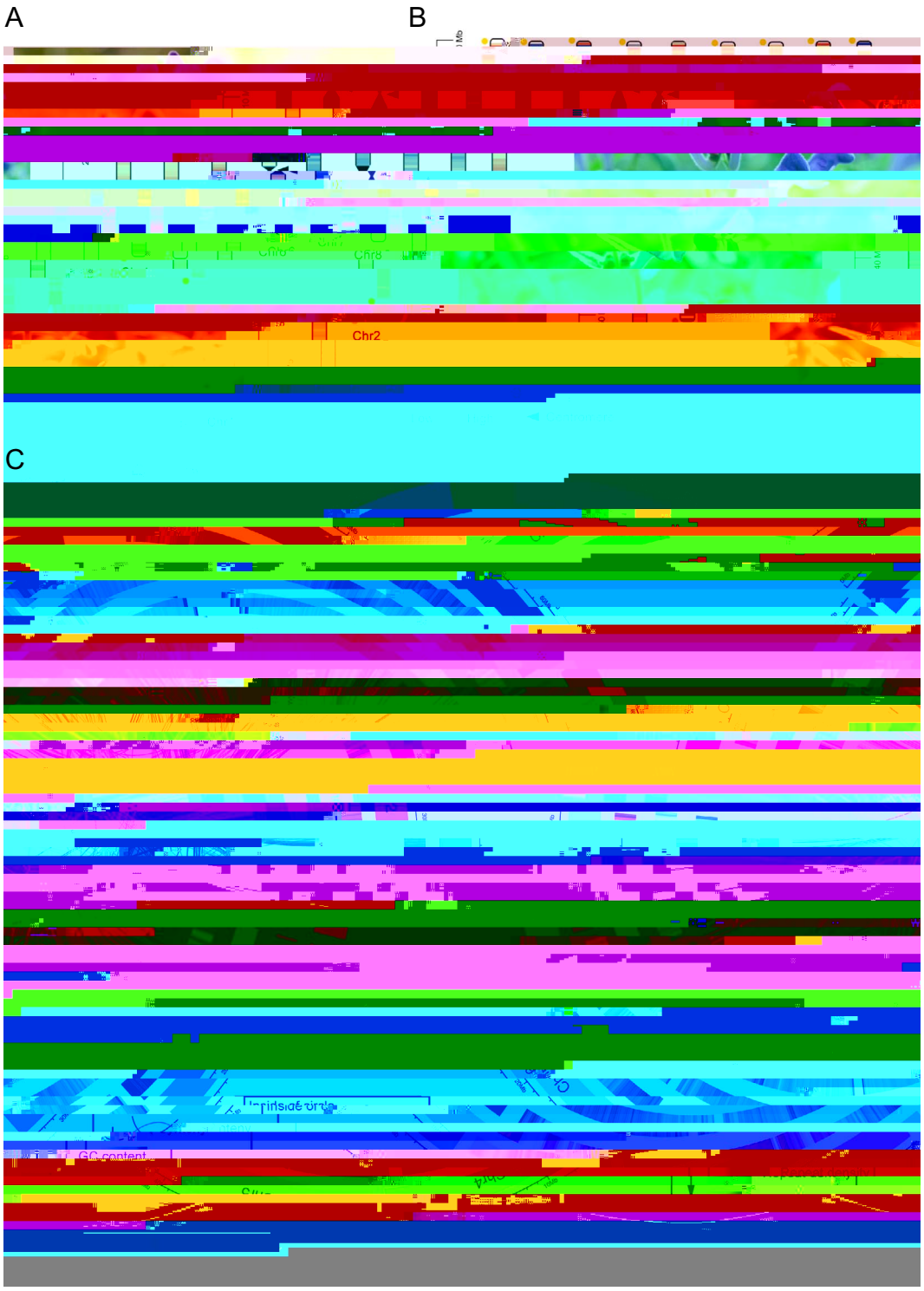
