Accepted Manuscript

Title: Production of Dammarane-type Sapogenins in Rice by Expressing the Dammarenediol-II Synthase Gene from *Panax ginseng* C.A.Mey.

Author: Zhiwei Huang Juncheng Lin Zuxin Cheng Ming Xu Xinying Huang Zhijian Yang Jingui Zheng

PII: S0168-9452(15)30027-3
DOI: http://dx.doi.org/doi:10.1016/j.plantsci.2015.07.021
Reference: PSL 9248

To appear in: *Plant Science*

Received date: 5-4-2015
Revised date: 10-7-2015
Accepted date: 25-7-2015

Please cite this article as: Zhiwei Huang, Juncheng Lin, Zuxin Cheng, Ming Xu, Xinying Huang, Zhijian Yang, Jingui Zheng, Production of Dammarane-type Sapogenins in Rice by Expressing the Dammarenediol-II Synthase Gene from Panax ginseng C.A.Mey., Plant Science http://dx.doi.org/10.1016/j.plantsci.2015.07.021

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Production of Dammarane-type Sapogenins in Rice by Expressing the Dammarenediol-II Synthase Gene from *Panax ginseng* C.A.Mey.

Zhiwei Huang[a,b], Juncheng Lin[b], Zuxin Cheng[b], Ming Xu[b], Xinying Huang[b], Zhijian Yang[b], Jingui Zheng[b,*]

[a] College of Food Science, Fujian Agriculture and Forestry University, 15 Shangxiadian Road, CangShan District, Fuzhou Fujian 350002, China

[b] Agricultural Product Quality Institute, Fujian Agriculture and Forestry University, 15 Shangxiadian Road, CangShan District, Fuzhou Fujian 350002, China

Zhiwei Huang: hzwfau@163.com

Juncheng Lin: linjc2011@126.com

Zuxin Cheng: chengzuxin@163.com

Ming Xu: xmfau@163.com

Xinying Huang: 764900262@qq.com

Zhijian Yang: yangzj41@163.com

Jingui Zheng: zjg231@126.com

*** Corresponding author.**

Phone and Fax: 0086-591-83789231.

Email: zjg231@126.com

Highlights

ACCEPTED MANUSCRIPT

- Rice cannot synthesize ginsenosides or sapogenins without DS and *β*AS.
- This is the first report on transforming ginseng *DS* gene into rice.
- The ginseng rice germplasm containing dammarane-type sapogenins was obtained.

ABSTRACT

Ginsenosides are the main active ingredients in Chinese medicinal ginseng; 2,3-oxidosqualene is a precursor metabolite to ginsenosides that is present in rice. Because rice lacks a key rate-limiting enzyme (dammarenediol-II synthase, DS), rice cannot synthesize dammarane-type ginsenosides. In this study, the ginseng (*Panax ginseng* CA Mey.) *DS* gene (GenBank: AB265170.1) was transformed into rice using agrobacterium, and 64 rice transgenic plants were produced. The Transfer-DNA (T-DNA) insertion sites in homozygous lines of the T_2 generation were determined by using high-efficiency thermal asymmetric interlaced PCR (hiTAIL-PCR) and differed in all tested lines. One to two copies of the T-DNA were present in each transformant, and real-time PCR and Western blotting showed that the transformed *DS* gene could be transcribed and highly expressed. High performance liquid chromatography (HPLC) analysis showed that the dammarane-type sapogenin 20(S)–protopanaxadiol (PPD) content was 0.35-0.59 mg/g dw and the dammarane-type sapogenin 20(S)–protopanaxatriol (PPT) content was 0.23-0.43 mg/g dw in the transgenic rice. LC/MS analysis confirmed production of PPD and PPT. These results indicate that a new "ginseng rice" germplasm containing dammarane-type sapogenins has been successfully developed by transforming the ginseng *DS* gene into rice.

Keywords: Dammarenediol-II synthase; Genetic transformation; Ginsenosides; Rice; Sapogenins

ACCEPTED MANUSCRIPT

1. INTRODUCTION

Rice is one of the world's principal food crops, and the hybridization of rice has increased rice yields dramatically. The demand for high-quality rice has also gradually gone up as the standard of living has increased. The development of transgenic technology provides a means of efficiently improving rice quality through the creation of rice germplasm that can synthesize active ingredients. For example, Ye *et al.* transformed key enzymes in the β-carotene biosynthesis pathway from daffodil and bacteria into rice, and by expressing the genes specifically in the endosperm, they produced "golden rice". With golden rice Vitamin A deficiencies can be prevented, as the β-carotene content reached substantial levels [1]. Paine *et al.* transformed the gene encoding phytoene synthase (PSY) in the pathway of β-carotene synthesis from maize instead of bacteria into rice and produced the second generation of "golden rice" with improved β-carotene content [2]. Lee *et al.* transformed the encoding gene of sesame methionine-rich 2S albumin into rice and produced "sesame nutrition rice", which substantially increased the levels of methionine, an essential and limiting amino acid [3].

P. ginseng, which belongs to *Panax* within Araliaceae family, is a rare traditional Chinese herbal medicine, and its active ingredient is ginsenosides [4]. All ginseng plants in the genus have relatively high medicinal values, and have roles in adaptation, immunity, tumor prevention, anti-atherosclerosis, and myocardial protection [4, 5]. Ginsenosides are triterpenoid saponins of plant secondary metabolites, i.e. the product of triterpenoid saponins biosynthesis branch in isoprenoid pathway (Fig. 1) [6–8].

As shown in Fig. 1, triterpenoid saponins are formed by the cyclization of squalene. The synthesis of squalene is a branch point in the central isoprenoid pathway entering the triterpenoid saponins biosynthesis branch [7, 9, 10]. Squalene epoxidase (SQE) catalyzes squalene to produce

ACCEPTED MANUSCRIPT

2,3-oxidosqualene, causing a linear shift from squalene to cyclic triterpenoids [11, 12]. 2,3-oxidosqualene cyclases (OSCs) further catalyze the cyclization of 2,3-oxidosqualene to form dammarane-type tetracyclic triterpenoids, oleanane-type pentacyclic triterpenoids, or other compounds. Then the triterpenoid skeleton forms triterpenoid saponins of many different structures through oxidation, replacement, glycosylation, and other forms of chemical modification. OSCs are the rate-limiting enzymes of the triterpenoid saponin biosynthesis pathway [6, 11, 13–15]. Although the synthetic precursor of both ginsenoside and rice sterols is 2,3-oxidosqualene, the type and catalytic function of OSCs are different between ginseng genus and rice [7, 9]. The OSCs of the ginseng genus are mainly dammarenediol-II synthase (DS) and *β*-amyrin synthase (*β*AS), which catalyze 2,3-oxidosqualene to produce dammarane-type and oleanane-type substances, respectively (specifically, dammarenediol-II and *β*-amyrin). Rice OSC is a cycloartenol synthase (CS), which catalyzes 2,3-oxidosqualene to produce cycloartenol [6, 11, 14], but due to the absence of DS and *β*AS, although rice contains ginsenosides precursor 2,3-oxidosqualene, it cannot synthesize ginsenosides.

The necessary knowledge base of ginsenoside synthesis is now present and provides an opportunity to improve staple rice quality using ginsenoside biosynthetic genes to allow the production of saponins. The *DS* gene has been made available through cloning in the ginseng genus, such as *P. ginseng* (GenBank: AB265170.1, GU183405.1, JN596111.1) [16–17], *P. quinquefolius* (GenBank: GU997679.1), *P. notoginseng* (GenBank: GU997680.1), etc. In the present work the ginseng *DS* gene was transformed into rice allowing the synthesis of dammarane-type sapogenins, creating a new germplasm called "ginseng rice" and providing germplasm resources for future germplasm improvements.

ACCEPTED MANUSCRIPT

2. MATERIALS AND METHODS

2.1. Materials

Rice cultivar 'Shuhui 527', *Escherichia coli* (*E. coli*) strain DH5α, *Agrobacterium tumefaciens* (*A. tumefaciens*) strain LBA4404, plasmids pMD-Gt1-AmA1, pBlue-Ubi, and pCD-AMA1-hpt are kept by the Agricultural Product Quality Institute of Fujian Agriculture and Forestry University.

In order to facilitate the subsequent construction of plant expression vector without changing the amino acid sequence of ginseng *DS* gene (GenBank: AB265170.1), the recognition sites of *Bam*H I, *Sac* I, *Hind* III, *Sma* I, and *Eco*R I were removed from the coding region of the *DS* gene. The optimized *DS* gene was designated as *OPDS* (2310 bp in length). The recognition sites of *Bam*H I and *Sac* I were added to the 5′ end and 3′ end of *OPDS*, respectively. The whole *OPDS* gene was synthesized and linked to the pMD®18-T Simple Vector (Takara, D103A) by Shanghai Invitrogen Biotechnology Co., and the constructed plasmid was designated as pMD-DS.

2.2. Construction of binary T-DNA plant expression vector containing *OPDS*

The main process is as follows [18]: After double digesting *Bam*H I and *Sac* I in the plasmid pMD-DS to release the *OPDS* gene fragment, the *OPDS* gene fragment was ligated to a *Bam*H I and *Sac* I double-digested pMD-Gt1-AmA1 vector to form pMD-Gt1-DS. After double digesting *Hind* III and *Bam*H I in the plasmid pMD-DS to release the Ubi promoter fragment, the Ubi promoter fragment was ligated to a *Hind* III and *Bam*H I double-digested pMD-Gt1-DS vector to form plasmid pMD-Ubi-DS. After double digesting *Hind* III and *Sac* I in the plasmid pMD-Ubi-DS to release the Ubi promoter-OPDS fragment, the Ubi promoter-OPDS fragment

ACCEPTED MANUSCRIPT

was recovered and ligated to *Hind* III and *Sac* I double-digested plasmid pCD-AMA1-hpt to form binary vector pCD-DS-hpt (Fig. 2). This binary vector was assessed using restriction endonuclease and the target gene in the plasmid was confirmed by sequencing.

2.3. Transformation of recombinant vector containing *OPDS* into rice

Inflorescences of rice 'Shuhui 527' were taken 10–15 d after pollination. Embryogenic calli and subculture calli were induced every 15 d with fresh medium using the method of Chen *et al*. [19]. Using the freeze-thaw method, the binary expression vector pCD-DS-hpt was transformed into *A. tumefaciens* LBA4404 [20], which was then used to transform embryogenic calli. Co-culture of calli with agrobacterium, screening of resistant calli, induction of differentiation of cultured calli, and rooting of resistant seedlings, etc. were conducted using a method described by Burkhardt *et al*. [21]. The basic medium used for genetic transformation of rice was NB medium (pH 5.8). The medium used for *Agrobacterium* infection was AAM medium (pH 5.2) [21]. In screening for resistant calli, hygromycin B (Hyg) was used at a concentration of 25–50 mg/L (the initial 25 mg/L in the first screening was gradually increased to 50 mg/L in subsequent generations).

