

Pathway mining-based integration of critical enzyme parts for *de novo* biosynthesis of steviolglycosides sweetener in *Escherichia coli*

Cell Research (2016) 26:258-261. doi:10.1038/cr.2015.111; published online 11 September 2015

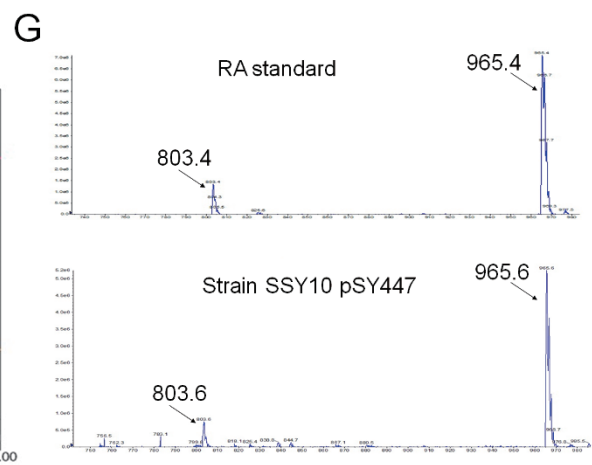
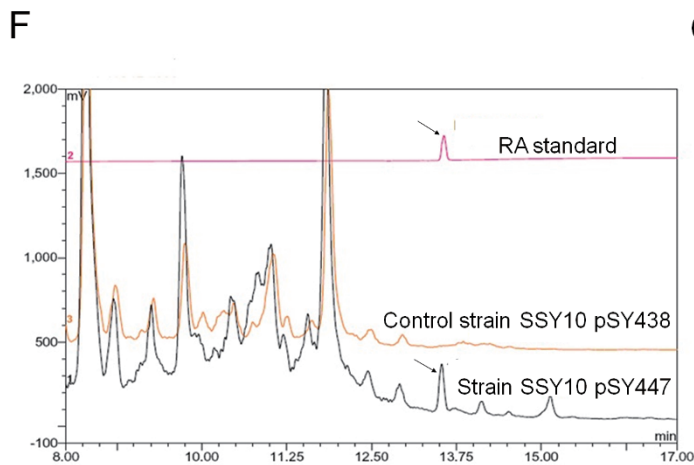
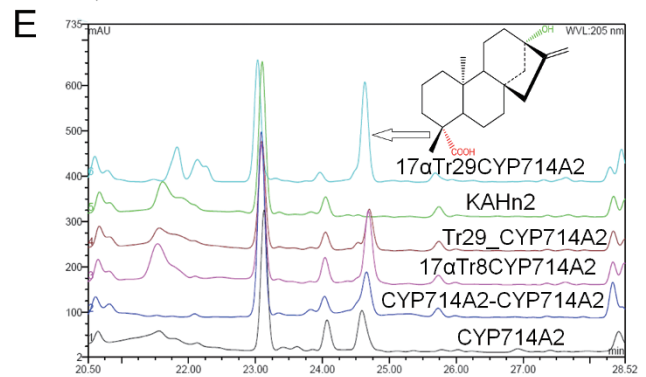
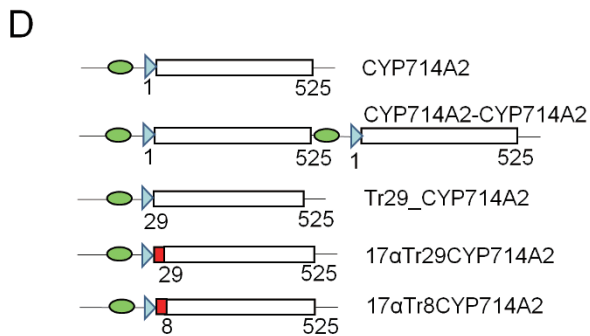
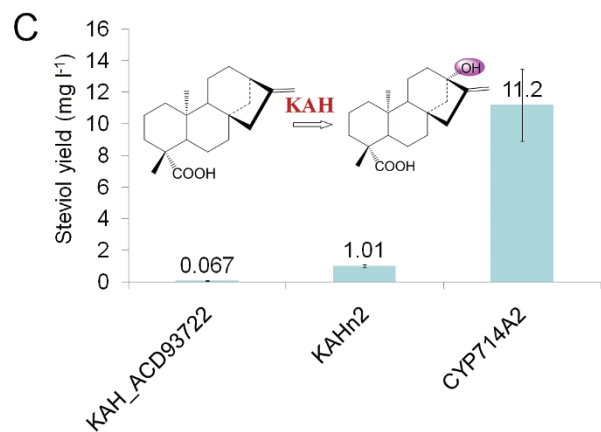
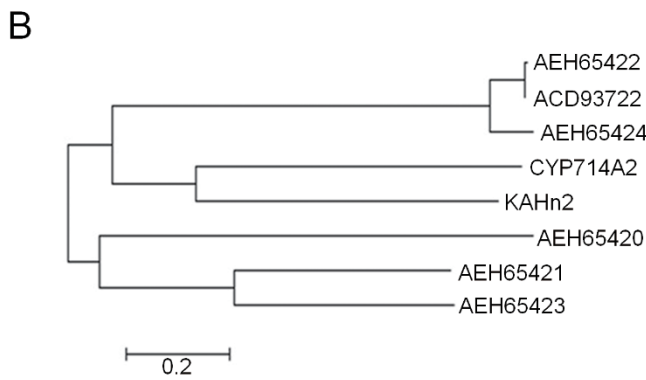
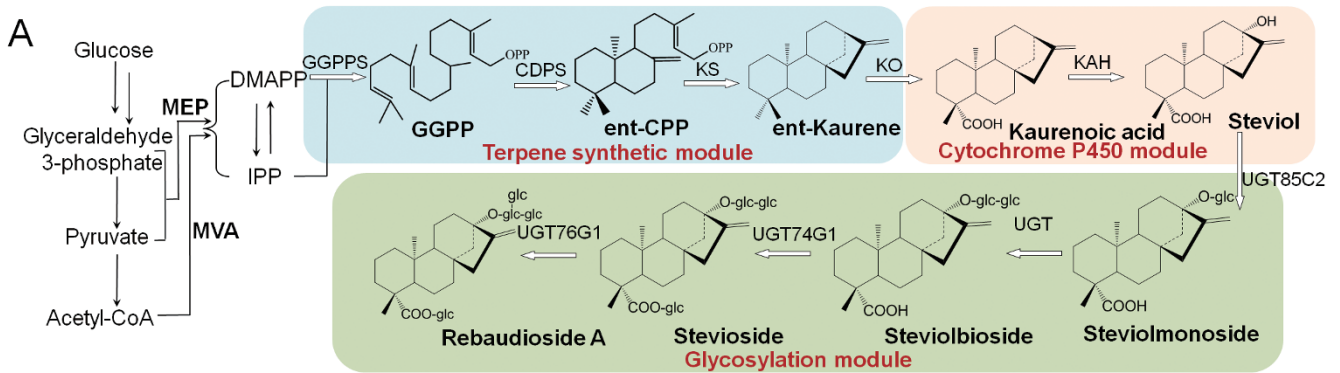
Dear Editor,

