

Table 3 The gene list annotation using the metagenomic database

| Genome Name | Gene Symbol | Length (bp) | Genome Name | Gene Symbol | Length (bp) |
|---|-------------|-------------|--|--------------|-------------|
| <i>Corynebacterium glutamicum</i> ATCC 21831 | zupT | 792 | <i>Marivirga tractuosa</i> H-43, DSM 4126 | mutS2 | 2397 |
| <i>Bacillus licheniformis</i> DSM 13 Goettingen | yyaS | 606 | <i>Frankia inefficax</i> Eul1c | murA | 1260 |
| <i>Bacillus licheniformis</i> DSM 13 Goettingen | ywrA | 537 | <i>Brevibacillus laterosporus</i> NRS 682, LMG 15441 | mtrB | 240 |
| <i>Bacillus licheniformis</i> 9945A | ywnC | 393 | <i>Desulfovibrio vulgaris vulgaris</i> DP4 | mtnA | 1053 |
| <i>Paenibacillus polymyxa</i> CICC 10580 | yvbA | 312 | <i>Acidipropionibacterium acidipropionici</i> F3E8 | msrA | 630 |
| <i>Paenibacillus</i> sp. lzh-N1 | yutG1 | 498 | <i>Bacillus paralicheniformis</i> Bac48 | msrA | 546 |
| <i>Bacillus paralicheniformis</i> Bac48 | yunB | 777 | <i>Modestobacter marinus</i> BC501 | mscS | 837 |
| <i>Bacillus paralicheniformis</i> Bac48 | yugT | 1683 | <i>Bacillus paralicheniformis</i> Bac84 | mrgA | 465 |
| <i>Bacillus licheniformis</i> DSM 13 Novozymes | yuel | 402 | <i>Enterobacter cloacae</i> A1137 | mreD | 489 |
| <i>Paenibacillus polymyxa</i> M1 | ytlV3 | 1155 | <i>Thioalkalivibrio paradoxus</i> ARh 1 | mrcB | 2268 |
| <i>Streptomyces</i> sp. RJA2910 | ytnA | 1434 | <i>Bradyrhizobium</i> sp. BM-T | moaE | 468 |
| <i>Bacillus paralicheniformis</i> 14DA11 | yrrS | 693 | <i>Bacillus paralicheniformis</i> Bac84 | moaE | 495 |
| <i>Bacillus paralicheniformis</i> 14DA11 | yrbB | 1143 | <i>Paenibacillus kribbensis</i> AM49 | moaA1 | 1014 |
| <i>Paenibacillus polymyxa</i> J | yqfU | 945 | <i>Thermoanaerobacter</i> sp. X514 | mnmE | 1383 |
| <i>Bacillus paralicheniformis</i> MDJK30 | ypzA | 267 | <i>Trichodesmium erythraeum</i> IMS101 | mnmA | 1080 |
| <i>Bacillus paralicheniformis</i> 14DA11 | ylbO | 606 | <i>Xanthomonas albilineans</i> XaFL07-1 | mltD | 1212 |
| <i>Paenibacillus polymyxa</i> SQR-21 | ykkC3 | 342 | <i>Klebsiella pneumoniae pneumoniae</i> RJF293 | mioC | 441 |
| <i>Klebsiella pneumoniae pneumoniae</i> RJF293 | yjhQ | 552 | <i>Priestia megaterium</i> SF185 | minJ | 1191 |
| <i>Paenibacillus polymyxa</i> SC2 | yjbR | 348 | <i>Sphingopyxis</i> sp. MG | mfeA | 903 |
| <i>Bacillus licheniformis</i> 9945A | yhzE | 87 | <i>Nitrosospora multiformis</i> NI14 | mfd | 3468 |
| <i>Paenibacillus polymyxa</i> E681 | yfmM | 1581 | <i>Halobacillus halophilus</i> HL2HP6 | metN2 | 1053 |
| <i>Paenibacillus polymyxa</i> J | yfiF5 | 786 | <i>Corynebacterium terpenotabidum</i> Y-11 | metI | 702 |
| <i>Paenibacillus polymyxa</i> J | yetL3 | 510 | Genome sequencing | | |
| <i>Bacillus licheniformis</i> 5NAP23 | yesE | 420 | <i>Phaeobacter inhibens</i> P88 | metF | 870 |
| <i>Lelliottia nimipressuralis</i> SGAir0187 | yejK | 1008 | <i>Bacillus</i> sp. FJAT-21351 | metB | 1146 |
| <i>Paenibacillus polymyxa</i> Sb3-1 | yclD | 462 | <i>Ligilactobacillus acidipiscis</i> ACA-DC 1533 | menB | 822 |
| <i>Bacillus paralicheniformis</i> BL-09 | ycgN | 1551 | <i>Enterobacter cloacae</i> A1137 | mdoG | 1602 |
| <i>Klebsiella pneumoniae pneumoniae</i> RJF293 | ycgB | 1533 | <i>Amphibacillus xylanus</i> NBRC 15112 | mcsB | 1068 |
| <i>Moorena producens</i> PAL-8-15-08-1 | ycf27 | 729 | <i>Deinococcus gobiensis</i> I-0, DSM 21396 | map | 744 |
| <i>Paenibacillus polymyxa</i> SC2 | ybbK | 462 | <i>Cutibacterium acnes</i> AE1 | map | 855 |
| <i>Caldanaerobacter subterraneus</i> MB4 | XylB | 1077 | <i>Bacillus paralicheniformis</i> 14DA11 | manP | 1944 |
| <i>Geodermatophilus turciae</i> DSM 44513 | xylA | 1185 | <i>Paenibacillus</i> sp. lzh-N1 | M1-957 | 789 |
