

Effects of parental genetic divergence on gene expression patterns in interspecific hybrids of *Camellia*

Min Zhang

Nanjing Forestry University

Yi-Wei Tang

Fudan University

Ji Qi

Fudan University

Xin-Kai Liu

Palm Eco-Town Development Co,Ltd

Dan-Feng Yan

Palm Eco-Town Development Co,Ltd

Nai-Sheng Zhong

Palm Eco-Town Development Co,Ltd

NAi-Qi Tao

Fudan University

Ji-Yin Gao

Palm Eco-Town Development Co,Ltd

Yu-Guo Wang

Fudan University

Zhi-Ping Song

Fudan University

Ji Yang

Fudan University

Wen-Ju Zhang (✉ wjzhang@fudan.edu.cn)

Fudan University <https://orcid.org/0000-0003-1105-9465>

Research article

Keywords: *Camellia*, allele-specific gene expression, hybridization, transcriptome shock, cis and trans regulation

Posted Date: September 18th, 2019

DOI: <https://doi.org/10.21203/rs.2.14500/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on November 8th, 2019. See the published version at <https://doi.org/10.1186/s12864-019-6222-z>.

Abstract

Background: The merging of two divergent genomes during hybridization can result in the remodeling of parental gene expression in hybrids. A molecular basis underlying expression change in hybrid is regulatory divergence, which may change with the parental genetic divergence. However, there still no unanimous conclusion for this hypothesis.

Results : Three species of *Camellia* with a range of genetic divergence and their F₁ hybrids were used to study the effect of parental genetic divergence on gene expression and regulatory patterns in hybrids by RNA-sequencing and allele-specific gene expression analysis. We found that though the proportion of differentially expressed genes (DEGs) between the hybrids and their parents did not increase, a greater proportion of DEGs would be non-additively (especially transgressively) expressed in the hybrids as genomes between the parents become more divergent . In addition, the proportion of genes with significant evidence of *cis*-regulatory divergence increased , whereas with *trans*-regulatory divergence decreased with parental genetic divergence.

Conclusions : The discordance within hybrid would intensify as the parents become more divergent, manifesting as more DEGs would be non-additively expressed. *Trans*-regulatory divergence contributed more to the additively inherited genes than *cis* , however, its contribution to expression difference would be weakened as *cis* mutations accumulated over time; and this might be an important reason for that the more divergent the parents are, the greater proportion of DEGs would be non-additively expressed in hybrid.

Introduction

Hybridization is an important power facilitating adaptive evolution [1]. In nature, hybridization is ubiquitous. It has been reported that over 25% of plant species and 10% of animal species are involved in hybridization or potential introgression with other species [2, 3]. Although most hybrids are infertile, some can possess novel phenotypic traits, like stronger stress tolerance and improved growth rate, which are better for their adaptation to hostile environments or expansion into new habitats; under natural selection, they also have the opportunity to evolve into new species [4–6].

Novel phenotypes can arise from changes of protein sequences. However, the variation of protein sequence is insufficient to explain so abundant morphological types present in nature [7]. Alternatively, the change of gene expression provides another source of phenotypic novelty [8]. There is growing evidence that merging of two divergent genomes during hybridization can result in the remodeling of parental gene expression patterns in hybrids, a phenomenon called “transcriptome shock” [9–12]. As manifestations, many genes would be non-additively expressed in hybrids (diverge from the mid-parental value), which contribute to their transgressive phenotypes at some extent [13, 14].

Broadly speaking, gene expression is controlled by the interactions between *cis*- and *trans*-acting elements, so transcriptome shock is likely in large part due to the variation of *cis*- and *trans*-regulation [15, 16]. *Cis*- and *trans*-regulatory divergence can be distinguished by measuring the allele-specific gene expression (ASE) between two genotypes and their F₁ hybrid. In F₁ hybrid, two parental alleles are exposed to a common cellular environment, so *trans*-regulatory change has same effect on the two alleles, and their imbalanced expression is a readout of the relative *cis*-regulatory divergence [17]. Based on this strategy, a substantial effort has been made and revealed variable roles that *cis*- and *trans*-regulatory changes would play in reshaping gene expression. Previous studies on *Drosophila* showed that *cis*-regulatory change tended to result in the additive inheritance of gene expression [18, 19], but opposite result appeared in plant for that *trans*-regulatory change contributed more to the additive expression of the *Cirsium* hybrids [20]. In addition, the relative frequency of *cis*- and *trans*-regulatory divergence among studies was always inconsistent. Shi et al.'s study on *Arabidopsis* found that a greater proportion of genes showed significant evidence of *cis*- than *trans*-regulatory divergence [21], whereas Combes et al.'s study on *Coffea* got the opposite result [22]. For better understanding these processes, studies based on more species are still needed.

Moreover, Tirosh et al. found that *cis*-regulatory divergence seemed to be more common between than within species [16]. That means regulatory patterns may change with the genetic divergence between two parents and may be even related to the degree of “transcriptome shock” in hybrids [18, 23, 24]. To validate these hypotheses, three species of *Camellia* L., including *C. azalea* Z. F. Wei, *C. chekiangoleosa* Hu and *C. amplexicaulis* (Pit.) Cohen-Stuart as well as their F₁ hybrids [*C. azalea* (♀) × *C. chekiangoleosa* (♂) and *C. azalea* (♀) × *C. amplexicaulis* (♂)] were used in this study to detect the influence of parental genetic divergence on gene expression and regulatory patterns in hybrids. Two crosses represent the intra- and inter-sectional hybridization of *Camellia*, respectively. Through RNA sequencing and ASE analysis, we are aiming to investigate how *cis*- and *trans*-regulations change with parental genetic divergence as well as their effect on gene expression in hybrid.

Results

Sequencing and mapping

As described above, two crosses representing intra- and inter-sectional hybridization of *Camellia* were used in this study (Figure 1). cDNA libraries were constructed using RNA extracted from flower buds of the F₁ hybrids and their parental species, and then sequenced using the Illumina HiSeq X-ten platform. For each species and hybrid, three biologic replicates were set up. Finally, 664.6 million clean reads were obtained from 15 libraries with a mean of 44.3 million for each library. The proportion of clean reads with quality better than Q20 was over 97%, and better than Q30 was over 92% for each library (Additional file 1: Table S1). Three pseudo-genomes, representing the female and the two male parents, were constructed. Clean reads from the parental species were then realigned to their pseudo-genomes. The mean mapping rates for *C. azalea*, *C. chekiangoleosa* and *C. amplexicaulis* were ~70%. Clean reads from the hybrids were mapped to the pseudo-genomes of their parents, respectively. Although the mapping rates for the hybrids were relatively lower (~60%), we chose the maximum value of the two mapping results for each allele and their sum as the total reads count, which could counteract the influence of low mapping rates on the subsequent analysis.

Changes of parental gene expression patterns in different F₁ hybrids

Over half of the analyzed genes (57.8% for *C. azalea* × *C. chekiangoleosa* and 51.7% for *C. azalea* × *C. amplexicaulis*) were significantly differentially expressed between the F₁ hybrids and at least one of their parents. Regardless of parental divergence, DEGs between the hybrids and their parents for each cross were further classified into eight clusters (Figure 2). For the cross of *C. azalea* × *C. chekiangoleosa*, the relative proportion of genes belonging to additivity (including additivity female > male and female < male), female dominance (including dominance up and down), male dominance (including dominance up and down) and transgressivity (over-dominance and under-dominance) was 4.56%, 37.09%, 27.38% and 30.97%, respectively; while for the cross of *C. azalea* × *C. amplexicaulis*, the proportions were 1.48%, 25.76%, 35.51% and 37.25%, respectively. Compared with the intra-sectional cross (95.44%), a greater proportion of DEGs between the hybrids and their parents exhibited a non-additively expressed pattern in the inter-sectional cross (98.52%) (Fisher's exact test, P -value < $2.2e^{-16}$). The relative proportion of DEGs with transgressive expression pattern was significantly higher in the inter-sectional hybrid (37.25%) than that in the intra-sectional hybrid (30.97%) (Fisher's exact test, P -value = $9.0e^{-11}$). Pearson correlation analysis showed that the total expression level of the F₁ hybrid of *C. azalea* × *C. chekiangoleosa* was more similar to its parents ($cor > 0.81$, P -value < $2.2e^{-16}$) than the hybrid of *C. azalea* × *C. amplexicaulis* ($cor < 0.79$, P -value < $2.2e^{-16}$) (Additional file 1: Figure S1).

