**Complete Genome Sequence of the Butirosin-producing *Bacillus vitellinus* NBRC 13296 and its reclassification to *Paenibacillus chitinolyticus***

**Kyung-A Hyun, Seung-Young Kim, Kyung-Hwan Boo,** **Won-Jae Chi and** **Chang-Gu Hyun**

\* Correspondence: wjchi76@korea.kr and cghyun@jejunu.ac.kr; Tel.: +82-64-754-3900

Supplementary Materials

Table S1. Comparison of 16S rRNA genes in *B. vitellinus* NBRC 13296 and Type strains of the genus *Paenibacillus1*

Table S2. Putative gene clusters coding for secondary metabolites in *B. vitellinus* NBRC 13296 derived using BAGEL 4.0.3

Table S3. Putative genetic organization of contig1.0.AOI\_02 derived using BAGEL 4.0. and BLASTP packages4

Table S4. Putative gene clusters coding for secondary metabolites in B. vitellinus NBRC 13296 derived using PRISM 4.4.5. 5

Table S5. Genetic organization and sequence homology of butirosin BGCs derived from *B. vitellinus* NBRC 13296 and *N. circulans* ATCC 215586

Figure S1. Maximum-likelihood phylogenetic tree based on the complete 16S rRNA gene sequence, extracted from the assembled genome8

Figure S2. Distribution of secondary metabolite gene clusters (1–18) in *B. vitellinus* NBRC 13296 as predicted by the antiSMASH package9

Figure S3. Genetic organization of BGC region 15 as predicted by the antiSMASH package10

