Short Genome Communications

Complete genome sequence of the nematicidal Bacillus thuringiensis MYBT18247

Jacqueline Hollensteinera, Anja Poehleinb, Cathrin Spröerb, Boyke Bunkb, Anna E. Sheppardc,1,
Philip Rosenstield, Hinrich Schulenburgc, Heiko Liesegang

a Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, University of Göttingen, Germany
b Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
c Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian-Albrechts University of Kiel, Kiel, Germany

ARTICLE INFO

Keywords:
Bacillus thuringiensis
Genome sequence
Nematicidal
Crystal toxins
Evolution

ABSTRACT

The Gram-positive spore forming bacterium Bacillus thuringiensis MYBT18247 encodes three cry toxin genes, (cry6Ba2, cry6Ba3 and cry21-1ike) which are active against nematodes. For a better understanding of the evolution of virulence and cry toxins, we present here the complete genome sequence of Bacillus thuringiensis MYBT18247. Various additional virulence factors such as bacteriocins, proteases and hemolysins were identified. In addition, the methylome and the metabolic potential of the strain were analyzed and the strain phylogenetically classified.

Bacillus thuringiensis is a ubiquitous, Gram-positive, spore-forming, bacterium (Schnepf et al., 1998). Strains of the species are used as a biopesticide because of their ability to produce parasporal protein crystals (Bechtel and Bulla, 1976; Ibrahim et al., 2010). These protein crystals consist of δ-endotoxins which are active against a broad spectrum of invertebrates including species of the orders Lepidoptera, Diptera, Coleoptera, Hymenoptera, Homoptera, Orthoptera, Mallophaga as well as mites, protozoa and nematodes (Feitelson, 1993; Schnepf et al., 1998). Here we present the annotated genome sequence of B. thuringiensis MYBT18247 that was isolated and used for single as well as multiple infection co-evolution experiments within the nematode Caenorhabditis elegans (Masri et al., 2015; Schulte et al., 2010).

Genomic DNA was isolated using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) and the Genomic-Tip 100 G Kit (Qiagen, Hilden, Germany). For 454 pyrosequencing the genomic DNA was sheared (∼700 bp), end repaired and universal barcode sequencing adaptors were ligated (Rapid-GSFLX Titanium, Roche 454, Branford CT). The library preparation was done with the GS Titanium Sequencing Kit XLR70t (Roche 454, Branford CA). Illumina sequencing DNA libraries were generated with the Nextera XT DNA Library Prep Kit (Illumina, San Diego, USA). For SMRT-sequencing (Pacific Biosciences, Menlo Park, USA) the C2/P4 chemistry was applied for three SMRT-cells and C4/P6 chemistry were used for two additional SMRT-cells, respectively. Whole-genome sequencing was performed using the 454 GS-FLX instrument, the Genome analyzer IIx (Illumina, San Diego, USA) and the PacBio RSII system (Pacific Biosciences, Menlo Park, USA). The 454 shotgun sequencing produced 335,141 single-end reads with an average read length of 420 bp. The Newbler 2.8 de novo assembler (Roche Diagnostics) assembled the reads into 411 contigs with a coverage of 18 x. A hybrid assembly was performed with Mira 4.0.3 (http://mira-assembler.sourceforge.net/docs/DefinitiveGuideToMIRA.html) by using 4,000,000 (112 bp) paired-end Illumina reads and 30,952 PacBio reads (C2-chemistry) with an average read length of 5053 bp. An HGAP 2.3.0 assembly using 67,045 PacBio reads (P6-chemistry) with a mean length of 13,802 bp resulted in an average coverage of 124.46 x. The assemblies were manually combined and contradictions were resolved by Sanger sequencing using BigDye 3.0 chemistry and an ABI3730XL capillary sequencer (Applied Biosystems, Life Technology GmbH, Darmstadt, Germany). All sequence positions were manually checked using Gap4 (v4.11) of the Staden package (Staden et al., 1999) to ensure the sequence quality.

