Isolation, Partial Purification and Characterization of a Novel Restriction Enzyme from *Pseudomonas Anguilliseptica*

Swarna Nirosha Jayasinghe Pathirana (niroshapathirana2010@gmail.com)
University of Colombo https://orcid.org/0000-0002-8013-2147
Don Anushka Sandaruwan Elvitigala
University of Colombo
Chandrika Malkanthi Nanayakkara
University of Colombo
Prashanth Suravajhala
Birla Institute of Scientific Research
Sanath Rajapakse
University of Peradeniya
Gardhi Hewage Chamari Madhu Hettiarachchi
University of Colombo
Naduviladath Vishvanath Chandrasekharan
University of Colombo

Research Article

Keywords: Restriction Enzymes, HindIII Isoschizomer, Pseudomonas angulliseptica, Whole-genome sequencing

Posted Date: June 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-600889/v1

Abstract

Type II restriction enzymes (REs) which can cleave double stranded DNA in a sequence specific manner have many applications in recombinant DNA technology and are considered the work horses of molecular biology. Soil and water samples were screened for isolation of bacteria, harboring restriction enzymes. Cell lysates of isolated bacteria were incubated with unmethylated λ DNA, followed by analysis by agarose gel electrophoresis. The presence of distinct banding patterns indicated the presence of REs. Nine putative isolates harboring REs were morphologically and molecularly characterized using 16S rRNA analysis and belonged to four different genera (*Acinetobacter, Lysinibacillus, Pseudomonas,* and *Brevibacillus*).

A *Hin*dIII like restriction digestion profile was observed in a lysate of a soil bacterium belonging to genus *Pseudomonas*. Based on 16S rRNA analysis, the bacterial species was identified as *P. angulliseptica.* The enzyme was partially purified and optimum conditions for enzyme activity and its recognition sequence were determined. The enzyme showed optimum activity at 40 ^{0}C and was stable at 40 $^{\circ}$C for 20 minutes without the DNA substrate. The Recognition sequence of the enzyme was determined and found to be 5'AAGCT 3' indicating it to be an isoschizomer of *Hin*dIII.

The whole genome of the *Pseudomonas* species was sequenced and the coding sequence of the gene for the putative *Hin*dIII isoschizomer was identified together with other genes encoding putative REs. The gene coding for the *Hin*dIII isoschizomer was analyzed *in silico* and its homology and evolutionary relationship to other known isoschizomers of *Hin*dIII were determined. The enzyme was tentatively designated as *Pan*I.

1.0 Introduction

All living organisms face the challenge of survival and threats posed by their pathogens. Bacteria exert different mechanisms for their survival, one of which is driven by 'restriction enzymes' (REs) or 'restriction endonucleases' (REases) which represent a type of bacterial immune system, dedicated to shielding them from phage attack. REs restricts growth of phages by cleaving double-stranded phage DNA at precise nucleotide positions. Methyltransferases (MTases) modify bacterial DNA at sequences recognized by REases to prevent self-cleavage of its own genome. Collectively, these two enzymes form the "restriction-modification (R-M) system" in bacteria (Williams 2003). The recognition sites of REases generally contain variable numbers of base pairs (four, six or eight) (Wilson and Murray 1991). There are four major types of REs, namely, type I, II, III and IV, categorized according to their structure, recognition site, cleavage site, cofactor(s), and activator(s) (Williams 2003; Sistla and Rao 2004; Loenen et al. 2014b). While they have the ability to cleave double-stranded DNA in a sequence-specific manner, irrespective of the source of DNA; with the advent of recombinant DNA technology, extensive research is being carried out worldwide for isolation, purification, and characterization of novel REs. Current trends in restriction enzyme research encompass the use of protein engineering and fusions to produce enzymes with novel endonuclease activities (Di Felice et al. 2019).

Type II REs represent the highest number of characterized REs to date. Over 300 Type II REs, with > 200 dissimilar sequence-specificities, are commercially available (Loenen et al. 2014b, a). These constitute a very diverse group of proteins in terms of size and amino acid sequence, organization of the domains and protein subunit composition, co-factor requirement for enzymatic activity and reaction mechanisms (Pingoud et al. 2005; Loenen et al. 2014a).

Type II REs are fundamental tools in DNA manipulation and play an important role in genetic engineering and molecular biology (Williams 2003), because they have the ability to cleave DNA at defined positions close to or within their recognition sequences producing distinct reproducible banding patterns on agarose gels (Roberts 1976). Generally, they identify short, palindromic, 4–8 base pair (bp) sequences (Pingoud and Jeltsch 2001; Pingoud et al. 2014) and cleave DNA in the presence of Mg^{2+}. Type II REases are the most commonly used group of enzymes in laboratories for gene cloning, DNA analysis, library preparation, diagnostic purposes, etc. Their sizes range from 250–350 amino acids and are the simplest and smallest among the REases. REases that identify the same DNA sequence, irrespective of the site at which they cleave the DNA are designated as 'isoschizomers'(Pingoud et al. 2014), while the first enzyme which revealed the recognition sequence is known as the prototype.

Among the type II REases, the enzyme *Hin*d III was isolated from the bacterium *Haemophilus influenzae*, serotype d (Pingoud et al. 2014) and is used extensively in genetic engineering and molecular biology. Isoschizomers of *Hin*d III REase reported to date are listed in the REBASE database (Kazennova et al. 1982; Mise and Nakajima 1984). In this study, an isolate of *Pseudomonas anguilliseptica* was identified from a soil sample and was found to produce an isoschizomer of *Hin*dIII. The enzyme was extracted from the bacterium, partially purified, functionally characterized and tentatively designated as *Pan*I. Furthermore, through whole-genome sequencing the putative gene encoding the enzyme was identified and molecularly characterized using different *in-silico* tools.

2.0 Materials And Methods

2.1 Sample collection, Isolation of bacteria from soil samples, Screening isolated bacteria for REs and Characterization of restriction enzyme-producing bacteria.

Soil samples were collected using sterile 50 mL conical tubes from different regions of the country, transported to the laboratory and stored at 4 ^{0}C to isolate bacteria. One gram of each soil sample was vigorously mixed with 5 mL of sterile PBS (1X: 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4) separately and a 100 µL aliquot was serially diluted (10^{-1} to 10^{-10}) using PBS. An aliquot (100 µL) from each dilution was spread on Luria-Bertani (LB) agar plates and incubated at 37 ^{0}C, and observed daily for a week. Well-isolated colonies were picked for screening.

Well isolated colonies were separately inoculated into 5 mL LB broth and grown overnight. The cultures were then centrifuged at 3500 rpm for 30 minutes and cell pellets were washed with TME buffer (50 mM Tris-HCl, 20 mM $MgCl_2$ and 0.1 mM EDTA, pH 7.5) and then centrifuged again using the same conditions.

Cell pellets were then re-suspended in 500 µL of TME buffer along with 0.5µL of β-mercaptoethanol. The cell suspensions were then disrupted by cold sonication (3x 2second pulses with 15 second intervals in between) and then centrifuged (18000 g /3 min/ 4°C) followed by careful separation of the supernatant. The supernatant of each cell lysate was then screened for REs as follows; A 30 µL of reaction mixture (digestion mixture) containing 5.0 µL of cell lysate, 1X Multicore (MC) buffer (Promega), 0.5 µg of un-methylated λ DNA and deionized water was incubated at 37ºC for 3 h. Aliquots (15 µL) from the above reaction mixtures were electrophoresed on a 0.8% agarose gel and visualized under a UV trans-illuminator for the presence of distinct banding patterns.

