Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches

Xiaonan Liu1,2, Jian Cheng1, Guanghui Zhang3, Wentao Ding1, Lijin Duan1, Jing Yang3,4, Ling Kui2,5, Xiaozhi Cheng1, Jiangxing Ruan1, Wei Fan3, Junwen Chen3, Guangqiang Long3, Yan Zhao3, Ping Cai6, Wen Wang5,7, Yanhe Ma1, Yang Dong3,4, Shengchao Yang3,4 & Huifeng Jiang1

The flavonoid extract from *Erigeron breviscapus*, breviscapine, has increasingly been used to treat cardio- and cerebrovascular diseases in China for more than 30 years, and plant supply of *E. breviscapus* is becoming insufficient to satisfy the growing market demand. Here we report an alternative strategy for the supply of breviscapine by building a yeast cell factory using synthetic biology. We identify two key enzymes in the biosynthetic pathway (flavonoid-7-O-glucuronosyltransferase and flavone-6-hydroxylase) from *E. breviscapus* genome and engineer yeast to produce breviscapine from glucose. After metabolic engineering and optimization of fed-batch fermentation, scutellarin and apigenin-7-O-glucuronide, two major active ingredients of breviscapine, reach to 108 and 185 mg l⁻¹, respectively. Our study not only introduces an alternative source of these valuable compounds, but also provides an example of integrating genomics and synthetic biology knowledge for metabolic engineering of natural compounds.
The traditional Chinese medicinal plant *Erigeron breviscapus* has been used for more than 1,000 years\(^1\). Breviscapine, the total flavonoid extract of *E. breviscapus*, was classified as a prescription drug of Chinese medicine 30 years ago for the clinical treatment of cardiovascular and cerebrovascular diseases\(^2\), which are the leading causes of death and disability for older people around the world\(^3\). Significantly, it has been proven to be effective in dilating vessels, inhibiting platelet aggregation, promoting blood circulation, alleviating myocardial ischemia-reperfusion injury, removing blood stasis and reducing oxidative damage\(^4\). At present, more than ten million patients use breviscapine and related drugs each year in China (Supplementary Note 1). Although 1200 ha are under cultivation with *E. breviscapus* annually, yielding about 4,500 tons of stover, agricultural producers are not able to fulfill the ever growing demand of the Chinese market\(^5\). It is therefore urgent to develop an alternative, sustainable way to secure the breviscapine supply.

In recent years, engineered yeast strains have been used to produce various natural products of plant origin, including artemisinic acid\(^6,10\), ginsenosides\(^11,12\), etoposide aglycone\(^13\), and opioids\(^14\). Furthermore, biosynthetic pathways of flavonoids such as naringenin\(^15\) and apigenin\(^16\) have also been engineered into yeasts. The biosynthetic pathway from l-phenylalanine to apigenin has been well characterized\(^17\), which comprises six consecutive steps, catalyzed by phenylalanine ammonia-lyase, tyrosine ammonia-lyase, dihydrodiphenylpyruvate synthase, flavone synthase, chalcone isomerase, and flavone synthase II. Although considerable efforts have been made to identify the key enzymes in the biosynthetic pathway of breviscapine\(^18,19\), the complete pathway from apigenin to breviscapine in *E. breviscapus* was still unclear before this study.

In this work, we set out to decode the biosynthetic pathway of breviscapine by combining genomic analysis and synthetic biology tools. Breviscapine mainly contains scutellarin, along with a small amount of apigenin 7-O-glucuronide\(^18\). Using genomic analysis, we identified two key enzymes flavonoid-7-O-glycosyltransferase (F7GAT), which converts apigenin into apigenin-7-O-glucuronide, and flavone-6-hydroxylase (F6H), which functions together with F7GAT to produce scutellarin from apigenin. Subsequently, we constructed a yeast cell factory to produce these compounds of breviscapine from glucose using synthetic biology tools. This work provides insights for the identification of biosynthetic pathways of natural products from traditional Chinese medicinal herbs and illustrates the potential of yeast fermentation system in producing natural compounds through synthetic biology approach.

**Results**

**Construction of apigenin-producing platform.** In order to decipher the biosynthetic pathway of breviscapine, the draft genome of *E. breviscapus* was re-annotated by combination of genomic and transcriptomic data (Methods). Multiple copies of the known genes in the flavonoid biosynthesis pathways from phenylalanine to apigenin in *E. breviscapus* were identified based on sequence homology\(^20-22\) (Supplementary Table 1). The candidate genes for hypothetical pathways from apigenin to the two major components of breviscapine were proposed (Fig. 1a). One component of breviscapine is apigenin-7-O-glucuronide, which was expected to be synthesized from apigenin by F7GAT. Another major component of breviscapine is scutellarin, which can plausibly be produced by two putative pathways. One would comprise hydroxylation by F6H to yield the aglycone scutellarin, followed by glycosylation catalyzed by F7GAT, while the other would comprise glycosylation yielding apigenin-7-O-glucuronide, followed by hydroxylation of the flavonoid at position six by F6H.

