Targeted enzyme gene re-positioning: A computational approach for discovering alternative bacterial enzymes for the synthesis of plant-specific secondary metabolites

Yuya Nakamura a, Shuichi Hirose b, Yuko Taniguchi b, Yuki Moriya c, Takuji Yamada a,d,e,*

a School of Life Science and Technology, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo, 152-8550, Japan
b NAGASE R&D Center, Nagase & Co., Ltd, Kobe High Tech Park 2-2-3 Marukami, Nishi-ku, Kobe, Hyogo, 651-2241, Japan
c Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Kashiwa, 277-0871, Japan
d PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama, 332-0012, Japan
e Metabologenomics Inc, 246-2 Kakuganji, Tsuruoka, Yamagata, 997-0052, Japan

ARTICLE INFO

ABSTRACT

Plant-biosynthesised secondary metabolites are unique sources of pharmaceuticals, food additives, and flavourings, among other industrial uses. However, industrial production of these metabolites is difficult because of their structural complexity, dangerousness and unfriendliness to natural environment, so the development of new methods to synthesise them is required. In this study, we developed a novel approach to identifying alternative bacterial enzyme to produce plant-biosynthesised secondary metabolites. Based on the similarity of enzymatic reactions, we searched for candidate bacterial genes encoding enzymes that could potentially replace the enzymes in plant-specific secondary metabolism reactions that are contained in the KEGG database (enzyme re-positioning). As a result, we discovered candidate bacterial alternative enzyme genes for 447 plant-specific secondary metabolic reaction. To validate our approach, we focused on the ability of an enzyme from Streptomyces coelicolor strain A3(2) strain to convert valencene to the grapefruit metabolite nootkatone, and confirmed its enzymatic activity by gas chromatography-mass spectrometry. This enzyme re-positioning approach may offer an entirely new way of screening enzymes that cannot be achieved by most of other conventional methods, and it is applicable to various other metabolites and may enable microbial production of compounds that are currently difficult to produce industrially.

1. Introduction

Plants biosynthesis leads to the production of various compounds known as secondary metabolites, which are not required for plant growth, development, or reproduction, but are important for human daily life (Bourgaud et al., 2001). For example, rose, vanilla and star anise extracts are used for perfumes and medicines. However, these plant-specific secondary metabolites are often expensive because of the rarity of the compounds within the plants or the plants themselves. Some of these compounds are produced industrially, with direct extraction from the plants being the most common approach (Srivastava and Srivastava, 2007). However, there is the problem of resource exhaustion with this strategy, particularly for rare materials. Indeed, Ashtawarga plants as an Ayurvedic and allopatic medicine in India have been included among 560 plants appearing in the red list of endangered species (Virk et al., 2015). In addition, it can be difficult to obtain a stable supply due to climate change, and a simulative research has reported that the impact of climate change on the agricultural markets would amount to about one-sixth of total crop value (Costinot et al., 2016).

To overcome the challenges of extracting metabolites from plants, several means of synthesising metabolites artificially have been developed. Chemical synthesis is popular methods of stably obtaining metabolites that do not involve extraction from plants. Nonetheless, organic synthesis is not always readily possible, because plant secondary metabolites have complex structures and large molecular sizes. In such cases, organic synthesis requires many steps to obtain target metabolites from simple substrates, constituting a great hurdle for industrial large-scale synthesis of plant-specific secondary metabolites (Bruggink et al., 2003; Laine, 2017). On the other hand, microbial fermentation which is another method for synthesising metabolites enables the supply of stable...
plant-specific secondary metabolites without an environmental impact (Daniell et al., 2002; Velasco et al., 2000).

Nootkatone is one of the valuable plant-specific secondary metabolite and a factor for the characteristic flavor of grapefruit. In nootkatone production, organic synthesis is a general method, however it involves eight reaction steps (Laine, 2017) and requires the use of t-BuOOH, which is a dangerous substance (Julien and Wallace, 2008), or the massive use of heavy metals (MnO2), which has an environmental impact (Serra, 2016). Thus, a method was developed for producing nootkatone from valencene or hydroxyvalencene, which is inexpensive and easily available, using Chloroflexus (algae) (Asakawa et al., 2003). Furthermore, methanolysis that employs laccase derived from microbes (Huang et al., 2001) and metabolism by available, using E-zyme2, a web-based tool for searching for enzyme sequences. In other words, we aimed to search for enzymes with similar enzymes based on similarity of enzyme reactions and not on gene sequence similarity to identify microbial genes homologous to plant genes from a set of reaction substrates and products using enzymatic reaction similarity that is generally used to search for orphan enzymes with a known reaction but unknown gene (Moriya et al., 2016). To identify candidate bacterial enzymes for nootkatone biosynthesis, we used KEGG OC database version 2016-01-16.