Table S1. Comparison of 16S rRNA genes in *B. vitellinus* NBRC 13296 and Type strains of the genus *Paenibacillus*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Accession** | **Description** | **Total Score** | **Query Cover** | **Per. Ident** |
| NR\_040854.1 | P. chitinolyticus HSCC 596 16S ribosomal RNA | 2717 | 95% | 99.66% |
| NR\_043219.1 | P. gansuensis B518 16S ribosomal RNA | 2435 | 97% | 95.71% |
| NR\_173496.1 | P. lutrae N10 16S ribosomal RNA | 2431 | 97% | 95.60% |
| NR\_029071.1 | P. favisporus GMP01 16S ribosomal RNA | 2422 | 99% | 94.96% |
| NR\_148738.1 | P. yonginensis DCY84 16S ribosomal RNA | 2410 | 100% | 94.68% |
| NR\_117730.2 | P. polymyxa DSM 36 16S ribosomal RNA | 2401 | 99% | 94.71% |
| NR\_043166.1 | P. rhizosphaerae CECAP06 16S ribosomal RNA | 2399 | 99% | 94.71% |
| NR\_178528.1 | P. swuensis DY6 16S ribosomal RNA | 2388 | 95% | 95.52% |
| NR\_136854.1 | P. wulumuqiensis Y24 16S ribosomal RNA | 2370 | 97% | 94.97% |
| NR\_029109.1 | P. xylanilyticus XIL14 16S ribosomal RNA | 2366 | 99% | 94.32% |
| NR\_114621.1 | P. turicensis MOL722 16S ribosomal RNA | 2364 | 99% | 94.28% |
| NR\_117012.1 | P. uliginis N3/975 16S ribosomal RNA | 2364 | 97% | 94.90% |
| NR\_117742.1 | P. peoriae KCTC 3763 16S ribosomal RNA | 2362 | 99% | 94.38% |
| NR\_148890.1 | P. bovis BD3526 16S ribosomal RNA | 2361 | 100% | 94.09% |
| NR\_040890.1 | P. apiarius DSM 5581 16S ribosomal RNA | 2361 | 98% | 94.48% |
| NR\_116789.1 | P. tianmuensis B27 16S ribosomal RNA | 2357 | 99% | 94.32% |
| NR\_122065.1 | P. sabinae T27 16S ribosomal RNA | 2353 | 99% | 94.10% |
| NR\_180160.1 | P. nuruki TI45-13ar 16S ribosomal RNA | 2350 | 100% | 93.97% |
| NR\_165675.1 | P. lutimineralis MBLB1234 16S ribosomal RNA | 2348 | 100% | 93.98% |
| NR\_133806.1 | P. relictisesami KB0549 16S ribosomal RNA | 2348 | 97% | 94.66% |
| NR\_040892.1 | P. validus JCM 9077 16S ribosomal RNA | 2346 | 96% | 95.00% |
| NR\_178645.1 | P. enshidis RP-207 16S ribosomal RNA | 2344 | 99% | 94.07% |
| NR\_115664.1 | P. aquaticus GPTSA 19 16S ribosomal RNA | 2344 | 95% | 95.24% |
| NR\_041929.1 | P. mendelii C/2 16S ribosomal RNA | 2344 | 99% | 94.12% |
| NR\_025391.1 | P. azoreducens CM1 16S ribosomal RNA | 2344 | 96% | 94.87% |
| NR\_112054.1 | P. ehimensis IFO 15659 16S ribosomal RNA | 2338 | 97% | 94.47% |
| NR\_181067.1 | P. rhizovicinus 14171R-81 16S ribosomal RNA | 2337 | 100% | 93.91% |
| NR\_180424.1 | P. protaetiae FW100M-2 16S ribosomal RNA | 2337 | 100% | 93.83% |
| NR\_118573.1 | P. lentus CMG1240 16S ribosomal RNA | 2337 | 99% | 94.00% |
| NR\_042189.1 | P. cineris LMG 18439 16S ribosomal RNA | 2337 | 96% | 94.63% |
| NR\_132304.1 | P. profundus Sl 79 16S ribosomal RNA | 2337 | 97% | 94.50% |
| NR\_179441.1 | P. phocaensis mt24 16S ribosomal RNA | 2333 | 97% | 94.36% |
| NR\_028817.1 | P. naphthalenovorans PR-N1 16S ribosomal RNA | 2333 | 99% | 93.95% |
| NR\_178252.1 | P. lemnae L7-75 16S ribosomal RNA | 2331 | 96% | 94.57% |
| NR\_040884.1 | P. illinoisensis JCM 9907 16S ribosomal RNA | 2331 | 96% | 94.62% |
| NR\_042009.1 | P. jamilae CECT 5266 16S ribosomal RNA | 2329 | 98% | 93.91% |
| NR\_025024.1 | P. chinjuensis WN9 16S ribosomal RNA | 2329 | 97% | 94.78% |
| NR\_134783.2 | P. nicotianae YIM h-19 16S ribosomal RNA | 2326 | 99% | 93.87% |
| NR\_040886.1 | P. macerans IAM 12467 16S ribosomal RNA | 2324 | 96% | 94.51% |
| NR\_025299.1 | P. borealis KK19 16S ribosomal RNA | 2324 | 98% | 93.91% |
| NR\_137363.1 | P. physcomitrellae XB 16S ribosomal RNA | 2322 | 97% | 94.29% |
| NR\_169361.1 | P. dakarensis FF9 16S ribosomal RNA | 2320 | 97% | 94.29% |
| NR\_147741.1 | P. terreus D33 16S ribosomal RNA | 2320 | 97% | 94.28% |
| NR\_151978.1 | P. aceti L14 16S ribosomal RNA | 2316 | 97% | 94.27% |
| NR\_145625.1 | P. yunnanensis YN2 16S ribosomal RNA | 2316 | 99% | 93.87% |
| NR\_040885.1 | P. chibensis JCM 9905 16S ribosomal RNA | 2316 | 96% | 94.23% |
| NR\_179515.1 | P. tuaregi Marseille-P2472 16S ribosomal RNA | 2314 | 97% | 94.09% |
| NR\_144710.1 | P. antibioticophila GD11 16S ribosomal RNA | 2314 | 97% | 94.39% |
| NR\_179433.1 | P. rubinfantis MT18 16S ribosomal RNA | 2313 | 97% | 94.22% |
| NR\_112162.1 | P. campinasensis JCM 11200 16S ribosomal RNA | 2313 | 96% | 94.37% |
| NR\_041380.1 | P. anaericanus Gsoil 1638 16S ribosomal RNA | 2313 | 96% | 94.60% |

Table S2. Putative gene clusters coding for secondary metabolites in *B. vitellinus* NBRC 13296 derived using BAGEL 4.0.

|  |  |  |  |
| --- | --- | --- | --- |
| **AOI** | **Class** | **Start** | **End** |
| Contig1.0.AOI\_01 | LAPs | 333668 | 354376 |
| Contig1.0.AOI\_02 | Sactipeptides | 5883692 | 5903692 |

