

Supplementary Information

Composite genome sequence of *Bacillus clausii*, a probiotic commercially available as Enterogermina®, and insights into its probiotic properties

Indu Khatri^{1,2*}, Gaurav Sharma^{1,3*}, Srikrishna Subramanian^{1#}

* Equal Contribution

Corresponding author

Affiliations:

¹ CSIR-Institute of Microbial Technology, Sector-39A, Chandigarh, India

² Leiden University Medical Center, Leiden, the Netherlands

³ Institute of Bioinformatics and Applied Biotechnology, Bengaluru, Karnataka, India

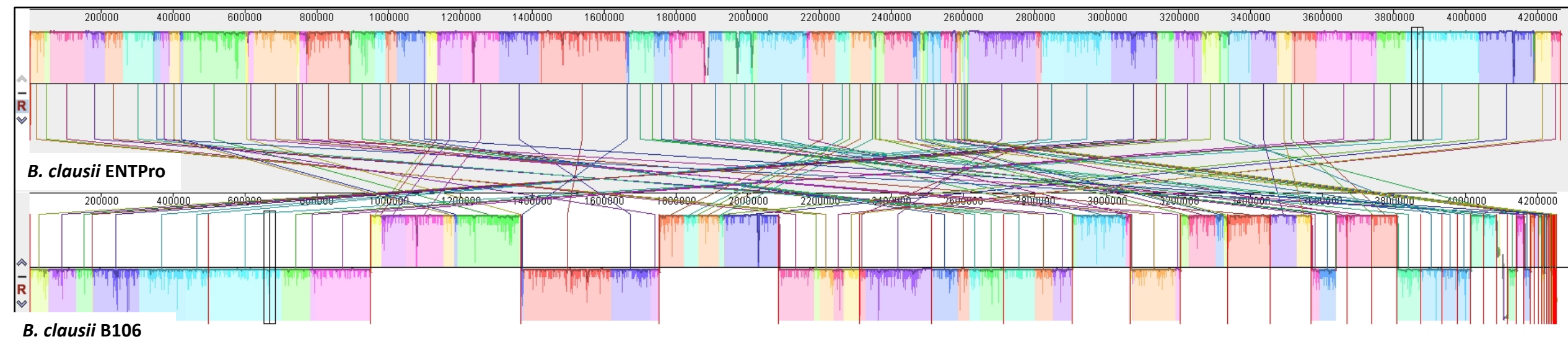
Contact Information:

Srikrishna Subramanian
Institute of Microbial Technology-CSIR
Sector-39A, Chandigarh
India-160036
Email: krishna@imtech.res.in
Telephone: +911726665483

Figure S1: Mauve alignment of *B. clausii* ENTPro with its closer genomes. A) *B. clausii* ENTPro with *B. clausii* B106

B) *B. clausii* ENTPro with *B. clausii* KSM K-16.

