en Guyane, il y a une différenciation forte entre les populations de l'Est et les populations de l'Ouest ($F_{ST}=0.26$, Casalis, *non publié*).

LA BIODIVERSITE DANS LES FORETS TROPICALES

Les forêts tropicales situées entre le tropique du Cancer au nord et le tropique du Capricorne au sud, contiennent 50% de la biodiversité mondiale, répartie sur seulement 7% de la surface planétaire[15]. Définies par une température et une pluviométrie de $\pm 22^{\circ}\text{C}$ et 3000mm de précipitations en moyenne annuelle, elles sont également définies comme des écosystèmes complexes et hétérogènes, et constituées de micro-habitats proches géographiquement[16]. Cette étonnante biodiversité questionne, et les hypothèses qui tentent d'expliquer cette biodiversité se basent sur des taux d'extinction plus faibles et/ou taux de spéciation plus importants (Figure 2.).

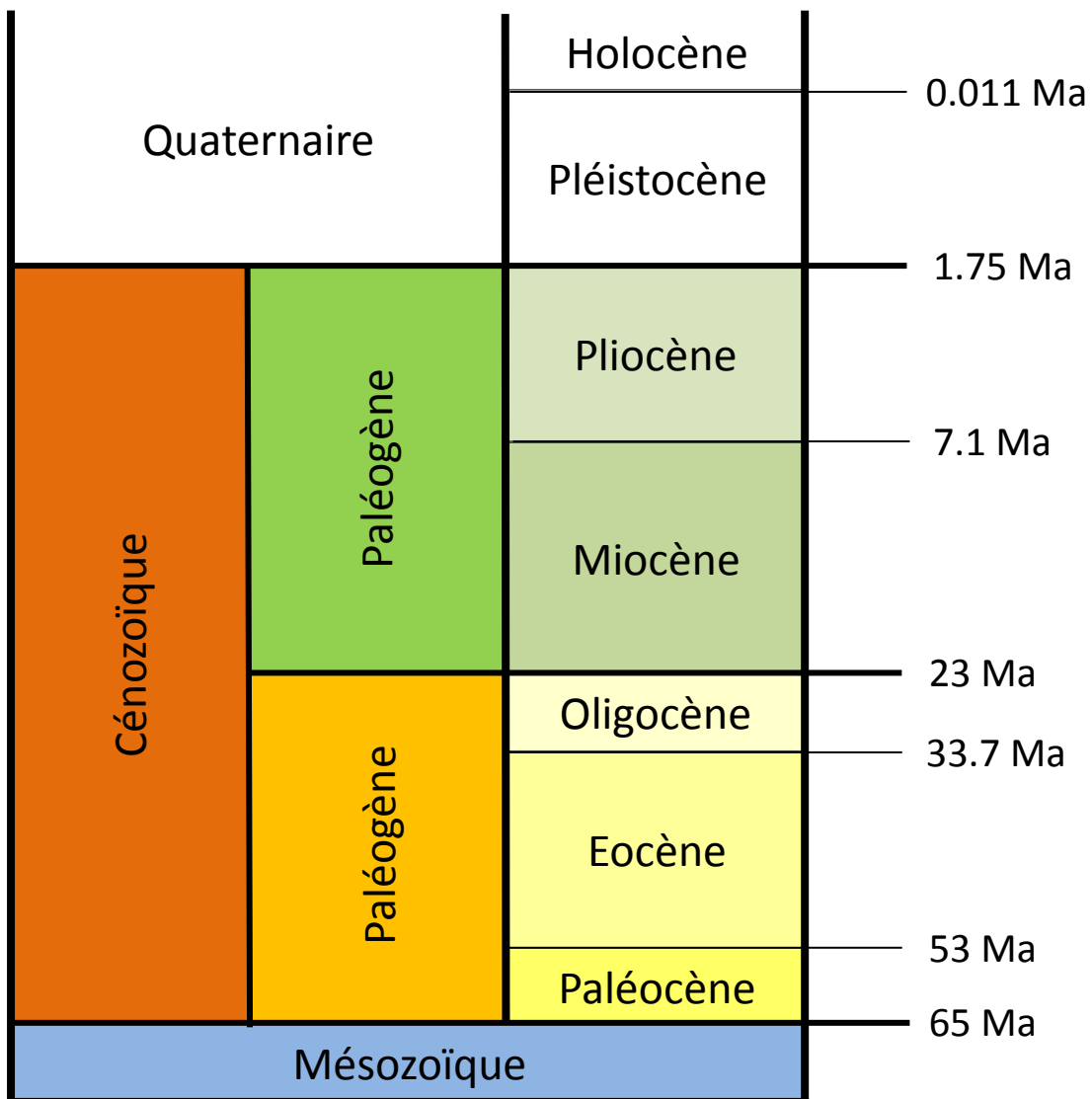


Figure 2. Echelle des temps géologiques du Mésozoïque au Quaternaire

HYPOTHESE DU BERCEAU ET THEORIE DE LA REINE ROUGE

L'hypothèse du berceau se traduit par un taux de spéciation plus fort dans les tropiques que dans les régions tempérées. Elle suppose qu'un taux de spéciation fort dans les tropiques par rapport aux zones tempérées, associé à un taux d'extinction constant des espèces, serait responsable de la forte biodiversité observée dans les milieux tropicaux[17]. L'hypothèse du berceau serait donc moteur à la diversification des espèces et pourrait être l'une des explications de la forte biodiversité végétale des forêts tropicales humides. Les espèces sont en perpétuelles compétitions et doivent être le mieux adaptées à leur habitat pour avoir la meilleure fitness ¹ qui leur permettra d'évoluer et de se reproduire dans cet habitat [18]. Les espèces cherchant à s'adapter le mieux à leurs habitats auront tendance à plus diverger.

HYPOTHESE DU MUSEE

La théorie du musée repose sur l'hypothèse que la forte biodiversité observée dans les tropiques soit ancienne, et pourrait dater du Cénozoïques, avec une accumulation des espèces du paléocène (~65Ma) au pléistocène (~2Ma)[19]. Associée à une diversité ancienne, cette hypothèse implique aussi un taux d'extinction très faible des espèces en forêts tropicales humides permettant la présence encore aujourd'hui d'une forte biodiversité. La majorité des espèces communes entre l'Asie, l'Afrique, l'Amérique centrale et l'Amérique du Sud, montrerait en effet un établissement « jeune » (durant le pléistocène), même si les espèces qui constituent en abondance ces forêts pourraient être plus ancienne et dater du cénozoïque[20]. C'est d'ailleurs pour cette raison que la biodiversité dans les forêts tropicales est parfois décrite comme étant une résultante des hypothèse du berceau et du musée[21], qui pourraient être complémentaires, plutôt qu'exclusives.

HYPOTHESE DU REFUGE

La théorie du refuge se base sur l'alternance de refroidissements (cycles glaciaires) et de réchauffement planétaires (interglaciaires) avec notamment les dernières glaciations durant le pléistocène. Elle suppose que durant ces cycles alternés de périodes de glaciation et de réchauffement planétaire, la distribution géographique des espèces a été impactée, donnant lieu à l'isolation d'espèces congénériques, et donc à la fragmentation des populations. Les populations isolées dans des environnements différents ne peuvent plus s'hybrider, ce qui provoque un arrêt de flux de gènes. Les divergences peuvent s'établir. Elles ont ainsi pu se différencier génétiquement les unes des autres, au point de constituer de nouvelles espèces et

¹ La fitness est une mesure de la sélection naturelle, et informe sur le succès reproducteur de l'organisme par le suivi des traits de la germination, la survie et la reproduction.

de permettre une plus forte proportion de spéciation dans les forêts tropicales humides. Depuis la dernière période de glaciation[22], il y a 2 millions d'années, les forêts tropicales ont non seulement été un lieu de refuge pour les espèces[23] mais également un lieu avec un important taux de spéciation.

LA GUYANE ET SON ENVIRONNEMENT HETEROGENE

Située en Amérique du Sud, entre le Suriname et le Brésil, la Guyane possède un climat tropical. Ainsi, la Guyane connaît une alternance de saisons des pluies et de saisons sèches[24]. La saison sèche s'étend de la mi-août à la mi-novembre, avec une petite saison sèche également en mars[25].

En Guyane française, le bloc forestier constitue 93.2% du territoire de 88150km²[26].

MICRO-HABITAT : BAS-FONDS ET TERRA FIRME

Dans la forêt guyanaise, un type de micro-habitat particulièrement étudié est la microtopographie qui désigne la présence, dans ces forêts, d'habitats de type « bas-fond » et « *terra firme* ».

Les habitats de type bas-fond sont désignés comme étant des habitats régulièrement inondés, dont le sol est saturé d'eau durant plusieurs semaines en saison des pluies. Aussi, il est caractérisé par une nappe phréatique inférieure à un mètre de profondeur en milieu de saison sèche[27].

Les habitats de type *terra firme* sont plus drainés en eau et présentent une nappe d'une profondeur supérieure à celle des bas-fonds. Aussi, la *terra firme* est associée à des populations végétales présentant une surface terrière moins importante que dans les habitats de type bas-fond [26]

VARIATIONS REGIONALES

La Guyane possède un gradient de pluviométrie important entre l'est et l'ouest de la Guyane[28]. A l'est, les précipitations sont fortes, de l'ordre de 3500mm par an avec une saison sèche de 24 jours environ ; tandis qu'à l'ouest, les précipitations sont plus modérées, avec une valeur de 2500 mm par an sur une période de saison sèche plus longue de 25 à 36 jours.

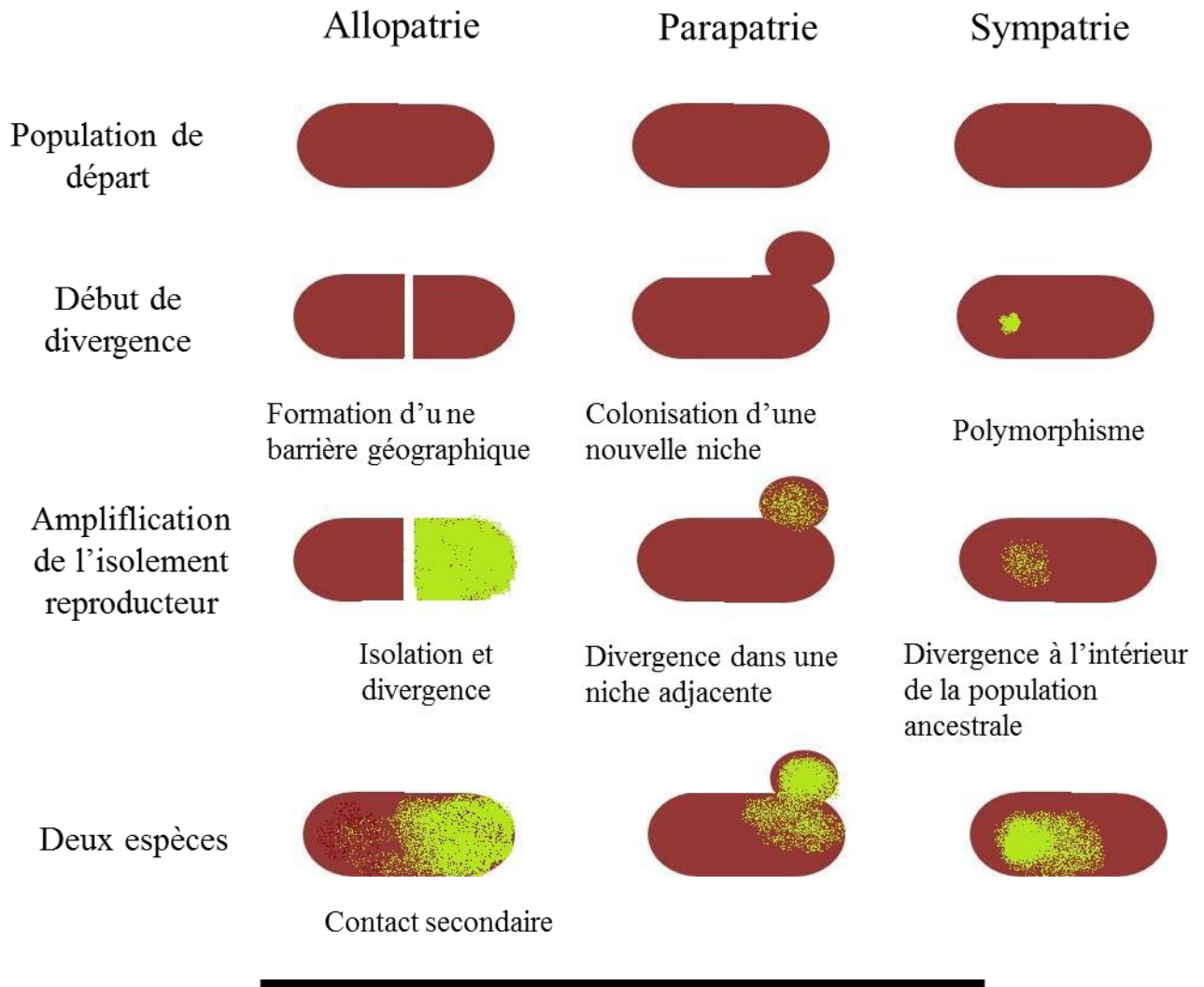
Si l'ouest de la Guyane est principalement composé de sables blancs, l'est, lui est composé d'un sol latéritique.

Box2. La spéciation

La spéciation est le processus par lequel de nouvelles espèces biologiques sont formées [29]. Les différents modes de spéciation sont caractérisés selon les différentes modalités de l'arrêt du flux de gènes (Figure 3.). L'isolement reproducteur peut se dérouler avant l'hybridation, (isolement prézygotique) en empêchant le croisement entre espèces, ou il peut se dérouler après l'hybridation (isolement post-zygotique), en engendrant des hybrides non viables.

Les modes de spéciation sont :

- La spéciation allopatrique qui est caractérisée par un isolement géographique d'une espèce, en réponse à un isolement géographique responsable de l'arrêt du flux de gènes. L'aire géographique de l'espèce est alors scindée par un événement climatique ou géologique (élévation d'une montagne, élargissement d'un cours d'eau, changement climatique, glaciations...). En l'absence de flux de gènes, et sous l'influence de la sélection naturelle et de la dérive, les populations vont diverger, jusqu'à devenir deux espèces à part entière.
- La spéciation sympatrique est caractérisée par une divergence des populations en l'absence de barrière géographique. La différenciation est alors la conséquence d'une sélection divergente qui sélectionne les caractères de façon opposée dans les populations. Ici, la spéciation se déroule en présence d'un flux de gènes. Dans ce modèle de spéciation, la sélection naturelle a un rôle très important, faisant diverger les populations jusqu'à l'établissement de la barrière reproductive.
- La spéciation parapatrique est un intermédiaire entre les deux modes de spéciation précédemment cités. Les espèces occupent des aires de répartition adjacentes, et sont donc potentiellement en contact dans un gradient environnemental. Dans ce mode de spéciation, il y a une forte importance de l'adaptation locale dans l'homogénéisation du pool génétique des deux nouvelles populations qui divergent.



- L'allopatric consiste en l'isolement géographique d'une espèce permettant un isolement reproductif.
- La parapatric consiste en un isolement reproductif permis par une migration dans une nouvelle niche
- La sympatrie consiste en une divergence de deux populations partageant un territoire commun, malgré la présence d'un flux de gènes

Figure 3. Mode de spéciation allopatric, parapatric et sympatrie

ORIGINE, MAINTIEN ET ÉVOLUTION DE LA DIVERSITE GENETIQUE ET PHENOTYPIQUE

ORIGINE DE LA DIVERSITE GENETIQUE

Les populations végétales sont génétiquement structurées dans l'espace et dans le temps, dans un environnement et à une génération donnée[29]. La structure génétique d'une population dépend alors non seulement de l'environnement dans lequel la population évolue, mais aussi de son fond génétique. L'évolution des espèces est le résultat de mécanismes génétiques (mutation, migration, dérive génétique et sélection) répétés sur plusieurs générations[30].

LE POLYMORPHISME

Selon la théorie neutre de la diversité moléculaire[30], les polymorphismes présents au sein des populations sont généralement neutres, car tout allèle apparu par mutation est rapidement fixé (si il est avantageux) ou rapidement éliminé (si il est délétère) laissant seulement les allèles « quasi » neutres dans un état polymorphe, et n'ont donc pas d'effet sur la fitness de l'individu et de la population. Une fraction des polymorphismes peut toutefois être soumise à la sélection naturelle [31-33] mais persister dans les populations, grâce à la sélection balancée ou à des régimes de sélection variables dans l'espace et dans le temps[34]. Une autre fraction du polymorphisme, transitoire, est représentée par des allèles en cours d'élimination ou de fixation [30]).

Chez les arbres forestiers le polymorphisme est généralement élevé en raison de leur biologie de la reproduction (allogame pour la plupart), un temps de génération long, peu ou pas de domestication, des populations de grandes tailles. Pour cet ensemble de critères, les arbres forestiers pourraient renfermer un réservoir de gènes importants pour répondre aux variations de l'environnement et s'adapter localement [35-37].

Les arbres tropicaux présentent aussi ses caractéristiques, avec de forts niveau de diversité génétique intra population [38-40] qui est un pré-requis pour l'action de la sélection naturelle en réponse aux variations de l'environnement.

LES MUTATIONS

La mutation permet, par l'introduction de nouveaux allèles, l'augmentation de la diversité génétique d'une population. Elle correspond à un changement héréditaire du matériel génétique de l'organisme. Quand les mutations se produisent sur un gène ou un ensemble de gènes importants pour la fitness, elles peuvent donner lieu à un polymorphisme adaptatif et ainsi avoir un rôle évolutif.

Une hypothèse complémentaire proposée pour expliquer la diversité spécifique en milieu tropical serait un taux d'évolution plus fort qu'en milieu tempéré[41]. Le taux de mutation serait plus fort en raison de l'absence d'arrêt de croissance en hiver, et d'un climat chaud et ensoleillé qui augmenterait les mutations. Il a par notamment été montré récemment[42] que le taux d'évolution des espèces tropicales est deux fois plus rapides que des espèces congénériques des climats tempérés.

LA MIGRATION

En génétique des populations la migration se caractérise par la transmission d'allèles, d'une population à une autre. Comme la mutation, elle peut permettre l'augmentation de la diversité génétique d'une population. Chez les plantes, la migration se fait par la dispersion, du pollen et des graines.

Le niveau de divergence génétique entre populations va ainsi dépendre de l'intensité de la diffusion du pollen, de sa dissémination et des distances parcourues, en moyenne, par le pollen et les graines. Des espèces à faible dispersion auront tendance à montrer une forte structuration génétique spatiale à l'intérieur des peuplements et une forte différenciation entre peuplements. Par contre, des forts taux de dispersion des graines et de longues distances de pollinisation engendreront une structure populationnelle beaucoup plus uniforme [43] et diminueront le risque de divergence des populations[29].

LA DERIVE GENETIQUE

La dérive génétique est la fluctuation des fréquences alléliques causée par l'échantillonnage des gamètes qui contribuent réellement à la génération suivante, à partir de l'ensemble des gamètes produits à une génération donnée. C'est un mécanisme de l'évolution qui entraîne une variation stochastique des fréquences des allèles à l'intérieur des populations, et au cours des générations. La dérive génétique, dans une population finie, peut rapidement fixer ou éliminer un allèle, si la population est petite, et à contrario, les chances de fixation ou d'élimination d'un allèle seront d'autant plus faibles si la taille de la population est importante.

La probabilité de fixation d'une mutation neutre qui vient d'apparaître est de $1/2N$, avec $2N$, le nombre total de gamètes[44]. La théorie prédit qu'à terme, en absence d'autres forces évolutives, la dérive entraîne la fixation de tous les locus d'une population.

LA SELECTION

La sélection est un mécanisme qui contribue à l'évolution des espèces par le changement des fréquences alléliques au sein des populations. Selon Darwin[45], elle agit comme un filtre favorisant les phénotypes les plus aptes à la survie et à la reproduction, en éliminant les phénotypes les moins adaptés. Elle agit donc au sein d'une population qui montre une diversité génétique héritable et influençant la fitness des individus. Les génotypes portant les mutations avantageuses auront une meilleure capacité reproductive qu'ils transmettront à leur descendance, conduisant à l'établissement des populations les mieux adaptés à leur environnement.

A l'intérieur d'une population, il existe trois types de sélection en fonction des effets qu'elles auront sur le devenir de la mutation. La sélection purifiante élimine les allèles délétères (mortels). La sélection positive va favoriser les allèles qui confèrent un avantage reproducteur. La sélection balancée va permettre le maintien de plusieurs allèles d'un gène et homogénéiser le pool génétique[46].

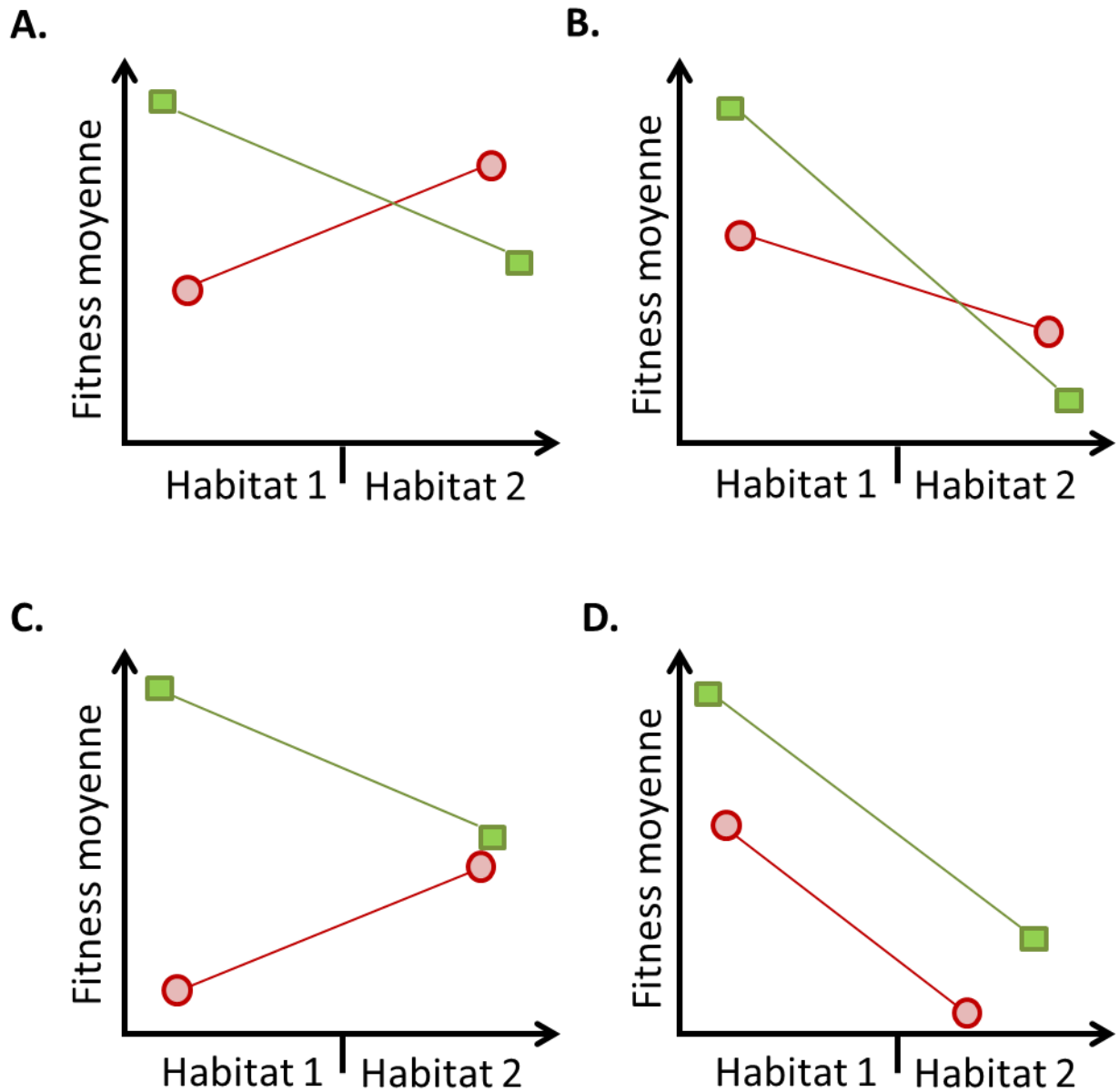
La sélection responsable de la divergence entre les populations est la sélection divergente ou diversifiante[47]. Elle va entraîner l'accumulation de la différenciation génétique entre les populations.

Box 3. Expérimentation de transplantations réciproques

L'interaction Génotype-Environnement peut donner lieu à de l'adaptation locale qui se traduit par le fait qu'un génotype est localement adapté à son environnement et va donc avoir une meilleure fitness. On peut définir un trait adaptatif comme étant un caractère (1) qui permet de caractériser le phénotype adapté à son environnement et (2) qui est positivement sélectionné dans une population et augmente la fitness de la population.

La norme de réaction est la relation entre les phénotypes possibles produits par un génotype donné et la gamme de valeurs d'une variable environnementale (Figure 4.). La décomposition de la variance phénotypique d'un trait, en composante génétique, environnementale, et l'interaction des deux, est possible en plaçant des copies de plusieurs génotypes dans des environnements différents, au sein d'un « jardin de transplantations réciproques ». L'impossibilité de cloner les individus dans la plupart des arbres forestiers (en tout cas tropicaux) ne permet pas de dupliquer les génotypes dans les différents environnements. On utilise alors des représentants d'une même famille ou d'une même population. Ce type de dispositif permet de distinguer la plasticité phénotypique « dite neutre » de la plasticité phénotypique dite « adaptative ». Si les phénotypes résultent de la plasticité phénotypique « neutre », la norme de réaction ne montrera pas d'interaction GxE car le phénotype résulte seulement de l'action de l'environnement (Figure 4. D). Lorsque les phénotypes résultent d'une interaction GxE, cela va se traduire graphiquement par une différence de pente entre les courbes de normes de réaction (A, B, C). Ceci n'est cependant pas une condition suffisante pour mettre en évidence de l'adaptation locale. Il faut aussi démontrer que la fitness d'un génotype donné (ou une population) varie en fonction des environnements et que le génotype (ou population) est mieux adapté dans son environnement local.

Il y a deux façons de regarder l'adaptation locale que l'on associe à deux schémas d'interaction[48]. Un premier schéma théorique est l'interaction « à domicile *versus* à l'extérieure » (Figure 4 A. et B.). Les individus montrent une meilleure fitness « à domicile » plutôt qu'à « l'extérieure ». Un deuxième schéma théorique est l'interaction « locaux *versus* étrangers » (Figure 4 A. et C.). Les individus qui sont chez eux ont une meilleur fitness que les « étrangers ».



(D'après Kawecki et al., 2004)

Figure 4. Normes de réactions théoriques montrant l'interaction population x habitat pour la fitness. Les carrés représentent la population originaire de l'habitat 1. Les cercles représentent la population originaire de l'habitat 2. **A.** et **B.** Les interactions répondent au critère « locaux *versus* étrangers ». **A.** et **C.** Les interactions répondent au critère « à domicile *versus* à l'extérieure ». **D.** L'habitat 1 est avantageux par rapport à l'habitat 2, aussi la population originaire de l'habitat 1 a une meilleure fitness que la population originaire de l'habitat 2, quelque soit l'habitat.

INTERACTION GENOTYPE-ENVIRONNEMENT

LA PLASTICITE PHENOTYPIQUE

La plasticité phénotypique est la capacité d'un organisme à produire plusieurs phénotypes à partir d'un seul génotype, en fonction des caractéristiques de l'environnement dans lequel il se trouve[49]. La plasticité phénotypique traduit donc l'influence de l'environnement sur le phénotype d'un organisme. La fonction qui relie le phénotype d'un génotype aux valeurs d'une variable environnementale est dite « norme de réaction ». Si deux génotypes ont des normes de réaction différentes (de sorte que l'écart entre leurs phénotypes varie en fonction de la valeur de la variable environnementale) on parle d' « interaction génotype x environnement » (GxE). Ce phénomène peut avoir une signification fonctionnelle et adaptative (Box 3). .

Lorsque plusieurs génotypes sont analysés dans plusieurs environnements, il est possible de tester les interactions GxE[48]. Lorsque l'interaction est forte on peut observer des changements d'ordre dans le classement des génotypes vis-à-vis des différents environnements (exemple figure 4 A : dans l'habitat H1, fitness carré meilleure que rond ; dans l'habitat H2, fitness rond meilleur que carré)

LA PLASTICITE PHENOTYPIQUE ADAPTATIVE

La plasticité phénotypique peut être adaptative si le phénotype engendré répond positivement à l'environnement et montre une meilleure fitness ; elle peut être mal-adaptative si le phénotype qui en résulte montre une réduction de la fitness. La plasticité phénotypique peut également être neutre, si elle n'a aucun effet sur l'adaptation ou la mal-adaptation de l'organisme.

La plasticité phénotypique adaptative peut permettre la colonisation d'un nouvel environnement ou peut permettre la persistance d'une population dans un environnement perturbé. Cela dépendra de la différence entre la valeur du trait du phénotype et l'optimum de la valeur du trait dans l'environnement colonisé[50].

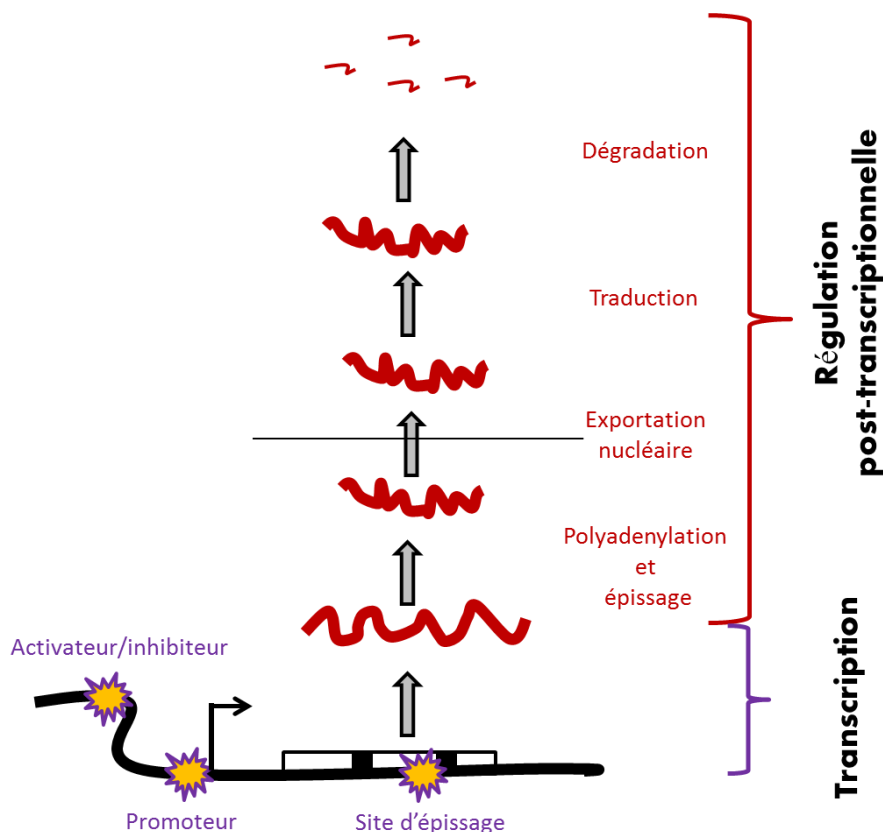
PLASTICITE PHENOTYPIQUE ADAPTATIVE ET EXPRESSION DES GENES

La transcription est le mécanisme qui permet de transcrire les séquences ADN en ARNs messager, elle est effectuée par l'ARN polymérase. Ces ARNm contiennent les informations nécessaires à la production des protéines, par le mécanisme de la traduction. De l'ADN à la production de la protéine, chaque étape est régulée, contrôlée et permet d'accroître ou de diminuer la quantité d'ARNs et de protéines produites. Aussi, une même séquence d'ADN peut donner lieu à divers combinaisons d'ARNs, par le biais de l'épissage alternatif.

Un gène est exprimé s'il produit des ARNs et des protéines (dans le cas d'ARNs codants). En fonction de la régulation de l'expression, un gène peut être surexprimé ou sous-exprimé par rapport à son niveau basal d'expression[51].

Les variations génétiques sont capables de réguler l'expression des gènes quand elles sont situées sur des séquences de régulation. Les promoteurs qui sont des séquences situées à l'avant du gène sont des zones sur lesquelles l'ARN polymérase peut se fixer. Les « enhancers » ou « silencers » sont situés aux abords du gène et vont permettre l'activation ou l'inhibition du promoteur. La présence de boîte de branchement permet l'épissage alternatif. Enfin, la structure secondaire de la molécule d'ADN peut le rendre accessible au non à la RNA polymérase (Figure 5.).

Enfin, l'environnement est capable d'influencer l'expression génétique globale des gènes dans un organisme[52] ; Cela donne lieu à la plasticité phénotypique[53, 54]. Tous les caractères ne sont pas plastiques, mais certains peuvent réagir à la pression exercée par l'environnement. La plasticité phénotypique adaptative peut contribuer à la divergence des espèces dans le cadre d'une spéciation en présence d'un flux de gènes comme la spéciation sympatrique ou écologique (Figure 3.).



(D'après Williams, Chan et al., 2007)

Figure 5. Les facteurs génétiques et cellulaire de la régulation de l'expression de l'ARNm

DETECTION DE LA DIVERGENCE ET OUTILS MOLECULAIRES

DIVERGENCE ET FREQUENCES ALLELIQUES

LES MARQUEURS MOLECULAIRES

Un marqueur génétique est un caractère individuel qui répond aux deux critères suivants : (a) être transmis de façon Mendélienne (b) être indépendant de l'environnement. Il existe des marqueurs codominants, qui permettent d'identifier les deux allèles d'un même locus (chez un organisme diploïdes) et permettent donc de distinguer les individus homozygotes des individus hétérozygotes. Il existe aussi des marqueurs dominants qui ne détectent que l'allèle dominant pour chaque locus et ne permettent donc pas de distinguer un individu hétérozygote d'un individu homozygote. Un marqueur moléculaire répond aux deux critères énoncés ci-dessus, et est obtenu par une manipulation des propriétés biochimiques de l'organisme. Typiquement, un marqueur moléculaire est obtenu à partir des caractéristiques du matériel génétique (ADN, ARN) de l'organisme selon une ou plusieurs des méthodes propres à la génétique et à la biologie moléculaires (coupe d'un acide nucléique par des enzymes de restriction, PCR, séquençage...).

Les marqueurs AFLPs (Amplified Fragment Length Polymorphism)[55, 56] ont le désavantage d'être dominants, mais ont l'avantage de s'étendre sur tout le génome sans cibler de séquences précises et permettent donc, sans connaissance au préalable du génome, d'obtenir des informations sur le polymorphisme de milliers de locus. Les marqueurs AFLPs renseignent sur des polymorphismes de présence/absence de fragments de tailles différentes ; ainsi, s'ils restent des marqueurs de choix dans l'appréhension des polymorphismes à l'intérieur de populations d'espèces dont on ne connaît pas encore le génome, ils sont sensibles à l'homoplasie [57] (tailles des allèles identiques qui ne correspondent pas à la même région du génome). Enfin, la répétabilité de ces marqueurs est particulièrement discutée[58] et nécessite une vérification du taux d'erreurs de phénotypage par des tests de répétabilité.

FREQUENCES ALLELIQUES ET TESTS DE SELECTION

Les fréquences des allèles neutres varient de manière stochastique à l'intérieur des populations. Ce n'est pas le cas des allèles ayant un rôle adaptatif et donc pouvant influencer la reproduction et la survie d'un organisme. Ainsi, les allèles avantageés par la sélection tendent à avoir une fréquence plus élevée que les allèles neutres dans la population. A contrario, les allèles contre-sélectionnés auront une fréquence plus faible qu'à l'attendu sous l'hypothèse de la neutralité[59].

La détection de locus soumis à la sélection divergente (sélection qui favorise des allèles différents entre des populations localisées dans des environnements différents) peut être effectuée par l'estimation de l'indice de différenciation génétique, le F_{ST} . Quand une différence de fréquences allélique est importante entre deux populations, la valeur du F_{ST} se rapproche de 1[60]. Des tests basés sur l'écart entre la valeur de F_{ST} individuelle à chaque locus et la valeur attendue sous un modèle neutre permettent d'identifier des locus potentiellement soumis à sélection (marqueurs « outliers »). Une série de tests ont été développés depuis les premiers tests proposés par Lewontin & Krakauer en 1973[61]. Une amélioration des tests a été la prise en compte de la relation entre le F_{ST} et l'hétérozygotie (Figure 6) qui est mieux adaptée pour les locus multialléliques[62]. En effet, il est clairement démontré que la distribution du F_{ST} neutre dépend fortement de l'hétérozygotie observée. Les tests se sont ensuite complexifiés en simulant des modèles neutres plus compliqués que les simples modèles en îles (utilisé par Beaumont & Nichols, 1996) qui prennent par exemple en compte des sous structuration de populations [63, 64].

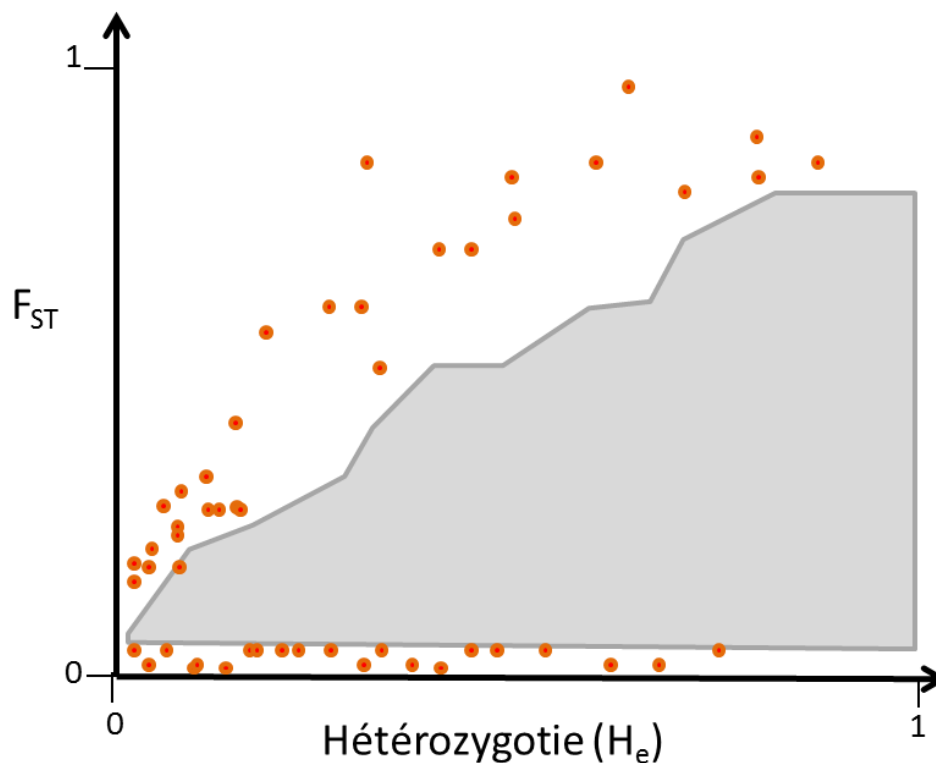


Figure 6. Identification des « outliers ». Les « outliers » (en orange), ont un F_{ST} supérieur à l'attendu du nuage neutre (en gris).

EXPRESSIONS DIFFERENTIELLES ET TECHNIQUES

OBSERVATION DE L'EXPRESSION DIFFERENTIELLE

OBJECTIFS

L'étude de l'expression des gènes (en particulier, de leur niveau de transcription) consiste à quantifier la quantité d'ARNm produits par un organisme, ou un organe, de façon à identifier les séquences produites et les niveaux d'expression des gènes dont elles sont issues. Cela permet l'observation du transcriptome labile à un moment précis ou dans une condition donnée que l'on souhaite étudier[65].

Diverses méthodes d'observation de l'expression existent[66]; les stratégies basées sur le séquençage à haut débit permettent d'estimer le taux d'expression des gènes par l'observation directe du nombre de molécules d'ARNs produites par chaque gène et la production d'une table de comptage contenant le nombre de copies des gènes[67, 68].

LE RNASEQ

L'ARN total va être extrait de l'échantillon d'intérêt avant la création de banques enrichies ou non en ARNm par l'utilisation ou non de PCR sélective de la queue Poly A. Ces banques vont être fixées sur des plateformes nommées *flowcell* afin de subir l'élongation des brins ADNc par une reverse transcriptase avec des amorces universelles permettant le séquençage de n'importe quel gène.

La technologie de séquençage Illumina en pair-end (adaptateurs sur les deux bords) ou en single end (adaptateur sur un seul bord) produit des centaines de millions de séquences de 30 à 100pb, que nous appellerons « *read* ». Ces *reads* obtenus représentent des copies plus ou moins nombreuses des gènes exprimés constituant le transcriptome de l'échantillon d'intérêt. Selon des critères d'homologie[69] les *reads* vont être assemblés en contigs par des logiciels d'assemblage tels que Velvet[69, 70], Cufflinks[71], ou Trinity[72] afin de reconstruire *de novo* le transcriptome de l'échantillon d'intérêt (Figure 7.). On s'attend à obtenir aux alentours de 20 000 gènes pour un arbre tel que le *S.globulifera*[73].

