

# Basil interspecific hybridization and transcriptome study indicates altered developmental and metabolic gene expression

**Saumya Shah**

Central Institute of Medicinal and Aromatic Plants CSIR

**Shubhra Rastogi**

Siksha O Anusandhan University

**Divya Vashisth**

Central Institute of Medicinal and Aromatic Plants CSIR

**Mytra .**

Central Institute of Medicinal and Aromatic Plants CSIR

**R K Lal**

Central Institute of Medicinal and Aromatic Plants CSIR

**Umesh Chandra Lavania**

University of Lucknow

**Ajit Shasany** (✉ [ak.shasany@cimap.res.in](mailto:ak.shasany@cimap.res.in))

Central Institute of Medicinal and Aromatic Plants CSIR

---

## Research article

**Keywords:** amphidiploid, flavonoid, intergenomic stability, interspecific hybrid F1, lignin, phenylpropanoid, polyploidy, RNAseq

**Posted Date:** August 1st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-44832/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 at Plants on July 18th, 2022. See the published version at <https://doi.org/10.3390/plants11141873>.

# Abstract

## Background

In order to understand the developmental modulation of transcriptome and associated gene expression in inter-genomic combinations, a systematic study was planned using two diverse yet closely related species of *Ocimum*, targeting their hybrid F1 and derived amphidiploid (colchiploid of F1 hybrid). The existing developmental alterations between F1 and amphidiploid through phenotypical and anatomical assessments were analyzed.

## Results

Study of several genes and transcription factors putatively involved in the growth and developmental processes of plants clearly amalgamates the transcriptome data linking the phenotypic differences in F1 and amphidiploid. Additionally, differentially expressed genes of stomatal patterning and development revealed their involvement leading to higher density of stomata in F1 while larger size of stomata in the amphidiploid. Absence of 8,330 transcripts of interspecific hybrid F1 in its amphidiploid and exclusive presence of two detected transcripts in amphidiploid provides a set of genes to analyze the suppressed or activated functions between F1 and amphidiploid. Estimation of chlorophyll, lignin, flavonoid and phenylpropanes (eugenol and methyleugenol) content were correlated with the average FPKM and digital gene expression values in F1 and amphidiploid.

## Conclusion

This is the first investigation which describes the genes and transcription factors influenced by interspecific hybridization leading to developmental changes and alleviation of intergenomic instability in amphidiploid.

## Background

All current day flowering plants that appear to be diploid have in fact undergone at least one round of ancient whole genome duplication, suggesting that all angiosperms have polyploidy lineages [1, 2]. It is estimated that up to 25–30% of angiosperms still continue to remain polyploid that have not yet diploidized [3, 4, 5, 6]. Further, it is suggested that autopolyploidy can remodel transcriptome and metabolome [7, 8]. As such polyploidy is considered to be the most important event in diversification and speciation of flowering plants [9]. It has been observed that polyploidy leads to increased genetic variation offering ecological and evolutionary advantages including robustness, increased environmental fitness and tolerance to a broader range of ecological and environmental conditions, enhanced photosynthetic efficiency, resistance to abiotic and biotic factors [6, 10, 11], enhanced productivity of secondary metabolites [12], increased cell size, stomata and vascular cells [13, 14], however, bestow

differential effect on body size [15, 14]. Also owing to enhanced variation and adaptation potential the polyploids could be 20% more invasive than their closely related diploids [16].

Since the events of polyploidization are preceded by hybridization in allopolyploids, therefore, genomic doubling could drive multiple changes in gene expression, including gene silencing or chromosomal changes [17, 18], loss and retention of duplicate genes [19], physiological divergence [20] and epigenetic modifications [14]. One of the important observations in this context is the possible dominance of one genome over the other and repeated patterns of genetic change in the natural and synthetic hybrids including detection of transcriptomic shock altering gene expression, DNA methylation and epigenetic modification [21, 22, 23]

However, the relative role of hybridization vs. genome doubling as drivers of genetic and genomic change has not yet been addressed [23], although such knowledge could be quite important in breeding of plants where genetic enhancement of vegetative organs, tissues and the secondary metabolites produced therein are the source of economic product. The present investigation was therefore planned on two closely related but diverse species of the model plant *Ocimum*, their interspecific hybrids and genomically doubled amphidiploids to elucidate developmental changes occurring during intergenomic stabilization and associated transcriptomic insights and gene expression using next generation sequencing approach.

*Ocimum* is an important and the largest genus of mint family, Lamiaceae. The genus *Ocimum* collectively called as basil is represented by many different varieties having incomparable curative properties and unique chemical compositions [24]. The genus shows high degree of morphological, chemical and genetic differences at inter and intra specific levels [25]. According to karyomorphological studies performed on different varieties of *Ocimum*, it was suggested that there also exist a great variation in the chromosome number across the genus, 'n' number ranging from 11–38 (<http://ccdb.tau.ac.il/search/Ocimum/>). This shows that polyploidy is a common event in *Ocimum* spp. and has played an important role in the course of evolution of this genus. Conventional breeding techniques and other ploidy manipulation tools have been actively used in the genetic improvement of *Ocimum* species for developing better plant having high yielding essential oils and other bioactive molecules [24]. In this study, an interspecific hybrid of *Ocimum basilicum* and *Ocimum kilimandscharicum* and its amphidiploid plants were used. Further assessment of the plants revealed that the interspecific hybrid F1 were sterile with smaller leaf area and taller plant height. However, the fertility was restored in genomically double amphidiploids which had larger leaf area but shorter plant height. These developmental variation in interspecific hybrids of *Ocimum* generated interest to carry out the high-throughput *de novo* transcriptome sequencing and digital gene expression profiling of parent plants (*O. basilicum* and *O. kilimandscharicum*), interspecific hybrid F1 and amphidiploid plants.

To elucidate changes in gene expression, the differential gene expression (DGE) profile of all four plants (*i.e.* the two progenitor diploids, their interspecific hybrid and derived amphidiploids) were compared. Detection of gene loss, silencing and activation showed that the hybridization and whole genome doubling triggers the gene expression *via* genetic and epigenetic alterations immediately upon

allopolyploid formation. The overall influence of hybridization and whole genome duplication on the genes related to chlorophyll metabolism as well as lignin and phenylpropanoid biosynthesis pathways was also investigated. Additionally, many transcripts related to stomatal patterning and development were differentially regulated in F1 vs. amphidiploid indicating the role of these genes in the higher stomatal density in interspecific hybrid F1 and larger size of stomata in amphidiploid. In particular, several transcription factors (TFs) having possible role in morphological/ anatomical characteristics and in different metabolic processes such as phenylpropanoid biosynthesis (flavonoid and lignin biosynthesis), chlorophyll biosynthesis and catabolism were also identified. This study provides deep understanding between parents, F1 and the genomically doubled amphidiploids. The results of this investigation explain the underlying mechanisms responsible for the developmental changes in interspecific hybrid F1 and amphidiploids underpinning the significance of associated changes in the gene expression.

## Results

### Comparative phenotype

Morphologically interspecific hybrid F1 and its colchicine induced amphidiploid were more similar to its parent 2 (OKP2). However, to identify the overall effect of polyploidization, different phenotypic features like plant height (**Additional Fig. S1 A-D**), leaf area, inflorescence (**Additional Fig. S1 E**), stem diameter (**Additional Fig. S1 F-I**), trichome density (**Fig. S2**), stomata *etc.* of amphidiploid, hybrid F1 and its parents (OBP1 and OKP2) were measured (**Table 1**). On comparing interspecific hybrid F1 with its parent plants and amphidiploid, F1 was found to be robust, rapidly growing, vigorous and tall ( $110.00 \pm 8.9$ ). However, the leaf area of F1 ( $3.5 \pm 0.18$ ) was lesser than its amphidiploid ( $9.63 \pm 0.75$ ) and its parent plants. The leaf of interspecific hybrid F1 was long, medium broad and thin. Also, the inflorescence and stem were weak in interspecific hybrid F1 but the length of inflorescence of interspecific hybrid F1 was longer (nearly two fold) than amphidiploid and its parents. In contrast, amphidiploid was slow growing and medium tall ( $101.50 \pm 9.30$ ). The leaf of amphidiploid was oval shaped, broad, thick and the leaf area was greater than its parents. Besides these characteristics, trichome density was more in the hybrid F1 (nearly threefold higher than amphidiploid). Scanning electron microscopy revealed that size of trichome and stomata (nearly threefold greater than F1) was greater in amphidiploid (**Fig. 1**). Furthermore, the density of stomata was more in interspecific hybrid F1. However, oil yield per 100gm of fresh leaf was more in amphidiploid ( $0.48 \pm 0.02$ ).

### Chromosome number

The results obtained from root-tip mitosis of the four target plants revealed the modal somatic chromosome number to be as:  $2n$  (AA) = 48 for OBP1,  $2n$  (BB) = 76 for OKP2,  $n+n$  (A+B) = 62 for interspecific hybrid F1, and  $2n+2n$  (AA+BB) = 124 (**Additional Fig. S3 A-D**). This is in conformity that the F1 hybrid and the amphidiploids constitute the genuine genomic combination of the two progenitor parents employed in the present study.

## Transcriptome Sequencing, Assembly and Annotation

Pair-end sequencing generated nearly 14,857,263 (OBP1), 13,017,783 (OKP2), 14,266,805 (interspecific hybrid F1), and 13,955,543 (amphidiploid) reads of 101bp average read length. After filtering and removing adapter sequences from raw reads, 14,431,491 (OBP1), 12,665,614 (OKP2), 13,866,646 (interspecific hybrid F1) and 13,479,705 (amphidiploid) high quality reads were acquired for further assembly (**Table S1**). High quality reads retained after filtering raw reads contained 1,437,320,230 (OBP1), 1,261,764,356 (OKP2), 1,382,322,028 (interspecific hybrid F1) and 1,334,646,724 (amphidiploid) high quality bases. High quality reads were used for merged assembly using Trinity assembler which generated total 243,647 numbers of transcripts of 148 Mbp in length. Total number of final transcripts obtained after the final assembly was 91,778 having final transcriptome length of 66 Mb with N50 value 942. BlastX search was carried out for 91,778 clustered transcripts against NRDB (Non redundant database) plant using e-value  $\leq 0.001$ . Out of the total only 78,928 transcripts got annotated while 12,850 remained un-annotated.

## Comparative gene expression patterns between parents and progenies

The parent plants (OBP1 vs. OKP2) were investigated for their differential gene expression patterns (**Fig. 2A**). Further, the differentially expressed genes between F1 hybrids or amphidiploid compared to their parents were also analyzed (**Fig. 2B**). Out of 172 differentially expressed genes between interspecific hybrid F1 and parent 1 (OBP1), 60 showed up-regulation and 112 down-regulation. From 123 differential genes between amphidiploid and OBP1, 51 were up-regulated and 72 down-regulated. Similarly, 155 differentially expressed genes (47 Up-regulated and 108 down regulated) were obtained from the analysis of interspecific hybrid F1 and parent 2 (OKP2) and 215 (82 up-regulated and 133 down regulated) for amphidiploid and parent 2 (OKP2). In all comparisons, it was observed that the proportion of genes displaying the differential expression between interspecific hybrid F1 and amphidiploid and their parents was asymmetric (FDR<0.05; BH multiple correction test).

In order to identify the non-additively expressed genes, the expression levels of amphidiploid and hybrid F1 was compared with the mid parent values (MPVs) derived from the base mean values of the two parents assuming that one-third of the total transcription is from the genome from parent 1 (OBP1), one-third from parent 2 (OKP2) and one-third is from the interspecific hybrid F1. On the other hand, interspecific hybrid F1 was compared to the parent 1 (OBP1) and parent 2 (OKP2) assuming that the half of the total transcripts is from the genome of each parent (OBP1 and OKP2), respectively. The non-additively expressed genes between the F1 hybrid and MPV of parent1 and parent2 (OBP1 and OKP2) was found to be 38, of which 10 genes were up-regulated and 28 genes were down-regulated (**Fig. 3A**; FDR< 0.05). Likewise, of 786 non-additively expressed genes observed between the amphidiploid and the MPV of Parent1, Parent2 and F1, 395 genes showed up-regulation and 391 genes exhibited down-regulation (**Fig. 3B**; FDR <0.05). GO analysis of 10 up-regulated genes in F1 hybrid and MPV of OBP1 and OKP2 indicated enrichment of photosynthesis, terpene biosynthesis including sesquiterpenoid metabolism and lipid biosynthesis (**Additional Fig. S4A**; FDR< 0.05; **Table S2**) whereas, the 28 down

regulated genes in F1 hybrid and MPV of OBP1 and OKP2 were mainly associated with lignan biosynthesis and metabolism, steroid metabolism (**Additional Fig. S4B**; FDR< 0.05; **Table S3**). In contrast, the 395 up-regulated genes in amphidiploid and the MPV of OBP1, OKP2 and F1 revealed the enrichment of developmental vegetative growth, regulation of leaf development and regulation of shoot apical meristem development (**Additional Fig. S4C**; FDR< 0.05; **Table S4**) while the 391 down regulated genes in the amphidiploid and MPV of OBP1, OKP2 and F1 were mainly enriched in flavonoid metabolism, starch catabolism, s-adenosylmethionine metabolism (**Additional Fig. S4D**; FDR< 0.05; **Table S5**).

### Hybridization induced transcript expression

To understand the reasons for developmental changes between interspecific hybrid F1 and amphidiploid, pair-wise comparison between OBP1 vs. F1&Amphid2, OKP2 vs. F1&Amphid2, F1 vs. OBP1&OKP2 and Amphid2 vs. F1,OKP2&OBP1 were analyzed (**Fig. 4A**). 244 transcripts out of 38,040 common transcripts (obtained from pair-wise comparison between OBP1 vs. F1&Amphid2, OKP2 vs. F1&Amphid2, F1 vs. OBP1&OKP2 and Amphid2 vs. F1,OKP2&OBP) exhibited antagonistic expression pattern ( $\log_2$  fold  $-1 \geq < 1$ ) in interspecific hybrid F1 and amphidiploid. Further analysis revealed similar expression pattern of these 244 transcripts in amphidiploid and parent 2 (OKP2). Among these 126 transcripts were up regulated in amphidiploid and parent 2 whereas these transcripts were down regulated in interspecific hybrid F1 (**Fig. 5A**; **Table S6**). In contrast to this, 118 transcripts were having higher expression in interspecific hybrid F1 in comparison to amphidiploid and parent 2 (**Fig. 4C**). These results indicated that the overall transcriptome of amphidiploid was more similar to its parent 2 (OKP2) and it matches to the morphological analysis which shows that amphidiploid plants were more similar to its parent 2 (OKP2). To identify the possible function of these transcripts GO enrichment analysis was conducted. This analysis revealed that these genes were mainly enriched to biological function such as shoot system development, flower development, cotyledon morphogenesis, embryonic morphogenesis, reproductive structure development, programmed cell death (**Fig. 4D**). It was also observed that many transcripts were reported as "PREDICTED" cellulose synthase-like protein G3 (proposed to synthesize non cellulosic polysaccharides that comprise plant cell walls)", "1-deoxy-D-xylulose 5-phosphate synthase (catalyzes the first step of the MEP pathway)", SWI/SNF complex component SNF12 homolog (activator of flower homeotic genes)", oligopeptide transporter 3 (essential for early embryo development)", casein kinase II subunit beta-like (involved in flowering-time regulation)", AMP deaminase-like isoform X2 (essential for the transition from zygote to embryo)", SNW/SKI-interacting protein (Splicing factor involved in post-transcriptional regulation of circadian clock and flowering time genes)," nuclear transcription factor Y subunit A-10 (positive Regulators of Photomorphogenesis)", "Calmodulin-domain protein kinase 5 isoform 1 (involved in the many aspects of plant growth and development)", "Epidermal patterning factor-like protein 4 (negative regulator of stomatal development)", "WAT1-related protein At4g19185-like (involved in the secondary wall formation)", "protein STRICTOSIDINE SYNTHASE-LIKE 11 (involved in anther development and pollen wall formation)" (**Table S7**). It was also noticed that many of the transcripts were predicted as "uncharacterized protein LOC105176273, hypothetical protein POPTR\_0001s256302g, hypothetical protein MIMGU\_mgv1a009003mg, hypothetical protein MIMGU\_mgv1a005332mg, hypothetical protein JCGZ\_24107, hypothetical protein M569\_05704,

uncharacterized protein LOC105177873, uncharacterized protein LOC105176091, uncharacterized protein LOC105165185, uncharacterized protein LOC105164617 isoform X1, *etc.*" and many of them were left unannotated, perhaps due lack of its annotation in *Ocimum*. Further investigation of these transcripts could provide good candidates for understanding the role of genome doubling in the correction of phenotypic weakness induced by hybridization. Interestingly, these transcripts could be utilized for the identification of novel genes probably having their role in morphological and anatomical differences between parents and hybrid and the amphidiploid plants.