A



B

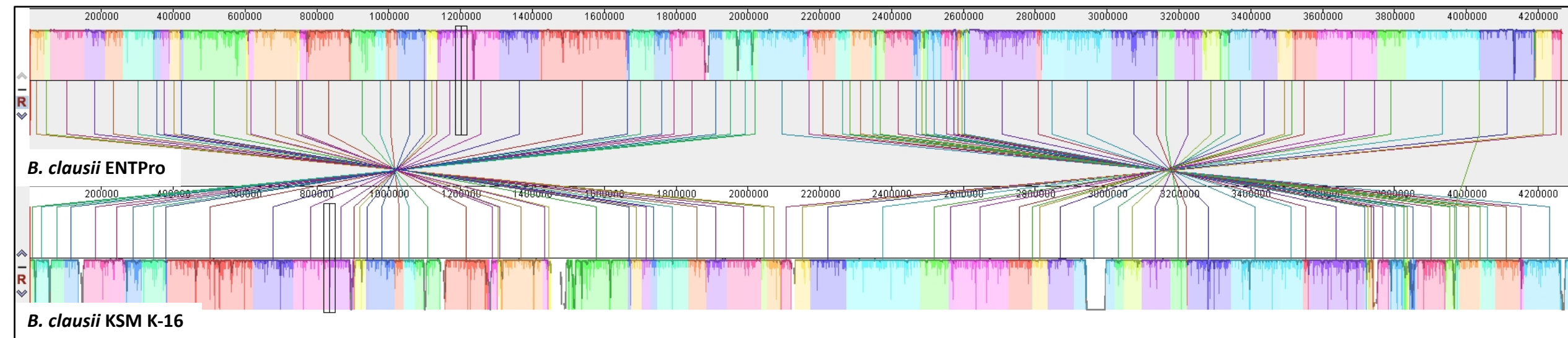


Figure S2: The 16S rRNA alignment of operons identified in *B. clausii* ENTPro and its parental strains. The V-region are identified using V-Xtractor v.2.1 67. The V regions are marked by red bars and mutated positions are marked in pink color.

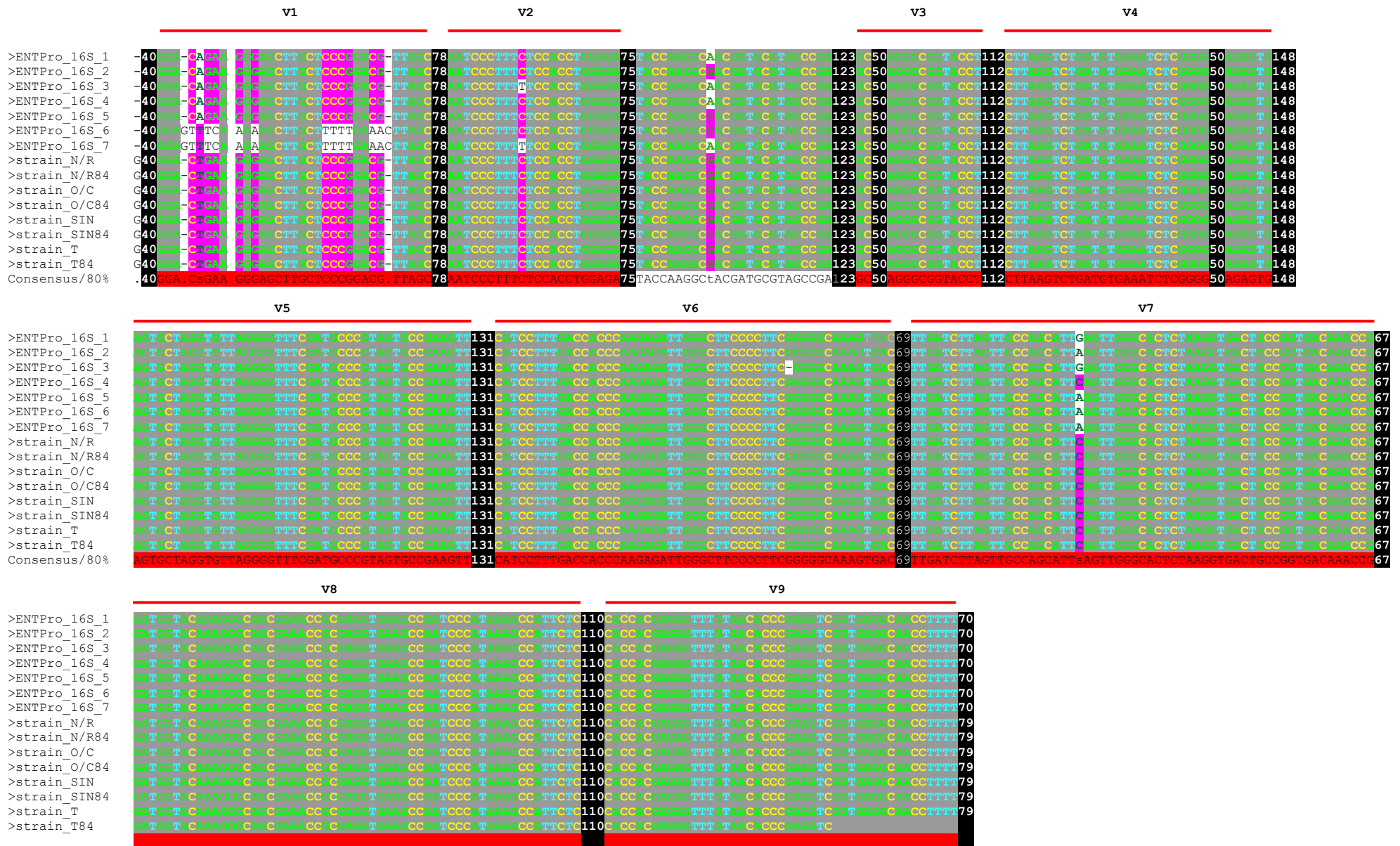


Figure S3: RpoB sequence alignment between *E. coli* and *B. clausii* genomes.

