



30(3): 33-62, 2021; Article no.IJBCRR.70674 ISSN: 2231-086X, NLM ID: 101654445

An Overview of the Proofreading Functions in Bacteria and in Severe Acute Respiratory Syndrome-Coronaviruses

Peramachi Palanivelu^{1*}

¹Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai – 625 021, India.

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/IJBCRR/2021/v30i330258 <u>Editor(s)</u>: (1) Prof. Cheorl-Ho Kim, Sungkyunkwan University, South Korea. (2) Dr. Chunying Li, Georgia State University, USA. (3) Dr. Muhammad Farhan Jahangir Chughtai, NUR International University, Pakistan. <u>Reviewers</u>: (1) Lea Spindler, Univerza V, Mariboru Rektorat, Slovenia. (2) James Mageto, Kenyatta University, Kenya. (3) Signe Christensen, Lund University, Sweden. (4) Talha Bin Emran, BGC Trust University, Bangladesh. Complete Peer review History: <u>https://www.sdiarticle4.com/review-history/70674</u>

Original Research Article

Received 25 May 2021 Accepted 30 July 2021 Published 06 August 2021

ABSTRACT

Aim: To understand the structure-function relationship of the proofreading (PR) functions in eubacteria and viruses with special reference to Severe Acute Respiratory Syndrome-Coronaviruses (SARS-CoVs) and propose a plausible mechanism of action for PR exonucleases of SARS-CoVs.

Study Design: Bioinformatics, biochemical, site-directed mutagenesis (SDM), X-ray crystallographic data were used to study the structure-function relationships of the PR exonucleases from bacteria and CoVs.

Methodology: The protein sequences of the PR exonucleases of various DNA polymerases, and RNA polymerases of SARS, SARS-related and human CoVs (HCoVs) were obtained from PUBMED and SWISS-PROT databases. The advanced version of Clustal Omega was used for protein sequence analysis. Along with the conserved motifs identified by the bioinformatics analysis, the data already available by biochemical, SDM experiments and X-ray crystallographic analysis on these enzymes were used to arrive at the possible active amino acids in the PR exonucleases of these crucial enzymes.

Results: A complete analysis of the active sites of the PR exonucleases from various bacteria and CoVs were done. The multiple sequence alignment (MSA) analysis showed many conserved amino acids, small and large peptide regions among them. Based on the conserved motifs, the PR exonucleases are found to fit broadly into two superfamilies, viz. DEDD and polymerase-histidinol phosphatase (PHP) superfamilies. The bacterial DNA polymerases I and II, RNase D, RNase T and ε-subunit of DNA polymerases III belong to the DEDD superfamily. The PR enzymes from SARS, SARS-related CoVs and other HCoVs also essentially belong to the DEDD superfamily. The DEDD superfamily either uses an invariant Tyr or a His as proton acceptor during catalysis. Depending on the proton acceptor, they are further classified into DEDHD and DEDYD subfamilies. RNase Τ, εsubunit of DNA polymerases III and the SARS, SARS-related CoVs and other HCoVs belong to DEDHD subfamily. However, the SARS, SARS-related CoVs and other HCoVs showed additional zinc finger motifs (ZFMs) in their active sites. DNA polymerases I, II and RNase D belong to DEDYD subfamily. The bacterial DNA polymerases X, YcdX phosphoesterases and the co-editing exonuclease of DNA polymerases III belong to the PHP superfamily. Based on the MSA, X-ray crystallographic analyses and SDM experiments, the proposed active-site proton acceptor is Tyr/His in DEDDY/H subfamilies and His in PHP superfamily of PR exonucleases.

Conclusions: Based on the similarities of active site amino acids/motifs, it may be concluded that the DEDD and PHP superfamilies of PR exonucleases should have evolved from a common ancestor but diverged very long ago. The biochemical properties of these enzymes, including the four conserved acidic amino acid residues in the catalytic core, suggest that the CoVs might have acquired the exonuclease function, possibly from a prokaryote. However, the presence of two zinc fingers in the PR active site of the SARS, SARS-related CoVs and other HCoVs sets their PR exonucleases apart from other homologues.

Keywords: Proofreading exonucleases; coronaviruses; SARS-CoVs; DNA polymerases; RNase D, RNase T; DNA polymerase X; ExoN active site; ExoN catalytic Mechanism.

1. INTRODUCTION

Maintenance of genome stability is very important for all living organisms and relies mainly on the DNA and RNA polymerases which replicate the genomes. They replicate the genomes faithfully and thus, preserve and maintain the blueprint of life in all living cells. An in-depth analysis of these crucial catalysts of life, not only reveal fundamental information about their emergence, but also on the evolution of life on earth. Interestingly, not only the living cells but also the non-living entities like DNA and RNA viruses also possess these important enzymes. The DNA and RNA polymerases exhibit strong discrimination for NTPs and dNTPs and rarely insert a wrong nucleotide during replication of the genomes and hence the error rate in DNA or RNA synthesis is very, very minimum and is usually in the order of $\sim 10^{-6} - 10^{-9}$ and $\sim 10^{-4}$ to 10^{-6} , respectively. Even one mistake in critical areas is detrimental to the survival of organisms. Therefore, these crucial enzymes are invariably associated with a PR mechanism to correct any insertion error(s) during genome replication. These PR exonucleases belong to $3' \rightarrow 5'$ types, and they excise any wrongly added nucleotide from the 3'-growing end, and thus, helping the polymerases to perform error-free genome

replication. When a mismatch is encountered by the DNA or RNA polymerases during replication, the polymerases stall/pause, which in-turn activates the PR function which promptly excises the mismatch. Following the excision of the wrong base, the correct base is inserted and replication proceeds. This important PR step in living organisms ensures the original DNA/RNA template is copied without any mistake and passed on to the next generation. These PR enzymes are located either as a part of the replicase on the same polypeptide as a multifunctional enzyme (MFE) or as an independent subunit of a multienzyme complex (MEC). For example, in bacterial DNA polymerases I, three different enzymes are found on a single polypeptide as three distinct domains and exhibit three different activities, viz. i) polymerization, ii) proofreading and iii) DNA repair. The second type of PR exonucleases exists as an independent subunit of a multienzyme complex (MEC), e.g., *ɛ*-subunit of the bacterial DNA polymerases III (also known as replicases) [1 and references therein, 2]. To have a holistic view on these important PR enzymes in biological systems, including the one from the SARS, SARS-related and HCoVs an overview of these enzvmes is presented in this communication.

1.1 PR Functions in Biological Systems

The PR exonucleases are an important class of exonucleases. They are ubiquitous in biological systems and are reported from viruses, bacteria. fungi, plants, animals, etc. The PR function is not only associated with nucleic acid polymerases like DNA and RNA polymerases, but also associated with other nucleic acid modifying enzymes. Based on the active site amino acids, they are broadly classified into two superfamilies. viz. DEDD and PHP [2,3] Most of the bacterial and CoV DNA/RNA polymerases-associated PR exonucleases use four acidic amino acids, DEDD, for metal-binding and catalysis and hence belong to the DEDD superfamily whereas the DNA polymerases X, DNA polymerase III co-YcdX exonuclease editing [4] and phosphoesterases [5] use essentially His residues for metal-binding and catalysis. In this communication, the PR exonucleases belonging to the two different superfamilies are discussed in detail.

1.1.1 PR Exonucleases of DEDD and PHP superfamilies

The DEDD superfamily consists of two subfamilies, viz. DEDDv and DEDDh, depending upon whether they employ an invariant Y or a H as the proton acceptor during catalysis [2]. At least three different DNA polymerases involve in DNA repair and replication processes in prokaryotes. They are DNA polymerase I (encoded by polA), DNA polymerase II (encoded by polB) and DNA polymerase III, (a MEC). The DNA polymerases I, II and RNase D, belong to the DEDDY subfamily whereas the proofreading ε-subunit of the DNA polymerase III, RNase T, Exons of the RNA-dependent RNA polymerases (RdRps) of SARS, SARS-related CoVs and other HCoVs belong to the DEDDH subfamily. These two subfamilies and the PHP exonuclease superfamily are analyzed and discussed in detail.

2. MATERILAS AND METHODS

The protein sequences of the PR exonucleases of various DNA polymerases, RNases D, RNases T and RNA polymerases of SARS, SARS-related and human CoVs (HCoVs) were obtained from PUBMED and SWISS-PROT databases. The advanced version of Clustal Omega was used for protein sequence analysis. Along with the conserved motifs identified by the bioinformatics analysis, the data already available by biochemical, SDM experiments and X-ray crystallographic analysis on these enzymes were used to arrive at the possible active amino acids in the PR exonucleases.

3. RESULTS AND DISCUSSION

The PR function in eubacterial DNA polymerases I exists as an independent domain on the same polypeptide. The DNA polymerase I of *E. coli* is studied in great detail [1 and references therein]. It is a MFE and consists of three enzymes viz. i) 5'-3' exonuclease (DNA repair function), ii) 3'-5' exonuclease (PR function) and iii) DNA polymerase and are located in three independent domains of the same polypeptide (Fig. 1). The last two domains are also known as Klenow polymerase, and the distance between them is found to be ~30 Å.

Based on the conserved active site amino acids. this PR exonuclease is classified into DEDDy type [2]. Fig. 2 shows the MSA of the DNA polymerases I from different bacteria. Only the PR domains, from amino acids from 324-517, (numberings from the E. coli DNA polymerases I) and the polymerase region are shown here. (The E. coli enzyme is highlighted in yellow and the possible metal-binding sites are highlighted in green). The MSA analysis shows that the PR domain is almost completely conserved in all bacteria, except for few minor variations. At least four metal-binding sites are observed, and all the four are found in the completely conserved blocks. Two of the metal-binding sites, viz. -DTE- (355-357) and -DAD- (501-503), were proved to be essential for PR exonuclease activity by SDM and X-ray crystallographic analyses in the E. coli DNA polymerases I (marked in red) [1 and references therein]. Interestingly, the -DTE- site is found to be a fusion site where two -DXD- types of motifs are -FDTETDS- (where fused as one single site the D is replaced by an equivalent amino acid E) and both the motifs are followed by a hydroxyl amino acid, T or S and the two Ts lie in-between. The typical -DEDD- with 4 invariant acidic amino acids is found in all the 3'-5' exonucleases from bacterial DNA polymerases I and highlighted. The invariant -YA- template-binding pair (highlighted in yellow) suggests that they are all strictly template-dependent enzymes. The PR exonuclease with the four invariant acidic amino pattern identifiable acids with an DEDD $DxE \rightarrow D \rightarrow Y \rightarrow D$ belongs to the exonuclease superfamily and to the dnaQ-Y subfamily (Fig. 2).





Clustal (O) of PR 3'-5' Exonucleases of bacterial DNA polymerases I (only the PR exonuclease and polymerase regions are shown).



// End of the polymerase I		
tr A0A514EYQ8 A0A514EYQ8 9ENTR	HELMENSTTLAVPLLVEVGSGENWDQAH	921
tr A0A1V0LLD5 A0A1V0LLD5_9ENTR	HELMESSTTLDVPLLVEVGSGQNWDQAH	921
tr A0A1C4GJM7 A0A1C4GJM7_9ENTR	HELMENSTTLDVPLLVEVGSGQNWDQAH	921
tr F0JW51 F0JW51_ESCFE	HQLMENCTRLDVPLLVEVGSGENWDQAH	921
tr V0Y9R8 V0Y9R8_ECOLX	HQLMENCTRLDVPLLVEVGSGENWDQAH	921
tr I6DHG2 I6DHG2_SHIBO	HQLMENCTRLDVPLLVEVGSGENWDQAH	921
sp P00582 DP01_ECOLI	HQLMENCTRLDVPLLVEVGSGENWDQAH	921
tr A0A3P6LP28 A0A3P6LP28_SHIDY	HQLMENCTRLDVPLLVEVGSGENWDQAH	921
sp Q9F173 DPO1_SALTY	HQLMENCTRIDVPLLVEVGSGENWDQAH	921
tr A0A2I8SCZ7 A0A2I8SCZ7_9ENTR	HQLMENETQIDVPLLVEVGSGENWDQAH	92
	* • * * * • * * * * * * * * * * * * * *	

Fig. 2. MSA of the PR 3'-5' Exonucleases of bacterial DNA polymerases I

A0A514EYQ8_9ENTR, Raoultella electrica A0A1C4GJM7_9ENTR, Kosakonia oryziphila V0Y9R8_ECOLX, Escherichia coli P00582[DP01_ECOLI, Escherichia coli (strain K12) Q9F173]DP01_SALTY, Salmonella typhimurium A0A1V0LLD5_9ENTR, Kosakonia radicincitans F0JW51_ESCFE, Escherichia fergusonii ECD227 I6DHG2_SHIBO, Shigella boydii A0A3P6LP28_SHIDY, Shigella dysenteriae A0A2I8SCZ7_9ENTR, Citrobacter freundii



Fig. 3. A Schematic diagram showing the subsites-A and -B of the PR exonuclease of E. coli DNA polymerase I. (Active sites amino acids are placed based on the crystallographic and SDM data)

Fig. 3B shows only the proposed amino acids at the exonuclease active site with a water molecule as the 4th ligand. (All the four acidic amino acids of the DEDD superfamily are shown here)

3.1 Active Site Analyses of the PR Exonuclease of DNA polymerase I

The PR exonuclease activity of the DNA polymerase I of *E. coli* is one of the most wellstudied enzymes in this class [6. 7]. The active site of the PR exonuclease of the *E. coli* DNA polymerase I was analysed by genetics and SDM experiments and also by crystallographic studies. It was found that the PR exonuclease active site (EAS) essentially consisted of two sites, viz. a dNMP site and a metal-binding site. Therefore, dNMPs could inhibit the exonuclease reaction by product inhibition. The metal-binding site consisted of two subsites, viz., subsite-A and subsite-B and thus, EAS can bind two divalent metal ions. The presence of two divalent metal ion binding sites was further confirmed by anomalous scattering difference Fourier analysis of the wild-type enzyme with the ligands [6].

The subsite-A is coordinated by three amino acids, viz., Asp^{355} , Glu^{357} and Asp^{501} and the dNMP-phosphate provides the fourth ligand. Usually a Zn²⁺ is associated to the subsite-A. The second metal-binding site, subsite-B, is mainly coordinated by Asp^{424} and to the divalent metal ion Mg^{2+} . The subsite-B is located between

dNMP-phosphate and the carboxylate of Asp⁴²⁴ (Fig. 3A). The Zn²⁺ binding subsite-A was found to be very close to the 3' O-of the susceptible bond to be cleaved, and the Mg²⁺ binding subsite-B is very close to subsite-A. X-ray crystallographic data showed that the distance between the two metal atoms is ~ 3.9 Å in *E. coli* PR exonuclease (Fig. 3A) [6].

Further insights into the amino acids that constitute the EAS, were provided by SDM experiments by Joyce and Steitz [8].

- a) In a double mutant with Asp³⁵⁵→Ala and Glu³⁵⁷→Ala, both the dNMP binding site and the metal-binding site A (Zn²⁺) were completely abolished. This mutant protein had lost the exonuclease activity, but exhibited the polymerase activity. This suggested that the dNMP site is coordinating by both the metal-binding sites (Fig. 3A) [8].
- b) In the second SDM experiment, the Asp⁴²⁴ was replaced by Ala (Asp⁴²⁴ \rightarrow Ala). In this mutant enzyme, the metal-binding site B (Mg²⁺) was abolished and exhibited no exonuclease activity (the mutant protein, D⁴²⁴ \rightarrow A, did not bind to the metal ion in subsite B). However, in this mutant enzyme also the polymerase activity was found to be preserved [8].

These data suggest that the metal ions play a direct role during PR activity. The SDM studies have further shown that both the metal-binding sites are functionally connected and in the absence of one, the other cannot function. The Zn²⁺ binding site possibly involves in catalysis and the Mg-binding site, bind to dNMPphosphate and link the dNMP site. The Zn^{2+} ligands, viz. $D^{355},\ E^{357},\ D^{501}$ with a water molecule were found within ~ 2.0 Å distances [8]. Furtehrmore, substitution of Asp³⁵⁵ and Glu³⁵⁷ in the E. coli polymerase I yielded an enzyme devoid (<0.01% remaining) of exonuclease activity, while retaining its overall structure [6]. The Tyr⁴⁹⁷ is placed as the proton acceptor as it is a completely conserved in the highly conserved block and is in the equivalent position to the His¹⁶² of the PR exonuclease of the DNA polymerase III ε-subunit (dnaQ-H family). Furthermore, the PR exonuclease of DNA polymerase I belongs to dnaQ-Y family where an invariant Y is proposed to involve in deprotonation of water molecule similar to the H¹⁶², an invariant amino acid in the dnaQ-H family performing the same function.

A nucleophilic attack on the phosphorous atom of the terminal nucleotide is postulated to be carried out by a hydroxide ion that is activated by one the divalent metal, while expected pentacoordinate transition state and the leaving oxyanion are stabilized by a second divalent metal ion that is placed 3.9 Å away from the first metal ion [9]. Of particular importance is the mutant protein $D^{424} \rightarrow A$, which showed no measurable exonuclease activity. Not only this implies an important catalytic role for the metal ion B, but it also allows the preparation of a stable complex with a single-stranded DNA substrate. Therefore, it was concluded that the chemical catalysis of the hydrolytic phosphoryl transfer reaction is promoted by the two metal ions and a water molecule which is coordinated to the Zn^{2+} [6].

Furthermore, the pH dependence of the 3'-5' exonuclease reaction is consistent with a mechanism in which nucleophilic attack on the terminal phosphodiester bond is initiated by a hydroxide ion coordinated to one of the enzymebound metal ions [7]. The properties of the mutant proteins suggest that one metal ion plays a role in substrate binding while the other is involved in catalysis [6]. It is interesting to note that a complete absence of any C in this PR domain suggests that the Zn²⁺ is not coordinated by Cs as reported in other PR exonucleases elsewhere. discussed Based on the crystallographic data and SDM analysis, the proposed amino acids at the PR exonuclease active site is shown in Fig. 3B

3.2 PR Function in DNA Polymerases II (DEDYD)

The second enzyme that shows an intrinsic PR exonuclease function is the bacterial DNA polymerases II. The DNA polymerase II from E. *coli* is one of the most well-studied enzyme among this class [1 and references therein, 7]. It is encoded by polB gene and consists of 783 amino acids with a molecular mass of ~90 kDa. It is a member of the Family B DNA polymerases, or otherwise known as repair polymerases. Like DNA polymerase I, it exists as a monomer and the catalytic core consists of the typical structural domains, viz. palm, fingers, and thumb as reported in other DNA/RNA polymerases. The enzyme exhibits both $5' \rightarrow 3'$ DNA synthesis and $3' \rightarrow 5'$ PR exonuclease activity. The DNA polymerases II can extend primers in a variety of lesions, which is known as translesion synthesis. Wang and Yang [10] have shown that the amino acids 147 to 367 comprise the 3'-5' PR exonuclease domain (highlighted in red) and 368 to 783 involve in polymerase function in *E. coli* polymerase II. As the $D^{335} \rightarrow N$ mutant lost its exonuclease activity, it is implicated in the catalysis. The MSA analysis shows that the exonuclease domain is highly conserved in all

bacterial DNA polymerases II (The *E. coli* sequence is highlighted in yellow). The PR exonuclease contains the four invariant acidic amino acids with an identifiable pattern $-DxE \rightarrow D \rightarrow Y \rightarrow D$ - and hence belongs to the DEDD exonuclease superfamily and to dnaQ-Y subfamily (Fig. 4).

CLUSTAL O (1.2.4) MSA of DNA Polymerases	II (only the PR exonuclease and polymerase active
site regions are shown)	
-	

