

RESEARCH

Open Access



# Whole genome sequencing of a novel *Bacillus thuringiensis* isolated from Assam soil

Mihir Rabha<sup>1,2</sup>, Debajit Das<sup>1,3</sup>, Trishna Konwar<sup>1,3</sup>, Sumita Acharjee<sup>1,3\*</sup> and Bidyut Kumar Sarmah<sup>1,3\*</sup>

## Abstract

**Background** *Bacillus thuringiensis* (Bt) is a gram-positive ubiquitous saprophytic bacterium that produces proteins (Crystal protein, Vegetative insecticidal protein, and Secreted insecticidal protein) toxic to insects during its growth cycle. In the present study, the whole genome of a locally isolated *B. thuringiensis* strain BA04 was sequenced to explore the genetic makeup and to identify the genes responsible to produce insecticidal proteins including the virulence factors. The strain was isolated from the soil sample of the Kaziranga National Park, Assam, North-Eastern part of India (Latitude: 26°34'39.11"N and Longitude: 93°10'16.04"E).

**Results** The whole genome sequencing (WGS) of the BA04 strain revealed that it has a circular genome of size 6,113,005 bp with four numbers of plasmids. A total of 6,111 genes including two novel crystal protein-encoding genes (MH753362.1 and MH753363.1) were identified. The BLASTn analysis of MH753362.1 showed 84% similarities (maximum identity) with *Cry1Ia* (KJ710646.1) gene, whereas MH753363.1 exhibited 66% identity with *Insecticidal Crystal Protein (ICP)-6* gene (KM053257.1). At the protein level, MH753362.1 and MH753363.1 shared 79% identity with *Cry1Ia* (AIW52613.1) and 40% identity with *Insecticidal Crystal Protein (ICP)-6* (AJW76687.1) respectively. Three-dimensional structures of these two novel protein sequences revealed that MH753362.1 have 48% structural similarity with *Cry8ea1* protein, whereas MH753363.1 showed only 20% structural similarity with *Cry4Aa* protein.

Apart from these insecticidal genes, the strain was also found to contain virulence and virulence-associated factors including the antibiotic resistance genes and Clustered regularly interspaced short palindromic repeat (CRISPR) sequences.

**Conclusion** This is the first report on the whole genome sequence of Bt strain BA04 isolated from Assam, a North-Eastern state of India. The WGS of strain BA04 unveils the presence of two novel types of insecticidal crystal protein-encoding genes which can be used for the development of insect-resistant transgenic crops. Additionally, the strain could be used for the formulations of effective biopesticides. The WGS provides the fastest and cheapest platform for a better understanding of the genetic makeup of a strain and helps to explore the role of virulence genes in pathogenicity against the insect host.

**Keywords** *Bacillus thuringiensis*, Insecticidal proteins, *Cry* genes, Whole Genome Sequencing, *B. thuringiensis* strain BA04

\*Correspondence:

Sumita Acharjee  
sumita.acharjee@aaau.ac.in  
Bidyut Kumar Sarmah  
bidyutsarmah@aaau.ac.in

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

*Bacillus thuringiensis* is a gram-positive, saprophytic, spore-forming, entomopathogenic bacteria that can potentially be used as a biocontrol agent. *B. thuringiensis* produces crystal proteins (delta endotoxins) during sporulation which confers toxicity against insect pests [1–3]. These crystal proteins are generally classified into different groups and sub-groups based on their amino acid sequence similarities and the specificity of toxicity against insect pests of different orders such as lepidopteran, coleopteran, dipteran, homopteran, hymenopteran, Mallophaga, and other organisms such as nematodes, protozoa and mites [4–6]. Due to the specificity of these crystal proteins, *B. thuringiensis* is one of the most widely studied biocontrol agent at the molecular level. These crystal proteins are encoded by *Cry* genes and to date, more than 800 *Cry* gene sequences have been identified from different *B. thuringiensis* strains and registered on the NCBI site [7]. Despite all these, the mining of novel *Cry* genes is remaining an interesting area of research to explore novel candidate genes having higher efficacy and toxicity against a wide range of insect pests.

Several molecular approaches have been adopted in the recent past to search for a novel type of *Cry* genes such as gene hybridization [8, 9], PCR-mediated techniques by using general or multi-primer [10], DNA library [11], and PCR followed by restriction fragment analysis [12]. However, all these techniques are labor-intensive, time-consuming, and inefficient compared to the whole genome sequencing approach.

Apart from these insecticidal crystal proteins, strains of *B. thuringiensis* are known to produce other insecticidal proteins such as Vegetative Insecticidal Protein (VIP), Secreted Insecticidal Protein (Sip), and Cytotoxic proteins (Cyt) at different stages of its growth cycle. Thus, Bt strains are highly copious of insecticidal genes which might help them to survive and proliferate in different ecological and geographical conditions. The whole genome sequence analysis of *B. thuringiensis* from different geographical locations and ecological habitats would help us to understand the integrity of their genome and the precise order of evolutions. The increasingly available online resources, databases, and archives of the WGS data along with the parallel progress in the field of bioinformatics have remarkably reduced the cost of genome sequencing. The large-scale availability of WGS data has facilitated the identification, characterization, and mapping of genomes of new organisms and detailed metabolic pathways of such organisms, fishing out new candidate genes having broad spectrum activity against insects and comparing genomes across multiple samples. The whole genome sequencing of *B. thuringiensis* provides an opportunity to investigate in detail the genetic

makeup for pathogenicity and toxicity against the host insects. In the present investigation, we identified and isolated novel types of insecticidal crystal protein genes from a *B. thuringiensis* strain of Assam soil. Since the diversity in the ecological habitat and geographical locations plays a vital role in finding variations in the *Cry* genes of *B. thuringiensis* [13, 14]. Assam is a bio-diversity hotspot endowed with a variety of natural flora and fauna, making it a unique natural habitat for novel *B. thuringiensis* strains with novel insecticidal genes. For the present investigation, the Bt strain was isolated from the soil samples of Assam, and we performed morphological, biochemical, and molecular characterization to identify novel genes [15, 16].

From the taxonomical classification point of view, *B. thuringiensis* belongs to the *B. cereus* group that consists of six species (*B. thuringiensis*, *B. anthracis*, *B. cereus*, *B. mycoides*, *B. pseudomycoides* and *B. weihenstephanensis*) [17]. Based on the previous reports, species of the *B. cereus* group share a very high level of protein homology and show similar orientations of conserved sequences in the genome [18]. Out of which, three species *B. thuringiensis*, *B. cereus*, and *B. anthracis* were found to have similar genetic makeup [19, 20]. However, among these three species, *B. cereus* and *B. thuringiensis* are considered closely related species with almost similar kind of genetic makeup. *B. cereus* is considered acrySTALLIFEROUS, whereas *B. thuringiensis* is known as crystal-liferous bacteria that produces crystal proteins. This is the only property that differentiates these two species. In case of loss of these crystal protein-encoding genes located on the plasmid (transfer through conjugation), it becomes nearly impossible to differentiate them. The previous report also concluded that these two species cannot be separated based on phylogenetic analysis [17]. Therefore, a whole genome study can help to understand their genetic makeup to differentiate these two closely related species more effectively.

## Results

### Insecticidal activity of the strain BA04

Larvae of *Helicoverpa armigera* of various instars raised on artificial diet were fed with spores of the *B. thuringiensis* strain BA04. After 7 days, considerable (98–100%) mortality of larvae was observed irrespective of the larval instars (Supplementary Fig. 2).

### Genome sequencing and assembly of *B. thuringiensis* strain BA04

A total of 9,293,122 reads were generated after filtering with a quality score of 94.89% and as many as 94 contigs were obtained. The total size of the genome was found to be 6,113,005 bp with an average GC% of

34.78164 (Table 1). The sequencing depth was 153X [Seq. depth = 938,605,322 no. of bases obtained / 6113005 bp size of the genome = 153.54] and the coverage was 99.99% [ Coverage (%) = (65,031 read length × 94 No.

of reads) / 6,113,005 genome length × 100]. The genome of the BA04 strain of Bt was comprised of 4 numbers of plasmids.