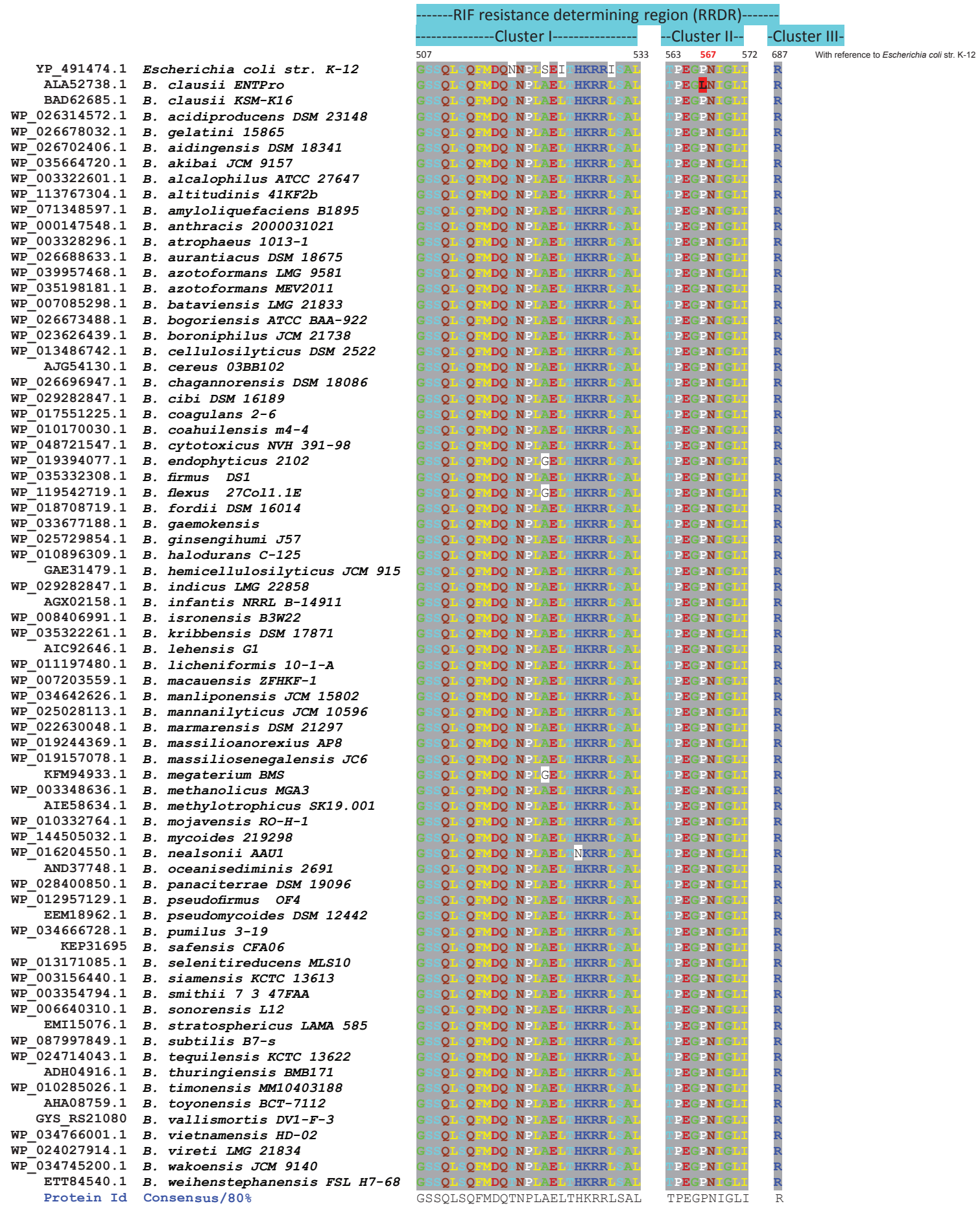


Figure S4: The Streptomycin synthesis pathway components present in *B. clausii*. Green box marks the presence of enzymes present in *B. clausii*.