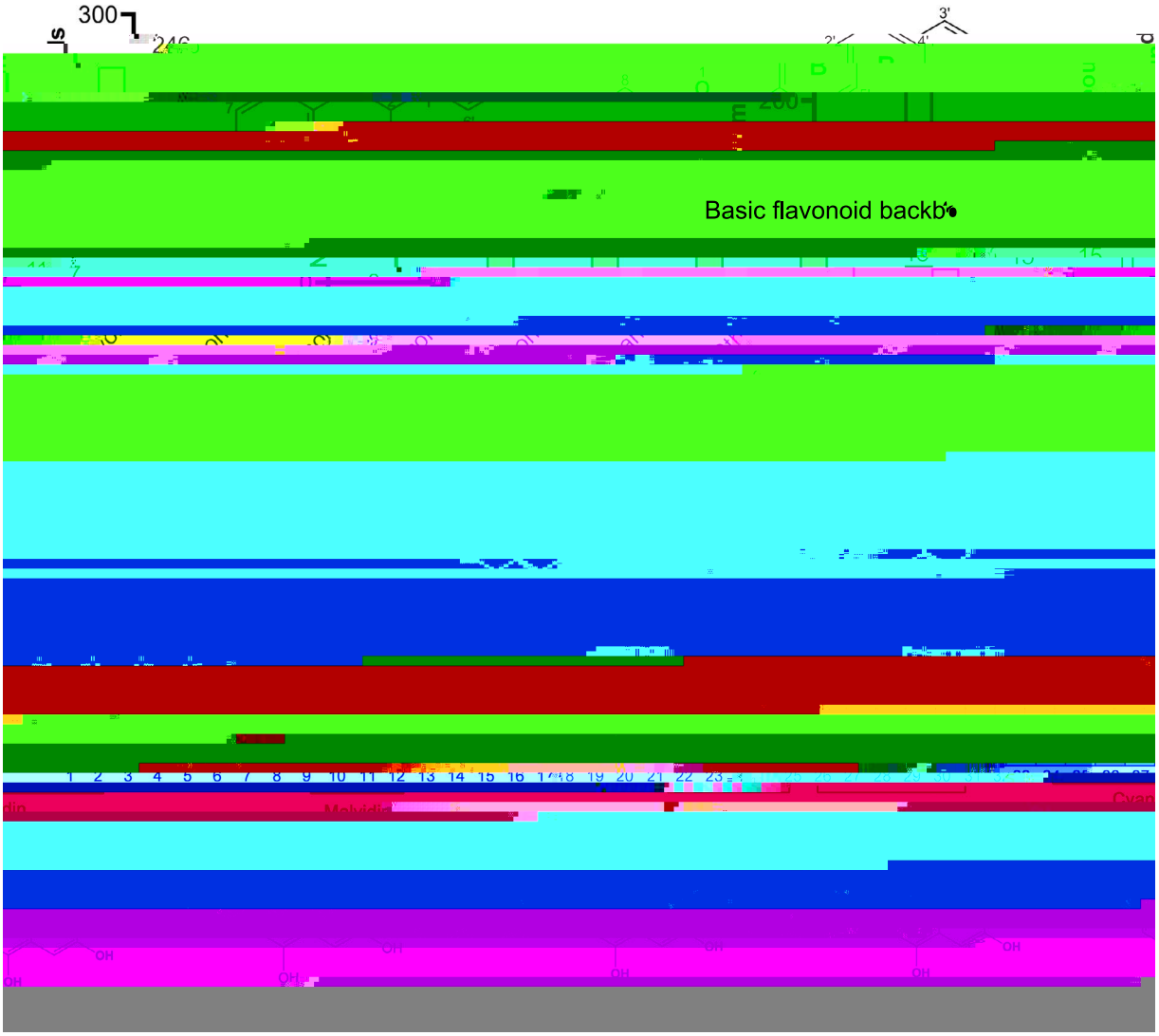


