In silico characterization of banana bunchy top virus (BBTV) and its genetic variation globally

ABSTRACT

Banana bunchy top disease is one of the most prevalent viral infections associated with banana farming. It is a rapidly spreading disease. Currently, there are very few fully sequenced isolate reports from India and around the world.When *in silico* analysis was conducted on various Banana Bunchy Top Virus (BBTV) various genome components at nucleotide level(DNA-R and DNA-U3), and amino acid level (DNA-C and DNA-U3) maximum genetic variability was observed in all reported strains of BBTV.When comparing DNA-R (Replicative protein) to DNA-S segment (Coat protein) most Indian isolates are in agreement with isolates from countries in East and Southeast Africa and belong to the PIO (Pacific-Indian Oceans) group of BBTV isolate classification.BBTV coat protein model demonstrated the highest degree of protein binding to NBS-LRR class resistance protein.Furthermore, the Ramachandran plot was used to validate the BBTV CP model, and Procheck and the PROSA web server were used to enhance the structure. The greatest binding affinity of the BBTV coat protein was seen with the NBS-LRR class resistance protein (receptor) when the BBTV coat protein model interacted with it.Three mutagenic epitope (IADEFYVERL, SKRFLLVLDD and WEFFKQCAFSS) were predicted from BBTV coat protein region. Consequently, our work presents the Indian subcontinent as a possible hotspot for rapid demographic development from an extremely small viral population size, adding significantly to the knowledge already known on BBTV worldwide.

Keywords: Banana, BBTV, Epitope, Bunchy top virus, Begomovirus, Ramachandran plot and Mutagenecity

1. INTRODUCTION

The banana (*Musa paradisica* Mill.), a fruit crop that is widely cultivated and has a high production value and economic benefits while requiring little maintenance, is a member of the *Musaceae* family. Owing to its exceptional nutritional value, it constitutes one of the most important horticultural food crops planted in tropical and subtropical countries [1,2,3].The Banana Bunchy Top Virus (BBTV), which is classified as one of the most devastating viruses on the top ranking of "World's worst invasive species list," belongs to the Genus *Babuvirus* and family *Nanoviridae*[4], causing limited production. The terrible viral disease that affects bananas is called banana bunchy top disease (BBTV). Since its discovery in 1889 in Fij, BBTV has catastrophically destroyed numerous banana varieties in all of the worldwide significant banana-growing locations over the past several years[3,5]. BBTV is transmitted in persistent, circulative manner by the black banana aphid (*Pentalonia nigronervosa*) [6,7]. The term "bunchy top" originates from the rosette-like structure of the diseased plants that exhibit severe signs, which have thin, erect, and gradually smaller leaflets. Often, the margins of the leaves curl upward and exhibit a slight browning. Midrib and petiole frequently have dark green stripes that run down into the pseudo stem. The multi-component, circular, DNA with one strand virus known as BBTV is a member of the family *Nanoviridae* and genus *Babuvirus*. Numerous molecular analyses of the BBTV genome have demonstrated that the virus possesses isometric virions [3,8], The isometric virion measures 18–20 nm in diameter. It is transmitted by banana black aphid (*P. nigronervosa*) in a persistent manner. The BBTV genome is made up of at least six essential parts, each measuring around 1 KB (BBTV DNA-R, -S, -M, -C, -N and -U) BBTV DNA-R encodes a master replication initiation protein (Rep), DNA-S encodes a viral coat protein (CP) for encapsulation, DNA-M encodes for movement protein, DNA-C encodes for cell cycle link protein, DNA-N encodes for nuclear shuttle protein while DNA-U encodes unknown function[9].

The Banana bunchy top disease has spread extensively over the globe, mostly to the continents of Africa, Australia, and Asia Pacific, as well as to other South Pacific Islands, however reports of it in American countries have not yet been received[10,11,12]. This has been made feasible by an increase in human-mediated travel, the global trade in agriculture in bananas, and inadequate sanitary laws[13]. Gujarat has the greatest production in India among the several states, averaging 177.5 metric tonnes compared to the national average of 35.50 metric tons. Numerous variables that limit productivity, including pests and illnesses, have an impact on bananas. Diseases such as banana bunchy top virus (BBTV), wilt (*Fusarium oxysporum*), and black sigatoka (*Mycosphaerella fijiensis*) result in large reductions in yield [14,15,16].

Despite being a large producer of bananas, only a small number of instances of BBTV infection have been thoroughly investigated in several of India's districts[17,18,19,20,21]. Lack genetic variation, sterility of most edible cultivars, long reproduction times, and varying ploidy levels have hindered the development of disease-resistant Musa through traditional breeding methods[22,23]. To control bunchy top disease in bananas, a number of methods have been tried, but none of them have been able to provide complete protection. Compared to many other severe viral diseases, research into the fundamental reason of resistance to this disease has progressed extremely slowly because of the challenges associated with purifying the virus[24,25,26]. Therefore, the present study was under taken on*in silico* characterization of banana bunchy top virus (BBTV). Our primary goal was to investigate attempts to genetically distinguish BBTV isolates. By comprehending the evolutionary tendency, we will undoubtedly be able to better disease management in the future and support protected agriculture practices worldwide.

2 Material and Methods:

2.1 Identified amino acid and nucleotide variation among each genome components of BBTV

Different segment or component wise sequence retrieved from NCBI database to find the variation in different nucleotide sequences for conserved sequence and motif using motif scan tool(https://www.genome.jp/tools/motif/) and MEGA X v.10.0.7[27,28]. Similarly different segment wise protein sequence were retrieved from NCBI database to find conserved domain using NCBI CCD(https://www.ncbi.nlm.nih.gov/Structure/ cdd/wrpsb.cgi) available respectively. Later analyzed to determine the optimal nucleotide and amino acid substitution model and generate a Neighbor joining phylogenetic evolutionary tree (1000 bootstrap replicates) using MEGA X v10.0.7.[27,28].

2.2 Structure Prediction of BBTV coat protein

Complete BBTV Coat-protein sequence was also used in a BLASTP-based homolog search with RCSBPDB (PDB; http://www.rcsb.org/pdb/home/home.do) to determine homologs. Following BLASTP, template sequences in fasta format with at least 30% sequence identity were obtained from RCSB-PDB. Afterwards, we decide to use the online I Tasser web server for abinitio modeling of coat protein (https://zhanggroup.org/I-TASSER/). The model's built-model quality was assessed by analyzing the amino acid area in the Ramachandran plot on the ProCheck web server. (https://servicesn.mbi.ucla.edu/PROCHECK/) [29,30,31]. The models were chosen for additional investigation based on the frequency of outliers and favoritism percentage. Later Refinement of model was done by PROSA web server(https://prosa.services.came.sbg.ac.at/prosa.php).

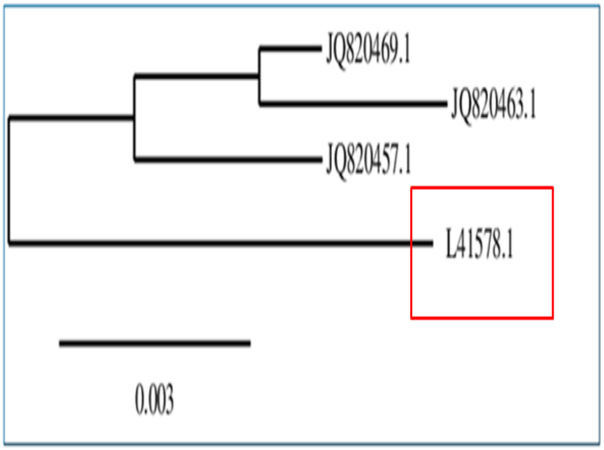
2.3 Prediction of antigenic mutagenecity peptide from BBTV coat protein

BBTV coat protein model was tested as interaction with *Musa acuminata* NBS-LRR class resistance protein using GalaxyPepDock web server(https://ardock.ibcp.fr/). Later, conserve epitope sequences from BBTV coat protein predicted using different web tool ieSVMTriP(http://sysbio.unl.edu/SVMTriP/prediction.php), NOVAPROLABS (https://www.novoprolabs.com/tools/peptide-antigen-design),PREDICTED ANTIGENIC PEPTIDES (http://imed.med.ucm.es/Tools/antigenic.pl)Emboss(http://www.bioinformatics.nl/emboss) available for antigenic mutagenecity from available amino acid sequences.

3. RESULTS AND DISCUSSION

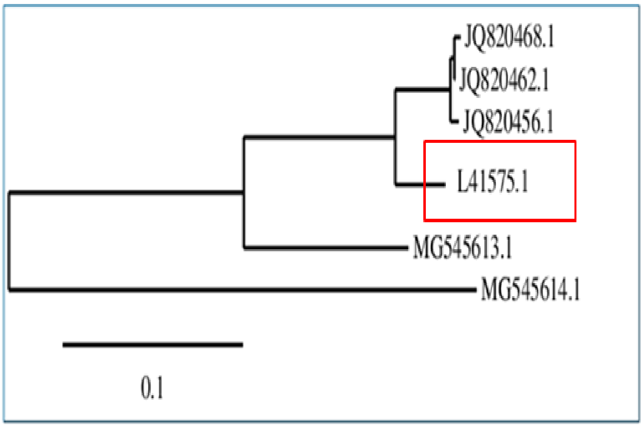
Total six segments ie DNA C,DNA M,DNA N, DNA R,DNA S and DNA U3 respectively different nucleotides sequences retrieved from NCBI taking BioProject: PRJNA485481 as reference . Segment DNA C which mainly encodes for cell cycle link protein in BBTV genome , where L41578.1 (reference), JQ820457.1 , JQ820463.1, JQ820469.1, Total 13 conserved sequence and total 32 conserved motif were found using motif scan tool. Phylogenic analysis revealed L41578.1 different cluster from JQ820457.1 , JQ820463.1, JQ820469.1 respectively. Segment DNA –M encodes for movement protein in BBTV genome, where L41575.1 (reference), JQ820456.1, JQ820462.1, JQ820468.1 , MG545613.1 , MG545614.1 (Figure.1, Figure. 7, Figure.8, Figure. 9, Figure. 10, Figure. 11, Figure. 12).

Phylogeny analysis of all DNA –C segments region (Nucleotide)



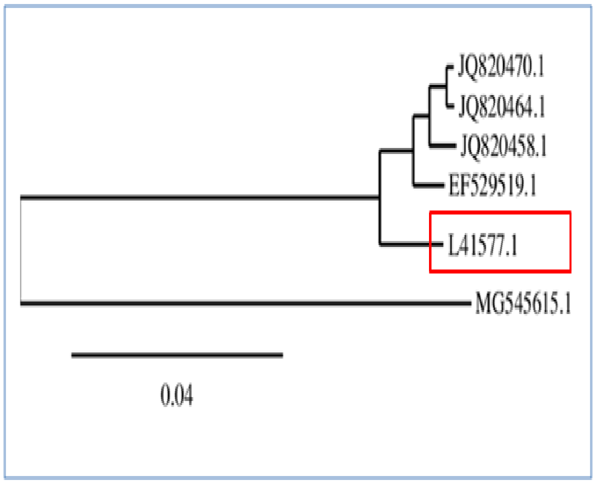
Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

Phylogeny analysis of all DNA –M segments region



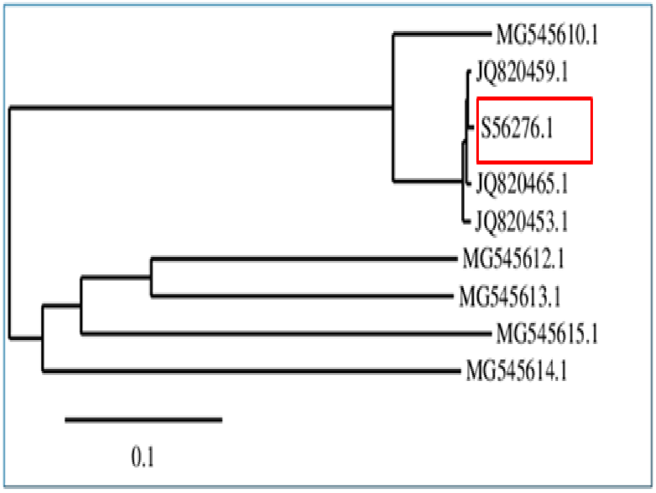
Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

Phylogeny analysis of all DNA –N segments region

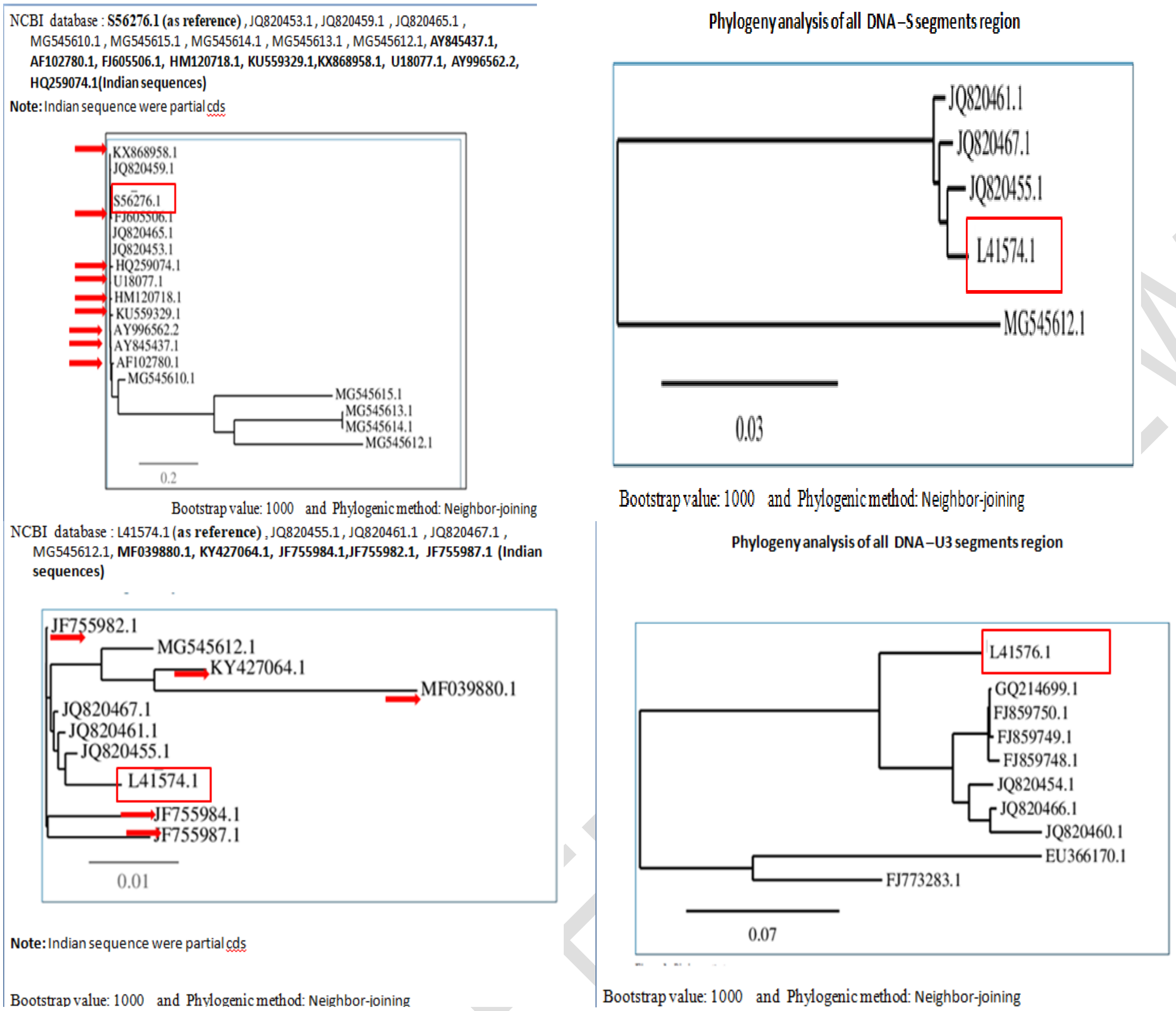


Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

Phylogeny analysis of all DNA –R segments region



Bootstrap value: 1000 and Phylogenic method: Neighbor-joining



NCBI database : **S56276.1 (as reference)** , JQ820453.1 , JQ820459.1 , JQ820465.1 , MG545610.1 , MG545615.1 , MG545614.1 , MG545613.1 , MG545612.1, **AY845437.1, AF102780.1, FJ605506.1, HM120718.1, KU559329.1,KX868958.1, U18077.1, AY996562.2, HQ259074.1(Indian sequences)**

Note: Indian sequence were partial cds

Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

Phylogeny analysis of all DNA –S segments region

Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

NCBI database : L41574.1 (**as reference**) , JQ820455.1 , JQ820461.1 , JQ820467.1 , MG545612.1, **MF039880.1, KY427064.1, JF755984.1,JF755982.1, JF755987.1 (Indian sequences)**

Note: Indian sequence were partial cds

Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

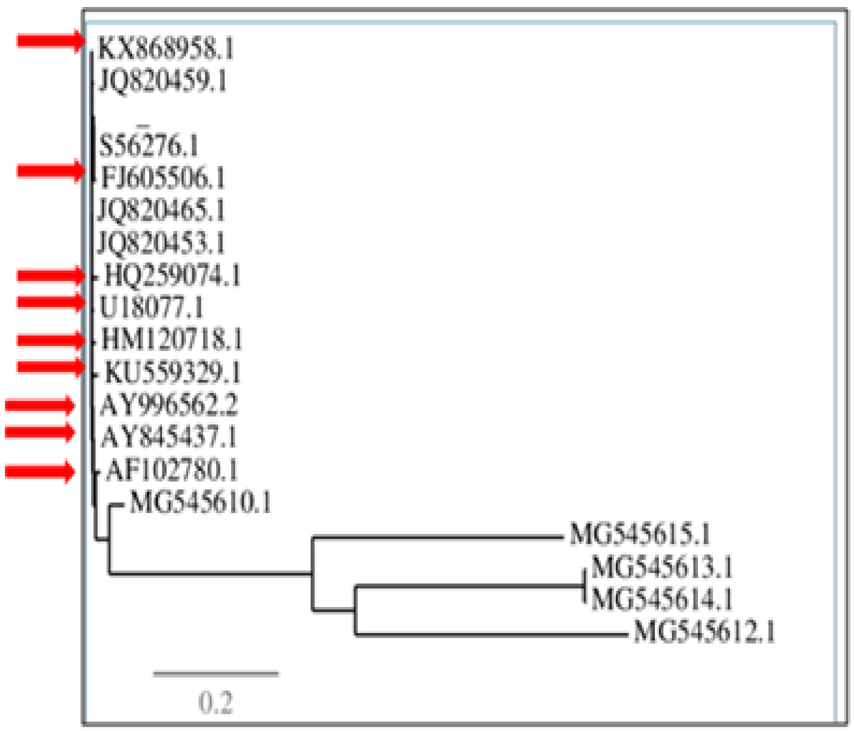
Phylogeny analysis of all DNA –U3 segments region

Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

Figure 1 Nucleotide variation of different genome components of BBTV

Total 2 conserved sequence and 54 motif were found using motif scan tool. Phylogenic analysis revealed L41575.1 found in different cluster compare to JQ820456.1 , JQ820462.1 , JQ820468.1 respectively.Segment DNA N encodes for nuclear shuttle protein in BBTV genome, where L41577.1 (reference), EF529519.1, JQ820458.1, JQ820464.1, JQ820470.1 , MG545615.1. Total 2 conserved sequences and 61 motif were found using motif scan tool. Phylogenic analysis revealed L41577.1 found in different cluster compare to EF529519.1, JQ820458.1, JQ820464.1, JQ820470.1 respectively. Segment DNA R encodes for Replicase protein in BBTV genome, where S56276.1 (reference) , JQ820453.1 , JQ820459.1 , JQ820465.1 , MG545610.1 , MG545615.1 , MG545614.1 , MG545613.1 , MG545612.1. Total 43 motif were found using motif scan tool. Phylogenic analysis revealed S56276.1 found in same cluster with JQ820459.1 and JQ820465.1. Segment DNA S encodes for Coat protein in BBTV genome, where L41574.1 (reference), JQ820455.1, JQ820461.1, JQ820467.1, MG545612.1. Total 6 conserved sequence and 49 motif found using motif scan tool. Phylogenic analysis revealed L41574.1 found in same cluster with JQ820455.1, JQ820461.1, JQ820467.1.Segment DNA U3: encodes unknown function in BBTV genome, where L41576.1 (reference) , GQ214699.1, FJ773283.1 , EU366170.1 , FJ859750.1 , FJ859749.1 , JQ820454.1 , JQ820460.1 , JQ820466.1 , FJ859748.1. In DNA- R (Replicase protein) majority of Indian isolates matches with isolates of region Rwanda(Country in East Africa) and Malawi(Southeastern Africa) compare to rest majority isolates from china whiles DNA-S segments (coat protein) majority of indian isolates matches with isolates of region Malawi(Southeastern Africa), Rwanda(Country in East Africa) and DR Congo(Central Africa) compare to rest isolates from DR Congo(Central Africa) . Total 61 motif and no conserved region found using motif scan tool. Phylogenic analysis revealed L41576.1 found in different cluster compare to all(Figure. 2, Figure.13,Figure. 14,Figure. 15,Figure. 16, Figure. 17, Figure. 18).

