Supplementary Information (SI) for Green Chemistry.

# Supplementary Materials for

# Discovery and green metabolic engineering of a self-sufficient genistein pathway in *Paenibacillus jilinensis*

Jia Bao Zhang *et al.*

CORRESPONDING AUTHOR: Yong Jun Yang, youngjune@jlu.edu.cn; Zhen Zhen Liu, zhenzhenliu@jlu.edu.cn

**THE PDF FILE INCLUDES:**

FIGURES S1 TO S5

TABLES S1 TO S5

**OTHER SUPPLEMENTARY MATERIAL FOR THIS MANUSCRIPT INCLUDES THE FOLLOWING:**

DATE FILES S1 to S3

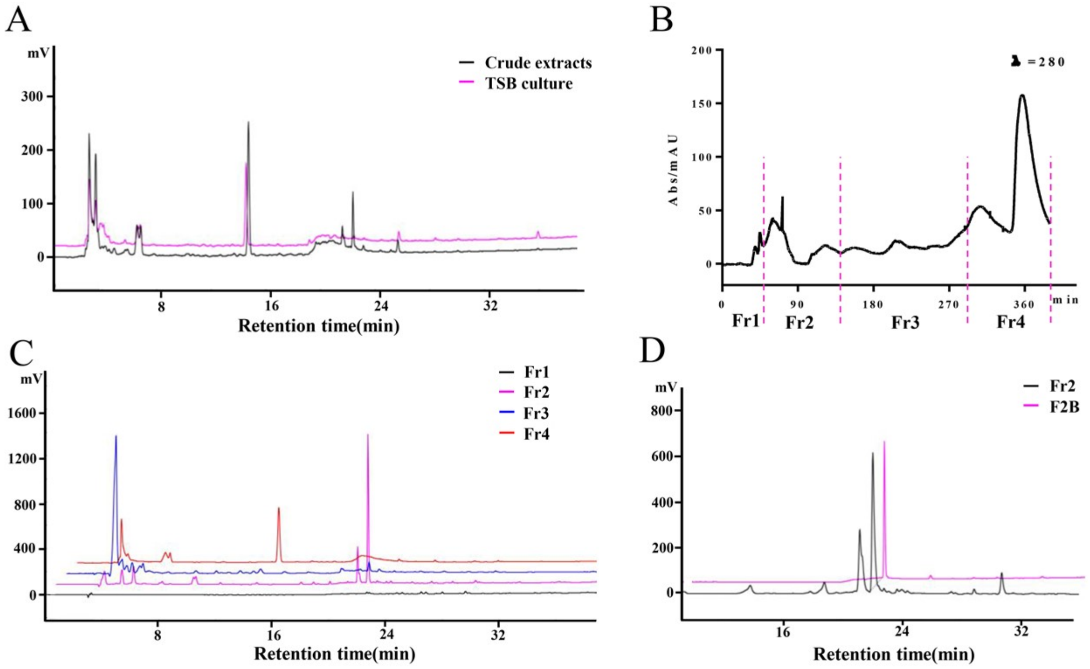

**Fig. S1. Purification of F2B from *P. jilinensis* YPG26.** (A) Comparison of analytical HPLC of crude extracts of *P. jilinensis* YPG26 and TSB culture medium. Both samples were obtained by acidification precipitation and methanol extraction. (B) Fr1-Fr4 components were separated by silica gel column chromatography. The filling material was 200-300 mesh silica gel. The mobile phase consisted of dichloromethane and methanol, and the elution ratios were 10:0 (Fr1), 8:2 (Fr2), 6:4 (Fr3), and 0:10 (Fr4), respectively. (C) Analytical HPLC chromatograms of Fr1-Fr4. (D) Sample preparation for F2B from Fr2 component by Semi-preparative HPLC chromatogram.

