

University of Groningen

## Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries

Kovacs, Akos T.; Smits, Wiep Klaas; Mironczuk, Aleksandra M.; Kuipers, Oscar; Mirończuk, Aleksandra M.

*Published in:*  
Environmental Microbiology

*DOI:*  
[10.1111/j.1462-2920.2009.01937.x](https://doi.org/10.1111/j.1462-2920.2009.01937.x)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2009

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Kovacs, A. T., Smits, W. K., Mironczuk, A. M., Kuipers, O. P., & Mirończuk, A. M. (2009). Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries. *Environmental Microbiology*, 11(8), 1911-1922. DOI: 10.1111/j.1462-2920.2009.01937.x

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries

Ákos T. Kovács, Wiep Klaas Smits, Aleksandra M. Mirończuk, Oscar P. Kuipers

### Legends to Supplementary Figures

#### Supplementary Figure 1

Presence of homologs of proteins involved in the regulation of competence in *Bacillus* and closely related species. Results of BLAST queries were visualized with Genesis 1.6 software: white is absent (with e-value of E-0), dark blue is present (e-value <E-20). BLAST analysis was performed with *B. subtilis* protein sequences against the translated protein database of a given genome. Question marks denote small ORFs where identification is uncertain using the available bioinformatic tools that can miss homologues. Where the search yielded no hit (or with an e-value below E-05) an additional tblastn was performed. Protein names are indicated on the right. Bsu (*B. subtilis*), Bam (*B. amyloliquefaciens*), Bli (*B. licheniformis*), Bpu (*B. pumilus*), Ban (*B. anthracis*), Bce (*B. cereus*), Bth (*B. thuringiensis*), Bwe (*B. weihenstephanensis*), Oih (*O. iheyensis*), Gka (*G. kaustophilus*), Gth (*G. thermodenitrificans*), Bcl (*B. clausii*), Bha (*B. halodurans*).

#### Supplementary Figure 2

Multiple alignments of ComG<sub>A-G</sub> homologues. Black background represents conserved amino acids and grey background represents similar amino acids. Alignment was performed using Clustal W (Thompson et al., 1994), and presented using the Boxshade 3.21 program. For abbreviations of species names see Supplementary Fig. 1. Conserved domains of *B. subtilis* ComG proteins are indicated with arrows above the alignment. AAA+, ATPase domain (Smart accession number: SM00382); GSPII-F, general secretion pathway domain (PFAM accession number: PF00482); TMS, transmembrane segment; N-met, N-methyl domain often found at N-terminus of pilins and proteins involved in secretion (PFAM accession number: PF07963); SSeq, signal peptide sequence.