2.4. PCR identification of transgenic rice

Genomic DNA was extracted from trefoil stage leaves of resistant seedlings, the T_1 and T_2 generation lines of transgenic rice; 2 × *Eco* Taq PCR SupperMix (TransGen Biotech, AS151) was used to detect *OPDS* and hygromycin resistance gene (*HPT*) in transgenic rice plants, primers were DSF1 (5′-TGCTGATATTGTGGTCGAGAGG-3′), DSR1 (5′-TTTGAGTTGCTGGTGCTTTGGC-3′) and Hpt-1 (5′-TACACAGCCATCGGTCCAGA-3′), Hpt-2 (5′-TAGGAGGGCGTGGATATGT C-3′). The PCR reaction system had a total volume of

ACCEPTED MANUSCRIPT

10 μl containing 5.0 μl 2 × *Eco* Taq PCR SupperMix (+ dye), 0.2 μl each of forward and reverse primers (20 μM), 0.8 μl DNA template, and 3.8 μl ddH_2O. The PCR reaction conditions were 94°C for 5 min at first; followed by 35 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 60 s/kb; and a final extension at 72°C for 10 min.

2.5. hiTAIL-PCR analysis of T-DNA insertion sites

Leaf genomic DNA was extracted from trefoil stage of the rice homozygous lines of the T_2 generation (derived from the transgenic plants obtained by transformation in different batches). Using the method described by Liu *et al*., *Premix Ex Taq*® Version 2.0 (Takara, D335A) was used to conduct hiTAIL-PCR (high-efficiency thermal asymmetric interlaced PCR) to amplify genomic sequences flanking T-DNA insertion sites in transgenic rice, primers for the 1st PCR were LAD1-1 (5′-ACGATGGACTCCAGAGCGGCCGC(G/C/A)N(G/C/A)NNNGGAA-3′), LAD1-3 (5′-ACGATGGACTCCAGAGCGGCCGC(G/C/A)(G/C/A)N(G/C/A)NNNCCAA-3′) and RB-0b (5′-CGTGACTGGGAAAACCCTGGCGTT-3′), primers for 2nd PCR were AC1 (5′-ACGATGGACTCCAGAG-3′) and RB-1b (5′-ACGATGGACTCCAGTCCGGCCCAACTT AATCGCCTTGCAGCACATC-3′), primers for 3rd PCR were AC1 and RB-2b (5′-GAAGAGG CCCGCACCGATCGCCCTT-3′) [22]. The clear bands that had good specificity and were greater than 200 bp in 3rd PCR product were sequenced. PCR confirmation of the T-DNA flanking sequences was conducted by using 2× *Eco* Taq PCR SuperMix (TransGen Biotech, AS151). The forward primer was RB-3b (5'-GATCGCCCTTCCCAACAGTTGC-3'), and the reverse primers were D6R (5'-GCAAGAACTGGAATGGCTGG-3'), D28R (5'-GGATCTGTAGCTCACCGAGG-3'), D35R1 (5'-CCCGCAATGGTAAAGTATGGTG-3'), and D35R2 (5'-ATTCTCCTATTGTCGTCGTTCATTC-3'). The PCR reaction system had a total

ACCEPTED MANUSCRIPT

volume of 50 μl containing 25.0 μl 2× *Eco* Taq PCR SuperMix, 1.0 μl each of the forward and reverse primers (20 μM), 2.0 μl DNA template, and 21.0 μl ddH_2O. The PCR reaction conditions were 94°C for 5 min; followed by 35 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 60 s/kb; and a final extension at 72°C for 10 min. The PCR products were then sequenced. The sequences obtained were aligned with the rice "9311" genome sequences (http://rise2.genomics.org.cn/page/rice/index.jsp) using BLASTN to determine the T-DNA insertion sites.

2.6. qRT-PCR of *OPDS* expression

TRIzol (Invitrogen, 15596026) was used to extract total RNA from trefoil stage leaves of homozygous rice lines of the T_2 generation and residual DNA was removed from RNA samples using DNA enzyme (Takara, D2210). A PrimeScript® RT Reagent Kit (Takara, DRR037A) was used to reverse-transcribe total RNA to synthesize first-strand cDNA. The conditions for reverse transcription reaction were as follows: 37°C for 15 min and 85°C for 5 s. The PCR reaction system had a total volume of 20 μl including 10.0 μl SYBR® *Premix Ex Taq*™ II (2 ×), 0.8 μl each forward and reverse primer (10 μM), 0.4 μl ROX Reference Dye II (50 ×), 4.0 μl first-strand cDNA (diluted 5 times), 4.0 μl ddH_2O. PCR reaction conditions were 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, 60°C for 34 s. SYBR® *Premix Ex Taq*™ II (Takara, RR041A) and the ABI 7500 Real-Time PCR System were used for *OPDS* real-time fluorescence quantitative PCR (qRT-PCR), primers were qDSF (5′-CCGAGTCAGACGACGGAAA-3′) and qDSR (5′-GGAAACGAGGAAGGGAACAAC-3′); *ACTIN* served as a reference gene: primers were qACTF (5′-CATCTTGGCATCTCTCAGCAC-3′) and qACTR (5′-AACTTTGTCCACGCTAATGAA-3′). Three replicates were done and the relative expression of the *OPDS* gene was $RQ = 2^{-\Delta\Delta Ct}$.

ACCEPTED MANUSCRIPT

2.7. Analysis of DS protein by Western blotting

RIPA lysis buffer (Beijing Ding Guo Changsheng Biotechnology Company, Ltd., China) was used to extract total protein from mature seeds of some of the T_2 generation homozygous rice lines and a Total Protein Assay Kit (biuret method, Shanghai Rongsheng Biotechnology Company, China) was used to measure concentrations of total proteins. Then 20–40 μg total protein was taken and Mini-PRO TEAN® 3 Cell (BioRad) was used for SDS-PAGE (Sodium dodecyl sulphate polyacrylamide gel electrophoresis), Trans-Blot SD semi-dry electrophoretic transfer cell (BioRad) was used to transfer proteins to a nitrocellulose membrane. Then Western blotting was performed to analyze the targeted protein [18]. A rabbit polyclonal antibody against dammarenediol-II synthase (primary antibody) was prepared by GL Biochem Co., Ltd. (Shanghai, China).

2.8. HPLC and LC/MS analysis of sapogenins and dammarenediol-II content in transgenic rice

Methanol was used as a solvent to reflux extract (12 h) sapogenins and dammarenediol-II from mature seeds of homozygous T_2 rice. LC-20A high-performance liquid chromatography (Shimadzu) was used to assess levels of sapogenins and dammarenediol-II with the Hypersil ODS2 C18 column (250 nm × 4.6 nm, 5 μm) (Dalian Elite Company, China). The chromatographic conditions were as follows: mobile phase methanol - water (85:15, v/v), the flow rate was 1.0 ml/min, the column temperature was 30°C, and the detection wavelength was 203 nm. Standards 20(S)-protopanaxadiol (PPD) (batch number: 111747-200501) and 20(S)–protopanaxatriol (PPT) (batch number: 111755-200601) were purchased from the Chinese Academy of Food and Drug Testing. Dammarenediol-II (DAD, CAS: 14351-29-2) was purchased from BioBioPha Co., Ltd, China.

ACCEPTED MANUSCRIPT

In order to determine PPD and PPT, methanol extracts (20 μl) of mature seeds of homozygous T_2 rice were analyzed by LC/MS using a 1260 HPLC system (Agilent) coupled to a 6520B Q-TOF mass spectrometer (Agilent) with a dual electrospray ionization (ESI) interface. Data acquisition and processing were done with the Agilent MassHunter Workstation Data Acquisition / Agilent MassHunter Qualitative Analysis B.04.00 software. For chromatographic separation, a Hypersil ODS2 C18 column (250 nm × 4.6 nm, 5 μm) (Dalian Elite Company, China) was used. The chromatographic conditions were as follows: mobile phase methanol - water (85:15, v/v), the flow rate 1.0 ml/min, the column temperature 30°C, and the detection wavelength 203 nm. The MS operating conditions were as follows: all spectra were obtained in positive mode over an *m*/*z* range of 100–1200; dry gas flow, 8.0 L/min; dry temperature, 350°C; nebulizer pressure, 35 psi; probe voltage +3500 V.

3. RESULTS

3.1. Development of transgenic rice

The binary plant expression vector pCD-DS-hpt was subjected to *Hind* III/*Sac* I and *Bam*H I/*Sac* digestion, as shown in Fig. 3. The sizes of digestion products were as expected. The target gene in pCD-DS-hpt was also sequenced and showed no mutations. These findings indicated that the binary plant expression vector pCD-DS-hpt was successfully constructed.

This binary vector was transformed into Rice ‘Shuhui 527’ and 410 anti-Hyg rice plants (which could grow on the screening medium containing 50 mg/L Hyg) to obtain the T_0 generation. Using genomic DNA from resistant rice seedlings as templates, the *HPT* gene and *OPDS* gene were detected by PCR (Fig. 4). The results showed that all resistant seedlings contained *HPT*; 64 plants contained *OPDS*, and the positive rate was 15.6%. There was not a significant visible difference in agronomic traits and appearance between the sprigs of transgenic

ACCEPTED MANUSCRIPT

and receiver "Shuhui527" rice. The homozygous lines of the T_2 generation rice were derived from the positive T_0 lines which were obtained by PCR detection (Fig. S1).

3.2. hiTAIL-PCR confirmation of T-DNA insertion sites

hiTAIL-PCR can be used to obtain genomic fragments flanking T-DNA insertion sites with three rounds of PCR amplification, and is a convenient way to locate insertion sites [22]. hiTAIL-PCR was conducted using genomic DNA of T_2 generation homozygous rice lines as templates. This produced clear bands (labeled by arrow, Fig. 5) with good specificity and greater than 200 bp in 3rd PCR products from three lines (Fig. 5). The 3rd PCR products of other lines only gave vague bands on the gel or were smaller than 200 bp. After Sanger sequencing, the resulting sequences (GenBank: KP687750, KR055668, KR055669, KR055670) were aligned with rice “9311” genome sequences (Fig. S2, Fig. S3 and Fig. S4), and confirmed by PCR (Fig. S5). We found that the T-DNA containing *OPDS* in transgenic rice lines D6 and D28 was a single-copy insertion, while line D35 had two copies of the transgene. The insertions in these three lines were located at 32503243 bp on the third chromosome, 21854153 bp on the first chromosome, and 33629139 bp and 33632516 bp on the second chromosome, respectively.