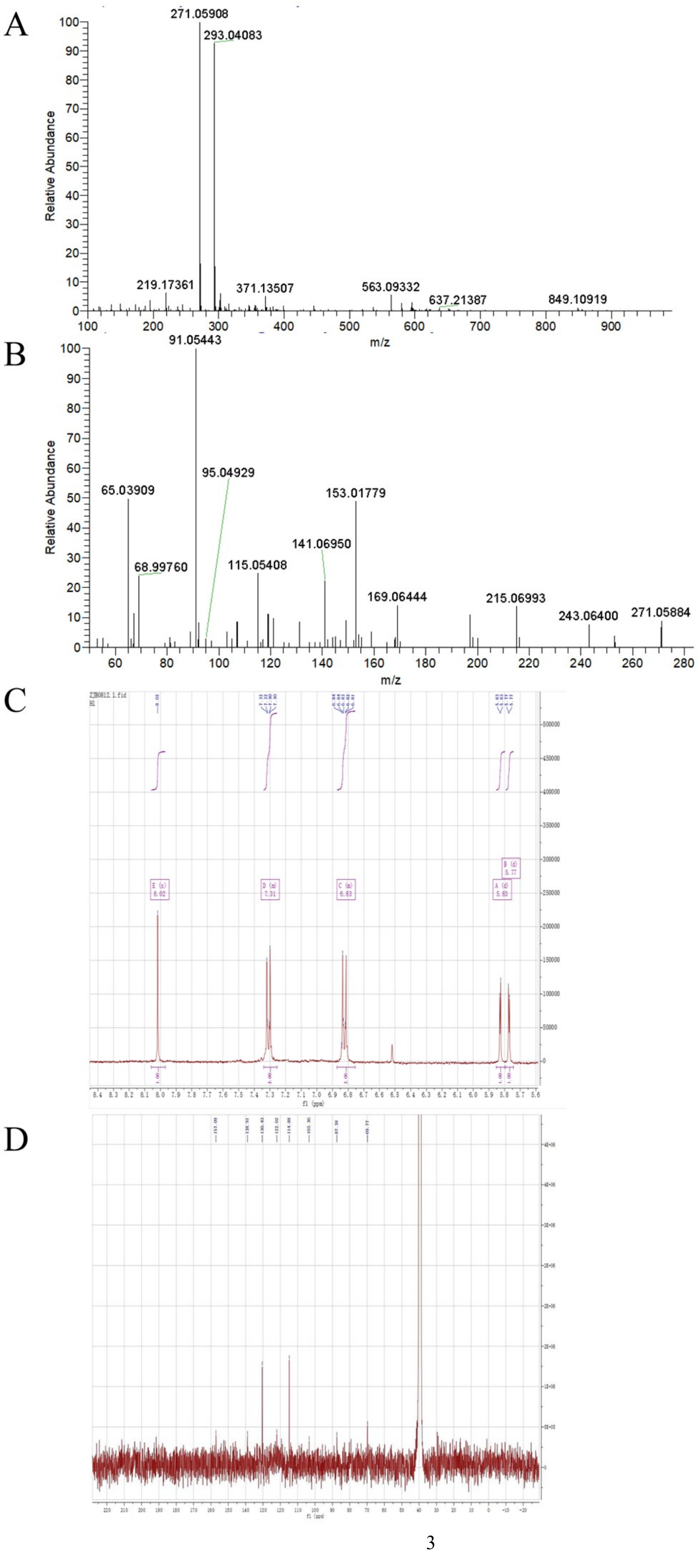 
A
Relative Abundance
271.05908
293.04083
219.17361
371.13507
563.09332
637.21387
849.10919
m/z
B
Relative Abundance
91.05443
95.04929
65.03909
68.99760
115.05408
141.06950
153.01779
169.06444
215.06993
243.06400
271.05884
m/z
C
ZJB0812.1.fid
H1
E (s) 8.02
D (m) 7.31
C (m) 6.83
A (d) 5.83
B (d) 5.77
1.00
2.00
2.06
f1 (ppm)
D
157.09
138.91
130.43
122.02
114.88
103.36
87.38
69.77
f1 (ppm)

**Fig. S2. Structural identification of F2B.** Compound F2B was isolated and purified as described in the method. (A)The first order ESI (+) Orbitrap FTMS mass spectrum of F2B. (B) The secondary order ESI (+) Orbitrap FTMS mass spectrum of F2B. (C) 13C NMR spectra of F2B. The chemical shifts are reported in ppm, and the coupling constant J is measured in Hertz. S represents a single peak, D represents double peaks, and M represents multiple peaks. (D) 1H NMR spectra of F2B.

