



Polyploïdie et adaptation des plantes : caractérisation et variation de l'expression des gènes homoélogues chez le caféier *Coffea arabica*

Marie-Christine Combes Gavalda

► To cite this version:

Marie-Christine Combes Gavalda. Polyploïdie et adaptation des plantes : caractérisation et variation de l'expression des gènes homoélogues chez le caféier *Coffea arabica*. Biologie végétale. Université Montpellier, 2015. Français. NNT : 2015MONTTS115 . tel-01871268

HAL Id: tel-01871268

<https://theses.hal.science/tel-01871268>

Submitted on 10 Sep 2018

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

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

THÈSE

Pour obtenir le grade de
Docteur

Délivré par
Université de Montpellier

Préparée au sein de l'école doctorale GAIA

Et de l'unité de recherche DIADE
« Diversité Adaptation Développement »

Filière : **Biologie, Interactions, Diversité adaptative des
plantes**

Présentée par **Marie-Christine COMBES GAVALDA**

**POLYPLOÏDIE ET ADAPTATION DES PLANTES :
CARACTERISATION ET VARIATION DE L'EXPRESSION
DES GENES HOMEOLOGUES CHEZ LE CAFEIER
*COFFEA ARABICA***

Soutenue le 12 octobre 2015 devant le jury composé de

Eric JENCZEWSKI
Directeur de Recherche, INRA Versailles-Grignon

Rapporteur

Christian PARISOD
Group Leader, P.D., Université Neuchâtel

Rapporteur

Malika AINOUCHE,
Professeur, Université Rennes 1

Examinatrice

Sandrine MAURICE
Maître de conférences, Université Montpellier

Examinatrice

Philippe LASHERMES
Directeur de Recherche, IRD Montpellier

Directeur de thèse

Je voudrais tout d'abord remercier Eric Jenczewski et Christian Parisod pour avoir accepté d'être rapporteur de mon travail de thèse, et Malika Ainouche et Sandrine Maurice pour avoir accepté de faire partie de ce jury. Merci à tous pour le temps que vous avez accordé à l'évaluation de mon travail.

Je veux remercier tout particulièrement Philippe Lashermes bien entendu pour l'encadrement de mon travail de thèse, mais surtout pour sa disponibilité et son soutien tout au long des nombreuses années que nous avons partagées au cours de nos vies professionnelles. Merci pour tout ce que tu m'as enseigné, merci pour la richesse de nos discussions, mais aussi pour m'avoir fait confiance lorsque j'ai entrepris le projet de cette thèse.

Je remercie les membres du comité de thèse, Karine Alix, Carine Charron, Armel Salmon et Jacques David pour les échanges lors des comités de thèse et pour les conseils qu'ils m'ont donné pour orienter mon travail.

Je souhaite remercier toutes les personnes qui ont pris part à la réalisation de ce travail, Alexis Dereeper, Yann Hueber, Marine Peralta pour les analyses bioinformatiques, Alberto Cenci pour les analyses génomiques des clones BAC, Hélène Baraille pour les analyses de biologie moléculaire, Benoît Bertrand pour les analyses statistiques, et Juan-Carlos pour la production des hybrides interspécifiques de *Coffea*.

Je remercie aussi les membres de l'ex-équipe DIVA, équipe dans laquelle j'ai débuté ce travail, et les membres de la nouvelle équipe ADvENS pour leur sympathie tout au long de ces années.

Enfin, je veux remercier « tous mes hommes », grands et petits pour m'avoir soutenue et encouragée tous les jours. Je n'oublie pas tous mes proches qui ont eu le souci et la gentillesse de se préoccuper de mon travail et se sont intéressés à l'avancée de ce projet.

Ces remerciements paraissent sans doute conventionnels mais ils n'en sont pas moins sincères.

TABLE DES MATIERES

TABLE DES MATIERES

INTRODUCTION GENERALE	1
CHAPITRE 1 : REVUE BIBLIOGRAPHIQUE	3
1 La polyplœdie	3
1.1 Palœpolyplœdie et différents types de polyplœdie.....	3
1.2 Formation, fréquence et installation des polyplœides.....	5
2 Etude de la polyplœdie	10
3 Les effets immédiats de la polyplœdie	11
3.1 Les modifications structurales du génome	11
3.1.1 Les changements génomiques	11
3.1.2 Activation des éléments transposables	14
3.2 Les modifications épigénétiques	15
3.2.1 Variation de la méthylation de l'ADN	16
3.2.2 Les modifications des histones	17
3.3 Les modifications de l'expression des gènes	18
3.3.1 Expression des gènes des allopolyploïdes par rapport aux espèces parentales	18
3.3.2 Contribution des gènes homœologues à l'expression totale des gènes	21
3.4 Les étapes de modification et régulation post-transcriptionnelles	24
3.4.1 L'épissage alternatif au sein de polyplœides	25
3.4.2 Le translatome et la polyplœdie	28
3.5 Les modifications du protéome	29
3.6 Les modifications transcriptionnelles par l'intermédiaire des petits ARN	

non codants.....	30
4 Les conséquences de la polyploïdie.....	32
4.1 Hétérosis, vigueur hybride	32
4.2 Evolution des gènes dupliqués	33
4.3 Evolution des polyploïdes.....	36
4.4 Polyploïdie et adaptation des plantes.....	39
5 Le genre <i>Coffea</i>, modèle biologique d'étude	41
5.1 Phylogénie et origine des caféiers	41
5.2 <i>Coffea arabica</i>	44
5.2.1 Origine et caractéristiques	44
5.2.2 Système de reproduction et caractéristiques génomiques	47
5.3 L'environnement naturel des caféiers et leur adaptation	48
5.4 Les hybrides interspécifiques néosynthétiques de <i>Coffea canephora</i> et <i>Coffea eugenioïdes</i>	49
6 Les objectifs de la thèse	52
 CHAPITRE 2 : ETUDE D'UNE ETAPE MAJEURE DE L'ALLOPOLYPLOÏDIE : L'HYBRIDATION GENOMIQUE.....	 54
1 Evolution et étude de l'expression des gènes	54
2 Régulation de l'expression des gènes	56
3 Les hybrides interspécifiques	58
4 L'étude d'hybrides interspécifiques de <i>Coffea</i>	61
4.1 L'apport de connaissances des hybrides interspécifiques	62

4.2 Méthodologie développée pour l'étude des hybrides interspécifiques	63
4.2.1 Hérité de l'expression des gènes dans les hybrides interspécifiques de <i>Coffea</i>	63
4.2.2 Analyse de l'ASE des hybrides interspécifiques de <i>Coffea</i>	64
CHAPITRE 3 : L'ALLOPOLYPLOÏDIE ET L'EXPRESSION RELATIVE DES GENES HOMEOLOGUES	79
1 Analyse et expression relative de gènes homéologues chez <i>Coffea arabica</i>	79
2 Caractérisation et variation de l'expression relative des gènes homéologues chez <i>Coffea arabica</i>	92
CHAPITRE 4 : DISCUSSION GENERALE.....	103
1 Synthèse des principaux résultats	103
1.1 Caractérisation de l'expression des gènes après un évènement d'hybridation	103
1.2 Hérité de l'expression des gènes dans les hybrides	104
1.3 Régulation de l'expression des allèles et des homéologues.....	105
1.3.1 <i>Cis</i> et <i>trans</i> régulation des allèles et des homéologues	105
1.3.2 Effets des facteurs de <i>trans</i> -régulation sur l'expression des allèles	105
1.3.3 Relation entre la régulation de l'expression des allèles et l'hérité d'expression des gènes chez les hybrides.....	106
1.4 Variation de l'expression relative des homéologues en fonction de la température	106
2 Discussion des résultats et présentation d'hypothèses.....	107
2.1 Caractérisation globale de l'expression relative des génomes après un Evènement d'hybridation	107
2.2 L'héritage parental et la régulation <i>cis</i> versus <i>trans</i>	109
2.3 Variation faible de l'expression relative des gènes homéologues	111
2.4 Modèle de l'expression des gènes homéologues chez <i>C. arabica</i>	113

3 Perspectives	117
3.1 Explication de l'expression différentielle des espèces de Coffea par des différences génomiques	117
3.2 Dynamisme de la régulation de l'expression des gènes : mécanismes d'adaptation phénotypique des plantes après un évènement d'hybridation ou d'allopolyploïdisation	119
3.3 Complexité de la régulation des gènes au sein d'hybrides et de polyploïdes.	122
LISTE BIBLIOGRAPHIQUE	124
ANNEXES	138

Liste des figures

Chapitre 1 : Revue bibliographique

Page 4	Figure 1 : Evènements anciens de polyploïdie chez les plantes à graines et chez les angiospermes
	Figure 2 : 20 évènements récents de WGD datés de la fin du Crétacé
Page 7	Figure 3: Voies de formation d'un tétraploïde.
Page 10	Figure 4 : Processus à la fois adaptatif et neutre contribuant à favoriser l'installation d'un polyploïde dans un environnement stressant.
Page 12	Figure 5 : Recombinaison et réarrangements chromosomiques.
Page 15	Figure 6 : Modèle de l'activation de TEs due à des différences épigénétiques et dépendant de l'effet dose, au sein d'un allopolyploïde
Page 16	Figure 7 : Facteurs génétiques et épigénétiques du modelage de la chromatine
Page 19	Figure 8 : Notions additivité et de non-additivité.
Page 20	Figure 9 : Profils d'expression de gènes d'un allopolyploïde comparé à ceux de ses espèces parentales.
Page 23	Figure 10 : Représentation des phénomènes principaux révélés par les deux approches d'analyse de l'expression des gènes au sein d'un allopolyploïde.
Page 24	Figure 11 : Explication du phénomène de dominance du niveau d'expression
Page 26	Figure 12 : Schéma illustrant l'impact de l'épissage alternatif sur la régulation de l'expression des gènes et sur le phénotype.
Page 27	Figure 13 : Fréquence et types d'épissage alternatif fréquent chez l'homme et chez <i>Arabidopsis thaliana</i> .
Page 31	Figure 14 : Classification hiérarchique des petits ARN non codants chez les plantes
Page 33	Figure 15 : Un modèle de mécanisme moléculaire de la vigueur hybride
Page 34	Figure 16 : Phases d'évolution conduisant à la préservation d'un gène à l'état dupliqué
Page 35	Figure 17 : Hypothèses expliquant la résistance de certains gènes à la duplication, pour des WGD ou des duplications de petite échelle (small scale duplication SSD)
Page 41	Figure 18 : Phylogénie moléculaire des caféiers sur la base de l'analyse de l'ADN chloroplastique

Figure 19 : Distribution des espèces du genre *Coffea*.

Page 42	Figure 20 : Dispersion du genre <i>Coffea</i> à partir d'un centre d'origine située en basse Guinée
Page 43	Figure 21 : Phylogénie et duplication pour les Eudicotylédones.
Page 45	Figure 22 : Origine de la seule espèce allotétraploïde <i>C. arabica</i>
Page 46	Figure 23 : Origine des deux variétés « mères » des principales variétés cultivées actuelles de <i>C. arabica</i>
Page 49	Figure 24 : A/ Origine et Diversité des caféiers diploïdes africains utilisés comme parents dans les croisements interspécifiques. B/ Viabilité pollinique des hybrides F1 décrivant leur fertilité
Page 50	Figure 25 : A/ <i>C. canephora</i> , B/ <i>C. eugenioides</i>
Page 51	Figure 26 : Diversité phénotypique des hybrides F1, en haut <i>C. canephora</i> x <i>C. eugenioides</i> , en bas <i>C. eugenioides</i> x <i>C. canephora</i> .

Chapitre 2 : Etude d'une étape majeure de l'allopolyploïdie : l'hybridation génomique

Page 54	Figure 27 : Structure du gène et présentation des régions codantes et de <i>cis</i> -régulation.
Page 56	Figure 28 : Représentations de la région <i>cis</i> -régulation d'un gène.
Page 58	Figure 29 : Dispositif expérimental utilisant l'estimation de l'ASE afin de déterminer les différences de <i>cis</i> - et <i>trans</i> -régulation entre deux individus diploïdes.
Page 59	Figure 30 : Schéma pour l'interprétation de résultats d'une étude d'ASE.
Page 61	Figure 31 : Dispositif expérimental de l'étude d'hybrides interspécifiques de <i>Coffea</i> .
Page 63	Figure 32 : Héritéité du niveau d'expression des gènes.
Page 64	Figure 33 : Corrélation de l'ASE entre les deux groupes d'hybrides.
Page 66	Figure 34 : Etude des facteurs de <i>trans</i> -régulation.
	Figure 35 : Révélation des divergences de <i>trans</i> -régulation sous-jacentes à l'héritéité de l'expression des gènes de l'hybride.

Article "Regulatory divergence between parental alleles determines gene expression patterns in hybrids"

Page 70	FIG. 1.—Differences in gene expression between <i>Coffea canephora</i> and <i>Coffea eugenioides</i> .
Page 71	FIG. 2.—Differentially expressed genes in each possible comparison between parents and hybrids

Page 72	FIG. 3.—Inheritance of gene expression level in hybrids between <i>Coffea canephora</i> and <i>Coffea eugenioides</i> .
Page 73	FIG. 4.—cis- and trans-regulatory divergence between parental species for 11,438 genes in C_E hybrids and 11,270 genes in E_C hybrids.
Page 75	FIG. 5.—Variation in specific allelic expression (ASE) for genes differentially expressed between parents and the hybrid in the C_E hybrids.
	FIG. 6.—Effects of trans-regulatory factors of the E and C genome on up- and downregulation of Chyb and Ehyb alleles in C_E and E_Chybrids.
Page 76	FIG. 7.—For the C_E hybrids, all the data collected on genes belonging to each category of expression inheritance are combined.

Chapitre 3 : L'alloploïdie et l'expression relative des genes homéologues

Article “Homeologous gene expression in response to growing temperature in a recent allopolyploid (*Coffea Arabica* L.)”

Page 84	Figure 1. Validation of the sequencing-based method to determine the relative contribution of the homeologs using 2 methods.
Page 86	Figure 2. Representation of the variation in the relative contribution of homeologs (Ea/Ea þ Ca) for the 13 genes as a function of the organ and the culture conditions.
Page 88	Figure 3. Representation of the average relative homeologous expression of 13 genes (Ea/Ea þ Ca) in different organs and for 2 growth condition.
	Figure 4. Variation in the relative contribution of homeologs between Caturra (Ea/Ea þ Ca) and an introgressed variety S795

Article “Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures .”

Page 96	Fig. 1 Relative homoeologous gene expression in <i>Coffea arabica</i> in twodifferent growing conditions
Page 97	Fig. 2 Variation in the relative homoeologous gene expression in <i>Coffea arabica</i> in the two growing conditions (warm and cold).
	Fig. 3 Correlation between the relative homoeologous gene expression measured in <i>Coffea arabica</i> in the two growing conditions (warm and cold).
Page 98	Fig. 4 Relationships between the relative homoeologous gene expression and total expression variations observed in <i>Coffea arabica</i> in two growing conditions.
Page 99	Fig. 5 Comparison of the relative homoeologous gene expression in <i>Coffea arabica</i> with the gene expression ratio between its diploid progenitor species.

Chapitre 4 : Discussion générale

Page 104	Figure 36 : Comparaison de la distribution des gènes dans les catégories de l'expression relative des gènes homéologues ou des allèles.
Page 114	Figure 37 : Modèle de l'expression des gènes homéologues au sein d'un allopolyploïde <i>C. arabica</i> , lien entre hérédité de l'expression des gènes et la <i>trans</i> -régulation des gènes homéologues.
Page 116	Figure 38 : Application du modèle aux données d'expression des gènes de <i>C. arabica</i> cultivé à des températures différentes

Liste des tableaux

Chapitre 1 : Revue bibliographique

Page 33	Tableau I : Différents types de duplication de genes selon <i>Wang et al.</i>
Page 37	Tableau II : Liste des remaniements génomiques et épigénétiques et leurs conséquences évolutives chez les allopolyploïdes

Chapitre 2 : Etude d'une étape majeure de l'allopolyploïdie : l'hybridation génomique

Page 64	Tableau III : Corrélation de l'ASE entre les différentes répétitions des deux groupes d'hybrides.
---------	---

Chapitre 3 : L'allopolyploïdie et l'expression relative des genes homéologues

Article « Homeologous gene expression in response to growing temperature in a recent allopolyploid (*Coffea Arabica* L.) »

Page 83	Table 1 Homeologous genes investigated, SNP, and pairs of oligonucleotides used to study their relative contribution to the transcriptome of <i>Coffea Arabica</i>
Page 87	Table 2 ANOVA of the relative expression of the homeologs (Ea/Ea þ Ca) for the 13 genes as a function of the organ and culture condition.

INTRODUCTION GENERALE

INTRODUCTION GENERALE

La plasticité phénotypique et l'évolution génétique sont des mécanismes de flexibilité dont disposent les plantes pour s'adapter à des changements environnementaux. La plasticité phénotypique qui se traduit par la capacité à produire rapidement de multiples phénotypes en réponse à des conditions environnementales changeantes, peut tamponner les effets de l'environnement, accélérer ou limiter l'adaptation génétique des organismes. L'évolution génétique dépendant de la diversité génétique et des modifications qui peuvent se produire sur les régions codantes et/ou régulatrices des génomes peut permettre l'adaptation à un environnement devenu défavorable. Ces deux mécanismes influant l'un sur l'autre fournissent à l'individu ou à l'espèce un potentiel adaptatif et évolutif. Ce potentiel résulte de la flexibilité de la chaîne de mécanismes moléculaires qui relient les stimuli environnementaux au phénotype en passant par l'information génétique, l'expression génique et protéique et leurs régulations. Mais il découle aussi de la variabilité génétique et des patrons de la sélection naturelle.

La polyploïdie est la caractéristique des espèces qui ont subi au cours de leur évolution un événement de duplication du génome associé ou non à une hybridation interspécifique. La fusion au sein d'une même cellule de deux génomes plus ou moins divergents et de leurs systèmes de régulation génique modifie l'environnement des gènes, ce qui peut provoquer des variations des mécanismes induisant le phénotype des plantes. L'état polyploïde est très fréquent chez les plantes et des marques de polyploïdie ancienne ont été observées pour tous les génomes de plantes. La polyploïdie est considérée comme un facteur majeur de la plasticité phénotypique et de l'évolution génétique pouvant modifier les capacités adaptatives des espèces végétales.

Depuis une dizaine d'années, les espèces polyploïdes ont fait l'objet de nombreuses études, autant pour déceler les conséquences de la polyploïdie sur la structure des génomes que sur la variation de l'expression des gènes. Dès les premières générations après un événement d'hybridation, des changements importants interviennent et modifient la structure et le fonctionnement des génomes. Les changements du fonctionnement des génomes se traduisent par des modifications de l'expression transcriptomique et/ou protéomique des espèces polyploïdes. Les études portant sur le transcriptome ont d'abord consisté à comparer l'expression génique des espèces polyploïdes et de leurs espèces

parentales. Avec l'essor technologique des méthodes de séquençage à haut débit, il est maintenant possible d'étudier au sein des polyploïdes, l'expression relative des gènes homéologues qui dérivent de chacune des espèces parentales. Même si progressivement, de plus en plus de modèles de polyploïdes sont décrits, aucune étude ne s'est intéressée à l'analyse de l'expression des gènes homéologues au sein d'une plante ligneuse et de plus en conditions environnementales changeantes. Pourtant face aux changements climatiques, des plantes des forêts tropicales et des espèces d'intérêt agronomique qui en sont originaires, adaptées à des variations environnementales de faible intensité, sont de plus en plus exposées à des contraintes environnementales fortes. Par conséquent, étudier et comprendre le potentiel d'adaptation de ces plantes constitue un enjeu majeur pour le développement d'une agriculture durable au Sud.

Le caféier et plus particulièrement *Coffea arabica*, espèce pérenne polyploïde des forêts tropicales, représente un modèle intéressant à la fois pour l'étude des capacités adaptatives d'un polyploïde et pour la connaissance d'une plante d'intérêt agronomique au Sud. En effet, le café est un produit de consommation international se rangeant parmi les exportations agricoles les plus importantes des pays du Sud. 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 essentiellement sur deux espèces, le café Arabica (*Coffea arabica*) et le café Robusta (*Coffea canephora*) qui représentent respectivement 65% et 35% de la production. La culture des caféiers est répandue dans toute la ceinture intertropicale.

Afin d'étudier les capacités adaptatives induites par la polyplœdie, nous proposons au cours de cette thèse, d'une part d'évaluer les conséquences de l'hybridation en analysant les variations de l'expression allélique et de sa régulation au sein d'hybrides F1 entre *C. canephora* et *C. eugenioides* mais aussi en examinant à l'échelle du génome l'expression relative des gènes homéologues de *C. arabica*. D'autre part nous présenterons une étude de la variation de l'expression des gènes homéologues et de son implication aux capacités adaptatives de l'allopolyplœdie en fonction de conditions de culture.

CHAPITRE 1

Revue Bibliographique

CHAPITRE 1 : REVUE BIBLIOGRAPHIQUE

1 La polyploïdie

1.1 Paléopolyploïdie et différents types de polyploïdie

Le terme de polyploïde qualifie les individus ou les cellules qui présentent plus de deux lots de chromosomes. La polyploïdie a été décrite chez différents organismes eucaryotes comme les plantes, les animaux (review Mable *et al.* 2011), les champignons (Albertin et Marullo 2012) ou encore les levures (Kellis *et al.* 2004). A l'exception des poissons et des amphibiens, ces événements sont rares chez les animaux (Otto 2007) alors que la polyploïdisation ou la duplication du génome entier (Whole Genome duplication WGD) est un phénomène récurrent chez les végétaux et plus particulièrement chez les Angiospermes. La sensibilité du développement embryonnaire aux variations de dose de copies de gènes, notamment ceux liés au déterminisme sexuel expliquerait en partie la fréquence faible d'événements de polyploïdisation au sein des animaux (Mable 2004, Chen et Ni 2006, Evans dans Soltis et Soltis 2012). L'étude de gènes dupliqués partagés par divers eucaryotes (animaux, champignons et plantes) indique que des événements de duplication se sont produits dès le début de l'histoire évolutive des Eucaryotes (Zhou *et al.* 2010). Plus récemment, Jiao *et al.* (2011) ont mis en évidence chez les plantes deux événements de duplication anciens, le premier ayant eu lieu il y a environ 320-350 millions d'années (Ma) est antérieur à la séparation des Angiospermes et des Gymnospermes et le second datant d'environ 200-230 millions d'années (Ma) est commun à tous les Angiospermes (Jiao *et al.* 2011) (Figure 1).

On considère à ce jour que l'ensemble des Angiospermes a une origine polyploïde et/ou paléopolyploïde, et la question actuelle les concernant ne consiste plus à démontrer leur origine polyploïde, mais plutôt à déterminer combien d'événements de polyploïdisation se sont produits pour les différentes lignées (Soltis *et al.* 2009). L'analyse de séquences de plusieurs génomes a démontré qu'un événement de triplication était commun à tous les Dicotylédones (Jaillon *et al.* 2007) et deux événements de duplication étaient communs à tous les Monocotylédones (Paterson *et al.* 2004) (Figure 1). Plusieurs autres événements de WGD, plus récents, ont été révélés, certains communs à plusieurs lignées d'Angiospermes et d'autres spécifiques à certaines lignées (Cui *et al.* 2006, Fawcett *et al.* 2009, Van de Peer *et al.* 2009, Soltis *et al.* 2009, Vanneste, Baele, *et al.* 2014, Vanneste, Maere, *et al.* 2014, Renny-Byfield et Wendel 2014) (Figure 2).

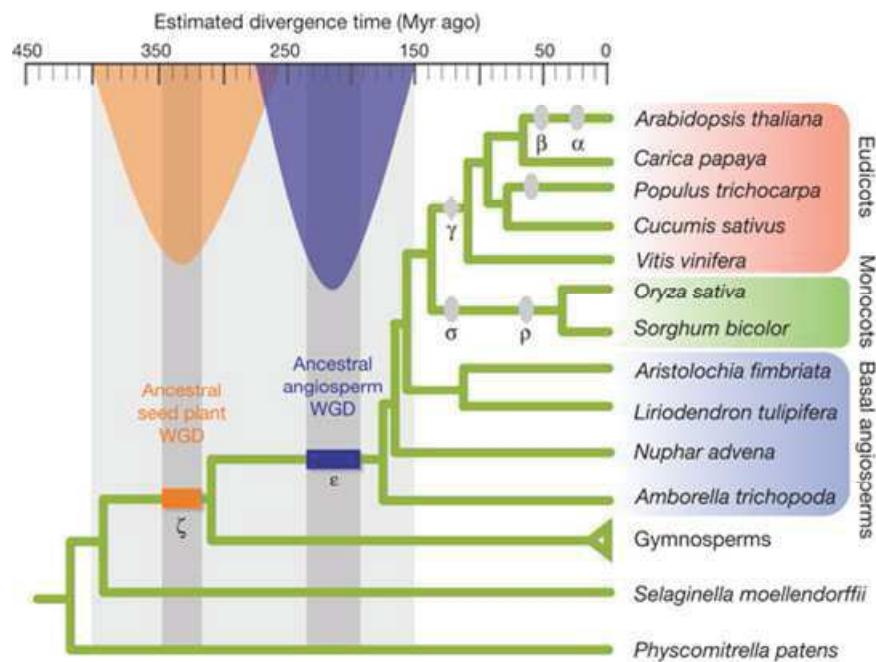


Figure 1 : Evènements anciens de polyploïdie chez les plantes à graines et chez les angiospermes Jiao *et al* (2011).

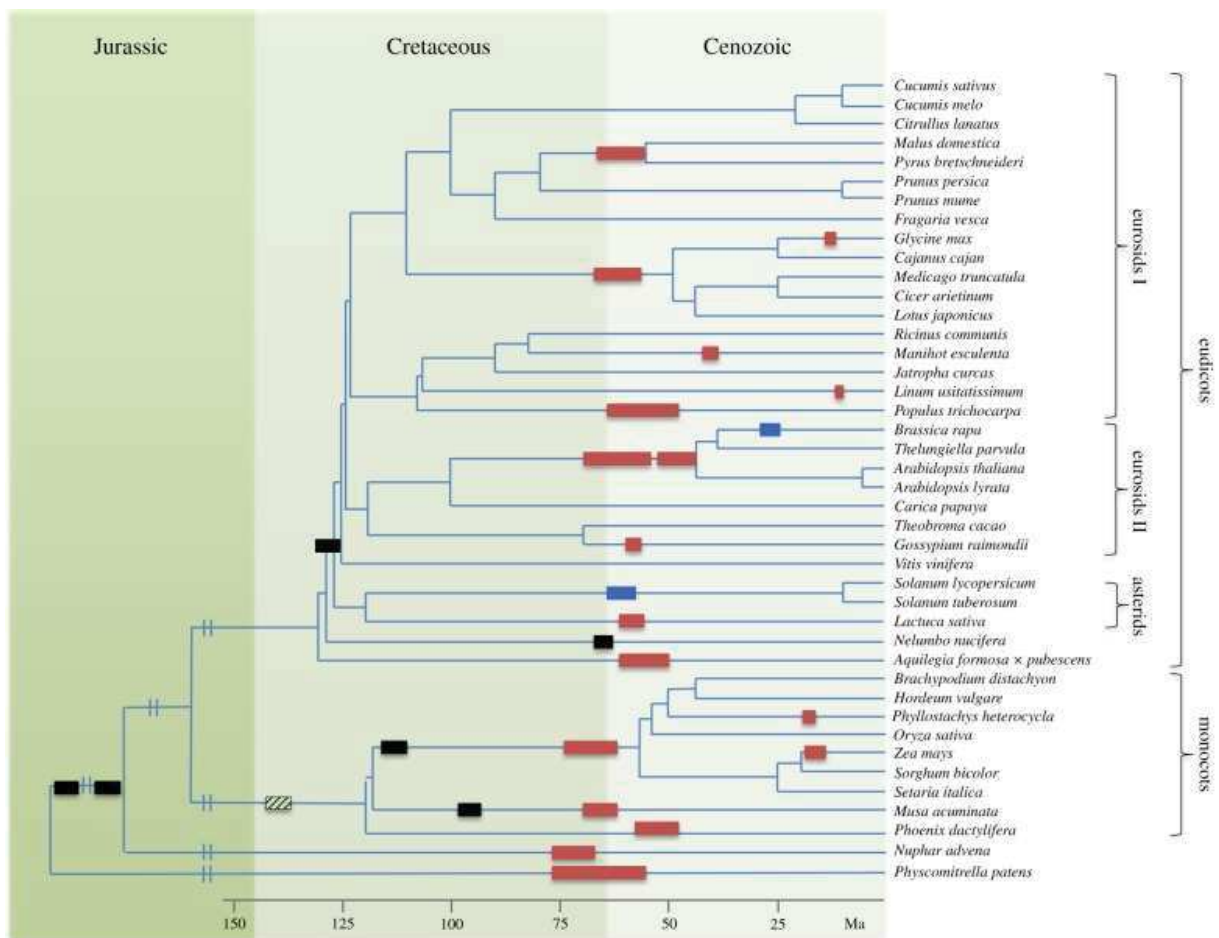


Figure 2 : 20 évènements récents de WGD datés de la fin du Crétacé, révélés au cours de l'analyse de 41 génomes de plantes (Vanneste, Baele, et al. 2014).

Plusieurs types de polyploïdes sont distingués selon les évènements qui sont à leur origine (Chen et Ni 2006, Chen 2007). Les autopolyploïdes au sein desquels des génomes provenant d'une même espèce sont juxtaposés, sont composés de plusieurs jeux de chromosomes homologues. Chez les végétaux, de nombreuses espèces autopolyploïdes sont répertoriées : des espèces cultivées comme la Pomme de terre, la Canne à sucre, la Luzerne et aussi des espèces sauvages ou ornementales comme les *Lantana*. Les allopolyploïdes qui résultent d'une hybridation interspécifique associée à un doublement de chromosomes présentent à la fois des paires de chromosomes homologues mais aussi des jeux de chromosomes apparentés (homéologues), plus ou moins divergents, provenant des espèces parentales. Des espèces cultivées comme le Coton, le Blé, le Colza, le Caféier, le Tabac, l'Arachide mais aussi des espèces sauvages comme la Spartine (*Spartina*), *Tragopogon* et *Senecio* sont des allopolyploïdes. Sur le plan cytogénétique, les deux types de polyploïdes se différencient par l'appariement des chromosomes au cours de la méiose. Alors qu'au sein des autopolyploïdes, tous les chromosomes peuvent s'apparier au cours de la méiose, conduisant à la formation de bivalents ou multivalents (alignement partiel ou total de plus de deux chromosomes), pour les allopolyploïdes, l'appariement des chromosomes est principalement disomique, des bivalents se produisent entre les chromosomes homologues et pas homéologues (Madlung et Wendel 2013). Au-delà de cette séparation stricte des types de polyploïdes, un continuum d'état cytologique au cours de la méiose est possible, et dépend de la divergence entre les chromosomes homéologues (Madlung et Wendel 2013, Soltis, Visger, et al. 2014). Pour des espèces ayant des chromosomes homéologues divergents, l'appariement pourraient se faire strictement entre chromosomes homologues, selon un mode d'hérédité disomique, alors que pour des espèces dont les chromosomes présentent une grande homologie entre leurs homéologues, des échanges entre chromosomes homéologues sont possibles au cours de la formation de multivalents (Ramsey et Schemske 2002, Madlung et Wendel 2013).

1.2 Formation, fréquence et installation des polyploïdes

Différents modèles de génération de polyploïdes sont représentés sur la figure 3. Un individu polyploïde peut provenir du doublement chromosomique spontané dans le zygote suite à une hybridation entre deux espèces différentes (voie de doublement somatique) (Figure 3B) ou bien par la production de gamètes non-réduits (Figure 3C et 3D) (Moghe et Shiu 2014). Lors d'une méiose régulière, deux divisions successives font suite à une seule étape de réplication de l'ADN, ce qui conduit à diviser de moitié le nombre de chromosome ; la méiose produit des cellules haploïdes à n chromosomes (gamètes) à partir d'un organisme diploïde à $2n$ chromosomes. Lors de la première division (réductionnelle), les chromosomes homologues sont séparés, tandis que dans la seconde division (équationnelle) ce sont les chromatides sœurs produites lors de la réplication qui sont séparés. Le déroulement de la méiose peut être perturbé et peut conduire à la formation de gamètes possédant

2n chromosomes, gamètes dits « non-réduits ». Ces gamètes ayant un rôle important dans le processus de polyploïdisation, des études ont été menées pour comprendre leur formation. Globalement deux types de gamètes non réduits ont été observés. Les gamètes du type FDR (pour «First-Division Restitution») qui héritent d'une chromatide (recombinée ou non) de chaque chromosome présent chez le parent. Le produit final est ainsi équivalent à une absence de première division méiotique. D'un point de vue mécanistique, plusieurs aberrations différentes du déroulement de la méiose peuvent conduire à la formation de ce type de gamètes (Bretagnolle et Thompson 1995). Le second type de gamètes non réduits sont les gamètes de type SDR pour «Second-Division Restitution», ils contiennent au contraire des paires de chromatides sœurs, résultant de problèmes ou d'une absence complète de deuxième division. Ce dernier type de gamètes non-réduits semblerait moins fréquent au sein des polyploïdes (Ramanna et Jacobsen 2003). Les études menées sur les bases génétiques de la production de gamètes non-réduits laissent supposer que des populations de plantes possèdent des caractères génétiques leur donnant la capacité de produire des gamètes non-réduits. (Bretagnolle et Thompson 1995, Ramsey et Schemske 1998). Le contrôle génétique de cette caractéristique a été exploré chez *Arabidopsis*, un gène lié à la formation de gamètes non-réduits à haute fréquence a été caractérisé (d' Erfurth et al. 2008, d' Erfurth et al. 2009) et dernièrement un QTL qui serait associé à la formation de gamètes non-réduits de type FDR a été identifié sur des hybrides hexaploïdes de blé (Hao et al. 2014).

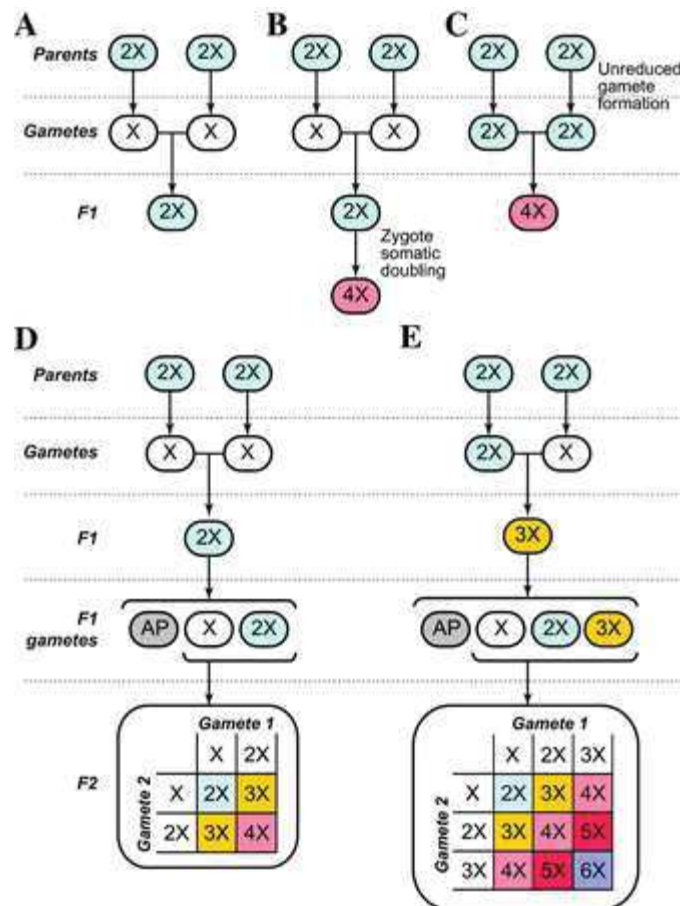


Figure 3: Voies de formation d'un tétraploïde. Le symbole X représente le nombre de chromosomes de base chez une espèce et 1X correspond aux gamètes haploïdes. **A** Voie normale de production d'hybrides F1 entre deux espèces diploïdes. **B** Voie de doublement somatique pour générer un tétraploïde 4X. **C** Voie de fusion de gamètes non-réduits. **D** Voie de production suite à la fusion de gamètes non-réduits d'hybrides F1 entre deux espèces diploïdes. **E** Voie du pont triploïde, fusion de gamètes de niveau de ploïdie variable. (Moghe et Shiu 2014).

Chez les plantes, on considère que la formation de polyploïdes est plus probable par la voie des gamètes non-réduits que par la voie somatique (Ramsey et Schemske 2002, Bretagnolle et Thompson 1995). Les gamètes non-réduits de diploïdes peuvent conduire à un tétraploïde par l'hybridation entre des gamètes non-réduits males et femelles (en une étape Figure 3C ou en deux étapes Figure 3D) ou par l'intermédiaire d'un triploïde (parfois appelé pont triploïde) (Figure 3E). Pour cette dernière voie de création de polyploïde, la plante triploïde va produire des gamètes aneuploïdes, certains viables vont pouvoir s'hybrider à d'autres gamètes aneuploïdes et générer une plante avec un niveau de polyploïdie supérieur. Actuellement il n'est pas démontré que l'une des voies soit plus fréquente (Soltis et al. 2010). Mais toutes les deux dépendent de la formation de gamètes non-réduits.

La grande variation de la fréquence de la polyplœdie suggère des tendances différentes des groupes taxonomiques à la polyplœdie et à la réussite de son établissement (Soltis et al. 2010, Weiss-Schneeweiss et al. 2013). La fréquence de formation de polyplœde dépend à la fois de facteurs génétiques mais aussi de facteurs environnementaux (Otto 2007).

Outre la proximité géographique entre deux espèces indispensable à l'hybridation interspécifique, la formation d'un polyplœde et d'une population qui en découlerait dépend de facteurs génétiques et génomiques qui seraient impliqués dans le contrôle de l'appariement des chromosomes au cours de la méiose. La ségrégation équilibrée des chromosomes au moment de la méiose détermine la normalité des gamètes produits et conditionne la fertilité d'un croisement interspécifique. Des systèmes génétiques contrôlant l'appariement des chromosomes décrits pour de nombreuses plantes (blé, *Aegilops*, avoine, coton, tabac, *Brassica*...) de façon plus ou moins détaillée selon les espèces ont été comparés (Jenczewski et Alix 2004). Il existe une diversité de systèmes génétiques impliqués au cours de différentes étapes du processus d'appariement des chromosomes. L'action de ces systèmes est parfois complétée, par exemple chez le blé, par des remaniements génomiques (délétion de séquences principalement d'un génome) qui augmenteraient la divergence entre les chromosomes homéologues et favoriseraient l'appariement des chromosomes homologues (Feldman et Levy 2012). L'origine et l'évolution du système d'appariement des chromosomes ne sont pas bien connues, plusieurs modèles sont proposés, l'un indiquant l'existence de tel système de contrôle d'appariement des gènes au sein des espèces diploïdes parentales, l'autre indiquant l'apparition progressive de ce contrôle après l'allopolyplœdisation (Jenczewski et Alix 2004). Pour d'autres auteurs, la divergence génétique entre les espèces parentales détermine la fertilité des hybrides après l'hybridation. Ils considèrent les règles de Darlington : pour des espèces parentales de faible divergence, des hybrides homoploïdes seront fertiles, alors que des allopolyplœdes auront une fertilité plus faible, en raison de la possibilité d'une ségrégation déséquilibrée des chromosomes. En revanche, pour des espèces parentales plus divergentes, les hybrides homoploïdes seront stériles, les chromosomes ne pourront pas s'apparier, mais des allopolyplœdes seront plus fertiles en raison de la duplication qui permettra la formation de bivalents à la méiose (Soltis et al. 2010). Ainsi Buggs et al. ont démontré que la formation de polyplœdes se faisait au hasard (n'était pas dépendante de la divergence génétique entre les espèces parentales), en revanche la formation d'hybrides homoploïdes se produisait avec plus de succès entre espèces parentales de faible divergence (Buggs et al. 2009).

En plus des facteurs génétiques, la formation de polyplœdes peut dépendre de facteurs environnementaux. Ainsi il a été observé que la production de gamètes non-réduits peut être augmentée suite à l'exposition à un stress ou à un environnement changeant (Ramsey et Schemske 1998, Mable 2004); notamment lié à la variation de la température (Mable et al. 2011, Pécir et al. 2011). Vanneste *et al.* (2014) interprète l'apparition de nombreux événements de polyplœdie à la fin du Crétacé, période caractérisée par une augmentation de stress et de changements environnementaux,

comme la conséquence de l'exposition des plantes à un environnement stressant (Vanneste, Baele, et al. 2014, Vanneste, Maere, et al. 2014). Au-delà des conditions favorables à la formation de gamètes non-réduits, d'autres facteurs de type physiologique ou écologique interviennent dans l'établissement d'individus polyploïdes (Arrigo et Barker 2012). En effet la polyploïdisation entraîne immédiatement des inconvénients (augmentation de la taille du génome, réduction de la fertilité) mais cependant dans certains cas, elle apporte des avantages (tolérance à la sécheresse, résistance à des pathogènes) qui vont permettre aux polyploïdes de se développer dans une niche écologique distincte (Otto 2007). Certains polyploïdes vont s'isoler des espèces diploïdes suite à l'acquisition de nouveaux modes de propagation (propagation végétative, apomixie, auto-compatibilité) qui vont permettre leur expansion (Weiss-Schneeweiss et al. 2013) ou en profitant dans certaines circonstances d'une descendance viable et fertile après un croisement inter-polyploïde (Arrigo et Barker 2012). La taille de la population de polyploïdes, dépendant de l'origine simple ou multiple des polyploïdes et de la pérennité des espèces, intervient aussi dans le succès de l'installation d'une population de polyploïdes (Arrigo et Barker 2012, Soltis et al. 2014). Pour les allopolyploïdes, en plus de la proximité géographique et de la compatibilité des caractéristiques de floraison des espèces parentales nécessaires à l'hybridation, la divergence génétique entre les espèces parentales influence aussi la réussite de l'installation des allopolyploïdes en réduisant le nombre d'anomalies de la méiose (Ramsey et Schemske 2002, Weiss-Schneeweiss et al. 2013) . A partir de l'observation des nombreux cas de polyploïdie de la fin du Crétacé, Vanneste et al. (2014) ont proposé un modèle pour schématiser la contribution de différents facteurs au succès de l'installation de polyploïdes (Vanneste, Maere, et al. 2014) (Figure 4). Pour Arrigo *et al.* (2012) même si les conditions propices à l'émergence de polyploïdes sont souvent réunies, peu de polyploïdes survivent et parviennent à durer à l'échelle de l'évolution (Mayrose et al. 2011, Arrigo et Barker 2012). L'importance et l'influence des polyploïdes dans l'évolution des Angiospermes sera abordée plus loin.

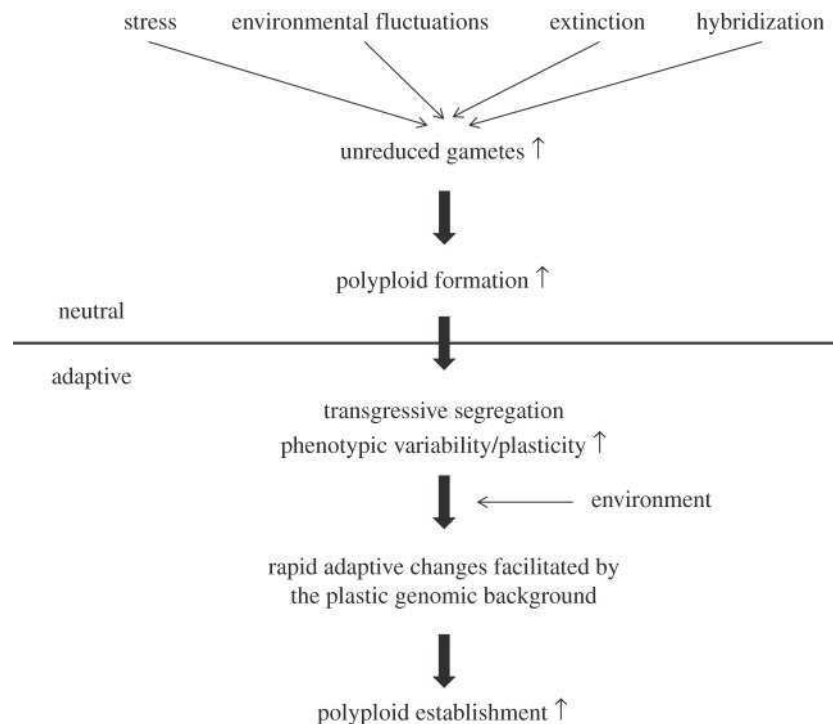


Figure 4 : Processus à la fois adaptatif et neutre contribuant à favoriser l’installation d’un polyplœide dans un environnement stressant. La fréquence de formation de gamètes non-réduits peut être augmentée par des facteurs environnementaux ou évolutifs des populations diploïdes. Plus de polyplœides sont formés. L’instabilité des génomes polyplœides et la ségrégation transgressive favorisent l’apparition de polyplœides avec des caractéristiques phénotypiques avantageuses dans un environnement changeant. Plus de polyplœides s’établissent dans un environnement qui a eu un rôle important sur cette installation (Vanneste, Maere, et al. 2014).

2 Etude de la polyplœidie

L’évolution des technologies de séquençage de ces dernières années a permis d’étudier la polyplœidie et ses conséquences sur le plan de l’évolution et de l’expression des gènes. Les modèles utilisés se différencient par le type de polyplœidie et par l’âge de la polyplœidisation, et les enseignements distincts qui en découlent sont complémentaires (Yang et al. 2010). La polyplœidie peut être examinée selon le type de polyplœidie des plantes considérées : le WGD dans le cas d’autopolyploïdes, la fusion des génomes divergents dans le cas d’allopolyploïdes. Les mécanismes de régulation de l’expression des gènes entre ces deux situations pourraient être variables (Wang et al. 2004). Quand l’âge de la polyplœidisation est considéré, elle informe sur l’importance et le type de phénomènes d’évolution que certains polyplœides ont subi depuis le WGD ou l’hybridation des génomes à leur origine. On distingue les paléopolyploïdes (par exemple : *Zea mays*), pour lesquels le processus de fractionnement (perte progressive de copies de gènes dupliqués) est très avancé, les polyplœides formés plus récemment et enfin les nouveaux polyplœides formés il y a moins de 150 ans. Selon la date de leur formation, il est possible d’étudier directement l’évolution génomique du jeune allopolyploïde *in situ*, dans les populations naturelles actuelles (par exemple : *Tragopogon miscellus/mirus* 4x, *Spartina anglica* 12x, *Senecio cambrensis* 4x). Pour les espèces polyplœides plus

âgées, dont les progéniteurs ne sont pas tous précisément connus, il faut prendre en compte l'évolution des génomes parentaux depuis la formation du polyploïde. De plus, après la polyploïdisation et une phase de stabilisation, les génomes de ces espèces subissent le processus de fractionnement. Par conséquent, pour ces espèces, il est souvent nécessaire de produire des hybrides ou des polyploïdes synthétiques pour analyser l'évolution génomique qui fait suite à l'hybridation. Quelques exemples de polyploïdes plus âgés sont le cotonnier (*Gossypium hirsutum/barbadense* 4x), le blé dur (*Triticum turgidum durum* 4x), le blé tendre (*Triticum aestivum* 6x), le caféier (*Coffea arabica* 4x), le tabac (*Nicotiana tabacum* 4x), le colza (*Brassica napus* 4x), *Arabidopsis suecica* (4x).

Les hybrides et polyploïdes synthétiques constituent un excellent matériel génétique pour étudier les changements de la structure du génome et de l'expression des gènes qui accompagnent la formation d'un hybride ou d'un polyploïde, par comparaison avec les génotypes parentaux. Ils présentent plusieurs avantages par rapport aux espèces polyploïdes naturelles. Les génotypes parentaux sont connus, ce qui est le cas seulement pour les polyploïdes formés très récemment. On peut étudier les phénomènes qui suivent l'hybridation et la duplication (pour les polyploïdes) et cela pour plusieurs générations. L'étude des conséquences de l'hybridation (hybrides interspécifiques) et de la duplication (allopolyploïdes) peut être dissociée. Plusieurs événements indépendants peuvent être générés pour savoir si les modifications observées sont aléatoires ou reproductibles. Et enfin le sens des croisements des hybrides ou des allopolyploïdes peut être étudié.

3 Les effets immédiats de la polyploïdie

Au cours des dernières dizaines d'années, de nombreuses études portant sur des polyploïdes et/ou des néopolyploïdes ont démontré que l'hybridation et la duplication des génomes ont des effets à court et à moyen terme sur les génomes. Ces effets sont de différente nature : pertes de séquences, recombinaisons entre séquences homéologues, changements des profils épigénétiques, activation des transposons, changements de l'expression des gènes, modifications post-transcriptionnelles et protéiques.

3.1 Les modifications structurales du génome

3.1.1 Les changements génomiques

Chez les plantes, au cours de la méiose, des mécanismes de recombinaisons génomiques comme des crossing-overs (entrecroisements de chromatides ou de portions chromosomiques) réciproques ou des conversions de gènes (remplacement d'une région d'un chromosome par une région homologue, ou homéologue pour les allopolyploïdes) peuvent survenir en raison de l'appariement des chromosomes, ou de régions dupliquées. Ces phénomènes de recombinaison se

différencient par la taille des régions concernées, alors que les échanges réciproques peuvent se produire sur des portions importantes de chromosomes, la taille des conversions de gènes varie de quelques centaines à quelques milliers de paires de base (Gaeta et Chris Pires 2010). Les phénomènes de recombinaison peuvent avoir diverses conséquences pour les génomes des plantes diploïdes ou polyploïdes : délétions, duplications, échanges réciproques entre homéologues ou translocations homéologues. Au sein des allopolyploïdes, les phénomènes de translocations homéologues correspondent à des combinaisons de chromosomes produites au cours des générations suivant un évènement de recombinaison homéologue (Figure 5).

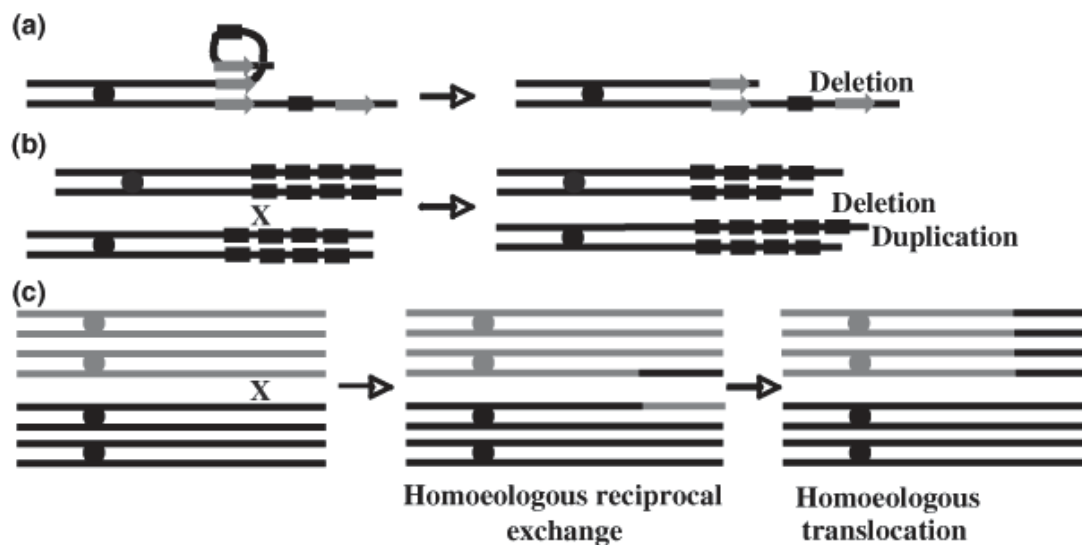


Figure 5 : Recombinaison et réarrangements chromosomiques. (a) Délétion due à une recombinaison intra-chromatide parmi des séquences dupliquées chez des diploïdes ou des polyploïdes. (b) Délétions et duplications causées par un crossing-over déséquilibrant le nombre de motifs répétés pour une région répétée sur un chromosome d'un diploïde ou d'un polyploïde. (c) Délétions et duplications intergénomiques qui résultent d'une recombinaison homéologues dans un allopolyploïde plusieurs générations précédemment d'après Gaeta et Pires (2010) (Gaeta et Chris Pires 2010).

Donc globalement, l'appariement et la recombinaison entre des chromosomes homéologues, ou des séquences paralogues contribuent aux délétions et aux translocations de gènes observées chez les polyploïdes, ils sont les mécanismes de la restucturation des génomes (Gaeta et Chris Pires 2010) et conditionnent l'histoire évolutive des polyploïdes. Par exemple, les translocations homéologues peuvent conduire à la désorganisation du mode d'hérédité disomique en augmentant la similarité entre chromosomes homéologues au cours des générations et en permettant l'appariement entre homéologues. Puisque le contrôle de l'appariement des chromosomes et la fertilité des polyploïdes sont souvent liés, il est alors possible d'imaginer que les polyploïdes disposent ou sont en cours d'acquisition d'un système de contrôle génétique de l'appariement des chromosomes (Gaeta et Chris Pires 2010).

Brassica napus fait partie des premiers polyploïdes pour lesquels les variations génomiques ont été étudiées. Gaeta *et al.* (2007) qui ont analysé plusieurs générations (S0-S5) d'une cinquantaine de lignées de polyploïdes synthétiques de *Brassica napus*, ont mis en évidence de nombreuses pertes de séquences d'ADN pour les générations les plus anciennes. Ces variations de type translocations homéologues, n'interviennent pas de façon aléatoire sur l'ensemble du génome. Leur fréquence est plus élevée aux extrémités des chromosomes. Pour ces polyploïdes, les variations génomiques sont corrélées avec des changements de l'expression de gènes (Gaeta *et al.* 2007). En 2011, une analyse cytogénétique menée sur le même matériel avec quelques générations supplémentaires a révélé des phénomènes d'aneuploïdie et d'importants remaniements inter et intragénomiques associés à une diminution de la fertilité et de la viabilité du pollen (Xiong *et al.* 2011). Szadkowsky *et al.* (2010) ont étudié la confrontation des génomes divergents au cours de la première méiose d'allopolyplloïdes synthétiques de *Brassica napus*. Ils ont observé que les réarrangements génomiques au moment de la première méiose entraînent des restructurations génomiques pour les générations suivantes (Szadkowski *et al.* 2010). A partir de la séquence du génome de *Brassica napus*, Chalhoub *et al.* (2014) ont confirmé l'abondance d'échanges homéologues entre les séquences des deux sous-génomes de ce polyploïde (Chalhoub *et al.* 2014).

Pour *Nicotiana tabacum*, c'est en comparant les séquences d'ADN répétées des génomes des polyploïdes naturels avec celles des génomes des espèces parentales diploïdes, que le phénomène de fractionnement des séquences répétées d'origine paternelle a été révélé à l'échelle du génome (Renny-Byfield *et al.* 2011). Dans une seconde étude, la présence de deux séquences hautement répétées au sein de polyploïdes naturels et synthétiques de *Nicotiana tabacum* ont été comparées, une d'entre elles a été préférentiellement éliminée des polyploïdes naturels (Renny-Byfield *et al.* 2012). Dans une étude plus récente Renny-Byfield *et al.* (2013) ont examiné deux espèces allopolyplloïdes naturelles, d'environ 5 millions d'années, *Nicotiana repanda* et *Nicotiana nudicaulis*, provenant de l'hybridation d'individus proches des espèces *Nicotiana sylvestris* et *Nicotiana obtusifolia*. Ces allopolyplloïdes présentent des tailles de génome divergentes. Alors que le processus de fractionnement des gènes dupliqués est en cours pour chacune d'elles, la divergence de taille de génome est liée à une accumulation et à une rétention différentielle des séquences d'ADN hautement répétées (Renny-Byfield *et al.* 2013).

Des phénomènes de pertes de séquences homéologues sont aussi rapportés pour *Tragopogon miscellus* (Tate *et al.* 2006, Tate *et al.* 2009, Richard J. A. Buggs *et al.* 2010). Le fractionnement pour ce jeune polyploïde semble avoir débuté rapidement (Tate *et al.* 2009) et se répète pour les différentes populations observés. Certaines classes de gènes ou bien des gènes sensibles à « l'effet dose » sont préférentiellement retenus (Tate *et al.* 2009, Buggs *et al.* 2012). Pour Chester *et al.* (2012) qui ont mené des analyses cytologiques sur *Tragopogon miscellus*, après 40 générations, des irrégularités chromosomiques et des phénomènes d'aneuploïdie sont encore visibles et peu de variations génomiques semblent fixées entre les populations (Chester *et al.* 2012). Des populations de

Tragopogon mirus, un allopolyploïde naturel récent entre *Tragopogon dubius* et *Tragopogon porrifolius*, ont été analysées par Koh et al. (2010), et ont montré elles aussi un processus d'élimination de séquences homéologues et préférentiellement pour certains loci. L'analyse conjointe des deux allopolyploïdes de *Tragopogon*, a indiqué que le fractionnement se produit en éliminant des séquences homéologues du parent partagé *Tragopogon dubius* (Koh et al. 2010).

Le blé est aussi un allopolyploïde pour lequel de nombreuses études (Feldman dans Soltis et Soltis 2012) ont observé l'élimination de séquences d'ADN à faible ou grand nombre de copies, codantes ou non-codantes. Cette élimination se produit pendant ou rapidement après la formation des polyploïdes (Eilam et al. 2008, Eilam et al. 2010) sur des chromosomes spécifiques ou sur une paire de chromosomes d'un sous-génome (Feldman et al. 1997, Feldman et Levy 2012).

Les modifications de la structure du génome ne concernent pas toutes les espèces polyploïdes, et tous les polyploïdes ne sont pas soumis aux mêmes phénomènes, pour la Spartine et le Coton, peu d'altérations sont rapportées, mais pour le coton, les événements d'échanges homéologues observés sont le résultat de conversions de gènes (Salmon et al. 2010). Les espèces qui ont subi des remaniements génomiques, montrent des conséquences importantes sur l'expression des gènes. Les modifications génomiques par délétion ou duplication, résultant de crossing-over méiotiques réciproques, induisent une variabilité du nombre de copies d'une même région génomique au sein du génome. L'expression des gènes peut donc être surexprimée (duplication du gène ou des séquences régulatrices) ou réprimée (délétion ou mutation de zones de régulation). Le déséquilibre de l'expression peut concerner un seul gène ou bien une cascade de gènes si le gène modifié est impliqué dans un réseau génique. Les effets de ces remaniements ne se limitent pas au génome, ils peuvent atteindre le niveau protéique et avoir des conséquences physiologiques.

3.1.2 Activation des éléments transposables (TE)

Les TEs constituent une fraction importante du génome des plantes. L'hybridation entre des génomes divergents et la duplication représentent un choc génomique qui pourrait activer les TEs. Ainsi l'insertion ou l'excision de TEs du génome peut conduire à un réarrangement de la structure des génomes chez les polyploïdes. Peu d'études menées sur des allopolyploïdes synthétiques ou de jeunes polyploïdes rapportent une transposition excessive des TEs après l'allopolyploïdisation. Le déclenchement de la mobilité des TEs n'est pas fréquent (Parisod et al. 2010), et il est limité à des familles spécifiques de TE (souvent de jeunes familles de TE) dans des allopolyploïdes particuliers : néopolyploïdes de *Nicotiana* (Petit et al. 2010), de blé (Kashkush et al. 2002). En revanche, ces études déclarent l'observation, dès l'allopolyploïdisation, de pertes de séquences liées à des TEs et de la modification des profils de méthylation. Ce qui peut impliquer pour le dernier point un impact sur l'expression des gènes et la stabilité des génomes. Pour *Spartina*, la perte de séquences de TE et

l'altération des profils de méthylation se produisent préférentiellement sur un sous-génome (Parisod et al. 2009). Pour des polyploïdes naturels et plus anciens, l'activation des TE ne semblent pas non plus fréquentes, parfois seulement observées que pour quelques transposons chez *Brassica* (Alix et al. 2008), *Gossypium* (Grover et al. 2008). Parisod et al. (2010) considèrent que la réorganisation des séquences de TE qui s'effectue après l'hybridation agiraient sur les premières étapes de l'évolution des allopolyploïdes (Parisod et al. 2010). Ils proposent deux hypothèses pour expliquer l'instabilité provoquée par les TEs. D'une part, les observations reflètent simplement ce qui se produit à l'échelle du génome. Le phénomène serait visible car les TEs représentent une grande portion des génomes. D'autre part, ils imaginent un modèle qui expliquerait l'activation des TEs due aux changements épigénétiques (Figure 6). Ils présentent le cas d'espèces parentales possédant des TEs réprimés par des petits ARNs, qui s'hybrideraient et subiraient une incompatibilité des facteurs de répression des TEs. Ainsi une réorganisation asymétrique de la répression des TEs se mettrait en place et l'apport en petits ARNs maternels ne seraient plus suffisant pour réprimer l'ensemble des TEs. Après l'hybridation entre deux génomes parentaux, qui peuvent se différencier par la nature de leurs TEs, des TEs maternels ou paternels pourraient échapper à la répression et être activés, et déclenchés des vagues de transpositions de TEs délétères (Parisod et al. 2010).

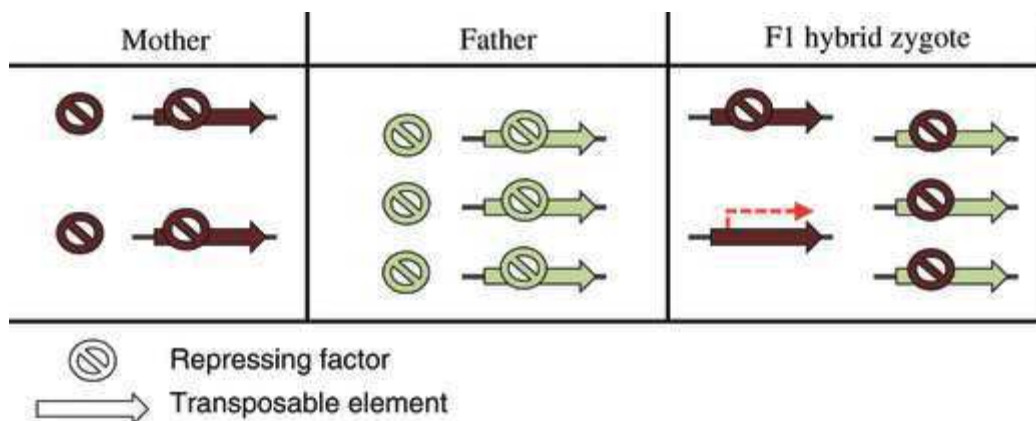


Figure 6 : Modèle de l'activation de TEs due à des différences épigénétiques et dépendant de l'effet dose, au sein d'un allopolyploïde (Parisod et al. 2010).

3.2 Les modifications épigénétiques

Les variations épigénétiques désignent les changements phénotypiques, morphologiques ou moléculaires qui sont héréditaires sans qu'il y ait eu une modification de la séquence du gène et de sa région régulatrice (Rapp et Wendel 2005). A partir des nombreuses études portant sur les variations épigénétiques, Ha (2013) propose le concept du modelage de la chromatine (chromatin remodeling). La chromatine qui est un complexe d'ADN, de protéines et d'ARN est la structure d'emballage du génome dans le noyau des cellules. C'est une structure dynamique avec parfois des particularités (méthylation des cytosines, histones modifiées) reconnues par des facteurs protéiques qui peuvent sous certaines conditions (stress environnemental ou lié à une attaque d'un pathogène...) induire la réplication et la transcription de l'ADN. Le modelage de la chromatine pourrait être le mécanisme moléculaire principal de l'épigénétique. Il pourrait être régulé par plusieurs phénomènes : la formation et la désagrégation des nucléosomes, les modifications des histones, la méthylation de la cytosine, les ARNs non-codants, séquences d'ADN spécifiques occasionnant une rigidité de la molécule d'ADN (Figure 7). L'analyse conjointe de l'ensemble de ces facteurs épigénétiques et génétiques pourraient révéler un code qui permettrait de comprendre les signaux de condensation et décondensation de la chromatine selon l'environnement des individus étudiés (Ha 2013). En raison du stress génomique dû à l'hybridation de deux génomes au cours de l'allopolyploidisation, les variations épigénétiques ont été examinées sur de nombreux polyploïdes. Jusqu'à présent les études portent la plupart du temps sur seulement un aspect épigénétique cité précédemment.

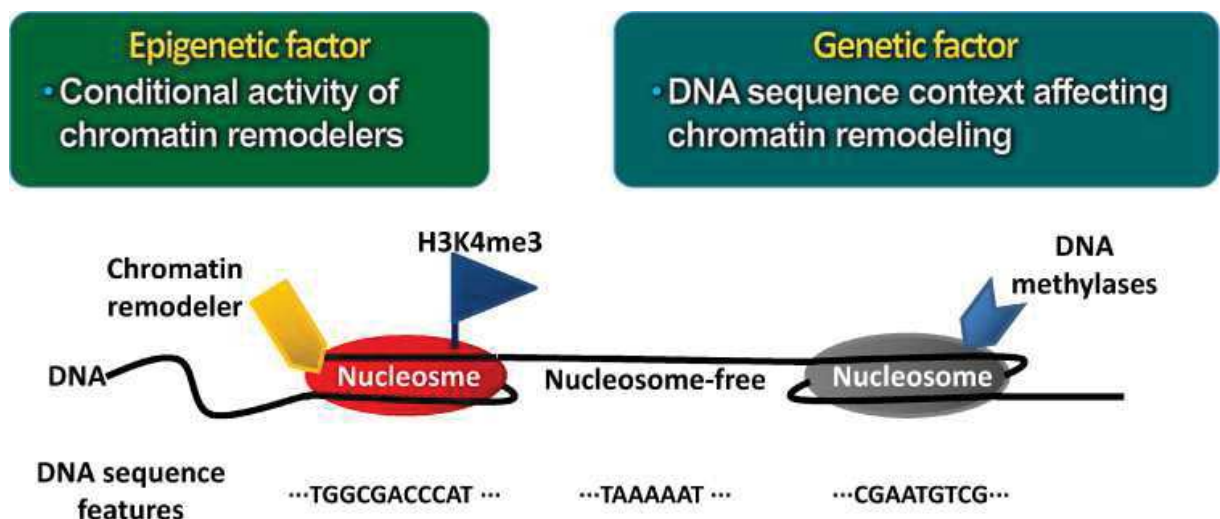


Figure 7 : Facteurs génétiques et épigénétiques du modelage de la chromatine (Ha 2013)

3.2.1 Variation de la méthylation de l'ADN

Les séquences d'ADN dont l'état de méthylation peut varier contiennent les motifs suivants : CG, CHG et CHH, H signale n'importe quel nucléotide autre que la guanine. Les variations des profils de méthylation des polyploïdes ont été examinées essentiellement au moyen de polyploïdes naturels récents et de polyploïdes synthétiques. Tout d'abord pour *Arabidopsis*, des altérations de méthylation de l'ADN de polyploïdes néosynthétiques ont été observées par rapport aux profils de méthylation des espèces parentales (Madlung et al. 2002). Il en est de même pour *Spartina*, Salmon et al. (2005) ont estimé que 30% des profils de méthylation des génomes parentaux (*Spartina alterniflora* et *Spartina maritima*) sont modifiés dans deux hybrides indépendants *Spartina towsendii* et *Spartina neyrautii* (Salmon et al. 2005).

L'ampleur des modifications de méthylation est très variable selon les espèces. Pour des colzas néosynthétiques, environ 6% des fragments étudiés par RFLP ont présenté une modification de leur profil de méthylation dès la première génération (Lukens et al. 2006). Certains auteurs ont ensuite analysé simultanément les variations des profils de méthylation et l'activité d'éléments transposables (TE) au sein de néopolyploïdes. Madlung *et al.* (2004) ont observé pour *Arabidopsis* que l'activité de certains TE est liée à l'état de méthylation de l'ADN aux loci correspondant (Madlung et al. 2004). Pour *Spartina*, Parisod et al. (2009) ont constaté que l'hybridation est accompagnée de plus de changements structuraux et de variations épigénétiques près des TE que la duplication. De plus ils ont observé que les variations épigénétiques apparaissent préférentiellement sur le sous-génome maternel induisant plus de changements génomiques pour ce sous-génome (Parisod et al. 2009). En 2009, Xu et al. (2009) ont observé eux aussi un biais parental, un des sous-génomes de néopolyploïdes de *Brassica napus* présentent plus de gènes réprimés et de modifications du profil de méthylation. Ils déduisent aussi de leur étude que les variations des profils de méthylation se produisent à l'échelle du génome (pas limitée aux régions présentant des TE) et immédiatement après l'hybridation (Xu et al. 2009). A partir de l'observation des variations épigénétiques et de l'activité de TE de néopolyploïdes de blé, Zhao et al. (2011) ont observé que les changements épigénétiques se produisent sur l'ensemble du génome et sont reproductibles entre plusieurs lignées. Ils font l'hypothèse que les changements épigénétiques héréditaires induisent une stabilité des génomes après l'hybridation et pourraient contribuer fortement à la spéciation des polyploïdes (Zhao et al. 2011). Salmon et al. (2005) ont souligné l'impact adaptatif des variations épigénétiques en expliquant la plasticité morphologique de *Spartina anglica* et son adaptabilité écologique par les variations épigénétiques (Salmon et al. 2005). Dernièrement une étude de l'état de méthylation de plusieurs organes de lignées de coton cultivé a révélé l'importance de la diversité de la méthylation et de son impact sur la régulation du développement de certains organes (Osabe et al. 2014).

