

Article

# De Novo Biosynthesis of *p*-Coumaric Acid in *E. coli* with a *trans*-Cinnamic Acid 4-Hydroxylase from the Amaryllidaceae Plant *Lycoris aurea*

# Yikui Li<sup>1,2,†</sup>, Jie Li<sup>1,†</sup>, Binbin Qian<sup>1</sup>, Li Cheng<sup>1</sup>, Sheng Xu<sup>1,2</sup> and Ren Wang<sup>1,2,\*</sup>

- <sup>1</sup> Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China; liyikui@cnbg.net (Y.L.); lj734820531@163.com (J.L.); qian\_binbin01@163.com (B.Q.); chengli9075@163.com (L.C.); xusheng@cnbg.net (S.X.)
- <sup>2</sup> The Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Nanjing 210014, China
- \* Correspondence: rwang@cnbg.net or jswangren@aliyun.com; Tel.: +86-25-84347111
- + These authors contributed equally to this work.

Received: 30 September 2018; Accepted: 29 November 2018; Published: 3 December 2018



Abstract: *p*-Coumaric acid is a commercially available phenolcarboxylic acid with a great number of important applications in the nutraceutical, pharmaceutical, material and chemical industries. *p*-Coumaric acid has been biosynthesized in some engineered microbes, but the potential of the plant CYP450-involved biosynthetic route has not investigated in Escherichia coli. In the present study, a novel trans-cinnamic acid 4-hydroxylase (C4H) encoding the LauC4H gene was isolated from Lycoris aurea (L' Hér.) Herb via rapid amplification of cDNA ends. Then, N-terminal 28 amino acids of LauC4H were characterized, for the subcellular localization, at the endoplasmic reticulum membrane in protoplasts of Arabidopsis thaliana. In E. coli, LauC4H without the N-terminal membrane anchor region was functionally expressed when fused with the redox partner of A. thaliana cytochrome P450 enzyme (CYP450), and was verified to catalyze the *trans*-cinnamic acid to *p*-coumaric acid transformation by whole-cell bioconversion, HPLC detection and LC-MS analysis as well. Further, with phenylalanine ammonia-lyase 1 of A. thaliana, p-coumaric acid was de novo biosynthesized from glucose as the sole carbon source via the phenylalanine route in the recombinant E. coli cells. By regulating the level of intracellular NADPH, the production of *p*-coumaric acid was dramatically improved by 9.18-fold, and achieved with a titer of 156.09 µM in shake flasks. The recombinant cells harboring functional LauC4H afforded a promising chassis for biological production of p-coumaric acid, even other derivatives, via a plant CYP450-involved pathway.

**Keywords:** *p*-coumaric acid; *trans*-cinnamic acid 4-hydroxylase; *Lycoris aurea*; *Escherichia coli*; synthetic biology

# 1. Introduction

*p*-Coumaric acid is a commercially available phenolcarboxylic acid with a great number of important applications in the nutraceutical, pharmaceutical, material and chemical industries. *p*-Coumaric acid possesses potent anti-oxidant, antibacterial and anti-inflammatory properties, and serves as a conventional precursor for the production of flavors and fragrances used in edible and daily chemical products. *p*-Coumaric acid is also a starting material for the preparation of environmentally degradable thermoplastics with liquid crystalline behavior [1]. Recently, *p*-coumaric acid has been found to have many novel bioactivities, such as antiproliferative effect [2], anxiolytic effect [3], a nephroprotective role [4], melanogenesis inhibition [5], and neuroprotective effects [6].



In addition, *p*-coumaric acid, as an upstream metabolite in the plant phenylpropanoid pathway, is a common precursor for biosynthesizing numerous derivatives, such as other phenylpropanoids [7,8], flavonoids [9,10], stilbenes [11] and anthocyanins [12].

*p*-Coumaric acid has been biosynthesized mainly from tyrosine in engineered microbes. *p*-Coumaric acid, as a component of lignin, is ubiquitously present in plants at a low concentration [13]. In plants, the biosynthesis of *p*-coumaric acid involves two biochemical processes: phenylalanine ammonia-lyase (PAL) firstly catalyzes the conversion of phenylalanine to *trans*-cinnamic acid, which is then hydroxylated at the *para* position under the action of *trans*-cinnamic acid 4-hydroxylase (C4H) (Figure 1) [14,15]. However, some PAL enzymes can accept tyrosine as an alternative substrate (PAL/TAL) and directly form *p*-coumaric acid from tyrosine without the intermediacy of *trans*-cinnamic acid (Figure 1) [16]. Also, there are some aromatic amino acid ammonia-lyase homolog enzymes specific for tyrosine deamination (TAL) (Figure 1) [17], so introducing a heterologous PAL/TAL or TAL, *p*-coumaric acid could be produced via the tyrosine route in recombinant cells such as *Escherichia coli* [18–20], *Saccharomyces cerevisiae* [21], *Streptomyces lividans* [22] and *Pseudomonas putida* [23].



**Figure 1.** The biosynthetic routes of *p*-coumaric acid. The phenylalanine route (PAL-C4H) presents ubiquitously in plants and the tyrosine route (PAL/TAL and TAL) exists in some microbes.

In virtue of the plant biosynthetic route PAL-C4H, *p*-coumaric acid has been biosynthesized from phenylalanine only in engineered *S. cerevisiae* [18]. The enzyme C4H is a member of the CYP73A subfamily in cytochrome P450 enzymes (CYP450s) catalyzing a series of oxidation reactions with a CYP450 reductase as redox partner for supplying electrons from NADPH, and is supposed to commonly localize at the cytoplasmic side of endoplasmic reticulum (ER) membrane [14]. Given that prokaryotic microbes such as *E. coli* do not possess compartmentalized organelle(s), the functional expression of CYP450s is difficult [24,25]. Vannelli and his co-workers have functionally co-expressed the C4H and the CYP450 reductase from *Helianthus tuberosus* with a fungal PAL enzyme in *S. cerevisiae* cells, and produced *p*-coumaric acid from the central metabolite L-phenylalanine via the PAL-C4H route [18].

Though *p*-coumaric acid can be produced directly from tyrosine, it was necessary to investigate the potential of the plant CYP450-involved biosynthetic route from phenylalanine in *E. coli* which could to some extent expand the biosynthetic pathway of *p*-coumaric acid and offer an alternative approach to produce *p*-coumaric acid from phenylalanine as well as tyrosine. In the present study, based on our previous transcriptome data of *Lycoris aurea* (L' Hér.) Herb [26], an ornamentally and

medicinally important plant of the *Lycoris* genus of the Amaryllidaceae family, a novel C4H encoding gene was isolated from *L. aurea*, and designated as *Lau*C4H. Then *Lau*C4H was expressed truncatedly at the *N*-terminus in protoplasts of *A. thaliana* to identify the amino acids responsible for the subcellular localization. Moreover, in *E. coli, Lau*C4H without the N-terminal membrane anchor region was heterogeneously expressed for functional identification. PAL1 of *A. thaliana* was further introduced in the recombinant *E. coli* for *p*-coumaric acid de novo biosynthesis from glucose via the phenylalanine route. By regulating the level of intracellular NADPH, the production of *p*-coumaric acid was further increased by 9.18-fold, and a titer of 156.09 µM was achieved in shake flasks.

# 2. Results and Discussion

# 2.1. C4H Homology of L. aurea Transcriptome

Previously, de novo transcriptome sequencing has been performed to produce a comprehensive expressed sequence tag (EST) dataset for *L. aurea* using high-throughput sequencing technology [26]. The EST dataset of *L. aurea* provides a platform be critical in the speeding-up identification of a large number of related genes of secondary metabolite products. Further batch alignment results revealed that 226 contigs and unigenes were annotated to be responsible for the phenylpropanoid biosynthetic pathway. Of them, one unigene, namely unigene CL5217, showing high similarity with plant C4Hs was retrieved. Unigene CL5217 was 1794 bp with a predicted 1518 bp open reading frame (ORF), and was selected for further molecular cloning and functional characterization for the biosynthesis of *p*-coumaric acid.

# 2.2. Cloning of Full-length C4H Genes in L. aurea

By quantitative real-time polymerase chain reaction (qRT-PCR) and rapid amplification of cDNA ends (RACE) approaches, a cDNA encoding C4H homology was isolated from *L. aurea*. Firstly the quantitative PCR primer pairs annealing at the 3'-terminus of the predicted ORF and 3'-untranslational region were designed and used to determine the abundance of the unigene CL5217 in various tissues of *L. aurea*. qRT-PCR results revealed that the level of unigene CL5217 was highest in scape (Figure 2).



**Figure 2.** The abundance of *C*4*H* candidate gene (unigene CL5217) in various tissues of *L. aurea*. The values and error bars represent the mean  $\pm$  standard error from three independent samples in three replicates per sample.

Then, based on the sequence of unigene CL5217, the predicted ORF was cloned with the cDNA pool of scape as the template. Four gene variants were obtained with above 99% identity to each other at the nucleic acid level (Appendix A Figure A1) referring to two protein variants at the amino acid level (Appendix A Figure A2). Based on the genome size observed by flow cytometry, the plant material *L. aurea* used in this study is diploid (2n = 16) [27]. Therefore, a number of C4H paralogs were expected in the genome of *L. aurea*. Moreover, the plant used in this study was flowering and the bulb has been vegetatively propagated. For these reasons, the existence of multiple similar *C4H* transcripts was not surprising. This phenomenon was also reported in cloning the coding genes of norbelladine 4'-O-methyltransferase and *para-para'* C–C phenol coupling CYP450 in the Amaryllidaceae plant *Narcissus sp. aff. pseudonarcissus* [28,29]. The deduced *Lau*C4H protein (variant 1) had a predicted molecular mass of 58.19 kDa and pI of 9.04.

*Lau*C4H possessed all the diagnostic features of the primary structure for CYP450s as well as the CYP73A subfamily. A hydrophobic membrane-spanning region was predicted in the *N*-terminus of *Lau*C4H between amino acid residues 9 and 26 by TMpred online program (Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland) and SignalP 3.0 server [30]. Following the *N*-terminal transmembrane sequence, a proline-rich region (PPGPLPVP) was present (Figure 3A). A conserved heme-binding motif (PFGVGRRSCPG) was also found near the *C*-terminus of *Lau*C4H (Figure 3A). In addition, the "PERF" consensus sequence and some conserved helices such as I-helix (AAIET), J-helix (PDIQQKLRNE), K-helix (KETLR) and K'-helix (AWWLANN) were also identified in *Lau*C4H sequence (Figure 3A). In the homologous tree, plant C4Hs were grouped into two classes as that implied by the differential N-terminus and *C*-terminus in the amino acid sequence alignment (Figure 3A,B), and *Lau*C4H showed higher identity with the C4Hs of Class I than those of Class II (Figure 3B).



**Figure 3.** Sequence alignment (**A**) and homology analysis (**B**) of the full-length amino acid sequence of *Lau*C4H with other C4Hs reported in *A. thaliana* (NP\_180607), *Glycine max* (NP\_001237317, XP\_003544934, XP\_003555891), *Nicotiana tabacum* (NP\_001312254, NP\_001312445, NP\_001313000, NP\_001313097, XP\_016493546, XP\_016505655), *Oryza sativa* Japonica (XP\_015623447, XP\_015626579, XP\_015635394, XP\_015639656), *Sorghum bicolor* (XP\_002452044, XP\_002458683, XP\_002461939) and *Zea mays* (NP\_001140681, NP\_001140726, NP\_001151365, XP\_008657127). GenBank accession numbers are indicated in parentheses. The threshold for shading in panel A was 100%. CYP98A3 (NP\_850337) from *A. thaliana* and CYP96T1 (AMO65741) from *Narcissus* sp. *aff. pseudonarcissus* were additionally analyzed in panel B.

#### 2.3. Subcellular Localization of LauC4H

Further, an investigation of the subcellular localization of LauC4H in plant cells was carried out. LauC4H was firstly predicted to localize in the ER using the WoLF PSORT program [31], and its N-terminal sequence was considered responsible for anchoring to the membrane using the online software TMpred (SIB) and SignalP 3.0 [30]. Then, the whole ORF and the truncated sequences encoding the N-terminal 28 amino acids (LauC4H<sup>1-28</sup>) and the LauC4H without the N-terminal 28 amino acids (LauC4H $^{\Delta 2-28}$ ), were fused with the enhanced green fluorescence protein (EGFP) gene, respectively. These fusion transgenes were transiently expressed in protoplasts prepared from the tender leaf of A. thaliana. Meanwhile, the fusion protein HDEL-mCherry with red fluorescence was used as the ER marker [32]. In Arabidopsis protoplasts containing LauC4H-EGFP and HDEL-mCherry, the green fluorescence was present overlapping with the red fluorescence of the ER marker (Figure 4), showing a typical fluorescence pattern of ER localization. A similar pattern was also observed for the LauC4H<sup>1-28</sup>-EGFP fusion protein overlapping with the ER marker, though there was a small portion of green fluorescence not overlapped with the red fluorescence (Figure 4). In Arabidopsis protoplasts containing  $LauC4H^{\Delta 2-28}$ -EGFP, the fluorescence pattern was distinctly different from those of LauC4H-EGFP and LauC4H<sup>1-28</sup>-EGFP, and there was scarcely any overlap between the green fluorescence and the red fluorescence of the ER marker (Figure 4). Thereby, our observation suggested that the N-terminal 28 amino acids spanning the hydrophobic transmembrane region was responsible for leading LauC4H to localize in ER of plant cells. Previous studies showed that the CYP450 redox partners from A. thaliana and hybrid poplar were localized at the ER membrane of Arabidopsis protoplasts [33,34], those supported the notion that the *N*-terminus maybe confer *Lau*C4H to co-localize with the redox partner at the ER membrane and guide the electron transfer in plant cells (Figure 5A).



