Fig. S1. An example image of a read depth change.

The positions showing drastic change of depth were regarded as transcriptional start sites (TSSs).
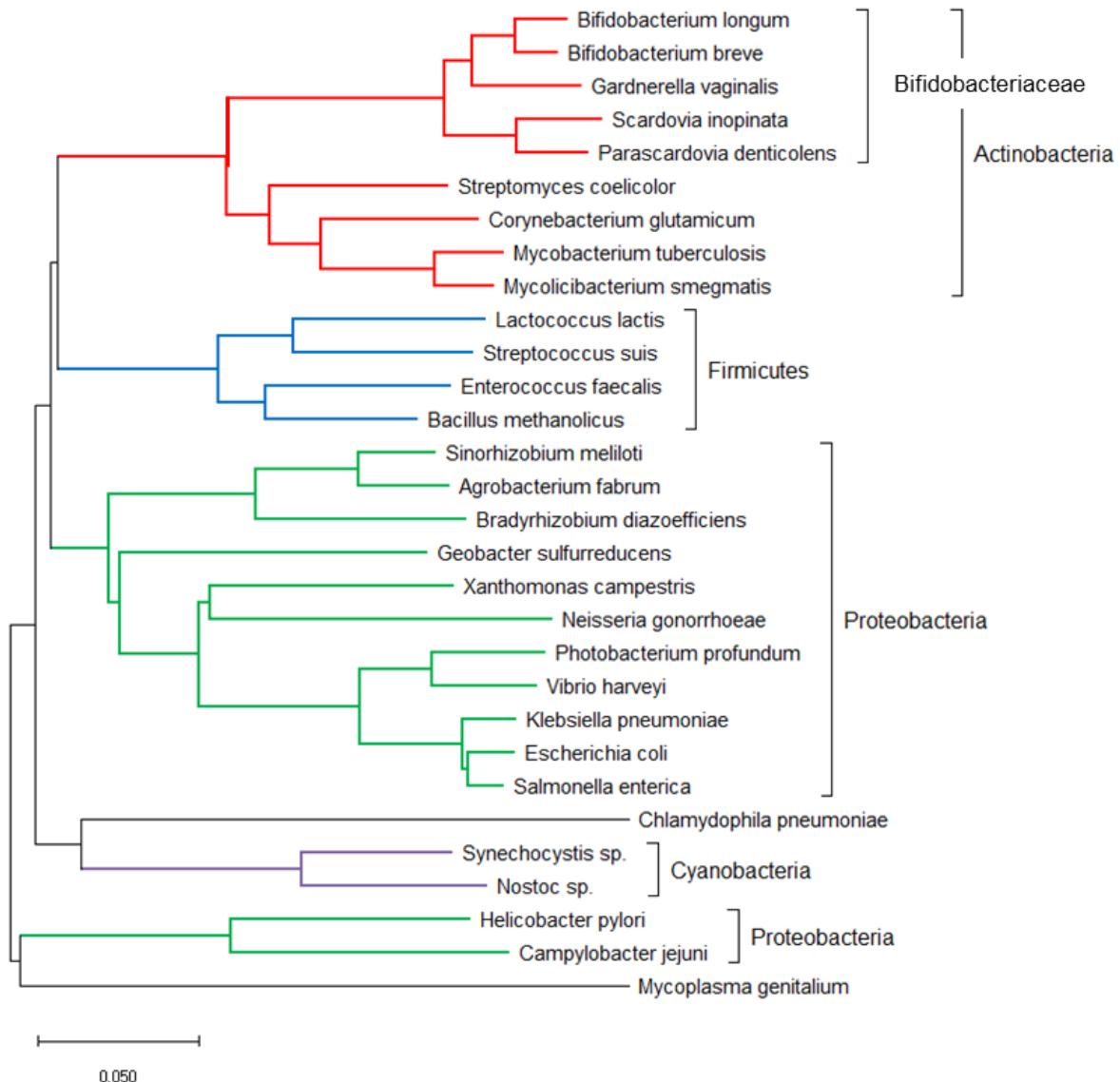
Fig. S2. The phylogenetic tree based on 16S ribosomal RNA gene sequences.

The phylogenetic classification of bacteria utilized in Table 1. Four bacteria in the phylum Actinobacteria; *C. glutamicum* ATCC 13032, *M. smegmatis* MC² 155, *M. tuberculosis* H37Rv, and *S. coelicolor* A3(2). Four bacteria in the phylum Firmicutes; *B. methanolicus* MGA3, *E. faecalis* V583, *S. suis* P1/7, and *L. lactis subsp. cremoris* MG1363. Two bacteria in the phylum Cyanobacteria; *Nostoc* sp. PCC 7120 and *Synechocystis sp.* PCC 6803. Thirteen bacteria in the phylum Proteobacteria; G. sulfurreducens PCA, *C. jejuni subsp. jejuni* NCTC 11168, *H. pylori* 26695, *B. japonicum* USDA110, *A. fabrum* C58, *S. meliloti* 1021, *S. enterica* serovar Typhimurium str. SL1344, *K. pneumoniae subsp. pneumoniae* MGH 78578, *E. coli* K-12 MG1655, *V. harveyi* FDAARGOS_107, *P. profundum* SS9, *X. campestris pv. vesicatoria* 85-10, and *N. gonorrhoeae* MS11. One bacterium in the phylum Chlamydiae; *C. pneumoniae* CWL029.
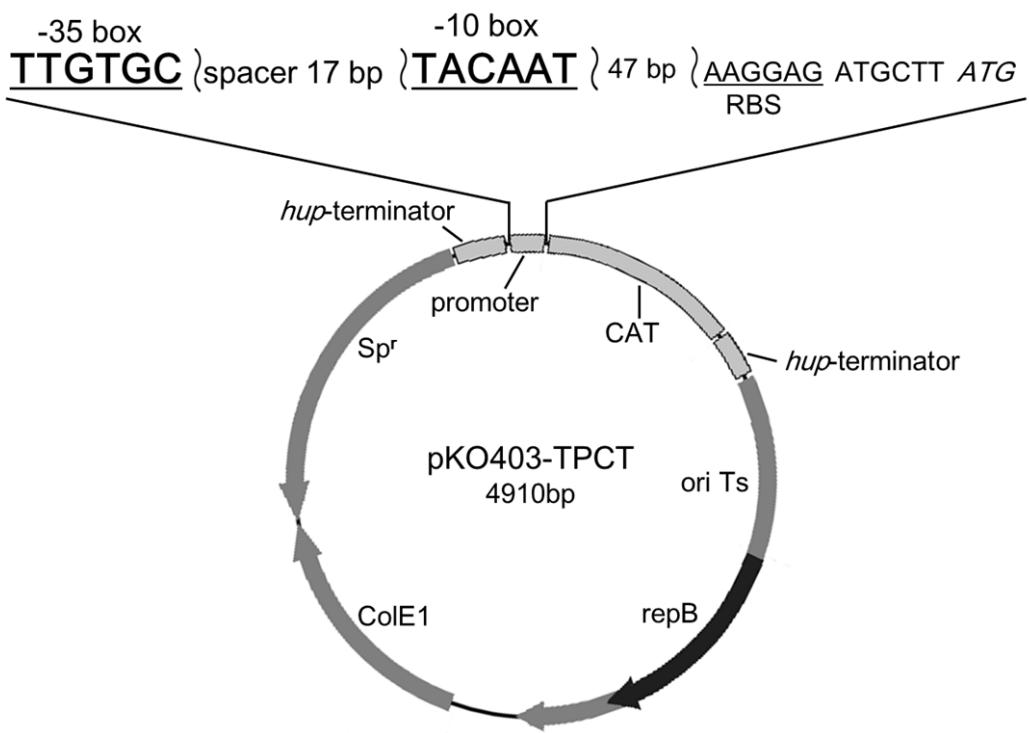
Fig. S3. The plasmid map of the pKO403-TPCTcon.