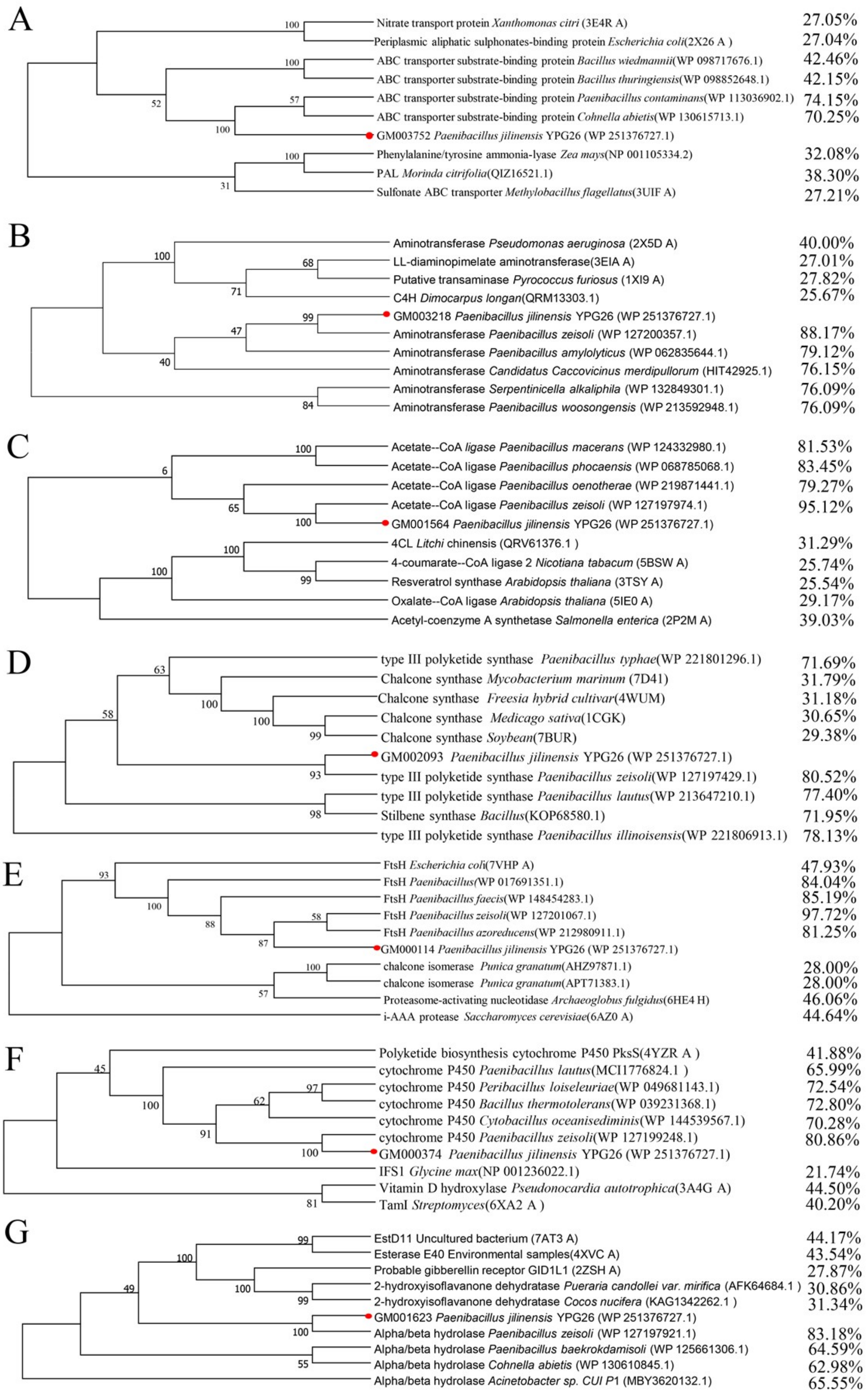

**Fig. S3. Phylogenetic trees of genistein biosynthesis-related enzymes of *P. jilinensis* YPG26.** (A) *Pj*PAL. (B) *Pj*C4H. (C) *Pj*4CL. (D) *Pj*CHS. (E) *Pj*CHI. (F) *Pj*IFS. (G) *Pj*HID. The percentage of protein sequence similarity is displayed to the right of the species name in the phylogenetic tree.

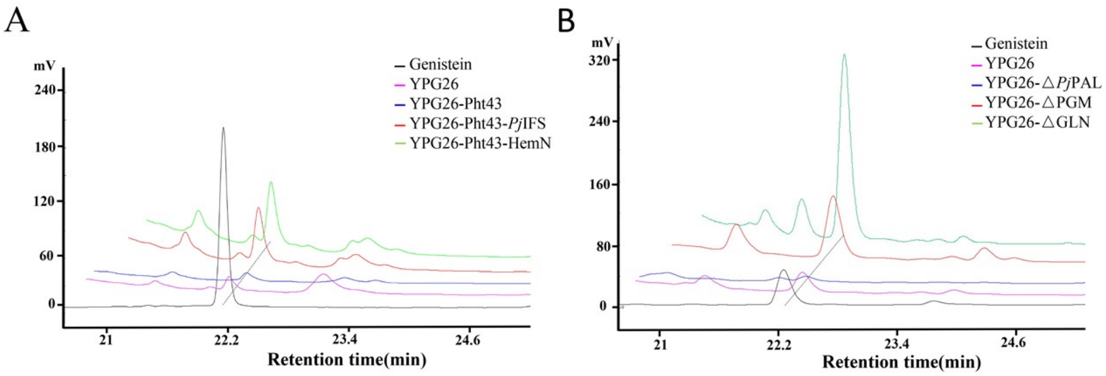

**Fig. S4. HPLC diagram of genistein content detection of parent strain and engineering strains.** (A) Detection of genistein content in gene overexpressing strains and parental strains by HPLC. (B) The content of genistein in gene knockout strains and parental strains was detected by HPLC.

