Comparative genomic analysis and mosquito larvicidal activity of four *Bacillus thuringiensis* serovar *israelensis* strains

Giselly B. Alves¹, Fernando L. Melo², Eugenio E. Oliveira³, Khalid Haddi⁴, Lara T. M. Costa¹, Marcelo L. Dias¹, Fabrício S. Campos³, Eliseu J. G. Pereira³, Roberto F. T. Corrêa¹, Sergio D. Ascêncio⁵, Gil R. Santos¹, Guy Smagghe⁶, Bergmann M. Ribeiro² & Raimundo W. S. Aguiar¹*¹

*Bacillus thuringiensis* serovar *israelensis* (*Bti*) is used to control insect vectors of human and animal diseases. In the present study, the toxicity of four strains of *Bti*, named T0124, T0131, T0137, and T0139, toward *Aedes aegypti* and *Culex quinquefasciatus* larvae was analyzed. The T0131 strain showed the highest larvicidal activity against *A. aegypti* (LC₅₀ = 0.015 µg/ml) and *C. quinquefasciatus* larvae (LC₅₀ = 0.035 µg/ml) when compared to the other strains. Furthermore, the genomic sequences of the four strains were obtained and compared. These *Bti* strains had chromosomes sizes of approximately 5.4 Mb with GC contents of ~35% and 5472–5477 putative coding regions. Three small plasmids (5.4, 6.8, and 7.6 kb) and three large plasmids (127, 235, and 359 kb) were found in the extrachromosomal content of all four strains. The SNP-based phylogeny revealed close relationship among isolates from this study and other *Bti* isolates, and SNPs analysis of the plasmids 127 kb did not reveal any mutations in δ-endotoxins genes. This newly acquired sequence data for these *Bti* strains may be useful in the search for novel insecticidal toxins to improve existing ones or develop new strategies for the biological control of important insect vectors of human and animal diseases.

During sporulation, the gram-positive bacterium *Bacillus thuringiensis* (*Bt*) produces crystalline inclusions consisting of δ-endotoxins (Cry or Cyt proteins) with insecticidal activity¹. Genomic analysis has contributed to the identification of new genes coding for toxins that are active against different insect species including orders such as Lepidoptera, Diptera²–⁷, and Coleoptera⁸. Proteins with nematicidal⁹–¹¹ and molluscicidal¹² activities have also been described. In addition, genome sequencing of *Bt* strains with diverse ecological functions has been conducted, including a endophytic strain with potential utility in the biocontrol of phytopathogens¹³. Sequencing of complete *Bt* genomes has allowed structural and functional analysis of new plasmids that enhance our knowledge of the pathogenic properties of *Bt* in targeting organisms¹⁴–¹⁷. One study reported the plasmid sequence of a *Bacillus thuringiensis* serovar *israelensis* (*Bti*) strain, and revealed that it may produce up to seven crystal-forming toxins, named Cry4A, Cry4B, Cry10A, Cry11A, Cyt1A, Cyt2Ba, and Cyt1Ca, which are all encoded by genes found in a single 127923 bp plasmid called pBtoxis¹⁸. The average size of the complete genome sequences of *Bti* is 6.1 Mb, with ~35% GC content of the chromosomal DNA and an average of 6132 coding sequences¹⁹,²⁰. Genome sequences of seven *Bti* isolates have been reported so far¹⁹,²¹–²⁴. In this study, we sequenced the genomes of four *Bti* strains, specifically T0124, T0131, T0137, and T0139 that were collected from the soil of the Tocantins state in Brazil and determined their larvicidal activity against larvae of two important mosquito species of *A. aegypti* and *C. quinquefasciatus*. Then, to better characterize these strains, we performed comparative and phylogenetic analyses among their different genomes and compared the

¹Departamento de Biotecnologia, Universidade Federal de Tocantins, Gurupi, TO, 77413-070, Brazil. ²Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF, 70910-900, Brazil. ³Departamento de Entomologia, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil. ⁴Departamento de Entomologia, Universidade Federal de Lavras, Lavras, MG, 37200-900, Brazil. ⁵Rede de Biodiversidade e Biotecnologia da Amazônia Legal (Rede Bionorte), Universidade Federal do Tocantins, Palmas, TO, 77413-070, Brazil. ⁶Department of Plants and Crops, Ghent University, 9000, Ghent, Belgium. *email: rwsa@uft.edu.br
potential insecticidal toxin genes and other virulence factors of the four *Bti* strains with the commercial *Bti* strain H14. In case we identify high anti-mosquito activity with these strains, we believe these new data are useful in the continuous search for new insecticidal toxins to improve the existing ones or develop new strategies for the biological control of important insect vectors of human and animal diseases.

### Results

#### Larvicidal activity and features of *δ*-endotoxins.

The spore-crystal mixtures of *Bti* strains T0124, T0131, T0137, and T0139 were tested against third instar larvae of *A. aegypti* and *C. quinquefasciatus*. The T0131 strain showed the highest larvicidal activity against *A. aegypti* ($LC_{50} = 0.015\mu g/ml$) compared to 14 other complete chromosomes of *B. thuringiensis* compared to 14 other complete chromosomes of *B. thuringiensis* and *C. quinquefasciatus*. Among the tested strains, two strains of *B. thuringiensis* showed similar toxicity to the reference strain H14 (Toxicity ratio $\chi^2 = 0.015$ against $\chi^2 = 0.028$ for H14). The T0131 strain presented similar toxicity to the reference strain H14 (Toxicity ratio $\chi^2 = 1.1$ against $\chi^2 = 0.028$ for H14). The T0131 strain showed lower toxicities compared to the H14 strain (Table 1). However, the SDS-PAGE analysis of crystal protein content revealed that all the strains have similar protein profiles. *δ*-endotoxins with molecular weights of 130, 70, and 27 kDa (Fig. 1A) and round morphology (Fig. 1B–E), characteristics of the *Bti* protein profile, were observed for the strains.

