Identification and functional analysis of cytochrome P450 complement in *Streptomyces virginiae* IBL14

Zhi-Zhen Li¹, Xiao-Fei Li¹†, Wei Yang¹, Xiang Dong¹, Jie Yu², Shu-Liang Zhu¹, Man Li¹, Li Xie¹ and Wang-Yu Tong¹*

**Abstract**

**Background:** As well known, both natural and synthetic steroidal compounds are powerful endocrine disrupting compounds (EDCs) which can cause reproductive toxicity and affect cellular development in mammals and thus are generally regarded as serious contributors to water pollution. *Streptomyces virginiae* IBL14 is an effective degradative strain for many steroidal compounds and can also catalyze the C25 hydroxylation of diosgenin, the first-ever biotransformation found on the F-ring of diosgenin.

**Results:** To completely elucidate the hydroxylation function of cytochrome P450 genes (CYPs) found during biotransformation of steroids by *S. virginiae* IBL14, the whole genome sequencing of this strain was carried out via 454 Sequencing Systems. The analytical results of BLASTP showed that the strain IBL14 contains 33 CYPs, 7 ferredoxins and 3 ferredoxin reductases in its 8.0 Mb linear chromosome. CYPs from *S. virginiae* IBL14 are phylogenetically closed to those of *Streptomyces sp.* Mg1 and *Streptomyces sp.* C. One new subfamily was found as per the fact that the CYP Svu001 in *S. virginiae* IBL14 shares 66% identity only to that (ZP_05001937, protein identifier) from *Streptomyces sp.* Mg1. Further analysis showed that among all of the 33 CYPs in *S. virginiae* IBL14, three CYPs are clustered with ferredoxins, one with ferredoxin and ferredoxin reductase and three CYPs with ATP/GTP binding proteins, four CYPs arranged with transcriptional regulatory genes and one CYP located on the upstream of an ATP-binding protein and transcriptional regulators as well as four CYPs associated with other functional genes involved in secondary metabolism and degradation.

**Conclusions:** These characteristics found in CYPs from *S. virginiae* IBL14 show that the EXXR motif in the K-helix is not absolutely conserved in CYP157 family and I-helix not absolutely essential for the CYP structure, too. Experimental results showed that both CYP Svh01 and CYP Svu022 are two hydroxylases, capable of bioconverting diosgenone into isonuatigenone and β-estradiol into estriol, respectively.

**Keywords:** Biotransformation, Cytochrome P450, Ferredoxin, Ferredoxin reductase, Gene sequencing, Secondary metabolism

**Background**

Cytochrome P450 (CYP) genes refer to such genes that encode a superfamily of iron-containing hemoproteins with a maximum absorption spectrum near 450 nm, often characterized by conserved Cys residue in hydrophobic pocket(s) [1]. Most of the ORFs of CYP have three distinct characteristics used often for their identification and analysis, i.e., the I-helix of putative CYPs (a highly conserved threonine involved in oxygen activation), the conserved EXXR motif located in the K-helix and the cytochrome P450 cysteine heme-iron ligand signature motif (GXXXXCXG, there are exceptions) [2]. According to a widely-accepted taxonomy, CYPs within a family share more than 40% amino acid identity and members of subfamilies share more than 55% amino acid identity [3]. Occasionally, the decision to accept a sequence in a known family depends greatly on how it clusters on a tree, not so much on the absolute amino acid identity [4].

© 2013 Li et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
CYPs have been confirmed existing in all eukaryotic (human, animals, plants, fungi, etc.) and prokaryotic organisms (bacteria, archaea, and even in virus) [5-8]. They often are monooxygenases involved in oxidation of a range of endogenous compounds, such as cholesterol, lipids and steroidal hormones, as well as xenobiotics such as drugs and toxic chemicals in environment [9-11]. CYPs catalyse diverse reactions, including C-H hydroxylation, epoxidation, hetero-atom oxidation, aromatic ring oxidation and dealkylation [11-13]. In the catalytic reaction process of P450 monooxygenase, one atom of O2 is inserted into substrate while the other is reduced to H2O. CYP genes responsible for secondary metabolism are often laid in antibiotic biosynthetic gene clusters to catalyze stereo- and region-specific reaction of substrates to related derivatives.

The biotransforming capabilities of bacterial CYPs have been widely elucidated. P450soy (CYP105D1) from Streptomyces griseus was involved in the degradation of a diverse array of complex agrochemicals and environmental pollutants [14]. CYP105C1 from Actinomycete spp. had the ability to transform benanomicin A into two derivatives, 10-hydroxybenanomicin A and 11-O-demethylbenanomicin [15]. The functions of related CYP107 family members have been reported. CYP107E from Micromonaspora griseorubida was found to govern the hydroxylation and epoxidization in mycinamic biosynthesis [16], P450 Terf (107 L) from Streptomyces platensis to catalyze hydroxylation of terfenadine [17] and hydroxylase PikC (107 L1) of Streptomyces venezuelae to convert narnbomycin A to picromycin [18]. CYP124 of Mycobacterium tuberculosis demonstrated omega-hydroxylation activity of relevant methyl-branched lipids [19]. YbdT (CYP152A) of Bacillus subtilis was involved in fatty acid beta-hydroxylation [20], CYP154 of Nocardia farcinica IFM10152 had the functions of the O-dealkylation and ortho-hydroxylation of formononetin [21] and 154H1 from Clostridium acetobutylicum performed biocatalytic reactions with different aliphatic and aromatic substrates [22].

Genome sequencing is an effective way to predict and annotate all the possible CYPs genes in an organism. Streptomyces coelicolor A3 (2), a typical strain which is often used for the study of physiological function and antibiotic production, is the first Streptomyces species sequenced in 2001. Its linear chromosome is 8.7 Mb [23] which contains 7825 open reading frames (ORFs) with 18 putative CYPs [24]. S. avermitilis, known for producing the antiparasitic agent avermectin, contains 7600 ORFs with 33 putative CYPs in the 9 Mb chromosomes [25,26]. The genome of Streptomyces peucetius ATCC27592 with the size of 8.7 Mb contains 19 putative CYPs [27].

