

# INTRODUCTION.....1

## REVUE BIBLIOGRAPHIQUE

<b>I. POLYPLOÏDIE : Modifications des génomes, du transcriptome et du phénotype.....</b>	<b>3</b>
<b>I.1 GENERALITES SUR LA POLYPLOÏDIE ET LES POLYPLOÏDES.....</b>	<b>3</b>
<b>I.1.1 Notion de spéciation.....</b>	<b>3</b>
I.1.1.1 <i>Définition.....</i>	<i>3</i>
I.1.1.2 <i>Les spéciations homoploïde et polyploïde.....</i>	<i>3</i>
<b>I.1.2 Le phénomène de polyploïdisation.....</b>	<b>4</b>
I.1.2.1 <i>Les différents types de polyploïdie.....</i>	<i>4</i>
I.1.2.2 <i>La formation des polyploïdies.....</i>	<i>5</i>
I.1.2.3 <i>Les différents comportements méiotiques chez les polyploïdes.....</i>	<i>6</i>
<b>I.1.3 Les espèces polyploïdes chez les êtres vivants.....</b>	<b>6</b>
I.1.3.1 <i>Fréquence et récurrence de la polyploïdie chez les végétaux...6</i>	<i>6</i>
I.1.3.2 <i>Fréquence et récurrence de la polyploïdie chez les animaux...7</i>	<i>7</i>
<b>I.1.4 Rôle de la polyploïdie dans l'adaptation et l'évolution des espèces.....</b>	<b>7</b>
I.1.4.1 <i>Diversification et évolution des espèces.....</i>	<i>8</i>
I.1.4.2 <i>Adaptation et plasticité phénotypique.....</i>	<i>8</i>
<b>I.2 ROLE DE LA POLYPLOÏDIE DANS L'EXPRESSION GENIQUE.....</b>	<b>10</b>
<b>I.2.1. Etudes du transcriptome.....</b>	<b>11</b>
I.2.2.1 <i>Notions d'additivité et de non-additivité.....</i>	<i>11</i>
I.2.2.2 <i>Observation de l'expression relative des gènes homéologues.....</i>	<i>12</i>
I.2.2.3 <i>Les autopolyploïdes et les variations d'expression génique...12</i>	<i>12</i>
I.2.2.4 <i>Les allopolyploïdes et les variations d'expression génique....13</i>	<i>13</i>
<b>I.2.2 Etude du protéome, métabolome.....</b>	<b>16</b>
<b>I.3 MECANISMES DE REGULATION CONTRIBUANT AUX VARIATIONS DANS L'EXPRESSION DES GENES CHEZ LES POLYPLOÏDES.....</b>	<b>17</b>
<b>I.3.1 Le devenir des copies de gènes dupliqués.....</b>	<b>17</b>
I.3.1.1 <i>La pseudogénéisation et la néofonctionnalisation.....</i>	<i>17</i>
I.3.1.2 <i>La Sous-fonctionnalisation.....</i>	<i>18</i>
I.3.1.3 <i>Conservation et équilibre de dosage.....</i>	<i>18</i>
<b>I.3.2 Les mécanismes de régulation.....</b>	<b>19</b>
I.3.2.1 <i>Augmentation des possibilités de variations dans l'expression « dose-dépendante » des gènes.....</i>	<i>20</i>
I.3.2.2 <i>Altération des interactions régulatrices.....</i>	<i>20</i>
I.3.2.3 <i>Mécanismes génétiques et épigénétiques.....</i>	<i>21</i>
<b>I.3.3 Evolution des espèces polyploïdes.....</b>	<b>25</b>
<b>II. COFFEA ARABICA : MODELE BIOLOGIQUE D'ETUDE.....</b>	<b>26</b>
<b>II.1 LE GENRE COFFEA.....</b>	<b>26</b>
<b>II.1.1 Classification.....</b>	<b>26</b>
<b>II.1.2 Relations phylogénétiques.....</b>	<b>27</b>

<b>II.2 LE CAS DE L'UNIQUE ALLOTETRAPLOÏDE, <i>COFFEA ARABICA</i> : ORIGINES ET CARACTERISTIQUES.....</b>	<b>28</b>
<b>II.2.1 Origine de <i>Coffea arabica</i>.....</b>	<b>28</b>
<b>II.2.2 Caractéristiques du génome et système de reproduction de <i>C. arabica</i>.....</b>	<b>28</b>
<b>II.2.3 Ressources génétiques et amélioration chez <i>Coffea arabica</i>.....</b>	<b>29</b>
<b>II.2.4 Les caféiers et leur milieu naturel.....</b>	<b>31</b>
<b>II.2.5 L'adaptation des caféiers en conditions de culture.....</b>	<b>32</b>
<b>III. OBJECTIFS DE LA THESE.....</b>	<b>33</b>
 <b>CHAPITRE I : MISE AU POINT DE LA PREMIERE PUCES A ADN CHEZ LE CAFE</b>	
<b>Introduction au chapitre.....</b>	<b>34</b>
<b>Article « The 'PUCE CAFÉ' project: the first 15K coffee microarray, a new tool for discovering candidate genes correlated to agronomic and quality traits. <i>BioMed central Genomics</i> 12: 5.».....</b>	<b>37</b>
 <b>CHAPITRE II : DIVERGENCE ET DISTANCES TRANSCRIPTOMIQUES SOUS CONTRAINTES THERMIQUE ENTRE ALLOPOLYPLOÏDE ET SES ESPECES PARENTALES DIPLOÏDES</b>	
<b>Introduction au chapitre.....</b>	<b>51</b>
<b>PARTIE I : DIVERGENCE TRANSCRIPTOMIQUE ET EXPRESSION NON-ADDITIVE CHEZ L'ALLOPOLYPLOÏDE ET SES ESPECES PARENTALES DIPLOÏDES</b>	
<b>Article « Genomic expression dominance in the natural allopolyploid <i>Coffea arabica</i> is massively affected by growth temperature. <i>New Phytologist</i>.».....</b>	<b>53</b>
<b>PARTIE II : DISTANCE DU TRANSCRIPTOME DE L'ESPECE ALLOPOLYPLOÏDE PAR RAPPORT A CELLES DE SES DEUX ESPECES PARENTALES SOUS CONTRAINTES THERMIQUES</b>	
<b>INTRODUCTION.....</b>	<b>68</b>
<b>RESULTATS.....</b>	<b>68</b>
<i>Mesures de la distance euclidienne basée sur les variations du transcriptome des espèces en fonction des conditions de températures.....</i>	<b>68</b>
<i>Détermination des unigènes communs aux trois espèces ou exclusifs en réponse au stress de chaleur.....</i>	<b>68</b>
<i>Les processus biologiques impliqués dans les groupes d'unigènes communs aux trois espèces ou 'spécifiques' à chacune d'elle.....</i>	<b>70</b>
<b>DISCUSSION et CONCLUSION.....</b>	<b>72</b>
 <b>CHAPITRE III : PLASTICITE DES ALLOPOLYPLOÏDES FACE A DIFFERENTS REGIME THERMIQUES : COMPARAISON ENTRE COFFEA ARABICA ET SES DEUX ESPECES PARENTALES</b>	
<b>INTRODUCTION.....</b>	<b>74</b>
<b>MATERIELS ET METHODES.....</b>	<b>78</b>
<i>Matériels biologiques.....</i>	<b>78</b>
<i>Croissance en phytotron.....</i>	<b>78</b>
<i>Analyses chimiques.....</i>	<b>79</b>
<i>Analyses de qRT-PCR (Quantitative Reverse Transcriptase Polymerase Chain Reaction).....</i>	<b>81</b>
<i>Analyses statistiques.....</i>	<b>83</b>

<b>RESULTATS</b> .....	<b>84</b>
<i>Influence des contraintes thermiques sur la croissance et la teneur en chlorophylle des trois espèces de caféiers étudiés</i> .....	<b>84</b>
<i>Abondance relative des principaux acides gras chez les trois espèces de caféiers étudiées en fonction des régimes thermiques</i> .....	<b>85</b>
<i>Variation de l'expression de gènes codant pour les oméga-3 et omega-6 désaturases en fonction des différents régimes thermiques</i> .....	<b>86</b>
<b>DISCUSSION</b> .....	<b>88</b>
<i>Croissance et teneur en chlorophylle chez les trois espèces de caféiers</i> .....	<b>89</b>
<i>Corrélation entre l'activité des enzymes désaturases et l'abondance relative des acides gras chez les trois espèces de caféiers suivant des régimes thermiques contrastés</i> .....	<b>90</b>
<b>CONCLUSION</b> .....	<b>92</b>
<b>CHAPITRE IV : COMPARAISON DU TRANSCRIPTOME DE COFFEA ARABICA AVEC LES TRANSCRIPTOMES DE TROIS GROUPES GENETIQUES DE COFFEA CANEPHORA</b>	
<b>INTRODUCTION</b> .....	<b>93</b>
<b>MATERIELS ET METHODES</b> .....	<b>95</b>
<i>Matériels biologiques</i> .....	<b>95</b>
<i>Extraction des ARN totaux</i> .....	<b>95</b>
<i>Séquençage</i> .....	<b>96</b>
<i>Analyses statistiques</i> .....	<b>96</b>
<b>RESULTATS</b> .....	<b>98</b>
<i>Estimation de la puissance et de la répétabilité de la technologie RNAseq</i> .....	<b>98</b>
<i>Variabilité des profils de transcriptome des trois groupes génétiques de C. canephora comparés à l'espèce allopolyploïde C. arabica</i> .....	<b>100</b>
<i>Distances entre les transcriptomes des caféiers étudiés</i> .....	<b>100</b>
<i>Divergences transcriptomiques entre l'allopolyploïde C. arabica et son espèce parentale C. canephora</i> .....	<b>101</b>
<i>Les gènes traduisant la plasticité de l'allopolyploïde, C. arabica</i> .....	<b>103</b>
<b>DISCUSSION ET CONCLUSION</b> .....	<b>104</b>
<b>DISCUSSION GENERALE ET PERSPECTIVES</b> .....	<b>107</b>
<b>BIBLIOGRAPHIE</b> .....	<b>119</b>
<b>ANNEXES</b> .....	<b>152</b>



# Liste des figures

## *Revue bibliographique*

Figure 1 : Les différents types de spéciation.

Figure 2 : Les différents types de polyploïde : Les autopolyploïdes et les allopolyploïdes.

Figure 3 : Les chromosomes chez les espèces diploïdes et allopolyploïdes.

Figure 4 : Transmission de la diversité parentale.

Figure 5 : Les différentes voies de formation des allopolyploïdes.

Figure 6 : Les deux types de formation de gamètes non réduits.

Figure 7 : Les grands événements de duplication au cours de l'évolution des espèces.

Figure 8 : Représentation de différents modèles d'étude au cours du temps.

Figure 9 : Notions additivité et de non-additivité.

Figure 10 : Profils d'expression de l'allopolyploïde comparé à ses espèces parentales.

Figure 11 : Profils des sous-génomes parentaux chez les allopolyploïdes.

Figure 12.A : Modèles théoriques sur le devenir des gènes dupliqués d'après Luchetta *et al.*, 2005.

Figure 12.B : Modèles théoriques sur le devenir des gènes dupliqués d'après Hahn *et al.*, 2009.

Figure 13 : Augmentation des possibilités de variations dans l'expression « dose-dépendante » des gènes.

Figure 14 : Augmentation de la complexité des interactions régulatrices.

Figure 15 : Mécanismes génétiques.

Figure 16 : Mécanismes épigénétiques.

Figure 17 : Les mécanismes de régulation.

Figure 18 : Effets des facteurs génétiques et épigénétiques sur le phénotype et l'évolution des polyploïdes.

Figure 19 : Classification des caféiers.

Figure 20 : *Coffea* et *Psilanthus*.

Figure 21 : Dispersion du genre *Coffea* sous genre *Coffea* à partir d'un centre d'origine en basse Guinée.

Figure 22 : Relation phylogénétique des espèces de caféiers basées sur des variations de séquences ADN chloroplastiques et ribosomales.

Figure 23 : Origine de la seule espèce allotétraploïde *Coffea arabica*.

Figure 24 : Origine des deux variétés mères de *Coffea arabica* à l'origine des variétés cultivées actuelles.

Figure 25 : Origine des trois variétés principales actuelles les plus cultivées.

Figure 26 : Diversité observée dans des populations de caféiers sauvages et cultivés chez l'allotétraploïde *Coffea arabica*.

Figure 27 : Croisement à l'origine d'hybrides interspécifiques artificiels triploïdes et tétraploïdes de caféiers.

Figure 28 : Croisement à l'origine des variétés Arabusta et Icatu.

## *Chapitre I*

Figure 1 : Représentation schématique des deux types de puces.

*Figures de l'article "The 'PUCE CAFÉ' project: the first 15K coffee microarray, a new tool for discovering candidate genes correlated to agronomic and quality traits"*

Figure 1: Plots for each Channel of Background Intensity versus Intensity

Figure 2: Density Plot and MA-plot for Both Channels Before and After Normalization.

Figure 3: Quality Analyses between Biological and Technical Hybridization Replicates for Experiment 1 (Bean, Flower and Leaf).

Figure 4: Transcriptome divergence between the three tissues (flower/leaf/mature bean).

Figure 5: Transcriptome divergence between *C. arabica*, *C. canephora* and *C. eugenioides*.  
Figure 6: Venn Diagrams indicated genes that are overexpressed specifically in each tissue (Bean, Flower and Leaf).

## **Chapitre II**

### **PARTIE I**

Figure 1 : Dispositif expérimental.

Figure 2 : Origine des modèles théoriques catégorisant les variations d'expression génique entre un allopolyploïde et ses espèces parentales.

*Figures de l'article "Genomic expression dominance in the natural allopolyploid Coffea arabica is massively affected by growth temperature"*

Fig. 1: Transcriptome divergence and comparison of nonadditive gene expression between *Coffea arabica* cv Java and parental diploid species.

Fig. 2: Venn diagram representing unigenes differentially expressed between the two diploid parents (*Coffea canephora* and *Coffea eugenioides*) at both temperatures.

Fig. 3: Patterns of differential expression between *Coffea arabica* cv Java allopolyploid and parental diploid species (*C. canephora* and *C. eugenioides*) at 26–22°C.

Fig. 4: Patterns of differential expression between *Coffea arabica* cv Java allopolyploid and parental diploid species at 30–26°C.

Fig. 5: Patterns of differential expression between *Coffea arabica* cv T18141 allopolyploid and parental diploid species at 26–22°C.

Fig. 6: Patterns of differential expression between *Coffea arabica* cv T18141 allopolyploid and parental diploid species at 30–26°C.

Fig. 7: The 13 possible expression patterns in two *Coffea arabica* allotetraploids relative to their progenitors (*C. canephora* and *C. eugenioides*).

Fig. 8: Behaviour at 30–26°C of the two majority unigene categories, that is, '*C. canephora*-like dominance' and 'no change'.

### **PARTIE II**

Figure 3 : Analyse hiérarchique descendante basée sur les distances euclidiennes entre les transcriptomes.

Figure 4 : Diagramme de Venn des unigènes sur-exprimés en condition de température de 30-26°C par rapport à la condition de température de 22-26°C pour chacune des espèces.

Figure 5 : Diagramme de Venn des unigènes sous-exprimés en condition de température de 30-26°C par rapport à la condition de température de 22-26°C pour chacune des espèces.

## **Chapitre III**

Figure 1 : Représentation des voies eucaryotique et procaryotique dans la synthèse de galactolipides et phospholipides dans une cellule de plante.

Figure 2 : Influence de la température sur la croissance des trois espèces de caféiers.

Figure 3 : Influence de la température sur la teneur en chlorophylle totale des jeunes feuilles des trois espèces de caféiers.

Figure 4 : Variation des proportions en 16 :0 et 18 :3 chez les trois espèces de caféiers en fonction de la température.

Figure 5 : Corrélation entre les variations observées de profils chez les acides gras et celles observées chez 4 gènes impliqués dans la désaturation des lipides de feuilles chez le caféier.

Figure 6 : *Coffea canephora* à 18-15°C.

## **Chapitre IV**

Figure 1 : Protocole expérimental généralisé de la technique RNAseq.

Figure 2 : Principaux groupes génétiques de *C. canephora* sur la base de marqueurs neutres.

Figure 3 : Assemblage et nettoyage des données de la référence *C. canephora*.

Figure 4 : Nombre de 'reads' et sa médiane sur les 55964 unigènes étudiés par RNAseq.

Figure 5 : Courbes représentant l'ajustement de la variance et la moyenne pour l'ensemble des 55964 unigènes étudiés.

Figure 6 : Clusters issus d'une classification ascendante hiérarchique basée sur les distances entre les transcriptomes (méthode RNAseq).

Figure 7 : Divergence du transcriptome entre *Coffea arabica* cv Caturra et trois groupes génétiques de l'espèce parentale diploïde *Coffea canephora*.

Figure 8 : Diagrammes de Venn des gènes sur-exprimés.

Figure 9 : Représentation schématique des variations du transcriptome du *C. canephora* (groupe Congolais) et de *C. arabica* en fonction de deux régimes thermiques.

### ***Discussion générale et perspectives***

Figure 1 : Scénarios possibles à l'origine de l'adaptation des espèces.

Figure 2 : Scénario sur la formation et la stabilisation de *C. arabica*.

Rapport-Gratuit.com





# Liste des tableaux

## *Chapitre I*

*Tableaux de l'article "The 'PUCE CAFÉ' project: the first 15K coffee microarray, a new tool for discovering candidate genes correlated to agronomic and quality traits"*

Table 1: In Experiment 1, for each comparison (six slides) 16,512 spots were examined.

Table 2: Mean, Median and P90 of the Coefficients of Variation (CV) of the Fluorescence Signal Intensity.

Table 3: Expression Patterns of a few Genes Involved in Well-Characterized Lipid Biosynthetic Pathways.

Table 4: Expression Patterns of a few Genes encoding Potential Storage Proteins.

## *Chapitre II*

### **PARTIE I**

*Tableaux de l'article "Genomic expression dominance in the natural allopolyploid Coffea arabica is massively affected by growth temperature"*

Table 1: Comparison of expressed unigenes relative to the background noise.

Table 2: Fluorescence intensity differential (DFI) and the variation index (IVar) between parental *Coffea canephora* diploid species and *C. arabica* allopolyploids (cv Java and cv T18141).

### **PARTIE II**

Tableau 1 : Matrice des distances euclidiennes.

Tableau 2 : Nombre de gènes sur- ou sous-exprimés à la condition de température 30-26°C au seuil  $P < 10^{-7}$ .

## *Chapitre III*

Tableau 1 : Liste des amorces utilisées lors de la qRT-PCR.

Tableau 2 : Variations de l'abondance relative des principaux acides gras en fonction des conditions de température chez les trois espèces étudiées.

## *Chapitre IV*

Tableau 1 : Matériels biologiques et plan expérimental.

Tableau 2 : Ajustement des données par le logiciel DESeq.

Tableau 3 : Nombre de gènes sur et sous exprimés dans les comparaisons intra groupes et inter-températures.

Tableau 4 : Recherche des 'gènes de la plasticité' chez *Coffea arabica*.



# **INTRODUCTION**

## **GENERALE**



La plasticité phénotypique se traduit par la capacité d'un génotype à produire de multiples phénotypes en réponse à des conditions environnementales variées. De nombreux changements phénotypiques à forte valeur adaptative, concernant la vigueur et la biomasse, la période de floraison, la résistance aux insectes nuisibles ou encore l'adaptation à des stress biotiques ou abiotiques, permettraient aux polyploïdes d'occuper de nouvelles niches écologiques et favoriseraient leur utilisation dans l'agriculture. Il s'agirait souvent d'espèces très vigoureuses avec une capacité à s'adapter à des environnements défavorables aux espèces parentales.

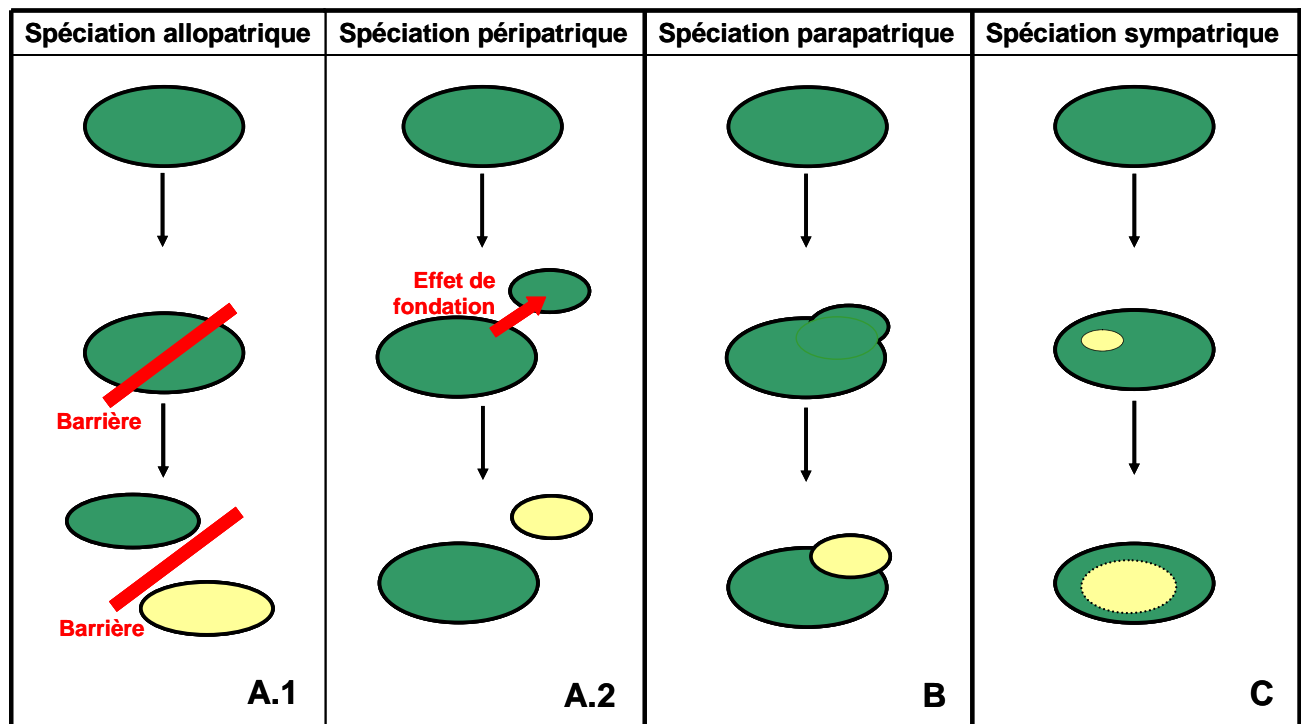
Les espèces polyploïdes sont issues d'une hybridation interspécifique réunissant deux génomes plus ou moins divergents ayant évolué indépendamment et/ou d'une duplication du génome. L'état polyploïde est très fréquent chez les plantes et des marques de polyploïdies anciennes ont été observées chez la plupart des génomes de plantes. La plupart des auteurs s'accorde sur la théorie que l'état polyploïde jouerait un rôle important dans la plasticité phénotypique et les capacités adaptatives de nouvelles espèces. Les génomes allopolyploïdes ne représentent donc pas la simple addition et duplication des génomes parentaux. Dès les premières générations, des changements importants interviennent et modifient la structure et le fonctionnement des génomes. Leur structure peut être modifiée par des éliminations de séquences et des réarrangements chromosomiques ou faire intervenir la méthylation de l'ADN ou des histones, l'acétylation des histones ou encore les petits ARNs. Des changements dans le fonctionnement des génomes se traduisent par des modifications de l'expression transcriptomique et/ou protéomique des génomes. Ces dernières années, les variations d'expression génique chez les allopolyploïdes ont fait l'objet de nombreuses études. Au niveau transcriptomique, l'expression des gènes peut être observée sous deux angles, soit en étudiant de façon globale la divergence d'expression entre l'allopolyploïde et ses espèces parentales soit en mesurant le ratio des transcrits dérivant de chaque gène homéologue. Pour observer ces variations, différentes techniques ont été développées telles que la PCR quantitative en temps réel (qRT-PCR) permettant une analyse de l'expression pour un seul gène à la fois, les Puces à ADN permettant une mesure simultanée des niveaux d'expression de milliers de gènes et plus récemment le séquençage à haut débit de fragments d'ARN (RNAseq) permettant une analyse complète du génome.

Le café est un produit de consommation international se rangeant parmi les exportations agricoles les plus importantes des pays en voie de développement. Sa production, son traitement et son commerce sont à l'origine de dizaine de millions d'emplois dans le monde ce qui en fait un enjeu économique majeur. La production mondiale repose principalement sur deux espèces, le café Arabica (*Coffea arabica*) et le café Robusta (*Coffea canephora*) représentant respectivement 65% et 35% de la production mondiale. La culture

des caféiers est répandue dans toute la ceinture intertropicale. *Coffea arabica*, la seule espèce polyploïde du genre *Coffea*, est issu d'une hybridation interspécifique et d'une duplication du génome entre deux espèces parentales africaines proches génétiquement, *C. eugenioides* et *C. canephora*. *Coffea arabica* est une espèce relativement récente (10-50000 ans) originaire d'une région limitée du Sud-ouest de l'Ethiopie.

Au cours de cette thèse, nous proposons d'étudier les variations du transcriptome chez l'allopolyploïde, *C. arabica*, et ses deux espèces parentales. Sachant que les trois espèces possèdent une écologie différentes, *C. arabica* est un modèle d'étude très intéressant pour tenter d'observer la plasticité d'une espèce et sa dépendance aux conditions environnementales pouvant entraîner une capacité adaptative de celle-ci à des milieux considérés défavorables pour ses espèces parentales.

**REVUE**  
**BIBLIOGRAPHIQUE**



**Figure 1 : Les différents types de spéciation.**

Cette différenciation entre les types de spéciation est fonction de la géographie. Le cercle vert représente la population d'origine et le cercle jaune, la population nouvellement formée.

Il existe trois types de spéciation. **A1.** La *spéciation allopatrique* qui résulte de l'évolution divergente de populations géographiquement isolées les unes des autres. Barrière : causes géographiques, écologiques, comportementales et/ou génétiques. **A2.** La *spéciation péripatrique* est un cas particulier de la spéciation allopatrique qui résulte de l'évolution divergente de populations isolées de façon périphérique. **B.** La *spéciation parapatrique* résulte de l'évolution divergente de populations qui sont géographiquement adjacentes les unes aux autres. **C.** La *spéciation sympatrique* résulte de l'évolution divergente de populations issues d'une même zone géographique.



# I. POLYPLOÏDIE : MODIFICATIONS DES GENOMES, DU TRANSCRIPTOME ET DU PHENOTYPE

## I.1 GENERALITES SUR LA POLYPLOÏDIE ET LES POLYPLOÏDES

### I.1.1 Notion de spéciation

#### I.1.1.1 Définition

La spéciation a un rôle majeur dans la diversification biologique (Rieseberg et Willis, 2007). Elle peut être définie par un isolement reproductif ayant des causes géographiques, écologiques, comportementales et/ou génétiques, par un arrêt d'échanges génétiques entre des populations qui deviendront des espèces différentes et par l'accumulation de différences génétiques et phénotypiques (Luchetta *et al.*, 2005). Ces trois aspects sont liés et peuvent se dérouler de manière simultanée ou successive (ordre variable). Il existe trois types de spéciation<sup>1</sup> :

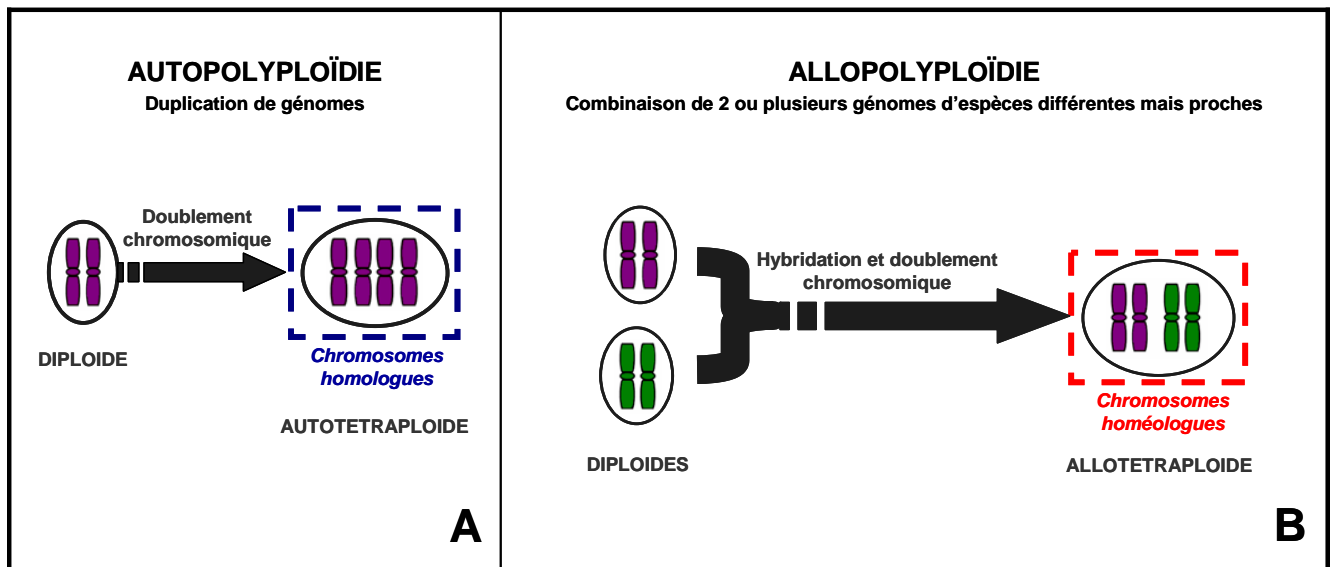
- La spéciation allopatrique est issue d'une divergence progressive de populations complètement isolées les unes des autres, entre lesquelles les échanges géniques ne peuvent avoir lieu (Figure 1.A.1). Au cours de la spéciation allopatrique, une population peut également se former par effet de fondation<sup>2</sup> (spéciation péripatrique ; Figure 1.A.2).
- La spéciation parapatrique est issue d'une divergence progressive de populations géographiquement adjacentes, entre lesquelles les échanges géniques sont possibles (Figure 1.B).
- La spéciation sympatrique est issue d'une divergence de populations entièrement chevauchantes originaires d'une même zone géographique entre lesquelles les échanges génétiques s'arrêtent (Figure 1.C).

#### I.1.1.2 Les spéciations homoploïde et polyploïde

La formation de nouvelles espèces peut se produire à la suite d'une hybridation interspécifique réunissant deux génomes plus ou moins divergents ayant évolué indépendamment (spéciation dite « homoploïde ») et/ou d'une duplication du génome (spéciation dite « polyploïde »). La spéciation homoploïde est à l'origine de nouvelles lignées hybrides n'ayant pas subi de changements de niveau de ploïdie (Buerkle *et al.*, 2000) alors que la spéciation polyploïde conduit à la duplication du génome entier de l'individu

<sup>1</sup> Définition des différents types de spéciation en fonction de la géographie.

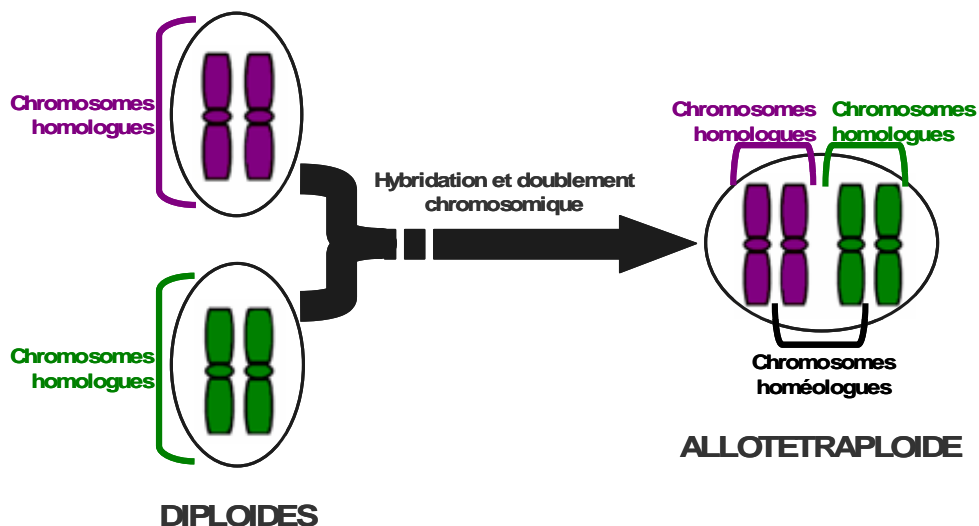
<sup>2</sup> Spéciation dite « péripatrique ». La population initiale ne se divise pas de façon équivalente en deux groupes d'individus mais en un petit groupe d'individus qui sera isolé soit par la mise en place d'une barrière géographique, soit par la migration de ce dernier.



**Figure 2 : Les différents types de polyplœide : Les autopolyploïdes et les allopolyploïdes.**

Pour plus de simplicité une seule paire de chromosomes est représentée (en mauve et en vert).

**A.** Les *autopolyploïdes* sont formés par la duplication de génomes. Dans cette figure est représentée la formation d'un autotétraploïde par doublement chromosomique d'une espèce diploïde. L'autotétraploïde est formé de chromosomes homologues, c'est-à-dire identiques entre eux. **B.** Les *allopolyploïdes* sont formés par la combinaison de deux ou plusieurs génomes d'espèces différentes mais proches. Dans cette figure est représentée la formation d'un allotétraploïde par hybridation et doublement chromosomique entre deux espèces diploïdes différentes mais proches. L'allopolyploïde est formé de chromosomes homologues (mauve ou vert) et homéologues (mauve et vert).



**Figure 3 : Les chromosomes chez les espèces diploïdes et allopolyploïdes.**

Pour plus de simplicité une seule paire de chromosomes est représentée (en mauve et en vert).

Les espèces diploïdes sont constituées de chromosomes *homologues*. Après hybridation et doublement chromosomique, ces chromosomes se retrouvent dans un même individu deviennent des chromosomes *homéologues*.

(Rieseberg et Willis, 2007). La spéciation polyploïde conduit soit à la formation d'espèces allopolyploïdes issues d'une hybridation interspécifique et d'une duplication du génome soit à la formation d'espèces autopolyploïdes issues uniquement d'une duplication du génome. Les hybrides interspécifiques ont une stérilité plus ou moins importante due à une divergence des génomes parentaux induisant des incompatibilités génomiques et des problèmes d'appariement chromosomique à l'origine de méioses irrégulières. Il a également été montré que le « choc génomique<sup>3</sup> » issue d'une hybridation interspécifique réunissant dans un même individu des réseaux hiérarchiques de régulation divergents (Riddle et Birchler, 2003) pouvait entraîner une dérégulation de l'expression des gènes et conduire ainsi à une stérilité des hybrides (Renaut *et al.*, 2009). Chez les espèces allopolyploïdes, la duplication du génome restaure une certaine fertilité, après une ou plusieurs générations.

### **I.1.2 Le phénomène de polyploïdisation**

Le terme de polyploïdie a été introduit, pour la première fois, par Winkler en 1916 et l'hypothèse de son importance dans l'évolution des angiospermes a été émise par Winger en 1917 (Chen et Ni, 2006). La polyploïdisation est un cas particulier de duplication au cours duquel le génome entier se duplique (whole-genome duplication, WGD). Elle peut être observée à différentes échelles, c'est à dire aussi bien pour une population, une espèce, un individu que pour un organe, un tissu ou encore une cellule.

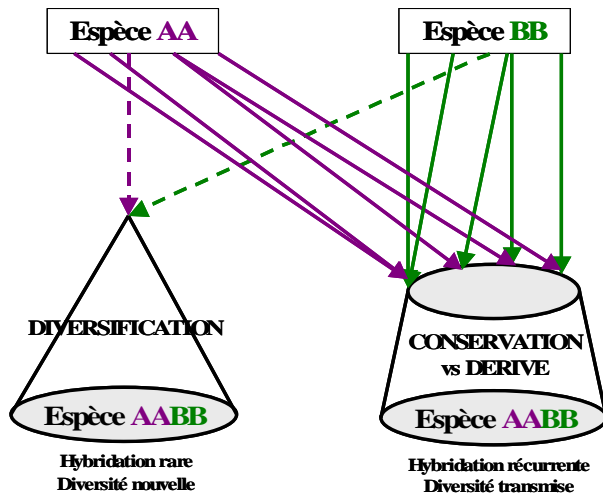
#### **I.1.2.1 Les différents types de polyploïdie**

En 1926, Kihara et Ohno décrivent, pour la première fois, les concepts d'auto- et d'allopolyploïdie. D'après Grant en 1971, on peut diviser les polyploïdes en différents « types » : les autopolyploïdes, les polyploïdes segmentaires, les polyploïdes disomiques ou amphidiploïdes et enfin les auto-allopolyploïdes. Par la suite, cette division a été simplifiée à deux « types » de polyploïdie (Stebbins, 1971 ; Osborn *et al.*, 2003 ; Chen, 2007) :

- Les autopolyploïdes, issus de la multiplication d'un seul génome ou de différents génomes d'une même espèce, possèdent des chromosomes dits homologues (schéma X.A). Ces chromosomes homologues, se retrouvant dans un même individu, sont identiques (Figure 2.A).
- Les allopolyploïdes, issus de la combinaison de deux ou plusieurs génomes d'espèces différentes, possèdent des chromosomes dits homéologues (Figure 2.B). Chez ces chromosomes homéologues (Figure 3), les gènes ne sont pas toujours conservés et leur ordre peut être différent.

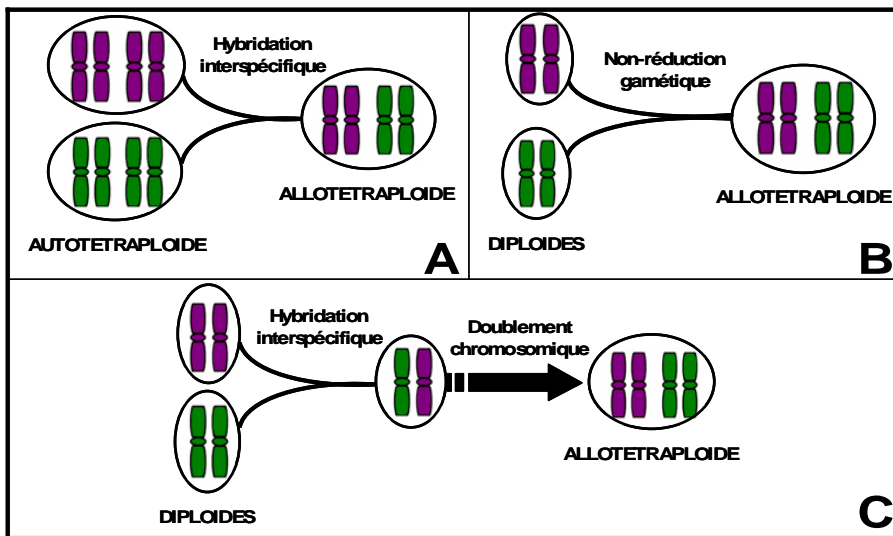
---

<sup>3</sup> Définition de MacClintock en 1984 exprimant l'évènement dû à la réunion de deux génomes plus ou moins divergents ayant évolué indépendamment.



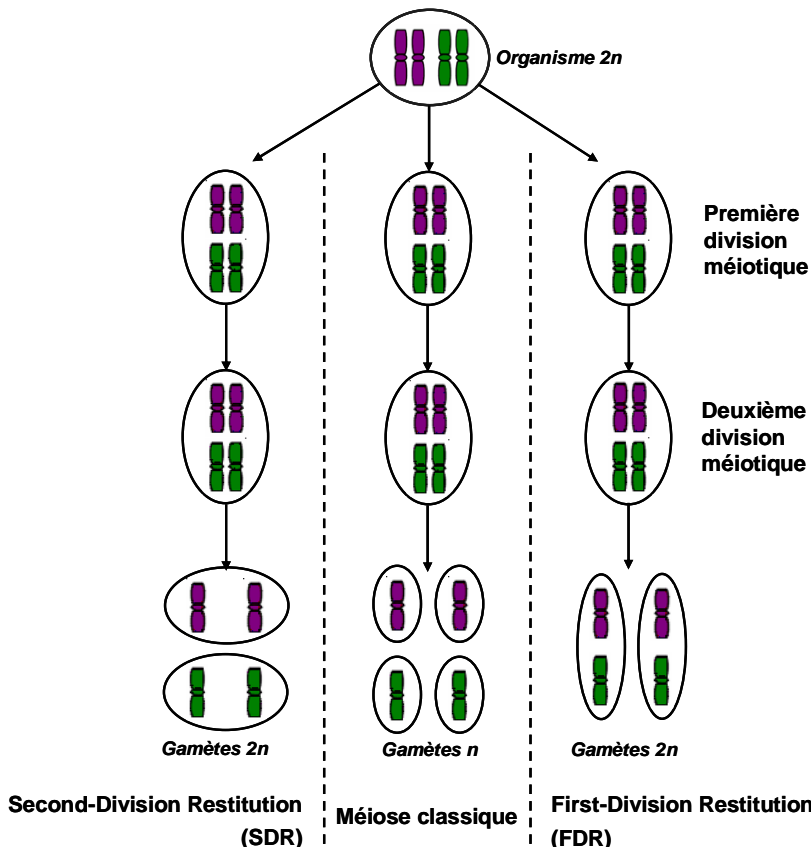
**Figure 4 : Transmission de la diversité parentale.**

Différentes hypothèses sur l'origine de la diversité des polyploïdes. Soit ces espèces sont issues d'une origine unique, soit elles issues d'une origine multiple.



**Figure 5 : Les différentes voies de formation des allopolyploïdes.**

Pour plus de simplicité une seule paire de chromosomes est représentée (en mauve et en vert).  
**A.** Formation d'un allotétraploïde par *hybridation interspécifique* entre deux autotétraploïdes.  
**B.** Formation d'un allopolyploïde par *non-réduction gamétique* des deux espèces diploïdes.  
**C.** Formation d'un allopolyploïde par *hybridation interspécifique* suivie d'un *doublement chromosomique* entre deux espèces diploïdes.



**Figure 6 : Les deux types de formation de gamètes non réduits.**

Pour plus de simplicité une seule paire de chromosomes est représentée (en mauve et en vert).

**SDR.** Les gamètes de type SDR possèdent des paires de chromatides sœurs. **FDR.** Les gamètes de type FDR possèdent une chromatide de chaque chromosome présent chez l'organisme 2n.

*D'après Bretagnolle et Thompson, 1995*

### I.1.2.2 La formation des polyploïdies

Lors de la formation des polyploïdes, la transmission de la diversité des espèces parentales peut avoir un rôle majeur de leur diversité. Certaines espèces ont une origine multiple et sont issues de croisements réciproques<sup>4</sup> tels que *Tragopogon* (Figure 4). D'autres espèces ont une origine associée à peu de croisements tel que *Spartina* voire unique tel que *Arabidopsis suecica* (Chen et Ni, 2006 ; Figure 4). La diversité génétique des espèces ayant une origine unique serait plus faible que celles des espèces ayant une origine multiple favorisant l'introduction d'une plus grande part de la diversité des espèces parentales. Cette variabilité initiale peut être altérée par des pressions de sélection anthropiques et naturelles. Cependant, les mécanismes concernant la régulation de l'expression des gènes dupliqués chez les allopolyploïdes peuvent influencer et générer de la diversité aussi bien génétique que morphologique.

Le passage de l'état diploïde à l'état allopolyploïde peut se faire soit par une hybridation interspécifique entre autopolyploïdes<sup>5</sup> (Figure 5.A), soit par la fusion de gamètes non-réduites (Figure 5.B), soit par d'un doublement chromosomique<sup>6</sup> (Figure 5.C), (Chen, 2007). La fusion de gamètes non-réduits serait le modèle prédominant<sup>7</sup> du passage de l'état diploïde à l'état polyploïde. Lors de la méiose, un organisme à  $2n$  chromosomes forme des cellules à  $n$  chromosomes. Pour obtenir ces cellules, une étape de réplication de l'ADN est suivie de deux divisions successives, réductionnelle<sup>8</sup> et équationnelle<sup>9</sup>. Cette méiose peut être perturbée et conduire à la formation de gamètes non-réduits ( $2n$  chromosomes), (Bretagnolle et Thompson, 1995). Deux types de gamètes non-réduits peuvent être observés, les « First-Division Restitution » et les « Second-Division Restitution » (Figure 6). Les gamètes de type FDR possèdent une chromatide (recombinée ou non) de chaque chromosome présent chez l'organisme  $2n$  (Figure 6). Les gamètes  $2n$  obtenus ne présentent pas de traces de la première division méiotique. A contrario, les gamètes de type SDR possèdent des paires de chromatides sœurs (Figure 6). Les gamètes obtenus ne présentent pas de traces de la seconde division méiotique.

---

<sup>4</sup> Populations avec les deux cytoplasmes parentaux possibles.

<sup>5</sup> Hybridation entre deux espèces autopolyploïdes proches mais différentes.

<sup>6</sup> Hybridation entre deux espèces proches mais différentes suivie d'un doublement du jeu de chromosomes de l'hybride.

<sup>7</sup> Cette voie est considérée comme le modèle prédominant sachant que de nombreuses espèces végétales présentent des anomalies de la méiose.

<sup>8</sup> La première division méiotique, ou division réductionnelle, conduit à la séparation des chromosomes homologues.

<sup>9</sup> La seconde division méiotique, ou division équationnelle, conduit à la séparation des chromatides sœurs.

### I.1.2.3 Les différents comportements méiotiques chez les polyploïdes

Le phénomène de polyploïdisation a un impact majeur sur la méiose. En général, les polyploïdes pairs ( $4n$ ,  $6n$  ...) sont dits « stables », c'est à dire fertiles et les polyploïdes impairs ( $3n$ ,  $5n$ ...) sont plus ou moins stériles. Cette stérilité est due à la difficulté de répartir de manière équitable, au cours de la méiose, les chromosomes entre les deux cellules filles (Luchetta *et al.*, 2005). En fonction de caractéristiques génétiques et cytogénétiques, des comportements méiotiques différents sont observés :

- Chez les polyploïdes polysomiques, des multivalents sont formés à la suite d'appariements entre plus de 2 chromosomes homologues. Les autopolyploïdes sont souvent polysomiques (Butruille et Boiteux, 2000). Toutefois, il faut noter que les polyploïdes polysomiques ne conduisent pas nécessairement à la formation de multivalents comme par exemple chez la Canne à sucre (Al-Janabi *et al.*, 1993)
- Chez les polyploïdes disomiques, seuls des bivalents sont formés à la suite d'appariements se faisant uniquement entre chromosomes homologues (Chen, 2007). En générale, les allopolyploïdes sont considérés comme disomiques. Toutefois, il faut différencier les amphidiploïdes pour lesquels la formation de bivalents est stricte et les allopolyploïdes segmentaires pour lesquels la formation de multivalents est possible. Si les chromosomes homéologues possèdent des segments homologues, l'appariement peut se faire entre les chromosomes homéologues (Chen, 2007).

### I.1.3 Les espèces polyploïdes chez les êtres vivants

Depuis longtemps, la polyploïdie a été reconnue comme un mécanisme de l'évolution des eucaryotes (Osborn *et al.*, 2003). Elle est présente chez les êtres vivants à différentes fréquences et sous des formes d'hybrides inter- et intra-spécifiques (Rieseberg *et al.*, 2003 ; Mavárez *et al.*, 2006 ; Chen, 2007). La polyploïdie a été observée chez les champignons, au cours d'études sur des levures (*Saccharomyces cerevisiae* ; Albertin *et al.*, 2009), ou desoomycètes (*Phytophthora sp.* ; Martens et De Peer, 2010), les végétaux et les animaux.

#### I.1.3.1 Fréquence et récurrence de la polyploïdie chez les végétaux

La polyploïdie est un mécanisme majeur d'adaptation et de spéciation chez les végétaux (Ramsey et Schemske, 1998, Otto et Whitton, 2000 ; Adams, 2007). Une étude récente<sup>10</sup> montre que, pour 15% des angiospermes et 31% des ptéridophytes, les évènements de spéciation seraient accompagnés d'une augmentation du niveau de ploïdie (Wood *et al.*,

<sup>10</sup> Etude de Wood *et al.* en 2009 sur la fréquence de la polyploïdie chez les plantes vasculaires. Les auteurs ont estimé la fréquence de la polyploïdie en repérant les changements dans le niveau de ploïdie au travers d'arbres phylogénétiques de plantes créés à partir de leur nom infragénérique (rang inférieur au genre).

2009). La plupart des plantes cultivées ou encore la quasi-totalité des plantes ornementales sont polyploïdes (Luchetta *et al.*, 2005). Des formes performantes de polyploïdes sont cultivées comme le blé, le coton, la banane, la canne à sucre ou encore le café, et de nombreuses études ont été menées sur ces diverses plantes. La polyploïdisation peut être très récente. C'est le cas du *Tragopogon*, où deux espèces allopolyploïdes (*T. miscellus* et *T. mirus*) se sont formées naturellement il y a moins de 80 ans. Il a été observé qu'un nouveau *Tragopogon* allotétraploïde apparaîtrait tous les ans. (Tate *et al.*, 2006 ; Chen, 2007). Des marques de polyploïdies plus anciennes sont fréquemment observées chez les génomes des plantes. Ces paléopolyploïdes ont subi au moins un évènement de duplication du génome entier (WGD) suivi d'une perte massive de gènes et d'une réorganisation génomique par diploïdisation (Wolfe, 2001 ; Chen, 2007). Parmi eux on trouve *Arabidopsis*, *Gossypium* ou encore *Zea mays* (Chen, 2007). Chez les angiospermes, Cui *et al.* (2006) ont montré que des traces de paléopolyploïdie étaient présentes dans presque toutes les espèces étudiées sauf *Amborella trichopoda*<sup>11</sup>. Pour nombre d'entre elles, plusieurs évènements de polyploïdie se sont superposés au cours du temps évolutif ce qui confirme le caractère récurrent de la polyploïdie dans l'évolution des espèces.

### I.1.3.2 Fréquence et récurrence de la polyploïdie chez les animaux

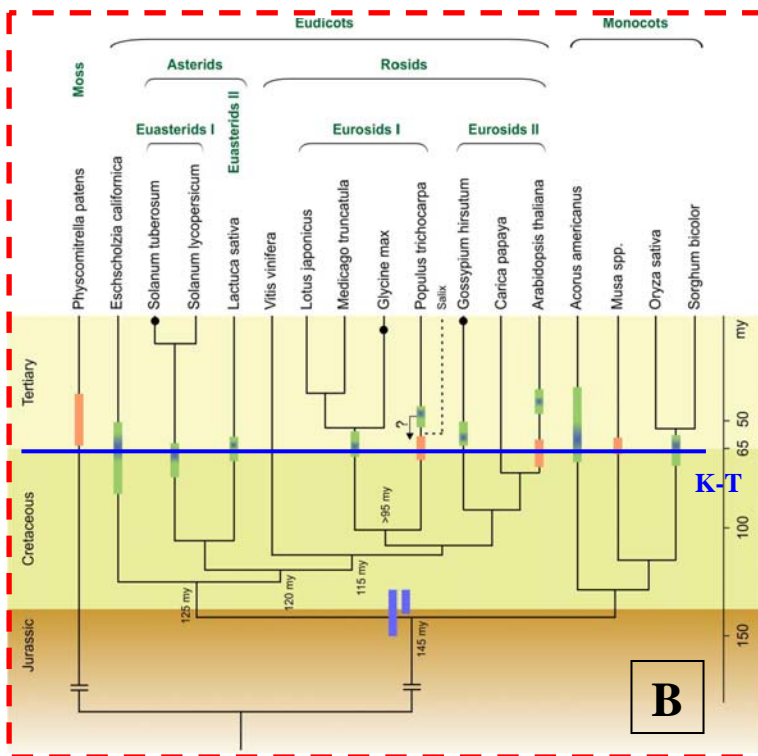
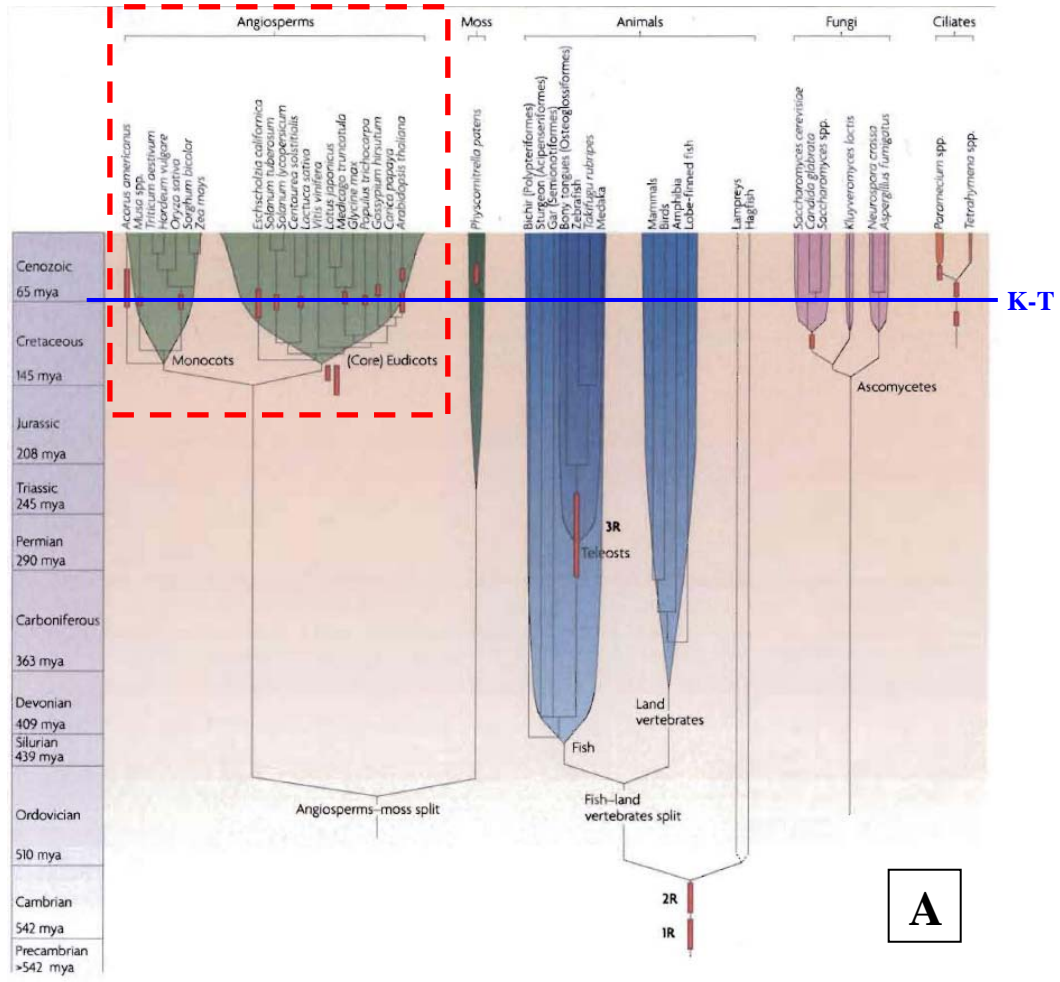
Du fait de l'impact de la polyploïdisation sur la détermination sexuelle et le développement de l'animal, la polyploïdie est plus rare chez les animaux que chez les plantes (Muller, 1925 ; Mable, 2004 ; Chen, 2006 et 2007). Malgré la faible fréquence de cet état chez les espèces animales comparées aux espèces végétales, on constate tout de même que la polyploïdie se retrouve chez des espèces animales très différentes les unes des autres (amphibiens, reptiles, insectes, crustacés, poissons et divers vertébrés ; Otto *et al.*, 2007). On peut citer comme exemple l'espèce de mammifère pour laquelle un évènement de polyploïdisation a été observé, les rats-viscaches<sup>12</sup> (Gallardo *et al.*, 1999 et 2004).

### I.1.4 Rôle de la polyploïdie dans l'adaptation et l'évolution des espèces

Les polyploïdes seraient à l'origine de phénotypes nouveaux, d'une diversification écologique et de conquêtes de nouvelles niches écologiques (Otto et Whitton, 2000 ; Adams, 2007). On suppose que l'état d'hétérozygotie fixé leur permettrait l'obtention d'un large panel de nouveaux allèles intervenant au cours de la sélection, de la mutation et de l'évolution génétique (Adams, 2007).

<sup>11</sup> Le genre *Amborella* est considéré comme le taxon le plus basal des angiospermes. Les *Amborella* seraient les uniques espèces sœurs de tous les angiospermes (Kim *et al.*, 2004 ; Soltis DE et Soltis PS, 2004).

<sup>12</sup> Il existe deux espèces sœurs, *Tympanoctomys barrerae* (2n=4x=102) et *Pipanaoctomys aureus* (2n=4x=92). *Tympanoctomys barrerae* est l'espèce mammifère au plus grand nombre de chromosomes. L'évènement d'allopolyploïdisation est relativement récent et aurait eu lieu il y a 6.5 Mya (Gallardo et Kirsh, 2001).



**Figure 7 : Les grands évènements de duplication au cours de l'évolution des espèces.**

**A.** Cinq grands groupes d'espèces sont représentés : les Angiospermes et les Mousses pour les Végétaux, les Animaux, les Champignons et les Ciliés. Cette représentation de Van de Peer *et al.* est fonction de l'échelle des temps (du Précambrien au Cénozoïque). **B.** Fawcett *et al.* (2009) représentent les évènements de duplication chez les Angiospermes observés au cours de l'évolution des espèces. Les auteurs remarquent que les différents évènements se sont produits au cours de la crise K-T.

Fawcett *et al.*, 2009  
 Van de Peer *et al.*, 2009



### I.1.4.1 *Diversification et évolution des espèces*

La polyploïdie serait la clef de la diversification des espèces végétales (Soltis, 2009) et de leur succès évolutif à long terme (Van de Peer *et al.*, 2009). D'après Fawcett *et al.* (2009), la majorité des WGD se regroupe au cours d'une unique période coïncidant à la crise crétacé-tertiaire<sup>13</sup> (KT ; Figure 7.B). Avec l'idée que les espèces polyploïdes auraient une adaptabilité supérieure et une tolérance accrue à des conditions environnementales variables, les auteurs ont émis l'hypothèse que la polyploïdie aurait contribué à la survie et à la diffusion des espèces de plantes pendant ou à la suite d'évènements d'extinction produits à la crise KT. Mais il faut tout de même nuancer le lien qu'il peut y avoir entre les extinctions massives d'espèces et les évènements de WGD car pour certains paléopolyploïdes tels que des champignons (*Saccharomyces cerevisiae*) ainsi que pour l'apparition des Eudicotylédones supérieurs<sup>14</sup> (core eudicotyledons), il n'y a aucune indication qui permette de relier ces deux évènements (Van de Peer *et al.*, 2009 ; Figure 7.A). Selon ces auteurs, il est également difficile de définir le rôle de la polyploïdie dans la diversification et l'évolution des espèces et de savoir si la polyploïdie a un rôle dans la survie des organismes lors des extinctions massives ou si elle facilite les transitions évolutives et la complexification biologique<sup>15</sup>.

### I.1.4.2 *Adaptation et plasticité phénotypique*

Dans des environnements adaptés aux parents diploïdes, les polyploïdes et diploïdes se développent et cohabitent alors que dans des environnements non adaptés aux parents diploïdes tels que les hautes altitudes et latitudes où les températures sont plus froides, les polyploïdes survivent mieux que leur parents diploïdes (Chen, 2007). L'adaptation à de nouvelles niches écologiques peut être illustrée par le cas des plantes invasives et des plantes arctiques. Comme le soulignent Prentis *et al.* (2008), la majorité des plantes invasives sont polyploïdes. Cette hypothèse peut s'expliquer par des différences de fitness ou de compétition entre des populations ayant différents niveaux de ploïdie et par une meilleure colonisation des polyploïdes dans des milieux extrêmes comme cela a été observé dans la flore arctique (Prentis *et al.*, 2008). Brochmann *et al.* (2003) soulignent que le succès de la colonisation des polyploïdes dans la zone arctique serait basé sur l'état d'hétérozygotie fixé de leur génome.

<sup>13</sup> La crise KT, survenue il y a environ 65 millions d'années, est la plus récente des extinctions massives d'espèces. L'origine de cette crise n'est pas encore clairement établie, un ou plusieurs évènements pourraient en être la cause comme l'impact d'astéroïdes ou une activité volcanique intense.

<sup>14</sup> Les Eudicotylédones supérieurs ou noyau des dicotylédones vraies appartiennent aux angiospermes. Ils forment un groupe monophylétique comprenant de grands clades comme les Rosidées et les Astéridées.

<sup>15</sup> Exemple de l'interaction angiospermes et insectes. Cette interaction aurait joué un rôle majeur dans la diversification des niches écologiques chez les angiospermes. Les angiospermes ont développé des innovations facilitant grandement la pollinisation par les insectes telles que l'association des organes reproducteurs mâles et femelles sur le même axe ou encore le développement de périanthe très coloré. Les phénomènes de polyploïdisation des angiospermes seraient associés à l'explosion de la pollinisation par les insectes à la même période qui auraient permis aux plantes de conquérir de nouvelles niches écologiques (Van de Peer *et al.*, 2009).

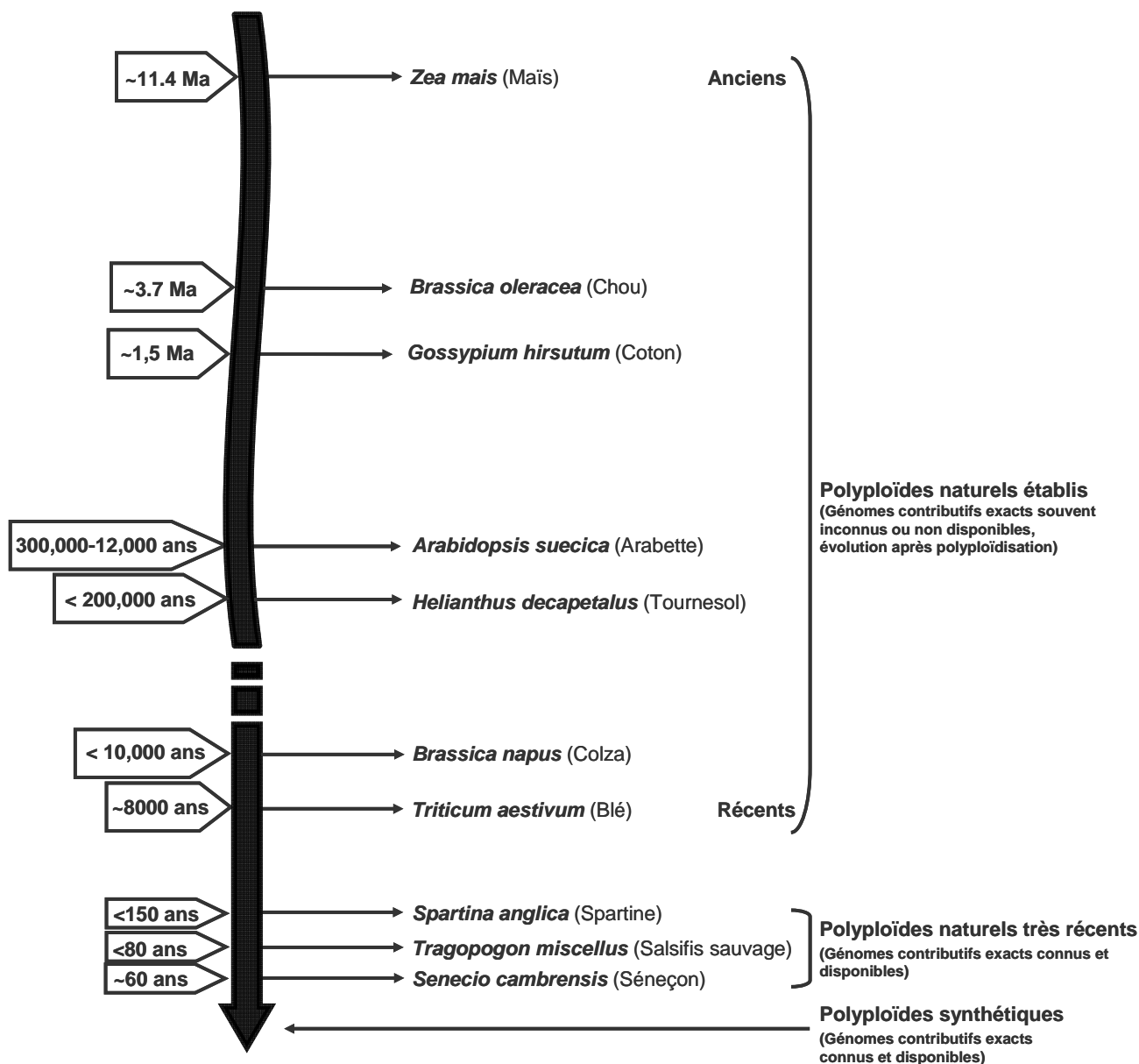
Cet état permettrait une meilleure survie des espèces pendant des périodes climatiques instables.

La plupart des auteurs s'accorde sur la théorie que l'état polyploïde jouerait un rôle important dans la plasticité phénotypique et les capacités adaptatives des nouvelles espèces. La plasticité phénotypique se traduit par la capacité d'un génotype à produire de multiples phénotypes en réponse à des conditions environnementales variées (Pfenning, 2010). L'acquisition de certains caractères par les polyploïdes, comme la tolérance aux fortes sécheresses, la production de graines asexuées, la résistance aux insectes nuisibles, le temps de floraison, la taille des organes et la biomasse, permettrait à ceux-ci d'occuper de nouvelles niches écologiques et favoriseraient leur utilisation dans l'agriculture. (Osborn *et al.*, 2003 ; Chen, 2007). Il a été observé chez plusieurs espèces polyploïdes une vigueur et une biomasse plus importantes que chez les espèces parentales diploïdes. Ni *et al.* (2009) soulignent que, dans des conditions environnementales contrôlées, il y a un accroissement de la vigueur assimilable à de la vigueur hybride<sup>16</sup> chez un allotétraploïde néosynthétique d'*Arabidopsis* (*A. thaliana* x *A. arenosa*) par rapport à ses deux parents. Dans cette étude, les auteurs montrent que l'allopolyploïde présenterait un meilleur contrôle du rythme circadien jouant un rôle dans la croissance et permettant un rôle la synthèse accrue de chlorophylle et d'amidon. Anssour *et al.* (2009) soulignent également, qu'au niveau cellulaire, des lignées allopolyploïdes de *Nicotiana* (*N* x *obtusata*) ont des phénotypes intermédiaires aux parents. Dans cette étude, les auteurs montrent que les auto- et allopolyploïdes naturels de *Nicotiana* subirait des changements morphologiques, génétiques et génomiques. Des études ont été menées sur d'autres espèces telles que *Brassica* (Lukens *et al.*, 2003) et *Triticum* (Takumi *et al.*, 2009) pour lesquelles des changements morphologiques et des variations dans la période de floraison ont été constatés. Chez *Spartina* (Ainouche *et al.*, 2004 et 2009 ; Maricle *et al.*, 2006), des différenciations morphologiques observées relèveraient d'une très grande plasticité phénotypique de l'allopolyploïde naturel (*Spartina anglica*) se traduisant par sa très large amplitude écologique et sa tolérance aux conditions réductrices et aux sédiments riches en sulfites.

---

<sup>16</sup> La vigueur hybride ou hétérosis est définie par une meilleure croissance et survie des hybrides comparés aux espèces parentales (Chen, 2010).

**Rapport-Gratuit.com**



**Figure 8 : Représentation de différents modèles d'étude au cours du temps.**

Sur cette figure sont cités des exemples de polyplœides naturels établis anciens ou récents et de polyplœides très récents (moins de 200 ans) souvent utilisés comme modèles d'étude de la polyplœidie.

## I.2 RÔLE DE LA POLYPLOÏDIE DANS L'EXPRESSION GÉNIQUE

Les phénotypes nouveaux<sup>17</sup> observés chez les polyploïdes suggèrent des variations dans l'expression des gènes chez les espèces polyploïdes. Les conséquences de la polyplœdie sur l'expression génique des polyploïdes ont été étudiées en fonction de l'espèce et du type de polyplœdie (auto- et allopolyploïde). Dans les différentes études traitant de l'expression génique, les polyploïdes sont divisés selon trois catégories (Figure 8) :

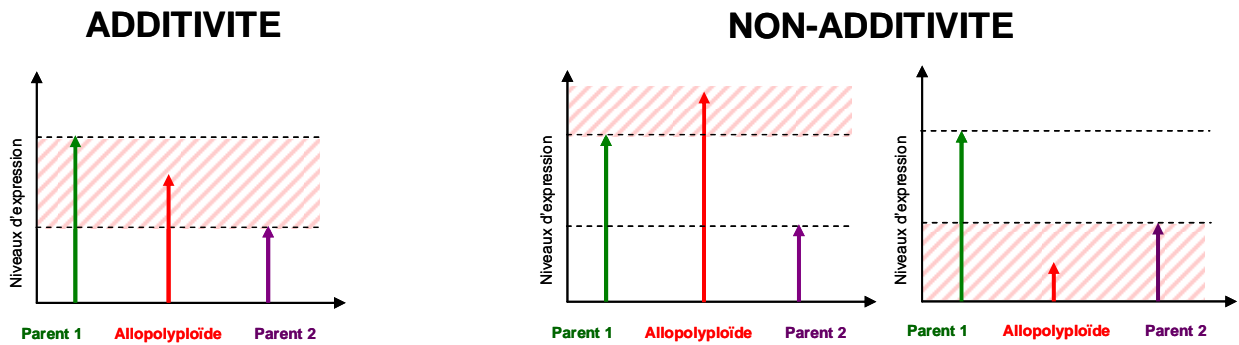
- Les polyploïdes naturels établis pour lesquels les génomes contributifs sont souvent inconnus ou non disponibles dû à l'évolution des espèces après le phénomène de polyplœdisation. Chez ces polyploïdes, on retrouve les espèces telles que *Zea mais*,
- *Gossypium hirsutum*, *Arabidopsis suecica*, *Helianthus decapetalus*, *Brassica napus*, *Brassica oleracea* ou encore *Triticum aestivum* (Figure 8). Ces polyploïdes peuvent être considérés comme anciens ou plus ou moins récents.
- Les polyploïdes naturels très récents pour lesquels les génomes contributifs sont connus et disponibles. Chez ces polyploïdes, on retrouve les espèces telles que *Spartina anglica*, *Tragopogon miscellus* ou encore *Senecio cambrensis* (Figure 8).

Les polyploïdes synthétiques pour lesquels les génomes contributifs sont connus et disponibles. Ce sont des formes analogues aux polyploïdes naturels qui sont formées par hybridation et doublement à la colchicine (Blakeslee et Avery, 1937 ; Figure 8). Comme les espèces naturelles récentes, les allopolyploïdes synthétiques sont considérés comme de bons modèles d'étude car ils permettraient des comparaisons exactes avec les parents diploïdes (Osborn *et al.*, 2003). Cependant, les hybrides ont une stérilité souvent très faible due au « choc génomique » qui entraîne des dérégulations de l'expression génique (voire paragraphe I.1.1.2).

L'analyse du transcriptome permet d'estimer le rôle de la polyplœdie dans l'expression des gènes. Cette analyse peut se faire soit en comparant le niveau d'expression global entre les allopolyploïdes et leurs parents soit en déterminant la contribution de chaque sous-génome homéologue au transcriptome de l'allopolyploïde.

---

<sup>17</sup> Phénotypes absents ou supérieurs aux deux parents.



**Figure 9 : Notions d'additivité et de non-additivité.**

Les espèces parentales sont représentées en vert et mauve et l'espèce allopolyploïde en rouge. Les zones rouges hachurées représentent les zones d'expression de l'allopolyploïde en situation d'additivité ou de non-additivité par rapport aux espèces parentales.

EXPRESSION GENIQUE		SUR-REGULATION	SOUS-REGULATION	
<p><b>'DOMINANCE'</b></p> <p>Le niveau d'expression de l'allopolyploïde est proche de celui d'un des deux parents</p>	Parent 1			
	Parent 2			
<p><b>'TRANSGRESSION'</b></p> <p>Le niveau d'expression de l'allopolyploïde est supérieur ou inférieur au niveau d'expression des deux parents</p>	Expression parentale non différentielle			
	Expression parentale différentielle	Parent 1		
		Parent 2		
<p><b>'ADDITIVITE'</b></p> <p>Le niveau d'expression de l'allopolyploïde se situe entre les niveaux d'expression des deux parents</p>				
<p><b>'NO CHANGE'</b></p> <p>Le niveau d'expression de l'allopolyploïde est égal au niveau d'expression des deux parents</p>				

**Figure 10 : Profils d'expression de l'allopolyploïde comparé à ses espèces parentales.**

P1 et P2 représentent les espèces parentales et A représente l'espèce allopolyploïde. Les quatre catégories sont représentées : 'dominance', 'transgression', 'additivité' et 'No change'. Pour les catégories de 'dominance' et de 'transgression', des subdivisions sont faites en fonction de la dominance vers l'un des deux parents et/ou en fonction de la sur-expression ou sous-expression du gène. Les catégories de 'dominance' et de 'transgression' représentent les gènes à expression non-additive.

### I.2.1. Etudes du transcriptome

#### I.2.2.1 Notions d'additivité et de non-additivité

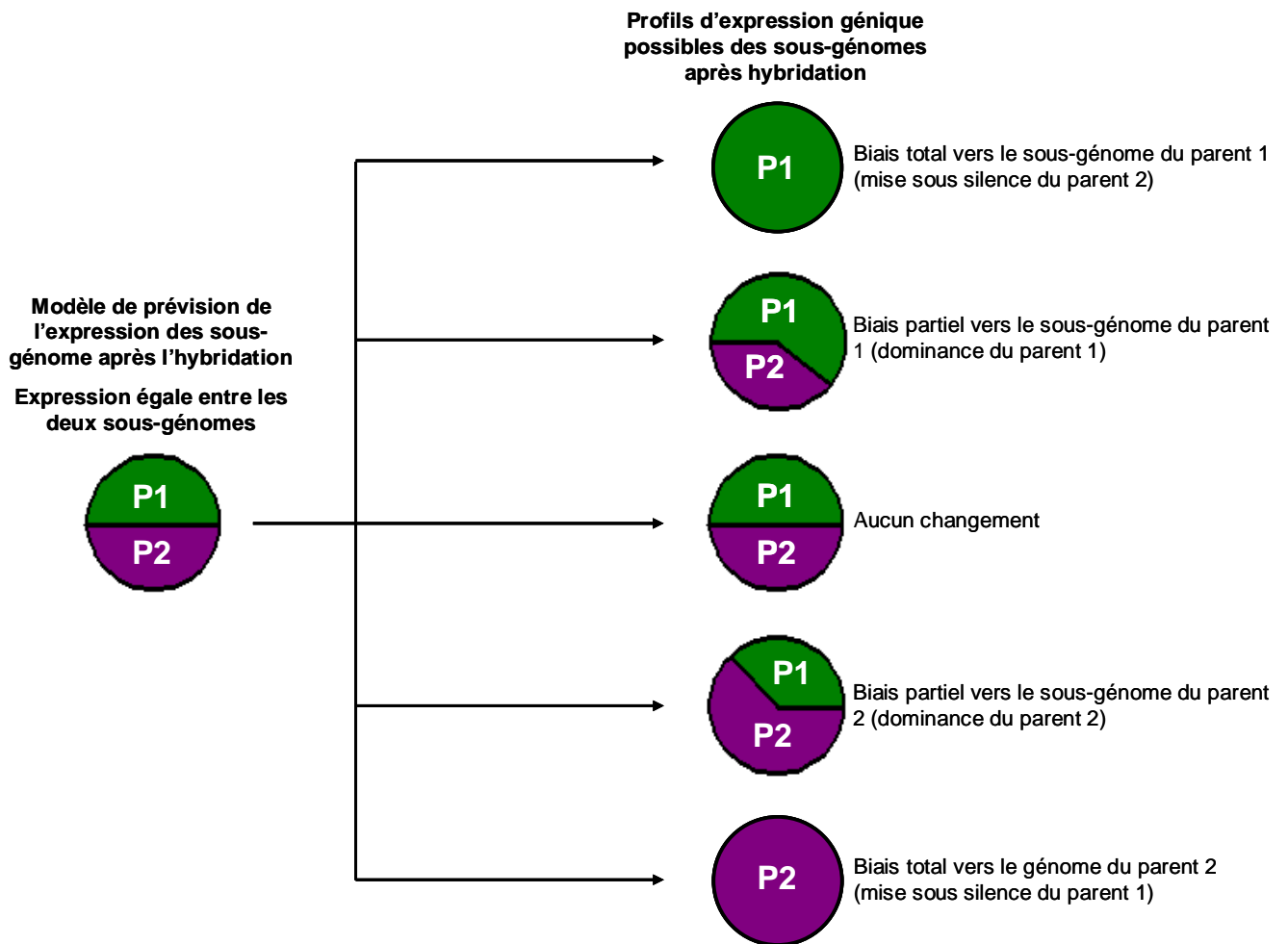
Les conséquences de la polyploïdie sur l'expression génique ont fait l'objet de nombreuses études qui se sont inspirées de modèles de génétique quantitative basés sur les notions d'additivité et de non-additivité (Figure 9) :

- L'additivité se définit par le niveau d'expression d'un gène chez un allotétraploïde qui est égal à la moyenne de l'expression du même gène chez les parents diploïdes (Figure 9).
- La non-additivité se définit par le niveau d'expression d'un gène chez un allotétraploïde qui n'est pas égal à la moyenne de l'expression du même gène chez les parents diploïdes. Les polyplodes montrent ainsi des niveaux d'expression intermédiaires, non présents et/ou 'excédants'<sup>18</sup> ceux de leurs parents diploïdes et des sur-expressions et sous-expressions du transcriptome (Figure 9).

Ces deux notions sont associées à des modèles théoriques qui tentent de catégoriser les différentes variations de l'expression génique entre l'allopolyplode et ses parents. Les modèles théoriques prévoient 3 grandes catégories, l'additivité, la dominance et la transgression (Figure 10 ; Rapp *et al.*, 2009). Ces catégories sont subdivisées en sous catégories (en général, 12 catégories au total ; Figure 10) suivant l'orientation du gène de l'allopolyplode vers une des espèces parentales et/ou la sur-régulation ou sous-régulation du gène d'intérêt chez l'allopolyplode comparé aux parents. Certaines études ajoutent une treizième catégorie « No change » pour laquelle aucune variation n'est observée (Figure 10). Il faut tout de même noter qu'il est difficile de classer l'expression d'un gène de façon exacte. En effet pour s'approcher des modèles théoriques, il est nécessaire d'avoir une approche statistique qui mesure quantitativement l'expression de l'allopolyplode comparée à celle des parents permettant ainsi de classer ce gène dans une des catégories. Au cours de ces analyses statistiques, le choix du seuil de probabilité est important car il détermine le nombre de gène pouvant entrer ou non dans les catégories d'additivité, de dominance et de transgression. Pour un seuil donné, il existe une catégorie de gènes dont le niveau d'expression sera défini comme « équivalent » entre les trois espèces et qui seront donc classés dans la catégorie « No change ». Cette catégorie traduirait l'écart existant entre les modèles théoriques et la mesure de l'expression. Cet écart est dépendant de la précision et de la répétabilité de la mesure. Dans la bibliographie, l'interprétation des modèles théoriques varie selon les auteurs. D'autre part les pourcentages d'additivité et de non-additivité sont observés en fonction de techniques

---

<sup>18</sup> Ces niveaux d'expression excédants les niveaux d'expression des parents diploïdes peuvent être positifs ou négatifs. On parle de sur-expression quand le niveau d'expression 'excédant' est positif et de sous-expression quand le niveau d'expression 'excédant' est négatif.



**Figure 11 : Profils des sous-génomes parentaux chez les allopolyploïdes.**

P1 et P2 représentent les espèces parentales. Les proportions en vert et en mauve représentent les proportions de transcrits homéologues de chaque sous-génome. Les trois catégories sont représentées, le biais total vers l'un des sous-génomes, le biais partiel vers l'un des sous-génomes et la catégorie « aucun changement ».

*D'après Flagel et al., 2008*



statistiques différentes utilisées par les auteurs. Il est faut souvent relativiser les comparaisons effectuées entre les espèces, les techniques et les seuils statistiques observées dans la bibliographie.

### **I.2.2.2 Observation de l'expression relative des gènes homéologues**

Le rapport des quantités de transcrits homéologues peut être exprimé en pourcentage de l'expression totale. Différents profils d'expression des sous-génomes peuvent être observés (Figure 9 ; Flagel *et al.*, 2008) :

- Le profil 'aucun changement' pour lequel les sous-génomes sont coexprimés de manière égale chez l'allopolypléide (Figure 11).
- Les profils de 'biais partiels vers l'un des sous-génomes parentaux' pour lesquels la co-expression des sous-génomes est inégale. Un des sous-génomes parentaux est le plus exprimé. La détermination de ce biais est fonction des analyses statistiques utilisées (Figure 11).
- Les profils de 'biais total vers l'un des sous-génomes parentaux' pour lesquels un des sous-génomes parentaux est mis sous silence (Figure 11).

### **I.2.2.3 Les autopolyploïdes et les variations d'expression génique**

Chez les autopolyploïdes, il existe peu d'études sur le plan expression de gènes (Parisod *et al.*, 2010). Dans la plupart des études, il a été observé que le phénomène d'autopolyploïdie a peu d'influence sur le transcriptome et le protéome. Au niveau du transcriptome, seulement 10% des variations d'expression ont été observés chez des autopolyploïdes synthétiques de *Solanum phureja* (Stupar *et al.*, 2007), 4.2% chez des autopolyploïdes synthétiques d'*Isatis indigotica* (Lu *et al.*, 2006) et 1.3% chez des autopolyploïdes synthétiques de *Paspalum notatum* (Martelotto *et al.*, 2005). En ce qui concerne les espèces autopolyploïdes naturelles, les observations sont identiques. Une étude de Church et Spaulding (2009) portant sur l'espèce *Helianthus decapetalus*, montre que seulement 6.6% des variations d'expression sont observés. Au niveau du protéome, des résultats similaires sont observés. Une étude d'Albertin *et al.* (2005) ne montre aucun changement significatif dans l'expression de gènes chez l'autopolyploïde *Brassica oleracea*. Comme le suggère Parisod *et al.* (2010), le doublement génomique induit peu de variations dans l'expression de gènes que ce soit à court terme (étude sur des autopolyploïdes synthétiques) ou à long terme (étude sur des autopolyploïdes naturels).

#### **I.2.2.4 Les allopolyploïdes et les variations d'expression génique**

Ces dernières années, le rôle de l'allopolyploïdie dans les variations d'expression génique a fait l'objet de nombreuses études. La majorité de ces études traite des variations d'expression au niveau transcriptomique. A ce niveau, l'expression des gènes peut être observée sous deux angles, soit en étudiant de façon globale la divergence d'expression entre l'allopolyploïde et ses espèces parentales soit en mesurant le ratio des transcrits dérivant de chaque gène homéologue. Des études ont également été menées sur le protéome et le métabolome.

##### **I.2.2.4.1 Divergence globale entre les espèces polyploïdes et leurs espèces parentales diploïdes**

La plupart des analyses transcriptomiques traitant de la divergence globale d'expression entre les allopolyploïdes et leurs parents ont été menées sur plusieurs espèces à partir de puces à ADN<sup>19</sup>. Chez *Arabidopsis*, Wang *et al.* (2006) montre qu'entre 5 à 38% (selon le mode de comparaison) des gènes ont une expression non-additive chez des allopolyploïdes synthétiques comparés à une valeur de MPV<sup>20</sup> (*A. Thaliana* et *A. arenosa*). Cette différence d'appréciation est due aux différentes analyses statistiques utilisées. Les auteurs obtiennent 5,2 à 5,8% avec une analyse de 'common variance'<sup>21</sup> et 32 à 38% avec une analyse de 'per-gene variance'<sup>22</sup>. Une étude similaire a été menée sur *Brassica* par Gaeta *et al.* en 2009 utilisant la même puce et les mêmes analyses statistiques que Wang *et al.* (2006). Les auteurs montrent que pour deux lignées d'allopolyploïdes synthétiques de *Brassica napus* (Lignées S0 et S5) comparées au MPV, des valeurs de 1,6 à 32% sont obtenues par une analyse de 'per gene variance' et des valeurs de 2,9 à 4,7% sont obtenues par une analyse de 'common variance'. Wang *et al.* (2006) montre également que les gènes ayant un rôle dans la biosynthèse de la chlorophylle et dans le métabolisme de l'amidon sont sur régulés et que les gènes répondant au stress et codant pour des phytohormones ont une expression non-additive chez l'allopolyploïde. Chez *Gossypium*, Rapp *et al.* (2009) montre que 19 à 30% des gènes ont une expression non-additive chez les allopolyploïdes (respectivement A2G1 et A2D1) comparés aux parents respectifs. En calculant un MPV (*G. arboreum*/*G. bickii* et *G. arboreum*/*G. thurberi*) in silico ils obtiennent des résultats très différents. Seulement 1 à 6% des gènes ont une expression non-additive chez les allopolyploïdes (respectivement A2G1 et

<sup>19</sup> Technique mise au point pour analyser les niveaux d'expression des gènes dans différents compartiments biologiques (les tissus, les organes ou encore les cellules). Son principe repose sur le fait que l'ADN dénaturé reforme sa double hélice à partir d'un brin complémentaire (réaction d'hybridation).

<sup>20</sup> Le « Mid-parent value » (MPV) représente l'expression moyenne parentale.

<sup>21</sup> Méthode pour laquelle tous les gènes sont considérés.

<sup>22</sup> Méthode pour laquelle pour chaque gène on estime les changements significatifs d'expression génétique.

A2D1) comparés au MPV in silico<sup>23</sup>. Les auteurs pensent que l'utilisation d'un MPV masquerait des gènes qui auraient une expression non-additive. Des gènes qui apparaissent exprimés au même niveau que l'expression parentale moyenne correspondraient en réalité à des situations de dominance de l'un des deux parents. Les auteurs montrent d'ailleurs que la majorité de l'expression des gènes non-additifs se traduit par un biais de dominance en faveur de l'un des parents. Dans la littérature, l'expression transgressive est décrite comme pouvant avoir d'importantes conséquences phénotypiques et adaptatives. Une étude récente de Flagel *et al.* (2010), portant sur cinq espèces allopolyploïdes naturels domestiqués et sauvages de *Gossypium* et d'un hybride F1 a été menée. Les auteurs observent que le nombre de gènes exprimant la dominance d'une des deux espèces parentales diminue entre les espèces naturelles et synthétiques tandis que la fréquence du nombre de gènes exprimant la transgression augmente. Cela soulignerait le fait que le transcriptome évoluerait après le phénomène de polyploïdisation. Chez *Triticum*, Chagué *et al.* (2010) montre que des gènes ont une expression non-additive chez les allopolyploïdes synthétiques de différentes générations comparés au MPV in vivo<sup>24</sup> (entre différents génotypes de *Triticum turgidum* et plusieurs accessions d'*Aegilops tauschii*). Des espèces naturelles très récentes ont également été étudiées. Chez *Spartina*, Chelaifa *et al.* 2010 montre que 6 à 5% des gènes sont différentiellement exprimés entre l'allopolyploïde naturel *Spartina anglica* et ses parents diploïdes (*Spartina maritima* et *Spartina alterniflora*). Les auteurs ont également comparé un MPV in silico à l'allopolyploïde et montre que 4,6% ont une expression non-additive. Chez *Senecio*, Hegarty *et al.* 2005 montre que 3% des gènes ont une expression non-additive chez l'allopolyploïde (*Senecio cambrensis*) comparé à ses parents diploïdes<sup>25</sup> (*Senecio squalidus* et *Senecio vulgaris*).

---

<sup>23</sup> Le MPV in silico correspond à des valeurs moyennes d'expression observées chez les espèces parentales (Chagué *et al.*, 2010)

<sup>24</sup> Le MPV in vivo correspond à un mélange équivalent d'ARN de chaque espèce parentale (Chagué *et al.*, 2010)

<sup>25</sup> L'auteur n'a pas utilisé de MPV, mais a fait des comparaisons deux à deux avec chacune des espèces étudiées.

#### **I.2.2.4.2 Etude de l'expression relative des gènes homéologues chez les espèces allopolyploïdes**

Les analyses transcriptomiques traitent également de la comparaison de l'expression des gènes homéologues chez les allopolyploïdes en mesurant le ratio des transcrits dérivant de chaque gène homéologue. Chez *Gossypium*, une puce spécifique (Udall *et al.*, 2006) permettant de distinguer<sup>26</sup> les niveaux d'expression des copies homéologues a été produite. Flagel *et al.* (2008) montrent que 24% des gènes auraient une expression dite biaisée chez l'allopolyploïde naturel (*Gossypium hirsutum*) ; c'est à dire une contribution inégale des génomes parentaux serait observée. Les auteurs montrent que le biais observé est en faveur de l'un des parents (le génome D ; *Gossypium raimondii*) et confirme les observations de Rapp *et al.* (2009). Une étude récente de Flagel *et al.* (2010) et complémentaire à celle de 2008, confirme le biais en faveur du génome D (54 à 60%) au travers de cinq espèces allopolyploïdes naturels domestiqués et sauvages de *Gossypium* et d'un hybride F1. Chez *Gossypium*, des biais en fonction du stade développement de la fibre de coton (Hovav *et al.*, 2008), de différents tissus (Chaudhary *et al.*, 2009) et de différents stress (Dong *et al.*, 2011) sont observés. Chaudhary *et al.* (2009) a menée une étude, chez des allopolyploïdes naturels et synthétiques ainsi que sur les espèces parentales diploïdes, comparant les niveaux d'expression relatifs de chaque copie homéologues des 63 paires de gènes obtenus à partir de 24 tissus. Les auteurs observent un biais d'expression d'environ 40% (des 63 paires de gènes homéologues étudiées) dans au moins un des tissus étudiés. Les auteurs mettent également en évidence différents niveaux de biais variables d'un tissu à un autre. Une étude récente de Dong *et al.* (2011), menée sur l'expression de gènes dupliqués d'un allopolyploïde naturel (*Gossypium hirsutum*) et d'un synthétique (*G. arboreum* x *G. davidsonii*) dans différents tissus et soumis à différents stress, montre que 11 paires d'homéologues ont une expression biaisée en fonction du stress et que, dans certains cas, cette expression biaisée serait organe-spécifique. D'autres études ont été menées chez différentes espèces telles que *Arabidopsis* (Chang *et al.*, 2010 ; Observation d'un biais parental), *Triticum* (Stamati *et al.*, 2009 ; Observation de variations tissu- et stress-spécifiques dans l'expression de gènes), ou encore *Tragopogon* (Buggs *et al.*, 2010 ; Observation de variations tissu-spécifiques dans l'expression de gènes).

---

<sup>26</sup> La méthodologie de cette puce est basée sur la détection des SNPs spécifiques du génome A et D ou sur des petites insertions/délétions.

#### **I.2.2.4.3 Rôle de l'hybridation interspécifique et de la duplication de génome des allopolyploïdes**

Il est difficile de connaître le rôle de l'hybridation et de la duplication de génome dans l'expression génique. Très peu d'études ont tenté d'estimer la contribution des deux phénomènes dans les changements d'expression génétique. Hergaty *et al.* (2006), Flagel *et al.* (2010) et Chelaifa *et al.* (2010) ont montré que l'hybridation jouait un rôle majeur dans les changements d'expression chez des allopolyploïdes synthétiques et naturels de *Senecio*, *Gossypium* et *Spartina*. Cependant chez *Senecio*, Hergaty *et al.* (2006) montrent que les changements s'atténuent après la duplication génomique. Alors que chez *Gossypium* et *Spartina*, les auteurs montrent que des changements additionnels apparaissent chez les allopolyploïdes suite à la duplication du génome (Chelaifa *et al.*, 2010 ; Flagel *et al.*, 2010).

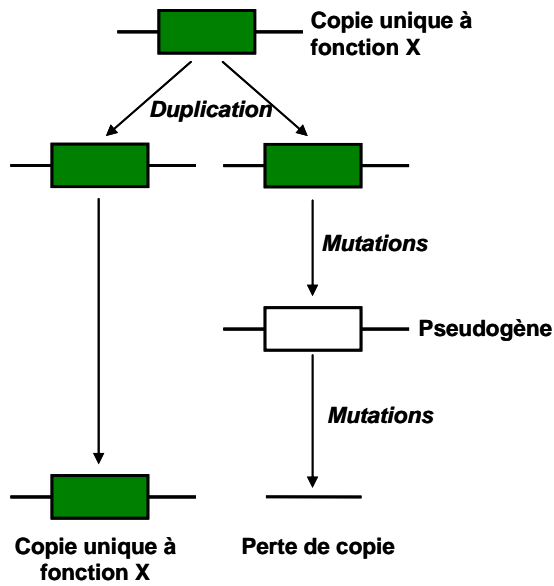
#### **I.2.2 Etude du protéome, métabolome**

Des études sur le protéome ont été menées chez deux espèces, *Brassica* et *Gossypium*. Albertin *et al.* (2006) montrent que, chez des allopolyploïdes synthétiques de *Brassica napus*, 25 à 38% des polypeptides ont une expression non-additive. Certaines expressions non-additives de protéines sont différentielles suivant les organes étudiés montrant ainsi une régulation organe-spécifique. Les auteurs ont également observé un biais d'environ 60% vers l'une des espèces parentales (*Brassica rapa*). Le rôle de l'hybridation et du doublement génomique a également été étudié et montre que 89% des variations sont associés au phénomène d'hybridation interspécifique. Dans une seconde étude, Albertin *et al.* (2007) souligne qu'environ 50% des isoformes et des composantes de complexes protéiques montrent des profils de régulation différents chez les allopolyploïdes. Chez *Gossypium*, une étude récente de Hu *et al.* (2011) révèlent que le protéome des allopolyploïdes (ici *Gossypium hirsutum*) est similaire au protéome de l'espèce parentale D (26% de spots 2-DE<sup>27</sup> détectés chez l'allopolyploïde sont présents uniquement chez le génome D et 11% chez le génome A) et qu'un biais de l'expression non-additive en faveur du génome D est observé (14% pour le génome D contre 10% pour le génome A). Ces résultats confirment ceux observés dans l'étude du transcriptome de Rapp *et al.* (2009).