3.3. Over-expression of exogenous *OPDS* gene in transgenic rice

qRT-PCR was conducted to analyze the expression of the *OPDS* gene in homozygous rice lines of the T_2 generation. The results showed that the *OPDS* gene can be highly transcribed in the T_2 lines D6, D28, D29, and D35. The relative expression was between 5589.2 and 6417.9 (Fig. 6, D45 was a negative line). There was not a significant difference in the expression of *OPDS* between the four homozygous lines.

ACCEPTED MANUSCRIPT

3.4. Western blot analysis of the DS protein

Total protein was extracted from the seeds of homozygous rice lines in the T_2 generation. After separation by SDS-PAGE, rabbit antiserum against DS was used as the primary antibody and horseradish peroxidase-labeled goat anti-rabbit as secondary antibody. DS protein was assessed by Western blotting. The results showed that the rice samples and the positive control (prokaryotic expression product of *OPDS*) had a specific hybridization band of about 88 kDa, as expected, while the negative sample D45 line did not (Fig. 7). These results indicate that seeds of *OPDS* transgenic rice successfully expressed dammarenediol-II synthase.

3.5. HPLC and LC/MS analysis of sapogenins and DAD content in transgenic rice

Using PPD, PPT and DAD as the standards, the levels of sapogenins and dammarenediol-II were assessed in mature seeds of homozygous rice lines in the T_2 generation by HPLC (Fig. 8). The PPD content was between 0.35 and 0.59 mg/g dw, PPT content was between 0.23 and 0.43 mg/g dw, and DAD content was between 0.08 and 0.44 mg/g dw in T_2 generation lines D6, D28, D29, and D35 (Fig. 9). LC/MS analysis of rice seed extraction confirmed the production of PPD (Fig. 10) and PPT (Fig. 11).

4. DISCUSSIONS

The genomes of the three homozygous rice lines, of which the genome sequences flanking T-DNA insertion sites were identified, each contained 1–2 copies of *OPDS*. This is consistent with results of previous studies on transgenic plants: the copy number of T-DNA insertion was low, the average was 1.2–2.0 copies, and most had one copy [23, 24]. In addition, T-DNA insertions are relatively random and insertion sites can vary across the genome [25, 26], although sometimes a preference is observed. For example, Zhang *et al.* analyzed the flanking sequences

ACCEPTED MANUSCRIPT

of 13804 T-DNA insertion sites and found that T-DNA preferred to insert into larger chromosomes in rice. While a few of them inserted into transposons, an insertion preference was seen 1 kb upstream and 500 bp downstream of coding region, but not within the coding region [27].

Because wild ginseng is becoming less abundant in its natural habitat, ginsenosides are mainly extracted from cultivated ginseng. But long periods of artificial cultivation coupled with obstacles in continuous cropping and other matters resulted in the failure to meet the community's increasing demand for higher ginsenoside production levels. New sources of medical-quality ginsenosides need to be found [4, 5]. Scientists are attempting to synthesize ginsenosides. On the basis of cloning and characterization of a key enzyme in the synthesis of PPD, Han *et al.* constructed a yeast cell capable of producing PPD [28]. Later, cytochrome P450 (CYP716A53v2) was found to catalyze PPD to produce PPT [29]. Dai *et al* constructed a "ginseng yeast" which can simultaneously produce three kinds of sapogenins, oleanolic acid, PPD and PPT, with the yields 21.4 mg/L, 17.2 mg/L and 15.9 mg/L, respectively [30, 31]. The rare ginsenoside compound K (CK) is the active compound that can be detected in the blood after oral administration of ginseng or ginsenosides, but it has not been detected in ginsenosides [32]. Yan *et al.* found that *P. ginseng* UDP- glycosyltransferase (PgUGT1) could specifically catalyze the glycosylation of C-20S hydroxyl of dammarane tetracyclic triterpenoids, and when PgUGT1 was co-expressed with the PPD synthesis pathway in yeast cells, the rare ginsenoside CK could be synthesized [33].

Rice is one of the world's major food crops, engineering metabolic pathways into rice has become an important means of creating new germplasms. [1–3, 34, 35]. In this study, the exogenous *DS* gene, a key enzyme in the synthesis of ginsenosides, was transformed into rice allowing the production of dammarenediol-II synthase, and thus producing a new type of rice germplasm containing dammarane-type sapogenins. The *OPDS* rice lines produced in this study

ACCEPTED MANUSCRIPT

are new ginsenoside donors, and may provide germplasm for future research on "ginseng rice" varieties. These results may have a profound impact on the genetic breeding of functional rice and research and development (R&D) of ginsenosides as a medicinal source.

ABBREVIATIONS

<xps:span class=deft>DS</xps:span> <xps:span class=defd> Dammarenediol-II synthase; T-DNA: Transfer-DNA; HPLC: High performance liquid chromatography; PPD: 20 (S)-Protopanaxadiol; PPT: 20 (S)–Protopanaxatriol; DAD: Dammarenediol-II; ESI: Electrospray ionization; EIC: Extracted ion chromatogram; PSY: phytoene synthase; *P. ginseng*: *Panax ginseng*; HMG-CoA: 3-hydroxy-3-methylglutaryl coenzyme A; MVA: mevalonic acid; IPP: Isopentenyl pyrophosphate; DMAPP: Dimethylallyl pyrophosphate; GPP: Geranyl pyrophosphate; FPP: Farnesyl pyrophosphate; GGPP: geranylgeranyl pyrophosphate; HMGR: 3-Hydroxy-3-methylglutaryl-CoA reductase; IPPS: Isoprenyl diphosphate synthase; MC: monoterpene cyclases; FPPS: Famesyl diphosphate synthase; SC: sesquiterpene cyclases; GGPPS: geranylgeranyl diphosphate synthase; SQS: Squalene synthase; SQE: Squalene epoxidase; CS: Cycloartenol synthase; LS: Lanosterol synthase; αAS: α-Amyrin synthase; LUS: Lupeol synthase; βAS: β-Amyrin synthase; OSCs: 2,3-Oxidosqualene cyclases; *P. quinquefolius*: *Panax quinquefolius*; *P. notoginseng*: *Panax notoginseng*; *E. coli*: *Escherichia coli*; *A. tumefaciens*: *Agrobacterium tumefaciens*; polyA T: Terminator of polyA; *HPT*: Hygromycin phosphotransferase gene; 35S P: 35S promoter; NOS T: Terminator of nopaline synthase gene; Ubi P: Ubiquitin promoter; LB: Left border; RB: Right border; Hyg: Hygromycin B; hiTAIL-PCR: High-efficiency thermal asymmetric interlaced PCR; qRT-PCR: real-time fluorescence quantitative PCR; SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis; PC: Positive control; CK: compound K; PgUGT1: *P. ginseng*

ACCEPTED MANUSCRIPT

UDP-glycosyltransferase.</xps:span>

ACKNOWLEDGMENTS

This research was supported financially by the National Transgenic Organism New Variety Culture Key Project (2009ZX08001-032B), the National Key Technology R&D Program (2013BAD01B05), and the Major Science and Technology Projects of Fujian Province (2013NZ0002-2) in China.

AUTHORS' CONTRIBUTIONS

JZ designed the study; ZH performed the experiment, conducted the analysis, and wrote the paper; JL assisted in performing the experiments and analyzing the data; ZC, MX, XH and ZY assisted in performing the experiments. All authors read and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

References

[1] X.D. Ye, S. Al-Babili, A. Klöti, J. Zhang, P. Lucca, P. Beye, I. Potrykus. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science. 287 (2000), 303-305.

ACCEPTED MANUSCRIPT

[2] J.A. Paine, C.A. Shipton, S. Chaggar, R.M. Howells, M.J. Kennedy, G. Vernon, S.Y. Wright, E. Hinchliffe, J.L. Adams, A.L. Silverstone, R. Drake. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol. 23 (2005), 482-487.

[3] T.T.T. Lee, M. Chung, Y. Kao, C. Wang, L. Chen, J.T.C. Tzen. Specific expression of a sesame storage protein in transgenic rice bran. J Cereal Sci. 41 (2005), 23-29.

[4] F.Y. Li. Chinese ginseng and American ginseng. China Agricultural Science and Technology Press, Beijing, 2002. (in Chinese)

[5] T.K. Yun. Panax ginseng—a non-organ-specific cancer preventive? The Lancet. Oncology. 2 (2001), 49-55.

[6] K. Haralampidis, M. Trojanowska, A.E. Osbourn. Biosynthesis of triterpenoid saponins in plants. Advances in biochemical engineering/biotechnology. 75 (2002), 31.

[7] H.L. Hao, X.F. Zhu, Y.Z. Zeng. The biosynthesis and regulation of isoprenoids. Journal of Zhejiang Agricultural University (Agric.& Life Sci.). 28 (2002), 224-230. (in Chinese)

[8] Y.L. Liang, S.J. Zhao. Progress in understanding of ginsenoside biosynthesis. Plant Biology. 10 (2008), 415-421.

[9] J. Chappell. Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. Annual Review of Plant Physiology and Plant Molecular Biology (1995), 521-547.

[10] D.H. Chen, H.C. Ye, G.F. Li, Y. Liu. Advances in molecular biology of plant isoprenoid metabolic pathway. Acta Botanica Sinica. 42 (2000), 551-558. (in Chinese)

[11] H. Suzuki, L. Achnine, R. Xu, S.P.T. Matsuda, R.A. Dixon. A genomics approach to the early stages of triterpene saponin biosynthesis in Medicago truncatula. The Plant Journal. 32

ACCEPTED MANUSCRIPT

(2002), 1033-1048.

[12] J. Wu, J.J. Zhong. Production of ginseng and its bioactive components in plant cell culture: current technological and applied aspects. J Biotechnol. 68 (1999), 89-99.

[13] J.M. Gao. Phytochemistry. Science Press, Beijing, 2003. (in Chinese)

[14] K. Haralampidis, G. Bryan, X. Qi, K. Papadopoulou, S. Bakht, R. Melton, A. Osbourn. A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. P Natl Acad Sci Usa. 98 (2001), 13431-13436.

[15] J.M. Augustin, V. Kuzina, S.B. Andersen, S. Bak. Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry. 72 (2011), 435-457.