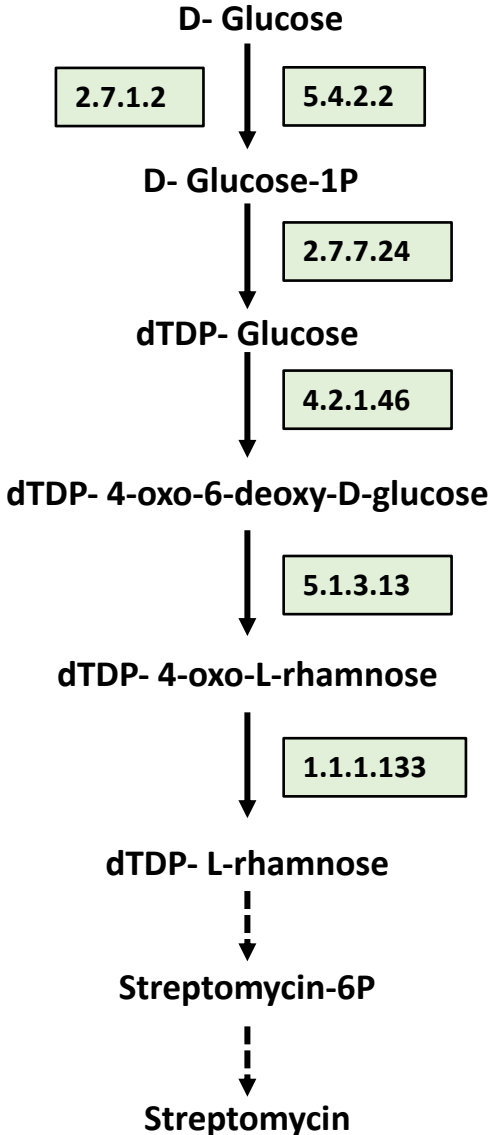


Figure S5: The Vancomycin resistance pathway components present in *B. toyonensis* BCT-7112.
Green box marks the presence of gene whereas the not filled box marks the absence.