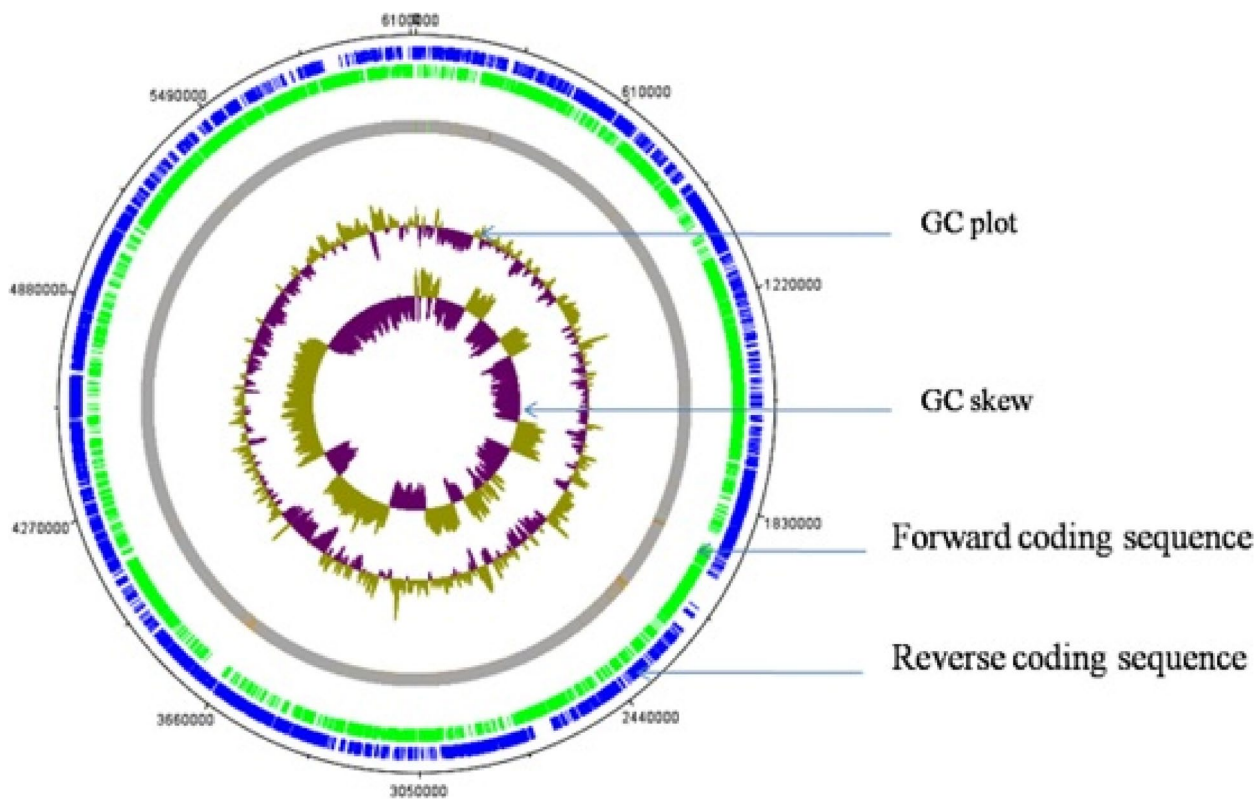
**Table 1** Genomic features of *B. thuringiensis* isolate BA04

Genome features	Value
Number of contigs	94
Genome size (bp)	6,113,005
GC percent	34.78164
Number of plasmid	4
Total genes	6111
Protein coding genes	6038
tRNA	65
rRNA	7
tmRNA	1

- Bases (bp) The number of bases in each contig.
- Gene The basic physical unit of heredity,
- CDS Coding Sequence
- tRNA Transfer RNA,
- rRNA Ribosomal RNA, a molecular component of ribosome.
- tmRNA Transfer-messenger RNA (dual tRNA-like and messenger RNA-like properties).

**Genome annotation**

Genome annotation revealed the presence of 6111 genes corresponding to a total of 6038 protein-encoding genes (98.80% of all the genes). The non-coding RNAs include 67 tRNA (~ 1.1% of all the genes), 7 rRNA genes (0.11% of all the genes), and one tmRNA (Transfer-messenger RNA), which is a bacterial RNA molecule with dual tRNA-like and messenger RNA-like properties (Table 1). The tmRNA was known to play a major role in the ribosome rescue process and quality protein synthesis [21]. The circular genome obtained through the DNA plotters showed the quality of the genome sequence (both forward and reverse sequences). Figure 1 has both forward sequences (blue circle) and reverse sequences (green circle) where the GC content and GC skew are also flaunted. The circular form of the whole genome of strain BA04 has been obtained using the DNA plotter tool available at <http://www.sanger.ac.uk/science/tools/dnaplotter>. DNA plotter produces the genome in a linear and circular form where the quality of the sequences can be observed in a graphical format [22].



**Fig. 1** Circular genome map of *Bacillus thuringiensis* isolate BA04 generated by DNA plotter tool

The Rapid Annotations using Subsystems Technology (RAST) annotation have distributed 6283 genes into 180 subsystems. The subsystems of the BA04 strain genome based on the RAST annotation server are depicted in Fig. 2. The most abundant genes annotated were associated with amino acids synthesis and their derivatives (156 genes; 2.48%) followed by energy and precursor metabolites generation (130 genes; 2.07%) and then cofactors, vitamins, and prosthetic groups (126 genes; 2.00%). The sketch of the KEGG (Kyoto Encyclopaedia of Genes and Genomes) metabolic pathway obtained through the RAST analysis was provided in the supplementary data (Additional file 1: Figure S1) and with the list of all the genes identified through the RAST server (Additional file 2: Table S1).

### Comparison with closely related *B. thuringiensis* strains

A comparative protein-encoding genome sequence analysis of BA04 strain with two *B. cereus* genomes (*B. cereus* ATCC 14,579 and *B. cereus* biovar *anthracis* str. CI) and a different sub-species of *B. thuringiensis* (*B. thuringiensis* serovar *konkukian* str. 97–27) revealed that BA04 genome has a significant variation with the *B. cereus* strains (Fig. 3). Interestingly, noticeable variations were also found between BA04 and *B. thuringiensis* serovar *konkukian* strain 97–27. The whole genome BLAST analysis of BA04 has shown close relation with another *B. thuringiensis* strains HS18-1 with 99% identity (82% query cover and maximum alignment score 4021 with

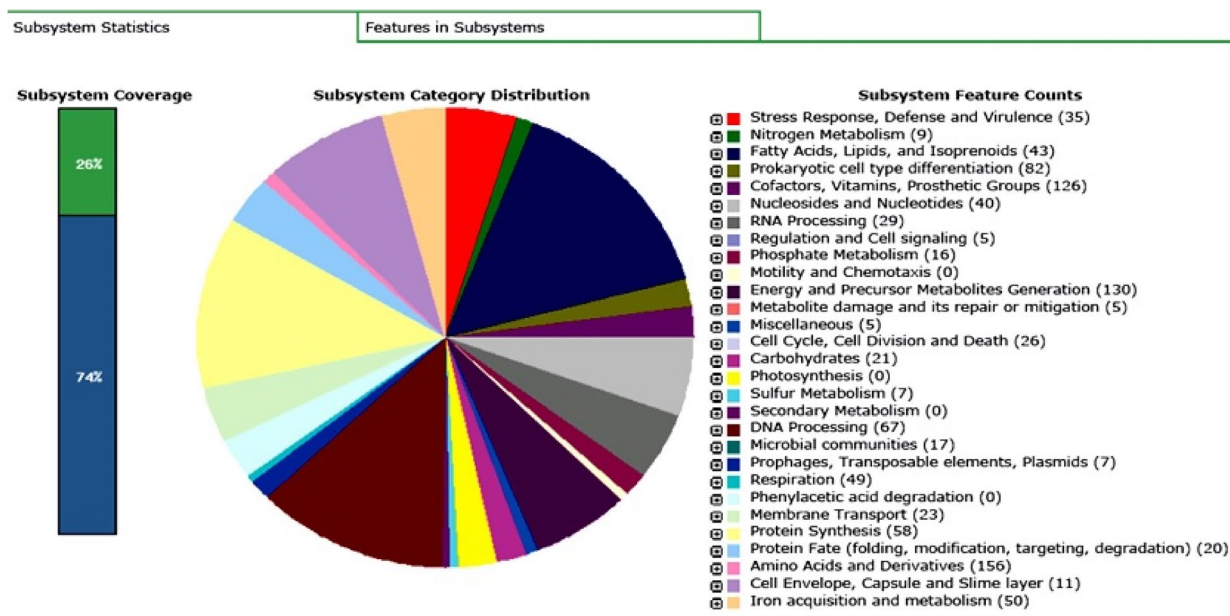
0.0 error value) rather than with *B. thuringiensis* serovar *konkukian* str. 97–27 (64.8% similarity). Also, BA04 showed 73.6 and 63.8% similarity with *B. cereus* ATCC 14,579 and *B. cereus* biovar *anthracis* strain, respectively.