#### Genome features.

The average size of the chromosomal draft sequences of the T0124, T0131, T0137, and T0139 isolates was 5.4 Mb, with GC contents of ~35%. Chromosomes of these isolates contained 5477 (T0124), 5473 (T0131) and 5472 (T0137 and T0139) protein-coding genes. The number of tRNA genes was consistent over the strains (122) while small variation was seen in the number of rRNA (39–42) among them (Table 2).

| Insecticide type | Strains     | No. of insects | LC50 (95% FI) μg/ml | LC95 (95% FI) μg/ml | TR50 (95% CL) | $\chi^2$ | P     |
|------------------|-------------|----------------|---------------------|---------------------|---------------|---------|-------|
| *A. aegypti*     | T0124       | 525            | 0.069 (0.061–0.077) | 0.243 (0.21–0.31)   | 5.2 (4.9–5.7) | 4.36    | 0.36  |
|                  | T0131       | 525            | 0.015 (0.012–0.018) | 0.045 (0.03–0.077)  | 1.1 (0.9–1.3)  | 0.56    | 0.13  |
|                  | T0137       | 525            | 0.165 (0.149–0.182) | 0.534 (0.45–0.68)   | 12.6 (11.2–13.6) | 5.06   | 0.28  |
|                  | T0139       | 525            | 0.123 (0.096–0.157) | 0.404 (0.28–0.79)   | 9.4 (8.2–10.3) | 9.4     | 0.05  |
|                  | H14         | 175            | 0.013 (0.011–0.016) | 0.037 (0.03–0.05)   | *              | 4.33    | 0.16  |
| *C. quinquefasciatus* | T0124     | 450            | 0.172 (0.157–0.188) | 0.467 (0.39–0.59)   | 6.4 (5.7–6.8)  | 1.97    | 0.58  |
|                  | T0131       | 450            | 0.035 (0.031–0.039) | 0.101 (0.08–0.14)   | 1.3 (1.2–1.4)  | 0.72    | 0.86  |
|                  | T0137       | 525            | 0.239 (0.219–0.261) | 0.630 (0.54–0.78)   | 8.6 (7.9–9.4)  | 7.07    | 0.13  |
|                  | T0139       | 375            | 0.250 (0.220–0.283) | 0.791 (0.63–1.10)   | 9.3 (8.2–9.8)  | 3.36    | 0.19  |
|                  | H14         | 175            | 0.028 (0.024–0.032) | 0.069 (0.06–0.10)   | *              | 6.49    | 0.16  |

Table 1. Lethal concentrations of *Bti* strains to larvae of *A. aegypti* and *C. quinquefasciatus*, isolated in the town of Gurupi-TO, Brazil. *FI = Fiducial Intervals; TR50 = Toxicity ratio determined by LC50 of given strain/LC50 of the reference strain H14 (*); 95% CL = 95% Confidence limits; $\chi^2$ = Chi-square for lack-of-fit to the probit model, and $P$ = Probability associated with the chi-square statistic.

#### Comparative genomic analysis.

The genome drafts of the isolates T0124, T0131, T0137, and T0139 were compared to 14 other complete chromosomes of *B. thuringiensis* (Table 5) by phylogenetic analysis and Mauve alignment (Fig. 4). The Mauve alignment showed collinearity of genes among the isolates from this study and the *Bti* strains AM65-52 and HD-789, forming 32 locally collinear blocks (LCB) (Fig. 4A). The SNP-based phylogeny revealed close relationship between the isolates T0124, T0131, T0137, T0139 and the *Bti* strains AM65-52 and HD-789 (Fig. 4B). Although a total of 2190 SNPs positions were found in all analyzed chromosomes, no SNPs were found in the chromosomes of the isolates used in this study compared to *Bti* strains AM65-52 and HD-789 (Fig. 4C). Furthermore, the plasmids with 127 kb found in the four isolates (T0124, T0131, T0137, T0139) differed only by minor nucleotide changes (1 to 7 mutations) from the pBtoxis plasmid (NC_010076), and none of the nucleotide changes was related to the *δ*-endotoxins (Table 6).

A functional gene ontology analysis was performed among the four strains (T0124, T0131, T0137, and T0139) and two strains of *B. thuringiensis* (HS18-1, and YWC2-8), which presented toxic bioactivity to dipteran insect and not associated to the serotype H14, followed by a summary from shared OrthoVenn clusters. The comparison of the inferred proteins among the strains of this study and the two other strains revealed 4829 proteins shared by the strains and a total of 231 orthologous clusters shared by HS18-1 and YWC2-8 (Fig. 5A). The HS18-1 and YWC2-8 strains presented specific genes with 6 and 64 single clusters, respectively (Fig. 5A). The analysis of all
Gene Ontology (GO) terms assigned to 4,829 orthologous clusters shared by the species showed 1,180 for metabolic processes, 1,001 for ion binding, and 1,317 for cell parts (GO-inferred terms) (Fig. 5B–D).

The important genes of the sporulation process previously described as variable and absent in some *Bacillus* species were analyzed. ORF sequences coding for the germination gene (GerB), small acid soluble proteins genes (SspP and SspH), sensor kinase (SerK) genes, coat gene of the spore (CoatB), and sigma factor genes (SigB, SigE, SigF, and SigH) were compared to the same genes present in other species of the *Bacilli* group (Fig. 6).

Higher sequence identity was observed for *B. thuringiensis* HD-789 and *B. thuringiensis* serovar *israelensis* AM65-52. The GerB and SspH genes showed the highest sequence variability when compared with the sequences acquired in this study.

**Discussion**

Here, the four new *Bti* strains T0124, T0131, T0137, and T0139, collected from the soil of the Tocantins state in Brazil, showed toxic activity to larvae of *A. aegypti* and *C. quinquefasciatus*. In addition to the fact that the mosquito strains were collected from locations not targeted by insecticide applications and hence presenting low risk of insecticide resistance build up. The *Bti* mode of action is distinct from neurotoxic or growth-regulating compounds used for mosquito control. These facts make *Bti* an effective alternative for controlling mosquito populations displaying or not resistance to these insecticides.

The *Bti* strains analyzed in this study presented different lethal concentrations among them and when compared to a reference commercial strain of *Bti* (H14). However, the δ-endotoxin gene content and toxin protein profiles assessed by SDS-PAGE were very similar. The T0131 strain presented the highest toxicity for both insect vectors and, therefore, it is probably the most promising strain for biological control among the four isolated strains.