To confirm our hypothetical pathway to breviscapine, we first proposed to build an engineered yeast platform to characterize functional genes in the genome of *E. breviscapus*. The highly expressed copies for each known gene were chosen to build biosynthetic pathway from phenylalanine to apigenin in yeast (Supplementary Table 2). The engineered DNA fragments with all known genes were integrated into the genome of *Saccharomyces cerevisiae* W303-1B using a modularized two-step chromosome integration approach (Fig. 1b, and Supplementary Tables 3, 4 and Methods\(^23,24\)). The product of apigenin in the engineered yeast was confirmed by high-performance liquid chromatography (HPLC) and mass spectrometry (MS) (Supplementary Fig. 1). Thus, we constructed an engineered yeast platform (SC1) to produce precursor apigenin.

**Functional characterization of F7GAT.** F7GATs, which typically transfer UDP-glucuronic acid to the C7 position of flavonoids, have been reported in Lamiales\(^24-27\), but the functional gene had not yet been confirmed in *E. breviscapus*. A total of 83 UDP-glycosyltransferase (UDPGT) genes were identified in the *E. breviscapus* genome and assigned into 15 gene families (Supplementary Fig. 2). Previous studies suggested that enzymes of the UGT88 family can transfer UDP-glucuronic acid to the C7 position of flavonoids\(^24,26\). Among the identified UDPGT genes, only one belongs to the UGT88 gene family\(^23,27\) (Supplementary Fig. 3), indicating that it may be the F7GAT of *E. breviscapus*. With UDP-glucuronic acid as sugar donor, *E. b. F7GAT* was indeed able to utilize both apigenin and scutellaren as substrates in vitro (Supplementary Fig. 3). The kinetic parameters of purified *E. b. F7GAT* were determined for apigenin (*K_m* 9.24 µM and *k_cat* 0.57 s⁻¹) and scutellaren (*K_m* 70.15 µM and *k_cat* 0.24 s⁻¹) (Supplementary Table 5). Furthermore, *E. b. F7GAT* also can convert many other flavonoid substrates such as naringenin, kaempferol, quercetin, chrysin, baicalin, luteolin, diosmetin, and chrysoeriol at high conversion rates in vitro (Supplementary Fig. 4 and Supplementary Table 6).

In order to confirm the function of *E. b. F7GAT* in vivo, we first identified the UDP-glucose dehydrogenase of *E. breviscapus* (EhUDPGDH) based on the orthologous gene from *Glycine max*\(^28\) (Supplementary Fig. 5), which provides substrate in vivo for *E. b. F7GAT*. Then a vector carrying *E. b. F7GAT* and *E. b. UDGPDH* was introduced into the engineered yeast strain SC1. A new chromatographic peak was detected among the fermentation products around 6.5 min, which was consistent with the retention time of the authentic apigenin-7-O-glucuronide reference standard (Fig. 1c). Furthermore, we confirmed that the molecular weight of the compound in the new peak was the same as that of apigenin-7-O-glucuronide by liquid chromatography-MS (LC-MS) (Fig. 1d). After 5 days of fermentation, the engineered yeast produced about 20 mg l⁻¹ of apigenin-7-O-glucuronide, of which 26.6% was located within the cells and 73.4% had secreted into the culture broth (Supplementary Table 7).

**Evolution and expression of the P450 genes.** It has been published that F6H should belong to the cytochrome P450 family\(^29-31\), which is the largest gene family in plants\(^32,33\). In order to identify the gene encoding F6H in *E. breviscapus*, a total of 312 putative P450 genes were annotated in the *E. breviscapus* genome based on known P450 genes (Supplementary Data 1 and Methods). To narrow down the list of candidates, we used a strategy of combining chemotaxonomy with evolutionary genomics. We analyzed the distribution of scutellaren production in five sequenced species in the Asteraceae family and eight sequenced species from three other representative plant taxa (Fig. 2a). All of the five studied species from the Asteraceae family (*E. breviscapus*, *Conyza canadensis*, *Lactuca sativa*, *Cynara cardunculus*, and...
**Carthamus tinctorius** produced scutellarin, but we did not detect scutellarin in non-Asteraceae species (Fig. 2b). These pieces of evidence indicated that the encoding gene of F6H in *E. breviscapus* may be an Asteraceae-specific gene.

In order to identify Asteraceae-specific P450 genes, we annotated all P450 genes in the five sequenced species from the Asteraceae family. Using eight non-Asteraceae species as the out group, we identified 134 Asteraceae-specific P450 genes in the *E. breviscapus* genome (Fig. 2c). However, after comparing the expression levels of all genes in the wild-type species and breviscapine-overproducing cultivars of *E. breviscapus*, we did not find significant differences between the Asteraceae-specific and non-Asteraceae-specific P450 genes (Fig. 2d). Interestingly, we found that genes in the flavonoid biosynthetic pathway showed very high gene expression levels in both the wild type and the cultivated breeds. It is therefore possible that F6H may also be highly expressed in *E. breviscapus*.