3.2.2 Les modifications des histones

Les modifications des histones et les différentes formes d'histones figurent aussi parmi les variations épigénétiques. Elles ont un effet important sur la transcription des gènes chez les plantes (Deal et Henikoff 2011). Pour étudier l'impact de l'allopolyploïdisation sur l'expression des gènes, Wang et al. (2006) ont analysé la variation du début de la floraison pour des allopolyploïdes synthétiques et naturels d'*Arabidopsis*. Ils ont constaté que les variations du début de la floraison entre les allopoloïdes naturels et synthétiques coïncident avec des changements de la méthylation d'histone H3K4me2 et de la diméthylation et l'acétylation d'histone H3K9 (Wang et al. 2006). Dans une étude portant elle aussi sur des polyploïdes de synthèse d'*Arabidopsis*, Ni et al. (2009) ont montré que les régions promotrices de plusieurs gènes du cycle circadien présentent des modifications des histones. Ces changements affectent la transcription des gènes et pourraient expliquer l'accumulation plus importante d'amidon et sucre dans les allopolyploïdes par rapport aux espèces parentales (Ni et al. 2009).

3.3 Les modifications de l'expression des gènes

C'est sans doute l'effet de la polyploïdisation qui a été et est encore le plus étudié. C'est en raison de la plus grande valeur sélective et/ou adaptative des allopolyploïdes par rapport aux espèces parentales que l'expression des gènes est examinée au sein des allopolyploïdes. Les études visant à examiner l'expression des gènes de polyploïdes ont été menées indépendamment par deux approches différentes :

- la première consiste à comparer pour chaque gène le niveau d'expression de l'allopolyploïde avec les niveaux d'expression des espèces parentales pour ce même gène,
- la seconde compare le niveau d'expression de chaque gène homéologue pour un gène considéré, autrement dit c'est l'estimation de la contribution relative de l'expression de chaque homéologue à l'expression totale de l'allopolyploïde pour le gène considéré.

3.3.1 Expression des gènes des allopolyploïdes par rapport aux espèces parentales

Les études qui ont comparé l'expression des gènes des allopolyploïdes avec l'expression des gènes des espèces parentales se sont inspirées de modèles de génétique quantitative basés sur les notions d'additivité et de non-additivité (Figure 8).

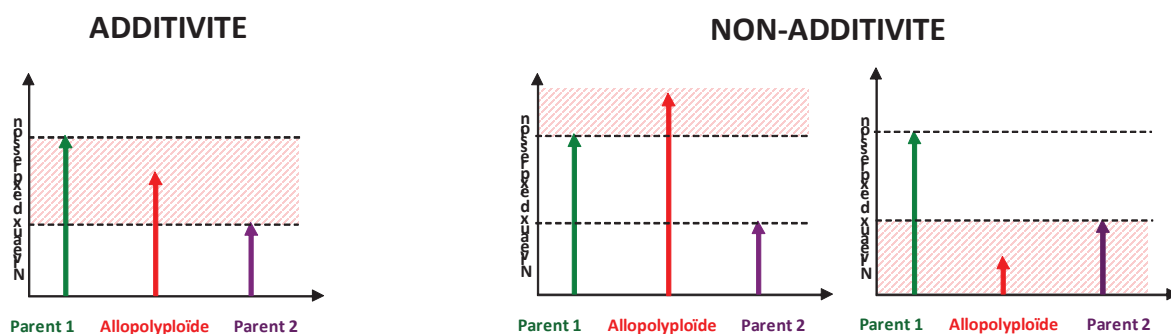


Figure 8 : Notions additivité et de non-additivité. Le niveau d'expression d'un gène chez les espèces diploïdes et parentales est représenté par une flèche. Les zones rouges hachurées représentent les zones d'expression de l'allopolyploïde en situation d'additivité (zone des niveaux d'expression correspondant à la moyenne des expressions des espèces parentales) ou de non-additivité (zone des niveaux d'expression ne correspondant pas à la moyenne des expressions des espèces parentales) par rapport aux espèces parentales.

- L'additivité se définit par le niveau d'expression d'un gène chez un allopolyploïde qui est égal à la moyenne de l'expression du même gène chez les parents diploïdes.
- La non-additivité se définit par le niveau d'expression d'un gène chez un allopolyploïde qui n'est pas égal à la moyenne de l'expression du même gène chez les parents diploïdes.

Les polyploïdes peuvent ainsi montrer des niveaux d'expression intermédiaires, supérieurs ou inférieurs à ceux de leurs parents.

Pour des allopolyploïdes synthétiques d'*Arabidopsis* (Wang, Tian, Lee, Wei, et al. 2006), et des allopolyploïdes de *Senecio* (Hegarty et al. 2006), la majorité des gènes étudiés présentent des profils additifs. En 2009, Rapp et al. étudient l'expression des gènes chez des polyploïdes synthétiques de *Gossypium*, une majorité de gènes sont aussi additifs. Pour les gènes non-additifs, qui pourraient être les gènes responsables des différentes capacités adaptatives des polyploïdes, Rapp et al. (2009) catégorisent l'expression des gènes en dominance et transgressivité, et sous-divisent ces catégories suivant le niveau d'expression du gène de l'allopolyploïde par rapport à celui des espèces parentales (catégorie dominance et catégorie transgressivité) (Figure 9).

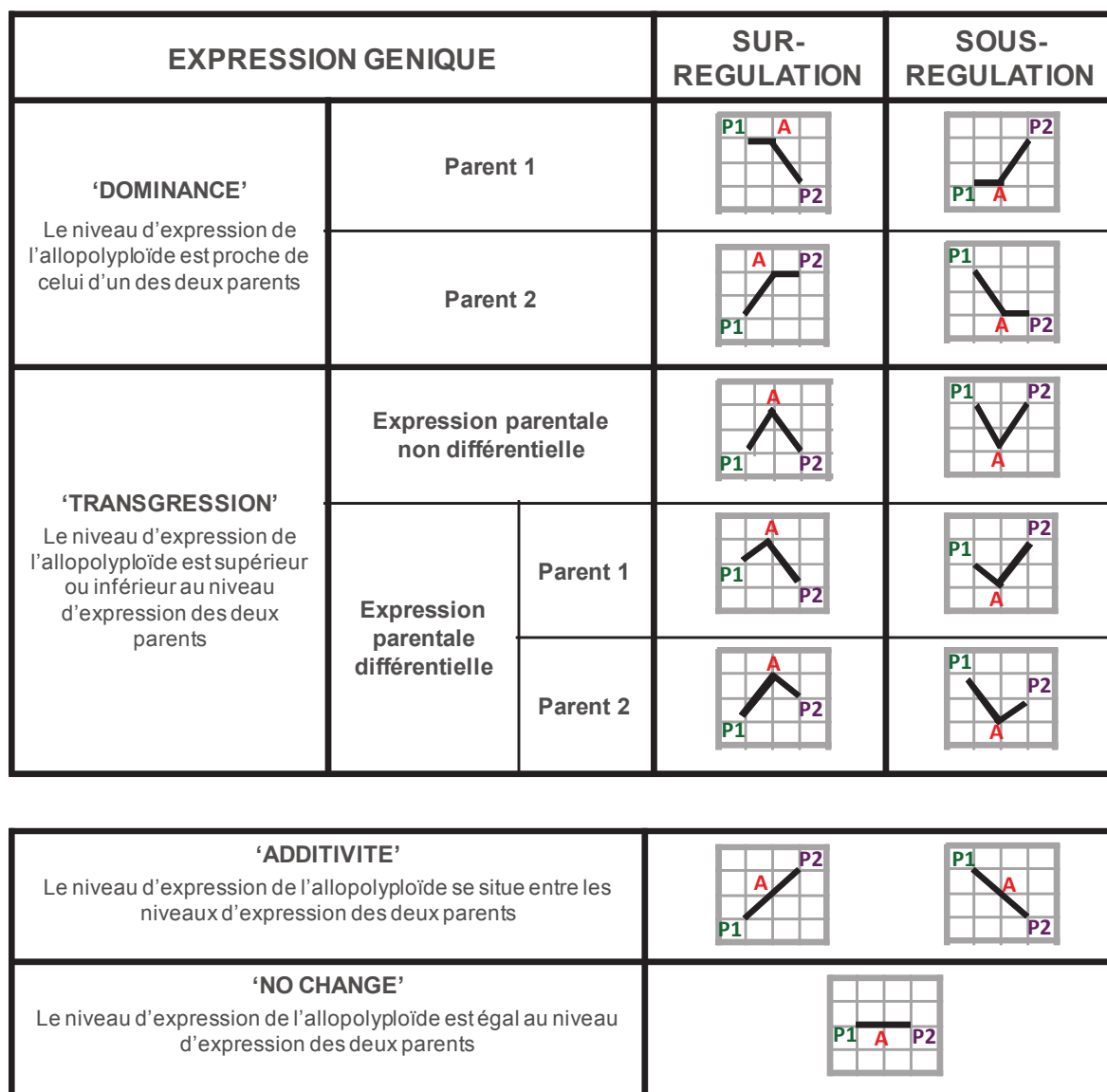


Figure 9 : Profils d’expression de gènes d’un allopolyploïde comparé à ceux de ses espèces parentales. P1 et P2 représentent les espèces parentales et A représente l’espèce allopolyploïde. Les quatre catégories sont représentées : deux catégories d’expression non-additive ‘dominance’, ‘transgression’, la catégorie d’expression additive ‘additivité’ et la catégorie dans laquelle sont rassemblées les gènes ne présentant pas de variation d’expression entre les parents diploïdes et l’allopolyploïde ‘No change’. Pour les catégories de ‘dominance’ et de ‘transgression’, des subdivisions sont faites en fonction de la dominance vers l’un ou l’autre des deux parents et/ou en fonction de la sur-expression ou sous-expression du gène (d’après Rapp et al. 2009).

Par cette méthode, Rapp et al. (2009) observent et décrivent la « genomic expression dominance », c'est-à-dire qu’une grande majorité des gènes non-additifs ont un niveau d’expression comparable à celui d’une espèce parentale, qu’elle soit plus ou moins exprimée par rapport à l’autre espèce parentale (Rapp et al. 2009).

En suivant cette méthode, Bardil et al. (2011) ont comparé l’expression des gènes de l’allopolyploïde *Coffea arabica* à ses espèces parentales et cela pour plusieurs températures de culture. Ils constatent comme pour le coton, qu’une majorité de gène non-additifs s’exprime dans l’allopolyploïde comme les gènes d’une des espèces parentales. Mais ils observent que ce phénomène « d’expression

génomique dominante » est sensible à la température, en effet selon la température de culture des allopolyploïdes, le nombre de gènes s'exprimant pareillement à l'une ou l'autre des espèces est variable. Ils estiment que ce phénomène fournit des indices pour l'explication de la plasticité des allopolyploïdes (Bardil et al. 2011).

3.3.2 Contribution des gènes homéologues à l'expression totale des gènes

Cette seconde approche est particulièrement difficile à mettre en œuvre, car elle nécessite une méthode qui permette de différencier les transcrits des gènes homéologues des transcrits de gènes homologues (sachant que les paires de gènes homologues peuvent présenter des variations alléliques). En raison de cette contrainte, cette étude a tout d'abord été menée à l'échelle d'un gène et d'une centaine de gènes pour plusieurs polyploïdes naturels. Ces travaux ont permis d'observer que la contribution des gènes homéologues à l'expression totale d'un gène pouvait être biaisée, c'est-à-dire qu'un homéologue peut contribuer plus à l'expression totale du gène. Ce fut le cas pour le blé (Mochida et al. 2003), (Nomura et al. 2005), (Stamati et al. 2009), pour le coton (Adams et al. 2003). Parfois les auteurs décrivent des biais extrêmes, qui vont jusqu'à l'absence de la participation d'un gène homéologue à l'expression totale de gènes, par exemple pour *Tragopogon miscellus* (Richard J. A. Buggs et al. 2010) et le blé (Mochida et al. 2003). Ensuite sur le coton de nombreuses études ont été menées pour examiner la variation de l'expression relative des gènes homéologues. Différents organes de polyploïdes naturels de coton exposés à des stress abiotiques ont révélé que la contribution des gènes homéologues peut varier en fonction des gènes, des organes et des stress (Adams et al. 2004, Liu et Adams 2007, Liu et al. 2011, Dong et Adams 2011). De plus il a été montré que pour certaines régions génomiques étudiées, le profil de la variation de l'expression relative des gènes homéologues proches se répète pour différents organes, laissant penser à une coordination de la régulation des gènes par des mécanismes à l'échelle de la région génomique. Pour d'autres régions, le profil de la variation de l'expression relative des gènes homéologues proches semble indépendant de la distance physique entre les gènes (Flagel et al. 2009). Pour ces différents auteurs, la variation de l'expression des gènes homéologues peut être un signe de sous-fonctionnalisation, ce phénomène maintient la rétention de ces gènes à l'état dupliqué au cours de l'évolution (Liu et Adams 2007, Dong et Adams 2011). En 2009, Chaudhary et al. ont comparé des polyploïdes naturels, synthétiques de coton, des hybrides F1 et les espèces parentales. Ils ont attribué le biais de la contribution des gènes homéologues à des différences de *cis*-régulation entre les espèces parentales (Chaudhary et al. 2009).

Plus récemment, les avancées technologiques et surtout les méthodes de séquençage de nouvelle génération (NGS) ont permis d'examiner l'expression des gènes homéologues à l'échelle génomique. Suite à l'observation de polyploïdes synthétiques de *Brassica napus* (Higgins et al. 2012) et de coton (Flagel et al. 2008), il est apparu que la modulation de l'expression des gènes homéologues

pouvait se produire immédiatement après la polyploïdisation et aussi se poursuivre au cours de l'évolution. Pour *Tragopogon mirus*, la répression de certains homéologues dans certains organes dénotant une initialisation de sous-fonctionalisation est aussi observé rapidement (Richard J A Buggs et al. 2010). Pour des polyploïdes de *Glycine*, plus anciens, l'utilisation de gènes dupliqués dans différentes voies biologiques confirme la sous-fonctionalisation (Ilut et al. 2012).

Mais l'augmentation du nombre d'études traitant de ce sujet et la variabilité du vocabulaire employé par les différents auteurs amènent Grover et al. (2012) à définir précisément les termes pour décrire les phénomènes observés sans confusion. Ainsi les deux phénomènes observés par chacune des deux approches sont décrits et schématisés (Figure 10) (Grover et al. 2012) :

- Dominance du niveau d'expression (ou expression level dominance) lorsqu'un gène de l'allopolypléide ne s'exprime pas de façon additive, mais s'exprime comme l'un des gènes des espèces parentales (plus ou moins exprimé par rapport à la moyenne de l'expression parentale), (l'approche analytique est la comparaison des niveaux d'expression d'un gène entre les espèces parentales et l'allopolypléide, cf paragraphe 2.3.1)
- Biais de l'expression des homéologues (ou homeolog expression bias), un des homéologues contribue de façon majoritaire à l'expression totale des gènes de l'allopolypléide. (dans ce cas-là, l'approche analytique est la détermination de la contribution de l'expression des homéologues à l'expression totale d'un gène pour l'allopolypléide, cf paragraphe 2.3.2)

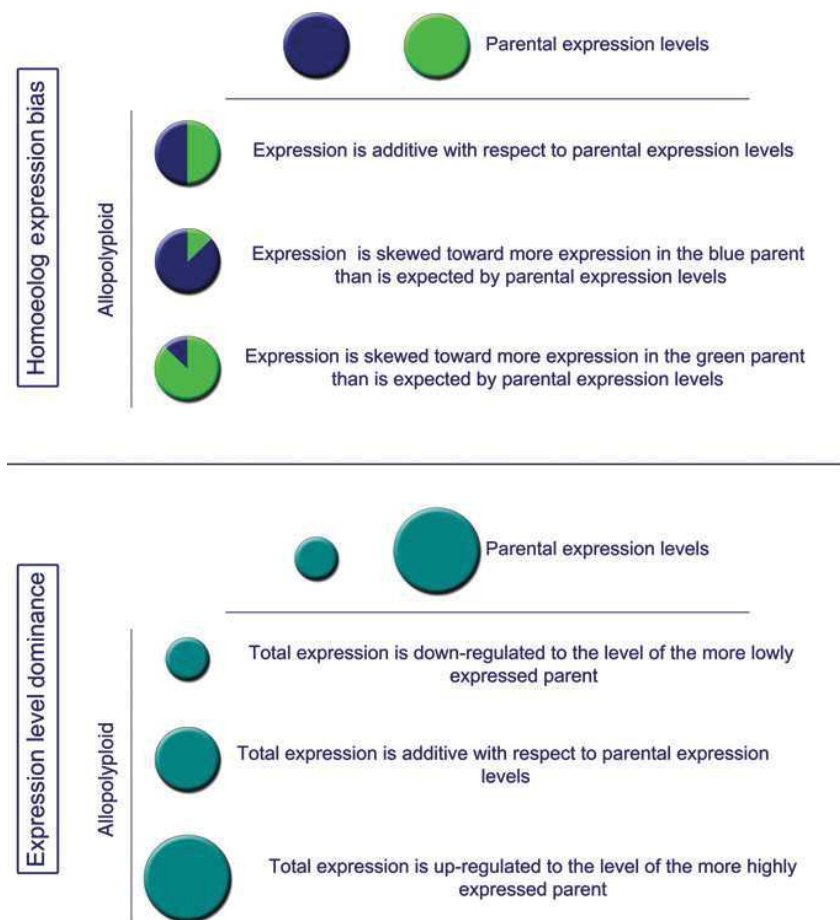


Figure 10 : Représentation des phénomènes principaux révélés par les deux approches d’analyse de l’expression des gènes au sein d’un allopolyploïde. Ces schémas représentent l’expression d’un couple de gènes homéologues. Partie haute : les niveaux d’expression des gènes homéologues peuvent être équivalents ou biaisés vers l’un ou l’autre des parents. Partie basse : l’expression globale des deux homéologues peut être plus faible, intermédiaire ou plus forte par rapport à l’expression des gènes orthologues dans les espèces parentales (Grover et al. 2012).

Dans une étude portant sur des allopolyploïdes synthétiques et naturels de *Gossypium*, mais aussi sur des hybrides interspécifiques F1 des espèces parentales du coton, Yoo et al. (2012) ont analysé l’expression des gènes homéologues par les deux approches et ont proposé un modèle du phénomène de « l’expression level dominance ». Ils ont observé que pour des gènes non-additifs, la dominance du niveau d’expression orientée vers un parent s’expliquait par la sur ou sous-régulation de l’autre parent (Figure 11) (Yoo et al. 2012).

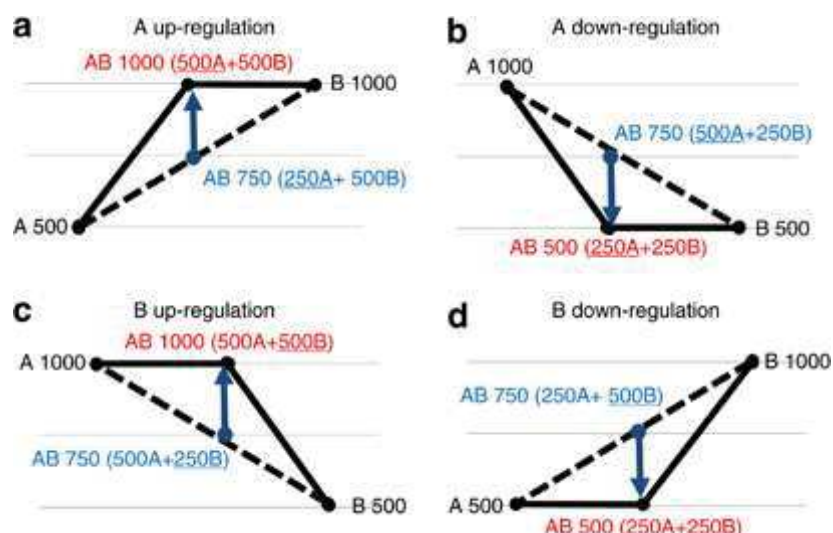


Figure 11 : Explication du phénomène de dominance du niveau d'expression : l'expression additive d'un gène d'un allopolyploïde AB qui maintient le niveau parental d'expression est représentée en bleu. Les situations de dominance du niveau d'expression soit orientées vers B sont schématisées en a (expression level dominance up) et en b (expression level dominance down), soit orientées vers A sont schématisées en c (expression level dominance up) et en d (expression level dominance down). A partir des résultats de l'analyse de la contribution des gènes homéologues dans chacune de ces situations, Yoo et al. (2012) observent fréquemment la sur-régulation ou la sous-régulation du parent non-dominant dans les cas d'expression level dominance up ou expression level dominance down (Yoo et al. 2012).

Enfin dans une étude d'allopolyploïdes synthétiques d'*Arabidopsis*, Shi et al. (2012) ont analysé les modifications de l'expression des gènes homéologues en comparant leurs niveaux d'expression à celui des gènes orthologues chez les espèces parentales. Alors qu'ils observent que la divergence d'expression entre les espèces parentales provient de façon majoritaire de divergences en *cis*-régulation, ils constatent qu'au sein des polyploïdes F1, les facteurs de *trans*-régulation provenant des espèces parentales affectent différemment l'expression des homéologues (Shi, Danny W.-K. Ng, et al. 2012).

3.4 Les étapes de modification et régulation post-transcriptionnelles

Les variations de la transcription des gènes sont sans doute un moyen rapide et efficace des plantes pour s'adapter à un environnement. Mais des étapes de régulation qui seraient postérieures à la transcription pourraient aussi contribuer à la réponse immédiate d'une plante à l'environnement ou à un stress. Ces mécanismes de régulation qui interviendraient sur des produits plus proches des protéines fonctionnelles dans la chaîne des processus moléculaires entre le génome et le phénotype pourraient permettre une réponse de la plante plus rapide. Par conséquent d'autres processus moléculaires complémentaires à la transcription (l'épissage alternatif, la traduction) et des éléments de régulation sont considérés et de plus en plus étudiés.

3.4.1 L'épissage alternatif au sein de polyploïdes

L'épissage alternatif est un processus complexe de régulation post-transcriptionnelle dont la fonction n'est pas encore bien connue mais qui vient s'ajouter aux mécanismes de régulation de l'expression des gènes déjà repérés (Syed et al. 2012, Reddy et al. 2013). Sous l'action de signaux environnementaux et /ou développementaux, ce processus peut réguler l'expression des gènes. Il peut diminuer le niveau d'expression d'un gène ou affecter la stabilité d'ARNm. Effectivement l'épissage alternatif peut provoquer la dégradation prématurée d'un transcript parce qu'il est lié aux processus de régulation des gènes par dégradation des transcrits non-sens (NMD nonsense-mediated decay) et par l'intermédiaire de petits ARN. Il peut augmenter la diversité protéique. En effet plusieurs ARNm matures peuvent être produit à partir d'un même gène. Parfois les protéines induites par ce processus peuvent être fonctionnelles. Par ces mécanismes, l'épissage alternatif peut entraîner des changements phénotypiques et impacter des processus physiologiques importants (Figure 12) (Zhou et al. 2011, Syed et al. 2012, Reddy et al. 2013). Par exemple, il a été démontré chez *Arabidopsis thaliana*, que l'épissage alternatif modifié par une élévation de la température pendant la culture intervenait dans la régulation du cycle circadien de la plante (James et al. 2012). Dernièrement, dans une étude visant à mieux comprendre comment l'épissage alternatif pouvait modifier l'expression des gènes, Palusa et al. (2015) ont montré chez *Arabidopsis thaliana*, qu'au cours du développement, les ARNm alternatifs étaient différemment retenus pour les polysomes en réponse à l'application d'un stress abiotique (Palusa et Reddy 2015).

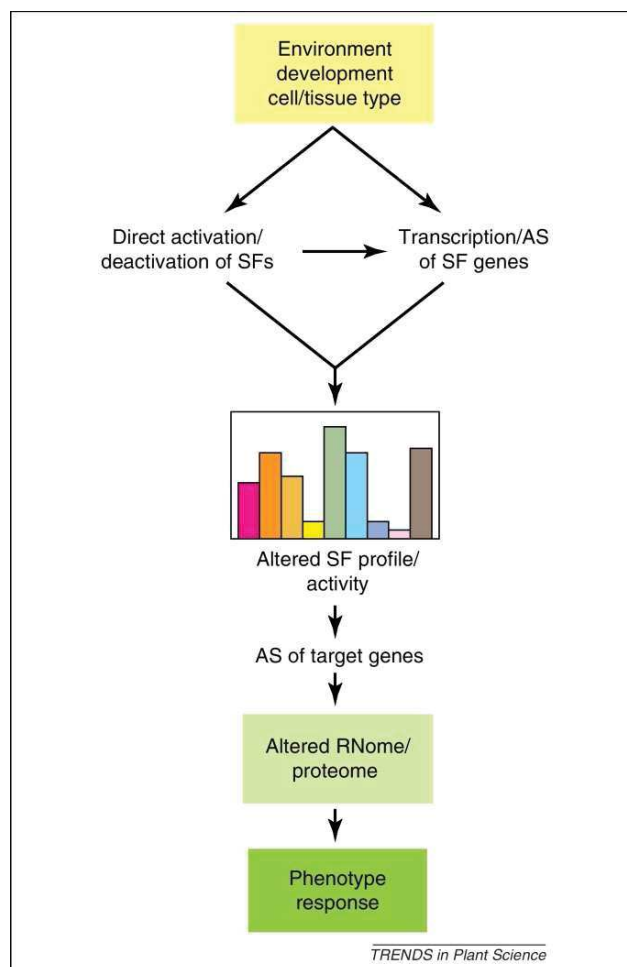
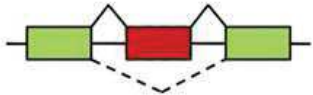
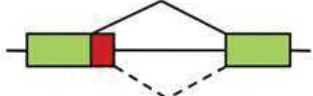


Figure 12 : Schéma illustrant l'impact de l'épissage alternatif sur la régulation de l'expression des gènes et sur le phénotype. Des signaux environnementaux et développementaux activent des facteurs d'épissage qui vont provoquer des changements dans la transcription des facteurs d'épissage. Ces derniers vont agir sur des gènes cibles en entraînant l'épissage alternatif. Les changements du protéome vont modifier le phénotype. (Syed et al. 2012).

Plusieurs types d'épissage alternatif ont été répertoriés. Ce phénomène étudié à l'échelle du génome pour deux plantes modèles, *Arabidopsis* et le riz, a révélé que chez les plantes la rétention d'intron est le phénomène dominant (Figure 13) (Wang et Brendel 2006, Reddy et al. 2013). Toujours chez les plantes on estimait qu'environ 33% des gènes pouvaient présenter des épissages alternatifs (Zhou et al. 2011), mais avec l'évolution des technologies permettant d'étudier le phénomène à l'échelle du génome, on estime maintenant que 60% des gènes multi-exoniques peuvent subir un épissage alternatif (Reddy et al. 2013). Mais l'estimation de la part des gènes présentant des événements d'épissage alternatif peut être difficile. En effet l'épissage alternatif des gènes peut être très variable selon les espèces, selon l'environnement et les tissus examinés. Chez *Arabidopsis thaliana* (Zhang et al. 2010) et un allopolyploïde naturel de *Brassica napus* (Zhou et al. 2011), les profils d'épissage alternatif ont montré des variations en fonction des organes étudiés et des stress abiotiques appliqués. Chez *Citrullus lanatus*, la pastèque, les profils d'épissage alternatif au sein

d'individus polyploïdes sont variables selon les organes au cours du développement (Saminathan et al. 2015).

	Alternative splicing event	Human	<i>Arabidopsis thaliana</i>
	Exon skipping/ inclusion	>40%	~ 8%
	Alternative 3' splice site	~18.4%	~15.5%
	Alternative 5' splice site	~7.9%	~7.5%
	Intron retention	<5%	~40%

 Constitutive region  Alternative region

Figure 13 : Fréquence et types d'épissage alternatif fréquent chez l'homme et chez *Arabidopsis thaliana*. (Reddy et al. 2013).

En outre, les profils d'épissage alternatif peuvent être modifiés par la polyploïdie. Pour *Arabidopsis thaliana*, des divergences importantes entre les profils d'épissage alternatif de gènes dupliqués par un événement ancien de polyploïdie ont été rapportées (Zhang et al. 2010). Pour *Brassica napus*, la comparaison des profils d'épissage d'allopolyplloïdes synthétiques et de ceux des espèces parentales a indiqué que, pour 20% des gènes, l'épissage alternatif était immédiatement affecté par l'allopolyplloïdisation, parfois différemment selon les homéologues et aussi de façon parallèle entre différents allopolyplloïdes synthétiques. Ce dernier point traduirait la mise en place de l'épissage alternatif des gènes de manière non-stochastique. Cette étude a aussi révélé qu'au cours de l'évolution, les variations des profils d'épissage alternatif s'atténuaient, en effet les allopolyplloïdes naturels *Brassica napus* présentaient moins de cas d'épissage alternatif que les allopolyplloïdes synthétiques (Zhou et al. 2011). Dernièrement une étude de l'épissage alternatif a été réalisée à l'échelle du génome de *Glycine max* pour plusieurs organes à différents stades de développement. Cette étude a révélé que 63% des gènes multi-exoniques présentaient des événements d'épissage alternatif, sans que la plante ait subi un stress au moment de la culture. Pour cette espèce également,

l'épissage alternatif est variable selon les organes étudiés, et il semblerait que ces événements se produisent plus pour les jeunes stades de développement (Shen et al. 2014).

Considérés globalement, les travaux déjà réalisés pour étudier l'épissage alternatif dans un contexte polyploïde, indiquent que ce phénomène impactant la régulation de l'expression de certains gènes, est rapidement affecté par la polyploïdisation. Il pourrait ainsi contribuer à l'évolution fonctionnelle (perte ou nouvelle fonction) de certains gènes et préserver certains gènes à l'état dupliqué (Zhang et al. 2010, Zhou et al. 2011, Shen et al. 2014).

3.4.1 Le translatome et la polyploïdie

Pour mieux évaluer l'impact de la régulation de l'expression des gènes au moment de la traduction, des études ont considéré le translatome, c'est-à-dire la part des ARNm à l'échelle du génome, qui est reliée aux polysomes pour la traduction. Pour l'allopolyploïde *Glycine dolichocarpa*, globalement la régulation du translatome n'induit pas d'importantes variations de l'expression des gènes. En effet, Coate et al. (2014) ont révélé que le transcriptome et le translatome étaient fortement corrélés même si un ¼ du transcriptome semblait régulé au moment de la traduction. De plus, peu de gènes, 1 à 5 % des gènes pour lesquels l'expression relative des gènes homéologues a pu être analysée (environ 4000 gènes), ont révélé une différence de régulation de la traduction entre homéologues. Pour ces gènes, la régulation de la traduction atténuait le biais d'expression relative des homéologues qui pouvait provenir de dérèglements de la régulation dus à la fusion des génomes lors de l'allopolyploïdisation. Enfin pour les autres gènes, la régulation de la traduction s'est faite de façon combinée sur les deux sous-génomes (Coate et al. 2014).

Dans une étude portant sur des hybrides entre *Saccharomyces cerevisiae* et *Saccharomyces bayanus*, Wang et al. (2015) ont aussi constaté que la traduction est plus conservative que la transcription. En effet la comparaison des translatomes des espèces parentales et des hybrides a révélé moins de divergences que la comparaison de leurs transcriptomes (Wang et al. 2015). De plus l'analyse de l'hérédité de la traduction des gènes au sein des hybrides est plus conservative que celle de l'hérédité de l'expression des gènes, c'est-à-dire qu'elle révèle moins de profils non-additifs au sein des hybrides par rapport aux espèces parentales (Wang et al. 2015).

3.5. Les modifications du protéome

Jusqu'à récemment, peu d'études s'étaient intéressées aux variations de l'expression des protéines chez des allopolyploïdes, peut être en raison de la difficulté de la mise en œuvre des méthodes de quantification de l'expression des protéines. Pourtant des travaux portant sur des colzas synthétiques ont montré que des variations importantes d'expression se font spécifiquement au niveau du protéome (Albertin et al. 2006, Marmagne et al. 2010). L'étude comparative des protéomes des premières générations de colzas synthétiques avec ceux de leurs parents diploïdes, a montré que plus d'un tiers des protéines exprimées présentent une non-additivité par rapport au niveau attendu provenant des parents. L'hybridation interspécifique serait principalement à l'origine de ces variations puisque 89% des variations ont été détectées dès l'hybride F1, le reste des variations sont intervenues au cours des générations suivantes (Albertin et al. 2006). L'essentiel des variations observées s'est avéré reproductible entre des lignées de colzas synthétiques obtenues indépendamment à partir des mêmes parents ce qui indique que la régulation post-transcriptionnelle de l'expression des gènes se fait de manière non-stochastique. Par contre, la nature des protéines présentant un écart à l'additivité n'est pas liée à la catégorie fonctionnelle ni à la localisation cellulaire (Albertin et al. 2007). Une étude de l'expression d'une centaine de gènes codant des protéines non-additives à partir des colzas synthétiques ont révélé que deux tiers des gènes ne montraient pas une modification dans leur expression à l'étape de transcription. Ainsi les régulations post-transcriptionnelles semblent être responsables d'une partie de la non-additivité au niveau du protéome et paraissent importantes (Marmagne et al. 2010).

Dernièrement des études ont été menées sur le protéome de modèles polyplœides supplémentaires en utilisant de nouvelles technologies (iTRAQ et spectrométrie de masse). Une étude portant sur la comparaison des protéomes d'auto et allopolyploïdes d'*Arabidopsis* et de leurs parents en lien avec leurs divergences transcriptomiques connues, a aussi révélé des changements rapides entre la transcription et la traduction qui pourraient être dus à la régulation post-transcriptionnelle (Ng et al. 2012). Pour des allopolyploïdes naturels et synthétiques de *Tragopogon mirus*, la comparaison de leurs protéomes avec celui des espèces parentales a montré que les variations du protéome par rapport aux espèces parentales étaient plus liées à l'hybridation qu'à la polyplœidisation (Koh et al. 2012). Pour un hexaploïde de *Brassica*, Shen et al. (2015) ont aussi observé que la variation de la régulation du niveau d'expression des protéines se produisait après l'allopolyploïdisation, pour environ $\frac{1}{4}$ des protéines étudiées (soit 452 protéines), le niveau d'expression entre l'allopolyploïde synthétique et ses espèces parentales étaient différents. Mais de plus ils ont aussi constaté que le niveau d'ARNm et celui de protéine n'étaient pas corrélés, ce qui suggérerait l'existence de phénomènes de régulation post-transcriptionnelle (Shen et al. 2015). Enfin très récemment, l'analyse protéomique de deux allopolyploïdes naturels de *Gossypium* et des espèces parentales a montré que bien qu'ayant des origines similaires, les deux allopolyploïdes présentaient des divergences protéomiques, ce qui

suggère que les conséquences de l'allopolyploïdisation sur leurs protéomes et par la suite sur leurs évolutions ont été différentes (Hu et al. 2015).

3.6 Les modifications transcriptionnelles par l'intermédiaire des petits ARN non codants

Les plantes contiennent une variété de petits ARN non codants, parmi lesquels figurent les miRNAs (microRNAs), les siRNAs (small interfering RNAs) et les ta-siRNAs (trans-acting small interfering RNAs). Du fait de la diversité des fonctions biologiques dans lesquelles les petits ARN sont impliqués, leur classification a généralement pris en compte comme premier critère leur mode de biogénèse. La classification la plus récente des petits ARN chez les plantes a été proposée par Axtell (Axtell 2013) (Figure 14). Les petits ARN sont impliqués dans le développement de la plante, les réponses au stress, ou les mécanismes de défense contre les virus et autres agents pathogènes et le maintien de la stabilité de régions hétérochromatiques dans le génome. Depuis quelques années, certaines études s'intéressent à l'effet de la polyploïdisation sur ces populations de petits ARN.

Dans une étude comparant les populations des petits ARN des allopolyploïdes synthétiques, naturels d'*Arabidopsis suecica* et des espèces parentales, Ha et al. (2009) ont observé que les siRNAs sont stables dans l'allopolyploïde naturel et ont conclu que les siRNAs des espèces parentales sont compatibles et permettent le maintien de l'intégrité du génome de l'allopolyploïde en réprimant les transposons (Ha et al. 2009).

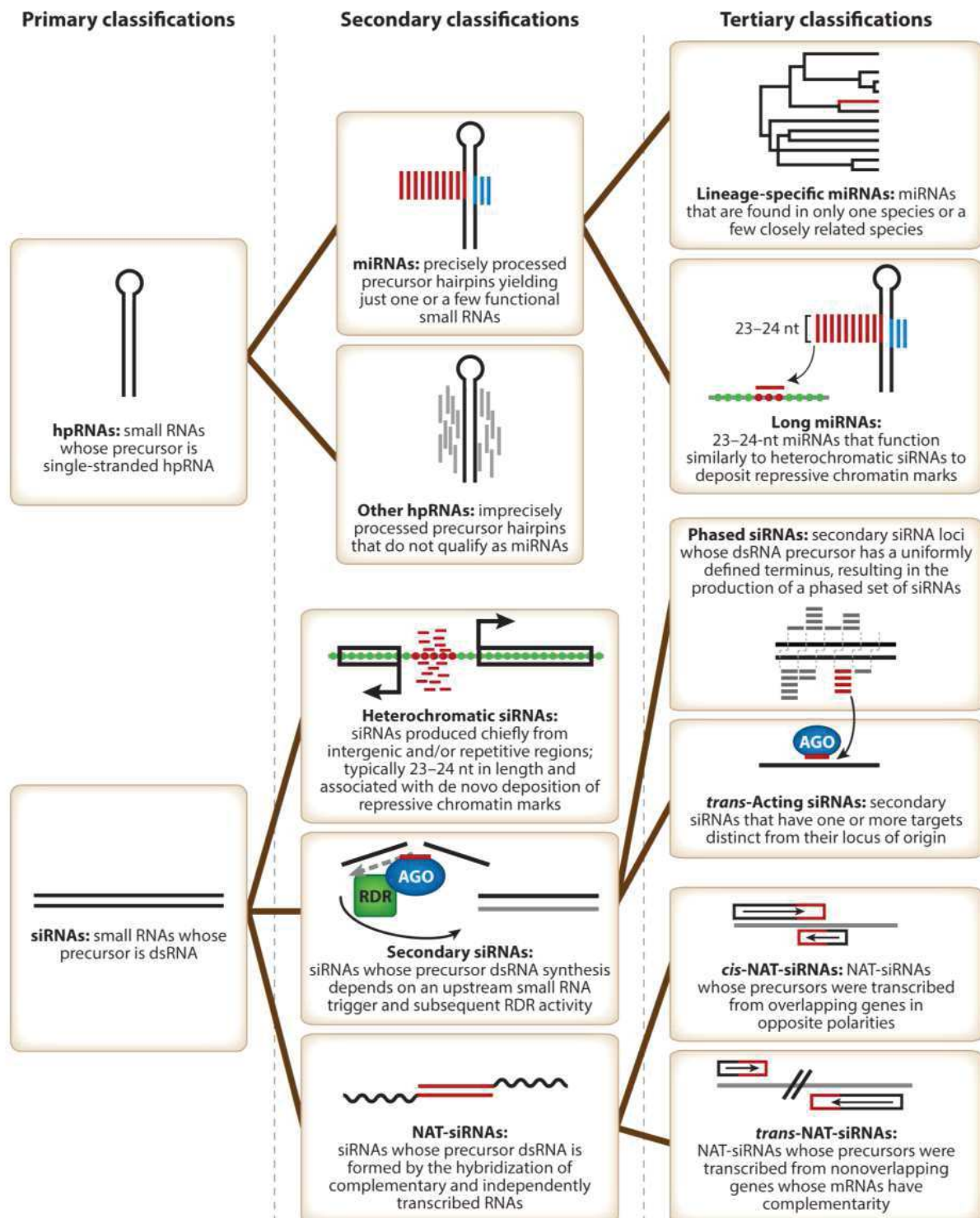


Figure 14 : Classification hiérarchique des petits ARN non codants chez les plantes (Axtell et al. 2013).

A l'inverse, les miRNAs et les ta-siRNAs ont subi des variations après la polyploïdisation, qui diminuent au cours de l'évolution. Ces variations pourraient être impliqués dans les changements adaptatifs des polyploïdes. A l'inverse, Kenan-Eichler et al. (2011) rapportent dans leur étude des petits ARN chez le blé en fonction de l'augmentation du niveau de polyploïdie, que le nombre de siRNAs diminue avec le niveau de ploïdie. Ils suggèrent que chez le blé, la polyploïdisation ou

l'hybridation pourraient entraîner la déstabilisation du génome (Kenan-Eichler et al. 2011). Dans une étude des petits ARNs d'allopolyploïdes synthétiques de colza, Marinez Palacios et al. (2014) démontrent que les miARNs sont globalement peu affectés, mais les siRNAs, qui ciblent spécifiquement les éléments transposables, présentent des variations d'expression spécifiques des familles des TE (Martinez Palacios 2014).

4 Les conséquences de la polyploïdie

4.1 Hétérosis, vigueur hybride

Depuis des décennies, il est bien connu que des hybrides peuvent mieux pousser, plus rapidement, avoir de meilleurs rendements et une meilleure fertilité que leurs espèces parentales. L'augmentation des capacités de ces plantes les rend forcément très intéressantes d'un point de vue agronomique, particulièrement dans l'amélioration variétale. Cette vigueur hybride ou hétérosis serait en grande partie due à l'hétérozygotie de l'hybride qui provient de l'hybridation entre deux individus génétiquement différents (Birchler et al. 2010). Les polyploïdes présentent aussi cette vigueur hybride et sont considérés comme des hybrides particuliers entre des espèces parentales génétiquement différentes. Chez les allopolyploïdes, l'hétérozygotie serait en quelque sorte fixée en raison de la présence fréquente des copies homéologues (Chen 2010). L'hétérozygotie est estimée entre génomes pour les polyploïdes, et entre allèles pour un hybride diploïde.

Malgré son importance dans l'agriculture, les mécanismes de la vigueur hybride chez les hybrides et les polyploïdes sont peu connus. Mais grâce aux avancées technologiques et au nombre croissant d'études sur les polyploïdes, le rôle de l'expression non-additive des gènes, de la variation de l'expression allélique, des petits ARNs commence à être décelé (Chen 2010, Blum 2013). Parmi ces études, Ni et al. (2009) ont montré sur des allopolyploïdes d'*Arabidopsis* que la variation épigénétique entre les espèces parentales diploïdes concernant les gènes du cycle circadien pourrait expliquer la vigueur hybride et l'accumulation de biomasse des polyploïdes (Ni et al. 2009) (Figure 15).

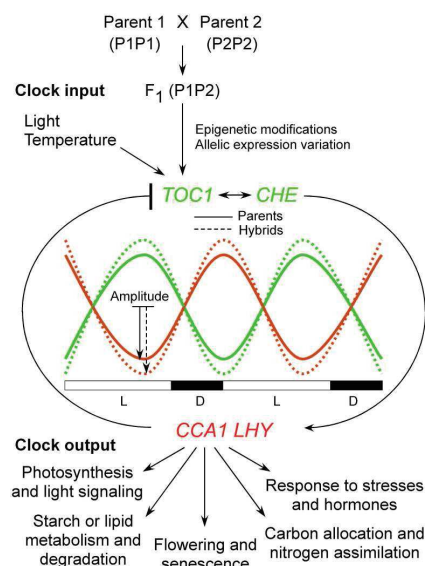


Figure 15 : Un modèle de mécanisme moléculaire de la vigueur hybride. Les horloges internes des plantes sont contrôlées par plusieurs boucles feedback, dont celle des deux répresseurs de transcription CCA et LHY combinés avec deux régulateurs TOC1 et CHE. Les plantes reçoivent des signaux (lumière, température) et contrôlent des caractères (photosynthèse, réponse aux stress...). Dans les hybrides, les interactions alléliques provoquent la répression épigénétique de CCA et LHY et la sur-régulation de TOC1. Cette situation induit l'expression de gènes impliqués dans différentes voies biologiques.

4.2 Evolution des gènes dupliqués

On a vu précédemment qu'avec l'augmentation du nombre de génomes de plante séquencés, pour beaucoup d'Angiospermes, des traces actuelles de duplication de gènes témoignent d'événements de WGD. Mais les analyses de ces génomes ont aussi révélées d'autres types de duplication, Wang *et al.* (2012) en ont répertorié six qu'ils ont classés selon la distance physique entre les copies des gènes dupliqués et le mécanisme à l'origine de la duplication (Tableau I).

Tableau I : Différents types de duplication de gènes selon Wang *et al.* (Wang et al. 2012)

Mode of gene duplication	Feature	Mechanism	Change of promoter
Whole genome or segmental	Duplication of all genes in a genome or chromosomal region.	Abnormal cell division	No
Tandem	Duplication of a single gene. The newly created gene is adjacent to its parental gene.	Unequal chromosomal crossing over	No
Proximal	Duplication of a single gene. The newly created gene is near, but not adjacent to, its parental gene.	Tandem duplicates interrupted by other genes, or induced by localized transposon activities	Yes
DNA-based transposed	Duplication of a single gene. The newly created gene is translocated to a different and nonhomologous chromosomal region.	Induced by DNA transposons	Yes
Retrotransposed	Duplication of a single gene. The newly created gene is translocated to a different and nonhomologous chromosomal region.	Induced by retrotransposons	Yes
Dispersed	Duplication of a single gene. The newly created gene is translocated to a different and nonhomologous chromosomal region.	Unclear	Unclear

La conséquence la plus probable de la duplication d'un gène est à long terme le retour du gène à l'état de simple copie. Ce processus est appelé « fractionation » ou « genome downsizing » (Renny-Byfield et Wendel 2014) lorsque l'état de duplication provient d'un événement de WGD. De nombreux modèles d'évolution des gènes dupliqués sont proposés (Hughes 1994, Force et al. 1999, Lynch et Force 2000), rassemblés et analysés dans plusieurs articles de synthèse (Innan et Kondrashov 2010, Conant et Wolfe 2008). Ces travaux permettent de décrire les tendances générales de l'évolution des gènes dupliqués (Figure 16).

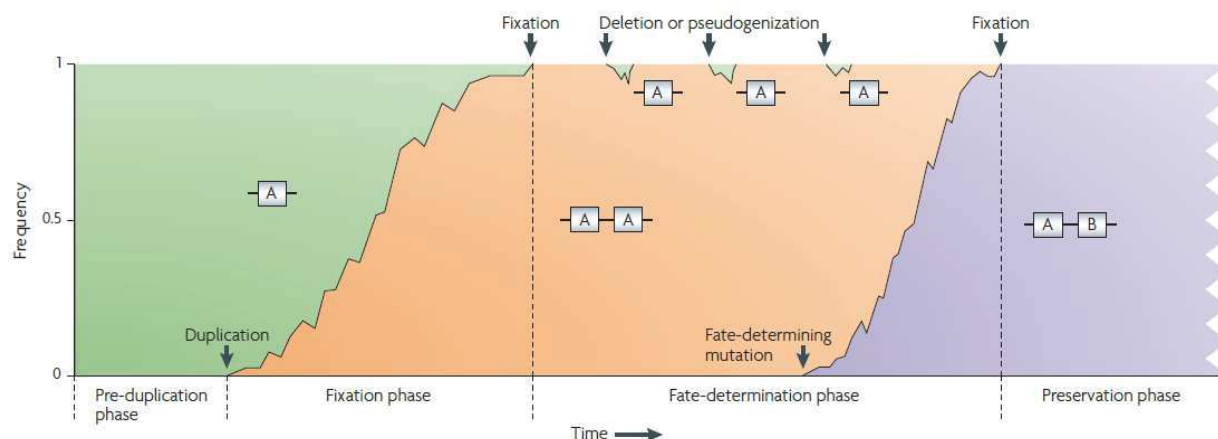


Figure 16 : Phases d'évolution conduisant à la préservation d'un gène à l'état dupliqué. Dès que la duplication d'un gène se produit, la phase de fixation débute. Le gène dupliqué peut être perdu ou fixé. Quand la fixation de la duplication est atteinte, la phase du destin de la duplication débute et se poursuit jusqu'à ce qu'une mutation détermine le destin d'une copie du gène. (d'après Innan et Kondrashov 2010).

En termes d'évolution, la redondance des gènes dupliqués n'est pas stable. Les gènes dupliqués vont être au travers de substitutions neutres soit fixés ou détruits (perdus ou transformés en pseudogènes). Après cette phase de fixation, les gènes dupliqués seront maintenus de façon stable par sélection naturelle (Innan et Kondrashov 2010, Wang et al. 2012). L'une des deux copies peut évoluer en faisant émerger une nouvelle fonction, il s'agit du modèle de la néofonctionalisation. Dans le cas de la sous-fonctionalisation, la fonction initiale du gène est assurée par les deux copies du gène (Hahn 2009).

Il a été constaté pour de nombreuses plantes polyploïdes et paléopolyploïdes : *Arabidopsis* (Thomas et al. 2006), maïs (Woodhouse et al. 2010), blé (Feldman et Levy 2012), que le fractionnement se faisait de façon biaisée (élimination préférentielle des gènes dupliqués provenant d'un sous-génome) et était accompagné d'un déséquilibre d'expression des sous-génomes (favorisant le sous-génome qui perdait le moins de gènes). Garsmeur *et al.* (2014) ont révélé que pour d'autres paléopolyploïdes (Poplar, Soybean, Banana) le fractionnement ne suivait pas les mêmes règles. Il ne présentait de biais ni pour la délétion des copies de gènes ni pour l'expression des sous-génomes (Garsmeur et al. 2014). Les auteurs attribuent ces différences d'évolution à l'origine allo ou autopolyploïdes de ces organismes.

La nature aléatoire du fractionnement a été aussi analysée. De Smet et al. (2013) ont examiné les caractéristiques évolutives et fonctionnelles de gènes uniques pour une vingtaine de polyploïdes et paléopolyploïdes. Ils ont observé que le fractionnement ne se produit pas au hasard : pour certains gènes impliqués dans le métabolisme de base cellulaire ou ayant des fonctions cellulaires essentielles (réparation et réplication de l'ADN), une pression évolutive convergente a provoqué l'élimination des duplications (pour les organismes étudiés) (De Smet et al. 2013). Ces gènes résistants à la duplication pourrait être des gènes sensibles à l'effet dose (impliqués dans un réseau d'interactions moléculaires de la cellule) ou pour lesquels la moindre mutation sur une copie du gène pourrait provoquer un effet dominant négatif ou délétère (Feldman et Levy 2012, Wang et al. 2012, De Smet et al. 2013) (Figure 17).

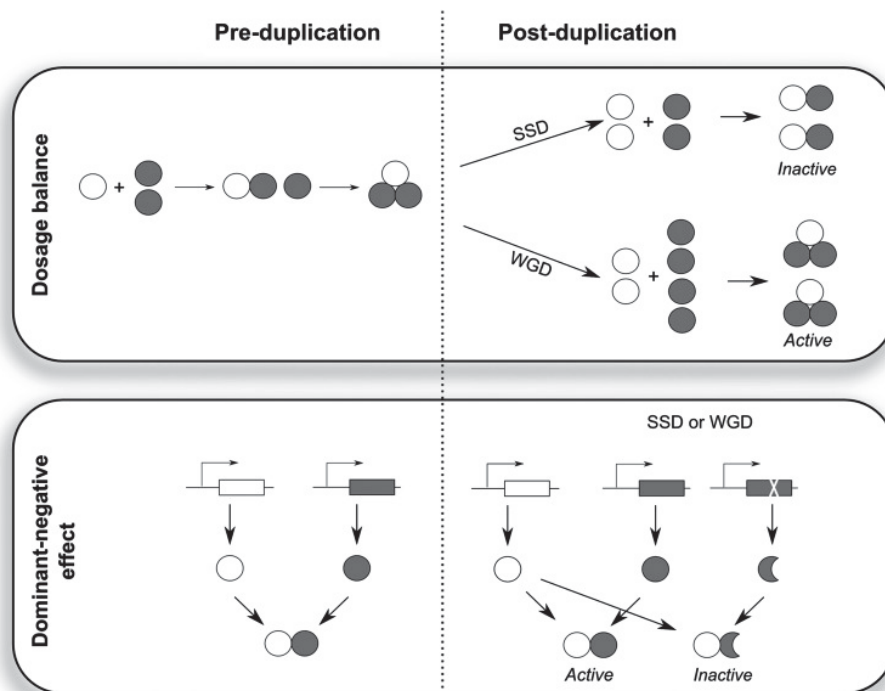


Figure 17 : Hypothèses expliquant la résistance de certains gènes à la duplication, pour des WGD ou des duplications de petite échelle (small scale duplication SSD). En cas de gènes nécessitant un équilibre stœchiométrique (en haut), la WGD n'aura pas forcément un effet négatif sur le produit du gène et ses sous-unités. Pour des duplications limitées, certains produits de gènes peuvent être défectueux. Pour la seconde hypothèse, mutation à effet dominant négatif, le gène dupliqué muté produit une sous-unité défectueuse dans les deux cas de duplication (d'après De Smet et al. 2013).

Chez le blé, le fractionnement entraîne des conséquences cytologiques. Des séquences spécifiques des chromosomes impliquées dans la reconnaissance et l'initiation de l'appariement d'homologues à la méiose, auraient été rapidement éliminées des paires d'homéologues. Ce processus aurait augmenté la divergence génétique entre homéologues et compléterait le contrôle génétique (Ph-genes) de l'appariement des chromosomes. Ainsi l'appariement restreint aux chromosomes homologues garantirait une stabilité génétique et une ségrégation régulière des chromosomes (Feldman et Levy 2012).

Même si la plupart des gènes dupliqués seront éliminés ou bien transformés en pseudogènes (Conant et al. 2014, Jiao et Paterson 2014), la rétention des gènes à l'état dupliqué fait aussi partie de leur évolution. Le processus de sous-fonctionalisation et la sélection pour maintenir un équilibre stœchiométrique entre les copies de gènes pourrait préserver les gènes dupliqués du fractionnement (Flagel et Wendel 2009, Conant et al. 2014). De même, des gènes appartenant à certaines catégories fonctionnelles sont retenus après duplication (Jiao et Paterson 2014, Soltis et al. 2012). Lors de l'étude des polyploïdes, les mécanismes de préservation des gènes dupliqués répertoriés sont nombreux : néofonctionalisation, sous-fonctionalisation, gènes de fonction particulière, gènes sensibles à l'effet dose, et sont difficiles à distinguer les uns des autres (Flagel et Wendel 2009), mais ils interagissent et font que la rétention et la diploïdisation suivront des schémas très variables selon les polyploïdes (Conant et al. 2014). Ainsi la duplication des gènes est considérée comme le mécanisme principal de création de nouveaux gènes mais il n'est pas illimité. Aussi l'évolution d'une copie de gènes dupliqués vers une nouvelle fonction avec un avantage sélectif doit se produire rapidement avant que la copie de gène ne soit perdue (Jiao et Paterson 2014).

4.3 Evolution des polyploïdes

Feldman et Levy (2012) ont décrit l'évolution du blé et des espèces proches, et en ont déduit un modèle d'évolution des polyploïdes en distinguant deux phases. Une première phase, la phase des « changements révolutionnaires » qui se produisent rapidement après la polyploïdisation en quelques générations et une seconde phase, la phase des « changements évolutifs » qui se produisent à l'échelle de l'évolution après des centaines ou des milliers de générations (Feldman et Levy 2012). Ce sont ces changements évolutifs qui ont associé la polyploïdie à l'histoire évolutive des Angiospermes. Ils résultent de l'évolution des gènes dupliqués selon les processus évoqués précédemment.

Soltis *et al.* (2014) définissent la polyploïdisation comme un événement de spéciation instantané, qui augmente la diversité et fournit de nouvelles combinaisons génétiques évolutives (Soltis, Visger, et al. 2014). Quelques exemples d'études menées pour comprendre la polyploïdie illustrent ces propos. Tout d'abord la polyploïdie répond aux conditions qui décrivent la spéciation : changements au niveau de l'ADN et du génome qui sont héréditaires et qui provoquent un changement brutal d'état (Heslop-Harrison 2012). Wood et al. estimaient en 2009 que 15% des événements de spéciation chez les plantes à fleur et 31% chez les fougères étaient provoqués par la polyploïdisation (Wood et al. 2009). Ensuite la polyploïdie favorise la diversification. La richesse en espèces des Angiospermes est liée à des événements de WGD (Schranz et al. 2012, Soltis et al. 2014) qui après un délai entre les événements de polyploïdisation et les processus de radiation serait à l'origine d'importantes diversifications. Enfin l'étude de Vanneste *et al.* (2014) illustre l'importance de la diversité génétique qui a contribué à la survie de certaines lignées au cours de la période du Crétacé (Vanneste, Maere, et al. 2014, Vanneste, Baele, et al. 2014).

Plusieurs facteurs vont participer à l'augmentation de la diversité des polyploïdes et permettre de nouvelles perspectives évolutives (Crawford et al. 2014). Tout d'abord la variabilité génétique des polyploïdes provenant d'origine multiple. La polyploïdisation ne correspond pas nécessairement à un goulet d'étranglement évolutif qui serait du à l'isolement reproductif des espèces polyploïdes (Jiao et Paterson 2014). En effet, les lignées polyploïdes appartiennent à des réseaux généalogiques qui permettent des flux de gènes et des recombinaisons (Soltis, Visger, et al. 2014). Mais l'origine multiple des polyploïdes n'est pas une règle, même si elle est fréquente, des exceptions existent, par exemple chez *Spartina anglica* (Ainouche dans Soltis et Soltis 2012), chez *Coffea arabica* (Lashermes et al. 2014). La nature dynamique des génomes des polyploïdes est un autre facteur d'augmentation de la diversité. Feldman et Levy (2012) ont répertorié et classé les remaniements génomiques et épigénétiques et ainsi que leurs conséquences évolutives (Tableau II), (Feldman et Levy 2012) qui vont être à l'origine de nombreux avantages pour les polyploïdes. En effet ces changements, rapides ou acquis au cours de l'évolution, vont avoir des conséquences sur les processus physiologiques et développementaux des polyploïdes. Ils augmentent l'adaptabilité et la vigueur des polyploïdes qui vont profiter de l'apparition de nouveautés phénotypiques (hétérosis inter-génomiques, colonisation de nouvelles niches écologiques, nouvelles interactions plantes microorganismes...) (Feldman et Levy 2012, Jiao et Paterson 2014, Soltis et al. 2014).

Tableau II : Liste des remaniements génomiques et épigénétiques et leurs conséquences évolutives chez les allopolyploïdes (Feldman et Levy 2012)

Structural	Functional
1. Revolutionary changes occur during or soon after allopolyploidization, lead to diploidization and are often reproducible Genetic • Elimination of low-copy DNA sequences • Elimination, reduction or amplification of high-copy sequences • Intergenomic invasion of DNA sequences • Elimination of rRNA and 5S RNA genes Epigenetic • Chromatin remodeling • Chromatin modification • Heterochromatinization • DNA methylation • Small RNA activation or repression 2. Evolutionary changes facilitated by allopolyploidy in wheat occur during the evolution of the species by promoting species biodiversity (current knowledge is limited to the genetic rather than epigenetic changes) • Chromosomal re-patterning (intra- and intergenomic translocations) • Introgression of chromosomal segments from alien genomes and production of recombinant genomes • Intergenomic transposons invasion	Genetic • Gene loss/loss of function • Rewiring of gene expression through novel intergenomic interactions • Novel dosage responses (positive, negative, dosage compensation) • Gene suppression or activation • Transcriptional activation of transposons that may affect nearby genes • New transposon insertion/excision Epigenetic • Gene silencing or activation through changes in methylation or small RNAs or through chromatin modifications • Transposon silencing or activation through demethylation or changes in small RNAs or through chromatin modifications • Subfunctionalizations • Neofunctionalizations • New dosage effects through copy number variation • New allelic variations

Mais tout comme l'origine multiple des polyploïdes, les remaniements chromosomiques ne sont pas une règle, certains polyploïdes en présentent peu : *Spartina Anglica* (Ainouche dans Soltis et Soltis 2012), *Gossypium hirsutum* (Salmon et al. 2010). Les réponses variables des polyploïdes aux

événements d'hybridation et de duplication des génomes, informent de la diversité des histoires évolutives et de la complexité à étudier la polyploïdisation.

Malgré ces études qui décrivent les perspectives évolutives favorables de la polyploïdisation, le rôle de la polyploïdie sur la spéciation et la diversification est toujours controversé. Certains auteurs développent des théories différentes et signalent que les polyploïdes ont un potentiel évolutif réduit. Ils rapportent des taux d'extinction plus importants pour des populations de néopolyploïdes que pour des populations de diploïdes (Mayrose et al. 2011). Ils attribuent ces taux d'extinction à l'inefficacité de la sélection des gènes dans un contexte de gènes multiples. De plus les polyploïdes se développant dans des populations entourées de diploïdes subiraient des pressions compétitives pour s'établir (Arrigo et Barker 2012). Pour Arrigo *et al.* (2012) les polyploïdes ne bénéficieraient pas vraiment d'avantages évolutifs. Les polyploïdes observés actuellement seraient les représentants des polyploïdes qui n'auraient pas disparu et seraient dus au fort taux de formation de WGD (Arrigo et Barker 2012). Dans une étude très récente, Estep et al. (2014) rapportent qu'ils n'ont pas observé une augmentation nette des taux de diversification après la polyploïdisation pour des espèces du groupe des Andropogonées, néanmoins comme mode de spéciation, la fréquence d'allopolyplôïdisation est élevée (Estep et al. 2014).

Bien que le débat au sujet de l'impasse évolutive de la polyploïdisation se poursuive, et que l'échantillonnage et les méthodes utilisées dans les études qui visent à examiner cette question soient longuement discutés (Soltis, Segovia-Salcedo, et al. 2014, Mayrose et al. 2014), il semblerait que l'on puisse considérer que les événements de polyploïdisation soient fréquents, qu'occasionnellement quelques polyploïdes aient résisté durablement au cours de l'évolution. Ces polyploïdes résultant de mutations adaptatives occasionnelles seraient à l'origine de la réussite de certaines lignées (par exemple pour les Angiospermes), car d'une part ils seraient apparus à des moments favorables à la diversification des polyploïdes et à leur installation (Fawcett et al. 2009) et d'autre part ils disposeraient d'un réseau de régulation transcriptionnel complexe (Weiss-Schneeweiss et al. 2013).

4.4 Polyplôïdie et adaptation des plantes

Dans le contexte du succès évolutif de la polyploïdie, la question d'une meilleure adaptation des polyploïdes à des environnements stressants a été très rapidement considérée. Comme des polyploïdes sont plus fréquemment observés dans des environnements extrêmes (arctiques et subarctiques, de hautes altitudes et aussi arides) (Madlung 2012), des modèles pour généraliser la distribution et la fréquence de la polyploïdie selon des variations éco-géographiques ont été proposés. Mais les études récentes basées sur plus d'individus représentant les genres des Angiospermes et avec plus de données ne confirment pas une relation entre la polyploïdie et la position dans un environnement climatique particulier, à l'exception des environnements d'altitude élevée (Soltis et al.

2010, Madlung 2012, Weiss-Schneeweiss et al. 2013). Mais comme cette relation peut dépendre de nombreux facteurs, l'effet adaptatif lié à la polyploïdie pourrait se révéler en considérant les espèces indépendamment les unes des autres (Moghe et Shiu 2014). Ainsi pour *Chamerion angustifolium*, la correspondance entre les distributions géographiques, les niches climatiques et la tolérance physiologique différente d'individus diploïdes et tétraploïdes à des paramètres climatiques a été établie (Thompson et al. 2014). Parmi les généralisations portant sur les polyploïdes, l'hypothèse de l'autofécondation plus fréquente pour les polyploïdes (par rapport aux diploïdes) a été confirmée.

Même en l'absence d'un lien écogéographie-polyploïdie, il est tout de même admis que les polyploïdes sont géographiquement et écologiquement différenciés de leurs parents diploïdes et peuvent se développer sur un plus large éventail d'environnements (Soltis et al. 2010, Weiss-Schneeweiss et al. 2013). La polyploïdisation augmenterait la tolérance à des conditions écologiques très diverses, Pandit *et al.* (2011) ont montré que les espèces en voie d'extinction étaient plutôt diploïdes alors que les espèces invasives avaient tendance à être polyploïdes (Pandit et al. 2011). Pour Ramsey (2011), la valeur adaptative provoquée par la polyploïdisation est incertaine tant qu'elle n'est pas prouvée par une étude qui permet de dissocier les effets de la polyploïdisation de ceux des variations génétiques accumulés après la duplication du génome. La comparaison du comportement de polyploïdes et de néo-polyploïdes fait la part des changements physiologiques provoqués par chacun des phénomènes (Ramsey 2011). Ainsi pour *Achillea borealis*, il est avéré que la duplication du génome confère aux populations d'individus hexaploïdes l'aptitude à se développer dans un environnement aride (Ramsey 2011). De même pour le blé, des changements moléculaires et physiologiques immédiats induit par l'allopolyploïdisation provoquent une modification de la régulation des gènes sous stress salin (Yang et al. 2014). En revanche pour *Chamerion angustifolium*, Maherali *et al.* (2009) ont montré que la tolérance à la sécheresse provenait de l'évolution des polyploïdes et était indépendante de la duplication des génomes (Maherali *et al.* 2009). Pour *Spartina anglica*, ce sont les conséquences de l'hétérosis du à la polyploïdisation qui provoquent le caractère invasif des plantes (Ainouche et al. 2009).

Parfois les changements physiologiques n'induisent pas une adaptation à un milieu mais plutôt une caractéristique physique qui va changer l'interaction des polyploïdes dans leur environnement. C'est le cas pour *Chamerion angustifolium*, dont le pollen est plus attractif pour les pollinisateurs que celui des espèces diploïdes (Baldwin et Husband 2011).

La diversité des résultats obtenus au travers des quelques études citées suggère la complexité du lien entre la polyploïdie et l'adaptation. Pour chaque espèce, ce lien peut dépendre simultanément ou indépendamment de la polyploïdisation, des variations génétiques qui suivent cet événement mais aussi de l'environnement des plantes (Moghe et Shiu 2014).

5 Le genre *Coffea*, modèle biologique d'étude

5.1 Phylogénie et origine des caféiers

Les caféiers appartiennent à l'une des plus grandes familles tropicales des Angiospermes, la famille des *Rubiaceae*. La tribu des *Coffeae* fait partie de la sous-famille des *Ixoroideae* (Figure 18) et était jusqu'en 2011 divisée en deux genres, *Coffea* et *Psilanthus*. En effet, à partir de caractéristiques florales et de croissance différentes, 104 espèces étaient regroupées dans le genre *Coffea* et 20 dans le genre *Psilanthus*. Mais plusieurs études phylogénétiques moléculaires démontraient un niveau faible de variation de séquences entre espèces de ces deux genres et dénonçaient cette classification (Lashermes et al. 1997, Cros et al. 1998, Maurin et al. 2007). A la suite d'une nouvelle analyse phylogénétique sur 4 régions d'ADN plastidial et une région ITS de l'ADN nucléaire ribosomal, Davis et al. (2011) ont confirmé les études précédentes, et ont intégré des espèces du genre *Psilanthus* dans le genre *Coffea*. Le genre *Coffea* est donc maintenant composé de 124 espèces qui se répartissent sur les zones intertropicales d'Afrique, et sont présentes à Madagascar, dans l'archipel des Comores et les îles Mascareignes mais aussi en Asie du Sud-Est et Océanie (Figure 19) (Davis et al. 2011).