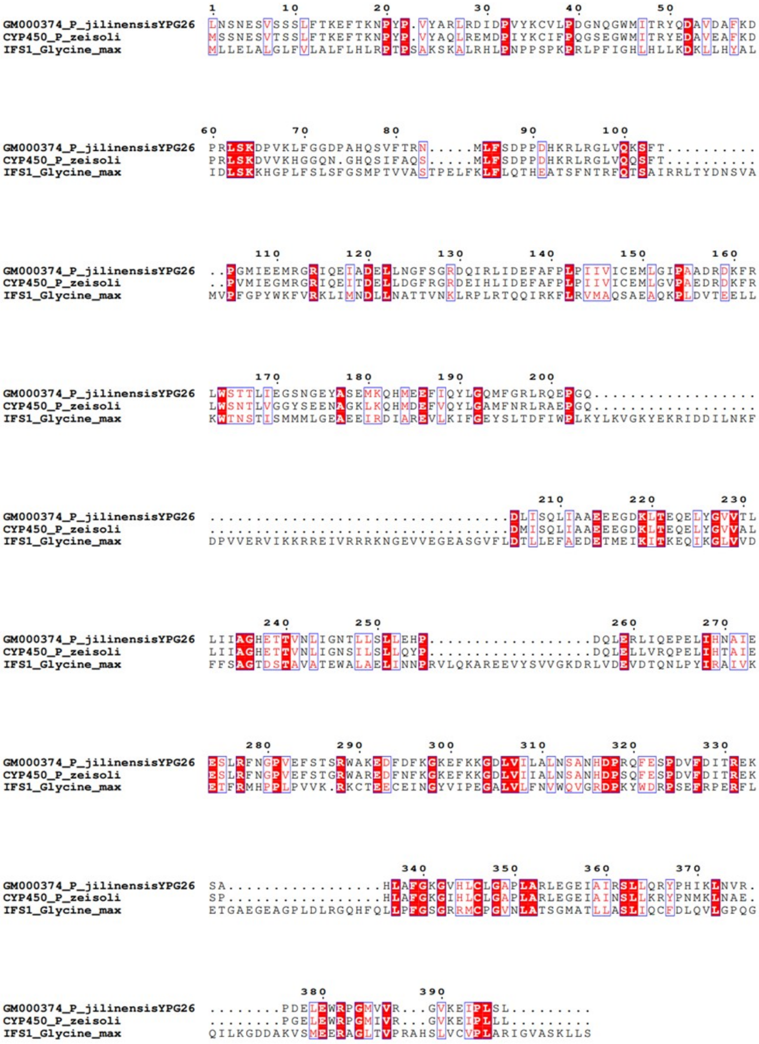

**Fig. S5. Multiple alignment of *Pj*IFS (*Paenibacillus jilinensis*YPG26, GM000374), cytochrome P450 (*Paenibacillus zeisoli*, WP_127199248.1), and IFS1 (*Glycine max*, NP_001236022.1).**

**Table. S1 Strains and plasmids used in this study.**

| **Strains or plasmids** | **Description** | **Source** |
|---|---|---|
| Strains | | |
| *Paenibacillus jilinensis* YPG26 | Wild type | This lab |
| *E. coli* DH5α | Host strain for plasmid construction | This lab |
| *E. coli* BL21 | Host strain for recombinant expression | This lab |
| *Pj*C4H-BL21 | pET28a carrying *Pj*C4H gene | This study |
| *Pj*PAL- BL21 | pET28a carrying *Pj*PAL gene | This study |
| *Pj*C4H-*Pj*PAL-BL21 | pET28a carrying *Pj*C4H and *Pj*PAL genes | This study |
| *Pj*CHS-BL21 | pET28a carrying *Pj*CHS gene | This study |
| *Pj*CHI-BL21 | pET28a carrying *Pj*CHI gene | This study |
| *Pj*4CL-BL21 | pET28a carrying *Pj*4CL gene | This study |
| *Pj*4CL-*Pj*CHI-*Pj*CHS-BL21 | pET28a carrying *Pj*4CL, *Pj*CHI, and *Pj*CHS genes | This study |
| *Pj*IFS-BL21 | pET28a carrying *Pj*IFS gene | This study |
| *Pj*HID-BL21 | pET28a carrying *Pj*HID gene | This study |
| *Pj*IFS-*Pj*HID-BL21 | pET28a carrying *Pj*IFS and 2HID genes | This study |
| YPG26-PHT43-*Pj*IFS | YPG26 derivate, overexpression of *Pj*IFS gene | This study |
| Y26-PHT43-HemN | YPG26 derivate, overexpression of HemN gene | This study |
| Y26-PHT43 | YPG26 derivate, PHT43 empty vector | This study |
| JM110 | Plasmid demethylation *Escherichia coli* | This lab |
| Plasmids | | |
| pET28a | Kan$^r$, recombinant expression vector | This lab |
| pMD18-T | Amp$^r$, TA Cloning | This lab |
| pHT43 | Amp$^r$, Cm$^r$, high expression vector | This lab |
| pJOE8999 | Kan$^r$, gene knockout vector | This lab |