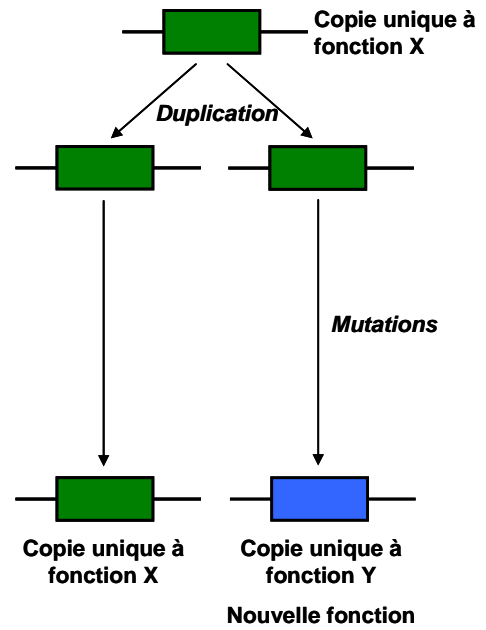
Une étude sur le métabolome de l'espèce *Nicotiana* a été menée par Pearse *et al.* en 2006. Cette étude porte sur trois lignées d'allopolyploïdes synthétiques (*Nicotiana miersi*) et leurs espèces parentales (*N. miersi* et *N. attenuata*). Les auteurs montrent que les profils de métabolites secondaires dits 'anti-herbivores' diffèrent entre les trois lignées allopolyploïdes ayant les mêmes espèces parentales et entre les allopolyploïdes et les espèces parentales. Les

<sup>27</sup> La technique 2-DE (two-dimensional gel electrophoresis) est une technique classique de protéomique. Des spots de protéines sont obtenus puis sont identifiés par spectrophotométrie de masse (MS).

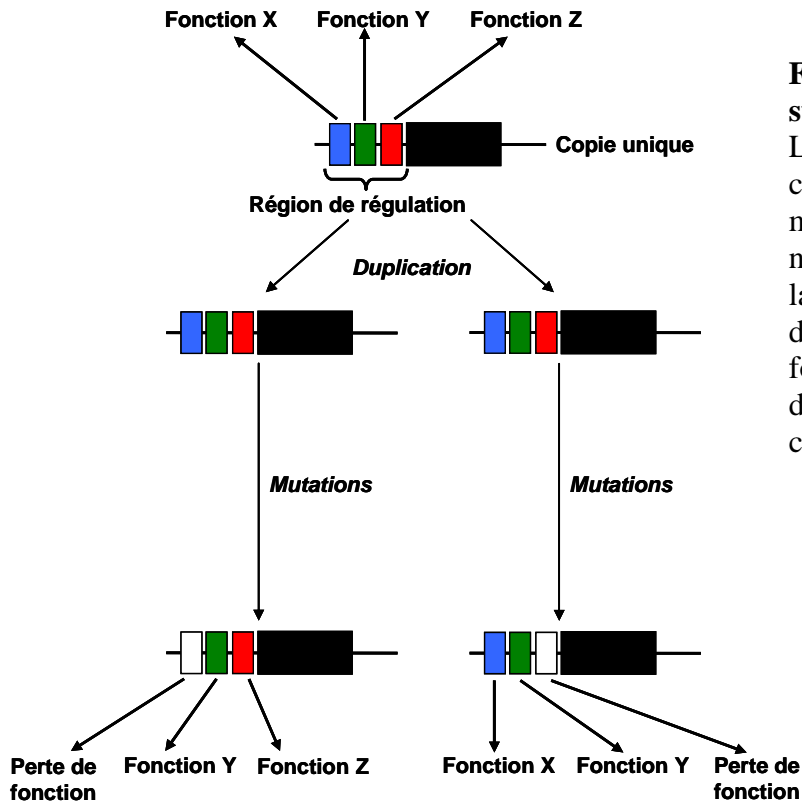
## PSEUDOGENISATION



## NEOFONCTIONNALISATION



## SOUS-FONCTIONNALISATION



### Figure 12.A : Modèles théoriques sur le devenir des gènes dupliqués.

Les trois modèles impliquent un changement génétique (une mutation). La pseudogénéisation et la néofonctionnalisation permettent soit la perte de copie soit l'apparition d'une nouvelle fonction. La Sous-fonctionnalisation permet le maintien de la fonction ancestrale sur les deux copies de gènes homéologues.

*D'après Luchetta et al., 2005*

auteurs suggèrent que les allopolyploïdes seraient capable d'intégrer deux systèmes physiologiques différents dans un seul système plastique.

### **I.3 MECANISMES DE REGULATION CONTRIBUANT AUX VARIATIONS DANS L'EXPRESSION DES GENES CHEZ LES POLYPLOÏDES**

Depuis quelques années, de nombreux auteurs soulignent le fait que la polyploïdie aurait des effets immédiats sur le phénotype en s'appuyant sur l'augmentation de la taille de la biomasse qui traduirait la vigueur hybride de l'individu (Osborn *et al.*, 2003). Ces effets sur le phénotype pourraient être liés à l'augmentation de la quantité d'ADN dans la cellule indépendamment de l'information qu'elle véhicule (Bretagnolle *et al.*, 1998), ce qui participerait aux changements de taille et de forme ou encore à un ralentissement des divisions cellulaires (Stebbins, 1971 ; Bretagnolle *et al.*, 1998) Cependant, Bretagnolle *et al.* (1998), soulignent que l'augmentation de la taille des cellules pourraient également être liés à l'information véhiculée par la molécule d'ADN qui, elle-même, peut être modifiée par des mécanismes de régulation résultant de la duplication des gènes. Depuis quelques années, les recherches sur les mécanismes de régulation et leur implication dans des variations de l'expression génique se multiplient. Pour comprendre ces mécanismes et leurs implications, il est nécessaire de se référer aux modèles théoriques élaborés pour expliquer le devenir des gènes dupliqués.

#### **I.3.1 Le devenir des copies de gènes dupliqués**

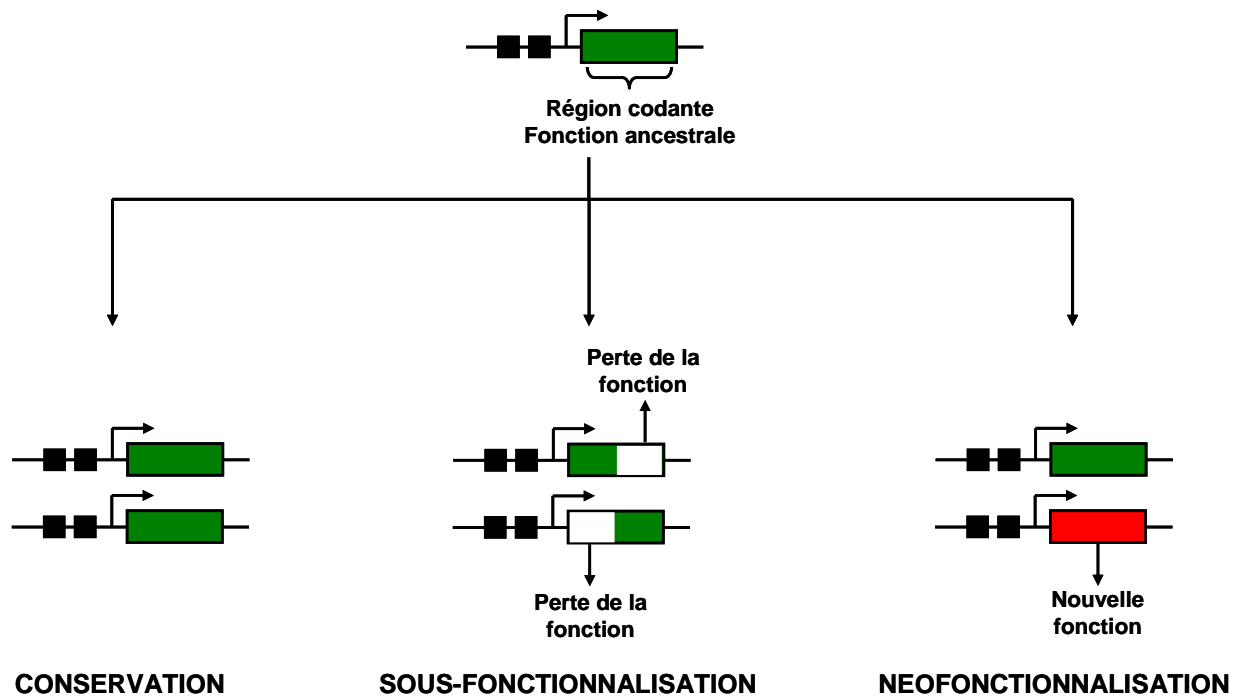
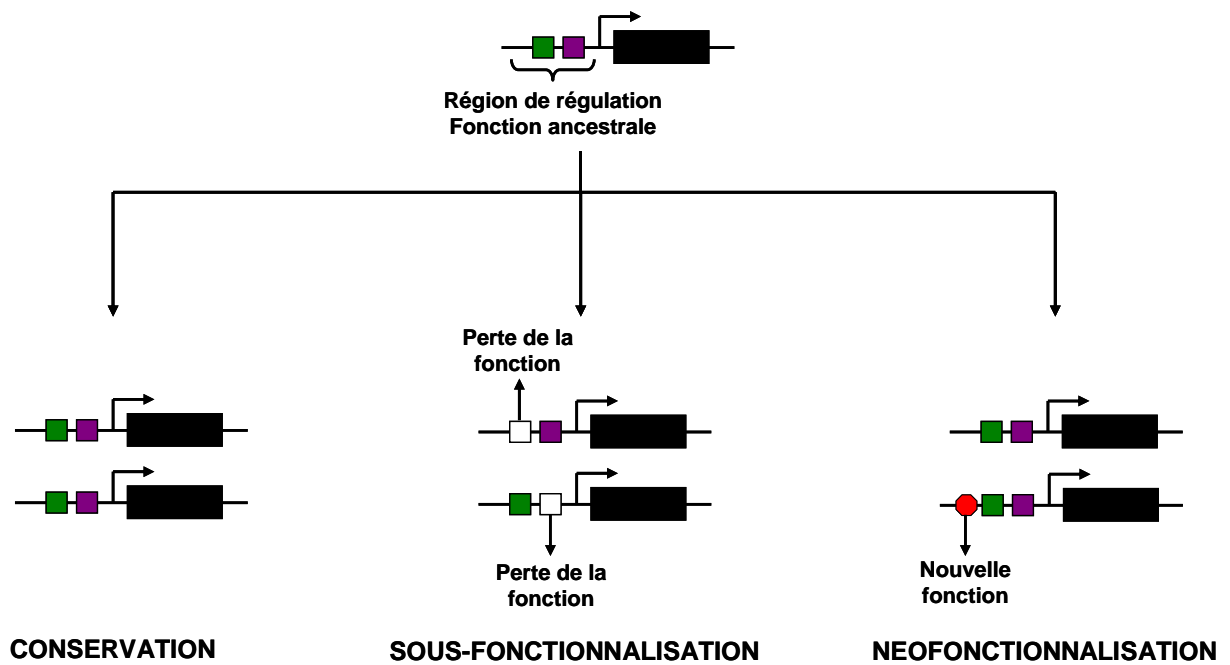
D'après Adams en 2004, le phénomène de polyploïdisation a été décrit comme étant à l'origine d'innovations morphologiques et de comportements originaux des polyploïdes à court terme et d'une évolution des gènes dupliqués vers de nouvelles fonctions à long terme. Sur le plan fonctionnel, plusieurs modèles concernant l'évolution des gènes dupliqués ont été proposés (Ohno, 1970 ; Force, 1999).

##### **I.3.1.1 La pseudogénéisation et la néofonctionnalisation**

Lors de la duplication, l'augmentation du nombre de copies des gènes entraîne des redondances qui peuvent s'avérer délétères pour les espèces. Comme une seule copie serait suffisante pour assurer la fonction ancestrale, l'autre ne subirait pas de pression de sélection et évoluerait ainsi sous l'action de mutations<sup>28</sup>. Deux voies sont possibles selon ce modèle<sup>29</sup> de Ohno (1970), soit une des deux copies est soumise à des mutations induisant une perte de fonction (Pseudogénéisation) et devient un pseudogène (Figure 12.A) ; soit cette copie acquiert une nouvelle fonction par divergence (Neofonctionnalisation ; Figure 12.A). La limite majeure

<sup>28</sup> Modifications de l'information contenue dans l'ADN quelles qu'en soit sa nature (Mutations au sens large).

<sup>29</sup> Modèle, dit « classique », suggéré par Ohno en 1970. Il associe les notions de pseudogénéisation et néofonctionnalisation.



**Figure 12.B : Modèles théoriques sur le devenir des gènes dupliqués.**

Les trois modèles n'impliquent pas nécessairement un changement génétique. Le changement de fonction peut s'observer soit dans la zone de régulation soit dans la zone codante.

*D'après Hahn et al., 2009*



de ce modèle est que seul un petit nombre de copies de gènes dupliqués peut subsister dû aux fortes contraintes sélectives associées à chacun des gènes. Or de nombreux auteurs ont montré que les génomes gardaient un grand nombre de gènes dupliqués fonctionnels (Luchetta *et al.*, 2005).

### I.3.1.2 *La Sous-fonctionnalisation*

Afin de comprendre l'intérêt que peut avoir le génome à garder autant de gènes dupliqués fonctionnels, un second modèle<sup>30</sup> a été proposé par Force *et al.* (1999). Ce modèle se base sur le fait que l'ensemble des gènes dupliqués et préservés ne dépendrait pas uniquement de l'acquisition de nouvelles fonctions mais aussi de la répartition des fonctions ancestrales. Après duplication, chacune des copies accumule de façon indépendante des mutations<sup>31</sup> rendant les deux copies nécessaires à l'obtention d'une fonction équivalente à celle du gène ancestral (Figure 12.A). Chaque copie est soumise aux contraintes sélectives favorisant leur rétention. Dans ce modèle, on retrouve la notion de néofonctionnalisation car à long terme, la préservation des gènes dupliqués augmente fortement la possibilité d'acquérir de nouvelles fonctions (Luchetta *et al.*, 2005).

Dans ces modèles il faut prendre en compte que la plupart des auteurs mettent en avant l'action de mutations, donc de changements génétiques, à l'origine de la néofonctionnalisation et de la sous-fonctionnalisation (Figure 12.A). Depuis les études d'expression génique, des modèles plus récents tels que ceux décrits par Hahn *et al.* (2009) mettent en avant le fait que les changements génétiques ne sont pas nécessairement à l'origine de la néofonctionnalisation et la sous-fonctionnalisation (Figure 12.B) mais qu'elles peuvent être induites par des modifications de mécanismes de régulation. Ces mécanismes de régulation peuvent conduire à la mise sous silence de l'une des deux copies homéologues et ainsi à une possible sous-fonctionnalisation des gènes dupliqués.

### I.3.1.3 *Conservation et équilibre de dosage*

L'équilibre de dosage s'appuie sur l'idée que, dans une cellule, le maintien à l'équilibre d'un nombre de copies de gènes sensibles aux effets doses influencerait sur l'expression des gènes (Vietia *et al.*, 2008). L'équilibre de dosage prendrait en compte un niveau d'interaction supplémentaire dépendant de la stœchiométrie<sup>32</sup> des molécules synthétisées d'une fonction donnée permettant ainsi le bon fonctionnement de celle-ci (Birchler *et al.*, 2005 ; Birchler et Veitia, 2007). En effet, il a été observé que si l'on supprime des copies de gènes codant pour une des sous unités d'un complexe protéique quelconque

<sup>30</sup> Modèle, dit « Duplication-Dégénérescence-Complémentation » (DDC), suggéré par Force *et al.*, en 1999.

<sup>31</sup> Ces mutations peuvent se produire soit dans la région régulatrice, soit dans la région codante.

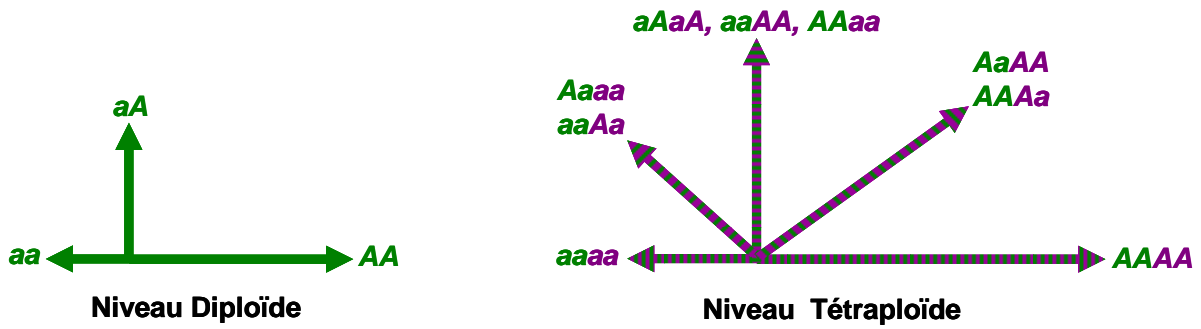
<sup>32</sup> La stœchiométrie correspond ici à un équilibre dans les proportions de molécules synthétisées (Birchler *et al.*, 2005).

cela conduit à un déséquilibre en quantité de ses sous-unités. Ce déséquilibre conduirait à un non-fonctionnement. Des études ont aussi montré que certains groupes de gènes dupliqués, pour la plupart impliqués dans des voies de régulations et ayant conservés leur fonction ancestrale après plusieurs évènements de polyploïdisation, étaient maintenus (Birchler et Veitia, 2010). Les classes de gènes retenus sont généralement représentées par des gènes ayant un rôle dans les complexes protéiques ou par des gènes interagissant avec d'autres protéines (Birchler et Veitia, 2010).

### **I.3.2 Les mécanismes de régulation**

Comme nous l'avons vu précédemment, les gènes dupliqués peuvent fonctionner ensemble dans un même génome par effets de sous-fonctionnalisation (sans mutation) ou de dosage. Il est donc nécessaire de comprendre le rôle des mécanismes de régulation impliqués. En effet, la plasticité des polyploïdes peut être corrélée avec les changements d'expression génétique à des niveaux transcriptionnels et post-transcriptionnels. Comme le souligne Jackson et Chen (2009), les changements d'expression génétiques sont contrôlés en grande partie par des mécanismes épigénétiques, lesquels jouent un rôle, direct ou indirect, de régulateurs de variations physiologiques et développementales chez les allopolyploïdes (Wang *et al.*, 2006 ; Ni *et al.*, 2009). La plupart de ces mécanismes surviendraient dès le début de la formation des polyploïdes alors que d'autres se mettraient en place au cours des générations qui suivent la première hybridation de l'espèce (Adams et Wendel, 2005). Jackson et Chen (2009) soulignent que ces mécanismes expliqueraient la plasticité supérieure des polyploïdes. Cette meilleure plasticité des allopolyploïdes résulterait d'effets sur le phénotype induits par des effets 'dose-dépendants' et des modifications des interactions régulatrices.

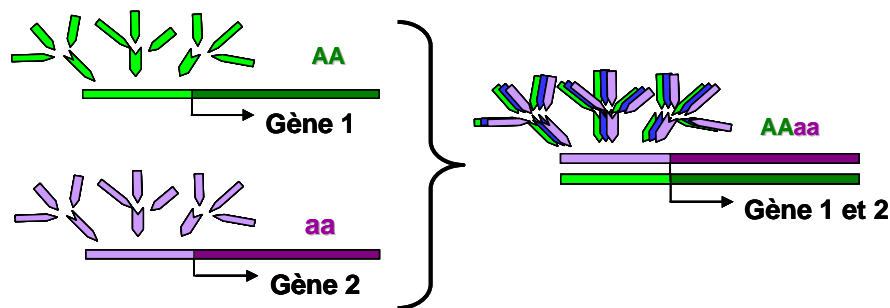




**Figure 13 : Augmentation des possibilités de variations dans l'expression « dose-dépendante » des gènes.**

Le nombre de génotype et les niveaux d'expression génétique varient suivant le niveau de ploïdie. Les génotypes passent de trois au niveau diploïde à neuf au niveau tétraploïde. Les niveaux d'expression génique passe de trois (représentés par les flèches) chez les diploïdes à cinq chez les allopolyploïdes. Chez les allopolyploïdes, on observe donc des classes phénotypiques intermédiaires.

*D'après Osborn et al., 2003*



**Figure 14 : Augmentation de la complexité des interactions régulatrices.**

Sur le premier schéma, AA et aa représentent les génomes des deux espèces diploïdes contributrices du génome allopolyploïdes AAaa. Les flèches représentent les facteurs de régulation (en vert et en mauve ; homodimères) affectant l'expression des gènes 1 et 2 de façon hiérarchique. Chez l'allopolyploïde, on observe une complexification des réseaux de régulation. Ses facteurs de régulation sont représentés par des homodimères venant de parents diploïdes (vert et mauve) et par la création suite à leur réunion dans l'allopolyploïde d'hétérodimère (bleu).

*D'après Osborn et al., 2003*

### **I.3.2.1 Augmentation des possibilités de variations dans l'expression « dose-dépendante » des gènes**

La polyploïdie aurait un effet sur l'augmentation des niveaux d'expression de gènes proportionnellement au nombre de copies des gènes induits par le niveau de ploïdie (Figure 13 ; Osborn *et al.*, 2003). Pour illustrer cette notion, une étude sur *Zea mays* (Guo *et al.*, 1996) a montré que l'augmentation de l'expression de la plupart des gènes au travers de séries de ploïdie (monoploïdes, diploïdes, triploïdes et tétraploïdes) est proportionnelle à l'augmentation du nombre de copies des gènes. La polyploïdie augmente ainsi les variations potentielles dans les niveaux d'expression des gènes « dose-dépendants » (Figure 13). Ce phénomène conduit, non pas au développement de la gamme de phénotypes, mais à une augmentation des classes phénotypiques intermédiaires dont certaines pourraient avoir un avantage sélectif (Osborn *et al.*, 2003).

### **I.3.2.2 Altération des interactions régulatrices**

L'expression de la plupart des gènes dépend de réseaux de régulateurs organisés de façon hiérarchique comme les facteurs de transcription. L'augmentation du niveau de ploïdie peut entraîner une complexification du réseau des gènes de régulation (Figure 14 ; Osborn *et al.*, 2003). Une différence peut être observée entre les allopolyploïdes et les autopolyploïdes. Suite à la réunion de génomes divergents chez les allopolyploïdes, Osborn *et al.* (2003) pensent que l'altération de ce fonctionnement de réseaux de régulation est plus importante que chez les autopolyploïdes où tous les allèles proviennent de la même espèce (Osborn *et al.* 2003). Certains produits hétérologues<sup>33</sup> peuvent ne pas fonctionner, fonctionner tout aussi bien ou même mieux fonctionner que les produits homologues<sup>34</sup> (Figure 14 ; Phillips *et al.*, 1995). Les produits hétérologues pourraient jouer un rôle dans l'augmentation de la vigueur hybride (Chen, 2007).

---

<sup>33</sup> Les produits hétérologues ou hétérodimères possèdent des sous unités différentes.

<sup>34</sup> Les produits homologues ou homodimères possèdent des sous unités identiques.

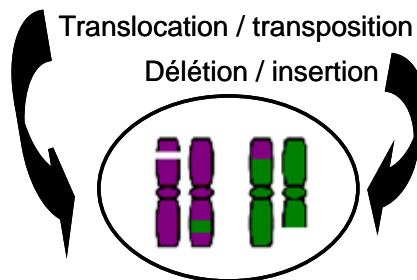
# MECANISMES GENETIQUES

## CAUSES POSSIBLES

Transposons  
Appariement de chromosomes homéologues



## EFFETS



**Figure 15 : Mécanismes génétiques.**

Les effets observés sur les chromosomes sont la translocation/transposition et la délétion/insertion. Ces modifications de la séquence de l'ADN sont notamment induites par des appariements de chromosomes homéologues et des transposons.

### I.3.2.3 Mécanismes génétiques et épigénétiques

De nombreux mécanismes affectent les espèces polyploïdes tels que des changements génétiques à l'origine de perte, mutation et divergence de certains gènes dupliqués et des mécanismes épigénétiques modifiant l'expression des gènes et les voies de développement (Chen, 2007). La contribution de ces mécanismes est variable suivant les espèces d'études. Chez des allopolyploïdes de *Triticum*, *Brassica* et *Tragopogon* (Song *et al.*, 1995 ; Feldman *et al.*, 1997 ; Shaked *et al.*, 2001 ; Tate *et al.*, 2006), les mécanismes génétiques prédominent alors que les mécanismes épigénétiques prédominent chez des allopolyploïdes de *Gossypium* et d'*Arabidopsis* (Lee et Chen, 2001 ; Adams *et al.*, 2003 ; Wang *et al.*, 2004 et 2006).

#### I.3.2.3.1 Les mécanismes génétiques

Les modifications de la séquence de l'ADN, c'est à dire les délétions/insertions et les translocations/transpositions peuvent être soit induites par l'appariement de chromosomes homéologues soit par l'action de transposons (Figure 15). Ces modifications ont été mises en évidence grâce à des méthodes d'analyse de fragments d'ADN comme la RFLP<sup>35</sup>et/ou l'AFLP<sup>36</sup> (Osborn *et al.*, 2003). Chez des allotétraploïdes synthétiques de *Brassica* (Song *et al.*, 1995), la plupart des changements sont dus à des translocations et transpositions entre chromosomes homéologues alors que des évènements d'élimination de séquences sont plus fréquents chez les allotétraploïdes synthétiques de *Triticum* (Feldman *et al.*, 1997 ; Shaked *et al.*, 2001) et les allotétraploïdes naturels de *Tragopogon* (Tate *et al.*, 2006). Ces modifications n'ayant lieu que sur un seul des gènes homéologues, peuvent affecter les niveaux d'expression des gènes « dose-dépendants » et la fonction des réseaux de régulation chez les allopolyploïdes récents (Osborn *et al.*, 2003) et conduire à de la néofonctionnalisation et sous-fonctionnalisation des gènes dupliqués. Elles peuvent s'effectuer immédiatement après l'allopolyploïdisation comme chez *Triticum* (Feldman *et al.*, 1997 ; Shaked *et al.*, 2001) ou tardivement comme chez *Brassica* (Tate *et al.*, 2006). Chez ces deux espèces, ces changements sont très fréquents, jusqu'à 14% du génome pour *Triticum* (Feldman *et al.*, 1997 ; Shaked *et al.*, 2001), alors que chez des allotétraploïdes synthétiques de *Gossypium* (Lui *et al.*, 2001) et des allopolyploïdes naturels de *Spartina* (Salmon *et al.*, 2005), la fréquence de

<sup>35</sup> Restriction Fragment Length Polymorphisms. Technique basée sur les variations de séquences homéologues d'ADN. Analyses portant sur plusieurs loci dans la même expérience.

<sup>36</sup> Amplified Fragment Length Polymorphisms. Technique basée sur les variations de séquences homéologues d'ADN. Analyses portant sur un seul locus par expérience mais où les variations alléliques peuvent être observées.

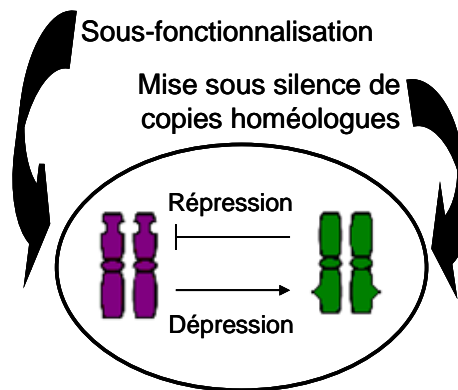
## MODIFICATIONS EPIGENETIQUES

### CAUSES POSSIBLES

Méthylation de l'ADN / Acétylation des histones  
ARNi



### EFFETS



**Figure 16 : Mécanismes épigénétiques.**

Les effets observés sur les chromosomes sont la sous-fonctionnalisation et la mise sous-silence de copie homéologue. Ces modifications de la structure de l'ADN sont induites par de la méthylation de l'ADN et des ARNi.



ces changements est plus faible voir très faible (environ 1%) comme chez des allopolyploïdes d'*Arabidopsis* (Comai *et al.*, 2000 ; Madlung *et al.*, 2002). Cette faible fréquence traduirait un haut niveau de tolérance à l'hybridation et au doublement génomique (Chen et Ni, 2006). L'appariement de chromosomes homéologues est contrôlé par des locus comme Ph1<sup>37</sup> observé chez un hexaploïde de *Triticum* (Riley et Chapman, 1958) ou PrBn<sup>38</sup> chez *Brassica napus* (Jenczewski *et al.*, 2003). Chez *Triticum*, le *Ph1* réprime l'appariement homéologue et résoudrait également des appariements incorrects (Moore, 2002). L'élimination de séquences augmenterait la différenciation entre les chromosomes homéologues entraînant un comportement méiotique de l'allopolyploïde similaire à ses parents diploïdes et la formation de bivalents uniquement (Ozkan *et al.*, 2001). Chez les allopolyploïdes, les réorganisations structurelles des génomes peuvent être également évitées par la mise sous silence de transposons. Chez deux espèces, il a été observé que des événements de transposition induisaient des hauts niveaux de délétions de séquences (*Triticum* ; Feldman *et al.*, 1997 ; Shaked *et al.*, 2001) et des échanges chromosomiques homéologues (*Brassica* ; Pires *et al.*, 2004 ; Udall *et al.*, 2005).

### I.3.2.3.2 Les mécanismes épigénétiques

Les modifications de la structure de l'ADN, c'est à dire la mise sous silence de gènes homéologues et les modifications issues de la sous-fonctionnalisation, peuvent être induites par la méthylation de l'ADN, l'acétylation des histones et par les ARNi<sup>39</sup> (Figure 16). La mise sous silence de gènes et la sous-fonctionnalisation sont liées. Le fait de mettre sous silence une des deux copies parentales induirait la sous-fonctionnalisation de ces copies homéologues permettant ainsi leur expression de façon alternative dans différents organes et/ou différents stades de développement. La mise sous silence d'une des copies parentales a été observé chez plusieurs espèces telles que des allotétraploïdes synthétiques de *Brassica* (Chen et Pikaard, 1997) et d'*Arabidopsis* (Comai *et al.*, 2000 ; Wang *et al.*, 2004) ainsi que des allotétraploïdes naturels de *Tragopogon* (Buggs *et al.*, 2009). Chez les *Brassica*, les gènes transcrits par l'ARN polymérase I sont mis sous silence et peuvent être réactivés par l'utilisation d'inhibiteur chimique de la méthylation de l'ADN et de l'acétylation d'histones (Chen et Pikaard, 1997). Cette mise sous silence est donc réversible et contrôlée par des mécanismes épigénétiques de méthylation et d'acétylation. Chez *Arabidopsis*, la mise sous silence de

<sup>37</sup>Le gène « Homeologous pairing locus 1 » (Ph1) est situé sur le chromosome arm 5BL du blé. Le gène Ph1 déterminerait des appariements exclusivement homologues (Riley and Chapman, 1958).

<sup>38</sup> Comme le Ph1, le gène « Pairing regulator in *B. napus* » (PrBn) aurait un rôle majeur dans le contrôle des appariements homéologues (Jenczewski *et al.*, 2003).

<sup>39</sup> Les ARNi sont des petits ARNs complémentaires d'ARN de gènes. Il s'hybride à l'ARN messager de gène et conduit à sa dégradation.

copies parentales a lieu dans la première et deuxième générations (Wang *et al.*, 2004). D'après l'étude de Wang *et al.* (2004), deux gènes montrent une sur-expression chez les parents diploïdes et une mise sous silence chez l'autotétraploïde *A. thaliana*. Ils sont ensuite réactivés chez l'allopolyploïde formé suite à l'hybridation entre *A. thaliana* et le diploïde *A. arenosa*. De plus un troisième gène mis sous silence chez les parents diploïdes est activé dès la formation de l'autopolyploïde. On peut donc supposer que l'expression des gènes peut être affecté par l'hybridation et la duplication génomique (Adams et Wendel, 2005). La mise sous silence a également été observée chez des allotétraploïdes naturels de *Tragopogon* alors qu'aucune présence n'a été observée chez les synthétiques. Ces évènements de mise sous silence ne sont pas fixés chez les allopolyploïdes naturels de *Tragopogon* (Buggs *et al.*, 2009). La mise sous silence et les variations des niveaux d'expression des gènes dupliqués peuvent varier dans différents organes de plantes indiquant une différence de régulation entre les deux gènes homéologues pendant leur développement (Adams et Wendel, 2005). Des études menées sur des allopolyploïdes synthétiques et naturels de *Gossypium* (Adams *et al.*, 2003 et 2004) montrent que 10 gènes sur 40 ont une expression inégale et organe-spécifique. Certains locus sont ainsi sur-exprimés dans un des sous-génomes appartenant à un organe et d'autres dans l'autre sous-génome appartenant à un autre organe. Les auteurs ont observé qu'un des homéologues de l'adhA<sup>40</sup> s'exprime dans les pétales et l'autre dans le style. Cette sous-fonctionnalisation est observée chez les allopolyploïdes synthétiques et chez les allopolyploïdes naturels formés il y a 1 à 2 MA. Ces observations suggèrent la possibilité de mécanismes épigénétiques qui se maintiennent au cours de l'évolution de la plante (Adams et Wendel, 2005). Une étude récente (Buggs *et al.*, 2010) menée chez des allopolyploïdes naturels de *Tragopogon* montre qu'au travers de tous les tissus étudiés, 20% des gènes sont mis sous silence pour l'une des copies homéologues et cela pour une seule plante et un tissu particulier et 8% des gènes sont mis sous silence pour l'un des gènes homéologues. Deux cas d'expression tissus-spécifiques réciproques entre homéologues ont été observés et indiquent une possible sous-fonctionnalisation (Buggs *et al.*, 2010).

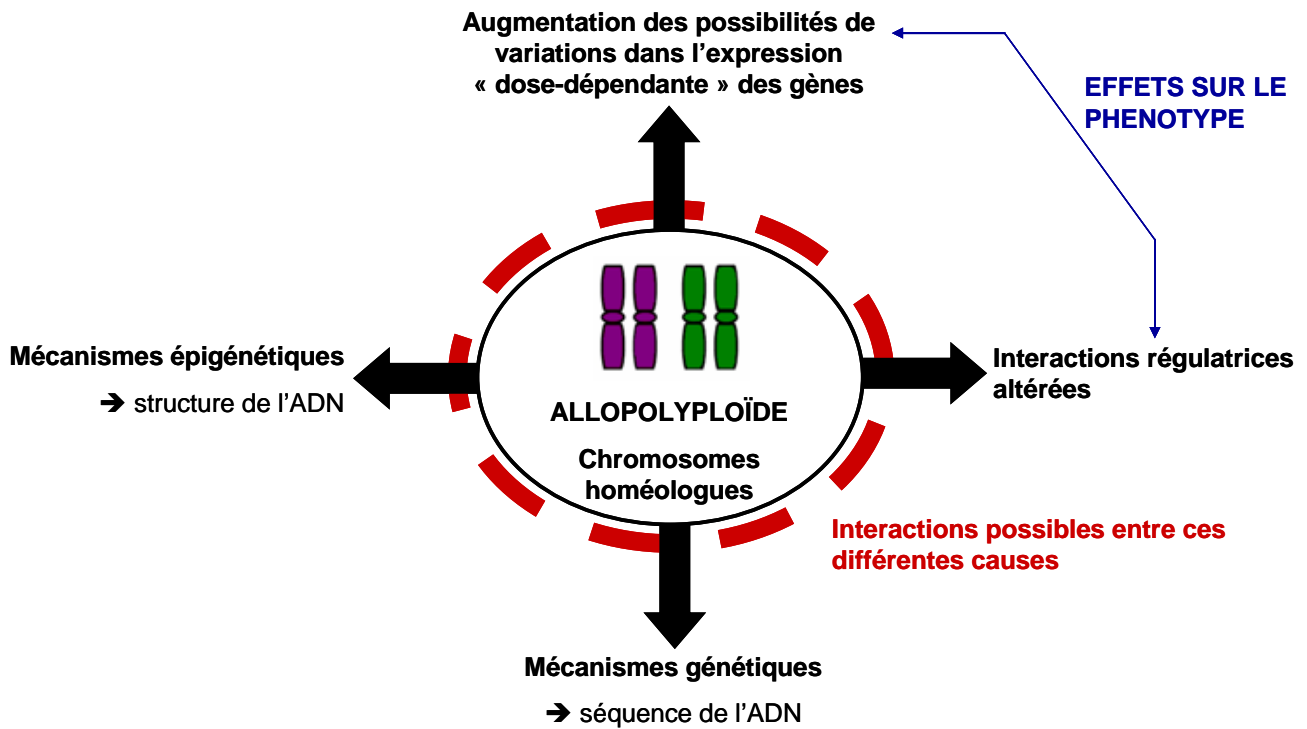
Le rôle de la méthylation de l'ADN et des ARNi dans la mise sous silence ou l'activation de gènes chez les allopolyploïdes a été mis en évidence chez des allotétraploïdes synthétiques d'*Arabidopsis* (Madlung *et al.*, 2002 ; Wang *et al.*, 2004). Madlung *et al.* (2002) ont observé que des changements de méthylation d'ADN étaient plus fréquents chez les allopolyploïdes que chez les parents. Un traitement à l'aza-dC<sup>41</sup> a révélé la sensibilité des allopolyploïdes aux

---

<sup>40</sup> L'Alcohol dehydrogenase A (adhA) est une protéine anaérobie qui catalyse la réduction du pyruvate en éthanol.

<sup>41</sup> Le 5-aza-2'-deoxycytidine (Aza-dC) est un inhibiteur chimique de la méthylation de l'ADN (Madlung *et al.*, 2002).





**Figure 17 : Les mécanismes de régulation.**

Figure récapitulatif des différents mécanismes de régulation. L'augmentation des possibilités de variations dans l'expression « dose-dépendante » des gènes et les interactions régulatrices altérées ont un effet direct sur le phénotype. Cette figure insiste sur le fait que tous ces mécanismes peuvent interagir entre eux.

modifications de la chromatine. Chez *Spartina*, 30% des profils de méthylation parentaux sont altérés chez les hybrides et allopolyploïdes (Salmon *et al.*, 2005). Le rôle de la méthylation de l'ADN dans la mise sous silence de copies de gènes parentaux a également été étudié dans des lignées transgéniques hypométhylées de allopolyploïde, *Arabidopsis suecica*, par l'intermédiaire d'ARNi (Wang *et al.*, 2004). Les auteurs ont observé que deux gènes sont réactivés dans des lignées transgéniques ddm1 et met1-RNAi par la déméthylation de l'ADN (Wang *et al.*, 2004). L'acétylation et la désacétylation des histones ne sont pas des mécanismes indépendants dans la mise sous silence de gènes (Chen et Tian, 2007). Une étude d'Earley *et al.* (2006), montre que lorsqu'une histone désacétylase (HDA6) est perturbée cela induit une baisse des niveaux de méthylation de promoteurs de la cytosine et provoque des modifications de plusieurs histones suggérant un rôle concerté de l'acétylation et de la méthylation d'histones dans la mise sous silence de gènes.

Toutefois, il faut noter que les mécanismes génétiques et épigénétiques peuvent avoir des interactions (Figure 17). Par exemple la réactivation de transposons par les voies de la méthylation et des ARNi conduisent à des délétions et réarrangements chromosomiques. Les délétions de séquence peuvent altérer les gènes 'dose-dépendants' et la structure de la chromatine (Chen et Ni, 2006). Une étude récente de Nasrallah *et al.* (2007), montre que les changements de méthylation ont par exemple un rôle dans l'évolution de l'auto-incompatibilité chez des allotétraploïdes récents d'*Arabidopsis*. Les auteurs ont ainsi montré que l'auto-fertilité des hybrides résultait de changements épigénétiques dans l'expression des gènes du locus S qui détermine la réponse SI<sup>42</sup>. L'hybridation interspécifique entraîne un non-fonctionnement du SI générant ainsi, à partir d'espèces auto-incompatibles, des hybrides auto-fertiles. Il est important de noter que l'autogamie est un système de reproduction particulièrement favorable pour maintenir l'intégrité des deux sous-génomes durant les premières générations des allopolyploïdes.

---

<sup>42</sup> Le système de Self-incompatibility (SI) est défini comme une barrière génétique empêchant la pollinisation pré-zygotique (Nasrallah *et al.*, 2007).

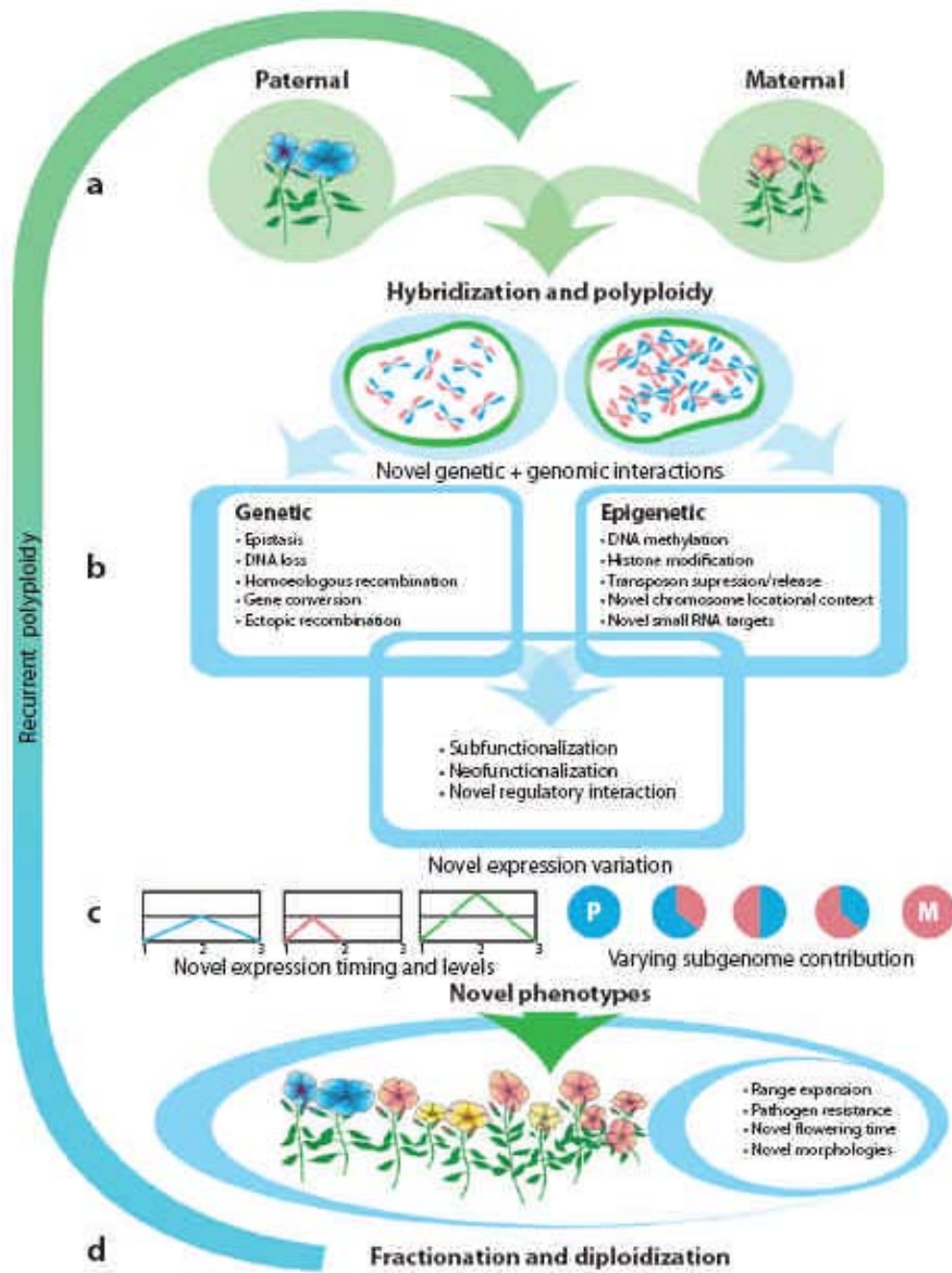
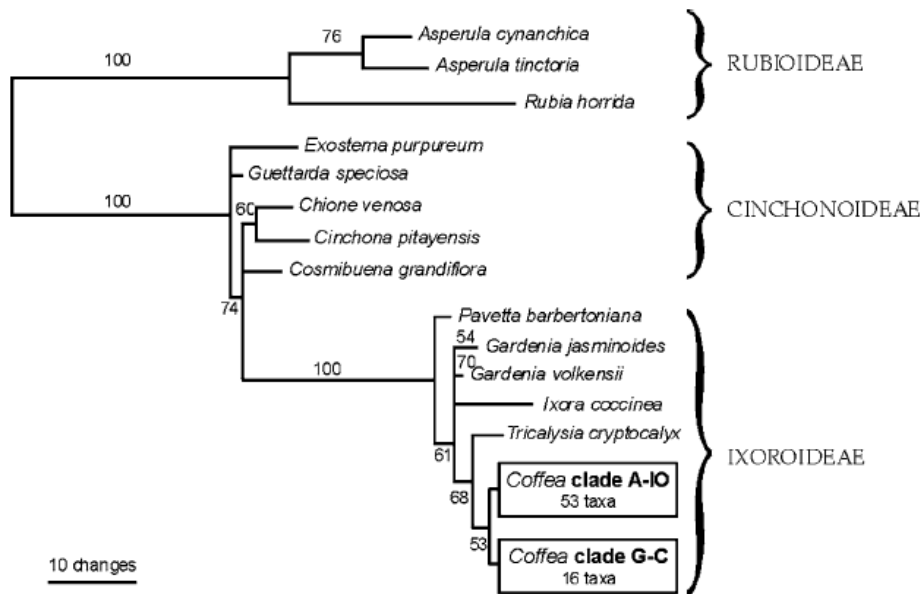


Figure 18 : Effets des facteurs génétiques et épigénétiques sur le phénotype et l'évolution des polyploïdes.

Doyle et al., 2008

### I.3.3 Evolution des espèces polyploïdes

Une étude de Ha *et al.* (2007), sur l'expression des gènes dupliqués d'*Arabidopsis thaliana* a mis en évidence que la divergence entre les gènes homéologues est plus rapide sous les stress environnementaux (facteurs externes) que sous l'action des processus de développement (Facteurs internes). La divergence rapide entre les gènes dupliqués en réponse à stress biotiques et abiotiques faciliteraient la sous-fonctionnalisation, la néofonctionnalisation et un mécanisme d'adaptation aux changements environnementaux (Ha *et al.*, 2007). Associé à cette divergence rapide, il faut prendre en compte que les divers changements génétiques et épigénétiques entraînent des modifications du génome par accumulation de mutations qui conduit l'allopolyploïde à revenir à un état diploïde (Figure 18 ; Doyle *et al.*, 2008). L'intérêt de ces cycles d'allopolyploïdisation-diploïdisation est que l'espèce allopolyploïde disposerait de nouvelles sources de diversité.



**Figure 19 : Classification des caféiers.**

Les caféiers (*Coffea* clades A-IO et G-C) appartiennent à la sous-famille des Ixoroideae.

Anthony et al., 2010



**Figure 20 : *Coffea* et *Psilanthus*.**

A. 1 et 2. Morphologie des fleurs du genre *Coffea*. B. 1 et 2. Morphologie des fleurs du genre *Psilanthus*.



## II. COFFEA ARABICA : MODELE BIOLOGIQUE D'ETUDE

### II.1 LE GENRE COFFEA

#### II.1.1 Classification

Les espèces de caféiers sont affiliées à l'une des plus grandes familles tropicales d'angiospermes, la famille des *Rubiaceae*<sup>43</sup>. Les *Coffeae* appartiennent à la sous-famille monophylétique des *Ixoroideae*, proche des *Gardenieae*, *Pavetteae* et *Vanguerieae* (Figure 19 ; Bremer et Jansen, 1991 ; Bremer et Eriksson, 2009). En 1934, seulement quatre espèces de caféiers étaient connues (Chevalier, 1929b ; Anthony et al., 2007). Depuis, plus d'une centaine d'espèces de caféiers ont été identifiées et de nouveaux taxons continuent à être découverts (Lashermes et al., 2008). Malgré cela, un grand nombre de taxons n'a pas encore été caractérisé et il est donc difficile de confirmer les relations entre les espèces (Anthony et Lashermes, 2006).

Dans les années 80, la classe *Coffeae* est divisée en deux genres, *Coffea* l. et *Psilanthus* hook f.. Cette distinction est basée sur la floraison et les caractéristiques florales (Leroy, 1980 ; Bridson, 1982 et 1988 ; Robbrecht and Puff, 1986 ; Robbrecht, 1988a et 1994). Les espèces appartenant au genre *Coffea* sont natives des forêts intertropicales d'Afrique, de Madagascar et des Mascareignes, alors que les espèces appartenant au genre *Psilanthus* sont originaires d'Afrique, d'Inde, de Malaisie, de la Nouvelle Guinée et d'Australie (Anthony et al., 2007). Ces espèces sont pérennes et peuvent être soit sous forme buissonnante soit sous forme arborescente (Lashermes et al., 2008).

Les deux genres, *Coffea* et *Psilanthus*, sont eux mêmes subdivisés en deux sous genres sur la base du type de croissance (monopodial ou sympodial), le type d'inflorescence (fleurs en position axillaires ou terminales) et la morphologie de la fleur (Figure 20). En ce qui concerne le genre *Coffea*, on distingue le sous genre *Coffea* avec 95 espèces et le sous genre *Baracoffea* avec 8 espèces (Davis, 2006). Le sous genre *Baracoffea* est restreint aux forêts saisonnièrement sèches à feuilles caduques et aux brousses situées à l'ouest de Madagascar, dans le nord-est du Kenya et dans le sud-est de la Somalie (Maurin, 2007). Le sous genre *Coffea* est, quant à lui, présent dans les forêts humides à feuilles persistantes où l'on retrouve les trois espèces cultivées pour la production de café, *Coffea arabica* (le café Arabica), *Coffea*

<sup>43</sup> La famille des Rubiaceae est divisée en 3 sous familles : Cinchonoideae, Ixoroideae et Rubioideae (Bremer et al., 2009).



**Figure 21 : Dispersion du genre *Coffea* sous genre *Coffea* à partir d'un centre d'origine en basse Guinée.**

*Anthony et al., 2010*

*canephora* (le café Robusta) et *Coffea liberica* (le café Libérien) (Maurin, 2007). En ce qui concerne le genre *Psilanthus*, il est également divisé en deux sous-genres, *Psilanthus* (2 espèces) qui est restreint à l'Afrique tropicale de l'ouest et centrale et *Afrocoffea* (20 espèces) que l'on retrouve en Afrique tropicale, au sud et sud-est de l'Asie et dans le nord de l'Australie tropicale (Maurin, 2007).

### II.1.2 Relations phylogénétiques

Les relations phylogénétiques entre les espèces de caféiers ont été à l'origine de nombreuses études utilisant des données de systématique de plusieurs sources incluant la morphologie (Stoffelen, 1998 ; Davis *et al.*, 2005), le RAPD (Lashermes *et al.*, 1993), des séquences provenant d'ADN plastidial (Cros, 1994 ; Lashermes *et al.*, 1996 ; Cros *et al.*, 1998) et des séquences ITS provenant d'ADNr nucléaire (Lashermes *et al.*, 1997).

A partir de ces premières analyses phylogénétiques, il a été observé qu'il y avait très peu de différences entre les espèces de caféiers et que le genre *Psilanthus* ne se distinguait pas du genre *Coffea*. Des études récentes (Davis *et al.*, 2006 and 2007 ; Maurin *et al.*, 2007) ont confirmé le fait que la plupart des genres appartenant au *Coffeae* sont monophylétiques à l'exception du genre *Coffea* identifié comme paraphylétique relativement à *Psilanthus* (Davis *et al.*, 2006 and 2007 ; Maurin *et al.*, 2007). Le faible niveau de divergence entre les séquences associées aux diverses observations morphologiques et moléculaires montre que *Coffea* et *Psilanthus* seraient compris dans le même clade et que cet argument confirme l'hypothèse d'un genre unique<sup>44</sup> (Lashermes *et al.*, 1997 ; Cros *et al.*, 1998 ; Davis *et al.*, 2007 ; Maurin *et al.*, 2007).

Les premières analyses phylogénétiques ont également suggéré un mode de spéciation radiale, une origine récente du genre *Coffea* en Afrique et une division du genre en quatre groupes majeurs en fonction d'une correspondance géographique forte : Madagascar, East Africa, Central Africa and West-central Africa. Une étude récente d'Anthony *et al.* (2010), a montré que la diversification du genre *Coffea* sous-genre *Coffea* aurait eu lieu dans la seconde moitié du Pleistocène moyen (460000–100000 BP<sup>45</sup>) et que le centre d'origine du genre se trouverait probablement en basse Guinée (Figure 21). Les auteurs ont également confirmé la division du genre en groupes à partir des origines biogéographiques<sup>46</sup> des espèces.

<sup>44</sup> Pour confirmer cette hypothèse, il serait, toutefois, nécessaire d'élargir les recherches ainsi que le nombre d'échantillonnages effectué au sein de la classe des *Coffeae*.

<sup>45</sup> Cet intervalle a été défini sur la base des différentes régions des ADNcp considérées (Anthony *et al.*, 2010).

<sup>46</sup> Dans cette étude menée sur des espèces Africaines, le genre *Coffea* sous-genre *Coffea* est divisé en deux clades, le clade A-IO qui couvre la totalité de la zone géographique du sous genre et le clade G-C (Guinéo-congolais) qui se restreint uniquement à la région Guinéo-congolaise. Ces clades sont divisés en cinq sous clades, EA (Est Africain), E-CA (Centre Est Africain), C (Congolais), LG (basse Guinée), et UG (Haute Guinée).

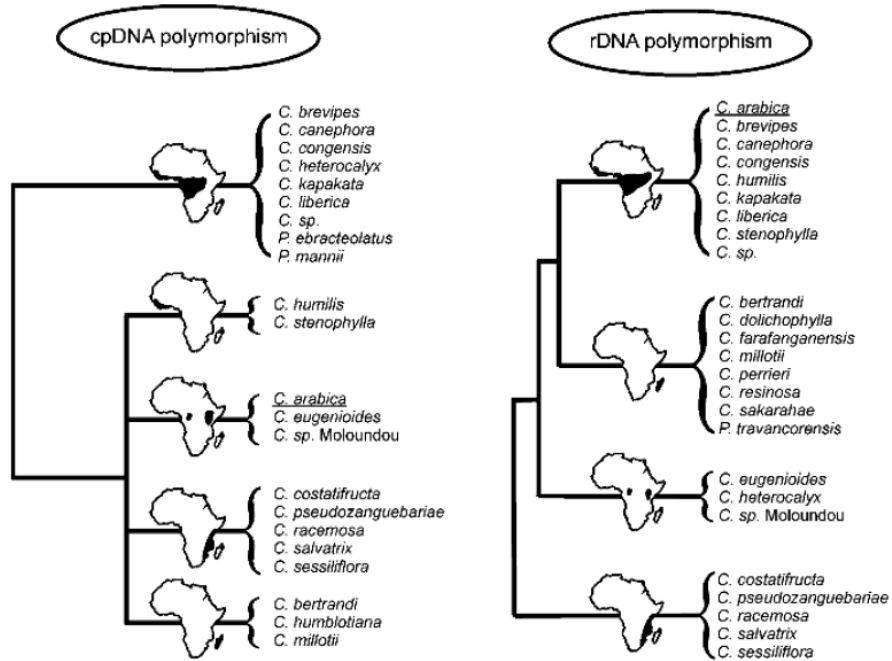


Figure 22 : Relation phylogénétique des espèces de caféiers basées sur des variations de séquences ADN chloroplastiques (cpDNA polymorphism) et ribosomales (rDNA polymorphism).

Lashermes et al., 1999

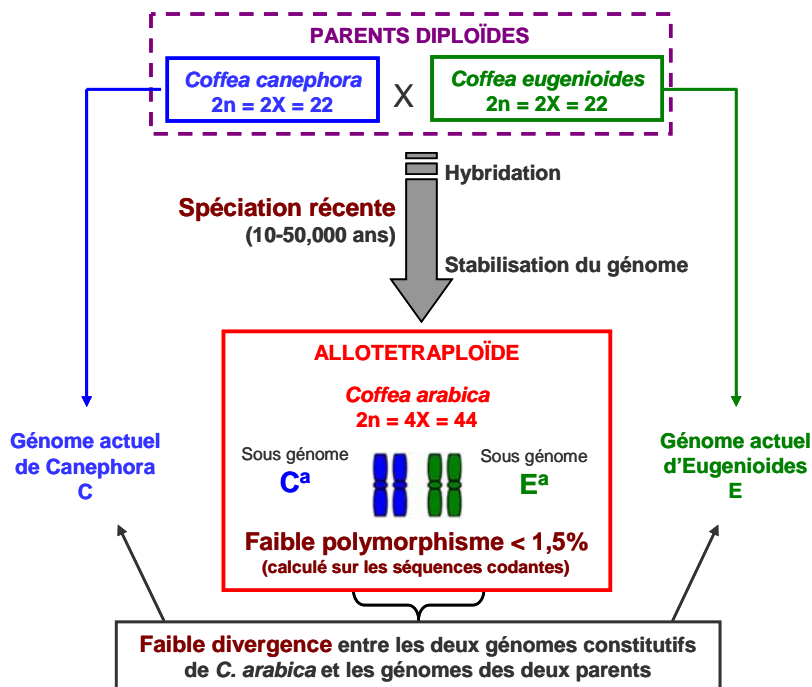


Figure 23 : Origine de la seule espèce allotétraploïde *Coffea arabica*.

D'après Lashermes et al., 1999

## II.2 LE CAS DE L'UNIQUE ALLOTETRAPLOÏDE, *COFFEA ARABICA* : ORIGINES ET CARACTERISTIQUES.

### II.2.1 Origine de *Coffea arabica*

L'utilisation de marqueurs chloroplastiques et ribosomiques (Lashermes *et al.*, 1996, 1997 et 1999) a mis en évidence une forte affinité de l'espèce tétraploïde *C. arabica* avec des espèces diploïdes originaires d'Afrique centrale et de l'ouest. Les analyses d'ADNcp ne montrent pas de différences entre l'ADNcp de *C. arabica* et celui de certaines accessions diploïdes, *C. eugenioides* et *C. Anthonyi* (Figure 22 ; Lashermes *et al.*, 1996). De plus une forte homologie est observée dans les séquences de la région ITS2 de *C. arabica* et des espèces du groupe des « canephoroides », *C. brevipes*, *C. canephora* et *C. congensis* (Figure 22 ; Lashermes *et al.*, 1997). Une étude plus approfondie appuyée sur des analyses RFLP<sup>47</sup> et GISH<sup>48</sup> a permis de démontrer la présence de deux lots de chromosomes E<sup>a</sup> et C<sup>a</sup>, qui proviendraient respectivement de formes ancestrales des espèces actuelles *C. eugenioides* (parent femelle, génome E<sup>a</sup>) et *C. canephora* (parent mâle, génome C<sup>a</sup>) ou d'écotypes proches de ces espèces diploïdes (Figure 23 ; Raina *et al.*, 1998 ; Lashermes *et al.*, 1999).

De nombreuses études se sont intéressées au degré de divergence de *C. arabica*. Comme le souligne Lashermes *et al.* (1999), une faible divergence est observée entre les deux génomes constitutifs de *C. arabica* (E<sup>a</sup> C<sup>a</sup>) et entre *C. arabica* et ses parents, suggérant une spéciation récente de l'espèce (Figure 23). Une étude récente de Cenci *et al.* (2011), a estimé une divergence moyenne de  $1,30\% \pm 0,67$ <sup>49</sup> observé dans les régions codantes, à partir de la comparaison entre le génome *C. canephora* (C) et celui de *C. eugenioides* constitutif de *C. arabica* (E<sup>a</sup>). Sur la base des fréquences de substitution des nucléotides, les auteurs ont également déterminé que la spéciation de *C. arabica* avait eu lieu relativement récemment (10-50000 BP).

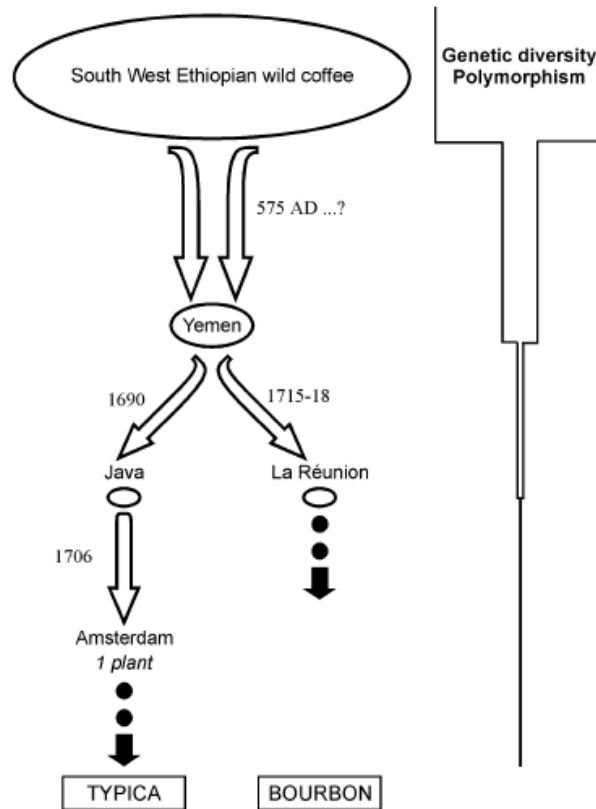
### II.2.2 Caractéristiques du génome et système de reproduction de *C. arabica*

Les espèces diploïdes ont un génome de base commun peu différencié comprenant 11 chromosomes (Bouharmont, 1963 ; Charrier 1978) et présentant une taille moyenne réduite (1 à 3µm). Chez les caféiers, toutes les espèces sont diploïdes (2n=2x=22) à l'exception de *C.*

<sup>47</sup> Utilisation de sondes spécifiques de locus.

<sup>48</sup> Hybridation de l'ADN génomique des espèces potentielles avec des chromosomes de *C. arabica*.

<sup>49</sup> Ce résultat obtenu en comparant le génome *C. canephora* (C) et celui de *C. eugenioides* constitutif de *C. arabica* (E<sup>a</sup>) correspond à la divergence maximale observable entre *C. arabica* et ses parents diploïdes.



**Figure 24 : Origine des deux variétés mères de *Coffea arabica* à l'origine des variétés cultivées actuelles.**

*Anthony et al., 2002*

*arabica* qui est tétraploïde ( $2n=4x=44$ ) (Charrier and Berthaud, 1985). La plupart des caféiers sont auto-incompatibles à l'exception de trois espèces auto-fertiles : *C. heterocalyx* Stoff. (Coulibaly *et al.*, 2003), *C. anthonyi* Stoff. & F. anthony (Stoffelen *et al.*, 2009) et l'unique espèce tétraploïde, *C. arabica*.

Chez *C. arabica*, il a été observé un mode d'hérédité de type disomique par analyse des ségrégations au sein d'une population F2 (Lashermes *et al.*, 2000). *Coffea arabica* possède un comportement méiotique comparable aux espèces diploïdes. A la méiose, les chromosomes homologues s'appariaient de façon systématique pour former des bivalents. Aucun des chromosomes homéologues de *C. arabica* ne s'apparie et cela est probablement dû au fonctionnement de facteurs génétiques régulant l'appariement (Lashermes *et al.*, 2000).

La cytométrie en flux a été utilisée pour estimer la quantité d'ADN nucléaire chez une quinzaine d'espèces de *Coffea* originaires d'Afrique (Cros *et al.*, en 1995 ; Noirot *et al.*, 2003). Chez les espèces diploïdes, les quantités en 2C d'ADN nucléaire sont très variables avec des valeurs comprises entre 0,95 et 1,78 pg par noyau. Ces variations observées seraient dues aux variations du nombre de copies de séquences d'ADN répétées (Lashermes *et al.*, 2008). En ce qui concerne le tétraploïde, *C. arabica*, sa quantité d'ADN a été estimée deux fois supérieure (2,61 pg) aux quantités des espèces diploïdes (Cros *et al.*, 1998). Sur la base des résultats obtenus en cytométrie en flux (Cros *et al.*, en 1995), la moyenne du génome haploïde des caféiers a été estimée à 700Mb. En ce qui concerne le seul génome dihaploïde, sa taille a été estimée à 1300Mb (Noir *et al.*, 2004). Comparé aux autres espèces de plantes, le génome des caféiers apparaît relativement de petite taille (Anthony and Lashermes, 2006).

### II.2.3 Ressources génétiques et amélioration chez *Coffea arabica*

Les espèces sauvages du centre de diversité de *C. arabica*, les espèces diploïdes et les hybrides spontanés constituent les principales ressources génétiques formant un réservoir de diversité utilisable pour l'amélioration de l'espèce *C. arabica*.

Le centre primaire de diversification génétique de *C. arabica* se situe en Ethiopie (Zone montagneuse du Sud-ouest) et au Soudan (Plateau du Boma). Chez *C. arabica*, les individus sont divisés en deux groupes, les caféiers sauvages<sup>50</sup> et les caféiers cultivés<sup>51</sup>. Les variétés actuelles de caféiers cultivés dérivent toutes de deux populations, *C. arabica* var. *Typica* et *C. arabica* var. *Bourbon* (Figure 24 ; Krug *et al.*, 1939). Peu de plantes sont à l'origine de ces

---

<sup>50</sup> Des espèces sauvages ont également été observées au Soudan (Mont Imatong) et au Kenya (Mont Marsabit) (Thomas 1942 ; Anthony *et al.*, 1987).

<sup>51</sup> Sa culture a débuté dans le Sud-ouest de l'Ethiopie il y aurait environ 1500 ans (Wellman, 1961).

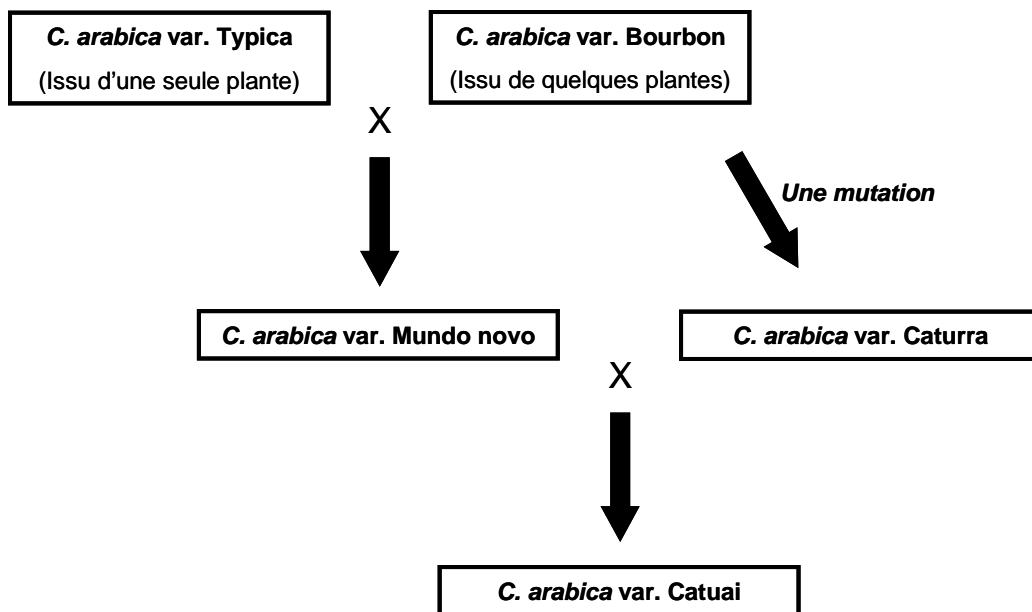


Figure 25 : Origine des trois variétés principales actuelles les plus cultivées.

*D'après Bertrand et al., 1999*

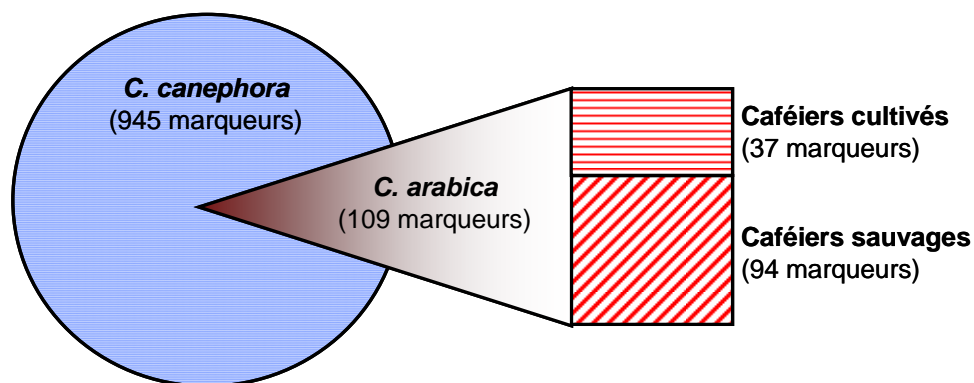


Figure 26 : Diversité observée dans des populations de caféiers sauvages et cultivés chez l'allotétraploïde *Coffea arabica*.

Cette diversité est estimée à partir du nombre de marqueurs polymorphes.

*D'après Lashermes et al., 2000 et Anthony et al., 2002*



deux populations. A partir de l'étroite base génétique de *C. arabica*, trois principales variétés (var. Caturra, var. Catuai et var. Mondo Novo) à bonne production et de bonne qualité ont largement été diffusées à travers le monde (Figure 25 ; Bertrand *et al.*, 1999). Cependant du fait de sa très faible diversité génétique, la caféiculture de *C. arabica* est particulièrement sensible à la plupart des maladies et espèces ravageuses<sup>52</sup>. Il a été observé que la diversité génétique chez *C. arabica* était faible entre les deux populations cultivées de base (Typica-Bourbon) mais que cette diversité était plus importante entre les caféiers cultivés et les caféiers sauvages (Figure 26 ; Lashermes *et al.*, 2000 ; Anthony *et al.*, 2002). Ces résultats montrent l'intérêt que peuvent avoir les caféiers sauvages dans l'amélioration de la base génétique des variétés cultivées.

Les espèces diploïdes présentent également un intérêt dans l'amélioration de *C. arabica*. Trois espèces ont montré des traits agronomiques pertinents. *C. canephora* est la source principale de caractères de résistance aux maladies et ravageurs dont la résistance à la rouille orangée, à l'antracnose des baies et aux nématodes (Bertrand *et al.*, 2003). *C. liberica* est une source de résistance à la rouille orangée (Srinivasan et Narasimhaswamy, 1975) et *C. racemosa* constitue une source de résistance contre la mineuse des feuilles (Guerreiro Filho *et al.*, 1999). Le croisement entre *C. arabica* et une espèce diploïde peut se produire de façon spontanée ou de façon artificielle. Lorsque *C. arabica* et une espèce diploïde se retrouvent dans une même zone de distribution, un hybride spontané peut se former. Plusieurs hybrides spontanés<sup>53</sup> issus de l'hybridation entre *C. arabica* et *C. canephora* ou *C. liberica* ont été répertoriés. On peut prendre l'exemple de l'hybride spontané entre *C. arabica* et *C. canephora*, appelé Hybride de Timor (HdT). L'Hybride de Timor a été identifié au sein d'une plantation sur l'île de Timor en Indonésie (Bettencourt, 1973). A partir de cet hybride, des rétrocroisements ont été effectués avec des variétés de *C. arabica*<sup>54</sup>. Ces descendants ont été largement diffusés dans le monde, notamment au Brésil, en Colombie ou au Kenya et constituent une source importante de gènes résistants aux maladies et ravageurs. Une étude de Lashermes *et al.*, en 2000, portant sur l'étude de marqueurs AFLP, a permis d'estimer un niveau d'introgression entre 8 et 27% du génome *C. canephora* dans les lignées dérivées de l'hybride HdT. Des hybrides artificiels ont également été créés entre *C. arabica* et

---

<sup>52</sup> La rouille orangée (*Hemileia vastatrix* Berk. and Br.), l'antracnose des baies (*Colletotrichum kahawae*), le scolyte des baies (*Hypothenemus hampei*) et les nématodes (*Meloidogyne spp.* et *Pratylenchus spp.*) (Lashermes, 2008).

<sup>53</sup> L'hybride de Timor (HdT) et Devamachy (répertorié en 1930 en Inde) issus du croisement entre *C. arabica* et *C. canephora*. Kalimas (répertorié en 1899 en Indonésie), Kawisari (répertorié en 1899 en Indonésie), S26 (répertorié en 1932 en Inde) et S288 (répertorié en 1932 en Inde) issus du croisement entre *C. arabica* et *C. liberica*.

<sup>54</sup> Des variétés ont été créées comme Catimor issu du croisement entre HdT et Caturra et Sarchimor issu de l'hybridation entre HdT et Villa Sarchi.

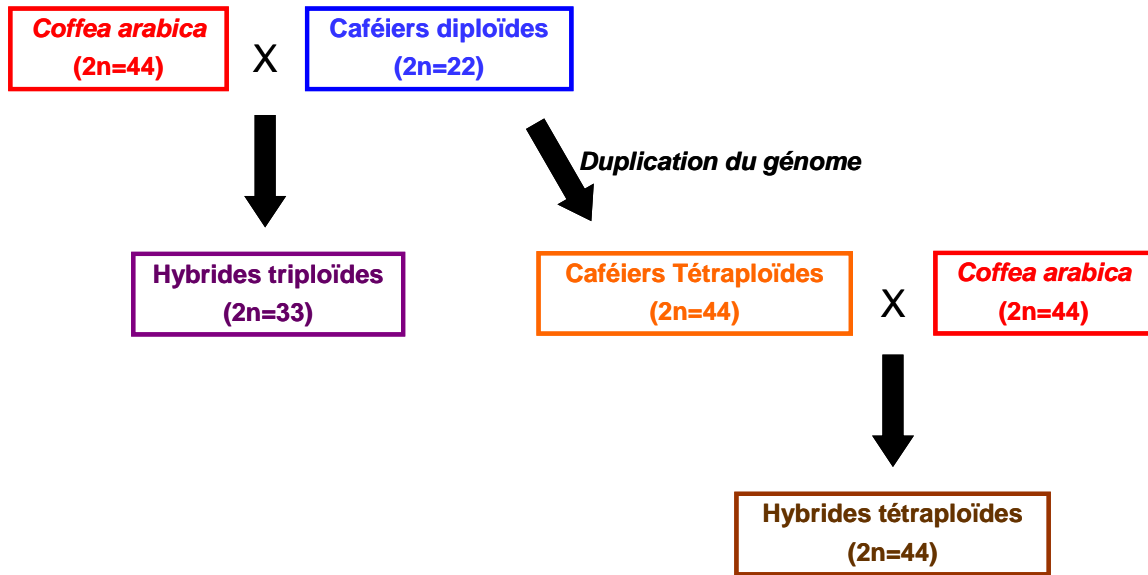


Figure 27 : Croisement à l'origine d'hybrides interspécifiques artificiels triploïdes et tétraploïdes de caféiers.

*D'après Lashermes et al., 2000*

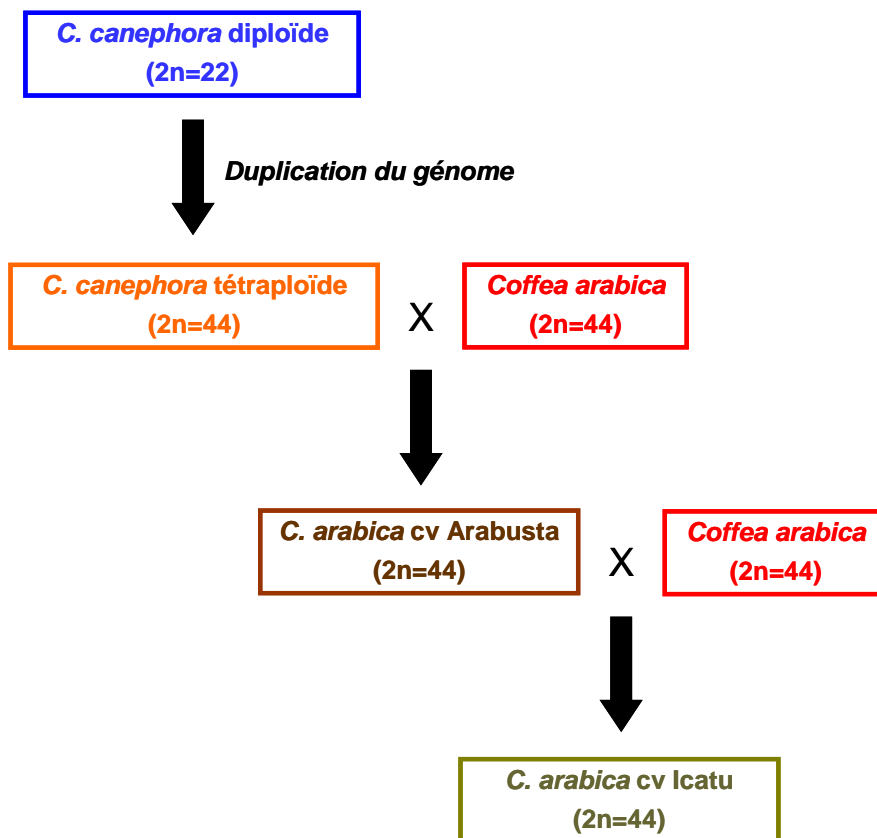


Figure 28 : Croisement à l'origine des variétés Arabusta et Icatu.

*D'après Bertrand et al., 1999*

l'espèce diploïde portant les gènes d'intérêt. Le descendant est ensuite rétrocroisé avec *C. arabica* et les gènes indésirables du parent diploïde sont progressivement éliminés par sélection. Cette procédure dite « conventionnelle » s'est avérée efficace pour l'amélioration de la production, la vigueur et la résistance aux maladies de *C. arabica*, tout en conservant les caractéristiques gustatives. Ainsi des hybrides interspécifiques artificiels triploïdes<sup>55</sup> et tétraploïdes<sup>56</sup> ont été créés (Figure 27). Chez les tétraploïdes, on peut prendre comme exemple la variété Arabusta issue d'un croisement entre *C. arabica* et *C. canephora* (Figure 24). Cette variété a été rétrocroisée avec l'espèce *C. arabica* formant ainsi la variété Icatu (Figure 28 ; Bertrand *et al.*, 1999). La production d'hybrides interspécifiques naturels ou artificiels a servi de relais dans le transfert de caractères d'intérêt agronomique par rétrocroisement avec *C. arabica* et produit de nombreuses lignées de variétés à grandes valeurs agronomiques et économiques (Bertrand *et al.*, 2003). Toutefois, il faut souligner que ces programmes d'hybridation interspécifique sont lents et coûteux<sup>57</sup>.

#### II.2.4 Les caféiers et leur milieu naturel

Les caféiers sont cultivés dans tous les pays d'Amérique du sud, d'Asie et d'Afrique situés dans la zone intertropicale humide<sup>58</sup>. Le confinement des caféiers dans cette zone est à des facteurs écologiques liés à la température et à l'humidité (Smith, 1989 ; Damatta, 2006). Les espèces de caféiers diffèrent suivant des adaptations morphologiques et écologiques. A l'exception des deux espèces à large distribution, *C. canephora* et *C. liberica* dont la distribution s'effectue de la Guinée à l'Ouganda ; les autres espèces de caféiers ont une zone de distribution plus ou moins restrictive liée à une adaptation plus spécifique, comme *C. congensis* situé dans des zones saisonnièrement inondées du bassin du Zaïre ou encore *C. racemosa* situé dans des zones très sèches de la région côtière du Mozambique (Anthony et Lashermes, 2006).

La distribution géographique de l'espèce allopolyploïde *C. arabica* et de ses espèces parentales (*C. canephora* et *C. eugenioides*) est très différente. A l'état sauvage, le territoire de *C. canephora* s'étend de l'Afrique Centrale, au golf de la Guinée et à l'Ouganda. Comparés au très vaste territoire occupé par *C. canephora*, les territoires de *C. arabica* et de *C. eugenioides* sont plus limités. Le territoire de *C. eugenioides* s'étend de l'Est au Sud-est de

<sup>55</sup> Les hybrides triploïdes (2n= 33) sont issus d'un croisement entre *C. arabica* et l'espèce diploïde d'intérêt.

<sup>56</sup> Les hybrides tétraploïdes (2n= 44) sont issus d'un croisement entre *C. arabica* et l'espèce diploïde d'intérêt après une duplication du nombre de chromosomes de ce dernier.

<sup>57</sup> Le temps de génération d'un caféier est de 5 ans et il faut au minimum 25 ans pour que l'hybride se stabilise génétiquement.

<sup>58</sup> Cette zone intertropicale va de 20-25°N à Hawaï à 24°S au Brésil.

l'Afrique (Ouganda, Tanzanie, Kenya, Rwanda et Burundi). *C. arabica*, quant à lui, est issu d'une région limitée du Sud-ouest de l'Ethiopie et aucune autre espèce de *Coffea* n'est présente dans cette zone. L'espèce polyploïde et ses deux espèces parentales se développent également à des altitudes et à des températures différentes. On retrouve l'espèce sauvage *C. arabica* dans des forêts tropicales de hautes altitudes, entre 1200-1950m (Davis *et al.*, 2006), avec un optimum de températures moyennes annuelles pour l'espèce de 18-21°C (Damatta & Ramalho, 2008). L'espèce sauvage *C. canephora* est issue de forêts équatoriales humides de basses à moyennes altitudes, entre 250-1500m (Davis *et al.*, 2006), avec un optimum de températures moyennes annuelles pour l'espèce de 22-26°C (Damatta & Ramalho, 2008). Enfin, l'espèce sauvage *C. eugenioides* se retrouve dans des forêts de hautes altitudes, entre 1000 et 2000m, avec un optimum de températures moyennes annuelles pour l'espèce de 18-23°C (Davis *et al.*, 2006). Comme la plupart des espèces tropicales, les caféiers sont particulièrement sensibles au froid (Bauer *et al.*, 1985; DaMatta *et al.*, 1997; Ramalho *et al.*, 2003, Fortunato *et al.*, 2009). De faibles températures limitent les zones géographiques de distribution des caféiers (Fortunato *et al.*, 2009). Comme le souligne Fortunato *et al.* (2009), chez les caféiers, des effets négatifs sur la croissance des plantes sont observés dans des zones où les températures moyennes mensuelles sont inférieures à 15-16°C (Barros *et al.*, 1999) et la photosynthèse est affectée sous 18°C (Ramalho *et al.*, 2003).

### II.2.5 L'adaptation des caféiers en conditions de culture

Deux espèces, *C. arabica* et *C. canephora*, sont cultivées dans toute la ceinture intertropicale. Les intervalles de températures des zones de culture de *C. arabica* et du *C. canephora* sont plus importants comparés aux températures des zones de diversification primaire ce qui témoignerait d'une certaine capacité adaptative de ces deux espèces. Il a été observé que pour des conditions de culture intensives, des variétés sélectionnées sont conduites à pousser dans des régions où les températures moyennes sont de 24 à 25°C (Damatta *et al.*, 2008). L'origine de cette capacité pourrait ne pas avoir la même origine entre les deux espèces. La culture de *C. canephora* est basée sur des variétés issues de populations présentant une diversité génétique importante. L'exploitation de la diversité génétique naturelle pourrait être à la base de l'adaptation variétale aux conditions environnementales et aux systèmes de culture. En revanche, la culture de *C. arabica* est basée sur une population à très faible base génétique. Par exemple, en Amérique Latine plus de 80% du verger, s'étendant d'Hawaï au Brésil, est constitué uniquement de deux variétés de *C. arabica* (Caturra et Catuai) issues d'une même base génétique très étroite. Pourtant *C. arabica* s'adapterait à des conditions de culture très variables. Comparé à ces espèces parentales, il semble montrer une certaine aptitude à tolérer des variations de température plus importantes.

### III. OBJECTIFS DE LA THESE

Ces dernières années, les conséquences de l'allopolyploïdie sur la structure et le fonctionnement des génomes ont fait l'objet de plusieurs études. Celles-ci ont mis en évidence que l'expression des génomes polyploïdes pouvait être additive par rapport aux profils d'expression des espèces parentales mais également que, dans plusieurs cas, les profils d'expression des gènes montraient un écart à l'additivité à l'origine de phénotypes nouveaux. Les génomes allopolyploïdes ne seraient pas dus à la simple addition des génomes parentaux. Les changements provoqués par l'hybridation et la duplication du génome contribueraient à une plus grande plasticité phénotypique et à de nouvelles capacités adaptatives des espèces nouvellement formées. Ces études mettent également en évidence une grande diversité de réponse des génomes parentaux face au phénomène de polyploïdisation. Cependant, celles-ci soulignent également les limites technologiques et les difficultés méthodologiques dans l'exploration et la compréhension des génomes allopolyploïdes.

Au cours de cette thèse, nous proposons d'observer et d'analyser le comportement d'une espèce allopolyploïde et d'établir sa plasticité face à différents régimes thermiques. Quatre objectifs majeurs ont été identifiés :

- ***Evaluer les technologies et les analyses statistiques employées pour l'étude du transcriptome.***

Le chapitre I et II traiteront de la mise au point d'une puce à ADN et de son utilisation dans l'analyse des transcriptomes de caféiers. Dans le chapitre II, la PCR quantitative en temps réel (qRT-PCR) sera utilisée dans le but d'analyser les niveaux d'expression de gènes spécifiques à une voie de biosynthèse et enfin dans le chapitre IV sera utilisée, pour la première fois chez le caféier, le séquençage à haut débit de fragments d'ARN (RNAseq). L'objectif sera d'évaluer si les résultats obtenus sont dépendants de la technologie et/ou sensibles aux analyses statistiques employées et dans quelle mesure.

- ***Etudier les variations entre le transcriptome d'un allopolyploïde et ceux des espèces parentales et établir si ces variations sont dépendantes des conditions environnementales.***

Le chapitre II traitera dans une première partie de la « divergence transcriptomique et de l'expression génétique non-additive chez l'allotétraploïde (*Coffea arabica*) et ses espèces parentales diploïdes (*Coffea canephora* et *Coffea eugenioides*) ». Ces variations seront observées dans deux conditions de température,

30-26°C et 26-22°C. Dans une seconde partie, les variations du transcriptome seront étudiées suivant une contrainte thermique imposée. La condition de température 30-26°C sera considérée comme un stress thermique chaud pour les trois espèces étudiées. La condition de température 26-22°C sera utilisée comme un témoin tout en sachant qu'il s'agit d'une condition de température adéquate pour les trois espèces.

- ***Etablir la plasticité des espèces allopolyploïdes face à différents régimes thermiques.***

Le chapitre III traitera de la « plasticité des allopolyploïdes face à différents régimes thermiques en comparant l'allopolyploïde *C. arabica* à ses deux espèces parentales diploïdes (*C. canephora* et *C. eugenioides*) ». Les caféiers sont des espèces tropicales sensibles aux fortes variations de température. La température est donc un enjeu majeur dans l'adaptation des caféiers à des changements climatiques globaux. Cette plasticité sera observée au cours de différents régimes thermiques contrastés, 18-15°C, 23-20°C, 28-25°C et 33-30°C. Les changements de composition en acides gras jouent un rôle dans la fluidité membranaire et les variations observées dans le degré d'insaturation des acides gras sont considérées comme l'un des facteurs majeurs dans l'adaptation métabolique des plantes tempérées aux stress thermiques. Les études portant sur la sensibilité des plantes à différentes conditions de température contrastées se sont intéressées aux lipides et à leurs effets sur le fonctionnement des membranes cellulaires. Nous étudierons plus particulièrement cette voie de biosynthèse chez les trois espèces tropicales.

- ***Comparer la plasticité observée chez les allopolyploïdes et la diversité de leurs espèces parentales.***

Le chapitre IV traitera de la « comparaison du transcriptome de *C. arabica* avec les transcriptomes de trois groupes génétiques de *C. canephora* ». Cette étude basée sur la technologie RNAseq, nous permettra de déterminer si les variations de divergence transcriptomique entre *C. arabica* et *C. canephora* observées sur la base d'un seul écotype de *C. canephora* à partir de la puce à ADN du café, s'observent en élargissant cette étude à d'autres écotypes de *C. canephora* adaptés à des zones écologiques différentes.

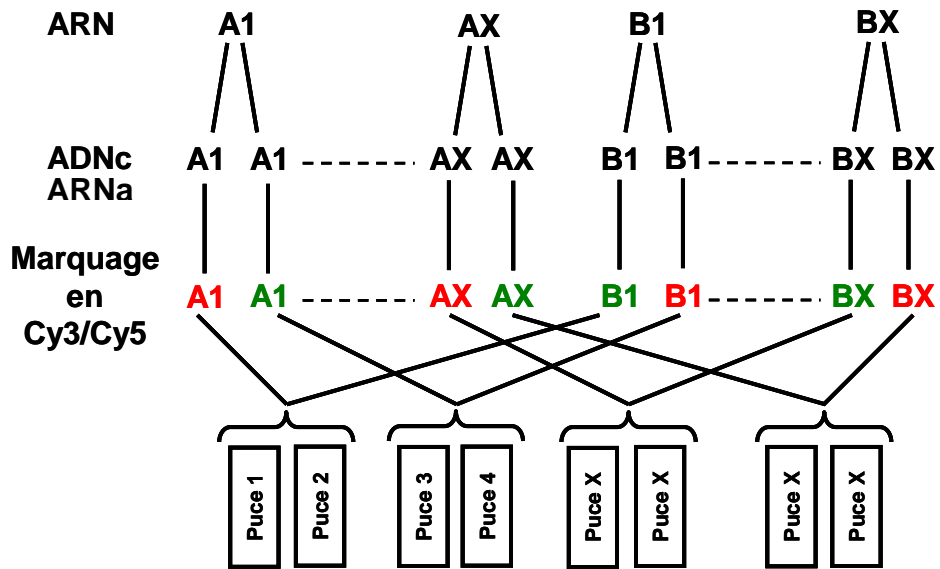
# CHAPITRE I

## MISE AU POINT DE LA PREMIERE PUCE A ADN CHEZ LE CAFE

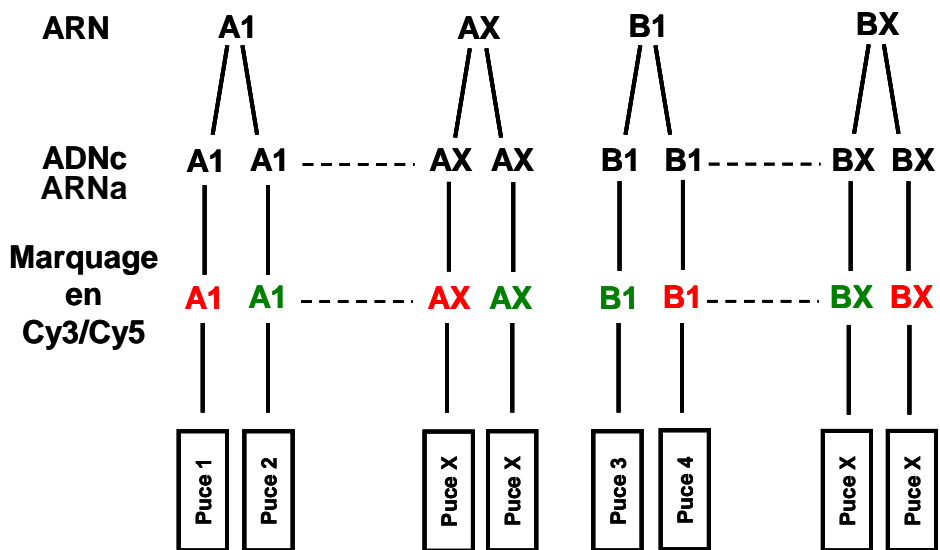
Ce chapitre a fait l'objet d'une publication:

**Privat I, Bardil A, Bombarely Gomez A, Severac D, Dantec C, Fuentes I, Mueller L, Joët T, Pot D, Foucrier S, Dussert S, Leroy T, Journot L, de Kochko A, Campa C, Combes M-C, Lashermes P, Bertrand B. 2011.** The 'PUCE CAFÉ' project: the first 15K coffee microarray, a new tool for discovering candidate genes correlated to agronomic and quality traits. *BioMed central Genomics* **12**: 5. doi: 10.1186/1471-2164-12-5

## PUCES ADNc et oligonucléotides



## PUCES Affymetrix



**Figure 1 : Représentation schématique des deux types de puces.**

Cette figure représente le plan expérimental des puces ADNc/oligonucléotides et des puces Affymetrix. En rouge et vert sont représentés les marquages de fluorescence en Cy3 et Cy5. Les puces ADNc et oligonucléotides sont des puces dites de « deux couleurs » où deux échantillons d'ADNc marqués sont co-hybridés sur la même puce alors que les puces « Affymetrix genechip » sont des puces dites d'« une couleur » où un seul échantillon est hybridé sur chacune des puces.



## INTRODUCTION AU CHAPITRE

Ces dernières années, le rôle de l'allopolyploïdie dans les variations d'expression génique a fait l'objet de nombreuses études. La majorité de ces études traite des variations d'expression au niveau transcriptomique. Différentes techniques permettant d'étudier l'expression des gènes ont été développées. En utilisant des techniques telles que les Northern blot ou les PCR quantitative en temps réel (qRT-PCR), l'analyse de l'expression ne se fait que pour un seul gène à la fois ce qui ne permet pas d'avoir une vue d'ensemble des profils d'expression des gènes d'un individu. Ces dernières années, le développement de Pucés à ADN a permis d'analyser, de façon comparative, les transcriptomes. Elles permettent une mesure simultanée des niveaux d'expression de milliers de gènes répartis sur l'ensemble du génome. Les analyses transcriptomiques par Pucés à ADN correspondent à des hybridations d'ADNc/ARNa présents dans différents tissus, différents stades de développement ou soumis à différents traitements déposés sur lames. Les premières Pucés fabriquées, appelées macropuces, utilisent la radioactivité pour marquer les sondes s'hybridant avec les molécules d'ADNc. Cependant deux inconvénients ont été identifiés, l'utilisation de la radioactivité et le nombre limité de gènes à étudier. Pour pallier cela, d'autres types de Pucés, les micropuces, ont été développées. Deux types de micropuces peuvent être utilisés, les Pucés ADNc/oligonucléotides et les Pucés affymetrix (Figure 1).