S. virginiae IBL14, isolated from activated sludge for treatment of waste from a steroidal drug factory, is an effective degradative strain of various steroidal compounds, including progesterone, isostestosterone, dihydrotestosterone, hydrocortisone, cholesterol and oestrone [28]. To comprehensively understand the function of CYPs of S. virginiae IBL14 in degradation and biotransformation of diosgenin, the whole genome sequencing of S. virginiae IBL14 isolated by our lab was carried out for the first time. Using in silico technology, we predict and annotate all of the putative CYPs of S. virginiae IBL14 and analyze these CYPs evolutionarily and functionally via comparison with those of other Streptomyces species. Furthermore, functions and characteristics of CYP genes svh01 and CYP svu022 in this strain are experimentally identified and analyzed.

Results and discussion
Genome sequencing and CYPs in S. virginiae IBL14
By in silico analysis of newly-sequenced S. virginiae IBL14 8.0 Mb genome, 8288 ORFs are identified and the total GC content exceeds 70%. The annotated results via Rpsblast display that there are a total of 33 putative CYPs in the genome of this strain IBL14, contributing to approximately 0.4% of all the coding sequences. The number of CYPs is identical to that in S. avermitilis and almost two times as that in S. coelicolor A3(2) and S. peucetius ATCC27952 (18 and 19 CYPs, respectively). Such high level of CYP diversity suggests the high diversity of the secondary metabolism pathways in S. virginiae IBL14.

The 32 out of 33 putative CYPs of S. virginiae IBL14 belong to 13 previously-reported CYP families, i.e., 105 (5), 107 (11), 121 (1), 124 (1), 147 (1), 152 (1), 154 (1), 157 (2), 185 (1), 191 (3), 197 (4), 247(1) and another to an unknown family, as shown in Table 1. Among the all, the CYP121A (Svu018), CYP124 (Svu19), CYP147 (Svu20), CYP152 (Svu021), CYP154H (Svu022), CYP157 (Svu023-024), CYP185 (Svu025), CYP191 (Svu26-28), CYP197 (Svu017,029-031) and CYP247 (Svu032) are firstly reported in S. virginiae, and especially, CYP107M, CYP185A and CYP247A have been found rarely in Streptomyces spp. The Svu025, Svu026 and Svu029 have lower identity with other family members (<50%) while others show more than 63% identity to CYPs of other organisms. It's worth noting that the Svu001 presumably belongs to a new CYP family since no close homologue is found in Genbank except that in Streptomyces sp. Mg1 with 66% identity.

Features of CYPs from S. virginiae IBL14
Table 2 displays the three characteristic motifs of CYPs of S. virginiae IBL14. The critical residues are highlighted with bold fonts, which are threonine (T) in GXXTT motif of I-helix, glutamic acid (E) and arginine (R) in EXXR motif of K-helix and cysteine (C) in the GXXXCXG heme-binding domain signature, respectively.

From the Table 2, we can find the I-helix is absent in Svu001(new family), and the I-helix and K-helix missing
Table 1  Putative cytochrome P450s in *S. virginiae* IBL14 with their closest homologs

| ID*                  | Size* | Species                          | Protein identifier | CYP family | AA overlap% | identity% |
|----------------------|-------|----------------------------------|--------------------|------------|-------------|-----------|
| Svu001               | 464   | *Streptomyces* sp. Mg1           | ZP_05001937        | new        | 598         | 66        |
|                      |       | *Photobacterium profundum* 3TCK  | ZP_01217946        |            | 113         | 25        |
| Svu002               | 361   | *Streptomyces violaceusniger* Tu 4113 | BAG16627          |            | 529         | 69        |
| Svu003               | 439   | *Streptomyces* sp. Mg1           | ZP_06679587        |            | 93.2        | 66        |
| Svu004               | 398   | *Streptomyces* sp. ACT50-5       | ZP_07284721        |            | 611         | 87        |
| Svu005               | 400   | *Streptomyces* sp. C             | ZP_07285089        | 105D       | 595         | 74        |
| Svu006               | 403   | *Streptomyces* sp. Mg1           | ZP_04999247        | 107 L14    | 484         | 59        |
| Svu007               | 351   | *Streptomyces* sp. C             | ZP_07286517        | 107 L      | 727         | 85        |
| Svu008               | 406   | *Streptomyces* sp. C             | ZP_04997607        | 107 L14    | 578         | 74        |
| Svu009               | 415   | *Streptomyces* sp. C             | ZP_05005324        | 107 L      | 601         | 75        |
| Svu010               | 396   | *Streptomyces* sp. C             | ZP_07287209        | 107 L      | 808         | 92        |
| Svu011               | 405   | *Streptomyces* sp. C             | ZP_0500207         | 107 L      | 780         | 91        |
| Svu012               | 430   | *Streptomyces* sp. C             | ZP_0500207         | 107 L      | 780         | 91        |
| Svu013               | 396   | *Streptomyces* sp. C             | ZP_07287353        | 107 L      | 780         | 91        |
| Svu014               | 395   | *Streptomyces* sp. C             | ZP_07287353        | 107 L      | 780         | 91        |
| Svu015               | 406   | *Streptomyces* sp. C             | ZP_0500207         | 107 L      | 780         | 91        |
| Svu016               | 395   | *Streptomyces* sp. C             | ZP_0500207         | 107 L      | 780         | 91        |
| Svu017               | 368   | *Streptomyces* sp. C             | ZP_0500207         | 107 L      | 780         | 91        |
| Svu018               | 393   | *Streptomyces* sp. C             | ZP_0500207         | 107 L      | 780         | 91        |
in Svu002 (105 L, often for hydroxylation activity) [29], which reflects I-helix is not absolutely essential for the CYP structure. The 2 members of CYP157 family Svu023 (E276VLW279)/157A and Svu024 (E284QILW288)/157C do not have arginine residue in K-helix like the CYP157C1 from S. coelicolor having a motif E297QSLW [30] and the CYP157A2 and CYP157C2 from S. avermitilis exhibiting a257EVLW motif and a257EQSLW motif [26]. The CYP157 family proteins that lack consensus EXXR motifs but genetically are linked to their upstream conservons imply that they have functions linked to the upstream pathway(s) [30]. Besides, Svu002, Svu018, Svu021, Svu023 and Svu031 do not strictly follow the GXXXCGXG motif of heme-binding.