## Supplementary Figure 2

### ComG<sub>A</sub>

```

Bsu168      1  --DSETEKVSKNLIEEAYLTASDIHIVPRBEDAIIHFRVVDHALIKRRDKKKEECVRLSHFKFLSAMDIGERRKPPONGSLTLKKEGNVHLRMSLPTI
BamFZB42   1  --MKNTEYYSRNLINEAYTARASDIHIVPRBEDAFHFRIGEAHMKRRVKKKEECVRLSHFKFLSAMDIGERRKPPONGSLPLPPTETVHLRMSLPTV
Bli        1  --YITTEISLGRLLDEAYRMRASDIHIVGCKEAVYFRIDDELFQKDCIENECSSLSHFKFLSMDIGERRKPPONGSLALYINREPVHLRMSLPTI
BpuSAFR    1  --LYGEYVLCGELLLEACRMRASDIHIVGCKEASYSFRVSDLLIQRTIDKKSGERLASHFKFLSMDIGERRKPPONGSLAMRISGQVHLRMSLPTV
Ban        1  --MNGTESFANMLKEACRVOASDLHIVPRCKDVVQLRIKQDLMTKQCIKKEGKELVSHFKFLSAMDIGERRKPPONGSLVLOHDGQEVHLRMSLPTV
Bth        1  --MNGTESFANMLKEACRVOASDLHIVPRCKDVVQLRIKQDLMTKQCIKKEGKELVSHFKFLSAMDIGERRKPPONGSLVLOHDGQEVHLRMSLPTV
Bce        1  --MNGTEIFANMLKEACRVOASDLHIVPRCKDVVQLRIKQDLMTKQCIKKEGKELVSHFKFLSAMDIGERRKPPONGSLVLOHDGQEVHLRMSLPTV
Bwe        1  --MNSVELFANMLKEACRVOASDLHIVPRCKDVVQLRIKQDLMTKQCIKKEGKELVSHFKFLSAMDIGERRKPPONGSLVLOHDGQEVHLRMSLPTV
Gth        1  VDDDEEQVANRLAEAVORASDIHIVPRRHDATIRRLDGMIDVGTISKEAERVIVHFKEFLASMDIGERRRPPONGSLVNESSETVHLRMSLPTI
Oih        1  --MNEKLLSDSLIHSVSLQSSDIHFVLYPTQTEYPRFHGRRLHRRKISSNOYRLLIYKFLSGMDIGERRRPPONGSLIWIWTEKSNYSLRMSLPTI
BhaC125   1  --ITTHNQSRKLLQLAQCASDIHLSSTREDEIQFRINGRVAPFQRIPLQTEERLASHFKFLSAMDIGERRKPPONGSLIIEQKIPVSLRMSLPTI
BclKSMK16 1  --MADTEGKSTAILVEAFQSSASDIHFVETSQKYLHHRALGETHAGQTLKAEICFRIVSFKACGMDIGERRKPPONGSLIISYDWOVNGYTLRMSLPTI
    
```

#### AAA+

```

Bsu168      99 NEESLVIRVMPQYNIPSHDLISLFFKTCATLLSFLKHSHGMLIFTGPTGSGKTTTLYSLVQYAKKHFNRNIITLEDVPEVREDDVLOVQNEKAGVTYS
BamFZB42   99 NDESLVIRLLEKROVPLDMLSLFRGACALLSFLKHSHGMLIFTGPTGSGKTTTLYSLVQYAKKHFNRNIITLEDVPEVREDDVLOVQNEKAGVTYS
Bli        99 HDESLVIRLLEKMSKPLTLKSLFPPSATFKLLSFLKHSHGMLIFTGPTGSGKTTTLYSLVQYAKKHFNRNIITLEDVPEVREDDVLOVQNEKAGVTYS
BpuSAFR    99 NEESLVIRLIPQDHPKIKHLSLFFKASTLLSFLKHSHGMLIFTGPTGSGKTTTLYSLVQYAKKHFNRNIITLEDVPEVREDDVLOVQNEKAGITYA
Ban        99 YQESLVIRLHLQASVQPLSLSLFPSTAKLLSFLVYSHGLLIFTGPTGSGKTTLYALLEVIRKKIRRIITLEDVPEVREDDVLOVQNEKAGITYE
Bth        99 YQESLVIRLHLQASVQPLSLSLFPSTAKLLSFLVYSHGLLIFTGPTGSGKTTLYALLEVIRKKIRRIITLEDVPEVREDDVLOVQNEKAGITYE
Bce        99 YQESLVIRLHLQASVQPLSLSLFPSTAKLLSFLVYSHGLLIFTGPTGSGKTTLYALLEVIRKKIRRIITLEDVPEVREDDVLOVQNEKAGITYE
Bwe        99 YQESLVIRLHLQASVQPLSLSLFPSTAKLLSFLKHSHGMLIFTGPTGSGKTTLYALLEVIRKKIRRIITLEDVPEVREDDVLOVQNEKAGITYK
Gth        100 YDESLVIRLIPQRFSLPLRSLFSHSTARLLSFLMPOQGLVLTGPTGSGKTTTLYLLVQCAERQNIITLEDVPEVREDDVLOVQNEKAGITYA
Oih        99 LLESVAIRLHPQSHQIEKLLFLFNQIRHKKELNSQAGILLTGTGSGKTTLYALLESMLAENSCQITLEDVPEVREDDVLOVQNEKAGITYH
BhaC125   100 LFEAVAIRLHPQVPPAFVESHPIKROVQQLSFLVTEVGLLIFTGPTGSGKTTLYALLEQALHMQNLEIITLEDVPEVREDDVLOVQNEKAGITYE
BclKSMK16 99 PSESVAIRLHPQVPPAFVESHPIKROVQQLSFLVTEVGLLIFTGPTGSGKTTLYALLEVQVQSSKNIITLEDVPEVREDDVLOVQNEKAGITYA
    
```