We performed whole genome sequencing of T0124, T0131, T0137, and T0139. However, differently from the toxicity results, the genomic analysis of these isolates indicated highly similar sequences. Previous genome

**Figure 1.** Crystal protein profile of *Bti* strains. Lane 1: molecular weight marker (Invitrogen); Lane 2: T0124; Lane 3: T0131; Lane 4: T0137; and Lane 5: T0139. Seven μg of solubilized crystals from each strain were analyzed by SDS-PAGE. The ultrastructural characterization of the spores and Cry proteins from T0124 (B), T0131 (C), T0137 (D), and T0139 (E) strains. All strains presented round crystals. Arrows indicate spores (S) and round crystals (R).

| General features | T0124 | T0131 | T0137 | T0139 |
|------------------|-------|-------|-------|-------|
| Average coverage (n° reads) | 10.2  | 10.3  | 17.5  | 11.0  |
| Chromosome size (bp) | 5,415,530 | 5,414,369 | 5,414,369 | 5,414,367 |
| Sites no cover (%) | 0.7   | 0.1   | 0.06  | 0.1   |
| GC content (%) | 35.2  | 35.3  | 35.3  | 35.3  |
| CDS | 5,477 | 5,473 | 5,472 | 5,472 |
| tRNA | 112  | 112  | 112  | 112  |
| rRNA | 39   | 42   | 42   | 42   |
| Plasmids (n°) | 6    | 6    | 6    | 6    |

**Table 2.** General features of the genome sequences of *Bti* T0124, T0131, T0137, and T0139 strains.
comparison among strains of *Bt* revealed that 80% of the genes of this species are conserved, and the variability among *Bt* strains can be attributed to the acquisition of essential or non-essential genes from other microorganisms residing in the same microbial community. In addition, *Bt* has an open pan-genome which is a characteristic of species that colonize different environments and have different genetic material exchange pathways. *Bt* species comprise different subspecies and comparative analysis of the same subspecies may reveal genetically identical or highly related strains, even from different geographic regions. Such findings could be explained by the emergence of clonal lineages of pathogens that successfully colonized the biosphere, undergoing limited genetic exchange, thus representing homogeneous subspecies. Similarly, studies have shown that *Bti* also presents genomically similar strains, indicating the close relationship among them and suggesting a high degree of genomic conservation. The SNP-based phylogeny revealed close relation among the four isolates and other *Bti* isolates (HD-789 and AM65-52), in agreement with previous study, reinforcing the close genetic relationship among these bacteria.

Table 3. The nucleotide identity among plasmids of T0124, T0131, T0137, and T0139 strains and plasmids references.

| Plasmids references | pTX14-1 | pTX14-2 | pTX14-3 | pBTHD789-2 | pBtoxis | pHD1002-1 |
|---------------------|---------|---------|---------|------------|---------|-----------|
| T0124-1: 99.8%      |         |         |         | pT0124-3: 99% |         | pT0124-5: 99.9% |
| T0131-1: 99.8%      |         |         |         | pT0131-3: 99% |         | pT0131-5: 99.9% |
| T0137-1: 99.8%      |         |         |         | pT0137-3: 99.5% |         | pT0137-5: 99.9% |
| T0139-1: 99.8%      |         |         |         | pT0139-3: 99.5% |         | pT0139-5: 99.9% |

Figure 2. Comparative chromosomal nucleotide analysis of *Bti* strains. The concentric rings represent the sequences of T0124, T0131, T0137, and T0139 against the reference CP003763 strain. The black circle represents GC content of CP003763. The blue circle represents T0124, red circle represents T0131, green circle represents T0137, and purple circle represents T0139. Regions with less than 80% identity appear as blanks on each ring. This circular map was generated using the BLAST Ring Image Generator (BRIG) tool.
Since the plasmids with 235 kb and 359 kb are not reported to show any crystal-forming proteins linked with toxic activity, they have not been described in this study. In fact, the plasmid with 235 kb has been described as a conjugative plasmid and the plasmid with 359 kb encodes various metabolite transporters. 

Since the 1980s, the direct relationship between plasmids and the pathogenicity of Bt was established, as they are responsible for carrying genes that express toxins active against target insects. Plasmids of 127 kb were found in all isolates of Bti containing cry and cyt genes involved in insect toxicity. This type of plasmid, termed pBtoxis, 

| Strains | Plasmids | Average coverage (n° reads) | Standard deviation | Plasmid Size (bp) | GC (%) | CDS | Access number |
|---------|----------|----------------------------|-------------------|-------------------|--------|-----|---------------|
| T0124   | pT0124-1 | 14326                      | 3312              | 5421              | 36.1   | 4   | CP037884      |
| T0124   | pT0124-2 | 15310                      | 3005              | 6824              | 36     | 3   | CP037885      |
| T0124   | pT0124-3 | 5897                       | 982               | 7697              | 35.3   | 9   | CP037886      |
| T0124   | pT0124-4 | 109                        | 58.6              | 127922            | 32.4   | 117 | CP037887      |
| T0124   | pT0124-5 | 41.5                       | 10                | 235425            | 36.6   | 242 | CP037888      |
| T0124   | pT0124-6 | 23.8                       | 14                | 358206            | 32.3   | 338 | CP037889      |
| T0131   | pT0131-1 | 14244                      | 2742              | 5415              | 36.3   | 3   | CP037453      |
| T0131   | pT0131-2 | 14870                      | 2408              | 6824              | 36     | 3   | CP037454      |
| T0131   | pT0131-3 | 5429                       | 794.6             | 7697              | 35.3   | 9   | CP037455      |
| T0131   | pT0131-4 | 76.9                       | 41.5              | 127923            | 32.4   | 117 | CP037456      |
| T0131   | pT0131-5 | 44.5                       | 8.9               | 235425            | 36.6   | 241 | CP037457      |
| T0131   | pT0131-6 | 20.2                       | 11                | 359437            | 32.3   | 336 | CP037458      |
| T0137   | pT0137-1 | 8118                       | 2540              | 5415              | 36.3   | 3   | CP037460      |
| T0137   | pT0137-2 | 8409                       | 2676              | 6824              | 36     | 3   | CP037461      |
| T0137   | pT0137-3 | 3704                       | 832               | 7697              | 35.3   | 9   | CP037471      |
| T0137   | pT0137-4 | 124.7                      | 64.4              | 127923            | 32.4   | 117 | CP037462      |
| T0137   | pT0137-5 | 55.9                       | 15.4              | 235425            | 36.6   | 241 | CP037463      |
| T0137   | pT0137-6 | 33.5                       | 19.5              | 359440            | 32.3   | 336 | CP037464      |
| T0139   | pT0139-1 | 4585                       | 48.6              | 5415              | 36.3   | 3   | CP037465      |
| T0139   | pT0139-2 | 12014                      | 3.013             | 6827              | 36     | 3   | CP037466      |
| T0139   | pT0139-3 | 4423                       | 874.9             | 7697              | 35.3   | 9   | CP037467      |
| T0139   | pT0139-4 | 106                        | 55.6              | 127930            | 32.3   | 117 | CP037468      |
| T0139   | pT0139-5 | 44                         | 12.1              | 235425            | 36.6   | 241 | CP037469      |
| T0139   | pT0139-6 | 28                         | 18.2              | 359438            | 32.3   | 336 | CP037470      |