### **Comparison of gene expression between the F1 hybrid and amphidiploid**

Between colchicine induced amphidiploid and F1 hybrids, 179 differentially expressed genes (DEGs) were identified including 132 up-regulated and 47 down-regulated genes when BH multiple test correction method was applied (Benjamini and Hochberg, 1995) having FDR < 0.05 (**Additional Fig. S5A**). These DEGs were mapped to reference canonical pathways in KEGG to find out their involvement in biological pathways and 44 out of 179 DEGs were assigned to 46 KEGG pathways (**Additional Fig. S6**). The largest cluster was of biosynthesis of secondary metabolites with 13 members and the second largest was of metabolic pathways with 9 members, indicating that many genes among these DEGs were involved in the biosynthesis of secondary metabolites. Thereafter, to investigate the probable function of these DEGs, GO enrichment analysis was performed. The results of GO enrichment analysis showed that they were mainly enriched in the secondary metabolic processes like sesquiterpenoid biosynthetic and metabolic process, isoprenoid biosynthetic process, jasmonic acid metabolic process in biological process category (**Additional Fig. S5B**) and were enriched in auxin:proton symporter activity, fatty-acyl-CoA reductase (alcohol-forming) activity, magnesium ion binding, lyase activity, cyclase activity *etc* (FDR < 0.05) (**Additional Fig. S5C**).

### **Detection of gene expression alterations in interspecific hybrid F1 and amphidiploid**

Gene alteration events were calculated by the occurrence of new transcripts (lacking in parents) or by the lack of some transcripts (existing in parents) in interspecific hybrid F1 and amphidiploids. For this analysis, total transcripts of parent 1 (67,770), parent 2 (73,265), interspecific hybrid F1 (76,917) and the amphidiploid (76,563) were examined (**Fig. 5**). The result of this analysis illustrated that 5,766 common transcripts of parents and interspecific hybrid F1 were not detected in amphidiploid. In addition to this, about 6,432 transcripts present in interspecific hybrid F1 and amphidiploid were missing in parents. However, of these 6,432 transcripts, only 3,868 transcripts were common in interspecific hybrid F1 and amphidiploid. Therefore, total 8,330 transcripts of interspecific hybrid F1 were absent in amphidiploid. On the other hand, only two transcripts were found to be exclusive in amphidiploid with respect to interspecific hybrid F1 and parents (OBP1 and OKP2). These alterations in gene expression may be because of gene silencing, activation or may be due to sequencing error, but here it was assumed that these transcripts were either suppressed or expressed in amphidiploid. Upon analyzing the annotations of these non-expressing 8,330 transcripts in amphidiploid, it was observed that these transcripts mainly included the genes which were involved in disease resistance, primary and secondary metabolism and

cell cycle. It also included many transcription factors (“basic helix-loop-helix transcription factor”, “MYB/MYB-related”, “MADS-box”, “APETALA”, “AP5/EREBP”, “WRKY” *etc*), cytochrome p450s and methyl-CpG-binding domain-containing proteins. Moreover, there were many transcripts which were predicted as “uncharacterized or unnamed protein” and many of them were left un-annotated also. In contrast, the 2 transcripts exclusive to amphidiploid were WNK lysine deficient protein kinase and geranylgeranyl transferase type-2 subunit beta 1-like proteins.

### **Chlorophyll content and DGE related to chlorophyll biosynthesis between F1 hybrid and amphidiploid**

The result of chlorophyll estimation showed that the amount of Chla, Chlb and the total chlorophyll was 0.35 mg/g, 0.13 mg/g and 0.54 mg/g, respectively in interspecific hybrid F1 and 0.30 mg/g, 0.11 mg/g and 0.37 mg/g in amphidiploid. Here, it was observed that the amount of Chla, Chlb and the total chlorophyll contents were higher in the interspecific hybrid F1 than its amphidiploid (**Fig.7A**). To find the probable reason for this content change, 167 transcripts in interspecific hybrid F1 and 169 transcripts in amphidiploid related to 27 classic enzymes involved in chlorophyll metabolic pathway were analyzed (**Table 2**). Chlorophyll metabolic pathway in plants consists of ALA, Proto IX, heme and chlorophyll formation/degradation steps. In ALA, Proto IX, heme and chlorophyll biosynthesis steps, higher number of transcripts for enzymes like *HemA*, *HemE*, *HemY*, *HemH*, *COX15*, *POR* and *CAO* were detected while fewer transcripts for *HemF*, *COX10*, *ChlD*, *ChlI*, *ChlM*, *ChlE*, *4VCR* and *CLH* were recorded. Single copy of enzyme ChlH and ChlG were identified from the transcriptome sequences of the interspecific hybrid F1 and amphidiploid. In the chlorophyll degradation steps, more than one transcript was identified for enzymes like *NYC1*, *HCAR*, *PPH*, *PAO* and *RCCR*. FPKM values and fold change values for 167 common transcripts in interspecific hybrid F1 and amphidiploid were averaged for the further analysis. Based on these identified transcripts, proposed chlorophyll metabolic pathway in interspecific hybrid F1 and amphidiploid was constructed (**Fig.6**).

The difference in the content change of Chla, Chlb and total chlorophyll were correlated with the average FPKM and the fold change values of the 27 classic enzymes involved in the chlorophyll metabolism. The trend of change in expression based on the average FPKM and fold change values ( $\log_2$  fold change  $-0.1 \geq < 0.1$ ) of the main enzymes (*HemA*, *HemL*, *HemC*, *HemE*, *HemF*, *HemY*, *HemH* and *COX10*) involved in the ALA, Proto IX, Heme and the key enzymes (*ChlI*, *ChlII*, *ChlM*, *ChlE*, *4VCR*) of chlorophyll formation showed increased expression in interspecific F1 compared to amphidiploid. On the other hand, several enzymes (*HCAR*, *PPH*, *PAO* and *RCCR*) involved in the chlorophyll degradation showed increased expression in amphidiploid compared to interspecific hybrid F1. These results clearly demonstrated that the enzymes involved in chlorophyll biosynthesis (ALA, Proto IX, Heme and in Chlorophyll formation) were positively regulated in interspecific hybrid F1 but negatively regulated in amphidiploid. In contrast, the enzymes involved in chlorophyll degradation (*NYC1*, *PPH*, *PAO* and *RCCR*) were negatively regulated in interspecific F1 while positively regulated in the amphidiploid. These results correlates with the results obtained from the chlorophyll estimation.

### **Phenylpropanoid profiling and DEG related biosynthesis in F1 hybrids and amphidiploid**

Phenylpropanoid biosynthetic pathway is the predominant pathway present in the different *Ocimum* spp. which produces different phenylpropenes, lignins and flavonoids. Therefore, to understand the effect of interspecific hybridization and whole genome duplication on phenylpropanoid biosynthesis, GC-MS profiling of essential oils of F1 and amphidiploid, total content change of lignin and flavonoid was analyzed (**Fig. 7 B-D**). The essential oil analysis showed that the amount of eugenol was 0.082 mg/g leaf in interspecific hybrid F1 and 0.063 mg/g leaf in amphidiploid whereas, the amount of methyleugenol was 0.087 mg/g leaf and 0.032 mg/g leaf in interspecific hybrid F1 and amphidiploid, respectively. In addition, the content of total flavonoid and total lignin was higher in the amphidiploid than interspecific hybrid F1. In addition, the amount of flavonoid and lignin was 0.26 mg/g leaf and 0.63 mg/g leaf, respectively in interspecific hybrid F1 and 0.41 mg/g leaf and 1.03 mg/g leaf, respectively in amphidiploid. Thus, it was found that amount of eugenol and methyleugenol was higher in interspecific hybrid F1 but the amount of total flavonoid and total lignin was higher in amphidiploid. Several transcripts corresponding to 16 enzymes directly involved in the general phenylpropanoid biosynthesis pathway were identified and analyzed to address the question of differential biosynthesis of phenylpropanoids, lignin and flavonoids in the interspecific hybrid F1 and amphidiploid (**Table 3**). Among these enzymes, *PAL*, *C4H* and *4CL* are mandatory enzymes catalyzing the initial three steps of the phenylpropanoid pathway, while *HCT*, *CCR*, *COMT*, *CCoAOMT* and *CAD* are downstream enzymes directly involved in the biosynthesis of lignin in plants. These enzymes including *PAL*, *C4H* and *4CL* belong to multigene family and hence, more than one transcript for these enzymes was identified. Similarly, for *EGS* and *EOMT* belonging to small gene families responsible for the production of eugenol and methyleugenol, respectively in *Ocimum*, only three transcripts for *EGS* and one transcript of *EOMT* were identified in the transcriptome. *CHS* is the first enzyme of flavonoid biosynthesis producing the first flavonoid naringenin chalcone with the involvement of 4-coumaroyl-CoA and three molecules of malonyl-CoA and *CHI*, the key enzyme of flavonoid biosynthesis pathway catalyses intramolecular cyclization of naringenin chalcone into naringenin. About 6 transcripts of *CHS* and 7 transcripts of *CHI* were identified in both interspecific hybrid F1 and amphidiploid. Reduction of dihydroflavanols at position 4 is catalyzed by enzyme belonging to *DFR* superfamily. About 14 and 15 transcripts of *DFR* were detected in interspecific F1 and amphidiploid, respectively. Likewise, 8 and 7 transcripts of *F3'H* and only 1 transcripts of *F3'5'H* and *F3H* were recorded in interspecific F1 and amphidiploid. About 4 transcripts of *UFGT* enzyme responsible for converting anthocyanidin to anthocyanin were found in the transcriptome sequences of interspecific F1 and amphidiploid. But the transcripts of *CCoA-3H*, *F5H*, *CAAT* and *ANS/LDOX* could not be annotated. The means of FPKM values and the values of differential gene expression of 184 common transcripts based on the above identified transcripts were used to correlate the trend of content change of phenylpropenes with its gene expression (**Fig. 8**).

The average of FPKM values and differential gene expression ( $\log_2$  fold change  $-0.1 \geq \leq 0.1$ ) showed that the expression of four crucial genes (*HCT*, *CCR*, *COMT* and *CAD*) directly involved in the biosynthesis of lignin were up-regulated in amphidiploid but down-regulated in interspecific F1 hybrids. Conversely, expression of *PAL*, *C4H*, *4CL*, *EGS*, *EOMT* gene were highest in the interspecific hybrid F1 while the expression of genes involved in flavonoid biosynthesis (*CHS*, *CHI*, *F3'H*, *DFR* and *UFGT*) were down-

regulated in interspecific hybrid F1. To validate the gene expression profile of RNAseq Data, qPCR of seven genes involved in general phenylpropanoid biosynthesis was performed (**Fig. 9**). The result of qPCR correlates the decreased expression of *COMT*, *CAD*, *CHS*, *DFR* genes involved in lignin and flavonoid biosynthetic pathway and increased expression of *PAL*, *4CL*, and *EGS* in interspecific hybrid F1 confirming the RNAseq data.

### **Identification of candidate DEGs involved in higher stomatal density in interspecific hybrid F1 and larger stomatal size in amphidiploid**

Higher stomatal density in interspecific hybrid F1 and larger stomatal size in amphidiploid are the two peculiar characteristics which were observed through anatomical analyses in mature leaves. For this reason, 172 candidate DEGs ( $\log_2$  fold change  $-0.1 \geq < 0.1$ ) putatively associated in the stomatal patterning and development were identified (**Table S8**). The means of FPKM values and DEGs of the identified genes showed that most of the identified genes including *TOO MANY MOUTHS (TMM)* which promotes cell fate progression in stomatal development, *EPIDERMAL PATTERNING FACTOR 4 (EPF 4)*; a negative regulator of stomatal development, receptor like proteins such as *CLAVATA1/ CLAVATA1-like*, *CLAVATA3/ CLAVATA3-like* receptor kinases, *ERECTA/ ERECTA-like*, various type mitogen-activated protein kinase (*MAPK*) such as *MAPK9/MAPK9-like*, *MAPK15/MAPK15-like*, *MAPK19/MAPK19-like*, *MAPK20/MAPK20-like*, mitogen-activated protein kinase homolog *MMK1-like*, mitogen-activated protein kinase homolog *NTF3*, mitogen-activated protein kinase kinase kinase *MAPKK2/MAPKK2-like*, *MAPKK5/MAPKK5-like*, *MAPKK6/MAPKK6-like* were down regulated in interspecific hybrid F1 and up-regulated in amphidiploid. Besides this, MAP Kinase Kinase Kinase *YODA (YDA)* which act as a molecular switch in stomatal patterning and development were also negatively regulated in the interspecific hybrid F1 and positively regulated in amphidiploid. In contrast, *EPIDERMAL PATTERNING FACTOR-like* protein 9 (*EPF9*)/Stomagen and subtilisin-like protease *SDD1*, which act a positive regulator of stomatal density and patterning were up-regulated in interspecific hybrid F1. Additionally, leucine-rich repeat receptor-like protein *CLAVATA2*, mitogen-activated protein kinase *MAPK4/MAPK4-like*, *MAPK7/MAPK7-like*, *MAPK10/MAPK10-like*, *MAPK16/MAPK16-like*, mitogen-activated protein kinase homolog *NTF4-like*, mitogen-activated protein kinase homolog *MMK2-like*, mitogen-activated protein kinase kinase kinase *MAPKKK1/MAPKKK1-like*, *MAPKKK3/MAPKKK3-like* etc were also up-regulated in interspecific hybrid F1. These results suggests that these differentially expressed genes (DEGs) in interspecific hybrid F1 and amphidiploid are involved in the higher stomatal density and larger stomatal size in interspecific hybrid F1 and amphidiploid, respectively.