Several isolates displayed distinct DNA banding patterns on agarose gel electrophoresis. The colony and cellular morphology, including shape, height, margin, surface refraction, opacity and color of the bacterial colony was observed as described elsewhere. (Bhumbla and Bhumbla 2018). An isolate designated 'MatS1' was selected for further analysis.

The MatS1 isolate was molecularly identified at the species level using 16S rRNA gene sequence analysis by employing the universal primers (RNABR1 – AGAGTTTGATCCTGGCTCAG and RNABR2-AAGGAGGTGATCCAGCC) in a standard PCR amplification reaction (Weisburg et al. 1991) followed by sequence verification of the amplicon.

2.2 Partial purification of MatS1 lysate, Determination of the recognition sequence, Endonuclease activity assays, Enzyme activity in commercially available buffers, Determination of optimum temperature, Thermal stability of the partially purified enzyme.

Partial purification of MatS1 by size exclusion filtration was carried out using molecular weight cut-off filters (Vivaspin 6 – Sigma-Aldrich) according to the manufacturer's instructions. Briefly, a cell free lysate of MatS1 was prepared as described above for screening (Sect. 2.3) and 2 mL was filtered through a 50 kDa molecular weight cut off filter and the enzyme activity of the retentate and filtrate determined. The filtrate was then passed through 30 kDa molecular weight cut off filter and enzyme activity were assayed in both the filtrate and retentate.

To determine the recognition sequence, λ DNA (0.5 µg) was incubated with 5.0 µL of a lysate of MatS1 at 37°C for 3 h. Several reactions were carried out and the digestion mixtures were pooled, phenol chloroform extracted, precipitated and resuspended in TE buffer (pH 7.9). An aliquot of this DNA was treated with Taq polymerase (5 U) in a reaction mixture containing 2mM dNTPs and 1X PCR buffer, for 20 minutes at 72 ^{0}C to fill in any 3' recessed ends and to add an 'A' residue to the 3' end of the DNA fragments to facilitate TA cloning. The DNA fragments were then resolved on an agarose gel and fragments less than 2000 bp in size were excised from the gel. Excised fragments were purified using Wizard® SV gel and PCR clean up system (Promega), according to manufacturer's instructions. Purified fragments were then ligated to pGEM-T easy vector (Promega) and transformed in *E.Coli* (JM109) competent cells. Recombinant plasmid DNA was prepared from recombinant clones and the ends of the insert sequenced using the dideoxy method (Macrogen- Korea).

Activity assays were carried out using the partially purified enzyme (5 µL) and unmethylated λ DNA (0.7 µg) as a substrate in a 30 µL reaction volume for 2 h. The cleavage products were then resolved on a 0.8 % agarose gel and the relative activities were evaluated based on the intensities of the DNA bands. Enzyme activity in several commercially available buffers (Promega - Table 1) were determined as follows; λ DNA (0.7 µg) was digested with 5.0 µL of the purified lysate in the respective buffer (1X) for 2 h at 37 ºC in a 30 µL reaction mixture. The resulting DNA fragments were then resolved on a 0.8 % agarose gel and the intensity of the resolved bands were determined using IMAGEJ software (https://imagej.nih.gov/ij/download.html). The relative enzyme activity was determined based on the intensities of the cleaved DNA fragments.

Table 1
Composition of commercially available restriction endonuclease buffers (1X) (Promega).

Buffer	pH at 37^0C	Tris HCl (mM)	$MgCl_2$ (mM)	NaCl (mM)	KCl (mM)	DTT (mM)
B	7.5	6	6	50	-	1
D	7.9	6	6	150	-	1
E	7.5	6	6	100	-	1
H	7.5	90	10	50	-	-
J	7.5	10	7	-	50	1
K	7.4	10	10	-	150	1
MC	25 mM Tris acetate (pH 7.5 at 37°C), 100 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT					

Lambda DNA (0.7 µg) was digested in a reaction mixture (30 µL) containing 1X MC buffer (Promega), 5 µL of the partially purified enzyme, at temperatures ranging from 0 to 90 ºC (0, 20, 37, 40, 50, 60, 70, 80 and 90 ºC) for 2 h. The digestion mixture was then electrophoresed on 0.8 % agarose gel and the optimum temperature was determined based on the intensity of the bands produced.

The partially purified enzyme was pre-incubated at different temperatures (0, 20, 37, 40, 50, 60, 70, 80 and 90^0C) for 20 minutes and restriction digestion reactions were carried out as described above (Sect. 2.8.2).

2.3 Whole-genome sequencing of MatS1 and In-silico characterization of the Pan I gene

Genomic DNA was extracted from the isolated *Pseudomonas sp*. using the DNeasy Blood & Tissue Kit, as per manufactures instructions (Qiagen – USA). The DNA was then *de novo* sequenced using PacBio technology (GeneWiz - China). Briefly, genomic DNA was sheared and 20 Kb double-stranded DNA fragments were selected, end repaired and ligated with universal hairpin adapters. SMRTbell library was constructed and sequenced in PacBio RSII SMRT (Mccarthy 2010) sequencer. PacBio reads were

assembled using PBcR of WGS-Assembler 8.2 (Berlin et al. 2015). The Glimmer (Delcher et al. 2007) gene-finding software was used for the identification of coding genes. The coding genes were annotated against the NR database of the National Center for Biotechnology Information (NCBI), using BLAST.

Based on the alignments obtained from NCBI Genbank sequence database with the contiguous sequences of MatS1 whole genome using BLAST, a homolog which codes a *Hin*dIII like endonuclease was identified. The corresponding amino acid sequence of the homolog was obtained by using Uni-pro-U-GENE software (Okonechnikov et al. 2012) and substrate (DNA) binding sites were predicted by COACH online server (Yang et al. 2013). Sequence comparison studies of predicted protein sequence with its homologs were performed using the Clustal-Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and EMBOSS-Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) sequence alignment programs. With the objective of investigating the evolutionary relationship with its counterpart molecules, phylogenetic analysis of the identified sequence (MatS1- *Hin*dIII) was carried out using the Neighbor-joining method with bootstrapping values taken from 1000 replicates, using Molecular Evolutionary Genetics Analysis (MEGA) software, version 6 (Tamura et al. 2013). Characteristic signatures of the MatS1 *Hin*dIII like protein sequence were predicted by the ExPASy-prosite server (http://prosite.expasy.org) and some of the physicochemical properties were identified using the ExPASy ProtParam tool (http://web.expasy.org/protparam). Tertiary structure of the enzyme was predicted by I-TASSER online server (Yang and Zhang 2015) along with the potential ligand binding sites and the corresponding ligands and visualized using PyMol (v1.3) molecular visualization software.

3.0 Results And Discussion

When the cell-free extracts of bacterial isolates were separately incubated with un-methylated lambda DNA (Promega), a banding pattern similar to that of lambda DNA cleaved with commercially available *Hin*dIII was observed (Fig. 1) with one isolate. This was designated as MatS1 and selected for further analysis.