The spacer sequence between the –35 and –10 boxes is in the upstream region of the CAT gene. AAGGAG was designed as the RBS and the 17 bp-long sequences as the spacers between the –35 and –10 boxes. *hup*-terminator; terminator of a *hup* gene in *Bifidobacterium longum* NCC2705 strain. *Sp^r*; a resistance gene of spectinomycin. *ColE1*; a replication origin for *Escherichia coli. repB*; a replication origin for *B. longum. ori Ts*; a temperature-sensitive replication origin. *CAT*; a gene coding region of chloramphenicol acetyltransferase.

A
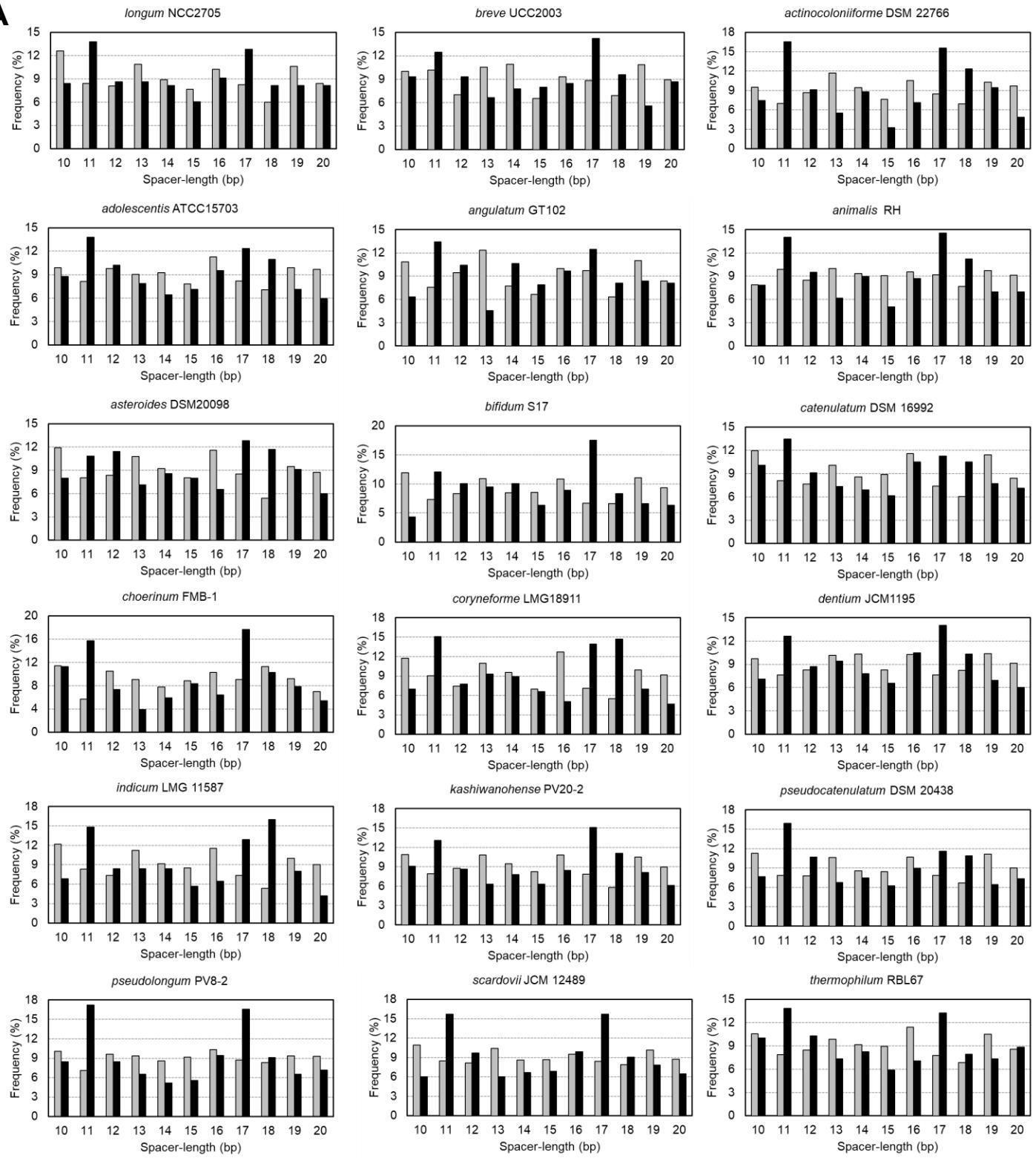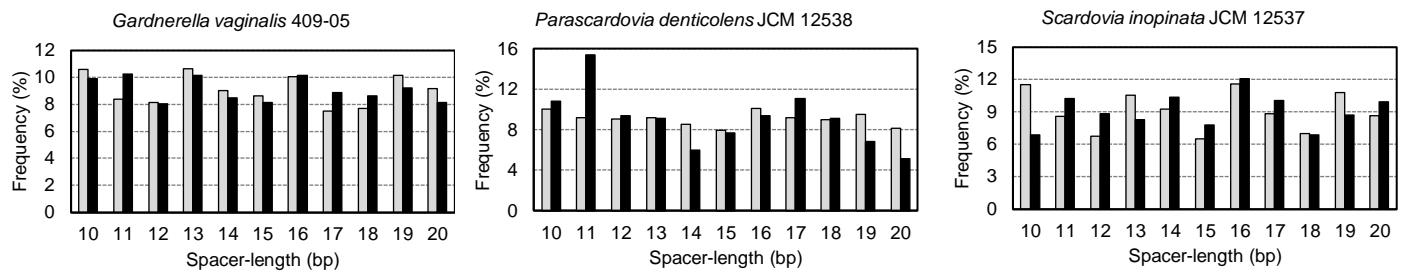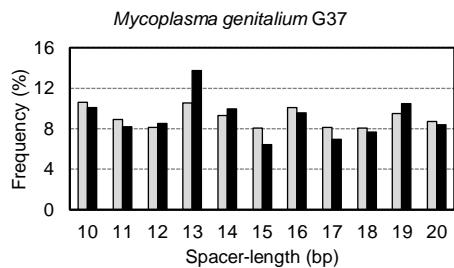
Fig. S4. Comparison of the frequency of spacer-lengths among bacteria.