TECHNIQUES DE DETECTION DES GENES DIFFERENTIELLEMENT EXPRIMES ET TESTS STATISTIQUES

Pour un gène donné, on veut savoir s'il est plus exprimé (ou moins exprimé) dans une condition, par rapport à une autre. Pour comparer les échantillons, la première étape consiste à la normalisation des données. Pour cela, il existe plusieurs méthodes dont la méthode de normalisation nommée Reads Per Kilobase and per Million (RPKM) où le nombre de *reads* obtenu pour le gène va être divisé par le produit de la taille du gène et le nombre de *reads*

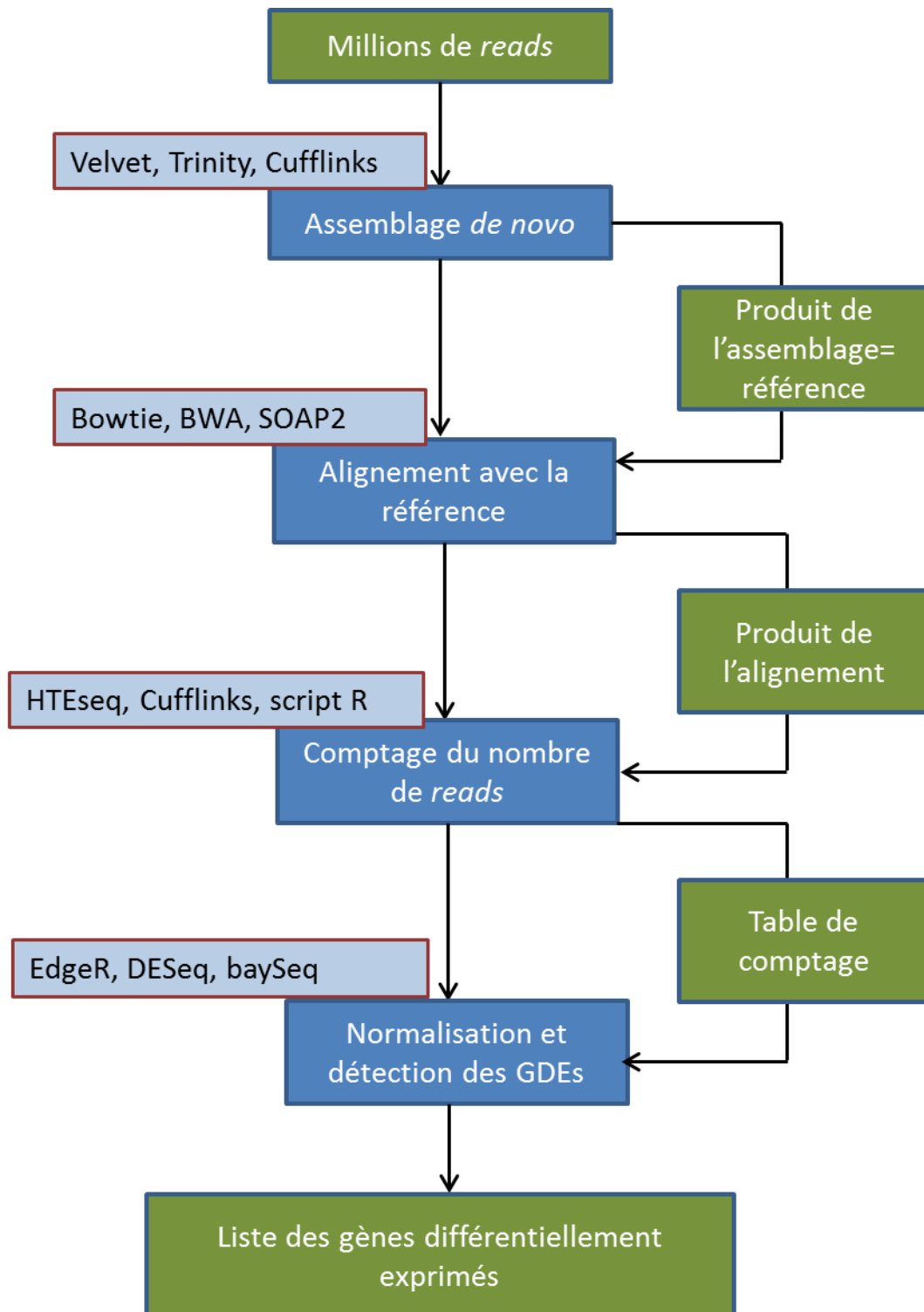


Figure 7 . Pipeline pour la détection de l'expression différentielle. Les rectangles bleus renseignent le nom des méthodes; les rectangles bleu ciel renseignent les logiciels permettant le traitements des données; les rectangles verts renseignent les produits.

obtenus pour le gène. Une méthode plus élaborée est implémentée dans EdgeR[74] et DESeq[75] qui sont les packages les plus utilisés dans la détection de gènes différentiellement exprimés (GDEs). La normalisation va s'effectuer en prenant la moyenne géométrique des échantillons et en les normalisant tous par cette référence [75-77]. En considérant un « read » comme un évènement d'échantillonnage parmi un nombre fixe de gènes, l'expérience peut être approximée par une loi de poisson [78], qui n'est pas assez sensible aux variations biologiques et conduit des faux positifs. Ainsi, EdgeR et DESeq modèlent cet échantillonnage par une binomiale négative [75, 77, 78] qui est la généralisation de la loi de poisson. Une fois la moyenne géométrique calculée, elle pourra permettre la normalisation des échantillons, puis la détection des gènes différentiellement exprimés. La variation de l'expression des gènes, considérée comme un trait caractérisant un phénotype, sera expliquée par l'utilisation de modèles linéaires multi-facteurs [74, 76, 79]. Ceci permet de prendre en compte tous les facteurs présent dans l'expérimentation et pouvant influencer la variation de l'expression des gènes (Figure 7.).

EPIGENETIQUE ET EXPRESSION DES GENES

LA METHYLATION DE L'ADN

La méthylation des îlots CpG consiste en l'ajout d'un groupe méthyl (CH_3) sur un couple de nucléotides constitués d'une cytosine accompagnée d'une guanine. Ceci est effectué par des enzymes appartenant à la famille des « Dna Méthyl-Transférases (DNMTs). La méthylation de la cytosine est la marque épigénétique la plus étudiée, avant la modification des histones, et joue un rôle important dans la structure de la chromatine et l'expression des gènes. Chez les plantes, il existe trois types de profils de méthylation. Le plus conservé est le motif CG, aussi appelé CpG. Il est présent en plus grande proportion chez les plantes que chez d'autres organismes[80].

La méthylation est associée à des processus de régulation de l'expression des gènes[81, 82] comme l'activation de retrotransposons, les modifications d'histones qui vont réguler la compaction de l'ADN[83], ou la méthylation de l'ADN qui sur un promoteur va diminuer l'expression alors qu'elle l'augmentera si elle est située sur une région codante [84].

Plusieurs méthodes permettent de suivre les états de méthylation de l'ADN [85-87], notamment de façon à différencier des populations génétiquement proches [88, 89].

EPIGENETIQUE ET EVOLUTION

La théorie de l'évolution sur l'établissement de nouveaux caractères au sein d'une population et sur l'évolution des espèces repose sur la sélection naturelle, agissant comme un filtre et favorisant les individus les mieux adaptés à leur environnement. En tenant compte de la génétique mendélienne, Fisher a développé la théorie de la génétique des populations avec Wright et Haldane[90]. Dès lors, les modèles de biologie évolutive tentant de comprendre les mécanismes de l'évolution des espèces s'appuient sur les variations de séquences comme principal facteur pouvant avoir un impact sur l'évolution des populations.

L'épigénétique est un domaine nouveau dans l'étude de l'évolution des espèces[91], et dont les conséquences sur l'évolution des espèces peut être importante puisque des marques épigénétiques ont déjà montré un fort potentiel à la réponse directe de l'environnement, agissant alors sur le phénotype de l'organisme[92-94]. La théorie qui présente un rôle important des marques épigénétiques dans l'évolution des espèces fait l'hypothèse que les marques épigénétiques serait le démarreur de l'adaptation locale permettant une adaptation rapide, qui sera par la suite suivie par des variations de séquences qui vont stabiliser le nouveau phénotype, par le biais de l'assimilation génétique [81, 92, 95-97]. Un point très important à la base de cette théorie est qu'une cytosine méthylée a beaucoup plus de chance

de subir une mutation [98, 99]. Ainsi, la méthylation pourrait agir comme un « recruteur » de mutation adaptative.

HYPOTHESES ET EXPERIMENTATIONS

OBJECTIFS ET HYPOTHESES

L'objectif de cette thèse est d'étudier les mécanismes génétiques à la base de la divergence et de la spéciation écologique chez les arbres forestiers.

Le modèle biologique est le *Symphonia globulifera*, il présente deux écotypes, l'un spécialiste de la *terra firme* et l'autre spécialiste des bas-fonds. Nous pourrions donc être en présence d'une spéciation écologique en présence de flux de gènes car il n'existe qu'une faible différenciation génétique entre le *S.globulifera* (écotype d'origine) et le *S.sp1* (nouvel écotype). Ce modèle biologique était déjà étudié dans le laboratoire. Aussi, en 2009, des jardins de transplantations réciproques avaient déjà été mis en place. Ainsi, le *Symphonia globulifera* paraît être un très bon modèle pour l'analyse des mécanismes moléculaires et fonctionnels mis en place pour l'adaptation à un nouvel environnement.

Mon hypothèse à défendre est la suivante : Il y a peu de différenciation génétique mais une expression des gènes différente entre les écotypes, responsable d'une adaptation locale et de la différence des traits observés en population naturelle.

Ces hypothèses ont d'abord été testées par des expérimentations de transplantations réciproques, dans lesquelles l'analyse en génétique quantitative de traits de croissance et de fitness permet de mettre en évidence la présence ou non d'une adaptation locale chez les écotypes.

Puis, l'expression des gènes ayant déjà été liée à des statuts de méthylation dans la littérature scientifique, j'ai effectué un génotypage par le biais de marqueurs AFLP (MSAP) sur l'ADN de deux populations d'arbres adultes, dans le but de mettre en évidence une différence de statuts de méthylations entre les populations de *S.sp1* et *S.globulifera*.

Enfin, dans un troisième volet, j'ai utilisé la technique de séquençage RNAseq sur des plantules issues des transplantations réciproques afin de vérifier l'hypothèse forte de ce projet, à savoir que, l'adaptation locale des populations de *S.sp1* et *S.globulifera*, et la différence de taille des organes observée puissent être liées à une expression différentielle des gènes chez les deux écotypes. Il s'agissait alors de voir si l'expression des gènes est significativement différente en fonction de l'environnement et/ou en fonction de l'écotype.

GUYANE FRANÇAISE : 3 SITES D'ETUDE

LES SITES DE LAUSSAT ET DE REGINA

Les sites expérimentaux de Laussat et de Régina sont des propriétés de l'Office National des Forêts (ONF) de Guyane.

Situé à l'ouest de la Guyane, à environ 60 Km de Saint-Laurent du Maroni (5°28'N, 53°34'W), le site expérimental de Laussat, est le lieu d'établissement de jardins de transplantation réciproques, et comprend également des installations de mesures météorologiques (température et pluviométrie).

Le site de Régina est situé à 20Km environ de la commune de Régina (4°18'N, 52°14'W). Il reçoit des expérimentations de transplantations réciproques et les arbres de l'espèce *Symphonia globulifera* ont été répertoriés sur ce site.

Sur les deux sites, des relevés pédologiques ont été effectués permettant de connaître la composition des sols.

Le site de Laussat est caractérisé par une couche de sable blanc recouvrant un sol charbonneux et riche en minéraux. La température annuelle est de 27°C, et la pluviométrie moyenne annuelle est d'environ 2500mm, tandis que celui de Régina est caractérisé par un sol constitué de latérite, une température annuelle de 27°C, une pluviométrie de 3000mm par an. En saison sèche, la nappe montre une hauteur de 0 à 40 cm en bas-fond à Laussat et est à plus de 1m de hauteur à Régina.

LE SITE DE PARACOU

Le dispositif expérimental de Paracou est une concession accordée au CIRAD par le Centre Spatial Guyanais (CSG), situé à proximité de Sinnamary (5°18'N, 52°53'W).

La température annuelle est de 26°C et la pluviométrie annuelle moyenne est de 2980mm[25]. Le site de Paracou comprend 16 parcelles en forêt primaire. Dans chaque parcelle, les arbres de plus de 10cm de largeur sont régulièrement recensés, cartographiés et identifiés botaniquement. Pour notre étude, nous avons effectué des récoltes sur les parcelles 14 et 15 uniquement. Ces parcelles n'ont subis aucuns traitements.

BIBLIOGRAPHIE

1. Abdul-Salim, K., *Systematics and Biology of Symphonia L.f. (Clusiaceae)*. 2002: Harvard University.
2. Degen, B., E. Bandou, and H. Caron, *Limited pollen dispersal and biparental inbreeding in Symphonia globulifera in French Guiana*. *Heredity*, 2004. **93**: p. 585-591.
3. Dick, Christopher W., K. Abdulah Salim, and E. Bermingham, *Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree*. *The American Naturalist*, 2003. **162**(6): p. 691-703.
4. Dick, C.W. and M. Heuertz, *The complex biogeographic history of a widespread tropical tree species*. *Evolution*, 2008. **62**(11): p. 2760-2774.
5. Aldrich, P.r., et al., *Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree Symphonia globulifera*. *Molecular Ecology*, 1998. **7**(8): p. 933-944.
6. Gill, G.E., R.T. Fowler, and S.A. Mori, *Pollination Biology of Symphonia globulifera (Clusiaceae) in Central French Guiana I*. *Biotropica*, 1998. **30**(1): p. 139-144.
7. Vinson, C., et al., *Characterization and isolation of DNA microsatellite primers for the tropical tree, Symphonia globulifera Linn. f.* *Molecular Ecology Notes*, 2005. **5**: p. 202-204.
8. Fromentin, Y., et al., *Symphonia globulifera, a widespread source of complex metabolites with potent biological activities*. *Planta Med*, 2015. **81**(2): p. 95-107.
9. Provine, W.B., *Ernst Mayr: Genetics and speciation*. *Genetics*, 2004. **167**(3): p. 1041-6.
10. Turrill, W.B., *THE ECOTYPE CONCEPT*. *New Phytologist*, 1946. **45**(1): p. 34-43.
11. Baraloto, C., et al., *Seasonal water stress tolerance and habitat associations within four neotropical tree genera*. *Ecology*, 2007. **88**(2): p. 478-489.
12. Loubry, D., *Déterminisme du comportement phénologique des arbres en forêt tropicale humide de Guyane française (5°lat. N.)*. PhD Thesis, University of Paris 6, 1994.
13. Budde, K.B., et al., *The ancient tropical rainforest tree Symphonia globulifera L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa*. *Heredity*, 2013. **111**(1): p. 66-76.
14. van Andel, T.R., *Floristic composition and diversity of three swamp forests in northwest Guyana*. *Plant Ecology*, 2003. **167**(2): p. 293-317.
15. Myers, N., *Tropical forests and their species*. *Biodiversity*. National Academy: Washington, 1988: p. 28-35.
16. Baraloto, C. and P. Couteron, *Fine-scale Microhabitat Heterogeneity in a French Guianan Forest*. *Biotropica*, 2010. **42**(4): p. 420-428.
17. Arita, H.T. and E. Vázquez-Domínguez, *The tropics: cradle, museum or casino? A dynamic null model for latitudinal gradients of species diversity*. *Ecology letters*, 2008. **11**(7): p. 653-663.
18. Chase, J.M. and T.M. Knight, *Community genetics: toward a synthesis*. *Ecology*, 2003. **84**(3): p. 580-582.
19. Pennington, R.T., M. Hughes, and P.W. Moonlight, *The Origins of Tropical Rainforest Hyperdiversity*. *Trends in Plant Science*.
20. Richardson, J.E., et al., *Rapid Diversification of a Species-Rich Genus of Neotropical Rain Forest Trees*. *Science*, 2001. **293**(5538): p. 2242-2245.

21. Rull, V., *Speciation timing and neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence*. Mol Ecol, 2008. **17**(11): p. 2722-9.
22. Connor, E.F., *The role of Pleistocene forest refugia in the evolution and biogeography of tropical biotas*. Trends in Ecology & Evolution, 1986. **1**(6): p. 165-168.
23. Duminil, J., et al., *Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis*. Journal of Biogeography, 2015. **42**(8): p. 1443-1454.
24. Gond, V., et al., *Broad-scale spatial pattern of forest landscape types in the Guiana Shield*. International Journal of Applied Earth Observation and Geoinformation, 2011. **13**(3): p. 357-367.
25. Epron, D., et al., *Spatial Variation of Soil Respiration across a Topographic Gradient in a Tropical Rain Forest in French Guiana*. Journal of tropical ecology, 2006. **22**(5): p. 565-574.
26. Koponen, P., et al., *Tree species diversity and forest structure in relation to microtopography in a tropical freshwater swamp forest in French Guiana*. Plant Ecology, 2004. **173**(1): p. 17-32.
27. Ferry, B., et al., *Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest*. Journal of Ecology, 2010. **98**(1): p. 106-116.
28. Baraloto, C., et al., *Disentangling stand and environmental correlates of aboveground biomass in Amazonian forests*. Global Change Biology, 2011. **17**(8): p. 2677-2688.
29. Loveless, M.D. and J.L. Hamrick, *Ecological Determinants of Genetic Structure in Plant Populations*. Annual Review of Ecology and Systematics, 1984. **15**(1): p. 65-95.
30. Kimura, M., *The neutral theory of molecular evolution*. 1985, Cambridge, UK: Cambridge University Press.
31. Paaby, A.B., et al., *Identification of a candidate adaptive polymorphism for Drosophila life history by parallel independent clines on two continents*. Molecular Ecology, 2010. **19**(4): p. 760-774.
32. Bierne, N., *The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population*. Evolution, 2010. **64**(11): p. 3254-3272.
33. Lei, Y.-K., et al., *Adaptive genetic variation in the smoke tree (Cotinus coggygria Scop.) is driven by precipitation*. Biochemical Systematics and Ecology, 2015. **59**: p. 63-69.
34. Delph, L.F. and J.K. Kelly, *On the importance of balancing selection in plants*. New Phytol, 2014. **201**(1): p. 45-56.
35. Neale, D.B. and A. Kremer, *Forest tree genomics: growing resources and applications*. Nature Reviews Genetics, 2011. **12**(2): p. 111-122.
36. Savolainen, O. and T. Pyhajarvi, *Genomic diversity in forest trees*. Current Opinion in Plant Biology, 2007. **10**(2): p. 162-167.
37. Savolainen, O., T. Pyhajarvi, and T. Knurr, *Gene Flow and Local Adaptation in Trees*. Annu Rev Ecol Evol Syst, 2007. **38**(1): p. 595 - 619.
38. Hamrick, J.L. and M. Godt, *Allozyme diversity in plant species*, in *Plant Population Genetics, Breeding, and Genetic Resources*, A.H.D. Brown, et al., Editors. 1990, Sinauer Associates: Sunderland, MA. p. 43-63.
39. Hamrick, J.L. and D.A. Murawski, *Levels of allozyme diversity in populations of uncommon Neotropical tree species*. Journal of tropical ecology, 1991. **7**(03): p. 395-399.
40. Hamrick, J.L., M. Godt, and S. Sherman-Broyles, *Factors influencing levels of genetic diversity in woody plant species*. New Forests, 1992. **6**(1-4): p. 95-124.

41. Rohde, K., *Latitudinal Gradients in Species Diversity: The Search for the Primary Cause*. *Oikos*, 1992. **65**(3): p. 514-527.
42. Wright, S., J. Keeling, and L. Gillman, *The road from Santa Rosalia: A faster tempo of evolution in tropical climates*. *Proceedings of the National Academy of Sciences of the USA*, 2006. **103**(20): p. 7718-7722.
43. Hamrick, J.L., D.A. Murawski, and J. Nason, *The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations*. *Vegetatio*, 1993. **107/108**: p. 281-297.
44. Wright, S., *EVOLUTION IN MENDELIAN POPULATIONS*. *Genetics*, 1931. **16**(2): p. 97-159.
45. Darwin, C., *De l'origine des espèces*. 1859.
46. Nielsen, R., *MOLECULAR SIGNATURES OF NATURAL SELECTION*. *Annual Review of Genetics*, 2005. **39**(1): p. 197-218.
47. Nosil, P., D.J. Funk, and D. Ortiz-Barrientos, *Divergent selection and heterogeneous genomic divergence*. *Molecular Ecology*, 2009. **18**(3): p. 375-402.
48. Kawecki, T.J. and D. Ebert, *Conceptual issues in local adaptation*. *Ecology letters*, 2004. **7**(12): p. 1225-1241.
49. Valladares, F., E. Gianoli, and J.M. Gómez, *Ecological limits to plant phenotypic plasticity*. *New Phytologist*, 2007. **176**(4): p. 749-763.
50. Ghalambor, C., et al., *Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments*. *Functional ecology*, 2007. **21**: p. 394-407.
51. Dekel, E. and U. Alon, *Optimality and evolutionary tuning of the expression level of a protein*. *Nature*, 2005. **436**: p. 588-592.
52. Babu, M.M. and L. Aravind, *Adaptive evolution by optimizing expression levels in different environments*. *Trends in microbiology*, 2006. **14**(1): p. 11-14.
53. Dayan, D.I., D.L. Crawford, and M.F. Oleksiak, *Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of *Fundulus heteroclitus**. *Molecular Ecology*, 2015. **24**(13): p. 3345-3359.
54. Healy, T.M. and P.M. Schulte, *Phenotypic plasticity and divergence in gene expression*. *Molecular Ecology*, 2015. **24**(13): p. 3220-3222.
55. Vos, P., et al., *AFLP: a new technique for DNA fingerprinting*. *Nucleic Acid Research*, 1995. **23**(21): p. 4407-4414.
56. Zabeau, M. and P. Vos, *Selective restriction fragment amplification : a general method for DNA fingerprinting*, 1993, Google Patents.
57. Caballero, A., H. Quesada, and E. Rolán-Alvarez, *Impact of Amplified Fragment Length Polymorphism Size Homoplasmy on the Estimation of Population Genetic Diversity and the Detection of Selective Loci*. *Genetics*, 2008. **179**(1): p. 539-554.
58. Ley, A.C. and O.J. Hardy, *Improving AFLP analysis of large-scale patterns of genetic variation – a case study with the Central African lianas *Haumania* spp (Marantaceae) showing interspecific gene flow*. *Molecular Ecology*, 2013. **22**(7): p. 1984-1997.
59. Fay, J.C. and C.-I. Wu, *Hitchhiking Under Positive Darwinian Selection*. *Genetics*, 2000. **155**(3): p. 1405-1413.
60. Luikart, G., et al., *The power and promise of population genomics: from genotyping to genome typing*. 2003. **4**(12): p. 981-994.
61. Lewontin, R.C. and J. Krakauer, *Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms*. *Genetics*, 1973. **74**: p. 175-195.
62. Beaumont, M.A. and R.A. Nichols, *Evaluating Loci for Use in the Genetic Analysis of Population Structure*. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 1996. **263**(1377): p. 1619-1626.

63. Beaumont, M.A. and D.J. Balding, *Identifying adaptive genetic divergence among populations from genome scans*. *Molecular Ecology*, 2004. **13**(4): p. 969-980.
64. Excoffier, L., T. Hofer, and M. Foll, *Detecting loci under selection in a hierarchically structured population*. *Heredity*, 2009. **103**(4): p. 285-298.
65. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. *Nat Rev Genet*, 2009. **10**: p. 57 - 63.
66. Alvarez, M., A.W. Schrey, and C.L. Richards, *Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution?* *Molecular Ecology*, 2015. **24**(4): p. 710-725.
67. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. *Genome Biol*, 2010. **11**: p. R106.
68. Wagner, G., K. Kin, and V. Lynch, *Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples*. *Theory in Biosciences*, 2012. **131**(4): p. 281-285.
69. Zerbino, D. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs*. *Genome Res*, 2008. **18**: p. 821 - 829.
70. Zerbino, D.R., *Using the Velvet de novo assembler for short-read sequencing technologies*. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]*, 2010. **CHAPTER**: p. Unit-11.5.
71. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks*. *Nat Protoc*, 2012. **7**(3): p. 562 - 578.
72. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome*. *Nat Biotech*, 2011. **29**(7): p. 644-652.
73. Brousseau, L., et al., *High-throughput transcriptome sequencing and preliminary functional analysis in four Neotropical tree species*. *BMC Genomics*, 2014. **15**(1): p. 238.
74. Robinson, M., D. McCarthy, and G. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. *Bioinformatics*, 2010. **26**: p. 139 - 140.
75. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. *Genome Biology*, 2010. **11**(10): p. R106.
76. McCarthy, D.J., Y. Chen, and G.K. Smyth, *Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation*. *Nucleic Acids Research*, 2012. **40**(10): p. 4288-4297.
77. Anders, S., et al., *Count-based differential expression analysis of RNA sequencing data using R and Bioconductor*. *Nat. Protocols*, 2013. **8**(9): p. 1765-1786.
78. Kvam, V.M., P. Liu, and Y. Si, *A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data*. *American Journal of Botany*, 2012. **99**(2): p. 248-256.
79. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. *Nucleic Acids Research*, 2015.
80. Feng, S., et al., *Conservation and divergence of methylation patterning in plants and animals*. *Proceedings of the National Academy of Sciences*, 2010. **107**(19): p. 8689-8694.
81. Lukens, L.N. and S. Zhan, *The plant genome's methylation status and response to stress: implications for plant improvement*. *Current Opinion in Plant Biology*, 2007. **10**(3): p. 317-322.
82. Zilberman, D., et al., *Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription*. *Nature Genet.*, 2007. **39**: p. 61-69.

83. Fedoroff, N.V., *Transposable Elements, Epigenetics, and Genome Evolution*. Science, 2012. **338**(6108): p. 758-767.
84. Flores, K.B., F. Wolschin, and G.V. Amdam, *The Role of Methylation of DNA in Environmental Adaptation*. Integrative and Comparative Biology, 2013. **53**(2): p. 359-372.
85. Cokus, S., et al., *Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning*. Nature, 2008. **452**: p. 215 - 219.
86. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics*. Nat Rev Genet, 2008. **9**(6): p. 465-476.
87. Harris, R., et al., *Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications*. Nat Biotechnol, 2010. **28**: p. 1097 - 1105.
88. Alonso, C., et al., *MSAP markers and global cytosine methylation in plants: a literature survey and comparative analysis for a wild-growing species*. Molecular Ecology Resources, 2015: p. n/a-n/a.
89. Rathore, M., S. Mastan, and P. Agarwal, *Evaluation of DNA methylation using methylation-sensitive amplification polymorphism in plant tissues grown in vivo and in vitro*. Plant Growth Regulation, 2015. **75**(1): p. 11-19.
90. Fisher, R., *The genetical theory of natural selection*. 1958, New York: Dover Publications.
91. Felsenfeld, G., *A Brief History of Epigenetics*. Cold Spring Harbor Perspectives in Biology, 2014. **6**(1).
92. Richards, C.L., O. Bossdorf, and M. Pigliucci, *What Role Does Heritable Epigenetic Variation Play in Phenotypic Evolution?* BioScience, 2010. **60**(3): p. 232-237.
93. Herrera, C.M. and P. Bazaga, *Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees*. Botanical Journal of the Linnean Society, 2013. **171**(3): p. 441-452.
94. Schulz, B., R.L. Eckstein, and W. Durka, *Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb*. Molecular Ecology, 2014. **23**(14): p. 3523-3537.
95. Pigliucci, M., C.J. Murren, and C.D. Schlichting, *Phenotypic plasticity and evolution by genetic assimilation*. Journal of Experimental Biology, 2006. **209**(12): p. 2362-2367.
96. Bossdorf, O., C.L. Richards, and M. Pigliucci, *Epigenetics for ecologists*. Ecology letters, 2008. **11**(2): p. 106-115.
97. Skinner, M.K., *Environmental Epigenetics and a Unified Theory of the Molecular Aspects of Evolution: A Neo-Lamarckian Concept that Facilitates Neo-Darwinian Evolution*. Genome Biology and Evolution, 2015. **7**(5): p. 1296-1302.
98. Norberg, J. and M. Vihinen, *Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides*. Journal of Molecular Structure: THEOCHEM, 2001. **546**(1-3): p. 51-62.
99. Kumar, S., et al., *Roles, and establishment, maintenance and erasing of the epigenetic cytosine methylation marks in plants*. J Genet, 2013. **92**(3): p. 629-66.

*PARTIE I: ANALYSE DES CARACTERES DE
L'ADAPTATION LOCALE DANS
UNE EXPERIENCE
DE TRANSPLANTATION RECIPROQUE*

<i>Introduction</i>	46
<i>Matériel et méthodes</i>	48
Echantillonnage et plan d'expérience	48
Mesure de traits phénotypiques	51
Les traits de croissance.....	51
Les traits de fitness.....	51
Données additionnelles	52
Trait de compétition dans les jardins.....	52
Année de germination	52
Analyses statistiques	52
Pour les traits de croissance	52
Pour les traits de fitness.....	53
<i>Résultats</i>	54
Effets environnementaux observés sur les traits de croissance et de fitness	54
Effet de l'habitat sur la hauteur et le nombre de feuilles par centimètre	54
Effet de la plantation régionale sur la survie des plantules	54
Effets génétiques observés sur les traits de croissance et de fitness	55
Effets de l'écotype sur le diamètre et le nombre de feuilles par centimètre	55
Effets de l'écotype sur l'herbivorie	55
Effet de la provenance régionale sur la survie des plantules	55
Interactions Génotype x Environnement sur les traits de croissance et de fitness	56
Effets d'interaction Ecotype*Habitat pour la hauteur	56
Interaction Ecotype*Habitat pour la germination	56
Interaction provenance régionale*site de plantation pour la survie des plantules	56
Effet d'interaction croisés	57
Effet d'interaction Provenance*Habitat sur les traits de croissance.....	57
Effet d'interaction Provenance*Habitat sur les traits de fitness.....	57
Effet d'interaction Ecotype*Sites sur les traits de croissance	58
Effet d'interaction Ecotype*Sites sur les traits de fitness	58
<i>Discussion</i>	59
Adaptation locale au sein du complexe <i>S. globulifera</i> et <i>S. sp1</i>	59
Sur les traits de fitness.....	59
Sur les traits de croissance	60
Adaptation régionale au sein du complexes d'espèces <i>S. globulifera</i> et <i>S. sp1</i>	60
Sur les traits de fitness.....	60
Sur les traits de croissance	61
<i>Limites de l'étude</i>	61
<i>Conclusion</i>	61
<i>Bibliographie</i>	72

INTRODUCTION

La germination et le recrutement sont les premières étapes de la vie de l'arbre[1] et constituent des phases critiques à l'établissement des populations des espèces dans les environnements[2]. Ce sont des acteurs directs de l'avantage reproducteur[3] et sont responsables du fait que les plantes ne sont pas actrices dans le choix de leur habitat mais que celui-ci leur est imposé[4]. Les graines et les plantules, dans les premières phases de recrutement montrent un taux important de mortalité[1] due au manque de ressources du sol[5], à la prédation[6, 7] et à la maladaptation[8, 9]. Pourtant, l'hétérogénéité de l'environnement dans les forêts tropicales humides[10], qui présentent des micro-habitats différents mais très proches géographiquement, a été associée à une biodiversité importante. Cette hétérogénéité ne semble pas être un frein à l'établissement des populations d'arbres, mais semble être responsable de la répartition des espèces [11, 12].

L'adaptation locale traduit l'avantage reproducteur d'un génotype dans un environnement donné par rapport à un autre et dépend ainsi de l'effet des interactions entre le génotype et l'environnement[13]. Le suivi de la fitness et de la variation des traits phénotypiques qui divergent entre les populations au sein d'un jardin de transplantation réciproque permet de déceler, par l'étude des normes de réaction et des interactions Génotype-Environnement, la présence d'une adaptation locale [13, 14]. Il existe deux types de test d'interaction se traduisant par une adaptation locale. Le test d'interaction « à domicile versus à l'extérieur » montre une meilleure fitness « à domicile » plutôt qu'à « l'extérieur » pour les populations adaptés, tandis que le test d'interaction « locaux versus étrangers » montre que les individus qui sont chez eux ont une meilleur fitness que les « étrangers »[13]. La plupart des génotypes peuvent répondre différemment à une pression de l'environnement, par le biais de la plasticité phénotypique dont l'effet est de faire varier le phénotype d'un génotype donné, en fonction de l'habitat dans lequel il se situe [15-17]. Ainsi, ce croisement d'habitat permet alors de déterminer si : (i) l'observation phénotypique que nous faisons est le résultat de la plasticité, avec un génotype unique qui répond différemment à deux types d'habitats, ou si (ii) il s'agit de deux génotypes, adaptés localement à un habitat en particulier. Dans le premier cas, il y a une réponse communes des deux écotypes et différente en fonction de l'habitat. Dans le deuxième cas, on observe deux réponses différentes des deux écotypes en fonction de l'habitat.

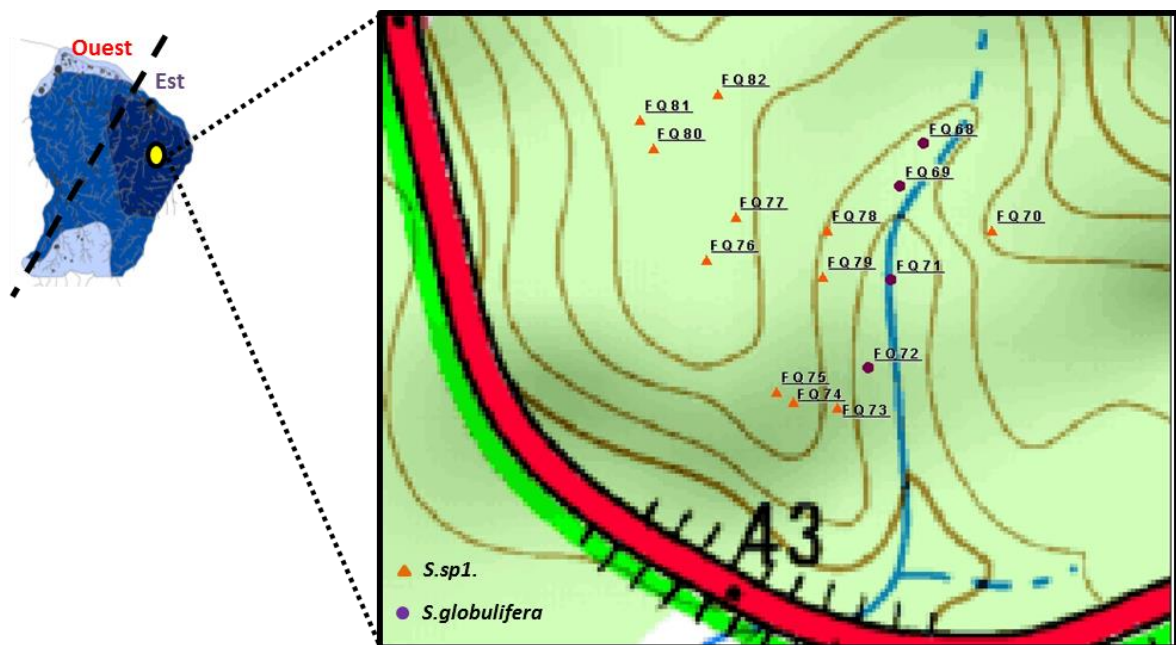


Figure 1: Répartition non-aléatoire des écotypes en bas-fond et terre-ferme. Carte montrant un agrandissement d'une portion du site de plantation de l'Est, avec les populations naturelles d'arbres matures de l'écotype *S.sp1* en terre-ferme (lignes très rapprochées) et l'écotype *S.globulifera* en bas-fond (lignes moins rapprochées).

Arbre Mère	Site	Latitude	Longitude	Provenance régionale	Provenance topographique	Morphotype	Nombre de graines récoltées
W426	Kaw	4°33'22.34"N	52°14'20.00"W	Est	Bas-fond	<i>S. globulifera</i>	36
W503	Régina	4°16'00.84"N	52°09'51.84"W				37
M837	Montagne Trésor	4°33'38.21"N	52°13'59.06"W		Plateau	<i>S. sp1</i>	36
W424	Kaw	4°33'20.93"N	52°14'21.12"W				38
W425	Kaw	4°33'22.07"N	52°14'22.76"W	Ouest	Bas-fond	<i>S. globulifera</i>	39
W463	St Laurent	5°23'28.70"N	53°39'07.83"W				36
W466	Crique Naï	5°23'47.03"N	53°42'25.95"W				36
W497	Apatou	5°09'27.57"N	54°20'16.22"W		Plateau	<i>S. sp1</i>	36
W498	Apatou	5°09'27.58"N	54°20'16.23"W				35
W465	Crique Naï	5°23'47.04"N	53°42'25.96"W				36
W474	Montagne de Fer	5°21'07.37"N	53°32'48.57"W				36
W475	Montagne de Fer	5°21'07.38"N	53°32'48.58"W	Plateau	<i>S. sp1</i>	36	
W476	Montagne de Fer	5°21'07.39"N	53°32'48.59"W			36	
W477	Montagne de Fer	5°21'07.40"N	53°32'48.60"W			37	

Table 1: Echantillonnage régional et topographique de 510 graines de *Symphonia globulifera* en bas-fonds et en *terra firme* à l'Est et à l'Ouest de la Guyane.

Dans cette étude, je veux déterminer la présence ou non d'une adaptation locale à l'intérieur des populations d'arbres, de l'espèce *Symphonia globulifera* présente en Guyane, permettant d'expliquer la répartition naturelle non aléatoire des écotypes (Figure 1.). J'utilise la génétique quantitative de façon mettre en évidence les facteurs responsables de la différence morphologique des populations et de la spécialisation des écotypes à leur environnement, dans une expérimentation de jardins de transplantations réciproques. Selon l'hypothèse d'une spéciation en cours, nous nous attendons à observer deux réponses différentes des deux écotypes adaptés localement à leur habitat, bas-fond ou *terra firme* ; qui est traduit par une interaction Ecotype*Habitat significative.

MATERIEL ET METHODES

ECHANTILLONNAGE ET PLAN D'EXPERIENCE

Des graines des deux écotypes *S.globulifera* et *S.sp1* ont été échantillonnées de Septembre 2008 à Avril 2009, à l'est et à l'ouest de la Guyane (Table 1.). Les semis provenant de six arbres mères de l'écotype *S.globulifera* en bas-fonds et 8 arbres mères de l'écotype *S.sp1* en *terra firme* ont été mis en germination dans des pots en serre avec les mêmes conditions utilisées par Brousseau et al.[18].

Par la suite, de Mai à Juillet 2009, les plantules germées et des graines non germées ont été transplantées dans les jardins, à raison de 3 graines ou plantules par famille, aléatoirement choisies et réparties dans chaque bloc.

Au préalable, 12 jardins de transplantations avaient été mis en place (Table 2). Pour chaque conditions, 3 blocs de jardins (9*10m), séparés de 50m ont été installés (3 jardins en bas-fonds et 3 jardins en *terra firme* situés à l'est et à l'ouest de la Guyane, Figure 2), et protégés à l'aide de grillage de façon à exclure les gros herbivores.

Le site de plantation à l'est (4°18'N, 52°14'W) est caractérisé par un sol constitué de latérite, une température annuelle minimale de 22.1°C, une température maximale de 31.2°C, et une pluviométrie de 3364.6mm par an (de 1981 à 2010, Météo France). Ce site est entouré des Montagnes Tortues et des montagnes de Kaw. La nappe peut-être à plus de 1m de profondeur en bas-fonds, en saison sèche.

Le site de plantation Ouest (5°28'N, 53°34'W) est caractérisé par une couche de sable blanc recouvrant un sol charbonneux et riche. La température annuelle est de 27°C, et la pluviométrie moyenne annuelle est de 2528.3 mm (de 1981 à 2010, Météo France). La nappe montre une profondeur entre 0 et 40cm en saison sèche, en bas-fonds.

Pour les deux sites, les habitats *terra firme* ont une nappe de profondeur supérieur à un mètre en milieu de saison sèche.

Numéro de jardin	Latitude	Longitude	Site de plantation	Habitat
Jardin 1	5°28'41.8980"N	53°34'27.6000"W	Ouest	Plateau
Jardin 2	5°28'49.9836"N	53°34'31.7905"W		
Jardin 3	5°28'51.2976"N	53°34'28.1923"W		
Jardin 4	5°28'39.7956"N	53°34'40.4156"W		Bas-fond
Jardin 5	5°28'42.1068"N	53°34'39.6138"W		
Jardin 6	5°28'37.2936"N	53°34'39.3984"W		
Jardin 7	4°18'55.0908"N	52°14'08.9144"W	Est	Bas-fond
Jardin 8	4°18'53.8848"N	52°14'09.3016"W		
Jardin 9	4°18'51.8364"N	52°14'06.8656"W		
Jardin 10	4°18'49.1364"N	52°14'04.8827"W		Plateau
Jardin 11	4°18'49.5936"N	52°14'04.4617"W		
Jardin 12	4°18'49.7268"N	52°14'03.3591"W		

Table 2: Emplacement des jardins de transplantations

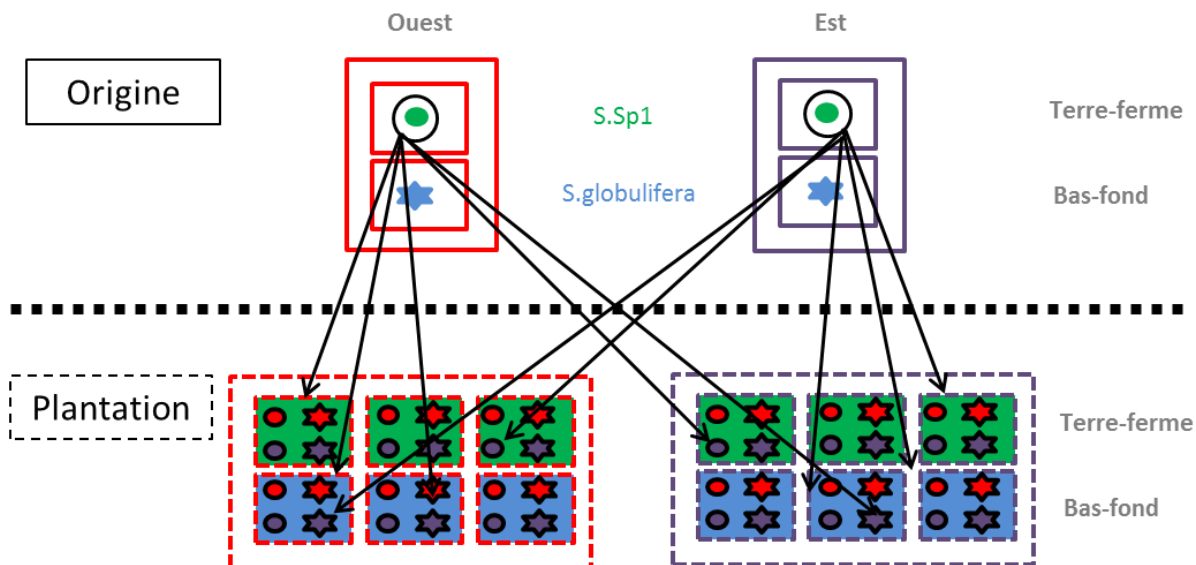


Figure 2: Design expérimental des jardins de transplantations. Les écotypes *Symphonia globulifera* et *S.sp1* provenant de l'Est et de l'Ouest de la Guyane, sont plantés en bas-fond ou terre-ferme à l'Est et à l'Ouest de la Guyane. A l'est et à l'ouest, 3 jardins en bas-fonds et 3 jardins en terre-ferme.

MESURE DE TRAITS PHENOTYPIQUES

Afin de suivre la croissance et la fitness des plantules, nous avons choisi de mesurer la hauteur, le diamètre, le nombre de feuilles, pour ce qui constitue les traits de croissance, et l'herbivorie, la survie et la germination, pour ce qui concerne les traits de fitness².