MatS1was found to be a Gram negative rod shaped bacterium, dark orange in color and the colony was sticky, shiny, circular, raised with entire margin and swarming. Analysis of the sequence of 16S rRNA gene using BLAST confirmed the isolate to be a strain of *Pseudomonas anguilliseptica*, evidenced by a 99.6% sequence identity with *Pseudomonas anguilliseptica* VITEPRRL6 strain (GenBank ID - KR149276). Therefore, the *Hin*dIII like enzyme of MatS1 was tentatively designated as *Pan*I. This is a first report of a *HindIII* like enzyme characterized from the genus *Pseudomonas*.

Partial purification of MatS1was carried out by size exclusion, using molecular weight cut off filters. Lambda DNA cleavage by concentrated lysate using 50 and 30 kDa molecular weight cut-off filters displayed two different banding patterns (Fig. 2). A banding pattern similar to that of λ DNA cleaved by *Hin*dIII was observed in the retentate of the 30 kDa filter and filtrate of the 50 kDa filter, suggesting the molecular weight of the *Hin*dIII like enzyme of MatS1 to lie between 30 and 50 kDa. Moreover, the observed cleavage pattern of λ DNA with the retentate of the 50 kDa filter indicates the presence of other

restriction enzymes with different specificities and molecular weights above 50 kDa. This observation was further confirmed by the whole genome sequence analysis of MatS1, which indicated the presence of other genes encoding putative restriction enzymes.

The recognition sequence of RE of MatS1 (*Pan*I) was determined by cloning and sequencing the ends of λ DNA fragments produced by digestion of λ DNA with partially purified *Pan*I. The recognition sequence was found to be “AAGCTT”.

Partially purified *PanI* was active in most of the commercially available buffers including buffer B, which is the optimum buffer for commercially available *Hin*dIII (Promega). However, optimum activity of *Pan*I was observed in multi-core (MC) buffer (Fig. 3), which contained 10 mM Mg^{2+} as a divalent cation and 100 mM K^{+} as a monovalent cation and a pH of 7.5 (Table 1).

The optimum temperature of the partially purified enzyme appeared to be between 37ºC to 50ºC (Fig. 4). The highest intensity of the banding pattern was observed at 40 ^{0}C. Above 70ºC the activity decreased and detectable activity was observed even at 80ºC, suggesting its functionality even under extreme temperatures. A previously reported, isoschizomers of *Hin*dIII; *Eco*VIII also showed optimal activity at 48°C, aligning with our observations (Mise and Nakajima 1984).

To determine the thermal stability of *Pan*I without the DNA substrate, partially purified enzyme was pre-incubated at different temperatures. The enzyme was found to be stable up to 40 ^{0}C for 20 minutes without the DNA substrate (Fig. 5).

Whole-genome sequence analysis of MatS1 revealed the presence of a complete coding sequence of a gene having a high degree of similarity to known genes encoding *Hin*dIII family enzymes, reported in NCBI-GenBank database; further confirming that partially purified *Pan*I in MatS1 is an isoschizomer of *Hin*d III. The sequence information of *Pan*I was deposited in NCBI GenBank database (GenBank ID - MW140018). The complete putative ORF of *Pan*I was 912 bp and encodes a protein of 304 amino acids with a predicted molecular weight of ~ 34.6 kDa and a theoretical iso-electric point of 6.18. This predicted molecular weight is in agreement with the empirically estimated range of molecular weight (30–50 kDa) of partially purified *Pan*I. The protein sequence has a region that resembles the partially conserved *Hin*dIII endonuclease superfamily signature (residues 13–292) and several conserved substrate binding sites (DNA) as predicted by COACH online server (Fig. 8 - (Yang et al. 2013). Based on the pairwise sequence alignment, *Pan*I shared significant sequence relatedness with its bacterial counterparts with a maximum similarity of 89.1 % and identity of 76.3 % with that of *Cylindrospermopsis raciborskii*, validating its homology to known *Hin*dIII counterparts (Table 2).

Table 2. Percentage similarity and identity of *Pan*I with different homologues.

Species name	Restriction Endonuclease (RE)	NCBI_GenBank Accession Number	Length (amino acids)	% Identity	% Similarity
1. *Cylindrospermopsis raciborskii*	*Hin*dIII family	WP071241953	304	76.3	89.1
1. *Pseudomonas mygdali*	*Hin*dIII family	WP057425488	304	73.1	85.6
1. *Pseudomonas meliae*	*Hin*dIII family	WP054991699	304	73.1	85.2
1. *Thermoflexibacter ruber*	*Hin*dIII family	WP091549217	304	69.2	82.6
1. *Chlorobi bacterium OLB4*	*Hin*dIII family	KXK03935	304	66.9	82.6
1. *Bacteroidetes bacterium RIFCSPLOWO2*	RE	OFY69087	304	66.6	82.6
1. *Ignavibacteria bacterium*	RE	OIO24181	304	65.2	82.3
1. *Bacteroidales bacterium Barb6XT*	*Hin*dIII family	WP066182127	304	62.3	80.7
1. *Planktothrix sp. PCC 11201*	*Hin*dIII family	WP079679536	304	61.3	78.7
1. *Geminocystis herdmanii*	*Hin*dIII family	WP017294927	304	60.3	78.0
1. *Arthrospira*	*Hin*dIII family	WP006621132	304	60.3	77.0
1. *Oscillatoria acuminate*	*Hin*dIII family	WP015150672	304	59.0	77.7
1. *Chlorobi bacterium OLB7*	*Hin*dIII family	KXK52534	303	58.6	79.3
1. *Enterobacteriaceae*	*Hin*dIII family	WP015059042	307	55.8	73.2

1. *Aquamicrobium aerolatum*	*Hin*dIII family	WP091525013	304	55.1	71.8
1. *Escherichia coli*	*Hin*dIII family	WP024235892	308	54.4	72.2
1. *Escherichia coli* (Plasmid)	*Eco*VIII	AAA91203	333	51.5	67.6

In the phylogenetic reconstruction, *PanI* was clustered with known bacterial *Hin*dIII family enzymes (Fig. 7). According to the tree topology, *Pan*I was closely clustered with *Pseudomonas Hin*dIII homologs. However, it showed its highest evolutionary relationship to its counterpart in C. *raciborskii*, by forming a sub-clade with it in the main cluster which also harbors *Hin*dIII family enzymes of *Pseudomonas* species, with high bootstrapping support (85).

This pattern of clustering suggests a possible horizontal gene transfer event between *P. anguliseptica and C. raciborskii* with respect to *Hin*dIII like protein coding gene, which is also further reinforced by the pronounced sequence identity of *Pan*I with *C. raciborskii* HindIII family protein, compared to those from other two *Pseudomonas* species (Table 2). However, further investigations are warranted for the validation of this likelihood.

I-TASSER online server predicted the tertiary structure of *Pan*I based on 10 threading templates identified from the research Collaboratory for Structural Bioinformatics (RCSB) protein data bank, of which the normalized Z score of the threading alignments was between 2.27 to 11.92, confirming the credibility of each alignment. The most reliable model with a substantial global accuracy, measured by C-score (1.00) with estimated TM-score of 0.85 ± 0.08 and RMSD of 4.2 ± 2.8 A was selected for visualization on PyMol software.