Outre la classification des espèces, les analyses moléculaires (RAPD, séquences d'ADN plastidial et séquences ITS d'ADNr nucléaire) ont permis de préciser la phylogénie et les modalités de diversification du genre *Coffea* (Lashermes et al. 1993, Lashermes et al. 1996, Lashermes et al. 1997, Cros et al. 1998, Maurin et al. 2007, Anthony et al. 2010). Ces analyses suggèrent que les espèces du genre *Coffea* résultent d'une spéciation radiale relativement récente, qui pourrait avoir débuté il y a moins de 0.5 million d'années en Afrique (Anthony et al. 2010). Sachant que les datations moléculaires et notamment l'étape de calibration sont des analyses délicates, cette estimation est à considérer avec précaution. A partir d'un centre d'origine située en Afrique centrale atlantique (Basse Guinée) (Anthony et al. 2010), les caféiers auraient colonisé les forêts d'une part vers l'Est et d'autre part de toute l'Afrique équatoriale puis de Madagascar (Figure 20). Cette dispersion aurait connu des phases d'expansion et de reflux des populations en fonction des variations climatiques et se serait accompagnée de spéciations associées à des adaptations écologiques. En effet les données phylogénétiques révèlent l'existence de plusieurs groupes d'espèces relativement peu différenciés correspondant globalement à des ensembles biogéographiques, et dont la distribution actuelle serait fortement liée aux fluctuations climatiques du quaternaire. Après l'intégration au genre *Coffea* des espèces appartenant anciennement au genre *Psilanthus*, une seconde voie de dispersion et diversification à partir de l'Afrique (Somalie et Nord-Est du Kenya) vers l'Asie est suggérée pour compléter l'histoire de l'évolution de ce genre (Davis et al. 2011).

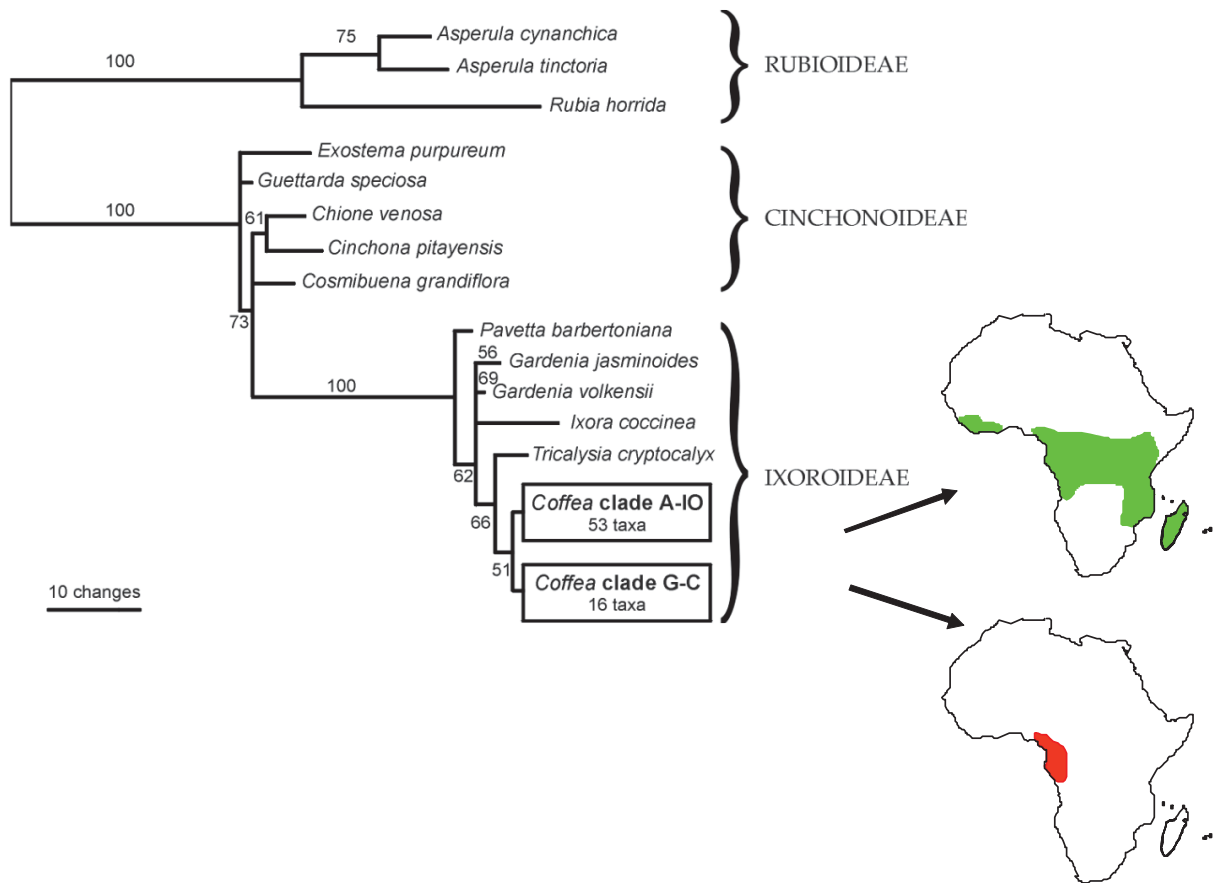


Figure 18 : Phylogénie moléculaire des caféiers sur la base de l'analyse de l'ADN chloroplastique
(Anthony et al. 2010)



Figure 19 : Distribution des espèces du genre *Coffea*. Le tracé pointillé indique les nouvelles aires de distribution de *Coffea*, suite à l'intégration des espèces *Psilanthus* de l'Asie et l'Océanie (Davis et al. 2011).



Figure 20 : Dispersion du genre *Coffea* à partir d'un centre d'origine située en basse Guinée (Anthony et al. 2010).

Grâce aux avancées technologiques récentes permettant de disposer très rapidement d'importantes données génomiques pour de nombreuses plantes, des études paléogénomiques basées sur la comparaison de génomes ou de portions de génomes sont dorénavant possibles. Dans le contexte de ces études, les astérides (dont *Coffea*) et les rosides partageraient le même ancêtre hexaploïde (Jiao et al. 2011, Vanneste, Baele, et al. 2014). En développant la stratégie comparative d'une région de *Coffea* avec les régions correspondantes de plusieurs rosides (*V. vinifera*, *Populus trichocarpa* et *Arabidopsis thaliana*), Cenci et al. (2010) ont confirmé ce fait et ont démontré qu'aucun événement de polyploïdisation supplémentaire ne se serait produit au sein de la lignée des *Coffea* (Cenci et al. 2010). Les études menées à l'échelle du génome suite au séquençage du génome de *C. canephora* (Denoeud et al. 2014) confirment aussi ce résultat et indiquent que *C. canephora* présente une forte conservation de la structure du génome hexaploïde commun aux Eudicotylédones (Figure 21). Cette dernière étude établit aussi le génome de *Coffea* comme l'espèce de référence pour la compréhension de l'évolution de la structure des génomes des astérides (Denoeud et al. 2014).

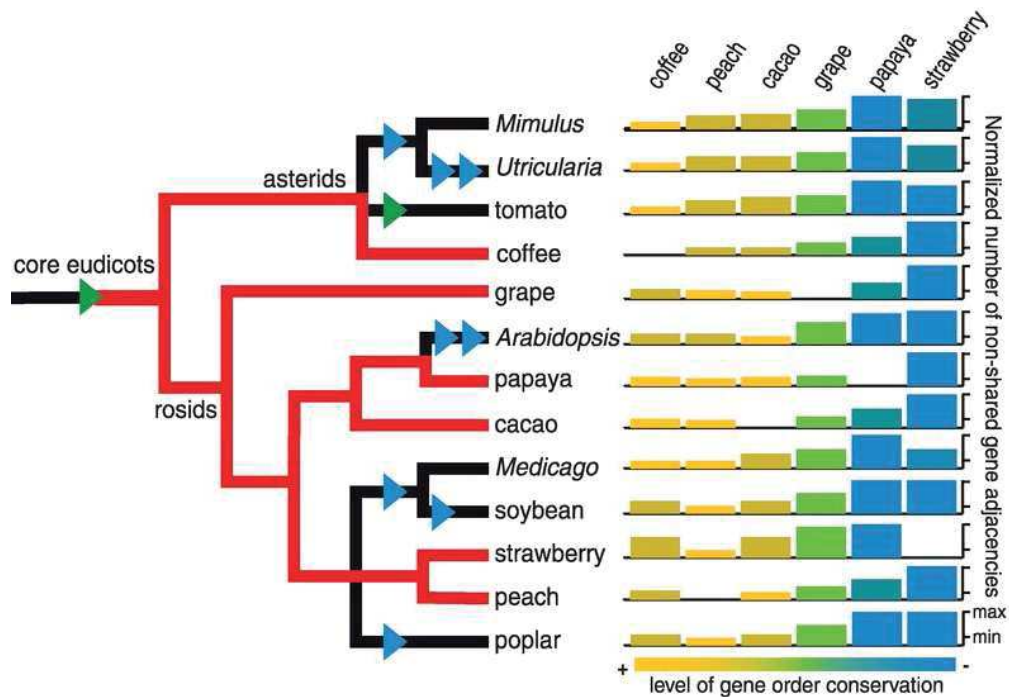


Figure 21 : Phylogénie et duplication pour les Eudicotylédones. Les flèches bleues et vertes indiquent respectivement les événements de tétra- et hexaploïdisation, les lignes rouges indiquent les lignées de 6 espèces qui n'ont pas subi d'autres événements de polyploïdisation. Les histogrammes colorés présentent pour ces 6 espèces, le niveau de conservation de l'ordre des gènes (Denoëud et al. 2014).

5.2 *Coffea arabica*

5.2.1 Origine et caractéristiques

Les espèces du genre *Coffea* sont toutes diploïdes ($2n = 2x = 22$) à l'exception de l'espèce *C. arabica* qui est tétraploïde ($2n = 4x = 44$). L'analyse des caryotypes du génome de caféier, rapportée par Sybenga (1959), met en évidence l'existence de 11 paires de chromosomes ($2n=22$) présentant une taille moyenne réduite (1 à 3 μm) et peu variable. En revanche la quantité d'ADN nucléaire est très variable entre les différentes espèces de caféier, avec des valeurs comprises entre 0,95 et 1,78 pg par noyau. Une étude suggère que ces différences résulteraient de l'amplification spécifique de certaines familles d'éléments transposables (Hamon et al. 2011). *C. arabica* avec le double de chromosomes présente 2,61 pg d'ADN pour un génome de 1300 Mb. (Cros et al. 1995, Anthony et Lashermes 2005 dans Sharma et Sharma 2005).

Une origine allopolyploïde a été suggérée pour l'espèce *C. arabica* (Carvalho 1950; Grassias et Kammacher 1975). Par des approches moléculaires et notamment l'utilisation de marqueurs chloroplastiques et ribosomiques (Berthou et al. 1983, Lashermes et al. 1997, Cros et al. 1998), il a été démontré une faible divergence génétique entre *C. arabica* et des espèces du groupe des diploïdes originaires d'Afrique de l'ouest (*C. canephora*, *C. congensis*, *C. brevipes*) et centrale (*C. eugenioides*,

et *Coffea* sp. « Mouloundou »). Une étude plus approfondie effectuée au moyen d'analyses RFLP (en utilisant des sondes spécifiques de locus) et cytogénétiques (par hybridation de l'ADN génomique des espèces potentielles avec les chromosomes de *C. arabica*) a démontré la présence de deux lots de chromosomes E^a et C^a, qui proviendraient respectivement de formes ancestrales des espèces actuelles *C. eugenoides* (parent femelle, Génome E^a) et *C. canephora* (parent mâle, génome C^a) (Lashermes et al. 1999) (Figure 22). Ainsi plusieurs déroulements de la polyploïdisation peuvent être envisagés. L'hybride ancestral tétraploïde pourrait résulter d'un croisement interspécifique faisant intervenir des diplo-gamètes (2n) dont la formation aurait été favorisée par des conditions climatiques particulières. Il pourrait également être issu de la formation d'un hybride interspécifique diploïde suivie d'un doublement chromosomique provenant de gamètes non réduits au cours de la méiose peut être mal régulée des hybrides interspécifiques. Et enfin, il pourrait aussi venir de la formation d'un hybride interspécifique triploïde suivie d'un autre croisement avec l'une des espèces parentales, l'hybridation entre gamètes aneuploïdes de l'hybride triploïde pourrait conduire à un changement de polyploïdie. Les connaissances actuelles génomiques de *C. arabica* ne permettent pas de favoriser l'une des voies de polyploïdisation.

Lashermes et al. (Lashermes et al. 1999) soulignent qu'une faible divergence est observée entre les deux sous-génomes constitutifs de *C. arabica* mais aussi entre *C. arabica* et ses parents, ce qui suggère une spéciation récente. A partir de l'étude de la fréquence de substitution nucléotidique d'une région génomique de *C. arabica* montrant un échange non réciproque homéologue, Cenci et al. (2012) ont estimé que la spéciation de *C. arabica* aurait eu lieu relativement récemment (10 000 à 50 000 ans) (Figure 22). Cette estimation est supposée à partir de la divergence génétique observée de 1 ‰ base nucléotidique entre les sous-génomes de *C. arabica* pour cette région, c'est-à-dire environ 10 fois moins importante que celle qui différencie les deux espèces parentales diploïdes divergeant depuis 500 000 ans (1,30 ‰ bases nucléotidiques $\pm$ 0,67) (Cenci et al. 2012). Dernièrement une étude génomique de plusieurs *C. arabica* sauvages et *C. arabica* var. Caturra a permis de préciser que l'espèce *C. arabica* proviendrait d'un seul événement d'allopolyploïdisation. En effet, des réarrangements génomiques (dont celui utilisé pour l'estimation de la période de spéciation de *C. arabica*) mis en évidence dans cette étude sont partagés par toutes les accessions, d'autres à l'inverse spécifiques d'accessions, contribueraient à l'origine de la diversité de l'espèce *C. arabica* (Lashermes et al. 2014).

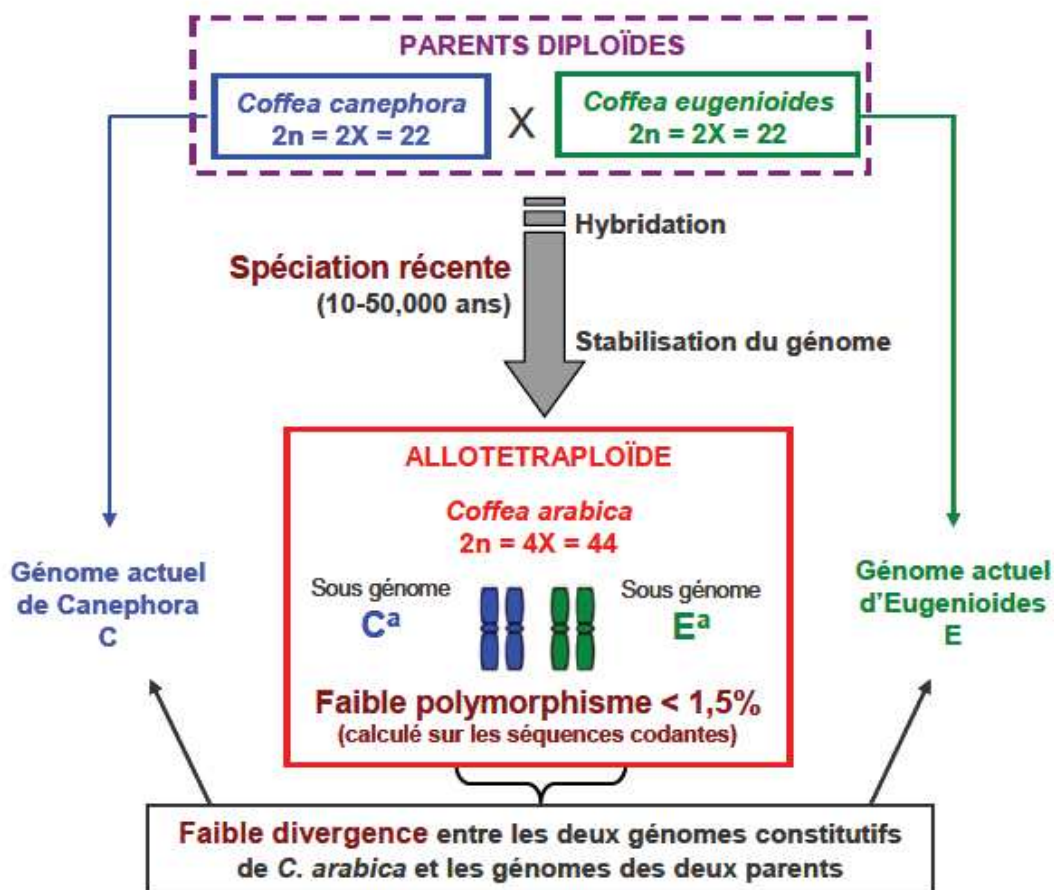


Figure 22 : Origine de la seule espèce allotétraploïde *C. arabica* (d'après (Lashermes et al. 1999)).

Les variations climatiques importantes pendant le quaternaire et leurs implications sur la distribution des différentes espèces de caféiers rendent difficiles l'interprétation de l'origine géographique de *C. arabica*. Toutefois, Lashermes et al. (Lashermes et al. 1999) ont suggéré le versant nord-ouest du rift africain comme zone d'origine possible de *C. arabica*, zone géographique où les espèces parentales *C. canephora* et *C. eugenioides* se chevauchent (Thomas 1944). Par la suite, l'espèce *C. arabica* aurait colonisé la région comprenant le sud du Soudan, le Nord du Kenya et le sud-ouest de l'Éthiopie qui est considéré comme son centre de diversité.

Pour l'espèce *C. arabica*, la diversité génétique est étroite et d'autant plus pour les variétés principales de caféiers cultivés. Une part importante des variétés actuelles de caféiers cultivés dérivent de deux populations, *C. arabica* var. Typica et *C. arabica* var. Bourbon (Krug et Mendes 1940). Peu de plantes sont à l'origine de ces deux populations qui ont été diffusées à travers le monde dans les pays actuellement producteurs (Figure 23).

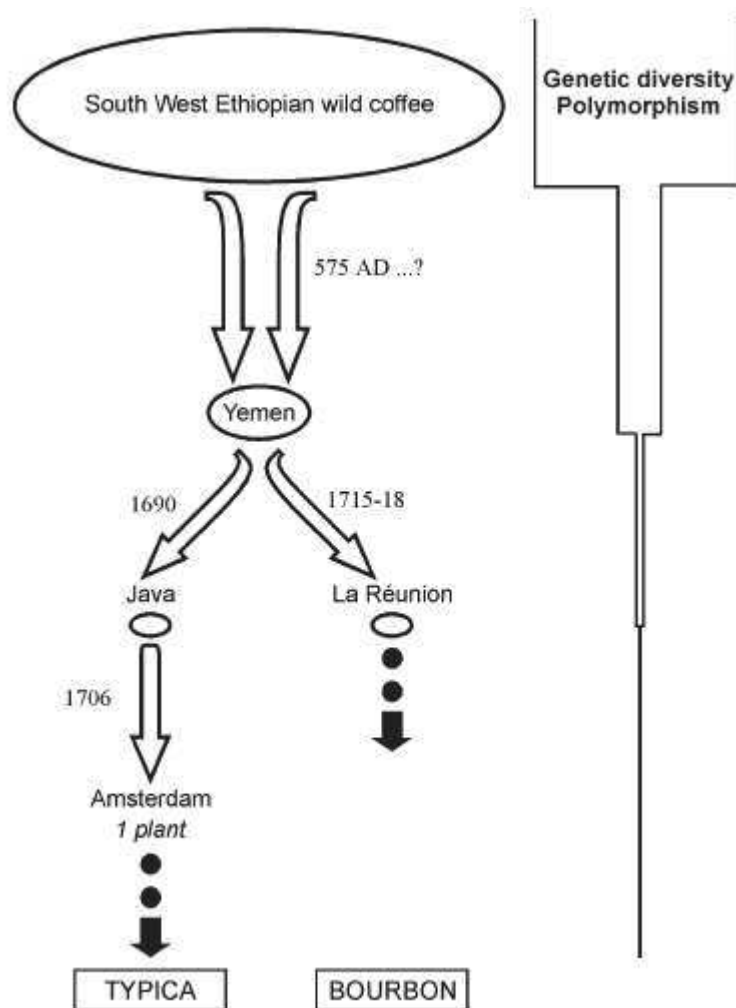


Figure 23 : Origine des deux variétés « mères » des principales variétés cultivées actuelles de *C. arabica* (Anthony et al. 2002).

5.2.2 Système de reproduction et caractéristiques génomiques

La plupart des espèces diploïdes de *Coffea* sont allogames et auto-incompatibles, à l'inverse *C. arabica* est autogame et auto-compatible. Pour cette espèce, il a été observé un mode d'hérédité de type disomique par l'analyse des ségrégations au sein d'une population F2 d'un croisement interspécifique de *C. arabica* et *C. canephora* (Lashermes 2000). *C. arabica* possède un comportement méiotique comparable à celui des espèces diploïdes. A la méiose, les chromosomes homologues s'appariaient de façon systématique pour former des bivalents. Probablement en raison de facteurs génétiques, les chromosomes homéologues de *C. arabica* ne s'appariaient pas même s'ils sont peu divergents entre les deux sous-génomes (Lashermes et al. 1999). Pourtant des anomalies peuvent se produire au cours de la méiose, des réarrangements génomiques supposant des échanges entre les sous-génomes homéologues ont été mis en évidence dans la dernière analyse génomique de

C. arabica citée précédemment (Lashermes et al. 2014). Dans le contexte de sa faible diversité génétique, ces échanges représenteraient une source majeure de diversification génétique.

5.3 L'environnement naturel des caféiers et leur adaptation

Toutes espèces de caféiers sont des plantes pérennes ligneuses, variant de la taille d'un arbuste à celle d'un arbre. Elles présentent une diversité considérable en termes de morphologie, de phénologie et d'adaptation écologique. Elles sont des marqueurs précieux de l'évolution de la forêt tropicale africaine car elles ont colonisées différents types de forêts (Anthony et al. 2010), allant de la forêt sempervirente humide aux forêts claires de savane plus sèches, en passant par des forêts saisonnièrement sèches à feuilles caduques, des forêts mixtes sempervirentes-caduques saisonnièrement sèches et des forêts riveraines parfois inondées (Davis et al. 2006). A l'exception de trois espèces *C. canephora*, *C. liberica* et *C. eugenoides*, toutes les espèces ont des aires de distribution plutôt restreintes, parfois limitées à quelques kilomètres carrés, par exemple pour *C. congensis* située dans des zones saisonnièrement inondées du bassin du Zaïre ou encore *C. racemosa* située dans des zones sèches de la région côtière du Mozambique (Davis et al. 2006).

Les distributions géographiques de l'espèce tétraploïde *C. arabica* et de ses espèces parentales (*C. canephora* et *C. eugenoides*) sont très différentes. A l'état sauvage, l'aire de répartition de *C. canephora* s'étend de l'Afrique Centrale, au golf de la Guinée et à l'Ouganda. Comparés à la vaste aire de répartition occupée par *C. canephora*, les espaces occupés par *C. arabica* et *C. eugenoides* sont plus limités. *C. eugenoides* est repérée de l'Est au Sud-est de l'Afrique (Ouganda, Tanzanie, Kenya, Rwanda et Burundi). *C. arabica*, quant à elle, est issue 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 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 20-24°C (DaMatta et al. 2007). L'espèce sauvage *C. canephora* est issue de forêts équatoriales humides de basses à moyennes altitudes, entre 250-1500m (Thomas 1944, Davis et al. 2006), avec un optimum de températures moyennes annuelles pour l'espèce de 22-26°C (DaMatta et Ramalho 2006, DaMatta et al. 2007). Enfin, l'espèce sauvage *C. eugenoides* 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 (Thomas 1944, Davis et al. 2006).

Dans le contexte de diversité phénotypique et phénologique associée à la faible divergence génétique (due à l'évolution radiale, rapide et récente) des caféiers, on peut imaginer que les modalités de régulation de l'expression des gènes jouent un rôle central dans la diversification et les capacités adaptatives des caféiers.

De plus, même en condition de culture, les espèces de caféiers témoignent d'une certaine capacité adaptative. Seules, les deux espèces, *C. arabica* et *C. canephora*, sont cultivées, mais elles sont adaptées aux conditions environnementales de toute la ceinture intertropicale où parfois les intervalles de température de leurs zones de culture sont plus importants comparés à ceux des zones de diversification primaire (DaMatta et al. 2007). Pour ces deux espèces, l'origine de cette capacité pourrait avoir des causes différentes. Pour *C. canephora*, la culture 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 (Montagnon et al. 2012). En revanche pour *C. arabica* comme nous l'avons évoqué précédemment, la culture est basée sur une population à très faible base génétique. Pourtant *C. arabica* s'adapte à des conditions de culture très variables. Comparée à ses espèces parentales, elle montre une aptitude à tolérer des variations de température plus importante, peut être en raison de son caractère polyploïde.

C'est dans ce contexte d'avantage adaptatif potentiellement lié à la polyploïdie, que les profils d'expression des gènes de *C. arabica* et de chaque espèce parentale ont été comparés en utilisant les puces à ADN (Privat et al. 2011) et en faisant varier les conditions de culture. L'expression différentielle des gènes de *C. arabica* par rapport aux espèces parentales a été caractérisée par une majorité de gènes qui montrait une dominance du niveau d'expression variable en fonction des conditions environnementales. Ces résultats suggèrent la plasticité transcriptomique de *C. arabica* (Bardil et al. 2011) et suggéreraient que même dans le cadre d'une faible diversité génétique, des mécanismes de régulation de l'expression des gènes pourraient intervenir pour permettre une adaptation environnementale.

5.4 Les hybrides interspécifiques néosynthétiques de *C. canephora* et *C. eugenoides*

Avant l'application des méthodes de marquage moléculaire aux espèces du genre *Coffea*, la structure génétique des caféiers diploïdes sauvages basée sur des caractères botaniques et écologiques faisait l'objet de discussions. Aussi au cours des années 1970, plusieurs programmes d'hybridation interspécifique entre caféiers diploïdes africains ont été entrepris à l'ORSTOM en Côte d'Ivoire. L'un d'eux avait pour objectif d'une part de repérer des voies d'amélioration de l'espèce cultivée *C. canephora* et d'autre part d'étudier les niveaux de différenciation entre espèces et d'établir une classification biosystématique des caféiers (Louarn 1976, 1993). Au sein du genre *Coffea*, toutes les espèces peuvent être hybridées. Des hybrides F1 ont été obtenus au moyen de croisements contrôlés entre caféiers diploïdes africains représentant la diversité d'origine (Figure 24). La fertilité des

hybrides F1 révélée entre autre par l'analyse de la viabilité pollinique des hybrides ont démontré l'affinité plus ou moins importante des caféiers diploïdes selon leur groupe d'origine (Figure 24).

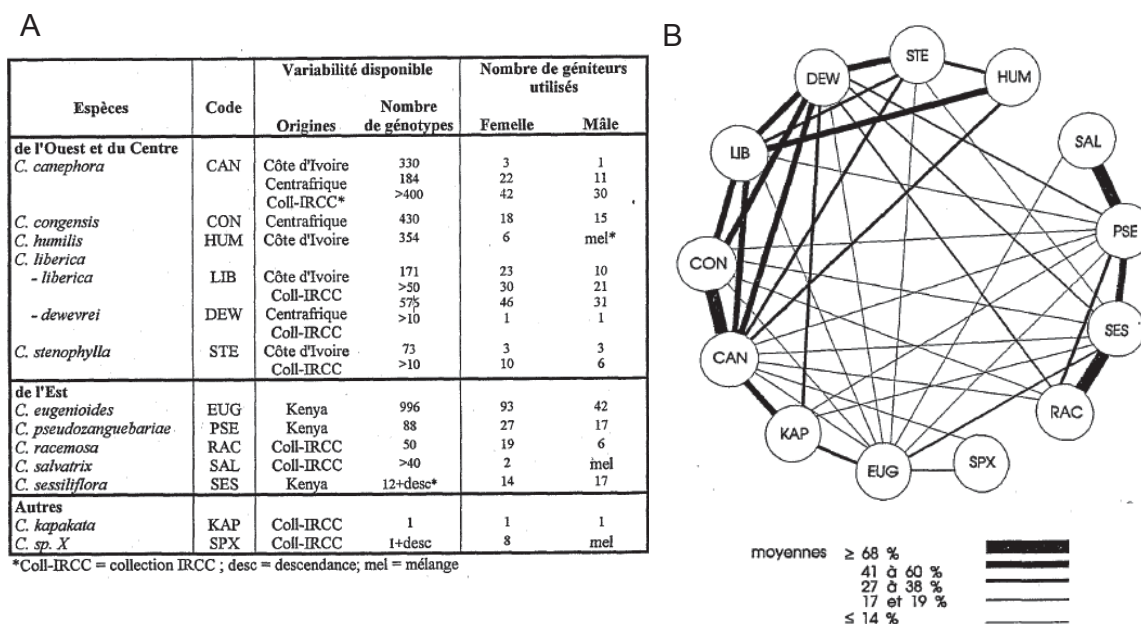


Figure 24 : A/ Origine et Diversité des caféiers diploïdes africains utilisés comme parents dans les croisements interspécifiques. B/ Viabilité pollinique des hybrides F1 décrivant leur fertilité (Louarn 1993).

Au cours de l'étude d'hybrides interspécifiques entre *C. canephora* et *C. eugenioides* (à l'origine de l'allotétraploïde naturel *C. arabica*), plusieurs caractéristiques morphologiques, caryotypiques et agronomiques ont été comparées entre les espèces parentales et les hybrides F1 (Louarn 1976). Le comportement méiotique et la fertilité des parents et des hybrides ont été aussi examinés. Les hybrides F1 ont été obtenus facilement, mais ils présentaient souvent une nature chétive. Selon les caractères considérés, ils ressemblaient à l'un des deux parents (ramification, taille des fleurs : comparable à *C. eugenioides*, granulométrie, inflorescence : comparables à *C. canephora*) ou bien étaient intermédiaires entre les deux espèces (teneur en caféine des graines, floraison, nouaison, maturité). L'analyse du comportement méiotique a révélé que des anomalies des appariements étaient fréquentes, seulement 35% des cellules montraient un appariement normal de onze bivalents. Ce résultat semblait lié à la baisse de la fertilité des hybrides par rapport à celle des parents, et permettait de conclure sur la faible différenciation entre les deux espèces (Louarn 1976)

Plus récemment, des hybrides entre les espèces *C. canephora* (C) et *C. eugenioides* (E) ont été réalisés à CENICAFE en Colombie (Centre National Colombien de Recherche sur le Café) (Figure 25). Ces hybrides F1, diploïdes ont été produits en utilisant chacune des espèces comme parent mâle et femelle. Ces plantes au phénotype très variables (Figure 26) ainsi que les génotypes des espèces parentales poussent en plein champ à la station expérimentale de Naranjal (Chinchina, Colombie).



Figure 25 : A/ *C. canephora*, B/ *C. eugenioides* (Photos : Juan-Carlos Herrera CENICAFE-Colombie)



Figure 26 : Diversité phénotypique des hybrides F1, en haut *C. canephora* x *C. eugenioides*, en bas *C. eugenioides* x *C. canephora*. Photos : Juan-Carlos Herrera CENICAFE-Colombie)

6 Les objectifs de la thèse

Les dix dernières années, les conséquences de l'allopolyploïdie sur la structure et le fonctionnement des génomes ont fait l'objet de plusieurs études. Les changements d'expression provoqués par l'évènement d'hybridation ont été examinés chez des polyploïdes, mais aussi chez de simples hybrides. Dans les deux cas, la comparaison de l'expression des génomes des espèces polyploïdes ou hybrides par rapport à celle des espèces parentales ont montré pour partie un écart à l'additivité, qui pourrait être à l'origine des phénotypes nouveaux qui caractérisent les espèces polyploïdes et les hybrides. Pour certaines études, l'analyse de l'expression totale des gènes a été complétée par l'étude de l'expression relative des allèles parentaux ou des gènes homéologues. Ainsi ces études initient l'observation de la régulation des gènes après la fusion des génomes et de leurs réseaux de régulation dans le même contexte cellulaire. Ces études ont montré une grande diversité de réponse des transcriptomes des polyploïdes et des hybrides face à l'hybridation. Certaines caractéristiques de la variation de l'expression des gènes sont partagées entre les modèles. Mais des particularités de certains modèles ne permettent pas d'établir un principe général de la variation de l'expression des gènes après un évènement d'hybridation et suggèrent de considérer les divergences génétiques et évolutives des espèces parentales.

En raison des limites technologiques rendant difficile la différenciation des gènes homéologues, l'étude de la variation de l'expression relative des gènes homéologues a été longtemps restreinte à quelques gènes. Récemment ces contraintes ont été réduites par l'essor des méthodes de séquençage à haut débit. Alors que pour quelques polyploïdes, l'expression relative des gènes homéologues a été décrite à l'échelle du génome, peu d'études ont été menées à cette échelle pour évaluer sa variation en fonction des organes de la plante ou de conditions variables (stress, conditions environnementales).

Au cours de cette thèse, nous proposons d'étudier un nouveau modèle de polyploïde : *C. arabica*, une plante pérenne issue de l'hybridation d'espèces parentales peu divergentes génétiquement mais adaptées à des environnements éco-géographiques différents. Son aptitude à tolérer de plus grandes variations thermiques que ses espèces parentales, lui ait permis de coloniser de nouvelles zones géographiques. D'une part nous étudierons les conséquences de l'hybridation sur la variation de l'expression et de la régulation des gènes et d'autre part nous analyserons la variation de l'expression des gènes homéologues chez *C. arabica* et son implication dans les capacités adaptatives de l'allopolyploïde.

- **Etude des conséquences de l'hybridation sur l'expression et la régulation des gènes**

Cette étude sera présentée dans le chapitre 2. Il s'agira d'analyser le transcriptome d'hybrides F1 des deux espèces parentales de *C. arabica* au moyen de la technologie RNA-seq. Une première étape consistera à caractériser l'expression des hybrides par rapport à celle des espèces parentales. L'analyse

de l'expression allélique spécifique permettra ensuite de mettre en évidence les divergences de régulation des espèces parentales et de décrire l'expression et la régulation des allèles au sein des hybrides.

- **Analyse de la variation de l'expression des gènes homéologues chez *C. arabica* et son implication dans les capacités adaptatives de l'allopolyploïde**

Cette analyse sera traitée dans le chapitre 3 au moyen de deux études. Une première étude consistera à caractériser pour *C. arabica* l'expression d'un petit nombre de gènes homéologues appartenant à une région génomique limitée. La variation de l'expression relative de ces gènes homéologues sera ensuite analysée en fonction de plusieurs tissus de la plante, et deux conditions de culture (température de culture différente). Une méthode consistant à comparer sur un chromatogramme, les hauteurs de pics des nucléotides d'un SNP différenciant les gènes homéologues sera utilisée pour traiter ce petit jeu de données.

Dans la seconde étude, la caractérisation de l'expression relative des gènes homéologues de *C. arabica* et l'étude de sa variation sera étendue à l'échelle du génome en utilisant la technologie RNA-seq pour analyser le transcriptome. L'implication de la variation de l'expression des gènes homéologues dans les capacités adaptatives de *C. arabica* sera traitée en comparant l'expression des gènes homéologues à deux conditions de culture (également température de culture différente).

Dans le chapitre 4, les résultats des trois études seront d'abord discutés et ensuite des perspectives de recherche de ce travail seront proposées.

CHAPITRE 2

Etude d'une étape majeure de l'allopolyploïdie :
l'hybridation génomique

CHAPITRE 2 :

Etude d'une étape majeure de l'allopolyploïdie : l'hybridation génomique

1 Evolution et étude de l'expression des gènes

Au cours de l'évolution, en raison de la sélection naturelle et de la dérive génétique, des espèces proches vont accumuler des différences génétiques pour des gènes orthologues. La coévolution des gènes codant pour des éléments cellulaires qui interagissent permet le maintien des fonctions cellulaires et moléculaires malgré l'accumulation de variations génétiques sur les séquences codantes ou régulatrices (Landry et al. 2007, Barreto et al. 2015). Depuis plusieurs dizaines d'années, les biologistes ont pris conscience que le polymorphisme fonctionnel des gènes, conduisant à des changements phénotypiques, caractéristiques d'espèces, n'est pas seulement lié à des variations des séquences codantes mais est aussi dépendant de la modulation de l'expression des gènes (Prud'homme et al. 2007, Emerson et Li 2010, Wolf et al. 2010). Les variations génétiques comme des substitutions peuvent donc survenir sur des régions codantes et modifier la séquence en acides aminés ou la séquence de l'ARN. Elles peuvent aussi se produire sur les régions *cis*-régulatrices en amont des gènes et dans ce cas là provoquer une variation de l'expression des gènes (Figure 27). Enfin en cas de changements génétiques de plus grande taille, par exemple délétion, insertion, réarrangement, les régions codantes et régulatrices peuvent être atteintes à la fois. Dans le domaine de la biologie évolutive du développement, plusieurs hypothèses ont été émises, « l'hypothèse de régulation » et «

l'hypothèse de *cis*-régulation » pour décrire l'impact des mutations sur la régulation des gènes et leur rôle sur les variations morphologiques des espèces (Stern et Orgogozo 2008). Dans « l'hypothèse de régulation », l'évolution phénotypique et morphologique est causée principalement par des changements de régulation de l'expression des gènes. Ces changements de régulation peuvent résulter de mutations des régions régulatrices (*cis*-régulation) mais aussi des régions codantes, par exemple d'un facteur de transcription qui régule le gène cible (Hoekstra et Coyne 2007). En revanche dans « l'hypothèse de *cis*-régulation », les mutations de la région *cis*-régulatrices sont les principales causes de l'évolution phénotypique (Prud'homme et al. 2007) à l'inverse d'autres changements des régions codantes, de l'épissage alternatif, d'événements de duplication ou de délétion...

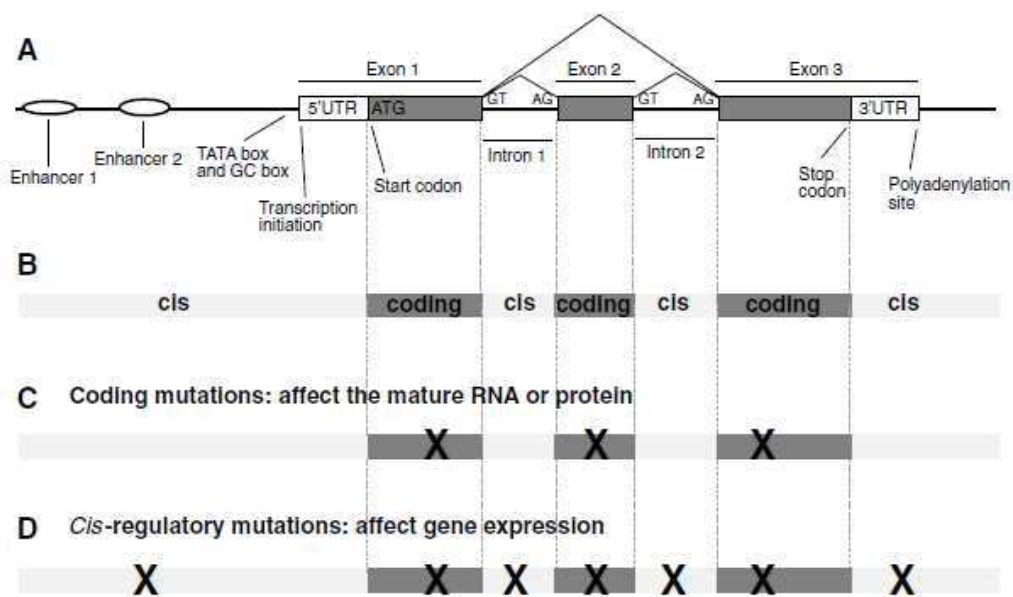


Figure 27 : Structure du gène et présentation des régions codantes et de *cis*-régulation. (A) Pour un gène simple, de nombreuses instructions sont codées sur la séquence d'ADN. La partie codante du gène est représentée par les bandes grises. L'ARNm est transcrit du codon start jusqu'au site de polyadénylation et les introns sont épissés. De nombreux gènes peuvent coder pour plusieurs ARNm d'épissage alternatif variable, générés par le choix différents d'exons utilisés. Les variants provenant d'épissage alternatif différent sont habituellement exprimés dans différents organes et à différentes périodes du développement. L'ARNm contient des régions 5' et 3' qui ne seront pas traduites et qui sont impliquées dans la stabilité de l'ARN, la localisation de l'ARNm et la traduction. (B) Les gènes sont divisés en séquences codantes comprenant les exons et en séquences de *cis*-régulation incluant les parties de l'ADN qui régulent l'expression. Les séquences de *cis*-régulation régulent la transcription, la stabilité de l'ARN, l'épissage et la traduction. (C) Les mutations de codage sont des mutations qui changent la séquence d'acides aminés codées par l'ARNm ou qui changent la séquence nucléotidique de l'ARNm mature. (D) Les mutations de *cis*-régulation se produisent la plupart du temps sur des séquences non-codantes et provoquent des changements de l'expression des gènes mais elles peuvent aussi se produire plus rarement sur des séquences codantes (Stern et Orgogozo 2008).

Les mutations des séquences régulatrices qui sont potentiellement plus malléables que les séquences codantes auraient un impact phénotypique important. Le dynamisme d'une transcription finement et continuellement régulée permet une adaptation et une évolution bien plus rapide qu'une mutation de structure (Prud'homme et al. 2007). Un argument important en faveur de la divergence adaptative en

cis-régulation est la spécificité de la régulation de la transcription en fonction des allèles, des tissus, d'un état particulier de développement, qui évite les effets pléiotropiques néfastes (Prud'homme et al. 2007, Wittkopp et Kalay 2012). Cependant un rôle important de mutations provenant de séquences codantes (par exemple les facteurs de *trans*-régulation) qui influent sur d'autres mécanismes de régulation ne peut pas être écarté (Hoekstra et Coyne 2007, Stern et Orgogozo 2008).

L'hybridation seule ou bien associée à un évènement de duplication génomique (allopolyploïdisation) est un phénomène très répandu parmi les plantes, qui peut avoir des conséquences importantes pour leur histoire évolutive. Le croisement entre espèces (interspécifique) ou entre populations (intraspécifique) rassemble dans un même organisme des génomes parentaux qui ont subi des évolutions indépendantes et n'ont jamais interagi auparavant. Lorsque l'hybridation se produit, elle provoque l'émergence de nouveaux phénotypes. Parmi toutes les raisons possibles à l'origine de ces nouveaux phénotypes, l'interaction des réseaux de régulation de la transcription des espèces parentales induirait de nouveaux profils d'expression des gènes au sein des hybrides (Landry et al. 2007).

De nombreuses études ont été menées pour analyser l'expression des gènes d'hybrides intra ou interspécifiques (Emerson et Li 2010, McManus et al. 2010, Shi et al. 2012, Bell et al. 2013, Coolon et al. 2014) et de jeunes allopolyploïdes (Buggs et al. 2014). La régulation de l'expression des gènes déterminant le phénotype de l'organisme apparaît en partie dépendante des différences d'expression des gènes provenant des espèces parentales.

Que cela soit pour étudier l'émergence d'individus présentant de nouveaux phénotypes après hybridation ou allopolyploïdisation, ou bien pour étudier les mécanismes d'évolution adaptative des espèces parentales, la compréhension des bases génétiques de l'hybridation est devenue une étape essentielle de l'étude de l'expression des gènes (Emerson et Li 2010). L'essor technologique des méthodes de séquençage d'acides nucléiques à haut débit permet depuis quelques années d'étudier la régulation des gènes en distinguant les effets de *cis*-régulation des effets de *trans*-régulation à l'échelle du génome, alors que ce n'était envisageable auparavant que pour quelques gènes.

2 Régulation de l'expression des gènes

L'expression des gènes débute avec la transcription d'une séquence d'ADN codant pour un gène. Elle est contrôlée par des interactions entre la séquence d'ADN en *cis*-régulation et des éléments de régulation agissant en *trans*. Littéralement *cis* et *trans* signifient du même côté et du côté opposé. La région en *cis*-régulation est située sur la même molécule d'ADN que le gène à réguler. Elle est à proximité de ce gène et est composée d'un promoteur basal positionné en amont du codon start, sur le site d'assemblage du complexe ARN polymérase et protéine. Ce complexe transcrit un niveau d'ARN

basique qui peut être modifié par des « enhancers » (amplificateurs). « L'enhancer » regroupe des sites de fixation d'éléments de *trans*-régulation qui peuvent augmenter ou réprimer la transcription. Ce sont des molécules qui diffusent dans la cellule et dont l'activité n'est spécifique ni du gène ni de l'allèle. Plusieurs gènes peuvent être régulés par le même facteur de *trans*-régulation, et de même, deux allèles à un locus peuvent être modulés par un facteur de régulation agissant en *trans*. En revanche les éléments de *cis*-régulation sont spécifiques des allèles. Par conséquent, la variation génétique pour l'expression d'un gène régulé en *cis* devrait ségréger avec l'allèle, alors que la variation génétique qui ségrégerait indépendamment d'un allèle devrait agir en *trans*. En effet les éléments de *trans*-régulation ne sont pas forcément codés par des séquences d'ADN proches du gène/de l'allèle à réguler. L'interaction entre des facteurs de transcription et « l'enhancer » peut provoquer des changements de la forme de la molécule de l'ADN. Ainsi la structure de la chromatine, la séquence de l'ADN en 5' et 3' des régions non traduites des gènes peuvent avoir aussi une activité de *cis*-régulation (donc spécifique d'un allèle). Par ailleurs pour la régulation en *trans*, des petits ARNs, des protéines et même l'environnement peuvent intervenir sur la disponibilité et l'activité des facteurs de transcription et ainsi avoir un effet sur la *trans*-régulation de la transcription. Ensemble, tous ces éléments et ces interactions entre les facteurs de *cis* et *trans*-régulation créent un réseau complexe qui régule la transcription (Figure 28).

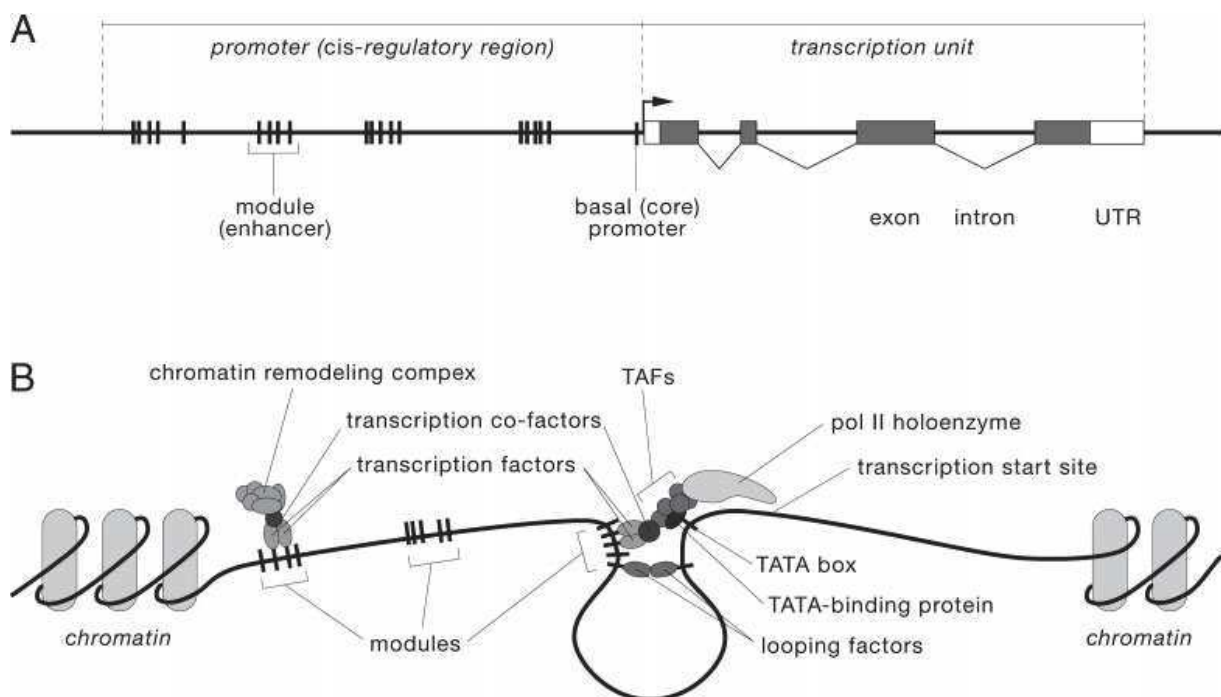


Figure 28 : Représentations de la région cis-régulation d'un gène. (A) Représentation linéaire. La région *cis*-régulation contient un promoteur de base d'où débute la transcription, plusieurs « enhancers » regroupant des sites de fixation de plusieurs facteurs de transcription. (B) Schéma représentant les effets de la structure sur la mise en route de la machinerie de la transcription. Un assemblage complexe composé d'ARN polymérase, de protéines, de facteurs de transcription (TAFs : TATA-associated factors) se fixe sur les sites de fixation en amont du site de départ de la transcription, remodele la chromatine en provoquant une boucle pour rapprocher des sites de fixation de facteurs de transcription (Coolon et Wittkopp 2013).

3 Les hybrides interspécifiques

En raison de la complexité de la régulation de l'expression des gènes, dans de nombreuses études, l'analyse de la variation de l'expression des gènes est abordée, à l'échelle de quelques gènes ou à l'échelle génomique, en classant les changements génétiques qui affectent l'expression des gènes soit en *cis*- soit en *trans*-régulation. Pour cela des hybrides F1 intra ou interspécifiques sont produits entre des espèces parentales proches. L'estimation de l'expression allélique spécifique (ASE : Allele-Specific Expression) est utilisée pour distinguer les effets de *cis*- et *trans*-régulation (Cowles et al. 2002, Wittkopp et al. 2004). Cette approche peut être employée pour étudier les mécanismes moléculaires à l'origine de divergences d'expression dans une espèce, mais aussi entre espèces si des hybrides interspécifiques peuvent être produits. En plus de révéler des différences génétiques des génotypes parentaux, l'étude des variations d'expression des gènes des hybrides F1 par rapport à l'expression des gènes des espèces parentales informe sur la conciliation des réseaux de régulation divergents des espèces parentales.

Après l'hybridation, les allèles *cis*-régulés de l'hybride (provenant directement des allèles des génotypes parentaux, $S1_{F1}$ et $S2_{F1}$) sont dans le même environnement cellulaire, soit le même environnement de *trans*-régulation. Les différences d'expression entre deux allèles indiquent des différences dues à des divergences de *cis*-régulation entre les allèles des espèces parentales, mais aussi à des divergences de *trans*-régulation, si la totalité de la variation d'expression allélique au sein des hybrides n'est pas expliqué par les divergences de *cis*-régulation entre les allèles des espèces parentales. Ainsi les relations entre ces deux origines de variation d'expression allélique peuvent être schématisées (Figure 29) et représentées par $P = T + C$. P est l'expression relative entre les espèces $S1$ et $S2$ pour un gène considéré ($S1/S2$), C est l'expression relative des allèles F1 ($S1_{F1}/S2_{F1}$) au sein de l'hybride (s'il n'y pas de divergences de *trans*-régulation entre les espèces parentales, $P = C$, ou $S1/S2 = S1_{F1}/S2_{F1}$) et enfin, T est l'évaluation de variations d'expression entre les espèces parentales dues à des différences de *trans*-régulation et est déduite de P et de C . Toutes les différences d'expression entre les espèces parentales qui ne sont pas dues à des différences de *cis*-régulation sont attribuées à des différences de *trans*-régulation.

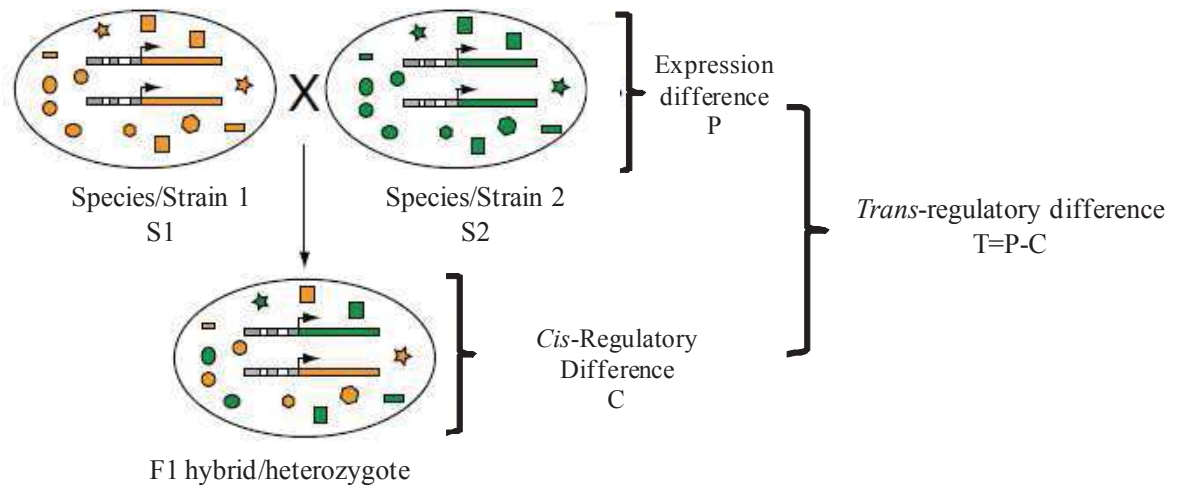


Figure 29 : Dispositif expérimental utilisant l'estimation de l'ASE afin de déterminer les différences de *cis*- et *trans*-régulation entre deux individus diploïdes. Deux espèces ou deux souches sont croisées pour produire des hybrides F1, qui auront un environnement de *trans*-régulation commun. Les molécules agissant en *trans*-régulation pour chacun des individus sont indiquées par des couleurs différentes. P est la différence d'expression entre les espèces ou les souches parentales, C indique les différences de *cis*-régulation et T désigne les différences de *trans*-régulation déduites de P et C (Coolon et Wittkopp 2013).

Sur le plan méthodologique, cette approche consiste à évaluer d'abord la divergence d'expression entre les espèces parentales due à des différences de *cis*-régulation. Pour chaque gène, le rapport de l'expression des allèles parentaux ($S1/S2$) est comparé avec l'expression relative des allèles de l'hybride ($S1_{F1}/S2_{F1}$). Les différences d'expression relative conservées entre les espèces parentales et les allèles de l'hybride témoignent de différences de *cis*-régulation entre les espèces parentales (zone noire sur la Figure 30). Des différences d'expression entre les allèles parentaux qui ne sont pas conservées et correspondant à une expression équivalente des allèles au sein de l'hybride indiquent l'activité de facteurs de *trans*-régulation (zone rouge sur la Figure 30). Pour les gènes n'entrant dans aucune des deux catégories décrites, la régulation de l'expression serait la combinaison de facteurs de *cis*- et *trans*-régulation (zone verte et bleue sur la Figure 30). La représentation de la comparaison de l'expression relative des allèles de l'hybride ($S1_{F1}/S2_{F1}$) et des gènes des espèces parentales ($S1/S2$) pour chaque gène sur un graphe révèle la diversité des combinaisons des éléments de *cis*- et *trans*-régulation (Figure 30).

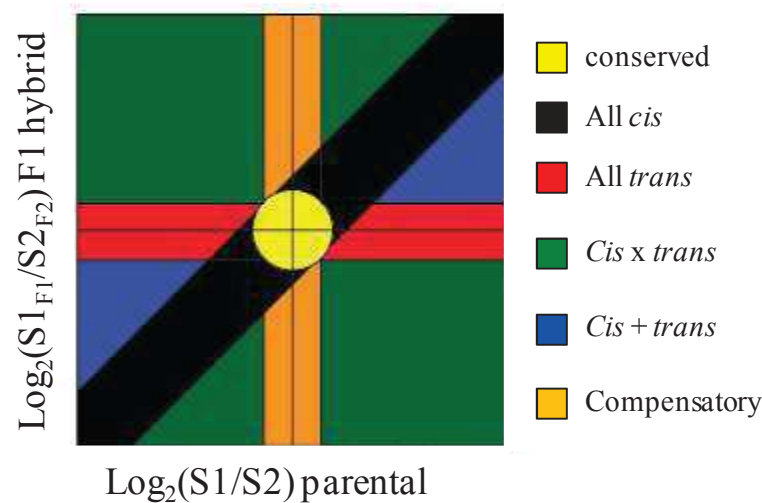


Figure 30 : Schéma pour l'interprétation de résultats d'une étude d'ASE. L'expression relative des espèces parentales (en $\log_2$) reportée sur l'axe des abscisses est comparée à l'expression relative des allèles dans l'hybride (en $\log_2$) reportée sur l'axe des ordonnées. Chaque point représentant un gène de l'étude se positionnera sur une des zones décrites en couleur.

En effet les changements de *cis*- et *trans*-régulation ne sont pas mutuellement exclusifs, et de nombreux gènes peuvent présenter à la fois des différences de *cis*- et *trans*-régulation entre espèces. Pour certains gènes, des divergences en *cis*- et *trans*-régulation vont se cumuler, favoriser l'expression du même allèle (*cis* + *trans*, « enhancing », zone bleue sur la Figure 30) ou bien de allèle opposé et dans ce cas là, compenser les différences d'expression (*cis* x *trans*, «compensating », (zone verte sur la Figure 30). Ces différentes interactions proviennent de mutations qui auraient affecté les éléments de *cis* et *trans*-régulation des gènes parentaux. Au sein de l'hybride, les gènes pour lesquels les divergences de *cis* et *trans*-régulation se cumulent révèlent des cas d'évolution non conservatrice des éléments de *cis* et *trans*-régulation, c'est-à-dire que le niveau d'expression d'un gène n'est pas conservé entre les espèces analysées, les éléments de *cis* et *trans*-régulation vont évoluer tous les deux en provoquant l'augmentation de l'expression d'un gène ou inversement la diminution. Mais les éléments de régulation évoluent dans la même direction. A l'inverse pour des gènes présentant des divergences de *cis* et *trans*-régulation qui se compensent, le niveau d'expression des gènes est conservé entre les deux espèces analysées, l'évolution des éléments de *cis* et *trans*-régulation va se produire en compensant les mutations d'un ou l'autre des éléments de régulation pour maintenir le niveau d'expression d'un gène (Tirosh et al. 2009). Chez la levure, un exemple de coévolution des éléments de *cis* et *trans*-régulation avec des mutations compensatoires a été décrit (Kuo et al. 2010). Plusieurs modèles de co-évolution compensatoire sont proposées, pour l'un d'eux des mutations se

succèderaient. Une première mutation pourrait se produire et être neutre pour le contexte génétique considéré. Mais en changeant la séquence, cette mutation changerait le contexte génétique pour les mutations suivantes qui seraient elles aussi neutres dans le nouveau contexte génétique. Mais si ces dernières mutations s'étaient produites dans le contexte génétique initial, elle auraient été délétères. De tels changements n'affecteraient pas le phénotype entre espèces mais pourraient avoir des effets sur l'expression au sein des hybrides après un événement d'hybridation. Ils dépendraient de l'architecture du réseau de transcription (nombre de sites de fixation de facteurs de transcription, abondance et affinité des facteurs de transcription) et de la nature épistatique des mutations (Landry et al. 2007).

L'estimation de l'ASE d'hybrides F1 a été très pratiquée pour étudier le polymorphisme de régulation dans et entre espèces pour des animaux, (Cowles et al. 2002, Wittkopp et al. 2004, Graze et al. 2009, McManus et al. 2010, Graze et al. 2012, Coolon et al. 2014), des levures (Tirosh et al. 2009), et des plantes (Bell et al. 2013), cette approche a été aussi appliquée à des allopolyploïdes de synthèse de plante (Chaudhary et al. 2009, Shi et al. 2012).

4 L'étude d'hybrides interspécifiques de *Coffea*

Les hybrides diploïdes F1 ont été produits en utilisant comme parent femelle C (*C. canephora*) pour les hybrides CxE et E (*C. eugenoides*) pour les hybrides ExC. Ces plantes ainsi que les génotypes des espèces parentales poussent en plein champ à la station expérimentale de Naranjal (Chinchina, Colombie). Pour réaliser une étude de la variation de l'expression des gènes suite à l'hybridation, six plantes provenant du groupe CxE et 4 du groupe d'hybrides réciproques ExC ont été sélectionnées. Trois génotypes représentatifs de l'espèce *C. canephora* et quatre de l'espèce *C. eugenoides* complètent le dispositif expérimental (Figure 31). De jeunes feuilles complètement développées ont été prélevées sur les plantes sélectionnées. L'ARN total de ces feuilles a été extrait et l'ARNm séquencé en vue de l'estimation de l'ASE des deux groupes d'hybrides.

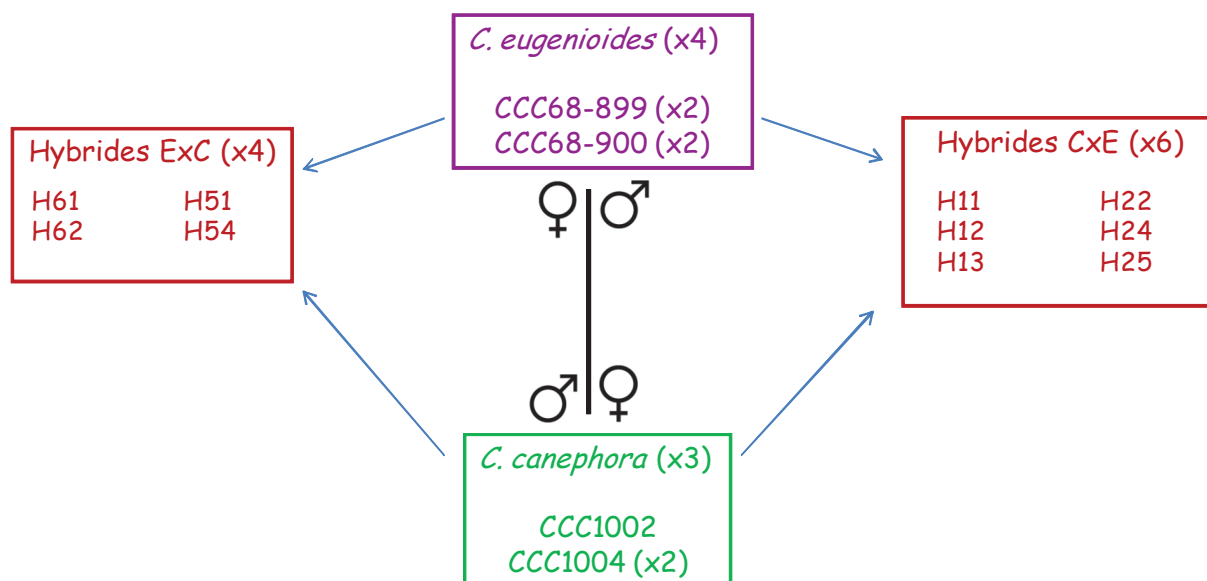


Figure 31 : Dispositif expérimental de l'étude d'hybrides interspécifiques de *Coffea*. Les espèces *C. canephora* et *C. eugenoides* sont réciproquement utilisées comme parent femelle et mâle. Le nombre de répétitions biologiques est indiqué entre parenthèses.

4.1 L'apport de connaissances des hybrides interspécifiques

Pour l'ensemble des études d'hybrides interspécifiques et même d'allopolyploïdes synthétiques, l'apport de connaissances des hybrides est double. Selon que l'estimation de l'ASE soit considérée par rapport à l'expression relative des espèces parentales ou bien par rapport à la différence d'expression entre les espèces parentales et les hybrides, l'étude informera sur la différence de régulation des gènes entre les espèces parentales ou sur les nouveaux profils d'expression des gènes au sein des hybrides.

Ainsi en ce qui concerne les hybrides interspécifiques de *Coffea*, les données analysées pourront indiquer des différences de régulation entre *C. canephora* et *C. eugenoides*, plusieurs informations peuvent être déduites de l'analyse de l'ASE : la fréquence relative des changements de *cis*- et *trans*-régulation entre *C. canephora* et *C. eugenoides* et la nature du processus d'évolution, neutre ou non, qui contribue à la divergence de *cis*- et *trans*-régulation entre ces espèces.

Ensuite, ces données permettront de caractériser l'expression des gènes au sein des hybrides de *Coffea*, c'est-à-dire de définir le niveau d'expression des gènes, l'hérédité de l'expression par rapport aux espèces parentales. En complétant ces observations par l'estimation de l'ASE, la réorganisation de la régulation de l'expression des gènes suite à l'hybridation et à la combinaison des réseaux de régulation des gènes des espèces parentales pourra être explorée. Les effets des facteurs de *trans*-régulation : leur importance et leur mode d'action seront définis. Enfin les relations entre l'hérédité de l'expression des gènes et l'ASE pourront être élucidées.

4.2 Méthodologie développée pour l'étude des hybrides interspécifiques

Les méthodes développées pour analyser les données de ce dispositif expérimental ont été de deux types : d'abord l'expression des gènes au sein de l'hybride a été considérée de façon globale sans distinguer l'expression particulière de chacun des allèles et ensuite dans un second temps l'expression de chaque allèle a été estimée (ASE).

4.2.1 Hérité de l'expression des gènes dans les hybrides interspécifiques de *Coffea*

Dans une première étape, il s'agissait d'une part de comparer pour chaque gène les niveaux d'expression des espèces parentales avec celui des hybrides en effectuant des tests binomiaux deux à deux au moyen du package DESeq (Anders et Huber 2010) et d'autre part de classer les gènes en fonction de l'hérité de l'expression des gènes dans l'hybride par rapport aux espèces parentales. Ce classement a été réalisé en combinant les résultats de trois tests binomiaux réalisés au moyen de DESeq (CxE/C, CxE/E et C/E). Mais pour de nombreux gènes (plus de 16%) par cette méthode, il n'était pas possible de définir l'hérité de l'expression des gènes dans l'hybride car le classement obtenu par deux des tests n'était pas cohérent avec le troisième test (C/E). Alors pour effectuer ce classement, la méthode utilisée a été inspirée de l'analyse de McManus et al (McManus et al. 2010). Cette méthode est limitée à deux comparaisons. Pour chaque gène, les changements d'expression au sein de l'hybride sont évalués par rapport aux espèces parentales en soustrayant le niveau d'expression du gène de l'hybride à celui de chacun des gènes des espèces parentales. Les gènes pour lesquels l'expression totale dans l'hybride dévie de plus de 1.25 fois par rapport à celle des espèces parentales présentent une hérité de l'expression des gènes non-conservée par rapport aux espèces parentales. Les classes d'hérité de l'expression des gènes de l'hybride par rapport aux espèces parentales sont ensuite révélées en reportant ces valeurs sur un graphe qui exprime les variations d'expression par rapport à l'espèce parentale *C. canephora* en fonction des variations d'expression par rapport à l'espèce parentale *C. eugenioides* (Figure 32, article GBE). Cette méthode réduit le nombre de classes d'hérité considérées (notamment celles qui permettent de décrire les variations d'expression entre les espèces parentales quand l'hybride s'exprime différemment d'elles deux), mais elle a eu l'avantage de permettre le classement de l'hérité de l'expression des gènes de l'hybride pour un plus grand nombre de gènes.

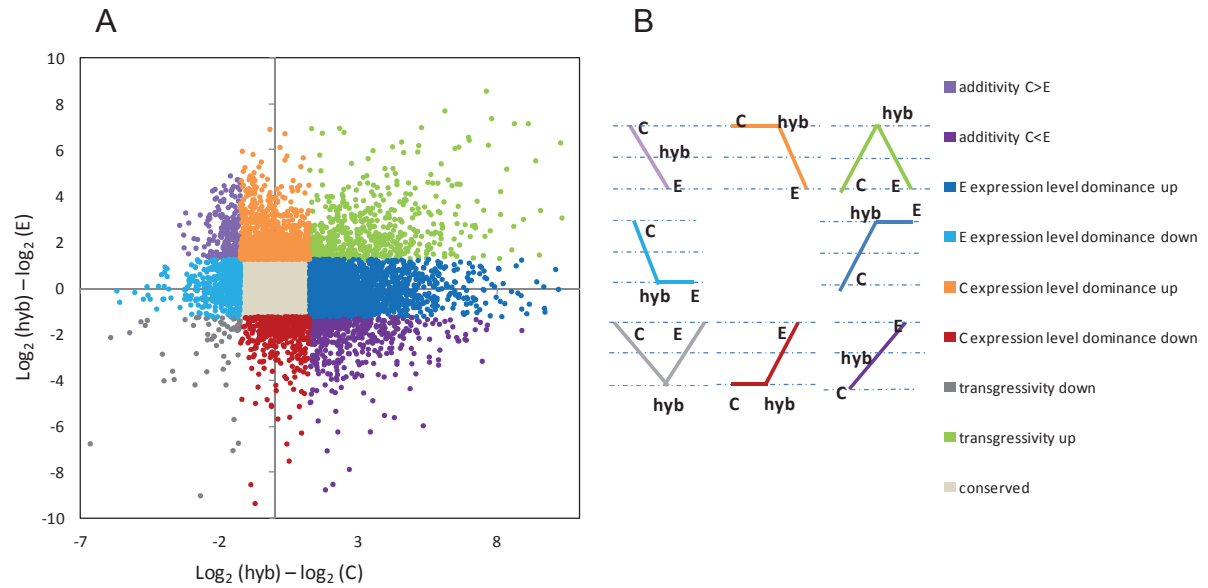


Figure 32 : Hérédité du niveau d'expression des gènes. (A) Diagramme représentant la distribution des catégories de l'hérédité du niveau d'expression des gènes, après comparaison des différences d'expression de l'hybride avec chacune des espèces parentales. (B) Profils hypothétiques d'expression d'un gène par rapport aux deux espèces parentales, chaque couleur correspondant à une catégorie est nommée dans la légende ci-contre, la catégorie de gènes « conserved » n'est pas représentée, elle rassemble les gènes pour lesquels il n'y a aucun changement significatif d'expression entre l'hybride et les espèces parentales.

