**Standards in
Genomic Sciences**

EXTENDED GENOME REPORT   **Open Access**

CrossMark

Complete Genome sequence of the nematicidal *Bacillus thuringiensis* MYBT18246

Jacqueline Hollensteiner[1], Anja Poehlein[1], Cathrin Spröer[2], Boyke Bunk[2], Anna E. Sheppard[3,4], Philip Rosentstiel[5], Hinrich Schulenburg[4] and Heiko Liesegang[1*]

Abstract

Bacillus thuringiensis is a rod-shaped facultative anaerobic spore forming bacterium of the genus *Bacillus*. The defining feature of the species is the ability to produce parasporal crystal inclusion bodies, consisting of δ-endotoxins, encoded by *cry*-genes. Here we present the complete annotated genome sequence of the nematicidal *B. thuringiensis* strain MYBT18246. The genome comprises one 5,867,749 bp chromosome and 11 plasmids which vary in size from 6330 bp to 150,790 bp. The chromosome contains 6092 protein-coding and 150 RNA genes, including 36 rRNA genes. The plasmids encode 997 proteins and 4 t-RNA's. Analysis of the genome revealed a large number of mobile elements involved in genome plasticity including 11 plasmids and 16 chromosomal prophages. Three different nematicidal toxin genes were identified and classified according to the Cry toxin naming committee as *cry*13Aa2, *cry*13Ba1, and *cry*13Ab1. Strikingly, these genes are located on the chromosome in close proximity to three separate prophages. Moreover, four putative toxin genes of different toxin classes were identified on the plasmids p120510 (Vip-like toxin), p120416 (Cry-like toxin) and p109822 (two Bin-like toxins). A comparative genome analysis of *B. thuringiensis* MYBT18246 with three closely related *B. thuringiensis* strains enabled determination of the pan-genome of *B. thuringiensis* MYBT18246, revealing a large number of singletons, mostly represented by phage genes, morons and cryptic genes.

Keywords: *Bacillus thuringiensis*, *Bacillus cereus* sensu *lato*, Prophages, Parasporal crystal protein, Pan-Core-genome

Introduction

Bacillus thuringiensis is an ubiquitously distributed, rod-shaped, Gram-positive, spore forming, facultative anaerobic bacterium [1, 2]. *Bacillus thuringiensis* has been isolated from various ecological niches, including soil, aquatic habitats, phylloplane and insects [3–7]. The defining property of the species is the ability to produce parasporal protein crystals consisting of δ-endotoxins, which are predominantly encoded on plasmids [1, 8, 9]. These proteins are toxic towards a wide spectrum of invertebrates of the orders *Lepidoptera, Diptera, Coleoptera, Hymenoptera, Homoptera, Orthoptera, Mallophaga* and other species like *Gastropoda*, mites, protozoa and especially nematodes [7, 10–12]. In addition, *B. thuringiensis* produce additional toxins such as Cyt, Vip, and Sip toxins [13]. Cry toxins represent the largest group

and can be subdivided into three different homology groups. In total, over 787 different Cry toxins have been identified, each exhibiting toxicity against a specific host organism [14]. It has been shown that *B. thuringiensis* strains can produce more than one Cry toxin resulting in a broad host range. As such, *B. thuringiensis* has been used widely as a biopesticide in agriculture for several decades [1, 2, 8, 13, 15, 16]. *Bacillus thuringiensis* is a member of the genus *Bacillus*, which are low GC-content, Gram-positive bacteria with a respiratory metabolism and the ability to form heat- and desiccation-resistant endospores [11, 17, 18]. Within this genus, *B. thuringiensis* is a member of the *Bacillus cereus* sensu *lato* species group which originally contained seven different species (*B. cereus, B. anthracis, B. thuringiensis, B. mycoides, B. pseudomycoides, B. weihenstephanensis, B. cytotoxicus* [17–25]). Historically, most pathogenic and phenotypic properties were used for strain classification. However, recent publications utilizing genomic criteria suggest that the species group should be extended by species *B. toyonensis* [26, 27]. Moreover, the three proposed species

* Correspondence: hlieseg@gwdg.de
[1]Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, University of Göttingen, Göttingen, Germany
Full list of author information is available at the end of the article

Hollensteiner *et al. Standards in Genomic Sciences* (2017) 12:48

Page 2 of 10

"Bacillus gaemokensis"[28], "Bacillus manliponensis"[29] and "Bacillus bingmayongensis" [30] have been isolated and effectively published. However, these names had not yet appeared on a Validation List at the time of pulbication [31]. Due to the very close phylogenetic relationships, it has also been proposed to assign the eleven species to a single extended Bcsl species [32, 33]. The genome of Bcsl-members contains a highly conserved chromosome with regard to gene content, sequence similarity and genome synteny, while variation can be observed within mobile genomic elements such as prophages, insertion elements, transposons, and plasmids [34]. Due to the significance of Bcsl group members in human health, the food industry and agriculture, resolving the phylogeny is of great importance. Because of the highly conserved 16S rRNA-genes, the classical 16S phylogeny of Bcsl strains is inconclusive. Thus, a combination of 16S and a seven gene multi-locus sequence typing scheme have been used to establish taxonomic relationships within species of the Bcsl-group [35, 36]. Comparative genomics of the *cry*-gene loci has revealed remarkable proximity to elements of genome plasticity such as plasmids, transposons, insertion elements and prophages [2, 37–39]. The activity of these mobile elements has resulted in a magnitude of highly diverse plasmid sizes through rearrangements such as deletions and insertions, as well as migration of *cry*-genes into the bacterial chromosome [40]. The worldwide distribution of *B. thuringiensis* and its capacity to adapt to a diverse spectrum of invertebrate hosts is explained by the formation of spores and a remarkable variability in crystal protein families [13]. This toxin arsenal, especially the copy number of individual toxin genes, can be shaped by reciprocal co-adaptation with a nematode host, as previously demonstrated using controlled evolution experiments in the laboratory [41, 42]. The *B. thuringiensis* strain MYBT18246 described herein and its host *Caenorhabditis elegans* have been selected as a model system for such co-evolution experiments [41]. One aim of this sequencing project was to provide a high-quality reference genome sequence for the original *B. thuringiensis* MYBT18246 in order to obtain a detailed phylogeny and shed light on the evolution of this microparasite, with a particular focus on the presence of virulence factors, elements of genome plasticity and host adaptation factors. Here we present the genome of the nematicidal *B. thuringiensis* MYBT18246 and its comparative analysis to the three closest relatives identified by MLST phylogeny.

Organism information
Classification and features

Bacillus thuringiensis belongs to the genus *Bacillus* and has been isolated in the end of the nineteenth century [17, 20] and used as a biocontrol agent for several decades [7, 18, 21]. The strain *B. thuringiensis* MYBT18246 is a Gram-positive, rod-shaped and spore forming bacterium (Fig.1a), as most *B. thuringiensis* [7]. *Bacillus thuringiensis* MYBT18246 was isolated in the Schulenburg lab by AS from a mixture of genotypes present in the strain NRRL B-18246, originally provided by the Agricultural Research Service Patent Culture Collection (United States Department of Agriculture, Peoria, IL, USA) [43–45]. As a member of the species *B. thuringiensis*, *B. thuringiensis* MYBT18246 is facultative anaerobe, motile and is able to produce parasporal crystal toxins, which is the characteristic feature of this species [2]. Growth occurred at temperatures ranging from 10 to 48 °C and optimal growth was monitored at mesophil temperatures ranging from 28 to 37 °C [46]. The pH range of *B. thuringiensis* strains varies from pH 4.9 to 8.0, with the optimum documented as pH 7 [47, 48]. Strain *B. thuringiensis* MYBT18246 exhibits flat, opaque colonies with undulate, curled margins and produced crystals during the stationary phase (Fig. 1a-b). Characteristic features of *B. thuringiensis* MYBT18246 are listed in Table 1.