+rla0a381c280la0a381c280 cTTAM	TO DUMUECETRNCA TUNARI KPHPDVR	PPLKWLS	DIFTERRETIVCTCLECCORTVVM	180
	I SE V WVEGETRNGAT VNARERE DIR	FF LIWLS.	IDIEITKIIGEETCIGEEGCGQKIVIM	100
tr A0A482PUL1 A0A482PUL1_CITRO	TAPVWVDGEARDGALVNARLKPHPDYR	PPLRWLS.	DIETTRHGELYCIGLEGCGQRIVYM	180
tr A0A7G1PJC3 A0A7G1PJC3 CITKO	TAPVWLEGDTKDGAIVNARLKPHPDYR	PPLKWVS	DIETTRHGELYCIGLOGCGORVVFM	180
+ x 1 3 0 3 7 3 4 UT 4 6 1 3 0 3 7 3 4 UT 4 6 S 3 T F B	TO DUMT DO EMPNOUTONA DI KOUDOVO	DDTKMUC	DIFTERRELICET VOTOT ECCOPTOVM	190
L ISOS SECTION STATUS	TOT WITDOBHING VIRWARDERT HTDIR	DDIWWW		100
tr AUA3 / 90 F.F.0 AUA3 / 90 F.F.0 SALER	TSPVWIDGEMRNGVIRNARLKPHPDYR	PPLKWVS.	DIELLEHGEFICIGFEGCGÖKLAIW	180
tr A0A2B7LUW5 A0A2B7LUW5 9ESCH	TAPVWVDGDFRDGAIVNARLKPNPDYR	PPLKWVS	DIETTRHGELYCIGLEGCGORIVYM	180
+r1202200186612022001866 SHTSO	TORWWECOMHNOTIVNARIKPHPDVR	PPLKWVS	DIFTERRET VOIGLECCORTVVM	180
	101 VWVEGEEIINGTI VWAREERT HE DIR	I I DIGWVO		100
tr A0A2S8DDW7 A0A2S8DDW7_SHIDY	TSPVWVEGDMHNGTIVNARLKPHPDYR	PPLKWVS.	DIETTRHGELYCIGLEGCGQRIVYM	180
sp P21189 DPO2_ECOLI	TSPVWVEGDMHNGTIVNARLKPHPDYR	PPLKWVSI	IETTRHGELYCIGLEGCGORIVYM	180 SDM
train a characteria characteria	THE DURING DATE OF THE PROPERTY OF THE PROPERT	DDIKENC	DIETERBUCEL VOICLECCCORTAVM	100
LT AUAONSKGUS AUAONSKGUS SHIFL	ISPVWVEGDMHNGIIVNARLKPHPDIR	PPLRWVS	DILITRIGELICIGLEGCGQRIVIM	100
tr B2U266 B2U266_SHIB3	TSPVWVEGDMHNGTIVNARLKPHPDYR	PPLKWVS	DIETTRHGELYCIGLEGCGQRIVYM	180
tr A0A7D6UTK4 A0A7D6UTK4 9ENTR	TSPVWVEGDMRNGAIVNARLKPHPDYR	PPLKWVS	DIETTRHGELYCIGLEGCGORIVYM	180
+*1303191NCD11303191NCD1 VIEOV	TO DUBUE COT DUDA TUNIA DI KOUDDVD	DDIKENC	DIETERDUCELYCICLECCCODIWYM	100
CT MONTOINSEI MONTOINSEI_KTEON	ISPVWVEGDIRNDAIVNARLKPHPDIR	PPLRWVS	DILITRIGELICIGLEGCGQRIVIM	100
tr B7LVT1 B7LVT1_ESCF3	TSPVWVEGDMHNGAIVNARLKPHPDYR	PPLKWVS	DIETTRHGELYCIGLEGCGQRIVYM	180
tr A0A7L6S4D3 A0A7L6S4D3 ESCFE	TSPVWVEGDMRNGAIVNARLKPHPDYR	PPLKWVS	DIETTRHGELYCIGLEGCGORIVYM	180
	* * * * * • * * * * * * * * * * * * * *	***•*•*	*********	
		L · · ·	• • •	
		▼ I		
			_	
+*1303201020013032010200 CTM3M	I CDENCOCON I DEOL EVUNCODOL LEVI		A T T CHNILL CRIPT DUL ORILAEDY	240
CI KOKSOIGZOO KOKSOIGZOO CIIKM	TGLENGD22MIDLÖTET MOKLÖTTEVI	NEW VAR IDEL	DITIGWIND VOT DIKV DOMIKERI	240
tr A0A482PUL1 A0A482PUL1_CITRO	LGPANGDASGLDFQLEYVTSRPQLLEKL	NEWIARHDPI)VIIGWNLV QFU LRMLQKHAERY	240
tr A0A7G1PJC3 A0A7G1PJC3 CITKO	LGPANGDASALDFQLEYAASRPLLLEKL	NDWFARHDPI	DVIIGWNVVOFILRVLOKHAERY	240
+ x 1 A 0 A 7 3 / UT / 6 1 A 0 A 7 3 / UT / 6 S A T F P			WTTCHNUM PTI DMIOKUAEDY	240
CI KOK/JAIII40 KOK/JAIII40 JALLEK	TGEWIGDDKÖTDLETAIAWSKEÖTTEVT	MAWE ABILDED	JATTOMIA A ALT TRUPOLITYPE L	240
tr A0A379QFF6 A0A379QFF6_SALER	LGPANGDDRQLDFELVYVASRPQLLEKL	NAWFAEHDPI	DVIIGWNVV QFI LRMLQKHAERY	240
trla0a2B7LUW5La0a2B7LUW5 9ESCH	LGPENGDASALNEDLEYVASRPOLLEKI.	NAWFATHOPI	OVIIGWNVVOFTLEMLOKHAEBY	240
				240
TT AUAZUULKG0 AUAZUULKG0_SHISU	LGPENGDASSLDFELEIVASKPLLLEKL	NAWFANHDPI	JVIIGWNVVQFILKMLQKHAERI	240
tr A0A2S8DDW7 A0A2S8DDW7 SHIDY	LGPENGDASSLDFELEYVASRPQLLEKL	NAWFANYDPI	DVIIGWNVV QFI LRMLQKHAERY	240
splP211891DP02_ECOLT	LGPENGDASSLDFELEYVASRPOLLEKU	VAWFANYDPI	OVITGWNVVOF LEMLOKHAERY	240 SDM
				240
TT AUAGNSRGJ9 AUAGNSRGJ9_SHIFL	LGPENGDASSLDFELEIVASRPQLLEKL	NAWFANIDPI	JVIIGWNVVQFLLKMLQKHAERY	240
tr B2U266 B2U266 SHIB3	LGPENGDASSLDFELEYVASRPQLLEKL	NAWFANYDPI	DVIIGWNVV QFI LRMLQKHAERY	240
trla0a7D6UTK4la0a7D6UTK4 9ENTR	LGPENGDASALDFELEYVASRPLLLEKL	NAWFATHDPI	OVITGWNVVOFTLEMLOKHAERY	240
				240
TT AUAISIWSPI AUAISIWSPI_KLEUX	LGPENGDASALDFELEIVASKPQLLEKL	NAWFANIDPI	JVIIGWNVVQFILKMLQKHAERI	240
tr B7LVT1 B7LVT1 ESCF3	LGPENGDASALDFELEYVASRPQLLEKL	NAWFANHDPI	DVIIGWNVV QFI LRMLQKHAERY	240
trla0a7L6s4D3la0a7L6s4D3_ESCFE	LGPENGDASALDFELEYVASRPOLLEKL	NAWFANHDPI	OVITGWNVVOFTLEMLOKHAERY	240
			pre se	210
		to also also also also also	 All also de also de la sola also de also de la de also de also de also de also de also 	
	*** *** *:*:* *. *** *****	* *.* :***	* : * * * * : * * * * * * * : * * * * *	
	*** *** *:*:* *: *** *****	* *.* :***	* * * * * * * * * <mark>* *</mark> * * • * * * * * * * * *	
trla0a381G2801a0a381G280_CTTAM	RIPLREGRONSELEWREHGEKNGVEFAO	* *.* :***	*:*****:** <mark>***</mark> **:*********************	30.0
tr A0A381G280 A0A381G280_CITAM	RIPLRFGRDNSELEWREHGFKNGVFFAQ	* *.* :***	* :***** :** *************************	300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVYFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDGJ	*:*****:** *** **:********* EEALKSAFWNFSSFSLETVSQEL EEALKSAFWNFSSFSLEAVSQEL	300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVYFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGI AKGRLIIDGI AKGRLIIDGI	*:*****:******************************	300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A734H746 A0A734H746_SALER	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAO RIPLRLGDNSELEWREHGFKNGVFFAO	* *.* :*** AKGRLIVDGI AKGRLIIDGI AKGRLIIDGI ARGRVIIDGI	* :***** :****************************	300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PU1 CITRO tr A0A7GPU1 A0A482PU1 CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A734HT46 A0A7G4HT46_SALER	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGI AKGRLIIDGI AKGRLIIDGI ARGRVIIDGI	*:************************************	300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A379QFF6 A0A379QFF6_SALER	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGI AKGRLIIDGI AKGRLIIDGI ARGRVIIDGI ARGRVIIDGI	*:******	300 300 300 300 300
tr A 0 A 3 8 1 G 2 8 0 A 0 A 3 8 1 G 2 8 0 _ CITAM tr A 0 A 4 8 2 P U 1 A 0 A 4 8 2 P U 1 _ CITRO tr A 0 A 7 G I P J 3 A 0 A 7 G I P J 3 _ CITKO tr A 0 A 7 3 4 H T 4 6 A 0 A 7 3 4 H T 4 _ SALER tr A 0 A 3 7 9 Q F 6 A 0 A 3 7 9 Q F 6 _ SALER tr A 0 A 2 P J U W 5 A 0 A 2 P J U W 5 9 E S C H	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGI AKGRLIIDGI AKGRLIIDGI ARGRVIIDGI AKGRLIIDGI	*:************************************	300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46 SALER tr A0A379QFF6 A0A379QFF6 SALER tr A0A287LUM5 A0A287LUM5 SECH tr A0A200LK66 A0A200LK66 SH1S0	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDGJ ARGRVIIDGJ AKGRLIIDGJ AKGRLIIDGJ	*:******	300 300 300 300 300 300 300
tr A 0 A 3 8 1 G 2 8 0 A 0 A 3 8 1 G 2 8 0 _ CITAM tr A 0 A 4 2 2 U L 1 A 0 A 4 8 2 P U L 1 _ CITRO tr A 0 A 7 G I P J C 3 A 0 A 7 G I P J C 3 _ CITKO tr A 0 A 7 3 4 H T 4 6 A 0 A 7 3 4 H T 4 _ SALER tr A 0 A 3 7 9 Q F 6 A 0 A 3 7 9 Q F 6 _ SALER tr A 0 A 2 0 L K G 6 A 0 A 2 7 J L W 5 _ 9 E SC H tr A 0 A 2 0 C L K G 6 A 0 A 2 0 C D L K G _ SU S Y	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDGJ AKGRLIIDGJ ARGRVIIDGJ AKGRLIIDGJ AKGRLIIDGJ	*:******	300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A379QFF6 A0A379QFF6_SALER tr A0A287LUW5 A0A287LUW5_9ESCH tr A0A287LUW5 A0A20LK66_SHISO tr A0A288DDW7 A0A288DDW7_SHIDY	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDGJ ARGRVIIDGJ ARGRVIIDGJ AKGRLIIDGJ AKGRLIIDGJ	*:*****: EALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PU11 A0A482PU11_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A379QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A200LKG6 A0A200LKG6_SHISO tr A0A288DDW7 A0A2S8DDW7_SHIDY sp P21189 DPO2_ECOLI	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGI AKGRLIIDGI ARGRVIIDGI ARGRVIIDGI AKGRLIIDGI AKGRLIIDGI AKGRLIIDGI	*:************************************	300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A379QFF6 A0A379QFF6 SALER tr A0A287LUW5 A0A287LUW5_9ESCH tr A0A280LUK5 A0A200LK56 SHISO tr A0A288DDW7 A0A288DDW7_SHIDY sp P21189 DPO2_ECOLI tr A0A8N3RG39 A0A6N3RG39 SHIFL	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDGJ ARGRVIIDGJ ARGRVIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ	*:************************************	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PU1 CITRO tr A0A7G1PU1 A0A482PU1 CITRO tr A0A7G1PJ3 A0A7G1PJ3 CITKO tr A0A734HT46 A0A734HT46 SALER tr A0A297LUN5 A0A297LUN5_9ESCH tr A0A297LUN5 A0A297LUN5_9ESCH tr A0A297LUN5 A0A297LUN5_9ESCH tr A0A298DDW7 A0A297LUN5_9ESCH tr A0A298DDW7 A0A298DDW7 SHIDY sp 221189 DPO2_ECOL1 tr A0A6N3RG39 A0A6N3RG39_SHIFL tr B01266 B21266 SHIB3	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNSVFFAQ RLPLRLGRDNSELEWREHGFKNSVFFAQ RLPLRLGRDNSELEWREHGFKNSVFFAQ RLPLRLGRDNSELEWREHGFKNSVFFAQ	* *.* :*** AKGRLIVGI AKGRLIIGJ ARGRVIIGG AKGRVIIGG AKGRLIIGG AKGRLIIGG AKGRLIIGG AKGRLIIGG AKGRLIIDG	*:************************************	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A761PJC3 A0A7G1PJC3_CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A29790FF6 A0A2970FF6 SALER tr A0A281LUW5 A0A287LUW5_9ESCH tr A0A200LKG6 A0A200LKG6 SHISO tr A0A288DDW7 A0A288DDW7_SHIDY sp P21189 DPO2_ECOLI tr A0A6N3RG39] A0A6N3RG39_SHIFL tr B20266 B20266 SHIB3	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIDGJ ARGRVIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ	*:******	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A422PU1 A0A482PU1 CITRO tr A0A72PU1 A0A7G1PJC3 CITRO tr A0A734HT46 A0A734HT46 SALER tr A0A3734HT46 A0A3730F76 SALER tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A288DDW7 A0A288DDW7 SHISO tr A0A288DDW7 A0A288DDW7 SHISO tr A0A683RG39 A0A683RG39 SHIFL tr B2U266 B2U266 SHIB3 tr A0A706UTK4 A0A706UTK4 9ENTR	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIDGJ ARGRVIDGJ ARGRVIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ	*:******	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A734H746 A0A734H746_SALER tr A0A734H746 A0A734H746_SALER tr A0A734H746 A0A734H746_SALER tr A0A29TLUW5 A0A29TLUW5_9ESCH tr A0A200LKG6 A0A200LKG6_SHISO tr A0A200LKG6 A0A200LKG6_SHISO tr A0A208DDW7 A0A288DDW7_SHIDY sp P21189 DPO2_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr A0A7D6UTK4 A0A7D6UTK4_9ENTR tr A0A181WSP1 A0A181WSP1 KLEOX	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDG AKGRUIDGJ AKGRVIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ	*:******	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A761PJC3 A0A7G1PJC3]CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A379QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DDW7_SHIDY sp 221189 DP02_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266 SHIB3 tr A0A7b6UTK4 A0A7D6UTK4_9ENTR tr A0A716UTK4 A0A7D6UTK4_9ENTR tr B0A716UTK4 DA181WSP1_KLEOX	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDG ARGRVIIDGJ ARGRVIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ VKGRLIIDGJ VKGRLIIDGJ	*:******	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A279CF6 A0A297LUW5_9ESCH tr A0A28DLUW5 A0A287LUW5_9ESCH tr A0A200LKC6 A0A200LKC6_SHISO tr A0A200LKC6 A0A200LKC6_SHISO tr A0A6N3RG39 A0A6N3RG39_SHIFL tr B2U266 B2U266 SHIS3 tr A0A7BUTK4 A0A7D6UTK4_9ENTR tr A0A181WSP1 A0A181WSP1_KLEOX tr B7LVT1 B7LVT1 ESCF3	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDG AKGRUIDGJ AKGRUIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ	*:******	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A734HT46 A0A7G4H745_SALER tr A0A279QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DDW7_SHIDY 39 P21189 DP02_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B20266 B20266_SHIB3 tr A0A7D6UTK4 A0A7D6UTK4_9ENTR tr A0A181WSP1 A0A181WSP1_KEOX tr B7LV1 B7LVT1_ESCT3 tr A0A7L6S4D3 A0A7L6S4D3_ESCFE	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGI AKGRLIDGI ARGRVIIDGI AKGRVIIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI	CALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A287LUW5 A0A2B7LUW5_9ESCH tr A0A20LK66 A0A200LK66_SHIS0 tr A0A200LK66 A0A200LK66_SHIS0 tr A0A288DDW7 A0A288DDW7_SHIDY sp P21189 DP02_ECOLI tr A0A6N3RG39 A0A6N3RG39_SHIFL tr B2U266 B2U266 SHIB3 tr A0A7B0UTK4 A0A7D6UTK4_9ENTR tr A0A181WSP1 A0A181WSP1_KLEOX tr B7LVT1 B7LVT1_ESCF3 tr A0A7L6S4D3 A0A7L6S4D3_ESCFE	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDG AKGRLIIDG AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A42PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A7G1PJC3 A0A7G1PJC3_CITKO tr A0A2B7LUM5 A0A2B7LUM5_9ESCH tr A0A2B7LUM5 A0A2B7LUM5_9ESCH tr A0A2B7LUM5 A0A2B7LUM5_9ESCH tr A0A2B7LUM5 A0A2S8DDW7_SHISO tr A0A2B8DDW7 A0A2S8DDW7_SHISO tr A0A2B8DDW7 A0A2S8DDW7_SHISO tr A0A2B61B2U26CS tr A0A7G601K4 A0A7D601K4_9ENTR tr A0A7G01K4 A0A7D601K4_9ENTR tr A0A7L6S4D3 A0A7L6S4D3_ESCFE	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* : : : : : : : : : : : : : : : : :	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A761PLC3 A0A761PLC3_CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A292LUN5 A0A379QFF6 SALER tr A0A292LUN5 A0A292LUN5]SECH tr A0A292LUN5 A0A292LUN5]SECH tr A0A202LKG6 A0A202LK66_SHISO tr A0A202LKG6 A0A202LK66_SHISO tr A0A202LKG6 A0A202LK66_SHISO tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266_SHIB3 tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266_SHIB3 tr A0A181WSP1 A0A181WSP1_KLEOX tr B7LVT1 B7LVT1_ESCF3 tr A0A7L6S4D3 A0A7L6S4D3_ESCFE	RI PLRFGRDNSELEWREHGFKNGVFFAQ RI PLRFGRDSELEWREHGFKNGVFFAQ RI PLRLGRDNTELEWREHGFKNGVFFAQ RI PLRLGRDNTELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ RI PLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIIDGJ AKGRVIIDGJ AKGRVIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDGJ AKGRLIIDG	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A279QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A2B7LUM5_9ESCH tr A0A287LUM5 A0A2B7LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DDW7_SHIDY mp P21189 DP02_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B20266 B20266_SHIB3 tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A16WSH1 A0A18UMSP1_KLEOX tr B7LV1 B7LVT1_ESCF3 tr A0A716S4D3 A0A7L6S4D3_ESCFE	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* :*** AKGRLIVDGJ AKGRLIDGJ ARGRVIDG AKGRUIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A761PLC3 A0A761PLC3_CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A379QFF6 A0A379QFF6_SALER tr A0A287LUW5 A0A287LUW5_9ESCH tr A0A287LUW5 A0A287LUW5_9ESCH tr A0A287LUW5 A0A287LUM5_9ESCH tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266_SHIB3 tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266_SHIB3 tr A0A71654D3 A0A71654D3_ESCFE tr A0A71654D3 A0A71654D3_ESCFE	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ	* *.* : : : : : : : : : : : : : : : : :	CALLSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSFSFSLETVAFWNFSFSFSLETVAQEL EALKSFFSFSFSLETVAFFSFSFSLETVAFWNFSFSFSLETVAFFSFSFSLETVAFFSFSFSLETVAFFSFSFSLETVAFFSFSFSFSLETVAFFSFSFSLETVAFFSFSFSFSFSLETVAFFSFSFSFSFSLETVAFFSFSFSLETVAFFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSF	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A279QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DDW7_SHIDY 39 221189 DP02_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266_SHIB3 tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A163WSP1 A0A18UMSP1_KEOX tr B7LV1 B7LVT1_ESCF3 tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ STATTATATATATATATATATATATATATATATATATAT	* *.* : : : : : : : : : : : : : : : : :	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CTRKO tr A0A7G1PJC3 A0A7G1PJC3_CTRKO tr A0A7G1PJC3 A0A734T46 SALER tr A0A287LUM5 A0A270LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DLW7 SHISO tr A0A638G20 A0A288DLW7 SHISO tr A0A638G20 A0A6N3RG39_SHIFL tr B2U266 B2U26 G_SHIB3 tr A0A764UX4 A0A764VA4_9ENTR tr A0A181WSP1 A0A181WSP1_KLEOX tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3_CTKKO	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ	* *.* : :*** AKGRLIVDGJ AKGRLIDGG AKGRVIDGJ AKGRVIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDG AKGRLIDG AKGRLIDG AKGRLIDG AKGRLIDG AKGRLIDG AKGRLIDG AKGRLIDG AKGRLIDG AKGRLIDG AYNLKDC AYNLKDC AYNLKDC AYNLKDC	CALLSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A2732UM5 A0A272UM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A760TA4A0A7040TK4_9ENTR tr A0A760TK4 A0A7040TK4_9ENTR tr A0A7160TK4 A0A7040TK4_9ENTR tr A0A71654D3 A0A71654D3_ESCFE tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A734HT46 A0A734HT46_SALER	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LIPLRLGRDNSELEWREHGFKNGVFFAQ LIPLRLGRDNSELEWREHGFKNGVFFAQ XIVIII X X X X X X X X X X X X X X X X X	* *.* : :*** AKGRLIVDGJ AKGRLIDGJ ARGRVIDGJ ARGRVIDGJ AKGRLIG AKGRLIG AKGRLIG AKGRLIG AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIDGJ AKGRLIG AKGRLIDGJ AKGRLIG AKGRLIG AKGRLIG AKGRLIG AKGRLIG AKGRLIG AKGRLIG AKGRLIG AKGRLIG	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A73QFF6 SALER tr A0A287LUM5 A0A279LUM5 SECH tr A0A287LUM5 A0A287LUM5 SECH tr A0A287LUM5 A0A287LUM5 SECH tr A0A287LUM5 A0A287LUM5 SECH tr A0A287LUM5 A0A287LUM5 SECH tr A0A688G79 A0A688G7 SH1F1 tr A0A688G79 A0A688G7 SH1F1 tr A0A688G79 A0A78G79 SH1F1 tr A0A766 B2U266 SH153 tr A0A78G79 A0A78G78 SH1F1 tr A0A78G79 A0A78G78 SH1F1 tr A0A78G79 A0A78G78 SH1F1 tr A0A78G78 A0A78G78 SH1F1 tr A0A78G78 A0A78G78 SH1F1 tr A0A78G78 A0A78G78 SH1F1 tr A0A778 A0A78G78 SH1F1 tr A0A778 A0A78 SH1F1 tr A0A778 A0A78 A0A78 SH1F2 tr A0A778 A0A78 SH1F2 SH1F1 tr A0A778 A0A78 SH1F2 SH1F1 tr A0A778 A0A78 SH1F2 SH1F1 tr A0A778 SH1F2 SH1F1 SH1F2	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDSSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRCGNSTDNFMDRMENDE RIPLRGFAC RIPLRCGNSTDNFMDRMENDE RIPLRGFAC RIPLRGNGVFNGVFFAQ RIPLRGNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFFAQ RIPLRGNGVFNGVFNGVFFAQ	* *.* : :*** AKGRLIVDGI AKGRLIDGI ARGRVIDGI ARGRVIDGI AKGRLIGI AKGRLIGI AKGRLIDGI	CALLSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFFLERATVNG VT RTHKTE IMPFLLERATNG	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A422PU1 A0A482PU1_CITRO tr A0A7G1PJC3]CTRKO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A292DLW5 A0A297LW5]9ESCH tr A0A292DLW5 A0A292LW5]9ESCH tr A0A292DLW5 A0A292BDW7_SHIDY 39 221189 DP02_ECOLI tr A0A97bGUTK4]A0A70bGUTK4_9ENTR tr A0A70bGUTK4 A0A70bGUTK4_9ENTR tr A0A716GUTK4 A0A70bGUTK4_9ENTR tr A0A716GVTK4]A0A716GUTK4_9ENTR tr A0A716GVTK4]A0A736GUTTK0 tr A0A734GTFG A0A734GUTA6_SALER tr A0A739QFF6 A0A734GUT6_SALER	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LIEGGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL	* *.* : : : : : : : : : : : : : : : : :	VIEWANNESSESLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSFSFSFSLETVAQEL EALKSFFSFSFSLETVAQEL EALKSFFSFSFSFSLETVAQEL EALKSFFSFSFSFSFSFSFSFS EALKSFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSFSF	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A2B7LUM5 A0A270LUM5 BESCH tr A0A2B7LUM5 A0A287LUM5 BESCH tr A0A2B7LUM5 A0A287LUM5 BESCH tr A0A288DDW7 A0A288DLW7 SHISO tr A0A288DDW7 A0A288DLW7 SHISO tr A0A688G0 A0A380G9 SHISO tr A0A688G0 A0A380G9 SHIFL tr A0A181WSP1 A0A181WSP1 KLEOX tr B7LVT1 B7LVT1 ESCF3 tr A0A71654D3 A0A71654D3 ESCFE tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1 CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A731P46 SALER tr A0A379QFF6 A0A379QFF6 SALER tr A0A287LUW5 A0A287LUM5 BESCH	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELEWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRLGRDNSELWREHGFNGVFFAQ RIPLRCHGNGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHGNG RIPLRCHG RIPLRCHG RIPLRCHG RIPLRCHG	* *.* : :*** AKGRLIVDGI AKGRLIDGI ARGRVIDGI ARGRVIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AYNIKDCEI AYNIKDCEI AYNIKDCEI	CALLSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A422PU1 A0A482PU1_CITRO tr A0A7G1PJC3]CTRKO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DDW7_SHISO tr A0A288DDW7 A0A288DDW7_SHISO tr A0A705UTK4 A0A705UTK4_9ENTR tr A0A716UTK4 A0A705UTK4_9ENTR tr A0A716UTK4 A0A705UTK4_9ENTR tr A0A716UTK4 A0A705UTK4_9ENTR tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A381G280 A0A381G280_CITAM tr A0A381G280 A0A381G280_CITAM tr A0A381G280 A0A381G280_CITAM tr A0A76UTK4 A0A736UTK4_9ENTR tr A0A716T454D3 A0A716T45_4D3_ESCFE tr A0A734T46 A0A734T46_SALER tr A0A379QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A281LUM5_9ESCH	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFACDKPAL LGEGKSIDNPWDRMDEIDRRFACDKPAL LGEGKSIDNPWDRMDEIDRRFACDKPAL	* *.* : : : : : : : : : : : : : : : : :	VERVERVE MER PELERATING VERVERVERVERVERVERVERVERVERVERVERVERVERV	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJG3 A0A7G1PJG3 CITKO tr A0A7G1PJG3 A0A7G1PJG3 CITKO tr A0A7G1PJG3 A0A7G1PJG3 CITKO tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A288DDW7 A0A288DLW7 SHISO tr A0A288DDW7 A0A288DLW7 SHISO tr A0A688G39 A0A688G39 SHISO tr A0A688G39 A0A78G19 SHIFL tr B02266 B2U266 SHIB3 tr A0A78UTK4 A0A705UTK4 9ENTR tr A0A181WSP1 A0A181WSP1 KLEOX tr B1UVT1 B7LVT1 ESCF3 tr A0A71654D3 A0A71654D3 ESCFE tr A0A761PJG3 A0A761PJG3 CITRM tr A0A761PJG3 A0A761PJG3 CITRO tr A0A761PJG3 A0A731926 SALER tr A0A379QFF6 A0A379QFF6 SALER tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A28DLW1 A0A288DLW7 SHISO tr A0A287LUM5 A0A28DTUN5 SECH tr A0A287LUM5 A0A28DTUN5 SHISO SHISO	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAC DKPAL LGEGKSIDNPWDRMDEIDRRFAC DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL	* *.* : :*** AKGRLIVDGI AKGRLIDGI AKGRLIDGI ARGRVIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI .:**::**** ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI ATYNLKDCEI	TEALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSFFSFSFSLETVA EALKSAFWNFSFSFSLETVA EALKSAFWNFSFSFS EALKSFFSFSFS EALKSFSFSFS EALKSFSFSFS EALKSFSFSFS EALKSFSFS EALKSFSFSFS EALKSFSFSFS EALKSFSFSSFS EALKSFSFS EALKSFSF	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A422PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITRO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DDW7_SHIDY 59 121189 DPO2_ECOLI 11 tr A0A705UTK4 A0A705UTK4_9ENTR tr A0A705UTK4 A0A705UTK4_9ENTR tr A0A716S4D3 A0A716S74D3_ESCFE tr A0A716S4D3 A0A716S74D3_ESCFE tr A0A716S4D3 A0A716S74D3_ESCFE tr A0A381G280 A0A381G280_CITAM tr A0A761PJC3 A0A736TF46_SALER tr A0A7374T46 A0A736TF46_SALER tr A0A379QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A282DUM7 A0A288DDW7_SHDY	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFACDKPAL LGEGKSIDNPWDRMDEIDRRFACDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL LGEGKSIDNPWDRMDEIDRRFAEDKPAL	* *.* : : : : : : : : : : : : : : : : :	VERVERSE MANY SET OF A STATE OF A	300 300 300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A288DDW7 A0A288DDW7 SHIDY gp P21189 DPO2 ECOLI tr A0A683RG3 A0A683RG3 SHIFL tr B02L66 B2U266 SHIB3 tr A0A764UTK4 A0A7050TK4 9ENTR tr A0A181WSP1 A0A181WSP1 KLEOX tr B1UVT1 B7LVT1 ESCF3 tr A0A71654D3 A0A71654D3 ESCFE tr A0A381G280 A0A381G280 CITAM tr A0A381G280 A0A381G280 CITAM tr A0A381G280 A0A381G280 CITAM tr A0A381G280 A0A381G280 CITAM tr A0A391G280 A0A381G280 CITAM tr A0A391G280 A0A381G280 CITAM tr A0A381G280 A0A381G280 CITAM tr A0A380 CITAM tr A0A3	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNTELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAC DKPAL LGEGKSIDNPWDRMDEIDRRFAC DKPAL LGEGKSIDNPWDRMDEIDRRFAC DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL	* *.* : :*** AKGRLIVDGI AKGRLIDGI ARGRVIDGI ARGRVIDGI AKGRLICGI AYNIKDCI AYNIKDCI AYNIKDCI AYNIKDCI	THE ALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFS EALKSAFWNFSFS EALKSAFWNFSFS EALKSAFWNFS EALKSAFWNFSFS EALKSAFWNFSFS EALKSAFWNFS EALKSAFWNFSFS EALKSAFWNFS EALKSAFWNFSFS EALKSAFWNFS	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A422PU1 A0A482PU1_CITRO tr A0A7G1PJC3]CTRKO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A292DLWS A0A297LUMS_9ESCH tr A0A292DLWS A0A297LUMS_9ESCH tr A0A292DLWS A0A297LUMS_9ESCH tr A0A298DDW7 A0A298DDW7_SHIDY sp 221189 DPC2_ECOLI tr A0A760UTK4 A0A706UTK4_9ENTR tr A0A706UTK4 A0A706UTK4_9ENTR tr A0A716UTK4 A0A706UTK4_9ENTR tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A381G280 A0A381G280_CITAM tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A29TLUMS A0A29TLUMS_9ESCH tr A0A20LK66 A0A200LK66_SHISO tr A0A728DDW7 A0A28BDW7_SHIDY sp 22189 DPC2_ECOLI	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAT LGEGKSIDNPWDRMDEIDRRFAZ LGEGKSIDNPWDRMDEIDRRFAZ LGEGKSIDNPWDRMDEIDRRFAZ LGEGKSIDNPWDRMDEIDRRFAZ LGEGKSIDNPWDRMDEIDRRFAZ DKPAL LGEGKSIDNPWDRMDEIDRRFAZ DKPAL LGEGKSIDNPWDRMDEIDRRFAZ DKPAL LGEGKSIDNPWDRMDEIDRRFAZ DKPAL LGEGKSIDNPWDRMDEIDRRFAZ DKPAL	* *.* : :*** AKGRLIVDG] AKGRLIDG] ARGRVIDG] ARGRVIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLIDG] AKGRLICG] ATYNLKDCE ATYNLKD	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSFFWNFSFSFSLETVAQEL EALKSFFWNFSFSFSTAFWNFSFSFS EALKSFWNFSFSFS EALKSFFWNFSFSFS EALKSFFWNFSFSF EALKSFWNFSFSF EALKSFWNFSFSF EALKSFWNFSFSF EALKSFWNF E	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A288DDW7 A0A288DLW7 SHISO tr A0A288DDW7 A0A288DLW7 SHISO tr A0A288DDW7 A0A288DLW7 SHISO tr A0A588DDW7 A0A288DLW7 SHIFL tr A0A181WSP1 A0A181WSP1 KLEOX tr A0A71654D3 A0A71654D3 SECFE tr A0A381G280 A0A381G280 CITAM tr A0A381G280 A0A381G280 CITAM tr A0A381G280 A0A381G280 CITAM tr A0A381G280 A0A381G280 CITAM tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A731PJC3 CITKO tr A0A7G1PJC3 A0A731PJC3 CITKO tr A0A379QFF6 A0A379QFF6 SALER tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A28DTU7 SHISO tr A0A287LUM5 A0A28DTU7 SHISO tr A0A287LUM5 A0A28DTU7 SHISO tr A0A287LUM5 A0A28DTU7 SHISO tr A0A287LUM5 SHISO tr A0A287JUA5 A0A288DW7 SHISO tr A0A287JUA5 A0A38GJ9 SHIFL tr B2U266 B2U26 SHIS3	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAZ DKPAL LGEGKSIDNPWDRMDEIDRRFAZ DKPAL	* *.* : :*** AKGRLIVDGI AKGRLIDGI AKGRLIDGI ARGRVIDGI AKGRLICGI AYNIKDCEI AYNIKDCEI AYNIKDCEI AYNIKDCEI AYNIKDCEI AYNIKDCEI AYNIKDCEI AYNIKDCEI AYNIKDCEI AYNIKDCEI	<pre>LEALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSSLETVAQEL EALKSAFWNFSFSSLETVAQEL EALKSAFWNFSFSSLETVAQEL EALKSAFWNFSFSSTETVAQEL EALKSAFWNFSFSSTETVAQEL EALKSAFWNFSFSSTETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSSTETVAQEL EALKSAFWNFSFSSTETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSFFFFFSFSFS ETVAQEL EALKSFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF</pre>	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A422PU1 A0A482PU1_CITRO tr A0A7G1PJC3]CTRKO tr A0A734HT46 A0A7G1PJC3]CTRKO tr A0A734HT46 A0A734HT46_SALER tr A0A292DUN5 A0A297LUN5_9ESCH tr A0A292DUN5 A0A297LUN5_9ESCH tr A0A292DUN5 A0A297LUN5_9ESCH tr A0A292DUN5 A0A298DDW7_SHIDY bg P21189 DP02_ECOLI tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A716UTK4 A0A7D6UTK4_9ENTR tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A200LK66 A0A200LK66_SHISO tr A0A728DDW7 A0A28DDW7_SHIDY bg P21189 DP02_ECOLI tr A0A8G29]A0A6N3RG39_SHIFL tr A0A8G29]A0A6N3RG39_SHIFL tr A0A8G39]A0A6N3RG39_SHIFL tr A0A76UTK4 A0A7D6UTK4_9ENTP	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAEDKPAL	* *.* : :*** AKGRLIVDG] AKGRLIDG] AKGRVIDG] ARGRVIDG] AKGRLIDG] AKGRLICKANANANANANANANANANANANANANANANANANANAN	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSFFFFENFSFSFS EALKSFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A287LUM5 A0A27LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A288DDW7 A0A288DLW7 SHIPL tr B0L266 B2U266 SHIB3 tr A0A764D1 A0A7050TK4 9ENTR tr A0A161WS91 A0A181WS91 KLEOX tr B1LVT1 B7LVT1 ESCF3 tr A0A761PJC3 A0A71654D3 ESCFE tr A0A761PJC3 A0A761PJC3 CITRM tr A0A761PJC3 A0A761PJC3 CITRO tr A0A761PJC3 A0A731PJC3 CITRO tr A0A761PJC3 A0A731PJC3 CITRO tr A0A761PJC3 A0A731PJC3 SHIER tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 SHIEN tr A0A381G280 A0A381G280 SHIEN tr A0A381S7 A0A287LUM5 SHIS0 T A0A287LUM5 A0A287LUM5 SHIS0 T A0A288DDW7 A0A288DW7 SHIS1 T A0A287LUM5 A0A287LUM5 SHIS1 T A0A6013RG39 A0A603RG39 SHIFL T B2U266 B2U266 SHIE3 T B2U266 B2U266 SHIE3 T B2U266 B2U266 SHIE3 T B2U266 B2U260 SHIE3 T B2U266 B2U260 SHIE3 T B2U26 B2U260 SHIE3 T B2U266 B2U260	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL	* *.* : :*** AKGRLIVDGI AKGRLIVDGI AKGRLIDGI ARGRVIDGI AKGRLIDGI AKGRLIGGI AKGRLIGGI AKGRLIDGI AKGRLIDGI AKGRLIDGI AKGRLIDGI A	<pre>LEALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSFFFSFSFS EALKSAFWNFSFSFS EALKSFFSFSFS EALKSFFSFSFS EALKSFFFSFS EALKSFFSFS EALKSFFSFS EALKSFFSFS EALKSFFSFS EALKSFFSFS EALKSFFSFS EALKSFFS EALKSFFS EALKSFFSFS EALKSFSFS EALKSFSFS EALK</pre>	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A422PU1 A0A482PU1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A239QFF6 A0A379QFF6_SALER tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A287LUM5 A0A287LUM5_9ESCH tr A0A288DDW7 A0A287LUM5_9ESCH tr A0A288DDW7 A0A288DDW7_SHIDY sp 221189 DP02_ECOLI tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A28DLUM5 A0A28DLUM5_9ESCH tr A0A28BDUM7 A0A28BDLUM5_9ESCH tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266_SHIB3 tr A0A716VTK4 A0A7D6UTK4_9ENTR tr A0A18LWSP1 A0A18LWSP1_KLEOX	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAEDKPAL	* *.* : :*** AKGRLIVDG] AKGRLIDG] AKGRVIDG AKGRVIDG AKGRUIDG AKGRLIDG] AKGRLIDG AYNIKDCEI	CALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSTA EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSTAF EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSAFWNFSFSFS EALKSFF EALKSFF EALKSFF EALKSFF EALKSFF EALKSFF EALKSFF EALKSFF EALKSFF EALKSFF EALKSFF EALKS	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A2B7LUM5 A0A287LUM5 9ESCH tr A0A2B7LUM5 A0A2B7LUM5 9ESCH tr A0A58BDW7 A0A288DUM7 SHIPL tr B0L66 B2U266 SHIB3 tr A0A76UTK4 A0A705UTK4 9ENTR tr A0A181WSP1 A0A181WSP1 KLEOX tr B1UVT1 B7LVT1 ESCF3 tr A0A761PJC3 A0A761PJC3 CITRM tr A0A381G280 A0A381G280 CITAM tr A0A761PJC3 A0A761PJC3 CITRO tr A0A761PJC3 A0A761PJC3 CITRO tr A0A761PJC3 A0A761PJC3 CITRO tr A0A761PJC3 A0A761PJC3 CITRO tr A0A3790F6 A0A3790F6 SALER tr A0A2B7LUM5 A0A2B7LUM5 9ESCH tr A0A2B7LUM5 A0A2B7LUM5 9ESCH tr A0A2B7LUM5 A0A2B7LUM5 SHIS0 tr A0A88DDW7 A0A2B7LUM5 SHIS1 tr A0A761PJC8 A0A761PJC5 SHIFL tr A0A761PJC8 A0A761PJC5 SHIFL tr A0A761PJC8 A0A781PJC5 SHIFL tr A0A761PJC8 A0A761PJC5 SHIFL tr A0A761PJC8 A0A761PJC7 SHIFL tr A0A761PJC8 A0A761PJC7 SHIFL tr A0A761PJC8 A0A761PJC7 SHIFL tr A0A761PJC7 A0A761PJC7 SHIFT STC7 SHIFL tr A0A761PJC7 A0A761PJC7 SHIFT STC7 SHIFL tr A0A761PJC7 A0A761PJC7 SHIFT STC7 SHIFT STC7 SHIFT STC7 SHIFT STC7 SHIFT STC7 SHIFT ST	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL	* *.* : :*** AKGRLIVDG] AKGRLIVDG] AKGRLIDG] ARGRVIDG] ARGRVIDG] AKGRLIDG] AKGRLIGAN A	<pre>LEALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFS ETVAQEL EALKSAFWNFSFSFS ETVAQEL EALKSAFWNFSFSFS ETVAQEL EALKSAFWNFSFSFS ETVAQEL EALKSAFWNFSFS EALKSAFWNFSFS EALKSAFWNFS</pre>	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280_CITAM tr A0A422PU1 A0A482PU1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A292DLW5 A0A297LUM5_9ESCH tr A0A292DLW5 A0A297LUM5_9ESCH tr A0A292DLW5 A0A297LUM5_9ESCH tr A0A292DLW5 A0A297LUM5_9ESCH tr A0A298DDW7 A0A298DDW7_SHIDY bg P21189 DP02_ECOLI tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A760UTK4 A0A7D6UTK4_9ENTR tr A0A716UTK4 A0A716S4D3_ESCFE tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A716S4D3 A0A716S4D3_ESCFE tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SALER tr A0A734HT46 A0A734HT46_SHER tr A0A290LKG6 A0A290LKG6_SHISO tr A0A734HT46 A0A734HT46_SHER tr A0A200LKG6 A0A200LKG6_SHISO tr A0A78G79]A0A6N3RG9_SHIFL tr A0A8G29]A0A6N3RG9_SHIFL tr A0A81SG9 A0A6N3RG9_SHIFL tr A0A81SW91 A0A181WSP1_KLEOX tr A0A181WSP1 A0A716SV43_9ENTR tr A0A181WSP1 A0A181WSP1_KLEOX tr A0A181WSP1 A0A181WSP1_KLEOX tr A0A181WSP1 A0A756T	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAEDKPAL	* *.* : :*** AKGRLIVDG] AKGRLIDG] AKGRVIDG] AKGRVIDG] AKGRVIDG] AKGRLIDG] AKGRLICG] AYNIKDCE] AYNIKNN	THE ALKSAFWNFSSFSLETVSQEL TEALKSAFWNFSSFSLEAVSQEL TEALKSAFWNFSSFSLEAVSQEL TEALKSAFWNFSSFSLEAVSQEL TEALKSAFWNFSSFSLEAVSQEL TEALKSAFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSSFSLETVAQEL TEALKSFFWNFSFSFSLETVAQEL TEALKSFFWNFSFSF TALKFFWNF TEALKFFWNF TEALKSFFWNF TEALKFFWNF TEALKSFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKSFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALKFFWNF TEALFFWNF TEALKFWNF TEALKFFWNF TEALKFFWNF TEALKFWNF TEAL	300 300 300 300 300 300 300 300
tr A0A381G280 A0A381G280 CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A7G1PJC3 A0A7G1PJC3 CITKO tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A288DDW7 A0A288DLW7 SHIFL tr B0L266 B2U26 CS SHIF3 tr A0A7640TK4 A0A7050TK4 9ENTR tr A0A181WSP1 A0A181WSP1 KLEOX tr B1LVT1 B7LVT1 ESCF3 tr A0A71654D3 A0A71654D3 ESCFE tr A0A381G280 A0A381G280 CITAM tr A0A761PJC3 A0A761PJC3 CITRO tr A0A761PJC3 A0A761PJC3 CITRO tr A0A761PJC3 A0A730PJC5 SALER tr A0A3790F6 A0A3790F6 SALER tr A0A3790F6 A0A3700F6 SALER tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A287LUM5 A0A287LUM5 9ESCH tr A0A380J9 A0A6N3RG3 SHIFL tr A0A380J9 A0A6N3RG3 SHIFL tr A0A705UTK4 A0A760TK4 9ENTR tr A0A7165 A0A380TK4 9ENTR tr A0A7165 A0A380TK4 9ENTR tr A0A7165 A0A380TK4 9ENTR tr A0A7165 A0A380TK4 9ENTR tr A0A7165 A0A380 A0A7165 A0A287LUM5 SALER tr A0A7105UTK4 A0A7165 A0A287LUM5 SALER tr A0A7165 A0A380 A0A380 A0A580 SALER Tr A0A7105 A0A287LUM5 A0A287LUM5 A0A287LUM5 A0A287LUM5 A0A287LUM5 A0A287LUM5 A0A287LUM5 A0A380 A0A7165 A0A3705 A0A37	RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRFGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RLPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ RIPLRLGRDNSELEWREHGFKNGVFFAQ LGEGKSIDNPWDRMDEIDRRFAE DKPAL LGEGKSIDNPWDRMDEIDRRFAE DKPAL	* *.* : :*** AKGRLIVDG] AKGRLIVDG] AKGRLIDG] ARGRVIDG] AKGRLIDG] AKGRLICG] AYNLKDC] AYNLYNLYAN	TEALKSAFWNFSSFSLETVSQEL EALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL DALKSAFWNFSSFSLEAVSQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFSLETVAQEL EALKSAFWNFSFSFS EITYTA	300 300 300 300 300 300 300 300