2.2. Bacterial strains and culturing conditions

Streptomyces lividans TK64 (NBRC 15678), which was purchased from National Institute of Technology and Evaluation (NITE, Chiba, Japan), was used as the host organism for bioconversion. S. lividans TK64 was cultured in TSB medium (17 g/L pancreatic digest of casein, 3 g/L papain digestive meal, 2.5 g/L glucose, 5 g/L sodium chloride, and 2.5 g/L dipotassium phosphate, BD® Tryptic Soy Broth (Catalogue no. 211825); BD, Heidelberg, Germany) supplemented with 50 μg/mL of thiostrepton at 28°C and 200 rpm for 72 h.

Escherichia coli JM109 (Takara Bio, Shiga, Japan) was used as the host organism for constructing plasmid clones. Escherichia coli JM109 was cultured in LB medium (Nacalai tesque, Kyoto, Japan) at 37°C for 16 h.

2.3. Plasmid construction

For constructing expression vectors, four genes derived from Streptomyces coelicolor A3(2) (GenBank No. NC 003888), SCO5223, SCO3099, SCO3770 and SCO3636 were cloned. The genomic DNA of S. coelicolor A3(2) was acquired by phenol-chloroform extraction. The coding sequences of four genes were amplified with primers (Supplementary Table 1) using the S. coelicolor A3(2) genomic DNA as a template. PCR was performed under the following conditions: 94°C for 2 min, followed by 30 cycles of 98°C for 10 s and 68°C for 1.5 min (KOD-Plus-Neo; Toyobo, Osaka, Japan). The PCR products were purified and digested with SpH1, except for the SCO3636 PCR product, which was digested with EcoT22I. plPJ702D, a derivative of plPJ702SSMP from which melC1 and melC2 have been removed (Supplementary Fig. 1A), was also digested with Stul and SpH1 (or EcoT22I for SCO3636) (Comalada et al., 2005). The PCR products were ligated into the digested vector to produce the expression plasmids. The ligation reaction was incubated with Ligation Mix (Takara Bio, Shiga, Japan) at 16°C for 1 h.

To construct the SCO3770 partially-knockout vector, nucleotides 97 to 1090 of SCO3770, which encode an inactive form of the enzyme, were amplified with primers (Supplementary Table 2) using the S. coelicolor A3(2) genome as a template. PCR was performed under the following conditions: 98°C for 30 s, followed by 30 cycles of 98°C for 30 s, 55°C for 30 s, and 68°C for 1 min (PrimeSTAR® GXL DNA Polymerase; Takara Bio, Shiga, Japan). The PCR product was purified and digested with EcoRI and KpnI. pk18mcE, a derivative of pk18mobs carrying melC1 and melC2 was digested with the same restriction enzyme, and the fragments were ligated together (Supplementary Fig. 1B). The ligation condition was same as that in expression vectors.

2.4. Transformation

The p450 gene expression vectors were introduced into S. lividans TK64 via protoplast transformation according to Keiser’s method (Dang et al., 2018). Transformants were selected based on thiostrepton resistance.

Similarly, SCO3770 was knocked out in S. lividans TK64 by introducing the plPJ702D·ΔSCO3770 vector. Transformants were selected based on kanamycin resistance. To construct the complemented strain, protoplasts from S. lividans TK64 ΔSCO3770 were transformed with plPJ702D·SCO3770 as above.

