

**Complete genome sequence of *Streptomyces* sp. HSG2 from rhizosphere soil of mangrove in Qingmei Gang, Sanya**

Shengxiang Pei<sup>1</sup>, Siwen Niu<sup>1</sup>, Fuquan Xie<sup>1</sup>, Shuang Zhang<sup>1</sup>, Wenjing Wang<sup>1</sup>, Gaiyun Zhang<sup>1\*</sup>

<sup>1</sup>Key Laboratory of Marine Biogenetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen 361005, Fujian, P. R. China

**\* Corresponding author:**

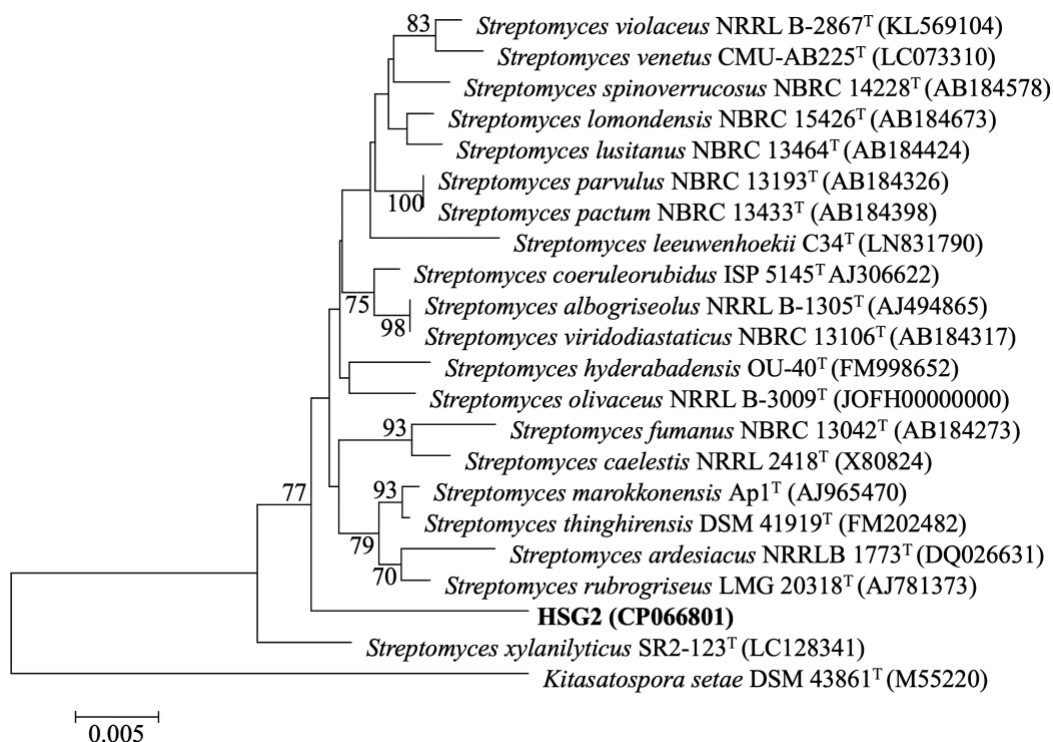
Gaiyun Zhang

[zhgyun@tio.org.cn](mailto:zhgyun@tio.org.cn)

## Content

|                                                                                                                                                                                            |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Fig. S1</b> Neighbor-joining phylogenetic tree based on 16S rRNA gene sequences showing the relationship between strain HGS2 and the members of the genus <i>Streptomyces</i> . . . . . | 3  |
| <b>Table S1</b> Predicted CRISPRs in <i>Streptomyces</i> sp. HSG2 . . . . .                                                                                                                | 4  |
| <b>Table S2</b> Predicted genomic islands in <i>Streptomyces</i> sp. HSG2 . . . . .                                                                                                        | 5  |
| <b>Table S3</b> Genes related to aerobic denitrification, phenol degradation and xenobiotics degradation as well as industrial important enzymes . . . . .                                 | 6  |
| <b>Table S4</b> Secondary metabolite biosynthesis gene clusters predicted in <i>Streptomyces</i> sp. HSG2. . . . .                                                                         | 9  |
| <b>Table S5</b> Putative genes identified in the nystatin A1 biosynthetic gene cluster of HSG2 . . . . .                                                                                   | 11 |
| <b>Table S6</b> Putative genes identified in the streptobactin biosynthetic gene cluster of HSG2. . . . .                                                                                  | 12 |