Allele-specific gene expression tests reveal *cis*- and *trans*-regulatory divergence in different crosses

Of the 7,629 genes detected in the cross of *C. azalea* × *C. chekiangoleosa*, 8.09% (617) showed significant evidence of *cis*-regulatory divergence (Figure 3). When it came to the cross of *C. azalea* × *C. amplexicaulis*, the proportion of genes with significant evidence of *cis*-regulatory divergence was 10.31% (986 of 9,566). Expression differences between species not attributable to *cis*-regulatory divergence could be caused by *trans*-regulatory divergence. In *C. azalea* × *C. chekiangoleosa*, 13.34% (1,018 of 7,629) of the genes showed significant evidence of *trans*-regulatory divergence, compared with 8.24% (629 of 9,566) in *C. azalea* × *C. amplexicaulis*. In addition, there were also 276 (3.62% of 7,629) genes in *C. azalea* × *C. chekiangoleosa* and 294 (3.07% of 9,566) genes in *C. azalea* × *C. amplexicaulis* showed significant evidence of both *cis*- and *trans*-regulatory divergence. Genes with significant evidence of both *cis*- and *trans*-regulatory divergence were further divided into three clusters, i.e. "*cis* + *trans*", "*cis* × *trans*" and "compensatory" (Additional file 1: Table S2). The proportion of genes belong to the above three clusters in the cross of *C. azalea* × *C. chekiangoleosa* was 1.15% (88), 1.19% (91) and 1.27% (97), respectively; while in *C. azalea* × *C. amplexicaulis* was 1.08% (103), 0.76% (73) and 1.23% (118), respectively.

Regulatory difference underling expression divergence between species

The median significant *trans*-regulatory difference between *C. azalea* and *C. chekiangoleosa* was 1.26 folds, which was significantly larger than the median *cis*-regulatory difference (0.94-fold, Wilcoxon's rank-sum test, P -value < $2.2e^{-16}$). Same pattern was also detected between *C. azalea* and *C. amplexicaulis* (Wilcoxon's rank-sum test, P -value = $1.0e^{-15}$), of which the median significant *trans*-regulatory difference was 1.30-fold, and the median significant *cis*-regulatory difference was 1.06-fold, respectively (Figure 4A). Kendall's test showed that, the expression differences between *C. azalea* and *C. chekiangoleosa* correlated more strongly with *trans*-regulatory divergence ($\tau = 0.34$, P -

value $< 2.2e^{-16}$) than with *cis*-regulatory divergence ($\tau = 0.12$, P -value $< 2.2e^{-16}$). Same pattern was also detected between *C. azalea* and *C. amplexicaulis*, of which *trans*-regulatory divergence contributed more to the expression divergence ($\tau = 0.21$, P -value $< 2.2e^{-16}$) than *cis*-regulatory divergence ($\tau = 0.18$, P -value $< 2.2e^{-16}$). The amount of total regulatory divergence explained by *cis*-regulatory difference (% *cis*) decreased with the absolute magnitude of expression divergence between *C. azalea* and the other two species (Figure 4B). However, the contribution of *cis*-regulatory difference to the expression divergence between *C. azalea* and *C. amplexicaulis* increased significantly compared with that between *C. azalea* and *C. chekiangoleosa* (Wilcoxon's rank-sum test, P -value $< 2.2e^{-16}$). We also compared the absolute magnitude changes of parental expression divergence with different regulatory categories. As shown in Figure 4C and 4D, "*trans* only" play a larger role than "*cis* only" in aggravating expression divergence between different species (Wilcoxon's rank-sum test, P -value < 0.001). Furthermore, the interaction effect of *cis*- and *trans*-regulations functioning in the same direction (*cis* + *trans*) could tremendously change the gene expression patterns between two species. However, when the two regulations worked in the opposite direction ("*cis* × *trans*" and "compensatory"), the divergence of gene expression would be relieved to a large extent.

Regulatory divergence underling gene expression patterns in different F₁ hybrids

To examine the potential relationship between regulatory divergence and gene expression patterns in hybrid, we compared the % *cis* between sets of genes with additive and non-additive expression patterns in different hybrids. As shown in Figure 5, in the F₁ hybrid of *C. azalea* × *C. chekiangoleosa*, the median % *cis* for genes with non-additive expression patterns was significantly higher than that with additive expression patterns (Wilcoxon's rank-sum test, P -value = $3.2e^{-7}$). However, different result was detected in the hybrid of *C. azalea* × *C. amplexicaulis* for that there was no significant difference in the median % *cis* for additively and non-additively expressed genes (Wilcoxon's rank-sum test, P -value = 0.1). In addition, % *cis* in the hybrid of *C. azalea* × *C. amplexicaulis* was significant higher than that in the hybrid of *C. azalea* × *C. chekiangoleosa* for either additively (Wilcoxon's rank-sum test, P -value = $2.8e^{-8}$) or non-additively inherited genes (Wilcoxon's rank-sum test, P -value $< 2.2e^{-16}$). Most DEGs between the hybrids and their parents were subjected to the effects of "conserved" and "ambiguous". Of the remaining DEGs with any expression patterns, a greater proportion were subjected to "*trans* only" than any other effects in the F₁ hybrid of *C. azalea* × *C. chekiangoleosa*, while in the hybrid of *C. azalea* × *C. amplexicaulis*, a greater proportion were regulated by "*cis* only" (Table 1).

Discussion

Transcriptome shock in hybrid intensifies with parental genetic divergence

As described above, the merging of two divergent genomes during hybridization can result in "transcriptome shock". Many studies reported the altered expression patterns in hybrids. The study on yeast found that 56.3% of the studied genes were differentially expressed between the F₁ hybrid and its parents, of which 81.8% were non-additively expressed [11, 25]. Bell et al.'s study on *Cirsium* found that 70.0% of the studied genes were differentially expressed between the F₁ hybrid and at least one of its parents, with 92.5% showed non-additive expression patterns [20]. In our study, over half of the genes were differentially expressed between the hybrids and at least one of their parents in either the intra-sectional or the inter-sectional hybridization, and most of them were non-additively expressed in the hybrids. Although the proportion of DEGs decreased to some extent in the inter-sectional hybrid, a greater proportion of DEGs would be non-additively expressed in the inter-sectional hybrid than that in the intra-sectional hybrid. Specifically, more DEGs were transgressively expressed in the inter-sectional hybrid than that in the intra-sectional hybrid. That means the relative proportion of non-additively (especially transgressively) expressed gene within DEGs in hybrids would increase with parental genetic divergence. Correspondingly, the total expression level of genes in the inter-sectional hybrid was more diverge from its parents than that in the intra-sectional hybrid as shown in Additional file 1: Figure S1. These results could serve as important evidence that transcriptome shock in hybrid would intensify with parental genetic divergence.