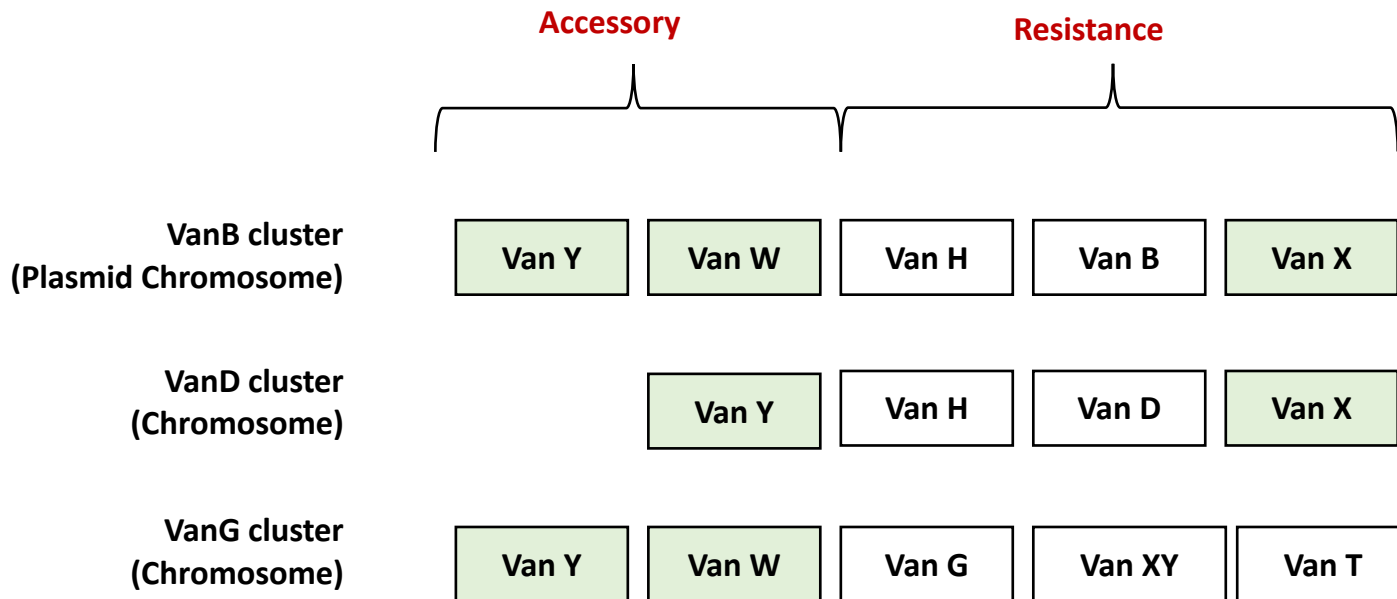


Table S1: Genome Similarity of *B. clausii* ENTPro with other *B. clausii* genomes. Formula I, II and III represents different methods used by GGDC to calculate the similarities.

| Query Genome (GCA_000737305.2) | Reference Genome | Assembly reference | Formula I | Formula II | Formula III |
|------------------------------------|----------------------------|--------------------|-----------|------------|-------------|
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7520-2 | GCA_002272085.1 | 85.6 | 62.4 | 84 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7522 | GCA_002272655.1 | 87.4 | 63.1 | 85.6 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7523-2 | GCA_002272645.1 | 76.9 | 62.8 | 76.7 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7529 | GCA_002272615.1 | 92.7 | 92.8 | 94.9 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7535-K | GCA_002272605.1 | 85.8 | 62.7 | 84.2 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7537-T | GCA_002272525.1 | 89.6 | 92.1 | 92.6 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7538 | GCA_002272545.1 | 93.3 | 93.5 | 95.4 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7539 | GCA_002272535.1 | 86.4 | 62.7 | 84.7 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7540-1 | GCA_002271585.1 | 87 | 63.4 | 85.3 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7540-2 | GCA_002271575.1 | 95.5 | 93.1 | 96.9 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7541 | GCA_002271555.1 | 94.5 | 92.4 | 96.1 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7543 | GCA_002272505.1 | 86.9 | 63.4 | 85.3 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7547-G | GCA_002272445.1 | 93.5 | 93.4 | 95.5 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> 7894-1 | GCA_002271695.1 | 86.7 | 63.4 | 85.1 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> AKU0647 | GCA_001598215.1 | 75.9 | 50.7 | 72.4 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> B106 | GCA_002266625.1 | 99.9 | 99.8 | 100 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> DSM-8716 | GCA_002250115.1 | 86.2 | 62.6 | 84.5 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> KSM-K16 | GCA_000009825.1 | 95.2 | 94.3 | 96.8 |
| <i>B. clausii</i> ENTPro | <i>B. clausii</i> UBBC-07 | GCA_000981315.1 | 93.1 | 92 | 95.1 |

Table S2: Distribution of proteins involved in probiotic properties in *B. clausii* ENTPro genome.