420

420

420

420

420

420

420

420 420

420

420 420

420

420

420

540

540

540

540

540

540

540

540

540

540

540

540

540

540

540

600

600

600

600

600

600

600

600

600

600

600

600

600

600

600

- A Polymerase region

tr A0A381G280 A0A381G280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A3790FF6 A0A3790FF6_SALER tr A0A297LUW5 A0A297LUW5_9ESCH tr A0A200LKG6 A0A200LKG6_SHISO tr A0A28DDW7 A0A288DDW7_SHIDY sp P21189 DP02_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B02C66 B2U266_SHIB3 tr A0A7D6UTK4 A0A7D6UTK4_9ENTR tr A0A181WSP1 A0A181WSP1_KLEOX tr B7LVT1 B7LVT1_ESCF3 tr A0A7L6S4D3 A0A7L6S4D3_ESCFE	LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG LPVDRHG	GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSQPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSQPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY GSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPRMHRAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPNMFAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPNMFAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFPNMFAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFFNMFAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFFNMFAGYVAPNLGEVPHASPGGYVMDSRPGLYDSVLVLDY SSVAAFGHLYFN
DNA polymerase II active site region	(by seque	ence similarity)
Lt A0.381G280 A0.381G280_CITAM tr A0.381G280 A0.381G280_CITAM tr A0.482PUL1 A0.482PUL1_CITRO tr A0.7G1PJC3 A0.A736HT46_SALER tr A0.3739QFF6 A0.A739QFF6_SALER tr A0.287LUW5 A0.A287LUW5_9ESCH tr A0.280DW7 A0.A288DUW7_SHIDY sp P21189 DP02_ECOLI tr A0.288DUW7 A0.A288DUW7_SHIDY sp P21189 DP02_ECOLI tr A0.46N3RGJ9 A0.46N3RGJ9_SHIFL tr A0.46N3RGJ9 A0.46N3RGJ9_SHIFL tr A0.7D6UTK4 A0.A7D6UTK4_9ENTR tr A0.181WSP1 A0.181WSP1_KLEOX tr A0.47L6S4D3 A0.47L6S4D3_ESCFE	KRHGNKPL KRQGNKPL KRQGNKPL KRQGNKPL KRQGNKPL KRQGNKPL KRQGNKPL KRQGNKPL KRQGNKPL KRQGNKPL KRQGNKPL	SOALKI IMNAFYGVLGTTACRFFDPRLASS ITMRGHAIMAOT KALIEAQGYD SOALKI IMNAFYGVLGTTACRFFDPRLASS ITMRGHAIMAOT KALIEAQGYD SOALKI IMNAFYGVLGTTACRFFDPRLASS ITMRGHAIMAOT KALIEAQGYD SOALKI IMNAFYGVLGTTACRFFDPRLASS ITMRGHAIMAOT KALIEAQGYD SOALKI IMNAFYGVLGTTACRFFDPRLASS ITMRGHOIMAOT KALIEAQGYD
tr A0A381G280 A0A381G280_CITAM tr A0A381G280 A0A381G280_CITAM tr A0A7G1PJC3]A0A7G1PJC3_CITKO tr A0A731PJC3]A0A7G1PJC3_CITKO tr A0A734PT46 A0A734HT46_SALER tr A0A790FF6 A0A3790FF6_SALER tr A0A287LUW5 A0A287LUW5_9ESCH tr A0A285DUW7 A0A285DUW7_SHIDY sp P21189 DP02_ECOLI tr A0A285DUW7 A0A285DUW7_SHIDY sp P21189 DP02_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B2U266 B2U266_SHIB3 tr A0A7D6UTK4 A0A7D6UTK4_9ENTR tr A0A181WSP1 A0A181WSP1_KLEOX tr B7LVT1 B7LVT1_ESCF3 tr A0A7L6S4D3 A0A7L6S4D3_ESCFE	VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS VIYGOTDS	TFVWLKRAHTEEEAAKIGRALVAHVNAWWTQSLQEKNLTSALELEYETHFCR TFVWLKRARSEEEAAQIGRSLVAHVNAWWQQELQRQNLTSALELEYETHFCR TFVWLKGAHSEDEAARIGRELVRHVNDWWTQSLQQQNLTSALELEFETHFCR TFVWLRAHSEADAAKIGHMLVRHVNEWWAQTLQQQNLTSALELEFETHFCR TFVWLRAHSEADAAKIGHMLVRHVNEWWAQTLQQQNLTSALELEFETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR TFVWLKGAHSEEEAAKIGRALVQHVNAWWAETLQKQRLTSALELEYETHFCR
// End of DNA polymerases II tr A0A3816280 A0A3816280_CITAM tr A0A482PUL1 A0A482PUL1_CITRO tr A0A761PJC3 A0A761PJC3_CITKO tr A0A734HT46 A0A734HT46_SALER tr A0A379QFF6 A0A379QFF6_SALER tr A0A2871UW5 A0A287LUW5_9ESCH tr A0A280DW7 A0A288DDW7_SHIDY sp F21189 DP02_ECOLI tr A0A6N3RGJ9 A0A6N3RGJ9_SHIFL tr B0A7D6UTK4 A0A706UTK4_9ENTR tr A0A181WSP1 ACA181WSP1_KLEOX tr B7LVT1 B7LVT1_ESCF3_	GLF 7 GLF 7	83 83 83 83 83 83 83 83 83 83

GLF

783

Fig. 4. MSA of the PR Exonuclease domain of bacterial DNA polymerases II

A0A381G280 CITAMCitrobacter amalonaticus A0A7G1PJC3_CITKO, Citrobacter koseri A0A379QFF6_SALER, Salmonella enterica A0A200LKG6_SHISO, Shigella sonnei P21189|DPO2_ECOLI, Escherichia coli B2U266_SHIB3, Shigella boydii A0A181WSP1_KLEOX, Klebsiella oxytoca A0A7L6S4D3_ESCFE, Escherichia fergusonii

tr|A0A7L6S4D3|A0A7L6S4D3_ESCFE

A0A482PUL1 CITRO, Citrobacter rodentium A0A734HT46_SALER, Salmonella enterica subsp. Salamae A0A2B7LUW5_9ESCH, Escherichia marmotae A0A2S8DDW7_SHIDY, Shigella dysenteriae A0A6N3RGJ9_SHIFL, Shigella flexneri A0A7D6UTK4_9ENTR, Enterobacter hormaechei B7LVT1_ESCF3, Escherichia fergusonii

3. 2.1 Active Site Analyses of the PR Exonuclease of DNA polymerase II

The probable active site amino acids in the bacterial DNA polymerases II are shown in Fig. 5. The exonuclease amino acids are very similar to the E. coli DNA polymerase I as both belong to the DEDD superfamily of exonucleases. The active site amino acids are placed based on the crystallographic and SDM data available on the enzyme (D¹⁵⁶ \rightarrow N, D³³⁵ \rightarrow N and D²²⁹ \rightarrow N are the exo⁻ mutants) [10]. The Tyr³³¹ is placed as the proton acceptor as it is a completely conserved amino acid in highly conserved block which is in the equivalent position to the His¹⁶² of the PR exonuclease of the DNA polymerase ε subunits. Furthermore, this PR function in DNA polymerase II could also belong to the dnaQ-Y family like DNA polymerase I where an invariant Y could involve in deprotonation of water molecule similar to the H¹⁶², an invariant amino acid in the dnaQ-H family performing the same function [10]. It is interesting to note that in both the DNA polymerases, (i.e.), I and II, the proton Tyr exhibits also a distance acceptor conservation, (i.e.), 4 amino acids from the last catalytic Asp (Fig. 5).

3.3 PR Function RNase D Exonucleases (DEDYD)

The RNase D (EC 3.1.13.5), one of the seven exoribonucleases, which involve in the 3'maturation of several stable RNAs like tRNA, 5S rRNA, and other small structured RNAs, also belong to the DEDD family [2]. They initiate hydrolysis at the 3'-terminus of an RNA molecule and releases 5'-mononucleotides. MSA of the RNase D from different bacteria is shown in Fig. 6. It is not highly conserved like the DNA polymerases I and II (Figs. 2, 4). However, the active site regions are completely conserved. The metal-binding regions are highlighted in light green and the proton acceptor is highlighted in orange. This class of enzyme also shows the four conserved acidic amino acids, viz. DEDD with the Y as the proton acceptor and hence belong to the DEDD superfamily of exonuclease. The pattern found $-DxE \rightarrow D \rightarrow Y \rightarrow D$ - is very similar to the other exonucleases of DEDYD subfamily of enzymes like DNA polymerases I and II.





CLUSTAL O (1.2.4) multiple sequence alignment RNase D (Only the exonuclease domain is shown)



Palanivelu; IJBCRR, 30(3): 33-62, 2021; Article no.IJBCRR.70674

sp A6V8R6 RND PSEA7	DPLLVRDWGPFAELLEDPRVVKVLHAC <mark>SEDL</mark> EVFLRLTGSLPVPLFDTQLAAAYLGMAHS	120
tr A0A6171490 A0A6171490 KLEPN	DPLLIQDWSPFAELLEDERVVKVLHAC <mark>SEDL</mark> EVFLRLTGSLPVPLFDTQLAAAYLGMAHS	118
tr B7LPP7 B7LPP7 ESCF3	DPLTIRDWSPLKSVLRDPAITKFLHAG <mark>SEDL</mark> EVFLNKFGEMPQPLIDTQVLAAFCGRPMS	115
tr A0A7H9K3L2 A0A7H9K3L2 9ESCH	DPLGITDWSPLKAILRDPSITKFLHAG <mark>SEDL</mark> EVFLNVFGELPQPLIDTQILAAFCGRPMS	111
tr A0A7D6UZ56 A0A7D6UZ56 9ENTR	DPLGISDWSPLKAILRDPSITKFLHAG <mark>SEDL</mark> EVFLNVFGELPQPLIDTQILAAFCGRPMS	111
tr A0A659ZP75 A0A659ZP75_SHISO	DPLGITDWSPLKAILRDPSITKFLHAG <mark>SEDL</mark> EVFLNVFGELPQPLIDTQILAAFCGRPIS	111
sp P09155 RND_ECOLI	DPLGITDWSPLKAILRDPSITKFLHAGSEDLEVFLNVFGELPQPLIDTQILAAFCGRPMS	115
tr A0A380AWJ5 A0A380AWJ5_SHIFL	DPLGITDWSPLKAILRDPSITKFLHAG <mark>\$EDL</mark> EVFLNVFGELPQPLIDTQILAAFCGRPMS	111
tr A0A6I1J9T0 A0A6I1J9T0_9ENTR	DPLGITDWSPLKAILRDPSITKFLHAG <mark>SEDL</mark> EVFLNVFGELPQPLIDTQILAAFCGRPMS	111
tr A0A5Q3TIV6 A0A5Q3TIV6_9ENTR	DPLGITDWSPLKAILRDPSITKFLHAG <mark>SEDL</mark> EVFLNVFGELPQPLIDTQILAAFCGRPMS	111
tr E7SVN6 E7SVN6_9ENTR	DPLGITDWSPLKSILRDPSITKFLHAG <mark>SEDL</mark> EVFLNVFGELPQPLIDTQILAAFCGRPMS	111
	*** : **.*: :*.* :.*.********* **.********	
		100
SPIAOVOROIKND_PSEA/		170
tr B7L DD7 B7L DD7 ESCE3	MGEA SMUEEVSCULL DKSESPEDMLAPDITEDOORVAADDVQHLADVALLDTAKIMAETEAS	173
trla0a7H9K3L2La0a7H9K3L2_9ESCH	WGFASHVEETSGVTLDKSESRTDWLARDIJTEROCEYAAADUWTLDTTAKIMA ETEAS	169
trla0a7D6UZ56La0a7D6UZ56_9ENTR	WGFASMVEEYSGVTLDKSESRTDWLARPLTEROCEYAAADVWYLLPITAKIMVETEAS	169
tr A0A6597P75 A0A6597P75 SHISO	WGFASMVEEYSGVTLDKSESSTDWLARPLTEROCEYAAADVWYLLPTTAKLMV-ETEAS	169
sp P09155 RND_ECOLT	WGFASMVEEYSGVTLDKSESRTDWLARPLTEROCEYAAADVWYLLPITAKIMVETEAS	173
tria0a380awj5la0a380awj5 SHIFL	WGFASMVEEYSGVTLDKSESBTDWLARPLTEROCEYAAADVWYLLPTTAKLMVETEAS	169
tr A0A6I1J9T0 A0A6I1J9T0 9ENTR	WGFASMVEEYSGVTLDKSESRTDWLARPLTEROCEYAAADVWYLLPITAKLMVETEAS	169
tr A0A5Q3TIV6 A0A5Q3TIV6 9ENTR	WGFASMVEEYSGVTLDKSESRTDWLARPLTERQC <mark>EYAAAD</mark> VWYLLPITAKLMVETEAS	169
tr E7SVN6 E7SVN6 9ENTR	WGFASMVEEYSGVTLDKSESRTDWLARPLTERQC <mark>EYAAAD</mark> VWYLLPITAKLMVETEAS	169
_	*::.:*:* .: * *.*:*:*** ***** * <mark>.*</mark> ** ** : * *	
// End of RNase D		
sp A6V8R6 RND PSEA7	SLRGWRRERMGOALLNALESA 376	
tr A0A6171490 A0A6171490 KLEPN	SLRGWRRERMGQALLDALESA 374	
tr B7LPP7 B7LPP7 ESCF3	MISGWRGDLMAERLNALLQEYPQ 375	
tr A0A7H9K3L2 A0A7H9K3L2 9ESCH	LISGWRGALMAEALHNLLQEYPQ 371	
tr A0A7D6UZ56 A0A7D6UZ56 9ENTR	LISGWRGELMAEALHNLLQEYPQ 371	
tr A0A659ZP75 A0A659ZP75_SHISO	LISGWRGELMAEALHNLLQEYPQ 371	
sp P09155 RND_ECOLI	LISGWRGELMAEALHNLLQEYPQ 375	
tr A0A380AWJ5 A0A380AWJ5_SHIFL	LISGWRGELMAEALHNLLQEYPQ 371	
tr A0A6I1J9T0 A0A6I1J9T0_9ENTR	LISGWRGELMAEALHNLLQEYPQ 371	
tr A0A5Q3TIV6 A0A5Q3TIV6_9ENTR	LISGWRGELMAEALHNLLQEYPQ 371	
tr E7SVN6 E7SVN6_9ENTR	LISGWRGELMAEALHNLLQEYPQ 371	
	· *** * · · * * · ·	