According to the generated model, *Pan*I consists of 15 α-helices and 1 β pleated sheet with five strands (Fig. 8A). However, the monomeric form of the empirically determined crystal structure of *Hin*dIII was found to be made up of 16 α-helices and 2 β pleated sheets; one with two strands, and the other with five strands (Watanabe et al. 2009). Thus, comparison of these two structures suggests that the five stranded β pleated sheet is conserved in *Pan* I. Moreover, in compliance with the *Hin*dIII crystal structure, the first strand of the five stranded β sheet is oriented in a parallel direction with the fifth strand in the predicted tertiary structure of *Pan*I

As a part of the computer based simulation performed by I-TASSER, potential DNA substrate binding sites were predicted based on the modeled tertiary structure of *Pan*I along with the three dimensional structure of the enzyme - cognate DNA complex, by using COBOLT and COFACTOR algorithms (Yang et al. 2013). The most reliable prediction with the highest C-score (0.55) and cluster size (35) consists of 8 potential DNA binding residues, namely from the N terminal, Ser-31, Thr-69, Asp-67, Lys-72, Ala-127, Asn-129, Lys-131 and Lys-284 with a binding probability of over 0.5 (Fig. 8B). The repeated occurrence of positively charged amino acid, Lys in the binding site indicates potentially strong interactions between negatively

charged DNA and the enzyme via formation of ionic bonds. As expected, the cognate stretches of DNA which overlaps with the active site of the enzyme was predicted to bear the consensus *Hin*dIII recognition sequence 'AAGCTT' validating the empirically determined recognition sequence of *Pan*I.

Conclusions

REs are powerful tools used in molecular biology and genetic engineering. In this study, several soil and water samples were screened for the isolation of restriction enzyme producing bacteria. Potent *Hin*dIII like activity was observed in the cell-free extract of an isolate designated MatS1 and was identified as a novel strain of *Pseudomonas anguilliseptica* through 16S rRNA analysis. The restriction enzyme isolated from this organism was designated as *Pan*I. The isolated enzyme was partially purified and characterized in relation to its recognition sequence and optimum reaction conditions for DNA digestion. The recognition sequence was found to be 5'AAGCTT 3' and revealed *Pan*1 to be an isoschizomer of *Hin*dIII. The whole genome of MatS1 was sequenced and the gene for *Pan*I was identified and characterized.

Declarations

Acknowledgments

We are grateful to Mr. H. M. J. C. B. Herath, Dr. Y.C. Guruge, Mr. Chatura Samarasinghe, Mr. Charitha Samarasinghe, Mrs. Padmini Wijenayake and Mr. K.M.Karunajeewa for their generous assistance in collecting soil and water samples for the study and Dr. (Mrs.) W.W.P. Rodrigo for her kind assistance in Bioinfomatic analysis. This research study was supported by a grant (RG/2014/BT/2) from National Science Foundation, Sri Lanka.

Funding

This research study was supported by a grant (RG/2014/BT/2) from the National Science Foundation, Sri Lanka.

Competing interests

The authors Swarna Nirosha Jayasinghe Pathirana, Don Anushka Sandaruwan Elvitigala, Chandrika Malkanthi Nanayakkara, Prashanth Suravajhala, Sanath Rajapakse, Gardhi Hettiarachchige Chamari Madhu Hettiarachchi and Naduviladath Vishwanath Chandrasekharan declare that they have no competing interests.

Availability of data and materials

Not applicable

Code availability

Not applicable

CRediT authorship contribution statement

Swarna Nirosha Jayasinghe Pathirana: Data curation, Investigations, Formal analysis, Methodology, Project administration, Writing original draft; **Don Anushka Sandaruwan Elvitigala**: Formal analysis; Methodology; Writing-review and editing; **Chandrika Malkanthi Nanayakkara:** Writing-review and editing, Supervision, Methodology; **Prashanth Suravajhala**: Writing-review and editing, methodology, Formal analysis; **Sanath Rajapakse:** Writing-review and editing, methodology, Resources, Supervision; **Gardhi Hettiarachchige Chamari Madhu Hettiarachchi:** Funding acquisition, Project administration, Supervision, Methodology; **Naduviladath Vishwanath Chandrasekharan:** Conceptualization, Funding acquisition, Data curation, Formal analysis, Methodology, Project administration, Resources, Writing-review and editing, Supervision

References

1. Berlin K, Koren S, Chin CS et al (2015) Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 33:623–630. https://doi.org/10.1038/nbt.3238
2. Bhumbla U, Bhumbla U (2018) Gram's Staining. In: Workbook for Practical Microbiology. Jaypee Brothers Medical Publishers (P) Ltd., pp 30–30
3. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. https://doi.org/10.1093/bioinformatics/btm009
4. Di Felice F, Micheli G, Camilloni G (2019) Review Restriction enzymes and their use in molecular biology: An overview. J Biosci 44:. https://doi.org/10.1007/s12038-019-9856-8
5. Kazennova EV, Tarasov AP, Mileikovskaia MM, Tsevetkova NV (1982) Isolation and purification of restriction endonuclease BpeI from Bordetella pertussis. ZH Mikrobiol Epidemiol immunnobiol 6:56–57
6. Loenen WAM, Dryden DTF, Raleigh EA et al (2014a) Highlights of the DNA cutters: A short history of the restriction enzymes. Nucleic Acids Res 42:3–19. https://doi.org/10.1093/nar/gkt990
7. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG (2014b) Type I restriction enzymes and their relatives. Nucleic Acids Res 42:20–44. https://doi.org/10.1093/nar/gkt847
8. Mccarthy A (2010) Third generation DNA sequencing: Pacific biosciences' single molecule real time technology. Chem Biol 17:675–676
9. Mise K, Nakajima K (1984) Isolation of restriction enzyme EcoVIII, an isoschizomer of *Hin*dIII, produced by *Escherichia coli* E1585-68. Gene 30:79–85. https://doi.org/10.1016/0378-1119(84)90107-0
10. Okonechnikov K, Golosova O, Fursov M et al (2012) Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics 28:1166–1167

11. Pingoud A, Fuxreiter M, Pingoud V, Wende W (2005) Type II restriction endonucleases: Structure and mechanism. Cell Mol Life Sci 62:685–707
12. Pingoud A, Jeltsch A (2001) Structure and function of type II restriction endonucleases. Nucleic Acids Res 29:3705–3727
13. Pingoud A, Wilson GG, Wende W (2014) Type II restriction endonucleases–a historical perspective and more. Nucleic Acids Res 42:7489–7527. https://doi.org/10.1093/nar/gku447
14. Roberts RJ (1976) Restriction endonucleases. CRC Crit Rev Biochem 4:123–164
15. Sistla S, Rao DN (2004) S-Adenosyl-L-methionine-dependent restriction enzymes. Crit Rev Biochem Mol Biol 39:1–19
16. Tamura K, Stecher G, Peterson D et al (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197
17. Watanabe N, Takasaki Y, Sato C et al (2009) Structures of restriction endonuclease *Hin*dIII in complex with its cognate DNA and divalent cations. Acta Crystallogr Sect D Biol Crystallogr 65:1326–1333. https://doi.org/10.1107/S0907444909041134
18. Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173:697–703. https://doi.org/10.1128/jb.173.2.697-703.1991
19. Williams RJ (2003) Restriction endonucleases: classification, properties, and applications. Mol Biotechnol 23:225–243
20. Wilson GG, Murray NE (1991) Restriction and modification systems. Annu Rev Genet 25:585–627. https://doi.org/10.1146/annurev.ge.25.120191.003101
21. Yang J, Roy A, Zhang Y (2013) BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res 41:. https://doi.org/10.1093/nar/gks966
22. Yang J, Zhang Y (2015) I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Res 43:W174–W181. https://doi.org/10.1093/nar/gkv342