[16] P. Tansakul, M. Shibuya, T. Kushiro, Y. Ebizuka. Dammarenediol-II synthase, the first dedicated enzyme for ginsenoside biosynthesis, in Panax ginseng. Febs Lett. 580 (2006), 5143-5149.

[17] W. Hu, N. Liu, Y.H. Tian, L.X. Zhang. Molecular cloning, expression, purification, and functional characterization of dammarenediol synthase from Panax ginseng. BioMed research international. 2013 (2013), 285740.

[18] J. Sambrook, D.W. Russell, translated by P.T. Huang. Molecular Cloning: A Laboratory Manual, 3rd ed. Science Press, Beijing, 2002. (in Chinese)

[19] L. Chen, S. Zhang, R.N. Beachy, C.M. Fauquet. A protocol for consistent, largescale production of fertile transgenic rice plants. Plant Cell Rep (1998), 25-31.

[20] R. Hofgen, L. Willmitzer. Storage of competent cells for Agrobacterium transformation. Nucleic Acids Res. 16 (1988), 9877.

ACCEPTED MANUSCRIPT

[21] P.K. Burkhardt, P. Beyer, J. Wunn, A. Kloti, G.A. Armstrong, M. Schledz, J. von Lintig, I. Potrykus. Transgenic rice (Oryza sativa) endosperm expressing daffodil (Narcissus pseudonarcissus) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. The Plant journal. 11 (1997), 1071-1078.

[22] Y.G. Liu, Y.L. Chen. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. Biotechniques. 43 (2007), 649.

[23] J.S. Jeon, S. Lee, K.H. Jung, S.H. Jun, D.H. Jeong, J. Lee, C. Kim, S. Jang, K. Yang, J. Nam, K. An, M.J. Han, R.J. Sung, H.S. Choi, J.H. Yu, J.H. Choi, S.Y. Cho, S.S. Cha, S.I. Kim, G. An. T-DNA insertional mutagenesis for functional genomics in rice. The Plant journal. 22 (2000), 561-570.

[24] A. Krishnan, E. Guiderdoni, G. An, Y.C. Hsing, C.D. Han, M.C. Lee, S.M. Yu, N. Upadhyaya, S. Ramachandran, Q.F. Zhang, V. Sundaresan, H. Hirochika, H. Leung, A. Pereira. Mutant Resources in Rice for Functional Genomics of the Grasses. Plant Physiol. 149 (2009), 165-170.

[25] R. Azpiroz-Leehan, K.A. Feldmann. T-DNA insertion mutagenesis in Arabidopsis: going back and forth. Trends Genet. 13 (1997), 152-156.

[26] P.J. Krysan, J.C. Young, M.R. Sussman. T-DNA as an insertional mutagen in Arabidopsis. The Plant cell. 11 (1999), 2283-2290.

[27] J. Zhang, D. Guo, Y. Chang, C. You, X. Li, X. Dai, Q. Weng, J. Zhang, G. Chen, X. Li, H. Liu, B. Han, Q. Zhang, C. Wu. Non-random distribution of T-DNA insertions at various levels of the genome hierarchy as revealed by analyzing 13 804 T-DNA flanking sequences

ACCEPTED MANUSCRIPT

from an enhancer-trap mutant library. The Plant Journal. 49 (2007), 947-959.

[28] J.Y. Han, H.J. Kim, Y.S. Kwon, Y.E. Choi. The Cyt P450 Enzyme CYP716A47 Catalyzes the Formation of Protopanaxadiol from Dammarenediol-II During Ginsenoside Biosynthesis in Panax ginseng. Plant Cell Physiol. 52 (2011), 2062-2073.

[29] J.Y. Han, H.S. Hwang, S.W. Choi, H.J. Kim, Y.E. Choi. Cytochrome P450 CYP716A53v2 Catalyzes the Formation of Protopanaxatriol from Protopanaxadiol During Ginsenoside Biosynthesis in Panax Ginseng. Plant Cell Physiol. 53 (2012), 1535-1545.

[30] Z.B. Dai, Y. Liu, X.N. Zhang, M.Y. Shi, B.B. Wang, D. Wang, L.Q. Huang, X.L. Zhang. Metabolic engineering of Saccharomyces cerevisiae for production of ginsenosides. METABOLIC ENGINEERING. 20 (2013), 146-156.

[31] Z.B. Dai, B.B. Wang, Y. Liu, M.Y. Shi, D. Wang, X.N. Zhang, T. Liu, L.Q. Huang, X.L. Zhang. Producing aglycons of ginsenosides in bakers' yeast. Scientific Reports. 4 (2014), 3698.

[32] H. Hasegawa. Proof of the mysterious efficacy of ginseng: Basic and clinical trials: Metabolic activation of ginsenoside: Deglycosylation by intestinal bacteria and esterification with fatty acid. Journal of Pharmacological Sciences. 95 (2004), 153-157.

[33] X. Yan, Y. Fan, W. Wei, P.P. Wang, Q.F. Liu, Y.J. Wei, L. Zhang, G.P. Zhao, J.M. Yue, Z.H. Zhou. Production of bioactive ginsenoside compound K in metabolically engineered yeast. Cell Res. 24 (2014), 770-773.

[34] F. Goto, T. Yoshihara, N. Shigemoto, S. Toki, F. Takaiwa. Iron fortification of rice seed by the soybean ferritin gene. Nat Biotechnol. 17 (1999), 282-286.

ACCEPTED MANUSCRIPT

[35] P. Lucca, R. Hurrell, I. Potrykus. Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. Theor Appl Genet. 102 (2001), 392-397.

FIGURE LEGENDS

Fig. 1 The isoprenoid biosynthetic pathway in plants [6-8]**.** Intermediates: HMG-CoA, 3-hydroxy-3-methylglutaryl coenzyme A; MVA, mevalonic acid; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; GPP, geranyl pyrophosphate; FPP, farnesyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate. Enzymes: HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; IPPS, isoprenyl diphosphate synthase; MC, monoterpene cyclases; FPPS, famesyl diphosphate synthase; SC, sesquiterpene cyclases; GGPPS, geranylgeranyl diphosphate synthase; SQS, squalene synthase; SQE, squalene epoxidase; CS, cycloartenol synthase; LS, lanosterol synthase; αAS, α-amyrin synthase; LUS, lupeol synthase; DS, dammarenediol-II synthase; βAS, β-amyrin synthase.

Fig. **2 T-DNA region of the plant expression vector pCD-DS-hpt.** Intermediates: LB, left border; RB, right border; 35S polyA, terminator of *CaMV* 35S gene; *HPT*, hygromycin phosphotransferase gene, selectable marker gene; 35S P, 35S promoter; NOS T, terminator of nopaline synthase gene; Ubi P, Ubiquitin promoter.

ACCEPTED MANUSCRIPT

Fig. **3 Identification of recombinant plant expression vector pCD-DS-hpt by enzyme digestion.** (M) DNA Marker DL15000. (1) Double digested by *Hin*d III and *Sac* I. (2) Double digested by *Bam*H I and *Sac* I.

Fig. **4 PCR analyses for the presence of *HPT* (845 bp) and *OPDS* (637 bp) in transgenic rice plants.** (M) DNA Marker DL 2000. (1–12) Transgenic rice plants. (13) Positive control (vector containing *OPDS* gene). (14) Negative control (rice cultivar 'Shuhui 527').

Fig. **5 Amplification of T-DNA flanking sequences from transgenic rice plants by hiTAIL-PCR.** (M) DNA Marker DL 2000. (1, 3, 5) Products of the 2nd round of PCR. (2, 4, 6) Products of the 3rd round of PCR.

Fig. **6 Real-time fluorescence quantitative PCR analysis of the expression of *OPDS* in transgenic rice plants.** (D6, D28, D29, D35) Homozygous lines of T_2 generation rice. (D45) Negative line of T_2 generation rice.

Fig. 7 Detection of the DS protein (about 88 kDa) in transgenic rice by Western blotting. (PC) positive control (the prokaryotic expression products of *OPDS*).

Fig. **8 HPLC chromatograms of PPD, PPT and DAD.** (A) Mixture of standards of PPD, PPT, and DAD. (B) HPLC chromatograms of PPD, PPT, and DAD in transgenic rice plants. (C) Negative control (the extraction of non-transgenic rice).

Fig. **9 PPD, PPT, and DAD production as analyzed by HPLC.** (ck) negative control (the extraction of non-transgenic rice).

Fig. **10 Identification of PPD production by LC/MS.** (A) The extracted ion chromatograms (EICs) from the LC/MS analysis of PPD standard. (B) The EICs from the LC/MS analysis of PPD extraction from transgenic rice. (C) Mass spectra of PPD standard. (D) Mass spectra of PPD extraction from transgenic rice.

ACCEPTED MANUSCRIPT

Fig. **11 Identification of PPT production by LC/MS.** (A) The EICs from the LC/MS analysis of PPT standard. (B) The EICs from the LC/MS analysis of PPT extraction from transgenic rice. (C) Mass spectra of PPT standard. (D) Mass spectra of PPT extraction from transgenic rice.

SUPPLEMENTARY INFORMATION

Fig. S1 PCR analyses for the presence of *HPT* (845 bp) and *OPDS* (637 bp) in T_2 generation. (M) DNA Marker DL 2000. (1–6) Homozygous lines of T_2 generation. (7) Positive control (vector containing *OPDS* gene). (8) Negative control (rice cultivar 'Shuhui 527').

Fig. **S2 Alignment of T-DNA flanking sequence (GenBank: KP687750) from line D6 with the reference sequence.**

Fig. **S3 Alignment of T-DNA flanking sequence (GenBank: KR055668) from line D28 with the reference sequence.**

Fig. **S4 Alignments of T-DNA flanking sequences (GenBank: KR055669, KR055670) from line D35 with the reference sequence.**

Fig. **S5 PCR analysis of T-DNA flanking sequences.** (M) DNA Marker DL 2000. (1-4) PCR products of the T-DNA flanking sequences.