**Functional identification of F6H.** To identify the gene encoding F6H from dozens of candidates, we developed a functional screening method for P450 genes (Fig. 3a and Methods). Recombinant vectors harboring each of 36 highly expressed candidate genes and a cytochrome P450 reductase gene (EbCPR), which was cloned from *E. breviscapus* to transport electrons for P450 catalysis, were constructed by Gibson assembly. The resulting vectors were transferred into the engineered yeast strain (SC1-FU) that produces apigenin-7-O-glucuronide (Supplementary Table 4). After fermentation for 5 days in baffled 24-well plates, metabolic products of the engineered yeasts were analyzed by HPLC. We screened the putative F6H candidate genes with priorities according to their expression levels. Fortunately, a highly expressed gene, CYP706X, induced a new peak with a retention time of 4.6 min, which was the same as that of the authentic scutellarin standard (Fig. 3b). Moreover, the molecular weight of the compound in the new peak was also the same as that of scutellarin, as determined by LC-MS (Fig. 3c). Furthermore, when the vector containing CYP706X and EbCPR was transferred into our starting strain W303-1B, the engineered yeast was able to convert apigenin into its 6-hydroxy derivative scutellarin (Supplementary Fig. 6a), whose structure was confirmed.
Engineered the host cells to produce the flavonoids scutellarin and breviscapine from S. flavus. To overcome this problem, we introduced all the genes of the corresponding biosynthetic pathways into yeast, the engineered yeast was not able to convert apigenin-7-O-glucuronide into its 6-hydroxy derivative scutellarin in vivo (Supplementary Fig. 6c).

In order to further confirm the function of CYP706X and EbCPR in vitro, we proposed to isolate microsomal proteins from the engineered strains expressing CYP706X and EbCPR. In vitro assays of microsomal proteins were conducted using apigenin or apigenin-7-O-glucuronide as the substrate. We found that the microsomal protein fractions containing CYP706X and EbCPR were able to catalyze the hydroxylation of apigenin at position six, but could not convert the hydroxyl group to apigenin-7-O-glucuronide (Fig. 3d, e). Thus, the P450 gene CYP706X was identified as EbF6H. Furthermore, we speculated that apigenin should first be converted into scutellarein by EbF6H, after which scutellarein could be converted into scutellarin by EbF7GAT in E. breviscapus.

**Biosynthesis of breviscapine in yeast.** Although we were able to produce both components of breviscapine when we introduced all genes of the corresponding biosynthetic pathways into yeast, the yield was still very low, with only 9.2 mg of scutellarin per gram of dry cell weight (DCW). To overcome this problem, we further engineered the host cells’ metabolism to increase the breviscapine yield. As three molecules of malonyl-CoA are needed to form the flavonoid backbone, we surmised it would be helpful to improve the intracellular concentration of malonyl-CoA (Fig. 4a). As the concentration of acetyl-CoA, the immediate precursor of malonyl-CoA, undergoes dynamic changes due to continuous generation and consumption, we first tried to decrease the consumption of acetyl-CoA by deleting the cytosolic malate synthase gene (MLS1), and thus preventing the associated oxidation of acetyl-CoA. This improved the yield of scutellarin from 9.2 to 11.6 mg g\(^{-1}\) DCW. In the next optimization step, we deleted the peroxisomal citrate synthase gene (CIT2) in the yeast strain, which has been published to block the consumption of acetyl-CoA. The scutellarin yield in this double-knockout strain was improved to 14.3 mg g\(^{-1}\) DCW (Fig. 4b).

Second, we proposed to increase the supply of acetyl-CoA in the metabolic network. Overexpression of the endogenous alcohol dehydrogenase gene (ADH2) can lead to the conversion of more ethanol into acetaldehyde and acetyl-CoA. Moreover, overexpression of the endogenous aldehyde dehydrogenase gene (ALD6) and introduction of an acetyl-CoA synthetase variant from *S. enterica* (ACS\(_{SE}\)) was reported to increase the metabolic flux towards malonyl-CoA. We designed a malonyl-CoA-producing multiple-gene module that integrated the genes ALD6, ACS\(_{SE}\), and ADH2 into the YPRC215 locus of the double-knockout strain. The resulting engineered strain was able to produce 15.5 mg g\(^{-1}\) DCW of scutellarin from 2% glucose in flask cultures. Thus, we obtained an engineered yeast strain with a 1.8-fold improved scutellarin yield compared with the starting strain (Fig. 4b).

Finally, the engineered strain was tested in scaled-up fermentations in a 3-L benchtop fermenter. During the
ferrmentation process, the dissolved oxygen concentration (DOC) is crucial for the proportion of breviscapine in total flavonoids, as EbF6H is a P450 enzyme, and therefore requires molecular oxygen for its catalytic reaction\(^\text{39}\). When we optimized the DOC during the fermentation, the maximum yields of scutellarin and apigenin-7-O-glucuronide after 7 days of fed-batch fermentation reached 108 and 185 mg L\(^{-1}\), respectively (Fig. 4c). Although the obtained yield is already quite high compared with the starting strain, it should be possible to further increase it via metabolic engineering and fermentation optimization in the future to utilize the full potential of this yeast fermentation system.