Annotation was conducted with Prokka v.1.9 (Seemann, 2014). The initial annotation was performed using Bacillus thuringiensis Bt407 as a high quality species reference (Sheppard et al., 2013) and a comprehensive toxin protein database (including all Cry, Cyt, Vip, Sip toxins) as a feature reference set. The annotations of detected cry toxin genes were manually corrected and confirmed by Crickmore and deposited at...
The genome size of *B. thuringiensis* MYBT18247 is 6,138,199 bp with an average GC-content of 35% (Table 1). The data set consists of a chromosome of 5.6 Mbp and six plasmids which vary in size from 12.5 kb to 175 kb. One plasmid contig of 130 kb could not be circularized due to large repetitive regions at both contig ends. The chromosome encodes 6210 protein-coding and 156 RNA genes, including 14 rRNA clusters (5S, 16S, 23S rRNA). The plasmids encode 556 protein coding genes. There were 3727 genes assigned to COG database (Table 2) and 1114 genes were identified in MetaCyc metabolic pathways (Caspi et al., 2016). Three nematicidal cry toxin genes have been identified on plasmid p174778 (cry6Ba2; BTI247_60340), p174778 (cry6Ba3; BTI247_62380) and p15092 (cry21-like; BTI247_64000) (Table 3).

### Table 1

| Genome Feature | Value     |
|----------------|-----------|
| Genome size (bp) | 6,138,199 |
| DNA coding (bp)  | 5,143,010 |
| DNA G + C (bp)   | 2,159,739 |
| DNA scaffolds    | 7         |
| Total genes      | 6366      |
| Protein coding genes | 6210     |
| RNA genes        | 156       |
| rRNA genes       | 42        |
| Genes in internal clusters | 536    |
| Genes with function prediction | 4767 |
| Genes assigned to COG | 1443    |
| Genes assigned to Mobilome | 268    |
| Genes with Pfam domains | 5054  |
| Genes with signal peptide | 424    |
| Genes with transmembrane helices | 1692 |
| CRISPR repeats   | -         |

### Table 2

| COG categories of *B. thuringiensis* MYBT18247. |
|-----------------------------------------------|
| **Letter** | **Name** | **Count** | **Percent [%]** |
| E           | Amino acid transport and metabolism | 365       | 8.83          |
| G           | Carbohydrate transport and metabolism | 227       | 5.49          |
| D           | Cell cycle control, cell division, chromosome partitioning | 57        | 1.38          |
| N           | Cell motility | 52        | 1.26          |
| M           | Cell wall/membrane/envelope biogenesis | 216       | 5.22          |
| B           | Chromatin structure and dynamics | 1         | 0.02          |
| H           | Coenzyme transport and metabolism | 213       | 5.15          |
| Z           | Cytoskeleton | 2         | 0.05          |
| V           | Defense mechanisms | 111       | 2.69          |
| C           | Energy production and conversion | 194       | 4.69          |
| W           | Extracellular structures | 4         | 0.1           |
| S           | Function unknown | 295       | 7.14          |
| R           | General function prediction only | 368       | 8.9           |
| P           | Inorganic ion transport and metabolism | 226       | 5.47          |
| U           | Intracellular trafficking, secretion, and vesicular transport | 32        | 0.77          |
| F           | Lipid transport and metabolism | 138       | 3.34          |
| F           | Nucleotide transport and metabolism | 126       | 3.05          |
| O           | Posttranslational modification, protein turnover, chaperones | 156       | 3.77          |
| L           | Replication, recombination and repair | 190       | 4.6           |
| Q           | Secondary metabolites biosynthesis, transport and catabolism | 89        | 2.15          |
| T           | Signal transduction mechanisms | 185       | 4.48          |
| K           | Transcription | 364       | 8.81          |
| J           | Translation, ribosomal structure and biogenesis | 255       | 6.17          |
| Not in COG |                               | 2907      | 42.92         |