Les Pucés ADNc et oligonucléotides sont des Pucés dites de « deux couleurs » où deux échantillons d'ADNc marqués sont co-hybridés sur la même Puce (Figure 1). On peut prendre en exemple deux études, celle de Wang *et al.* (2006) qui étudient la divergence globale entre les espèces polyploïdes et leurs espèces parentales diploïdes et celle d'Udall *et al.* (2006) qui étudie la comparaison de l'expression de copies homéologues chez les espèces allopolyploïdes. Wang *et al.* (2006) ont développé une Puce à oligonucléotides longs (70-mer) où un marquage en Cy3- et Cy5-dCTP<sup>59</sup> a été utilisé. 192 paires d'ADNc d'*Arabidopsis thaliana* ont été utilisées et deux dye-swaps (même quantité en ADNc marquée en Cy3 et Cy5) ont été effectués pour chacun des réplicats biologiques. Udall *et al.* (2006) ont créé une Puce à oligonucléotides courts basée sur des SNPs spécifiques du génome D et A de *Gossypium*. Des ESTs de trois espèces de *Gossypium*, dont un allopolyploïde (*G. gossypium*) et ses deux espèces parentales (*G. arboreum* et *G. raimondii*), ont été utilisés. A contrario, les Pucés « Affymetrix genechip » sont des Pucés dites d'« une couleur » où un seul échantillon

---

<sup>59</sup> Ce sont deux fluorochromes Cyanines : Cy5 (émission de fluorescence à une longueur d'onde de 662 - 670 nm, couleur rouge) et Cy3 (émission de fluorescence à une longueur d'onde de 563 - 570 nm, couleur vert). Ils permettent de comparer les niveaux d'expression relatifs de deux transcriptomes différents sur une seule puce.

est hybridé sur chacune des Puces (Figure 1). On peut prendre en exemple la Puce de Chagué *et al.* (2010) où différents génotypes de *Triticum* sont étudiés.

Ces Puces peuvent également être utilisées de façon « hétérologue ». Elles sont développées pour une espèce donnée mais utilisées pour une autre. Ainouche *et al.* (2009) utilisent une Puce dessinée pour *Oryza* (44 K rice Agilent array (Agilent G2519F)) afin d'étudier le transcriptome d'une espèce polyploïde invasive, *Spartina anglica*. Les auteurs soulignent le fait que 70% des 21510 gènes d'*Oryza* ont été hybridés par quatre individus *Spartina*. Ces proportions seraient en accord avec les hauts niveaux de conservation des séquences codantes (environ 90% d'identité) estimés entre *Oryza* et *Spartina*. Enfin l'utilisation des Puces peuvent se faire à partir de sondes anonymes d'ADNc pour comparer les transcriptomes, comme chez *Senecio*, des parents, hybrides et allopolyploïdes (Hegarty *et al.* 2005, 2006, 2008). Les transcrits montrant une expression différentielle peuvent être séquencés et identifiés a posteriori.

Ce premier chapitre sera consacré à la mise au point de la première Puce à ADN chez le café. Cette Puce a été produite à partir de séquences d'EST de café publiées et provenant essentiellement des différentes étapes du développement du fruit et des feuilles de *Coffea canephora* cv. Robusta. La Puce café a été développée sur une Puce à oligonucléotides longs à haute densité sur la base de 15721 unigènes dérivant des séquences d'ESTs de café publiées. Des échantillons de deux expériences indépendantes ont été hybridés sur la PuceCafé. La première expérience permet la comparaison des profils d'expression génique chez trois tissus de *Coffea arabica* cv. Caturra (graines, feuilles et fleurs). La seconde expérience permet la comparaison des profils d'expression génique chez des feuilles de trois espèces de caféiers, l'allotétraploïde (*Coffea arabica*) et ses deux espèces parentales diploïdes (*Coffea eugenoides* et *Coffea canephora*). Des analyses statistiques et une validation par qRT-PCR ont été effectuées.

METHODOLOGY ARTICLE

Open Access

# The 'PUCE CAFE' Project: the First 15K Coffee Microarray, a New Tool for Discovering Candidate Genes correlated to Agronomic and Quality Traits

Isabelle Privat<sup>1\*</sup>, Amélie Bardil<sup>4</sup>, Aureliano Bombarely Gomez<sup>3</sup>, Dany Severac<sup>2</sup>, Christelle Dantec<sup>2</sup>, Ivanna Fuentes<sup>2</sup>, Lukas Mueller<sup>3</sup>, Thierry Joët<sup>5</sup>, David Pot<sup>6</sup>, Séverine Foucrier<sup>1</sup>, Stéphane Dussert<sup>5</sup>, Thierry Leroy<sup>5</sup>, Laurent Journot<sup>2</sup>, Alexandre de Kochko<sup>5</sup>, Claudine Campa<sup>5</sup>, Marie-Christine Combes<sup>4</sup>, Philippe Lashermes<sup>4</sup>, Benoit Bertrand<sup>4</sup>

## Abstract

**Background:** Understanding the genetic elements that contribute to key aspects of coffee biology will have an impact on future agronomical improvements for this economically important tree. During the past years, EST collections were generated in Coffee, opening the possibility to create new tools for functional genomics.

**Results:** The "PUCE CAFE" Project, organized by the scientific consortium NESTLE/IRD/CIRAD, has developed an oligo-based microarray using 15,721 unigenes derived from published coffee EST sequences mostly obtained from different stages of fruit development and leaves in *Coffea Canephora* (Robusta). Hybridizations for two independent experiments served to compare global gene expression profiles in three types of tissue matter (mature beans, leaves and flowers) in *C. canephora* as well as in the leaves of three different coffee species (*C. canephora*, *C. eugenioides* and *C. arabica*). Microarray construction, statistical analyses and validation by Q-PCR analysis are presented in this study.

**Conclusion:** We have generated the first 15 K coffee array during this PUCE CAFE project, granted by Génoplante (the French consortium for plant genomics). This new tool will help study functional genomics in a wide range of experiments on various plant tissues, such as analyzing bean maturation or resistance to pathogens or drought. Furthermore, the use of this array has proven to be valid in different coffee species (diploid or tetraploid), drastically enlarging its impact for high-throughput gene expression in the community of coffee research.

## Background

In recent years, microarray technology has demonstrated the power of the high-throughput study of gene expression in unravelling key processes in plant biology [1-3]. Microarrays have become especially relevant for species where little genome information is available and where intensive laboratory work is necessary to gain insight into a particular biological process, as well as to identify candidate target genes for future breeding programs [4,5].

The genus *Coffea* (Rubiaceae family) encompasses approximately 100 species, all of which are native to the African continent, Madagascar and the Mascarene

Islands [6]. Two of these species *Coffea canephora* (robusta) and *Coffea arabica*, are widely cultivated for the production of coffee beverages. The former is diploid ( $2n = 2 \times = 22$ ) and allogamous, the latter, allotetraploid ( $2n = 4 \times = 44$ ) and preferentially autogamous. Approximately 60% of the world coffee production comes from *C. arabica* versus 40% for *C. canephora*. In terms of cup quality, consumers appreciate *C. arabica* (Arabica) more due to its taste, which is less bitter and more flavourful compared with *C. canephora* (Robusta). While it is not widely known, coffee is one of the most valuable international exchange commodities in agricultural trade. This is reflected in the fact that raw coffee values rank fourth on the international stock market only after wheat, sugar, and soya [7]. Furthermore, over 25 million people worldwide are linked to coffee cultivation and processing. Despite these economic aspects,

\* Correspondence: isabelle.privat@rdto.nestle.com

<sup>1</sup>Nestlé R&D Tours, 101 Avenue Gustave Eiffel, Notre Dame d'Oé, BP 49716, 37097 Tours Cedex 2, France

Full list of author information is available at the end of the article

coffee research suffers from a lack of both scientific and financial investment. Also, coffee is a perennial plant which only begins to bear seed after about five years, which makes genetic studies more complicated and time-consuming. While some genomic information is publicly available for coffee (e.g., an expressed sequence tag (EST) database), it lags far behind what is available for many other agricultural species. As a result, coffee researchers have only limited access to the plethora of genomic resources available for most major crop species.

During the past few years, aiming to develop genomic tools to assist future coffee research, various scientific groups have produced large scale sets of *Coffea* EST sequences. However, the number of publicly available ESTs remains low because many of the sequences discovered fall under the domain of private property and are not published. At the time when the PUCE CAFE Project began, two large coffee EST databases were available: the NESTLE/Cornell and IRD databases with respectively 62,877 and 8782 sequences. Those sequences were mainly cDNA derived from leaves, fruit (whole cherries), pericarp and beans at different stages of maturation in *Coffea canephora* (robusta) [8,9]. The research aimed to catalogue as many genes as possible which are involved in the bean-filling period of fruit development in order to better understand the final composition of the beans which constitute the commercial product. The purpose of the PUCE CAFE Project was to develop a long oligonucleotide array based on available sequences and thus to use this new tool to perform large-scale transcriptomic analyses in different areas such as bean/fruit development, polyploidy or drought resistance in *Coffea canephora* or *Coffea arabica*. To assess its utility, we ran a comparison between three different tissues, i.e. mature beans, flowers and fully-expanded leaves, in *Coffea arabica* in order to catalogue genes specifically expressed in each tissue. We analyzed in particular the genes involved in fatty acid synthesis and storage proteins and compared our results with those in recent publications on *Coffea* [10] and also with exalbuminous bean species. Then we tested the usability of our 15k microarray for three coffee species.

## Methods

### The Coffee Gene Assembly (Build II)

To create the SGN Coffee Unigene Build II <http://solgenomics.net/>, 71,659 EST (Expressed Sequence Tag) chromatograms were processed from the following *C. canephora* sequence libraries: cccl (coffee leaf, 11,655 chromatograms), ccp (coffee pericarp, 10,849 chromatograms), cccs18w (coffee early-stage bean, 1,972 chromatograms), cccs30w (coffee middle-stage bean, 15,318

chromatograms), cccs42w (coffee late-stage bean, 42 weeks after pollination, 469 chromatograms), cccs46w (coffee late-stage bean, 46 weeks after pollination, 10,907 chromatograms), cccwc22w (coffee early-stage whole fruit, 11,660 chromatograms), irdccf (IRD coffee cherry in various developmental stages, 5,089 chromatograms), irdccl (IRD, young leaves, 3,693 chromatograms) and nDav1 (Nestle Dav1, 47 chromatograms), using PHRED software <http://www.phrap.org/phredphrap-consed.html>[11]. The sequences were processed to remove vector, adaptors and low complexity sequences using an SGN-developed Perl script). Chimeric sequences were screened by processing the BLAST results [12] using *Arabidopsis thaliana* <ftp://ftp.arabidopsis.org/home/tair/Sequences/> as reference dataset and a SGN Perl script. A total of 55,539 sequences passed the filter tests and were used in the assembly. The unigene assembly was created in two steps. First, using a self-BLAST and an SGN Perl script (precluster.pl), we implemented a pre-clustering phase of the EST sequences with a minimum identity percentage of 90% and a minimum alignment length of 30 bp. Secondly, we used CAP3 software <http://seq.cs.iastate.edu/>[13] for each cluster with the following parameters: -e 5000 -p 90 -d 10,000 -b 60. The -e, -d and -b options were set so that the assembler would disregard them or minimize their effect. The -p option increased the sequence identity necessary with overlaps to 90 from a default of 75, and thus was found to be lacking in stringency.

Concerning unigene annotations, we first made a homology search using the BLAST program against GenBank <ftp://ftp.ncbi.nih.gov/genbank/> and *Arabidopsis thaliana* <ftp://ftp.arabidopsis.org/home/tair/Sequences/datasets>, setting an e-value of 1e-10 as the cutoff value. Secondly, we implemented a prediction of protein sequences based on unigene sequences using ESTScan software <http://estscan.sourceforge.net/>[14] and an SGN Perl script (longest6frame.pl), which simply determines the longest open reading frame and translates it into a protein sequence. Thirdly, we set up a protein domain homology search on predicted protein sequences using InterProScan software <http://www.ebi.ac.uk/Tools/InterProScan/>[15]. All the information concerning the different scripts used to perform EST assembly are available on [https://github.com/solgenomics/sgn-home/tree/master/aure/scripts/old\\_sgn\\_transcript/](https://github.com/solgenomics/sgn-home/tree/master/aure/scripts/old_sgn_transcript/).

### Long Oligonucleotide Microarray Design and Synthesis

The *Coffea canephora* long oligonucleotide set was designed and synthesized by Operon (Cologne, Germany) based on the SGN Coffee Build II (15,721 unigenes; <http://solgenomics.net/>). An amino linker was attached to the 5'-end of each oligonucleotide. The oligonucleotides, selected to limit secondary structure,

have a melting temperature of  $67 \pm 3^\circ\text{C}$ , length  $65 \pm 5$  bases, GC content  $43 \pm 5\%$ . More than 98% of the oligonucleotides were within 1000 bases from the 3'-end of the available gene sequence. For 195 unigenes, no adequate oligonucleotide could be designed and therefore correspond to "missing genes" (Additional File 1). BLAST alignments were performed to identify oligonucleotides that could cross-hybridize with other sequences of the SGN Coffee Build II. Finally out of 15,522 oligonucleotides designed, there are 371 oligonucleotides which have  $> 70\%$  of overall identity to another unigene and have a contiguous identical length of over 20 nt common to another unigene (Additional File 2).

### Plant Material and RNA Extraction

In a first experiment (Experiment 1), we compared three tissues, namely fully-expanded leaves, open flowers and mature beans. They were collected from *C. arabica* L. cv. *Caturra T 2308* grown in greenhouse conditions in Tours, France.

In a second experiment (Experiment 2), we compared fully-mature leaves of three species (namely *C. arabica*, *C. canephora* and *C. eugenioïdes*) to determine if our microarray could be used for different coffee species. *C. arabica* was represented by the cv. 'Java' issued from the Arabica woodland Ethiopian pool and by one genotype representing the Arabic-cultivated pool. *C. canephora* was represented by the cv. 'Nemaya' derived from the cross of two Congolese genotypes. Finally, *C. eugenioïdes* was represented by several genotypes, collected in Kenya at the Mount Elgon. The coffee seedlings were grown in a greenhouse with natural daylight and a constant temperature of  $24^\circ\text{C}$  and watered as necessary. After 120 days, the plants were transferred for an additional 60 days to a phytotron chamber (CRYONEXT, France, model RTH 1200L). The standard conditions in the phytotron were 12-hr light ( $600 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ,  $26^\circ\text{C}$ ), 12-hr dark ( $22^\circ\text{C}$ ), with 80% to 99% relative humidity. In each growth chamber three plants for each species were cultivated. Each plant represented one replicate. Two fully-developed leaves were collected from each plant (i.e. two leaves/replicate) at noon (6-8 hours after lights on) and then flash-frozen in liquid nitrogen.

Tissues were ground into a powder and total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen; Valencia, Cal., USA), then treated with DNase following the manufacturer's instructions. Total RNAs were finally eluted from the columns with RNase-free water ( $2 \times 30 \mu\text{L}$ ). For each tissue, three independent RNA extractions were performed. All RNA samples were analyzed by formaldehyde agarose gel electrophoresis to assess their integrity. To test for contamination by polyphenols, carbohydrates and proteins, a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies; Wilmington,

Delaware, USA) was used. Only RNA samples with OD 260/280  $> 1.8$  and OD 260/230  $> 2$  were used for further analysis.

### RNA Labelling

For the preparation of the labelled Cy3- and Cy5- aRNA target, one microgram of the total RNA samples were amplified and labelled using the Amino Allyl Message Amp II aRNA Amplification Kit (Ambion; Austin, Texas, USA), according to the manufacturer's instructions.

### Microarray Printing

The synthesized oligonucleotides were arranged in 384-well plates, and dissolved at  $20 \mu\text{M}$  in a phosphate buffer ( $150 \text{ mM}$ , pH 8.5). The oligonucleotide probes were printed on reflective epoxysilane-coated slides (Amplislide, Genewave, Ecole Polytechnique, France) using a Lucidea Array printer (GE HealthCare, St. Catharines, Ontario). The oligonucleotides library also included sets of positive and negative control points that were used for verifying, for example, the quality of the microarray and mRNA, the sensitivity and linearity of the signal, or the consistency of the assay. In addition, the expected dye ratios were determined and the differences in signal intensities due to the differences in dye incorporation and quantum yield were estimated.

### Hybridization

Prior to hybridization, oligonucleotides were cross linked to the slides by UV irradiation at 100 mJ and the excess was removed from the arrays by washing them twice in one minute in 0.2% sodium dodecyl sulphate (SDS). Arrays were then washed twice in distilled water. The two labelled aRNA were added to Microarray Hybridization Buffer Version 2 (GE HealthCare, St Catharines, Ontario) in a final concentration of 50% formamide, denaturated at  $95^\circ\text{C}$  for three minutes and applied to the microarrays in individual chambers of an automated slide processor (GE HealthCare, St Catharines, Ontario). Hybridization was carried out at  $37^\circ\text{C}$  for 12 hours. Hybridized slides were washed at  $37^\circ\text{C}$  successively with  $1\times$  Saline Sodium Citrate, 0.2% SDS for 10 minutes, twice with  $0.1\times$  SSC, 0.2% SDS for 10 min, with  $0.1\times$  SSC for one minute and with isopropanol before air drying.

### Data Acquisition

Microarrays were immediately scanned at  $10 \mu\text{m}$  resolution in both Cy3 and Cy5 channels with GenePix 4200AL Scanner (Molecular Devices, Silicon Valley, Cal., USA) with variable photo multiplier tube (PMT) settings to obtain maximal signal intensities with  $< 0.1\%$  probe saturation. ArrayVision<sup>®</sup> software (GE HealthCare,

St Catharines, Ontario) was used for feature extraction. Spots with high local background or contamination fluorescence were flagged manually. A local background was calculated for each spot as the median values of the fluorescence intensities of four squares surrounding the spot.

### Real-time PCR

We carried out reverse transcription of total RNA using random hexamer oligonucleotides and SuperScript II Kit (Invitrogen, Carlsbad, Cal., USA) according to the manufacturer's instructions. Real-time PCR was performed on a LightCycler<sup>®</sup> 480 equipped with a 384-well block using the LightCycler<sup>®</sup> 480 SYBR Green I Master Mix (Roche Diagnostics, Indianapolis, Ind., USA) according to the manufacturer's instructions. The primer sequences used for the determination of gene expression levels are given in Additional File 3. The selection of appropriate housekeeping genes was performed using geNorm [16]. The level of expression of each gene *X* was normalized to the geometric mean of the expression levels of 3 reference genes (Spermidine synthase 1, Cyclophilin and Actin-11), according to the formula

$$\frac{X}{\sqrt[3]{R1 \times R2 \times R3}} = 2^{\left( Ct(X) - \left( \frac{Ct(R1) + Ct(R2) + Ct(R3)}{3} \right) \right)}$$

where *Ct* is the threshold cycle and *R1*, *R2*, *R3* are the 3 reference genes.

Additional information concerning the Q-PCR experiment can be found in the MIQE document (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) (See Additional File 4).

### Experimental Design and Data Analysis

For the first experiment, for qRT-PCR as for microarray, three biological replicates were made for each tissue analyzed (i.e. leaves, flowers and mature beans). The following comparisons were made: Bean-Flower, Leaf-Flower and Leaf-Bean. In all, we performed microarray analyses on 18 slides [3 (replicates) × 2 (dyes) × 3 (organs)]. For qRT-PCR we performed 3 technical replicates × 3 biological replicates × 3 organs for 108 genes and three reference genes (*R1*, *R2* and *R3*) for each tissue.

For the second experiment, three biological replicates were done, each of one containing two leaves. In total we used 36 slides [3 (replicates) × 2 (dye) × 6 (comparisons)]. All microarray analyses were performed using Bioconductor <http://www.bioconductor.org>, the open development software project for the analysis and comprehension of genomic data.

### Preprocessing

A quality analysis was made by generating image plots (MA-plots, boxplot, visualization of the array).

No background correction was performed. Few spots were flagged and controls were removed for the normalization. Loess normalization was performed for each microarray to correct the dye effect and technical bias. Then the microarray data were filtered keeping the spot intensity above a median of 90 percent of the control spots (Empty/Negative Control - *NC*) plus twice the deviation standard, applied to both channels (Red and Green).

### Statistical Analysis

Two tests of differential expression were conducted simultaneously: the first test fit a linear model for the expression data for each gene by using the Limma Package (Linear Models for Microarray Data) [17], the second test SAM used repeated permutations of the data to determine significant genes [18]. It was conducted with the Siggenes package from Bioconductor. Multiple testing adjustments were performed by using a false discovery rate approach [19]. These two analyses allowed us to rank significantly expressed genes. The Bioarray Software Environment (BASE - [20]) (local installation: <http://baseprod.igf.cnrs.fr/index.phtml>) was used to visualize the differential expression for each gene.

### Reproducibility of Biological Replicates and Specificity of the Microarray

For the first experiment, coefficients of variation were calculated for the mean signal intensity for the 6 slides (i.e. two dyes × three biological replicates). Using the procedure rank in SAS 9.2 (SAS Institute, Cary, NC), each CV received a rank according to the size of the CV value. These rank values were then expressed on a 100-based scale.

The effect of sample size on the power of statistical tests for different CVs was estimated using the sample size estimate procedure for a two-sample t-test in SAS 9.2 with  $\alpha = 0.01$ , group 1 mean = 1, and group 2 mean = 1.5 or 2.5. The effect of the CV on the minimally detectable expression ratio (threshold expression ratio) was iteratively estimated for a power of 0.9 and  $\alpha = 0.01$  using the sample size estimate procedure for a two-sample t-test.

To validate the expression changes found in Microarray Experiment 1, transcript levels of 108 genes and three reference genes were quantified by Q-PCR (with validated primers [10]). Results obtained by both techniques (microarray and Q-PCR) (Additional File 3) were compared by calculating the Pearson correlation coefficients (SAS 9.2). For the microarray, the data input into the correlation analysis was the Log<sub>2</sub> ratio value of the weighted average for each gene on the composite array representing all replicates. For qRT-PCR, we used the mean Log<sub>2</sub> ratio value representing all replicate plants.

All correlation analyses carried an alpha value of 0.01 and were performed using SAS 9.2.

For the second experiment, background noise was defined from the intensity of “negative” controls (NC). The median of the negative controls was calculated. A gene was considered expressed if the intensity exceeded twice the standard deviation. If a gene is significantly expressed during a comparison (6 hybridizations), its signal should be superior to the highest background noise in each hybridization (maximum 6 times). We chose to fix this threshold to 5, and for each species we screened the number of genes that reached or surpassed this value (Additional File 5).

A hierarchical tree-clustering support method was performed using TMeV 4.0 software from TIGR. The Euclidean distance was used as a measure of similarity or distance between hybridizations. As a rule, the Average-Link Method was used for linking clusters. With this method, distance calculations are based on pairs of clusters: taking the average between the distance of objects from the first cluster and of objects from the second cluster. The averages are performed for all pairs to determine the actual distance between clusters.

Venn diagrams were generated using the online Venny tool <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

#### Gene Ontology Functional Enrichment Analysis

Computational annotation was also performed using Blast2GO software v2.4.4. (<http://www.blast2go.org> website) [21]. The annotation step was performed using the BlastX algorithm, the NCBI nr database and a Blast expectation value threshold of 1E-3. The Blast2GO tool was then used to obtain GO information from retrieved database matches. Mapping of all sequences was performed using default parameters. An InterPro Scan was also performed to find functional patterns and related GO terms by using the specific tool implemented in the Blast2GO software with the default parameters. Finally, an enrichment analysis was completed for the sets of up-regulated unigenes in each of the three tissues studied using the corresponding Blast2GO module, which is based on Fisher’s Exact Test and FDR statistics. Additional File 6 shows GO terms showing a significantly higher or lower frequency in tissue-specific unigene sets in comparison with the full set of unigenes of the PUCE CAFE array.

#### Availability of the Microarray Data

Microarray data are publicly available at <http://www.ncbi.nlm.nih.gov/geo/>. The GEO accession number is GSE24754 for Experiment 1 and GSE24682 for Experiment 2. The array is referenced as GPL10928.

## Results and Discussion

### Chip Quality

To estimate the quality of the work performed by the MGX platform (Montpellier, France), one validation experiment (Experiment 1) was implemented. Expression was compared in three tissues, namely fully-expanded leaves (L), open flowers (F) and mature bean (B) in *Coffea Arabica*. Three biological replicates were used for each tissue. The following comparisons were studied: Bean-Flower, Leaf-Flower and Leaf-Bean.

Raw quantification and background noise values were represented for each chip (data not shown). Visually, the flags indicated invalidated spots. In this way it was possible to visualize whether there were any particular artefacts on a slide, due for example to washing impurities or to the presence of dust. The distribution of raw intensities, background noise and log-ratios were uniform. Very few spots were flagged (Table 1) and background noise was low and virtually constant when signal intensity increased (Figure 1) indicating that the chips were of very good quality.

### Signal Distribution for each Hybridization and Data Standardization

Gene expression was compared as a function of the dye (Cy3 or Cy5). Background noise was defined from the intensity of “negative” controls (NC). The median of the negative controls was calculated. A gene was considered expressed if the intensity exceeded twice the standard deviation. We compared efficiency for Cy3 and for Cy5 incorporation. The dye bias was greater for low-value signals. Of 15,998 genes, 40.76% were always significantly expressed compared to background noise for the red dye and 44.70% for the green dye. This bias was largely corrected by standardizing the data using the Loess regression method.

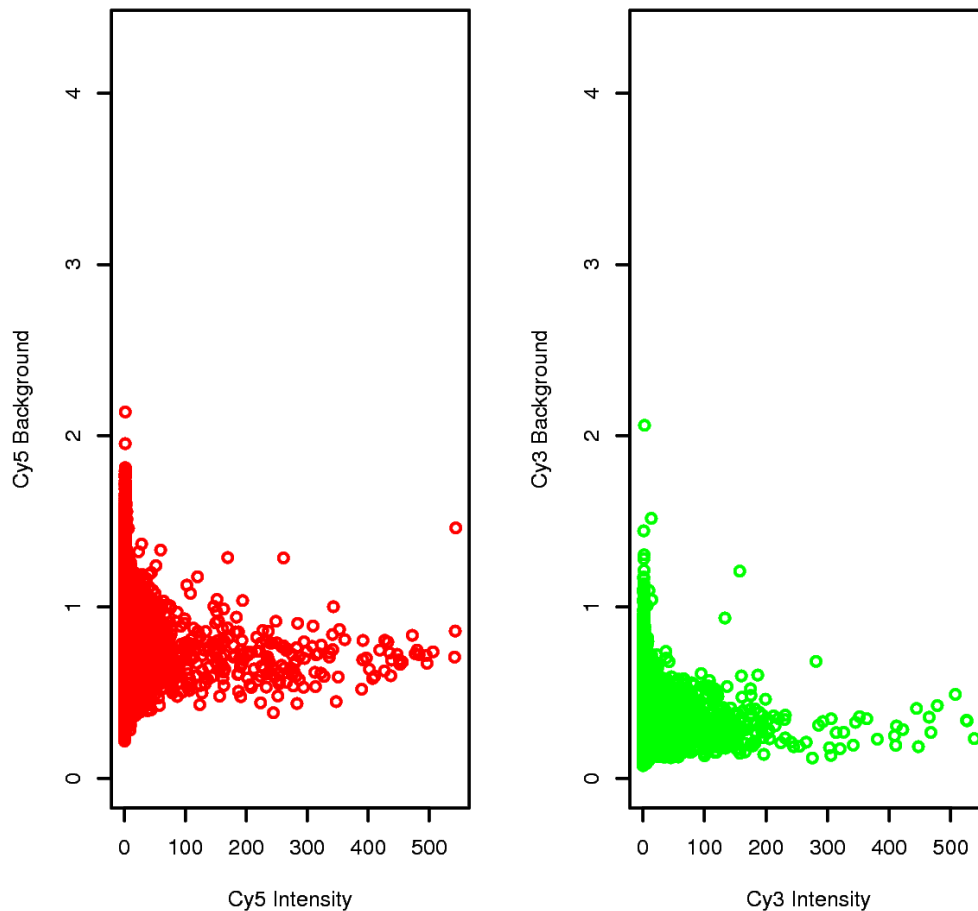
After standardization, the curves for the Cy3 and Cy5 signals were superposed upon the density graphs (Figure 2). The MA-plot scatterplots did not display any striking differences as most of the technical biases were reduced by using automated protocol as well as specific slides for reducing differences due to dye incorporation.

Box-plots of both pre- and post-normalization (Figure 3A and 3B) confirmed that our data were successfully

**Table 1 In Experiment 1, for each comparison (six slides) 16,512 spots were examined**

Comparisons	Number of Spots invalidated by visual examination (Flags)
Leaf/bean	16
Leaf/flower	34
Bean/flower	7

The number of genes invalidated by visual examination is indicated.



**Figure 1** Plots for each Channel of Background Intensity versus Intensity. The background is uncorrelated and does not increase along with the intensity. Flower vs. leaf comparison data are presented here.

normalized. Data quality was assessed by comparing the signal intensity data from each array to that obtained from the technical or biological replicates. Pearson correlation between replicates was calculated for every gene in all the arrays, resulting in a very high correlation level, with a coefficient of  $>0.89$  for every independent experiment in a pairwise comparison (Figure 3D). This high coefficient is indicative of the precision level in which the microarray is able to process transcriptomic data reliability.

A support tree-clustering method with bootstrapping using expression data was performed to statistically validate the tool. Figure 3C showed a high level of similarity between the replicates.

#### Reproducibility of Biological Replicates

The variability between expression profiles derived from the two dyes and the three biological replicates was estimated. For each experiment we calculated the coefficient of variation (CV) for the mean signal intensities for six

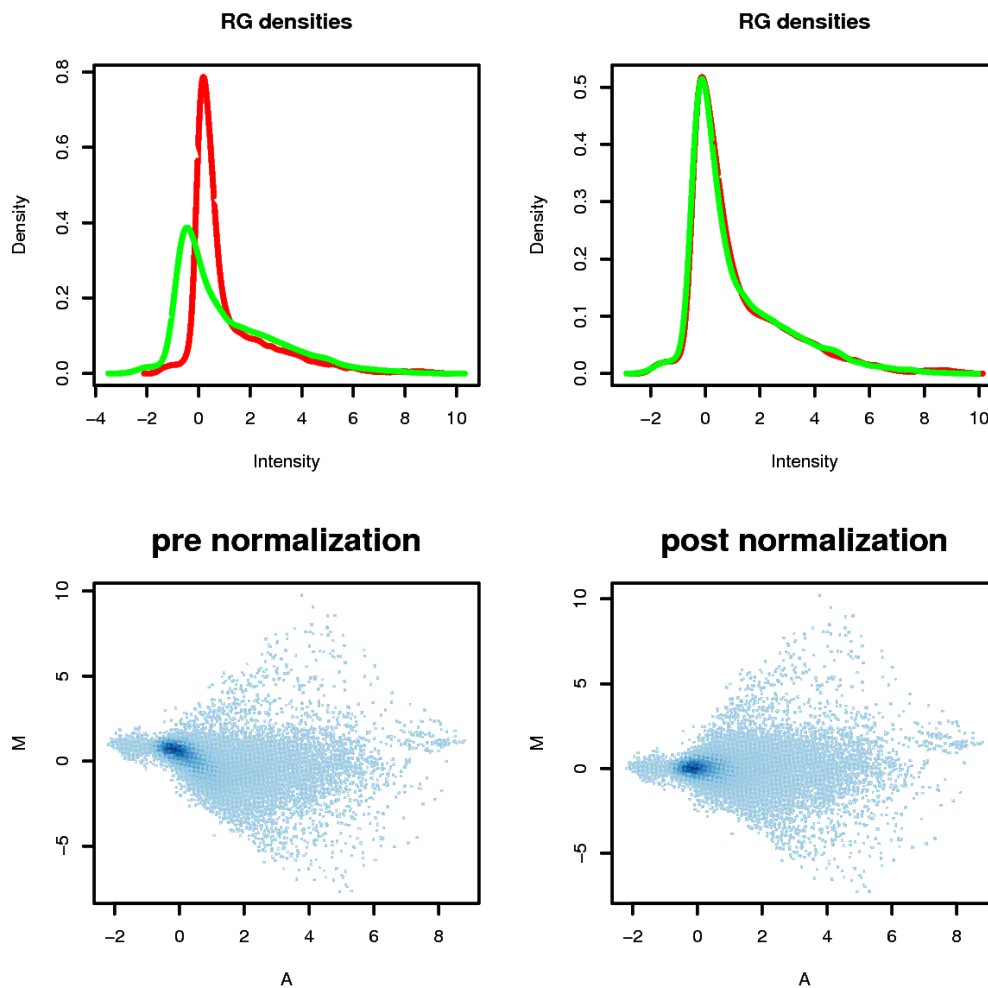
slides (i.e. 2 dyes  $\times$  3 biological replicates). Ninety percent of the spots on the arrays could be determined with a CV of less than 42% (Table 2). These values are consistent with those obtained previously [22]. The CV (%) was then plotted against the relative rank of the CV (Additional file 7). Based on this estimated variance, a power analysis was performed for a two-sample t- test (Additional file 8).

With a CV of 35%, a sample size of six slides can detect a 2-fold change in gene expression with a power of 90%, with a Type I error rate of 10%. The detection of 1.5-fold changes with a Type I error rate of 0.1, a power of 90% and sample size of six slides requires the CV to be below 20%. We therefore concluded that a sample size of six slides would allow acceptable control of both Type I and Type II errors.

#### Analysis of Differential Expressions

Two statistical analyses were performed on normalized data from Experiments 1 and 2, one by the Limma





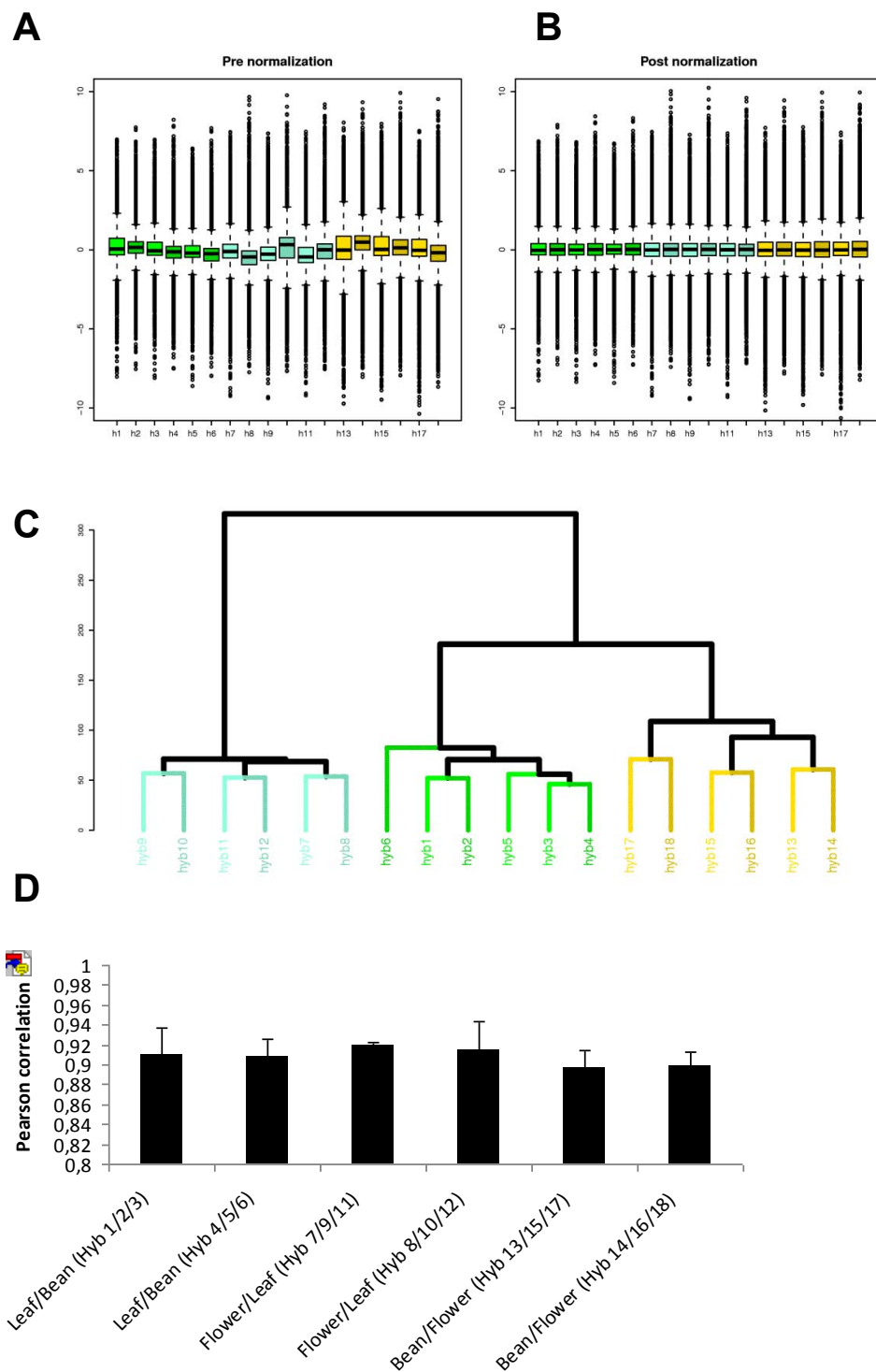
**Figure 2** Density Plot and MA-plot for Both Channels Before and After Normalization. *Density plot*: green and red curves represent the densities of the intensity of the green and the red channel. After normalization, the curves are similar. *MA-plot*: After normalization, the data fall to a straight horizontal line along 0; before normalization, there is a slight upward curve. Data from leaf vs. bean comparison (Hybridization No. 1, Experiment 1) are presented in this example.

Method (linear model), the other by the SAM Method (significance analysis for microarray) with two thresholds  $P = 0.01$  and  $P = 0.05$ . For each comparison, lists of differentially-expressed genes were generated. One list was derived from the “Limma” analysis, the other from the “SAM” analysis. The files are available on BASE <http://baseprod.igf.cnrs.fr/index.phtml>. The Limma Method was more restrictive than the SAM Method, so the number of genes determined as being significantly differentially expressed was therefore smaller. Nevertheless, the totality of these genes was also detected by the SAM Method in the three comparisons conducted. These genes were therefore validated by two independent methods (results not shown). The lists

resulting from the SAM 0.01 analysis were used to compare differentially-expressed genes for each comparison (Figure 4 for Experiment 1 and Figure 5 for Experiment 2). The lists resulting from the Limma  $P = 0.01$  analysis was used to identify over-expressed genes in one specific tissue (bean flower or leaf) (Figure 6 Additional file 9).

#### Comparison of the Three Tissues (Mature Bean, Flower, Leaf)

The number of genes that were differential and significantly expressed when comparing different tissues was between 9,021 and 9,999 genes (Figure 4), i.e. between 58% and 64% of genes spotted on the array. For each comparison, the percentages of up-regulated genes



**Figure 3 Quality Analyses between Biological and Technical Hybridization Replicates for Experiment 1 (Bean, Flower and Leaf).** Hybridizations 1-6 correspond to Leaf-Bean comparisons. Hybridizations 7-12 correspond to Flower-Leaf comparisons and hybridizations, and 13-18 correspond to Bean-Flower comparisons. The box-plots allow us to compare the log<sub>2</sub> (ratio) distribution of all the hybridizations before (A) and after (B) normalization. The distribution of log<sub>2</sub> ratios for all the comparisons is shown on one plot. The baseline is set to a similar raw expression level, allowing the inter-chips comparison. (C) Hierarchical clustering of samples using Euclidean Distance on normalization data. The samples cluster primarily by replicates. In blue is the flower-leaf comparison; in green, the leaf-bean comparison and in brown, the bean-flower comparison. (D) Pearson correlation at gene-level for all the probes in the replicates of the microarray. All the replicates showed a correlation value greater than  $p > 0.89$  thus showing a high level of similarity.

**Table 2 Mean, Median and P90 of the Coefficients of Variation (CV) of the Fluorescence Signal Intensity**

Comparison	Tissue	Mean	Median	P90
Leaf/bean	Leaf	27.31	24.15	41.58
	Bean	25.66	21.82	38.93
Leaf/flower	Leaf	23.38	21.76	35.68
	Flower	22.92	20.46	32.59
Bean/flower	Bean	23.28	20.58	37.56
	Flower	19.90	17.96	32.26

varied between 30 and 33%. The three tissues therefore behaved globally in an identical manner.

**Correlation between qRT-PCR and Microarray Results for 111 Genes**

Consistent with the previous results [23], fold change results determined by qRT-PCR were significantly greater than fold change assessed for the same genes by microarray analysis. Correlations for the data sets (i.e., bean/flower, flower/leaf, leaf/bean) ranged from 0.78-0.81 ( $p < 0.01$ ) for the 108 genes analyzed (Additional file 3). The direction of change was similar for both qRT-PCR and microarray for 70-75% of the genes analyzed. Furthermore, when we correlated only the significantly-expressed genes (SAM 0.05) ( $> 1.5$  fold change), the degree of correlation between microarray and qRT-PCR results was higher, ranging from 0.85-0.87 ( $p < 0.01$ ) for 83 genes. The lack of congruence between both methods for genes exhibiting low levels of variation ( $< 1.5$  fold change) has been commonly reported [24]. These results validate the implicit assumption that there

is a good correlation between the microarray data and the mRNA levels in the tissue under investigation.

**Specifically Over-Expressed Genes in each Tissue (Mature Bean, Flower, Leaf)**

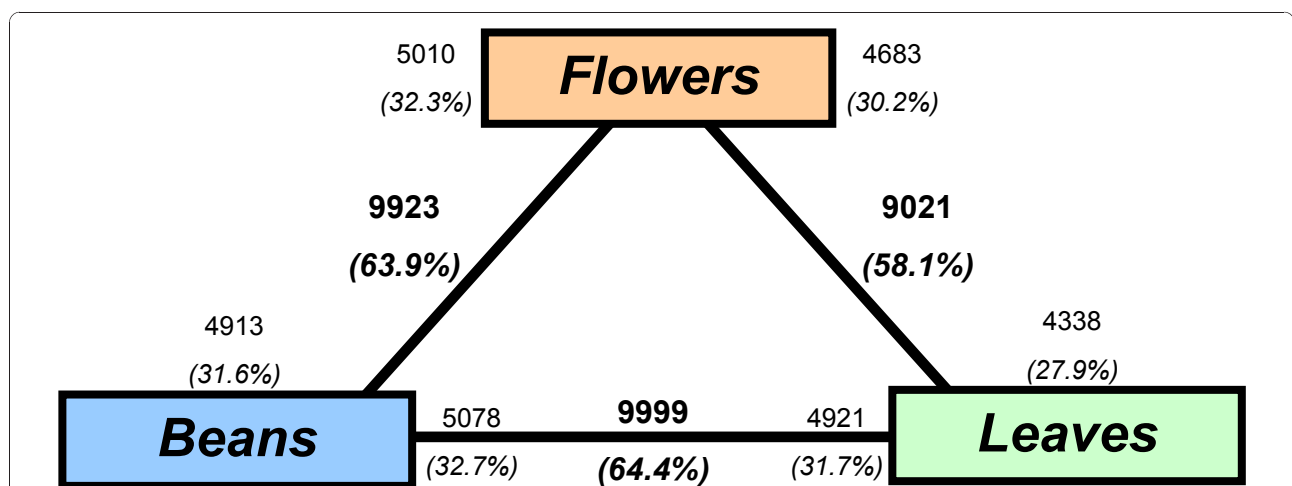
Based on the lists of deregulated genes identified for each comparison (Limma Analysis;  $P = 0.01$ ), Venn diagrams were constructed (Figure 6). Over-expressed genes in each specific tissue were so listed (Additional file 9). 1,565 genes were significantly over-expressed in the bean compared to the leaves and flowers (Figure 6A and Additional file 9). Likewise, 220 “flower-specific” (Figure 6B) and 176 “leaf-specific” (Figure 6C) genes were identified.

The number of genes which are over-expressed in the bean is significantly higher than those identified in flower and leaf. This observation is quite normal since the PUCE CAFE array is mainly based on genes expressed in the grain during fruit maturation.

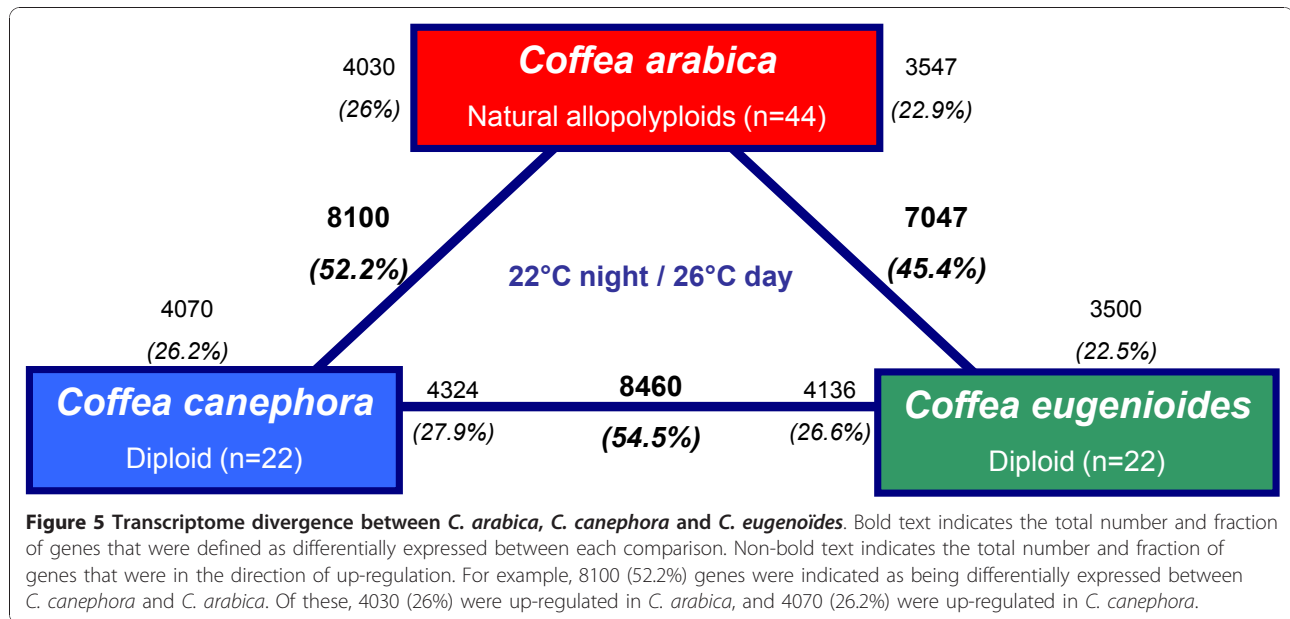
This analysis is quite interesting and clearly identified the genes involved in different metabolic pathways specific to each organ. In order to shed light onto the processes involved under the conditions studied, we enriched the Gene Ontology (GO terms) among up-regulated genes in the three different organs. Additional file 6 shows GO terms showing a significantly higher or lower frequency in tissue-specific unigene sets in comparison with the full set of unigenes of the PUCE CAFE array.

**Validity of Microarray Results related to Biosynthesis Pathways of Lipids or Storage Proteins**

To test the accuracy of the results obtained with the 15k coffee microarray a bit further, we compared expression



**Figure 4 Transcriptome divergence between the three tissues (flower/leaf/mature bean).** The total number and fraction of genes diagnosed as differentially expressed in each contrast are indicated in bold text. Also shown for each contrast is the partitioning of the total number of differentially expressed genes in the direction of up-regulation; these numbers are indicated in non-bold text. For example, 9923 genes are indicated as being differentially expressed between flowers and leaves. Of these, 5010 (32.3%) were up-regulated in flowers, and 4913 (31.6%) were up-regulated in beans. Around 58-65% of the 15,522 unigenes were differentially expressed between the three tissues.



patterns of a few genes involved in well-characterized biosynthetic pathways of lipids or storage proteins with those described in studies on *Coffea* and model plants. As storage tissue, the mature endosperm accumulates nutrient reserves (mainly cell-wall polysaccharides, sucrose, proteins and oils) which are mobilized by the embryo during germination and seedling growth. As expected, most of the genes involved in the accumulation of these storage compounds displayed enhanced transcriptional activity in the bean compared to leaves and flowers.

#### Study of Different Genes Involved in the Biosynthesis of Lipids

In coffee leaves, linolenic acid (18:3) is the predominant FA (fatty acid) [25], whereas it represents only a small percentage of the total FA in beans [26]. Accordingly, the gene encoding the enzyme involved in converting linoleic acid to linolenic acid ( $\omega$ -3 desaturase, *FAD8*) was significantly over-expressed in leaves compared to beans (Table 3). Similarly, although waxes can be detected in beans and flowers, these compounds predominantly accumulate in leaves. The first step of wax biosynthesis involves a  $\beta$ -Ketoacyl-CoA Synthase (KCS) activity, which initiates the biosynthesis of a very-long chain of fatty acids specific to waxes. Again, the micro-array showed evidence of a higher KCS gene expression in leaves compared to other tissues (Table 3).

The acyl-ACP thioesterase (encoded by the *Fat B* gene) was described as the control point of the remarkably high palmitic acid content of Arabidopsis flowers in comparison with other organs [27]. A similar expression pattern was observed for the putative *FatB* gene in

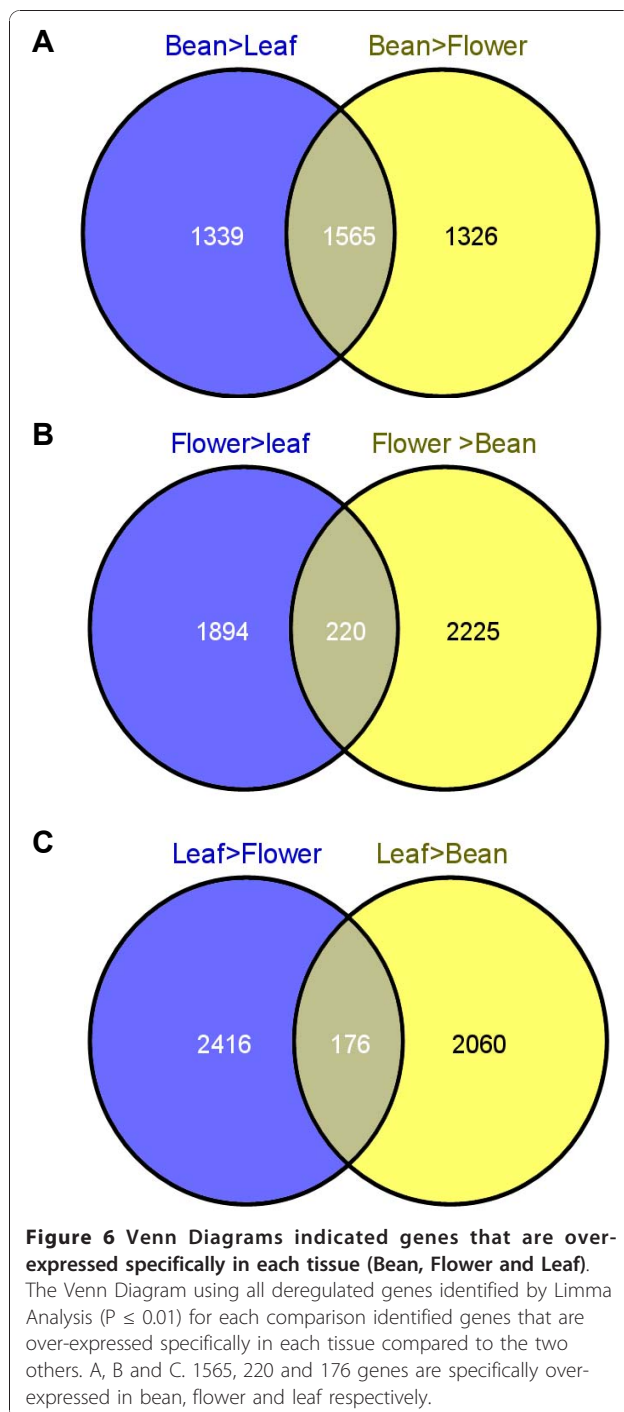
coffee, suggesting that the coffee flower could also be highly rich in this fatty acid. Finally, since coffee flowers are well known for their jasmine fragrance, we investigated the expression pattern of a putative *JMT* gene that encodes a jasmonate O-methyl transferase. Indeed, the volatile plant hormone jasmonate and methyl-jasmonate are also directly involved in flower fragrance [28]. As expected, *JMT* expression appeared to be highly specific in coffee flowers (Table 3).

Finally, the bean-specific expression of genes encoding DGAT (Table 3), the enzyme catalyzing the last step of triglyceride synthesis [29], and Oleosin-2, a structural component of oil bodies [30], provided that we had a good signature of storage lipid accumulation in the coffee bean.

#### Study of Genes Encoding Main Proteins Stored in Mature Beans

The storage proteins in the bean constitute the major portion of the proteins found in ripe beans. The expression of these proteins is temporally regulated during the coffee cherry ripening period and is restricted to bean tissues such as cotyledons or endosperm [31]. The coffee storage protein 1 (*csp1*) mRNA encoding 11 S globulin is highly accumulated in ripe beans and poorly detected in leaves or flowers (Table 4) as shown in previous publications [10,32]. The coffee storage protein 2 (*csp2*) mRNA is also detected in the ripe bean but considerably less than *csp1*, suggesting that among the *csp* gene family a strong difference of expression can be observed from one member to another.

The late embryogenesis abundant (LEA) proteins, a diverse class of highly abundant, heat-stable proteins,



accumulate late in embryo maturation or in endosperm. This accumulation coincides with the acquisition of desiccation tolerance that occurs also during coffee bean ripening. These proteins can be detected in vegetative organs, especially under stress conditions such as cold, drought, or high salinity [33].

*CcLEAP2* and *CcLEAP3* are highly expressed in ripe beans. While *CcLEAP3* is not detected in flowers and

leaves (Table 4), *CcLEAP2* is significantly expressed in flowers. *CcLEAP4* is expressed significantly in the three tissues analyzed but its manifestation in beans is quite low compared to *CcLEAP2* and *CcLEAP3*. This wide difference of expression is also largely observed in *Arabidopsis* [33] among the 51 LEA proteins identified in the genome, suggesting different functions for each member of this superfamily.

#### Utility of the 15k Microarray for Different Coffee Species

Although the long oligonucleotides spotted on the 15K coffee microarray were defined from *Coffea canephora* EST sequences principally derived from genes expressed during coffee fruit development and leaves. We tested the microarray utility tool for two other coffee species, namely *C. arabica* and *C. eugenioides*. In these two species and in *C. canephora*, evolutionary (divergence) is quite recent (< 100 000 to 10 000 years) and their nucleotide divergence was recently estimated at below 5% [34]. Furthermore, *C. canephora* and *C. eugenioides* are considered to be the diploid parents of *C. arabica* (tetraploid) [34].

Transcriptomic differences between the two Arabica genotypes (data not shown) appeared minor, so for the present study we considered the average response of both genotypes for the Arabica species. We observed that 8226, 8270 and 8530 genes were significantly expressed in comparison to the background noise in *C. Canephora*, *C. Arabica* and *C. Eugenioides* respectively (Additional File 5). These last results indicate that about 53% of the genes represented on the chip are specifically expressed in leaves. We also calculated that 97.4% of these expressed genes are common in the three species.

A large difference was observed between the transcriptomes of the parental diploids *C. eugenioides* and *C. canephora* as 54.5% of the 15,522 genes were differentially expressed (Figure 5). Among the differentially-expressed genes, equivalent proportions were up-regulated in each parent 27.9% for *C. canephora* versus 26.6% for *C. eugenioides* (Figure 5). A high fraction of genes was differentially expressed between *C. arabica* and their parents, between 7047 and 8100 genes were indicated as being differentially expressed in *C. arabica* and *C. eugenioides* and *C. canephora* respectively. Of these, 23 to 26% were up-regulated in *C. arabica* and 23-26% were up-regulated in the two diploid species (respectively *C. eugenioides* and *C. canephora*). Finally, a low variation (~ 5.4%) in percentages of differentially expressed genes was observed between the three comparisons.

It can be stated that our microarray tool may be used to analyze global expression not only in *Coffea canephora* but also in other important species such as *Coffea arabica* (which represents 70% of the coffee market) or wild species such as *Coffea eugenioides*.

**Table 3 Expression Patterns of a few Genes Involved in Well-Characterized Lipid Biosynthetic Pathways**

Gene	SGN Accession	Putative Function	E value	% Id	B/L Ratio	B/F Ratio	L/F Ratio	Tissue Specificity
<i>DGAT</i>	SGN-U349452	Acyl-CoA Diacylglycerol acyltransferase (At2g19450)	1E-103	78	4	8.7	2	Bean > L > F
<i>OLE-2</i>	SGN-U350187	Oleosin CcOLE-2 (AY841272)	0	100	9.6	9.1	1.3	Bean > L-F
<i>FAD8</i>	SGN-U349395	Plastidial Linoleate Desaturase FAD8 (At5g05580)	0	73	0.021	0.085	3.8	Leaf > F > B
<i>KCS</i>	SGN-U359520	-Ketoacyl-CoA Synthase (At1g68530)	2E-87	64	0.023	0.526	21	Leaf > F > B
<i>Fat B</i>	SGN-U350529	Acyl-ACP Thioesterase (At1g08510)	1E-153	78	1.1	0.196	0.213	Flower > L-B
<i>JMT</i>	SGN-U349158	SAM:jasmonic acid carboxyl methyltransferase (At1G19640)	3E-34	42	4.6	0.0025	0.0087	Flower > L-B

Comparison: bean to leaves (B/L), bean to flower (B/F), leaf to flower (L/F).

**Table 4 Expression Patterns of a few Genes encoding Potential Storage Proteins**

Gene	SGN accession	Putative function	E value	% Id	B/L ratio	B/F ratio	L/F ratio	Tissue specificity
<i>csp1</i>	SGN-U350946	11 S plant bean storage protein <i>Coffea arabica</i> (Y16975)	1-e134	100	103	93	0.925	B > F-L
<i>csp2</i>	SGN-U347807	11 S plant bean storage protein (At2G28490)	1e-121	50	28.56	12.75	0.315	B>F > L
<i>CcLEAP2</i>	SGN-U350577	Late embryogenesis abundant protein (At1G52690)	1e-12	60	257	6.97	0.01	B>F > L
<i>CcLEAP3</i>	SGN-348605	Late embryogenesis abundant protein (At2G40170)	8e-24	73	592.76	339.53	0.38	B > F-L
<i>CcLEAP4</i>	SGN-347291	Late embryogenesis abundant protein (At4g02380)	2e-15	50	6.26	1.82	0.32	B>F > L

Comparison: bean to leaves (B/L), bean to flower (B/F), leaf to flower (L/F); csp (coffee storage protein); LEAP (late embryogenesis abundant protein).

## Conclusions

We present here the creation and validation of the first coffee oligonucleotide-based microarray tool for functional genomic studies in coffee. Our results reveal that this new tool applies to high-throughput gene expression analyses in various *Coffea* species. Furthermore, the use of the array has proven to be valid for genomic studies on different plant tissues. As proof of principle, we have reported changes in gene expression generated by this microarray in two independent experiments. The statistical analyses of our microarray data, the correct correlation between Q-PCR and the microarray data validate our chip. Overall the coffee microarray (designated as "PUCE CAFE") offers the possibility to carry out functional genomic studies in a wide variety of research areas such as plant development, biotic and abiotic stress response or fruit quality traits. This new tool will be valuable for researchers interested in *Coffea* transcriptomics and will be available through the MGX platform.

## Additional material

**Additional file 1: The Missing Genes.** List of 195 unigenes for which the design of specific oligonucleotides could not be achieved and that are therefore absent from the PUCE CAFE microarray.

**Additional file 2: Oligonucleotides that may cross-hybridize with several sequences.** List of 371 oligonucleotides spotted on the microarray which have > 70% of overall identity to another unigene and have a contiguous identical length of more than 20 nt common to another unigene.

**Additional file 3: Q-PCR and Microarray Data for 111 Genes of Interest Extracted from Experiment 1.** Microarray and Q-PCR were compared by calculating the Pearson correlation coefficients. For the microarray, the data input into the correlation analysis was the Log<sub>2</sub> ratio value of the weighted average and for each gene on the composite array representing all replicates. For qRT-PCR, we used the mean Log<sub>2</sub> ratio value representing all replicate plants. All correlation analyses carried an alpha value of 0.01 and were performed using SAS 9.2. The list of the primers and relative sequence used for the qPCR are indicated as well as the efficiency of amplification for each couple of primers.

**Additional file 4: MIQE document (Minimum Information for Publication of Quantitative Real-Time PCR Experiments).**

**Additional file 5: Genes Significantly Expressed in Experiment 2.** For each species analyzed in Experiment 2 (*C. arabica*, *C. canephora*, *C. eugenoides*), genes showing a significant hybridization signal compared to the background noise are indicated.

**Additional file 6: Gene Ontology Functional Enrichment.** Functional enrichment using gene ontology terms and Fisher's Exact Test was performed using Blast2GO for the sets of up-regulated unigenes in each of the three tissues (bean, leaf and flower). The full set of unigenes of the PUCE CAFE array was used as the reference set. The resulting p-values are indicated after adjusting for FDR multiple testing results. The over- or under-expressed functions at p ≤ 0.01 are specified.

**Additional file 7: Reproducibility of Replicates for the Leaf Tissue in the Leaf-Flower Comparison.** Coefficient of variation (CV%) for all cDNAs spotted on the array based on raw data mean fluorescence values plotted against the relative rank of the CV. CVs were estimated from raw data derived from six replicates (i.e. 2 dye × 3 biological replicates). For this tissue and for this experiment, 90% of the spots on the arrays could be determined with a CV of less than 35%. For the other experiments, 90% of the spots on the arrays could be determined with CV between 32 and 42%.

**Additional file 8: Theoretical Power Analysis for a Two-sample T-test to Detect a 1.25 to 2.50 Fold Change in Gene Expression as a Function of CV%.** Indicated sample size = 6 and a Type I error (false positive rate) of 0.1 were used as input values to determine the fraction of changes in gene expression that would be detected at a given CV%.

A power of 1.0 denotes a Type II error (false negative rate) of zero, i.e. 100% of all changes that occurred were detected.

**Additional file 9: Lists of Genes Specifically Over-expressed in each Tissue Compared to the Two Others (Bean, Flower and Leaf).** The first 3 datasheets correspond to the lists of genes differentially expressed for each comparison identified by the Limma Analysis ( $p = 0.01$ ). The last datasheet corresponds to the lists of genes specifically over-expressed in each tissue compared to the two others.

#### Acknowledgements

This work was completed thanks to a grant from Génoplante, the French consortium for plant genomics. We would like to thank Maud Lepelley and Philippe Schaeffer.

#### Author details

<sup>1</sup>Nestlé R&D Tours, 101 Avenue Gustave Eiffel, Notre Dame d'Oé, BP 49716, 37097 Tours Cedex 2, France. <sup>2</sup>Montpellier Genomix, IGF, 141 rue de la Cardonille, 34094 Montpellier Cedex 05, France. <sup>3</sup>Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, New York 14853-1801, USA. <sup>4</sup>IRD-CIRAD, UMR RPB, BP 64501, 34394 Montpellier, France. <sup>5</sup>IRD, UMR DIAPC, BP 64501, 34394 Montpellier, France. <sup>6</sup>CIRAD avenue Agropolis, 34398 Montpellier Cedex 5, France.

#### Authors' contributions

IP and BB have contributed equally to this work. The project was coordinated by IP, PL and BB. All the authors have read the manuscript and agree with the contents.

#### Competing interests

The authors declare that they have no competing interests.

Received: 24 June 2010 Accepted: 5 January 2011

Published: 5 January 2011

#### References

- Aharoni A, Vorst O: DNA microarrays for functional plant genomics. *Plant Mol Biol* 2002, **48**:99-118.
- Galbraith DW: DNA microarray analyses in higher plants. *OMICS* 2006, **10**:455-473.
- Clarke JD, Zhu T: Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives. *Plant J* 2006, **45**:630-650.
- Mascarell-Creus A, Canizares J, Vilarrasa-Blasi J, Mora-García S, Blanca J, Gonzalez-Ibeas D, Saladie M, Roig C, Deleu W, Pico-Silvent B, Lopez-Bigas N, Aranda MA, Garcia-Mas J, Nuez F, Puigdomenech P, Cano-Delgado AI: An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon (*Cucumis melo* L.). *BMC Genomics* 2009, **10**:467.
- Martinez-Godoy MA, Mauri N, Juarez J, Marques MC, Santiago J, Forment J, Gadea J: A genome-wide 20 K citrus microarray for gene expression analysis. *BMC Genomics* 2008, **9**:318.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P: An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* 2006, **152**:165-512.
- Pendergrast M: Coffee second only to oil? Is coffee really the second largest commodity? *Tea and Coffee Trade Journal* 2009, **181**:38-41.
- Lin C, Mueller LA, Mc CJ, Crouzillat D, Petiard V, Tanksley SD: Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theor Appl Genet* 2005, **112**:114-130.
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de KA, Hamon P: SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol Genet Genomics* 2006, **276**:436-449.
- Joët T, Laffargue A, Salmons J, Doubeau S, Descroix F, Bertrand B, de Kochko A, Dussert S: Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. *New Phytol* 2009, **182**:146-162.
- Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998, **8**:175-185.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, **215**:403-410.
- Huang X, Madan A: CAP3: A DNA sequence assembly program. *Genome Res* 1999, **9**:868-877.
- Iseli C, Jongeneel CV, Bucher P: ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol* 1999, **138**-148.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimmma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: InterPro: the integrative protein signature database. *Nucleic Acids Res* 2009, **37**:D211-D215.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002, **3**:RESEARCH0034.
- Wettenhall JM, Smyth GK: LimmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* 2004, **20**:3705-3706.
- Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
- Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser* 1995, **B 57**:289-300.
- Troein C, Vallon-Christersson J, Saal LH: An introduction to BioArray Software Environment. *Methods Enzymol* 2006, **411**:99-119.
- Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 2008, **36**:3420-3435.
- Degenkolbe T, Hannah MA, Freund S, Hinch DK, Heyer AG, Kohl KI: A quality-controlled microarray method for gene expression profiling. *Anal Biochem* 2005, **346**:217-224.
- Yuen T, Wurmbach E, Pfeffer RL, Sealfon SC: Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res* 2002, **30**:e48.
- Morey JS, Ryan JC, Van Dolah FM: Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biol Proced Online* 2006, **8**:175-193.
- Campos PS, Quartín V, Ramalho JC, Nunes MA: Electrolyte leakage and lipid degradation account for cold sensitivity in leaves of *Coffea* sp. plants. *J Plant Physiol* 2003, **160**:283-292.
- Nikolova-Damyanova B, Velikova R, Jham GN: Lipid classes, Fatty Acid Composition and Triacylglycerol Molecular Species in Crude Coffee Beans Harvested in Brazil. *Food research international* 1998, **31**:479-486.
- Dormann P, Voelker TA, Ohlrogge JB: Accumulation of palmitate in Arabidopsis mediated by the acyl-acyl carrier protein thioesterase FATB1. *Plant Physiol* 2000, **123**:637-644.
- Cheong JJ, Choi YD: Methyl jasmonate as a vital substance in plants. *Trends Genet* 2003, **19**:409-413.
- Beaudoin F, Napier JA: Biosynthesis and compartmentation of triacylglycerol in higher plants. In *Lipid Metabolism and Membrane Biogenesis (Topics in Current Genetics, Vol. 6)*. Edited by: Daum G. Berlin: Springer-Verlag; 2004:267-287.
- Simkin AJ, Qian T, Caillet V, Michoux F, Ben AM, Lin C, Tanksley S, McCarthy J: Oleosin gene family of *Coffea canephora*: quantitative expression analysis of five oleosin genes in developing and germinating coffee grain. *J Plant Physiol* 2006, **163**:691-708.
- Shewry PR, Napier JA, Tatham AS: Seed storage proteins: structures and biosynthesis. *Plant Cell* 1995, **7**:945-956.
- Marraccini P, Deshayes A, Pétiard V, Rogers WJ: Molecular cloning of the complete 11 S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants. *Plant Physiology and Biochemistry* 1999, **37**:273-282.

33. Hundertmark M, Hinch DK: LEA (Late Embryogenesis Abundant) Proteins and their Encoding Genes in *Arabidopsis Thaliana*. *BMC Genomics* 2008, **9**:118.
34. Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A: Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol Gen Genet* 1999, **261**:259-266.

doi:10.1186/1471-2164-12-5

**Cite this article as:** Privat *et al.*: The 'PUCE CAFE' Project: the First 15K Coffee Microarray, a New Tool for Discovering Candidate Genes correlated to Agronomic and Quality Traits. *BMC Genomics* 2011 **12**:5.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



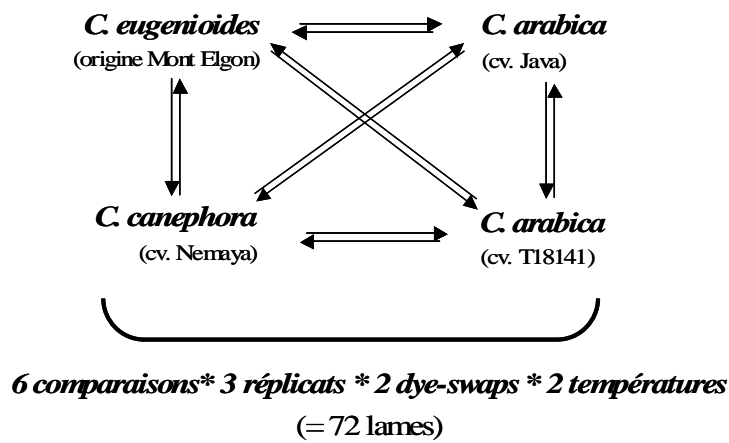


# CHAPITRE II

## DIVERGENCE ET DISTANCES TRANSCRIPTOMIQUES SOUS CONTRAINTE THERMIQUE ENTRE L'ALLOTETRAPLOÏDE (*C.* *ARABICA*) ET SES ESPECES PARENTALES DIPLOÏDES (*C. CANEPHORA* ET *C. EUGENIOIDES*)

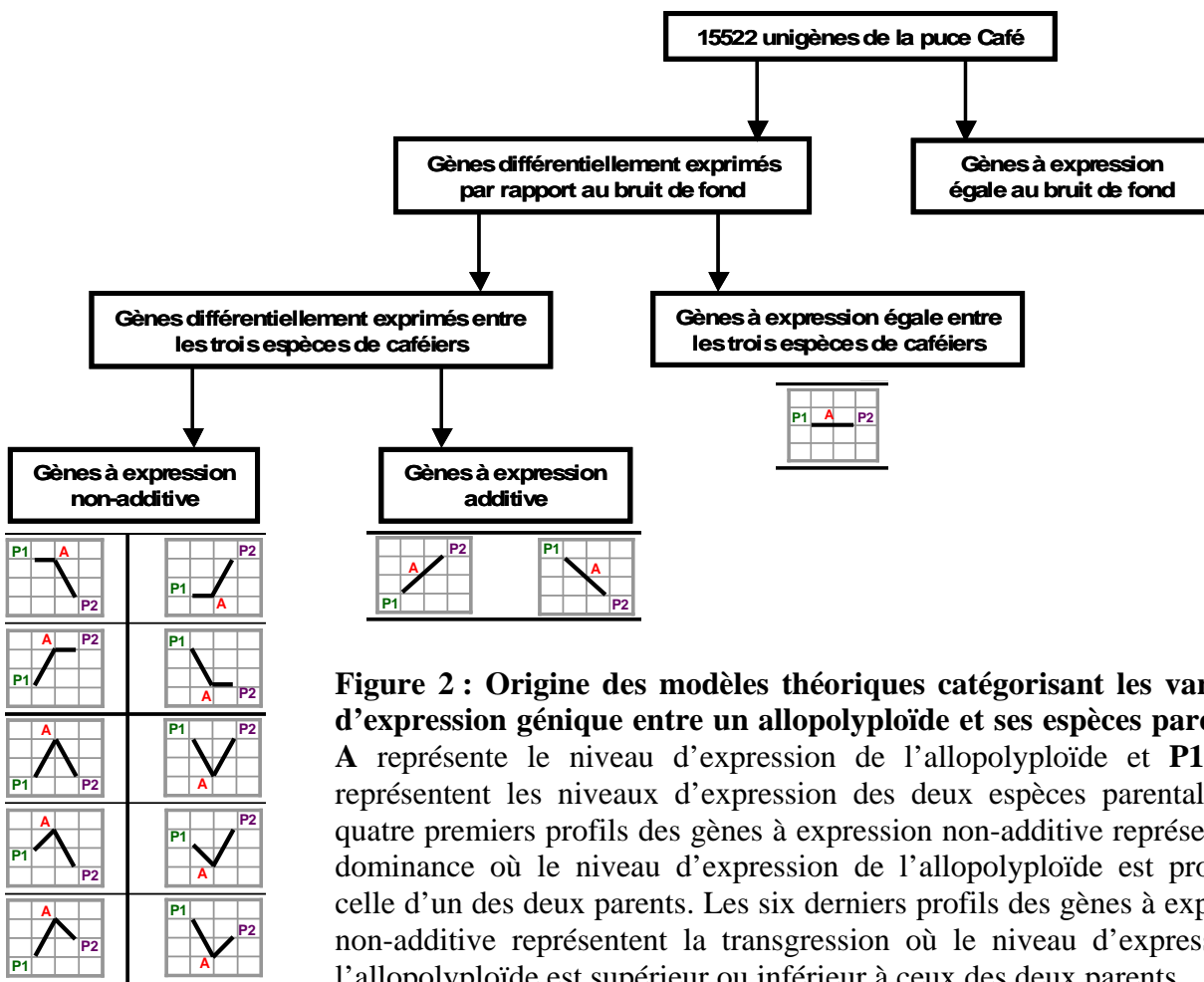
La première partie de ce chapitre a fait l'objet d'une publication:

**Bardil A, Dantas de Almeida J, Combes M-C, Lashermes P, Bertrand B. 2011.** Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytologist*. doi: 10.1111/j.1469-8137.2011.03833.x



**Figure 1 : Dispositif expérimental.**

La figure représente le dispositif expérimental de l'étude. Les quatre accessions et les comparaisons effectuées (flèches noires) sont représentées. Au cours de cette étude six comparaisons en deux dye-swaps pour trois réplicats sont effectuées. Ces comparaisons sont faites à deux conditions de température, 30-26°C et 26-22°C.



## INTRODUCTION AU CHAPITRE

Les phénotypes nouveaux suggèrent des variations dans l'expression des gènes chez les espèces polyploïdes. L'analyse du transcriptome permet d'estimer l'effet de la polyploïdie dans l'expression des gènes. Cette analyse peut se faire soit en comparant le niveau d'expression global entre les allopolyploïdes et leurs parents soit en déterminant la contribution de chaque sous-génome homéologue au transcriptome de l'allopolyploïde. Les conséquences de la polyploïdie sur l'expression génique ont fait l'objet d'études s'inspirant de modèles de génétique quantitative basés sur les notions d'additivité et de non-additivité. L'additivité se définit par le niveau d'expression d'un gène chez un allotétraploïde qui est égal à la moyenne de l'expression du même gène chez les parents diploïdes. La non-additivité se définit, quant à elle, par le niveau d'expression d'un gène chez un allotétraploïde qui n'est pas égal à la moyenne de l'expression du même gène chez les parents diploïdes. Les polyploïdes montrent ainsi des niveaux d'expression génique intermédiaires, non présents et/ou 'excédents' ceux de leurs parents diploïdes équivalent à des sur-expressions où sous-expressions. Ces deux notions sont associées à des modèles théoriques qui tentent de catégoriser les différentes variations de l'expression génique entre l'allopolyploïde et ses parents. Cependant l'application des modèles théoriques varie selon les auteurs. Les pourcentages d'additivité et de non-additivité sont établis en fonction de techniques statistiques et d'hybridation différentes. Il faut donc relativiser les comparaisons effectuées entre les espèces, les techniques et les seuils statistiques que l'on observe parfois dans la bibliographie.

Notre expérience porte sur l'analyse de transcriptome de feuilles de trois espèces de caféiers, un allotétraploïde (*Coffea arabica*) et ses deux espèces parentales diploïdes (*Coffea eugenioides* et *Coffea canephora*) à deux conditions de température, 26-22°C et 30-26°C. Quatre accessions (*C. eugenioides*, *C. canephora* et deux *C. arabica*, cv. Java et cv. T18141) ont été utilisés. *Coffea arabica* cv. Java provient du pool sauvage Ethiopien et *C. arabica* cv. T18141 provient d'un rétrocroisement entre *C. arabica* et un hybride interspécifique entre *C. arabica* et *C. canephora*. *Coffea canephora* cv. « Nemaya » provient d'un croisement entre deux génotypes congolais sauvages. *Coffea eugenioides* provient des forêts du mont Elgon au Kenya.

A partir de la PuceCafé à oligonucléotides longs à haute densité, décrite dans le chapitre I, les comparaisons entre chaque accession ont été faites pour chacune des conditions de

température (Figure 1). L'utilisation de répliquats de lame permet d'estimer la variabilité inter-lames alors que l'utilisation de dye-swaps permet d'estimer la variabilité intra-expérimentale. Les gènes de la Puce sont ainsi différenciés du bruit de fond et un rapport de fluorescence Cy3/Cy5 normalisé est utilisé.

Ce chapitre II sera divisé en deux parties. Une première partie portant sur l'analyse transcriptomique de l'allopolyploïde, *C. arabica* et de ses espèces parentales diploïdes, *C. eugenioides* et *C. canephora*. Au cours de cette partie on déterminera les taux de divergence transcriptomique<sup>1</sup> entre l'allopolyploïde et ses espèces parentales mais aussi les proportions des gènes à expression non-additive. On observera si ces variations sont dépendantes des conditions de température. Une deuxième partie portera sur une analyse in silico de la Puce permettant des comparaisons pour chacune des espèces entre conditions de températures. Au cours de cette partie on déterminera si les distances transcriptomiques<sup>2</sup> entre les espèces étudiées sont température-dépendantes et quels seraient les processus biologiques et gènes remarquables impliqués dans ces variations. Il faut noter que pour cette partie, la condition de température 30-26°C sera considérée comme un stress thermique chaud pour les trois espèces étudiées. La condition de température 26-22°C sera utilisée comme un témoin tout en sachant qu'il s'agit d'une condition de température adéquate pour les trois espèces.

---

<sup>1</sup> Le nombre de gènes différentiellement exprimé commun aux espèces comparées ou spécifiques à chacune des espèces.

<sup>2</sup> Distance basée sur l'abondance des transcrits dans le cas des Puces ADN sur l'intensité de fluorescence.

# CHAPITRE II

## PARTIE I

**DIVERGENCE TRANSCRIPTOMIQUE ET EXPRESSION  
GENETIQUE NON-ADDITIVE CHEZ L'ALLOTETRAPLOÏDE (*C.*  
*ARABICA*) ET SES ESPECES PARENTALES DIPLOÏDES (*C. CANEPHORA* ET *C.*  
*EUGENIOIDES*)**

# Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature

Amélie Bardil<sup>1\*</sup>, Juliana Dantas de Almeida<sup>2,3</sup>, Marie Christine Combes<sup>2</sup>, Philippe Lashermes<sup>2</sup> and Benoît Bertrand<sup>1\*</sup>

<sup>1</sup>CIRAD, UMR RPB, équipe DIVA, 911 Avenue Agropolis, BP 64501, 34394 Montpellier, France; <sup>2</sup>IRD, UMR RPB, équipe DIVA, 911 Avenue

Agropolis, BP 64501, 34394 Montpellier, France; <sup>3</sup>EMBRAPA, Recursos Genéticos e Biotecnologia, PO BOX 02372, CEP 70770-900, Brasília-DF, Brazil

## Summary

Author for correspondence:

Benoît Bertrand

Tel: +33 467 416273

Email: benoit.bertrand@cirad.fr

Received: 26 April 2011

Accepted: 14 June 2011

*New Phytologist* (2011) **192**: 760–774

doi: 10.1111/j.1469-8137.2011.03833.x

**Key words:** *Coffea arabica*, genomic expression dominance, microarrays, natural allopolyploid, plasticity, temperatures.

• Polyploidy occurs throughout the evolutionary history of many plants and considerably impacts species diversity, giving rise to novel phenotypes and leading to ecological diversification and colonization of new niches. Recent studies have documented dynamic changes in plant polyploid gene expression, which reflect the genomic and functional plasticity of duplicate genes and genomes.

• The aim of the present study was to describe genomic expression dominance between a relatively recently formed natural allopolyploid (*Coffea arabica*) and its ancestral parents (*Coffea canephora* and *Coffea eugenioides*) and to determine if the divergence was environment-dependent. Employing a microarray platform designed against 15 522 unigenes, we assayed unigene expression levels in the allopolyploid and its two parental diploids. For each unigene, we measured expression variations among the three species grown under two temperature conditions (26–22°C (day–night temperatures) and 30–26°C (day–night temperatures)).

• More than 35% of unigenes were differentially expressed in each comparison at both temperatures, except for *C. arabica* vs *C. canephora* in the 30–26°C range, where an unexpectedly low unigene expression divergence (< 9%) was observed.

• Our data revealed evidence of transcription profile divergence between the allopolyploid and its parental species, greatly affected by environmental conditions, and provide clues to the plasticity phenomenon in allopolyploids.

## Introduction

Allopolyploidy has long been recognized as an important mechanism in eukaryote evolution (Osborn *et al.*, 2003), especially in flowering plants, including many important agricultural crops such as wheat (*Triticum aestivum*), cotton (*Gossypium hirsutum*), sugarcane (*Saccharum officinarum*) and coffee (*Coffea arabica*) (Chen & Ni, 2006; Jackson & Chen, 2009). Polyploidy considerably impacts plant species diversity, giving rise to novel phenotypes and leading to ecological diversification and colonization of new niches (Otto & Whitton, 2000; Adams, 2007). As highlighted by Hegarty & Hiscock (2009), the processes by which two genomes adapt to coexistence within the same nucleus are

complex. Recent studies have documented dynamic changes in plant polyploid gene expression, which reflect the genomic and functional plasticity of duplicate genes and genomes (Jackson & Chen, 2009). To investigate the effects of genomic merger and doubling, transcriptomic divergence between parents and synthetic nascent allopolyploids or natural allopolyploids has been assessed in several recent studies using genome-wide approaches to measure the deviation from additivity. Indeed, in the additive model, allotetraploid expression would be expected to be equivalent to the average expression of the parental species. Studies in a variety of allopolyploids have revealed a tendency for non-additive gene expression in *Arabidopsis* allotetraploids (Wang *et al.*, 2006), in *Gossypium* allotetraploids (Chaudhary *et al.*, 2009), in *Senecio* interspecific hybrids and allohexaploids (Hegarty *et al.*, 2006, 2008), and in *Triticum*

\*These authors contributed equally to this work.

allohexaploids (Pumphrey *et al.*, 2009). More recently, Rapp *et al.* (2009) have shown in nascent *Gossypium* allopolyploids that only 4–11% of genes exhibit additive expression while 82–92% of genes show up- or down-regulation relative to the level of one of the two parents, reflecting massive expression dominance. Most studies have compared synthetic allopolyploids with their parents. As highlighted by Buggs (2008), only a few natural polyploid species of known parentage and for which genomic resources are available have been well characterized. In a recent study comparing five *Gossypium* allotetraploids that have diverged after a million or so years of evolution, the authors (Flagel & Wendel, 2010) revealed that the magnitude of dominance remains but that the bias in the direction of dominance disappears.

Among the 104 *Coffea* species so far described (Davis & Rakotonasolo, 2008), all are diploid ( $2n = 2x = 22$ ) and generally self-incompatible, except *Coffea arabica* which appears to be the only one that is tetraploid ( $2n = 4x = 44$ ) and self-fertile (Charrier & Berthaud, 1985). Molecular analyses (Lashermes *et al.*, 1999) have shown that *Coffea arabica* is an allopolyploid resulting from the hybridization between *Coffea eugenioides* (E genome) and *Coffea canephora* (C genome) or ecotypes related to those diploid species. The *C. arabica* coffee tree model displays some interesting features when compared with *Gossypium* or *Arabidopsis*. While diversification of the *Coffea* subgenus *Coffea* probably occurred in the second half of the Middle Pleistocene (450 000–100 000 yr before present (BP)), it is most likely that the allopolyploid speciation of *C. arabica* took place in relatively recent times (10–50 000 BP; A. Cenci *et al.*, unpublished). Note also that *C. arabica* displays little genetic diversity and has experienced spatial isolation in Ethiopia, its centre of primary diversity. As highlighted by Lashermes *et al.* (2010), little divergence has been observed between the two constitutive genomes of *C. arabica* (E<sup>a</sup> C<sup>a</sup>) and those of its parental species. Hybridization followed by polyploidization is sufficiently recent that differences between parental genomes in the polyploid are not eroded (Leitch & Leitch, 2008). However, compared with a neo-synthetic allopolyploid, the *C. arabica* species is old enough for successive generations since 30 000 BP to have been subjected to natural selection, thus allowing stabilization of its genome. Consequently, the *C. arabica* allopolyploid is considered as a 'recent' natural polyploid model, like *Triticum aestivum* (c. 8000 BP; Pumphrey *et al.*, 2009). The origin of *C. arabica* occurred between the origins of more ancient allopolyploid species, such as *Gossypium* (c. 1.5 million BP; Senchina *et al.*, 2003) and *Arabidopsis suecica* (12 000–300 000 BP; Jakobsson *et al.*, 2006), and very recent allopolyploid species, such as *Spartina anglica* (< 150 yr ago; Ainouche *et al.*, 2003), *Tragopogon miscellus* (< 80 yr ago; Soltis *et al.*, 2004) and *Senecio cambrensis* (c. 60 yr ago; Abbott & Lowe, 2004).

We recently generated the first 15K coffee microarray, a spotted 70-mer oligo-gene microarray, based on publicly available expressed sequence tags (ESTs; Privat *et al.*, 2011). In a series of experiments, we demonstrated that this microarray (called 'PUCECAFE') enabled reproducible global expression analysis with different tissues (seeds, leaves and flowers) and with different coffee species (*C. eugenioides*, *C. arabica* and *C. canephora*). In a gene expression analysis using a 15K coffee microarray in which *C. arabica* was compared with its related parents, we assessed: the extent of genomic expression dominance between the *C. arabica* transcriptome and its parents; and whether the divergence between the allopolyploid transcriptome and its parents is modulated by temperature.

## Materials and Methods

### Plant material

Fresh mature seeds of *Coffea arabica* (L.) and *Coffea canephora* (Pierre) were provided by the Centre de Coopération Internationale en Recherche Agronomique pour le Développement from La Cumpulida (Matagalpa, Nicaragua). *Coffea eugenioides* mature seeds were provided by the Coffee Research Foundation, Kenya. *Coffea arabica* was represented by two accessions, that is, cv Java derived from the wild *Coffea arabica* Ethiopian pool and cv T18141, a *C. arabica* homozygous line derived from a backcross between *C. arabica* and a natural interspecific hybrid between *C. arabica* and *C. canephora*, which thus is introgressed by *C. canephora* unigenes. *Coffea canephora* was represented by cv Nemaya, derived from a cross between two wild Congolese genotypes. Finally, *Coffea eugenioides* accession seeds were collected from trees originating from Mount Elgon forest in Kenya (1°8'N 34°33'E). The coffee seedlings were grown in a glasshouse under natural daylight, at a constant temperature of 24°C, and watered as necessary. After 120 d, the plants were transferred to a phytotron chamber (CRYONEXT, Montpellier, France; model RTH 1200L).

### Cultivation in phytotron chambers

The study involved comparing the transcriptomes of different species under two sets of growing conditions, which differed in terms of the diurnal and nocturnal temperatures. In the two phytotron chambers, the photoperiod, humidity and luminosity were set at 12 h : 12 h, light : dark, 80–90% and 600  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , respectively. The plants were subjected to a diurnal temperature of 26°C and a nocturnal temperature of 22°C in one phytotron (referred to as the coldest temperature) and 30°C and 26°C, respectively, in a second phytotron (referred to as the hottest temperature). In each phytotron chamber, from each accession, three plants were grown in a randomized complete block design.

After 45 d for the hottest temperature and 60 d for the coldest, plant size and number of newly formed leaves were similar. Two young leaves were then collected from each plant at midday (6–8 h after lights on), and then flash-frozen in nitrogen and stored at  $-80^{\circ}\text{C}$  until extraction.

### Experimental design

Two leaves were pooled per plant from three plants to form three biological replicates, which were subjected to RNA extraction and hybridized in microarray. We used 72 slides (i.e. 3 (replicates)  $\times$  2 (dye-swaps)  $\times$  6 (comparisons)  $\times$  2 (growth temperature conditions)) following a saturated design as described in Supporting Information Fig. S1.

### RNA extraction

Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations, but with slight modifications. A DNase treatment (RNase free DNase set; Qiagen) was carried out on RNA during the extraction protocol and the total RNA was eluted in a smaller volume of 65  $\mu\text{l}$ . Each RNA sample was quantified on a Nanodrop (ThermoFisher Scientific Inc., Waltham, MA, USA). RNA quality was confirmed on a Bioanalyser (Agilent Technologies, Foster City, CA, USA).

### Microarray hybridization

The microarrays were performed at the Genomix transcriptomic platform (Institut de Génomique Fonctionnelle, Montpellier) using the protocols described previously (Privat *et al.*, 2011).

**Coffee gene assembly (Build II)** To create the SGN Coffee Unigene Build II (<http://solgenomics.net/>), 71 659 EST chromatograms were processed from the following *C. canephora* sequence libraries: cccl (coffee leaf; 11 655 chromatograms), cccp (coffee pericarp; 10 849 chromatograms), cccs18w (coffee early-stage bean; 1972 chromatograms), cccs30w (coffee middle-stage bean; 15 318 chromatograms), cccs42w (coffee late-stage bean, 42 wk after pollination; 469 chromatograms), cccs46w (coffee late-stage bean, 46 wk after pollination; 10 907 chromatograms), cccwc22w (coffee early-stage whole fruit; 11 660 chromatograms), irdccf (IRD coffee cherry in various developmental stages; 5089 chromatograms), irdccl (IRD, young leaves; 3693 chromatograms) and nDav1 (Nestle Dav1; 47 chromatograms), using PHRED software (<http://www.phrap.org/phredphrapconsed.html>) (Ewing *et al.*, 1998). The sequences were processed to remove vector, adaptors and low-complexity sequences using an SGN-developed Perl script. A total of 55 539 sequences passed the filter tests and were used in the assembly. The

unigene assembly was created in two steps. First, we used a self-BLAST and an SGN Perl script (precluster.pl) and, secondly, we used CAP3 software (<http://seq.cs.iastate.edu/>) (Huang & Madan, 1999) for each cluster.

**Long oligonucleotide microarray design and synthesis** The *C. canephora* long oligonucleotide set was designed and synthesized by Operon (Cologne, Germany) based on the SGN Coffee Build II (15 721 unigenes; <http://solgenomics.net/>). An amino linker was attached to the 5' end of each oligonucleotide. The oligonucleotides, selected to limit the secondary structure, have a  $67 \pm 3^{\circ}\text{C}$  melting temperature,  $65 \pm 5$  base length, and  $43 \pm 5\%$  GC content. More than 98% of the oligonucleotides were within 1000 bases from the 3' end of the available gene sequence. For 195 unigenes, no adequate oligonucleotide could be designed and they were therefore classified as 'missing unigenes' (Table S3). BLAST alignments were performed to identify oligonucleotides that could cross-hybridize with other sequences of the SGN Coffee Build II. Finally, out of 15 522 oligonucleotides designed, 371 oligonucleotides had  $> 70\%$  overall identity to another unigene and had a contiguous identical length of over 20 nt common to another unigene (Table S4). For the preparation of labelled Cy3- and Cy5-aRNA target, 1  $\mu\text{g}$  of total RNA sample was amplified using the Amino Alkyl Message Amp II aRNA amplification kit (Ambion), according to the manufacturer's instructions. The oligonucleotide probes were printed on reflective epoxysilane-coated slides (Amplislide; Genewave, Ecole Polytechnique, France) using a Lucidea Array printer (GE Healthcare, Bio-Sciences Corp., Piscataway, NJ, USA). The oligo library also included sets of positive and negative controls used for quality control. The two labelled aRNA were added to Microarray Hybridization Buffer Version 2 (GE Healthcare) and applied to the microarrays in the individual chambers of an automated slide processor (GE Healthcare). Hybridization was carried out at  $37^{\circ}\text{C}$  for 12 h. Hybridized slides were washed and were immediately scanned at 10- $\mu\text{m}$  resolution in both Cy3 and Cy5 channels with a GenePix 4200AL scanner (Molecular Devices, Sunnyvale, California, USA). ARRAYVISION (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA) software was used for feature extraction. Spots with high local background or contamination fluorescence were flagged manually. A local background was calculated for each spot as the median fluorescence intensity of four squares surrounding the spot. This background was subtracted from the foreground fluorescence intensity.

**Validation of microarray results** The 15K coffee microarray was created from *C. canephora* EST libraries (Privat *et al.*, 2011). For *C. arabica*, public accessibility to EST collections was limited (Lashermes *et al.*, 2008) and no EST resources were available for *C. eugenioides*. Despite the high genetic similarity of these three species, we wanted to ensure



that there was no bias at the time of hybridization in favour of *C. canephora*. Our choice of no-bias indicator was the number of unigenes significantly expressed compared with background noise. The signals emitted by the negative controls defined the background noise. A unigene was considered expressed if the intensity exceeded twice the median negative control standard deviation. If a unigene was significantly expressed during a comparison (six hybridizations), its signal should be superior to the highest background noise in each hybridization (maximum six times). We chose to set this threshold at 5, and we screened the number of genes that reached or surpassed this value for each species. Consequently, a detection limit threshold of 5 meant that, for a given gene in the six hybridizations, we observed a signal that was significantly greater than the background noise at least five times. Correlation analyses were performed in order to check for similarities between biological and technical replicates through Pearson's moment correlation analysis.

**Analyses of expression changes between species and growth conditions** A significant analysis of microarray (SAM) test was run using the Bioconductor SIGGENES package (<http://bioconductor.case.edu/bioconductor/2.5/bioc/html/siggenes.html>). Repeated permutations of the data were carried out to identify significant unigenes (Tusher *et al.*, 2001). Multiple testing adjustments based on Benjamini and Hochberg's false discovery method (1995) were performed ( $FDR \leq 0.05$ ), allowing a stringent analysis with no false positive identification of differentially regulated unigenes. These analyses provided the ranking of significantly expressed unigenes. For each unigene, we had the log 2-fold expression ratio of three contrasts: the diploid parents to each other and each diploid parent to the allopolyploid. For each comparison, we determined the number of differentially expressed unigenes. That number of unigenes was compared with the total number of unigenes of the chip (15 522 unigenes) to obtain percentages calculated with the same denominator, thereby following the conventions adopted by Rapp *et al.* (2009). This calculation method enables exact percentage comparisons under different conditions (i.e. temperatures and species).

$$DFI = \frac{\left| \sum_{i=1}^n (\text{Mean intensity hottest} - \text{Mean intensity coldest}) \right|}{n}$$

Eqn 2

**Comparisons between species and growth conditions as a function of 12 possible expression pattern categories** Unigenes significantly differentially expressed were binned into 12 possible expression pattern categories, as defined by Rapp *et al.* (2009), between two diploids and their derived allopolyploid. To assess the distribution of expression inten-

sities between them, we mapped the kernel density of expression for each species using the density estimator (Proc SGPLOT) in the SAS (Statistical Analysis System) software package (SAS Institute, Cary, NC, USA). These were plotted on a standardized scale against the experimental mean to illustrate allopolyploid vs diploid comparisons.