ACCEPTED MANUSCRIPT

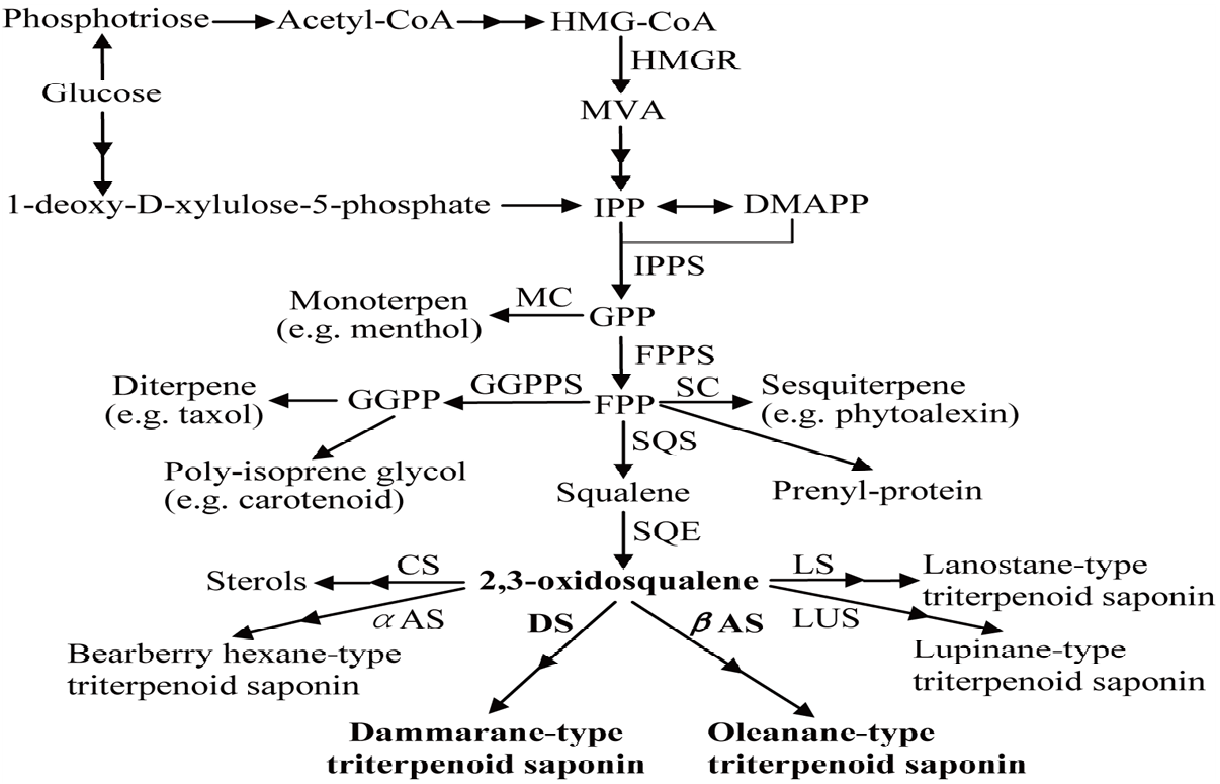


Fig1 .

ACCEPTED MANUSCRIPT

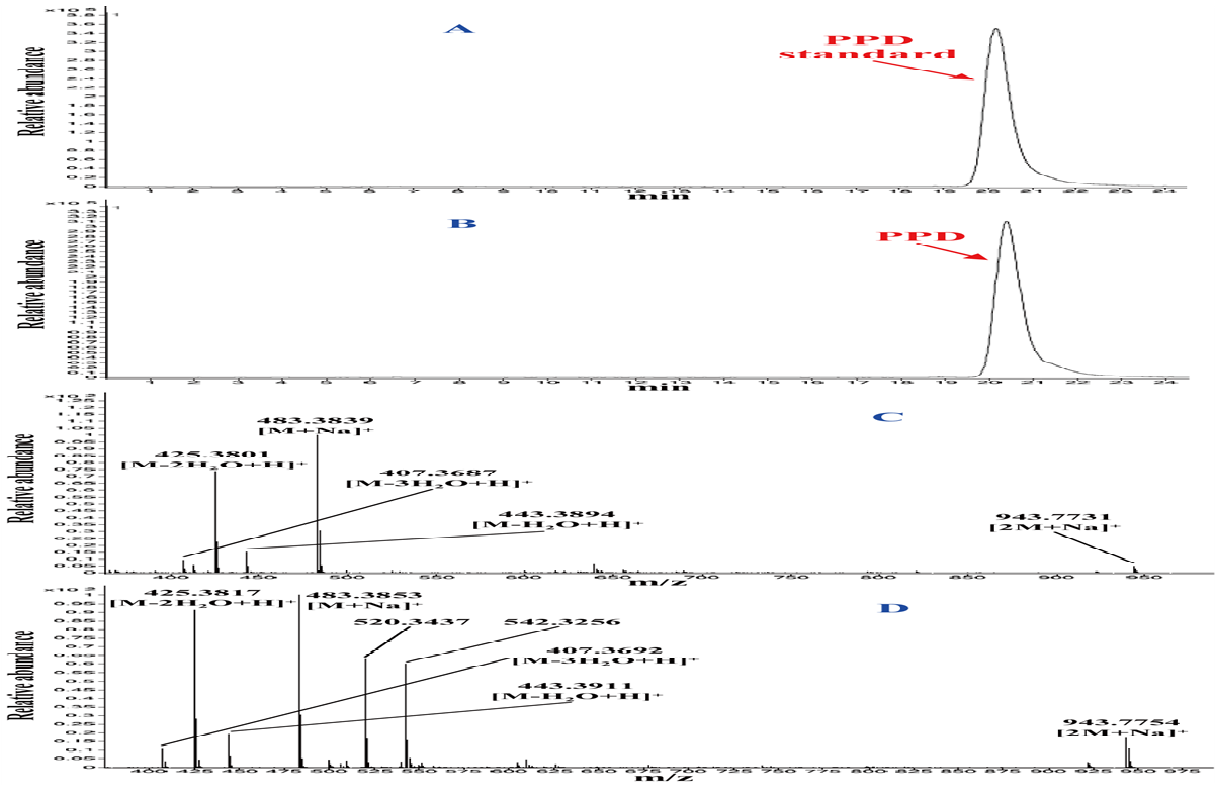


Fig10 .

ACCEPTED MANUSCRIPT

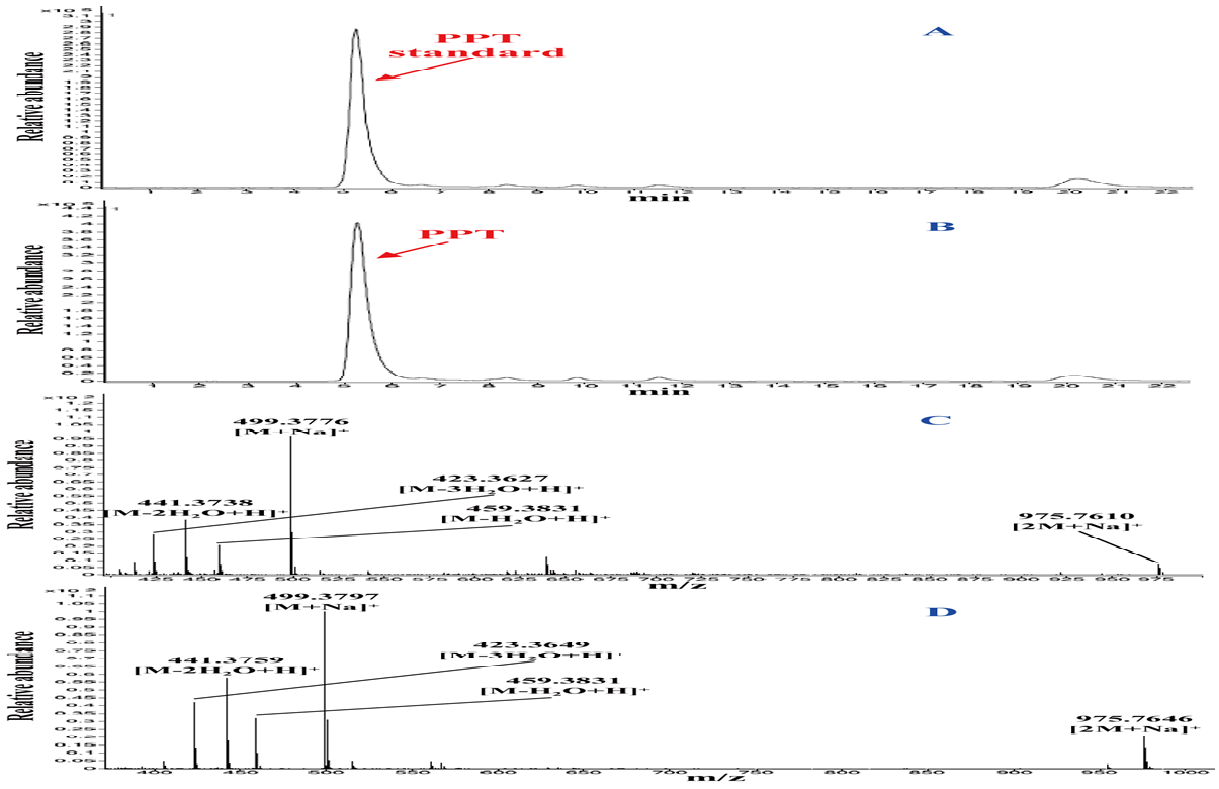


Fig11 .

ACCEPTED MANUSCRIPT

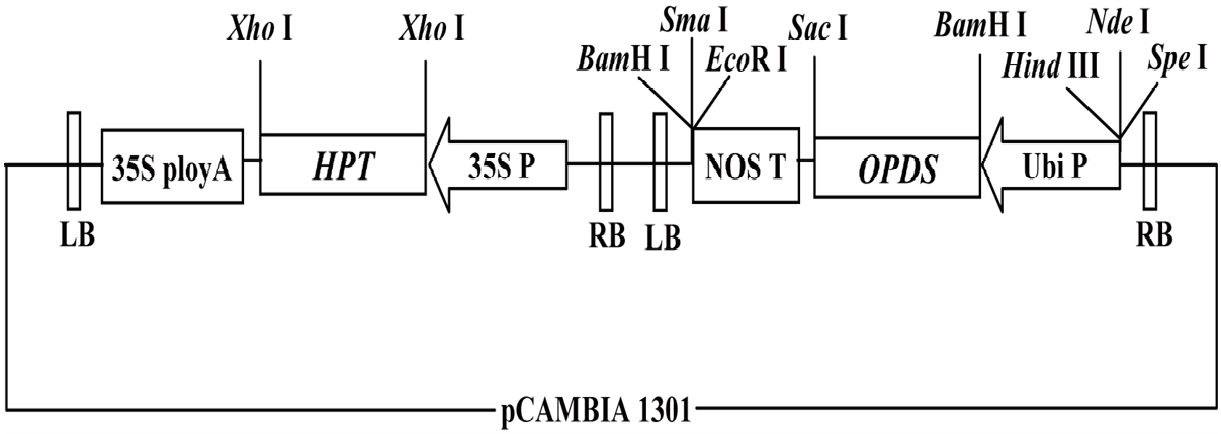


Fig2 .

