Regulatory Divergence between Parental Alleles Determines Gene Expression Patterns in Hybrids

Marie-Christine Combes[1],*, Yann Hueber[1], Alexis Dereeper[2], Stéphanie Rialle[3], Juan-Carlos Herrera[4], and Philippe Lashermes[1]

[1]IRD, UMR DIADE, Montpellier Cédex 5, France

[2]IRD, UMR IPME, Montpellier Cédex 5, France

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

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

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

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

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 ($C_{hyb} + E_{hyb}$) was determined and expressed as the percentage of the C_{hyb} allele (%C_{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: $C/E = C_{hyb}/E_{hyb}$, 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: $C_{hyb}/E_{hyb} = 1$, 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 ($C_{hyb}/E_{hyb} > 1$ and $C/E < C_{hyb}/E_{hyb}$ or $C_{hyb}/E_{hyb} < 1$ and $C/E > C_{hyb}/E_{hyb}$). If the *cis* and *trans* effects of one gene acted in the same direction ($C_{hyb}/E_{hyb} > 1$ and $C/E > C_{hyb}/E_{hyb}$ or $C_{hyb}/E_{hyb} < 1$ and $C/E < C_{hyb}/E_{hyb}$), 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
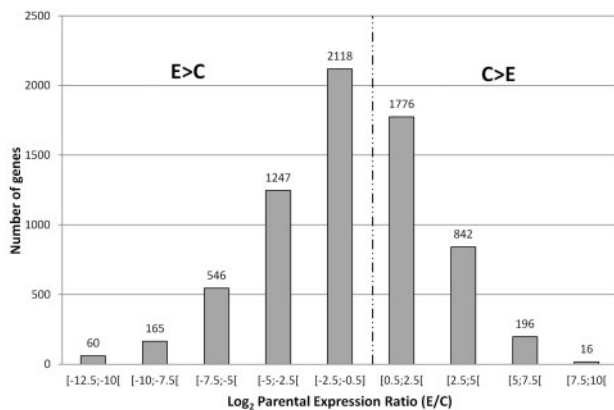
FIG. 1.—Differences in gene expression between *Coffea canephora* and *Coffea eugenioides*. The histogram shows the direction and magnitude of changes in expression in genes exhibiting divergent parental expression. Negative values indicate higher expression in *C. eugenioides* (E > C), and positive 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 × E and E × 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
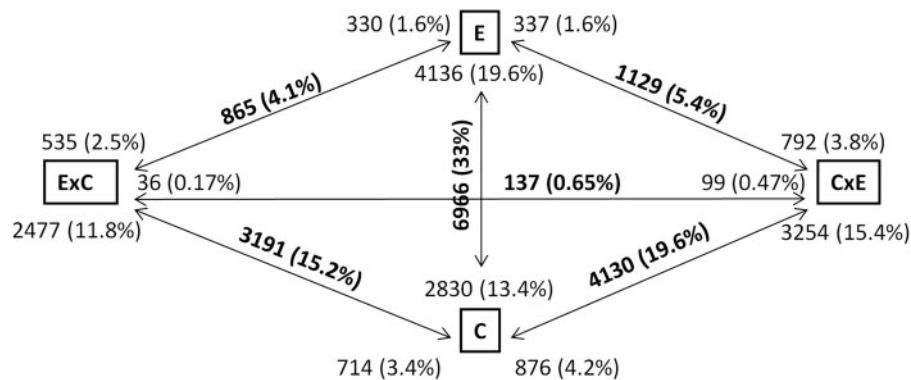
FIG. 2.—Differentially expressed genes in each possible comparison between parents and hybrids (DESeq analysis, see supplementary Methods, Supplementary Material online). Numbers in bold indicate the total number and fraction of genes differentially expressed in each comparison among the 