Table 4. General features of the assembly of complete plasmids of T0124, T0131, T0137 and T0139 strains.

Figure 3. Comparative sequence map of pT0124-4, pT0131-4, pT0137-4 and pT0139-4 plasmids. The circles starting with the outermost ring are as follows: circle 1 (pT0124-4) showing the position of δ-endotoxins; circle 2 (pT0131-4), circle 3 (pT0137-4), and circle 4 (pT0139-4) show regions of sequence similarity representing darker regions detected by BLASTN in the primary sequence (pT0124-4). Circle 5 shows GC content (deviation from average) and circle 6 illustrates the GC skew in green (+) and purple (−). The circle with δ-endotoxins and the map was generated using the Geneious tool.

Since the plasmids with 235 kb and 359 kb are not reported to show any crystal-forming proteins linked with toxic activity, they have not been described in this study. In fact, the plasmid with 235 kb has been described as a conjugative plasmid and the plasmid with 359 kb encodes various metabolite transporters.

Since the 1980s, the direct relationship between plasmids and the pathogenicity of Bt was established, as they are responsible for carrying genes that express toxins active against target insects. Plasmids of 127 kb were found in all isolates of Bti containing cry and cyt genes involved in insect toxicity. This type of plasmid, termed pBtoxis,
that may lead to variability in the detection of stress conditions, spore resistance, and germination. These genes have been described as variable and absent in some species as a result of niche-specific constraints.

Bacilli utilize carbon, and energy substances for sporulation and massive synthesis of crystal toxins. Important roles in the insecticidal activity of Bacilli are attributed to the Cry proteins, because metabolic pathways are regulated to produce amino acid, carbon, and energy substrates for sporulation and massive synthesis of crystal toxins. is widely studied and described as the only plasmid capable of encoding the crystal-forming toxins of this bacterium. In addition, pBtoxis also presents sequences with functions predicted to increase crystal formation and toxicity. Within the same serotype may share a highly related plasmid pattern; this relationship is most evident in different strains of Bacillus thuringiensis (serotypes H14) which, although isolated from different geographic origins, have the same basic plasmid pattern, sometimes even identical. Therefore, our results, that show a high degree of genomic conservation among the strains T0124, T0131, T0137, and T0139, are consistent with previous studies.

The functional gene ontology analysis from shared clusters showed a unique set of proteins identified only within the same serotype may share a highly related plasmid pattern; this relationship is most evident in different strains of Bacillus thuringiensis (serotypes H14) which, although isolated from different geographic origins, have the same basic plasmid pattern, sometimes even identical. Therefore, our results, that show a high degree of genomic conservation among the strains T0124, T0131, T0137, and T0139, are consistent with previous studies.

The comparative analysis of four new genomes of Bacillus thuringiensis carried out in the present study revealed their very high identity of nucleotide sequence. Furthermore, the results presented here are important for evolutionary studies of this species and potentially may contribute to the improvement of existing strategies or the development of new approaches in biological control that use these bacteria. Further investigations aiming to evaluate potential differences at transcriptomic/proteomic levels during specific phases (e.g., middle vegetative, early sporulation) of the four Bacillus thuringiensis strains will contribute to clarify the higher larvicidal activity described here for the T0131 strain.

### Methods

#### Isolation of Bacillus thuringiensis strains.

*B. thuringiensis* serovar *israelensis* (serotypes H14) T0124, T0131, T0137, and T0139 strains were isolated from a soil sample collected in Tocantins state (Brazil) according to the previously described protocol. The bacterial strains were cultured at 28 °C for 12 h using the streak plate method on Luria-Bertani (LB) solid medium (10 g L⁻¹ tryptone, 5 g L⁻¹ yeast extract, 10 g L⁻¹ NaCl, and 20 g L⁻¹ Agar). Single bacterial colonies of each strain were inoculated in the LB liquid medium at 28 °C with shaking for 16 h.

### Table 5. General features of chromosomes of Bacillus thuringiensis strains used in Mauve alignment and SNP-phylogenetic analysis.