**Table. S2 Sequences of the primers used in this study.**

| Primers | Sequences (5’-3’) | Notes |
|---|---|---|
| RT-PAL-F | GAGTACATTAATACCCCGGTTAC | qRT - PCR |
| RT-PAL-R | ACTGATATCAACCGCTTCTC | qRT - PCR |
| RT-C4H-F | CATATTATTGATGCGCTGTGCAC | qRT - PCR |
| RT-C4H-R | CGGGTCGAGCTCTACTCC | qRT - PCR |
| RT-4CL-F | AAGGCGGAGATTTCCAAGTTTG | qRT - PCR |
| RT-4CL-R | TGCTTTAAGTACCCGCCGC | qRT - PCR |
| RT-CHS-F | ATTACCCACCTTATTACTGTCAGC | qRT - PCR |
| RT-CHS-R | TGCGGCGCATCCTTGGAATATG | qRT - PCR |
| RT-CHI-F | GGACGTGTAGCTGAGGAGC | qRT - PCR |
| RT-CHI-R | ACCCAGCTTCTCACTCATACC | qRT - PCR |
| RT-IFS-F | AAGCTGACTGAGCAGGAGC | qRT - PCR |
| RT-IFS-R | CAAGCGCTCAAGCTGATCC | qRT - PCR |
| RT-CPR-F | CAGGATCTGATCAGTCAGCTTATTG | qRT - PCR |
| RT-CPR-R | GATCAGATTCACCGTCGTCTCATG | qRT - PCR |
| RT-HID-F | TCCAGCGGGCTTAGAGTACTG | qRT - PCR |
| RT-HID-R | GATCTGACCGGCGGAATAACC | qRT - PCR |
| RT-16S-F | ACTCCTACGGGAGGCAGC | qRT - PCR |
| RT-16S-R | GTAGTTAGCCGGGGCTTTCTTC | qRT - PCR |
| C4H-BamHI-F | CTGGGATCCATGAAGTTTGCTACAAG | Heterologous expression |
| C4H-SalI-R | CTGGTCGACCTATAATTCAACCTTCAG | Heterologous expression |
| PAL -HindIII-F | CTGAAGCTTATGGCTATGATTACTGCCCTG | Heterologous expression |
| PAL -XhoI-R | CTGCTCGAGTTATTTTAGGGCTTTCTC | Heterologous expression |
| CHI-SalI-F | CTGGTCGACATGAATCGGTTCATCCGGAATTCTG | Heterologous expression |
| CHI- XhoI -R | CTGCTCGAGTTACAGCGGCGTGCTGCCAC | Heterologous expression |
| CHS-SacI-F | CTGGAGCTCATGACGTCAACGATCAG | Heterologous expression |
| CHS-Sal I-R | CTGGTCGACCTAAAGAAGGGCATGATC | Heterologous expression |
| 4CL-BamhI-F | CTGGGATCCATGAACCAACTGCAAGG | Heterologous expression |
| 4CL-SalI-R | CTGGTCGACTTAGTCTTCCTCAATTG | Heterologous expression |
| 4CL-NotI-R | ATTTGCGGCCGCTTAGTCTTCCTCAATTG | Heterologous expression |
| 2HID-SacI-F | CTGGAGCTCATGACGGCAAGAATATC | Heterologous |