21,052 genes 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, 2.5% and 11.8% of upregulated genes in E × C compared with 1.6% and 3.4% 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
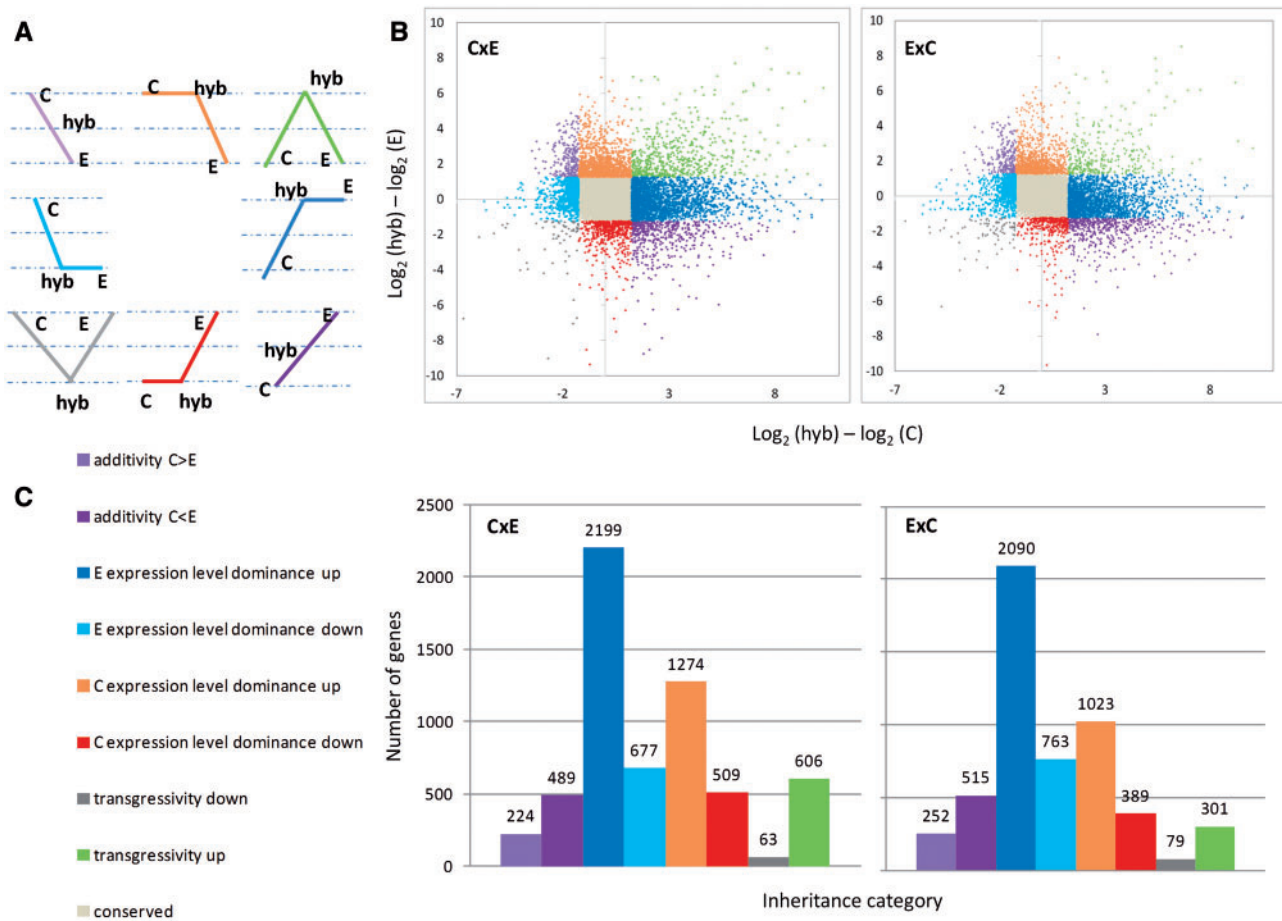
Fig. 3.—Inheritance of gene expression level in hybrids between *Coffea canephora* and *Coffea eugenioides*. (A) Hypothetical patterns of gene expression in *C. canephora* (C), *C. eugenioides* (E), and F₁ hybrids; C × E or E × C (hyb) illustrating additivity, transgressivity, and dominance are shown. (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,
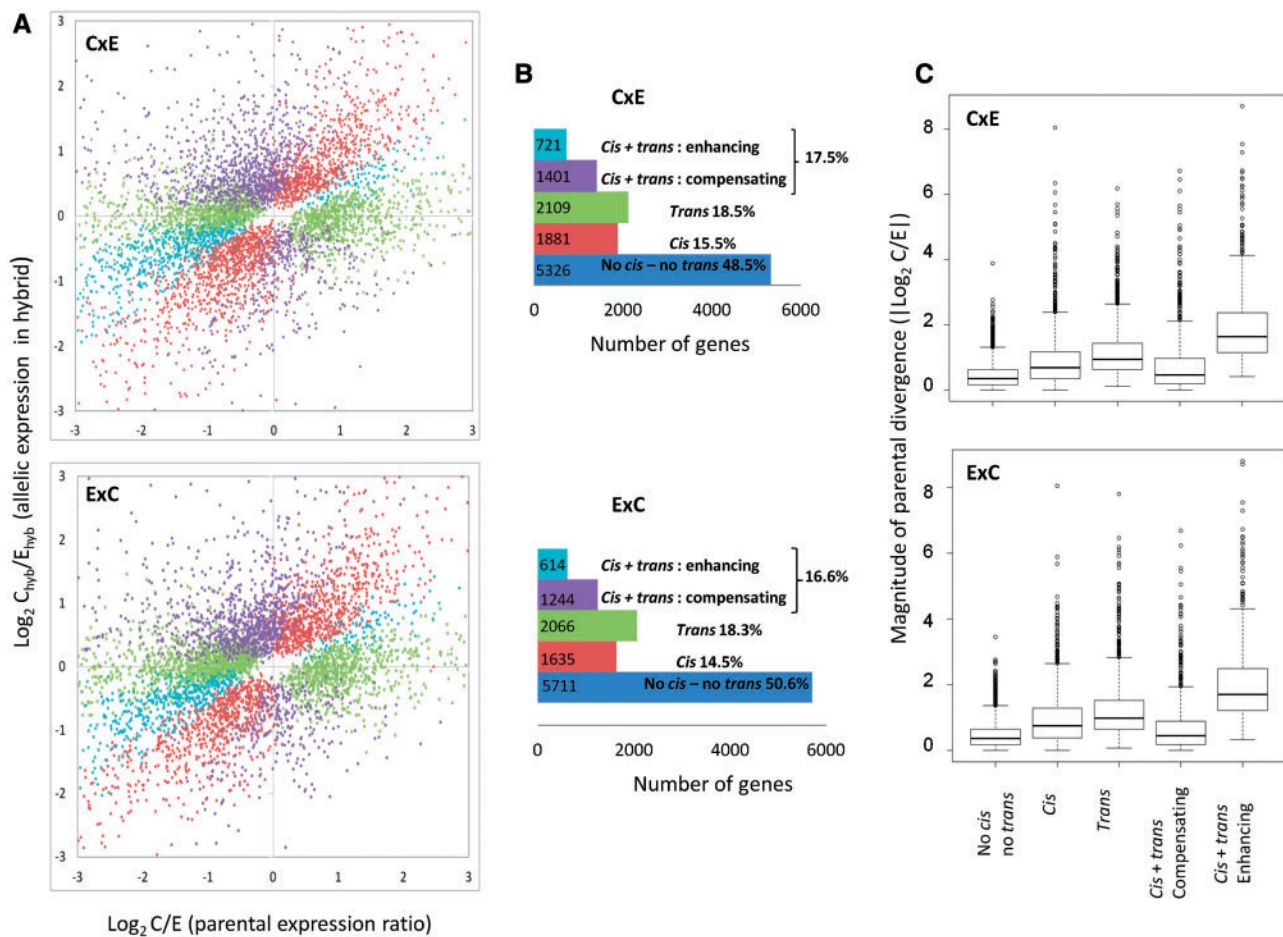
FIG. 4.—cis- and trans-regulatory divergence between parental species for 11,438 genes in C × E hybrids and 11,270 genes in E × C hybrids. (A) The plots summarize the relative allele-specific expression levels in parents and hybrids (C/E on the x axis; C_{hyb}/E_{hyb} on the y axis; C, E and C_{hyb}, 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 × E and E × 