| Strain | Status of assembly | Chromosome size (bp) | GC (%) | CDS | Description | Access number | Reference |
|--------|-------------------|----------------------|--------|-----|-------------|---------------|------------|
| HS18-1 | Complete          | 5 292 526            | 35.4   | 5234 | Toxicity to Lepidoptera and Diptera | CP012099.1   | Li et al.38 |
| MYBT18246 | Complete       | 6 752 490            | 35.4   | 6413 | Toxicity to nematode     | CP015350.1   | Unpublished |
| YC-10  | Complete          | 5 675 007            | 34.9   | 6028 | Toxicity to nematode     | CP011349.1   | Cheng et al.39 |
| YWC2-8 | Complete          | 5 674 369            | 35.29  | 5692 | Toxicity to Lepidoptera and Diptera | CP013055.1   | Zhu et al.40 |
| Bc601  | Complete          | 5 627 121            | 35.30  | 5485 | Used in fermentation for the production of vitamin C | CP015150.1   | Jia et al.41 |
| KNU-07 | Complete          | 5 344 151            | 35.30  | 5111 | Used in agriculture | CP016588.1   | Unpublished |
| Bt185  | Complete          | 5 243 635            | 35.30  | 4981 | Toxicity to Lepidoptera | CP014282.1   | Li et al.42 |
| HD1011 | Complete          | 5 232 696            | 35.5   | 5245 | Medical relevance | CP009335.1   | Johnson et al.43 |
| HD682  | Complete          | 5 213 295            | 35.5   | 5201 | Medical relevance | CP009720.1   | Johnson et al.44 |
| 97-27  | Complete          | 5 235 838            | 35.4   | 5216 | Medical relevance | CP010088.1   | Johnson et al.44 |
| HD571  | Complete          | 5 256 240            | 35.4   | 5219 | Medical relevance | CP009600.1   | Johnson et al.44 |
| CTC    | Complete          | 5 327 397            | 35.4   | 5268 | High producer of S-layer protein | CP013274.1   | Dong et al.44 |
| HD789  | Complete          | 5 495 278            | 35.3   | 5551 | Commercial insecticide isolate | CP003763.1   | Doggett et al.45 |
| AM65-52| Complete          | 5 499 731            | 35.0   | 5463 | Toxicity to Diptera | CP013275.1   | Bolotin, et al.46 |
| T0124  | Draft             | 5 415 530            | 35.2   | 5477 | Toxicity to Diptera | CP037890     | This study |
| T0131  | Draft             | 5 414 369            | 35.3   | 5473 | Toxicity to Diptera | CP035735     | This study |
| T0137  | Draft             | 5 414 369            | 35.3   | 5472 | Toxicity to Diptera | CP035736     | This study |
| T0139  | Draft             | 5 414 367            | 35.3   | 5472 | Toxicity to Diptera | CP035737     | This study |
Spore-crystal protein preparation and crystal analysis by SDS–PAGE. Spore-crystal mixtures were obtained according to the protocol described previously. For SDS–PAGE analysis, the crystals were purified using hexane and low speed centrifugation according to the previously described method. Proteins were suspended in a small volume of phosphate-buffered saline (136 mM NaCl, 1.4 mM KH₂PO₄, 2.6 mM KCl, 8 mM Na₂HPO₄, and 4.2 ml H₂O; pH 7.4), and fractionated by electrophoresis on 12% SDS-PAGE gels.

Scanning electron microscopy. The characterization of the spores and Cry proteins from the T0124, T0131, T0137, and T0139 strains was performed by scanning electron microscopy. The strains were cultivated in NYSM agar medium at 30 °C for 72 h, then a loop of the isolate was collected and diluted in sterile water. A volume of 100 µL of this dilution was deposited over metallic supports and dried for 24 h at 37 °C, covered with gold for 180 s using an Emitech apparatus (model K550; Quorum Technologies, Lewes, UK), and observed under a Zeiss scanning electron microscope (model DSM 962; Carl Zeiss AG, Oberkochen, Germany) at 10 or 20 Kv.

Mosquitoes and toxicity bioassays. The colonies of A. aegypti and C. quinquefasciatus were established from insects collected from the field in regions of transition between the urban and rural areas in the state of Tocantins, Brazil, (11°40′55.7″ latitude S, 49°04′3.9″ longitude W), where no insecticides have been used for the control of mosquitoes. The insects were maintained in the Entomology Laboratory of the Federal University of Tocantins, Gurupi Campus, according to the methodology described previously. Adult mosquitoes were maintained on a 10% aqueous sucrose solution and the blood of live Wistar rats (Rattus norvegicus albinus). The larvae

Figure 4. Comparative chromosome sequences of the isolates T0124, T0131, T0137, and T0139 with 14 genomes of other strains of B. thuringiensis. (A) Reciprocal LCBs in different sequences are indicated by the same colors and are connected by lines. (B) Phylogenetic tree based on the concatenated SNPs. The SNPs were called by CSI phylogeny 1.4 using HD-789 strain as reference. The branch structure was confirmed by a bootstrap consensus tree inferred from 1,000 replicates in MEGA 10. The scale bar indicates the evolutionary distance between the sequences determined by 0.10 substitutions per nucleotide at the variable positions. Red dots indicate the genomes of strains from the present study and blue dots indicate other genomes of Bti from the GenBank database. (C) The matrix shows the 2190 SNPs after pairwise comparison between isolates.
Table 6. The SNPs of the pT0124-4, pT0131-4, pT0137-4, and pT0139-4 using pBltoxis (NC_010076) as reference.

| Plasmids          | Name                  | Position | Nucleotide Change | Amino Acid Change | Codon Change | Coverage | Polymorphism Type | Protein Effect | Variant Frequency | Variant P-Value (approximate) |
|-------------------|-----------------------|----------|-------------------|-------------------|--------------|----------|-------------------|----------------|-------------------|-----------------------------|
| pT0124-4          | hypothetical protein CDS | 99670    | C -> T            | S -> F            | TCC -> TTC   | 486      | SNP (transversion) | Substitution    | 75.2%             | 0.0                        |
|                   | hypothetical protein CDS | 99354    | G -> A            | H -> Y            | CAT -> TAT   | 290      | SNP (transversion) | Substitution    | 76.9%             | 0.0                        |
|                   | hypothetical protein CDS | 99343    | A -> C            | G -> Y            | GGT -> GGG   | 304      | SNP (transversion) | None           | 77.6%             | 7.6E-194                   |
|                   | hypothetical protein CDS | 99255    | AA -> CC          | F -> G            | TTT -> GGT   | 367      | Substitution       | Substitution    | 75.2%             | 0.0                        |
|                   | hypothetical protein CDS | 99240    | GA -> TG          | S -> Q            | TCG -> CAG   | 381      | Substitution       | Substitution    | 75.9%             | 0.0                        |
|                   | hypothetical protein CDS | 99234    | T -> G            | AGG -> GGG        | 412         | SNP (transversion) | None           | 76.7%             | 0.0                        |
|                   | hypothetical protein CDS | 58097    | G -> T            | ACC -> ACA        | 241         | SNP (transversion) | None           | 100.0%            | 7.9E-25                     |
| pT0131-4          | hypothetical protein CDS | 99670    | C -> T            | S -> F            | TCC -> TTC   | 238      | SNP (transversion) | Substitution    | 79.8%             | 0.0                        |
|                   | hypothetical protein CDS | 99282    | T -> G            | I -> L            | ATA -> CTA   | 289      | SNP (transversion) | Substitution    | 75.8%             | 0.0                        |
|                   | hypothetical protein CDS | 99255    | AA -> CC          | F -> G            | TTT -> GGT   | 274      | Substitution       | Substitution    | 78.1%             | 0.0                        |
|                   | hypothetical protein CDS | 99240    | GA -> TG          | S -> Q            | TCG -> CAG   | 282      | Substitution       | Substitution    | 78.4%             | 0.0                        |
|                   | hypothetical protein CDS | 99234    | T -> G            | AGG -> GGG        | 312         | SNP (transversion) | None           | 80.8%             | 0.0                        |
|                   | hypothetical protein CDS | 99207    | T -> C            | I -> V            | ATT -> GTT   | 299      | SNP (transversion) | Substitution    | 77.6%             | 0.0                        |
|                   | hypothetical protein CDS | 58097    | G -> T            | ACC -> ACA        | 183         | SNP (transversion) | None           | 100.0%            | 5.0E-19                     |
| pT0137-4          | hypothetical protein CDS | 58097    | G -> T            | ACC -> ACA        | 323         | SNP (transversion) | None           | 99.7%             | 4.8E-63                     |
| pT0139-4          | hypothetical protein CDS | 99670    | C -> T            | S -> F            | TCC -> TTC   | 302      | SNP (transversion) | Substitution    | 75.5%             | 0.0                        |
|                   | hypothetical protein CDS | 58097    | G -> T            | ACC -> ACA        | 265         | SNP (transversion) | None           | 100.0%            | 3.2E-27                     |