| | | |
|---|---|---|
| | C | expression |
| 2HID-SalI-R | CTGGTCGACTTATCGATTACAATCTAATATC | Heterologous expression |
| IFS-BamhI-F | CTGGGATCCTTGAATTCCAATGAATC | Heterologous expression |
| IFS-SacI-R | CTGGAGCTCCTATAATGACAGCGGAATTTC | Heterologous expression |
| T7-NotI-F | ATTTGCGGCCGCTAATACGACTCACTATAG | Heterologous expression |
| RBS -NotI-F | ATTTGCGGCCGCAATAATTTTGTTTAAC | Heterologous expression |
| T7-SalI-F | CTGGTCGACTAATACGACTCACTATAG | Heterologous expression |
| XbaI-HemN-F | CTGTCTAGAATGGAATATTTAAAAGCGCT | Gene overexpression |
| NotI-HemN-R | ATTTGCGGCCGCCTAGCCGTTATTAGCGGT | Gene overexpression |
| XbaI-IFS-F | CTGTCTAGATTGAATTCCAATGAATCCGT | Gene overexpression |
| NotI-IFS-R | ATTTGCGGCCGCCTATAATGACAGCGGAAT | Gene overexpression |
| pHT43-CX-F | ATTCAAAAACGAAAGCGGAC | Gene overexpression |
| pHT43-CX-R | CCATTTGTTCCAGGTAAGGTAT | Gene overexpression |
| L-gln-sgRNA-F | TACGACTGTGATATTGACTGCTAC | Gene knockout |
| L-gln-sgRNA-R | AAACGTAGCAGTCAATATCACAGT | Gene knockout |
| gln-L-F | cgactcactatagggtcgacGTACGAGGAGTTAGGGCTTCCG | Gene knockout |
| gln-L-R | tccgccgtcggttccGTTGGTTACCCCTCCAATTC | Gene knockout |
| gln-R-F | ggaggggtaaccaacGGAACCGACGGCGGAGCAAATC | Gene knockout |
| gln-R-R | agattatttcttaatctagaGCACAGATAACACCCCAGCGGTC | Gene knockout |
| gln-test-F | CGAGCTGAACACGGACAATC | Gene knockout |
| gln-test-R | CCCGGTCCCATCAGAACACG | Gene knockout |
| PGM -sgRNA-F | TACGAAGATGCAGGAGGCATCCAG | Gene knockout |
| PGM -sgRNA-R | AAACCTGGATGCCTCCTGCATCTT | Gene knockout |
| PGM -L-F | cgactcactatagggtcgacCTATGCTTCAGAAGGTGTGAAC | Gene knockout |
| PGM -L-R | aatttctacttataaGCTAACGGCTCCTCTCACCTTG | Gene knockout |
| PGM -R-F | agaggagccgttagcTTATAAGTAGAAATTT | Gene knockout |

| | | |
|---|---|---|
| | AGAAT | |
| PGM -R-R | agattatttcttaatctagaAGTCGGGAATCTCACCTCGG | Gene knockout |
| PGM -test-F | GCGCGTACCGTGGGCATCGTG | Gene knockout |
| PGM -test-R | CGGCAGCAGAATCGTCTCGGC | Gene knockout |
| PAL-sgRNA-F | TACGGGGTATTAATGTACTCCGTA | Gene knockout |
| PAL-sgRNA-R | AAACTACGGAGTACATTAATACCC | Gene knockout |
| PAL-L-F | cgactcactatagggtcgacCATAACACCGGCGGCGATCATC | Gene knockout |
| PAL-L-R | atcaccgtgctattaATGCAACTGAGAGGGTGAGAC | Gene knockout |
| PAL-R-F | ccctctcagttgcatTAATAGCACGGTGATTTTGG | Gene knockout |
| PAL-R-R | agattatttcttaatctagaTAGATTAATAGACGGGCTTATC | Gene knockout |
| PAL-test-F | CAGATGACCCACAATCTCCTG | Gene knockout |
| PAL-test-R | CCAGCTCCTGCTCCAGCTTGG | Gene knockout |
| pjoe8999-Spacer-F | AACCATCACTGTACCTCCCA | Gene knockout |
| pjoe8999-Spacer-R | GAGCGTTCTGAACAAATCCA | Gene knockout |
| Donor-test-R | GGAGGTGACTGAAGTATATT | Gene knockout |
| Donor-test-F | GCAAGTTAAAATAAGGCTAGTC | Gene knockout |