Relative frequency of *cis*- and *trans*-regulatory divergence in different hybrids

According to previous studies, *cis*- and *trans*-regulatory divergence have their own ways in affecting gene expression [19]. So, the relative frequency of *cis*- and *trans*-regulatory divergence has great influence on the inheritance of gene expression patterns in hybrid [18]. The relative frequency of *cis*- and *trans*-regulatory divergence revealed by different studies is always variable. Taking *Drosophila* for example, Wittkopp et al. [15, 26] and Landry et al.'s [27] studies on the hybrids of *D. melanogaster* × *D. simulans* found that more genes showed

significant evidence of *cis*-than *trans*-regulatory divergence. However, studies on the hybrids of *D. melanogaster* × *D. simulans* got the opposite result [19, 24]. In plants, Zhuang and Adams researched the gene regulatory patterns in the F₁ hybrid of *Populus trichocarpa* × *P. deltoids* [28]. Their result showed that a greater proportion of genes exhibited significant evidence of *cis*-than *trans*-regulatory change. Same result was also detected in the study of maize [29]. However, when it came to *Cirsium*, more genes were significantly influenced by *trans*-rather than *cis*-regulatory divergence [20].

Denver et al. speculated that natural selection would eliminate most *trans*-acting mutations and accumulate *cis*-regulatory mutations over time [30]. That means the relative frequency of *cis*- and *trans*-regulatory changes in hybrids may be related to the divergence time between the parental species. In our study, the *cis*- and *trans*-regulatory divergence in different crosses were distinguished using the same method with unified criterions. However, the results were completely different for that the proportions of genes with significant evidence of *cis*- and *trans*-regulatory divergence in the intra-sectional cross (*C. azalea* × *C. chekiangoleosa*) were 8.09% and 13.34%, respectively, whereas in the inter-sectional cross of *C. azalea* × *C. amplexicaulis* were 10.31% and 8.24%, respectively. In other words, *trans*-regulatory divergence was more prevailing than *cis*- in the intra-sectional cross, while in the inter-sectional cross was just the opposite. These results indicate that the proportion of genes with significant evidence of *cis*-regulatory divergence would increase, while with significant evidence of *trans*-regulatory divergence would decrease with genetic divergence between species. A potential reason for this phenomenon may be that *cis*-regulatory mutations are more likely to be fixed than *trans*- under natural selection. This seems to be inconsistent with a neutral model assuming equal probabilities of fixation for *cis*- and *trans*-regulatory polymorphisms. In fact, *cis*-acting mutations in the promoter region may simply alter the transcript levels of gene(s) downstream, whereas a *trans*-acting mutation in a transcription factor may result in multiple downstream regulatory changes [31]. For selection, it must act more strongly against mutations with pleiotropic effects to operate more efficiently [32]. So *trans*-regulatory mutations with multiple effects are more likely to be eliminated. Specifically, as Wittkopp et al. [26] speculated, *trans*-acting mutations may include both highly pleiotropic changes as well as some with limited effects, the former ones are more likely to be eliminated, while the later ones could be accumulated by mutation-selection balance. This may be an important pattern for the evolution of *cis*- and *trans*-regulation. We speculate the variation of parental genetic divergence may be also an important reason for that the relative frequency of *cis*- and *trans*-regulatory divergence revealed by different studies was divergent.

***cis*- and *trans*-regulatory differences underlying expression divergence between species**

McManus et al.'s study on the hybrid of *Drosophila* showed that the median significant of *trans*-regulatory divergence was larger than that of *cis*-regulatory divergence between species, and *trans*-regulatory divergence correlated more highly with the expression difference between species [19]. Same profile also appeared in the study of *Cirsium* [20]. Similarly, our results showed that *trans*-regulatory change contributed more to the expression divergence between *C. azalea* and the other two species than *cis*-regulatory change. Correspondingly, the expression differences between *C. azalea* and the other two species correlated more highly with *trans*-regulatory changes, too. That means *trans*-regulatory change plays a larger role than *cis*-regulatory change in promoting the differentiation of gene expression between species. We also detected the changes of the % *cis* with the absolute magnitude of total expression divergence between species. As shown in Figure 4B, the relative percent of *cis*-regulatory divergence decreased with the absolute magnitude of total expression divergence between *C. azalea* and the other two species. In other words, genes which were more deeply affected by *trans*-regulatory change would be more divergently expressed between species. This was consistent with the result generated from previous studies [20, 24]. However, as *cis*-regulatory mutation accumulates over time, its influence on expression divergence increases, too. This could be deduced from the fact that the contribution of *cis*-regulatory change (% *cis*) to the expression divergence between *C. azalea* and *C. amplexicaulis* at any level was obviously higher than that between *C. azalea* and *C. chekiangoleosa* (Figure 4B).

Cis- and *trans*-regulatory divergence are not mutually exclusive, many genes would be significantly influenced by both *cis*- and *trans*-regulatory changes [19, 20, 22, 24, 27]. In our study, 3.62% of the studied genes between *C. azalea* and *C. chekiangoleosa* and 3.07% between *C. azalea* and *C. amplexicaulis* showed significant evidence of both *cis*- and *trans*-regulatory divergence. Interactions between *cis*- and *trans*-regulatory divergences can result in quite divergent expression patterns between species. As shown in Figure 4C and 4D, *cis*- and *trans*-regulatory changes promoting expression of the same allele (*cis* + *trans*) could greatly stimulate the expression divergence between species. Conversely, if two regulatory categories act on the alternate alleles (*cis* × *trans*), the divergence of gene expression between species would be relieved to some extent. Specifically, the compensatory effect of *cis*- and *trans*-regulatory changes tended to eliminate expression divergence between species. These findings are consistent with the results generated from *Coffea* [22] and *Arabidopsis* [21]. According to previous studies, genes significantly influenced by "*cis* + *trans*" might be driven by directional selection [16], whereas regulated by "*cis* × *trans*" as well as "compensatory" are likely to be driven by stabilizing selection [33]. In our study, the proportion of genes with significant evidence of both *cis*- and *trans*-regulatory divergence was lower in the cross of *C. azalea* × *C. amplexicaulis* than that in *C.*

azalea × *C. chekiangoleosa*, mainly because of that fewer genes were affected by "cis × trans" (Figure 3). This may just reflect the evolutionary history of the three *Camellia* species: *C. azalea* and *C. amplexicaulis*, as two closely related species, have more genes experienced stabilizing selection; while for species from two divergent sections (*C. azalea* and *C. amplexicaulis*), fewer genes between them are driven by stabilizing selection.

Contribution of regulatory divergence to gene expression patterns in hybrid

As described above, molecular basis underlying gene-expression difference is the variation of *cis*- and *trans*-regulations. Previous studies on *Drosophila* [19] and yeast [34] showed that *cis*-regulatory divergence appeared to result in additive inheritance of gene expression more often than *trans*-regulatory divergence. However, latest studies based on transcriptome analysis reported the opposite result, for that *trans*-regulatory divergence in these studies accounted for a greater proportion of the regulatory divergence at sites with additive than that with non-additive inheritance patterns [20, 24]. In our study, in the F₁ hybrid of *C. azalea* × *C. chekiangoleosa*, the median of % *trans* was significant higher for genes showing additive expression pattern than that showing non-additive expression pattern (Figure 5). However, in the hybrid of *C. azalea* × *C. amplexicaulis*, there was no significant difference in the relative percent of *cis*- and *trans*-regulatory divergence for neither additively nor non-additively expressed genes. We speculate that the relative contribution of *cis*- and *trans*-regulatory divergence (% *cis* and % *trans*) to the inheritance of gene expression may depend on the parental genetic divergence. A potential mode is that *trans*-regulatory divergence is more likely to lead to additive inheritance than *cis*-regulatory divergence. For hybrids whose parents are closely related species, the relative frequency of *trans*-regulatory divergence is higher than that of *cis*-regulatory divergence; however, as genomes between the two parents become more divergent, *trans*-regulatory mutations are eliminated to some extent and *cis*-regulatory divergence becomes dominant. This could be used to explain why a higher proportion of genes would be non-additively (especially transgressively) expressed in the F₁ hybrid of inter-sectional than that of intra-sectional hybridization.