### Phylogenetic assessments

The Genome-to-genome distance calculator (GGDC) analyses indicated that the strain BA04 is closely clustered with *B. thuringiensis* strain HS18-1 (with 92.3% similarity) followed by *B. thuringiensis* strain HD12 (71.5% similarity). The distance of the tree was about 0.005. However, the *B. cereus* strain showed a distant relation with strain BA04 and clustered in separate lineages (Fig. 4). The numbers above the branches are greedy-with-trimming pseudo-bootstrap support values above 80% are shown. *thuringiensis*.

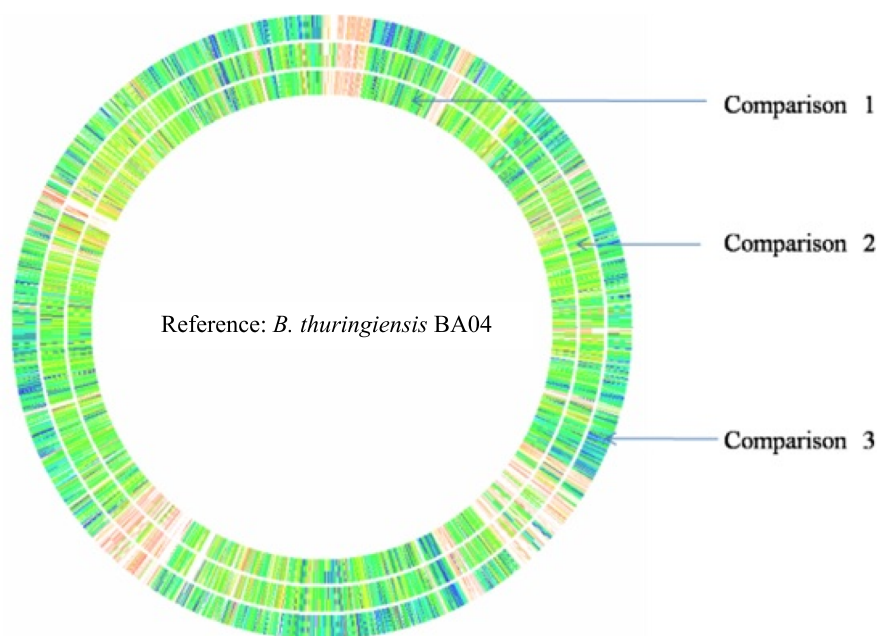
### Genomic islands (GIs) of *B. thuringiensis* strain BA04

A total of 17 genomic islands mostly containing genes of unknown function i.e. hypothetical proteins (222) with 191 clustered genes have been identified in the genome when compared with the reference genome of *B. thuringiensis* strain HD789 using the Island Viewer 4 (<http://www.pathogenomics.sfu.ca/islandviewer/browse/>) (Fig. 5). Moreover, genes encoding putative transposase (4), PD-(D/E)XK nuclease family transposase (2), putative prophage phiRv2 integrase (1) were also represented in the GIs. These genomic islands are the regions of the genome that provide evidence about the horizontal gene



**Fig. 2** The RAST analysis-based subsystem distribution of whole genome sequence of isolates BA04. Each colour in the pi graph represents a particular group of genes mentioned in the right site of the graph

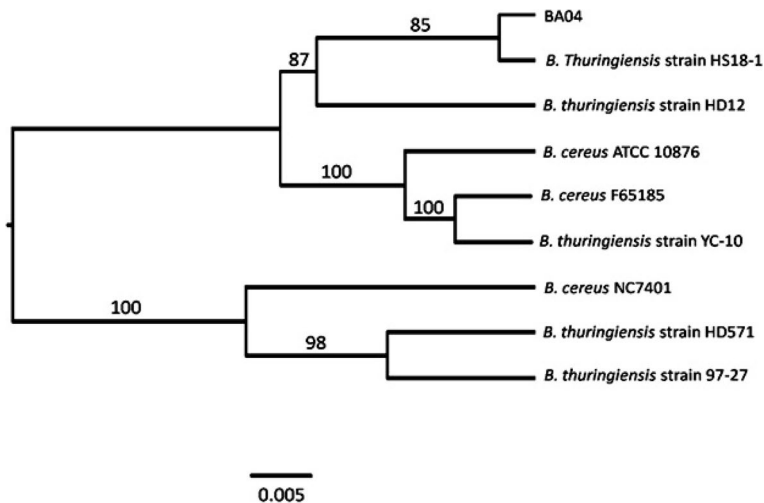




Percent protein sequence identity

Bidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10
Unidirectional best hit	100	99.9	99.8	99.5	99	98	95	90	80	70	60	50	40	30	20	10

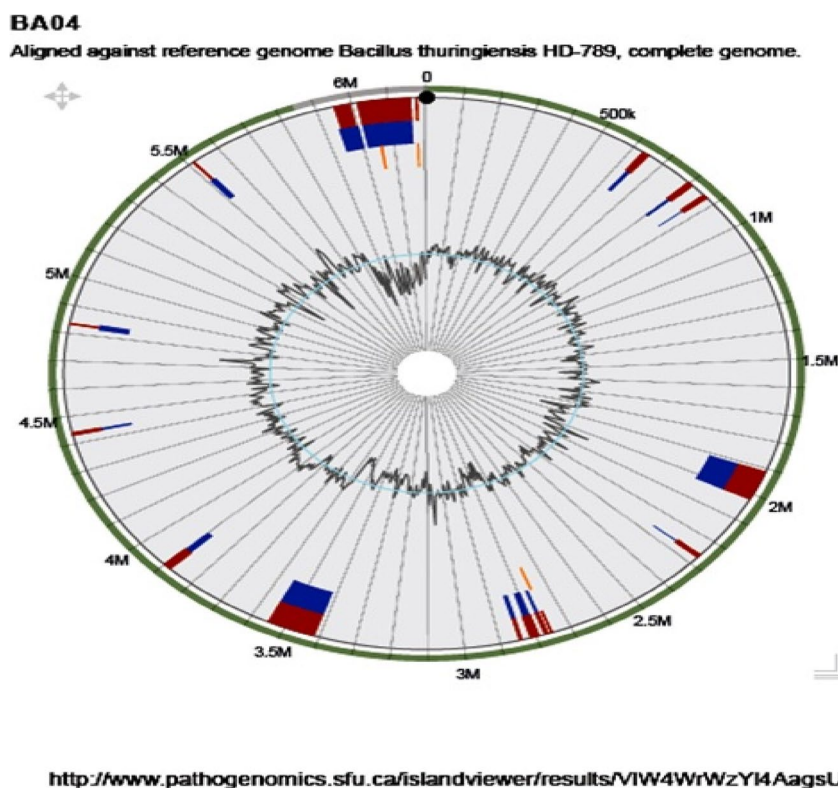
**Fig. 3** Comparative genome analysis of isolate BA04 under RAST server. The first circle Comparison 1 is with *B. cereus* strain ATCC 14,579. The second circle Comparison 2 is with *B. cereus* biovar *anthracis* str. CI (637,380.6) and the third circle Comparison 3 is with *B. thuringiensis* serovar konkukian str. 97-27 (281,309.3)



**Fig. 4** Phylogenetic tree obtained using the Genome-to-Genome Distance Calculator 2.1 (GGDC)

transfer which play important role in the evolution, and diversification of pathogenic microbes and the adaptation of bacteria to different environments [23]. The list of

the genes that were clustered in the genomic islands and their locus in the genome are included as supplementary data (Additional file 3: Table S2).