2.5. Conversion of valencene into nootkatone by Streptomyces

Single colonies of transformants were pre-cultured in 5 mL TSB medium supplemented with 50 μg/mL of thiostrepton at 28°C and 200 rpm
for 72 h. A 1.5-mL aliquot of this pre-culture was inoculated into a baffled flask containing 50 mL SSMP medium (5% glucose, 0.8% K2HPO4, 0.05% MgSO4·7H2O, 0.5% yeast extract, 0.5% polypeptone) supplemented with 10 μg/mL of thiostrepton and incubated at 28 °C and 160 rpm. After 72 h, glucose and valencene (Catalogue no. 75056-10-G-F; SIGMA-ALDRICH, Missouri, USA) were added at final concentrations of 10% and 0.025%, respectively. Cultivation was then continued for another 168 h.

2.6. Confirmation of nootkatone production

Following culture of the P450-overexpressing strains, cells were harvested from 2 mL whole broth by centrifugation at 14,000×g for 20 min. The cell pellet was resuspended in 1.5 mL ethanol, and 0.25 mL distilled water and 0.5 mL n-hexane were added to 1.0 mL of the supernatant and mixed by shaking. The n-hexane layer was collected by centrifugation at 14,000×g for 1 min. Nootkatone in the supernatant was added to 1.0 mL of the supernatant and mixed by shaking. The n-hexane layer was collected by centrifugation at 14,000×g for 1 min. Nootkatone in the extract was analysed by GC/MS (GC-2010 and GCMS-QP2010 Plus, Shimadzu; Kyoto, Japan). The peak of nootkatone was at approximately 12.0 min. Measurements were performed under the conditions indicated in Supplemental Table 2. Purified purchased nootkatone (Catalogue no. N0920; Tokyo Chemical Industry, Tokyo, Japan) was used for nootkatone standard.

2.7. Quantification of nootkatone

Four strains (S. lividans TK64, S. lividans TK64/pIJ702DSCO3770, S. lividans TK64 ΔSCO3770, and S. lividans TK64 ΔSCO3770/pIJ702DSCO3770) were cultured, and cells were harvested from 2 mL whole broth by centrifugation at 14,000×g for 20 min. Next, 0.5 mL n-hexane was added to 1.0 mL of the supernatant and mixed by shaking, after which the n-hexane layer was separated by centrifugation at 14,000×g for 1 min. The cell pellet was resuspended in 1.5 mL ethanol, and 0.25 mL distilled water and 0.5 mL n-hexane were added to 1.0 mL of the supernatant and mixed by shaking. Nootkatone in both extracts was analysed by GC/MS (GCMS-QP2010 Plus; Shimadzu, Kyoto, Japan) under the conditions indicated in Supplemental Table 3. The amount of nootkatone was estimated based on the standard curve for β-carophyllene (Catalogue no. C0796; Tokyo Chemical Industry, Tokyo, Japan).

2.8. Data analysis

We used TogoWS (Katayama et al., 2010) and KEGG API (Kanehisa et al., 2019) to extract data from KEGG and format the data. Python was used for summarising and visualising the enzyme search results. iTol4 (Letunic and Bork, 2016) was used to visualise the phylogenetic distribution of bacterial enzymes that could potentially replace plant-specific enzymes.

The amino acid sequence of the nootkatone biosynthetic enzyme was obtained from the UniProt database (Bateman et al., 2017). Homology searches of SCO3770 and nootkatone biosynthetic enzymes were performed using BLAST (Altschul et al., 1990) with default options. COBALT (Papadopoulos and Agarwala, 2007) was employed for the multiple alignment.

3. Results

3.1. Identification of candidate bacterial enzyme genes for plant-specific secondary metabolite biosynthesis

To determine whether bacterial enzymes could be identified that perform similar functions to plant enzymes involved in secondary metabolite production, we first obtained 880 substrate/product compound pairs of plant-specific metabolic reactions from KEGG database (Kanehisa et al., 2019), which is an integrated database of genes, metabolic pathways and metabolites. Next, we used the 880 substrate/product pairs as a query for E-zyme2 (Moriya et al., 2016) to search for bacterial enzyme genes that appeared to replace the plant enzyme. E-zyme2 searches for “similar reactions” in KEGG using a substrate/product compound pair as the query and returns orthologue gene candidates and reaction similarity scores. The “similar reactions” does not mean that a substrate/product pair of query is chemically and structurally similar to substrate/product pairs of predicted enzyme but that the patterns of chemical changes occurring during the reactions (referred to as “RDM” patterns) is similar (Moriya et al., 2016). This tool enabled us to search for bacterial alternative enzyme genes for plant-specific metabolic reactions based on reaction similarity and not gene sequence similarity. E-zyme2 identified 447 of 880 plant-specific metabolic reactions as a reaction that might be catalysed by bacterial enzymes, with a similarity score >0.4 (Fig. 1A, Supplementary Table 4). Of these 447 reactions, 41 reactions involved orphan enzymes in plant (Supplementary Table 5). In addition, 19 reactions of these belong to flavonoid-related metabolic pathways.