**Fig. S1** Neighbor-joining phylogenetic tree based on 16S rRNA gene sequences showing the relationship between strain HGS2 and the members of the genus *Streptomyces*. The tree was reconstructed using mega 7.0 with Kimura's two-parameter model. Bootstrap values (expressed as percentages of 1000 replications) over 50 % are shown at branching nodes. Bar, 0.005 substitutions per nucleotide position



**Table S1** Predicted CRISPRs in *Streptomyces* sp. HSG2

| CRISPR<br>Id | Start   | End     | Length (bp) | Spacer | Repeat consensus               |
|--------------|---------|---------|-------------|--------|--------------------------------|
| 1            | 2201159 | 2201264 | 105         | 1      | TCACCGACTCTACCAGACCGTTTC       |
| 2            | 2746510 | 2746598 | 88          | 1      | CGTGGTGGGCGTTCGTTTCGAGCGTCGTGA |
| 3            | 2778270 | 2778375 | 105         | 1      | GAAACGGTCTGGTAGAGTCGGAGA       |
| 4            | 3679446 | 3680388 | 942         | 15     | GAGCAACCCCCGCAGGCGCGGGGACGAC   |
| 5            | 3690859 | 3692655 | 1796        | 29     | GTCGTCCCCGCGCCTGCGGGGGTAGCTC   |
| 6            | 3694669 | 3695672 | 1003        | 16     | GTCGTCCCCGCGCCTGCGGGGGTAGCTC   |
| 7            | 4380343 | 4380448 | 105         | 1      | GAAACGGTCTGGTAGAGTCGGAGA       |

**Table S2** Predicted genomic islands in *Streptomyces* sp. HSG2

| Genomic islands | Start     | End       | Length (bp) |
|-----------------|-----------|-----------|-------------|
| GI-1            | 30,554    | 37,710    | 7,156       |
| GI-2            | 60,889    | 163,706   | 102,817     |
| GI-3            | 229,181   | 252,029   | 22,848      |
| GI-4            | 710,228   | 729,730   | 19,502      |
| GI-5            | 785,192   | 794,271   | 9,079       |
| GI-6            | 937,152   | 946,658   | 9,506       |
| GI-7            | 1,078,710 | 1,092,226 | 13,516      |
| GI-8            | 1,417,032 | 1,423,721 | 6,689       |
| GI-9            | 1,577,111 | 1,583,255 | 6,144       |
| GI-10           | 1,990,561 | 1,996,472 | 5,911       |
| GI-11           | 2,033,987 | 2,043,398 | 9,411       |
| GI-12           | 2,105,927 | 2,138,582 | 32,655      |
| GI-13           | 2,109,461 | 2,116,087 | 6,626       |
| GI-14           | 2,218,022 | 2,225,264 | 7,242       |
| GI-15           | 2,292,776 | 2,303,802 | 11,026      |
| GI-16           | 2,336,758 | 2,343,232 | 6,474       |
| GI-17           | 2,380,193 | 2,388,700 | 8,507       |
| GI-18           | 2,424,252 | 2,433,787 | 9,535       |
| GI-19           | 2,520,622 | 2,528,383 | 7,761       |
| GI-20           | 2,793,913 | 2,810,046 | 16,133      |
| GI-21           | 3,050,801 | 3,058,126 | 7,325       |
| GI-22           | 3,116,866 | 3,126,525 | 9,659       |
| GI-23           | 3,136,792 | 3,156,244 | 19,452      |
| GI-24           | 3,697,297 | 3,713,909 | 16,612      |
| GI-25           | 3,876,657 | 3,952,357 | 75,700      |
| GI-26           | 3,923,593 | 3,939,707 | 16,114      |
| GI-27           | 4,457,537 | 4,476,140 | 18,603      |
| GI-28           | 4,557,012 | 4,567,418 | 10,406      |
| GI-29           | 5,186,073 | 5,214,210 | 28,137      |
| GI-30           | 5,223,951 | 5,245,259 | 21,308      |