Fig. 6. MSA of RNase D exonucleases from different bacteria

A6V8R6|RND_PSEA7, Pseudomonas aeruginosa B7LPP7_ESCF3, Escherichia fergusonii A0A7D6UZ56_9ENTREnterobacter hormaechei P09155|RND_ECOLI, Escherichia coli (K12) A0A6I1J9T0_9ENTR, Enterobacteriaceae bacterium E7SVN6_9ENTR, Shigella boydii A0A6I7I490_KLEPN, Klebsiella pneumonia A0A7H9K3L2_9ESCH, Escherichia marmotae A0A659ZP75_SHISO, Shigella sonnei A0A380AWJ5_SHIFL, Shigella flexneri A0A5Q3TIV6_9ENTR, Salmonella



Fig. 7. Proposed amino acids at the active site of the exonuclease of RNase D of E. coli

CLUSTAL O (1.2.4) MSA of the ε-subunits of DNA polymerases III

	←− →	
tr A0A0F6RDN9 A0A0F6RDN9_CITAM	MTAMSTAITRQIVI <mark>, DTE</mark> TTGMNQIGAHYEGHKIIEIGAVEVVNRRLTGNNFHVYLKPDRL	60
sp P03007 DPO3E ECOLI	MSTAITRQIVLDTETTGMNQIGAHYEGHKIIEIGAVEVVNRRLTGNNFHVYLKPDRL	57
tr A0A1Q8NNQ1 A0A1Q8NNQ1 SHIDY	MSTAITRQIVIDTETTGMNQIGAHYEGHKIIEIGAVEVVNRRLTGNNFHVYLKPDRL	57
tr F3WE50 F3WE50 9ENTR	MSTAITRQIVI.DTETTGMNQIGAHYEGHKIIEIGAVEVVNRRLTGNNFHVYLKPDRL	57
tr E7SZS7 E7SZS7 9ENTR	MTAMSTAITRQIVI <mark>DTE</mark> TTGMNQIGAHYEGHKIIEIGAVEVVNRRLTGNNFHVYLKPDRL	60
tr A0A5F1HT22 A0A5F1HT22 9ESCH	MSTAITROIVIDTETTGMNOIGAHYEGHKIIEIGAVEVVNRRLTGNNFHVYLKPDRL	57
sp P0A1G9 DP03E SALTY	MSTAITROIVIDTETTGMNOIGAHYEGHKIIEIGAVEVINRRLTGNNFHVYLKPDRL	57
tr A0A6C7I1V1 A0A6C7I1V1 SALPK	MSTAITROIVIDTETTGMNOIGAHYEGHKIIEIGAVEVINRRLTGNNFHVYLKPDRL	57
trla0a0F6aX64la0a0F6aX64 SALT1	MSTATTROTVIDTETTGMNOTGAHYEGHKITETGAVEVINRRLTGNNFHVYLKPDRL	57
trla0a3V9NPW9la0a3V9NPW9 SALGL	MSTAITROIVIDTETTGMNOIGAHYEGHKIIEIGAVEVINRRITGNNFHVYLKPDRI	57
	*****	-
	← →	
tr A0A0F6RDN9 A0A0F6RDN9 CITAM	VDPEAFGVHGIADEFLLDKPTFAEVADEFLDYIRGAELVIHNASF <mark>D</mark> IGFMDYEFGKLNRD	120
sp P03007 DPO3E ECOLI	VDPEAFGVHGIADEFLLDKPTFAEVADEFMDYIRGAELVIHNAAFDIGFMDYEFSLLKRD	117
tr A0A108NN01 A0A108NN01 SHIDY	VDPEAFGVHGIADEFLLDKPTFAEVADEFMDYIRGAELVIHNAAFD GFMDYEFSLLKRD	117
tr F3WE50 F3WE50 9ENTR	VDPEAFGVHGIADEFLLDKPTFAEVADEFMDYIRGAELVIHNAAFDIGFMDYEFSLLKRD	117
tr E7SZS7 E7SZS7 9ENTR	VDPEAFGVHGIADEFLLDKPTFAEVADEFMDYIRGAELVIHNAAFDIGFMDYEFSLLKRD	120
trlA0A5F1HT22LA0A5F1HT22 9ESCH	VDPEAFGVHGIADEFLIDKPTFAEVADEFMDY IRGAELVIHNAAFD	117
SDIPOA1G9IDPO3E SALTY	VDPEAFGVHGTADEFLLDKPVFADVVDEFLDYTRGAELVTHNASFDTGFMDYEFGLLKRD	117
trla0a6c7t1V1la0a6c7t1V1 SALPK	VDPEAFGVHGTADEFLLDKPVFADVVDEFLDYTRGAELVTHNASFDTGFMDYEFGLLKRD	117
+rla0a0F6ay64la0a0F6ay64 Sat T1		117
+ x A 0 A 3V 9 N D W 9 A 0 A 3V 9 N D W 9 S A 1 G 1		117
CT [AOAS V JALWS [AOAS V JALWS _SALGE		1 1 /
	· · · · · · · · · · · · · · · · · · ·	
tr A0A0F6RDN9 A0A0F6RDN9 CITAM	IPKTNTFCKVTDSLALARKMFPGKRNSLDALCSRYEIDNSKRT <mark>LH</mark> GAL LDAQ ILADVYLM	180
sp P03007 DPO3E ECOLI	IPKTNTFCKVTDSLAVARKMFPGKRNSLDALCARYEIDNSKRTLHGALLDAQILAEVYLA	177
tr A0A1Q8NNQ1 A0A1Q8NNQ1 SHIDY	IPKTNTFCKVTDSLAVARKMFPGKRNSLDALCARYEIDNSKRT <mark>LH</mark> GAL <mark>LDAQ</mark> ILAEVYLA	177
tr F3WE50 F3WE50 9ENTR	IPKTNTFCKVTDSLAVARKMFPGKRNSLDALCARYEIDNSKRT <mark>LH</mark> GAL LDAG ILAEVYLA	177
tr E7SZS7 E7SZS7 9ENTR	IPKTNTFCKVTDSLAVARKMFPGKRNSLDALCARYEIDNSKRT <mark>LH</mark> GAL LDAG ILAEVYLA	180
tr A0A5F1HT22 A0A5F1HT22 9ESCH	IPKTNTFCKVTDSLAVARKMFPGKRNSLDALCARYEIDNSKRT <mark>LH</mark> GALLDAG <mark>ILAEVYLA</mark>	177
sp P0A1G9 DP03E SALTY	IPKTNTFCKVTDSLALARKMFPGKRNSLDALCSRYEIDNSKRT <mark>LH</mark> GALLDAG <mark>ILAEVYLA</mark>	177
tr A0A6C7I1V1 A0A6C7I1V1 SALPK	IPKTNTFCKVTDSLALARKMFPGKRNSLDALCSRYEIDNSKRTLHGALLDAGILAEVYLA	177
tr A0A0F6AX64 A0A0F6AX64 SALT1	IPKTNTFCKVTDSLALARKMFPGKRNSLDALCSRYEIDNSKRTLHGALLDAGILAEVYLA	177
tr A0A3V9NPW9 A0A3V9NPW9 SALGL	IPKTNTFCKVTDSLALARKMFPGKRNSLDALCSRYEIDNSKRT <mark>LH</mark> GALLDACILAEVYLA	177

tr A0A0F6RDN9 A0A0F6RDN9 CITAM	MTGGOTTMAFSMEGET00-OGNTGIORLVROASKLRVVFATDEEVAAHESRLDLVEKKGG	239
sp P03007 DPO3E ECOLI	MTGGOTSMAFAMEGET0000GEATIORIVROASKLRVVFATDEEIAAHEARLDLVOKKGG	237
tr A0A108NN01 A0A108NN01 SHIDY	MTGGOTSMAFAMEGET0000GEATIORIVROASKLRVVFATDEELAAHEARLDLVEKKGG	237
trlF3WE501F3WE50 9ENTR	MTGGOT SMAFAMEGETOOOOGEATIORIVROASKI.RVVFATDEELAAHEARI.DI.VEKKGG	237
trlE7SZS7/E7SZS7 9ENTB	MTGGOTSMAFAMEGETOOOOGEATIORIVROASKI.RVVFATDEELAAHEARI.DI.VEKKGG	240
trla0a5E1HT221a0a5E1HT22 9ESCH	MTGGOTSMAFAMEGETOOOOGEATIORIVEOASKLEVVENTEELAAHEARIDI.VOEKGG	237
		237
+*10006C7T1V110006C7T1V1 SALDK		237
	WACCULOWAEDWECEAUDUUCEDAIUDIIADU CDI DIVIED CEEEI Y AREGDI DI NUMACU NICOAT DUILUIRGEI AVAACULULULULULULULULULULULULULULULULULULU	237
+ + 1 0 2 20 0 0 0 0 1 2 0 2 20 0 0 0 0 0 0	WACCOMMAENWECEAUDOUCENALOEIMEUN GEI BIMEN GEEET NIMEUD DI MOMACC MACCOMMAENWECEAUDOUCENALOEIMEUN GEI BIMEN GEEET NIMEUD DI MOMACC	237
CT LYON2 A ANEMA LYON2 A ANEMA SYTCP	**************************************	231
trla0a0F6RDN9la0a0F6RDN9 CITAM	SCIWRA 245	
SplP030071DP03E_ECOLT	SCLWRA 243	
tria0a108NN011a0a108NN01 SHIDY	SCLWBA 243	
+*IE3ME501E3ME50 GENTE	SCIMPA 243	

sp P03007 DPO3E ECOLI	SCLWRA 243
tr A0A1Q8NNQ1 A0A1Q8NNQ1_SHIDY	SCLWRA 243
tr F3WE50 F3WE50 9ENTR	SCLWRA 243
tr E7SZS7 E7SZS7 9ENTR	SCLWRA 246
tr A0A5F1HT22 A0A5F1HT22 9ESCH	SCLWRA 243
sp P0A1G9 DPO3E SALTY	SCLWRA 243
tr A0A6C7I1V1 A0A6C7I1V1 SALPK	SCLWRA 243
tr A0A0F6AX64 A0A0F6AX64 SALT1	SCLWRA 243
tr A0A3V9NPW9 A0A3V9NPW9 SALGL	SCLWRA 243
—	* * * * * *

Fig. 8. MSA of PR ε-subunits of DNA polymerases III from different organisms

A0A0F6RDN9_CITAM, Citrobacter amalonaticus A0A1Q8NNQ1_SHIDY, Shigella dysenteriae E7SZS7_9ENTR, Shigella boydii P0A1G9\DPO3E_SALTY, Salmonella typhimurium A0A0F6AX64_SALT1, Salmonella typhimurium

The proposed active site amino acids are shown in Fig. 7. Interestingly, all the three DEDDY subfamily of exonucleases maintain a distance conservation between the last D and the proton acceptor Y, and it is only 4 amino acids.

P03007|DP03E_ECOLI, Escherichia coli (strain K12) F3WE50_9ENTR, Shigella boydii A0A5F1HT22_9ESCH, Escherichia, sp. E4385 A0A6C711V1_SALPK, Salmonella paratyphi A0A3V9NPW9_SALGL, Salmonella gallinarum

3.4 PR Function in DNA Polymerases III (DEDHD)

Bacterial DNA polymerases III are the third type of polymerases where PR function is reported. The PR function is essential for these types of enzymes, as they are the replicative enzymes. Unlike the other two polymerases, viz. the polymerases I and II, the polymerases III always exist as MECs with about 10 different subunits. For example, the DNA polymerase III holoenzyme from E. coli, the most well characterized in this category, is composed of 10 subunits (α , β , ϵ , θ , δ , δ' , γ , τ , χ , ψ), that together with the helicase (DnaB) and the RNA primase (DnaG) form the replisome with a combined molecular weight of ~1 MDa [1 and references therein]. The PR activity is associated with the εsubunit of the MEC. The ϵ -subunit which is made up of ~240 amino acid residues, consists of two domains, the N-terminal domain (1-186) with the PR activity and the C-terminal domain bind to the α -subunit of the polymerase. The ϵ -subunit encoded by dnaQ and contains the 3'-5' PR exonuclease catalytic site to edit any misinserted nucleotides by the α -subunit during the synthesis. Fig. 6 shows the MSA of the ε-subunit from different bacteria. The ɛ-subunits from different bacteria are highly conserved and exhibit over 99% identical residues over the entire sequence. The N-terminal domain contains three conserved exo motifs as marked by arrows. The metal-binding motifs are completely conserved (highlighted in light green) and the possible proton acceptor, H is highlighted in orange. The PR exonuclease contains all of the four invariant acidic amino acids with an identifiable pattern $-DxE \rightarrow D \rightarrow H \rightarrow D$ - and hence belongs to the DEDD exonuclease superfamily and to dnaQ-H subfamily (Fig. 8).

3.4.1 Active Site Analyses of the PR Exonuclease of DNA polymerase III

Fijalkowska and Schaaper [11] found that modification of the two conserved amino acid residues, viz. $Asp^{12} \rightarrow Ala$ and $Glu^{14} \rightarrow Ala$, by SDM experiments resulted in the loss of the exonuclease function and hence suggested to play a role in the coordination of an essential metal ion. Further analysis of the enzyme by Cisneros et al [12] has shown that a water molecule bound to the catalytic metal acts as the nucleophile for the hydrolysis of the phosphate bond. Initially, they observed a direct proton transfer to H^{162} . In a two metal mechanism, the catalytic metal (Me1) is proposed to form an attacking metal-hydroxide which performs a nucleophilic attack on the α -phosphate of the nucleotide base to be excised. The second metal (Me2) is termed as the nucleotide binding metal. These observations were further confirmed by Xray crystallographic analysis of the *ε*-186 by Hamdan et al [13]. Their results showed that the active site was composed of three residues, D¹², E^{14} and $D^{167},$ and bind to two divalent metals. In addition, H^{162} hydrogen bonds to a water molecule that is coordinated to the catalytic metal (Fig. 9).



Fig. 9. Proposed amino acids at the active site of the ε-subunits of *E. coli* DNA polymerase III

The ε- exonuclease belongs to the DnaQ-H family with the four active site carboxylates (Asp¹², Glu¹⁴, Asp¹⁰³, and Asp¹⁶⁷) with the invariant His¹⁶². The H¹⁶² acts as a general base to deprotonate the active site nucleophile. Most important is the substitution of the highly conserved active site tyrosine in enzymes of the DnaQ-Y family (Tyr497 in DNA polymerase I) as the His¹⁶² in the DnaQ-H family [14].

3.5 PR type 3'-5' Exonuclease Activity in RNases T

Another interesting PR type 3' exonuclease was found not in polymerases, but in the tRNA processing enzyme, RNase T (EC 3.1.13.-) [15]. They are single-strand specific exonucleases which trim short 3' overhangs of a variety of RNA species with extreme sequence specificity, discriminating against cytosine at the 3' end of the substrate and thus leaving one or two nucleotide(s) 3' overhang. They are also responsible for the important end processing reaction in tRNAs, and thus they specifically remove the terminal **AMP** residue from uncharged tRNAs (tRNA-C-C-**A**). Therefore, they play a key role in the maturation of tRNAs. Fig. 8 shows the MSA of the RNase T form different organisms. They are almost completely conserved from N- to C-terminal, with only a few amino acid modifications in the entire sequence. (The *E. coli* enzyme is highlighted in yellow). The metal-binding amino acids are highlighted in light green and the proton acceptor amino acid is highlighted in orange. The MSA analysis shows that they also belong to the DEDD superfamily of 3'-5'exonucleases and are characterized by the presence of four acidic residues, DEDD, making the active site.





Fig. 10. MSA of RNase T from different bacteria

A0A564UR04_ESCFE Ribonuclease T, Escherichia fergusonii; P30014|RNT_ECOLI Ribonuclease T, Escherichia coli (strain K12) P66683|RNT_SHIFL Ribonuclease T, Shigella flexneri Q3Z208|RNT_SHISS Ribonuclease T, Shigella sonnei Q32FB8|RNT_SHIDS Ribonuclease T, Shigella dysenteriae A0A181WQU7_KLEOX Ribonuclease T, Klebsiella oxytoca A0A2I8S6X4_9ENTR Ribonuclease T, Citrobacter freundii A0A482PCL8_CITRO Ribonuclease T, Citrobacter rodentium



Fig. 11. The proposed amino acids at the active site of RNase T (E. coli)

As the exonuclease contains all of the four invariant acidic amino acids with an identifiable pattern $-DxE \rightarrow D \rightarrow H \rightarrow D$ - they are also classified under the DEDD exonuclease superfamily (Fig. 10).

3.5.1 Active Site Analyses of the Exonuclease of RNase T

The E coli RNase T has been extensively studied both by X-ray crystallography and SDM experiments [2,15]. Four conserved acidic residues, viz. Asp²³ and Glu²⁵, Asp¹²⁵ and Asp¹⁸⁶ were identified in the active site of the enzyme and are also found to be essential for the exonuclease activity. In addition to the above essential amino acid residues, modification of His¹⁸¹ by an SDM experiment also abolished the exonuclease activity, suggesting that His¹⁸¹ also play an important role in catalysis. Thus, in E. *coli* RNase T, at least five amino acid residues, viz. Asp²³, Glu²⁵, Asp¹²⁵, His¹⁸¹ and Asp¹⁸⁶ are found in the active site (Fig. 11). These residues, together with the substrate, are known to bind two divalent metal ions. The structures of RNase T from Pseudomonas aeruginosa and E. coli have been solved by Zuo et al [15]. The site A metal ion is coordinated by the three conserved acidic residues, and the site B metal ion is coordinated by the conserved aspartate residue. The B site metal ion in P. aeruginosa has an octahedral coordination typical for a magnesium ion whereas the A site metal ion is occupied by a non-magnesium ion (used here a Zn²⁺). In fact, a water molecule (Water-164) occupying the A site in P. aeruginosa RNase T might mimic a nonmagnesium metal ion (Zn^{2+}) with 5 potential coordination ligands [15]. Interestingly, all the three DEDDH subfamily of exonucleases maintains a distance conservation between the last D and the proton acceptor H and it is only 5 amino acids. This distance conservation is also maintained in SARS, SARS-related and HCoVS PR exonucleases as well.

4. PR FUNCTION BY PHP SUPERFAMILY OF EXONUCLEASES

A different type of PR function is reported from Family B DNA polymerases. They belong to the

PHP superfamily of exonucleases and include DNA polymerases X, DNA polymerases III (coediting), eukaryotic, YcdX phosphatases, etc. There are at least 4 different X-type of DNA polymerases are also reported from eukaryotes, viz. the terminal transferases and DNA polymerases β , λ , and μ [16, 17]. Unlike the DEDD superfamily, these enzymes use two invariant Hs followed by three acidic amino acids with the general pattern -HxH \rightarrow E \rightarrow H \rightarrow D-. These enzymes which belong to the PHP superfamily are analyzed further for their active site structure(s).

4.1 Bacterial DNA Polymerases X

The DNA polymerases X are ubiquitous like the DNA polymerases I and are reported in wide variety of organisms like viruses, protozoa, archaea, eubacteria, and eukaryotes [3]. They are strictly template-directed DNA polymerases and preferentially act on DNA structures containing gaps from one to a few nucleotides with a phosphate group at the 5' -end of the downstream of DNA fragments. Therefore, they are suggested to participate in the later stages of DNA synthesis like in base excision repair (BER) and in error-prone non-homologous end joining (NHEJ) activity to repair double-stranded breaks [3]. It is interesting to note that these enzymes structurally different from the DNA are polymerases I and II, as the members of this family possess a different type of PR domain, known as the PHP domain. The polymerase domain is found at their N-terminal, whereas the PHP domain is present at the C-terminal end of the polypeptide. The polymerase domains also found to harbour a HNH motif (Fig. 12). Unlike in DNA polymerases I where the PR function is localized in front of the polymerase domain, in X DNA polymerases, it is localized behind the polymerase domain (Figs. 1, 12), i.e., the PR domains are reversed. In addition to the polymerase and PR domains, a HNH endonuclease type of motif (highlighted in red) is also observed towards at the end of the polymerase domain [3].



Fig. 12. A schematic diagram of the bacterial DNA Polymerases X

CLUSTAL O (1.2.4) MSA of DNA polymerases X (1-315 Polymerase & 336-570 Exo domains in *B Subtilis*).