Extended feature descriptions

The cell size of *Bacillus thuringiensis* can vary from 0.5×1.2 µm - 2.5×10 µm [11]. Categorization into the
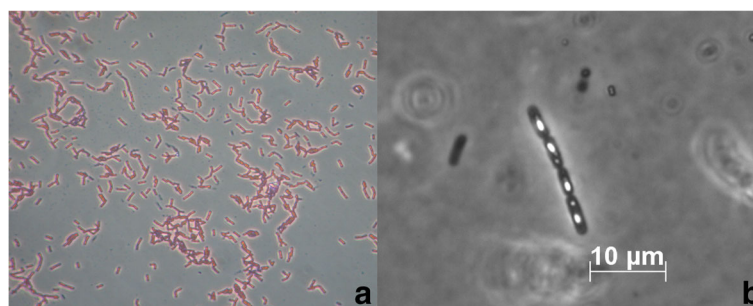
Fig. 1 Microscopic characteristics of *Bacillus thuringiensis* MYBT18246. **a** Light microscope analysis of Gram stained *B. thuringiensis* MYBT18246 cells (40×). **b** Phase contrast microscope analysis of sporulated and Cry-toxin producing cells of *B. thuringiensis* MYBT18246 (40×)

Hollensteiner *et al. Standards in Genomic Sciences* (2017) 12:48

Page 3 of 10

Table 1 Classification and general features of *B. thuringiensis* MYBT18246 [54]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [86]
		Phylum *Firmicutes*	TAS [47]
		Class *Bacilli*	TAS [87, 88]
		Order *Bacillales*	TAS [18, 89]
		Family *Bacillaceae*	TAS [18, 90]
		Genus *Bacillus*	TAS [17, 18]
		Species *Bacillus thuringiensis*	TAS [46]
		Strain MYBT18246	IDA
	Gram stain	positive	IDA
	Cell shape	rod-shaped	IDA
	Motility	Motile	TAS [46]
	Sporulation	Spore-forming	IDA
	Temperature range	10–48 °C	TAS [46]
	Optimum temperature	28–37 °C	TAS [46]
	pH range; Optimum	4.9–8.0; 7.0	TAS [47, 48]
	Carbon source	Organic carbon source	NAS
MIGS-6	Habitat	Worldwide	TAS [7]
MIGS-6.3	Salinity	Salt tolerant	TAS [7]
MIGS-22	Oxygen requirement	Aerobic, facultative anaerobic	TAS [11]
MIGS-15	Biotic relationship	Free-living, microparasite of *C. elegans*	TAS [41]
MIGS-14	Pathogenicity	Nematode pathogen	TAS [41]
MIGS-4	Geographic location	not reported	
MIGS-5	Sample collection	not reported	
MIGS-4.1	Latitude	unreported	
MIGS-4.2	Longitude	unreported	
MIGS-4.4	Altitude	unreported	

[a]Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project

group of Gram-positive organisms was confirmed by Gram staining, as shown in Fig. 1a. In Fig. 1b the production of Cry toxins can be observed. These toxins accumulate during the sporulation phase next to the endospore and build phase-bright inclusions [7]. *Bacillus thuringiensis* MYBT18246 exhibited 99% 16S rRNA sequence identity to other published Bcsl-members [49]. As a result of the high sequence similarity, a phylogenetic differentiation of *B. thuringiensis* MYBT18246 based on 16S phylogenetic differentiation of Bcsl group members is impossible (Fig. 2a). As an alternative, 23 *B. thuringiensis* strains, and a representative of each of the Bcsl group

species were chosen for phylogenetic analysis using multi-locus sequence typing as previously developed by Priest [36] (Fig. 2b). *Bacillus subtilis subsp. subtilis* str. 168 was selected as an outgroup to root the tree [17, 18]. The phylogenies were generated using the Neighbor-Joining method [50] and evolutionary distances were computed by the Maximum Composite Likelihood method [51]. In total, 217 MLST gene sequences were compared with 1000 bootstrap replicates. Phylogenetic analysis was conducted in MEGA7 [52]. All used reference sequences were retrieved from GenBank hosted at NCBI.

Genome sequencing information
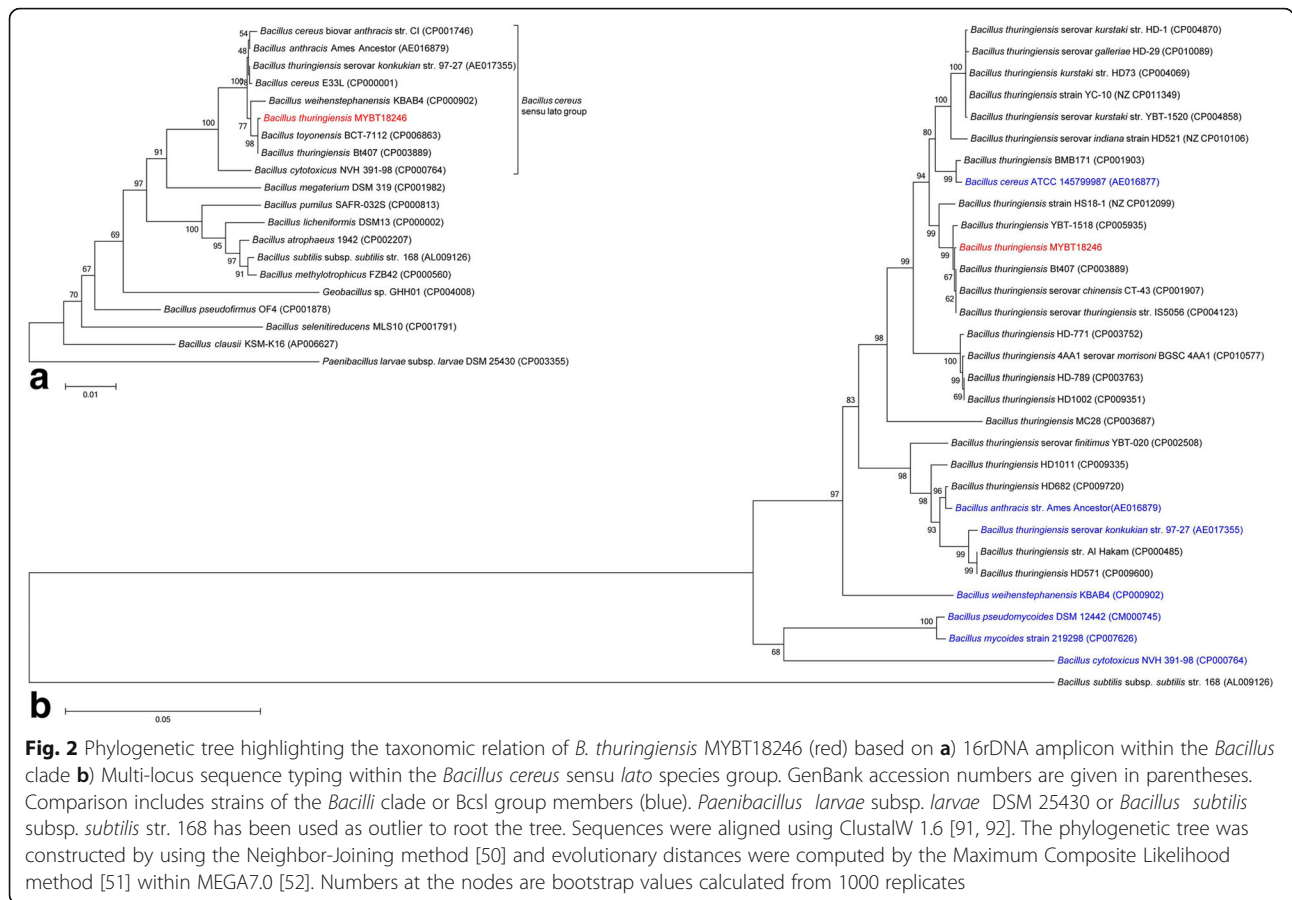### Genome project history