LES TRAITS DE CROISSANCE

La hauteur (H) et le diamètre (D) ont été mesurés en centimètres, avec une précision au millimètre. La mesure de la hauteur a été prise à l'aide d'un mètre entre le collet et l'apex de la plantule. Le diamètre a été mesuré au collet à l'aide de pieds à coulisse dans deux sens orthogonaux, de façon à prendre en compte l'irrégularité de la largeur de la tige. Nous avons également compté le nombre total de feuilles sur chaque plantule. Cette donnée a été divisée par la hauteur de la plantule de façon à obtenir un nombre de feuilles par centimètre (TNLH).

LES TRAITS DE FITNESS

L'herbivorie (Herb) a été définie par classe. Pour chaque feuille de chaque plantule, une observation visuelle a permis d'attribuer à chaque feuille un pourcentage de surface consommée. Ensuite, en fonction de ce pourcentage, j'ai pu les répartir selon 5 classes numériques. J'ai attribué une valeur de 10 si la feuille avait subi moins de 20% d'attaques sur toute sa surface, 30 si le pourcentage d'attaque sur la feuille était compris entre 20 et 40%, 50 s'il était entre 40 et 60% d'attaques, 70 s'il était entre 60 et 80% d'attaques et enfin, pour les feuilles qui avaient subi plus de 80% d'attaques sur toute leur surface, je leur attribuais une valeur de 90. Pour obtenir une valeur de l'herbivorie par plantule, une moyenne d'herbivorie par plantule a été calculée en divisant la somme des classes numériques pour chaque feuille par le nombre total de feuilles présentes cette année-là sur la plantule.

La survie des plantules (Survie) a été codée : 1 si la plantule était vivante, 0 si la plantule est morte et NA si la plantule était morte avant l'année d'observation. Les résultats donnés seront sous forme de pourcentage de plantules vivantes.

La germination des graines (Germ) a été codée : 1 si la graine avait germé, 0 si elle n'avait pas encore germé, et NA si la graine avait germé avant l'année d'observation. Les résultats donnés seront sous forme de pourcentage de graines germées.

²

La fitness informe sur le succès reproducteur et la capacité des organismes à survivre.

Les mesures ont été effectuées en septembre de chaque année, de 2009 à 2014, sauf en 2012. Nous avons dans cette analyse les résultats de 5 campagnes de mesures, pour chacun des traits présentés.

DONNEES ADDITIONNELLES

TRAIT DE COMPETITION DANS LES JARDINS

La surface terrière à l'intérieur et jusqu'à 10m autour des jardins a été mesurée en Septembre 2014.

ANNEE DE GERMINATION

La date de germination des plantules et des graines est donnée par l'année de la première observation de la plantule.

ANALYSES STATISTIQUES

L'analyse des effets de l'écotype, de la provenance régionale, de l'habitat ou du site de plantation régional, a été effectuée à l'aide du Modèle Linéaire Généralisé (GLM) avec une décomposition de la somme des carrés des écarts à la moyenne de type III afin de prendre au mieux en compte les interactions dans le modèle.

POUR LES TRAITS DE CROISSANCE

Les effets des facteurs propres et interactions sur les phénotypes des plantules ont été calculés avec une erreur de type gaussienne avec une fonction de lien « identity », concernant les traits de croissance. Le jeu de données concerne les plantules vivantes, plantées sous forme de graines et de plantules. Ainsi, nous prenons en compte dans le modèle de croissance, l'état de plantation (graine ou plantule) ainsi que l'interaction Année de mesure*Année de germination, de façon à ce que les graines germées tardivement ne soient pas traités dans le modèle comme les plantules plus âgées. D'autre part, il a été ajouté au modèle l'interaction écotype*Année, pour tenir compte du fait que d'une année à l'autre un écotype peut être plus performant que l'autre pour le trait étudié.

La moyenne de l'herbivorie par plantule a été analysée comme les traits de croissance.

Les analyses se font selon le modèle suivant :

$Y = \mu + \text{effet écotype} + \text{effet provenance régionale} + \text{effet habitat} + \text{effet site régional de plantation} + \text{interactions (effet écotype*effet habitat, effet écotype*effet site de plantation régionale, effet provenance régionale*effet site régional de plantation, effet provenance régionale*effet habitat)} + \text{état de plantation (graines ou plantules)} + \text{surface terrière} + \text{Année de mesure} + \text{interaction écotype*Année de mesure} + \text{interaction Année de mesure*Année de germination} + \text{résidus}$

Avec μ la moyenne entre tous les individus pour le trait observé.

POUR LES TRAITS DE FITNESS

Les effets propres et interactions sur les phénotypes des plantules ont été calculés avec une erreur de type binomial et une fonction de lien « logit ». Concernant la germination et la survie, les jeux de données concernent les « graines » et les « plantules », séparément.

$Y = \mu + \text{effet écotype} + \text{effet provenance régionale} + \text{effet habitat} + \text{effet site régional de plantation} + \text{interactions (effet écotype*effet habitat, effet écotype*effet site de plantation régionale, effet provenance régionale*effet site régional de plantation, effet provenance régionale*effet habitat)} + \text{surface terrière} + \text{Année de mesure} + \text{interaction écotype*Année de mesure} + \text{résidus}$

Avec μ la moyenne entre tous les individus pour le trait observé

RESULTATS

Le modèle linéaire a permis de mettre en avant des effets propres et d'interactions, de l'habitat et de l'écotype, mais aussi du site de plantation et de la région de provenance sur les traits de fitness (Table 3) ainsi que sur les traits de croissance (Table 4).

EFFETS ENVIRONNEMENTAUX OBSERVES SUR LES TRAITS DE CROISSANCE ET DE FITNESS

EFFET DE L'HABITAT SUR LA HAUTEUR ET LE NOMBRE DE FEUILLES PAR CENTIMETRE

J'ai pu mettre en évidence un effet de l'habitat sur la hauteur des plantules (Df=1, F=3.9165, $p < 0.05^*$). En bas-fond, les plantules sont plus grandes (moyenne=20.275 \pm écart-type=13.279) qu'en terra firme (moyenne=19.191 \pm écart-type 8.570). En 2014, la hauteur moyenne d'une plantule en bas-fond est de 31.903 (\pm 17.815) et de 25.717 cm (\pm 9.822), en *terra firme* (Figure 3,A).

C'est également le cas pour le nombre de feuilles par centimètre (Df=1, F=3.8792, $p < 0.05^*$). En bas-fond, les plantules ont plus de feuilles en moyenne (moyenne=0.573 \pm 0.299) qu'en terra firme (moyenne=0.556 \pm écart-type=0.287). L'évolution du nombre de feuilles par centimètre est irrégulière, mais de 2009 à 2011, les plantules de bas-fond ont montré plus de feuilles (moyenne=0.606 \pm écart-type=0.310) qu'en terra firme (moyenne=0.549 \pm écart-type=0.288). Par contre, à partir de 2013, les plantules en *terra firme* ont en moyenne plus de feuilles par centimètre (moyenne=0. 566 \pm écart-type=0.287) qu'en bas-fond (moyenne=0.510 \pm écart-type=0.267) (Figure 3, B).

EFFET DE LA PLANTATION REGIONALE SUR LA SURVIE DES PLANTULES

Les plantules survivent mieux quand elles sont plantées à l'est de la Guyane (Df=1, F=13.699, $p < 0.001^{***}$). En effet, à l'est les plantules ont un taux de survie moyen de 0.676 (\pm 0.468), tandis qu'à l'ouest ce taux descend à 0.576 (\pm 0.494) (Figure 3,C.).

EFFETS GENETIQUES OBSERVES SUR LES TRAITS DE CROISSANCE ET DE FITNESS

EFFETS DE L'ECOTYPE SUR LE DIAMETRE ET LE NOMBRE DE FEUILLES PAR CENTIMETRE

Dès le plus jeune âge, les plantules de *S.globulifera* ont un diamètre significativement supérieur (3.763 ± 1.658) à celui des plantules de *S. sp1* (2.646 ± 1.146) (Df=1, F=5.8378, $p<0.05^*$). En 2014, dernière année de mesure, le *S.globulifera* a en moyenne un diamètre de $4.824 (\pm 1.777)$ contre $3.653 (\pm 1.400)$ pour les plantules de l'écotype *S. sp1*, soit un diamètre 1.32 fois supérieur (Figure 4,A.).

Par contre, concernant le nombre de feuilles par centimètre, c'est l'écotype *S. sp1* qui a significativement plus de feuilles (moyenne= $0.687 \pm$ écart-type= 0.285) que le *S. globulifera* (moyenne $0.390 \pm$ écart-type 0.202 ; Df=1, F=6.1236, $p<0.05^*$). En 2009, le *S. sp1* a en moyenne 0.635 feuilles par centimètres (± 0.276), tandis que le *S. globulifera* a 0.360 feuilles par centimètres en moyenne (± 0.159). En 2014, je note un nombre de feuilles par centimètres de 0.754 ± 0.265 pour le *S. sp1* et de 0.341 ± 0.154 , pour le *S. globulifera* (Figure 4, B).

EFFETS DE L'ECOTYPE SUR L'HERBIVORIE

Les feuilles de *S. globulifera* sont significativement plus consommées (9.550 ± 9.814) que les feuilles de *S. sp1* (7.724 ± 6.948 ; Df=1, F=45.5676, $p<0.001^{***}$). Suivant la courbe d'évolution de l'herbivorie, ce n'est que depuis 2013 que le *S. globulifera* subit plus d'attaques que le *S. sp1*. Entre 2009 et 2013, la proportion moyenne d'attaques pour une plantule est de $5.646 (\pm 7.025)$ pour le *S. globulifera* et $5.473 (\pm 6.788)$ pour le *S. sp1*, contre $19.223 (\pm 9.034)$ et $13.115 (\pm 3.519)$, respectivement, depuis 2013(Figure 4, C.).

EFFET DE LA PROVENANCE REGIONALE SUR LA SURVIE DES PLANTULES

Les plantules provenant de l'ouest de la Guyane survivent mieux (0.649 ± 0.4783) que les plantules dont les graines proviennent de l'est de la Guyane (moyenne= $0.633 \pm$ écart-type= 0.482 ; Df=1, F=19.9048, $p<0.001^{***}$). (Figure 4, D.).

INTERACTIONS GENOTYPE X ENVIRONNEMENT SUR LES TRAITS DE CROISSANCE ET DE FITNESS

EFFETS D'INTERACTION ECOTYPE*HABITAT POUR LA HAUTEUR

Un effet significatif de l'interaction Ecotype*Habitat sur le trait de la hauteur a été mis en évidence (Df=1, F=4.2822, p<0.05*). De 2011 à 2014, quand le *S. globulifera* est planté en *terra firme*, il est moins haut (moyenne=22.664 ± écart-type=7.742) que le *S. globulifera* planté en bas-fond (moyenne=29.191 ± écart-type=13.737). Depuis 2013, il est également moins haut (moyenne=23.719 ± écart-type=8.023) en *terra firme* que le *S. sp1* en bas-fond (moyenne=26.644 ± écart-type=17.165) et en *terra firme* (moyenne=24.308 ± écart-type=9.819 ; Figure 5,A).

INTERACTION ECOTYPE*HABITAT POUR LA GERMINATION

Les graines de *S.sp1*, en *terra firme*, ont un taux de germination 1.98 fois supérieur à celui des graines de *S. sp1* plantés en bas-fond, avec une moyenne de 0.242 (soit 24.2%, ± 0.256) et de 0.0701 (± 0.430) respectivement (Df=1, F=3.6908, p<0.05*). Le taux de germination moyen des graines de *S. globulifera* en *terra firme* est de 0.093 (± 0.291), et en bas-fond de 0.082 (± 0.275). Ces chiffres sont comparables aux taux des graines de *S. sp1* plantés en bas-fond (figure 5,B). La courbe d'évolution de la germination des écotypes dans les habitats montre que les graines de *S.sp1* ont un pic de germination en 2013 à 20 %, alors que dans les autres conditions la germination est presque nulle.

INTERACTION PROVENANCE REGIONALE*SITE DE PLANTATION POUR LA SURVIE DES PLANTULES

Les effets de l'interaction provenance régionale*site de plantation sur le trait de survie des plantules sont significatifs (Df=1, F=9.8514, p<0.01**). La provenance ouest montre un meilleur taux de survie à domicile (0.708 ± 0.457 contre 0.614 ± 0.488). C'est également le cas pour la provenance est (0.695 ± 0.460 contre 0.541± 0.499 ; Figure 5, C.).

EFFET D'INTERACTION CROISES

EFFET D'INTERACTION PROVENANCE*HABITAT SUR LES TRAITS DE CROISSANCE

Les plantules ayant pour provenance l'ouest et plantées en bas-fond ($23.392\text{cm} \pm 14.216$; Figure 6, A.) montrent des hauteurs significativement plus importantes que celles des plantules à provenance ouest plantée en *terra firme* ($20.833\text{cm} \pm 9.051$ cm) ; qui sont également plus hautes que les plantules à provenance est plantées en bas-fond ($17.548\text{cm} \pm 11.767$ cm) et en *terra firme* ($17.138\text{cm} \pm 7.452$ cm ; Df=1, F=7.238, $p<0.01^{**}$).

De façon similaire, la provenance régionale de l'ouest plantée en bas-fond ($3.503\text{cm} \pm 1.655$ cm ; Figure 6. B.) est significativement plus large au collet que lorsqu'elle est plantée en *terra firme* ($3.052\text{cm} \pm 1.190$ cm) et plus large encore que les plantules provenant de l'est, qui ont un diamètre moyen de $2.881\text{cm} (\pm 1.595$ cm) et $3.048\text{cm} (\pm 1.433$ cm) (Df=1, F=15.0193, $p<0.001^{***}$), respectivement, en bas-fond et en *terra firme*.

Enfin, concernant le nombre de feuilles par centimètre, la provenance de l'est montrent aussi le plus de feuilles par centimètres quand elle est en bas-fond (0.661 ± 0.322 contre 0.627 ± 0.301 , en *terra firme*), tandis que pour la provenance de l'ouest, cela s'observe pour la *terra firme* (0.498 ± 0.263 contre 0.473 ± 0.232 , Df=1, F=3.3077, $p<0.05^*$). Je remarque tout de même qu'à partir de 2014 la provenance de l'est montre plus de feuilles en *terra firme* (0.615 ± 0.285 , de 2009 à 2013 et 0.683 ± 0.363 , en 2014 ; Figure 6, C.).

EFFET D'INTERACTION PROVENANCE*HABITAT SUR LES TRAITS DE FITNESS

Les graines provenant de l'est ont significativement un meilleur taux de germination en bas-fond (moyenne= $0.258 \pm \text{écart-type}=0.441$) qu'en *terra firme* (moyenne= $0.105 \pm \text{écart-type}=0.311$; Df=1, F=14.7962, $p<0.001^{***}$). Celles de l'ouest montrent un meilleur taux de germination en *terra firme* (moyenne= $0.181 \pm \text{écart-type}=0.386$) qu'en bas-fond (moyenne= $0.034 \pm \text{écart-type}=0.183$; Figure 7,A.). Par contre, le taux de survie des plantules provenant de l'ouest et plantées en bas-fond (moyenne= $0.806 \pm \text{écart-type}=0.396$) est significativement supérieur à celui de la *terra firme* (moyenne= $0.400 \pm \text{écart-type}=0.493$; Df=1, F=15.3986, $p<0.001^{***}$). Aussi, le taux de survie des plantules provenant de l'est et plantées en bas-fond est de $0.689 (\pm 0.463)$, et en *terra firme*, et de $0.581 (\pm 0.494$; Figure 7, B.) en bas-fond.

EFFET D'INTERACTION ECOTYPE*SITES SUR LES TRAITS DE CROISSANCE

L'écotype *S. globulifera* est plus haut (moyenne=22.388 ± écart-type=10.791) que l'écotype *S. sp1* (moyenne=17.843 ± écart-type=1.070) avec une tendance à arborer une hauteur moins importante quand il est planté à l'est (moyenne=21.654 ± écart-type=9.483). Cette hauteur est du même ordre de celle du *S. sp1* dans les deux conditions : est (moyenne=18.301 ± écart-type=12.132) et ouest (moyenne=17.394 ± écart-type=9.918, Df=1 ; F=9.7104, p<0.01** ; Figure 8. A.).

Le diamètre au collet est plus grand pour le *S. globulifera* quand il est à l'ouest (moyenne 3.871 ± écart-type=1.932) qu'à l'est (moyenne 3.718 ± écart-type=1.528) de la Guyane (Figure 8. B.), tandis que dans le cas du *S. sp1*, les valeurs de diamètre à l'ouest (moyenne 2.532 ± écart-type=1.066) sont inférieures (moyenne 2.763 ± écart-type=1.214) à celles de l'est (Df=1, F=7.5239, p<0.01**). Enfin, concernant le nombre de feuilles par centimètres, dans l'interaction écotYPE*site de plantation, le *S.globulifera* fait plus de feuilles quand il est planté à l'est (moyenne 0.401± écart-type=0.209 contre moyenne 0.364 ± écart-type=0.181) de la Guyane, alors que le *S. sp1* fait, à l'inverse, plus de feuilles quand il est planté à l'ouest de la Guyane (moyenne 0.701 ± écart-type=0.293 contre une moyenne de 0.673 ± écart-type=0.277 ; Df=1, F=4.8396, p<0.05* ; Figure 8.C).

EFFET D'INTERACTION ECOTYPE*SITES SUR LES TRAITS DE FITNESS

Les plantules de *Symphonia globulifera* ont un taux de survie de 0.585 (± 0.495) quand elles sont plantées à l'ouest et de 0.728 ± 0.445 quand elles sont plantées à l'est (Df=1, F=4.1852, p<0.05*). Concernant le *S. sp1*, les plantules ont un taux de survie moyen de 0.573 (± 0.495) quand elles sont plantées à l'ouest et de 0.628 ± 0.484, à l'est. On remarque ici, que la différence de survie des écotypes entre les sites de plantations est plus importante quand les plantules sont dans les jardins de l'est (0.155) que lorsqu'elles sont plantées dans les jardins de l'ouest (0.012 ; Figure 9.).

DISCUSSION

ADAPTATION LOCALE AU SEIN DU COMPLEXE *S. GLOBULIFERA* ET *S. sp1*

SUR LES TRAITS DE FITNESS

L'interaction Ecotype*Habitat sur le trait de la germination montre que l'Ecotype *S.sp1* germe mieux en *terra firme* qu'en bas-fond. Cet effet étant observé sur un trait de fitness important, tel que la germination[13], il suggère une adaptation locale du *S. sp1* pour la *terra firme*, et peut contribuer à expliquer la répartition naturelle non aléatoire des écotypes, avec le *S. sp1* localisé sur la *terra firme* et le *S. globulifera* dans les bas-fonds[12]. Du point de vue de l'évolution des espèces et pour revenir sur l'hypothèse de spéciation écologique faite en introduction de cette thèse, la mise en évidence d'une adaptation locale est un facteur clé en faveur de la spéciation écologique [19, 20]. Par ailleurs, en considérant la faible différenciation génétique entre les écotypes[21], cette spéciation se réaliserait en présence de flux de gènes, et s'apparenterait à un cas de spéciation écologique en condition de sympatrie. Ainsi mes résultats vont dans le sens d'un processus de spéciation incipiente³, où les deux espèces localement adaptées ne sont pour autant pas encore différenciées génétiquement. Nous serions dans un cas de spéciation écologique [19, 22-24]: les deux espèces se côtoient même si elle ne partagent pas le même micro-habitat ; malgré un décalage de floraison montré par Loubry en 1994[25], il existe toujours un flux de gènes ou le flux de gènes s'est arrêté il y a peu de temps.

De façon surprenante, je ne détecte pas d'interaction Ecotype*Habitat sur la survie des plantules de 2009 à 2014, ce qui signifierait qu'une fois germée, une plantule de *S. sp1* ou de *S. globulifera* plantée en bas-fond et en *terra firme*, aurait la même chance de survie. Il semblerait ainsi, que la répartition des écotypes en population naturelle est uniquement due à la facilité de recrutement du *S. sp1* en *terra firme*. Ce résultat que l'on observe sur le trait de la germination n'est pas étonnant pour un arbre, car ces organismes ont un taux de mortalité très important lors de l'établissement des plantules[1].

³ La spéciation incipiente est le début de formation de nouvelles espèces, encore capables de s'hybrider.

SUR LES TRAITS DE CROISSANCE

Le *S. globulifera* se montre plastique quand il est planté en *terra firme*, et à l'ouest de la Guyane. Il montre une hauteur inférieure à celle du *S.sp1* en *terra firme* et similaire à celle du *S.sp1* planté à l'ouest. Dans l'hypothèse que le *S.globulifera* représente l'espèce ancestrale, elle posséderait encore des gènes qui pourraient être à l'origine de la taille du *S. sp1*. Ainsi, quand le *S. globulifera* est planté en *terra firme* ou à l'ouest, en réponse aux caractéristiques environnementale de l'habitat [26-28], il ferait varier l'expression de ses gènes, donnant lieu à un phénotype qui montre une hauteur similaire à celle du *S. sp1*.

ADAPTATION REGIONALE AU SEIN DU COMPLEXES D'ESPECES *S. GLOBULIFERA* ET *S. SP1*

SUR LES TRAITS DE FITNESS

Dans l'interaction provenance régionale x site de plantation, les plantules provenant de l'est et de l'ouest ont un meilleur taux de survie quand elles sont plantées dans leur propre site. Si en 2009, au début de l'expérimentation, il a été choisi de prendre en compte les effets génétiques et effets d'environnements régionaux, cela ne représentait qu'une duplication prenant en compte l'hétérogénéité du climat entre les régions est et ouest de la Guyane. La région ouest possède un sol charbonneux et une pluviométrie moins importante que la région Est, possédant un sol ferrique de type latéritique. Durant sa thèse, Maxime Casalis a pu mettre en évidence, à l'aide de marqueurs chloroplastiques *trnH/psbA* et *trnC/ycf6*, la présence de 34 haplotypes partagés par les deux écotypes, qui montrent une faible différenciation ($F_{ST}=0.015$) ; une plus forte différenciation est détectée par contre entre les provenances est et ouest de la Guyane ($F_{ST}=0.26$). Cette divergence génétique semble avoir un effet sur la survie des plantules et montre une adaptation régionale au sein de l'espèce *Symphonia globulifera*. Au vu des résultats de l'étude de Dick en 2008 [29], montrant la répartition complexe d'haplotypes en Amérique centrale et du sud, ainsi qu'aux Caraïbes, nos résultats vont également dans le sens de cette complexité à l'intérieur de l'espèce et montre une diversité également à un niveau régional, au sein de la Guyane.

Mon analyse a également permis de mettre en évidence des effets inattendus d'interaction Provenance*Habitat sur la survie des plantules. Ainsi, une plantule provenant de l'ouest de la Guyane montre un meilleur taux de survie quand elle est plantée en *terra firme*. Je fais ici l'hypothèse que cela est dû à la teneur en eau du sol. Les plantules provenant de l'ouest, où les précipitations sont moindres et les bas-fonds moins profonds, s'accommoderaient plus facilement à un habitat de type *terra firme* qu'à un habitat de type bas-fond. Ces résultats sont

obtenus sur le trait de la survie, et montrent donc en plus d'une adaptation régionale, une adaptation plus complexe mettant en interaction la provenance régionale et l'habitat.

SUR LES TRAITS DE CROISSANCE

Les trois traits de croissance ont montré un effet d'interaction de la provenance et de l'habitat. Pour la hauteur et le diamètre, les plantules provenant de l'ouest sont les plus grands et larges en bas-fond ; tandis que pour le nombre de feuilles par centimètres, la provenance de l'est arbore plus de feuilles, en *terra firme*. Il est intéressant de remarquer que la hauteur et le diamètre au collet des plantules montrent les mêmes réponses dans l'interaction Provenance régionale*Habitat et l'interaction Ecotypes*Sites de plantations, avec le nombre de feuilles par centimètre qui lui a une réponse opposée à ces derniers traits de croissance.

LIMITES DE L'ETUDE

Dans cette étude, j'ai cherché à observer une adaptation locale dans le cadre d'une expérimentation de transplantations réciproques en jardins naturels. Le *Symphonia globulifera* est une espèce tolérante à l'ombre et possède donc une croissance lente. Sur l'analyse du trait de survie des plantules, après 5 ans d'expérimentation, nous ne pouvons observer qu'une tendance et pas d'interaction Ecotype * habitat. Il paraît donc important, dans ce type d'étude, d'intégrer le trait de la germination dans les analyses. Bien que difficile à intégrer de par le phénomène de dormance, ce trait a permis d'observer une adaptation locale après cinq ans d'étude, plus tôt que le trait de la survie des plantules.

CONCLUSION

En population naturelle, les écotypes du complexe *Symphonia globulifera* / *S. sp1* se distribuent de la façon suivante : le *S.globulifera* se retrouve en bas-fonds, possède des pneumatophores et montre des organes de taille plus importante que le *S.sp1*, rencontré en *terra firme* et ne possédant que très rarement des pneumatophores, remplacés par des racines adventives. Au vu de la faible différenciation génétique entre les écotypes ($F_{ST}=0.03$), malgré la présence de deux morphotypes, le laboratoire a choisi d'effectuer une expérimentation de transplantations réciproques afin de déterminer la part génétique et la part environnementale expliquant le phénotype observé. Dans un modèle linéaire, j'ai alors analysé les interactions génotype*habitat et j'ai mis en évidence des résultats peu attendus, sur le trait de la survie des

plantules qui ont montré une adaptation régionale des écotypes. L'écotype *S. sp1* montre un meilleur taux de survie que le *S. globulifera* quand il est planté à l'Ouest de la Guyane. Ce résultat est tout de même compatible avec l'histoire complexe de l'établissement du *Symphonia globulifera* dans les Néotropiques.

Plus important, j'ai pu mettre en évidence une adaptation locale du *S. sp1* pour la *terra firme* sur le trait de la germination qui peut expliquer la répartition des écotypes dans leurs habitats naturels. Aussi, le suivi des traits de croissance a permis de mettre en évidence des traits descriptifs des écotypes. Le *S. globulifera* montre une circonférence plus importante que l'écotype *S.sp1*, qui possède plus de feuilles par centimètre. Même si un décalage de floraison a déjà été mis en évidence, la différenciation génétique entre les écotypes est faible, suggérant un début de spéciation écologique en présence de flux de gènes, où seules quelques régions du génome sont la cible de la sélection naturelle divergente, le reste du génome étant encore perméable aux flux de gènes[30] ; ou encore il peut s'agir d'une spéciation écologique où le flux de gènes s'est arrêté trop récemment pour que l'on puisse en observer les conséquences sur la différenciation génétique.

TABLES ET FIGURES

Facteurs propres	Ecotype	Effet génétique en réponse à la variation topographique de l'environnement
	Plantation régionale	Effet de l'environnement régional (macro-environnement)
	Habitat	Effet de l'environnement topographique
	Provenance régionale	Effet génétique d'ordre historique
Interaction Génétique x Environnement	Ecotype x Habitat	Renseigne sur la présence d'adaptation locale des écotypes
	Provenance x Plantation	Renseigne sur la présence d'une adaptation macro-environnementale associée au gradient de pluviométrie, par exemple
	Ecotype x Plantation	Renseigne sur la présence d'un effet d'interaction entre l'écotype et le site de plantation régional
	Provenance x Habitat	Renseigne sur la présence d'un effet d'interaction entre la provenance régionale et l'habitat
Interaction Génétique x Génétique	Ecotype x Provenance	Renseigne sur la présence d'un effet de la provenance régionale sur les différences entre les écotypes

Table 3: Récapitulatif des facteurs testés et interprétation de l'effet

Traits	Effets	Df	F statistic	Erreur standart	LR Chisq	Pr(>Chisq)
Herb	Habitat	1	0.0502	0.6994	0.05	NS
	Ecotype	1	45.5676	391.5	45.568	***
	Region de provenance	1	0.1364	0.7461	0.136	NS
	Site de plantation	1	0.3794	0.762	0.379	NS
	Année	1	280.6475	0.2224	280.647	***
	Surface terrière	1	0.0809	0.01922	0.081	NS
	état de plantation	1	0.1686	0.5165	0.169	NS
	Ecotype*Habitat	1	0.1567	0.7444	0.157	NS
	Provenance*Site de plantation	1	0.0006	0.7704	0.001	NS
	Ecotype*Site de plantation	1	0.3232	0.7902	0.323	NS
	Provenance*Habitat	1	0.1182	0.7518	0.118	NS
	Ecotype*Année	1	45.6487	0.1946	45.649	***
	Année*Année de germination	1	0.4991	0.0000801	0.499	NS
Survie	Habitat	1	0.5993	0.305586	0.5917	NS
	Ecotype	1	0.4623	183.322798	0.4564	NS
	Région de provenance	1	19.9048	0.333325	19.6501	***
	Site de plantation	1	13.699	0.312327	13.5237	***
	Année	1	29.8126	0.072356	29.4311	***
	Surface terrière	1	1.402	0.0078	1.3841	NS
	Ecotype*Habitat	1	0.3546	0.345644	0.3501	NS
	Provenance*Site de plantation	1	9.8514	0.417447	9.7253	**
	Ecotype*Site de plantation	1	4.1852	0.354539	4.1316	*
	Provenance*Habitat	1	15.3986	0.410148	15.2015	***
Ecotype*Année	1	0.4678	0.091122	0.4618	NS	
Germ	Habitat	1	2.0842	0.83313	2.4319	NS
	Ecotype	1	2.0755	361.00409	2.4218	NS
	Année	1	0.7562	0.56666	0.8824	NS
	Région de provenance	1	0.705	0.03473	0.8226	NS
	Surface terrière	1	3.5626	0.1492	4.1571	*
	Ecotype*Habitat	1	3.6908	0.58443	4.3066	*
	Provenance*Habitat	1	14.7962	0.72803	17.265	***
	Ecotype*Année	1	2.084	0.17957	2.4317	NS

Table 4: Table d'Anova pour les traits de fitness. Dans ce tableau sont rapportées les valeurs des tests statistiques pour les traits de l'herbivorie (Herb), pour le trait de la survie (Survie) et pour le trait de la germination (Germ).

Traits	Effets	Df	F statistic	Erreur standart	LR Chisq	Pr(>Chisq)
H	Habitat	1	3.9165	1.08	3.916	*
	Ecotype	1	1.6407	628.8	1.641	NS
	Region de provenance	1	0.993	1.199	0.993	NS
	Site de plantation	1	6.9303	1.182	6.93	**
	Année	1	89.2659	0.3746	89.266	***
	Surface terrière	1	1.6468	0.03023	1.647	NS
	état de plantation	1	32.0032	0.7979	32.003	***
	Ecotype*Habitat	1	4.2822	1.166	4.282	*
	Provenance*Site de plantation	1	1.0274	1.211	1.027	NS
	Ecotype*Site de plantation	1	9.7104	1.233	9.71	**
	Provenance*Habitat	1	7.238	1.186	7.238	**
	Ecotype*Année	2	1.6327	0.3126	1.633	NS
	Année*Année de germination	1	5.939	0.0001327	5.939	*
D	Habitat	1	0.0291	0.1398	0.029	NS
	Ecotype	1	5.8378	81.39	5.838	*
	Region de provenance	1	8.063	0.1552	8.063	**
	Site de plantation	1	2.0289	0.153	2.029	NS
	Année	1	125.4528	0.04849	125.453	***
	Surface terrière	1	2.2841	0.003912	2.284	NS
	état de plantation	1	10.2119	0.1033	10.212	**
	Ecotype*Habitat	1	1.8144	0.1509	1.814	NS
	Provenance*Site de plantation	1	1.543	0.1567	1.543	NS
	Ecotype*Site de plantation	1	7.5239	0.1596	7.524	**
	Provenance*Habitat	1	15.0193	0.1535	15.019	***
	Ecotype*Année	2	5.8898	0.04046	5.89	*
	Année*Année de germination	1	5.0072	0.00001717	5.007	*
TNLH	Habitat	1	3.8792	0.02961	3.8792	*
	Ecotype	1	6.1236	17.22	6.1236	*
	Region de provenance	1	1.7978	0.03277	1.7978	NS
	Site de plantation	1	0.0041	0.03249	0.0041	NS
	Année	1	0.0856	0.01032	0.0856	NS
	Surface terrière	1	0.1369	0.0008268	0.1369	NS
	état de plantation	1	10.2352	0.02181	10.2352	**
	Ecotype*Habitat	1	0.6164	0.03192	0.6164	NS
	Provenance*Site de plantation	1	3.9817	0.03315	3.9817	*
	Ecotype*Site de plantation	1	4.8396	0.03385	4.8396	*
	Provenance*Habitat	1	3.3077	0.03243	3.3077	.
	Ecotype*Année	2	6.2058	0.008559	6.2058	*
	Année*Année de germination	1	0.2415	0.000003656	0.2415	NS

Table 5: Table d'Anova pour les traits de croissance. Dans ce tableau sont rapportées les valeurs des tests statistiques pour les traits de la hauteur (H), pour le trait du diamètre au collet (D) et pour le trait du nombre de feuilles par centimètre (TNLH)

Figures

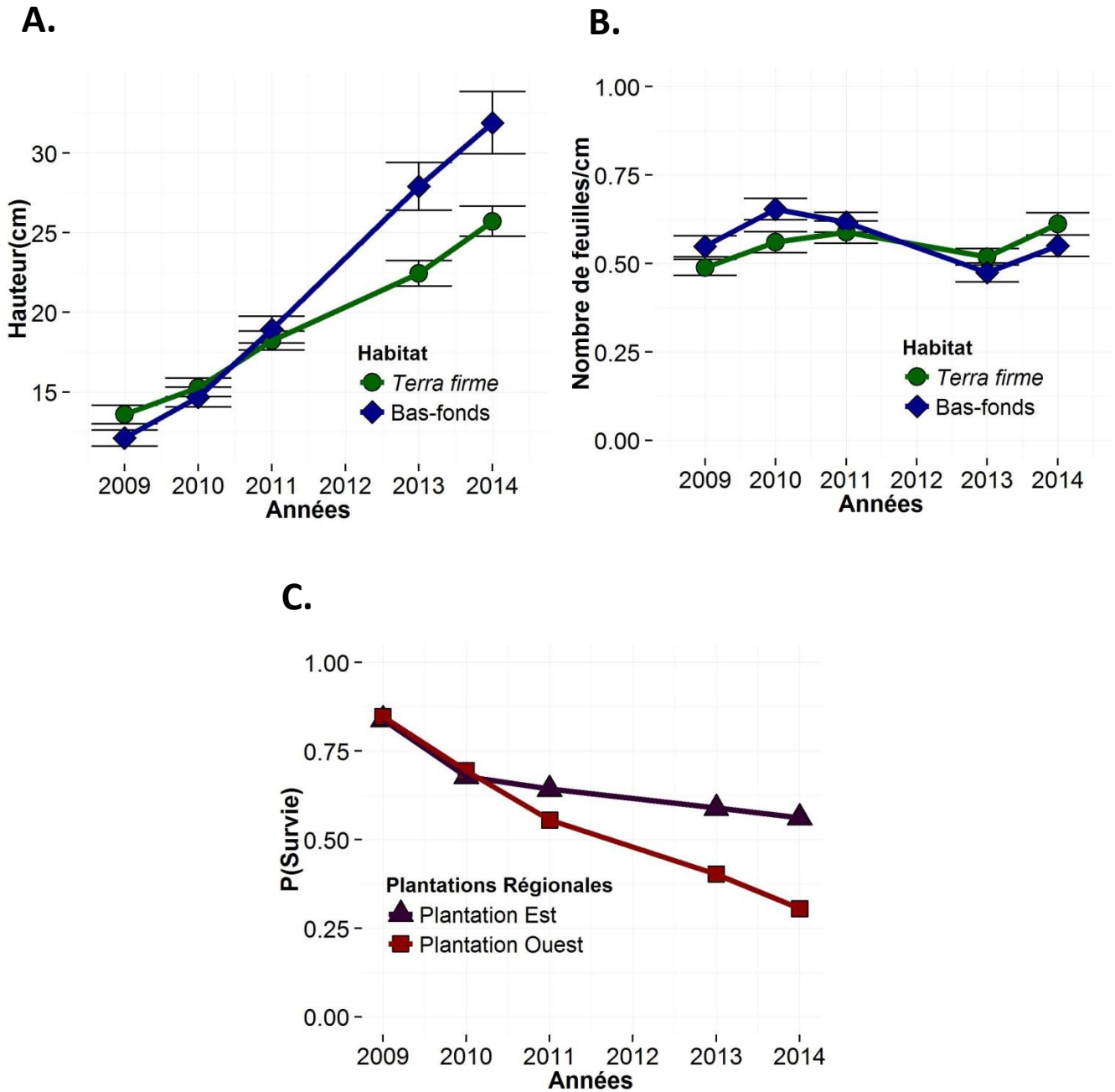


Figure 3: Effets environnementaux sur les traits de croissances et de fitness. **A.** La hauteur, **B.** Le nombre de feuille par centimètre, **C.** La survie des plantules

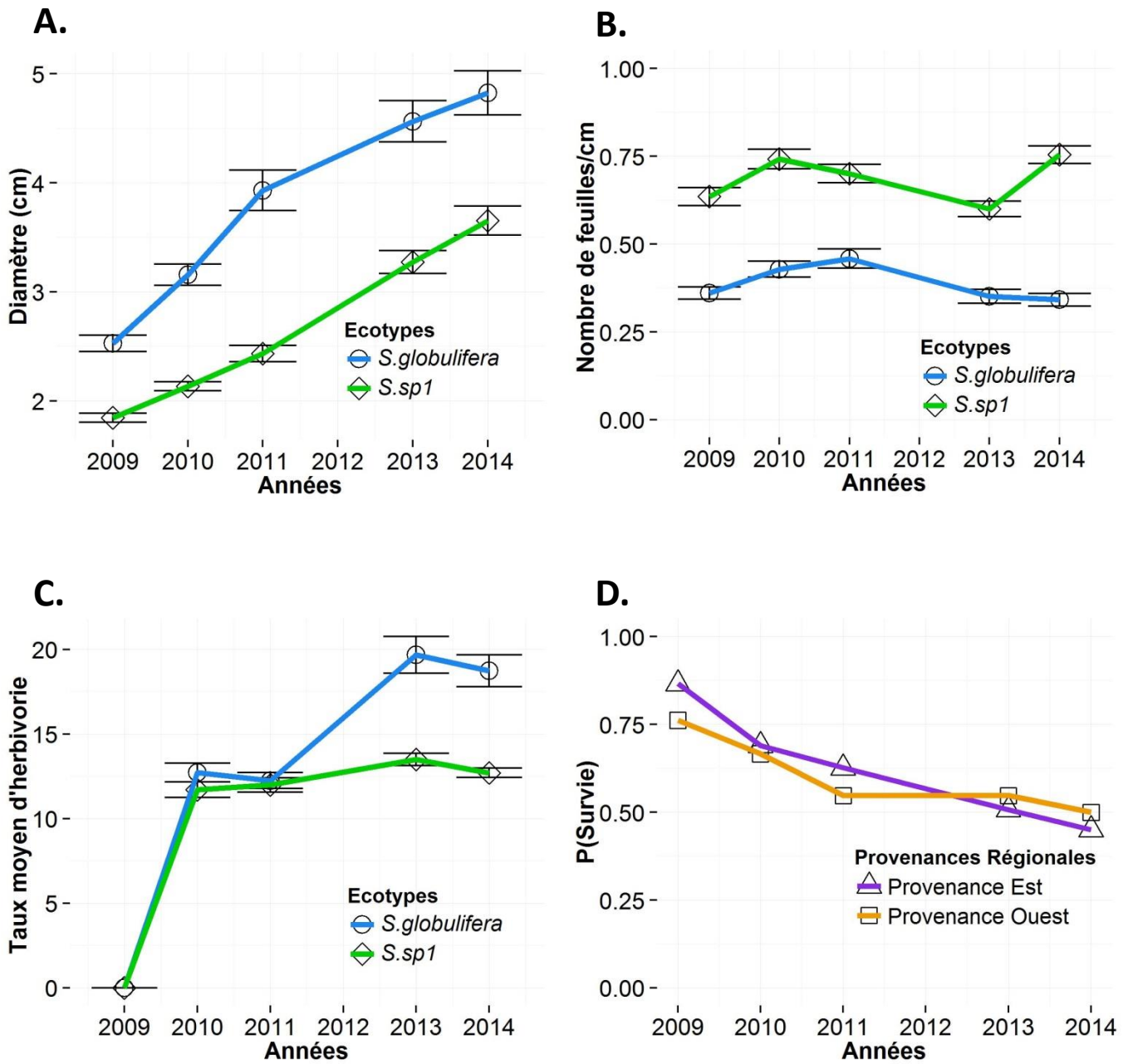


Figure 4: Effets génétiques sur les traits de croissances et de fitness. **A.** Le diamètre, **B.** Le nombre de feuille par centimètre, **C.** Le taux d'herbivorie et **D.** La survie des plantules

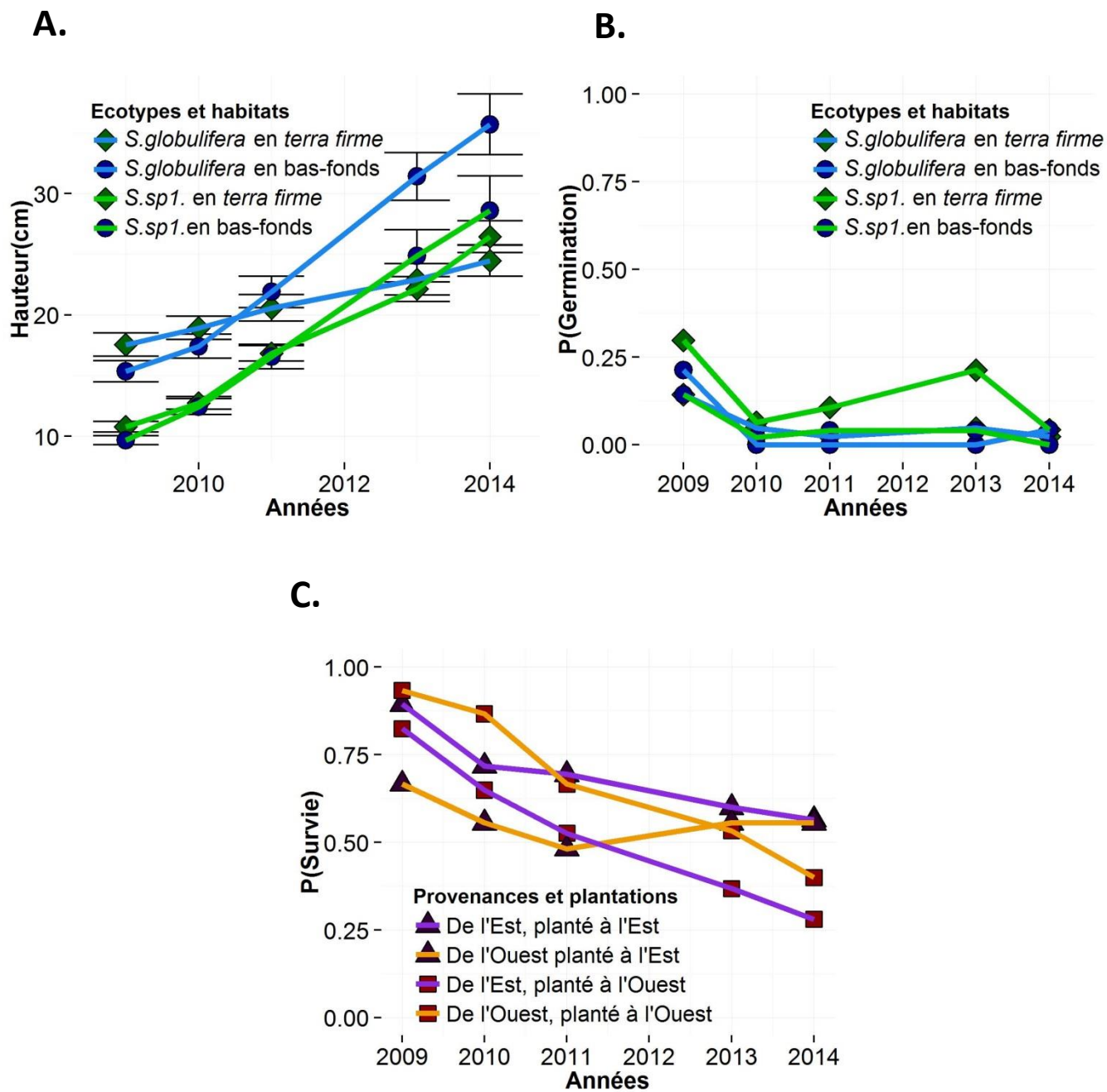


Figure 5: Effets d'interactions sur les traits de croissances et de fitness. A. La hauteur, B. La germination et C. La survie des plantules

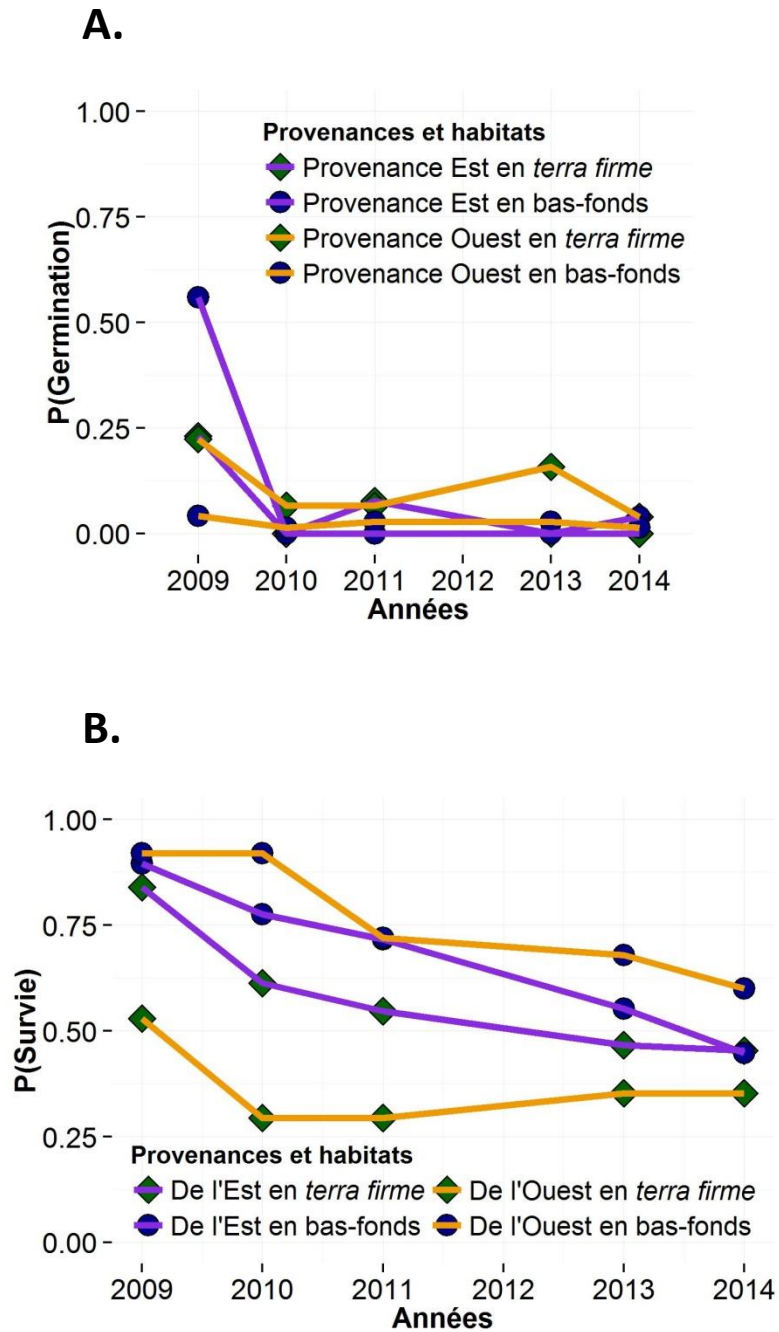


Figure 7: Effet d'interactions Provenance Régionale*Habitats sur la valeur des traits de fitness. A. la germination, B. La survie des plantules.