**Table. S3 The protein sequences of genistein-related synthase in *P. jilinensis* YPG26.**

| Name | Protein sequence |
|---|---|
| *Pj*PAL | >YPG26_GM003752 locus=Chr1:3841888:3842865:-<br>MAMITALILSACNSTGSTSNGASNSSSSKKEATVNIAINGGLNLLTIAKQKGWFEEEFSKVNAKVQWHEFQSSVPLLEGLVSGRVDFSFIGDGTVVTGKSAKMPFTVISTTGVQGNQNSVIVKPDSPIKSIADLKGKTIAVAKGSSGHIFLIKALQKNNMAEADVKLVDLQPGEGNPAFQTGKVDAWAIWDPFVTTEVQAKRARIVESVDSLGIVAPAVMIGRDEFIKNNPDLTSAYLKVYEKTVKWVNANKDEAAAILAKEKKMELELVKTLVNRTEYINTPVTGEVQAAMQSTADVLLKSGTIREAVDISKVFDNSYIEKALK* |
| *Pj*C4H | >YPG26_GM003218 locus=Chr1:3280037:3281212:-<br>MKFATRMNHFGEGIFTKLLEIKRKRLEQGQPVIDLSVGTPNIPPAPHIIDALCTAAADESNYIYAVNDQSALLEAAGSWYLTRYGVELDPKTEICSLLGSQEGLAHISLSIIDEGDLVLVPDPCYPVFADGPRLAGAELYFMPQRKKNDYLIQLQDIPEDIAARAKFMIVSYPNNPTAALAPDSFYTDLIRFAKKYDIIVLHDNAYSELVFDGRTCGSFLAFPGAMEVGVEFNSLSKTYGLAGARIGFCLGNAEVVARLKTLKSNMDYGMFIPIQQAAIAAITGDQGSVEKTAKAYERRRDILCEGFTRIGWNMEKPAATMFIWTRIPAHYETSENFAMDMVSKAGVIVTPGSAFGPSGEGHVRLALVQDEEMLRHAVQAVEASGILKVEL* |
| *Pj*4CL | >YPG26_GM001564 locus=Chr1:1589220:1590947:+<br>MNQLQGEVISVVAPSSNLGDYESTYDSFNWKDVEKEFSWYETGKLNLAYEAIDRHVDAGRGGRTALLYSDATRDEAFTFEELSGLSNQFANVLRTIGIHKGDRVFVFMPRTPELYVSILGTIKVGAVVGPLFEAFMETAVKDRLEDSGAVAIVTTPALLPRIKRDELPALQHIIVVGDEVEPGEGIIDFKSAMAEASPEAQIEWVDREDGLIIHYTSGSTGKPKGVHHVHNAMIQHYYTGKVVLDLQEGDVYWCTADPGWVTGTSYGIFAPWLNGATNVIRGGRFSPADWYGTIQKYGVTVWYSAPTAFRMLMGAGEDVISQFDLSSLRHVLSVGEPLNPEVVRWGLKVYGKRIHDTWWMTETGAQLICNYPSLPMKPGSMGRPLPGIEAAIIDDSGNILPPYRMGNLAIKTPWPSMMRKIWNNPAKYEEYFRIQGWYISGDSAYMDEDGYFWFQGRIDDVINSSGERIGPFEVESKLVEHPAVAEAGVIGKPDPTRGEIIKAFISLREGFEPSEELKAEISKFVKVGLSAHAAPREIEFKDKLPKTRSGKIMRRVLKAWELQLPVGDLSTIEED* |
| *Pj*CHS | >YPG26_GM002093 locus=Chr1:2117249:2118406:-<br>MTSTISTSQAAILGIGTALPVHSIAQSDIAEFIAHSLQDSPDLARFARRVFRSCGVETRYTCEPAFLGSPEECRYLPEGDPSDIPTTQDRMDTYKREAAPLGLKAAQAAINDAGILPDSITHLITVSCTGQYLPGLDVMLIRQLGLSPRTNRLPLIFQGCAAGLKAIQMARDVVQGAPGSQVLVVCVELCSLHFQPAKEREALFAASFFGDAAAACVIGSPEPHH |