Bacillus thuringiensis MYBT18246 was used in a co-evolution study with a *Caenorhabditis elegans* host. The original strain MYBT18246 was selected for sequencing in order to generate a reliable reference sequence for subsequent experiments [41, 42]. The genome sequence was analyzed to identify virulence factors and fitness factors contributing to the efficient infection of *C. elegans*. Additionally, the phylogenetic position of *B. thuringiensis* MYBT18246 in the Bcsl group was determined [41]. The complete genome sequence has been deposited in GenBank with the accession numbers (CP015350-CP015361) and in the integrated Microbial Genomes database with the Taxon ID 2671180122 [53]. A summary of the project information and its association with MIGS version 2.0 compliance [54] is shown in Table 2.

Growth conditions and genomic DNA preparation

Genomic DNA was isolated from *B. thuringiensis* MYBT18246 using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) for 454 pyrosequencing [55] and the Genomic-Tip 100/G Kit (Qiagen, Hilden, Germany) for Single Molecule real-time sequencing [56] according to the manufacturer's instructions. For SMRT-sequencing the procedure and Checklist: Greater than 10 kb Template Preparation Using AmPure PB Beads was used and blunt end ligation was applied overnight. Whole-genome sequencing was performed using a 454 GS-FLX system (Titanium GS70 chemistry; Roche Life Science, Mannheim, Germany) and on one SMRT Cell on the PacBio *RSII* system using P6-chemistry (Pacific Biosciences, Menlo Park, CA, USA).

Genome sequencing and assembly

A summary of the project information can be found in Table 2. 454-pyrosequencing was carried out at the Institute of Clinical Molecular Biology in Kiel, Germany and SMRT-sequencing at the DSMZ Braunschweig. First, approximately 331,000,454-reads with an average length of 600 bp were assembled using the Newbler 2.8 de novo assembler (Roche Diagnostics), resulting in 729

Hollensteiner *et al. Standards in Genomic Sciences* (2017) 12:48

Page 4 of 10

Fig. 2 Phylogenetic tree highlighting the taxonomic relation of *B. thuringiensis* MYBT18246 (red) based on **a**) 16rDNA amplicon within the *Bacillus* clade **b**) Multi-locus sequence typing within the *Bacillus cereus* sensu *lato* species group. GenBank accession numbers are given in parentheses. Comparison includes strains of the *Bacilli* clade or Bcsl group members (blue). *Paenibacillus larvae* subsp. *larvae* DSM 25430 or *Bacillus subtilis* subsp. *subtilis* str. 168 has been used as outlier to root the tree. Sequences were aligned using ClustalW 1.6 [91, 92]. The phylogenetic tree was constructed by using the Neighbor-Joining method [50] and evolutionary distances were computed by the Maximum Composite Likelihood method [51] within MEGA7.0 [52]. Numbers at the nodes are bootstrap values calculated from 1000 replicates

Table 2 Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Complete
MIGS-28	Libraries used	Two genomic libraries: 454 pyrosequencing shotgun library, PacBio library
MIGS 29	Sequencing platforms	454 GS FLX system, PacBioRSII
MIGS 31.2	Fold coverage	18 × 454; 50 × PacBio
MIGS 30	Assemblers	Newbler 2.8; HGAP v2.3.0
MIGS 32	Gene calling method	Prodigal 2.6
	Locus Tag	BT246
	Genbank ID	CP015350-CP015361
	GenBank Date of Release	2016–07–15
	GOLD ID	Gp0020852
	BIOPROJECT	PRJNA290307
MIGS 13	Source Material Identifier	Department of Evolutionary Ecology and Genetics, CAU, Kiel
	Project relevance	Evolution

contigs with a coverage of 18 x. Repeats were resolved and gaps between contigs were closed using PCR with Sanger sequencing of the products with BigDye 3.0 chemistry and an ABI3730XL capillary sequencer (Applied Biosystems, Life Technology GmbH, Darmstadt, Germany). Manually editing in Gap4 (version 4.11) software of the Staden package [57] was performed to improve the sequence quality. For final gap closure PacBio sequencing was used. A total of 27,870 PacBio reads with a mean length of 14,053 bp were assembled using HGAP 2.0 [58], resulting in a coverage of 50 x, with further analysis using SMRT Portal (v2.3.0) [59]. Finally, both assemblies were combined, resulting in 12 contigs including a closed circular chromosome sequence of 5,867,749 bp. Eight additional contigs exhibited overlapping ends and were circularized to plasmid sequences ranging from 6.3 kb to 150 kb (Table 3). The assembly was checked for coverage drop downs and extremes of disparities including GC, AT, RY, and MK. Moreover, we determined the origin of replication of *B. thuringiensis* MYBT18246 by comparative analysis with OriC of eight other *B. thuringiensis* strains available in DoriC [60, 61]. These strains varied in chromosome size from 5.2 Mb to 5.8 Mb but all shared a similar GC-

Table 3 Summary of genome: one chromosome and 11 plasmids

Label	Size (Mb)	Topology	INSDC identifier	RefSeq ID
Chromosome	5.8	Circular	CP015350	NZ_CP015350.1
Plasmid 1	0.151	Circular	CP015351	NZ_CP015351.1
Plasmid 2	0.142	Circular	CP015352	NZ_CP015352.1
Plasmid 3	0.121	Circular	CP015353	NZ_CP015353.1
Plasmid 4	0.120	Circular	CP015354	NZ_CP015354.1
Plasmid 5	0.110	Circular	CP015355	NZ_CP015355.1
Plasmid 6	0.101	Circular	CP015356	NZ_CP015356.1
Plasmid 7	0.055	Linear	CP015357	NZ_CP015357.1
Plasmid 8	0.047	Circular	CP015358	NZ_CP015358.1
Plasmid 9	0.017	Linear	CP015359	NZ_CP015359.1
Plasmid 10	0.014	Linear	CP015360	NZ_CP015360.1
Plasmid 11	0.006	Circular	CP015361	NZ_CP015361.1

content of 35%. In total, including *B. thuringiensis* MYBT18246, two OriC regions were identified using the ORF-Finder [62]. One region was highly conserved with regard to OriC length (178/179 nt), OriC AT content (~0.69) and number of DnaA boxes (4). The second region varied in OriC length (564–767 nt) and OriC AT content (~0.67–0.7), but all had the same number of DnaA boxes (9). *B. thuringiensis* MYBT18246 showed the highest OriC similarities with both OriC regions of *B. thuringiensis* Bt407.