Multiple alignments and phylogenetic analysis

The phylogenetic tree of the combined CYPs of S. virginiae IBL14, S. avermitilis MA-4680, S. venezuelae ATCC 10712 and Streptomyces sp. Mg1 is presented in Figure 1. From Figure 1, we can find almost of all the CYPs in S. virginiae IBL14 are closely related to their homologues. More than 10 of CYPs from S. virginiae IBL14 are close to those from Streptomyces sp. Mg1 and the member (Svu001) of new CYP family found in S. virginiae IBL14 is only close to Streptomyces sp. Mg1. These results indicate that the CYPs from S. virginiae IBL14 are closer to those from Streptomyces sp. Mg1 than those from other Streptomyces spp, including S. avermitilis MA-4680 and S. venezuelae ATCC 10712. For the four species of S. virginiae IBL14, sp. Mg1, avermitilis MA-4680 and S. venezuelae ATCC 10712, the families CYP 107 and CYP157 (labeled with circle A and B in Figure 1, respectively) have more closely evolutionary relationship. Further, the paralogous relationship of the 33 CYPs in S. virginiae IBL14 was generated with the neighbor-joining methods (Clustal W and MEGA 5.0). From Figure 2, we can find that svh01 and svu03 and svu04 as well as svu022 and svu005 in S. virginiae IBL14 have the closest homologous evolutionary relationship, respectively. It’s worth noting that most members belonging to the same CYP family are clustered together as expected, e.g., the 11 members of CYP107 family.

The prediction of functions of CYPs in S. virginiae IBL14

A high identity over 70% among different protein sequences reasonably suggests that they may hold...
similar function [26]. As shown in the Table 1, we can find a sum of 26 CYP sequences of *S. virginiae* IBL14 have best matches to those of other *Streptomyces*, which are helpful in function prediction.

CYP105 and CYP107 are the most studied bacterial cytochromes which are associated with the degradation and biotransformation of a diverse array of xenobiotics and antibiotic biosynthesis. Analysis of CYPs sequence of *S. virginiae* IBL14 shows that there are 11 CYPs belonging to CYP107, five to CYP105, four to CYP197, three to CYP191, two to CYP157 and one to each other family, which indicates the diversity and importance of the two groups CYP105 and CYP107. The predicted functions of several putative CYPs in *S. virginiae* IBL14, combined with reported experimental evidences, were listed in Table 3.

**CYPs in *S. virginiae* IBL14 and their ferredoxin reductase and ferredoxin**

The catalytic activity of CYPs depends greatly on individual ferredoxin or/and ferredoxin reductase associated with. It was reported that there are three, six and four
ferredoxin reductase genes and six, nine and two ferredoxin genes in *S. coelicolor* A3 (2), *S. avermitilis* and *S. peucetius*, respectively. In *S. coelicolor* A3 (2) only CYP105D5 is arranged in an operon with a ferredoxin gene [24]. In *S. peucetius* CYP147F is clustered with ferredoxin reductase [27]. In *S. avermitilis* both CYP105P1 and CYP105D6 are clustered with ferredoxin, CYP147B1 is arranged in an operon with a ferredoxin and ferredoxin reductase, CYP105Q1 is associated in an operon containing both a ferredoxin and ferredoxin reductase, and CYP102 is fused to a P450 reductase [26].

Three ferredoxin reductase genes and seven ferredoxin genes are found in *S. virginiae* IBL14 after annotation of *S. virginiae* IBL14 genome. That is, the activities of many of the CYPs in *S. virginiae* IBL14 are supported by different combinations with the three ferredoxin reductases and seven ferredoxins. Also in *S. virginiae* IBL14, *svu005* (CYP105D), *svh01* (CYP105C) and *svu019* (CYP124B) is found to cluster with ferredoxin *svf03*, *svf09* and *svf07*, respectively and *svu020* (CYP147A) clustered with ferredoxin reductase *svfr03* and ferredoxins *svf06*. The facts suggest that the functional realization of CYPs

![Figure 1 Phylogenetic tree of the CYPs from *S. virginiae* IBL14 and three related bacteria.](http://www.biomedcentral.com/1471-2164/14/130)
Svu005, Svh01, Svu019 and Svu020 needs the participation of electron transfer. The result of homology analysis by Blast-searching the Genbank are listed in the Table 4.

Regulatory elements and functional genes clustered with CYPs

The CYPs in S. peucetius ATCC27952 clustered with regulatory elements were reported [27]. In the annotations of gene arrangement around the putative CYPs on the S. virginiae IBL14 chromosome, svu022, svu023 and svu024 were found to cluster with the genes of ATP/GTP binding proteins (having a phosphate-binding loop for energy requiring metabolic reactions) [34], svu001, svu015 to cluster with LysR-family transcriptional regulator (regulating a diverse set of genes, including those involved in virulence, metabolism, quorum sensing and motility) [35], svu011 to cluster with two component transcriptional regulators and LuxR family (quorum sensing signals in Gram-negative bacteria often regulated by acylated homoserine lactones) [36], svu018 to cluster with a transcriptional regulator, AraC family (transcriptional regulators having diverse functions ranging from carbon metabolism to stress responses to virulence) [37] and two component transcriptional regulators, LuxR family and svu020 to cluster with the ATP-binding protein fbpC and TetR-family transcriptional regulators (among bacteria with an HTH DNA-
binding motif for the transcriptional control of multidrug efflux pumps, pathways for the biosynthesis of antibiotics, response to osmotic stress and toxic chemicals, control of catabolic pathways, differentiation processes, and pathogenicity [38].