#### AAA+

```

Bsu168      198 AGLKAILRHDPDMIIIGEIRDAETAEI AVRAAMTGHVLSLHTRDAKGAIVRLIEFCINNEIEQTVIAIAAQRVLVLAQPCFCENGCSVYCRQSRNTR
BamFZB42   198 AGLKAILRHDPDMIIIGEIRDAETAEI AVRAAMTGHVLSLHTRDAKGAIVRLIEFCINNEIEQTVIAIAAQRVLVLAQPCFCKEENCSVYCRMYRTR
Bli        198 AGLKAILRHDPDMIIIGEIRDAETAEI AVRAALTGHVLSLHAKNAGAIYRLLEIGVDNTEIEQTLVAIAAQRVLVLAQPCFCG-DSCLYCRLSRVVR
BpuSAFR    198 AGLKAILRHDPDMIIIGEIRDAETAEI AVRAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-ETCQLYCRLSRVVR
Ban        198 AGLKAILRHDPDMIIIGEIRDAETAEI AVRAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-RKCSFLCNSRQVR
Bth        198 AGLKAILRHDPDMIIIGEIRDAETAEI AVRAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-RKCSFLCNSRQVR
Bce        198 TGLKAILRHDPDMIIIGEIRDAETAEI AVRAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-RKCSFLCNSRQVR
Bwe        198 TGLKAILRHDPDMIIIGEIRDAETAEI AVRAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-RKCSFLCNSRQVR
Gth        199 TGLKAILRHDPDMIIIGEIRDRHTAATAVRSALTGHLVSTHRAADAVGAVYRLHEFGVIPGDLAETLLAVSAQRVLELRCPLCG-DNCHPACRRVQRRR
Oih        199 TGLKAILRHDPDMIIIGEIRDRKTAQPALEAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-----RGKQR
BhaC125   199 VGLKAILRHDPDMIIIGEIRDTTAVAVRAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-----CSBYETWVNRK
BclKSMK16 198 TGLKAILRHDPDMIIIGEIRDRNTAEALCAALTGHVLSLHAKNAGAIYRMLIEFGVIMIEIEQTMVAIAAQRVLVLAQPCFCG-----KKN
    
```

```

Bsu168      298 RASVYELLYGKNIQCFQEAKEGNHAN--YQYQLRRLIRKGHALGYLITNNYDRWVYHEKD-----
BamFZB42   298 RASVYELLYGKNIQCFQEAKEGNHAN--YQYQLRRLIRKGHALGYLITNNYDRWVYHEAD-----
Bli        297 RASVYELLYGKSNLQCFQEAKEGCGG--IKTDLRLIRKGHALGYLPSKTYERWIGHED-----
BpuSAFR    297 RTNVEELLECKELGECI KEAGEYAH--SSYBTLQRLIRKGVALGYLSKNTYHRWVYEEASL-----
Ban        297 QASVYELLYGYELKQAIKEANGEVCT--YKHEFLQSSIRKGVALGFLLEDVWV-----
Bth        297 QASVYELLYGYELKQAIKEANGEVCT--YKHEFLQSSIRKGVALGFLLEDVWV-----
Bce        297 QASVYELLYGYELKQAIKEANGEVCT--YKHEFLQSSIRKGVALGFLLEDVWV-----
Bwe        297 QASVYELLYGYELKQAIKEANGEVCT--YKHEFLQSSIRKGVALGFLLEDVWV-----
Gth        298 RAAVYELLYGSAFDVINSLSGEGEQLRPHLILARIRKGHALGYLPTRALELVGGGER-----
Oih        285 RAAVYELLYGSAFDVINSLSGEGEQLRPHLILARIRKGHALGYLPTRALELVGGGER-----
BhaC125   299 RAAVYELLYGSAFDVINSLSGEGEQLRPHLILARIRKGHALGYLPTRALELVGGGER-----
BclKSMK16 280 RVAFYELLYGSAFDVINSLSGEGEQLRPHLILARIRKGHALGYLPTRALELVGGGER-----
    
```