Genome annotation

Annotation was performed with Prokka v1.9 [63] using the manually curated *Bacillus thuringiensis* strain Bt407 [64] as a species reference and a comprehensive toxin protein database (including Cry, Cyt, Vip, and Sip toxins) as feature references. The Prokka pipeline was applied using prodigal for gene calling [65]. RNAmmer 1.2 [66] and Aragorn [67] were used for rRNA gene and t-RNA identification, respectively. Additionally, signal leader peptides were identified with SignalP 4.0 [68] and non-coding RNAs with an Infernal 1.1 search against the Rfam database [69]. Annotation of *cry* toxin genes were manually corrected and named according to the standards of the Cry toxin nomenclature by Crickmore [70]. Identified toxins were deposited at the *Bacillus thuringiensis* Toxin nomenclature database [14].

Genome properties

The genome of *B. thuringiensis* MYBT18246 consists of 12 replicons with a circular chromosome of 5,867,749 bp (Table 3). The GC content of the chromosome is 35% and the GC content of the plasmids ranges from 32 to 37%. The total number of protein coding genes is 7089 with 6092 genes on the chromosome and 997 genes on the plasmids. The genome harbors 12 rRNA clusters, 111 t-RNA genes, 5274 predicted protein-coding genes

with assigned function and 1815 genes encoding proteins with unknown function (Table 4). All gene products have been assigned to COGs (Table 5). The genome sequence of *B. thuringiensis* MYBT18246 is available in GenBank (CP015350 for the chromosome and CP015351 - CP015361 for the plasmids).

Insights from the genome sequence

To investigate the phylogeny of *B. thuringiensis* MYBT18246 two approaches were used. First, nineteen *Bacillus* strains were chosen for 16S rRNA analysis within the *Bacillus* clade (Fig. 2a). The 16S rRNA phylogeny shows that *B. thuringiensis* MYBT18246 clusters with other Bcsl group members within the *Bacillus* clade. However, the low bootstrap values confirm the limitations of 16S rRNA as a discriminatory marker within the Bcsl species group. Second, we applied an MLST approach based on the scheme by Priest et al. [36]. This revealed that MYBT18246 clusters with the toxin cured *B. thuringiensis* Bt407, insecticidal *B. thuringiensis* serovar *chinensis* CT-43, and with the nematicidal *B. thuringiensis* YBT-1518 within the Bcsl phylogeny (Fig. 2b). Based on this phylogeny and the phenotypic defining feature of the *B. thuringiensis* species group (the ability to produce crystal toxins against invertebrates and nematodes), the strain *B. thuringiensis* MYBT18246 can be safely classified as nematicidal *B. thuringiensis*.

For a detailed analysis of encoded toxins in *B. thuringiensis* MYBT18246, we generated a local database consisting of all available Cry, Cyt, Vip and Sip protein sequences from UniProtKB [71] and GenBank [72]. The database was curated to generate a set of non-redundant reference toxins. In total, we identified three different

Table 4 Genome statistics

Attribute	Value	% of Total
Genome size (bp)	6,752,488	100
DNA coding (bp)	5,623,665	83.28
DNA G + C (bp)	2,389,665	35.39
DNA scaffolds	12	100
Total genes	7239	100
Protein coding genes	7089	97.9
RNA genes	151	2.09
Genes in internal clusters	2694	37.22
Genes with function prediction	5274	72.86
Genes assigned to COGs	4662	64.40
Genes with Pfam domains	5503	76.02
Genes with signal peptides	500	6.91
Genes with transmembrane helices	1863	25.74
CRISPR repeats	0	0

Hollensteiner *et al. Standards in Genomic Sciences* (2017) 12:48

Page 6 of 10

Table 5 Number of protein encoding genes associated with general COG functional categories

Code	Value	%	Description
J	226	3.19	Translation, ribosomal structure and biogenesis
A	0	0	RNA processing and modification
K	487	6.87	Transcription
L	625	8.82	Replication, recombination and repair
B	1	0.01	Chromatin structure and dynamics
D	59	0.83	Cell cycle control, Cell division, chromosome partitioning
V	141	1.99	Defense mechanisms
T	218	3.08	Signal transduction mechanisms
M	270	3.81	Cell wall/membrane biogenesis
N	64	0.90	Cell motility
U	79	1.12	Intracellular trafficking and secretion
O	121	1.71	Posttranslational modification, protein turnover, chaperones
C	214	3.02	Energy production and conversion
G	263	3.71	Carbohydrate transport and metabolism
E	420	5.93	Amino acid transport and metabolism
F	128	1.81	Nucleotide transport and metabolism
H	177	2.50	Coenzyme transport and metabolism
I	129	1.82	Lipid transport and metabolism
P	243	3.43	Inorganic ion transport and metabolism
Q	83	1.17	Secondary metabolites biosynthesis, transport and catabolism
R	653	9.22	General function prediction only
S	505	7.13	Function unknown
-	1978	27.9	Not in COGs

[a]The total number is based on the total number of protein coding genes in the genome