### **Analysis of putative transcription factors involved in the altered phenotype of interspecific F1 and its improvement in amphidiploid**

TFs play an important role in the growth and development of all organisms. In the present work, about 1,504 transcripts from interspecific hybrid F1 and 1,537 transcripts from amphidiploid were identified as TFs. These were categorized under 58 families as per plant transcription factor database PlantTFDB 4.0 (<http://planttfdb.cbi.pku.edu.cn/>) [26]. Of these, MYB and MYB-related transcription factors were found

to be most abundant TF family, followed by WRKY, AP2/ERF and NAC. Total 8 TFs were related to bZIP superfamily which is reported to be associated in various biological processes such as flower and vascular development, embryogenesis, organ differentiation, seed maturation. TFs which were annotated to have transcription factor activity but do not fall in any of the families as classified by plant transcription factor database PlantTFDB 4.0 were specified as 'other' (**Fig.10**). The differential gene expression ( $\log_2$  fold change  $-0.1 \geq \leq 0.1$ ) and average FPKM values indicated highest number for MYB and MYB related super-family (35) followed by WRKY (33), NAC (17), MADS-BOX (14) and ethylene-responsive transcription factor (13) which are differentially expressed between interspecific hybrid F1 and amphidiploid. Also, a large number of TFs in WRKY (25) superfamily, MYB and MYB-related superfamily (22), NAC superfamily (12), MADS-BOX (7) and ethylene-responsive transcription factor (7) were down-regulated in interspecific hybrid F1 and up-regulated in amphidiploid. All the differentially expressed TFs were summarized according to their probable involvement in the floral development, leaf development, trichome development, seed development and xylem formation. Many other TFs putatively showing their role in different metabolic pathways such as chlorophyll biosynthesis/ catabolism, phenylpropanoid biosynthesis, flavonoid biosynthesis and anthocyanin biosynthesis were also analyzed and listed (**Table S9**). In addition, many TFs involved in growth and developmental processes such as flowering time (MADS7, SOC1, FAR1-RELATED SEQUENCE 8 and 6-like, CO-LIKE 9 *etc*), floral organ development (AP2, GAMYB, AGL9, SPL11, AGL8), cell expansion in leaf and shape (GRF1-interacting factor 2-like, GRF1-interacting factor 3, BEL1-like homeodomain protein 4), seed development (NAC025, bZIP11, DOF3.7-like) and xylem formation (NAC07, MYB48, MYB 86, GATA 12-like) were down-regulated in interspecific hybrid F1 and up-regulated in amphidiploid. Similarly, many TFs related to the repression of phenylpropanoids and regulation of flavonoids (NAC078, MYB3-like, MYB4) also showed down-regulation in interspecific hybrid F1. TFs involved conferring flowering time delay (RAP2-7), trichome development (MYB-like transcription factor *ETC3*, trihelix transcription factor GTL1, WRKY 44-like) were up-regulated in interspecific hybrid F1 and down-regulated in amphidiploid. Additionally, the TFs having their contribution in chlorophyll degradation (NAC029, NAC100-like) and anthocyanin biosynthesis (MYB10) were negatively regulated in the interspecific hybrid F1 and positively regulated in amphidiploid. Thus, the expression of these TFs specifies their role in the altered phenotype of interspecific hybrid F1 and amphidiploid.

## Discussion

### Amphidiploidy alleviates genomic instability in the inter-specific hybrid

Polyploidization is an important event of speciation and evolution leading to generation of new forms adapting to new ecological niches. Often this occurs naturally but can be realized through breeding. Many of the modern day crop plants have evolved through polyploidization. This technique is being extensively used by the plant breeders and such events are thought to have a profound effect on genome structure and gene expression [26]. It is well accepted that hybridization between diverse genotypes often generates hybrid vigour and combination of characters, but may sometimes lead to hybrid dysgenesis.

However, such interspecific hybrids that involve diverse genomes often suffer from developmental and reproductive deficiency on account of cell cycle incompatibility and meiotic disturbances. Nevertheless, the latter deficiency could be overcome by genomic duplication / amphidiploidization, as observed in the instant inter-specific F1 hybrid and its derived amphidiploids.

Polyploidization often causes developmental changes in plants such as increased plant height, enlarged cell and organ size, higher biomass and other phenotypic variation. In Ma bamboo (*Dendrocalamus latiflorus* Munro) different polyploidy levels displayed altered anatomical, physiological and growth characteristics, such as leaf thickness, fusoid cell and stomatal size, shoot number, photosynthesis and respiration rate *etc* [27]. *Senecio cambrensis* generated through chromosome doubling in the sterile triploid hybrid *S. × baxteri* produced fruits [28]. Though these phenomena are age old and very well known, yet the changes occurring at expression levels has not been systematically studied with respect to the parents. Hence, this study for the first time tends to describe the genes and transcription factors influenced by interspecific hybridization leading to developmental changes and alleviation of intergenomic instability in amphidiploids. Here, it was found that hybridization and genome duplication have abrupt but noticeable effects on the gene expression patterns. The hybridization event strongly alters the parental gene expression patterns which get ameliorated after genome duplication. Therefore, transcriptome was analysed as the dynamics of expression changes keeping the nature of genome constant.

### **Gene silencing and activation in interspecific hybrid F1 and amphidiploid**

Chromosome doubling after interspecific or intergeneric hybridization, leads to the development of new allopolyploid species [29]. Newly formed allopolyploids must overcome the reduced fertility (occurred due to improper chromosome pairing and segregation) in order to prove themselves as a successful species, and associated alteration of gene expressions [30, 18]. These alterations in gene expression may occur because of gene silencing and activation. However, the type of gene affected and the probable mechanism involved in the ploidy regulation of gene expression are still challenging [31]. Very few works have addressed such responses in gene alterations in eukaryotic system. In yeast, Galitski *et al.* [32], showed that ploidy-regulated activation and silencing of genes were mainly related to cell growth and development. In a newly synthesized wheat allotetraploid the silenced/lost genes included rRNA genes and genes involved in metabolism, disease resistance, and cell cycle regulation but the activated genes were of known function and all were retroelements [29]. In this study, it was found that disappearance/suppression of genes in amphidiploids from interspecific hybrid F1 was mainly associated with the disease resistance, primary and secondary metabolism and cell cycle. Suppression of many transcription factors like basic helix-loop-helix transcription factor", "MYB/ MYB-related", "MADS-box", "APETALA", "AP5/EREBP", "WRKY" and cytochrome P450 might have helped interspecific hybrid to overcome the reduced fertility in amphidiploid. Silencing of methyl-CpG-binding domain-containing proteins in amphidiploid plants indicated that the formation of amphidiploid is associated with the epigenetic changes. This suggests that the hybridization and allopolyploidy causes rapid changes in gene structure and expression which contributes to the novel type of expression profiles. Further investigation of

unknown/uncharacterized genes which got suppressed/ lost expression in amphidiploid could provide a better understanding of gene affected and the mechanism involved in the ploidy regulation of gene expression upon chromosome doubling.

### **Enhanced chlorophyll biosynthesis in interspecific hybrid F1**

Increase in chlorophyll content serves as an indicator of hybrid vigor as it is believed that increase in chromosome number tends to increase the number of chloroplast in cells and hence increases the chlorophyll content [33, 34, 35]. However, this tendency is not always anticipated as in *Atriplex confertifolia*, which remains constant in plants of different ploidy levels [36]. In the present investigation, increased chlorophyll content (chl<sub>a</sub>, chl<sub>b</sub> and total chlorophyll) of hybrid F1 compared to amphidiploid correlates with respective FPKM and the differential gene expression values. The increased level of chlorophyll in F1 also correlates with the increased expression of chlorophyll biosynthetic genes and decreased expression of the degradation genes in contrast to amphidiploid. Relative chlorophyll content was found to be similar in haploid, diploid and tetraploid plants of *Ricinus communis* [37]. Sometimes increase in chlorophyll content in interspecific hybrid F1 could be related to sterility [38, 39]. F1 hybrids of sorghum raised from the CMS lines having P614 genome had increased chlorophyll a content because of the sterile M35-1A cytoplasm. There was also increase in the total chlorophyll content in F1 hybrids obtained with CMS lines with the Zh10 genome and the P35 pollen parent which was because of A4 cytoplasm. Ectopic over expression of bol-miR171b in *Brassica oleracea* L var. *ital* also led to increase in chlorophyll content which was also sterile [40]. These findings suggest a relationship between chlorophyll biosynthesis and the regulation of sterility in plant as in the present investigation. However, the literature support, does not provide any concrete foundation of this suggestion which require further validation of chlorophyll metabolism genes (*HemA, HemL, HemC, HemE, HemF, HemY, HemH, COX10 Chl, ChII, ChIM, ChIE, 4VCR, HCAR, PPH, PAO, RCCR*) and transcription factors involved in chlorophyll metabolism such as NAC transcription factor *ANAC046, ANAC087, and ANAC100* in plant fertility, floral timing, floral development and seed development.

### **Reduced lignin biosynthesis in interspecific hybrid F1**

Biosynthesis of lignin plays a major a role in the developmental changes in interspecific hybrid F1 and amphidiploid as it is an integral component of plant cell wall providing strength and rigidity to the cell wall. Lignin also provide armory to the plant against various biotic and abiotic stresses [41]. Highest lignin content and the differential gene expression of the genes involved lignin biosynthesis in the amphidiploid suggests the thicker stem, leaf and its ability to withstand towards various abiotic and biotic stresses is due to up-regulation of *HCT, CCR, COMT* and *CAD* genes of the lignin biosynthesis pathway. Hence, contribution of a robust lignin biosynthesis mechanism in the processes of heterosis and enhanced adaptability in amphidiploid cannot be ignored. Lignin biosynthesis is shared by general phenylpropanoid pathway [42] which requires deamination of phenylalanine, successive hydroxylation and O-methylation of aromatic ring, followed by the conversion of the side-chain carboxyl to an alcohol group [43]. The development of plant is severely affected by disruption of *CCR* and *CAD* gene [44] and ccc

mutants of *Arabidopsis* showed male sterility due to lack of lignification in anther endothecium causing the failure of anther dehiscence and pollen discharge. In contrast, down-regulation of *COMT* to low activity levels reduces 30% lignin content in alfalfa and maize while 17% in poplar. The lignin content *vis a vis* the genes of lignin biosynthesis could be correlated to the F1 and the amphidiploid in this investigation as the expression of biosynthetic genes are depressed in the F1 compared to amphidiploid.

### **Higher amount of eugenol and methyleugenol in interspecific hybrid F1 is possibly associated with reduced lignin and flavonoid biosynthesis**

Higher amount of phenylpropanes (eugenol and methyleugenol) in interspecific hybrid F1 (and not in amphidiploid) is possibly due the down regulation of the genes involved in the production of lignin and flavonoids, diverting some flux towards the production of higher amount of phenylpropanes in interspecific hybrid F1. Earlier studies on the modification of lignin and flavonoid biosynthetic pathway in plants suggest that the down-regulation of lignin pathway alters the carbon flux within the phenylpropanoid pathway and indirectly influencing production of other secondary metabolites [45]. For example, in *Petunia hybrida*, suppression of cinnamoyl-CoA reductase (*CCR1*) and up regulation of expression of cinnamate-4-hydroxylase (*C4H*) increased the fluxes through the phenylpropanoid pathway [13]. Down-regulation of chalcone synthase (*CHS*) gene in Flax also showed decreased lignin synthesis and significant plant morphology, modulating the flux towards tannins [46]. Despite the suppressed differential gene expression of *CCR*, *COMT*, *CAD*, *CHS*, and *DFR* genes was positively correlated with reduced content of lignin and flavonoids in interspecific hybrid F1.

### **Effect of hybridization on stomatal patterning and development associated genes**

Stomatal density, guard cell length and stomatal plastid number have frequently been used as morphological markers to test ploidy levels in many plants [47]. In *Coffea canephora*, significant differences in stomatal frequency, guard cell length were noticed between diploid to tetraploid [47]. In this work, it was found that stomatal density in interspecific hybrid F1 ( $2n=62$ ) was nearly two fold higher than its amphidiploid ( $2n=124$ ). In contrast to this, the length of stomata of amphidiploid was found to be nearly threefold greater than its interspecific hybrid F1. Therefore, to understand the genetic basis underlying such variation in stomatal frequency and stomatal length upon change in ploidy levels, several genes putatively involved in stomatal patterning and development were analyzed. In *Arabidopsis*, a number of components in the series of stomatal patterning and development have been identified which include, putative receptors *TOO MANY MOUTHS (TMM)* gene, Erecta-gene family, *CLAVATA*, stomatal density and distribution 1 (*sdd1*) and several *EPIDERMAL PATTERNING FACTORS (EPFs)* [48, 49]. Besides, *SDD1*-like protease (which shares high level of identity with *SDD1*) is concerned with the epidermal development. Also, downstream MAP kinase signaling cascade negatively regulates stomatal development. Similarly, defect in *YODA* gene (*YDA*), a putative MAP kinase kinase kinase (*MAPKKK*), results in an excessive number of clustered stomata [50, 51].

In this study, transcripts associated with the stomatal patterning and development such as *EPIDERMAL PATTERNING FACTOR*-like protein 9 (*EPF9*), subtilisin-like protease *SDD1*, leucine-rich repeat receptor-like

protein *CLAVATA2*, mitogen-activated protein kinase 4/mitogen-activated protein kinase 4-like (*MAPK4/MAPK4*-like) were upregulated indicating the involvement of these genes for the higher number of stomata in interspecific hybrid F1. In contrast, transcripts homologous to *TOO MANY MOUTHS (TMM)* gene, Erecta-gene family, *YODA* gene *etc* were down regulated in interspecific hybrids F1, implying the involvement of these transcripts in determining larger size and less stomata number in amphidiploid.

### Differentially expressed Transcription factors

Transcription factors play a critical role in the regulation of gene expression of all the vital processes in all living organism. They are involved in the regulation of variety of processes that ranges from development to differentiation, metabolism to defense [52]. To integrate transcriptome data with the phenotypic and metabolic profiles, transcription factors involved in flower, trichome, seed and leaf development were identified and analyzed. Some transcription factors involved in chlorophyll metabolism, phenylpropanoid biosynthesis and xylem development were also recovered. Earlier studies in model plants such as *Arabidopsis* suggest the involvement of *MYB*, *MAD*, *NAC*, *WRKY*, *bHLH* and *bZIP* transcription factors in plant growth and development processes [53, 54]. In *Arabidopsis*, majority of TFs belonging to MADS box family were specifically involved in floral developmental processes [55]. Similarly, *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL)* gene family and *CONSTANS* genes also promotes flowering in *Arabidopsis* [56]. *MYB* transcription factors are regarded as a master regulator of phenylpropanoid pathway which is also implicated in the formation of trichome in *Arabidopsis* [57]. Several other *MYB* TFs (*MYB20*, *MYB42*, *MYB43*, *MYB46*, *MYB52*, *MYB54*, *MYB58*, *MYB69*, *MYB61*, *MYB63*, *MYB83*, *MYB85*, and *MYB103*) as well as some *GATA* like TFs are considered to be important regulatory factors for the formation of secondary wall in *Arabidopsis* [58]. Some *NAC* TFs including *ANAC046*, *ANAC087*, and *ANAC100*, directly bind to the promoter regions of *NYC1*, *SGR1*, *SGR2*, and *PaO* suggesting the existence of TFs coordinating the expression of a number of *Chl* catabolic genes [59].