F . (**1.** T2T gap-free genome assembly and genomic features of

genes, . 3 24 and . 3 2 , encoding flavone synthase II-1 (FNSII-1) and FNSII-2, respectively. FNSII-1 can catalyze both naringenin and pinocembrin to produce apigenin and chrysin, respectively, whereas FNSII-2 can only accept pinocembrin as substrate [5]. In addition, two CYP82D enzymes act as flavone 6-hydroxylase (F6H/CYP82D1.1) and F8H (CYP82D2) which are involved in the synthesis of baicalein and wogonin, respectively [18]. However, genome-wide investigation of the . 4 0 family


300
Basic flavonoid backb
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 18 19 20 21 22 23
OH
OH
OH
OH
OH
OH

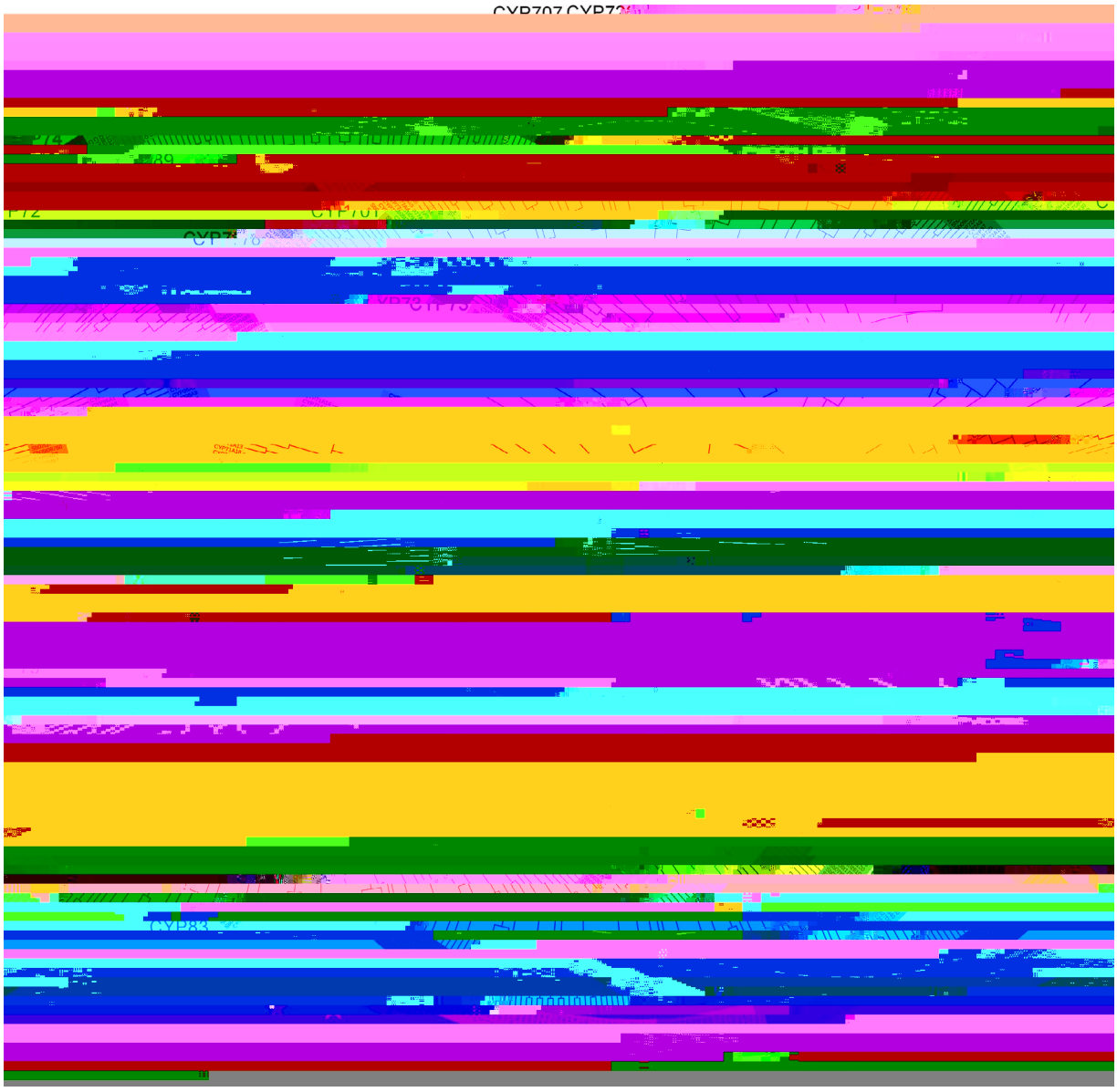


Fig. 4. Phylogenetic tree of CYP450s from and . Subfamilies were annotated using CYP450s from . The phylogenetic tree was constructed using the neighbor-joining (NJ) method with a bootstrap test (=1000 replications). The star indicates the CYP75 enzymes for further study.

potentially responsible for the 3′- or 5′-hydroxylation of flavones in .