cry toxin genes in the *B. thuringiensis* MYBT18246 genome and classified them as *cry*13Aa2 (>95%), *cry*13Ba1 (<78%) and *cry*13Ab1 (<95%), based on the similarity scheme from the Cry-toxin naming committee by Crickmore [13, 70]. Notably, these *cry* toxin genes are encoded on the chromosome and not on extra-chromosomal elements as has been previously reported for the vast majority of *cry* toxin genes [7, 73, 74]. The toxin gene analysis revealed four additional putative toxin-like genes on plasmids with sequence similarity to *cry* genes and *vip* genes. A Pfam domain analysis using
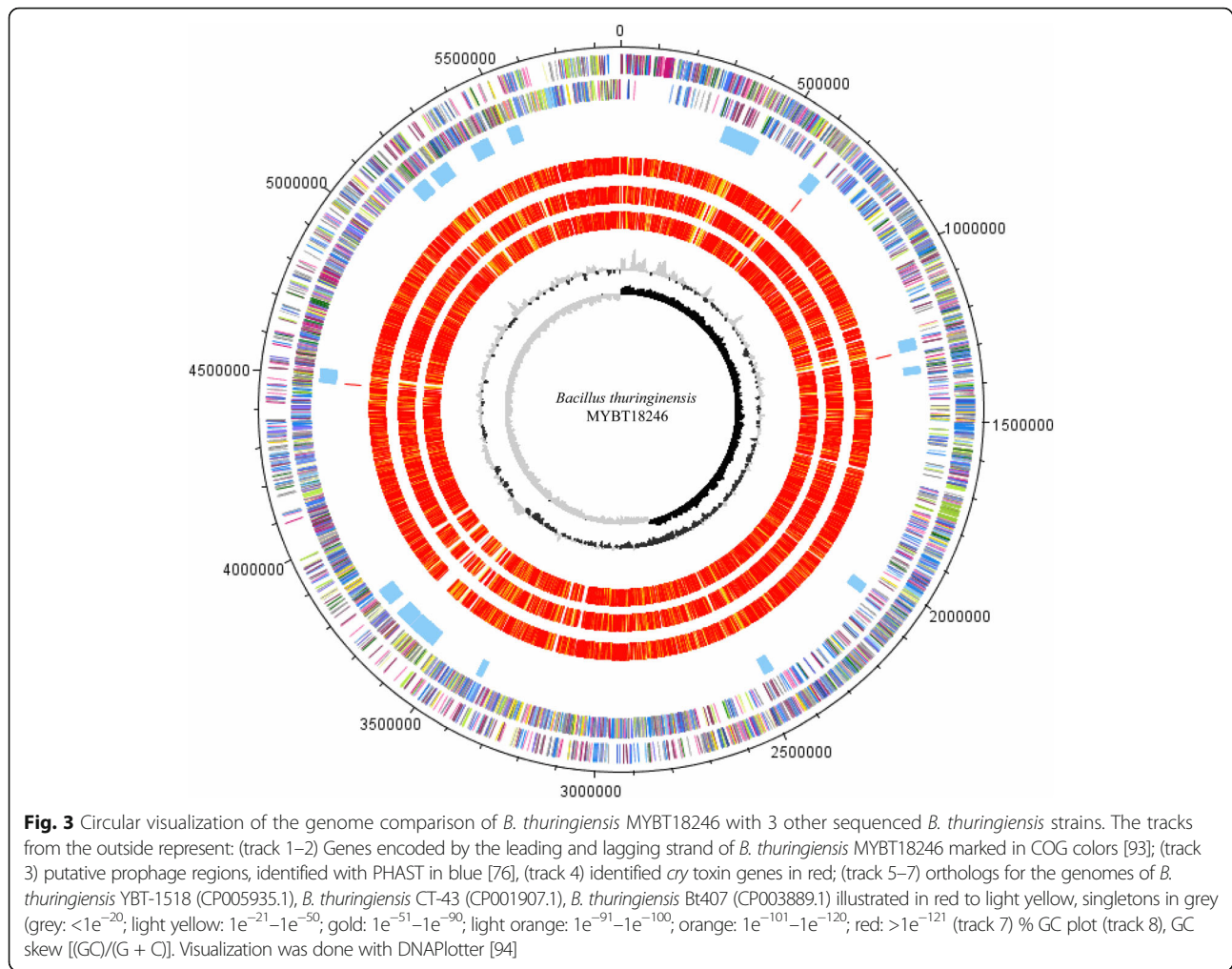
InterPro [75] revealed a p120510 encoded putative Vip-like toxin, a p120416 encoded putative Cry-like toxin and two p109822 encoded putative Bin-like toxins with potential for future studies.

Additionally, the *B. thuringiensis* MYBT18246 chromosome was screened for prophage regions by using the Phage Search Tool with default parameters. PHAST identifies prophage regions based on key genes from a reference database and defines the boundaries using a genomic composition-based algorithm. For a more detailed description see [76]. A total of 16 putative prophage loci were identified in the chromosome, including three that were associated with the previously identified chromosomally encoded *cry* toxin genes. As shown in Fig. 3, the *cry* toxins (displayed in red, track 4) are located in close proximity to identified prophage regions (displayed in blue, track 3). Furthermore, all *B. thuringiensis* MYBT18246 extra-chromosomal elements were also screened for prophages to check whether we could identify phages that reside in a linear or circular state in the host, as has been reported in 2013 by Fortier et al. [77]. Apparently, intact phage regions were identified according to the PHAST score system on p150790, p120416, p109822, p101287 and p46701.

The finding of prophage associated *cry* genes in strain MYBT18246 indicates that phages may serve as vectors for the transmission of virulence factors within the species *B. thuringiensis*. This resembles the previously described lysogenic conversion of pathogens by phages [78], supporting the idea that phages may represent a driving force for the distribution of fitness factors as well as virulence factors [78–80]. The finding that toxins, which are generally specific for a certain type of host organism, are located within a mobile genomic element in the chromosome of this bacterium, suggests that phages of strain MYBT18246 may contribute to adaptation to different hosts [81–83].

Extended insights
Based on the proximity within the tree (Fig. 2b), the genomes of *B. thuringiensis* Bt407, *B. thuringiensis* serovar *chinensis* CT-43 and *B. thuringiensis* YBT-1518 were identified as closest relatives and selected for an in depth comparative analysis. Shared gene contents were determined, visualized and compared, with a focus on known virulence factors such as *cry* toxins and pathogenic driving forces such as phages. The analysis revealed unique as well as shared gene contents for each strain (Fig. 3). In Fig. 3 the outer rings represent the genes on the leading and lagging strand with COG classification. The inner rings (track 5–7) illustrate the orthologous genes of *B. thuringiensis* YBT-1518, *B. thuringiensis* CT-43, *B. thuringiensis* Bt407 in red (high similarity) to light yellow (low similarity), and white (no similarity). The

Hollensteiner *et al. Standards in Genomic Sciences* (2017) 12:48

Page 7 of 10

Fig. 3 Circular visualization of the genome comparison of *B. thuringiensis* MYBT18246 with 3 other sequenced *B. thuringiensis* strains. The tracks from the outside represent: (track 1–2) Genes encoded by the leading and lagging strand of *B. thuringiensis* MYBT18246 marked in COG colors [93]; (track 3) putative prophage regions, identified with PHAST in blue [76], (track 4) identified *cry* toxin genes in red; (track 5–7) orthologs for the genomes of *B. thuringiensis* YBT-1518 (CP005935.1), *B. thuringiensis* CT-43 (CP001907.1), *B. thuringiensis* Bt407 (CP003889.1) illustrated in red to light yellow, singletons in grey (grey: $<1e^{-20}$; light yellow: $1e^{-21}$–$1e^{-50}$; gold: $1e^{-51}$–$1e^{-90}$; light orange: $1e^{-91}$–$1e^{-100}$; orange: $1e^{-101}$–$1e^{-120}$; red: $>1e^{-121}$ (track 7) % GC plot (track 8), GC skew [(GC)/(G + C)]. Visualization was done with DNAPlotter [94]

circular representation of the chromosome comparison revealed that prophages are a major source of regional differences between the strains (Fig. 3). Additionally, the pan-genome of *B. thuringiensis* MYBT18246 compared to the three closest relatives was determined (Fig. 4). Orthologous genes between all four organisms were identified by comparing the whole genomes using Proteinortho [84] with a similarity cutoff of 50% and an E-value of $1e^{-10}$. Gbk-files were downloaded from NCBI and the protein sequences were extracted using cds_extractor v0.7.1 [85]. Detected paralogous genes are displayed in the Venn diagram in Fig. 4. All four strains share a core genome of 4298 genes. This is equivalent to 67% of each genome. *Bacillus thuringiensis* MYBT18246 shares 4 additional genes exclusively with *B. thuringiensis* Bt407, 17 genes with *B. thuringiensis* serovar *chinensis* CT-43 and 327 genes with *B. thuringiensis* YBT-1518. *Bacillus thuringiensis* serovar *chinensis* CT-43 and *B. thuringiensis* Bt407 share 398 orthologous genes. Notably, the genome of *B. thuringiensis* MYBT18246 contains 1242 orphan genes and thus two to threefold