4.2.2 Analyse de l'ASE des hybrides interspécifiques de *Coffea*

Pour la deuxième étape de l'analyse, seuls les gènes pour lesquels il a été possible d'estimer l'expression de chacun des deux allèles ont été retenus. Les données d'ASE pour chacun des individus de deux groupes d'hybrides ont été normalisées par DESeq sur le nombre total de séquences de chaque allèle en prenant en compte la variance des répétitions biologiques. Pour évaluer la qualité des répétitions biologiques, les données de l'ASE des hybrides au sein de chaque groupe d'hybrides et ensuite entre les deux groupes d'hybrides ont été comparées. Toutes les paires d'hybrides comparés ainsi que la comparaison entre les deux groupes d'hybrides présentent des coefficients de corrélation significatifs (Tableau III, Figure 33). Les estimations de l'ASE des hybrides comparées deux à deux présentent des coefficients de corrélation significatifs (Tableau III), les valeurs de l'ASE des hybrides sont alors cumulées pour chaque groupe d'hybrides. De même les estimations de l'ASE entre les deux groupes d'hybrides sont aussi significativement corrélées (Figure 33).

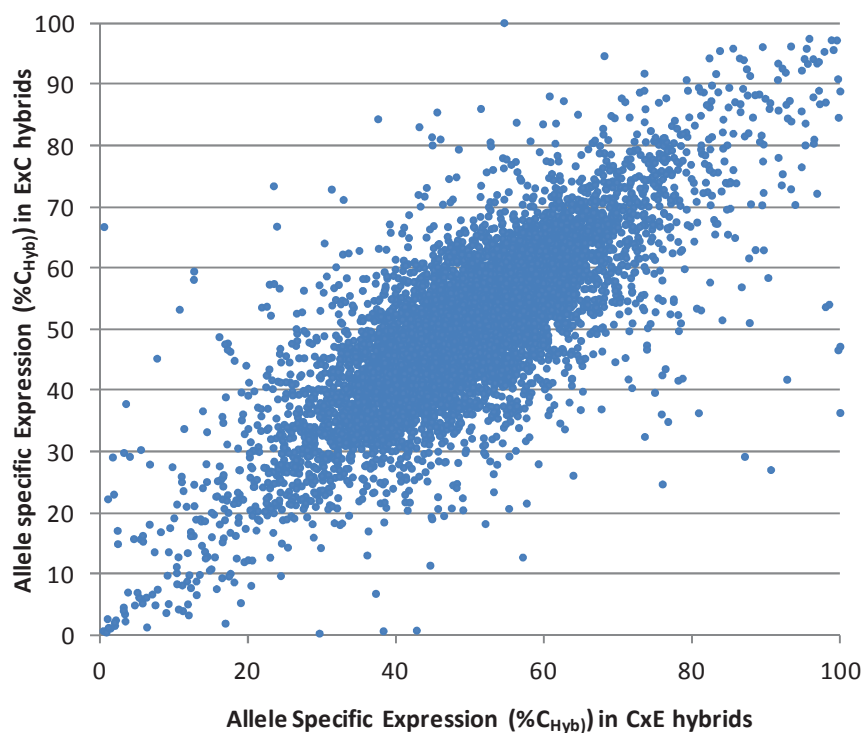
Tableau III : Corrélation de l'ASE entre les différentes répétitions des deux groupes d'hybrides.

CxE hybrids	H11	H12	H13	H22	H24	H25
H11	1					
H12	0.705	1				
H13	0.660	0.770	1			
H22	0.547	0.613	0.604	1		
H24	0.487	0.535	0.544	0.662	1	
H25	0.514	0.587	0.565	0.793	0.642	1

$P < 2.2 \times 10^{-16}$ in each comparison

ExC hybrids	H51	H54	H61	H62
H51	1			
H54	0.621	1		
H61	0.559	0.554	1	
H62	0.593	0.522	0.658	1

$P < 2.2 \times 10^{-16}$ in each comparison

**Figure 33 : Corrélation de l'ASE entre les deux groupes d'hybrides. ($r = 0.78$, $P > 0.001$)**

Comme nous l'avons vu dans ce chapitre précédemment, la première analyse, classique pour les études d'hybrides interspécifiques, a consisté à comparer l'expression relative des allèles dans l'hybride ($C_{\text{hyb}}/E_{\text{hyb}}$) avec l'expression relative des allèles dans les espèces parentales (C/E) pour révéler les différences de *cis*- et *trans*-régulation entre les espèces parentales, mais aussi l'interaction

des réseaux de régulation des gènes des espèces parentales. Une deuxième analyse plus originale a été ensuite menée sur les gènes qui présentaient une expression régulée par des facteurs de *trans*-régulation. Il s'agissait de déterminer pour chaque gène de cette catégorie, l'effet des facteurs de *trans*-régulation, sur ou sous-régulation, et ensuite la cible allélique des facteurs de *trans*-régulation. Pour cela, l'analyse de la variation de l'expression des allèles des deux génomes entre les contextes parentaux et hybrides a été effectuée. Pour chaque gène, la variation allélique pour l'allèle C entre l'hybride et *C. canephora* (C_{hyb}/C) et la variation allélique pour l'allèle E entre l'hybride et *C. eugenioides* (E_{hyb}/E) ont été comparées. Sachant que les facteurs de *cis*-régulation spécifique de l'allèle et donc des génomes sont similaires entre le contexte parental et le contexte hybride, les effets de *trans*-régulation ainsi que leur origine ont pu ainsi être révélés. Sur un graphe à l'échelle logarithmique représentant l'expression relative allélique du génome *C. canephora* en fonction de l'expression relative allélique du génome *C. eugenioides*, des valeurs positives des expressions relatives (C_{hyb}/C et E_{hyb}/E) indiquent la sur-régulation des allèles par les facteurs de *trans*-régulation du génome homologue, des valeurs négatives indiquent la sous-régulation des allèles. La position du gène représenté par un point par rapport aux deux axes révèle l'origine génomique des facteurs de *trans*-régulation et leur action simple ou combinée (Figure 34). Enfin, en reportant pour chaque gène, sur cette même figure (Figure 35), la catégorie d'hérédité de l'expression à laquelle il appartient, les divergences de *trans*-régulation sous-jacentes à l'hérédité de l'expression des gènes peuvent être révélées.

Ce travail a été publié dans l'article suivant : **Regulatory divergence between parental alleles determines gene expression patterns in hybrids**. Marie Christine Combes, Yann Hueber, Alexis Dereeper, Stéphanie Rialle, Juan-Carlos Herrera, Philippe Lashermes. 2015. *Genome Biology Evolution*. 7 : 1110-1121.

Annexe 1: Figures supplémentaires

Annexe 2 : Tableau supplémentaire

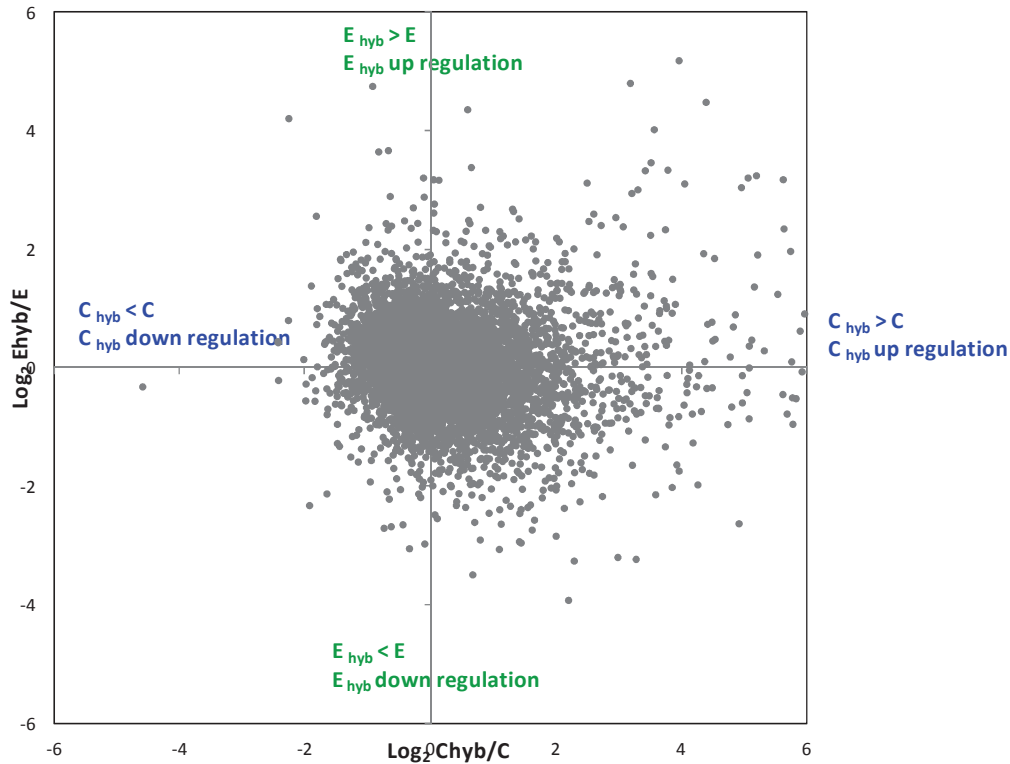


Figure 34 : Etude des facteurs de *trans*-régulation. Pour les gènes présentant des divergences de *trans*-régulation, la comparaison de la variation allélique entre les contextes hybride et parentaux pour les deux génomes révèle l'origine de l'allèle régulé et sa sur ou sous-régulation. Chaque gène représenté par un point indique par sa position sur le diagramme l'action simple (d'un génome) ou combinée (des deux génomes) de facteurs de *trans*-régulation.

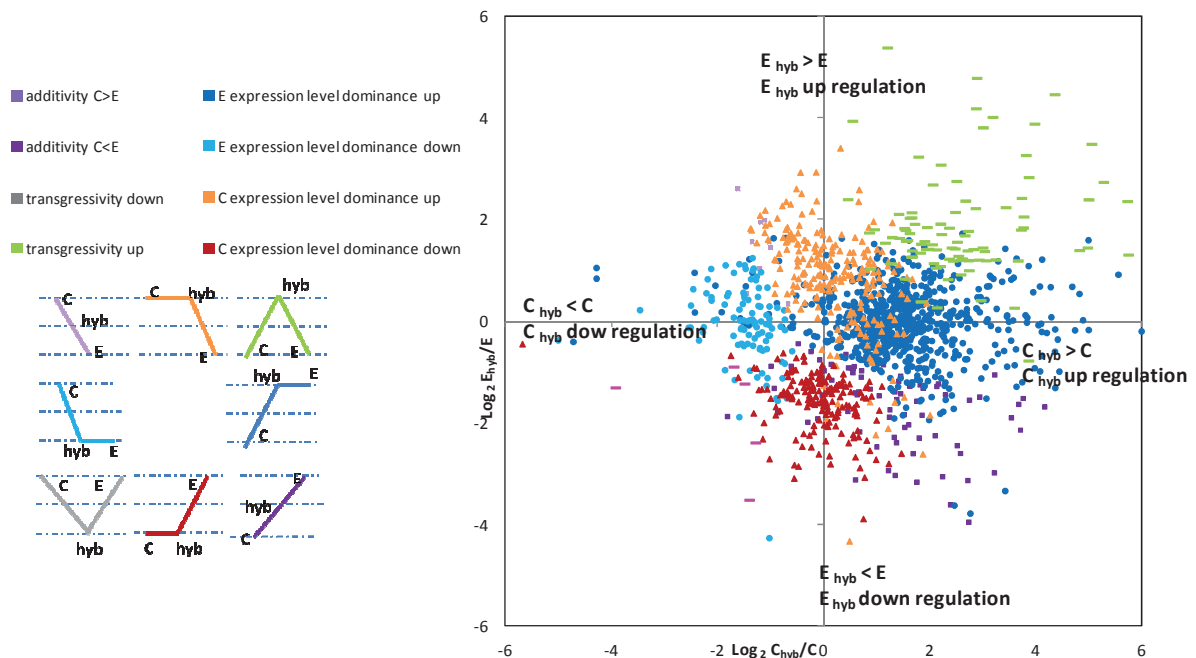


Figure 35 : Révélation des divergences de *trans*-régulation sous-jacentes à l'hérédité de l'expression des gènes de l'hybride. Représentation distincte des catégories d'hérédité d'expression des gènes pour les gènes ayant révélé des divergences de *trans*-régulation.

Regulatory Divergence between Parental Alleles Determines Gene Expression Patterns in Hybrids

Marie-Christine Combes^{1,*}, Yann Hueber¹, Alexis Dereeper², Stéphanie Rialle³, Juan-Carlos Herrera⁴, and Philippe Lashermes¹

¹IRD, UMR DIADE, Montpellier Cédex 5, France

²IRD, UMR IPME, Montpellier Cédex 5, France

³MGX-Montpellier GenomiX, Institut de Génomique Fonctionnelle, Montpellier Cédex 5, France

⁴Centro Nacional de Investigaciones de Cafe, CENICAFE – FNC, Manizales, Colombia

*Corresponding author: E-mail: marie-christine.combes@ird.fr.

Accepted: March 18, 2015

Data deposition: This project has been deposited at European Nucleotide Archive under the accession PRJEB7565.

Abstract

Both hybridization and allopolyploidization generate novel phenotypes by conciliating divergent genomes and regulatory networks in the same cellular context. To understand the rewiring of gene expression in hybrids, the total expression of 21,025 genes and the allele-specific expression of over 11,000 genes were quantified in interspecific hybrids and their parental species, *Coffea canephora* and *Coffea eugenioides* using RNA-seq technology. Between parental species, *cis*- and *trans*-regulatory divergences affected around 32% and 35% of analyzed genes, respectively, with nearly 17% of them showing both. The relative importance of *trans*-regulatory divergences between both species could be related to their low genetic divergence and perennial habit. In hybrids, among divergently expressed genes between parental species and hybrids, 77% was expressed like one parent (expression level dominance), including 65% like *C. eugenioides*. Gene expression was shown to result from the expression of both alleles affected by intertwined parental *trans*-regulatory factors. A strong impact of *C. eugenioides* *trans*-regulatory factors on the upregulation of *C. canephora* alleles was revealed. The gene expression patterns appeared determined by complex combinations of *cis*- and *trans*-regulatory divergences. In particular, the observed biased expression level dominance seemed to be derived from the asymmetric effects of *trans*-regulatory parental factors on regulation of alleles. More generally, this study illustrates the effects of divergent *trans*-regulatory parental factors on the gene expression pattern in hybrids. The characteristics of the transcriptional response to hybridization appear to be determined by the compatibility of gene regulatory networks and therefore depend on genetic divergences between the parental species and their evolutionary history.

Key words: hybridization, *cis*- and *trans*-regulation, allele-specific expression, allopolyploidy.

Introduction

Hybridization and allopolyploidization, both prominent events in plant evolution, generate novel hybrid phenotypes (Hegarty and Hiscock 2009). Changes in gene expression have been observed in hybrids that reconcile divergent genomes and regulatory networks in the same cellular context (Landry et al. 2007). Elucidating the regulatory changes at the origin of the variability of gene expression in hybrids is a key to understanding gene-expression novelty and its role in adaptive evolution (Wolf et al. 2010; Buggs et al. 2014). In recent years, several genome-wide studies have investigated gene regulatory divergences and/or inheritance of gene expression

in hybrids and allopolyploids (Chagué et al. 2010; McManus et al. 2010; He et al. 2012; Qi et al. 2012; Shi et al. 2012; Bell et al. 2013; Yoo et al. 2013; Cox et al. 2014). In these studies, *cis* and *trans* effects have been identified in gene expression differences between parental alleles or homoeologs (McManus et al. 2010; He et al. 2012; Shi et al. 2012; Bell et al. 2013; Xu et al. 2014). In addition, nonadditive gene expression variations in hybrids or allopolyploids relative to parental expression level have been shown. In particular, dominance expression was observed for numerous genes, the expression level in the hybrid or in the allopolyploid being similar to that exhibited by one of the parent, up or down regulated

© The Author(s) 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

relative to the other parent. Furthermore, for most of these studies, the genome-wide expression level dominance was skewed toward one parent (termed as biased expression level dominance; Grover et al. 2012). This phenomenon appears to be common and was reported in various organisms (flies, yeast, plants) (McManus et al. 2010; Qi et al. 2012; Bell et al. 2013; Yoo et al. 2013; Cox et al. 2014). Although overall patterns in gene expression changes in hybrids were described, the relationships between inheritance of gene expression and gene regulatory divergences have not been completely explored and the genetic bases of the variations in allelic expression are poorly understood. To understand the rewiring of gene expression in hybrids, we investigated interspecific hybrids between two *Coffea* species, *Coffea canephora* (C) and *Coffea eugenioides* (E). The two species are perennial woody trees and display considerable variation in morphology, size, and ecological adaptation. Nevertheless, they are closely related, display low sequence divergence (i.e., 1.3% average difference for genes) with very high gene synteny (Cenci et al. 2012), and are characterized by similar genome sizes, 710 Mbp for C and 660 Mbp for E (Noirot et al. 2003). In addition, they hybridize readily (Louarn 1976), and are parental species of the cultivated natural allotetraploid *Coffea arabica* (Lashermes et al. 1999). To understand the relationships between the inheritance of gene expression and allele-specific expression (ASE), using RNA-seq technology, we analyzed inheritance of gene expression, *cis*- and *trans*-regulatory divergences, and variations in ASE in the hybrids.

Materials and Methods

Plant Materials

Interspecific hybrids were generated between C, widespread species in West and Central Africa (0–1,500 m asl.), and E, endemic species in East Africa at much higher altitudes (1,000–2,200 m asl.). The hybrids were produced by crossing C, used as the female parent, with E (C × E) and reciprocally using E as the female parent (E × C) following the standard hybridization technique (Couturon et al. 1998). All the plants belong to the core collection of the Colombian National Center of Coffee Research (CENICAFE: www.cenicafe.org). For this study, we selected 17 plants including three plants of *C. canephora* species, four plants of *C. eugenioides* species, six plants of the F1 interspecific hybrids C × E, and four plants of the reciprocal F1 crosses E × C. All plant materials were grown in common garden growing conditions at Naranjal experimental station (Chinchina, Colombia). Special care was taken to sample, young and just fully expanded leaves on each plant and to frozen the material immediately in liquid nitrogen.

RNA Extraction and RNA Sequencing

Total RNA was extracted using the Qiagen RNeasy kit (Qiagen, Stanford, CA) according to the manufacturer's instructions.

The quality and concentration of extracted RNAs were determined using the Agilent DNA 1000 (Agilent, Santa Clara, CA). The messenger RNA (mRNA) libraries were constructed with the Illumina "RNA-seq sample prep" kit (Illumina, San Diego, CA) and sequenced separately on a HiSeq 2000 (Illumina) at the MGX platform (Montpellier Genomix, France). After quality filtering (Phred score higher than 28 and removal of reads less than 50 bp), a total of 375 million single-end reads (~72 nt) were retained (ranging from 12 to 31 million reads per library), including 80 million, 73 million, 135 million, and 87 million for the *C. canephora*, *C. eugenioides*, C × E hybrid, and E × C hybrid libraries, respectively.

RNA-seq Data Processing

The data concerning the parents and the reciprocal interspecific hybrids were treated using the same method. Owing to the low genetic divergence between parental species, the 72-nt reads of each library were mapped to a *C. canephora*-coding DNA sequence reference (25574 CDS) as transcriptome reference (Denoeud et al. 2014) using Burrows Wheeler Aligner MEM (BWA MEM) (Li and Durbin 2010) with the default parameters. The RSeqC v2.4 software package was used to check the quality of RNA-seq data from BAM files (.bam) (Wang et al. 2012). Gene coverage profiles that were slightly biased toward the 3' extremity of the genes were observed uniformly in all the samples used in the study. As previously (Combes et al. 2013) described, the gene expression estimates for the E parent were assumed to be slightly underestimated, due to a slight technical bias caused by the nature of the transcriptome reference. The aligned sequences of each library were then analyzed to find SNPs (single nucleotide polymorphism) with the GATK toolkit (<http://www.broadinstitute.org/gatk/>) using the Unified-Genotyper module with default parameters to obtain a list of SNPs and allelic data, and the depth-of-coverage module to obtain information on depth coverage. To avoid artifacts due to reads from pseudogenes or repeat sequences, only the CDS identified as single copy was used for subsequent analyses. A species-specific database of 82,916 SNPs, representing 14,206 CDS, was produced by retaining only divergent polymorphic nucleotide sites at which accessions of both parents were homozygous for differences.

Differential Expression Analysis

Mapped reads of each sample were analyzed with the "DESeq" package (Anders and Huber 2010) in R software version 3.0 (R Foundation for Statistical Computing, Vienna, Austria). To remove the negative effect of background expression noise on differential expression estimation, we restricted analysis to genes with minimum cumulated read counts (depending on the number of replicates) of 30, 40, and 60 for the *C. canephora* samples, the *C. eugenioides*, and E × C and C × E hybrid samples, respectively, in at least one parental

and hybrid group. For genes whose allele was not detected in interspecific hybrids, we changed the number of mapped reads from 0 to 1 in order to calculate ratios and perform statistical analyses (Marioni et al. 2008). Data were first normalized on the total number of counts with the DESeq package (using “per condition” and “maximum” as arguments), taking the variance and the mean of the biological replicates into account. Next, pairwise comparisons were performed of the total level of gene expression of the parental species, and those of the hybrids with each of the parental species (negative binomial test). The distribution of *P* values was checked for a false discovery rate using the Benjamini and Hochberg method at $\alpha = 0.05$.

Inheritance Classifications

Expression inheritance was determined for differentially expressed genes in hybrids. Log-transformed and DESeq-normalized expression values of each parent were subtracted from those of the hybrid to examine changes in expression. As described by McManus et al. (2010), genes whose total expression in hybrids deviated more than 1.25-fold from that of either parent were considered to have nonconserved inheritance. Based on the magnitude and the direction of the changes, the genes were classified as displaying additivity (with *E* expression lower or higher than *C* expression), *E* or *C* expression level dominance (lower or higher than either parent: down and up), and transgressivity (lower or higher than both parent: down and up).

ASE and Assignment of *cis*- and *trans*-Regulatory Divergence

Using the species-specific SNP database, the reads of hybrids with SNPs were sorted into *C* or *E* allele-specific bins using Custom Perl Scripts. Genes with more than 5% of reads exhibiting ambiguous assignment (not consistent with the species-specific SNP database) due to the potential presence of paralogs were discarded (around 1.5% of genes). To examine the possibility of biased ASE, genes with low coverage (cumulative numbers of *C*-specific reads and *E*-specific reads was ≤ 80) were filtered out. ASE of replicates of each hybrid and parental expression of replicates of parents were then treated and normalized using the DESeq package. The count data were normalized on the total number of counts taking the variance and the mean of the biological replicates was taken into account. Relative allelic expression (RAE) corresponding to allele-specific read counts (C_{hyb} or E_{hyb}) among the total read counts ($C_{\text{hyb}} + E_{\text{hyb}}$) was determined and expressed as the percentage of the C_{hyb} allele ($\%C_{\text{hyb}}$) in the total gene expression of the hybrid.

At the transcription level, gene expression is governed by interactions between *cis*- and *trans*-regulatory factors. In hybrids, both parental alleles are in the same cellular context and exposed to a common set of *trans*-regulatory factors; allelic

expression can therefore be altered due to *cis*- and/or *trans*-regulatory divergence between parental species. ASE in the hybrids compared with allelic expression in their parents makes it possible to distinguish between *cis*- and *trans*-regulatory divergences. Several tests were performed (Wittkopp et al. 2004; Shi et al. 2012), first to compare the ratio of expression of the two parental alleles in the hybrids with the relative expression of the same alleles in their parents (two-sided prop.test in R, $H_0: C/E = C_{\text{hyb}}/E_{\text{hyb}}$, *P* value adjusted for multiple testing; Benjamini–Hochberg method) and then to compare the RAE with the balance parental expression (two-sided prop.test in R, $H_0: C_{\text{hyb}}/E_{\text{hyb}} = 1$, *P* value adjusted for multiple testing; Benjamini–Hochberg method). Based on these tests, genes were categorized as conserved, or as belonging to *cis*-regulatory, *trans*-regulatory, or *cis*- + *trans*-regulatory divergence categories. A similar and balanced allelic expression between parents and hybrids indicated conserved regulation, whereas a conserved unbalanced allelic expression between parents and hybrids was the signature of parental *cis*-regulatory divergences, and a balanced allelic expression only in hybrids revealed parental *trans*-regulatory divergences. For the remaining genes, parental expression divergences were the result of a combination of *cis* and *trans* effects. Compensating and enhancing *cis* and *trans* interactions were identified by comparing the directions of *cis* and *trans* effects. A compensating *cis* + *trans* interaction was inferred if the *cis* and *trans* effects of one gene were in the opposite direction ($C_{\text{hyb}}/E_{\text{hyb}} > 1$ and $C/E < C_{\text{hyb}}/E_{\text{hyb}}$ or $C_{\text{hyb}}/E_{\text{hyb}} < 1$ and $C/E > C_{\text{hyb}}/E_{\text{hyb}}$). If the *cis* and *trans* effects of one gene acted in the same direction ($C_{\text{hyb}}/E_{\text{hyb}} > 1$ and $C/E > C_{\text{hyb}}/E_{\text{hyb}}$ or $C_{\text{hyb}}/E_{\text{hyb}} < 1$ and $C/E < C_{\text{hyb}}/E_{\text{hyb}}$), then the interaction was categorized as enhancing *cis* + *trans* interaction.

Up- and Downregulation of Hybrid Alleles

For genes whose ASE was available, the expression regulation of both alleles was evaluated. For each allele, the variation in expression in the parents and in the hybrid was plotted on a graph (C_{hyb}/C on the *x* axis and E_{hyb}/E on the *y* axis). Because for each allele, the genome-specific *cis*-regulator factors are similar between the parental and hybrid cellular contexts, this graph provides information about changes in allelic expression due to *trans*-regulatory factors of the other genome. The position of each gene depends on the up- and downregulation of both alleles and on the origin of *trans*-regulatory effects. Although negative values of the $\log_2$ transformed allele-specific ratio (C_{hyb}/C or E_{hyb}/E) indicate downregulation of alleles in hybrids, positive values indicate upregulation.

Analysis of Functional Enrichment of Gene Ontology (GO)

Computational mapping and plant GO-slim annotation were performed using BLAST2GO PRO software (Conesa et al. 2005; Conesa and Gotz 2008) with default parameters for

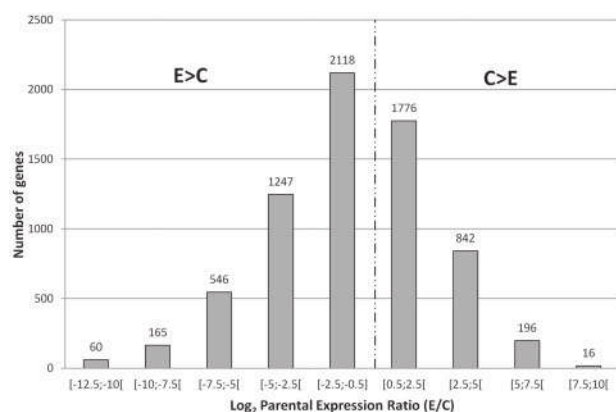


FIG. 1.—Differences in gene expression between *Coffea canephora* and *Coffea eugenioides*. The histogram shows the direction and magnitude of changes in expression in genes exhibiting divergent parental expression. Negative values indicate higher expression in *C. eugenioides* ($E > C$), and positive values indicate higher expression in *C. canephora* ($C > E$). A total of 33% of genes were differentially expressed between parental species (6,966 genes among the 21,052 genes analyzed), regardless of the magnitude of the change (DESeq analysis, see [supplementary Methods, Supplementary Material](#) online). Of these genes, significantly more (59%, $n = 4,136$) were highly expressed in *C. eugenioides* than in *C. canephora* (41%, $n = 2,830$) ($E/C > 1$), (binomial test, $P < 0.001$).

25,574 CDS of the *C. canephora* reference. For the genes selected at different steps of the analysis, molecular and cellular functions were searched for using an enrichment test (False Discovery Rate (FDR) < 0.05) to evaluate statistical significance.

Results and Discussion

Differential Gene Expression among Parental Species

Before studying gene expression patterns in hybrids, we first characterized pre-existing differential gene expression between the two parental species; 33% of the 21,052 genes we examined were significantly differentially expressed between parental species. Although in E, the gene expression was slightly underestimated due to the use of the C transcriptome as mapping reference (Combes et al. 2013), 59% of differentially expressed genes were more highly expressed in E than in C (fig. 1). In addition, after setting a threshold ($\log_2$ fold change ± 1.25) so as to consider only markedly differentially expressed genes, the percentage of differentially expressed genes between parents was reduced to 23%, but 60% of differentially expressed genes was still more highly expressed in E than in C. Thus, the estimate of the level of expression divergence between low genetic divergent *Coffea* species is slightly higher than those between intraspecific *Cirsium* parents (Bell et al. 2013) and considerably lower than those between divergent *Drosophila*

interspecific parents (McManus et al. 2010). In the GO term enrichments of differentially expressed genes, the processes of “regulation” and “response to stimulus” were over represented, whereas the “metabolic process” and “binding” function were under represented ([supplementary table S1, Supplementary Material](#) online). GO term enrichments of the most expressed genes in each parental species were also identified, whereas enrichments of GO terms linked with binding function, metabolic process, “regulation of metabolic process” and “development” process ([supplementary table S2, Supplementary Material](#) online) were observed in C. GO terms associated with response to stimulus appeared over represented in E ([supplementary table S3, Supplementary Material](#) online). These data suggest that the expression profiles of the parental species are quite similar for basic cellular functions, but differ for genes involved in the regulation of metabolic processes and in the response to stimulus. However, more differences between both species may exist, indeed subtle changes (less than $\log_2$ fold change ± 1.25) for some parental genes that could considerably affect the phenotype were not considered in the analysis. Nevertheless, the observed transcriptomic divergences for genes involved in the regulation of metabolic processes and in response to stimulus are not surprising for *Coffea* species which result from a recent and rapid speciation (Cros et al. 1998; Anthony et al. 2010). Under different environmental pressure conditions, each species could have rapidly evolved by regulating gene expression leading to morphological variations and ecological adaptations as observed for species resulting from rapid adaptive radiation (Barrier et al. 2001; Kapralov et al. 2013). The role of gene expression variation in rapid evolution has already been suggested for *Senecio aethnensis* and *Senecio chrysanthemifolius*, two closely related species that recently diverged and are adapted to different altitude environments. In these species, regulatory differences in genes involved in environmental adaptation appear to determine their phenotypic differences (Chapman et al. 2013; Muir et al. 2013). In addition, variations in gene expression inducing changes in cellular function with relatively little change in genomic sequence are part of evolutionary processes (Whitehead and Crawford 2006; Ames and Lovell 2011).

Global Pattern of Gene Expression in Hybrids Biased Toward *Coffea eugenioides*

We characterized changes in gene expression in ten hybrids resulting from distinct hybridization events between accessions of C and E (3 and 4, respectively), and subdivided them into two groups according to the crossing direction ($C \times E$ and $E \times C$). To assess the reproducibility of the results, we examined the two groups separately throughout the study. In all comparisons of the total level of gene expression of the hybrids with those of each parental species, we

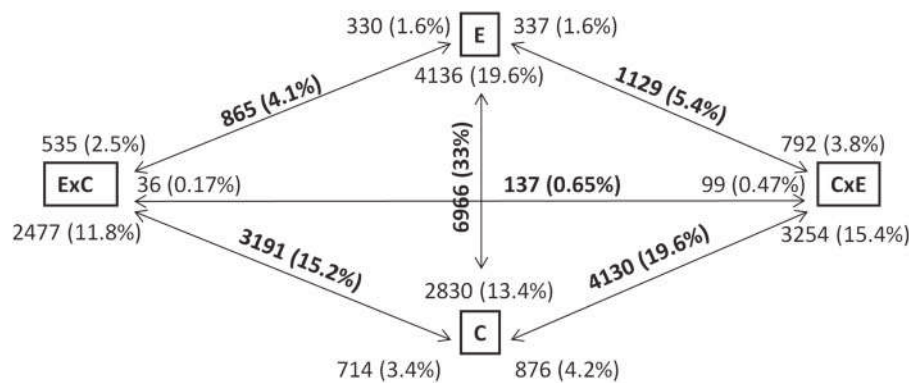


FIG. 2.—Differentially expressed genes in each possible comparison between parents and hybrids (DESeq analysis, see [supplementary Methods, Supplementary Material](#) online). Numbers in bold indicate the total number and fraction of genes differentially expressed in each comparison among the 21,052 genes analyzed. For each comparison, the number and fraction of upregulated genes in each group of plants are indicated at the ends of the arrows. In pairwise comparisons between gene expression in the hybrids and their parental species, the percentages of genes showing differential expression between hybrids and *Coffea eugenioides* (4.1% and 5.4 % for the comparison E × C versus E and C × E versus E, respectively) were lower than those showing differential expression between hybrids and *Coffea canephora* (15.2% and 19.6 % for the comparison E × C versus E and C × E versus E, respectively, $P < 0.001$, Fisher's exact tests). In each of the four comparisons of hybrids with their parents, the percentage of upregulated genes was higher in the hybrids than in the parents (3.8% and 15.4% of upregulated genes in C × E compared with 1.6% and 4.2% in *C. eugenioides* and *C. canephora*, respectively, $P < 0.001$, Fisher's exact tests).

observed fewer differentially expressed genes relative to the E parent than relative to the C parent. Moreover, for each pairwise comparison between hybrids and one of the parents, most differentially expressed genes were more highly expressed in the hybrids than in the parental species, providing evidence for changes in the regulation of gene expression (mainly upregulation) after the genome merger that resulted from hybridization (fig. 2).

Inheritance of Gene Expression in Hybrids and Expression Level Dominance of *Coffea eugenioides*

Changes in gene expression in hybrids were also evidenced by classifying gene expression inheritance patterns. Indeed, in both groups of hybrids, among divergently expressed genes between parental species and hybrids (around 27%), we observed all the different forms of altered gene expression (additivity, up and down dominance, and up and down transgressivity) (fig. 3). Although around 13% and 10% of genes were binned in additivity and transgressivity categories, respectively, the majority of categories contained expression level dominance (the remaining 77%) and the E expression level dominance “up” accounted for almost half the genes. Although we applied different methods of gene classification, the distribution of genes in categories was close to that previously reported for hybrids and allopolyploids (McManus et al. 2010; Qi et al. 2012; Bell et al. 2013; Yoo et al. 2013). Furthermore, the expression level dominance where the total expression of homeologs for a given gene in an allopolyploid is statistically equivalent to the expression level of that

gene in only one of the parents has been described many times before (Grover et al. 2012), but the molecular mechanisms of this phenomenon are still poorly understood.

Contribution of *cis* and *trans* Effects to Gene Expression Divergence

For genes with species-specific SNPs, RAE was determined and expressed as the percentage of the C_{hyb} allele ($\%C_{hyb}$) in the total gene expression of the hybrid ([supplementary fig. S1, Supplementary Material](#) online). In all the genes we analyzed, both alleles were always expressed in both groups of hybrids and 70% contributed equally to the transcriptome of hybrids, that is, the average percentage of the C_{hyb} allele was around 50%. For genes showing unbalanced RAE, the C_{hyb} allele was no more frequently expressed than the E_{hyb} allele. For *C. arabica*, the genome-wide analysis of the relative homoeologous gene expression has also revealed that neither of the two subgenomes was preferentially expressed (Combes et al. 2013).

In hybrids, gene expression differences between parental alleles can result from *cis*- and/or *trans*-regulatory changes. To elucidate the genetic basis of gene expression divergence, we compared the ASE in parental and hybrid samples to identify *cis*- and *trans*-regulatory divergences (fig. 4A). The RAE in hybrids, in which parental alleles and regulatory networks interact in the same cellular environment, provides readout of relative *cis*-regulatory divergences. Allelic expression divergences not explained by *cis*-regulatory divergences between the parental species are inferred to be caused by

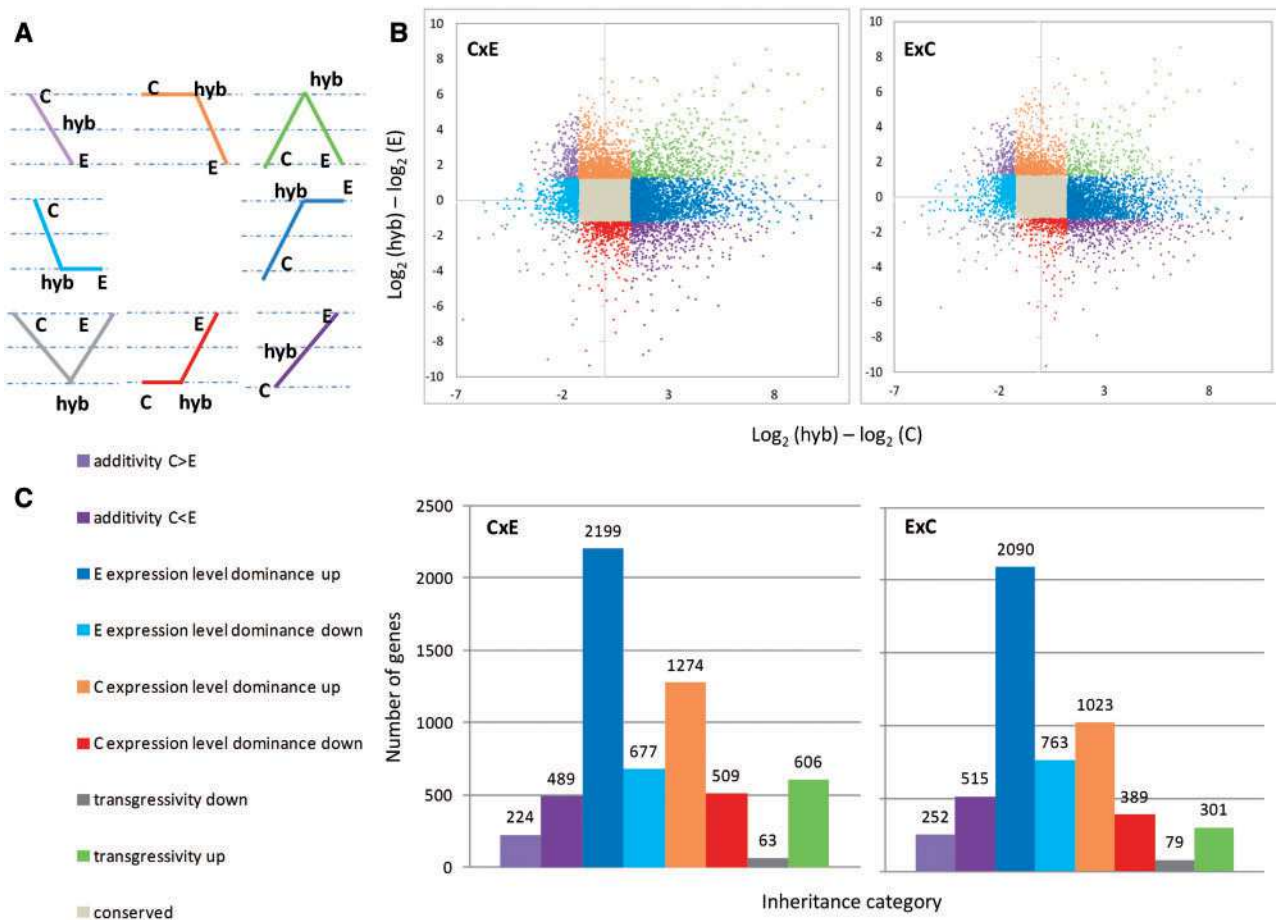


FIG. 3.—Inheritance of gene expression level in hybrids between *Coffea canephora* and *Coffea eugenioides*. (A) Hypothetical patterns of gene expression in *C. canephora* (C), *C. eugenioides* (E), and F₁ hybrids; C × E or E × C (hyb) illustrating additivity, transgressivity, and dominance are shown. (B) For the 21,052 genes in C × E and E × C hybrids examined, the scatterplots compare the level of gene expression of F₁ hybrids and those of the parental species (*C. canephora* on the x axis and *C. eugenioides* on the y axis). (C) The bar graphs show the frequency of genes in each expression inheritance category for C × E and E × C hybrids.

trans-regulatory divergences (Cowles et al. 2002; Wittkopp et al. 2004). In our experimental conditions, half the genes showed no expression divergence between parental species and no allelic expression changes in hybrids. In the remaining genes, divergence in allelic expression was classified in three categories with no clear majority among the three: on average, 15.5% in the *cis*-regulatory category alone, 18.5% in the *trans*-regulatory category alone, and 17.5% in the *cis* + *trans*-regulatory categories (fig. 4B), in the two groups of hybrids. The absolute magnitude of parental gene expression divergence in each regulatory divergence category was evaluated. *Trans* effects appeared to play a larger role than *cis* effects in parental gene expression divergence in both groups of hybrids (fig. 4C). The data indicate that divergent evolution between diploid species may have impacted the regulatory sequences of genes (*cis*-regulatory factors) and abundance, as well as the activity of upstream regulators throughout the genome (*trans*-regulatory factors). In addition, the data also provided

evidence for the compatibility of diverged parental regulatory networks and of intertwined regulation of alleles. Indeed for all genes showing *trans*-regulatory expression divergence (combined with *cis*-regulatory or not), ASE was controlled by *trans*-regulatory factors from both parental genomes that are able to cause up- or downregulation of one or both alleles. Intertwined regulation of parental species has been already reported in gene expression in *C. arabica* (Combes et al. 2013). In several studies considering intra- or interspecific hybrids of variable organisms (flies, yeast, plants), *cis*-regulatory divergences predominated and explained more of the expression differences between species than within species (Emerson and Li 2010; Shi et al. 2012). With the exception of the study of McManus et al. (2010), *trans*-regulatory divergences contributed preferentially to the expression differences within species (Emerson et al. 2010; Bell et al. 2013; Xu et al. 2014). Others observations, such as the relative proportion of genes with evidence of *cis*- and/or *trans*-regulatory divergences,

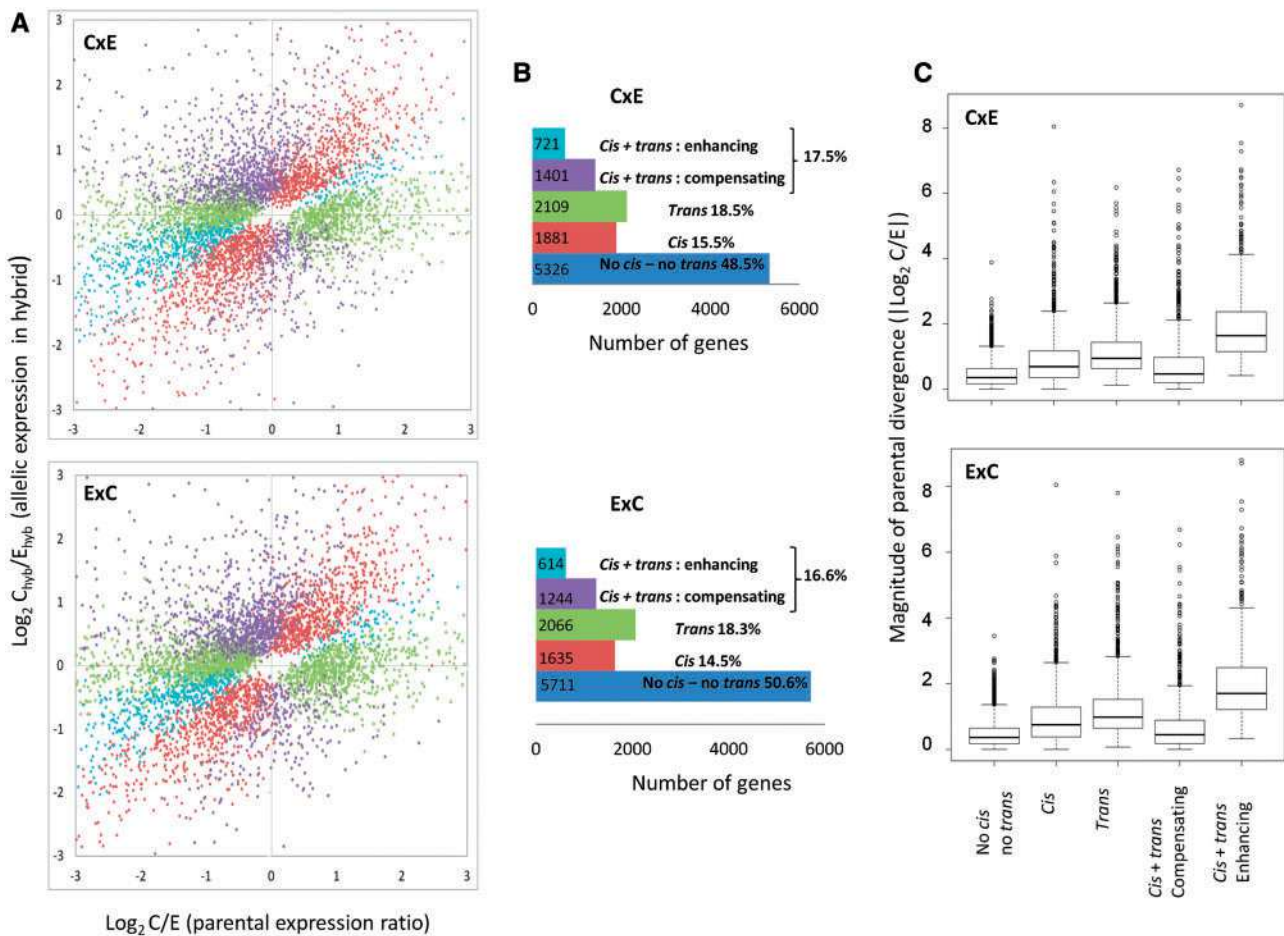


FIG. 4.—*cis*- and *trans*-regulatory divergence between parental species for 11,438 genes in $C \times E$ hybrids and 11,270 genes in $E \times C$ hybrids. (A) The plots summarize the relative allele-specific expression levels in parents and hybrids (C/E on the x axis; $C_{\text{hyb}}/E_{\text{hyb}}$ on the y axis; C , E and C_{hyb} , E_{hyb} corresponding to parental and allelic read counts, respectively). Each point represents a single gene on a logarithmic scale, and is color coded according to the regulatory divergence categories inferred from a hierarchical series of statistical tests (see Material and Methods). (B) The bar graphs depict the frequency of genes in each regulatory divergence category for $C \times E$ and $E \times C$ hybrids. The exact number of genes in each category is indicated at the extremity of each bar. The numbers in bold on the bar indicate the percentage of genes relative to the total number of genes analyzed. (C) The box plots summarize the absolute magnitude (fold-change) of parental expression divergence resulting from *cis*, *trans*, *cis* and *trans* (compensating or enhancing) interactions. In $C \times E$ and $E \times C$ hybrids, each category of regulatory divergence showed increased levels of expression divergence (Wilcoxon's rank-sum test, $P < 0.001$). The median significant *trans*-regulatory difference between species (1.15-fold for $C \times E$ and 1.22-fold for $E \times C$ hybrids) was larger than the median *cis*-regulatory difference between species (0.88-fold for $C \times E$ and 0.95-fold for $E \times C$ hybrids, Wilcoxon's rank-sum test, $P < 0.001$).

were more variable among studies. For Coolon et al. (2014) who studied expression differences within and between species of *Drosophila*, the proportion of genes with *cis*-regulatory divergences increases with divergence time but also depends on the mode of regulatory evolution of the species. In the study of *Drosophila melanogaster* and *Drosophila sechellia* hybrids, McManus et al. (2010) interpreted the higher proportion of *trans*-regulatory divergences than expected for an interspecific hybrid as an effect of particular demographic and ecological history of *D. sechellia* (McManus et al. 2010). In the case of *Coffea* species, gene expression differences result notably from *cis*-regulatory divergences but also from

a larger part than expected of *trans*-regulatory divergences associated or not with *cis*-regulatory divergences. This finding could be related to the recent speciation and the low genetic divergence between parental species. In addition, the perennial habit of *Coffea* species and the long generation time (i.e., 10–20 years) of coffee trees in natural conditions associated with a slow molecular evolution rate could be considered (Cenci et al. 2013).

The GO term enrichments of genes belonging to *cis*- or *trans*-regulatory divergences categories have revealed that GO terms related to “cellular metabolic process” were over represented among genes displaying *cis*-regulatory changes,

while those associated with “regulation of cellular metabolic process,” development, and response to stimulus were over represented among genes displaying *trans*-regulatory changes (supplementary table S4, Supplementary Material online). This result is consistent with the observations of Tirosh et al. (2009) in yeast. Indeed, these authors showed that divergent expression of genes responding to the environment may be preferentially caused by *trans*-regulatory changes. They also established that *trans* effects were attributable to differential interpretation of sensory signals and not to mutations in direct transcription regulators (Tirosh et al. 2009). Therefore, the observed *trans*-regulatory divergences between *Coffea* species could contribute to their different abilities to growth in contrasted environments. A study of gene expression of these same hybrids in variable controlled environments could be suggested to explore the phenomenon of *Coffea* species adaptation.

In this study, the analysis of the fraction of enhancing (*cis* and *trans* effects in the same direction) versus compensating interactions (*cis* and *trans* effects in opposite directions) in the category of *cis*- + *trans*-regulatory divergences makes possible to predict evolutionary models. Although stabilizing selection would be characterized by compensating *cis* and *trans* effects, diversifying selection would be revealed by enhancing *cis* and *trans* effects (Tirosh et al. 2009; Shi et al. 2012). For both groups of hybrids, the fractions of genes with compensating *cis*- + *trans*-regulatory effects among genes with *cis*- + *trans*-regulatory divergences were dominant (prop.test, $P < 0.001$). In addition, the absolute magnitude of parental gene expression divergence in this category was quite similar to the absolute magnitude of parental gene expression divergence in *cis*- and *trans*-regulatory categories. These data suggest a prevalent role for stabilizing selection that maintains gene expression levels.

A Clear Relationship between Regulation of Allelic Expression and Expression Inheritance in Hybrids

In hybrids, *trans*-regulatory factors from both parental origins could bind *cis*-regulatory regions of alleles differently and have different effects on expression between alleles. Analysis of the expression variation of alleles from both genomes between the parental and the hybrid contexts reveals the roles of *trans*-regulatory factors, and hence should elucidate the evolution of the parental species (Shi et al. 2012). To further analyze regulation of alleles in the differentially expressed genes in the hybrids compared with the parental species, we assessed variation in ASE in the two groups of hybrids. For each gene, we estimated variations in ASE in either the C genome (C-ASE) with the RAE C_{hyb}/C or the E genome (E-ASE) with the RAE E_{hyb}/E between the parental and hybrid contexts (fig. 5 and supplementary fig. S2, Supplementary Material online, for C $\times$ E and E $\times$ C hybrids, respectively). Given that the genome-specific *cis*-regulatory factors are similar in the

parents and the hybrids, *trans*-regulatory effects on the variation in ASE were revealed as well as their genome origin. The RAE of the C genome (C_{hyb}/C) was plotted against the RAE of the E genome (E_{hyb}/E) on a logarithmic scale. $\log_2$ positive values of the RAE $C_{hyb} > C$ or $E_{hyb} > E$ indicate upregulation of the alleles by the *trans*-regulatory factors of the homologous genome, while negative values of this ratio indicate downregulation of the alleles. The gene position relative to both axes reveals the single or combined genome origin of the *trans*-regulatory factors acting on that gene. Although alleles of both C and E genomes displayed both upregulated and downregulated allelic expression, the effects of *trans*-regulatory factors on both genomes appeared to be genome dependent. The E_{hyb} alleles were up- and downregulated by C *trans*-regulatory factors, whereas C_{hyb} alleles were mainly upregulated by E *trans*-regulatory factors. What is more, the E *trans*-regulatory factors had a greater impact on C-ASE than the C *trans*-regulator factors on E-ASE; the magnitude of variation in the expression of the C_{hyb} allele was higher than that of the E_{hyb} allele. These results suggest asymmetric effects of the activation of alleles of one genome by *trans*-regulatory factors of the other.

To refine this observation, the effects of C and E *trans*-regulatory factors on the C_{hyb} and E_{hyb} alleles in hybrids were analyzed. For genes in the *cis*- + *trans*-regulatory categories and *trans*-regulatory divergence categories, C_{hyb} and E_{hyb} alleles were classified according to the magnitude of the variation in allele expression between the hybrid and the parents (fig. 6). More E_{hyb} alleles than C_{hyb} alleles displayed only slight changes in ASE between the hybrids and their parental species, whereas more C_{hyb} alleles than E_{hyb} alleles showed marked ASE changes. These observations confirm that the E *trans*-regulatory factors have greater effects than C *trans*-regulatory factors on ASE variation between parental species and hybrids, in agreement with the observed overall E expression level dominance. As in the study of He et al. (2012) where asymmetry of ASE in *Arabidopsis* hybrids has been explained by differences in the efficiency of epigenetic gene silencing (He et al. 2012), one can also imagine epigenetic causes of the asymmetric effects of the *trans*-regulatory factors of one genome on the activation of alleles of the other. Indeed, the more effective E *trans*-regulatory factors observed in both the C $\times$ E and E $\times$ C hybrids for genes showing *trans*-regulatory divergences could indicate divergent epigenetic states between the two parental species.

To elucidate the regulatory mechanisms at the root of the gene expression inheritance, the up- and downregulation of the expression of alleles and the expression inheritance of genes were interpreted jointly (fig. 5 for C $\times$ E and supplementary fig. S2, Supplementary Material online, for E $\times$ C hybrids). For the majority of genes classified in the E expression level dominance “up” category, the C_{hyb} alleles were upregulated by E *trans*-regulatory factors, whereas for genes in the E expression level dominance “down” category, the C_{hyb} alleles

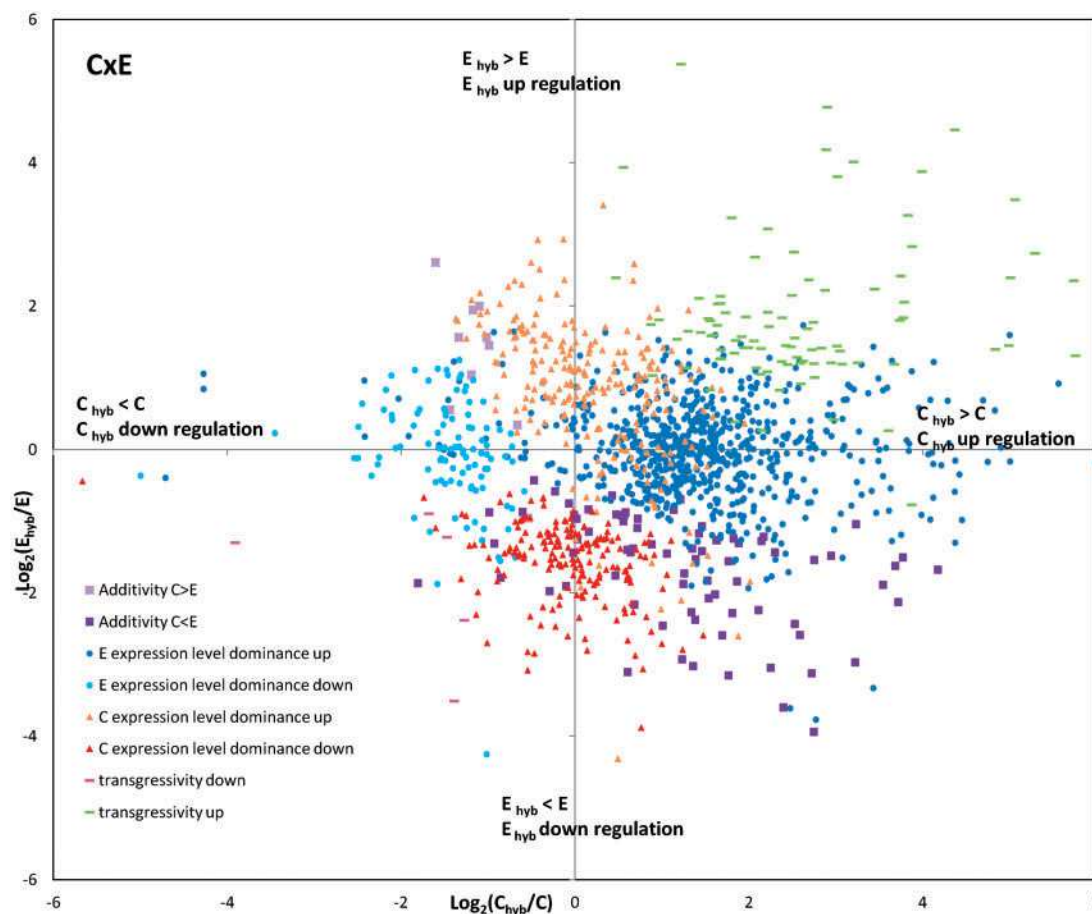


FIG. 5.—Variation in specific allelic expression (ASE) for genes differentially expressed between parents and the hybrid in the $C \times E$ hybrids. Each point represents a single gene, its position displays the combined up- and downregulation of both alleles (C_{hyb}/C on the x axis; E_{hyb}/E on the y axis, C_{hyb} , E_{hyb} and C , E corresponding to allelic and parental read counts, respectively), and its color code, the inheritance category of the gene. In this way, the combined regulation of alleles at the root of expression inheritance was elucidated for each gene analyzed (chi-square test, $P < 0.001$).

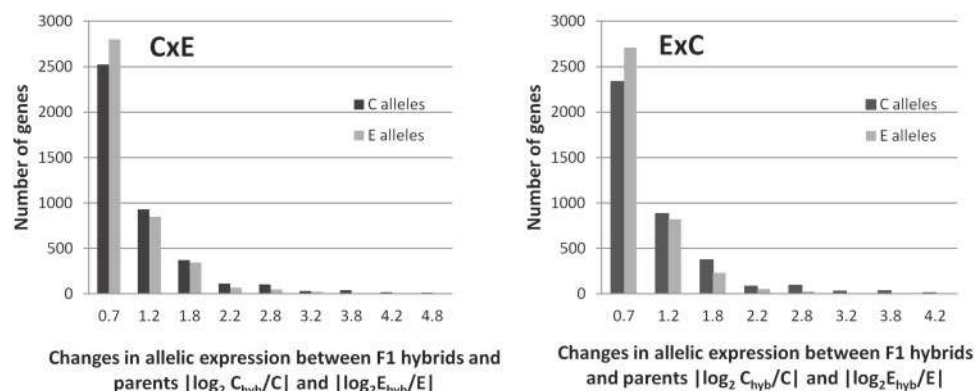


FIG. 6.—Effects of *trans*-regulatory factors of the E and C genome on up- and downregulation of C_{hyb} and E_{hyb} alleles in $C \times E$ and $E \times C$ hybrids. In line with the method of Shi *et al.* (2012), the genes subjected to either only *trans*- or only *cis* + *trans* effects were categorized according to differences in the absolute values of allelic expression between $C \times E$ and $E \times C$ hybrids and each parental species (on the x axis, dark gray for C_{hyb}/C and light gray for E_{hyb}/E). Although more E_{hyb} alleles showed a low level of expression difference between the parents and the hybrid, more C_{hyb} alleles displayed a high level of expression difference (chi-squared test, $P < 0.001$).

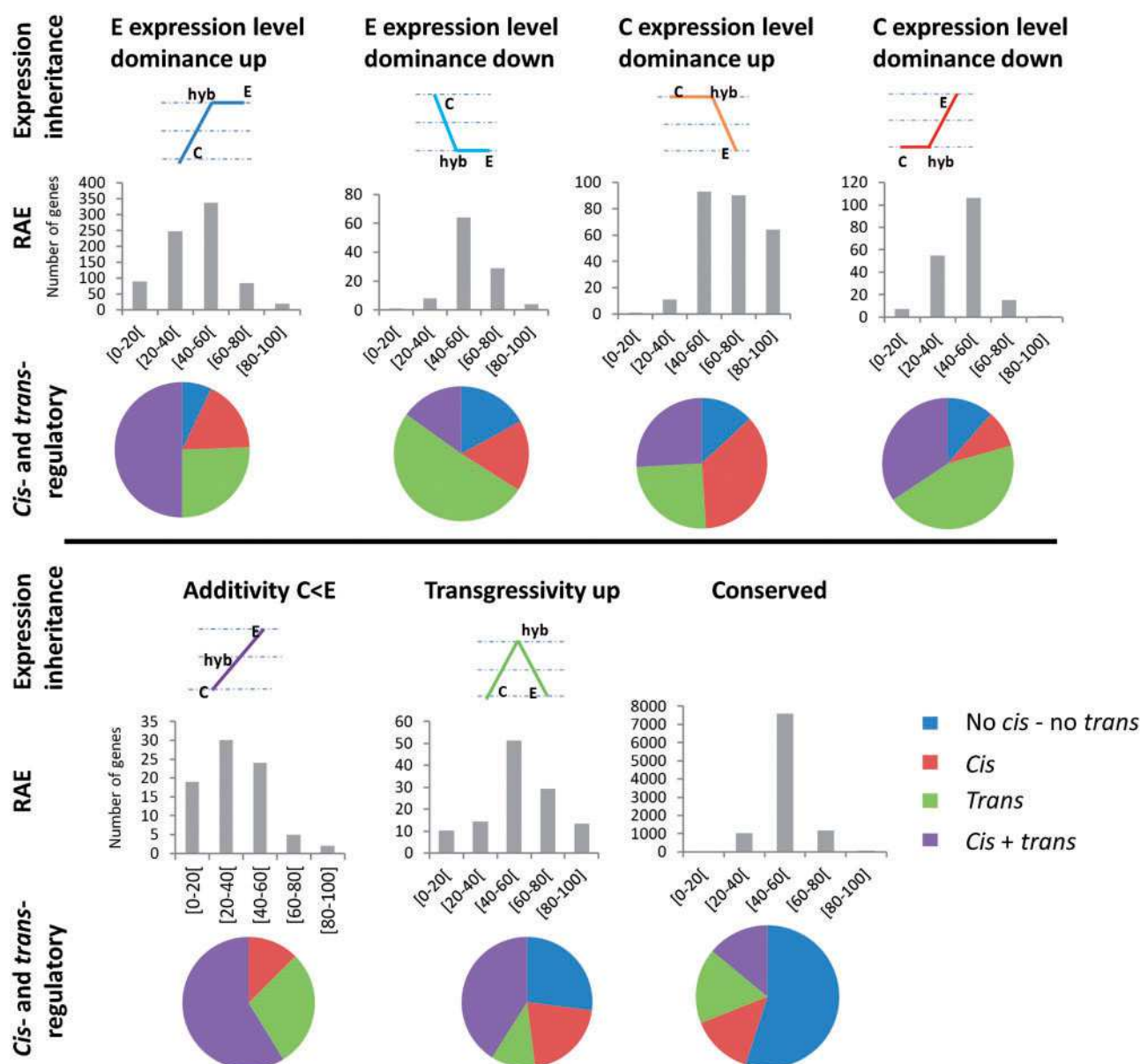


FIG. 7.—For the $C \times E$ hybrids, all the data collected on genes belonging to each category of expression inheritance are combined. The distributions of genes according to RAE and to *cis*- and *trans*-regulatory divergences are shown in the form of histograms and pie charts.

were downregulated. The same was observed for the C expression level dominance up and down categories, the E_{hyb} alleles being up- and downregulated by the C *trans*-regulatory factors. For the additivity and transgressivity categories, gene expression appeared to result from the combined effects of E and C *trans*-regulatory factors. Each category of expression inheritance of genes in hybrids depends on particular combinations of up- and downregulation of both alleles. In agreement with two previous studies that investigated both homoeologous and total gene expression in allopolyploids (Yoo et al. 2013; Cox et al. 2014), the present data confirmed

that the biased expression level dominance toward one parent was mainly caused by up- or downregulation of the allele of the other parent. In addition, in this study, the observed biased expression level dominance was attributed to asymmetric effects of *trans*-regulators factors of parental genomes. Similarly, between *Arabidopsis thaliana* and *Arabidopsis arenosa*, Shi et al. (2012) observed higher effects of *trans*-regulatory factors of *A. arenosa* on allelic expression in hybrids. Moreover, an overall expression dominance of *A. arenosa* genes over *A. thaliana* genes was previously reported in synthetic allopolyploids between *A. arenosa* and *A. thaliana*

(Wang et al. 2006). For other studies in various biological organisms (Bell et al. 2013; Yoo et al. 2013; Cox et al. 2014), the phenomenon of expression level dominance would be also attributed to the unbalanced effects of parental *trans*-regulatory factors. Accordingly, we hypothesize that in hybrids, gene expression inheritance could depend on the effects of *trans*-regulatory factors of one genome on the other genome, and hence the regulatory divergence between parental species. Thus, the divergent evolutionary histories of parental species could partly explain the different patterns of gene expression inheritance in hybrids reported in previous studies (McManus et al. 2010; Shi et al. 2012; Bell et al. 2013).

Particular RAE Distribution for Each Category of Expression Inheritance

Next, we refined the description of regulatory mechanisms of differentially expressed genes in interspecific hybrids relative to their parental species in the two groups of hybrids. For each category of expression inheritance, the distribution of genes according to RAE and regulatory divergence were compared with the distribution of genes in the conserved genes category, and significant differences were revealed (fig. 7 for $C \times E$ and supplementary fig. S3, Supplementary Material online, for $E \times C$ hybrids). Whatever the category of expression inheritance, the distributions of RAE were always unbalanced and biased toward the parental allele corresponding to the parent more expressed for the considered category. In the transgressivity up category, in which the two parental species were not differentially expressed, the RAE was more balanced than in other categories. With regard to the distributions of regulatory divergence mechanisms, the proportion of genes classified in each category varied between the categories of expression inheritance. The data showed that both *cis*- and *trans*-regulatory divergences were associated with variable allelic expression and consequently involved in variations of gene expression in hybrids. These findings reveal the intertwined allelic expression regulation and the complex relationship between the *cis*- and *trans*-regulatory factors that modulate gene expression. In addition, *trans*-regulatory divergences between parental species can have different origins: Changes in chromatin structures but also in transcription factors, changes in effectors, and RNA interference (RNAi) (Groszmann et al. 2011; Chodavarapu et al. 2012; He et al. 2012; Shen et al. 2012; Donoghue et al. 2014; Greaves et al. 2014).

Conclusions

Our genome-wide analysis enabled us to characterize both the genetic divergence between parental species and new gene expression patterns in hybrids. Gene expression changes between *C. canephora* and *C. eugenioides* resulted from both *cis*- and *trans*-divergences. The importance of *trans*-regulatory divergences observed between these two species could be related to their low genetic divergence and perennial habit.

In *Coffea* hybrids, the pattern of gene expression inheritance was characterized by a genome-wide expression level dominance biased toward the *C. eugenioides* parent. Moreover, gene expression was shown to result from the expression of both alleles and was affected by intertwined parental *trans*-regulatory factors indicating compatibility of parental regulatory networks. Hence, gene expression inheritance appeared to be determined by a complex mix of *cis*-regulatory factors and asymmetric effects of divergent *trans*-regulatory parental factors on up- and downregulation of alleles. In particular, the phenomenon of biased expression level dominance was attributed to unbalanced effects of parental *trans*-regulatory factors. More generally in hybrids, the extent and relative rates of *cis*- and *trans*-divergence among the parental species would have a significant effect on the gene expression pattern. In addition, the more similar the parental genomes, the more likely it is that the transcription factors of one genome might be compatible with transcription factor binding sites on the other genome. It appears that in hybrids, the characteristics of the transcriptional response to genome merger depend on genetic divergences between the parental species and their evolutionary history.