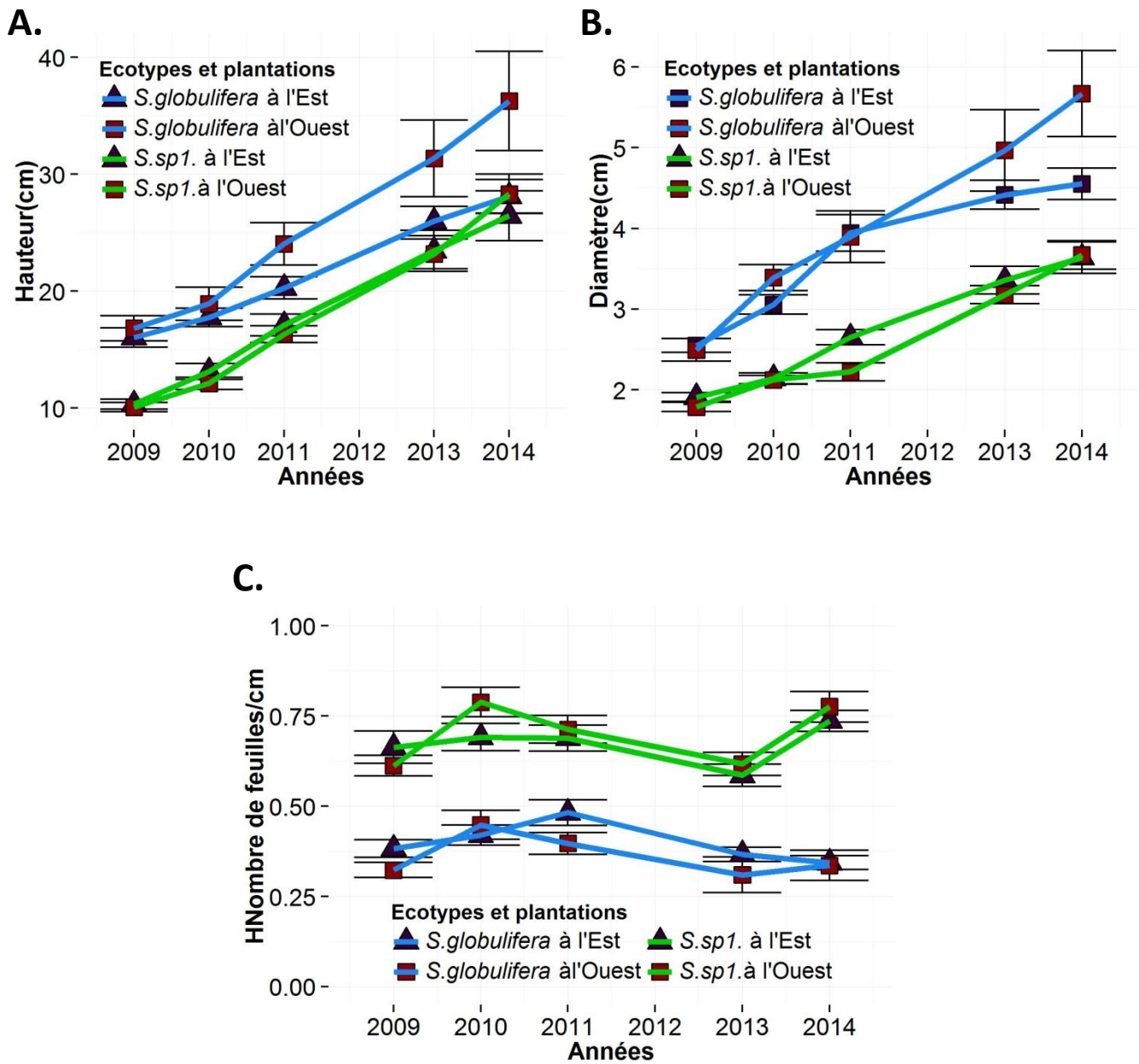


Figure 8: Effet d'interactions Ecotypes*Sites de plantations sur la valeur des traits de croissance. **A.** La hauteur, **B.** Le diamètre, **C.** Le nombre de feuilles par centimètres des plantules, dans l'interaction Ecotypes*Sites de plantations.

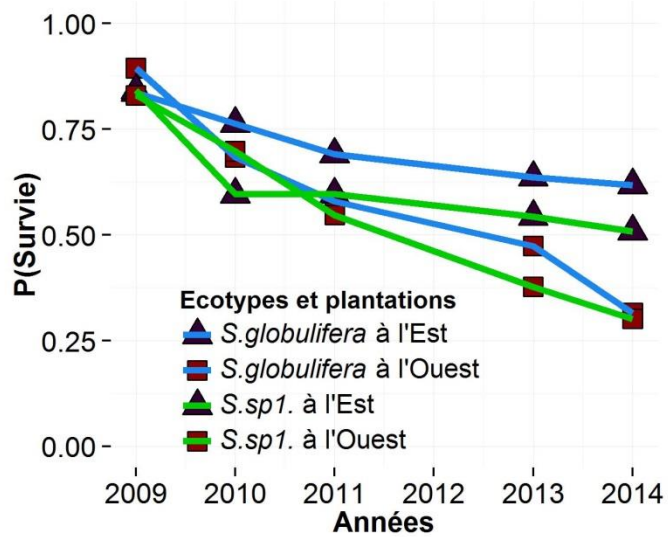


Figure 9: Effet d'interactions Ecotypes*Sites de plantations sur la survie des plantules.

BIBLIOGRAPHIE

1. Petit, R.J. and A. Hampe, *Some Evolutionary Consequences of Being a Tree*. Annual Review of Ecology, Evolution, and Systematics, 2006. **37**(1): p. 187-214.
2. Hamrick, J.L., D.A. Murawski, and J. Nason, *The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations*. Vegetatio, 1993. **107/108**: p. 281-297.
3. Hedrick, P., *Genetics of populations*. 2nd ed. 2000, Sudbury, MS: Jones and Bartlett.
4. Schupp, E.W., *Seed-Seedling Conflicts, Habitat Choice, and Patterns of Plant Recruitment*. American Journal of Botany, 1995. **82**(3): p. 399-409.
5. Slot, M. and L. Poorter, *Diversity of tropical tree seedling responses to drought*. Biotropica, 2007. **39**(6): p. 683-690.
6. Louda, S.M., *Distribution ecology: variation in plant recruitment over a gradient in relation to insect seed predation*. Ecological monographs, 1982. **52**(1): p. 25-41.
7. Bricker, M., D. Pearson, and J. Maron, *Small-mammal seed predation limits the recruitment and abundance of two perennial grassland forbs*. Ecology, 2010. **91**(1): p. 85-92.
8. Breed, M., et al., *Which provenance and where? Seed sourcing strategies for revegetation in a changing environment*. Conservation genetics, 2013. **14**(1): p. 1-10.
9. Bower, A.D., J.B.S. Clair, and V. Erickson, *Generalized provisional seed zones for native plants*. Ecological applications, 2014. **24**(5): p. 913-919.
10. Baraloto, C. and P. Coutron, *Fine-scale Microhabitat Heterogeneity in a French Guianan Forest*. Biotropica, 2010. **42**(4): p. 420-428.
11. Baraloto, C., D. Goldberg, and D. Bonal, *Performance trade-offs among tropical tree seedlings in contrasting microhabitats*. Ecology, 2005. **86**(9): p. 2461-2472.
12. Baraloto, C., et al., *Seasonal water stress tolerance and habitat associations within four neotropical tree genera*. Ecology, 2007. **88**(2): p. 478-489.
13. Kawecki, T.J. and D. Ebert, *Conceptual issues in local adaptation*. Ecology letters, 2004. **7**(12): p. 1225-1241.
14. Leimu, R. and M. Fischer, *A Meta-Analysis of Local Adaptation in Plants*. PLOS One, 2008. **3**(12): p. e4010.
15. Nicotra, A.B., et al., *Plant phenotypic plasticity in a changing climate*. Trends in Plant Science, 2010. **15**(12): p. 684-692.
16. Gratani, L., *Plant Phenotypic Plasticity in Response to Environmental Factors*. Advances in Botany, 2014. **2014**: p. 17.
17. Palacio-López, K., et al., *The ubiquity of phenotypic plasticity in plants: a synthesis*. Ecology and Evolution, 2015. **5**(16): p. 3389-3400.
18. Brousseau, L., et al., *Highly local environmental variability promotes intrapopulation divergence of quantitative traits: an example from tropical rain forest trees*. Ann Bot, 2013. **112**(6): p. 1169-79.
19. Rundle, H.D. and P. Nosil, *Ecological speciation*. Ecology letters, 2005. **8**(3): p. 336-352.
20. Schluter, D., *Ecology and the origin of species*. Trends in ecology & evolution (Personal edition), 2001. **16**(7): p. 372-380.
21. Degen, B., E. Bandou, and H. Caron, *Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana*. Heredity, 2004. **93**: p. 585-591.
22. Nosil, P., *Ecological Speciation*. 2012: Oxford University Press.
23. Hendry, A.P., P. Nosil, and L.H. Rieseberg, *The speed of ecological speciation*. Functional ecology, 2007. **21**(3): p. 455-464.

24. Seehausen, O., *Integrating ecology and genetics in speciation research*. Trends in Ecology & Evolution. **28**(1): p. 12-13.
25. Loubry, D., *Déterminisme du comportement phénologique des arbres en forêt tropical humide de Guyane française (5°lat. N.)*. PhD Thesis, University of Paris 6, 1994.
26. Dayan, D.I., D.L. Crawford, and M.F. Oleksiak, *Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of *Fundulus heteroclitus**. Molecular Ecology, 2015. **24**(13): p. 3345-3359.
27. Healy, T.M. and P.M. Schulte, *Phenotypic plasticity and divergence in gene expression*. Molecular Ecology, 2015. **24**(13): p. 3220-3222.
28. Meier, K., et al., *Local Adaptation at the Transcriptome Level in Brown Trout: Evidence from Early Life History Temperature Genomic Reaction Norms*. PLOS One, 2014. **9**(1): p. e85171.
29. Dick, C.W. and M. Heuertz, *The complex biogeographic history of a widespread tropical tree species*. Evolution, 2008. **62**(11): p. 2760-2774.
30. Wu, C.-I., *The genic view of the process of speciation*. Journal of Evolutionary biology, 2001. **14**(6): p. 851-865.

PARTIE II: GENOTYPAGES
ET METHYLATIONS
CHEZ DES ARBRES ADULTES DE
SYMPHONIA GLOBULIFERA

Introduction.....	77
Matériels et méthodes.....	79
Echantillonnage	79
Extraction ADN	79
Digestion enzymatique, ligation et amplification PCR.....	81
Analyse de répétabilité et géotypage.....	83
Analyses statistiques de la divergence	83
Test sur les fréquences alléliques	84
Détection des locus soumis à sélection	85
Résultats	86
Comparaison du Delta de fréquence entre les enzymes.....	87
Test sur les phénotypes : Analyse en composantes Principales	87
Détection des locus soumis à sélection	88
Discussion.....	89
Fréquences de bandes dans les populations et les écotypes	89
Variation épigénétique entre les écotypes et les populations	89
Variation génétique entre les écotypes et les populations.....	90
Conclusion	91
Tables et figures	92
Bibliographie	98

INTRODUCTION

Même s'il existe quelques exceptions, comme le ver *Caenorhabditis elegans* ou la levure *Saccharomyces Cerevisiae*, la méthylation des cytosines de l'ADN est un mécanisme très conservé de la régulation de l'expression des gènes, autant chez les eucaryotes que chez les procaryotes ; mais la proportion de la méthylation du génome est très variable entre les espèces [1, 2]. Il existe plusieurs motifs de méthylation du génome, en fonction des nucléotides qui suivent la cytosine. Le plus conservé est le motif CG, aussi appelé CpG. Il est présent en plus grande proportion chez les plantes que chez d'autres organismes[2] et est héritable [3-6], contrairement aux motifs CHH et CHG (où H représente une Adénine, une Thymine ou une Cytosine).

L'adaptation locale est souvent associée à l'établissement de nouveaux phénotypes [7-14], qui ne peuvent pas toujours être reliés à la divergence des fréquences alléliques, dans le cas des écotypes avec peu ou pas de différenciation génétique. Les marques de méthylation de l'ADN varient en fonction des caractéristiques de l'environnement, et induisent de nouveaux phénotypes [2, 15-17] et des organismes adaptés à leur environnement. Dans ce contexte-là, les marques épigénétiques pourraient jouer un rôle décisif dans l'adaptation des espèces[6, 18-21] en permettant la colonisation de nouveaux environnements[22] ou en permettant la persistance de l'espèce par une adaptation rapide dans la génération subissant une variation de l'environnement, et un transfert vertical induisant l'adaptation locale des générations futures [3, 23, 24].

La technique MSAP (*Methyl Sensitive Amplification Polymorphism*) est un génotypage AFLP utilisant des isoschizomères sensibles à la méthylation des îlots CpG. Un couple d'enzymes de restriction, *HpaII* et *MspI*, reconnaît la même séquence tétranucléotidique 5'-CCGG-3' [25, 26]. Selon le profil de méthylation des séquences, on observe un pattern de coupe et donc d'amplifications des ADNs qui est différent [25, 27, 28]. Les deux enzymes sont sensibles à des profils de méthylation différents (Figure 1.). Une méthylation sur le C interne (CmCGG) va empêcher *HpaII* de couper le brin d'ADN [26], n'empêchant pas *MspI* de couper (Figure 1. C.). *MspI* digère les séquences non-méthylées ainsi que les séquences *hémi*-méthylées ou méthylées sur le C interne (CmCGG), sur les deux brins complémentaires. A contrario, une méthylation sur le C externe (mCCGG) empêchera *MspI* de couper le brin d'ADN [29, 30], mais pas *HpaII* qui est capable de digérer des séquences non-méthylées ou des séquences *hémi* méthylées sur le C externe (mCCGG) (Figure 1. B.). Le MSAP a déjà permis de différencier

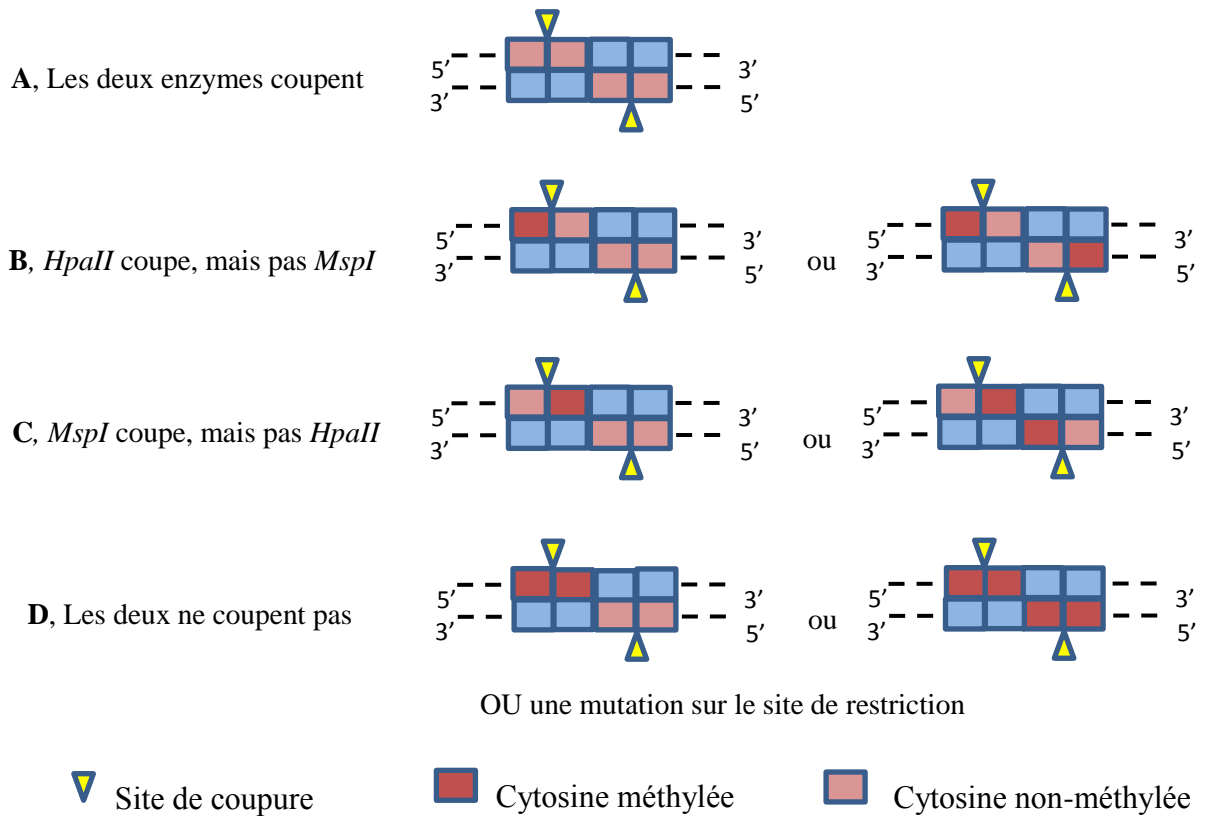


Figure 1. Pattern de digestion des enzymes *MspI* et *HpaII*; sur le site de restriction 5'-CCGG-3', en fonction de l'état de méthylation des cytosines.

des écotypes montrant peu ou pas de différenciation génétique [22, 31-36]. Aussi, une étude récente a utilisé le séquençage génomique de l'ADN au bisulfite [37], qui permet la distinction des cytosines méthylées et non-méthylées. Cette étude a montré que les forces de la sélection naturelle agissent non seulement sur la variation de séquences, mais aussi sur les sites CpG de la méthylation de l'ADN. Ainsi, les sites de variation de l'état de la méthylation peuvent contribuer à l'adaptation locale.

Dans mon cas d'étude, j'ai auparavant (Chapitre 1) montré la présence d'une adaptation locale de l'écotype *S.Sp1* à la *terra firme* et la présence de traits morphologiques divergents entre les populations, malgré une faible différenciation génétique (M. Casalis, *non publié*). En échantillonnant des arbres adultes sur les sites de Paracou et Régina, je veux à présent tester la présence d'un patron de méthylation différent entre les écotypes, en population naturelle, et savoir dans quelle mesure les variants génétiques et les sites de méthylation des cytosines sont impliqués dans l'adaptation locale du *S.globulifera* aux bas-fonds et du *S.sp1* à la *terra firme*. Dans ce contexte exploratoire, l'utilisation de la technique MSAP est particulièrement intéressante.

MATERIELS ET METHODES

ECHANTILLONNAGE

Sur les sites de Régina et de Paracou, de Mars à Avril 2014, 61 arbres ont été échantillonnés (Table 1.). Avec la participation d'un grimpeur professionnel, il a pu être récolté des feuilles fraîches, à même l'arbre. Dans l'objectif de limiter au maximum les changements d'état de méthylation associés au stress[24], les feuilles ont tout de suite été placées par individu dans un sac *easy-lock* rempli au tiers d'eau glacée. Par la suite, les échantillons ont été transportés en sac isotherme jusqu'à l'arrivée au laboratoire où ils ont été rincés à l'eau, essuyés avec du papier absorbant puis surcongelés quelques secondes dans de l'azote liquide avant d'être conservés dans un congélateur à -80°C.

EXTRACTION ADN

Pour tous les individus, il a été extrait deux doublons. Pour chacun des doublons, il a été pesé 150 à 200mg de feuilles qui ont permis des extractions ADN au CTAB[38]. En moyenne, il a été obtenu 1mg d'ADN par individus.

ID	Ecotype	Site	Parcelle	Carré	Numéro_arbre	X (UTM)	Y (UTM)
BW04	S.globulifera	Paracou	P13	2	931	NA	NA
FQ92	S.globulifera	Paracou	P13	1	922	NA	NA
FQ91	S.globulifera	Paracou	P13	1	1085	NA	NA
FR09	S.globulifera	Paracou	P13	1	782	NA	NA
FB19	S.globulifera	Paracou	P14	2	318	NA	NA
FB22	S.globulifera	Paracou	P14	2	459	NA	NA
BW13	S.globulifera	Paracou	P14	2	67	NA	NA
FQ93	S.globulifera	Paracou	P14	1	14	NA	NA
FQ96	S.globulifera	Paracou	P14	1	576	NA	NA
BW10	S.globulifera	Paracou	P14	1	587	NA	NA
BW22	S.globulifera	Paracou	P14	4	770	NA	NA
BW47	S.globulifera	Paracou	P15	1	198	NA	NA
FB24	S.globulifera	Paracou	P15	1	575	NA	NA
FB25	S.globulifera	Paracou	P15	1	637	NA	NA
FR04	S.globulifera	Paracou	P15	4	1	NA	NA
FR05	S.globulifera	Paracou	P15	4	235	NA	NA
BW40	S.globulifera	Paracou	P15	4	938	NA	NA
BW41	S.globulifera	Paracou	P15	4	1098	NA	NA
FQ83	S.globulifera	Régina	NA	NA	NA	362676	476839
FQ88	S.globulifera	Régina	NA	NA	NA	362624	476737
FQ90	S.globulifera	Régina	NA	NA	NA	362544	476675
FQ89	S.globulifera	Régina	NA	NA	NA	362583	476676
FQ49	S.globulifera	Régina	NA	NA	NA	362834	477278
FQ51	S.globulifera	Régina	NA	NA	NA	362878	477174
FQ52	S.globulifera	Régina	NA	NA	NA	362905	477188
FQ48	S.globulifera	Régina	NA	NA	NA	362848	477180
FQ54	S.globulifera	Régina	NA	NA	NA	362889	477084
FQ56	S.globulifera	Régina	NA	NA	NA	362924	477008
FQ47	S.globulifera	Régina	NA	NA	NA	362852	477056
FQ46	S.globulifera	Régina	NA	NA	NA	362848	477048
FQ72	S.globulifera	Régina	NA	NA	NA	363018	476304
FQ69	S.globulifera	Régina	NA	NA	NA	363037	476415
FQ67	S.globulifera	Régina	NA	NA	NA	362924	476587
FQ71	S.globulifera	Régina	NA	NA	NA	363032	476358
FR08	S.sp1.	Paracou	P13	4	654	NA	NA
FR12	S.sp1.	Paracou	P13	4	149	NA	NA
FR10	S.sp1.	Paracou	P13	3	113	NA	NA
FR13	S.sp1.	Paracou	P13	1	235	NA	NA
FR11	S.sp1.	Paracou	P13	1	655	NA	NA
FQ97	S.sp1.	Paracou	P14	2	87	NA	NA
FQ94	S.sp1.	Paracou	P14	1	311	NA	NA
FQ95	S.sp1.	Paracou	P14	1	495	NA	NA
BW07	S.sp1.	Paracou	P14	3	445	NA	NA
FQ99	S.sp1.	Paracou	P14	4	651	NA	NA
FQ98	S.sp1.	Paracou	P14	4	460	NA	NA
FR01	S.sp1.	Paracou	P15	2	669	NA	NA
FR00	S.sp1.	Paracou	P15	2	1093	NA	NA
FR03	S.sp1.	Paracou	P15	4	13	NA	NA
FR06	S.sp1.	Paracou	P15	4	40	NA	NA
FR07	S.sp1.	Paracou	P15	4	458	NA	NA
FR02	S.sp1.	Paracou	P15	4	1118	NA	NA
FQ87	S.sp1.	Régina	NA	NA	NA	362510	476796
FQ60	S.sp1.	Régina	NA	NA	NA	363004	476888
FQ81	S.sp1.	Régina	NA	NA	NA	362879	476455
FQ78	S.sp1.	Régina	NA	NA	NA	362993	476388
FQ80	S.sp1.	Régina	NA	NA	NA	362887	476438
FQ82	S.sp1.	Régina	NA	NA	NA	362926	476471
FQ74	S.sp1.	Régina	NA	NA	NA	362972	476284
FQ70	S.sp1.	Régina	NA	NA	NA	363093	476388
FQ73	S.sp1.	Régina	NA	NA	NA	362999	476280
FQ76	S.sp1.	Régina	NA	NA	NA	362919	476370

Table 1. Liste et emplacements des arbres échantillonnés. Les emplacements des arbres de la parcelle Paracou sont représentés par des codes par parcelle, puis par carré dans la parcelle et enfin par arbres. Les emplacements des arbres de la parcelle de Régina sont donnés par les points GPS en format UTM.

Après une incubation de 15mn à 37°C, en présence de RNase (10 ng/μl), la propreté de l'ADN est vérifié par électrophorèse sur gel d'agarose 1% (0,7g d'agarose, 70mL de TAE 0,5X, 0,7μL de Gel Red) qui a migré durant 30 à 40min à 100V/cm. Ensuite, les concentrations et ratio 260/280 des échantillons ont été déterminées sur Nanodrop (Thermo Fisher). Les échantillons qui ont servis pour les amplifications avaient un ratio 260/280 de 1.7 au minimum.

DIGESTION ENZYMATIQUE, LIGATION ET AMPLIFICATION PCR

Les premiers articles utilisant la digestion enzymatique dans le but d'étudier la méthylation d'ADN, datent de la fin des années 1990 [25], peu après le développement de la technique de génotypage AFLP [27, 39]. En suivant ces protocoles, pour chaque individu, 500ng d'ADN sont utilisés pour la digestion enzymatique, par les couples d'amorces *EcoR1-MseI*, *EcoR1-MspI* et *EcoR1-HpaII*. Il a été utilisé 3U pour *EcoR1* et 0.5U pour *MseI*, *MspI* et *HpaII*. Après une digestion de 4h à 37°C, la ligation par les adaptateurs a été faite à 5 μM pour *EcoR1* et à 50μM pour les autres enzymes. A suivi une incubation à 100°C durant 10mn. Ensuite, les échantillons ont été laissés à refroidir jusqu'à température ambiante.

L'amplification pré-sélective a été effectuée avec des adaptateurs à 0.33μM dans le mix PCR selon le cycle thermique suivant : 94°C durant 5mn ; 30 cycles de 94°C durant 30s ; 56°C durant 30s ; 72°C durant 2mn et une extension finale à 72°C durant 15mn. Enfin, pour un total de 12 combinaisons pour *MspI* et *HpaII*, et 6 combinaisons pour *MseI* (Figure 1. A. et B.), la PCR sélective a été effectuée avec 0.046μM de *EcoR1*+AAG marqué au fluorochrome FAM (bleu), *EcoR1*+ACC marqué au fluorochrome HEX (vert) et *EcoR1*+ATG marqué au fluorochrome NED (jaune). La PCR a été effectuée selon le programme suivant : 94°C durant 5mn ; 12 cycles de 94°C durant 30s suivi de 1mn de 65°C à 78.3°C avec un pas de 0.8°C par cycles et d'une polymérisation à 72°C durant 2mn ; 22 cycles de 94°C durant 30s, 56°C durant 1mn, 72°C durant 2 mn ; avant une extension finale de 72°C durant 15mn.

A.

		EcoRI						
		A			C			
		AAG	ACC	ATG	CAA	CCC	CTG	
MspI ou HpaII	A	AAT	X	X	X			
		AGT	X	X	X			
	C	CAT				X	X	X
		CTT				X	X	X

B.

		EcoRI						
		A			C			
		AAG	ACC	ATG	CAA	CCC	CTG	
MseI	A	AAT	X	X	X			
		AGT	X	X	X			

Figure 2. Présentation des combinaisons utilisées.

A., Les profils de digestion des enzymes *MspI* et *HpaII* ont été amplifiés selon 12 combinaisons. **B.**, Les profils de digestion de l'enzyme *MseI* ont été amplifiés selon 6 combinaisons

Le génotypage a été effectué sur ABIPRISM 3130XL dans le polymère pop7 avec le marqueur de taille GZ500LIZ_3130.

La lecture des amplifications a été effectuée de 50 à 500pb à l'aide du logiciel GeneMapper. Pour chaque *run* et chaque couleur, une hauteur minimale de détection de la présence de bande a été établie, en deçà de laquelle tous les pics sont mis à zéro (considérés comme du bruit de fond). En moyenne, sur l'ensemble des *runs*, sous la hauteur minimum de 10, les pics n'ont pas été pris en compte pour les analyses.

ANALYSE DE REPETABILITE ET GENOTYPAGE

La répétabilité des amplifications entre les doublons a été déterminée selon la méthode utilisée par Ley AC en 2013 [40]. Le calcul d'un F_{ST} [41] pour chaque marqueurs obtenu avec le logiciel de structuration spatiale SPAGEDI [42] permet de déterminer un degré d'héritabilité des marqueurs.

Un marqueur est répétable s'il est reproductible entre les doublons. Il est parfaitement reproductible si la variation de la présence ou absence de bandes n'est due qu'à une variation génétique. Ainsi, un locus est parfaitement répétable s'il possède une héritabilité (H^2) de 1.

A l'image du F_{ST} qui est le ratio entre la variance expliquée par l'effet de la population et la variance totale.

$$F_{ST} = V_{entre\ les\ populations} / (V_{entre\ les\ populations} + V_{à\ l'interieur\ des\ populations})$$

La répétabilité est le ratio de la variance expliquée entre les individus et la variance totale, qui inclue la variance environnementale (extraction ADN, position sur la plaque..)

$$H^2 = V_{entre\ les\ individus} / (V_{entre\ les\ individus} + V_{intra\ individus})$$

En considérant les doublons comme des individus à l'intérieur des populations, et les individus comme des populations, le calcul du F_{ST} par SPAGEDI permet d'avoir une estimation de l'héritabilité à chaque locus. Les locus pour lesquels la F_{ST} avait une valeur strictement inférieure à 0.3 avec une p-value supérieure ou égale à 0.05, ont été exclus de l'analyse. En dessous de cette valeur seuil de 0.3, le bruit est trop important.

ANALYSES STATISTIQUES DE LA DIVERGENCE

Afin d'éliminer les risques d'homoplasie, qui ne traduit pas l'héritabilité des caractères, les locus dont la taille étaient inférieure à 70pb n'ont pas été pris en compte dans les analyses.

Aussi, les locus ayant une fréquence de bande inférieure à 0.03 ont été sortis du jeu de données d'analyse.

TEST SUR LES FREQUENCES ALLELIQUES

Afin de mettre en évidence des profils d'amplification différents entre les écotypes, entre les populations et entre les enzymes, pour chaque locus obtenus par la digestion des enzymes *MseI*, *MspI* et *HpaII*, il a été calculé les fréquences de présence de bandes. Aussi les différences de fréquences (appelées Delta de fréquences par la suite) entre écotypes et entre populations ont été calculées pour chaque locus, et pour chaque enzyme, permettant de comparer la faculté des enzymes à différencier les populations et les écotypes. Sur ces données de fréquences et de Delta de fréquences, des tests de rangs de Wilcoxon, appariés dans le cas des analyses de fréquences de bandes, ont permis de tester la différence de distributions entre écotypes, entre populations et entre enzymes pour les écotypes et les populations de l'est et de l'ouest de la Guyane.

Ainsi, j'ai appliqué un test de rang de Wilcoxon apparié avec l'option « alternative=greater » en testant l'hypothèse d'une plus grande fréquence de présence de bande pour Régina par rapport à Paracou et pour le *S.sp1* par rapport au *S.globulifera*. Ensuite, j'ai effectué des tests non appariés sur les Delta de fréquence entre Paracou et Régina (Freq(Paracou)-Freq(Régina)) et entre *S.sp1* et *S.globulifera* (Freq(*S.sp1*)-Freq(*S.globulifera*)), en testant l'hypothèse que l'enzyme *MseI* a un Delta plus petit que les enzymes méthylsensibles, puis que *MspI* a un Delta de fréquence plus petit que *HpaII*.

Aussi, afin d'observer les relations entre les individus, expliquées par les profils d'amplifications obtenus par chaque enzyme, il été effectué des Analyses en Composantes Principales (ACP) sur les données de présence (1) et d'absence de bandes (0) pour les locus de chaque enzymes, à l'aide du packages R « adegenet »[43]. Pour chaque enzyme, les données manquantes ont été remplacées par un arrondi à l'unité de la moyenne des valeurs 1 et 0 sur l'ensemble des individus pour le locus concerné. Avec la fonction « stat_ellipse » du package « ggplot2 » qui se base sur une loi de Student multivariée, des ellipses ont été dessinées pour observer la dissimilarité entre les groupes.

DETECTION DES LOCUS SOUMIS A SELECTION

Afin de déterminer la contribution des locus à variation de séquences et les locus sensibles à la méthylation dans l'adaptation locale du *S.globulifera* à la *terra firme*, il a été créé un jeu de données contenant tous les locus obtenus pour les trois enzymes, pour les deux sites (Paracou et Régina).

La détection des locus sélectionnés a été effectuée à l'aide de l'algorithme DFDIST[44, 45] incorporé dans la plateforme Mcheza [46]. Les locus qui montrent une différenciation génétique et qui s'écarte de l'intervalle de confiance à 99,5% de la distribution obtenue avec un modèle neutre (modèle en îles) sont potentiellement soumis à sélection (44 48). Plutôt que d'utiliser le F_{ST} seul pour la recherche d'outliers, DFDIST utilise la relation entre le F_{ST} et H_e . Le F_{ST} est un estimateur de la différenciation génétique entre populations, tandis que le taux d'hétérozygotie H_e renseigne sur la proportion d'hétérozygotes attendue dans une population et permet de tenir compte de la diversité génétique intra population[47] [44, 48]. A partir des fréquences alléliques calculées par DFDIST [49], des modèles en île sont simulés par coalescence pour calculer un nuage neutre, avec Θ , le taux de migration entre les îles, fixés à 0.04. Les locus n'appartenant pas au nuage neutre sont les locus potentiellement soumis à sélection. 50 000 itérations ont été utilisées pour construire l'intervalle de confiance à 99.5 % de l'enveloppe neutre et les *outliers* ont été acceptés avec un taux de faux positif (FDR) de 0.01. Pour l'analyse sur les deux sites, la moyenne simulée a été forcée à se rapprocher de la moyenne du jeu de données empiriques.

RESULTATS

TEST SUR LES FREQUENCES DE BANDES EN FONCTION DES TYPES DE DIGESTION

COMPARAISON ENTRE POPULATIONS ET ECOTYPES

Sur un total de 792 loci pour *MseI* obtenus avec 6 combinaisons d'amorces; 2985 loci pour *HpaII* et 3241 loci pour *MspI*, obtenus avec 12 combinaisons d'amorces, j'ai calculé les fréquences de la présence de bande pour chaque locus pour chacune des enzymes et effectué un test de rangs de façon à observer comment se répartissent les amplifications dans les populations de l'est et de l'ouest et entre les écotypes.

Quelle que soit l'enzyme, les populations provenant de Paracou (Table 3, Figure 2. en violet), montrent toujours une fréquence de bandes moyenne inférieure à celle des populations de Régina (en rouge), avec une fréquence moyenne de bandes par locus maximale de 0.27 pour *MseI* (Wilcoxon : $v=18577$, $p=0.8172$) (Table 3. A.). L'enzyme *MspI* montre la plus faible fréquence moyenne de bandes avec une moyenne de 0.14 (Wilcoxon : $v=90781$, $p=1$) pour la population de Paracou. Aussi, les profils d'amplification des individus de Paracou semblent avoir beaucoup de locus avec une fréquence proche de 0; plus que les profils d'amplification des populations d'arbres provenant de Régina.

Pour toutes les enzymes, l'écotype *S.spI*. (Figure 2 en bleu) a montré les plus fortes moyennes de fréquences de bandes (Table 3. B., Figure 2.). L'enzyme *MseI* a une fréquence de bandes pour *S.spI* (moyenne=0.26 ; Wilcoxon : $v=38310.5$, $p=1 \times 10^{-5}$) supérieure aux autres enzymes, pour lesquelles les valeurs de fréquences moyenne des bandes pour *S.spI*. sont de 0.18 (Wilcoxon : $v=141883.5$, $p=1.87 \times 10^{-12}$) et 0.16 (Wilcoxon : $v=169093$, $p=0.0007$), pour les enzymes *HpaII* et *MspI*, respectivement.

COMPARAISON DU DELTA DE FREQUENCE ENTRE LES ENZYMES

Sur les données de fréquences de bandes, il a été calculé la différence de fréquences entre écotypes et entre populations ; ce qui m'a permis de comparer la capacité des enzymes à différencier les écotypes et les populations de l'est et de l'ouest. Une enzyme plus discriminante aura une différence de fréquence de bandes significativement plus importante en valeur absolue.

Les enzymes sensibles à la méthylation ont systématiquement montré une différence des fréquences de bandes entre populations, supérieure en valeur absolue à celle de l'enzyme *MseI* (Table 4., A.), avec une différence moyenne de -0.02 (Wilcoxon : $v=107541.5$, $p=0.03662$) pour *MspI* et -0.018 (Wilcoxon : $v=87757.5$, $p=0.119$) pour *HpaII*, contre une différence de 0.01 pour *MseI*, en valeur absolue. Aussi, l'enzyme *HpaII* montre une plus forte sensibilité dans la différenciation des écotypes que *MseI* et *MspI*, avec une différence de fréquences moyenne de 0.03 (moyenne=0.04, Wilcoxon : $v=119459.5$, $p=0.408$, pour *MseI* et moyenne=0.01, Wilcoxon : $v=240654$, $p=0.9996$, pour *MspI*, Table 4. B.).

Sur la Figure 4., quand elle est comparée à *HpaII* et à *MspI*, *MseI* est en rouge ou en bleu, et montre une distribution moins large que les enzymes sensibles à la méthylation.

TEST SUR LES PHENOTYPES : ANALYSE EN COMPOSANTES PRINCIPALES

A l'aide du packages R « adegenet », j'ai pu décorréler la variance du profil d'amplification entre les individus pour chaque populations et écotypes en deux axes. Les enzymes sensibles à la méthylations ont des axes de première dimension qui explique moins bien la variance (40.27% pour *MspI* et 37.30% pour *HpaII*, Figure 4. C. et B.), avec des écarts compris entre -5 et 5. L'enzyme *MseI* a un axe de première dimension qui explique 75.80% de la variance, avec des écarts compris entre -7 et 8. En deuxième axe, les enzymes *MspI* (33.86%) et *HpaII* (27.44%) ont un pourcentage similaire de variance expliquée à celui du deuxième axe de *MseI* (31.69%).