**Discussion**

P450 enzymes catalyze key steps in the biosynthetic pathways of many highly valuable plant natural products. In our study, we identified 36 new P450 subfamilies in the genome of *E. breviscapus* (Supplementary Data 1). One of subfamilies that performs the F6H function appears to have originated from the ancient P450 gene family CYP706A, whose members can catalyze the hydroxylation of terpenoids\(^\text{40,41}\). This subfamily is profoundly different from the reported F6H in other plants, such as the CYP82D subfamily from the Lamiaceae *Ocimum basilicum*\(^\text{31}\), *Mentha piperita*\(^\text{31}\), *Scutellaria baicalensis*\(^\text{42}\), and *Salvia miltiorrhiza*\(^\text{43}\), as well as the less closely related to CYP71D9 from *G. max*\(^\text{30}\). The orthologous genes of these reported F6H in *E. breviscapus* genome did not perform F6H function at all (Supplementary Table 8). Furthermore, based on evolutionary analysis, we found that CYP706A was present in four copies in the common ancestor of the Asteraceae (Supplementary Fig. 8). However, only one of the CYP706X copies evolved the flavone 6-hydroxylation activity (Supplementary Fig. 8). Interestingly, this function was present in all the studied species of Asteraceae, as the EbF6H orthologs from the studied Asteraceae species (*CcF6H*, *LsF6H*, *CcarF6H*, and *CfF6H*) also showed the same function (Supplementary Fig. 9). Moreover, this was consistent with our observations regarding the distribution of the scutellarin production phenotype (Fig. 2b). We therefore speculated that F6H may have independently originated in the common ancestor of the Asteraceae, while more research needs to be conducted to illustrate the underlying evolutionary mechanisms. Our study also illustrates the power of combining genomic analysis and synthetic biology knowledge for metabolic engineering of natural compounds. Otherwise, it would be extremely difficult to pinpoint a single novel gene out of several hundred P450 candidates.

The vigorous development of sequencing technologies has enabled great breakthroughs in the identification of biosynthetic pathways of natural products\(^\text{44}\). In this study, we combined...
genomic and transcriptomic technologies to uncover the complete biosynthetic pathways of two major components of breviscapine—scutellarin and apigenin-7-O-glucuronide. In this pathway, the precursor apigenin is directly converted into apigenin-7-O-glucuronide by EbF7GAT and can also be used to synthesize scutellarin by EbF6H and EbF7GAT. Accordingly, both compounds were produced by an engineered yeast strain expressing EbF6H and EbF7GAT. Currently, the total yield of breviscapine in the engineered yeast is about 300 mg l\(^{-1}\) (the sum of 108 mg l\(^{-1}\) scutellarin and 185 mg l\(^{-1}\) apigenin-7-O-glucuronide), which is comparable to the yields of many other flavonoids such as naringenin\(^{15,45}\) and apigenin\(^{16,46}\). Moreover, it is feasible to modulate the ratio of scutellarin to apigenin-7-O-glucuronide in the final fermentation product according to the real ratio of both compounds in breviscapine by regulating the gene expression levels of EbF6H and EbF7GAT in engineered yeasts. It should therefore be possible to directly produce the drug breviscapine using engineered yeasts in future.

To our best knowledge, this work represents the first biosynthesis of breviscapine in a heterologous host starting from glucose. By combining genomic and transcriptomic analysis and high-throughput functional screening, we identified all of the enzymes of the breviscapine biosynthetic pathway in *E. breviscapus*, including EbF6H and EbF7GAT. Our work also illustrates the potential of the yeast fermentation system to increase the yield of breviscapine through metabolic engineering and the optimization of fermentation conditions. Moreover, this work provides new insights for the identification of biosynthetic pathways of natural products from thousands of traditional Chinese medicinal herbs and a tremendous potential to cheaply produce these compounds through synthetic biology.

**Methods**

**Gene re-annotation.** The draft genome of *E. breviscapus* was released as Data Note in our previous study\(^7\). However, the total genome assembly is ~80% of the estimated genome size. In order to detect all genes in the biosynthetic pathway of breviscapine, we re-annotated genes in *E. breviscapus* by combining the genomic data and the transcriptomic data, which include 40.73 GB RNA-seq data from the leaves of five wild individuals, 39.74 GB RNA-seq data from the leaves of six cultivated individuals and 37.03 GB RNA-seq data from the leaf, root, stem, and flower tissues of a cultivated individual\(^{17}\). These RNA-seq data were respectively assembled to predict the protein-coding genes. Then all predicted genes were merged together to form a comprehensive and non-redundant reference gene set using OrthoMCL\(^{38}\) and EVidenceModeler\(^{49}\). Finally, 51,715 protein-coding genes in *E. breviscapus* were annotated (Supplementary Data 2), which performed about 96% completeness using BUSCO\(^{50}\).

**Annotation and analysis of P450 genes.** The P450 genes in the *E. breviscapus* genome were predicted using hmmsearch in conjunction with the Pfam\(^{51,52}\) and EVidenceModeler\(^{49}\). Putative P450 proteins were screened by amino acid length (350 < length < 650). A total of 312 P450 genes were predicted in the *E. breviscapus* genome (Supplementary Data 1).