As described above, the CYPs in *S. virginiae* IBL14 chromosome are responsible for the transcriptional regulation of many functional genes related with primary, secondary metabolism, as well as the responses to environmental factors as expected. Besides, CYPs are clustered with other functional genes. *svh*01 is adjacent to the genes of MdB, ABC-type multidrug transport system, ATPase and permease components, which may be involved in the transportation of substrates [39]. *svu*009 lies next to alcohol dehydrogenase, suggesting that *svu*009 may take part in alcohol bioconversion and biodegradation. *svu*013 is next to 4, 5-DOPA dioxygenase which is a member of the class III extradiol dioxygenase family (a group of enzymes which use a non-heme Fe (II) to cleave aromatic rings between a hydroxylated carbon and an adjacent non-hydroxylated carbon), suggesting that the combination of *svu*013 and 4, 5-DOPA dioxygenase may be responsible in biodegradation of substrates with aromatic rings. *svu*026 is adjacent to MbtH-like protein which is found in known antibiotic synthesis gene clusters [40]. The cholesterol oxidase ChoL from *S. virginiae* IBL14 in the bioconversion and biodegradation of diosgenin responsible for the conversion of diosgenin to diosgenone (a 4-ene-3-keto steroid) via a couple of C3-dehydrogenation and C4-5-isomerization was reported [41].

Table 3 Prediction of functions of several putative CYPs in *S. virginiae* IBL14

| ID  | Functions                              | Reference |
|-----|----------------------------------------|-----------|
| Svu03 | Hydroxylation & O-demethylation        | [15]      |
| Svu05 | N-demethylation & Hydroxylation        | [31]      |
| Svh01 | Hydroxylation                          | [32] & this study |
| Svu06 | Hydroxylation                          | [17]      |
| Svu07 | Hydroxylation                          | [16]      |
| Svu09 | Hydroxylation                          | [19]      |
| Svu21 | Hydroxylation & Decarboxylation        | [33]      |
| Svu22 | Hydroxylation & O-dealkylation         | This study |

Table 4 Putative ferredoxin reductases and ferredoxins in *S. virginiae* IBL14 with their closest homologs

| ID  | Accession numbers | NO. nucleic acids | Species | Accession | Identity% |
|-----|-------------------|-------------------|---------|-----------|-----------|
| svf01 | JX119052          | 453               | Streptomyces sp. C | ZP_07290734 | 94        |
|       |                   |                   | Streptomyces pristinaespiralis ATCC 25486 | ZP_06911868 | 92        |
| svf02 | JX119053          | 463               | Streptomyces sp. C | ZP_07285271 | 87        |
|       |                   |                   | Streptomyces sp. Mg1 | ZP_05002250 | 84        |
| svf03 | JX119054          | 464               | Streptomyces sp. C | ZP_07289558 | 94        |
|       |                   |                   | Streptomyces peucetius ATCC 27952 | CAF33360 | 84        |
| svf04 | JX119056          | 1143              | Streptomyces sp. C | ZP_07285090 | 84        |
|       |                   |                   | Streptomyces vinidochromogenes DSM 40736 | ZP_07308348 | 79        |
|       |                   |                   | Streptomyces sp. Mg1 | ZP_04996989 | 83        |
| svf05 | JX119057          | 234               | Streptomyces sp. C | ZP_07286537 | 88        |
|       |                   |                   | Streptomyces sp. Mg1 | ZP_05002165 | 79        |
| svf06 | JX119058          | 231               | Streptomyces peucetius ATCC 27952 | ACE73829 | 62        |
|       |                   |                   | Streptomyces hygroscopicus subsp | AYE87986 | 61        |
| svf07 | JX119059          | 600               | Streptomyces sp. C | ZP_07287304 | 98        |
|       |                   |                   | Streptomyces venezuetae ATCC 10712 | CCA56325 | 94        |
| svf08 | JX119060          | 315               | Streptomyces sp. C | ZP_07285869 | 89        |
|       |                   |                   | Streptomyces peucetius ATCC 27952 | ACE73824 | 88        |
| svf09 | JX119061          | 243               | Streptomyces castellia NRRL 8057 | YP_004920089 | 63       |
|       |                   |                   | Streptomyces diastaticus | AAR16520 | 61        |

*a* The name of gene in *S. virginiae* IBL14.

*b* Homologues searched in Genbank.
gene encoding Svu004 (CYP105L) clusters with the genes of putative ferredoxin Svf02 and cholesterol oxidase (ChoL), suggesting that the cytochrome P450 joins with the cholesterol oxidase ChoL to catalyze the oxidation of cholesterol and its structural analogs. In conclusion, CYPs from *S. virginiae* IBL14 may have multiple functions in secondary metabolism, including hydroxylation, dehydrogenation, ring-cleavage, transportation, etc.

**Functional identification and characteristics of svh01 and svu022**

To elucidate all putative CYPs' functions in *S. virginiae* IBL14, four CYP genes of the strain IBL14 were firstly selected. Among them, the functional identities of CYP genes *Svh01* (105C1) and *svu022* (154H) have been finished.

The cytochrome P450 *Svh01* (responsible for the C25-hydroxylation of diosgenin) [32] belongs to the class I (prokaryotic/mitochondrial) P450 system based on a taxonomic split, in which electrons are transferred from NADPH or NADH to ferredoxin reductase and ferredoxin. Sequence analysis revealed the complete sequence of *svh01* with ATG as the start codon has 70% G + C content. The sequence of possible ribosome-binding site is located on the upstream of *svf09* (a coenzyme of *Svh01*).

