DOI 10.66073/ReFocus-Conf/3010034

A BRIEF COMPARATIVE BIOINFORMATICS ANALYSIS OF A KatB-LIKE CATALASE PROTEIN CANDIDATE FROM *BACILLUS THURINGIENSIS Bt26* AND REFERENCE KatB PROTEINS FROM OTHER *B. THURINGIENSIS* STRAINS

[1]Aliyev Zafar Zokirovich, **[1]Khalilov Ilxom Mamatkulovich**, **[2]Boboqulov M.Sh.**, **[2]Abdullayev A.K.**

[1] Institute of Microbiology, Academy of Sciences of the Republic of Uzbekistan
[2] Rice Research Institute aliyevzafar24@gmail.com

Abstract: This study comparatively analyzed the catalase protein candidate corresponding to the *JEEBFNLG_01767* locus in the genome of *Bacillus thuringiensis* 26 against reference KatB protein sequences from 50 other *B. thuringiensis* strains. BLASTP analysis revealed amino acid sequence identity ranging from 94.672% to 100%, with 27 reference proteins showing 100% identity. BLASTN analysis identified 26 reference sequences with 100% nucleotide identity. Multiple sequence alignment revealed 40 variable amino acid positions and a total of 125 recorded differences, with the most frequent differences occurring at positions 254, 169, and 291. Phylogenetic analysis indicated close sequence relationships between the protein candidate and many reference proteins. However, despite being annotated as *katA_1*, the locus exhibited high similarity to reference KatB proteins. Therefore, the protein is provisionally described as a KatB-like catalase candidate, and further studies are required to clarify its functional identity.

Keywords: *Bacillus thuringiensis*, *Bt*26, catalase, KatB protein, amino acid substitutions, comparative bioinformatics.

Annotatsiya: Ushbu tadqiqotda *B. thuringiensis* 26 shtammi genomidagi *JEEBFNLG_01767* lokusiga mos katalaza oqsili nomzodi boshqa 50 ta *B. thuringiensis* larning KatB oqsillari referens ketma-ketlik bilan qiyosiy tahlil qilindi. BLASTP tahlilida aminokislota identikligi 94,672–100% oralig‘ida bo‘lib, 27 ta referens oqsil bilan 100% identiklik aniqlandi. BLASTN tahlilida 26 ta referens ketma-ketlik bilan 100% nukleotid identikligi qayd etildi. Ko‘p ketma-ketlikli tekislash natijasida 40 ta o‘zgaruvchan aminokislota pozitsiyasi va jami 125 ta farq qaydi aniqlandi; eng ko‘p farqlar 254-, 169- va 291-pozitsiyalarda kuzatildi. Filogenetik tahlil oqsil nomzodining ko‘plab referens oqsillar bilan yaqin ketma-ketlik munosabatlariga ega ekanini ko‘rsatdi. Biroq lokus *katA_1* sifatida annotatsiya qilinganiga qaramay, referens KatB oqsillari bilan yuqori o‘xshashlik namoyon etdi. Shu sababli, ushbu oqsil hozircha KatB-ga o‘xshash katalaza nomzodi sifatida tavsiflanadi; uning funksional mansubligini aniqlashtirish uchun qo‘shimcha tadqiqotlar talab etiladi.

Kalit so‘zlar: *B. thuringiensis*, *Bt*26, katalaza, KatB oqsili, aminokislota almashinishlari, qiyosiy bioinformatika.

Аннотация: В настоящем исследовании проведён сравнительный анализ белка-кандидата каталазы, соответствующего локусу JEEBFNLG_01767 в геноме *Bacillus thuringiensis* 26, с референсными аминокислотными последовательностями белков KatB из 50 других штаммов *B. thuringiensis*. Анализ BLASTP выявил идентичность аминокислотных последовательностей в диапазоне от 94,672 до 100%; при этом 27 референсных белков показали 100%-ную идентичность. В результате анализа BLASTN были выявлены 26

FOR CITATION
Aliyev Z.Z., Khalilov I.M., Boboqulov M.Sh., Abdullayev A.K. A brief comparative bioinformatics analysis of a KatB-like catalase protein candidate from Bacillus thuringiensis Bt26 and reference KatB proteins from other B. thuringiensis strains // Proceedings of the Republican Scientific and Practical Conference “Status and Development Prospects of Fundamental and Applied Microbiology: The Viewpoint of Young Scientists”. – Tashkent, 2026. – P. 127–132.
DOI: https://doi.org/10.66073/ReFocus-Conf/3010034

референсных последовательностей со 100%-ной идентичностью нуклеотидных последовательностей. Множественное выравнивание последовательностей выявило 40 вариабельных позиций аминокислот и в общей сложности 125 зарегистрированных различий. Наиболее часто различия наблюдались в позициях 254, 169 и 291. Филогенетический анализ показал близкое родство последовательностей белка-кандидата со многими референсными белками. Однако, несмотря на аннотацию локуса как *katA_1*, он продемонстрировал высокую степень сходства с референсными белками KatB. Таким образом, данный белок предварительно охарактеризован как кандидат на KatB-подобную каталазу, а для уточнения его функциональной идентичности необходимы дальнейшие исследования.