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 × E and E × 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 × E and 1.22-fold for E × C hybrids) was larger than the median cis-regulatory difference between species (0.88-fold for C × E and 0.95-fold for E × 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 × E and E × 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 × E and E × 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 × E and supplementary fig. S2, Supplementary Material online, for E × 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
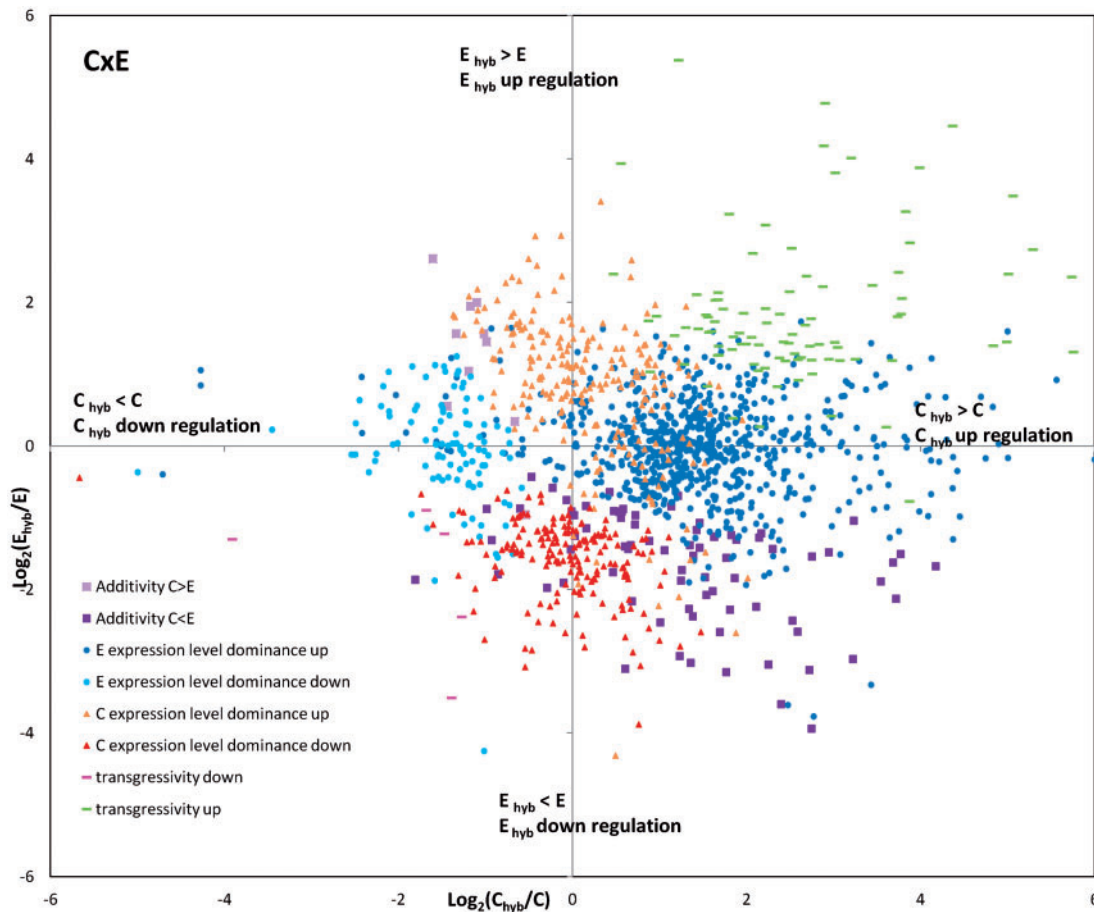
Fig. 5.—Variation in specific allelic expression (ASE) for genes differentially expressed between parents and the hybrid in the C × E hybrids. Each point represents a single gene, its position displays the combined up- and downregulation of both alleles (C_{hyb}/C on the x axis; E_{hyb}/E on the y axis, 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, $P < 0.001$).
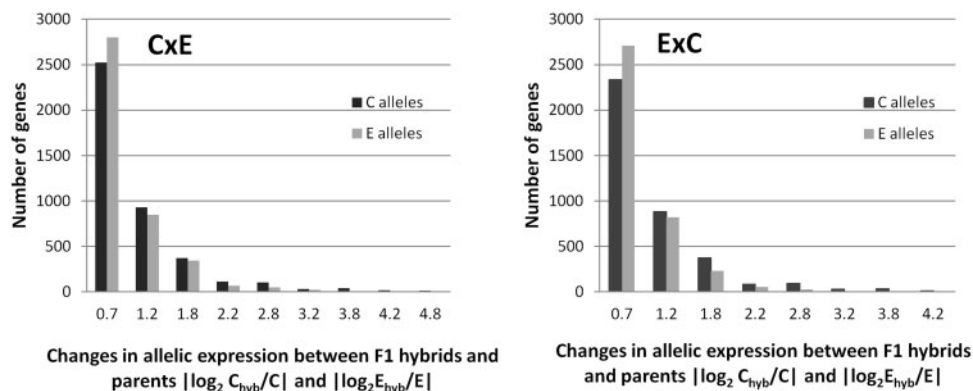