**Stability of *C. arabica* cultivars under different temperature conditions** In order to be able to compare differences in fluorescence intensity in the same genotype as a function of the per-species growth temperature conditions, a lowess normalization method was used to normalize M-values (log-ratios) for dye-bias within each array, and quantile normalization ensured that the intensities had the same empirical distribution across arrays. These two methods were implemented in the limma R package (Smyth & Speed, 2003; Smyth, 2005) of the Bioconductor project (Gentleman *et al.*, 2004). This data normalization method enabled us to conduct a global analysis of the 72 slides.

Estimates of gene expression were used to fit a linear model in the SAS software package, taking the form:

$$Y_{ij} = \mu + \delta_i + S_j + e_{ij}, \quad \text{Eqn 1}$$

( $Y_{ij}$ , the normalized expression intensity of a unigene;  $\mu$ , the intercept;  $\delta_i$ , the fixed effect of treatment  $i$  (i.e. genotype observed for a growth temperature condition) with the random effect of replication  $s_j$  and the random error term  $e_{ij}$ .) Resulting  $P$ -values were adjusted for multiple testing using the Benjamini & Hochberg's (1995) method for controlling the FDR. For a highly stringent analysis with no false positive identification of differentially regulated genes, we filtered for a  $P$ -value  $\leq 10^{-6}$ .

In order to estimate the expression stability between the two temperature conditions, we calculated two indices, called the 'fluorescence intensity differential' and the 'variation index'.

The fluorescence intensity differential was calculated between hybridizations at 26–22°C (coldest temperature) and 30–26°C (hottest temperature) using the following equation.

Fluorescence intensity differential:

( $n$ , the number of unigenes for which we noted a significant treatment effect at  $P \leq 10^{-6}$ .)

This differential was calculated for *C. canephora* and for the two allopolyploids.

An intensity variation index was also calculated by the following equation:

Variation index (in %)

$$= \left( \frac{\text{DFI}}{\text{Mean intensity coldest } T^{\circ}\text{C}} \right) \times 100 \quad \text{Eqn 3}$$

**Gene ontology functional enrichment analysis** Computational annotation was performed using BLAST2 GO software v2.4.4. (<http://www.blast2go.org>). The annotation step was performed using the BlastX algorithm, the NCBI nr database and a Blast expectation value threshold of  $1 \text{ E}^{-3}$ . The BLAST2GO tool was then used to obtain GO information from retrieved database matches. All sequences were mapped using default parameters. An InterPro Scan was also performed to find functional patterns and related GO terms using the specific tool implemented in the BLAST2GO software with the default parameters. We utilized gene ontology classifications for molecular and cellular function, coupled with Fisher's exact test, to identify processes under- or over-represented for unigenes partitioned into 'transgression' and 'dominance' categories as a function of temperature.

**Microarray data deposition** Original microarray data files have been submitted in a MIAME-compliant format to Gene Expression Omnibus under GEO Accession Number GSE24682 (<https://www.ncbi.nlm.nih.gov/geo/>).

## Results

We hybridized labelled leaf aRNAs to a 15K coffee microarray, assaying 15 522 unigenes for their relative expression levels and determined, for each unigene, the level of expression variation between parents, and between each parent and the allopolyploid. Our results are displayed according to the temperature conditions.

A new set of leaves formed in the three species after growing for 45 d under the hottest conditions and 60 d under the coldest conditions. This tallied with what is commonly found under natural conditions. The vegetative cycles of the coffee trees were more rapid at the hottest temperature. The phytotron growth conditions seemed to mimic natural growing conditions well. Leaf RNAs were then extracted and used to interrogate microarrays. We used 72 slides for this study. There were three replicates and two dye swaps

for each comparison, with a total of six comparisons carried out per temperature (Fig. S1). The analysis was therefore focused on a transcriptome comparison of coffee tree leaves from three species, *C. canephora*, *C. eugenioides* and *C. arabica*. The last species was represented by two different accessions, that is, cv Java and cv T18141.

## Assessment of DNA microarray quality

Raw quantification and background noise values were calculated for each chip (data not shown). Flags highlighted invalidated spots, which made it possible to see if there were any unusual artefacts on a slide caused, for example, by washing impurities or dust. The distributions of raw intensities, background noise and log-ratios were uniform. Few spots were flagged and background noise was low and constant when the signal intensity increased, indicating that the chips were good quality.

As 70-mer oligonucleotides designed from *C. canephora* may not hybridize well to *C. eugenioides* or *C. arabica* unigenes, we conducted a comparison of the number of unigenes expressed relative to the background noise for each species. Detection limit data were used to determine the quantity of unigenes significantly expressed in comparison to the background noise. We found that more than 8200 unigenes out of the 15 522 unigenes targeted by the microarray were significantly expressed relative to the background noise, irrespective of the species and growth temperature conditions (Table 1). We calculated that from 86.3% to 90.3% unigenes were common to the three species (Table 1). Moreover, a Pearson correlation between replicates was calculated for each unigene in all arrays, with coefficients ranging from 0.87 to 0.96 in pairwise comparisons for each independent experiment (data not shown). This high coefficient demonstrates the high level of repeatability at which the microarray is able to detect transcriptomic data.

## Transcriptome divergence between diploid parents

We used the SAM test for statistical analyses of the microarray data, with the probability set at  $P < 0.05$  for the

**Table 1** Comparison of expressed unigenes relative to the background noise

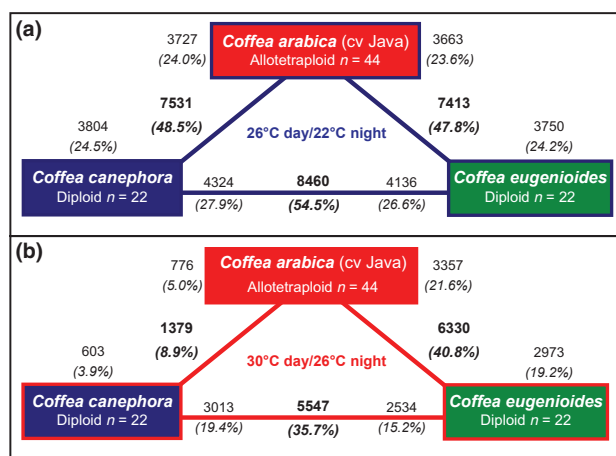
Growth temperature conditions	Detection limit threshold	Expressed unigenes relative to the background noise				Average	CV (%)	Common unigenes among the three species (%)
		<i>C. arabica</i> cv Java	<i>C. arabica</i> cv T18141	<i>C. canephora</i>	<i>C. eugenioides</i>			
26–22°C	5	8270	8324	8226	8530	8337.5	1.6	86.3
30–26°C	5	8663	9280	8990	8930	8965.7	2.8	90.3

Expressed unigenes were observed in leaves for the two *Coffea arabica* cultivars and the two parental species. Detection limit thresholds of 5 mean that, for a given gene on the six hybridizations (three replicates  $\times$  two dye-swaps), we found a signal significantly greater than the background noise at least five times.

CV, coefficient of variation.

purposes of identifying unigene expression differences between parents. For both temperatures, high levels of expression divergence were observed between parental diploids. At 26–22°C, amongst the 15 522 unigenes of the microarray, 8460 (54.5%) unigenes were differentially expressed between *C. eugenioides* and *C. canephora* (Fig. 1a). At 30–26°C, 5547 unigenes (35.7%) were differentially expressed between *C. eugenioides* and *C. canephora* (Fig. 1b). Of the differentially expressed unigenes, equivalent proportions were up-regulated in each parent, that is, 27.9% for *C. canephora* vs 26.6% for *C. eugenioides* at 26–22°C and 19.4% for *C. canephora* vs 15.2% for *C. eugenioides* at 30–26°C (Fig. 1a,b).

We sought to estimate the number of unigenes for which differential expression remained stable under both temperature conditions. The intersection of the Venn diagram in the Fig. 2 represents unigenes whose expression differential was common to both temperature conditions (3880 unigenes; Fig. 2). Of these 3880 unigenes, we found that 3033 unigenes maintained the same pattern (i.e. genes that were



**Fig. 1** Transcriptome divergence and comparison of nonadditive gene expression between *Coffea arabica* cv Java and parental diploid species. Bold text indicates the total number and fraction of unigenes that were defined as differentially expressed between species in each comparison. Nonbold text indicates the total number and fraction of unigenes that were up-regulated. For example, at 26–22°C, 7531 unigenes (48.5%) were indicated as being differentially expressed between *C. canephora* and the natural allotetraploid *C. arabica*. Of these, 3727 unigenes (24%) were up-regulated in the allotetraploid, and 3804 unigenes (24.5%) were up-regulated in *C. canephora*. (a) *Coffea canephora* and *C. eugenioides* were compared to the allotetraploid *C. arabica* cv Java at 26–22°C. Around 54.5% of the 15 522 unigenes were differentially expressed between diploids, 48.5% between *C. canephora* and the allotetraploid and 47.8% between the allotetraploid and *C. eugenioides*. (b) *Coffea canephora* and *C. eugenioides* were compared with the allotetraploid *C. arabica* cv Java at 30–26°C. Around 35.7% of the 15 522 unigenes were differentially expressed between diploids, 8.9% between *C. canephora* and the allotetraploid and 40.8% between the allotetraploid and *C. eugenioides*.

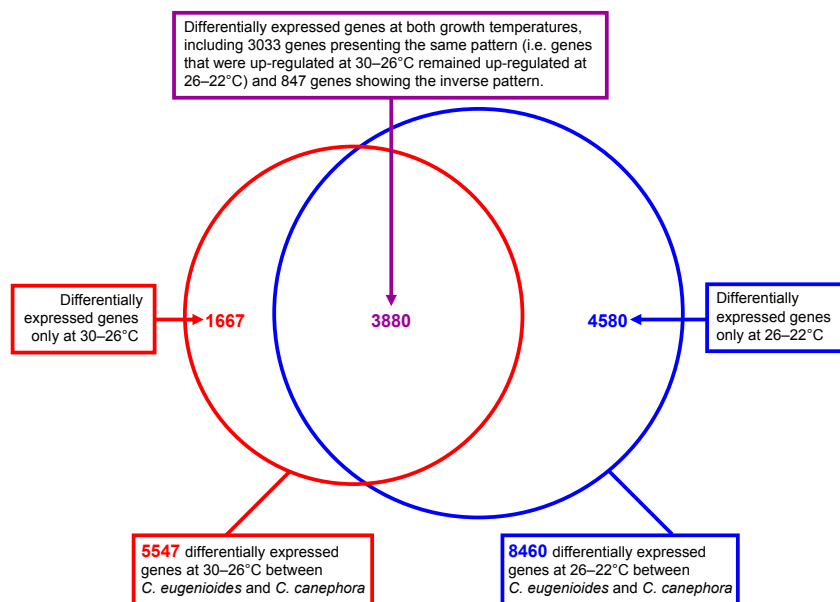
up-regulated at 30–26°C remained up-regulated at 26–22°C) under both temperature conditions (Fig. 2).

### Transcriptome divergence between diploid parents and natural allotetraploid species

In Fig. 1(a,b), we only show the results pertaining to cv Java compared with the diploid parents. The data obtained for cv T18141 were very similar and are represented in Fig. S2. At 26–22°C, the percentages of unigenes differentially expressed between the allotetraploid and its parents were high. Overall, 48.5% and 47.8% of unigenes were differentially expressed between the allotetraploid and *C. canephora*, and between the allotetraploid and *C. eugenioides*, respectively (Fig. 1a). Equivalent proportions were found at 30–26°C for the comparison between the allotetraploid and *C. eugenioides* (40.8%; Fig. 1b). However, an unexpected divergence of only 8.9% in gene expression was noted for the comparison between the allotetraploid and *C. canephora* at 30–26°C (Fig. 1b). Also, we observed that the proportions of up- and down-regulated unigenes for the same comparison varied little. For example, in *C. arabica* vs *C. canephora* at 26–22°C, 3727 unigenes (24%) were up-regulated in the allotetraploid, and 3804 unigenes (24.5%) were up-regulated in *C. canephora*, that is, a difference of 0.5% (Fig. 1a). This difference peaked at 4.2% for *C. canephora* vs *C. eugenioides* at 30–26°C (Fig. 1b).

### Patterns of differential expression between both *C. arabica* cultivars and parental species

As defined by Rapp *et al.* (2009), the unigenes were partitioned into 12 differential expression patterns. This analysis – unlike that used in the previous section – was specifically focused on the two *C. arabica* cultivars. We have presented the details of these results in Figs 3–6. These results are summarized in Fig. 7 to facilitate the analysis. At 26–22°C, the ‘*C. eugenioides*-like dominance’ and ‘*C. canephora*-like dominance’ ranged from 13 to 17% and from 14 to 8% for the allotetraploid cv Java and for the allotetraploid cv T18141, respectively (Fig. 7). There was also substantial transgressive up- and down-regulation in the allotetraploids (16–18%; Fig. 7). Lastly, only 12% of unigenes were in the ‘additivity’ category in the two allotetraploids (Fig. 7). At 30–26°C, we found that the proportion of unigenes in the ‘no change’ category increased considerably from 13–17% to 31–33% (Fig. 7). However, it was the virtual disappearance of some categories and the very marked increase in ‘*C. canephora*-like dominance’ unigenes that were most surprising. The proportion of unigenes in the ‘*C. eugenioides*-like dominance’ category declined drastically to only 0.2 to 2 for cv Java and for cv T18141, respectively (Fig. 7). The proportion of unigenes in the ‘transgression’ category also



**Fig. 2** Venn diagram representing unigenes differentially expressed between the two diploid parents (*Coffea canephora* and *Coffea eugenioides*) at both temperatures. Among the unigenes differentially expressed between the two diploid parents, we found 3880 unigenes that were common to both growth temperature conditions. 1667 and 4580 differentially expressed unigenes only at 30–26°C and 26–22°C, respectively, were specific to each of the growth temperature conditions.

decreased drastically to 3% (Fig. 7). Lastly, the proportion of unigenes in the ‘additivity’ category was under 2% for both allopolyploids (Fig. 7). Under these conditions (30–26°C), we noted dominance of the *C. canephora* transcription profile. Fig. 8 shows that the 3250 unigenes in the ‘*C. canephora*-like dominance’ category in the allopolyploid cv Java were derived from all categories of unigenes in the coldest conditions. This was also the case for unigenes in the ‘no change’ category (Fig. 8).

We also sought to determine the proportion of unigenes common to both of the studied allopolyploids. The calculation was focused on categories for which there were over 70 unigenes. At 26–22°C, the two allopolyploids shared 45–56% unigenes irrespective of the category considered (data not shown). At 30–26°C, these proportions were over 70% for the ‘*C. canephora*-like dominance’ category (i.e. the only one having a significant number of unigenes; data not shown). This high proportion of unigenes common to two very different representatives of the allopolyploid species indicates that the results obtained here could be applied to the entire species.

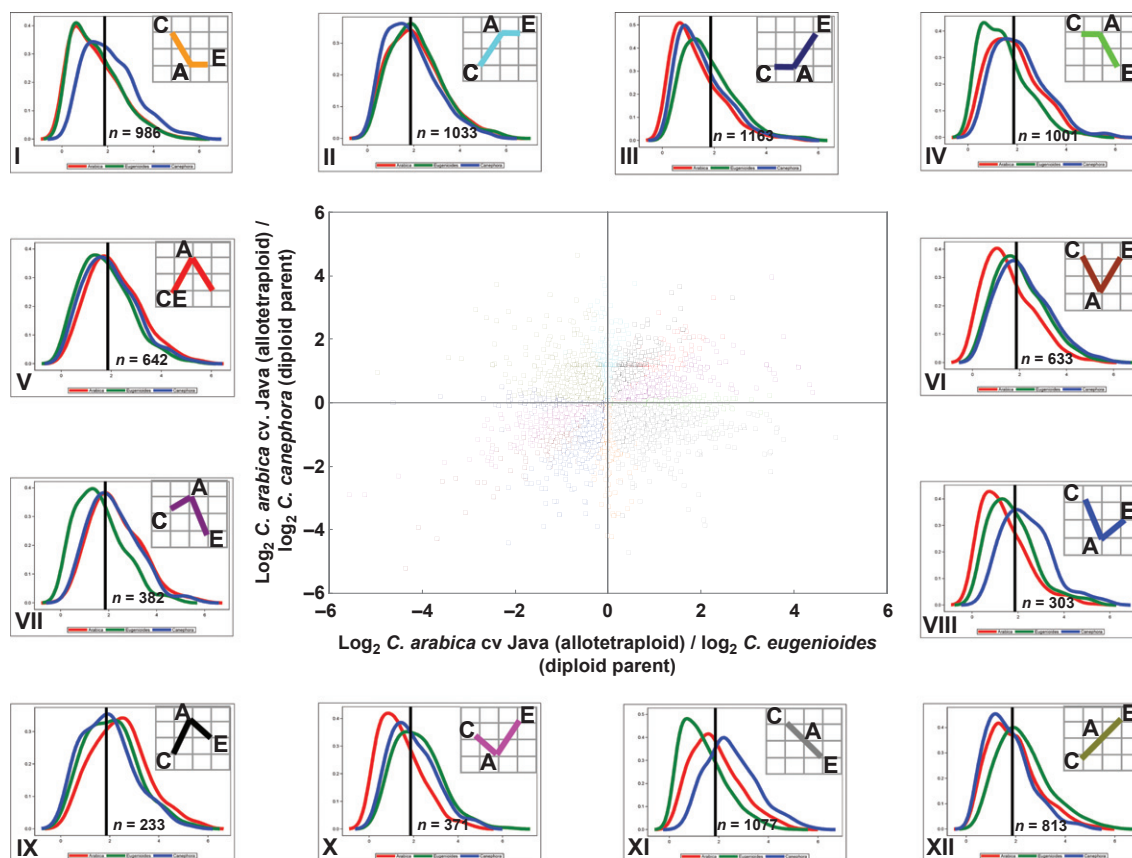
### Impact on physiology and metabolic pathways in allopolyploids

We focused our study in particular on 66 unigenes involved in several major metabolic pathways (sugar and starch degradation, lipid, phenylpropanoid and ethylene biosynthesis, stress response and circadian rhythm). For each of these unigenes, it was thus possible to monitor changes in category according to the growth temperature conditions. We noted that there were category changes for all biosynthetic pathways. However, of the 66 studied unigenes, 10 had an atypical behaviour as they remained in the same

category or shifted to a minority category when the environmental conditions changed (Table S1). For instance, this was the case for lipoxygenase (SGN-U347607) and glutathione peroxidase (SGN-U349893), which are major pathways in the oxygenation of fatty acids, acylglycerol-phosphate-acyltransferase (SGN-U351326), which is involved in membrane phospholipid biosynthesis, or caffeoyl-CoA-3-O-methyltransferase (SGN-U351599), which is involved in lignin biosynthesis and always in a transgressive situation under both hot and cold conditions (Table S1). Circadian clocks are known to affect many physiological and developmental processes, including various metabolic pathways and fitness traits in photosynthesis and starch metabolism in plants. The late elongated hypocotyl (LHY; SGN-U351840), which is one of the negative regulators of the central oscillators of the circadian clock in *Arabidopsis*, was in the ‘additivity’ category in the coldest conditions and in the ‘*C. eugenioides*-like dominance’ category in the hottest conditions (Table S1).

### Stability of *C. arabica* cultivars in different temperature conditions

The question then arises as to whether regulatory changes at high temperatures affect *C. arabica* more than *C. canephora*. Of the 15 522 unigenes, we found 4244 unigenes for which there was a significant between-treatment effect (i.e. genotype observed for a growth temperature condition; Table 2). For this subset of 4244 unigenes, a fluorescence intensity differential (DFI) was calculated for hybridizations conducted at 26–22°C and 30–26°C. This differential was calculated for Canephora and both allopolyploids. The variation index (IVar) was also calculated. On average, the absolute intensity differential value was 3.22 in *C. arabica* cv Java and 2.74 in *C. arabica* cv T18141 as compared with



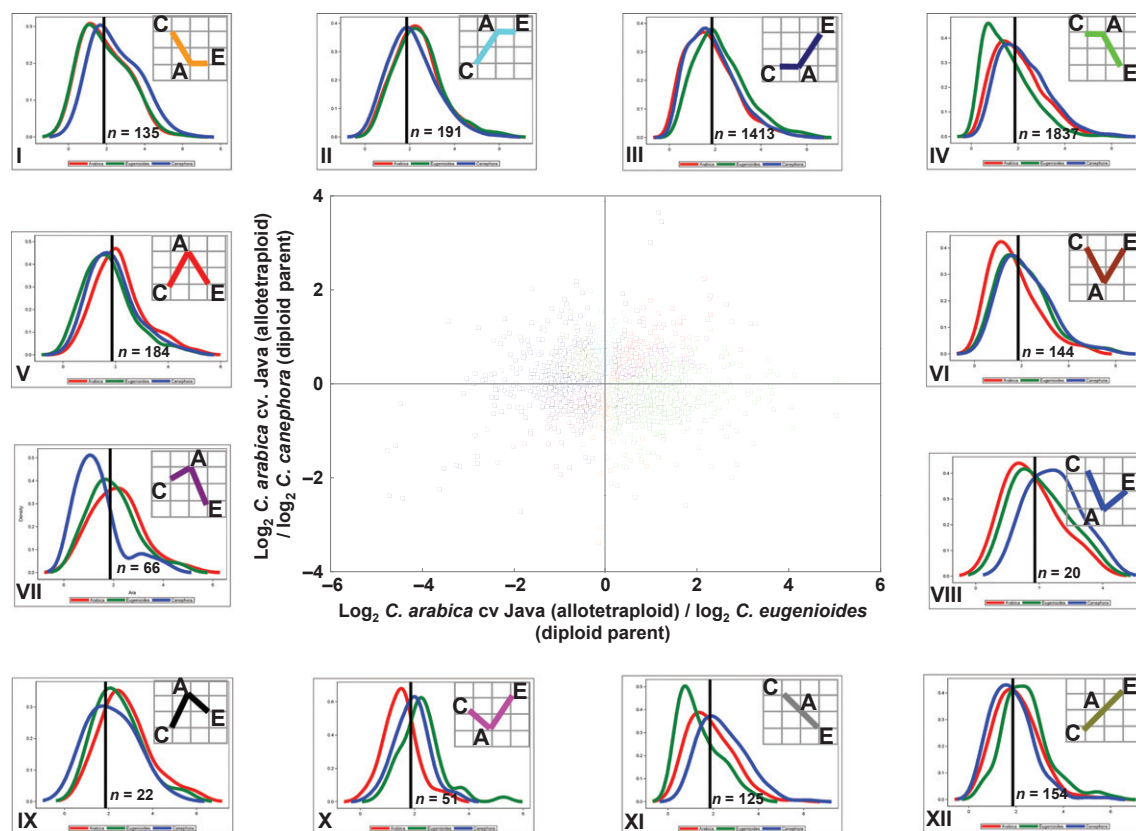
**Fig. 3** Patterns of differential expression between *Coffea arabica* cv Java allopolyploid and parental diploid species (*C. canephora* and *C. eugenioides*) at 26–22°C. Graphs split into panels I–XII represent possible expression patterns in comparisons between parents and the *C. arabica* cv Java allopolyploid, where statistically differential expression is indicated by different vertical levels in each graph. For example, panel I shows unigenes for which *C. arabica* cv Java and *C. eugenioides* were not differentially expressed and unigenes for which *C. canephora* were up-regulated relative to the allopolyploid. The total number of unigenes falling into each category is shown (*n*) and the average for the entire experiment is shown as a black horizontal line in each panel. In each of panels I to XII, the y-axis represents density distributions of expression levels for unigenes involved in that particular expression pattern (*C. canephora*, blue; *C. eugenioides*, green; cv Java allopolyploid, red) and the x-axis represents the standardized expression intensity. The ‘dominance’ group is shown in panels I–IV. Most of the unigenes (4183 unigenes of the 8637 differentially expressed in 12 patterns) belong to this group. The expression dominance of *C. eugenioides* is shown in panels I and II, whereas that of *C. canephora* is shown in panels III and IV. Variation between the two dominance expressions was only observed for a few unigenes (145 genes of the 4183 genes observed). The ‘transgression’ group is shown in panels V–X. 2564 unigenes of the 8637 differentially expressed in 12 patterns belong to this group. Panels V, VII and IX show transgressive up-regulation in allopolyploid and panels VI, VII and X show transgressive down-regulation in allopolyploids. Low variation of 50 unigenes (of the 2564 unigenes observed) was observed between the two transgressive expressions. The ‘additive’ group is shown in panels XI and XII. 1890 unigenes of the 8637 differentially expressed in 12 patterns belong to this group. The middle panel shows the distribution of expression divergence of each parent relative to the polyplody, with the *C. canephora* contrast on the y-axis and *C. eugenioides* on the x-axis, with colours corresponding to those used in the insets in panels I–XI.

3.78 in *C. canephora* (Table 2). The mean intensity variation was 26.48% in *C. arabica* cv Java and 25.26% in *C. arabica* cv T18141 as compared with 33.21% in *C. canephora* (Table 2). A second subset of 385 unigenes was created that corresponded to a ‘transgressive’ situation for allopolyploids at 26–22°C and to a ‘*C. canephora*-like dominance’ situation at 30–26°C (Table 2). On average, the absolute intensity differential value was 6.14 in *C. arabica* cv Java and 9.41 in *C. arabica* cv T18141 as compared with 16.45 in *C. canephora* (Table 2). The intensity variation between 30 and 26°C and 26–22°C for the 385 unigenes relative to the intensity at 26–22°C was, on

average, 35.60% in *C. arabica* cv Java and 53.34% in *C. arabica* cv T18141 as compared with 89.32% for *C. canephora* (Table 2). Taken together, these results suggest better homeostasis of the allopolyploids as compared with the diploid.

#### Functional enrichment analysis

In order to shed light on the processes involved under the conditions studied, we enriched the gene ontology (GO terms) within ‘dominance’ and ‘transgression’ categories for both growth temperature conditions. Table S2 shows GO

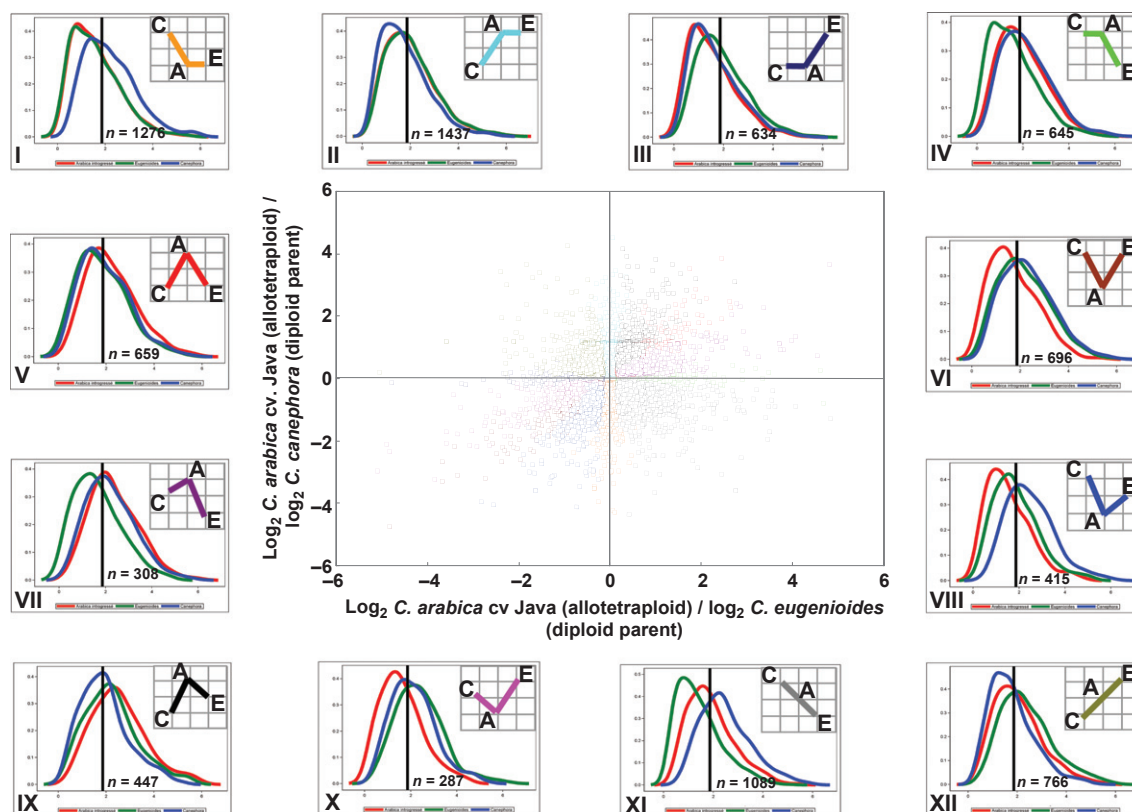


**Fig. 4** Patterns of differential expression between *Coffea arabica* cv Java allotetraploid and parental diploid species at 30–26°C. Graphs inset into panels I–XII represent possible expression patterns in comparisons between parents and the *C. arabica* cv Java allotetraploid, where statistically differential expression is indicated by different vertical levels in each graph. The total number of unigenes falling into each category is shown ( $n$ ) and the average for the entire experiment is shown as a black horizontal line in each panel. In each of panels I to XII, the y-axis represents density distributions of expression levels for unigenes involved in that particular expression pattern (*C. canephora*, blue; *C. eugenioides*, green; cv Java allotetraploid, red) and the x-axis represents the standardized expression intensity. The 'dominance' group is shown in panels I–IV. Most of the unigenes (3576 unigenes of the 4342 differentially expressed in 12 patterns) belong to this group. The expression dominance of *C. eugenioides* is shown in panels I and II, whereas the expression dominance of *C. canephora* is shown in panels III and IV. High variation of 2924 unigenes (of the 4342 unigenes observed) was observed between the two dominance expressions. Most of genes are in the '*C. canephora*-like dominance' group. The 'transgression' group is shown in panels V–X. 487 genes of the 4342 differentially expressed in 12 patterns belong to this group. Panels V, VII and IX show transgressive up-regulation in allotetraploids and panels VI, VIII and X show transgressive down-regulation in allotetraploids. Low variation of 57 unigenes (of the 487 unigenes observed) is observed between the two transgressive expressions. The 'additive' group is shown in panels XI and XII. 279 unigenes of the 4342 differentially expressed in 12 patterns belong to this group.

terms with a significantly higher frequency in category unigene sets in comparison with the full set of unigenes of the PUCE CAFE array. Unigenes dominantly transcribed in the allotetraploid were enriched for GO terms pertaining to chloroplasts, membrane activities and activities that regulate DNA replication, or which more specifically are involved in photorespiration (Table S2). This is the case of the respiratory chain complex I, which forms part of the mitochondrial respiratory chain, including the NADH dehydrogenase complex. For unigenes transgressively transcribed in the allotetraploid at the coldest temperatures, there were many enrichments of GO terms, especially for functions pertaining to chloroplasts, sugars, lipids and membranes (Table S2).

## Discussion

The tetraploid species *C. arabica* is an agricultural species of prime importance (65% of world coffee production) that is grown in > 30 intertropical countries. It is native to the tropical forests on the Abyssinian Plateau, at elevations of 1200–1950 m, with an optimum mean annual temperature range of 18–21°C (DaMatta & Cochicho Ramalho, 2006; Davis *et al.*, 2006). The diploid species *C. canephora* (35% of world coffee production) is native to lowland equatorial rainforests of the Congo River basin and tropical West Africa, extending up to Lake Victoria in Uganda at elevations of 250–1500 m (Davis *et al.*, 2006). In those regions, the annual mean temperature ranges from 22 to 26°C,

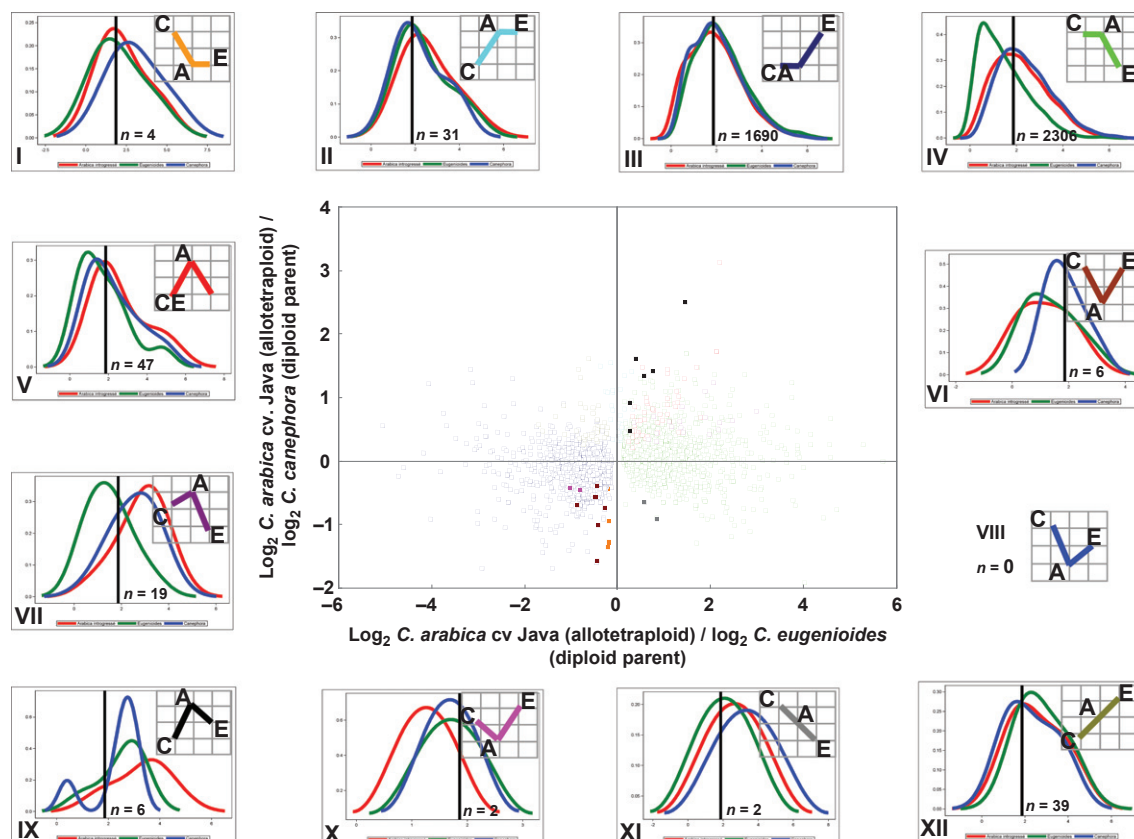


**Fig. 5** Patterns of differential expression between *Coffea arabica* cv T18141 allopolyploid and parental diploid species at 26–22°C. Graphs inset into panels I–XII represent possible expression patterns in comparisons between parents and the *C. arabica* cv T18141 allopolyploid, where statistically differential expression is indicated by different vertical levels in each graph. The total number of unigenes falling into each category is shown ( $n$ ) and the average for the entire experiment is shown as a black horizontal line in each panel. In each of panels I–XII, the y-axis represents density distributions of expression levels for unigenes involved in that particular expression pattern (*C. canephora*, blue; *C. eugenioides*, green; cv T18141 allopolyploid, red) and the x-axis represents the standardized expression intensity. The 'dominance' group is shown in panels I–IV. Most of the unigenes (3992 unigenes of the 8659 differentially expressed in 12 patterns) belong to this group. The expression dominance of *C. eugenioides* is shown in panels I and II, whereas the expression dominance of *C. canephora* is shown in panels III and IV. High variation of 1434 unigenes (of the 8659 unigenes observed) was observed between the two dominance expressions. Most of unigenes are in the '*C. eugenioides*-like dominance' group. The 'transgression' group is shown in panels V–X. 2812 unigenes of the 8659 differentially expressed in 12 patterns belong to this group. Panels V, VII and IX show transgressive up-regulation in allopolyploids and panels VI, VIII and X show transgressive down-regulation in allopolyploids. Low variation of 16 unigenes (of the 2812 unigenes observed) was observed between the two transgressive expressions. The 'additive' group is shown in panels XI and XII. 1855 unigenes of the 8637 differentially expressed in 12 patterns belong to this group.

without substantial oscillations (DaMatta & Cochicho Ramalho, 2006). The wild diploid species *C. eugenioides* is found in large or small relict forests in highland areas, at elevations of 1000–2000 m, with a mean annual temperature ranging from 18 to 23°C (Davis *et al.*, 2006). In order to monitor gene expression in the natural allopolyploid *C. arabica*, we grew the two diploid parent species (*C. canephora* and *C. eugenioides*) and two *C. arabica* cultivars in two separate growth chambers in which diurnal and nocturnal temperature conditions were set at 26–22°C for the coldest growth temperature conditions and 30–26°C for the hottest. The coldest temperatures were similar to the hottest conditions in tropical environments under which both *C. arabica* and *C. eugenioides* can be found, while the hottest temperature conditions corresponded to extreme

growth conditions under which *C. arabica* and *C. canephora* can be grown. The moderate lighting conditions were suitable for growing coffee trees, which naturally grow under canopy.

We hybridized labelled leaf aRNAs to a 15K coffee microarray, assaying 15 522 unigenes for their relative expression levels, and determined, for each unigene, the level of expression variation between parents, and between each parent and the allopolyploid. This approach is often used to estimate changes in the expression of several thousand genes in many different species (Hegarty *et al.*, 2006; Wang *et al.*, 2006; Rapp *et al.*, 2009; Fligel & Wendel, 2010). The original feature of our approach is that we compared the three species in varying environmental conditions. We varied the temperature because it is a key parameter affecting most physico-chemical reactions in



**Fig. 6** Patterns of differential expression between *Coffea arabica* cv T18141 allopolyploid and parental diploid species at 30–26°C. Graphs inset into panels I to XII represent possible expression patterns in comparisons between parents and the *C. arabica* cv T18141 allopolyploid, where statistically differential expression is indicated by different vertical levels in each graph. The total number of unigenes falling into each category is shown ( $n$ ) and the average for the entire experiment is shown as a black horizontal line in each panel. In each of panels I to XII, the y-axis represents density distributions of expression levels for unigenes involved in that particular expression pattern (*C. canephora*, blue; *C. eugenioides*, green; cv T18141 allopolyploid, red) and the x-axis represents the standardized expression intensity. Most of the unigenes (4031 unigenes of the 4152 differentially expressed in 12 patterns) belong to this group. The expression dominance of *C. eugenioides* is shown in panels I and II, whereas the expression dominance of *C. canephora* is shown in panels III and IV. High variation of 3961 unigenes (of the 4152 unigenes observed) was observed between the two dominance expressions. Most genes are in the '*C. canephora*-like dominance' group. The 'transgression' group is shown in panels V–X. Eighty unigenes of the 4125 differentially expressed in 12 patterns belong to this group. Panels V, VII and IX show transgressive up-regulation in allopolyploids and panels VI, VIII and X show transgressive down-regulation in allopolyploids. High variation of 64 unigenes (of 80 observed) was observed between the two transgressive expressions. The 'additive' group is shown in panels XI and XII. Forty-one unigenes of the 4152 differentially expressed in 12 patterns belong to this group.

plants. Our results are displayed according to the temperature conditions.

We confirmed an absence of hybridization bias, as the number of unigenes expressed relative to the background noise was the same whatever the species. The 70-mer oligonucleotides designed from *C. canephora* thus seemed to hybridize well to the *C. eugenioides* and *C. arabica* unigenes. This result could be explained by the high genic sequence identity among these three species (> 98%; A. Cenci *et al.*, unpublished).

#### Transcriptomic divergence between parents and between each parent and the allopolyploids

As all plants were grown under common controlled conditions, we expected only modest expression divergence among

diploids, but high levels of expression divergence were observed between *C. eugenioides* and *C. canephora* at both temperatures (Fig. 1a,b). Of the differentially expressed unigenes, equivalent proportions were up-regulated in each parent. These initial results showed that the transcriptomic regulation of the two species markedly differed despite the high sequence identity between the genomes (A. Cenci *et al.*, unpublished) and was modulated by growth temperature conditions. Our findings were in the same range as those obtained by Rapp *et al.* (2009), for comparisons between *Gossypium arboreum* and *Gossypium bickii* and between *G. arboreum* and *Gossypium thurberi*. In our study, equivalent proportions of the differentially expressed genes were up-regulated in each parent – this balance was also reported by Wang *et al.* (2006) and Rapp *et al.* (2009).



		No change	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
26–22°C	cv Java	2686	986	1033	1163	1001	642	633	382	303	233	371	1077	813	
	cv T18141	2130	1276	1437	634	645	659	696	308	415	447	287	1089	766	
	% by category	13–17	13–17		8–14		16–18						12		
30–26°C	cv Java	5112	135	191	1413	1837	184	144	22	20	66	51	125	154	
	cv T18141	4948	4	31	1690	2307	47	6	19	6	0	2	2	39	
	% by category	31–33	0.2–2		21–26		0.5–3						0.2–2		

**Fig. 7** The 13 possible expression patterns in two *Coffea arabica* allotetraploids relative to their progenitors (*C. canephora* and *C. eugenioides*). Each of the 13 patterns involving differential expression is labelled with a Roman numeral using the same classification as Rapp *et al.* (2009) and Flagel & Wendel (2010), and including a graphical depiction, where *C. canephora* and *C. eugenioides* are on the outer edges and the polyploid patterns in the middle. Expression values on the same horizontal line indicate statistically equivalent expression, whereas expression values on higher or lower horizontal lines represent statistically significant up- and down-regulation, respectively. For example, genes of the XI categories are those in which expression in the *C. canephora* diploid genome is significantly higher than in the *C. eugenioides* diploid genome, with gene expression in the allopolyploid being significantly different from both but with an intermediate value. The number of genes falling into the XIII gene expression category, where no significance was noted in any comparison among the two diploids and the allopolyploid, is named 'no change'. The 13 patterns are binned in five categories representing 'no change', '*C. eugenioides*-like dominance', '*C. canephora*-like dominance', 'transgression' and 'additivity'. Here these are represented by the following colour codes: yellow ('no change'), green (I and II), blue (III, IV), purple (V–X) and orange (XI, XII). Contributions per category (% per category) are given for the two allopolyploids as a percentage of the 15 522 unigenes of the microarray. For example, the green category represented 13–17% (cv Java and cv T18141, respectively) relative to the unigenes of the microarray.

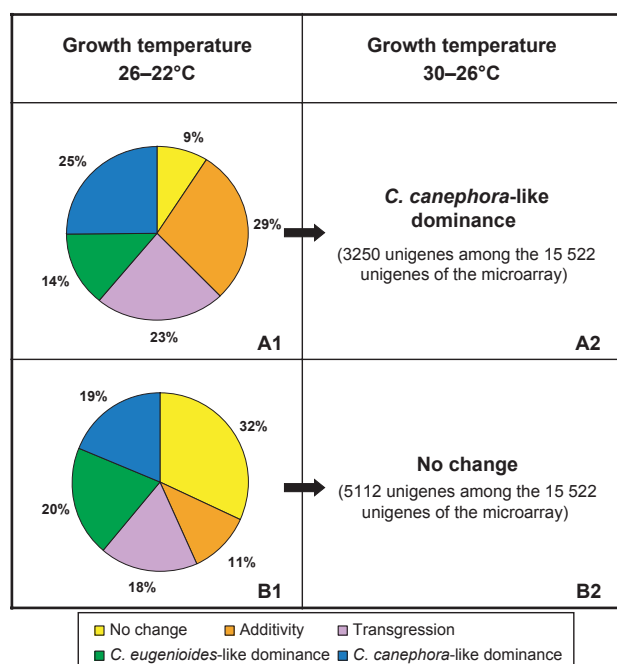
When the allopolyploid was compared to its two diploid parents, it was found that the transcriptomic divergence between two species could vary considerably depending on growing conditions, with almost half of unigenes differentially expressed between the allopolyploid and its two parental species at 26–22°C (Fig. 1a) and between the allopolyploid and *C. eugenioides* at 30–26°C (Fig. 1b), whereas unexpectedly low unigene expression divergence (< 9%) was noted for the comparison between the allopolyploid and *C. canephora* at 30–26°C (Fig. 1b). The overall results showed that the *C. arabica* and *C. canephora* transcription profiles converged as the temperature increased. It is worth noting that there was little variation in the proportions of up- and down-regulated unigenes for the same comparison.

### Impact of temperature on gene expression in allopolyploids

We used a categorically partitioned analysis of the full set of unigenes, as defined by Rapp *et al.* (2009), and were able to characterize all unigenes studied in each of the comparisons. This analysis was specifically focused on the two *C. arabica* cultivars. The virtual disappearance of some categories and the very marked increase in '*C. canephora*-like dominance' and 'no change' was most surprising. The new phenomenon revealed by our study was the drastic change in proportions when environmental conditions (i.e. growth temperature) were modified. This clearly suggests that parental regulatory genes might be modulated by allopolyploid species.

Among the lines very widely disseminated in Latin America, introgression of widely chromosome fragments from *C. canephora* by interspecific hybridization seems to have made the selected lines more adapted to hot temperature conditions (22–26°C), as compared with natural *C. arabica* lines. This is the case for cv T18141, which we studied here and whose rate of introgression by *C. canephora* was estimated using molecular markers at over 20% of the *C. canephora* genome (Lashermes *et al.*, 2000). We noted that the transcription profile of *C. arabica* cv T18141 seems to be more similar to that of its parent *C. canephora* than that of the pure *C. arabica* variety 'Java', which makes sense considering the introgression of T18141 by chromosome fragments from *C. canephora*.

We found that the behaviour at 30–26°C of the two majority unigene categories (i.e. '*C. canephora*-like dominance' and 'no change') derived from all categories of unigenes in the coldest conditions (Fig. 8). We particularly focused our study on 66 unigenes involved in several major metabolic pathways (Wang *et al.*, 2006; Salmona *et al.*, 2008; Joët *et al.*, 2009; Ni *et al.*, 2009) and showed that no biosynthesis pathway was uniformly classified in any category and that, when the growth temperature conditions varied, there were category changes for all biosynthesis pathways. However, there was a minority of genes showing atypical behaviour. Those genes would be interesting targets for testing the hypothesis that the higher plasticity of allopolyploids might be explained by a few genes in particular biological pathways, as suggested by Ni *et al.* (2009). In the



**Fig. 8** Behaviour at 30–26°C of the two majority unigene categories, that is, '*C. canephora*-like dominance' and 'no change'. The proportions observed in blue, green, orange, purple and yellow represent the '*C. canephora*-like dominance', '*C. eugenioides*-like dominance', 'additivity', 'transgression' and 'no change' groups, respectively. (A1, A2) The proportion of unigenes of the different categories at 26–22°C which became '*C. canephora*-like dominance' at 30–26°C. Of the 3250 unigenes belonging to the category '*C. canephora*-like dominance' at 30–26°C, 29% were from the category 'additivity', 25% from '*C. canephora*-like dominance', 23% from 'transgression', 14% from '*C. eugenioides*-like dominance' and 9% from 'no change' at 26–22°C. (B1, B2) The proportion of unigenes of the different categories at 26–22°C which became 'no change' at 30–26°C. Of the 5112 unigenes belonging to the category 'no change' at 30–26°C, 32% were from the category 'no change', 20% from '*C. eugenioides*-like dominance', 19% from '*C. canephora*-like dominance', 18% from 'transgression' and 11% from 'additivity' at 26–22°C.

functional enrichment analysis, we found that the enrichments varied at the coldest temperature and that there was considerable depletion at the highest temperature in both of the categories studied ('dominance' and 'transgression'), which suggests stressful growing conditions.

Lastly, we observed the stability of *C. arabica* cultivars under different temperature conditions. The allopolyploids seemed to display a higher level of homeostasis than their *C. canephora* parent. It is generally accepted that allopolyploids exhibit greater phenotypic plasticity than their parents. Our results seem to suggest that their greater phenotypic plasticity is based on better homeostasis of gene expression. To explain this mechanism, we proposed that the relative contribution of homeologues to the transcriptome varied with growth temperature conditions, but maintained the same global genetic expression level between the two temperatures. This would mean that at the cold temperature one

**Table 2** Fluorescence intensity differential (DFI) and the variation index (IVar) between parental *Coffea canephora* diploid species and *C. arabica* allopolyploids (cv Java and cv T18141)

	<i>C. arabica</i> cv Java	<i>C. arabica</i> cv T18141	<i>C. canephora</i>
Subset of 4244 unigenes			
DFI	3.22	2.74	3.78
IVar (%)	26.48	25.26	33.21
Subset of 385 unigenes			
DFI	6.14	9.41	16.45
IVar (%)	35.60	53.34	89.32

Of the 15 522 unigenes, a significant between-treatment effect was noted for 4244 unigenes (i.e. genotype observed for a growth temperature condition). A second subset of 385 unigenes corresponding to a 'transgressive' situation for allopolyploids at 26–22°C and to a '*C. canephora*-like dominance' situation at 30–26°C was created. DFI, fluorescence intensity differential between hybridizations at 26–22°C and those at 30–26°C in absolute values (see equation in the Materials and Methods section). IVar (Eqn 3), variation index in % (see equation in the Materials and Methods section).

homeologue gene is mostly recruited to be expressed whereas the other is recruited little or not at all, and that the opposite situation would be observed at the hot temperature. However, the global level of expression would vary little in the allopolyploid and a certain 'stability' would be observed.

## Conclusion

We studied the transcription profile of a relatively recent natural allopolyploid species and compared this profile with that of its parental species under two temperature conditions during plant growth. By employing a microarray technology for the analysis of 15 522 coffee unigenes, amounting to around half of all coffee tree unigenes, we reached the conclusion that transcriptome divergence in *C. arabica*, in comparison with its two diploid parents, is modulated by the environment. This finding was not completely unexpected, as it was previously reported that the pattern of parental dominance varies from tissue to tissue (Chaudhary *et al.*, 2009) and under stress (Liu & Adams, 2007). However, it is the extent of the phenomenon that is remarkable. The capacity of an allopolyploid to play on two subgenomes and to adapt the score to the prevailing environmental conditions distinguishes it from its parents. The *C. arabica* allopolyploid species showed higher homeostasis depending on environmental conditions as compared with the variability noted in the diploid *C. canephora* species. In our view, this phenomenon could explain the greater plasticity of allopolyploids compared with their parental species in coping with environmental variations and might ultimately explain their better adaptation to environmental conditions.

Our global gene expression analysis did not allow us to determine the relative contributions of homeologues to the

transcriptome. To be able to do this, a microarray technology needs to be developed using homeologue-specific (i.e. parent-specific) probe sets or by sequencing as was done by Vidal *et al.* (2010). Such technologies would enable us to differentiate cases in which similar levels of expression are attained via biased parental expression or via equal expression of the homeologous parental genomes. It would thus be possible to determine whether or not one subgenome is massively expressed to the detriment of the other subgenome, as suggested by Flagel & Wendel (2010), and whether this balance of subgenomic contributions is environment-dependent.

It is still unclear what mechanisms are involved in the modulation of parental regulatory genes in allopolyploid species. Several nonexclusive hypotheses could be put forward. For instance, as suggested by Ni *et al.* (2009), a few regulatory genes whose activity is dependent on environmental conditions could lead to a cascade of changes in downstream genes and physiological pathways in polyploids. But how are these regulatory genes activated? Ha *et al.* (2009) suggested that expression variation of miRNAs leads to changes in gene expression, growth vigor and adaptation.

## Acknowledgements

We thank C. Dantec and D. Severac (CNRS, Institut de Génomique Fonctionnelle, Montpellier) for microarray generation and quality control. This research was supported by a grant from the Agence Nationale de la Recherche (ANR; Génoplatte GPLA 06010 G). All the authors have read the manuscript and agree with the contents.

## References

- Abbott RJ, Lowe AJ. 2004. Origins, establishment and evolution of new polyploid species: *Senecio cambrensis* and *Senecio eboracensis* in the British Isles. In: Leitch AR, Soltis DE, Soltis PS, Leitch IJ, Pires JC, eds. *Biological relevance of polyploidy: ecology to genomics*. *Biological Journal of the Linnean Society* 82: 467–474.
- Adams KL. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. *Journal of Heredity* 98: 136–141.
- Ainouche M, Baumel A, Salmon A, Yannic G. 2003. Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytologist* 161: 165–172.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57: 289–300.
- Buggs RJA. 2008. Towards natural polyploid model organisms. *Molecular Ecology* 17: 1875–1876.
- Charrier A, Berthaud J. 1985. Botanical classification of coffee. In: Clifford MN, Wilson KC, eds. *Coffee: botany, biochemistry and production of beans and beverage*. London, UK: Croom Helm, 13–47.
- Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF. 2009. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* 182: 503–517.
- Chen ZJ, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* 28: 240–252.
- DaMatta FM, Cochicho Ramalho JD. 2006. Impacts of drought and temperature stress on coffee physiology and production: a review. *Brazilian Journal Plant Physiology* 18: 55–81.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P. 2006. An annotated taxonomic conspectus of genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* 152: 465–512.
- Davis AP, Rakotonasolo F. 2008. A taxonomic revision of the baracoffea alliance: nine remarkable *Coffea* species from western Madagascar. *Botanical Journal of the Linnean Society* 158: 355–390.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using *phred*. I. Accuracy assessment. *Genome Research* 8: 175–185.
- Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytologist* 186: 184–193.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al.* 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5: R80.
- Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ. 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy Sciences, USA* 106: 17835–17840.
- Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ. 2008. Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. *Philosophical Transactions of the Royal Society B* 363: 3055–3069.
- Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. 2006. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Current Biology* 16: 1652–1659.
- Hegarty MJ, Hiscock SJ. 2009. The complex nature of allopolyploid plant genomes. *Heredity* 103: 100–101.
- Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Research* 9: 868–877.
- Jackson S, Chen ZJ. 2009. Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* 13: 1–7.
- Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C, Nordborg M. 2006. A Unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Molecular Biology and Evolution* 23: 1217–1231.
- Joët T, Laffargue A, Salmons J, Doubeau S, Descroix F, Bertrand B, De Kochko A, Dussert S. 2009. Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. *New Phytologist* 182: 146–162.
- Lashermes P, Andrzejewski S, Bertrand B, Combes MC, Dussert S, Grasiozi G, Trouslot P, Anthony F. 2000. Molecular analysis of introgressive breeding in coffee (*Coffea arabica*). *Theoretical and Applied Genetics* 100: 139–146.
- Lashermes P, Carvalho Andrade A, Etienne H. 2008. Genomics of coffee, one of the world's largest traded commodities. In: Moore PH, Ming R, eds. *Genomics of tropical crop plants*. New York, NY, USA: Springer, 203–224.
- Lashermes P, Combes MC, Ribas A, Cenci A, Mahé L, Etienne H. 2010. Genetic and physical mapping of the SH3 region that confers resistance to leaf rust in coffee tree (*Coffea arabica* L.). *Tree Genetics and Genomes* 6: 973–980.
- Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A. 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics* 261: 259–266.
- Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* 320: 481–483.

- Liu Z, Adams KL. 2007. Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Current Biology* 17: 1669–1674.
- Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457: 327–331.
- Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V *et al.* 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends Genetics* 19: 141–147.
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.
- Privat I, Bardil A, Bombarely Gomez A, Severac D, Dantec C, Fuentes I, Mueller L, Joët T, Pot D, Foucrier S *et al.* 2011. The ‘PUCE CAFÉ’ project: the first 15K coffee microarray, a new tool for discovering candidate genes correlated to agronomic and quality traits. *BioMed Central Genomics* 12: 5.
- Pumphrey M, Bai M, Laudencia-Chinguanco D, Anderson O, Gill BS. 2009. Nonadditive expression of homeologous genes is established upon polyploidisation in hexaploid wheat. *Genetics* 181: 1147–1157.
- Rapp RA, Udall JA, Wendel JF. 2009. Genomic expression dominance in allopolyploids. *BioMed Central Biology* 7: 18.
- Salmona J, Dussert S, Descroix F, de Kochko A, Bertrand B, Joët T. 2008. Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. *Plant Molecular Biology* 66: 105–124.
- Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution* 20: 633–643.
- Smyth GK. 2005. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, eds. *Bioinformatics and computational biology solutions using R and Bioconductor*. New York, NY, USA: Springer, 397–420.
- Smyth GK, Speed TP. 2003. Normalization of cDNA microarray data. *Methods* 31: 265–273.
- Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E. 2004. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. In: Leitch AR, Soltis DE, Soltis PS, Leitch IJ, Pires JC, eds. *Biological relevance of polyploidy: ecology to genomics*. *Biological Journal of Linnean Society* 82: 485–501.
- Tusher V, Tibshirani R, Chu C. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy Sciences, USA* 98: 5116–5121.
- Vidal RO, Costa Mondego JM, Pot D, Ambrosio AB, Carvalho Andrade A, Protasio Pereira LF, Colombo CA, Gonzaga Esteves Vieira L, Falsarella Carazzolle M, Amarante Guimarães Pereira G. 2010. A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiology* 154: 1053–1066.
- Wang JL, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L *et al.* 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172: 507–517.

## Supporting Information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Microarray experimental design.

**Fig. S2** Transcriptome divergence and nonadditive gene expression between *Coffea arabica* cv T18141 and parental diploid species.

**Table S1** Monitoring selection of differential responses of 66 unigenes in two growth temperature conditions

**Table S2** BLAST2GO annotation for genes differentially expressed in the ‘dominance’ and ‘transgression’ categories

**Table S3** The missing genes

**Table S4** Oligonucleotides that may cross-hybridize with several sequences

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

# CHAPITRE II

## PARTIE II

**DISTANCE DU TRANSCRIPTOME DE L'ESPECE  
ALLOPOLYPLOÏDE (*C. ARABICA*) PAR RAPPORT SES DEUX ESPECES  
PARENTALES (*C. CANEPHORA* ET *C. EUGENIOIDES*) SOUS CONTRAINTES  
THERMIQUES**

	C. eugenioides 30-26°C	C. eugenioides 26-22°C	C. canephora 30-26°C	C. canephora 26-22°C	C. arabica cv. Java 30- 26°C	C. arabica cv. Java 26- 22°C	C. arabica cv. T18141 30-26°C	C. arabica cv. T18141 26-22°C
C. eugenioides 30-26°C	0	1137	834	1603	898	1240	763	1136
C. eugenioides 26-22°C	2.30	0	1358	1354	1393	978	1275	885
C. canephora 30-26°C	3.06	2.30	0	1564	775	1173	506	1031
C. canephora 26-22°C	2.38	2.26	2.56	0	1727	1122	1465	1240
C. arabica cv. Java 30-26°C	2.86	2.44	2.96	2.38	0	1305	595	1201
C. arabica cv. Java 26-22°C	2.84	2.52	2.39	3.37	2.30	0	1072	574
C. arabica cv. T18141 30- 26°C	3.07	2.22	3.41	2.56	2.08	2.32	0	976
C. arabica cv. T18141 26- 22°C	2.95	2.88	2.39	3.01	2.43	2.54	2.27	0

**Tableau 1 : Matrice des distances euclidiennes.**

En vert sont indiquées les distances euclidiennes. En rouge sont indiqués les coefficients de variation calculés par bootstrap.

## INTRODUCTION

Comme indiqué dans le paragraphe « Stability of *C. arabica* cultivars under different temperature conditions » de l'article paru dans *New Phytologist*, le dispositif expérimental de la Puce ne prévoit pas de comparaisons entre les conditions de température. Les données ont donc été normalisées *in silico* pour pouvoir les réaliser. Sur les 15522 unigènes de la Puce, nous avons sélectionné uniquement les unigènes pour lesquels nous obtenions des signaux stables et répétables. Pour chaque espèce, nous avons calculé le coefficient de variation sur l'ensemble des réponses (2 dye-swaps x 6 lames=12 réponses) obtenues pour chaque unigène à une condition de température. Nous avons conservé les unigènes pour lesquels le coefficient de variation du niveau d'expression était inférieur à 50%. Au total, 11986 unigènes ont été sélectionnés avec, pour certains, des niveaux d'expression strictement équivalents entre espèces et pour d'autres une variation significative entre la condition 26-22°C et la condition 30-26°C et cela pour une espèce donnée ou entre deux espèces. Au cours de cette deuxième partie de chapitre, la condition de température 30-26°C qui sera considérée comme un stress thermique chaud pour les trois espèces étudiées et la condition de température 26-22°C qui sera utilisée comme un témoin tout en sachant qu'il s'agit d'une condition de température adéquate pour les trois espèces.

## RESULTATS

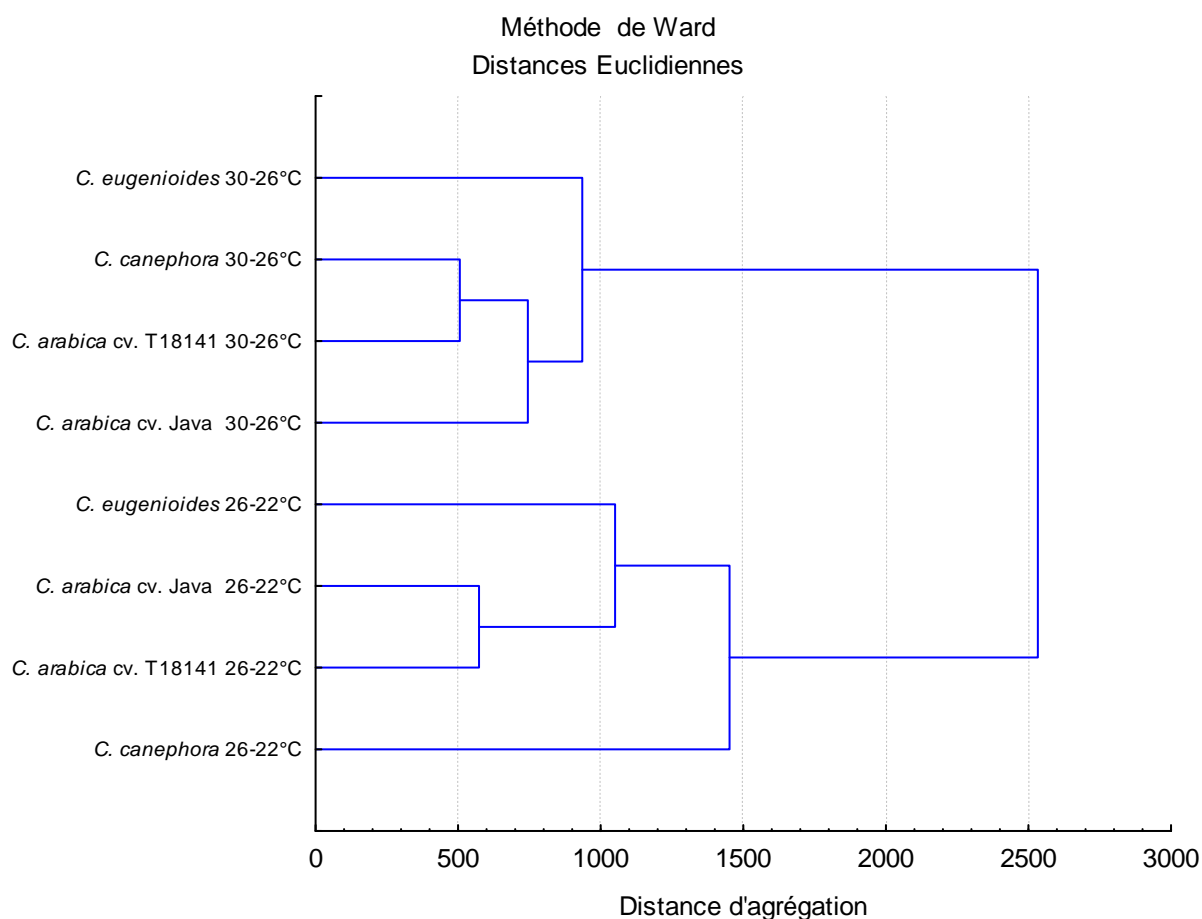
### Mesures de la distance euclidienne basée sur les variations du transcriptome des espèces en fonction des conditions de températures

Dans une récente étude de Stupar *et al.* (2008), les auteurs montrent que la diversité génétique et les variations du transcriptome peuvent être reliées. En effet, l'abondance des transcrits représente un ensemble de données quantitatives qui caractérise un génotype. Dans notre étude, cette abondance est représentée par une espèce et pour une condition de température de croissance. Les variations du transcriptome sont utilisées pour calculer les distances euclidiennes entre les espèces en fonction des conditions de température. Nous avons choisi de faire des mesures (méthode de Ward) basées sur la distance entre deux classes (barycentres au carré) et pondérées par les effectifs des deux clusters.

La distance euclidienne est calculée de la façon suivante :

$$\sqrt{\sum_{i=1}^n |x_i - y_i|^2}$$

La matrice de distance a été calculée par bootstrap (Tableau. 1).



**Figure 3 : Analyse hiérarchique descendante basée sur les distances euclidiennes entre les transcriptomes.**

Les distances sont déterminées sur la base des 11986 unigènes exprimés par rapport au bruit de fond.

Gènes différentiellement exprimés au seuil de probabilité $P < 10^{-7}$	<i>C. eugenioides</i>	<i>C. canephora</i>	<i>C. arabica</i> cv. Java	<i>C. arabica</i> cv. T18141
Gènes sur-exprimés à 30-26°C	904	1086	1316	875
Gènes sous-exprimés à 30-26°C	844	1151	1176	968

**Tableau 2 : Nombre de gènes sur- ou sous-exprimés à la condition de température 30-26°C au seuil  $P < 10^{-7}$**

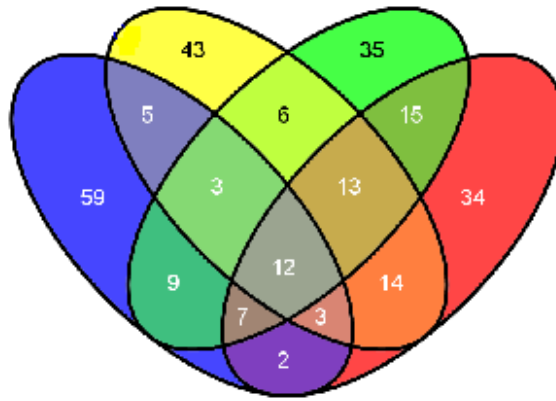
Le classement est effectué par espèce et par variation de l'expression génique.



Les distances calculées ont des coefficients de variation faibles (cv maximum = 3%) traduisant la répétabilité de l'estimation. Les distances calculées permettent d'établir une classification ascendante hiérarchique (CAH) robuste. La CAH permet de distinguer, au sein d'un échantillon, des sous-populations aussi homogènes que possible et d'en mesurer les proximités à partir de variables quantitatives. La classification obtenue montre que les transcriptomes divergent en fonction des conditions de température de croissance (Figure 3). On obtient donc deux groupes de profils transcriptomiques en fonction des conditions de température. Dans ces deux groupes distincts, les relations entre les espèces sont modifiées. En ce qui concerne les deux variétés de *C. arabica*, elles sont proches à la condition de température 22-26°C. A contrario, on note une variation à la température 30-26°C, montrant que l'espèce diploïde, *C. canephora* est plus proche de *C. arabica* cv T18141 que de *C. arabica* cv Java. On observe également que le transcriptome de l'espèce allopolyploïde est plus proche du transcriptome de *C. canephora* que de celui de *C. eugenioides* à la condition de température 30-26°C, alors qu'un comportement inverse est observé à la condition de température 26-22°C.

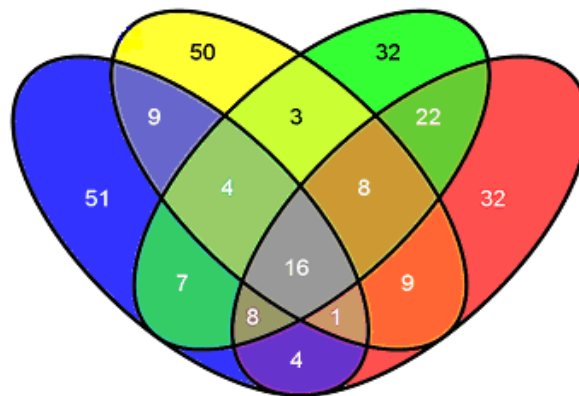
### **Détermination pour les unigènes les plus exprimés de ceux qui sont communs ou exclusifs aux trois espèces**

Parmi les 11986 unigènes exprimés par rapport au bruit de fond, nous avons déterminés les unigènes pour lesquels on observe une sur-expression ou une sous-expression à la condition la plus chaude 30-26°C par rapport à la température la plus froide 26-22°C en imposant un seuil de probabilité élevé ( $p < 10^{-7}$ ). On constate qu'il y a autant d'unigènes qui sont sur-exprimés que sous-exprimés et que leur nombre est du même ordre de grandeur quelle que soit l'espèce étudiée (Tableau 2). Ce qui semble indiquer qu'il n'y a pas de biais. En effet, en présence d'un biais dû à une normalisation inefficace des données ou s'attendrait à voir plus de gènes sur-ou sous-exprimés à une même condition de température pour chacune des espèces. Nous avons choisi d'étudier la centaine d'unigènes très fortement sur- ou sous-exprimée à la condition de température 30-26°C pour chacune des espèces. Nous observons des ratios  $\text{Log}_2$  (foldchange) qui vont de 2.34 à 27.85 et dont les seuils de probabilité sont toujours inférieurs à  $9.99 \times 10^{-8}$  (allant jusqu'à  $9.99 \times 10^{-22}$ ).



**Figure 4. : Diagramme de Venn des unigènes sur-exprimés en condition de température de 30-26°C par rapport à la condition de température de 22-26°C pour chacune des espèces.**

Le cercle bleu représente les unigènes sur-exprimés de *C. canephora*, le cercle jaune représente les unigènes sur-exprimés de *C. eugenioides*, le cercle vert représente les unigènes sur-exprimés de *C. arabica* cv. Java et le cercle rouge représente les unigènes sur-exprimés de *C. arabica* cv. T18141. On constate que 12 unigènes sont communs aux trois espèces étudiées. On obtient également 59 unigènes 'spécifiques à l'espèce' chez *C. canephora*, 43 unigènes 'spécifiques à l'espèce' chez *C. eugenioides*, 35 unigènes 'spécifiques à l'espèce' chez *C. arabica* cv. Java et 34 unigènes 'spécifiques à l'espèce' chez *C. arabica* cv. T18141.



**Figure 5. : Diagramme de Venn des unigènes sous-exprimés en condition de température de 30-26°C par rapport à la condition de température de 22-26°C pour chacune des espèces.**

Le cercle bleu représente les unigènes sous-exprimés de *C. canephora*, le cercle jaune représente les unigènes sous-exprimés de *C. eugenioides*, le cercle vert représente les unigènes sous-exprimés de *C. arabica* cv. Java et le cercle rouge représente les unigènes sous-exprimés de *C. arabica* cv. T18141. On constate que 16 unigènes sont communs aux trois espèces étudiées. On obtient également 51 unigènes 'spécifiques à l'espèce' chez *C. canephora*, 50 unigènes 'spécifiques à l'espèce' chez *C. eugenioides*, 32 unigènes 'spécifiques à l'espèce' chez *C. arabica* cv. Java et 32 unigènes 'spécifiques à l'espèce' chez *C. arabica* cv. T18141.

Les unigènes filtrés (les gènes à sur- ou sous-expression les plus fortes) ont été classés dans des ensembles de Venn. Deux diagrammes ont été produits en fonction de la sous- ou sur-expression dans la condition de température 30-26°C (Figures 4 et 5). On constate que la plupart des unigènes sont 'espèce-spécifiques' et que très peu d'unigènes sont communs à toutes les espèces. Seulement 12 unigènes fortement sur-exprimés à 30-26°C apparaissent communs aux trois espèces et 16 unigènes fortement sous-exprimés à 30-26°C (Figures 4 et 5).

### **Les processus biologiques impliqués dans les groupes d'unigènes communs aux trois espèces ou 'spécifiques' à chacune d'elle**

A partir de la centaine de gènes choisi pour chacune des espèces, nous voulons vérifier si la transcription d'unigènes sur ou sous-exprimés en conditions de stress de chaleur (30-26°C) révèle des processus biologiques spécifiques liés aux stress thermiques. Pour observer ces processus biologiques, nous les avons classés en deux groupes :

- Les unigènes communs aux trois espèces dont 12 unigènes sur-exprimés à la condition de stress de chaleur (30-26°C) et 16 unigènes sous-exprimés à la même condition de température.
- Les unigènes spécifiques à chacune des espèces. Pour l'espèce *C. arabica* nous avons choisi d'étudier les unigènes communs aux deux variétés dont 15 unigènes sur-exprimés à la condition de stress de chaleur (30-26°C) et 22 unigènes sous-exprimés à la même condition de température.

Les informations, générées par l'analyse d'enrichissement, figurent dans les tableaux de données de l'annexe 10. Les résultats obtenus sont riches et difficiles à interpréter. Nous analyserons uniquement les processus biologiques les plus représentés pour chacune des catégories analysées.

#### **Unigènes communs aux trois espèces de caféiers étudiés**

Dans le groupe des 12 unigènes sur-exprimés sous stress de chaleur (30-26°C), on trouve un grand nombre d'unigènes liés au processus de la photosynthèse dont les chromophores contrôlant plusieurs processus tels que le réglage de l'horloge interne, le contrôle des rythmes endogènes ainsi que la biosynthèse de diverses substances telles que les anthocyanes, flavones, ou encore des protéines. On trouve également un grand nombre

d'unigènes liés aux processus de la voie de biosynthèse des diterpènes avec des unigènes ayant un rôle dans la voie de signalisation de l'acide gibbérellique (diterpène tétracyclique) favorisant l'élongation cellulaire et la croissance racinaire. Dans le groupe des 16 unigènes sous-exprimés, on observe que la plupart des unigènes appartiennent à des processus biologiques liés à la régulation négative de la transcription par le promoteur de la polymérase III.

#### **Unigènes spécifiques à l'espèce parentale diploïde, *C. canephora***

Dans le groupe des 59 unigènes sur-exprimés sous stress de chaleur (30-26°C), on note quatre processus principaux. La biosynthèse des polysaccharides des parois (ex : glucan et pyrimidine biosynthetic process), la biosynthèse de la chlorophylle (ex : porphyrin et tetrapyrrole biosynthetic process), la biosynthèse des acides gras et des lipides (ex : dicarboxylic acid metabolic process) et des processus biologiques liés à la résistance (ex : herbicide, toxin, oxoacid and chorismate). Dans le groupe des 51 unigènes sous-exprimés sous stress de chaleur (30-26°C), on trouve des processus biologiques qui concernent essentiellement la réponse aux stimuli abiotiques ainsi que la voie de biosynthèse des lipides.

#### **Unigènes spécifiques à l'espèce parentale diploïde, *C. eugenoides***

Dans le groupe des 43 unigènes sur-exprimés sous stress de chaleur (30-26°C), on trouve des processus liés aux stress oxydatifs et la voie de biosynthèse des alcaloïdes. Dans le groupe des 50 unigènes sous-exprimés sous stress de chaleur (30-26°C), on trouve quatre principaux processus liés aux acides aminés (proline, sérine), aux vitamines (notamment riboflavine), aux polyamines dont l'accumulation est souvent mentionnée lors de stress abiotiques et de stress oxydatifs (oxydo-réduction or electron carrier activity).

#### **Unigènes 'spécifiques' à l'espèce allopolyploïde, *C. arabica***

Dans le groupe des 15 unigènes sur-exprimés sous stress de chaleur (30-26°C), on retrouve, des processus liés à la biosynthèse de la chlorophylle et des triglycérides comme chez *C. canephora*, et des processus liés aux hormones (acide gibbérellique et acide abscissique) et à la régulation de la méthylation, des transporteurs ou encore des voies de signalisation. Dans le groupe des 22 unigènes sous-exprimés sous stress de chaleur (30-26°C), on trouve des processus liés à la biosynthèse des lipides et notamment des phospholipides et de facteurs de transcription (RNA elongation).

## DISCUSSION et CONCLUSION

Les variations observées entre les transcriptomes des trois espèces de l'étude ont permis de calculer des distances euclidiennes entre ces espèces en fonction des deux conditions de température. Les résultats obtenus montrent que les distances observées entre les conditions de température sont plus importantes que celles observées entre les espèces et que ces variations dans les distances forment deux groupes 'température-dépendants' bien distincts. Dans ces deux groupes distincts, les relations entre les espèces sont modifiées. En ce qui concerne les deux variétés de *C. arabica*, elles sont proches de la condition de température 22-26°C. On note une variation à la température 30-26°C, montrant que l'espèce diploïde, *C. canephora* est plus proche de *C. arabica* cv T18141 que de *C. arabica* cv Java. *C. arabica* cv T18141 est issu de lignées homozygotes *C. arabica* dérivant d'un rétrocroisement entre *C. arabica* et un hybride interspécifique entre *C. arabica* et *C. canephora*. L'introggression par *C. canephora* est estimée, par marqueurs moléculaires, à plus de 20% du génome de *C. canephora* (Lashermes *et al.*, 2000). Ces lignées cultivées en Amérique latine montrent une meilleure adaptation que les lignées pures *C. arabica* cv Java aux températures élevées. Dans le chapitre II partie I, nous avons déjà souligné que le profil transcriptomique de *C. arabica* cv T18141 semblait plus proche du profil transcriptomique de *C. canephora* que du profil transcriptomique de *C. arabica* cv Java. Des modifications de distances entre l'allopolyplôïde et ses espèces parentales aux deux conditions de température ont également été observées. Le transcriptome de l'espèce allopolyplôïde est plus proche du transcriptome de *C. canephora* que de celui de *C. eugenioides* à la condition de température 30-26°C, alors qu'un comportement inverse est observé à la condition de température 26-22°C. Ce résultat a également été observé dans le chapitre II partie I, par une approche sur la divergence transcriptomique entre *C. arabica* et *C. canephora* à la température la plus chaude (30-26°C) était faible (<9%) comparé à la divergence observée à la température froide (26-22°C ; >40%).

Parmi les 11986 unigènes exprimés par rapport au bruit de fond, autant de gènes sont sous- et sur-exprimés à la condition de stress thermique (30-26°C) quelles que soit l'espèce étudiée. Cela montre une absence de biais due à la normalisation des 15522 unigènes de la Puce. Malgré tout, il faut tout de même prendre ces résultats obtenus avec précaution sachant que la Puce utilisée n'a pas été conçue pour des comparaisons entre espèces entre conditions de température. Lorsqu'on étudie les gènes différentiellement exprimés entre les deux conditions

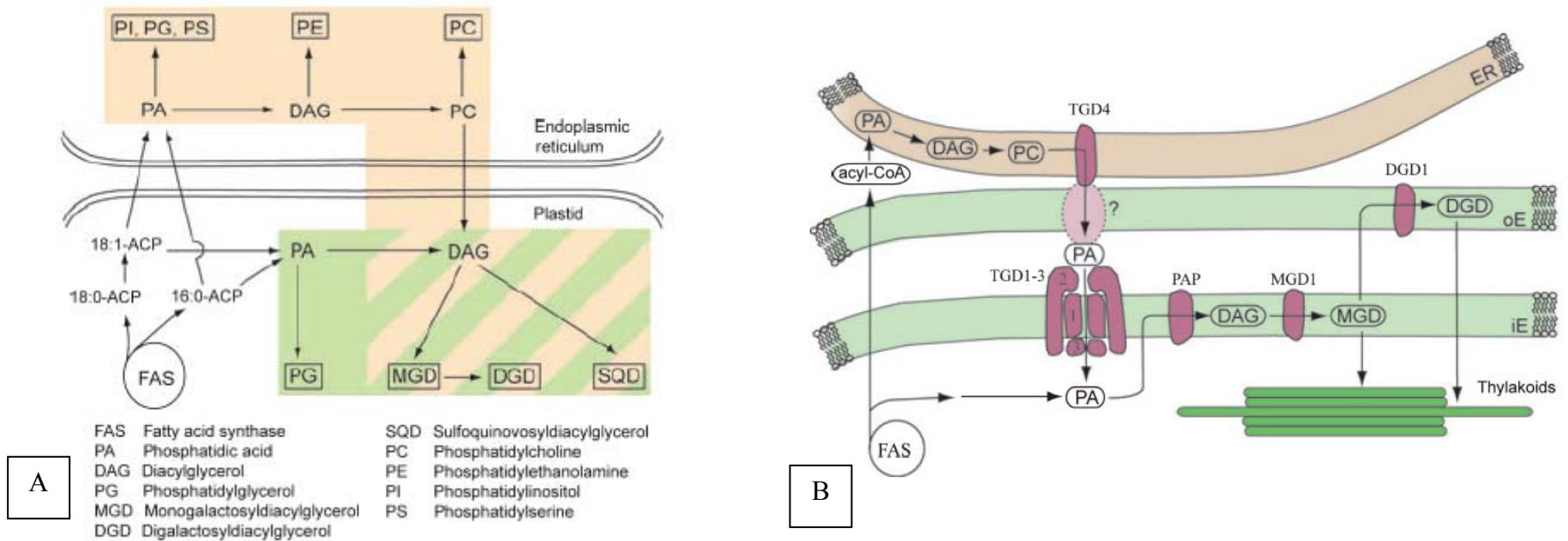
de température, on constate que peu de gènes sont communs aux trois espèces et entre les espèces parentales. Les cultivars de *C. arabica* possèdent une proportion plus importante de gènes en commun mais la tendance reste tout de même faible comparée au nombre important de gènes dépendants de l'espèce.

Au niveau des processus biologiques, les unigènes communs aux trois espèces de caféiers étudiés sous stress thermiques (30-26°C), sont tous liés à la régulation négative de la transcription par un promoteur de la polymérase III. Cette RNA Polymérase est connue pour réguler l'expression des gènes en fonction des stades de développement et des tissus de la plante mais aussi en fonction de l'environnement. Plusieurs facteurs de transcription liés au rythme circadien agissant sur le développement de la plante et aux réponses aux stress sont notablement sous-exprimés à la condition de température de stress de chaleur (30-26°C). On note cependant un comportement différent des espèces parentales face aux stress thermiques. Chez *C. canephora*, les gènes sur-exprimés couvrent un grand nombre de processus biologiques tels que la voie de biosynthèse de la chlorophylle, des lipides, des polysaccharides et des processus lié à la résistance. Un comportement inverse est observé chez *C. eugenioides* où les gènes sur-exprimés couvrent peu de processus biologiques et que ceux-ci sont essentiellement liés aux stress oxydatifs. L'espèce allopolyploïde, *C. arabica* s'apparente plus à *C. canephora* avec des gènes surexprimés qui couvrent un grand nombre de processus biologiques tels que la voie de biosynthèse de la chlorophylle et des triglycérides mais présente aussi de façon originale des processus liés aux hormones et à la régulation de la méthylation, des transporteurs ou encore des voies de signalisation. Parmi les gènes liés à ces processus biologiques, la majorité ont un rôle dans les stress biotiques et abiotiques. Ces gènes sont pour la plupart sur-exprimés en condition de stress de chaleur chez les trois espèces de caféiers. Par exemple les gènes codant pour les HSPs, MAPK et MYB sont des gènes qui ont un rôle majeur dans la régulation transcriptionnelle induite par des stress thermiques (Sung *et al.*, 2003).

# CHAPITRE III

## **PLASTICITE DES ALLOPOLYPLOÏDES FACE A DIFFERENTS REGIMES THERMIQUES: COMPARAISON ENTRE *COFFEA* *ARABICA* ET SES DEUX ESPECES PARENTALES (*C. CANEPHORA* ET *C.* *EUGENIOIDES*)**

Ce chapitre est rédigé en français et fera l'objet d'un article ultérieurement.



**Figure 1. Représentation des voies eucaryotique et procaryotique dans la synthèse de galactolipides et phospholipides dans une cellule de plante.**

La voie eucaryotique est représentée en orange et la voie procaryotique est représentée en vert. Ces représentations sont basées sur des études réalisées chez *Arabidopsis*. **A.** Représentation schématique des deux voies, procaryotique et eucaryotique, avant le séquençage du génome d'*Arabidopsis*. **B.** Représentation schématique des nouvelles découvertes basées sur le génome d'*Arabidopsis*. ER représente le réticulum endoplasmique, oE représente la membrane plastidiale externe et iE représente la membrane plastidiale interne.

Wallis et Browse, 2010



## INTRODUCTION

La polyploïdie joue un rôle majeur dans la diversité des espèces végétales, l'apparition de nouveaux phénotypes et dans la conquête de nouvelles niches écologiques (Otto et Whitton, 2000; Adams, 2007). Dans des environnements non adaptés aux parents diploïdes tels que les hautes altitudes et latitudes où les températures sont plus froides, les polyploïdes survivent mieux que leurs parents diploïdes (Chen, 2007). L'adaptation à de nouvelles niches écologiques peut être illustrée par le cas des plantes invasives (Prentis *et al.*, 2008) et des plantes arctiques (Brochmann *et al.*, 2003). Ces derniers soulignent que le succès de l'évolution des polyploïdes dans la zone arctique serait basé sur l'état d'hétérozygotie fixé de leur génome. Cet état serait à l'origine de la meilleure survie des espèces pendant des périodes climatiques instables. Ces études suggèrent ainsi un avantage évolutif des allopolyploïdes comparés à leurs espèces parentales diploïdes. La plasticité phénotypique se traduit par la capacité d'un génotype à produire de multiples phénotypes en réponse à des conditions environnementales variées (Pfenning, 2010). La plasticité des polyploïdes peut être corrélée avec des changements d'expression à des niveaux transcriptionnels et post-transcriptionnels (Jackson et Chen, 2009). Comme le souligne Jackson et Chen (2009), de tels changements dans l'expression génique sont contrôlés en grande partie par des mécanismes épigénétiques, lesquels jouent un rôle direct ou indirect, de régulateurs de variations physiologiques et développementales chez les allopolyploïdes (Wang *et al.*, 2006 ; Ni *et al.*, 2009).

Les stress thermiques ont des effets sur la croissance de la plante et sur la photosynthèse par des processus de photoinhibition (Na *et al.*, 2007) et également sur le fonctionnement des membranes cellulaires. Les phospholipides et galactolipides composent la bicouche lipidique des membranes cellulaires (Figure 1). Chez les plantes, les lipides les plus abondants dans les membranes des thylakoïdes sont les galactolipides MGDG (monogalactosyldiacylglycerol) et DGDG (digalactosyldiacylglycerol) et le phospholipide PG (phosphatidylglycerol) (Na *et al.*, 2007). Lors de stress thermiques, les changements de composition en acides gras des phospholipides et galactolipides membranaires jouent un rôle dans la fluidité de la membranaire (Robert et Upchurch, 2008). La modification de la fluidité des membranes est estimée par le taux de conversion de l'acide

oléique (C18 :1) en linoléique (C18 :2) puis en linoléique (C18 :3), (Iba 2002). Les changements observés dans le degré d'insaturation des acides gras sont considérés comme l'un des facteurs majeurs dans l'adaptation métabolique des plantes aux stress thermiques (Murakami *et al.*, 2000 ; Wakita *et al.*, 2001; Anai *et al.*, 2003; Orlova *et al.*, 2003; Sakurai *et al.*, 2003). Malgré les variations de température, les plantes maintiennent la fluidité de la membrane constante (Penfield, 2008). L'accumulation d'acide gras polyinsaturés dans les lipides contribue au maintien de la fluidité de la membrane (Murakami *et al.*, 2000 ; Liu *et al.*, 2006) et au fonctionnement régulier des chloroplastes (Routaboul *et al.* 2000; Iba, 2002) aux basses températures. Chez des espèces, telles que le coton et le soja, il a été montré que les niveaux en 18 :2 augmentent en fonction des baisses de température (Heppard *et al.*, 1996 ; Horiguchi *et al.*, 2000 ; Kargiotidou *et al.*, 2008). Pour les acides gras triénoïques, tels que l'acide hexadecatriénoïque (16:3) et l'acide linoléique (18:3), il a été montré que leur augmentation dans les membranes des chloroplastes induisaient une tolérance à des faibles températures au cours des premières étapes de leur développement (Iba, 2002).

Un groupe d'enzymes, les acides gras désaturases (FADs) sont à l'origine des conversions en acides gras insaturés. Elles permettent la conversion de l'acide oléique (C18 :1 ; monoinsaturés) en acide linoléique (C18 :2 ; polyinsaturés ou diénoïque) et la conversion de l'acide linoléique (18 :2) en acide linoléique (C18 :3 ; polyinsaturés ou triénoïque). Ces enzymes se situent dans différents compartiments cellulaires. On retrouve les désaturases FAD2 (ou  $\omega$ -6 désaturase) et FAD3 (ou  $\omega$ -3 désaturase) au niveau des membranes cytoplasmiques du réticulum endoplasmique et les désaturases FAD6 (ou  $\omega$ -6 désaturase) et FAD8 (ou  $\omega$ -3 désaturase) au niveau des membranes chloroplastiques des chloroplastes. Ces FADs, ainsi que d'autres enzymes intervenant dans la voie de biosynthèse des acides gras, sont responsables, pour partie, des modifications lipidiques et par conséquent à l'origine des réarrangements membranaires. La FAD2 et la FAD6 ( $\omega$ -6 désaturase) sont responsables de la conversion de l'acide oléique (C18 :1) en acide linoléique (C18 :2). Il a été observé, chez le mutant FAD2 d'*Arabidopsis*, une forte diminution des quantités en acides gras polyinsaturés dans les membranes cytoplasmiques lipidiques. Les cultures à long terme de ces mutants à basses températures conduisent à leur flétrissement et à leur mort (Miquel *et al.*, 1993).

Chez *Gossypium* et *Portulaca*, une expression génique différentielle a été observée. Chez *Gossypium* (Kargiotidou *et al.*, 2008), il a été observé une augmentation du niveau de C18:2 associée à des hauts niveaux de transcription du gène codant pour la FAD2 à basses températures alors que chez *Portulaca* (Teixeira *et al.*, 2009), il a été observé de faibles niveaux de transcription du gène codant pour la FAD2 à basses températures. Chez le mutant FAD6 d'*Arabidopsis*, il a été observé une forte activité d'enzyme désaturase et de hauts niveaux d'accumulation d'acides palmitoléique (16:1) et oléique (18:1), associée à une diminution des niveaux d'acides gras polyinsaturés dans les galactolipides des chloroplastes (Iba, 2002). Les plantules de ces mutants, soumis à des stress thermiques, deviennent chlorotiques et possèdent une teneur en chlorophylle moitié moins importante que chez les types sauvages. De plus, le nombre de thylakoïdes dans leurs membranes diminuent et leurs chloroplastes sont plus petits. La FAD8 et la FAD3 ( $\omega$ -3 désaturase) sont responsables de la conversion de l'acide linoléique (C18 :2) en acide linoléique (C18 :3). L'expression chez les gènes codant pour la FAD8 a été étudiée chez de nombreuses espèces dont *Arabidopsis*, *Gossypium* (Kargiotidou *et al.*, 2008) et *Descurainia* (Tang *et al.*, 2007). Chez *Arabidopsis*, le gène codant pour la désaturase FAD8 peut être réprimé ou non en fonction d'une différence de quelques degrés (Iba, 2002). Ce gène est identifié comme un gène « température-dépendant » (McConn *et al.*, 1994). Chez *Gossypium* et *Descurainia*, il a été observé des hauts niveaux de transcrits du gène codant pour la FAD8 et donc une forte expression à basses températures (Tang *et al.*, 2007 ; Kargiotidou *et al.*, 2008 ). Les gènes codant pour l'enzyme FAD3 ont été isolés chez plusieurs espèces de plantes incluant la *Glycine max* et *Brassica napus* (Yadav *et al.*, 1993), *Nicotiana tabacum* (Hamada *et al.*, 1996), *Arabidopsis thaliana* (Arondel *et al.*, 1992; Nishiuchi *et al.*, 1994), et *Oryza sativa* L. (Kodama *et al.*, 1997). Chez *B. napus*, les niveaux d'acides linoléique et linoléique augmentent aux basses températures mais les auteurs précisent que les variations de l'activité enzymatique ne seraient pas dues aux niveaux d'activité des transcrits de FAD3 (Dyer *et al.*, 2001). La production d'acide linoléique (C18 :3) est catalysée par la FAD3 (Yadav *et al.*, 1993) et la FAD8 (Lemieux *et al.*, 1990; Iba *et al.*, 1993; Somerville et Browse, 1996). Les transferts lipidiques entre compartiments sont possibles (Ohlrogge et Browse, 1995; Horiguchi *et al.*, 1996; McConn et Browse, 1996; Kodama *et al.*, 1997) et seraient dus à la contribution totale de C18 :3 provenant des deux formes de l' $\omega$ -3 désaturase (la FAD3 et la FAD8).

*Coffea arabica* est un allopolyploïde relativement récent (10-50000 ans) constitué de sous-génomes (E<sup>a</sup> et C<sup>a</sup>) provenant d'espèces parentales africaines diploïdes proches génétiquement, *C. eugenioides* et *C. canephora*. *Coffea arabica* est issu d'une région limitée

du Sud-ouest de l’Ethiopie où aucune autre espèce de *Coffea* n’est présente. L’espèce polyploïde et ses deux espèces parentales se développent également à des altitudes et à des températures différentes. On retrouve *C. arabica* dans des forêts tropicales de hautes altitudes, entre 1200-1950m (Davis *et al.*, 2006), avec un optimum de températures moyennes annuelles pour l’espèce de 18-21°C (Damatta et Ramalho, 2008). *Coffea canephora* est issu de forêts équatoriales humides de basses à moyennes altitudes, entre 250-1500m (Davis *et al.*, 2006), avec un optimum de températures moyennes annuelles pour l’espèce de 22-26°C (Damatta et Ramalho, 2008). Enfin, *C. eugenioides* se retrouve dans des forêts de hautes altitudes, entre 1000 et 2000m, avec un optimum de températures moyennes annuelles pour l’espèce de 18-23°C (Davis *et al.*, 2006). Comme la plupart des espèces tropicales, les caféiers sont particulièrement sensibles au froid (Bauer *et al.*, 1985; DaMatta *et al.*, 1997; Ramalho *et al.*, 2003, Fortunato *et al.*, 2009) limitant leurs zones géographiques de distribution (Fortunato *et al.*, 2009). Les deux espèces de caféiers cultivées (*C. arabica* et *C. canephora*) ont montré qu’elles pouvaient s’adapter à des intervalles de températures plus importants dans leur zone de culture comparés aux températures des zones de diversification primaire. Malgré ces intervalles plus importants pour les espèces cultivées, la température reste tout de même une limite à l’exploitation des variétés de *C. arabica* et *C. canephora* (Barros *et al.*, 1997; Silva *et al.*, 2004; Damatta *et al.*, 2006). Les variations de la composition lipidique en fonction des stress thermiques constituent a priori un bon modèle pour étudier la plasticité des plantes.