A; Comparison among the genus *Bifidobacterium*. Two peaks of 11 and 17 bp was shown in the genus *Bifidobacterium*. Fig. 2D is based on these values. B; Comparison among the family Bifidobacteriaceae. Two peaks of 11 and 17 bp was shown in *Gardnerella*, *Parascardovia*, and *Scardovia*. C; the frequency of spacer-length in Mycoplasma, showing a specific peak in 13 bp. Gray and black bars represent frequencies in other regions and regions 150 bp upstream of genes.

```
Domain 2_Alternative sigma   K R A R F E Q L A M P A V N A L Y R Q A M R L T N D P D D A Q D L V Q D T - - - F E R G F K A F D S F Q P G S N F E A W
Domain 2_primary sigma       - - K A K D H L L E A N L R L V V S L A K R Y T G R G M L F L D L I Q E G N L G L I R A V E K F D - W K K G F K F S T Y
                                 : : *       .   : .   :           *     *     * .           * * : * :         :   * . . :   * *   : :     *     : * . : :

Domain 2_Alternative sigma   M T T I E R N A Y F N Q Y A K A K R R P Q R A N D S T G E Y D D W D I Y D A A E H
Domain 2_primary sigma       A T W W I R Q A I T R A M A D Q A R - T I R V P V H M V E V I - - - - - - - - - -
                                 *           * : *       .       * .       *       . * .               *

Domain 4_Alternative sigma   T S D G L K S A E E E Y L D A F A P E E I M A A L A K L S P E R R Q V F I D A A I D G K S Y Q Q - V A D E Q G V K I G T
Domain 4_primary sigma       P S D A V A F S - - - - - - - L L Q E Q F K Q V L E T L S P - R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V
                               . * * . :   :             :       * : :     . *   . * * *   * .     . *         .     : .   *       .     * :   *         * .

Domain 4_Alternative sigma   V M S R L N R A R T Q L K R E L A S Y A K
Domain 4_primary sigma       T R E R I R - - - - Q I E S K T M S K L R
                               .   . * : .           * : :   :       *           :
```

Fig. S5. The alignment of sigma factors of the genus *Bifidobacterium*.

Alternative sigma; alternative sigma factor in *B. longum* NCC2705 (BL1357). Primary sigma; primary sigma factor in *B. longum* NCC2705 (BL1428), corresponding to sigma 70 in *E. coli*. Domain 2 and 4 relate to the recognition of -10 and -35 boxes.

```
longum              - - - - - - - - - - - - - M A T K E T T A T K Q T A E S S - - - - - - - - - E E T E T K K K A S S T T T A R K T S A K K
breve               - - - - - - - - - - - - - M A T K E T T A T K Q T A E S S - - - - - - - - - E E T E T K K K A S S - T T A R K T S A K K
catenulatum         - - - - - - - - - - - - - M A T K E T T A S M E - - - - - - - - - - - - - Q A D L T E K T T S T K K A A T R K A S T K
kashiwanohense      - - - - - - - - - - - - - M A T K E T T A S M E - - - - - - - - - - - - - Q A D L T E K T T S T K K A A T R K A S T K
pseudocatenulatum   - - - - - - - - - - - - - M A T K E T T A S M E - - - - - - - - - - - - - Q A D L T E K T T S T K K A A T R K A S T K
dentium             - - - - - - - - - - - - - M A T K E T T A S T E - - - - - - - - - - - - - Q A D L T D K T T S T K K A P S R K A S T K
adolescentis        M V Y A S G G I N R R I S L A T K E T T V N T E - - - - - - - - - - - - - Q T D L T E K T T S G K K A S T R K P A A K
thermophilum        - - - - - - - - - - - - - - M A - K E T T A S K K - - - - - - - - - - - - - D A D Q T E K T T S T R K S A A K K A A T P
bifidum             - - - - - - - - - - - - M A T K D T M A T T V P D D T D A K T K S E K K G T A A R T A S T A D K K T A A G R K S A A G
scardovii           - - - - - - - - - - - M A K K D A K A T T D - - - - - - - - - - - - - - - - N A E Q L D K A T S T K R K A A S A
angulatum           - - - - - - - - - - - - M A T K S T T A A E N - - - - - - - - - - - - - - - T E Q V Q D E A K A V S A K K A S A T R
animalis            - - - - - - - - - - - M V A K T T S T E K A A K K K A - - - - - - S E S A H D E A A V E E K G K T T R R R A S S K
choerinum           - - - - - - - - - - - M A T K R T A N T K D K S E D K D - - - - - - T S T V E G A S A T K E T S A K P A K R T R - K
pseudolongum        - - - - - - M W R R N S L V T K Q T S N A A K K S E G K T - - - - - S S T K A D E A Q Q E D E T T T A K K P A K - -
indicum             - - - - - - - - - - - M A T K D T K S V K T A E V D A E - - - - - - - - - E A K T P S T S G K D R S Q K G D K A K
coryneforme         - - - - - - - - - - - M A T K D T K S V K T A E V D A E - - - - - - - - - E A K T P S T S G K D R S Q K G D K A K
asteroides          - - - - - - - - - - - M A T K D T K S A K T A A V D T G - - - - - - - - - K G R T S S T A G - - - - - - - - K T K
actinocoloniiforme  - - - - - - - - - - - M A A P A K T K A R K V P S K G G A - - - - - - - - - - K S S S S S T R K A K G S S K S A A S S
                                                                                        :

longum              S K S A E - - P S K K T T R K S A A K K T A K K A K D D E L V K D E D T Q V D D E L D D D A E V - - - - - - - - - - - -
breve               T K D T K G T T A A K T G E K T S R K T S A K K V K K E E E L P Q N D I S Q S D D P E D - - - - - - - - - - - -
catenulatum         K S A A K K T P A T K T P R K T T A K K S A P A K K - E E E Q I E D D V L D D E D A Q - - - - - - - - - - - -
kashiwanohense      K S A A K K T P A T K T P R K T T A K K S A P A K K - E E E Q I E N D A L D D E D A Q - - - - - - - - - - - -
pseudocatenulatum   K S A A K K T T A T K T P R K T T A K K S A P A K K - E E P - L E D D A L D D E D A Q - - - - - - - - - - - -
dentium             K A A D K K T P A T K T P R K T T A K K K A P A K R Q E E P I E D E D T L D G E E D Q - - - - - - - - - - - -
adolescentis        K S E S S K T - A T K T S R K T A A K K T S A A K K S K K A E E E E E N L I E Q D A P E D - - - - - - - - - - - -
thermophilum        A K A T K R S A T S A S A K K T A A A K S A K T S K A A K A S G G E S A A P K A K K K A S R T A T A T K T - - - - - - -
bifidum             K K T S S R S A A S K S T V R E S S A K A A A P A K P T R K R A A S K S A K A V K D E T E E M A - - - - - - - - - - - -
scardovii           G R K P S R K G A T K K A A S D T A R K V T A K E D P I T A N D P D V G E D V L D D E - - - - - - - - - - - - - -