**Table S3** Genes related to aerobic denitrification, phenol degradation and xenobiotics degradation as well as industrial important enzymes

| Locus tag                      | Start   | End     | Strand | Length (bp) | Annotation                                      |
|--------------------------------|---------|---------|--------|-------------|-------------------------------------------------|
| <b>Enzymes</b>                 |         |         |        |             |                                                 |
| HSG2_0612                      | 800255  | 802717  | +      | 2462        | alpha-amylase                                   |
| HSG2_0792                      | 1002981 | 1004339 | -      | 1358        | protease                                        |
| HSG2_0795                      | 1008567 | 1009766 | -      | 1199        | citrate synthase                                |
| HSG2_0799                      | 1014290 | 1015126 | -      | 836         | serine protease                                 |
| HSG2_0906                      | 1157506 | 1158375 | -      | 869         | putative phospholipase                          |
| HSG2_1051                      | 1326738 | 1328783 | +      | 2045        | neutral zinc metalloprotease                    |
| HSG2_1880                      | 2297443 | 2298894 | +      | 1451        | prohead protease protein                        |
| HSG2_2292                      | 2747593 | 2749398 | -      | 1805        | protease                                        |
| HSG2_3072                      | 3592804 | 3594090 | +      | 1286        | ATP-dependent protease ATP-binding subunit ClpX |
| HSG2_3206                      | 3753432 | 3754103 | +      | 671         | lipase, class 2                                 |
| HSG2_3733                      | 4353501 | 4355036 | -      | 1535        | ssp-like protease                               |
| HSG2_3770                      | 4396198 | 4398000 | -      | 1802        | glycosyl hydrolase                              |
| HSG2_4066                      | 4720229 | 4721029 | -      | 800         | glycosyl hydrolase                              |
| <b>Aerobic denitrification</b> |         |         |        |             |                                                 |
| HSG2_0481                      | 648856  | 649365  | +      | 509         | nitrate reductase delta chain NarJ              |
| HSG2_0595                      | 783328  | 785019  | +      | 1691        | nitrite/sulphite reductase                      |
| HSG2_1457                      | 1792450 | 1793862 | -      | 1412        | putative nitrate extrusion protein              |
| HSG2_3203                      | 3748355 | 3748849 | +      | 494         | nitrite/sulphite reductase small subunit        |
| HSG2_3204                      | 3748846 | 3751464 | +      | 2618        | nitrite reductase large subunit NirB            |
| HSG2_3205                      | 3751464 | 3752774 | +      | 1310        | nitrite reductase NirB                          |
| <b>Hydrocarbon degradation</b> |         |         |        |             |                                                 |
| HSG2_0407                      | 568250  | 569278  | +      | 1028        | alcohol dehydrogenase                           |
| HSG2_1197                      | 1494394 | 1495374 | -      | 980         | alcohol dehydrogenase                           |
| HSG2_1434                      | 1764269 | 1765414 | +      | 1145        | 4-hydroxyphenylpyruvate dioxygenase             |
| HSG2_0164                      | 281299  | 282225  | -      | 926         | haloalkane dehalogenase                         |
| HSG2_2091                      | 2511597 | 2512607 | +      | 1010        | phenylacetate-CoA oxygenase subunit PaaA        |

|                                      |         |         |   |      |                                                     |
|--------------------------------------|---------|---------|---|------|-----------------------------------------------------|
| HSG2_2092                            | 2512604 | 2512891 | + | 287  | phenylacetate-CoA oxygenase subunit PaaB            |
| HSG2_2093                            | 2512888 | 2513622 | + | 734  | phenylacetic acid degradation protein               |
| HSG2_2094                            | 2513619 | 2514173 | + | 554  | phenylacetic acid degradation protein               |
| HSG2_2095                            | 2514173 | 2515237 | + | 1064 | phenylacetic acid degradation NADH oxidoreductase   |
| HSG2_2555                            | 3034466 | 3035932 | + | 1466 | aldehyde dehydrogenase                              |
| HSG2_3292                            | 3854282 | 3855790 | + | 1508 | aldehyde dehydrogenase                              |
| HSG2_3293                            | 3855787 | 3856056 | + | 269  | ferredoxin                                          |
| HSG2_3391                            | 3970863 | 3972125 | - | 1262 | 2-polyprenylphenol hydroxylase-like protein         |
| HSG2_3807                            | 4437382 | 4438698 | - | 1316 | homogentisate 1,2-dioxygenase                       |
| HSG2_3900                            | 4536007 | 4537128 | + | 1121 | phenylalanyl-tRNA synthetase subunit alpha          |
| HSG2_3901                            | 4537128 | 4539650 | + | 2522 | phenylalanyl-tRNA synthetase subunit beta           |
| HSG2_4452                            | 5162590 | 5163213 | + | 623  | phenazine biosynthesis protein PhzD                 |
| HSG2_4453                            | 5163210 | 5165171 | + | 1961 | phenazine biosynthesis protein PhzE                 |
| <b>Other xenobiotics degradation</b> |         |         |   |      |                                                     |
| HSG2_0027                            | 21004   | 22500   | + | 1496 | cyclohexanone monooxygenase                         |
| HSG2_0028                            | 22542   | 24050   | + | 1508 | putative oxygen-dependent FAD-linked oxidoreductase |
| HSG2_0070                            | 61609   | 63117   | - | 1508 | monooxygenase                                       |
| HSG2_0104                            | 163703  | 164905  | + | 1202 | cytochrome P450                                     |
| HSG2_0106                            | 165204  | 166388  | - | 1184 | cytochrome P450                                     |
| HSG2_0110                            | 219147  | 220355  | - | 1208 | cytochrome P450                                     |
| HSG2_0164                            | 281299  | 282225  | - | 926  | haloalkane dehalogenase                             |
| HSG2_0341                            | 493392  | 495038  | + | 1646 | FAD-binding monooxygenase                           |
| HSG2_1070                            | 1351750 | 1352118 | - | 368  | antibiotic biosynthesis monooxygenase               |
| HSG2_1166                            | 1455359 | 1456681 | - | 1322 | FAD-binding monooxygenase                           |
| HSG2_1922                            | 2341327 | 2342298 | - | 971  | histidine kinase                                    |
| HSG2_2026                            | 2439552 | 2440409 | + | 857  | tryptophan 2,3-dioxygenase                          |