Supplementary Material

Supplementary figures S1–S3, tables S1–S4, and Methods are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgment

This work was partially supported by a grant from the National Centre for Coffee Research and the Agriculture Minister of Colombia (contract FNC 217/2011).

Literature Cited

- Ames RM, Lovell SC. 2011. Diversification at transcription factor binding sites within a species and the implications for environmental adaptation. *Mol Biol Evol.* 28:3331–3344.
- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.
- Anthony F, Diniz LEC, Combes M-C, Lashermes P. 2010. Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. *Plant Syst Evol.* 285:51–64.
- Barrier M, Robichaux RH, Purugganan MD. 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proc Natl Acad Sci U S A.* 98:10208–10213.
- Bell GD, Kane NC, Rieseberg LH, Adams KL. 2013. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol Evol.* 5:1309–1323.
- Buggs RJ, et al. 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos Trans R Soc Lond B Biol Sci.* 369. pii:20130354.
- Cenci A, Combes MC, Lashermes P. 2012. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant Mol Biol.* 78:135–145.

- Cenci A, Combes M-C, Lashermes P. 2013. Differences in evolution rates among eudicotyledon species observed by analysis of protein divergence. *J Hered*. 104:459–464.
- Chagué V, et al. 2010. Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytol*. 187:1181–1194.
- Chapman MA, Hiscok SJ, Filatov DA. 2013. Genomic divergence during speciation driven by adaptation to altitude. *Mol Biol Evol*. 30:2553–2567.
- Chodavarapu RK, et al. 2012. Transcriptome and methylome interactions in rice hybrids. *Proc Natl Acad Sci U S A*. 109:12040–12045.
- Combes M-C, Dereeper A, Severac D, Bertrand B, Lashermes P. 2013. Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol*. 200:251–260.
- Conesa A, Gotz S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. 2008:619832.
- Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.
- Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res*. 24:797–808.
- Couturon E, Lashermes P, Charrier A. 1998. First intergeneric hybrids (*Psilanthus ebracteolatus* Hiern *Coffea arabica* L.) in coffee trees. *Can J Bot*. 76:542–546.
- Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. 2002. Detection of regulatory variation in mouse genes. *Nat Genet*. 32:432–437.
- Cox MP, et al. 2014. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. *PLoS Genet*. 10:e1004180.
- Cros J, et al. 1998. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. *Mol Phylogenet Evol*. 9:109–117.
- Denoeud F, et al. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345:1181–1184.
- Donoghue MTA, et al. 2014. C(m)CGG methylation-independent parent-of-origin effects on genome-wide transcript levels in isogenic reciprocal F1 triploid plants. *DNA Res*. 21:141–151.
- Emerson JJ, Li WH. 2010. The genetic basis of evolutionary change in gene expression levels. *Philos Trans R Soc Lond B Biol Sci*. 365:2581–2590.
- Emerson JJ, et al. 2010. Natural selection on *cis* and *trans* regulation in yeasts. *Genome Res*. 20:826–836.
- Greaves IK, Groszmann M, Wang A, Peacock WJ, Dennis ES. 2014. Inheritance of trans chromosomal methylation patterns from *Arabidopsis* F1 hybrids. *Proc Natl Acad Sci U S A*. 111:2017–2022.
- Groszmann M, et al. 2011. Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci U S A*. 108:2617–2622.
- Grover CE, et al. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol*. 196:966–971.
- He F, et al. 2012. Genome-wide analysis of *cis*-regulatory divergence between species in the *Arabidopsis* genus. *Mol Biol Evol*. 29:3385–3395.
- Hegarty MJ, Hiscok SJ. 2009. The complex nature of allopolyploid plant genomes. *Heredity* 103:100–101.
- Kapralov MV, Votintseva AA, Filatov DA. 2013. Molecular adaptation during a rapid adaptive radiation. *Mol Biol Evol*. 30:1051–1059.
- Landry CR, Hartl DL, Ranz JM. 2007. Genome clashes in hybrids: insights from gene expression. *Heredity* 99:483–493.
- Lashermes P, et al. 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol Gen Genet*. 261:259–266.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
- Louarn J. 1976. Hybrides interspécifiques entre *Coffea canephora* pierre et *C. eugenioides* Moore. *Café Cacao Thé*. 20:433–452.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. 18:1509–1517.
- McManus CJ, et al. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res*. 20:816–825.
- Muir G, Osborne OG, Sarasa J, Hiscok SJ, Filatov DA. 2013. Recent ecological selection on regulatory divergence is shaping clinal variation in senecio on Mount Etna. *Evol Int J Org Evol*. 67:3032–3042.
- Noirot M, et al. 2003. Genome size variations in diploid African *Coffea* species. *Ann Bot*. 92:709–714.
- Qi B, et al. 2012. Global transgenerational gene expression dynamics in two newly synthesized allohexaploid wheat (*Triticum aestivum*) lines. *BMC Biol*. 10:3.
- Shen H, et al. 2012. Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell* 24:875–892.
- Shi X, et al. 2012. *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat Commun*. 3:950.
- Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324:659–662.
- Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28:2184–2185.
- Wang J, et al. 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172:507–517.
- Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol*. 15:1197–1211.
- Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430:85–88.
- Wolf JBW, Lindell J, Backström N. 2010. Speciation genetics: current status and evolving approaches. *Philos Trans R Soc B Biol Sci*. 365:1717–1733.
- Xu C, et al. 2014. Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Mol Biol Evol*. 31:1066–1076.
- Yoo M-J, Szadkowski E, Wendel JF. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110:171–180.

Associate editor: Patricia Wittkopp

CHAPITRE 3

L'allopolyploïdie et l'expression relative des gènes
homéologues

CHAPITRE 3 :

L'ALLOPOLYPLOÏDIE ET L'EXPRESSION RELATIVE DES GENES HOMEOLOGUES

Dans le cadre de l'étude de l'allopolyploïdie et de ses effets sur l'adaptation et l'évolution des plantes, l'expression relative des gènes n'avait été décrite, au moment de la réalisation des études présentées ci-dessous, que sur quelques modèles, coton (Adams et al. 2003, Adams et al. 2004, Dong et Adams 2011, Flagel et al. 2009) , blé (Mochida et al. 2003, Nomura et al. 2005, Stamati et al. 2009), *Arabidopsis suecica* (Chang et al. 2010), *Tragopogon miscellus* (Richard J A Buggs et al. 2010). Ces études rapportent des variations de la contribution des gènes homéologues en fonction des gènes, des organes examinés et du stress appliqué aux plantes (cf chapitre 2, paragraphe 2.3.2). Varier et augmenter le nombre de modèles en observant de nouvelles espèces, d'autres organes ou d'autres conditions environnementales est nécessaire pour distinguer les tendances de la variation de l'expression des gènes homéologues et approfondir les connaissances au sujet du dynamisme du transcriptome qui serait à l'origine de la diversité et des changements phénotypiques des allopolyploïdes.

1 Analyse et expression relative de gènes homéologues chez *Coffea arabica*

Pour *C. arabica*, comme pour d'autres espèces ne disposant pas alors d'importantes ressources génomiques, l'expression des gènes homéologues a d'abord été étudiée pour quelques gènes situés sur une région génomique pour laquelle l'équipe disposait de données de séquençage de clones BAC. A partir des séquences de clones de chacun des sous-génomes, des SNP différenciant les gènes homéologues ont été repérés sur des exons et des amorces les encadrant ont été définies pour amplifier l'ADNc de *C. arabica*. A partir de chromatogrammes provenant du séquençage Sanger de ces

amplicons, la différence d'expression entre les gènes homéologues a été estimée en comparant les hauteurs des pics superposés correspondants aux deux bases du SNP. Cette méthode a permis d'analyser l'expression de 13 gènes pour 4 organes et deux conditions de culture des plants de *C. arabica*. Cette étude a montré que l'expression relative des gènes homéologues varie aussi pour *C. arabica* en fonction des organes et des conditions de culture. Néanmoins ces variations sont modérées, bien que significatives, elles ne révèlent pas de biais extrême de l'expression d'un homéologue par rapport à l'autre, ni de « silencing » (absence d'expression d'un homéologue) ou signes de sous-fonctionnalisation. L'expression des deux homéologues toujours détectée suggère l'intervention de mécanismes de *trans*-régulation provenant des deux sous-génomes sur la régulation de l'expression des gènes homéologues. Ce travail a donné lieu à une publication, **Homeologous gene expression in response to growing temperature in a recent allopolyploid (*Coffea Arabica* L.)**. Marie-Christine Combes, Alberto Cenci, Hélène Baraille, Benoît Bertrand, Philippe Lashermes. 2012. Journal of Heredity 103 : 36-46.

Homeologous Gene Expression in Response to Growing Temperature in a Recent Allopolyploid (*Coffea arabica* L.)

MARIE-CHRISTINE COMBES, ALBERTO CENCI, HÉLÈNE BARAILLE, BENOÎT BERTRAND, AND PHILIPPE LASHERMES

From the Institut de Recherche pour le Développement (IRD), UMR RPB (IRD, Centre de coopération Internationale en Recherche Agronomique pour le Développement [CIRAD] Université Montpellier II), BP 64501, 34394 Montpellier Cédex 5, France (Combes, Cenci, Baraille, and Lashermes); and the CIRAD, UMR RPB (IRD, CIRAD, Université Montpellier II), BP 64501, 34394 Montpellier Cédex 5, Montpellier, France (Bertrand).

Address correspondence to Marie-Christine Combes at the address above, or e-mail: marie-christine.combes@ird.fr.

Abstract

Allopolyploidy is considered as a major factor contributing to speciation, diversification, and plant ecological adaptation. In particular, the expression of duplicate genes (homeologs) can be altered leading to functional plasticity and to phenotypic novelty. This study investigated the influence of growing temperatures on homeologous gene expression in *Coffea arabica* L., a recent allopolyploid involving 2 closely related diploid parental species. The relative expression of homeologs of 13 genes all located in the same genomic region was analyzed using an SNP ratio quantification method based on dideoxy-terminated sequences of cDNA amplicons. The relative expression of homeologous genes varied depending on the gene, the organ, and the growing condition. Nevertheless, expression of both homeologs was always detected (i.e., no silencing). Although the growing conditions were suitable for one or other of the parental species, neither subgenome appeared preferentially expressed. Furthermore, relative homeologous expression showed moderate variations across organs and conditions and appeared uncorrelated between adjacent genes. These results indicate the absence of signs of subfunctionalization suggesting *C. arabica* has not undergone noticeable diploidization. Furthermore, these results suggest that the expression of homeologous genes in *C. arabica* is regulated by a shared *trans*-regulation mechanism acting similarly on the 2 subgenomes and that the observed biases in the relative homeolog expression may result from *cis* fine-scale factors.

Key words: adaptation, coffee, duplicate gene expression, introgression, polyploidy, transcription

Polyploidy, a feature of organisms that have multiple sets of chromosomes, occurs in eukaryotes and is particularly widespread in flowering plants (angiosperms) including many major crops (wheat, maize, sugarcane, potato, soybean, coffee, and cotton) (Leitch AR and Leitch IJ 2008). Whereas an autopolyploid results from doubling a diploid genome, an allopolyploid is formed by the combination of 2 or more sets of distinct genomes (homeologous chromosomes) after interspecific hybridization and chromosome doubling (Chen 2007). In the last decade, in numerous studies on polyploids formed at different timescales (from a few years up to several million years ago and most recently on synthetic polyploids), dynamic and stochastic changes in genomic and epigenetic organization and in gene expression have been documented (Chen and Ni 2006; Chen 2007). Polyploidy is considered as a major factor contributing to speciation, diversification, and ecological adaptation (Osborn et al. 2003; Adams 2007; Peng et al. 2008; Jackson and Chen 2010). Genomic

plasticity (Leitch AR and Leitch IJ 2008) enables polyploids to combine multiple sets of genetic material in the same nucleus (Chen 2007; Hegarty and Hiscock 2009), whereas functional plasticity of duplicate genes enables them to regulate gene expression to adapt to their environment (Comai 2005; Jackson and Chen 2010; Buggs et al. 2011).

In several plant allopolyploid species such as *Gossypium hirsutum*, *Triticum aestivum*, *Tragopogon miscellus*, and *Arabidopsis suecica*, variations in gene expression have been investigated by either comparing the expression of the diploid parents and the allopolyploids or comparing the ratio of transcripts derived from each homeolog. Changes in homeologous expression have been explored at the genome scale in a single organ (Flagel et al. 2008; Hovav et al. 2008; Vidal et al. 2010) and in one to several dozen genes in different organs (Adams et al. 2003, 2004; Mochida et al. 2003; Nomura et al. 2005; Chaudhary et al. 2009; Flagel et al. 2009; Buggs et al. 2010). However, few studies have explored homeologous gene expression in several organs in relation to different types of

stress (Liu and Adams 2007; Stamati et al. 2009; Dong and Adams 2011). In the allopolyploid, locally or throughout the genome, homeologous genes may exhibit unequal expression patterns. The homeologous expression ratio can vary with the organ, its development, and in response to abiotic stresses. In some instance, complete silencing of one copy or reciprocal silencing (silencing of a homeologous gene in one organ and silencing of the other homeologous gene in another organ) is observed. Complex genetic and epigenetic regulatory mechanisms govern genome-specific expression biases in allopolyploid plant species. No single unifying factor has been reported suggesting that each allopolyploid plant has its own regulatory mechanism (Chaudhary et al. 2009). Along these lines, the homeologous expression ratio contributes to the fate of duplicated genes and to the evolution of allopolyploid plants after polyploidization. For instance, studies of gene expression patterns in different organs revealed adaptation mechanisms for evolution (neofunctionalization, subfunctionalization, and nonfunctionalization) (Buggs et al. 2010). All these findings on changes in homeologous gene expression attest to the capacity of polyploids for functional and adaptive diversification.

Coffee is one of the most important agricultural commodities and provides the main livelihood for more than 80 million people worldwide. Among the around 100 *Coffea* species identified (Davis et al. 2006), 2, *Coffea arabica* and *C. canephora*, are cultivated and account for 65% and 35% of world coffee production, respectively. *Coffea* is a relatively young genus (whose origin was estimated at 100 000–450 000 years before present; Anthony et al. 2010), and all *Coffea* species are diploid, except for *C. arabica*, which is allotetraploid ($2n = 4 \times = 44$) and derived from a recent (10 000–50 000 years ago) interspecific hybridization between 2 diploid species: *C. eugenioides* and *C. canephora* (Cenci et al. forthcoming; Lashermes et al. 1999). Homeologous genomes in *C. arabica* are designated E^a and C^a according to their parental origin. The 2 parental species exhibit different agroecological adaptations. *Coffea canephora* and *C. eugenioides* are endemic in regions where the annual mean temperature ranges from, respectively, 22 to 26 °C and 18 to 23 °C with no substantial oscillations (DaMatta and Cochicho Ramalho 2006; Davis et al. 2006). In contrast, *C. arabica*, whose optimum mean annual temperature range from 20 to 24 °C, can be grown in regions with marked variations in thermal amplitude. The 2 parental species are closely related, and the 2 subgenomes exhibit low sequence divergence (Cenci et al. forthcoming). Nevertheless, *C. arabica* displays diploid like meiotic behavior (Lashermes 2000). Preliminary results of a comparison of genome-wide expression patterns in *C. arabica* and its ancestral parents (Bardil et al. 2011) showed genomic expression dominance like that reported in nascent *Gossypium* allopolyploids (Rapp et al. 2009). Moreover, an expressed sequence tag (EST)-based study of differential homeologous gene expression in *C. arabica* (Vidal et al. 2010) showed both equal and biased homeologous genes expression.

To investigate the influence of growing temperatures on homeologous gene expression, we analyzed homeologous transcript ratio in different organs of *C. arabica* grown in 2

contrasted culture conditions known to be stressful for one or the other parental species. The relative expression of homeologs of 13 genes, all located in the same genomic region (Cenci et al. 2010; Lashermes et al. 2010), was determined using the SNP ratio quantification method based on dideoxy-terminated sequences of cDNA amplicons. Among previous studies on polyploids, this study completes the analysis of the variations in homeologous gene expression and informs about the behavior of duplicated genes in a recent allopolyploid. This study also provides new evidences regarding the *C. arabica*'s ability to adapt to environmental variations.

Materials and Methods

Plant Materials, RNA Extractions, and Reverse Transcription

Individuals of 2 *C. arabica* inbred varieties, Caturra and S795, were used. Caturra is a regular cultivar (genotype E^aE^aC^aC^a), whereas S795 was derived from an interspecific cross between *C. arabica* and *C. liberica* and carries several *C. liberica*-introgressed segments (Prakash et al. 2004). In particular, an introgression occurred on the subgenome C^a in a region named S_H3 (Lashermes et al. 2010), resulting locally in a genotype E^aE^aLL.

For the first study, seedlings of *C. arabica* var. Caturra were cultivated for 2 months in 2 sets of contrasted growing conditions with different day/night temperatures: 33 °C/30 °C (warmer conditions) and 23 °C/20 °C (colder conditions). The conditions were chosen to suit either *C. canephora* or *C. eugenioides*. In the 2 climatic chambers, the photoperiod, the hygrometry, and the luminosity were set at 12 h per day, 80–90%, and 600 μE/m²/s¹, respectively. In each climatic chamber, the plants were grown in a randomized complete block design. The cotyledons, young leaves, leaves, stems, and roots were harvested on 3 plants (biological replicates) for each culture condition.

For the second study, *C. arabica* var. Caturra and S795 were cultivated in a greenhouse. The temperatures were set at 26 ± 3 °C in the day and at 24 ± 3 °C at night. The hygrometry was maintained over 60%. Young leaves, leaves, stems, and roots were collected from 3 S795 plants and 2 Caturra plants of similar size.

In both studies, the collected samples were immediately flash frozen in liquid nitrogen and stored at –80 °C until RNA extraction.

RNA was extracted from all tissues using the Qiagen (Valencia, CA) RNeasy kit according to the manufacturer's instructions with a few modifications. The tissues (100 mg) were ground in liquid nitrogen. The tubes were centrifuged at 4 °C and kept in ice throughout extraction and DNase treatment. RNA was eluted by impregnating the purification column with 65 and 25 μl of sterile water for, respectively, 30 and 5 min before centrifugation. All RNA samples were quantified and visually assessed for degradation and DNA contamination using a Nanodrop (ThermoFisher Scientific Inc, Waltham, MA) and gel analysis.

Reverse Transcription was performed using 1 µg of total RNA with the ImProm-II Reverse Transcription System (Promega, Madison, WI). As controls for DNA contamination, reactions were performed without reverse transcriptase.

Design of Primers for Gene-Specific Amplification and Sequencing

Based on both *C. arabica* subgenome sequences from the S_{H3} region (Cenci et al. 2010, forthcoming, Lashermes et al. 2010), SNPs differentiating single-copy homeologous genes were identified. For 13 different genes spanning 350 kb of the S_{H3} region, primer pairs were designed on a single exon to amplify a fragment larger than 100 bp containing at least one SNP (Table 1).

PCR reactions were performed twice in a volume of 20 µl with either 1 µl of the diluted (one-fifth) cDNA generated by

the first-strand synthesis or 25 ng of genomic DNA, 0.4 µM of each primer, 2.5 mM $MgCl_2$, and 1 U of *Taq* DNA polymerase. Cycling was done in a GeneAmp PCR 9700 thermocycler for 2 min at 94 °C followed by 5 cycles of 10 s at 94 °C and 30 s at 60–55 °C (–1 °C per cycle) and by 30 cycles of 10 s at 94 °C, 30 s at 55 °C, 30 s at 72 °C, and a final 8-min extension at 72 °C. Sequencing was performed by Beckman Coulter Genomics (Takeley, UK).

Relative Expression of Homeologs

The relative expression of homeologs was estimated using a method similar to that of Rauscher et al. (2002) on sequence chromatograms of exon portions amplified on cDNA. Chromatogram files were analyzed using BIOEDIT (Hall 1999). The height of both peaks at an SNP position indicating the presence of the 2 homeoalleles was measured

Table 1 Homeologous genes investigated, SNP, and pairs of oligonucleotides used to study their relative contribution to the transcriptome of *Coffea arabica*

Genes	Putative function (BLASTP e value)	Primer sequences 5' > 3'	Size of the amplified product (bp)	Position of SNPs (differentiating bases E ^a and C ^a)
S_{H3} -44	Serine–threonine protein kinase, plant type ($E < 0.05$)	For: ggctacattgccctgaata Rev: caagcaagtgcgacagtcac	269	173 bp (A/G)
S_{H3} -48	Calcineurin B-like calcium sensor protein-interacting serine/threonine protein kinase (8×10^{135})	For: aaaacgggatgcttcatacg Rev: gtcggattaggatccagcaa	265	154 bp T/C
S_{H3} -51a	Beta-ketoacyl-CoA synthase ($E < 0.05$)	For: aaattgcatcttctggatgg Rev: aactaatggcccgagtgttg	247	63 bp G/A, 83 bp A/G
S_{H3} -55	Hypothetical protein (2×10^{123})	For: gacaacggttctctctctaaa Rev: ttgatacttcgcccgaatc	317	131 bp C/T
S_{H3} -56	Pyruvate kinase ($E < 0.05$)	For: gagcacacggaggactaagc Rev: tcctcattcaacctccgaac	171	101 bp A/G
S_{H3} -61	Chloroplast inner envelope protein (9×10^{156})	For: tcaaggccagtttctcaacc Rev: tagcaagttgatgcggtgac	274	84 bp T/C, 131 bp G/A
S_{H3} -66	Sulfate adenylyltransferase ($E < 0.05$)	For: gagaccactgtggtatctatatt Rev: gtaccttcttcccatgatctg	187	129 bp C/G
S_{H3} -67	Eukaryotic translation initiation factor 2c ($E < 0.05$)	For: agtgccagagctcaatg Rev: caccttgatcagatttccg	171	85 bp T/C, 129 bp T/C
S_{H3} -74	Putative early light-induced protein 2 (2×10^{151})	For: gccatgatagggtttgttgc Rev: aagggcacctctttgagat	270	119 bp C/T
S_{H3} -76	DNA-binding protein (2×10^{99})	For: gctaataagggttcttaaggg Rev: catcatctgctgtgtttat	296	134 bp C/T
S_{H3} -78	Acireductone dioxigenase (4×10^{91})	For: gatgaggtgtttgtgggaga Rev: gcactcccaccttctcagtg	110	75 bp G/A
S_{H3} -81	Chlorophyllase-2, chloroplast (5×10^{122})	For: cttatcaaagctggcgcta Rev: gcctgatccgatgactaatac	217	147 bp A/G
S_{H3} -83	Ice-binding protein (3×10^{176})	For: tctgcacagggaagga Rev: agctgttaaggtcatgatgtc	186	142 bp C/T, 147 bp T/C

For, forward; Rev, reverse.

using the coordinate system. Because the sequence process generated variations in peak size for a given position, the height of the different haplotypes was calibrated using chromatograms obtained from the amplification of *C. arabica* genomic DNA, where the E^a and C^a haplotypes are expected to be equally represented. For each SNP position, correction factor thus determined was applied to the peak heights obtained from different cDNA samples. The contribution of the E^a homeolog was expressed as the height of the E^a peak over the sum of the heights of the E^a and C^a peaks and conversely.

Two-way ANOVA and correlation analysis, performed with R software, were used to identify significant differences in homeologous gene expression as a function of the organ and growing condition concerned. The matrices of distances between genes based on pairwise correlations of relative homeologous gene expression and physical distances were compared by the Mantel test (Mantel 1967; Piepho 2005) using GENETIX software version 4.01 (Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France).

Method Validation

Two experiments were performed to validate the sequencing-based method used to determine the relative contribution of the homeologs. First, the relative proportion of SNP haplotypes was measured for all genes in amplicons from DNA samples showing different genome combinations: *C. arabica* var. Caturra (tetraploid, 1/2 E^a, 1/2 C^a), H855 triploid hybrid: *C. arabica* × *C. canephora* (1/3 E^a, 2/3 C^a + C₁), and Arabusta tetraploid hybrid: *C. arabica* × “tetraploid” *C. canephora* (1/4 E^a, 3/4 C^a + C₁). Second, DNAs from bacterial artificial chromosome (BAC) clones containing the S_H3 region derived from both E^a and C^a subgenomes were mixed with different proportions of the targeted region (0.84 E^a/0.16 C^a, 0.63 E^a/0.37 C^a, and 0.36 E^a/0.64 C^a). The relative proportion of the 2 SNP haplotypes in the amplicons was measured after amplification of the mixed solutions for 3 genes: S_H3-66, S_H3-76, and S_H3-81.

Results

Assessment of the Relative Contribution of Homeologs to the Transcriptome

In order to evaluate the reliability of the quantitative analysis of SNP peaks on sequence chromatograms, the expected and estimated proportion of SNP haplotypes in the amplicons were compared in 2 experiments. In the first experiment, the proportion of SNP haplotypes was analyzed in amplicons from genomic DNA extracted from genotypes with different genomic composition; the results obtained for the S_H3-66 gene are presented to illustrate the experiment (Figure 1A). In the second experiment, the proportion of SNP haplotypes for the S_H3-66, S_H3-76, and S_H3-81 genes was estimated in amplicons from mixed homeologous BAC DNA solutions (Figure 1B). In both experiments, the observed proportions of SNP were very close to expected values (in the first experiment, *r* ranged from 0.94 to 0.98 depending on the

genes considered with *P* < 0.001, in the second experiment for S_H3-66, *r* = 0.99 with *P* < 0.003, for S_H3-76, *r* = 0.99 with *P* < 0.002, for S_H3-81, *r* = 0.99 with *P* < 0.002), indicating

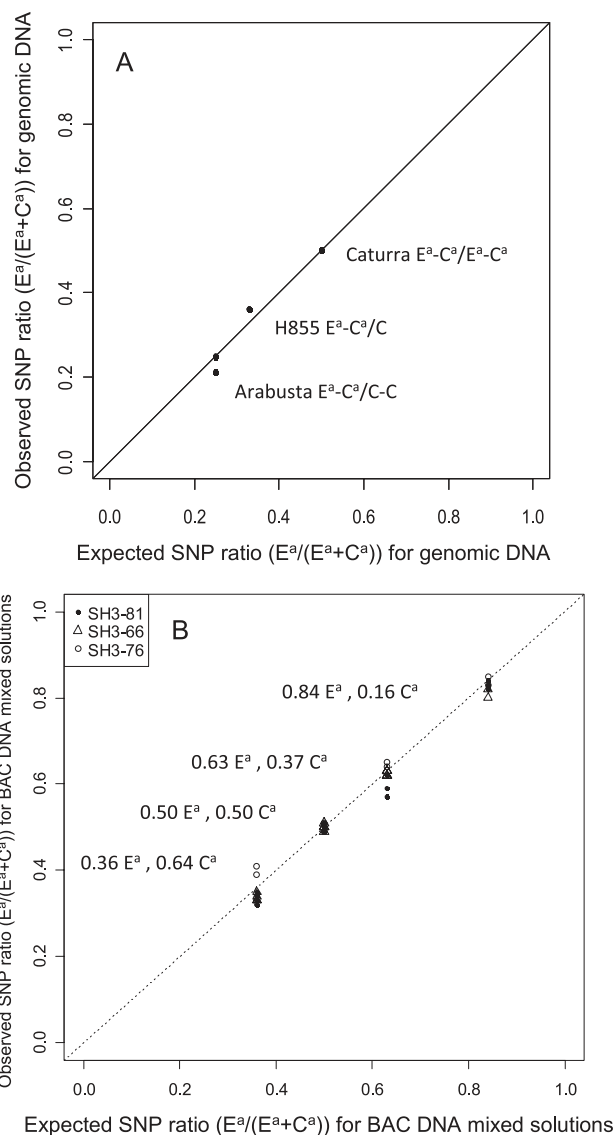


Figure 1. Validation of the sequencing-based method to determine the relative contribution of the homeologs using 2 methods. (A) Example (gene S_H3-66) for measure of proportion of SNP haplotypes in amplicons from 3 types of genomic DNA: *Coffea arabica* Caturra (tetraploid, 1/2 E^a, 1/2 C^a), H855 triploid hybrid *C. arabica* × *C. canephora* (2/3 C^a + C₁, 1/3 E^a), and Arabusta tetraploid hybrid from a cross between *C. arabica* and “tetraploid” *C. canephora* (3/4 C^a + C₁, 1/4 E^a). (B) The relative proportion of 2 types of the SNP in amplicons from different mixed BAC DNA solutions (0.84 E^a/0.16 C^a, 0.63 E^a/0.37 C^a, and 0.36 E^a/0.64 C^a) and *C. arabica* var. Caturra (tetraploid, 1/2 E^a, 1/2 C^a) for 3 genes: S_H3-66, S_H3-76, and S_H3-81. In the 2 figures, the values obtained were compared with the theoretical ratio curve.

that the method based on dideoxy-terminated sequencing accurately determined the ratios of homeologs in an amplicon.

For 3 genes (S_{H3-51a} , S_{H3-61} , and S_{H3-83}), which displayed several SNPs in the analyzed amplicons, the homeologous gene expression ratio for the overall set of samples (60 samples) was estimated on 2 SNPs. The relative contributions of homeologs to the transcriptome at both SNP positions were compared. The 2 measures were found to be significantly correlated (S_{H3-51a} : $r = 0.93$ with $P < 2.2 \times 10^{-16}$, S_{H3-61} : $r = 0.80$ with $P < 9.4 \times 10^{-11}$, and S_{H3-83} : $r = 0.87$, $P < 2.2 \times 10^{-16}$).

Finally, the variation coefficients of the estimation of the homeologous gene expression on replicates for all samples were calculated. The deviation between the technical replicates was less than 10% for 90% of the measures.

Homeologous Expression of Genes in the S_{H3} Region of *C. arabica*

The relative expression of the homeologs was assayed for the overall set of samples representing 5 organs and 2 growing conditions (Figure 2). The relative expression of the homeologs was represented by the E^a relative expression ratio [ratio $E^a/(E^a + C^a)$]. For all 13 genes analyzed in the S_{H3} region, both homeologous genes appeared to be expressed although their respective contribution varied. When all the 13 genes were considered, neither of the 2 subgenomes was preferentially expressed. Based on the overall relative expression of the homeologs for each gene, 4 categories of genes were distinguished. The first category contained genes S_{H3-51a} , S_{H3-66} , S_{H3-74} , and S_{H3-67} whose relative contribution to the transcriptome was biased toward the E^a genome. Inversely, the second category contained genes S_{H3-44} , S_{H3-48} , S_{H3-81} , and S_{H3-83} for whose relative contribution was biased toward the C^a genome. For genes S_{H3-55} , S_{H3-56} , and S_{H3-61} , the expression of the homeolog E^a was practically equal to the expression of the homeolog C^a . Finally, for genes S_{H3-76} and S_{H3-78} , the predominantly expressed homeolog differed with the organ and the culture condition concerned.

Effect of the Organ and Growing Conditions

The relative expression of homeologs varied with the organ analyzed. ANOVA (Table 2) indicated that 10 genes out of 13 had the statistically significant probability of the “organ” factor ($P < 0.001$). Depending on the gene, the variation in the homeologous gene expression depended either on one organ (S_{H3-66}) or on overall heterogeneity between organs (S_{H3-76} and S_{H3-78}). No organ-specific effect was observed.

According to the ANOVA, 4 and 6 genes out of 13 showed statistically significant differences between “growing conditions” at probabilities $P < 0.001$ and $P < 0.05$, respectively. For some genes (Figure 2), the E^a relative expression ratio was stronger or weaker in colder and warmer conditions depending on the organ concerned, whereas for other genes (S_{H3-55} and S_{H3-76}) exhibited a higher contribution by the homeolog E^a in colder conditions for all organs than in warmer conditions or conversely (S_{H3-44}).

Comparison of Neighboring Genes

To compare the behavior of genes spread over 350 kb of the S_{H3} region, expression profiles representing the homeologous gene expression for the 13 genes per organ were superimposed on the warmer and colder conditions (Figure 3). The correlation between these organ expression profiles was estimated. In warmer conditions, all correlation coefficients were statistically significant (5 coefficients, $P < 0.001$; 4 coefficients, $P < 0.01$; 1 coefficient, $P < 0.05$). In colder conditions, the situation was different, 6 correlation coefficients were statistically significant (1 coefficient, $P < 0.001$; 3 coefficients, $P < 0.01$; 2 coefficients, $P < 0.05$). The deviations between tissue expression profiles were greater in colder conditions (S_{H3-48} , S_{H3-74} , S_{H3-76} , S_{H3-78} , and S_{H3-81}) than in warmer conditions.

We performed statistical analyses on the S_{H3} region to determine if physical distance had an effect on pairwise correlations of relative homeologous gene expression across tissues and culture conditions. The significance of this factor was assessed using permutation tests with physical distance measured in base pairs. These distance data were subjected to 1000 random permutations, and physical distance was found to be nonsignificantly correlated with the relative homeologous gene expression for the S_{H3} region (Pearson's $r = 0.098$; $P = 0.21$).

Homeologous Gene Expression in an Introgressed Genotype

To analyze the impact of introgression on homeologous gene expression, the relative contribution of homeologs to the transcriptome was analyzed in genes S_{H3-81} and S_{H3-83} in 2 *C. arabica* inbred varieties carrying (var. S795) or not (var. Caturra) a *C. liberica*-derived introgression. The ratios of the relative homeologous gene expression $E^a/(E^a + C^a)$ for var. Caturra and $E^a/(E^a + L)$ for the introgressed var. S795 were compared in 4 organs (Figure 4). Depending on the gene, introgressed homeologs were expressed differently from the E^a homeolog in almost all organs. For gene S_{H3-81} , the contribution to the transcriptome of the E^a homeolog was higher in the introgressed variety than in var. Caturra. In contrast, gene S_{H3-83} showed a lower contribution of the homeolog E^a in the introgressed plants than in var. Caturra. As in the previous experiment, the organ factor was statistically significant ($P < 0.001$ for the gene S_{H3-81} and $P < 0.01$ for the gene S_{H3-83}).

Discussion

Variation in the Homeologous Gene Expression

In an allopolyploid, 2 diverged genomes merge in a common nucleus and contribute either equally or disproportionately to the transcriptome (Adams et al. 2003; Chaudhary et al. 2009). In this study, in *C. arabica*, we analyzed the relative contribution of 13 homeologous gene pairs to the transcript pool in 5 organs and in 2 contrasted growth conditions. The genes analyzed in the present study are located in a 350-kb interval of the S_{H3}

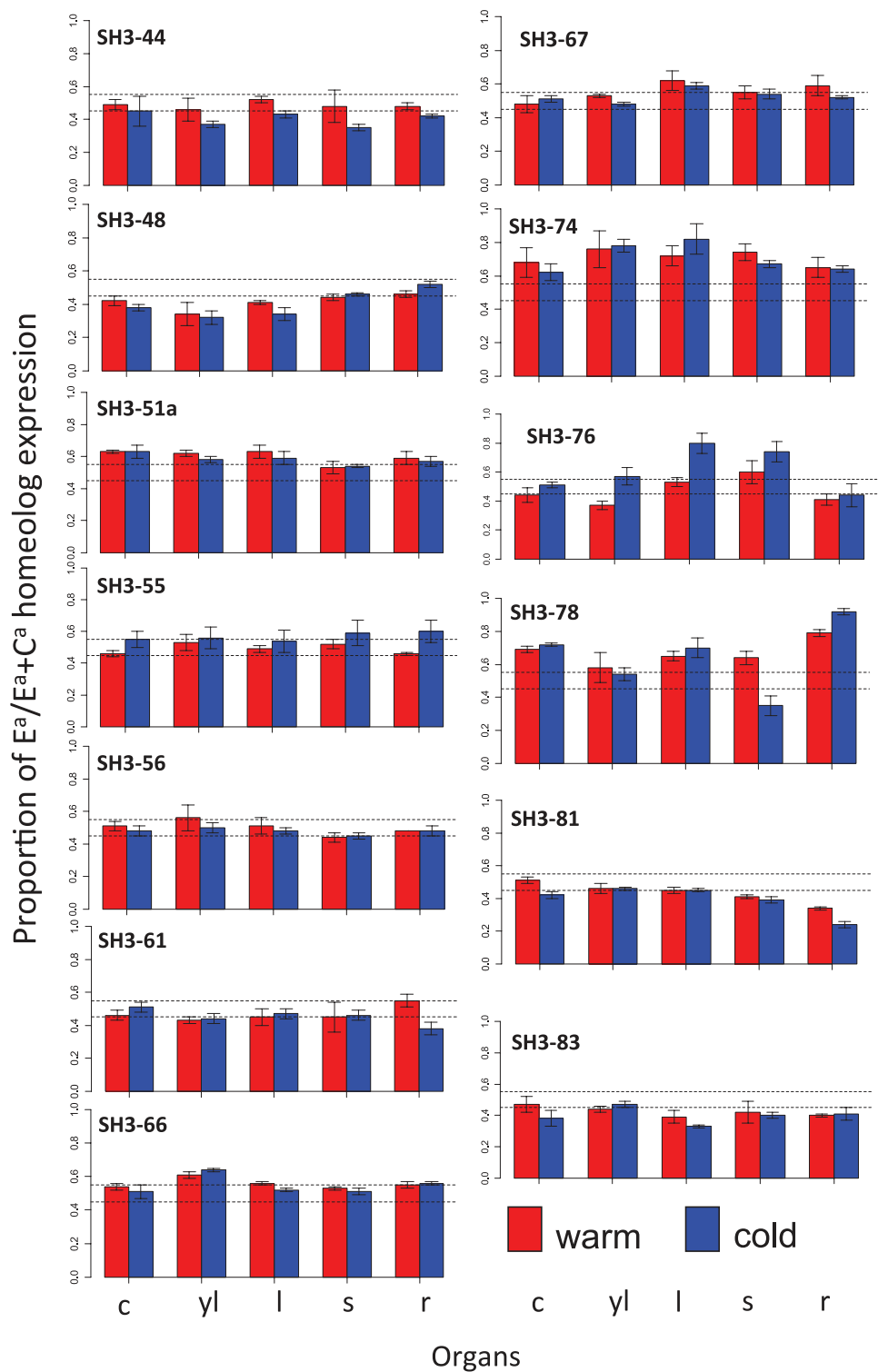


Figure 2. Representation of the variation in the relative contribution of homeologs ($E^a/E^a + C^a$) for the 13 genes as a function of the organ (c, cotyledon; yl, young leaf; l, leaf; s, stem; r, root) and the culture conditions.

region of *Coffea* spp. Because the sequence was available for both the subgenomes, it was possible to unequivocally assign each SNP variant to its respective subgenome (Lashermes et al. 2010). The relative contribution of the

Table 2 ANOVA of the relative expression of the homeologs ($E^a/E^a + C^a$) for the 13 genes as a function of the organ and culture condition.

Studied factors	S _{H3-44}	S _{H3-48}	S _{H3-51a}	S _{H3-55}	S _{H3-56}	S _{H3-61}	S _{H3-66}	S _{H3-67}	S _{H3-74}	S _{H3-76}	S _{H3-78}	S _{H3-81}	S _{H3-83}
Organs													
Cotyledon	0.47	0.40	0.63	0.51	0.50	0.49	0.53	0.49	0.65	0.47	0.71	0.46	0.42
Young leaf	0.41	0.33	0.60	0.55	0.52	0.43	0.63	0.50	0.77	0.49	0.56	0.46	0.46
Leaf	0.47	0.37	0.61	0.52	0.49	0.46	0.54	0.61	0.77	0.65	0.67	0.45	0.36
Stem	0.42	0.45	0.54	0.55	0.45	0.45	0.52	0.55	0.70	0.67	0.49	0.40	0.41
Root	0.44	0.50	0.58	0.54	0.48	0.45	0.56	0.54	0.65	0.43	0.87	0.28	0.41
Probability <i>F</i>	0.019*	0.000***	0.000***	NS	0.000***	NS	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
Conditions													
Warm	0.49	0.40	0.60	0.50	0.50	0.46	0.56	0.55	0.71	0.48	0.66	0.44	0.42
Cold	0.40	0.41	0.58	0.57	0.48	0.45	0.55	0.53	0.71	0.61	0.65	0.39	0.40
Probability <i>F</i>	0.000***	NS	NS	0.000***	0.024*	NS	0.016*	0.036*	NS	0.000***	0.029*	0.000***	0.013*
Interactions													
Warm × cotyledon	0.49	0.42	0.63	0.46	0.51	0.46	0.54	0.48	0.68	0.44	0.69	0.51	0.47
Warm × young leaf	0.46	0.34	0.62	0.53	0.56	0.43	0.61	0.53	0.76	0.37	0.58	0.46	0.44
Warm × leaf	0.52	0.41	0.63	0.49	0.51	0.45	0.56	0.62	0.72	0.53	0.65	0.45	0.39
Warm × stem	0.48	0.44	0.53	0.52	0.44	0.45	0.53	0.55	0.74	0.60	0.64	0.41	0.42
Warm × root	0.48	0.46	0.59	0.46	0.48	0.55	0.55	0.59	0.65	0.41	0.79	0.34	0.40
Cold × cotyledon	0.45	0.38	0.63	0.55	0.48	0.51	0.51	0.51	0.62	0.51	0.72	0.42	0.38
Cold × young leaf	0.37	0.32	0.59	0.56	0.50	0.44	0.64	0.48	0.78	0.57	0.54	0.46	0.47
Cold × leaf	0.43	0.34	0.59	0.54	0.48	0.47	0.52	0.59	0.82	0.80	0.70	0.45	0.33
Cold × stem	0.35	0.46	0.54	0.59	0.45	0.46	0.51	0.54	0.67	0.74	0.35	0.39	0.40
Cold × root	0.42	0.52	0.57	0.60	0.48	0.38	0.56	0.52	0.64	0.44	0.92	0.24	0.41
Probability <i>F</i>	NS	0.01**	NS	NS	NS	NS	0.002**	0.029*	0.021*	0.017*	0.000***	0.000***	0.024*

****P* < 0.001, ***P* < 0.01, **P* < 0.05. NS, not significant.

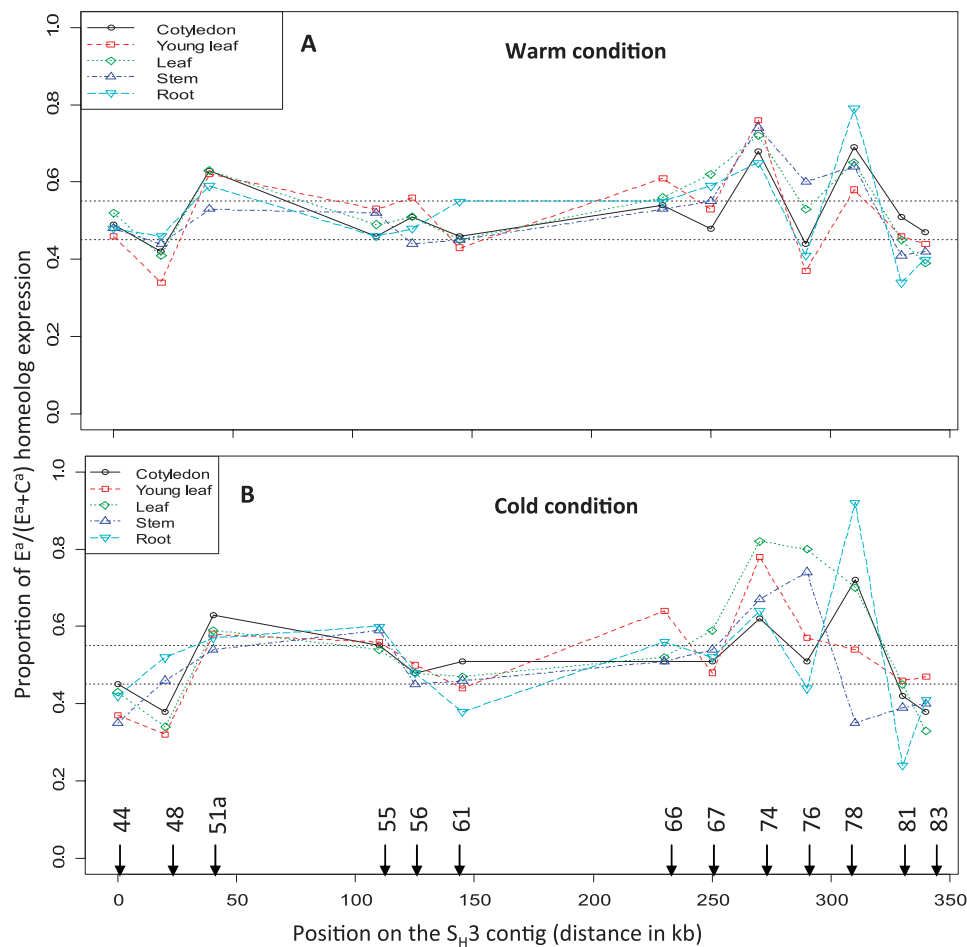


Figure 3. Representation of the average relative homeologous expression of 13 genes ($E^a/E^a + C^a$) in different organs and for 2 growth condition (A, warm; B, cold).

homeologous genes was estimated by the SNP ratio quantification method on dideoxy-terminated sequences (Rauscher et al. 2002). Comparison of the ratio of 2 SNPs

on the same gene and determination of the coefficients of variation of replicated samples of each SNP studied confirmed that the method is accurate and reliable.

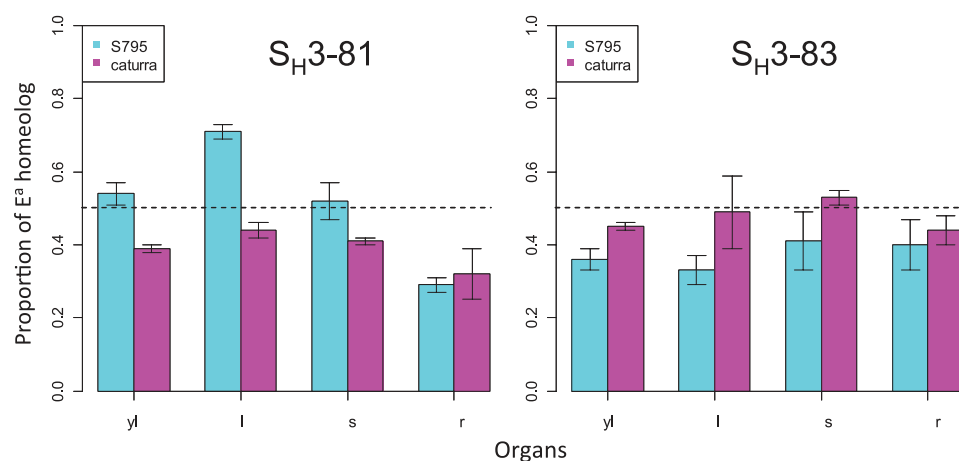


Figure 4. Variation in the relative contribution of homeologs between Caturra ($E^a/E^a + C^a$) and an introgressed variety S795 ($E^a/E^a + L$) for the genes S_H3-81 and S_H3-83 as a function of the organ (yl, young leaf; l, leaf; s, stem; r, root).

Although time consuming and more expensive per gene than methods enabling genome-wide analysis, this method appears to be suitable for the analysis of a limited gene sampling in a wide range number of conditions.

In our experiments involving combinations of 130 gene pairs/growing conditions/organs, the expression of both homeologs was always detected. As reported in several polyploids (Adams et al. 2003; Chaudhary et al. 2009; Flagel et al. 2009), the relative contribution of the homeologous genes to the transcriptome varied from equivalent to significantly biased expression depending on the organ and on the growing conditions. It is noteworthy that although the growing conditions were suitable for one or other of the parental species, neither subgenome appeared preferentially expressed. Similarly, the comparison between *C. arabica* and the *C. liberica*-introgressed *C. arabica* showed that both E^a and L homeologs were significantly expressed.

In our study, no case of gene silencing or organ-specific silencing was detected. Nevertheless, 10 out of 13 sampled genes exhibited biased expression, 4 genes toward E^a, 4 genes toward C^a, and 2 genes toward E^a or C^a depending on the organ considered. In the natural allopolyploid *G. hirsutum* (Adams et al. 2003), the proportion of biased homeologous genes was estimated at 25% of 40 genes and in the hexaploid wheat *T. aestivum* (Mochida et al. 2003) at 83% of 90 genes. In *C. arabica*, 22% of 2069 EST contigs analyzed showed biased expression (10% overexpressed C^a and 12% E^a) (Vidal et al. 2010). The difference between these results and ours could be due to the limited number of genes analyzed in the present study and to the different sensitivity and/or accuracy of the methods used to estimate relative homeologous gene expression. In particular, our approach based on SNP ratio quantification on dideoxy-terminated sequences was able to detect significant bias in homeolog expression even when the bias was small (10%), whereas only expression biases higher than 2/1 were considered in the in silico analysis performed by Vidal et al. (2010). In addition, because homeologous gene expression varies with the organ, the probability of detecting gene silencing in a study of several organs is greater than in a study of only one organ. Biologically, these divergent results indicate that homeologous gene expression after polyploidization may occur differently depending on the polyploid species. The biased gene expression may result from a combination of factors: inherited differences from diverged parental genomes and postdivergence structural changes between homeologs. (Lin et al. 2010).

Organ divergence of duplicate gene expression patterns provides information on the evolution of the plant species (Buggs et al. 2010). Indeed, after the polyploidization event, the duplicated genes appear to evolve in 2 steps at different timescales (Adams and Wendel 2005; Flagel et al. 2008). The first step occurs immediately after the genomic merger of the diploids, and some of the duplicated genes disappear in genomic rearrangements. This is followed by a long-term evolution step during which a race between subfunctionalization (which enables preservation of duplicated genes) and pseudogenization (inactivation of a gene copy) starts for the

duplicated genes (Rapp and Wendel 2005). Reciprocal epigenetic silencing or biased expression of homeologs in different organs or tissues may enhance the probability of retention of homeologous genes (Lynch and Force 2000; Chaudhary et al. 2009), thus preserving the availability of these genes for variation in expression and phenotypic diversification (Rapp and Wendel 2005; Chaudhary et al. 2009; Flagel et al. 2009). Thus, each allopolyploid lineage resulting from an independent event evolves in its own unique context according its evolutionary history (Chen and Ni 2006; Hufton and Panopoulou 2009; Jackson and Chen 2010). *Coffea arabica* is a recent stable allopolyploid involving little divergent diploid genomes. Our results showing limited variation of relative homeologous gene expression between organs and the absence of signs of subfunctionalization suggest that *C. arabica* has not undergone diploidization.

Regulation of the Homeologous Gene Expression

In polyploids, a variety of changes in gene expression have been observed in comparison with their diploid parents (Adams et al. 2004; Chaudhary et al. 2009). Some of these changes could be explained by interactions of homeologous regulatory factors with their target genes. Several regulatory models of the homeologous gene expression have been proposed (Riddle and Birchler 2003; Pignatta and Comai 2009). According to these models, a slight variation in the ratio of expression of the homeologous genes could mean that both transcription factors are able to bind almost equally well to both homeologous target genes. In wheat (Mochida et al. 2003), the homeologous gene expression and silencing were shown to be regulated at the level of the gene and not at the chromosome or genome level. In cotton, 2 regions appeared to be regulated very differently (Flagel et al. 2009): coordinated for 40 kb and fine-scale control (gene by gene), whereas Adams et al. (2003) reported that epigenetic factors act at the gene level together with other underlying controls.

In *C. arabica*, nonsignificant correlations between physical distance and relative homeologous gene expression suggest a gene-specific regulatory control. Because the 2 subgenomes exhibit low sequence divergence and variations in homeologous gene expression between the 2 subgenomes are limited and consistent across organs, it is likely that homeologous regulatory factors interoperate and control the regulation of the transcription of both subgenomes. Furthermore, depending on the growing temperature, other regulatory mechanisms including localized divergence among homeologous *cis*-regulatory regions or epigenetic marks could influence the homeologous gene contribution. In addition, comparison of homeologous gene expression of 2 genes in *C. arabica* and in an introgressed *C. arabica* suggests the participation of similar *trans*-regulation factors. In this particular context of allopolyploidy, homeologous newly introgressed genes from another species of the genus *Coffea* seem regulated like other homeologous genes in the genome of *C. arabica*. These regulation mechanisms could explain why *C. arabica* is able to adapt to different growing conditions better than its diploid parental species. It would be interesting to compare gene expression of *C. arabica* with its

diploid progenitor to specify the relative roles of *cis*- and *trans*-regulation (Chaudhary et al. 2009).

References

- Adams KL. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. *J Hered*. 98(2):136–141.
- Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci U S A*. 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. 2005. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol*. 8(2):135–141.
- 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(1–2):51–64.
- Bardil A, Dantas de Almeida J, Combes MC, Lashermes P, Bertrand B. 2011. Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytol*. doi: 10.1111/j.1469-8137.2011.03833.x.
- 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 Phytol*. 186(1):175–183.
- Buggs RJA, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE. 2011. Transcriptomic shock generates evolutionary novelty in a newly formed natural allopolyploid plant. *Curr Biol*. 21:551–556.
- Cenci A, Combes MC, Lashermes P. 2010. Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleohexaploid ancestral genome. *Mol Genet Genomics*. 283(5):493–501.
- Cenci A, Combes MC, Lashermes P. Forthcoming 2011. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant Mol Biol*. *Mol Genet Genomics*.
- 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(2):503–517.
- Chen ZF. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol*. 58:377–406.
- Chen ZF, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays*. 28(3):240–252.
- Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat Rev Genet*. 6(11):836–846.
- DaMatta FM, Cochicho Ramalho JD. 2006. Impacts of drought and temperature stress on *Coffea* physiology and production: a review. *Braz J Plant Physiol*. 18(1):55–81.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P. 2006. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot J Linn Soc*. 152:465–512.
- Dong S, Adams KL. 2011. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytol*. 190(4):1045–1057.
- Flagel L, Udall J, Nettleton D, Wendel J. 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol*. 6:1–16.
- Combes et al. • Homeologous Gene Expression in Allopolyploid
- Flagel LE, Chen L, Chaudhary B, Wendel JF. 2009. Coordinated and fine-scale control of homeologous gene expression in allotetraploid cotton. *J Hered*. 100(4):487–490.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp*. 41:95–98.
- Hegarty MJ, Hiscock SJ. 2009. The complex nature of allopolyploid plant genomes. *Heredity*. 103(2):100–101.
- Hovav R, Udall J, Chaudhary B, Rapp R, Flagel L, Wendel JF. 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc Natl Acad Sci U S A*. 105(16):6191–6195.
- Hufton AL, Panopoulou G. 2009. Polyploidy and genome restructuring: a variety of outcomes. *Curr Opin Genet Dev*. 19(6):600–606.
- Jackson S, Chen ZJ. 2010. Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol*. 13(2):153–159.
- Lashermes P. 2000. Brief communication. Single-locus inheritance in the allotetraploid *Coffea arabica* L. and interspecific Hybrid *C. arabica* x *C. canephora*. *J Hered*. 91(1):81–85.
- 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 Genet Genomes*. 6(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. *Mol Gen Genet*. 261(2):259–266.
- Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science*. 320(5875):481–483.
- Lin JY, Stupar RM, Hans C, Hyten DL, Jackson SA. 2010. Structural and functional divergence of a 1-Mb duplicated region in the soybean (*Glycine max*) genome and comparison to an orthologous region from *Phaseolus vulgaris*. *Plant Cell*. 22:2545–2561.
- Liu Z, Adams KL. 2007. Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr Biol*. 17(19):1669–1674.
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics*. 154(1):459–473.
- Mantel N. 1967. The detection of disease clustering and generalized regression approach. *Cancer Res*. 27:209–220.
- Mochida K, Yamazaki Y, Ogihara Y. 2003. Discrimination of homeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Genet Genomics*. 270(5):371–377.
- Nomura T, Ishihara A, Yanagita RC, Endo TR, Iwamura H. 2005. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proc Natl Acad Sci U S A*. 102(45):16490–16495.
- 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 Genet*. 19(3):141–147.
- Peng H, Zhang J, Wu X. 2008. The ploidy effects in plant gene expression: progress, problems and prospects. *Sci China Ser C Life Sci*. 51(4):295–301.
- Piepho HP. 2005. Permutation tests for the correlation among genetic distances and measures of heterosis. *Theor Appl Genet*. 111:95–99.
- Pignatta D, Comai L. 2009. Parental squabbles and genome expression: lessons from the polyploids. *J Biol*. 8(4):43.
- Prakash NS, Marques DV, Varzea VMP, Silva MC, Combes MC, Lashermes P. 2004. Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *Coffea arabica* L. *Theor Appl Genet*. 109(6):1311–1317.
- Rapp RA, Udall JA, Wendel JF. 2009. Genomic expression dominance in allopolyploids. *BMC Biol*. 7:18.
- Rapp RA, Wendel JF. 2005. Epigenetics and plant evolution. *New Phytol*. 168(1):81–91.

Rauscher JT, Doyle JJ, Brown AHD. 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (*Leguminosae*) polyploid complex. *Mol Ecol.* 11(12):2691–2702.

Riddle NC, Birchler JA. 2003. Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. *Trends Genet.* 19(11):597–600.

Stamati K, Mackay I, Powell W. 2009. A quantitative genomic imbalance gene expression assay in a hexaploid species: wheat (*Triticum aestivum*). *Genome.* 52(1):89–94.

Vidal RO, Mondego JMC, Pot D, Ambrosio AB, Andrade AC, Pereira LFP, Colombo CA, Vieira LGE, Carazzolle MF, Pereira GAG. 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 Physiol.* 154(3):1053–1066.

Received June 20, 2011; Revised September 21, 2011;
Accepted September 30, 2011

Corresponding Editor: Reid Palmer

2 Caractérisation et variation de l'expression relative des gènes homéologues chez *Coffea arabica*

Toujours avec l'idée de varier les modèles et les dispositifs expérimentaux, la seconde étude menée sur *C. arabica* visait à caractériser l'expression relative des gènes homéologues, et sa variation en fonction de la température de culture des plants de *C. arabica*. Pour cela, des plants de *C. arabica* ont été cultivés à deux températures de culture convenant chacune à l'une des espèces parentales de l'allopolyploïde. En raison de l'évolution des méthodes de séquençage, cette étude a été menée à l'échelle du génome en utilisant des données de séquençage d'ARN par la méthode Illumina (Illumina, San Diego, CA, USA). Les données de séquençage ont été mappées sur une référence correspondant à un assemblage d'EST (Expressed Sequence Tag) qui représentait 56216 unigènes. Dans cette analyse, l'expression relative environ 10000 paires de gènes homéologues a été examinée. Cette étude a montré que plus de la moitié des gènes (65%) présente un niveau équivalent de l'expression des gènes homéologues, et cela quel que soit la température. En effet l'expression relative des gènes homéologues varie peu, pour 92% des gènes, elle présente moins de 10% de variation entre les températures de culture. La contribution des deux sous-génomes à l'expression de l'ensemble des gènes et la légère variation en fonction de la température de culture suggèrent les effets de mécanismes de *trans*-régulation provenant des deux sous-génomes sur la régulation de l'expression des gènes homéologues. Ce travail est présenté dans l'article suivant : **Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures**. Marie-Christine Combes, Alexis Dereeper, Dany Severac, Benoît Bertrand, Philippe Lashermes. 2013. New phytologist 200: 251-260.

Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures

Marie-Christine Combes¹, Alexis Dereeper¹, Dany Severac², Benoît Bertrand³ and Philippe Lashermes¹

¹IRD, UMR RPB (IRD, CIRAD, Université Montpellier II), 911 avenue Agropolis, BP 64501, 34394, Montpellier Cédex 5, France; ²MGX-Montpellier GenomiX, Institut de Génétique Fonctionnelle, 141 rue de la Cardonille, 34094, Montpellier Cédex 5, France; ³CIRAD, UMR RPB (IRD, CIRAD, Université Montpellier II), 911 avenue Agropolis, BP 64501, 34394, Montpellier Cédex 5, France

Summary

Author for correspondence:
Marie-Christine Combes
Tel: +033 467416434
Email: marie-christine.combes@ird.fr

Received: 5 April 2013
Accepted: 14 May 2013

New Phytologist (2013) 200: 251–260
doi: 10.1111/nph.12371

Key words: adaptation, *cis trans*-regulation, coffee (*Coffea arabica*), homoeologous gene expression, polyploidy, RNA-sequencing, transcriptome.

- Polyploidy has occurred throughout the evolutionary history of plants and led to diversification and plant ecological adaptation. Functional plasticity of duplicate genes is believed to play a major role in the environmental adaptation of polyploids. In this context, we characterized genome-wide homoeologous gene expression in *Coffea arabica*, a recent allopolyploid combining two subgenomes that derive from two closely related diploid species, and investigated its variation in response to changing environment.
- The transcriptome of leaves of *C. arabica* cultivated at different growing temperatures suitable for one or the other parental species was examined using RNA-sequencing. The relative contribution of homoeologs to gene expression was estimated for 9959 and 10 628 genes in warm and cold conditions, respectively.
- Whatever the growing conditions, 65% of the genes showed equivalent levels of homoeologous gene expression. In 92% of the genes, relative homoeologous gene expression varied < 10% between growing temperatures.
- The subgenome contributions to the transcriptome appeared to be only marginally altered by the different conditions (involving intertwined regulations of homeologs) suggesting that *C. arabica*'s ability to tolerate a broader range of growing temperatures than its diploid parents does not result from differential use of homoeologs.

Introduction

Polyploidy has occurred throughout the evolutionary history of eukaryotes, especially in flowering plants, which have all undergone at least one round of polyploidization (Jiao *et al.*, 2011). Many major agricultural crop plants including wheat (*Triticum aestivum*), cotton (*Gossypium hirsutum*), sugarcane (*Saccharum officinarum*) and coffee (*Coffea arabica*) are allopolyploid, that is, they combine two or more sets of distinct genomes (homoeologous chromosomes) after hybridization of diverged genomes and genome duplication. Other crops such as potato (*Solanum tuberosum*) and alfalfa (*Medicago sativa*) are autopolyploids derived from duplication of the same genome (Chen & Ni, 2006; Jackson & Chen, 2010). Polyploidy considerably influences plant species diversity, giving rise to novel phenotypes and leading to ecological diversification and colonization of new niches (Otto & Whitton, 2000; Adams, 2007; Madlung, 2013). Recent studies documented rapid and dynamic changes in genomic structure and gene expression in plant polyploids (Comai, 2005; Buggs *et al.*, 2011). Genomic plasticity (Leitch & Leitch, 2008) enables polyploids to combine two divergent genomes of

the parental diploid species in the same cell (Chen, 2007; Hegarty & Hiscock, 2009), whereas functional plasticity of duplicate genes enables them to regulate gene expression to adapt to their environment (Jackson & Chen, 2010).

Numerous studies have investigated variations in gene expression by comparing the expression of the diploid parents and of the allopolyploids. In the additive model, allopolyploids are supposed to display mid-parent expression value. However many exceptions have been found in *Arabidopsis* (Wang *et al.*, 2006), *Gossypium* (Chaudhary *et al.*, 2009; Rapp *et al.*, 2009), *Senecio* (Hegarty *et al.*, 2006, 2008) and coffee (Bardil *et al.*, 2011), suggesting that differential regulation of gene expression is a common feature of allopolyploid plants. Changes in homoeologous expression have been investigated in one to several dozen genes in different organs of allopolyploids (Adams *et al.*, 2003, 2004; Mochida *et al.*, 2003; Nomura *et al.*, 2005; Chaudhary *et al.*, 2009; Flagel *et al.*, 2009; Buggs *et al.*, 2010a) and linked to different types of stress (Liu & Adams, 2007; Stamati *et al.*, 2009; Dong & Adams, 2011; Combes *et al.*, 2012). More recently, studies have been performed at the genome scale in a single organ (Flagel *et al.*, 2008; Hovav *et al.*, 2008; Buggs *et al.*, 2010a; Vidal

et al., 2010; Higgins *et al.*, 2012; Ilut *et al.*, 2012; Yoo *et al.*, 2012). In allopolyploids, homoeologous genes may exhibit unequal expression patterns (i.e. homoeolog expression bias; Grover *et al.*, 2012) either locally or throughout the genome. The ratio of homoeologous expression can vary with the organ, its development, and in response to abiotic stresses. At the genome scale, differentially expressed homoeologous genes have been observed in some allopolyploids, preferentially toward one of the two subgenomes (i.e. unbalanced homoeolog expression bias; Grover *et al.*, 2012) (Buggs *et al.*, 2010b; Vidal *et al.*, 2010; Ilut *et al.*, 2012). The differences observed in the studies of allopolyploids indicate that no single unifying factor governs genome-specific expression biases. Evolution of genome-specific expression occurs via a unique *ad hoc* mixture of genetic and epigenetic regulatory mechanisms within different species (Chaudhary *et al.*, 2009). A multiple model approach is needed to study how allopolyploid plants respond to the merger of genomes (Hegarty, 2011).

Coffee is one of the most important agricultural commodities and ensures the livelihood of > 80 million people worldwide. Among the around one hundred *Coffea* species identified (Davis *et al.*, 2006), two, *C. arabica* and *C. canephora*, are cultivated species and account for 65% and 35% of world coffee production, respectively. *Coffea* is a relatively recent genus estimated to have originated 0.5 million yr ago (Anthony *et al.*, 2010). *Coffea* species have variable phenotypes and are able to grow in specific environments. The low sequence divergence and phylogenetic analysis of *Coffea* species suggest a rapid and radial mode of speciation (Lashermes *et al.*, 1997; Cros *et al.*, 1998; Anthony *et al.*, 2010). All *Coffea* species are diploid, except *C. arabica*, which is allotetraploid ($2n=4 \times=44$) and derived from a recent (< 50 000 yr ago) interspecific hybridization between two diploid species: *C. eugenoides* and *C. canephora* (Lashermes *et al.*, 1999; Cenci *et al.*, 2012). Homoeologous genomes in *C. arabica* are designated E^a and C^a according to their parental origin. The two parental species are closely related, and the two subgenomes have low sequence divergence (i.e. 1.3% average difference for genes, Cenci *et al.*, 2012). The evolution of the *C. arabica* genome after the polyploidization event has not yet been investigated but almost perfect gene synteny between the two subgenomes has been shown for several genomic regions (Yu *et al.*, 2011; Cenci *et al.*, 2012). Nevertheless *C. arabica* displays disomic inheritance with bivalent pairing of homoeologous chromosomes (Lashermes, 2000). The two parental species exhibit different agroecological adaptations. *C. canephora* and *C. eugenoides* are endemic in regions where the annual mean temperature ranges from respectively 22 to 26°C and from 18 to 23°C with no major oscillations (DaMatta & Ramalho, 2006; Davis *et al.*, 2006). By contrast, *C. arabica*, whose optimum mean annual temperature ranges from 20 to 24°C, can be grown in regions with marked variations in thermal amplitude. Studies have already been conducted to analyze the expression of genes in *C. arabica* in different growing conditions. First, the genome-wide expression pattern of *C. arabica* and its related parental species were compared and for some genes, genomic expression dominance (i.e. genomic expression level dominance; Grover *et al.*, 2012)

modulated by the growing conditions (Bardil *et al.*, 2011) was revealed. Second, the relative expression of homoeologs of *C. arabica* was analyzed in 13 genes located in the same genomic region. Analysis of relative homoeologous expression revealed moderate variations across organs and growing conditions (Combes *et al.*, 2012).

The aim of this study was to assess the contribution of subgenomes to the *C. arabica* transcriptome and to identify variations related to adaptation to different growing conditions. Transcriptome of leaves of *C. arabica* plants grown in contrasted temperature conditions was sequenced. RNA-sequencing (RNA-seq) data enabled detection of SNPs (single nucleotide polymorphisms) between the two subgenomes, and quantification of transcribed homoeolog abundance. To this end, the allopolyploid reads were first aligned against a *C. canephora* reference transcriptome of 56 216 unigenes resulting from deep RNA-sequencing. Then the genome origin of the allopolyploid SNPs was inferred by comparing the SNPs with available sequences of alleles of the two current diploid progenitor species. These data enabled characterization of genome-wide homoeologous gene expression in *C. arabica*, and investigation of the influence of growing temperatures on homoeologous gene expression. The relationships between homoeologous gene expression and total gene expression were also explored. Thanks to the recent origin of *C. arabica*, the relative homoeologous gene expression in the allopolyploid was then compared with differential gene expression in modern diploid progenitor species. This comparison provided information on changes in gene expression that were attributed to genome merger and caused by *cis*- and *trans*-regulatory divergences between the two parental diploid species and, hence, between the two subgenomes (Chaudhary *et al.*, 2009; Shi *et al.*, 2012). To our knowledge, this study of variation in homoeologous gene expression in two contrasted growing conditions is the first to be carried out at the genome scale in an allopolyploid. In addition to advancing knowledge on overall homoeologous gene expression in allopolyploids, new features connected with *C. arabica*'s ability to adapt to environmental variations were also revealed.