angulatum           K T A A A K K T T A K K T A K S G E G T G S V R K K S A A T K S T K N S K T K K R S E E D - - - - - - - - - - - - -
animalis            A S S A K K G S A G K T T K R A T A R K K K D A A E S E S T P A R K T R A A S K T G A K K A I S - - - - - - - - -
choerinum           S T S A K S G - - - - - A R K G T A K K K S A S A Q A E A T K Q P R K S T A K K K T A K K E T T G - - - - - - -
pseudolongum        - - - - - - - - - R T T R K S T A K K T A K A A S - - - - - - Q T K A P K K T T A R K S A S K - - - - - - - - - -
indicum             T G T R A K G K T A T P T K K S A A K K T A A A K A S S N K S G A T A E K E S V E G S S T G K S K S T G A R K S S A K K
coryneforme         T G T R A K G K T A T P T K K S A A K K T A A A K A S S N K S G A T A E K E S V E G S S T G K S K S T G A R K S S A K K
asteroides          T A T R S K T S S S K A S H K T S S R K T A T A S T S T A K - - - - - - - - - - - - - - - - - - - - - - - - - - -
actinocoloniiforme  S K A R S S K T T T K A T V K T K A S K - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
                                      .                                 .

longum              - - - - - - - - - - - - - - - - - - - - - - - - - - - - - N D D D L D V D D L D E D L D D V D E D - - - - - - - - - - - - D
breve               - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - L D V D N L D E D L D D V E S E - - - - - - - - - - - D
catenulatum         - - - - - - - - - - - - - - - - - - - - - - - - - - - - V D D P D L D D F D D D L D D V D S - - - - - - - - - - -
kashiwanohense      - - - - - - - - - - - - - - - - - - - - - - - - - - - - V D D P D L D D F D D D L D D V D S - - - - - - - - - - -
pseudocatenulatum   - - - - - - - - - - - - - - - - - - - - - - - - - - - - V D D A D L D D F D A D L D D V D S - - - - - - - - - - -
dentium             - - - - - - - - - - - - - - - - - - - - - - - - - - - - V D E A D L D D F D D D L D D V D A E - - - - - - - - - - -
adolescentis        - - - - - - - - - - - - - - - - - - - - - - - - - D P V E D A D L D D F D D D L D D V D D E - - - - - - - - - - -
thermophilum        - - - - - - - - - - - - - P R K R T A K R T E P A A A E H D E T Q L D D E D E D I N D E D I E N - - - - - F D D L D E
bifidum             - - - - - - - - - - - - - - - - - - - - - - - K T E E L A N E D S S D D L D D V Q V D E D G L E N L D D D L D D V D D G G L
scardovii           - - - - - - - - - - - - - - - - - - - - - - - D Q D G E A D L D D L E D D E S G L D - - - - - - - - - - - - -
angulatum           - - - - - - - - - - - - - - - - - - - - - - - L D E V D E F D E D A D L D D A K D E D - - - - - - - - - L D
animalis            - - - - - - - - - - - - - - - - - - - - - - - K S E E I D D E D M I G - - - E D V I D E E - - - - D L D L D E D
choerinum           - - - - - - - - - - - - - - - - - - - - - - H A A Q P E D D E D V M L D D D D E Q I D E D - - - - S L D L D E D
pseudolongum        - - - - - - - - - - - - - - - - - - - - - - P K D D E P V D D D E V M I D E E D I D L D D - - - - D L E D V D D
indicum             S A A K T G T T K K T S G K S T S T K A G K T T D - K P V D D A I L E D Q T G E D E F D V T D I D E N G T P E S D E L D
coryneforme         S A A K T G T T K K T S G K S T S T K A G K T T D N K P V D D A I L E D Q T G E D E F D V T D I D E N G T P E S D E L D
asteroides          - - - - - - - - - - - - - - - - - T N R R P A S K T S K K V E E E P I L Q Q E E Q E P D E D L L E D Q E D - - - - D D Q L D
actinocoloniiforme  - - - - - - - - - - - - - - - - - S S E H D E A L E D K A P E E L E E D Q E L D I D D L E E G - - - - - - - - -

longum              A D - - - - - D E D L D E D D A D D D V E D D E D E E E A K S K A P E Q - - - - P K E K G A Y V V S D T D D E E E - N
breve               V D D L E D A D A E S D D E D I D D E D D E D E E D E G K A A S K A P E Q - - - P K E K G A Y V V S D T D D E E E - N
catenulatum         - D D D D D L D D D E E S E P E D D - - - D D E - - E E R K K P E E - - - - - P K E K G A F V V R D D D D D E - - N
kashiwanohense      - D D D D D L D D D E E S E P E D D - - - D D E D E E R K K P E E - - - - - P K E K G A F V V R D D D D D E - - N
pseudocatenulatum   - D D E D D L D D D E E S E P E D D - - - D D E D E E R K K P E E - - - - - P K E K G A F V V R D D D D D E - - N
dentium             - D D E A D L D E D E E S E P E D D - - - E D E D D E E R K K P E E - - - - - P K E K G A F V V R D D D D D D - - N
adolescentis        - G E D D D L D E D N I D E D D D S E L - - E D D D E D K R K K P E E - - - - - P K E K G A F V V R D D D D D D - - N
thermophilum        G D D P D N L D D D I D S D D E D D E G D D S E D D E E N K P K H R E P E V L P K A K G A F V M R D D D D D D - - N
bifidum             D G D D D D I D D E S G E D D D V P E D E E D D E G T E P G R K E P E Q - - - - P K E K G A F V V R D D D D D E - - N
scardovii           - - D E D E G D E E L D E D S D I A D D E D - - - - - E V K R K E P E T - - - - - P K E K G A F V V R D D D D D D - - N
angulatum           V D D V D D L D G I N G E S D D D D D D Y D D D D D D D D E E E D D K S D A P D I S K Q K G A F V V S D D D D D D E - D
animalis            L - D A D E N D D D L E D E E S E D E P E P E E D D - D E E P Q A P K E - - - - - P K E K G A F V V S D N E D D E A L
choerinum           L G D V D G Q D V D A D D E D D E S E P E P E E D D E D E E E K K P E E - - - - - P K E K G A F V V R D D D D D E - - T
pseudolongum        V D D H D H M D D D E D E H D T D D D A V E D D E D D E D E P K P E E - - - - - P K A K G A F V V R D D D D D E - - T
indicum             K D Q D Q L E D E P G S D D E E E H E A R E V S E E A E E D R K A A H E L A - - - A K V K G A F V V D D S E D D E - - S
coryneforme         K D Q D Q L E D E P G S D D E E E H E A R E V S E E A E E D R K A A H E L A - - - A K V K G A F V V D D S E D D E - - S
asteroides          K S E D Q L E D E P D E D D Q E E D E V S R V E D E A E E D R K A A H A L A - - - A K V K G A F V V D D S D D D E - - N
actinocoloniiforme  - Q E D Q V E D E E P D G E V D - - - - - - - G N E D E E E S P K P V V T P - - - P D I K G A F V V S D S D D D E - - N
                        *           .                                          . .  * * * : * : *   : : : :
```