| <i>Paenibacillus polymyxa</i> Mc5Re-14 | xkdM | 408 | <i>Corynebacterium stationis</i> ATCC 21170 | lysC | 1266 |
| <i>Xanthomonas albilineans</i> GPE PC73 | XCC3509 | 972 | <i>Bacillus</i> sp. IHB B 7164 | lysC | 1233 |
| <i>Caldicellulosiruptor bescii</i> RKC130 | valS | 2625 | <i>Agromyces</i> sp. AR33 | Lxx16240 | 321 |
| <i>Trichormus variabilis</i> NIES-23 | valS | 3045 | <i>Methylobacillus flagellatus</i> KT | lpxA | 783 |
| <i>Chloracidobacterium thermophilum</i> B | valS | 2685 | <i>Xanthomonas albilineans</i> HVO005 | lptC | 567 |
| <i>Candidatus Mycoplasma haemolamae</i> Purdue | valS | 2508 | <i>Vitis vinifera</i> PN40024 | LOC100257077 | 1054 |
| <i>Caldicellulosiruptor bescii</i> RKC121 | valS | 2625 | <i>Vitis vinifera</i> PN40024 | LOC100244859 | 1192 |
| <i>Cupriavidus</i> sp. NP124 | uvrD3 | 2094 | <i>Rhodococcus jostii</i> RHA1 | lipA | 930 |
| <i>Rudanella lutea</i> DSM 19387 | uvrB | 2022 | <i>Anabaenopsis circularis</i> NIES-21 | leuS | 2640 |
| <i>Corynebacterium crudilactis</i> JZ16 | ureG | 618 | <i>Caldanaerobacter subterraneus</i> MB4 | LepA | 1812 |
| <i>Actinoplanes</i> sp. SE50/110 | ureB | 408 | <i>Spiribacter vilamensis</i> DSM 21056 | lepA | 1824 |
| | | | <i>Lactobacillus delbrueckii lactis</i> KCCM 34717 | Ldb2189 | 843 |

Table 3 (continued)

| Genome Name | Gene Symbol | Length (bp) | Genome Name | Gene Symbol | Length (bp) |
|--|-------------|-------------|--|-------------|-------------|
| <i>Stigmatella aurantiaca</i> DW4/3-1 | uppP | 894 | <i>Lactobacillus delbrueckii</i> jakobsenii ZN7a-9 | Ldb1360 | 2235 |
| <i>Chloracidobacterium thermophilum</i> B | uppP | 855 | <i>Lactobacillus delbrueckii</i> lactis KCTC 3035 | Ldb1010 | 1002 |
| <i>Paenibacillus kribbensis</i> AM49 | ugpB5 | 1305 | <i>Lactobacillus delbrueckii</i> jakobsenii ZN7a-9 | Ldb0854 | 1254 |
| <i>Ochrobactrum</i> sp. MYb15 | ubiE | 792 | <i>Lactobacillus delbrueckii</i> lactis KCTC 3035 | Ldb0761 | 234 |
| <i>Rhodopseudomonas palustris</i> TIE-1 | trxC | 438 | <i>Cupriavidus taiwanensis</i> LMG 19424 | kynU | 1257 |
| <i>Rhodococcus</i> sp. MTM3W5.2 | trxB | 990 | <i>Micromonospora aurantiaca</i> ATCC 27029 | kptA | 543 |
| <i>Solidesulfovibrio magneticus</i> RS-1 | truB | 969 | <i>Micromonospora aurantiaca</i> DSM 45487 | kptA | 543 |
| <i>Collimonas arenae</i> Ter282 | trpS2 | 1035 | <i>Paenibacillus polymyxa</i> CICC 10580 | kinA | 1740 |
| <i>Leisingera caerulea</i> DSM 24564 | trpE | 1512 | <i>Laribacter hongkongensis</i> HLHK9 | kcy | 666 |
| <i>Brevibacillus laterosporus</i> B9 | troA | 963 | <i>Phaeobacter inhibens</i> P72 | iscA | 360 |
| <i>Pedobacter ginsengisoli</i> T01R-27 | trmD | 678 | <i>Nostoc</i> sp. Moss6 | invB | 1452 |
| <i>Geobacillus subterraneus</i> KCTC 3922 | topA | 2076 | <i>Ligilactobacillus salivarius</i> salivarius UCC118 | infC | 525 |
| <i>Candidatus Protochlamydia amoebophila</i> UWE25 | tolA | 1068 | <i>Ligilactobacillus salivarius</i> salivarius UCC118 | infA | 219 |
| <i>Mycoplasmopsis fermentans</i> PG18 | tmk | 663 | <i>Phaeobacter gallaeciensis</i> P11 | hutG | 786 |
| <i>Virgibacillus</i> sp. SK37 | tilS | 1389 | <i>Candidatus Symbiobacter mobilis</i> CR (contamination screened) | htpG | 2013 |
| <i>Virgibacillus halodenitrificans</i> PDB-F2 | thyA | 957 | <i>Kutzneria albida</i> DSM 43870 | hrcA | 1023 |
| <i>Janthinobacterium lividum</i> NCTC 8661 | thrS | 1908 | <i>Candidatus Protochlamydia amoebophila</i> UWE25 | hisH | 594 |
| <i>Xanthomonas albilineans</i> GPE PC73R | thrS | 1905 | <i>Sphingopyxis</i> sp. YR583 | hisH | 609 |
| <i>Arcobacter</i> sp. L | thrS | 1809 | <i>Streptomyces</i> sp. Wb2n-11 | hisG | 858 |