| | |
|---|---|
| | RHYLDLGTGYSVLLPDSTEDMTWEVGDTGYDLYLSPRIPKLLGEHLETELASLLEGDELPELWAIHPGGRGIVDSVEEVMNLTGDQTKYSREILRTVGNVSSVTIIFVLHAMREDMKSLDQARSEGVAMAFGPGLTAELMRFTYMPAQSVAVKEQDHALL |
| *Pj*CHI | >YPG26_GM000114 locus=Chr1:59255:61366:+<br>MNRFIRNSGFYLILFLVVVGIVQFLSGGSEAADTPRYDQLRQQLQANNVKELTGQFDGYAYRVTGKYKNKVDNKESFQTYVPYDTNVVAELTDYSEKNKISYSWDKMEGDSIWLTLLTSIVPLAIMFILFFFLFNQAQGGGGKVMNFGKSRARLYNEEKKKVTFEDVAGADEEKQELVEVVEFLKDPRKFAAVGARIPKGVLLVGPPGTGKTLLARAVAGEAGVPFFSISGSDFVEMFVGVGASRVRDLFENAKKNAPCIIFIDEIDAVGRQRGAGLGGGHDEREQTLNQLLVEMDGFGGNEGIIIVAATNRADILDPALLRPGRFDRQITVDRPDVKGREAVLKVHARNKPLTKDVRMDVIAKRTTGFTGADLENLLNEAALLAARRNRKDISMREVDEAIDRVIVGTEKRSRVVSDREKRIVAFHEAGHTIVGYFLEHADMVHKVTIIPRGRAGGYVIMMPKEDRMLVTKQELLDKVTGLLGGRVAEELFIGEIGTGAYSDFQQATSIVRSMIVEYGMSEKLGPMQFGTSQGQVFLGRDIGHEQNYSDAIAYEIDQEMQRFISESYERCKQLLMKHSKEVHLIAETLLEVETLELDQIKQLIETGKLTPDPDGDGSGSSESGAPIVDTIGDVNVRIQARDEEAQPQGPSGDIPNDVPGSSANDIRGDHPVDTPEGTVKDAPTTPPQPKDQDGSGNGGSTPL* |
| *Pj*IFS | >YPG26_GM000374 locus=Chr1:337716:338909:+<br>LNSNESVSSSLFTKEFTKNPYPVYARLRDIDPVYKCVLPDGNQGWMITRYQDAVDAFKDPRLSKDPVKLFGGDPAHQSVFTRNMLFSDPPDHKRLRGLVQKSFTPGMIEEMRGRIQEIADELLNGFSGRDQIRLIDEFAFPLPIIVICEMLGIPAADRDKFRLWSTTLIEGSNGEYASEMKQHMEEFIQYLGQMFGRLRQEPGQDLISQLIAAEEEGDKLTEQELYGVVTLLIIAGHETTVNLIGNTLLSLLEHPDQLERLIQEPELIHNAIEESLRFNGPVEFSTSRWAKEDFDFKGKEFKKGDLVILALNSANHDPRQFESPDVFDITREKSAHLAFGKGVHLCLGAPLARLEGEIAIRSLLQRYPHIKLNVRPDELEWRPGMVVRGVKEIPLSL* |
| *Pj*HID | >YPG26_GM001623 locus=Chr1:1658464:1659108:+<br>MTARISYSSGLRVLTFEYRKAPEDPYPAANEDALTAYYWLKEAGYSAGQIIFGGDSVGATLALMTLITLRDREEALPAGAFLISPHADLVHLDGDSYVTNRENDPTGSLEGNRRLIAAYLGSWQGEPPEILSPLKMNLAALPPLLIQVGSLEVLLSDAVRLSEWAKQAGVDVTMEIWDNMWSVFHFLAYMLPEAEQAIRNIGAFNKAKILDCNR* |

**Table. S4 Growth simulation of *P. jilinensis* on a single carbon or nitrogen source.**

| Substrate | In *silico*[a] | In *vivo*[a] | Reference |
|---|---|---|---|
| Carbon source | | | |
| Fructose | + | + | This study |
| Glucose | + | + | This study |
| Glycerol | + | + | This study |
| Lactose | + | + | This study |
| Maltose | + | + | This study |
| Mannose | + | + | This study |
| Sorbitol | + | + | This study |
| Starch | + | + | This study |
| Sucrose | + | + | This study |
| Nitrogen Source | | | |
| $NH_4^+$ | + | + | This study |
| Urea | + | + | This study |
| Potassium nitrate | + | + | This study |

[a] (+ for growth/− for non-growth).

**Table. S5 Comparison between models YPG26, iJYQ746, and iYO844.**

| **Characteristic** | **YPG26** | **iYO844** | **iJYQ746** |
|---|---|---|---|
| Characteristics of genome | | | |
| Genome size (Mb) | 3.97 | 4.22 | 4.06 |
| Total genes | 3871 | 4536 | 4048 |
| GC% | 49.3 | 43.5 | 46.4 |
| Model of metabolism | | | |
| Genes | 716 | 844 | 746 |
| Reactions | 1637 | 1250 | 1736 |
| Metabolites | 1512 | 990 | 1611 |
| Gene coverage rate% | 18.5 | 18.6 | 18.4 |
| Biochemical reactions | 1387 | 769 | 1463 |
| Exchange reactions | 119 | 229 | 134 |
| Transport reactions | 131 | 251 | 137 |

**S1 Data. GSMM YPG26 in xls.** This document includes reaction list and metabolite list. **(Separate file)**
**S2 Data. Comparison results with DEG database. (Separate file)**
**S3 Data. Computer simulation of up-regulated and down-regulated targets. (Separate file)**