# ComG<sub>B</sub>

**GSP11-F**

```
Bsu168      1  --MKKIRKRWMLKQARLLKRLGEMTAGSYLLDGRRLVLEQNNKQOADDTSVTCLEKRECAPFYQVVKLSLSPFKKAVGTCYFAETHGCLPASVLOSGBLL
BamFZB42   1  --MKKIRKRWMLKQARLLKRLGEMTAGSYLLDGRRLVLEQNNKQOADDTSVTCLEKRECAPFYQVVKLSLSPFKKAVGTCYFAETHGCLPASVLOSGBLL
Bli        1  --MKIKMKMPVREQADLLKRLGEMTAGSYLLDGRRLVLEQNNKQOADDTSVTCLEKRECAPFYQVVKLSLSPFKKAVGTCYFAETHGCLPASVLOSGBLL
BpuSAFR    1  MKKQASLAWSKKQDQAECHSLKRLGEMTAGSYLLDGRRLVLEQNNKQOADDTSVTCLEKRECAPFYQVVKLSLSPFKKAVGTCYFAETHGCLPASVLOSGBLL
Ban        1  ---MFKIWSLSDQVLLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
Bth        1  ---MFKIWSLSDQVLLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
Bce        1  ---MFKIWSLSDQVLLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
Bwe        1  ---MFKIWSLSDQVLLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
Gth        1  ---MFKIWSLSDQVLLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
BclKSMK16  1  --VGVREKQVQVQVPLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
BhaC125    1  --MFKIWSLSDQVLLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
Oih        1  --MFKIWSLSDQVLLKRLGELLKRGYLLQALEFTIRPOLLEKRVQVQRORIDGLKDKKSLHDSFHQLKRFHCEMLSYLYFAEOYGDTSFALQGSALL
```