more singletons than the compared genomes. This result confirms the high degree of conservation of the four *Bacillus thuringiensis* strains (Fig. 2a and b) and it also refines the phylogenetic relationship of the strains to each other based on non-orthologous regions. Singletons are located on the chromosome as well as on extra-chromosomal elements. The density of singletons is higher (2.5 fold) on the plasmids. Notably, all major chromosomal differences can be attributed to prophage regions. All gene products were assigned to COG categories and investigated for PFAM domains and Signal peptides (Table 6). In detail, those genes code for: (i) phage proteins, (ii) morons (virulence factors), (iii) a vast majority of proteins with cryptic function. This is supported by Fig. 3 which clearly shows that the regions of differences (track 5–7) directly correspond to the regions of identified phages (track 3). Moreover, the identified *cry* toxins (track 4) are adjacent to identified prophage regions and could be suggested as morons. Additionally, the singletons were screened for further virulence factors and genes encoding type-IV secretion system, C5-
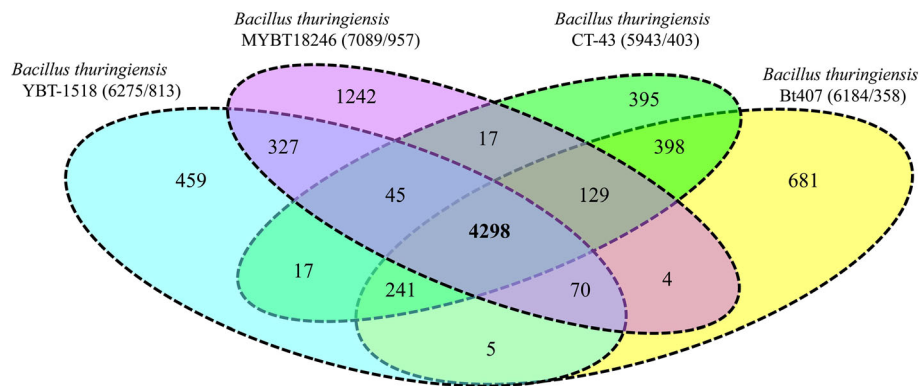
Fig. 4 Venn diagram of the genome comparison of *B. thuringiensis* MYBT18246 with other *B. thuringiensis* strains. Venn diagram displays the orthologous genes between *B. thuringiensis* MYBT18246 (CP015350-CP015361), *B. thuringiensis* YBT-1518 (CP005935-CP002486), *B. thuringiensis* serovar *chinensis* CT-43 (CP001907-CP001917) and *B. thuringiensis* Bt407 (CP003889-CP003898). Ortholog detection was performed with Proteinortho [84] including protein blast with a similarity cut-off of (50%) and an E-value of 1e^{-10}. The total number of genes and paralogs are depicted under the corresponding species name. Open reading frames that were classified as pseudogenes were not included in this analysis

methyltransferase, type-restriction enzymes, sporulation, resistance and genes involved in genetic competence were identified. In particular, the finding of restriction-modification systems indicates a protection mechanism against other phages and plasmids and thus forms a putative barrier against further genomic modification.

Table 6 General genome features of *B. thuringiensis* MYBT18246 and close relatives

Genome features	Genome name			
	B. thuringiensis MYBT18246[a]	*B. thuringiensis* 407[b]	*B. thuringiensis* YBT-1518[c]	*B. thuringiensis* CT-43[d]
Sequencing status	Finished	Finished	Finished	Finished
Genome size (Mbp)	6.75	6.13	6.67	6.15
DNA coding (bp)	5,623,665	5,133,026	5,421,574	5,079,667
GC (%)	35.4	35.02	35.29	35.12
DNA scaffolds	12	10	7	11
Total gene count	7239	6442	6738	6252
Protein coding genes (%)	97.9	95.9	98.0	95.1
RNA genes	151	180	139	124
Genes in internal clusters	2694	489	370	334
Genes with function prediction	5274	4615	5193	4211
Genes assigned to COGs	4662	3634	3746	3505
Genes with Pfam domains	5503	4991	5333	4809
Genes with signal peptides	500	447	471	418
Genes with transmembrane helices	1863	1750	1854	1698
CRISPR repeats	0	2	0	2

Accession numbers: [a]CP015350, [b]CP003889, [c]CP005935, [d]CP001907

Conclusion

In this work we present the whole-genome sequence of *B. thuringiensis* MYBT18246 and its specific genome features. The genome includes three nematicidal *cry*13 gene variants located on the chromosome, which were named according to sequence similarity as stated by the Cry Toxin Nomenclature Committee, as *cry*13Aa2, *cry*13Ba1, and *cry*13Ab1. Four additional putative toxin genes were identified with low sequence similarity to other known toxins on plasmids: p120510 (Vip-like toxin), p120416 (Cry-like toxin) and p109822 (two Bin-like toxins). These toxins contained complete toxin domains, yet the activity against potential hosts should be elucidated in future studies. The genome comprises a large number of mobile elements involved in genome plasticity including eleven plasmids and sixteen chromosomal prophages. Both plasmids and prophages are important HGT elements indicating that they are an important driving force for the evolution of pathogens. The most striking finding is the close proximity of the chromosomal nematicidal *cry* toxin genes to three distinct prophages indicating a contribution of phages in defining the host range of this strain. *B. thuringiensis* MYBT18246 may show potential as a biocontrol agent against nematodes which should be addressed in future experiments.

Hollensteiner *et al. Standards in Genomic Sciences* (2017) 12:48

Page 9 of 10

Authors' contributions
HL and HS designed the study. AP supervised the genome analysis. JH corrected the annotation, analyzed the genome with focus on comparative genomics. JH, AP and HL wrote the manuscript. BB, CS and PR performed the genome sequencing. HL performed the assembly and the annotation of the sequence data. JH performed the microscopy of strain MYBT18246. AS and CS isolated the genomic DNA. All authors read and approved the final manuscript.

Competing interest
The authors declare that they have no competing interests.

Author details
[1]Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, University of Göttingen, Göttingen, Germany. [2]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. [3]Present address: Nuffield Department of Medicine, University of Oxford, Oxford, UK. [4]Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian-Albrechts University of Kiel, Kiel, Germany. [5]Institute of Clinical Molecular Biology, Christian-Albrechts University of Kiel, Kiel, Germany.