Ключевые слова: *Bacillus thuringiensis*, штамм 26, каталаза, белок KatB, аминокислотные замены, сравнительный биоинформатический анализ.

Introduction

Catalases (EC 1.11.1.6) are enzymes that protect cells against oxidative damage by decomposing hydrogen peroxide into water and oxygen. They play an important role in the bacterial response to oxidative stress. Based on their structural and catalytic properties, catalases are classified into monofunctional catalases, catalase-peroxidases, and manganese catalases. Because proteins with similar sequences may differ in their enzymatic properties, functional classification requires the evaluation of multiple lines of evidence [1–3].

Different catalases with distinct physiological roles occur in the genus *Bacillus*. For example, in *Bacillus subtilis*, KatA is the major catalase in growing cells, whereas KatX is associated with sporulation and contributes to protection against hydrogen peroxide during spore germination [4]. The enzymatic activity and stability under acidic conditions of KatB isolated from *Bacillus altitudinis* SYBC hb4 have also been characterized experimentally [5]. More recently, the catalase-peroxidase activity of KatB from *B. thuringiensis* strain 97–27 and its involvement in Mn(II) oxidation were experimentally demonstrated [6]. These studies highlight the importance of comparing bacterial catalases in terms of both sequence characteristics and functional properties.

Studies of *B. thuringiensis* genomics have reported that genome annotations do not always fully correspond to protein expression data [7]. Comparative genomic analyses among strains also enable the investigation of genetic similarities and evolutionary relationships [8]. In our previous study, 60 genes associated with abiotic stress were identified in the genome of *B. thuringiensis* Bt-26, including several genes associated with oxidative stress [9].

The aim of this study was to compare the catalase protein candidate corresponding to the *JEEBFNLG_01767* locus in the *B. thuringiensis Bt26* genome with 50 reference sequences, identify amino acid differences, and assess phylogenetic relationships. The locus was annotated as *katA_1*; however, preliminary comparisons revealed high similarity to KatB proteins. This discrepancy was also an important aspect of the study.

Materials and Methods

The previously assembled and annotated genome sequence of *B. thuringiensis* strain *Bt26* was used in this study. Prokka v1.14.6 was used for the previous genome annotation [9, 10]. The nucleotide and amino acid sequences corresponding to the *JEEBFNLG_01767* locus were retrieved. Fifty reference sequences selected from the NCBI RefSeq database were used for comparison [11].

Protein and nucleotide sequence similarity was assessed using the BLASTP and BLASTN algorithms in BLAST+ v2.12.0+ [12]. Protein comparisons were evaluated based on sequence identity, alignment length, query coverage, E-value, and bit score, whereas nucleotide comparisons were evaluated based on sequence identity, coverage, and alignment length.

The *Bt26* protein candidate and reference proteins were subjected to multiple sequence alignment using MAFFT v7.526 [13]. Variable amino acid positions were identified from the aligned sequences. Phylogenetic analysis was performed using the maximum-likelihood method in IQ-TREE v3.1.4 [14]. The evolutionary model was selected using ModelFinder based on the Bayesian

information criterion (BIC) [15], and branch support was assessed using 1,000 ultrafast bootstrap replicates [16].

Results

Similarity of the *Bt26* Catalase Protein Candidate to Reference Proteins

The protein candidate corresponding to the *JEEBFNLG_01767* locus in the *B. thuringiensis* 26 genome consisted of 488 amino acids. BLASTP analysis identified 50 reference hits, with amino acid sequence identity ranging from 94.672% to 100%. Of these, 27 hits showed 100% identity.

One of the highest-scoring matches was the reference KatB protein with accession number WP_001069180.1, for which complete alignment across all 488 amino acids was observed. The query coverage was 100%, the E-value was 0.0, and the bit score was 1024. These results indicate a very high degree of sequence similarity between the *Bt*26 protein candidate and this reference protein.

Nucleotide Sequence Similarity.

In the BLASTN analysis, the nucleotide sequence corresponding to the *Bt26* catalase candidate was compared with 50 reference sequences. Complete nucleotide sequence identity (100%) was observed in 26 references. Two references showed identity values between 99% and 100%, 21 showed values between 95% and 99%, and one reference exhibited identity below 95%.