In order to analyze the evolutionary process of the P450 family in the *E. breviscapus* genome, we compared the P450 genes of *E. breviscapus* with those of its close relatives and other typical eudicotyledons. Four extra Asteraceae
species (i.e., C. canadensis, L. sativa, C. cardunculus, and C. tinctorius), and eight non-Asteraceae species that are typical representatives of seven eudicotyledon orders (i.e., Catharanthus roseus, Daucus carota, Actinidia chinesis, Nicotiana sylvestris, Solanum tuberosum, Malus domestica, G. max, and Arabidopsis thaliana) were used. The protein sequences of D. carota, N. sylvestris, S. tuberosum, M. domestica, and A. thaliana were directly downloaded from the NCBI website (www.ncbi.nlm.nih.gov). The protein sequences of C. canadensis, L. sativa, C. cardunculus, C. tinctorius, A. chinesis, and C. roseus were predicted using AUGUSTUS39, based on genome sequences downloaded from NCBI website. All potential P450 proteins in the twelve species were respectively predicted using the same process above. We found a total of 3663 P450 proteins in thirteen plant genomes (including *E. breviscapus*, Supplementary Table 9). All 3663 P450 proteins were further classified into 411 P450 subfamilies based on 2 criteria: (1) the proteins in a subfamily had relatively closer phylogenetic relationship in the phylogenetic tree, which was constructed by the alignment of all 414 P450 proteins; (2) the identity between two protein sequences in a subfamily was higher than 55%54. Among 411 P450 subfamilies, the subfamilies including 414 P450 proteins in 5 Asteraceae species, were regarded as Asteraceae-specific P450 subfamilies. All 312 P450 enzymes in the *E. breviscapus* genome could be classified into 60 non-Asteraceae specific subfamilies including 178 P450 genes and 36 Asteraceae-specific subfamilies including 134 P450 genes (Supplementary Data 1). It should be noted that “Asteraceae-specific” is a relative definition, as there were only eight non-Asteraceae species in the outgroup, so that some bona fide Asteraceae-specific P450 proteins in our study may be present in other non-Asteraceae species. The true Asteraceae-specific P450 proteins are thus likely a subset of our bona fide Asteraceae-specific P450 proteins.

**Extraction of flavonoids.** Plant materials from *E. breviscapus*, *C. canadensis*, *L. sativa*, *C. cardunculus*, *C. tinctorius*, *A. chinesis*, *D. carota*, *C. roseus*, *N. sylvestris*, *S. tuberosum*, *M. domestica*, *G. max*, and *A. thaliana* were obtained from Yunnan Agricultural University. For preparation of flavonoids, the plant materials were ground in a mortar under liquid nitrogen. The resulting powder was dissolved in 80% methanol (W/V = 1:5), ultrasonicated for 30 min and the cell debris was removed by centrifugation at 13,000 × g for 30 min. The supernatant was used for HPLC analysis.

**Chemicals and media.** Yeast nitrogen base without amino acids and ammonium sulphate (YNB), Bacto peptone, Bacto yeast extract, Luria Broth (LB), agar, lithium acetate, ssDNA, and glucose were obtained from Solarbio, China. Kanamycin, ampicillin, amino acids, adenine, histidine, leucine, tryptophan, and uracil were obtained from Sigma-Aldrich, USA. Chromatography-grade methanol and acetonitrile were obtained from EMD Chemicals, USA. Chromatography-grade formic acid and isopropanol were obtained from Thermofisher Scientific, USA. Authentic reference standards of apigenin, apigenin-7-O-glucuronide, scutellarin, and scutellarein were obtained from Yunnan Agricultural University. For preparation of flavonoids, the plant materials were ground in a mortar under liquid nitrogen. The resulting powder was dissolved in 80% methanol (W/V = 1:5), ultrasonicated for 30 min and the cell debris was removed by centrifugation at 13,000 × g for 30 min. The supernatant was used for HPLC analysis.

**Construction of apigenin producing yeast strain.** To construct an apigenin producing strain, the corresponding biosynthetic pathway was introduced into yeast *S. cerevisiae*. The *EBPAL*, *EB4CL*, *EBC4H*, *EBCHI*, and *EBFSII* coding sequences from *E. breviscapus* were synthesized by GENWIZ with codon optimization for *S. cerevisiae*. The M2S integration method was applied for protein expression. The main cultures were grown at 30 °C and 250 r.p.m. for 5 days and used for HPLC analysis.

**P450 gene screening method.** By analyzing the genome of *E. breviscapus*, 36 highly expressed P450 genes were found as *Fdh* candidates. First, the terminators (*TDH1* and *ICYC1*) with *BsaI* digestion sites were fused and integrated into the plasmid YCplac33, resulting in Y33-GP. Second, head-to-head promoters (p*TDH3-pADH1*) were cloned into the vector YCplac33 through Golden Gate cloning. The resulting plasmid harbored the expression cassette of P450 and EB-CPR. Then, the P450 gene in the vector was substituted by other candidates through Gibson assembly33. All the resulting P450 expression plasmids were individually introduced into SC1-FU for product identification. All primers used for P450 gene screening are listed in Supplementary Table 11.

Three colonies were picked for each genotype and used to inoculate 2 ml of SD medium (minus histidine, tryptophan, and uracil) in a 24-well plate. The cells were grown in a shaker at 30 °C and 800 r.p.m. for 48 h, after which the resulting seed cultures were transferred into fresh medium at a ratio of 1:30 and fermented under the same condition for 5 days. The products were analyzed by HPLC.