The interactions between *cis*- and *trans*-regulatory divergences can greatly affect gene expression patterns between species. There were studies showed that "cis × trans" regulatory divergence was more common in transgressively expressed genes [19, 27]. However, study on the hybrids of *Cirsium* found that genes with transgressive expression pattern were mainly regulated by "cis + trans" [20]. In our study, neither "cis × trans" nor "cis + trans" regulation was the major reason leading to the transgressive expression patterns in hybrids (Table 1). Instead, most of the DEGs between the hybrids and their parents followed a "conserved" or "ambiguous" manner. In addition, compared with the intra-sectional hybridization, a greater proportion of DEGs were subjected to "cis only" effect in the inter-sectional hybridization for any expression patterns. So, inheritance of gene expression patterns is more likely to be the result of the comprehensive effects of different regulatory mechanisms, and the change of relative frequency of *cis*- and *trans*-regulatory divergence plays an important role in the formation of divergent expression patterns in hybrid.

Conclusion

In this study, by comparing the gene expression patterns between the *Camellia* hybrids and their parents, we found that the relative proportion of DEGs with non-additively expressed patterns in hybrid would increase with parental genetic divergence, which indicated that the discordance within hybrid would intensify as the parents become more divergent. Meanwhile, the proportion of genes with significant evidence of *cis*-regulatory divergence increased, while with *trans*-regulatory divergence decreased with parental genetic divergence. *Trans*-regulatory change contributes more to the additively inherited genes in hybrid than *cis*. So, the weakening of *trans*-regulatory effect and the strengthen of *cis*-regulatory effect provide a major reason for the phenomenon that the more divergent the parents are, the greater proportion of DEGs would be non-additively expressed in hybrid.

Materials And Methods

Plant materials and hybridization

Three species including *C. azalea* (2n = 30), *C. chekiangoleosa* (2n = 30) and *C. amplexicaulis* (2n = 30) were used in this study (Figure 1). According to morphological and molecular studies, both *C. azalea* and *C. chekiangoleosa* belong to the Sect. *Camellia* of *Camellia*, while *C. amplexicaulis* belongs to the Sect. *ArcheCamellia* of *Camellia* [35, 36]. Hybridizations were carried out by Palm Eco-Town Development Co. Ltd in 2007 following the technique described by Gao et al. [37]. For all the hybridization experiments, *C. azalea* was served as the female parent, and the other two species were served as the male parents. All the plants in this study are grown in a same greenhouse of Palm Eco-Town Development Co. Ltd at Guangzhou, China. To improve pollination efficiency, pollens from different individuals of the two paternal species were collected together, respectively. The mixed pollens were then used to pollinate the flowers of *C. azalea* plants. So, the

F₁ hybrids may be not from the identical parents, but their parents came from individuals of one wild population, respectively. Finally, two F₁ hybrid series, *C. azalea* (♀) × *C. chekiangoleosa* (♂) and *C. azalea* (♀) × *C. amplexicaulis* (♂), representing intra-sectional and inter-sectional hybrids, were successfully obtained. Flower buds of the F₁ hybrids and their parents at same stage were harvested and frozen in liquid nitrogen immediately, then transferred to −80°C refrigerator for storage.

RNA extraction and sequencing

Total RNA was extracted from the flower buds using the RNAprep Puree Plant Kit DP441 (TIANGEN, Beijing, China) according to the manufacturer's instructions. For each species and hybrid, three biologic replicates (from three individuals, respectively) were set up as parallel experiments. Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) was used to detect the quantity and quality of RNA. RNA-seq library was constructed for each sample. In total, 15 libraries were constructed, and then paired-end (2×150 bp) sequenced using Illumina HiSeq X-ten platform (Illumina Inc, San Diego, CA, USA) by Beijing Genomics Institute (BGI, Shenzhen, China) with the standard Illumina RNA-seq protocols.

Mapping transcriptome reads to the reference genome

Clean reads were obtained by removing reads with adapter contamination and ploy-Ns ($\geq 5\%$) as well as low quality reads with over 20% of low-quality bases (Phred < 15). Since the reference genomes of the species in our study were not available, filtered reads from the parental libraries were first mapped to the genome of *C. sinensis* var. *assamica* [38] using STAR software [39] with default parameters, and only uniquely mapped reads were retained. Then, SAMtools [40] and VarScan [41] software were orderly used for SNP calling. SNP sites at which three replicates were consistent were marked. For allele-specific gene expression research, a tough problem deserving consideration is mapping bias. To relieve mapping bias, three pseudo-genomes, representing the female and the two male parents, were constructed by replacing the reference alleles in the *C. sinensis* genome with the corresponding alternative alleles at the SNP sites, respectively. Then, transcriptome reads from the parental libraries were realigned to their pseudo-genomes using the same parameters to obtain the final read counts at the SNP sites. When it came to the F₁ hybrid series, reads from each library were mapped to the two pseudo-genomes of their parents, respectively. To relieve mapping bias, for each parental allele in the hybrids, we chose the maximum value of the two mapping results as the final reads count at each SNP site, and the sum of the two alleles as the total reads count at one site. To identify reliable SNP sites, quality controls were applied as follows: (i) the SNP sites in the two parents must be homozygous for difference; (ii) each SNP site in the F₁ hybrid must consist of only two alleles (one for the male parent, another for the female parent); (iii) the read count of the minor parental allele in the hybrid at each SNP site must be ≥ 2 ; and (iv) the total read count at each SNP site must be ≥ 20 .

Gene expression quantitation

We wrote a R script to identify species-specific SNP sites from the mapping results. Finally, 37,078 SNPs, representing 7,629 genes were identified from the cross of *C. azalea* × *C. chekiangoleosa*; and 81,477 SNPs, representing 9,566 genes, were identified from the cross of *C. azalea* × *C. amplexicaulis*. Transcript abundances of genes were evaluated as the normalized reads mapped per SNP site. Trimmed Mean of M-values (TMM) method [42] implemented in the edgeR package [43] was used for data normalization across libraries based on the assumption that most genes are not differentially expressed. Gene expression level was independently quantified for each cross, taking the biological replicates into consideration. The normalized gene expression for each cross is provided in the supporting information (Additional file 1: Table S3 and Additional file 1: Table S4). Cluster analysis was then carried out to examine the repeatability of the three biological replicates. As shown in Additional file 1: Figure S2, nearly all the biological replicates for each species and hybrid were clustered together (with $R^2 > 0.90$) except for amp1 and aza_che3, and these two samples were removed in the following analyses.

Classification of gene expression patterns

The edgeR package [43] was used for pairwise expression comparison, taking the three biological replicates into consideration. A fold change of 1.25 and the FDR < 0.05 were used as thresholds for differentially expressed gene (DEG) identification. DEGs between the hybrids and their parental species were further classed into eight clusters according to previous studies. Specifically, DEGs whose expression in the hybrid were higher than one of the parents but lower than another were classified as additivity (including additivity male > female and male < female); DEGs which were up/down-regulated in the hybrid compared with one of the parents but not differentially

expressed with another were classified as male/female expression level dominance-up/down; DEGs whose expression level in the hybrid were significantly higher/lower than both of the parents were classified as transgressivity (overdominance/underdominance).