Based on fragments per kilobase of exon model per million mapped fragments (FPKM) values obtained from RNA-seq of different tissues in , it was observed that 2, and exhibited comparatively high expression levels in both flowers and flower buds (Fig. 5C), which were likely involved in the pigmentation process of flowers. The average FPKM value of in flower buds was 22.16 and 16.63 times higher than those in 2 and , respectively (Table S20, see online supplementary material), suggesting that SbFBH7 was the key enzyme involved in the coloration of flowers. Transcripts of 1 and 4 were equally accumulated in plant aerial parts, while 3 and exhibited high expression levels in leaves, roots, and stems. Additionally, the transcripts of 4, , and were up-regulated in JA-treated roots, indicating their potential involvement in the hydroxylation of flavones in the roots.

Functional characterization of SbFBHs

To further analyse the enzymatic functions, full-length ORFs of were successfully constructed into the expression vectors and transformed into yeast. The strains were then fermented with different substrates. Novel peaks (Peak I) corresponding to the retention time of the eriodictyol standard were detected when SbFBH1, 2, 5, and 7 were incubatedI) corresponding

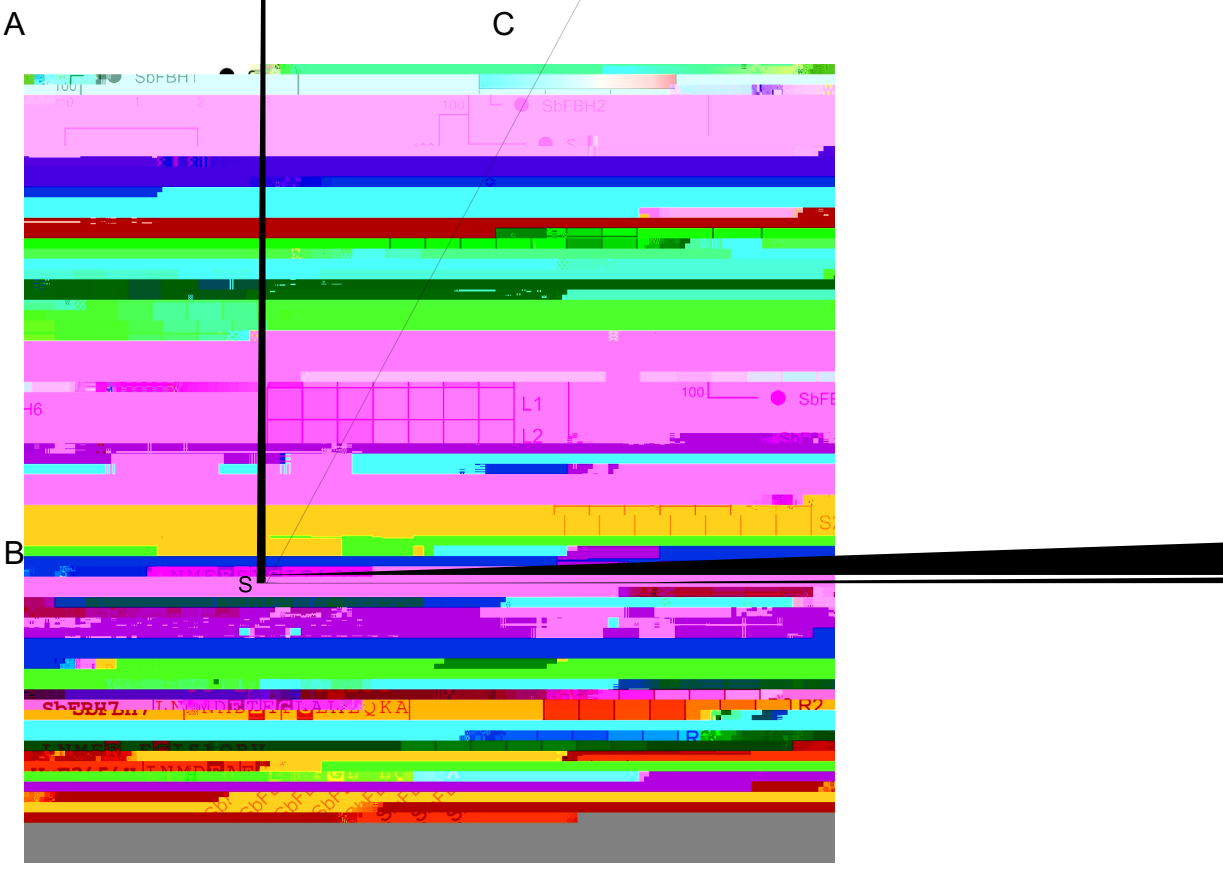

A
C
B
S
SbFBH2
L1
L2
100

Figure 6. Yeast enzyme assays of SbFBHs. A–J HPLC analysis and the reaction catalyzed by SbFBHs using naringenin (A, B), dihydrokaempferol (C, D), kaempferol (E, F), apigenin (G, H), and scutellarein (I, J) as substrates in yeast enzyme assays, respectively.

genome in the analysis of biosynthetic pathways for specialized metabolites.

Approximately 300 flavonoids have been identified in . [4]. Previous studies have primarily focused on the 4′-deoxyflavones, such as baicalin, wogonoside, baicalein, and wogonin, which are highly accumulated in the roots. These compounds are known to contribute to the majority of the health benefits associated with . , including a broad

range of antitumor properties [3]. Anthocyanins are an important subgroup of flavonoids possessing antioxidant, anticancer, and antibacterial properties, in addition to serving as flower pigments [32]. Anthocyanins can be classified into six major classes: cyanidin, delphinidin, malvidin, petunidin, pelargonidin, and peonidin, which contribute to a diverse range of colors observed in flowers. For instance, delphinidin-based anthocyanins contribute to blue and purple colors, while cyanidin-based anthocyanins give rise to magenta and red colors. The color spectrum of anthocyanins is influenced by the number of hydroxyl groups present on the B-ring, with greater hydroxylation resulting in bluer color [33]. Metabolome analysis has revealed that purple and purple-red flowers of . variants contained