Table S3. Putative genetic organization of contig1.0.AOI\_02 derived using BAGEL 4.0. and BLASTP packages.

|  |  |
| --- | --- |
| **ORFs** | **Putative Functions of contig1.0.AOI\_02** |
| 5332 | Carbon-nitrogen hydrolase family protein |
| 5333 | Class I SAM-dependent methyltransferase |
| 5334 | MerR family transcriptional regulator |
| 5335 | MFS transporter |
| 5336 | UDP-glucose--hexose-1-phosphate uridylyltransferase |
| 5337 | UDP-glucose 4-epimerase GalE |
| 5338 | Galactokinase |
| 5339 | AraC family transcriptional regulator |
| 5340 | Prolipoprotein diacylglyceryl transferase |
| 5341 | Radical SAM protein |
| 5342 | Hypothetical protein |
| 5343 | Hypothetical protein |
| 5344 | Hypothetical protein |
| 5345 | ABC transporter permease |
| 5346 | ABC transporter permease |
| 5347 | ABC transporter ATP-binding protein |
| 5348 | BMP family ABC transporter substrate-binding protein |
| 5349 | Threonine ammonia-lyase IlvA |

Table S4. Putative gene clusters coding for secondary metabolites in *B. vitellinus* NBRC 13296 derived using PRISM 4.4.5.

|  |  |
| --- | --- |
| **Region** | **Type** |
| Cluster 1 | Nonribosomal peptide |
| Cluster 2 | Unknown thiotemplated cluster type |
| Cluster 3 | Polyketide/nonribosomal peptide |
| Cluster 4 | Polyketide/nonribosomal peptide |
| Cluster 5 | Polyketide/nonribosomal peptide |
| Cluster 6 | Phosphonate |
| Cluster 7 | 2-Deoxy-streptamine derived aminoglycoside |
| Cluster 8 | Nonribosomal peptide |
| Cluster 9 | Nonribosomal peptide |
| Cluster 10 | Polyketide |
| Cluster 11 | Nonribosomal peptide |

Table S5. Genetic organization and sequence homology of butirosin BGCs derived from *B. vitellinus* NBRC 13296 and *N. circulans* ATCC 21558.