Allele-specific gene expression (ASE) patterns and *cis*- and *trans*-regulatory divergence assignment

Based on the species-specific SNP information, relative expression of the parental alleles in hybrids was evaluated. For each allele, the mean value of the three biological replicates was used for ASE as well as the subsequent regulatory divergence analysis. Expression divergence between the parental species is mainly caused by the combination of *cis*- and *trans*-regulatory changes, which could be quantified as $\log_2(\text{male/female})$. In F_1 hybrid, two parental alleles are exposed to a common trans-regulatory environment, and are equally affected by the trans-regulatory change. So, the \log_2 -transformed ratio of allelic expression in hybrid was used to quantify the degree of *cis*-effect: $cis = \log_2(F_1A_{\text{male}}/F_1A_{\text{female}})$. Binomial exact test with FDR correction (FDR: 5%) was used to determine the significant *cis*-effect with a null hypothesis $F_1A_{\text{male}} = F_1A_{\text{female}}$. Then *trans*-regulatory divergence was calculated as the difference between \log_2 -transformed ratios of species-specific reads in the parents and the hybrids: $trans = \log_2(\text{male/female}) - \log_2(F_1A_{\text{male}}/F_1A_{\text{female}})$. Fisher's exact test with FDR corrections (FDR: 5%) was used to identify the statistically significant *trans*-effects with a null hypothesis $\text{Male/Female} = F_1A_{\text{male}}/F_1A_{\text{female}}$. The relative proportion of total regulatory divergence attributable to *cis*-regulatory divergence (% *cis*) was calculated as $(\% cis) = [|cis|/(|cis|+|trans|)] \times 100\%$, similarly, % *trans*. In addition, binomial exact test (FDR: 5%) was used to detect the significantly different expression between the two parental species. Regulatory divergence for different genes was then identified based on the results of binomial and Fisher's tests as well as the direction of changes. According to previous studies, seven regulatory types were further identified (Additional file 1: Table S2). (i) *cis* only: the parental alleles were unequally expressed in the same ratio in F_1 hybrid and between the two parents. (ii) *trans* only: the parental alleles were equally expressed in F_1 hybrid but unequally expressed between the two parents. (iii) *cis* + *trans*: the parental alleles were unequally expressed both in F_1 hybrid and between the two parents, but have the same direction (species with higher expression contributed the higher expressing allele in the F_1 hybrid). (iv) *cis* × *trans*: the parental alleles are unequally expressed both in F_1 hybrid and between the two parents, but have the opposite direction (species with higher expression contribute the lower expressing allele in the F_1 hybrid). (v) Compensatory: the two parental alleles are equally expressed between the two parents but unequally in the F_1 hybrid. (vi) conserved: the parental alleles are equally expressed both between the two parents and within the F_1 hybrid. (vii) ambiguous: other situations not included in the above six categories.

Statistical test

Pearson correlation analysis was carried out to detect the relationship of gene expression between the hybrids and their parents. Wilcoxon's rank-sum test was performed to compare the median parental expression divergence attributable to *cis* and *trans*-regulation. Kendall's test was used to detect the relative contribution of *cis*- and *trans*-regulation to the divergent gene expression between different species. All the test statistics were calculated in R programme (v 3.3.2, CRAN). The main scripts used in this study are available in the supporting information.

Abbreviations

ASE: allele-specific gene expression

DEGs: differentially expressed genes

Declarations

Acknowledgements

Not applicable

Funding

This work was supported by the National Natural Science Foundation of China (grant number 31670223 and 31270407) and the Key Research and Development Program of Guangdong Province (grant number: 2018B0202020002).

Availability of data and materials

All the sequencing data is available at the NCBI Sequence Read Archive (SRA) database with the accession number of SRP144621.

Author's contributions

MZ, WJZ, and JQ. conceived and designed the study. XKL, DFY, NSZ, and JYG. were responsible for the cultivation of all the materials. MZ and NQT performed the experiments. MZ and YWT analyzed and interpreted the data. WJZ supervised the research. MZ, WJZ, JQ, YGW, ZPS, and JY wrote, reviewed, and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, et al. Hybridization and speciation. *J Evol Biol.* 2013;26(2):229–246.
2. Mallet J. Hybridization as an invasion of the genome. *Trends Ecol Evol.* 2005; 20(5):229–237.
3. Whitney KD, Ahern JR, Campbell LG, Albert LP, King MS. Patterns of hybridization in plants. *Perspect Plant Ecol.* 2010;12(3):175–182.
4. Rieseberg LH. Homoploid reticulate evolution in *Helianthus* (asteraceae): Evidence from ribosomal genes. *Am J Bot.* 1991;78(9):1218–1237.
5. Rieseberg LH, Beckstrom-Sternberg SM, Liston A, Arias DM. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asteraceae). *Syst Bot.* 1991;16(1):50–76.
6. Tate JA, Symonds VV, Doust AN, Buggs RJA, Mavrodiev E, Majure LC, et al. Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (Asteraceae): 60 years after Ownbey's discovery. *Am J Bot.* 2009;96(5):979–988.
7. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science.* 1975;188(4184):107–116.
8. Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. Regulatory variation within and between species. *Annu Rev Genomics Hum Genet.* 2011;12(1):327–346.
9. Han H, Sun X, Xie Y, Feng J, Zhang S. Transcriptome and proteome profiling of adventitious root development in hybrid larch (*Larix kaempferi* × *Larix olgensis*). *BMC Plant Biol.* 2014;14(1):305.
10. Zhang J, Li G, Li H, Pu X, Jiang J, Chai L, et al. Transcriptome analysis of interspecific hybrid between *Brassica napus* and *B. rapa* reveals heterosis for oil rape improvement. *Int J Genomics.* 2015;2015.
11. Wu Y, Sun Y, Wang X, Lin X, Sun S, Shen K, et al. Transcriptome shock in an interspecific F₁ triploid hybrid of *Oryza* revealed by RNA sequencing. *J Integr Plant Biol.* 2016;58(2):150–164.

12. Taliercio E, Eickholt D, Rouf R, Carter T. Changes in gene expression between a soybean F₁ hybrid and its parents are associated with agronomically valuable traits. *PLoS One*. 2017;12(5):e0177225.
13. Combina M, Pérez-Torrado R, Tronchoni J, Belloch C, Querol A. Genome-wide gene expression of a natural hybrid between *Saccharomyces cerevisiae* and *S. kudriavzevii* under enological conditions. *Int J Food Microbiol*. 2012;157(3):340–345.
14. Zhuang Y, Tripp EA. Genome-scale transcriptional study of hybrid effects and regulatory divergence in an F₁ hybrid *Ruellia* (Wild Petunias: Acanthaceae) and its parents. *BMC Plant Biol*. 2017;17(1):15.
15. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in *cis* and *trans* gene regulation. *Nature*. 2004;430(6995):85–88.
16. Tirosh I, Reikhav S, Levy AA, Barkai N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*. 2009;324(5927):659–662.
17. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. Detection of regulatory variation in mouse genes. *Nat Genet*. 2002;32(3):432–437.
18. Lemos B, Araripe LO, Fontanillas P, Hartl DL. Dominance and the evolutionary accumulation of *cis*- and *trans*-effects on gene expression. *Proc Natl Acad Sci USA*. 2008;105(38):14471–14476.
19. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res*. 2010;20(6):816–825.
20. Bell GDM, Kane NC, Rieseberg LH, Adams KL. RNA-Seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol Evol*. 2013;5(7):1309–1323.
21. Shi X, Ng DWK, Zhang C, Comai L, Ye W, Jeffrey CZ. *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat Commun*. 2012;3:950.
22. Combes MC, Hueber Y, Dereeper A, Rialle S, Herrera JC, Lashermes P. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biol Evol*. 2015;7(4):1110–1121.
23. Gordon KL, Ruvinsky I. Tempo and mode in evolution of transcriptional regulation. *PLoS Genet*. 2012;8(1):e1002432.
24. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res*. 2014;24(5):797–808.
25. Schaefer B, Emerson JJ, Wang TY, Lu MYJ, Hsieh LC, Li WH. Inheritance of gene expression level and selective constraints on *trans*- and *cis*-regulatory changes in yeast. *Mol Biol Evol*. 2013;30(9):2121–2133.
26. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet*. 2008;40:346.
27. Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics*. 2005;171(4):1813–1822.
28. Zhuang Y, Adams KL. Extensive allelic variation in gene expression in *Populus* F₁ hybrids. *Genetics*. 2007;177(4):1987–1996.
29. Springer NM, Stupar RM. Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell*. 2007;19(8):2391–2402.
30. Denver DR, Morris K, Strelman JT, Kim SK, Lynch M, Thomas WK. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat Genet*. 2005;37(5):544–548.
31. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, et al. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*. 2003;20(9):1377–1419.
32. Haygood R. Mutation rate and the cost of complexity. *Mol Biol Evol*. 2006;23(5):957–963.