| <i>Bacillus paralicheniformis</i> MDJK30 | thiM | 810 | <i>Geobacter metallireducens</i> GS-15 | hisE | 330 |
| <i>Mycobacterium bovis</i> BCG Tokyo 172 | thiL | 1002 | <i>Corynebacterium stationis</i> ATCC 21170 | hisE | 264 |
| <i>Deferribacter desulfuricans</i> SSM1 | thiH | 1119 | <i>Micromonospora viridifaciens</i> DSM 43909 | hisE | 264 |
| <i>Phaeobacter inhibens</i> P80 | thiB | 978 | <i>Moorella thermoacetica</i> ATCC 39073 | hisC | 1161 |
| <i>Candidatus Saccharimonas aalborgensis</i> | tgt | 1248 | <i>Xanthomonas albilineans</i> FU080 | hflD | 615 |
| <i>Enterobacter ludwigii</i> P101 | tesB | 861 | <i>Phaeobacter gallaeciensis</i> P128 | hemN1 | 1356 |
| <i>Corynebacterium striatum</i> NCTC 9755 | tcsR4 | 723 | <i>Corynebacterium variabile</i> DSM 44702 | hemB | 1038 |
| <i>Corynebacterium striatum</i> 216 | tcsR3 | 639 | <i>Bacillus licheniformis</i> DSM 13 Novozymes | hemA | 1362 |
| <i>Bacillus</i> sp. H15-1 | tasA | 795 | <i>Caldicellulosiruptor changbaiensis</i> CBS-Z | hcp | 1650 |
| <i>Xanthomonas albilineans</i> HVO005 | suxR | 1032 | <i>Paenibacillus polymyxa</i> CICC 10580 | guxA | 2154 |
| <i>Streptomyces</i> sp. 57 | sseA | 840 | <i>Cupriavidus metallidurans</i> CH34 | gtrA | 447 |
| <i>Paenibacillus polymyxa</i> SC2 | srrA1 | 1275 | <i>Geobacter sulfurreducens</i> KN400 | GSU2289 | 1446 |
| <i>Enterobacter cloacae</i> A1137 | srIE | 960 | <i>Geobacter sulfurreducens</i> PCA | GSU0207 | 201 |
| <i>Priestia megaterium</i> WSH-002 | spolliAD | 384 | <i>Geobacter sulfurreducens</i> KN400 | GSU0201 | 2172 |
| <i>Phaeobacter piscinae</i> P71 | SPO1017 | 756 | <i>Phaeobacter inhibens</i> P92 | grxC | 258 |
| <i>Sphaerobacter thermophilus</i> 4ac11, DSM 20745 | smpB | 483 | <i>Rubrobacter radiotolerans</i> RSPS-4 | grpE | 708 |
| <i>Xanthomonas albilineans</i> GPE PC17 | Smlt3089 | 936 | <i>Novosphingobium pentaromativorans</i> US6-1 | groS | 315 |
| <i>Ligilactobacillus salivarius</i> GJ-24 | smc | 3537 | <i>Singulisphaera</i> sp. GP187 | greA | 480 |
| <i>Paenibacillus kribbensis</i> AM49 | sigW11 | 549 | <i>Xanthomonas albilineans</i> GPE PC17 | gpmA | 750 |
| <i>Priestia megaterium</i> WSH-002 | sigI | 720 | <i>Thioalkalivibrio</i> sp. ALRh | gmhA | 600 |
| <i>Nostoc</i> sp. Moss3 | serS | 1281 | <i>Priestia megaterium</i> Q3 | gltB | 1482 |
| <i>Actinoplanes</i> sp. N902-109 | selA | 1257 | <i>Xanthomonas albilineans</i> GPE PC17 | glpK | 1500 |
| <i>Comamonas</i> sp. 26 | secF | 957 | <i>Streptomyces viridosporus</i> T7A, ATCC 39115 | glpD2 | 1656 |
| <i>Sulfurospirillum deleyianum</i> 5175, DSM 6946 | secE | 180 | <i>Nostoc</i> sp. PCC 7107 | gloB | 774 |

Table 3 (continued)

| Genome Name | Gene Symbol | Length (bp) | Genome Name | Gene Symbol | Length (bp) |
|--|-------------|-------------|---|-------------|-------------|
| <i>Phaeobacter piscinae</i> P42 | secA | 2700 | <i>Xanthomonas albilineans</i> REU209 | glnB2 | 339 |
| <i>Phaeobacter piscinae</i> P18 | scpA | 792 | <i>Xanthomonas sacchari</i> LMG 476 | glmU | 1368 |
| <i>Lysinibacillus</i> sp. YS11 | scpA | 780 | <i>Caldicellulosiruptor bescii</i> MACB1021 | glmS | 1836 |
| <i>Streptomyces</i> sp. RJA2910 | SCO5669 | 930 | <i>Paenibacillus polymyxa</i> M1 | gldF | 723 |
| <i>Streptomyces noursei</i> ATCC 11455 | SCO5590 | 591 | <i>Actinoplanes</i> sp. N902-109 | glcA | 1248 |
| Genome sequencing | | | | | |
| <i>Streptomyces</i> sp. 3214.6 | SCO5167 | 729 | <i>Geobacillus</i> sp. C56-T3 | GK3260 | 1290 |
| <i>Streptomyces viridosporus</i> T7A, ATCC 39115 | SCO0254 | 738 | <i>Geobacillus</i> sp. 12AMOR1 | GK3216 | 963 |
| <i>Sorangium cellulosum</i> So ce 56 | sce9191 | 993 | <i>Geobacillus kaustophilus</i> HTA426 | GK3038 | 1371 |
| <i>Sorangium cellulosum</i> So ce 56 | sce0166 | 909 | <i>Geobacillus</i> sp. 12AMOR1 | GK3036 | 900 |