Pour les enzymes *HpaII* et *MseI*, la population de *S.spI* provenant de Régina (en vert) montre la plus faible variance, contrairement au cas de l'enzyme *MspI* (Figure 4. C.) pour laquelle cette population d'écotype montre la plus forte variance sur les axes 1 et 2. La plus faible variance chez *MspI* est retrouvée par contre pour le *S.globulifera* de Paracou (en jaune), avec un écart de -1.5 à 2 qui au contraire montre la plus forte variance pour *HpaII*, avec un écart de -5 à 5, environ, et pour *MseI*, avec un écart de -2.5 à 6.5 environ.

DETECTION DES LOCUS SOUMIS A SELECTION

La plateforme Mcheza qui implémente la méthode DFDIST a permis de détecter 129 outliers entre les écotypes à Paracou et 243 outliers entre les écotypes à Régina avec un intervalle de confiance de 99.5% (Figure 6). Les locus de chaque enzyme ont une couleur différente, permettant de les différencier. Ainsi, je peux observer que les locus à variation de l'état de méthylation sont sélectionnés sur les deux sites. A Paracou, 5.78% des locus de *HpaII*, et 2.58% des locus de *MspI* sont sélectionnés avec un F_{ST} moyen pour les enzymes de 0.005 pour *HpaII* et 0.0007 pour *MspI*. A Régina, 8.43% des locus de *HpaII*, et 12.29% des locus de *MspI* sont sélectionnés avec un F_{ST} moyen pour les enzymes de 0.0000961 pour *HpaII* et 0.00 pour *MspI*.

Aussi, à Paracou, les points rouges qui représentent les locus à variations de séquence ADN sortent rarement du nuage neutre (5.25% des locus de *MseI*) et sont donc rarement outliers malgré un F_{ST} moyen de 0.012, comparé aux locus obtenus par la digestion avec les enzymes méthyl sensibles. Ce n'est pas le cas à Régina, où ils sortent très souvent du nuage de F_{ST} neutres (11.93% des locus de *MseI*, F_{ST} moyen= 0.097).

DISCUSSION

FREQUENCES DE BANDES DANS LES POPULATIONS ET LES ECOTYPES

Quel que soit le couple d'enzyme utilisé, l'écotype *S.sp1* a systématiquement montré une fréquence de la présence de bandes inférieure par rapport à l'écotype *S.globulifera*. De même, les individus localisés sur le site de Paracou ont systématiquement montré une fréquence de la présence de bandes inférieure par rapport aux individus présents sur le site de Régina. Ces individus ont pourtant été échantillonnés, traités dans les mêmes conditions et mélangés sur la même plaque PCR. Aussi, les digestions et amplifications PCR ont été faites par deux techniciens différents. Une hypothèse alternative à celle d'un biais technique, qui paraît à ce stade peu vraisemblable, serait une hypothèse biologique. Les résultats suggèrent que le génome du *S. S.sp1* serait plus méthylé (moins de fragments digérés), par contre, lorsque les fragments sont digérés (pas méthylés), ils le sont en plus forte fréquence que dans le *S. globulifera*. On observe bien, sur la Figure 3, que l'écotype *S.globulifera* (en vert) montre moins de locus avec une fréquence de bandes supérieure ou égale à 0.8 que l'écotype *S.sp1*. L'effet régional sur la différence de fréquence de bandes, pourrait s'expliquer par une plus faible proportion de locus avec des fréquences importantes pour les arbres provenant de Paracou (en violet) que pour ceux provenant de Régina (en rouge). Le *S.sp1* est en *terra firme* (sol drainés, souvent sec, par rapport aux bas-fonds). Le site de Paracou est à l'ouest de la Guyane, qui a un niveau de pluviométrie faible, dans le gradient pluviométrique entre l'est et l'ouest de la Guyane. Les faibles ressources en eau ayant déjà été associées avec de la dé-méthylation du génome[36] cela pourrait contribuer à expliquer le pattern de méthylation que l'on observe ici.

VARIATION EPIGENETIQUE ENTRE LES ECOTYPES ET LES POPULATIONS

Les enzymes MSAP *HpaII* et *MspI* ont montré qu'elles sont capables de mettre en avant la différence entre les écotypes et les populations de l'est et l'ouest de la Guyane. Aussi, un sous ensemble des locus obtenus par leur digestion de l'ADN sont potentiellement soumis à la sélection divergente en contribuant à la divergence entre les écotypes qui sont localement adaptés à leur environnement. Néanmoins, le F_{ST} sur l'ensemble des locus est très faible, suggérant que les écotypes *S.globulifera* et *S.sp1* sont peut-être en train de subir une spéciation écologique, avec le *S.sp1* en *terra firme* et le *S.globulifera* en bas-fond. Ici, l'échantillonnage d'arbres adultes a permis d'effectuer des profils de l'état de la variation de la méthylation chez les écotypes, dans leur population naturelle. Le fait que ces enzymes

méthyl sensibles génèrent des locus qui discriminent les écotypes, dont une minorité est potentiellement soumise à sélection et la grande majorité présentent de faibles valeurs de F_{ST} à travers le génome, indiqueraient que la méthylation ne joue qu'un faible rôle dans l'adaptation locale des écotypes de l'espèce *Symphonia globulifera*. D'autres études ont montré une capacité des enzymes méthyl sensibles à différencier des écotypes [22, 25, 32-34, 36, 50-53] et peu sont celles qui ont testé la possibilité que ces locus soient sélectionnés à travers le génome et pourrait permettre l'adaptation des écotypes [34, 37, 54].

Aussi, les enzymes *MspI* et *HpaII* ont une plus grande capacité dans la distinction des écotypes et des populations que *MseI*, même si les faibles valeurs de F_{ST} indiquent une variance de la présence de bandes plus importante au sein des populations d'écotypes plutôt que entre les populations d'écotypes. Ces enzymes sont sensibles à la méthylation mais sont également sensibles aux variations génétiques, puisque pour digérer ou non en fonction d'un statut de méthylation, le site de restriction doit être présent. C'est peut être cette double sensibilité qui leur permet de différencier plus souvent les populations que les enzymes uniquement sensibles à la présence ou absence d'un site de restriction. Enfin, dans l'adaptation locale du *Symphonia* de Guyane, le suivi du profil de méthylation des écotypes indiquerait que la méthylation n'apporte pas de contribution claire à la divergence, qui s'ajouterait à l'effet de la divergence des fréquences alléliques.

VARIATION GENETIQUE ENTRE LES ECOTYPES ET LES POPULATIONS

La différenciation observée aux locus obtenus avec l'enzyme *MseI* est globalement aussi forte que celle observée avec des enzymes méthyl-sensibles ; dans une analyse en composante principale les locus dérivés de *MseI* expliquent mieux les variations entre les écotypes et les populations que les enzymes méthyl sensibles. Aussi, les tests de sélection ont montré des F_{ST} s faibles mais supérieurs à ceux obtenus avec les enzymes méthyl sensibles. Ici, nous avons voulu observer la part que pouvait prendre la méthylation de l'ADN et la variation de séquence dans l'adaptation locale au sein de l'espèce *Symphonia globulifera*. Ainsi, les variations de séquences de l'ADN distinguent plus globalement les populations et les écotypes à l'intérieur des populations, suggérant que même en présence d'un faible F_{ST} , et de l'action de la méthylation, la variation de séquence garde un rôle important dans la différenciation des populations et des écotypes. En effet, il peut suffire de quelques gènes variant entre les écotypes pour permettre leur différenciation.[55, 56].

CONCLUSION

Dans cette étude, j'ai utilisé la technique de génotypage MSAP qui permet d'observer l'état de méthylation du génome, et cela dans le but de vérifier l'hypothèse que les écotypes pourraient avoir des profils de méthylation différents. J'ai également effectué un test de sélection de façon à observer un possible rôle des locus à variations de méthylation dans l'adaptation locale du *Symphonia globulifera* à la *terra firme*. Et de façon à comparer l'implication des locus à variation de séquences et les locus à variation de méthylation dans l'adaptation locale, le nuage neutre a été calculé avec l'ensemble des enzymes, et a montré qu'une minorité des locus obtenus avec les enzymes méthyl sensibles est sélectionnée à travers le génome et entre en jeu dans l'adaptation locale du *Symphonia* ; toutefois, leur niveau de divergence n'est pas supérieur, et peut être sensiblement inférieur, par rapport à celui observé par des marqueurs non méthyl-sensibles. La méthylation – telle qu'elle peut être appréhendée par une approche MSAP – ne contribue pas à la différenciation entre écotypes.

TABLES ET FIGURES

A.

Enzymes	Populations			
	Paracou	Régina	v	p-value
<i>Msel</i>	0.25	0.27	18577.5	0.8172
<i>Hpall</i>	0.16	0.18	69768	0.9998
<i>Mspl</i>	0.14	0.17	90781	1

B.

Enzymes	Ecotypes			
	<i>S.globulifera</i>	<i>S.sp1.</i>	v	p-value
<i>Msel</i>	0.25	0.26	38310.5	1.161e-05
<i>Hpall</i>	0.15	0.18	141883.5	1.876e-12
<i>Mspl</i>	0.14	0.16	169093	0.0006608

Table 3. Résumé du test de Wilcoxon effectué sur la fréquence de bandes, donné pour (A.) les populations et (B.) les écotypes .

Options utilisées: alternative="greater",paired=T; pour Paracou-Régina et *S.sp1.-S.globulifera*.

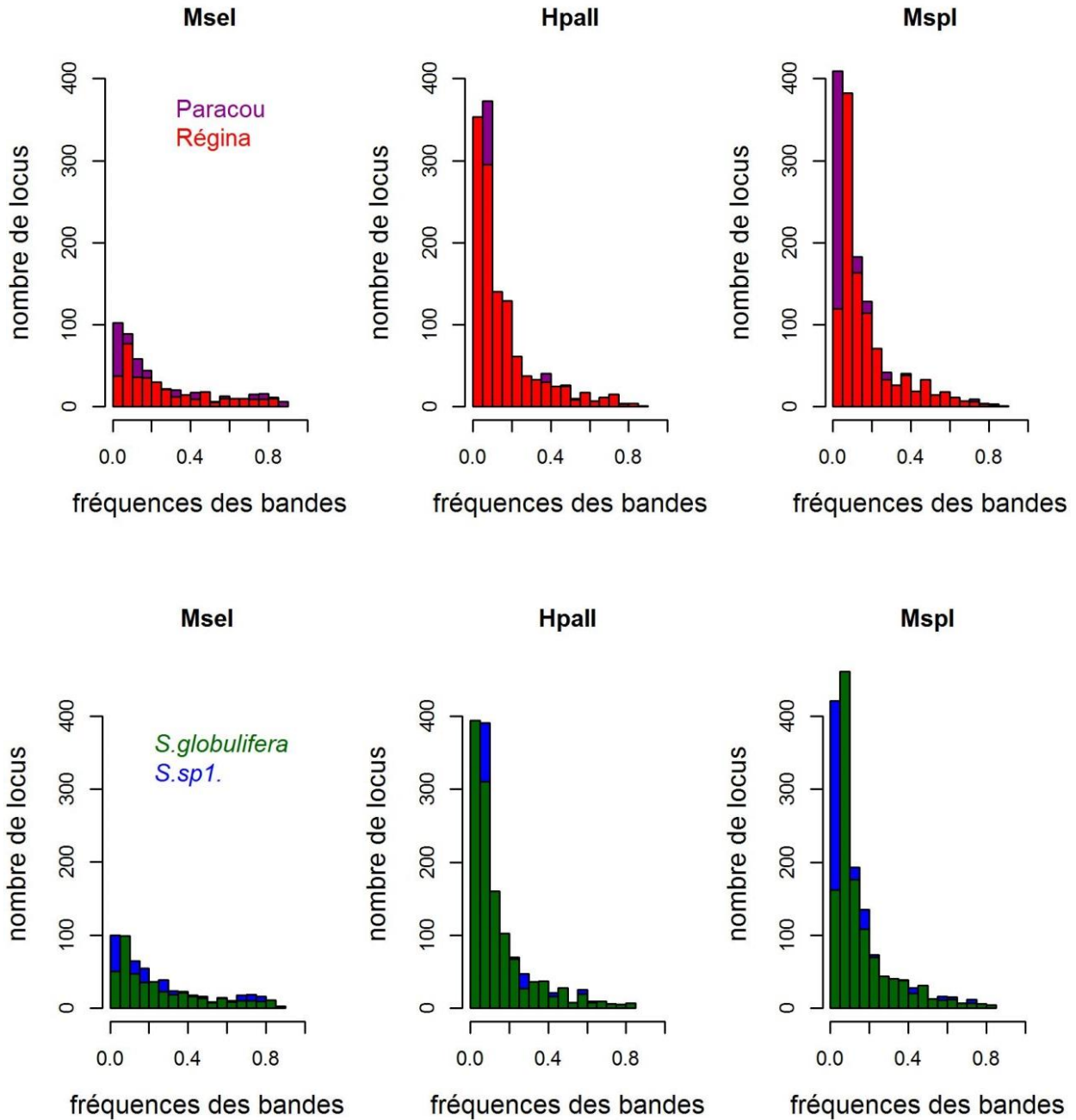


Figure 3. Fréquences de bandes; pour **A.**, les populations (Paracou et Régina) et pour **B.**, les écotypes (*S.sp1* et *S.globulifera*)

A.

Enzymes	Populations			p-value
	Moyennes		w	
<i>Msel</i> ^ᵃ - <i>Mspi</i> ^ᶲ	-0.01 ^ᵃ	-0.02 ^ᶲ	107541.5	0.03662
<i>Msel</i> ^ᵃ - <i>Hpal</i> ^ᶜ	-0.01 ^ᵃ	-0.018 ^ᶜ	87757.5	0.119
<i>Mspi</i> ^ᶲ - <i>Hpal</i> ^ᶜ	-0.02 ^ᶲ	-0.018 ^ᶜ	203394.5	0.7501

B.

Enzymes	Ecotypes			p-value
	Moyennes		w	
<i>Msel</i> ^ᵃ - <i>Mspi</i> ^ᶲ	0.04 ^ᵃ	0.01 ^ᶲ	154394.5	0.007132
<i>Msel</i> ^ᵃ - <i>Hpal</i> ^ᶜ	0.04 ^ᵃ	0.03 ^ᶜ	119459.5	0.408
<i>Mspi</i> ^ᶲ - <i>Hpal</i> ^ᶜ	0.01 ^ᶲ	0.03 ^ᶜ	240654	0.9996

Table 4. Résumé du test de Wilcoxon effectué sur la différence de fréquences de bandes entre enzymes, entre A., les populations (Paracou-Régina) et B., les écotypes (*S.spl.*-*S.globulifera*).

Options utilisées: alternative= "greater" .

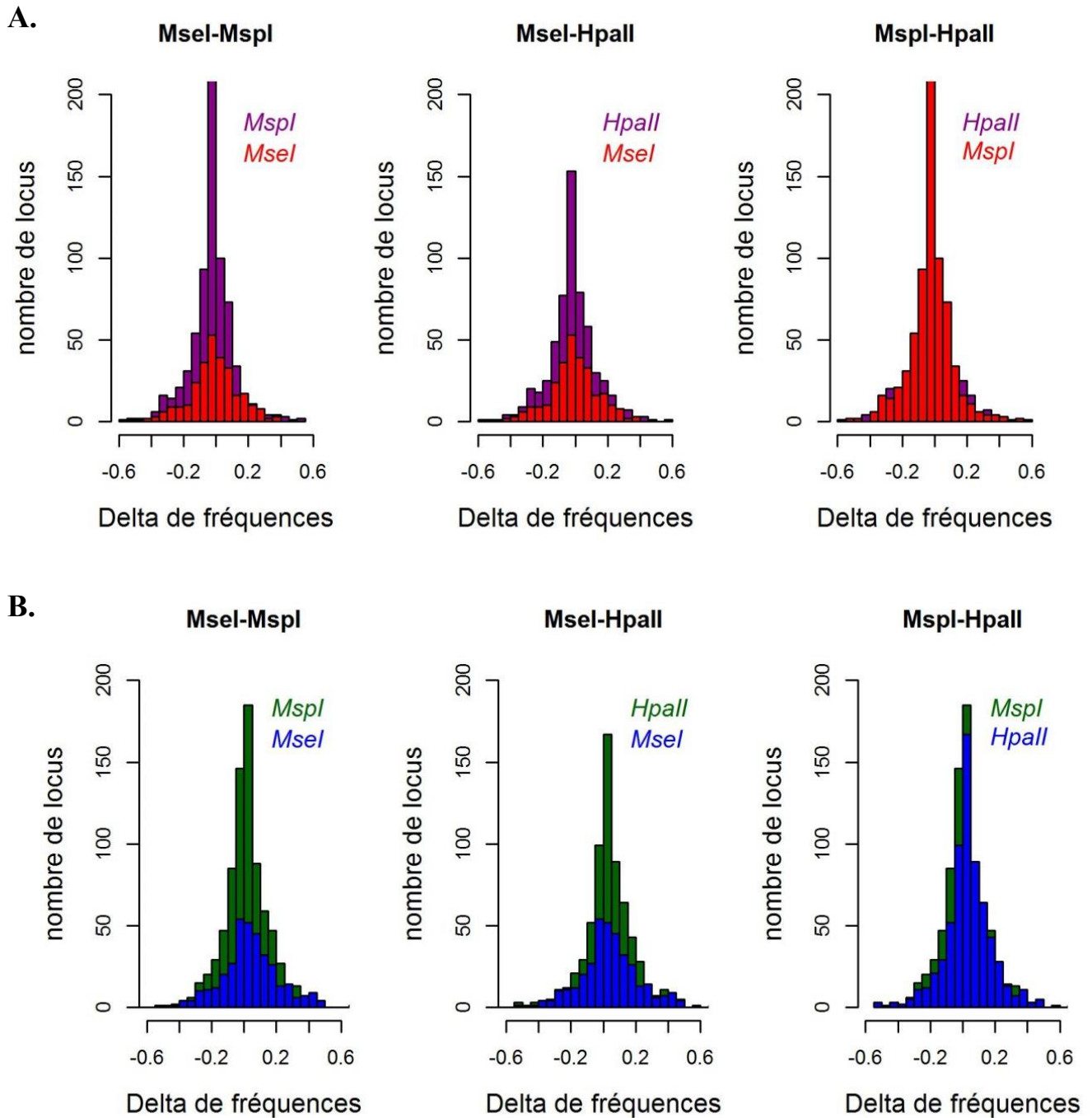


Figure 4. Différence de fréquence de bandes entre les enzymes;

A., entre les populations (Paracou et Régina) et **B.**, entre les écotypes (*S.spl* et *S.globulifera*) .

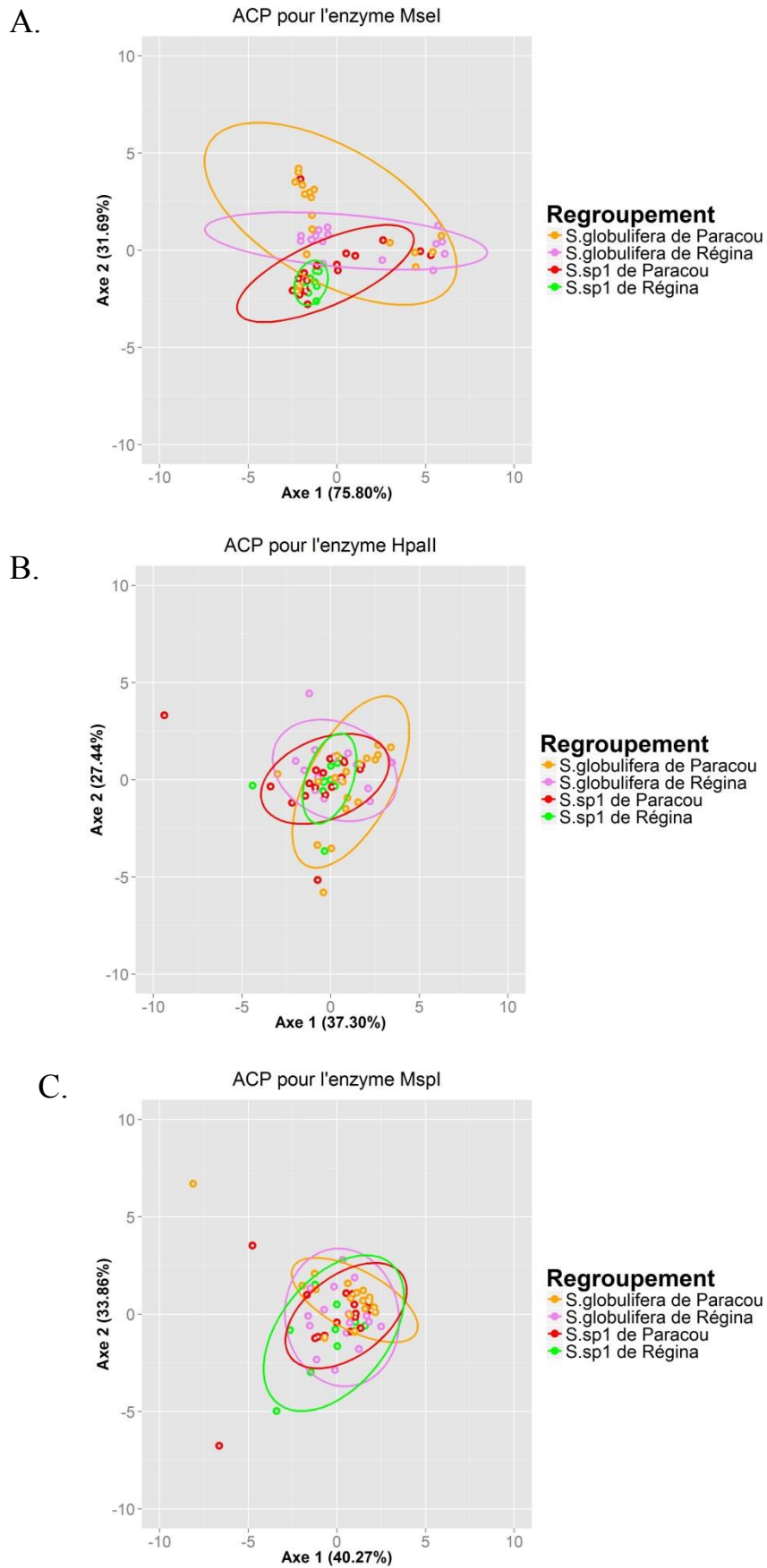


Figure 5. Analyse en composante principale pour les enzymes (A.) *MseI*, (B.) *HpaII* et (C.) *MspI*.

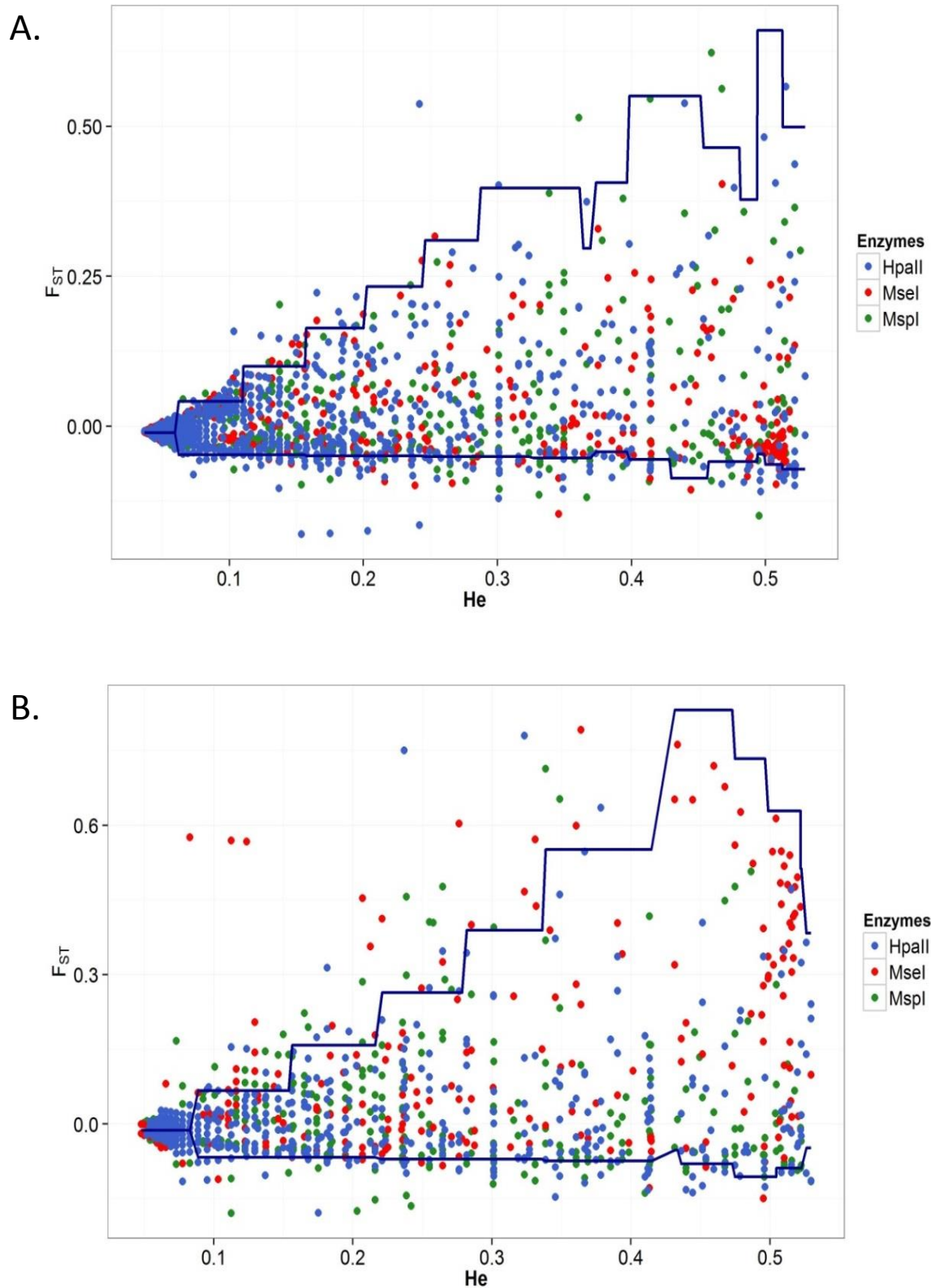


Figure 6. Recherche d'outliers. Pour chaque sites, **A**, à Paracou et **B**, à Régina, il a été effectué une recherche d'outliers, représentant les locus soumis à sélection; parmi les locus de *HpaII*, *MspI* et *MseI*. Le nuage neutre a été créé avec les locus des trois enzymes.

BIBLIOGRAPHIE

1. Feng, S., et al., *Conservation and divergence of methylation patterning in plants and animals*. Proceedings of the National Academy of Sciences, 2010. **107**(19): p. 8689-8694.
2. Flores, K.B., F. Wolschin, and G.V. Amdam, *The Role of Methylation of DNA in Environmental Adaptation*. Integrative and Comparative Biology, 2013. **53**(2): p. 359-372.
3. Vaughn, M.W., et al., *Epigenetic Natural Variation in *Arabidopsis thaliana**. PLoS Biol, 2007. **5**(7): p. e174.
4. Johannes, F., et al., *Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits*. PLoS Genet, 2009. **5**(6): p. e1000530.
5. Jablonka, E. and G. Raz, *Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution*. Q Rev Biol, 2009. **84**(2): p. 131-76.
6. Richards, C.L., O. Bossdorf, and M. Pigliucci, *What Role Does Heritable Epigenetic Variation Play in Phenotypic Evolution?* BioScience, 2010. **60**(3): p. 232-237.
7. De Jong, G., *Evolution of phenotypic plasticity: patterns of plasticity and the emergence of ecotypes*. New Phytologist, 2005. **166**(1): p. 101-118.
8. Kawecki, T.J. and D. Ebert, *Conceptual issues in local adaptation*. Ecology letters, 2004. **7**(12): p. 1225-1241.
9. Galindo, J., J.W. Grahame, and R.K. Butlin, *An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis**. Journal of Evolutionary biology, 2010. **23**(9): p. 2004-2016.
10. Street, N.R., et al., *The genetics and genomics of the drought response in *Populus**. The Plant Journal, 2006. **48**(3): p. 321-341.
11. Li, X.-M., et al., *Life history trait differentiation and local adaptation in invasive populations of *Ambrosia artemisiifolia* in China*. Oecologia, 2014: p. 1-9.
12. Dayan, D.I., D.L. Crawford, and M.F. Oleksiak, *Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of *Fundulus heteroclitus**. Molecular Ecology, 2015. **24**(13): p. 3345-3359.
13. Osunkoya, O.O., R. Boyne, and T. Scharaschkin, *Coordination and plasticity in leaf anatomical traits of invasive and native vine species*. American Journal of Botany, 2014. **101**(9): p. 1423-1436.
14. Migdałek, G., et al., *Morphological differences between violets growing at heavy metal polluted and non-polluted sites*. Flora - Morphology, Distribution, Functional Ecology of Plants, 2013. **208**(2): p. 87-96.
15. Nicotra, A.B., et al., *Plant phenotypic plasticity in a changing climate*. Trends in Plant Science, 2010. **15**(12): p. 684-692.
16. Herrera, C.M. and P. Bazaga, *Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees*. Botanical Journal of the Linnean Society, 2013. **171**(3): p. 441-452.
17. Bossdorf, O., C.L. Richards, and M. Pigliucci, *Epigenetics for ecologists*. Ecology letters, 2008. **11**(2): p. 106-115.
18. Pigliucci, M., C.J. Murren, and C.D. Schlichting, *Phenotypic plasticity and evolution by genetic assimilation*. Journal of Experimental Biology, 2006. **209**(12): p. 2362-2367.
19. Richards, C., K.F. Verhoeven, and O. Bossdorf, *Evolutionary Significance of Epigenetic Variation*, in *Plant Genome Diversity Volume 1*, J.F. Wendel, et al., Editors. 2012, Springer Vienna. p. 257-274.
20. Pfennig David, W. and R. Servedio Maria, *The role of transgenerational epigenetic inheritance in diversification and speciation*, in *Non-Genetic Inheritance* 2013. p. 17.
21. Van Oosten, M.J., et al., *The Role of the Epigenome in Gene Expression Control and the Epimark Changes in Response to the Environment*. Critical Reviews in Plant Sciences, 2014. **33**(1): p. 64-87.
22. Richards, C.L., A.W. Schrey, and M. Pigliucci, *Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation*. Ecology letters, 2012. **15**(9): p. 1016-1025.
23. Boyko, A. and I. Kovalchuk, *Epigenetic control of plant stress response*. Environmental and Molecular Mutagenesis, 2008. **49**(1): p. 61-72.
24. Lukens, L.N. and S. Zhan, *The plant genome's methylation status and response to stress: implications for plant improvement*. Current Opinion in Plant Biology, 2007. **10**(3): p. 317-322.

25. Reyna-López, G.E., J. Simpson, and J. Ruiz-Herrera, *Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms*. *Molecular and General Genetics MGG*, 1997. **253**(6): p. 703-710.
26. Walder, R.Y., et al., *Cloning of the MspI modification enzyme. The site of modification and its effects on cleavage by MspI and HpaII*. *Journal of Biological Chemistry*, 1983. **258**(2): p. 1235-41.
27. Zabeau, M. and P. Vos, *Selective restriction fragment amplification : a general method for DNA fingerprinting*, 1993, Google Patents.
28. Yaish, M., M. Peng, and S. Rothstein, *Global DNA Methylation Analysis Using Methyl-Sensitive Amplification Polymorphism (MSAP)*, in *Arabidopsis Protocols*, J.J. Sanchez-Serrano and J. Salinas, Editors. 2014, Humana Press. p. 285-298.
29. Schulz, B., R.L. Eckstein, and W. Durka, *Scoring and analysis of methylation-sensitive amplification polymorphisms for epigenetic population studies*. *Molecular Ecology Resources*, 2013. **13**(4): p. 642-653.
30. Fulnecek, J. and A. Kovarik, *How to interpret Methylation Sensitive Amplified Polymorphism (MSAP) profiles?* *BMC Genetics*, 2014. **15**(1): p. 2.
31. Schrey, A.W., et al., *Ecological Epigenetics: Beyond MS-AFLP*. *Integrative and Comparative Biology*, 2013. **53**(2): p. 340-350.
32. Lira-Medeiros, C.F., et al., *Epigenetic Variation in Mangrove Plants Occurring in Contrasting Natural Environment*. *PLOS One*, 2010. **5**(4): p. e10326.
33. Saez-Laguna, E., et al., *Epigenetic variability in the genetically uniform forest tree species Pinus pinea L.* *PLOS One*, 2014. **9**(8).
34. Schulz, B., R.L. Eckstein, and W. Durka, *Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb*. *Molecular Ecology*, 2014. **23**(14): p. 3523-3537.
35. Medrano, M., C.M. Herrera, and P. Bazaga, *Epigenetic variation predicts regional and local intraspecific functional diversity in a perennial herb*. *Molecular Ecology*, 2014. **23**(20): p. 4926-4938.
36. Tang, X.-M., et al., *Analysis of DNA methylation of perennial ryegrass under drought using the methylation-sensitive amplification polymorphism (MSAP) technique*. *Molecular Genetics and Genomics*, 2014. **289**(6): p. 1075-1084.
37. Platt, A., et al., *Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations*. *Molecular Ecology*, 2015. **24**(15): p. 3823-3830.
38. Doyle, J. and J. Doyle, *A rapid procedure for DNA purification from small quantities of fresh leaf tissue*. *Phytochem Bull*, 1987. **19**: p. 11 - 15.
39. Vos, P., et al., *AFLP: a new technique for DNA fingerprinting*. *Nucleic Acids Research*, 1995. **23**(21): p. 4407-4414.
40. Ley, A.C. and O.J. Hardy, *Improving AFLP analysis of large-scale patterns of genetic variation – a case study with the Central African lianas Haumania spp (Marantaceae) showing interspecific gene flow*. *Molecular Ecology*, 2013. **22**(7): p. 1984-1997.
41. Weir, B.S. and C.C. Cockerham, *Estimating F-Statistics for the Analysis of Population Structure*. *Evolution*, 1984. **38**(6): p. 1358-1370.
42. Hardy, O.J. and X. Vekemans, *SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels*. *Molecular Ecology Notes*, 2002. **2**(4): p. 618-620.
43. Jombart, T., *ade4: a R package for the multivariate analysis of genetic markers*. *Bioinformatics*, 2008. **24**(11): p. 1403-1405.
44. Beaumont, M.A. and R.A. Nichols, *Evaluating Loci for Use in the Genetic Analysis of Population Structure*. *Proceedings of the Royal Society of London B: Biological Sciences*, 1996. **263**(1377): p. 1619-1626.
45. Caballero, A., H. Quesada, and E. Rolán-Alvarez, *Impact of Amplified Fragment Length Polymorphism Size Homoplasmy on the Estimation of Population Genetic Diversity and the Detection of Selective Loci*. *Genetics*, 2008. **179**(1): p. 539-554.
46. Antao, T. and M.A. Beaumont, *Mchaza: A workbench to detect selection using dominant markers*. *Bioinformatics*, 2011.
47. Holsinger, K.E. and B.S. Weir, *Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}*. *Nat Rev Genet*, 2009. **10**(9): p. 639-650.
48. Weir, B.S. and C.C. Cockerham, *Estimating F-statistics for the analysis of population structure*. *Evolution*, 1984. **38**(6): p. 1358-1370.
49. Zhivotovskiy, L.A., *Estimating population structure in diploids with multilocus dominant DNA markers*. *Molecular Ecology*, 1999. **8**(6): p. 907-913.

50. Labra, M., et al., *Analysis of Cytosine Methylation Pattern in Response to Water Deficit in Pea Root Tips*. *Plant Biology*, 2002. **4**(6): p. 694-699.
51. Rathore, M., S. Mastan, and P. Agarwal, *Evaluation of DNA methylation using methylation-sensitive amplification polymorphism in plant tissues grown in vivo and in vitro*. *Plant Growth Regulation*, 2015. **75**(1): p. 11-19.
52. Fang, J.G. and C.T. Chao, *Methylation-Sensitive Amplification Polymorphism in Date Palms (*Phoenix dactylifera* L.) and their Off-Shoots*. *Plant Biology*, 2007. **9**(4): p. 526-533.
53. Bobadilla Landey, R., et al., *High Genetic and Epigenetic Stability in *Coffea arabica* Plants Derived from Embryogenic Suspensions and Secondary Embryogenesis as Revealed by AFLP, MSAP and the Phenotypic Variation Rate*. *PLOS One*, 2013. **8**(2): p. e56372.
54. Herrera, C.M. and P. Bazaga, *Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis**. *New Phytologist*, 2010. **187**(3): p. 867-876.
55. Cominelli, E., et al., *A Guard-Cell-Specific MYB Transcription Factor Regulates Stomatal Movements and Plant Drought Tolerance*. *Current Biology*, 2005. **15**: p. 1196-1200.
56. Chan, Y.F., et al., *Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a *Pitx1* Enhancer*. *Science*, 2010. **327**(5963): p. 302-305.

*PARTIE III: PROFIL D'EXPRESSION
DU SYMPHONIA GLOBULIFERA DANS
UNE EXPERIENCE
DE TRANSPLANTATION RECIPROQUE*

Introduction.....	105
Matériels et Méthodes.....	106
Echantillonnage	106
Extraction ARN et séquençage	108
Assemblage et alignements	109
Analyse de l'expression différentielle	111
Résultats	113
Assemblage et read-mapping	113
Analyse exploratoire	113
La relation entre les individus.....	114
Analyse des facteurs propres	114
Analyse des interactions topologiques.....	115
Analyse des interactions régionales	115
Analyse des interactions croisées topographiques et régionales.....	116
L'écotype et la provenance régionale	116
L'habitat et la provenance régionale	117
L'écotype et la plantation régionale	117
Discussion.....	119
Assemblage et Alignements	119
Synthèse des effets qui influencent l'expression des gènes	119
Le S.globulifera est plastique	120
Conclusion	121
Tables et figures	123
Bibliographie	138

INTRODUCTION

Les interactions Génotype x Environnement ont un rôle très important dans l'évolution et la diversification des espèces, elles sont également le facteur clé de l'adaptation locale[1]. Dans un environnement hétérogène, les populations peuvent exprimer des phénotypes différents en réponse aux pressions environnementales[2] d'origine biotique (taille de la population, prédation, symbiose..) et/ou abiotique (température, composition chimique du sol, quantité de lumière perçue..)[3]. La comparaison de la variation phénotypique de traits de fitness de plusieurs génotype dans des environnements changeants (plasticité phénotypique ou polyphénisme⁴(4)) permet de tester directement la présence ou non d'une adaptation locale [3].

Aussi, la plasticité phénotypique peut aboutir à la colonisation d'un nouvel environnement[4] ou la persistance d'une population dans un environnement qui change[5], et permettre une adaptation locale.

Si la plasticité phénotypique se caractérise classiquement par l'étude de la variation des traits, les avancées techniques en biologie moléculaire permettent aujourd'hui de caractériser les bases moléculaires de la plasticité phénotypique par l'étude de la variation de l'expression des gènes dans des environnements changeants. Ces approches moléculaires ont notamment permis de mieux comprendre le lien phénotype-environnement[6], en particulier le rôle de l'expression des gènes dans l'adaptation locale[7] et l'adaptation rapide à un nouvel environnement[8, 9]. Des études ont ainsi montré, sur différents modèles biologiques, qu'il peut exister un niveau d'expression différent des gènes entre deux écotypes [6, 10-12].

La variation de l'expression des gènes peut avoir deux origines : génétique et épigénétique. Les facteurs génétiques transmis d'une génération à l'autre, et impliqués dans la variation de l'expression des gènes, peuvent être l'action d'un facteur de transcription[10] ; ou ils peuvent être aussi d'ordre épigénétique [14], par le biais de la méthylation [15, 16] ou la modification des histones [17-20]. Enfin, ils peuvent être une combinaison des deux [21, 22].

Parmi les méthodes moléculaires qui permettent l'analyse de l'expression des gènes, le RNA-Seq est une technique de séquençage du transcriptome à haut-débit [23-25], qui s'est popularisée durant ces dernières années, en défaveur des puces ADN[26] et de la technique de qRT-PCR[27] (quantitative Reverse Transcription - Polymerase Chain Reaction). Le RNAseq

⁴ Cas particulier de la plasticité qui dans lequel un génotype fait correspondre plusieurs phénotypes discrets en fonctions de signaux environnementaux récurrents, tels que la saisonnalité.

permet l'observation du transcriptome à un moment t dans une condition donnée. Cette technique présente deux avantages : (i) le coût du séquençage par base de plus en plus faible associé à une augmentation de la quantité de données. Pour exemple, une puce ADN permet l'étude de l'expression de milliers de séquences[26] alors que le RNA-Seq permet l'étude d'une centaine de millions de séquences[28]. Le second avantage (ii) est la possibilité de méconnaître le génome au préalable en effectuant des assemblages *de novo*. Au contraire, l'utilisation des puces ADN ou de la qRT-PCR nécessitent de créer des ancrages pour l'hybridation de l'ADN complémentaire (puce ADN) ou la production d'amorces (qRT-PCR). Afin de pouvoir dissocier les principaux facteurs qui interviennent dans la variation de l'expression des gènes du *Symphonia globulifera*, nous avons utilisé le dispositif de transplantations réciproques décrit dans le chapitre 1 (Table 1.,2. Et Figure 1)

Grace à ce dispositif, j'ai pu montrer que la variation de certains traits d'adaptation est d'origine génétique et que pour certains traits de fitness (germination de graines), il existe une interaction génotype x environnement témoignant d'une adaptation locale des deux écotypes de *Symphonia*.

Dans ce dernier chapitre, j'ai pour objectif d'identifier les facteurs (génétiques, environnementaux, ou appartenant à l'interaction génétique x environnement) qui expliquent la variation de l'expression des gènes, afin de proposer des hypothèses sur le rôle de l'expression des gènes dans la divergence entre les écotypes (effet génétique) et dans l'adaptation locale (interaction génotype x environnement). La réplication des jardins de transplantation réciproques dans deux régions climatiquement et écologiquement contrastées de la Guyane (est et ouest) me permettra par ailleurs de tester l'effet de variations macro-environnementales sur l'expression des gènes.