**Figure 4.** Subcellular localization of *LauC*4H. "Green" panels show the fluorescence of *LauC*4H variants fused with the enhanced green fluorescent protein (EGFP), "mCherry" panels indicate the fluorescence of endoplasmic reticulum marker (HDEL), and "Chlorophyll" panels indicate the auto-fluorescence of chloroplast. "Merged" panels represent the combined fluorescence from EGFP and mCherry. Bars = 10 µm.

# 2.4. Functional Identification of LauC4H

Given that the oxygenation process of CYP450s requires a redox partner as electron supplier, the *A. thaliana* CYP450 redox partner ATR2 was employed for the functional exploration of *Lau*C4H in *E. coli*. Since there was no internal ER-like membrane in *E. coli*, both *Lau*C4H and ATR2 were expressed heterogenously without the *N*-terminal membrane anchor region, rather than in the full-length sequence, thus to avoid the so-called incompatibility between the membrane recognition signal

of heterogenous proteins and the prokaryotic host [25]. In detail, the truncated ATR2 (ATR2<sup> $\Delta$ 2-74</sup>) was fused in the N-terminus of truncated *Lau*C4H (*Lau*C4H<sup> $\Delta$ 2-28</sup>) via a flexible octapeptide linker for electron supply (Figure 5B). The linker and the following proline-rich region in *Lau*C4H could orient optimally between ATR2 and *Lau*C4H [25,35]. When the recombinant *E. coli* cells harboring the chimeric ATR2<sup> $\Delta$ 2-74</sup>-*Lau*C4H<sup> $\Delta$ 2-28</sup> protein were incubated with the substrate *trans*-cinnamic acid, a new product peak with the same retention time of *p*-coumaric acid was detected by HPLC analysis (Figure 5C). UV-Vis absorption spectra confirmed that the new product was identical to *p*-coumaric acid (Figure 5D).



**Figure 5.** Functional identification of *Lau*C4H. (**A**) Diagram of endoplasmic reticulum co-localization of *Lau*C4H and *A. thaliana* cytochrome P450 reductase 2 (ATR2) in plant cells. (**B**) Heterologous fusion expression of *Lau*C4H and ATR2 via a flexible linker in *E. coli* without both N-terminal membrane anchor region. (**C**) HPLC analysis of reaction products from pET29a-ATR2<sup> $\Delta$ 2-74</sup> (1), pET29a (2), pET29a-ATR2<sup> $\Delta$ 2-74</sup>*Lau*C4H.1<sup> $\Delta$ 2-28</sup> (3) and pET29a-ATR2<sup> $\Delta$ 2-74</sup>*Lau*C4H.2<sup> $\Delta$ 2-28</sup> (4) using *trans*-cinnamic acid as the substrate. Standards (5) refer to HPLC analysis result of standard substance mixture of *p*-coumaric acid (a) and *trans*-cinnamic acid (b). (**D**) UV absorption spectra of the new product and the standard *p*-coumaric acid. (**E**,**F**) LC-MS identification of reaction products generated by recombinant *Lau*C4H.1 (**E**) and *Lau*C4H.2 (**F**).

The corresponding product was further examined using a liquid chromatography-mass spectrometry (LC-MS) in positive ion mode. LC-MS analysis of the targeted peak of the two LauC4H paralogs displayed the  $[M + H]^+$  ions at m/z 165 (Figure 5E,F), corresponding to the calculated molecular weight of *p*-coumaric acid (MW = 164.16). In contrast, no any visible peak with the same retention time of p-coumaric acid was observed in the conversion systems with ATR2<sup> $\Delta$ 2-74</sup> only as well as the empty vector pET29a, even though, with trans-cinnamic acid as the substrate. These results indicated that the new product synthesized with the chimeric ATR2<sup> $\Delta$ 2-74</sup>-LauC4H<sup> $\Delta$ 2-28</sup> protein was *p*-coumaric acid. Thus, the LauC4H is an authentic trans-cinnamic acid 4-hydroxylase able to participate the conversion of trans-cinnamic acid into p-coumaric acid with the plant CYP450 redox partner (Figure 1).

## 2.5. p-Coumaric Acid De Novo Biosynthesis Using LauC4H in E. coli

For the application of LauC4H to de novo biosynthesize p-coumaric acid, the E. coli cells harboring the functional chimera  $ATR2^{\Delta 2-74}$ -LauC4H $^{\Delta 2-28}$  were cultured in mineral medium added with glucose as the sole carbon source. However, no p-coumaric acid was detected (Table 1), indicating that E. coli could not produce trans-cinnamic acid. When A. thaliana phenylalanine ammonia-lyase 1 (AthPAL1) was further introduced to the E. coli cells mentioned above, p-coumaric acid could be produced and accumulated in the medium of the recombinant cells with  $ATR2^{\Delta 2-74}$ -LauC4H $^{\Delta 2-28}$  and AthPAL1 (Table 1). Therefore, the plant CYP450-involved p-coumaric acid biosynthesis pathway has been successfully established in procaryotic E. coli. Moreover, within a duration of 42 h induction with IPTG, the recombinant E. coli produced 17.22 µM of p-coumaric acid. The expression level and solubility of the chimeric ATR2<sup> $\Delta 2-74$ </sup>-LauC4H<sup> $\Delta 2-28$ </sup> fusion protein were detected in Ec/LauC4H and Ec/LauC4H-AthPAL (Figure A3). As shown in Figure A3, the expression level of the chimeric  $ATR2^{\Delta 2-74}$ -LauC4H $^{\Delta 2-28}$  fusion protein in Ec/LauC4H was more than that in Ec/LauC4H-AthPAL. However, the level of the soluble  $ATR2^{\Delta 2-74}$ -LauC4H $^{\Delta 2-28}$  fusion protein was comparable between Ec/LauC4H and Ec/LauC4H-AthPAL. Comparably, E. coli cells only with AthPAL1 or PAL/TAL from the yeast *Rhodotorula glutinis* (*RgIPAL*/TAL) [18] were also cultured as controls. In cells only with AthPAL1, trans-cinnamic acid rather than p-coumaric acid was detected in the medium (Table 1). In cells only with RglPAL/TAL, both trans-cinnamic acid and p-coumaric acid were accumulated, of which the concentration was 91.94  $\mu$ M and 46.09  $\mu$ M, respectively (Table 1). Notably, there was 342.82  $\mu$ M of *trans*-cinnamic acid detected in the medium of the recombinant cells with ATR2<sup> $\Delta 2-74$ </sup>-LauC4H<sup> $\Delta 2-28$ </sup> and AthPAL1 (Table 1).

Strain	Biomass (OD <sub>600</sub> )	Glucose Utilization (g/L)	<i>trans</i> -Cinnamic Acid (μM)	<i>p-</i> Coumaric Acid (μM)
Ec/LauC4H	$5.56\pm0.72$	$13.50\pm0.72$	ND <sup>2</sup>	ND <sup>2</sup>
Ec/AthPAL	$8.92\pm0.85$	$24.50\pm0.65$	$241.32\pm13.24$	ND <sup>2</sup>
Ec/ <i>Rgl</i> PAL(TAl)	$9.73 \pm 1.24$	$25.08\pm0.75$	$91.94 \pm 6.32$	$46.09\pm2.96$
Ec/LauC4H-AthPAL	$7.66\pm0.83$	$24.78 \pm 1.26$	$342.82\pm15.80$	$17.22\pm1.23$

Table 1. p-Coumaric acid de novo biosynthesis using LauC4H in E. coli.<sup>1</sup>

 $^1$  The data represent the mean  $\pm$  the standard error from three independent experiments of each strain.  $^2$  ND, not detected.

# 2.6. p-Coumaric Acid Production Improved by Intracellular NADPH Regulation

The biochemical process of *Lau*C4H catalysis of the conversion of *trans*-cinnamic acid into *p*-coumaric acid needs two electrons per mole of substrate. ATR2 as a CYP450 redox partner is NADPH-dependent [33,34]. Since we observed that about 20-fold the substrate of *Lau*C4H still existed in the medium of strain Ec/*Lau*C4H-*Ath*PAL (Table 1), we were assuming that the output of *p*-coumaric acid was subjected to the level of intracellular NADPH (Figure 1). For testing this assumption, we attempted to elevate the level of intracellular NADPH. Using synthetic small regulatory RNA

(srRNA) anti(sthA) [36] to specific repression of the translation of the soluble transhydrogenase SthA (also referred to as UdhA) [37], the conversion of NADPH to NADH may be down-regulated when treated with anti(sthA) so that the content of NADPH would be relatively enhanced. As shown in Figure 6, the introduction of the srRNA anti(sthA) accelerated cell growth and glucose utilization, and resulted in about 2-fold increase of the *p*-coumaric acid production. In addition, we also pursued overexpression of the membrane-bound transhydrogenase PntAB catalyzing the NADH to NADPH conversion [37]. When the overexpression of five-copy *pntAB* driven by a T7 promoter took place, the production of *p*-coumaric acid was dramatically improved by 7.93-fold up to 136.53  $\mu$ M (Figure 6C). Both over-expressed PntAB and srRNA anti(sthA) resulted in a 9.18-fold increase, along with relative lower biomass and slower glucose consumption. Thus over-expressed PntAB and srRNA anti(sthA) played a synergetically positive effect on the de novo biosynthetic production of LauC4H-mediated *p*-coumaric acid. The expression level and solubility of the chimeric ATR2<sup> $\Delta 2-74$ </sup>-LauC4H<sup> $\Delta 2-28$ </sup> fusion protein in those *p*-coumaric acid producers were comparable (Figure A3). Under such circumstance, there was a considerable amount of *trans*-cinnamic acid yet in the medium, indicating that the chimera ATR2<sup> $\Delta 2-74$ </sup>-LauC4H<sup> $\Delta 2-28$ </sup> was involved in a rate-limiting step for the formation of *p*-coumaric acid besides NADPH. Subsequently, there should be more approaches to be tested for improving the output of *p*-coumaric acid in *E. coli*. For example, the turnover of *Lau*C4H could be enhanced either by increasing the expression level of the chimera or by modulating the spatial structure of the chimera using modularized bioengineering tools in synthetic biology.



**Figure 6.** Regulation of intracellular NADPH to improve the *p*-coumaric acid production. (**A**) cell growth; (**B**) glucose utilization; (**C**) *p*-coumaric acid biosynthesis; (**D**) *trans*-cinnamic acid accumulation. The *p*-coumaric acid producers were Ec/*Lau*C4H-*Ath*PAL-sRNA (black square), Ec/*Lau*C4H-*Ath*PAL-anti(sthA) (red circle), Ec/*Lau*C4H-*Ath*PAL-PntAB (magenta triangle), and Ec/*Lau*C4H-*Ath*PAL-PntAB-anti(sthA) (blue diamond). The error bars represent the standard error of the mean from three independent experiments of each strain.

#### 3. Materials and Methods

#### 3.1. Plant Materials and Chemicals

The *Lycoris aurea* plants used in this study were collected from Nanjing Botanical Garden Mem. Sun Yat-Sen (Nanjing, China), and were about three-year old when the flowers bloom unless otherwise stated. *A. thaliana* wild-type (Columbia ecotype) plants used in this study were grown at 22 °C for four weeks after germination. Chemicals and reagents used in this study were purchased from either Sigma-Aldrich (St. Louis, MO, USA) or Sangon Biotech (Shanghai, China).

#### 3.2. RNA Extraction and Isolation of LauC4H Genes

Total RNA from different tissues of L. aurea were extracted using the RNAprep pure Plant Kit (TIANGEN, Beijing, China). The cDNA pool was then synthesized by PrimerScript<sup>TM</sup> RT reagent Kit (TaKaRa, Dalian, China). To quantify the unigene CL5217 expression in different tissues, quantitative real-time polymerase chain reaction (qRT-PCR) was performed using AceQ® qPCR SYBR® Green Master Mix (High ROX Premixed) (Vazyme Biotech, Nanjing, China) with the gene LauTIP41 as internal reference [38]. To obtain desired sequences, 5'- and 3'- rapid-amplification of cDNA ends (RACE) were carried out using the SMARTer<sup>TM</sup> RACE cDNA Amplification Kit (Clontech Laboratories, Inc., Mountain View, CA, USA) according to the manufacturer's manual. The full-length cDNA were verified by re-amplification of the open reading frame (ORF) using forward primer LauC4H-ORF-PF and reverse primer LauC4H-ORF-PR. PCR protocols were as follows: one cycle of 5 min at 98 °C; 30 cycles with a denaturing time of 45 s at 95  $^{\circ}$ C, an annealing time of 45 s at 56  $^{\circ}$ C, and an elongation time of 90 s at 72 °C; and a final elongation step of 10 min at 72 °C. The PCR products of expected size were excised from the EtBr-stained 1% (w/v) agarose gels, purified using the DNA Gel Extraction Kit (BioTeke Corporation, Beijing, China) and ligated into the pMD19-T vector (TaKaRa) for sequencing. Each of the transcripts was obtained from at least three different monoclones. The primers used in this experiment are listed in Appendix A Table A1.

#### 3.3. Sequence Aanalysis of LauC4H

The open reading frames (ORFs) in the obtained cDNA sequences were searched and the correspondingly encoded protein sequences were exported on NCBI ORFfinder webpage (https://www.ncbi.nlm.nih.gov/orffinder/) (Bethesda, MD, USA). Then, homology search of the predicted protein were performed using regular blastp on NCBI BLAST webpage (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Bethesda, MD, USA). Molecular mass and isoelectric point of translated proteins were predicted on the SIB Bioinformatics Resource Portal ExPASy (https://www.expasy.org/) (Lausanne, Switzerland). The subcellular localization in plant cells was predicted using the WoLF PSORT program (https://wolfpsort.hgc.jp/) [31]. The prediction of signal sequence and transmembrane sequence was carried out using the SignalP 3.0 Server (http://www.cbs.dtu.dk/services/SignalP-3.0/) [30] and the TMpred online program (https://embnet.vital-it.ch/software/TMPRED\_form.html) (SIB). Amino acid sequences of full-length C4Hs were aligned and analyzed by the BioEdit Sequence Alignment Editor Version 7.1.3.0 [39] and the ClustalW program (SIB), respectively. The homologous tree was constructed via the software DNAMAN Version 8 (Lynnon Corporation, San Ramon, CA, USA).