| <i>Geobacillus kaustophilus</i> HTA426 | SAM | 109 | <i>Geobacillus vulcani</i> PSS1 | GK2801 | 1860 |
| <i>Mycobacterium bovis</i> BCG Tokyo 172 | Rv0075 | 1173 | <i>Geobacillus kaustophilus</i> Et7/4 | GK2160 | 186 |
| <i>Geobacter sulfurreducens</i> AM-1 | ruvA | 600 | <i>Geobacillus thermoleovorans</i> FJAT-2391 | GK1813 | 486 |
| <i>Klebsiella pneumoniae</i> FDAARGOS_127 | rsxA | 582 | <i>Geobacillus vulcani</i> PSS1 | GK1582 | 699 |
| <i>Nostoc</i> sp. PCC 7524 | rsgA | 1062 | <i>Geobacillus</i> sp. GHH01 | GK1316 | 1146 |
| <i>Hydrogenobaculum</i> sp. 3684 | rpsZ | 189 | <i>Geobacillus</i> sp. GHH01 | GK1185 | 885 |
| <i>Mageeibacillus indolicus</i> UPII9-5 | rpsU | 174 | <i>Geobacillus vulcani</i> PSS1 | GK1101 | 789 |
| <i>Micromonospora citrea</i> DSM 43903 | rpsT | 267 | <i>Geobacillus kaustophilus</i> HTA426 | GK0983 | 675 |
| <i>Priestia megaterium</i> DSM 319 | rpsS | 279 | <i>Geobacillus thermoleovorans</i> FJAT-2391 | GK0603 | 657 |
| <i>Spiribacter curvatus</i> UAH-SP71 | rpsP | 270 | <i>Geobacillus kaustophilus</i> Et7/4 | GK0572 | 249 |
| <i>Caldicellulosiruptor bescii</i> RKC8122 | rpsP | 246 | <i>Geobacillus kaustophilus</i> HTA426 | GK0545 | 369 |
| <i>Methylobacillus flagellatus</i> KT | rpsN | 306 | <i>Geobacillus kaustophilus</i> HTA426 | GK0418 | 225 |
| <i>Acidobacteriaceae</i> sp. KBS 146 | rpsK | 423 | <i>Geobacillus</i> sp. LC300 | GK0324 | 1233 |
| <i>Xanthomonas albilineans</i> XaFL07-1 | rpsG | 468 | <i>Anaeromyxobacter</i> sp. Fw109-5 | gcvT | 1083 |
| <i>Phaeobacter gallaeciensis</i> P73 | rpsF | 354 | <i>Candidatus Endomicrobium trichonymphae</i> Rs-D17 | gcvPB | 1437 |
| <i>Sphingopyxis terrae ummariensis</i> UI2 | rpsE | 714 | <i>Geobacillus subterraneus</i> KCTC 3922 | gcvH | 384 |
| <i>Agromyces</i> sp. 23-23 | rpsC | 753 | <i>Cutibacterium acnes</i> PA_15_1_R1 | gcvH | 372 |
| <i>Bacillus licheniformis</i> DSM 13 Goettingen | rpoZ | 201 | <i>Corynebacterium striatum</i> 216 | gatA | 1485 |
| <i>Desulfovibrio vulgaris vulgaris</i> DP4 | rpoD | 1773 | <i>Bacillus sonorensis</i> SRCM101395 | ganA | 2055 |
| <i>Deinococcus</i> sp. NW-56 | rpoB | 3459 | <i>Paenibacillus polymyxa</i> CF05 | ganA | 1053 |
| <i>Sphingopyxis granuli</i> TFA | rpmJ | 126 | <i>Frankia</i> sp. QA3 | galE | 1041 |
| <i>Saccharopolyspora erythraea</i> DSM 40517 | rpmI | 195 | <i>Caldicellulosiruptor bescii</i> RKC8131 | fusA | 2076 |
| <i>Brevibacillus laterosporus</i> DSM 25 | rpmE | 201 | <i>Phaeobacter inhibens</i> P54 | fur | 414 |
| <i>Janthinobacterium svalbardensis</i> PAMC 27463 | rpmE | 270 | <i>Dyadobacter fermentans</i> NS114, DSM 18053 | fumC | 1404 |
| <i>Lactobacillus delbrueckii bulgaricus</i> ATCC BAA-365 | rpmC | 198 | <i>Aulosira laxa</i> NIES-50 | ftsH | 1938 |
| <i>Sphingopyxis lindanitolerans</i> WS5A3p | rpmB | 294 | <i>Nitrospira multiformis</i> NI4 | ftsH | 1896 |
| <i>Bifidobacterium kashiwanohense</i> PV20-2 | rplX | 336 | <i>Frankia</i> sp. QA3 | FRAAL6651 | 570 |
| <i>Terriglobus roseus</i> AB35.6 | rplW | 294 | <i>Frankia alni</i> ACN14a | FRAAL2500 | 804 |
| <i>Sphaerobacter thermophilus</i> 4ac11, DSM 20745 | rplV | 345 | <i>Frankia alni</i> ACN14a | FRAAL2444 | 270 |
| <i>Candidatus Protochlamydia naegleriophila</i> KNic | rplV | 336 | <i>Frankia alni</i> ACN14a | FRAAL1235 | 1128 |
| <i>Micromonospora aurantiaca</i> ATCC 27029 | rplR | 390 | <i>Frankia alni</i> ACN14a | FRAAL0999 | 912 |
| <i>Hydrogenobaculum</i> sp. HO | rplQ | 357 | <i>Modestobacter marinus</i> BC501 | folA | 579 |
| <i>Luteibacter</i> sp. 329MFSHa | rplQ | 387 | <i>Thermoclostridium stercorarium</i> stercorarium DSM 8532 | folA | 501 |