| | Probiotic nature | Accession Numbers | Domain-tag | Domains-identified | Domain-wise E value (hmmscan) |
|--|-----------------------------|--------------------------|-------------------|---------------------------|--------------------------------------|
| Environmental stress resistance | Chaperonins GroEL | ALA51946 | PF00118 | Cpn60_TCP1 | 4.30E-104 |
| | Chaperonins GroES | ALA51947 | PF00166 | Cpn10 | 7.30E-37 |
| | Clp protease | ALA52850 | PF00574 | CLP_protease | 2.90E-79 |
| | | ALA54099 | PF00574 | CLP_protease | 3.40E-89 |
| | | ALA54894 | PF00574 | CLP_protease | 5.10E-36 |
| | Cold shock protein | ALA52002 | PF00313 | CSD | 1.10E-31 |
| | | ALA54053 | PF00313 | CSD | 3.50E-33 |
| | Heat shock protein | ALA52772 | PF01430 | HSP33 | 4.20E-90 |
| | Heat resistance | ALA51177 | PF01025 | GrpE | 6.20E-52 |
| | Hyperosmotic stress | ALA51175 | PF00226 | DnaJ | 4.40E-25 |
| | Oxidative stress | ALA51081 | PF01625 | PMSR | 4.20E-60 |
| | Universal Stress | ALA54975 | PF01625 | PMSR | 2.50E-60 |
| | | ALA52046 | PF00582 | Usp | 1.20E-07 |
| | | ALA52818 | PF00582 | Usp | 4.20E-06 |
| ALA53908 | PF00582 | Usp | 3.20E-32 | | |
| Adhesion to gut mucosa | Mucus binding protein | ALA51653 | PF00746 | Gram_pos_anchor | 9.10E-05 |
| | Collagen binding protein | ALA53512 | PF05737 | Collagen_bind | 1.60E-30 |
| | Fibronectin binding protein | ALA54776 | PF05833 | FbpA | 1.10E-166 |
| | Sortase | ALA51654 | PF04203 | Sortase | 4.50E-24 |
| | | ALA51828 | PF04203 | Sortase | 1.00E-26 |
| | | ALA53120 | PF04203 | Sortase | 4.50E-28 |
| | | ALA53513 | PF04203 | Sortase | 5.60E-26 |
| | | ALA53680 | PF04203 | Sortase | 1.50E-22 |
| | Flagellin | ALA53337 | PF00669 | Flagellin_N | 1.90E-40 |
| | | ALA54045 | PF00669 | Flagellin_N | 4.90E-16 |
| Triosephosphate isomerase | ALA54107 | PF00121 | TIM | 3.00E-95 | |

| | | | | | |
|-----------------------|-----------------------------------|----------|---------|---------------|-----------|
| Bile tolerance | Ornithine decarboxylase | ALA52835 | PF01276 | OKR_DC_1 | 1.60E-60 |
| | | ALA54693 | PF01276 | OKR_DC_1 | 1.90E-101 |
| | Sodium bile acid symporter family | ALA51223 | PF01758 | SBF | 1.10E-14 |
| | | ALA51591 | PF01758 | SBF | 1.40E-41 |
| | | ALA52219 | PF01758 | SBF | 3.80E-25 |
| Acid tolerance | F0F1 ATP_synthase | ALA53170 | PF00119 | ATP-synt_A | 3.20E-49 |
| | | ALA53171 | PF00137 | ATP-synt_C | 7.10E-14 |
| | | ALA53172 | PF00430 | ATP-synt_B | 3.10E-40 |
| | | ALA53173 | PF00213 | OSCP | 1.70E-49 |
| | | ALA53174 | PF00006 | ATP-synt_ab | 2.10E-73 |
| | | ALA53175 | PF00231 | ATP-synt | 1.20E-91 |
| | | ALA53176 | PF00006 | ATP-synt_ab | 1.70E-68 |
| | | ALA53177 | PF02823 | ATP-synt DE N | 1.40E-25 |

Table S3: Proteins and domains involved in conferring antibiotic resistance in *B. clausii* ENTPro.

| Resistance to | Accession Numbers | Domain-tag | Domains-identified | Domain-wise E value (hmmScan) | Homolog in KSM-K16 (% identity) |
|-----------------------|-------------------|------------|--------------------|-------------------------------|---------------------------------|
| Aminoglycosidase | ALA51464 | PF01636 | APH | 8.1E-19 | ABC_RS07395 (97.31) |
| Aminoglycosidase | ALA51859 | PF01636 | APH | 9.9E-25 | ABC_RS05095 (100) |
| Aminoglycosidase | ALA51976 | PF04439 | Adenyl_transf | 1.8E-117 | ABC_RS04555 (99.65) |
| Aminoglycosidase | ALA52626 | PF07827 | KNTase_C | 2E-62 | ABC_RS01520 (96.09) |
| Penicillin repressor | ALA52627 | PF03965 | Penicillinase_R | 1.2E-34 | ABC_RS01515 (98.41) |
| Penicillin regulatory | ALA52628 | PF05569 | Peptidase_M56 | 6.2E-57 | ABC_RS01510 (94.35) |
| Penicillin | ALA52629 | PF13354 | Beta-lactamase2 | 8.4E-34 | ABC_RS01505 (91.69) |
| Aminoglycosidase | ALA52671 | PF01636 | APH | 6.6E-22 | ABC_RS01175 (98.63) |
| Rifampicin | ALA52738 | PF00562 | RNA_pol_Rpb2_6 | 5E-167 | ABC_RS00825 (100) |
| Tetracycline | ALA53309 | cd00881 | TetM_like | 1.07E-131 | ABC_RS19270 (98.92) |
| Aminoglycosidase | ALA53421 | PF04439 | Adenyl_transf | 1.7E-124 | ABC_RS18770 (97.23) |
| Aminoglycosidase | ALA53555 | PF01636 | APH | 9.5E-21 | ABC_RS18315 (97.8) |
| Erythromycin | ALA53582 | PF00398 | RrnaAD | 1.7E-51 | ABC_RS18170 (99.64) |
| Aminoglycosidase | ALA54163 | PF01636 | APH | 2.5E-27 | ABC_RS15455 (99.65) |
| Aminoglycosidase | ALA54297 | PF01636 | APH | 2.1E-19 | ABC_RS14455 (99.62) |
| Chloramphenicol | ALA55212 | PF00302 | CAT | 3E-78 | ABC_RS10335 (98.6) |