Both *svh01* and *svf09* contain 1200 bp and 243 bp, respectively, based on sequence analysis. To obtain the expressed products of them, both *svh01* and *svf09* sequences were first ligated into a pET22b vector in a cluster to generate the expression plasmid pET22b-*svh01-svf09* that was then cloned into *E. coli* JM109 (DE3) to form a recombinant strain *E. coli* IBL161 [JM109 (DE3)/pET22b-*svh01-svf09*]. The PCR results of *svh01* and *svf09* from the recombinant strain *E. coli* IBL161 were analyzed by gel electrophoresis (Figure 3A and B) and also confirmed by gene sequencing.

The *svu022* with a G + C content of 73% (clustering with the gene of ATP/GTP binding protein) consists of 1239 nucleotides. Similarly, the complete sequence of *svu022* was first inserted to the shuttle plasmid pHCMC05 to form the recombinant plasmid pHCMC05-*svu022*, and then cloned in *B. subtilis* WB800N (improving the extra-cellular expression level of *svu022* for the analysis of enzymatic biotransformation) to produce the recombinant strain *B. subtilis* IBL 241 [WB800N/pHCMC05-*svu022*]. The PCR result of *svu022* from the recombinant strain *B. subtilis* IBL 241 is shown in Figure 3C.

*Svh01* (105C1) is a peptide of 399 amino acids, with a molecular weight of 44.04 kDa and a pl value of 4.97 estimated by the ExPASy (a computing pl/MW tool). To obtain its expressed product and study product characteristics, the recombinant strain *E. coli* IBL161 was incubated and induced. The expression of *Svh01* was shown in Figure 4A. From the SDS-PAGE, we can find that the two distinctly additional protein bands should be *Svh01* with an about MW of 44 kDa and *Svf09* with an about MW of 8.0 kDa, respectively. The further functional identification of the *Svh01/*FcpC of *S. virginiae* IBL14, hydroxylation the C25-tertiary carbon of diosgenin to form isonuatinone, was experimentally confirmed [32].

*Svu022* (154H) is a deduced protein of 412 amino acids which shares 91% identity with that in *Streptomyces sp. Mg1*. The estimation of MW and pl of SVU022 are 44.59 kDa and 5.00, respectively. Similarly, the recombinant strain *B. subtilis* IBL 241 was incubated and induced to study the product expression and its characteristics. The expressed result of *Svu022* from the recombinant strain *B. subtilis* IBL 241 was shown in Figure 4B. The SDS-PAGE displays a distinct protein band with about MW of 45.0 kDa as expected. The further experimental results from TLC, HPLC and LC/MS indicated that the CYP *Svu022* enables to biotransform β-estradiol into estradiol. Figure 5 shows the profiles of the biotransformation of β-estradiol by strains *B. subtilis* WB800N and *B. subtilis* IBL 241 in HPLC. The functional identification of the Svu005 (CYP105D) and Svu019 (CYP124B) is in progress.

**Conclusion**

*S. virginiae* IBL14 contains 33 putative CYPs, 7 ferredoxins and 3 ferredoxin reductases in its 8.0 Mbp linear chromosome. Most of the CYPs in *S. virginiae*
IBL14 belong to the CYP107 (11 members) family and CYP105 (5 members) family. Compared phylogenetically with CYPs from 3 typical Streptomycete spp., S. virginiae IBL14 appears to be closest to those of Streptomyces sp. Mg1.

Further analysis showed that among all of the 33 CYPs in S. virginiae IBL14, three CYPs are clustered with ferredoxins, one with ferredoxin and ferredoxin reductase and three CYPs with ATP/GTP binding proteins, four CYPs arranged with transcriptional regulatory genes and one CYP locates on the upper of ATP-binding protein and transcriptional regulators as well as four CYPs associated with other functional genes involved in secondary metabolism and degradation.

The new characteristics found in CYPs from S. virginiae IBL14 suggest that the EXXR motif in the K-helix is not absolutely conserved in CYP157 family as reported [30] and I-helix not absolutely essential for the CYP structure. Particularly, one new family was found based on the CYP svu001 in S. virginiae IBL14 which shares 66% identity only to that from Streptomyces sp. Mg1.

Two recombinant strains E. coli IBL161 [JM109 (DE3)/pET22b-svh01-svf09] and B. subtilis IBL 241 [WB800N/pHCMC05-svu022] were constructed and subsequently their functions were identified, respectively. Experimental results showed that both CYP Svh01 and CYP Svu022 are two hydroxylases, capable of bioconverting diosgenone into isonuatigenone and β-estradiol into estriol, respectively.
Methods

Strains and plasmids

*S. virginiae* IBL-14 (CCTCCM 206045) [42] as the strain of interest was used for the Cytochrome P450 gene identification and functional analysis. *E. coli* JM109, JM109 (DE3) and *B. subtilis* WB800N were used as the host for plasmid construction and target protein expression in the functional identification of the CYPs, respectively. The vector pET22b was used for cloning and expression of genes of interest in *E. coli*. The shuttle plasmid pHCMC05 was used for the expression of target proteins in *B. subtilis* (a GRAS strain by FDA). The features of the bacterial strains and plasmids used in this study are listed in Table 5.