The lowest sequence identity was 92.371%. Although query coverage was reported as 100% for all hits, one alignment had a length of 1,468 nucleotides, whereas the query sequence was 1,467 nucleotides long.

Evaluating sequence identity at both the nucleotide and protein levels is important. Nucleotide substitutions may occur without changing the encoded amino acid because of the properties of the genetic code. Therefore, nucleotide sequence identity and amino acid sequence identity were considered distinct measures.

Variable Positions in the Amino Acid Sequence

Multiple sequence alignment of the *Bt26* KatB protein candidate and 50 reference proteins identified 40 variable amino acid positions. A total of 125 reference-specific differences were recorded across these positions. This value represents the cumulative number of differences observed across the reference sequences rather than 125 distinct variable positions.

The most frequent difference occurred at position 254: the *Bt26* protein contained tyrosine (Y), whereas histidine (H) was present in 22 reference sequences. At position 169, serine (S) in *Bt26* was replaced by threonine (T) in 16 references. At position 291, isoleucine (I) was observed in 12 references instead of methionine (M) in the *Bt26* protein.

Additionally, proline (P) was observed in nine references at position 73, where the *Bt26* protein contained glutamine (Q). At position 221, serine (S) or aspartate (D) was present in eight references instead of glycine (G) in Bt-26. At position 373, methionine (M) occurred in eight references instead of valine (V). At position 395, proline (P) or serine (S) was found in seven references instead of lysine (K) in the *Bt26* protein (Table 1).

Table 1. Conservation and Variable Positions in the Protein Sequence

Amino acid position	*Bt26* residue	Amino acid residue(s) observed in reference sequences	Number of reference sequences showing a difference
254	Y	H	22
169	S	T	16
291	M	I	12
73	Q	P	9
221	G	S (7), D (1)	8
373	V	M	8
395	K	P (6), S (1)	7
338	Y	C	4

DOI: 10.66073/ReFocus-Conf/3010034

352	Q	K	3
414	E	D	3

Conservation and Variable Positions in the Protein Sequence

These results indicate that although most of the protein sequence is conserved, differences occur at certain positions among the reference sequences. However, the effects of these substitutions on enzymatic activity, protein stability, or responses to oxidative stress cannot be determined from sequence data alone.

Phylogenetic Relationships

Phylogenetic analysis was performed using a total of 51 sequences, comprising the *Bt26* protein candidate and 50 reference proteins. The alignment contained 488 amino acid positions. IQ-TREE analysis identified 448 constant sites, 29 singleton sites, and 11 parsimony-informative sites.

Phylogenetic distances among many reference sequences were small. The terminal branch length of the *Bt26* protein candidate was zero, consistent with an exact sequence match to one or more reference proteins. The distance matrix indicated zero distance for 27 reference sequences.

The largest distance among the reference sequences was observed for GCF_025947955.1, with a value of 0.0630343. This result indicates that this protein sequence was more divergent from the *Bt26* candidate than the other reference sequences examined. The JTT+I model was selected for phylogenetic analysis, and 1,000 ultrafast bootstrap replicates were specified. During the analysis, the split-frequency correlation reached 0.990. This value reflects the convergence of the phylogenetic tree search and does not indicate that all internal nodes had support values of 99%. Therefore, the support value of each individual node should be considered when evaluating the reliability of specific groups.

Overall, the results indicate that the *Bt26* KatB protein candidate has very close sequence relationships with many reference proteins. However, the phylogenetic tree alone does not definitively establish the functional identity or biological activity of this protein.

Discussion

BLASTP analysis of the *Bt26* KatB protein candidate revealed high sequence identity, including complete matches with several reference proteins. These results indicate that the candidate is closely related to reference catalases. However, because the catalase family includes groups with distinct structural and functional properties, sequence similarity alone is insufficient to establish the definitive functional classification of the enzyme [1–3].

Mo et al. experimentally demonstrated the catalase-peroxidase activity of KatB from *B. thuringiensis* strain 97–27 and its involvement in Mn(II) oxidation [6]. The length of the *Bt26* protein candidate and its high similarity to reference KatB proteins provide a basis for comparison with these findings. Nevertheless, the present bioinformatics analysis does not confirm that the *Bt26* protein also oxidizes Mn(II) or possesses the same enzymatic activities.

An important issue is the discrepancy between the *katA_1* annotation assigned to the locus and its similarity to reference proteins annotated as *katB*. The distinct physiological roles of KatA and KatX in *B. subtilis* [4], together with the different biochemical properties reported for KatB proteins from other *Bacillus* strains [5], indicate that catalase genes should not be classified solely on the basis of their names. Therefore, the genomic context of the *JEEBFNLG_01767* locus, its conserved domains, and the annotations of the reference sequences require further examination.