|  |  |  |  |
| --- | --- | --- | --- |
| ***B. vitellinus* NBRC 13296** | **Amino acids** | **Putative Function** | ***N. circulans*** **ATCC 21558 Identity/Similarity (%)** |
| ORF 3151 | 421 | RNA polymerase sigma factor | 88.73%/92.72% |
| ORF 3152 | 144 | YciI family protein  | 90.97%/93.75% |
| ORF 3153 (BtrZ) | 279 | Alpha/beta hydrolase | 81.36%/89.25% |
| ORF 3154 (BtrY) | 422 | MFS transporter | 83.45%/90.54% |
| ORF 3155 (BtrX) | 673 | ABC transporter ATP-binding protein/permease | 87.24%/94.36% |
| ORF 3156 (BtrW) | 582 | ABC transporter transmembrane domain-containing protein | 85.67%/93.52% |
| ORF 3157 (BtrQ) | 504 | GMC family oxidoreductase | 92.87%/98.22% |
| ORF 3158 (BtrV) | 82 | 4-(gamma-L-glutamylamino)butanoyl-ACP monooxygenase DUF6137 domain-containing protein | 80.23%/88.37% |
| ORF 3159 (BtrP) | 213 | 5''-Phosphoribostamycin phosphatase Histidine phosphatase family protein | 82.16%/89.67% |
| ORF 3160 (BtrO) | 338 | 4-(gamma-L-glutamylamino)butanoyl-[BtrI acyl-carrier protein] monooxygenaseLLM class flavin-dependent oxidoreductase | 86.73%/91.74% |
| ORF 3161 (BtrK) | 428 | L-glutamyl-[BtrI acyl-carrier protein] decarboxylaseType III PLP-dependent enzyme  | 82.94%/92.06% |
| ORF 3162 (BtrJ) | 419 | [Butirosin acyl-carrier protein]--L-glutamate ligaseATP-grasp domain-containing protein  | 75.53%/85.75% |
| ORF 3163 (BtrI) | 87 | Acyl carrier protein | 87.36%/94.25% |
| ORF 3164 (BtrH) | 302 | Ribostamycin:4-(gamma-L-glutamylamino)-(S)-2-hydroxybutanoyl-[BtrI acyl-carrier protein] 4-(gamma-L-glutamylamino)-(S)-2-hydroxybutanoate transferaseBtrH N-terminal domain-containing protein | 74.03%/83.44% |
| ORF 3165 (BtrG) | 139 | Gamma-L-glutamyl-butirosin B gamma-glutamyl cyclotransferaseBranched-chain alpha-keto acid dehydrogenase | 76.60%/84.40% |
| ORF 3166 (BtrF) | 232 | Xylostasin dehydrogenaseSDR family oxidoreductase | 81.20%/89.74% |
| ORF 3167 (BtrE) | 349 | Ribostamycin dehydrogenaseZinc-binding alcohol dehydrogenase  | 92.29%/95.71% |
| ORF 3168 (BtrD) | 275 | 2'-N-acetylparomamine deacetylasePIG-L family deacetylase | 77.70%/85.97% |
| ORF 3169 (btrC) | 368 | 2-deoxy-scyllo-inosose synthase | 90.54%/96.49% |
| ORF 3170 (BtrB) | 432 | Neamine transaminaseAminotransferase class III-fold pyridoxal phosphate-dependent enzyme | 85.88%/94.68% |
| ORF 3171 (BtrA) | 1224 | Hypothetical protein | 77.80%/87.09% |
| ORF 3172 (BtrL) | 604 | Phosphoribostamycin synthaseNeamine phosphoribosyltransferase | 67.81%/82.19% |
| ORF 3173 (BtrM) | 412 | 2-Deoxystreptamine N-acetyl-D-glucosaminyltransferaseGlycosyltransferase family 4 protein | 91.75%/95.87% |
| ORF 3174 (BtrN) | 250 | S-adenosyl-L-methionine-dependent 2-deoxy-scyllo-inosamine dehydrogenaseSPASM domain-containing protein | 92.80%/98.00% |
| ORF 3175 (BtrS) | 418 | L-glutamine:2-deoxy-scyllo-inosose aminotransferaseDegT/DnrJ/EryC1/StrS family aminotransferase | 86.84%/94.26% |
| ORF 3176 (BtrT) | 56 | Hypothetical protein  |  |
| ORF 3177 (BtrR2) | 353 | Suppressor of fused domain protein  | 81.87%/89.80% |
| ORF 3178 (BtrU) | 205 | NAD(P)H-dependent oxidoreductase | 81.46%/87.32% |
| ORF 3179 (BtrR1) | 217 | TetR/AcrR family transcriptional regulaton | 84.02%/92.24% |
| ORF 3180 (BtrT2) | 334 | Iron-hydroxamate ABC transporter substrate-binding protein |  |
| ORF 3181 (BtrT3) | 181 | DUF6022 family protein | 78.69%/89.62% |
| ORF 3182 (BtrT4) | 188 | NADAR family protein  | 82.45%/90.43% |
| ORF 3183 (BtrT5) | 296 | NUDIX hydrolase |  |
| ORF 3184 (BtrT6) | 194 | YdeI/OmpD-associated family protein |  |
| ORF 3185 (BtrT7) | 188 | UbiX family flavin prenyltransferase |  |
| ORF 3186  | 497 | UbiD family decarboxylase | 　 |

**Figure S1**



Figure S1. Phylogenetic tree constructed using Maximum-likelihood methods based on the complete 16S rRNA gene sequence, extracted from the assembled genome. The relationship among closely related species of Type cultures from the genus *Paenibacillus* are illustrated. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches in decimal form. The analysis, performed using MEGA 11, demonstrates *Paenibacillus chitinolyticus* KCCM 41400T as the closest neighbor.

**Figure S2.**



Figure S2. Distribution of secondary metabolite gene clusters (1–18) in *B. vitellinus* NBRC 13296, as predicted by the antiSMASH package. Details of the predicted BGCs (1–18) are provided in Table 1.

**Figure S3.**



Figure S3. Genetic organization of BGC region 15 as predicted by the antiSMASH package. Region 15 contains two-copies (a and b) of an opine-like metallophore gene cluster, consisting of opine metallophore biosynthesis dehydrogenase, nicotianamine synthase family protein, and a nickel/cobalt ABC transporter substrate-binding protein/permease.