Media and cultivation

Luria-Bertani (LB) medium was used for plasmid construction and protein expression. A final concentration of 70 μg/ml ampicillin was supplemented into the medium when *E. coli* IBL161 [JM109 (DE3)/pET22b-svh01-svu09] and *E. coli* IBL152 [JM109/pHCMC05-svu022] were cultivated. A final concentrations of 25 μg/ml chloramphenicol was added to the medium when *B. subtilis* IBL 241 [WB800N/pHCMC05-svu022] was cultivated. The cultivating procedure of *S. virginiae* IBL-14 has been described previously [42]. Diosgenin in 95% purity (J&K Chemical Ltd, China) was dissolved in anhydrous ethanol before adding into medium. Sequencing and *in-silico* identification analyses of CYPs

The *S. virginiae* IBL14 genome sequencing was performed at 454 platform (Encode Genomics Co. Ltd., Suzhou, China) for the first time (sequence data will be published step by step). All of the ORFs of this genome were predicted using glimmer3.0 and prodigal, respectively. To dig out all possible CYP gene function information in *S. virginiae* IBL14, the genome sequence of the strain was compared with the SWISSPROT, TrEMBL, KEGG databases by using Blastp and the CDD and COG databases by using Rpsblast, respectively. The deduced amino acid sequences of the putative CYPs of *S. virginiae* IBL14 were aligned with the CYPs from *S. avermitilis* MA-4680, *S. venezuelae* ATCC 10712 and *Streptomyces* sp. Mgl1 by using ClustalW [43]. Then the molecular evolution and phylogenetic analyses by neighbor-joining methods were carried out using MEGA5.0 [44]. To forecast the possible functions involved in secondary metabolism, comparison between all putative CYPs of *S. virginiae* IBL14 with those in other organisms based on homologues was done by using Blastp too.

Using the three motifs as described above as criteria, the CYP gene candidates of *S. virginiae* IBL14 were blast searched against GenBank non-redundant protein database to identify their closest bacterial homologues and tentatively distribute all of the CYPs of *S. virginiae* IBL14 into the corresponding family or subfamily [26]. Similar procedure was performed to the putative ferredoxin and ferredoxin reductase genes to identify their closest bacterial homologues.

Construction and cloning of expression plasmids

The genes of *svh*01, *svf*09 and *svu*022 from the genomic DNA of *S. virginiae* IBL14 were amplified by using PCR method (Pfu DNA Polymerase, Fermentas, Thermo Fisher Scientific Inc.) and the primers used are listed in Table 6. The PCR products of *svh*01 and *svf*09 were digested with *NdeI/ EcoRl* and *EcoRl/Hind III*, respectively, then ligated

| Strains | Relevant properties | Source |
|---------|--------------------|--------|
| Escherichia coli | | |
| JM109 | Cloning host, genotype: endA1, recA1, gyrA96, thi, hsdR17 (Rk-, mk-), relA1, supE44, (lac-proAB), [F' traD36, proAB, lacI(q22)Δ(115) | Promega |
| JM109 (DE3) | Expression host, genotype: endA1, recA1, gyrA96, thi, hsdR17 (Rk-, mk-), relA1, supE44, (lac-proAB), [F’ traD36, proAB, lacI(q22)Δ(115)] | Promega |
| Bacillus subtilis | | |
| WB800N | Secretion host with resistance to neomycin, genotype: rpsE- aprE- epr bpr mpr - ble rpsB + bsr vpr wspA - hos cm + neos | Mo Bi Tec |
| Streptomyces virginiae | | |
| IBL14 | Wild type | Our lab |

Plasmids

pET22b | Expression vector in *E. coli* | Novagen |
pHCMC05 | Shuttle plasmid | BGSC |
pET22b-svh01-svu09 | The fragment of *svh*01 and *svu*022 were digested with *NdeI/EcoRl* and *EcoRl/Hind III*, respectively, and ligated into the *NdeI* and *Hind III* sites of pET22b | This study |
pHCMC05-svu022 | The gene of *svu*022 digested with *BamHl/SmaI* was ligated into pHCMC05 | This study |

Table 6 The PCR primers used in this study

| Primer | Primer sequence* (5’ to 3’) | Restriction site |
|--------|-----------------------------|----------------|
| pSVH01F | GCCCCCCATATGAGTGAGTCCCTCACCCCTGC | *NdeI* |
| pSVH01R | GGAGGAATTCATTCCGCCTCCACAGGTG | *EcoRI* |
| pSVF09F | CCGGAATTCCGGGGCGAGTAGGGCCGGG | *EcoRI* |
| pSVF09R | CCCAAGCTTTCGAGCGGGATTCGGGG | *Hind III* |
| pSVU022F | CTGAGATCAGGCTGCGGATGACC | *BamHl* |
| pSVU022R | CCTAAGCTTCCGAGGGCTGAGC | *SmaI* |

*The underlined sequence are recognition sites of restriction enzymes and the nucleotides before it are the protected bases. All primers are designed by Primer Premier 5.0 and verified by Oligo 7.0.*
into a pET22b vector, and finally transformed to the host bacterium *E. coli* JM109 (DE3). Similarly, the PCR product of *svu022* was digested with *Bam*H*I*/Sma*I, then ligated into a shuttle plasmid pHCMC05 and finally transformed to *B. subtilis* WB800N.

**Expression and analysis of target proteins**

0.3 ml (inoculation ratio of 1%) of the overnight culture of *E. coli* IBL161 as seed was inoculated in 30 ml LB medium (containing 70 µg/ml ampicillin) and then cultivated at a shaking speed of 200 rpm at 37°C. The expression of target protein was induced by adding 0.2 mM IPTG when the OD value reached 0.5 ~ 0.6 at 600 nm. Then the culture was continuously cultivated for another 24 h at 25°C at a speed of 200 rpm in a rotary shaker. Similarly, the overnight culture of *B. subtilis* IBL 241 was inoculated with 1% ratio in 30 ml LB medium (25 µg/ml chloramphenicol, 200 rpm at 30°C). After adding 0.2 mM IPTG in logarithmic growth phase, the culture was continuously cultivated for another 48 h at the same conditions. The harvested recombinant cells were resuspended and subjected to ultrasonication in 50 mM PBS (pH 7.4), and then centrifuged at 6000 rpm for 5 min. The supernatant was analysed by SDS-PAGE.

**Biotransformation and product extraction**

One milliliter of β-estradiol/diosgenin (a final concentration of 0.2 mg/ml) for each flask was added for biotransformation after *E. coli* IBL161 was induced by IPTG at 25°C for 2 h. After cultivated for another 24 h under the same conditions, the cultures were extracted two times with a half volume of 100% ethyl acetate (Sinopharm Chemical Reagent Co., Ltd). The extracts were evaporated to dryness, then re-dissolved in 1 ml anhydrous ethanol, and finally detected and analyzed (thin layer chromatography/TLC, high performance liquid chromatography/HPLC and liquid chromatography–mass spectrometry LC-MS).