ACCEPTED MANUSCRIPT

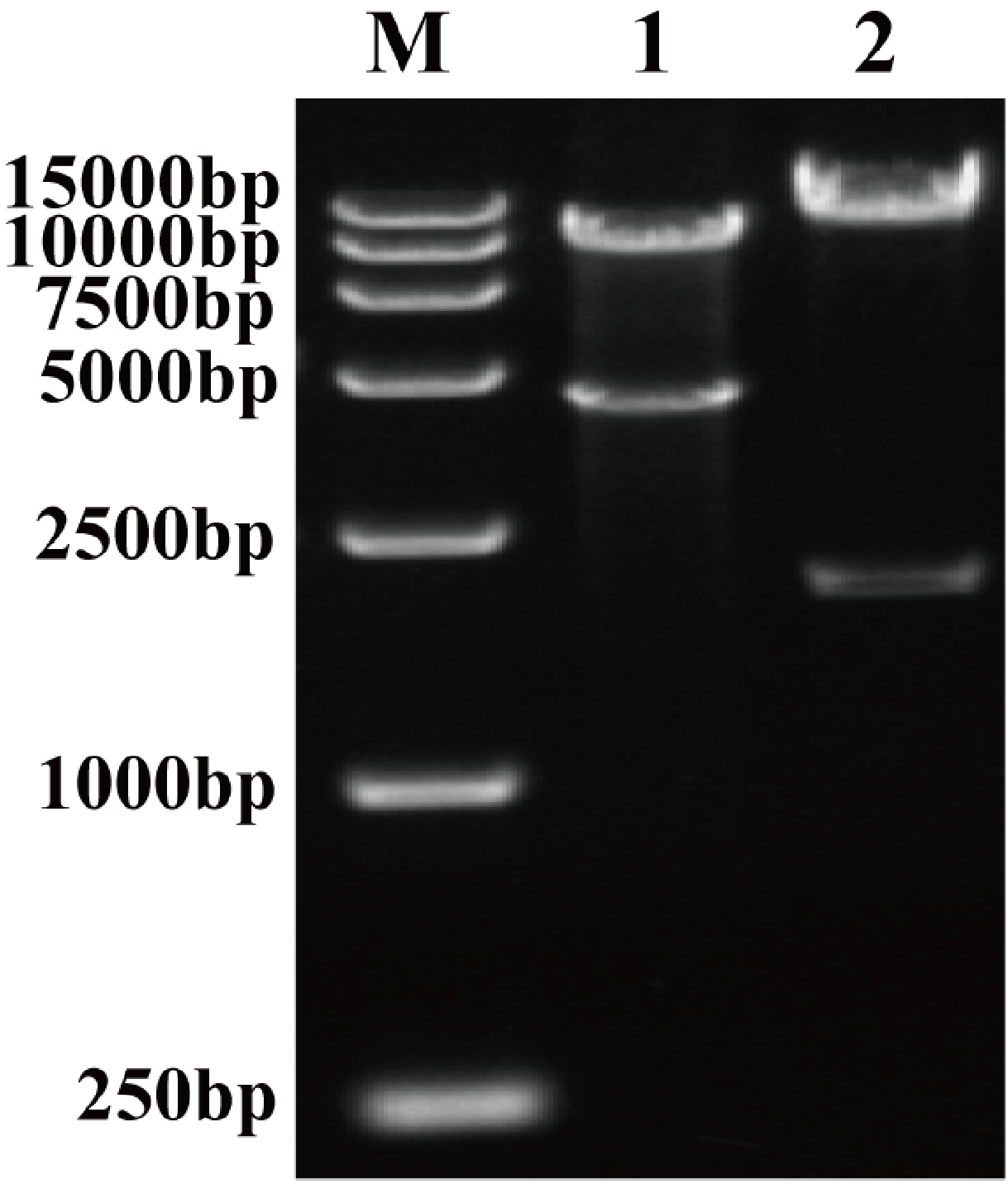


Fig3 .

ACCEPTED MANUSCRIPT

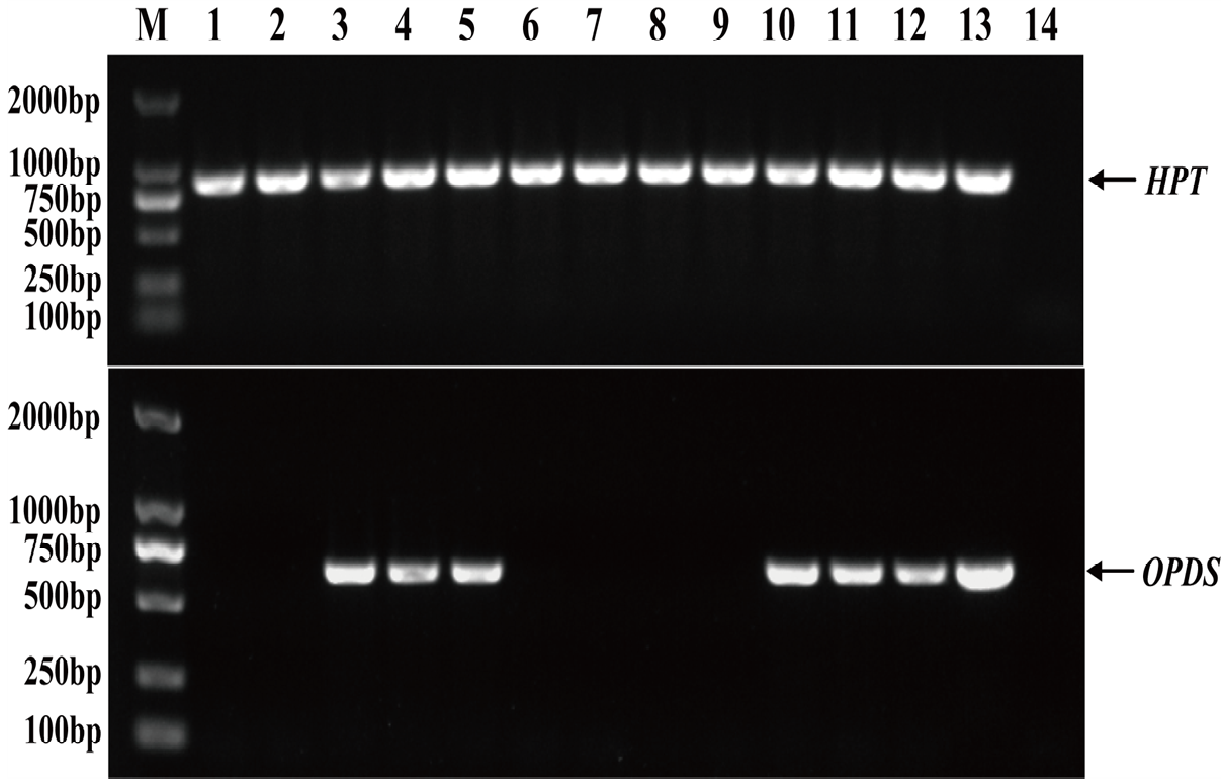


Fig4 .

ACCEPTED MANUSCRIPT

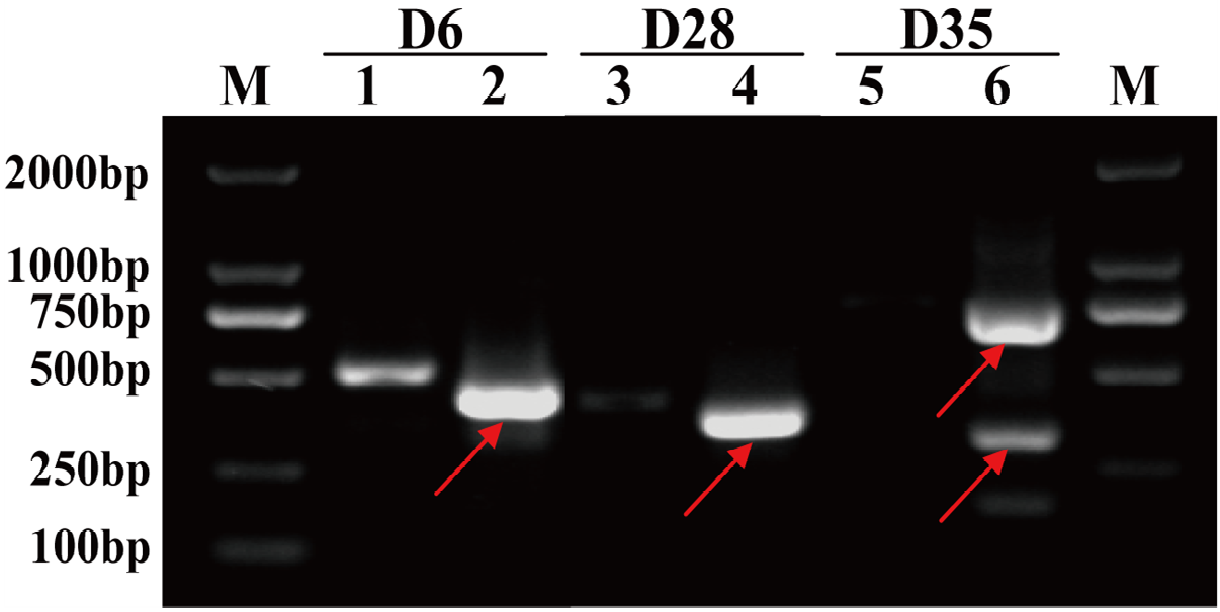


Fig5 .