| <i>Novosphingobium</i> sp. P6W | rplP | 435 | <i>Thermoclostridium stercorarium</i> stercorarium DSM 8532 | flieE | 300 |

Table 3 (continued)

| Genome Name | Gene Symbol | Length (bp) | Genome Name | Gene Symbol | Length (bp) |
|--|-------------|-------------|--|-------------|-------------|
| <i>Corynebacterium striatum</i> 216 | rplI | 453 | <i>Cupriavidus nantongensis</i> X1 | flgK | 1920 |
| <i>Propionibacterium freudenreichii</i> freudenreichii DSM 20271 | rplE | 663 | <i>Phaeobacter inhibens</i> P92 | flgC | 393 |
| <i>Nitrosomonas</i> sp. IS79A3 | rplD | 621 | <i>Priestia megaterium</i> QM B1551 | flbD | 216 |
| <i>Solitalea canadensis</i> USAM 9D, DSM 3403 | rplD | 630 | <i>Paenibacillus</i> sp. lzh-N1 | fhuD1 | 960 |
| <i>Nitrospira defluvii</i> | rplC | 621 | <i>Arthrosira platensis</i> C1 | ffh | 1416 |
| <i>Rhodopseudomonas palustris</i> TIE-1 | RPA4706 | 582 | <i>Corynebacterium striatum</i> KC-Na-01 | fda | 1035 |
| <i>Rhodopseudomonas palustris</i> CGA009 | RPA3762 | 873 | <i>Rhodococcus jostii</i> DSM 44719 | fadE26 | 1182 |
| <i>Rhodopseudomonas palustris</i> TIE-1 | RPA3585 | 513 | <i>Simkania negevensis</i> Z, ATCC VR-1471 | fabG-B | 744 |
| <i>Rhodopseudomonas palustris</i> CGA009 | RPA2908 | 645 | <i>Phaeobacter inhibens</i> P80 | fabB | 1230 |
| <i>Rhodopseudomonas palustris</i> TIE-1 | RPA2269 | 1803 | <i>Bacillus</i> sp. IHB B 7164 | eutC | 714 |
| <i>Rhodopseudomonas palustris</i> CGA009 | RPA0026 | 1398 | <i>Phaeobacter gallaeciensis</i> P11 | edd | 1824 |
| <i>Syntrophus gentianae</i> DSM 8423 | rny | 1569 | <i>Rubrobacter radiotolerans</i> RSP5-4 | dut | 456 |
| <i>Paenibacillus polymyxa</i> J | rluD3 | 999 | <i>Laribacter hongkongensis</i> LHGZ1 | dut | 450 |
| <i>Paenibacillus polymyxa</i> E681 | rimM | 516 | <i>Pseudodesulfovibrio profundus</i> 500-1 | dsrK | 1668 |
| <i>Xanthomonas albilineans</i> XaFL07-1 | rimK | 876 | <i>Aliarcobacter butzleri</i> NCTC 12481 | dprA | 777 |
| <i>Comamonas</i> sp. 26 | recR | 591 | <i>Desulfotalea psychrophila</i> Lsv54 | DP2200 | 201 |
| <i>Phaeobacter gallaeciensis</i> P128 | recR | 600 | <i>Streptomyces</i> sp. 1222.2 | dnaQ2 | 726 |
| <i>Klebsiella pneumoniae pneumoniae</i> RUF293 | recF | 1074 | <i>Propionibacterium freudenreichii</i> freudenreichii DSM 20271 | dnaN | 1161 |
| <i>Phaeobacter inhibens</i> P83 | recA | 1068 | <i>Frankia</i> sp. EUN1f | dnaJ | 1146 |
| <i>Actinoplanes</i> sp. SE50 | rhsA | 1512 | <i>Frankia</i> sp. EUN1f | dnaJ | 1179 |
| <i>Xanthomonas albilineans</i> HVO082 | rbfA | 411 | <i>Streptomyces</i> sp. Root1310 | dnaE1 | 3540 |
| <i>Nitrosococcus watsoni</i> C-113 | queA | 1035 | <i>Bacillus licheniformis</i> 9945A | dnaB | 1425 |
| <i>Limnospira indica</i> PCC 8005 | pyrR | 534 | <i>Caldithrix abyssi</i> LF13, DSM 13497 | dnaA | 1401 |
| <i>Bacillus licheniformis</i> DSM 13 Novozymes | pyrP | 1305 | <i>Sulfurovum</i> sp. NBC37-1 | dnaA | 1329 |
| <i>Paenibacillus polymyxa</i> SC2 | pyrC | 1323 | <i>Janthinobacterium</i> sp. 61 | dkSA | 453 |
| <i>Thioalkalivibrio</i> sp. ALJ5 | pyrC | 1296 | <i>Geobacter sulfurreducens</i> PCA | divC | 336 |
| <i>Streptomyces</i> sp. 1222.2 | pyrAA | 1143 | <i>Priestia megaterium</i> ATCC 14581 | desR | 603 |
| <i>Streptomyces</i> sp. 1222.2 | puuC | 1458 | <i>Deinococcus proteolyticus</i> MRP, DSM 20540 | der | 1326 |
| <i>Paenibacillus polymyxa</i> Mc5Re-14 | purR11 | 1041 | <i>Micromonospora</i> sp. CNZ295 | deoD | 708 |
| <i>Corynebacterium doosanense</i> CAU 212, DSM 45436 | purN | 567 | <i>Actinoplanes teichomyceticus</i> DSM 43866 | deoA | 1278 |
| <i>Lactobacillus delbrueckii lactis</i> KCCM 34717 | purL | 2223 | <i>Bacillus paralicheniformis</i> Bac84 | degQ | 141 |
| <i>Deinococcus proteolyticus</i> MRP, DSM 20540 | purK | 1119 | <i>Phaeobacter piscinae</i> P71 | def1 | 519 |