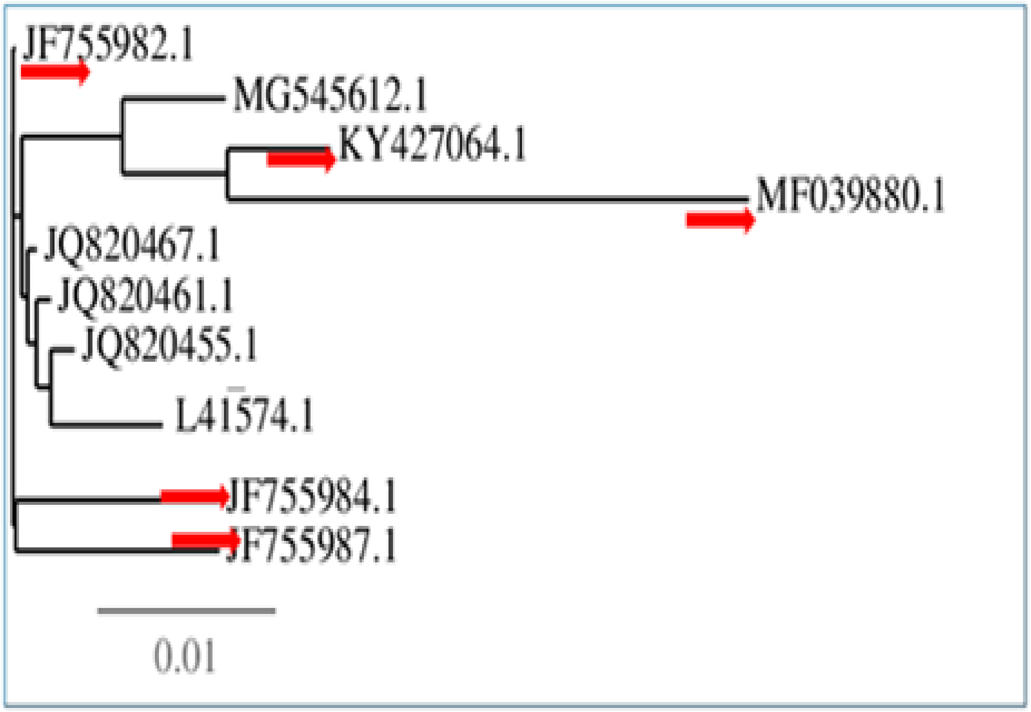
Phylogeny analysis of all DNA–R segments region (Protein)



On base of percent identity in nucleotide sequence and protein sequence of Replicase protein

Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

Phylogeny analysis of all DNA–S segments region (Protein)



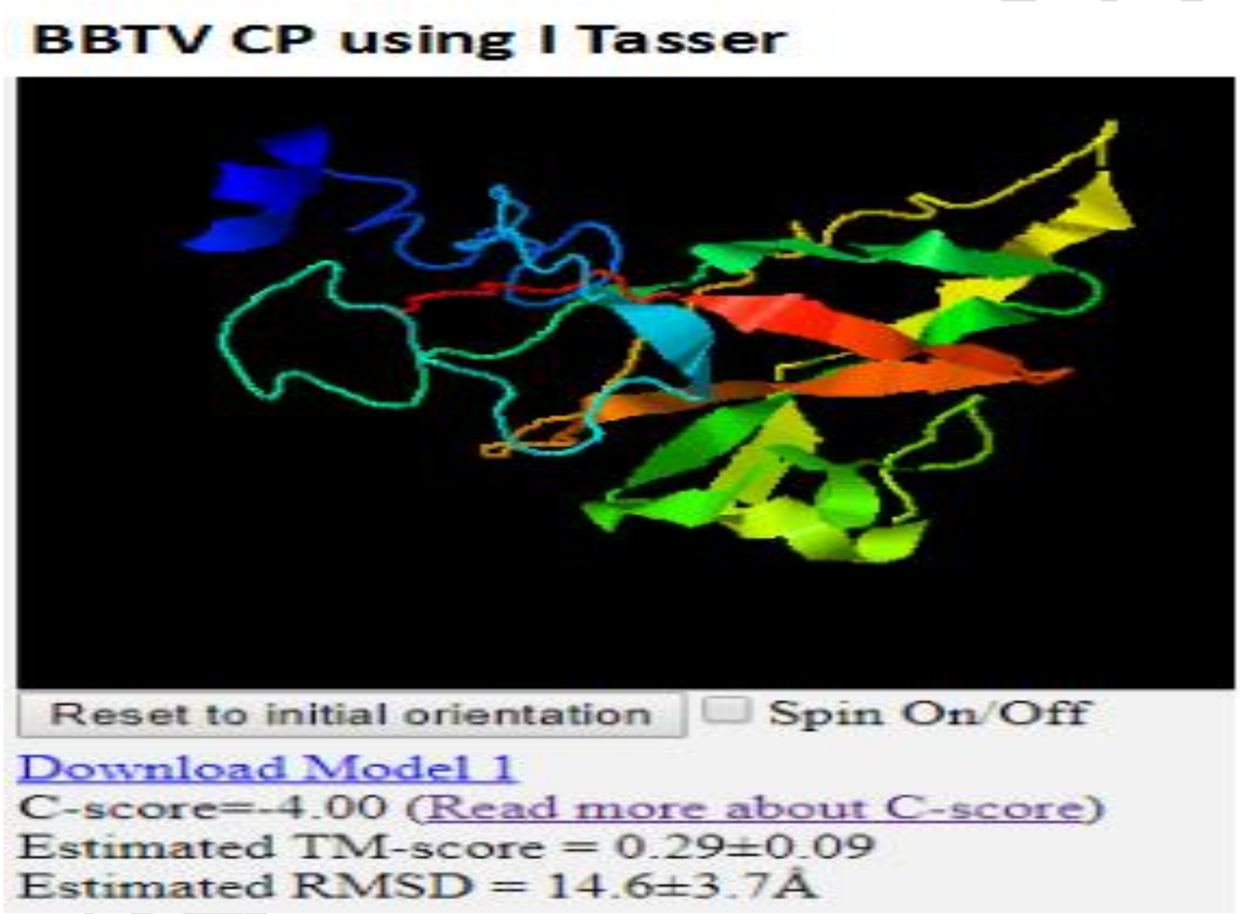
On base of percent identity in nucleotide sequence and protein sequence of coat protein

Bootstrap value: 1000 and Phylogenic method: Neighbor-joining

Figure 2 Amino acid variation of replicase and coat protein components of BBTV

Similar work was carried out [31] to find HIV genome wide variation. Similar research was carried out by [32] to find nucleotide variations in human genome .Similarly Belmabroukand his co worker [33]carried out nucleotide variation in different BBTV genome component. Conserved Domain was present in segment DNA M as Babuvirus MP super family,DNA U3 as Nanovirus C8 super family, DNA R as Viral Rep and P loop NTPase super family, DNA S as nanovirus coat super family. Homology of BBTV coat protein was carried out by search with pdb database for similarity was 36% (>50) then decide to go for abinitio modeling. Predicted BBTV CP model validation with Ramachandran plot using Rampage revealed that 83.3% aa in favoured region. Further Prosa (refinement and validation of protein structures prediction and modelling based on energy criteria validation) result revealed that 89.1% aa in favoured region. Similar research was carried out by Prajapati and his co worker [34] predicted 3D structure of AC1 Proteins of Begomovirus. Similarly Patel and kalaria [29] carried out papaya ring spot virus coat protein prediction.

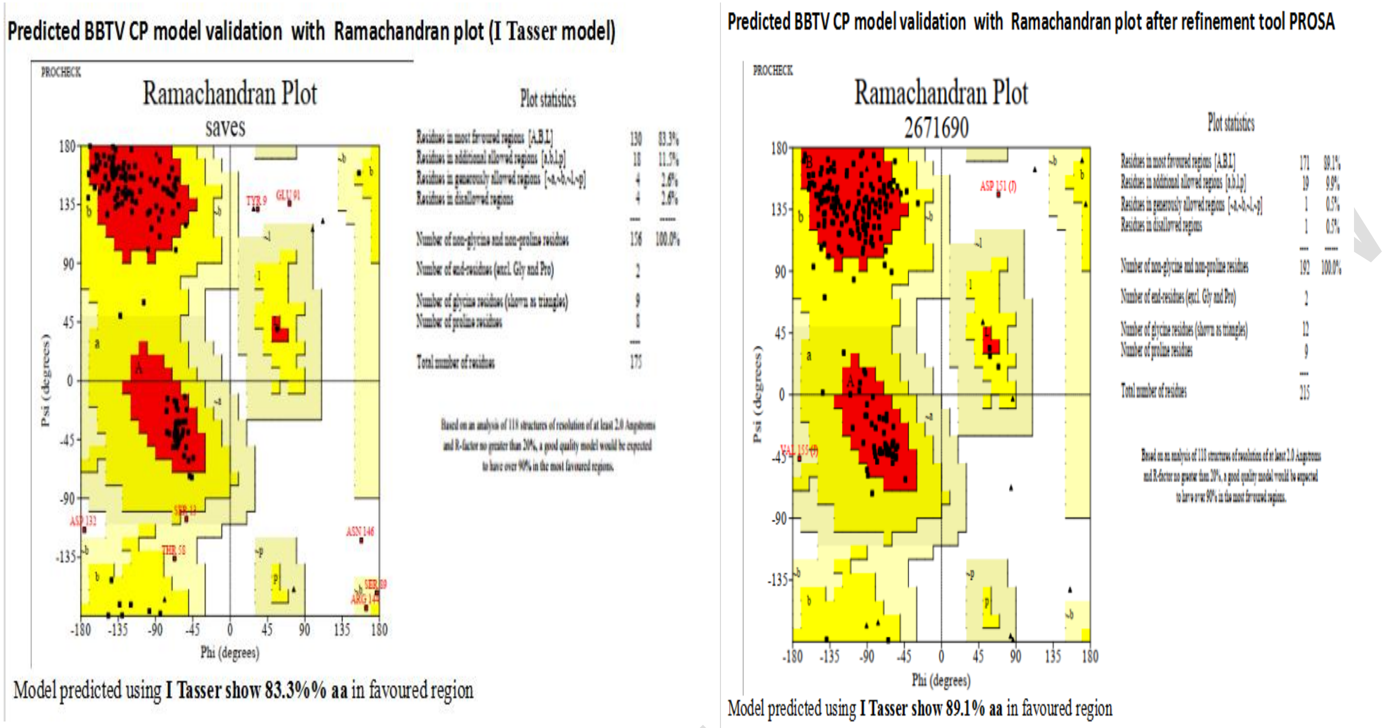
Later Identification of coat protein binding receptor in banana genome or any other plant genome and its interaction with BBTV coat protein revealed HQ704837.1 as result of BLASTP. Due to less similarity in PDB database later sequence was taken for 3D structure prediction using I Tasser web server (Figure 3). Similarly Ravindra and Kalaria [31] carried out TLCV coat protein prediction using homology modelling.



C-score=-4.00 (Read more about C-score)
Estimated TM-score = 0.29±0.09
Estimated RMSD = 14.6±3.7Å

Figure 3 Prediction of BBTV coat protein structure using I Tasser server

Vigna mungo disease resistance protein CYR1 (CYR1) mRNA, complete cds (aa) then Check above sequence similarity with (Musa (taxid:4640)) in NCBI then show 79.55% sequence similarity with Musa acuminata AAA Group NBS-LRR class resistance protein (BR-4) mRNA, partial cds (EF515836.1). An essential phase in the structural modeling process is evaluating the model's performance. After the model is constructed, it needs to be examined using the checking tools provided to determine whether the stereochemistry of the model is relatively dependable and that the usual values come from crystal structures. The accuracy of the simulated structure provided by I Tasser web server model computations was assessed using a Ramachandran plot calculation in the PROCHECK tool validation package. The result revealed 83.3% amino acid in most favourable region with 130 amino acid. Later the model was configure in PROSA web server result in 89.1% amino acid in most favourable region with 171 amino acid(Figure 4). The GalaxyPepDock web service was utilized to test the BBTV coat protein model in order to identify any interactions with the *Musa acuminata* NBS-LRR class resistance protein engaging in molecular docking with both interacting residues(Figure 5).


Predicted BBTV CP model validation with Ramachandran plot (I Tasser model)

Predicted BBTV CP model validation with Ramachandran plot after refinement tool PROSA

Model predicted using **I Tasser** show **83.3%%** **aa** in favoured region

Model predicted using **I Tasser** show **89.1%** **aa** in favoured region

Figure 4 Predicted BBTV CP model validation with Ramachandran plot

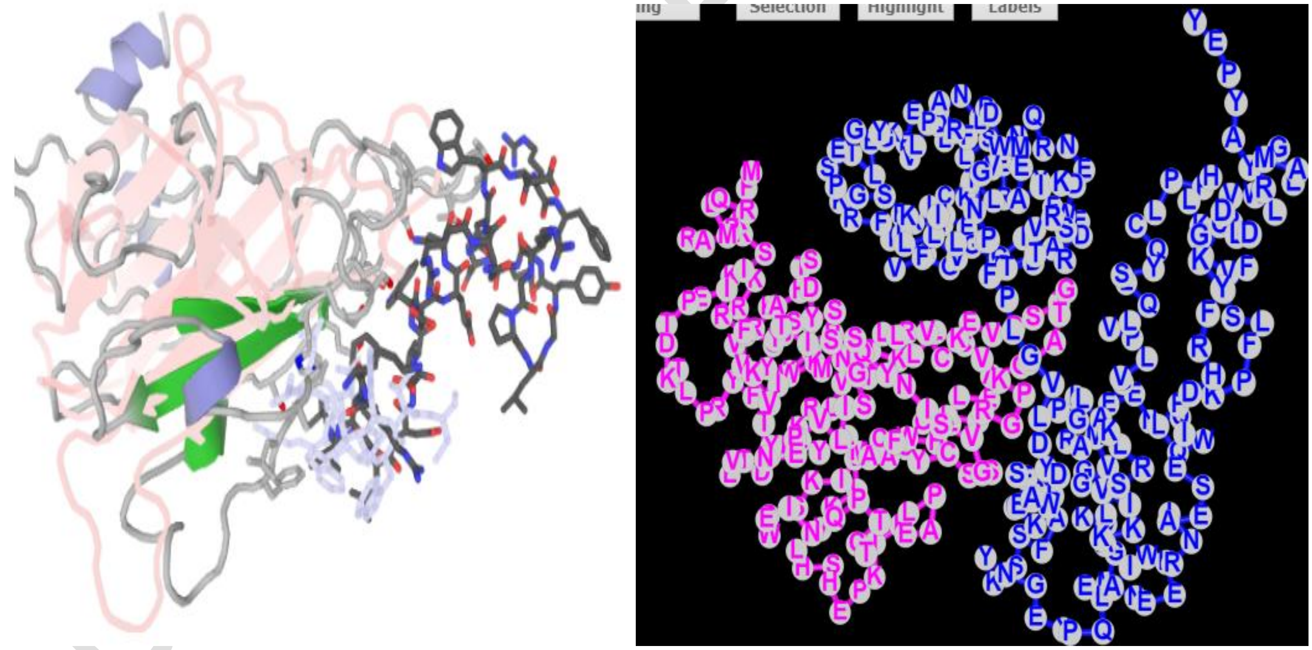
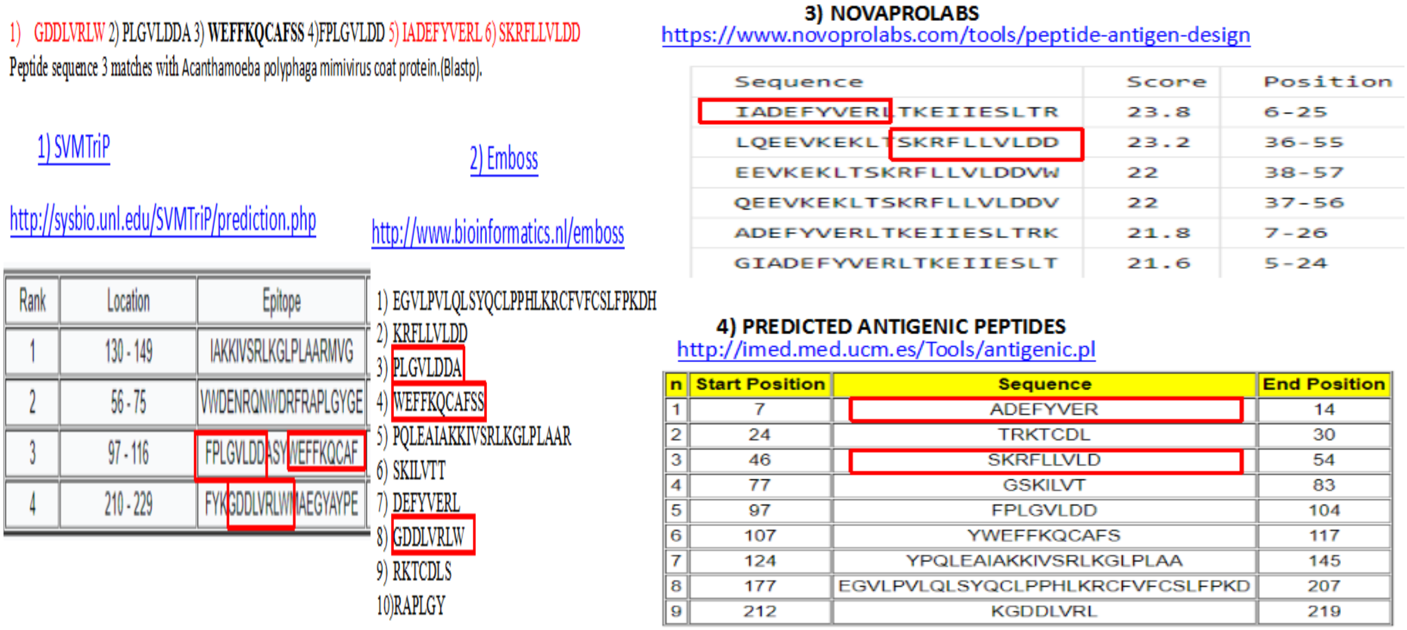


Figure. 5 Predicted BBTV coat protein model interaction with *Musa acuminata* NBS-LRR class resistance protein

Similar research was carried out by meena and his co worker [35] to predicted *in silico* homology of Garlic Common Latent Virus Coat-protein. Similarly Ravindra and Kalaria [31] carried out TLCV coat protein model refined using PROCHECK tool where number of amino acid were increased in favourable region. Similarly, Likhith and Peter also carried out *in silico* analysis of Coat Protein (CP) of Tomato Leaf Curl Virus (ToLCV) and Tomato Yellow Leaf Curl Virus

(TYLCV) .Later these models interacted with GroEL Protein of *BemisiaTabaci*to find out the interactive amino acid respectively[36].