**Fig. 5** Genomic Islands prediction and genome visualization of isolates BA04 against Bt strain HD789 revealed 17 genomic islands

### Identification of two novel *Cry* genes

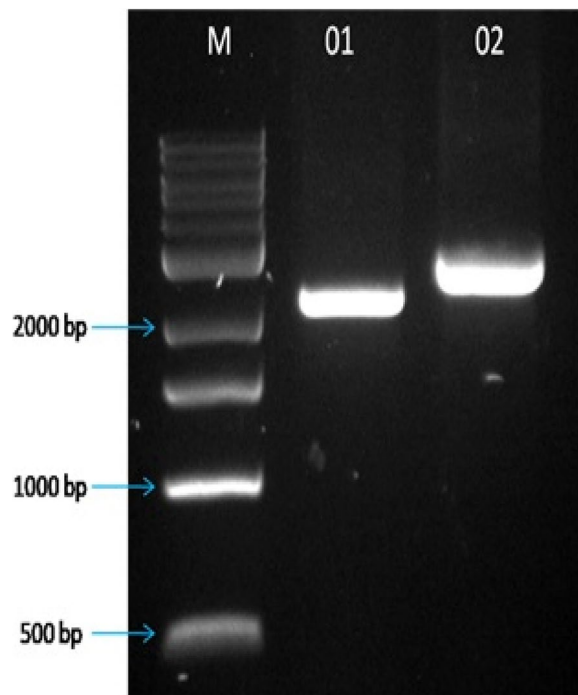
The genome annotations of the strain BA04 have led to the identification of two novel types of *Cry* genes which were deposited at the NCBI (MH753362.1 and MH753363.1), the size of MH753362.1 and MH753363.1 were found to be 2.1 and 2.7 kb, respectively. Full-length primers were designed for both genes and were successfully PCR amplified (Fig. 6). Subsequently, the purified PCR products were cloned into a pGEM<sup>®</sup>-T easy cloning vector for further analysis. The BLASTn analysis revealed that MH753362.1 has a maximum sequence similarity of about 84% with *Cry11a* (KJ710646.1), while MH753363.1 showed 66% identity with the *ICP-6* gene (KM053257.1). The pBlast of MH753362.1 showed a maximum of 79% identity with *Cry11a* (AIW52613.1) like protein, and MH753363.1 exhibited 40% identity with *ICP-6* (ID AJW76687.1). These results indicated significant variations in the sequences of both MH753362.1 and MH753363.1 and could be considered novel types of crystal protein-encoding genes, however, the toxicity assay through expression analysis and insect bioassay could add to their efficacy towards target insects.

Analysis of the three-dimensional structures of these two novel proteins using the online server Phyre2

revealed 48% structural identities with *Cry8Ea1* for MH753362.1, whereas MH753363.1 was only 20% similar to *Cry4Aa* proteins. The 3D analysis of both sequences denoted the presence of three functional domains of a typical crystal protein along with a few additional domains (Fig. 7).

### Genome properties

The genome of *B. thuringiensis* strain BA04 also has many virulence factors (Table 2) apart from the *Cry* toxins. During the genome annotation, two different types of chitinases (*chiA* and *chiD*) along with different types of protease encoding genes including four bacillolysin, five collagenase, four different types of immune inhibitors, and three phospholipase encoding genes were identified (Table 2). Three different lactonase encoding genes, alveolysin, zeta toxin (*Streptococcus pyogenes*), and toxin A were also found in BA04. Also, different types of hemolysin-encoding genes were identified during the genome annotation (Table 2). We also found numerous virulence-associated factors (Table 3) including antitoxin gene, extracellular metalloprotease, spore photoproduct lyase, capsule biosynthesis genes, sporulation protein genes, spore germination protein genes, and genes encoding three different potential bacteriocins such *asalbA*, *albE*,

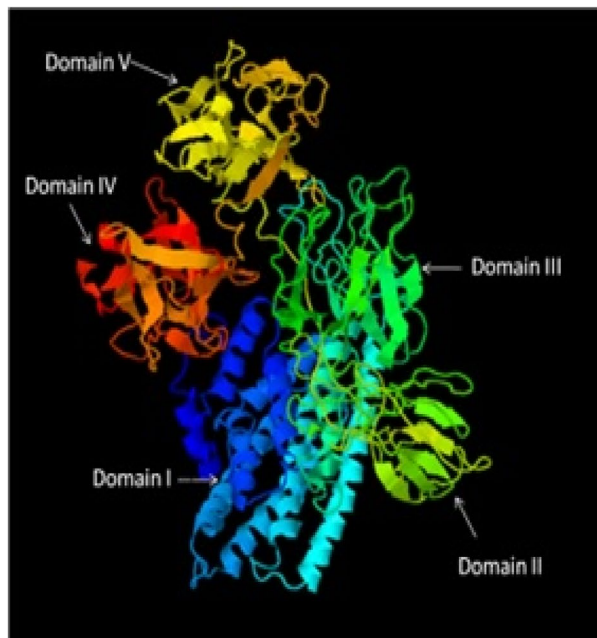


**Fig. 6** Amplification of full length identified *Cry* gene sequences from isolate BA04. Lane M- 1 kb DNA ladder; Lane 1: amplified product of MH753362.1; Lane 2: Amplified product of MH753363.1. The gel image was cropped to show the desired amplicons. Original gel image was included as additional figure (Additional file 1\_Figure S1)

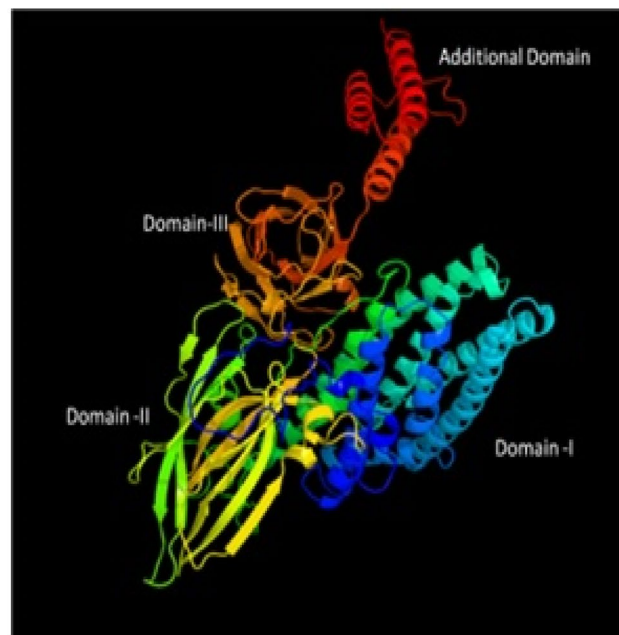
and *albF* (Table 3). The strain was also found to carry genes associated with nitrogen metabolism (Table 3). In all, 8 nitrogen metabolism-related genes were identified corresponding to *ureA*, *ureB*, *ureC*, *ureD*, *ureE*, *ureF*, *ureG*, *ureH*, and *ureI* which play a crucial role in the synthesis of urease. Furthermore, the genome of the BA04 strain was found to encode several antibiotic resistance genes such as vancomycin, tetracycline, polymyxin, bicyclomycin, fosfomicin, polymyxin, and fosmidomycin resistance genes (Table 4). We identified 11 genes associated with multidrug resistance genes of which 5 genes are related to ABC transporter. The strain was also found to contain five different types of CRISPR sequences (Table 5).