Table S4: Accession numbers of the genomes used to generate the phylogenies.

| Genome | GenBank assembly accession |
|---|-----------------------------------|
| <i>B. altitudinis</i> SGAir0031 | GCA_002443015.2 |
| <i>B. amyloliquefaciens</i> B1895 | GCA_000696285.1 |
| <i>B. amyloliquefaciens</i> H57 | GCA_001617995.1 |
| <i>B. amyloliquefaciens</i> -DSM7 | GCA_000196735.1 |
| <i>B. anthracis</i> str. Ames | GCA_000007845.1 |
| <i>B. anthracis</i> str. Sterne | GCA_000008165.1 |
| <i>B. atrophaeus</i> SRCM101359 | GCA_002173495.1 |
| <i>B. beveridgei</i> MLTeJB | GCA_001721685.1 |
| <i>B. bombysepticus</i> str. Wang | GCA_000831065.1 |
| <i>B. cellulolyticus</i> DSM2522 | GCA_000177235.2 |
| <i>B. cereus</i> ATCC 14579 | GCA_000007825.1 |
| <i>B. coagulans</i> GB-I30 | GCA_000756285.1 |
| <i>B. coagulans</i> HM-08 | GCA_000876545.1 |
| <i>B. coagulans</i> S-lac | GCA_001039495.1 |
| <i>B. coagulans</i> Unique IS-2 | GCA_001578455.1 |
| <i>B. coagulans</i> DSM1 | GCA_000832905.1 |
| <i>B. cohnii</i> DSM-6307 | GCA_001591425.1 |
| <i>B. cytotoxicus</i> NVH 391-98 | GCA_000017425.1 |
| <i>B. endophyticus (filamentous)</i> Hbe603 | GCA_000972245.3 |
| <i>B. flexus</i> KLBMP 4941 | GCA_002024265.1 |
| <i>B. gibsonii</i> FJAT-10019 | GCA_001719185.1 |
| <i>B. glycinifermentans</i> BGLY | GCA_900093775.1 |
| <i>B. gobiensis</i> FJAT-4402 | GCA_001278705.1 |
| <i>B. halodurans</i> C-125 | GCA_000011145.1 |
| <i>B. horikoshii</i> 20a | GCA_002157855.1 |
| <i>B. infantis</i> NRRL-B-14911 | GCA_000473245.1 |
| <i>B. kochii</i> BDGP4 | GCA_002276165.1 |
| <i>B. krulwichiae</i> AM31D | GCA_002109385.1 |
| <i>B. lehensis</i> G1 | GCA_000706725.1 |
| <i>B. licheniformis</i> ATCC 14580 | GCA_000011645.1 |
| <i>B. megaterium</i> NBRC 15308 | GCA_000832985.1 |
| <i>B. methanolicus</i> MGA3 | GCA_000724485.1 |
| <i>B. muralis</i> G25-68 | GCA_001645685.2 |
| <i>B. mycoides</i> ATCC 6462 | GCA_000832605.1 |
| <i>B. oceanisediminis</i> 2691 | GCA_000294775.2 |