#### 3.4. Subcellular Localization Analysis of LauC4H

To determine the subcellular localization of *Lau*C4H, enhanced green fluorescent protein (EGFP) was in frame fused to the C-terminus of the *Lau*C4H protein sequence under the control of the dual cauliflower mosaic virus (CaMV) 35S promoter in the pAN580 vector. The plasmid pAN580 was digested by *NcoI* (TaKaRa, Dalian, China). The DNA fragments without termination codon encoding full-length *Lau*C4H, *N*-terminal 28 amino acids of *Lau*C4H (*Lau*C4H<sup>1–28</sup>) and N-terminal truncated *Lau*C4H (*Lau*C4H<sup>Δ2–28</sup>) were prepared with primer pairs pAN580-*Lau*C4H-PF and EGFP-*Lau*C4H(N28)-PR, and primer

pairs pAN580-*Lau*C4H( $\Delta$ N28)-PF and EGFP-*Lau*C4H-PR, respectively. The DNA fragments were assembled to linear pAN580 by ClonExpress One Step Cloning Kit (Vazyme Biotech) to create pDual35S::*Lau*C4H-EGFP, pDual35S::*Lau*C4H<sup>1–28</sup>-EGFP and pDual35S::*Lau*C4H<sup> $\Delta$ 2–28</sup>-EGFP. Primers used to make these constructs were designed according to the product manual and are listed in Appendix A Table A1. The well-established fluorescent protein marker mCherry-HDEL [32] was used for the indication of the endoplasmic reticulum (ER). All transient expression constructs were transformed separately into *Arabidopsis* protoplasts with the mCherry-HDEL construct according to the method [40]. The transformed samples were incubated for 16–18 h before examination. Fluorescent images were observed by a laser scanning confocal microscope using a Zeiss LSM780 camera (Carl Zeiss Microscopy GmbH, Jena, Germany).

## 3.5. Heterologous Expression of LauC4H in Escherichia coli

Escherichia coli Top10 was used as the host for molecular cloning. For gene expression in E. coli, Novagen®'s pET29a vector (EMD Millipore Corporation, Darmstadt, Germany) was used and digested by *NdeI* and *XhoI* (TaKaRa). The DNA fragment encoding *N*-terminal truncated *Lau*C4H (*Lau*C4H $^{\Delta 2-28}$ ) were amplified with primer pairs  $29aNdeI-LauC4H(\Delta N28)$ -PF and 29aXhoI-LauC4H-PR. The DNA encoding the cytochrome P450 redox partner ATR2 without N-terminal 74 amino acids (ATR2 $^{\Delta 2-74}$ ) was obtained with the primer pairs 29aNdeI-ATR2(ΔN74)-PF and 29aXhoI-ATR2-PR from A. thaliana cDNA pool as the template. The DNA fragments was assembled to linear pET29a by ClonExpress One Step Cloning Kit (Vazyme Biotech) to create pET29a-*Lau*C4H<sup> $\Delta$ 2-28</sup> and pET29a-ATR2<sup> $\Delta$ 2-74</sup>. To express LauC4H functionally in E. coli, the truncated ATR2 with a short flexible octapeptide (GSTSSGSG) was prepared with the primer pairs  $29aNdeI-ATR2(\Delta N74)-PF$  and  $LauC4H(\Delta N28)ATR2-PR$ , and was in frame fused to the N-terminus of  $LauC4H^{\Delta 2-28}$  into the NdeI site of pET29a-LauC4H<sup> $\Delta 2-28$ </sup> by ClonExpress One Step Cloning Kit (Vazyme Biotech) to create pET29a-ATR2 $^{\Delta 2-74}LauC4H^{\Delta 2-28}$ . Primers used to make these constructs are listed in Appendix A Table A1. E. coli Rosetta (DE3) (Vazyme Biotech) was used as the host for protein expression. Freshly transformed E. coli Rosetta (DE3) cells harboring pET29a-derived plasmids were grown at 37 °C and 220 rpm in M9Y minimal medium [41] with 20 g L<sup>-1</sup> glucose as carbon source to A<sub>600</sub> of 0.6–0.8, and then were induced with 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) at 28 °C overnight.

#### 3.6. Bioconversion of LauC4H with trans-Cinnamic Acid

Whole-cell bioconversion strategy was adopted to identify the function of *Lau*C4H. 50 mL cell cultures were harvested by centrifugation at 8000 rpm and 4 °C for 5 min, washed twice and re-suspended in 40 mL nitrogen-free M9Y mineral medium. The re-suspended cells were added 20 g L<sup>-1</sup> glucose and 100  $\mu$ M *trans*-cinnamic acid, and incubated at 28 °C with a stirring rate of 250 rpm for the whole-cell bioconversion. Samples were taken with an interval of 12 h, an equal volume of methanol was added to terminate the reaction and the samples were then subjected to HPLC analysis.

#### 3.7. p-Coumaric Acid De Novo Biosynthesis in E. coli

To in vivo biosynthesize the substrate of *Lau*C4H in *E. coli* cells, the *AthPAL1* gene encoding *A. thaliana* phenylalanine ammonia lyase 1 was overexpressed in pACYC184 vector [42,43] under a trc promoter induced by IPTG. For gene overexpression, pACYC184 was digested by *Nco*I and *Eco*RI (TaKaRa, Dalian, China) to get the plasmid skeleton with p15A origin of replication and tetracycline-resistant gene. Primer pairs 184-trc-lacO-PF and BBa\_B0034-lacO-PR were annealed and elongated, and then purified to obtain the DNA fragment trcO-RBS containing the core-trc promoter (–10 box and –35 box), lacI binding site and the BioBrick ribosome binding site (RBS) BBa\_B0034 chosen from the MIT Registry of Standard Biological Parts (http://parts.igem.org/Main\_Page) (The International Genetically Engineered Machine (iGEM), Cambridge, MA, USA). Primer pairs BBa\_B0034-BBa\_B0015-PF and 184-BBa\_B0015-PR were annealed and elongated, and then purified to obtain the DNA fragment BBa\_B0034-BBa\_B0015 containing the BioBrick terminator BBa\_B0015 also chosen

from the MIT Registry of Standard Biological Parts (iGEM, Cambridge, MA, USA). The three fragments, plasmid skeleton, trcO-RBS and BBa\_B0015 terminator, sharing the overlap one by one, were assembled together by ClonExpress Ultra One Step Cloning Kit (Vazyme Biotech). The new plasmid was named as p15A-trcO3415. Then, the *AthPAL1* gene was amplified from the *A. thaliana* cDNA pool with primers BBa\_B0034-*AthPAL1*-PF and BBa\_B0015-*AthPAL1*-PR and assembled into the linear p15A-trcO3415 by *Nde*I restriction endonuclease (TaKaRa) via ClonExpress One Step Cloning Kit (Vazyme Biotech) to create p15A-trcO*Ath*PAL1. Then p15A-trcO*Ath*PAL1 along with the plasmid pET29a-ATR2<sup> $\Delta$ 2-74</sup>LauC4H<sup> $\Delta$ 2-28</sup> were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/LauC4H-AthPAL. The empty vector p15A-trcO3415 along with the plasmid pET29a-ATR2<sup> $\Delta$ 2-74</sup>LauC4H<sup> $\Delta$ 2-28</sup> were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/LauC4H. The plasmid p15A-trcO*Ath*PAL1 along with the empty vector pET29a were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/LauC4H. The plasmid p15A-trcO*Ath*PAL1 along with the empty vector pET29a were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/LauC4H. The plasmid p15A-trcO*Ath*PAL1 along with the empty vector pET29a were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/LauC4H. The plasmid p15A-trcO*Ath*PAL1 along with the empty vector pET29a were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/LauC4H. The plasmid p15A-trcO*Ath*PAL1 along with the empty vector pET29a were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/AthPAL. The primers involved in this experiment are summarized in Appendix A Table A1.

The DNA sequence encoding the PAL/TAL from the yeast *Rhodotorula glutinis* [18] was optimized according to the codon preference of *E. coli*, and synthesized on the clone vector pUC57 to obtain pUC57-Opt*Rgl*PAL/TAL plasmid by Sangon Biotech (Shanghai) Co., Ltd (Shanghai, China). The *RglPAL/TAL* gene was amplified from the plasmid pUC57-Opt*Rgl*PAL/TAL with primers BBa\_B0034-*RglPAL/TAL*-PF and BBa\_B0015-*RglPAL/TAL*-PR and assembled into the linear p15A-trcO3415 by *Ndel* restriction endonuclease (TaKaRa) via ClonExpress One Step Cloning Kit (Vazyme Biotech) to create p15A-trcO*Rgl*PAL/TAL. Then p15A-trcO*RglPAL/TAL* plasmid along with the empty vector pET29a were transformed into *E. coli* Rosetta (DE3) cells to obtain the recombinant strain Ec/*Rgl*PAL(TAL). The primers are summarized in Appendix A Table A1 and the optimized DNA sequence of *Rgl*PAL/TAL was shown in Appendix A Figure A4.

A single clone of the recombinant *E. coli* strains was incubated in 3 mL LB medium for 24 h. Then 1 mL of the cell cultures was collected, the supernatant was discarded, and the cell pellet was transferred into 50 mL fermentation medium and grown at 37 °C and 200 rpm. The fermentation medium contains 100 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS), 28.71 mM K<sub>2</sub>HPO<sub>4</sub>, 25.72 mM KH<sub>2</sub>PO<sub>4</sub>, 26.50 mM (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub>, 10.00 mM citric acid, 5.00 mM MgSO<sub>4</sub>, 1 µg L<sup>-1</sup> vitamin B1, 0.5 µg L<sup>-1</sup> biotin, 166.67 mM glucose, 0.02 mM phenol red, and 10 mL trace element solution. The trace element solution was prepared in 100 mM HCl with 36.00 mM FeSO<sub>4</sub>, 7.82 mM ZnSO<sub>4</sub>, 4.00 mM CuSO<sub>4</sub>, 2.00 mM MnSO<sub>4</sub>, 0.60 mM Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub>, 13.60 mM CaCl<sub>2</sub>, and 0.08 mM (NH<sub>4</sub>)<sub>6</sub>MO<sub>7</sub>O<sub>24</sub>. The cells were induced with 0.1 mM IPTG at 30 °C when grown to A<sub>600</sub> of 0.6–0.8. The pH in the medium was maintained at about 7.0 by adding ammonium hydroxide aperiodically. Samples were taken with an interval of 6 h and analyzed by HPLC.

#### 3.8. Intracellular NADPH Regulation

For down-regulation of *EcoSthA* gene, the synthetic sRNA-based strategy [36] was applied. The plasmid skeleton with pSC101 origin of replication and spectinomycin-resistant gene was obtained from pCL1920 [44] as previously described [45]. The P<sub>R</sub>-MicC fragment including the constitutive P<sub>R</sub> promoter and the MicC scaffold was amplified from the wild-type *E. coli* MG1655 genome with primer pairs BBa\_R0051-MicC-PF and BBa\_B0015-MicC-PR. The BBa\_B0015 terminator was amplified from the plasmid p15A-trcO3415 with primer pairs BBa\_B0015-PF and BBa\_B0015-PR. The P<sub>R</sub>-MicC-BBa\_B0015 fragment was amplified with primer pairs pCL-BBa\_R0051-PF and pCL-BBa\_B0015-PR from both of P<sub>R</sub>-MicC fragment and BBa\_B0015 terminator as the template. Then, the plasmid skeleton and the P<sub>R</sub>-MicC-BBa\_B0015 fragment were assembled by ClonExpress One Step Cloning Kit (Vazyme Biotech) to construct the new plasmid pSC101-sRNA for synthetic sRNA production. Subsequently, the *N*-terminal 24 bp of *EcoSthA* gene was obtained by annealing and elongation with primer pairs anti(*sthA*)-PF and anti(*sthA*)-PR, and assembled into the linear pSC101-sRNA by *Nsi*I restriction endonuclease (TaKaRa) via ClonExpress One Step Cloning Kit (Vazyme Biotech) to create pSC101-anti(sthA).

For appropriate overexpression of NAD(P) transhydrogenase, the two genes *EcopntAB* were overexpressed with their native ribosome binding sites under a constitutive T7 promoter and transcriptionally terminated by the BioBrick terminator BBa\_B1006 also chosen from the MIT Registry of Standard Biological Parts (iGEM, Cambridge, MA, USA). The *EcopntAB* DNA fragment was amplified from the wild-type *E. coli* MG1655 genome with primer pairs pCL-T7-*pntA*-PF and pCL-BBa\_B1006-*pntB*-PR, and assembled into the *KpnI* (TaKaRa) digested pSC101-sRNA and pSC101-anti(sthA) via ClonExpress One Step Cloning Kit (Vazyme Biotech) to create pSC101-OE(pntAB) and pSC101-anti(sthA)-OE(pntAB), respectively. Then pSC101-sRNA, pSC101-anti(sthA), pSC101-OE(pntAB) and pSC101-anti(sthA)-OE(pntAB), respectively, was transformed into strain Ec/*Lau*C4H-*Ath*PAL cells to obtain the recombinant strain Ec/*Lau*C4H-*Ath*PAL-PntAB-anti(sthA).