The development of microbial synthetic biology has been revolutionizing the industrial production of important plant terpenoids since the dawn of the century [1, 2]. By reconstructing the downstream biosynthetic pathways and systematically engineering the supply of common terpenoid precursors, the engineered microorganism chassis (e.g., *Escherichia coli* and *Saccharomyces cerevisiae*) can produce desired terpenoids in a more controllable manner. This strategy is also adapted to enrich the knowledge of cell metabolism via either discovery of novel or proof of postulated pathways [3].

Diterpenoid steviolglycosides (SGs) from *Stevia rebaudiana* are intense natural sweeteners with various health benefits. To date, all of the commercialized SGs have been manufactured by *Stevia* extraction in which, at least 30 compounds of this class with different glycosylation patterns have already been characterized, including the main components, steviosides and rebaudioside A, and many trace components [4, 5]. Various SGs are produced naturally in different quantities and their glycosylation patterns substantially influence their tastes. The production of SGs with a consistent taste profile remains a major challenge. Much effort has already been made to elucidate the biosynthetic pathway of SGs [5]. Taking RA, the major component of SGs, as an example, its biosynthesis in *Stevia* plants was proposed and largely characterized to involve nine enzyme-catalyzed reactions from isopentenyl diphosphate/dimethylallyl diphosphate [4, 5], which can be divided into three metabolic modules (Figure 1A). Although limited success in constructing yeast cells for *de novo* biosynthesis of SGs has been claimed in some patents (e.g., see reference [6]), we report here the efforts in reconstruction of the system in *E. coli* cells for systematic characterization of critical enzymes with proper activities in this tractable microbial chassis taking the advantage of its efficient manipulation technology and higher theoretical maximum yield of

isoprene from glucose for heterologous biosynthesis of terpenoids [1, 2].