**Metabolic engineering for scutellarin production.** MLSI and CIT2, encoding enzymes for acetyl-CoA consumption, were selected for deletion in this study. For gene knockout, the *Ura3* selectable marker was first amplified from YCplac33 and flankled with 1,000 bp upstream and downstream homologous sequences of the target genes. The resulting fragment was integrated into the yeast genome by homologous recombination. Then a fragment with fused 1,000 bp of upstream and downstream homologous sequences was integrated into the yeast genome by homologous recombination to rescue the *Ura3* marker; 5-fluoroacetic acid32 was added to the SD plate to select for the loss of the *Ura3* marker. All primers used for yeast genomic knockouts are listed in Supplementary Table 12. ACS*$_{P641L}$, ALD6, and ADH2, encoding an acetyl-CoA synthetase pathway from ethanol, were overexpressed to enhance the acetyl-CoA supply for flavonoid production. ACS*$_{P641L}$ form *S. cerevisiae* was synthesized by GENWIZ, and ALD6 and ADH2 were amplified from the genome of W303-1B. ACS*$_{P641L}$ and ALD6 were cloned into the vector T4 through Golden Gate assembly, and ADH2 was analogously cloned into the vector T5. ACS*$_{P641L}$, ALD6, and ADH2 were amplified from *pTEF2, pPGK1*, and *pHXT7* respectively (primers: L4-F/L5-R, L5-F/L6-R). The expression cassettes of these genes were amplified with flanking sequences and co-transformed with a flanked *Leu2* (151stet-Leu2-L4 and L6-151atet2) gene into ΔMC-FU-FC, with genomic integration by the YW4135 strain, resulting in the strain ΔMC-FU-FC-AAA.

**EBf7GAT purification and enzyme assays.** The Eb7F7GAT coding gene was ligated into the expression vector pET-28a via NdeI and Xhol restriction sites into the plasmid pET28a-Eb7F7GAT, and the plasmid was transformed into *E. coli* BL21 (DE3). The engineered strain was grown to an OD$_{600}$ of 0.6 at 37 °C in LB medium supplemented with 100 μg/ml of kanamycin and then induced at 16 °C. The cDNA of IPTG (isopropyl β-1-thiogalactopyranoside) to a final concentration of 0.5 mM, and continued overnight. The cells were harvested by centrifugation at 13,000 × g for 30 min and stored at −80 °C until analysis.

The frozen cell paste was thawed on ice and suspended in 35 ml of isosm buffer (50 mM Tris-HCl, pH 7.5). The cells were disrupted on ice using a high-pressure homogenizer (NBIO, China), and the cell debris was removed by centrifugation at 13,000 × g for 20 min. To bind the recombinant enzyme, which was expressed as a fusion protein containing the 6-His tag, the supernatant was passed through a 0.45 μm filter and loaded onto a Ni$_2^+$-chelating affinity chromatography column (GE
were 5 using GraphPad Prism 5 (GraphPad Software, USA). tryptophan and grown for about 30 h until the OD600 reached 2.0. Cells were initiated by the addition of 1 0 min pre-incubation of the mixture without the enzyme at 37 °C, the reaction protein concentration was determined using a BCA Protein Assay Reagent Kit ( Pierce, USA) with BSA as the standard.

The standard reaction mixture (100 µl) consisted of 50 mM Tris-HCl pH 7.5, 100 µM glycyl acceptor, 1 mM UDPGA, and purified EB-F7GAT enzyme. After a 10 min incubation at 37 °C for 8 h, the reaction was stopped by the addition of 100 µl of methanol. The enzyme activity was assessed by measuring the generation of the corresponding glycosylation products via HPLC.

Enzyme kinetics with apigenin as substrate were determined in assays with 0.05 µg of purified recombinant protein, apigenin concentrations of 0.1–20 µM and a reaction time of 10 min; for scutellarein, the substrate was performed in enzyme assays with 0.5 µg of purified recombinant protein and scutellarein concentrations of 5–300 µM. Kinetic parameters were determined from triplicate experiments using GraphPad Prism 5 (GraphPad Software, USA).

Microsome isolation and enzyme assays. A single colony of W303-FC or W303-C was used to inoculate a 5 ml tube and incubated at 30 °C for 24 h. A 4 ml aliquot of the resulting culture was transferred to 200 ml of SD medium lacking tryptophan and grown for about 30 h until the OD600 reached 2.0. Cells were centrifuged for 10 min at 13,000 × g, and the cell debris was removed by centrifugation at 13,000 × g for 10 min. The supernatant was analyzed using a LC-20ADXR HPLC system (Shimadzu, Japan) equipped with a photodiode-array detector. Apigenin, scutellarein, and apigenin-7-O-glucuronide were measured at 335 nm using an amicon ultracentrifuge filtration using an amicon ultracentrifuge system.

Proportion of flavonoids inside and outside of cells. For the detection of flavonoids in the culture supernatant, the yeast cells were collected by centrifugation at 13,000 × g for 10 min and the supernatant was analyzed by HPLC. For the detection of flavonoids within the cells, the yeast cells were washed twice with double distilled water, disrupted on ice in 30 ml of 30% methanol using a high-pressure homogenizer (JNBIO), and the cell debris were removed by centrifugation at 13,000 × g for 30 min. The supernatant was collected and analyzed by HPLC. The detection of total flavonoids was conducted as mentioned above.