**TMS**

```
Bsu168      100 ERKIAQADQKRVRYRPFLLFTTAAAFYMLQSGITIPQSGTYSNMNMTSRSDMLFAFFQHDLVILLVFTAGGTYVIVVFKKSAEAOQLICER
BamFZB42   100 QRKLMQTDQKRMRYRPFLLISSVCMFYLLSLLIIPQFAGTYSCNMNMTSGATFTTAFRRHFHEACALAAAFACFLYVWVFLKPKKSPQDKMLIVVK
Bli        99  HKFAQSEKHKAAARYRPFLLVLTVCVITSMKSAIVPOFAGTYSCNMNMTSFAATLTIYVDFHYLFFAGPLVAAAALSHLYLCSFRHKPPEKMAFLIT
BpuSAFR    101 EKKIQTDTFKRMRYRPFLLFTVSTVILVQOMIIPQFAGTYSCNMNMTSFAATLTIYVDFHYLFFAGPLVAAAALSHLYLCSFRHKPPEKMAFLIT
Ban        98  YKKDKYRDKMIMQYPMFLAIFLLIMLLEFNRIILPQVDMYSSFGSTABLETEQLLSTKILPYLIIISTFIIIVICGVYVYFRKLPBMQVAILLR
Bth        98  YKKDKYRDKMIMQYPMFLAIFLLIMLLEFNRIILPQVDMYSSFGSTABLETEQLLSTKILPYLIIISTFIIIVICGVYVYFRKLPBMQVAILLR
Bce        100 YKKDKYRDKMIMQYPMFLAIFLLIMLLEFNRIILPQVDMYSSFGSTABLETEQLLSTKILPYLIIISTFIIIVICGVYVYFRKLPBMQVAILLR
Bwe        100 YKKDKYRDKMIMQYPMFLAIFLLIMLLEFNRIILPQVDMYSSFGSTABLETEQLLSTKILPYLIIISTFIIIVICGVYVYFRKLPBMQVAILLR
Gth        98  AQKARLGEORLFRYPFLFALLVIMLLEMERLWLPQERTAAALSPNEQTSLLTALAHAFMAVVAASFAVSWLVAFAVFRWETAQLQSLDS
BclKSMK16  99  EKRAYTKKQCKMRYRPFLLWGAADVIVLQFVYPRKSLYARMBAPFWLTQFFQFITYPYMLTGLAVVGFASLWIKQKRRSHQQLAVPR
BhaC125    100 KREEMKRWKWEAMRYRPFLLIIVTLLMFLIMYFVLPQHOTVYRSIQLELPIVTKLITACSEKLPWFTAFFVVALVIVTVEHFFPETHKQVIVLR
Oih        98  QNRINWQKRFQVRYRPFLLIITFTLLIYFYVNOLESEQFLEVSD--LSPAIQLTFFVQWFSYFLIIVILLILLFIFWSVKKIISTDLKIKFISR
```

**GSP11-F**

```
Bsu168      200 IPLVCKLMLFNSYFSLQSLSSLSGLSYDSNAFKHTRLPFFYCEAEQIERLRKGESESAACGSHLYETLISKVISHGOLSRLDRELTYSQF
BamFZB42   200 IPLLKAAVDFNSYFSLQSLSSLSGLSYDSNAFKHTRLPFFYCEAEQIERLRKGESESAACGSHLYETLISKVISHGOLSRLDRELTYSQF
Bli        199  IPLLQTFDFNSYFSLQSLSSLSGLSYDSNAFKHTRLPFFYCEAEQIERLRKGESESAACGSHLYETLISKVISHGOLSRLDRELTYSQF
BpuSAFR    201 FFFIKGMRIMHYFSLQSLSSLSGLSYDSNAFKHTRLPFFYCEAEQIERLRKGESESAACGSHLYETLISKVISHGOLSRLDRELTYSQF
Ban        198  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
Bth        198  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
Bce        200  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
Bwe        200  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
Gth        197  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
BclKSMK16  199  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
BhaC125    200  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
Oih        195  IPLVKTFLIKKSHNFATQSLGSLGGLSVLEALIMMEOKYVPPFFQYEAERQLIAGEPFCSTIAKSYEYEBELSYIITHGOANGNLAIELGYSDL
```

**N-met**

```
Bsu168      300  ILQRLERKAKQWTGELQPMIYGFVAAMILLVYISMLVPMVQMMNQV
BamFZB42   300  MMERLEQNAARYTGELQPMIYGFVAAMILLVYISMLVPMVQMMNQV
Bli        299  LMERLEKAKQWTGELQPMIYGFVAAMILLVYISMLVPMVQMMNQV
BpuSAFR    301  LLENABLKIKKMWVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
Ban        298  IMEKMERKIKKMLVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
Bth        298  IMEKMERKIKKMLVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
Bce        300  IMEKMERKIKKMLVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
Bwe        300  IIEKVEKIKKMLVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
Gth        297  LLOVNEKIKKMLVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
BclKSMK16  299  LYCEMDRILKQVIVVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
BhaC125    200  QIEEFERVKKMLVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
Oih        295  SIEDVYQLNKQVIVVLIQPIILFICIGGIIVVLMYLAMIMPFOQMNNSI
```