ACCEPTED MANUSCRIPT

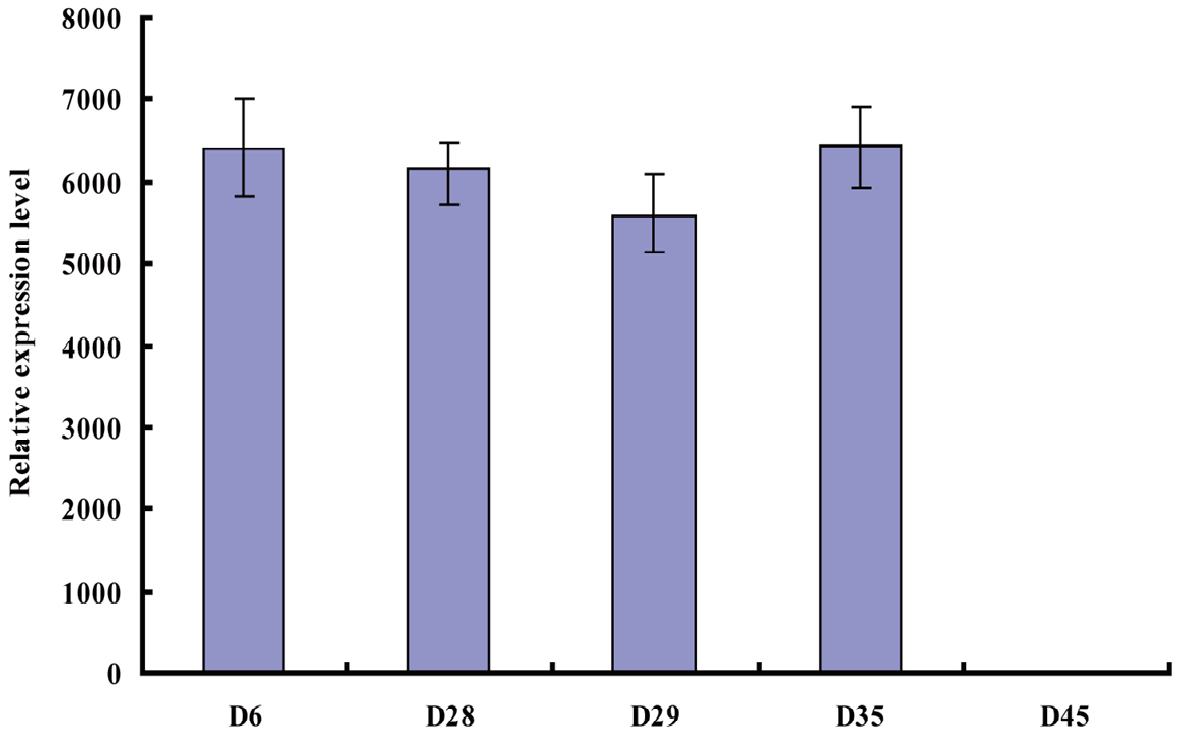


Fig6 .

ACCEPTED MANUSCRIPT

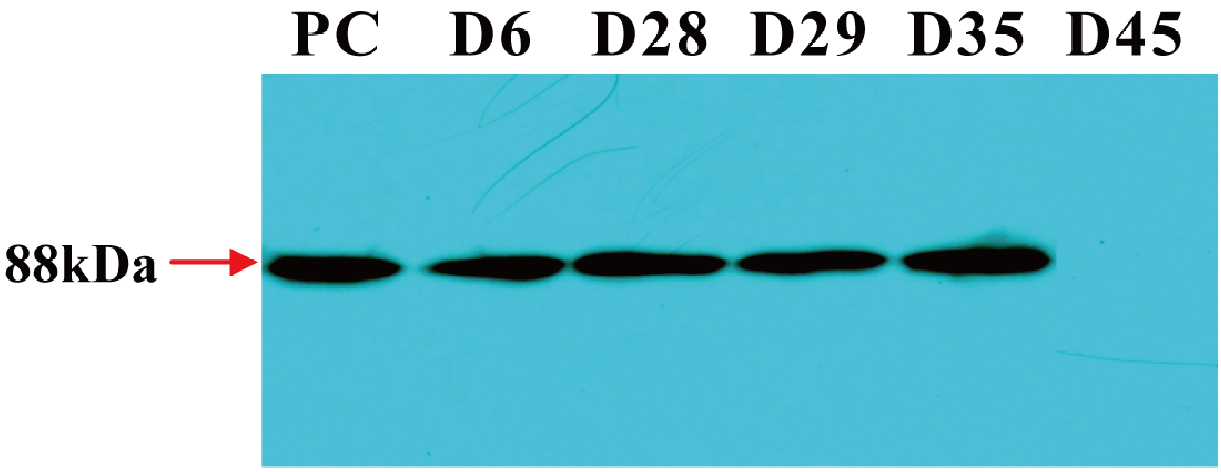


Fig7 .

ACCEPTED MANUSCRIPT

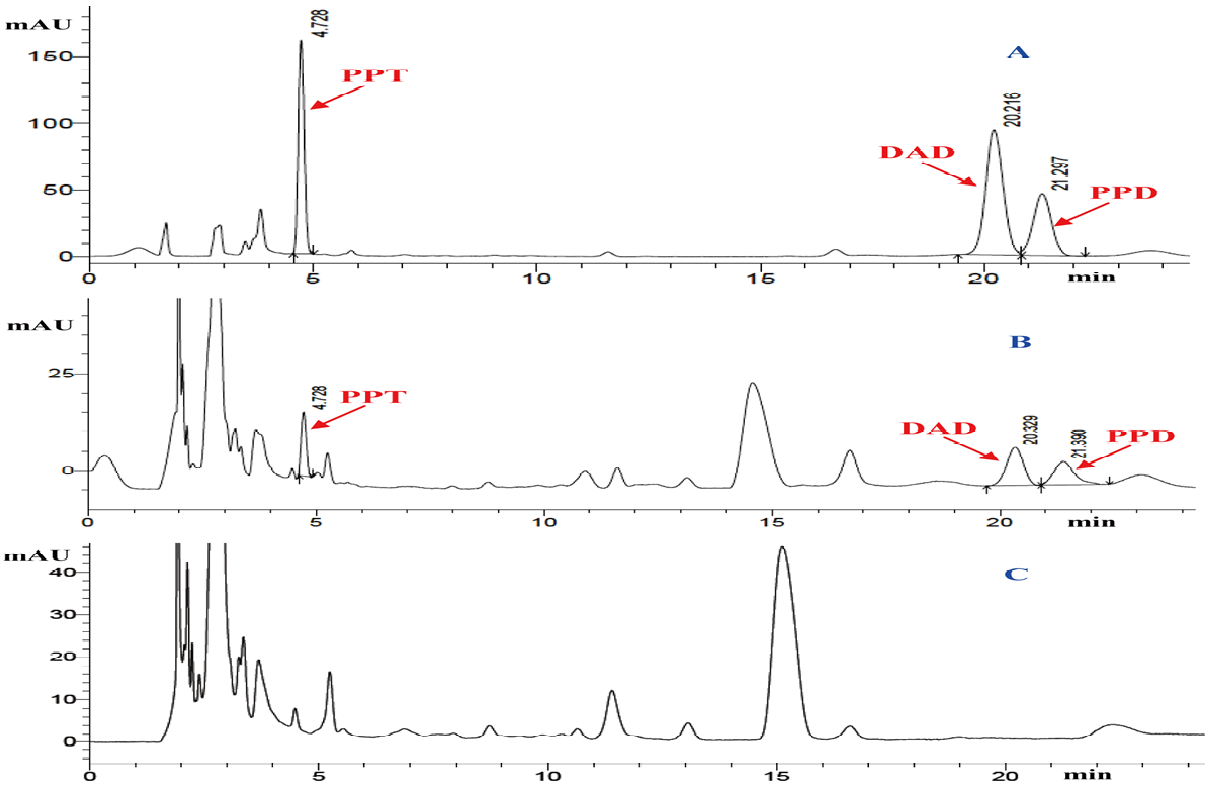


Fig8 .

ACCEPTED MANUSCRIPT

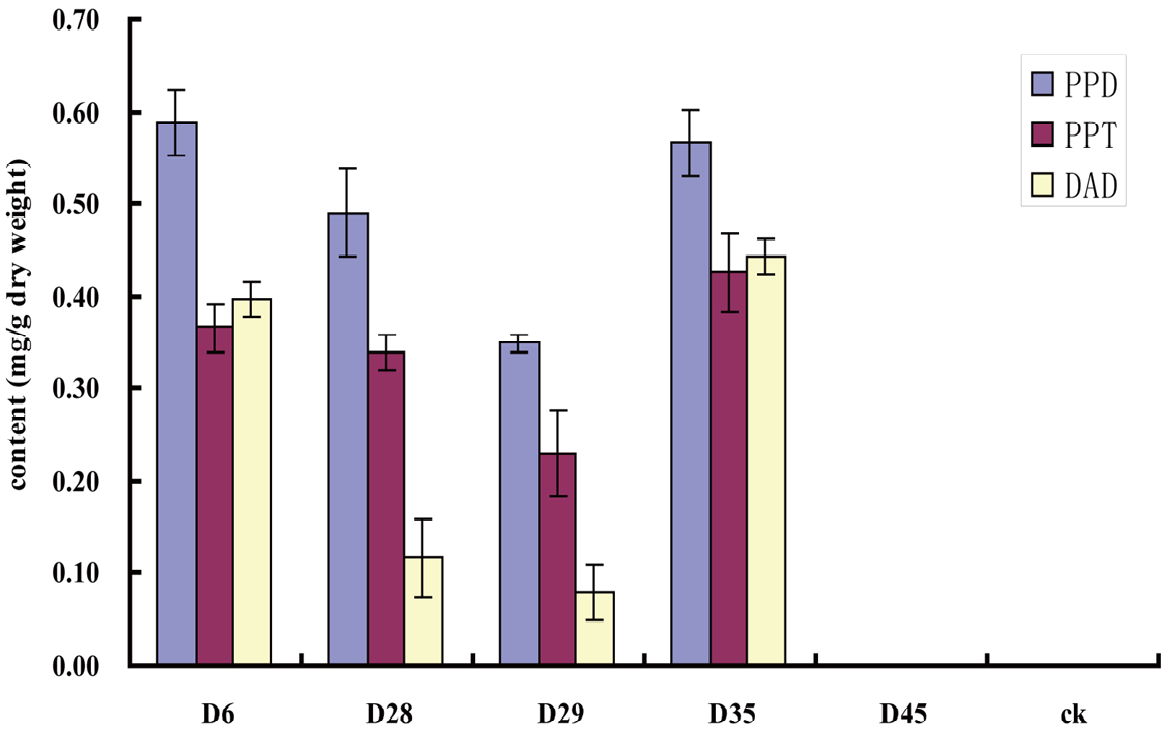
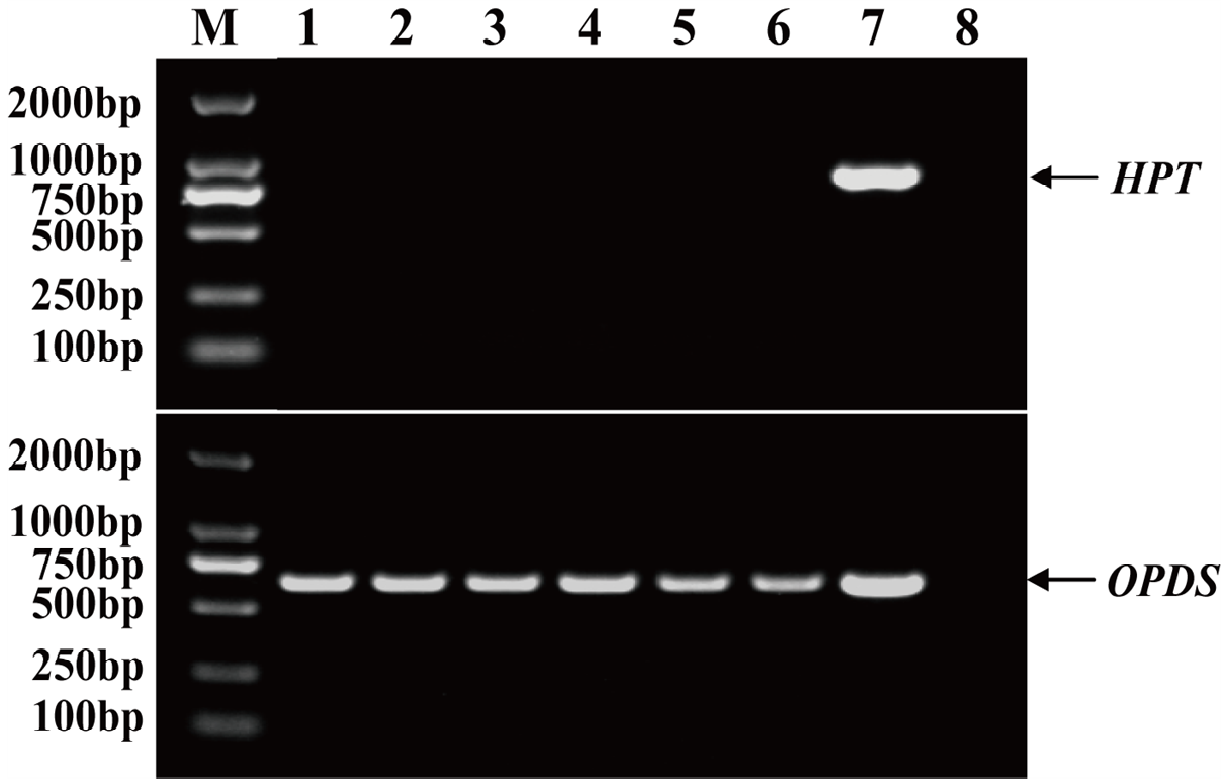