Here, it was found that many transcription factors responsible for floral development (*MADS-box* transcription factor 7-like isoform X3, *MADS-box* protein SOC1-like, *FAR1-RELATED SEQUENCE 8* isoform X1, *CONSTANS-LIKE 13*, transcription factor *VOZ1*, *FAR1-RELATED SEQUENCE 6*-like, *GAMYB*-binding protein, partial, agamous-like *MADS-box* protein AGL9 homolog, agamous-like *MADS-box* protein AGL8 homolog), trichome development (*WRKY* transcription factor 44-like), xylem formation (*NAC* domain-containing protein 7, *MYB48*, *GATA* transcription factor 12-like, PREDICTED: transcription factor *MYB86*-like), seed development (*bZIP* transcription factor family protein 11, *NAC* transcription factor 25,) and cell expansion (*GRF1*-interacting factor 2-like, *GRF1*-interacting factor 3, *BEL 1*-like homeodomain protein 4) in leaf were negatively regulated in interspecific hybrid F1 and positively regulated in amphidiploid proves the potential role of TFs in phenotypic variability in interspecific hybrid F1 and amphidiploid. *MYB4* and *MYB3* like TFs which act as repressor of phenylpropanoid biosynthesis were also negatively regulated in interspecific hybrids suggests the higher amount of phenylpropenes (methyl eugenol and eugenol) in interspecific hybrid F1. However, down regulation of *NAC* transcription factors, positive regulator of flavonoid biosynthesis and chlorophyll catabolism in interspecific hybrid F1 is indicative of the lower amount total flavonoids and higher amount of chlorophyll in interspecific hybrid F1. Although the above

discussion suggests the involvement of TFs in the phenotypic variability and metabolic variation in these two hybrids, there in depth functional characterization could bring out clear understanding of regulation of these TFs in developmental alteration and metabolic variation in interspecific hybrid F1 and amphidiploid.

## Conclusions

This investigation reveals the significance of critical changes in gene expressions associated with biological functions and secondary metabolite biosynthesis in the progenitors and interspecific hybrids of Basil. In addition, the study provides a systematic understanding of the role and regulation of different metabolic pathways responsible in the intergenomic stabilization of interspecific hybrid upon chromosome doubling and underpins the significance of associated changes in gene expression.

## Methods

### Plant Materials

The two well identified clones of *Ocimum* species, namely (i) *Ocimum basilicum* L. OBP1 *i.e.* parent 1, (ii) *Ocimum kilimandscharicum* Gürke – OKP2 *i.e.* parent 2, were used as the starting source material. These clones represent the local landraces and are maintained at the experimental farm of the CSIR-Central Institute of Medicinal and Aromatic Plants, Lucknow 226015, India.

### Development of F1 hybrid and the amphidiploid

An inter-specific hybrid was produced between the two target species by hand pollination taking OBP1 as the female parent. F1 Hybrid was raised from the seeds obtained from the fertilized ovules borne on the female parent. This hybrid thus obtained was seed sterile. Therefore, it was multiplied vegetatively to raise its clonal progenies. Shoot tips of the fast growing hybrid at 10-12 leaf stage were administered with 0.2% aqueous solution of Colchicine (Sigma Aldrich) following cotton swab and intermittent colchicine dropping method for 24 hours. Further, colchicine treatment was stopped by eliminating the cotton swab, and the colchicine treated shoots were washed carefully with the help of water sprayer; and shoots were permitted to grow naturally. The colchicine affected shoots were excised after growth of 10-12 leaf whorls, and multiplied vegetatively to raise the amphidiploids. The amphidiploids were seed fertile.

### Chromosome count

To ascertain chromosome status of progenitor parents, their hybrid and the amphidiploid, somatic chromosome analysis procedure was performed on the four sets of plants. For this the shoot cuttings were planted in sand, and the fast growing roots emerging from the nodes were excised, then pretreated in saturated aqueous solution of *para*-dichlorobenzene for three hours at 12-14<sup>0</sup>C, followed by thorough washing in water and then fixation overnight in Carnoy's solution (6:3:1, Absolute alcohol: Chloroform:

Acetic Acid). Root-tips were stained overnight at 37<sup>0</sup>C in 2% Aceto-Orcein: 1N HCl. Fixed root-tips were squashed in 45% acetic acid to observe chromosome count under the microscope. Only the intact cells were considered to count the chromosome number, while modal number was taken into account to ascertain the somatic chromosome number.

### **Phenotypic changes and its statistical analysis**

Parents (OBP1 and OKP2), interspecific hybrid F1 and amphidiploid (Amphid2) were cloned from cutting and planted in the BT-2 field of CSIR-CIMAP, Lucknow. Field experiments were carried out for the trait measurement and phenotype assessments. Plant height, stem diameter, leaf area and length of the inflorescence, trichome density, length of trichome, length of stomata were calculated from the leaves of the mature plants (six month old) and analyzed statistically. The digital pictures of each mature plant, leaves and inflorescence (OBP1, OKP2, F1 and Amphid2) depicting the phenotypic changes were taken. Glandular trichome density was observed at 40X magnification using a compound microscope (Leica DM750). Ten biological replicates were taken for the trait measurement and phenotype assessments.

### **Scanning electron Microscopy**

Mature leaves (six month old) of all four plants (OBP1, OKP2, F1 and Amphid2) were cut into small pieces and fixed in 100% methanol for 10 min. Fixed leaves were dehydrated through acetone series (30, 50, 70, 80, 90, 100%), each for 10 min and then dried using critical point dryer (K850 from Quorum Technology, UK). After this, the dried leaves were mounted on the scanning electron microscopy stubs for sputter-coating with gold using Q150TES from Quorum Technology, UK and observed under scanning electron microscope (Quanta 250, FEI, Thermo Fisher Scientific). Observation of trichomes and stomata under scanning electron microscope was performed in ten different replicates.

### **Total RNA isolation and library preparation**

Total RNA from the apical leaves of mature plants (six month old) of *O. basilicum* (OBP1) and *O. kilimandscharicum* (OKP2), interspecific hybrid F1 and amphidiploid were isolated using Spectrum Plant Total RNA Kit (Sigma Aldrich) according to manufacturer's protocol. The quantitative and qualitative analysis of the isolated RNA was performed using Nanodrop and 1% agarose gel electrophoresis. Equal amount of RNA isolated from parent plants, F1 and amphidiploid were used to construct and sequence RNAseq libraries. Illumina HiSeq 2000 platform was used for sequencing RNAseq library generating 144, 127, 139 and 135 Mbp of 101bp paired end reads for OBP1, OKP2, F1 and Amphidiploid plants, respectively. All the contaminants and low quality reads were removed for obtaining processed data. For the evaluation of the quality of raw reads and filtering high quality reads, a NGSQC Toolkit was used.

### **De novo assembly and functional annotation**

*De novo* transcriptome assembly of high quality reads was carried out by Trinity Assembler (Trinityrnaseq-2.0.6) using 200bp minimum contig length and 2bp minimum count for K-mers as per

Inchworm algorithm for the assembly of primary transcriptome. CD-HIT-EST (v4.6.1): a fast program for clustering and comparing large sets of protein or nucleotide sequences and TransImprove-2.0.1 (Bionivid) tools were used for the final transcriptome assembly. Validated and ameliorated reads with average depth  $\geq 5$  and coverage  $\geq 70\%$  were used as an input for CD-Hit-EST. Basic local alignment search tool program (Blast X) (E value  $\leq 0.001$ ) was used against NCBI non-redundant (NR) protein database for the annotation of 91,778 total transcripts obtained from the final transcriptome assembly. Total number of transcripts retrieved were 67,770 (OBP1), 73,265 (OKP2), 76,917 (interspecific hybrid F1) and 76,563 (amphidiploids). Out of 67,770 transcripts retrieved from OBP1, 52,474 were annotated and 15,296 remained un-annotated. Similarly, out of 73,265 retrieved in OKP2, 56,879 got annotated and 16,386 were un-annotated. Among the total 76,917 transcripts of interspecific hybrid F1, 60,726 were annotated while 16,191 un-annotated. On the other side, from total 76,563 transcripts of amphidiploid, 60,954 were annotated and 15,609 remained un-annotated (**Table S10**).

### Differential Gene Expression (DGE) Analysis

Determination of differentially expressed genes (DEGs) between progenies and their progenitors was performed using DESeq R Package (60). The numbers of DEGs were compared with the total number of expressed genes obtained from the DGE analysis of 91,778 clustered transcripts. FPKM (Fragments per Kilobase of transcript per Million) values were calculated to measure the expression level of each assembled transcript sequence. For the estimation of genotype specific transcriptional differences following pair-wise contrasts was carried out: OBP1 Vs. OKP2, OBP1 Vs. F1, OBP1 Vs. Amphid2, OBP1 Vs. F1&Amphid2, OKP2 Vs. F1, OKP2 Vs. Amphid2, OKP2 Vs. F1&Amphid2, F1 Vs. Amphid2, F1 Vs. OBP1&OKP2 and Amphid2 Vs. OBP1,OKP2&F1 (**Table S11-S20**). Up regulated and down regulated genes with fold change 2 and above with FDR adjusted p-value  $\leq 0.05$  were considered to be statistically significant and others were considered to be baseline expressed genes. However, significant up and down regulated genes with  $\log_2$  fold change  $-1 \geq -1$  was considered for the analysis of specific transcriptional differences in different pathways.

### Quantitative RT-PCR for gene expression analysis

To validate the results of RNA seq data, qRT-PCR was conducted as per the protocol described by Rastogi et al. [61]. 2 $\mu$ g of total RNA isolated (Spectrum Plant Total RNA Kit, Sigma Aldrich) from all the genotype was used for the synthesis of cDNA (cDNA synthesis Kit, ThermoScientific, USA). These cDNAs were taken as template for the quantification of relative mRNA levels following the SYBR green chemistry (Maxima SYBR Green 2 x PCR Master Mix, Thermo Scientific, Waltham MA,US), performed on Fast Real time PCR system (7900HT applied Biosystems,USA). The gene sequences were selected from the RNAseq data of all 4 *Ocimum* genotypes used in this study. The PCR reactions for each genotype of *Ocimum* were run in triplicate and mean value of each sample was statistically analyzed. Actin was used as an endogenous control. Primers for qRT-PCR were designed through Primer express software (Agilent technologies) (**Table S21**)

## Gene ontology enrichment analysis

Differentially expressed genes were analyzed using KEGG Orthology-Based Annotation System KOBAS for the determination of significantly enriched pathways. The statistically enriched DEGs were considered significant with false discovery rate (FDR) <0.05. Further, AgriGO (<http://bioinfo.cau.edu.cn/>) was used to summarize and visualize Gene Ontology terms using *Arabidopsis thaliana* as a background species. The color of boxes indicates corrected P-value < 0.05.

## Biochemical Assays for estimation of total chlorophyll, total lignin and total flavonoid contents

Total chlorophyll content of both interspecific hybrid F1 and amphidiploid were calculated as described by Arnon [62]. The absorbance of the pigment was estimated at 645 and 663nm wavelength against 80% acetone. Three biological replicates were taken for the analysis and absorbance of each replicate was repeated thrice.

Total lignin content from the mature leaves of three biological replicate samples of interspecific hybrid F1 and amphidiploid were estimated using the method as described by Kumar et al. [63].

Total flavonoid content was quantified from the dried sample of mature leaves of interspecific hybrid F1 and amphidiploid using Dowd method as illustrated by Sankhalkar and Vernekar. [64]

## Metabolite Analysis

For relative oil analysis, 100g of fresh leaves of interspecific hybrid F1 and amphidiploid were hydro-distilled in the Clevenger's apparatus. The oil samples of both plants were collected into the micro-centrifuge tubes (MCTs) separately. 1µl of dehydrated (using anhydrous sodium sulphate) oil diluted with hexane in the ration of 1:10 was injected to GC-MS system (MSD 7890A, Agilent Technologies) equipped with autosampler, HP-5MS column and 7977A mass detector. The oil samples were run in splitless mode as described by Akhtar et al. [65]. All the samples were run in three biological replicates and analyzed statistically ( $\pm$  standard deviation). Mass spectra acquisition was carried out in scan mode and analyzed with the help of Mass hunter workstation software (Agilent technologies) by comparing with the NIST11 library.

## Statistical measurement

Microsoft Office Excel 2007 was used for the calculation of mean values and standard error (SE). Graphpad prism software was used for the calculation of significant differences between the samples. \*\*\* represents  $p < 0.001$ , \*\* represents  $p < 0.01$  and \* represents  $p < 0.05$ .

## Abbreviations

OBP1: *Ocimum basilicum* parent 1

OKP2: *Ocimumkilimandscharicum* parent 2

F1: Interspecific hybrid first filial

Amphid2: Amphidiploid

FPKM: Fragments per kilobase of exon model per million reads mapped

DEGs: Differentially expressed genes

KEGG: Kyoto Encyclopedia of Genes and Genomes

DGE: Differential gene expression

GO: Gene Ontology

Chl: Chlorophyll

NIST: National Institute of Standards and Technology

## **Declarations**

### **Availability of data and materials**

The NGS data of the present investigation is submitted to the NCBI SRA database under the bioproject ID PRJNA520976.

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

The NGS data of the present investigation is submitted to the NCBI SRA database under the bioproject ID PRJNA520976

### **Competing interests**

The authors declare no conflict of interest.

### **Funding**

This work was supported by CSIR (Council of Scientific and Industrial Research) funded projects [BSC 0203 and HCP 0007]. CSIR-UGC for fellowship of SS and DV, and INSA Senior Scientist Scheme of UCL.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Authors' contributions

SS performed the experiments and drafted the manuscript, SR, DV and M helped in data analysis and manuscript review, RKL analyzed the breeding and plant traits, UCL performed the cytological studies, AKS planned the present investigation and finalized the manuscript. All authors have read and approved the final manuscript.

### Acknowledgements

We acknowledge the efforts of Dr. PK Rout for lignin analysis and Dr. PV Ajaykumar for electron microscopy experiments. The authors express their heartfelt thanks to the Director, CSIR-CIMAP for his great interest and providing facilities for the experiments. We thank Bionivid Pvt. Ltd., Bengaluru, India for the transcriptome sequencing.