The 13 α -hydroxylation of kaurenoic acid to generate steviol is the first divergent step specific to the SGs pathway. Although the critical kaurenoic acid 13 α -hydroxylase (KAH) was reported to be partially purified and its N-terminal sequence was described, efforts in cloning its encoding cDNA have been unsuccessful [4, 5]. On the other hand, sequences of several potential KAHs from *S. rebaudiana* have been deposited in Genbank database, including KAH_ACD93722, KAH_AEH65420, KAH_AEH65421, KAH_AEH65422, KAH_AEH65423, and KAH_AEH65424, without any functional characterization [5]. To comprehensively survey the genes associated with SG biosynthesis, RNA-seq was performed for two samples of *Stevia* leaves, sr-1 (30-day-old leaves with a low SG content) and sr-2 (90-day-old leaves with a high SG content). All these potential KAHs mentioned above were detected in our *Stevia* RNA-seq data. Phylogenetic analysis showed that KAH_ACD93722, KAH_AEH65422, and KAH_AEH65424 belong to the CYP716 family, KAH_AEH65420 belongs to the CYP701 family, and KAH_AEH65421 and KAH_AEH65423 belong to the CYP82 family (Figure 1B). When the resting cells of recombinant *E. coli* BL21(DE3) co-expressing various KAH and its *NADPH*-dependent cytochrome P450 reductase (CPR) were used to transform the substrate kaurenoic acid, only KAH_ACD93722 exhibited low activity, producing 0.067 mg l⁻¹ steviol (Figure 1B and 1C, Supplementary information, Figure S1). Recently, a CYP714 family enzyme CYP714A2, which contributes to the production of diverse GA compounds through various oxidations of C and D rings of the ent-kaurene scaffold in *Arabidopsis thaliana*, was reported to bear a KAH activity when expressed in yeast [7]. We thus screened for sequences similar to CYP714A2 in our RNA-seq database, and identified a novel KAH belonging to the CYP714 family designated as KAHn2, showing 40% identity with that of CYP714A2 (Figure 1B). KAHn2



was nearly 15-fold more efficient than KAH_ACD93722 and the steviol yield reached 1.01 mg l⁻¹ (Figure 1C). Interestingly, contrary to other SG pathway genes (Supplementary information, Figure S2), the transcription of KAHn2 in mature leaves was downregulated rather than upregulated in our RNA-seq data.

Steviol is further converted to different kinds of SGs by four consecutive glycosylation steps (Figure 1A). In *S. rebaudiana*, more than a dozen of UDP-glycosyltransferases (UGTs) have been detected, yet only three have so far been clearly shown to contribute to the biosynthesis of steviol glycosides. The second UGT (referred to as steviol-13-monoglucoside-1,2-glycosyltransferase) has so far eluded functional characterization, although a candidate has been suggested (UGT91D2) [5, 6]. Our RNA-seq data (Supplementary information, Figure S3) revealed 35 UGTs differentially expressed in *Stevia* leaves collected in two different developmental stages. Among the 23 upregulated UGTs, because 10 of them were known to have no desired activity [8], the rest 13 unknown UGTs were tested. Full length cDNAs of 9 genes were isolated and expressed successfully in *E. coli*. However, no desired activity was detected. Using UGT91D2 from *S. rebaudiana* [5, 6] as the reference sequence, cDNA clones of UGT91D2 from a different cultivar of *S. rebaudiana* (the cultivar was planted in Shandong province, which has been used to extract SGs sweetener for export) were obtained by PCR amplification. Sequencing analyses of cDNA clones amplified from five individual plants identified three UGT91D2 cDNA variants, with small variations in the encoded polypeptides (designated as UGT91D2_#1, UGT91D2_#2, and UGT91D2_#3). We expressed these three protein-coding sequences in *E. coli* BL21 (DE3) to test their activity. By adding steviolmonoside and UDP-glucose as substrates, the end product steviolbioside was successfully detected. The activity of UGT91D2_#2 was apparently the highest. The