L’objectif de ce chapitre sera de confirmer la plasticité de l’allopolyploïde *C. arabica* sous stress thermiques et d’estimer si cette plasticité est supérieure ou égale à celle de ses espèces parentales diploïdes, *C. eugenioides* et *C. canephora*. Pour répondre à cette question, nous avons donc mesuré la croissance, la teneur en chlorophylle et la composition en acides gras ainsi que l’expression de gènes majeurs de la voie de biosynthèse des acides gras pour chacune des trois espèces cultivées dans des conditions de températures contrastées.

## MATERIELS ET METHODES

### Matériels biologiques

Les semences de *Coffea arabica* et *Coffea canephora* ont été produites à La station expérimentale ECOM à la Cumplida, Matagalpa, Nicaragua. Les semences de *Coffea eugenioides* ont été produites par la « Coffee Research Foundation » du Kenya. La variété de *Coffea arabica* utilisée est la variété 'Caturra' issue, par mutation, de la variété *C. arabica* cv bourbon. *Coffea canephora* est représenté par la variété 'Nemaya' issue d'un croisement entre deux génotypes congolais sauvages. Et enfin, les graines de *Coffea eugenioides* ont été collectées à partir d'arbres originaires de la forêt du Mont Elgon au Kenya.

### Croissance en phytotrons

Après cinq semaines de croissance dans des boîtes magenta à température constante (27°C), les poids secs (103°C pendant 24h) ont été mesurés pour trente plantules de chaque espèce. Les cent vingt plantules ont été réparties dans douze mini serres. Dans chacune des mini serres, ont été plantées trois lignes de dix plantules de chacune des trois espèces étudiées. Chaque mini serre est considérée comme une unité expérimentale. Les mini serres sont transférées dans des phytotrons (CRYONEXT, France, model RTH 1200L), où les conditions environnementales sont contrôlées. La photopériode, l'humidité et la luminosité sont respectivement fixées à 12:12 LD (Long Day), 80-90% et 600  $\mu$  mole photons  $m^{-2} s^{-1}$ . Pour cette étude, quatre conditions de température ont été utilisées et pour lesquelles des températures diurnes et nocturnes ont été définies : 18-15°C, 23-20°C, 28-25°C, 33-30°C. Après 11 semaines de culture dans les phytotrons, deux jeunes feuilles ont été échantillonnées à la mi-journée (6-8 h après la nuit) pour chacune des plantes. Ces feuilles sont directement congelées dans l'azote liquide puis stockées à -80°C jusqu'à l'extraction de l'ARN en vue des analyses d'expression génique par la PCR quantitative en temps réel (qRT-PCR). Les jeunes plantules sont ensuite prélevées et les racines lavées. Les poids secs (103°C pendant 24h) ont été de nouveau mesurés pour cinq plantules par espèce et par unité expérimentale. Pour chaque unité expérimentale, les plantules restantes sont lyophilisées et stockées.

## **Analyses chimiques**

Pour chacune des trois espèces et chacune des quatre conditions de température, des dosages de la chlorophylle et la composition en acides gras ont été effectués. Les analyses ont été faites sur trois répétitions correspondant aux plantes d'une même espèce dans les trois mini serres. Chacune des répétitions est constituée d'un pool de cinq plantes.

### ***Dosage de la chlorophylle***

Un dosage de la teneur en chlorophylle est effectué suivant la méthode de Mc Kiney (1941) et Holden (1975) par l'utilisation d'acétone à 80% pour l'extraction et d'un spectrophotomètre à double cuve pour les mesures du spectre d'absorption. L'extraction de la chlorophylle est réalisée à partir d'un broyage manuel de 20mg d'échantillon lyophilisé dans 2ml d'acétone à 80%. La solution est complétée pour un volume final de 10 ml. Les tubes sont centrifugés (30240g, 10min, 4°C). Pour obtenir une DO < 1, une dilution de ½ est effectuée pour 1ml de chaque échantillon. Préalablement aux mesures effectués, une gamme étalon est réalisée à partir d'extrait pur de chlorophylle a et b d'épinard (SIGMA ALDRICH C5753 et C5878 1mG). Le spectre d'absorption pour chacun des échantillons est mesuré pour des longueurs d'ondes comprises entre 500 et 800nm. Les concentrations en chlorophylle sont déterminées par les formules suivantes :

$$\text{Chl totale} = F(0,0202 * \text{DO}_{645} + 0,008 * \text{DO}_{663}) * 1000 \text{ en mg.L}^{-1}$$

Avec F le Facteur de dilution

### ***Extraction des lipides***

Les lipides totaux sont extraits à partir de 100 mg d'échantillons lyophilisés et réduits en poudre en utilisant la méthode de Folch *et al.* (1957) modifiée par Laffargue *et al.* (2007). Chaque échantillon est placé dans un pilulier en verre de 15 ml à capuchon vissant (bakélite et joint téflon). Les poudres de feuille sont homogénéisées avec un IKA T25 Ultra-turrax (Werke, Allemagne) pendant 30s dans 10ml de solvant (dichlorométhane/méthanol (CH<sub>2</sub>CL<sub>2</sub>/MeOH, 2/1)). Les échantillons sont filtrés sous vide (porosité 4µ). Les poudres d'extraction sont collectées et l'étape est renouvelée une deuxième fois. Le filtrat est transféré dans une ampoule à décanter où est ajoutée 4ml d'une solution de 0.73% de chlorure de sodium. La solution est mélangée manuellement pendant 1min et laissée à décanter toute la nuit. Après décantation (présence de deux phases), la phase la plus basse contenant les lipides totaux est récupérée. Après évaporation sous flux d'azote (37°C) de l'ensemble du solvant à

l'aide d'un évaporateur N-EVAP 11155-RT (Organomation, USA), les lipides totaux extraits sont pesés puis conservés dans 2 ml de solvant à -20°C.

#### ***Procédure de saponification, méthylation et chromatographie en phase gazeuse***

La composition des acides gras des lipides totaux a été déterminée à partir du protocole de Laffargue *et al.* (2007). Pour la méthylation, les acides gras estérifiés sont saponifiés avec de l'hydroxyde de sodium méthanolique (MeOH). Les savons ainsi formés et les acides gras libres sont ensuite convertis en EMAG par réaction avec un complexe de trifluorure de bore (BF<sub>3</sub>)/méthanol. Pour la saponification, chaque extrait lipidique est transféré dans un tube de 50ml à capuchon vissant (bakélite et joint téflon). Le solvant est ensuite évaporé sous flux d'azote dans un bain marie à 37°C. Les lipides extraits sont saponifiés avec 4 ml de solution de NaOH à 0,5M. Les échantillons sont vortexés et placés au bain marie, sous agitation (90°C, 5 min). Les lipides sont ensuite méthylés avec 5 mL d'une solution BF<sub>3</sub> à 14% à chaud. Les échantillons sont de nouveau vortexés et placés au bain marie, sous agitation (90°C, 3 min). Les échantillons sont vortexés et placés au bain marie (90°C, 5min). trois ml d'une solution d'hexane et 20ml d'une solution de chlorure de sodium saturée (NaCl) sont ajoutés. Après agitation manuelle, on rajoute de la solution de NaCl saturée jusqu'au col du tube. Après 2h de décantation, la phase supérieure d'hexane est récupérée et mise dans un pilulier en verre avec capuchon en bakélite et joint téflon de 10ml. Les traces d'eau résiduelles sont piégées en ajoutant quelques mg de sulfate de sodium (Na<sub>2</sub>SO<sub>4</sub>) anhydre. L'échantillon est prêt pour être injecté en CPG (chromatographie en phase gazeuse).

La phase hexane est récoltée et les FAMES sont directement analysés par chromatographie en phase gazeuse (GC) avec un chromatographe HP 5890 avec détecteur à ionisation de flamme (FID) et des colonnes capillaires Wax pour FAME (RESTEK, France) de 30 m \* 0.25 mm \* 0.25 m. Les analyses ont été réalisées à partir d'un mode augmentant la température de 185°C à 225°C à 4°C/ min et d'un mode isotherme de 225°C pendant 10min. Le gaz transporteur est l'hélium à 40 cm s<sup>-1</sup>. L'injection et la détection se font à 230°C. Les FAMES ont été identifiés en comparant les temps de rétentions avec ceux des FAMES standard (Supelco) et sont quantifiés en pourcentage d'acides gras total (w/w).

La chromatographie en phase gazeuse (CPG) est une méthode d'analyse physico-chimique qui sépare les constituants d'un mélange afin de les quantifier individuellement. Une fois vaporisés dans l'injecteur, les constituants sont entraînés par un gaz vecteur le long d'une

phase stationnaire emprisonnée dans une colonne. Dans le cas des esters méthyliques d'acides gras (EMAG), ils parcourent la colonne avec des temps proportionnels à leurs propriétés intrinsèques (nombre d'atomes de carbone, nombre de doubles liaisons) et à leur affinité avec la phase stationnaire (polarité). A leur arrivée en bout de colonne, le détecteur mesure en continu la quantité de chacun des EMAG du mélange. Le tracé d'un chromatogramme correspond à l'enregistrement de l'intensité du signal généré par le détecteur en fonction du temps.

L'identification des EMAG a été réalisée avec des standards commerciaux. Chaque standard est d'abord caractérisé par son temps de rétention (TR). Tous les temps de rétention obtenus ont été convertis en temps de rétention relatif (TRR) par division par le TR de l'acide oléique (18:1). Si les TR peuvent évoluer légèrement au cours des analyses, principalement à cause de la dégradation de la phase stationnaire de la colonne, les TRR restent constants. L'identification d'un composé dans un chromatogramme est ainsi réalisée par son TRR. La quantification de chaque composé est ensuite réalisée à l'aide de la surface de son pic. Le pourcentage d'un acide gras dans le mélange est calculé par le rapport entre la surface de son pic et la somme des surfaces de tous les pics identifiés.

## **Analyses de qRT-PCR (Quantitative Reverse Transcriptase Polymerase Chain Reaction)**

### ***Extraction des ARN totaux***

Pour chacun des échantillons, environ 1g de feuilles a été broyé dans de l'azote liquide et homogénéisé dans 20ml d'une solution tampon contenant 5 M d'isothiocyanate de guanidium, 31 mM d'acetate de sodium (pH 8), 1% de  $\beta$ -mercaptoéthanol, 0.88% (w/v) de *N*-lauroyl sarcosine and 1% (w/v) de polyvinylpyrrolidone 40. Après centrifugation (15000g, 4°C, 20min), le surnageant est déposé sur un coussin de chlorure de césium (5.7 M) contenant 25mM d'acétate de sodium (pH 6) et centrifugé de nouveau (180000g, 20°C, 20h). Après élimination du surnageant, les culots d'ARN totaux sont lavés deux fois avec 70% (v/v) d'éthanol puis séchés pendant 1h. Les échantillons sont ensuite suspendus dans 100 $\mu$ l d'eau RNase free. Le dosage des ARN totaux a été effectué en mesurant la densité optique d'1 $\mu$ l de solution à 260nm à l'aide d'un spectrophotomètre adapté aux petits volumes (Nanodrop, ND, Labtech, France).



**La PCR quantitative en temps réel (qRT-PCR)**

La synthèse du premier brin d'ADNc, a été effectuée à l'aide du kit « the ImProm-II™ Reverse Transcription System » (Promega, Madison, WI, USA). Pour chacun des échantillons, 1 µg d'ARN total et 1 µl d'oligonucléotides (Oligo(dT)) à 0.5 µg/µl sont utilisés pour un volume réactionnel de 10 µl. Un contrôle négatif sans matrice d'ADNc est effectué. Les solutions d'ARN totaux sont dénaturées (70°C, 5min) et stockés dans de la glace (5 min). La deuxième étape consiste en la synthèse d'ADNc par la reverse transcriptase à partir des oligonucléotides fixés sur les ARN totaux. Au mélange de 10 µl est ajouté 4 µl de tampon de la reverse transcriptase, 4 µl de solution de MgCl<sub>2</sub> à 25mM, 1 µl de solution de dNTP à 500 µM, 1 µl de RNase inhibitor à 20U/µl et enfin 1 µl de reverse transcriptase. La réaction est incubée (25°C pendant 5min, 42°C pendant 1 h et 70°C pendant 15min). Le premier palier de température correspond à l'hybridation des oligonucléotides dT sur les queues poly-A des ARNm, le second palier correspond à la synthèse de l'ADNc par l'enzyme à partir des oligonucléotides et le dernier palier correspond à la phase de désactivation de l'enzyme.

L'analyse Q-PCR est effectuée à partir du kit « LightCycler® 480 SYBR Green I Master » (Version February 2008, Roche, Mannheim, Germany). La préparation du mélange PCR se fait sur la glace. Pour une réaction, le mélange PCR contient 1 µl d'eau (Water PCR Grade), 4 µl d'amorces (10X conc., 2 µl de 'reverse' et 2 µl de 'forward') et 10 µl d'un tampon « Master Mix » (2X conc.). 5 µl d'échantillon d'ADN sont ajoutés au 15 µl du mélange PCR dans chaque puits de la plaque 384 (LightCycler® 480 Multiwell Plate 384). La plaque est centrifugée (1500g, 2 min) et analysée à l'aide du protocole « LightCycler® 480 SYBR Instrument Protocol ». Quatre étapes sont effectuées : la *pré-incubation* afin d'activer la polymérase (FastStart Taq DNA polymérase) et dénaturer l'ADN (Paramètres: 1 cycle; 95°C, 10 min), l'*amplification* de l'ADN (Paramètres: 45 cycles; 95°C pendant 15 s, 60°C pendant 1 min et 72°C pendant 30 s), la génération des *courbes de fusion* (Paramètres: 1 cycle; 95°C pendant 5 s, 65°C pendant 1 min et 97°C pour 5 acquisitions par °C) permettant d'identifier les produits PCR et le *refroidissement* de la plaque (Paramètres: 40°C, 10 s).

Une gamme de dilution de la RT (pour chaque échantillon) a été effectuée afin de déterminer la meilleure dilution de la RT. Cette gamme permettra également de déterminer l'efficacité des amorces pour chaque échantillon. Avec cette méthode, l'efficacité du couple amorces/échantillons est comprise entre 1.60 et 2. Les valeurs de l'efficacité seront prises en compte au cours des analyses statistiques.

Pour éviter des effets liés au run de la PCR, les quatre gènes de référence sont utilisés comme une référence interne dans chacune des plaques.

	Gènes	Noms	N° SGN	Amorce Reverse	Amorce Foward
<b>Gènes cibles</b>	Isoforme 1 oméga-6 désaturase	<b>ω-3.1</b>	SGN-U348085	TCGTAATGAGGCAACGAAGG	CTTGCTGGGTTGTCTGTG
	Isoforme 2 oméga-6 désaturase	<b>ω-3.2</b>	SGN-U342489	CAGGAGAACTGGGGCAAGTA	GACTCTTCAGGGGCGATTTC
	Isoforme 1 oméga-3 désaturase	<b>ω-6.1</b>	Pas de SGN Amorces dessinées	ACCTGGTCGGAAGCATCAGT	TCTGGCCTGCAACTCGTAGA
	Isoforme 2 oméga-3 désaturase	<b>ω-6.1</b>	SGN-U349395	CAATGCCCAATTCCTGTCCT	CAAAACCAAAAAGTGGCCTGA
<b>Gènes de référence</b>	GADPH	<b>GADPH</b>	SGN-U347734	AACATGGGTGCATCCTTGCT	TTGAAGGGCGGTGCAAA
	AP47	<b>AP47</b>	SNG-U359260	AAATGCACAAGCAACATTGG	CCAGATTGGAGGATGCTCTT
	RP17	<b>RP17</b>	SGN-U351477	TGTCTCAGGCGCAGAAGCT	CATTCGAGGTATCAATGCTATGCA
	14-3-3	<b>14-3-3</b>	SGN-U356404	CTTCACGAGACATATTGTCTTACTCAA	TGTGCTCTTTAGCTTCAAACG

**Tableau. 1 : Liste des amorces utilisées lors de la qRT-PCR.**

Sont représenté les quatre gènes cibles et les quatre gènes de référence.

L'expression relative des transcrits des gènes ciblés est calculée de la manière suivante : (Joët *et al.*, 2009)

$$E^{(-\Delta Ct)}$$

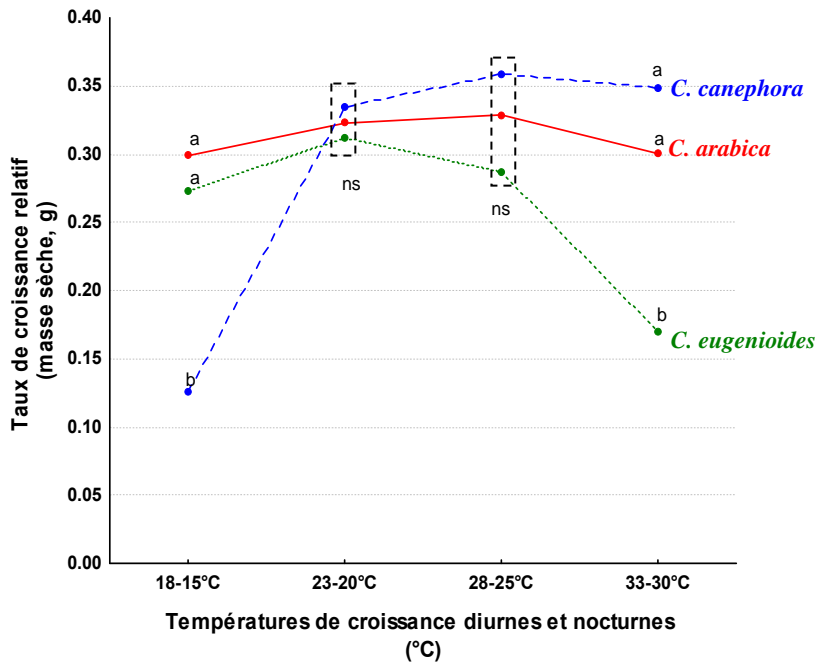
Avec E correspondant à l'efficacité

$\Delta Ct = (Ct \text{ gène cible} - Ct \text{ référence})$  où le Ct référence correspond à la moyenne des quatre gènes de référence.

Le pool de quatre gènes de référence a été établi à partir de la bibliographie (Barsalobres *et al.*, 2009 ; Cruz *et al.*, 2009 ; Tableau. 1). Ces quatre gènes de référence: GAPDH (glyceraldehyde-3-phosphate dehydrogenase), RP17 (60S ribosomal protein L7), 14-3-3 (14-3-3 protein), AP47 (clathrin-associated protein) semblent être adaptés aux espèces de caféiers cultivés. D'après ces auteurs, ces quatre gènes sont des gènes relativement stables chez les caféiers. L'expression de la stabilité de ces quatre gènes de référence a été déterminée en utilisant le logiciel geNorm software (Vandesompele *et al.*, 2002 ; <http://medgen.ugent.be/~jvdesomp/genorm/>). Les quatre gènes codant pour les enzymes Omega-3 ( $\omega$ -3.1 et  $\omega$ -3.2) et omega-6 ( $\omega$ -6.1 et  $\omega$ -6.2) ont été déterminés à partir du logiciel Primer3 software (<http://frodo.wi.mit.edu/primer3/>) avec les critères suivants : 80bp à 150bp, Tm de 62°C, GC% entre 40 et 60. La spécificité de chaque couple d'amorces a été testée par l'analyse des courbes de dissociation (qui indique la séparation des deux brins d'ADN de l'espèce) et l'intégrité des échantillons d'ARN a été vérifiée par une électrophorèse sur gel d'agarose.

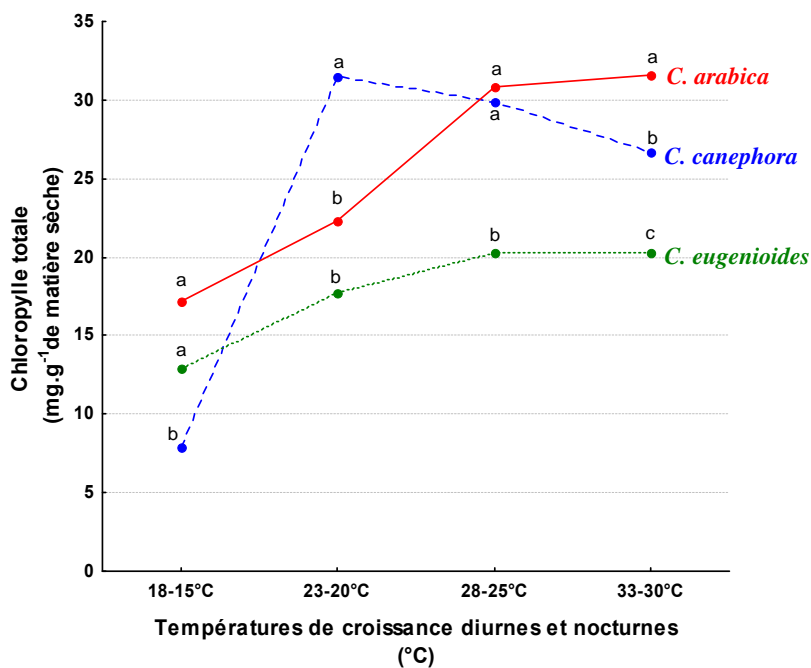
### ***Analyses statistiques***

Un taux de croissance relative a été calculé par la formule :  $\log(D60-D0)$ . Le taux de croissance relative, la teneur en chlorophylle, la composition en acides gras et le niveau d'expression des gènes sont analysés par des ANOVA suivies de tests HSD (Honest Significant Difference) de Tukey.



**Figure. 2 : Influence de la température sur la croissance des trois espèces de caféiers.**

*C. arabica*, *C. canephora* et *C. eugenioides* sont représentés, respectivement, en rouge, bleu et vert. Pour chacune des conditions de température, les trois espèces sont comparées par une ANOVA suivie d'un test HSD de Tukey. Deux lettres différentes pour une même température de croissance signifient qu'il y a une différence significative au seuil de 0.05 entre deux espèces et 'ns' signifie qu'il n'y a pas de différences entre espèces pour la température considérée. *C. arabica* apparaît stable quelque soit la température.



**Figure. 3 : Influence de la température sur la teneur en chlorophylle totale des jeunes feuilles des trois espèces de caféiers.**

*C. arabica*, *C. canephora* et *C. eugenioides* sont représentés, respectivement, en rouge, bleu et vert. Pour chacune des conditions de température, les trois espèces sont comparées par une ANOVA suivie d'un test HSD de Tukey. Deux lettres différentes pour une même température de croissance signifient qu'il y a une différence significative au seuil de 0.05 entre deux espèces et 'ns' signifie qu'il n'y a pas de différences entre espèces pour la température considérée. La teneur en chlorophylle totale de *C. arabica* est proche de celle de *C. eugenioides* aux températures froides et proche de celle de *C. canephora* aux températures les plus chaudes.

## RESULTATS

### **Influence des contraintes thermiques sur la croissance et la teneur en chlorophylle des trois espèces de caféiers étudiés**

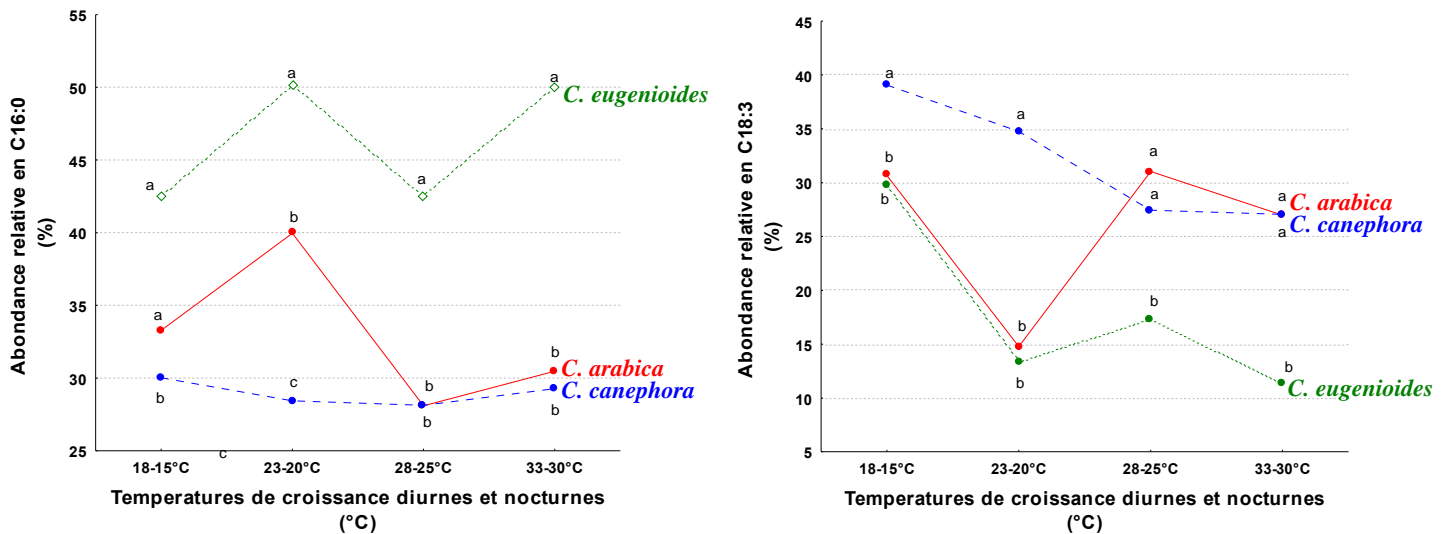
Pour étudier la croissance des trois espèces de caféiers en fonction des conditions de températures, un taux de croissance relatif a été calculé à partir de la différence entre les poids secs estimés avant la mise en phytotrons et à leur sortie (à T0 et à T11 semaines). Les trois espèces ne se développent pas de la même façon en fonction des conditions de températures. Pour l'espèce *C. canephora*, la condition la plus froide (18-15°C) apparaît très défavorable à la croissance des plantules (Figure. 2). A contrario, la croissance des plantules de *C. eugenoides* est affectée à la température la plus chaude (30-33°C ; Figure. 2). A cette condition de température, on a constaté un taux de mortalité de 52% des plantules de *C. eugenoides*. Le taux de croissance de l'espèce allopolyploïde apparaît stable quelle que soit la température. Ce taux de croissance est du même ordre de grandeur que celui du parent *C. canephora* aux températures les plus chaudes (28-25°C et 33-30°C) et est similaire à celui de *C. eugenoides* aux températures les plus froides (18-15°C et 23-20°C ; Figure. 2).

Quelle que soit l'espèce considérée, on observe que plus les conditions de température sont chaudes, plus la teneur en chlorophylle augmente (Figure. 3). A la condition de température de 18-15°C, on observe une très forte diminution (environ 25mg.g<sup>-1</sup> de matière sèche) de la teneur en chlorophylle chez *C. canephora*. La teneur en chlorophylle de l'espèce allopolyploïde est du même ordre de grandeur que celle de *C. eugenoides* aux températures les plus froides (18-15°C ; 23-20°C) alors qu'aux températures les plus chaudes (28-25°C ; 33-30°C) sa teneur est du même ordre de grandeur que celle de *C. canephora* (Figure. 3).

Espèces	T°C	Acide Palmitique (C16 :0)	Acide Palmitoléique (C16 :1)	Acide Stéarique (C18 :0)	Acide Oléique (C18 :1)	Acide Linoléique (C18 :2)	Acide Linoléique (C18 :3)	Acide Arachidique (C20 :0)
<i>C. canephora</i>	18-15	30	3	6.5	4.3	12.7	39.1	1.9
	23-20	28.5	5.8	8.1	3.5	13.2	34.8	1.9
	28-25	28.1	9.4	10.5	5.2	10.8	27.5	2.6
	33-30	29.3	7.5	8.7	5.5	14.6	27	2.8
<i>C. eugenioides</i>	18-15	42.5	3.1	6.7	3.5	10.8	29.8	1.3
	23-20	50.2	5	8.3	5.9	12.3	13.4	1.7
	28-25	42.5	6.2	10.4	7.3	10.2	17.3	2.3
	33-30	50	4.6	10.7	5.7	12.1	11.4	3.2
<i>C. arabica</i>	18-15	33.3	5.2	8.2	4	13.1	30.7	1.3
	23-20	40	7	11.1	4.2	15.9	14.8	2.2
	28-25	28.1	7.5	9.1	7.2	11	31	1.5
	33-30	30.5	5.3	9.7	6.8	14.7	27	2.6

**Tableau. 2 : Variations de l'abondance relative des principaux acides gras en fonction des conditions de température chez les trois espèces étudiées.**

Sept acides gras sont présentés : l'acide palmitique (C16 :0), l'acide palmitoléique (C16 :1), l'acide stéarique (C18 :0), l'acide oléique (C18 :1), l'acide linoléique (C18 :2), l'acide linoléique (C18 :3) et l'acide arachidique (C20 :0). Les variations de l'abondance relative s'observent tant au niveau des espèces qu'au niveau des conditions de températures. Les colonnes rouges représentent les deux acides gras les plus abondants, l'acide palmitique (C16 :0) et l'acide linoléique (C18 :3). Les résultats sont exprimés en pourcentage.



**Figure 4 : Variation des proportions en 16 :0 et 18 :3 chez les trois espèces de caféiers en fonction de la température.**

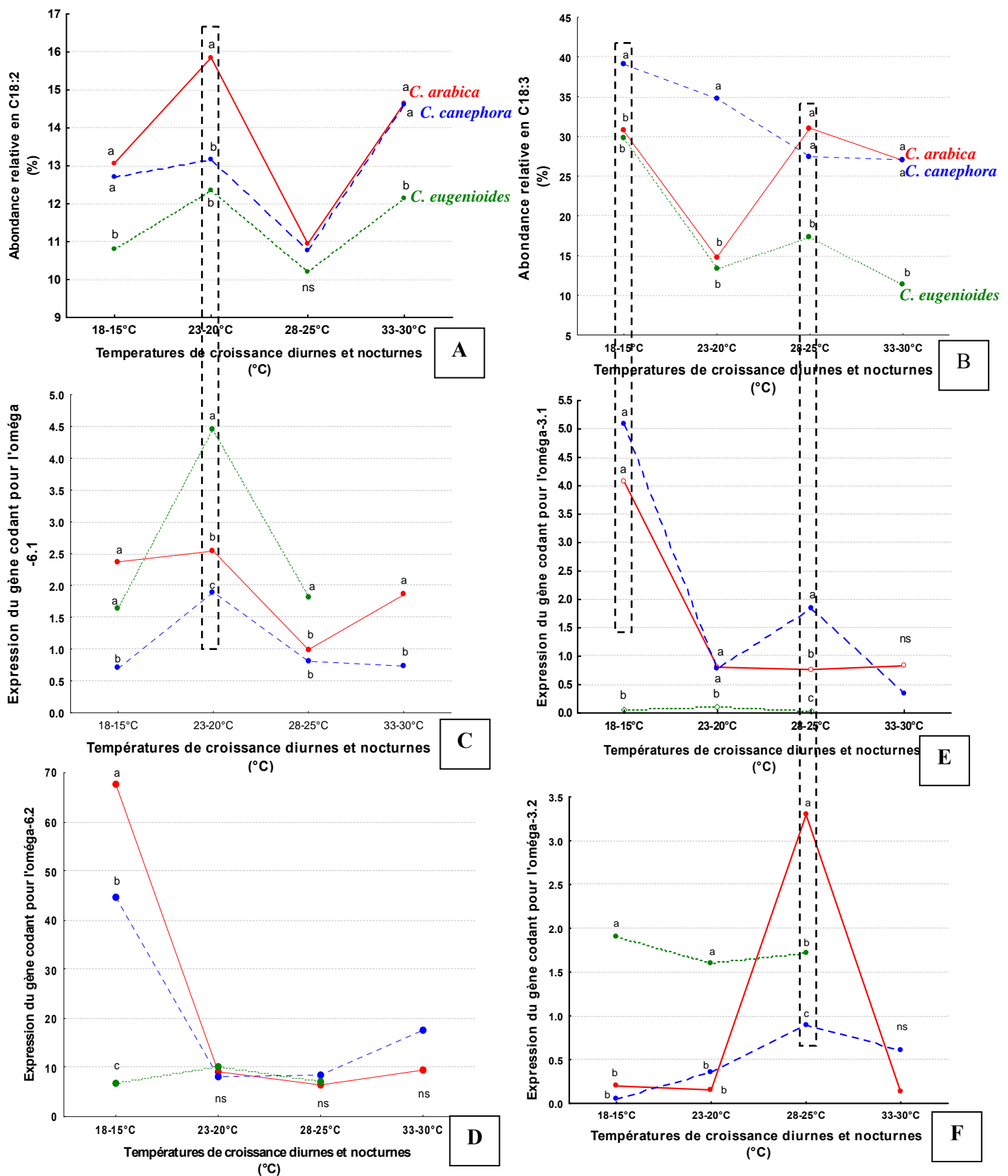
Pour chacune des conditions de température, les trois espèces sont comparées par une ANOVA suivie d'un test HSD de Tukey. Deux lettres différentes pour une même température de croissance signifient qu'il y a une différence significative au seuil de 0.05 entre deux espèces et 'ns' signifie qu'il n'y a pas de différences entre espèces pour la température considérée. **4A.** Graphique représentant les variations en palmitique (C 16:0). **4B.** Graphique représentant

## Abondance relative des principaux acides gras chez les trois espèces de caféiers étudiées en fonction des régimes thermiques

Les acides gras ont été convertis en esters méthyliques d'acides gras et analysés par chromatographie en phase gazeuse. Sept principaux acides gras ont été quantifiés. Chez les trois espèces de caféiers étudiées, l'acide palmitique (C16 :0) et l'acide linoléique (C18 :3) sont les deux acides gras les plus abondants. Les variations de l'abondance relative de ces acides gras s'observent tant au niveau des espèces qu'au niveau des conditions de températures. Suivant les conditions de température, on observe des proportions allant de 28.1 à 50% pour C16 :0 et de 11.4 à 39.1% pour C18 :3 (Tableau. 2). Les cinq autres acides gras ne représentent que 25 à 39% de l'abondance totale (Tableau. 2). Pour ceux-ci, on remarque peu de variations que ce soit entre les trois espèces ou entre les conditions de température.

En comparant les espèces parentales, on constate une différence d'abondance relative chez les deux acides gras les plus abondants en fonction des régimes thermiques. Pour les quatre régimes thermiques, on observe que les proportions en C16:0 sont plus élevées chez *C. eugenioïdes* (42.5-50.2%) que chez *C. canephora* (>30%) (Tableau. 2). A contrario, la proportion en C18:3 est plus importante chez *C. canephora* (27-39.1%) que chez *C. eugenioïdes* (11.4-29.8%) sur l'ensemble des régimes thermiques étudiés (Tableau. 2). On observe que, pour les deux espèces parentales, les proportions en acide linoléique augmentent lorsque les températures baissent (Figure. 4). En ce qui concerne l'acide palmitique, on observe une stabilité de l'abondance relative, donc peu de variabilité en fonction des conditions de température pour l'espèce *C. canephora* et une variabilité de celle-ci en fonction des conditions de température pour l'espèce *C. eugenioïdes* (Figure. 4).

Concernant *C. arabica*, on observe une forte variabilité des proportions en 16 :0 et 18 :3 en fonction des conditions de température. Comparé aux trois autres conditions de température pour lesquelles les proportions en 16 :0 et 18 :3 sont du même ordre de grandeur, on observe que la proportion en 16 :0 (40%) est plus importante que la proportion en 18 :3 (14.8%) à la condition de température 23-20°C (Tableau. 2). Aux températures les plus chaudes (28-25°C et 33-30°C), on constate que ces proportions en 16 :0 et 18 :3 sont du même ordre de grandeur que celles de *C. canephora* alors qu'aux températures les plus froides (18-15°C et 23-20°C), ces proportions relatives sont intermédiaires aux deux espèces parentales pour le C16 :0 et du même ordre de grandeur que celles de *C. eugenioïdes* pour le C18 :3 (Figure. 4).



**Figure 5 : Corrélation entre les variations observées de profils chez les acides gras et celles observées chez 4 gènes impliqués dans la désaturation des lipides de feuilles chez le caféier.**

En rouge est représenté l'espèce allopolyploïde *C. arabica*, en bleu l'espèce parentale *C. canephora* et en vert l'espèce parentale *C. eugenioides*. **5A. B** Graphiques représentant, respectivement, les pourcentages relatifs en 18:2 et en 18:3. **5C. D** Graphiques représentant l'expression génique des gènes codant pour l'enzyme oméga-6 désaturase. La figure 5C représente l'isoforme 1 et la figure 5D l'isoforme 2. **5E. F** Graphiques représentant l'expression génique des gènes codant pour l'enzyme oméga-3 désaturase. La figure 5E représente l'isoforme 1 et la figure 5F l'isoforme 2. Les encadrés noirs en pointillés représentent les correspondances entre les pics observés en pourcentages relatifs d'acides gras et les niveaux d'expression génique des quatre gènes codant pour les enzymes de désaturation.



## Variation de l'expression de gènes codant pour les oméga-3 et oméga-6 désaturases en fonction des différents régimes thermiques.

Nous avons choisi d'étudier quatre gènes codant pour des enzymes appartenant au groupe des désaturases. En ce qui concerne ces gènes, on n'est pas capable de déterminer avec exactitude leur correspondance avec les gènes de FADs décrits dans la littérature. Les séquences des unigènes de caféiers ne permettent pas de les attribuer aux gènes décrits chez *Arabidopsis*. Ne pouvant pas déterminer si ces gènes codant pour des enzymes de désaturase sont spécifiques d'un compartiment cellulaire (plaste ou reticulum endoplasmique), nous définirons ces quatre gènes comme des oméga-3 et oméga-6 désaturases. L'étude porte donc sur deux isoformes de l'oméga-3 désaturase ( $\omega$ -3.1 et  $\omega$ -3.2) et deux isoformes de l'oméga-6 désaturase ( $\omega$ -6.1 et  $\omega$ -6.2). Sur les quatre gènes codant pour les oméga-3 et oméga-6 désaturases, seul l' $\omega$ -6.2 s'exprime à des niveaux élevés, entre des valeurs de 6.5 à 71,5, pour les trois espèces étudiées et les quatre conditions de température. L'expression des trois autres gènes est plus faible avec des valeurs comprises entre 0.02 à 5.5.

Pour trois des quatre gènes étudiés ( $\omega$ -6.2 et  $\omega$ -3.1/ $\omega$ -3.2), on constate que les profils d'expression génique de *C. eugenioides* suivant les régimes thermiques sont très différents des profils d'expression génique des deux autres espèces, *C. canephora* et *C. arabica*. Chez cette espèce, les profils montrent peu de variations suivant les différents régimes thermiques. Ces faibles variations observées ne sont pas corrélées avec les fortes variations des pourcentages relatifs observées chez 18 :2 et 18 :3. Il faut tout de même noter que pour *C. eugenioides*, seulement trois des quatre conditions de température sont représentées. La condition 33-30°C n'a pas pu être représentée suite à la perte importante de plantules à cette température. Les échantillons prélevés étaient en très faible quantité et l'ARN s'est avéré de mauvaise qualité, ce qui n'a pas permis l'analyse en q-RT PCR.

Les oméga-6 désaturases permettent la conversion du C18:1 (acide oléique) en C18:2 (acide linoléique). On remarque que les profils d'expression génique observés pour l' $\omega$ -6.1 sont bien corrélés aux profils des pourcentages relatifs en 18 :2 et cela pour les trois espèces étudiées (Figure. 5A. et 5C). Les pics importants de pourcentage relatif des trois espèces observés à la condition 23-20°C sont associés aux pics d'expression génique de l' $\omega$ -6.1 à la même condition (Figure. 5A. et 5C). Pour l' $\omega$ -6.2, on remarque que les profils d'expression génique observés ne sont pas corrélés aux profils des pourcentages relatifs en 18 :2 et cela pour les trois espèces étudiées (Figure. 5A. et 5D).

Notamment, les forts pics d'expression génique observés à la condition 18-15°C pour *C. arabica* et *C. canephora*, ne sont pas associés avec les pourcentages relatifs observés à la même condition (Figure. 5A. et 5D).

Les désaturases oméga-3 permettent la conversion du C18:2 (acide linoléique) en C18:3 (acide linoléique). A la condition 18-15°C, les pics d'expression génique observés pour l'ω-3.1 chez *C. arabica* et *C. canephora* sont corrélés aux forts pourcentages relatifs observés à cette même condition (Figure. 5B. et 5E). Chez *C. arabica*, la forte expression génique observée pour l'ω-3.2 à la condition 28-25°C est corrélée au fort pourcentage relatif observé à cette même condition (Figure. 5B. et 5F). A cette condition (28-25°C), le comportement original de l'allopolyploïde est confirmé par le niveau d'expression génique bien supérieur aux espèces parentales et par la très forte augmentation du pourcentage relatif de 18 :3 (Figure. 5B. et 5F).

## DISCUSSION

*Coffea arabica* et *Coffea canephora* sont des espèces tropicales cultivées d'intérêt économique majeur. Ces caféiers sont cultivés dans tous les pays d'Amérique du sud, d'Asie et d'Afrique situés dans la zone intertropicale humide (Doré, 2006). Leur culture s'est étendue à des zones où les températures sont plus chaudes ou plus froides que celles rencontrées dans les régions d'origine des espèces sauvages. Les intervalles de températures des zones de culture de *C. arabica* et de *C. canephora* sont donc plus importants comparés aux températures des zones de diversification primaire reflétant ainsi une certaine capacité adaptative de ces deux espèces. Il faut tout de même noter que, chez les caféiers, des effets négatifs sur la croissance des plantes sont observés dans des zones où les températures moyennes mensuelles sont inférieures à 15-16°C (Barros *et al.*, 1999) et la photosynthèse est affectée sous 18°C (Ramalho *et al.*, 2003). Les processus biologiques à l'origine de cette capacité à s'adapter à des conditions considérées comme défavorables sont encore très peu décrits. Quelques études se sont intéressées aux stress thermiques chez les caféiers (Fortunato *et al.*, 2009) mais ces données sont encore insuffisantes pour comprendre les réponses des caféiers aux fortes et faibles températures. D'un point de vue agronomique, il serait intéressant de mieux comprendre l'impact de ces stress thermiques sur la production de café. De plus l'étude de ces effets thermiques devient un enjeu dans le développement de stratégies contre les effets potentiel des changements climatiques globaux.

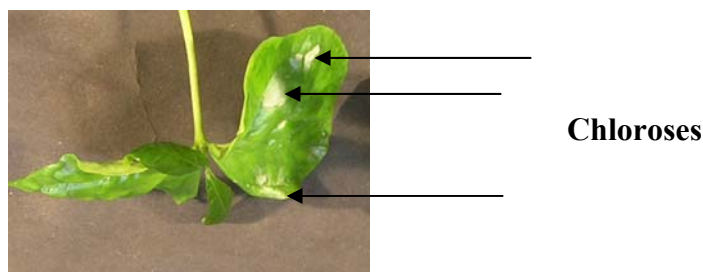
Les phénomènes de polyploïdisation sont considérés comme les clefs de la diversification des espèces végétales (Soltis, 2009) et de leur succès évolutif à long terme (Van de Peer *et al.*, 2009). Avec l'idée que les espèces polyploïdes auraient une adaptabilité supérieure et une tolérance accrue à des conditions environnementales variables, on peut émettre l'hypothèse que la polyploïdie aurait contribué à la survie et à la diffusion des espèces de plantes pendant ou à la suite d'évènements d'extinction produits lors de grandes crises climatiques et/ou géologiques (Fawcett *et al.*, 2009 ; Van de Peer *et al.*, 2009). En ce qui concerne notre modèle d'étude *C. arabica*, il est l'unique espèce allopolyploïde issue de l'hybridation entre deux espèces diploïdes de caféiers proches génétiquement, *C. canephora* et *C. eugenioides*. L'espèce allopolyploïde provient du Sud-ouest de l'Ethiopie ou aucune autre espèce de *Coffea* n'est présente et se développe à des altitudes et à des amplitudes thermiques différentes de celles de ses espèces parentales. Ce phénomène de colonisation d'une nouvelle niche écologique par *C. arabica* serait souvent observé chez les allopolyploïdes,

comme par exemple chez les plantes invasives (Prentis *et al.*, 2003), et témoignerait de la plasticité accrue observé chez ces espèces lors de stress abiotiques. Par exemple chez l'espèce invasive *Spartina*, une très grande plasticité phénotypique de l'allopolyploïde naturel (*Spartina anglica*) a été observée et se traduirait par une très large amplitude écologique et une tolérance aux conditions réductrices et riches en sulfites des sédiments (Ainouche *et al.*, 2004 et 2009 ; Maricle *et al.*, 2006).

Pour mettre en évidence la plasticité supérieure de l'allopolyploïde *C. arabica* sur ses deux espèces parentales (*C. eugenioides* et *C. canephora*), nous avons choisi d'étudier les variations en acides gras des feuilles et des gènes codant pour des enzymes de désaturation pour des régimes thermiques contrastés. Ce type d'étude est très fréquent chez les plantes tempérées avec une gamme plus large de température alors qu'elles sont rares chez les plantes tropicales. Nous avons choisi une gamme suffisamment importante pour l'étude d'une espèce tropicale. Nous constatons que même si cette gamme paraît peu importante, les teneurs en chlorophylle et les taux de croissance sont très affectés entraînant des pourcentages de mortalité très forts pour certaines espèces aux températures extrêmes.

#### ***Croissance et teneur en chlorophylle chez les trois espèces de caféiers***

En comparant l'allopolyploïde avec ses deux espèces parentales, il a été montré que le taux de croissance relatif diminuait chez *C. eugenioides* à la condition la plus chaude (33-30°C) et chez *C. canephora* à la condition la plus froide (18-15°C). A la condition la plus chaude, *C. eugenioides* a des difficultés d'adaptation pouvant conduire à la mort du plantule. A cette condition de température, on a constaté un taux de mortalité de 52% des plantules de *C. eugenioides*. A 18-15°C, *C. canephora* est particulièrement mal adapté se qui se traduit par une croissance très faible voire quasi nulle. A cette condition, on constate que *C. canephora* présente de forts symptômes de chlorose (Photo1).



**Figure 6 : *Coffea canephora* à 18-15°C**

Au cours des différents régimes thermiques imposés, *C. arabica* a montré un comportement beaucoup plus stable. Ce résultat est confirmé par les teneurs en chlorophylle. Pour la plupart des régimes thermique étudiés, la teneur en chlorophylle de *C. arabica* est supérieure à celles des deux espèces parentales. A la température la plus froide, la teneur en chlorophylle de *C. arabica* est également supérieure à celles des deux espèces parentales et notamment à celle de *C. canephora*, qui chute brutalement. Cette très faible teneur en chlorophylle expliquerait les chloroses observées chez les plantules témoignant d'une destruction de la chlorophylle. Comme le soulignent Na *et al.* 2007, les stress thermiques ont des effets sur la photosynthèse activant des processus de photo-inhibition. Selon ces auteurs, les complexes protéiques photosynthétiques des membranes de chloroplastes sont fortement affectés par une exposition à de fortes températures (Iba, 2002) et le fonctionnement régulier des chloroplastes est affecté aux basses températures (Routaboul *et al.* 2000; Iba, 2002).

***Corrélation entre l'activité des enzymes désaturases et l'abondance relative des acides gras chez les trois espèces de caféiers suivant des régimes thermiques contrastés***

Au cours de cette discussion, nous comparerons nos données observées à la bibliographie existante. Il faut toutefois noter que la grande partie des études sur les stress thermiques ont été effectuées sur des plantes tempérées. Comparées aux espèces tropicales, ces espèces supportent des extrêmes de température très importants (de 0 à 40°C). Les études associées à ces espèces utilisent souvent de très forts stress thermiques tel que le gel. La notion de stress thermique est donc différente entre les espèces tempérées et tropicales. Chez les espèces tropicales, un stress thermique peut être induit par un très faible écart de température par rapport sa température optimale de croissance.

Chez nos trois espèces caféiers, on observe la présence de mécanismes de désaturation. Aux températures les plus basses, la désaturation induite par l'augmentation des proportions en acide linoléique aux basses températures à également été observé chez des plantes tempérées. Iba *et al.* (2002) soulignent le fait que l'augmentation en acides gras triénoïques, tels que l'acide hexadécatriénoïque (16:3) et l'acide linoléique (18:3), dans les membranes des chloroplastes induisaient une tolérance à des faibles températures au cours des premières étapes de leur développement. Ces mécanismes de désaturation, pouvant témoigner d'une plasticité phénotypique et qui sont bien connus chez les plantes tempérées, existent aussi chez les plantes tropicales.

Au cours de l'étude des gènes codant pour des désaturases, on a mis en évidence une régulation transcriptionnelle en fonction des conditions de température. Pour une température donnée, les pics liés à l'activité transcriptionnelle sont corrélés aux pics liés à une

accumulation de métabolites. Cette relation entre l'activité transcriptionnelle et l'accumulation de métabolites a également été observé chez la graine de caféier *C. arabica* (Joët *et al.*, 2009) et pour trois grandes voies de biosynthèses (lipides, sucres et acides chlorogéniques). En ce qui concerne les gènes codant pour l'enzyme omega-6 désaturase, on peut dire que malgré les pics observés de l' $\omega$ -6.2 chez *C. canephora* et *C. arabica* à la condition de température la plus froide, son activité n'est pas suffisante pour assurer un développement normal de *C. canephora*. Malgré cette forte activité de l' $\omega$ -6.2, *C. canephora* n'arrive pas à s'acclimater à la température la plus froide. A la condition 23-20°C, on observe que l'activité de l' $\omega$ -6.1 joue un rôle majeur dans l'accumulation du 18 :2 permettant aux espèces de s'adapter à l'environnement. Il a été montré chez *Gossypium*, qu'une augmentation du niveau de C18:2 est associé à des hauts niveaux de transcription du gène codant pour la FAD2 (Oméga-6 désaturase) à basses températures (Kartigidiou *et al.*, 2008).

En ce qui concerne les gènes codant pour l'enzyme omega-3 désaturase, on observe que les pics de l' $\omega$ -3.1 chez les espèces *C. canephora* et *C. arabica* à la condition 18-15°C aurait un rôle dans l'accumulation du 18 :3 à cette même température. Il en va de même pour le pic de l' $\omega$ -3.2 observé chez *C. arabica* à 28-25°C qui induirait le pic de 18 :3 observé à la même température.

De façon globale, nous avons montré que des relations entre les quantités de métabolites (acides gras) et les quantités de transcrits (gènes codant pour les enzymes Oméga3- et Oméga6- désaturase) sont bien observées. Pour *C. arabica*, nous confirmons son aptitude à s'adapter à différents régimes thermiques. La forte expression génique de l' $\omega$ -3.2 spécifique à *C. arabica* corrèle fortement avec l'augmentation des quantités en 18 :3. Finalement, en ce qui concerne *C. canephora*, on observe que la sensibilité de l'espèce aux conditions de température les plus froides ne semble pas être reliée à un fort taux d'acide linoléinique ou à une activité transcriptionnelle particulière. Cette sensibilité devrait peut être pouvoir s'observer à partir d'une étude portant sur d'autres isoformes ou d'autres enzymes de la voie de biosynthèse des acides gras.

Pour l'espèce *C. eugenioides* on a observé très peu de variations entre les conditions de température pour trois des quatre gènes étudiés. On peut émettre l'hypothèse que l'isoforme permettant d'observer son activité n'a pas été étudiée ou que chez cette espèce la régulation génique est pourrait être post-transcriptionnelle.

## CONCLUSION

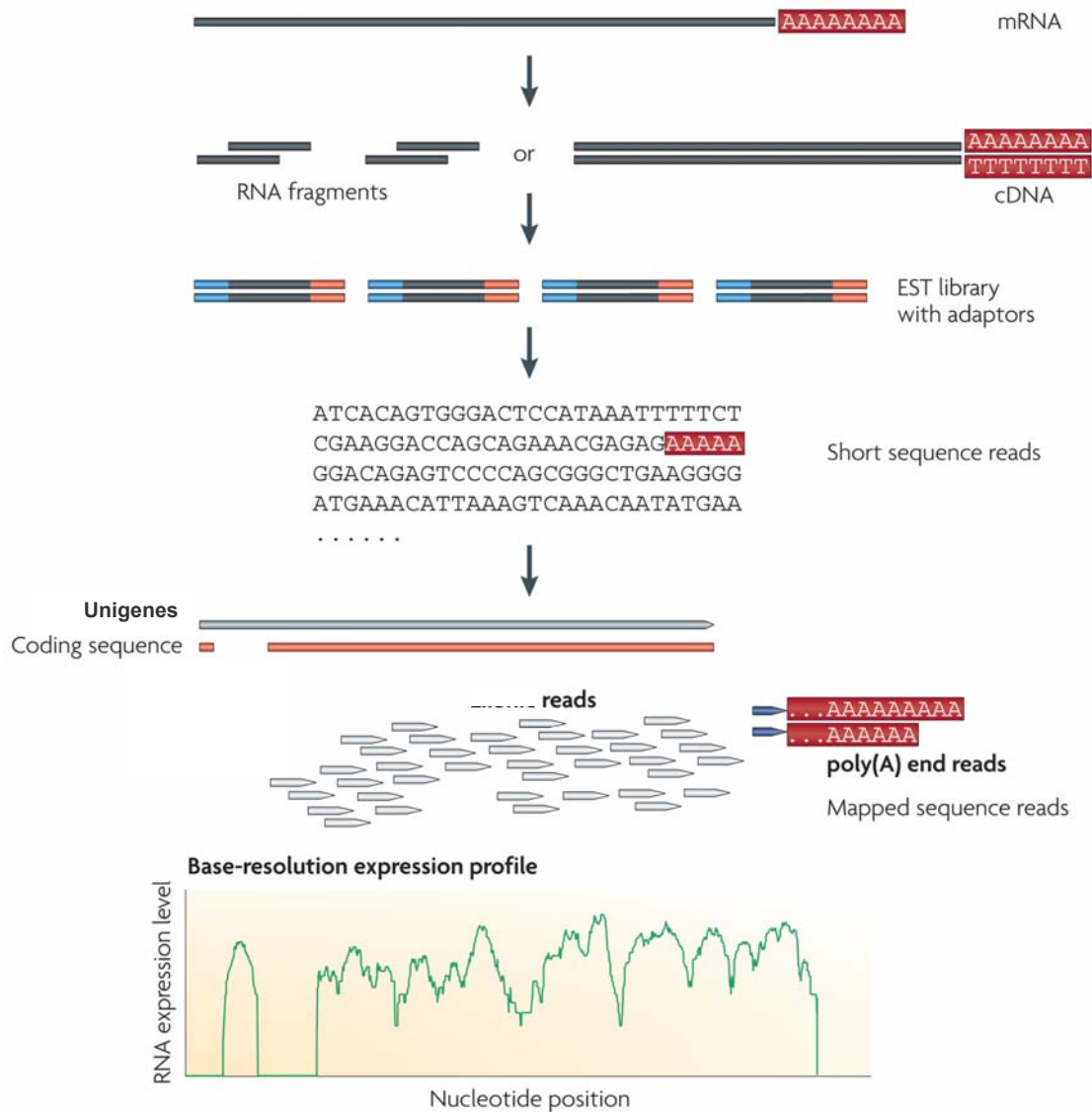
Cette étude, portant sur des analyses de la biomasse, de la teneur en chlorophylle ainsi que sur les variations en acides gras et l'expression génique codant pour les FADs, montre que l'espèce l'allopolyploïde *C. arabica* montre une plasticité supérieure à celle de ses deux espèces parentales. Quelles que soient les analyses effectuées, l'allopolyploïde apparaît souvent du même niveau que l'espèce parentale adaptée à la condition donnée ou supérieur à ses deux espèces parentales. En ce qui concerne la croissance on a observé une meilleure homéostasie de l'espèce allopolyploïde alors que la teneur en chlorophylle a montré que l'espèce allopolyploïde disposait d'une meilleure plasticité phénotypique que ses espèces parentales car il est capable d'adapter son comportement en fonction des conditions froides ou chaudes. La plasticité accrue de l'allopolyploïde est confirmée par les résultats de variations en acides gras et l'expression de gènes codant pour des enzymes désaturases. Les relations observées entre les quantités de métabolites et les quantités de transcrits montrent le comportement parfois original que peut adopter l'espèce allopolyploïde face à un stress thermique imposé. Ces analyses nous ont également permis de conclure que *C. canephora* montre des difficultés à s'adapter à des températures froides malgré le fonctionnement important de certaines enzymes de désaturation. Nos résultats montrent une relation entre l'activité transcriptionnelle et l'accumulation de métabolites mais que cette relation ne saurait à elle seule expliquer les variations enzymatique et génique observées ou le manque de variation sous stress abiotique. On peut émettre les hypothèses suivantes : que d'autres isoformes ou d'autres enzymes de la voie de biosynthèse des acides gras entrent en jeu ou qu'il y aurait une activité post-transcriptionnelle. Il faut aussi garder à l'esprit que la voie de biosynthèse des acides gras est un bon indicateur de la plasticité phénotypique et adaptative d'une espèce mais que d'autres facteurs tels que les variations observées chez les hormones ou chez d'autres voies de biosynthèse influent fortement sur l'adaptation des plantes au stress thermiques.





# **CHAPITRE IV**

**COMPARAISON DU TRANSCRIPTOME DE *COFFEA ARABICA*  
AVEC LES TRANSCRIPTOMES DE TROIS GROUPES GENETIQUES  
DE *COFFEA CANEPHORA* :  
PREMIERE UTILISATION DE LA TECHNIQUE RNASEQ CHEZ LE  
CAFEIER**



**Figure. 1: Protocole expérimental généralisé de la technique RNAseq**

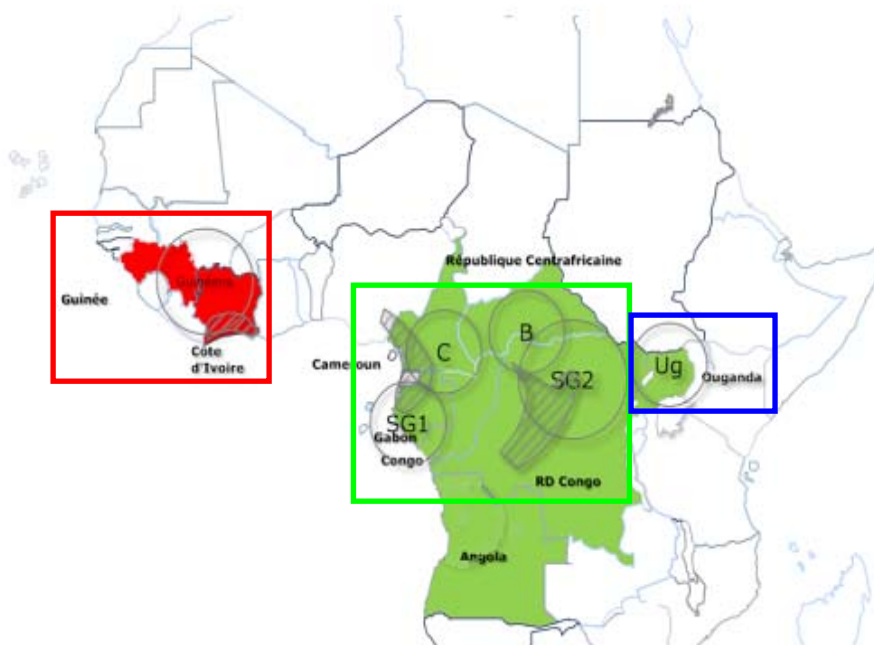
Les ARN sont dans un premier temps fragmentés pour construire la banque d'ADNc. Des adaptateurs (en bleu) sont alors attachés à chaque fragment d'ADNc, à partir desquels le séquençage sera initié. Après le séquençage, les 'reads' sont alignés à un génomeou transcriptome de référence et classés pour obtenir un profil d'expression avec une résolution à la base près.

Wang et al., 2009

## INTRODUCTION

Ces dernières années, le développement de Puces à ADN a permis d'analyser, de façon comparative, les transcriptomes. Elles permettent une mesure simultanée des niveaux d'expression de milliers de gènes répartis sur l'ensemble du génome. Le principe général des Puces à ADN est l'hybridation d'échantillons d'ADNc marqués par radioactivité ou fluorescence. Cette technique a été utilisée à de nombreuses reprises pour observer les variations d'expression génique chez les allopolyploïdie. La majorité de ces études traite des variations d'expression observées au niveau transcriptomique chez l'espèce allopolyploïde en comparaison à ses espèces parentales (Wang *et al.*, 2006 ; Ainouche *et al.*, 2009 ; Rapp *et al.*, 2009). Récemment, une nouvelle génération de techniques de séquençage à haut-débit (RNAseq), permettant d'aligner et de quantifier les transcriptomes, est apparue.

Le principe de base de la technique RNAseq, consiste en la conversion d'une population d'ARN en une librairie de fragments d'ADNc associés à des adaptateurs se trouvant à l'une ou aux deux extrémités (Figure 1). Chaque molécule, avec ou sans amplification, est séquencée afin d'obtenir des petites séquences ou 'reads' (variation due à la technologie utilisée). Les 'reads' sont alignés contre un génome ou transcriptome de référence et classés pour obtenir un profil d'expression (Figure 1 ; Wang *et al.*, 2009). Comparé à la technique de Puces à ADN souvent utilisée, la technique RNAseq n'est pas limitée dans la détection des transcrits correspondant à des séquences génomiques existantes (Wang *et al.*, 2009). Cette détection ouvre des possibilités d'études sur les organismes non modèles pour lesquels les séquences génomiques sont à déterminer. La technique RNAseq a un niveau de bruit de fond très faible comparé aux Puces à ADN car les séquences ADN sont alignées à une unique région du génome (Wang *et al.*, 2009). La technique RNAseq permet également d'observer les niveaux d'expression de gènes très fortement ou très faiblement exprimés. Par conséquent, elle permet la détection de transcrits de gènes que les techniques qRT-PCR ou Puces à ADN ne détecteraient pas. Des travaux préliminaires ont montré que les résultats de séquençage sont hautement reproductibles (Marioni *et al.*, 2008). Les niveaux d'expression génétique déterminés par cette méthode se corrèlent aux résultats obtenus par qRT-PCR. Cette technique permettrait des analyses dans différents tissus ou différentes conditions environnementales.



**Figure 2 : Principaux groupes génétiques de *C. canephora* sur la base de marqueurs neutres.**

Le cadre rouge représente le groupe génétique des Guinéens, le cadre vert représente le groupe génétique des Congolais (SG1, B, SG2 et C) et le cadre bleu représente le groupe Ougandais (Ug). La variété ‘Nemaya’, utilisée dans les études précédentes (Chapitre I, II et III), fait partie du groupe Congolais SG2. Dans ce chapitre, les individus utilisés proviennent du groupe génétique des Guinéens (clones GUI58 et GUI55), du groupe génétique des Congolais SG2 (CONG54 et CONG65) et du groupe génétique des Ougandais Ug (OUGB et OUGJ).

*Carte de C. Montagnon, 2011 en cours d'édition pour les Cahiers de l'Agriculture.*

Dans le chapitre II partie I, nous avons montré que les divergences transcriptomiques entre l'allopolyploïde et son espèce parentale *C. canephora* est variable en fonction des conditions de température. Pour représenter le parent *C. canephora* sur la PuceCafé, la variété 'Nemaya' appartenant au groupe génétique des congolais a été utilisée. Cependant, la grande diversité naturelle de *C. canephora* n'a pas été prise en compte et notamment pour les individus de *C. canephora* adaptés à des zones de haute altitude et des températures plutôt fraîches pour l'espèce (Figure 2). Le groupe génétique des Ougandais est bien adapté aux conditions de moyenne altitude et donc à des températures moyennes plus basses que celles des Guinéens et des Congolais. Au cours de ce chapitre, nous comparons le transcriptome de l'allopolyploïde *C. arabica* aux transcriptomes de trois groupes génétiques de *C. canephora* (Congolais, Ougandais et Guinéen) représentant une grande partie de la diversité naturelle de l'espèce. Comme pour le dispositif PuceCafé, différents régimes thermiques (30-26°C et 23-19°C) ont été utilisés.

Au cours de ce chapitre, nous tenterons : 1) d'une part, d'estimer la puissance de la technique RNAseq; 2) d'autre part, d'estimer la diversité des transcriptomes des *C. canephora* et de situer l'allopolyploïde *C. arabica* par rapport à cette diversité en terme de distance transcriptomique et de divergence transcriptomique.

<i>Espèces</i>	<i>Variété/Clones (Groupes génétiques)</i>	<i>Conditions de température</i>	<i>Nombre d'individus</i>	<i>Codes utilisés lors de l'étude</i>
<b>C. arabica</b>	<b>Caturra</b>	23-19°C	2	ARA1F et ARA2F
		30-26°C	2	ARA3C et ARA4C
<b>C. canephora</b>	<b>B/2/1 et J/1/1 (Ougandais)</b>	23-19°C	2	OUGBF et OUGJF
		30-26°C	2	OUGBC et OUGJC
	<b>BD54 et BD65 (Congolais)</b>	23-19°C	2	CONG54F et CONG65F
		30-26°C	2	CONG54C et CONG65C
	<b>BA55 et BA58 (Guinéen)</b>	23-19°C	2	GUI55F et GUI58F
		30-26°C	2	GUI55C et GUI58C

**Tableau 1 : Matériels biologiques et plan expérimental.**

Les seize individus sont représentés et classés suivant l'espèce, la variété et/ou le groupe génétique et la condition de température de croissance. Au cours de ce chapitre, des codes associés à chaque individu seront utilisés.

## MATERIELS ET METHODES

### Matériels biologiques

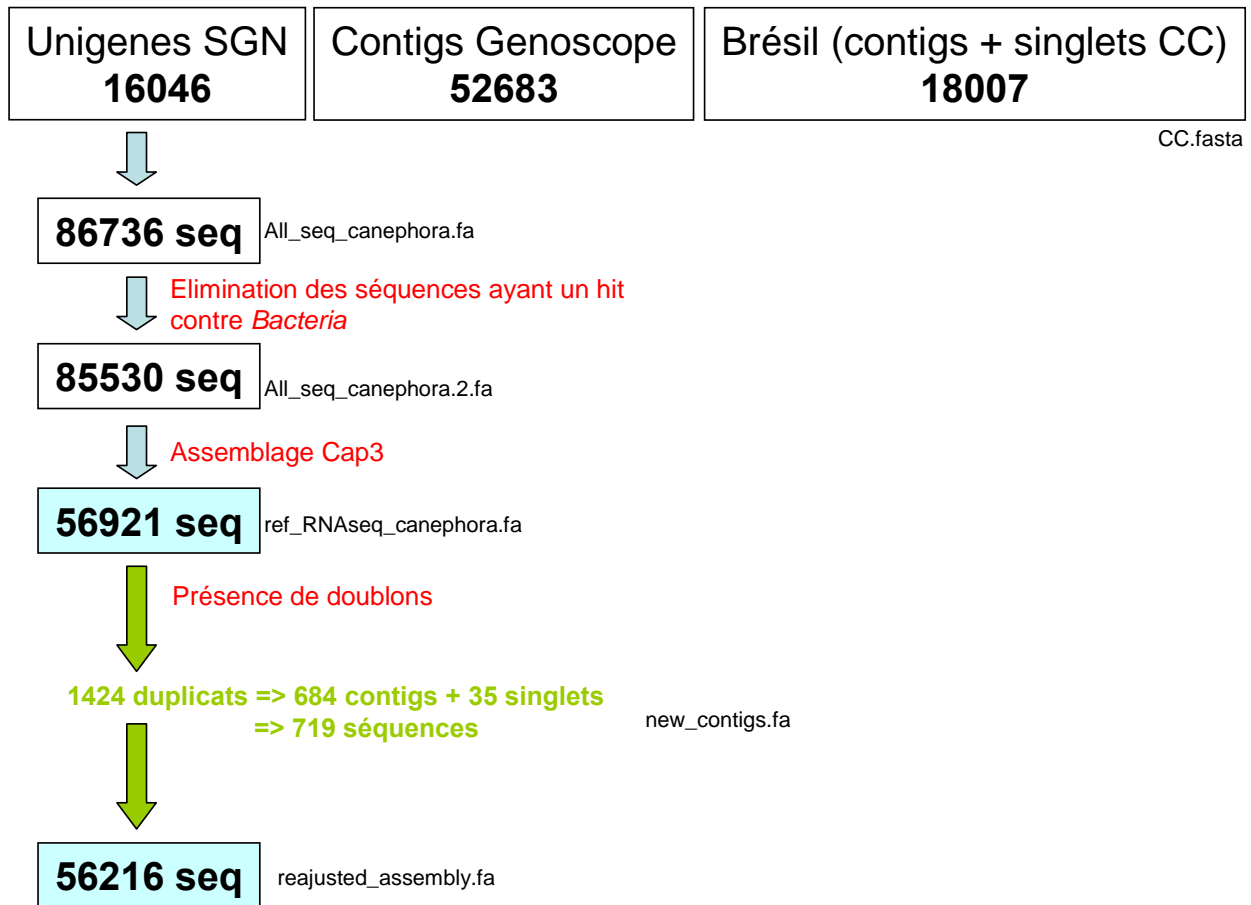
La variété de *Coffea arabica* utilisée est la variété Caturra issue, par mutation, de la variété *C. arabica* cv bourbon. *Coffea canephora* est représenté par des variétés sauvages issues de groupes génétiques différents, les Congolais, les Guinéens et les Ougandais. Les Congolais sont représentés par les clones BD54 et BD65, les Guinéens par les clones BA55 et BA58 et les Ougandais par les clones B/2/1 et J/1/1 (Tableau 1). La variété Caturra est représentée par quatre plantes et les clones de chaque groupe génétique de *C. canephora* par deux plantes.

Après 12 semaines de culture en serre (27°C), les seize plantes matures sont transférées dans des phytotrons (CRYONEXT, France, model RTH 1200L), où les conditions environnementales sont contrôlées. La photopériode, l'humidité et la luminosité sont respectivement fixées à 12:12 LD (Long Day), 80-90% et 600  $\mu$  mole photons  $m^{-2} s^{-1}$ . Pour cette étude, deux conditions de température ont été utilisées pour lesquelles des températures diurnes et nocturnes ont été définies : 30-26°C et 23-19°C. La variété Caturra est représentée par quatre plantes avec deux plantes par condition de température. Chacun des clones des trois groupes génétiques de *C. canephora* est représenté par deux plants avec une plante par condition de température.

Après 8 semaines de culture en phytotron, quatre jeunes feuilles ont été échantillonnées à la mi-journée (6-8 h après la nuit) pour chacune des plantes. Ces feuilles ont été directement congelées dans l'azote liquide puis stockées à -80°C jusqu'à l'extraction de l'ARN en vue d'analyses par RNAseq.

### Extraction des ARN totaux

Les ARN totaux ont été extraits à l'aide du protocole de Corre *et al.* (1996). Pour chacun des échantillons, environ 1g de feuilles a été broyé dans de l'azote liquide et homogénéisé dans 20ml d'une solution tampon contenant 5 M d'isothiocyanate de guanidium, 31 mM d'acetate de sodium (pH 8), 1% de  $\beta$ -mercaptoéthanol, 0.88% (w/v) de *N*-lauroyl sarcosine and 1% (w/v) de polyvinylpyrrolidone 40. Après centrifugation (15000g, 4°C, 20min), le surnageant est déposé sur un coussin de chlorure de césium (5.7 M) contenant 25mM d'acétate de sodium (pH 6) et centrifugé de nouveau (180000g, 20°C, 20h). Après élimination du surnageant, les culots d'ARN totaux sont lavés deux fois avec 70% (v/v) d'éthanol puis séchés pendant 1h. Les échantillons sont ensuite suspendus dans 100 $\mu$ l d'eau RNase free. Le dosage des ARN totaux a été effectué



**Figure 3 : Assemblage et nettoyage des données de la référence *C. canephora***

Trois jeux de données sont assemblés (16046+52683+18007). Les séquences matchant avec des séquences de *Bacteria* sont éliminées. Un assemblage des séquences est effectué. Après cet assemblage, on a constaté que des doublons étaient encore présents. Un deuxième nettoyage est effectué et 1424 doublons sont éliminés. La référence *C. canephora* est donc constituée de 56216 unigènes.



en mesurant la densité optique d'1 µl de solution à 260nm à l'aide d'un spectrophotomètre adapté aux petits volumes (Nanodrop, ND, Labtech, France).

## **Séquençage**

Le séquençage RNAseq a été effectué par la plateforme MGX (Montpellier Genomix, Institut de Génomique Fonctionnelle, Montpellier ; [www.mgx.cnrs.fr/](http://www.mgx.cnrs.fr/)).

Différentes ressources d'EST de *C. canephora* ont été utilisées pour construire le transcriptome de référence: le « SGN coffee unigene build III » (16046 clusters ; SOL, ref à indiquer), les clusters générés par la « Brazilian Coffee Initiative » (16801 clusters ; Vidal *et al.*, 2011) et des EST assemblés par le « French *C. canephora* sequencing consortium » (52683 clusters). Alors que les deux premiers sont basés sur des séquences Sanger, la dernière provient de 'reads' Solexa (76 nt par 'reads'). L'assemblage des trois ressources est faite à partir du logiciel 'CAP3 software' (<http://seq.cs.iastate.edu/>; Huang et Madan, 1999) avec un seuil minimal d'identité de 90% sur un minimum de 40 bases chevauchantes. Un total de 56216 (Figure 3) clusters est obtenu incluant 14222 contigs et 2584, 2294 et 37116 singlets respectivement de SGN, BCI et FCSC.

## **Préparation de la banque ARNm et hybridation**

La banque d'ARNm est faite à l'aide d'un kit 'RNA Seq sample prep.' (Illumina ; <http://www.illumina.com>). Les ARN sont fragmentés et des adaptateurs sont fixés à l'extrémité de chaque fragment d'ADNc. La validation de la banque est réalisée grâce à une quantification (concentration et taille des fragments) de l'ADN sur Puce Agilent DNA1000. Chacune des banques est dénaturée puis diluée à 5,4 pM avant d'être injectée (90 µl par banque) et hybridée à la Flow-cell. L'hybridation à la flow-cell est effectuée à l'aide du kit 'cluster génération kit' (Illumina ; <http://www.illumina.com>). La flow-cell est une lame de verre sur laquelle sont coatées de façon covalente des séquences d'ADN complémentaires à celles des adaptateurs ajoutées aux fragments d'ADN lors de la création de la banque. Chaque Flow-cell possède 8 pistes traitant 8 banques en même temps de façon indépendante. Une des pistes est réservée au contrôle PhiX.

## **Séquençage**

Le séquençage est réalisé sur un HiSeq 2000 (Illumina) grâce à la technique SBS (Sequence By Synthesis). Cette technique permet l'incorporation et la détection séquentielle de

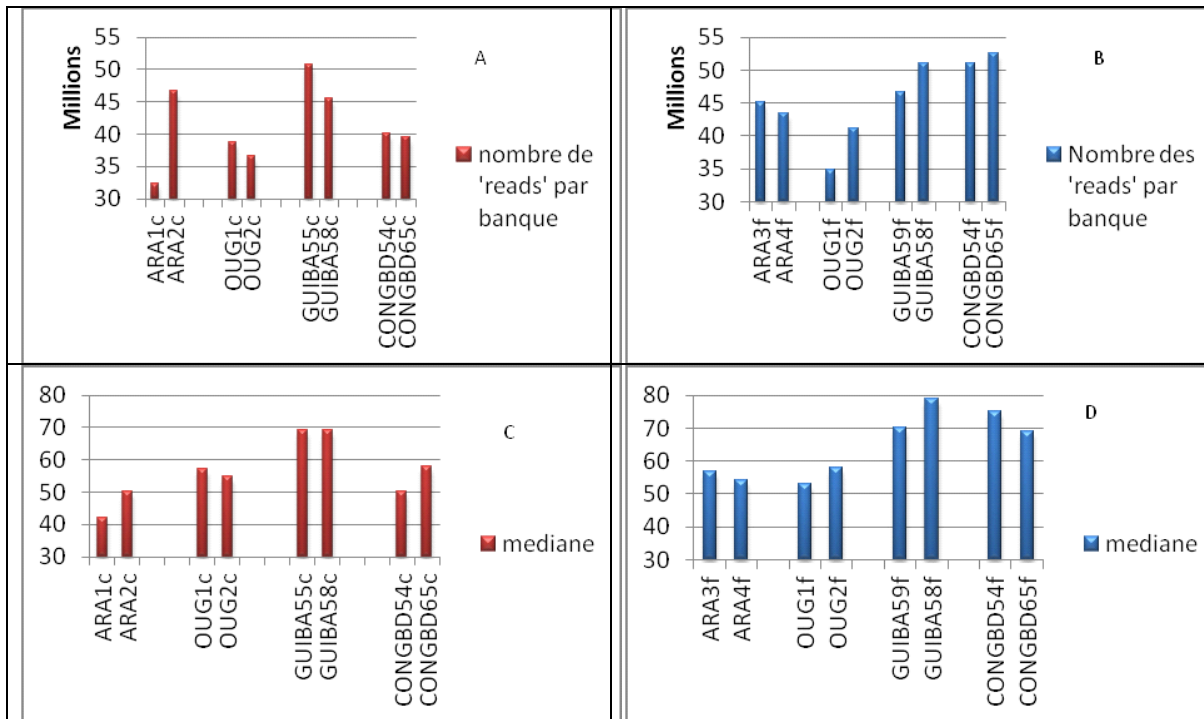
nucléotides est marqué avec un fluorochrome et un agent bloquant (empêche l'incorporation de plusieurs nucléotides au cours d'un même cycle de SBS). Le contrôle qualité du run s'effectue, à la fin du premier cycle, une fois que la première base est incorporée. Ce contrôle permet d'estimer la densité en cluster de chacune des pistes de la Flow-cell, l'intensité du signal généré et la qualité du focus de l'appareil.

### *Analyses statistiques*

Le séquençage de ces fragments a généré des dizaines de millions de 'reads' de 72 bp. Ces 'reads' ont été alignés sur une référence contenant 56921 unigènes de *C. canephora*. Au cours de cette expérience, seize échantillons ont été séquencés en 72 cycles. Pour une piste (piste 1, run 160, échantillon ARA1C), on obtient moins de séquences exploitables que les autres pistes. Ce résultat s'expliquerait par une trop forte densité en clusters bruts (chevauchements de clusters). Des fichiers indiquant le nombre de tags par contig ont été générés. Nous travaillerons uniquement sur les gènes s'exprimant au moins une fois, soit 55964 unigènes de *C. canephora*.

Dans notre dispositif, des deux répétitions biologiques par condition de température ont été utilisées. Dans la littérature, peu de dispositifs publiés utilisent des répétitions. Pour l'analyse statistique des résultats, nous avons comparé les logiciels édités récemment et disponibles sous R. Seul le logiciel DESeq, créé par Anders et Huber en 2010 et fonctionnant sous langage R (Bioconductor), prend en compte des répétitions dans les dispositifs expérimentaux. Ce logiciel a été sélectionné pour les analyses effectuées au cours de ce chapitre.





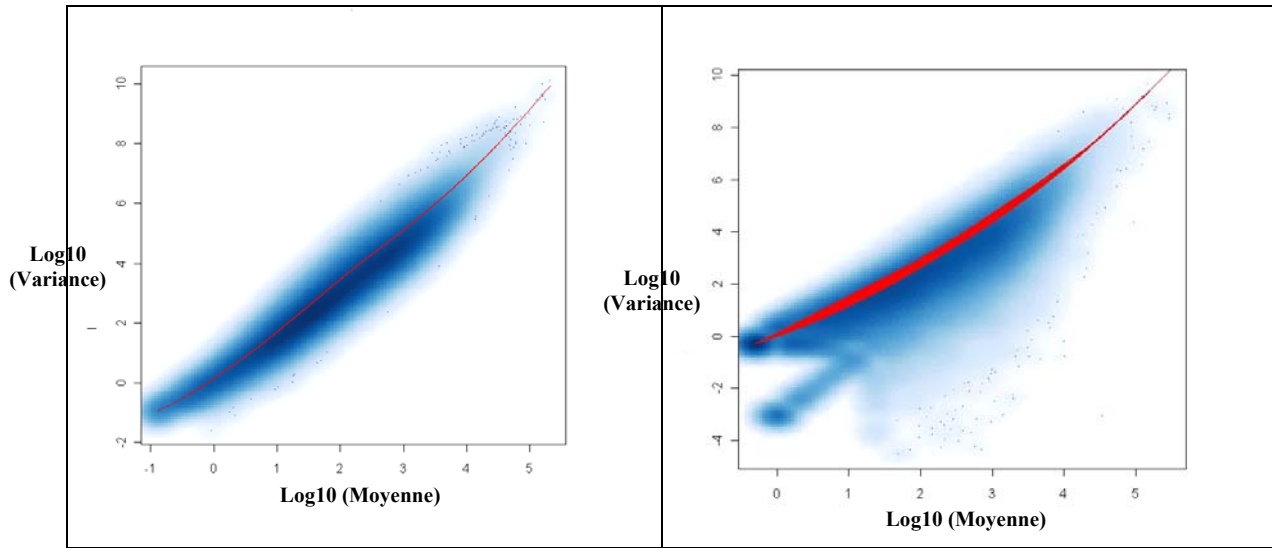
**Figure 4. :** Nombre de 'reads' et sa médiane sur les 55964 unigènes étudiés par RNAseq. Seize bibliothèques sont obtenues correspondant chacune aux huit individus étudiés dans les deux conditions de température, 23-19°C et 30-26°C. **A. B.** Nombre de 'reads' par bibliothèque aux conditions de température 30-26°C et 19-23°C. **C. D.** Médiane du nombre de 'reads' aux conditions de température 30-26°C et 19-23°C. On observe que les variations entre bibliothèques sont importantes au sein des tailles de bibliothèque et également au sein des médianes du nombre de 'reads' calculées par gène.

## RESULTATS

### Estimation de la puissance et de la répétabilité de la technologie RNAseq

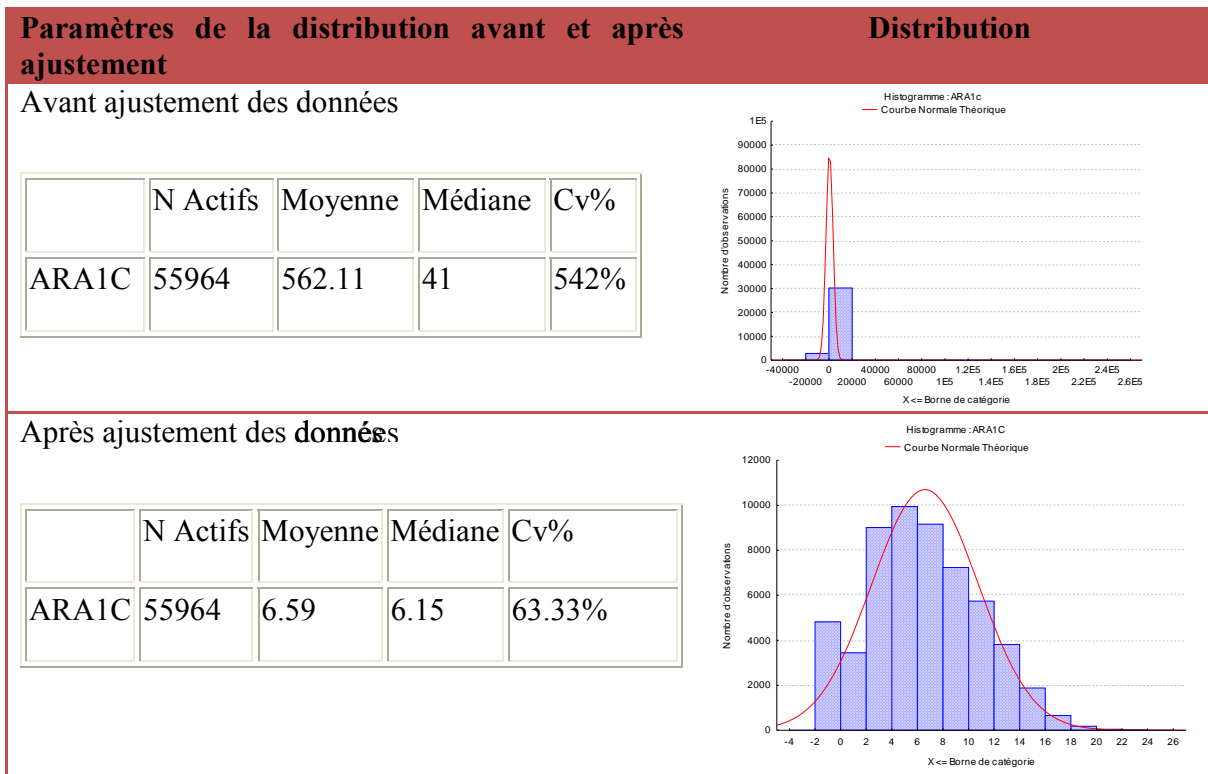
Après séquençage, seize librairies sont obtenues correspondant chacune aux huit individus étudiés dans les deux conditions de température, 23-19°C et 30-26°C. La taille des seize librairies et la médiane du nombre de 'reads' sont calculées par gène pour chacune des librairies (Figure 4). On observe que les variations entre librairies sont importantes pour les tailles de librairie et pour les médianes du nombre de 'reads' calculées par gène. Par librairies, on obtient entre 32 à 53 millions 'reads' (variation de 65%) et des médianes comprises entre 42 et 70 (Figure 4). On constate que ces variations entre librairies sont essentiellement dues aux variations de la librairie d'un individu de *C. arabica* à la condition 30-26°C (ARA1C). Comparé aux autres librairies pour lesquelles on obtient au minimum 37 millions de 'reads' avec des médianes à plus de 50, l'individu ARAC1 présente moins de 33 millions de 'reads' et une médiane à 42 (Figure 4).

L'intérêt de cette technique est basée sur la linéarité entre le comptage du nombre de 'reads' par gène et l'abondance des transcrits obtenus. Pour comparer entre elles deux librairies, il est préférable que le nombre total de 'reads' par librairie soit du même ordre de grandeur. Or dans notre étude on observe qu'entre deux librairies les ordres de grandeur sont différents. Pour pallier à cela, on rapporte le nombre de 'reads' observés par unigène et par condition au nombre total de 'reads' de la librairie. Si l'on considère que chaque librairie est indépendante des autres, cela revient par exemple à comparer la fraction du nombre de 'reads' pour le gène 1 de la librairie 1 qui est égale à  $100/33000000$ , à celle du même gène pour la librairie 6 qui est égale à  $120/53000000$ . Avec ce type de comparaison, on peut observer la présence de biais dans l'analyse. La distribution théorique de Poisson a été fréquemment utilisée pour estimer les probabilités mais présente des inconvénients. Anders et Huber (2010) montrent que les tests basés sur la loi de Poisson peuvent sévèrement sous-estimer les effets de la variabilité biologique en particulier pour les gènes hautement exprimés. Ils proposent d'utiliser des distributions basées sur la loi binomiale négative avec la variance estimée par des méthodes de régression locale. Après une réflexion sur les logiciels disponibles permettant ce type d'analyses, nous avons sélectionné le logiciel DESeq développé par Anders et Huber en 2010 et nous l'utilisons donc pour la suite des analyses.



**Figure 5. : Courbes représentant l'ajustement de la variance et la moyenne pour l'ensemble des 55964 unigènes étudiés.**

Le premier graphique (à gauche) représente les données de *C. canephora* (CONGF) à la condition de température 30-26°C. Le second graphique (à droite) représente les données de *C. arabica* (ARAC) à la condition de température 30-26°C.

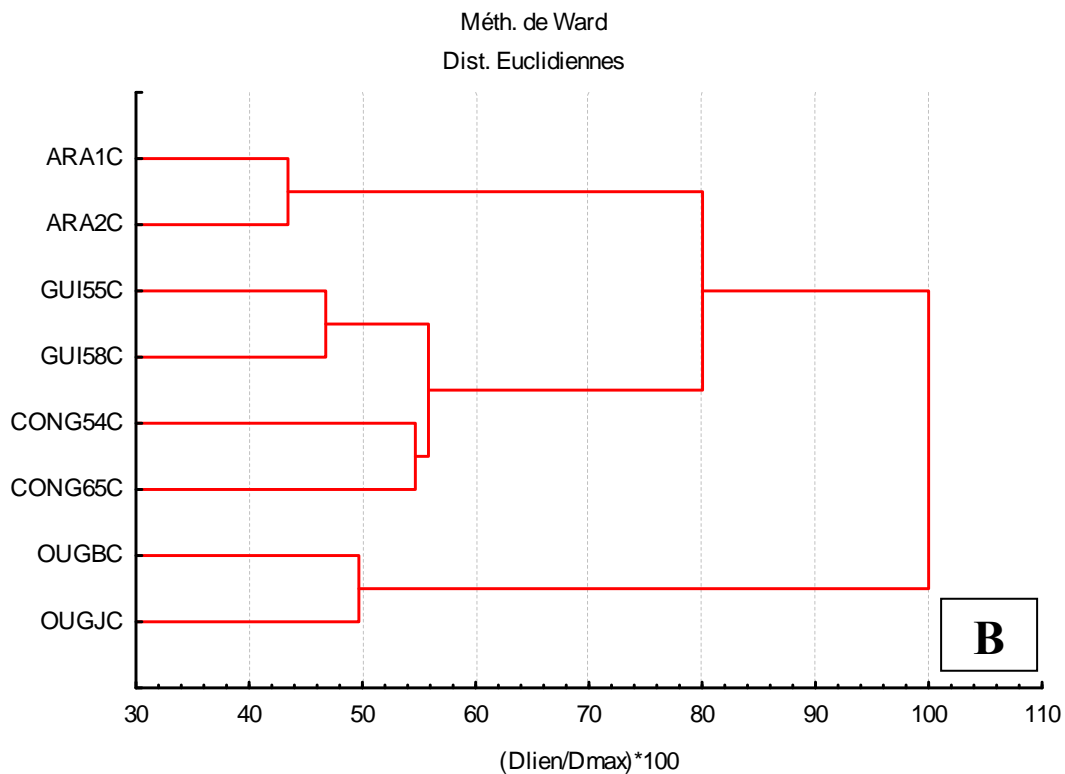
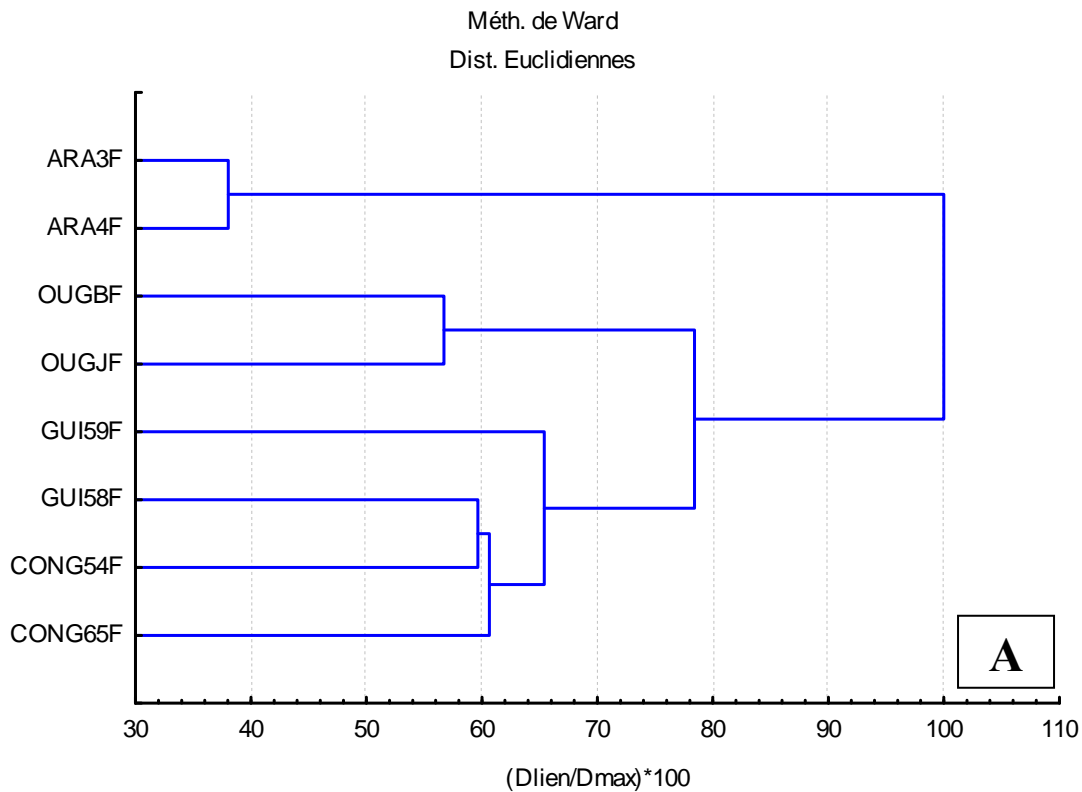


**Tableau 2. : Ajustement des données par le logiciel DESeq.**

Sur la première ligne, les statistiques de base avant correction et la distribution des données (histogramme). Sur la deuxième ligne, les statistiques de base après correction et la distribution des données suivant une loi normale.

Le logiciel DESeq prend en compte les répétitions pour estimer la variance. Le principe de cette méthode repose sur l'hypothèse que la moyenne est un bon prédicteur de la variance, c'est-à-dire que des gènes ayant un niveau d'expression similaire présenteront également une variance similaire. On calcule donc pour chaque gène la moyenne et la variance entre répétitions biologiques et on ajuste une régression sur ces données. Nous avons vérifié l'ajustement par la relation entre la variance et la moyenne (Figure. 5). Sur la figure 5, nous comparons le meilleur ajustement obtenu (bibliothèque de l'espèce *C. canephora* congolaise à froid 'CONGF') avec l'ajustement le moins bon (bibliothèque de l'espèce *C. arabica* à froid 'ARA1C'). Du aux faibles répétitions, la variance gène par gène dévie, parfois de beaucoup, de la valeur ajustée. Nous observons que l'ajustement suit l'estimation faite gène par gène, même quand cette dernière est très étendue. On peut vérifier également les qualités d'ajustement à partir d'un test de  $\chi^2$  et de sa probabilité associée, calculé sur l'écart entre la valeur ajustée et la variance gène par gène (données non montrées). Nous avons vérifié sur ces données que la probabilité associée aux tests de  $\chi^2$  n'était pas significativement supérieure à  $P < 0.05$  dans 98% des cas.