| <i>Propionibacterium freudenreichii</i> shermanii JS | purE | 561 | <i>Halorhodospira halophila</i> SL1 | ddl | 915 |
| <i>Cytophaga hutchinsonii</i> ATCC 33406 | purB | 1371 | <i>Qipengyuania flava</i> VG1 | dcd | 555 |
| <i>Paenibacillus polymyxa</i> SC2 | pucR | 1665 | <i>Frankia</i> sp. QA3 | dapB | 753 |
| <i>Phaeobacter inhibens</i> P88 | ptsP | 2241 | <i>Paenibacillus polymyxa</i> M1 | dapB | 804 |
| <i>Janthinobacterium</i> sp. 13 | pssA | 852 | <i>Propionibacterium freudenreichii</i> shermanii JS | dapB | 741 |
| <i>Candidatus Accumulibacter regalis</i> UW-1 | psd | 858 | <i>Saccharopolyspora erythraea</i> NRRL 2338 | dapA1 | 924 |
| <i>Vitis vinifera</i> PN40024 | psaJ | 132 | <i>Bacillus</i> sp. FJAT-21351 | dacF | 1167 |
| <i>Actinoplanes friuliensis</i> DSM 7358 | prsA2 | 981 | <i>Laribacter hongkongensis</i> LHGZ1 | cysB1 | 939 |
| <i>Desulfohalobium retbaense</i> HR100, DSM 5692 | prmA | 891 | <i>Bacillus licheniformis</i> SNAP23 | cydC | 1725 |
| <i>Thermoanaerobacter wiegelii</i> Rt8.B1 | priA | 2199 | <i>Corynebacterium glyciniphilum</i> AJ 3170 | ctaC | 1113 |
| <i>Thioalkalivibrio</i> sp. AKL12 | prfA | 1086 | <i>Xanthomonas albilineans</i> GPE PC17 | cspA2 | 246 |

Table 3 (continued)

| Genome Name | Gene Symbol | Length (bp) | Genome Name | Gene Symbol | Length (bp) |
|--|-------------|-------------|--|-------------|-------------|
| <i>Mycobacterium bovis</i> BCG Tokyo 172 | ppsC | 6567 | <i>Thermoanaerobacter</i> sp. X514 | crcB | 405 |
| <i>Modestobacter multiseptatus</i> DSM 44402 | ppiB | 537 | <i>Geobacter sulfurreducens</i> KN400 | corA-2 | 954 |
| <i>Cutibacterium acnes</i> PA_30_2_L1 | PPA1529 | 468 | <i>Anabaenopsis circularis</i> NIES-21 | corA | 1143 |
| <i>Cutibacterium acnes</i> PA_30_2_L1 | PPA0469 | 1068 | <i>Priestia megaterium</i> DSM 319 | comGF | 438 |
| <i>Cutibacterium acnes</i> AE1 | PPA0083 | 2217 | <i>Priestia megaterium</i> WSH-002 | comGB | 1047 |
| <i>Paenibacillus polymyxa</i> Mc5Re-14 | potD1 | 1074 | <i>Phaeobacter inhibens</i> P78 | codA | 1281 |
| <i>Priestia megaterium</i> ATCC 14581 | ponA | 2835 | <i>Salinispora pacifica</i> DSM 45543 | cobD | 1002 |
| <i>Xanthomonas albilineans</i> GPE PC17 | pntA-2 | 318 | <i>Frankia alni</i> ACN14a | coaX | 753 |
| <i>Candidatus Saccharimonas aalborgensis</i> | pnp | 2124 | <i>Corynebacterium variabile</i> DSM 44702 | cmk | 657 |
| <i>Actinoplanes</i> sp. SE50 | pk3A | 2271 | <i>Deferribacter desulfuricans</i> SSM1 | clpX | 1233 |
| <i>Actinoplanes</i> sp. SE50/110 | phy1 | 1215 | <i>Bifidobacterium actinocoloniiforme</i> DSM 22766 | clpP | 633 |
| <i>Paenibacillus polymyxa</i> CICC 10580 | phnX | 837 | <i>Arthrosira platensis</i> YZ | chlL | 867 |
| <i>Nostoc</i> sp. PCC 7120 | phnD | 1002 | <i>Desulfovibrio</i> cf. <i>magneticus</i> IFRC170 | cheW | 477 |
| <i>Desulfosudis oleivorans</i> Hxd3 | pheT | 2412 | <i>Corynebacterium glutamicum</i> ATCC 21831 | Cgl3047 | 156 |
| <i>Cylindrospermum stagnale</i> PCC 7417 | pheT | 2436 | <i>Corynebacterium glutamicum</i> ATCC 21831 | Cgl2611 | 1485 |
| <i>Janthinobacterium</i> sp. 67 | pgsA | 585 | <i>Corynebacterium flavum</i> ZL-1 | Cgl2418 | 252 |
| <i>Xanthomonas albilineans</i> GPE PC86 | pglA | 1173 | <i>Corynebacterium flavum</i> ZL-1 | Cgl2255 | 294 |
| <i>Priestia megaterium</i> DSM 319 | pfyP | 645 | <i>Corynebacterium flavum</i> ZL-1 | Cgl1238 | 1143 |
| <i>Verrucospora</i> sp. CNZ293 | pfp | 1029 | <i>Corynebacterium glutamicum</i> ATCC 21831 | Cgl1220 | 1146 |
| <i>Propionibacterium freudenreichii</i> freudenreichii DSM 20271 | pf456 | 1536 | <i>Corynebacterium glutamicum</i> ATCC 13869 | Cgl1195 | 564 |