Figures

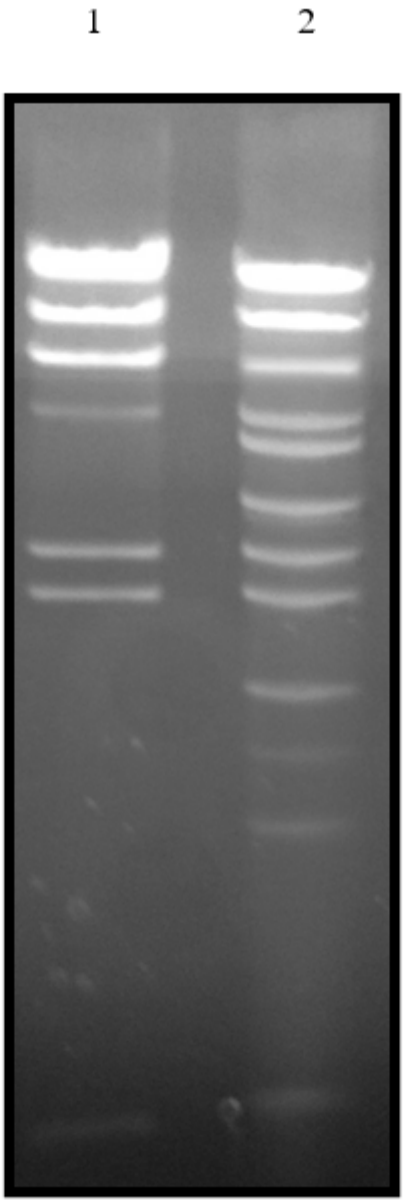

Figure 1

Screening of MatS1 for REs. Lane 1: Lambda DNA/ HindIII marker. Lane 2: Lambda DNA fragments produced by the REs in MatS1

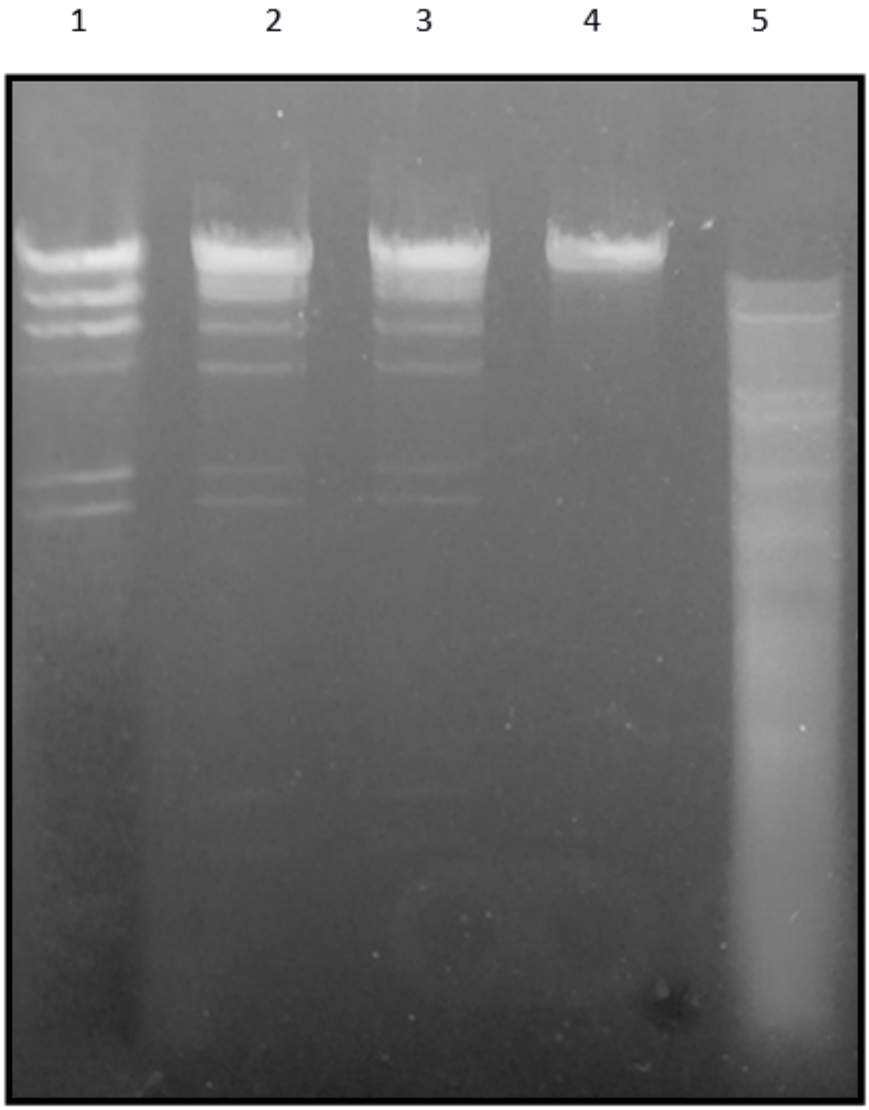

Figure 2

Partial purification of PanI. Lane 1: Lambda DNA/HindIII marker, Lane2: Restriction digestion of λ DNA treated with MatS1 lysate, concentrated using a 30 kDa molecular weight cut off filter. Lane 3: λ DNA cleaved using the filtrate of the 50 kDa molecular weight cut off filter. Lane 4: λ DNA cleaved using the filtrate of the 30 kDa molecular weight cut off filter. Lane 5: Lambda DNA cleaved with concentrated lysate (retentate) using a 50 kDa molecular weight cut off filter.