MATERIELS ET METHODES

ECHANTILLONNAGE

Pour cette étude, j'ai utilisé des plantules de *S.globulifera* et de *S.sp1* provenant des jardins de transplantations réciproques. Pour rappel, il a été planté en 2009, 510 plantules des deux écotypes de *Symphonia globulifera*, répartis en bas-fond et en *terra firme*, à l'est et à l'ouest de la Guyane (Table 1 et Figure 1) ; ce dispositif est le même que celui décrit au chapitre 1 sur l'analyse de la divergence phénotypique entre écotypes. Sur 190 individus vivants en 2014, les feuilles de 2 à 3 plantules dans chaque conditions (Figure 2), à savoir en bas-fond et *terra*

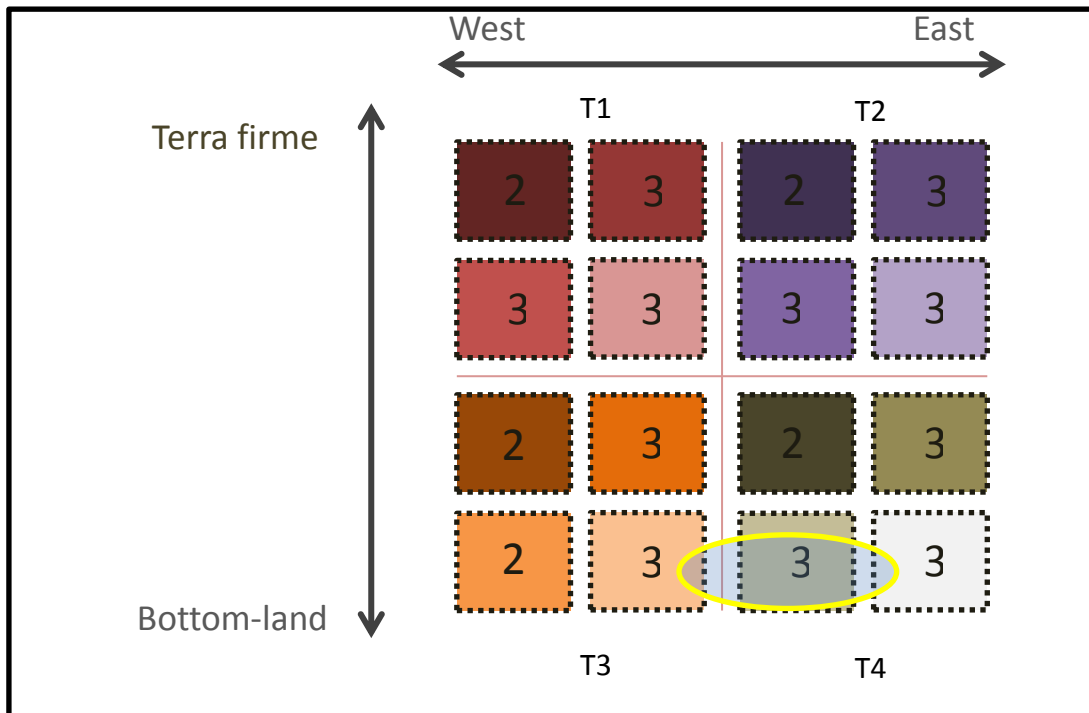


Figure2: Design expérimental. T1 étant le « common garden » en *terra firme* à Laussat (Ouest), T3 le « common garden » en bas-fond, T2 étant le « common garden » en *terra firme* à Régina (Est) et T4 le « common garden » en bas-fond, pour chaque common garden, la case du haut à gauche représente la provenance en *terra firme* de l’Ouest, celle de droite, la provenance en *terra firme* de l’Est. La case du bas à gauche représente la provenance bas-fonds de l’Ouest et celle de droite la provenance bas-fonds de l’Est. Le prélèvement de feuilles à pu être fait sur 41 plantules au lieu de 43. Deux plantules sélectionnées sont morts (cercle jaune).

	LANE 1	LANE 2	LANE 3	
Symphonia	W465_3	W476_10	W475_10	Répliques d'indexage
	W497_2	W424_17	W498_1	
	M837_23	W475_34	W425_25	
	W466_27	W497_36	W476_24	Répliques entre les <i>lanes</i> 1 et 3
	M837_14	W424_19	W425_44	
	W474_4	W475_11	W476_9	
	M837_39	W424_26	W425_30	Répliques entre les <i>lanes</i> 1 et 2
	W477_42	W474_27	W475_25	
	W463_27	W497_35	W466_21	
	M837_12	W424_36	W425_16	
	W426_20	W503_38	W475_10	
	W426_16	W503_17	W426_16	
	W497_3	W497_3	W498_1	
	W426_36	W426_36	W425_25	
	W426_7	W503_2	W426_7	

Figure3: Schéma de séquençage qui montre comment ont été répartis les échantillons entre les lanes. Toutes les conditions: *S.globulifera* et *S.sp1* provenant de l’Est et de l’Ouest et plantés à l’Est ou à l’Ouest en bas-fond et en *terra firme* sont mélangées par *lanes* sur chaque *lanes*. Les répliqués de *lanes* et de libraires sont mis en évidences.

firme, à l'est (Régina) et à l'ouest (Laussat) de la Guyane (Figure 2); et en minimisant les pertes du à un échantillonnage de feuilles trop important pour des plantules d'une hauteur moyenne de 28cm, j'avais prévu de prélever des feuilles sur 43 plantules. Entre la prospection et le jour de prélèvement, il a été constaté deux morts(tous les deux des plantules de l'écotype *S.globulifera* plantées en bas-fonds à l'Est de la Guyane, l'un provenait de l'Est et l'autre de l'Ouest (cercle jaune)). Ainsi, seulement 41 plantules ont été prélevées.

Concernant l'échantillonnage, nous nous sommes également souciés de la différence d'attaque herbivore pouvant être occasionnée sur les différentes parties (hautes et basses) des plantules. Des études ont montré que les feuilles jeunes sont plus soumises à l'herbivorie que les feuilles plus matures situées au bas de la plantule[29]. Afin de prendre en compte cette hétérogénéité, trois stades de maturité des feuilles ont été prélevés.

Sur le terrain, trois feuilles (une mature, au bas ; une modérément jeune, à mi-hauteur ; et une jeune à l'apex) ont été échantillonnées.

Les prélèvements ont été effectués le 14 et le 15 Mai 2014.

Les échantillons prélevés sur chaque plantule ont été mis dans un sac hermétique contenant de l'eau fraîche. Ces sacs ont été transportés au frais, dans une glacière, en présence de glace. Une fois arrivé au laboratoire, chaque feuille ($123=41*3$) a été rapidement nettoyée à l'eau, séchée manuellement entre deux feuilles de papier absorbant puis plongée quelques secondes dans de l'azote liquide avant d'être conservée au congélateur à -80°C .

EXTRACTION ARN ET SEQUENÇAGE

Les extractions ont été effectuées à l'aide du kit de NORGEN « Plant/Fungi Total RNA Purification Kit », à partir de 100mg de matériel frais. Les extractions ont été faites sur les 3 feuilles indépendamment et pour chaque individu. Les extractions ont été exécutées dans un ordre aléatoire sur chaque feuille de chaque individu de façon à éviter d'introduire des biais durant l'extraction ARN, qui pourraient se répercuter sur l'analyse de l'expression différentielle [30, 31].

Pour chaque extraction, les concentrations et qualités des ARNs ont été estimées sur Nanodrop. Globalement, plus d'un microgramme d'ARNs par feuille a été obtenu et présentait un peu de dégradation (RIN entre 3 et 7, LabChip GX (PerkinElmer)). Les échantillons ont ensuite été regroupés par individus de façon à obtenir 41 échantillons contenant la même quantité d'ARNs provenant des 3 stades de maturité. Les ARNs regroupés par individus ont été séchés au SpeedVac en présence de RNstable (Biomatrix, San Diego, Etats-Unis d'Amérique), qui permet le séchage et la conservation d'un ARN intègre durant 12

ans à température ambiante. Les échantillons secs ont ensuite été envoyés pour un séquençage total « Total RNAseq » en paired-end de 100pb à la plateforme de séquençage Illumina de la société IGA (Udine, Italie) qui s'est chargée de la préparation des banques en indexant individuellement chaque échantillon de façon à pouvoir les différencier les uns des autres. Lors des contrôles de qualité effectués par la société IGA, trois ont dû être retirés du lot à cause d'un RIN inférieur à 4, réduisant le nombre à séquencer à 38.

Les banques ont été séquencées sur 3 *lanes* de façon à obtenir une quantité de 20 à 30 millions de *reads* par individu. Afin de contrôler la présence d'un possible effet d'indexage, nous avons répliqué 3 individus sur la même *lane* en leur associant 2 index différents (3x2). D'autre part, afin de contrôler un possible effet *lane*, les librairies de 4 individus ont été scindées en deux et distribuées sur deux *lanes* différentes. Ainsi, l'étude est réalisée sur un total de 45 librairies indexées (38 individus + 7 répliques). Enfin, pour s'affranchir totalement d'un possible effet de *lane*, qui empêcherait la comparaison entre provenance, plantation et interaction entre les deux, nous avons choisi de distribuer chaque condition de provenances et de plantations sur toutes les *lanes*, à raison de 15 librairies par *lane* (Figure 3).

ASSEMBLAGE ET ALIGNEMENTS

Afin de créer une table de comptage, contenant les informations sur le nombre de *reads* par séquences, obtenues lors du séquençage RNAseq, il est tout d'abord nécessaire d'avoir une référence sur laquelle les *reads* de séquençage pourront être alignés. Cette référence a été obtenue par l'assemblage *de novo* des *reads* des 45 transcriptomes séquencés.

Après fusion des paires de *reads*, 45 librairies ont été assemblées avec l'algorithme Velvet en utilisant le script perl VelvetOptimiser.pl. La démarche a été itérative car l'assemblage est réalisé en prenant comme référence le fichier consensus de l'assemblage précédent (Figure 4). La première librairie a été sélectionnée aléatoirement, sans a priori. Son assemblage a permis de produire un fichier consensus « Premier consensus », contenant la séquence consensus de chaque contigs produits. Celui-ci a ensuite été incorporé dans l'assemblage d'une deuxième librairie grâce à l'option `-long`. Cette opération a été effectuée 45 fois, une fois par librairie. Le dernier assemblage a été effectué sur la librairie de départ, en incorporant les séquences consensus obtenues de l'assemblage de toutes les autres librairies assemblée par la suite. C'est ainsi, que j'ai pu obtenir un fichier .fa, nommé « Consensus final », contenant les séquences consensus d'un ensemble de 46 assemblages qui ont été effectués sur les 45 librairies.

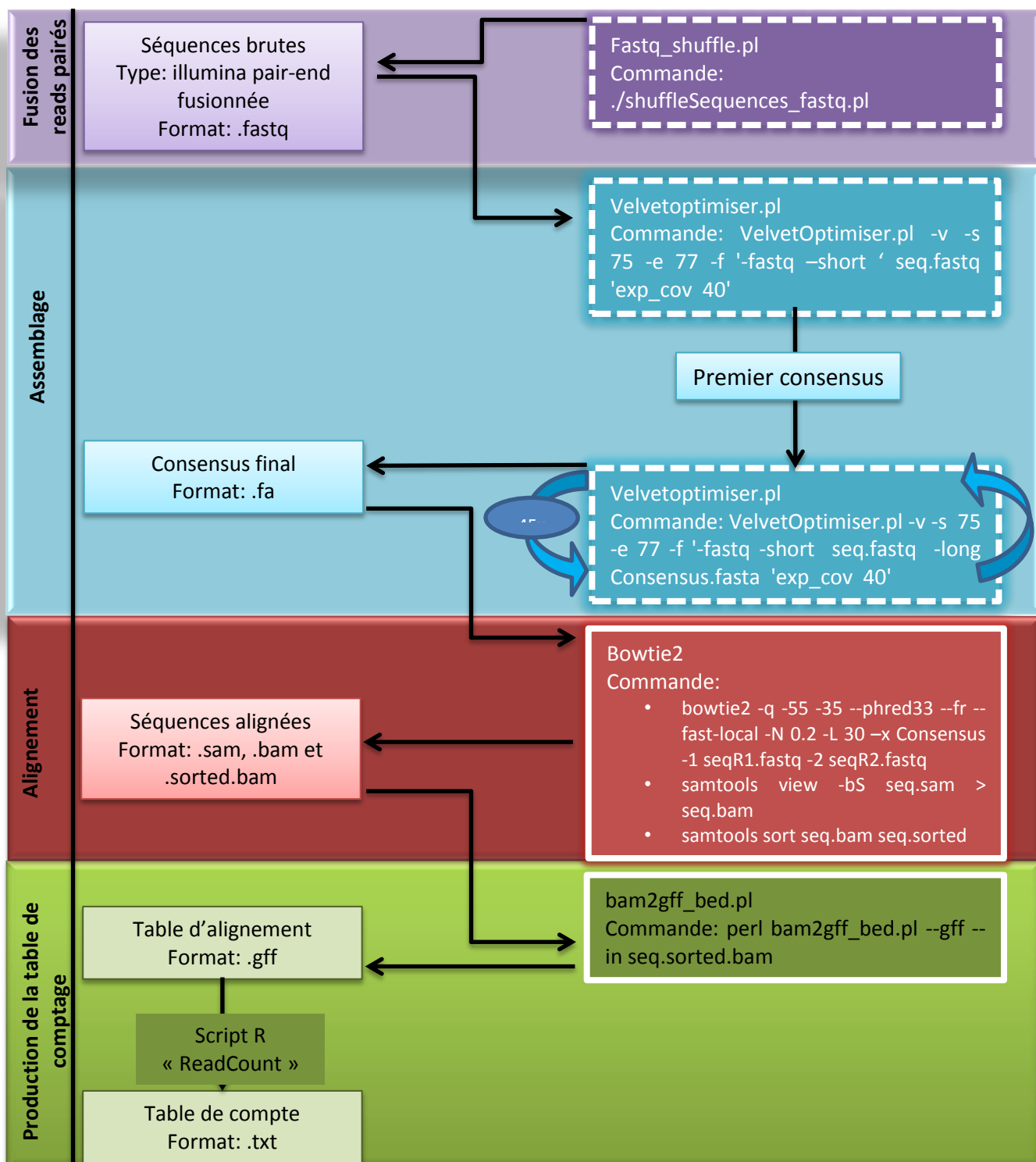


Figure 4: Etapes de production des séquences brutes à la table de comptage. Le séquençage en paired-end produit par la société IGA sur une plateforme Illumina, à permis d'obtenir 90 fichier .fastq, l'ensemble représentant les 2 sens de séquençage pour 45 librairies séquencées. La première étape permet la fusion des fichiers. Ainsi 45 fichiers, 1 par librairie, ont pu être assemblés de façon itérative. Cela a permis de produire un fichier .fa, contenant 18673 contigs. Par la suite, les 45 librairie ont été alignées avec ces contigs consensus, de façon à pour créer la table de comptage.

Par la suite, chaque librairie (45) a été alignée, à l'aide du logiciel Bowtie2[32] sur les séquences consensus uniques du fichier « Consensus final », permettant la production de fichiers .sam (Sequence Alignment/Map), .bam (version binaire du fichier .sam) et enfin .sorted.bam (version trié du .bam). Les 45 fichiers .sorted.bam ont alors été convertis à l'aide d'un script perl bam2gff_bed.pl du package Bio-ToolBox-1.15, en 45 fichiers .gff (General Feature Format), contenant les informations d'alignement, sous forme de table. A partir de ce point, un script R [33] *ad hoc* « ReadCount » (Annexe 1 p.133) a permis d'obtenir la table de comptage du nombre de *reads* par contigs pour chaque librairie. La première colonne des fichiers de tables .gff contient le nom des séquences de la référence (Consensus final, Figure 4)). En comptabilisant le nombre de lignes de la table par séquences de références, j'ai pu obtenir le nombre de *reads* par séquences. Ainsi, dans la table de comptage pour chaque individu, j'avais une colonne contenant à chaque ligne le nombre de *reads* par séquences ; une ligne représentant une séquence. Par la suite, une fonction « merge() » a été utilisée de façon à réunir dans un tableau, l'information du nombre de *reads* par séquences pour tous les individus.

ANALYSE DE L'EXPRESSION DIFFERENTIELLE

Une première visualisation des similarités et dissimilarités entre les individus a été effectuée par un positionnement multidimensionnel (MDS, pour MultiDimensional scaling). La fonction *plotMDS()* du package EdgeR permet le calcul d'une matrice de distance à partir du taux de changement d'expression des individus. Il est admis que dans le transcriptome, seul une minorité de gènes montrent un changement d'expression important [34], c'est-à-dire que la majorité des gènes ont un niveau d'expression neutre. Afin de mettre toutes les distances sur une même échelle, cette fonction effectue une transformation logarithmique de base 2 sur les distances calculées.

L'analyse de l'expression différentielle des séquences assemblées a été effectuée avec le package « edgeR »[35] du projet Bioconductor en utilisant des Modèles Linéaires Généralisés GLM avec deux facteurs au maximum en mettant en relation l'écotype et la provenance régionale avec l'habitat régional et topologique (Annexe 2). L'analyse a été effectuée avec un $FDR^5 < 0.01$. Ainsi, tous les résultats présentés mettent en évidence des séquences qui ressortent différemment exprimées, selon cette valeur de FDR.

⁵ Le False Discovery Rate représente la probabilité de rejeter H_0 alors que l'hypothèse H_0 est vraie (erreur de type I)

Dans la suite du manuscrit, l'acronyme GDE représentera les gènes différentiellement exprimés entre les conditions.

RESULTATS

ASSEMBLAGE ET READ-MAPPING

Le séquençage illumina des 45 librairies en paired-end a permis d'obtenir 25 millions de *reads* de 100 pb par librairie, en moyenne. Un total de 18 673 séquences uniques de 645pb en moyenne et communes entre les librairies ont été obtenues par assemblage *de novo*. L'alignement des librairies sur les séquences consensus des contigs m'a permis par la suite d'obtenir en moyenne 12.07 millions de *reads* alignés par librairie. Un maximum a été atteint à 19.82 millions pour la banque W497_3, ainsi qu'un minimum à 3.27 millions de *reads* pour la banque W426_7 correspondant à deux *S.globulifera* planté en bas-fond (Figure 5).

ANALYSE EXPLORATOIRE

Le graphique MDS en Figure 6 montre la relation entre les individus sur la base du ratio du nombre de *reads* obtenus après alignement. Les *lanes* sont représentées par des couleurs différentes. Une première observation est la présence de deux échantillons, qui s'écartent des autres points et qui expliquent la majorité de la variance du premier axe de l'analyse multidimensionnelle. Ces deux points correspondent à deux répliques de l'individu W426_7, une plantule de l'écotype *S.globulifera* plantée en bas-fond. Le premier réplique séquencé sur la *lane* 1 a aligné 6.67 millions de *reads* ; le deuxième, séquencé sur la *lane* 3 a aligné 3.27 millions de *reads*.

Une deuxième observation est que les individus ne sont pas regroupés par *lane* suggérant l'absence d'effet *lane*. Pourtant, l'effet *lane* sur le nombre de *reads* par librairie est significatif (ANOVA : $F=15.2$, $P<0.001$). On peut voir sur la Figure 7 que les librairies séquencées sur les trois *lanes* n'ont pas le même nombre moyen de *reads* par librairies (L1 : 11.4 millions, L2 : 14.6 millions et L3 : 10.2 millions). Enfin, la fonction *decideTestsDGE*, appliquée à la suite d'un *exactTest* qui calcule le taux de changement par paire de *lanes*, montre une quantité de 588 séquences sous exprimées entre la *lane* 1 et la *lane* 2 (Figure 8, Colonne de gauche). Lorsque l'individu W426_7 est éliminé du jeu de données, il n'y a plus que deux séquences sous exprimées. D'autre part, le graphique MDS (Figure 9) montre une plus grande homogénéité des écarts entre les échantillons et confirme l'absence d'effet de *lane* regroupant les échantillons. Sur ce graphique, on voit une bonne superposition des librairies qui sont représentées en rouge pour les répliques des *lanes* et en bleu pour les répliques d'indexage. Pour la suite des analyses, l'individu W426_7 est exclu.

LA RELATION ENTRE LES INDIVIDUS

Les figures 10 et 11 correspondent au même graphique MDS qu'en figure 9. Les couleurs indiquent respectivement : l'écotype dans sa topologie de plantation (bas-fond et *terra firme*) et la provenance régionale dans son site régional de plantation.

En figure 10, il y a globalement deux groupes de similarité. Les points rouges et oranges, représentent les plantules de l'écotype *S.globulifera* plantées respectivement en bas-fond et en *terra firme*, et sont regroupés sur la droite du graphique. Les points bleu ciel et bleu foncé représentent les plantules de l'écotype *S.sp1* plantées en bas-fond et en *terra firme* respectivement. Ils sont regroupés sur la partie droite du graphique.

On peut néanmoins constater que deux semis de l'arbre mère W424 (*S.sp1* provenant de l'est) s'éloignent de leur groupe, qui par ailleurs comprend deux autres individus de ce même arbre mère (réunis par un triangle bleu transparent)

Aucun regroupement par provenance ou site de plantation ne se distingue dans la figure 11.

ANALYSE DES FACTEURS PROPRES

Les figures 12 et 13 nous renseignent sur le nombre de gènes différentiellement exprimés et montrent la façon dont ils se répartissent entre les facteurs propres : l'écotype (Ecot), l'habitat (Hab), la provenance régionale (Prov) et le site de plantation régional (Plant). Parmi les 18673 séquences uniques, 553 gènes (3%) présentent un niveau d'expression différent entre le *S.globulifera* et le *S.sp1*, avec 340 séquences qui sont sous exprimées et 213 séquences qui sont surexprimées (figure 12). Parmi 18673 contigs, sur l'ensemble des quatre facteurs principaux, j'ai pu détecter 718 GDEs, dont 540 différencient uniquement les écotypes. **Ainsi, 75.21% des gènes différentiellement exprimés sur l'ensemble des facteurs simples sont ceux qui différencient les écotypes.** Aussi, 5 gènes différenciant les écotypes sont en commun avec la provenance régionale qui comptabilise 92 séquences dont l'expression est différente entre l'est et l'ouest.

Concernant l'habitat, 14 gènes montrent un écart d'expression. Deux séquences sont sous exprimées et 12 séquences sont surexprimées en *terra firme* par rapport aux bas-fonds. Sur cet ensemble de gènes qui se différencient entre les habitats, 8 (57.14% de 14) sont partagés avec l'écotype et 5 (35.71% de 14) sont partagés avec le site de plantation régional. Une seule séquence représentant 7.14% de l'ensemble des séquences différentiellement exprimées entre bas-fond et *terra firme* ne montre un différentiel d'expression que pour le facteur habitat.

Entre la provenance régionale de l'est et la provenance régionale de l'ouest, 92 gènes sont exprimés différemment, contre 91 pour la plantation régionale, parmi lesquels 9 séquences sont en communs.

Aucune séquence différentiellement exprimée n'est partagée entre l'écotype et la plantation ou l'habitat et la provenance.

ANALYSE DES INTERACTIONS TOPOLOGIQUES

Le modèle linéaire a été utilisé pour tester l'effet du facteur écotype en fonction des habitats. Au total, 1027 séquences sont différentiellement exprimées entre le *S.globulifera* et le *S.sp1* en bas-fonds (EcotBF), le *S.globulifera* et le *S.sp1* en *terra firme* (EcotTF), les bas-fonds et la *terra firme* pour le *S.globulifera* (GloHab) et entre les bas-fonds et la *terra firme* pour le *S.sp1* (Sp1Hab) (Figure 14 et 15). Les gènes différenciant les écotypes, que ce soit en bas-fonds ou en *terra firme*, sont au nombre de 1024, ce qui représente 98.44% du nombre total de gènes différenciés dans les interactions topologiques. En bas-fond, on dénombre 553 séquences exprimées différemment entre le *S.globulifera* et le *S.sp1*. 213 séquences sont surexprimées et 340 sont sous exprimées. En *terra firme*, 703 séquences sont différentiellement exprimées entre les écotypes, et parmi lesquelles, 233 sont surexprimées et 470 sont sous exprimées. Sur ces 1024 séquences, 230 (22%) sont différentiellement exprimés entre le *S.globulifera* et le *S.sp1* dans les deux types d'habitats, *terra firme* et bas-fond.

S.sp1 (Sp1Hab) et *S.globulifera* (GloHab) ne partagent aucune séquence exprimée différemment dans le changement d'habitat bas-fonds à *terra firme*. L'écotype *S.sp1* n'a que deux séquences surexprimées que l'on ne retrouve dans aucune autre condition. Par contre dans la transition bas-fonds à *terra firme*, le *S.globulifera* a 14 séquences dont le niveau d'expression diffère également entre les écotypes en bas-fonds (EcotBF : 6), puis en *terra firme* (EcotTF : 5). Aussi, deux séquences sont pareillement en commun avec les écotypes en bas-fonds et en *terra firme* (EcotBF et EcotTF).

ANALYSE DES INTERACTIONS REGIONALES

Les figures 16 et 17 montrent comment se répartissent les gènes différentiellement exprimés entre les sites de plantation, quand les plantules proviennent de l'est (PlantPrEst), puis de l'ouest (PlantPrOuest) et entre les provenances, pour les plantules disposées à l'est (ProvPIEst) et à l'ouest (PlantPIOuest).

208 séquences (1.1%) sont différentiellement exprimées dans les emboîtements des sites de plantations avec la provenance régionale des plantules (Figure 16 et Figure 17). 76 séquences différencient les plantules à l'est et à l'ouest de la Guyane quand elles proviennent de l'est ; et 72 séquences différencient les plantules provenant de l'est et de l'ouest de la Guyane, quand elles sont plantées à l'est. Ces gènes sont globalement sous exprimés. 21 gènes sont surexprimés contre 70 sous exprimés pour la transition de plantation est à ouest quand les plantules proviennent de l'ouest. Par contre, 30 gènes sont surexprimés contre 62 sous exprimés pour la transition entre les plantules provenant de l'est à ouest et plantées dans le site de plantation de l'est.

Pour les plantules à l'ouest, il n'y a que 29 gènes différentiellement exprimés contre 9 pour les plantules provenant de l'ouest.

S'il n'y a aucun gène différentiellement exprimé entre les provenances est (PlantPrEst) et ouest(PlantPrOuest), deux gènes sont en commun entre les sites de plantation est (ProvPIE) et ouest (ProvPIO). Les sites de provenances et de plantations de l'est ont en commun 14 gènes différentiellement exprimés contre 1 seul gène entre les provenances et site de plantations de l'ouest.

ANALYSE DES INTERACTIONS CROISEES TOPOGRAPHIQUES ET REGIONALES

L'ECOTYPE ET LA PROVENANCE REGIONALE

Les interactions écotipe x provenance régionale sont présentées dans les figures 18 et 19. Elles renseignent le nombre de séquences différentiellement exprimées entre les écotypes quand ils proviennent de l'est (EcotPrE) ou de l'ouest (EcotPIO), mais aussi entre les provenances régionales du *S.globulifera* (GloPROV) et du *S.sp1* (Sp1PROV). Parmi toutes ces conditions, 2511 séquences sont différentiellement exprimées (13.5%).

Le nombre de gènes dont l'expression est différente entre les écotypes provenant de l'est et de l'ouest est de 2358 séquences, ce qui représente un pourcentage de 93.9% des GDEs détectés dans l'interaction écotipe x provenance régionale.

1599 séquences sont différentiellement exprimées entre les deux écotypes provenant de l'est (836 séquences surexprimées et 763 sous exprimées) alors que 1007 séquences sont différentiellement exprimées pour la provenance de l'ouest (185 séquences surexprimées et 822 sous exprimées). Parmi ces séquences, 239 sont communes aux deux provenances.

Parmi les gènes exprimés différemment entre les écotypes provenant de l'est, 150 (9.38% de 1599) sont partagés avec le *S.globulifera* (GloPROV), qui a également en commun 60 gènes

qui sont différentiellement exprimés entre les écotypes provenant de l'ouest. Quand on compare le nombre de GDEs observés pour les écotypes, le nombre de séquences différentiellement exprimées entre les régions de provenances est beaucoup plus faible pour l'écotype *S.sp1* (337 séquences pour *S. globulifera* et 80 séquences pour *S. sp1*)

Le *S.globulifera* (GloPROV) et le *S.sp1* (Sp1PROV) ont respectivement 119 séquences et 31 séquences non partagées et exprimées différemment entre les provenances régionales.

L'HABITAT ET LA PROVENANCE REGIONALE

L'interaction croisée entre habitats et provenances régionales a comptabilisée seulement 3 gènes différentiellement exprimés (Figure 20). Les provenances régionales surexpriment la même séquence, quand elles sont plantées en bas-fond et en *terra firme*.

Par contre, dans la différence d'expression entre les habitats, les séquences ne sont pas les mêmes ; une séquence est sous exprimée quand les plantules proviennent de l'Est et une autre séquence est surexprimées quand elles proviennent de l'Ouest.

L'ECOTYPE ET LA PLANTATION REGIONALE

Les figures 21 et 22 montrent le nombre de gènes différentiellement exprimés dans l'interaction entre l'écotype et de la plantation régionale ; et comment ils sont répartis entre les *S.globulifera* et les *S.sp1* quand ils sont plantés à l'est (EcotPIE), entre les *S.globulifera* et les *S.sp1* quand ils sont plantés à l'ouest (EcotPIO), entre les sites de plantations de l'est à l'ouest pour le *S.globulifera* (GloPLANT) et entre les sites de plantations de l'est à l'ouest pour le *S.sp1* (Sp1PLANT).

Dans l'interaction entre l'écotype et la plantation régionale, 1200 séquences sont différentiellement exprimées entre les écotypes quand ils sont plantés à l'est, contre 629 quand ils sont plantés à l'ouest. Quand les écotypes sont plantés à l'est (EcotPIE), on dénombre 636 séquences surexprimées contre 564 séquences sous exprimées. Quand ils sont plantés à l'ouest, il n'y a plus que 142 gène surexprimés et 487 sous exprimés (Figure 21). Parmi ces séquences, 265 sont en commun entre les écotypes qu'ils soient plantés à l'est ou à l'ouest (Figure 22).

L'effet de la plantation à l'intérieur d'un écotype est moins important avec seulement 127 séquences différentiellement exprimées entre les *S. globulifera* plantés à l'est et à l'ouest, et seulement 42 séquences différentiellement exprimées entre les *S. sp1* plantés à l'est et l'ouest. Parmi les séquences exprimées différentiellement pour les écotypes *S.sp1* (Sp1PLANT) et *S.globulifera* (GloPLANT) dans les sites de plantation régionaux, 127 séquences différencient

le *S.globulifera* planté à l'Est et à l'Ouest parmi lesquelles 62 sont en commun avec les gènes différentiellement exprimés entre les écotypes plantés à l'est (EcotPIE) et 11 pour ceux qui sont à l'ouest (EcotPIO). Dans le cas du *S.sp1*, il n'y a qu'une séquence en commun avec celles qui différencient les écotypes à l'Est de la Guyane et 3 pour le site de plantation Ouest.

DISCUSSION

ASSEMBLAGE ET ALIGNEMENTS

La technique de séquençage illumina a permis d'obtenir un grand nombre de *reads* (20 à 25 millions de *reads* pour chaque librairie). En l'associant à une démarche d'assemblage itérative, j'ai pu obtenir 18673 séquences de 645pb en moyenne, dans l'assemblage *de novo* du *S. globulifera*. Cette méthode itérative a permis la création de séquences « consensus » basées à la fois sur les séquences obtenues à partir de l'ARN provenant de l'écotype *S. globulifera* et de l'ARN provenant de l'écotype *S.sp1*, permettant d'éviter la possibilité d'un biais différentiel dans l'alignement des séquences des deux écotypes avec les séquences consensus.

J'ai choisi d'utiliser la technologie illumina pour cette étude d'expression différentielle, ce qui me permet d'obtenir un nombre important de *reads*, avec plus de contigs et de plus grande taille que le précédent séquençage du transcriptome effectué avec du pyroséquençage 454[36]. La technologie de séquençage à haut-débit est très sensible durant l'étape de la préparation des banques et durant toute les étapes qui interviennent avant le séquençage [30, 37]. En réalisant une analyse multidimensionnelle présentant les relations entre les individus j'ai mis en évidence la présence d'un échantillon (W426_7) qui présentait un nombre de *reads* très faible par rapport au reste du jeu de données. L'homogénéité entre les autres librairies (44 librairies) montre que le protocole spécifiquement mis au point pour la préparation des banques ADNc de *Symphonia* est reproductible et efficace. Enfin, il est important d'insister sur l'importance du plan d'expérience (ajout de répliquas de *lanes* et d'indexage) qui a permis de tester et contrôler les effets environnementaux liés à la technique de la méthode.

SYNTHESE DES EFFETS QUI INFLUENCENT L'EXPRESSION DES GENES

Cette étude a permis de montrer un fort effet de l'écotype sur l'expression différentielle des gènes (qui implique plus de 75 % des séquences différentiellement exprimées) alors que l'effet génétique dû à la provenance régionale ainsi que les effets environnements (habitats et plantation régionale) semblent marginaux (figure12). Cela signifie que la divergence phénotypique entre les écotypes, d'une part, et l'adaptation locale, d'autre part, présents dans le complexe *Symphonia globulifera* peuvent être expliqués par une différence du niveau d'expression global entre les écotypes. L'effet de l'écotype est maintenu à l'intérieur de chaque provenance régionale (figure 18) avec une plus forte différence entre les deux écotypes provenant de l'est de la Guyane (1599 vs 1007 séquences, figure 18). La présence

d'une différenciation génétique significative entre l'est et l'ouest de la Guyane ($F_{ST} = 0.26$ au niveau du génome chloroplastique, thèse Maxime CASALIS) pourrait refléter une histoire évolutive différente des *Symphonia* dans ces deux régions.

Le dispositif expérimental a aussi permis d'estimer les effets emboîtés « génétique dans environnement » et « environnement dans génétique » révélant que quel que soit l'habitat de plantation (bas-fond ou terre ferme) les différences entre les deux écotypes restent fortes (553 et 703 séquences différentiellement exprimées en bas- fond et *terra firme* respectivement, figure 14). Au contraire, à l'intérieur d'un écotype, que ce soit *S. globulifera* ou *S. sp1*, l'habitat de plantation a peu d'effet (14 et 2 séquences différentiellement exprimées, Figure 14) suggérant qu'un faible nombre de gènes différentiellement exprimés est nécessaire à la plasticité phénotypique. L'expression différentielle des gènes est bien connue dans l'étude du développement de l'organisme, dans les situations de stress ou encore dans l'établissement de la plasticité [6, 38, 39]. Quelques études ont également permis de mettre en évidence une expression des gènes différente entre des écotypes[6, 10, 12, 40] et notamment dans le cas de domestication[11, 41, 42].Ceci peut être dû à plusieurs facteurs, d'ordre génétique[8, 43-45] ou épigénétique[46], mais ces résultats ajoutent de la crédibilité à l'hypothèse que l'expression différentielle des gènes pourrait contribuer à l'adaptation locale[47, 48] par le renforcement de trait adaptatifs agissant sur la fitness des individus[6, 14, 15, 44, 49].

LE *S.GLOBULIFERA* EST PLASTIQUE

Lorsque l'on regarde de près les différences de niveau d'expression entre le *S. globulifera* et le *S. sp1* en interaction avec l'environnement, le *S. globulifera* présente plus de séquences différenciées que le *S. sp1* (14 vs 2 entre les habitats, 127 vs 42 entre les plantations régionales). Cet ajustement de l'expression des gènes en fonction de l'environnement de plantation suggérerait une plasticité du *S. globulifera* qui viendrait conforter les hypothèses faites dans le chapitre 1, ou j'ai pu mettre en avant la plasticité du *S.globulifera* qui montre une hauteur similaire à celle du *S.sp1* quand il est planté en *terra firme*, et quand il est planté à l'Ouest de la Guyane.

La plasticité a déjà été associée à une expression différentielle des gènes affectant le phénotype [4, 6, 50].

L'EST, UN SITE PLUS STABLE ?

Le site de plantation à l'est de la Guyane a montré deux fois plus de séquences différentiellement exprimées que celui de l'ouest (1200 contre 629), suggérant un environnement plus hétérogène à l'est qu'à l'ouest de la Guyane. Sous l'hypothèse d'un mécanisme adaptatif, l'expression différentielle est une réponse aux variations de l'environnement. La région Régina-Kaw, correspondant au site de plantation Est, a probablement subi de plus faibles modifications climatiques au cours du Pléistocène et est considérée comme une zone refuge pour les écosystèmes forestiers. Le plus fort signal de différenciation d'expression pourrait alors être la marque d'une plus fine adéquation entre écotypes et environnement dans ces conditions écologiquement plus stables.

CONCLUSION

Cette étude visait à tester l'hypothèse d'un rôle de l'expression des gènes dans l'adaptation locale de l'espèce *Symphonia globulifera*, avec l'écotype *S.globulifera* spécialiste des bas-fonds et l'écotype *S.sp1* spécialiste de la *terra firme*. Afin de contrôler les interactions Génotype x Environnement, nous avons séquencé le transcriptome de plantules de 5 ans provenant d'une expérimentation de jardins de transplantation réciproque, en gardant l'information de l'individu dans un séquençage RNAseq assemblé *de novo*. Mes résultats montrent tout d'abord un effet très important de l'écotype sur l'expression différentielle des gènes consensus, qui a permis la détection de 553 gènes différentiellement exprimés entre les écotypes. Aussi, l'étude des interactions Génotype x Environnement montre une expression différentielle du *S.globulifera* quand il est placé en *terra firme* par rapport au bas-fond, mais aussi quand il est planté à l'ouest par rapport à l'est ; des résultats qui correspondent à la plasticité du *S.globulifera* précédemment montrée, en *terra firme* et quand il est planté à l'Ouest dans les jardins de transplantation réciproque. Mon étude est d'actualité[6] et montre l'intérêt qui doit être porté sur l'expression des gènes dans l'adaptation locale, la spéciation écologique et la diversification des espèces. Enfin, la nouveauté qui est apporté ici réside dans le fait d'étudier le rôle de l'expression des gènes dans un processus d'adaptation locale avec pour modèle biologique une espèce d'arbre tropical, en milieu naturel et contrôlé.

TABLES ET FIGURES

Arbre Mère	Site	Latitude	Longitude	Provenance régionale	Provenance topographique	Morphotype	Nombre de graines récoltées
W426	Kaw	4°33'22.34"N	52°14'20.00"W				36
W503	Régina	4°16'00.84"N	52°09'51.84"W		Bas-fond	<i>S. globulifera</i>	37
M837	Montagne Trésor	4°33'38.21"N	52°13'59.06"W	Est			36
W424	Kaw	4°33'20.93"N	52°14'21.12"W		Plateau	<i>S. sp1</i>	38
W425	Kaw	4°33'22.07"N	52°14'22.76"W				39
W463	St Laurent	5°23'28.70"N	53°39'07.83"W				36
W466	Crique Naï	5°23'47.03"N	53°42'25.95"W		Bas-fond	<i>S. globulifera</i>	36
W497	Apatou	5°09'27.57"N	54°20'16.22"W				36
W498	Apatou	5°09'27.58"N	54°20'16.23"W				35
W465	Crique Naï	5°23'47.04"N	53°42'25.96"W	Ouest			36
W474	Montagne de Fer	5°21'07.37"N	53°32'48.57"W				36
W475	Montagne de Fer	5°21'07.38"N	53°32'48.58"W		Plateau	<i>S. sp1</i>	36
W476	Montagne de Fer	5°21'07.39"N	53°32'48.59"W				36
W477	Montagne de Fer	5°21'07.40"N	53°32'48.60"W				37

Table 1: Rappel de l'échantillonnage régional et topographique

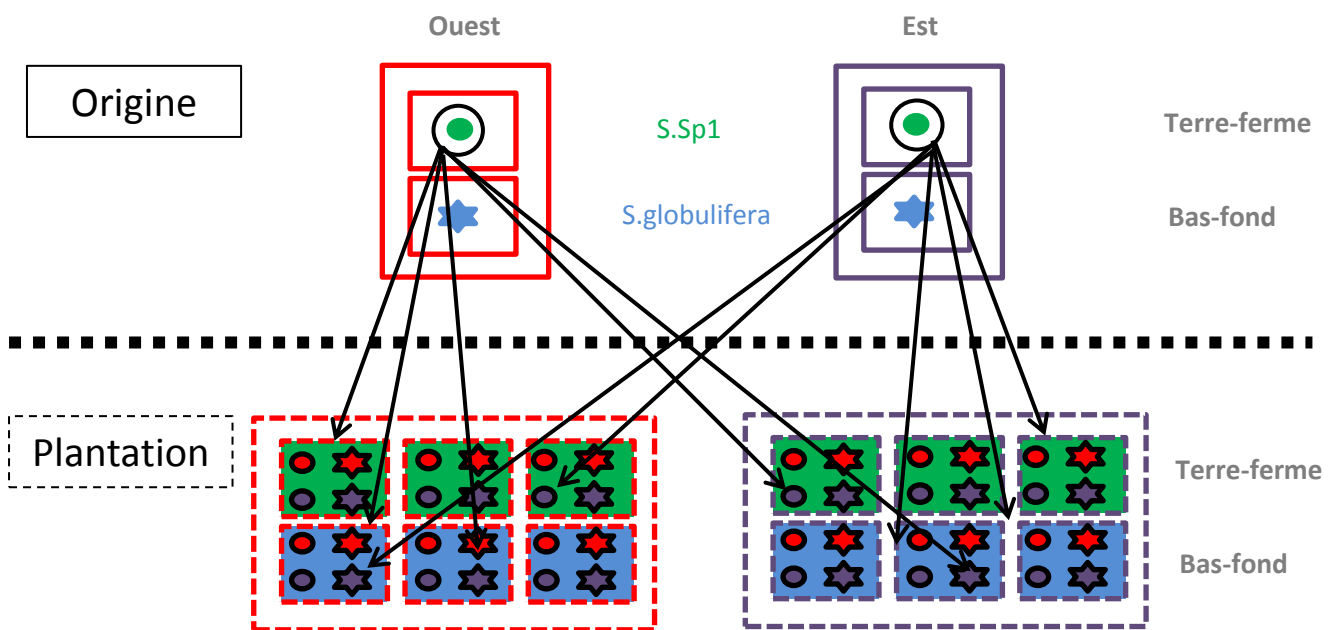


Figure 1: Rappel du design expérimental des jardins de transplantations. Les écotypes *Symphonia globulifera* et *S.sp1* provenant de l'Est et de l'Ouest de la Guyane, sont plantés en bas-fond ou terre-ferme à l'Est et à l'Ouest de la Guyane. A l'est et à l'ouest, 3 jardins en bas-fonds et 3 jardins en terre-ferme.