| <i>Pseudodesulfovibrio indicus</i> J2 | pcm | 642 | <i>Corynebacterium crudilactis</i> JZ16 | Cgl0754 | 681 |
| <i>Candidatus Protochlamydia amoebophila</i> UWE25 | pc0987 | 336 | <i>Corynebacterium glutamicum</i> ATCC 13869 | Cgl0388 | 1833 |
| <i>Candidatus Protochlamydia amoebophila</i> UWE25 | pc0116 | 1773 | <i>Corynebacterium flavum</i> ZL-1 | Cgl0337 | 642 |
| <i>Bacillus</i> sp. 1 s-1 | parC | 2424 | <i>Actinoplanes</i> sp. SE50 | celA | 1506 |
| <i>Dakarella massiliensis</i> ND3 | parC | 2556 | <i>Cupriavidus taiwanensis</i> LMG 19424 | cca | 1248 |
| <i>Micromonospora</i> sp. L5 | parA | 924 | <i>Thermoanaerobacter kivui</i> DSM 2030 | carB | 3219 |
| <i>Xanthomonas albilineans</i> PNG130 | panD | 381 | <i>Bradyrhizobium sachari</i> BR 10556 | bll7821 | 2097 |
| <i>Thioalkalivibrio paradoxus</i> ARh 1 | panC | 855 | <i>Bacillus paralicheniformis</i> Bac84 | BLI02578 | 438 |
| <i>Fibrella aestuarina</i> BUZ 2 | pacB | 432 | <i>Bifidobacterium longum longum</i> CCUG30698 | BL0873 | 1596 |
| <i>Rhodococcus opacus</i> ATCC 51882 | paaN | 2040 | <i>Bifidobacterium longum longum</i> CCUG30698 | BL0422 | 1977 |
| <i>Collimonas arenae</i> Ter10 | paaF | 777 | <i>Bacillus</i> sp. 1 s-1 | BL03510 | 618 |
| <i>Streptomyces scabiei</i> 87.22 | oppD3 | 996 | <i>Bacillus licheniformis</i> SNAP23 | BL03504 | 1314 |
| <i>Bacillus sonorensis</i> SRCM101395 | oppD | 1068 | <i>Bacillus licheniformis</i> SNAP23 | BL03493 | 2433 |
| <i>Mycoplasma fermentans</i> M64 | oppC-1 | 999 | <i>Bifidobacterium longum longum</i> CCUG30698 | BL0349 | 627 |
| <i>Paenibacillus polymyxa</i> Sb3-1 | occM1 | 660 | <i>Bacillus paralicheniformis</i> ATCC 12759 | BL03105 | 543 |
| <i>Rubrobacter radiotolerans</i> RSPS-4 | nusG | 540 | <i>Bacillus paralicheniformis</i> 14DA11 | BL02837 | 1857 |
| <i>Xanthomonas sacchari</i> LMG 476 | nusG | 558 | <i>Bacillus</i> sp. H15-1 | BL02416 | 660 |
| <i>Delftia</i> sp. GW456-R20 | nuoN | 1494 | <i>Bacillus paralicheniformis</i> 14DA11 | BL01721 | 1995 |
| <i>Phaeobacter gallaeciensis</i> P129 | nuoM | 1548 | <i>Bacillus licheniformis</i> SNAP23 | BL01149 | 327 |
| <i>Phaeobacter inhibens</i> P92 | nuoL | 2130 | <i>Bacillus sonorensis</i> SRCM101395 | BL00226 | 393 |
| <i>Thioalkalivibrio paradoxus</i> ARh 1 | nuoI | 489 | <i>Paenibacillus polymyxa</i> CF05 | bioB | 1008 |
| <i>Kribbella flavida</i> IFO 14399, DSM 17836 | nuoD | 1206 | <i>Xanthomonas translucens</i> pv. <i>cerealis</i> CFBP 2541 | bfr | 471 |
| <i>Acidovorax</i> sp. 93 | nuoD | 1254 | <i>Lelliottia nimipressuralis</i> SGAir0187 | betA | 1665 |

Table 3 (continued)

| Genome Name | Gene Symbol | Length (bp) | Genome Name | Gene Symbol | Length (bp) |
|---|-------------|-------------|--|-------------|-------------|
| <i>Isoptericola variabilis</i> 225 | nuoA | 363 | <i>Frateriuria aurantia</i> Kondo 67, DSM 6220 | bamB | 1209 |
| <i>Rhodococcus koreensis</i> DSM 44498 | nuoA | 360 | <i>Thauera chlorobenzoica</i> 3CB1 | azo1431 | 852 |
| <i>Streptomyces</i> sp. 57 | nucS | 672 | <i>Saccharopolyspora erythraea</i> NRRL 2338 | atzB | 1392 |
| <i>Sphingopyxis</i> sp. C-1 | nodQ | 1902 | <i>Delftia</i> sp. 60 | atpH | 540 |
| <i>Nitrospira defluvii</i> | NIDE4034 | 333 | <i>Thioalkalivibrio</i> sp. ALgr1 | atpH | 537 |
| <i>Nitrospira defluvii</i> | NIDE1341 | 573 | <i>E. coli</i> 2886–75 | atpG | 795 |
| <i>Cupriavidus metallidurans</i> CH34 | nemA | 1110 | <i>Nitrosococcus watsoni</i> C-113 | atpG | 870 |
| <i>Geobacillus thermocatenulatus</i> KCTC 3921 | ndoA | 351 | <i>Leuconostoc gelidum gasicomitatum</i> LMG 18811 | atpC | 450 |
| <i>Phaeobacter inhibens</i> P80 | ndk | 423 | <i>Phaeobacter gallaeciensis</i> P129 | atpC | 414 |