Later NBS-LRR class resistance protein of *Musa acuminata* search for conserved amino acid residues with pfam database 1) KRFLLVLDDVWDENRQNWDRFRAPLGYGEP 2) VLPVLQLSYQCLPPHLKRCFVFCSLFPKD. Later prediction of antigenic mutagenecity peptide for BBTV coat protein result revealed 1)GDDLVRLW 2) PLGVLDDA 3) WEFFKQCAFSS 4)FPLGVLDD 5) IADEFYVERL 6) SKRFLLVLDD(Figure 6).



1) GDDLVRLW 2) PLGVLDDA 3) WEFFKQCAFSS 4)FPLGVLDD 5) IADEFYVERL 6) SKRFLLVLDD

Peptide sequence 3 matches with Acanthamoeba polyphaga mimivirus coat protein.(Blastp).

1) SVMTriP

http://sysbio.unl.edu/SVMTriP/prediction.php

Rank	Location	Epitope
1	130 - 149	IAKKIVSRLKGLPLAARMVG
2	56 - 75	VWDENRQNWDRFRAPLGYGE
3	97 - 116	FPLGVLDDASYWEFFKQCAF
4	210 - 229	FYKGDDLVRLWMAEGYAYPE

2) Emboss

http://www.bioinformatics.nl/emboss

1) EGVLPVLQLSYQCLPPHLKRCFVFCSLFPKDH
2) KRFLLVLDD
3) PLGVLDDA
4) WEFFKQCAFSS
5) PQLEAIAKKIVSRLKGLPLAAR
6) SKILVTT
7) DEFYVERL
8) GDDLVRLW
9) RKTCDLS
10)RAPLGY

3) NOVAPROLABS

https://www.novoprolabs.com/tools/peptide-antigen-design

Sequence	Score	Position
IADEFYVERLTKEIIESLTR	23.8	6-25
LQEEVKEKLTSKRFLLVLDD	23.2	36-55
EEVKEKLTSKRFLLVLDDVW	22	38-57
QEEVKEKLTSKRFLLVLDDV	22	37-56
ADEFYVERLTKEIIESLTRK	21.8	7-26
GIADEFYVERLTKEIIESLT	21.6	5-24

4) PREDICTED ANTIGENIC PEPTIDES

http://imed.med.ucm.es/Tools/antigenic.pl

n	Start Position	Sequence	End Position
1	7	ADEFYVER	14
2	24	TRKTCDL	30
3	46	SKRFLLVLD	54
4	77	GSKILVT	83
5	97	FPLGVLDD	104
6	107	YWEFFKQCAFS	117
7	124	YPQLEAIAKKIVSRLKGLPLAA	145
8	177	EGVLPVLQLSYQCLPPHLKRCFVFCSLFPKD	207
9	212	KGDDLVRL	219

Figure 6. Prediction of antigenic mutagenecity peptide for BBTV coat protein

Later three peptides sequence ie (1) GDDLVRLW, 2) IADEFYVERL,3) SKRFLLVLDD found as interacting amino acids of BBTV coat protein with NBS-LRR class resistance protein. Tian and his coworker [37] predicted mutagenic peptide using Potato virus Y Coat Protein Epitopes for Commercial monoclonal antibodies. This expands our knowledge of the mechanisms involved in antibody binding and helps create epitope-based diagnostic instruments, disease resistance, and management plans for bananas against BBTV, an economically significant crop species that causes significant production loss.

4. CONCLUSION

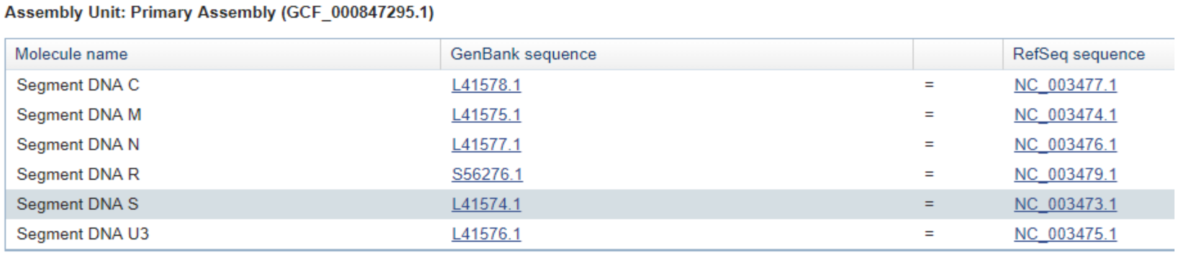
To conclude, our research findings for the first time has reported at nucleotide level of BBTV genome component, DNA –C showed maximum sequence similarity with 13 conserved region with 32 conserved motif while DNA-M and DNA-N both showed least conserved region 2 with 54 and 61 conserved motif respectively while no conserved region found in DNA-R and DNA-U3 among the different strains.At amino acid level of BBTV genome component, DNA –R showed 2 conserved domain while DNA –M , DNA –N , DNA –S each showed 1 conserved domain respectively while DNA-C and DNA-U3 showed no conserved domain among the different strains.In DNA- R (Replicase protein) majority of Indian isolates matches with isolates of region Rwanda(Country in East Africa) and Malawi(Southeastern Africa) compare to rest majority isolates from china whiles DNA-S segments (coat protein) majority of indian isolates matches with isolates of region Malawi(Southeastern Africa), Rwanda(Country in East Africa) and DR Congo(Central Africa) compare to rest isolates from DR Congo(Central Africa). Further the BBTV CP model was validated using Ramachandran plot and refinement of structure was done using Procheck and PROSA web server. BBTV coat protein model were interacted with NBS-LRR class resistance protein where maximum binding affinity of BBTV coat protein was observed with NBS-LRR class resistance protein (receptor) was predicted. Total 6 mutagenic epitope 1) GDDLVRLW 2) PLGVLDDA 3) WEFFKQCAFSS 4) FPLGVLDD 5) IADEFYVERI 6) SKRFLLVLDD of BBTV CP matches with *Acanthamoeba polyphagamimivirus* coat protein out of which 3 mutagenic epitope (1) GDDLVRLW, 2) IADEFYVERL,3) SKRFLLVLDD found as interacting amino acids with NBS-LRR class resistance protein. Generally more than 10 aa peptide are considered as epitope peptide.Furthermore, since the BBTV population is expanding globally, this study will motivate researchers to carry out more studies in other uncharted areas where banana farming is common. It will also be helpful in laying the groundwork for a more advanced and targeted detection mechanism for upcoming diagnostic tools.

REFERENCE:

1. Chakraborty, S, Dutta, S, Barman, M ,Samanta,S,Sarkar, K P, Poorvasandhya, R, Tarafdar, J.Detection and in silico characterization of banana bunchy top virus in West Bengal, India: relevance to global genetic diversity and population structure. *VirusDisease*,2023:34, 221–235
2. Anonymous. retrieved from FAO.2020. https://www.fao.org/faostat/en/#data/QCL. Accessed 25 April 2023.
3. Chandrasekar, A, Kalaiponmani, K and Angappan, K.. Banana Bunchy Top Viral Coat Protein (CP) Gene Expression Studies at Molecular Level in Hill Banana cv. Sirumalai (AAB). *International Journal of Current Microbiology and Applied Sciences,*2017:6(6): 398-411.
4. IPPC (International Plant Protection Convention), 2010. Banana Bunchy Top Virus. http://www.ippc.int/file_uploaded/1162364592814_List_of_Regulated_pests.doc. Accessed 12 Oct 2023.
5. Jones DR. Introduction to banana abaca and enset. In: Jones DR, editor. Diseases of banana abaca and enset. CABI Publishing Wallingford; 2000. p. 1–27.
6. Singh, RP, Kurz, J, Boiteau, G.. Detection of stylet-borne and circulative potato viruses in aphids by duplex reverse transcription polymerase chain reaction. *The Journal of Virological Methods*,1996:,59:189-196
7. Valerie, K, Thomas, JE, Sharman, M, Mademba-Sy F.. First record of banana bunchy top disease in New Caledonia. *Australasian Plant Pathology* ,2001:30:71
8. Su, HJ, Tsao, LY, Wu, ML, Hung, TH.. Biological and molecular categorization of strains of BBTV. *The Journal of Phytopathology*,2000:151:290-296
9. Thomas, JE, Dietzgen, RG. Purification, characterization and serological detection of virus-like particles associated with banana bunchy top disease in Australia. *Journal of General Virology*,1991:72:217-224
10. Amin, I, Qazi, J, Mansoor, S, Ilyas, M, Briddon, RW. Molecular characterization of banana bunchy top virus from Pakistan. *Virus Genes*, 2008;36:191–8.
11. Diekmann, M, Putter, CAJ. editors. Musa spp. Bioversity International. Rome: FAO/IPGRI; 1996. pp. 28
12. Chakraborty, S, Barman, M, Samanta, S, Roy, M, Tarafdar, J. Effect of banana bunchy top virus on the heat shock protein genes of *Pentalonia nigronervosa* during temperature susceptibility and its effect on virus transmission. *Agronomy Journal*. 2021;11(9):1866.
13. Stainton, D, Martin, DP, Muhire, BM, Lolohea, S, Halafihi, MI, Lepoint, P, Blomme, G, Crew, KS, Sharman, M, Kraberger, S, Dayaram, A. The global distribution of Banana bunchy top virus reveals little evidence for frequent recent human-mediated long distance dispersal events. *Virus Evolution*, 2015;1(1):1.
14. Dale, JL. Banana bunchy top: an economically important tropical plant virus disease. *Advances in Virus Research*,1987: 33: 301- 325
15. Tripathi, L, Tripathi, JN, Tushemereirwe, WK. Strategies for resistance to bacterial wilt disease of bananas through genetic engineering.*African Journal of. Biotechnology*:3 (12):688-692
16. Chandrasekar, A, Kalaiponmani, K, Elayabalan, S, Kumar, KK, Angappan, K, Balasubramanian, P. Screening of banana bunchy top virus through multiplex PCR approach. *Archives of Phytopathology and Plant Protection*,2001: 19:1920- 1925.
17. Banerjee, A, Roy, S, Behere, GT, Roy, SS, Dutta, SK, Ngachan, SV. Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian oceans group from North-East India. *Virus Research*. 2014;183:41–9.
18. Das, T, Banerjee, A. Distribution molecular characterization and diversity of banana bunchy top virus in Tripura India. *Virus Diseases*. 2018;29(2):157–66.
19. Islam, MN, Naqvi, AR, Jan, AT, Haq, QMR. Genetic diversity and possible evidence of recombination among banana bunchy top virus (BBTV) isolates. *International Research Journal of Microbiology*, 2010;1(1):1–12.
20. Kavino, M, Harish, S, Kumar, N, Saravanakumar, D, Samiyappan, R. Molecular characterization of coat protein and nuclear shuttle protein genes of banana bunchy top virus from Western Ghats in India. *Archives of phytopathology and plant protection*. 2011;44(5):405–11.

21. Selvarajan R, Sheeba M, Balasubramanian V, Rajmohan R, Dhevi NL, Sasireka T. Molecular characterization of geographically different banana bunchy top virus isolates in India. *Indian Journal of Virology*,2010;21(2):110–6.
22. Karan, M, Harding, R.M, Dale*, JL.. Evidence for two groups of banana bunchy top virus isolates. Journal* of General Virology,1994:75:3541-3546.
23. Horser, CL, Karan, M, Harding RM, Dale JL.. Additional Rep-encoding DNAs associated with banana bunchy top virus.*Archives of Virology*,2001:146:71-86
24. Elnifro, EM, Ashshi, AM, Cooper, RJ, Klapper, PE. Multiplex PCR: Optimization and Application in Diagnostic Virology. *Clinical Microbiology Reviews*,2000:13:559-570
25. Nassuth, A, Pollari, E, Helmeczy, K, Stewart, S, Kofalvi, SA. Improved RNA extraction and one-tube RT-PCR assay for simultaneous detection of control plant RNA plus several viruses in plant extracts. *Journal of Virological Methods*,2000:90:37- 49.
26. Stellrecht, KA, Woron, M, Mishrik, NG, Venezia, RA. . Comparison of Multiplex PCR Assay with Culture for Detection of Genital Mycoplasmas. *Journal of Clinical Microbiology*,2004 42 (4):1528-1533.
27. Tamura, K, Stecher, G, Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11, *Molecular Biology and Evolution*, 2021:38(7) 3022–3027
28. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*,2018:35(6):1547.
29. Patel, J and Kalaria, R K. In-silico Analysis & Homology of Papaya Ring Spot Virus & Papaya Leaf curl virus coat protein, *Bulletin of Environment, pharmacology and life science*,2018: 7(8), 19- 28.
30. Kumar, SP, Patel, S, K. Kapopara, R G, Jasrai, YT and Pandya, H A. Evolutionary andMolecular Aspects of Indian Tomato Leaf Curl Virus Coat .*International Journal of Plant Genomics*:2022: 1-12
31. Solanki, Ravindra and Kalaria, RK. Molecular Identification and in silico Characterization of Coat Protein in Tomato Leaf Curl Virus Associated in Tomato from South Gujarat Region of India. *International Journal of Current Microbiology and Applied Sciences*,2019: 8(07): 456-466.
32. Li G, Piampongsant S, Faria NR, Voet A, Pineda-Pena AC, Khouri R, Lemey P, Vandamme A-M, Theys K. An integrated map of HIV genome-wide variation from a population perspective. Retrovirology ,2015:12:18.
33. Belmabrouk, S, Kharrat, N, Abdelhedi, R, Amine, BA, Riadh, B, Ahmed, R.. Screening of nucleotide variations in genomic sequences encoding charged protein regions in the human genome. *BMC Genomics* 2017:18, 588.
34. Prajapati, R, Marwal, A,Gaur, RK. Recognition of Errors in the Refinement and Validation of Three-Dimensional Structures of AC1 Proteins of Begomovirus Strains by Using ProSA-Web", *Journal of Viruses*, 2014:6.
35. Meena, R.K. and S. Thakur, 2016. In-silico Analysis and Homology Modelling of Coat-protein of Garlic Common Latent Virus Isolates from India. *Indian Journal of Biotechnology*, .2016;26(4)45-57.
36. Likhith,R K and Peter A. Comparative in silico Analysis of Coat Protein (CP) of Tomato Leaf Curl Virus (ToLCV) and Tomato Yellow Leaf Curl Virus (TYLCV) and their Molecular Docking with GroEL Protein of Hamiltonella an Endosymbiont of their Vector *BemisiaTabaci*.*Mysore Journal of Agricultural Sciences,2023*,57(2):195.
37. Tian, YP, Hepojoki, J, Ranki, H, Lankinen, H, Valkonen, JPT. Analysis of Potato virus Y Coat Protein Epitopes Recognized by Three Commercial Monoclonal Antibodies. PLoS ONE,2014: 9(12):3456-3478.

APPENDIX



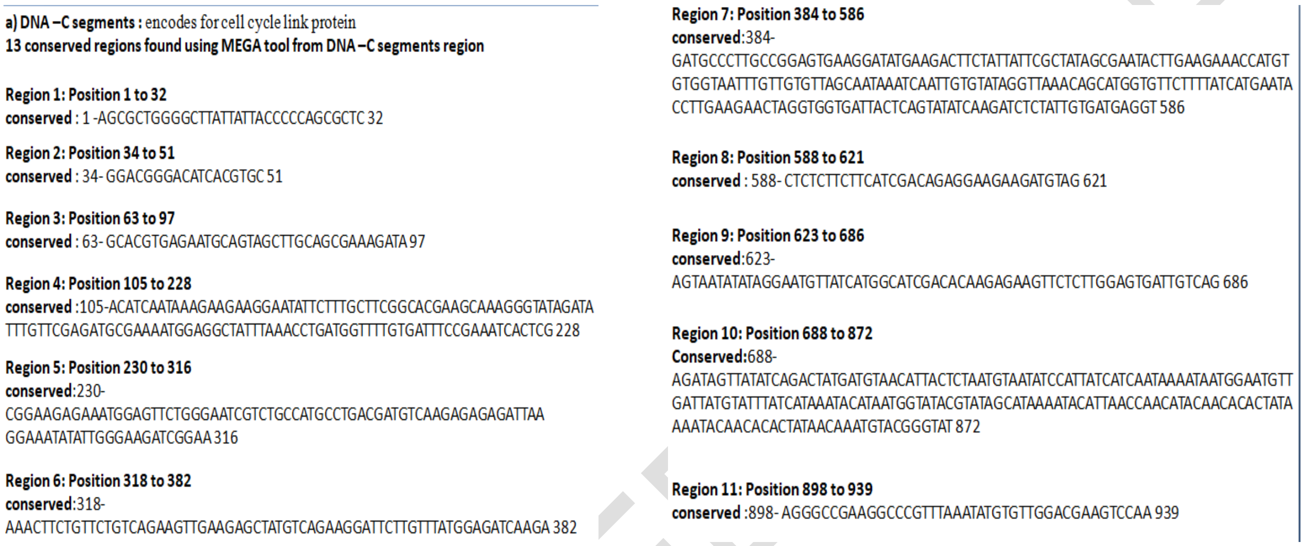
Assembly Unit: Primary Assembly (GCF_000847295.1)

Molecule name	GenBank sequence		RefSeq sequence
Segment DNA C	L41578.1	=	NC_003477.1
Segment DNA M	L41575.1	=	NC_003474.1
Segment DNA N	L41577.1	=	NC_003476.1
Segment DNA R	S56276.1	=	NC_003479.1
Segment DNA S	L41574.1	=	NC_003473.1
Segment DNA U3	L41576.1	=	NC_003475.1

ViralMultiSegProj14621BioProject: PRJNA485481

Segment DNA C: **encodes for cell cycle link protein**

NCBI database : L41578.1 (as reference), JQ820457.1 , JQ820463.1, JQ820469.1



a) DNA –C segments : encodes for cell cycle link protein

13 conserved regions found using MEGA tool from DNA –C segments region

Region 1: Position 1 to 32

conserved : 1 -AGCGCTGGGGCTTATTATTACCCCCAGCGCTC 32

Region 2: Position 34 to 51

conserved : 34- GGACGGGACATCACGTGC 51

Region 3: Position 63 to 97

conserved : 63- GCACGTGAGAATGCAGTAGCTTGCAGCGAAAGATA 97

Region 4: Position 105 to 228

conserved :105-ACATCAATAAAGAAGAAGGAATATTCTTTGCTTCGGCACGAAGCAAAGGGTATAGATA TTTGTTCGAGATGCGAAAATGGAGGCTATTTAAACCTGATGGTTTTGTGATTTCCGAAATCACTCG 228

Region 5: Position 230 to 316

conserved:230- CGGAAGAGAAATGGAGTTCTGGGAATCGTCTGCCATGCCTGACGATGTCAAGAGAGAGATTAA GGAAATATATTGGGAAGATCGGAA 316

Region 6: Position 318 to 382

conserved:318- AAACTTCTGTTCTGTCAGAAGTTGAAGAGCTATGTCAGAAGGATTCTTGTTTATGGAGATCAAGA 382

Region 7: Position 384 to 586

conserved:384- GATGCCCTTGCCGGAGTGAAGGATATGAAGACTTCTATTATTCGCTATAGCGAATACTTGAAGAAACCATGT GTGGTAATTTGTTGTGTTAGCAATAAATCAATTGTGTATAGGTTAAACAGCATGGTGTTCTTTTATCATGAATA CCTTGAAGAACTAGGTGGTGATTACTCAGTATATCAAGATCTCTATTGTGATGAGGT 586

Region 8: Position 588 to 621

conserved : 588- CTCTCTTCTTCATCGACAGAGGAAGAAGATGTAG 621

Region 9: Position 623 to 686

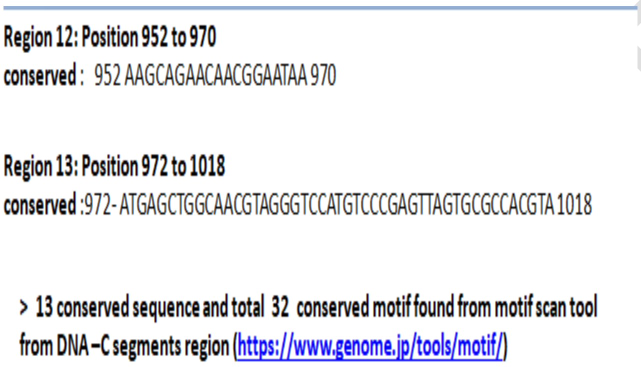
conserved:623- AGTAATATATAGGAATGTTATCATGGCATCGACACAAGAGAAGTTCTCTTGGAGTGATTGTCAG 686