Materials and Methods

Plant material

Seedlings of *Coffea arabica* L. var. Caturra, an inbred line, were cultivated for 2 months in two sets of contrasted growing conditions with different diurnal/nocturnal temperatures: 30°C : 26°C (warm conditions) and 23°C : 19°C (cold conditions). The conditions were chosen to suit either *Coffea canephora* or *C. eugenoides* and did not prevent normal growth of the *C. arabica* plants, even though the conditions were more stressful for one or the other of the parent species. A different lower temperature regime was used than in the study of Bardil *et al.* (2011) to suit better requirements of *C. eugenoides*, and consequently to compare more contrasting temperature regimes. In the two climatic chambers, the photoperiod was set at 12 h d⁻¹, humidity at 80–90% and luminosity at 600 µmol photon m⁻² s⁻¹. In each

climatic chamber, the plants were grown in a randomized complete block design. At the same time, 3 h after the beginning of the diurnal period, young, fully expanded leaves were harvested from two plants (biological replicates) in each growing chamber. Plants of *C. eugenoides* (genotype DA58 from the IRD collection) were cultivated in a glasshouse (26°C ± 3°C daytime and 24°C ± 3°C night-time temperatures, humidity maintained at over 60%). After collection, the samples were immediately flash frozen in liquid nitrogen and stored at –80°C until RNA extraction.

RNA sequencing

Total RNA was isolated from 1 g of material from each plant. Harvested samples were ground in liquid nitrogen and the powders suspended in 20 ml of an extraction buffer (5 M guanidinium isothiocyanate, 31 mM sodium acetate (pH 8), 1% β-mercaptoethanol, 0.88% (w/v) N-lauroyl sarcosine and 1% (w/v) polyvinylpyrrolidone 40). The solutions were centrifuged at 15 000 g for 20 min at 4°C. RNA was purified on a 5.7 M cesium chloride layer (18 000 g for 20 h at 20°C). The RNA pellets were rinsed twice with 70% (v/v) ethanol and dissolved in 100 µl of RNase-free water. The quality and the concentration of extracted RNA samples were determined using the Agilent DNA1000 (Agilent, Santa Clara, CA, USA).

The five mRNA libraries were built and sequenced at the MGX platform (Montpellier Genomix, *Institut de Génomique Fonctionnelle*, Montpellier France). The RNA libraries were constructed using the 'RNA-seq sample prep' kit (Illumina, San Diego, CA, USA). RNA was sequenced using the Illumina SBS (sequence by synthesis) technique on a HiSeq 2000 (Illumina). Image analysis, base calling and quality filtering were performed using Illumina software. For the *C. eugenoides* library, 32 million reads were obtained, while for the four *C. arabica* libraries, a total of 326 million reads (ranging from 64 to 88 million reads per library) was retained.

RNA-seq data processing

For each sample, the 72-nt single reads were aligned using BWA (Li & Durbin, 2009) against a *C. canephora* expressed sequence tag (EST) assembly (with 56 216 unigenes) as transcriptome reference. This reference base was built with transcripts from 'SGN coffee unigene build III' (16 046 clusters, SOL), 'Brazilian Coffee Initiative' (16 801 clusters) (Vidal *et al.*, 2010) and 'French *C. canephora* sequencing consortium' (52 683 clusters). Two former resources are the result of Sanger sequencing (*c.* 78 470 ESTs from various tissues), the latter of deep Illumina sequencing (*c.* 118 million 76 nt reads from total RNA isolated from leaf, stem and flower tissues). These three resources were assembled using CAP3 software (<http://seq.cs.iastate.edu/>) with default parameters. Redundancy among contigs was investigated using BLAST (*i.e.* all-against-all sequence comparison). Duplicates (*i.e.* > 95% identity on > 97% of sequence length) were removed, and only the longest sequences kept. A total of 56 216 nonredundant clusters with an average length of 663 bp were retained. A

maximum of four mismatched nucleotides (including gaps) between the read and the reference transcriptome sequence was allowed. The reads were also mapped to a *C. eugenoides* transcriptome reference (a subset of 8351 unigenes) obtained from the *C. eugenoides* transcriptome sequenced by Illumina technology and mapped to the *C. canephora* transcriptome reference.

Data analysis

From the selection of the unigenes up to the data analysis, all statistical analyses and data processing were performed with R software.

Estimation of transcript level Mapped sequence counts were processed using the 'DESeq' package (Anders & Huber, 2010) to estimate the transcript level. This method is particularly suitable for experiments with biological replicates. The count data were normalized on the total number of counts taking the variance and the mean of the biological replicates into account. For each sample, the normalized expression level for each unigene of the reference transcriptome was expressed as reads per kilobase of transcript sequence per million reads (RPKM) (Mortazavi *et al.*, 2008). For each unigene, the fold change of the gene expression between the both growing conditions was analyzed and the statistical significance estimated using an adjusted p-value for multiple testing (Benjamini–Hochberg method).

Estimation of homoeologous gene expression Different categories of SNPs can be found in allopolyploids that present a high level of genome complexity, including sequence variations between subgenomes (homoeologous-SNPs) that co-exist with allelic variations within subgenomes (homologous-SNPs) (Kaur *et al.*, 2012). For SNP discovery, the aligned allopolyploid reads were analyzed with the GATK toolkit (McKenna *et al.*, 2010) using the Unified Genotyper module to obtain SNPs list and allelic data, and the Depth Of Coverage module to obtain depth coverage information. To infer the SNP genome origin and distinguish homoeologous SNPs, the detected allopolyploid SNPs were compared to the corresponding nucleotides in both modern parental diploid genomes, *C. canephora* and *C. eugenoides*, using SNIploid, a newly developed tool (M. Peralta *et al.*, unpublished, <http://sniploy.cirad.fr/cgi-bin/sniploid.cgi>). Notably, SNIploid allows the user to set a minimum depth coverage to consider a sequence position. Minimum depth coverages of 20 and 4 were required for the samples of *C. arabica* and *C. eugenoides*, respectively. A read cutoff of 20 was selected to account for the possibility of a bias in homoeolog expression in the allopolyploid.

For each sample, homoeologous SNP sites were used to calculate a site homoeologous expression ratio. For each unigene, a weighted mean of site homoeologous expression ratios was calculated. Unigenes containing only one homoeologous SNP site, or with different site homoeologous expression ratios (prop.test with Bonferroni correction, $P < 0.001$) were discarded. Finally the biological replicates were combined by averaging the weighted means of each unigene. Unigenes displaying > 20% of variation between the two homoeologous gene expression estimations, or

< 80 total reads per unigene were discarded. The parameters were selected to ensure a reliable estimate of homoeologous gene expression. The data and the unigenes of the warmer and the cooler conditions were treated using the same method. Homoeologous gene expression was deduced from the results by considering unigenes as equivalent to genes. The homoeologous gene expression corresponding to homoeolog-specific read counts of the total read counts was expressed as the percentage of the C^a homoeolog in the total gene expression.

Comparison of subgenome and genome specific expression between the allopolyploid and its diploid progenitor species By comparing subgenome-specific expression in the allopolyploid with the genome-specific expression in the diploids, using the method applied by Wittkopp *et al.* (2004) to alleles of interspecific hybrids in the group of *Drosophila* species, it is possible to separate expression variation into *cis*- and *trans*-origin. Even though we had no synthetic F₁ hybrids of *Coffea* diploid species, these analyses were possible because a slight sequence divergence was observed between the two constitutive genomes of *C. arabica* and those of its modern parental species (Lashermes *et al.*, 1999, 2010; Cenci *et al.*, 2012). In theory, differences in genome-specific expression due to *cis*-regulatory divergences between the diploids and shared by the allopolyploid will fall along a 1 : 1 diagonal, whereas, when present in the allopolyploid, *trans*-regulatory divergences will balance genomic-specific expression and instead fall along on equivalently expressed horizontal line. Genes that do not fall along either of these lines are affected by both *cis*- and *trans*-regulatory divergences in the allopolyploid and are assumed to be modified by a combination of changes in *cis*- and *trans*-factors. Microarray data from *C. canephora* and *C. eugenoides* obtained in warmer growing conditions (Bardil *et al.*, 2011) were used to apply this method to *C. arabica*. Microarray and RNA-seq data were found to be highly correlated ($r=0.63$ $P<0.001$) for 2900 unigenes in the two studies. A set of 1434 unigenes displaying significant difference in expression between the two diploid species (t -test, H₀: C/E = 1, $P<0.01$) were selected. The expression of these genes was analyzed using statistical tests: two-sided t -test (H₀: C^a/E^a = 1, $P<0.01$) and two-sided Wilcoxon test (H₀: C/E = C^a/E^a, $P<0.01$) to identify any *cis*- and *trans*-regulatory changes that occurred in the allopolyploid.

In order to improve our knowledge of the relationships between gene expression and homoeologous gene expression in an allopolyploid, we compared the mean of relative homoeologous gene expression of genes showing expression level dominance (i.e. that expressed similarity with one of the two diploids, and were up- or downregulated with respect to the other parent, 891 genes common to the current study and the study of Bardil *et al.* (2011) to the mean of relative homoeologous gene expression of all the genes studied (t -test, $P<0.01$). Likewise, to test if the expression level dominance characteristic of these genes is linked to a particular type of regulatory divergence between diploid species, the distribution of gene expression divergence (*cis*, *trans*, *cis* + *trans* combination) was compared to the distribution of gene expression divergence of the whole set of genes.

Gene ontology functional enrichment analysis Computational mapping and Plant GO-slim annotation were performed using Blast2GO software v2.6.4 (Conesa *et al.*, 2005; Conesa & Götz, 2008) with default parameters of 12 060 unigene sequences for which the relative expression of homoeologs was estimated either in the warmer or in the cooler conditions. For the genes selected at different steps of the analysis, molecular and cellular functions were catalogued using enrichment tests (FDR < 0.05) to evaluate statistical significance.

Results

Identification of homoeologous SNPs by RNA-seq

Among all the SNPs detected (from 218 538 to 271 947 SNPs depending on the sample concerned) a total of 82 844 and 91 528 (77 920 found in both growing conditions) were considered as homoeologous SNPs for the warmer and cooler conditions, respectively. Following the selection steps described in the material and methods section, 9959 and 10 628 genes were analyzed for homoeologous gene expression analysis in the warmer and colder growing conditions, respectively; including 8527 genes found in both growing conditions.

Homoeologous gene expression throughout the genome

The primary goal of this study was to investigate homoeologous gene expression among plant samples cultivated in two different growing conditions. For each of the two growing conditions, the genes were classified as the percentage of the C^a homoeolog relative to total gene expression in ten categories ranging from 0% to 100% (Fig. 1). Relative homoeologous gene expression was evaluated with an average confidence interval length of 8.5% at the 95% confidence level. The distribution patterns of genes in different categories were similar in the two growing conditions (χ^2 test = 10.5217, df = 7, $P=0.1609$). The proportion of genes

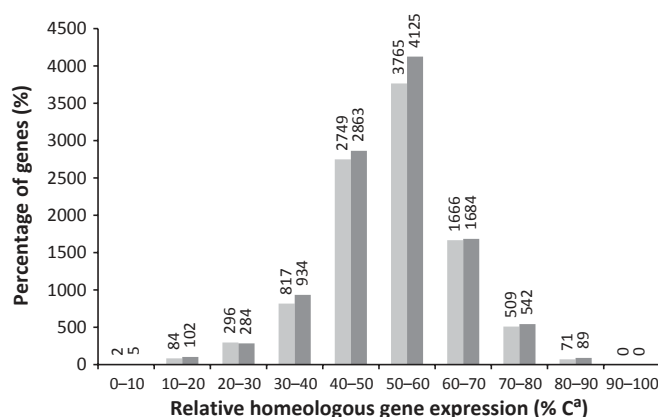


Fig. 1 Relative homoeologous gene expression in *Coffea arabica* in two different growing conditions (warm, light grey bars; cold, dark grey bars). *C. arabica* genes were classified according to the relative expression of their homoeologs represented by the contribution of subgenome C^a to total gene expression (% C^a). The exact number of genes in each category is indicated at the top of each column.

exhibiting a bias whatever the direction of the bias was almost the same: 37.5% and 34% for the warmer and cooler conditions, respectively (prop.test with Bonferroni correction, $H_0: C^a/E^a = 1$, $P < 0.001$). A very low percentage of genes (0.02% for the warmer temperature and 0.04% for the cooler temperature) showed extremely biased homoeologous gene expression (categories: 0–10% and 90–100% of C^a) toward one subgenome. The average percentage of the C^a homoeolog was 52.5% and 52.0% for the warmer and cooler temperatures, respectively.

Comparing the symmetric categories of the distribution revealed that, whatever the growing conditions, more genes showed biased homoeologous gene expression toward C^a than toward E^a . A mapping-related technical bias was investigated; the analysis of the reads of *C. arabica* cultivated in the cooler condition was repeated using a *C. eugenoides* reference transcriptome instead of a *C. canephora* reference transcriptome. The nature of the transcriptome reference used for mapping induced a slight technical bias (Supporting Information Fig. S1); however, the two relative homoeologous gene expression estimates were comparable and highly correlated (Fig. S2).

Gene ontology enrichment analysis was performed on genes showing biased and extremely biased homoeologous gene expression toward E^a or C^a . This analysis provided little information: a few GO terms showing statistically significant enrichment were revealed (Table S1); however, GO terms linked to 'enzymatic activity' were over-represented, whereas genes with GO terms linked to 'binding functions' were under-represented.

Homoeologous gene expression in response to different growing conditions

Variation in homoeologous gene expression between the two growing conditions For each of the 8527 genes analyzed in both growing conditions, the relative homoeologous gene expression in the warmer condition was compared to the relative homoeologous gene expression in the cooler condition. Figure 2 shows the distribution of the variation in the observed relative expression of the homoeologs in the two growing conditions whatever the direction of the variation. In 92% of the genes, homoeologous gene expression varied <10% between the two growing conditions; of the rest, variation was between 10% and 20% in 7.6% of genes and >20% in 0.4% of the genes. For the last two subsets of genes, shifts in the relative homoeologous gene expression favored the C^a or E^a homoeolog in 3.9% and 3.7% of genes, respectively. Homoeologous gene expression between the two growing conditions was highly correlated ($r = 0.87$ with $P < 0.001$) (Fig. 3).

Gene ontology enrichment analysis was performed on genes showing >10% of variation in relative homoeologous gene expression between the two growing conditions. This analysis revealed GO terms showing statistically significant enrichment. As in the previous analysis, genes described by GO terms linked to 'enzymatic activity' were over-represented, while genes described by GO terms linked to 'binding functions' were under-represented (Table S2). Gene ontology enrichment analyses performed on genes whose relative homoeologous gene expression

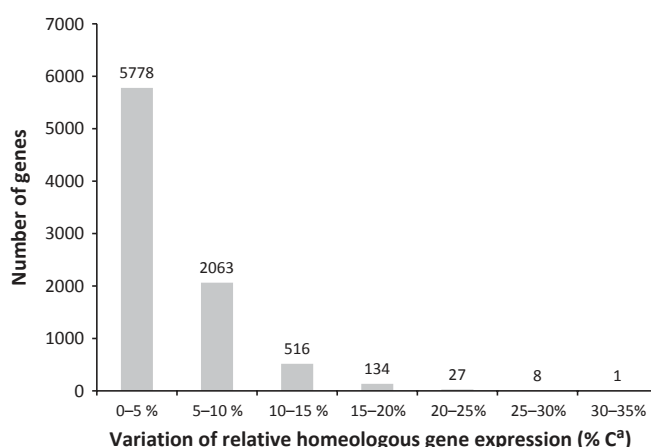


Fig. 2 Variation in the relative homoeologous gene expression in *Coffea arabica* in the two growing conditions (warm and cold). The genes were sorted by differences in relative homoeologous gene expression whatever the direction of the variation. The exact number of genes in each category is indicated at the top of each column.

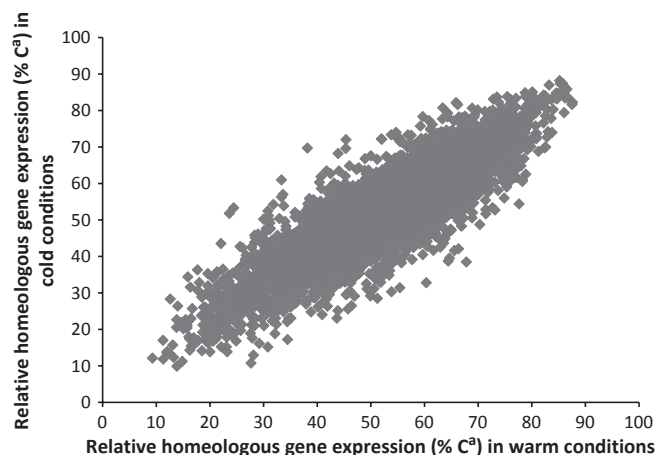


Fig. 3 Correlation between the relative homoeologous gene expression measured in *Coffea arabica* in the two growing conditions (warm and cold).

varied either toward C^a or E^a did not reveal statistically significant Go term enrichment.

Relationships between variations in relative homoeologous gene expression and total gene expression To look for a relationship between homoeologous gene expression and total gene expression, homoeologous gene expression and gene expression were investigated in 9959 and 10 628 genes in warmer and cooler growing conditions, respectively (Fig. S3). Whatever the growing conditions, no correlation was found between these two variables (for the warmer condition: $r = -0.01$, $P = 0.2988$, for the cooler condition: $r = -0.02$, $P = 0.03304$). In addition, variations in the level of gene expression ($\log_2$ fold change between the ratio of total expression in the warmer and cooler growing conditions) and of homoeologous gene expression between the same growing conditions (% C^a in the warmer condition vs % C^a in the cooler condition) were compared. A very low correlation coefficient

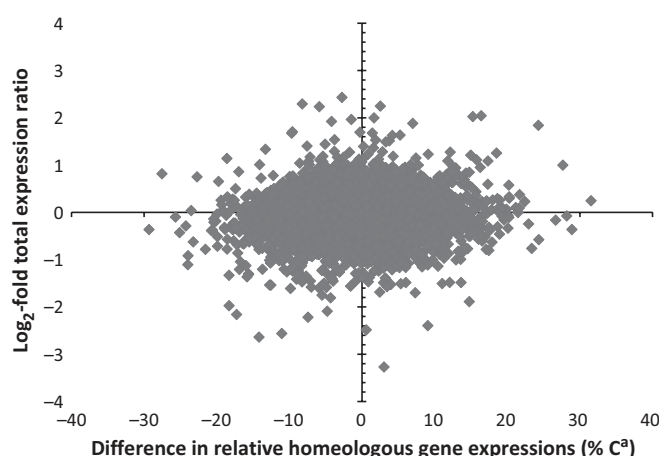


Fig. 4 Relationships between the relative homeologous gene expression and total expression variations observed in *Coffea arabica* in two growing conditions. For each *C. arabica* gene studied in two growing conditions, the difference in the relative homeologous gene expression (%C^a in warmer condition – %C^a in cooler condition) is plotted against the change in total expression (Log₂ expression in colder condition / expression in warmer condition) ($r = 0.06$, P -value = 2.135×10^{-7}).

($r = 0.06$, $P < 0.001$) was observed, suggesting that the variation in total gene expression is not dependent on the variation in relative homeologous gene expression (Fig. 4). No statistically significant GO term enrichment was revealed by gene ontology enrichment analysis performed on subsets of genes showing variations in the level gene expression and variations in the relative homeologous gene expression.

Homeologous gene expression and gene expression in the diploid progenitor species

Cis- and trans-regulatory homeologous gene expression In 1434 genes, we examined the *cis*- and *trans*-acting regulatory changes by plotting the ratio of gene expression of the diploid species (C/E) against the relative expression of homeologous genes in the allopolyploid (C^a/E^a) (Fig. 5). Amongst these genes, 32% and 14% showed *cis*-regulatory and *trans*-regulatory divergences, respectively, that occurred at a time between the two diploid species and the two subgenomes. Finally, 54% were affected by a combination of *cis*- or *trans*-regulatory divergences that occurred at a time between diploid species and both subgenomes. The last group was composed of 21% of genes showing a combination of *cis*- and *trans*-regulation divergences and 33% of genes for which it was difficult to statistically disentangle the *cis*- and *trans*-regulatory divergences.

Homeologous gene expression and patterns of differential expression between the allopolyploid and the diploid parental species The average percentage of the C^a homeolog in the subset of genes showing expression level dominance like the diploid species *C. eugenoides* was 50%, not significantly lower ($t = -1.7115$, $df = 57.632$, $P = 0.09237$) than 52%, the average percentage of homeologous C^a of all the genes studied. In the subset of genes showing expression level dominance like the

diploid species *C. canephora*, the average percentage of the C^a homeolog was 56%, significantly higher ($t = 8.0928$, $df = 523.093$, $P = 0.001$) than the average percentage of homeologous C^a of all the genes studied. Genes in the expression level dominance category did not present a particular type of regulation divergence (χ^2 test = 6.8686, $df = 3$, $P = 0.0762$).

Discussion

Homeologous gene expression throughout the genome

Even though RNA-seq analysis of polyploid plant transcriptomes is challenging, both the genome and the gene expression of polyploids have been successfully analyzed using strategies based on sequence similarity to determine the origin of the transcript (Buggs *et al.*, 2010b, 2012; Bombarely *et al.*, 2012; Higgins *et al.*, 2012; Ilut *et al.*, 2012). Here, we report results of genome-wide homeologous gene expression analysis of allopolyploid *C. arabica* plants in two contrasting growing conditions. First, the sequences were mapped using a transcriptome reference from one of its diploid progenitors; then, homeologous SNPs were identified by comparing allopolyploid SNPs with the allele sequences of its modern diploid parental species. Due to the use of a diploid transcriptome reference, this method was shown to induce a technical bias, but one which appeared to have no consequences for the relative homeologous gene expression estimates. For the first time, relative homeologous gene expression was assessed in 9959 pairs of homeologous genes in warm growing conditions (day : night temperatures: 30°C : 26°C) and in 10 628 pairs of homeologous genes in cold growing conditions (day : night temperatures: 23°C : 19°C). An earlier study analyzed homeologous gene expression in 2646 contigs of *C. arabica* by combining EST data from various libraries produced from seeds/berries, leaves and flowers in uncontrolled growing conditions (Vidal *et al.*, 2010). However in most cases, the number of ESTs that cover homeologous genes was too low to reliably assess relative homeologous gene expression.

The present characterization of the genome-wide homeologous gene expression of *C. arabica* complements previous studies performed in different allopolyploids (Buggs *et al.*, 2010b; Bombarely *et al.*, 2012; Higgins *et al.*, 2012; Ilut *et al.*, 2012). Whatever the growing conditions, 65% of genes studied here showed an equivalent level of homeologous gene expression, and 35% showed biased homeologous expression. Taking into consideration possible variations due to different methodological parameters between studies, the proportion of genes exhibiting biased homeologous gene expression is similar to that previously reported (i.e. 22%) for *C. arabica* by Vidal *et al.* (2010), as well as to that estimated for several natural allopolyploids: 25% in *Gossypium hirsutum* (Adams *et al.*, 2003), 27% in *Nicotiana tabacum* (Bombarely *et al.*, 2012) and 22% in *Tragopogon miscellus* (Buggs *et al.*, 2010b). But these estimates are higher than the 15% observed in synthetic *Brassica napus* allopolyploids (Higgins *et al.*, 2012). This discrepancy could be related to an increase in homeologous expression bias over evolutionary time as reported by Flagel & Wendel (2010). In some allopolyploids,

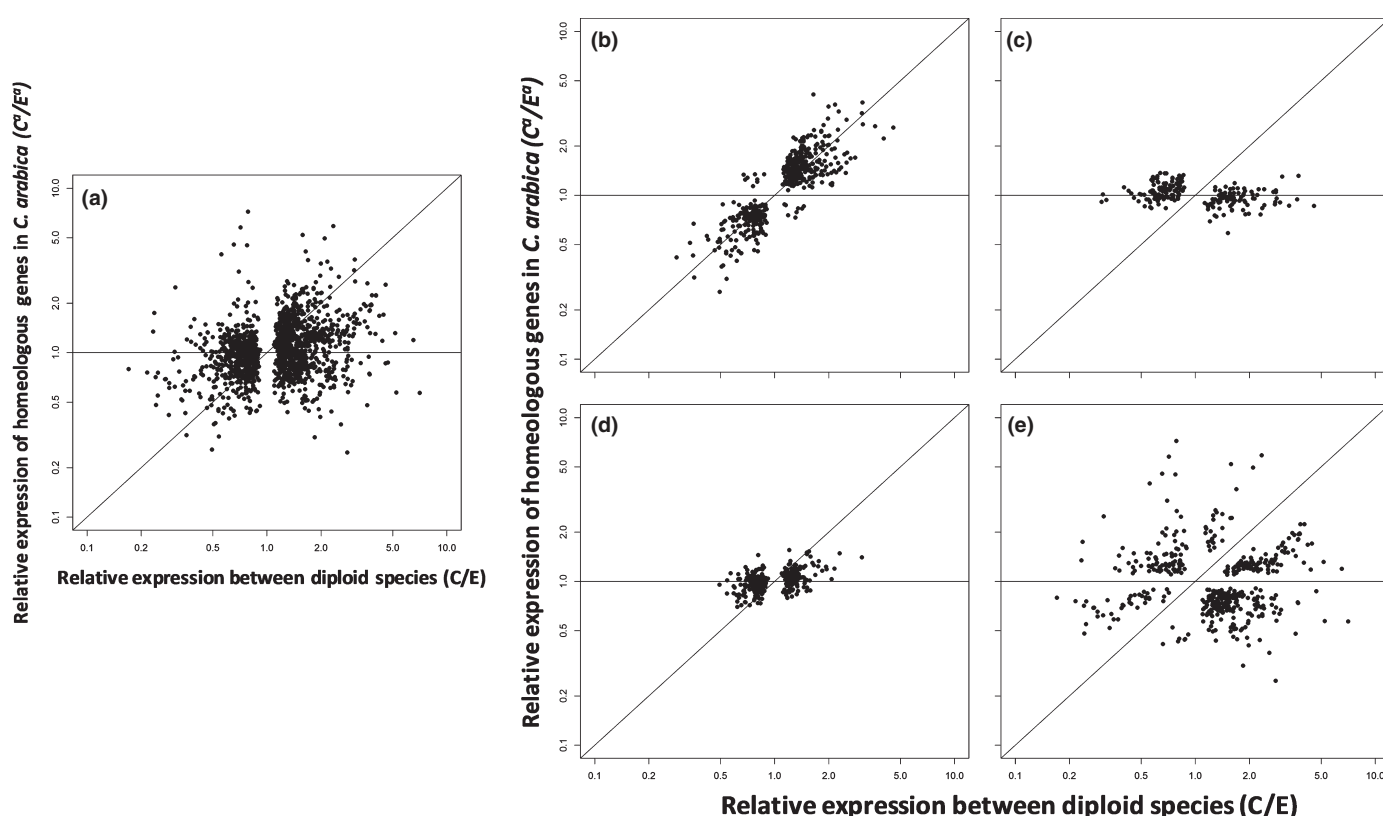


Fig. 5 Comparison of the relative homoeologous gene expression in *Coffea arabica* with the gene expression ratio between its diploid progenitor species. For differentially expressed genes of the diploid species, the ratio of gene expression of the diploid species (C/E) is plotted against the relative expression of homoeologous genes in the allopolyploid (C^3/E^3) on a logarithmic scale. (a) All genes studied are represented irrespective of the regulatory changes. Genes showing *cis*-regulatory divergences, *trans*-regulatory divergences, undetermined regulatory differences, and a combination of *cis* and *trans*-regulatory divergences are presented in (b), (c), (d), (e), respectively. The diagonal lines indicate 100% *cis*-regulatory divergence; the horizontal lines indicate 100% *trans*-regulatory divergence.

a subgenome can contribute preferentially to the transcriptome of the allopolyploid (i.e. unbalanced homoeolog expression bias). In *Glycine dolichocarpa*, Ilut *et al.* (2012) showed that the same subgenome was mostly affected by unbalanced homoeologous gene expression. In *Tragopogon miscellus* analyzed using a method based on a diploid transcriptome reference, a similar unbalanced homoeolog expression was reported; however, this result should be interpreted with caution due to a possible mapping-related technical bias, as mentioned by the authors (Buggs *et al.*, 2010b). In *C. arabica*, neither of the two subgenomes was preferentially expressed. In both growing conditions, the mean relative homoeologous gene expression was almost 50%.

Gene ontology enrichment analysis of biased homoeologous genes did not distinguish genes with specific functions preferentially expressed by one or other of the two subgenomes. However, when all genes exhibiting biased expression were considered together, gene ontology enrichment analysis did reveal slight over-representation of genes linked to 'enzymatic activity' and slight under-representation of genes linked to 'binding functions'. In *C. arabica*, a recent allopolyploid, observation of biased homoeologous gene expression may indicate a first step in an evolutionary process such as subfunctionalization. Indeed functional diversification of duplicated genes may be a major feature of the

long-term evolution of polyploids (Blanc & Wolfe, 2004). Even if most duplicate genes are rapidly lost at an evolutionary time-scale (Lynch & Conery, 2000), some homoeologous genes are retained depending on the dosage sensitivity of the gene and the gene function (Birchler *et al.*, 2005; Freeling & Thomas, 2006; Birchler & Veitia, 2007; Veitia *et al.*, 2008). In *Arabidopsis thaliana*, the transcription factors and signal transduction genes have been preferentially retained, which is not the case of other genes with enzymatic functions (Blanc & Wolfe, 2004).

Homoeologous gene expression regulation

In a new allopolyploid, both the level and the pattern of gene expression are affected by the merging of two divergent genomes of the parental diploid species in the same cell (Osborn *et al.*, 2003; Chen & Ni, 2006; Chen, 2007). Differences in gene expression between homoeologs can result from *cis*- and *trans*-regulatory changes. Several *cis*- and *trans*-regulatory models of homoeologous gene expression in allopolyploids have been proposed (Riddle & Birchler, 2003; Pignatta & Comai, 2009). *Cis*- and *trans*-regulatory changes have been distinguished in *C. arabica*, as well as in cotton by Chaudhary (Chaudhary *et al.*, 2009), and in *Arabidopsis* by Shi *et al.* (2012). Whereas in cotton

and *Arabidopsis*, predominant *cis*-regulatory effects mediating gene-expression divergence were observed, in *C. arabica*, the differential expression between homoeologous genes are explained exclusively by *cis*-regulatory differences in less than half the genes studied. In the remaining genes, the differential expression between homoeologous genes is explained by a combination of *cis*-regulatory effects and *trans*-regulatory effects. Therefore, because the two subgenomes of *C. arabica* display low sequence divergence and the variations in homoeologous gene expression are limited and consistent across growing conditions, it is likely that homoeologous *trans*-regulatory factors interact and control the regulation of the transcription of the two subgenomes. For most genes, overall expression appeared to be the result of two homoeolog expressions regulated by intertwined mechanisms. This high proportion of combined *cis*- and *trans*-regulatory divergences between diploid species and between the two subgenomes observed in *C. arabica* is congruent with the rapid radial speciation of *Coffea* species (Cros *et al.*, 1998; Anthony *et al.*, 2010). During adaptive radiation, rapid morphological diversification may be accompanied by accelerated rates of evolution of regulatory genes (Barrier *et al.*, 2001) in which *trans*-regulatory divergences between diploid species are more often observed (Wittkopp *et al.*, 2004).

In allopolyploids preferentially expressing one of the two subgenomes (i.e. unbalanced homoeolog expression bias), such as *Gossypium* allopolyploids and *Glycine dolichocarpa*, regulation models have been proposed which suggest that *cis*- and *trans*-regulatory divergences between subgenomes may play a role in the transgressive expression detected in allopolyploids in comparison with diploids (the level of expression in allopolyploids is higher or lower than in diploids; Rapp *et al.*, 2009). In studies comparing *Gossypium* allopolyploids and diploid F₁ hybrids with their diploid progenitors, it was also hypothesized that the genome preferentially expressed in allopolyploids is causally connected with the unbalanced expression level dominance of one parental species after genome merger, (Flagel & Wendel, 2010; Yoo *et al.*, 2012). In *C. arabica*, for genes previously classified in the expression level dominance category by a microarray study (Bardil *et al.*, 2011), we found no relationship between this characteristic and biased homoeologous gene expression or specific variations in *cis*- or *trans*-regulatory effects, perhaps because of intertwined regulation mechanisms between homoeologous genes.

Homoeologous gene expression in response to different growing conditions

Even though our understanding of gene expression in polyploids has increased in the last few years, the mechanisms that lead to increased responsiveness of polyploids in stressful environments remain to be investigated. The application of abiotic stress considerably impacts homoeologous gene expression in polyploids (Liu & Adams, 2007; Dong & Adams, 2011). The homoeologs in the allopolyploid could be used differentially for different or variable responses to an array of stressful conditions (Yoo *et al.*, 2012; Madlung, 2013). The present work was a whole genome

study of the variation in homoeologous gene expression in *C. arabica* in contrasted growing conditions that involve adaptation constraints for *C. arabica*. Unlike previously cited studies, neither growing conditions disrupted the relative homoeologous gene expression or produced harmful effects. Whatever the growing conditions, neither of the two subgenomes was preferentially expressed and no cases of gene silencing were observed, with the caveat that in our analytical conditions, only gene silencing that depended on growing conditions would be revealed. We did not observe relative homoeologous gene expression biased toward C^a in the warmer condition or toward E^a in the cooler condition. Furthermore, using gene ontology enrichment analysis, it was not possible to differentiate groups of genes showing a variation in biased homoeologous gene expression toward E^a or toward C^a. At the genome-wide scale, *C. arabica*'s ability to tolerate a broader temperature range than its diploid parents appears not to be related to the overall higher expression of the homoeologous genes deriving from the parental species best suited to one or the other temperature. Furthermore, for most of the genes, no relationship appears to link variation in total gene expression and variation in relative homoeologous gene expression. Whereas total gene expression changed, almost no change in relative homoeologous gene expression was observed. Given the intertwined regulation mechanisms between homoeologous genes, this outcome underlines the complexity of the regulation of homoeologous genes in a doubled network of interconnected genes and regulatory elements. To further advance our understanding of the adaptive abilities of polyploids, regulatory mechanisms of genes that simultaneously exhibit variation in total gene expression and in relative homoeologous gene expression could be investigated, and the relative homoeologous gene expression of specific genes on adaptation could be finely observed using ecological and controlled approaches. The analysis of post-transcriptional events can also be envisaged, as studies suggest that mRNA stability may be involved in the adaptation of polyploids to stress (Kim & Chen, 2011).

Taken together, the main results of the present study compared to those of previous studies on allopolyploids underline the particularities of *C. arabica*, a recent allopolyploid between two closely related diploid species. This outcome illustrates, on the one hand, the importance of considering the time of allopolyploidization and the evolution of progenitor species when interpreting the results of analyses; and, on the other, the need to examine multiple allopolyploid species in order to fully understand the mechanisms that lead to the adaptive abilities of allopolyploids.

References

- Adams KL. 2007. Evolution of duplicate gene expression in polyploid and hybrid plants. *Journal of Heredity* **98**: 136–141.
- 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, USA* **100**: 4649–4654.
- Adams KL, Percifield R, Wendel JF. 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* **168**: 2217–2226.

- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* 11: R106.
- Anthony F, Diniz LEC, Combes MC, Lashermes P. 2010. Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. *Plant Systematics and Evolution* 285: 51–64.
- Bardil A, Dantas de Almeida J, Combes MC, Lashermes P, Bertrand B. 2011. Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytologist* 192: 760–774.
- Barrier M, Robichaux RH, Purugganan MD. 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the National Academy of Sciences, USA* 98: 10 208–10 213.
- 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.
- Blanc G, Wolfe KH. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16: 1667–1678.
- Bombarely A, Edwards KD, Sanchez-Tamburino J, Mueller LA. 2012. Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: a phylogenomic perspective. *BMC Genomics* 13: 406.
- Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis DE, Barbazuk WB. 2010b. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Molecular Ecology* 19: 132–146.
- Buggs RJA, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, Soltis PS. 2010a. Tissue-specific silencing of homeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytologist* 186: 175–183.
- Buggs RJA, Renny-Byfield S, Chester M, Jordan-Thaden IE, Facio Viccini L, Chamala S, Leitch AR, Schnable PS, Barbazuk WB, Soltis PS *et al.* 2012. Next-generation sequencing and genome evolution in allopolyploids. *American Journal of Botany* 99: 372–382.
- Buggs RJA, Zhang L, Miles N, Tate JA, Gao L, Wu W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE. 2011. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Current Biology* 21: 551–556.
- Cenci A, Combes MC, Lashermes P. 2012. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant Molecular Biology* 78: 135–145.
- 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. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* 58: 377–406.
- Chen ZJ, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28: 240–252.
- Comai L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* 6: 836–846.
- Combes MC, Cenci A, Baraille H, Bertrand B, Lashermes P. 2012. Homeologous gene expression in response to growing temperature in a recent allopolyploid (*Coffea arabica* L.). *Journal of Heredity* 103: 36–46.
- Conesa A, Götz S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008: 1–13.
- Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A, Lashermes P. 1998. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. *Molecular Phylogenetics and Evolution* 9: 109–117.
- DaMatta FM, Ramalho JDC. 2006. Impacts of drought and temperature stress on coffee physiology and production: a review. *Brazilian Journal of Plant Physiology* 18: 55–81.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P. 2006. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* 152: 465–512.
- 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.
- Flagel L, Udall J, Nettleton D, Wendel JF. 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biology* 6: 16.
- Flagel LE, Chen L, Chaudhary B, Wendel JF. 2009. Coordinated and fine-scale control of homoeologous gene expression in allotetraploid cotton. *Journal of Heredity* 100: 487–490.
- Flagel LE, Wendel JF. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytologist* 186: 184–193.
- Freeling M, Thomas BC. 2006. Gene-balanced duplication, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Research* 16: 805–814.
- Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytologist* 196: 966–971.
- Hegarty MJ. 2011. Hybridization: expressing yourself in a crowd. *Current Biology* 21: R254–R255.
- 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 L, 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.
- Higgins J, Magusin A, Trick M, Fraser F, Bancroft I. 2012. Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. *BMC Genomics* 13: 247.
- Hovav R, Udall J, Chaudhary B, Rapp R, Flagel L, Wendel JF. 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proceedings of the National Academy of Sciences, USA* 105: 6191–6195.
- Ilut DC, Coate JE, Luciano AK, Owens TG, May GD, Farmer A, Doyle JJ. 2012. A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. *American Journal of Botany* 99: 383–396.
- Jackson S, Chen ZJ. 2010. Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* 13: 153–159.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.
- Kaur S, Francki MG, Forster JW. 2012. Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. *Plant Biotechnology Journal* 10: 125–138.
- Kim E-D, Chen ZJ. 2011. Unstable transcripts in *Arabidopsis* allotetraploids are associated with nonadditive gene expression in response to abiotic and biotic stresses. *PLoS ONE* 6: e24251.
- Lashermes P. 2000. Single-locus inheritance in the allotetraploid *Coffea arabica* L. and interspecific hybrid *C. arabica* × *C. canephora*. *Journal of Heredity* 91: 81–85.
- 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 & 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.
- 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. *Theoretical and Applied Genetics* 94: 947–955.
- Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* 320: 481–483.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

- Liu Z, Adams KL. 2007. Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Current Biology* 17: 1669–1674.
- Lynch M, Conery J. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.
- Madlung A. 2013. Polyploidy and its effects on evolutionary success: old questions revisited with new tools. *Heredity* 110: 99–104.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al.* 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.
- Mochida K, Yamazaki Y, Ogihara Y. 2003. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Molecular Genetics and Genomics* 270: 371–377.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5: 621–628.
- Nomura T, Ishihara A, Yanagita RC, Endo TR, Iwamura H. 2005. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proceedings of the National Academy of Sciences, USA* 102: 16 490–16 495.
- 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 in Genetics* 19: 141–147.
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.
- Pignatta D, Comai L. 2009. Parental squabbles and genome expression: lessons from the polyploids. *Journal of Biology* 8: 43.
- Rapp RA, Udall JA, Wendel JF. 2009. Genomic expression dominance in allopolyploids. *BMC Biology* 7: 18.
- Riddle NC, Birchler JA. 2003. Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. *Trends in Genetics* 19: 597–600.
- Shi X, Ng DWK, Zhang C, Comai L, Ye W, Chen ZJ. 2012. *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nature Communications* 3: 950.
- Stamati K, Mackay I, Powell W. 2009. A quantitative genomic imbalance gene expression assay in a hexaploid species: wheat (*Triticum aestivum*). *Genome/ National Research Council Canada* 52: 89–94.
- Veitia RA, Bottani S, Birchler JA. 2008. Cellular reaction to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* 24: 390–397.
- Vidal RO, Mondego JMC, Pot D, Ambrosio AB, Andrade AC, Pereira LFP, Colombo CA, Vieira LGE, Carazzolle MF, Pereira GAG. 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 J, 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.
- Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis*- and *trans*-gene regulation. *Nature* 430: 85–88.
- Yoo MJ, Szadkowski E, Wendel JF. 2012. Homeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110: 171–180.
- Yu Q, Guyot R, de Kochko A, Byers A, Navajas-Pérez R, Langston BJ, Dubreuil-Tranchant C, Paterson AH, Poncet V, Nagai C *et al.* 2011. Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allopolyploid coffee species (*Coffea*). *Plant Journal* 67: 305–317.

Supporting Information

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

Fig. S1 Relative homoeologous gene expression distribution after mapping against different reference transcriptomes.

Fig. S2 Correlation between the relative homoeologous gene expressions estimated after mapping using both reference transcriptomes.

Fig. S3 Relationships between relative homoeologous gene expression and total gene expression in two growing conditions.

Table S1 Gene ontology enrichment analysis for genes showing biased homoeologous gene expression toward E^a or C^a

Table S2 Gene ontology enrichment analysis for genes showing > 10% of variation in homoeologous gene expression between the two growth conditions

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 4

Discussion générale

CHAPITRE 4 : DISCUSSION GENERALE

Que les études portent sur les hybrides interspécifiques ou les allopolyploïdes, les résultats des travaux présentés dans les chapitres précédents peuvent être interprétés et discutés conjointement. Ils fournissent de nouvelles connaissances au sujet de l'expression et de la régulation des gènes après un évènement d'hybridation ancien ou récent et permettent d'initier l'étude des capacités adaptatives des plantes allopolyploïdes par rapport à leur environnement.

1 Synthèse des principaux résultats

1.1 Caractérisation de l'expression des gènes après un évènement d'hybridation

Bien que pour les deux études, l'expression des gènes ait été analysée en suivant la même stratégie qui repose sur l'évaluation de la contribution des allèles ou des gènes homéologues à l'expression totale d'un gène considéré (Figure 1 Combes et al. 2013, Figure S1 Combes et al. 2015), les données de séquençage d'ARNm n'ont pas été analysées avec la même référence transcriptomique. Aussi dans une analyse préliminaire, les données concernant l'étude de *C. arabica* var. Caturra ont été traitées une seconde fois en utilisant la référence de 25574 CDS provenant du séquençage du génome de *C. canephora* comme référence transcriptomique (Denoëud et al. 2014). La Figure 36 présente la distribution des gènes dans les différentes catégories de l'expression relative des gènes homéologues. Combinés avec les résultats concernant les hybrides F1 réciproques CxE et ExC, ces résultats permettent d'une part de caractériser l'expression des allèles et des gènes homéologues après un évènement d'hybridation récent et ancien et d'autre part de faire une ébauche de comparaison entre ces caractéristiques.

Dans les conditions méthodologiques des études, pour les hybrides F1 réciproques CxE, ExC et l'allopolyploïde naturel, les deux allèles ou gènes homéologues sont apparus toujours exprimés, parfois de façon biaisée en faveur d'un allèle (environ 30% des gènes) ou d'un sous-génome parental (environ 35% des gènes) et beaucoup plus rarement de façon très biaisée (autour de 1.5% des gènes pour les hybrides, moins de 0.5% des gènes pour les allopolyploïdes). Au sein des hybrides, le biais

d'expression vers un allèle ou un sous-génome n'est pas plus dirigé vers le parent *C. canephora* que vers le parent *C. eugenioides*, en revanche pour les allopolyploïdes un léger biais vers les gènes homéologues C^a est observé. La comparaison de la distribution des gènes dans les différentes catégories de l'expression relative indique des différences significatives entre les hybrides F1 et l'allopolyploïde naturel (chisq.test, $P < 0.01$). Pour ce dernier, plus de gènes présentent un biais vers l'homéologue C^a et moins de gènes présentent une expression relative équilibrée (Figure 36). Bien que les moyennes de l'expression relative des gènes homéologues C^a (52 %) et des allèles C_{hyb} (50%) soient proches, elles sont significativement différentes (t.test, $P < 0.01$).

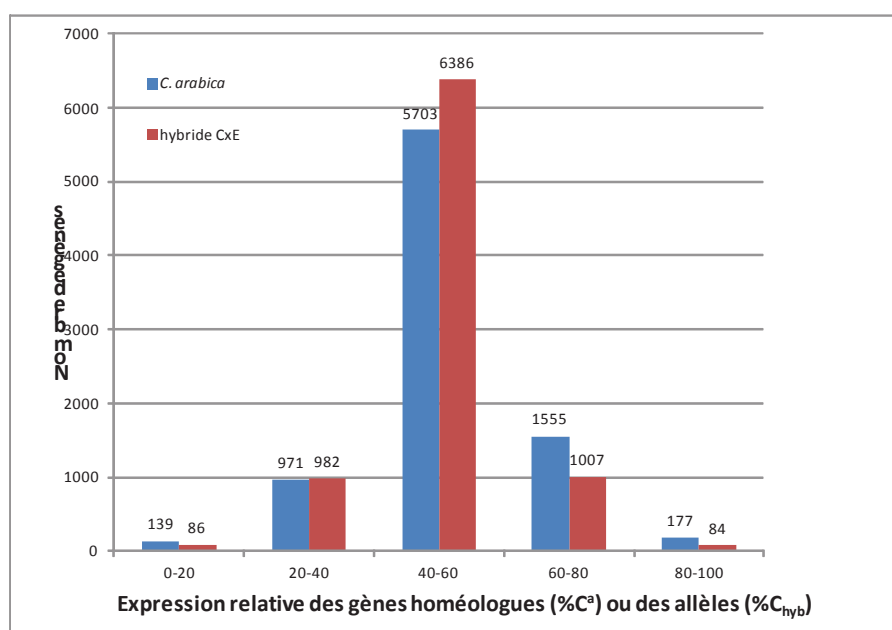


Figure 36 : Comparaison de la distribution des gènes dans les catégories de l'expression relative des gènes homéologues ou des allèles.

1.2 Hérité de l'expression des gènes dans les hybrides

La comparaison de l'expression des gènes des hybrides avec celles des espèces parentales a permis de révéler qu'au sein des hybrides interspécifiques, 30% des gènes analysés présentaient des variations d'expression par rapport aux espèces parentales. Même si toutes les formes de variation de l'expression des gènes (additivité, sur et sous dominance, sur et sous transgressivité) sont observées, une majorité de gènes (plus des $\frac{3}{4}$ des gènes différemment exprimés) montre un phénomène de « level expression dominance », c'est-à-dire que le niveau d'expression des gènes au sein des hybrides est comparable à celui de l'une des espèces parentales. Parmi ces gènes, 75% sont exprimés comme le parent *C. eugenioides*, il s'agit du phénomène de « level expression dominance » vers *C. eugenioides*.

1.3 Régulation de l'expression des allèles et des homéologues

1.3.1 *Cis*- et *trans*-régulation des allèles et des homéologues

Pour caractériser et étudier les divergences d'expression entre les allèles au sein des hybrides interspécifiques, l'expression des gènes a été au préalable comparée entre les espèces parentales. Sur les 21052 gènes annotés sur le génome de *C. canephora*, 33% présentent des différences d'expression entre *C. eugenioides* et *C. canephora*. Parmi ces gènes, 59% montrent des niveaux d'expression plus élevés pour *C. eugenioides* que pour *C. canephora*.

Au sein d'hybrides, l'étude de l'expression spécifique des allèles informe à la fois sur les divergences d'expression des espèces parentales (divergences de *cis*-régulation) mais aussi sur l'expression des allèles après l'hybridation. En effet les allèles partagent dans la cellule un environnement de régulation de l'expression qui combine les différences de *trans*-régulation des deux espèces parentales.

L'étude portant sur les hybrides interspécifiques a révélé que la variation de l'expression des gènes au sein des hybrides par rapport aux espèces parentales est due pour 15% à des divergences de *cis*-régulation entre les espèces parentales, pour 18% à des divergences de *trans*-régulation et pour 17% à des divergences liées à la fois à des différences de *cis*- et de *trans*-régulation. Ces résultats témoignent de la compatibilité des réseaux de régulation divergents des espèces parentales et de la régulation entrecroisée des allèles au sein des hybrides. L'expression spécifique des allèles est contrôlée par des facteurs de *trans*-régulation qui proviennent des deux génomes parentaux et qui peuvent sur ou sous-réguler simultanément l'un ou les deux allèles. Dans l'étude comparable menée sur les homéologues de *C. arabica*, une régulation entrecroisée des gènes homéologues a été aussi mise en évidence.

1.3.2 Effets des facteurs de *trans*-régulation sur l'expression des allèles

Pour les gènes présentant des divergences de *trans*-régulation entre les espèces parentales, une étude de la variation de l'expression de chacun des allèles entre les espèces parentales et les hybrides interspécifiques a permis de caractériser les effets des facteurs de *trans*-régulation. Puisque pour un allèle donné, les facteurs de *cis*-régulation spécifique des génomes sont identiques dans le contexte parental et hybride, la variation d'expression d'un allèle d'un génome parental considéré est due à l'action des facteurs de *trans*-régulation provenant de l'autre génome parental. Cette analyse a révélé que les facteurs de *trans*-régulation de *C. eugenioides* et *C. canephora* présentent des effets asymétriques sur la régulation des allèles. Alors que les allèles de *C. eugenioides* sont sur et sous-régulés par les facteurs de *trans*-régulation de *C. canephora*, les allèles de *C. canephora* sont principalement sur-régulés par les facteurs de *trans*-régulation de *C. eugenioides*.

1.3.3 Relation entre la régulation de l'expression spécifique des allèles et l'hérédité d'expression des gènes chez les hybrides

Pour les gènes présentant des divergences de *trans*-régulation entre les espèces parentales, l'analyse conjointe des effets des facteurs de *trans*-régulation et de l'hérédité de l'expression a permis de révéler que les catégories d'hérédité d'expression des gènes dépendent des combinaisons particulières des facteurs de *trans*-régulation sur la sur ou sous-régulation des deux allèles. Les phénomènes de « level expression dominance » et notamment celui dirigé vers *C. eugenioides* est attribué aux effets asymétriques des facteurs de *trans*-régulation des espèces parentales.

1.4 Variation de l'expression relative des homéologues en fonction de la température

Dans le cadre de l'étude des capacités adaptatives des polyploïdes, la variation de l'expression relative des gènes homéologues en fonction de la température de culture des plants de *C. arabica* a été analysée. Quel que soit la température (température du jour et de la nuit pour les conditions chaudes : 30°C et 26 °C et pour les conditions froides : 23°C et 19°C), l'expression relative des gènes homéologues ne semble pas corrélée à l'expression totale des gènes. Même si les espèces parentales sont endémiques de régions pour lesquelles les températures annuelles moyennes varient de 18°C à 23°C pour *C. eugenioides* et de 22° à 26°C pour *C. canephora*, l'expression relative des gènes à l'échelle du génome, n'est pas plus biaisée vers l'homéologue provenant de *C. eugenioides* à température froide que vers l'homéologue provenant de *C. canephora* à température chaude. Pour 92% des gènes analysés, l'expression relative des gènes homéologues varie de moins de 10% entre les deux températures. Comme précédemment il n'y a pas de corrélation directe entre la variation de l'expression totale des gènes et la variation de l'expression relative des gènes homéologues, ce qui aurait indiqué un mécanisme simple de régulation des gènes homéologues, faisant intervenir peu d'intermédiaires. De plus, il faut noter pour les gènes dont l'expression totale varie, que l'expression relative des gènes homéologues reste stable, ce qui signifie que l'expression des gènes homéologues est régulée simultanément et dans le même sens pour les deux homéologues. De même, lorsque l'expression relative des gènes homéologues varie, l'expression totale n'est pas forcément modifiée.

2 Discussions des résultats et présentation d'hypothèses

La possibilité d'évaluer indépendamment l'expression de chaque homéologue au sein de polyploïdes, ou de chaque allèle au sein d'hybrides a ouvert de nouvelles perspectives d'études des effets de l'hybridation des génomes sur l'expression des gènes (McManus et al. 2010, Shi et al. 2012, Bell et al. 2013, Xu et al. 2014). Pour les hybrides, après la fusion des génomes et de leurs réseaux de régulation de gènes, les modifications subies par chacun d'eux peuvent être dissociées. En revanche pour les allopolyploïdes, les effets de l'évolution depuis l'évènement de polyploïdisation sont mêlés à ceux de l'hybridation et de la duplication. De nouveaux enseignements concernant l'expression et la régulation des allèles ou des homéologues ont été révélés après l'application d'une telle analyse d'une part à des hybrides interspécifiques F1 entre *C. canephora* et *C. eugenioides* et d'autre part à *C. arabica*.

2.1 Caractérisation globale de l'expression relative des génomes après un évènement d'hybridation

Dans le cadre de l'étude de l'allopolyploïdie, il est considéré que l'hybridation et la duplication peuvent induire un choc transcriptomique. Ces évènements ont été analysés en comparant l'expression des gènes d'hybrides et de polyploïdes synthétiques par rapport à celle des espèces parentales mais aussi par rapport à celle de polyploïdes naturels pour prendre en compte les effets de l'évolution sur ces évènements. Selon les modèles polyploïdes considérés, les résultats des observations conduisent parfois à des conclusions variables. Pour *Arabidopsis*, les effets de l'hybridation sur l'expression des gènes semblent plus importants que ceux de la duplication (Wang, Tian, Lee, Wei, et al. 2006). En revanche pour des hybrides intraspécifiques et des polyploïdes synthétiques et réciproques du genre *Oryza*, les effets de la duplication sont apparus beaucoup plus importants que ceux de l'hybridation. Cette étude a aussi révélé que le sens de l'hybridation n'avait pas d'incidence particulière sur la variation de l'expression des gènes (Xu et al. 2014). Pour *Senecio*, la comparaison de l'expression des gènes entre des hybrides triploïdes et des allopolyploïdes hexaploïdes indiquent que la duplication liée à l'allopolyploïdisation semblent diminuer les différences d'expression des gènes provoquées par l'hybridation (Hegarty et al. 2006). Pour *Tragopogon*, Buggs et al. (2011) ont observé que le profil d'expression des gènes résultant du choc transcriptomique évoluait au cours des générations. Ils ont suggéré que ces changements génèrent au sein des allopolyploïdes de nouveaux phénotypes qui affectent l'avantage évolutif des allopolyploïdes (Buggs et al. 2011). Pour *Spartina*, des effets de l'hybridation sur l'expression des gènes (notamment une dominance maternelle) observés au sein d'hybrides sont moins importants au sein d'allopolyploïdes suite à la duplication (Chelaifa et al. 2010). Alors que dans les études citées

précédemment, les perturbations de l'expression totale ont été observées, les effets de l'hybridation ont été aussi examinés au travers de l'analyse de l'expression des gènes homéologues. Pour le coton, la modulation de l'expression des gènes homéologues au sein d'allopolyploïdes résultent des conséquences de l'hybridation des génomes et des effets de l'hybridation, mais aussi des changements dus à l'évolution sur les effets de l'hybridation, notamment le biais d'expression vers un sous-génome augmentent au cours du temps (Flagel et al. 2008). Plus récemment, Yoo et al. (2012) ont confirmé ces résultats en comparant l'expression d'environ 2000 gènes homéologues au sein d'un hybride interspécifique F1 et d'allopolyploïdes naturels et synthétiques. Ils ont observé que l'expression relative des gènes homéologues était biaisée vers un sous-génome parental dès l'hybridation pour les hybrides F1 diploïdes, vers le même sous-génome parental mais de façon plus biaisée pour les allopolyploïdes naturels et vers le sous-génome opposé pour les polyploïdes synthétiques. Pour le coton, la sélection naturelle aurait permis la conciliation de disparités de régulation des génomes parentaux après l'hybridation, au cours de l'évolution de nouvelles possibilités d'expression des gènes seraient générées (Yoo et al. 2012) comme cela a été suggéré dans le cas de *Tragopogon* (Buggs et al. 2011). Dans notre cas, en ce qui concerne le sens de l'hybridation, nos résultats sont comparables à ceux obtenus sur les hybrides intra-spécifiques de riz (Xu et al. 2014), très peu de gènes (0,65% des 21052 gènes analysés) étaient différemment exprimés entre les deux groupes d'hybrides CxE et ExC. Si l'on compare les distributions de l'expression relative des allèles dans les hybrides F1 et des homéologues dans les allopolyploïdes, on constate que pour les allopolyploïdes, plus de gènes (environ 8%) présentent un biais d'expression supérieur à 20% dirigé vers C^a par rapport aux hybrides. Yoo et al. (2012) ont aussi observé une accentuation du biais d'expression des homéologues vers un parent entre des hybrides diploïdes synthétiques et des allopolyploïdes naturels (Yoo et al. 2012). Sur la base de nos résultats préliminaires, des événements qui présenteraient des similarités entre le coton et le caféier, pourraient s'être produit au cours de l'évolution au sein des deux allopolyploïdes. L'évolution depuis l'évènement d'hybridation, et peut être aussi l'état de duplication (il n'est pas possible, avec cette simple analyse, de distinguer les deux) de *C. arabica* pourraient avoir eu un effet sur l'expression relative des homéologues. Dans notre cas, pour valider cette hypothèse, des analyses supplémentaires sont nécessaires. La légère différence entre les moyennes de l'expression relative des gènes homéologues (52%) et des allèles (50%) entraîne à considérer la variabilité de l'expression relative des gènes homéologues au sein de génotypes de *C. arabica* et de l'expression relative des allèles au sein de plusieurs hybrides. Pour déterminer l'impact de la duplication, notre étude devrait être complétée par une analyse de l'expression des gènes homéologues au sein d'allopolyploïdes synthétiques.

2.2 L'héritage parental et la régulation *cis* versus *trans*

Buggs et al. (2014) rapportent que très tôt dans l'histoire de l'étude de la polyploïdie, même si la technologie ne permettait d'étudier que quelques gènes, les chercheurs se sont intéressées à définir dans l'expression de l'allopolyplœide, l'importance des différences d'expression des gènes qui provenaient de l'héritage de divergences d'expression déjà existantes entre les espèces parentales diploïdes (Buggs et al. 2014). Actuellement un moyen pour aborder cette question est de considérer à l'échelle du génome l'expression des gènes d'un hybride ou d'un allopolyplœide en distinguant les caractéristiques de *cis*- et *trans*-régulation des gènes. Ce type d'analyse a déjà été menée sur des allopolyplœides de synthèse (Shi et al. 2012, Yoo et al. 2012). Ces études ont montré que l'expression divergente des gènes au sein d'allopolyplœides de synthèse provenait de divergences de *cis*-régulation et *trans*-régulation des espèces parentales. En menant cette étude sur les hybrides interspécifiques F1 *Coffea*, couplée à une analyse de l'hérédité de l'expression des gènes, quelques éléments de compréhension viennent compléter les connaissances de l'expression des gènes dans les allopolyplœides.

Tout d'abord, au sujet de l'hérédité de l'expression des gènes, les hybrides F1 *Coffea* montrent un phénomène d'« expression level dominance » qui est un effet de l'hybridation partagé par d'autres modèles d'hybrides ou allopolyplœides (McManus et al. 2010, Qi et al. 2012, Yoo et al. 2012, Bell et al. 2013). Ensuite pour l'étude des divergences de *cis*- et *trans*-régulation entre les espèces parentales, les hybrides *Coffea*, se distinguent d'autres modèles d'hybrides par la proportion importante de gènes présentant des variations d'expression dues à des divergences de *trans*-régulation entre les espèces parentales. Dans ce contexte, les effets des facteurs de *trans*-régulation ont été étudiés et une asymétrie des effets des facteurs de *trans*-régulation des génomes parentaux a été révélée. En faisant pour chaque gène, une analyse conjointe de sa catégorie d'expression et de la régulation de son expression allélique par les facteurs de *trans*-régulation de chacun des génomes, il a été possible d'établir une relation entre l'hérédité d'expression des gènes et la régulation de l'expression allélique par les facteurs de *trans*-régulation. Ce résultat vient confirmer le travail de Yoo et al. (2013) qui ont observé pour le coton, que les gènes caractérisés par le phénomène d'« expression level dominance » vers le parent le plus ou le moins exprimé, présentaient une sur ou sous-régulation de l'allèle du parent non-dominant. Pour Buggs et al. (2014) le phénomène d'« expression level dominance » observé chez *Arabidopsis* et *Tragopogon* proviendrait également de la régulation des gènes par des effets de *trans*-régulation (Buggs et al. 2014). Nos résultats complètent cette conclusion en démontrant que pour toutes les catégories de l'hérédité de l'expression des gènes, la combinaison des différences des effets des facteurs de *trans*-régulation des espèces parentales déterminent la variation d'expression du gène. Au sujet de l'asymétrie des facteurs de *trans*-régulation, Shi et al. (2012) ont déjà observé qu'au sein d'allopolyplœides synthétiques entre *A. thaliana* et *A. arenosa*, les facteurs de *trans*-régulation de *A. arenosa* avaient un effet plus fort sur la variation de l'expression allélique que les facteurs de *trans*-

régulation d'*A. thaliana* (Shi et al. 2012). Si entre ces deux espèces une dominance d'expression des gènes d'*A. arenosa* par rapport à *A. thaliana* avait déjà été rapportée (Wang, Tian, Lee, Wei, et al. 2006), l'asymétrie des facteurs de *trans*-régulation n'a pas encore été expliquée. Dans notre cas, la différence d'effets des facteurs de *trans*-régulation de *C. canephora* et *C. eugenioides* reste aussi sans explication, une différence de nature épigénétique des génomes de ces deux espèces pourrait expliquer la différence d'efficacité des facteurs de *trans*-régulation. Au sein d'hybrides entre *A. thaliana* et *A. lyrata*, la différence de méthylation des deux génomes expliquerait des différences de silencing de gènes entre les deux génomes (He et al. 2012).

De façon générale, la difficulté à retirer de l'observation des multiples modèles d'hybrides décrits, un enseignement général concernant la régulation des gènes vient du fait que chaque hybride est un cas particulier qui concilie au sein de ses cellules les différences de régulation de gènes de ses espèces parentales et que chacune d'elles à sa propre histoire évolutive. Les caractéristiques d'hérédité d'expression des gènes au sein des hybrides dépendent de la divergence génétique entre les espèces parentales. Néanmoins au sein d'allopolypléïdes la régulation de l'expression des gènes a pu changer au cours de l'évolution et ne serait plus simplement l'héritage des systèmes de régulation des espèces parentales (Buggs et al. 2014). D'une part les contraintes résultant de l'hybridation, et de la duplication des génomes pourraient au cours du temps modifier la régulation de l'expression des gènes. Des différences d'expression liées à des divergences de *cis*-régulation entre les espèces parentales pourraient plus probablement être maintenues. Par contre des différences d'expression liées à des divergences de *trans*-régulation pourraient plus évoluer en raison de pression de sélection due à la coévolution moléculaire des réseaux de transcription (Landry et al. 2007). Mais d'autre part, plus les génomes des espèces parentales d'un allopolypléïde sont proches, plus il est probable que les facteurs de transcription d'un génome soient compatibles avec les sites de fixation des facteurs de transcription de l'autre génome (Buggs et al. 2014) et ces compatibilités contraindraient moins la variation de la régulation de l'expression des gènes que dans le cas inverse. Les conditions environnementales sont aussi des contraintes qui pourraient faire varier la régulation de l'expression des gènes. Dans une étude de la variation de la régulation de l'expression des gènes chez la levure, la fréquence des différences de *trans*-régulation varie selon les conditions de culture des levures et ces différences de régulation liées à des facteurs *trans* seraient à l'origine du potentiel de variation de l'expression des gènes (Tirosh et al. 2009). Dans cette même étude, les auteurs étudient l'hypothèse de deux types de facteurs de *trans*-régulation : les *trans*-régulateurs qui seraient à l'origine de différences de la transcription et des régulateurs de la chromatine et les *trans*-sensoriels qui seraient des composés en amont de la transcription qui transmettraient des signaux internes ou environnementaux aux *trans*-régulateurs. Tirosh et al. (2009) démontrent que chez la levure, les facteurs de *trans*-régulation *trans*-sensoriels ont un rôle dominant parmi les gènes présentant des effets de facteurs de *trans*-régulation. Un pourcentage important de gènes ayant une réponse similaire aux variations environnementales, comparés deux à

deux, présentent une corrélation des divergences de *trans*-régulation. Ce résultat soutient l'hypothèse de facteurs de *trans*-régulation *trans*-sensoriels. En effet si les divergences de facteurs de *trans*-régulation étaient majoritairement dues à l'activité de facteurs *trans*-régulateurs, les gènes comparés deux à deux auraient divergé indépendamment et ils n'auraient pas montré de corrélation des divergences de *trans*-régulation. Pour ce modèle, la variation d'expression entre les espèces parentales liée à des divergences en *trans*-régulation est plus dépendante de leur interprétation différentielle des signaux sensoriels qu'à des mutations des régulateurs de la transcription (Tirosh et al. 2009). De telles conclusions sont à considérer dans notre cas pour interpréter et comprendre qu'une forte proportion de gènes avec des divergences de *trans*-régulation entre les espèces parentales soient observées pour les hybrides F1 *Coffea*. Les espèces parentales *C. eugenoides* et *C. canephora* génétiquement proches et adaptées à différents écosystèmes pourraient se différencier par une activité différentielle des facteurs *trans*-sensoriels et ainsi induire au sein des hybrides, pour certains gènes, une expression non-additive par rapport aux espèces parentales.

2.3 Variation faible de l'expression relative des gènes homéologues

La variation de l'expression relative des gènes homéologues a été étudiée à plusieurs reprises chez le coton, d'une part pour différents organes, d'abord pour quelques gènes (Adams et al. 2003, Adams et al. 2004, Flagel et al. 2009) ensuite à plus grande échelle (jusqu'à environ 1500 gènes) (Chaudhary et al. 2009, Flagel et Wendel 2010, Hovav et al. 2008) et d'autre part par rapport à l'application de stress abiotiques (Liu et Adams 2007, Dong et Adams 2011). Pour le blé, l'expression relative des gènes homéologues a été aussi observée en fonction des organes et en réponse à un stress à l'échelle de quelques gènes (Mochida et al. 2003, Nomura et al. 2005, Stamati et al. 2009). Toutes ces études rapportent des variations de la contribution des gènes homéologues à l'expression totale. Avec l'avancée technologique des méthodes de séquençage, l'expression relative des gènes homéologues a été explorée à l'échelle génomique pour plusieurs allopolyploïdes : *Brassica napus* (Higgins et al. 2012), *Glycine dolichocarpa* (Ilut et al. 2012) et *Nicotiana tabaccum* (Bombarely et al. 2012). Mais aucune de ces études n'inclut une observation de la variation du biais d'expression des homéologues en fonction de la température. Notre étude effectuée à l'échelle du génome sur *C. arabica* a permis de comparer l'expression relative des gènes homéologues après culture des plantes à deux températures. Nos résultats qui ont montré à l'échelle du génome que les deux homéologues étaient toujours exprimés ont aussi révélé une étonnante stabilité de la contribution des homéologues à l'expression totale des gènes en fonction de la température. Quelle que soit la température qui convenait plus à l'une ou à l'autre des espèces parentales de l'allopolyploïde, aucun biais favorisant les homéologues d'un génome n'a été noté. Ce résultat est assez cohérent avec les résultats obtenus lors de l'analyse de l'expression relative des gènes homéologues d'une région génomique de *C. arabica* (Combes et al. 2012). En effet le biais de la contribution des gènes homéologues en fonction des organes et de la

température présentaient des variations modérées et aucun biais extrême qui auraient pu suggérer que *C. arabica* présentaient des signes de néo ou sous-fonctionalisation.