References
1. Bechtel DB, Bulla LA, Kramer KJ, Bechtel DB, David- LI. Electron microscope study of Sporulation and Parasporal crystal formation in *Bacillus thuringiensis*. J Bacteriol. 1976;127:1472–81.
2. Ibrahim MA, Griko N, Junker M, Bulla LA. *Bacillus thuringiensis*: a genomics and proteomics perspective. Bioeng Bugs. 2010;1:31–50.
3. Carozzi NB, Kramer VC, Warren G, Evola S, Koziel MG. Prediction of insecticidal activity of *Bacillus thuringiensis* strains by polymerase chain reaction product profiles. Appl Environ Microbiol. 1991;57:3057–61.
4. Iriarte J, Porcar M, Lecadet M-M, Caballero P. Isolation and characterization of *Bacillus thuringiensis* strains from aquatic environments in Spain. Curr Microbiol. 2000;40:402–8.
5. Smith RA, Couche GA. The phylloplane as a source of *Bacillus thuringiensis* variants. Appl Environ Microbiol. 1991;57:311–5.
6. Burges D, Hurst JA. Ecology of *Bacillus thuringiensis* in storage moths. J Invertebr Pathol. 1977;30:131–9.
7. Schnepf E, Crickmore N, Van RJ, Lereclus D, Baum J, Feitelson J, et al. *Bacillus thuringiensis* and its Pesticidal crystal proteins. Microbiol Mol Biol Rev. 1998;62:775–806.
8. Sanahuja G, Banakar R, Twyman RM, Capell T, Christou P. *Bacillus thuringiensis*: a century of research, development and commercial applications. Plant Biotechnol J. 2011;9:283–300.
9. Vidal-Quist JC, Castañera P, González-Cabrera J. Diversity of *Bacillus thuringiensis* strains isolated from citrus orchards in Spain and evaluation of their insecticidal activity against Ceratitis Capitata. J Microbiol Biotechnol. 2009;19:749–59.
10. Feitelson J. The *Bacillus thuringiensis* family tree. In: Kim L, editor. Advanced engineered pesticides. 1st ed. New York: CRC Press; 1993. p. 63–71.
11. De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, et al. Bergey's manual of systematic bacteriology. In: Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H, Whitman W, editors. Volume 3: the Firmicutes. 2nd ed. New York: Springer; 2009. p. 1–1317.
12. Sabha ME-S, Adayel SA, El-Masry SA, Alazazy H. *Bacillus thuringiensis* (Bt) toxin for the control of citrus trees snails. Researcher. 2013;5:26–32.
13. Palma L, Muñoz D, Berry C, Murillo J, Caballero P. *Bacillus thuringiensis* toxins: an overview of their Biocidal activity. Toxins (Basel). 2014;6:3296–325. doi:10.3390/toxins6123296.
14. Crickmore N. *Bacillus thuringiensis* Toxin nomenclature. http://www. btnomenclature.info/. Accessed 12 Oct 2016.
15. Aronson AI. The two faces of *Bacillus thuringiensis*: insecticidal proteins and post-exponential survival. Mol Microbiol. 1993;7:489–96.
16. Sanchis V, Bourguet D. *Bacillus thuringiensis*: applications in agriculture and insect resistance management. A review Agron Sustain Dev. 2008;28:11–20.
17. Cohn F. Untersuchungen über Bakterien. Beiträge zur Biologie der Pflanzen. 1872;1:127–224.
18. Skerman V, McGowan V, Sneath P. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–420.
19. Scott E, Dyer DW. Divergence of the SigB regulon and pathogenesis of the *Bacillus cereus* Sensu Lato group. BMC Genomics. 2012;13:564. doi:10.1186/1471-2164-13-564.
20. Frankland GC, Frankland PF. Studies on some New Micro-Organisms obtained from Air. Philosophical Transactions of the Royal Society of London. 1887. p. 257–87.
21. Ernst B. Ueber die Schlaffsucht der Mehlmottenraupe (Ephestia kuhniella) und ihren Erreger *Bacillus thuringiensis* n. sp. Zeitschrift für Angew Entomol. 1915;2:21–56.
22. Flügge C. Die Mikroorganismen. 2.Auflage. Leipzig, Germany: F.C.W. Vogel; 1886.
23. Nakamura LK. *Bacillus pseudomycoides* sp. nov. Int J Syst Bacteriol. 1998;48:1031–5. doi:10.1099/00207713-48-3-1031.
24. Lechner S, Mayr R, Francis KP, Prüss BM, Kaplan T, Wießner-Gunkel E, et al. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. Int J Syst Bacteriol. 1998;48:1373–82. doi:10.1099/00207713-48-4-1373.
25. Guinebretière MH, Auger S, Galleron N, Contzen M, de Sarrau B, de Buyser ML, et al. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* group occasionally associated with food poisoning. Int J Syst Evol Microbiol. 2013;63:31–40.
26. Jiménez G, Urdiain M, Cifuentes A, López-López A, Blanch AR, Tamames J, et al. Description of *Bacillus toyonensis* sp. nov., a novel species of the *Bacillus cereus* group, and pairwise genome comparisons of the species of the group by means of ANI calculations. Syst Appl Microbiol. 2013;36:383–91. doi:10.1016/j.syapm.2013.04.008.
27. Oren A, Garrity GM. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2014;64:1–5. doi:10.1099/ijsem.0.000737.
28. Jung MY, Paek WK, Park IS, Han JR, Sin Y, Paek J, et al. *Bacillus gaemokensis* sp. nov., isolated from foreshore tidal flat sediment from the Yellow Sea. J Microbiol. 2010;48:867–71.
29. Jung MY, Kim JS, Paek WK, Lim J, Lee H, Kim PI, et al. *Bacillus manliponensis* sp. nov., a new member of the *Bacillus cereus* group isolated from foreshore tidal flat sediment. J Microbiol. 2011;49:1027–32.
30. Liu B, Liu GH, Hu GP, Cetin S, Lin NQ, Tang JY, et al. *Bacillus bingmayongensis* sp. nov., isolated from the pit soil of emperor Qin's Terra-cotta warriors in China. Antonie Van Leeuwenhoek. 2014;105:501–10.
31. NL; Validation of the publication of new names and new combinations previously effectively published outside the IJSB. List No. 65. Int J Syst Bacteriol. 1998;48(Pt 2):627.
32. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, et al. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*–one species on the basis of genetic evidence. Appl Environ Microbiol. 2000;66:2627–30.
33. Okinaka TR, Keim P, Okinaka RT, Keim P. The Phylogeny of *Bacillus cereus* sensu lato. Microbiol Spectr. 2016;4 doi:10.1128/microbiolspec.
34. Rasko DA, Altherr MR, Han CS, Ravel J. Genomics of the *Bacillus cereus* group of organisms. FEMS Microbiol Rev. 2005;29:303–29.
35. Schmidt TR, Scott EJ II, Dyer DW. Whole-genome phylogenies of the family Bacillaceae and expansion of the sigma factor gene family in the *Bacillus cereus* species-group. BMC Genomics. 2011;12:430. doi:10.1186/1471-2164-12-430.
36. Priest FG, Barker M, Baillie LWJ, Holmes EC, Maiden MCJ. Population structure and evolution of the *Bacillus cereus* group. Society. 2004;186:7959–70.
37. Yu H, Zhang J, Huang D, Gao J, Song F. Characterization of *Bacillus thuringiensis* strain Bt185 toxic to the Asian cockchafer: *Holotrichia parallela*. Curr Microbiol. 2006;53:13–7.
38. Ackermann H-WW, Azizbekyan RR, Bernier RL, Barjac H, Saindoux S, Valero JR, et al. Phage typing of *Bacillus subtilis* and *B. thuringiensis*. Res Microbiol. 1995;146:643–57.
39. Mahillon J, Rezsöhazy R, Hallet B, Delcour J. IS231 and other *Bacillus thuringiensis* transposable elements: a review. Genetica. 1994;93:13–26.
40. Kronstad JW, Whiteley HR. Inverted repeat sequences flank a *Bacillus thuringiensis* crystal protein gene. J Bacteriol. 1984;160:95–102.
41. Masri L, Branca A, Sheppard AE, Papkou A, Laehnemann D, Guenther PS, et al. Host–pathogen Coevolution: the selective advantage of *Bacillus*

Hollensteiner *et al. Standards in Genomic Sciences* (2017) 12:48

Page 10 of 10

thuringiensis virulence and its cry toxin genes. PLoS Biol. 2015;13:e1002169. doi:10.1371/journal.pbio.1002169.