|           |         |         |   |      |                                   |
|-----------|---------|---------|---|------|-----------------------------------|
| HSG2_2524 | 3000088 | 3001152 | + | 1064 | fatty acid desaturase             |
| HSG2_2551 | 3029741 | 3030706 | + | 965  | monooxygenase                     |
| HSG2_2552 | 3030765 | 3031511 | + | 746  | oxidoreductase                    |
| HSG2_2553 | 3031609 | 3032727 | + | 1118 | monooxygenase                     |
| HSG2_2885 | 3384967 | 3386562 | + | 1595 | histidine ammonia-lyase           |
| HSG2_2932 | 3430154 | 3430708 | + | 554  | alkyl hydroperoxide reductase     |
| HSG2_3051 | 3563446 | 3564906 | + | 1460 | monooxygenase FAD-binding protein |
| HSG2_3491 | 4077723 | 4078532 | + | 809  | cytochrome c heme-binding subunit |
| HSG2_4410 | 5112594 | 5113676 | + | 1082 | FAD dependent oxidoreductase      |
| HSG2_4411 | 5113673 | 5114980 | + | 1307 | L-lysine 6-monooxygenase (NADPH)  |
| HSG2_4456 | 5167163 | 5168857 | + | 1694 | monooxygenase FAD-binding protein |
| HSG2_4553 | 5254293 | 5255789 | - | 1496 | cyclohexanone monooxygenase       |



**Table S4** Secondary metabolite biosynthesis gene clusters predicted in *Streptomyces* sp. HSG2

| Region           | Cluster assignment | From          | To            | Most similar known cluster (percent of similarity)              | Reference strain                              | MIBiG accession   |
|------------------|--------------------|---------------|---------------|-----------------------------------------------------------------|-----------------------------------------------|-------------------|
| Region 1         | Ectoine            | 4049          | 13725         | Kosinostatin (11%)                                              | <i>Micromonospora</i> sp. TP-A0468            | BGC0001073        |
| <b>Region 2</b>  | <b>Type I PKS</b>  | <b>69715</b>  | <b>238373</b> | <b>Nystatin A1 (72%)</b>                                        | <b><i>Streptomyces noursei</i> ATCC 11455</b> | <b>BGC0000115</b> |
| Region 3         | Indole             | 300623        | 321422        | 5-isoprenylindole-3-carboxylate $\beta$ -D-glycosyl ester (23%) | <i>Streptomyces</i> sp. RM-5-8                | BGC0001483        |
| Region 4         | Ladderane          | 382412        | 424114        | Metatricycloene (6%)                                            | uncultured bacterium                          | BGC0001369        |
| Region 5         | Oligosaccharide    | 439305        | 495553        | Staurosporine (100%)                                            | <i>Streptomyces</i> sp. TP-A0274              | BGC0000825        |
| Region 6         | Bacteriocin        | 535637        | 547571        | -                                                               | -                                             | -                 |
| Region 7         | Terpene            | 576247        | 602424        | Hopene (76%)                                                    | <i>Streptomyces coelicolor</i>                | BGC0000663        |
| Region 8         | Terpene            | 733290        | 759199        | Isorenieratene (100%)                                           | <i>Streptomyces argillaceus</i>               | BGC0001456        |
| Region 9         | Terpene            | 796131        | 816654        | Geosmin (100%)                                                  | <i>Streptomyces coelicolor</i>                | BGC0001181        |
| <b>Region 10</b> | <b>NRPS</b>        | <b>849030</b> | <b>946839</b> | <b>Streptobactin (76%)</b>                                      | <b><i>Streptomyces</i> sp. ATCC 700974</b>    | <b>BGC0000368</b> |
| Region 11        | Siderophore        | 1039265       | 1051175       | -                                                               | -                                             | -                 |
| Region 12        | Type II PKS        | 1322438       | 1394935       | Spore pigment (75%)                                             | <i>Streptomyces avermitilis</i>               | BGC0000271        |