### Discussion

In the present study, the whole genome sequencing of a new *B. thuringiensis* strain BA04 isolated from Assam soil was carried out. The WGS revealed that the genome of the BA04 strain is substantially large. It is widely documented that the strains with large genome sizes have higher toxicity than that of smaller genomes against the target pest [18] since the larger genomes contain a high copy number of plasmids having insecticidal activity. Moreover, the strains with higher toxicity generally produce different types of virulence factors which enhance the toxicity of the insecticidal genes [24]. The virulence genes are mostly reported



**A**



**B**

**Fig. 7** Three-D analysis for *Cry* protein like sequences under online server Phyre2. (A) 3D structure obtained from sequence MH753362.1, (B) 3D structure obtained from sequence MH753363.1

**Table 2** A list of virulence factors encoding genes other than Cry toxin of *B. thuringiensis* isolate BA04

Virulence factor	Number	Locus tag
Bacteriocin	3	
albE		s15_02010
albF		s15_02009
albA		s15_06070
Proteases		
Bacillolysin	4	
npr1		s15_00620
npr_2		s15_00919
npr_3		s15_04094
nprM		s15_05144
Collagenases	5	
colA_1		s15_00088
colA_2		s15_01985
colA_3		s15_03845
colA_4		s15_03865
colA_5		s15_05435
Immune inhibitor	4	
ina_1		s15_01231
ina_2		s15_04166
ina_3		s15_04315
ina_4		s15_05414
Phospholipase	3	
plc_1		s15_02179
ytpA		s15_03179
plc_2		s15_04172
Chitinases	3	
chiA1_1		s15_00514
chiA1_2		s15_01673
chid		s15_01944
N-acyl homoserine lactonase	3	
attM_1		s15_00124
attM_2		s15_00836
aiiA		s15_01744
Enterotoxin	1	
hblA_1		s15_00184
Hemolysin A	1	
tlyA		s15_02850
Hemolysin C	1	
tlyC		s15_01423
Hemolysin BL	6	
hblA_1		s15_00185
hblA_2		s15_00186
hblA_3		s15_01601
hblA_4		s15_01602
hblA_5		s15_02546
hblA_6		s15_02547
Hemolysin II	1	
Hly		s15_03857
Gamma hemolysin	1	

**Table 2** (continued)

Virulence factor	Number	Locus tag
hlgB		s15_03550
Alveolysin	1	
Alv		s15_05712
Zeta toxin	1	
Toxin A		s15_00623
toxA		s15_03685

**Table 3** List of genes encoding virulence associated factors of *B. thuringiensis* isolate BA04

Genes	Numbers	Locus tag
<b>Antitoxin protein</b>	4	
ndoAI		s15_00663
yokJ		s15_05447
hipB		s15_05774
yezG		s15_01132
<b>Spore photoproduct lyase</b>	1	
splG		s15_02820
<b>Extracellular metalloprotease precursor</b>	1	
mpr		s15_02435
<b>Putative metalloprotease</b>	1	
ypwA		s15_04590
<b>ATP-dependent zinc metalloprotease</b>	1	
ftsH		s15_01911
<b>Genes involved in nitrogen metabolism</b>	8	
ureA (gamma)		s15_03980
ureB (beta)		s15_03979
ureC (alpha subunit)		s15_03978
ureD		s15_03974
ureE		s15_03977
ureF		s15_03976
ureG		s15_03975
ureI_1		s15_03973
<b>Capsule biosynthesis protein</b>	2	
CapA		s15_01142
CapA_2		s15_01629
<b>Stage V sporulation protein T</b>	2	
spoVT		s15_01898
spoVS_2		s15_01999
<b>Stage II sporulation protein E</b>	2	
spollE		s15_01908
spollGA		s15_02128
<b>Spore germination protein</b>	5	
gerD		s15_02579
yndE		s15_02226
gerBA_2		s15_02225
gerAA_2		s15_03155
gerPF_2		s15_03223



**Table 4** A list of antibiotic resistance genes identified in *B. thuringiensis* isolate BA04

Antibiotic resistance gene	Numbers	Locus tag
<b>Multidrug resistance</b>	19	
mdtH_1		s15_00672
mdtH_2		s15_01211
mdtH_3		s15_02501
mdtH_4		s15_03959
mdtH_5		s15_05299
stp_1		s15_01506
norM_1		s15_03626
norM_2		s15_04412
ykkC		s15_02811
ykkD_1		s15_02494
ykkD_2		s15_02810
bmr3_1		s15_00395
bmr3_2		s15_03132
bmr3_3		s15_04158
bmr3_4		s15_05294
bmrA_1		s15_01552
bmrA_2		s15_02657
emrK		s15_02986
emrY		s15_01657
<b>Vancomycin B-type resistance</b>	3	
vanW_1		s15_03872
vanW_2		s15_04098
vanW_3		s15_05415
<b>Putative multidrug resistance ABC</b>	5	
yheH_1		s15_02508
yheH_2		s15_04023
yhel_1		s15_02507
yhel_2		s15_02528
yhel_3		s15_04024
<b>Tetracycline resistance protein, class C</b>	6	
tetA_1		s15_00274
tetA_2		s15_01626
tetA_3		s15_02496
tetA_4		s15_04169
tetA_5		s15_04827
tetO		s15_05454
<b>Daunorubicin/doxorubicin resistance ATP-binding protein</b>	6	
drA_1		s15_01763
drA_2		s15_02952
drA_3		s15_04767
drA_4		s15_04937
drA_5		s15_05968
drB		s15_01764
<b>Tellurite resistance protein</b>	1	
TehB		s15_04950
<b>fosfomycin resistance protein</b>	1	
FosB		s15_05424

**Table 4** (continued)

Antibiotic resistance gene	Numbers	Locus tag
<b>Multiple antibiotic resistance protein</b>	2	
marA		s15_05470
marR		s15_01821
<b>Bifunctional polymyxin resistance protein</b>	1	
arnA		s15_00573
<b>Bicyclomycin resistance protein</b>	1	
bcr		s15_00728
<b>Fosmidomycin resistance protein</b>	1	
Fsr		s15_01651

to be involved either in adhesion, recognition or degradation that increases the pathogenicity of the bacteria [18]. These virulence factors help to invade the host atrociously, for example, chitinases are involved in the degradation of chitin present in exoskeletons of insects, bacillolysin (Zinc metal endopeptidase) which performs the hydrolysis of the amino leucine and phenylalanine that elicit the innate immune system [25]. Collagenases and phospholipases were reported to be associated with the disruption of the intestine and mid-gut epithelial cells that help the pathogen to colonize inside the host [26, 27]. Additionally, the strains were found to carry different types of immune inhibitors (4 genes) which are metalloprotease causing the degradation of antibacterial proteins produced by the host insect and helping the bacteria to invade the host cells [28]. Capsule biosynthetic genes play a role in the synthesis of the polymeric capsule that protects it from the pathogen [29]. In *B. anthracis*, capA, capB, and capC proteins are linked to the synthesis of polyglutamic acid capsules which are vital for host invasion [30].

Most bacteria produce bacteriocins that have broad-spectrum antimicrobial properties against viruses, fungi, and cancer cells [31] that help the bacteria to survive and compete with other microbes. BA04 strain possesses three genes encoding for bacteriocins which might have similar roles to play. The bacterial strains were also reported to carry sequences for Zeta toxins, which were known to cause toxicity against the gram-positive and gram-negative bacteria and were reported to involve in programmed cell death in bacteria [32, 33]. Zwittermycin (ZmA) another important antimicrobial protein identified from the strain found to enhance the toxicity of the crystal protein against the host insects. Also, the above antimicrobial genes and numerous multi-drug resistance genes were identified in bacteria which depict that the strain can survive extreme exposure to multiple antibiotics.

**Table 5** Details of CRISPR sequences of BA04 genome obtained through the CRISPRfinder

CRISPR candidate	CRISPR start position	CRISPR end position	CRISPR length	Direct Repeat consensus sequences	DR length	No. Of Spacers
1	615,849	615,958	109	TACGTAGATAAAGAAGGAAAAGAAATCGCA CAACGCAATA	40	1
2	2,288,456	2,288,577	121	TGAACGCCCTGAGGTCCAGTAGG	23	2
3	4,184,871	4,184,944	73	CATCACCATGGAGGACACAATCA	23	1
4	4,622,492	4,622,808	316	TTTCTTAGCTTCTTCTCAGCTTTTTTCT	29	5
5	5,727,442	5,727,519	77	CGCTGGAGCTGGATATAACTAAAG	25	1

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats, DR Direct repeats

The WGS also revealed a cluster of urease synthase genes in the BA04 strain. Previous studies have demonstrated that these genes are involved in the recycling of environmental nitrogen and act as a virulence factor in pathogenic microorganisms associated with gastric ulceration and urinary stone formation [34]. The urease gene family creates favorable conditions for the bacterial pathogen by neutralizing the gastric acid in the guts of the insect host and helping the bacterial pathogen to use the ammonia for protein synthesis [34]. The insecticidal activity of *B. thuringiensis* is generally attributed to the production of compounds such as chitinase that degrade chitin. These insecticidal chitinases are mostly involved in the perforation of the peritrophic membrane that helps the microbes and their toxins to enter inside the peritrophic membrane of the host insect, thereby enhancing the activity of membrane binding toxins which increases the virulence of the pathogen [35, 36].