#### 3.9. Protein Detection

Cells induced for 12 h were harvested by centrifugation at 8000 rpm for 5 min at 4 °C, washed twice with TNG buffer (20 mM Tris-HCl pH 7.9, 0.5 M NaCl, 10% Glycerol), and re-suspended in the same buffer containing 1 mM phenylmethylsulfonyl fluoride (PMSF). The suspended cells were sonicated on ice-bath followed by centrifugation at 12,000 rpm for 30 min at 4 °C. The supernatant was transferred out, and the sediment was re-suspended in the PMSF-contained TNG buffer. Proteins were assessed by SDS-PAGE.

# 3.10. High Performance Liquid Chromatography (HPLC) Analysis

All samples taken from the cultures were centrifuged at 12,000 rpm for 2 min and the supernatants were filtered through a 0.22 µm polytetrafluorethylene (PTFE) filter and analyzed by HPLC (Shimadzu LC-20A, Kyoto, Japan) with a reverse phase Shimadzu InertSustain C18 column (5 µm, 4.6 mm × 250 mm) and a Shimadzu SPD-M20A photodiode array detector. The mobile phase used was a gradient of solvent A (H<sub>2</sub>O containing 1.3% acetic acid) and solvent B (100% acetonitrile) applied as following time procedure: 0–20 min, 10–100% B linear; 20–20.5 min, 100%–10% linear; 20.5–30 min, 10% B isocratic. The flow rate was set at 1.0 mL·min<sup>-1</sup>, and the injection volume was 10 µL. The column was maintained at 35 °C, and the eluted compounds were monitored at 309 nm for *p*-coumaric acid and at 274 nm for *trans*-cinnamic acid respectively. The concentration of *trans*-cinnamic acid and *p*-coumaric acid were quantified by fitting the peak area with a standard curve (R<sup>2</sup> > 0.999) of the corresponding standard.

#### 4. Conclusions

In the present study, *p*-coumaric acid was de novo produced via the plant CYP450-involved biosynthetic route in *Escherichia coli*. Firstly, a novel CYP73A from the Amaryllidaceae plant *Lycoris aurea* was cloned based on the transcriptome data, of which the *N*-terminal 28 amino acids were characterized responsible for localizing at the endoplasmic reticulum membrane in protoplasts of *Arabidopsis thaliana*. Then, *Lau*C4H was expressed functionally in *E. coli* when fused with the CYP450 redox partner from *A. thaliana*, and shown to catalyze the conversion of *trans*-cinnamic acid into *p*-coumaric acid. Further, *p*-coumaric acid was de novo biosynthesized via introducing a phenylalanine ammonia-lyase from *A. thaliana* into the recombinant *E. coli* cells. The production of *p*-coumaric acid was dramatically improved via regulation of the intracellular NADPH level of the constructed cell factory. The producer reported herein afforded a promising chassis for the biosynthesis of *p*-coumaric acid-derived molecules with great application value.

Author Contributions: Conceptualization, Y.L. and R.W.; Data Curation, Y.L. and J.L.; Formal Analysis, Y.L. and J.L.; Funding Acquisition, Y.L. and R.W.; Investigation, Y.L., J.L., B.Q. and L.C.; Methodology, Y.L., J.L., B.Q., L.C. and S.X.; Project Administration, R.W.; Resources, J.L., B.Q. and L.C.; Software, Y.L. and S.X.; Supervision,

Y.L. and R.W.; Validation, J.L. and B.Q.; Visualization, Y.L. and J.L.; Writing-Original Draft Preparation, Y.L.; Writing-Review & Editing, Y.L., J.L. and R.W. All authors revised and approved the manuscript.

**Funding:** This work was financially supported by a research grant from the National Natural Science Foundation of China (NOs. 31600074 and 31770387), a grant from the Natural Science Foundation of Jiangsu Province (NO. BK20160599) and a grant from the Jiangsu Key Laboratory for the Research and Utilization of Plant Resources (NO. JSPKLB201837).

Conflicts of Interest: The authors declare no conflict of interest.