Region 10: Position 688 to 872

Conserved:688- AGATAGTTATATCAGACTATGATGTAACATTACTCTAATGTAATATCCATTATCATCAATAAAATAATGGAATGTT GATTATGTATTTATCATAAATACATAATGGTATACGTATAGCATAAAATACATTAACCAACATACAACACACTATA AAATACAACACACTATAACAAATGTACGGGTAT 872

Region 11: Position 898 to 939

conserved :898- AGGGCCGAAGGCCCGTTTAAATATGTGTTGGACGAAGTCCAA 939



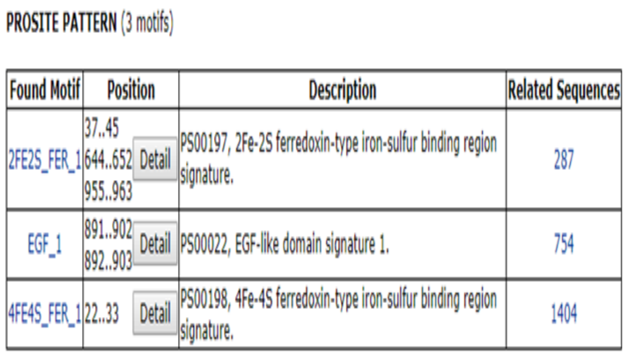
Region 12: Position 952 to 970

conserved : 952 AAGCAGAACAACGGAATAA 970

Region 13: Position 972 to 1018

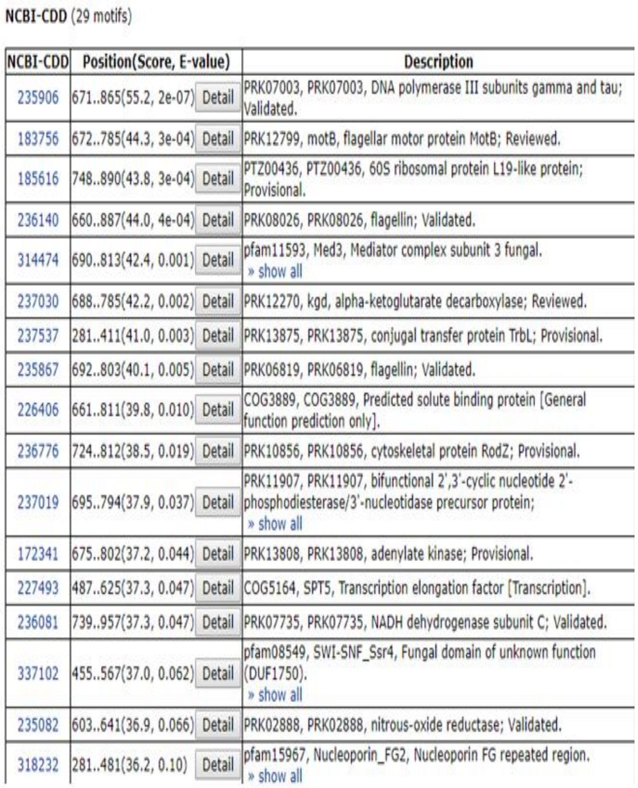
conserved :972- ATGAGCTGGCAACGTAGGGTCCATGTCCCGAGTTAGTGCGCCACGTA 1018

> 13 conserved sequence and total 32 conserved motif found from motif scan tool from DNA –C segments region (https://www.genome.jp/tools/motif/)



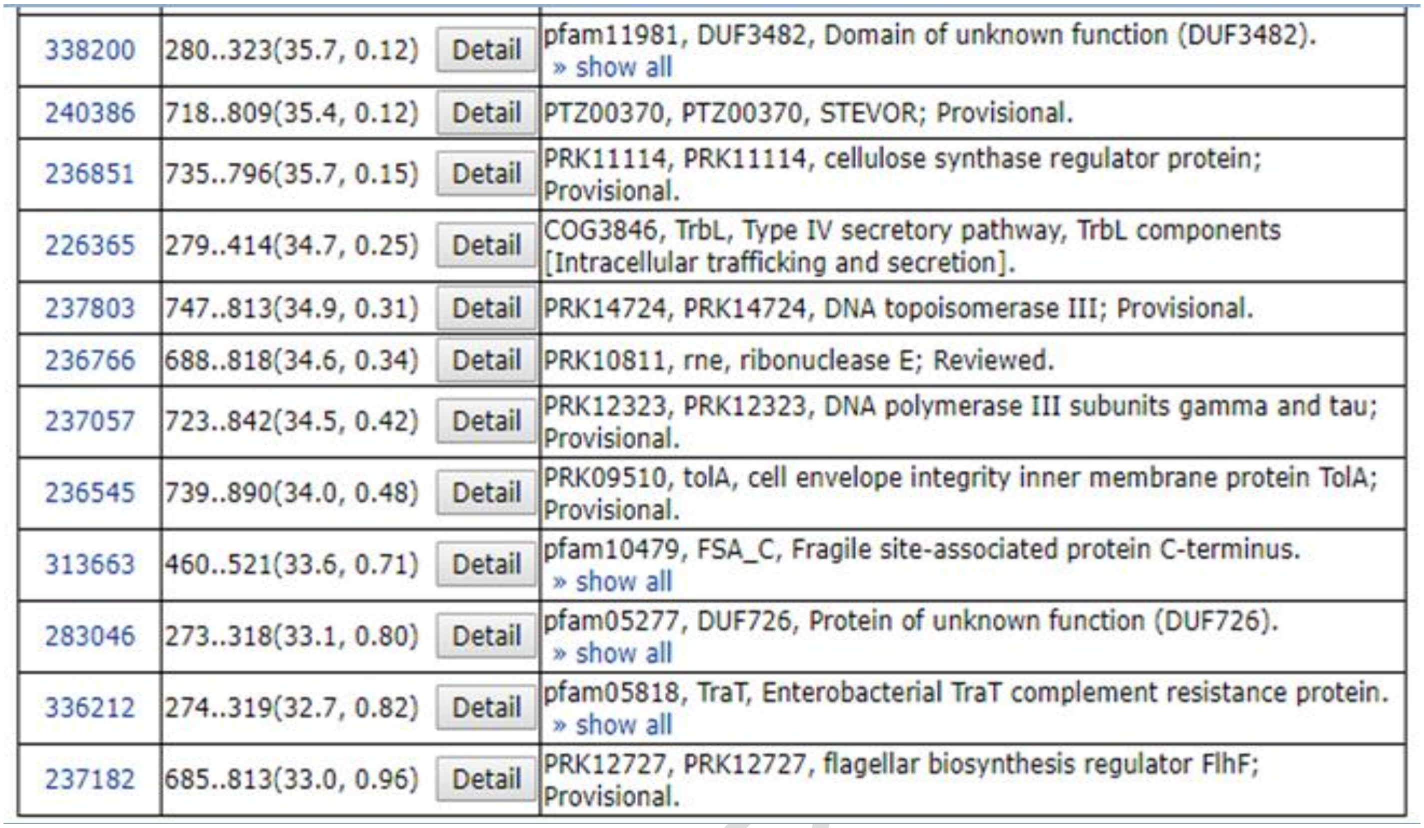
PROSITE PATTERN (3 motifs)

Found Motif	Position	Description	Related Sequences
2FE2S_FER_1	37..45 644..652 955..963 Detail	PS00197, 2Fe-2S ferredoxin-type iron-sulfur binding region signature.	287
EGF_1	891..902 892..903 Detail	PS00022, EGF-like domain signature 1.	754
4FE4S_FER_1	22..33 Detail	PS00198, 4Fe-4S ferredoxin-type iron-sulfur binding region signature.	1404



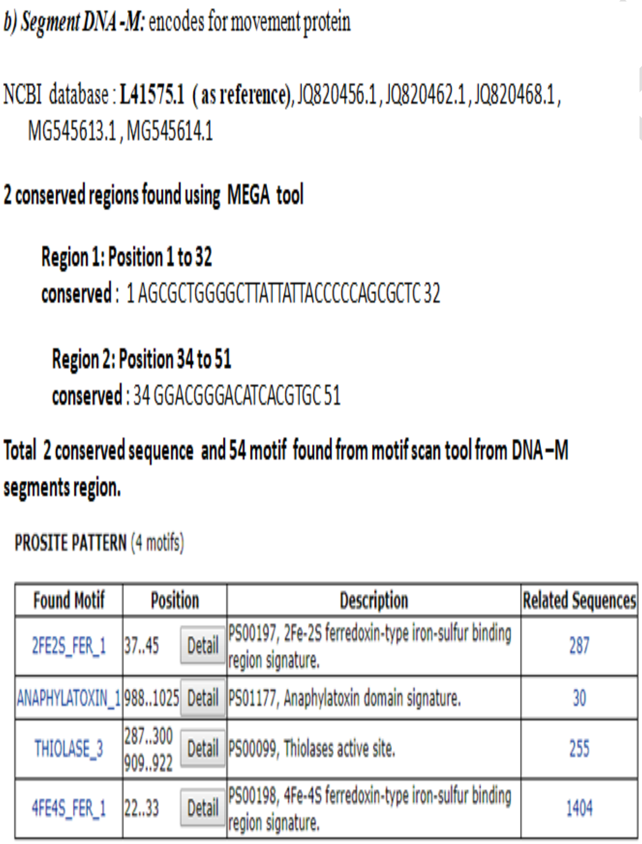
NCBI-CDD (29 motifs)

NCBI-CDD	Position(Score, E-value)	Description
235906	671..865(55.2, 2e-07) Detail	PRK07003, PRK07003, DNA polymerase III subunits gamma and tau; Validated.
183756	672..785(44.3, 3e-04) Detail	PRK12799, motB, flagellar motor protein MotB; Reviewed.
185616	748..890(43.8, 3e-04) Detail	PTZ00436, PTZ00436, 60S ribosomal protein L19-like protein; Provisional.
236140	660..887(44.0, 4e-04) Detail	PRK08026, PRK08026, flagellin; Validated.
314474	690..813(42.4, 0.001) Detail	pfam11593, Med3, Mediator complex subunit 3 fungal. » show all
237030	688..785(42.2, 0.002) Detail	PRK12270, kgd, alpha-ketoglutarate decarboxylase; Reviewed.
237537	281..411(41.0, 0.003) Detail	PRK13875, PRK13875, conjugal transfer protein TrbL; Provisional.
235867	692..803(40.1, 0.005) Detail	PRK06819, PRK06819, flagellin; Validated.
226406	661..811(39.8, 0.010) Detail	COG3889, COG3889, Predicted solute binding protein [General function prediction only].
236776	724..812(38.5, 0.019) Detail	PRK10856, PRK10856, cytoskeletal protein RodZ; Provisional.
237019	695..794(37.9, 0.037) Detail	PRK11907, PRK11907, bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase precursor protein; » show all
172341	675..802(37.2, 0.044) Detail	PRK13808, PRK13808, adenylate kinase; Provisional.
227493	487..625(37.3, 0.047) Detail	COG5164, SPT5, Transcription elongation factor [Transcription].
236081	739..957(37.3, 0.047) Detail	PRK07735, PRK07735, NADH dehydrogenase subunit C; Validated.
337102	455..567(37.0, 0.062) Detail	pfam08549, SWI-SNF_Ssr4, Fungal domain of unknown function (DUF1750). » show all
235082	603..641(36.9, 0.066) Detail	PRK02888, PRK02888, nitrous-oxide reductase; Validated.
318232	281..481(36.2, 0.10) Detail	pfam15967, Nucleoporin_FG2, Nucleoporin FG repeated region. » show all



338200	280..323(35.7, 0.12)	Detail	pfam11981, DUF3482, Domain of unknown function (DUF3482). » show all
240386	718..809(35.4, 0.12)	Detail	PTZ00370, PTZ00370, STEVOR; Provisional.
236851	735..796(35.7, 0.15)	Detail	PRK11114, PRK11114, cellulose synthase regulator protein; Provisional.
226365	279..414(34.7, 0.25)	Detail	COG3846, TrbL, Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion].
237803	747..813(34.9, 0.31)	Detail	PRK14724, PRK14724, DNA topoisomerase III; Provisional.
236766	688..818(34.6, 0.34)	Detail	PRK10811, rne, ribonuclease E; Reviewed.
237057	723..842(34.5, 0.42)	Detail	PRK12323, PRK12323, DNA polymerase III subunits gamma and tau; Provisional.
236545	739..890(34.0, 0.48)	Detail	PRK09510, tolA, cell envelope integrity inner membrane protein TolA; Provisional.
313663	460..521(33.6, 0.71)	Detail	pfam10479, FSA_C, Fragile site-associated protein C-terminus. » show all
283046	273..318(33.1, 0.80)	Detail	pfam05277, DUF726, Protein of unknown function (DUF726). » show all
336212	274..319(32.7, 0.82)	Detail	pfam05818, TraT, Enterobacterial TraT complement resistance protein. » show all
237182	685..813(33.0, 0.96)	Detail	PRK12727, PRK12727, flagellar biosynthesis regulator FlhF; Provisional.

Figure7Nucleotide variation of DNA-C genome components of BBTV



b) Segment DNA-M: encodes for movement protein

NCBI database : **L41575.1 (as reference)**, JQ820456.1 , JQ820462.1 , JQ820468.1 , MG545613.1 , MG545614.1

2 conserved regions found using MEGA tool

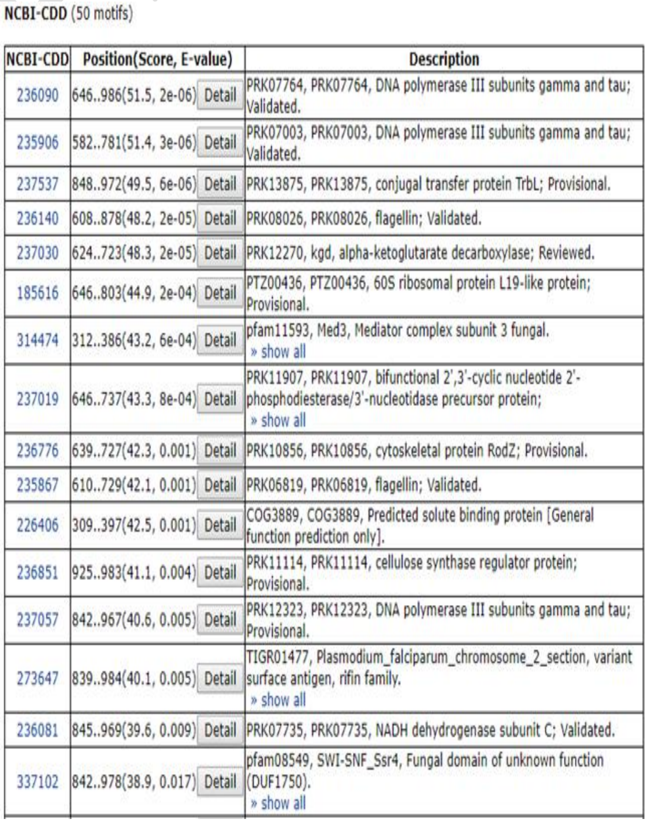
Region 1: Position 1 to 32
conserved : 1 AGCGCTGGGGCTTATTATTACCCCCAGCGCTC 32

Region 2: Position 34 to 51
conserved : 34 GGACGGGACATCACGTGC 51

Total 2 conserved sequence and 54 motif found from motif scan tool from DNA–M segments region.

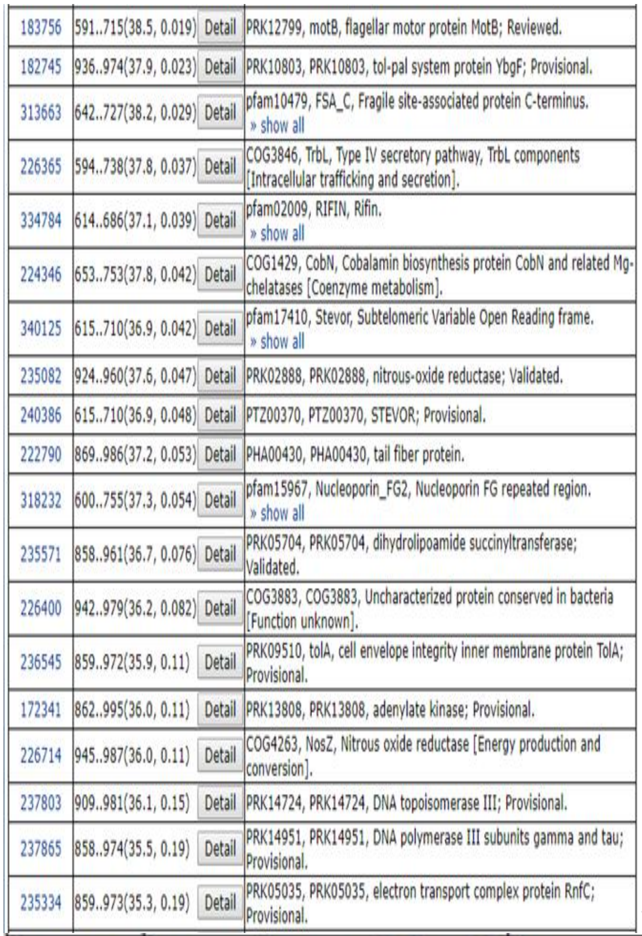
PROSITE PATTERN (4 motifs)

Found Motif	Position		Description	Related Sequences
2FE2S_FER_1	37..45	Detail	PS00197, 2Fe-2S ferredoxin-type iron-sulfur binding region signature.	287
ANAPHYLATOXIN_1	988..1025	Detail	PS01177, Anaphylatoxin domain signature.	30
THIOLASE_3	287..300 909..922	Detail	PS00099, Thiolases active site.	255
4FE4S_FER_1	22..33	Detail	PS00198, 4Fe-4S ferredoxin-type iron-sulfur binding region signature.	1404

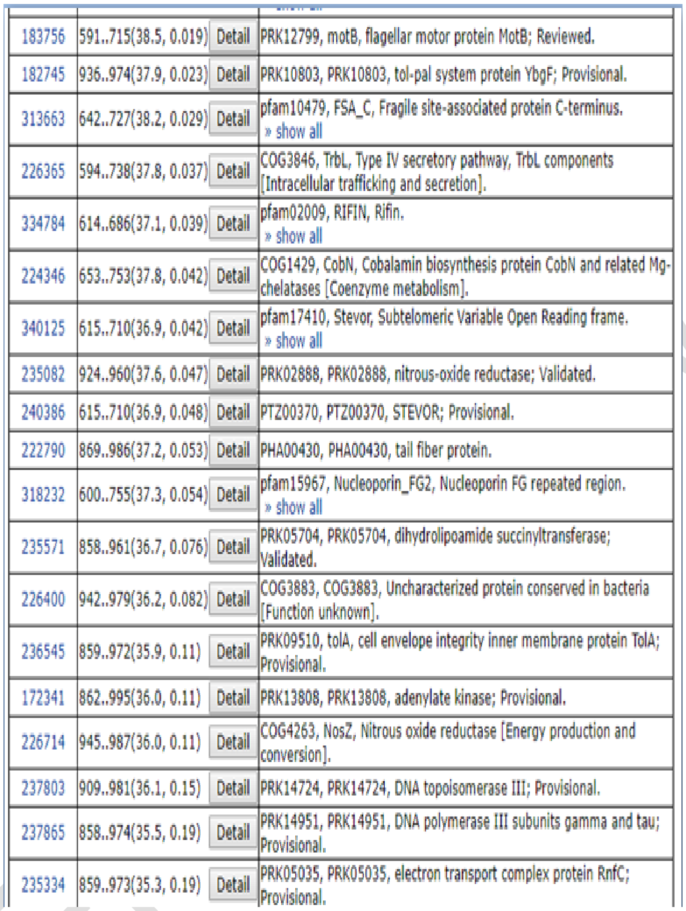


NCBI-CDD (50 motifs)