	tr A0A3P4ARX8 A0A3P4ARX8_THETH PCQ20452.1 tr A0A0D6HTC1 A0A0D6HTC1_STAAU AAW38032.1 tr A0A4V0A0W1 A0A4V0A0W1_STAHY tr A0A348BAD9 A0A348BAD9 9STAP tr A0A348BAD9 A0A348BAD9 9STAP tr A0A77UIA1 A0A77UIA1_9STAP tr A0A2K4ADG4 A0A2K4ADG4_9STAP AAP11460.1 tr A0A5M8S3G5 A0A5M8S3G5_BACAT tr A0A5M8S3G5 A0A5M8S3G5_BACAT tr A0A5M2JTG9 A0A6H2JTG9 BACM0 tr A0A5M1CG75 A0A7H1CG75 BACM1 tr A0A6M4JMH3 A0A6M4JMH3_BACSU tr A0A6M0WNP9 A0A6H0WNP9_9BAC1 tr A0A410WE37 A0A410WE37_BACVA	PGVERAELCGSARRYKDTVG EDINQYQVAGSFRRWKEMSK DDIOQYASAGSFRRYKEMSK NYIDQYSSAGSFRRFKEMSK NYIDQYSSAGSFRRFKEMSK EGIDQYSSAGSFRRYKEMSK EGIDQYSSAGSFRYKEMSK EGIDQYSSAGSFRYKEMSK AEVIRFSRAGSLRRAETVK TDIIKYSRAGSLRRAETVK TDIIKYSRAGSLRRAETVK THIIKFSRAGSLRRAETVK TDIIKFSRAGSLRRAETVK TDIIKFSRAGSLRRAETVK TDIIKFSRAGSLRRAETVK TDIIKFSRAGSLRRAETVK	LUFLVASREGKAVVEGFVRLPQVKEIYAKGKERATVFLN LUFLISTEEPTKVQQALLEFPDIKEQIAVQQTKVSDLQ LUFTISTNHFEKVQQULLDIPNKVKEVAVQTKVSLELA LUFTISTDNPKAVQQULNIPNKVKEVAVQNTKVSLELA LUFTISTDNPKAVQQULNIPNKVKEVAVQNTKVSLELA LUFTISTDNPKAVQQULLDIPNKVKEVAVQNTKVSLELS LUFTISTNNPKEVQQULLDIPNKVKEVAVQNTKVSLELS LUFTISTNNPKEVQQULLDIPNKVKEVAVQNTKVSLELS LUFTISTNNPKEVQQULLDIPNKVKEVAVQNTKVSLELS LUFTISTNNPKEVQQULLDIPNKVKEVAVQNTKVSLELS LUFTIATTEPAAVREHLLQFDNMIEVIASGDTKVSVLLS LUFIIATDPAEVREQLLALPNIKSVIASGDTKVSVLLS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS LUFIIATDHPAEVREQLLALPNIKSVIASGDTKVSVILS	237 233 232 232 232 232 232 232 232 232
		:: .** ** :: *	**:::: * : : : : * : : : *	
<pre>tri AGABYARNE AGABYARNE _THETH I.I. AGABYARNE _ THETH I.I. AGABYARNE AGASYARNE _ THETH I.I. AGAAYARNE AGAYARNE & THETH I.I. AGAAYARNE AGAYARNE & THETH I.I. AGAAYARNE AGAYARNE & THETH I.I. AGAAYARNE AGAYARNE &</pre>	tr A0A3P4ARX8 A0A3P4ARX8_THETH pcQ20452.1 COE05251.1 r A0A0D6HTC1 A0A0D6HTC1_STAAU AAW38032.1 r A0A346BAD9 A0A346BAD9_STAP tr A0A346BAD9 A0A346BAD9_STAP tr A0A770IA1 A0A070IA1_95TAP tr A0A770IA1 A0A070IA1_95TAP tr A0A5M53G5 A0A5M63G5 BACAT tr A0A6H2JTG9 A0A5M423G5 BACAT tr A0A6H2JTG9 A0A5M423G5 BACAT tr A0A6H2JTG9 A0A5M423G5 BACAT tr A0A6H2JTG9 A0A7H1CC%_9BAC1 r A0A6H2MH3 A0A7M1CC%_9BAC1 r A0A6H0MP9 A0A6H0MNP9_9BAC1 tr A0A6H0MP9 A0A6H0MNP9_BAC1 tr A0A6H0WP5]A0A410WE37 bacva tr A0A10WE37 BACVA tr A0A109FQD6 A0A109FQD6_BACL1	TLEELWEAAKALGYQYLG IT KLEEMIEAAIERQYQYIG IT SIRDMVEANIARGYLFMVIT SIRDMVEANIARGYLFMVIT SIRDMVEANIARGYLFMVIT SIRDMVEANIARGYLFMVIT SIRDMVEANIARGYLFMVIT SIRDMEANIARGYDYMVIT SIRDMEANIARGYDYMIT SIRDMEANIARGYLFMAIT SIRDMARGYLFWAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT SIREMAEACHKGYQYMAIT	AVRVAGGPSPEEALKRIEAIRRFNETHGPPYLLAGA HSGSLAVANGLSIERLLEQNERIKKLNESYKEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKQLNKEYDEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYSEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYKEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYKEIDIYSGT HSGSLRVANGLQVERLLRQNEEIKALDKEYKEIDIYSGT HSGSLRVANGLQVERLLQNKEIDALDKEFKFI HSGVLKVNGLTAERLKQQAKEIDALNAEFENFFIKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFIKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKQQAKEIDALNAEFENFFILKGV HSGYLKVANGLTAERLKGQAKEIDALNAEFENFFILKGV HSGYLKVAN	412 SDM 408 407 407 407 407 407 407 407 402 409 409 409 409 409 409 409 409 409 409
LE DAFATS DACHS AND LE JARAFENDEZIN DACHS AND LE JARAFENDEZIN DACHS AND LE JARAFENDEZIN DACHS AND LE JARAFENDEZIN DALOSTOL AND LE JARAFENDEZIN AND	tr A0A3P4ARX8 A0A3P4ARX8_THETH PCQ20452.1 COE05251.1 tr A0A0D6HTC1 A0A0D6HTC1_STAAU AAW38032.1 tr A0A348BAD9 A0A348BAD9 95TAP tr A0A348BAD9 A0A348BAD9 95TAP tr A0A077UIA1 A0A077UIA1_95TAP tr A0A2K4ADG4 A0A2K4ADG4 _ 95TAP tr A0A5M853G5 A0A5M853G5 BACAT tr A0A5M853G5 A0A5M853G5 BACAT tr A0A5M853G5 A0A5M853G5 BACAT tr A0A5H21G9 A0A6H21G9 BACM0 tr A0A7H1CCW5 A0A7H1CCW5 9BAC11 - A0A7H1CCW5 A0A7H1CW5 PBAC11 - A0A7H1CW5 A0A7H1CW5 PBAC11 - A0A7H1CW5 A0A5H21G9 BACH1	VVI I HPDGTLDYPDWVLRELD EMI I KPDGSLDYPDDVLRELD EMI I LPCSLDYDDEILAQLD EMI I LPCSLDYDDEVLAEN EMI I LPCTLDYDDVJLAEMI EMI I LPCTLDYDDVJLAEMI	ULVLVSVHSRFNLPKAEQTKRLLKALENPFVHVA A TAR YVIAAIHQSFNQSEEIMNRLKTACENQVVRH AH TCR YVIAAIHQSFNQSEAIMGRLENACHNPVVRH AH TCR YVIGAIHQSFNQSEQIMERLANACRNPVVRH AH TCR YVIGAIHQSFNQSEQIMERLANACRNPVVRH AH TCR YVIGAIHQSFNQSEQIMERLANACRNPVVRH AH TCR YVIAAIHQSFNQSEQIMERLANACRNPVVRH AH TCR YVIAAIHQSFNQSEQIMERLANACRNPVRH AH TCR YVIAAIHQSFNQSEQIMERLANACRNPVRH AH TCR YVIAAIHQSFNQSEQIMERLANACRNPVRH AH TCR YVIAAIHQSFNQSEQIMERLANACRNPVRH AH TCR YVIAAIHQSFNQSEQIMERLANACRNPVRH TCR YVIAAIHQSFNQSEQIMERLANACRNPVRH TCR YVIAAIHQSFNQSEQIMERLANACRNPVRH AH TCR	472 2SDM 468 467 467 467 467 467 467 472 469 469 469 469
tr A0A3P4ARX8 A0A3P4ARX8 THETH NGLQUDLRVVPPSYGASLQVITGSE STR LAQEKGLRLS VPRGK-RL 292 1 1 0 DONGTEKGBORGTET ADAOCHTCH STAU 1 1	tr D4F2T5 D4F2T5_BACNB tr AOA6H0WNP9 AOA6H0WNP9_9BACI tr AOA410W23'LAOA410WE37_BACVA tr AOA1Q9FQD6 AOA1Q9FQD6_BACLI	EMI ILPDGTLDYDDDMLAEMD EMI ILPDGTLDYDDDVLAEMD EMI ILPDGTLDYDDVLAAMD EMI ILPDGTLDYDDVLAEMD ************************************	DIVIASIHSSFNOPEHVIMKRLETALTNKHVDI AH TGR IVIASIHSSFNOPEHVIMKRLETALTSKHVDI AH TGR IVIASIHSSFNOPEHVIMKRLETALANKHVDI AH TGR DIVIASIHSSFNOPEHVIMKRLETALANKHVDI AH TGR *: ::* *	469 469 469 469
A0A0975243 9CAUD HNH Endo 9DANGPIPKDSDGRETE II KOGNNENDLDNLMCLSIGETVDIHLACKDYQACHAT 64 PCQ0452.1 IEDDVIGVPFLIGPEAPYHILOFTOSKILIK KENSKILI	tr A0A3P4ARX8 A0A3P4ARX8_THETH	NGLQVDLRVVPPESYGAG	lqyltgsk <mark>a</mark> isir <mark>lr</mark> alaqekglkls <mark>eyg</mark> vfrgekrl	292
PCQ20452.1 IEDD/IGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKGKNEKVSCH.IEKA-NGNII 292 COE0525.1 TIAOADD6HTC.JADADD6HTC.STAAU YDDETIGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKGNEKVSCH.IEKA-NGNII 292 TIAOADD6HTC.JADADD6HTC.JADADD6HTC.STAAU YDDETIGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKGNEKVSCH.IEKA-NGNII 291 TIAOAAVOAOWJIAOAAVOAOWJ.STAHY YDDETIGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKADDKVSCH.IEQA-DGTLI 291 TIAOAAVAADGGIAOAZKADAGGISTAA YDDETIGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKADDKVSCH.IEQA-DGTLI 291 TIAOAAVAADGGIAOAZKADAGGISTAA FDDETIGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKADDKVSCH.IEQA-DGTLI 291 TIAOAAVAADGGIAOAZKADAGGISTAA FDDETIGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKANDKVSCH.IEQA-DGTLI 291 TIAOAAVAADGGIAOAZKADAGGISTAA FDDETIGVDFRLIPPAR/HTC.FTGSK.DUIKEDLAKANDKVSCH.IEQA-DGTLI 291 TIAOAAVAADGIAOAACKADGISTAA FDDETIGVDFRLIPPAR/HTC.FTGSK.DUKK.UKE.TAKADKKVSCH.IEQA-DGTLI 291 TIAOAAVAADGIAOAANANASA FEVE-TSVDFRLVTEDGFFTT.HITTARFTGSK.DUKK.GKNEKSEN.VK.WETVETGEVK 291 TIAOAAVAADGIAOAAA10KB3GIAOAANANANANANANANANANANANANANANANANANAN	A0A097J243_9CAUD HNH Endo	9DANGPIPKDSDGRTDE	H N H I <mark>HH</mark> KDGNREN <mark>N</mark> DLDNLMCLSIQE H YDIHLAQKDYQACHAI	64
CUDADALL CUDADALL <td< td=""><td>PCQ20452.1</td><td>IEDDVIGVDFRLIQPEAFYHT</td><td>LQHFTGSKDIN IKIRQLAKQKNEKVSEYGIEEA-NGNII</td><td>292</td></td<>	PCQ20452.1	IEDDVIGVDFRLIQPEAFYHT	LQHFTGSKDIN IKIRQLAKQKNEKVSEYGIEEA-NGNII	292
AAM38032.1 VDETIGVDFPLIEPSAFYHT.0 ITGSKIN IFLOLAKARDEWS YTEQA-DGTLI 291 tr A0A348BAD9]A0A348BAD9_9STAP tr A0A077UIA1 A0A077UIA1 SAAD404_9STAP tr A0A2K4ADC4 A02X4ADC4_9STAP tr A0A2K4ADC4 A02X4ADC4_9STAP tr A0A2K4ADC4 A02X4ADC4_9STAP tr A0A359AARX85]A0A5M833G5_BACAT tr A0A6H2JTG9 A0A42H2G5_BACAT tr A0A5M853G5 A0A5M833G5_BACAT tr A0A6H2JTG9 A0A42H2G5_BACAT tr A0A374LCCM5_9BACI tr A0A394ARX8]A0A394ARX8_THETH PC220452.1 COB5251.1 tr A0A394ARX8 A0A394ARX8_THETH PC220452.1 COB5251.1 tr A0A394ARX8 A0A394ARX8_THETH tr A0A394ARX8 A0A394ARX8_THETH tr A0A348BAD9 A0A348BAD9_9STAP tr A0A3794ARX8 A0A394ARX8_THETH PC220452.1 COB5251.1 tr A0A348BAD9 A0A348BAD9 ASTAP tr A0A3794ARX8 A0A394ARX8_THETH tr A0A3794ARX8 A0A394ARX8_THETH tr A0A394ARX8 A0A394ARX8_THETH tr A0A394ARX8 A0A394ARX8_THETH tr A0A394ARX8 A0A394ARX8_THETH tr A0A348BAD9 A0A348BAD9 ASTAP tr A0A3794ARX8 A0A394ARX8_THETH tr A0A3794ARX8 A0A394ARX8_THETH tr A0A394ARX8 A0A394ARX8_THETH tr A0A3794ARX8 A0A394ARX8_THETH tr A0A394ARX8 A0A394ARX8_THETH tr A0A394BAR39 A0A348BAD9 ACTI_STAAU AA738032.1 tr A0A348BAD9 A0A348BAD9 ASTAP tr A0A3794ARX8 A0A394ARX8_THETH tr A0A348BAD9 A0A348BAD9 ASTAP tr A0A3794ARX8 A0A394ARX8_THETH tr A0A348BAD9 A0A348BAD9 ASTAP tr A0A348BAD9 A0A348BAD9 ASTAP tr A0A348BAD9 A0A348BAD9 ASTAP tr A0A348BAD9 ASTAP tr A0A348BAD9 ASTAP tr A0A348BAD9 ASTAP tr A0A348BAD9 ASTAP tr A0A544ADC4 ASTAP tr A0A348BAD9 ASTAP tr A0A543MBAD9 ASTAP tr A0A643MB3 ASSSSABACAT tr A0A543MBAD9 ASTAP tr A0A643MB3 ASSSSABACAT tr A0A543MBAD9 ASTAP tr A0A643MB3 ASSSSABAD9 ASTAP tr A0A643MB3 ASSSSABABAD9 ASTAP tr A0A643MB3 ASSSSABABAD9 ASTAP tr A0A643MB3 ASABAD9 ASTAP tr A0A643MB3 ASABAD9 ASTAP tr A0A643MB3 ASABAD9 ASTAP tr A0A643MB3 ASABAD9 ASTAP tr A0A543MBABAB9 A	tr A0A0D6HTC1 A0A0D6HTC1 STAAU	YDDETIGVDFRLIEPVAFIHI YDDETIGVDFRLIEPSAFYHT	LQEFTGSKEIN IRIROLAKARDEKVSEYGIEQA-DGTLI	291
tr A0A4Y0A0Wi A0A4V0AWi _ STAHY TDETIGVDFRLIEPSAFYHT O FTGSKEN IRLD LARAQUEKS STIEQA-DGTLI 291 tr A0A9770LAI A0A0770LAI _ 9STAP FDDETIGVDFRLIEPSAFYHT O FTGSKEN IRLD LARAQUEKS STIEQA-DGTLI 291 tr A0A84BAD9 A0A2K4ADG4 _ 9STAP FDDETIGVDFRLIEPSAFYHT O FTGSKEN IRLD LARAQUEKS STIEQA-DGTLI 291 tr A0A84BAD9 A0A2K4ADG4 _ 9STAP FDDETIGVDFRLIEPSAFYHT O FTGSKEN IRLD LARAQUEKS STIEQA-DGTLI 291 tr A0A84B3AD9 A0A6K2JTG9 A0A6K2JTG9 BACKT FEVE-TSVDFRLITEOFFTT H FTGSKEN IRLD LARAQUEKS STIVENCETGEK 291 tr A0A6M3MB1 A0A6K4JMH3 A0A6K4JMH3 BACSU FEVE-TSVDFRLITEOFFTT H FTGSKEN IRLM LARAGUEKS STIVENCETGEK 291 tr A0A6H0WNP9 A0A6H0WNP9 BACI FEVE-TSVDFRLITEOFFTT H FTGSKEN IRLM LARAGUEKS STIVENCETGEK 291 tr A0A6H0WNP9 A0A6H0WNP9 BACI FEVE-TSVDFRLITEOFFTT H FTGSKEN IRLM LARGERIES STIVENCETGEK 291 tr A0A6H0WNP9 A0A6H0WNP9 - 9BACI FEVE-TSVDFRLITEOFFTT H FTGSKEN IRLM LARGERIES STIVENCETGEK 291 tr A0A6H0WNP9 A0A340WAS3 THETH FEVE-TSVDFRLITEOFFTT H FTGSKEN IRLM LARGERIES STIVENCETGEK 291 tr A0A6H0WNP9 A0A340WAS3 THETH FEVE-TSVDFRLITEOFFTT H FTGSKEN IRLM LARGERIES STIVETVETGEK 201	AAW38032.1	YDDETIGVDFRLIEPSAFYHT	LQIFTGSKEN IR IR DLAKARDEKVSEYGIEQA-DGTLI	291
tr A0A077UIA1 SACATURAL 95TAP tr A0A02K4ADG4 SACATURAL 95TAP tr A0A5M833G5 A0A5M833G5 BACATURAL 95TAP tr A0A5M833G5 A0A5M833G5 BACATURAL 95TAP tr A0A5M833G5 A0A5M833G5 BACATURAL 95TAP tr A0A6M2JTG9 BACM tr A0A10WE37 BACABURAL 95TAP tr A0A6M2JTG9 BACM tr A0A10WE37 BACABURAL 95TAP tr A0A6M2JTG9 BACM tr A0A10WE37 BACABURAL 95TAP tr A0A10WE37 BACABURAL 95TAP tr A0A10WE37 BACABURAL 95TAP tr A0A10WE37 A0A410WE37 BACABURAL 95TAP tr A0A10WE37 A0A410WE37 BACABURAL 95TAP tr A0A3294ARX8 A0A3294ARX8 _ THETH tr A0A3046HTC1 STAAU AAM38032.1FEVE-TSUDFLUTEOPFTT HETTGSKDH KHC LIAKERGENES EV VETVETGEIK 291 FEVE-TSUDFLUTEOPFTT HETTGSKDH KHC LIAKERGENES EV VETVETGENK 291 FEVE-TSUDFLUTEOPFTT HETTGSKDH KHC LIAKERGENES EV VETVETGENK 291 FEVE-TSUDFLUTEOPFTT HET	tr A0A4V0A0W1 A0A4V0A0W1_STAHY	FDDETIGVDFRLIEPSAFYHT	LOHFTGSKEIN IRLEDLAKARDERVSEYGIEQA-DGTLI	291
tr A0A2K4ADG4 A0A2K4ADG4_STAP PDDETIGVDFRLIEPSAFVHICQLTGSKEH KEP LAKARNEWS KYCHEQA-DGTLI 291 tr A0A5M853G5 A0A5M853G5 BACAT YEVD-TISUPRLYTEROFFTT.HELTGSKEH KKP LIAKARNEWS KYCHEQA-DGTLI 291 tr A0A6H2JTG9 A0A6H2JTG9 BACMD YEVD-TISUPRLYTEROFFTT.HELTGSKEH KKP LIAKERGERLS KYCHEVETGEKK 291 tr A0A6H2JTG9 A0A6H2JTG9 BACMD FEYE-TSVDFRLYTEROFFTT.HELTGSKEH KKP LIAKERGERLS KYCHEVETGEIK 291 tr A0A6H0MNP9 A0A6H0MNP9 BACSU FEYE-TSVDFRLYTEROFFTT.HELTGSKEH KKP LIAKERGERLS KYCHEVETGEIK 291 tr A0A6H0MNP9 A0A6H0MNP9 BACSU FEYE-TSVDFRLYTEROFFTT.HELTGSKEH KKP LIAKERGERLS KYCHEVETGEIK 291 tr A0A3P4ARX8 A0A3P4ARX8_THETH FEYE-TSVDFRLYTEROFFTT.HELTGSKEH KKP LIAKERGERLS KYCHEVETGEIK 291 tr A0A3P4ARX8 A0A3P4ARX8_THETH AGGTEECVYAALGLFPI PPHKED FEYE-TSVDFRLYTEROFFTT.HELTGSKEH KKP LIAKERGERLS KYCHEVETGEIK 291 tr A0A3P4ARX8 A0A3P4ARX8_THETH AGGTEECVYAALGLFPI PPHKED GSEFDKDLSQIIQLDIK GDINK TYSDGAF 347 tr A0A3P4ARX8 A0A3P4ARX8_THETH AGGTEECVYAALGLFPI PPHKED GSEFDKDLSQIIQLDIK GDINK TYSDGAF 347 tr A0A3ABBAD9 SAABW TYSDERKYKERFYKER GSEFDKDLSQIIQLDIK GDINK TYSDGAF 347 tr A0A3ABBAD9 SAABW GYSEKI YHENWYNFIPPAMED GSEFDKDLSNIITLDDIN GDINK TYSDGAF 347 <td>tr A0A077UIA1 A0A077UIA1_9STAP</td> <td>FDDETIGVDFRLIEPSAFYHT</td> <td>LÕHFTGSKEHNIR<mark>IR</mark>ÕLAKARNEKVSEYGIEÕA-DGTLI</td> <td>291</td>	tr A0A077UIA1 A0A077UIA1_9STAP	FDDETIGVDFRLIEPSAFYHT	LÕHFTGSKEHNIR <mark>IR</mark> ÕLAKARNEKVSEYGIEÕA-DGTLI	291
Init: In	tr A0A2K4ADG4 A0A2K4ADG4_9STAP	FDDETIGVDFRLIEPSAFYHT	LQHFTGSKEIN IR IRDLAKARNEKVSEYGIEQA-DGTLI	291
tr A0A6H2JTG9 A0A6H2JTG9 BACMOFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A6H4JMH3 A0A6H4JMH3 BACSUFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A6H4JMH3 A0A6H4JMH3 BACSUFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A109F701 A0A6H0WNP9 BACIFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A109F706 A0A109F706 BACLIFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A109F706 A0A109F706 BACLIFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A109F706 A0A109F706 BACLIFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A3P4ARX8 A0A3P4ARX8 _ THETHFEYE-TSUDFRLUTEE0FTTLHFTGSKDN IK KED TAKERGERTS YG VETVETGEIK291tr A0A06H1C1 A0A006HTC1 _ STAAUAGETEEGYAALGLPF1PPLREDGTEFF1KDLSQIIQLDIKGDIMN TTYSDGAFAAW33032.1COE05251.1GOIDAGHTTYSDGAF347tr A0A040A00M1 A0A4V0A0M1 STAHYGYDSEAKIYEHFNVNF1PPAMREDGSEF1KDLSNIITIDDINGDIMN TTYSDGAFQYDSEAKIYEHFNVF1PPAMREDGSEF1KDLSNIITIDDINGDIMN TTYSDGAF347GYDSEAKIYEHFNVF1PPAMREDGSEF1KDLSNIITLDDINGDIMN TTYSDGAF347tr A0A6H2/MA3 A0A6H4/MA3 BACSUTFFSEREYAHFGLPL1PPAMREDGSEF1KDLSNIITLDDINGDIMN TTYSDGAFtr A0A6H2/MA3 A0A6H4/MA3 BACSUTFFSEREYAHFGLPL1PPAMREDGSEF1KDLSNIITLDDINGDIMN TTYSDGAFtr A0A6H2/MA3 A0A6H4/MA3 BACSUTFFSEREYAHFGLPL1PPAMREDGSEF1KDLS	tr A0A5M8S3G5 A0A5M8S3G5 BACAT	FEYE-TSVDFRLVTEEQFPTT	LHEFTGSKDEN IKMRDIAKERGERISEYGVETVETGEVK	291
tr A0A /HICCWS 190A/HICCWS 190A/L FEYE-TSUDFRLUTEE0FTTLIFTGSK DAILKERGELS FOUNTVETGELK 291 tr A0A /HICCWS 10A /HICCWS 190A/L FEYE-TSUDFRLUTEE0FTTLIFTGSK DAILKERGELS FOUNTVETGELK 291 tr A0A 10WE37 A0A 10WE37 BACVA FEYE-TSUDFRLUTEE0FTTLIFTGSK DAILKERGELS FOUNTVETGELK 291 tr A0A 10WE37 A0A 10WE37 BACVA FEYE-TSUDFRLUTEE0FTTLIFTGSK DAILKERGELS FOUNTVETGELK 291 tr A0A 10WE37 A0A 10WE37 BACVA FEYE-TSUDFRLUTEE0FTTLIFTGSK DAILKERGELS FOUNTVETGELK 291 tr A0A 3P4ARX8 A0A 3P4ARX8_THETH FCYE-TSUDFRLUTEE0FTTLIFTGSK DAILKERGELS FOUNTVETE0FTS 291 tr A0A 3P4ARX8 A0A 3P4ARX8_THETH AGETEEGYVALGLPF1PPLRED GETEEGYVALGLPF1PPLRED GETEGYVALGLPF1PPLRED GETEGYVALGLPF1PL 31-5' EXORUCL000C 352 SDM tr A0A 3P4ARX8 A0A 3P4ARX8_THETH GOLONG TYSDGQN 352 SDM GYESEKEIY DEPNXSY IPPTMRED GEFEDKDLSQIIQLEDIN GDILM TYSDGAF 347 QYDSEAKIYEHFNVSY IPPARED GEFEDKDLSQIIQLEDIN GDILM TYSDGAF 347 QYDSEAKIYEHFNVNFI PPARED GSEFDKDLSNIITIDDIN GDILM TYSDGAF 347 QYDSEAKIYEHFNVNFI PPARED GSEFDKDLSNIITIDDIN GDILM TYSDGAF 347 QYDSEAKIYEHFNVNFI PPARED GSEFDKDLSNIITIDDIN GDILM TYSDGAF 347 </td <td>tr A0A6H2JTG9 A0A6H2JTG9_BACMO</td> <td>FEYE-TSVDFRLVTEEQFPTT</td> <td>LHEFTGSKDENIK<mark>MR</mark>DIAKERGERISEYGVETVETGEIK</td> <td>291</td>	tr A0A6H2JTG9 A0A6H2JTG9_BACMO	FEYE-TSVDFRLVTEEQFPTT	LHEFTGSKDENIK <mark>MR</mark> DIAKERGERISEYGVETVETGEIK	291
tr A04275 [D4F275 BACNB FEYE-TSUDFRLUTEEOFTTLHFTGSKDH IK HE LAKERGERLS EVENTUTEGEIK 291 tr A0A10WE37 A0A10WE37 BACVA FEYE-TSUDFRLUTEEOFTTLHFTGSKDH IK HE LAKERGERLS EVENTUTEGDIK 291 tr A0A10WE37 A0A10WE37 BACVA FEYE-TSUDFRLUTEEOFTTLHFTGSKDH IK HE LAKERGERLS EVENTUTEGDIK 291 tr A0A10WE37 A0A10WE37 BACVA FEYE-TSUDFRLUTEEOFTTLHFTGSKDH IK HE LAKERGERLS EVENTUTEGDIK 291 tr A0A309F006 A0A10WE37 BACVA FEYE-TSUDFRLUTEEOFTTLHFTGSKDH IK HE LAKERGERLS EVENTUTEGDIK 291 tr A0A30400WE37 BACVA FEYE-TSUDFRLUTEEOFTTLHFTGSKDH IK HE LAKERGERLS EVENTUTEGDIK 291 tr A0A3040A304ARX8 _ THETH FEYE-TSUDFRLUTEEOFTTLHFTGSKDH IK HE LAKERGERLS EVENTUTEGEIK 291 tr A0A306HTL A0A006HTC1 _ STAPU AGETTEGTVALGLEFIPPLKED HEEIBAALAGLEFIPLEEISIK GDIMM TYSDGAF 347 tr A0A4V0A0MI A0A4V0A0MI _ STAPU GEFTD=KDLSNIITLDDIN GDIMM TYSDGAF 347 347 tr A0A348BAD9 A0A348BAD9 _ STAPU GYDEFKLYEHPNVNFIPPAMRED GSEFDKDLSNIITLDDIN GDIMM TYSDGAF 347 tr A0A642JTC9 A0A4KBZGE BACXD TFFSEREFYAHFGLPLIPPENKED GSEFDKDLSNIITLDDIN GDIMM TYSDGAF 347 tr A0A642JTC9 A0A6KBZGE BACXD TFFSEREFYAHFGLPLIPPENKED GSEFDKDLSNIITLDDIN GDIMM TYSDGAF 347 tr A0A642JTC9 A0A6KBZGE BACXD TFF	tr A0A6M4JMH3 A0A6M4JMH3 BACSU	FEYE-TSVDFRLVTEEQFPTT FEYE-TSVDFRLVTEEOFPTT	LHHFTGSKDHNIKMROIAKERGERISEYGVETVETGEIK	291
tr A0A6HOWNP9 A0A6HOWNP9 _ BACKI FEYE-TSUDFRLUTEEQFTTLHEFTOSKDHINKMEDTAKERGERLSEVEVETUETGDIK 291 tr A0A1Q9FQD6 A0A1Q9FQD6 _ BACLI FEYE-TSUDFRLUTEEQFTTLHEFTOSKDHINKMEDTAKERGERLSEVEVETUETGDIK 291 tr A0A3P4ARX8 A0A3P4ARX8 _ THETH FOLMEDTAKERGERLSEVEVETUETGDIK 291 r A0A0D6HTC1 A0A0D6HTC1 _ STAAU AGETEEGVSALEPIPTED HGEIEALAGRIPRLEISEIK GDIDVISTYSDCQN 352 SDM AAW38032.1 QYDSEAKIYEHFNVNFIPPAMRED GSEFDKDLSNIITIDDIN GDILMM TYSDGAF 347 QYDSEAKIYEHFNVNFIPPAMRED GSE	tr D4FZT5 D4FZT5_BACNB	FEYE-TSVDFRLVTEEQFPTT	LHEFTGSK <mark>DEN</mark> IK <mark>MR</mark> DIAKERGERISEYGVETVETGEIK	291
tr A0AlQ9FQD6 A0AlQ9FQD6 BACLI File TSVDFRLVTEEQFTILLEFFGSX.DELIDERGENESE VETIEGAIX 291 C0E05251.1 Tr A0AdV0A0M1 A0A4V0A0M1 STAHY AGTESEVIENTERVINFIPPAMRED GSEFDKDLSNIITIDDIN GDIMM TTYSDGAF 347 QYDSEAKIYEHENVNFIPPAMRED GSEFDKDLSNIITIDDIN GDIMM TTYSDGAF 347 Tr A0ASM833G5 A0A5M833G5 BACAT TFFEFEQAYAHGLPLIPPENNES GGEVET	tr AOA6HOWNP9 AOA6HOWNP9 _ 9BACI	FEYE-TSVDFRLVTEEQFPTT	LHETGSKOHNIKMEDIAKERGERISEYGVETVETGDIK	291
i: :::::::::::::::::::::::::::::::::::	tr A0A1Q9FQD6 A0A1Q9FQD6_BACLI	FEYE-TSVDFRLVTEEQFPTT	LHIFTGSKDEN IK <mark>MR</mark> QIAKERGERIS <mark>EYG</mark> VETIETGAIK	291
Image: Constraint of the constra		: :*:*:: :	<u>*:</u> :***** <mark>*</mark> :::*************************	
AAW38032.1	tr A0A3P4ARX8 A0A3P4ARX8_THETH PCQ20452.1 COE05251.1 tr A0A0D6HTC1 A0A0D6HTC1 STAAU	Polymerase Ageteegvyaalglpfipppl Tygsekeiydhfnvsyipptm Qfnseaeiyehfgvswiepsm Qydseakiyehfnvnfippam	Linker 3'-5' Exonucles RED HGEIALGRLPRLLEISEIK GDIØV STYSDGQN RED GTEFDKDIQDIQLEDIN GDIANN TTYSDGAF RED GSEFDKDLSNIITIDDI GDIANN TTYSDGAF	352 <mark>SDM</mark> 348 347 347
tr A0A4V0A0W1 A0A4V0A0W1 STMHYQYDSEAKIYEHENVNFIPPAMREDGSEFDKDLSNIITLDDINGDIMM TYSDGAF347tr A0A348baD9 A0A348baD9 A0A348baD9 STAPQYDSEAKIYEHENVNFIPPAMREDGSEFDKDLSNIITLDDINGDIMM TYSDGAF347tr A0A077UIA1 A0A077UIA1 STAPQYDSEAKIYEHENVNFIPPAMREDGSEFDKDLSNIITLDDINGDIMM TYSDGAF347tr A0A2K4ADC4 A0A2K4ADC4 STAPQYDSEAKIYEHENVNFIPPAMREDGSEFDKDLSNIITLDDINGDIMM TYSDGAF347tr A0A5M8S3G5 BACATTFFEEDFFAHFGLPFIPPAMREDGSEFDKDLSNIITLDDINGDIMM TYSDGAF349tr A0A6H3JTG9 A0A6H2JTG9 BACATTFFSEQAFYAHFGLPLIPPEIRESGQEVDTYHDGIELIEHDIKGDIMM STWSDGAF349tr A0A7H1CCW5 A0A7H1CCW5 BACITTFFSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEHQIKGDIMM STWSDGAF349tr A0A6H3VH3 A0A6H3VH3 BACSUTFPSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEHQIKGDIMM STWSDGAF349tr A0A6H0WH59 A0A10WE37 BACXATFPSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEPGQIKGDIMM STWSDGAF349tr A0A10WE37 A0A410WE37 BACXATFPSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEPGQIKGDIMM STWSDGAF349tr A0A10QFQD6 A0A1Q9FQD6_BACLITFPSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEPGQIKGDIMM STWSDGAF349tr A0A10QFQD6 A0A1Q9FQD6_BACLITFPSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEPGQIKGDIMM STWSDGAF349tr A0A10QFQD6 A0A1Q9FQD6_BACLITFPSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEPGQIKGDIMM STWSDGAF349tr A0A10QFQD6 A0A1Q9FQD6_BACLITFPSEREFYAHFGLPLIPPEIRESGQEVETYNDSIELIEPGQIK <td>AAW38032.1</td> <td>QYDSEAKIYEHFNVNFIPPAM</td> <td>RED GSEFDKDLSNIITIDDIN GDIHMHTTYSDGAF</td> <td>347</td>	AAW38032.1	QYDSEAKIYEHFNVNFIPPAM	RED GSEFDKDLSNIITIDDIN GDIHMHTTYSDGAF	347
tr A0A3*05BAD3 AVA3*05BAD3 _STAFQUDSEANI IENENVNEIPPAMRED GSEFDKDLSNIITLDDIN GDIMMETTYSDGAF347tr A0A077ULTAI A0A077ULTAI A0A077ULTAI A0A077ULTAI STAFQUDSEANI YEHFNVNEIPPAMRED GSEFDKDLSNIITLDDIN GDIMMETTYSDGAF347tr A0A2K4ADG4 A0A2K4ADG4 _STAFQUDSETKI YEHFNVNEIPPAMRED GSEFDKDLSNIITLDDIN GDIMMETTYSDGAF347tr A0A5M8S3G5 A0A5M8S3G5 BACATTFFEEDFFAHFGLPIIPPENESGGEVDTYHDGIELIETHDIKGDIMMETTYSDGAF349tr A0A6H2JTG9 A0A6H2JTG9 BACMOTFFEEREFYAHFGLPIIPPELRESGGEVETYNDSIELIEHEQIKGDIMMETTWSDGAF349tr A0A6H2JH3 A0A6H4JMH3 BACSUTFFEEREFYAHFGLPIIPPELRESGGEVETYNDSIELIEHEQIKGDIMMETWSDGAF349tr A0A6H0WNP9 A0A6H0WNP9 BACITFFSEREFYAHFGLPIIPPELRESGGEVETYSDSIELIEPGQIKGDIMMESTWSDGAF349tr A0A6H0WNP9 A0A6H0WNP9 BACITFFSEREFYAHFGLPLIPPELRESGGEVETYSDSIELIEPGQIKGDIMMESTWSDGAF349tr A0A6H0WNP9 A0A6H0WNP9 BACITFFSEREFYAHFGLPLIPPELRESGGEVETYSDSIELIEPGQIKGDIMMESTWSDGAF349tr A0A10WE37 A0A410WE37 BACVATFFSEREFYAHFGLPLIPPELRESGGEVETYSDSIELIEPGQIKGDIMMESTWSDGAF349tr A0A10WE37 A0A410WE37 BACVATFFSEREFYAHFGLPLIPPELRESGGEVETYSDSIELIEPGQIKGDIMMESTWSDGAF349tr A0A10WE37 A0A410WE37 BACVATFFSEREFYAHFGLPLIPPELRESGGEVETYSDSIELIEPGQIKGDIMMESTWSDGAF349tr A0A10WE37 A0A410WE37 BACVATFFSEREFYAHFGLPLIPPELRESGGEVETYSDSIELIEPGGIKGDIMMESTWSDGAF349tr A0A10WE37 A0A410WE37 BACVATFFSEREFYAHFGLPLIPPELRESGGEVETYSDSIELIEPGGIKGDIMMESTWSDGAF	tr A0A4V0A0W1 A0A4V0A0W1_STAHY	QYDSEAKIYEHFNVNFIPPAM	RED GSEFDKDLSNIITLDDIN GDIHMHTTYSDGAF	347
tr A0A2K4ADG4 A0A2K4ADG4 9STAPQYDSETKIYEHFNVRFIPPAMREDGSEFDKDLSNIITLDDINGDIMSTTYSDGAF347AAP11460.1TFFSEGAFYAHFGLPIPPAMREDGSEFDKDLSNIITLDDINGDIMSTTYSDGAF352tr A0A5M853G5 A0A5M853G5 BACATTFFSEGAFYAHFGLPIIPPEVREDGQEVDTYHDGIELIETHDIKGDIMSTTYSDGAF352tr A0A642JTG9 A0A6H2JTG9 BACMOTFFSEREFYAHFGLPIIPPELRESGQEVDTYHDGIELIETHDIKGDIMSSTWSDGAF349tr A0A6H2JTG9 A0A6H2JTG9 BACMOTFFSEREFYAHFGLPIIPPELRESGQEVDTYNDSIELIEHDQIKGDIMSSTWSDGAF349tr A0A6H4JHH3 A0A6H4JHH3 BACSUTFFSEREFYAHFGLPIIPPELRESGQEVETYNDSIELIEHDQIKGDIMSSTWSDGAF349tr A0A6H0WH9 A0A6H0WH99 BACITFFSEREFYAHFGLPIIPPELRESGQEVETYSDSIELIEHQGIKGDIMSSTWSDGAF349tr A0A6H0WH99 A0A6H0WH99 BACITFFSEREFYAHFGLPIIPPELRESGQEVETYSDSIELIEHQGIKGDIMSSTWSDGAF349tr A0A10WE37 A0A10WE37 BACVATFFSEREFYAHFGLPIIPPELRESGQEVETYSDSIELIEFQGIKGDIMSSTWSDGAF349tr A0A10PFQD6 A0A10PFQD6 BACLITFFSEREFYAHFGLPIIPPELRESGQEVETYSDSIELIEFQGIKGDIMSSTWSDGAF349tr A0A10PFQD6 A0A10PFQD6 BACLITFFSEREFYAHFGLPIIPPELRESGQEVETYSDSIELIEFQGIKGDIMSSTWSDGAF349tr A0A10PFQD6 A0A10PFQD6 BACLITFFSEREFYAHFGLPIIPPELRESGQEVETYSDSIELIEFQGIKGDIMSSTWSDGAF349tr A0A10PFQD6 A0A10PFQD6 BACLITFFSEREFYAHFGLPIIPPELRESGQEVETYSDSIELIEFQGIKGDIMSSTWSDGAF	tr A0A077UIA1 A0A077UIA1 9STAP	QYDSEAKIYEHFNVNFIPPAM	RED GSEFDKDLSNIITLDDIN GDIHMHTTISDGAF	347
AAP11460.1 TFETEEDFFAHFGLPIPFEVRED GKEIELIKEYPNLLQFSDIC GDILMNITWSDCAF 352 triA0A5M853G5 A0A5M853G5_BACAT TFPSEREFYAHFGLPLIPFEVRED GGEVETVHDGIELIETHDIK GDILMNISWSDCAF 349 triA0A642Jr09 A0A642Jr09_BACMO TFPSEREFYAHFGLPLIPPELRES GGEVETVNDSIELIEHEDIK GDILMNISWSDCAF 349 triA0A642Jr109 A0A6412Jr09_BACMO TFPSEREFYAHFGLPLIPPELRES GGEVETVNDSIELIEHEDIK GDILMNISWSDCAF 349 triA0A642J4104H3 BACSMO TFPSEREFYAHFGLPLIPPELRES GGEVETVSDSIELIEHGOIK GDILMNISWSDCAF 349 triA0A644J4H3 A0A644J4H3 BACSMO TFPSEREFYAHFGLPLIPPELRES GGEVETVSDSIELIEHGOIK GDL MISTWSDCAF 3492SDM triA0A644J4H3 A0A6440HH3 BACSMO TFPSEREFYAHFGLPLIPPELRES GGEVETVSDSIELIEHGOIK GDL MISTWSDCAF 3492SDM triA0A6400MF9 A0A6410MH59 DAGAK0AMFP TFPSEREFYAHFGLPLIPPELRES GGEVETVSDSIELIEFGQIK GDL MISTWSDCAF 349 triA0A10WE37 A0A410WE37_BACVA TFPSEREFYAHFGLPLIPPELRES GGEVETVSDSIELIEFGQIK GDLIMNISTWSDCAF 349 triA0A10PFQD6 A0A10PFQD6_BACLI TFPSEREFYAHFGLPLIPPELRES GGEVETVSDSIELIEFGQIK GDLIMNISTWSDCAF 349 <tri::::::::::::::::::::::::::::::::::::< td=""><td>tr A0A2K4ADG4 A0A2K4ADG4_9STAP</td><td>QYDSETKIYEHFNVNFIPPAM</td><td>IRED GSEFDKDLSNIITLDDIN GDIHMHTTYSDGAF</td><td>347</td></tri::::::::::::::::::::::::::::::::::::<>	tr A0A2K4ADG4 A0A2K4ADG4_9STAP	QYDSETKIYEHFNVNFIPPAM	IRED GSEFDKDLSNIITLDDIN GDIHMHTTYSDGAF	347
tr A0A5M853G5 BACAT TFP5EQAFYAHFGLPLIPPEIRES GQEVDTYHDGIELIETHDIK GDIIMM SAWSDGAF 349 tr A0A6H2JTG9 A0A6H2JTG9 A0A6H2JTG9 A0A6H2JTG9 A0A6H2JTG9 A0A6H2JTG9 A0A6H2JTG9 A0A6H2JTG9 A0A6H2JTG9 A0A7H1CCW5 BAOTH1CCW5 BAOTH1CW5 BAOTH1CCW5 BAOTH	AAP11460.1		DED CRETET TREVENUL OFEDIO CDI UMU TEMEDOAE	352
tr A0AH1COUS A0AH1COUS_DEACH IFFSEREFIAREGELIFELESS GUEVET-INDSIELEENQIA GDIMESTWSDGAF 349 tr A0AH1COUS A0AH1COUS_DEACH TFPSEREFYAHFGLPLIPPELRES GQEVET-YDDSIELEENQIK GDIMESTWSDGAF 349 tr A0AH1COUS A0AH1COUS_DEACH TFPSEREFYAHFGLPLIPPELRES GQEVET-YDDSIELEENQIK GDIMESTWSDGAF 349 tr A0AH1COUS A0AH1COUS_DEACH TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELEENQIK GDIMESTWSDGAF 349 tr A0AH1COUS]A0AH1COUST_BACVA TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELEENQIK GDIMESTWSDGAF 349 tr A0AH1COUST A0AH1COUST_BACVA TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELEENGIK GDIMESTWSDGAF 349 tr A0AH1COUST A0AH1COUST_BACVA TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELIEPGQIK GDIMESTWSDGAF 349 tr A0AH1COUST BACH100E37 BACVA TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELIEPGQIK GDIMESTWSDGAF 349 tr A0AH1COUST BACH100E37 BACVA TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELIEPGQIK GDIMESTWSDGAF 349 tr A0AH1COUST BACH100E37 BACVA TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELIEPGQIK GDIMESTWSDGAF 349 tr A0AH1Q9FQD6 A0AH1Q9FQD6_BACLI TFPSEREFYAHFGLPLIPPELRES GQEVET-YSDSIELIEPGQIK GDIMESTWSDGAF 349 tr A0AH1Q9FQD6 A0AH1Q9FQD6_BACLI tr A0AH1COUST_STANSCAF 349 449		TFETEEDFFAHFGLPFIPPEV	KED GREIEL-IREIFRILIGFSDIG GDIARATIWSDGAF	
Lr A0A6M4 JMH3 A0A6M4 JMH3 BACSU TFPSEREFYAHFGLPLIPPEIRES GQEVET-YSDSIELIELGQIK GDL M STWSDGAF 34 92SDM tr D4F275 D4F275 _ BACNB TFPSEREFYAHFGLPLIPPEIRES GQEVETYSDSIELIEPGQIK GDL M STWSDGAF 34 9 tr A0A6H0WNP9 A0A6H0WNP9 _ 9BACI TFPSEREFYAHFGLPLIPPEIRES GQEVETYSDSIELIEPGQIK GDL M STWSDGAF 34 9 tr A0A10WE37 A0A6H0WNP3 _ BACVA TFPSEREFYAHFGLPLIPPEIRES GQEVETYSDSIELIEPGQIK GDLIMM STWSDGAF 34 9 tr A0A10WE37 A0A410WE37 _ BACVA TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDLIMM STWSDGAF 34 9 tr A0A10WE37 A0A410WE37 _ BACVA TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDLIMM STWSDGAF 34 9 tr A0A10WE37 A0A10WE37 _ BACVA TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDLIMM STWSDGAF 34 9 tr A0A1Q9FQD6 A0A1Q9FQD6 _ BACLI TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDLIMM STWSDGAF 34 9 tr A0A1Q9FQD6 A0A1Q9FQD6 _ BACLI TFPSEREFYAHFGLPLIPPELRES GVEVETYSDSIELIEPGQIK STWSDGAF 34 9 tr A0A1Q9FQD6 A0A1Q9FQD6 _ BACLI tr :	tr A0A5M8S3G5 A0A5M8S3G5 BACAT	TFETEEDFFAHFGLPFIPPEV TFPSEQAFYAHFGLPLIPPEI	RES GQEVDTYHDGIELIETHDIK GDIHMESAWSDGAF	349
tr D4FZT5 D4FZT5_BACNB TFPSEREFYAHFGLPLIPPEIRES GQEVETYSDSIELIEPGQIK GDINM:STWSDGAF 349 tr A0A6H0WNP9 A0A6H0WNP9]BACI TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDINM:STWSDGAF 349 tr A0A10WE37 A0A410WE37 BACVA TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDINM:STWSDGAF 349 tr A0A10WE37 A0A410WE37 BACVA TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDINM:STWSDGAF 349 tr A0A109F0D6 A0A109F0D6_BACLI TFPSEREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDINM:STWSDGAF 349 :* :: :* :: :* :: :* :: :*	tr A0A5M8S3G5 A0A5M8S3G5 BACAT tr A0A6H2JTG9 A0A6H2JTG9 BACMO tr A0A7H1CCW5 A0A7H1CCW5 9BACT	TFETEEDFFAHFGLPFIPPEV TFPSEQAFYAHFGLPLIPPEL TFPSEREFYAHFGLPLIPPEL TFPSEREFYAHFGLPLIPPEI	RES GQEVDTYHDGIELIETHDIK GDIHMI SAWSDGAF RES GQEVETYNDSIELIEHEQIK GDIHMI STWSDGAF RES GOEVETYNDSIELIEHEOIK GDIHMMSTWSDGAF	349 349 349
tr A0A6H0WNP9]A0A6H0WNP9_9BACI TFP5EREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDIMMSTWSDGAF 349 tr A0A10WE37 A0A10WE37_BACVA TFP5EREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDIMMSTWSDGAF 349 tr A0A10PF06 A0A10PF06_BACLI TFP5EREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDIMMSTWSDGAF 349 tr A0A10PF06 A0A10PF06_BACLI TFP5EREFYAHFGLPLIPPELRES GQEVETYSDSIELIEPGQIK GDIMMSTWSDGAF 349 tr .: .: * *: .: .: *: *: 349	tr A0A5M8S3G5 A0A5M8S3G5_BACAT tr A0A6H2JTG9 A0A6H2JTG9_BACMO tr A0A7H1CCW5 A0A7H1CCW5 BACI tr A0A6M4JMH3 A0A6M4JMH3_BACSU	TETTEEDFFAHFGLPFIPPEV TFPSEQAFYAHFGLPLIPPEI TFPSEREFYAHFGLPLIPPEL TFPSEREFYAHFGLPLIPPEL	RES GOEVETYHOGIELIETHDIK GDIAMMISAWSDGAF RES GOEVETYHOGIELIETHDIK GDIAMMISAWSDGAF RES GOEVETYNDSIELIEHOQIK GDIAMMISTWSDGAF RES GOEVETYSDSIELIELGOIK GDIAMMISTWSDGAF	349 349 349 34 9 34 9
tr a0a1Q9FQD6 a0a1Q9FQD6_BACLI TFFSEREYAPFSEREYAPFSERES GQEVETYSDSIELEFGGIQ GDLMM STWSDGAF 349 :* :: :: * * :***	tr AOA5M853G5 AOA5M853G5_BACAT tr AOA6H2JTG9 BACAM tr AOA7H1CCW5 AOA7H1CCW5_ 9BACI tr AOA7H1CCW5 AOA5H4JCW5_9BACI tr AOA5M4JMH3 AOA6M4JMH3_BACSU tr D4F2T5 D4F2T5_BACNB	TFFTEEDFFAHFGLPFIPEU TFFSEQAFYAHFGLPLIPPEI TFFSEREFYAHFGLPLIPPEL TFFSEREFYAHFGLPLIPPEI TFFSEREFYAHFGLPLIPPEI	RES GQEVETYHOSIELIETHDIK GDIHMISAWSDGAF RES GQEVETYHOSIELIEHQIK GDIHMISAWSDGAF RES GQEVETYHOSIELIEHQIK GDIHMISTWSDGAF RES GQEVETYSDSIELIELGQIK GDIHMISTWSDGAF RES GQEVETYSDSIELIELGQIK GDIHMISTWSDGAF	349 349 349 349 349 349 349
:* .: :.: * * :**.♥*: ::	tr AOA5M853G5 AOA5M853G5_BACAT tr AOA6H2JTG9 BACAT tr AOA6H2JTG9 BACAT tr AOA7H1CCM5 AOA7H1CCM5 BBACI tr AOA6M4JMH3 AOA6M4JMH3_BACSU tr AO4F2T5 D4F2T5_BACNB tr AOA6H0WNP9 AOA6H0WNP9 BBACI tr AOA6H0WNP9 AOA410WF3 BACI Tr AOA6H0WNP9 AOA610WF3 BACI Tr AOA6H0WF3 AOA410WF3 AOA410WF3 BACI Tr AOA6H0WF3 AOA6H0WF3 AOA6H0WF3 BACAF Tr AOA6H0WF3 AOA6H0WF3 AOA6H0WF3 AOA6H0WF3 BACAF Tr AOA6H0WF3 AOA	TFFETEDFFAHFGLPIPEV TFFSEQAFYAHFGLPLIPEL TFPSEREFYAHFGLPLIPEL TFPSEREFYAHFGLPLIPEL TFPSEREFYAHFGLPLIPEL TFPSEREFYAHFGLPLIPEL TFPSEREFYAHFGLPLIPEL	RED GREIED-THEIRBLUEYSDUG GOLANN INNSOAP RES GQEVDTYHDGIELIETHDIK GDIANN SAWSDGAP RES GQEVETYNDSIELIEHQIK GDIANN STWSDGAP RES GQEVETYNDSIELIEHQIK GDIANN STWSDGAP RES GQEVETYSDSIELIEPGQIK GDIANN STWSDGAP RES GQEVETYSDSIELIEPGQIK GDIANN STWSDGAP RES GQEVETYSDSIELIEPGQIK GDIANN STWSDGAP	349 349 349 349 349 349 349 349
	tr AOA5M853G5 AOA5M853G5_BACAT tr AOA6H2JTG9 BACAT tr AOA6H2JTG9 BACAT tr AOA6H2JTG9 AOA6H2JTG9 BACAT tr AOA6M4JMH3 AOA6M4JMH3_BACSU tr D4F2T5 D4F2T5_BACNB tr AOA6H0MNP9 AOA6H0MNP9 BBACI tr AOA6H0MNP9 AOA6H0MNP9 BBACI tr AOA6H0M27 AOA410WE37_BACVA tr AOA109FQD6 AOA109FQD6_BACLI	TFFTEDFAHRGLPIPEV TFFSEQAFYAHRGLPIPEL TFFSEREFYAHRGLPIPEL TFFSEREFYAHRGLPIPEL TFFSEREFYAHRGLPIPEL TFFSEREFYAHRGLPIPEL TFFSEREFYAHRGLPIPEL TFFSEREFYAHRGLPIPEL	RES GOEVDTYHGIRLIETHDIK GDIAMAINABGAP RES GOEVDTYHDGIRLIETHDIK GDIAMASAWSDGAP RES GOEVETYNDSIELIEHOQIK GDIAMASTWSDGAP RES GOEVETYNDSIELIEHOQIK GDIAMASTWSDGAP RES GOEVETYSDSIELIEPGQIK GDIAMASTWSDGAP RES GOEVETYSDSIELIEPGQIK GDIAMASTWSDGAP RES GOEVETYSDSIELIEPGQIK GDIAMASTWSDGAP RES GOEVETYSDSIELIEPGQIK GDIAMASTWSDGAP	349 349 349 349 349 349 349 349 349 349