33. Birchler JA, Veitia RA. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 2010;186(1):54–62.
34. Gruber JD, Vogel K, Kalay G, Wittkopp PJ. Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. *PLoS Genet.* 2012;8(2):e1002497.
35. Ming TL. A systematic synopsis of the genus *Camellia*. *Acta Botanica Yunnanica.* 1998;21:149–159.
36. Vijayan K, Zhang WJ, Tsou CH. Molecular taxonomy of *Camellia* (Theaceae) inferred from nrITS sequences. *Am J Bot.* 2009;96(7):1348–1360.
37. Gao J, Liu X, Zhao Q. Illustrations of the new *Camellia* hybrids that bloom year-round. Zhejiang: Zhejiang Science and technology publishing house; 2016.
38. Xia EH, Zhang HB, Sheng J, Li K, Zhang QJ, Kim C, et al. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol Plant.* 2017;10(6):866–877.
39. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2012;29(1):15–21.
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079.
41. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics.* 2009;25(17):2283–2285.
42. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11(3):R25.
43. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009;26(1):139–140.

Tables

Table 1. Contributions of regulatory divergence to gene expression patterns in F₁ hybrids

	<i>Camellia azalea</i> × <i>C. chekiangoleosa</i>				<i>C. azalea</i> × <i>C. amplexicaulis</i>			
	Additivity	Female dominance	Male dominance	Transgressivity	Additivity	Female dominance	Male dominance	Transgressivity
Conserved	0.00%	39.69%	40.93%	69.01%	0.00%	43.72%	52.56%	76.55%
Ambiguous	22.39%	33.27%	35.29%	18.40%	32.87%	34.23%	29.45%	16.01%
<i>Cis</i> only	13.43%	4.71%	6.79%	2.12%	36.99%	12.32%	8.71%	2.82%
<i>Trans</i> only	47.26%	18.17%	13.50%	5.71%	17.81%	6.51%	5.81%	2.01%
<i>Cis</i> + <i>trans</i>	14.43%	1.53%	1.33%	0.66%	12.33%	2.12%	1.82%	0.33%
<i>Cis</i> × <i>trans</i>	2.49%	1.96%	1.33%	1.54%	0.00%	0.86%	0.85%	0.71%
Compensatory	0.00%	0.67%	0.83%	2.56%	0.00%	0.24%	0.80%	1.57%

Figures

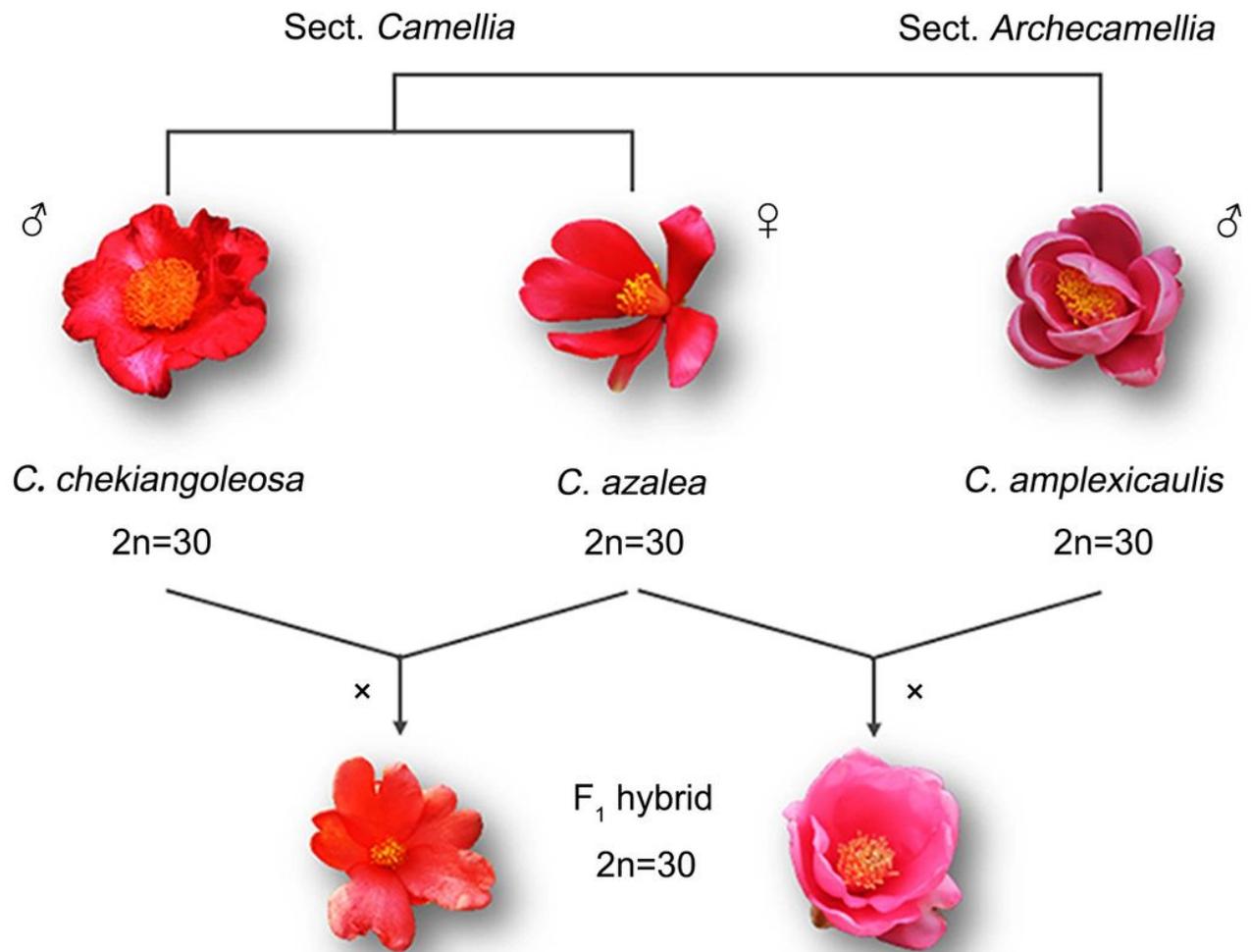


Figure 1

Diagram showing construction of the Camellia hybrids as well as materials used in this study.