### Additional information

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

## References

1. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE. Ancestral polyploidy in seed plants and angiosperms. *Nature*, 2011; 473(7345):97-100.
2. Wendel JF, Lisch D, Hu G, Mason AS. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr Opin Genet Dev*. 2018; 49:1-7.
3. Scarpino SV, Levin DA, Meyers LA. Polyploid formation shapes flowering plant diversity. *Nat*. 2014; 184(4):456-65.
4. Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. *New Phytol*. 2016; 210(2):391-8.
5. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. *Rev. Genet*. 2017; 18(7):411..
6. Lavania, U.C. Plant speciation and polyploidy: in habitat divergence and environmental perspective. *Nucleus*2020; 63(1): 1–5.
7. Fasano C, Diretto G, Aversano R, D'Agostino N, Di Matteo A, Frusciante L, Giuliano G, Carputo D. Transcriptome and metabolome of synthetic *Solanum* autotetraploids reveal key genomic stress

- events following polyploidization. *New Phytol.* 2016; 210(4):1382-94.
8. Sun Y, Liu Z, Fei Z, Li C, Chun Y, Zhang A. Synergistic effect and degradation mechanism on Fe-Ni/CNTs for removal of 2, 4-dichlorophenol in aqueous solution. *Sci. Pollut. Res.* 2019; 26(9):8768-78.
  9. Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Opin. PlantBiol.* 2016; 30:159-65.
  10. Levin DA. The role of chromosomal change in plant evolution. Oxford University Press; 2002.
  11. Schoenfelder KP, Fox DT. The expanding implications of polyploidy. *Cell Biol.* 2015; 209(4):485.
  12. Lavania UC. Genomic and ploidy manipulation for enhanced production of phyto-pharmaceuticals. *Plant Genet. Res.* 2005; 3(2):170.
  13. mishra MK. Stomatal characteristics at different ploidy levels in Coffea L. *Bot.* 1997; 80(5):689-92.
  14. Lavania UC, Srivastava S, Lavania S, Basu S, Misra NK, Mukai Y. Autopolyploidy differentially influences body size in plants, but facilitates enhanced accumulation of secondary metabolites, causing increased cytosine methylation. *Plant Journal.* 2012; 71(4):539-49.
  15. Riddle NC, Kato A, Birchler JA. Genetic variation for the response to ploidy change in Zea mays L. *Theor appl genet.* 2006; 114(1):101-11.
  16. Pandit MK, Pocock MJ, Kunin WE. Ploidy influences rarity and invasiveness in plants. *J. Ecol.* 2011; 99(5):1108-15.
  17. Levin DA. The role of chromosomal change in plant evolution. Oxford University Press; 2002 Nov 28.
  18. Ozkan H, Levy AA, Feldman M. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell.* 2001; 13(8):1735-47.
  19. Wendel JF. The wondrous cycles of polyploidy in plants. *J. Bot.* 2015; 102(11):1753-6.
  20. Hegarty MJ, Abbott RJ, Hiscock SJ. Allopolyploid speciation in action: The origins and evolution of *Senecio cambrensis*. In Polyploidy and genome evolution 2012 (pp. 245-270). Springer, Berlin, Heidelberg.
  21. Soltis DE, Buggs RJ, Barbazuk WB, Chamala S, Chester M, Gallagher JP, Schnable PS, Soltis PS. The early stages of polyploidy: rapid and repeated evolution in *Tragopogon*. In Polyploidy and genome evolution 2012 (pp. 271-292). Springer, Berlin, Heidelberg.
  22. Soltis DE, Visger CJ, Marchant DB, Soltis PS. Polyploidy: pitfalls and paths to a paradigm. *J. Bot.* 2016; 103(7):1146-66.
  23. Shasany AK, Kole C. Genomic Resources of *Ocimum*. In The *Ocimum* Genome 2018 (pp. 99-110). Springer, Cham.
  24. Viña A, Murillo E. Essential oil composition from twelve varieties of basil (*Ocimum* spp) grown in Colombia. *J Brazil Chem Soc.* 2003; 14(5):744-9.
  25. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 2016: gkw982.
  26. Doyle JJ, Egan AN. Dating the origins of polyploidy events. *New Phytol.* 2010; 186(1):73-85.

27. Qiao G, Li H, Liu M, Jiang J, Yin Y, Zhang L, Zhuo R. Callus induction and plant regeneration from anthers of *Dendrocalamus latiflorus* *In Vitro Cell. Dev. Biol.* 2013; 49(4):375-82.
28. Bidère N, Snow AL, Sakai K, Zheng L, Lenardo MJ. Caspase-8 regulation by direct interaction with TRAF6 in T cell receptor-induced NF- $\kappa$ B activation. *Bio.* 2006; 16(16):1666-71.
29. Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell.* 2001; 13(8):1749-59.
30. Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. Phenotypic instability and rapid gene silencing in newly formed. *Plant Cell.* 2000; 12: 1551–1567.
31. Kashkush K, Feldman M, Levy AA. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*, 2002; 160(4):1651-9.
32. Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. Ploidy regulation of gene expression. *Science*, 1999; 285(5425):251-4.
33. Meyers SP, Nichols SL, Baer GR, Molin WT, Schrader LE. Ploidy effects in isogenic populations of alfalfa: I. Ribulose-1, 5-bisphosphate carboxylase, soluble protein, chlorophyll, and DNA in leaves. *Plant Physiol.* 1982; 70(6):1704-9.
34. Leech RM, Leese BM, Jellings AJ. Variation in cellular ribulose-1, 5-bisphosphate-carboxylase content in leaves of *Triticum* genotypes at three levels of ploidy. *Planta*, 1985; 166(2):259-63.
35. Warner DA, Edwards GE. Effects of polyploidy on photosynthetic rates, photosynthetic enzymes, contents of DNA, chlorophyll, and sizes and numbers of photosynthetic cells in the C4 dicot *Atriplex confertifolia*. *Plant physiol.* 1989; 91(3):1143-51.
36. Timko MP, Vasconcelos AC. Euploidy in *Ricinus*: euploidy effects on photosynthetic activity and content of chlorophyll-proteins. *Plant physiol.* 1981; 67(6):1084-9.
37. Kibalnik OP, Elkonin LA. Effect of the cytoplasmic male sterility type on the chlorophyll content in the leaves of grain *sorghum Russ.* *J. Genet.* 2016; 6(5):520-6.
38. Li H, Zhang Q, Li L, Yuan J, Wang Y, Wu M, Han Z, Liu M, Chen C, Song W, Wang C. Ectopic overexpression of bol-miR171b increases chlorophyll content and results in sterility in broccoli (*Brassica oleracea* L var. italica). *Agri. Food Chem.* 2018; 66(37):9588-97.
39. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin biosynthesis and structure. *Plant physiol.* 2010; 153(3):895-905.
40. Whetten R, Sederoff R. Lignin biosynthesis. *Plant cell.* 1995; 7(7):1001.
41. Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. *Rev. Plant Biol.* 2003; 54, 519-546.
42. Thévenin J, Pollet B, Letarnec B, Saulnier L, Gissot L, Maia-Grondard A, Lapierre C, Jouanin L. The simultaneous repression of CCR and CAD, two enzymes of the lignin biosynthetic pathway, results in sterility and dwarfism in *Arabidopsis thaliana*. *Molecular Plant*, 2011; 4(1):70-82.
43. Campbell MM, Ellis BE. Fungal elicitor-mediated responses in pine cell cultures. *Planta*, 1992; 186(3):409-17.

44. Muhlemann JK, Woodworth BD, Morgan JA, Dudareva N. The monolignol pathway contributes to the biosynthesis of volatile phenylpropenes in flowers. *New Phytol.* 2014; 204(3):661-70.
45. Zuk M, Działo M, Richter D, Dymińska L, Matuła J, Kotecki A, Hanuza J, Szopa J. Chalcone synthase (CHS) gene suppression in flax leads to changes in wall synthesis and sensing genes, cell wall chemistry and stem morphology parameters. *Plant Sci.* 2016; 7:894.
46. Casson S, Gray JE. Influence of environmental factors on stomatal development. *New phytol.* 2008; 178(1):9-23.
47. Bergmann DC, Lukowitz W, Somerville CR. Stomatal development and pattern controlled by a MAPKK kinase. *Science* 2004; 304(5676):1494-7.
48. Wang H, Ngwenyama N, Liu Y, Walker JC, Zhang S. Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in *Arabidopsis*. *Plant Cell.* 2007; 19(1):63-73.
49. Coupe SA, Palmer BG, Lake JA, Overy SA, Oxborough K, Woodward FI, Gray JE, Quick WP. Systemic signalling of environmental cues in *Arabidopsis* leaves. *Exp. Bot.* 2006; 57(2):329-41.
50. Qiao S, Sun S, Wang L, Wu Z, Li C, Li X, Wang T, Leng L, Tian W, Lu T, Wang X. The RLA1/SMOS1 transcription factor functions with *OsBZR1* to regulate brassinosteroid signaling and rice architecture. *Plant Cell.* 2017; 29(2):292-309.
51. Ramachandran S, Hiratsuka K, Chua NH. Transcription factors in plant growth and development. *Curr opin genet dev.* 1994; 4(5):642-6.
52. Yanagisawa S. Transcription factors in plants: physiological functions and regulation of expression. *Plant Research.* 1998; 111(3):363-71.
53. Rounsley SD, Ditta GS, Yanofsky MF. Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell.* 1995; 7(8):1259-69.
54. Preston JC, Hileman L. Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Plant Sci.* 2013; 4:80.
55. Putterill J, Robson F, Lee K, Simon R, Coupland G. The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell.* 1995; 80(6):847-57.
56. Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *plant.* 2015; 8(5):689-708.
57. Xun H, Zhang Z, Zhou Y, Qian X, Dong Y, Feng X, Pang J, Wang S, Liu B. Identification and functional characterization of R3 MYB transcription factor genes in soybean. *Plant Biol.* 2018; 61(2):85-96.
58. Cassan-Wang H, Goué N, SAIDI MN, Legay S, Sivadon P, Goffner D, Grima-Pettenati J. Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. *Plant Sci.* 2013; 4:189.
59. Oda-Yamamizo C, Mitsuda N, Sakamoto S, Ogawa D, Ohme-Takagi M, Ohmiya A. The NAC transcription factor ANAC046 is a positive regulator of chlorophyll degradation and senescence in

- Arabidopsis leaves. *Rep.* 2016; 6(1):1-3.
60. Anders S, Huber W. Differential expression of RNA-Seq data at the gene level–the DESeq package. Heidelberg, Germany: European Molecular Biology Laboratory (EMBL). 2012; 10:f1000research.
61. Rastogi S, Kumar R, Chanotiya CS, Shanker K, Gupta MM, Nagegowda DA, Shasany AK. 4-Coumarate: CoA ligase partitions metabolites for eugenol biosynthesis. *Plant cell physiol.* 2013; 54(8):1238-52.
62. Arnon DI. Copper enzymes in isolated chloroplast. Polyphenoloxidase in *Beta vulgaris*. *Plant physiol.* 1949; 24:1-5.
63. Kumar R, Vashisth D, Misra A, Akhtar MQ, Jalil SU, Shanker K, Gupta MM, Rout PK, Gupta AK, Shasany AK. RNAi down-regulation of cinnamate-4-hydroxylase increases artemisinin biosynthesis in *Artemisia annua*. *Rep.* 2016; 6:26458.
64. Sankhalkar S, Vernekar V. Quantitative and Qualitative analysis of Phenolic and Flavonoid content in *Moringa oleifera* Lam and *Ocimum tenuiflorum* *Pharmacogn Res.* 2016; 8(1):16.
65. Akhtar MQ, Qamar N, Yadav P, Kulkarni P, Kumar A, Shasany AK. Comparative glandular trichome transcriptome-based gene characterization reveals reasons for differential

## Tables

**Table 1** Phenotypic comparison of parents (OBP1 and OKP2), interspecific hybrid F1 and Amphidiploids

Trait	OBP1	OKP2	F1	AMPHID2
Plant height (cm)	87.40 ± 7.80	112.70 ± 6.24	82.30 ± 8.90	136.50 ± 9.30
Leaf area (cm <sup>2</sup> )	7.61 ± 0.61	6.85 ± 0.45	3.50 ± 0.18	9.63 ± 0.75
Stem diameter (cm)	3.40 ± 0.60	3.47 ± 0.50	2.92 ± 0.40	6.02 ± 0.70
Length of inflorescence (cm)	12.06 ± 1.70	13.30 ± 1.20	24.30 ± 1.90	17.60 ± 1.90
Trichome density	42.20 ± 2.87	29.80 ± 1.18	62.30 ± 1.87	26.20 ± 0.54
Trichome length (µm)	69.85 ± 2.75	72.54 ± 1.87	70.92 ± 1.32	88.30 ± 2.71
Stomata length (µm)	12.44 ± 0.65	13.00 ± 1.08	7.90 ± 0.61	21.20 ± 1.08
Oil yield (gm/100g leaves)	0.43 ± 0.03	0.30 ± 0.03	0.41 ± 0.04	0.48 ± 0.02

Data are presented as mean ± standard mean error, n ≥ 10

**Table 2** Transcripts related to chlorophyll metabolism in interspecific F1 and amphidiploid

Function	Gene	Enzyme	KO ID (EC No.)	F1	AMPHID2
ALA formation	<i>HemA</i>	Glutamyl-tRNA reductase	K02492 (1.2.1.70)	12.00	12.00
	<i>HemL</i>	Glutamate-1-semialdehyde 2,1-aminomutase	K01845 (5.4.3.8)	1	1
Proto IX formation	<i>HemB</i>	Porphobilinogen synthase	K01698 (4.2.1.24)	4	5
	<i>HemC</i>	Hydroxymethylbilane synthase	K01749 (2.5.1.61)	5	5
	<i>HemD</i>	Uroporphyrinogen-III synthase	K01719 (4.2.1.75)	2	3
	<i>HemE</i>	Uroporphyrinogen decarboxylase	K01599 (4.1.1.37)	10	9
	<i>HemF</i>	Coproporphyrinogen-III oxidase	K02495 (1.3.99.22)	5	5
	<i>HemY</i>	Protoporphyrinogen oxidase	K00231 (1.3.3.4)	10	11
Heme formation	<i>HemH</i>	Ferrochelatase	K01772 (4.99.1.1)	14	13
	<i>COX10</i>	Protoheme IX farnesyltransferase	K02257 (2.5.1.-)	4	4
	<i>COX15</i>	Cytochrome c oxidase assembly protein subunit 15	K02259 (1.9.3.1)	7	8
Chlorophyll formation	<i>ChIH</i>	Magnesium chelatase subunit H	K03403 (6.6.1.1)	1	1
	<i>ChID</i>	Magnesium chelatase subunit D	K03404 (6.6.1.1)	2	2
	<i>ChII</i>	Magnesium chelatase subunit I	K03405 (6.6.1.1)	3	5
	<i>ChIM</i>	Magnesium protoporphyrin IX methyltransferase	K03428 (2.1.1.11)	5	5
	<i>ChIE</i>	Magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase	K04035 (1.14.13.81)	4	4
	<i>4VCR</i>	Divinyl chlorophyllide a 8-vinyl-reductase	K19073 (1.3.1.75)	6	6
	<i>POR</i>	Protochlorophyllide reductase	K00218 (1.3.1.33)	7	7