**DNA and protein analytical methods**

DNA electrophoresis for recombinant plasmid analysis was carried out in agarose gels at 110 V for 30 min [45]. SDS-PAGE with a 15% (w/v) acrylamide gel for expressed protein analysis was run at 110 V for 2 h according to Schagger’s publication [46]. The bands were visualized by Coomassie R-250 staining.

**HPLC analysis of biotransformation products**

To identify and analyze the metabolites, high performance liquid chromatography (HPLC) was carried out. Simply, the sample of 10 µl was first loaded onto 250 mm Symmetry C 18 (4.6 mm × 250 mm, Waters Co., USA) and eluted with ethanol/water (60/40, v/v). The flow rate, the wavelength for UV-detection and the temperature of the column on the HPLC system (Breeze 1525 series, Waters Co., USA) were set at 1 ml/min, 245 nm and 35°C, respectively. The products after biotransformation were qualitatively and quantitatively analyzed by comparing with corresponding standard material.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

Z-GL carried out the bioinformatic and genomic analyses and experiments and drafted the manuscript. X-FL carried out the experiments and bioinformatic and genomic analyses, WY, XD, JY, S-LZ, ML, LX participated in the experiments or interpretation of results. W-YT contributed to study design, data analysis and writing of the manuscript. All authors read and approved the final manuscript.

**Acknowledgments**

Preparation of this manuscript was supported by the National Natural Science Foundation of China (20976001) and by the Key Scientific Research Projects of Department of Education of Anhui Province of China (KJ2009A161) and by the Academic Leader Foundation of 211 Project (phase-3) of Anhui University of China (AU02030154). Also, we are grateful to Hui Peng and Ya-Zhong Xiao, School of Life Science, Anhui University, Hefei, China, for generous gift of plasmids and to Hong Yu, Encode Genomics Co. Ltd, Suzhou China, for the genome sequencing of the strain IBL14.

**Author details**

1Integrated Biotechnology Laboratory, Institute of Health Science, School of Life Science, Anhui University, 111 Jiulong Road, Hefei 230601, China.
2Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada.

Received: 25 October 2012 Accepted: 21 February 2013
Published: 27 February 2013

**References**

1. Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J: Structure and function of cytochromes P450: A comparative analysis of three crystal structures. *Structure (London, England)* 1993, 3(1):41–62.
2. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotch O, Coon MJ, Estabrook RW, et al: P450 superfAMILY: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 1996, 6(1):1–42.
3. Nelson DR: Cytochrome P450 nomenclature. 2004, *Methods Mol Biol* (Clifton, NJ) 2005, 320:1–10.
4. Nelson DR: Cytochrome P450 nomenclature. *Methods Mol Biol* 1998, 107:15–24.
5. Danielson PB: The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Current drug metabolism* 2002, 3(6):561–597.
6. Mogi T, Kita K: Diversity in mitochondrial metabolic pathways in parasitic protists *Plasmodium* and *Cryptosporidium*. *Parasitol Int* 2010, 59(3):305–312.
7. Fu C, Xiong J, Mao W: Genome-wide identification and characterization of cytochrome P450 monoxygenase genes in the eutelate *Tetrahymena thermophila*. *BMC Genomics* 2009, 10(1):208.
8. Lamb DC, Lei L, Warnlow AGS, Lepesheva GI, Mullins JLG, Waterman MR, Kelly SL: The first virally encoded cytochrome P450. *J Virol* 2009, 83(16):8266–8269.
9. Guengerich FP: Cytochrome p450 and chemical toxicology. *Chem Res Toxicol* 2008, 21(1):70–83.
10. Pryor WA: Cytochrome P450: Structure, mechanism, and biochemistry. *New York: Montellano Plenum Press*; 1996.
11. McLean KJ, Hans M, Munro AW: Cholesterol, an essential molecule: diverse roles involving cytochrome P450 enzymes. *Biochem Soc Trans* 2012, 40(3):587–593.
12. Dujellet H, Johnston JB, Ortiz de Montellano PR: The *Myxococccun tuberculosis* cytochrome P450 system. *Arch Biochem Biophys* 2010, 493(1):82–95.
13. Johnston JB, Ouellet H, Podust LM, Ortiz de Montellano PR: Structural control of cytochrome P450-catalyzed ω-hydroxylation. Arch Biochem Biophys 2011, 507(1):86–94.

14. Trower MK, Lenstra R, Omer C, Buchholz SE, Sariaslani FS: Characterization and expression of a P-450-like mycinamicin biosynthesis gene using a novel microcomposonosa-Escherichia coli shuttle cosmid vector. Mol Gen Metab 1994, 245(4):456–464.

15. Lombard M, Salard I, Sari M-A, Mansuy D, Buisson D: Control of cytochrome P450-catalyzed ω-hydroxylation from Streptomyces griseus and ortho-hydroxylation of formononetin. Biochem Biophys Res Commun 2009, 387(2):931–937.

16. Lombard M, Salard I, Sari M-A, Mansuy D, Buisson D: Expression of the genes controlling cytochrome P450-catalyzed ω-hydroxylation in Nocardia farcinica and ortho-hydroxylation of formononetin. Arch Biochem Biophys 2009, 487(1):141–147.

17. Johnston JB, Kells PM, Podust LM, Ortiz de Montellano PR: Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from Mycobacterium tuberculosis. Proc Nat Acad Sci United States Am 2009, 106(49):20687–20692.

18. Youssef NH, Wofford N, McMenney MJ: Identification of the long-chain fatty acid beta-hydroxylating cytochrome P450 enzyme YbdT for lipopeptide biosynthesis in Bacillus subtilis strain OX105. Int J Mol Sci 2011, 12(3):1767–1786.