A recent toxin gene co-occurrence network study has revealed that different strains of *B. thuringiensis* are capable of accumulating multiple toxins with similar targets in a single cell [37], which serves as one of the most powerful strategies for delaying the development of host resistance. All these clusters of information unveil the complexity and the use of different pathways to cause toxicity against the insect host by different *B. thuringiensis* strains. Hence, the information generated through the WGS of *B. thuringiensis* strain BA04 would be helpful to understand the underlying mechanisms of these pathways. Furthermore, WGS would also facilitate comparative genomic studies to elucidate microbial evolutionary relationships.

The whole genome sequencing has revealed that a single *B. thuringiensis* strain may contain more than one type of crystal protein-encoding gene. Previously the Bt strain HD-1 was found to contain six crystal protein-encoding genes (viz. *Cry1Aa*, *Cry1Ab*, *Cry1Ac*, *Cry1Ia*, *Cry2Aa*, and *Cry2Ab*) [18]. In the present investigation, we identified two new types of *Cry* genes MH753362.1 and MH753363.1 are similar to *Cry1Ia* (KJ710646.1) and

*ICP-6*-like gene (KM053257.1) respectively. However, both these sequences have significant variations which suggest that these are novel crystal protein-encoding genes. Previous studies have shown that proteins grouped in CryII are effective against insects belonging to Plutellidae, Chrysomelidae, Tortricidae, Noctuidae families, etc. [38]. However, the precise mode of action of ICP-6-like proteins is yet to be unknown.

The 3D analysis showed that apart from the normal active three domains (Domain-I, Domain-II, and Domain-III) of Cry protein, both the sequences have a few additional domains that could be part of the protoxin or non-truncated protein. Based on the previous reports, domain-I having a cluster of seven  $\alpha$ -helices is responsible for pore formation in the membrane [39, 40], whereas domain-II is composed of three anti-parallel  $\beta$ -sheets and domain-III has a  $\beta$ -sandwich of two antiparallel  $\beta$ -sheets involved in receptor recognition and binding [41–45]. In most cases, the activated toxin binds to receptor molecules of midgut epithelial cells and forms non-selective pores near the vicinity of the receptor binding site that cause the lysis of cells and finally kill the insects [46, 47]. The sequence MH753362.1 had a total of four numbers of domains (Fig. 7A) which includes one additional domain which is similar to that of the protoxin of Cry1Ac [39]. Likewise, the sequence MH753363.1 showed two additional domains besides the active three-domain structure (Fig. 7B). In the present study, the 3D analysis of both MH753362.1 and MH753363.1 genes were found to have similar structural domains.

The phylogenetic tree obtained through the genome-to-genome distance comparison (GGDC) tool placed the BA04 strain close to *B. thuringiensis* strain HS18-1. This strain isolated from the Sichuan basin of China was found to confer high toxicity against both lepidopteran and dipteran insects. The strain contained a few potential insecticidal genes (*Cry30Ga*, *Cry4Cb1*, *Cry50Aa1*, *Cry69Ab1*, *Cry30Ea*, *Cry54Ba*, *Cry70Aa*, *Cry71Aa*, *Cry72Aa*, and *Cry56Aa*) based on the WGS [20]. The WGS provided information about the genetic makeup of BA04 and the

possibility of using it as a biocontrol agent against target insect pests. Therefore, knowing the genomic evolution of BA04 was pertinent.

The genomic islands (GIs) represent mostly the group of genes related to the horizontal origin that involves genetic exchange in bacteria and the archaeal genome [23]. GIs play a major role in genomic evolution and adaptation in a particular habitat. The Island viewer 4 is a tool based on a comparative genomics approach that helps to identify the genomic islands. The genes which are found in GIs of BA04 are mostly responsible for encoding virulence factors, antimicrobial compounds, and metal resistance genes that are effective against pathogen outbreaks [48–50].

In the present investigation, the CRISPR finder (<http://crispr.i2bc.paris-saclay.fr/Server/>) is used to identify the CRISPR sequences. CRISPR finder is an efficient tool as it allows the identification of the CRISPR sequences and their characteristics with their precise locations in the genome [51]. CRISPR elements play a major role in the bacterial immune system which helps to eliminate foreign genetic materials [52]. The presence of CRISPR elements in the BA04 strain confirmed the role of these sequences to resist any exogenous DNA of bacteriophages.

The whole genome sequencing of strain BA04 has made it easy to understand the genetic makeup along with the identification of two new insecticidal genes with other virulence factors. The availability of these genome sequences in the database will further help in genome annotation and evolutionary studies.

## Conclusion

Whole genome sequencing is a rapid way to characterize a microbe that can explore the genetic makeup with accuracy. Databases like NCBI, DDBJ, and EMBL played a crucial role in annotating the whole genome sequences of BA04. The whole genome sequence analysis of our local strain BA04 helped to generate information on two novel types of *Cry* genes which can be further used for the development of bioformulations and insect-resistant transgenic crops. The availability of the WGS of BA04 from India contributed to enriching the Bt database and could be used as a reference strain for the characterization of the Bt strain from NE of India in the future.

## Methods

### Bioassay

BA04 was cultured and incubated at 37°C for 72 h, then spores were collected and washed with sterile water and diluted to an OD of 0.5 to 0.6. The artificial diets of size 1 cm squares were dipped into the spore solution and feed to various instars of larvae of *Helicoverpa armigera*

as described by El-kersh and co-workers in 2016 [53]. The mortality data were recorded after one week.

### Growth condition and genomic DNA isolation

The total DNA was isolated according to the protocol mentioned by Sullivan and Klaenhammer [54] with slight modification. The culture was incubated at 37°C (120 rpm) overnight in 100 ml of T3 medium [55] and the pallet was collected through centrifugation at 6000xg in an Eppendorf tube. The pallet was resuspended in 200 µl of lysozyme buffer (25% sucrose, 30 mg/ml lysozyme) and incubated at 37 °C for 30 min., then 400 µl of SDS buffer (1% SDS, 0.2N NaOH) was added to the mixed solution and incubated for 7 min. Then 300 µl of ice-cold sodium acetate (pH-4.8) was mixed thoroughly and incubated on ice for 5 min. The lysate was centrifuged at 12000xg (4°C) for 15 min and the supernatant was collected and treated with phenol three times. The DNA was precipitated with two volumes of ethanol followed by centrifugation at 13000xg for 15 min at 4°C. The pallet was washed with 70% alcohol, dried, and dissolved in 40 µl of Tris–EDTA (Tris–CL; 10 mM and EDTA; 1 mM) buffer and the quality was observed under 0.8% agarose gel.

### Library construction

High-quality DNA was used for constructing the library with the help of the TruSeq Nano DNA kit (Cat. No. FC-121–4001). The sequencing library is prepared by random fragmentation of the DNA sample followed by 5' and 3' adapter ligation. Alternatively, "tagmentation" combines the fragmentation and ligation reactions into a single step which greatly increases the efficiency of the library preparation process. Adapter-ligated fragments are then PCR amplified and gel purified. The purified products were subjected to sequencing.

### Sequencing and annotation

The whole genome sequencing (WGS) was done on the Illumina SBS (sequencing by synthesis) platform. This sequencing technology utilizes a proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands. As all 4 reversible, terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces raw error rates compared to other technologies. The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive sequence regions and homopolymers. The quality of the raw data (sequences) were analyzed under Fast QC which allows for performing simple quality control checks on raw sequence data obtained from high throughput sequencing. The sequence reads were

filtered before assembly so that for a pair of PE (Pair End) reads, each read has more than 90% of bases with base quality greater than or equal to Q20. The value of K-mer was analyzed using the JELLYFISH. These K-mers are the sequences of length K that were obtained during the sequencing of DNA. The De-novo assembly was done on the SOAPdenovo, a novel short-read assembly method where locations of protein-coding sequences, tRNA genes, and rRNA genes were identified.