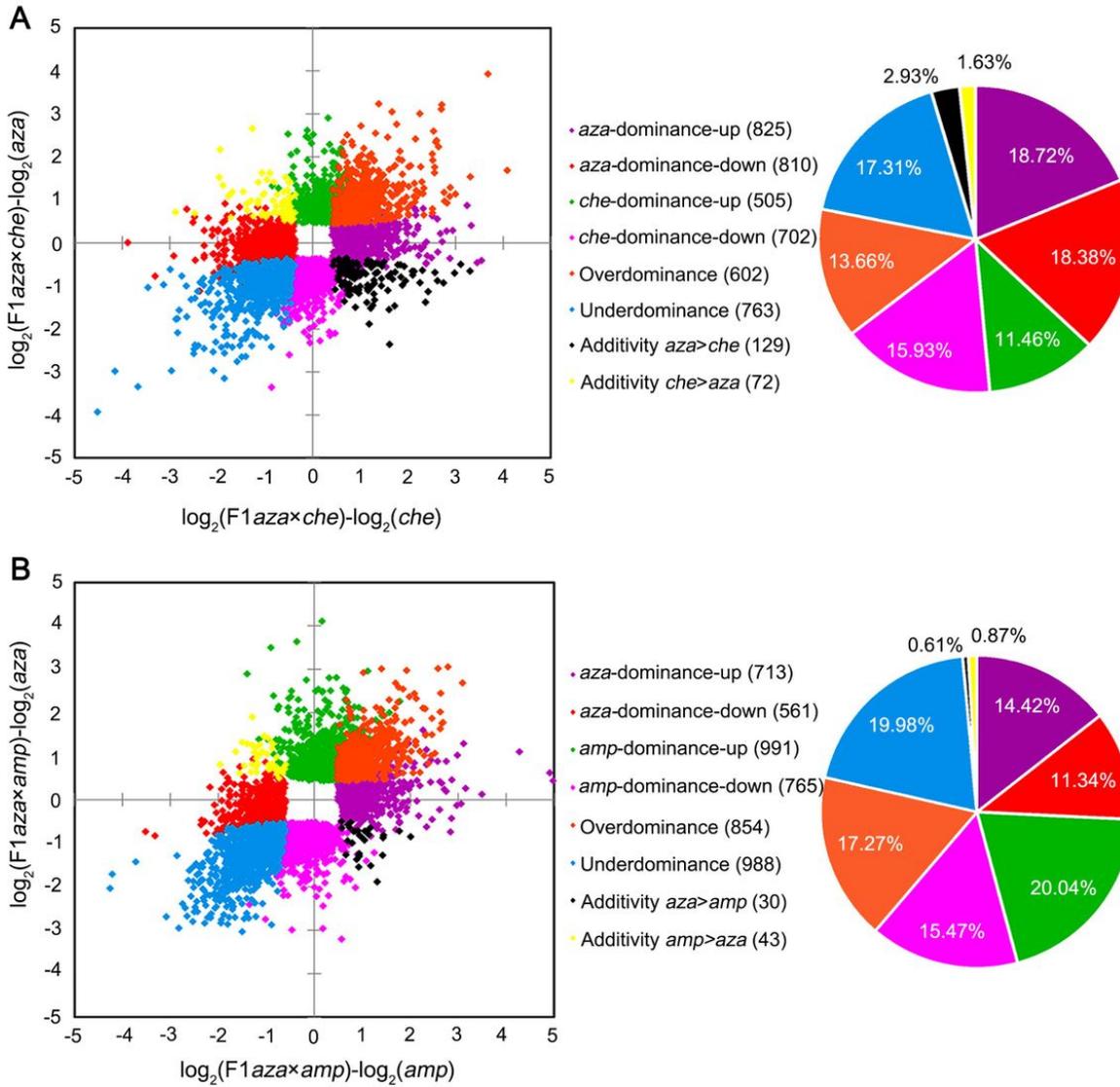


Figure 2

Classification of differentially expressed genes (DEGs) between the F1 hybrids and their parents. According to expression patterns, DEGs detected from the intra- (A) and inter-sectional (B) crosses were further classed into eight clusters as listed in the center of the images, respectively. Numbers in the brackets show genes included in each cluster, and pie charts show the relative proportions of DEGs for each cluster. *aza*, *Camellia azalea*; *che*, *C. chekiangoleosa*; *amp*, *C. amplexicaulis*; F1 *aza* × *che*, F1 hybrid of *C. azalea* × *C. chekiangoleosa*; F1 *aza* × *amp*, F1 hybrid of *C. azalea* × *C. amplexicaulis*. A fold-change of 1.25 combining with FDR < 0.05 were used as threshold for DEGs detection.

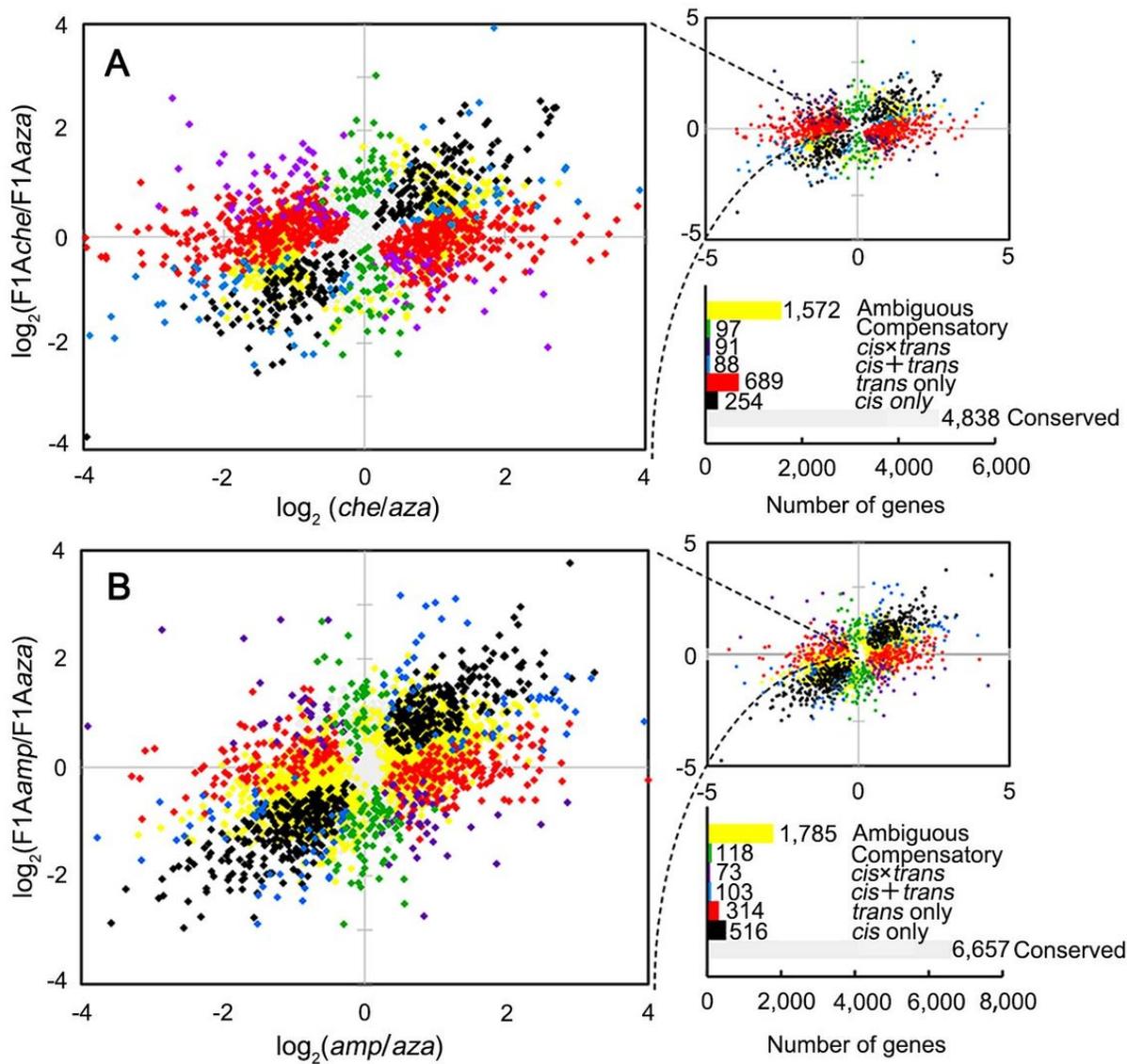


Figure 3

Plots summarize the relative allele-specific gene expression as well as gene regulation patterns in different crosses. (A) The cross of *Camellia azalea* × *C. chekiangoleosa*. (B) The cross of *C. azalea* × *C. amplexicaulis*. Each point represents a single gene and is color-coded according to the regulatory type (as shown in the bar graphs) it is regulated by. *aza*, *C. azalea*; *che*, *C. chekiangoleosa*; *amp*, *C. amplexicaulis*; F1Aaza, allele from *C. azalea* in the F1 hybrid; F1Ache, allele from *C. chekiangoleosa* in the F1 hybrid. F1Aamp, allele from *C. amplexicaulis* in the F1 hybrid.