# Appendix A

		10	20	30	40	50	60	70	80 .	90
LauCL5217 ORF	ATGGATCT	CATTCTACTA	GAAAGTCAC	TCCTCGCAGTT	TTTTTCGCT.	AŤCATCŤTTA	GTATCATAATA	TCCAAACTC	CĠCGGGĂAGAA	Ġ
LanC4H.2							T			
LauC4H.3 LauC4H.4							т			
24110-1224										
		100	110	120	130	140	150	160	170	180
LauCL5217 ORF	CTTAAGCT	céceccéeec	CCGCTTCCGG	гісссвівттс	G G G A A C T G G	CTCCAGGTCG	GCGACGÁCCTC	AACCACCGC.	айтетейсете	ċ
LanC4H.2										
LanC4H.3 LanC4H.4										
		190	200	210	220	230	240	250	260	270
LauCL5217 ORF	CTCGCCAA	AAATTĊGGC	GACTTATTCT	T T C T G A G A A T G	GGTCAGAGG	AATCTCGTAG	IGGTTTCGTCG	CCCGAGCTT	G C C C G T G A T G T	Ť
LanC4H.2										
LawC4H.3 LawC4H 4						т				
2400 412.4						• • • • • • • •				
		290	290	300	310	320	330	340	350	360
LauCL5217 ORF	TTGCACAC	A C A A G G G G T C	GAGTTCGGGT	CGCGGACGAGG	AACGTCGTT	TTCGATATAT	TTACTGGAAAG	GGTCAGGAC	ATGGTCTTTAC	Ġ
LauC4H.1 LauC4H.2										
LauC4H3										
2010-4124										
		370	380	390	400	410	420	130	440	450
LauCL5217 ORF	GTGTATGG	TGAGCACTGG	CGCAAGATGC	GCCGGATCATG	ACGGTGCCG	TTTTTCACCA	A C A A A G T T G T G	AATCAATAC.	AGGTTTGGTTG	Ġ
LanC4H.1 LanC4H.2										
LauC4H.3										
Lanc 41.4										
		460	470	480	490	500	510	520	530	540
LauCL5217 ORF	GAAGACGA	A G C G G C T A G G	GTTGTGGAGG	ATGTGAAGAAG	GACCCCCAG	GCCGCGACTG	AGGGGATTGTT	TTGAGGAGG	AGGTTGCAATT	Ġ
LauC4H.1 LauC4H.2	G					A				
LauC4H.3			<u>G</u>							
Lanc 4H.4										
		550	560	570	580	590	600	510	620	630
LanCL5217 ORF	ATGATGTA	CAACAATATG	TATAGGATTA	T G T T T G A T A G G	AGATTCGAG	AGCGAGGAGG	ATCCATTGTTC	TTGAGACTG	AAGGCCCTGAA	Ť
LauC4H.1 LauC4H.2										
LauC4H.3										
LanC 4H.4										
		640	650	660	670	680	690	700	710	720
LauCL5217 ORF	GGAGAGCG	GAGCAGACTC	GCACAGAGTT	TCGAGTACAAT	TACGGAGAT	TTTATCCCGA	TTTTGAGGCCG	TTCTTGAGA	GGATATTTGAA	ģ
LauC4H.1 LauC4H.2					T					
LauC4H.3									<mark>c</mark>	
LauC4H.4										
		730	740	750	760	770	780	790	800	810
LauCL5217 ORF	ATTTGCAA	GCAAGTCAAG	GAGAGGAGGT	TGCAATTGTTC	AAAGACTAT	TTTCTTGAGG	AGAGGAAGAAG	TTGGCAAGC	ACAAAACCAAC	Å
Low CAU 1										
LauC4H.2					11111111					
LanC4H.2 LanC4H.3					$\ldots, \overset{T}{\underset{T}{}} \ldots$					
LanC4H.2 LanC4H.3 LanC4H.4					T					
LanC4H.2 LanC4H.3 LanC4H.4		\$20	830	840	T	860	\$70	550	\$20	200
LanC4H.2 LanC4H.3 LanC4H.4 LanC4H.4	GATAATGC	820 	830 	840 A T C A C A T C T T G	150 GATGCTGAA	860 	\$70	880 	SPO CTTTACATTGT	- 
LanC4H.2 LanC4H.2 LanC4H.3 LanC4H.4 LanC4H.4 LanC4H.2	GATAATGC	820 	830 TGTGCAATTG.	\$40 A T C A C A T C T T G	T T GATGCTGAA	560 	£70 : 1	580 GACAATGTT	sso CTTTACATTGT	900 G
LanC4H.2 LanC4H.2 LanC4H.3 LanC4H.4 LanC4S217 ORF LanC4H.1 LanC4H.2 LanC4H.2 LanC4H.4	GATAATGC	E20 I G G A C T A A A G	830   TGTGCAATTG.	840 A T C A C A T C T T G	100 T T T T T T T T T T T T T T T T T T	560 A A G A A G G G A G J	\$70	BBD GACAATGTT	SSD CTTTACATTGT	900 G
LanC4H.2 LanC4H.2 LanC4H.2 LanC4H.3 LanC4H.4 LanC4H.4 LanC4H.2 LanC4H.2 LanC4H.2 LanC4H.4	GATAATGC	E20 TGGACTAAAG	530 T G T G C A A T T G	840 A T C A C A T C T T G	50 G A T G C T G A A	560 	870	SSD GACAATGTT	800 CTTTACATIGT	900 G
LanC4H.2 LanC4H.2 LanC4H.2 LanC4H.4 LanC4H.4 LanC4H.1 LanC4H.2 LanC4H.2 LanC4H.2 LanC4H.4	GATAATGC	220 T G G A C T A A A G 910	833 T G T G C A A T T G . 933	840 A T C A C A T C T T G 930	50 G A T G C T G A A	560 	\$70 A A A T T A A T G A G	500 G A C A A T G T T G A C A A T G T T	890 CTTTACATTGT 980	G 900
LanC4H.2 LanC4H.2 LanC4H.3 LanC4H.4 LanC4H.4 LanC4H.2 LanC4H.2 LanC4H.3 LanC4H.3 LanC4H.4 LanCL5217 ORF	GATAATGC	200 1 G G A C T A A A G 910 C A A T G T T G C A	830 T G T G C A A T T G. 930 	840 A T C A C A T C T T G 930 C A A C A C T C T C G G	510 GATGCTGAA 940	₽50 A A G A A G G G A G , 950 	\$60 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.	500 GACAATGT 770 AACCACCCT	900 910 920	900 G
LanCH1.1 LanCH1.2 LanCH2.2 LanCH3 LanC4H.4 LanC4H.4 LanC4H.1 LanC4H.2 LanC4H.3 LanC4H.4 LanC4H.1 LanC4H.2 LanC4H.2	GATAATGC	220 I G G A C T A A A G 910 C A A T G T T G C A	530 T G T G C A A T T G 920 G C C A T C G A G A	SHO ATCACATCTTG 990 CAACACTCTGG	559 GATGCTGAA.	₽50 A A G A A G G G A G A 950 T G G G G T G T A G G	870 A A A T T A A T G A G 990 C A G A G C T A G T G	580 G A C A A T G T T 770 A A C C A C C C T	800 CTTTACATTGT 900 GACATCCAACA	G G A A
LanCH2 LanCH2 LanCH3 LanCH3 LanCH3 LanCH3 LanCH1 LanCH3 LanCH3 LanCH3 LanCH3 LanCH4 LanCH3 LanCH4 LanCH4 LanCH3 LanCH4 LanCH3 La	GATAATGC	ED G G A CT A A A G 910 C A A T G T T G C A	830 1 GT G C A A T T G 900 G C C A T C G A G A 1	840 ATCACATCTTG 990 CAACACTCTGG	50 GATGCTGAA 940 TCCATCGAA	EO A G A A G G G A G 90 T G G G G T G T A G	50 4 A A T T A A T G A G 90 6 A G A G C T A G T G	889 	500 CTTTACATTGT 910 GACATCCAACA	G G A A
LanCH2 LanCH2 LanCH3 LanCH3 LanCH3 LanCH3 LanCH1 LanCH1 LanCH3 LanCH4 La	GATAATGC	E0 I G G A C T À A A G 90 C A A T G T I G C A	500 T G T G C A A T T G . 900 G C C A T C G A G A	540 T C A C A T C T T G 990 C A A C A C T C T G G	50 G A T G C T G A A. S40 T C C A T C G A A	60 1	570	500 G A C A A T G T T 7 770 A A C C A C C T	B0 TTTACATTGT GACATCCAACA	
Lanc4H2 Lanc4H2 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4	GATAATGC GAGAACAT T	E0 G G A C T A A A G 910 C A A T G T T G C A 1000	800 T G T G C AA T T G 900 G C C A T C G A G A 1000	840 ATCACATCTTG 990 CAACACTCTGG 1000	550 G T G C T G A A 540 T C C A T C G A A T C C A T C G A A	800 A G G A G G G A G 970 G G G G T G T A G G 1040	50 AAATTAATGAG 500 CAGAGCTAGTG 1920 1	580 	500 CT TT A C A T T G T 940 G A C A T C C A A C A 1070	
Lanc4H2 Lanc4H2 Lanc4H3 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4 Lanc4H4	GAGAACAT GAGACAT AAGCTCCG	500 1 G G A C T A A A G 910 A A T G T T G C A 1000 1000	830 T G T G C A A T T G 923 G C C A T C G A G A G 930 B000 G C C A C C A T C C	540 ATCACATCTTG 590 CAACACTCTGG 1000	50 GATGCTGAA 940 TCCATCGAA 500 GCCAGATC	100 10 G G G G G G G G G G G G G G G G G G G	E0	50 GACAATGTT 775 AACCACCCT 600 CTCCCGTAC	90 CTTTACATTOT 90 00 1070	
LancH2 LancH2 LancH2 LancH4 LancH4 LancH4 LancH4 LancH2 LancH2 LancH2 LancH3 LancH3 LancH1 LancH1 LancH1 LancH1 LancH1 LancH1 LancH1 LancH2 LancH2 LancH4 LanCH4 La	GATAATGC GAGAACATU T AAGCTCCG	E30 1 G G A C T À À À G 910 1 A À T G T T G C À 1000 1 A À C G A G C T T	830 T T T G C A T T G . 920 G C A T C G A G A 920 G C A T C G A G A	540 1 C A C A T C T T G 930 1020 1020 T G G C C T G G T	550 GATGCTGAA 940 TCCATCGAA GTCCAGATC	500 700 700 700 700 700 700 700 700 700	EQ MAATTAATGAG 900 1 A G A G C T A G T G 1000 1 T A C C A C C C C C C	880   G C A A T G T T ' G A C A T G T T ' A A C C A C C C T A A C C A C C C T C T C C C G T A C	90 90 1070 I TTCCAAGCCGT	
LancH2 LancH2 LancH2 LancH2 LancH2 LancH2 LancH2 LancH3 LancH3 LancH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 LanCH4 La	GATAATGC GAGAACAT T AAGCTCCG	E0 G G A C T A A A G 90 A A T G T G C A 1000 C A A C G A G C T T	833 T G T G C A Á T T G G C C A T C G A G A G C C A T C G A G A 1900 G A C A C C A T C C	840 1 C A C A T C T T G 940 A A C A C T C T C G 1000 1 G G C C T G G	50 GATGCTGAA 90 TCCATCGAA GTCCAGATC	50 A A G A A G G G A G . 70 T G G G G T G T A G ( 1940 A C G G A G C C C G .	570 57 A A T T A A T O A G 960 7 A G A G C T A G T G 5500 1 T A C C A C C G C	50 GACAATGTT AACCACCCT AACCACCCT CTCCCGTAC	90 90 90 90 90 90 90 90 90 90 90 90 90 9	
LancH12 LancH12 LancH13 LancH13 LancH14 LancH11 LancH12 LancH12 LancH13 LancH14 LancH14 LancH14 LancH14 LancH13 LancH14 LancH13 LancH14 LancH13 LancH1	GATAATGC GAGAACAT T	E20 G G A C T À A A G P10 C À A T G T I G C A 1900 1000 1000 1 A A C G A C T T	830 T G T G C A A T T G C G C C A T C G A G A 900 G C C A T C G A G A 900 G C C A T C C A T C C	540 T C A C A T C T T G 700 700 1000 11 G G C C T G G T	550 GATGCTGAA TCCATCGAA TCCATCGAA	50 A G G A A G G G A G A T G G G G G T G T A G 1940 A G G G A G C C C G	500 600 600 600 600 600 600 600	90 GACAATGTT 773 AACCACCCT 900 CTCCCGTAC	500 540 1977	
LmC4H2 LmC4H3 LmC4H3 LmC4H4 LmC4H4 LmC4H4 LmC4H2 LmC4H3 LmC4H3 LmC4H3 LmC4H3 LmC4H3 LmC4H4 LmC4H3 LmC4H4 LmC4H3 LmC4H4 LmC4H3 LmC4H4 LmC4H3 Lm	GATAATGC GAGAACAT T AAGCTCCG	E0 G G A C T A A A G 90 	500 T G T G C A A T T G G C A T C G A G A G C A T C G A G A G A C A C C A T C C H000	840 AT C A C A T C T T G 930 T G G C C T G G G 1100	50 GATGCTCAA GATGCTCAA CCATCGAA GTCCAGATC GTCCAGATC	500 A G A A G G G A G B G G G G G G T A G C G G G G G G C C G G A C G G A G C C C G	E3 1 A A T T A A T G A G 900 1 A G A G C T A G T G 1000 1 T A C C C A C C G C 1100 1	90 90 90 90 10 10 10 10 10 10 10 10 10 1	590 TTTACATTGT 980 ACATCCAACA 1970 TCCCAAGCCGT 1100	
LmC4H12 LmC4H3 LmC4 LmC4 LmC4 LmC4 LmC4 LmC4 LmC4 LmC4	GATAATGC GAGAACAT T AAGCTCCG	E3 1GGACTAAAG 99 2AATGTTGCA 3000 3000 3000 3000 3000 3000 3000 30	500 107 G C A A T T G . 500 G C C A T C G A G A 1000 C T C C G C A T C C 1000 C T C C C C A T C C	99 299 caacactctcc 100 110 ccatcccccccc 110 ccatccccccccc 110 ccatcccccccccc 100 100 100 100 100 100	100 100 100 100 100 100 100 100	50 A G G A G G G G G G G 1 G G G G G G T A G C G G G G C C C G C G G A G C C C G	E3 1 AAATTAATGAG 560 1 CAGAGCTAGTG 1000 1 1 ATACCCACCGC 1010 1 1 TCTACGATGCC	99 A A C C A C C C C T T 90 10 10 10 10 10 10 10 10 10 1	590 590 500 500 500 500 500 500 500 500	
LmcR412 LmcR413 LmcR43 LmcR43 LmcR43 LmcR41 LmCR41 LmCR41 LmcR41 LmcR41 LmCR41	GAGAÁCAT T AAGCTCCG	100 10 G A C T A A A G 190 10 A A T G T I G C A 100 100 100 100 100 100 100 100 100 10	80 50 50 100 66 67 67 67 67 67 67 67 67 67 67 67 67	на 99 99 100 110 оссот о о о 110 оссот о о о о	58 56 56 A T G C T G A A 58 7 C C A T C G A A 50 7 C C A G G A T C 5 C C A T C G A A 5 C C A T C G A A 5 C C A T C C A G A T C 5 C C C C C C C C C C C C 5 C C C C C	60 99 16 6 6 6 7 6 7 4 6 10 6 6 6 7 6 7 4 6 10 6 10 6 10 6 10 6 10 6 10 6 10 6 10	10 10 10 10 10 10 10 10 10 10	99 GACAATGTT 97 AACCACCCT TCCCGTAC 10 AAGCTCGGC	90 GACATCCAACA 1070 CTCCAAGCCGT 1000 T GGCTATGACAT	
Lanc(H1) Lanc(H2) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H4) Lan	GAGAACAT AAGCTCCG	59 6 G G G C T A A G 59 6 A A T G T T G G A 7 A A T G T T G G A 7 A A C G A G C T T 7 A A C G A G C T T 7 A A C G A G C T T	00 10 1 0 C A A T I G 0 C A T C A G A G 0 C A T C A G A G 0 A C A T C C A T C C 0 A C A T C C A T C C 100 100	100 100 100 100 100 100 100 100 100 100	100 100 100 100 100 100 100 100	ы 90 100 100 100 100 100 100 100	13 I I I I I I I I I I I I I I I I I I I	980 0 A C A A T G T T 193 193 195 195 195 195 195 195 195 195 195 195	10 10 10 10 10 10 10 10 10 10	· · · · · · · · · · · · · · · · · · ·
Lanc(H1) Lanc(H2) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H4) Lan	GATAATGC GAGAACAT T. AAGCTCCG	ра 199 199 199 199 1990 1990 1990 1990 19	50 50 60 CATCAGA 60 CATCAGA 60 CATCAGA 100 100 100 100 100 100 100 100 100 10	100 110 110 110 110 110 110 110 110 110	58 GATGCTGAA TCCATCGAA TCCATCGAA TCCATCGAA TCCATCGAA TCCATCGAA TCCATCGAA GTCCAGATC GCCCC GC	60 59 10 C C G T G T A G 10 C C G A G C C C G 100 100 100 100 100 100 100 100 100 10	19 19 10 10 10 10 10 10 10 10 10 10 10 10 10	293 GACAATGTT 793 AACCACCCTT 600 CTCCCGTAC 193 AACCTCGGC	199 194 194 195 1 195 1 195 1 196 1 196 1 196 1 196 1 196 1 196 1 197 1	
Lanc(H12) Lanc(H13) Lanc(H14) Lanc(H15)7) OBF Lanc(H17) OBF Lanc(H12) Lanc(H12) Lanc(H13)7) OBF Lanc(H13)7) OBF Lanc(H12) Lanc(H14) Lanc(H13)7) OBF Lanc(H14) Lanc(H14	GATAATGC GAGAACAT T AAGCTCCG	59 6 G A C T A A A G 90 A A T G T T G C A 90 A A C G A G C T T 90 A A C C G A G C T T 90 A A C C C T C C G 90 A A C C T C C C 90 A A C C T C C A 90 A C C T A A A G 90 A C T A A C T A C T A C A A A G 90 A C T A C T A C T A C A A A G 90 A C T A C T A C T A C A	53 55 6 C C A T T G 6 C A T G A G A 56 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	на 393 салсастото с 110 сол сосото с 110 сол сососто с 110 сол сососто с 110 сол сососто с 110 сососто с с с с с с с с с с с с с с с с с с с	1000 100 1000 1	во 350 10 0 0 0 1 0 0 0 1 0 0 10 0 0 0 1 0 1	19 10 10 10 10 10 10 10 10 10 10 10 10 10	88 G A C A A T O T T A A C C A C C C T T C C C G T A C T C C C G G C A A G C T C G G C	100 CTTTAAAITOT 90 DAAATCCAACA 100 CTTCCAACCCT 100 CTTCCAACCCT 100 CTTCCAACCCT	
Lunc242) Lunc243 Lunc2443 Lunc2443 Lunc2443 Lunc2443 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc2447 Lunc247	GÁ TÁÁ TGC GÁ GÁÁ CÁT T T A Á GC TCCG	500 500 C C T A A A G 500 C A A T G T T G C A 500 C A A C G A G C T T C A A C G A G C T T C A A C G A G C T T 100 C C C C C C C C C C C C C C C C C C C	90 90 0 C A T C A T T C 90 0 C C A T C A A T A 0 C C C C C A C A T C C 100 10	100 100 100 100 100 100 100 100 100 100	113 113 113 113 113 113 113 113	89 99 103 103 103 103 103 103 103 103 103 103	F3 1 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	20 20 20 20 20 20 20 20 20 20	199 199 199 1070 1 1 1 A C A T T G T 199 1070 1 1 1070 1 1070 1 1 1070 1 100 1 100 1 100 1 100 1 100 1 1	
LuncH12 LuncH13 LuncH13 LuncH13 LuncH13 LuncH12 LuncH12 LuncH12 LuncH12 LuncH13 LuncH1	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA CCCGCTGA	59 50 A A T G T T G C A 300 C A A C G A G A C C T T 300 C A A C G A G A C C T 100 C C 100 C C	50 50 50 50 50 50 50 50 50 50 50 50 50 5	50 50 50 50 50 50 50 50 50 50 50 50 50 5	113 113 113 113 113 113 113 113	100 100 100 100 100 100 100 100 100 100	193   1 4444TTAATGAG 999   1 46646CTA6TG 199   1 199   1 190   1 10 10 10 10 10 10 10 10 10 10 10 10 10	99 GACAATOTT AACCACCT CTCCCGTAC AAGCTCGGC	нр СТТТАСАТТОТ БАСАТССААСА 1000 П СТССААССС 300 СТАТСААСА 1 1 100 П СССССААСАС	