### Table 3

| Predicted virulence factors of *B. thuringiensis* MYBT18247. |
|---------------------------------------------------------------|
| **Virulence factor** | **No. of proteins** | **Locus,Taq** | **Location** |
| Cry toxins          | 3                   | BTI247_60340  | p174778     |
| Pesticidal crystal protein |           |               |             |
| Cry6Ba2             | BTI247_62380       | p113275      |
| Pesticidal crystal protein |               |               |             |
| Cry6Ba3             | BTI247_64000       | p15092       |
| Putative pesticidal crystal protein |         |               |             |
| Cry21-like          | BTI247_64000       | p174778      |
| Bacteriocins        | 4                   |               |             |
| Bacteriocin > 10 kDa | (bacteriocinIII)    |               |             |
| LA1 (linear azol(ine) containing peptides) | 5    |
| Head to tail cyclized bacteriocin | IId   |               |             |
| Head to tail cyclized bacteriocin | IId   |               |             |
| Proteases           | 4                   |               |             |
| Bacillolysin        | 4                   |               |             |
| npr1                | BTI247_07440       | chromosome   |
| npr2                | BTI247_24310       | chromosome   |
| nprM                | BTI247_29880       | chromosome   |
| npr3                | BTI247_36120       | chromosome   |
| Collagenases        | 6                   |               |             |
| colA1               | BTI247_06950       | chromosome   |
| colA2,1             | BTI247_32790       | chromosome   |
| colA4               | BTI247_36420       | chromosome   |
| colA3               | BTI247_37990       | chromosome   |
| colA5               | BTI247_46580       | chromosome   |
| colA2,2             | BTI247_59250       | chromosome   |
| Immune inhibitor A  | 5                   |               |             |
| inui1               | BTI247_08150       | chromosome   |
| inui2               | BTI247_08170       | chromosome   |
| inui1,2             | BTI247_15830       | chromosome   |
| inui3               | BTI247_32580       | chromosome   |
| inui2,2             | BTI247_33630       | chromosome   |
| Phospholipases      | 4                   |               |             |
| cerA                | BTI247_08230       | chromosome   |
| phospholipase, patatin family | plcA |             |
| plcA                | BTI247_22620       | chromosome   |
| ytpA                | BTI247_50630       | chromosome   |
| Chitinases          | 2                   |               |             |
| chaA                | BTI247_05680       | chromosome   |
| chiD                | BTI247_39810       | chromosome   |
| Camelysins          | 4                   |               |             |
| calF1,1             | BTI247_07820       | chromosome   |
| calF1,2             | BTI247_15800       | chromosome   |
| calF1,3             | BTI247_39500       | p174778      |
| calF1,4             | BTI247_63500       | p81952       |
| N-acetyl homoserine lactone | 1 |             |
| AiiA                | BTI247_36860       | chromosome   |
| Enterotoxins        |                    |               |             |
| Hemolysins          |                    |               |             |
| Hemolysin BL        |                    |               |             |
| hhlA1,1             | BTI247_21380       | chromosome   |
| hhlA2               | BTI247_21390       | chromosome   |
| hhlB                | BTI247_26830       | chromosome   |
| hhlA3               | BTI247_26840       | chromosome   |
| hhlA4,1             | BTI247_26850       | chromosome   |
| hhlA4,2             | BTI247_26880       | chromosome   |
| hhlA1,2             | BTI247_60570       | p130548      |
| hhlA1,3             | BTI247_61620       | p130548      |
| Hemolysin II        | 3                   |               |             |
| hly,1               | BTI247_37800       | chromosome   |
| hly,2               | BTI247_59850       | p174778      |
| hly,3               | BTI247_63990       | p15092       |
| Hemolysin III       | 2                   |               |             |
| hlyIII,1            | BTI247_24750       | chromosome   |

http://www.btnomenclature.info/(Crickmore et al., 1998).
proteases including four camelysins, six collagenases, and four phospholipases were identified and may play an important role in activating protoxins (Nisnevitch et al., 2010), destruction of the intestine of pathogenic microbes. In the B. thuringiensis MYBT18247 with antiSMASH3.0.5 (Weber et al., 2015) and BAGEL3 (van Heel et al., 2013) were used for identification.