Le logiciel mis au point par Anders et Huber (2010) permet également de stabiliser les variances afin qu'elle soit indépendante de la moyenne. Lorsqu'on applique cette transformation aux données brutes, on obtient des valeurs avec des variances du même ordre de grandeur quelle que soit leur grandeur. Pour observer la puissance de la mesure de cette stabilisation, on a pris l'exemple de *Coffea arabica* à la condition 23-19°C (ARA1C) dont l'ajustement apparaît de moins bonne qualité dans la figure 5 (Tableau 2). Avant l'ajustement des données, on constate que la distribution ne suit pas une loi normale, que la moyenne et la médiane sont très différentes et que le coefficient de variation est très élevé (>500%), (Tableau 2). Après ajustement des données, la distribution suit une loi normale avec une moyenne et une médiane identiques et un coefficient de variation de 63% (Tableau 2). La correction des données brutes est donc très satisfaisante et la méthode s'avère adaptée à ce type d'analyse. Cette transformation est utilisée lorsqu'il s'agit de déterminer les distances entre transcriptomes.



**Figure 6. : Clusters issus d'une classification ascendante hiérarchique basée sur les distances entre les transcriptomes (méthode RNAseq).**

Les distances sont établies sur les variations de 'reads' des 55954 unigènes. **A.** Relations entre les clones et groupes génétiques à la température 23-19°C. **B.** Relations entre les clones et groupes génétiques à la température 30-26°C.

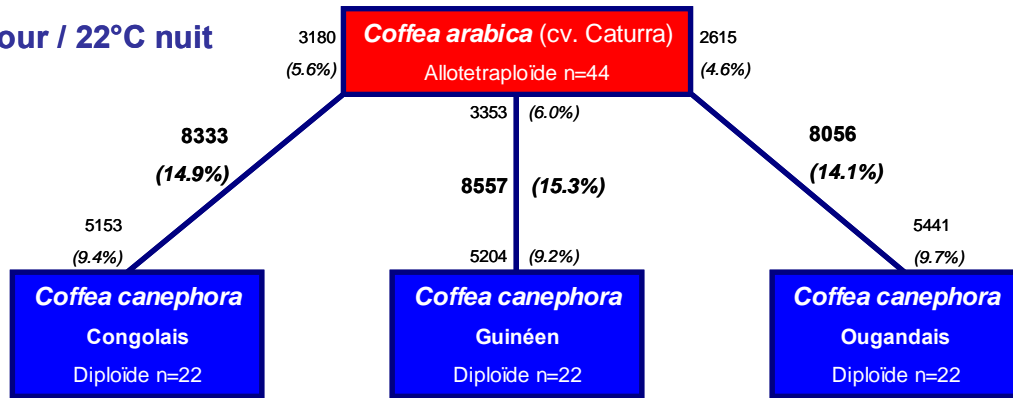


## **Variabilité des profils de transcriptome des trois groupes génétiques de *C. canephora* comparés à l'espèce allopolyploïde *C. arabica***

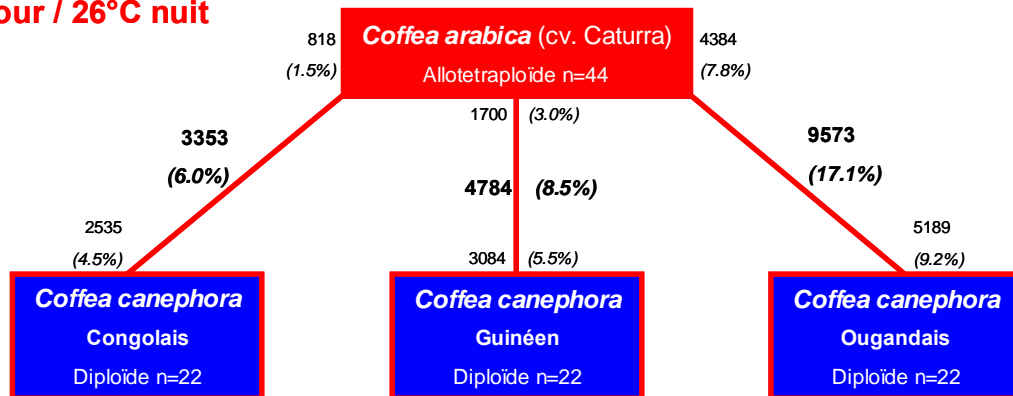
### ***Distances entre les transcriptomes des caféiers étudiés***

Pour estimer la diversité entre les groupes génétiques de *C. canephora* et leur relation avec *C. arabica*, nous calculons les distances qui existent entre les transcriptomes sur la base des données ajustées. La méthode utilisée est celle de Ward (Chapitre II, partie II) pour laquelle une métrique euclidienne permet de calculer des clusters suivant une classification ascendante hiérarchique (CAH). Les distances sont calculées sur l'ensemble des unigènes (différentiellement exprimés ou non). Les groupes génétiques des Guinéens et des Congolais apparaissent proches, quelle que soit la condition de température, formant un sous groupe homogène (Figure 6). A contrario, la distance transcriptomique du groupe génétique des Ougandais est plus importante quelle que soit la condition de température (Figure 6). A la température 23-19°C, le groupe génétique des Ougandais est plus proche des deux autres groupes génétiques de *C. canephora* que *C. arabica* (Figure 6). A cette condition de température, on a bien les groupes génétiques de *C. canephora* face à l'allopolyploïde *C. arabica*. A la condition de température 30-26°C, le groupe génétique des Ougandais est distant du groupes génétiques de *C. canephora* et de *C. arabica* (Figure 6). A cette condition, l'allopolyploïde est proche des deux groupes génétiques, Guinéen et Congolais, traduisant un comportement de *C. arabica* proche de certains écotypes de *C. canephora*. Les distances entre le groupe génétique de l'Ougandais et les deux groupes génétiques de *C. canephora* varient selon la condition de température. Les distances entre le groupe génétique de *C. arabica* et les trois groupes génétiques de *C. canephora* varient selon la condition de température.

**A 26°C jour / 22°C nuit**



**B 30°C jour / 26°C nuit**



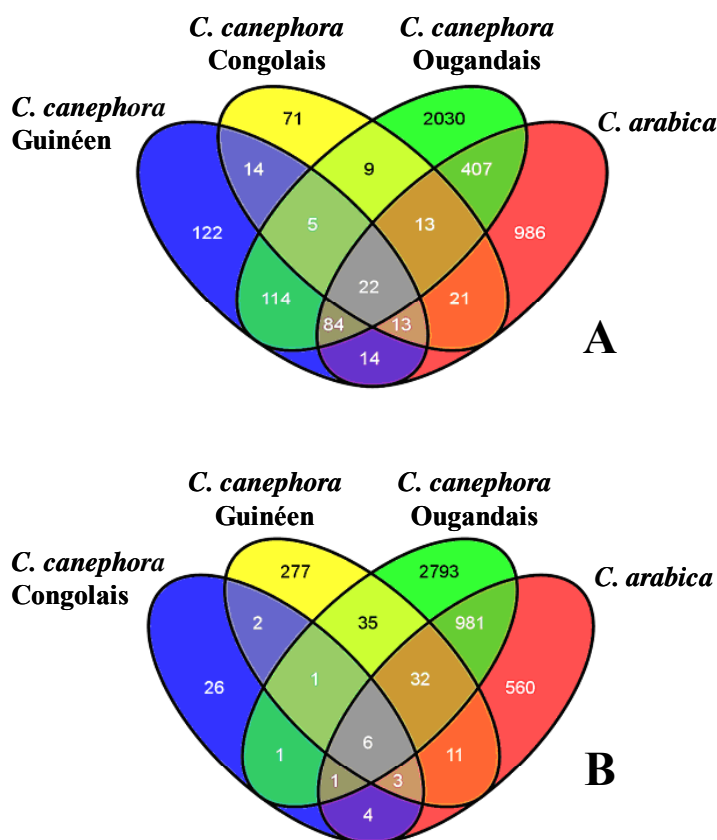
**Figure 7. : Divergence du transcriptome entre *Coffea arabica* cv Caturra et trois groupes génétiques de l'espèce parentale diploïde *Coffea canephora*.** Le texte en caractère gras indique le nombre total et la proportion d'unigènes qui sont définis comme significativement différenciellement exprimés entre *C. arabica* et les groupes génétiques du *C. canephora*. Le texte normal indique le nombre total d'unigènes et la proportion sur-exprimée. Par exemple à 23-19°C, 14.9% des unigènes sont indiqués comme différenciellement exprimés entre le groupe Congolais et l'allopolyploïde *C. arabica*. De ces, 8333 unigenes, 5.6% sont sur-exprimés chez l'allopolyploïde, et 9.4% chez le *C. canephora*. **A.** A la température la plus froide 23-19°C, environ 15% des unigènes sont différenciellement exprimés entre l'Arabica et les trois groupes. **B.** A la température la plus chaude 30-26°C, la situation est plus contrastée. Les pourcentages de divergence sont plus faibles entre *C. arabica* et les groupes Congolais et Guinéens (6 et 8.5%). En revanche, la divergence entre l'allopolyploïde et le groupe Ougandais est maintenue (17%).

***Divergences transcriptomiques entre l'allopolyploïde C. arabica et son espèce parentale C. canephora***

Nous avons calculé les divergences transcriptomiques entre *C. arabica* et les trois principaux groupes génétiques (Guinéen, Congolais et Ougandais) représentant la diversité de *C. canephora*. Les calculs ont été faits sur la base des 55965 unigènes appartenant au transcriptome de référence. Dans les deux conditions de température, on note que les unigènes de *C. arabica* sont sous-exprimés dans toutes les comparaisons effectuées avec les groupes génétiques de *C. canephora* (Figure 7). La divergence entre *C. arabica* et les groupes génétiques Congolais et Guinéens est plus importante à la condition de température 26-22°C. A la condition de température 29-13°C, on observe, respectivement, 14,9% et 15,3% de gènes divergents entre *C. arabica* et les groupes Congolais et Guinée (Figure 7). A la condition de température 30-26°C, on observe, respectivement, 6,0% et 8,5% de gènes divergents entre *C. arabica* et les groupes Congolais et Guinée (Figure 7). A contrario, on constate peu de différence entre les deux conditions de température pour les gènes divergents entre *C. arabica* et le groupe génétique des Ougandais.

	C. ARABICA	C. CANEPHORA Ougandais	C. CANEPHORA Congolais	C. CANEPHORA Guinéen
Nombre d'unigènes sur-exprimés à 30-26°C	1560	2684	168	388
Nombre d'unigènes sous-exprimés à 30-26°C	1638	3813	44	337
Total des unigènes différentiellement exprimés	3198	6497	212	773
Total exprimé en pourcentage (sur les 55965 unigènes)	5.7%	11.6%	0.4%	1.3%

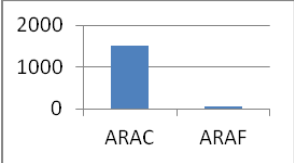
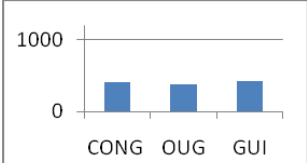
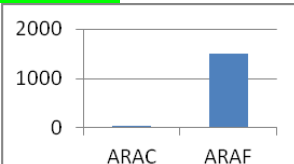
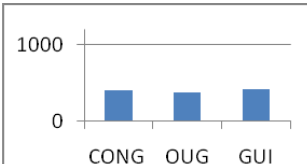
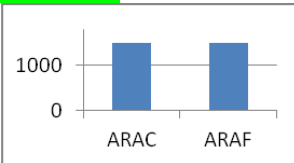
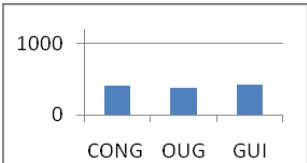
**Tableau 3. : Nombre de gènes sur et sous exprimés dans les comparaisons intra groupes et inter-températures.** Exemple pour le groupe Ouganda: les transcriptomes des deux clones OUGB et OUGJ à 23-19°C (OUGF) sont comparés aux transcriptomes des clones OUGB et OUGJ à 30-26°C (OUGC). Sur la base de  $p < 0.05$ , on détermine le nombre de gènes différentiellement exprimés et le nombre d'unigènes dans les deux sous-catégories sur et sous exprimées.



**Figure 8. : Diagrammes de Venn des gènes sur-exprimés. A.** en conditions de 'stress de chaleur' (30-26°C) par rapport à leur niveau d'expression à 23-19°C. On constate que 22 gènes sont communs aux trois espèces en réponse au stress de chaleur. **B.** en conditions de 'stress de froid' (23-19°C) par rapport à leur niveau d'expression à 30-26°C. On constate que seulement 6 gènes sont communs aux trois espèces.

Le nombre de gènes significativement sur-exprimés ou sous-exprimés à la condition de température 30-26°C (comparée à la condition de température 23-19°C) diffèrent suivant les groupes génétiques (Tableau 3). Pour le groupe des Ougandais, on estime qu'au moins 11% des gènes, soit 6497 unigènes, sont modifiés entre les deux conditions de température. Pour les groupes des Congolais et Guinéens, la divergence est très faible, seulement 0.37% à 1.38% (respectivement 212 et 773 unigènes pour les groupes génétiques des Congolais et Guinéens) des gènes varient de façon significative en fonction des conditions de température. Pour l'allopolyploïde *C. arabica*, on trouve plus de 5.6% des gènes qui varient entre les deux conditions de température.

Nous nous sommes intéressés aux gènes communs entre les quatre accessions et les gènes spécifiques à chacune des accessions. Pour les deux conditions, le nombre de gènes communs aux quatre accessions est très faible. Les gènes partagées entre le groupe génétique Ougandais et *C. arabica* sont de 14.10% à la condition 30-26°C et de 15.86% à la condition de température 23-19°C (Figure 8). Il faut souligner que, pour les groupes génétiques des Congolais et Guinéen, les gènes sous- et sur-exprimés à la condition de température 30-26°C (comparé à la condition de température 23-19°C) sont en très faible nombre ce qui conduit logiquement à partager peu de gènes.

Situation Arabica	Situation Canephora à chaud et à froid	Filtres proposés sur les fichiers générés par DESeq	Nombre d'unigènes sélectionnés sur les 55965 unigènes du génomme de référence
<b>Scénario 1</b>  	ARAC > ARAF & CF=CC & ARAC > (CC & CF)	90	
<b>Scénario 2</b>  	ARAF > ARAC & CC=CF & ARAC > (CC & CF)	151	
<b>Scénario.3</b>  	ARAC=ARAF & CF=CC & ARAC ou ARAF > (CC & CF)	276	

**Tableau 4. : Recherche des ‘gènes de la plasticité’ chez *Coffea arabica*.** Scénarios recherchés chez *C. arabica* et dans le groupe des *C. canephora*. Exemple du scénario 1: on cherche les unigènes d’Arabica (ARAC) fortement exprimés à la plus haute température (30-26°C) par rapport à l’Arabica (ARAF) à la plus basse température (23-19°C) et pour lesquels les Canephora sont faiblement exprimés. Deux quantités (nombre de reads) sont considérées similaires si  $\text{padj} > 0.10$  et  $\log_2(\text{foldchange})$  compris entre -1 et 1 ; deux quantités sont considérées différentes si  $\text{padj} < 0.05$  et  $\log_2(\text{foldchange}) > \pm 1.5$ .

## Les gènes traduisant la plasticité de l'allopolyploïde, *C. arabica*

A partir des données exhaustives obtenues par RNAseq, les recherches sur les gènes dits de plasticité sont facilitées. L'utilisation du logiciel DESeq, nous permet d'observer les variations géniques de l'espèce allopolyploïde, *C. arabica* aux deux conditions de température face aux variations géniques des trois groupes génétiques de *C. canephora* aux deux conditions de température. Pour cela, la variance est ajustée sur les six bibliothèques de *C. canephora* à 23-19°C et à 30-26°C et sur les deux bibliothèques de *C. arabica* à 23-19°C et à 30-26°C. Sur la base de ces comparaisons, nous avons sélectionné des cas de figures les plus contrastés illustrant la plasticité supérieure de *C. arabica* face à aux trois écotypes de *C. canephora* mais aussi l'aptitude de *C. arabica* à s'adapter aux conditions de température (Tableau 4). Ces cas de figures impliquent à chaque fois une expression génique similaire entre les trois écotypes de *C. canephora* permettant une comparaison exacte entre les variations géniques de l'allopolyploïde et du parent *C. canephora*. Deux seuils ont été fixés, un seuil permettant de regrouper les gènes qui sont similaires ( $p_{adj} > 0.10$  et  $\log_2$  du foldchange compris entre -1 et 1) et un seuil permettant de regrouper les gènes qui sont différents ( $p_{adj} < 0.05$  et le  $\log_2$  du foldchange supérieur à  $|1.5|$ ). A partir de l'utilisation de ces deux seuils, nous obtenons un nombre de gènes limité, soit 517 gènes (environ 1% du total des gènes). On observe que 90 gènes sont sur-exprimés chez *C. arabica* à la condition de température 30-26°C comparé à la condition de température 23-19°C et pour lesquels l'expression génique est faible chez les écotypes de *C. canephora*. On observe que 151 gènes sont sous-exprimés chez *C. arabica* à la condition de température 30-26°C comparé à la condition de température 23-19°C et pour lesquels l'expression génique est faible chez les écotypes de *C. canephora*. Et enfin 276 gènes ont une expression égale chez *C. arabica* entre les deux conditions de température et pour lesquels l'expression génique est plus faible chez les écotypes de *C. canephora*.

L'annotation du génome du caféier est pour l'instant partielle. Environ 18000 gènes sont annotés sur les 30000 attendus après le séquençage du génome. Il n'a donc pas toujours été possible d'annoter les unigènes que nous avons sélectionnés. Nous obtenons une liste de 139 processus biologiques (Annexe 11). Parmi les gènes dits de 'plasticité' on trouve environ la moitié des gènes qui codent pour des enzymes ayant un rôle dans différentes voies de biosynthèses telles que lipides, les sucres, les flavonoides, les acides aminées, de la chlorophylle et un groupe d'enzymes

« polymérase ». Des gènes codant pour des protéines de voie de signalisation sont également très nombreux. Comparé à ces deux groupes, on observe peu de facteurs de transcription.

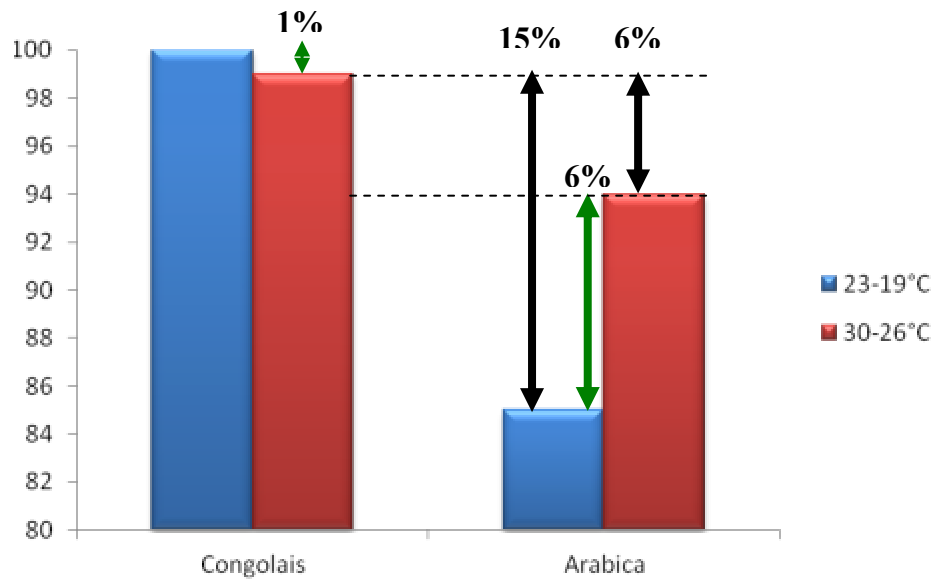
## DISCUSSION ET CONCLUSION

Au cours de ce chapitre, l'analyse du transcriptome d'espèces de caféiers au travers d'une nouvelle technique, RNAseq a été abordée. Contrairement aux Puces à ADN, le RNAseq ne contraint pas à un dispositif expérimental établi. Pour une expérience donnée, toutes les comparaisons sont possibles. Cette technique permet également de détecter des transcrits faiblement ou fortement exprimés avec précision. Néanmoins, la difficulté majeure est la gestion des réplicats biologiques. Pour cette étude, nous avons mis en place un dispositif expérimental à deux réplicats biologiques par condition de température. Pour pallier aux problèmes d'analyse liés à l'utilisation de réplicats biologiques, nous avons choisi un logiciel récemment développé, le DESeq, qui permet l'analyse de dispositifs expérimentaux avec répétitions. Ce logiciel permet d'ajuster les données en calculant, tout au long de la gamme dynamique, des variances estimées par régression locale s'ajustant à la moyenne et de vérifier que les données brutes, transformées par une formule basée sur la 'variance ajustée', suivent une loi normale (avec une déviation standard de l'erreur constante et non dépendante du nombre de reads). Au travers d'une étude basée sur les distances transcriptomiques entre espèces de caféiers, la répétabilité de la technique a pu être confirmée. Les distances observées entre réplicats biologiques d'une même condition de température (i.e deux individus d'un même groupe génétique x conditions de température) sont toujours plus faibles que les variations observées entre les groupes génétiques ou entre les conditions de température. L'ensemble de ces résultats confirme la puissance de la technique dans le cadre d'étude globale du génome.

L'étude des divergences et distances transcriptomiques a été menée sur un allopolyploïde, *C. arabica*, et trois écotypes de *C. canephora* issus de trois principaux groupes de diversité connus (Ougandais, Congolais et Guinéens). L'étude des distances transcriptomiques a mis en évidence des variations suivant les conditions de température. A la condition de température dite 'chaude' (30-26°C), le transcriptome de *C. arabica* apparaît proche de celui des groupes génétiques des Guinéens et Congolais. A la condition de température dite 'froide' (23-19°C), les individus de *C. arabica* forment un groupe bien distinct des individus appartenant aux trois groupes génétiques de *C. canephora*. Ces résultats montrent un comportement similaire entre l'allopolyploïde et les groupes génétiques Congolais et Guinéen à la condition de température la plus chaude (30-26°C).



La variabilité des transcriptomiques en fonction des conditions de température est démontrée chez l'espèce allopolyploïde, *C. arabica* et le groupe génétique Ougandais. Les résultats obtenus à partir des distances transcriptomiques sont confirmés par l'étude de la divergence transcriptomique entre l'allopolyploïde et les trois groupes génétiques de *C. canephora*. La divergence transcriptomique entre l'allopolyploïde et les groupes génétiques des Congolais et Guinéens est plus importante à la condition de température dite 'froide' (23-19°C) qu'à la condition de température dite 'chaude' (30-26°C). Ces résultats confirment également les résultats obtenus par la Puce à ADN, la PuceCafé (Chapitre II partie I). A partir des données Puce à ADN, on observe que la divergence transcriptomique passe de 49% à 8% à la condition de température 30-26°C. Cette différence est de nouveau observée à partir de la technique RNAseq (on passe de 15% à 6% à la condition de température 30-26°C). Une différence de proportion plus faible est observée pour la technique RNAseq. Cette différence d'estimation est probablement liée aux technologies utilisées. Cependant on note un comportement original du groupe génétique des Ougandais. La divergence transcriptomique entre *C. arabica* et les Ougandais est du même ordre de grandeur quelles que soit la condition de température. Les trois groupes génétiques de *C. canephora* ont une répartition écologique différente. L'espèce sauvage *C. canephora* est issue de forêts équatoriales humides de basses à moyennes altitudes, entre 250-1500m (Davis *et al.*, 2006), avec un optimum de températures moyennes annuelles pour l'espèce de 22-26°C (Damatta & Ramalho, 2008). Comparé aux écotypes Congolais et Guinéen, le groupe génétique des Ougandais est un écotype de *C. canephora* bien acclimaté aux zones d'altitude équatoriale et à des températures considérées fraîches pour l'espèce. A partir des données obtenues et de l'écologie des Ougandais, on peut émettre l'hypothèse que la condition de température la plus chaude modifie fortement son transcriptome et que les plantes issues de ce groupe pourraient être en conditions de stress. En ce qui concerne l'espèce *C. arabica*, on la retrouve dans des forêts tropicales de hautes altitudes, entre 1200-1950m (Davis *et al.*, 2006), avec un optimum de températures moyennes annuelles pour l'espèce de 18-21°C (Damatta & Ramalho, 2008). Comparée à l'écologie des groupes génétiques des Congolais et Guinéens, l'écologie de l'espèce allopolyploïde est plus proche de celle des Ougandais. Cependant, les résultats confirment tout de même le fait que l'allopolyploïde possède un comportement transcriptomique similaire aux groupes génétiques de *C. canephora* adaptés à des conditions de température relativement chaudes (30-26°C).



**Figure 9 : Représentation schématique des variations du transcriptome du *C. canephora* (groupe Congolais) et de *C. arabica* en fonction de deux régimes thermiques.** Les flèches noires indiquent les divergences entre *C. arabica* et le groupe des Congolais pour une condition de température. Les flèches vertes indiquent les divergences entre température pour une même espèce. Les pourcentages de divergence observés (cf Tableau 3 et Figure 7) sont reportés sur ce schéma. On a représenté la variation du transcriptome sur une échelle de 100%

Pour interpréter l'ensemble des résultats, nous représenté schématiquement des variations du transcriptome du *C. canephora* (groupe Congolais) et de *C. arabica* en fonction de deux régimes thermiques (Figure 9). Nous donnons uniquement l'exemple des Congolais car nous obtenons les même variations du transcriptome avec les Guinéens et le groupe Ougandais n'a pas été pris en compte du fait de son 'état de stress' présumé face à la condition de température 30-26°C. Sur la base de librairies indépendantes les unes des autres, nous constatons que la variabilité du transcriptome des Congolais est extrêmement faible (<1% des gènes). Face à la stabilité observée du transcriptome du groupe génétique Congolais, le transcriptome de *C. arabica* montrerait plus de variabilité dans les niveaux d'expression génique en fonction des différents régimes thermiques. Le transcriptome de *C. arabica* apparaît donc moins stable que celui de *C. canephora*. Ce qui semble traduire une plasticité transcriptionnelle plus importante de *C. arabica* que chez les Guinéens et les Congolais.



**DISCUSSION**

**ET**

**PERSPECTIVES**

## Les principaux résultats établis au cours de la thèse

### *Synthèse des résultats obtenus*

A partir des données Puce, on a montré que l'environnement modifie la divergence transcriptomique entre l'espèce allopolyploïde, *C. arabica*, et l'espèce parentale *C. canephora*. On observe que la divergence entre les deux espèces passe de 49% à la condition de température froide (26-22°C) à 9% à la condition de température chaude (30-26°C). Parmi les gènes divergents, une forte proportion se trouve en situation de non-additivité (37%). A l'aide d'un modèle génétique, l'expression a été catégorisée : l'additivité, la dominance 'Eugenioides-like', la dominance 'Canephora-like', la transgression et la catégorie 'No change'. On a observé qu'il y avait peu de différence en proportion de gènes entre les cinq catégories à 26-22°C, alors que les gènes sont répartis seulement dans deux catégories, les 'No change' et les gènes dominants 'Canephora-like' à 30-26°C. Le comportement original de *C. arabica* observé à la condition froide (16-18% de gènes dans la catégorie « transgressive ») laisse place à un comportement proche de *C. canephora* à la condition chaude (<3% de gènes dans la catégorie « transgressive »). Les données Puce ont également été utilisées in silico pour permettre la comparaison des distances transcriptomiques entre les conditions de température. A 26-22°C, les deux accessions de *C. arabica* sont assez éloignées de l'espèce *C. canephora* alors qu'à 30-26°C, on constate que *C. canephora* se retrouve très proche des deux accessions et même plus proche de la variété cv. T18141 que ne l'est la variété sauvage cv. Java. Comme pour l'étude comparant les espèces, on confirme un comportement de l'allopolyploïde proche de l'une de ses espèces parentales considérée comme l'espèce la mieux adaptée à la condition de température la plus chaude.

A partir de cette première approche transcriptomique nous permettant d'observer un comportement différentiel de l'allopolyploïde face à des conditions de température différentes, nous avons étudié la plasticité de cette espèce au travers d'une gamme thermique plus contrastée en s'appuyant sur des mesures biochimiques et génétiques. On observe pour les températures extrêmes, une forte mortalité de plantules de *C. eugenioides* à 33-30°C et des chloroses sur les plantules de *C. canephora* à 18-15°C. Lors de l'étude la biomasse, on observe que l'allopolyploïde montre très peu de variations de sa biomasse entre les quatre conditions de température. A contrario, *C. canephora* montre une chute de sa biomasse à 18-15°C et *C. eugenioides* une chute de sa biomasse à 33-30°C. On montre que l'allopolyploïde est moins affecté par les conditions de la gamme thermique que ses espèces parentales. En ce qui concerne

la teneur en chlorophylle et la composition en acide gras, on observe des comportements différentiels de l'allopolyploïde face à la gamme thermique. Nous avons observé que les proportions en acide palmitique (C16 :0) et en acide linoléique (C18 :3) sont proches de l'espèce diploïde la mieux adaptée à la condition de température donnée. L'étude des gènes associés à la voie de biosynthèse des acides gras confirme le comportement souvent original de l'espèce allopolyploïde comparée à ses espèces diploïdes.

Dans cette dernière étude, nous n'avons pas pris en compte la diversité de l'espèce *C. canephora*. Nous avons développé une étude à partir de la technique RNAseq afin de comparer le transcriptome de l'allopolyploïde et les transcriptomes de trois groupes génétiques de *C. canephora* (Congolais, Ougandais et Guinéens). Comme dans l'étude à partir de la Puce à ADN, on observe qu'à la température la plus chaude (30-26°C), *C. arabica* est plus proche des *C. canephora* qu'à la température froide (23-19°C). Cependant, cette observation n'est valable que pour deux des trois groupes génétiques. On observe un comportement différentiel du groupe Ougandais qui à 30-26°C s'éloigne des deux groupes génétiques Guinéens et Congolais ainsi que de *C. arabica*. Comme pour l'étude avec la Puce à ADN, nous avons observé le taux de divergence entre *C. arabica* et les groupes de *C. canephora*. Pour les taux de divergence entre deux des groupes (Guinéens et Congolais) de *C. canephora* et *C. arabica*, on observe des résultats qui vont dans le même sens que ceux obtenus grâce à la Puce à ADN. A 23-19°C, on obtient des pourcentages de divergence plus importants entre les groupes génétiques Congolais et Guinéen et l'allopolyploïde alors qu'à 30-26°C, les taux de divergence baissent. Le comportement différentiel du groupe des Ougandais lors de l'étude des distances génétique est de nouveau observé. En comparant ce groupe génétique et l'allopolyploïde, on constate peu de variation du taux de divergence entre les deux conditions de température. Nous émettons l'hypothèse que le comportement transcriptomique particulier du groupe des Ougandais au sein de *C. canephora* traduirait son adaptation génétique à des conditions de température fraîches. On peut donc penser que cette adaptation particulière à des zones plus froides confère une originalité au transcriptome des Ougandais. Nous émettons l'hypothèse qu'à la température la plus chaude ce groupe est mal adapté et que la divergence constatée représente en fait un stress thermique important de ce groupe à cette condition de température.

### ***Critiques et perspectives associées aux résultats***

Par l'étude de la divergence et de la distance transcriptomique et l'étude de la composition lipidique, nous avons établi la plasticité et l'originalité de l'allopolyploïde. Cependant des critiques peuvent être apportées à l'ensemble du travail.

### ***Résultats technologie-dépendants?***

La divergence transcriptomique entre les espèces a été observée à partir de la technique de Puce à ADN et de RNAseq. Comparé au RNAseq, la Puce à ADN que nous avons choisi permet des comparaisons deux à deux suite à une double hybridation. Cette double hybridation pourrait entraîner un biais vers l'une des espèces co-hybridées. Toutefois, nous avons montré que l'on n'observait pas de biais (Chapitre I). Cependant dans le chapitre II partie II, le plan expérimental de la Puce ne permettait pas la comparaison entre conditions de température et les données ont donc été normalisées *in silico*. Par ailleurs, nous observons qu'il y a autant d'unigènes qui sont surexprimés que sous-exprimés et que leurs quantités sont du même ordre de grandeur quelle que soit l'espèce étudiée. S'il y a un biais dû à la normalisation, il n'a pas été détecté lors de l'étude.

Lors du chapitre IV, nous avons utilisé des données du séquençage des transcrits par RNAseq pour déterminer la divergence et la distance transcriptomique entre les espèces et groupes génétiques. Ces données ont été obtenues à partir d'un transcriptome de référence obtenu à partir d'une des deux espèces parentales, *Coffea canephora*. Le principe de base de la technique RNAseq, consiste en la conversion d'une population d'ARN en une librairie de fragments d'ADNc associés à des adaptateurs se trouvant à l'une ou aux deux extrémités. Chaque molécule est séquencée afin d'obtenir des petites séquences qui sont alignés à un génome ou transcriptome de référence. L'obtention d'un génome ou transcriptome de référence complet permettant des comparaisons exactes entre allopolyploïde et espèces parentales est en cours. En ce qui concerne notre modèle d'étude, la référence est constituée de 56216 unigènes appartenant à une seule des espèces de caféiers étudiés, *C. canephora*. Lors de l'alignement, chacune des séquences obtenues est alignée au transcriptome de référence. Si les séquences ne s'alignent pas au transcriptome de référence, elles sont éliminées de l'analyse. On peut donc supposer que la perte de séquences de *C. arabica* est plus importante que la perte de séquence des groupes génétiques de *C. canephora*. Cette observation pourrait être expliquée le fait qu'il y a systématiquement plus de gènes surexprimés pour les trois groupes de *C. canephora* que chez *C. arabica*. La technique RNAseq est moins coûteuse que les Pucés à ADN, plus facile à analyser, plus précise et plus répétable



surtout lorsqu'on prévoit des répétitions comme nous l'avons fait. Toutefois des comparaisons entre espèces même proches, peuvent conduire à un biais des résultats lorsque le transcriptome de référence est représenté par une seule espèce, en l'occurrence dans notre étude par *C. canephora*. S'agissant d'étude sur les distances transcriptomiques, une sous-estimation quasi-systématique et aléatoire du nombre de 'reads' chez *C. arabica* conduit à un biais sur l'estimation des distances entre espèces étudiées. Le même biais peut être observé chez la PuceCafé car l'ensemble des ESTs de l'étude proviennent de l'espèce *C. canephora*. Cependant comparé à la technique RNAseq, ce biais n'a pas été détecté.

### Les régimes thermiques utilisés

Pour établir la plasticité de *C. arabica*, nous avons choisi de faire varier les conditions de température. Notre choix a été déterminé par le fait que la température est probablement le facteur de l'environnement qui a le plus favorisé la sélection de nouvelles espèces. Il s'agit d'autre part d'un paramètre fondamental pour la plupart des réactions physico-chimiques de la plante et d'un facteur relativement facile à contrôler expérimentalement. Aujourd'hui, il s'agit de nouveau d'un enjeu majeur pour l'adaptation de la culture des caféiers face à des changements climatiques globaux.

Etant en présence d'espèces tropicales cultivées dans toute la ceinture intertropicale, nous avons utilisé une gamme thermique adaptée mais qui peut paraître faible si on la compare à celles utilisées chez les plantes tempérées. Cependant, nous pensons que cette gamme convient à notre étude. Dans l'expérience avec la Puce à ADN, les températures étaient relativement chaude (30-26°C et 26-22°C) pour les espèces de caféiers. Ces températures, ne permettaient pas d'observer un stress au froid mais ont toutefois permis d'observer qu'aux températures chaudes, le transcriptome de *C. arabica* est proche de celui d'une des espèces parentales, *C. canephora* (appartenant au groupe des Congolais). Pour le RNAseq, nous avons élargi l'écart entre les deux conditions environnementales à 23-19°C et 30-26°C. Les résultats RNAseq confirment les résultats obtenus avec la Puce, c'est-à-dire qu'à la température la plus chaude *C. arabica* est proche du groupe Congolais de *C. canephora*. Pour les analyses par qRT-PCR, la gamme thermique la plus large a été utilisée, 15°C à 30°C de nuit et 18°C à 33°C de jour. A partir de cette gamme nous avons pu établir la plasticité de *C. arabica*. Pour améliorer l'interprétation des résultats sur gamme thermique il serait nécessaire d'obtenir et de réunir plus d'information sur

l'écologie des trois espèces notamment pour l'espèce sauvage *C. eugenioides* qui a été très peu étudiée et dont la diversité génétique est inconnue.

Dans le chapitre III, malgré de forts niveaux transcriptionnels de certaines désaturases associés à de fortes proportions d'acides gras insaturés, nous avons constaté que *C. canephora* était mal adapté à la température la plus froide. Chez *C. eugenioides* on observe un cas de figure très différent. On a constaté un manque d'activité des gènes chez *C. eugenioides*. On peut émettre les hypothèses suivantes : soit nous n'avons pas étudié les bon isoformes, soit d'autres voies de biosynthèse sont affectées soit la régulation génique est de nature post-transcriptionnelle.

Pour compléter ces études, il serait intéressant :

- D'étudier plus d'isoformes de gènes codant pour les enzymes étudiées, de généraliser l'étude à d'autres gènes codant pour les enzymes intervenant dans la voie de biosynthèse des acides gras et d'autres voies de biosynthèse pouvant avoir un rôle dans l'adaptation à des conditions environnementales extrêmes.

- D'étudier la plasticité des différents groupes de *C. canephora* sur la gamme thermique utilisée pour *C. arabica* et montrer si ces groupes sont adaptés de façon spécifique à des températures de croissance particulière.

- D'améliorer le dispositif expérimental en utilisant une référence transcriptomique plus adaptée à la comparaison d'espèces de caféiers, c'est-à-dire construite à partir d'unigènes provenant des trois espèces étudiées. Ce dispositif expérimental permettrait de confirmer les comportements originaux parfois observés chez *C. arabica* lors de l'étude des acides gras et l'étude de gènes supplémentaires de la voie de biosynthèse des acides gras par RNAseq.

On pourrait également s'intéresser à la vigueur hybride. Ni *et al.* (2009) soulignent que, dans des conditions environnementales contrôlées, il y a un accroissement de la vigueur assimilable à de la vigueur hybride chez un allotétraploïde néosynthétique d'*Arabidopsis* (*A. thaliana* x *A. arenosa*) par rapport à ses deux parents. Dans cette étude, les auteurs montrent que l'allopolyploïde présenterait un meilleur contrôle du rythme circadien jouant un rôle central dans la croissance se traduisant par la synthèse accrue de chlorophylle et d'amidon. A notre connaissance, c'est la première fois que des auteurs décrivent un système de régulation particulier et en démontrent les mécanismes pour expliquer la plus grande vigueur hybride des allopolyploïdes et des hybrides. Il serait d'étudier ce système de régulation chez les Caféiers.

- D'obtenir plus d'information sur l'espèce *C. eugenioides* et sur sa diversité génétique. Pour l'instant, n'ayant peu de ressources génétiques sur *C. eugenioides*, les expériences de la Puce et celle du RNAseq ont été faites sur des ESTs ou sur un ensemble de données génétiques appartenant à un seul parent, *C. canephora*. Dans le chapitre IV, on constate que d'intégrer la diversité des espèces dans nos modèles d'études est important. Il serait donc intéressant de faire une étude de diversité comme nous l'avons faite pour *C. canephora*.

### **Contribution de nos résultats à la compréhension des mécanismes de spéciation des allopolyploïdes**

La polyploïdie est un mécanisme majeur d'adaptation et de spéciation chez les végétaux (Ramsey et Douglas, 1998, Otto et Whitton, 2000 ; Adams, 2007). Des marques de polyploïdies plus anciennes sont fréquemment observées chez les génomes des plantes. Ces paléopolyploïdes ont subi au moins un évènement de duplication du génome entier (WGD) suivi d'une perte massive de gènes et d'une réorganisation génomique par diploïdisation (Wolfe, 2001 ; Chen, 2007). Chez les angiospermes, des traces de paléopolyploïdie sont présentes dans presque toutes les espèces étudiées (Cui *et al.* 2006) et pour nombre d'entre elles, plusieurs évènements de polyploïdie se sont superposés au cours du temps évolutif confirmant le caractère récurrent de cet état dans l'évolution des espèces.

Dans des environnements adaptés aux parents diploïdes, les polyploïdes et diploïdes se développent et cohabitent alors que dans des environnements non adaptés aux parents diploïdes tels que les hautes altitudes et latitudes où les températures sont plus froides, les polyploïdes survivent mieux que leur parents diploïdes (Chen, 2007). L'adaptation à de nouvelles niches écologiques peut être illustrée par le cas des plantes invasives et des plantes arctiques. Comme le souligne Prentis *et al.* (2008), la majorité des plantes invasives sont polyploïdes. Cette hypothèse peut s'expliquer par des différences de fitness ou de compétition entre des populations ayant différents niveaux de ploïdie et par une meilleure colonisation des polyploïdes dans des milieux extrêmes comme cela a été observé dans la flore arctique (Prentis *et al.*, 2008). Brochmann *et al.* (2004) souligne que le succès de l'évolution des polyploïdes dans la zone arctique serait basé sur l'état d'hétérozygotie fixé de leur génome. Cet état permettrait une meilleure survie des espèces pendant des périodes climatiques instables.

Pour décrire la capacité que peuvent avoir les espèces à s'adapter à des environnements variables, trois concepts sont mis en avant (Whitman et Agrawal, 2009 ; Pfenning *et al.*, 2010) :

- L'adaptation génétique qui est définie par la spécialisation d'un génotype ou d'un phénotype à une condition environnementale donnée.

- La notion de « Jack-of-all-trades » qui se définit par le fait qu'un génotype produit un phénotype généraliste. Ce phénotype est en général moins efficace pour chaque environnement mais produit tout de même une efficacité moyenne supérieure.

- La plasticité phénotypique qui est définie par la capacité d'un génotype à produire des phénotypes différents chacun adapté à un environnement correspondant. L'allopolyploïde serait un génotype très plastique qui serait sélectionné lorsque les conditions environnementales sont défavorables aux espèces parentales. Sous ces conditions défavorables, les espèces parentales ont un temps trop court pour leur permettre une adaptation génétique face à des changements climatiques majeurs tels que les grandes crises biologiques.

Les espèces allopolyploïdes sont issues d'une hybridation interspécifique et d'un doublement génétique entre des espèces parentales proches génétiquement. Cette hybridation interspécifique ne peut avoir lieu que si les barrières génétique, comportementale, écologique et/ou géographique pouvant exister sont faibles, c'est-à-dire que les espèces sont génétiquement proches, qu'elles partagent une même zone géographique (vivent en sympatrie). Pour que les hybrides interspécifiques nouvellement formés soient sélectionnés pour évoluer en tant qu'espèce polyploïde, il faut que les conditions environnementales leurs soient favorables (c'est-à-dire défavorable aux espèces parentales), que la sélection naturelle les favorise notamment par rapport aux back-cross possibles avec des espèces diploïdes proches génétiquement et qu'un système de reproduction efficace et une méiose régulée se mettent en place puis sont sélectionnés. Ces conditions permettraient de sélectionner un phénotype allopolyploïde conduisant à la formation d'une nouvelle espèce mieux adaptée à une niche écologique nouvelle.

A partir de l'étude, nous avons tenté d'établir que,

- *Coffea arabica* présenterait une capacité adaptative supérieure à ses espèces parentales lui permettant de faire face à des changements environnementaux ce qui expliquerait qu'il ait été sélectionné puis qu'il ait évolué comme une espèce à part entière.

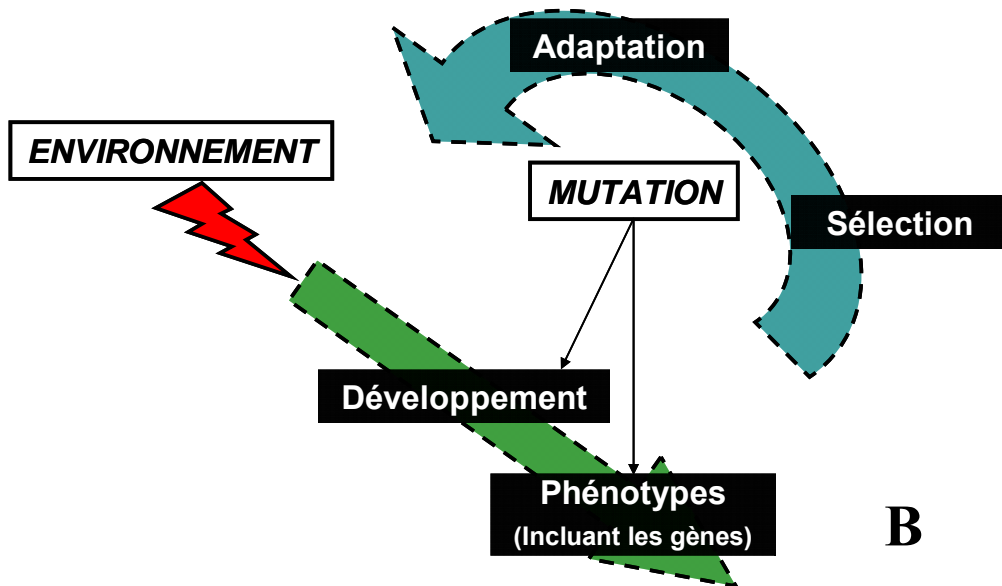
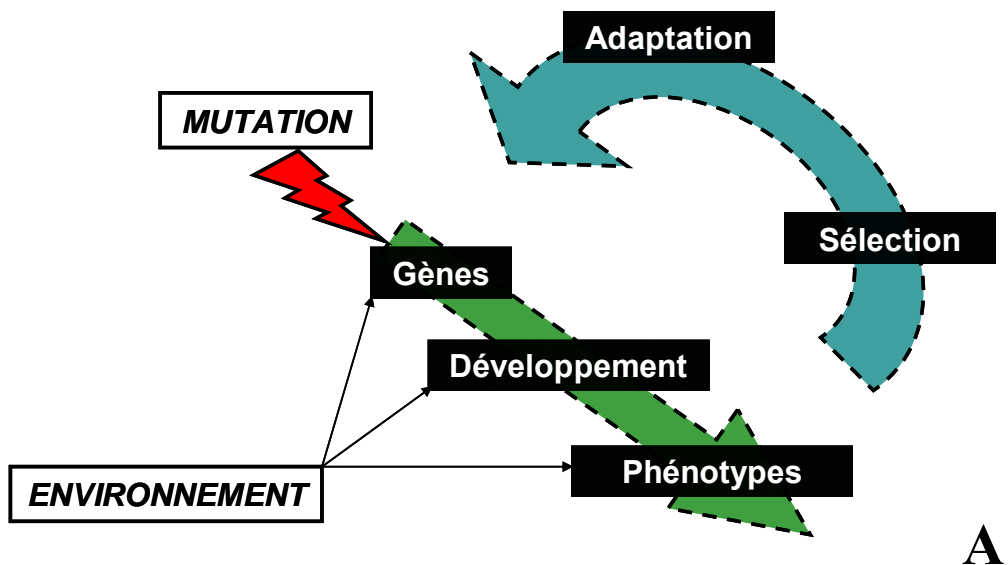
- Les capacités adaptatives de l'espèce seraient dues majoritairement à des variations des niveaux d'expression en fonction des conditions environnementales. C'est-à-dire que le phénotype allopolyploïde serait plastique suite à des modifications de régulation de son expression.

### ***Notions de plasticité phénotypique, de plasticité adaptative et de canalisation***

La plasticité phénotypique est définie par la capacité d'un génotype à produire des phénotypes différents chacun adapté à un environnement correspondant (Travis, 1994 ; West-Eberhard, 2003). Pour notre modèle nous observons que pour l'étude de la biomasse il y a peu de variations observées sur la gamme thermique alors que pour la teneur en chlorophylle, l'abondance relative en AG et l'expression génique des gènes codants pour des enzymes de la voie de biosynthèse des acides gras un comportement original de l'allopolyploïde est parfois observé. L'hypothèse permettant d'expliquer ce phénomène est l'association de phénomène de canalisation et de plasticité phénotypique.

La canalisation est considérée comme le phénomène opposé de la plasticité phénotypique. Les facteurs internes lors du développement, la physiologie ou le comportement réduisent l'influence des stimuli environnementaux pour produire un seul phénotype quel que soit l'environnement. (Waddington 1940, 1942 ; David *et al.*, 2004). Cependant, la canalisation et la plasticité phénotypique ne sont pas incompatibles. Chez un organisme complexe, on observe fréquemment de la 'canalisation' pour un des caractères et de la plasticité phénotypique pour d'autres (Whitman et Agrawal, 2009). Par exemple, chez les insectes il est fréquent d'observer un phénomène de canalisation de la taille des œufs en condition de malnutrition associé à un phénomène de plasticité phénotypique de la taille de la 'cuvée' (Stearns 1992, Nylin and Gotthard 1998, Fox and Czesak 2000). En définitive ce qui s'avérait une réponse non adaptative passive (diminution de la taille des couvées) s'avère en fait une réponse plastique pour maintenir la taille des œufs (Whitman et Agrawal, 2009). La canalisation qui peut être observée d'un point de vue physiologique et/ou développemental se met souvent en place en parallèle au phénomène de plasticité phénotypique.

La plasticité phénotypique associée à la canalisation confirmerait une plasticité adaptative de l'espèce (Ghalambor *et al.*, 2007). La plasticité adaptative permet établir l'espèce de façon définitive dans de nouveaux environnements et de contribuer à l'évolution des capacités adaptatives des espèces face à la modification de environnement ou à de nouveaux environnements sur de longues durées. Comme nous les avons décrit précédemment, on peut prendre en exemple les espèces invasives (Prentis *et al.*, 2008) ou les plantes arctiques (Brochmann *et al.*, 2004).



**Figure 1 : Scénarios possibles à l'origine de l'adaptation des espèces.**

**A.** Scénario incluant une modification génétique à l'origine de l'adaptation d'une espèce. **B.** Scénario centré sur l'influence des conditions environnementales sur l'adaptation des espèces.

*D'après West-Eberhard, 2005*

***Théories sur l'apparition et l'adaptation de nouvelles espèces à des milieux défavorables aux espèces parentales.***

La théorie existante de l'évolution est basée sur les modifications génétiques apportées aux espèces et a toujours considérée la plasticité comme un bruit de fond. De nos jours la plasticité phénotypique est reconnue comme ayant un fort potentiel d'adaptation, confirmant son rôle majeur dans l'évolution. D'après West-Eberhard (2005), on peut représenter de manière cyclique ce qui se passe chez les espèces. Deux théories sont mises en avant opposant les modifications génétiques (Figure 1.A) et le rôle de l'environnement dans la sélection et l'adaptation des espèces (Figure 1.B). Dans le cas particulier de l'allopolypléidie, nous pensons que l'hybridation entre deux espèces proches est rendue possible par des modifications temporaires ou permanentes de l'environnement (Figure 1B) qui influent momentanément ou définitivement sur les barrières qui étaient en place entre les espèces. Des phénotypes allopolypléïdes nouveaux, dont la plasticité est supérieure à celles des espèces parentales, apparaissent et sont ensuite sélectionnés favorisant leur maintien sous forme de nouvelles espèces. La nouvelle espèce formée pourra soit perdurer, soit se croiser de nouveau avec l'une de ses espèces parentales. Si la nouvelle espèce se maintient dans un environnement en allopatrie, alors elle pourra subir des mutations et conduire à la diploïdsation.

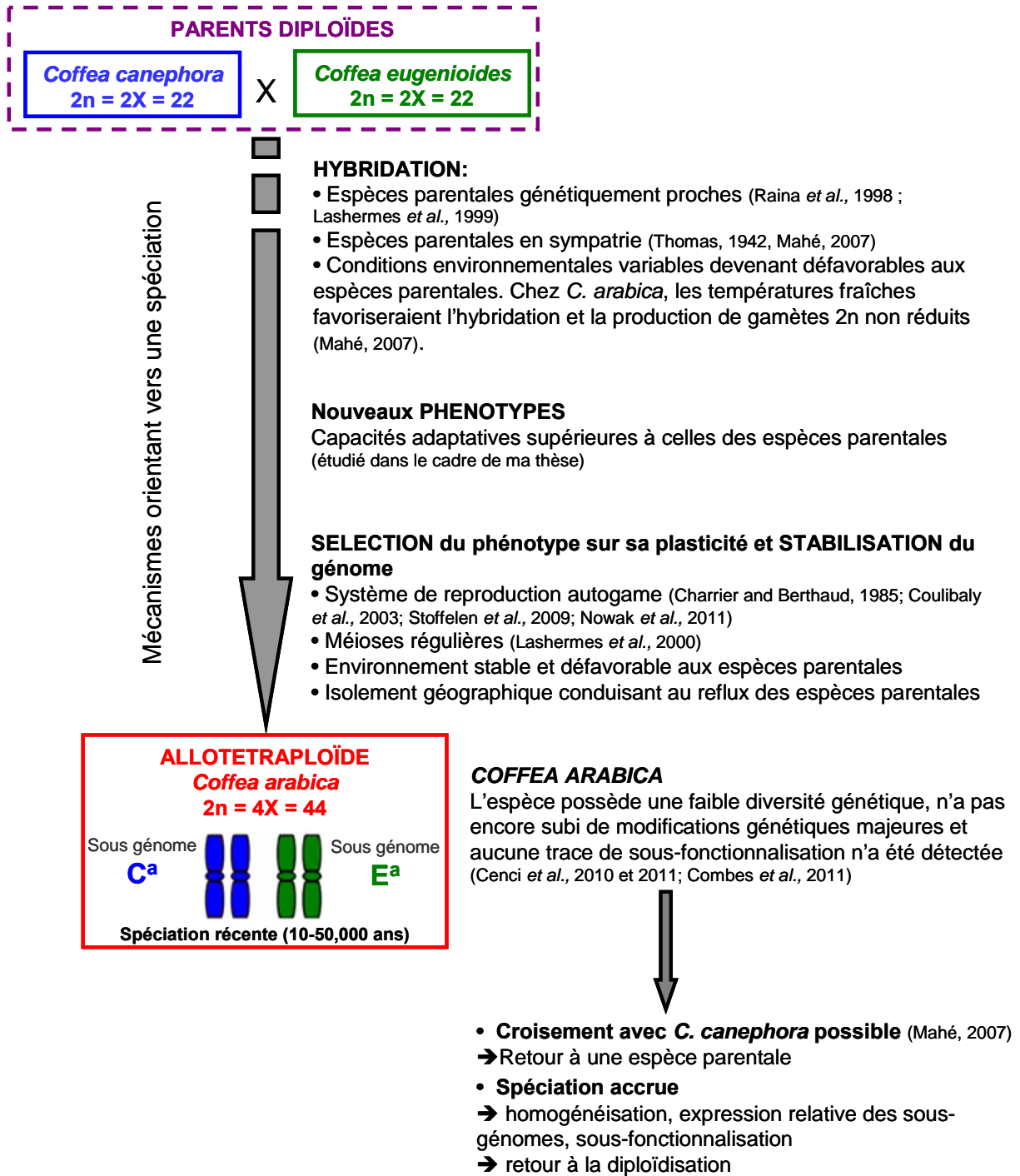


Figure 2 : Scénario sur la formation et la stabilisation de *C. arabica*.



### **Scénario sur la formation de notre modèle d'étude, *C. arabica***

La diversification du genre *Coffea* sous-genre *Coffea* s'est produite par radiation rapide et récente (460000–100000BP) avec un centre d'origine du genre se trouvant probablement en basse Guinée (Anthony *et al.*, 2010). Les caféiers sont des plantes pérennes appartenant à la famille des Rubiaceae. L'analyse comparative d'une région génomique de *Coffea* et des régions orthologues d'espèces du groupe des Rosideae (*Vitis vinifera*, *populus Trichocarpa* et *Arabidopsis thaliana*) a montré que ces deux groupes proviennent de même génome ancestral hexaploïde pour lequel le processus de diploïdisation était bien avancé au moment de la divergence entre les deux groupes et qu'aucun évènement supplémentaire de polyploïdisation ne s'est produit au cours de l'évolution du genre *Coffea* (Cenci *et al.*, 2010).

*Coffea arabica*, la seule espèce polyploïde du genre *Coffea*, est issu d'une hybridation interspécifique et d'une duplication du génome entre deux espèces parentales africaines proches génétiquement, *C. eugenioides* et *C. canephora*. L'analyse des fréquences de substitution de régions génomiques orthologues des espèces *C. arabica*, *C. canephora* et *C. eugenioides* a révélé que cet évènement d'hybridation entre les deux espèces diploïdes ne s'est pas produit avant 10000 à 50000ans (probablement lors de la dernière glaciation survenue il y a 18000 ans). La faible divergence de ces mêmes régions génomiques, la présence d'éléments transposables partagés et une synténie des gènes presque parfaite indiquent que les génomes diploïdes à l'origine de *C. arabica* sont très peu divergents (Cenci *et al.*, 2011). La présence d'hybrides interspécifiques entre l'espèce diploïde *C. canephora* et l'espèce allopolyploïde apparentée *C. arabica* indique que les barrières génétiques post-zygotiques ne sont pas infranchissables dans les conditions environnementales de Nouvelle-Calédonie. La présence en sympatrie de *C. eugenioides* et de *C. canephora* dans les montagnes Ougandaises a également été rapportée dès 1942 par Thomas. Dans une étude menée par Mahé en 2007, il a été observé en Nouvelle-Calédonie où les températures fraîches des nuits et plus particulièrement les variations importantes de température entre le jour et la nuit observées dans les montagnes sont probablement à l'origine de la production de gamètes 2n non réduits permettant le doublement chromosomique. Des zones de sympatrie ont permis des flux de gènes et plus particulièrement l'incorporation dans le patrimoine génétique d'une population de gènes d'espèces différentes (Mahé, 2007). L'importance des flux de gènes entre *C. arabica* et son parent *C. canephora* en conditions naturelles, nous conduirait à penser qu'il y a pu avoir une phase de coexistence stable entre *C. arabica* et ses espèces parentales ancestrales diploïdes (Thomas, 1942 ; Mahé, 2007).

Notre scénario prévoit que sous les conditions environnementales défavorables aux espèces parentales, de nouveaux phénotypes allopolyploïdes sont créés. Ces phénotypes montrent parfois des capacités adaptatives supérieures les conduisant à la colonisation du nouveau milieu. Les individus possédant une méiose stable et un système de reproduction autogame sont sélectionnés. Ce système de reproduction est le seul qui permet facilement et rapidement une reproduction à l'identique du meilleur phénotype. Dans l'article de Nowak *et al.* (2011), il a été montré que les allèles du gène S-RNase responsable du système d'incompatibilité chez les caféiers sont présents chez *C. arabica* mais ne sont pas fonctionnels. On peut donc imaginer une mise sous silence du gène entraînant une inactivation du système d'incompatibilité chez l'espèce allopolyploïde. *Coffea arabica* possède un comportement méiotique comparable aux espèces diploïdes. A la méiose, aucun des chromosomes homéologues de *C. arabica* ne s'apparie et cela est probablement dû au fonctionnement de facteurs génétiques régulant l'appariement (Lashermes *et al.*, 2000). L'environnement se stabilise vers des conditions défavorables pour les deux espèces parentales. On peut penser que, dans le cas de *C. arabica*, les espèces parentales ont reflué progressivement de la zone de sympatrie laissant *C. arabica* en allopatrie dans une niche écologique nouvelle pour l'espèce. Des études récentes montrent que *Coffea arabica* n'a pas encore subi de modifications génétiques majeures (article Cenci *et al.*, 2010 et 2011), et qu'aucune trace de sous-fonctionnalisation n'a été constatée (Combes *et al.*, 2011).

#### ***Perspectives associées à l'observation de la plasticité chez C. arabica***

- Une étude préliminaire sur le fonctionnement des sous-génomes chez *C. arabica* a été menée sur une région de 1200kb du génome, SH3 (Combes *et al.*, 2011). Cette étude a permis d'estimer l'expression relative de 13 paires de gènes homéologues de la même région génomique à partir de différents organes. Les auteurs ont montré que l'expression des gènes homéologues au sein de *C. arabica* est régulée par un mécanisme de trans-régulation partagé qui n'est pas spécifique du sous-génome mais que par contre les biais d'expression relative des gènes homéologues résulteraient des facteurs de régulation à l'échelle du gène. Cette régulation semble intervenir sur les capacités adaptatives de l'espèce allopolyploïde. Cette étude sera étendue à l'ensemble du génome par l'exploitation de données d'un projet de séquençage du transcriptome de *C. arabica* cultivé dans deux conditions de température différentes (30-26°C et 23-19°C). Pour cela, des données de séquençage du transcriptome de *C. arabica* associées aux données sur les

transcriptomes des espèces diploïdes obtenues par la Puce à ADN, nous permettrons d'observer l'impact de la fusion des sous-génomes sur leur expression dans *C. arabica* et d'identifier les modes de régulation intervenant sur les sous-génomes. Afin de traiter les données de séquençage par RNAseq, un outil de bioinformatique a été développé (SNIploid). Il permettra, à partir de données SNP, d'estimer le ratio de l'expression relative entre les sous-génomes.

- Nous nous sommes intéressés à l'influence des conditions de température sur la croissance de *C. arabica* et ses deux espèces parentales, cependant on peut se demander si les conditions environnementales qui prévalaient lors de la formation de l'allopolyploïde n'auraient pas orientées vers des phénotypes particuliers. Pour valider cette hypothèse, il serait intéressant de travailler sur du matériel néosynthétique pour observer comment l'environnement influe pendant les étapes fondamentales de la spéciation. Nous suggérons d'étudier l'influence de l'environnement sur l'hybridation et plus précisément sur le choc génomique et sur les processus liés à la méiose. L'hybridation entre les deux espèces parentales s'effectuerait dans deux conditions environnementales contrastées (par ex une population à hautes altitudes dans des températures fraîches et une population à basses altitudes dans des températures plus chaudes) permettant de constater si le choc génomique est de même nature en fonction de l'environnement et l'influence sur la première méiose. A partir de ces deux populations, deux sous populations seraient créées dont chacune sera exposée à une des deux conditions environnementales. Ceci nous permettrait d'étudier l'influence de l'environnement sur la méiose, sur l'émergence d'un système de reproduction autogame et sur le phénotype. Cependant, il est sans doute difficile de mettre cette expérience en œuvre sur les caféiers car l'obtention de générations est longue. On pourrait utiliser une autre espèce à générations moins longue, par exemple en utilisant des individus polyploïdes appartenant également aux Rubiacées.

# Bibliographie

## A

**Abbott RJ, Lowe AJ. 2004.** Origins, establishment and evolution of newpolyploid species: *Senecio cambrensis* and *Senecio eboracensis* in the British Isles. In: Leitch AR, Soltis DE, Soltis PS, Leitch IJ, Pires JC, eds. Biological relevance of polyploidy: ecology to genomics. *Biological Journal of the Linnean Society* **82**: 467–474.

**Adams KL, Cronn R, Percifield R, Wendel JF. 2003.** Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences of the United States of America* **100(8)**: 4649-4654.

**Adams KL, Percifield R, Wendel JF. 2004.** Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* **168(4)**: 2217-2226.

**Adams KL, Wendel JF. 2004.** Exploring the genomic mysteries of polyploidy in cotton. *Biological Journal of the Linnean Society* **82(4)**: 573-581.

**Adams KL, Wendel JF. 2005.** Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8(2)**: 135-141.

**Adams KL. 2007.** Evolution of duplicate gene expression in polyploid and hybrid plants. *Journal of Heredity* **98**: 136-141.

**Aharoni A, Vorst O. 2002.** DNA microarrays for functional plant genomics. *Plant Molecular Biology* **48**:99–118.

**Ainouche M, Baumel A, Salmon A, Yannic G. 2003.** Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytologist* **161**: 165–172.

**Ainouche ML, Baumel A, Salmon A, Yannic G. 2004a.** Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytologist* **161(1)**: 165-172.

**Ainouche ML, Baumel A, Salmon A. 2004b.** *Spartina anglica* C. E. Hubbard: a natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biological Journal of the Linnean Society* **82(4)**: 475-484.

**Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien MA, Fukunaga K, Ricou M, Misset MT. 2009.** Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biological invasions* **11**: 1159–1173.

**Albertin W, Brabant P, Catrice O, Eber F, Jenczewski E, Chevre AM, Thiellement H. 2005.** Autopolyploidy in cabbage (*Brassica oleracea* L.) does not alter significantly the proteomes of green tissues. *Proteomics* **5(8)**: 2131-2139.

**Albertin W, Balliau T, Brabant P, Chevre AM, Eber F, Malosse C, Thiellement H. 2006.** Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics* **173(2)**: 1101-1113.

**Albertin W, Alix K, Balliau T, Brabant P, Davanture M, Malosse C, Valot B, Thiellement H. 2007.** Differential regulation of gene products in newly synthesized *Brassica napus* allotetraploids is not related to protein function nor subcellular localization. *Bmc Genomics* **8**.

**Albertin W, Marullo P, Aigle M, Bourgeois A, Bely M, Dillmann C, De Vienne D, Sicard D. 2009.** Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species *J. evol. biol.* **22**: 2157–2170

**Al-Janabi SM, Honeycutt RJ, McClelland M, Sobral BWS. 1993.** A genetic linkage map of *Saccharum spontaneum* L. ‘SES 208.’ *Genetics* **134**:1249-1260

**Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search tool. *J Mol Biol.* **215**:403–410.

**Albertin W, Marullo P, Aigle M, Bourgais A, Bely M, Dillman C, Vienne DD et Sicard D. 2009.** Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. *Journal of Evolutionary Biology* **22**: 2157-2170.

**Anssour S, Krugel T, Sharbel TF, Saluz HP, Bonaventure G, Baldwin IT. 2009.** Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. *Annals of Botany* **103(8)**: 1207-1217.

**Anders S, Huber W. 2010.** Differential expression analysis for sequence count data. *Genome Biology*, **11**:R106

**Anai T, Koga M, Tanaka H, Kinoshita T, Rahman SM, Takagi Y. 2003.** Improvement of rice (*Oryza sativa* L.) seed oil quality through introduction of a soybean microsomal omega-3 fatty acid desaturase gene. *Plant Cell Reports*, **21**:10

**Anthony F, Berthaud J, Guillaumet JL, Lourd M. 1987.** Collecting wild *Coffea* species in Kenya and Tanzania. *Plant Genet Resour News* **1** **69**:23–29

**Anthony F, Combes MC, Astorga C, Bertrand B, Graziosi G, Lashermes P. 2002.** The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor Appl Genet* **104**: 894–900

**Anthony F, Lashermes P. 2006.** Origin, evolution and diversity of coffee (*Coffea arabica* L) genome. In: plant Genome, Biodiversity and Evolution. Ed. By K.A Sharma and A. Sharma. Science Publishers Inc. Plymouth, UK. pp. 208-228

**Anthony F, Dussert S, Dulloo E. 2007.** The coffee genetic resources. In: Engelmann F, Dulloo E, Astorga C, Dussert S, Anthony F (eds) Complementary strategies for ex situ conservation of

*Coffea arabica* genetic resources. A case study in CATIE, Costa Rica. Topical Reviews in Agricultural Biodiversity. Bioersity International, Rome, pp 12–22

**Anthony F, Diniz LEC, Combes MC, Lashermes P. 2010.** Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. *Plant Syst Evol* **285**:51–64

**Aronel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR. 1992.** Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. *Science* **258**:1353–55

## B

**Bauer H, Wierer R, Hatheway WH, Larcher W. 1985.** Photosynthesis of *Coffea arabica* after chilling. *Physiol Plant*, **64**:449–54.

**Barros RS, Maestri M, Rena AB. 1999.** Physiology of growth and production of the coffee tree - a review. *J. Coffee Res.* **27**:1-54.

**Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG 2009.** Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. *BMC Molecular Biology* 2009, **10**:1

**Beaudoin F, Napier JA. 2004.** In: *Lipid Metabolism and Membrane Biogenesis (Topics in Current Genetics, Vol. 6)* Daum G, editor. Berlin: Springer-Verlag; Biosynthesis and compartmentation of triacylglycerol in higher plants; pp. 267–287.

**Benjamini Y, Hochberg Y. 1995.** Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**: 289–300.

**Bertrand B., Aguilar, G., Santacreo, R. & Anzueto, F. 1999.** El mejoramiento genético en América Central. pp. 407–456, in: B. Bertrand and B. Rapidel (editors). Desafíos de la Caficultura Centroamericana. IICA/PROMECAFE-CIRAD-IRD-CCCR France, San José.

**Bertrand B, Guyot B, Anthony F, Lashermes P. 2003.** Impact of *Coffea canephora* gene introgression on beverage quality of *C. arabica*. *Theor Appl Genet* **107**:387–394

**Bettencourt A. 1973.** Considerações gerais sobre o ‘Híbrido de Timor’. Circular no 31. Instituto Agronômico de Campinas

**Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005.** Dosage balance in gene regulation: biological implications. *Trends in Genetics* **21**: 219–226.

**Birchler JA, Veitia RA. 2007.** The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* **19**: 395–402.

**Birchler JA, Veitia RA. 2010.** The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytologist*, **186**:54–62.

**Blakeslee AF, Avery AG. 1937.** Methods of inducing doubling of chromosomes in plant: By Treatment With Colchicine *J Hered*, **28 (12)**: 393-411.

**Bouharmont J. 1963.** Somatic chromosomes of some *Coffea* species *Euphytica*, **12(3)**: 254-257

**Bremer B, Jansen RK. 1991.** Comparative restriction site mapping of chloroplast DNA implies new phylogenetic relationships within *Rubiaceae*. *Am J Bot* **78**:198–213

**Bremer B, Eriksson T. 2009.** Time tree of Rubiaceae phylogeny and dating the family, subfamilies and tribes. *Int. J. Plant Sci.* **170(6)**:766–793.

**Bretagnolle F, Thompson JD. 1995.** Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of polyploid plants. *New Phytologist* **129**: 1–22.



**Bretagnolle F, Felber F, Calame FG, Küpfer P. 1998.** La polyplœdie chez les plantes. *Bot Helv* **108**:5–37

**Bridson D. 1982.** Studies in *Coffea* and *Psilanthus* (*Rubiaceae* subfam. *Cinchonoideae*) for Part 2 of ‘Flora of Tropical East Africa’: *Rubiaceae*. *Kew Bull* **36**:817–859

**Bridson D, Verdcourt B. 1988.** *Coffea*. In: Polhill RM (ed) Flora of Tropical East Africa. Rubiaceae (Part 2). Balkema, Rotterdam, The Netherlands, pp 703–727

**Brochmann C, Brysting AK, Alsos IG, Borgen L, Liu JX, Grundt HH, Scheen A, Elven R. 2004.** Polyploidy in arctic plants. *Biological Journal of the Linnean Society*. **82**:521 -536.

**Butruille DV, Boiteux LS. 2000.** Selection-mutation balance in polysomic tetraploids: impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. *Proc. Natl. Acad. Sci. USA* **97**: 6608–6613.

**Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH. 2000.** The likelihood of homoploid hybrid speciation. *Heredity* **84**(4): 441-451.

**Buggs RJA. 2008.** Towards natural polyploid model organisms. *Molecular Ecology* **17**:1875–1876.

**Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE. 2009.** Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* **103**(1): 73-81.

**Buggs RJA, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, Soltis PS. 2010.** Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytologist*, **186**:175–183.

## C

**Campos PS, Quartin V, Ramalho JC, Nunes MA. 2003.** Electrolyte leakage and lipid degradation account for cold sensitivity in leaves of *Coffea* sp. plants. *J Plant Physiol.*; **160**:283–292.

**Cenci A, Combes MC, Lashermes P. 2010.** Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. *Molecular Genetics and Genomics*. **283(5)**:493-501.

**Cenci A, Combes MC, Lashermes P. 2011.** Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome *Molecular Genetics and Genomics* **283(5)**: 493-501

**Chagué V, Just J, Mestiri I, Balzergue S, Tanguy A.M, Huneau C, Huteau V, Belcram H, Coriton O, Jahier J, Chalhoub B. 2010.** Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytologist*, **187**:1181–1194.

**Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV. 2010.** Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners *Genome Biology*, **11**:R125

**Charrier A. 1978.** La structure génétique des caféiers spontanés de la Région Malgache (Mascarocoffea). Mémoires ORSTOM no. **87**. Paris: ORSTOM.

**Charrier A, Berthaud J. 1985.** Botanical classification of coffee. In: Clifford MN, Wilson KC, eds. Coffee: botany, biochemistry and production of beans and beverage. London, UK: Croom Helm, 13–47.

**Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF. 2009.** Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* **182**: 503–517.

**Chaudhary B, Hovav R, Flagel L, Mittler R, Wendel JF. 2009.** Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (*Gossypium*). *Bmc Genomics* **10**.

**Chelaifa H, Mahé F, Ainouche M. 2010.** Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora* (Poaceae). *Molecular Ecology*, **19**:2050–2063.

**Chelaifa H, Monnier A, Ainouche M. 2010.** Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* × *townsendii* and *Spartina anglica* (Poaceae). *New Phytologist*, **186**:161–174.

**Chen ZJ, Pikaard CS. 1997.** Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes & Development* **11(16)**: 2124-2136.

**Chen ZJ, Ni Z. 2006.** Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays* **28**: 240–252.

**Chen ZJ. 2007.** Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58**: 377-406.

**Chen ZJ, Tian L. 2007.** Roles of dynamic and reversible histone acetylation in plant development and polyploidy. *Biochimica et Biophysica Acta* **1769**: 295–307

**Chen ZJ. 2010.** Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science* **15 (2)**: 1360-1385

**Cheong JJ, Choi YD. 2003.** Methyl jasmonate as a vital substance in plants. *Trends Genet.*, **19**:409–413.

**Chevalier A, Dagon M. 1928.** Recherches historiques sur les débuts de la culture du caféier en Amérique. *Communications et Actes de l'Académie des Sciences Coloniales* (Paris) **5**:1–38

**Church SA, Spaulding EJ. 2009.** Gene Expression in a Wild Autopolyploid Sunflower Series. *Journal of Heredity* **100**(4): 491-495.

**Clarke JD, Zhu T. 2006.** Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives. *Plant J.* **45**:630–650.

**Comai L. 2000.** Genetic and epigenetic interactions in allopolyploid plants. *Plant Molecular Biology* **43**(2-3): 387-399.

**Combes MC, Cenci AI, Baraille H, Bertrand B, Lashermes P. 2011.** Homeologous Gene Expression in Response to Growing Temperature in a Recent Allopolyploid (*Coffea arabica* L.). *Journal of Heredity*; doi: 10.1093/jhered/esr120

**Coulibaly I, Revol B, Noirot M, Poncet V, Lorieux M, et al. 2003.** AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx* x *C. canephora*) x *C. canephora*]. *Theor Appl Genet* **107**:1148–1155

**Cros, J. 1994.** Implications phylogénétiques des variations de l'ADN chloroplastique chez les caféiers (genres *Coffea* L. et *Psilanthus* Hook. f.). Ph.D. dissertation, Université Montpellier II, Montpellier, France

**Cros J., Combes MC, Chabrilange N, Hamon S, Duperray C, Monnot des Angles A. 1995.** Nuclear DNA content in the subgenus *Coffea* (Rubiaceae): inter- and intra-specific variation in African species. *Revue canadienne de botanique*, **73**(1): 14-20

**Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A, Lashermes P. 1998.** Phylogenetic relationships of *Coffea* species: new evidence based on the chloroplast DNA variation analysis. *Mol Phylogenet Evol*, **9**:109–117

**Cui LY, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW. 2006.** Widespread genome duplications throughout the history of flowering plants. *Genome Research* **16(6)**: 738-749.

**Cruz F, Kalaoun S, Nobile P, Colombo C, Almeida J, Barros LMG, Romano E, Grossi-de-Sa MF, Vaslin M, Alves-Ferreira M. 2009.** Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR *Mol Breeding*, **23**:607–616

## D

**DaMatta FM, Maestri M, Mosquim PR, Barros RS. 1997.** Photosynthesis in coffee (*Coffea arabica* and *C. canephora*) as affected by winter and summer conditions. *Plant Sci.* **128**:43-50.

**DaMatta FM, Cochicho Ramalho JD. 2006.** Impacts of drought and temperature stress on coffee physiology and production: a review. *Brazilian Journal Plant Physiology* **18**: 55–81.

**DaMatta FM, Ronchi CP, Maestri M, Barros RS. 2008.** Ecophysiology of coffee growth and production. *Braz. J. Plant Physiol.*, **19(4)**:485-510

**David JR, Gibert P, Moreteau B. 2004.** Evolution of reaction norms. pp. 50-63 in T. J. DeWitt and S. M. Scheiner (eds.). Phenotypic Plasticity: Functional and Conceptual Approaches. *Oxford University Press, Oxford.*

**Davis AP, Bridson D, Rakotonasolo F. 2005.** A reexamination of *Coffea* subgenus *Baracoffea* and comments on the morphology and classification of *Coffea* and *Psilanthus* (Rubiaceae-Coffeae). In: Keating RC, Hollowell VC, Croat T (eds) Festschrift for William G. D'Arcy: the legacy of a taxonomist (Monograph in Syst Bot 104). MBG Press, St. Louis, pp 398–420

**Davis AP, Govaerts R, Bridson DM, Stoffelen P. 2006.** An annotated taxonomic conspectus of genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* **152**: 465–512.

**Davis AP, Chester M, Maurin O, Fay MF. 2007.** Searching for the relatives of *Coffea* (Rubiaceae, Ixoroideae): the circumscription and phylogeny of Coffeae based on plastid sequence data and morphology. *Am J Bot* **94**:313–329

**Davis AP, Rakotonasolo F. 2008.** A taxonomic revision of the baracoffea alliance: nine remarkable *Coffea* species from western Madagascar. *Botanical Journal of the Linnean Society* **158**: 355–390.

**Degenkolbe T, Hannah MA, Freund S, Hinch DK, Heyer AG, Kohl KI. 2005.** A quality-controlled microarray method for gene expression profiling. *Anal Biochem.*; **346**:217–224.

**Dong S, Adams KL. 2011.** Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids, *New Phytologist* **190**:1045–1057

**Dormann P, Voelker TA, Ohlrogge JB. 2000.** Accumulation of palmitate in Arabidopsis mediated by the acyl-acyl carrier protein thioesterase FATB1. *Plant Physiol.*; **123**:637–644. doi: 10.1104/pp.123.2.637.

**Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008.** Evolutionary Genetics of Genome Merger and Doubling in Plants. *Annual Review of Genetics* **42**: 443-461.

**Dyer JM., Mullen RT. 2001.** Immunocytological localization of two plant fatty acid desaturases in the endoplasmic reticulum. *FEBS Letters* **494**: 44-47

## E

**K Earley, RJ Lawrence, O Pontes, R Reuther, AJ Enciso, M Silva, N Neves, M Gross, W Viegas, CS. Pikaard. 2006.** Erasure of histone acetylation by *Arabidopsis HDA6* mediates large-scale gene silencing in nucleolar dominance. *Genes et Developement* **20**:1283–1293

**Ewing B, Hillier L, Wendl MC, Green P. 1998.** Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**: 175–185.

## F

**Fawcett JA, Maere S, Van de peer YV. 2009.** Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of National Academy of Sciences*. **106**: 5737- 5742.

**Flagel L, Udall J, Nettleton D, Wendel J. 2008.** Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *Bmc Biology* **6**.

**Flagel L, Wendel J 2010.** Evolutionary rate variation, genomic dominance, and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytologist*, **186**:184–193.

**Feldman M, Liu B, Segal G, Abbo S, Levy AA, et al. 1997.** Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics*, **147**:1381–1387.

**Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999.** Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151**: 1531–1545

**Fortunato AS, Lidon FC, Batista-Santos P, Leitao AE, Pais IP, Ribeiro AI, Ramalho JC. 2009.** Biochemical and molecular characterization of the antioxidative system of *Coffea sp.* Under cold conditions in genotypes with contrasting tolerance *Journal of Plant Physiology* **167**: 333–342

**Fox CW, Czesak ME. 2000.** Evolutionary ecology of progeny size in arthropods. *Annu. Rev. Entomol.* **45**: 341–369.

## G

**Gaeta RT, Yoo S-Y, Pires JC, Doerge RW, Chen ZJ, Osborn TC. 2009.** Analysis of gene expression in resynthesized *Brassica napus* allopolyploids using *Arabidopsis* 70 mer Oligo Microarrays. *PLoS ONE* **4**: e4760.

**Galbraith DW. 2006.** DNA microarray analyses in higher plants. *OMICS.* **10**:455–473.

**Gallardo MH, Bickham JW, Honeycutt RL, Ojeda R, Köhler N. 1999.** Discovery of tetraploidy in a mammal. *Nature* **401**: 341.

**Gallardo MH, Kirsch JAW. 2001.** Molecular relationships among, octodontidae (Mammalia: Rodentia: Caviomorpha). *J. Mamm. Evol.* **8**: 73–89.

**Gallardo MH, Kausel G, Jiménez A, Bacquet C, González C, Figueroa J, Köhler N, Ojeda R. 2004.** Whole-genome duplications in South American desert rodents (Octodontidae). *Biol. J. Linn. Soc.* **82**: 443–451

**Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al. 2004.** Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80.

**Ghalambor CK, McKay JK, Carroll SP, Reznick DN. 2007.** Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct. Ecol.* **21**: 394–407



**Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. 2008.** High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.*; **36**:3420–3435.

**Grant V. 1971.** Plant Speciation. *New York: Columbia University Press.*

**Guerrero Filho O, Silvarolla MB, Eskes AB. 1999.** Expression and mode of inheritance of resistance in coffee to leaf miner *Perileucoptera coffeella*. *Euphytica* **105**:7–15

**Guo M, Davis D, Birchler JA. 1996.** Dosage Effects on Gene Expression in a Maize Ploidy Series. *Genetics* **142**: 1349-1355.

## H

**Ha M, Li WH, Chen ZJ. 2007.** External factors accelerate expression divergence between duplicate genes. *Trends in Genetics* **23(4)**: 162-166.

**Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ. 2009.** Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proceedings of the National Academy Sciences, USA* **106**: 17835–17840.

**Hahn MW. 2009.** Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. *Journal of Heredity*, **100(5)**:605–617.

**Hamada T, Nishiuchi T, Kodama H, Nishimura M, Iba K. 1996.** cDNA cloning of a wounding-inducible gene encoding a plastid  $\Delta^3$  fatty acid desaturase from tobacco. *Plant Cell Physiol.* **37**:606–11

**Hegarty MJ, Jones JM, Wilson ID, Barker GL, Coghill JA, Sanchez-Baracaldo P, Liu GQ, Buggs RJA, Abbott RJ, Edwards KJ, Hiscock SJ. 2005.** Development of anonymous cDNA

microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. *Molecular Ecology* **14(8)**: 2493-2510.

**Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. 2006.** Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Current Biology* **16**: 1652-1659.

**Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ. 2008.** Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. *Philosophical Transactions of the Royal Society B* **363**: 3055–3069.

**Hegarty MJ, Hiscock SJ. 2009.** The complex nature of allopolyploid plant genomes. *Heredity* **103**: 100–101.

**Heppard EP, Kinney AJ, Stecca K, Guo-Hua M. 1996.** Developmental and Growth Temperature Regulation of Two Different Microsomal  $\omega$ -6 Desaturase Genes in Soybeans. *Plant Physiol.* **110**: 311-319

**Horiguchi G, Iwakawa H, Kodama H, Kawakami N, Nishimura M, Iba K. 1996.** Expression of a gene for plastid  $\omega$ -3 fatty acid desaturase and changes in lipid and fatty acid compositions in light- and dark- grown wheat leaves. *Physiologia Plantarum* **96(2)**:275–283

**Hu G, Houston NL, Pathak D, Schmidt L, Thelen JJ, Wendel JF. 2011.** Genomically biased accumulation of seed storage proteins in allopolyploid cotton 10.1534/genetics.111.132407

**Huang X, Madan A. 1999.** CAP3: A DNA sequence assembly program. *Genome Res.* **9**:868–877.

**Hovav R, Udall JA, Chaudhary B, Hovav E, Flagel L, Hu GJ, Wendel JF. 2008.** The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. *Plos Genetics* **4**(2).

**Hundertmark M, Hinch DK. 2008.** LEA (Late Embryogenesis Abundant) Proteins and their Encoding Genes in *Arabidopsis Thaliana*. *BMC Genomics*. **9**:118.

**Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. 2009.** InterPro: the integrative protein signature database. *Nucleic Acids Res*. **37**:D211–D215.

## I

**K Iba.2002.** Acclimative response to temperature stress in higher plants: Approaches of Gene Engineering for Temperature Tolerance. *Annu. Rev. Plant Biol*, **53**:225–45

**Iba K, Gibson S, Nishiuchi T, Fuse T, Nishimura M, et al. 1993.** A gene encoding a chloroplast  $\Delta^3$  fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the *fad7* mutant of *Arabidopsis thaliana*. *J. Biol. Chem*. **268**:24099–105

**Iseli C, Jongeneel CV, Bucher P. 1999.** ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol*. pp. 138–148.

## J

**Jackson S, Chen ZJ. 2009.** Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* **13**: 1–7.

**Jakobsson M, Hagenblad J, Tavare´ S, Sa´ll T, Hallde´n C, Lind-Hallde´n C, Nordborg M. 2006.** A Unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Molecular Biology and Evolution* **23**: 1217–1231.

**Jenczewski E, Eber F, Grimaud A, Huet S, Lucas MO, et al. 2003.** PrBn, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. *Genetics*, **164**:645–653.

**Joët T, Laffargue A, Salmona J, Doulebeau S, Descroix F, Bertrand B, De Kochko A, Dussert S. 2009.** Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. *New Phytologist* **182**: 146–162.

## K

**Kargiotidou A, Deli D, Galanopoulou D, Tsaftaris A, Farmaki T. 2008.** Low temperature and light regulate delta 12 fatty acid desaturases (FAD2) at a transcriptional level in cotton (*Gossypium hirsutum*). *Journal of Experimental Botany*, **59(8)**: 2043–2056

**Kihara H, Ono T. 1926.** Chromosomenzahlen und systematische gruppierung der Rumex arten. *Zeitschr Zellf Mikrosk Anat*, **4**:475–481.

**Kim S, Yoo MJ, Albert VA, Farris JS, Soltis PS, Soltis DE. 2004.** Phylogeny and diversification of B-function MADS-box genes in angiosperms: Evolutionary and functional implications of a 260-million-year-old duplication. *Am. J. Bot.* **91**: 2102–2118.

**Kodama H, Akagi H, Kusumi K, Fujimura T, Iba K. 1997.** Structure, chromosomal location and expression of a rice gene encoding the microsomal Δ<sup>3</sup> fatty acid desaturase *Plant Molecular Biology* **33**: 493–502,

**Krug CA, Mendes JET, Carvalho A. 1939.** Taxonomia de *Coffea arabica* L. Technical bulletin no 62. Instituto Agronômico do Estado, Campinas

## L

**Laffargue A, de Kochko A, Dussert S. 2007.** Development of solid-phase extraction and methylation procedures to analyse free fatty acids in lipid-rich seeds, *Plant Physiology and Biochemistry* **45 (3-4):** 250-257

**Lashermes P, Cros J, Marmey P, Charrier A. 1993.** Use of random amplified DNA markers to analyze genetic variability and relationships of *Coffea* species. *Genet Resour Crop Evol*, **40:91–99**

**Lashermes P, Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A. 1996a.** Inheritance and restriction fragment length polymorphism of chloroplastDNA in the genus *Coffea* L. *Theor Appl Genet*, **93:626–632**

**Lashermes P, Trouslot P, Anthony F, Combes MC, Charrier A. 1996b.** Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica* **87:59–64**

**Lashermes P, Combes MC, Trouslot P, Charrier A. 1997.** Phylogenetic relationships of coffee tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theor Appl Genet*, **94:947–955**

**Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A. 1999.** Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol Gen Genet*, **261:259–266**

**Lashermes P, Andrzejewski S, Bertrand B, Combes MC, Dussert S, Grasiozi G, Trouslot P, Anthony F. 2000.** Molecular analysis of introgressive breeding in coffee (*Coffea arabica*). *Theoretical and Applied Genetics* **100**: 139–146.

**Lashermes P, Carvalho Andrade A, Etienne H. 2008.** Genomics of coffee, one of the world's largest traded commodities. In: Moore PH, Ming R, eds. Genomics of tropical crop plants. New York, NY, USA: Springer, 203–224.

**Lashermes P, Combes MC, Ribas A, Cenci A, Mahe' L, Etienne H. 2010.** Genetic and physical mapping of the SH3 region that confers resistance to leaf rust in coffee tree (*Coffea arabica* L.). *Tree Genetics and Genomes* **6**: 973–980.

**Lee HS, Chen ZJ. 2001.** Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc Natl Acad Sci USA*, **98**:6753–58.

**Leitch AR, Leitch IJ. 2008.** Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481–483.

**Lemieux B, Miquel M, Somerville C, Browse J. 1990.** Mutants of *Arabidopsis* with alterations in seed lipid fatty acid composition. *Theor. Appl. Genet.* **80**:234–40

**Leroy J-F (1980)** Evolution et taxogenèse chez les caféiers: hypothèse sur l'origine. *C R Acad Sci Paris* **291**:593–596

**Lin C, Mueller LA, Mc CJ, Crouzillat D, Petiard V, Tanksley SD. 2005.** Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theor Appl Genet.* **112**:114–130.

**Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF. 2001.** Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* **44(3)**: 321-330.

**Liu XY, Yang JH, Li B, Yang XM, Meng QW. 2006.** Antisense-Mediated Depletion of Tomato Chloroplast Omega-3 Fatty Acid Desaturase Enhances Thermal Tolerance *Journal of Integrative Plant Biology*, **48 (9)**: 1096–1107

**Liu Z, Adams KL. 2007.** Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Current Biology* **17**: 1669–1674.

**Lu B, Pan X, Zhang L, Huang B, Sun L, Li B, Yi B, Zheng S, Yu X, Ding R, Chen W. 2006.** A Genome-wide Comparison of Genes Responsive to Autopolyploidy in *Isatis indigotica* Using *Arabidopsis thaliana* Affymetrix Genechips, *Plant Molecular Biology Reporter* **24**: 197–204.

**Lukens L, Zou F, Lydiate D, Parkin I, Osborn T. 2003.** Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* **164(1)**: 359-372.4

**Luchetta P, Higuete D, Maurel MC, Vervoort M. 2005.** Évolution moléculaire : cours & questions de revision, 10-2005 © 1995-2011 LAVOISIER S.A.S.

## M

**Mable BK. 2004.** Evolutionary genetics of self-incompatibility in a new ‘model’ plant: *Arabidopsis lyrata*. In: QCCronk, JWhitton, IEPTaylor, eds. Plant adaptation: molecular biology and ecology. Ottawa: NRC Research Press.

**Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002.** Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology* **129(2)**: 733-746.

**Mahé L. 2007.** Contribution à l’amélioration génétique de la résistance des caféiers (*Coffea arabica* L.) à la rouille (*Hemileia vastatrix*) : De l’étude des hybrides interspécifiques naturels de Nouvelle-Calédonie à la cartographie d’un locus de résistance. Thèse, *centre international d’études supérieures en sciences agronomiques, Montpellier SUP AGRO.*, pp. 1-134

**Marioni JC, Mason CE, Mane SM, Stephens, Gilad Y. 2008.** RNA-seq : an assessment of technical reproducibility and comparison with gene expression arrays' *Genome Res*, **18**: 1509-1517

**Marraccini P, Deshayes A, Pétiard V, Rogers WJ. 1999.** Molecular cloning of the complete 11 S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants. *Plant Physiology and Biochemistry*, **37**:273–282.

**Maricle BR, Crosier JJ, Bussiere BC, Lee RW. 2006.** Respiratory enzyme activities correlate with anoxia tolerance in saltmarsh grasses. *J. Exp. Mar. Biol. Ecol.* **337**: 30–37

**Martelotto LG, Ortiz JPA, Stein J, Espinoza F, Quarín CL, Pessino SC. 2007.** Genome rearrangements derived from autopolyploidization in *Paspalum* sp. *Plant Science* **172**: 970–977.

**Martens C, De Peer YV. 2010.** The hidden duplication past of the plant pathogen phytophthora and its consequences for infection. *BMC Genomics* **11**: 353.

**Martinez-Godoy MA, Mauri N, Juarez J, Marques MC, Santiago J, Forment J, Gadea J. 2008.** A genome-wide 20 K citrus microarray for gene expression analysis. *BMC Genomics*, **9**:318.

**Mascarell-Creus A, Canizares J, Vilarrasa-Blasi J, Mora-Garcia S, Blanca J, Gonzalez-Ibeas D, Saladie M, Roig C, Deleu W, Pico-Silvent B, Lopez-Bigas N, Aranda MA, Garcia-Mas J, Nuez F, Puigdomenech P, Cano-Delgado AI. 2009.** An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon (*Cucumis melo* L.) *BMC Genomics*.**10**:467.

**Maurin O, Davis AP, Chester M, Mvungi EF, Jaufeerally-Fakim Y, Fay MF. 2007.** Towards a phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. *Ann Bot Lond* **100**:1565–1583



**Mavárez J, Salazar CA, Bermingham ES, Jiggins CD, Linares M. 2006.** Speciation by Hybridization in Heliconius Butterflies. *Nature*. **441**: 868-871.

**McClintock B. 1984.** The significance of responses of the genome to challenge. *Science* **226**: 792– 801.

**McConn M, Hugly S, Browse J, Somerville C. 1994.** A mutation at the *fad8* locus of *Arabidopsis* identifies a second chloroplast  $\Delta^3$  desaturase. *Plant Physiol*. **106**:1609–14

**McConn M, Browse J. 1996.** The Critical Requirement for Linolenic Acid Is Pollen Development, Not Photosynthesis, in an Arabidopsis Mutant. *The plant cell* **8(3)**: 403-416

**Miquel M, James DJ, Dooner H, Browse J. 1993.** Arabidopsis requires polyunsaturated lipids for low-temperature survival. *Proc. Natl. Acad. Sci. USA* **90**: 6208-6212.

**Moore, G. 2002.** Meiosis in allopolyploids – the importance of ‘Teflon’ chromosomes. *Trends Genet*. **18**: 456–463.

**Morey JS, Ryan JC, Van Dolah FM. 2006.** Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biol Proced Online*. **8**:175–193.

**Muller HJ. 1925.** Why polyploidy is rarer in animals than in plants. *The american naturalist*, **59 (663)**: 346-353.

**Murakami Y, Tsuyama M, Kobayashi Y, Kodama H, Iba K. 2002.** Trienoic Fatty Acids and Plant Tolerance of High Temperature. *Science* **287**:476

## N

**Na S, Meng L, Shi-Jie Z, Feng L, Hui L, Qing-Wei M. 2007.** Overexpression of glycerol-3-phosphate acyltransferase gene improves chilling tolerance in tomato *Planta* **226**:1097–1108

**Nasrallah JB, Liu P, Sherman-Broyles S, Schmidt R, Nasrallah ME. 2007.** Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. *Genetics* **175(4)**: 1965-1973.