Whole genome sequencing, assembly, and annotation. Total genomic DNA was extracted and purified using a Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. DNA concentration and purity were measured using a NanoDrop™ 8000 (Thermo Fisher Scientific, Waltham, MA, USA). Whole genome sequencing was performed on the Illumina MiSeq™ platform using a paired-end application (2 × 150 bp) (Illumina, San Diego, CA, USA). The read quality of the sequenced libraries was analyzed using FastQC software v 0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and sequence reads were trimmed to yield a minimum Phred quality score > 20. The genome assembly was performed using Geneious v 10.1.3 (https://www.geneious.com/). The de novo assembly was performed using Geneious assembler with medium sensitivity settings and allowing contigs with matching ends to circularize. The linear contigs were extended. For this, the reads were mapped back to the linear contigs and the resulting contigs were used as seeds for another attempted assembly until no further extension. Finally, the extended linear contigs were aligned and reordered using as reference the genome Bti HD-789 (accession number CP003763) from the “map to reference” tool with minimum overlap identity of 85%. The circular contigs were used to investigate plasmid-like sequences, by matching them against plasmid bank with custom BLAST tool. Genome annotation was added by the NCBI Prokaryotic Annotation.

Comparative genomic and phylogenetic analysis. Comparative chromosome-sequences analysis among T0124, T0131, T0137, T0139 and reference HD-789 was performed using BRIG (BLAST Ring Image
The comparative analysis of the gene sequence of sporulation for the strains considered in this study and other species from the Bacilli group was performed using blastx, and the heatmaps were generated using the MeV tool version 4.9. Venn diagrams generation and orthologous cluster annotation for biological process, molecular function, and cellular component categories were achieved using OrthoVenn. The collinearity and phylogenetic analysis among T0124, T0131, T0137, T0139, HS18-1, and YWC2-8 was performed. The collinear analysis and display of results were conducted using Mauve with the parameters reported previously. The phylogenetic tree based on single nucleotide polymorphisms (SNPs) was performed by CSI phylogeny 1.4 web using the default parameters and HD-789 as reference. The SNPs were named, concatenated and aligned, and the tree was constructed using the maximum likelihood method. The phylogeny tree inferred was reviewed using MEGA X software with 1000 replicates. The pBtoxis (NC_010076) was used as reference for the SNPs analysis of pT0124-4, pT0131-4, pT0137-4, and pT0139-4 using Geneious v 10.1.3, “Find SNPs/InDels” tool, with minimum coverage of 10, minimum variance frequency 0.75.

Nucleotide sequence accession number. The Whole Genome Shotgun projects of PRJNA521267, PRJNA521275, PRJNA521276, and PRJNA521307 Bti strains were deposited in DDBJ/ENA/GenBank under the accession numbers CP037890, CP035735, CP035736, and CP035737.

Statistical analyses. Concentration–mortality curves were estimated via probit analysis using the PROBIT procedure in the SAS statistical software package. The differential susceptibility among mosquito species to H14 and the T0124, T0131, T0137, and T0139 Bti strains was assessed based on the estimated LC50 (i.e., the lethal concentration capable of killing 50% of tested mosquito species) of each strain and the toxicity ratios (TR50) were estimated by dividing the LC50 value obtained for the T0124, T0131, T0137, and T0139 Bti strains by the LC50 value obtained for the H14 standard strain. The 95% confidence intervals estimated for these toxicity rates were considered to be significantly different if they did not include the value 1.

Ethical approval. All applicable international, national, and institutional guidelines for the care and use of animals were considered in the present investigation.
Figure 6. Heatmap comparison of the distribution of sporulation (GerB, SspP, SspH, SerK, CoatB) and sigma factor (SigB, SigE, SigF, and SigH) gene sequences among some species from the Bacillus group. Each column and line represents a gene and a Bacillus lineage, respectively, and percentage sequence identities between these species and the strains from this study were shown as colors ranging from 70% (dark blue) to 100% (red) as shown in the bottom. Undetected gene sequences are shown in black. The comparative analysis was performed using blastx and the Heatmap was generated using version 4.9.0 of the MeV tool.

Informed consent. All the authors of this manuscript accepted that the paper is submitted for publication in the Scientific Reports journal, and report that this paper has not been published or accepted for publication in another journal, and it is not under consideration at another journal.