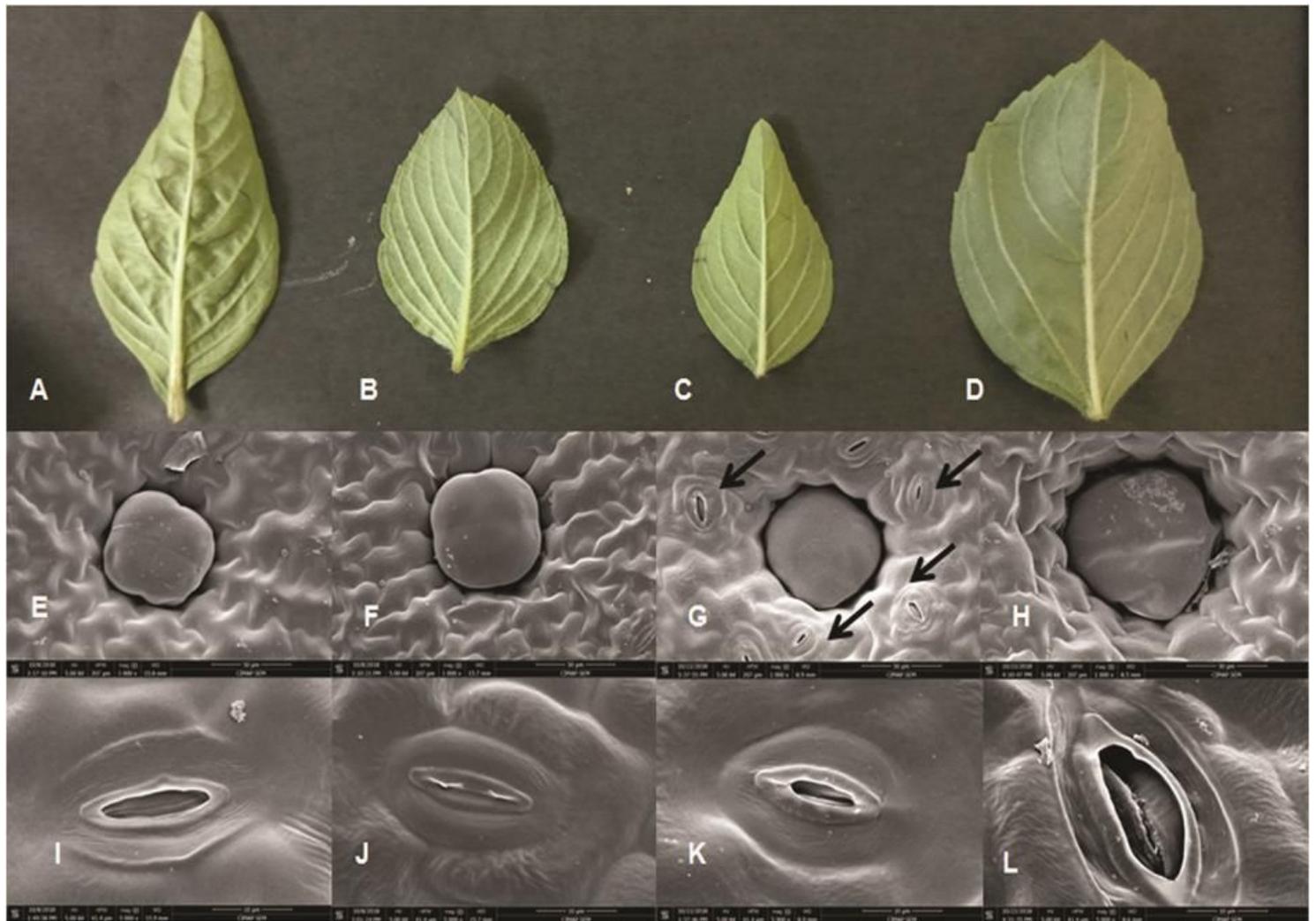
	<i>CAO</i>	Chlorophyllide a oxygenase	K13600 (1.14.13.122)	9	7
	<i>ChlG</i>	Chlorophyll synthase	K04040 (2.5.1.62)	1	1
	<i>CLH</i>	Chlorophyllase	K08099 (3.1.1.14)	9	8
Chlorophyll degradation	<i>NYC1</i>	Chlorophyll (ide) b reductase	K13606 (1.1.1.294)	2	2
	<i>HCAR</i>	7-Hydroxymethyl chlorophyll a reductase	K18010 (1.17.7.2)	5	5
	<i>PPH</i>	Pheophytinase	K08099 (3.1.1.14)	32	33
	<i>PAO</i>	Pheophorbide a oxygenase	K13071 (1.14.15.17)	3	3
	<i>RCCR</i>	Red chlorophyll catabolite reductase	K13545 (1.3.7.12)	4	4

**Table 3** Transcripts related to general phenylpropanoid biosynthetic pathway

Pathway	Gene	Enzyme	KO ID (E.C.no)	F1	AMPHID2
Phenylpropanoid/Lignin biosynthesis	<i>HCT</i>	Hydroxycinnamoyl transferase	K13065 (2.3.1.133)	10	15
	<i>PAL</i>	Phenylalanine ammonia-lyase	K10775 (4.3.1.24)	19	20
	<i>C4H</i>	Cinnamate 4-hydroxylase	K14974 (1.14.13.11)	5	4
	<i>CCR</i>	Cinnamoyl-CoA reductase	K09753 (1.2.1.44)	44	41
	<i>4CL</i>	4-Coumarate:coenzyme A ligase	K01904 (6.2.1.12)	43	41
	<i>COMT</i>	Caffeic acid 3-O-methyltransferase	K13066 (2.1.1.68)	5	5
	<i>CCoAOMT</i>	Caffeoyl CoA O-methyltransferase	K00588 (2.1.1.104)	2	6
	<i>CAD</i>	Cinnamyl alcohol dehydrogenase	K00083 (1.1.1.195)	12	9
	<i>EOMT</i>	Eugenol O-methyltransferase	K17058 (2.1.1.146)	1	1
	<i>EGS</i>	Eugenol synthase 1	K17055 (1.1.1.318)	3	3
Flavonoid Biosynthesis	<i>CHS</i>	Chalcone synthase	K00660 (2.3.1.74)	6	7
	<i>CHI</i>	Chalcone Isomerase	K01859 (5.5.1.6)	6	7
	<i>DFR</i>	Dihydroflavone reductase	1.1.1.219	14	15
	<i>F3'H</i>	Flavonoid 3'-monooxygenase	K05280 (1.14.14.82)	8	7
	<i>F3'5'H</i>	Flavonoid 3'5' hydroxylase	K13083 (1.14.14.81)	1	1
	<i>F3H</i>	flavonol synthase/flavanone 3-hydroxylase	K05278 (1.14.20.6)	1	1
Anthocyanin modification	<i>UF3GT</i>	flavonoid 3-O-glucosyltransferase	K10757 (2.4.1.91)	4	4

<i>UGT75C1</i>	Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase	K12338 (2.4.1.298)	1	1
<i>5MaT1</i>	anthocyanin 5-O-glucoside-6 <sup>'''</sup> -O-malonyltransferase	K12934 (2.3.1.172)	4	3

## Figures



**Figure 1**

Morphological characterization of parents (OBP1 and OKP2), interspecific hybrid F1 and Amphidiploids grown under same condition. (A-D) Leaves of OBP1, OKP2, interspecific hybrid F1 and Amphidiploids. (E-H) Glandular trichome of OBP1, OKP2, interspecific hybrid F1 and Amphidiploids. (I-L) Stomata of OBP1, OKP2, interspecific hybrid F1 and Amphidiploids. Arrows in black colour shows density of stomata at 50 $\mu$ m magnification in interspecific F1.

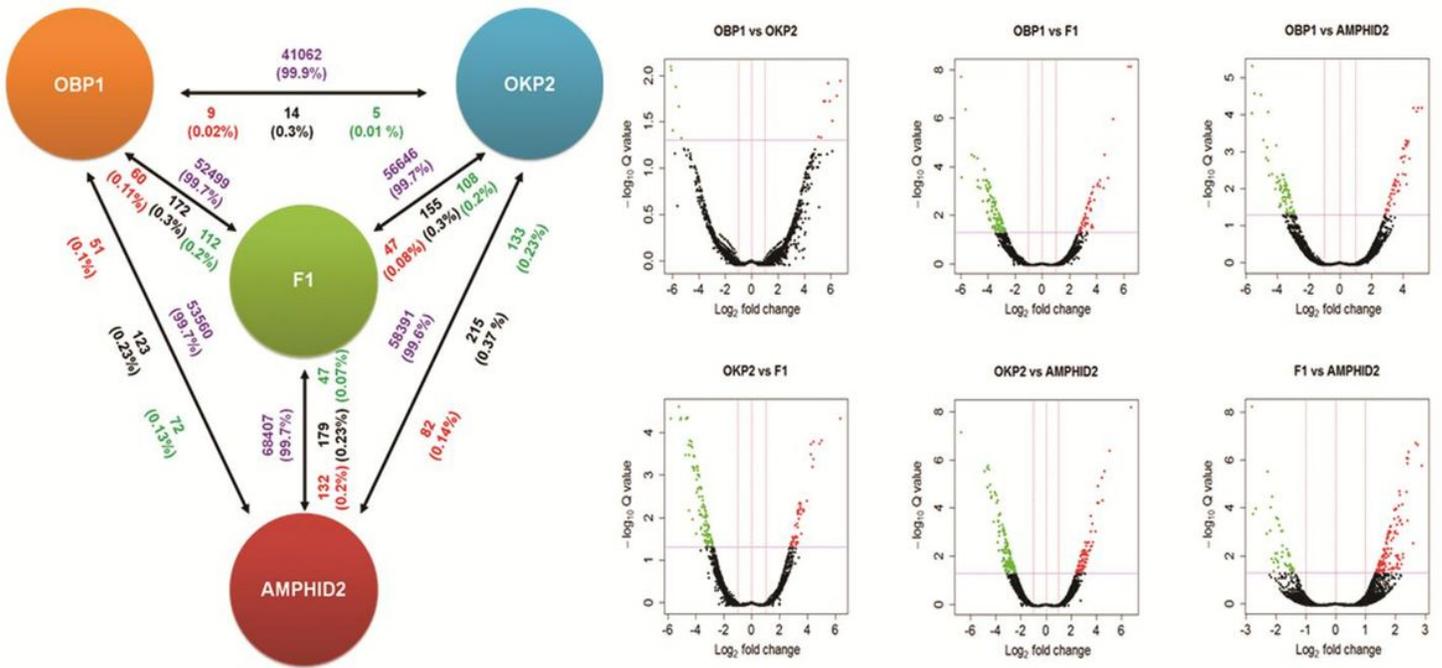
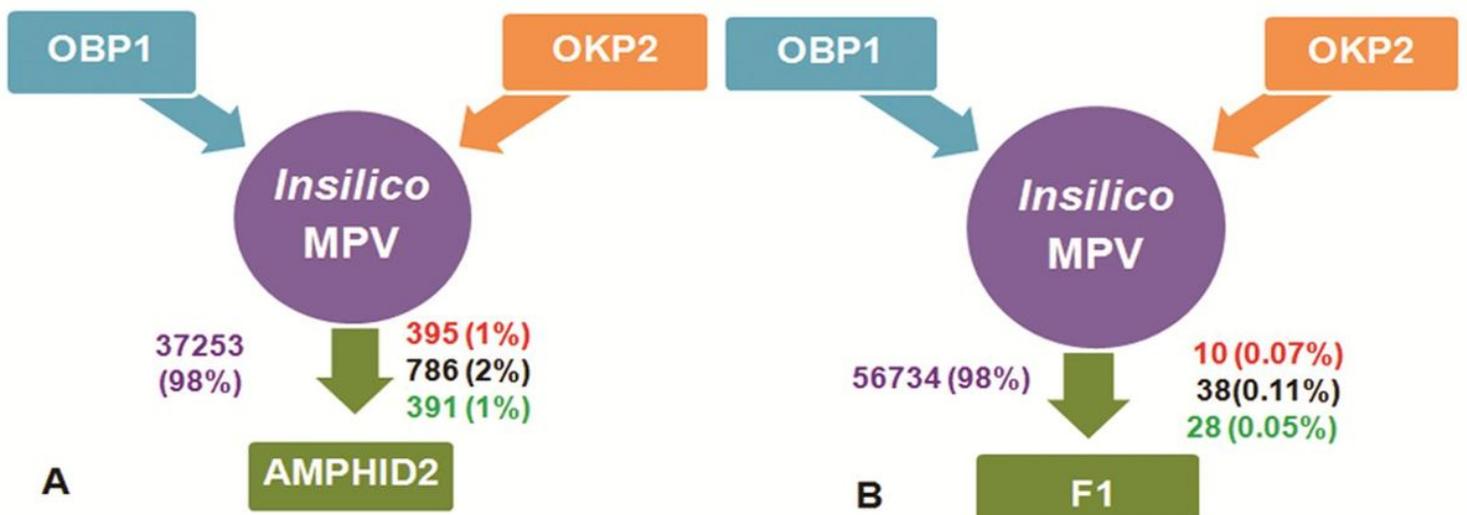


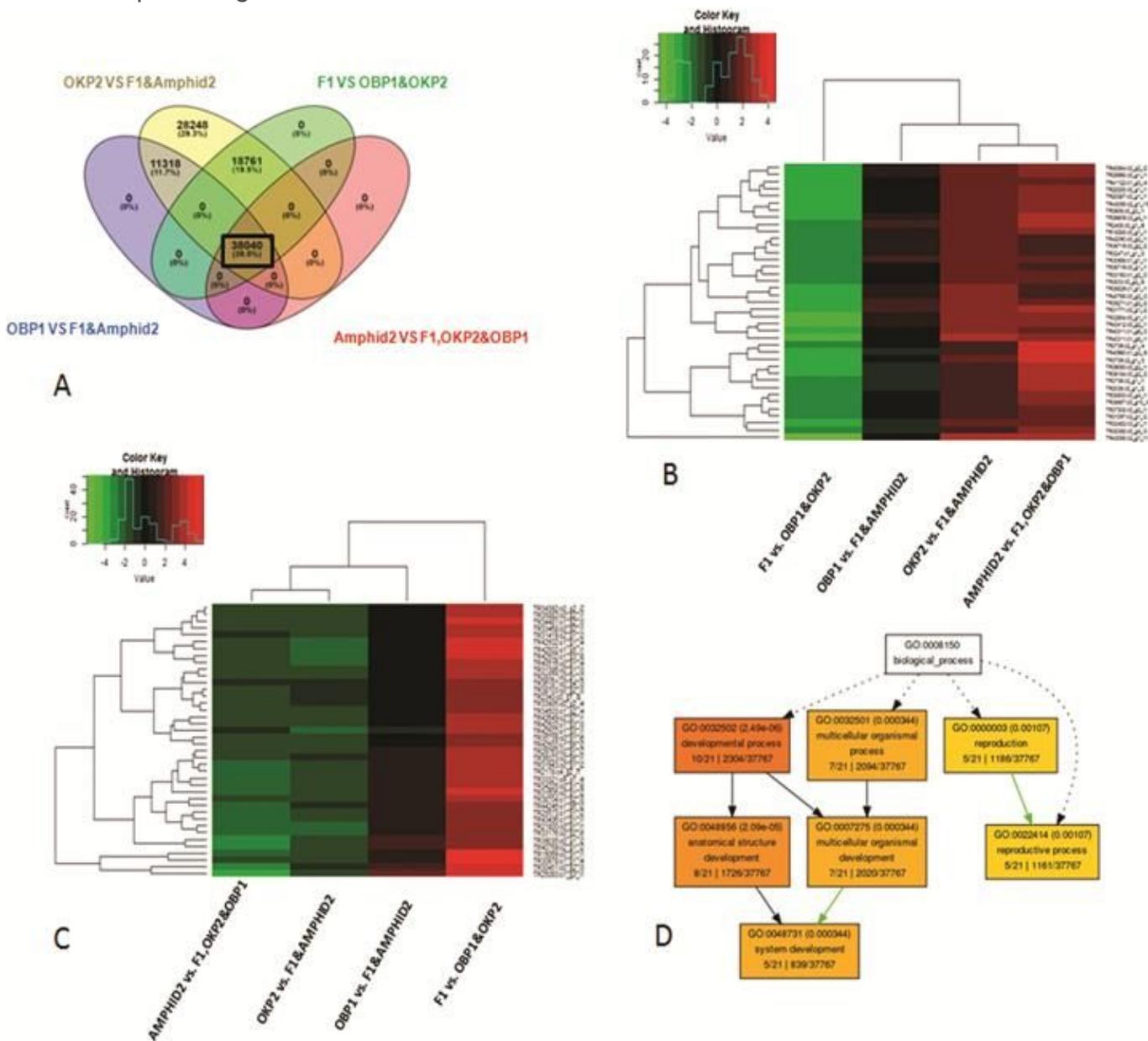
Figure 2

(A) Diagrammatic representation of RNA-seq Differential gene expression of Parent 1 (OBP1), Parent 2 (OKP2), its interspecific hybrid F1 and its colchicine treated Amphidioid. Text at lefthand side indicates the total number and fraction (%) of Up-regulated genes while text written at righthand side indicates the total number and fraction (%) of Down regulated genes. Total number and fraction (%) of differentially expressed genes are represented by the text written in the middle and below of the arrow while the text written at above the arrow highlights the number and fraction (%) of baseline expressed genes. (B) Volcano plots showing the differentially expressed gene between the parents (OBP1 and OKP2) and their progenies (interspecific hybrid F1 and amphidioid). Dots above the horizontal line shows the differentially significant transcripts (FDR< 0.05) and the dots in right and left handside but above the horizontal line represents the highly significant up regulated and down regulated transcripts respectively.