Dans le contexte de l'explication de l'adaptabilité des allopolyploïdes par une variation de l'expression totale simplement régulé par la contribution des gènes homéologues, la faible variation de l'expression relative des gènes homéologues en fonction de la température a été un résultat surprenant. Pourtant maintenant, avec la connaissance actuelle de l'expression systématique des deux allèles et de la prédominance de facteurs de *trans*-régulation au sein d'hybrides F1 *coffea*, ce résultat est plus compréhensible. En effet, en fonction de la température, des variations de l'expression des homéologues peuvent se produire mais elles se produisent de façon simultanée sur les deux sous-génomes et ne provoquent pas de variation de l'expression relative des gènes homéologues. A l'échelle du génome entre les deux températures de culture, il n'y a pas de changement général de l'expression relative des gènes homéologues qui soit lié à une variation de l'expression totale des gènes, mais pourtant quelques gènes (environ 8% des gènes analysés) montrent une variation de l'expression relative supérieure à 10%. Dans le contexte de l'hypothèse des *trans*-sensoriels de Tirosh et al. (2009) qui décrit une cascade de facteurs de *trans*-régulation, un faible nombre de gènes qui présenteraient des variations de l'expression relative des homéologues pourrait avoir une incidence sur un nombre plus important de gènes, mais comme les gènes régulés par les facteurs de *trans*-régulation des deux sous-génomes sont très nombreux, cette incidence n'a que peu d'effet sur la variation générale de l'expression relative des homéologues et sur la variation générale de l'expression totale. Cet état de l'expression relative des gènes homéologues reflète la réponse de l'allopolyploïde à la variation de température. Quel que soit la température, *C. arabica* aurait l'avantage de disposer d'une paire d'homéologues pour laquelle les deux homéologues peuvent être exprimés puisque leur régulation est assurée par les facteurs de *trans*-régulation des deux sous-génomes et au pire de seulement un des deux sous-génomes.

Depuis la publication de nos travaux au sujet de la variation de l'expression relative des gènes homéologues en fonction de la température (Combes et al. 2013), une étude portant sur un allopolyploïde naturel *Arabidopsis kamchatica* a aussi rapporté la stabilité de la contribution des gènes homéologues. Dans cette étude, l'expression relative des gènes homéologues à l'échelle du génome a été comparée entre des plantes ayant subi ou non un stress (une culture à 4°C pendant 7 jours). L'expression relative des homéologues entre les deux traitements est fortement corrélée (Akama et al. 2014) et les auteurs concluent sur la similarité des sites de fixation des facteurs de transcription et de l'état épigénétique des homéologues.

2.4 Modèle de l'expression des gènes homéologues chez *C. arabica*

Dans le contexte maintenant connu pour *C. arabica* de l'expression systématique des deux gènes homéologues et de leur régulation entrecroisée par des facteurs de *trans*-régulation des deux sous-génomes, un modèle de la régulation de l'expression des gènes homéologues pour cet allopolyploïde inspiré de celui que Tirosh et al. (2009) ont présenté pour les hybrides de levure (Tirosh et al. 2009) peut être proposé (Figure 37). Dans ce modèle, des facteurs *trans*-sensoriels spécifiques de chaque espèce parentale seraient regroupés au sein de l'allopolyploïde. Ils percevraient des variations de l'environnement et transmettraient des signaux à des facteurs *trans*-régulateurs provenant des deux sous-génomes qui eux-mêmes agiraient sur les gènes homéologues. Ce modèle reflète bien l'importance de la régulation liée à des facteurs *trans* mise en évidence par l'étude des hybrides F1 et des allopolyploïdes, mais aussi la stabilité et l'équilibre (dans la majorité des cas) de l'expression relative des gènes homéologues même en cas de variation de la température. La réponse (disparition, diminution ou augmentation) de facteurs *trans*-sensoriels provenant d'une espèce parentale qui seraient sensibles à la température serait compensée par l'action des effets *trans*-sensoriels de l'autre parent et aussi par le maintien de l'activité des facteurs *trans*-régulateurs des deux parents. Enfin ce modèle combine la connaissance que l'on a maintenant du lien entre l'hérédité de l'expression des gènes et les mécanismes de *trans*-régulation des allèles (que l'on pourrait extrapoler aux gènes homéologues).

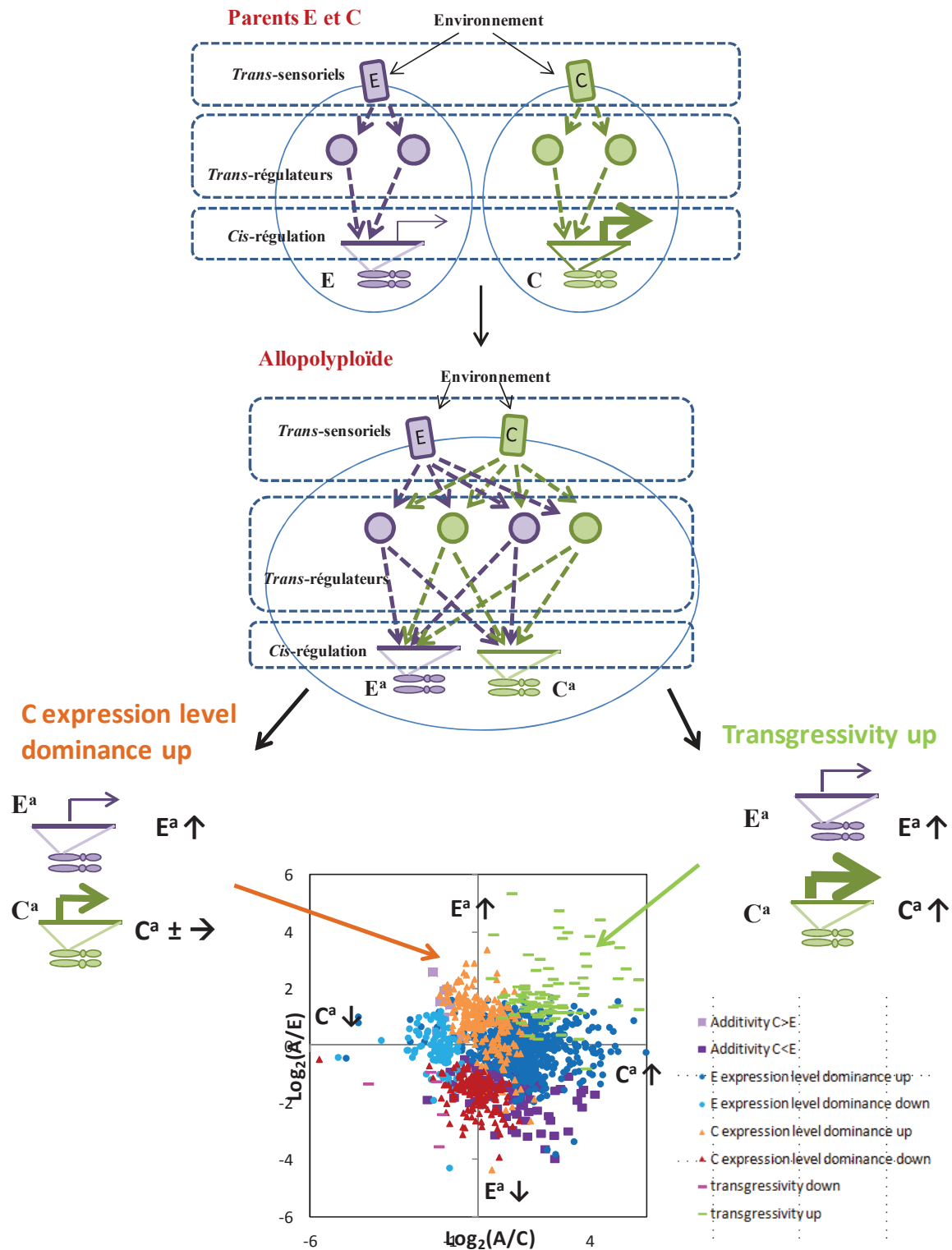


Figure 37 : Modèle de l'expression des gènes homéologues au sein d'un allopolyploïde *C. arabica*, lien entre hérédité de l'expression des gènes et la *trans*-régulation des gènes homéologues. La partie haute du schéma présente un exemple d'expression d'un gène pour les deux espèces parentales C et E. Au sein de l'allopolyploïde (partie intermédiaire du schéma), les deux gènes homéologues (C^a et E^a) s'expriment et sont régulés de façon entrecroisée par des facteurs de *trans*-régulation des deux sous-génomes. Enfin la partie basse du schéma présente deux exemples de variation de l'expression des gènes dans l'allopolyploïde par rapport aux espèces parentales. Les effets des facteurs de *trans*-régulation des deux sous-génomes sur la régulation de l'expression des gènes homéologues sont illustrés pour chaque catégorie d'hérédité de l'expression des gènes de l'allopolyploïde.

A partir de ce modèle, les résultats de Bardil et al. (2011) obtenus lors de l'étude de la variation de l'hérédité de l'expression des gènes de plantes de *C. arabica* cultivées à différentes températures (Bardil et al. 2011) pourraient être mieux appréhendés (Figure 38). Dans cette étude, l'hérédité de l'expression des gènes de *C. arabica* est caractérisée par le phénomène d'«expression level dominance» biaisé vers *C. eugenoides* lorsque les plantes sont cultivées à température moyenne pour le caféier, ce résultat est cohérent avec les observations effectués sur les hybrides F1 de *Coffea*. En revanche, lorsque les plantes sont cultivées à une température élevée, le biais d'«expression level dominance» s'inverse et est dirigé vers *C. canephora* et les effectifs des gènes dans les différentes catégories d'expression transgressive et additive diminuent fortement. En prenant l'hypothèse que le mécanisme qui relie les classes d'hérédité d'expression et l'action des facteurs de *trans*-régulation est similaire entre les hybrides et les allopolyploïdes, les gènes classés par Bardil et al. (2011) indiqueraient que cette classification («expression level dominance» dirigé vers *C. canephora* et augmentation du nombre de gènes sans changement d'expression entre les espèces parentales et l'allopolyploïdes) correspond à une absence de régulation des homéologues C^a par les facteurs de *trans*-régulation du génome E^a. A une température de culture extrême, élevée qui dépasse la température maximale tolérée par l'espèce *C. eugenoides*, les facteurs *trans*-sensoriels provenant du génome E^a ne seraient pas activés et la régulation des facteurs *trans*-régulateurs serait seulement assurée par les facteurs *trans*-sensoriels du sous-génome C^a. En revanche la régulation de l'expression des gènes homéologues serait maintenue par les facteurs *trans*-régulateurs des deux sous-génomes. Mais elle serait fortement perturbée par la diminution de l'action de facteurs *trans* de *C. eugenoides*.

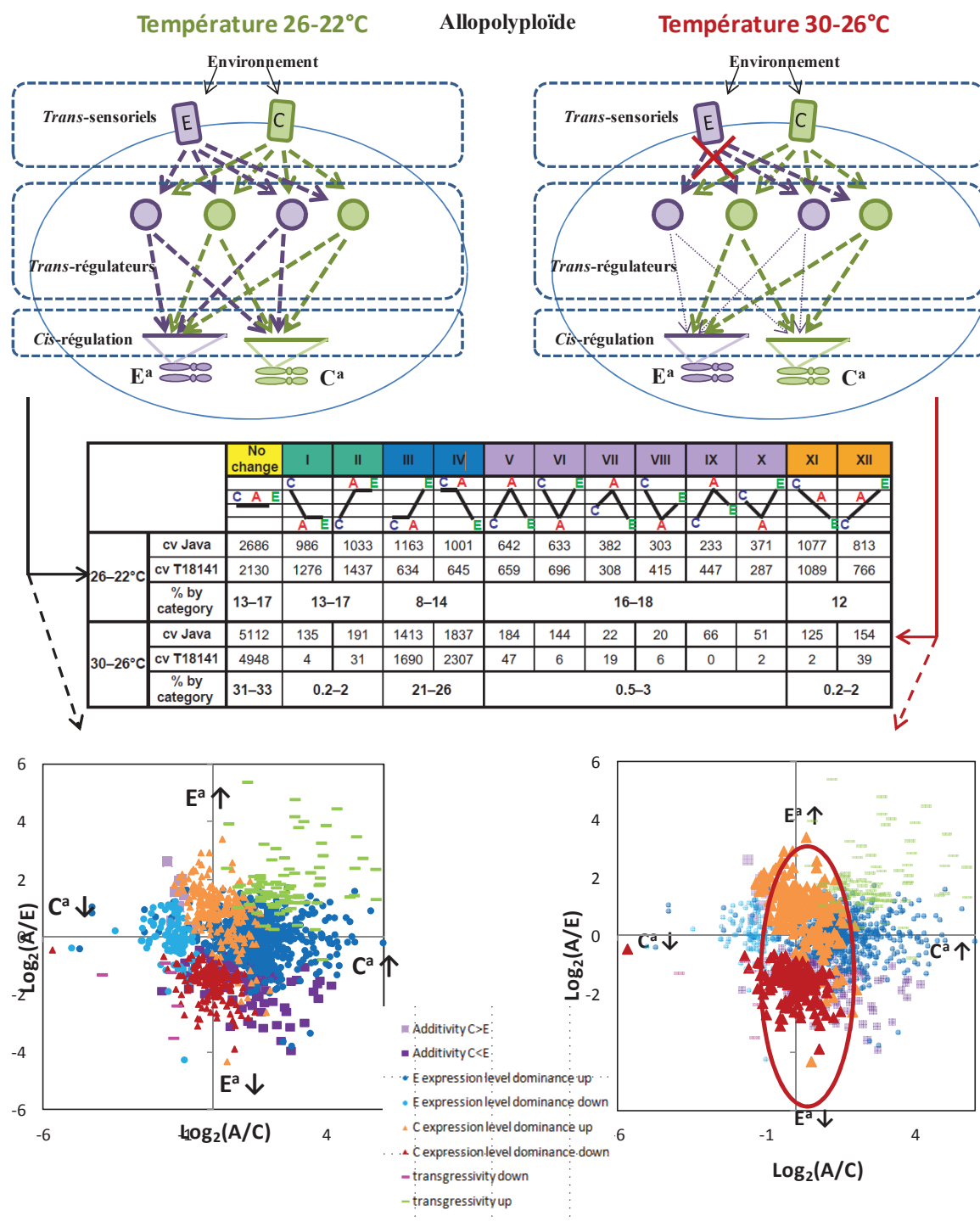


Figure 38 : Application du modèle aux données d'expression des gènes de *C. arabica* cultivé à des températures différentes (Bardil et al. 2011). La condition de culture à 30-26°C (condition extrême pour *C. eugenoides*) inactiverait les facteurs *trans*-sensoriels de E, la régulation de l'expression des gènes homéologues serait perturbée (partie haute du schéma). Le tableau provient de l'étude de Bardil et al. (2011), il présente le classement des gènes selon la catégorie d'hérédité de l'expression des gènes de l'allopolyploïde (deux génotypes, cv Java, et cv T18141) pour les deux températures de l'étude (26-22°C et 30-26°C). Toutes les catégories d'hérédité d'expression présentent une diminution de l'effectif sauf les catégories présentant le phénomène de « level expression dominance » dirigé vers C et aucune variation de l'expression par rapport aux espèces parentales. En reportant virtuellement les effectifs des catégories d'hérédité d'expression sur le diagramme de la variation de l'expression des gènes homéologues selon le lien proposé par le modèle, les catégories de « level expression dominance » dirigé vers C seraient majoritaires et témoigneraient de la diminution des effets des facteurs de *trans*-régulation provenant de E, les homéologues de C^a ne seraient que peu augmentés ou diminués (partie basse du schéma).

3 Perspectives

Les travaux menés sur *C. arabica* et des hybrides F1 de ses espèces parentales apportent de nouvelles connaissances concernant l'expression des gènes homéologues et des allèles après un événement d'hybridation, mais aussi de nouveaux éléments pour comprendre les mécanismes d'adaptation des plantes polyploïdes. En s'ajoutant aux travaux réalisés précédemment sur d'autres polyploïdes, ces résultats viennent compléter la vision actuelle des conséquences génétiques et génomiques de la polyploïdie. Pourtant de nouvelles questions apparaissent, très liées aux résultats obtenus lors de cette étude ou plus générales et complexes :

- Quelles sont les raisons de l'asymétrie des effets des facteurs de *trans*-régulation ? Et ont-elles un lien avec la diversité des espèces diploïdes ?
- L'adaptation des plantes à leur environnement après un événement d'hybridation ou de polyploïdisation dépend-elle du dynamisme de régulation de l'expression des gènes en fonction des conditions environnementales ?

Des études peuvent être proposées pour répondre à ces questions, mais elles doivent être replacées dans le contexte extrêmement complexe qui relie l'expression des gènes à un phénotype.

3.1 Explication de l'expression différentielle des espèces de *Coffea* par des différences génomiques

Pour répondre à la question concernant l'asymétrie des effets des facteurs de *trans*-régulation chez les hybrides F1 de *Coffea*, les différences génomiques entre les espèces *C. canephora* et *C. eugenioides* peuvent être considérées. Il a été évoqué à plusieurs reprises que ces espèces sont génétiquement peu divergentes, pourtant elles présentent des variations phénotypiques importantes et sont adaptées à différents écosystèmes. Les différences d'adaptation entre ces deux espèces pourraient être dues à des variations de plasticité phénotypique. En effet, la plasticité phénotypique est reconnue comme un moyen de modification adaptative d'un phénotype en réponse à des changements environnementaux (Schlichting et Wund 2014). Les biologistes de l'évolution portent un intérêt croissant à ce phénomène. Ils lui accordent un rôle grandissant dans les nouveaux modèles d'évolution car ce phénomène relie le génotype et le phénotype en modifiant l'expression des gènes. La plasticité peut dépendre de plusieurs causes, et parmi elles, les modifications épigénétiques. Même s'il est difficile d'assimiler à un même processus ces deux phénomènes, la plasticité et les modifications épigénétiques partagent la particularité de provoquer des changements du phénotype ou de l'expression des gènes sans changer la séquence de l'ADN (Schlichting et Wund 2014). Des variations épigénétiques peuvent être induites par l'environnement et transmises aux générations suivantes (Schlichting et Wund 2014, Jablonka et Lamb 2015). Ces variations se produiraient sur des loci dont

l'activité change en fonction de variations de l'environnement (Jablonka et Lamb 2015). De telles variations contribueraient à l'adaptation des plantes à de nouvelles niches écologiques et joueraient un rôle important dans la spéciation (Jablonka et Lamb 2015).

Parmi les variations épigénétiques (méthylation de l'ADN et modification des histones), la méthylation de l'ADN a été particulièrement étudiée. Au travers de plusieurs études portant sur des hybrides F1 intraspécifiques d'*Arabidopsis thaliana*, des variations du niveau de méthylation de l'ADN des hybrides F1 par rapport à celui des accessions parentales ont été observées. Ces variations sont apparues sur des loci pour lesquels le niveau de méthylation des parents se différenciait. De plus l'étude intégrative du profil de méthylation de l'ADN et de petits ARNs (siRNA de 24 nucléotides) a aussi indiqué que les variations du niveau de méthylation de l'ADN étaient corrélées à des différences du niveau de production des siRNA des accessions parentales. Au sein des hybrides F1, des mécanismes de méthylation et déméthylation chromosomiques gérés par des effets *trans* se produisent pour des régions différemment méthylées entre les parents, proches de régions géniques et corrélées avec des changements de niveau de siRNAs (Groszmann et al. 2011, Greaves et al. 2012, Shen et al. 2012, Greaves et al. 2014). Des observations similaires ont été faites sur des hybrides F1 entre sous-espèces d'*Oryza sativa japonica*. En effet Chodavarapu et al. (2012) ont mené une étude intégrative du méthylome, du transcriptome, et des siRNAs des hybrides F1. Ils suggèrent que les variations du niveau d'expression des gènes au sein des hybrides sont causées par des variations du profil de méthylation provoqués par des effets *trans* qui correspondraient à l'action de siRNAs (Chodavarapu et al. 2012). Alors que dans les études citées précédemment, les variations du profil de méthylation au sein des hybrides F1 semblent liés au phénomène d'hétérosis, pour des hybrides interspécifiques entre *Brassica napus* et *Brassica rapa* le profil de méthylation d'hybrides F1 ressemble à celui de l'une des espèces parentales. Ce résultat indiquerait une asymétrie génomique possible du profil de méthylation entre les espèces parentales (Xiong et al. 2013).

En s'appuyant sur ces observations, on peut imaginer que les espèces *C. canephora* et *C. eugenoides* présentent des variations du profil de méthylation. Alors pour certains gènes, l'accessibilité de l'ADN au moment de la transcription serait différente entre les deux espèces. Pour des gènes ayant des niveaux de méthylation différentiels entre les espèces parentales, le niveau de méthylation au sein des hybrides pourrait se distinguer de celui des espèces parentales suite à des effets *trans* provoqués par d'éventuels siRNAs parentaux. Par conséquent, pour ces gènes, le niveau de transcription des allèles pourrait varier par rapport aux espèces parentales. Pour évaluer cette hypothèse, il s'agit d'effectuer une analyse des hybrides F1 de *Coffea* et des espèces parentales *C. canephora* et *C. eugenoides* qui intégrera le séquençage de l'ADN génomique traité au bisulfite de sodium, le séquençage de petits ARNs (de 24 nucléotides) et du transcriptome. Le dispositif expérimental sera complété par *C. arabica*.

Une telle étude permettrait de différencier les différents types de méthylation de l'ADN des espèces *Coffea* (^mCG, ^mCHG, ^mCHH, H indiquant n'importe quel nucléotide) et surtout celui dépendant de siRNAs pour son établissement. Des différences de méthylation entre les espèces diploïdes de *Coffea* seront révélées. Elles pourraient avoir un lien avec les divergences d'évolution de ces espèces en raison de leur impact sur la plasticité phénotypique des espèces. Des hypothèses au sujet de la spéciation de ce genre pourraient être émises.

Combinés avec l'analyse du transcriptome, ces résultats permettraient de vérifier le lien potentiel entre la variation de la méthylation de l'ADN entre les espèces parentales et les variations de l'expression au sein des hybrides F1. La connaissance de la variation de l'expression allélique spécifique des hybrides F1 permettrait d'évaluer l'impact des variations potentielles de la méthylation de l'ADN et de la production des siRNAs sur la régulation de l'expression des allèles. Les mécanismes d'interactions entre les génomes seraient ainsi mieux décrits et compris.

De telles connaissances auraient des retombées intéressantes sur la compréhension de la régulation de l'expression des gènes de *C. arabica* et de sa diversité phénotypique. Plus généralement, tous ces éléments seraient utiles pour appréhender la diversité des espèces *Coffea* et son utilisation dans le cadre d'introggression pour l'amélioration de *C. arabica*.

Au-delà de cet intérêt, cette étude permettrait de compléter la description des hybrides F1 de *Coffea*. Dans le domaine de l'étude de la polyploïdie, la présentation d'un nouveau modèle d'hybridation permettrait de confirmer et généraliser des observations déjà effectuées sur d'autres modèles ou bien de limiter les conclusions de certaines études à des cas particuliers. Et enfin dans le cadre de l'étude de la polyploïdie, la comparaison des résultats concernant les hybrides F1 avec ceux de l'analyse intégrative de *C. arabica* permettrait d'évaluer les variations des profils de méthylation que la duplication et l'évolution depuis l'évènement de polyploïdisation auraient induites.

3.2 Dynamisme de la régulation de l'expression des gènes : mécanismes d'adaptation phénotypique des plantes après un évènement d'hybridation ou d'allopolypléisation

Au cours des dernières décennies, la diversité de la structure des génomes décrite par les biologistes a conduit à l'idée d'un génome dynamique et évolutif. Ils ont observé un contexte génétique considérablement redondant et complexe. Ils ont révélé que les réseaux de gènes sont constitués d'une profusion de gènes qui contrôlent l'expression d'éléments (gènes ou protéines) en aval, et que seulement une petite fraction de gènes a finalement un impact direct sur le phénotype (Parisod 2012). Récemment des hybrides intraspécifiques de *Drosophila melanogaster* ont été étudiés pour examiner la variation de l'hérédité de l'expression des gènes selon la température de

développement. Il a été d'abord observé, que la variation d'expression des gènes était associée à des variations de régulation dues à des changements de *trans*-régulation en conditions extrêmes. Ensuite il a été révélé que peu de facteurs de transcription impliqués dans la réponse à la température pouvaient être la cause de la coordination de l'expression de centaines de gènes en aval (Chen et al. 2015). Les auteurs ont évoqué la cohérence de cette observation avec le modèle de régulation (*trans*-régulateur et *trans*-sensoriel) exposé par Tirosh et al. (Tirosh et al. 2009).

Dans le contexte de l'aspect évolutif de la polyploïdie, le potentiel d'adaptation des polyploïdes à des environnements variables et limitants est toujours débattu, et les mécanismes moléculaires de l'adaptation, en cours d'observation, sont encore très mal compris. Les approches expérimentales qui se sont intéressées au rôle adaptatif de la polyploïdie en réponse à des variations environnementales sont peu nombreuses. Dans les études présentées dans les articles de synthèse de Madlung (Madlung 2012) et Parisod (Parisod 2012) les phénotypes des espèces polyploïdes et de leurs espèces parentales, sont souvent comparés pour démontrer un avantage adaptatif, mais il semblerait que la connection entre les changements d'expression des gènes et le rôle adaptatif de ce changement en réponse à un stress ne soit pas toujours faite. Pour comprendre les capacités adaptatives des plantes après un événement d'hybridation ou de polyploïdisation, il devient essentiel de définir le lien qui unit les variations d'expression et de régulation des gènes avec les variations des phénotypes en cas de stress environnemental ou d'adaptation à un environnement particulier.

Pour parvenir à cet objectif, des traitements environnementaux variables seront appliqués aux hybrides F1 de *Coffea* bouturés, à leurs parents *C. canephora* et *C. eugenioides* ainsi qu'à *C. arabica*. Des gammes de conditions normales à extrêmes concernant deux facteurs de croissance importants : la température et le pourcentage d'humidité du sol seront contrôlées indépendamment en phytotron. Ces deux contraintes environnementales, maîtrisables en phytotron, correspondent à des facteurs écologiques qui varient selon l'altitude, et qui pourraient permettre de bien distinguer les caractéristiques de chacune des espèces diploïdes et de *C. arabica*.

Après la culture des plantes dans de telles conditions, on s'intéressera d'une part au transcriptome de ces plantes et d'autre part à leurs réponses physiologiques aux conditions environnementales. Pour ce dernier point, plusieurs paramètres visant à évaluer la réponse physiologique seront mesurés : la morphologie foliaire (rapports d'allométrie, surface massique, densité stomatique) et le fonctionnement de la plante pour la photosynthèse (potentiel hydrique des feuilles, delta C13).

Le déroulement de cette analyse inclura plusieurs étapes.

- Identification des gènes impliqués dans l'adaptation aux conditions environnementales. Dans cette première partie, les analyses de l'expression totale des gènes et de l'expression allélique spécifique seront menées sur l'ensemble des individus du dispositif expérimental. Ces analyses permettront d'une

part d'identifier l'hérédité de l'expression des gènes pour chaque condition environnementale, et d'autre part d'évaluer les divergences de *cis* et *trans*-régulation entre les espèces parentales et les hybrides également pour chaque condition environnementale. Les gènes présentant des variations d'hérédité d'expression, ou des variations des divergences de *cis* et *trans*-régulation entre les différentes conditions appliquées aux plantes seront potentiellement inclus dans les mécanismes d'adaptation.

- Etude de la structure des réseaux de régulation de l'expression des gènes. Des gènes qui ont des fonctions liées dans un réseau de régulation, ou qui ont le même processus de régulation vont montrer un dynamisme d'expression et de régulation similaire selon les conditions expérimentales. A partir des données de co-expression, l'association fonctionnelle de gènes peut être déduite (Gehan et al. 2015, Lee et al. 2015). Ainsi les profils de co-expression des gènes des échantillons dans les différentes conditions expérimentales seront utilisés pour détecter l'association fonctionnelle de certains gènes et en déduire des réseaux de régulation. Cette démarche sera effectuée pour les hybrides et pour chacune des espèces représentées dans le dispositif expérimental.

- Identification et implication des facteurs de transcription. Comme cela a été effectué dans l'étude de l'effet de la température sur l'expression des gènes de *Drosophila melanogaster* (Chen et al. 2015), les gènes classés dans les catégories d'hérédité d'expression perturbées par les conditions de culture, seront analysés *in silico* pour révéler d'éventuels enrichissements en sites de liaison de facteurs de transcription. Les facteurs de transcription repérés par cette analyse *in silico* comme intervenant dans les processus d'adaptation à la température ou au degré d'hygrométrie, seront examinés pour d'éventuelles co-expressions pour les individus à une même condition, ou inversion de co-régulation entre condition. Ainsi des profils d'expression différentielle de gènes en réponse aux variations de la température et au degré d'hygrométrie seront repérés.

- Intégration des données de structure des réseaux de régulation et des données phénotypiques. Pour les variations environnementales étudiées, les données représentant la diversité de la réponse physiologique des échantillons du dispositif expérimental seront mises en relation avec les données de structure des réseaux de régulation. Cette démarche pourrait permettre d'identifier des réseaux de régulation spécifique de phénotypes.

Cette étude serait intéressante pour plusieurs raisons. D'abord elle permettrait d'observer et de décrire s'il y a lieu, le dynamisme de la régulation des gènes lorsque les plantes hybrides ou polyploïdes sont soumises à des stress environnementaux. C'est-à-dire que la caractérisation des bases génétiques de l'adaptation aux conditions environnementales limitantes serait possible par l'identification de variations des interactions entre les allèles. Ces variations pourraient être dues à des divergences de *trans*-régulation entre les espèces parentales des hybrides F1 pour les gènes sensoriels ou régulateurs de l'adaptation. Ces observations apporteraient des éléments pour évaluer la spéciation adaptative différentielle des espèces *C. canephora* et *C. eugenioides* et pour comprendre l'adaptation de ces

espèces à des environnements différents. De plus, l'intégration de *C. arabica* dans le dispositif expérimental, donnerait la possibilité de comparer les variations phénotypiques entre l'allopolyploïde et ses espèces parentales. Toutefois, l'analyse des divergences de *cis* et *trans*-régulation menée sur *C. arabica* présenterait certaines limites, puisque les espèces diploïdes actuelles ont évolué depuis l'évènement d'allopolyploïdisation. Malgré tout, l'identification de différences de régulation entre les espèces diploïdes pour les gènes impliqués dans l'adaptation à ces deux contraintes améliorerait la connaissance de l'adaptation environnementale de *C. arabica*, et exposerait un modèle polyploïde supplémentaire.

Ensuite ce travail permettrait de décrire les réseaux de régulation des gènes mis en place lors de l'adaptation à la température et à l'humidité du sol. Des gènes d'adaptation communs ou spécifiques aux deux contraintes environnementales pourraient être identifiés. Ces gènes permettraient de faire des hypothèses sur le processus global d'adaptation des plantes (les effecteurs et les voies de l'adaptation sont-ils propres à une contrainte ? ou bien est-ce un processus qui dépend d'éléments communs, par exemple, les capteurs des variations environnementales ?). Ces résultats pourraient confirmer ou infirmer le modèle de régulation des gènes proposé par Tirosh et al (Tirosh et al. 2009).

3.3 Complexité de la régulation des gènes au sein d'hybrides et de polyploïdes

Les études proposées permettraient en partie d'éclaircir pour des hybrides ou un polyploïde, le mécanisme d'expression différentielle des gènes entre espèces ou en réponse à des environnements variables. Mais il est évident que les conclusions de ces études seraient à considérer dans le contexte extrêmement complexe qui relie l'expression des gènes au phénotype d'une plante. De nombreux processus de régulation se produisent de la transcription de l'ARN à la synthèse des protéines. Certains de ces processus interviennent avant la transcription comme les différences épigénétiques, d'autres se produisent après la transcription et/ou la traduction des ARNs et d'autres enfin comme l'épissage alternatif en modifiant la nature de l'ARNm. Pour des hybrides et des polyploïdes, les études à l'échelle du génome, ont porté pendant de nombreuses années sur les variations d'expression des gènes, plus récemment et en raison de l'évolution technologique, des étapes initiales et intermédiaires de la régulation des gènes ont été examinées.

La structure de la chromatine, et plus particulièrement l'état des nucléosomes en amont des sites de transcription pour des hybrides de levure (Tirosh et al. 2010), la concordance entre le transcriptome et le translatome pour *Glycine dolichocarpa* (Coate et al. 2014), la variation de l'épissage alternatif pour *Brassica napus* (Zhou et al. 2011) ont été étudié dans le contexte d'hybridation ou d'allopolyploïdisation. Dernièrement, plusieurs études ont comparé le protéome de polyploïdes et de leurs parents (Ng et al. 2012 pour *Arabidopsis*, Koh et al. 2012 pour *Tragopogon mirus*, Shen et al. 2015 pour *Brassica*).

Selon les phénomènes étudiés, les résultats de ces études révèlent des variations plus (pour le protéome) ou moins (pour le transcriptome) importantes par rapport au transcriptome.

LISTE BIBLIOGRAPHIQUE

Liste bibliographique

Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U.S.A.* 100:4649–4654.

Adams KL, Percifield R, Wendel JF. 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics*. 168:2217–2226.

Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien M-A, Fukunaga K, Ricou M, Misset M-T. 2009. Hybridization, polyploidy and invasion: lessons from *Spartina* (*Poaceae*). *Biol. Invasions*. 11:1159–1173.

Ainouche M, Chelaifa H, Ferreira J, Bellot S, Ainouche A, Salmon A. 2012 Polyploid evolution in *Spartina* : dealing with highly redundant hybride genomes. in *Polyploidy and genome evolution*. Soltis PS and Soltis DE. Springer Berlin Heidelberg.

Akama S, Shimizu-Inatsugi R, Shimizu KK, Sese J. 2014. Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. *Nucleic Acids Res.* 42:e46

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*. 8:56.

Albertin W, Balliau T, Brabant P, Chèvre A-M, Eber F, Malosse C, Thiellement H. 2006. Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics*. 173:1101–1113.

Albertin W, Marullo P. 2012. Polyploidy in fungi: evolution after whole-genome duplication. *Proc. R. Soc.* 279:2497–2509

Alix K, Joets J, Ryder CD, Moore J, Barker GC, Bailey JP, King GJ, Pat Heslop-Harrison JS. 2008. The CACTA transposon Bot1 played a major role in *Brassica* genome divergence and gene proliferation. *Plant J. Cell Mol. Biol.* 56:1030–1044.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106.

Anthony F, Combes C, Astorga C, Bertrand B, Graziosi G, Lashermes P. 2002. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 104:894–900.

- Anthony F, Lashermes P.** 2005. Origin, evolution, and diversity of the Coffee (*Coffea arabica* L.) genome in Plant Genome: Biodiversity and Evolution, Vol. 1, Part B: Phanerogams (Higher Groups) Sharma AK and Sharma A. (Hardback) - Taylor & Francis.
- Anthony F, Diniz LEC, Combes M-C, Lashermes P.** 2010. Adaptive radiation in *Coffea* subgenus *Coffea* L. (*Rubiaceae*) in Africa and Madagascar. *Plant Syst. Evol.* 285:51–64.
- Arrigo N, Barker MS.** 2012. Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* 15:140–146.
- Axtell MJ.** 2013. Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* 64:137–159.
- Baldwin SJ, Husband BC.** 2011. Genome duplication and the evolution of conspecific pollen precedence. *Proc. R. Soc. Lond. B Biol. Sci.* 278:2011–2017.
- Bardil A, de Almeida JD, Combes MC, Lashermes P, Bertrand B.** 2011. Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytol.* 192:760–774.
- Barreto FS, Pereira RJ, Burton RS.** 2015. Hybrid dysfunction and physiological compensation in gene expression. *Mol. Biol. Evol.* 32:613–622.
- Bell GDM, Kane NC, Rieseberg LH, Adams KL.** 2013. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol. Evol.* 5:1309–1323.
- Berthou F, Mathieu C, Vedel F.** 1983. Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. *Theor. Appl. Genet.* 65:77–84.
- Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA.** 2010. Heterosis. *Plant Cell Online* 22:2105–2112.
- Blum A.** 2013. Heterosis, stress, and the environment: a possible road map towards the general improvement of crop yield. *J. Exp. Bot.* doi: 10.1093/jxb/ert289.
- Bombarely A, Edwards KD, Sanchez-Tamburrino J, Mueller LA.** 2012. Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: A phylogenomic perspective. *BMC Genomics.* 13:406.
- Bretagnolle F, Thompson JD.** 1995. Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytol.* 129:1–22.
- Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB.** 2010. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol. Ecol.* 19:132–146.
- Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB.** 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Curr. Biol. CB* 22:248–252.
- 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 Phytol.* 186:175–183.

- Buggs RJA, Soltis PS, Soltis DE.** 2009. Does hybridization between divergent progenitors drive whole-genome duplication? *Mol. Ecol.* 18:3334–3339.
- Buggs RJA, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE.** 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130354.
- Buggs RJA, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE.** 2011. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol. CB* 21:551–556.
- Carvalho A, Krug CA.** (1950), Genética de *Coffea* XIII -Hereditariedade da característica erecta em *Coffea arabica* L. *Bragantia.* 10: 321-328.
- Cenci A, Combes M-C, Lashermes P.** 2010. Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. *Mol. Genet. Genomics.* 283:493–501.
- Cenci A, Combes M-C, Lashermes P.** 2012. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant Mol. Biol.* 78:135–145.
- Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al.** 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950–953.
- 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 Biol.* 11:R125.
- 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*). *Genetic.* 182:503–517.
- Chelaifa H, Monnier A, Ainouche M.** 2010. Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica* (*Poaceae*). *New Phytol.* 186:161–174.
- Chen J, Nolte V, Schlötterer C.** 2015. Temperature stress mediates decanalization and dominance of gene expression in *Drosophila melanogaster*. *PLoS Genet.* 11:e1004883.
- Chen ZJ.** 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in Pplant polyploids. *Annu. Rev. Plant Biol.* 58:377–406.
- Chen ZJ.** 2010. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15:57–71.
- Chen ZJ, Ni Z.** 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays.* 28:240–252.
- Chester M, Gallagher JP, Symonds VV, Silva AVC da, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE.** 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (*Asteraceae*). *Proc. Natl. Acad. Sci.* 109:1176–1181.
- Chodavarapu RK, Feng S, Ding B, Simon SA, Lopez D, Jia Y, Wang G-L, Meyers BC, Jacobsen SE, Pellegrini M.** 2012. Transcriptome and methylome interactions in rice hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 109:12040–12045.

- Coate JE, Bar H, Doyle JJ.** 2014. Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *Plant Cell*. 26:136–150.
- Combes M-C, Cenci A, Baraille H, Bertrand B, Lashermes P.** 2012. Homeologous gene expression in response to growing temperature in a recent Allopolyploid (*Coffea arabica* L.). *J. Hered.* 103:36–46.
- Combes M-C, Dereeper A, Severac D, Bertrand B, Lashermes P.** 2013. Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol.* 200:251–260
- Combes M-C, Hueber Y, Dereeper A, Rialle S, Herrera JC, Lashermes P.** 2015. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biology Evolution*. 7:1110–1121
- Conant GC, Birchler JA, Pires JC.** 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* 19:91–98.
- Conant GC, Wolfe KH.** 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* 9:938–950.
- Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ.** 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* 24:797–808.
- Coolon JD, Wittkopp PJ.** 2013. *Cis*- and *trans*-Regulation in *Drosophila* Interspecific Hybrids. In: Chen ZJ, Birchler JA, editors. *Polyploid and Hybrid Genomics*. John Wiley & Sons.
- Cowles CR, Hirschhorn JN, Altshuler D, Lander ES.** 2002. Detection of regulatory variation in mouse genes. *Nat. Genet.* 32:432–437.
- Crawford DJ, Doyle JJ, Soltis DE, Soltis PS, Wendel JF.** 2014. Contemporary and future studies in plant speciation, morphological/floral evolution and polyploidy: honouring the scientific contributions of Leslie D. Gottlieb to plant evolutionary biology. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130341.
- Cros J, Combes MC, Chabrilange N, Duperray C, Monnot des Angles A, Hamon S.** 1995. Nuclear DNA content in the subgenus *Coffea* (*Rubiaceae*): inter- and intra-specific variation in African species. *Can. J. Bot.* 73:14–20.
- Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A, Lashermes P.** 1998. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. *Mol. Phylogenet. Evol.* 9:109–117.
- Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, et al.** 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16:738–749.
- DaMatta FM, Ramalho JDC.** 2006. Impacts of drought and temperature stress on coffee physiology and production: a review. *Braz. J. Plant Physiol.* 18:55–81.
- DaMatta FM, Ronchi CP, Maestri M, Barros RS.** 2007. Ecophysiology of coffee growth and production. *Braz. J. Plant Physiol.* 19:485–510.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P.** 2006. An annotated taxonomic conspectus of the genus *Coffea* (*Rubiaceae*). *Bot. J. Linn. Soc.* 152:465–512.
- Davis AP, Tosh J, Ruch N, Fay MF.** 2011. Growing coffee: *Psilanthus* (*Rubiaceae*) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot. J. Linn. Soc.* 167:357–377.

- Deal RB, Henikoff S.** 2011. Histone variants and modifications in plant gene regulation. *Curr. Opin. Plant Biol.* 14:116–122.
- Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G, et al.** 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345:1181–1184.
- Dong S, Adams KL.** 2011. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytol.* 190:1045–1057.
- Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M.** 2008. Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum*. *Genome.* 51(8):616–27.
- Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M.** 2010. Genome Size in Diploids, Allopolyploids, and Autopolyploids of Mediterranean *Triticeae*. *J. Bot.* 2010:e341380.
- Emerson JJ, Li W-H.** 2010. The genetic basis of evolutionary change in gene expression levels. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365:2581–2590.
- Evans BJ, Pyron AR, Wiens JJ.** 2012 Polyploidization and Sex Chromosome Evolution in Amphibians in Polyploidy and genome evolution. Soltis PS and Soltis DE. Springer Berlin Heidelberg.
- D' Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R.** 2009. Turning meiosis into mitosis. *PLoS Biol* 7:e1000124.
- D' Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Simon M, Jenczewski E, Mercier R.** 2008. Mutations in AtPS1 (*Arabidopsis thaliana* parallel spindle 1) lead to the production of diploid pollen grains. *PLoS Genet.* 4:e1000274.
- Estep MC, McKain MR, Diaz DV, Zhong J, Hodge JG, Hodkinson TR, Layton DJ, Malcomber ST, Pasquet R, Kellogg EA.** 2014. Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl. Acad. Sci.* 111:15149–15154.
- Fawcett JA, Maere S, Van de Peer Y.** 2009. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl. Acad. Sci.* 106:5737–5742.
- Feldman M, Levy AA.** 2012. Genome evolution due to allopolyploidization in wheat. *Genetics.* 192:763–774.
- Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM.** 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics.* 147:1381–1387.
- Flagel LE, Chen L, Chaudhary B, Wendel JF.** 2009. Coordinated and fine-scale control of homoeologous gene expression in allotetraploid cotton. *J. Hered.* 100:487–490.
- Flagel LE, Wendel JF.** 2009. Gene duplication and evolutionary novelty in plants. *New Phytol.* 183:557–564.
- Flagel LE, Wendel JF.** 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* 186:184–193.
- Flagel L, Udall J, Nettleton D, Wendel J.** 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* 6:16.

- 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.
- Gaeta RT, Chris Pires J.** 2010. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol.* 186:18–28.
- Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC.** 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell*. 19:3403–3417.
- Garsmeur O, Schnable JC, Almeida A, Jourda C, D’Hont A, Freeling M.** 2014. Two evolutionarily distinct classes of paleopolyploidy. *Mol. Biol. Evol.* 31:448–454.
- Gehan MA, Greenham K, Mockler TC, McClung CR.** 2015. Transcriptional networks – crops, clocks, and abiotic stress. *Curr. Opin. Plant Biol.* 24:39–46.
- Grassias M, Kammacher P.** 1975. Observations sur la conjugaison chromosomique de *Coffea arabica* L. *Café Cacao Thé*. 19: 177-190.
- Graze RM, McIntyre LM, Main BJ, Wayne ML, Nuzhdin SV.** 2009. Regulatory divergence in *Drosophila melanogaster* and *Drosophila simulans*, a genomewide analysis of Allele-Specific Expression. *Genetics*. 183:547–561.
- Graze RM, Novelo LL, Amin V, Fear JM, Casella G, Nuzhdin SV, McIntyre LM.** 2012. Allelic imbalance in *Drosophila* hybrid heads: exons, isoforms, and evolution. *Mol. Biol. Evol.* 29:1521–1532.
- Greaves IK, Groszmann M, Wang A, Peacock WJ, Dennis ES.** 2014. Inheritance of *Trans* Chromosomal Methylation patterns from *Arabidopsis* F1 hybrids. *Proc. Natl. Acad. Sci.*:201323656.
- Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES.** 2012. *Trans* Chromosomal Methylation in *Arabidopsis* hybrids. *Proc. Natl. Acad. Sci.* 109:3570–3575.
- Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES.** 2011. Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proc. Natl. Acad. Sci. U.S.A.* 108:2617–2622.
- Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF.** 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* 196:966–971.
- Grover CE, Yu Y, Wing RA, Paterson AH, Wendel JF.** 2008. A phylogenetic analysis of indel dynamics in the cotton genus. *Mol. Biol. Evol.* 25:1415–1428.
- Hahn MW.** 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* 100:605–617.
- Ha M.** 2013. Understanding the chromatin remodeling code. *Plant Sci.* 211:137–145.
- Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang X-J, Chen ZJ.** 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc. Natl. Acad. Sci.* 106:17835–17840
- Hamon P, Duroy P-O, Dubreuil-Tranchant C, Mafra D’Almeida Costa P, Duret C, Razafinarivo NJ, Couturon E, Hamon S, de Kochko A, Poncet V, Guyot R.** 2011. Two novel Ty1-copia retrotransposons isolated from coffee trees can effectively reveal evolutionary relationships in the *Coffea* genus (*Rubiaceae*). *Mol. Genet. Genomics.* 285:447–460.

- Hao M, Luo J, Zeng D, Zhang L, Ning S, Yuan Z, Yan Z, Zhang H, Zheng Y, Feuillet C, Choulet F, Yen Y, Zhang L.** 2014. QTug.sau-3B Is a Major Quantitative Trait Locus for Wheat Hexaploidization. *Genes Genomes Genetics*. 4:1943–1953.
- He F, Zhang X, Hu J, Turck F, Dong X, Goebel U, Borevitz J, de Meaux J.** 2012. Genome-wide analysis of *cis*-regulatory divergence between species in the *Arabidopsis* genus. *Mol. Biol. Evol.* 29:3385–3395.
- 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. *Curr. Biol.* 16:1652–1659.
- Heslop-Harrison JSP.** 2012. Genome evolution: extinction, continuation or explosion? *Curr. Opin. Plant Biol.* 15:115–121.
- Higgins J, Magusin A, Trick M, Fraser F, Bancroft I.** 2012. Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. *BMC Genomics*. 13:247.
- Hoekstra HE, Coyne JA.** 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evol. Int. J. Org. Evol.* 61:995–1016.
- Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel L, Wendel JF.** 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc. Natl. Acad. Sci.* 105:6191–6195.
- Hughes AL.** 1994. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* 256:119–124.
- Hu G, Koh J, Yoo M-J, Chen S, Wendel JF.** 2015. Gene-expression novelty in allopolyploid cotton: a proteomic perspective. *Genetics*. 200:91–104.
- Ilut DC, Coate JE, Luciano AK, Owens TG, May GD, Farmer A, Doyle JJ.** 2012. A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. *Am. J. Bot.* 99:383–396.
- Innan H, Kondrashov F.** 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11:97–108.
- Jablonka E, Lamb MJ.** 2015. The inheritance of acquired epigenetic variations. *Int. J. Epidemiol.* doi: 10.1093/ije/dyv020.
- Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al.** 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 449:463–467.
- James AB, Syed NH, Bordage S, Marshall J, Nimmo GA, Jenkins GI, Herzyk P, Brown JWS, Nimmo HG.** 2012. Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell*. 24:961–981.
- Jenczewski E, Alix K.** 2004. From Diploids to Allopolyploids: The emergence of efficient pairing control genes in plants. *Crit. Rev. Plant Sci.* 23:21–45.
- Jiao Y, Paterson AH.** 2014. Polyploidy-associated genome modifications during land plant evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130355.

- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al.** 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature*. 473:97–100.
- Kashkush K, Feldman M, Levy AA.** 2002. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* 33:102–106.
- Kellis M, Birren BW, Lander ES.** 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*. 428:617–624.
- Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, Levy AA.** 2011. Wheat hybridization and polyploidization results in deregulation of small RNAs. *Genetics*. 188:263–272.
- Koh J, Chen S, Zhu N, Yu F, Soltis PS, Soltis DE.** 2012. Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (*Asteraceae*) and its parents. *New Phytol.* 196:292–305.
- Koh J, Soltis PS, Soltis DE.** 2010. Homeolog loss and expression changes in natural populations of the recently and repeatedly formed allotetraploid *Tragopogon mirus* (*Asteraceae*). *BMC Genomics*. 11:97.
- Krug CA, Mendes AJT.** 1940. Cytological observations in *Coffea*. IV. *J. Genet.* 39:189–203.
- Kuo D, Licon K, Bandyopadhyay S, Chuang R, Luo C, Catalana J, Ravasi T, Tan K, Ideker T.** 2010. Coevolution within a transcriptional network by compensatory *trans* and *cis* mutations. *Genome Res.* 20:1672–1678.
- Landry CR, Hartl DL, Ranz JM.** 2007. Genome clashes in hybrids: insights from gene expression. *Heredity*. 99:483–493.
- Lashermes P.** 2000. Brief communication. Single-locus inheritance in the allotetraploid *Coffea arabica* L. and interspecific Hybrid *C. arabica* x *C. canephora*. *J. Hered.* 91:81–85.
- Lashermes P, Combes M-C, Hueber Y, Severac D, Dereeper A.** 2014. Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. *Plant J.* 78:674–685.
- Lashermes P, Combes M-C, 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, 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, Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A.** 1996. Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. *Theor. Appl. Genet.* 93:626–632.
- Lashermes P, Cros J, Marmey P, Charrier A.** 1993. Use of random amplified DNA markers to analyse genetic variability and relationships of *Coffea* species. *Genet. Resour. Crop Evol.* 40:91–99.
- Lee T, Kim H, Lee I.** 2015. Network-assisted crop systems genetics: network inference and integrative analysis. *Curr. Opin. Plant Biol.* 24:61–70.

- Liu S-L, Baute GJ, Adams KL.** 2011. Organ and cell type-specific complementary expression patterns and regulatory neofunctionalization between duplicated genes in *Arabidopsis thaliana*. *Genome Biol. Evol.* 3:1419–1436.
- Liu Z, Adams KL.** 2007. Expression Partitioning between Genes Duplicated by Polyploidy under Abiotic Stress and during Organ Development. *Curr. Biol.* 17:1669–1674.
- Louarn J.** 1976. Hybrides interspécifiques entre *Coffea canephora* Pierre et *Coffea eugenioides* Moore. *Café-Cacao-Thé.* 20:433–452.
- Louarn J.** 1993. Structure génétique des caféiers africains diploïdes basée sur la fertilité des hybrides interspécifiques. Association Scientifique Internationale du Café, 15eme Colloque. Montpellier
- Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T.** 2006. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiol.* 140:336–348.
- Lynch M, Force A.** 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics.* 154:459–473.
- Mable BK.** 2004. “Why polyploidy is rarer in animals than in plants”: myths and mechanisms. *Biol. J. Linn. Soc.* 82:453–466.
- Mable BK, Alexandrou MA, Taylor MI.** 2011. Genome duplication in amphibians and fish: an extended synthesis. *J. Zool.* 284:151–182.
- Madlung A.** 2012. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity.* 110:99–104.
- 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 Physiol.* 129:733–746.
- Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L.** 2004. Genomic changes in synthetic *Arabidopsis* polyploids. *Plant J.* 41:221–230.
- Madlung A, Wendel JF.** 2013. Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet. Genome Res.* 140:270–285.
- Maherali H, Walden AE, Husband BC.** 2009. Genome duplication and the evolution of physiological responses to water stress. *New Phytol.* 184:721–731.
- Marmagne A, Brabant P, Thiellement H, Alix K.** 2010. Analysis of gene expression in resynthesized *Brassica napus* allotetraploids: transcriptional changes do not explain differential protein regulation. *New Phytol.* 186:216–227.
- Martinez Palacios P.** Réponse des composants non codants du génome - éléments transposables et petits ARNs - à un événement allopolyploïde : le génome du colza (*Brassica napus*) comme modèle d'étude. Thèse de doctorat en Sciences du Végétal. Université Paris-Sud. 2014. 189 pages
- 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.* 100:1565–1583.

- Mayrose I, Zhan SH, Rothfels CJ, Arrigo N, Barker MS, Rieseberg LH, Otto SP.** 2014. Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis et al. (2014). *New Phytol.* 206:27-35.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP.** 2011. Recently formed polyploid plants diversify at lower rates. *Science.* 333:1257-1257.
- McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ.** 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* 20:816-825.
- Mochida K, Yamazaki Y, Ogihara Y.** 2003. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol. Genet. Genomics.* 270:371-377.
- Moghe GD, Shiu S-H.** 2014. The causes and molecular consequences of polyploidy in flowering plants. *Ann. N. Y. Acad. Sci.* 1320:16-34.
- Montagnon C, Cubry P, Leroy T.** 2006. *Coffea canephora* Pierre genetic improvement: Acquired knowledge, strategies and perspectives. *Cah. Agric.* 21:143-153.
- Ng DW-K, Zhang C, Miller M, Shen Z, Briggs SP, Chen ZJ.** 2012. Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. *Heredity.* 108:419-430.
- Ni Z, Kim E-D, 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.
- Nomura T, Ishihara A, Yanagita RC, Endo TR, Iwamura H.** 2005. Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proc. Natl. Acad. Sci. U. S. A.* 102:16490-16495.
- Osabe K, Clement JD, Bedon F, Pettolino FA, Ziolkowski L, Llewellyn DJ, Finnegan EJ, Wilson IW.** 2014. Genetic and DNA Methylation Changes in Cotton (*Gossypium*) Genotypes and Tissues. *PLoS ONE* 9:e86049.
- Otto SP.** 2007. The Evolutionary Consequences of Polyploidy. *Cell.* 131:452-462.
- Palusa SG, Reddy ASN.** 2015. Differential recruitment of splice variants from SR pre-mRNAs to polysomes during development and in response to stresses. *Plant Cell Physiol.* 56:421-427.
- Pandit MK, Pocock MJO, Kunin WE.** 2011. Ploidy influences rarity and invasiveness in plants. *J. Ecol.* 99:1108-1115.
- Parisod C.** 2012. Polyploids integrate genomic changes and ecological shifts. *New Phytol.* 193:297-300.
- Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien M-A.** 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* 186:37-45.
- Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M.** 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytol.* 184:1003-1015.
- Paterson AH, Bowers JE, Chapman BA.** 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. U.S.A.* 101:9903-9908.

- Pécrix Y, Rallo G, Folzer H, Cigna M, Gudin S, Le Bris M.** 2011. Polyploidization mechanisms: temperature environment can induce diploid gamete formation in *Rosa sp.* *J. Exp. Bot.* 62:3587–3597.
- Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K.** 2009. The flowering world: a tale of duplications. *Trends Plant Sci.* 14:680–688.
- Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien M-A, et al.** 2010. Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytol.* 186:135–147.
- Privat I, Bardil A, Gomez AB, Severac D, Dantec C, Fuentes I, Mueller L, Joët T, Pot D, Foucrier S, 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.* 12:5–5.
- Prud’homme B, Gompel N, Carroll SB.** 2007. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci.* 104:8605–8612.
- Qi B, Huang W, Zhu B, Zhong X, Guo J, Zhao N, Xu C, Zhang H, Pang J, Han F, et al.** 2012. Global transgenerational gene expression dynamics in two newly synthesized allohexaploid wheat (*Triticum aestivum*) lines. *BMC Biol.* 10:3.
- Ramanna MS, Jacobsen E.** 2003. Relevance of sexual polyploidization for crop improvement – A review. *Euphytica.* 133:3–8.
- Ramsey J.** 2011. Polyploidy and ecological adaptation in wild yarrow. *Proc. Natl. Acad. Sci.* 108:7096–7101.
- Ramsey J, Schemske DW.** 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29:467–501.
- Ramsey J, Schemske DW.** 2002. Neopolyploidy in Flowering Plants. *Annu. Rev. Ecol. Syst.* 33:589–639.
- Rapp RA, Udall JA, Wendel JF.** 2009. Genomic expression dominance in allopolyploids. *BMC Biol.* 7:18.
- Rapp RA, Wendel JF.** 2005. Epigenetics and plant evolution. *New Phytol.* 168:81–91.
- Reddy ASN, Marquez Y, Kalyna M, Barta A.** 2013. Complexity of the alternative splicing landscape in plants. *Plant Cell.* 25:3657–3683.
- Renny-Byfield S, Chester M, Kovařík A, Comber SCL, Grandbastien M-A, Deloger M, Nichols R, Macas J, Novák P, Chase MW, et al.** 2011. Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. *Mol. Biol. Evol.* 28:2843–2854.
- Renny-Byfield S, Kovařík A, Chester M, Nichols RA, Macas J, Novák P, Leitch AR.** 2012. Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. *PLoS ONE* 7:e36963.
- Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien M-A, Leitch AR.** 2013. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *Plant J.* 74:829–839.
- Renny-Byfield S, Wendel JF.** 2014. Doubling down on genomes: Polyploidy and crop plants. *Am. J. Bot.* 101:1711–1725.

- Salmon A, Ainouche ML, Wendel JF.** 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (*Poaceae*). *Mol. Ecol.* 14:1163–1175.
- Salmon A, Flagel L, Ying B, Udall JA, Wendel JF.** 2010. Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol.* 186:123–134.
- Saminathan T, Nimmakayala P, Manohar S, Malkaram S, Almeida A, Cantrell R, Tomason Y, Abburi L, Rahman MA, Vajja VG, et al.** 2015. Differential gene expression and alternative splicing between diploid and tetraploid watermelon. *J. Exp. Bot.* 66:1369–1385.
- Schlichting CD, Wund MA.** 2014. Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. *Evolution.* 68:656–672.
- Schranz ME, Mohammadin S, Edger PP.** 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.* 15:147–153.
- Shen H, He H, Li J, Chen W, Wang X, Guo L, Peng Z, He G, Zhong S, Qi Y, et al.** 2012. Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell.* 24:875–892.
- Shen Y, Zhang Y, Zou J, Meng J, Wang J.** 2015. Comparative proteomic study on *Brassica* hexaploid and its parents provides new insights into the effects of polyploidization. *J. Proteomics.* 112:274–284.
- Shen Y, Zhou Z, Wang Z, Li W, Fang C, Wu M, Ma Y, Liu T, Kong L-A, Peng D-L, et al.** 2014. Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell.* 26:996–1008.
- Shi X, Ng DW-K, Zhang C, Comai L, Ye W, Chen ZJ.** 2012. *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat. Commun.* 3:950.
- Shi X, Ng DW-K, Zhang C, Comai L, Ye W, Jeffrey Chen Z.** 2012. *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat. Commun.* 3:950.
- De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y.** 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 110:2898–2903.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, de Pamphilis CW, Wall PK, Soltis PS.** 2009. Polyploidy and angiosperm diversification. *Am J Bot.* 96:336–348.
- Soltis DE, Buggs RJA, Doyle JJ, Soltis PS.** 2010. What we still don't know about polyploidy. *Taxon.* 59:1387–1403.
- Soltis DE, Buggs RJA, Barbazuk WB, Chamala S, Chester M, Gallagher JP, Schnable PS, Soltis PS.** 2012 The early stages of polyploidy : rapid and repeated evolution in *Tragopogon*. in *Polyploidy and genome evolution*. Soltis PS and Soltis DE. Springer Berlin Heidelberg.
- Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, Mei W, Cortez MB, Soltis PS, Gitzendanner MA.** 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytol.* 202:1105–1117.
- Soltis DE, Visger CJ, Soltis PS.** 2014. The polyploidy revolution then...and now: Stebbins revisited. *Am. J. Bot.* 101:1057–1078.

- 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.
- Stern DL, Orgogozo V.** 2008. The loci of evolution: how predictable is genetic evolution? *Evolution*. 62:2155–2177.
- Syed NH, Kalyna M, Marquez Y, Barta A, Brown JWS.** 2012. Alternative splicing in plants-coming of age. *Trends Plant Sci*. 17:616–623.
- Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, et al.** 2010. The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytol*. 186:102–112.
- Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE.** 2009. On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). *BMC Plant Biol*. 9:80.
- Tate JA, Ni Z, Scheen A-C, 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*. 173:1599–1611.
- Thomas AS.** 1944. The wild coffees of Uganda. *Empire journal of Experimental Agriculture*. 12(45):1-12.
- Thomas BC, Pedersen B, Freeling M.** 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res*. 16:934–946.
- Thompson KA, Husband BC, Maherali H.** 2014. Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *Am. J. Bot*. 101:1868–1875.
- Tirosh I, Sigal N, Barkai N.** 2010. Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. *Mol. Syst. Biol*. 6:365.
- Tirosh I, Reikhav S, Levy AA, Barkai N.** 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*. 324:659–662.
- Vanneste K, Baele G, Maere S, Van de Peer Y.** 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Research*. doi:10.1101/gr.168997.113
- Vanneste K, Maere S, Van de Peer Y.** 2014. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 369(1648).
- Wang B-B, Brendel V.** 2006. Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U.S.A.* 103:7175–7180.
- Wang J, Tian L, Lee H-S, Chen ZJ.** 2006. Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in *Arabidopsis* allopolyploids. *Genetics*. 173:965–974.
- Wang J, Tian L, Lee H-S, 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.

- Wang J, Tian L, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ. 2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* 167:1961–1973.
- Wang Y, Wang X, Paterson AH. 2012. Genome and gene duplications and gene expression divergence: a view from plants. *Ann. N. Y. Acad. Sci.* 1256:1–14.
- Wang Z, Sun X, Zhao Y, Guo X, Jiang H, Li H, Gu Z. 2015. Evolution of gene regulation during transcription and translation. *Genome Biol. Evol.* 7:1155–1167.
- Weiss-Schneeweiss H, Emadzade K, Jang T-S, Schneeweiss GM. 2013. Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenet. Genome Res.* 140:137–150.
- Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature*. 430:85–88.
- Wittkopp PJ, Kalay G. 2012. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13:59–69.
- Wolf JBW, Lindell J, Backström N. 2010. Speciation genetics: current status and evolving approaches. *Philos. Trans. R. Soc. B Biol. Sci.* 365:1717–1733.
- Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biol* 8:e1000409.
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U.S.A.* 106:13875–13879.
- Xiong W, Li X, Fu D, Mei J, Li Q, Lu G, Qian L, Fu Y, Disi JO, Li J, et al. 2013. DNA methylation alterations at 5'-CCGG sites in the interspecific and intraspecific hybridizations derived from *Brassica rapa* and *Brassica napus*. *PLoS ONE* 8:e65946.
- Xiong Z, Gaeta RT, Pires JC. 2011. Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. *Proc. Natl. Acad. Sci. U.S.A.* 108:7908–7913.
- Xu C, Bai Y, Lin X, Zhao N, Hu L, Gong Z, Wendel JF, Liu B. 2014. Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Mol. Biol. Evol.* 31:1066–1076.
- Xu Y, Zhong L, Wu X, Fang X, Wang J. 2009. Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. *Planta* 229:471–483.
- Yang C, Zhao L, Zhang H, Yang Z, Wang H, Wen S, Zhang C, Rustgi S, Wettstein D von, Liu B. 2014. Evolution of physiological responses to salt stress in hexaploid wheat. *Proc. Natl. Acad. Sci.* 111:11882–11887.
- Yang X, Ye C-Y, Cheng Z-M, Tschaplinski TJ, Wulschleger SD, Yin W, Xia X, Tuskan GA. 2010. Genomic aspects of research involving polyploid plants. *Plant Cell Tissue Organ Cult.* 104:387–397.
- Yoo M-J, Szadkowski E, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity*. 110:171–180.
- Zhang PG, Huang SZ, Pin A-L, Adams KL. 2010. Extensive divergence in alternative splicing patterns after gene and genome duplication during the evolutionary history of *Arabidopsis*. *Mol. Biol. Evol.* 27:1686–1697.

Zhao N, Zhu B, Li M, Wang L, Xu L, Zhang H, Zheng S, Qi B, Han F, Liu B. 2011. Extensive and heritable epigenetic remodeling and genetic stability accompany allohexaploidization of wheat. *Genetics*. 188:499–510.

Zhou R, Moshgabadi N, Adams KL. 2011. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc. Natl. Acad. Sci.* 108:16122–16127.

Zhou X, Lin Z, Ma H. 2010. Phylogenetic detection of numerous gene duplications shared by animals, fungi and plants. *Genome Biol.* 11(4):R38.

ANNEXES

Annexes

Annexe 1: Figures supplémentaires de l'article

Regulatory divergence between parental alleles determines gene expression patterns in hybrids.

Marie Christine Combes, Yann Hueber, Alexis Dereeper, Stéphanie Rialle, Juan-Carlos Herrera, Philippe Lashermes. 2015. *Genome Biology Evolution*. 7 : 1110-1121.

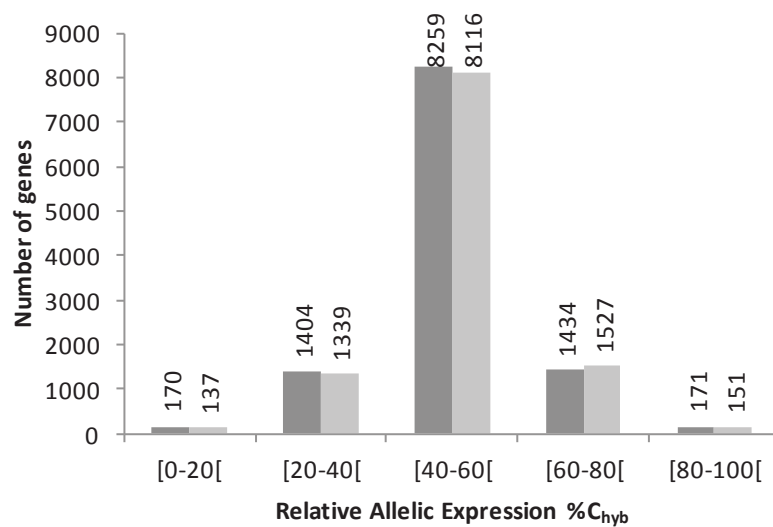


Figure S1: Relative allele expression (RAE) of 11,438 genes in CxE hybrids (dark gray bars) and of 11,270 genes in ExC hybrids (light gray bars). Genes were classified according to the relative expression of their alleles represented by the contribution of allele C to total gene expression (%C_{hyb}). The exact number of genes in each category is given at the top of each column. More than 70% of the genes in both subsets of hybrids showed balanced allelic expression (72.2% for CxE hybrids and 72 % for ExC hybrids) and the average percentage of the C_{hyb} allele was 50.2% for the CxE and 50.6% for the ExC hybrids. The allelic expression was similarly unbalanced toward the E allele or C allele. In CxE hybrids, 13.8% of genes showed allelic expression biased toward the E allele and 14% toward the C allele, ($P > 0.05$, Fisher's exact test). In ExC hybrids, despite a significant statistical test ($P = 0.0007$), allelic expression of genes was almost equally biased toward the E allele (13.1%) and the C allele (14.9%).

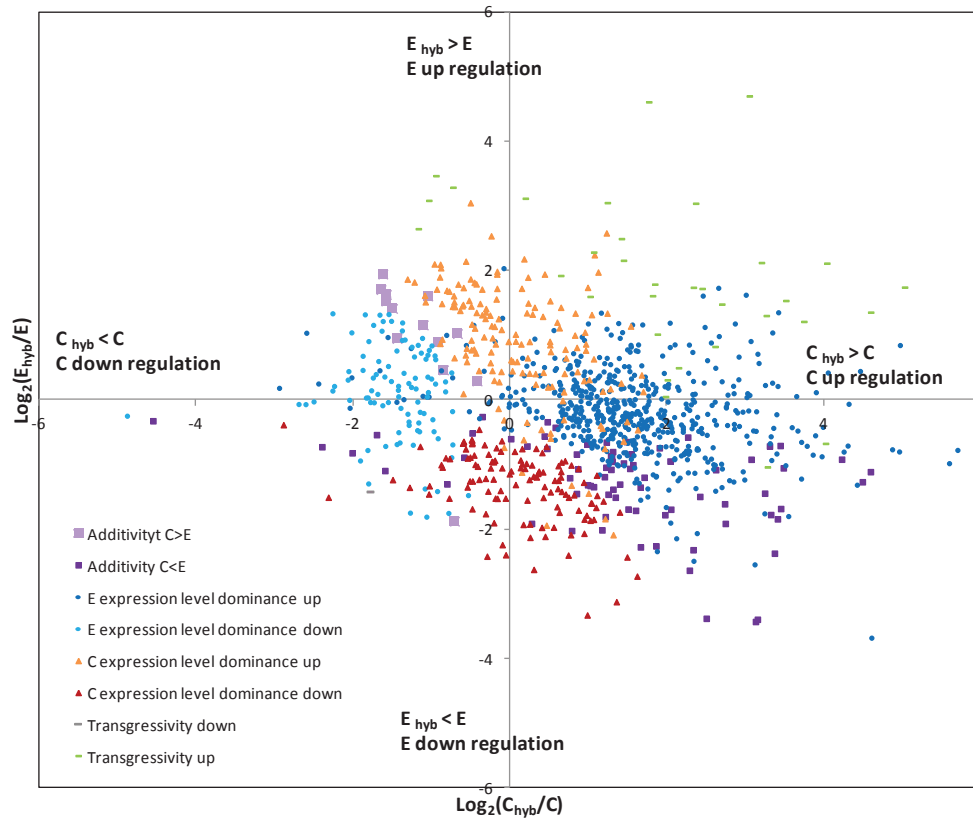


Figure S2: Variation of ASE for genes differentially expressed between the parents and the hybrid in ExC hybrids. Each point represents a single gene, and displays the combined up- and down-regulation of both alleles (C_{hyb}/C on the x-axis; E_{hyb}/E on the y-axis, C_{hyb} , E_{hyb} and C , E corresponding to allelic and parental read counts respectively) through its position on the one hand, and through its color code on the other hand, the inheritance category of the gene (Chi-squared test, $P < 0.001$).