**Ni ZF, Kim ED, Ha MS, Lackey E, Liu JX, Zhang YR, Sun QX, Chen ZJ. 2009.** Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* **457(7227)**: 327 - U327.

**Nikolova-Damyanova B, Velikova R, Jham GN. 1998.** Lipid classes, Fatty Acid Composition and Triacylglycerol Molecular Species in Crude Coffee Beans Harvested in Brazil. *Food research international*. **31**:479–486.

**Nishiuchi T, Hamada T, Kodama H, Iba K. 1997.** Wounding changes the special expression pattern of the Arabidopsis plastid  $\Delta^3$  fatty acid desaturase gene (*FAD7*) through differential signal transduction pathways. *Plant Cell* **9**:1701–12

**Noir S, Patheyron S, Combes MC, Lashermes P, Chalhoub B. 2004.** Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). *Theor Appl Genet*. **109**:225–230

**Noirot M, Poncet V, Barre P, Hamon P, Hamon S, De Kochko A. 2003.** Genome Size Variations in Diploid African *Coffea* Species *Ann Bot* **92 (5)**: 709-714.

**Nowak MD, Davis AP, Anthony F, Yoder AD. 2011.** Expression and Trans-Specific Polymorphism of Self-Incompatibility RNases in *Coffea* (Rubiaceae). *PLoS ONE* **6(6)**: e21019.

**Nylin S, Gotthard K. 1998.** Plasticity in life-history traits. *Annu. Rev. Entomol.* **43**:63–93.

## O

**Ohlrogger J, Browser J. 1995.** Lipid Biosynthesis. *The Plant Cell*, **7**: 957-970

**Ohno S. 1970.** Evolution by gene duplication. *Springer- Verlag, New York*.

**Oliveros JC. 2007.** VENNY. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

**Orlova IV, Serebriiskaya TS, Popov V, Merkulova N, Nosov AM, Trunova TI, Tsydendambaev VD, Los DA. 2003.** Transformation of Tobacco with a Gene for the Thermophilic Acyl-Lipid Desaturase Enhances the Chilling Tolerance of Plants. *Plant Cell Physiol.* **44(4)**: 447–450

**Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V et al. 2003.** Understanding mechanisms of novel gene expression in polyploids. *Trends Genetics* **19**: 141–147.

**Otto SP, Whitton J. 2000.** Polyploid incidence and evolution. *Annual Review of Genetics* **34**: 401–437.

**Otto SP. 2007.** The evolutionay consequences of polyploidy. *Cell* **131**:452-462.

**Ozkan H, Levy AA, Feldman M. 2001.** Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13(8)**: 1735-1747.

## P

**Parisod C, Holderegger R, Brochmann C. 2010b.** Evolutionary consequences of autopolyploidy. *New Phytologist*, **186**:5–17.

**Pearse IS, Krügel T, Baldwin IT. 2006.** Innovation in anti-herbivore defense systems during neopolyploidy – the functional consequences of instantaneous speciation *The Plant Journal* **47**: 196–210

**Pendergrast M. Coffee second only to oil? Is coffee really the second largest commodity? 2009.** *Tea and Coffee Trade Journal*. **181**:38–41.

**Penfield S. 2008.** Temperature perception and signal transduction in plants *New Phytologist* **179**: 615–628

**Pfennig DW, Wund MA, Snell-Rood EC, Ruickshank TC, Schlichting CD, Moczek AP. 2010.** Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology and Evolution* **25**: 459–467

**Phillips JP, Tainer JA, Getzoff ED, Boulianne GL, Kirby K, Hilliker AJ. 1995.** Subunit-destabilizing mutations in *Drosophila* copper/zinc superoxide dismutase: neuropathology and a model of dimer dysequilibrium. *Proc Natl Acad Sci USA*; **92**:8574–78.

**Pires JC, Zhao JW, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC. 2004.** Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biological Journal of the Linnean Society* **82(4)**: 675–688.

**Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de KA, Hamon P. 2006.** SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol Genet Genomics*. **276**:436–449.

**Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ. 2008.** Adaptive evolution in invasive species *Trends in Plant Science*, **13 (6)**:288-294

**Privat I, Bardil A, Bombarely Gomez A, Severac D, Dantec C, Fuentes I, Mueller L, Joe't T, Pot D, Foucrier S et al. 2011.** The 'PUCE CAFE' project: the first 15K coffee microarray, a

new tool for discovering candidate genes correlated to agronomic and quality traits. *BioMed Central Genomics* **12**: 5.

**Pumphrey M, Bai M, Laudencia-Chingcuanco D, Anderson O, Gill BS. 2009.** Nonadditive expression of homeologous genes is established upon polyploidisation in hexaploid wheat. *Genetics* **181**: 1147–1157.

## R

**Raina SN, Mukai Y, Yamamoto M. 1998.** *In situ* hybridisation identifies the diploid progenitor of *Coffea arabica* (Rubiaceae). *Theor Appl Genet* **97**:1204–1209.

**Ramalho JC, Quartin V, Fahl JI, Carelli ML, Leitao AE, Nunes MA. 2003.** Cold acclimation ability of photosynthesis among species of the tropical *Coffea* genus. *PlantBiol* , **5**:631–41.

**Ramsey J, Schemske DW. 1998.** Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467-501.

**Rapp RA, Udall JA, Wendel JF. 2009.** Genomic expression dominance in allopolyploids. *BMC Biology* **7**: 18.

**Renaut S, Nolte AW, Bernatchez L. 2009.** Gene Expression Divergence and Hybrid Misexpression between Lake Whitefish Species Pairs (*Coregonus spp. Salmonidae*). *Molecular Biology and Evolution* **26(4)**: 925-936.

**Rieseberg LH, Widmer A, Arntz AM, Burke JN. 2003.** The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Phil. Trans. R. Soc. Lond. B*, **358**: 1141-1147.

**Rieseberg LH, Willis JH. 2007.** Plant speciation. *Science* **317(5840)**: 910-914.

**Riddle NC, Birchler JA. 2003.** Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. *Trends in Genetics* **19(11)**: 597-600.

**Riley R, Chapman V. 1958.** Genetic control of cytologically diploid behaviour of hexaploid wheat. *Nature*, **182**:713–715.

**Robbrecht E, Puff C. 1986.** A survey of the Gardenieae and related tribes (Rubiaceae). *Botanische Jahrbücher für Systematik* **108**: 63–137.

**Robbrecht E. 1988a.** Studies in tropical African Rubiaceae. 13. Petitiocodon, a new genus to accommodate *Didymosalpinx parviflora* (Gardenieae-Diplosporinae). *Bulletin du Jardin Botanique National de Belgique* **58**: 109–120.

**Robbrecht E. 1988b.** Tropical woody Rubiaceae. *Opera Botanica Belgica* **1**: 1–271.

**Robbrecht E. 1994.** Supplement to the 1988 outline of the classification of the Rubiaceae. *Opera Botanica Belgica* **6**: 173–196.

**Routaboul J-M, Fischer SF, Browse J. 2000.** Trienoic fatty acids are required to maintain chloroplast function at low temperatures. *Plant Physiol* **124**:1697–1705

## S

**Sakurai I, Hagio M, Gombos Z, Tyystjäarvi T, Paakkanen V, Aro EM, Wada H. 2003.** Requirement of Phosphatidylglycerol for Maintenance of Photosynthetic Machinery1. *Plant Physiology*, **133**: 1376–1384

**Salmona J, Dussert S, Descroix F, de Kochko A, Bertrand B, Joeıt T.2008.** Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. *Plant Molecular Biology* **66**: 105–124.

**Salmon A, Ainouche ML, Wendel JF. 2005.** Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* **14(4)**: 1163-1175.

**Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF. 2003.** Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution* **20**: 633–643.

**Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001.** Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell*, **13**:1749–1759.

**Shewry PR, Napier JA, Tatham AS. 1995.** Seed storage proteins: structures and biosynthesis. *Plant Cell.*, **7**:945–956.

**Smith RF. 1985.** A history of coffee. In: Clifford MN, Willson KC (eds) Coffee botany, biochemistry and production of beans and beverage. *Croom Helm, London Sydney*, pp 1–12

**Simkin AJ, Qian T, Caillet V, Michoux F, Ben AM, Lin C, Tanksley S, McCarthy J. 2006.** Oleosin gene family of *Coffea canephora*: quantitative expression analysis of five oleosin genes in developing and germinating coffee grain. *J Plant Physiol.*, **163**:691–708.

**Smyth GK. 2005.** Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, eds. Bioinformatics and computational biology solutions using R and Bioconductor. New York, NY, USA: Springer, 397–420.

**Smyth GK, Speed TP. 2003.** Normalization of cDNA microarray data. *Methods* **31**: 265–273.

**Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E. 2004.** Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. In: Leitch AR, Soltis DE, Soltis PS, Leitch IJ, Pires JC, eds. Biological relevance of polyploidy: ecology to genomics. *Biological Journal of Linnean Society* **82**: 485–501.

**Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E. 2004.** Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society* **82(4)**: 485-501.

**Soltis PS, Soltis DE. 2009.** The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology* **60**: 561-588.

**Song K, Lu P, Tang K, Osborn TC. 1995.** Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc Natl Acad Sci USA*, **92**:7719–23.

**Srinivasan KH, Narasimhaswamy RL. 1975.** A review of coffee breeding work done at the Government coffee experiment station, Balehonnur. *Indian Coffee* **34**:311–321

**Stamati K, Mackay I, Powell W. 2009.** A quantitative genomic imbalance gene expression assay in a hexaploid species: wheat (*Triticum aestivum*). *Genome* **52**: 89–94

**Stearns SC. 1992.** The Evolution of Life Histories. *Oxford Univ. Press, Oxford*.

**Stebbins GL. 1971.** Chromosomal evolution in higher plants. London, UK: Edward Arnold.

**Stoffelen P. 1998.** *Coffea* and *Psilanthus* (Rubiaceae) in tropical Africa: a systematic and palynological study, including a revision of the West and Central African species. Doctoral thesis, Katholieke Universiteit Leuven, Faculteit Wetenschappen, Belgium

**Stoffelen P, Noirot M, Couturon E, Bontems S, De Block P, Anthony F. 2009.** *Coffea anthonyi* Stoff. & F. Anthony, a new selfcompatible Central African coffee species, closely related to an ancestor of *C. arabica* L. *Taxon* **58**:133–140

**Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, Veilleux RE, Busse JS, Erhardt RJ, Buell CR, Jiang JM. 2007.** Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics* **176(4)**: 2055-2067.



**Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM. 2008.** Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis *BMC Plant Biology*, **8**:33

**Somerville C, Browse J. 1991.** Plant lipids: metabolism, mutants, and membranes. *Science* **252**:80–87

**Song KM, Lu P, Tang KL, Osborn TC. 1995** Rapid genome changes in synthetic polyploides of Brassica and its implications for polyploidy evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92(17)**: 7719-7723.

## T

**Takumi S, Nishioka E, Morihito H, Kawahara T, Matsuoka Y. 2009.** Natural variation of morphological traits in wild wheat progenitor *Aegilops tauschii* Coss. *Breeding Science* **59**: 579–588

**Tate JA, Ni ZF, Scheen AC, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE. 2006.** Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173(3)**: 1599-1611.

**Teixeira MC, Coelho N, Olsson ME, Brodelius PE, Carvalho IS, Brodelius M. 2009.** Molecular cloning and expression analysis of three omega-6 desaturase genes from purslane (*Portulaca oleracea* L.). *Biotechnol Lett* **31**:1089–1101

**Thomas AS. 1942.** The wild arabica coffee on the Boma Plateau, Anglo-Egyptian Sudan. *Emp Exp Agric* **10**:207–212

**Troein C, Vallon-Christersson J, Saal LH. 2006.** An introduction to BioArray Software Environment. *Methods Enzymol.*, **411**: 99–119.

**Tusher V, Tibshirani R, Chu C. 2001.** Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy Sciences, USA* **98**: 5116–5121.

## U

**Udall JA, Quijada PA, Osborn TC. 2005.** Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. *Genetics* **169(2)**: 967-979.

**Udall JA, Swanson JM, Nettleton D, Percifield RJ, Wendel JF. 2006.** A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* **173(3)**: 1823- 1827.

**Upchurch RG. 2008.** Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress *Biotechnol Lett.* **30**:967–977

## V

**Van de Peer Y, Maere S, Meyer A. 2009.** Opinion the evolutionary significance of ancient genome duplications. *Nature Reviews Genetics* **10**: 725- 732.

**Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002.** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**:RESEARCH0034.

**Veitia RA, Bottani S, Birchler JA. 2008.** Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* **24(8)**: 390-397.

**Vidal RO, Costa Mondego JM, Pot D, Ambrosio AB, Carvalho Andrade A, Protasio Pereira LF, Colombo CA, Gonzaga Esteves Vieira L, Falsarella Carazzolle M, Amarante Guimaraes Pereira G. 2010.** A high-throughput data mining of single nucleotide

polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiology* **154**: 1053–1066.

## W

**Waddington CH. 1940.** Organisers and Genes. *Cambridge University Press, Cambridge*.

**Waddington CH. 1942.** Canalisation of development and the inheritance of acquired characters. *Nature* **150**: 563–565.

**Wakita Y, Otani M, Hamada T, Mori M, Iba K, Shimada T. 2001.** A tobacco microsomal  $\Delta^3$  fatty acid desaturase gene increases the linolenic acid content in transgenic sweet potato (*Ipomoea batatas*). *Plant Cell Reports* **20(3)**: 244-249.

**Wallis JG, Browse J. 2010.** Lipid biochemists salute the genome. *The Plant Journal* **61**: 1092–1106.

**Wang JL, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ. 2004.** Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* **167(4)**: 1961-1973.

**Wang JL, Tian L, Lee HS, Wei NE, Jiang HM, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ. 2006.** Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* **172(1)**: 507-517.

**Wang Z, Gerstein M, Snyder M. 2009.** RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10 (1)**: 57–63

**Wellman FL. 1961.** Coffee, botany, cultivation and utilization. *Leonard Hill Books, London*

**West-Eberhard MJ. 2005.** Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 6543–6549

**Wettenhall JM, Smyth GK. 2004.** LimmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics*, **20**:3705–3706.

**Whitman DW, Agrawal AA. (2009).** What is phenotypic plasticity and why is it important? In Phenotypic Plasticity of Insects (Whitman, D.W. and Ananthakrishnan, T.N., eds), pp. 1–63, *Science Publishers*

**Winkler H. 1916.** Über die experimentelle Erzeugung von Pflanzen mit abweichenden Chromosomenzahlen. *Zeitschr f Bot*, **8**:417–531.

**Wolfe KH. 2001.** Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* **2(5)**: 333-341.

**Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Doyle JJ, Rieseberg LH. 2009.** The frequency of polyploid speciation in vascular plants. *Proceedings of National Academy of Sciences of the United State of America* **106**: 13875-13879.

## Y

**Yadav NS, Wierzbicki A, Aegerter M, Caster CS, Pe´rez-Grau L, Kinney AJ, Hitz WD, Booth Jr JR, Schweiger B, Stecca KL, Allen SM, Blackwell M, Reiter RS, Carlson TJ, Russell SH, Feldmann KA, Pierce J, Browse J. 1993.** Cloning of higher plant omega-3 fatty acid desaturases. *Plant Physiol* **103**: 467–476

**Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC. 2002.** Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res.*, **30**:e48.

# Liste des annexes

## ***Chapitre I***

*Tableaux de l'article "The 'PUCE CAFÉ' project: the first 15K coffee microarray, a new tool for discovering candidate genes correlated to agronomic and quality traits"*

**Add 1:** The Missing Genes.

**Add 2:** Oligonucleotides that may cross-hybridize with several sequences.

**Add 3:** Q-PCR and Microarray Data for 111 Genes of Interest Extracted from Experiment 1.

**Add 4:** MIQE document (Minimum Information for Publication of Quantitative Real-Time PCR Experiments).

**Add 5:** Genes Significantly Expressed in Experiment 2.

**Add 6:** Gene Ontology Functional Enrichment.

**Add 7:** Reproducibility of Replicates for the Leaf Tissue in the Leaf-Flower Comparison.

**Add 8:** Theoretical Power Analysis for a Two-sample T-test to Detect a 1.25 to 2.50 Fold Change in Gene Expression as a Function of CV%.

**Add 9:** Lists of Genes Specifically Over-expressed in each Tissue Compared to the Two Others (Bean, Flower and Leaf).

## ***Chapitre II***

*Tableaux et figures de l'article "Genomic expression dominance in the natural allopolyploid Coffea arabica is massively affected by growth temperature"*

Fig. S1 Microarray experimental design.

Fig. S2 Transcriptome divergence and nonadditive gene expression between Coffea arabica cv T18141 and parental diploid species.

Table S1 Monitoring selection of differential responses of 66 unigenes in two growth temperature conditions

Table S2 BLAST2GO annotation for genes differentially expressed in the 'dominance' and 'transgression' categories

Table S3 The missing genes

Table S4 Oligonucleotides that may cross-hybridize with several sequences

## ***Chapitre III***

Annexe 16: Analyse d'enrichissement à l'aide du logiciel Blast 2Go (enrichment of Gene Ontology (GO) terms)

## ***Chapitre IV***

Annexe 17 : Liste de 139 processus biologiques obtenus à partir des gènes dits de 'plasticité' chez *Coffea arabica*

SGN-U359404  
SGN-U353149  
SGN-U359186  
SGN-U359155  
SGN-U359152  
SGN-U359114  
SGN-U359096  
SGN-U358918  
SGN-U358888  
SGN-U358880  
SGN-U358790  
SGN-U358758  
SGN-U358519  
SGN-U358452  
SGN-U358434  
SGN-U352281  
SGN-U358415  
SGN-U358406  
SGN-U358341  
SGN-U358288  
SGN-U358228  
SGN-U358209  
SGN-U352065  
SGN-U358140  
SGN-U351978  
SGN-U351957  
SGN-U358097  
SGN-U358053  
SGN-U358044  
SGN-U358012  
SGN-U357998  
SGN-U357876  
SGN-U351731  
SGN-U357868  
SGN-U351609  
SGN-U357742  
SGN-U357737  
SGN-U357736  
SGN-U357698  
SGN-U357691  
SGN-U357680  
SGN-U357671  
SGN-U357658  
SGN-U357587  
SGN-U357393  
SGN-U357373  
SGN-U351225  
SGN-U357255  
SGN-U357240

**Additional file 1:**

**The Missing Genes.** List of 195 unigenes for which the design of specific oligonucleotides could not be achieved and that are therefore absent from the PUCE CAFE microarray.

SGN-U357222  
SGN-U351033  
SGN-U357025  
SGN-U356995  
SGN-U350843  
SGN-U356938  
SGN-U356907  
SGN-U356886  
SGN-U356760  
SGN-U356700  
SGN-U356695  
SGN-U362832  
SGN-U362800  
SGN-U350506  
SGN-U362789  
SGN-U362780  
SGN-U356628  
SGN-U362738  
SGN-U362728  
SGN-U356569  
SGN-U350392  
SGN-U356529  
SGN-U356524  
SGN-U362656  
SGN-U362652  
SGN-U362637  
SGN-U362622  
SGN-U356462  
SGN-U362592  
SGN-U356439  
SGN-U362561  
SGN-U356411  
SGN-U350254  
SGN-U356396  
SGN-U356389  
SGN-U362516  
SGN-U356351  
SGN-U356328  
SGN-U362468  
SGN-U356324  
SGN-U356286  
SGN-U362422  
SGN-U356248  
SGN-U356246  
SGN-U356234  
SGN-U356217  
SGN-U350068  
SGN-U362328  
SGN-U356137

SGN-U362279  
SGN-U356114  
SGN-U362256  
SGN-U362228  
SGN-U362175  
SGN-U349834  
SGN-U355954  
SGN-U362057  
SGN-U349764  
SGN-U349724  
SGN-U361994  
SGN-U355845  
SGN-U361954  
SGN-U355808  
SGN-U361836  
SGN-U361831  
SGN-U361825  
SGN-U361792  
SGN-U361788  
SGN-U361786  
SGN-U361781  
SGN-U355627  
SGN-U361760  
SGN-U361711  
SGN-U361683  
SGN-U361615  
SGN-U361604  
SGN-U361560  
SGN-U361476  
SGN-U361475  
SGN-U361458  
SGN-U355305  
SGN-U355300  
SGN-U361429  
SGN-U349122  
SGN-U349114  
SGN-U361397  
SGN-U361337  
SGN-U361327  
SGN-U355183  
SGN-U349014  
SGN-U361285  
SGN-U361273  
SGN-U361229  
SGN-U361207  
SGN-U355015  
SGN-U348871  
SGN-U361134  
SGN-U361094



SGN-U361067  
SGN-U361060  
SGN-U348718  
SGN-U354841  
SGN-U360976  
SGN-U354804  
SGN-U360936  
SGN-U360898  
SGN-U360896  
SGN-U360858  
SGN-U360850  
SGN-U360823  
SGN-U354677  
SGN-U354670  
SGN-U360787  
SGN-U354640  
SGN-U360741  
SGN-U360723  
SGN-U354576  
SGN-U360686  
SGN-U354422  
SGN-U354358  
SGN-U354334  
SGN-U360477  
SGN-U360475  
SGN-U360468  
SGN-U354306  
SGN-U354305  
SGN-U360431  
SGN-U347984  
SGN-U360248  
SGN-U360232  
SGN-U360200  
SGN-U353989  
SGN-U360096  
SGN-U353884  
SGN-U359857  
SGN-U353632  
SGN-U353568  
SGN-U353567  
SGN-U359704  
SGN-U353550  
SGN-U353518  
SGN-U359634  
SGN-U353444  
SGN-U353375  
SGN-U353286  
SGN-U359428

Oligonucleotide	Unigene
Cof::opCocaV10015513	SGN-U356510
Cof::opCocaV10015511	SGN-U347161
Cof::opCocaV10015451	SGN-U355453
Cof::opCocaV10015421	SGN-U356181
Cof::opCocaV10015414	SGN-U361311
Cof::opCocaV10015390	SGN-U361983
Cof::opCocaV10015386	SGN-U356573
Cof::opCocaV10015256	SGN-U351191
Cof::opCocaV10015232	SGN-U356863
Cof::opCocaV10015177	SGN-U357764
Cof::opCocaV10015115	SGN-U348189
Cof::opCocaV10015091	SGN-U349720
Cof::opCocaV10014980	SGN-U356136
Cof::opCocaV10014913	SGN-U362019
Cof::opCocaV10014874	SGN-U361725
Cof::opCocaV10014868	SGN-U356226
Cof::opCocaV10014866	SGN-U353457
Cof::opCocaV10014858	SGN-U358775
Cof::opCocaV10014832	SGN-U349544
Cof::opCocaV10014831	SGN-U348401
Cof::opCocaV10014819	SGN-U354028
Cof::opCocaV10014799	SGN-U358661
Cof::opCocaV10014770	SGN-U349087
Cof::opCocaV10014736	SGN-U357341
Cof::opCocaV10014607	SGN-U350427
Cof::opCocaV10014477	SGN-U349855
Cof::opCocaV10014404	SGN-U352949
Cof::opCocaV10014380	SGN-U358507
Cof::opCocaV10014340	SGN-U352985
Cof::opCocaV10014250	SGN-U360229
Cof::opCocaV10014236	SGN-U359875
Cof::opCocaV10014217	SGN-U357810
Cof::opCocaV10014207	SGN-U349583
Cof::opCocaV10014195	SGN-U354407
Cof::opCocaV10014152	SGN-U351253
Cof::opCocaV10014100	SGN-U354864
Cof::opCocaV10014025	SGN-U351733
Cof::opCocaV10013957	SGN-U348100
Cof::opCocaV10013953	SGN-U359747
Cof::opCocaV10013918	SGN-U357952
Cof::opCocaV10013900	SGN-U356839
Cof::opCocaV10013795	SGN-U356370

**Additional file 2:**

**Oligonucleotides that may cross-hybridize with several sequences.** List of 371 oligonucleotides spotted on the microarray which have > 70% of overall identity to another unigene and have a contiguous identical length of more than 20 nt common to another unigene.

Cof::opCocaV10013711	SGN-U359987
Cof::opCocaV10013671	SGN-U353587
Cof::opCocaV10013656	SGN-U359869
Cof::opCocaV10013649	SGN-U352247
Cof::opCocaV10013565	SGN-U360453
Cof::opCocaV10013510	SGN-U355562
Cof::opCocaV10013455	SGN-U361840
Cof::opCocaV10013440	SGN-U358639
Cof::opCocaV10013406	SGN-U353294
Cof::opCocaV10013402	SGN-U350075
Cof::opCocaV10013398	SGN-U347400
Cof::opCocaV10013333	SGN-U357641
Cof::opCocaV10013283	SGN-U360902
Cof::opCocaV10013252	SGN-U348461
Cof::opCocaV10013201	SGN-U361169
Cof::opCocaV10013174	SGN-U358858
Cof::opCocaV10013168	SGN-U352653
Cof::opCocaV10013133	SGN-U358420
Cof::opCocaV10013104	SGN-U352208
Cof::opCocaV10013003	SGN-U353792
Cof::opCocaV10012981	SGN-U360728
Cof::opCocaV10012932	SGN-U358613
Cof::opCocaV10012910	SGN-U358691
Cof::opCocaV10012801	SGN-U360935
Cof::opCocaV10012797	SGN-U353307
Cof::opCocaV10012781	SGN-U356313
Cof::opCocaV10012721	SGN-U358064
Cof::opCocaV10012575	SGN-U356228
Cof::opCocaV10012513	SGN-U357629
Cof::opCocaV10012504	SGN-U358449
Cof::opCocaV10012497	SGN-U355002
Cof::opCocaV10012430	SGN-U358893
Cof::opCocaV10012374	SGN-U351940
Cof::opCocaV10012348	SGN-U353422
Cof::opCocaV10012329	SGN-U347354
Cof::opCocaV10012267	SGN-U361511
Cof::opCocaV10012263	SGN-U355945
Cof::opCocaV10012261	SGN-U356571
Cof::opCocaV10012226	SGN-U360587
Cof::opCocaV10012191	SGN-U361883
Cof::opCocaV10012157	SGN-U357422
Cof::opCocaV10012098	SGN-U347783
Cof::opCocaV10012060	SGN-U358703

Cof::opCocaV10012019	SGN-U360171
Cof::opCocaV10011985	SGN-U353076
Cof::opCocaV10011808	SGN-U355890
Cof::opCocaV10011789	SGN-U358087
Cof::opCocaV10011782	SGN-U357435
Cof::opCocaV10011727	SGN-U360661
Cof::opCocaV10011711	SGN-U360044
Cof::opCocaV10011640	SGN-U356926
Cof::opCocaV10011637	SGN-U359877
Cof::opCocaV10011590	SGN-U351454
Cof::opCocaV10011513	SGN-U358673
Cof::opCocaV10011466	SGN-U357152
Cof::opCocaV10011454	SGN-U348230
Cof::opCocaV10011442	SGN-U353500
Cof::opCocaV10011398	SGN-U358425
Cof::opCocaV10011388	SGN-U354872
Cof::opCocaV10011346	SGN-U350560
Cof::opCocaV10011306	SGN-U359744
Cof::opCocaV10011287	SGN-U359970
Cof::opCocaV10011244	SGN-U360559
Cof::opCocaV10011217	SGN-U353073
Cof::opCocaV10011212	SGN-U360025
Cof::opCocaV10011156	SGN-U356227
Cof::opCocaV10011092	SGN-U357182
Cof::opCocaV10011044	SGN-U355821
Cof::opCocaV10011024	SGN-U357271
Cof::opCocaV10010887	SGN-U358066
Cof::opCocaV10010881	SGN-U347969
Cof::opCocaV10010829	SGN-U356754
Cof::opCocaV10010822	SGN-U353493
Cof::opCocaV10010816	SGN-U348430
Cof::opCocaV10010776	SGN-U358311
Cof::opCocaV10010750	SGN-U352977
Cof::opCocaV10010740	SGN-U348991
Cof::opCocaV10010618	SGN-U355099
Cof::opCocaV10010611	SGN-U356810
Cof::opCocaV10010562	SGN-U350158
Cof::opCocaV10010491	SGN-U358603
Cof::opCocaV10010489	SGN-U349677
Cof::opCocaV10010483	SGN-U349705
Cof::opCocaV10010420	SGN-U360175
Cof::opCocaV10010398	SGN-U359833
Cof::opCocaV10010364	SGN-U350569

Cof::opCocaV10010351	SGN-U358185
Cof::opCocaV10010340	SGN-U356588
Cof::opCocaV10010334	SGN-U352492
Cof::opCocaV10010324	SGN-U358450
Cof::opCocaV10010311	SGN-U360706
Cof::opCocaV10010301	SGN-U357615
Cof::opCocaV10010284	SGN-U354462
Cof::opCocaV10010264	SGN-U357505
Cof::opCocaV10010263	SGN-U349961
Cof::opCocaV10010256	SGN-U353289
Cof::opCocaV10010121	SGN-U357030
Cof::opCocaV10010119	SGN-U349818
Cof::opCocaV10010102	SGN-U353386
Cof::opCocaV10010089	SGN-U350550
Cof::opCocaV10009982	SGN-U351923
Cof::opCocaV10009952	SGN-U352456
Cof::opCocaV10009931	SGN-U356603
Cof::opCocaV10009921	SGN-U348794
Cof::opCocaV10009774	SGN-U354751
Cof::opCocaV10009722	SGN-U347128
Cof::opCocaV10009710	SGN-U352947
Cof::opCocaV10009690	SGN-U352866
Cof::opCocaV10009664	SGN-U349792
Cof::opCocaV10009604	SGN-U356260
Cof::opCocaV10009547	SGN-U359911
Cof::opCocaV10009349	SGN-U358766
Cof::opCocaV10009294	SGN-U357630
Cof::opCocaV10009281	SGN-U358823
Cof::opCocaV10009258	SGN-U354266
Cof::opCocaV10009256	SGN-U358429
Cof::opCocaV10009183	SGN-U357715
Cof::opCocaV10009166	SGN-U362186
Cof::opCocaV10009161	SGN-U358121
Cof::opCocaV10009125	SGN-U357882
Cof::opCocaV10009117	SGN-U361741
Cof::opCocaV10009099	SGN-U352739
Cof::opCocaV10009077	SGN-U353591
Cof::opCocaV10009031	SGN-U353688
Cof::opCocaV10008885	SGN-U356188
Cof::opCocaV10008883	SGN-U356194
Cof::opCocaV10008728	SGN-U358055
Cof::opCocaV10008692	SGN-U352270
Cof::opCocaV10008660	SGN-U358455

Cof::opCocaV10008649	SGN-U357204
Cof::opCocaV10008617	SGN-U360355
Cof::opCocaV10008535	SGN-U362506
Cof::opCocaV10008326	SGN-U350575
Cof::opCocaV10008281	SGN-U362507
Cof::opCocaV10008273	SGN-U351126
Cof::opCocaV10008242	SGN-U349563
Cof::opCocaV10008221	SGN-U361618
Cof::opCocaV10008178	SGN-U360471
Cof::opCocaV10008076	SGN-U357466
Cof::opCocaV10008006	SGN-U357750
Cof::opCocaV10008000	SGN-U360361
Cof::opCocaV10007978	SGN-U360380
Cof::opCocaV10007934	SGN-U347831
Cof::opCocaV10007894	SGN-U356912
Cof::opCocaV10007877	SGN-U357191
Cof::opCocaV10007808	SGN-U352760
Cof::opCocaV10007800	SGN-U360110
Cof::opCocaV10007764	SGN-U356677
Cof::opCocaV10007684	SGN-U351075
Cof::opCocaV10007678	SGN-U357960
Cof::opCocaV10007670	SGN-U357007
Cof::opCocaV10007664	SGN-U359234
Cof::opCocaV10007587	SGN-U354049
Cof::opCocaV10007525	SGN-U351195
Cof::opCocaV10007521	SGN-U357561
Cof::opCocaV10007449	SGN-U357291
Cof::opCocaV10007430	SGN-U360795
Cof::opCocaV10007400	SGN-U356338
Cof::opCocaV10007398	SGN-U356040
Cof::opCocaV10007382	SGN-U357533
Cof::opCocaV10007352	SGN-U350347
Cof::opCocaV10007276	SGN-U354593
Cof::opCocaV10007252	SGN-U362174
Cof::opCocaV10007237	SGN-U354805
Cof::opCocaV10007212	SGN-U359867
Cof::opCocaV10007198	SGN-U356619
Cof::opCocaV10007080	SGN-U357940
Cof::opCocaV10007069	SGN-U353198
Cof::opCocaV10006984	SGN-U359021
Cof::opCocaV10006956	SGN-U353748
Cof::opCocaV10006935	SGN-U356184
Cof::opCocaV10006924	SGN-U354553

Cof::opCocaV10006846	SGN-U357394
Cof::opCocaV10006732	SGN-U361061
Cof::opCocaV10006712	SGN-U358366
Cof::opCocaV10006642	SGN-U357844
Cof::opCocaV10006622	SGN-U358409
Cof::opCocaV10006606	SGN-U362694
Cof::opCocaV10006604	SGN-U358811
Cof::opCocaV10006574	SGN-U357716
Cof::opCocaV10006541	SGN-U357058
Cof::opCocaV10006533	SGN-U359095
Cof::opCocaV10006479	SGN-U353364
Cof::opCocaV10006411	SGN-U347896
Cof::opCocaV10006396	SGN-U350609
Cof::opCocaV10006392	SGN-U357160
Cof::opCocaV10006375	SGN-U358715
Cof::opCocaV10006374	SGN-U355198
Cof::opCocaV10006333	SGN-U358352
Cof::opCocaV10006307	SGN-U353716
Cof::opCocaV10006229	SGN-U357247
Cof::opCocaV10006206	SGN-U347946
Cof::opCocaV10006142	SGN-U348835
Cof::opCocaV10006048	SGN-U362830
Cof::opCocaV10006028	SGN-U351718
Cof::opCocaV10005983	SGN-U357129
Cof::opCocaV10005970	SGN-U361064
Cof::opCocaV10005878	SGN-U360124
Cof::opCocaV10005717	SGN-U362378
Cof::opCocaV10005639	SGN-U361265
Cof::opCocaV10005632	SGN-U361538
Cof::opCocaV10005597	SGN-U361033
Cof::opCocaV10005585	SGN-U355113
Cof::opCocaV10005575	SGN-U358818
Cof::opCocaV10005553	SGN-U358159
Cof::opCocaV10005501	SGN-U361215
Cof::opCocaV10005470	SGN-U352782
Cof::opCocaV10005468	SGN-U362133
Cof::opCocaV10005464	SGN-U360351
Cof::opCocaV10005422	SGN-U347211
Cof::opCocaV10005411	SGN-U356598
Cof::opCocaV10005344	SGN-U356970
Cof::opCocaV10005343	SGN-U355589
Cof::opCocaV10005330	SGN-U356538
Cof::opCocaV10005283	SGN-U357056

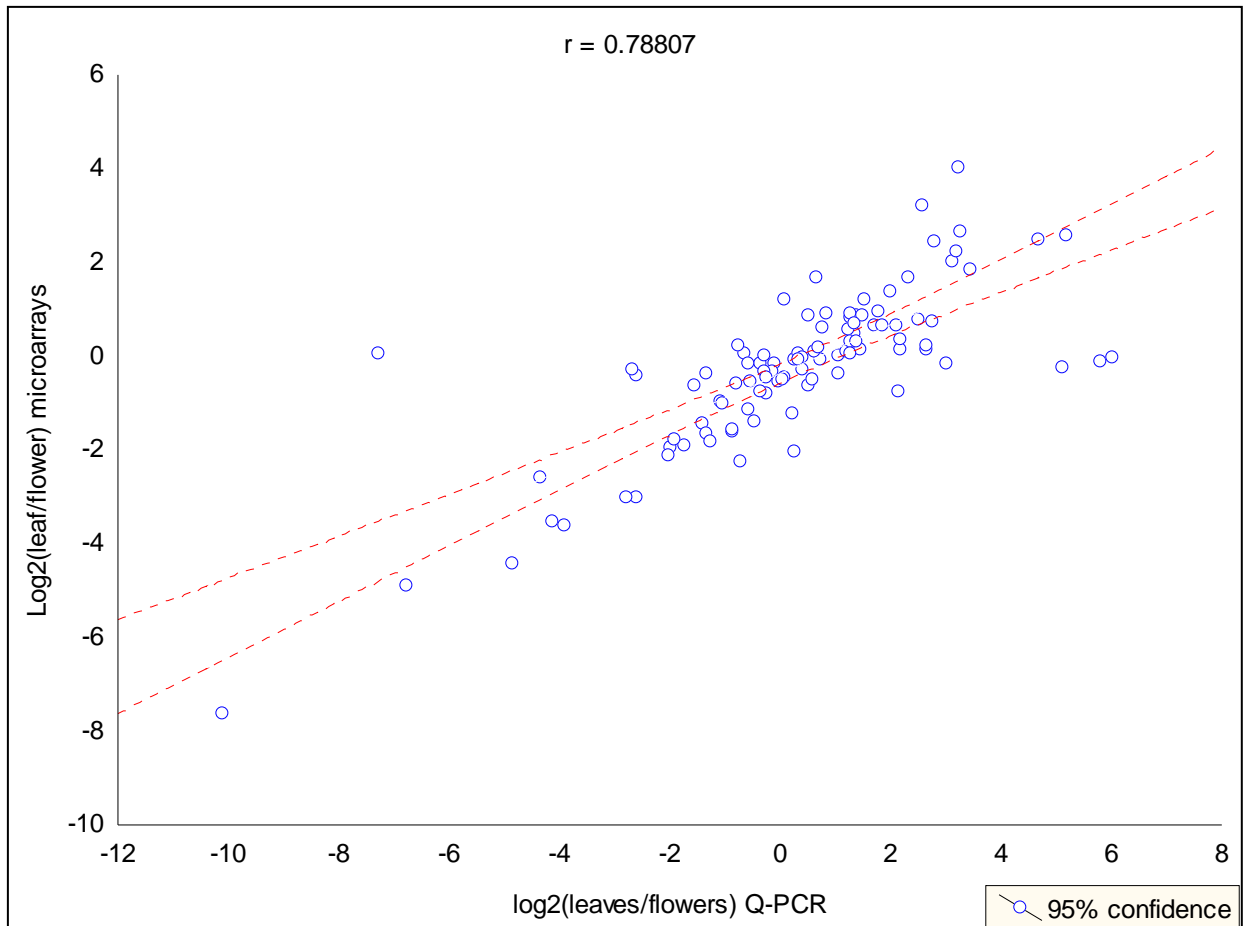
Cof::opCocaV10005249	SGN-U362323
Cof::opCocaV10005241	SGN-U356770
Cof::opCocaV10005139	SGN-U353784
Cof::opCocaV10005078	SGN-U354001
Cof::opCocaV10005074	SGN-U361179
Cof::opCocaV10004924	SGN-U359850
Cof::opCocaV10004917	SGN-U352557
Cof::opCocaV10004874	SGN-U356487
Cof::opCocaV10004841	SGN-U358713
Cof::opCocaV10004748	SGN-U352224
Cof::opCocaV10004736	SGN-U356304
Cof::opCocaV10004723	SGN-U357120
Cof::opCocaV10004685	SGN-U356681
Cof::opCocaV10004674	SGN-U352971
Cof::opCocaV10004653	SGN-U358173
Cof::opCocaV10004640	SGN-U357382
Cof::opCocaV10004606	SGN-U349520
Cof::opCocaV10004494	SGN-U358821
Cof::opCocaV10004480	SGN-U358927
Cof::opCocaV10004356	SGN-U360327
Cof::opCocaV10004345	SGN-U348454
Cof::opCocaV10004335	SGN-U361084
Cof::opCocaV10004313	SGN-U357625
Cof::opCocaV10004260	SGN-U350608
Cof::opCocaV10004176	SGN-U359712
Cof::opCocaV10004173	SGN-U362518
Cof::opCocaV10004159	SGN-U355842
Cof::opCocaV10004140	SGN-U362400
Cof::opCocaV10004038	SGN-U349710
Cof::opCocaV10003990	SGN-U356971
Cof::opCocaV10003985	SGN-U353635
Cof::opCocaV10003977	SGN-U357567
Cof::opCocaV10003901	SGN-U358379
Cof::opCocaV10003891	SGN-U358227
Cof::opCocaV10003697	SGN-U362108
Cof::opCocaV10003639	SGN-U361455
Cof::opCocaV10003551	SGN-U361105
Cof::opCocaV10003486	SGN-U351159
Cof::opCocaV10003464	SGN-U360911
Cof::opCocaV10003462	SGN-U351291
Cof::opCocaV10003446	SGN-U361097
Cof::opCocaV10003372	SGN-U353561
Cof::opCocaV10003313	SGN-U362159

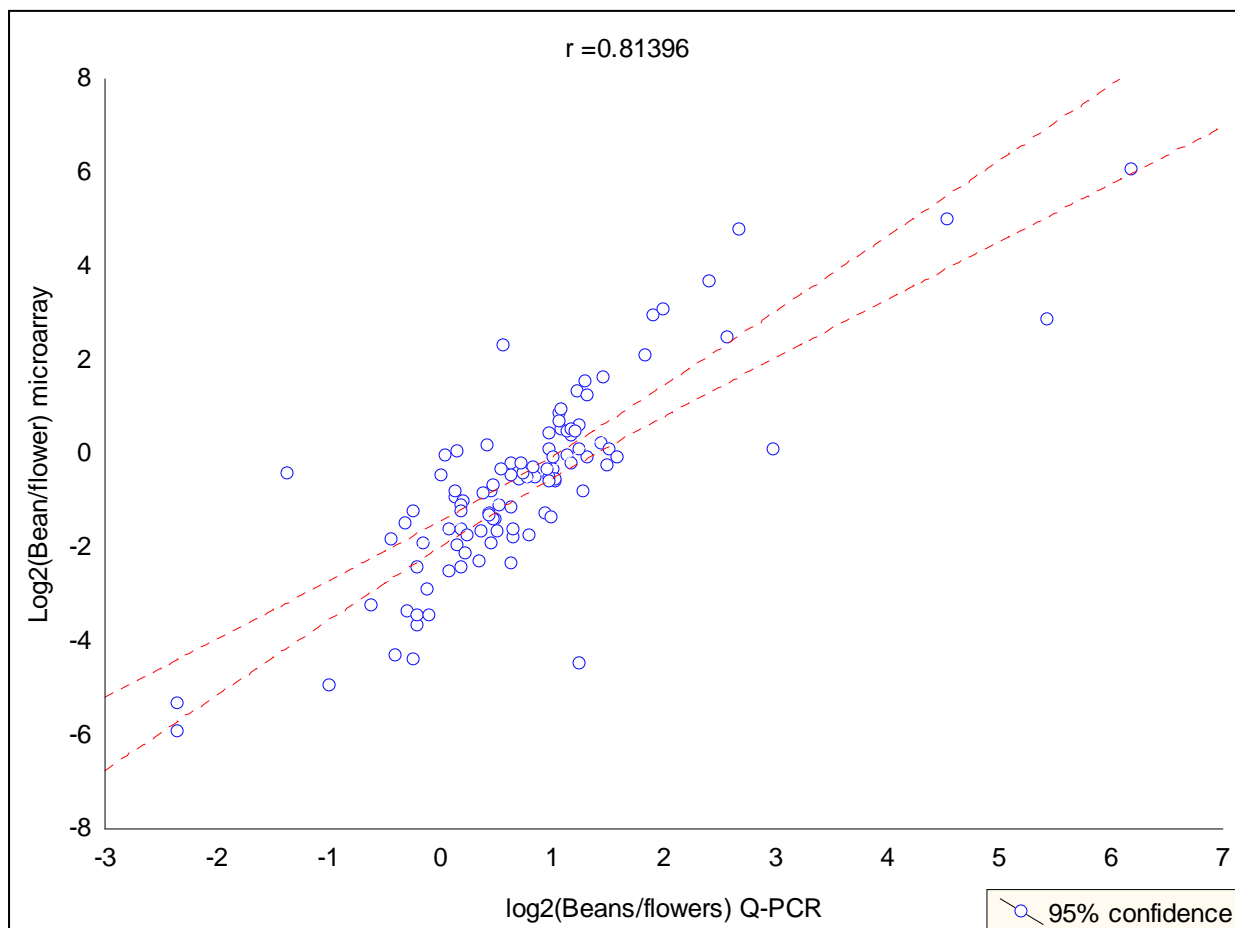
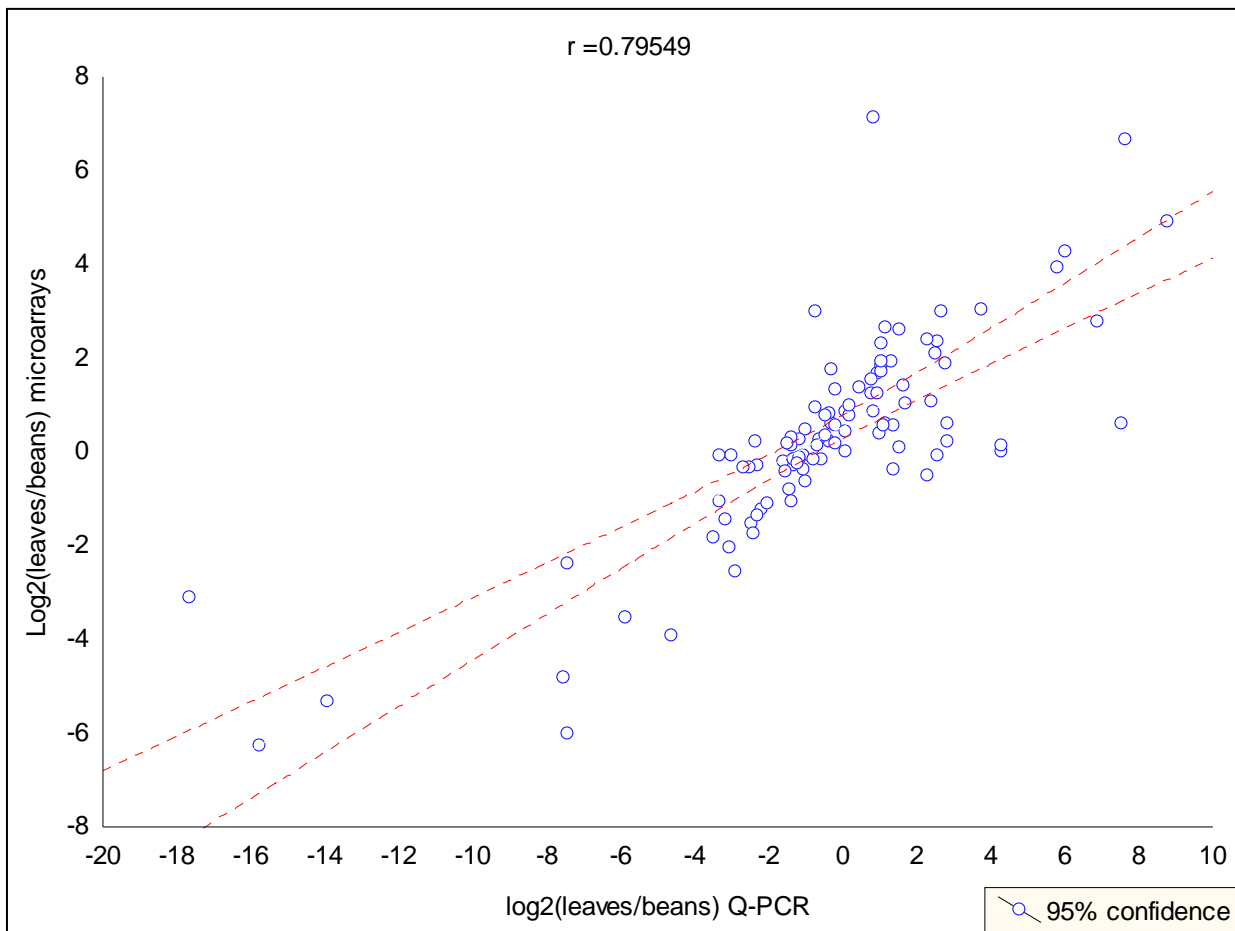


Cof::opCocaV10003296	SGN-U348439
Cof::opCocaV10003160	SGN-U350886
Cof::opCocaV10003156	SGN-U360205
Cof::opCocaV10003076	SGN-U347605
Cof::opCocaV10003075	SGN-U359902
Cof::opCocaV10003036	SGN-U361891
Cof::opCocaV10002997	SGN-U357609
Cof::opCocaV10002949	SGN-U356626
Cof::opCocaV10002850	SGN-U356583
Cof::opCocaV10002832	SGN-U359072
Cof::opCocaV10002668	SGN-U362402
Cof::opCocaV10002667	SGN-U360072
Cof::opCocaV10002663	SGN-U357463
Cof::opCocaV10002606	SGN-U362251
Cof::opCocaV10002595	SGN-U356026
Cof::opCocaV10002571	SGN-U360539
Cof::opCocaV10002545	SGN-U362692
Cof::opCocaV10002397	SGN-U348160
Cof::opCocaV10002280	SGN-U360775
Cof::opCocaV10002241	SGN-U361916
Cof::opCocaV10002239	SGN-U354502
Cof::opCocaV10002233	SGN-U355913
Cof::opCocaV10002214	SGN-U357762
Cof::opCocaV10002201	SGN-U358161
Cof::opCocaV10002192	SGN-U355538
Cof::opCocaV10002158	SGN-U350127
Cof::opCocaV10001921	SGN-U353486
Cof::opCocaV10001701	SGN-U360448
Cof::opCocaV10001688	SGN-U356355
Cof::opCocaV10001577	SGN-U353410
Cof::opCocaV10001440	SGN-U358143
Cof::opCocaV10001430	SGN-U362049
Cof::opCocaV10001409	SGN-U362650
Cof::opCocaV10001406	SGN-U353912
Cof::opCocaV10001401	SGN-U361005
Cof::opCocaV10001390	SGN-U353727
Cof::opCocaV10001347	SGN-U352923
Cof::opCocaV10001303	SGN-U361434
Cof::opCocaV10001298	SGN-U362726
Cof::opCocaV10001283	SGN-U356939
Cof::opCocaV10001282	SGN-U349028
Cof::opCocaV10001267	SGN-U360407
Cof::opCocaV10001221	SGN-U355523

Cof::opCocaV10001220	SGN-U360237
Cof::opCocaV10001193	SGN-U351324
Cof::opCocaV10001181	SGN-U352230
Cof::opCocaV10001176	SGN-U357044
Cof::opCocaV10001116	SGN-U362824
Cof::opCocaV10000912	SGN-U357915
Cof::opCocaV10000895	SGN-U354579
Cof::opCocaV10000821	SGN-U358384
Cof::opCocaV10000795	SGN-U354464
Cof::opCocaV10000789	SGN-U352328
Cof::opCocaV10000771	SGN-U361071
Cof::opCocaV10000759	SGN-U350183
Cof::opCocaV10000756	SGN-U361395
Cof::opCocaV10000654	SGN-U360159
Cof::opCocaV10000644	SGN-U358909
Cof::opCocaV10000566	SGN-U361390
Cof::opCocaV10000460	SGN-U355208
Cof::opCocaV10000448	SGN-U354288
Cof::opCocaV10000424	SGN-U362476
Cof::opCocaV10000409	SGN-U357442
Cof::opCocaV10000408	SGN-U361842
Cof::opCocaV10000389	SGN-U362548
Cof::opCocaV10000306	SGN-U358588
Cof::opCocaV10000181	SGN-U356557
Cof::opCocaV10000154	SGN-U357335
Cof::opCocaV10000139	SGN-U359098
Cof::opCocaV10000106	SGN-U361974
Cof::opCocaV10000017	SGN-U358616

**Additional file 3: Q-PCR and Microarray Data for 111 Genes of Interest Extracted from Experiment 1.** Microarray and Q-PCR were compared by calculating the Pearson correlation coefficients. For the microarray, the data input into the correlation analysis was the Log<sub>2</sub> ratio value of the weighted average and for each gene on the composite array representing all replicates. For qRT-PCR, we used the mean Log<sub>2</sub> ratio value representing all replicate plants. All correlation analyses carried an alpha value of 0.01 and were performed using SAS 9.2. The list of the primers and relative sequence used for the qPCR are indicted as well as the efficiency of amplification for each couple of primers.







UNIGENES	q-PCR Log2 (bean/flower)	q-PCR Log2(leaf/flower)	q-PCR Log2(leaves/beans)	microarray log2(BEAN/FLOWER )	microarray log2(LEAF/BEAN)	microarray Log2(LEAF/FLOWER)
SGN-U347142	3,20	6,01	2,82	-0,22	0,24	-0,02
SGN-U347219	-6,31	1,19	7,50	-0,39	0,60	0,11
SGN-U347362	-6,80	-4,05	2,75	-5,95	3,01	-2,69
SGN-U347468	-0,23	0,82	1,05	-1,88	2,35	0,92
SGN-U347498	0,41	-2,80	-3,21	-1,77	-1,44	-3,00
SGN-U347527	-0,31	2,32	2,63	-1,28	3,03	1,69
SGN-U347527	1,39	0,64	-0,76	-1,28	3,03	1,69
SGN-U347613	-9,54	-0,80	8,74	-5,90	4,94	-0,57
SGN-U347729	-1,70	5,18	6,88	-0,46	2,82	2,59
SGN-U347733	1,47	1,27	-0,20	-0,57	1,34	0,94
SGN-U347742	-1,25	1,26	2,51	-1,94	2,12	0,21
SGN-U347744	-1,00	0,32	1,32	-2,09	1,94	-0,05
SGN-U347752	0,87	-0,31	-1,18	-0,49	0,28	0,04
SGN-U347793	0,40	0,58	0,18	-1,61	1,02	-0,50
SGN-U347880	0,01	0,75	0,74	-1,10	1,55	0,63
SGN-U347922	2,57	1,51	-1,06	1,56	-0,37	1,23
SGN-U347948	2,34	-0,58	-2,92	1,34	-2,56	-1,12
SGN-U347974	0,15	-7,31	-7,45	2,31	-2,38	0,06
SGN-U348060	-2,13	0,40	2,53	-2,86	2,36	-0,27
SGN-U348063	1,77	-0,27	-2,04	0,70	-1,09	-0,43
SGN-U348076	-2,06	-0,54	1,53	-3,42	2,64	-0,53
SGN-U348085	-2,40	-1,36	1,05	-3,45	1,72	-1,63
SGN-U348109	1,84	1,24	-0,60	0,53	-0,13	0,33
SGN-U348109	2,16	1,38	-0,78	0,53	-0,13	0,33
SGN-U348135	2,62	-4,84	-7,46	1,27	-5,97	-4,42
SGN-U348168	2,04	1,26	-0,78	-0,01	0,97	0,85
SGN-U348210	3,04	0,72	-2,32	0,23	-0,26	-0,06
SGN-U348226	-2,76	-4,34	-1,58	-1,46	-0,42	-2,57
SGN-U348283	-0,07	-1,29	-1,22	-1,64	-0,24	-1,80
SGN-U348284	-1,23	1,04	2,27	0,07	-0,47	-0,34
SGN-U348291	-9,57	-6,77	2,80	-5,32	0,61	-4,86
SGN-U348350	0,36	0,05	-0,31	-0,45	1,78	1,24
SGN-U348442	-0,46	0,37	0,83	-0,85	0,86	-0,03
SGN-U348462	-2,44	1,32	3,76	-2,41	3,05	0,72
SGN-U348474	-1,13	-0,36	0,77	-1,62	1,27	-0,15
SGN-U348474	-1,49	-0,58	0,91	-1,62	1,27	-0,15
SGN-U348506	1,68	0,07	-1,61	-0,55	-0,20	-0,44
SGN-U348516	13,30	-0,66	-13,96	5,02	-5,32	0,06
SGN-U348531	-0,31	-0,28	0,03	-1,31	0,89	-0,33
SGN-U348571	10,95	1,03	-9,92	7,42	-7,61	-0,04
SGN-U348624	-2,57	3,24	5,81	-1,23	3,94	2,69
SGN-U348647	4,59	3,13	-1,47	2,96	-0,79	2,03
SGN-U348664	-3,05	-2,02	1,03	-4,29	1,96	-2,11
SGN-U348687	1,55	0,22	-1,33	-1,33	-0,15	-1,19
SGN-U348695	2,64	-0,88	-3,51	0,39	-2,06	-1,43
SGN-U348702	0,09	-2,60	-2,70	-0,32	-0,33	-0,38
SGN-U348705	1,02	1,97	0,95	-0,29	1,71	1,41
SGN-U348730	3,83	0,73	-3,11	2,44	-2,32	0,06
SGN-U348789	0,35	-0,16	-0,50	-1,14	0,81	-0,31
SGN-U348799	2,04	-0,27	-2,31	0,49	-1,36	-0,79
SGN-U348800	-0,16	1,46	1,63	-0,67	1,43	0,87
SGN-U348803	-1,07	1,70	2,76	-1,00	1,91	0,66
SGN-U348813	-2,42	-2,01	0,41	-3,65	1,40	-1,93
SGN-U349042	-1,28	-0,12	1,16	-0,77	0,60	-0,15
SGN-U349054	2,28	1,26	-1,02	0,49	-0,60	0,08
SGN-U349452	4,88	1,78	-3,10	3,10	-2,01	0,95
SGN-U349469	1,69	0,49	-1,20	-0,56	-0,16	-0,59
SGN-U349924	1,63	-0,76	-2,39	-0,05	0,25	0,26
SGN-U350103	3,07	0,61	-2,46	1,66	-1,51	0,10
SGN-U350118	1,50	1,31	-0,19	0,46	0,58	0,89
SGN-U350187	16,29	-1,36	-17,65	2,89	-3,09	-0,33
SGN-U350275	-0,18	-2,62	-2,44	-1,37	-1,72	-3,00
SGN-U350503	7,15	2,50	-4,65	4,81	-3,92	0,82
SGN-U350529	0,35	-0,74	-1,09	-2,31	-0,05	-2,22
SGN-U350674	1,86	0,50	-1,36	0,96	-0,15	0,87
SGN-U350714	2,38	-1,11	-3,49	0,64	-1,79	-0,95
SGN-U350724	1,42	1,45	0,03	-0,33	0,44	0,14
SGN-U350807	-1,09	0,25	1,35	-1,20	-0,35	-2,01
SGN-U350946	18,79	2,99	-15,81	6,10	-6,26	-0,12
SGN-U351021	-1,12	-1,75	-0,63	-2,41	0,30	-1,88
SGN-U351081	2,47	0,98	-1,49	0,09	-0,38	-0,18
SGN-U351087	1,81	0,40	-1,41	0,90	-1,03	-0,03

**Additional file 4****MIQE document (Minimum Information for Publication of Quantitative Real-Time PCR Experiments).**

<b>EXPERIMENTAL DESIGN</b>
Definition of experimental and control groups
Number within each group
<b>Assay carried out by core lab or investigator's lab?</b>
<b>Acknowledgement of authors' contributions</b>
<b>SAMPLE</b>
Description
Volume/mass of sample processed
<b>Microdissection or macrodissection</b>
Processing procedure
If frozen - how and how quickly?
If fixed - with what, how quickly?
Sample storage conditions and duration (especially for FFPE samples)

Experimental design is provided in the material and method section. Tissues (5 g) were harvested in the greenhouse and immediately frozen in liquid nitrogen. Tissues were stored in -80°C freezer until RNA extraction would be performed. Generally, storage did not exceed one week. Assays (Q-PCR experiment) were carried out by core lab.

<b>NUCLEIC ACID EXTRACTION</b>
Procedure and/or instrumentation
Name of kit and details of any modifications
<b>Source of additional reagents used</b>
Details of DNase or RNase treatment
Contamination assessment (DNA or RNA)
Nucleic acid quantification
Instrument and method
<b>Purity (A260/A280)</b>
<b>Yield</b>
RNA integrity method/instrument
RIN/RQI or Cq of 3' and 5' transcripts
<b>Electrophoresis traces</b>
Inhibition testing (Cq dilutions, spike or other)

Tissues were ground into a powder in liquid nitrogen and total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Cat# 74904), then treated with RQ1 DNase (Promega, Cat#M6101) following the manufacturer's instructions. Total RNAs were finally eluted from the columns with RNase-free water (2x30µL). For each tissue, three independent RNA extractions were performed. All RNA samples were analyzed by formaldehyde agarose gel electrophoresis to assess their integrity. To test for contamination by polyphenols, carbohydrates and proteins, a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE, USA) was used. Only RNA samples with OD 260/280 > 1.8 and OD 260/230 > 2 were used for further analysis. Exact values for each RNA can be provided upon request. The yield is about 10 µg of RNA per 100 mg of frozen powder.

### REVERSE TRANSCRIPTION

Complete reaction conditions

Amount of RNA and reaction volume

Priming oligonucleotide (if using GSP) and concentration

Reverse transcriptase and concentration

Temperature and time

**Manufacturer of reagents and catalogue numbers**

**Cqs with and without RT**

**Storage conditions of cDNA**

M-MLV Reverse Transcriptase from Invitrogen (Cat#28025-021; 200U/µl) was used for the generation of first strand cDNA in a 20µl reaction volume. 1µg of RNA (variable concentration), 1µL of random hexamer (50µM) and 1µl of dNTP mix (Invitrogen) were mixed, incubated at 65°C for 10min, and quickly chilled on ice. 4µL of 5x Buffer, 2µL of DTT 100mM, and 1µL of RNAsin (40U/uL) were added, and the reaction mix was incubated at 42°C for 2min. 1µL of the M-MLV were added, and the reaction mix was incubated at 42°C for 60min, then at 70°C for 15min. cDNA was stored in low adhesion tubes at -20°C.

### qPCR TARGET INFORMATION

If multiplex, efficiency and LOD of each assay.

Sequence accession number

**Location of amplicon**

Amplicon length

*In silico* specificity screen (BLAST, etc)

**Pseudogenes, retropseudogenes or other homologs?**

**Sequence alignment**

**Secondary structure analysis of amplicon**

Location of each primer by exon or intron (if applicable)

What splice variants are targeted?



Multiplex qPCR was not performed. Sequence accession numbers and amplicon lengths are included in Additional file 3. Primers were designed in exons or UTR regions close to the 3' end of the gene. No splice variants were targeted.

<b>qPCR OLIGONUCLEOTIDES</b>
Primer sequences
<b>RTPrimerDB Identification Number</b>
<b>Probe sequences</b>
Location and identity of any modifications
<b>Manufacturer of oligonucleotides</b>
<b>Purification method</b>

Primer sequences are included in the manuscript as Additional file 3. No modifications were used. Primers were purchased from Eurofins MWG Operon and are salt-free.

<b>qPCR PROTOCOL</b>
Complete reaction conditions
Reaction volume and amount of cDNA/DNA
Primer, (probe), Mg <sup>++</sup> and dNTP concentrations
Polymerase identity and concentration
Buffer/kit identity and manufacturer
<b>Exact chemical constitution of the buffer</b>
Additives (SYBR Green I, DMSO, etc.)
<b>Manufacturer of plates/tubes and catalog number</b>
Complete thermocycling parameters
<b>Reaction setup (manual/robotic)</b>
Manufacturer of qPCR instrument

The PCR reactions were performed using Roche's LC480 and white 384 well-plates. 5 µl of primers at 300 nM were distributed into wells and lyophilized. 2.5 µl of 2x SybrGreen PCR mix (Roche; #4887352001) and 2.5 µl of cDNA at 0.8 ng/µl were then added to each well. PCR parameters were as follow: 10 min. at 95°C, followed by 15 sec. at 95°C and 30 sec. at 95°C for 45 cycles. For the dissociation curves, fluorescence was recorded continuously during a ramp from 60°C to 95°C.

<b>qPCR VALIDATION</b>
<b>Evidence of optimisation (from gradients)</b>
Specificity (gel, sequence, melt, or digest)
For SYBR Green I, Cq of the NTC
Standard curves with slope and y-intercept
PCR efficiency calculated from slope
r <sup>2</sup> of standard curve
Linear dynamic range
Cq variation at lower limit
Evidence for limit of detection
If multiplex, efficiency and LOD of each assay.

The specificity of the amplification products have been confirmed by analyzing their melting curves, by size estimations on a 4% agarose gel, and finally by sequencing of the products. Without a template, no Cq could be determined since it never passed the threshold line. PCR efficiencies are included in Additional file 3.

<b>DATA ANALYSIS</b>
qPCR analysis program (source, version)
Cq method determination
Outlier identification and disposition
Results of NTCs
Justification of number and choice of reference genes
Description of normalisation method
<b>Number and concordance of biological replicates</b>
Number and stage (RT or qPCR) of technical replicates
Repeatability (intra-assay variation)
<b>Reproducibility (inter-assay variation, %CV)</b>
<b>Power analysis</b>
Statistical methods for result significance
Software (source, version)
<b>Cq or raw data submission using RDML</b>

qPCR analysis program (source, version): Roche  
Cq's were determined by setting the threshold to -1.0 using a log scale,  
No data have been exclude from the calculations  
Results of NTCs: no amplification products present thus no Cqs

Justification of number and choice of reference genes: 3 genes have been selected

Description of normalisation method: endogenous reference gene

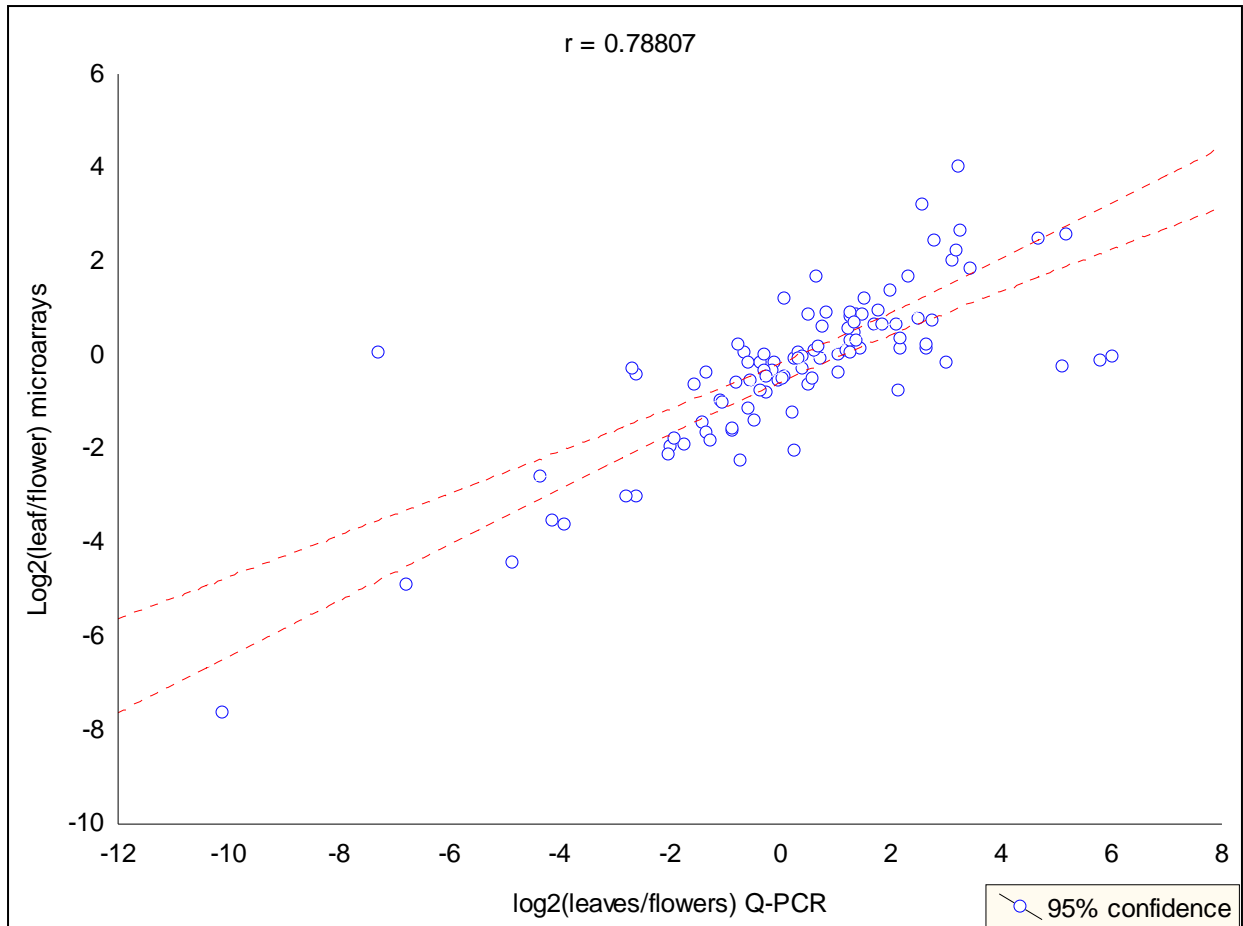
Number and concordance of biological replicates: 3

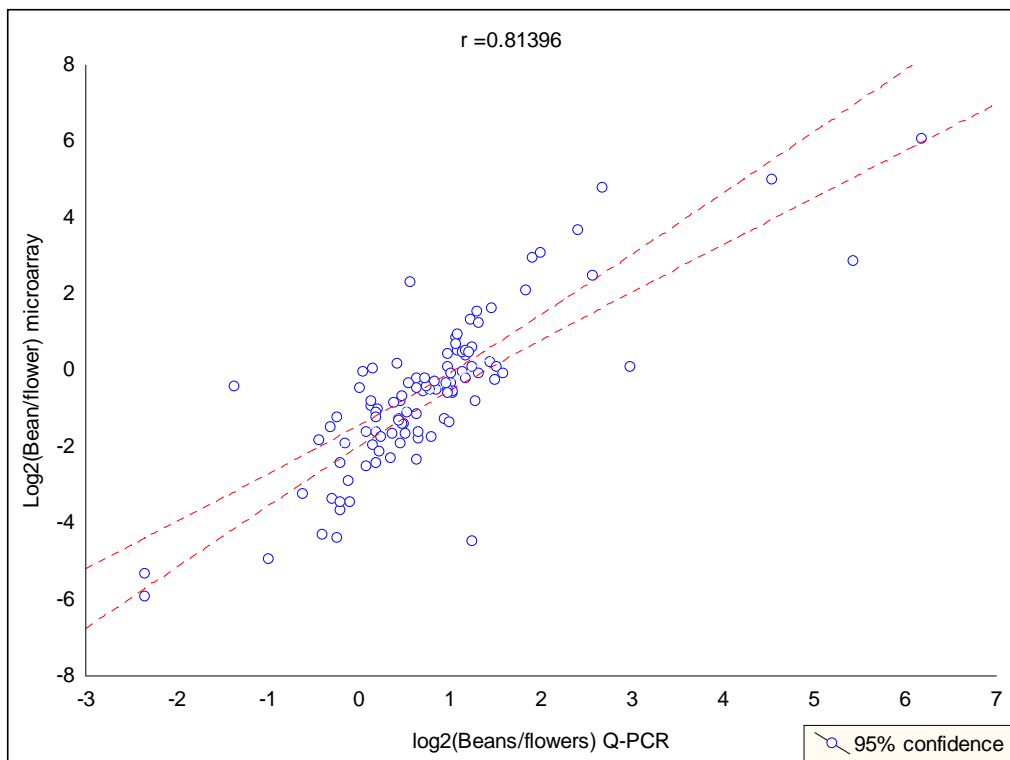
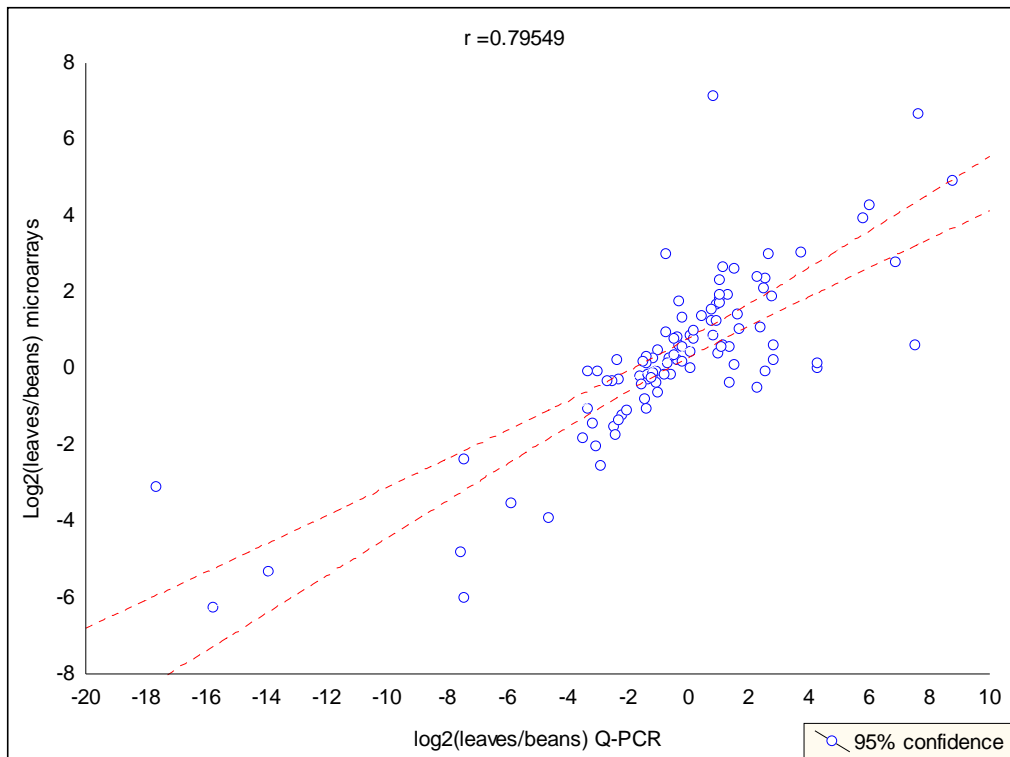
Number and stage (RT or qPCR) of technical replicates: 3 at qPCR level, 1 for RT (Mn samples)

Repeatability (intra-assay variation): was below one Cq

**Additional file 5**

**Genes Significantly Expressed in Experiment 2.** For each species analyzed in Experiment 2 (*C. arabica*, *C. canephora*, *C. eugenoides*), genes showing a significant hybridization signal compared to the background noise are indicated.







UNIGENES	q-PCR Log2 (bean/flower)	q-PCR Log2(leaf/flower)	q-PCR Log2(leaves/beans)	microarray log2(BEAN/FLOWER)	microarray log2(LEAF/BEAN)	microarray Log2(LEAF/FLOWER )
SGN-U347142	3.20	6.01	2.82	-0.22	0.24	-0.02
SGN-U347219	-6.31	1.19	7.50	-0.39	0.60	0.11
SGN-U347362	-6.80	-4.05	2.75	-5.95	3.01	-2.69
SGN-U347468	-0.23	0.82	1.05	-1.88	2.35	0.92
SGN-U347498	0.41	-2.80	-3.21	-1.77	-1.44	-3.00
SGN-U347527	-0.31	2.32	2.63	-1.28	3.03	1.69
SGN-U347527	1.39	0.64	-0.76	-1.28	3.03	1.69
SGN-U347613	-9.54	-0.80	8.74	-5.90	4.94	-0.57
SGN-U347729	-1.70	5.18	6.88	-0.46	2.82	2.59
SGN-U347733	1.47	1.27	-0.20	-0.57	1.34	0.94
SGN-U347742	-1.25	1.26	2.51	-1.94	2.12	0.21
SGN-U347744	-1.00	0.32	1.32	-2.09	1.94	-0.05
SGN-U347752	0.87	-0.31	-1.18	-0.49	0.28	0.04
SGN-U347793	0.40	0.58	0.18	-1.61	1.02	-0.50
SGN-U347880	0.01	0.75	0.74	-1.10	1.55	0.63
SGN-U347922	2.57	1.51	-1.06	1.56	-0.37	1.23
SGN-U347948	2.34	-0.58	-2.92	1.34	-2.56	-1.12
SGN-U347974	0.15	-7.31	-7.45	2.31	-2.38	0.06
SGN-U348060	-2.13	0.40	2.53	-2.86	2.36	-0.27
SGN-U348063	1.77	-0.27	-2.04	0.70	-1.09	-0.43
SGN-U348076	-2.06	-0.54	1.53	-3.42	2.64	-0.53
SGN-U348085	-2.40	-1.36	1.05	-3.45	1.72	-1.63
SGN-U348109	1.84	1.24	-0.60	0.53	-0.13	0.33
SGN-U348109	2.16	1.38	-0.78	0.53	-0.13	0.33
SGN-U348135	2.62	-4.84	-7.46	1.27	-5.97	-4.42
SGN-U348168	2.04	1.26	-0.78	-0.01	0.97	0.85
SGN-U348210	3.04	0.72	-2.32	0.23	-0.26	-0.06
SGN-U348226	-2.76	-4.34	-1.58	-1.46	-0.42	-2.57
SGN-U348283	-0.07	-1.29	-1.22	-1.64	-0.24	-1.80
SGN-U348284	-1.23	1.04	2.27	0.07	-0.47	-0.34
SGN-U348291	-9.57	-6.77	2.80	-5.32	0.61	-4.86
SGN-U348350	0.36	0.05	-0.31	-0.45	1.78	1.24
SGN-U348442	-0.46	0.37	0.83	-0.85	0.86	-0.03
SGN-U348462	-2.44	1.32	3.76	-2.41	3.05	0.72
SGN-U348474	-1.13	-0.36	0.77	-1.62	1.27	-0.15
SGN-U348474	-1.49	-0.58	0.91	-1.62	1.27	-0.15
SGN-U348506	1.68	0.07	-1.61	-0.55	-0.20	-0.44
SGN-U348516	13.30	-0.66	-13.96	5.02	-5.32	0.06
SGN-U348531	-0.31	-0.28	0.03	-1.31	0.89	-0.33
SGN-U348571	10.95	1.03	-9.92	7.42	-7.61	-0.04
SGN-U348624	-2.57	3.24	5.81	-1.23	3.94	2.69
SGN-U348647	4.59	3.13	-1.47	2.96	-0.79	2.03
SGN-U348664	-3.05	-2.02	1.03	-4.29	1.96	-2.11
SGN-U348687	1.55	0.22	-1.33	-1.33	-0.15	-1.19
SGN-U348695	2.64	-0.88	-3.51	0.39	-2.06	-1.43
SGN-U348702	0.09	-2.60	-2.70	-0.32	-0.33	-0.38
SGN-U348705	1.02	1.97	0.95	-0.29	1.71	1.41
SGN-U348730	3.83	0.73	-3.11	2.44	-2.32	0.06
SGN-U348789	0.35	-0.16	-0.50	-1.14	0.81	-0.31
SGN-U348799	2.04	-0.27	-2.31	0.49	-1.36	-0.79
SGN-U348800	-0.16	1.46	1.63	-0.67	1.43	0.87
SGN-U348803	-1.07	1.70	2.76	-1.00	1.91	0.66
SGN-U348813	-2.42	-2.01	0.41	-3.65	1.40	-1.93
SGN-U349042	-1.28	-0.12	1.16	-0.77	0.60	-0.15
SGN-U349054	2.28	1.26	-1.02	0.49	-0.60	0.08
SGN-U349452	4.88	1.78	-3.10	3.10	-2.01	0.95
SGN-U349469	1.69	0.49	-1.20	-0.56	-0.16	-0.59
SGN-U349924	1.63	-0.76	-2.39	-0.05	0.25	0.26
SGN-U350103	3.07	0.61	-2.46	1.66	-1.51	0.10
SGN-U350118	1.50	1.31	-0.19	0.46	0.58	0.89
SGN-U350187	16.29	-1.36	-17.65	2.89	-3.09	-0.33
SGN-U350275	-0.18	-2.62	-2.44	-1.37	-1.72	-3.00
SGN-U350503	7.15	2.50	-4.65	4.81	-3.92	0.82
SGN-U350529	0.35	-0.74	-1.09	-2.31	-0.05	-2.22
SGN-U350674	1.86	0.50	-1.36	0.96	-0.15	0.87
SGN-U350714	2.38	-1.11	-3.49	0.64	-1.79	-0.95
SGN-U350724	1.42	1.45	0.03	-0.33	0.44	0.14
SGN-U350807	-1.09	0.25	1.35	-1.20	-0.35	-2.01
SGN-U350946	18.79	2.99	-15.81	6.10	-6.26	-0.12
SGN-U351021	-1.12	-1.75	-0.63	-2.41	0.30	-1.88
SGN-U351081	2.47	0.98	-1.49	0.09	-0.38	-0.18
SGN-U351087	1.81	0.40	-1.41	0.90	-1.03	-0.03

**Additional file 6**

**Gene Ontology Functional Enrichment.** Functional enrichment using gene ontology terms and Fisher's Exact Test was performed using Blast2GO for the sets of up-regulated unigenes in each of the three tissues (bean, leaf and flower). The full set of unigenes of the PUCE CAFE array was used as the reference set. The resulting p-values are indicated after adjusting for FDR multiple testing results. The over- or under-expressed functions at  $p \leq 0.01$  are specified.

## FLOWER

GO term	Ontology domains	FDR	p-Value	Number of genes in the set test	Number of genes in the reference set	Number of gene not annotated in the set test	Number of gene not annotated in the reference test	Over/Under
GO:0044260	cellular macromolecule metabolic process	0	2,08E-04	7	1450	111	6827	under
GO:0043170	macromolecule metabolic process	0	0,00261383	12	1684	106	6593	under
GO:0009507	chloroplast	0	0,00422744	1	523	117	7754	under
GO:0016817	hydrolase activity, acting on acid anhydrides	0	0,00580003	0	356	118	7921	under
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	0	0,00588642	0	355	118	7922	under
GO:0016462	pyrophosphatase activity	0	0,00597409	0	354	118	7923	under
GO:0017111	nucleoside-triphosphatase activity	0	0,00702818	0	343	118	7934	under
GO:0019538	protein metabolic process	0	0,0108581	8	1168	110	7109	under
GO:0009059	macromolecule biosynthetic process	0	0,0114444	1	447	117	7830	under
GO:0034645	cellular macromolecule biosynthetic process	0	0,011593	1	446	117	7831	under
GO:0044267	cellular protein metabolic process	0	0,012754	5	868	113	7409	under
GO:0005198	structural molecule activity	0	0,0151058	0	291	118	7986	under
GO:0032991	macromolecular complex	0	0,0158564	5	846	113	7431	under
GO:0003723	RNA binding	0	0,0174905	0	281	118	7996	under
GO:0044435	plastid part	0	0,019664	0	273	118	8004	under
GO:0032555	purine ribonucleotide binding	0	0,0222571	6	905	112	7372	under
GO:0032553	ribonucleotide binding	0	0,0222571	6	905	112	7372	under
GO:0000166	nucleotide binding	0	0,0244491	10	1251	108	7026	under
GO:0017076	purine nucleotide binding	0	0,0286282	7	968	111	7309	under
GO:0006259	DNA metabolic process	0	0,0296017	0	245	118	8032	under
GO:0005886	plasma membrane	0	0,0311281	1	368	117	7909	under
GO:0006412	translation	0	0,0379205	0	228	118	8049	under
GO:0090304	nucleic acid metabolic process	0	0,0418872	2	452	116	7825	under
GO:0003735	structural constituent of ribosome	0	0,0419852	0	221	118	8056	under
GO:0043228	non-membrane-bounded organelle	0	0,0478518	0	212	118	8065	under
GO:0043232	intracellular non-membrane-bounded organelle	0	0,0478518	0	212	118	8065	under
GO:0043234	protein complex	0	0,0559386	3	524	115	7753	under
GO:0016787	hydrolase activity	0	0,0582703	12	1302	106	6975	under
GO:0030529	ribonucleoprotein complex	0	0,0654375	1	307	117	7970	under
GO:0016788	hydrolase activity, acting on ester bonds	0	0,0662227	1	306	117	7971	under
GO:0009987	cellular process	0	0,0692509	33	2887	85	5390	under
GO:0032559	adenyl ribonucleotide binding	0	0,0695614	6	773	112	7504	under
GO:0005524	ATP binding	0	0,0706921	6	771	112	7506	under
GO:0016887	ATPase activity	0	0,0718146	0	184	118	8093	under
GO:0044422	organelle part	0	0,0742949	7	851	111	7426	under
GO:0044446	intracellular organelle part	0	0,0760188	7	848	111	7429	under
GO:0044464	cell part	0	0,0776851	72	5600	46	2677	under
GO:0001883	purine nucleoside binding	0	0,0845186	7	834	111	7443	under
GO:0001882	nucleoside binding	0	0,0845186	7	834	111	7443	under
GO:0030554	adenyl nucleotide binding	0	0,0845186	7	834	111	7443	under
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0	0,0864856	4	570	114	7707	under
GO:0023052	signaling	0	0,0931619	0	166	118	8111	under
GO:0023033	signaling pathway	0	0,0958918	0	164	118	8113	under
GO:0044434	chloroplast part	0	0,0987009	0	162	118	8115	under
GO:0016779	nucleotidyltransferase activity	0	0,104566	0	158	118	8119	under
GO:0051234	establishment of localization	0	0,106111	6	719	112	7558	under
GO:0006810	transport	0	0,106916	6	718	112	7559	under
GO:0019001	guanyl nucleotide binding	0	0,107627	0	156	118	8121	under
GO:0032561	guanyl ribonucleotide binding	0	0,109191	0	155	118	8122	under
GO:0005525	GTP binding	0	0,109191	0	155	118	8122	under
GO:0016020	membrane	0	0,121507	11	1108	107	7169	under
GO:0051649	establishment of localization in cell	0	0,122539	0	147	118	8130	under
GO:0005634	nucleus	0	0,123661	4	528	114	7749	under
GO:0042623	ATPase activity, coupled	0	0,127951	0	144	118	8133	under
GO:0044237	cellular metabolic process	0	0,131505	28	2387	90	5890	under
GO:0051716	cellular response to stimulus	0	0,1336	0	141	118	8136	under
GO:0046907	intracellular transport	0	0,137502	0	139	118	8138	under
GO:0009057	macromolecule catabolic process	0	0,143569	0	136	118	8141	under
GO:0003676	nucleic acid binding	0	0,146667	12	1155	106	7122	under
GO:0005840	ribosome	0	0,149902	0	133	118	8144	under
GO:0044425	membrane part	0	0,151612	7	752	111	7525	under
GO:0010319	stromule	0	0,170608	0	124	118	8153	under
GO:0016791	phosphatase activity	0	0,175581	0	122	118	8155	under
GO:0044424	intracellular part	0	0,18079	63	4800	55	3477	under
GO:0033279	ribosomal subunit	0	0,183312	0	119	118	17858	under



## BEAN

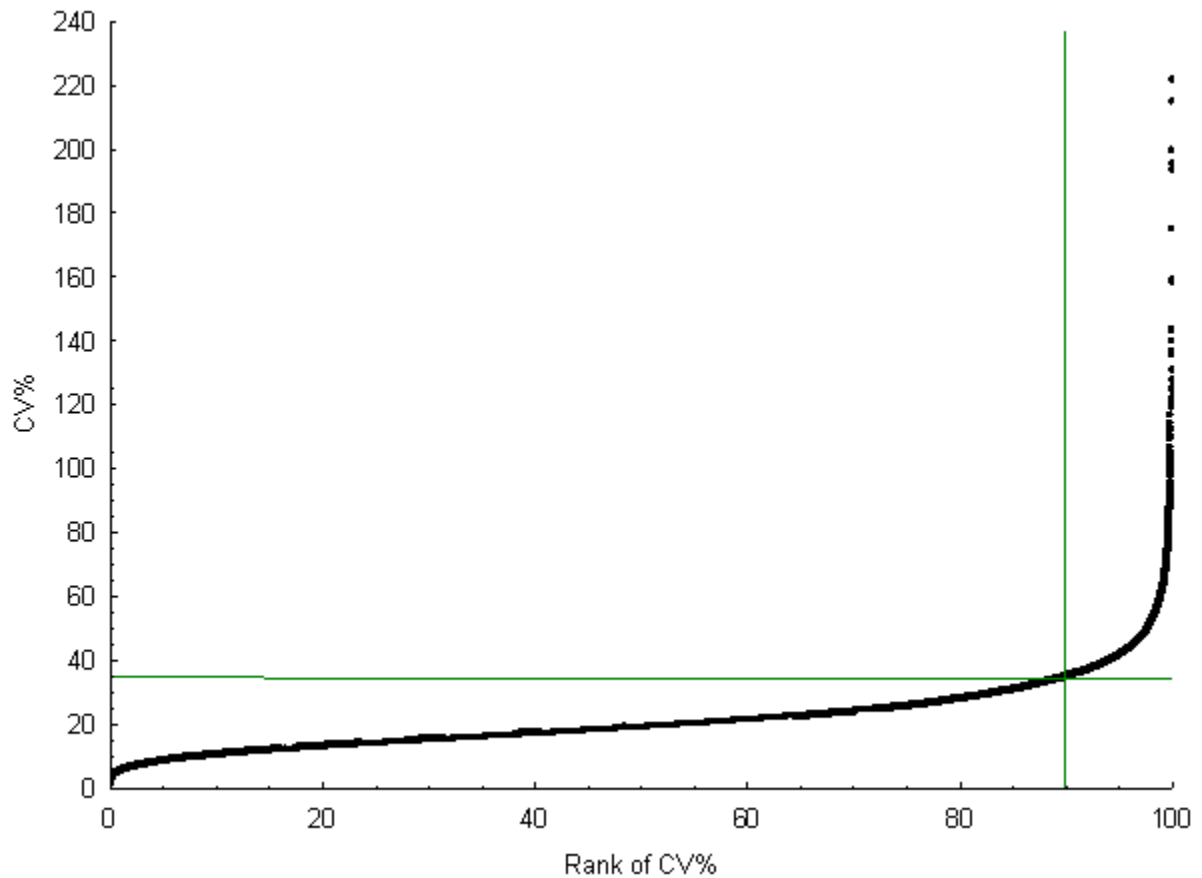
GO term	Ontology domains	FDR	p-Value	Number of genes in the set test	Number of genes in the reference set	Number of gene not annotated in the set test	Number of gene not annotated in the reference test	Over/Under
GO:0009507	chloroplast	1,44E-04	1,23E-07	21	503	819	7052	under
GO:0032991	macromolecular complex	1,44E-04	2,90E-07	46	805	794	6750	under
GO:0005198	structural molecule activity	1,44E-04	2,95E-07	7	284	833	7271	under
GO:0009987	cellular process	1,44E-04	3,79E-07	228	2692	612	4863	under
GO:0044237	cellular metabolic process	1,45E-04	1,38E-05	190	2225	650	5330	under
GO:0019538	protein metabolic process	1,45E-04	1,46E-05	79	1097	761	6458	under
GO:0051234	establishment of localization	1,45E-04	1,72E-05	42	683	798	6872	under
GO:0044267	cellular protein metabolic process	1,45E-04	1,78E-05	54	819	786	6736	under
GO:0005215	transporter activity	1,45E-04	1,80E-05	21	428	819	7127	under
GO:0006810	transport	1,45E-04	1,81E-05	42	682	798	6873	under
GO:0003735	structural constituent of ribosome	1,68E-04	2,38E-05	6	215	834	7340	under
GO:0006412	translation	3,06E-04	5,10E-05	7	221	833	7334	under
GO:0005886	plasma membrane	4,59E-04	8,42E-05	17	352	823	7203	under
GO:0043234	protein complex	4,89E-04	9,53E-05	29	498	811	7057	under
GO:0044238	primary metabolic process	5,13E-04	1,04E-04	208	2337	632	5218	under
GO:0008152	metabolic process	5,47E-04	1,13E-04	302	3218	538	4337	under
GO:0043170	macromolecule metabolic process	5,98E-04	1,33E-04	130	1566	710	5989	under
GO:0044260	cellular macromolecule metabolic process	6,77E-04	1,65E-04	109	1348	731	6207	under
GO:0043232	intracellular non-membrane-bounded organelle	6,77E-04	1,74E-04	7	205	833	7350	under
GO:0043228	non-membrane-bounded organelle	6,77E-04	1,74E-04	7	205	833	7350	under
GO:0016020	membrane	9,98E-04	2,39E-04	80	1039	760	6516	under
GO:0009058	biosynthetic process	0,00106476	2,68E-04	72	952	768	6603	under
GO:0044425	membrane part	0,00141738	3,59E-04	50	709	790	6846	under
GO:0044435	plastid part	0,00163428	4,40E-04	12	261	828	7294	under
GO:0022857	transmembrane transporter activity	0,00163428	4,42E-04	13	274	827	7281	under
GO:0003824	catalytic activity	0,00189099	5,04E-04	342	3530	498	4025	under
GO:0044434	chloroplast part	0,00254865	7,09E-04	5	157	835	7398	under
GO:0009059	macromolecule biosynthetic process	0,00269295	8,07E-04	26	422	814	7133	under
GO:0044249	cellular biosynthetic process	0,00269295	8,27E-04	65	849	775	6706	under
GO:0004557	alpha-galactosidase activity	0,00281848	5,76E-06	9	7	831	7548	over
GO:0005681	spliceosomal complex	0,00281848	8,97E-06	7	3	833	7552	over
GO:0034645	cellular macromolecule biosynthetic process	0,0028253	8,49E-04	26	421	814	7134	under
GO:0044436	thylakoid part	0,00328469	9,95E-04	2	105	838	7450	under
GO:0022891	substrate-specific transmembrane transporter activity	0,00359537	0,00108038	11	234	829	7321	under
GO:0022804	active transmembrane transporter activity	0,00359537	0,00109904	7	180	833	7375	under
GO:0044422	organelle part	0,00393867	0,00120106	61	797	779	6758	under
GO:0044446	intracellular organelle part	0,00419956	0,00134431	61	794	779	6761	under
GO:0032555	purine ribonucleotide binding	0,00431601	0,00141401	66	845	774	6710	under
GO:0032553	ribonucleotide binding	0,00431601	0,00141401	66	845	774	6710	under
GO:0030529	ribonucleoprotein complex	0,00433532	0,00145414	16	292	824	7263	under
GO:0008324	cation transmembrane transporter activity	0,00433532	0,00146008	6	162	834	7393	under
GO:0009579	thylakoid	0,00470274	0,00156356	1	82	839	7473	under
GO:0033279	ribosomal subunit	0,00485575	0,00161028	3	116	837	7439	under
GO:0017076	purine nucleotide binding	0,00485575	0,00164705	72	903	768	6652	under
GO:0004721	phosphoprotein phosphatase activity	0,00552311	0,00194412	0	59	840	7496	under
GO:0005840	ribosome	0,00552311	0,00195906	4	129	836	7426	under
GO:0015075	ion transmembrane transporter activity	0,00552311	0,00200685	8	185	832	7370	under
GO:0034357	photosynthetic membrane	0,00568753	0,00208121	1	79	839	7476	under
GO:0042651	thylakoid membrane	0,00568753	0,00208121	1	79	839	7476	under
GO:0008270	zinc ion binding	0,00618385	2,63E-05	78	418	762	7137	over
GO:0046914	transition metal ion binding	0,00636089	4,36E-05	118	719	722	6836	over
GO:0022892	substrate-specific transporter activity	0,00716099	0,00269202	16	281	824	7274	under
GO:0016740	transferase activity	0,00824596	0,00312294	101	1177	739	6378	under
GO:0032559	adenyl ribonucleotide binding	0,0111536	0,00402301	57	722	783	6833	under
GO:0022890	inorganic cation transmembrane transporter activity	0,0114532	0,00406638	2	89	838	7466	under
GO:0005524	ATP binding	0,0118363	0,00431621	57	720	783	6835	under
GO:0009535	chloroplast thylakoid membrane	0,0120496	0,00443349	1	71	839	7484	under
GO:0055035	plastid thylakoid membrane	0,0120496	0,00443349	1	71	839	7484	under
GO:0044430	cytoskeletal part	0,0120496	0,00454504	0	51	840	7504	under
GO:0001882	nucleoside binding	0,0120733	0,00495638	63	778	777	6777	under
GO:0001883	purine nucleoside binding	0,0120733	0,00495638	63	778	777	6777	under
GO:0030554	adenyl nucleotide binding	0,0120733	0,00495638	63	778	777	6777	under
GO:0030234	enzyme regulator activity	0,0131216	0,00534766	1	69	839	7486	under
GO:0006629	lipid metabolic process	0,0131216	0,00543959	12	220	828	7335	under

## LEAF

GO term	Ontology domains	FDR	p-Value	Number of genes in the set test	Number of genes in the reference set	Number of gene not annotated in the set test	Number of gene not annotated in the reference test	Over/Under
GO:0016772	transferase activity, transferring phosphorus-containing groups	0	0,00402792	2	680	111	7602	under
GO:0016788	hydrolase activity, acting on ester bonds	0	0,0144268	0	307	113	7975	under
GO:0006508	proteolysis	0	0,0154778	0	302	113	7980	under
GO:0008233	peptidase activity	0	0,0163731	0	298	113	7984	under
GO:0019538	protein metabolic process	0	0,016722	8	1168	105	7114	under
GO:0016301	kinase activity	0	0,0241865	2	523	111	7759	under
GO:0016020	membrane	0	0,026796	8	1111	105	7171	under
GO:0044238	primary metabolic process	0	0,0328791	25	2520	88	5762	under
GO:0070011	peptidase activity, acting on L-amino acid peptides	0	0,0334512	0	247	113	8035	under
GO:0004674	protein serine/threonine kinase activity	0	0,0368847	0	240	113	8042	under
GO:0006629	lipid metabolic process	0	0,0412381	0	232	113	8050	under
GO:0044281	small molecule metabolic process	0	0,0418137	4	678	109	7604	under
GO:0004672	protein kinase activity	0	0,0493244	1	345	112	7937	under
GO:0003824	catalytic activity	0	0,0501283	43	3829	70	4453	under
GO:0006796	phosphate metabolic process	0	0,0522728	1	340	112	7942	under
GO:0006793	phosphorus metabolic process	0	0,0522728	1	340	112	7942	under
GO:0043412	macromolecule modification	0	0,0608524	3	538	110	7744	under
GO:0016773	phosphotransferase activity, alcohol group as acceptor	0	0,061969	2	434	111	7848	under
GO:0043170	macromolecule metabolic process	0	0,0634309	16	1680	97	6602	under
GO:0044249	cellular biosynthetic process	0	0,0645846	7	907	106	7375	under
GO:0044237	cellular metabolic process	0	0,0687183	25	2390	88	5892	under
GO:0006464	protein modification process	0	0,0706173	3	522	110	7760	under
GO:0015075	ion transmembrane transporter activity	0	0,0709285	0	193	113	8089	under
GO:0009056	catabolic process	0	0,077175	1	306	112	7976	under
GO:0043687	post-translational protein modification	0	0,0813456	2	407	111	7875	under
GO:0016310	phosphorylation	0	0,0835205	1	299	112	7983	under
GO:0016787	hydrolase activity	0	0,0835234	12	1302	101	6980	under
GO:0044464	cell part	0	0,0843742	69	5603	44	2679	under
GO:0044267	cellular protein metabolic process	0	0,0872773	7	866	106	7416	under
GO:0044422	organelle part	0	0,0971026	7	851	106	7431	under
GO:0006468	protein amino acid phosphorylation	0	0,0987836	1	284	112	7998	under
GO:0044446	intracellular organelle part	0	0,0991739	7	848	106	7434	under
GO:0004175	endopeptidase activity	0	0,100278	0	168	113	8114	under
GO:0008324	cation transmembrane transporter activity	0	0,100278	0	168	113	8114	under
GO:0006811	ion transport	0	0,100278	0	168	113	8114	under
GO:0016740	transferase activity	0	0,103884	12	1266	101	7016	under
GO:0006066	alcohol metabolic process	0	0,10895	0	162	113	8120	under
GO:0016779	nucleotidyltransferase activity	0	0,115141	0	158	113	8124	under
GO:0044255	cellular lipid metabolic process	0	0,115141	0	158	113	8124	under
GO:0005506	iron ion binding	0	0,116586	1	269	112	8013	under
GO:0009987	cellular process	0	0,123425	33	2887	80	5395	under
GO:0008152	metabolic process	0	0,129159	41	3479	72	4803	under
GO:0065007	biological regulation	0	0,133737	6	718	107	7564	under
GO:0042180	cellular ketone metabolic process	0	0,145019	2	347	111	7935	under
GO:0006812	cation transport	0	0,147601	0	140	113	8142	under
GO:0006082	organic acid metabolic process	0	0,147739	2	345	111	7937	under
GO:0042578	phosphoric ester hydrolase activity	0	0,149649	0	139	113	8143	under
GO:0019752	carboxylic acid metabolic process	0	0,150502	2	343	111	7939	under
GO:0043436	oxoacid metabolic process	0	0,150502	2	343	111	7939	under
GO:0019222	regulation of metabolic process	0	0,150623	4	526	109	7756	under
GO:0044260	cellular macromolecule metabolic process	0	0,151275	15	1442	98	6840	under
GO:0005515	protein binding	0	0,151395	10	1036	103	7246	under
GO:0022891	substrate-specific transmembrane transporter activity	0	0,152876	1	244	112	8038	under
GO:0043234	protein complex	0	0,154126	4	523	109	7759	under
GO:0009057	macromolecule catabolic process	0	0,155964	0	136	113	8146	under
GO:0008610	lipid biosynthetic process	0	0,155964	0	136	113	8146	under
GO:0006091	generation of precursor metabolites and energy	0	0,162544	0	133	113	8149	under
GO:0060255	regulation of macromolecule metabolic process	0	0,16379	4	515	109	7767	under
GO:0009058	biosynthetic process	0	0,171548	10	1014	103	7268	under
GO:0044262	cellular carbohydrate metabolic process	0	0,171852	1	233	112	8049	under
GO:0080090	regulation of primary metabolic process	0	0,172639	4	508	109	7774	under
GO:0005996	monosaccharide metabolic process	0	0,174126	0	128	113	8154	under
GO:0044283	small molecule biosynthetic process	0	0,175827	2	326	111	7956	under
GO:0032991	macromolecular complex	0	0,178029	8	843	105	7439	under

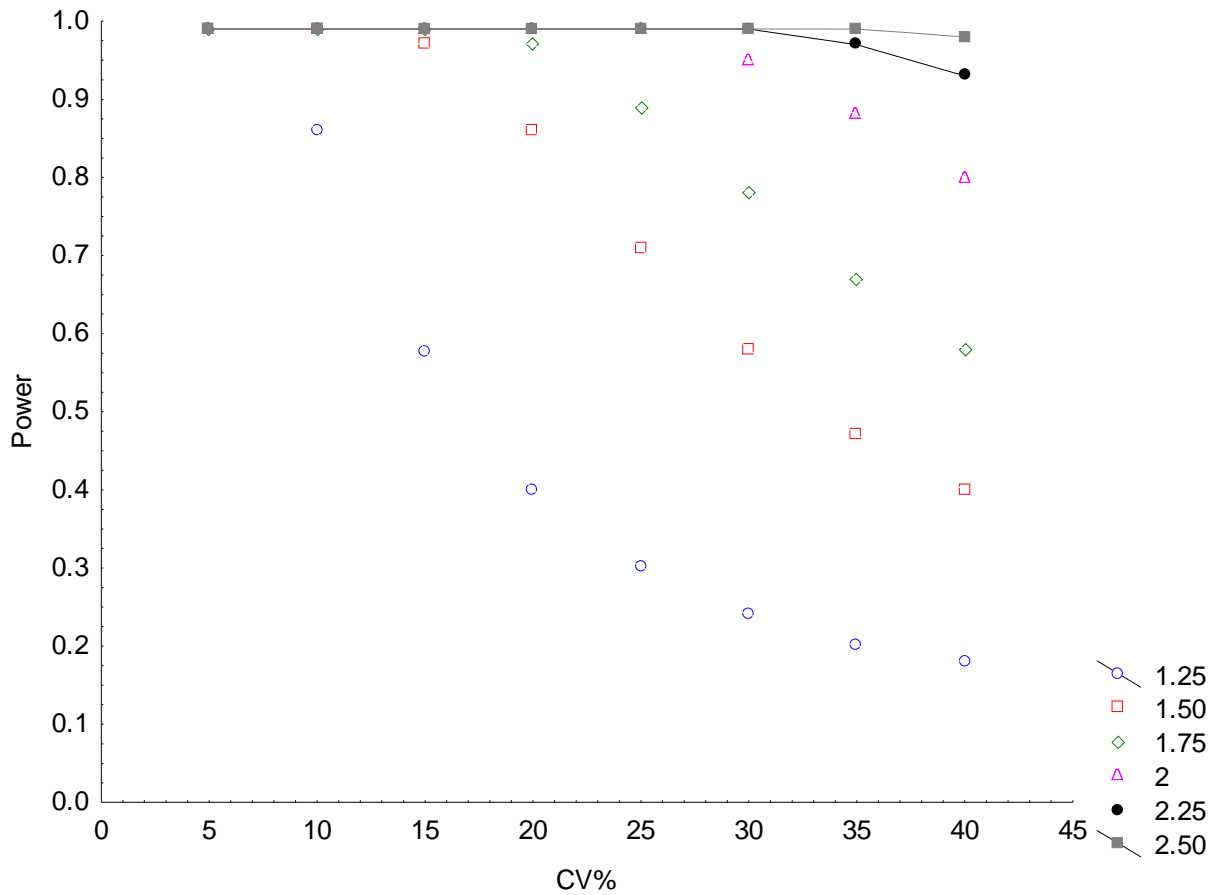
**Additional file 7****Reproducibility of Replicates for the Leaf Tissue in the Leaf-Flower Comparison.**

Coefficient of variation (CV%) for all cDNAs spotted on the array based on raw data mean fluorescence values plotted against the relative rank of the CV. CVs were estimated from raw data derived from six replicates (i.e. 2 dye  $\times$  3 biological replicates). For this tissue and for this experiment, 90% of the spots on the arrays could be determined with a CV of less than 35%. For the other experiments, 90% of the spots on the arrays could be determined with CV between 32 and 42%.