# ComG<sub>C</sub>

**N-met**

```
Bsu168      1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKAADVCPNGKRIHIT--GG
BamFZB42   1  --VLRKRNQDGFLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
Bli        1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
BpuSAFR    1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
Ban        1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
Bth        1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
Bce        1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
Bwe        1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
Gth        1  -----MNEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
BclKSMK16  1  VRIKIKKDEKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
BhaC125    1  --MKKPFHKGKRGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
Oih        1  ---MFKKESGFTLEMLIVMVISLILLIIPNVTKRNSQCHKKCGEGLQNVVRAQVTAELDDEGGQPSLAFLOSRYV--KKNLACPNKRIHIT--GG
```

```
Bsu168      93  EVKVEH-
BamFZB42   97  EVTVFQ-
Bli        93  KAQAD--
BpuSAFR    93  HKKHE--
Ban        93  TVTSGKL
Bth        93  TVTSGKL
Bce        93  AVSTRAL
Bwe        93  TVSVGKS
Gth        92  DVSEIGS
BclKSMK16  99  VVIVSG-
BhaC125    97  VVALSE-
Oih        96  RVDAS--
```

## ComG<sub>D</sub>

← N-met →

Bsu168	1	MNLIKNEKGFLLLESLLVLSLASVLLVAVFTTTPPAPDNTAARCAASOLKNDILTQCTAISRQQRKTLFHRKKE--YQVI----GDVVEREYATG
BamFZB42	1	MNNRRRTENGFTLLLESLLVLSLASVLLVAVFTTTPPAPDNTAARCAASOLKNDILTQCTAISRQQRKTLFHRKKE--YQVI----GDVVEREYATG
Bli	1	-----MNRKGFLLLESLLVLSLASVLLVAVFTTTPPAPDNTAARCAASOLKNDILTQCTAISRQQRKTLFHRKKE--YQVI----GDVVEREYATG
BpuSAFR	1	MNKLIFHEKGFLLLESLLVLSLASVLLVAVFTTTPPAPDNTAARCAASOLKNDILTQCTAISRQQRKTLFHRKKE--YQVI----GDVVEREYATG
Ban	1	-----VKOKGFLLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
Bth	1	-----VKOKGFLLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
Bwe	1	-----VKOKGFLLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
Bce	1	-----VKOKGFLLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
Gth	1	-----LARNNGFTLLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
Oih	1	-----MPPSKGFLLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
BhaC125	1	-----VSSNGFTLLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
BclKSMK16	1	MLSSFRADDGFTLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS

  

Bsu168	94	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
BamFZB42	95	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
Bli	95	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
BpuSAFR	98	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
Ban	94	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
Bth	94	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
Bwe	94	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
Bce	94	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
Gth	95	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
Oih	93	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
BhaC125	94	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----
BclKSMK16	97	---SLELLILKDRREFNEKCHENAGGKIRVKGH-AVYDITVYLGSGRVNVERK-----

## ComG<sub>E</sub>

← SSeq →

Bsu168	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
BamFZB42	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
Bli	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
BpuSAFR	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
Ban	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
Bth	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
Bce	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
Bwe	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
Gth	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
BpuSAFR	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
BhaC125	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
Oih	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE
BclKSMK16	1	---MWRNKGFSTTEITMSALSLWLFVLLTVVPLMDKILMAIE-----KMAESRETYOMNESISKYVMSGE--GAASKTITKNNHIDAMWEEE

  