Facteurs propres	Ecotype	Effet génétique en réponse à la variation topographique de l'environnement
	Plantation régionale	Effet de l'environnement régional (macro-environnement)
	Habitat	Effet de l'environnement topographique
	Provenance régionale	Effet génétique d'ordre historique
Interaction Génétique x Environnement	Ecotype x Habitat	Renseigne sur la présence d'adaptation locale des écotypes
	Provenance x Plantation	Renseigne sur la présence d'une adaptation macro-environnementale associée au gradient de pluviométrie, par exemple
	Ecotype x Plantation	Renseigne sur la présence d'un effet d'interaction entre l'écotype et le site de plantation régional
	Provenance x Habitat	Renseigne sur la présence d'un effet d'interaction entre la provenance régionale et l'habitat
Interaction Génétique x Génétique	Ecotype x Provenance	Renseigne sur la présence d'un effet de la provenance régionale sur les différences entre les écotypes

Table 2: Récapitulatif des facteurs testés et interprétation de l'effet

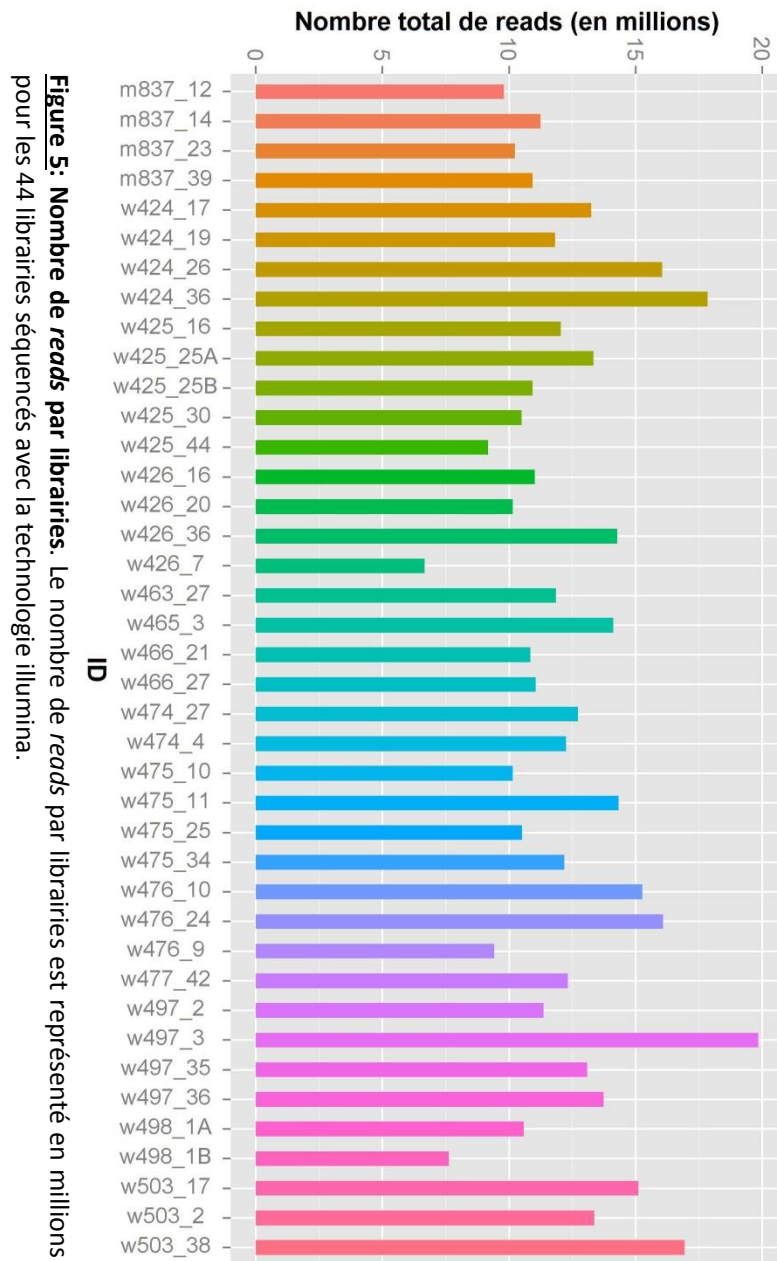


Figure 5: Nombre de reads par librairies. Le nombre de reads par librairies est représenté en millions pour les 44 librairies séquencées avec la technologie illumina.

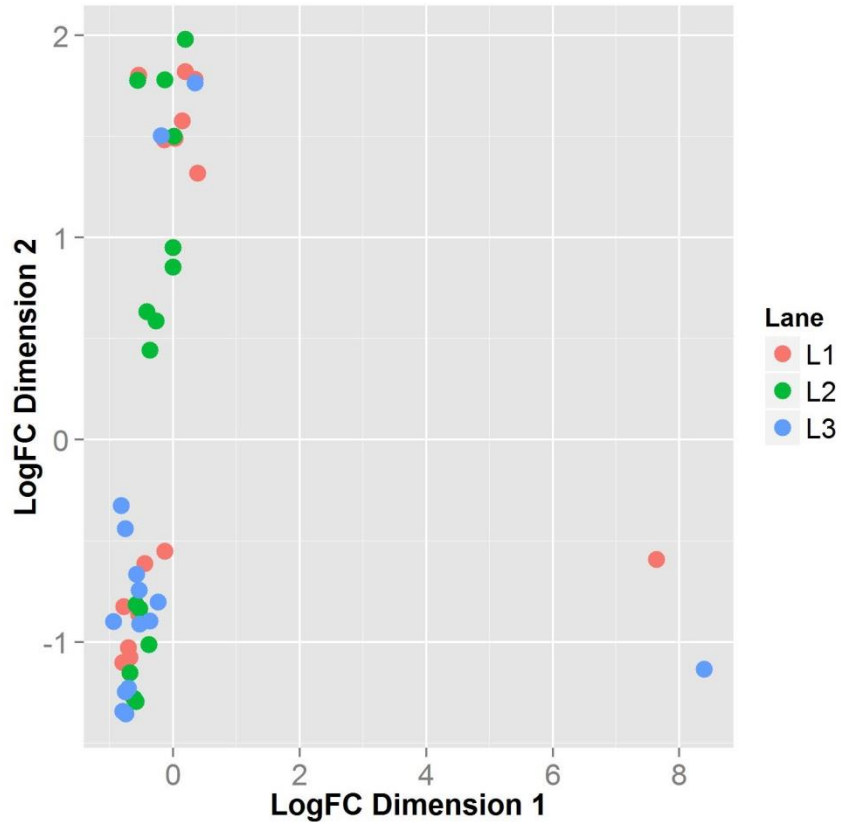


Figure 6: Relation entre les échantillons. Dans cette analyse multidimensionnelle, la distance entre les échantillons a été calculée sur la base du nombre de *reads* par librairies, en mettant en évidence l'information de la *lane* sur laquelle l'échantillon a été séquencé.

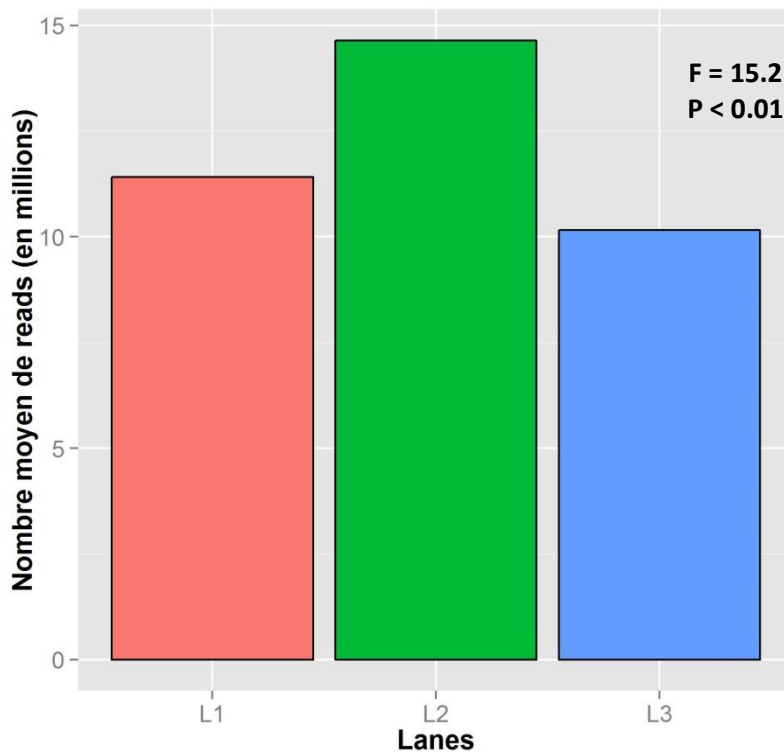


Figure 7: Nombre moyen de *reads* des librairies par *lanes*. Pour chaque *lanes*, le nombre moyen de *reads* par librairies est représenté en millions de *reads*. Une ANOVA à un facteur à permis de mettre en évidence un effet *lane* sur le nombre de *reads* par librairies

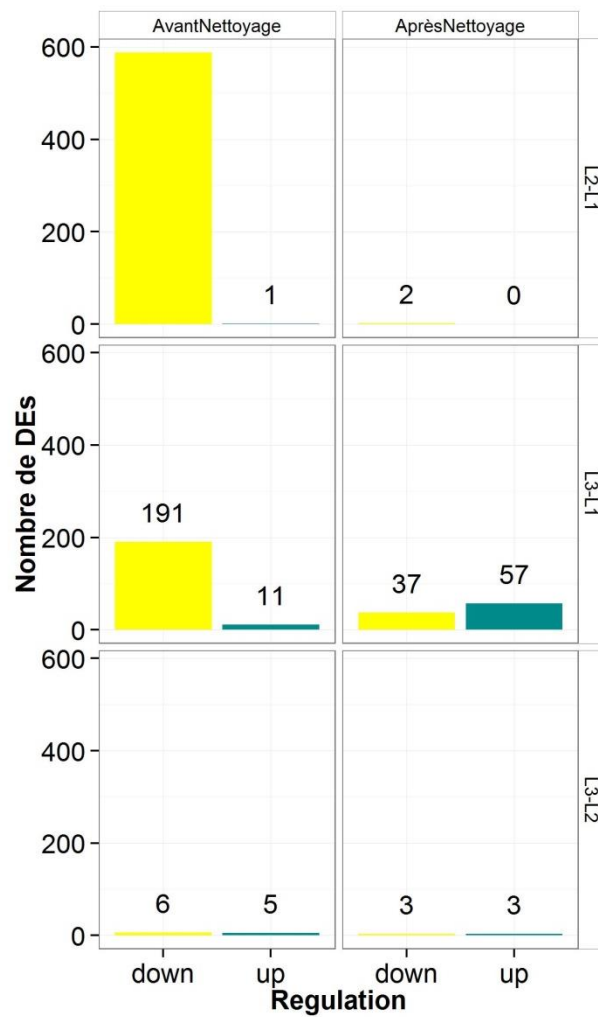


Figure 8: Nombres de séquences uniques Différentiellement Exprimés (DEs) entre les lanes. A l'aide de la fonction *decideTestsDGE* de EdgeR, le nombre de séquences uniques différenciellement exprimées par pairs de *lanes* a été calculé, avant et après le retrait des répliques par *lanes* de l'individu W426_7.

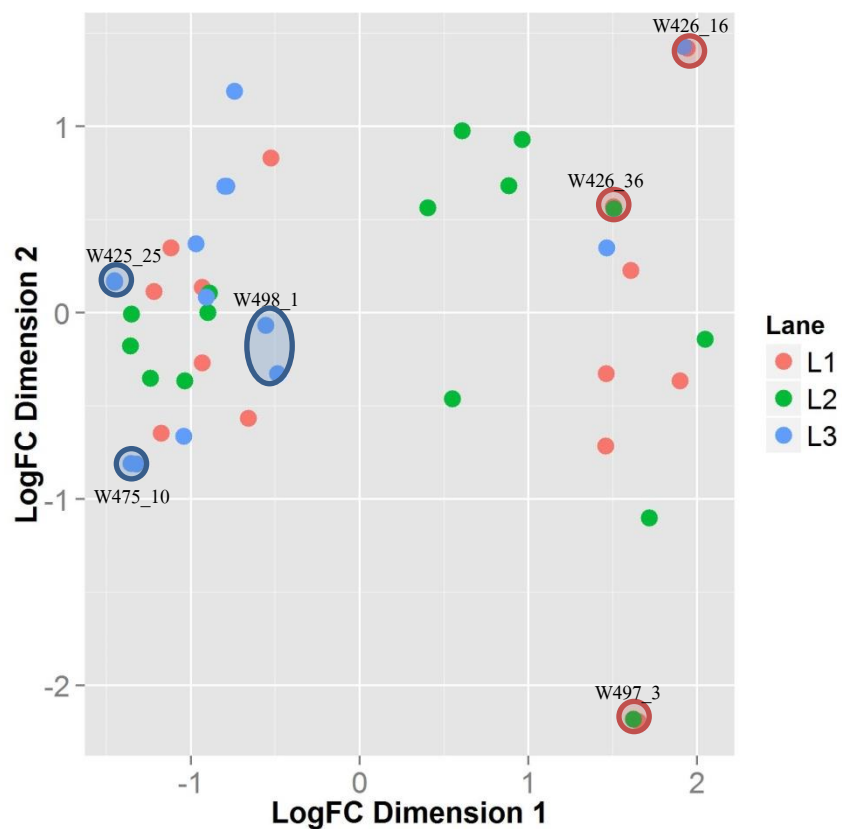


Figure 9: Relation entre les échantillons après nettoyage. Dans cette analyse multidimensionnelle, après retrait de l'individu W426_7, la distance entre les échantillons a été calculée sur la base du nombre de *reads* par librairies, en mettant en évidence l'information de la *lane* sur laquelle l'échantillon a été séquencé. Les cercles rouge indiquent les répliques de *lanes*. Les cercles bleu indiquent les répliques de librairies.

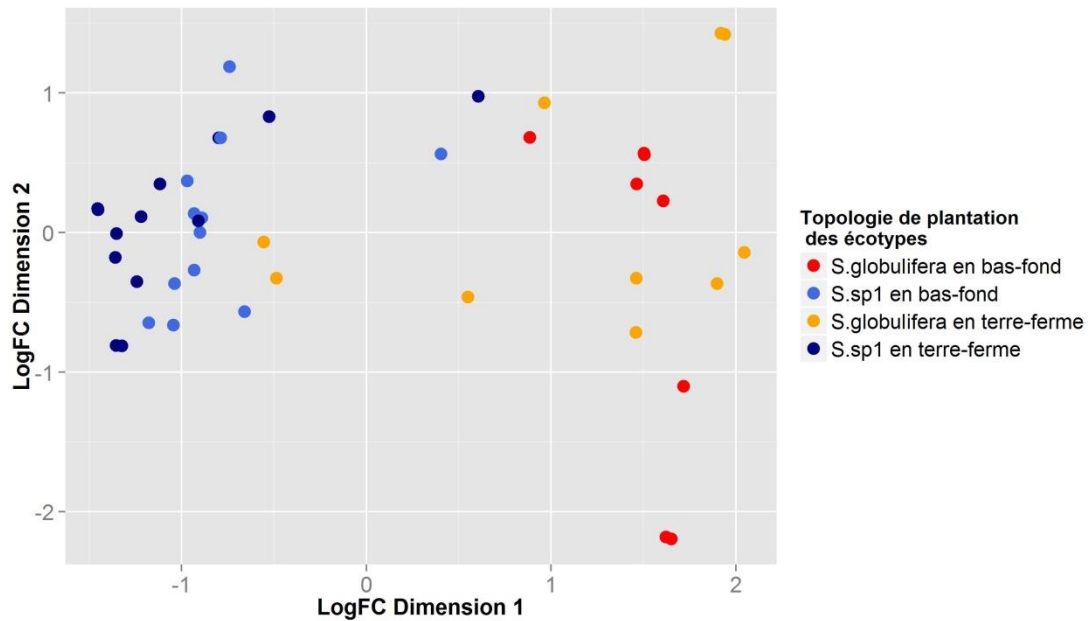


Figure 10: Relation entre les échantillons et mise en évidence de la topologie. Dans cette analyse multidimensionnelle, après retrait de l'individu W426_7, la distance entre les échantillons a été calculée sur la base du nombre de *reads* par librairies, en mettant en évidence l'écotype et la topologie de plantation (bas-fond ou *terra firme*) de l'individu. Les indications donnent le nom des individus qui s'éloignent de leur groupe, et le triangle bleu transparent indique les deux autres individus dont l'arbre mère est W424.

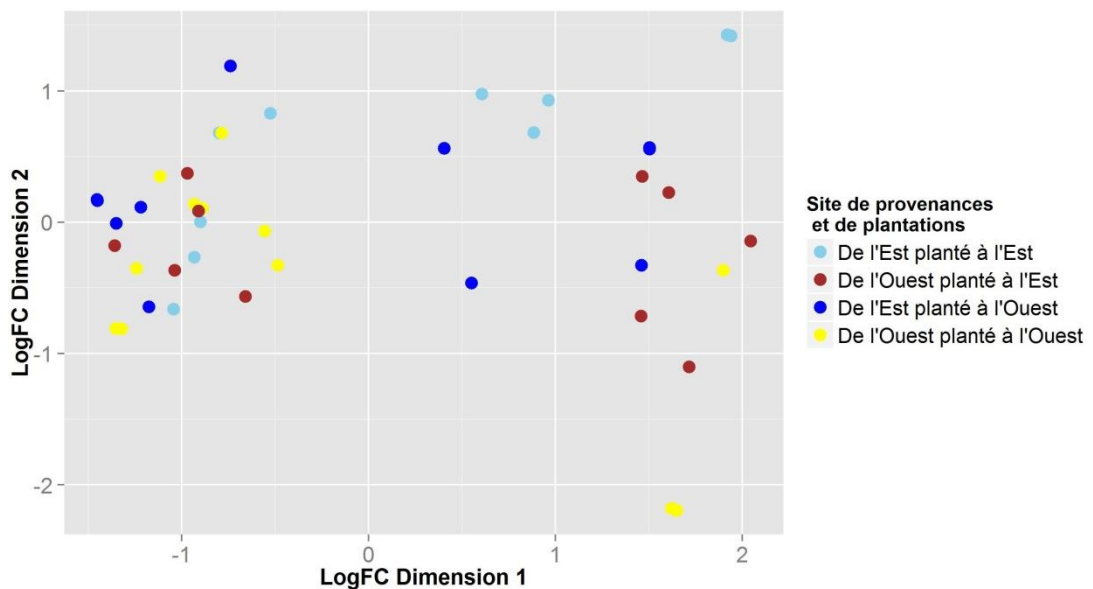


Figure 11: Relation entre les échantillons et mise en évidence de la région de provenance et de plantation. Dans cette analyse multidimensionnelle, après retrait de l'individu W426_7, la distance entre les échantillons a été calculée sur la base du nombre de *reads* par librairies, en mettant en évidence les provenances, et sites de plantations régionaux des individus

Facteurs simples

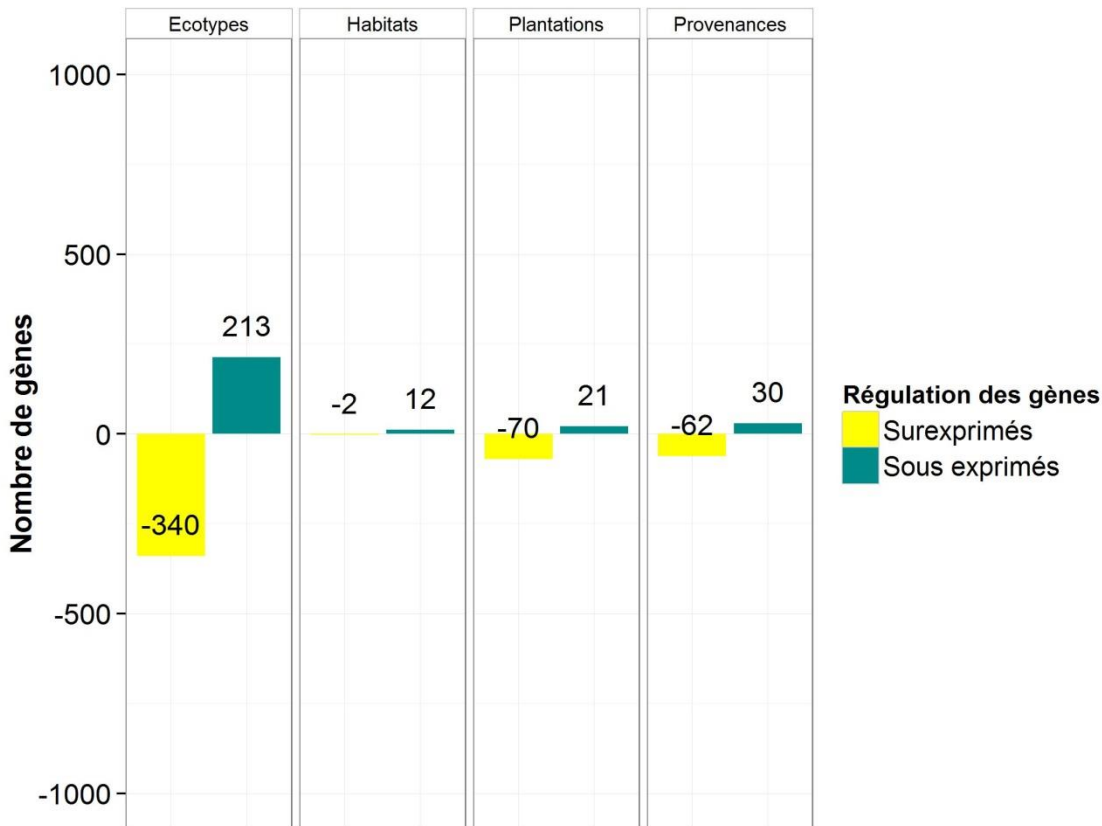


Figure 12: Compte du nombre de GDEs par facteurs simples. De gauche à droite, est présenté le nombre de séquences sous exprimés (en jaune) et surexprimés (en cyan foncé) pour le passage de *S.globulifera* à *S.sp1* (Ecotypes), le passage de bas-fond à terra firme (Habitats), le passage de site régional de plantation Est à Ouest (Plantations) et le passage de région de provenance Est à Ouest (Prov).

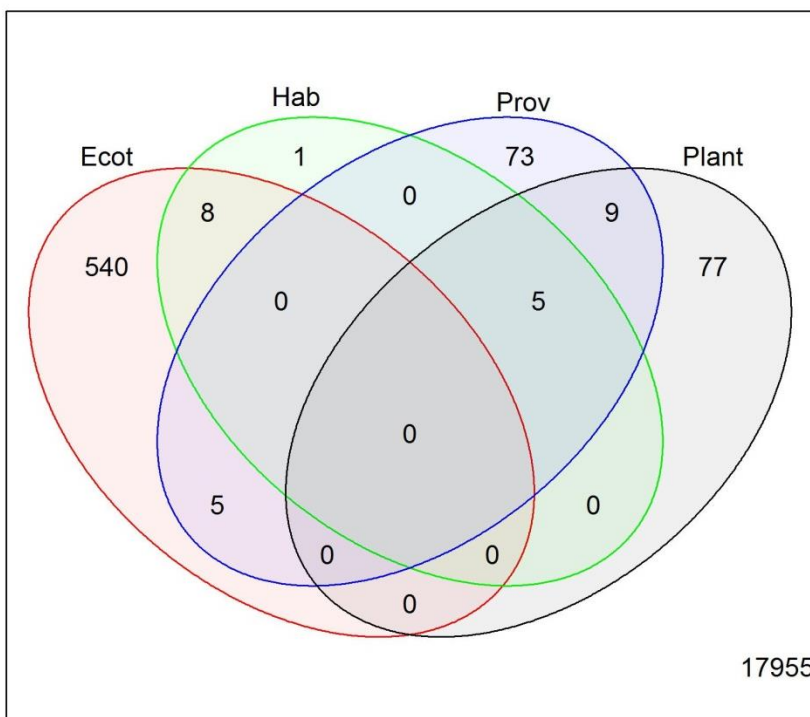


Figure 13: Diagramme de Venn des facteurs simples. Sur l'ensemble des 18673 séquences uniques, 17955 ne sont pas différemment exprimées pour les facteurs simples. Ce diagramme montre comment les 718 séquences différemment exprimées se répartissent entre les facteurs simples: l'écotype (**Ecot**), l'habitat (**Hab**), la provenance régionale (**Prov**) et le site de plantation régional (**Plant**).

Interactions topologiques

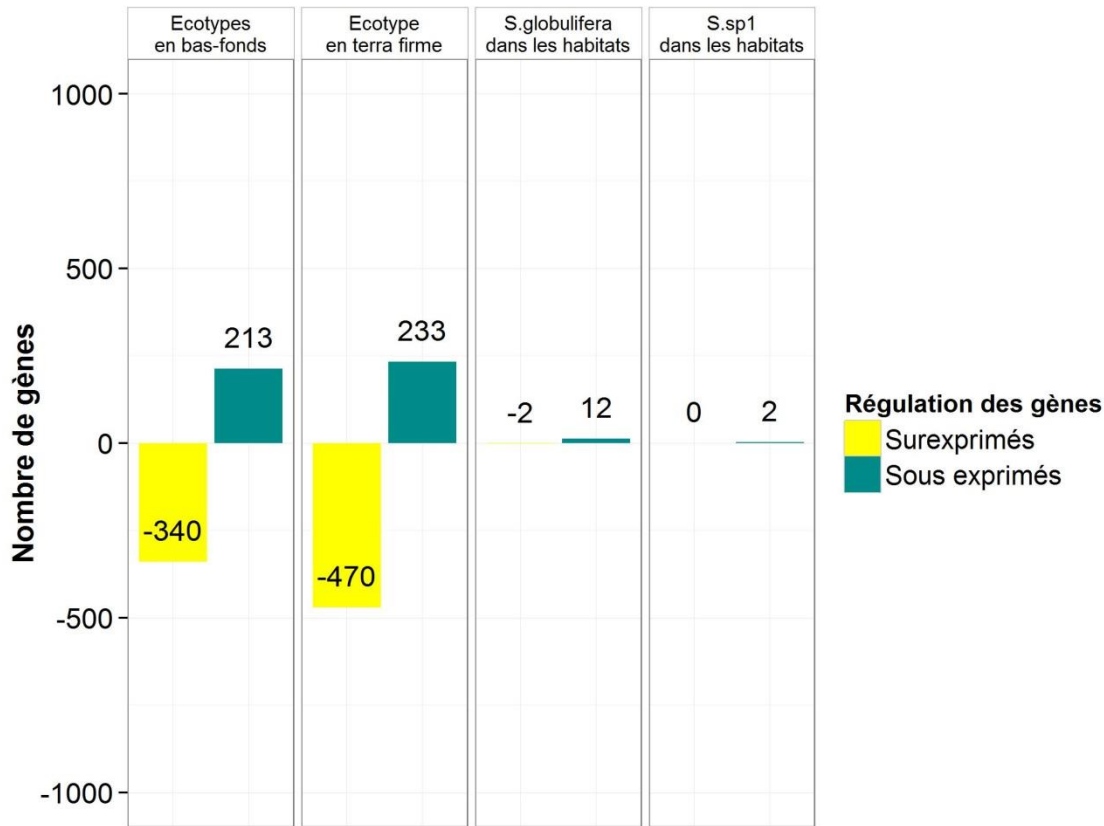


Figure 14: Compte du nombre de GDEs dans les interactions topologiques. De gauche à droite, est présenté le nombre de séquences sous exprimés (en jaune) et surexprimés (en cyan foncé) pour le passage de *S.globulifera* à *S.sp1* en bas-fonds, *S.globulifera* à *S.sp1* en terra firme, le passage de bas-fonds à terra firme pour le *S.globulifera* et le passage de bas-fonds à terra firme pour le *S.sp1*.

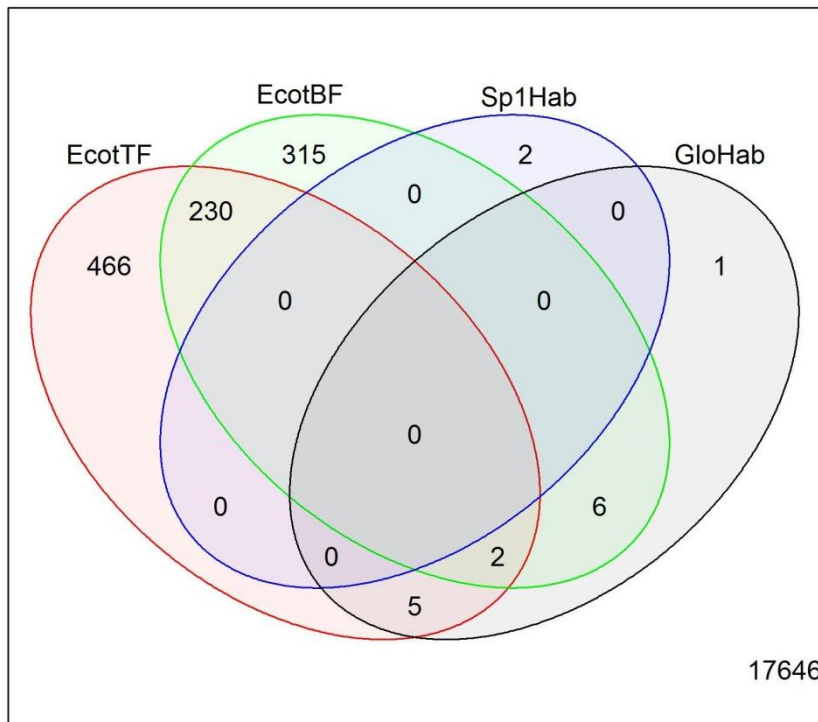


Figure 15: Diagramme de Venn des interactions topologiques. Sur l'ensemble des 18673 séquences uniques, 17646 ne sont pas différentiellement exprimées dans les interactions topologiques. Ce diagramme montre comment les 1027 séquences différentiellement exprimées se répartissent entre les facteurs emboîtés: les écotypes en bas-fonds (**EcotBF**), les écotypes en terra firme (**EcotTF**), les habitats pour le *S.globulifera* (**GloHab**) et les habitats pour le *S.sp1* (**Sp1Hab**).

Interactions régionales

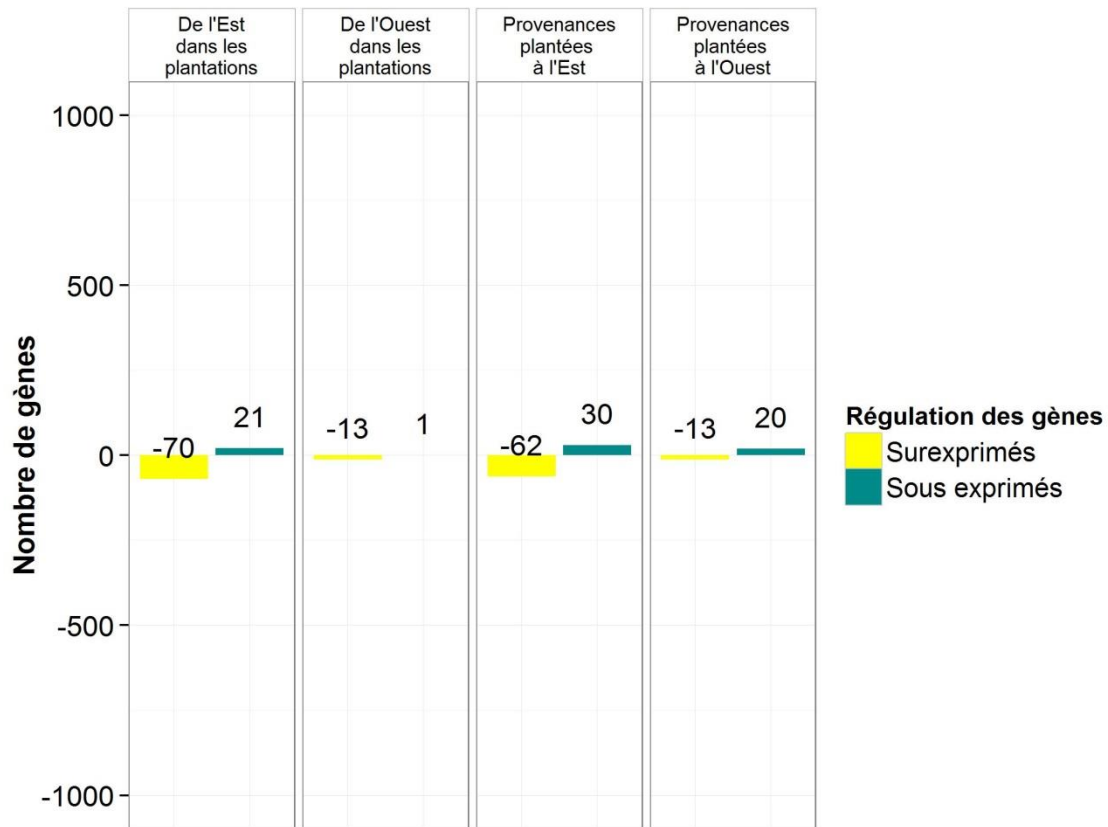


Figure 16: Compte du nombre de GDEs dans les interactions régionales. De gauche à droite, est présenté le nombre de séquences sous exprimés (en jaune) et surexprimés (en cyan foncé) entre la plantation Est à Ouest des plantules provenant de l'Est, entre la plantation Est à Ouest pour les plantules provenant de l'Ouest, entre la provenance Est à Ouest plantée à l'Est et entre la provenance Est à Ouest plantée à l'Ouest.

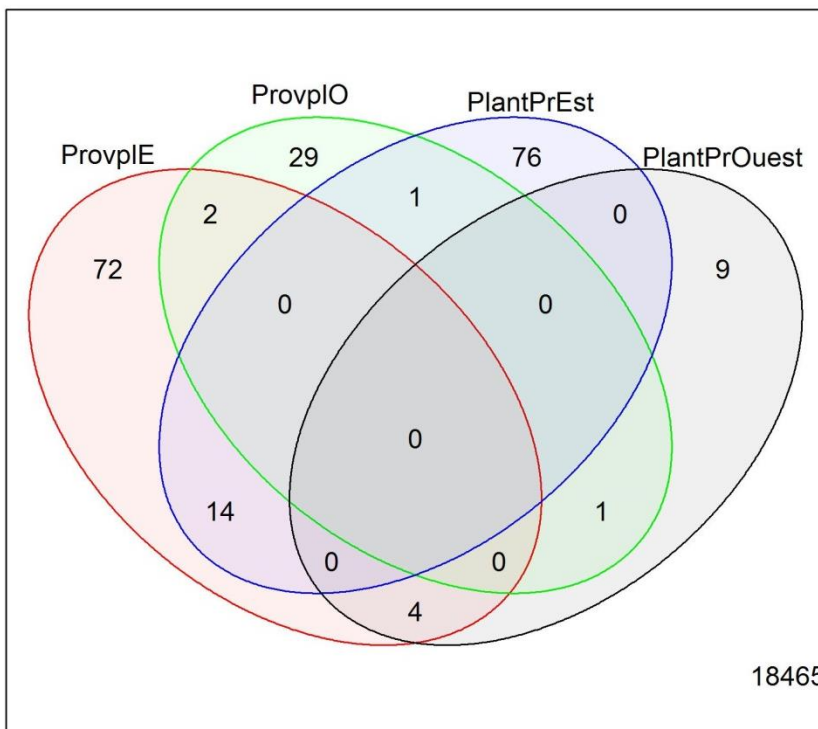


Figure 17: Diagramme de Venn des interactions régionales. Sur l'ensemble des 18673 séquences uniques, 18465 ne sont pas différentiellement exprimées dans les interactions régionales. Ce diagramme montre comment les 208 séquences différentiellement exprimées se répartissent entre les provenances régionales plantées à l'Est (**ProvpIE**), les provenances régionales plantées à l'Ouest (**ProvpIO**), les sites de plantations des plantules provenant de l'Est (**PlantPrEst**) et les sites de plantations des plantules provenant de l'Ouest (**PlantPrOuest**).

Interactions croisées _ Ecotypes et Provenances

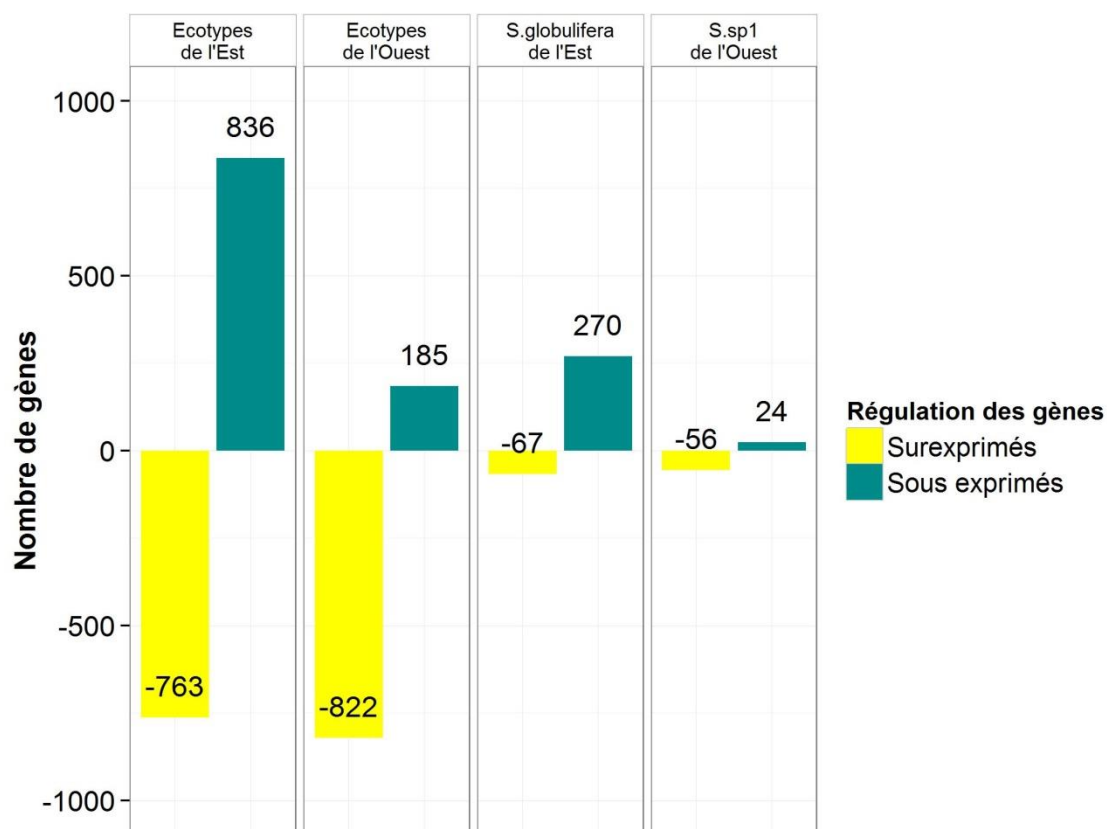


Figure 18: Compte du nombre de GDEs dans les interactions croisées entre l'écotype et de la provenance régionale des plantules. De gauche à droite, est présenté le nombre de séquences sous exprimées (en jaune) et surexprimées (en cyan foncé) entre les *S.globulifera* et les *S.sp1* quand ils proviennent de l'Est, entre les *S.globulifera* et les *S.sp1* quand ils proviennent de l'Ouest, entre les provenances régionales Est à Ouest du *S.globulifera* et entre les provenances régionales Est à Ouest du *S.sp1*.

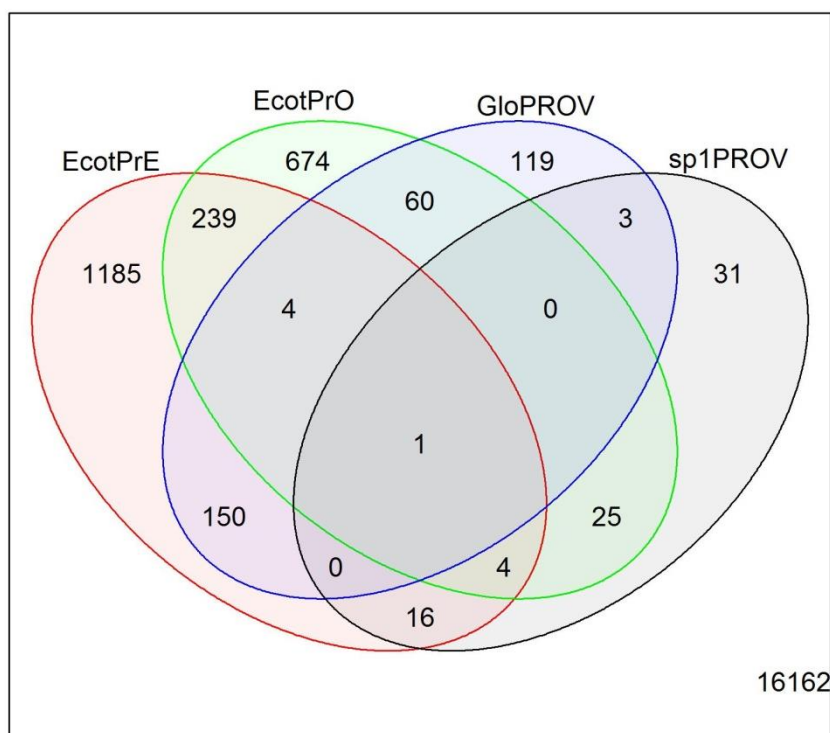


Figure 19: Diagramme de Venn des interactions croisées entre l'écotype et de la provenance régionale des plantules. Sur l'ensemble des 18673 séquences uniques, 16162 ne sont pas différentiellement exprimées dans les interactions croisées entre l'écotype et de la provenance régionale des plantules. Ce diagramme montre comment les 2511 séquences différentiellement exprimées se répartissent entre les *S.globulifera* et les *S.sp1* quand ils proviennent de l'Est (**EcotPrE**), entre les *S.globulifera* et les *S.sp1* quand ils proviennent de l'Ouest (**EcotPrO**), entre les provenances régionales Est à Ouest du *S.globulifera* (**GloPROV**) et entre les provenances régionales Est à Ouest du *S.sp1* (**Sp1PROV**).

Interactions croisées _ Habitats et Provenances



Figure 20: Compte du nombre de GDEs dans les interactions croisées entre l'habitat et de la provenance régionale des plantules. De gauche à droite, est présenté le nombre de séquences sous exprimés (en jaune) et surexprimés (en cyan foncé) entre les provenances régionales Est à Ouest quand les plantules sont en bas-fonds, entre les habitats quand les plantules proviennent de l'Est, entre les habitats quand les plantules proviennent de l'Ouest et entre les provenances régionales quand les plantules sont en terre-ferme. Les GDEs des provenances en bas-fonds et *terra firme* sont les mêmes. Il est par contre différent quand il s'agit du changement d'habitat.