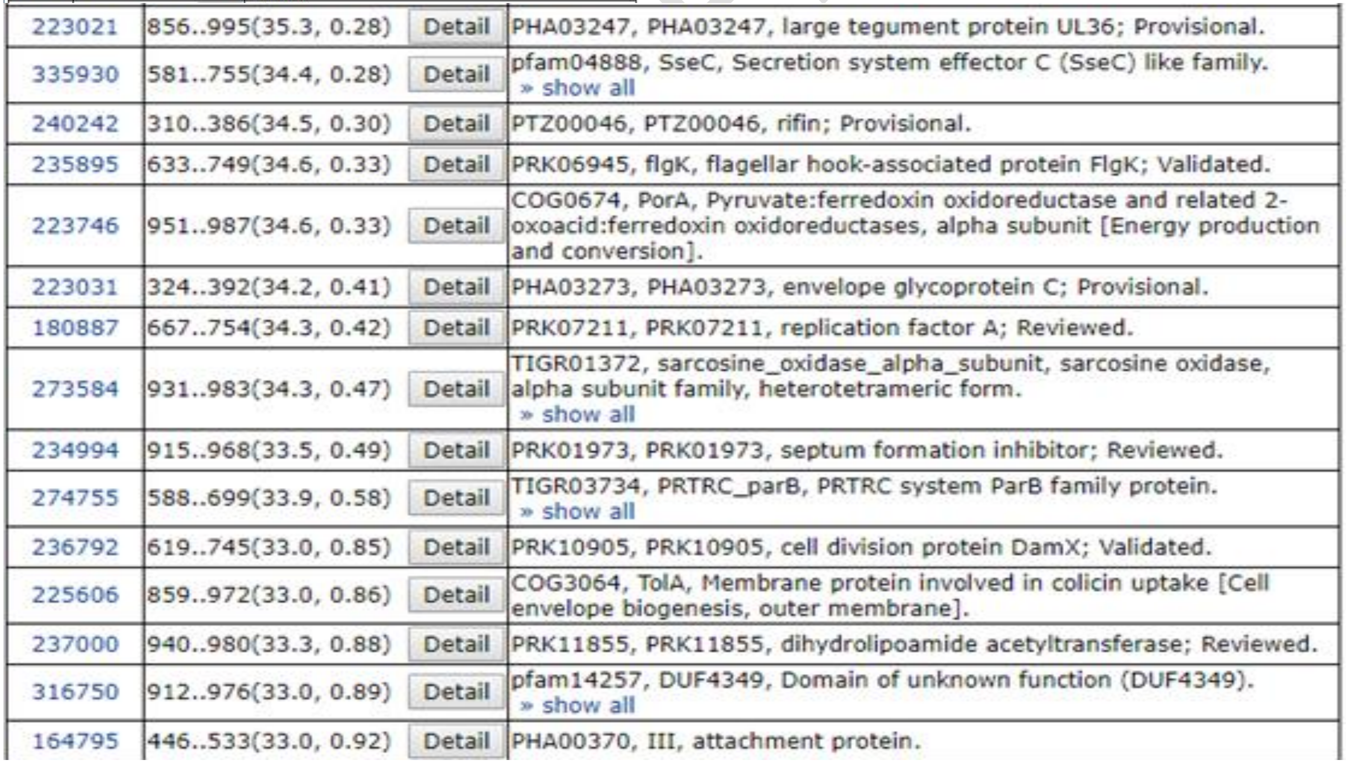
NCBI-CDD	Position(Score, E-value)		Description
236090	646..986(51.5, 2e-06)	Detail	PRK07764, PRK07764, DNA polymerase III subunits gamma and tau; Validated.
235906	582..781(51.4, 3e-06)	Detail	PRK07003, PRK07003, DNA polymerase III subunits gamma and tau; Validated.
237537	848..972(49.5, 6e-06)	Detail	PRK13875, PRK13875, conjugal transfer protein TrbL; Provisional.
236140	608..878(48.2, 2e-05)	Detail	PRK08026, PRK08026, flagellin; Validated.
237030	624..723(48.3, 2e-05)	Detail	PRK12270, kgd, alpha-ketoglutarate decarboxylase; Reviewed.
185616	646..803(44.9, 2e-04)	Detail	PTZ00436, PTZ00436, 60S ribosomal protein L19-like protein; Provisional.
314474	312..386(43.2, 6e-04)	Detail	pfam11593, Med3, Mediator complex subunit 3 fungal. » show all
237019	646..737(43.3, 8e-04)	Detail	PRK11907, PRK11907, bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase precursor protein; » show all
236776	639..727(42.3, 0.001)	Detail	PRK10856, PRK10856, cytoskeletal protein RodZ; Provisional.
235867	610..729(42.1, 0.001)	Detail	PRK06819, PRK06819, flagellin; Validated.
226406	309..397(42.5, 0.001)	Detail	COG3889, COG3889, Predicted solute binding protein [General function prediction only].
236851	925..983(41.1, 0.004)	Detail	PRK11114, PRK11114, cellulose synthase regulator protein; Provisional.
237057	842..967(40.6, 0.005)	Detail	PRK12323, PRK12323, DNA polymerase III subunits gamma and tau; Provisional.
273647	839..984(40.1, 0.005)	Detail	TIGR01477, Plasmodium_falciparum_chromosome_2_section, variant surface antigen, rifin family. » show all
236081	845..969(39.6, 0.009)	Detail	PRK07735, PRK07735, NADH dehydrogenase subunit C; Validated.
337102	842..978(38.9, 0.017)	Detail	pfam08549, SWI-SNF_Ssr4, Fungal domain of unknown function (DUF1750). » show all



183756	591..715(38.5, 0.019)	Detail	PRK12799, motB, flagellar motor protein MotB; Reviewed.
182745	936..974(37.9, 0.023)	Detail	PRK10803, PRK10803, tol-pal system protein YbgF; Provisional.
313663	642..727(38.2, 0.029)	Detail	pfam10479, FSA_C, Fragile site-associated protein C-terminus. » show all
226365	594..738(37.8, 0.037)	Detail	COG3846, TrbL, Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion].
334784	614..686(37.1, 0.039)	Detail	pfam02009, RIFIN, Rifin. » show all
224346	653..753(37.8, 0.042)	Detail	COG1429, CobN, Cobalamin biosynthesis protein CobN and related Mg-chelatases [Coenzyme metabolism].
340125	615..710(36.9, 0.042)	Detail	pfam17410, Stevor, Subtelomeric Variable Open Reading frame. » show all
235082	924..960(37.6, 0.047)	Detail	PRK02888, PRK02888, nitrous-oxide reductase; Validated.
240386	615..710(36.9, 0.048)	Detail	PTZ00370, PTZ00370, STEVOR; Provisional.
222790	869..986(37.2, 0.053)	Detail	PHA00430, PHA00430, tail fiber protein.
318232	600..755(37.3, 0.054)	Detail	pfam15967, Nucleoporin_FG2, Nucleoporin FG repeated region. » show all
235571	858..961(36.7, 0.076)	Detail	PRK05704, PRK05704, dihydrolipoamide succinyltransferase; Validated.
226400	942..979(36.2, 0.082)	Detail	COG3883, COG3883, Uncharacterized protein conserved in bacteria [Function unknown].
236545	859..972(35.9, 0.11)	Detail	PRK09510, tolA, cell envelope integrity inner membrane protein TolA; Provisional.
172341	862..995(36.0, 0.11)	Detail	PRK13808, PRK13808, adenylate kinase; Provisional.
226714	945..987(36.0, 0.11)	Detail	COG4263, NosZ, Nitrous oxide reductase [Energy production and conversion].
237803	909..981(36.1, 0.15)	Detail	PRK14724, PRK14724, DNA topoisomerase III; Provisional.
237865	858..974(35.5, 0.19)	Detail	PRK14951, PRK14951, DNA polymerase III subunits gamma and tau; Provisional.
235334	859..973(35.3, 0.19)	Detail	PRK05035, PRK05035, electron transport complex protein RnfC; Provisional.

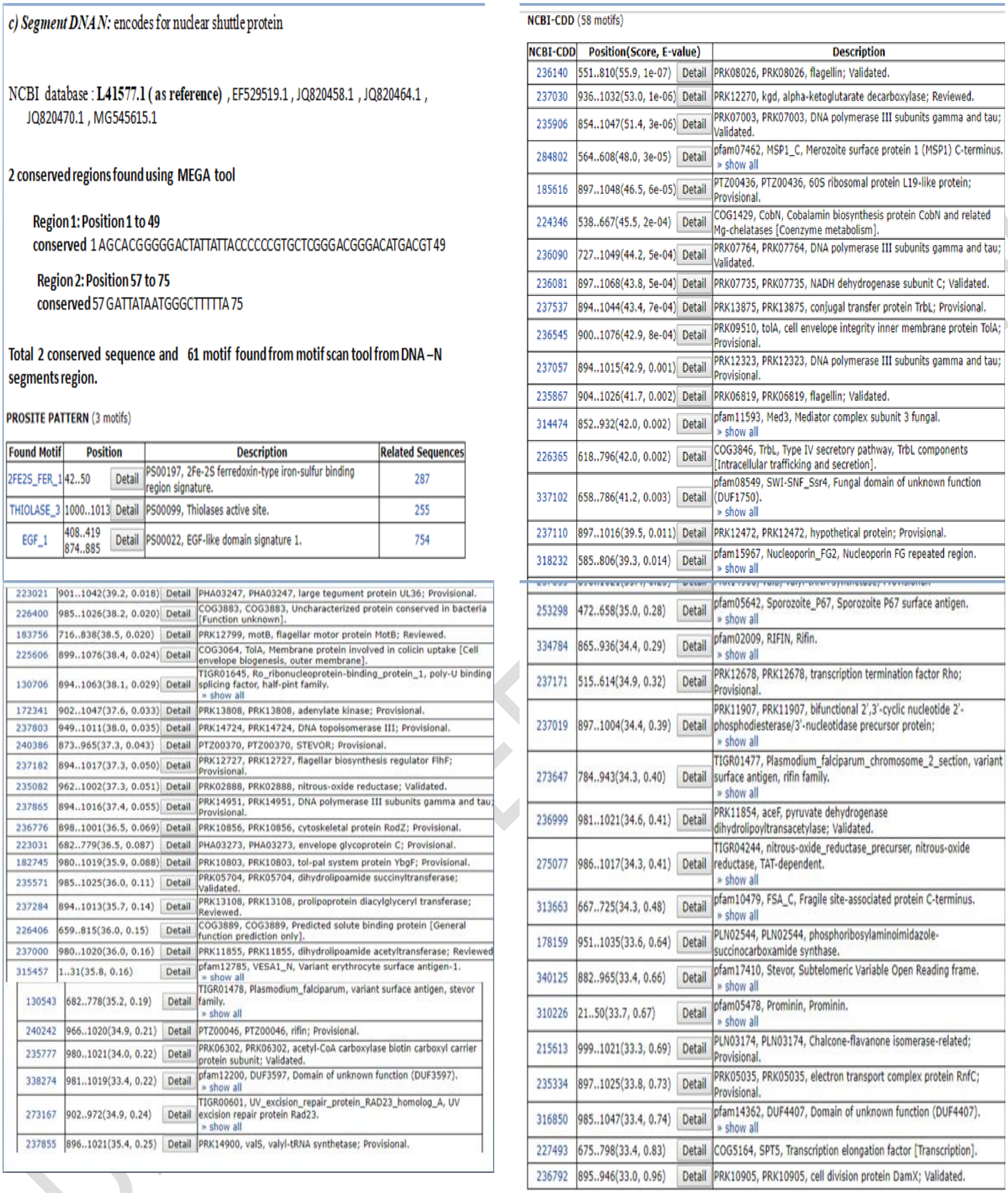


183756	591..715(38.5, 0.019)	Detail	PRK12799, motB, flagellar motor protein MotB; Reviewed.
182745	936..974(37.9, 0.023)	Detail	PRK10803, PRK10803, tol-pal system protein YbgF; Provisional.
313663	642..727(38.2, 0.029)	Detail	pfam10479, FSA_C, Fragile site-associated protein C-terminus. » show all
226365	594..738(37.8, 0.037)	Detail	COG3846, TrbL, Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion].
334784	614..686(37.1, 0.039)	Detail	pfam02009, RIFIN, Rifin. » show all
224346	653..753(37.8, 0.042)	Detail	COG1429, CobN, Cobalamin biosynthesis protein CobN and related Mg-chelatases [Coenzyme metabolism].
340125	615..710(36.9, 0.042)	Detail	pfam17410, Stevor, Subtelomeric Variable Open Reading frame. » show all
235082	924..960(37.6, 0.047)	Detail	PRK02888, PRK02888, nitrous-oxide reductase; Validated.
240386	615..710(36.9, 0.048)	Detail	PTZ00370, PTZ00370, STEVOR; Provisional.
222790	869..986(37.2, 0.053)	Detail	PHA00430, PHA00430, tail fiber protein.
318232	600..755(37.3, 0.054)	Detail	pfam15967, Nucleoporin_FG2, Nucleoporin FG repeated region. » show all
235571	858..961(36.7, 0.076)	Detail	PRK05704, PRK05704, dihydrolipoamide succinyltransferase; Validated.
226400	942..979(36.2, 0.082)	Detail	COG3883, COG3883, Uncharacterized protein conserved in bacteria [Function unknown].
236545	859..972(35.9, 0.11)	Detail	PRK09510, tolA, cell envelope integrity inner membrane protein TolA; Provisional.
172341	862..995(36.0, 0.11)	Detail	PRK13808, PRK13808, adenylate kinase; Provisional.
226714	945..987(36.0, 0.11)	Detail	COG4263, NosZ, Nitrous oxide reductase [Energy production and conversion].
237803	909..981(36.1, 0.15)	Detail	PRK14724, PRK14724, DNA topoisomerase III; Provisional.
237865	858..974(35.5, 0.19)	Detail	PRK14951, PRK14951, DNA polymerase III subunits gamma and tau; Provisional.
235334	859..973(35.3, 0.19)	Detail	PRK05035, PRK05035, electron transport complex protein RnfC; Provisional.



223021	856..995(35.3, 0.28)	Detail	PHA03247, PHA03247, large tegument protein UL36; Provisional.
335930	581..755(34.4, 0.28)	Detail	pfam04888, SseC, Secretion system effector C (SseC) like family. » show all
240242	310..386(34.5, 0.30)	Detail	PTZ00046, PTZ00046, rifin; Provisional.
235895	633..749(34.6, 0.33)	Detail	PRK06945, flgK, flagellar hook-associated protein FlgK; Validated.
223746	951..987(34.6, 0.33)	Detail	COG0674, PorA, Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, alpha subunit [Energy production and conversion].
223031	324..392(34.2, 0.41)	Detail	PHA03273, PHA03273, envelope glycoprotein C; Provisional.
180887	667..754(34.3, 0.42)	Detail	PRK07211, PRK07211, replication factor A; Reviewed.
273584	931..983(34.3, 0.47)	Detail	TIGR01372, sarcosine_oxidase_alpha_subunit, sarcosine oxidase, alpha subunit family, heterotetrameric form. » show all
234994	915..968(33.5, 0.49)	Detail	PRK01973, PRK01973, septum formation inhibitor; Reviewed.
274755	588..699(33.9, 0.58)	Detail	TIGR03734, PRTRC_parB, PRTRC system ParB family protein. » show all
236792	619..745(33.0, 0.85)	Detail	PRK10905, PRK10905, cell division protein DamX; Validated.
225606	859..972(33.0, 0.86)	Detail	COG3064, TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].
237000	940..980(33.3, 0.88)	Detail	PRK11855, PRK11855, dihydrolipoamide acetyltransferase; Reviewed.
316750	912..976(33.0, 0.89)	Detail	pfam14257, DUF4349, Domain of unknown function (DUF4349). » show all
164795	446..533(33.0, 0.92)	Detail	PHA00370, III, attachment protein.

Figure8Nucleotide variation of DNA-M genome components of BBTV



c) Segment DNA N: encodes for nuclear shuttle protein

NCBI database : **L41577.1 (as reference)** , EF529519.1 , JQ820458.1 , JQ820464.1 , JQ820470.1 , MG545615.1

2 conserved regions found using MEGA tool

Region 1: Position 1 to 49
conserved 1 AGCACGGGGGACTATTATTACCCCCCGTGCTCGGGACGGGACATGACGT 49

Region 2: Position 57 to 75
conserved 57 GATTATAATGGGCTTTTTA 75

Total 2 conserved sequence and 61 motif found from motif scan tool from DNA –N segments region.

PROSITE PATTERN (3 motifs)

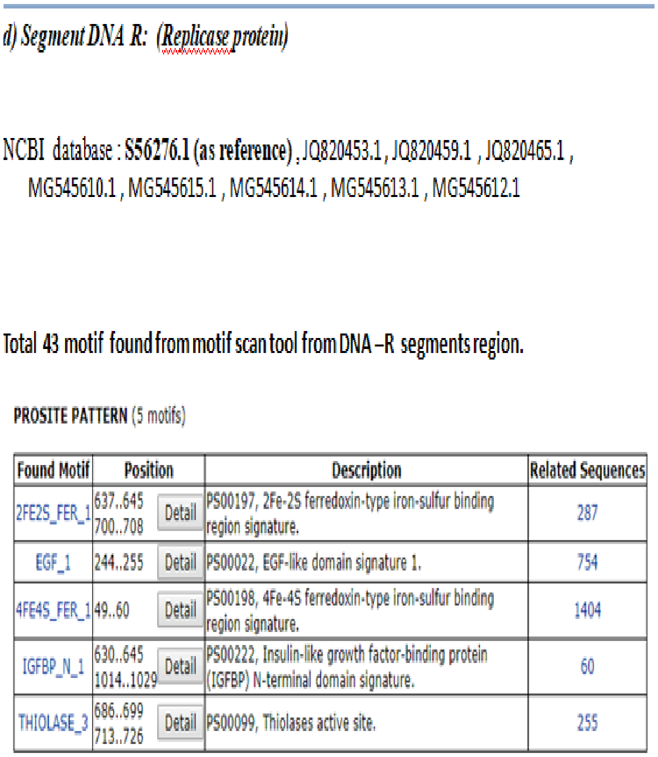
Found Motif	Position		Description	Related Sequences
2FE2S_FER_1	42..50	Detail	PS00197, 2Fe-2S ferredoxin-type iron-sulfur binding region signature.	287
THIOLASE_3	1000..1013	Detail	PS00099, Thiolases active site.	255
EGF_1	408..419 874..885	Detail	PS00022, EGF-like domain signature 1.	754

NCBI-CDD	Position(Score, E-value)		Description
223021	901..1042(39.2, 0.018)	Detail	PHA03247, PHA03247, large tegument protein UL36; Provisional.
226400	985..1026(38.2, 0.020)	Detail	COG3883, COG3883, Uncharacterized protein conserved in bacteria [Function unknown].
183756	716..838(38.5, 0.020)	Detail	PRK12799, motB, flagellar motor protein MotB; Reviewed.
225606	899..1076(38.4, 0.024)	Detail	COG3064, TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].
130706	894..1063(38.1, 0.029)	Detail	TIGR01645, Ro_ribonucleoprotein-binding_protein_1, poly-U binding splicing factor, half-pint family. » show all
172341	902..1047(37.6, 0.033)	Detail	PRK13808, PRK13808, adenylate kinase; Provisional.
237803	949..1011(38.0, 0.035)	Detail	PRK14724, PRK14724, DNA topoisomerase III; Provisional.
240386	873..965(37.3, 0.043)	Detail	PTZ00370, PTZ00370, STEVOR; Provisional.
237182	894..1017(37.3, 0.050)	Detail	PRK12727, PRK12727, flagellar biosynthesis regulator FlhF; Provisional.
235082	962..1002(37.3, 0.051)	Detail	PRK02888, PRK02888, nitrous-oxide reductase; Validated.
237865	894..1016(37.4, 0.055)	Detail	PRK14951, PRK14951, DNA polymerase III subunits gamma and tau; Provisional.
236776	898..1001(36.5, 0.069)	Detail	PRK10856, PRK10856, cytoskeletal protein RodZ; Provisional.
223031	682..779(36.5, 0.087)	Detail	PHA03273, PHA03273, envelope glycoprotein C; Provisional.
182745	980..1019(35.9, 0.088)	Detail	PRK10803, PRK10803, tol-pal system protein YbgF; Provisional.
235571	985..1025(36.0, 0.11)	Detail	PRK05704, PRK05704, dihydrolipoamide succinyltransferase; Validated.
237284	894..1013(35.7, 0.14)	Detail	PRK13108, PRK13108, prolipoprotein diacylglyceryl transferase; Reviewed.
226406	659..815(36.0, 0.15)	Detail	COG3889, COG3889, Predicted solute binding protein [General function prediction only].
237000	980..1020(36.0, 0.16)	Detail	PRK11855, PRK11855, dihydrolipoamide acetyltransferase; Reviewed
315457	1..31(35.8, 0.16)	Detail	pfam12785, VESA1_N, Variant erythrocyte surface antigen-1. » show all
130543	682..778(35.2, 0.19)	Detail	TIGR01478, Plasmodium_falciparum, variant surface antigen, stevor family. » show all
240242	966..1020(34.9, 0.21)	Detail	PTZ00046, PTZ00046, rifin; Provisional.
235777	980..1021(34.0, 0.22)	Detail	PRK06302, PRK06302, acetyl-CoA carboxylase biotin carboxyl carrier protein subunit; Validated.
338274	981..1019(33.4, 0.22)	Detail	pfam12200, DUF3597, Domain of unknown function (DUF3597). » show all
273167	902..972(34.9, 0.24)	Detail	TIGR00601, UV_excision_repair_protein_RAD23_homolog_A, UV excision repair protein Rad23. » show all
237855	896..1021(35.4, 0.25)	Detail	PRK14900, valS, valyl-tRNA synthetase; Provisional.