Bsu168	85	GEYQNVCHKAAYKE-----KSFCLSLIQTWTHAS----
BamFZB42	76	GEYQNVCHKAAYKE-----KSFCLSLIQTWTHAS----
Bli	85	GEYQNVCHKAAYKE-----KSFCLSLIQTWTHAS----
Ban	84	---EVCVMNKDVKKE-----KMBECCFYAGEKIN-----
Bth	84	---EVCVMNKDVKKE-----KMBECCFYAGEKIN-----
Bce	84	---EVCVMNKDVKKE-----KMBECCFYAGEKIN-----
Bwe	84	---EVCVMNKDVKKE-----KMBECCFYAGEKIN-----
Gth	83	GQMKVVCVRNRYVVE-----KMBECCFYAGEKIN-----
BpuSAFR	84	---EVCVMNKDVKKE-----KMBECCFYAGEKIN-----
BhaC125	83	LGAVQVCHRWVNRKE-----QAYEPCRLAKL-----
Oih	86	ESTLIGCAQWRNINY-----RDEEYCLHCKYOR-----
BclKSMK16	101	VIMAEVAVREKCEKGYGLSCLSTFENGKTTTRSFVAKDVAFNPHKK-----

## ComG<sub>F</sub>

← N-met →

Bsu168	1	-----KNHFASAFRCQNGYTLNMFSLVSLFLLSGSAAALHFLSRQ--QEHDCETQCEMSTIEGMNNECKS---QAVKTAEHGSLI
BamFZB42	1	MVSITKCVRLTAAKKCASAFSKQTFGLNMFSLVSLVCFLLSGTGLFHLHFLSNR--ADNGCDSREHQAIVQIMNECKQA---EAVKPADGGGRADA
Bli	1	MMKYASLLKKTAKGTVPFCGQRCTLNNMFSLVSLVCFLLSGTGLFHLHFLSNR--ADNGCDSREHQAIVQIMNECKQA---EAVKPADGGGRADA
Ban	1	-----MLLCRRKNKLEAGFTLLEMITVCFPLLSVFF--VLFPRLHFHIGEN--KQLKGLNDWEVDVFLGQVQLERBRVSVFVEKIVPENRESMLR
Bth	1	-----MLLCRRKNKLEAGFTLLEMITVCFPLLSVFF--VLFPRLHFHIGEN--KQLKGLNDWEVDVFLGQVQLERBRVSVFVEKIVPENRESMLR
Bce	1	-----MLLCRRKNKLEAGFTLLEMITVCFPLLSVFF--VLFPRLHFHIGEN--KQLKGLNDWEVDVFLGQVQLERBRVSVFVEKIVPENRESMLR
Bwe	1	-----MQRKNKREAGFTLLEMITVCFPLLSVFF--VLFPRLHFHIGEN--KQLKGLNDWEVDVFLGQVQLERBRVSVFVEKIVPENRESMLR
BpuSAFR	1	-----MNGECPFTLLEMITVCFPLLSVFF--VLFPRLHFHIGEN--KQLKGLNDWEVDVFLGQVQLERBRVSVFVEKIVPENRESMLR
Gth	1	-----VVRVQAVAGGSGFTLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
BhaC125	1	MRRVGLONCRGFTLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
Oih	1	-----MVRNCRGFTLLEMLLVLFASVLSVWTFNPHSHIYEKCKEQLRPFNSDILYMOOALNRQKHYLRWKKDHHMYVIGS--SDDLITKREYVDS
BclKSMK16	1	-----MWSLIFIKNERC-----ALPLPLLFFVLGGYFSYQVHYER-----ERLFEQCLQLELLEN-----ALQALALD

  

Bsu168	83	CINLSGDFRPIVHSM-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
BamFZB42	96	CRKTCGEEVREHYHKM-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
Bli	96	MTRNGDITVRYEYQTV-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
Ban	85	FRLRTGDEVYKLNH-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
Bth	85	FRLRTGDEVYKLNH-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
Bce	85	FRLRTGDEVYKLNH-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
Bwe	85	FRLRTGDEVYKLNH-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
BpuSAFR	78	LQKMNCRVPSLVKSN-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
Gth	82	LQKMNCRVPSLVKSN-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
BhaC125	79	LYQPSGDFRPIVHSM-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
Oih	76	PSKINQOTIYQVGNL-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--
BclKSMK16	63	ILASEGDFRPIVHSM-IRKRWDCGKGVPLDHTAKAD-ENGWLLKHESEDQKLYQTAFVYSVGGG--