Phylogenetic classification of B. thuringiensis MYBT18247 within the Bacillus cereus sensu lato group, was performed by multi-locus-sequence typing (MLST) according to Priest et al. (Priest et al., 2004) (Fig. 1). The strain clusters within a subgroup, comprising B. thuringiensis MYBT18246, B. thuringiensis YBT-1518, B. thuringiensis Bt407, and B. thuringiensis serovar chinenis CT-43. Strikingly, all members of the cluster except B. thuringiensis 407, which is an artificially cured laboratory strain, encode at least one nematicidal or insecticidal cry toxin gene.

Summarized, B. thuringiensis MYBT18247 comprises a variety of promising genes, including the rare cry6Ba which show high potential for the development of new biotechnological relevant nematicidal and insecticidal control agents.

Nucleotide sequence accession numbers

The whole genome sequence has been deposited at the DDBJ/EMBL/GenBank with the accession numbers CP015250.1-CP015256.1. The strain is available from DSMZ (Braunschweig, Germany) under accession 104068.

Acknowledgements

We thank Dr. Andrea Thürmer, Nicole Heyer and Simone Severitt for sequencing support. This project was funded by the German Loguercio, 2014; Kim et al., 2015). However, a test on Columbia agar revealed a hemolytic negative phenotype. This might correlate to the observation that the corresponding genome locus misses the hemolsyn B1 lytic component L2, one essential part of the tripartite toxin. The plasmids encode a number of genes dedicated to genome fluidity such as transposases, insertion sequences, transcriptional regulators and re-combinases.

The metabolic versatility of B. thuringiensis MYBT18247 has been evaluated using BIOLOG Phenotypic Microarrays (PM1-PM2). The strain utilized substrates assigned feeding into glycolysis, citrate cycle and pentose phosphate pathway. It metabolizes various carbon sources including sugars, sugar alcohols and sugar acids which have been found in fruits, fungi, insect compartments or plant compartments. Apparently, the strain is a generalist which is able to adapt to various ecological niches.

The methylome analysis of B. thuringiensis MYBT18247 was performed using the SMRT Portal v2.3.0 analysis platform. Four methylated N4-methyladenine (m6A) motifs were observed. The non-palindromic motif (CCTANNNNNNRTNC/GNAAYNNNNNNTAYG) was found to be methylated in more than 65% of all instances. Additionally, three N4-methylcytosin (m4C) motifs were detected with a methylation grade of 6–21% and in four putative motifs the identification of the methylation was not possible due to insufficient coverage. The SMRT data DNA methyltransferase recognition motifs are deposited in REBASE at http://rebase.neb.com/rebase/private/pacbio_Liesegang23.html (Roberts et al., 2015).

Table 3

| Virulence factor | No. of proteins | Locus_Taq | Location |
|------------------|----------------|-----------|----------|
| hlyII 2          | 1              | BT1247,58090 | chromosome |
| Hemolysin A      | 1              | BT1247,44490 | chromosome |
| Gamma hemolysin  | 1              | BT1247,13860 | Chromosome |
| Non-hemolysins   | 4              | BT1247,21370 | chromosome |
| nhe 1            | 1              | BT1247,60560p | p130548 |
| nhe 2            | 1              | BT1247,61630 | p130548 |
| Alveolysin       | 1              | BT1247,34170 | chromosome |

a all protein sequences of the genome were scanned with generic HMM models constructed from representative sequences of the known Cry/Cyt/VIP-toxins extracted from SwissProt and the Bt toxin database (http://www.bto.ncl.ac.uk). Identified sequences were characterized by a procedure described by the international committee for cry-Toxins (Crickmore et al., 1998), verified by Crickmore and deposited at the BT toxin database.