Palanivelu; IJBCRR, 30(3): 33-62, 2021; Article no.IJBCRR.70674



Fig. 13. MSA of the DNA polymerases X from different organisms

A0A3P4ARX8 THETH DNA polymerase beta Thermus thermophilus PCQ20452.1 DNA polymerase/3'-5' exonuclease PolX [Klebsiella pneumoniae] COE05251.1 DNA polymerase/3'-5' exonuclease PolX [Staphylococcus warneri] A0A0D6HTC1_STAAU DNA polymerase beta, Staphylococcus aureus AAW38032.1 DNA-dependent DNA polymerase family X, Staphylococcus aureus subsp. aureus A0A4V0A0W1 STAHY DNA polymerase beta, Staphylococcus hyicus A0A348BAD9_9STAP DNA polymerase beta, Staphylococcus argenteus A0A077UIA1_9STAP DNA polymerase beta, Staphylococcus schweitzeri A0A2K4ADG4 9STAP DNA polymerase beta, Staphylococcus schweitzeri AAP11460.1 DNA polymerase X family, Bacillus cereus A0A5M8S3G5 BACAT DNA-directed DNA polymerase, Bacillus atrophaeus A0A6H2JTG9_BACMO DNA-directed DNA polymerase, Bacillus mojavensis A0A7H1CCW5_9BACI DNA polymerase/3'-5' exonuclease PolX, Bacillus halotolerans A0A6M4JMH3_BACSU DNA-directed DNA polymerase, Bacillus subtilis D4FZT5 BACNB DNA-directed DNA polymerase. Bacillus subtilis subsp. Natto A0A6H0WNP9 9BACI DNA-directed DNA polymerase, Bacillus tequilensis A0A410WE37_BACVA DNA-directed DNA polymerase, Bacillus vallismortis A0A1Q9FQD6 BACLI DNA-directed DNA polymerase, Bacillus licheniformis A0A097J243_9CAUD Homing endonuclease, Enterobacteria phage RB5

4.1.1 Active Site Analyses of PR Exonuclease Domain in Bacterial DNA polymerases X

The PHP domain has been shown to possess the 3'-5' exonuclease activity and perform the PR function in DNA polymerases X [16-19]. Fig. 13 shows the MSA of the DNA polymerases X form different organisms. Unlike the DNA polymerases I and II, only there are a few conserved motifs among them, suggesting that they are highly diverged during evolution. However, the active site regions are completely conserved (Fig. 13). The X polymerases follow the completely conserved pattern -GSKD H⁵xxxx**RQ**¹IAKERGERISEY¹³GVsuggesting the repair is strictly template strand based with the template-binding YG pair with the NTP selecting invariant basic amino acid H at -5 from the catalytic R. (In the DNA polymerase I family, the polymerase catalytic region is slightly different from the X polymerases and is -SEQ $R^4xxxKAINFGLIY^{766}GI/L/M-$). However, the RQ and YG pairs are followed by I//V as in DNA polymerases I. The metal-binding site -DLD- is completely conserved in all X polymerases (highlighted in light green). Within the polymerase domain а HNH homing endonuclease domain is also observed as suggested by Nagpal and Nair [3]. (The active site region of the HNH homing endonuclease domain of Enterobacteria phage RB5 is highlighted in grey). Though the HH---N are completely conserved in all, the last H is not conserved and is replaced with another basic amino acid, K/R in DNA polymerases X, but preceded by an E in all the cases (Fig. 13).

The DNA polymerase domain in X polymerases is located from amino acids 1 to 315 and the 3'-5' exonuclease PR domain is located from amino acids 337-570 (numberings from Bacillus subtilis enzyme and highlighted in different colours) (Fig. 13). SDM experiments have shown that the polymerase and PR activities were independent of each other. For example, in a double mutant where both the Ds are replaced by As $(^{193}\text{DLD}\rightarrow\text{ALA}),$ abolished the polymerase activity [18] whereas in a similar experiment where both the Hs are replaced by As $(^{339}\text{HMH}\rightarrow\text{AMA})$, the exonuclease activity was abolished. The metal-dependent exonuclease activity is further confirmed by deletion mutants too. For example, in a deletion mutant 316-570 (mutant ΔPHP), the exonuclease activity was completely abolished [18].

DNA polymerases X of *Thermus thermophilus* was extensively studied by Nakane et al [19]. It

possessed both the DNA polymerase (ttPOLXc) and the exonuclease activities belonging to the PHP family, as two independent domains. The enzyme was subjected to SDM experiments by them. The ttPOLXc domain showed Mg²⁺dependent DNA polymerase activity but no 3'-5' exonuclease activity. They could hardly detect any 3'–5'exonuclease activity of the mutant enzymes with H³⁴⁴ \rightarrow A, H³⁷⁴ \rightarrow A, H⁴⁶⁸ \rightarrow A and D⁵²⁹ \rightarrow A mutations. Therefore, they concluded that His³⁴⁴ (-QVH³⁴⁴-), His³⁷⁴ (-DH³⁷⁴SP-), His⁴⁶⁸ (-AH⁴⁶⁸P-) and Asp⁵²⁹ (-D⁵²⁹AH-) are the most important residues for the 3'–5'exonuclease activity (marked in the margin).