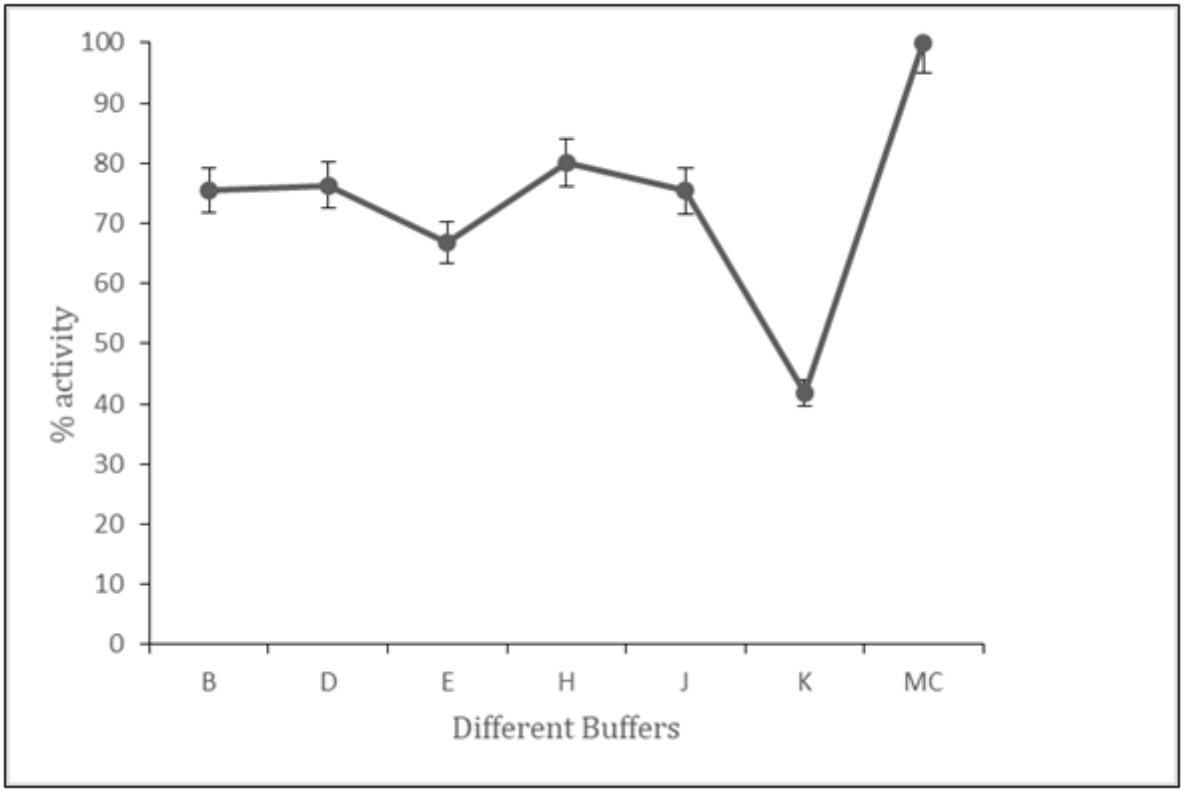

Figure 3

Determination of the activity of partially purified PanI in different commercially available buffers (Table 1). Percentage (%) activity of the partially purified PanI in different commercially available buffers was determined based on the intensities of the cleaved DNA fragments resolved on a 0.8 % agarose gel and analysed by image J software. Error bars represent SD (n=3)

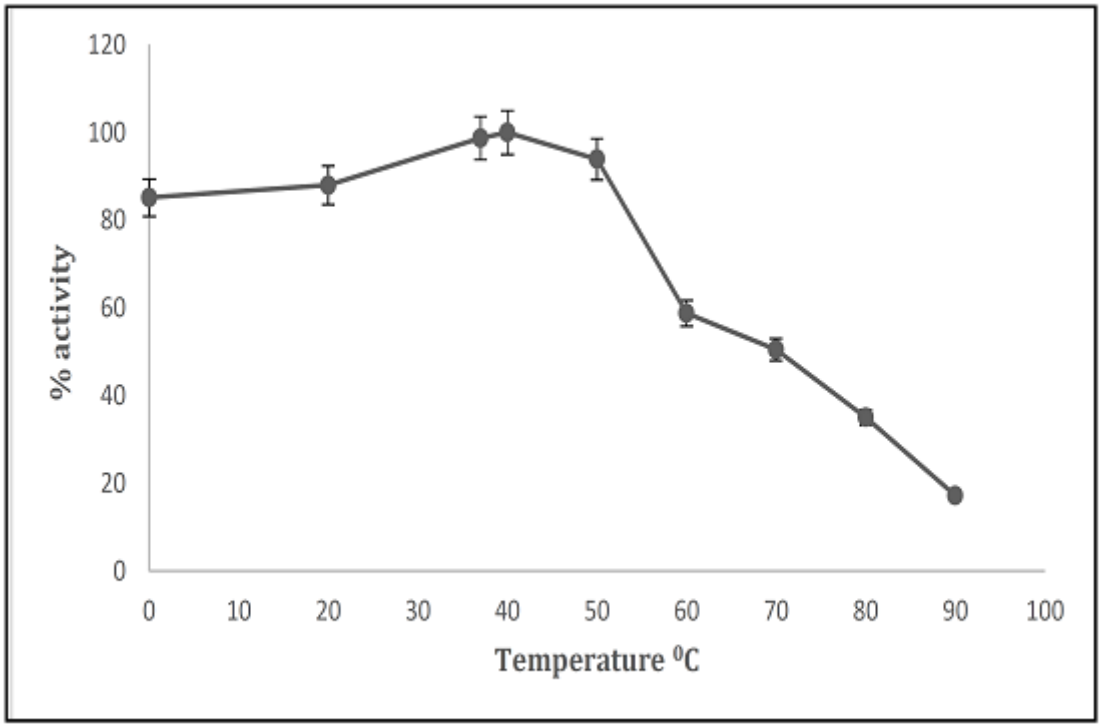

Figure 4

Determination of optimum temperature of partially purified PanI. Percentage (%) activity of partially purified enzyme at different temperatures varying from 0-90 oC was determined by analysing the intensities of the cleaved DNA fragments resolved on a 0.8 % agarose gel using image J software, Error bars represent SD (n=3)