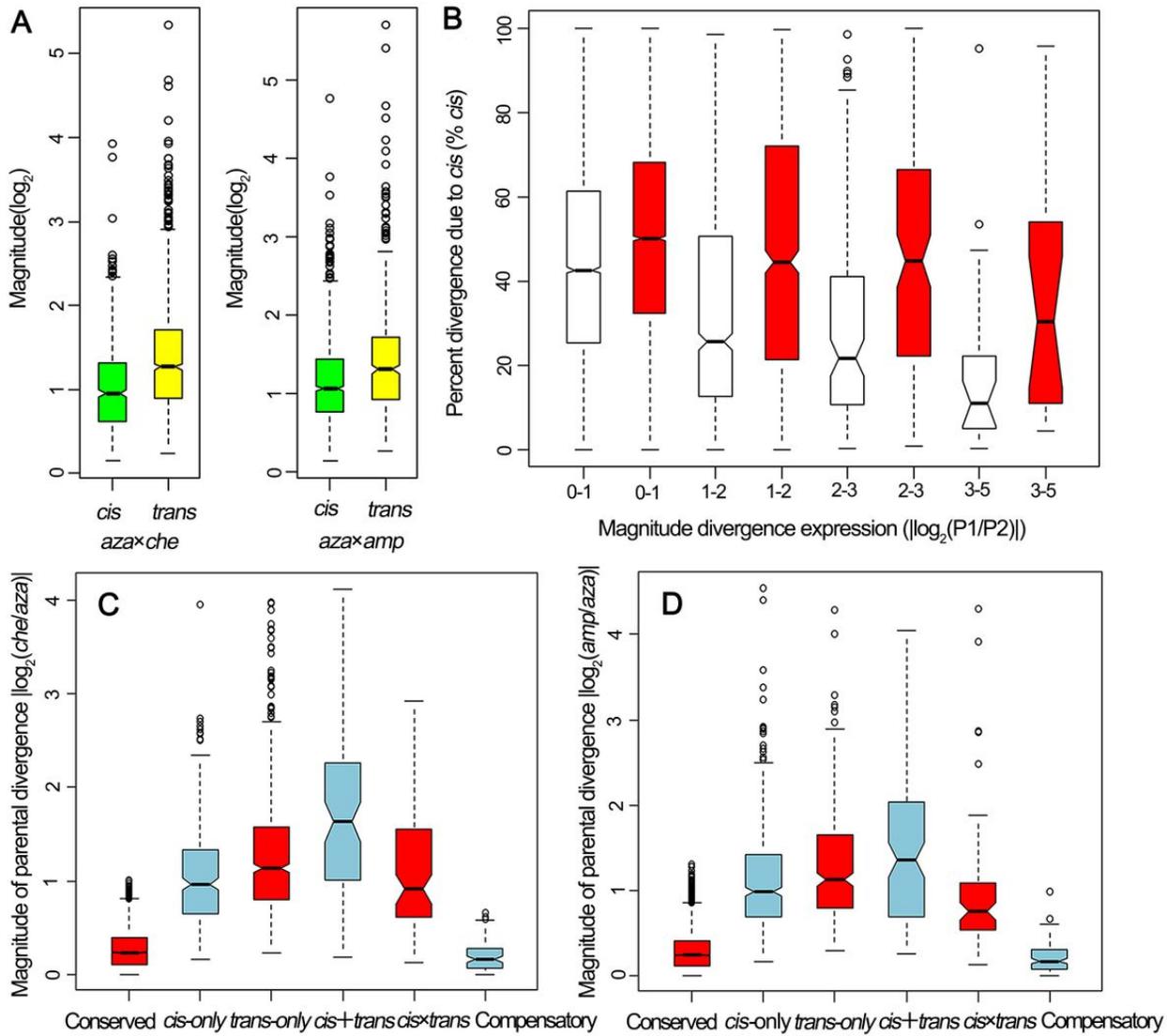


Figure 4

Influence of regulatory types on the expression divergence between the parental species. (A) Absolute magnitude (fold-change) of parental expression divergence resulting from cis- and trans-regulatory changes. aza×che, Comparison between *Camellia azalea* and *C. chekiangoleosa*; aza×amp, Comparison between *C. azalea* and *C. amplexicaulis*. (B) Percentage of total regulatory divergence attributable to cis-regulatory changes (% cis) for genes with different magnitudes of expression divergence between parents. P1, parent1; P2, parent2; Blank, comparison between *C. azalea* and *C. chekiangoleosa*; Red, comparison between *C. azalea* and *C. amplexicaulis*. (C, D) Absolute magnitude (fold-change) of parental expression divergence resulting from different regulatory types. aza, *C. azalea*; che, *C. chekiangoleosa*; amp, *C. amplexicaulis*.

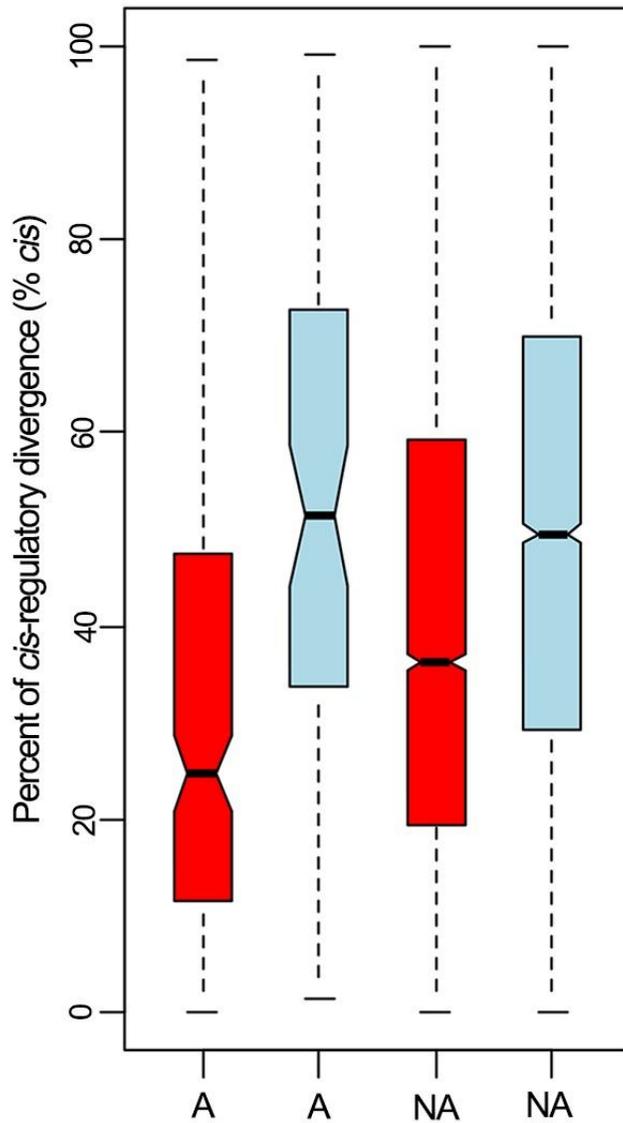


Figure 5

Percent of cis-regulatory divergence for genes showing additive and non-additive expression in *Camellia* F1 hybrids. A, additively expressed genes; NA, nonadditively expressed genes. Red, F1 hybrid of *Camellia azalea* × *C. chekiangoleos*; Blue, F1 hybrid of *C. azalea* × *C. amplexicaulis*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1 TableS3.txt](#)
- [Additionalfile.docx](#)
- [Additionalfile1 TableS4.txt](#)