NMR analysis. For the isolation of scutellarein, 61 of fermentation broth was centrifuged and the cells were extracted with methanol three times. The resulting methanolic extract was collected and completely evaporated. The residue was dissolved in 50 ml methanol and further purified on an Agilent 1260 preparative HPLC with MWD detector (solvent A: 0.1% formic acid; solvent B: methanol; solvent A:B=45:55; flow rate 10 ml min−1). A reverse-phase C18 column (21.2 × 250 mm, 5 µm, Welch, China) was used for the preparation. H and 13C NMR spectra were obtained on a 400 MHz Bruker Avance III spectrometer in DMSO-d6. Chemical shifts (δ) were expressed in p.p.m. and coupling constants (J) in hertz (Hz). The results were listed as follows: H NMR (400 MHz, DMSO-d6): δ 12.81 (1 H, 5-0H), 10.16 (1 H, 7-0H), 7.92 (2 H, J = 8.9 Hz, H-2’, 6’), 6.93 (2 H, J = 8.8 Hz, H-3’, 5’), 6.75 (1 H, 1-H), 6.59 (1 H, 1-H), 13C NMR (100 MHz, DMSO-d6): δ 182.5 (C-2), 163.9 (C-2’), 161.4 (C-4), 153.8 (C-7), 153.0 (C-8a), 147.5 (C-5), 129.6 (C-6), 121.9 (C-1’), 116.4 (C-3’, 5’), 104.4 (C-4a), 102.7 (C-3’), 94.3 (C-8); ESI-MS: m/z 287.0550 [M + H]+.

Fermentation. The fermentation media used for this work were based on the previous study58. The trace metal solution contained: 5.75 g l−1 ZnSO4·7H2O, 0.32 g l−1 MnCl2·4H2O, 0.47 g l−1 CoCl2·6H2O, 0.48 g l−1 Na2MoO4·2H2O, 9.2 g l−1 CaCl2·2H2O, 2.8 g l−1 FeSO4·7H2O, and 80 ml l−1 0.5 M EDTA, pH 8.0. The vitamin solution contained: 0.05 g l−1 thiamine, 1 g l−1 calcium pantothenate, 1 g l−1 nicotinic acid, 25 g l−1 inositol, 1 g l−1 thiamine HCI, 1 g l−1 pyridoxal HCl, and 0.2 g l−1 p-aminobenzoic acid. The feed-batch fermentation medium contained: 20 g l−1 glucose, 15 g l−1 (NH4)2SO4, 8.5 g l−1 KH2PO4, 6.2 g l−1 MgSO4·7H2O, 12 ml l−1 vitamin solution, 10 ml l−1 trace metal solution, and 2.5 g l−1 adenine. The seed medium was used as the feed-batch fermentation medium. All the medium components were purchased from Solarbio. HCl and H2SO4 were purchased from Sinopharm, China.

Glycerol stocks of strain ΔMC-FU-FC-AAA were used to inoculate 40 ml of SD medium in a 250 ml shake flask and incubated at 30 °C under constant orbital shaking at 220 r.p.m. for 24 h. The entire culture volume was transferred into 400 ml of fresh seed medium and incubated at 30 °C and 220 r.p.m. for 36 h. The seed medium was used to inoculate 1 l of fermentation medium on a 250 l bioreactor (Baoxing, China) with a maximal working volume of 3 l. The OD600 of the culture was monitored and the reactors were inoculated to an initial OD600 of 1.2–1.5. The vitamin solution, the trace metal solution, and adenine were added into the autoclaved medium after filter sterilization. The fermentations were performed at 30 °C. The pH was maintained at 5.0 with automatic addition of ammonium hydroxide or 1 M H2SO4. The agitation rate was kept between 300 and 800 r.p.m., and the air flow was set to 2.1 min−1. The DOC was kept above 40%.

The feed was active until residual ethanol produced from the glucose phase was completely depleted. A quasi-exponent feed strategy was adopted as described by Nielsen et al.39. The feed contained: 386 g l−1 glucose, 9 g l−1 KH2PO4, 5.12 g l−1 MgSO4·7H2O, 3.5 g l−1 K2SO4, 0.28 g l−1 Na2SO4, 5 g l−1 adenine, 12 ml l−1 vitamin solution, and 10 ml l−1 trace metal solution. The vitamin solution, adenine, and trace metal solution were added into the feed solution in the same way as into the fermentation medium. Samples were taken at regular intervals to measure OD600, and their remaining volume was mixed with an equal volume of absolute methanol, ultrasonicated for 30 min, and centrifuged at 13,000 × g for 10 min. The supernatant was stored at −20 °C until HPLC analysis.

Data availability. The raw data that supports this study are available from the corresponding authors (Huifeng Jiang and Shengchao Yang) upon reasonable request.