| <i>Flavobacterium johnsoniae</i> UW101, ATCC 17061 | nbaC | 540 | <i>Nostoc</i> sp. PCC 7107 | asr1559 | 252 |
| <i>Streptomyces</i> sp. 57 | nagB | 786 | <i>Nostoc</i> sp. PCC 7107 | asr0064 | 237 |
| <i>Fimbrimonas ginsengisoli</i> Gsoil 348 | nadK | 849 | <i>Xanthomonas albilineans</i> MTQ032 | aspS | 1752 |
| <i>Nitrospira briensis</i> Nsp8 | nadA | 1101 | <i>Geobacillus</i> sp. C56-T3 | aroA | 1083 |
| <i>Paenibacillus polymyxa</i> M1 | arnT | 2355 | <i>Micromonospora</i> sp. CNZ295 | alaS | 2679 |
| <i>Bacillus paralicheniformis</i> MDJK30 | argJ | 1221 | <i>Acidovorax</i> sp. 93 | ahcY | 1434 |
| <i>Sphaerobacter thermophilus</i> 4ac11, DSM 20745 | argH | 1371 | <i>Modestobacter marinus</i> BC501 | adk | 624 |
| <i>Janthinobacterium svalbardensis</i> PAMC 27463 | argA | 1320 | <i>Lactobacillus delbrueckii bulgaricus</i> ND02 | addA | 3684 |
| <i>Ligilactobacillus salivarius salivarius</i> UCC118 | apt | 519 | <i>Collimonas arenae</i> Ter282 | aceK | 1785 |
| <i>Priestia megaterium</i> QM B1551 | amt | 1227 | <i>Lysinibacillus</i> sp. YS11 | accA | 957 |
| <i>Nostoc</i> sp. PCC 7120 | alr4917 | 1689 | <i>Cupriavidus necator</i> NH9 | aat | 759 |
| <i>Nostoc</i> sp. PCC 7120 | alr3663 | 1050 | <i>Phaeobacter inhibens</i> P88 | aat | 633 |
| <i>Nostoc</i> sp. PCC 7120 | alr2594 | 435 | <i>Nostoc</i> sp. PCC 7120 | all4101 | 384 |
| <i>Trichormus variabilis</i> ATCC 29413 | alr0203 | 480 | <i>Nostoc</i> sp. Moss5 | all3116 | 738 |
| <i>Nostoc</i> sp. PCC 7120 | all5344 | 468 | <i>Nostoc</i> sp. PCC 7120 | all1863 | 864 |
| <i>Trichormus variabilis</i> NIES-23 | all4824 | 798 | <i>Trichormus variabilis</i> ATCC 29413 | all0781 | 1590 |
| | | | <i>Trichormus variabilis</i> ATCC 29413 | all4426 | 1254 |

high-throughput experiments. In this study, pathway analysis was used to identify 16 significant pathways enriched in the predicted genes. The network of pathway enrichment of the metagenome data has been shown in Fig. 10. These pathways are involved in a variety of essential cellular processes, including biosynthesis, energy production, and signaling. The CMP-KDO biosynthesis II (from D-arabinose 5-phosphate) pathway is one of the most significant pathways identified in this study [39]. It is involved in the biosynthesis of lipopolysaccharide (LPS), an essential component of the outer membrane of gram-negative bacteria (Fig. 1S(i)). Two sequences, aconitate hydratase (K01681) and citrate synthase (K01647), are associated with the TCA cycle pathway. They have been found in *Glycine max* (soybeans) and *Saccharomyces cerevisiae* (yeast). The TCA cycle is essential for optimal functioning of primary carbon metabolism in plants (Fig. 1S(ii-iv)). Aconitate hydratase catalyzes the

isomerization of citrate to isocitrate in the TCA cycle. The function of aconitate hydratase has been well studied in model plants, such as *Arabidopsis thaliana*. The TCA cycle is a metabolic process that occurs in plants, animals, fungi, and other bacteria. It is a series of chemical reactions that converts acetyl-CoA into carbon dioxide and energy. The TCA cycle is an important source of energy for cells and also plays a role in the synthesis of other molecules such as amino acids and fatty acids [40]. The next pathway involved the biosynthesis of fatty acids (Fig. 1S(v)). This is essential for the formation of membranes, which are necessary for the viability of all cells, except Archaea. Fatty acids are also a compact energy source for seed germination. Enenoyl-[acyl-carrier protein] reductase I (K00208) is an enzyme involved in fatty acid biosynthesis, prodigiosin biosynthesis, and biotin metabolism pathways. Another significant pathway that has been identified is biotin metabolism (Fig. 1S(vi)). The