The amino acid differences identified by MAFFT characterize sequence diversity among the proteins but do not directly demonstrate effects on enzymatic activity. The site-directed mutagenesis experiments reported by Mo et al. show that the functional significance of specific amino acid residues must be established experimentally [6]. Similarly, phylogenetic proximity reflects evolutionary relationships between sequences but does not prove functional equivalence [8, 14].

In our previous study, *katE*, *katA-1* and other stress-associated genes were identified in the *Bt26* genome [9]. The present study extends this work by characterizing the sequence-level properties of another catalase candidate and contributes to the investigation of catalase diversity in this strain.

However, the presence of a gene in the genome does not necessarily indicate its expression or activity under stress conditions. Such discrepancies between genomic predictions and proteomic evidence have also been reported in studies of *B. thuringiensis* [7].

Conclusion

Comparative analysis of the catalase protein candidate corresponding to the JEEBFNLG_01767 locus in the *Bacillus thuringiensis Bt26* genome revealed 94.672–100% amino acid identity across 50 reference proteins, including 27 exact matches. BLASTN analysis identified 100% nucleotide identity with 26 reference sequences. Multiple sequence alignment revealed 40 variable amino acid positions and 125 recorded differences. These findings indicate high sequence similarity between the candidate and many reference proteins. However, given the discrepancy between the *katA_1* annotation and its high similarity to KatB proteins, the protein should be described as a KatB-like catalase candidate, and its functional identity requires further investigation.

References

1. Chelikani P., Fita I., Loewen P.C. Diversity of structures and properties among catalases. *Cellular and Molecular Life Sciences*. 2004;61:192–208. https://doi.org/10.1007/s00018-003-3206-5
2. Zámocký M., Furtmüller P.G., Obinger C. Evolution of catalases from bacteria to humans. *Antioxidants & Redox Signaling*. 2008;10:1527–1548. https://doi.org/10.1089/ars.2008.2046
3. Imlay J.A. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nature Reviews Microbiology*. 2013;11:443–454. https://doi.org/10.1038/nrmicro3032
4. Bagyan I., Casillas-Martinez L., Setlow P. The *katX* gene, which codes for the catalase in spores of *Bacillus subtilis*, is a forespore-specific gene controlled by σF, and KatX is essential for hydrogen peroxide resistance of the germinating spore. *Journal of Bacteriology*. 1998;180:2057–2062. https://doi.org/10.1128/jb.180.8.2057-2062.1998
5. Zhang Y., Li X., Xi R., Guan Z., Cai Y., Liao X. Characterization of an acid-stable catalase KatB isolated from *Bacillus altitudinis* SYBC hb4. *Annals of Microbiology*. 2016;66:131–141. https://doi.org/10.1007/s13213-015-1089-y
6. Mo W., Mei H., Liu S., Cheng S., Wang Y., Wang J., He J., Su J. Bifunctional catalase-peroxidase KatB drives direct and indirect Mn(II) oxidation in genetically engineered bacteria: synergistic biochar immobilization enhances wastewater remediation. *Environment International*. 2025;205:109917. https://doi.org/10.1016/j.envint.2025.109917
7. Rang J., He H., Wang T., Ding X., Zuo M., Quan M., et al. Comparative analysis of genomics and proteomics in *Bacillus thuringiensis* 4.0718. *PLOS ONE*. 2015;10. https://doi.org/10.1371/journal.pone.0119065
8. Zhang W. Comparative genomic analysis of *Bacillus thuringiensis* strains for uncovering evolutionary mechanisms. *Bt Research*. 2024;15(1):20–29. https://doi.org/10.5376/bt.2024.15.0003
9. Aliyev Z.Z., Xalilov I.M., Boboqulov M.Sh. *Bacillus thuringiensis* 26 shtammi genomida mavjud abiotik stressga chidamlilikni ta’minlovchi genlarining funksional tavsifi. “Zamonaviy biotexnologiya: barqaror rivojlanish va ekologik muammolarga ilmiy yechimlar” konferensiyasi materiallari, 152–155-betlar. https://doi.org/10.5281/zenodo.22890680
10. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
11. O’Leary N.A., Wright M.W., Brister J.R., et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*. 2016;44–D745. https://doi.org/10.1093/nar/gkv1189
12. Camacho C., Coulouris G., Avagyan V., et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421

13. Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*. 2013;30:772–780. https://doi.org/10.1093/molbev/mst010
14. Wong T.K.F., Ly-Trong N., Ren H., et al. IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *Molecular Biology and Evolution*. 2026;43(5). https://doi.org/10.1093/molbev/msag117
15. Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*. 2017;14:587–589. https://doi.org/10.1038/nmeth.4285
16. Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*. 2018;35:518–522. https://doi.org/10.1093/molbev/msx281