Received: 10 July 2017 Accepted: 5 January 2018
Published online: 31 January 2018

References
1. Lan, M. Austroyunnanese Mater. Med. 2, 300-302 (1977).
2. Chinese Pharmacopoeia Commission. Pharmacopoeia of the People’s Republic of China. China Medical Science and Technology Press part 1, 138, 379, 848 (2014).
3. Prince, M. J. et al. The burden of disease in older people and implications for health policy and practice. Lancet 385, 549–562 (2015).
4. Ma, Y., Li, H. & Guan, S. Enhancement of the oral bioavailability of breviscipe by nanoemulsions drug delivery system. Drug. Dev. Ind. Pharm. 41, 177–182 (2015).
5. Jia, J. H. et al. Breviscipe, a traditional Chinese medicine, alleviates myocardial ischaemia repercussion injury in diabetic rats. Acta Cardiol. 63, 757 (2008).
6. Wang, M. et al. Studies on antioxidant activities of Breviscipe in the cell-free system. Am. J. Chin. Med. 36, 1199–1207 (2012).
7. Huang, Z. J., He, S. A. & Lei, B. Clinical analysis of acute cerebral infarction by Dengzhuhania injection and Xiongguin injection combined with Xuesatongtang treatment. China. Herbal. Med. 37, 1093 (2014).

ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-02883-z
8 NATURE COMMUNICATIONS naturecommunications www.nature.com/naturecommunications
16. Leonard, E., Yan, Y., Lim, K. H. & Koffas, M. A. Investigation of two distinct pathways of (+)-delta-cadinene-8-hydroxylase, a cytochrome P450 monoxygenase (CYP706B1) of cotton sesquiterpene biosynthesis. Plant J. Cell Mol. Biol. 28, 95 (2001).

17. Cankar, K. et al. Valencene oxidase CYP706M1 from Alaska cedar (Calitroptis nootkatensis). PLoS ONE 9, e115149 (2014).

18. Zhao, Q. et al. Two CYP82D11 Gene Functions as Flavone Hydrolases in the Biosynthesis of Root-Specific 4′-Deoxylavones in Scutellaria baicalensis. Molecular Plant 11, 135–148 (2017).

19. Chen, H., Wu, B., Nelson, D. R., Wu, K. & Liu, C. Computational identification and systematic classification of novel cytochrome P450 genes in Salvia miltiorrhiza. PLoS ONE 9, e115149 (2014).

20. Liu, X., Ding, W. & Jiang, H. Engineering microbial cell factories for the production of plant natural products: from design principles to industrial-scale production. Microb. Cell Fact. 16, 125–133 (2017).

21. Wu, J., Du, G., Chen, J. & Zhou, J. Enhancing flavonoid production by systematically tuning the central metabolic pathways: based on a CRISPR interference system in Escherichia coli. Sci. Rep. 5, 13477 (2015).

22. Lee, H., Kim, B. G., Kim, M. & Ahn, J. H. Biosynthesis of two flavones, Apigenin and Genkwanin, in Escherichia coli. J. Microbiol. Biotechnol. 25, 1442–1448 (2015).

23. Yang, J. et al. Hybrid de novo genome assembly of the Chinese herbal feedbale Erigeron breviscapus. Gigantia 6, 1–7 (2017).

24. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178 (2003).

25. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble aligned splice alignments. Genome Biol. 9, R7 (2008).

26. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriptsentsa, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212 (2015).

27. Punta, M. et al. The Pfam protein families database. Nucleic Acids Res. 40, D290–D301 (2012).

28. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W23–W31 (2011).

29. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32, W309–W312 (2004).

30. Chen, H., Wu, B., Nelson, D. R., Kai, W. & Chang, L. Computational identification and systematic classification of novel cytochrome P450 genes in Salvia miltiorrhiza. PLoS ONE 9, e115149 (2014).

31. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. Cell 56, 619 (1989).

32. Yao, L. et al. Strategies for enhancing resveratrol production and the expression of pathway enzymes. Appl. Microbiol. Biotechnol. 100, 7407 (2016).

33. Langle-Rouault, F. & Jacobs, E. A method for performing precise alterations in the yeast genome using a recyclable selectable marker. Nucleic Acids Res. 23, 3079 (1995).

34. Padddon, C. J. et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature Communications 9, 528–532 (2013).

35. Scalcinati, G. et al. Dynamic control of gene expression in Saccharomyces cerevisiae engineered for the production of plant sesquiterpene alpha-santalenin in a fed-batch mode. Metab. Eng. 14, 91–110 (2012).

Acknowledgements
We thank Professors Qinhong Wang and Yin Li from Tianjin Institute of Industrial Biotechnology for reading the manuscript and for their helpful comments. We thank Dr. Yibin Zhuang and Yi Cai from Tianjin Institute of Industrial Biotechnology for NMR analysis. This work was supported by grants from the 973 Program (2015CB557704), the National Natural Science Foundation of China (NSFC; Grant Number 31670100), the Key Research Program of the Chinese Academy of Science (Grant No. KFZD-SW-215) and The Hundred Talents Program of the Chinese Academy of Sciences to H.J., as well as grants by the NSFC (81560621), the Major Science and Technology Programs in Yunnan Province (2017ZF002), and the project of young and middle-aged talent of Yunnan province (2014HF011) to S.Y.

Author contributions
X.L., J. Cheng, and H.J. conceived, designed, and drafted the manuscript. X.L. designed and performed the experiments with the help from G.Z., W.D., L.D., X.C., and J.J. J. Cheng performed the bioinformatics analysis with the help from J.Y., L.K., W.F., J. Chen.,
G.L., and Y.Z. H.J., J. Cai, W.W., Y.M., Y.D., and S.Y. outlined the structure and reviewed the manuscript. All authors read and approved the final manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-02883-z.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018