NCBI-CDD (58 motifs)

NCBI-CDD	Position(Score, E-value)		Description
236140	551..810(55.9, 1e-07)	Detail	PRK08026, PRK08026, flagellin; Validated.
237030	936..1032(53.0, 1e-06)	Detail	PRK12270, kgd, alpha-ketoglutarate decarboxylase; Reviewed.
235906	854..1047(51.4, 3e-06)	Detail	PRK07003, PRK07003, DNA polymerase III subunits gamma and tau; Validated.
284802	564..608(48.0, 3e-05)	Detail	pfam07462, MSP1_C, Merozoite surface protein 1 (MSP1) C-terminus. » show all
185616	897..1048(46.5, 6e-05)	Detail	PTZ00436, PTZ00436, 60S ribosomal protein L19-like protein; Provisional.
224346	538..667(45.5, 2e-04)	Detail	COG1429, CobN, Cobalamin biosynthesis protein CobN and related Mg-chelatases [Coenzyme metabolism].
236090	727..1049(44.2, 5e-04)	Detail	PRK07764, PRK07764, DNA polymerase III subunits gamma and tau; Validated.
236081	897..1068(43.8, 5e-04)	Detail	PRK07735, PRK07735, NADH dehydrogenase subunit C; Validated.
237537	894..1044(43.4, 7e-04)	Detail	PRK13875, PRK13875, conjugal transfer protein TrbL; Provisional.
236545	900..1076(42.9, 8e-04)	Detail	PRK09510, tolA, cell envelope integrity inner membrane protein TolA; Provisional.
237057	894..1015(42.9, 0.001)	Detail	PRK12323, PRK12323, DNA polymerase III subunits gamma and tau; Provisional.
235867	904..1026(41.7, 0.002)	Detail	PRK06819, PRK06819, flagellin; Validated.
314474	852..932(42.0, 0.002)	Detail	pfam11593, Med3, Mediator complex subunit 3 fungal. » show all
226365	618..796(42.0, 0.002)	Detail	COG3846, TrbL, Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion].
337102	658..786(41.2, 0.003)	Detail	pfam08549, SWI-SNF_Ssr4, Fungal domain of unknown function (DUF1750). » show all
237110	897..1016(39.5, 0.011)	Detail	PRK12472, PRK12472, hypothetical protein; Provisional.
318232	585..806(39.3, 0.014)	Detail	pfam15967, Nucleoporin_FG2, Nucleoporin FG repeated region. » show all
253298	472..658(35.0, 0.28)	Detail	pfam05642, Sporozoite_P67, Sporozoite P67 surface antigen. » show all
334784	865..936(34.4, 0.29)	Detail	pfam02009, RIFIN, Rifin. » show all
237171	515..614(34.9, 0.32)	Detail	PRK12678, PRK12678, transcription termination factor Rho; Provisional.
237019	897..1004(34.4, 0.39)	Detail	PRK11907, PRK11907, bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase precursor protein; » show all
273647	784..943(34.3, 0.40)	Detail	TIGR01477, Plasmodium_falciparum_chromosome_2_section, variant surface antigen, rifin family. » show all
236999	981..1021(34.6, 0.41)	Detail	PRK11854, aceF, pyruvate dehydrogenase dihydrolipoyltransacetylase; Validated.
275077	986..1017(34.3, 0.41)	Detail	TIGR04244, nitrous-oxide_reductase_precurser, nitrous-oxide reductase, TAT-dependent. » show all
313663	667..725(34.3, 0.48)	Detail	pfam10479, FSA_C, Fragile site-associated protein C-terminus. » show all
178159	951..1035(33.6, 0.64)	Detail	PLN02544, PLN02544, phosphoribosylaminoimidazole-succinocarboxamide synthase.
340125	882..965(33.4, 0.66)	Detail	pfam17410, Stevor, Subtelomeric Variable Open Reading frame. » show all
310226	21..50(33.7, 0.67)	Detail	pfam05478, Prominin, Prominin. » show all
215613	999..1021(33.3, 0.69)	Detail	PLN03174, PLN03174, Chalcone-flavanone isomerase-related; Provisional.
235334	897..1025(33.8, 0.73)	Detail	PRK05035, PRK05035, electron transport complex protein RnfC; Provisional.
316850	985..1047(33.4, 0.74)	Detail	pfam14362, DUF4407, Domain of unknown function (DUF4407). » show all
227493	675..798(33.4, 0.83)	Detail	COG5164, SPT5, Transcription elongation factor [Transcription].
236792	895..946(33.0, 0.96)	Detail	PRK10905, PRK10905, cell division protein DamX; Validated.

Figure9Nucleotide variation of DNA-N genome components of BBTV



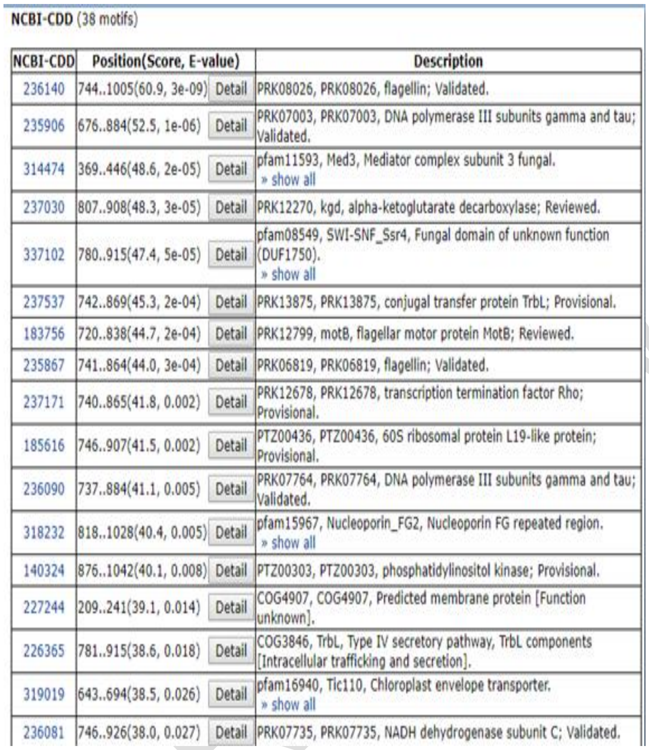
d) Segment DNA R: (Replicase protein)

NCBI database : **S56276.1 (as reference)** , JQ820453.1 , JQ820459.1 , JQ820465.1 , MG545610.1 , MG545615.1 , MG545614.1 , MG545613.1 , MG545612.1

Total 43 motif found from motif scan tool from DNA –R segments region.

PROSITE PATTERN (5 motifs)

Found Motif	Position		Description	Related Sequences
2FE2S_FER_1	637..645 700..708	Detail	PS00197, 2Fe-2S ferredoxin-type iron-sulfur binding region signature.	287
EGF_1	244..255	Detail	PS00022, EGF-like domain signature 1.	754
4FE4S_FER_1	49..60	Detail	PS00198, 4Fe-4S ferredoxin-type iron-sulfur binding region signature.	1404
IGFBP_N_1	630..645 1014..1029	Detail	PS00222, Insulin-like growth factor-binding protein (IGFBP) N-terminal domain signature.	60
THIOLASE_3	686..699 713..726	Detail	PS00099, Thiolases active site.	255



NCBI-CDD (38 motifs)

NCBI-CDD	Position(Score, E-value)		Description
236140	744..1005(60.9, 3e-09)	Detail	PRK08026, PRK08026, flagellin; Validated.
235906	676..884(52.5, 1e-06)	Detail	PRK07003, PRK07003, DNA polymerase III subunits gamma and tau; Validated.
314474	369..446(48.6, 2e-05)	Detail	pfam11593, Med3, Mediator complex subunit 3 fungal. » show all
237030	807..908(48.3, 3e-05)	Detail	PRK12270, kgd, alpha-ketoglutarate decarboxylase; Reviewed.
337102	780..915(47.4, 5e-05)	Detail	pfam08549, SWI-SNF_Ssr4, Fungal domain of unknown function (DUF1750). » show all
237537	742..869(45.3, 2e-04)	Detail	PRK13875, PRK13875, conjugal transfer protein TrbL; Provisional.
183756	720..838(44.7, 2e-04)	Detail	PRK12799, motB, flagellar motor protein MotB; Reviewed.
235867	741..864(44.0, 3e-04)	Detail	PRK06819, PRK06819, flagellin; Validated.
237171	740..865(41.8, 0.002)	Detail	PRK12678, PRK12678, transcription termination factor Rho; Provisional.
185616	746..907(41.5, 0.002)	Detail	PTZ00436, PTZ00436, 60S ribosomal protein L19-like protein; Provisional.
236090	737..884(41.1, 0.005)	Detail	PRK07764, PRK07764, DNA polymerase III subunits gamma and tau; Validated.
318232	818..1028(40.4, 0.005)	Detail	pfam15967, Nucleoporin_FG2, Nucleoporin FG repeated region. » show all
140324	876..1042(40.1, 0.008)	Detail	PTZ00303, PTZ00303, phosphatidylinositol kinase; Provisional.
227244	209..241(39.1, 0.014)	Detail	COG4907, COG4907, Predicted membrane protein [Function unknown].
226365	781..915(38.6, 0.018)	Detail	COG3846, TrbL, Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion].
319019	643..694(38.5, 0.026)	Detail	pfam16940, Tic110, Chloroplast envelope transporter. » show all
236081	746..926(38.0, 0.027)	Detail	PRK07735, PRK07735, NADH dehydrogenase subunit C; Validated.

313663	402..468(38.2, 0.030)	Detail	pfam10479, FSA_C, Fragile site-associated protein C-terminus. » show all
237019	746..833(37.9, 0.035)	Detail	PRK11907, PRK11907, bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase precursor protein; » show all
253298	259..441(37.7, 0.041)	Detail	pfam05642, Sporozoite_P67, Sporozoite P67 surface antigen. » show all
315457	54..102(36.9, 0.070)	Detail	pfam12785, VESA1_N, Variant erythrocyte surface antigen-1. » show all
215598	176..241(36.8, 0.087)	Detail	PLN03138, PLN03138, Protein TOC75; Provisional.
184603	596..649(36.3, 0.097)	Detail	PRK14282, PRK14282, chaperone protein DnaJ; Provisional.
182745	188..226(35.9, 0.10)	Detail	PRK10803, PRK10803, tol-pal system protein YbgF; Provisional.
172341	774..907(36.0, 0.11)	Detail	PRK13808, PRK13808, adenylate kinase; Provisional.
275181	650..845(36.0, 0.16)	Detail	TIGR04388, hypothetical_protein_LinasL1_02240, putative large structural protein. » show all
226406	405..454(36.0, 0.16)	Detail	COG3889, COG3889, Predicted solute binding protein [General function prediction only].
235082	551..588(35.7, 0.16)	Detail	PRK02888, PRK02888, nitrous-oxide reductase; Validated.
336212	188..230(34.6, 0.22)	Detail	pfam05818, TraT, Enterobacterial TraT complement resistance protein. » show all
236776	755..855(35.0, 0.24)	Detail	PRK10856, PRK10856, cytoskeletal protein RodZ; Provisional.
338200	188..240(34.1, 0.35)	Detail	pfam11981, DUF3482, Domain of unknown function (DUF3482). » show all
222890	769..930(34.7, 0.39)	Detail	PHA02584, 34, long tail fiber, proximal subunit; Provisional.
334784	387..459(34.0, 0.43)	Detail	pfam02009, RIFIN, Rifin. » show all
240386	356..439(33.9, 0.49)	Detail	PTZ00370, PTZ00370, STEVOR; Provisional.
184612	595..649(33.3, 0.80)	Detail	PRK14298, PRK14298, chaperone protein DnaJ; Provisional.
273167	769..855(33.3, 0.84)	Detail	TIGR00601, UV_excision_repair_protein_RAD23_homolog_A, UV excision repair protein Rad23. » show all
130543	356..450(32.9, 0.92)	Detail	TIGR01478, Plasmodium_falciparum, variant surface antigen, stevor family. » show all
188700	92..101(32.3, 0.94)	Detail	cd08746, RGS_RGS20, Regulator of G protein signaling (RGS) domain found in the RGS20 protein. » show all

Figure10Nucleotide variation of DNA-R genome components of BBTV

e) Segment DNA S: (Coat protein)

NCBI database : L41574.1 (**as reference**), JQ820455.1 , JQ820461.1 , JQ820467.1 , MG545612.1

6 conserved regions found using MEGA tool

Region 1: Position 13 to 56
conserved : 13 TATTATTACCCCCAGCGCTCGGGACGGGACATGGGCTAATGGAT 56

Region 2: Position 59 to 73
conserved : 59 TGGATATAGGGCCCA 73

Region 3: Position 86 to 101
conserved 86 AGATGGGTTTTGGGCT 101

Region 4: Position 112 to 144
conserved 112 ATCCAGAAGACCAAAAACAGGCGGGAACCGTCC 144

Region 5: Position 159 to 175
conserved : 159 GATTGCTTGCCCTGCAA 175

Region 6: Position 178 to 193
conserved : 178 CATCTAGAAGTCTATA 193

Total 6 conserved sequence and 49 motif found from motif scan tool from DNA –S segments region.

PROSITE PATTERN (5 motifs)

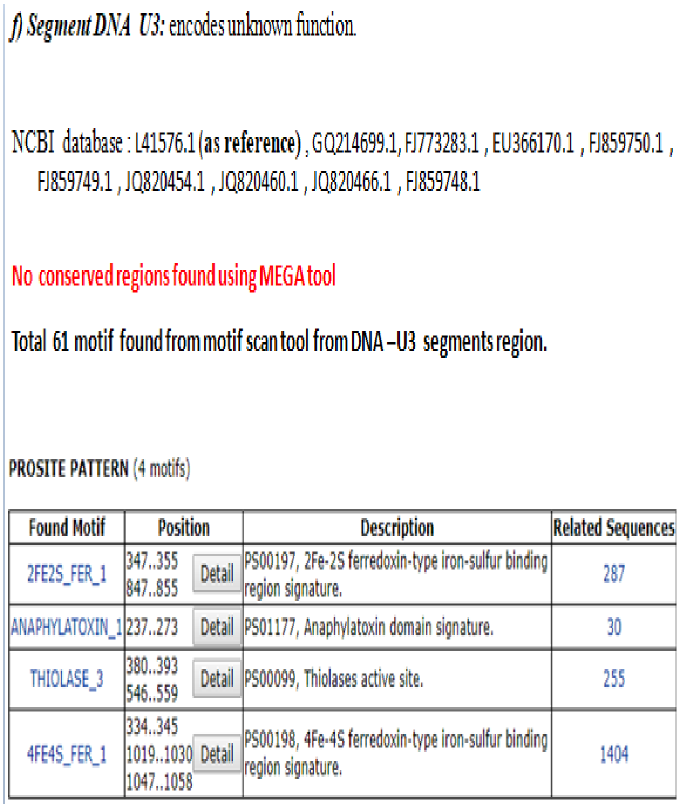
Found Motif	Position	Description	Related Sequences
4FE4S_FER_1	22..33 Detail	PS00198, 4Fe-4S ferredoxin-type iron-sulfur binding region signature.	1404
CTCK_1	275..311 Detail	PS01185, C-terminal cystine knot signature.	48
2FE2S_FER_1	290..298 Detail	PS00197, 2Fe-2S ferredoxin-type iron-sulfur binding region signature.	287
THIOLASE_3	117..130 941..954 Detail	PS00099, Thiolases active site.	255
EGF_1	590..601 Detail	PS00022, EGF-like domain signature 1.	754

NCBI-CDD (44 motifs)

NCBI-CDD	Position(Score, E-value)	Description
237537	876..1004(53.0, 6e-07) Detail	PRK13875, PRK13875, conjugal transfer protein TrbL; Provisional.
235906	756..993(51.8, 2e-06) Detail	PRK07003, PRK07003, DNA polymerase III subunits gamma and tau; Validated.
236140	646..910(50.1, 6e-06) Detail	PRK08026, PRK08026, flagellin; Validated.
237030	924..1001(48.3, 3e-05) Detail	PRK12270, kgd, alpha-ketoglutarate decarboxylase; Reviewed.
185616	866..1018(45.3, 1e-04) Detail	PTZ00436, PTZ00436, 60S ribosomal protein L19-like protein; Provisional.
314474	721..791(45.1, 2e-04) Detail	pfam11593, Med3, Mediator complex subunit 3 fungal. » show all
226406	640..777(43.7, 6e-04) Detail	COG3889, COG3889, Predicted solute binding protein [General function prediction only].
236081	877..1000(41.9, 0.002) Detail	PRK07735, PRK07735, NADH dehydrogenase subunit C; Validated.
237057	874..999(42.2, 0.002) Detail	PRK12323, PRK12323, DNA polymerase III subunits gamma and tau; Provisional.
235867	651..779(40.9, 0.003) Detail	PRK06819, PRK06819, flagellin; Validated.
253298	547..705(41.2, 0.003) Detail	pfam05642, Sporozoite_P67, Sporozoite P67 surface antigen. » show all
236090	686..1018(40.7, 0.005) Detail	PRK07764, PRK07764, DNA polymerase III subunits gamma and tau; Validated.
182745	968..1006(38.2, 0.015) Detail	PRK10803, PRK10803, tol-pal system protein YbgF; Provisional.
172341	909..1018(38.3, 0.018) Detail	PRK13808, PRK13808, adenylate kinase; Provisional.
222790	901..1007(38.3, 0.026) Detail	PHA00430, PHA00430, tail fiber protein.
236851	957..1011(38.4, 0.026) Detail	PRK11114, PRK11114, cellulose synthase regulator protein; Provisional.
183756	641..765(37.8, 0.033) Detail	PRK12799, motB, flagellar motor protein MotB; Reviewed.
235571	890..993(37.9, 0.033) Detail	PRK05704, PRK05704, dihydrolipoamide succinyltransferase; Validated.