amino-acid sequence of UGT91D2_#2 (designated as UGT91D2w in the following section), which is the same as that of UGT91D2e_No.5 [6], contains two variations (V241I and T444A) compared with that of UGT91D2. It is a bit surprising to note that the UGT91D2 cDNA was not detected among the differentially expressed UGTs in our RNA-seq analysis of *S. rebaudiana* leaves collected from two different growing stages.

With these two new parts available, the SGs *de novo* biosynthetic pathway was reconstructed in *E. coli* based on a heterologous ent-kaurene pathway (plasmid pZQ03) previously assembled [9] and a combinatorial mevalonic acid pathway consisting of two expression cassettes introduced into *E. coli* (BL21(DE3) pXL17/pXL13 (Supplementary information, Figures S4 and S5), designated as strain SSY10) to improve isoprenoid precursor supply. Under optimized conditions of IPTG (0.02 mM) and fermentation temperature (30 °C), the maximum yield of ent-kaurene was improved from 2.16 mg l⁻¹ to 194.12 mg l⁻¹ (in shake flask, Supplementary information, Figure S6) and 1.872 g l⁻¹ (in 5-l bioreactor, Supplementary information, Figure S7).

Of the cytochrome P450 module, ent-kaurene oxidase (KO) and KAH are responsible for the sequential C19 oxidation and C13 hydroxylation of ent-kaurene. The KO from *S. rebaudiana* (KO_Sr) was incorporated into plasmid pSY400 along with its electron transfer partner CPR_Sr from *S. rebaudiana*. Although functionally expressing plant membrane-bound CYP enzymes in the prokaryotic host *E. coli* is usually a great challenge, the strain SSY10 pSY414 produced 42.49 mg l⁻¹ kaurenoic acid at 30 °C with 0.02 mM IPTG added (Supplementary information, Figure S8). To enhance the functional expression of KO_Sr, we further optimized the fermentation temperature and IPTG concentration. Encouragingly, the highest yield of kaurenoic acid (22 °C and 0.1 mM IPTG) reached 100.23 mg l⁻¹. Subsequently, the

Figure 1 Functional characterization of unknown SG biosynthetic enzymes and the stepwise construction of *de novo* SG biosynthesis pathway in *E. coli*. **(A)** The biosynthetic pathway of RA, the main component of SGs produced in *Stevia rebaudiana*. The proposed pathway contains three metabolic modules including: the terpene module, consisting of geranylgeranyl diphosphatase (GGPPs, EC:2.5.1.29), ent-copalyl diphosphate synthase (CDPS, EC:5.5.1.13), and ent-kaurene synthase (KS, EC:4.2.3.19); the cytochrome P450 (CYP) module, consisting of ent-kaurene oxidase (KO, EC:1.14.13.78) and kaurenoic acid 13 α -hydroxylase (KAH); and the UDP-glycosylation (UGT) module, consisting of four UDP-glycosyltransferases UGT85C2 (EC 2.4.1.17), an unknown UGT, UGT74G1 (EC 2.4.1.17), and UGT76G1 (EC 2.4.1.17). **(B)** The phylogenetic tree of KAHs and related P450 enzymes from *Stevia* and other plants. The potential KAHs can be grouped into different CYP families: KAH_ACD93722, KAH_AEH65422, and KAH_AEH65424 belonging to the CYP716 family; KAH_AEH65420 belonging to the CYP701 family; KAH_AEH65421 and KAH_AEH65423 belonging to the CYP82 family; and KAHn2 and CYP714A2 belonging to the CYP714 family. **(C)** Activity of different KAHs analyzed by resting cell experiment. Among the reported potential KAHs, only KAH_ACD93722 generated 0.067 mg l⁻¹ steviol. The novel KAHn2 and *Arabidopsis thaliana*-derived CYP714A2 were shown to be more active KAH here. **(D)** Strategies for CYP714A2 engineering. Ellipse, ribosomal-binding site; triangle, translation starting amino acid "M"; white rectangle, amino-acid residues of native CYP714A2; red rectangle, the eight-residue peptide "MALLLAVI" of the N terminus of bovine 17 α -hydroxylase. **(E)** HPLC analysis of steviol production using different or engineered KAHs. **(F)** HPLC analysis of RA production in engineered strains. **(G)** MS profiles of standard RA or RA produced by engineered strains.