Fig. 6.—Effects of *trans*-regulatory factors of the E and C genome on up- and downregulation of C_{hyb} and E_{hyb} alleles in C × E and E × C hybrids. In line with the method of Shi et al. (2012), the genes subjected to either only *trans*- or only *cis* + *trans* effects were categorized according to differences in the absolute values of allelic expression between C × E and E × 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, $P < 0.001$).
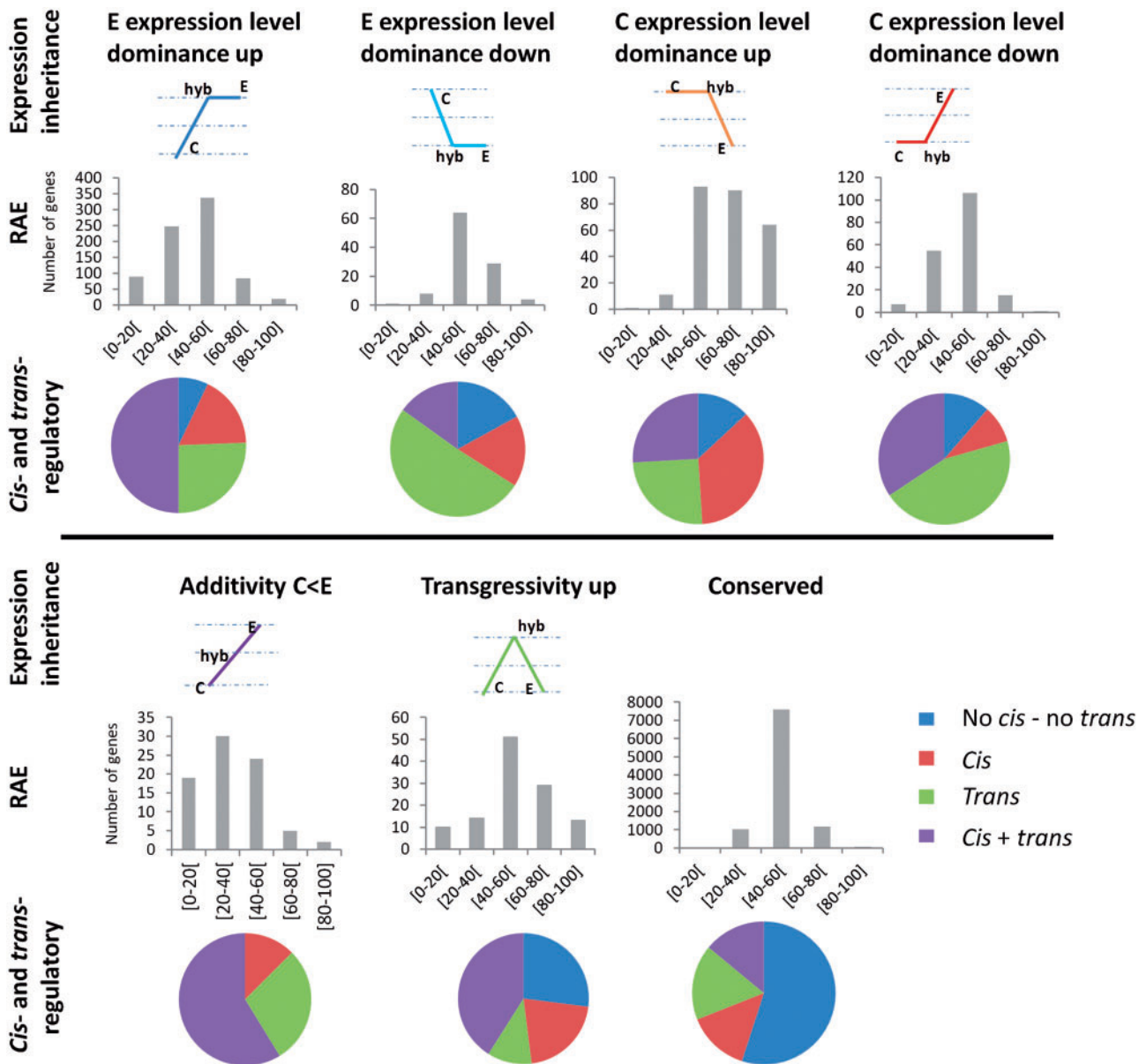
Fig. 7.—For the C × E hybrids, all the data collected on genes belonging to each category of expression inheritance are combined. The distributions of genes according to RAE and to cis- and trans-regulatory divergences are shown in the form of histograms and pie charts.

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-regulatoty 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 × E and supplementary fig. S3, Supplementary Material online, for E × 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

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, Hiscock 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, Hiscock 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, Hiscock 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