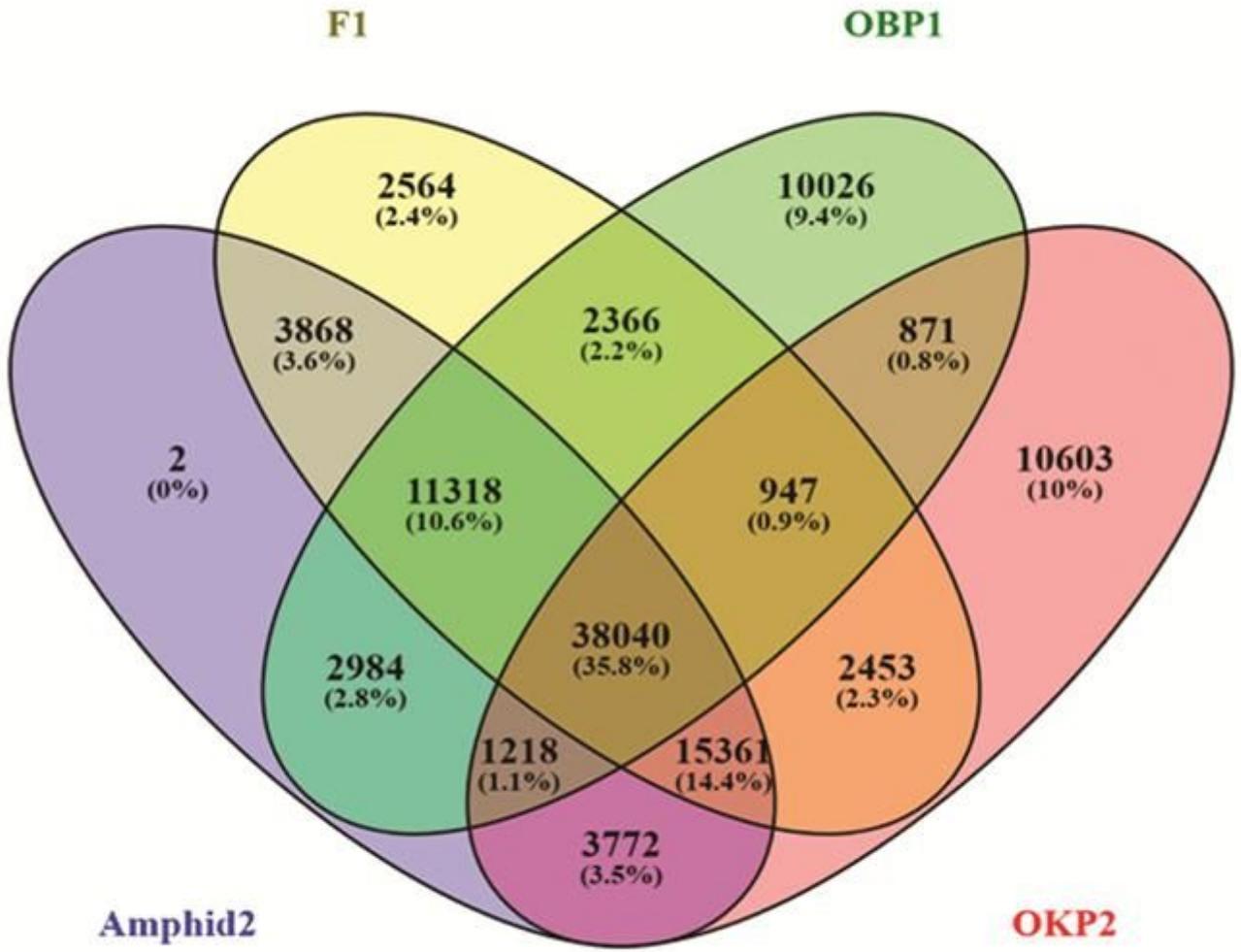
**Figure 3**

(A-B) Non-additively expressed genes in interspecific amphidiploid and F1. Differentially significant transcripts between interspecific hybrid F1 and Amphidiploid and its mid parent value (MPV) (FDR<0.05). The amphidiploids were compared with parent 1 (OBP1), parent 2 (OKP2) and interspecific of F1 assuming that 1/3 of the total transcription is from the genome from parent 1 (OBP1), 1/3 of the total transcription is from parent 2 (OKP2) and 1/3 of the total transcription is from the interspecific hybrid F1 while the interspecific hybrid F1 was compared with the parent 1 (OBP1) and parent 2 (OKP2) assuming that the ½ of the total transcription is from the genome of each parent (OBP1 and OKP2). The numbers written at the right hand side represents significantly up-regulated, total number of expressed genes and down regulated genes respectively (FDR< 0.05). The numbers written at the righthand side represents total number of baseline expressed genes respectively. Percentage indicates the proportion of total number of expressed genes.



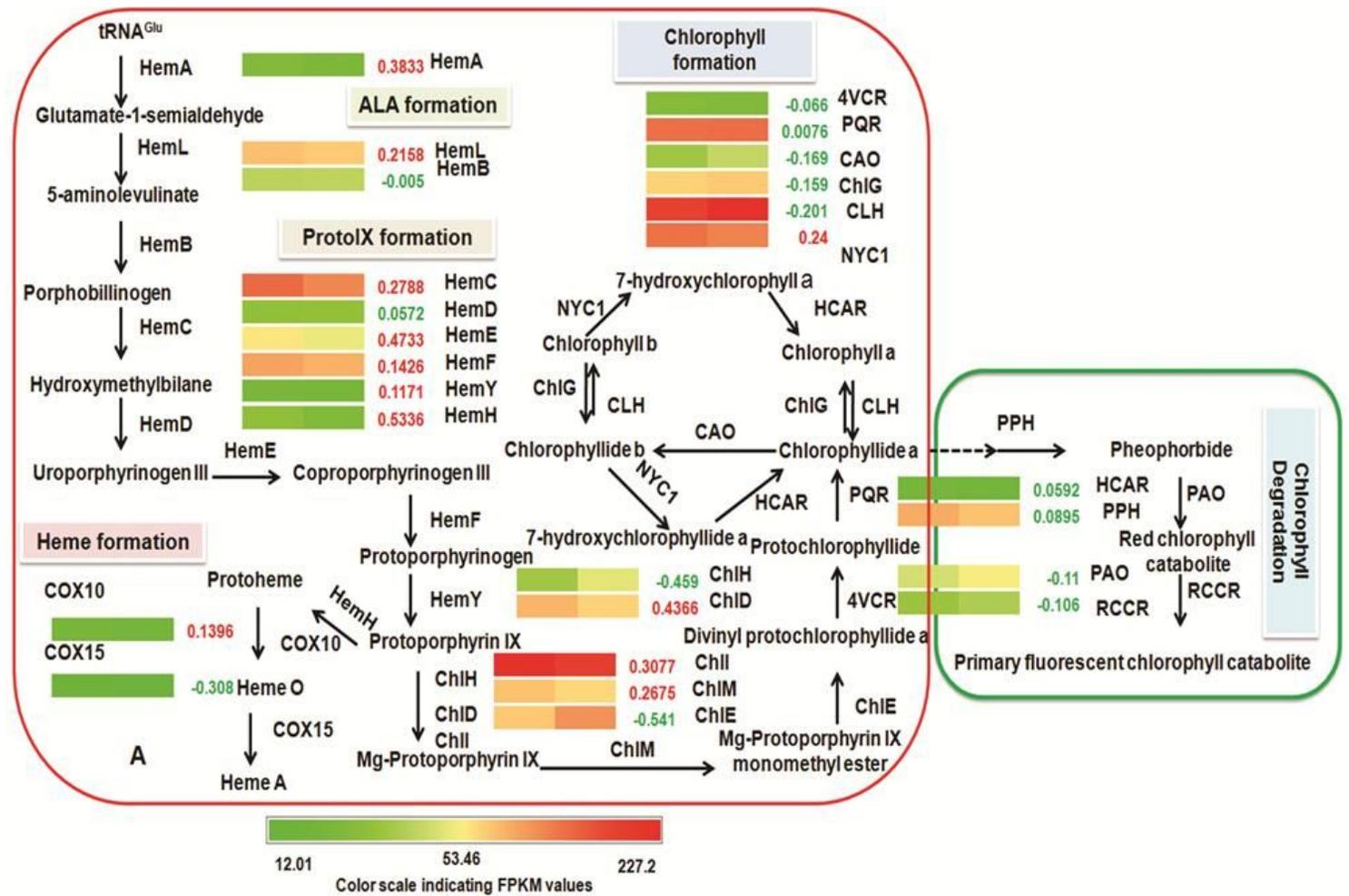
## Figure 4

(A) Venn diagram of differentially expressed genes between the pairwise comparison of OBP1 vs. F1 & Amphid2, OKP2 vs. F1 & Amphid2, F1 vs. OBP1 & OKP2 and Amphid2 vs. F1, OKP2 & OBP1. (B) Gene Ontology enrichment analysis of differentially expressed gene between the pairwise comparison of OBP1 Vs. F1 & Amphid2, OKP2 Vs. F1 & Amphid2, F1 Vs. OBP1 & OKP2 and Amphid2 Vs. F1, OKP2 & OBP1 using agriGo. FDR value less than 0.05 was considered to be significant. (C) Heat map of top 40 transcripts showing reprogramming of transcriptome (the transcripts which was significantly Up regulated in AMPHID2 vs. F1, OKP2 & OBP1 and OKP2 vs. F1 & HID2 was found to be significantly down regulated in F1 vs. OBP1/OKP2 but baseline expressed in OBP1 vs. F1 & HID2) between the pairwise comparison of OBP1 Vs. F1 & Amphid2, OKP2 Vs. F1 & Amphid2, F1 Vs. OBP1 & OKP2 and Amphid2 Vs. F1, OKP2 & OBP1. (D) Heat map of top 40 transcripts showing reprogramming of transcriptome (the transcripts which was significantly down regulated in AMPHID2 vs. F1, OKP2 & OBP1 and OKP2 vs. F1 & HID2 was found to be Up regulated in F1 vs. OBP1 & OKP2 but baseline expressed in OBP1 vs. F1 & HID2) between the pairwise comparison of OBP1 Vs. F1 & Amphid2, OKP2 Vs. F1 & Amphid2, F1 Vs. OBP1 & OKP2 and Amphid2 Vs. F1/OKP2/OBP1 (Red colour shows the significantly Up regulated transcripts and green colour shows the significantly down regulated transcripts  $\log_2$  fold change  $-0.1 \leq 0.1$  ).



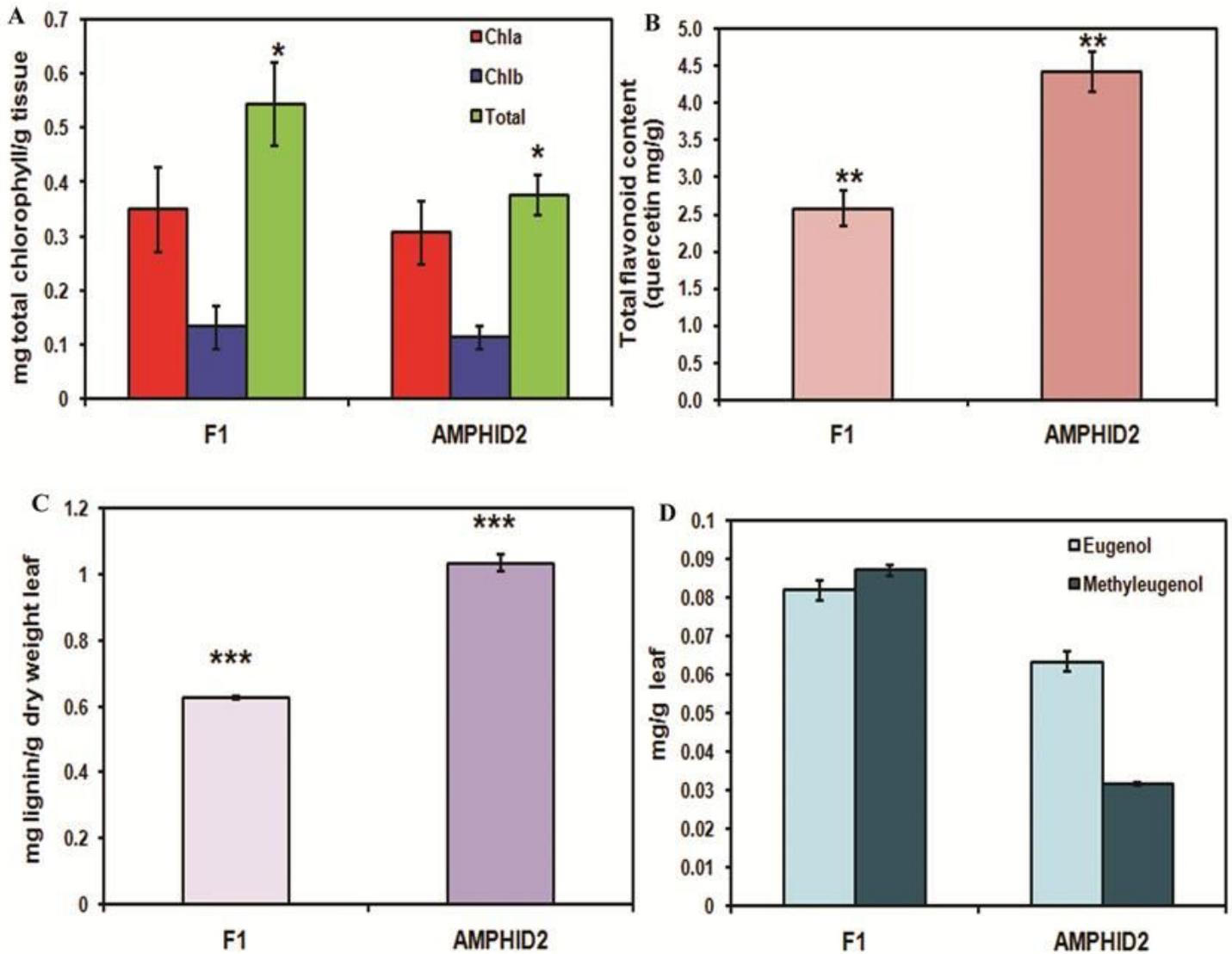
**Figure 5**

Vein diagram representing the total transcripts of parent1 (67,770), parent2 (73,265), interspecific hybrid F1 (76917) and amphidiploids (76563).