LancH12 LancH2 LancH32	64744760 64644047 44607000 47044404 47044404	59 6 G G G C T A A G 59 6 A A T G T T G C A 50 7 A A C G A G C T T 50 7 A A C G A G C T T 50 7 A A C G A G C T T 50 7 A A C G A G C T T 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A C A G 50 7 A A C G A G C T A A G 50 7 A A C G A G C T A C A G 50 7 A A C G A G C T A C A G 50 7 A A C G A G C T A C A G 50 7 A A C G A G C T T C A A C G A G C T T 50 7 A A C G A G C T T C A A C G A G C T T 50 7 A A C G A G C T T C A A C A G A G C T T 50 7 A A C G A G C T T C A A C A G A G C T T 50 7 A A C G A G C T T C A A C A G A G C T T 50 7 A A C G A G C T T C A A C A G A G C T T 50 7 A A C G A G C T T C A A C A G A G C T T 50 7 A A C G A G A C T T C A A C A G A G C T T 50 7 A A C G A G A C T T C A C A G A C T T C A C A G A C T T C A C A G A C T T C A C A G A C T T C A C A C A G A C T T C A C A C A G A C T C A C A C A C A C A C A C A C A C	83 60 C C A T C C A T C C 80 C C A T C C A C A C C 80 C C A T C C A C A C C 80 C C C A T C C C 80 C C C C C A T C C 80 C C C C C C C C C C C C C C C C C C C	100 100 100 100 100 100 100 100 100 100	на а т о с т о а а а а т о с т о а а а т с с а т с с а а с с а а с с т т о с с а а с с а а с с т т о с с а а с а а с с с т о с с а а с а а с	ы) 90 1 а а а а а а а а а а а 1 а а а а а а а	13 14 A T T A A T G A G 190 1 C A G A G G T A G G 1 T A C C A C C G G 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C 1 T A C C A T C C C C C 1 T A C C A T C C C C C C C 1 T A C C A T C C C C C C C C C C C C C C	99 79 A A C C A T C T T A A C C A C C C T T C C C C G T A C T C C C C G T A C A A G C T C C G G C A A G C T C C G G C	100 100 100 100 100 100 100 100 100 100	· · · · · · · · · · · · · · · · · · ·
LuncH12 LuncH13 LuncH13 LuncH13 LuncH12 LuncH12 LuncH12 LuncH12 LuncH12 LuncH13 LuncH1	GÁTAÁTGĆ GÁGAÁCAT T AAĞCTĊĊĠ ATĊAÁAĞA CCCGĊTĞA, T	50 50 A CT A A A G 50 A A T G T O C A 50 A A C G A G A C T T 50 A A C G A G A C T 100 100 100 100 100 100 100 100 100 10	59 60 C C C C C C C C C C C C C C C C C C C	ы нтсасатотто на састатото по састато са тоссоста от по са саста от са тоссоста от са са са са са са са на са са са са са са на са са са са са са са са са на са са са са са са са са са на са са са са са са са са са на са са са са са са са са са на са са са са са са са са са на са са са са са са са са са са на са са са са са са са са са на са са са са са са са са са на са са са са са са са са са на са са са са са са са са са на са са са са са са са са са са на са	100 100 100 100 100 100 100 100	ы 59 Табаасаа Сасатаа Сасатаа Сасатаа Сасатаа Сасатаа Сасатаа Сасатаа Сасатаа Сасатаа	193	979 779 779 770 770 770 770 770 770 770	HP HP GACATCCAACA HP HP GACATCCAACA CCCCAAGCGGT HD HD HD HD HC TCCCAAGCGGT T H HD HC HC HC HC HC HC HC HC HC HC HC HC HC	
LancH2 LancH2 LancH3 La	6474476C 64644C47 7 446c7CC6 47C44464 CCC6C764	55 56 A CT A A A G A A T G T T G C A 58 58 58 50 50 50 50 50 50 50 50 50 50 50 50 50	53 55 6 C C A T C G A G A 55 6 C C A T C G A G A 56 7 C C C C A T C G A 57 7 C C C C A T G C 7 C C C C A T G C 7 C C C C A T G C 7 C C C C C C C C 7 C C C C C C C C C	на 199 299 СААСАСТСТОС 199 199 199 199 199 199 199 19	1005 1005	во ма б А А Ф Ф Ф А А С 99 10 6 6 6 7 6 1 А С 100 100 100 100 100 100 100 10	13 14 15 15 15 15 15 15 15 15 15 15	100 100 100 100 100 100 100 100	109 TTTTACATOT 100 100 100 100 100 100 100 10	
Lanc(H2) Lanc(H2) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H3) Lanc(H4) Lanc(H3) Lanc(H4) Lan	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA ATCAAAGA T TTCTTTGA	ну солсталас латот осл зассотосос с с цр цр залосласт с с цр залосласт с с с	50 60 C C A C C A C A C A C A C A C A C A C	100 100 100 100 100 100 100 100 100 100	103 103 103 103 103 103 103 103	59 199 1995 1995 1995 1995 1995 1995 19	F3 1 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	293 293 294 295 295 296 297 297 298 298 299 299 299 299 299 299	на на басатесьаса изо сталаса осталаса изо стала и изо сталаса и и стала и и стала и и и и и и и и и и и и и и и и и и	
LancH2 LancH2 LancH2 LancH2 LancH2 LancH2 LancH3 LancH2 La	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA CCCGCTGA T T	59 6 GACTAAAG 99 AATGTTGCA 999 CAACGAGCTT CAACGAGCTT CC 109 109 CAACGAGATC C 109 109 C	50 60 10 C A A T I G 60 C A T C G A G A 60 C C A T C G A G A 100 C C C C C C A T C C 100 C C C C C C A T C C 100 C C C C C C A T C C 100 C C C C C A T C C 100 C C C C A T C C 100 C C C C A T C C 100 C C C C C C C C 100 C C C C C C C C C 100 C C C C C C C C C C 100 C C C C C C C C C C C 100 C C C C C C C C C C C C 100 C C C C C C C C C C C C C C C C C C C	50 100 100 100 100 100 100 100 100 100 1	100 100 100 100 100 100 100 100	на 590 10 сосот а сосо 10 сосот а сосот а сосо 10 сосот а сосот а сосот 10 сосот а сосот а сосот а сосот 10 сосот а сосот а сосот 10 сосот а сосот а сосот а сосот а сосот 10 сосот а сосот а сосот а сосот а сосот 10 сосот а сосот а сосот а сосот а сосот 10 сосот а сосот а сосот а сосот а сосот а сосот 10 сосот а сосот а сосот а сосот а сосот а сосот 10 сосот а сосот 10 соста сост	19 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	200 200 200 200 200 200 200 200	на на на на на на на на на на	
Lanc(H2) Lan	GĂTĂĂTGC GĂĞĂĂĊĂŢ T AĂĞĊŢĊĊĞ AŢĊĊĂĠĂĂ Ţ ŢŢŢŢŢŢĠĂ	59 6 G A C T À A A G 59 7 A A T G T T G C A 7 A A T G T T G C C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A 7 A A T G T T G C A A A A T G T T G C A A A A T G T T G C A A A A T G T T G C A A A A A T G T T G C A A A A A A A A A A A A A A A A A A	83 60 C A T C A C A T C 80 60 C A T C A C A C A 100 60 C A C C A T C C 100 60 C A C C A T C C 100 60 C A C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 60 C A T C C A T C C 100 70 C A T C C A T C C 100 70 C A T C C A T C C 100 70 C A T C C A T C C 100 70 C A T C C A T C C 100 70 C C C A T C C A T C C 100 70 C C C C A T C C C 100 70 C C C C A T C C C 100 70 C C C C A T C C C 100 70 C C C C C A T C C C 100 70 C C C C C C C C C C C C C C 100 70 C C C C C C C C C C C C C C C C C C C	100 100 100 100 100 100 100 100 100 100	100 G T C C A T C C A A G T C C A T C C A A G T C C A A C C A A G T C C A A C A A C G C C A C A A C G C C A C A A C G C C A C A C A C G C C A C C A C C A C G C C A C C A C C C G C C C C C C C C C C C C G C C C C C C C C C C C C C C C C C C C	ы) 99 1 а а а а а а а а а а а 1 а а а а а а а	13 14 A T T A A T G A G 15 A G G T A G G 14 A T T A A T G A G 15 A G G T A G G 16 A G G T A G G 17 A C G A T G C G 16 A A A A A C C C C 10 B 1 10 G 10 G	203 203 204 205 205 205 205 205 205 205 205	100 100 100 100 100 100 100 100 100 100	
LancH2 LancH2 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 La	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA CCCGCTGA TTCTTTGA	59 6 G A CT A A A G A A T G T T G C A 3950 1950 1950 1950 1950 1950 1950 1950 1	59 60 C C A C C A C A C A C A C A C A C A C	56 57 59 59 50 50 50 50 50 50 50 50 50 50 50 50 50	103 103 103 103 103 103 103 103	на 590 10 сосотота с 10 сосот	193   1 40 A A T T A A T G A A 40 A A C T A G G 40 A A C T A G G 40 A C T A C T A C T A G G 40 A C T A C	200 200 200 200 200 200 200 200	на СТТТАСАТТОТ БАСАТССААСА НИЗО П СТССААССА 100 П 100 П 100 П 100 П 100 П 100 П	
LancH2 LancH2 LancH2 LancH2 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 La	GATALTGC GAGALCAT T AAGCTCCG ATCALAGA CCCGCTGA T TTCTTTGA	55 56 G G C T A A A G A A T G T T G C A 58 58 58 50 50 50 50 50 50 50 50 50 50 50 50 50	53 53 55 55 55 55 55 55 55 55 55 55 55 5	на 199 СААСАСТСТО 199 190 100 100 100 100 100 100	100 100 100 100 100 100 100 100	100 100 100 100 100 100 100 100	19 19 10 10 10 10 10 10 10 10 10 10 10 10 10	200 200 200 200 200 200 200 200	109 TTTTACA ITOT 100 00CCATCCAACA 110 110 110 110 110 110 110 11	
Lanc(H2) Lan	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA CCCGGCTGA T T TTCTTTGA	59 50 A C T A A A G C A A T G T I G C A 50 50 50 50 50 50 50 50 50 50 50 50 50	59 60 C C C C C C C C C C C C C C C C C C C	100 100 100 100 100 100 100 100 100 100	100 100 100 100 100 100 100 100	ыр 750 16 бой бой бой 10 бой бой бой 10 бой бой бой 10 бой бой бой 10 б	193	779 779 779 770 770 770 770 770	HP HP GACATCCAACA HP HP GACATCCAACA HP HP HP HP HP HP HP HP HP HP	
LancH2 LancH2 LancH2 LancH2 LancH3 La	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA T TTCTTTGA	59 00ACTAAAG AATGTTGCA AATGTTGCA AAGGAGGCTT LU9 LU9 LU9 LU9 LU9 LU9 LU9 LU9 LU9 LU9	53 1010CATTG CCATCGAGA 100 GACACCATCC 100 CTCCCCATGG 100 CTTGCAATG 100 CTTGCAATG	на 99 Салсаст стоб 100 100 100 100 100 100 100 100 100 10	109 109 109 109 109 109 109 109	на 595 10 абай абай 10 абай абай 10 абай абай 10 абай абай 10 абай абай 10 абай 10 абай 10 абай 10 абай 10 абай 10 абай 10 абай 10 абай 10 10 10 10 10 10 10 10 10 10	19 1 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	200 200 200 200 200 200 200 200	на статаса тот на оссатссааса по стоссааса по стоссааса по стоссааса по по стоссааса по по стоссааса по по по по по по по по по по по по по	· · · · · · · · · · · · · · · · · · ·
Lanc(H2) Lan	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA T TTCTTTGA	59 50 C C C C C C C C C C C C C C C C C C C	50 60 TO TO CÁATTO. 50 60 CATCAGA 60 CATCAGA 100 100 100 100 100 100 100 100 100 10	100 100 100 100 100 100 100 100 100 100	103 103 103 103 103 103 103 103	190 590 ТС СС С	110 1 110 1 11	279 379 370 370 370 370 370 370 370 370	ни ни басатссааса што стосааса што статасааса щто постатасааса щто постатасааса щто постатасааса щто постатасааса щто постатасааса щто постатасааса щто постатасааса постатасааса щто постатасааса постатаса поссе постатаса постатаса постатаса поссе по	
LancH2 LancH2 LancH3 LancH4 La	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA CCCGCTGA TTCTTTGA	59 6 G A CT A A A G A A T G T T G C A A A T G T T G C A A T C U U U U U U SG A G G A G G A G C T U U SG A G G A G G A G C T U U SG A G G A G G A G C T C U U SG A G G A G G A G C T C U U SG A G G A G G A G C T C U SG A G A G A G A G A G C T C U SG A G A G A G A G A G C T C U SG A G A G A G A G A G C T C U SG A G A G A G A G A G C T C U SG A G A G A G A G A G A G C T C U SG A G A G A G A G A G A G C T C U SG A G A G A G A G A G A G C T C U SG A G A G A G A G A G A G C T C U SG A G A G A G A G A G A G A G A G A G A	50 50 50 50 50 50 50 50 50 50 50 50 50 5	50 50 50 50 50 50 50 50 50 50 50 50 50 5	103 103 103 103 103 103 103 103	на 59 10 соста соста с 100 100 100 100 100 100 100 100 100 10	на і і і і і і і і і і і і і і і і і і і	200 200 200 200 200 200 200 200	но Басатссааса 100 п 100 п 10	· · · · · · · · · · · · · · · · · · ·
Lanc(H2) Lanc(H2) Lanc(H3)Lanc(H3)Lanc(H3)Lanc(H3)Lanc(H3)Lanc(H3)Lanc(H3)Lanc(H3)La	GĂTĂĂTGC GĂĞĂĂĊĂŢ Ţ Ţ Ţ Ţ Ţ Ţ	59 59 CATGACTAAAG CATGTCCA CACGAGCCT CACGAGCCTGCG CC CC CC CC CC CC CC CC CC CC CC CC	ва Тотослатта, 527 GCCATCAGA 100 CTCCGCATCC 100 CTCCCGCATCC 100 CTTCCAATC 100 CTTCCAATC 100 CTTCCAATC 100 CTTCCAATC 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAATCA 100 CTTCCAACAACAACAACAACAACAACAACAACAACAACAA	на 393 САСАСАТСТСО 100 110 110 110 110 110 110 11	100 100 100 100 100 100 100 100	100 100 100 100 100 100 100 100	13 14 14 14 14 14 14 14 14 14 14	20 20 20 20 20 20 20 20 20 20	ня мр ассатссалса на на на на на на на на на н	
LancH22 LancH23 LancH32 LancH34 LancH34 LancH32 LancH32 LancH32 LancH32 LancH32 LancH32 LancH33 LancH4	GATAATGC GAGAACAT T AAGCTCCG ATCAAAGA CCCGCTGAA ATACTTCC	599 60 A CT TA A G A A T CT T C C A 3990 A A C G A G C T T 3990 A C C C T C C G C C C C C C C C C C C C C	59 60 TO C A A TT C 59 60 C A C C A TT C 100 100 100 100 100 100 100 10	50 50 50 50 50 50 50 50 50 50	1133 1135 1135	100 100 100 100 100 100 100 100 100 100	193 1 AAATTAATGAA 99 1 AGAGCTAGTG AGAGCTAGTG 119 1 TCCCCACCCC 110 1 100 100	20 20 20 20 20 20 20 20 20 20	на на басатссааса на сссааса на на на на на на на на на н	· · · · · · · · · · · · · · · · · · ·
LancH2 LancH2 LancH2 LancH2 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH4 LancH3 LancH3 LancH4 La	GĂTĂĂTGC GĂĞĂĂĊĂŢ Ţ Ţ Ţ Ţ Ţ Ţ Ţ Ţ Ţ	59 6 G A C T A A A G A A T G T T G C A 1990 A A C G A G G C T T 1990 A A C C A G G G C T T 1990 A A C C A G G A G C T 1990 A A C C A G A G A G C A 1990 A A C C A G A G A G C A 1990 A A C C A G A G A G C A 1990 A A C C A G A G A G C A 1990 A A C C A G A G A G C A 1990 A A C C A G A G A G C A 1990 A A C C A G A G C A G C A 1990 A A C C A G C A G C A 1990 A A C C A G C A G C A 1990 A A C C A G C A G C A C C 1990 A A C C A G C A G C A C C 1990 A A C C A G C A G C A C C 1990 A A C C A G C A G C A C C 1990 A A C C A G C A C C A C C 1990 A A C C A G C A G C A C C A 1990 A A C C A G C A G C A C C A 1990 A A C C A G C A G C A C C C 1990 A A C C A G C A G C A C C C 1990 A A C C A G C A G C A C C C C C C C C C	53 10 10 C A A T I G C C A T G A G A G A C A T C A G A G A C A C C A T C C C C C C C A T G C C T C C C A T G C C T G C C A T G C C T G C G A A C C T G C G A A C C T G G G G A T C A 100 C T G G G G A T C A 100 C C C C C C A T G C C C C C C A T C C C C C C C C C C C C C C C C C C C	на 59 СААСАСТСТО 100 100 100 100 100 100 100 10	109 109 109 109 109 109 109 109	на 99 10 сосос сосос 100 100 100 100 100 100 100 10	19 1 10 10 1 10	100 100 100 100 100 100 100 100	199 CTTTACATOT 190 0ACATCCAACA 190 100 100 100 100 100 100 100	· · · · · · · · · · · · · · · · · · ·
Lanc(H2) Lan	GATAATGC GAGAACAT T AAGCTCCG CCCGCTGA T TTCTTTGA ATACTTGC	599 597 598 598 599 599 599 599 599 599 599 599	59 60 C C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C	100 100 100 100 100 100 100 100 100 100	100 100 100 100 100 100 100 100	ыр 590 Табаалааса, 109 Сасатаа, Сасата, Сасата, Саса, Сасата, Саса, Саса, Саса, Саса, Саса, Саса, Сас	193   AAAATTAATGAA AAAATTAATGAA AAAAGCTAGTG 199   AGAGCTAGTGGA 199   199   AGAGCTAGAGCTAGA 199   190   190	20 20 20 20 20 20 20 20 20 20	на на басат ссалса на по стесалеса по стесалеса на т т соссеба соссалеса на на на на на на на на на н	····· \$200-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0