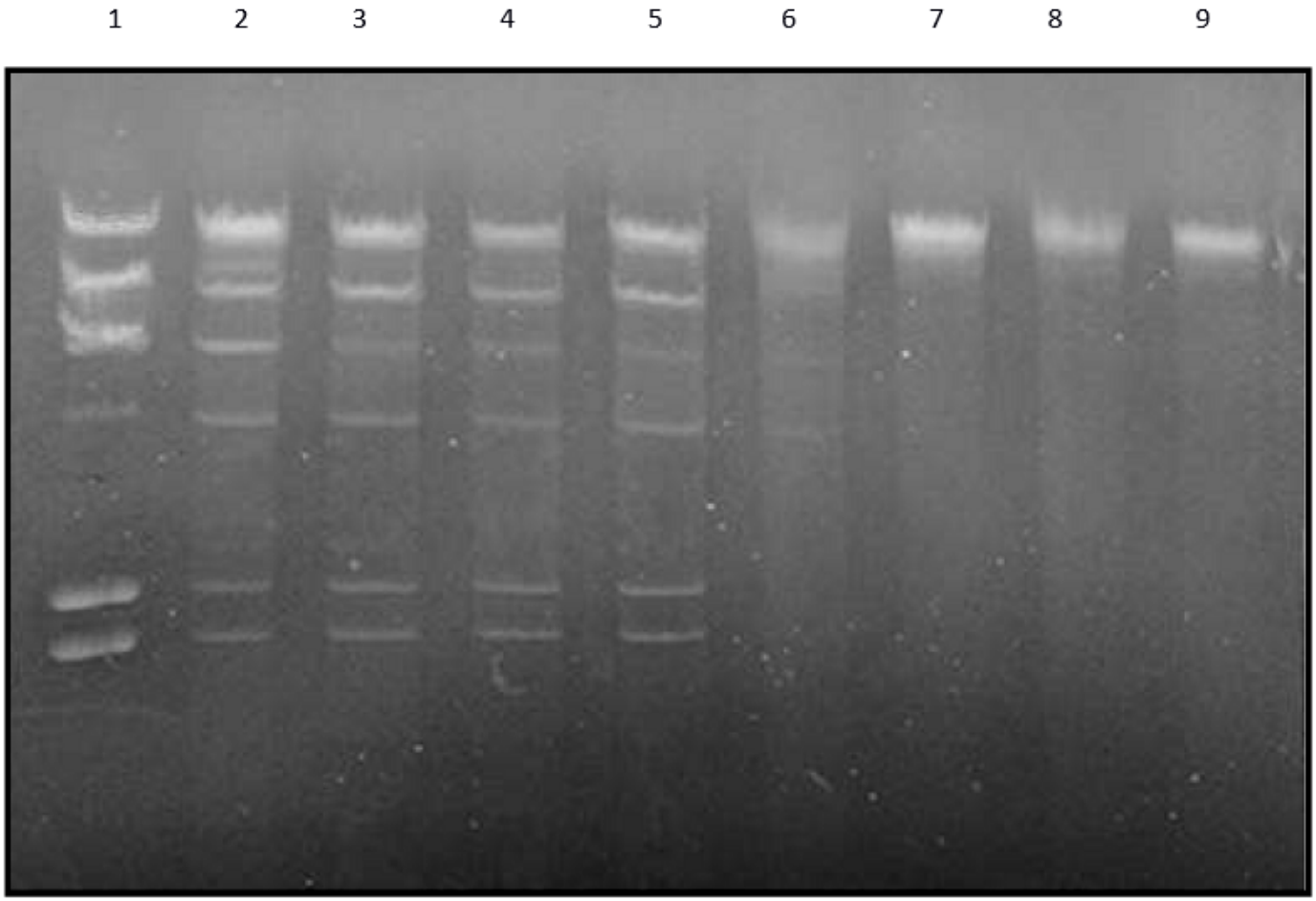

Figure 5

Determination of thermal stability of the partially purified enzyme. Lane 1: λ/HindIII marker, Lanes 2 to 9; fragments of λ DNA produced by restriction digestion reactions performed at 40 0C using partially purified PanI, which was pre-incubated at 0, 20, 37, 40, 50, 60, 70, 80 and 90 0C for 20 minutes, respectively

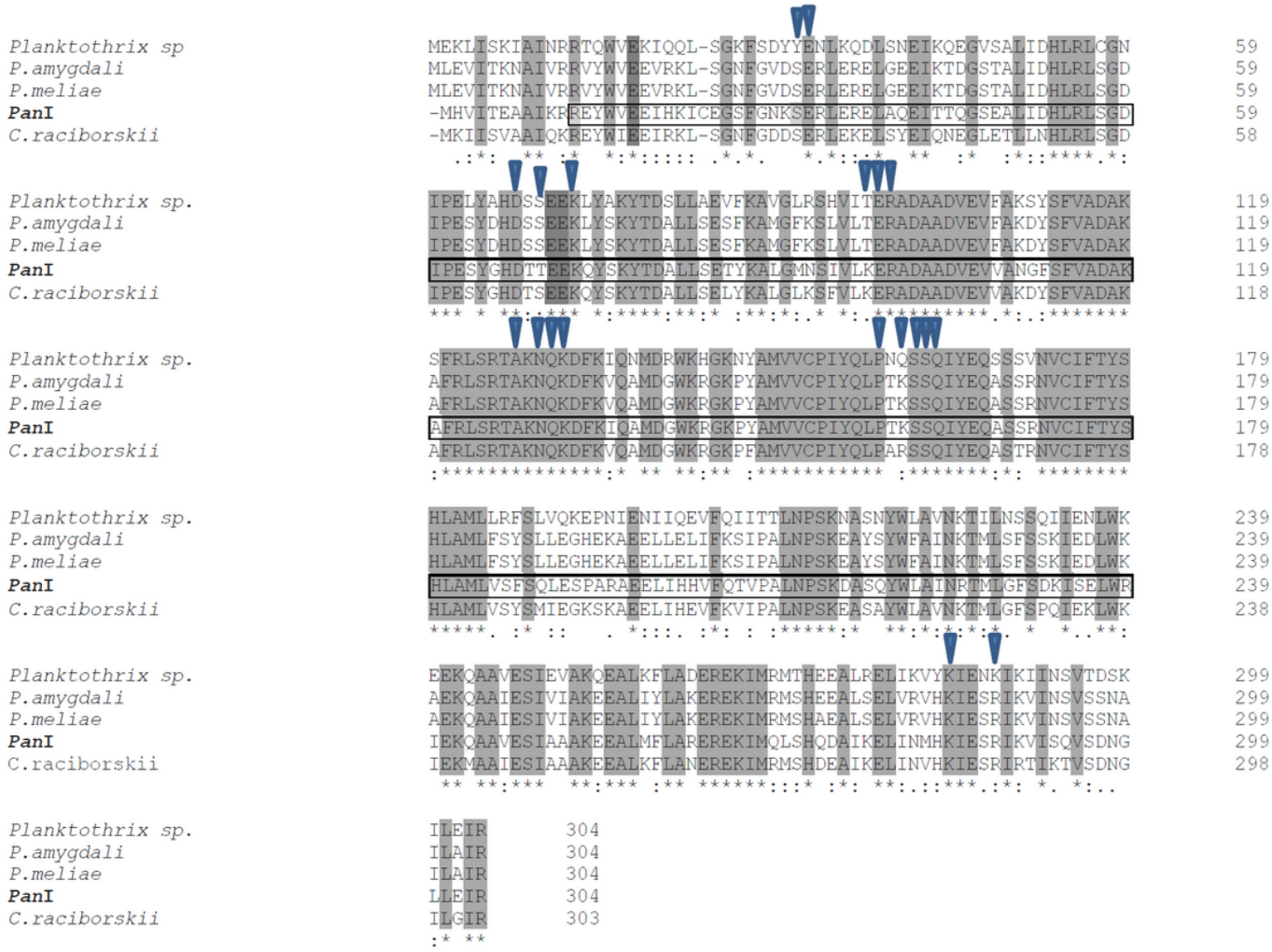

Figure 6

Multiple protein sequence alignment of selected bacterial HindIII like REs including PanI. Completely conserved and partially conserved residues are denoted by (*)/ highlighted in gray and (:) symbols respectively. Hind III superfamily signature is boxed on MatS1 HindIII. Blue color arrows depict the predicted substrate-binding sites.