226365	855..1018(37.8, 0.038) Detail	COG3846, TrbL, Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion].
235082	956..992(37.6, 0.040) Detail	PRK02888, PRK02888, nitrous-oxide reductase; Validated.
273647	871..1010(37.0, 0.051) Detail	TIGR01477, Plasmodium_falciparum_chromosome_2_section, variant surface antigen, rifin family. » show all
236545	891..1021(37.1, 0.057) Detail	PRK09510, tolA, cell envelope integrity inner membrane protein TolA; Provisional.
337102	874..1010(37.3, 0.058) Detail	pfam08549, SWI-SNF_Ssr4, Fungal domain of unknown function (DUF1750). » show all
226714	977..1019(37.2, 0.060) Detail	COG4263, NosZ, Nitrous oxide reductase [Energy production and conversion].
275061	973..1009(36.1, 0.068) Detail	TIGR04219, hypothetical_protein, outer membrane protein. » show all
236776	899..987(36.5, 0.073) Detail	PRK10856, PRK10856, cytoskeletal protein RodZ; Provisional.
313663	647..693(37.0, 0.073) Detail	pfam10479, FSA_C, Fragile site-associated protein C-terminus. » show all
226728	222..297(35.7, 0.12) Detail	COG4278, COG4278, Uncharacterized conserved protein [Function unknown].
237865	890..1006(36.2, 0.12) Detail	PRK14951, PRK14951, DNA polymerase III subunits gamma and tau; Provisional.
318232	570..778(36.2, 0.14) Detail	pfam15967, Nucleoporin_FG2, Nucleoporin FG repeated region. » show all
237803	964..1019(36.1, 0.14) Detail	PRK14724, PRK14724, DNA topoisomerase III; Provisional.
237019	653..768(36.0, 0.16) Detail	PRK11907, PRK11907, bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase precursor protein; » show all
223021	888..1006(36.1, 0.17) Detail	PHA03247, PHA03247, large tegument protein UL36; Provisional.
316750	944..1008(34.9, 0.20) Detail	pfam14257, DUF4349, Domain of unknown function (DUF4349). » show all
225606	891..1022(35.3, 0.21) Detail	COG3064, TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].
235334	891..1005(35.3, 0.23) Detail	PRK05035, PRK05035, electron transport complex protein RnfC; Provisional.
234994	947..1003(34.3, 0.27) Detail	PRK01973, PRK01973, septum formation inhibitor; Reviewed.
184543	940..1008(33.0, 0.68) Detail	PRK14157, PRK14157, heat shock protein GrpE; Provisional.
227122	964..1000(33.5, 0.70) Detail	COG4783, COG4783, Putative Zn-dependent protease, contains TPR repeats [General function prediction only].
338200	970..1008(33.0, 0.74) Detail	pfam11981, DUF3482, Domain of unknown function (DUF3482). » show all
319019	911..966(33.5, 0.75) Detail	pfam16940, Tic110, Chloroplast envelope transporter. » show all
237110	891..1024(33.3, 0.90) Detail	PRK12472, PRK12472, hypothetical protein; Provisional.
130706	895..1002(33.1, 0.92) Detail	TIGR01645, Ro_ribonucleoprotein-binding_protein_1, poly-U binding splicing factor, half-pint family. » show all

Figure11Nucleotide variation of DNA-S genome components of BBTV



f) Segment DNA U3: encodes unknown function.

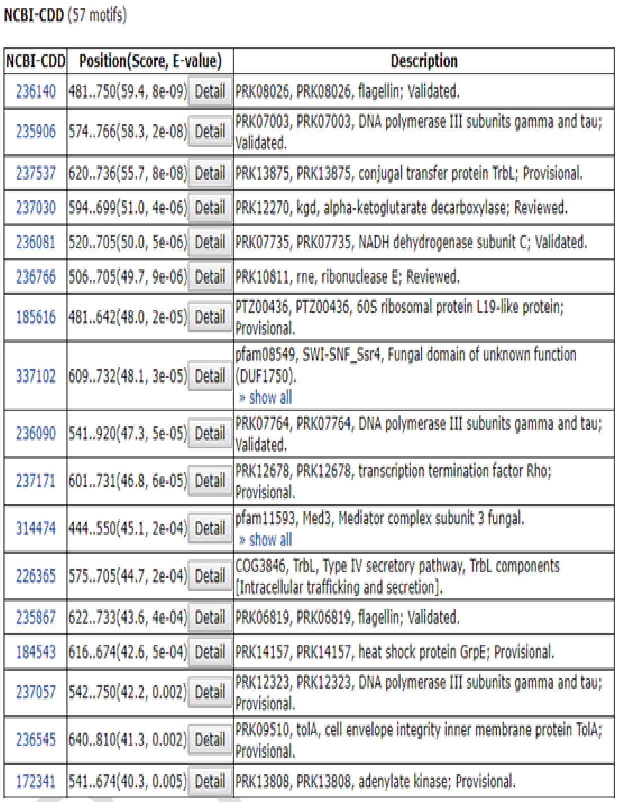
NCBI database : L41576.1 (**as reference**) , GQ214699.1, FJ773283.1 , EU366170.1 , FJ859750.1 , FJ859749.1 , JQ820454.1 , JQ820460.1 , JQ820466.1 , FJ859748.1

No conserved regions found using MEGA tool

Total 61 motif found from motif scan tool from DNA –U3 segments region.

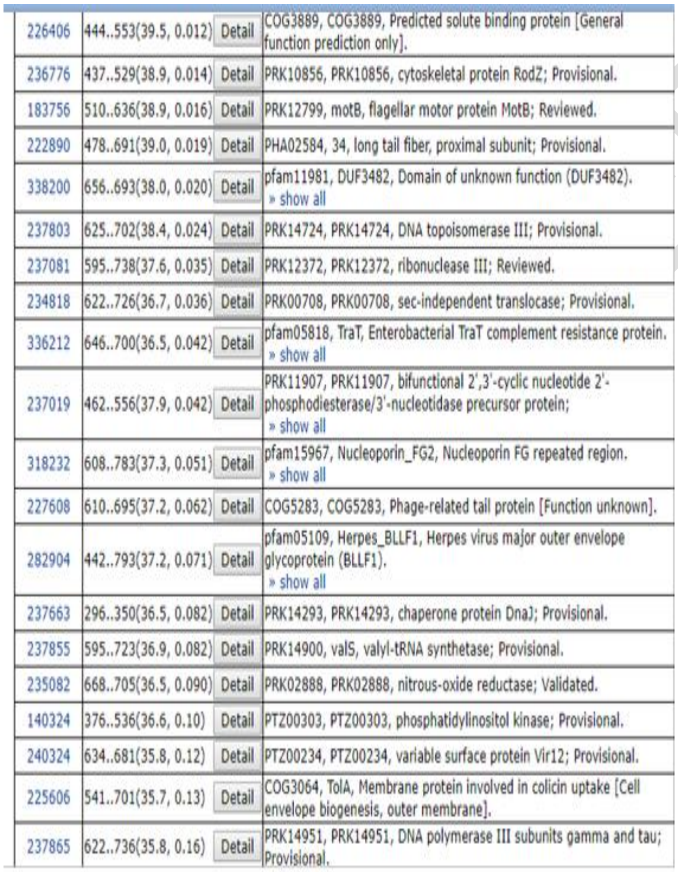
PROSITE PATTERN (4 motifs)

Found Motif	Position		Description	Related Sequences
2FE2S_FER_1	347..355 847..855	Detail	PS00197, 2Fe-2S ferredoxin-type iron-sulfur binding region signature.	287
ANAPHYLATOXIN_1	237..273	Detail	PS01177, Anaphylatoxin domain signature.	30
THIOLASE_3	380..393 546..559	Detail	PS00099, Thiolases active site.	255
4FE4S_FER_1	334..345 1019..1030 1047..1058	Detail	PS00198, 4Fe-4S ferredoxin-type iron-sulfur binding region signature.	1404

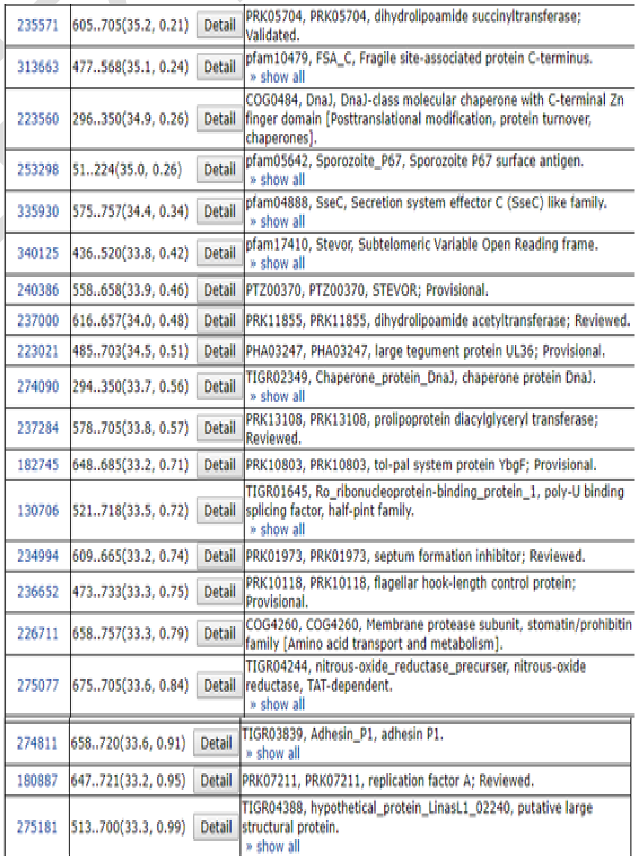


NCBI-CDD (57 motifs)

NCBI-CDD	Position(Score, E-value)		Description
236140	481..750(59.4, 8e-09)	Detail	PRK08026, PRK08026, flagellin; Validated.
235906	574..766(58.3, 2e-08)	Detail	PRK07003, PRK07003, DNA polymerase III subunits gamma and tau; Validated.
237537	620..736(55.7, 8e-08)	Detail	PRK13875, PRK13875, conjugal transfer protein TrbL; Provisional.
237030	594..699(51.0, 4e-06)	Detail	PRK12270, kgd, alpha-ketoglutarate decarboxylase; Reviewed.
236081	520..705(50.0, 5e-06)	Detail	PRK07735, PRK07735, NADH dehydrogenase subunit C; Validated.
236766	506..705(49.7, 9e-06)	Detail	PRK10811, rne, ribonuclease E; Reviewed.
185616	481..642(48.0, 2e-05)	Detail	PTZ00436, PTZ00436, 60S ribosomal protein L19-like protein; Provisional.
337102	609..732(48.1, 3e-05)	Detail	pfam08549, SWI-SNF_Ssr4, Fungal domain of unknown function (DUF1750). » show all
236090	541..920(47.3, 5e-05)	Detail	PRK07764, PRK07764, DNA polymerase III subunits gamma and tau; Validated.
237171	601..731(46.8, 6e-05)	Detail	PRK12678, PRK12678, transcription termination factor Rho; Provisional.
314474	444..550(45.1, 2e-04)	Detail	pfam11593, Med3, Mediator complex subunit 3 fungal. » show all
226365	575..705(44.7, 2e-04)	Detail	COG3846, TrbL, Type IV secretory pathway, TrbL components [Intracellular trafficking and secretion].
235867	622..733(43.6, 4e-04)	Detail	PRK06819, PRK06819, flagellin; Validated.
184543	616..674(42.6, 5e-04)	Detail	PRK14157, PRK14157, heat shock protein GrpE; Provisional.
237057	542..750(42.2, 0.002)	Detail	PRK12323, PRK12323, DNA polymerase III subunits gamma and tau; Provisional.
236545	640..810(41.3, 0.002)	Detail	PRK09510, tolA, cell envelope integrity inner membrane protein TolA; Provisional.
172341	541..674(40.3, 0.005)	Detail	PRK13808, PRK13808, adenylate kinase; Provisional.

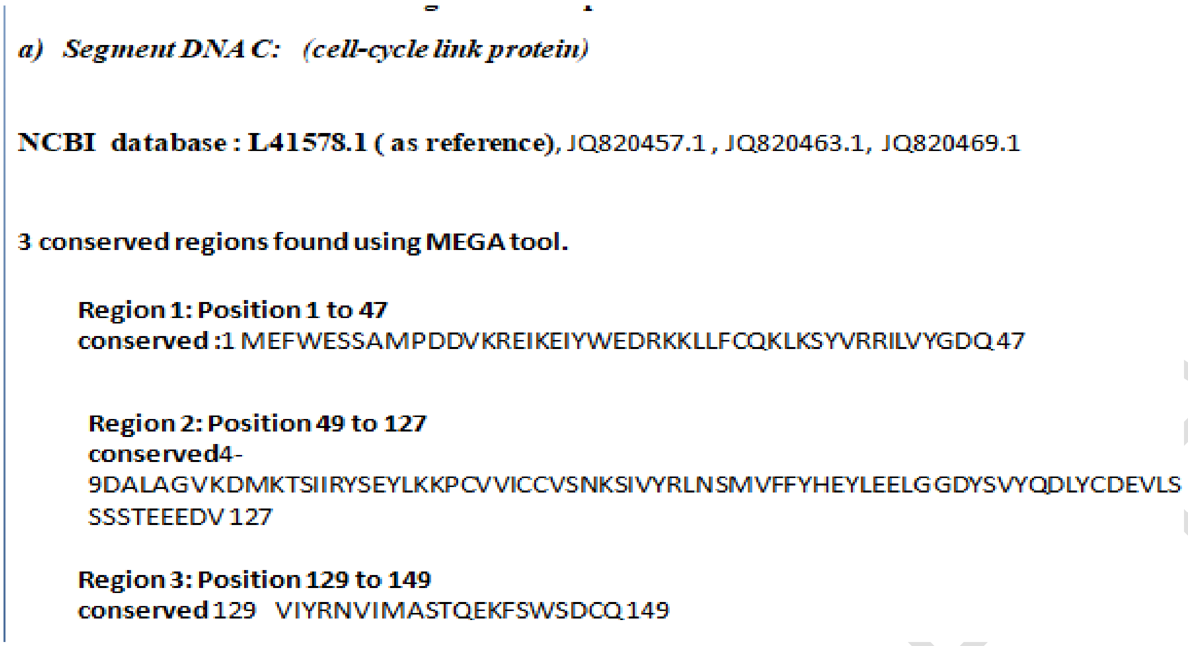


226406	444..553(39.5, 0.012)	Detail	COG3889, COG3889, Predicted solute binding protein [General function prediction only].
236776	437..529(38.9, 0.014)	Detail	PRK10856, PRK10856, cytoskeletal protein RodZ; Provisional.
183756	510..636(38.9, 0.016)	Detail	PRK12799, motB, flagellar motor protein MotB; Reviewed.
222890	478..691(39.0, 0.019)	Detail	PHA02584, 34, long tail fiber, proximal subunit; Provisional.
338200	656..693(38.0, 0.020)	Detail	pfam11981, DUF3482, Domain of unknown function (DUF3482). » show all
237803	625..702(38.4, 0.024)	Detail	PRK14724, PRK14724, DNA topoisomerase III; Provisional.
237081	595..738(37.6, 0.035)	Detail	PRK12372, PRK12372, ribonuclease III; Reviewed.
234818	622..726(36.7, 0.036)	Detail	PRK00708, PRK00708, sec-independent translocase; Provisional.
336212	646..700(36.5, 0.042)	Detail	pfam05818, TraT, Enterobacterial TraT complement resistance protein. » show all
237019	462..556(37.9, 0.042)	Detail	PRK11907, PRK11907, bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase precursor protein; » show all
318232	608..783(37.3, 0.051)	Detail	pfam15967, Nucleoporin_FG2, Nucleoporin FG repeated region. » show all
227608	610..695(37.2, 0.062)	Detail	COG5283, COG5283, Phage-related tail protein [Function unknown].
282904	442..793(37.2, 0.071)	Detail	pfam05109, Herpes_BLLF1, Herpes virus major outer envelope glycoprotein (BLLF1). » show all
237663	296..350(36.5, 0.082)	Detail	PRK14293, PRK14293, chaperone protein DnaJ; Provisional.
237855	595..723(36.9, 0.082)	Detail	PRK14900, valS, valyl-tRNA synthetase; Provisional.
235082	668..705(36.5, 0.090)	Detail	PRK02888, PRK02888, nitrous-oxide reductase; Validated.
140324	376..536(36.6, 0.10)	Detail	PTZ00303, PTZ00303, phosphatidylinositol kinase; Provisional.
240324	634..681(35.8, 0.12)	Detail	PTZ00234, PTZ00234, variable surface protein Vir12; Provisional.
225606	541..701(35.7, 0.13)	Detail	COG3064, TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].
237865	622..736(35.8, 0.16)	Detail	PRK14951, PRK14951, DNA polymerase III subunits gamma and tau; Provisional.



235571	605..705(35.2, 0.21)	Detail	PRK05704, PRK05704, dihydrolipoamide succinyltransferase; Validated.
313663	477..568(35.1, 0.24)	Detail	pfam10479, FSA_C, Fragile site-associated protein C-terminus. » show all
223560	296..350(34.9, 0.26)	Detail	COG0484, DnaJ, DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones].
253298	51..224(35.0, 0.26)	Detail	pfam05642, Sporozoite_P67, Sporozoite P67 surface antigen. » show all
335930	575..757(34.4, 0.34)	Detail	pfam04888, SseC, Secretion system effector C (SseC) like family. » show all
340125	436..520(33.8, 0.42)	Detail	pfam17410, Stevor, Subtelomeric Variable Open Reading frame. » show all
240386	558..658(33.9, 0.46)	Detail	PTZ00370, PTZ00370, STEVOR; Provisional.
237000	616..657(34.0, 0.48)	Detail	PRK11855, PRK11855, dihydrolipoamide acetyltransferase; Reviewed.
223021	485..703(34.5, 0.51)	Detail	PHA03247, PHA03247, large tegument protein UL36; Provisional.
274090	294..350(33.7, 0.56)	Detail	TIGR02349, Chaperone_protein_DnaJ, chaperone protein DnaJ. » show all
237284	578..705(33.8, 0.57)	Detail	PRK13108, PRK13108, prolipoprotein diacylglyceryl transferase; Reviewed.
182745	648..685(33.2, 0.71)	Detail	PRK10803, PRK10803, tol-pal system protein YbgF; Provisional.
130706	521..718(33.5, 0.72)	Detail	TIGR01645, Ro_ribonucleoprotein-binding_protein_1, poly-U binding splicing factor, half-pint family. » show all
234994	609..665(33.2, 0.74)	Detail	PRK01973, PRK01973, septum formation inhibitor; Reviewed.
236652	473..733(33.3, 0.75)	Detail	PRK10118, PRK10118, flagellar hook-length control protein; Provisional.
226711	658..757(33.3, 0.79)	Detail	COG4260, COG4260, Membrane protease subunit, stomatin/prohibitin family [Amino acid transport and metabolism].
275077	675..705(33.6, 0.84)	Detail	TIGR04244, nitrous-oxide_reductase_precurser, nitrous-oxide reductase, TAT-dependent. » show all
274811	658..720(33.6, 0.91)	Detail	TIGR03839, Adhesin_P1, adhesin P1. » show all
180887	647..721(33.2, 0.95)	Detail	PRK07211, PRK07211, replication factor A; Reviewed.
275181	513..700(33.3, 0.99)	Detail	TIGR04388, hypothetical_protein_LinasL1_02240, putative large structural protein. » show all

Figure12Nucleotide variation of DNA-U3 genome components of BBTV



a) Segment DNA C: (cell-cycle link protein)

NCBI database : L41578.1 (as reference), JQ820457.1 , JQ820463.1, JQ820469.1

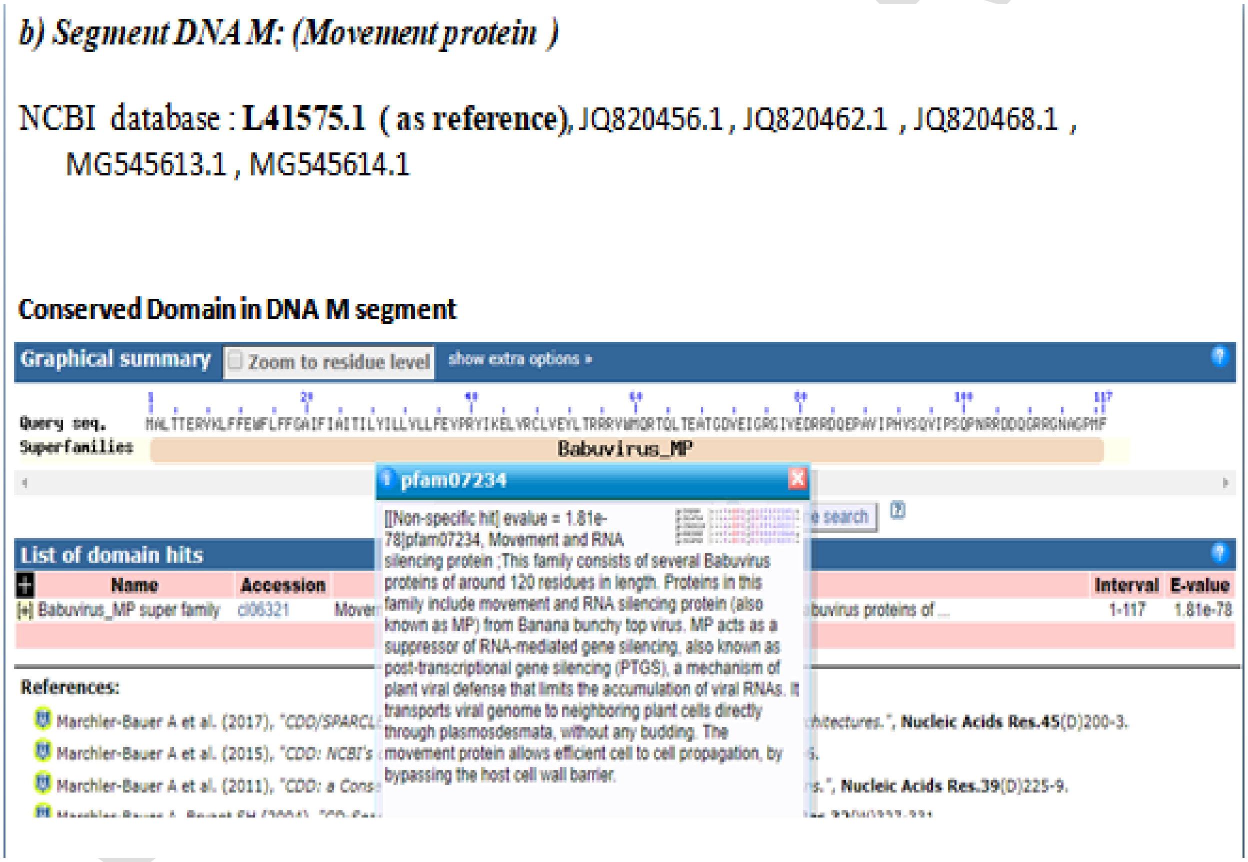
3 conserved regions found using MEGA tool.