Fig9 .

ACCEPTED MANUSCRIPT



FigS1 .

ACCEPTED MANUSCRIPT

```
Query=  D6
Length=398
```

KP687750

```
                                                                      Score     E
Sequences producing significant alignments:                          (Bits)  Value
  Chr03  2003-08-01 BGI                                                736     0.0

> Chr03  2003-08-01 BGI
Length=41884883

 Score =  736 bits (398),  Expect = 0.0
 Identities = 398/398 (100%), Gaps = 0/398 (0%)
 Strand=Plus/Plus

Query  1         GGGGCCCACAGGAATGAAGATGCTGAATTAAATTATAATCCCCATTCTCTGGATCGAAGT  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  32503243  GGGGCCCACAGGAATGAAGATGCTGAATTAAATTATAATCCCCATTCTCTGGATCGAAGT  32503302

Query  61        GTATCTGTACATGATCTTCATTATTCACTGCTTGCTACAGCCTTAAATCCTCTTTCTGGA  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  32503303  GTATCTGTACATGATCTTCATTATTCACTGCTTGCTACAGCCTTAAATCCTCTTTCTGGA  32503362

Query  121       AACAAAAATTTCCCTGTCCCCTTCTACATCCAACAGATATATAAAAGGGGGAAAACTGAA  180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  32503363  AACAAAAATTTCCCTGTCCCCTTCTACATCCAACAGATATATAAAAGGGGGAAAACTGAA  32503422

Query  181       GCAGTAAGGAATCCCTAACaaaaggaaaaaaaaagaaaagaaagaaagaaCTACTAGGGAA  240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  32503423  GCAGTAAGGAATCCCTAACAAAAGGAAAAAAAAGAAAAGAAAGAAAGAACTACTAGGGAA  32503482

Query  241       CTGGAAGTACCAATCTGTCACCCACCAAATCACAAAATTTACAGATGCCTCTGTCGTGGC  300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  32503483  CTGGAAGTACCAATCTGTCACCCACCAAATCACAAAATTTACAGATGCCTCTGTCGTGGC  32503542

Query  301       TGCCTCTGAAAATCCAGCCATTCCAGTTCTTGCAACGAATTTGCATCAAACCGCTTTGCA  360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  32503543  TGCCTCTGAAAATCCAGCCATTCCAGTTCTTGCAACGAATTTGCATCAAACCGCTTTGCA  32503602

Query  361       AACCCGATTACCTGTGCTCCTGGGTTGAGGCTTTACCG  398
                 |||||||||||||||||||||||||||||||||||||
Sbjct  32503603  AACCCGATTACCTGTGCTCCTGGGTTGAGGCTTTACCG  32503640
```

FigS2 .

ACCEPTED MANUSCRIPT

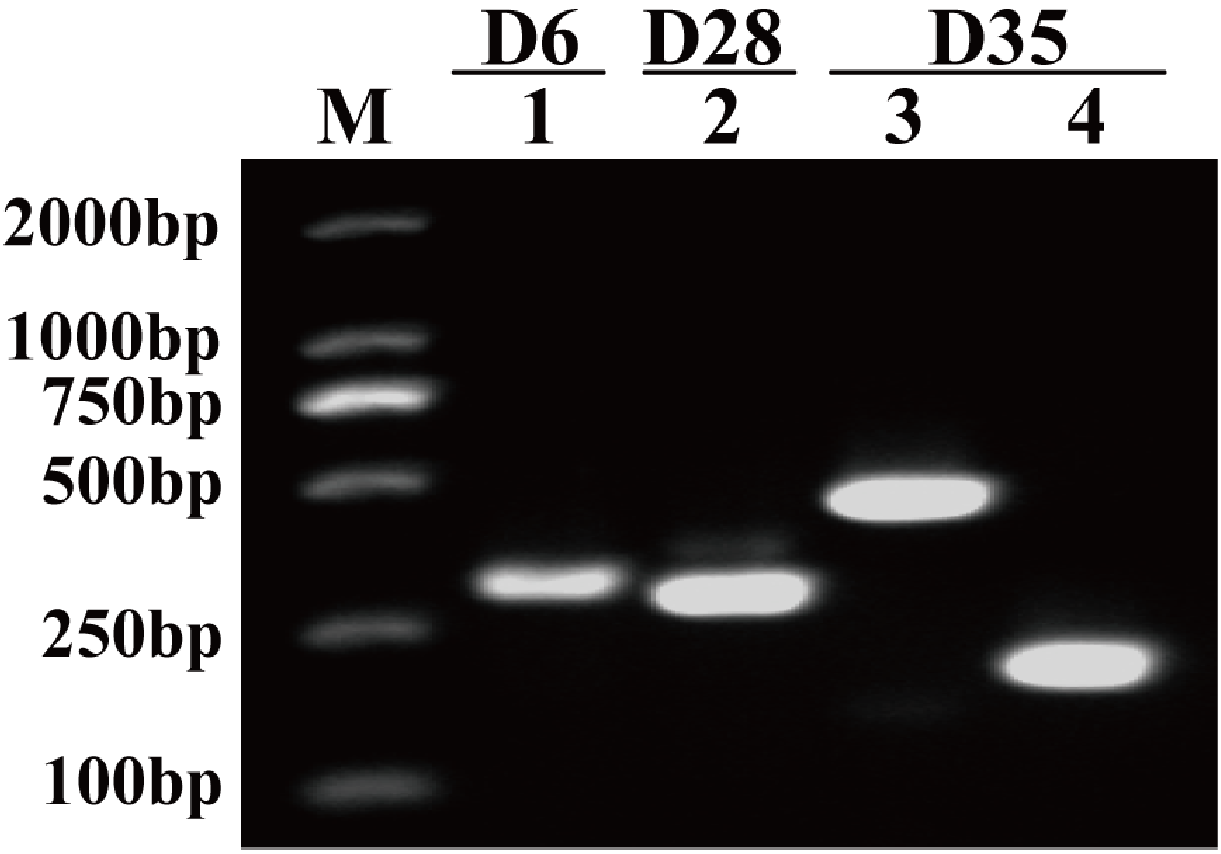
```
Query= D28
Length=253                                          KR055668
                                                                Score     E
Sequences producing significant alignments:                    (Bits)  Value
  Chr01   2003-08-01 BGI                                          468    9e-131

> Chr01   2003-08-01 BGI
Length=45064769

 Score =  468 bits (253),  Expect = 9e-131
 Identities = 253/253 (100%), Gaps = 0/253 (0%)
 Strand=Plus/Plus

Query  1         CCTCGGCTATTCCTCCCCTCGCTTTGTGCCATCGTCGCCTCACCCTCCTCCTATTGTCGT  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21854153  CCTCGGCTATTCCTCCCCTCGCTTTGTGCCATCGTCGCCTCACCCTCCTCCTATTGTCGT  21854212

Query  61        TGCGCCGGCGCCGCCCTCCTTGCCAGTGGATCCAGCCATCCGGAGGAAGGAAGGGTCACG  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21854213  TGCGCCGGCGCCGCCCTCCTTGCCAGTGGATCCAGCCATCCGGAGGAAGGAAGGGTCACG  21854272

Query  121       AGGCCATCAAGCCATCGCTGCCCTCTTCTCCCCTCTCATCTCCAGACCACTACCGCCTTC  180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21854273  AGGCCATCAAGCCATCGCTGCCCTCTTCTCCCCTCTCATCTCCAGACCACTACCGCCTTC  21854332

Query  181       CTCTTCCTCTCCCTATCCTCGGTGAGCTACAGATCCGGTTTGGGGTCATTGGAATCGATG  240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21854333  CTCTTCCTCTCCCTATCCTCGGTGAGCTACAGATCCGGTTTGGGGTCATTGGAATCGATG  21854392

Query  241       GTCTCGGCCTCCC  253
                 |||||||||||||
Sbjct  21854393  GTCTCGGCCTCCC  21854405
```

FigS3 .

ACCEPTED MANUSCRIPT

FigS4 .

ACCEPTED MANUSCRIPT



FigS5 .