# ComG<sub>G</sub>

		← SSeq →	
Bsu168	1	-----MYRTRGFYIPAVLFFVSALVLLIIVNEVAAQYVSRCMFEKSTKELYIGENLLONGVLLSIRHVLEER-RGQEGTQQLVGRVSYIHDTSIKR-Q	
BamFZB42	1	-----MYKSDGFYIPAVLFFVSAAVLLVSYTISDFITRRTFAKAGEYVIGENLLONGVLLSSRHVTQGG-RVQTGTQRFPVGTVSFHIIGSDRRE-T	
Bli	1	-----VKNSKGFYIPAVFAAVVCMLCASHSAALLTNORITFAQOTKEFHVQONLLONGVLLHALSHHEET-REKVQKQ---HGLISYSVS-PAHQQ-L	
Bce	1	-----MRKNGFTMPETLIFLVLTSFFVYENMLLETKKFTVTEOKPILLLELNRSITDKKDLQQ---REKEGTFPQVEKGEASGNVVFEND-L	
Bth	1	-----MRKQDGFIMPETLIFLVLTSFFVYENMLLIDKKEVTEOKPVLLELNRSITDKKDLKE---REKEDTFPQVEKGEASGNVVFEND-V	
Bwe	1	-----MRKQDGFAMPETLIFLVLTSFFVYENMLLIDKKEVTEOKPVLLELNRSITDKKDLKE---REKEDVFLQVEKGEVSGKYVFEND-V	
Ban	1	-----MRKQDGFIMPETLIFLVLTSFFVYENMLLIDKKEVTEOKPVLLELNRSITDKKDLKE---AKRKILFSVHMKRKRE-EITFLKT-M	
BpuSAFR	1	-----VGRGGFYIPQVAVVIFPFLITIGAVTIGFQKRLRSVELTISFVKKQCFRVGVSEAVTLLHLC-EKQHTPIDTEEGVTLKQVTCQDKKNR	
Gth	1	-----MCRQDGFVVEFLVAVVAVLAFSVTHIVLYEVEQAAHADRQAAEAELVQVAVFDVKEQALSDPTMTVREGVTVPRGTAVYRWEKEGETS	
Oih	1	MKKLSVFLKTEQGFLEIYVWISFLVVSLSLIIYHHQQIITFHLKQLQSETEFQMSRIDVEEDDELII--DTKSPKYVSPDGEVIRINKIEEHL	
BhaC125	1	----VKKLHNNGFALFLMLCLMLLAAASLYWGESLIMKRRIVQQQEVVMQLSLLVQVCKTLTEENG----GLENAVLRGQDVVAYQVRSSITIT	
Bsu168	92	KENVRSVDSCTERAAQIVDQKQKLRMTL---	
BamFZB42	92	VQVTLQAEITTGTRREAHLLDCKKQLIQMTL---	
Bli	88	TRVQVQVTLASCTETFAQVQDLNCKITQNKY---	
Bce	90	ILVTLQCTLKQEAFFVIGERRRKRKDKKIVDMIEG---	
Bth	90	ILVSLQCKIKQEVFYKVSERRRKRKDKKIFDMIEG---	
Bwe	90	ILVSLQCVIKQRFVYVVSERRRKRKDKKIFDMIEG---	
Ban	89	IFS-----HHSVSSSKFSIK---	
BpuSAFR	93	ISVLSVKKDLSDERELVLDGQTGEVRRANRE---	
Gth	94	VRVALSVRSSCFRRVVEVVMFLPALLETEVSEQND	
Oih	99	VTLTFIIFLDQEPFYVYEHFVKVNE-----	
BhaC125	92	ATLQVTAELGSCQRKVVVVSIDLAKGTIRDRVNP	