Alignment — Region 1

	1.2
longum	I I P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
breve	I I P A - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
catenulatum	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
kashiwanohense	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
pseudocatenulatum	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
dentium	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
adolescentis	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V N L L N A E Q E V D L S E R I E A G L Y A Q H
thermophilum	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
bifidum	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
scardovii	L T P S - - - - - - G N P K R R V I A A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
angulatum	V K P R R R R G K I G N P K R R V V T A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
animalis	L A N P G - - - - - V N P K R R V V T T G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
choerinum	L A P T G - - - - - S N P R R R V V T T G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
pseudolongum	L A P T G - - - - - K N P K R R V V T S G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
indicum	V T P S - - - - - - G N P K R R V V T A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
coryneforme	V T P S - - - - - - G N P K R R V V T A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
asteroides	V T P S - - - - - - G N P K R R V V T A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H
actinocoloniiforme	V T P S - - - - - - G N P K R R V V T A G A T A D P V K D Y L K Q I G R V S L L N A E Q E V D L S E R I E A G L Y A Q H

: . * * : * * * : : : * . * * * * * * * * * * * * * * * * *

Alignment — Region 2

		2.1	2.2
longum	L L D T E S E G M E F K R K R E L K W A A A D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
breve	L L D T E S E G M E F K R K R E L K W A A A D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
catenulatum	L L D T E S D Q M D F K R K R E L K W A A N D G R	R A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
kashiwanohense	L L D T E S D Q M D F K R K R E L K W A A N D G R	R A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
pseudocatenulatum	L L D T E S D Q M D F K R K R E L K W A A N D G R	R A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
dentium	L L D T E A S Q M D F K R R R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
adolescentis	L L D T E S D K M D F K R R R E L K W A A N D G K	R A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
thermophilum	L L D T E S E G M S F K R K R E L K W A A A D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
bifidum	L L D T E S D G M D F K R K R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
scardovii	L L D T Q S E G M D F K R K R E L K W A A S D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
angulatum	L L D T E G D T M D F K R K R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
animalis	L L D T E G D T M E F R R K R E L K W A A A D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
choerinum	L L D T E G D K L D F K R K R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
pseudolongum	L L E T E G D T L D F K R K R E L K W A A S D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
indicum	L L D T D D E G M D F K R R R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
coryneforme	L L D T D D E G M D F K R R R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
asteroides	L L D T E G E G M D F K R R R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G
actinocoloniiforme	L L D T E G E D M T F K R K R E L K W A A N D G K	K A K D H L L E A N L R L V V S L A K R Y T G R	G M L F L D L I Q E G

* * : * : . : * : : * * * * * * * * * : : *

Alignment — Region 3

		2.3	2.4	2.5	
longum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
breve	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
catenulatum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
kashiwanohense	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
pseudocatenulatum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
dentium	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
adolescentis	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
thermophilum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L A R V Q R
bifidum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
scardovii	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
angulatum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
animalis	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
choerinum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
pseudolongum	N L G L I R A V E K	F D W K K G F K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
indicum	N L G L I R A V E K	F D Y T K G Y K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
coryneforme	N L G L I R A V E K	F D Y T K G Y K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
asteroides	N L G L I R A V E K	F D Y T K G Y K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R
actinocoloniiforme	N L G L I R A V E K	F D Y T K G Y K F S T Y A T W W	I R Q A I T R A M A D Q A R T I R V P	V H M V E V	N K L S R V Q R

* * * * * * * * * * * * : : . * * : * : . * * * *

Alignment — Region 4

	3.1	3.2
longum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
breve	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
catenulatum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
kashiwanohense	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
pseudocatenulatum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
dentium	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
adolescentis	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
thermophilum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
bifidum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
scardovii	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
angulatum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
animalis	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
choerinum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
pseudolongum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
indicum	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
coryneforme	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
asteroides	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A
actinocoloniiforme	Q M L Q D L G R E P T P D E L A R E L D M P V E K V Q E V	Q K Y G R E P I S L H T P L G E D G D S E F G D L I E D T D A

* *

```
                    4.1                                                                    4.2
longum              I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
breve               I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
catenulatum         I A P S D A V A F S L L Q E Q F R Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
kashiwanohense      I A P S D A V A F S L L Q E Q F R Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
pseudocatenulatum   I A P S D A V A F S L L Q E Q F R Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
dentium             I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
adolescentis        I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
thermophilum        I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
bifidum             I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
scardovii           I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I R M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
angulatum           I A P S D A V A F S L L Q D Q F K Q V L E T L S P R E S G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
animalis            I A P S E A V A F S L L Q E Q F R Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
choerinum           I A P S D A V A F S L L Q E Q F R Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
pseudolongum        I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
indicum             I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
coryneforme         I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
asteroides          I A P S D A V A F S L L Q E Q F Q Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
actinocoloniiforme  I A P S D A V A F S L L Q E Q F K Q V L E T L S P R E A G V I K M R Y G L E D G Q P K T L D D I G R V Y G V T R E R I R
                    * * * * : * * * * * * * : * * : * * * * * * * * * * * : * * * : * * * * * * * * * * * * * * * * * * * * * * * * * *
```

```
longum              Q I E S K T M S K L R H P S R S Q T L R D F L D Q
breve               Q I E S K T M S K L R H P S R S Q T L R D F L D Q
catenulatum         Q I E S K T M S K L R H P S R S Q T L R D F L D Q
kashiwanohense      Q I E S K T M S K L R H P S R S Q T L R D F L D Q
pseudocatenulatum   Q I E S K T M S K L R H P S R S Q T L R D F L D Q
dentium             Q I E S K T M S K L R H P S R S Q T L R D F L D Q
adolescentis        Q I E S K T M S K L R H P S R S Q T L R D F L D Q
thermophilum        Q I E S K T M S K L R H P S R S Q T L R D F L D Q
bifidum             Q I E S K T M S K L R H P S R S Q T L R D F L D Q
scardovii           Q I E S K T M S K L R H P S R S Q T L S D F L D Q
angulatum           Q I E S K T M S K L R H P S R S Q T L R D F L D Q
animalis            Q I E S K T M S K L R H P S R S Q T L R D F L D Q
choerinum           Q I E S K T M S K L R H P S R S Q T L R D F L D Q
pseudolongum        Q I E S K T M S K L R H P S R S Q T L R D F L D Q
indicum             Q I E S K T M S K L R H P S R S Q T L R D F L D Q
coryneforme         Q I E S K T M S K L R H P S R S Q T L R D F L D Q
asteroides          Q I E S K T M S K L R H P S R S Q T L R D S L D Q
actinocoloniiforme  Q I E S K T M S K L R H P S R S Q T L R D F L D Q
                    * * * * * * * * * * * * * * * * * * *   *   * * *
```

Fig. S6. The alignment of sigma factors of the genus *Bifidobacterium*.

The alignment of primary sigma factors of the 18 bacteria belonging to the genus *Bifidobacterium*. Amino acids in N-terminal regions are colored as the positively charged amino acids K, R, H (blue), the negatively charged amino acids D, E (red), and the non-charged polar amino acids N, S, T, and Q (yellow).

Bifidobacterium longum
```
  1 MATKETTATKQTAESSEETETKKKASSTTTARKTSAKKSKSAEPSKKTTRKSAAKKTAKK   60
 61 AKDDELVKDEDTQVDDELDDDAEVNDDDLDVDDLDEDLDDVDEDDADDDLDEDDADDDD   120
121 VEDDEDEEEAKSKAPEQPKEK                                         141
```

Bifidobacterium breve
```
  1 MATKETTATKQTAESSEETETKKKASSTTARKTSAKKTDTKGTTAAKTGEKTSRKTSAK    60
 61 KVKKEEELPQNDISQSDDPEDLDVDNLDEDLDDVESEDVDDLEDADAESDDEDIDDEDDE  120
121 DEEEDEGKAASKAPEQPKEK                                         139
```

Gardnerella vaginalis
```
  1 MTSEVGLTVDSDEMQDDTKPVKTRKASVARKTSTRTSSADSKTRSSRSKKTEDNKIDTES   60
 61 VSADAEHDETEDSVDIESQDQLSIDDPLDDDSHLDVGDEHLNLDDLDDIQPEDISDVDDL  120
121 SDIDVADDLDDEESEDLDEIDDHDEIKHDAPQVPQAK                        157
```

Scardovia inopinata
```
  1 MVAAKKTDAGIGEEPVSSGTTVSKTGKKVKSSAKKSSAKTGTAQKSTAKTSSAAKKSTAKT  60
 61 ASTKKTSSSAKKTASTRKKTVKKKPAVEDDQKKTSQSTSAHSAEGKSSQKTASKSKKAASKK 120
121 SSAIKSANEEPQTSSRKTHKKQDDLSHVPEKGSESQTGTHDEDIRSLDSLDAGNQTDTDI  180
181 PVEDDDIDEGIGPGDERSNALDHIDDSVELDENGDPISPDDDSDILADAEEKDLEEEEED  240
241 EGRAEEKQPVEE                                                 252
```