* AntiSMASH3.0.5 (Weber et al., 2015) and BAGEL3 (van Heel et al., 2013) were used for identification.

Table 4

Secondary metabolite clusters identified in B. thuringiensis MYBT18247 with antiSMASH3.0.

| Cluster | Type | Location | Start | Stop | Similarity to known cluster | MIBIG BGC-ID |
|---------|------|----------|-------|------|----------------------------|---------------|
| 1       | Other| chromosome | 496,393 | 539,974 | –                          | –             |
| 2       | Bacteriocin| chromosome | 1,444,119 | 1,457,984 | –                          | –             |
| 3       | Siderophore| chromosome | 2,120,707 | 2,134,414 | Petrobactin biosynthetic gene cluster (83% of genes show similarity) | BGC0000942_c1 |
| 4       | NRPS| chromosome | 2,433,678 | 2,483,403 | Bacillibactin biosynthetic gene cluster (46% of genes show similarity) | BGC0000309_c1 |
| 5       | Bacteriocin| chromosome | 2,666,725 | 2,677,045 | –                          | –             |
| 6       | NRPS| chromosome | 2,736,144 | 2,783,151 | –                          | –             |
| 7       | Bacteriocin| chromosome | 2,800,158 | 2,810,424 | –                          | –             |
| 8       | NRPS| chromosome | 3,472,892 | 3,532,849 | –                          | –             |
| 9       | Terpene| chromosome | 3,659,931 | 3,681,784 | Molybdenum cofactor biosynthetic gene cluster (11% of genes show similarity) | BGC0000916_c1 |
| 10      | Bacteriocin| chromosome | p174778 | 132,176 | –                          | –             |
| 11      | Alveolysin| chromosome | p81952 | 56,165 | –                          | –             |
Research Foundation (DFG-SPP1399, Grant Li 1690/2-1 to HL; SCHU 1415/9 to HS, and RO 2994/3 to PR). We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the University of Goettingen.

References

Argolo-Filho, R.C., Loguercio, L.L., 2014. Bacillus thuringiensis is an environmental pathogen and host-specificity has developed as an adaptation to human-generated ecological niches. Insects 5, 62–91. http://dx.doi.org/10.3390/insects5010062.

Bechtel, D.B., Bulla, L.A., 1976. Electron Microscope Study of Sporulation and Parasporal Crystal Formation in Bacillus thuringiensis. J. Bacteriol. 127, 1472–1481.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 1–9. http://dx.doi.org/10.1186/1471-2105-10-421.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., et al., 2015. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285. http://dx.doi.org/10.1093/nar/gkv1164.

Dean, D.H., 1998. Revision of the nomenclature for the Bacillus thuringiensis pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. 62, 775–797. http://dx.doi.org/10.1128/MMBR.62.4.775-797.1998.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., et al., 2015. Various enterotoxin and other virulence factor genes widespread among Bacillus cereus and Bacillus thuringiensis strains. J. Microbiol. Biotechnol. 25, 672–879. http://dx.doi.org/10.4161/mmb.30502.

Kim, M.-J., Han, J.-K., Park, J.-S., Lee, J.-S., Lee, S.-H., Cho, J.-I., Kim, K.-S., 2015. Various enterotoxin and other virulence factor genes widespread among Bacillus cereus and Bacillus thuringiensis strains. J. Microbiol. Biotechnol. 25, 672–879. http://dx.doi.org/10.4161/mmb.30502.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 7, 1870–1874. http://dx.doi.org/10.1093/molbev/msw114.

Lereclus, D., Baum, J., Schlegel, H., 2001. Bacillus thuringiensis: a genomics and proteomics perspective. Bioeng. Bugs 1, 31–50. http://dx.doi.org/10.4161/bbug.1.1.5019.