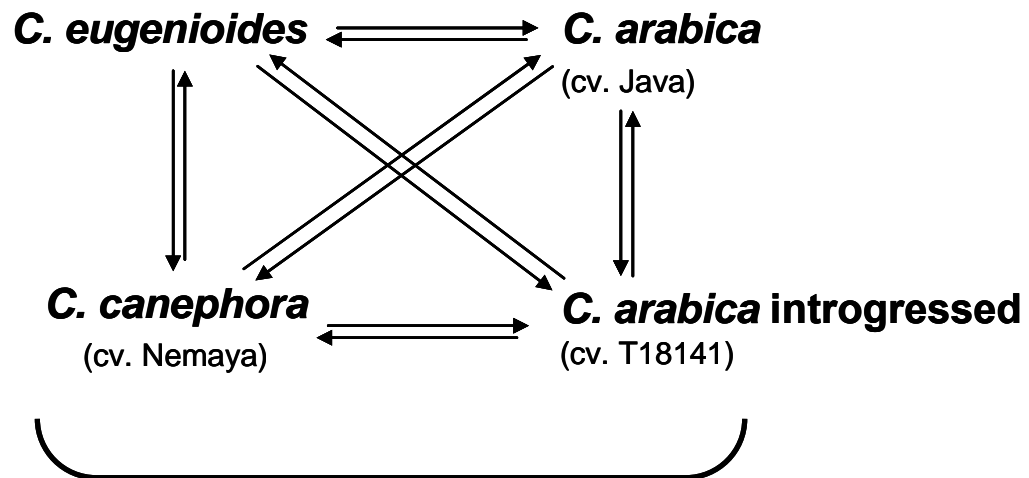


**Additional file 8**

**Theoretical Power Analysis for a Two-sample T-test to Detect a 1.25 to 2.50 Fold Change in Gene Expression as a Function of CV%.** Indicated sample size = 6 and a Type I error (false positive rate) of 0.1 were used as input values to determine the fraction of changes in gene expression that would be detected at a given CV%. A power of 1.0 denotes a Type II error (false negative rate) of zero, i.e. 100% of all changes that occurred were detected.

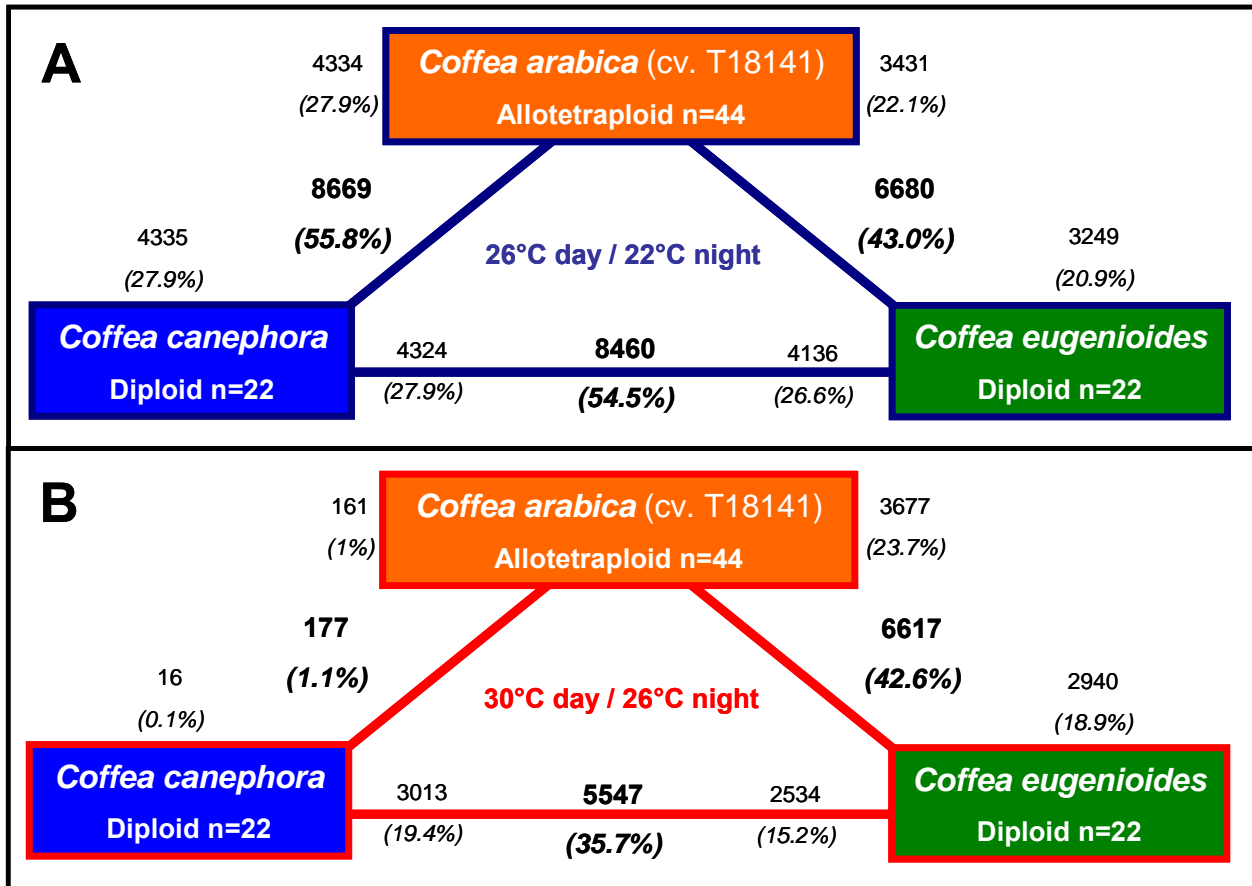






**6 comparisons \* 3 replicates \* 2 dye swaps \* 2 temperatures  
(= 72 slides)**

**Fig. S1 Microarray experimental design.**



**Fig. S2** Transcriptome divergence and nonadditive gene expression between *Coffea arabica* cv T18141 and parental diploid species.





## Positive dominance 30-26°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0015074	DNA integration	4,29E-04	1,07E-04	4,35E-06	0	76	1253	7066	under
GO:0009507	chloroplast	0,0108868	0,002718	1,67E-05	113	411	1140	6731	over
GO:0010319	stromule	0,0121194	0,00604139	3,87E-05	36	88	1217	7054	over
GO:0006259	DNA metabolic process	0,0134474	0,00670127	1,94E-04	18	227	1235	6915	under
GO:0003677	DNA binding	0,0185422	0,0138108	3,56E-04	64	554	1189	6588	under
GO:0044427	chromosomal part	0,0382165	0,0374982	9,01E-04	2	69	1251	7073	under
GO:0006310	DNA recombination	0,0384828	0,0469689	0,00106783	1	56	1252	7086	under
GO:0006325	chromatin organization	0,0399877	0,0582242	0,00129548	3	77	1250	7065	under

**Table S2 BLAST2GO annotation for genes differentially expressed in the ‘dominance’ and ‘transgression’**

## Negative dominance 30-26°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0003676	nucleic acid binding	7,69E-07	1,92E-07	2,48E-08	86	1081	919	6309	under
GO:0003677	DNA binding	2,36E-05	1,18E-05	8,92E-07	39	579	966	6811	under
GO:0006259	DNA metabolic process	6,68E-05	5,01E-05	3,49E-06	9	236	996	7154	under
GO:0090304	nucleic acid metabolic process	0,00156445	0,00156323	7,45E-05	30	424	975	6966	under
GO:0009987	cellular process	0,00291293	0,00363458	1,36E-04	298	2622	707	4768	under
GO:0044260	cellular macromolecule metabolic process	0,00317298	0,00474823	1,96E-04	135	1322	870	6068	under
GO:0044237	cellular metabolic process	0,00335554	0,00585508	2,24E-04	242	2173	763	5217	under
GO:0034061	DNA polymerase activity	0,00519679	0,0115642	4,04E-04	2	95	1003	7295	under
GO:0009059	macromolecule biosynthetic process	0,00519679	0,0116251	4,05E-04	32	416	973	6974	under
GO:0034645	cellular macromolecule biosynthetic process	0,00539534	0,0133983	4,29E-04	32	415	973	6975	under
GO:0006807	nitrogen compound metabolic process	0,00687815	0,0214802	6,57E-04	75	790	930	6600	under
GO:0015074	DNA integration	0,00687815	0,0223756	6,76E-04	1	75	1004	7315	under
GO:0019222	regulation of metabolic process	0,00687815	0,0240073	7,23E-04	41	489	964	6901	under
GO:0060255	regulation of macromolecule metabolic process	0,00687815	0,0242463	7,58E-04	40	479	965	6911	under
GO:0003964	RNA-directed DNA polymerase activity	0,00687815	0,0254648	7,78E-04	2	89	1003	7301	under
GO:0006278	RNA-dependent DNA replication	0,00713497	0,0281384	8,67E-04	2	88	1003	7302	under
GO:0080090	regulation of primary metabolic process	0,00829378	0,0355938	0,00109723	40	472	965	6918	under
GO:0034641	cellular nitrogen compound metabolic process	0,00829378	0,038531	0,00116805	72	751	933	6639	under
GO:0043170	macromolecule metabolic process	0,00829378	0,040312	0,00122178	167	1529	838	5861	under
GO:0003723	RNA binding	0,00829378	0,0406247	0,00122505	18	263	987	7127	under
GO:0005634	nucleus	0,0120678	0,0615875	0,00181733	43	489	962	6901	under
GO:0006260	DNA replication	0,0120678	0,0642276	0,00185545	4	107	1001	7283	under
GO:0044445	cytosolic part	0,0131084	0,0726147	0,00205623	5	118	1000	7272	under
GO:0016779	nucleotidyltransferase activity	0,0155051	0,0888514	0,00245221	8	150	997	7240	under
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,0162002	0,096314	0,00255868	48	526	957	6864	under
GO:0050789	regulation of biological process	0,017754	0,109016	0,00288615	58	610	947	6780	under
GO:0023052	signaling	0,0193477	0,12246	0,00330384	9	157	996	7233	under
GO:0023033	signaling pathway	0,0230203	0,148874	0,00385234	9	155	996	7235	under
GO:0000166	nucleotide binding	0,0242169	0,161074	0,00419191	123	1138	882	6252	under
GO:0031323	regulation of cellular metabolic process	0,0295917	0,19911	0,0051075	40	441	965	6949	under
GO:0010468	regulation of gene expression	0,0312633	0,215258	0,00540543	39	431	966	6959	under
GO:0022625	cytosolic large ribosomal subunit	0,0359943	0,268518	0,00670198	1	55	1004	7335	under
GO:0031326	regulation of cellular biosynthetic process	0,0359943	0,270305	0,00682524	39	426	966	6964	under
GO:0010556	regulation of macromolecule biosynthetic process	0,0359943	0,270305	0,00682524	39	426	966	6964	under
GO:0009889	regulation of biosynthetic process	0,0359943	0,270305	0,00682524	39	426	966	6964	under
GO:0016817	hydrolase activity, acting on acid anhydrides	0,0367119	0,2882	0,00692587	28	328	977	7062	under
GO:0051171	regulation of nitrogen compound metabolic process	0,0367119	0,300434	0,00722895	38	416	967	6974	under
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,0367119	0,300434	0,00722895	38	416	967	6974	under
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	0,0367119	0,301042	0,00729316	28	327	977	7063	under
GO:0016462	pyrophosphatase activity	0,0375229	0,316056	0,00767822	28	326	977	7064	under
GO:0044249	cellular biosynthetic process	0,0375229	0,319458	0,00774776	87	827	918	6563	under
GO:0032555	purine ribonucleotide binding	0,0381715	0,348055	0,00857997	87	824	918	6566	under
GO:0032553	ribonucleotide binding	0,0381715	0,348055	0,00857997	87	824	918	6566	under
GO:0017076	purine nucleotide binding	0,0381715	0,348055	0,00861072	94	881	911	6509	under
GO:0045449	regulation of transcription	0,0381715	0,349323	0,00868915	38	412	967	6978	under
GO:0051716	cellular response to stimulus	0,041552	0,380117	0,00922343	8	133	997	7257	under
GO:0007166	cell surface receptor linked signaling pathway	0,0478675	0,430468	0,0100464	0	36	1005	7354	under

Table S2 BLAST2GO annotation for genes differentially expressed in the ‘dominance’ and ‘transgression’

## Positive dominance 26-22°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0006810	transport	0,0349534	0,0309938	1,77E-04	147	577	1154	6517	over
GO:0051234	establishment of localization	0,0349534	0,0324541	1,89E-04	147	578	1154	6516	over
GO:0016192	vesicle-mediated transport	0,0349534	0,0329168	1,94E-04	29	67	1272	7027	over
GO:0016020	membrane	0,0349534	0,0409798	2,52E-04	214	905	1087	6189	over
GO:0043234	protein complex	0,0349534	0,0427518	2,73E-04	111	416	1190	6678	over
GO:0045271	respiratory chain complex I	0,0356923	0,0605526	3,74E-04	9	8	1292	7086	over
GO:0030964	NADH dehydrogenase complex	0,0356923	0,0605526	3,74E-04	9	8	1292	7086	over
GO:0048037	cofactor binding	0,0421319	0,0188092	4,76E-04	31	303	1270	6791	under
GO:0050662	coenzyme binding	0,0421319	0,0208465	5,21E-04	18	207	1283	6887	under
GO:0006897	endocytosis	0,0421537	0,0921598	4,64E-04	5	1	1296	7093	over
GO:0010324	membrane invagination	0,0421537	0,0921598	4,64E-04	5	1	1296	7093	over
GO:0044425	membrane part	0,0421537	0,100026	5,52E-04	150	609	1151	6485	over

## Negative dominance 26-22°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0070001	aspartic-type peptidase activity	0,00439411	0,00219466	6,42E-05	2	84	1304	7005	under
GO:0004190	aspartic-type endopeptidase activity	0,00439411	0,00219466	6,42E-05	2	84	1304	7005	under
GO:0016779	nucleotidyltransferase activity	0,0157768	0,0117631	3,10E-04	10	148	1296	6941	under
GO:0006396	RNA processing	0,0398171	0,0225378	1,11E-04	29	64	1277	7025	over
GO:0048519	negative regulation of biological process	0,0398171	0,045691	2,92E-04	14	20	1292	7069	over
GO:0000166	nucleotide binding	0,0398171	0,0543735	3,19E-04	238	1023	1068	6066	over
GO:0017076	purine nucleotide binding	0,0398171	0,0579294	3,63E-04	189	786	1117	6303	over
GO:0032553	ribonucleotide binding	0,0398171	0,0579786	3,65E-04	178	733	1128	6356	over
GO:0032555	purine ribonucleotide binding	0,0398171	0,0579786	3,65E-04	178	733	1128	6356	over

Table S2 BLAST2GO annotation for genes differentially expressed in the ‘dominance’ and ‘transgression’

## Positive transgressive 30-26°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0005488	binding	0,00645507	0,0096826	0,00188751	71	4358	100	3866	under
GO:0003677	DNA binding	0,00645507	0,0096826	0,00369976	4	614	167	7610	under
GO:0003676	nucleic acid binding	0,00645507	0,0096826	0,0147457	14	1153	157	7071	under
GO:0005886	plasma membrane	0,00645507	0,0096826	0,0173633	2	367	169	7857	under
GO:0008270	zinc ion binding	0,00645507	0,0096826	0,0231827	4	492	167	7732	under
GO:0016020	membrane	0,00645507	0,0096826	0,0244413	14	1105	157	7119	under
GO:0006259	DNA metabolic process	0,0162576	0,0281818	0,0374431	1	244	170	7980	under
GO:0031982	vesicle	0,028436	0,0755754	0,0414987	8	701	163	7523	under
GO:0031410	cytoplasmic vesicle	0,028436	0,0755754	0,0414987	8	701	163	7523	under
GO:0016023	cytoplasmic membrane-bounded vesicle	0,028436	0,0755754	0,041969	8	700	163	7524	under
GO:0031988	membrane-bounded vesicle	0,028436	0,0755754	0,041969	8	700	163	7524	under
GO:0044425	membrane part	0,0296128	0,085411	0,0467691	9	750	162	7474	under
GO:0005515	protein binding	0,0309703	0,0962168	0,0499864	14	1032	157	7192	under
GO:0006812	cation transport	0,0443439	0,154001	0,0547207	0	140	171	8084	under
GO:0022892	substrate-specific transporter activity	0,0443439	0,154001	0,0549665	2	295	169	7929	under

## Negative transgressive 30-26°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0016491	oxidoreductase activity	0	0	0,00116262	3	764	132	7496	under
GO:0044249	cellular biosynthetic process	0	0	0,00218063	5	909	130	7351	under
GO:0009058	biosynthetic process	0	0	0,00484939	7	1017	128	7243	under
GO:0005506	iron ion binding	0	0	0,0116853	0	270	135	7990	under
GO:0032991	macromolecular complex	0	0	0,0128094	6	845	129	7415	under
GO:0009055	electron carrier activity	0	0	0,0166051	0	249	135	8011	under
GO:0009987	cellular process	0	0	0,016846	35	2885	100	5375	under
GO:0044283	small molecule biosynthetic process	0	0	0,0290027	1	327	134	7933	under
GO:0008152	metabolic process	0	0	0,0367707	46	3474	89	4786	under
GO:0030529	ribonucleoprotein complex	0	0	0,0384465	1	307	134	7953	under
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0	0	0,0412234	4	570	131	7690	under
GO:0005198	structural molecule activity	0	0	0,0487115	1	290	134	7970	under
GO:0044237	cellular metabolic process	0	0	0,0524606	30	2385	105	5875	under
GO:0006807	nitrogen compound metabolic process	0	0	0,0537744	8	857	127	7403	under
GO:0003723	RNA binding	0	0	0,0559113	1	280	134	7980	under
GO:0046483	heterocycle metabolic process	0	0	0,0587775	0	173	135	8087	under
GO:0008324	cation transmembrane transporter activity	0	0	0,0638481	0	168	135	8092	under
GO:0046914	transition metal ion binding	0	0	0,0684257	8	829	127	7431	under
GO:0034641	cellular nitrogen compound metabolic process	0	0	0,076948	8	815	127	7445	under
GO:0044271	cellular nitrogen compound biosynthetic process	0	0	0,0817733	1	252	134	8008	under
GO:0046906	tetrapyrrole binding	0	0	0,104787	0	138	135	8122	under
GO:0006629	lipid metabolic process	0	0	0,108083	1	231	134	8029	under
GO:0008610	lipid biosynthetic process	0	0	0,108299	0	136	135	8124	under
GO:0055114	oxidation reduction	0	0	0,112117	4	472	131	7788	under
GO:0006412	translation	0,00910294	0,0560841	0,113905	1	227	134	8033	under
GO:0003735	structural constituent of ribosome	0,0135049	0,08524	0,124792	1	220	134	8040	under

Table S2 BLAST2GO annotation for genes differentially expressed in the ‘dominance’ and ‘transgression’

## Positive transgressive 26-22°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0031225	anchored to membrane	0,00744587	0,00185974	1,11E-05	13	23	770	7589	over

## Negative transgressive 26-22°C

go	name	FDR	FWER	p-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0044435	plastid part	2,79E-06	6,98E-07	6,38E-09	59	214	782	7340	over
GO:0010319	stromule	9,37E-04	4,69E-04	3,30E-06	30	94	811	7460	over
GO:0009507	chloroplast	0,0026556	0,00198972	1,13E-05	83	441	758	7113	over
GO:0044434	chloroplast part	0,00796845	0,00793682	5,20E-05	33	129	808	7425	over
GO:0003857	3-hydroxyacyl-CoA dehydrogenase activity	0,00803898	0,00999844	5,49E-05	5	1	836	7553	over
GO:0009536	plastid	0,00976931	0,0145472	8,79E-05	228	1608	613	5946	over
GO:0006805	xenobiotic metabolic process	0,0124419	0,0215383	1,00E-04	4	0	837	7554	over
GO:0016868	intramolecular transferase activity, phosphotransferases	0,0145248	0,0286321	1,46E-04	6	4	835	7550	over
GO:0016846	carbon-sulfur lyase activity	0,0181511	0,048691	2,71E-04	8	11	833	7543	over
GO:0016137	glycoside metabolic process	0,0181511	0,048691	2,71E-04	8	11	833	7543	over
GO:0005984	disaccharide metabolic process	0,0181511	0,048691	2,71E-04	8	11	833	7543	over
GO:0009526	plastid envelope	0,0182999	0,057742	3,14E-04	11	23	830	7531	over
GO:0009941	chloroplast envelope	0,0182999	0,057742	3,14E-04	11	23	830	7531	over
GO:0003824	catalytic activity	0,0233207	0,0783827	4,48E-04	434	3438	407	4116	over
GO:0045261	proton-transporting ATP synthase complex, catalytic core F(1)	0,0246294	0,0928043	4,60E-04	4	1	837	7553	over
GO:0006457	protein folding	0,0246294	0,0938237	4,77E-04	24	93	817	7461	over
GO:0009311	oligosaccharide metabolic process	0,035889	0,145728	8,72E-04	8	14	833	7540	over
GO:0044438	microbody part	0,035889	0,156751	8,90E-04	5	4	836	7550	over
GO:0044439	peroxisomal part	0,035889	0,156751	8,90E-04	5	4	836	7550	over
GO:0009620	response to fungus	0,035889	0,164277	9,18E-04	6	7	835	7547	over
GO:0019509	L-methionine salvage from methylthioadenosine	0,0405063	0,227857	0,00100214	3	0	838	7554	over
GO:0043102	amino acid salvage	0,0405063	0,227857	0,00100214	3	0	838	7554	over
GO:0005576	extracellular region	0,0405063	0,229427	0,00103228	28	124	813	7430	over
GO:0016853	isomerase activity	0,0405063	0,235361	0,00113669	24	100	817	7454	over
GO:0031967	organelle envelope	0,0405063	0,238109	0,0011519	11	28	830	7526	over
GO:0031975	envelope	0,0405063	0,238109	0,0011519	11	28	830	7526	over
GO:0006996	organelle organization	0,0405063	0,247856	0,00119929	13	38	828	7516	over
GO:0016860	intramolecular oxidoreductase activity	0,0405063	0,251205	0,00122115	8	15	833	7539	over
GO:0033178	proton-transporting two-sector ATPase complex, catalytic domain	0,0405063	0,284717	0,00127144	4	2	837	7552	over
GO:0043227	membrane-bounded organelle	0,0405063	0,287766	0,00133394	441	3544	400	4010	over
GO:0043231	intracellular membrane-bounded organelle	0,0405063	0,287766	0,00133394	441	3544	400	4010	over
GO:0051082	unfolded protein binding	0,0405063	0,292061	0,00139228	18	66	823	7488	over
GO:0019321	pentose metabolic process	0,0405063	0,309077	0,0014708	6	8	835	7546	over
GO:0005991	trehalose metabolic process	0,0405063	0,329643	0,00163448	5	5	836	7549	over
GO:0046351	disaccharide biosynthetic process	0,0405063	0,329643	0,00163448	5	5	836	7549	over
GO:0016138	glycoside biosynthetic process	0,0405063	0,329643	0,00163448	5	5	836	7549	over
GO:0048285	organelle fission	0,0405063	0,329643	0,00163448	5	5	836	7549	over
GO:0005992	trehalose biosynthetic process	0,0405063	0,329643	0,00163448	5	5	836	7549	over
GO:0000097	sulfur amino acid biosynthetic process	0,0405063	0,334098	0,00167304	8	16	833	7538	over
GO:0048046	apoplast	0,0405063	0,350544	0,00186752	22	92	819	7462	over
GO:0070279	vitamin B6 binding	0,0405063	0,351217	0,00188068	19	74	822	7480	over
GO:0030170	pyridoxal phosphate binding	0,0405063	0,351217	0,00188068	19	74	822	7480	over
GO:0016491	oxidoreductase activity	0,0405063	0,353062	0,00192485	101	666	740	6888	over
GO:0042579	microbody	0,0412617	0,369476	0,00215627	15	52	826	7502	over
GO:0043229	intracellular organelle	0,0412617	0,384173	0,00224655	449	3638	392	3916	over
GO:0043226	organelle	0,0412617	0,384173	0,00224655	449	3638	392	3916	over
GO:0044444	cytoplasmic part	0,0412617	0,384245	0,00224752	450	3647	391	3907	over
GO:0000096	sulfur amino acid metabolic process	0,0436099	0,407506	0,00257487	9	22	832	7532	over
GO:0006635	fatty acid beta-oxidation	0,0483965	0,46227	0,00273164	4	3	837	7551	over
GO:0034440	lipid oxidation	0,0483965	0,46227	0,00273164	4	3	837	7551	over
GO:0019395	fatty acid oxidation	0,0483965	0,46227	0,00273164	4	3	837	7551	over
GO:0004372	glycine hydroxymethyltransferase activity	0,0483965	0,479778	0,00275095	5	6	836	7548	over
GO:0009514	glyoxysome	0,0483965	0,479778	0,00275095	5	6	836	7548	over
GO:0016861	intramolecular oxidoreductase activity, interconverting aldoses and ketoses	0,0483965	0,479778	0,00275095	5	6	836	7548	over
GO:0006519	cellular amino acid and derivative metabolic process	0,0491004	0,493139	0,00299331	43	238	798	7316	over
GO:0009579	thylakoid	0,0491004	0,497533	0,00310454	17	66	824	7488	over
GO:0044281	small molecule metabolic process	0,0491004	0,503341	0,00318831	90	592	751	6962	over

SGN Number  
SGN-U359404  
SGN-U353149  
SGN-U359186  
SGN-U359155  
SGN-U359152  
SGN-U359114  
SGN-U359096  
SGN-U358918  
SGN-U358888  
SGN-U358880  
SGN-U358790  
SGN-U358758  
SGN-U358519  
SGN-U358452  
SGN-U358434  
SGN-U352281  
SGN-U358415  
SGN-U358406  
SGN-U358341  
SGN-U358288  
SGN-U358228  
SGN-U358209  
SGN-U352065  
SGN-U358140  
SGN-U351978  
SGN-U351957  
SGN-U358097  
SGN-U358053  
SGN-U358044  
SGN-U358012  
SGN-U357998  
SGN-U357876  
SGN-U351731  
SGN-U357868  
SGN-U351609  
SGN-U357742  
SGN-U357737  
SGN-U357736  
SGN-U357698  
SGN-U357691  
SGN-U357680  
SGN-U357671  
SGN-U357658  
SGN-U357587  
SGN-U357393  
SGN-U357373  
SGN-U351225  
SGN-U357255  
SGN-U357240

**Table S3 The missing genes**

SGN-U357222  
SGN-U351033  
SGN-U357025  
SGN-U356995  
SGN-U350843  
SGN-U356938  
SGN-U356907  
SGN-U356886  
SGN-U356760  
SGN-U356700  
SGN-U356695  
SGN-U362832  
SGN-U362800  
SGN-U350506  
SGN-U362789  
SGN-U362780  
SGN-U356628  
SGN-U362738  
SGN-U362728  
SGN-U356569  
SGN-U350392  
SGN-U356529  
SGN-U356524  
SGN-U362656  
SGN-U362652  
SGN-U362637  
SGN-U362622  
SGN-U356462  
SGN-U362592  
SGN-U356439  
SGN-U362561  
SGN-U356411  
SGN-U350254  
SGN-U356396  
SGN-U356389  
SGN-U362516  
SGN-U356351  
SGN-U356328  
SGN-U362468  
SGN-U356324  
SGN-U356286  
SGN-U362422  
SGN-U356248  
SGN-U356246  
SGN-U356234  
SGN-U356217  
SGN-U350068  
SGN-U362328  
SGN-U356137  
SGN-U362279

SGN-U356114  
SGN-U362256  
SGN-U362228  
SGN-U362175  
SGN-U349834  
SGN-U355954  
SGN-U362057  
SGN-U349764  
SGN-U349724  
SGN-U361994  
SGN-U355845  
SGN-U361954  
SGN-U355808  
SGN-U361836  
SGN-U361831  
SGN-U361825  
SGN-U361792  
SGN-U361788  
SGN-U361786  
SGN-U361781  
SGN-U355627  
SGN-U361760  
SGN-U361711  
SGN-U361683  
SGN-U361615  
SGN-U361604  
SGN-U361560  
SGN-U361476  
SGN-U361475  
SGN-U361458  
SGN-U355305  
SGN-U355300  
SGN-U361429  
SGN-U349122  
SGN-U349114  
SGN-U361397  
SGN-U361337  
SGN-U361327  
SGN-U355183  
SGN-U349014  
SGN-U361285  
SGN-U361273  
SGN-U361229  
SGN-U361207  
SGN-U355015  
SGN-U348871  
SGN-U361134  
SGN-U361094  
SGN-U361067  
SGN-U361060



SGN-U348718  
SGN-U354841  
SGN-U360976  
SGN-U354804  
SGN-U360936  
SGN-U360898  
SGN-U360896  
SGN-U360858  
SGN-U360850  
SGN-U360823  
SGN-U354677  
SGN-U354670  
SGN-U360787  
SGN-U354640  
SGN-U360741  
SGN-U360723  
SGN-U354576  
SGN-U360686  
SGN-U354422  
SGN-U354358  
SGN-U354334  
SGN-U360477  
SGN-U360475  
SGN-U360468  
SGN-U354306  
SGN-U354305  
SGN-U360431  
SGN-U347984  
SGN-U360248  
SGN-U360232  
SGN-U360200  
SGN-U353989  
SGN-U360096  
SGN-U353884  
SGN-U359857  
SGN-U353632  
SGN-U353568  
SGN-U353567  
SGN-U359704  
SGN-U353550  
SGN-U353518  
SGN-U359634  
SGN-U353444  
SGN-U353375  
SGN-U353286  
SGN-U359428

Oligonucleotide	Unigene
Cof::opCocaV10015513	SGN-U356510
Cof::opCocaV10015511	SGN-U347161
Cof::opCocaV10015451	SGN-U355453
Cof::opCocaV10015421	SGN-U356181
Cof::opCocaV10015414	SGN-U361311
Cof::opCocaV10015390	SGN-U361983
Cof::opCocaV10015386	SGN-U356573
Cof::opCocaV10015256	SGN-U351191
Cof::opCocaV10015232	SGN-U356863
Cof::opCocaV10015177	SGN-U357764
Cof::opCocaV10015115	SGN-U348189
Cof::opCocaV10015091	SGN-U349720
Cof::opCocaV10014980	SGN-U356136
Cof::opCocaV10014913	SGN-U362019
Cof::opCocaV10014874	SGN-U361725
Cof::opCocaV10014868	SGN-U356226
Cof::opCocaV10014866	SGN-U353457
Cof::opCocaV10014858	SGN-U358775
Cof::opCocaV10014832	SGN-U349544
Cof::opCocaV10014831	SGN-U348401
Cof::opCocaV10014819	SGN-U354028
Cof::opCocaV10014799	SGN-U358661
Cof::opCocaV10014770	SGN-U349087
Cof::opCocaV10014736	SGN-U357341
Cof::opCocaV10014607	SGN-U350427
Cof::opCocaV10014477	SGN-U349855
Cof::opCocaV10014404	SGN-U352949
Cof::opCocaV10014380	SGN-U358507
Cof::opCocaV10014340	SGN-U352985
Cof::opCocaV10014250	SGN-U360229
Cof::opCocaV10014236	SGN-U359875
Cof::opCocaV10014217	SGN-U357810
Cof::opCocaV10014207	SGN-U349583
Cof::opCocaV10014195	SGN-U354407
Cof::opCocaV10014152	SGN-U351253
Cof::opCocaV10014100	SGN-U354864
Cof::opCocaV10014025	SGN-U351733
Cof::opCocaV10013957	SGN-U348100
Cof::opCocaV10013953	SGN-U359747
Cof::opCocaV10013918	SGN-U357952

**Table S4 Oligonucleotides that may cross-hybridize with several sequences**

Cof::opCocaV10013900	SGN-U356839
Cof::opCocaV10013795	SGN-U356370
Cof::opCocaV10013711	SGN-U359987
Cof::opCocaV10013671	SGN-U353587
Cof::opCocaV10013656	SGN-U359869
Cof::opCocaV10013649	SGN-U352247
Cof::opCocaV10013565	SGN-U360453
Cof::opCocaV10013510	SGN-U355562
Cof::opCocaV10013455	SGN-U361840
Cof::opCocaV10013440	SGN-U358639
Cof::opCocaV10013406	SGN-U353294
Cof::opCocaV10013402	SGN-U350075
Cof::opCocaV10013398	SGN-U347400
Cof::opCocaV10013333	SGN-U357641
Cof::opCocaV10013283	SGN-U360902
Cof::opCocaV10013252	SGN-U348461
Cof::opCocaV10013201	SGN-U361169
Cof::opCocaV10013174	SGN-U358858
Cof::opCocaV10013168	SGN-U352653
Cof::opCocaV10013133	SGN-U358420
Cof::opCocaV10013104	SGN-U352208
Cof::opCocaV10013003	SGN-U353792
Cof::opCocaV10012981	SGN-U360728
Cof::opCocaV10012932	SGN-U358613
Cof::opCocaV10012910	SGN-U358691
Cof::opCocaV10012801	SGN-U360935
Cof::opCocaV10012797	SGN-U353307
Cof::opCocaV10012781	SGN-U356313
Cof::opCocaV10012721	SGN-U358064
Cof::opCocaV10012575	SGN-U356228
Cof::opCocaV10012513	SGN-U357629
Cof::opCocaV10012504	SGN-U358449
Cof::opCocaV10012497	SGN-U355002
Cof::opCocaV10012430	SGN-U358893
Cof::opCocaV10012374	SGN-U351940
Cof::opCocaV10012348	SGN-U353422
Cof::opCocaV10012329	SGN-U347354
Cof::opCocaV10012267	SGN-U361511
Cof::opCocaV10012263	SGN-U355945
Cof::opCocaV10012261	SGN-U356571
Cof::opCocaV10012226	SGN-U360587

Cof::opCocaV10012191	SGN-U361883
Cof::opCocaV10012157	SGN-U357422
Cof::opCocaV10012098	SGN-U347783
Cof::opCocaV10012060	SGN-U358703
Cof::opCocaV10012019	SGN-U360171
Cof::opCocaV10011985	SGN-U353076
Cof::opCocaV10011808	SGN-U355890
Cof::opCocaV10011789	SGN-U358087
Cof::opCocaV10011782	SGN-U357435
Cof::opCocaV10011727	SGN-U360661
Cof::opCocaV10011711	SGN-U360044
Cof::opCocaV10011640	SGN-U356926
Cof::opCocaV10011637	SGN-U359877
Cof::opCocaV10011590	SGN-U351454
Cof::opCocaV10011513	SGN-U358673
Cof::opCocaV10011466	SGN-U357152
Cof::opCocaV10011454	SGN-U348230
Cof::opCocaV10011442	SGN-U353500
Cof::opCocaV10011398	SGN-U358425
Cof::opCocaV10011388	SGN-U354872
Cof::opCocaV10011346	SGN-U350560
Cof::opCocaV10011306	SGN-U359744
Cof::opCocaV10011287	SGN-U359970
Cof::opCocaV10011244	SGN-U360559
Cof::opCocaV10011217	SGN-U353073
Cof::opCocaV10011212	SGN-U360025
Cof::opCocaV10011156	SGN-U356227
Cof::opCocaV10011092	SGN-U357182
Cof::opCocaV10011044	SGN-U355821
Cof::opCocaV10011024	SGN-U357271
Cof::opCocaV10010887	SGN-U358066
Cof::opCocaV10010881	SGN-U347969
Cof::opCocaV10010829	SGN-U356754
Cof::opCocaV10010822	SGN-U353493
Cof::opCocaV10010816	SGN-U348430
Cof::opCocaV10010776	SGN-U358311
Cof::opCocaV10010750	SGN-U352977
Cof::opCocaV10010740	SGN-U348991
Cof::opCocaV10010618	SGN-U355099
Cof::opCocaV10010611	SGN-U356810
Cof::opCocaV10010562	SGN-U350158

Cof::opCocaV10010491	SGN-U358603
Cof::opCocaV10010489	SGN-U349677
Cof::opCocaV10010483	SGN-U349705
Cof::opCocaV10010420	SGN-U360175
Cof::opCocaV10010398	SGN-U359833
Cof::opCocaV10010364	SGN-U350569
Cof::opCocaV10010351	SGN-U358185
Cof::opCocaV10010340	SGN-U356588
Cof::opCocaV10010334	SGN-U352492
Cof::opCocaV10010324	SGN-U358450
Cof::opCocaV10010311	SGN-U360706
Cof::opCocaV10010301	SGN-U357615
Cof::opCocaV10010284	SGN-U354462
Cof::opCocaV10010264	SGN-U357505
Cof::opCocaV10010263	SGN-U349961
Cof::opCocaV10010256	SGN-U353289
Cof::opCocaV10010121	SGN-U357030
Cof::opCocaV10010119	SGN-U349818
Cof::opCocaV10010102	SGN-U353386
Cof::opCocaV10010089	SGN-U350550
Cof::opCocaV10009982	SGN-U351923
Cof::opCocaV10009952	SGN-U352456
Cof::opCocaV10009931	SGN-U356603
Cof::opCocaV10009921	SGN-U348794
Cof::opCocaV10009774	SGN-U354751
Cof::opCocaV10009722	SGN-U347128
Cof::opCocaV10009710	SGN-U352947
Cof::opCocaV10009690	SGN-U352866
Cof::opCocaV10009664	SGN-U349792
Cof::opCocaV10009604	SGN-U356260
Cof::opCocaV10009547	SGN-U359911
Cof::opCocaV10009349	SGN-U358766
Cof::opCocaV10009294	SGN-U357630
Cof::opCocaV10009281	SGN-U358823
Cof::opCocaV10009258	SGN-U354266
Cof::opCocaV10009256	SGN-U358429
Cof::opCocaV10009183	SGN-U357715
Cof::opCocaV10009166	SGN-U362186
Cof::opCocaV10009161	SGN-U358121
Cof::opCocaV10009125	SGN-U357882
Cof::opCocaV10009117	SGN-U361741

Cof::opCocaV10009099	SGN-U352739
Cof::opCocaV10009077	SGN-U353591
Cof::opCocaV10009031	SGN-U353688
Cof::opCocaV10008885	SGN-U356188
Cof::opCocaV10008883	SGN-U356194
Cof::opCocaV10008728	SGN-U358055
Cof::opCocaV10008692	SGN-U352270
Cof::opCocaV10008660	SGN-U358455
Cof::opCocaV10008649	SGN-U357204
Cof::opCocaV10008617	SGN-U360355
Cof::opCocaV10008535	SGN-U362506
Cof::opCocaV10008326	SGN-U350575
Cof::opCocaV10008281	SGN-U362507
Cof::opCocaV10008273	SGN-U351126
Cof::opCocaV10008242	SGN-U349563
Cof::opCocaV10008221	SGN-U361618
Cof::opCocaV10008178	SGN-U360471
Cof::opCocaV10008076	SGN-U357466
Cof::opCocaV10008006	SGN-U357750
Cof::opCocaV10008000	SGN-U360361
Cof::opCocaV10007978	SGN-U360380
Cof::opCocaV10007934	SGN-U347831
Cof::opCocaV10007894	SGN-U356912
Cof::opCocaV10007877	SGN-U357191
Cof::opCocaV10007808	SGN-U352760
Cof::opCocaV10007800	SGN-U360110
Cof::opCocaV10007764	SGN-U356677
Cof::opCocaV10007684	SGN-U351075
Cof::opCocaV10007678	SGN-U357960
Cof::opCocaV10007670	SGN-U357007
Cof::opCocaV10007664	SGN-U359234
Cof::opCocaV10007587	SGN-U354049
Cof::opCocaV10007525	SGN-U351195
Cof::opCocaV10007521	SGN-U357561
Cof::opCocaV10007449	SGN-U357291
Cof::opCocaV10007430	SGN-U360795
Cof::opCocaV10007400	SGN-U356338
Cof::opCocaV10007398	SGN-U356040
Cof::opCocaV10007382	SGN-U357533
Cof::opCocaV10007352	SGN-U350347
Cof::opCocaV10007276	SGN-U354593

Cof::opCocaV10007252	SGN-U362174
Cof::opCocaV10007237	SGN-U354805
Cof::opCocaV10007212	SGN-U359867
Cof::opCocaV10007198	SGN-U356619
Cof::opCocaV10007080	SGN-U357940
Cof::opCocaV10007069	SGN-U353198
Cof::opCocaV10006984	SGN-U359021
Cof::opCocaV10006956	SGN-U353748
Cof::opCocaV10006935	SGN-U356184
Cof::opCocaV10006924	SGN-U354553
Cof::opCocaV10006846	SGN-U357394
Cof::opCocaV10006732	SGN-U361061
Cof::opCocaV10006712	SGN-U358366
Cof::opCocaV10006642	SGN-U357844
Cof::opCocaV10006622	SGN-U358409
Cof::opCocaV10006606	SGN-U362694
Cof::opCocaV10006604	SGN-U358811
Cof::opCocaV10006574	SGN-U357716
Cof::opCocaV10006541	SGN-U357058
Cof::opCocaV10006533	SGN-U359095
Cof::opCocaV10006479	SGN-U353364
Cof::opCocaV10006411	SGN-U347896
Cof::opCocaV10006396	SGN-U350609
Cof::opCocaV10006392	SGN-U357160
Cof::opCocaV10006375	SGN-U358715
Cof::opCocaV10006374	SGN-U355198
Cof::opCocaV10006333	SGN-U358352
Cof::opCocaV10006307	SGN-U353716
Cof::opCocaV10006229	SGN-U357247
Cof::opCocaV10006206	SGN-U347946
Cof::opCocaV10006142	SGN-U348835
Cof::opCocaV10006048	SGN-U362830
Cof::opCocaV10006028	SGN-U351718
Cof::opCocaV10005983	SGN-U357129
Cof::opCocaV10005970	SGN-U361064
Cof::opCocaV10005878	SGN-U360124
Cof::opCocaV10005717	SGN-U362378
Cof::opCocaV10005639	SGN-U361265
Cof::opCocaV10005632	SGN-U361538
Cof::opCocaV10005597	SGN-U361033
Cof::opCocaV10005585	SGN-U355113

Cof::opCocaV10005575	SGN-U358818
Cof::opCocaV10005553	SGN-U358159
Cof::opCocaV10005501	SGN-U361215
Cof::opCocaV10005470	SGN-U352782
Cof::opCocaV10005468	SGN-U362133
Cof::opCocaV10005464	SGN-U360351
Cof::opCocaV10005422	SGN-U347211
Cof::opCocaV10005411	SGN-U356598
Cof::opCocaV10005344	SGN-U356970
Cof::opCocaV10005343	SGN-U355589
Cof::opCocaV10005330	SGN-U356538
Cof::opCocaV10005283	SGN-U357056
Cof::opCocaV10005249	SGN-U362323
Cof::opCocaV10005241	SGN-U356770
Cof::opCocaV10005139	SGN-U353784
Cof::opCocaV10005078	SGN-U354001
Cof::opCocaV10005074	SGN-U361179
Cof::opCocaV10004924	SGN-U359850
Cof::opCocaV10004917	SGN-U352557
Cof::opCocaV10004874	SGN-U356487
Cof::opCocaV10004841	SGN-U358713
Cof::opCocaV10004748	SGN-U352224
Cof::opCocaV10004736	SGN-U356304
Cof::opCocaV10004723	SGN-U357120
Cof::opCocaV10004685	SGN-U356681
Cof::opCocaV10004674	SGN-U352971
Cof::opCocaV10004653	SGN-U358173
Cof::opCocaV10004640	SGN-U357382
Cof::opCocaV10004606	SGN-U349520
Cof::opCocaV10004494	SGN-U358821
Cof::opCocaV10004480	SGN-U358927
Cof::opCocaV10004356	SGN-U360327
Cof::opCocaV10004345	SGN-U348454
Cof::opCocaV10004335	SGN-U361084
Cof::opCocaV10004313	SGN-U357625
Cof::opCocaV10004260	SGN-U350608
Cof::opCocaV10004176	SGN-U359712
Cof::opCocaV10004173	SGN-U362518
Cof::opCocaV10004159	SGN-U355842
Cof::opCocaV10004140	SGN-U362400
Cof::opCocaV10004038	SGN-U349710



Cof::opCocaV10003990	SGN-U356971
Cof::opCocaV10003985	SGN-U353635
Cof::opCocaV10003977	SGN-U357567
Cof::opCocaV10003901	SGN-U358379
Cof::opCocaV10003891	SGN-U358227
Cof::opCocaV10003697	SGN-U362108
Cof::opCocaV10003639	SGN-U361455
Cof::opCocaV10003551	SGN-U361105
Cof::opCocaV10003486	SGN-U351159
Cof::opCocaV10003464	SGN-U360911
Cof::opCocaV10003462	SGN-U351291
Cof::opCocaV10003446	SGN-U361097
Cof::opCocaV10003372	SGN-U353561
Cof::opCocaV10003313	SGN-U362159
Cof::opCocaV10003296	SGN-U348439
Cof::opCocaV10003160	SGN-U350886
Cof::opCocaV10003156	SGN-U360205
Cof::opCocaV10003076	SGN-U347605
Cof::opCocaV10003075	SGN-U359902
Cof::opCocaV10003036	SGN-U361891
Cof::opCocaV10002997	SGN-U357609
Cof::opCocaV10002949	SGN-U356626
Cof::opCocaV10002850	SGN-U356583
Cof::opCocaV10002832	SGN-U359072
Cof::opCocaV10002668	SGN-U362402
Cof::opCocaV10002667	SGN-U360072
Cof::opCocaV10002663	SGN-U357463
Cof::opCocaV10002606	SGN-U362251
Cof::opCocaV10002595	SGN-U356026
Cof::opCocaV10002571	SGN-U360539
Cof::opCocaV10002545	SGN-U362692
Cof::opCocaV10002397	SGN-U348160
Cof::opCocaV10002280	SGN-U360775
Cof::opCocaV10002241	SGN-U361916
Cof::opCocaV10002239	SGN-U354502
Cof::opCocaV10002233	SGN-U355913
Cof::opCocaV10002214	SGN-U357762
Cof::opCocaV10002201	SGN-U358161
Cof::opCocaV10002192	SGN-U355538
Cof::opCocaV10002158	SGN-U350127
Cof::opCocaV10001921	SGN-U353486

Cof::opCocaV10001701	SGN-U360448
Cof::opCocaV10001688	SGN-U356355
Cof::opCocaV10001577	SGN-U353410
Cof::opCocaV10001440	SGN-U358143
Cof::opCocaV10001430	SGN-U362049
Cof::opCocaV10001409	SGN-U362650
Cof::opCocaV10001406	SGN-U353912
Cof::opCocaV10001401	SGN-U361005
Cof::opCocaV10001390	SGN-U353727
Cof::opCocaV10001347	SGN-U352923
Cof::opCocaV10001303	SGN-U361434
Cof::opCocaV10001298	SGN-U362726
Cof::opCocaV10001283	SGN-U356939
Cof::opCocaV10001282	SGN-U349028
Cof::opCocaV10001267	SGN-U360407
Cof::opCocaV10001221	SGN-U355523
Cof::opCocaV10001220	SGN-U360237
Cof::opCocaV10001193	SGN-U351324
Cof::opCocaV10001181	SGN-U352230
Cof::opCocaV10001176	SGN-U357044
Cof::opCocaV10001116	SGN-U362824
Cof::opCocaV10000912	SGN-U357915
Cof::opCocaV10000895	SGN-U354579
Cof::opCocaV10000821	SGN-U358384
Cof::opCocaV10000795	SGN-U354464
Cof::opCocaV10000789	SGN-U352328
Cof::opCocaV10000771	SGN-U361071
Cof::opCocaV10000759	SGN-U350183
Cof::opCocaV10000756	SGN-U361395
Cof::opCocaV10000654	SGN-U360159
Cof::opCocaV10000644	SGN-U358909
Cof::opCocaV10000566	SGN-U361390
Cof::opCocaV10000460	SGN-U355208
Cof::opCocaV10000448	SGN-U354288
Cof::opCocaV10000424	SGN-U362476
Cof::opCocaV10000409	SGN-U357442
Cof::opCocaV10000408	SGN-U361842
Cof::opCocaV10000389	SGN-U362548
Cof::opCocaV10000306	SGN-U358588
Cof::opCocaV10000181	SGN-U356557
Cof::opCocaV10000154	SGN-U357335

Cof::opCocaV10000139	SGN-U359098
Cof::opCocaV10000106	SGN-U361974
Cof::opCocaV10000017	SGN-U358616

## Annexe 16 : Analyse d'enrichissement à l'aide du logiciel Blast 2Go (enrichment of Gene Ontology (GO) terms)

Fisher's Exact Test			
Test-Set: only in can3026.txt			
Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).			
GO Term	Name	Type	Over/Under
<a href="#">GO:0009505</a>	plant-type cell wall	C	over
<a href="#">GO:0016851</a>	magnesium chelatase activity	F	over
<a href="#">GO:0051002</a>	ligase activity, forming nitrogen-metal bonds	F	over
<a href="#">GO:0051003</a>	ligase activity, forming nitrogen-metal bonds, forming coordination complexes	F	over
<a href="#">GO:0015995</a>	chlorophyll biosynthetic process	P	over
<a href="#">GO:0015994</a>	chlorophyll metabolic process	P	over
<a href="#">GO:0015926</a>	glucosidase activity	F	over
<a href="#">GO:0051188</a>	cofactor biosynthetic process	P	over
<a href="#">GO:0006779</a>	porphyrin biosynthetic process	P	over
<a href="#">GO:0006778</a>	porphyrin metabolic process	P	over
<a href="#">GO:0033014</a>	tetrapyrrole biosynthetic process	P	over
<a href="#">GO:0033013</a>	tetrapyrrole metabolic process	P	over
<a href="#">GO:0044271</a>	cellular nitrogen compound biosynthetic process	P	over
<a href="#">GO:0046148</a>	pigment biosynthetic process	P	over
<a href="#">GO:0042440</a>	pigment metabolic process	P	over
<a href="#">GO:0005618</a>	cell wall	C	over
<a href="#">GO:0004558</a>	alpha-glucosidase activity	F	over
<a href="#">GO:0030312</a>	external encapsulating structure	C	over
<a href="#">GO:0009058</a>	biosynthetic process	P	over
<a href="#">GO:0009374</a>	biotin binding	F	over
<a href="#">GO:0051186</a>	cofactor metabolic process	P	over
<a href="#">GO:0000175</a>	3'-5'-exoribonuclease activity	F	over
<a href="#">GO:0003989</a>	acetyl-CoA carboxylase activity	F	over
<a href="#">GO:0016885</a>	ligase activity, forming carbon-carbon bonds	F	over
<a href="#">GO:0016421</a>	CoA carboxylase activity	F	over
<a href="#">GO:0004532</a>	exoribonuclease activity	F	over
<a href="#">GO:0016896</a>	exoribonuclease activity, producing 5'-phosphomonoesters	F	over
<a href="#">GO:0016796</a>	exonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters	F	over
<a href="#">GO:0046483</a>	heterocycle metabolic process	P	over
<a href="#">GO:0044249</a>	cellular biosynthetic process	P	over
<a href="#">GO:0018130</a>	heterocycle biosynthetic process	P	over
<a href="#">GO:0030244</a>	cellulose biosynthetic process	P	over
<a href="#">GO:0033293</a>	monocarboxylic acid binding	F	over
<a href="#">GO:0016053</a>	organic acid biosynthetic process	P	over
<a href="#">GO:0046394</a>	carboxylic acid biosynthetic process	P	over
<a href="#">GO:0008408</a>	3'-5' exonuclease activity	F	over
<a href="#">GO:0030243</a>	cellulose metabolic process	P	over
<a href="#">GO:0016023</a>	cytoplasmic membrane-bounded vesicle	C	over
<a href="#">GO:0031988</a>	membrane-bounded vesicle	C	over
<a href="#">GO:0031982</a>	vesicle	C	over
<a href="#">GO:0031410</a>	cytoplasmic vesicle	C	over
<a href="#">GO:0009073</a>	aromatic amino acid family biosynthetic process	P	over
<a href="#">GO:0046417</a>	chorismate metabolic process	P	over
<a href="#">GO:0016874</a>	ligase activity	F	over
<a href="#">GO:0004127</a>	cytidylate kinase activity	F	over
<a href="#">GO:0004849</a>	uridine kinase activity	F	over
<a href="#">GO:0004648</a>	O-phospho-L-serine:2-oxoglutarate aminotransferase activity	F	over
<a href="#">GO:0019438</a>	aromatic compound biosynthetic process	P	over
<a href="#">GO:0004527</a>	exonuclease activity	F	over
<a href="#">GO:0005975</a>	carbohydrate metabolic process	P	over
<a href="#">GO:0003866</a>	3-phosphoshikimate 1-carboxyvinyltransferase activity	F	over
<a href="#">GO:0009129</a>	pyrimidine nucleoside monophosphate metabolic process	P	over
<a href="#">GO:0009173</a>	pyrimidine ribonucleoside monophosphate metabolic process	P	over
<a href="#">GO:0030054</a>	cell junction	C	over
<a href="#">GO:0008883</a>	glutamyl-tRNA reductase activity	F	over
<a href="#">GO:0004764</a>	shikimate 5-dehydrogenase activity	F	over
<a href="#">GO:0009041</a>	uridylate kinase activity	F	over
<a href="#">GO:0009250</a>	glucan biosynthetic process	P	over
<a href="#">GO:0009072</a>	aromatic amino acid family metabolic process	P	over
<a href="#">GO:0004540</a>	ribonuclease activity	F	over
<a href="#">GO:0043648</a>	dicarboxylic acid metabolic process	P	over
<a href="#">GO:0008152</a>	metabolic process	P	over
<a href="#">GO:0050535</a>	beta-primeverosidase activity	F	over
<a href="#">GO:0044281</a>	small molecule metabolic process	P	over
<a href="#">GO:0044283</a>	small molecule biosynthetic process	P	over
<a href="#">GO:0033692</a>	cellular polysaccharide biosynthetic process	P	over
<a href="#">GO:0016051</a>	carbohydrate biosynthetic process	P	over
<a href="#">GO:0005507</a>	copper ion binding	F	over
<a href="#">GO:0000271</a>	polysaccharide biosynthetic process	P	over
<a href="#">GO:0008652</a>	cellular amino acid biosynthetic process	P	over
<a href="#">GO:0006725</a>	cellular aromatic compound metabolic process	P	over
<a href="#">GO:0019206</a>	nucleoside kinase activity	F	over
<a href="#">GO:0004765</a>	shikimate kinase activity	F	over
<a href="#">GO:0006564</a>	L-serine biosynthetic process	P	over
<a href="#">GO:0006073</a>	cellular glucan metabolic process	P	over
<a href="#">GO:0044444</a>	cytoplasmic part	C	over
<a href="#">GO:0003824</a>	catalytic activity	F	over
<a href="#">GO:0044237</a>	cellular metabolic process	P	over
<a href="#">GO:0005737</a>	cytoplasm	C	over

**Fisher's Exact Test**

Test-Set: only in Eu3026.txt

Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).

GO Term	Name	Type	Over/Under
<a href="#">GO:0006979</a>	response to oxidative stress	P	over
<a href="#">GO:0008378</a>	galactosyltransferase activity	F	over
<a href="#">GO:0042221</a>	response to chemical stimulus	P	over
<a href="#">GO:0003755</a>	peptidyl-prolyl cis-trans isomerase activity	F	over
<a href="#">GO:0016859</a>	cis-trans isomerase activity	F	over
<a href="#">GO:0005730</a>	nucleolus	C	over
<a href="#">GO:0047216</a>	inositol 3-alpha-galactosyltransferase activity	F	over
<a href="#">GO:0035250</a>	UDP-galactosyltransferase activity	F	over
<a href="#">GO:0016844</a>	strictosidine synthase activity	F	over
<a href="#">GO:0016843</a>	amine-lyase activity	F	over
<a href="#">GO:0047274</a>	galactinol-sucrose galactosyltransferase activity	F	over
<a href="#">GO:0004109</a>	coproporphyrinogen oxidase activity	F	over
<a href="#">GO:0006950</a>	response to stress	P	over
<a href="#">GO:0009987</a>	cellular process	P	under
<a href="#">GO:0004601</a>	peroxidase activity	F	over
<a href="#">GO:0016684</a>	oxidoreductase activity, acting on peroxide as acceptor	F	over
<a href="#">GO:0016688</a>	L-ascorbate peroxidase activity	F	over
<a href="#">GO:0009821</a>	alkaloid biosynthetic process	P	over
<a href="#">GO:0009820</a>	alkaloid metabolic process	P	over
<a href="#">GO:0016209</a>	antioxidant activity	F	over
<a href="#">GO:0042277</a>	peptide binding	F	over
<a href="#">GO:0006783</a>	heme biosynthetic process	P	over
<a href="#">GO:0042168</a>	heme metabolic process	P	over
<a href="#">GO:0031981</a>	nuclear lumen	C	over
<a href="#">GO:0055114</a>	oxidation reduction	P	over
<a href="#">GO:0016840</a>	carbon-nitrogen lyase activity	F	over
<a href="#">GO:0016634</a>	oxidoreductase activity, acting on the CH-CH group of donors, oxygen as acceptor	F	over
<a href="#">GO:0010200</a>	response to chitin	P	over
<a href="#">GO:0009743</a>	response to carbohydrate stimulus	P	over

<b>Fisher's Exact Test</b>									
Test-Set: Ai et A 30-26 V2.txt									
Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).									
GO Term	Name	Type	FDR	single test p-Value	# in test group	# in reference group	# non annot test	# non annot reference group	Over/Under
<a href="#">GO:0015995</a>	chlorophyll biosynthetic process	P	6,4E-1	3,5E-4	2	37	5	9273	over
<a href="#">GO:0006779</a>	porphyrin biosynthetic process	P	6,4E-1	6,8E-4	2	52	5	9258	over
<a href="#">GO:0033014</a>	tetrapyrrole biosynthetic process	P	6,4E-1	7,1E-4	2	53	5	9257	over
<a href="#">GO:0019903</a>	protein phosphatase binding	F	6,4E-1	7,5E-4	1	0	6	9310	over
<a href="#">GO:0019902</a>	phosphatase binding	F	6,4E-1	7,5E-4	1	0	6	9310	over
<a href="#">GO:0015994</a>	chlorophyll metabolic process	P	6,4E-1	7,8E-4	2	56	5	9254	over
<a href="#">GO:0006778</a>	porphyrin metabolic process	P	6,4E-1	1,3E-3	2	73	5	9237	over
<a href="#">GO:0033013</a>	tetrapyrrole metabolic process	P	6,4E-1	1,3E-3	2	74	5	9236	over
<a href="#">GO:0003939</a>	L-iditol 2-dehydrogenase activity	F	6,4E-1	1,5E-3	1	1	6	9309	over
<a href="#">GO:0019432</a>	triglyceride biosynthetic process	P	6,4E-1	1,5E-3	1	1	6	9309	over
<a href="#">GO:0046504</a>	glycerol ether biosynthetic process	P	6,4E-1	1,5E-3	1	1	6	9309	over
<a href="#">GO:0046460</a>	neutral lipid biosynthetic process	P	6,4E-1	1,5E-3	1	1	6	9309	over
<a href="#">GO:0046463</a>	acylglycerol biosynthetic process	P	6,4E-1	1,5E-3	1	1	6	9309	over
<a href="#">GO:0046406</a>	magnesium protoporphyrin IX methyltransferase activity	F	6,4E-1	1,5E-3	1	1	6	9309	over
<a href="#">GO:0004853</a>	uroporphyrinogen decarboxylase activity	F	6,4E-1	1,5E-3	1	1	6	9309	over
<a href="#">GO:0046148</a>	pigment biosynthetic process	P	6,9E-1	1,8E-3	2	86	5	9224	over
<a href="#">GO:0006638</a>	neutral lipid metabolic process	P	6,9E-1	2,3E-3	1	2	6	9308	over
<a href="#">GO:0006639</a>	acylglycerol metabolic process	P	6,9E-1	2,3E-3	1	2	6	9308	over
<a href="#">GO:0006641</a>	triglyceride metabolic process	P	6,9E-1	2,3E-3	1	2	6	9308	over
<a href="#">GO:0009931</a>	calcium-dependent protein serine/threonine kinase activity	F	6,9E-1	2,3E-3	1	2	6	9308	over
<a href="#">GO:0010857</a>	calcium-dependent protein kinase activity	F	6,9E-1	2,3E-3	1	2	6	9308	over
<a href="#">GO:0042440</a>	pigment metabolic process	P	7,6E-1	2,6E-3	2	104	5	9206	over
<a href="#">GO:0051188</a>	cofactor biosynthetic process	P	1,0E0	4,2E-3	2	133	5	9177	over
<a href="#">GO:0010359</a>	regulation of anion channel activity	P	1,0E0	5,2E-3	1	6	6	9304	over
<a href="#">GO:0006636</a>	unsaturated fatty acid biosynthetic process	P	1,0E0	6,0E-3	1	7	6	9303	over
<a href="#">GO:0032409</a>	regulation of transporter activity	P	1,0E0	6,0E-3	1	7	6	9303	over
<a href="#">GO:0034765</a>	regulation of ion transmembrane transport	P	1,0E0	6,0E-3	1	7	6	9303	over
<a href="#">GO:0022898</a>	regulation of transmembrane transporter activity	P	1,0E0	6,0E-3	1	7	6	9303	over
<a href="#">GO:0032412</a>	regulation of ion transmembrane transporter activity	P	1,0E0	6,0E-3	1	7	6	9303	over
<a href="#">GO:2000377</a>		N	1,0E0	6,7E-3	1	8	6	9302	over

<b>Fisher's Exact Test</b>									
Test-Set: common 3026.txt									
Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).									
<b>GO Term</b>	<b>Name</b>	<b>Type</b>	<b>FDR</b>	<b>single test p-Value</b>	<b># in test group</b>	<b># in reference group</b>	<b># non annot test</b>	<b># non annot reference group</b>	<b>Over/Under</b>
<a href="#">GO:0009522</a>	photosystem I	C	1,0E0	4,6E-4	2	32	7	9276	over
<a href="#">GO:0046906</a>	tetrapyrrole binding	F	1,0E0	6,2E-4	3	185	6	9123	over
<a href="#">GO:0009535</a>	chloroplast thylakoid membrane	C	1,0E0	1,2E-3	3	235	6	9073	over
<a href="#">GO:0055035</a>	plastid thylakoid membrane	C	1,0E0	1,2E-3	3	235	6	9073	over
<a href="#">GO:0042651</a>	thylakoid membrane	C	1,0E0	1,4E-3	3	246	6	9062	over
<a href="#">GO:0034357</a>	photosynthetic membrane	C	1,0E0	1,6E-3	3	258	6	9050	over
<a href="#">GO:0031984</a>	organelle subcompartment	C	1,0E0	1,8E-3	3	267	6	9041	over
<a href="#">GO:0031976</a>	plastid thylakoid	C	1,0E0	1,8E-3	3	267	6	9041	over
<a href="#">GO:0009534</a>	chloroplast thylakoid	C	1,0E0	1,8E-3	3	267	6	9041	over
<a href="#">GO:0009521</a>	photosystem	C	1,0E0	1,8E-3	2	66	7	9242	over
<a href="#">GO:0044436</a>	thylakoid part	C	1,0E0	2,0E-3	3	280	6	9028	over
<a href="#">GO:0052617</a>	ent-kaurene oxidase activity	N	1,0E0	2,9E-3	1	2	8	9306	over
<a href="#">GO:0010241</a>	ent-kaurene oxidase activity	F	1,0E0	2,9E-3	1	2	8	9306	over
<a href="#">GO:0033331</a>	ent-kaurene metabolic process	P	1,0E0	2,9E-3	1	2	8	9306	over
<a href="#">GO:0052616</a>		N	1,0E0	2,9E-3	1	2	8	9306	over
<a href="#">GO:0052615</a>		N	1,0E0	2,9E-3	1	2	8	9306	over
<a href="#">GO:0044434</a>	chloroplast part	C	1,0E0	3,1E-3	4	707	5	8601	over
<a href="#">GO:0044435</a>	plastid part	C	1,0E0	3,4E-3	4	725	5	8583	over
<a href="#">GO:0009579</a>	thylakoid	C	1,0E0	3,5E-3	3	340	6	8968	over
<a href="#">GO:0051087</a>	chaperone binding	F	1,0E0	5,8E-3	1	5	8	9303	over
<a href="#">GO:0009686</a>	gibberellin biosynthetic process	P	1,0E0	7,7E-3	1	7	8	9301	over
<a href="#">GO:0016102</a>	diterpenoid biosynthetic process	P	1,0E0	7,7E-3	1	7	8	9301	over
<a href="#">GO:0046872</a>	metal ion binding	F	1,0E0	9,1E-3	5	1557	4	7751	over
<a href="#">GO:0016101</a>	diterpenoid metabolic process	P	1,0E0	9,6E-3	1	9	8	9299	over
<a href="#">GO:0009685</a>	gibberellin metabolic process	P	1,0E0	9,6E-3	1	9	8	9299	over
<a href="#">GO:0020037</a>	heme binding	F	1,0E0	9,7E-3	2	158	7	9150	over
<a href="#">GO:0015979</a>	photosynthesis	P	1,0E0	1,0E-2	2	163	7	9145	over
<a href="#">GO:0009740</a>	gibberellic acid mediated signaling pathway	P	1,0E0	1,1E-2	1	10	8	9298	over
<a href="#">GO:0071370</a>	cellular response to gibberellin stimulus	P	1,0E0	1,1E-2	1	10	8	9298	over
<a href="#">GO:0010476</a>	gibberellin mediated signaling pathway	P	1,0E0	1,1E-2	1	10	8	9298	over
<a href="#">GO:0004497</a>	monooxygenase activity	F	1,0E0	1,1E-2	2	166	7	9142	over
<a href="#">GO:0046914</a>	transition metal ion binding	F	1,0E0	1,2E-2	4	1018	5	8290	over
<a href="#">GO:0043169</a>	cation binding	F	1,0E0	1,2E-2	5	1649	4	7659	over
<a href="#">GO:0043167</a>	ion binding	F	1,0E0	1,2E-2	5	1650	4	7658	over
<a href="#">GO:0009707</a>	chloroplast outer membrane	C	1,0E0	1,6E-2	1	16	8	9292	over
<a href="#">GO:0009055</a>	electron carrier activity	F	1,0E0	1,9E-2	2	223	7	9085	over
<a href="#">GO:0009527</a>	plastid outer membrane	C	1,0E0	1,9E-2	1	19	8	9289	over
<a href="#">GO:0018298</a>	protein-chromophore linkage	P	1,0E0	2,1E-2	1	21	8	9287	over
<a href="#">GO:0005792</a>	microsome	C	1,0E0	2,6E-2	1	26	8	9282	over
<a href="#">GO:0042598</a>	vesicular fraction	C	1,0E0	2,6E-2	1	26	8	9282	over
<a href="#">GO:0016168</a>	chlorophyll binding	F	1,0E0	2,6E-2	1	26	8	9282	over
<a href="#">GO:0005506</a>	iron ion binding	F	1,0E0	2,9E-2	2	281	7	9027	over
<a href="#">GO:0009765</a>	photosynthesis, light harvesting	P	1,0E0	3,2E-2	1	33	8	9275	over
<a href="#">GO:0005626</a>	insoluble fraction	C	1,0E0	3,4E-2	1	35	8	9273	over

<b>Fisher's Exact Test</b>			
Test-Set: SGN 26-22 ONLY CAN POUR bLASDTTOGO.txt			
Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).			
<b>GO Term</b>	<b>Name</b>	<b>Type</b>	<b>Over/Under</b>
<a href="#">GO:0046520</a>	sphingoid biosynthetic process	P	over
<a href="#">GO:0000170</a>	sphingosine hydroxylase activity	F	over
<a href="#">GO:0009628</a>	response to abiotic stimulus	P	over
<a href="#">GO:0008138</a>	protein tyrosine/serine/threonine phosphatase activity	F	over
<a href="#">GO:0046519</a>	sphingoid metabolic process	P	over
<a href="#">GO:0030148</a>	sphingolipid biosynthetic process	P	over
<a href="#">GO:0004725</a>	protein tyrosine phosphatase activity	F	over
<a href="#">GO:0009640</a>	photomorphogenesis	P	over
<a href="#">GO:0044434</a>	chloroplast part	C	over
<a href="#">GO:0004721</a>	phosphoprotein phosphatase activity	F	over
<a href="#">GO:0004869</a>	cysteine-type endopeptidase inhibitor activity	F	over
<a href="#">GO:0009535</a>	chloroplast thylakoid membrane	C	over
<a href="#">GO:0055035</a>	plastid thylakoid membrane	C	over
<a href="#">GO:0009579</a>	thylakoid	C	over
<a href="#">GO:0006665</a>	sphingolipid metabolic process	P	over
<a href="#">GO:0042651</a>	thylakoid membrane	C	over
<a href="#">GO:0031984</a>	organelle subcompartment	C	over
<a href="#">GO:0031976</a>	plastid thylakoid	C	over
<a href="#">GO:0009534</a>	chloroplast thylakoid	C	over
<a href="#">GO:0046467</a>	membrane lipid biosynthetic process	P	over
<a href="#">GO:0004866</a>	endopeptidase inhibitor activity	F	over
<a href="#">GO:0061134</a>		N	over
<a href="#">GO:0030414</a>	peptidase inhibitor activity	F	over
<a href="#">GO:0061135</a>		N	over
<a href="#">GO:0009639</a>	response to red or far red light	P	over



Fisher's Exact Test			
Test-Set: SGN 26-22 ONLY EU POUR BLASDTTOGO.txt			
Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).			
GO Term	Name	Type	Over/Under
<a href="#">GO:0004604</a>	phosphoadenylyl-sulfate reductase (thioredoxin) activity	F	over
<a href="#">GO:0019344</a>	cysteine biosynthetic process	P	over
<a href="#">GO:0008686</a>	3,4-dihydroxy-2-butanone-4-phosphate synthase activity	F	over
<a href="#">GO:0003935</a>	GTP cyclohydrolase II activity	F	over
<a href="#">GO:0003933</a>	GTP cyclohydrolase activity	F	over
<a href="#">GO:0006534</a>	cysteine metabolic process	P	over
<a href="#">GO:0009070</a>	serine family amino acid biosynthetic process	P	over
<a href="#">GO:0000097</a>	sulfur amino acid biosynthetic process	P	over
<a href="#">GO:0044271</a>	cellular nitrogen compound biosynthetic process	P	over
<a href="#">GO:0009309</a>	amine biosynthetic process	P	over
<a href="#">GO:0000096</a>	sulfur amino acid metabolic process	P	over
<a href="#">GO:0019238</a>	cyclohydrolase activity	F	over
<a href="#">GO:0044283</a>	small molecule biosynthetic process	P	over
<a href="#">GO:0009231</a>	riboflavin biosynthetic process	P	over
<a href="#">GO:0042726</a>	riboflavin and derivative metabolic process	P	over
<a href="#">GO:0042727</a>	riboflavin and derivative biosynthetic process	P	over
<a href="#">GO:0006771</a>	riboflavin metabolic process	P	over
<a href="#">GO:0016671</a>	oxidoreductase activity, acting on sulfur group of donors, disulfide as acceptor	F	over
<a href="#">GO:0009069</a>	serine family amino acid metabolic process	P	over
<a href="#">GO:0016667</a>	oxidoreductase activity, acting on sulfur group of donors	F	over
<a href="#">GO:0044272</a>	sulfur compound biosynthetic process	P	over
<a href="#">GO:0016814</a>	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines	F	over
<a href="#">GO:0009055</a>	electron carrier activity	F	over
<a href="#">GO:0016053</a>	organic acid biosynthetic process	P	over
<a href="#">GO:0046394</a>	carboxylic acid biosynthetic process	P	over
<a href="#">GO:0008652</a>	cellular amino acid biosynthetic process	P	over
<a href="#">GO:0003824</a>	catalytic activity	F	over
<a href="#">GO:0055114</a>	oxidation reduction	P	over
<a href="#">GO:0006790</a>	sulfur metabolic process	P	over
<a href="#">GO:0060089</a>	molecular transducer activity	F	over
<a href="#">GO:0004871</a>	signal transducer activity	F	over
<a href="#">GO:0016740</a>	transferase activity	F	over
<a href="#">GO:0008152</a>	metabolic process	P	over
<a href="#">GO:0046906</a>	tetrapyrrole binding	F	over
<a href="#">GO:0051539</a>	4 iron, 4 sulfur cluster binding	F	over
<a href="#">GO:0033741</a>	adenylyl-sulfate reductase (glutathione) activity	F	over
<a href="#">GO:0044106</a>	cellular amine metabolic process	P	over
<a href="#">GO:0016830</a>	carbon-carbon lyase activity	F	over
<a href="#">GO:0009308</a>	amine metabolic process	P	over
<a href="#">GO:0016491</a>	oxidoreductase activity	F	over
<a href="#">GO:0046524</a>	sucrose-phosphate synthase activity	F	over
<a href="#">GO:0004350</a>	glutamate-5-semialdehyde dehydrogenase activity	F	over
<a href="#">GO:0004349</a>	glutamate 5-kinase activity	F	over
<a href="#">GO:0019202</a>	amino acid kinase activity	F	over
<a href="#">GO:0003842</a>	1-pyrroline-5-carboxylate dehydrogenase activity	F	over
<a href="#">GO:0009044</a>	xylan 1,4-beta-xylosidase activity	F	over
<a href="#">GO:0006767</a>	water-soluble vitamin metabolic process	P	over
<a href="#">GO:0042364</a>	water-soluble vitamin biosynthetic process	P	over
<a href="#">GO:0006766</a>	vitamin metabolic process	P	over
<a href="#">GO:0009110</a>	vitamin biosynthetic process	P	over
<a href="#">GO:0005506</a>	iron ion binding	F	over
<a href="#">GO:0006561</a>	proline biosynthetic process	P	over
<a href="#">GO:0048307</a>	ferredoxin-nitrite reductase activity	F	over
<a href="#">GO:0016664</a>	oxidoreductase activity, acting on other nitrogenous compounds as donors, iron-sulfur protein as acceptor	F	over
<a href="#">GO:0016810</a>	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	F	over
<a href="#">GO:0016829</a>	lyase activity	F	over
<a href="#">GO:0020037</a>	heme binding	F	over
<a href="#">GO:0016656</a>	monodehydroascorbate reductase (NADH) activity	F	over
<a href="#">GO:0004014</a>	adenosylmethionine decarboxylase activity	F	over
<a href="#">GO:0006597</a>	spermine biosynthetic process	P	over
<a href="#">GO:0008215</a>	spermine metabolic process	P	over
<a href="#">GO:0006560</a>	proline metabolic process	P	over
<a href="#">GO:0043436</a>	oxoacid metabolic process	P	over
<a href="#">GO:0019752</a>	carboxylic acid metabolic process	P	over

## Fisher's Exact Test

Test-Set: SGN 26-22 ONLY Ai et Ar POUR bLASDTTOGO.txt

Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).

GO Term	Name	Type	Over/Under
<a href="#">GO:0006354</a>	RNA elongation	P	over
<a href="#">GO:0030528</a>	transcription regulator activity	F	over
<a href="#">GO:0005942</a>	phosphoinositide 3-kinase complex	C	over
<a href="#">GO:0016303</a>	1-phosphatidylinositol-3-kinase activity	F	over
<a href="#">GO:0035004</a>	phosphoinositide 3-kinase activity	F	over
<a href="#">GO:0048015</a>	phosphoinositide-mediated signaling	P	over
<a href="#">GO:0048017</a>	inositol lipid-mediated signaling	P	over
<a href="#">GO:0046854</a>	phosphoinositide phosphorylation	P	over
<a href="#">GO:0046834</a>	lipid phosphorylation	P	over
<a href="#">GO:0016301</a>	kinase activity	F	over
<a href="#">GO:0004749</a>	ribose phosphate diphosphokinase activity	F	over
<a href="#">GO:0019932</a>	second-messenger-mediated signaling	P	over
<a href="#">GO:0010033</a>	response to organic substance	P	over
<a href="#">GO:0016778</a>	diphosphotransferase activity	F	over
<a href="#">GO:0006471</a>	protein amino acid ADP-ribosylation	P	over
<a href="#">GO:0016772</a>	transferase activity, transferring phosphorus-containing groups	F	over
<a href="#">GO:0001727</a>	lipid kinase activity	F	over
<a href="#">GO:0046488</a>	phosphatidylinositol metabolic process	P	over
<a href="#">GO:0010200</a>	response to chitin	P	over
<a href="#">GO:0003702</a>	RNA polymerase II transcription factor activity	F	over
<a href="#">GO:0009743</a>	response to carbohydrate stimulus	P	over
<a href="#">GO:0030258</a>	lipid modification	P	over
<a href="#">GO:0006650</a>	glycerophospholipid metabolic process	P	over
<a href="#">GO:0009156</a>	ribonucleoside monophosphate biosynthetic process	P	over
<a href="#">GO:0046486</a>	glycerolipid metabolic process	P	over
<a href="#">GO:0009161</a>	ribonucleoside monophosphate metabolic process	P	over
<a href="#">GO:0009124</a>	nucleoside monophosphate biosynthetic process	P	over
<a href="#">GO:0009123</a>	nucleoside monophosphate metabolic process	P	over
<a href="#">obsolete_molecular_function</a>		N	over
<a href="#">GO:0004428</a>	inositol or phosphatidylinositol kinase activity	F	over
<a href="#">GO:0003746</a>	translation elongation factor activity	F	over
<a href="#">GO:0050794</a>	regulation of cellular process	P	over
<a href="#">GO:0009116</a>	nucleoside metabolic process	P	over

<b>Fisher's Exact Test</b>			
Test-Set: SGN 26-22 COMMON POUR bLASDTTOGO.txt			
Test all Gene Ontology terms if they are enriched in a test group when compared to a reference group using Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg).			
<b>GO Term</b>	<b>Name</b>	<b>Type</b>	<b>Over/Under</b>
<a href="#">GO:0016480</a>	negative regulation of transcription from RNA polymerase III promoter	P	over
<a href="#">GO:0006359</a>	regulation of transcription from RNA polymerase III promoter	P	over
<a href="#">GO:0006383</a>	transcription from RNA polymerase III promoter	P	over
<a href="#">GO:0051253</a>	negative regulation of RNA metabolic process	P	over
<a href="#">GO:0045892</a>	negative regulation of transcription, DNA-dependent	P	over
<a href="#">GO:0051172</a>	negative regulation of nitrogen compound metabolic process	P	over
<a href="#">GO:0045934</a>	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	P	over
<a href="#">GO:0010629</a>	negative regulation of gene expression	P	over
<a href="#">GO:0009890</a>	negative regulation of biosynthetic process	P	over
<a href="#">GO:0031327</a>	negative regulation of cellular biosynthetic process	P	over
<a href="#">GO:0010558</a>	negative regulation of macromolecule biosynthetic process	P	over
<a href="#">GO:2000113</a>		N	over
<a href="#">GO:0008967</a>	phosphoglycolate phosphatase activity	F	over
<a href="#">GO:0031324</a>	negative regulation of cellular metabolic process	P	over
<a href="#">GO:0010605</a>	negative regulation of macromolecule metabolic process	P	over
<a href="#">GO:0008270</a>	zinc ion binding	F	over
<a href="#">GO:0009892</a>	negative regulation of metabolic process	P	over
<a href="#">GO:0003677</a>	DNA binding	F	over
<a href="#">GO:0048523</a>	negative regulation of cellular process	P	over
<a href="#">GO:0048519</a>	negative regulation of biological process	P	over
<a href="#">GO:0005634</a>	nucleus	C	over
<a href="#">GO:0005730</a>	nucleolus	C	over

helix-loop-helix dna-binding
dna binding
prokaryotic dna
dna topoisomerase i
dna polymerase alpha catalytic
dna polymerase lambda
dna polymerase epsilon subunit 1
dna repair helicase
dna-damage-inducible protein
copalyl diphosphate synthase
allene oxide synthase
anthranilate synthase component
class iii chitinase
nadph:protochlorophyllide oxidoreductase a
1-deoxyxylulose 5-phosphate synthase
acetylornithine deacetylase
phosphoribosylformylglycinamide synthase
pyruvate kinase
polyphenol oxidase
isochorismate synthase
purine permease
gibberellin 20-oxidase
8-amino-7-oxononanoate synthase
class iii chitinase
ferric reductase-like transmembrane component
rna polymerase sigma factor
ripening-related p-450 enzyme
cinnamyl alcohol dehydrogenase
aspartyl-trna synthetase
phospholipase c
beta-amyrin synthase
anthranilate n-benzoyltransferase
ent-kaurenoic acid oxidase
serine carboxypeptidase ii
6-phosphofructokinase 2
glutamate decarboxylase
beta-amyrin synthase
lanatoside 15 -o-acetylerase-like
purple acid phosphatase
2-oxoglutarate-dependent dioxygenase
polyphenol oxidase
enoyl-acp reductase
udp-glucose:flavonoid glucoside -glucosyltransferase
prolyl-trna synthetase
fructose-bisphosphate aldolase
ubiquitin-conjugating enzyme 2
pyrophosphate--fructose-6-phosphate 1-phosphotransferase
nadh ubiquinone oxidoreductase b22-like subunit
ubiquitin-conjugating enzyme e2 36

**Annexe 17 :**

Liste de 139 processus biologiques obtenus à partir des gènes dits de 'plasticité' chez *Coffea arabica*.

carboxypeptidase d
phosphoethanolamine n-methyltransferase
pectate lyase
peroxidase 64
peroxidase 43
heat shock protein 70
act domain-containing small subunit of acetolactate synthase protein
proton-dependent oligopeptide transport family protein
mlp-like protein 28
33kda precursor protein of oxygen-evolving complex
retrotransposon protein
hsc70-interacting protein
gag-pol polyprotein
heparanase-like protein 3
calcium-binding protein
acyl- -binding protein
cbl-interacting protein kinase 1
pectin methylesterase inhibitor protein 1
ap2 domain cbf protein
afg1-like protein
kinase family protein
bacterial transferase hexapeptide repeat-containing protein
disease resistance-like protein
signal recognition particle 19 kda protein
ring fyve phd zinc finger-containing protein
cytosolic class i small heat shock protein 2a
minichromosome maintenance protein
glycosyl hydrolase family protein with chitinase insertion domain
udp-glycosyltransferase-like protein
silencing group a protein
apyrase-like protein
pentatricopeptide repeat-containing protein
alpha- -glucan-protein synthase
gl1 protein
protein wax2
50s ribosomal protein l21
thiamine biosynthesis protein
beta-glucosidase-like protein
biofilm formation regulatory protein
rnase h domain-containing protein
isoflavone reductase-like protein
subtilase family protein
cytochrome p450
alpha-expansin 1
60s ribosomal
sulfate transporter
gibberellin 20
erd6-like transporter

short chain alcohol
desacetoxyvindoline 4-
cyclic nucleotide-gated ion
polyubiquitin 10
pak inhibitor
nitrate transporter
mip nip subfamily
transferring glycosyl
glucosyl glucuronosyl
ap2 domain-containing transcription factor
amp dependent
auxin efflux carrier component
indole-3-acetic acid inducible 33
nac domain ipr003441
oleosin low molecular weight isoform
thioredoxin y
retrotransposon ty3-gypsy subclass
anthocyanin 5-aromatic
cytochrome p450
alpha-expansin 1
60s ribosomal
sulfate transporter
gibberellin 20
erd6-like transporter
short chain alcohol
desacetoxyvindoline 4-
cyclic nucleotide-gated ion
polyubiquitin 10
pak inhibitor
nitrate transporter
mip nip subfamily
transferring glycosyl
glucosyl glucuronosyl
ap2 domain-containing transcription factor
amp dependent
auxin efflux carrier component
indole-3-acetic acid inducible 33
nac domain ipr003441
oleosin low molecular weight isoform
thioredoxin y
retrotransposon ty3-gypsy subclass
anthocyanin 5-aromatic