Parascardovia denticolens
```
  1 MATKKTVTTDANPTVTVSEIDASVKGSEVKISRTKTKKATGTKKATASKKTTVKKPSSKA   60
 61 ASAKATSAKAKSDKTESGKVEPVAEKDVASTQKPAKAKKSVASDKAEKSAESNKGKSAKT  120
121 VKSDKTEKAEDSAKTVKSTKTSATKAPKTKAASRSSSTSKAKKVAAKTSEKTGKKSSAKSAAEPT 180
181 AAKTGKSRKTDETEKEEGKASAKKTTSKKASAKKVAAKTSEKTGKKSSAKSAEKTEDEDL  240
241 PAEEASAEEGKVEEDKVDESTSASDEDDDQSLNDDDDMDGDLGPGSHAVNALDHIDDSVEL 300
301 DENGDPIQEDDENIVTDSEEKDLEDEGDEEEEAPVEKKPLEE                   342
```

Mycobacterium tuberculosis
```
  1 MAATKASTATDEPVKRTATKSPAASASGAKTGAKRTAAKSASGSPPAKRATKPAARSVKP   60
 61 ASAPQDTTTSTIPKRKTRAAAKSAAAKAPSARGHATKPRAPKDAQHEAATDPEDALDSVE  120
121 ELDAEPDLDVEPGEDLDLDAADLNLDDLEDDVAPDADDDLDSGDDEDHEDLEAEAVAPG   180
181 QTADDDEEIAEPTEKDKAS                                          199
```

Mycolicibacterium smegmatis
```
  1 MAATKASPATEEPVKRTATKTPAKKAPAKRAAKSAAAKAGGKAPAKKAPAKRAAKGTAAK   60
 61 PEDGVTDDLEVTDDLEAEPGEDLDVEDTDLELDDLDSDDDTAVEDEEEEADAATPAVATA  120
121 KAADDDIDEPSEKDKAS                                            137
```

Corynebacterium glutamicum
```
  1 MESSMVENNVAKKTVAKKTARKTARKAAPRVATPLGVASESPISATPARSIDGTSTPVEA   60
 61 ADTIETTAPAAKAPAAKAPAKKTVAKKTARKAPAKKTVAKKATTAKAAPATAKDENAPVDD  120
121 DEENLAQDEQDFDGDDDFVDGIEDEEDEDGVEALGEESEDDEED                 163
```

Streptomyces coelicolor
```
  1 MSASTSRTLPPEIAESVSVMALIERGKAEGQIAGDDVRRAFEADQIPATQWKNVLRSLNQ   60
 61 ILEEEGVTLMVSAAEPKRTRKSVAAKSPAKRTATKAVAAKPVTSRKATAPAAPAATEP   120
121 AAVEEEAPAKKAAAKKTTAKKATAKKTTAKKEDGELLEDEATEEPKAATE          180
181 EPEGTEN                                                      187
```

Fig. S7. The amino acid sequences of N-terminal regions of sigma factors of the phylum Actinobacteria.

Alignment of N-terminal sequences of primary sigma factors of the nine bacteria belonging to the phylum Actinobacteria. Amino acids in N-terminal regions are colored as the positively charged amino acids K, R, H (blue), the negatively charged amino acid D, E (red), and the non-charged polar amino acids N, S, T, and Q (yellow).

Table S1. The plasmids and primers.

Type		Properties	Reference
Plasmids	pKO403-*lac*Z	4,473 bp, Ori^{ts}, *lac*Z, Sp^r	(1)
	pKO403-TPCTcon	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (17 bp)-TACAAT	This study
	pKO403-TPCT1-3	4,913 bp, Ori^{ts}, Sp^r, Cm^r, vTGTGC-spacer (17 bp)-TACAAT	This study
	pKO403-TPCT4-6	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TvGTGC-spacer (17 bp)-TACAAT	This study
	pKO403-TPCT7-9	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TThTGC-spacer (17 bp)-TACAAT	This study
	pKO403-TPCT10-12	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGvGC-spacer (17 bp)-TACAAT	This study
	pKO403-TPCT13-15	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGThC-spacer (17 bp)-TACAAT	This study
	pKO403-TPCT16-18	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGd-spacer (17 bp)-TACAAT	This study
	pKO403-TPCT19-21	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (17 bp)-vACAAT	This study
	pKO403-TPCT22-24	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (17 bp)-TbCAAT	This study
	pKO403-TPCT25-27	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (17 bp)-TAvAAT	This study
	pKO403-TPCT28-30	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (17 bp)-TACbAT	This study
	pKO403-TPCT31-33	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (17 bp)-TACAbT	This study
	pKO403-TPCT34-36	4,913 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (17 bp)-TACAAv	This study
	pK0403-TPCT-10bp	4,906 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (10 bp)-TATAAT	This study
	pK0403-TPCT-11bp	4,907 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (11 bp)-TATAAT	This study
	pK0403-TPCT-12bp	4,908 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (12 bp)-TATAAT	This study
	pK0403-TPCT-15bp	4,911 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (15 bp)-TATAAT	This study
	pK0403-TPCT-16bp	4,912 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (16 bp)-TATAAT	This study
	pK0403-TPCT-18bp	4,914 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (18 bp)-TATAAT	This study
	pK0403-TPCT-19bp	4,915 bp, Ori^{ts}, Sp^r, Cm^r, TTGTGC-spacer (19 bp)-TATAAT	This study
Primers (5' → 3')	Fw_-35	TCACTACAGAGTCGGACTACAATCA	This study
	Rv_-35_mu1-3	GCACAbTGCCCGCGTTTTATA	This study
	Rv_-35_mu4-6	GCACbATGCCCGCGTTTTATA	This study
	Rv_-35_mu7-9	GCAdAATGCCCGCGTTTTATA	This study
	Rv_-35_mu10-12	GCbCAATGCCCGCGTTTTATA	This study
	Rv_-35_mu13-15	GdACAATGCCCGCGTTTTATA	This study
	Rv_-35_mu16-18	hCACAATGCCCGCGTTTTATA	This study
	Rv_-10	GAGCACAATGCCCGCGTTTTA	This study
	Fw_-10_mu1-3	ACTACAGAGTCGGACvACAATCATTGA	This study
	Fw_-10_mu4-6	ACTACAGAGTCGGACTbCAATCATTGA	This study
	Fw_-10_mu7-9	ACTACAGAGTCGGACTAdAATCATTGA	This study
	Fw_-10_mu10-12	ACTACAGAGTCGGACTACbATCATTGA	This study
	Fw_-10_mu13-15	ACTACAGAGTCGGACTACAbTCATTGA	This study
	Fw_-10_mu16-18	ACTACAGAGTCGGACTACAAvCATTGA	This study
	Fw_-10	ACTACAGAGTCGGACTACAATCATTGA	This study

Fw_spacer19	GTACCGGACTATAATCATTGATG	This study
Fw_spacer18	GTACGGACTATAATCATTGATG	This study
Fw_spacer16	GCGGACTATAATCATTGATG	This study
Fw_spacer15	CGGACTATAATCATTGATG	This study
Rv_spacer15-19	TCTGTAGTGAGCACAATG	This study
Fw_spacer10-12	CGGACTATAATCATTGATGACAACA	This study
Rv_spacer12	GTAGTGAGCACAATGCCC	This study
Rv_spacer11	TAGTGAGCACAATGCCCG	This study
Rv_spacer10	AGTGAGCACAATGCCCG	This study

Table S2. The bacterial strains of the genus *Bifidobacterium*.