The corresponding amino acids in *B. subtilis* enzyme are placed in the active site (marked in red and highlighted in blue) (Fig. 13). Nakane et al [19] also found that the POLX-core and PHP domains interacted with each other and a mixture of the two domains had Mn^{2+} -dependent 3'–5' exonuclease activity. Importantly, the DNA polymerase exhibited Mg^{2+} -dependent activity and the PHP domain exhibited no exonuclease activity in the presence of Mg^{2+} or Mn^{2+} but exhibited exonuclease activity only with Zn^{2+} , further corroborating the involvement of Zn^{2+} in the PR functions.

The Zn²⁺ atom is coordinated by three invariant His residues, and the fourth ligand is occupied by a water molecule (Fig. 14). Under the polymerase assay conditions, in the presence of Mn^{2+} the Q^{342→}A, D^{349→}A, E^{413→}A and H^{531→}A mutant exhibited stronger 3'-5' exonuclease activity.

The *Bacillus subtilis* DNA polymerases X was also analyzed by SDM experiments by Banos et al. [18]. They found that H^{339} and H^{341} (highlighted in light green) in the HxH motif are shown to be essential for the PR exonuclease active site in the DNA polymerase X. Fig. 14 shows the proposed active site at the PR site in *Bacillus subtilis* DNA polymerases X. The active site amino acids are proposed based on the SDM analysis of *T. thermophilus* and *B. subtilis* X polymerases. They follow the general active site pattern as $-HxH\rightarrow E\rightarrow H\rightarrow D^-$.

4.2 Bacterial YcdX class of Exonucleases

The second group of enzyme is the YcdX types which exhibits a phosphoesterase activity. It is interesting to note, these phosphoesterases do not form a part of polymerases, but also belong to the PHP superfamily because of their structural and sequence similarities [5]. Like the other PHP enzymes, the YcdX also consists of four conserved sequence motifs that contain invariant histidine and aspartate residues, which are implicated in metal ion coordination and catalysis. X-ray crystallographic data on the E.

coli YcdX is available [5]. The X-ray crystallographic studies have shown that the catalytic site of YcdX of E. coli consists of three Zn atoms and is similar to those enzymes which hydrolyze phosphoester bonds.



Fig. 14. Proposed amino acids at the active site of DNA polymerases X from B. Subtilis

CLUSTAL O (1.2.4) MSA of bacterial YcdX phosphoesterases of PHP superfamily

MYPVDI HNH TVASTHAYSTLSDYIAQAKQKGIKLFA TDUG DMEDA PH WHFINMRIWP MYPVDI HNH TVASTHAYSTLSDYIAQAKQKGIKLFA TDUG DMEDA PH WHFINMRIWP MYPVDI HNH TVASTHAYSTLSDYIAQAKQKGIKLFA TDUG PMEDA PH WHFINMRIWP	60 60 60 60 60 60 60
MY PVDI MMH TVASTHAYSTLSDYIAQARQKCIKLFA TDBG POMEDA HMWHFINMRIWP MY PVDI MMT TVASTHAYSTLSDYIAQARQKGIKLFA TDBG POMEDA HMWHFINMRIWP	60 60 60 60
RVD DQVGSLR II AN KNUGGE IDCSGKMEFSLDLI IA THE PVPAPHDKATNITAMIAA RVDGVGLIR II AN KNUGGE IDCSGKMEFSLDLI IA THE PVPAPHDKATNITQAHIAT RVDGVGLIR II AN KNUGGE IDCSGKMEFSLDLI IA THE PVPAPHDKATNITQAHIAT	120 120 120 120 120 120
RVDGVGILRFICAN KNUGGEIGESGKMFDSLDLIAGHLFVEPAPHDKATNTQANIAT RVDGVGILRFICAN KNUGGEIGESGKMFDSLDLIAGHLFVEPAPHDKATNTQANIAT RVDGVGILRFICAN KNUGGEIGESGKMFDSLDLIAGHLFVEPAPHDKATNTQANIAT RVDGVGILRFICAN KNUGGEIGESGKMFDSLDLIAGHLFVEPAPHDKATNTQANIAT RVDGVGILRFICAN KNUGGEIGESGKMFDSLDLIAGHLFVEPAPHDKATNTQANIAT	120 120 120 120 120 120
MASGHVHMY UP COMPKYPYDPKATAEAAAEYQVALEINNSSTLHSRKGSEDDICRALAAAY IASGNVHII D COMPKYEIDVAVAEAAAKYQVALEINNSSTLHSRKGSEDICRAVAAAY IASGNVHII D PONPKYEIDVKAVAEAAAKYQVALEINNSSTLHSRKGSEDDICRAVAAAY IASGNVHII D PONPKYEIDVKAVAEAAAKYQVALEINNSSTLHSRKGSEDDICRAVAAAY IASGNVHII D PONPKYEIDVKAVAEAAAKHQVALEINNSSTLHSRKGSEDDICRAVAAAY	180 180 180 180 180 180
IASCNYH II EPON PKYE IDVKAVAEAAAKHQVALE INNS SFLHSRKSSEDNCRAVAAAY IASCNYH II EPON PKYE IDVKAVAEAAAKHQVALE INNS SFLHSRKSSEDNCRAVAAAY	180 180 180 180 180 180
RAGGWALG OUT AFTMGEFECCKILDAVDF9CERLLNVSPRLLSFLESRGMSPIA RAGGWALG OUST AFTMGFFECKLIDAVDF9CERLINVSPRLLSFLESRGMSPIA RAGGWALG OUST AFTMGFFECLKILDAVDF9LERLINVSPRLLNFLESRGMAPIA RAGGWALG OUST AFTMGFFECLKILDAVDF9PERLINVSPRLLNFLESRGMAPIA RAGGWALGOUST AFSMGFFECLKILDAVDF9PERLINVSPRLLNFLESRGMAPIA RAGGWALGOUST AFSMGFFECLKILDAVDF9PERLINVSPRLLNFLESRGMAPIA	240 240 240 240 240 240
RDAGGWALGS STAFTMGFFECLKILDAVDFPPERILNYSPRLLNPLESRGMAPIA RDAGGWALGS STAFTMGFFECLKILDAVDFPPERILNYSPRLLNPLESRGMAPIA RDAGGWALGS STAFTMGFFECLKILDAVDFPPERILNYSPRLLNPLESRGMAPIA RDAGGWALGS STAFTMGFFECLKILDAVDFPPERILNYSPRLLNPLESRGMAPIA RDAGGWALGS STAFTMGFFECLKILDAVDFPPERILNYSPRLLNPLESRGMAPIA	240 240 240 240 240 240 240
FFADL 245	
	MYPUD INH TVASTHAYSTLSDYIAQAKQKGIKLEA TDIG DHEDA HEWHFINNEN MYPUD INH TVASTHAYSTLSDYIAQAKQKGIKLEA TDIG MYPUD INH TVASTHAYSTUD TO TAGAYAY MYPUD INH TVASTHAYSTUD TAGAYAY MYPUD INH TYSTUD TAGAYAY MYPUD INH TYSTUD TAGAYAYAY MYPUD

Fig. 15 MSA of YcdX phosphoesterases belonging to PHP superfamily

B1EL68_ESCAT, Escherichia albertii Q0T606 YCDX_SHIF8, Shigella flexneri A0A0A6ZQ83_SHIDY, Shigella dysenteria A0A5Q3TIN6_9ENTR, Salmonella sp. Q3Z392|YCDX_SHISS, Shigella sonnei A0A7D6YTS2 9ENTR, Enterobacter hormaechei
> A0A2B7MK98_9ESCH, Escherichia marmotae A0A6I1J2J7 9ENTR. Enterobacteriaceae bacterium P75914|YCDX_ECOLI, Escherichia coli (strain K12) Q31Z90|YCDX_SHIBS, Shigella boydii E7T3G3 9ENTR, Shigella boydii ATCC 9905 A0A163VYN4 KLEOX, Klebsiella oxytoca



Fig. 16. Proposed amino acids at the active site of E. coli YcdX phosphoesterase

They have found that the YcdX had an unusual type of topology with a $\beta_7 \alpha_7$ barrel type of structure whose C-terminal side had a deep cleft that contained three metal-binding sites which were ligated to the imidazole group of residues His⁷, His⁹ and His¹⁵ from motif I; His⁴⁰ from motif I; His¹⁰¹ from motif III, His¹³¹ and His¹⁹⁴ from motif IV; as well as to the carboxylate group of Glu⁷³, Glu¹⁵⁶ and Asp¹⁹², the latter belonged also to Motif IV. Fig. 15 shows the MSA of the YcdX phosphoesterases which belong to PHP superfamily with the active site pattern as - $HxH\rightarrow E\rightarrow H\rightarrow D$ -. Furthermore, these enzymes are almost completely conserved in their entire sequence only with a few changes. A probable HNH motif is also observed (highlighted in light red) suggesting their possible lateral transfer between genomes of bacterial species.

4.2.1 Active Site Analyses of the YcdX Phosphoesterase

Out of the three Zn atoms, the Zn2 which is coordinated by H⁷, H⁹, E⁷³ and D¹⁹² and also coordinated to a water molecule could possibly involve in the excision of the nucleotide (Fig. 16). Zn1 is the "high-affinity" site occupied by zinc in the native structure bound to H^{15} , H^{40} , H^{194} and H₂0 (marked in red) could likely play the structural role. The proposed active site is based on the X-ray crystallographic data. The catalytic mechanism of YcdX may proceed through the nucleophilic attack of the susceptible phosphorus atom by the water molecule bridging Zn2 and Zn3, which is presumably a hydroxide ion [5 and references therein]. Interestingly, no Ma² or other divalent metal ions are found in the X-ray crystallographic data. The conserved amino acids, viz. H^{131} of the motif – $SH^{131}P$ - and E^{156} of the motif - $E^{156}IN$ - are based on the sequence similarity to DNA polymerases Х.

4.3. Intrinsic PR Activity in the Replicative DNA Polymerases III

As shown in the earlier section, in the bacterial DNA polymerases III the PR function is performed by an independent subunit of the enzyme, (i.e.), by the ε -subunit which is a part of the MEC. In the MEC, the catalytic α -subunit of the polymerase III performs the polymerase function, whereas the ϵ -subunit performs the PR However, a novel PHP type PR function. exonuclease activity was reported in the DNA polymerase III α-subunit itself from the thermophilic bacterium, Thermus thermophilus [4]. They have reported that the PR exonuclease domain was in the same polypeptide as the polymerase domain and was located in the Nterminal region of the polymerase. They also found that the PR exonuclease was a typically a Zn²⁺ dependent enzyme, as the Zn²⁺ chelator Ophenanthroline inhibited the enzyme activity drastically even in the presence of 10 mM Mg²⁺. Therefore, it was suggested that the functions of these two exonuclease activities could be complementary, i.e., the PHP enzyme might be more active on mismatches not preferred by the ε-exonuclease. From the MSA analysis it is clear that the active site of this type of enzyme consists of 4 motifs similar to other exonucleases of the PHP superfamily (Fig. 17). For example, the motif I has a dyad of histidines, which are separated by a single amino acid as HxH, which apparently coordinating the metal ion zinc. Motif II has an Asp/Glu, motif III has a His and motif IV has an Asp residue. It was proposed that the motifs II to IV might be involved in catalysis by participating in proton transfer and/or through metal ion coordination. The MSA shows the general active site pattern $-HxH\rightarrow E\rightarrow H\rightarrow D$ -, which is very similar to the pattern found in other PHP enzymes like DNA polymerases X and YcdX. The presence of the highly conserved template-binding YG pair (highlighted in yellow) suggests that this PR function is templatedependent activity.

Mix and Match analysis of the α -subunit of DNA polymerases III of *E. coli* and other mesophilic enzymes along with the thermophilic counterpart is shown in Fig. 15 (only the N-terminal PHP domain is shown here). It shows that the mesophilic and thermophilic enzymes are not highly conserved but for the active site amino acids. However, some of the active site amino acids are not completely aligned in both the group of enzymes. Such a shift in the active site regions is already reported for invertases by Palanivelu [20]. Like other PHP enzymes this enzyme is also a strictly template–dependent enzyme as the template-binding YG pair is

completely conserved in all (YK in the thermophilic enzyme) (highlighted in yellow). The metal-binding regions are highlighted in light green (Fig. 17).

However, the MSA shows that the conserved amino acids that are found in *T. thermophilus* enzyme are also found in other α -subunits of the mesophilic DNA polymerases III, but with slight modifications, e.g., **HLH** \rightarrow **HL***R*; **D**HG \rightarrow **D**FT; **EMG** \rightarrow **EM**/LT; **DAR** \rightarrow **DAH** (similar conserved sequences in mesophiles are shown in red) (Fig. 16). The first and the last triads are highly conserved with similar amino acids motif. These data suggest that the active site amino acids are still intact in the α -subunits of the DNA polymerases III in both the groups of organisms.

CLUSTAL O (1.2.4) Mix and Match analysis of the α -subunit of DNA polymerases III

tr Q72GF2 Q72GF2_THET2 tr A0A071LXE8 A0A071LXE8 9ENTR sp P10443 DP03A ECOLI tr A0A7H8SQF7 A0A7H8SQF7_SHISO tr A0A0M7N2M0 A0A0M7N2M0_9BURK tr A0A6B8XLK6 A0A6B8XLK6_SHIBO tr A0A3C0H056 A0A3C0H056_9ENTR tr A0A47ML45 A0A447ML45_KLUIN tr A0A2F5GML4 A0A2F5GML4_9ENTR	MGRKLRFA HLPQHTQFSLLDGAAKLSDLLKWVKETTPEDPALAM TD HGNLFGAVEFYK -MAEPRFVHLSIHSDYSMIDGLAKUGPLVKKAASLMPAMAITDFTNLCGLVKFYGGA -MSEPRFVHLFVHSDYSMIDGLAKTAPLVKKAAALGMPALAITDFTNLCGLVKFYGAG -MSEPRFVHLFVHSDYSMIDGLAKTAPLVKKAAALGMPALAITDFTNLCGLVKFYGAG -MSEPRFVHLFVHSDYSMIDGLAKTAPLVKKAAALGMPALAITDFTNLCGLVKFYGAG -MAEPRFVHLFVHSDYSMIDGLAKTAPLVKKAASLGMPALAITDFTNLCGLVKFYGAG -MAEPRFVHLFVHSDYSMIDGLAKTAPLVKKAASLGMPALAITDFTNLCGLVKFYGAG -MAEPRFVHLFVHSDYSMIDGLAKTAPLVKKAASLGMPALAITDFTNLCGLVKFYGAG -MAEPRFVHLFVHSDYSMIDGLAKTCPLVKKAASLGMPALAITDFTNLCGLVKFYGAG -MAEPRFVHLFVHSDYSMIDGLAKTCPLVKKAASLGMPALAITDFTNLCGLVKFYGAG	60 57 57 57 57 57 57 57 57
tr A0ASB7XQ40 A0ASB7XQ40_9ENTR tr A0ASB7XQ40 A0ASB7XQ40_9ENTR tr A0A2N0CMK4 A0A2N0CMK4_9ENTR tr A0A5D4YAD4 A0A5D4YAD4_9ENTR tr A0A3S5XXT2 A0A3S5XXT2_LELAM	MAEPRFVHLFVHSDYSMIDGLAKTGPLVKKAAALGMPALAITFTHLCGLVKFYGTA -MAEPRFVHLFVHSDYSMIDGLAKTGPLVKKAASLGMPALAITFTHLCGLVKFYGTA -MAEPRFVHLFVHSDYSMIDGLAKTGPLVKKAASLGMPALAITFFTHLCGLVKFYGTA -MAEPRFVHLFVHSDYSMIDGLAKTGPLVKKAASLGMPALAITFFTHLCGLVKFYGAA : ** **: **: *:::*:** ** . *:* . : **:*:** . ** ** . 12MGIKPILGYEAYVAAESRFDRKRGKGLD <u>GC</u> YFHLTLLAKDFFGYQNIVRLASRAYLEG	57 57 57 57 57 57
tr A0A071LXE8 A0A071LXE8_9ENTR	HGVGMKPIVGADLQMTSEISGDENTQLTVLAVNNTGYQNITLLISHAYQRG	108
sp P10443 DPO3A_ECOLI	HGAGIKPIVGADFNVQCDLLGDELTHLTVLAANNTGYQNLTLLISKAYQRG	108
tr A0A7H8SQP7 A0A7H8SQP7_SHISO	HGAGIKPIVGADFNVQCDLLGDELTHLTVLAANNTGYQNITLLISKAYQRG	108
tr A0A0M7N2M0 A0A0M7N2M0_9BURK	HGAGIKPIVGADFNVQCDLLGDELTHLTVLAANNTGYQNITLLISKAYQRG	108
tr A0A6B8XLK6 A0A6B8XLK6_SHIBO	HGAGIKPIVGADFNVQCDLLGDELTHLTVLAANNTGYQNITLLISKAYQRG	108
tr A0A3C0H056 A0A3C0H056_9ENTR	HGAGMKPIIGADFNVHNPIMGDELNELTVLAADNTGYQNITLLISRAYQRG	108
tr A0A447ML45 A0A447ML45 KLUIN	HGAGIKPIVGADFHVQSELFGDELTQLTVLAANNVGYQNITLLISKAYQRG	108
tr A0A2P5GML4 A0A2P5GML4 9ENTR	HGAGIK PIVGADFHVONDLLADEVTOLTVLAMNNVGYONITLLISRAYORG	108
tr A0A5R9ING3 A0A5R9ING3 9ENTR	HGAGMK PVIGAD FHVOSELLADEVTOITVLAMNNTGYONITLLISRAYORG	108
tr A0A5B7X040 A0A5B7X040 9ENTR	HGAGMKPVIGADFHLOSELLADEMTOITVLAMNNTGYONITLLISRAYORG	108
tr A0A2N0CMK4 A0A2N0CMK4 9ENTR	HGSGLKPIVGADFHVOCDLIGDELTOISVLAMNNTGYONITLLSRAYORG	108
tr A0A5D4YAD4 A0A5D4YAD4 9ENTR	HGSGLKPIVGADFHVOCDLIGDELTOISVLAMNNTGYONITLLSRAYORG	108
tr A0A3S5XXT2 A0A3S5XXT2_LELAM	HGAGLKPI I GADFHVOSELLGDENTOI SVLAMNNTGYONI TLL SRAYORG	108
_	*:**::* : : : : : : : : : : : : : : : :	
tr Q72GP2 Q72GP2_THET2	FYE-KPRIDREILREHAEGLIALSGCLGAEIPQFILQDRLDLAEARLNEYLSIFKDRFFI	179
tr A0A071LXE8 A0A071LXE8_9ENTR	YGDAGPWIDREWLVELSEGLIILSGGRMGEVGKAILRGNDVQVEQSLAFWKQHFPDRFYL	168
sp P10443 DPO3A_ECOLI	YGAAGPIIDRDWLIELNEGLILLSGGRMGDVGRSLLRGNSALVDECVAFYEEHFPDRYFL	168
tr A0A7H8SQP7 A0A7H8SQP7_SHISO	YGAAGPIIDRDWLIELNEGLILLSGGRMGDVGRSLLRGNSALVDECVAFYEEHFPDRYFL	168
tr A0A0M7N2M0 A0A0M7N2M0_9BURK	YGAAGPIIDRDWLIELNEGLILLSGGRMGDVGRSLLRGNSALVDECVAFYEEHFPDRYFL	168
tr A0A6B8XLK6 A0A6B8XLK6_SHIBO	YGAAGPIIDRDWLIELNEGLILLSGGRMGDVGRSLLRGNSALVDECVAFYEEHFPDRYFL	168
tr A0A3C0H056 A0A3C0H056 9ENTR	YGAAGPFIDLDWLAEHRDGLILLSGARKGDVGKSLLRGNMALVDQCLSFYQQHFADRFYL	168
tr A0A447ML45 A0A447ML45 KLUIN	YGAAGPIIDRDWLVELKEGLILLSGARMGDVGRALLRGNMALVEOCAEFYETHFPNAYFL	168
tr A0A2P5GML4 A0A2P5GML4 9ENTR	YGALGPWIDRDWLAELGEGLILLSGGRLGDVGRSVMRGNNTLVDQCLSFYETHFPDRFYL	168
tr A0A5R9ING3 A0A5R9ING3 9ENTR	YGPQGPWIDREWLAELNEGLLLISGGRMGDIGRCLLRGNTALVEQCVDFYKEYFPDRFYL	168
tr A0A5B7X040 A0A5B7X040 9ENTR	YGPOGPWIDREWLAELNEGLLLISGGRMGDIGRCLLRGNTALVEOCVDFYKEYFPDRFYL	168
trla0a2N0CMK4LA0a2N0CMK4 9ENTR	YGAAGPWIDREWLAELNEGLLLISGGRIGDVGKSLLRGNSALVDOCVAFYEEHFADRFYL	168
trla0a5D4YAD4LA0a5D4YAD4 9ENTR	YGAAGPWIDEEWLAELNEGLLLISGGELGDVGKSLLEGNSALVDOCVAEVEEHEADEEVL	169
+r a0a3s5xxr2 a0a3s5xxr2 I.ET AM	YCAACPWIDEDWIAEINECLLISCORMODUCKSLIECNNILVEOCVAFTEENFEDDEVI	169
CT AUADODAAT2 AUADODAAT2 LELIAM		100

Palanivelu; IJBCRR, 30(3): 33-62, 2021; Article no.IJBCRR.70674

+x1072CP21072CP2 THET2	ETONUCI DEOKKUNEUI KEENDKYCI CMUNTNOCUVURKTONE NUEUI I N TOSKSTI DOB	230
+~13030711VE913030711VE9 0ENTE		239
CI AURO / ILLEO AURO / ILLEO JENIK		220
+*170778800717077880077 84180		220
+ + 1 0 0 0 M 7 N 2 M 0 1 2 0 2 0 M 7 N 2 M 0 0 D U D K		220
		220
		220
		220
tr AUA44 / ML45 AUA44 / ML45 _ KLUIN	ELIRTGRODEEAY LHAAVRLAETRGLPVVATNDVRFLEPGDTDAREIRVATHDGFTLDDP	228
tr AUA2P5GML4 AUA2P5GML4_9ENTR	ELIRTGRODEESYLHAAIKLAEERGLPVVATNDVRFISADDIDAREIRVAIHDGFILDDP	228
tr A0A5R9ING3 A0A5R9ING3_9ENTR	ELIRTGRPDEENYLHAAVALAEEQGLPVVASNDVRFINADDFDAHEIRVAIHDGFTLDDP	228
tr A0A5B/XQ40 A0A5B/XQ40_9ENTR	ELIRTGRPDEENYLHAAVALAEEQGLPVVASNDVRFINADDFDALEIRVAIHDGFTLDDP	228
tr A0A2N0CMK4 A0A2N0CMK4_9ENTR	ELIRTGRPDEESYLHAAVALAEERGLPVVATNDVRFLEPDDFDAHEIRVAIHDGFTLDDP	228
tr A0A5D4YAD4 A0A5D4YAD4_9ENTR	ELIRTGRPDEESYLHAAVALAEERGLPVVATNDVRFLEPGDFDAHEIRVAIHDGFTLDDP	228
tr A0A3S5XXT2 A0A3S5XXT2_LELAM	ELIRTGRPDEENY <mark>LHA</mark> AVALAEQRGLPVVATNDVRFINAED <mark>PDAH</mark> EIRVAIHDGFTLDDP	228
	*: . * ::: . ** :** :** ::: * <mark>**</mark> *: :**:. ****	
// Brobable polymorope potive site region		
TT FTODADIE POLYTIELASE ACTIVE SILE TEGIOT		0.4.0
tr AUAU/ILXE8 AUAU/ILXE8_9ENTR		840
SP PI0443 DPO3A_ECOLI		840
tr AUA / H8SQP / AUA / H8SQP / SHISO	EFMAAVMTADMDNTEKVVGLVDECWHMGLKILPPDINSGLYHFHVNDDGEIVYGIGAIKG	840
tr AUAUM/N2MU AUAUM/N2MU_9BURK	EFMAAVMTADMDNTERVVGLVDECWHMGLKILPPDINSGLYHFHVNDDGEIVYGIGAIKG	840
tr A0A6B8XLK6 A0A6B8XLK6_SHIBO	EFMAAVMTADMDNTEKVVGLVDECWHMGLKILPPDINSGLYHFHVNDDGEIVYGIGAIKG	840
tr A0A3C0H056 A0A3C0H056_9ENTR	EFMAAVMTADMDNTDKVVGLVDECWHMG <mark>LKI</mark> LPPDINSGLYHFHVNDDGEI <mark>VYG</mark> IGAIKG	840
tr A0A447ML45 A0A447ML45_KLUIN	EFMAAVMTADMDNTEKVVGLVDECW <mark>H</mark> MG <mark>LKI</mark> LPPDINSGLYHFHVNQDGEI <mark>VYG</mark> IGAIKG	840
tr A0A2P5GML4 A0A2P5GML4_9ENTR	EFMAAVMTADMDNTEKVVGLVDEC <mark>WH</mark> MG <mark>LKI</mark> LPPDINSGLYHFHVNDDGEI <mark>VYG</mark> IGAIKG	840
tr A0A5R9ING3 A0A5R9ING3_9ENTR	EFMAAVMTADMDNTEKVVGLVDECW <mark>H</mark> MG <mark>LKI</mark> LPPDINSGQYHFHVNDDGEI <mark>VYG</mark> IGAIKG	840
tr A0A5B7XQ40 A0A5B7XQ40_9ENTR	EFMAAVMTADMDNTEKVVGLVDEC <mark>WF</mark> MG <mark>LKI</mark> LPPDINSGQYHFHVNDDGEI <mark>VYG</mark> IGAIKG	840
tr A0A2N0CMK4 A0A2N0CMK4_9ENTR	EFMAAVMTADMDNTEKVVGLVDECW <mark>H</mark> MG <mark>LKI</mark> LPPDINAGMYHFHVNDDGEI <mark>VYG</mark> IGAIKG	840
tr A0A5D4YAD4 A0A5D4YAD4_9ENTR	EFMAAVMTADMDNTEKVVGLVDECW <mark>H</mark> MG <mark>LKI</mark> LPPDINAGMYHFHVNDDGEI <mark>VYG</mark> IGAIKG	840
tr A0A3S5XXT2 A0A3S5XXT2_LELAM	EFMAAVMTADMDNTEKVVGLVDECW <mark>H</mark> MG <mark>LKI</mark> LPPDINAGMYHFHVNDDGEI <mark>VYG</mark> IGAIKG	840

// End of polymerase α -subunit		
tr1072GP21072GP2 THET2	EAAGERAYLLPDREVLLOGGOAGEAOEAVPE 2067	
trIA0A071LXE8IA0A071LXE8 9ENTR	1160	
splP1044310P033_ECOLT	1160	
trla0a7H8SOP7La0a7H8SOP7_SHISO	1160	
+ x A 0A 0M7N2M0 A 0A 0M7N2M0 QPUDK	1100	
	1160	
	1160	
LI AUASCUHUSO AUASCUHUSO 9ENIR		
LI AUA44 /ML45 AUA44 /ML45 KLUIN		
tr AUA2P5GML4 AUA2P5GML4_9ENTR	1160	
tr AUASK91NG3 AUASK91NG3_9ENTR	1160	
tr AUASB / XQ4U AUASB / XQ4U _ 9ENTR	1160	
tr AUA2NUCMK4 AUA2NUCMK4_9ENTR	1160	
tr A0A5D4YAD4 A0A5D4YAD4_9ENTR	1160	
tr A0A3S5XXT2 A0A3S5XXT2_LELAM	1160	

Fig. 17 Mix and Match analysis of the α-subunit of the thermophilic DNA polymerase III with mesophilic DNA polymerase III α-subunits

Q72GP2_THET2, Thermus thermophilus	A0A071LXE8_9ENTR, Mangrovibacter sp.
P10443 DPO3A_ECOLI, Escherichia coli (K12)	A0A7H8SQP7_SHISO, Shigella sonnei
A0A0M7N2M0_9BURK, Achromobacter sp.	A0A6B8XLK6_SHIBO, Shigella boydii
A0A3C0H056_9ENTR, Enterobacteriaceae bacterium	A0A447ML45_KLUIN, Kluyvera intermedia
A0A2P5GML4_9ENTR, Superficieibacter electus	A0A5R9ING3_9ENTR, Enterobacter sp.
A0A5B7XQ40_9ENTR, Leclercia adecarboxylata	A0A2N0CMK4_9ENTR, Cedecea lapagei
A0A5D4YAD4_9ENTR, Lelliottia nimipressurali;	A0A3S5XXT2_LELAM, Lelliottia amnigena

Therefore, it is intriguing to know how *E. coli* and other mesophilic organisms have lost this activity or not detectable in their α -subunits even though both possess similar active site amino acids and motifs. Therefore, it is tempting to speculate that these mesophilic enzymes may also have the intrinsic co-editing function but possibly unexplored.