**Figure 6**

Metabolic pathway of chlorophyll biosynthesis in interspecific hybrid F1 and Amphidiploid. Colored palette represents the FPKM values and numbers in green color displays down regulated genes and numbers in red color shows up regulated genes. Abbreviations: HemA, Glutamyl-tRNA reductase; HemL, Glutamate-1-semialdehyde 2, 1-aminomutase; HemB, Porphobilinogen synthase; HemC, Hydroxymethylbilane synthase; HemD, Uroporphyrinogen-III synthase; HemE, Uroporphyrinogen decarboxylase; HemF, Coproporphyrinogen-III oxidase; HemN, Oxygen-independent coproporphyrinogen-III oxidase; HemY, Oxygen-dependent protoporphyrinogen oxidase; HemH, Ferrochelatase; COX10, Protoheme IX farnesyltransferase; COX15, Cytochrome c oxidase assembly protein subunit 15; ChIH, Magnesium chelatase subunit H; ChID, Magnesium chelatase subunit D; ChII, Magnesium chelatase subunit I; ChIM, Magnesium protoporphyrin IX methyltransferase; ChIE, Magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase; 4VCR, Divinyl chlorophyllide a 8-vinyl-reductase; PQR, Protochlorophyllide reductase; CAO, Chlorophyllide a oxygenase; ChIG, Chlorophyll synthase; CLH, Chlorophyllase; NYC1, Chlorophyll (ide) b reductase; HCAR, 7-Hydroxymethyl chlorophyll a reductase; PPH, Pheophytinase; PAO, Pheophorbide a oxygenase; RCCR Red chlorophyll catabolite reductase.



**Figure 7**

(A) Total Content change of chlorophyll metabolite in interspecific hybrid F1 and Amphidipiold. The significant difference was compared between interspecific F1 and Amphidipiold using t- test. \* indicates  $P < 0.05$ . (B) Total content change of flavonoid in interspecific hybrid F1 and Amphidipiold. The significant difference was compared between interspecific F1 and Amphidipiold using t- test. \*\* signifies  $P < 0.01$ . (C) Total content change of lignin in interspecific hybrid F1 and Amphidipiold. The significant difference was compared between interspecific hybrid F1. \*\*\* signifies  $P < 0.001$ . (d) Total amount of eugenol and methyl eugenol in the oil extracted from the 100 gram leaves of interspecific hybrid F1 and Amphidipiold.

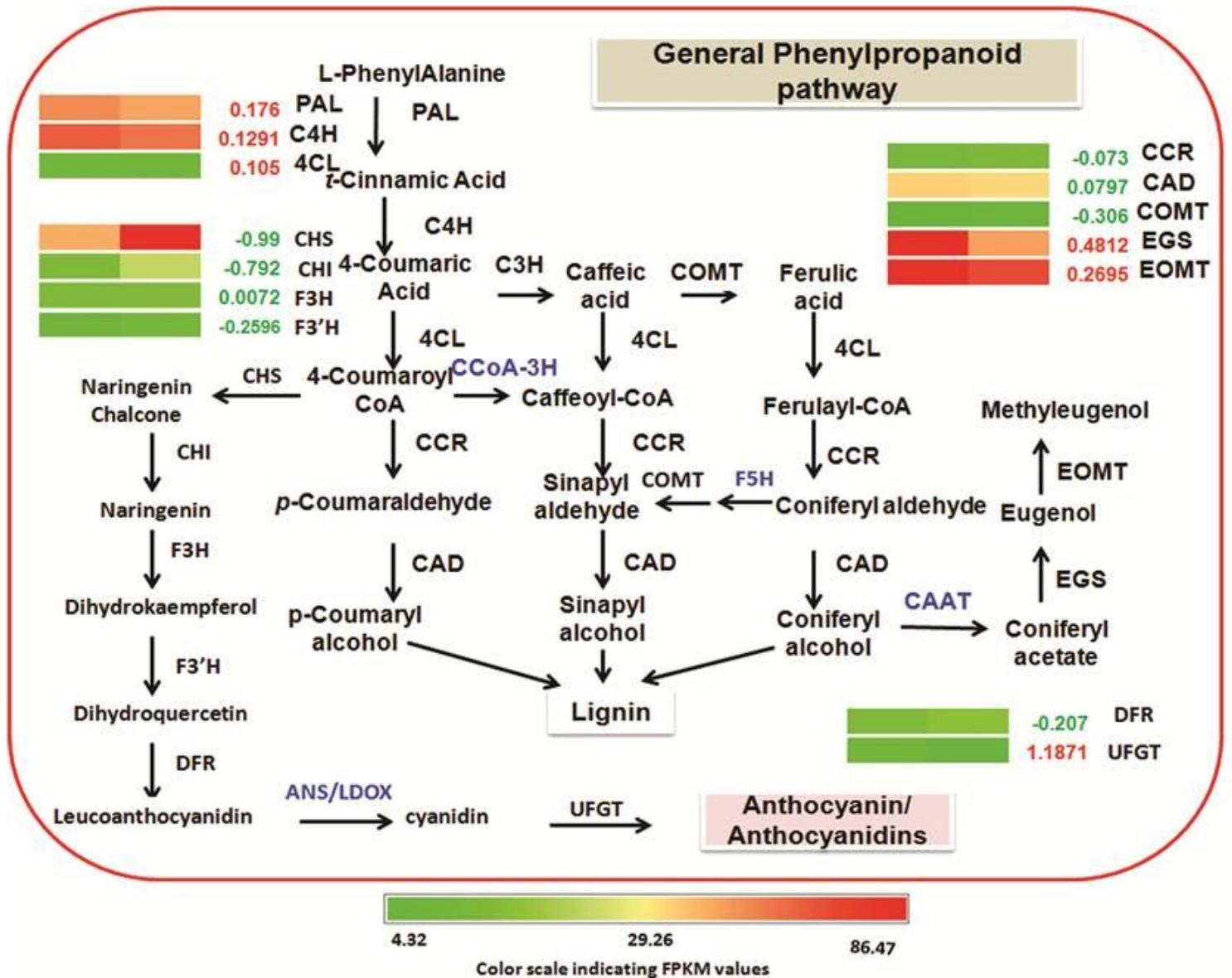
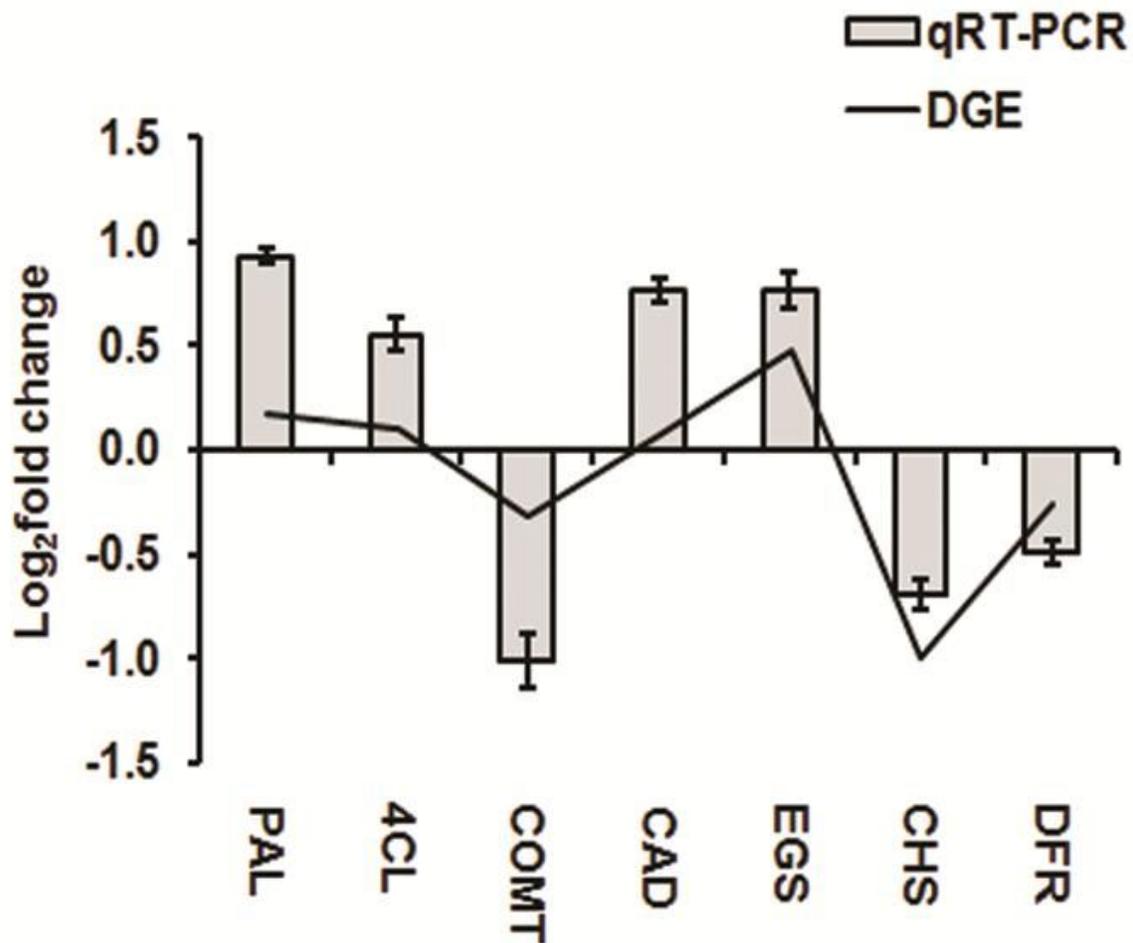


Figure 8

Metabolic pathway of general phenylpropanoid pathway in interspecific hybrid F1 and Amphidiploid. Colored palette represents the FPKM values and numbers in green colour displays down regulated genes whereas numbers in red color shows up regulated genes. Genes highlighted in the blue color were not annotated in the transcriptome. Abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase; C3H, p-coumarate 3-hydroxylase; COMT, caffeoyl O-methyl transferase; CCoAOMT, caffeoyl-CoA O-methyl transferase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; CAAT, coniferyl alcohol acetyl transferase; EGS, eugenol synthase; EOMT, eugenol O-methyl transferase; F5H, Ferulate 5-hydroxylase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS/LDOX, anthocyanidin synthase; UFGT, UDP-glucose: flavonoid 7-O-glucosyltransferase.



**Figure 9**

Validation of DEG results of phenylpropanoid biosynthetic gene with qRT-PCR. Columns in the graph shows the expression pattern (log<sub>2</sub> fold change values) of selected transcripts of phenylpropanoid biosynthetic gene using total RNA isolated from leaf tissues of interspecific hybrid F1 and amphidiploid through quantitative Real time PCR while line represents the digital gene expression values of the same transcripts.

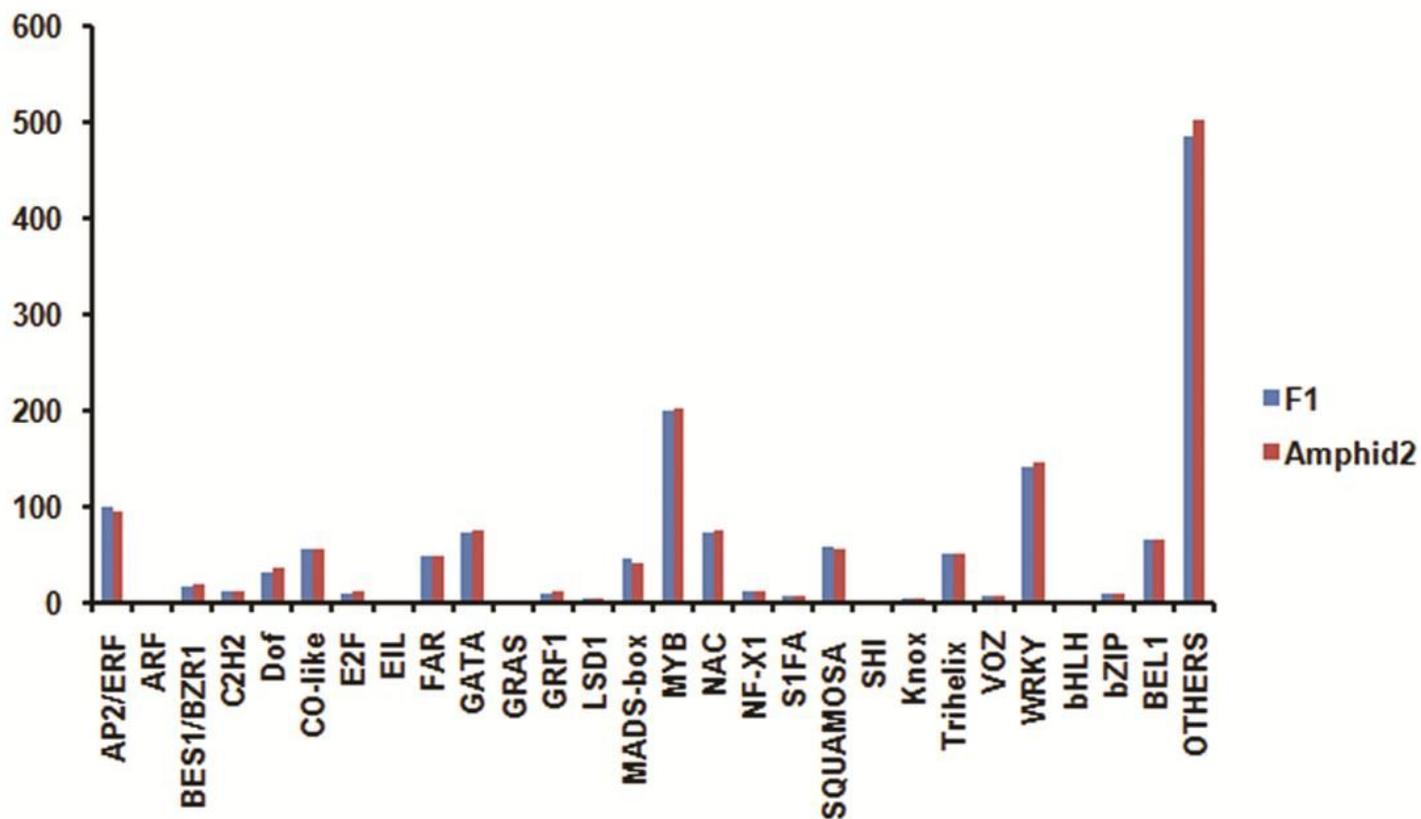


Figure 10

Transcript abundance of putative transcription factors in interspecific hybrid F1 and Amphidiploid.

## Supplementary Files

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

- [AdditionalfileS22.pdf](#)
- [TableS21.docx](#)
- [TableS20DGEAMPHID2vsF1OBP1OKP2.xlsx](#)
- [TableS19DGEF1Vs.OBP1andOKP2.xlsx](#)
- [TableS18DGEF1vsAMPHID2.XLSX.xlsx](#)
- [TableS17DGEOKP2Vs.F1andAMPHID2.xlsx](#)
- [TableS16DGEOKP2vsAMPHID2.xlsx](#)
- [TableS15DGEOKP2vsF1.xlsx](#)
- [TableS14DGEOBP1vsF1Amphid2.xlsx](#)
- [TableS13DGEOBP1vsAmphid2.XLSX.xlsx](#)
- [TableS12DGEOBP1vsF1.xlsx](#)

- [TableS11DGE0BP1vsOKP2.xlsx](#)
- [TableS10AnnotationSummary.xlsx](#)
- [TableS9.docx](#)
- [TableS8.docx](#)
- [TableS7GOanalysisUpregulatedgenes.XLSX](#)
- [TableS6GOanalysisofHybridizationinducedtranscripts.xlsx](#)
- [TableS5Hybridizationinducedtranscriptexpression.xlsx](#)
- [TableS4GOanalysisDownregulatedgenes.xlsx](#)
- [TableS3GOanalysisDownregulatedgenesF1vsOBP1OKP2.XLSX](#)
- [TableS2GOanalysisUpregulatedgenesF1vsOBP1OKP2.XLXS.xlsx](#)
- [TableS1AssemblyReport.xlsx](#)