| | |
|---|-----------------|
| <i>B. paralicheniformis</i> BL09 | GCA_000876525.1 |
| <i>B. paralicheniformis</i> 14DA11 | GCA_002393225.1 |
| <i>B. paralicheniformis</i> ATCC 9945a | GCA_000408885.1 |
| <i>B. pseudofirmus</i> OF4 | GCA_000005825.2 |
| <i>B. pseudomycooides</i> DSM 12442 | GCA_000161455.1 |
| <i>B. pumilus</i> RI06-95 | GCA_001183525.1 |
| <i>B. pumilus</i> SH-B9 | GCA_001578205.1 |
| <i>B. safensis</i> KCTC 12796BP | GCA_001895885.1 |
| <i>B. simplex</i> SH-B26 | GCA_001578185.1 |
| <i>B. smithii</i> DSM4216 | GCA_001050115.1 |
| <i>B. sonorensis</i> SRCM101395 | GCA_002202015.1 |
| <i>B. subtilis</i> CU1 | GCA_002093785.1 |
| <i>B. subtilis</i> KATMIRA1933 | GCA_000691185.1 |
| <i>B. subtilis</i> TO-A-JPC | GCA_001037985.1 |
| <i>B. subtilis</i> 168 | GCA_000009045.1 |
| <i>B. thuringiensis</i> YBT-1518 | GCA_000497525.2 |
| <i>B. thuringiensis</i> serovar konkukian str.97-27 | GCA_000008505.1 |
| <i>B. toyonensis</i> BCT-7112 | GCA_000496285.1 |
| <i>B. vallismortis</i> NBIF-001 | GCA_002113805.1 |
| <i>B. velezensis</i> AS43.3 | GCA_000319475.1 |
| <i>B. velezensis</i> SB1216 | GCA_001634675.1 |
| <i>B. velezensis</i> SRCM10072 | GCA_002201935.1 |
| <i>B. weihaiensis</i> Alg07 | GCA_001889165.1 |
| <i>B. xiamenensis</i> VV3 | GCA_001857925.1 |
| <i>B. kaustophilus</i> HTA426 | GCA_000009785.1 |
| <i>L. innocua</i> Clip11262 | GCA_000195795.1 |
| <i>L. ivanovii</i> WSLC3009 | GCA_000565155.1 |
| <i>L. marthii</i> FSL S4-120 | GCA_000183865.1 |
| <i>L. monocytogenes</i> EGD-e | GCA_000196035.1 |
| <i>L. seeligeri</i> serovar 1/2b str. SLCC3954 | GCA_000027145.1 |
| <i>L. welshimeri</i> NCTC11857 | GCA_900187315.1 |
| <i>B. clausii</i> 7520-2 | GCA_002272085.1 |
| <i>B. clausii</i> 7522 | GCA_002272655.1 |
| <i>B. clausii</i> 7523-2 | GCA_002272645.1 |
| <i>B. clausii</i> 7529 | GCA_002272615.1 |
| <i>B. clausii</i> 7535-K | GCA_002272605.1 |
| <i>B. clausii</i> 7537-T | GCA_002272525.1 |
| <i>B. clausii</i> 7538 | GCA_002272545.1 |
| <i>B. clausii</i> 7539 | GCA_002272535.1 |

| | |
|----------------------------|-----------------|
| <i>B. clausii</i> 7540-1 | GCA_002271585.1 |
| <i>B. clausii</i> 7540-2 | GCA_002271575.1 |
| <i>B. clausii</i> 7541 | GCA_002271555.1 |
| <i>B. clausii</i> 7543 | GCA_002272505.1 |
| <i>B. clausii</i> 7547-G | GCA_002272445.1 |
| <i>B. clausii</i> 7894-1 | GCA_002271695.1 |
| <i>B. clausii</i> AKU0647 | GCA_001598215.1 |
| <i>B. clausii</i> B106 | GCA_002266625.1 |
| <i>B. clausii</i> DSM-8716 | GCA_002250115.1 |
| <i>B. clausii</i> KSM-K16 | GCA_000009825.1 |
| <i>B. clausii</i> UBBC-07 | GCA_000981315.1 |
| <i>B. clausii</i> ENTPro | GCA_000737305.2 |