19. Choi K-Y, Park H-Y, Kim B-G: Characterization of bi-functional CYP154 from Nocardia farcinica IFM10152 in the O-dealkylation and ortho-hydroxylation of fomorenone. Enzyme Microb Technol 2010, 47(7):327–334.

20. Schallmayer A, den Besten G, Teune KJP, Kembare RF, Jansen DB: Characterization of cytochrome P450 monoxygenase CYP154H1 from the thermophilic soil bacterium Thermobifida fusca. Appl Microbiol Biotechnol 2011, 89(2):1475–1485.

21. Bentlej MC, Kealey JT, Ashley GW, McDaniel R: Characterization of the macroline P-450 hydroxylase from Streptomyces venezuelae which converts nARBomycin to picromycin. Biochem Biophys Res Commun 1998, 378(2):14437–14942.

22. Tanaka Y, Muto N, Beppu T, Horinouchi S: Identification of a P-450-like mycinamicin biosynthesis gene using a novel microcomposonosa-Escherichia coli shuttle cosmid vector. Mol Gen Metab 1994, 245(4):456–464.

23. Choi K-Y, Park H-Y, Kim B-G: Characterization of bi-functional CYP154 from Nocardia farcinica IFM10152 in the O-dealkylation and ortho-hydroxylation of fomorenone. Enzyme Microb Technol 2010, 47(7):327–334.

24. Lamb DC, Skaug T, Song H-L, Jackson CJ, Podust LM, Waterman MR, Kell DB, Harris DE, Quail MA, Kieser H, Harper D, Monaghan RL, Olson G, Putter I, Streptomyces coelicolor producer and comparison to that of Streptomyces virginiae IBL-14. BMC Genomics 2009, 10(10):2731–2739.

25. Zeng X, Gallegos MT, Brennan R, Tobeas R, The TetR Family of transcriptional repressors. Microb Mol Biol Rev 2005, 69(2):326–356.

26. Li B, Wang F-Q, Wei D-Z: Cholesterol oxidase chol is a critical enzyme that catalyzes the conversion of diosgenin to 4-ene-3-keto steroids in Streptomyces virginiae IBL-14. Applied Microbiol Biotechnol 2010, 85(5):1831–1838.

27. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994, 22(22):4673–4680.

28. Kamura K, Petersen D, Petersen N, Stecher G, Nei M, Kumar S: MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

29. Sambrook J, Russel D: Molecular cloning: a laboratory manual. 3rd edition, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2011.

30. Schägger H, von Jagow G: Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Analytical Biochem 1987, 150(1):356–362.

31. Wang W, Wang F-Q, Wei D-Z: Characterization of P450 FcpC, the enzyme responsible for bioconversion of diosgenone to isoungaetone in Streptomyces virgulinae IBL-14. Applied environ microb 2009, 75(12):4202–4205.

32. Li B, Wang F-Q, Wei D-Z: Characterization of P450 FcpC, the enzyme responsible for bioconversion of diosgenone to isoungaetone in Streptomyces virgulinae IBL-14. Applied environ microb 2009, 75(12):4202–4205.

33. Rude MA, Baron TS, Brubaker S, Allibha M, Del Cardaye SB, Schirmer A: Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid desaturase from Jeotgalicoccus species. Appl Environ Microbiol 2011, 77(5):1718–1727.

34. Saraste M, Sibbald PR, Wittinghofer A: The P-loop a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci 1990, 15(11):430–434.

35. Maddocks SE, Oyston PFC: Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiol (Reading, England) 2008, 154(Pt 12):3609–3623.

36. Li B, Wang F-Q, Wei D-Z: Cholesterol oxidase Chol is a critical enzyme that catalyzes the conversion of diosgenin to 4-ene-3-keto steroids in Streptomyces virginiae IBL-14. Applied Microbiol Biotechnol 2010, 85(5):1831–1838.

37. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994, 22(22):4673–4680.

38. Kamura K, Petersen D, Petersen N, Stecher G, Nei M, Kumar S: MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.

39. Parajuli N, Basset DB, Chan Lee H, Sching JK, Liou K: Genome analyses of Streptomyces peucetius ATCC 27952 for the identification and comparison of cytochrome P450 complement with other Streptomyces. Archives Biochem Biophys 2004, 425(2):233–241.

40. Yu Y, Yan L, Li Z, Wang Y: New biological activities and bioconversion reactions of steroids by Streptomyces virgulinae IBL-14. Environ Sci Technol 2009, 43(15):5967–5974.

41. Toya N, Kubota K, Tsuchiya Y, Itoh N, Fujigaki T, Shindo Y, Kikuchi N: Structure of cytochrome P450 MoxA from Nonomuraea reacticatenica (CYP105). Biochem Biophys Res Commun 2007, 361(4):876–882.

42. Rupasinghe S, Schuler MA, Kagawa N, Yuan H, Lee L, Zhao B, Kelly SL, Waterman MR, Lamb DC: The cytochrome P450 gamete fertility gene CYP157 does not contain EXXR in the Kel-helix reducing the absolute conserved P450 residues to a single cysteine. FEBS Lett 2006, 580(27):6338–6342.

43. Taylor M, Lamb DC, Cannel R, Dawson M, Kelly SL: Cytochrome P45001D1(CYP105D1) from Streptomyces griseus: heterologous expression, activity, and activation effects of multiple xenobiotics. Biochem Biophys Res Commun 1999, 263(3):836–842.

44. Watanabe K, Akamatsu Y: Identification and functional analysis of a P-450-like mycinamicin biosynthesis gene using a novel microcomposonosa- Escherichia coli shuttle cosmid vector. Mol Gen Metab 1994, 245(4):456–464.

45. Zhang X, Gallegos MT, Brennan R, Tobes R, The TetR Family of transcriptional repressors. Microb Mol Biol Rev 2005, 69(2):326–356.

46. Schägger H, von Jagow G: Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Analytical Biochem 1987, 150(1):356–362.