Region 1: Position 1 to 47
conserved :1 MEFWESSAMPDDVKREIKEIYWEDRKKLLFCQKLKSYVRRILVYGDQ 47

Region 2: Position 49 to 127
conserved4-9DALAGVKDMKTSIIRYSEYLKKPCVVICCVSNKSIVYRLNSMVFFYHEYLEELGGDYSVYQDLYCDEVLSSSSTEEEDV 127

Region 3: Position 129 to 149
conserved 129 VIYRNVIMASTQEKFSWSDCQ 149

Figure13Amino acid variation of DNA-C genome components of BBTV



b) Segment DNA M: (Movement protein)

NCBI database : **L41575.1 (as reference),** JQ820456.1 , JQ820462.1 , JQ820468.1 , MG545613.1 , MG545614.1

Conserved Domain in DNA M segment

[[Non-specific hit] evalue = 1.81e-78]pfam07234, Movement and RNA silencing protein ;This family consists of several Babuvirus proteins of around 120 residues in length. Proteins in this family include movement and RNA silencing protein (also known as MP) from Banana bunchy top virus. MP acts as a suppressor of RNA-mediated gene silencing, also known as post-transcriptional gene silencing (PTGS), a mechanism of plant viral defense that limits the accumulation of viral RNAs. It transports viral genome to neighboring plant cells directly through plasmosdesmata, without any budding. The movement protein allows efficient cell to cell propagation, by bypassing the host cell wall barrier.

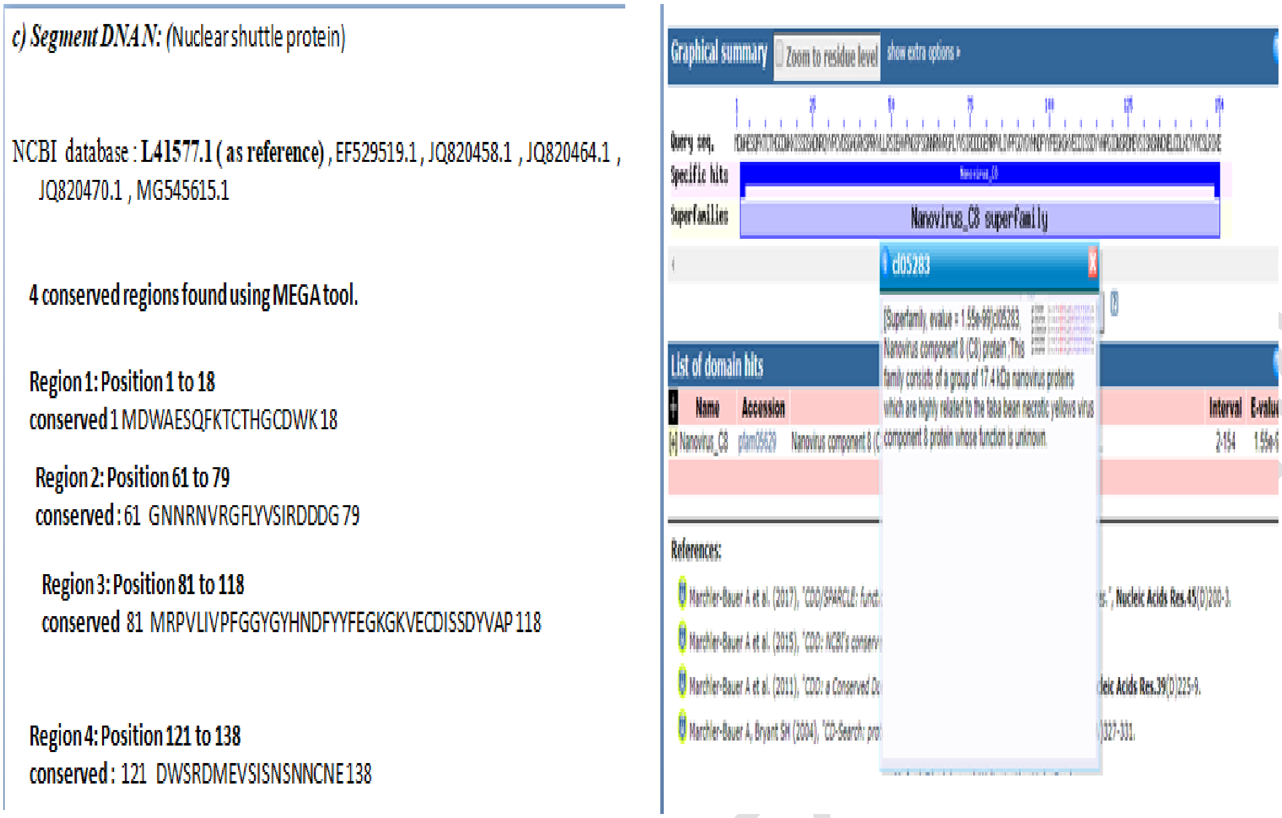
Name	Accession			Interval	E-value
[+] Babuvirus_MP super family	cl06321	Mover	buvirus proteins of ...	1-117	1.81e-78

Marchler-Bauer A et al. (2017), "CDD/SPARCL hitectures.", **Nucleic Acids Res.45**(D)200-3.

Marchler-Bauer A et al. (2015), "CDD: NCBI's 5.

Marchler-Bauer A et al. (2011), "CDD: a Cons s.", **Nucleic Acids Res.39**(D)225-9.

Figure14Amino acid variation of DNA-M genome components of BBTV



c) Segment DNA N: (Nuclear shuttle protein)

NCBI database : **L41577.1 (as reference)** , EF529519.1 , JQ820458.1 , JQ820464.1 , JQ820470.1 , MG545615.1

4 conserved regions found using MEGA tool.

Region 1: Position 1 to 18
conserved 1 MDWAESQFKTCTHGCDWK 18

Region 2: Position 61 to 79
conserved : 61 GNNRNVRGFLYVSIRDDDG 79

Region 3: Position 81 to 118
conserved 81 MRPVLIVPFGGYGYHNDFYYFEGKGKVECDISSDYVAP 118

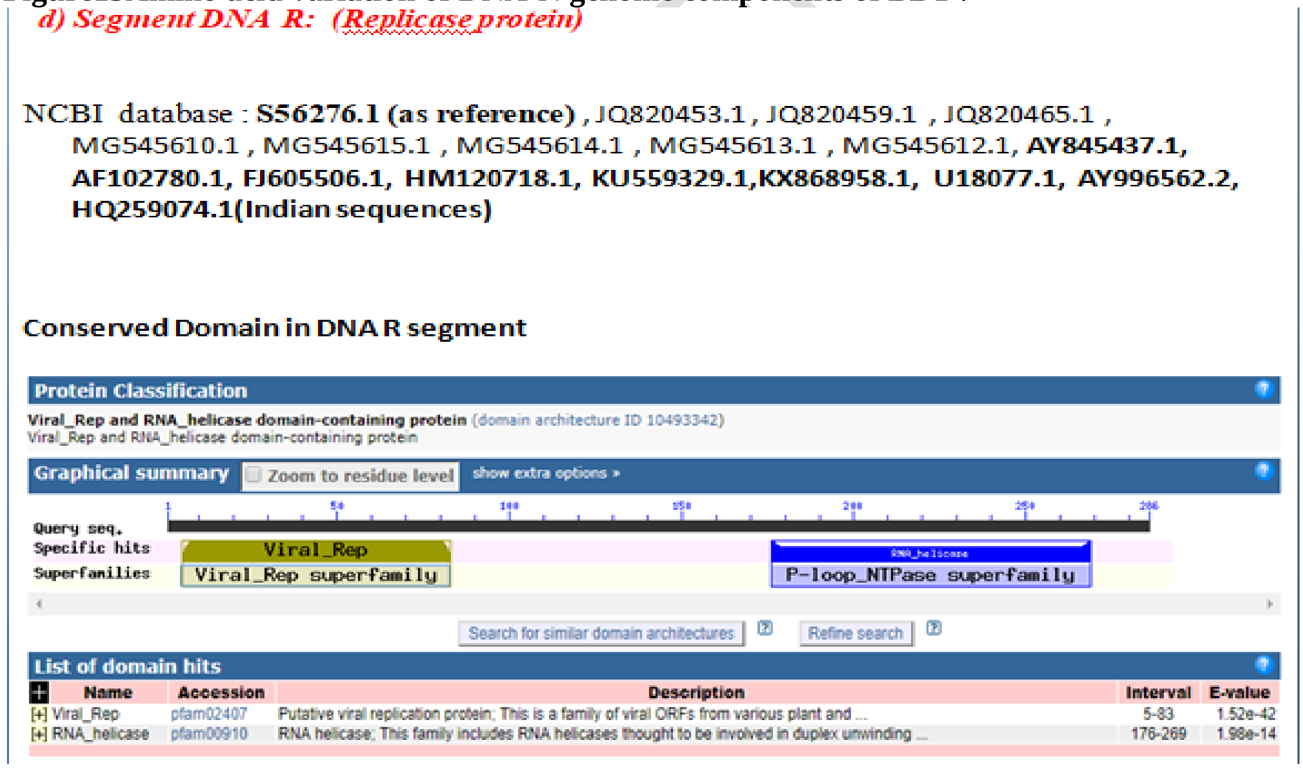
Region 4: Position 121 to 138
conserved : 121 DWSRDMEVSISNSNNCNE 138

[Superfamily, evalue = 1.55e-99] cl05283, Nanovirus component 8 (C8) protein ; This family consists of a group of 17.4 kDa nanovirus proteins which are highly related to the faba bean necrotic yellows virus component 8 protein whose function is unknown.

Name	Accession	Description	Interval	E-value
[+] Nanovirus_C8	pfam05629	Nanovirus component 8 (C	2-154	1.55e-9

Marchler-Bauer A et al. (2017), "CDD/SPARCLE: funct … s.", Nucleic Acids Res.45(D)200-3.
Marchler-Bauer A et al. (2015), "CDD: NCBI's conserv
Marchler-Bauer A et al. (2011), "CDD: a Conserved Do … leic Acids Res.39(D)225-9.
Marchler-Bauer A, Bryant SH (2004), "CD-Search: pro …)327-331.

Figure15Amino acid variation of DNA-N genome components of BBTV



d) Segment DNA R: (Replicase protein)

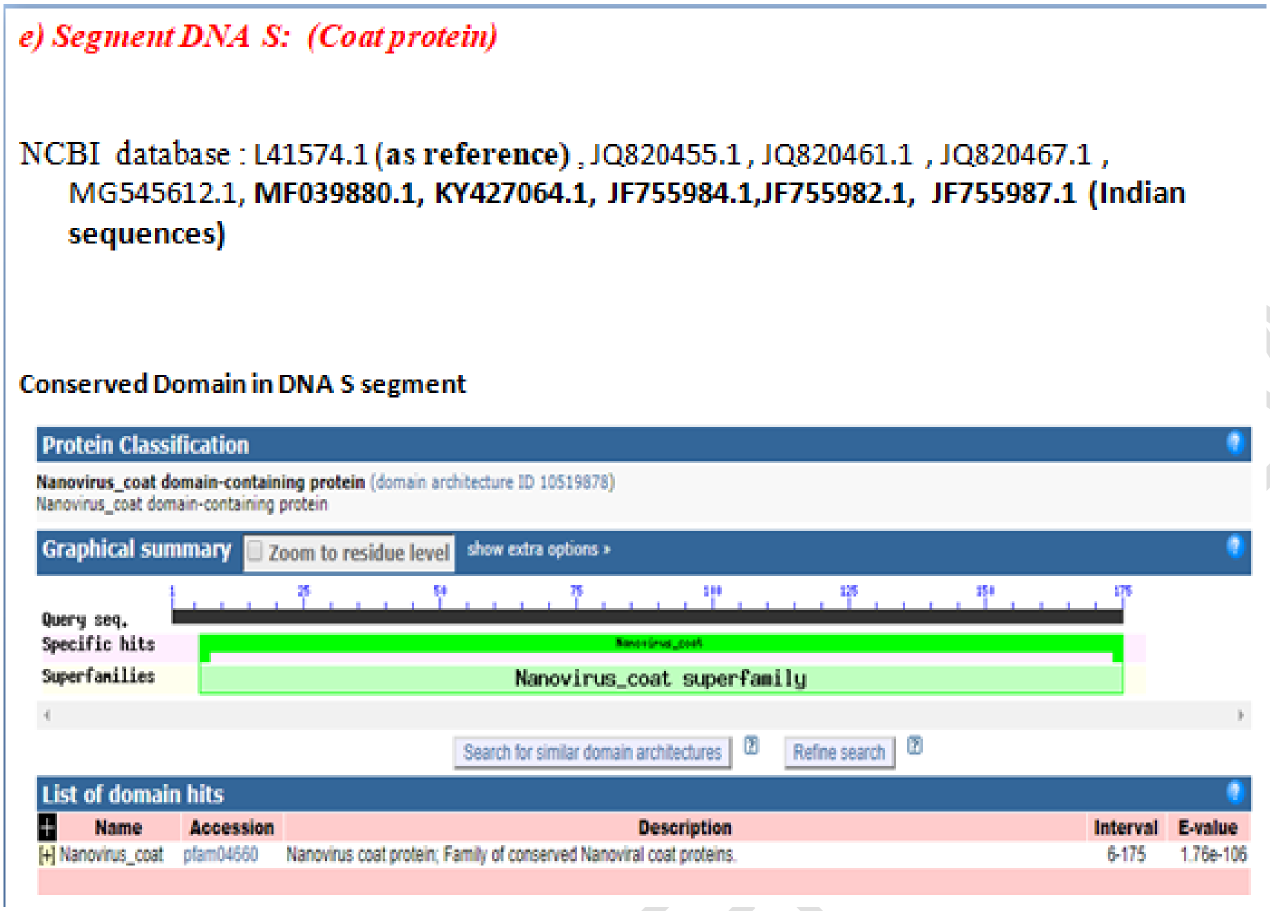
NCBI database : **S56276.1 (as reference)** , JQ820453.1 , JQ820459.1 , JQ820465.1 , MG545610.1 , MG545615.1 , MG545614.1 , MG545613.1 , MG545612.1, **AY845437.1, AF102780.1, FJ605506.1, HM120718.1, KU559329.1,KX868958.1, U18077.1, AY996562.2, HQ259074.1(Indian sequences)**

Conserved Domain in DNA R segment

Viral_Rep and RNA_helicase domain-containing protein (domain architecture ID 10493342)
Viral_Rep and RNA_helicase domain-containing protein

Name	Accession	Description	Interval	E-value
[+] Viral_Rep	pfam02407	Putative viral replication protein; This is a family of viral ORFs from various plant and ...	5-83	1.52e-42
[+] RNA_helicase	pfam00910	RNA helicase; This family includes RNA helicases thought to be involved in duplex unwinding ...	176-269	1.98e-14

Figure16Amino acid variation of DNA-R genome components of BBTV



e) Segment DNA S: (Coat protein)

NCBI database : L41574.1 (**as reference**) , JQ820455.1 , JQ820461.1 , JQ820467.1 , MG545612.1, **MF039880.1, KY427064.1, JF755984.1,JF755982.1, JF755987.1 (Indian sequences)**

Conserved Domain in DNA S segment

Protein Classification

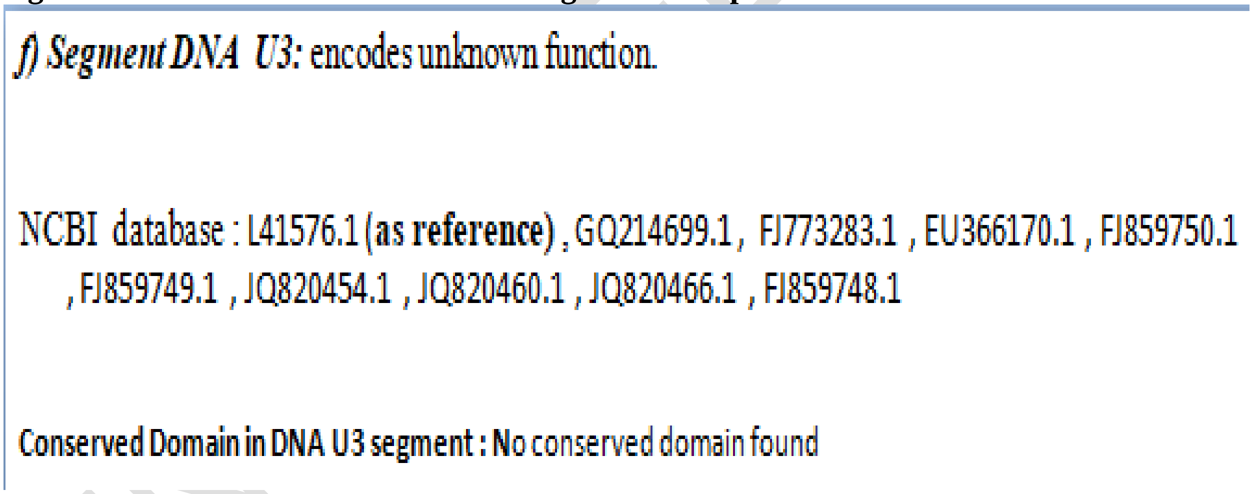
Nanovirus_coat domain-containing protein (domain architecture ID 10519878)
Nanovirus_coat domain-containing protein

Graphical summary ☐ Zoom to residue level show extra options »

List of domain hits

+	Name	Accession	Description	Interval	E-value
[+]	Nanovirus_coat	pfam04660	Nanovirus coat protein; Family of conserved Nanoviral coat proteins.	6-175	1.76e-106

Figure17Amino acid variation of DNA-S genome components of BBTV



f) Segment DNA U3: encodes unknown function.

NCBI database : L41576.1 (**as reference**) , GQ214699.1 , FJ773283.1 , EU366170.1 , FJ859750.1 , FJ859749.1 , JQ820454.1 , JQ820460.1 , JQ820466.1 , FJ859748.1

Conserved Domain in DNA U3 segment : No conserved domain found

Figure18Amino acid variation of DNA-U3 genome components of BBTV