	Type	Properties	GenBank
Bifidobacterium	*longum* NCC2705	RNA-Seq sample, Transformation host, Informatics analysis	Nantes Culture Collection, AE014295.3
	actinocoloniiforme DSM 22766^T	Informatics analysis	CP011786
	adolescentis ATCC 15703	Informatics analysis	AP009256
	angulatum JCM 1194^T	Informatics analysis	AP012322
	animalis strain RH	Informatics analysis	CP007755
	asteroides PRL 2011	Informatics analysis	CP003325
	bifidum S17	Informatics analysis	CP002220
	breve UCC 2003	Informatics analysis	CP000303
	catenulatum JCM 1194^T	Informatics analysis	AP012325
	choerinum FMB-1	Informatics analysis	CP018044
	coryneform strain LMG 18911^T	Informatics analysis	CP007287
	dentium JCM 1195^T	Informatics analysis	AP012326
	indicum LMG 11587^T	Informatics analysis	CP006018
	kashiwanohense PV20-2	Informatics analysis	CP007456
	pseudocatenulatum JCM 1200^T	Informatics analysis	AP012330
	pseudolongum PV8-2	Informatics analysis	CP007457
	scardovii JCM 12489^T	Informatics analysis	AP012331
	thermophilum RBL67	Informatics analysis	CP004346

Table S3. The genes with 11 bp type spacer in the upstream regions.

Gene	Function
BL0001	cold-shock protein
BL0004	hypothetical protein
BL0017	histidyl-tRNA synthetase
BL0047	polyphosphate glucokinase
BL0060	trans-aconitate 2-methyltransferase
BL0065	elongation factor P
BL1292	morphine 6-dehydrogenase
BL1422	anthranilate phosphoribosyltransferase
BL1431	ribokinase
BL1432	ATP-dependent helicase II
BL1444	EamA-like transporter family
BL0702	excinuclease ABC subunit A
BL0712	lanine-synthesizing transaminase
BL0724	threonyl-tRNA synthetase
BL0739	UTP-glucose-1-phosphate uridylyltransferase
BL0745	hypothetical protein
BL0766	hypothetical protein
BL0815	hypothetical protein
BL0854	hypothetical protein
BL0872	ABC transporter
BL0969	acetate kinase
BL0992	30S ribosomal protein S1
BL0993	methenyltetrahydrofolate cyclohydrolase
BL1026	peptidyl-tRNA hydrolase
BL1107	phosphoribosylaminoimidazole-succinocarboxamide synthase
BL1118	amino acid ABC transporter substrate-binding protein
BL1151	hypothetical protein
BL1164	sugar ABC transporter solute-binding protein
BL1175	glucosamine--fructose-6-phosphate aminotransferase
BL1194	4-hydroxy-tetrahydrodipicolinate reductase
BL1229	tyrosine kinase
BL0294	50S ribosomal protein L32
BL0319	phospho-2-dehydro-3-deoxyheptonate aldolase
BL0330	50S ribosomal protein L28
BL0458	hypothetical protein
BL0477	hypothetical protein

BL0492	aspartate-semialdehyde dehydrogenase
BL0525	LacI-type transcriptional regulator
BL0557	hypothetical protein
BL0564	modification methylase Sau3AI
BL0600	tryptophanyl-tRNA synthetase
BL0617	putative hemolysin
BL0681	transporter
BL0145	hypothetical protein
BL0160	MarR-type transcriptional regulator
BL0181	alpha-arabinofuranosidase I
BL0202	hypothetical protein
BL1530	IclR-type transcriptional regulator
BL1567	hypothetical protein
BL1609	hypothetical protein
BL1652	arabinose polymer utilization protein
BL1657	two-component system response regulator
BL1661	hypothetical protein
BL1694	sugar ABC transporter substrate-binding protein
BL1723	oligoribonuclease
BL1728	prolyl-tRNA synthetase
BL1732	methionine aminopeptidase
BL1745	hypothetical protein
BL1765	MarR-type transcriptional regulator