Received: 20 May 2019; Accepted: 16 January 2020; Published online: 26 March 2020

References

1. Ramírez-Lepe, M. & Ramírez-Suero, M. Biological Control of Mosquito Larvae by Bacillus thuringiensis subsp. israelensis. (Ed. Perven, F.), Insecticides-P Eng., https://doi.org/10.5772/29139 (2012).
2. Liu, G. et al. Complete genome sequence of Bacillus thuringiensis subsp. kurstaki strain HD73. Genome Announc. 1, e0080–13, https://doi.org/10.1128/genomea.00800-13 (2013).
3. Barbosa, L. C. B. et al. Draft genome sequence of Bacillus thuringiensis 147, a Brazilian strain with high insecticidal activity. Mem. Inst. Oswaldo Cruz. 110, 822–823 (2015).
4. He, J. et al. Complete genome sequence of Bacillus thuringiensis subsp. chinenensis strain CT-43, J. Bacteriol. 193, 3407–3408 (2011).
5. Palma, L., Muñoz, D., Berry, C., Murillo, J. & Caballero, P. Draft genome sequences of two Bacillus thuringiensis strains and characterization of a putative 41.9-kDa insecticidal toxin. Toxins. 6, 1490–1504, https://doi.org/10.3390/toxins6051490 (2014).
6. Guan, P. et al. Complete genome sequence of Bacillus thuringiensis serovar sibiricus strain MC28. J. Bacteriol. 194, 6975–6975 (2012).
7. Rusconi, B., Chen, Y., Koenig, S. S., El-Helou, E. R. & Eppinger, M. Sequence of Bacillus thuringiensis strain Btm27, an Egyptian isolate highly toxic to cotton leafworm. Genome Announc. 3, e00446–15, https://doi.org/10.1128/genomeA.00446-15 (2015).
8. Zorzetti, J. et al. Genome sequence of the mosquitoicidal Bacillus thuringiensis strain BR85, a biopesticide product effective against the coffee berry borer (Hypothenemus hampei). Genome Announc. 2, e01232–15, https://doi.org/10.1128/genomeA.01232-15 (2015).
9. Iatsenko, I., Corton, C., Pickard, D. J., Dougan, G. & Sommer, R. J. Draft genome sequence of highly nematocidal Bacillus thuringiensis DB27. Genome Announc. 2, e00101–14, https://doi.org/10.1128/genomea.00101-14 (2014).
10. Wang, P. et al. Complete genome sequence of Bacillus thuringiensis YBT-1518, a typical strain with high toxicity to nematodes. J. Biotechnol. 171, 1–2, https://doi.org/10.1016/j.jbiotec.2013.11.023 (2014).
11. Liu, X. et al. Draft genome sequence of Bacillus thuringiensis NBIN-866 with high nematocidal activity. Genome Announc. 2, e00429–14, https://doi.org/10.3390/genome6082229 (2014).
12. 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. 1, e00175–12, https://doi.org/10.1128/genomeA.00175-12 (2013).
13. Jeong, H., Jo, S. H., Hong, C. E. & Park, J. M. Genome sequence of the endophytic bacterium Bacillus thuringiensis strain KB1, a potential biocontrol agent against phytopathogens. Genome Announc. 4, e00279–16, https://doi.org/10.1128/genomeA.00279-16 (2016).
14. Loza-Lara, P. D. et al. The plasmid pBMBt1 from Bacillus thuringiensis subsp. darmstadensis (INTA Mo14–4) replicates by the rolling-circle mechanism and encodes a novel insecticidal crystal protein-like gene. Plasmid. 54, 229–240, https://doi.org/10.1016/j.plasmid.2005.04.003 (2005).
15. Huang, J., Han, D., Yu, Z. & Sun, M. A novel cryptic plasmid pBMB175 from Bacillus thuringiensis subsp. tenebrionis YBT-1765. Arch. Microbiol. 188, 47–53 (2007).
16. Sheppard, A. E., Poelein, A., Rosenstiel, P., Liesegang, H. & Schulenburg, H. Complete genome sequence of Bacillus thuringiensis strain 407 Cry, Genome Announc. 1, e00158–12, https://doi.org/10.1128/genomeA.00158-12 (2013).
17. Zoratti, J. et al. Genome sequence of the mosquitocidal Bacillus thuringiensis strain BR58, a biopesticide product effective against the coffee berry borer (Hypothenemus hampei). Genome Announc. 3, e01232–13, https://doi.org/10.1128/genomea.01232-15 (2015).
18. Berry, C. et al. Complete sequence and organization of pBtox, the toxin-coding plasmid of Bacillus thuringiensis subsp. israelensis. Appl. Environ. Microbiol. 68, 5082–5095 (2002).
19. Zghal, R. Z., Ghedira, K., Elleuch, J., Kharrat, M. & Tounsi, S. Genome sequence analysis of a novel Bacillus thuringiensis strain BL5806 active against Aedes aegypti larvae, a novel potential bioinsecticide. Int. J. Biol. Macromol. 116, 1153–1162 (2018).
20. Anderson, J. et al. Comparative genome analysis of Bacillus cereus group genomes with Bacillus subtilis. FEMS Microbiol. Lett. 250, 175–184 (2005).
21. Doggett, N. A. et al. Complete genome sequence of Bacillus thuringiensis serovar israelensis strain HD-789. Genome Announc. 1, e01023–13, https://doi.org/10.1128/genomea.01023-13 (2013).
22. Alcaraz, L. F., Leitão, F. S. & Almeida, M. A. Characterization of Bacillus thuringiensis and Bacillus yuraensis reveals separate clustering and a distinct population structure of psychrotrophic strains. Appl. Environ. Microbiol. 72, 1569–1578 (2006).
23. Frost, I. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. 3, 722–732, https://doi.org/10.1038/nrmicro1235 (2005).
24. Rasko, D. A., Altherr, M. R., Han, C. S. & Ravel, J. Genomics of the Bacillus cereus group of organisms. FEMS Microbiol. Lett. 29, 303–329 (2005).
25. Fayed, N. et al. Characterization and Whole Genome Sequencing of AR23, a Highly Toxic Bacillus thuringiensis Strain Isolated from Lebanese Soil. Curr. Microbiol. 76, 1503–1511 (2019).
26. Galvão, P., Gitathy, P. D. M., Vidal, M., Araujo, J. L. S. & Baldani, J. Plasmidios em Bacillus thuringiensis: tamanho, mecanismo de replicação e papel na atividade entomopatógena. Embrapa Agrobiologia - Infoteca-e. 262, 1–32 (2009).