Metabolic reactions in KEGG have a functional hierarchical structure consisting of reactions, pathway and biological functional category (Kanehisa et al., 2019; Uchihama et al., 2015). To confirm the multi-functional ability of bacterial enzymes, we projected target plant reactions and the native reactions of the detected bacterial enzymes by extrapolating the plant reactions onto the KEGG functional hierarchy. As a result, bacterial alternative enzyme candidates for plant reactions were different from the query plant reactions (Fig. 1B).

To identify bacterial strains that contain a high proportion of candidate enzyme genes for producing plant secondary metabolites, we aggregated the number of reactions for each bacterial species and visualised the phylogenetic distribution using iTol4 (Letunic and Bork, 2016) (Supplementary Table 6, Fig. 1C). As shown, there was bias in the number of alternative gene candidates among taxa, with bacteria belonging to Actinobacteria and Proteobacteria possessing the highest number of alternative gene candidates that might replace genes involved in plant-specific metabolism. In particular, Streptomycetaceae, Pseudonocardia, Mycobacteriaceae, Myxococcales and Burkholderiaceae harbours more alternative gene candidates compared to the other bacteria. Of these bacteria, the Streptomyces coelicolor A3(2), which is a representative Streptomyces species and generally used for enzyme screening, exhibited 365 reactions with a similarity score >0.4, and 43 reactions with a similarity score ≥0.9 (Fig. 1D). Therefore, we focused on sco for subsequent experiments and analysis.

3.2. A Streptomyces coelicolor A3(2) gene product metabolises valencene to nootkatone

To validate the ability of our enzyme re-positioning approach to accurately predict bacterial gene products that can substitute for plant-specific enzymes, we assayed enzymatic activity for S. coelicolor gene products predicted to catalyse synthesis of nootkatone from valencene. The reason for choosing this reaction was that the product, nootkatone, is a valuable compound in terms of its medicinal and agricultural properties. Therefore, to confirm that these four genes was discovered based on the sequence independence of our approach, we performed a multiple alignment of these four genes by BLAST and the three nootkatone synthetases reported previously (Cankar et al., 2011; Fraatz et al.,...
For the eukaryotic-derived genes E1B2Z9 (Cichorium intybus) and B8ZIU7 (Pleurotus sapidus), the total scores (BLAST bit score) compared to four genes were low, less than 54.6 and identity was also low, less than 30%. For the prokaryotic-derived gene U5U1Z3 (Bacillus subtilis), the BLAST bit score was 99–229, and some homology with four genes was observed, even though identity was only 24–36%. Thus, it could never be considered that these scores were top hits for sequence similarity based screening.

To assess the activity of these enzymes, we overexpressed them in S. lividans TK64 using p450 overexpression vectors. S. lividans TK64 was a model species and a closely-related to S. coelicolor, and commonly used in genetic engineering (Comalada et al., 2005; Miyashita et al., 1991; 2009; Girhard et al., 2009) (Table 2). For the eukaryotic-derived genes E1B2Z9 (Cichorium intybus) and B8ZIU7 (Pleurotus sapidus), the total scores (BLAST bit score) compared to four genes were low, less than 54.6 and identity was also low, less than 30%. For the prokaryotic-derived gene U5U1Z3 (Bacillus subtilis), the BLAST bit score was 99–229, and some homology with four genes was observed, even though identity was only 24–36%. Thus, it could never be considered that these scores were top hits for sequence similarity based screening.
by the standard nootkatone was observed in the spectrometry (GC-MS). A peak of nootkatone equal to a peak identified by gas chromatography-mass spectrometry (GC-MS) was detected by the TK64 strain effectively metabolized nootkatone from valencene. TK64 strain metabolized nootkatone at a 1:5 ratio. Nootkatone is formed from valencene, and TK64 strain is capable of metabolizing nootkatone. This result shows the involvement of bacterial enzyme re-positioning for the production of nootkatone. SCO3770 had low sequence homology with existing enzymes. This result showed that our bacterial enzyme re-positioning approach has a novel enzyme screening space which is not considered by sequence based similarity approach, and it may enable us to produce compounds that are currently difficult to produce industrially by microbial fermentation.