|           |               |         |         |                                            |                                                                 |            |
|-----------|---------------|---------|---------|--------------------------------------------|-----------------------------------------------------------------|------------|
| Region 13 | Blactam       | 1454604 | 1479343 | Clavulanic acid (29%)                      | <i>Streptomyces clavuligerus</i><br>ATCC 27064                  | BGC0000845 |
| Region 14 | Terpene       | 1544280 | 1565365 | Julichrome Q3-3 / julichrome<br>Q3-5 (22%) | <i>Streptomyces afghaniensis</i><br>772                         | BGC0002012 |
| Region 15 | TfuA-related  | 2447310 | 2470050 | -                                          | -                                                               | -          |
| Region 16 | Linaridin     | 2752709 | 2769496 | Pentostatine / vidarabine (6%)             | <i>Streptomyces antibioticus</i>                                | BGC0001735 |
| Region 17 | Lanthipeptide | 3062313 | 3085162 | Venezuelin (100%)                          | <i>Streptomyces venezuelae</i><br>ATCC 10712                    | BGC0000563 |
| Region 18 | Type I PKS    | 3526458 | 3576326 | Salinosporamide A (28%)                    | <i>Salinispora tropica</i>                                      | BGC0000145 |
| Region 19 | NRPS          | 3757068 | 3803679 | SCO-2138 (21%)                             | <i>Streptomyces coelicolor</i><br>A3(2)                         | BGC0000595 |
| Region 20 | NRPS          | 3901644 | 3952885 | Streptomycin (22%)                         | <i>Streptomyces griseus</i> subsp.<br><i>griseus</i> NBRC 13350 | BGC0000724 |
| Region 21 | Ectoine       | 4311837 | 4321983 | Ectoine (100%)                             | <i>Streptomyces anulatus</i>                                    | BGC0000853 |
| Region 22 | NRPS-like     | 4439196 | 4479982 | -                                          | -                                                               | -          |
| Region 23 | Thiopeptide   | 5022179 | 5158602 | Triacsins (84%)                            | <i>Kitasatospora aureofaciens</i>                               | BGC0001983 |
| Region 24 | Ectoine       | 5262642 | 5273016 | Kosinostatin (11%)                         | <i>Micromonospora</i> sp. TP-<br>A0468                          | BGC0001073 |