ExC

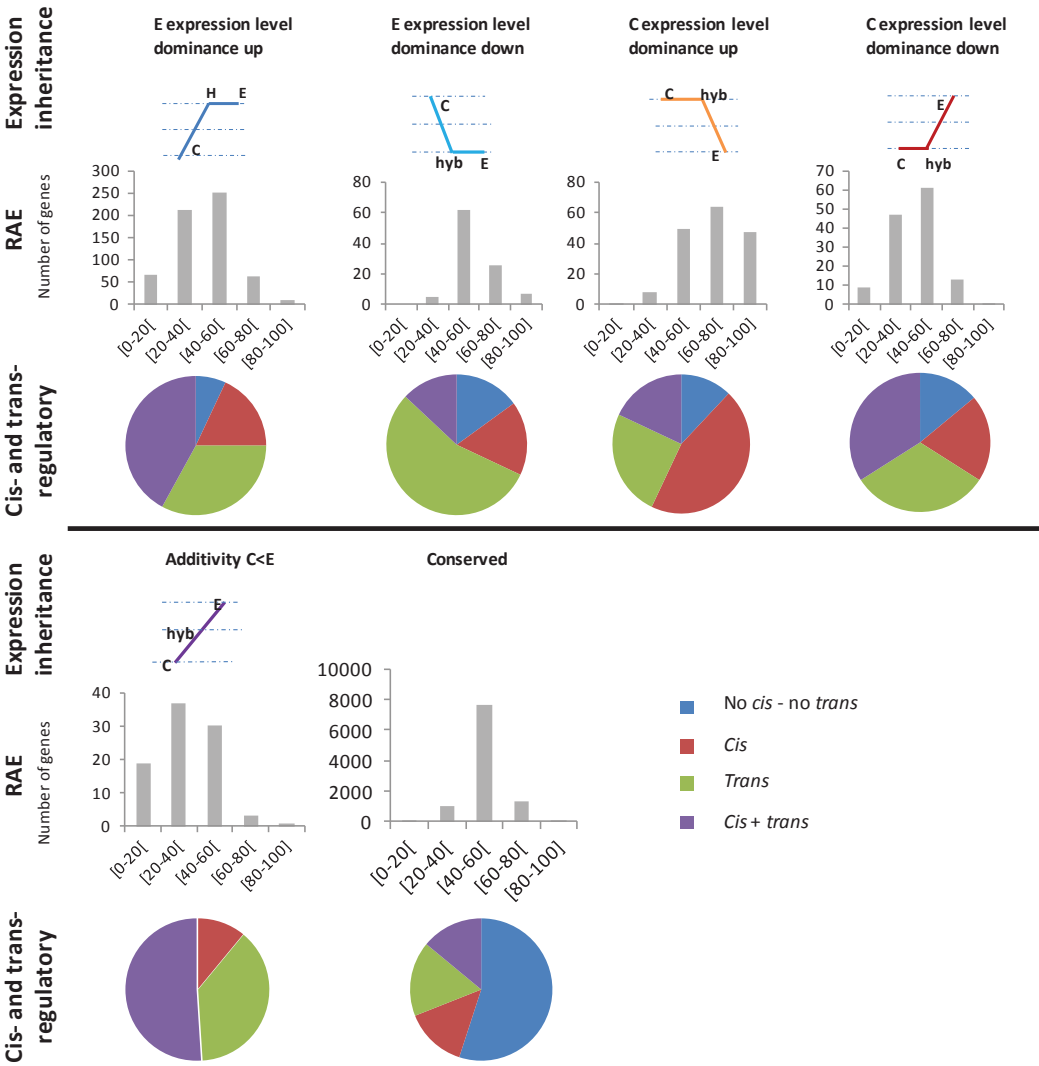


Figure S3: For the ExC hybrids, assembly of data on genes belonging to each inheritance category.

Annexe 2: Tableaux supplémentaires de l'article

Regulatory divergence between parental alleles determines gene expression patterns in hybrids.

Marie Christine Combes, Yann Hueber, Alexis Dereeper, Stéphanie Rialle, Juan-Carlos Herrera, Philippe Lashermes. 2015. *Genome Biology Evolution*. 7 : 1110-1121.

Table S1 : Gene ontology enrichment analysis for genes differentially expressed between parental species

Table S2 : Gene ontology enrichment analysis for genes more expressed in C than in E

Table S3 : Gene ontology enrichment analysis for genes more expressed in E than in C

Table S4 : Gene ontology enrichment analysis for genes belonging to divergence regulatory category

Table S1 Gene ontology enrichment analysis for genes differentially expressed between parental species

LevelGO	Term(Acc)	Term(Name)	#Seq	Score	Parents(Acc)	Parents(Name)	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
7	GO:0005856	cytoskeleton	105	105	GO:0043232	intracellular non-mem	C	1,08E-07	4,8E-10	105	116	5652	14881	over
2	GO:0005576	extracellular region	297	293,52	GO:0005575	cellular_component	C	2,84E-05	2E-06	297	550	5460	14447	over
6	GO:0005773	vacuole	242	240,72	GO:0044444, GO:0043231	cytoplasmic part, intra	C	0,020117	0,00536	242	508	5515	14489	over
5	GO:0005618	cell wall	199	199	GO:0030312	external encapsulatin	C	0,021114	0,00572	199	408	5558	14589	over
4	GO:0030312	external encapsulating structure	201	121,4	GO:0044464, GO:0071944	cell part, cell peripher	C	0,02202	0,00607	201	414	5556	14583	over
2	GO:0032991	macromolecular complex	88	31,68	GO:0005575	cellular_component	C	8,82E-05	6,7E-06	88	380	5669	14617	under
6	GO:0005840	ribosome	88	88	GO:0030529, GO:0044444, GO:0	ribonucleoprotein com	C	9,73E-05	8,2E-06	88	379	5669	14618	under
5	GO:0030529	ribonucleoprotein complex	88	52,8	GO:0032991, GO:0044424	macromolecular comp	C	9,73E-05	8,2E-06	88	379	5669	14618	under
3	GO:0043227	membrane-bounded organelle	1958	880,68	GO:0043226	organelle	C	0,000121	1,2E-05	1958	5591	3799	9406	under
6	GO:0043231	intracellular membrane-bounded organelle	1958	1467,79	GO:0043229, GO:0043227	intracellular organelle	C	0,000121	1,2E-05	1958	5591	3799	9406	under
6	GO:0005739	mitochondrion	362	362	GO:0044444, GO:0043231	cytoplasmic part, intra	C	0,000184	2,3E-05	362	1200	5395	13797	under
2	GO:0031974	membrane-enclosed lumen	114	15,42	GO:0005575	cellular_component	C	0,000184	2,3E-05	114	455	5643	14542	under
7	GO:0031981	nuclear lumen	114	71,4	GO:0044428, GO:0070013	nuclear part, intracell	C	0,000184	2,3E-05	114	455	5643	14542	under
4	GO:0043233	organelle lumen	114	25,7	GO:0044422, GO:0031974	organelle part, membe	C	0,000184	2,3E-05	114	455	5643	14542	under
6	GO:0070013	intracellular organelle lumen	114	42,84	GO:0044446, GO:0043233	intracellular organelle	C	0,000184	2,3E-05	114	455	5643	14542	under
5	GO:0044444	cytoplasmic part	1598	1306,63	GO:0005737, GO:0044424	cytoplasm, intracellul	C	0,00036	5,1E-05	1598	4593	4159	10404	under
3	GO:0044422	organelle part	130	29,16	GO:0005575, GO:0043226	cellular_component, c	C	0,000419	6,3E-05	130	494	5627	14503	under
4	GO:0044446	intracellular organelle part	130	48,6	GO:0044422, GO:0043229, GO:0	organelle part, intrace	C	0,000419	6,3E-05	130	494	5627	14503	under
2	GO:0043226	organelle	2053	570,09	GO:0005575	cellular_component	C	0,000429	6,9E-05	2053	5797	3704	9200	under
5	GO:0043229	intracellular organelle	2053	950,16	GO:0043226, GO:0044424	organelle, intracellula	C	0,000429	6,9E-05	2053	5797	3704	9200	under
6	GO:0044428	nuclear part	130	81	GO:0005634, GO:0044446	nucleus, intracellular	C	0,000443	7,5E-05	130	493	5627	14504	under
4	GO:0044424	intracellular part	2401	961,63	GO:0044464, GO:0005622	cell part, intracellular	C	0,000889	0,00017	2401	6690	3356	8307	under
6	GO:0005794	Golgi apparatus	138	138	GO:0044444, GO:0043231	cytoplasmic part, intra	C	0,001069	0,0002	138	507	5619	14490	under
4	GO:0005622	intracellular	2511	848,98	GO:0044464	cell part	C	0,001214	0,00027	2511	6964	3246	8033	under
6	GO:0005654	nucleoplasm	34	34	GO:0044428, GO:0031981	nuclear part, nuclear l	C	0,001214	0,00028	34	169	5723	14828	under
5	GO:0005737	cytoplasm	1926	1155,98	GO:0044424	intracellular part	C	0,005756	0,00138	1926	5373	3831	9624	under
7	GO:0005634	nucleus	756	685,6	GO:0043231	intracellular membran	C	0,016379	0,00422	756	2202	5001	12795	under
6	GO:0005730	nucleolus	85	85	GO:0044428, GO:0031981, GO:0	nuclear part, nuclear l	C	0,020117	0,00533	85	309	5672	14688	under
6	GO:0005768	endosome	41	41	GO:0044444, GO:0043231	cytoplasmic part, intra	C	0,023438	0,00667	41	170	5716	14827	under
6	GO:0005829	cytosol	329	329	GO:0044444	cytoplasmic part	C	0,043763	0,01342	329	998	5428	13999	under
8	GO:0003774	motor activity	44	44	GO:0017111	nucleoside-triphospha	F	2,25E-06	1,4E-07	44	34	5713	14963	over
6	GO:0016462	pyrophosphatase activity	44	15,84	GO:0016818	hydrolase activity, act	F	2,25E-06	1,4E-07	44	34	5713	14963	over
4	GO:0016817	hydrolase activity, acting on acid anhydrides	44	5,7	GO:0016787	hydrolase activity	F	2,25E-06	1,4E-07	44	34	5713	14963	over
5	GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	44	9,5	GO:0016817	hydrolase activity, act	F	2,25E-06	1,4E-07	44	34	5713	14963	over
7	GO:0017111	nucleoside-triphosphatase activity	44	26,4	GO:0016462	pyrophosphatase acti	F	2,25E-06	1,4E-07	44	34	5713	14963	over
2	GO:0005198	structural molecule activity	85	85	GO:0003674	molecular_function	F	2,85E-07	1E-08	85	418	5672	14579	under
4	GO:0003676	nucleic acid binding	644	426,16	GO:1901363, GO:0097159	heterocyclic compound	F	2,64E-05	1,8E-06	644	2047	5113	12950	under

Table S1 Gene ontology enrichment analysis for genes differentially expressed between parental species														
3	GO:0097159	organic cyclic compound binding	1669	679,06	GO:0005488	binding	F	0,00023	3,1E-05	1669	4796	4088	10201	under
3	GO:1901363	heterocyclic compound binding	1669	679,06	GO:0005488	binding	F	0,00023	3,1E-05	1669	4796	4088	10201	under
5	GO:0003723	RNA binding	140	126,6	GO:0003676	nucleic acid binding	F	0,000267	3,7E-05	140	531	5617	14466	under
2	GO:0005488	binding	3292	2518,7	GO:0003674	molecular_function	F	0,0024	0,00055	3292	8971	2465	6026	under
2	GO:0001071	nucleic acid binding transcription factor activity	136	81,6	GO:0003674	molecular_function	F	0,027068	0,00794	136	457	5621	14540	under
3	GO:0003700	sequence-specific DNA binding transcription factor activity	136	136	GO:0001071	nucleic acid binding tr	F	0,027068	0,00794	136	457	5621	14540	under
3	GO:0006950	response to stress	1091	1091	GO:0050896	response to stimulus	P	2,01E-07	6,2E-09	1091	2334	4666	12663	over
4	GO:0007049	cell cycle	228	228	GO:0044763	single-organism cellul	P	5,58E-07	2,2E-08	228	369	5529	14628	over
6	GO:0006259	DNA metabolic process	235	235	GO:0090304, GO:0044260	nucleic acid metabolic	P	0,000121	1,2E-05	235	429	5522	14568	over
5	GO:0090304	nucleic acid metabolic process	235	141	GO:0043170, GO:0006139	macromolecule metal	P	0,000121	1,2E-05	235	429	5522	14568	over
2	GO:0050896	response to stimulus	1449	1540,08	GO:0008150	biological_process	P	0,000462	8E-05	1449	3384	4308	11613	over
4	GO:0008219	cell death	103	103	GO:0016265, GO:0044763	death, single-organis	P	0,000574	0,0001	103	163	5654	14834	over
3	GO:0016265	death	103	61,8	GO:0044699	single-organism proce	P	0,000574	0,0001	103	163	5654	14834	over
6	GO:0010468	regulation of gene expression	111	66,6	GO:0010467, GO:0060255	gene expression, regu	P	0,001124	0,00024	111	185	5646	14812	over
4	GO:0019222	regulation of metabolic process	111	23,98	GO:0050789, GO:0008152	regulation of biological	P	0,001124	0,00024	111	185	5646	14812	over
7	GO:0040029	regulation of gene expression, epigenetic	111	111	GO:0010468	regulation of gene exp	P	0,001124	0,00024	111	185	5646	14812	over
5	GO:0060255	regulation of macromolecule metabolic process	111	39,96	GO:0043170, GO:0019222	macromolecule metal	P	0,001124	0,00024	111	185	5646	14812	over
3	GO:0044763	single-organism cellular process	845	545,4	GO:0044699, GO:0009987	single-organism proce	P	0,001169	0,00025	845	1909	4912	13088	over
3	GO:0009605	response to external stimulus	199	161	GO:0050896	response to stimulus	P	0,002523	0,00059	199	383	5558	14614	over
4	GO:0009991	response to extracellular stimulus	76	76	GO:0009605	response to external s	P	0,007939	0,00194	76	125	5681	14872	over
3	GO:0009628	response to abiotic stimulus	543	541,4	GO:0050896	response to stimulus	P	0,023438	0,00664	543	1236	5214	13761	over
3	GO:0009607	response to biotic stimulus	341	341	GO:0050896	response to stimulus	P	0,037608	0,01137	341	755	5416	14242	over
6	GO:0006412	translation	120	120	GO:0034645, GO:0044267, GO:0	cellular macromolecu	P	1,17E-07	2,6E-09	120	549	5637	14448	under
5	GO:0009059	macromolecule biosynthetic process	120	43,2	GO:1901576, GO:0043170	organic substance bio	P	1,17E-07	2,6E-09	120	549	5637	14448	under
5	GO:0034645	cellular macromolecule biosynthetic process	120	72	GO:0044260, GO:0044249, GO:0	cellular macromolecu	P	1,17E-07	2,6E-09	120	549	5637	14448	under
4	GO:0044249	cellular biosynthetic process	120	43,2	GO:0044237, GO:0009058	cellular metabolic pro	P	1,17E-07	2,6E-09	120	549	5637	14448	under
4	GO:1901576	organic substance biosynthetic process	121	26,92	GO:0071704, GO:0009058	organic substance me	P	1,35E-07	3,6E-09	121	549	5636	14448	under
4	GO:0019538	protein metabolic process	1023	597,8	GO:0044238, GO:0043170	primary metabolic pro	P	0,000436	7,2E-05	1023	3030	4734	11967	under
5	GO:0044267	cellular protein metabolic process	765	468	GO:0019538, GO:0044260	protein metabolic pro	P	0,001122	0,00022	765	2297	4992	12700	under
4	GO:0043170	macromolecule metabolic process	1187	467,26	GO:0071704	organic substance me	P	0,01106	0,00275	1187	3381	4570	11616	under
3	GO:0044237	cellular metabolic process	1373	511,78	GO:0008152, GO:0009987	metabolic process, ce	P	0,013569	0,00344	1373	3873	4384	11124	under
5	GO:0010467	gene expression	226	111,96	GO:0043170	macromolecule metal	P	0,027464	0,00818	226	716	5531	14281	under
4	GO:0044260	cellular macromolecule metabolic process	919	421,8	GO:0044237, GO:0043170	cellular metabolic pro	P	0,045142	0,01404	919	2609	4838	12388	under

Table S2 Gene ontology enrichment analysis for genes more expressed in C than in E									
GO-ID	Term	Category	FDR	P-Value	#Test	#Ref	notAnnotTes	notAnnotRe	Over/Under
GO:0005856	cytoskeleton	C	3,61E-14	1,61E-15	55	166	1515	19018	over
GO:0005634	nucleus	C	1,08E-02	3,82E-03	263	2695	1307	16489	over
GO:0044444	cytoplasmic part	C	2,22E-18	6,52E-20	314	5877	1256	13307	under
GO:0016020	membrane	C	3,26E-15	1,10E-16	255	4845	1315	14339	under
GO:0005737	cytoplasm	C	1,80E-12	1,51E-13	420	6879	1150	12305	under
GO:0009536	plastid	C	2,42E-08	2,34E-09	146	2786	1424	16398	under
GO:0030529	ribonucleoprotein complex	C	2,32E-06	3,32E-07	10	457	1560	18727	under
GO:0005840	ribosome	C	2,32E-06	3,32E-07	10	457	1560	18727	under
GO:0032991	macromolecular complex	C	2,32E-06	3,32E-07	10	458	1560	18726	under
GO:0005794	Golgi apparatus	C	5,55E-05	1,03E-05	22	623	1548	18561	under
GO:0005886	plasma membrane	C	6,43E-05	1,24E-05	129	2256	1441	16928	under
GO:0071944	cell periphery	C	6,97E-05	1,38E-05	153	2590	1417	16594	under
GO:0009579	thylakoid	C	1,07E-04	2,25E-05	13	446	1557	18738	under
GO:0005829	cytosol	C	2,03E-04	4,35E-05	64	1263	1506	17921	under
GO:0043231	intracellular membrane-bounded organelle	C	3,97E-04	9,50E-05	500	7049	1070	12135	under
GO:0043227	membrane-bounded organelle	C	3,97E-04	9,50E-05	500	7049	1070	12135	under
GO:0044464	cell part	C	6,07E-04	1,56E-04	724	9800	846	9384	under
GO:0005623	cell	C	9,57E-04	2,61E-04	734	9889	836	9295	under
GO:0044424	intracellular part	C	3,80E-03	1,13E-03	626	8465	944	10719	under
GO:0005768	endosome	C	7,05E-03	2,28E-03	5	206	1565	18978	under
GO:0005622	intracellular	C	7,43E-03	2,44E-03	659	8816	911	10368	under
GO:0043229	intracellular organelle	C	1,02E-02	3,47E-03	540	7310	1030	11874	under
GO:0043226	organelle	C	1,02E-02	3,47E-03	540	7310	1030	11874	under
GO:0005783	endoplasmic reticulum	C	3,30E-02	1,33E-02	24	481	1546	18703	under
GO:0005739	mitochondrion	C	3,59E-02	1,48E-02	94	1468	1476	17716	under
GO:0016462	pyrophosphatase activity	F	3,61E-14	2,12E-15	31	47	1539	19137	over
GO:0003774	motor activity	F	3,61E-14	2,12E-15	31	47	1539	19137	over
GO:0017111	nucleoside-triphosphatase activity	F	3,61E-14	2,12E-15	31	47	1539	19137	over
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	F	3,61E-14	2,12E-15	31	47	1539	19137	over
GO:0016817	hydrolase activity, acting on acid anhydrides	F	3,61E-14	2,12E-15	31	47	1539	19137	over
GO:0036094	small molecule binding	F	8,68E-14	6,20E-15	459	3944	1111	15240	over
GO:1901265	nucleoside phosphate binding	F	8,68E-14	6,20E-15	459	3944	1111	15240	over
GO:0000166	nucleotide binding	F	8,68E-14	6,20E-15	459	3944	1111	15240	over
GO:0097159	organic cyclic compound binding	F	7,87E-13	6,28E-14	624	5841	946	13343	over
GO:1901363	heterocyclic compound binding	F	7,87E-13	6,28E-14	624	5841	946	13343	over
GO:0005488	binding	F	5,95E-07	6,75E-08	1028	11235	542	7949	over
GO:0003677	DNA binding	F	6,43E-05	1,24E-05	159	1344	1411	17840	over
GO:0005515	protein binding	F	5,24E-03	1,67E-03	405	4275	1165	14909	over
GO:0003676	nucleic acid binding	F	3,13E-02	1,24E-02	236	2455	1334	16729	over
GO:0005198	structural molecule activity	F	3,57E-06	5,56E-07	12	491	1558	18693	under
GO:0005215	transporter activity	F	5,15E-05	9,09E-06	56	1191	1514	17993	under

Table S2 Gene ontology enrichment analysis for genes more expressed in C than in E									
GO:0003824	catalytic activity	F	8,78E-04	2,36E-04	712	9629	858	9555	under
GO:0016301	kinase activity	F	1,04E-02	3,61E-03	104	1675	1466	17509	under
GO:0016772	transferase activity, transferring phosphorus-containing groups	F	1,04E-02	3,61E-03	104	1675	1466	17509	under
GO:0016740	transferase activity	F	4,60E-02	1,94E-02	265	3699	1305	15485	under
GO:0006259	DNA metabolic process	P	4,50E-37	1,89E-39	157	507	1413	18677	over
GO:0007049	cell cycle	P	3,71E-35	3,12E-37	144	453	1426	18731	over
GO:0010468	regulation of gene expression	P	1,39E-19	3,52E-21	75	221	1495	18963	over
GO:0019222	regulation of metabolic process	P	1,39E-19	3,52E-21	75	221	1495	18963	over
GO:0040029	regulation of gene expression, epigenetic	P	1,39E-19	3,52E-21	75	221	1495	18963	over
GO:0060255	regulation of macromolecule metabolic process	P	1,39E-19	3,52E-21	75	221	1495	18963	over
GO:0090304	nucleic acid metabolic process	P	2,47E-12	2,18E-13	240	1766	1330	17418	over
GO:0006950	response to stress	P	4,54E-10	4,19E-11	356	3069	1214	16115	over
GO:0006807	nitrogen compound metabolic process	P	1,84E-06	2,17E-07	403	3840	1167	15344	over
GO:1901360	organic cyclic compound metabolic process	P	2,32E-06	3,52E-07	345	3216	1225	15968	over
GO:0046483	heterocycle metabolic process	P	2,32E-06	3,52E-07	345	3216	1225	15968	over
GO:0006139	nucleobase-containing compound metabolic process	P	2,32E-06	3,52E-07	345	3216	1225	15968	over
GO:0006725	cellular aromatic compound metabolic process	P	2,32E-06	3,52E-07	345	3216	1225	15968	over
GO:0034641	cellular nitrogen compound metabolic process	P	2,32E-06	3,52E-07	345	3216	1225	15968	over
GO:0050896	response to stimulus	P	1,19E-05	1,96E-06	444	4389	1126	14795	over
GO:0016043	cellular component organization	P	2,54E-05	4,38E-06	245	2220	1325	16964	over
GO:0071840	cellular component organization or biogenesis	P	2,54E-05	4,38E-06	245	2220	1325	16964	over
GO:0009908	flower development	P	8,36E-05	1,72E-05	72	497	1498	18687	over
GO:0048367	shoot system development	P	8,36E-05	1,72E-05	72	497	1498	18687	over
GO:0061458	reproductive system development	P	2,09E-04	4,75E-05	72	515	1498	18669	over
GO:0048608	reproductive structure development	P	2,09E-04	4,75E-05	72	515	1498	18669	over
GO:0003006	developmental process involved in reproduction	P	2,09E-04	4,75E-05	72	515	1498	18669	over
GO:0048731	system development	P	2,11E-04	4,87E-05	72	516	1498	18668	over
GO:0044763	single-organism cellular process	P	4,38E-04	1,07E-04	260	2494	1310	16690	over
GO:0044767	single-organism developmental process	P	1,88E-03	5,36E-04	98	819	1472	18365	over
GO:0044699	single-organism process	P	1,89E-03	5,47E-04	357	3662	1213	15522	over
GO:0009791	post-embryonic development	P	2,03E-03	5,96E-04	138	1240	1432	17944	over
GO:0000003	reproduction	P	8,25E-03	2,74E-03	144	1356	1426	17828	over
GO:0044707	single-multicellular organism process	P	1,65E-02	6,04E-03	217	2197	1353	16987	over
GO:0007275	multicellular organismal development	P	2,22E-02	8,20E-03	212	2157	1358	17027	over
GO:0022414	reproductive process	P	2,38E-02	9,00E-03	83	749	1487	18435	over
GO:0032501	multicellular organismal process	P	3,13E-02	1,25E-02	221	2285	1349	16899	over
GO:0032502	developmental process	P	3,32E-02	1,35E-02	219	2263	1351	16921	over
GO:0050789	regulation of biological process	P	4,40E-02	1,83E-02	129	1269	1441	17915	over
GO:0009838	abscission	P	4,60E-02	1,95E-02	6	23	1564	19161	over
GO:0006810	transport	P	3,97E-07	4,34E-08	137	2563	1433	16621	under
GO:0051234	establishment of localization	P	3,97E-07	4,34E-08	137	2563	1433	16621	under
GO:0051179	localization	P	3,97E-07	4,34E-08	137	2563	1433	16621	under

Table S2 Gene ontology enrichment analysis for genes more expressed in C than in E									
GO:0005975	carbohydrate metabolic process	P	7,32E-06	1,17E-06	94	1826	1476	17358	under
GO:0006091	generation of precursor metabolites and energy	P	5,16E-05	9,33E-06	16	518	1554	18666	under
GO:0008152	metabolic process	P	4,49E-04	1,11E-04	845	11290	725	7894	under
GO:0009056	catabolic process	P	4,49E-04	1,13E-04	136	2267	1434	16917	under
GO:0007154	cell communication	P	6,18E-04	1,61E-04	74	1369	1496	17815	under
GO:0009607	response to biotic stimulus	P	6,19E-04	1,64E-04	52	1044	1518	18140	under
GO:0006629	lipid metabolic process	P	1,17E-03	3,26E-04	74	1346	1496	17838	under
GO:0006412	translation	P	1,59E-03	4,47E-04	28	641	1542	18543	under
GO:0051716	cellular response to stimulus	P	3,97E-03	1,24E-03	59	1084	1511	18100	under
GO:0050794	regulation of cellular process	P	3,97E-03	1,24E-03	59	1084	1511	18100	under
GO:0007165	signal transduction	P	3,97E-03	1,24E-03	59	1084	1511	18100	under
GO:0015979	photosynthesis	P	4,00E-03	1,26E-03	10	309	1560	18875	under
GO:0044700	single organism signaling	P	1,22E-02	4,39E-03	64	1108	1506	18076	under
GO:0023052	signaling	P	1,22E-02	4,39E-03	64	1108	1506	18076	under
GO:0009605	response to external stimulus	P	2,30E-02	8,61E-03	28	554	1542	18630	under
GO:0065008	regulation of biological quality	P	2,45E-02	9,59E-03	8	234	1562	18950	under
GO:0019725	cellular homeostasis	P	2,45E-02	9,59E-03	8	234	1562	18950	under
GO:0042592	homeostatic process	P	2,45E-02	9,59E-03	8	234	1562	18950	under
GO:0044710	single-organism metabolic process	P	4,76E-02	2,04E-02	157	2294	1413	16890	under

Table S3 Gene ontology enrichment analysis for genes more expressed in E than in C									
GO-ID	Term	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
GO:0005576	extracellular region	C	5,92E-11	4,97E-13	172	675	2291	17616	over
GO:0005618	cell wall	C	3,41E-03	1,08E-04	103	504	2360	17787	over
GO:0030312	external encapsulating structure	C	3,41E-03	1,15E-04	104	511	2359	17780	over
GO:0003824	catalytic activity	F	1,14E-14	4,77E-17	1421	8920	1042	9371	over
GO:0030246	carbohydrate binding	F	1,02E-04	1,28E-06	50	162	2413	18129	over
GO:0016740	transferase activity	F	6,50E-04	1,37E-05	549	3415	1914	14876	over
GO:0016301	kinase activity	F	4,59E-02	3,09E-03	248	1531	2215	16760	over
GO:0016772	transferase activity, transferring phosphorus-containing groups	F	4,59E-02	3,09E-03	248	1531	2215	16760	over
GO:0009607	response to biotic stimulus	P	5,61E-04	9,42E-06	177	919	2286	17372	over
GO:0009991	response to extracellular stimulus	P	1,55E-03	3,90E-05	44	157	2419	18134	over
GO:0009605	response to external stimulus	P	7,65E-03	3,24E-04	97	485	2366	17806	over
GO:0016265	death	P	7,65E-03	3,53E-04	51	215	2412	18076	over
GO:0008219	cell death	P	7,65E-03	3,53E-04	51	215	2412	18076	over
GO:0009875	pollen-pistil interaction	P	1,05E-02	5,75E-04	22	67	2441	18224	over
GO:0048610	cellular process involved in reproduction	P	1,05E-02	5,75E-04	22	67	2441	18224	over
GO:0008152	metabolic process	P	3,66E-02	2,16E-03	1506	10629	957	7662	over

ANNEXES

Table S4 Gene ontology enrichment analysis for genes belonging to divergence regulatory category

<i>cis</i> -regulatory category														
LevelGO	Term(Acc)	Term(Name)	#Seq	Score	Parents(Acc)	Parents(Name)	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
7	GO:0044444	cytoplasmic part	702	576,36	GO:0005737, GO:0005737	cytoplasm, intracellular part	C	4,72066E-30	1,98347E-32	702	5489	927	13636	over
6	GO:0005737	cytoplasm	785	443,82	GO:0044424	intracellular part	C	1,66656E-27	1,62183E-29	785	6514	844	12611	over
3	GO:0044464	cell part	1042	417,28	GO:0005623, GO:0005623	cell, cellular component	C	1,66656E-27	2,1007E-29	1042	9482	587	9643	over
4	GO:0005623	intracellular	960	288	GO:0044464	cell part	C	1,82279E-27	3,06351E-29	960	8515	669	10610	over
2	GO:0005623	cell	1046	262,37	GO:0005575	cellular component	C	8,1764E-27	1,71773E-28	1046	9577	583	9548	over
5	GO:0044424	intracellular part	925	356,66	GO:0044464, GO:0044464	cell part, intracellular part	C	1,78196E-26	4,49235E-28	925	8166	704	10959	over
7	GO:0043231	intracellular membrane-bounded organelle	789	598,02	GO:0043229, GO:0043229	intracellular organelle	C	6,48048E-24	2,17831E-25	789	6760	840	12365	over
3	GO:0043227	membrane-bounded organelle	789	360,88	GO:0043226	organelle	C	6,48048E-24	2,17831E-25	789	6760	840	12365	over
6	GO:0043229	intracellular organelle	809	383,67	GO:0043226, GO:0043226	organelle, intracellular part	C	6,63421E-23	2,78748E-24	809	7041	820	12084	over
2	GO:0043226	organelle	809	230,2	GO:0005575	cellular component	C	6,63421E-23	2,78748E-24	809	7041	820	12084	over
8	GO:0009536	plastid	351	351	GO:0044444, GO:0044444	cytoplasmic part	C	3,3551E-16	1,55068E-17	351	2581	1278	16544	over
2	GO:0016020	membrane	478	476,2	GO:0005575	cellular component	C	3,2104E-05	2,69781E-06	478	4622	1151	14503	over
8	GO:0005783	endoplasmic reticulum	67	67	GO:0044444, GO:0044444	cytoplasmic part	C	0,000133228	1,56738E-05	67	438	1562	18687	over
6	GO:0009579	thylakoid	62	62	GO:0044424	intracellular part	C	0,000152411	1,85711E-05	62	397	1567	18728	over
8	GO:0005829	cytosol	141	141	GO:0044444	cytoplasmic part	C	0,000865803	0,000116411	141	1186	1488	17939	over
8	GO:0042579	microbody	31	18,6	GO:0044444, GO:0044444	cytoplasmic part	C	0,001945043	0,000290887	31	173	1598	18952	over
8	GO:0005739	mitochondrion	159	159	GO:0044444, GO:0044444	cytoplasmic part	C	0,002034571	0,000343313	159	1403	1470	17722	over
4	GO:0071944	cell periphery	261	158,88	GO:0044464	cell part	C	0,002155947	0,000380461	261	2482	1368	16643	over
5	GO:0005886	plasma membrane	227	227	GO:0016020, GO:0016020	membrane, cell part	C	0,004674889	0,000942835	227	2158	1402	16967	over
8	GO:0005794	Golgi apparatus	71	71	GO:0044444, GO:0044444	cytoplasmic part	C	0,01055704	0,002350937	71	574	1558	18551	over
8	GO:0005773	vacuole	78	78	GO:0044444, GO:0044444	cytoplasmic part	C	0,02466264	0,006424721	78	672	1551	18453	over
8	GO:0005840	ribosome	51	51	GO:0044444, GO:0044444	cytoplasmic part	C	0,03727	0,01033538	51	416	1578	18709	over
6	GO:0030529	ribonucleoprotein complex	51	30,6	GO:0032991, GO:0032991	macromolecular complex	C	0,03727	0,01033538	51	416	1578	18709	over
2	GO:0032991	macromolecular complex	51	18,36	GO:0005575	cellular component	C	0,03815983	0,01074247	51	417	1578	18708	over
9	GO:0005777	peroxisome	31	31	GO:0042579	microbody	C	0,001945043	0,000290887	31	173	1598	18952	over
3	GO:0044710	single-organism metabolic process	280	156	GO:0008152	metabolic process	P	3,4069E-10	1,71776E-11	280	2171	1349	16954	over
3	GO:0044238	primary metabolic process	831	490,75	GO:0008152	metabolic process	P	2,14316E-07	1,17063E-08	831	8376	798	10749	over
3	GO:0071704	organic substance metabolic process	831	388,52	GO:0008152	metabolic process	P	2,34613E-07	1,38008E-08	831	8383	798	10742	over
2	GO:0009987	cellular process	979	858,49	GO:0008150	biological process	P	3,79402E-07	2,39119E-08	979	10149	650	8976	over
3	GO:0009058	biosynthetic process	478	354,07	GO:0008152	metabolic process	P	8,54279E-07	5,74305E-08	478	4467	1151	14658	over
3	GO:0009628	response to abiotic stimulus	199	198,6	GO:0050896	response to stimulus	P	1,65124E-06	1,17946E-07	199	1580	1430	17545	over
4	GO:0006629	lipid metabolic process	164	164	GO:0071704, GO:0071704	organic substance metabolic process	P	3,11083E-06	2,35273E-07	164	1256	1465	17869	over
2	GO:0008152	metabolic process	1047	982,61	GO:0008150	biological process	P	4,44067E-06	3,54507E-07	1047	11088	582	8037	over
4	GO:0005975	carbohydrate metabolic process	204	204	GO:0071704, GO:0071704	organic substance metabolic process	P	3,56661E-05	3,14701E-06	204	1716	1425	17409	over
7	GO:0019752	carboxylic acid metabolic process	116	0	GO:0043436	oxoacid metabolic process	P	5,37957E-05	6,10287E-06	116	869	1513	18256	over
6	GO:0043436	oxoacid metabolic process	116	0	GO:0006082	organic acid metabolic process	P	5,37957E-05	6,10287E-06	116	869	1513	18256	over
5	GO:0006082	organic acid metabolic process	116	0	GO:0071704, GO:0071704	organic substance metabolic process	P	5,37957E-05	6,10287E-06	116	869	1513	18256	over
4	GO:0044281	small molecule metabolic process	116	0	GO:0044710	single-organism metabolic process	P	5,37957E-05	6,10287E-06	116	869	1513	18256	over
4	GO:1901564	organonitrogen compound metabolic process	116	0	GO:0071704, GO:0071704	organic substance metabolic process	P	5,37957E-05	6,10287E-06	116	869	1513	18256	over
3	GO:0009056	catabolic process	241	241	GO:0008152	metabolic process	P	0,000196859	2,48141E-05	241	2162	1388	16963	over

Table S4 Gene ontology enrichment analysis for genes belonging to divergence regulatory category

4	GO:0006091	generation of precursor metabolites and energy	69	69	GO:0044237	cellular metabol	P	0,000215426	2,80597E-05	69	465	1560	18660	over	
4	GO:0006810	transport	258	258	GO:0051234	establishment o	P	0,001945043	0,000318726	258	2442	1371	16683	over	
4	GO:0007154	cell communication	149	103	GO:0044763	single-organism	P	0,001945043	0,000281827	149	1294	1480	17831	over	
3	GO:0006807	nitrogen compound metabolic process	388	83,38	GO:0008152	metabolic proce	P	0,001945043	0,000300897	388	3855	1241	15270	over	
3	GO:0051234	establishment of localization	258	154,8	GO:0008150, GO	biological_proce	P	0,001945043	0,000318726	258	2442	1371	16683	over	
2	GO:0051179	localization	258	92,88	GO:0008150	biological_proce	P	0,001945043	0,000318726	258	2442	1371	16683	over	
4	GO:0015979	photosynthesis	43	43	GO:0044237	cellular metabol	P	0,002034571	0,000350493	43	276	1586	18849	over	
4	GO:0019748	secondary metabolic process	96	96	GO:0044710	single-organism	P	0,002555137	0,000461642	96	777	1533	18348	over	
3	GO:0044237	cellular metabolic process	595	215,23	GO:0008152, GO	metabolic proce	P	0,003678742	0,000680104	595	6231	1034	12894	over	
5	GO:0007165	signal transduction	119	119	GO:0044700, GO	single organism	P	0,00431007	0,000851148	119	1024	1510	18101	over	
4	GO:0050794	regulation of cellular process	119	71,4	GO:0009987, GO	cellular process,	P	0,00431007	0,000851148	119	1024	1510	18101	over	
4	GO:0051716	cellular response to stimulus	119	71,4	GO:0050896, GO	response to stim	P	0,00431007	0,000851148	119	1024	1510	18101	over	
3	GO:0044700	single organism signaling	121	72,6	GO:0023052, GO	signaling, single-	P	0,005001521	0,00105074	121	1051	1508	18074	over	
2	GO:0023052	signaling	121	43,56	GO:0008150	biological_proce	P	0,005001521	0,00105074	121	1051	1508	18074	over	
2	GO:0065007	biological regulation	159	32,74	GO:0008150	biological_proce	P	0,007643769	0,00163795	159	1460	1470	17665	over	
3	GO:0050789	regulation of biological process	139	45,56	GO:0065007, GO	biological regula	P	0,009230386	0,002016723	139	1259	1490	17866	over	
3	GO:0044763	single-organism cellular process	254	156,76	GO:0044699, GO	single-organism	P	0,01164231	0,002641533	254	2500	1375	16625	over	
2	GO:0044699	single-organism process	356	198,85	GO:0008150	biological_proce	P	0,02108481	0,00487254	356	3663	1273	15462	over	
5	GO:0006139	nucleobase-containing compound metabolic process	317	231,6	GO:0034641, GO	cellular nitrogen	P	0,02419465	0,006201149	317	3244	1312	15881	over	
4	GO:0006725	cellular aromatic compound metabolic process	317	138,96	GO:0044237	cellular metabol	P	0,02419465	0,006201149	317	3244	1312	15881	over	
4	GO:0034641	cellular nitrogen compound metabolic process	317	138,96	GO:0006807, GO	nitrogen compou	P	0,02419465	0,006201149	317	3244	1312	15881	over	
4	GO:0046483	heterocycle metabolic process	317	138,96	GO:0044237	cellular metabol	P	0,02419465	0,006201149	317	3244	1312	15881	over	
4	GO:1901360	organic cyclic compound metabolic process	317	138,96	GO:0071704	organic substanc	P	0,02419465	0,006201149	317	3244	1312	15881	over	
3	GO:0009605	response to external stimulus	63	52,2	GO:0050896	response to stim	P	0,02419465	0,005796768	63	519	1566	18606	over	
4	GO:0009653	anatomical structure morphogenesis	102	102	GO:0048856, GO	anatomical struc	P	0,03600701	0,009682559	102	933	1527	18192	over	
3	GO:0048856	anatomical structure development	151	101,4	GO:0032502	developmental p	P	0,03937876	0,01125107	151	1459	1478	17666	over	
7	GO:0006412	translation	70	70	GO:0044267, GO	cellular protein t	P	0,03127085	0,008277577	70	599	1559	18526	over	
Trans-regulatory category															
Level	GO	Term(Acc)	Term(Name)	#Seq	Score	Parents(Acc)	Parents(Name)	Category	FDR	P-Value	#Test	#Ref	#notAnnotTest	#notAnnotRef	Over/Under
3	GO:0044464	cell part		1166	513,49	GO:0005623, GO	cell, cellular_cor	C	1,02456E-32	4,30486E-35	1166	9358	641	9589	over
2	GO:0005623	cell		1170	318,1	GO:0005575	cellular_compor	C	6,13395E-32	5,15458E-34	1170	9453	637	9494	over
4	GO:0005622	intracellular		1061	362,09	GO:0044464	cell part	C	2,37482E-29	2,99348E-31	1061	8414	746	10533	over
5	GO:0044424	intracellular part		1013	416,82	GO:0044464, GO	cell part, intrace	C	5,59146E-26	9,3974E-28	1013	8078	794	10869	over
6	GO:0043229	intracellular organelle		895	424,63	GO:0043226, GO	organelle, intrac	C	1,41746E-24	3,57342E-26	895	6955	912	11992	over
2	GO:0043226	organelle		895	254,78	GO:0005575	cellular_compor	C	1,41746E-24	3,57342E-26	895	6955	912	11992	over
7	GO:0043231	intracellular membrane-bounded organelle		850	640,15	GO:0043229, GO	intracellular org	C	8,11215E-21	2,72677E-22	850	6699	957	12248	over
3	GO:0043227	membrane-bounded organelle		850	386,39	GO:0043226	organelle	C	8,11215E-21	2,72677E-22	850	6699	957	12248	over
6	GO:0005737	cytoplasm		807	494,68	GO:0044424	intracellular par	C	9,57759E-17	3,62178E-18	807	6492	1000	12455	over
7	GO:0044444	cytoplasmic part		698	592,8	GO:0005737, GO	cytoplasm, intra	C	1,40289E-15	5,8945E-17	698	5493	1109	13454	over
8	GO:0005856	cytoskeleton		56	56	GO:0043232	intracellular nor	C	2,41236E-12	1,52039E-13	56	165	1751	18782	over
2	GO:0016020	membrane		564	537,2	GO:0005575	cellular_compor	C	3,01365E-10	2,02598E-11	564	4536	1243	14411	over

Table S4 Gene ontology enrichment analysis for genes belonging to divergence regulatory category														
5	GO:0005886	plasma membrane	292	292	GO:0016020, GO	membrane, cell	C	4,69064E-09	4,53297E-10	292	2093	1515	16854	over
4	GO:0071944	cell periphery	326	196,2	GO:0044464	cell part	C	1,08314E-08	1,13775E-09	326	2417	1481	16530	over
6	GO:0009579	thylakoid	76	76	GO:0044424	intracellular par	C	4,46965E-07	5,25841E-08	76	383	1731	18564	over
7	GO:0043232	intracellular non-membrane-bounded organelle	130	84,6	GO:0043229, GO	intracellular org	C	5,69729E-07	7,18146E-08	130	803	1677	18144	over
3	GO:0043228	non-membrane-bounded organelle	130	50,76	GO:0043226	organelle	C	5,69729E-07	7,18146E-08	130	803	1677	18144	over
8	GO:0005634	nucleus	323	277,91	GO:0043231	intracellular me	C	3,79521E-05	7,49474E-06	323	2635	1484	16312	over
8	GO:0005829	cytosol	153	153	GO:0044444	cytoplasmic part	C	0,00088476	0,000282529	153	1174	1654	17773	over
8	GO:0005794	Golgi apparatus	81	81	GO:0044444, GO	cytoplasmic part	C	0,00227495	0,000814648	81	564	1726	18383	over
8	GO:0005773	vacuole	91	91	GO:0044444, GO	cytoplasmic part	C	0,003137926	0,001186611	91	659	1716	18288	over
8	GO:0005783	endoplasmic reticulum	60	60	GO:0044444, GO	cytoplasmic part	C	0,02513152	0,01309373	60	445	1747	18502	over
8	GO:0005768	endosome	28	28	GO:0044444, GO	cytoplasmic part	C	0,04772258	0,02606696	28	183	1779	18764	over
8	GO:0009536	plastid	371	371	GO:0044444, GO	cytoplasmic part	C	1,57874E-13	7,29671E-15	371	2561	1436	16386	over
7	GO:0017111	nucleoside-triphosphatase activity	28	16,8	GO:0016462	pyrophosphatas	F	3,1524E-10	2,78153E-11	28	50	1779	18897	over
6	GO:0016462	pyrophosphatase activity	28	10,08	GO:0016818	hydrolase activit	F	3,1524E-10	2,78153E-11	28	50	1779	18897	over
5	GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-conta	28	6,05	GO:0016817	hydrolase activit	F	3,1524E-10	2,78153E-11	28	50	1779	18897	over
4	GO:0016817	hydrolase activity, acting on acid anhydrides	28	3,63	GO:0016787	hydrolase activit	F	3,1524E-10	2,78153E-11	28	50	1779	18897	over
3	GO:0005515	protein binding	477	477,6	GO:0005488	binding	F	0,000217763	5,39832E-05	477	4203	1330	14744	over
5	GO:0003677	DNA binding	167	167	GO:0003676	nucleic acid bind	F	0,002378869	0,000869587	167	1336	1640	17611	over
4	GO:0003676	nucleic acid binding	279	184,16	GO:0097159, GO	organic cyclic co	F	0,003663885	0,001416292	279	2412	1528	16535	over
3	GO:0004871	signal transducer activity	26	26	GO:0060089	molecular transce	F	0,0071113	0,002838544	26	136	1781	18811	over
2	GO:0060089	molecular transducer activity	26	15,6	GO:0003674	molecular functi	F	0,0071113	0,002838544	26	136	1781	18811	over
2	GO:0003824	catalytic activity	830	612,77	GO:0003674	molecular functi	F	0,00172129	0,00056412	830	9511	977	9436	under
5	GO:0016301	kinase activity	117	117	GO:0016772	transferase activ	F	0,002068772	0,000695386	117	1662	1690	17285	under
4	GO:0016772	transferase activity, transferring phosphorus-containing groups	117	70,2	GO:0016740	transferase activ	F	0,002068772	0,000695386	117	1662	1690	17285	under
3	GO:0016740	transferase activity	296	223,12	GO:0003824	catalytic activity	F	0,004921998	0,001923302	296	3668	1511	15279	under
8	GO:0003774	motor activity	28	28	GO:0017111	nucleoside-triph	F	3,1524E-10	2,78153E-11	28	50	1779	18897	over
2	GO:0009987	cellular process	1124	981,39	GO:0008150	biological_proce	P	2,83749E-13	1,43067E-14	1124	10004	683	8943	over
3	GO:0016043	cellular component organization	318	318	GO:0071840	cellular compon	P	1,47833E-12	8,69604E-14	318	2147	1489	16800	over
2	GO:0071840	cellular component organization or biogenesis	318	190,8	GO:0008150	biological_proce	P	1,47833E-12	8,69604E-14	318	2147	1489	16800	over
2	GO:0044699	single-organism process	455	261,66	GO:0008150	biological_proce	P	2,48344E-09	2,29561E-10	455	3564	1352	15383	over
3	GO:0009058	biosynthetic process	540	370,76	GO:0008152	metabolic proce	P	6,88534E-09	6,9432E-10	540	4405	1267	14542	over
3	GO:0044763	single-organism cellular process	323	210,8	GO:0044699, GO	single-organism	P	5,36767E-08	5,86384E-09	323	2431	1484	16516	over
3	GO:0006807	nitrogen compound metabolic process	467	88,14	GO:0008152	metabolic proce	P	5,52475E-08	6,26758E-09	467	3776	1340	15171	over
5	GO:0006139	nucleobase-containing compound metabolic process	394	244,84	GO:0034641, GO	cellular nitrogen	P	6,62234E-07	9,73874E-08	394	3167	1413	15780	over
4	GO:0006725	cellular aromatic compound metabolic process	394	146,9	GO:0044237	cellular metabol	P	6,62234E-07	9,73874E-08	394	3167	1413	15780	over
4	GO:0034641	cellular nitrogen compound metabolic process	394	146,9	GO:0006807, GO	nitrogen compou	P	6,62234E-07	9,73874E-08	394	3167	1413	15780	over
4	GO:0046483	heterocycle metabolic process	394	146,9	GO:0044237	cellular metabol	P	6,62234E-07	9,73874E-08	394	3167	1413	15780	over
4	GO:1901360	organic cyclic compound metabolic process	394	146,9	GO:0071704	organic substanc	P	6,62234E-07	9,73874E-08	394	3167	1413	15780	over
5	GO:0009791	post-embryonic development	177	178,2	GO:0007275, GO	multicellular org	P	6,99881E-07	1,05864E-07	177	1201	1630	17746	over
3	GO:0009628	response to abiotic stimulus	215	214,6	GO:0050896	response to stim	P	2,51538E-06	3,98642E-07	215	1564	1592	17383	over
2	GO:0065007	biological regulation	199	38,12	GO:0008150	biological_proce	P	2,51538E-06	4,01615E-07	199	1420	1608	17527	over

Table S4 Gene ontology enrichment analysis for genes belonging to divergence regulatory category														
4	GO:0007275	multicellular organismal development	271	279,7	GO:0044707, GO	single-multicellu	P	7,61471E-06	1,24779E-06	271	2098	1536	16849	over
2	GO:0032502	developmental process	282	249,89	GO:0008150	biological_proce	P	8,18616E-06	1,37583E-06	282	2200	1525	16747	over
3	GO:0048856	anatomical structure development	194	143,28	GO:0032502	developmental p	P	1,17157E-05	2,01826E-06	194	1416	1613	17531	over
3	GO:0044237	cellular metabolic process	685	237,7	GO:0008152, GO	metabolic proce	P	1,51532E-05	2,6741E-06	685	6141	1122	12806	over
3	GO:0044707	single-multicellular organism process	273	172,62	GO:0032501, GO	multicellular org	P	1,7845E-05	3,2241E-06	273	2141	1534	16806	over
3	GO:0044767	single-organism developmental process	121	65,74	GO:0044699, GO	single-organism	P	1,90657E-05	3,52475E-06	121	796	1686	18151	over
6	GO:0048367	shoot system development	82	49,2	GO:0044767, GO	single-organism	P	2,69986E-05	5,21821E-06	82	487	1725	18460	over
7	GO:0048608	reproductive structure development	83	49,56	GO:0048856, GO	anatomical struc	P	4,84904E-05	1,01871E-05	83	504	1724	18443	over
6	GO:0061458	reproductive system development	83	29,74	GO:0048731	system developr	P	4,84904E-05	1,01871E-05	83	504	1724	18443	over
4	GO:0003006	developmental process involved in reproduction	83	29,74	GO:0022414, GO	reproductive pro	P	4,84904E-05	1,01871E-05	83	504	1724	18443	over
5	GO:0048731	system development	83	17,84	GO:0048856, GO	anatomical struc	P	4,89587E-05	1,04912E-05	83	505	1724	18442	over
4	GO:0006810	transport	296	296	GO:0051234	establishment o	P	6,72016E-05	1,52474E-05	296	2404	1511	16543	over
3	GO:0051234	establishment of localization	296	177,6	GO:0008150, GO	biological_proce	P	6,72016E-05	1,52474E-05	296	2404	1511	16543	over
2	GO:0051179	localization	296	106,56	GO:0008150	biological_proce	P	6,72016E-05	1,52474E-05	296	2404	1511	16543	over
3	GO:0050789	regulation of biological process	167	51,29	GO:0065007, GO	biological regula	P	9,79446E-05	2,26343E-05	167	1231	1640	17716	over
4	GO:0009653	anatomical structure morphogenesis	129	129	GO:0048856, GO	anatomical struc	P	0,000114845	2,70224E-05	129	906	1678	18041	over
5	GO:0010467	gene expression	242	53,28	GO:0043170	macromolecule	P	0,000135703	3,25002E-05	242	1924	1565	17023	over
2	GO:0032501	multicellular organismal process	274	107,61	GO:0008150	biological_proce	P	0,000179336	4,37037E-05	274	2232	1533	16715	over
3	GO:0009607	response to biotic stimulus	134	134	GO:0050896	response to stim	P	0,000218646	5,51209E-05	134	962	1673	17985	over
4	GO:1901576	organic substance biosynthetic process	230	14,61	GO:0071704, GO	organic substanc	P	0,000254526	6,52356E-05	230	1837	1577	17110	over
8	GO:0032774	RNA biosynthetic process	174	0	GO:0009059, GO	macromolecule	P	0,000315096	9,53232E-05	174	1331	1633	17616	over
7	GO:0016070	RNA metabolic process	174	0	GO:0044260, GO	cellular macrom	P	0,000315096	9,53232E-05	174	1331	1633	17616	over
6	GO:0034645	cellular macromolecule biosynthetic process	229	37,8	GO:0009059, GO	macromolecule	P	0,000315096	9,13005E-05	229	1837	1578	17110	over
6	GO:0034654	nucleobase-containing compound biosynthetic process	174	0	GO:0019438, GO	aromatic compo	P	0,000315096	9,53232E-05	174	1331	1633	17616	over
5	GO:0009059	macromolecule biosynthetic process	229	22,68	GO:1901576, GO	organic substanc	P	0,000315096	9,13005E-05	229	1837	1578	17110	over
5	GO:0018130	heterocycle biosynthetic process	174	0	GO:0046483, GO	heterocycle met	P	0,000315096	9,53232E-05	174	1331	1633	17616	over
5	GO:0019438	aromatic compound biosynthetic process	174	0	GO:0006725, GO	cellular aromati	P	0,000315096	9,53232E-05	174	1331	1633	17616	over
5	GO:0044271	cellular nitrogen compound biosynthetic process	174	0	GO:0034641, GO	cellular nitrogen	P	0,000315096	9,53232E-05	174	1331	1633	17616	over
5	GO:1901362	organic cyclic compound biosynthetic process	174	0	GO:1901576, GO	organic substanc	P	0,000315096	9,53232E-05	174	1331	1633	17616	over
4	GO:0044249	cellular biosynthetic process	229	22,68	GO:0009058, GO	biosynthetic pro	P	0,000315096	9,13005E-05	229	1837	1578	17110	over
4	GO:0007049	cell cycle	80	80	GO:0044763	single-organism	P	0,000378932	0,000116227	80	517	1727	18430	over
4	GO:0006091	generation of precursor metabolites and energy	73	73	GO:0044237	cellular metabol	P	0,000401268	0,000124764	73	461	1734	18486	over
6	GO:0090304	nucleic acid metabolic process	220	41,4	GO:0043170, GO	macromolecule	P	0,000766507	0,000241546	220	1786	1587	17161	over
2	GO:0000003	reproduction	170	117,58	GO:0008150	biological_proce	P	0,000911723	0,000294969	170	1330	1637	17617	over
7	GO:0040029	regulation of gene expression, epigenetic	43	43	GO:0010468	regulation of gen	P	0,00227495	0,000822041	43	253	1764	18694	over
6	GO:0010468	regulation of gene expression	43	25,8	GO:0010467, GO	gene expression	P	0,00227495	0,000822041	43	253	1764	18694	over
5	GO:0060255	regulation of macromolecule metabolic process	43	15,48	GO:0019222, GO	regulation of me	P	0,00227495	0,000822041	43	253	1764	18694	over
4	GO:0019222	regulation of metabolic process	43	9,29	GO:0050789, GO	regulation of bio	P	0,00227495	0,000822041	43	253	1764	18694	over
3	GO:0071704	organic substance metabolic process	870	386,27	GO:0008152	metabolic proce	P	0,00227495	0,000820868	870	8344	937	10603	over
3	GO:0044238	primary metabolic process	869	488,76	GO:0008152	metabolic proce	P	0,002422427	0,000895687	869	8338	938	10609	over
3	GO:0044710	single-organism metabolic process	257	144,6	GO:0008152	metabolic proce	P	0,003134082	0,001171989	257	2194	1550	16753	over
4	GO:0015979	photosynthesis	45	45	GO:0044237	cellular metabol	P	0,003286518	0,00125661	45	274	1762	18673	over
7	GO:0019752	carboxylic acid metabolic process	112	0	GO:0043436	oxoacid metabol	P	0,007324217	0,003123183	112	873	1695	18074	over

ANNEXES

Table S4 Gene ontology enrichment analysis for genes belonging to divergence regulatory category														
6	GO:0043436	oxoacid metabolic process	112	0	GO:0006082	organic acid met	P	0,007324217	0,003123183	112	873	1695	18074	over
5	GO:0006082	organic acid metabolic process	112	0	GO:0071704, GO	organic substance	P	0,007324217	0,003123183	112	873	1695	18074	over
4	GO:0044281	small molecule metabolic process	112	0	GO:0044710	single-organism	P	0,007324217	0,003123183	112	873	1695	18074	over
4	GO:1901564	organonitrogen compound metabolic process	112	0	GO:0071704, GO	organic substance	P	0,007324217	0,003123183	112	873	1695	18074	over
2	GO:0040007	growth	70	59,8	GO:0008150	biological_proce	P	0,007324217	0,00313895	70	498	1737	18449	over
3	GO:0009056	catabolic process	248	248	GO:0008152	metabolic proce	P	0,007934873	0,003434	248	2155	1559	16792	over
3	GO:0044700	single organism signaling	130	78,6	GO:0023052, GO	signaling, single-	P	0,008875713	0,003915756	130	1042	1677	17905	over
2	GO:0023052	signaling	130	47,16	GO:0008150	biological_proce	P	0,008875713	0,003915756	130	1042	1677	17905	over
3	GO:0009719	response to endogenous stimulus	105	105	GO:0050896	response to stim	P	0,00902688	0,004020375	105	815	1702	18132	over
5	GO:0007165	signal transduction	127	127	GO:0044700, GO	single organism :	P	0,00909648	0,004166035	127	1016	1680	17931	over
4	GO:0050794	regulation of cellular process	127	76,2	GO:0009987, GO	cellular process,	P	0,00909648	0,004166035	127	1016	1680	17931	over
4	GO:0051716	cellular response to stimulus	127	76,2	GO:0050896, GO	response to stim	P	0,00909648	0,004166035	127	1016	1680	17931	over
5	GO:0019725	cellular homeostasis	34	34	GO:0044763, GO	single-organism	P	0,01173776	0,005523653	34	208	1773	18739	over
4	GO:0042592	homeostatic process	34	20,4	GO:0065008	regulation of bio	P	0,01173776	0,005523653	34	208	1773	18739	over
3	GO:0065008	regulation of biological quality	34	12,24	GO:0065007	biological regula	P	0,01173776	0,005523653	34	208	1773	18739	over
4	GO:0007154	cell communication	155	104,8	GO:0044763	single-organism	P	0,01208719	0,005738875	155	1288	1652	17659	over
5	GO:0009790	embryo development	60	60	GO:0044767, GO	single-organism	P	0,01477002	0,007091374	60	428	1747	18519	over
4	GO:0043170	macromolecule metabolic process	537	173,78	GO:0071704	organic substance	P	0,01477002	0,007136774	537	5067	1270	13880	over
4	GO:0008219	cell death	36	36	GO:0016265, GO	death, single-org	P	0,01607581	0,008172995	36	230	1771	18717	over
3	GO:0016265	death	36	21,6	GO:0044699	single-organism	P	0,01607581	0,008172995	36	230	1771	18717	over
2	GO:0050896	response to stimulus	466	542,88	GO:0008150	biological_proce	P	0,01855223	0,009509968	466	4367	1341	14580	over
3	GO:0022414	reproductive process	93	21,87	GO:0008150, GO	biological_proce	P	0,0231485	0,0119633	93	739	1714	18208	over
4	GO:0019748	secondary metabolic process	96	96	GO:0044710	single-organism	P	0,03171305	0,01665601	96	777	1711	18170	over
5	GO:0044260	cellular macromolecule metabolic process	442	146,16	GO:0044237, GO	cellular metabol	P	0,03365663	0,01781821	442	4170	1365	14777	over
4	GO:0009991	response to extracellular stimulus	27	27	GO:0009605	response to exte	P	0,04254446	0,02287396	27	174	1780	18773	over
5	GO:0009856	pollination	11	11,2	GO:0044703, GO	multi-organism r	P	0,01502156	0,007510782	11	248	1796	18699	under
4	GO:0044703	multi-organism reproductive process	11	6,72	GO:0022414	reproductive pro	P	0,01502156	0,007510782	11	248	1796	18699	under
3	GO:0044706	multi-multicellular organism process	11	6,72	GO:0032501, GO	multicellular org	P	0,01502156	0,007510782	11	248	1796	18699	under
2	GO:0051704	multi-organism process	11	4,03	GO:0008150	biological_proce	P	0,01502156	0,007510782	11	248	1796	18699	under
8	GO:0009908	flower development	82	82	GO:0048608, GO	reproductive strological proc	P	2,69986E-05	5,21821E-06	82	487	1725	18460	over

POLYPLÔIDIE ET ADAPTATION DES PLANTES : CARACTERISATION ET VARIATION DE L'EXPRESSION DES GENES HOMEOLOGUES CHEZ LE CAFEIER *COFFEA ARABICA*

Résumé : La plupart des espèces végétales sont des polyploïdes récents ou anciens. L'allopolyplôidisation qui résulte d'une hybridation interspécifique couplée à une duplication génomique joue un rôle fondamental dans l'évolution des plantes. Elle provoquerait des changements de l'organisation du génome et de la régulation de l'expression des gènes qui seraient à l'origine de nouvelles aptitudes de ces plantes. Dans ce contexte, l'étude des mécanismes de conciliation des réseaux parentaux de régulation des gènes semble essentielle pour définir leur rôle dans la variation de l'expression des gènes.

Le genre *Coffea*, composé d'espèces diploïdes pouvant s'hybrider et d'une espèce allotétraploïde *C. arabica*, constitue un modèle approprié pour cette étude. L'allopolyplôïde présente deux sous-génomes homéologues peu différenciés, provenant de l'hybridation de *C. eugenioides* et *C. canephora*. A l'inverse de ses espèces parentales qui se caractérisent par des écosystèmes différents et d'amplitude thermique faible, *C. arabica* se développe en tolérant des variations de température plus importantes. Ce travail a consisté à étudier chez *C. arabica* les conséquences de l'hybridation sur la variation de l'expression et de la régulation des gènes et à analyser la variation de l'expression des gènes homéologues et son implication dans les capacités adaptatives.

L'expression et la régulation des gènes au sein d'hybrides F1 entre *C. canephora* et *C. eugenioides* ont été analysées en utilisant la technologie RNA-seq. L'étude de ces hybrides a montré que les espèces parentales se distinguent par une proportion importante de divergences de *trans*-régulation. Pour 77% des gènes différemment exprimés entre les espèces parentales et les hybrides, le niveau d'expression de l'hybride est comparable à celui de l'une des espèces parentales, c'est le phénomène de «level expression dominance». L'analyse de l'expression allélique spécifique a permis de caractériser la régulation de l'expression des gènes, les deux allèles sont exprimés, parfois de façon biaisée vers un génome. L'expression des gènes est déterminée par des combinaisons complexes de divergences de *cis*-régulation des espèces parentales et des effets entrecroisés de leurs facteurs de *trans*-régulation. Le niveau d'expression des gènes dépend de la sur ou sous-régulation simultanée des deux allèles et en particulier, de l'asymétrie des effets des facteurs de *trans*-régulation qui semble être à l'origine du phénomène de «level expression dominance».

Pour l'allopolyplôïde, l'étude de l'expression relative des gènes homéologues à l'échelle du génome a révélé que les gènes homéologues des deux sous-génomes s'inter-régulent et contribuent au transcriptome.

Le transcriptome de jeunes feuilles de *C. arabica* cultivés à deux conditions de températures, chacune convenant à l'une des espèces parentales a été analysé pour étudier la variation de l'expression des gènes homéologues chez *C. arabica* et son implication dans les capacités adaptatives. La contribution des sous-génomes au transcriptome est faiblement modifiée par les conditions de culture et n'apparaît pas directement impliquée dans la capacité de *C. arabica* à tolérer des amplitudes thermiques plus élevées que les espèces parentales.

Ces études ont permis de caractériser la réponse transcriptionnelle à un événement d'hybridation récent et ancien. Elles mettent en évidence les bases génétiques de la variation de l'expression allélique suite à la fusion des réseaux divergents de régulation des gènes d'espèces parentales. Elles ont permis de proposer un modèle de la régulation de l'expression des gènes homéologues pour *C. arabica*.

Mots clés : allopolyplôïdie, hybridation, homéologue, expression allélique spécifique (ASE), *cis- trans*-régulation, transcriptome, RNA-seq, adaptation, caféier.

POLYPLOIDY AND PLANT ADAPTATION : CHARACTERIZATION AND VARIATION OF HOMEOLOGOUS GENE EXPRESSION IN *COFFEA ARABICA*

Abstract : Polyploidy is a prominent mode of speciation and a recurrent process during plant evolution. Allopolyploidization, that involves inter-species hybridization and genome doubling, can induce an extensive array of genomic rearrangements and gene expression changes generating plants with new adaptive abilities. The study of the merger of divergent gene expression regulatory networks seems fundamental to elucidate the role of conciliation processes in the gene expression variations.

The genus *Coffea* that contains diploid species able to hybridize and *C. arabica* a recent allopolyploid between two low divergent diploid species *C. eugenioides* and *C. canephora*, represents an appropriate model for this study. Indeed *C. arabica* can be grown in regions with marked variations in thermal amplitude while the parental species are less adapted to temperature variations. The aims of the present work are, on one hand, the study of the effects of hybridization on the expression and regulation of genes and on the other hand, the analysis of homeologous gene expression variation in response to changing environment.

To examine the immediate effects of hybridization, the expression and regulation of genes in F1 hybrids between *C. canephora* and *C. eugenioides* were analyzed by genome-wide RNA-seq technology. Parental species are distinguished by an important proportion of *trans*-regulatory divergences. In hybrids, among divergently expressed genes between parental species and hybrids, 77% are expressed like one parent (expression level dominance). Gene expression was shown to result from the expression of both alleles, with occasional bias toward one genome. The gene expression patterns appear determined by complex combinations of *cis*- and *trans*-regulatory divergences of parental species and by intertwined parental *trans*-regulatory factors. The gene expression level depends on the simultaneous up and down-regulation of both alleles and the observed biased expression level dominance seems to be derived from the asymmetric effects of *trans*-regulatory parental factors on regulation of alleles. In the allopolyploid, at the genomic scale, both homeologous genes are also inter-regulated and contribute to the transcriptome.

The transcriptome of leaves from *C. arabica* cultivated at different growing temperatures suitable for one or the other parental species was examined to analyze the variation of homeologous gene expression in variable conditions. The relative subgenome contributions to the transcriptome appear to be only marginally altered by the growing conditions. *C. arabica*'s ability to tolerate a broader range of growing temperatures than its diploid parents does not result from differential use of homeologs.

The transcriptional response after a recent or old hybridization event was characterized by these studies. The genetic bases of the variations in allelic expression after the merger of parental gene expression regulatory networks, were elucidated and a model of regulation of homeologous gene expression in *C. arabica* is proposed.

Keywords : allopolyploidy, hybridization, homeolog, Allelic Specific Expression (ASE), *cis- trans*-regulation, transcriptome, RNA-seq, adaptation, coffee tree.