**Figure A1.** Coding sequence alignment of *Lau*C4H variants. Four unique variants of *Lau*C4H gene are aligned against the original sequence predicted by the de novo assembled transcriptome. Dots are identical bases.

		10	2	0	30	40	50	60
LauC4H.1	MDLILLEK	SĹLAVI	FAIIFS	IIISKLR	GKKLKLPPG	PLPVPVFGN	WLQVĠDDLNI	HRNLTS
LauC4H.2								
LauC4H.3								
LauC4H.4								
		70				100		120
			8 		.			120
LauC4H.1	LAKKFGDL	FFLRM	ORNLVV	VSSPELA	RDVLHTQGV	EFGSRTRNV	VFDIFTGKG	DMVFT
LauC4H.2								
LauC4H.3								
LauC4H.4								
		130	14	10	150	160	170	180
LauC4H.1	VYGEHWRK	MRRIM	VPFFTN	KVVNOYR	FGWEDEAAR	VVEDVKKDP	OAATEGIVL	RRRLOL
LauC4H.2					E			
LauC4H.3								
LauC4H.4					E			
					#			
		190	20	0	210	220	230	240
LanC4H 1	MMYNNMYR	IMEDRE	FESEET	PIFIRIK	AINGERSRI	AOSEEVNYG	DETRIERE	IRGVIK
LauC4H 2			(FESEEL	FEFERER	RENOERSKE	AUSTEINIU	DETETERE	LKOILK
LauC4H.3								
LauC4H.4								
		250	26	50	270	280	290	300
					· ·   · · · ·   ·			
LauC4H.1	ICKQVKER	RLQLFF	DYFLEE	KKKLASI	K P I D N A G L K	CAIDHILDA	EKKGEINEDI	NVLYIV
LauC4H 3								
LauC4H.4								
		310	32	20	330	340	350	360
I COLL					.	· · ·   · · · ·		
LauC4H.1	ENINVAAI	EIILWS	SIEWGVA	ELVNHPD	IQQKLKNEL	DITLGPGVQ	TIEPDIAKL	PILQAV
LauC4H.3								
LauC4H.4								
		370	38	80	390	400	410	420
ICIII I				ND AKI CC			· · · ·   · · · · ·	
LauC4H.1	TKETLKLK	MAIPLI	V P HMN I	I DAKLGG	TDIPAESKI	LVNAWWLAN	NPAHWKNPE	EFKPER
LauC4H 3								
LauC4H.4								
		430	44	10	450	460	470	480
		-						
LauC4H.1	FFEEEAEV	EANGNI	JFRYIPF	GVGRRSC	PGIILALPI	LGITIGRLV	QNFELLPPP	GQGKLD
LauC4H.2		• • • • • •						
LauC4H.4								
		490	50	00				
		490	50   · · ·	00				
LauC4H.1	 T T <b>E K </b> G G Q F	490 .   S L H I L K	50     CHSTIVA	00       K P R L F				
LauC4H.1 LauC4H.2 LauC4H.3	T T <b>E K G G Q F</b>	490 .   S L H I L K	50     CHSTIVA	00 A K P R L F				
LauC4H.1 LauC4H.2 LauC4H.3 LauC4H 4	T T E K G G Q F	490 -   S L H I L F	50     CHSTIVA	X P R L F				

**Figure A2.** Protein sequence alignment of *Lau*C4H variants. Four unique variants of *Lau*C4H protein are aligned against the original sequence predicted by the de novo assembled transcriptome. Dots are identical residues.



**Figure A3.** The expression level and solubility of the chimeric  $ATR2^{\Delta 2-74}$ -*Lau*C4H<sup> $\Delta 2-28$ </sup> fusion protein in the *p*-coumaric acid producers. Strain 1, Ec/*Lau*C4H; Strain 2, Ec/*Lau*C4H-*Ath*PAL; Strain 3, Ec/*Lau*C4H-*Ath*PAL-sRNA; Strain 4, Ec/*Lau*C4H-*Ath*PAL-PntAB; Strain 5, Ec/*Lau*C4H-*Ath*PAL-anti(sthA); Strain 6, Ec/*Lau*C4H-*Ath*PAL-PntAB-anti(sthA). KDa, kilodalton; M, protein marker; WP, whole-cell proteins; IP, insoluble proteins; SP, soluble proteins. The arrows indicated the chimeric  $ATR2^{\Delta 2-74}$ -*Lau*C4H<sup> $\Delta 2-28$ </sup> fusion protein.

ATGGCGCCGCGTCCGACCAGCCAGTCTCAGGCTCGTACCTGTCCGACCACCCAGGTTACCCAGGTTGATATTGTTGAAAAAATGCTGGCG GCGCCGACCGACCCCCTGGAACTGGACGGCTATAGCCTGAACCTGGGTGATGTAGTGAGCGCCGCCGCTCGTAAAGGCCGTCCGGTGCGT ACCGGCTTCGGTGGCAGCGCTGATACTCGTACCGAAGATGCGATCTCCCTGCAGAAAGCCCTGCTGGAACATCAGCTGTGCGGCGTTCTG ccgrtttcctttgatagctrccgcctgggtcgtggcaaaacagcctgccgctggaagttgttcgtggcgcgatgactatccgtgtg AACAGCCTGACCCGTGGTCACAGCGCTGTTCGCCTGGTGGTGCTGGAAGCTCTGACCAACTTCCTGAACCACCGGTATCACCCCCGATCGTG ccgctgcggtaccattagcgcgtctggtgatctgagcccgctgtcttacatcgcagcggctattagcggtcacccggattccaaagta CACGTTGTTCACGAGGGTAAAGAAAAATCCTGTATGCCCGTGAAGCGATGGCCCTGTTTAACCTGGAACCGGTGTGTGGGCCCGAAA GAGGGCCTGGGTCTGGTGAACGGTACCGCGGTTTCTGCAAGCATGGCAACCCTGGCACTTCATGACGCGCACATGCTGAGCCTGCTGAG CAGTETETEGACCGCCGATGACCGTTGAGGCCGATGGTTGGCCACGCGGGTTETTTTCACCCCGTTTETGCACGATGTAACCCCGCCCGCACCCG A CT CA GA T C GA A GT G G CT G C CA A C T C C G T A A A C T G C T G G G A C G T T C G C T C G C T C G C A C G A G G A G G T G A G G T T A A G A C 1040 1050 1020 1030 GACGAAGGCATCCTGCGTCAGGATCGTTACCCGCTGCGTACCTCCCCGCAATGGCTGGGTCCGCTGGTGTCTGACCTGATCCACGCTCAC 1110 1120 1130 1140 GCGGTTCTGACTATTGAAGCCGGTCAGAGCACCACCGATAACCCGCTGATCGATGTTGAAAACAAAACTAGTCATCACGGCGGTAACTTC 1210 1220 1230 CAGGET G C G G C T G C G A C A C C A T G G A A A A A C C C G T C T G G G C T C A G T C G G C A A C T G A C T T C A C C C A A C T G A C C G A A A T G 1310 1320 CTGAATGCGGGGATGAACCCTGGTCTTCCGAGCTGTCTGGCAGCTGAAGACCCGTCCCTGTCCTACCACTGCAAAGGTCTGGACATTGCG 1410 1420 GCAGCAGCGTACACCTCTGAACTGGGTCACCTGGCAAACCCGGTGACCACCCATGTTCAGCCGGCGAGATGGCAAACCAGGCGGTTAAC 1470 1490 1500 1510 TCCCTGGCGTTGATTTCTGCTCGTCGCACCACCGAAAGCAACGATGTTCTGAGCCTGCTGCTGCCGACCCATCTGTACTGCGTTCTGCAG 1570 1580 GCGATCGACCTGCGTGCTATCGAATTCGAATTCAAAAAACAGTTCGGCCCGGCGATCGTTAGCTTAATTGATCAGCACTTTGGCTCCGCT 1640 1650 1670 1680 1660 ATGACCGGCTCTAACCTGCGCGACGAACTGGTCGAAAAAGTCAACAACCCTGGCTAAACGCCTGGAACAGACCAACTCTTACGATCTG 1740 1760 GTTCCGCGCGGCATGACGCGTTTTCCTTTGCGGCGGGCACCGTCGTTGAAGTTCTGTCCTCTACCAGCCTGAGCCTGGCGGCAGTGAAC 1830 1840 1850 GCGTGGAAAGTGGCAGCCGCTGAAAGCGCGATTAGCCTGACCCGTCAGGTTCGTGAAACTTTTTTGGTCCGCTGCGAGCACCTCTAGCCCG 1940 1950 1930 CTGGGTAAACAGGAAGTGACCATCGGTAGCAACGTTTCTAAAATCTATGAAGCAATTAAAAGCGGTCGTATCAACAACGTTCTGCTGAAA ATGCTGGCGTAA

Figure A4. The codon-optimized DNA sequence of the yeast *Rhodotorula glutinis* PAL/TAL.

Name	Sequence (5' to 3')
RT-PCR	
RT-LauC4H-PF	AAGGGAAGCTTGATACCACTGAGAA
RT-LauC4H-PR	GACAAATTAGAACAGTCTAGGCTTGGC
RT-LauTIP41-PF	GCAACCATCCAAAGTTTAACTGCT
RT-LauTIP41-PR	AATGTGCAAGCAGGGCTAGTAA
cDNA cloning	
LauC4H-ORF-PF	ATCCTCCTCCGACGAAATG
LauC4H-ORF-PR	AGAGTACATTGCATGGGTAATAAGGAG
LauC4H-5'RACE-PR	GAACTCGACCCCTTGTGTGTGC
LauC4H-3'RACE-PF	GGCTATGACATCCCCGCTGA
Subcellular localization	
pAN580-LauC4H-PF	AGGACCGGTCCCGGGGGATCCATGGATCTCATTCTACTAGAAAAGTCACTC
EGFP-LauC4H-PR	TCCTCGCCCTTGCTCACCATGAACAGTCTAGGCTTGGCCACAATG
EGFP-LauC4H(N28)-PR	TCCTCGCCCTTGCTCACCATCCCGCGGAGTTTGGATATTATGAT
pAN580-LauC4H(ΔN28)-PF	AGGACCGGTCCCGGGGGATCCATGAAGAAGCTTAAGCTCCCCCGG
Functional expression	
29aNdeI-LauC4H(∆N28)-PF	AACTTTAAGAAGGAGATATACATATGAAGAAGCTTAAGCTCCCCCGG
	AGTGGTGGTGGTGGTGGTGCTCGAGTTAGAACAGTCTAGGCTTGGCCAC
29aXnol-LauC4H-PK	AATG
29aNdeI-ATR2( $\Delta$ N74)-PF	AACTTTAAGAAGGAGATATACATATGTCCGGTTCTGGGAATTCAAAA
29aXhoI-ATR2-PR	AGTGGTGGTGGTGGTGGTGGTGCTCGAGTGAGTGTGTGGCTTCAATAGTTTCG
	CCGGGGGGGGGCTTAAGCTTCTTCATACCAGAACCAGAAGAGGTAGAACC
$LauC4H(\Delta N28)A1R2-PR$	CCATACATCTCTAAGATATCTTCCACTCGTTTG
De novo biosynthesis	
184-trc-lacO-PF	<u>GCCTTGCGTATAATATTTGCCCATGG</u> TTGACAATTAATCA
101 40 1400 11	TCCGGCTCGTATAATGTGTGGGAATTGTGAGCGG
BBa B0034-lacO-PR	<u>CATATGTATTTCTCCTCTTTAGAT</u> CTGGAATTGTTATCCGCTCA
	CAATTCCACACATTATAC
	<u>ATCTAAAGAGGAGAAATACATATG</u> CTCGAGTAACCAGGCATCAA
BBa_B0034-BBa_B0015-PF	ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAT
	CTG TTGTTTGTCGGTGAACGCTCTCTACTA
184-BBa B0015-PR	<u>GCCCGCCTGATGATGCTCATCCGGAATTC</u> TATAAACGCAGAAAGGCCC
	ACCCGAAGGTGAGCCAGTGTGACTCTAGTAGAGAGCGTTCACCGACAAA
BBa_B0034-AthPAL1-PF	ATCTAAAGAGGAGAAATACATATGGAGATTAACGGGGGCACACAA
BBa_B0015-AthPAL1-PR	
BBa_B0034-Rg/PAL/TAL-PF	
BBa_B0015- <i>Rg</i> /PAL/TAL-PR	<u>TTATITGATGCCTGGTTACTCGAG</u> TTACGCCAGCATTTTCAGCAGA
NADPH regulation	
BBa_R0051-MicC-PF	
BBa B0015-MicC-PR	ΓΩΤΤΤΑΤΤΤΩΑΤΩΓΟΤΩΩΟΤΩΔΩΑΔΑΔΑΔΑΔΟΩΟΟΩΑΟΩΑΟΤΩΤΤΟ
BBa B0015-PE	CCCACCCATCAAATAAAACCeeooAcoacoacoacoacoacoacoacoacoacoacoacoacoac
BB2 B0015-PR	
bba_b0013-1 K	
pCL-BBa_R0051-PF	GACTATITTACCTCTGGCGGTGAT
pCL-BBa_B0015-PR	ACGTTGTAAAACGACGGCCAGTGGTACCTATAAACGCAGAAAGGCCCAC
anti( <i>sthA</i> )-PF	CCTCTGGCGGTGATAATGGTTGCATCGTAATCGTAGGAATGTGGCAT
anti(sthA)-PR	ATGCAATGGCCCAACAGAAAATGCCACATTCCTACGATTACGAT
	GGGCCTTTCTGCGTTTATAGGTACCAAATTAATACGACTCACTATAG
pCL-17-pntA-PF	GGGAATCTAGAGGGAATATCATGCGAATTGGCATACC
CL DD- D100((D DD	TTGTAAAACGACGGCCAGTGGATCCAAAAAAAACCCCCGCCCTGTCA
pCL-DDa_D1006-pntB-PK	GGGGCGGGGTTTTTTTTTAAGCTTGGGTTACAGAGCTTTCAGGATTGC

Table A1. Oligonucleotides used in this study. <sup>1</sup>

<sup>1</sup> The sequences with single underline and double underline are the homologous arms for one-step cloning and the flexible linker for the protein fusion, respectively.