Then the functions of these sequences were annotated by using the Prokka, an online platform for rapid annotation of prokaryotic genomes that performs a series of the process automatically (<http://www.vicbioinformatics.com/software/prokka.shtml>). The whole genome sequence of the BA04 strain was also annotated under RAST (Rapid annotation using subsystem technology) online annotation server (<http://rast.theseed.org/FIG/rast.cgi>) [56]. This annotation service determines the tRNA, rRNA, and protein-coding genes. Apart from these it also predicts the metabolic pathway involved in the genome by comparing it with other annotated genomes through the KEGG (Kyoto Encyclopaedia of Genes and genome) pathway. The Island Viewer 4 was used to predict the virulence and resistance genes and their location in the genome. (<http://www.pathogenomics.sfu.ca/islandviewer/results/PBvHBZk27SXKvY1oH92c64/>). The CRISPR elements were identified under CRISPR finder online web server based on the CRISPR database [51].

The whole genome phylogenetic analysis was performed on GGDC (Genome to genome distance calculator) platform, which compares the distance between genomes with pairwise alignment between the query and the searched database sequences. It also gives high support values and insignificant subspecies conflicts thereby enabling genome-based species delineation analogous to the traditional DNA-DNA hybridization method.

#### Amplification and cloning of *Cry* gene

To amplify the identified full-length *Cry* gene sequences primers have been synthesized. Both forward and reverse primers were used to amplify MH753362 (Fw-5'-ATGAAA CCCAAAAATCAAATAAGTG-3'; Rv-5'-CTAAATGTT ACGCTCAATATTGAGTTG-3') and MH753363.1 (Fw-5'-ATGGAACCTTATGCTGTATTATC-3'; Rv-5'-TTA ACTTTTTGACACTTGAATTAAGT-3'). PCR program was set up for 35 cycles with pre-denaturation at 95 °C for 5 min, denaturation at 95 °C for 1 min, annealing 1.5 min at 48 °C for sequence MH753362.1 and 52 °C for sequence MH753363.1, extension at 72 °C for 2 min and final extension at 72 °C for 10 min. The PCR-amplified products were analyzed on 1% Agarose gel. The products were purified and cloned into the pGEM-T easy cloning vector.

The two novel types of *Cry* genes sequences were also processed for 3D analysis using the online server Phyre2 to study the structural similarities with the other crystal proteins (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>).

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-023-02821-0>.

**Additional file 1: Supplementary figures**

**Additional file 2: Supplementary table 1**

**Additional file 3: Supplementary table 2**

#### Acknowledgements

This work was supported by grants from the Department of Biotechnology (DBT), Government of India. The funders had no role in the study design, collection of samples, analysis of data, interpretation of data, the writing of this manuscript, or the decision to submit this work for publication.

#### Authors' contributions

The experiment was designed and conceived by SA and MR, MR executed and performed bioinformatic analyses of the WGS data, DD performed a few bioinformatic analyses, TK performed PCR and isolate maintenance. SA and MR co-wrote the manuscript, and BKS reviewed the manuscript.

#### Funding

The project was funded by the Department of Biotechnology, Govt. of India.

#### Availability of data and materials

The whole genome sequences have been submitted to the NCBI [Accession number PUWY00000000(PUWY01000001-PUWY01000094) with Bioproject number PRJNA436085].

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

There is no competing interest between the authors.

#### Author details

<sup>1</sup>Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat-13, Assam, India. <sup>2</sup>Silkworm Pathology Section, Central Sericultural Research and Training Institute, Central Silk Board, Ministry of Textile, Govt of India, Berhampore, West Bengal 7421 01, India. <sup>3</sup>Department of Biotechnology-Northeast Centre for Agricultural Biotechnology (DBT-NECAB), Assam Agricultural University, Jorhat-13, Assam, India.

Received: 13 September 2022 Accepted: 13 March 2023

Published online: 31 March 2023

#### References

- Schnepf E, Crickmore N, Rie VJ, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH. *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol Mol Biol Rev.* 1998;62(3):775–806.
- Bechtel DB, Bulla LA. Electron Microscope Study of Sporulation and Parasporal Crystal Formation in *Bacillus thuringiensis*. *J Bacteriol.* 1976;127:1472–81.