Our bacterial enzyme gene re-positioning approach detected candidate bacterial alternative enzymes for 447 plant-specific secondary metabolism reactions, and 41 of 447 (9.2%) of the plant-specific secondary metabolism reactions for which re-positioning candidates were identified involved "orphan enzymes", whose reactions are reported but the genes are not discovered (Yamada et al., 2012). Although we call this method "enzyme re-positioning", we can identify genes that encode alternative enzymes for target reactions, even if the target enzymes are orphan enzymes. It is pretty obvious, but sequence based enzyme gene screening required the gene sequence. However, E-zyme2 was originally designed to be used as a tool for orphan enzyme discovery (Moriya et al., 2016). Thus, it can also be used to search for candidate enzymes using only a substrate and product pair.

Of the plant-specific orphan enzymatic reactions for which we found alternative enzymes in this study, about half of the 19 reactions belonged to flavone-related metabolic pathways. Flavone-related compounds reportedly have anti-tumour (Cincin et al., 2014; Dória et al., 2016; Martino et al., 2016; Sung et al., 2016; Yan et al., 2017) and anti-inflammatory (Comalada et al., 2005; Dang et al., 2018) effects and high medical usefulness. In particular, apigenin has been reported to function in cancer prevention by inducing apoptosis and autophagy (Sung et al., 2016), and it is expected to act as an anticancer agent (Yan et al., 2017). If alternative enzymes enable inexpensive production of these types of compounds, this would be of great benefit in the research and development of new anti-tumour drugs.

Our result of phylogenetic distribution analysis for discovered candidate bacterial alternative enzymes indicated that the bacterial enzyme candidate genes clustered into a few taxa. The bacterial enzyme candidates were phylogenetically biased, with Pseudonocardiaceae, Streptomyctecaceae, Mycobacteriaceae, Myxococcales and Burkholderiaceae possessing more alternative enzyme candidates than other taxa. Since Streptomyctecae and Burkholderiaceae have particularly many secondary metabolic functions among all of bacteria, it is possible that such bacterial secondary metabolic related enzymes have been detected as alternative candidates for plant secondary metabolism. Moreover, all these bacteria live in soil (Kim et al., 2015; Walters et al., 2018), and it is expected to act as an anticancer agent (Yan et al., 2017). If alternative enzymes enable inexpensive production of these types of compounds, this would be of great benefit in the research and development of new anti-tumour drugs.
catalysed by the enzymes. the reactions.

5. Conclusions

Taken together, we developed a novel method, bacterial enzyme gene re-positioning, to efficiently screen bacterial alternative enzymes for plant-specific secondary metabolism based on reaction similarity and found a nootkatone biosynthetic enzyme and many alternative enzyme candidates for other reactions. This bacterial enzyme gene re-positioning method may offer an entirely new way of screening enzymes that cannot

Fig. 2. Gas chromatography-mass spectrometry analysis of the SCO3770-overexpressing strain.
A: Gas chromatographic chart of the purified purchased nootkatone standard (1) and S. lividans TK64/pIJ702D::SCO3770 (2). B: Mass spectrometry result for the nootkatone standard. C: Mass spectrometry result for S. lividans TK64/pIJ702D::SCO3770.
be achieved by most of other conventional methods. Overall, this method will enable the production of compounds that are dif
be achieved by most of other conventional methods. Overall, this method will enable the production of compounds that are dif
be achieved by most of other conventional methods. Overall, this method will enable the production of compounds that are dif
be achieved by most of other conventional methods. Overall, this method will enable the production of compounds that are dif
be achieved by most of other conventional methods. Overall, this method will enable the production of compounds that are dif