42. Schulte RD, Makus C, Hasert B, Michiels NK, Schulenburg H. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. Proc Natl Acad Sci U S A. 2010;107:7359–64. doi:10.1073/pnas.1003113107.

43. Payne J. Isolates of *Bacillus thuringiensis* that are active against nematodes. 1992.

44. Payne J, Cannon R, Bagley A. *Bacillus thuringiensis* isolates for controlling acarides. 1993.

45. Schnepf H, Schwab G, Payne J, Narva K, Foncerrada L. Nematicidal proteins. 2001.

46. Barjac H, Frachon E. Classification of *Bacillus thuringiensis* strains. Entomophaga. 1990;35:233–40. doi:10.1007/BF02374798.

47. Gibbons NE, Murray RGE. Proposals concerning the higher Taxa of bacteria. Int J Syst Bacteriol. 1978;28:1–6. doi:10.1099/00207713-28-1-1.

48. West AW, Burges HD, Dixon TJ, Wyborn CH. Survival of *Bacillus thuringiensis* and *Bacillus cereus* spore inocula in soil: effects of pH, moisture, nutrient availability and indigenous microorganisms. Soil Biol Biochem. 1985;17:657–65.

49. Ash C, Farrow JA, Dorsch M, Stackebrandt E, Collins MD. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int J Syst Bacteriol. 1991; 41(3):343–6.

50. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing Phylogenetic trees. Mol Biol Evol. 1987;4:406–25.

51. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 2004;101:11030–5.

52. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;7:1870–4. doi:10. 1093/molbev/msw054.

53. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 2012;40:115–22.

54. Field D, Garrity G, Gray T, Morrison N, Selengut J, Tatusova T, et al. The minimum information about a genome sequences (MIGS) specification. Nat Biotechnol. 2008;26:541–7. doi:10.1038/1360.

55. Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F. Accuracy and quality assessment of 454 GS-FLX titanium pyrosequencing. BMC Genomics. 2011;12:245. doi:10.1186/1471-2164-12-245.

56. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics, Proteomics Bioinforma. 2015;13:278–89. doi:10.1016/j.gpb.2015.08.002.

57. Staden R, Beal KF, Bonfield JK. The Staden package, 1998. In: Misener S, Krawtz AS, editors. Methods in molecular biology. New York: Humana Press; 1999. p. 115–30.

58. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10:563–9. doi:10.1038/nmeth.2474.

59. PacBio Software Downloads. 2016. http://www.pacb.com/support/software-downloads/. Accessed 12 Apr 2016.

60. Gao F, Luo H, Zhang C-T. DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. Nucleic Acids Res. 2013;41 Database issue:D90–3. doi:10.1093/nar/gks990.

61. DoriC: an updated database of bacterial and archaeal replication origins. http://tubic.tju.edu.cn/doric/. Accessed 15 Dec 2016.

62. Gao F, Zhang C-T. Ori-finder: a web-based system for finding oriCs in unannotated bacterial genomes. BMC Bioinformatics. 2008;9:79.

63. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.

64. Sheppard AE, Poehlein A, Rosenstiel P, Liesegang H, Schulenburg H. Complete genome sequence of *Bacillus thuringiensis* strain 407 cry. Genome Announc. 2013;1:158–12.

65. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119. doi:10.1186/1471-2105-11-119.

66. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8.

67. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32:11–6.

68. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6. doi:10.1038/nmeth.1701.

69. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. Nucleic Acids Res. 2003;31:439–41.

70. Crickmore N, Zeigler DR, Feitelson J, Schnepf E, Van Rie J, Lereclus D, et al. Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiol Mol Biol Rev. 1998;62:807–13.

71. Bateman A, Martin MJ, O'Donovan C, Magrane M, Apweiler R, Alpi E, et al. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43:D204–12.

72. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res. 2013;41:36–42.

73. Carlson CR, Kolstø A-BB. A complete physical map of a *Bacillus thuringiensis* chromosome. J Bacteriol. 1993;175:1053–60.

74. Wang A, Pattemore J, Ash G, Williams A, Hane J. Draft genome sequence of Bacillus thuringiensis strain DAR 81934, which Exhibits Molluscicidal Activity. Genome Announc. 2013;1:e0017512. doi:10.1128/genomeA.00175-12.

75. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 2015;43:D213–21.

76. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39(SUPPL. 2):347–52.

77. Fortier L-C, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence. 2013;4:354–65. doi:10.4161/viru.24498.

78. Brüssow H, Canchaya C, Hardt W-D, Bru H. Phages and the evolution of bacterial pathogens: from genomic rearrangements to Lysogenic conversion. Microbiol Mol Biol Rev. 2004;68:560–602.

79. Barksdale L, Arden SB. Persisting bacteriophage infections, lysogeny, and phage conversions. Annu Rev Microbiol. 1974;28:265–99.

80. Freeman VJ. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. J Bacteriol. 1951;61:675–88.

81. De Maagd RA, Bravo A, Crickmore N. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends Genet. 2001;17:193–9.

82. Agaisse H, Lereclus D. How does *Bacillus thuringiensis* produce so much insecticidal crystal protein? J Bacteriol. 1995;177:6027–32.

83. Ben-Dov E. *Bacillus thuringiensis* subsp. *israelensis* and its Dipteran-specific toxins. Toxins (Basel). 2014;6:1222–43.

84. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:124. doi:10.1186/1471-2105-12-124.

85. Leimbach A. bac-genomics-scripts: Bovine *E. coli* mastitis comparative genomics editio. 2016. https://github.com/aleimba/bac-genomics-scripts/tree/master/cds_extractor ,Accessed 14 Mar 2016.

86. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.

87. Euzéby J. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2010;60:469–72.

88. Ludwig W, Schleifer K-H, Whitman WB. Class I. *Bacillus* class nov. Bergey's Man Syst Bacteriol. 2009;3:19–20.

89. Prévot AR. Dictionnaire des Bactéries Pathogènes In: Hauderoy P, Ehringer G, Guillot G, Magrou. J., Prévot AR, Rosset D, Urbain A (eds), Dictionnaire des Bactéries Pathogènes, Second Edition, Masson et Cie, Paris, 1953, p. 1–692.

90. Fischer A. Untersuchungen über bakterien. Jahrbücher für Wissenschaftliche Botanik. 1895;27:1–163.

91. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–80.

92. Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie Van Leeuwenhoek. 64:253–60. http://www.ncbi.nlm.nih.gov/pubmed/8085788. Accessed 6 Jun 2017.

93. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4:41. doi:10.1186/1471-2105-4-41.

94. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics. 2009;25:119–20.