Interactions croisées _ Ecotypes et Plantations

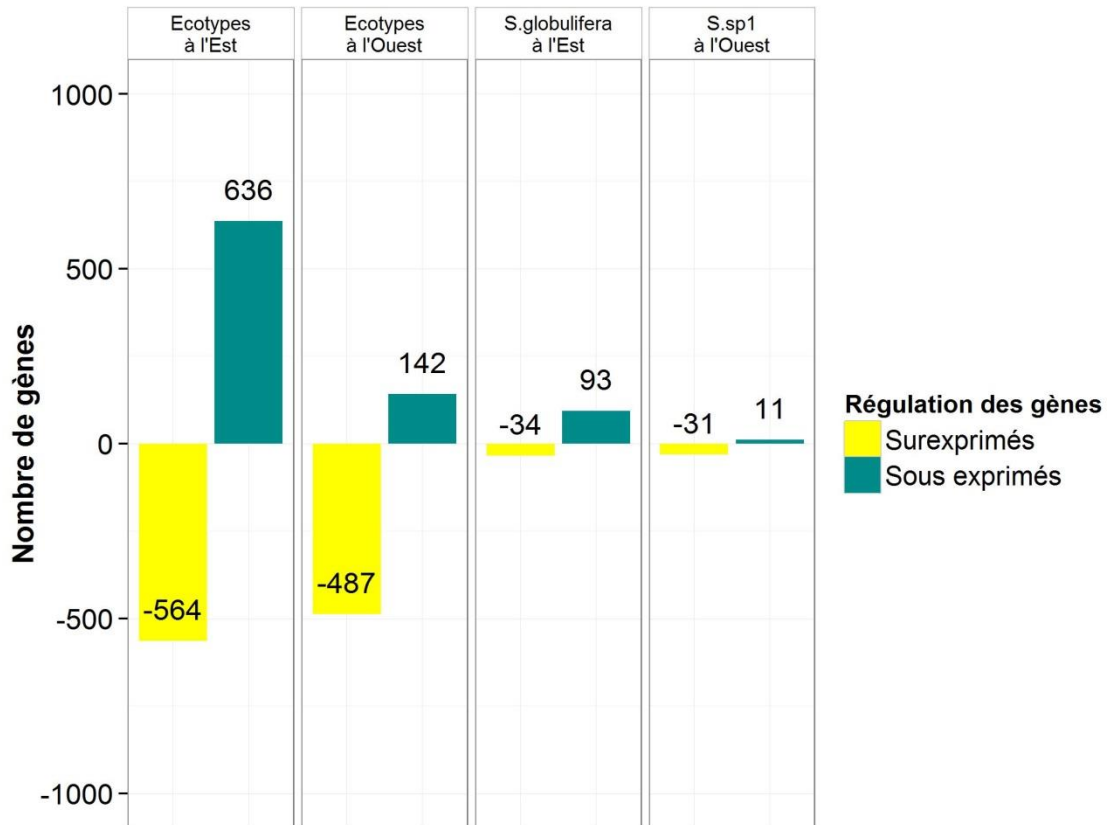


Figure 21: Compte du nombre de GDEs dans les interactions croisées entre l'écotype et le site de plantation régional. De gauche à droite, est présenté le nombre de séquences sous exprimés (en jaune) et surexprimés (en cyan foncé) entre les *S.globulifera* et les *S.sp1* quand ils sont plantés à l'Est, entre les *S.globulifera* et les *S.sp1* quand ils sont plantés à l'Ouest, entre les sites de plantations Est à Ouest pour le *S.globulifera* et entre les sites de plantations Est à Ouest pour le *S.sp1*.

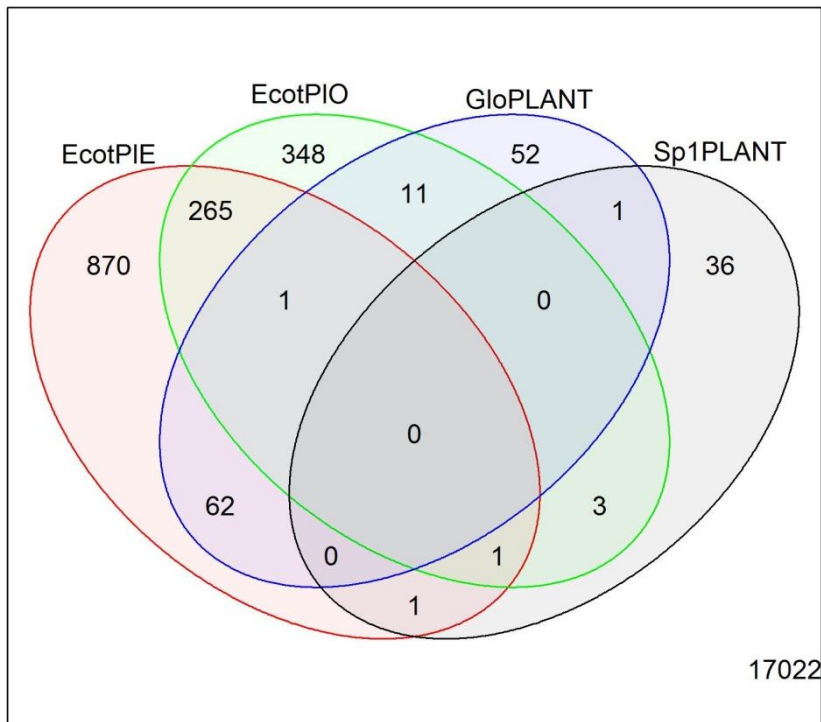


Figure 22: Diagramme de Venn des interactions croisées entre l'écotype et le site de plantation régional. Sur l'ensemble des 18673 séquences uniques, 17022 ne sont pas différentiellement exprimées dans les interactions de l'écotype avec le site de plantation régional. Ce diagramme montre comment les 1651 séquences différentiellement exprimées se répartissent entre les *S.globulifera* (**EcotPIE**) et les *S.sp1* quand ils sont plantés à l'Est, entre les *S.globulifera* et les *S.sp1* quand ils sont plantés à l'Ouest (**EcotPIO**), entre les sites de plantations Est à Ouest pour le *S.globulifera* (**GloPLANT**) et entre les sites de plantations Est à Ouest pour le *S.sp1* (**Sp1PLANT**).

```

simpleReadCounts<-function(fileList)
#fileList (character): fichier .txt contenant la liste de noms des fichiers .gff
#usage: someObjectName <- simpleReadCounts(someFileName)
{
#Téléchargement du package data.table
library(data.table)
#Chargement de la liste de noms
tableNames<-scan(fileList,what="character")
#Création de la table de comptage
d2<-list()
for (i in 1:length(tableNames))
{
d_temp<-fread(input=tableNames[i],select=1,header=T)
d1_temp<-table(d_temp)
d2_temp<-as.data.frame(d1_temp)
d2[[i]]<-d2_temp
}
for (i in 1:length(tableNames)) names(d2[[i]])[2] <-tableNames[i]
merged.data.frame <- Reduce(function(...) merge(..., all=T), d2)
}

```

Annexe 1: Script R « ReadCount » utilisé pour la production de la table de comptage.

	Modèles utilisés
Topologie	~Ecot+Ecot:Hab
	~Hab+Ecot:Hab
Région	~PROV+PROV:PLANT
	~PLANT+PROV:PLANT
Topologie x Région	~Ecot+Ecot:PROV
	~PROV+Ecot:PROV
	~Ecot+Ecot:PLANT
	~PLANT+Ecot:PLANT
	~Hab+Hab:PROV
	~PROV+Hab:PROV

Annexe 2: Modèles utilisés lors des analyses EdgeR. Le symbole ~ désigne la variable sur laquelle sera créée l'intercepte. **Ecot** représente le facteur « Ecotypes », **Hab**, le facteur « Habitats », **PROV**, le facteur « Provenances régionales » et **PLANT**, le facteur « Provenances régionales ».

BIBLIOGRAPHIE

1. Kawecki, T.J. and D. Ebert, *Conceptual issues in local adaptation*. Ecology letters, 2004. **7**(12): p. 1225-1241.
2. Pigliucci, M., *Evolution of phenotypic plasticity: where are we going now?* Trends in Ecology & Evolution, 2005. **20**(9): p. 481-486.
3. Palacio-López, K., et al., *The ubiquity of phenotypic plasticity in plants: a synthesis*. Ecology and Evolution, 2015. **5**(16): p. 3389-3400.
4. Pavey, S.A., et al., *The role of gene expression in ecological speciation*. Annals of the New York Academy of Sciences, 2010. **1206**(1): p. 110-129.
5. Merilä, J. and A.P. Hendry, *Climate change, adaptation, and phenotypic plasticity: the problem and the evidence*. Evolutionary Applications, 2014. **7**(1): p. 1-14.
6. Alvarez, M., A.W. Schrey, and C.L. Richards, *Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution?* Molecular Ecology, 2015. **24**(4): p. 710-725.
7. De Jong, G., *Evolution of phenotypic plasticity: patterns of plasticity and the emergence of ecotypes*. New Phytologist, 2005. **166**(1): p. 101-118.
8. Chalancon, G., et al., *Interplay between gene expression noise and regulatory network architecture*. Trends Genet, 2012. **28**(5): p. 221-32.
9. Stapley, J., A.W. Santure, and S.R. Dennis, *Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species*. Molecular Ecology, 2015. **24**(9): p. 2241-2252.
10. Padovan, A., et al., *Transcriptome Sequencing of Two Phenotypic Mosaic *Eucalyptus* Trees Reveals Large Scale Transcriptome Re-Modelling*. PLOS One, 2015. **10**(5): p. e0123226.
11. Rapp, R., et al., *Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication*. BMC Biology, 2010. **8**(1): p. 139.
12. Guo, J., et al., *Widespread and adaptive alterations in genome-wide gene expression associated with ecological divergence of two *Oryza* species*. Molecular Biology and Evolution, 2015: p. msv196.
13. Pelgas, B., et al., *QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments*. BMC Genomics, 2011. **12**(1): p. 145.
14. Robertson, M. and C. Richards, *Non-genetic inheritance in evolutionary theory - the importance of plant studies*, in *Non-Genetic Inheritance* 2015.
15. Skinner, M.K., et al., *Epigenetics and the Evolution of Darwin's Finches*. Genome Biology and Evolution, 2014. **6**(8): p. 1972-1989.
16. Bossdorf, O., C.L. Richards, and M. Pigliucci, *Epigenetics for ecologists*. Ecology letters, 2008. **11**(2): p. 106-115.
17. Liu, X., et al., *Regulation of Histone Methylation and Reprogramming of Gene Expression in the Rice Inflorescence Meristem*. The Plant Cell, 2015: p. tpc. 15.00201.
18. Tian, L., et al., *Reversible histone acetylation and deacetylation mediate genome-wide, promoter-dependent and locus-specific changes in gene expression during plant development*. Genetics, 2005. **169**(1): p. 337-345.
19. Li, X., et al., *High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression*. The Plant Cell, 2008. **20**(2): p. 259-276.
20. Tian, L. and Z.J. Chen, *Blocking histone deacetylation in Arabidopsis induces pleiotropic effects on plant gene regulation and development*. Proceedings of the National Academy of Sciences, 2001. **98**(1): p. 200-205.
21. Dubin, M.J., et al., *DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation*. eLife, 2015. **4**.
22. Fedoroff, N.V., *Transposable Elements, Epigenetics, and Genome Evolution*. Science, 2012. **338**(6108): p. 758-767.
23. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nat Rev Genet, 2009. **10**(1): p. 57-63.
24. Wilhelm, B.T. and J.-R. Landry, *RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing*. Methods, 2009. **48**(3): p. 249-257.
25. Oshlack, A., M. Robinson, and M. Young, *From RNA-seq reads to differential expression results*. Genome Biology, 2010. **11**(12): p. 220.

26. Hegde, P., et al., *A concise guide to cDNA microarray analysis*. BioTechniques, 2000. **29**(3): p. 548-563.
27. Jozefczuk, J. and J. Adjaye, *Chapter Six - Quantitative Real-Time PCR-Based Analysis of Gene Expression*, in *Methods in Enzymology*, M.V. Daniel Jameson and V.W. Hans, Editors. 2011, Academic Press. p. 99-109.
28. Mortazavi, A., et al., *Mapping and quantifying mammalian transcriptomes by RNA-Seq*. Nat Methods, 2008. **5**: p. 621 - 628.
29. Coley, P.D., *Effects of leaf age and plant life history patterns on herbivory*. Nature, 1980. **284**(5756): p. 545-546.
30. Auer, P. and R. Doerge, *Statistical design and analysis of RNA sequencing data*. Genetics, 2010. **185**: p. 405 - 416.
31. Alberti, A., et al., *Comparison of library preparation methods reveals their impact on interpretation of metatranscriptomic data*. BMC Genomics, 2014. **15**(1): p. 912.
32. Langmead, B. and S. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Methods, 2012. **9**(4): p. 357 - 359.
33. R Development Core Team, *R: A language and environment for statistical computing*, 2008, R Foundation for Statistical Computing: Vienna, Austria.
34. Ayroles, J.F. and G. Gibson, *Analysis of variance of microarray data*. Methods Enzymol, 2006. **411**: p. 214-33.
35. Robinson, M., D. McCarthy, and G. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 2010. **26**: p. 139 - 140.
36. Brousseau, L., et al., *High-throughput transcriptome sequencing and preliminary functional analysis in four Neotropical tree species*. BMC Genomics, 2014. **15**(1): p. 238.
37. t Hoen, P.A.C., et al., *Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories*. Nat Biotech, 2013. **31**(11): p. 1015-1022.
38. Van Oosten, M.J., et al., *The Role of the Epigenome in Gene Expression Control and the Epimark Changes in Response to the Environment*. Critical Reviews in Plant Sciences, 2014. **33**(1): p. 64-87.
39. Jablonka, E.V.A. and M.J. Lamb, *The Changing Concept of Epigenetics*. Annals of the New York Academy of Sciences, 2002. **981**(1): p. 82-96.
40. Dalziel, A.C., et al., *Adaptation and acclimation of aerobic exercise physiology in Lake Whitefish ecotypes (Coregonus clupeaformis)*. Evolution, 2015. **69**(8): p. 2167-2186.
41. Wang, R.-L., et al., *The limits of selection during maize domestication*. Nature, 1999. **398**: p. 236-239.
42. Hua, L., et al., *LABA1, a Domestication Gene Associated with Long, Barbed Awns in Wild Rice*. The Plant Cell, 2015. **27**(7): p. 1875-1888.
43. Doebley, J. and L. Lukens, *Transcriptional Regulators and the Evolution of Plant Form*. The Plant Cell, 1998. **10**(7): p. 1075-1082.
44. Pigliucci, M., C.J. Murren, and C.D. Schlichting, *Phenotypic plasticity and evolution by genetic assimilation*. Journal of Experimental Biology, 2006. **209**(12): p. 2362-2367.
45. Rockman, M.V. and L. Kruglyak, *Genetics of global gene expression*. Nat Rev Genet, 2006. **7**(11): p. 862-872.
46. Bräutigam, K., et al., *Epigenetic regulation of adaptive responses of forest tree species to the environment*. Ecology and Evolution, 2013. **3**(2): p. 399-415.
47. Meier, K., et al., *Local Adaptation at the Transcriptome Level in Brown Trout: Evidence from Early Life History Temperature Genomic Reaction Norms*. PLOS One, 2014. **9**(1): p. e85171.
48. Dayan, D.I., D.L. Crawford, and M.F. Oleksiak, *Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of Fundulus heteroclitus*. Molecular Ecology, 2015. **24**(13): p. 3345-3359.
49. Skinner, M.K., *Environmental Epigenetics and a Unified Theory of the Molecular Aspects of Evolution: A Neo-Lamarckian Concept that Facilitates Neo-Darwinian Evolution*. Genome Biology and Evolution, 2015. **7**(5): p. 1296-1302.
50. Healy, T.M. and P.M. Schulte, *Phenotypic plasticity and divergence in gene expression*. Molecular Ecology, 2015. **24**(13): p. 3220-3222.

Discussion

<i>Plan d'expérience et méthodes.....</i>	<i>143</i>
L'expérimentation en jardins de transplantations réciproques	143
Le <i>Symphonia globulifera</i>	143
Milieu naturel et contrôlé	144
Le RNAseq	144
Etude de la méthylation et choix de l'utilisation des AFLPs.....	145
Un plus en épigénétique	145
Le choix du MSAP	145
<i>Conséquences évolutives pour le modèle Symphonia</i>	<i>147</i>
Divergence écologique chez <i>Symphonia globulifera</i>	147
L'histoire complexe du <i>Symphonia globulifera</i>	148
<i>Conclusion</i>	<i>149</i>
<i>Bibliographie</i>	<i>150</i>

Durant ce projet de thèse, je me suis attelée à étudier les mécanismes génétiques impliqués dans l'accumulation de la divergence phénotypique entre des écotypes et qui pourraient être impliqués dans la spéciation écologique chez les arbres forestiers tropicaux, en prenant comme modèle l'espèce *Symphonia globulifera*. Dans cette discussion générale, je discute d'abord le choix du plan d'expérimentation et des techniques que j'ai utilisées ou mises en œuvre afin de tester mes hypothèses, ainsi que leurs caractéristiques innovantes. Dans une deuxième partie, je donne mon interprétation de l'histoire de la divergence du *S. globulifera* en proposant une ouverture sur de futures études qui pourraient être effectuées sur cette espèce.

PLAN D'EXPERIENCE ET METHODES

L'EXPERIMENTATION EN JARDINS DE TRANSPLANTATIONS RECIPROQUES

LE SYMPHONIA GLOBULIFERA

Malgré l'étonnante biodiversité qu'arborent les forêts tropicales humides, peu d'espèces d'arbres tropicaux sont utilisées dans les études de divergence[1] et de spéciation[2, 3], laissant une place importante aux espèces modèles et espèces évoluant dans les milieux tempérés. Les espèces modèles ont généralement un temps de génération court et leur génome est souvent connu, facilitant les expérimentations ; d'autre part, les espèces tempérées sont beaucoup plus accessibles. D'un autre côté, il me paraît très difficile de comprendre les mécanismes de divergence entre espèces sans y inclure les espèces tropicales qui ont la particularité d'occuper des milieux très hétérogènes[4], constitués de microclimats et de micro-habitats contribuant à la coexistence des espèces[5] et au maintien de la biodiversité à l'intérieur des forêts tropicales[4, 6].

Mon modèle d'étude est le *Symphonia globulifera*, qui fait l'objet de nombreuses publications concernant ses propriétés chimiques [7-11]. Ce modèle biologique est également intéressant pour l'étude des mécanismes de dispersion en milieu tropical [12-15], pour la reconstruction de l'histoire ancienne des forêts [1, 16-18] et de la divergence entre les espèces[19-22]. En effet, étant présent en Afrique, en Amérique centrale et en Amérique du Sud, il fait un très bon modèle pour l'étude des dispersions anciennes et de la divergence des populations et des espèces, d'autant plus qu'en Guyane et au Suriname[23] et peut-être même en Afrique[17], il existe deux écotypes à l'intérieur de l'espèce, l'un spécialiste des bas-fonds et l'autre de la *terra firme*.

MILIEU NATUREL ET CONTROLE

Deux expérimentations de ce projet reposent sur des jardins de transplantations réciproques. La première partie de mes travaux m'a permis de montrer une adaptation locale du *S.sp1* à la *terra firme*; tandis que la troisième partie de mon projet m'a permis de montrer des profils d'expression génique différents entre les écotypes. Le jardins de transplantations en milieu naturel permet de maîtriser les facteurs génétiques et environnementaux. Il me paraît être l'outil idéal pour détecter l'adaptation locale. Dans le milieu naturel, beaucoup de facteurs entrent en jeu dans les interactions Genotype-Environnement. Tant et si bien qu'il est difficile de les recréer en serre. Ainsi, une expérimentation en milieu naturel est plus propice à l'observation de la réponse plastique ou adaptative, dont le but est de comprendre les mécanismes de divergence et les facteurs qui en sont responsables[24]. De plus, les patrons de croissance et mortalité observés en conditions naturelles sont interprétables directement en termes de *fitness*, ce qui permet une interprétation « naturelle » des résultats. Par un exemple, l'étude de Baraloto et al.[6] sur le suivi de la croissance et la survie de plantules des écotypes *S.sp1* et *S.globulifera*, menée en serre, a montré une association entre le bas-fond et le *S.globulifera*, mais pas entre le *S.sp1* et la *terra firme*.

Enfin, lors de l'établissement de l'expérimentation, il n'a pas seulement été planté des plantules dans les différentes conditions, mais aussi des graines. Cela m'a permis d'observer non seulement les patrons de croissance et survie, mais également les taux de germination, qui contribuent à la compréhension de la répartition des espèces dans leur milieu naturel ; cela me paraît être un avantage très important dans les expérimentations de jardins de transplantations menées sur des arbres. En effet, les arbres en général, et les arbres forestiers tropicaux non héliophiles en particulier, ont de très longs temps de génération (de l'ordre de plusieurs décennies) et une très forte mortalité lors des premières années de vie[25]. Sachant qu'il est donc impossible de mesurer la fécondité dans des jardins de transplantations réciproques chez ces espèces, il me paraît très important d'observer également la germination des graines.

LE RNASEQ

Dans le cadre de ce projet, j'ai choisi d'observer le niveau d'expression des écotypes à l'intérieur des jardins de transplantations réciproques. Afin de montrer une expression constitutive différente entre les écotypes, l'utilisation d'un « jardin commun » en serre aurait pu être suffisante. En étant dans le même milieu et subissant les mêmes contraintes, j'aurais maîtrisé le facteur environnemental et pu mettre en évidence des différences d'expression génique entre les écotypes, qui gardent leurs différences morphologiques en milieu contrôlé

[6, 26]. Il est toutefois beaucoup plus informatif d'étudier l'expression différentielle en transplantations réciproques, car cela permet de faire la part entre l'effet du fond génétique et de l'environnement. Au final, mes résultats montrent que peu de séquences diffèrent les écotypes dans les habitats, malgré un nombre de séquences important les différenciant quand tous les habitats sont confondus, et donc quel que soit le milieu.

Aussi, l'utilisation du RNAseq en combinaison avec des jardins de transplantations réciproques est innovante et cela d'autant plus que le matériel biologique est une espèce tropicale.

ETUDE DE LA METHYLATION ET CHOIX DE L'UTILISATION DES AFLPS

UN PLUS EN EPIGENETIQUE

Depuis une dizaine d'années, l'épigénétique émerge notamment dans les domaines de l'étude de la divergence écologique et de la spéciation [27-33]. Aussi, très intéressée par le rôle que peut jouer l'épigénétique dans ces domaines de recherche, j'ai voulu introduire cette hypothèse dans mon étude. La méthylation de l'ADN est la plus étudiée des marques épigénétiques, devant l'acétylation des histones et la condensation de l'ADN[34], car depuis beaucoup plus longtemps étudiée dans les recherches sur le développement[35]. De ce fait, de nombreuses techniques ont été développées pour l'observation de la méthylation[36].

Dans notre étude, des Analyses en Composantes Principales nous ont permis d'observer que les variations génomiques entre des populations de l'est et de l'ouest de *S.globulifera* et *S.sp1* sont expliquées par des variations de la séquence d'ADN plutôt que par des variations de l'état de méthylation du génome. Pourtant, les locus à variant de méthylation sont sélectionnés, et en plus grand nombre à Paracou qu'à Régina. Je pense que même si elle ne semble jouer qu'un petit rôle dans la divergence des écotypes de *Symphonia globulifera*, la méthylation de l'ADN est un paramètre important à prendre en compte dans les études de la divergence des espèces en présence de flux de gènes, car elle a déjà été beaucoup impliquée dans la divergence à l'intérieur d'autres complexes d'espèces[28, 29, 37-40], mais aussi dans la réponse à l'environnement[27].

LE CHOIX DU MSAP

Dans l'étude du rôle de la méthylation dans la divergence à l'intérieur de l'espèce *Symphonia globulifera*, j'ai choisi d'utiliser le génotypage MSAP, qui repose sur des polymorphismes de présence/absence de fragments PCR produits par la digestion de l'ADN par des enzymes

méthyl-sensibles, et se basant sur un protocole AFLP. Je voulais suivre les variations de méthylation sur tout le génome de façon à pouvoir effectuer une analyse de génétique des populations et comparer les profils de méthylation des écotypes dans leur environnement naturel. Ainsi comme le montre cet arbre de décision (Figure 1), j'avais le choix entre l'utilisation d'anticorps, le MSAP et la conversion au bisulfite[39]. Par rapport aux habitudes et compétences du laboratoire ainsi qu'au faible coût de la technique MSAP, cette dernière me paraît être le meilleur choix pour une première analyse exploratoire. Les analyses ont été effectuées sur des arbres adultes en population naturelle. Je pense qu'il aurait été intéressant de suivre l'état de méthylation du génome dans les jardins de transplantations réciproques, mais échantillonner les plantules pour le génotypage MSAP était un risque trop grand en raison du petit nombre de feuilles portées par certaines plantules dans les jardins de transplantations. Cette étude était exploratoire, car les résultats issus d'une analyse MSAP ne sont ni suffisamment reproductibles (comme pour les AFLPs, dont les MSAP sont une application), ni directement interprétables en termes d'étude de l'impact de la méthylation sur l'état d'expression des gènes. Je pense qu'il serait intéressant d'effectuer une conversion au bisulfite suivi d'un séquençage à haut-débit, à présent que nous avons vu qu'il existe un profil de méthylation différents entre les écotypes, et qui participe à l'adaptation. Cela permettrait de confirmer les résultats du génotypage MSAP sur un nombre de locus beaucoup plus important, et avec une méthode très reproductible[40].

CONSEQUENCES EVOLUTIVES POUR LE MODELE *SYMPHONIA*

DIVERGENCE ECOLOGIQUE CHEZ *SYMPHONIA GLOBULIFERA*

En démontrant une origine génétique à la divergence entre les traits phénotypiques qui différencient les deux écotypes du genre *Symphonia* (hauteur, nombre des feuilles et diamètre), ce travail de thèse a permis, d'une part, (1) d'exclure l'hypothèse d'écotypes « labiles » répondant seulement aux variations de l'environnement, et d'autre part, (2) de rendre crédible l'hypothèse que cette divergence écotypique puisse être le résultat d'une adaptation locale à l'environnement, ceci étant déduit de la présence d'interactions Génotype x Environnement pour des traits de croissance et de fitness. Du point de vue génomique, des gènes spécifiques seraient la cible de la sélection divergente qui agirait, en fonction des milieux, sur des allèles différents avec des effets contrastés sur le phénotype. L'augmentation de la fréquence allélique pour des allèles différents à l'intérieur de chaque écotype se caractérise par des valeurs de différenciation génétique significativement supérieure à l'attendue sous un modèle neutre. L'analyse de génome scan sur les AFLPs non sensibles à la méthylation va dans ce sens puisque j'ai identifié 68 locus potentiellement soumis à la sélection divergente, soit seulement 8%. Ce pourcentage est dans la gamme de ceux obtenus dans d'autres couples d'espèces divergentes [41-43]

La présence d'une divergence phénotypique inter-écotypes qui résultent de l'adaptation locale associée à la présence de quelques régions génomiques impliquées dans la différenciation génétique s'apparente bien aux premières phases d'une spéciation écologique en condition de sympatrie, c'est-à-dire sans barrière géographique aux flux de gènes [44-46].

Aussi, la différence de l'expression des gènes semble être un mécanisme important et nécessaire dans ce type modèle de spéciation [24, 47-49]. Mon étude va dans ce sens, avec un niveau d'expression des gènes qui est différent entre les écotypes *S.globulifera* et *S.sp1*, qui montrent une faible différenciation génétique, et cela alors qu'il y a bien une adaptation locale des écotypes à leur milieu (bas-fonds et *terra firme*). L'adaptation locale, réalisée par le biais du changement de niveau d'expression, peut-être de nature génétique ou épigénétique, car les locus soumis à variation épigénétique (méthylation de l'ADN) peuvent être sélectionnés dans le génome [50, 51], comme le montre notre analyse MSAP.

Enfin, dans la « théorie unifiée de l'évolution » [52], les marques épigénétiques répondent rapidement aux variations de l'environnement à l'intérieur d'une génération et sont héréditaires [53, 54]. De plus, la méthylation des cytosines augmente le taux de mutation des cytosines en thymines [55]. Ainsi, cette théorie propose une divergence engendrée d'abord par

des marques épigénétiques qui vont permettre aux individus de répondre rapidement aux variations de l'environnement, avant que les divergences d'expression ne soient stabilisées dans le génome par l'établissement de variations génétiques. Mes résultats semblent aller dans ce sens. Le génotypage des écotypes du *Symphonia* a montré que des loci soumis à la variation de la méthylation de l'ADN sont en minorité sélectionnés dans le génome, et avec un faible F_{ST} moyen. Ainsi, les méthylations de l'ADN sont nombreuses dans le génome des écotypes et seules quelques-unes, celle qui répondent correctement aux pressions de l'environnement sont maintenues et pourraient permettre des mutations de séquences qui renforceraient la différenciation génétique des écotypes.

L'HISTOIRE COMPLEXE DU *SYMPHONIA GLOBULIFERA*

Au Costa-Rica, les populations de *Symphonia globulifera* sont des arbres d'une dizaine de mètres présents dans les régions montagneuses[56]. En Guyane, nous avons deux écotypes. L'écotype *S.sp1* est spécialiste des *terra firme* et est plus petit en taille que le *S.globulifera*, spécialiste des bas-fonds[23]. Au Brésil, il semble y avoir deux morphotypes du *Symphonia globulifera*[57]. En Afrique, il y a des *Symphonia globulifera* en bas-fonds et dans les montagnes[17]. Ainsi, peut-il y avoir une histoire plus complexe de l'espèce *Symphonia globulifera* qui est originaire d'Afrique ? Est-ce que l'écotype *S.sp1* peut avoir une origine africaine également, et que les deux écotypes aient subi les mêmes routes de dispersions ? Ou est-ce que la divergence est récente dans les différentes régions, auquel cas nous pourrions faire l'hypothèse de divergences parallèles comme dans le cas de l'épinoche[58]. Nous avons vu dans les jardins de transplantations réciproques que le *S.globulifera* germe mal en *terra firme* par rapport au *S.sp1*. Il se pourrait alors que, s'il n'existe pas d'écotype du type *S.globulifera* au Costa-Rica, c'est parce que les conditions environnementales du Costa-Rica ne sont pas ou n'étaient pas favorables à son établissement. Ainsi, il serait intéressant de comparer génétiquement (génotypage microsatellites ou génotypage haut-débit) les populations de tous ces sites, en incluant l'information des morphotypes, et sur le milieu dans lequel ils apparaissent, de façon à tester l'hypothèse que le *S.sp1* que l'on retrouve en Guyane et le *Symphonia* que l'on retrouve au Costa-Rica ou au Brésil, pourraient provenir de l'écotype spécialiste des montagnes en Afrique. Dans l'hypothèse d'une adaptation parallèle au milieu dans les différentes régions du monde, il serait envisageable d'identifier les gènes communs à l'accumulation de divergence adaptative[58].

CONCLUSION

Cette thèse en s'appuyant sur des techniques de génétique quantitative dans les jardins de transplantation réciproques, sur du génotypage AFLP appliqués sur des arbres adultes en population naturelles et sur du séquençage à haut débit du transcriptomes (RNAseq) de plantules en jardins de transplantations réciproques a permis de montrer la présence d'une adaptation locale au sein de l'espèce *Symphonia globulifera*, avec deux écotypes, l'un spécialiste du bas-fond, le *S.globulifera* et l'autre spécialiste de la *terra firme*, le *S.sp1*. Ces résultats appuient l'hypothèse d'une spéciation écologique en cours. Aussi, les amplifications AFLP mettant en évidence des profils de méthylation de l'ADN et des profils de variations de séquences, confirment les résultats auxquels nous nous attendions et déjà obtenu pour ces écotypes à savoir un calcul de F_{ST} faible indiquant la présence d'un flux de gènes entre les écotypes. Les marqueurs de la méthylation ont été « outliers » dans les tests de sélections mais ont montré une F_{ST} plus faible que le marqueur à variation de la séquence de l'ADN, suggérant une faible participation de leur part à l'adaptation locale des écotypes. Nous avons donc montré une adaptation locale en présence de flux de gènes ou avec un flux de gènes interrompu récemment, signifiant que nous pourrions être en début de spéciation des écotypes qui divergent pour correspondre au mieux à leur habitat. Aussi, les résultats du séquençage RNAseq, effectués sur les plantules en jardins de transplantations réciproques, montre un niveau d'expression différentiel des gènes entre les écotypes sans relation avec l'habitat dans lequel ils évoluent.

Cette thèse est innovante car elle s'appuie sur un modèle d'arbre tropical et associe jardins de transplantation au séquençage à haut-débit. C'est également une thèse qui se situe dans les thèmes de recherche actuels en combinant les études de l'adaptation locale à l'analyse de l'expression des gènes, en jardins de transplantation réciproques.

BIBLIOGRAPHIE

1. Dauby, G., et al., *Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover*. *Molecular Ecology*, 2014. **23**(9): p. 2299-2312.
2. Leimu, R. and M. Fischer, *A Meta-Analysis of Local Adaptation in Plants*. *PLOS One*, 2008. **3**(12): p. e4010.
3. Chen, G.F. and D.W. Schemske, *Ecological differentiation and local adaptation in two sister species of Neotropical Costus (Costaceae)*. *Ecology*, 2014.
4. Baraloto, C. and P. Couteron, *Fine-scale Microhabitat Heterogeneity in a French Guianan Forest*. *Biotropica*, 2010. **42**(4): p. 420-428.
5. Baraloto, C., D. Goldberg, and D. Bonal, *Performance trade-offs among tropical tree seedlings in contrasting microhabitats*. *Ecology*, 2005. **86**(9): p. 2461-2472.
6. Baraloto, C., et al., *Seasonal water stress tolerance and habitat associations within four neotropical tree genera*. *Ecology*, 2007. **88**(2): p. 478-489.
7. Cottet, K., et al., *Comparative LC-MS-based metabolite profiling of the ancient tropical rainforest tree Symphonia globulifera*. *Phytochemistry*, 2014.
8. Fromentin, Y., et al., *Symphonia globulifera, a widespread source of complex metabolites with potent biological activities*. *Planta Med*, 2015. **81**(2): p. 95-107.
9. Cottet, K., et al., *Isolation of Guttiferones from Renewable Parts of Symphonia globulifera by Centrifugal Partition Chromatography*. *Planta Med*, 2015. **21**: p. 21.
10. Nkengfack, A.E., et al., *Globulixanthes A and B, Two New Cytotoxic Xanthenes with Isoprenoid Groups from the Root Bark of Symphonia globulifera*. *Journal of Natural Products*, 2002. **65**(5): p. 734-736.
11. Ngouela, S., et al., *Anti-plasmodial and antioxidant activities of constituents of the seed shells of Symphonia globulifera Linn f.* *Phytochemistry*, 2006. **67**(3): p. 302-306.
12. Dick, C.W. and M. Heuertz, *The complex biogeographic history of a widespread tropical tree species*. *Evolution*, 2008. **62**(11): p. 2760-2774.
13. da Silva Carneiro, F., et al., *High levels of pollen dispersal detected through paternity analysis from a continuous Symphonia globulifera population in the Brazilian Amazon*. *Forest Ecology and Management*, 2009. **258**: p. 1260-1266.
14. Aldrich, P.R. and J.L. Hamrick, *Reproductive Dominance of Pasture Trees in a Fragmented Tropical Forest Mosaic*. *Science*, 1998. **281**(5373): p. 103-105.
15. Poelchau, M.F. and J.L. Hamrick, *Palaeodistribution modelling does not support disjunct Pleistocene refugia in several Central American plant taxa*. *Journal of Biogeography*, 2013. **40**(4): p. 662-675.
16. Dick, Christopher W., K. Abdulah Salim, and E. Bermingham, *Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree*. *The American Naturalist*, 2003. **162**(6): p. 691-703.
17. Budde, K.B., et al., *The ancient tropical rainforest tree Symphonia globulifera L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa*. *Heredity*, 2013. **111**(1): p. 66-76.
18. Tokuoka, T. and H. Tobe, *Phylogenetic analyses of Malpighiales using plastid and nuclear DNA sequences, with particular reference to the embryology of Euphorbiaceae sens. str.* *Journal of Plant Research*, 2006. **119**(6): p. 599-616.
19. Hardy, O.J., et al., *Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species*. *Molecular Ecology*, 2006. **15**(2): p. 559-571.
20. Vinson, C., et al., *Characterization and isolation of DNA microsatellite primers for the tropical tree, Symphonia globulifera Linn. f.* *Molecular Ecology Notes*, 2005. **5**: p. 202-204.
21. Brousseau, L., et al., *High-throughput transcriptome sequencing and preliminary functional analysis in four Neotropical tree species*. *BMC Genomics*, 2014. **15**(1): p. 238.
22. Da Silva Carneiro, F., et al., *Low Interannual Variation of Mating System and Gene Flow of Symphonia globulifera in the Brazilian Amazon*. *Biotropica*, 2007. **39**(5): p. 628-636.
23. Loubry, D., *Déterminisme du comportement phénologique des arbres en forêt tropicale humide de Guyane française (5°lat. N.)*. PhD Thesis, University of Paris 6, 1994.
24. Alvarez, M., A.W. Schrey, and C.L. Richards, *Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution?* *Molecular Ecology*, 2015. **24**(4): p. 710-725.

25. Petit, R.J. and A. Hampe, *Some Evolutionary Consequences of Being a Tree*. Annual Review of Ecology, Evolution, and Systematics, 2006. **37**(1): p. 187-214.
26. Brousseau, L., et al., *Highly local environmental variability promotes intrapopulation divergence of quantitative traits: an example from tropical rain forest trees*. Ann Bot, 2013. **112**(6): p. 1169-79.
27. Lukens, L.N. and S. Zhan, *The plant genome's methylation status and response to stress: implications for plant improvement*. Current Opinion in Plant Biology, 2007. **10**(3): p. 317-322.
28. Bräutigam, K., et al., *Epigenetic regulation of adaptive responses of forest tree species to the environment*. Ecology and Evolution, 2013. **3**(2): p. 399-415.
29. Schulz, B., R.L. Eckstein, and W. Durka, *Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb*. Molecular Ecology, 2014. **23**(14): p. 3523-3537.
30. Durand, S., et al., *Rapid Establishment of Genetic Incompatibility through Natural Epigenetic Variation*. Current Biology, 2012. **22**(4): p. 326-331.
31. Bossdorf, O., C.L. Richards, and M. Pigliucci, *Epigenetics for ecologists*. Ecology letters, 2008. **11**(2): p. 106-115.
32. Rapp, R.A. and J.F. Wendel, *Epigenetics and plant evolution*. New Phytologist, 2005. **168**(1): p. 81-91.
33. Skipper, M., *An epigenetic cause of adaptive variation*. Nat Rev Genet, 2006. **7**(12): p. 900-900.
34. Van Oosten, M.J., et al., *The Role of the Epigenome in Gene Expression Control and the Epimark Changes in Response to the Environment*. Critical Reviews in Plant Sciences, 2014. **33**(1): p. 64-87.
35. Jablonka, E.V.A. and M.J. Lamb, *The Changing Concept of Epigenetics*. Annals of the New York Academy of Sciences, 2002. **981**(1): p. 82-96.
36. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics*. Nat Rev Genet, 2008. **9**(6): p. 465-476.
37. Pruneda-Paz, J.L., et al., *A Functional Genomics Approach Reveals CHE as a Component of the Arabidopsis Circadian Clock*. Science, 2009. **323**(5920): p. 1481-1485.
38. Richards, E.J., *Natural epigenetic variation in plant species: a view from the field*. Current Opinion in Plant Biology, 2011. **14**(2): p. 204-209.
39. Richards, C.L., A.W. Schrey, and M. Pigliucci, *Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation*. Ecology letters, 2012. **15**(9): p. 1016-1025.
40. Alonso, C., et al., *MSAP markers and global cytosine methylation in plants: a literature survey and comparative analysis for a wild-growing species*. Molecular Ecology Resources, 2015: p. n/a-n/a.
41. Scotti-Saintagne, C., et al., *Genome Scanning for Interspecific Differentiation Between Two Closely Related Oak Species [Quercus robur L. and Q. petraea (Matt.) Liebl.]*. Genetics, 2004. **168**: p. 1615-1626.
42. Nosil, P., S.P. Egan, and D.J. Funk, *Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection*. Evolution, 2008. **62**(2): p. 316-336.
43. Stolting, K.N., et al., *Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species*. Mol Ecol, 2013. **22**(3): p. 842-55.
44. Nosil, P. and J.L. Feder, *GENOME EVOLUTION AND SPECIATION: TOWARD QUANTITATIVE DESCRIPTIONS OF PATTERN AND PROCESS*. Evolution, 2013. **67**(9): p. 2461-2467.
45. Feder, J.L., et al., *Geographic Mode of Speciation and Genomic Divergence*. Annual Review of Ecology, Evolution, and Systematics, 2013. **44**(1): p. 73-97.
46. Pinho, C. and J. Hey, *Divergence with gene flow: models and data*. Annual Review of Ecology, Evolution, and Systematics, 2010. **41**(1): p. 215-230.
47. Pavey, S.A., et al., *The role of gene expression in ecological speciation*. Annals of the New York Academy of Sciences, 2010. **1206**(1): p. 110-129.
48. Duhoux, A., et al., *RNA-Seq analysis of rye-grass transcriptomic response to an herbicide inhibiting acetolactate-synthase identifies transcripts linked to non-target-site-based resistance*. Plant molecular biology, 2015. **87**(4-5): p. 473-487.
49. Dayan, D.I., D.L. Crawford, and M.F. Oleksiak, *Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of Fundulus heteroclitus*. Molecular Ecology, 2015. **24**(13): p. 3345-3359.
50. Herrera, C.M. and P. Bazaga, *Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis*. New Phytologist, 2010. **187**(3): p. 867-876.
51. Platt, A., et al., *Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations*. Molecular Ecology, 2015. **24**(15): p. 3823-3830.

52. Skinner, M.K., *Environmental Epigenetics and a Unified Theory of the Molecular Aspects of Evolution: A Neo-Lamarckian Concept that Facilitates Neo-Darwinian Evolution*. *Genome Biology and Evolution*, 2015. **7**(5): p. 1296-1302.
 53. Flatscher, R., et al., *Environmental Heterogeneity and Phenotypic Divergence: Can Heritable Epigenetic Variation Aid Speciation?* *Genetics Research International*, 2012.
 54. Johannes, F., et al., *Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits*. *PLoS Genet*, 2009. **5**(6): p. e1000530.
 55. Lutsenko, E. and A.S. Bhagwat, *Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells: A model, its experimental support and implications*. *Mutation Research/Reviews in Mutation Research*, 1999. **437**(1): p. 11-20.
 56. Gill, G.E., R.T. Fowler, and S.A. Mori, *Pollination Biology of Symphonia globulifera (Clusiaceae) in Central French Guiana*. *Biotropica*, 1998. **30**(1): p. 139-144.
 57. Bittrich, V. and M.E. Amaral, *Pollination biology of Symphonia globulifera (Clusiaceae)*. *Plant Systematics and Evolution*, 1996. **200**(1-2): p. 101-110.
 58. Jones, F.C., et al., *The genomic basis of adaptive evolution in threespine sticklebacks*. *Nature*, 2012. **484**(7392): p. 55-61.
-