3. Ibrahim MA, Griko N, Junker M, Bulla LA. *Bacillus thuringiensis*: A genomics and proteomics perspective. *Bioeng Bugs*. 2010;1:31–50. <https://doi.org/10.4161/bbug.1.1.10519>.
4. Beegle CC, Yamamoto T. History of *Bacillus thuringiensis* Berliner research and development. *Can Entomol*. 1992;124:587–616.
5. Feitelson JS, Payne J, Kim L. *Bacillus thuringiensis*: insects and beyond. *Bio/Technology*. 1992;10:271–5. <https://doi.org/10.1038/nbt0392-271>.
6. Feitelson JS. The *Bacillus thuringiensis* family tree. In: Kim L, editor. *Advanced engineered pesticides*. New York, N.Y: Marcel Dekker, 1993;Inc. 63–71.
7. Crickmore N. *Bacillus thuringiensis* toxin classification. *Bacillus thuringiensis and Lysinibacillus sphaericus*. 2017; p 41–52.
8. Beard CE, Ranasinghe C, Akhurst RJ. Screening for novel *Cry* genes by hybridization. *Lett Appl Microbiol*. 2001;33(3):241–5.
9. Jones GW, Nielsen-Leroux C, Yang Y, Yuan Z, Dumas VF, Monnerat RG, Berry C. A new *Cry* toxin with a unique two-component dependency from *Bacillus sphaericus*. *FASEB J*. 2007;21(14):4112–20.
10. Kalman S, Kiehne KL, Libs JL, Yamamoto T. Cloning of a novel *CryI/C*-type gene from a strain of *Bacillus thuringiensis* subsp. *galleriae*. *Appl Environ Microbiol*. 1993;59(4):1131–7.
11. Shu C, Yu H, Wang R, Fen S, Su X, Huang D, Zhang J, Song F. Characterization of two novel *Cry8* genes from *Bacillus thuringiensis* strain BT185. *Curr Microbiol*. 2009;58(4):389–92.
12. Tan F, Zhu J, Tang J, Tang X, Wang S, Zheng A, Li P. Cloning and characterization of two novel crystal protein genes, *Cry54Aa1* and *Cry30Fa1*, from *Bacillus thuringiensis* strain BTMC28. *Curr Microbiol*. 2009;58(6):654–9.
13. Uribe D, Martínez W, Cerón J. Distribution and diversity of *Cry* genes in native strains of *Bacillus thuringiensis* obtained from different ecosystems from Colombia. *J Invertebr Pathol*. 2003;82(2):119–27.
14. Vilas-Bôas GT, Lemos MV. Diversity of *Cry* genes and genetic characterization of *Bacillus thuringiensis* isolated from Brazil. *Can J Microbiol*. 2004;50(8):605–13.
15. Rabha M, Sharma S, Acharjee S, Sarmah BK. Isolation and characterization of *Bacillus thuringiensis* strain native to Assam soil of Northeast India. *3Biotech*. 2017;7:303–10.
16. Rabha M, Acharjee S, Sarmah BK. Multilocus sequence typing and *VIP* gene profile of locally isolated *B. thuringiensis* native to Northeast India of Assam. *World J Microb Biotechnol*. 2018; 34:103. <https://doi.org/10.1007/s11274-018-2489-5>
17. Rasko DA, Altherr MR, Han CS, Ravel J. Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol Rev*. 2005;29:303–29.
18. Zhu L, Peng D, Wang Y, Ye W, Zheng J, Zhao C, Han D, Geng C, Ruan L, He J, Yu Z, Sun M. Genomic and transcriptomic insights into the efficient entomopathogenicity of *Bacillus thuringiensis*. *Sci Rep*. 2015;5:14129.
19. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto AB. *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis* one species on the basis of genetic evidence. *Appl Environ Microbiol*. 2000;66:2627–30.
20. Li Q, Zoua T, Ai P, Pana L, Fua C, Li P, Zheng A. Complete genome sequence of *Bacillus thuringiensis* HS18-1. *J Biotechnol*. 2015;214:61–2.
21. Janssen BD, Hayes CS. The tmRNA ribosome rescue system. *Adv Protein Chem Struct Biol*. 2012;86:151–91.
22. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*. 2009;25:119–20. <https://doi.org/10.1093/bioinformatics/btn578>.
23. Bertelli C, Laird M R, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoag G, Winsor GL, Brinkman FSL. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res*. 2017; 45:W30–W35; doi:<https://doi.org/10.1093/nar/gkx343>.
24. Nielsen-Leroux C, Gaudriault S, Ramarao N, Lereclus D, Givaudan A. How the insect pathogen bacteria *Bacillus thuringiensis* and *Xenorhabdus/Photorhabdus* occupy their hosts. *Curr Opin Microbiol*. 2012;15:220–31.
25. Huang S, Ding X, Sun Y, Yang Q, Xiao X, Cao Z, Xia L. Proteomic analysis of *Bacillus thuringiensis* at different growth phases by using an automated online two-dimensional liquid chromatography-tandem mass spectrometry strategy. *Appl Environ Microbiol*. 2012;78(15):5270–9. <https://doi.org/10.1128/AEM.00424-12>.
26. Peng D, Lin J, Huang Q, Zheng W, Liu G, Zheng J, Zhu L, Sun M. A novel metalloproteinase virulence factor is involved in *Bacillus thuringiensis* pathogenesis in nematodes and insects. *Environ Microbiol*. 2016;18(3):846–62. <https://doi.org/10.1111/1462-2920.13069>.
27. Schwan WR, Demuth A, Kuhn M, Goebel W. Phosphatidylinositol-specific phospholipase C from *Listeria monocytogenes* contributes to intracellular survival and growth of *Listeria innocua*. *Infect Immun*. 1994;62:4795–803.
28. Dalhammar G, Steiner H. Characterization of inhibitor A, a protease from *Bacillus thuringiensis* which degrades attacins and cecropins, two classes of antibacterial proteins in insects. *Eur J Biochem*. 1984;139:247–52.
29. Han CS, Xie G, Challacombe JF, Altherr MR, Bhotika SS, Bruce D, Campbell CS, et al. Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis*. *J Bacteriol*. 2005;188(9):3382–90.
30. Ezzell JW, Welkos SL. The capsule of *Bacillus anthracis*, a review. *J Appl Microbiol*. 1999;87:250.
31. López-Meza JE, Ochoa-Zarzosa A, Barboza-Corona JE, Bideshi DK. Antimicrobial peptides: current and potential applications in biomedical therapies. *Biomed Res Int*. 2015;2015:367243. <https://doi.org/10.1155/2015/367243>.
32. Zielenkiewicz U, Ceglowski P. The toxin-antitoxin system of the streptococcal plasmid pSM19035. *J Bacteriol*. 2005;187(17):6094–105.
33. Mutschler H, Gebhardt M, Shoeman RL, Meinhardt A. A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biol*. 2011;9:3:e1001033.
34. Mobley HL. The role of *Helicobacter pylori* urease in the pathogenesis of gastritis and peptic ulceration. *Aliment Pharmacol and Ther*. 1996;10:57–64.
35. Huber M, Cabib E, Miller L. Malaria parasite chitinase and penetration of the mosquito peritrophic membrane. *Proc Nat Acad Sci USA*. 1991;88:2807–10.
36. Sampson MN, Gooday GW. Involvement of chitinases of *Bacillus thuringiensis* during pathogenesis in insects. *Microbiology*. 1998;144:2189–94.
37. Zheng J, Gao Q, Liu L, Liu H, Wang Y, Peng D, Ruan L, Raymond B, Suna M. Comparative Genomics of *Bacillus thuringiensis* reveals a path to specialized exploitation of multiple invertebrate hosts. *M bio*. 2017;8:4e00822-17.
38. Ruiz de Escudero I, Estela A, Porcar M, Martínez C, Oguiza JA, Escriche B, Ferré J, Caballero P. Molecular and Insecticidal Characterization of a *Cry1I* Protein Toxic to Insects of the Families Noctuidae, Tortricidae, Plutellidae, and Chrysomelidae. *Appl Environ Microbiol*. 2006; p.4796–4804; doi:<https://doi.org/10.1128/AEM.02861-05>.
39. Evdokimov AG, Moshiri F, Sturman EJ, Rydel TJ, Zheng M, Seale JW, Franklin S. Structure of the full-length insecticidal protein *Cry1Ac* reveals intriguing details of toxin packaging into in vivo formed crystals. *Protein-Science*. 2014;23:1491–7.
40. Gómez I, Sánchez J, Miranda R, Bravo A, Soberón M. Cadherin-like receptor binding facilitates proteolytic cleavage of helix  $\alpha$ -1 in domain-I and oligomer pre-pore formation of *Bacillus thuringiensis* *Cry1Ab* toxin. *FEBS Lett*. 2002;513:242–6.
41. Li JD, Carroll J, Ellar DJ. Crystal structure of insecticidal delta-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. *Nature*. 1991;353(6347):815–21.
42. Grochulski P, Masson L, Borisova S, Pusztai-Carey M, Schwartz JL, Brousseau R, Cygler M. *Bacillus thuringiensis* *CryIA(a)* insecticidal toxin: crystal structure and channel formation. *J Mol Biol*. 1995;254(3):447–64.
43. Morse RJ, Yamamoto T, Stroud RM. Structure of *Cry2Aa* suggests an unexpected receptor-binding epitope. *Structure*. 2001;9(5):409–17.
44. Galitsky N, Cody V, Wojtczak A, Ghosh D, Luft JR, Pangborn W, English L. Structure of the insecticidal bacterial delta-endotoxin *Cry3Bb1* of *Bacillus thuringiensis*. *Acta Crystallogr D Biol Crystallogr*. 2001;57(8):1101–9.
45. Bravo A, Gill SS, Soberón M. *Bacillus thuringiensis* mechanisms and use. *Comp. Mole Insect Sci*. Elsevier BV. 2005; pp.175–206.
46. Vachon V, Laprade R, Schwartz JL. Current models of the mode of action of *Bacillus thuringiensis* insecticidal crystal proteins: A critical review. *J Invertebr Pathol*. 2012;111:1–12. <https://doi.org/10.1016/j.jip.2012.05.001>.
47. Kirouac M, Vachon V, Noel JF, Girard F, Schwartz JL, Laprade R. Amino acid and divalent ion permeability of the pores formed by the *Bacillus thuringiensis* toxins *Cry1Aa* and *Cry1Ac* in insect midgut brush border membrane vesicles. *Bba-Biomembranes*. 2002;1561:171–9. [https://doi.org/10.1016/S0005-2736\(02\)00342-5](https://doi.org/10.1016/S0005-2736(02)00342-5).
48. Ho Sui SJ, Fedynak A, Hsiao WW, Langille MG, Brinkman FS. The association of virulence factors with genomic islands. *PLoS ONE*. 2009;4(12): e8094.

49. Hall RM. Salmonella genomic islands and antibiotic resistance in *Salmonella enterica*. *Future Microbiol.* 2010;5(10):1525–38.
50. Pagano M, Martins AF, Barth AL. Mobile genetic elements related to carbapenem resistance in *Acinetobacter baumannii*. *Braz J Microbiol.* 2016;47(4):785–92.
51. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 2007;35:W52–7. <https://doi.org/10.1093/nar/gkm360>.
52. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Sci.* 2007;315(5819):1709–12. <https://doi.org/10.1126/science.1138140>.
53. El-kersh TA, Ahmed AM, Al-sheikh YA, et al. Isolation and characterization of native *Bacillus thuringiensis* strains from Saudi Arabia with enhanced larvicidal toxicity against the mosquito vector *Anopheles gambiae* (s.l.). *Parasites Vectors.* 2016;9:647–61.
54. Sullivan DJ, Klaenhammer TR. Rapid Mini-Prep Isolation of High-Quality Plasmid DNA from *Lactococcus* and *Lactobacillus* spp. *Appl Environ Microbiol.* 1993;59(8):2730–3.
55. Travers RS, Martin PAW, Reichelderfer CF. Selective process for efficient isolation of soil *Bacillus* sp. *Appl Environ Microbiol.* 1987;53:1263–6.
56. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genom.* 2008;9:75. <https://doi.org/10.1186/1471-2164-9-75>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