Materials and methods

Plant materials, DNA extraction, library construction and sequencing

All sequencing materials were sourced from an individual plant that was cultivated and maintained at Shang-

removal of redundant sequences, the final repetitive sequence set was obtained.

Gene structure was annotated using a combined strategy including homologous prediction, v gene prediction, as well as gene prediction based on RNA-seq and PacBio data. For homologous prediction, protein sequences were mapped to the reference genome, including .. h , . . , v , and . . . h using TblastN v2.7.1 [54], then Exonerate v2.4.0 was used to predict the transcripts and coding region [55]. In addition, genes predicted by BUSCO (which was performed in genome quality assessment) were also used as the homolog prediction results [56]. v gene structure identification was based on Augustus v3.3.2 [57] and GlimmerHMM v3.0.4 [58]. RNA-seq reads were filtered using fastp v0.21.0 [42], and were aligned to the genome using HISAT2 v2.1.0 [59]. Alignment results were then used as input for Stringtie v2.1.4 to obtain transcripts [60], and then predicted using TransDecoder v5.1.0. Nanopore RNA-seq reads were filtered using NanoFilt v2.8.0. Full-length sequences were identified using Pychopper v2.7.2. After error correction using racon v1.4.21, these full-length sequences were aligned to genome using minimap2 v2.17-r941 [61]. Alignment results were then used as input for Stringtie v2.1.4 [60], and then predicted using TransDecoder v5.1.0. All predicted gene sets were then merged into a gene set via MAKER v2.31.10 [62

for extraction, and candidates with a length <400 and >600

5. Zhao Q, Zhang Y, Wang G. . . A specialized flavone biosynthetic pathway has evolved in the medicinal plant, ... v. 2016;2:1501780
6. Pei TL, Yan MX, Huang YB. . . Specific flavonoids and their

44. Huson DH, Beier S, Flade I. . MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. . 2016;12:e1004957
45. Guan D, McCarthy SA, Wood J. . Identifying and removing haplotypic duplication in primary genome assemblies. - . 2020;36:2896–8