# References

- 1. Kaneko, T.; Thi, T.H.; Shi, D.J.; Akashi, M. Environmentally degradable, high-performance thermoplastics from phenolic phytomonomers. *Nat Mater.* **2006**, *5*, 966–970. [CrossRef] [PubMed]
- 2. Jaganathan, S.K.; Supriyanto, E.; Mandal, M. Events associated with apoptotic effect of *p*-coumaric acid in HCT-15 colon cancer cells. *World J. Gastroenterol.* **2013**, *19*, 7726–7734. [CrossRef] [PubMed]
- 3. Scheepens, A.; Bisson, J.-F.; Skinner, M. *p*-Coumaric acid activates the GABA-A receptor in vitro and is orally anxiolytic in vivo. *Phytother Res.* **2014**, *28*, 207–211. [CrossRef] [PubMed]

- 4. Navaneethan, D.; Rasool, M. *p*-Coumaric acid, a common dietary polyphenol, protects cadmium chloride-induced nephrotoxicity in rats. *Ren. Fail.* **2014**, *36*, 244–251. [CrossRef] [PubMed]
- Yoon, H.S.; Lee, N.H.; Hyun, C.G.; Shin, D.B. Differential effects of methoxylated *p*-coumaric acids on melanoma in B16/F10 cells. *Prev. Nutr. Food Sci.* 2015, 20, 73–77. [CrossRef] [PubMed]
- 6. Sakamula, R.; Thong-Asa, W. Neuroprotective effect of *p*-coumaric acid in mice with cerebral ischemia reperfusion injuries. *Metab. Brain Dis.* **2018**, *33*, 765–773. [CrossRef] [PubMed]
- 7. Huang, Q.; Lin, Y.; Yan, Y. Caffeic acid production enhancement by engineering a phenylalanine over-producing *Escherichia coli* strain. *Biotechnol. Bioeng.* **2013**, *110*, 3188–3196. [CrossRef] [PubMed]
- Kawaguchi, H.; Katsuyama, Y.; Danyao, D.; Kahar, P.; Nakamura-Tsuruta, S.; Teramura, H.; Wakai, K.; Yoshihara, K.; Minami, H.; Ogino, C.; et al. Caffeic acid production by simultaneous saccharification and fermentation of kraft pulp using recombinant *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 2017, 101, 5279–5290. [CrossRef]
- 9. Fowler, Z.L.; Koffas, M.A. Biosynthesis and biotechnological production of flavanones: Current state and perspectives. *Appl. Microbiol. Biotechnol.* **2009**, *83*, 799–808. [CrossRef]
- Zhao, S.; Jones, J.A.; Lachance, D.M.; Bhan, N.; Khalidi, O.; Venkataraman, S.; Wang, Z.; Koffas, M.A. Improvement of catechin production in *Escherichia coli* through combinatorial metabolic engineering. *Metab. Eng.* 2015, *28*, 43–53. [CrossRef]
- Li, M.; Kildegaard, K.R.; Chen, Y.; Rodriguez, A.; Borodina, I.; Nielsen, J. *De novo* production of resveratrol from glucose or ethanol by engineered *Saccharomyces cerevisiae*. *Metab. Eng.* 2015, 32, 1–11. [CrossRef] [PubMed]
- 12. Zha, J.; Koffas, M.A.G. Production of anthocyanins in metabolically engineered microorganisms: Current status and perspectives. *Synth. Syst. Biotechnol.* **2017**, *2*, 259–266. [CrossRef] [PubMed]
- Vanholme, R.; Storme, V.; Vanholme, B.; Sundin, L.; Christensen, J.H.; Goeminne, G.; Halpin, C.; Rohde, A.; Morreel, K.; Boerjan, W. A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis. Plant Cell* 2012, 24, 3506–3529. [CrossRef] [PubMed]
- Achnine, L.; Blancaflor, E.B.; Rasmussen, S.; Dixon, R.A. Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* 2004, *16*, 3098–3109. [CrossRef] [PubMed]
- 15. Ehlting, J.; Hamberger, B.; Million-Rousseau, R.; Werck-Reichhart, D. Cytochromes P450 in phenolic metabolism. *Phytochem. Rev.* **2006**, *5*, 239–270. [CrossRef]
- Vannelli, T.; Xue, Z.; Breinig, S.; Qi, W.W.; Sariaslani, F.S. Functional expression in *Escherichia coli* of the tyrosine-inducible tyrosine ammonia-lyase enzyme from yeast *Trichosporon cutaneum* for production of *p*-hydroxycinnamic acid. *Enzyme Microb. Technol.* 2007, 41, 413–422. [CrossRef]
- Jendresen, C.B.; Stahlhut, S.G.; Li, M.; Gaspar, P.; Siedler, S.; Förster, J.; Maury, J.; Borodina, I.; Nielsen, A.T. Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and *Saccharomyces cerevisiae*. *Appl. Environm. Microbiol.* 2015, *81*, 4458–4476. [CrossRef]
- 18. Vannelli, T.; Qi, W.W.; Sweigard, J.; Gatenby, A.A.; Sariaslani, F.S. Production of *p*-hydroxycinnamic acid from glucose in *Saccharomyces cerevisiae* and *Escherichia coli* by expression of heterologous genes from plants and fungi. *Metab. Eng.* **2007**, *9*, 142–151. [CrossRef]
- Vargas-Tah, A.; Martínez, L.M.; Hernández-Chávez, G.; Rocha, M.; Martínez, A.; Bolívar, F.; Gosset, G. Production of cinnamic and *p*-hydroxycinnamic acid from sugar mixtures with engineered *Escherichia coli*. *Microb. Cell Fact.* 2015, 14, 6. [CrossRef]
- 20. Trotman, R.J.; Camp, C.E.; Ben-Bassat, A.; DiCosimo, R.; Huang, L.; Crum, G.A.; Sariaslani, F.S.; Haynie, S.L. Calcium alginate bead immobilization of cells containing tyrosine ammonia lyase activity for use in the production of *p*-hydroxycinnamic acid. *Biotechnol. Prog.* **2007**, *23*, 638–644. [CrossRef]
- 21. Rodriguez, A.; Kildegaard, K.R.; Li, M.; Borodina, I.; Nielsen, J. Establishment of a yeast platform strain for production of *p*-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. *Metab. Eng.* **2015**, *31*, 181–188. [CrossRef]
- 22. Kawai, Y.; Noda, S.; Ogino, C.; Takeshima, Y.; Okai, N.; Tanaka, T.; Kondo, A. *p*-Hydroxycinnamic acid production directly from cellulose using endoglucanase- and tyrosine ammonia lyase-expressing *Streptomyces lividans*. *Microb. Cell Fact.* **2013**, *12*, 45. [CrossRef] [PubMed]

- Nijkamp, K.; Westerhof, R.G.M.; Ballerstedt, H.; Bont, J.A.M.D.; Wery, J. Optimization of the solvent-tolerant *Pseudomonas putida* S12 as host for the production of *p*-coumarate from glucose. *Appl. Microb. Biotechnol.* 2007, 74, 617. [CrossRef] [PubMed]
- 24. Leonard, E.; Yan, Y.; Koffas, M.A.G. Functional expression of a P450 flavonoid hydroxylase for the biosynthesis of plant-specific hydroxylated flavonols in *Escherichia coli*. *Metab. Eng.* **2006**, *8*, 172–181. [CrossRef] [PubMed]
- 25. Leonard, E.; Koffas, M.A. Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Appl. Environm. Microb.* **2007**, *73*, 7246–7251. [CrossRef]
- 26. Wang, R.; Xu, S.; Jiang, Y.; Jiang, J.; Li, X.; Liang, L.; He, J.; Peng, F.; Xia, B. *De novo* sequence assembly and characterization of *Lycoris aurea* transcriptome using GS FLX Titanium platform of 454 pyrosequencing. *PLoS ONE* **2013**, *8*, e60449. [CrossRef] [PubMed]
- 27. Jiang, Y.; Xu, S.; Han, X.; Wang, R.; He, J.; Xia, B.; Wang, R. Investigation of nuclear DNA contents of *Lycoris* species (Amaryllidaceae) with different chromosome number by flow cytometry. *Pak. J. Botany.* **2017**, *49*, 2197–2200.
- Kilgore, M.B.; Augustin, M.M.; May, G.D.; Crow, J.A.; Kutchan, T.M. CYP96T1 of *Narcissus sp. aff. pseudonarcissus* catalyzes formation of the *para-para*' C–C phenol couple in the Amaryllidaceae alkaloids. *Front Plant Sci.* 2016, 7, 225. [CrossRef]
- 29. Kilgore, M.B.; Augustin, M.M.; Starks, C.M.; O'Neil-Johnson, M.; May, G.D.; Crow, J.A.; Kutchan, T.M. Cloning and characterization of a norbelladine 4'-O-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in *Narcissus sp. aff. pseudonarcissus*. *PLoS ONE* **2014**, *9*, e103223. [CrossRef]
- Bendtsen, J.D.; Nielsen, H.; von Heijne, G.; Brunak, S. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 2004, 340, 783–795. [CrossRef] [PubMed]
- 31. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [CrossRef] [PubMed]
- 32. Nelson, B.K.; Cai, X.; Nebenführ, A. A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J.* **2007**, *51*, 1126–1136. [CrossRef] [PubMed]
- Urban, P.; Mignotte, C.; Kazmaier, M.; Delorme, F.; Pompon, D. Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. J. Biol. Chem. 1997, 272, 19176–19186. [CrossRef] [PubMed]
- Ro, D.K.; Ehlting, J.; Douglas, C.J. Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar. *Plant Physiol.* 2002, *130*, 1837–1851. [CrossRef] [PubMed]
- 35. Yamazaki, S.; Sato, K.; Suhara, K.; Sakaguchi, M.; Mihara, K.; Omura, T. Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. *J. Biochem.* **1993**, *114*, 652–657. [CrossRef] [PubMed]
- 36. Na, D.; Yoo, S.M.; Chung, H.; Park, H.; Park, J.H.; Lee, S.Y. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat. Biotechnol.* **2013**, *31*, 170–174. [CrossRef] [PubMed]
- 37. Sauer, U.; Canonaco, F.; Heri, S.; Perrenoud, A.; Fischer, E. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in nadph metabolism of *Escherichia coli*. *J. Biol. Chem.* **2004**, *279*, 6613. [CrossRef]
- 38. Ma, R.; Xu, S.; Zhao, Y.; Xia, B.; Wang, R. Selection and validation of appropriate reference genes for quantitative real-time PCR analysis of gene expression in *Lycoris aurea*. *Front Plant Sci.* **2016**, *7*, 536. [CrossRef]
- 39. Hall, T.A. Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
- 40. Chen, S.; Tao, L.; Zeng, L.; Vega-Sanchez, M.E.; Umemura, K.; Wang, G.L. A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. *Mol. Plant Pathol.* **2006**, *7*, 417–427. [CrossRef]
- 41. Fasan, R.; Crook, N.C.; Peters, M.W.; Meinhold, P.; Buelter, T.; Landwehr, M.; Cirino, P.C.; Arnold, F.H. Improved product-per-glucose yields in p450-dependent propane biotransformations using engineered *Escherichia coli. Biotechnol. Bioeng.* **2011**, *108*, 500–510. [CrossRef] [PubMed]
- 42. Chang, A.C.; Cohen, S.N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J. Bacteriol.* **1978**, *134*, 1141–1156. [PubMed]
- 43. Rose, R.E. The nucleotide sequence of pACYC184. Nucleic Acids Res. 1988, 16, 355. [CrossRef]

- 44. Lerner, C.G.; Inouye, M. Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability. *Nucleic Acids Res.* **1990**, *18*, 4631. [CrossRef] [PubMed]
- 45. Li, Y.; Li, M.; Zhang, X.; Yang, P.; Liang, Q.; Qi, Q. A novel whole-phase succinate fermentation strategy with high volumetric productivity in engineered *Escherichia coli*. *Biores. Technol.* **2013**, *149*, 333–340. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds are not available from the authors.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).