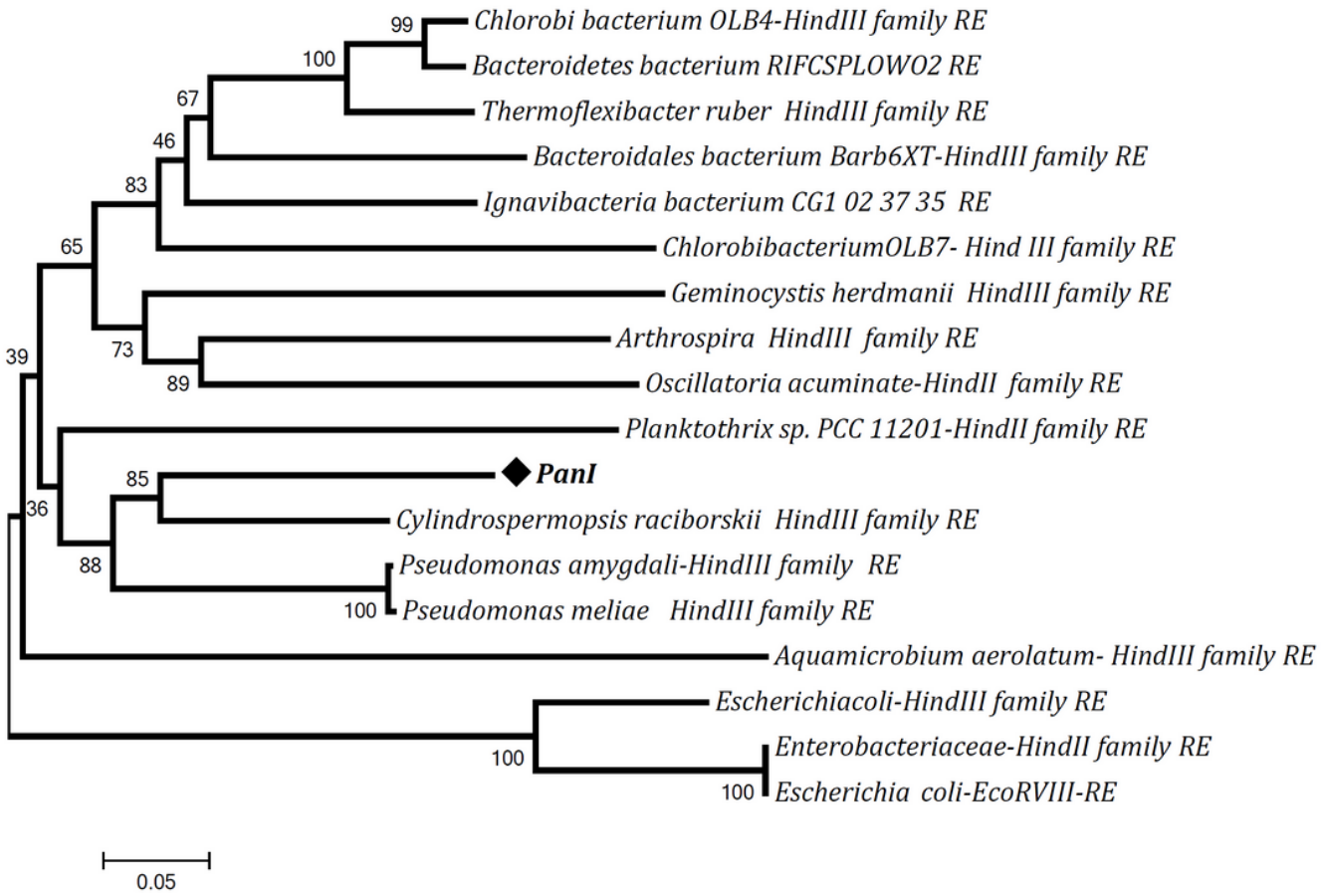

Figure 7

Phylogenetic position of PanI was analyzed using MEGA version 6.0 software based on Clustal omega multiple sequence alignment of different bacterial and cyanobacterial HindIII like REs under the neighbor-joining platform. Bootstrap supporting values are denoted at the tree branches and NCBI-GenBank accession numbers of used homologs are listed in Table 2

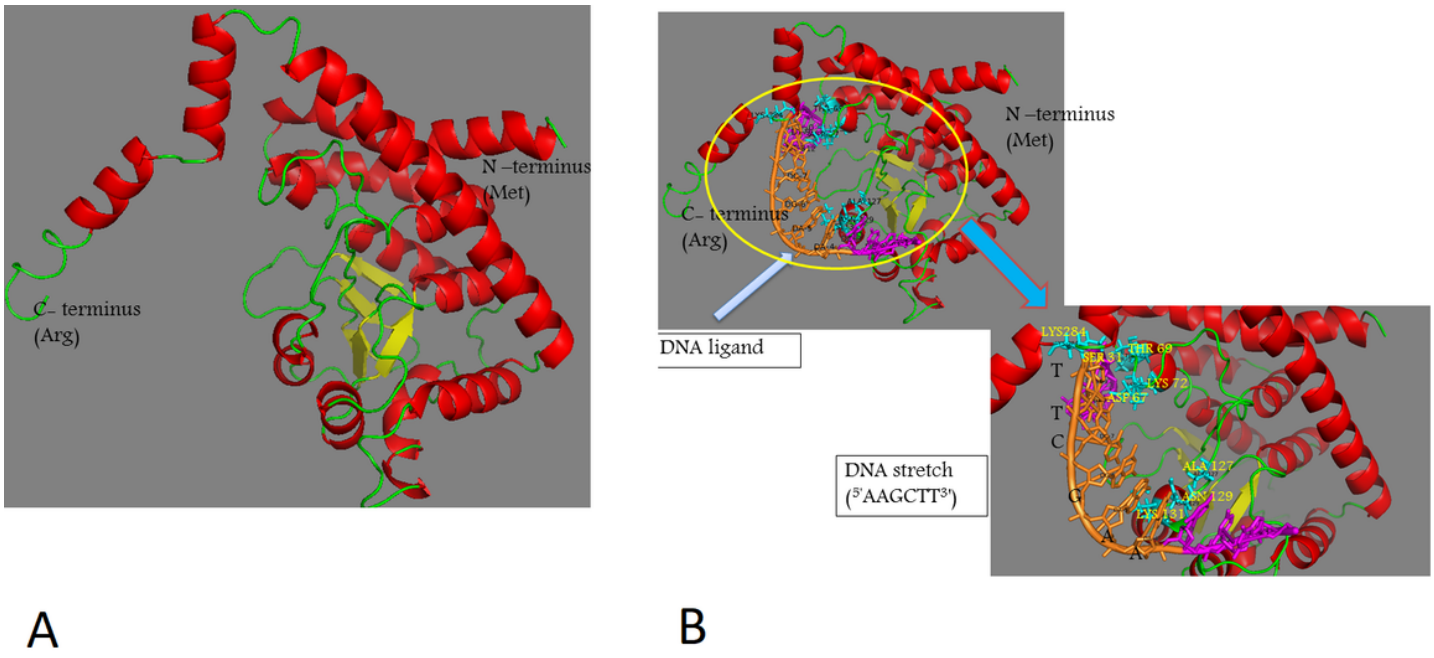

Figure 8

In silico structural and functional annotation of PanI using I-TASSER online server (A) Tertiary structure of PanI predicted by I-TASSER server and visualized by PyMoL v 3.1 software. α-Helices, β –strands and Loops/coil structures are denoted in red, yellow and green, respectively. (B) The predicted complex of PanI-cognate DNA. An enlarged version of active site of the predicted overall structural complex in the top panel is indicated in the bottom panel of the figure. Side chains of the residues in the active site are depicted as stick models in cyan and labeled. The cognate DNA stretch is shown in orange and violet, to indicate sites harboring the recognition sequence (5'AAGCTT3') and rest of the sequence, respectively