KAH enzyme was further incorporated to construct the steviol pathway (strain SSY10 pSY426). Although only trace amounts of steviol could be detected by employing KAH_ACD93722, with the more active KAHn2 (strain SSY10 pSY427), steviol yield reached 0.605 mg l⁻¹ with the kaurenoic acid accumulated to 85.4 mg l⁻¹. Further engineering the N terminus of KAHn2 by truncation or substitution with the N terminus of bovine 17 α -hydroxylase did not relieve this bottleneck. The heterologous CYP714A2 from *A. thaliana* was then tested to replace *S. rebaudiana*-derived KAHs. Steviol yield in strain SSY10 pSY429 with CYP714A2 reached 9.47 mg l⁻¹, which represented a 15.7-fold improvement compared with that of KAHn2. However, the level of kaurenoic acid in this strain still reached 78.52 mg l⁻¹. More active KAH was explored through various engineering methodologies. When we engineered the N terminus of CYP714A2, the construct 17atr29CYP714A2 produced 15.47 mg l⁻¹ of steviol (strain SSY10 pSY438; Figure 1D and 1E), which was improved by 63% compared with that of native CYP714A2.

Afterwards, the UGT module UGT85C2/UGT91D2w/UGT74G1/UGT76G1 (strain SSY10 pSY447; Supplementary information, Figure S9) was further assembled and incorporated into the highest steviol producer with 17atr29CYP714A2 to achieve the complete biosynthesis of SGs in *E. coli*. After 5-day cultivation, the components of SGs were detected by LC-MS in the fermentation broth. Yield of the major component RA reached 10.03 mg l⁻¹ (Figure 1F and 1G).

In summary, we identified a novel KAH KAHn2 from *S. rebaudiana* based on mining of the RNA-seq data and characterized KAH_ACD93722 with the activity of 13 α -hydroxylase and a UDP-glucosyltransferase UGT91D2w with the activity of steviol-13-monoglucoside-1,2-glucosyltransferase in *E. coli*. We have successfully established a novel pathway and constructed an *E.*

coli strain for the production of SGs. This study has laid a solid foundation for developing *E. coli* as a host for heterologous production of SGs, and provided an example that *E. coli* can also serve as a promising chassis for heterologous biosynthesis of complex terpenoids highly modified by P450 enzymes.

Acknowledgments

This work was supported by the National Basic Research Program of China (“973” Program, 2012CB721104), the National High Technology Research and Development Program (“863” Program, 2012AA02A701), and the Key Projects of Shanghai Committee of Science and Technology (14JC1406900).

Jianfeng Wang¹, Shiyuan Li¹, Zhiqiang Xiong¹,
Yong Wang¹

¹Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

Correspondence: Yong Wang

Tel/Fax: 86-21-54924295

E-mail: yongwang@sibs.ac.cn

References

- 1 Carothers JM, Goler JA, Keasling JD. *Curr Opin Biotechnol* 2009; **20**:498-503.
- 2 Julleson D, David F, Pflieger B, *et al.* *Biotechnol Adv* 2015; **33**:1395-1402.
- 3 Yan X, Fan Y, Wei W, *et al.* *Cell Res* 2014; **24**:770-773.
- 4 Brandle JE, Telmer PG. *Phytochemistry* 2007; **68**:1855-1863.
- 5 Ceunen S, Geuns JM. *J Nat Prod* 2013; **76**:1201-1228.
- 6 Kishore GM, Motion M, Hicks PM, *et al.* PCT Pat. Appl. WO 2011/153378, 2011.
- 7 Nomura T, Magome H, Hanada A, *et al.* *Plant Cell Physiol* 2013; **54**:1837-1851.
- 8 Richman A, Swanson A, Humphrey T, *et al.* *Plant J.* 2005; **41**:56-67.
- 9 Wang JF, Xiong ZQ, Li SY, *et al.* *Appl Microbiol Biotechnol* 2013; **97**:8057-8067.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)