**Table S5** Putative genes identified in the nystatin A1 biosynthetic gene cluster of HSG2

| Locus tag | Gene          | Protein size (no. of amino acids) | Strand | Product                                             |
|-----------|---------------|-----------------------------------|--------|-----------------------------------------------------|
| HSG2_0115 | <i>orf1</i>   | 270                               | -      | Probable cadicidin biosynthesis thioesterase        |
| HSG2_0114 | <i>orf2</i>   | 344                               | -      | GDP-mannose 4,6 dehydratase 1                       |
| HSG2_0113 | <i>nysG</i>   | 632                               | +      | ABC transporter                                     |
| HSG2_0112 | <i>nysH</i>   | 576                               | +      | ABC transporter                                     |
| HSG2_0111 | <i>orf3</i>   | 452                               | -      | Crotonyl-CoA reductase                              |
| HSG2_0110 | <i>orf4</i>   | 403                               | -      | Cytochrome P450                                     |
| HSG2_0109 | <i>nysI</i>   | 9011                              | -      | Nystatin PKS (modules 9–14)                         |
| HSG2_0108 | <i>nysJ</i>   | 6731                              | -      | Nystatin PKS (modules 15–17)                        |
| HSG2_0107 | <i>nysK</i>   | 1723                              | -      | Nystatin PKS (modules 18+TE)                        |
| HSG2_0106 | <i>nysJ</i>   | 395                               | -      | Cytochrome P450                                     |
| HSG2_0105 | <i>nysK</i>   | 65                                | +      | Ferredoxin-1                                        |
| HSG2_0104 | <i>nysL</i>   | 401                               | +      | Cytochrome P450                                     |
| HSG2_0103 | <i>nysDII</i> | 354                               | +      | Spore coat polysaccharide biosynthesis protein      |
| HSG2_0102 | <i>nysDI</i>  | 461                               | +      | Glucuronosyltransferase                             |
| HSG2_0101 | <i>orf5</i>   | 921                               | -      | Putative HTH-type transcriptional regulator Mb0914c |
| HSG2_0100 | <i>orf6</i>   | 541                               | +      | 3-hydroxycinnamic acid hydroxylase                  |
| HSG2_0099 | <i>orf7</i>   | 343                               | +      | Hypothetical protein                                |
| HSG2_0098 | <i>nysA</i>   | 2621                              | -      | Nystatin PKS (loading module)                       |
| HSG2_0097 | <i>nysB</i>   | 9294                              | -      | Nystatin PKS (modules 1 and 2)                      |
| HSG2_0096 | <i>nysC</i>   | 9936                              | -      | Nystatin PKS (modules 3–8)                          |
| HSG2_0095 | <i>orf8</i>   | 565                               | -      | 3-hydroxybutyryl-CoA dehydrogenase                  |
| HSG2_0094 | <i>orf9</i>   | 492                               | +      | Putative transposase YncI                           |
| HSG2_0093 | <i>orf10</i>  | 173                               | -      | Hypothetical protein                                |
| HSG2_0092 | <i>orf11</i>  | 136                               | -      | Putative transposase YncI                           |
| HSG2_0091 | <i>orf12</i>  | 231                               | +      | Hypothetical protein                                |

**Table S6** Putative genes identified in the streptobactin biosynthetic gene cluster of HSG2

| Locus tag | Gene        | Protein size (no. of amino acids) | Strand | Product                                                       |
|-----------|-------------|-----------------------------------|--------|---------------------------------------------------------------|
| HSG2_0664 | <i>orf1</i> | 510                               | +      | Carboxyvinyltransferase                                       |
| HSG2_0665 | <i>orf2</i> | 327                               | -      | Transcriptional regulator YobV                                |
| HSG2_0666 | <i>orf3</i> | 315                               | +      | Epoxide hydrolase                                             |
| HSG2_0667 | <i>dhbG</i> | 76                                | -      | Isochorismatase                                               |
| HSG2_0668 | <i>dhbB</i> | 211                               | -      | Isochorismatase                                               |
| HSG2_0669 | <i>dhbE</i> | 552                               | -      | DHBA-AMP ligase                                               |
| HSG2_0670 | <i>dhbC</i> | 402                               | -      | Isochorismate synthase DHBC                                   |
| HSG2_0671 | <i>dhbA</i> | 290                               | -      | 2,3-Dihydro-DHBA dehydrogenase                                |
| HSG2_0672 | <i>griA</i> | 758                               | +      | Fe <sup>3+</sup> -citrate import system permease protein YfmD |
| HSG2_0673 | <i>griB</i> | 415                               | +      | Enterochelin esterase                                         |
| HSG2_0674 | <i>griC</i> | 528                               | -      | Multidrug resistance protein B                                |
| HSG2_0675 | <i>griD</i> | 70                                | -      | Protein MbtH                                                  |
| HSG2_0676 | <i>griE</i> | 2396                              | -      | Nonribosomal peptide synthase                                 |
| HSG2_0677 | <i>griF</i> | 398                               | +      | Peptidase M20 domain-containing protein 2                     |
| HSG2_0678 | <i>griG</i> | 335                               | -      | Probable siderophore-binding lipoprotein YfiY                 |
| HSG2_0679 | <i>griH</i> | 290                               | +      | Fe <sup>3+</sup> -citrate import ATP-binding protein YfmF     |
| HSG2_0680 | <i>orf4</i> | 64                                | -      | Hypothetical protein                                          |
| HSG2_0681 | <i>orf5</i> | 77                                | +      | Hypothetical protein                                          |
| HSG2_0682 | <i>orf6</i> | 118                               | +      | Hypothetical protein                                          |