27. González, J. M., Dulmage, H. T. & Carlton, B. C. Correlation between specific plasmids and β-endotoxin production in Bacillus thuringiensis. Plasmid. 5, 351–365, https://doi.org/10.1016/0147-619X(81)90010-X (1981).
28. Stein, C., Jones, G. W., Chalmers, T. & Berry, C. Transcriptional analysis of the toxin-coding plasmid pBtox from Bacillus thuringiensis subsp. israelensis. Appl. Environ. Microbiol. 72, 1771–1776 (2006).
29. Reyes-Ramírez, A. & Ibarral, J. E. Plasmid patterns of Bacillus thuringiensis type strains. Appl. Environ. Microbiol. 74, 125–129 (2008).
30. Deng, C., Peng, Q., Song, F. & Lereclus, D. Regulation of cry gene expression in Bacillus thuringiensis. Toxins. 6, 2194–2209, https://doi.org/10.3390/toxins6072194 (2014).
31. Wang, J. et al. The metabolic regulation of sporulation and parasporal crystal formation in Bacillus thuringiensis revealed by transcriptomics and proteomics. Mol. Cell Proteomics. 12, 1363–1376 (2013).
32. Monnerat, R. G., Silva, S. F. & Silva-werneck, J. O. Catálogo do banco de germoplasma de bactérias entomopatogênicas do gênero Bacillus. Documentos, Brasília: Embrapa-Cenargen. 60, 65, https://doi.org/10.1016/j.biocrit.2006.11.008 (2001).
33. Lecadet, M. M., Chaufaux, J., Ribier, J. & Lereclus, D. Construction of novel Bacillus thuringiensis isolates with different insecticidal activities by transduction and transformation. Appl. Environ. Microbiol. 58, 840–849 (1992).
34. Mounief, J. R. et al. A simple method for the separation of Bacillus thuringiensis spores and crystals. J. Microbiol. Methods. 107, 147–149 (2014).
35. Sambrook, J. & Russell, D. W. Molecular cloning: A laboratory manual, (Ed 3 Lab). Cold Springs Harbour Press. 224 p. (2007).
36. Aguiar, R. W. S., Martins, E. S., Ribeiro, B. M. & Monnerat, R. G. Cry10Aa protein is highly toxic to Anthonomus grandis Bheman (Coleoptera: Curculionidae), an important insect pest in Brazilian cotton crop fields. Bt Research. 3, 20–28, https://doi.org/10.1093/bresa/30.1.26 (1984).
37. Andrews S. FastQC: a quality-control tool for high-throughput sequence (2015).
38. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28, 1647–1649 (2012).
39. Alikhan, N. F., Petty, N. K., Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC genomics. 12, 402, https://doi.org/10.1186/1471-2164-12-402 (2011).
40. Saeed, A. et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 34, 374–378 (2003).
41. Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 43, W78–W84 (2015).
42. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403 (2004).
43. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. Plos One. 9, e104984, https://doi.org/10.1371/journal.pone.0104984 (2014).
44. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 35, 1547–1549 (2018).
45. SAS Institute. SAS/STAT User's Guide, Cary, NC, USA. (2008).
46. Robertson, J. L., Jones, M. M., Olguín, E., & Alberts, B. Pesticide Bioassays with Arthropods. Boca Raton: CRC Press. 224 p. (2007).
47. Stothard, P. & Wishart, D. S. Circular genome visualization and exploration using CGView. Bioinformatics. 21, 537–539 (2004).
48. Gene ontology consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32, D258–D261 (2004).
49. Li, Q. et al. Complete genome sequence of Bacillus thuringiensis HS18-1. J. Biotechnol. 214, 61–62 (2015).
50. Cheng, F. et al. Complete genome sequence of Bacillus thuringiensis YT-10, a novel active strain against plant-parasitic nematodes. J. Biotechnol. 210, 17–18 (2015).
51. Li, Q. et al. The complete genome sequence of Bacillus thuringiensis serovar hiitisau YWCW-8. J. Biotechnol. 217, 38–39 (2016).
52. Jia, N., Ding, M. Z., Gao, F. & Yuan, Y. J. Comparative genomics analysis of the companion mechanisms of Bacillus thuringiensis Bc601 and Bacillus endophyticus Hbe603 in bacterial consortium. Sci. Rep. 6, 28794, https://doi.org/10.1038/srep28794 (2016).
62. Li, Y. et al. Complete genome sequence of Bacillus thuringiensis Bt185, a potential soil insect biocontrol agent. J. Integr. Agric. 16, 749–751 (2017).
63. Johnson, S. L. et al. Complete genome sequences for 35 biothreat assay-relevant Bacillus species. Genome ann. 3, e00151–15, https://doi.org/10.1128/genomeA.00151-15 (2015).
64. Dong, Z. et al. Complete genome sequence of Bacillus thuringiensis CTC—a typical strain with high production of S-layer proteins. J. Biotechnol. 220, 100–101 (2016).

Acknowledgements
This work was supported by grants from the Secretaria do Desenvolvimento Econômico, Ciência, Tecnologia e Inovação do Estado do Tocantins; the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (Finance Code 001); the National Council of Scientific and Technological Development (CNPq); the Minas Gerais State Foundation for Research Aid (FAPEMIG); and Fundação de Amparo a Pesquisa do Estado de Tocantins (FAPT-SESAU/TO-DECIT/SCTIE/MS_CNPQ/No 01/2017).

Author contributions
G.B.A., L.T.M.C. and M.L.D. performed the experiments, data analysis. F.L.M. and F.S.C. helped with the data analysis and performed the experiments (Bioinformatics); R.F.T.C., K.H., R.W.S.A., E.E.O., S.A.D, G.S. and E.J.G.P helped with the data analysis and review the manuscript and Statistic analysis. G.R.S. and B.M.R. helped with manuscript revision, laboratory infrastructure and reagents.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-60670-7.

Correspondence and requests for materials should be addressed to R.W.S.A.

Reprints and permissions information is available at www.nature.com/reprints.

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

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 Author(s) 2020