Perl script for spacer analysis

```perl
open (IN,"<genome.txt"); #genome.txt is FASTA format file of bacterium whole-genome.
open (OUT,">A_total-genome.txt");
open (OUT2,">B_total-genome_rev.txt");
while (<IN>) {
        s/^>.+¥n/¥n/;
    print OUT;
    tr/TAGC/ATCG/;
    print OUT2;
}
close(IN);
close(OUT);
close(OUT2);


open (IN, "<cds_GB.txt"); #cds_GB.txt is GB format file of bacterium.
open (OUT, ">C_gene_locus.txt");
while ($file = <IN>) {
        if($file =~ /^ +CDS +complement¥([0-9]+¥.¥.([0-9]+)¥)/){
        $cds_comp = $1;
        print OUT "gene¥t";
        print OUT $cds_comp;
        print OUT "¥t-¥n";
        }
        elsif($file =~ /^ +CDS +([0-9]+)¥.¥.[0-9]+/){
        $cds_comp = $1;
        print OUT "gene¥t";
        print OUT $cds_comp;
        print OUT "¥t¥+¥n";
        }
}
close(IN);
close(OUT);


$X = 150; #$X is the length of upstream regions you extract.
open (IN,"<A_total-genome.txt");
open (IN2,"<B_total-genome_rev.txt");
open (IN3,"<C_gene_locus.txt");
open (OUT,">D_up_150_region.txt");
@genome = <IN>;
chomp @genome;
```

```perl
$genome = join("",@genome);
@genome_rev = <IN2>;
chomp @genome_rev;
$genome_rev = join("",@genome_rev);
@line = <IN3>;
chomp @line;
for($i=0;$i<@line;$i++){
    if(@line[$i] =~ /(gene\t)(\d+)(\t\+)/){
        $start = $2 - $X - 1;
        $sequense = substr($genome,$start,$X);
        print OUT $sequense;
        print OUT "\n";
    }
    if(@line[$i] =~ /(gene\t)(\d+)(\t-)/){
        $start = $2 ;
        $sequense =reverse substr($genome_rev,$start,$X);
        print OUT $sequense;
        print OUT "\n";
    }
}
close(IN);
close(IN2);
close(IN3);
close(OUT);

open(OUT, '> F_spacer-count_upstream.txt');
print OUT "Seq\tNo\n";
open(IN, '< D_up_150_region.txt') or die;
while (<IN>){
$Zero++ while (/TTG[A-Z]{3}TA[A-Z]{3}T/g);
$One++ while (/TTG[A-Z]{4}TA[A-Z]{3}T/g);
$Two++ while (/TTG[A-Z]{5}TA[A-Z]{3}T/g);
$Three++ while (/TTG[A-Z]{6}TA[A-Z]{3}T/g);
$Four++ while (/TTG[A-Z]{7}TA[A-Z]{3}T/g);
$Five++ while (/TTG[A-Z]{8}TA[A-Z]{3}T/g);
$Six++ while (/TTG[A-Z]{9}TA[A-Z]{3}T/g);
$Seven++ while (/TTG[A-Z]{10}TA[A-Z]{3}T/g);
$Eight++ while (/TTG[A-Z]{11}TA[A-Z]{3}T/g);
$Nine++ while (/TTG[A-Z]{12}TA[A-Z]{3}T/g);
$Ten++ while (/TTG[A-Z]{13}TA[A-Z]{3}T/g);
```

```perl
$Eleven++ while (/TTG[A-Z]{14}TA[A-Z]{3}T/g);
$Twelve++ while (/TTG[A-Z]{15}TA[A-Z]{3}T/g);
$Thirteen++ while (/TTG[A-Z]{16}TA[A-Z]{3}T/g);
$Fourteen++ while (/TTG[A-Z]{17}TA[A-Z]{3}T/g);
$Fifteen++ while (/TTG[A-Z]{18}TA[A-Z]{3}T/g);
$Sixteen++ while (/TTG[A-Z]{19}TA[A-Z]{3}T/g);
$Seventeen++ while (/TTG[A-Z]{20}TA[A-Z]{3}T/g);
$Eighteen++ while (/TTG[A-Z]{21}TA[A-Z]{3}T/g);
$Nineteen++ while (/TTG[A-Z]{22}TA[A-Z]{3}T/g);
$Twenty++ while (/TTG[A-Z]{23}TA[A-Z]{3}T/g);
$Twe1++ while (/TTG[A-Z]{24}TA[A-Z]{3}T/g);
$Twe2++ while (/TTG[A-Z]{25}TA[A-Z]{3}T/g);
$Twe3++ while (/TTG[A-Z]{26}TA[A-Z]{3}T/g);
$Twe4++ while (/TTG[A-Z]{27}TA[A-Z]{3}T/g);
$Twe5++ while (/TTG[A-Z]{28}TA[A-Z]{3}T/g);
}
print OUT "0bp\t$Zero\n";
print OUT "1bp\t$One\n";
print OUT "2bp\t$Two\n";
print OUT "3bp\t$Three\n";
print OUT "4bp\t$Four\n";
print OUT "5bp\t$Five\n";
print OUT "6bp\t$Six\n";
print OUT "7bp\t$Seven\n";
print OUT "8bp\t$Eight\n";
print OUT "9bp\t$Nine\n";
print OUT "10bp\t$Ten\n";
print OUT "11bp\t$Eleven\n";
print OUT "12bp\t$Twelve\n";
print OUT "13bp\t$Thirteen\n";
print OUT "14bp\t$Fourteen\n";
print OUT "15bp\t$Fifteen\n";
print OUT "16bp\t$Sixteen\n";
print OUT "17bp\t$Seventeen\n";
print OUT "18bp\t$Eighteen\n";
print OUT "19bp\t$Nineteen\n";
print OUT "20bp\t$Twenty\n";
print OUT "21bp\t$Twe1\n";
print OUT "22bp\t$Twe2\n";
print OUT "23bp\t$Twe3\n";
```

```perl
print OUT "24bp¥t$Twe4¥n";
print OUT "25bp¥t$Twe5¥n";
close(IN);
close(OUT);

open (IN,"< A_total-genome.txt");
open (IN2,"< B_total-genome_rev.txt");
open (OUT,"> E_total-genome_both.txt");
@genome = <IN>;
chomp @genome;
$genome = join("",@genome);
@genome_rev = <IN2>;
chomp @genome_rev;
$genome_rev = join("",@genome_rev);
$genome_rev = reverse($genome_rev);
print OUT "$genome¥n$genome_rev";
close(IN);
close(IN2);
close(OUT);

open(OUT, '> G_spacer-count-whole.txt');
print OUT "Seq¥tNo¥n";
open(IN, '< E_total-genome_both.txt') or die;
while (<IN>){
$Zero++ while (/TTG[A-Z]{3}TA[A-Z]{3}T/g);
$One++ while (/TTG[A-Z]{4}TA[A-Z]{3}T/g);
$Two++ while (/TTG[A-Z]{5}TA[A-Z]{3}T/g);
$Three++ while (/TTG[A-Z]{6}TA[A-Z]{3}T/g);
$Four++ while (/TTG[A-Z]{7}TA[A-Z]{3}T/g);
$Five++ while (/TTG[A-Z]{8}TA[A-Z]{3}T/g);
$Six++ while (/TTG[A-Z]{9}TA[A-Z]{3}T/g);
$Seven++ while (/TTG[A-Z]{10}TA[A-Z]{3}T/g);
$Eight++ while (/TTG[A-Z]{11}TA[A-Z]{3}T/g);
$Nine++ while (/TTG[A-Z]{12}TA[A-Z]{3}T/g);
$Ten++ while (/TTG[A-Z]{13}TA[A-Z]{3}T/g);
$Eleven++ while (/TTG[A-Z]{14}TA[A-Z]{3}T/g);
$Twelve++ while (/TTG[A-Z]{15}TA[A-Z]{3}T/g);
$Thirteen++ while (/TTG[A-Z]{16}TA[A-Z]{3}T/g);
$Fourteen++ while (/TTG[A-Z]{17}TA[A-Z]{3}T/g);
$Fifteen++ while (/TTG[A-Z]{18}TA[A-Z]{3}T/g);
```

```perl
$Sixteen++ while (/TTG[A-Z]{19}TA[A-Z]{3}T/g);
$Seventeen++ while (/TTG[A-Z]{20}TA[A-Z]{3}T/g);
$Eighteen++ while (/TTG[A-Z]{21}TA[A-Z]{3}T/g);
$Nineteen++ while (/TTG[A-Z]{22}TA[A-Z]{3}T/g);
$Twenty++ while (/TTG[A-Z]{23}TA[A-Z]{3}T/g);
$Twe1++ while (/TTG[A-Z]{24}TA[A-Z]{3}T/g);
$Twe2++ while (/TTG[A-Z]{25}TA[A-Z]{3}T/g);
$Twe3++ while (/TTG[A-Z]{26}TA[A-Z]{3}T/g);
$Twe4++ while (/TTG[A-Z]{27}TA[A-Z]{3}T/g);
$Twe5++ while (/TTG[A-Z]{28}TA[A-Z]{3}T/g);
}
print OUT "0bp¥t$Zero¥n";
print OUT "1bp¥t$One¥n";
print OUT "2bp¥t$Two¥n";
print OUT "3bp¥t$Three¥n";
print OUT "4bp¥t$Four¥n";
print OUT "5bp¥t$Five¥n";
print OUT "6bp¥t$Six¥n";
print OUT "7bp¥t$Seven¥n";
print OUT "8bp¥t$Eight¥n";
print OUT "9bp¥t$Nine¥n";
print OUT "10bp¥t$Ten¥n";
print OUT "11bp¥t$Eleven¥n";
print OUT "12bp¥t$Twelve¥n";
print OUT "13bp¥t$Thirteen¥n";
print OUT "14bp¥t$Fourteen¥n";
print OUT "15bp¥t$Fifteen¥n";
print OUT "16bp¥t$Sixteen¥n";
print OUT "17bp¥t$Seventeen¥n";
print OUT "18bp¥t$Eighteen¥n";
print OUT "19bp¥t$Nineteen¥n";
print OUT "20bp¥t$Twenty¥n";
print OUT "21bp¥t$Twe1¥n";
print OUT "22bp¥t$Twe2¥n";
print OUT "23bp¥t$Twe3¥n";
print OUT "24bp¥t$Twe4¥n";
print OUT "25bp¥t$Twe5¥n";
close(IN);
close(OUT);
```

REFERENCE

1.   Sakaguchi K, He J, Tani S, Kano Y, Suzuki T. 2012. A targeted gene knockout method using a newly

      constructed temperature-sensitive plasmid mediated homologous recombination in *Bifidobacterium*

      longum. Appl Microbiol Biotechnol 95:499–509.