The T. *thermophilus* α -subunit of the DNA polymerase III and *B. subtilis* X polymerase shows – very similar active site patterns. For example, the ⁹HLH-----⁴⁶DHG----⁶¹EMG-----²²⁰DAR pattern at the very N-terminal region of the α -subunit of the DNA polymerase III is very similar

to the ---³³⁹HMH------³⁷⁰DHS--⁴²⁸EMD----⁵²⁶DAHpattern found in DNA polymerases X (numbering from *B. subtilis* DNA polymerase X) (Fig. 17).

5. PR FUNCTION IN CORONAVIRAL RNA POLYMERASES

5.1 PR Function in RdRps of SARS-CoVs

Majority of the human diseases are caused by RNA viruses [21]. All these RNA viruses replicate their genome as well as transcribe their genes using the single enzyme, the RdRp. Therefore, major advancements towards the design and development of antivirals are expected to come from studies on the structure-function relationships of the enzymes of the RTC (RdRp. Primase, etc.) and PR exonuclease, also known as ExoN in CoVs. Active site analyses of the RdRps and Primases in SARS, SARS-related CoVs and HCoVs have been already reported by [22, 23 and references therein]. Therefore, in this communication, the ExoN which plays an important role in the replication of the viral genome is analysed to understand its active-site architecture and catalytic mechanism. As the ExoN of the SARS-CoVs directly interferes with the replication of the genome, it is also considered as a new target for drug design in addition to the RdRps and primases.

The ExoNs are found in nidoviruses only when their genome-size exceed beyond a certain threshold, i.e., >20 kb and thus, all nidoviruses with genome sizes above this threshold invariably possess the ExoN, whereas below the threshold do not possess the ExoN. [24]. Therefore, in all the four families of the *Nidovirales*, only the *Arteriviridae* (infects vertebrates) with a genome size ~15 kb do not possess the ExoN. The other three families, viz. *Coronaviridae* (~30 kb, infect vertebrates, mostly mammals), and the *Roniviridae* (~26 kb, infect mainly crustaceans) and Mesoniviridae (~20 kb, infect invertebrates) possess the Exon [25]. Therefore, the Coronaviridae with larger genome sizes, like the SARS-CoVs, SARSrelated CoVs and HCoVs, invariably possess the ExoNs. It is suggested that the ExoNs might have been acquired by the three nidoviruses during evolution for the maintenance and stability of their larger genomes. The ExoN is encoded by the nonstructural protein 14 (NSP14) in CoVs and is found to be a bifunctional enzyme harbouring two enzyme activities on the same polypeptide, viz. a PR exonuclease activity at the N-terminal region (from amino acids 1 to 287), and a guanine-N7-methyltransferase (N7-MTase) activity for mRNA capping at the C-terminal region (from amino acids 288 to 527, numbering from SARS-CoV-1). Amino acids from 288 to 301 make a convoluted loop and a break in the loop resulted in the abolishment of the N7-MTase activity. The PR exonuclease and the N7-MTase domains perform PR function and viral mRNA capping, respectively, during genome replication and transcription processes. [26, 27] have found that these two domains function independently as in all ExoN knockout mutants the PR activity was severely affected, whereas the N7-MTase activity was not affected.

CLUSTAL O (1.2.4) MSA of ExoN and PR-exonucleases (Amino acids 1-287 represent PR exonuclease and 288-527 represent N7-MTase)

	PR exonuclease	
HCoV-NL63 NSP14	SSQVCGLFKNCTRTPLNLPPTHAHTFLSLSDQFKTTGDLAVQIGSN-NVCTYEHVISFMG	59
HCoV-229E NSP14	SESSCGLFKDCARNPIDLPPSHATTYLSLSDRFKTSGDLAVQIGNN-NVCTYEHVISYMG	59
YP 460021.1	CTTNLFKDCSKSCLGYHPAHAPSFLAVDDKYKVNENLAVNLNICEPVLTYSRLISLMG	58
YP_009555255.1	CSTNLFKDCSKSYSGYHPAHAPSFLAVDDKYKATGDLAVCLGIGDSAVTYSRLISLMG	58
MERS-CoV_NSP14	SQIVTGLFKDCSRETSGLSPAYAPTYVSVDDKYKTSDELCVNLNLP-ANVPYSRVISRMG	59
SARS-CoV-1	AENVTGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGIP-KDMTYRRLISMMG	59
Civet-CoV NSP14	AENVTGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGLCVDIPGIP-KDMTYRRLISMMG	59
Pangolin-CoV NSP14	AENVTGLFKDCSKVITGLHPTQAPTYLSVDTKFKTEGLCVDIPGIP-KDMTYRRLISMMG	59
SARS-CoV-2	AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIP-KDMTYRRLISMMG	59
Bat-CoV RaTG13	AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIP-KDMTYRRLISMMG	59
_	▼ .***:*:: . *: * :.::: ::* * <u>::**</u> **	
HCoV-NL63 NSP14	FRFDISIPGSHSLFCTRDFAIRNVRGWLGM <mark>DVE\$</mark> AHVCGDNIGTNVPLQVGFSNGVNFVV	119
HCoV-229E NSP14	FRFDVSMPGSHFCTRDFAMRHVRGWLGM <mark>DVEC</mark> AHVTGDNVGTNVPLQVGFSNCVDFVA	117
YP 460021.1	FKLDLTLDGYSKLFITKDEAIKRVRGWVGF <mark>DVEC</mark> AHATRENIGTNFPLQIGFSTCVDFVV	118
YP_009555255.1	FKLDVTLDGYCKLFITKEEAVKRVRAWVGF <mark>DAEC</mark> AHATRDSIGTNFPLQLGFSTGIDFVV	118
MERS-CoV NSP14	FKLDATVPGYPKLFITREEAVRQVRSWIGF <mark>DVEC</mark> AHASRNACGTNVPLQLGFSTCVNFVV	119
SARS-CoV-1	FKMNYQVNGYPNMFITREEAIRHVRAWIGF VF CHATRDAVGTNLPLQLGFSTGVNLVA	119 <mark>2sdm</mark>
Civet-CoV NSP14	FKMNYQVNGYPHMFITREEAIRHVRAWIGF <mark>DVE¢</mark> CHATRDAVGTNLPLQLGFS T ¢VNLVA	119
Pangolin-CoV NSP14	FKMNYQVNGYPNMFITREEAIKHVRAWVGF <mark>DVEC</mark> CHATREAVGTNLPLQLGFSTCVNLVA	119
SARS-CoV-2	FKMNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHATREAVGTNLPLQLGFSTGVNLVA	119
Bat-CoV_RaTG13	FKMNYQVNGYPNMFITREEAVRHVRA <mark>WIGF<mark>DVEG</mark>CHATREAI<mark>GTNLP</mark>LQLGFSTGVNLVA</mark>	119
-	*::: : * * *:: *::.**. <mark>*:*</mark> .*. : ***. <mark>*</mark> **.*.*	

Palanivelu; IJBCRR, 30(3): 33-62, 2021; Article no.IJBCRR.70674

HCoV-NL63_NSP14ILVFVLWAGSI BLTTMRYFYKIGPIK-YCYCGNFATCYNSVSNE C.FKKALGQDYVYNP238HCoV-229E_NSP14VLVFVLWAGGI BLTTMRYFYKIGPIK-YCYCGNFATCYNSVSNE C.FKKALGQDYVYNP238YP_009555255.1CVVLVTWAANEBLTCLRYFAKUGELSCHVCTKRATVYNSRTGY CCWERSVTQDYLNP238Givet-CoV_NSP14YCTFVCWARGELTSASYFGKIGKEGKCMCNFRAAAVSSPLGS ALWTBSCG DYVYNP239Givet-CoV_NSP14RVVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENVGDYVNPP239Pangolin-CoV_NSP14RVVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENVGDYVNPP239Bat-CoV_RATG13RVVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENVGDYVNPP239RVVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENVGDYVNPP239RVVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENVGDYVNPP239Bat-CoV_RATG13RVVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENVGDYVNPP239RVVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENVGDYVNPP239KrvDVFVLWARGELTSMKYFVKIGPERT CLDKRATCFSTSSDT ACMENFGFDYVNPP239RVFVUWARGSTSLSSNIN FILMSFFNTER ACMENTYR239YP_00955255.1LIVDIQWGYGSLSSNIN FILMYFNKIGPERT CLDKRATCFSTSSDT ACMENFGFDYVPP239SARS-COV-1FMI DUQWGYGSLSSNIN FILMYFNKIGPERT CLDKRATCFSTSSDT ACMENFYR239SARS-COV-1FMI DUQWGYGSLSSNIN FILMYFNKIGPERT CLDKRATCFSTSSDT ACMENFYR239SARS-COV-1LIVDIQWGYGSLSSNIN FILMYFNKIGPERT CLDKRATCFSTSSDT ACMENFYR239SARS-COV-1FMI DUQWGYGSLSSNIN FILMYFNKIGPERT CLDKRATCFSTSSDT ACMENFYR239SARS-COV-1LIVDIQWGYGSLSSNIN FILMYFNKIGPERT CLDKRATCFSTSSDT ACMENFYR239SARS-COV-1FMI DUQWGYGSLSSNIN FILMYFNKIGPERT CLDKRATCFSTSSDT ACME	8 6 8
HCoV-NL63_NSP14YAFDIQQWG'VGSLSQNHHTFONTHRNEHDASG DAVMTRCIAVHDCFVKNVDWTVTYFFI298HCoV-229E_NSP14YVIDIQQWG'VGSLSTNHHAIDNYHRNEHDASG DAVMTRCIAVHDCFVKNVDWSTTYPMI296YP_460021.1LIVDIQQWG'TGSLTSNHDIIDNYHKGAHVASG DAIMTRIAVYDCFCKNVDWSTTYPMI296YP_009555255.1LIVDIQQWG'IGSLSSNHDLYSYHGGAHVASSDAIMTRIAVYDCFCNNINWNEYPII298SARS-CoV-1FMIYVQQWG'TGNLQSNHDCHSYNGGAHVASDAIMTRIAVYDCFCNNINWNEYPII299Civet-CoV_NSP14FMIDVQQWG'TGNLQSNHDQHSYNGSNAVASCDAIMTRCIAVHECFVKRVDWSVEYPII299SARS-CoV-2FMIDVQQWG'TGNLQSNHDQHSYNGSNAVASCDAIMTRCIAVHECFVKRVDWSVEYPII299Bat-CoV_RaTG13FMIDVQQWG'TGNLQSNHDLYCQVHGNAHVASCDAIMTRCIAVHECFVKRVDWTIEYPII299+:******:***********************************	8 9 9 <mark>2SDM</mark> 9 9 9 9
HCoV-NL63_NSP14ANEKFINGCGRNVQGHVVRAALKLYKPSVIHDIGNPKGVRCAVT-DAKWYCYDKQPVNSN357HCoV-229E_NSP14ANENAINKGGRTVQSHIMRAAIKLYNPKAIHDIGNPKGIRCAVT-DAKWYCYDKNPINSN355YP_460021.1SNEVSINTSCRLLQRVMLKAAMLCNRYNLCYDIGNPKGIACVKDYEFKFYDAFPVAKS356MERS-CoV_NSP14SNELSINTSCRVLQRVILKAAMLCNRYNLCYDIGNPKGIACVKDFDFKFYDAQPIVKS356SARS-CoV-1GDELRVNSACRKVQHMVVKAALLADSFPVLHDIGNPKGIPVOPVVMHYFDAQPCSDK359Civet-CoV_NSP14GDELRVNSACRKVQHMVVKSALLADKFPVLHDIGNPKAIKCVPQAEVEWKFYDAQPCSDK359Bargolin-CoV_NSP14GDELKINAACRKVQHMVVKAALLADKFPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDK359Bars-CoV-2GDELKINAACRKVQHMVVKAALLADKFPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDK359Bat-CoV_RaTG13GDELKINAACRKVQHMVVKAALLADKFPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDK359*****	8 6 8 9 9 9 9 9 9 9 9 9
··* :* * :: ::::*: :******:: · · ·: :* ** ·	7 5 6 9 9 9 9 9 9 9 9 9
HCoV-NL63_NSP14 VKLLDYDYATHGQLDGLCLFWNCNVDMYPEFSIVCRFDTRTRSVFNLEGVNGGSL 412 HCoV-229E_NSP14 VKTLEYDYMTHGQMDGLLFWNCNVDMYPEFSIVCRFDTRTRSTLNLEGVNGGSL 410 YP_460021.1 VKQLFYVYDVHKDNFKDGLCMFWNCNVDKYPSNSIVCRFDTRVLNKLNLPGCNGGSL 413 YP_009555255.1 VKTLLYSFEAHKDSFKDGLCMFWNCNVDKYPNNAVVCRFDTRVLNKLNLPGCNGGSL 413 MERS-CoV_NSP14 VQQLFYT-EDMASRFADGLCLFWNCNVDKYPNNAVVCRFDTRVLNKLNLPGCDGGSL 415 SARS-CoV-1 AYKIEELFYSYATHHDKFTDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSL 419 Civet-CoV_NSP14 AYKIEELFYSYATHHDKFTDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSL 419 Pangolin-CoV_NSP14 AYKIEELFYSYATHSDKFKDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSL 419 Bar-CoV_NSP14 AYKIEELFYSYATHSDKFKDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSL 419 Bar-CoV_NSP14 AYKIEELFYSYATHSDKFKDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSL 419 BARS-CoV-2 AYKIEELFYSYATHSDKFTDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSL 419 SARS-CoV-1 AYKIEELFYSYATHSDKFTDGVCLFWNCNVDRYPANAIVCRFDTRVLSNLNLPGCDGGSL 419 Bat-CoV RATG13 AYKIEELFYSYATHSDKFTDGVCLFWNCNVDRYPANSIVDRFDTRVLSNLNLPGCDGGSL 419	2 0 3 5 9
.: * ***: **********************************	9 9 9 9 9

HCoV-NL63 NSP14	IGGAV <mark>C</mark>	SK <mark>H</mark>	NLYQKYVEAYNTFTQAGFNIWVPHSFDVYNLWQIFIETNLQ	518
HCoV-229E NSP14	IGGAV <mark>C</mark>	SKH.	NLYRAYVESYNIFTQAGFNIWVPTTFDCYNLWQTFTEVNLQ	516
YP 460021.1	lggav <mark>c</mark>	SKH.	EEYCNYLESYNIVTTAGFTFWVYKNFDFYNLWNTFTTLQ	521
YP 009555255.1	lggav <mark>c</mark>	I.KH.	EEYREYLESYNTATTAGFTFWVYKTFDFYNLWNTFTKLQ	521
MERS-CoV NSP14	lggav <mark>c</mark>	RKH.	TEYREYMEAYNLVSASGFRLWCYKTFDIYNLWSTFTKVQ	524
SARS-CoV-1	lggav <mark>c</mark>	R.H.H.	NEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQ	527
Civet-CoV_NSP14	lggav <mark>c</mark>	R.H.H.	NEYRQYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQ	527
Pangolin-CoV_NSP14	lggav <mark>c</mark>	R.H.H.	NEYRLYLDAYNMMISAGFSLWIYKQFDTYNLWNTFTRLQ	527
SARS-CoV-2	LGGAVC	RHH	ANEYRLYLDAYNMMISAGFSLWVYKQFDTYNLWNTFTRLQ	527
Bat-CoV_RaTG13	lggav <mark>c</mark>	R.H.H.	NEYRLYLDAYNMMISAGFSLWVYKQFDTYNLWNTFTRLQ	527
—	:****	: *	* * * : : : * * : * * * * * * * :	

Fig. 18 MSA of the PR exonucleases from SARS, SARS-related CoVs and other HCoVs

 Amino acids highlighted in light blue indicates the crystallographic data

 JX104161.1 HCoV-NL63 (CBJ 037)
 NC_002645, HCoV-229E

 YP_460021.1 HCoV-HKU1
 YP_009555255.1 HCoV-OC43

 NC_019843, MERS-CoV
 NC_004718, SARS-CoV-1

 AAU04645, Civet-CoV
 QIQ54047, Pangolin-CoV

 NC_045512.2, SARS-CoV-2
 MN996532.2, Bat (RaTG13)-CoV

Furthermore, they also found that the SARS-CoV-2 ExoN knockout mutant was unable to replicate, suggesting a possibility for development of antivirals for ExoNs.

Fig. 18 shows the MSA of the ExoNs from SARS, SARS-related CoVs and HCoVs. The PR exonuclease region is highlighted in yellow and the MTase region is highlighted in green. But for few peptide regions, PR exonuclease regions are not highly conserved. Interestingly, the active site amino acids and the Zn binding motifs are completely conserved in all, suggesting their importance in the structure and function of these enzymes. The PR exonuclease domains of CoVs show that they all belong to the member of the DEDD superfamily of exonucleases [2]. The ExoN domains possess four possible Zn binding motifs (ZFMs) (two in the PR domain with the patterns-CxxC \rightarrow CxxH- and -HxxxCxxH \rightarrow C- and two in the N7-MTase domain with the patterns -Cx4C \rightarrow CxxD- and –Cx3C \rightarrow CxxH- (Fig. 18).

5.2 Active Site Analyses of the Exons of SARS-CoVs

The Exon's PR exonuclease active sites of SARS-CoV-1 and MERS-CoV and SARS-CoV-2 were studied by different investigators [28-29]. Minskaia et al [29] have analyzed the SARS-CoV-1 ExoN by SDM experiments and identified the residues D^{90}/E^{92} (motif I), D^{243} (motif II), and D^{273} (motif III) as the putative active-site residues. They have further demonstrated that modification of the ExoN active site amino acids resulted in the failure to recover infectious viral progenies. Ma et al [28] have analyzed the crystal structures of SARS-CoV-1 ExoN, in complex with its activator (NSP10) and functional ligands. They found that the amino acid residues

Cys²⁰⁷, Cys²¹⁰, Cys²²⁶, and His²²⁹ constituted the first zinc finger whereas the second zinc finger was consisted of His²⁵⁷, Cys²⁶¹, His²⁶⁴, and Cys²⁷⁹. Simultaneously mutating Asp⁹⁰ and Glu⁹² to Ala, impaired the ExoN activity drastically, whereas E¹⁹¹→A, H²⁶⁸→A, or D²⁷³→A mutants severely affected their ability to degrade RNA, confirming their importance in the exonuclease function. Asp²⁴³ is the fifth highly conserved amino acid identified in motif II. The ExoN activity of D²⁴³→A mutant was completely lost, suggesting its possible role in the catalysis.

Ma et al. [28] have also found that the catalytic core of the SARS-CoV-1 ExoN was very similar to other DEDD superfamily of PR exonucleases but starkly differed from other PR exonucleases by the presence of two zinc fingers. Furthermore, SDM studies indicated that both these zinc fingers are essential for the function of the ExoN's PR function. For example, a set of mutations generated by SDM experiments on the PR domain of the ZFMs ($C^{210} \rightarrow H$ in ZF1 and $C^{261} \rightarrow A$ and $H^{264} \rightarrow R$ in ZF2) abolished replication of SARS-CoV-1 genome, suggesting the importance of both ZFMs in the genome viability of SARS-CoV-1. MSA shows that the 2 ZFMs (highlighted in orange) are highly conserved in all SARS, SARS-related CoVs and other HCoVs suggesting, one ZFM may play the structural role and the other one could possibly involve in catalysis (Fig. 19).

The MERS-CoV and SARS-CoV-2 ExoNs were analyzed by SDM experiments and knockout mutations by Ogando et al [27]. They subjected all the five predicted active-site amino acid residues of MERS-CoV's ExoN domain (D^{90} , E^{92} , E^{191} , D^{273} , and H^{268}) by replacing them with Ala or with more conservative substitutions like D to E or Q; E to D or Q. This SDM experiment yielded a total of 14 ExoN active-site mutants, including the $D^{90} \rightarrow A/E^{92} \rightarrow A$ (motif I), double mutant, which was frequently used as a prototypic viable ExoN knockout mutant in SARS-CoV-1 studies. The following SDM experiments of the MERS-CoV ExoN yielded non-viable phenotypes: $D^{90} \rightarrow A/Q/E$; $E^{92} \rightarrow A/D/Q$; $E^{191} \rightarrow A/Q$; $D^{273} \rightarrow A/E/Q$; $H^{268} \rightarrow A$, suggesting their importance in the activity of the PR exonuclease. Furthermore, the ZF mutations ($C^{210} \rightarrow H$ in ZF1 and $C^{261} \rightarrow A$ and $H^{264} \rightarrow R$ in ZF2) abolished MERS-CoV replication, further establishing the importance of both the ZFMs for MERS-CoV viability. These results are also in close agreement with the SARS-CoV-1 results obtained by Ma et al [28]. They further evaluated the impact of ExoN inactivation (using a $D^{90} \rightarrow A/E^{92} \rightarrow A$ ExoN motif I double mutant) on replication and SARS-CoV-2 viability. Surprisingly, they could not rescue any viable progeny in which the two key residues of the ExoN active site amino acids were mutated. Interestingly, all ExoN exonuclease knockout mutations that proved lethal in reverse genetics were found to severely decrease ExoN activity without affecting N7-MTase activity. The SDM and crystallographic analyses of the PR exonuclease domains of the ExoNs of SARS-CoV-1, MERS-CoV and SARS-CoV-2 have clearly established that the same set of amino acids are making the active site as shown in Fig. 19.



Fig. 19 Proposed amino acids at the active site for the PR exonuclease of SARS-CoV-2

Table 1 Summary of the PR 3'-5' exonucleases from different bacteria and CoVs

Family	Consensus As* Pattern	Proton Acceptor	Catalytic Metal ion**	Zn-Binding site(s)
DEDD				
DNA Pol I	-DxE-D-Y-D-	Tyr	Zn ²⁺	1
DNA Pol II	-DxE-D-Y-D-	Tyr	Zn ²⁺	1
RNase D	-DxE-D-Y-D-	Tyr	Zn ²⁺	1
DNA Pol III, ε-subunit	-DxE-D-H-D-	His	Zn ²⁺	1
RNase T	-DxE-D-H-D-	His	Zn ²⁺	1
^SARS-CoV-1 ExoN/ACE2\$	-DxE-D-H-D-	His	Zn ²⁺	3
MERS-CoV ExoN/DPP4#	-DxE-D-H-D-	His	Zn ²⁺	3
SARS-CoV-2 ExoN//ACE2	-DxE-D-H-D-	His	Zn ²⁺	3
HCoV-NL63 ExoN/ACE2	-DxE-D-H-D-	His	Zn ²⁺	3
PHP				
DNA Pol X (B. Subtilis)	-HxH-E-H-D-	His	Zn ²⁺	1
YcdX (E. coli)	-HxH-E-H-D-	His	Zn ²⁺	3
DNA Pol III (Tth) co-editing	-HxH-E-H-D-	His	Zn ²⁺	1
*As. Active site: Pol. polymerase				

**proposed catalytic metal ions, water bound Zn²⁺

[^]Similar active sites and structural features are found in SARS-related CoVs and HCoVs too

\$ACE2, Angiotensin-Converting Enzyme 2; #DPP4, Dipeptidyl peptidase 4

Table 1 summarizes the minimum participating amino acids in both the superfamilies of exonucleases. The first two amino acids are acidic (DxE) in DEDD superfamily, whereas they are basic (HxH) in PHP superfamily. The next three amino acids are functionally equivalent in both.

5.3 Role of ExoN in the Functioning of the RdRp in CoVs

The ExoN activity is stimulated by binding of the NSP10. The activated ExoN is now placed, likely next to the RdRp's polymerization site during replication. As soon as the elongation starts, the N7-MTase caps the RNA to mimic the host mRNAs and to avoid degradation by host exonucleases. By a mechanism of stuttering at the slippery stop-site present at the end of viral genome, it adds the poly-A chain and thus, making the genome replication process complete. Only this faithfully replicated RNA, also known as the genomic RNA, is encapsidated and assembled into virions (Fig. 20).

Even though the replication and transcription processes are accomplished by the same set of enzymes, mechanistically both appears to be different. For example, the viral the genomic RNA is replicated by a continuous process from end to end, but the transcription process is not a continuous process, but a discontinuous process and a complex one. For example, the SARS- CoV-2 makes 1 genomic (1ab) and 9 subgenomic mRNAs (S, 3a, E, M, 6, 7a, 8, 7b and N, 10?). All the mRNAs are capped at their 5'-ends and tailed at their 3'-ends like the host mRNAs for subsequent translations. Invariably, all the genomic and the subgenomic RNAs in CoVs contain a transcription-regulating sequence (TRS) at their 5'-ends, which are located immediately adjacent to the ORFs. Each subgenomic mRNA contains the common 5-"leader" sequence (~70 nt) fused to the "body" TRS also known as TRS-B. The TRS-L and TRS-B sites have a conserved core sequence (CS) of 7 to 8 nt (-AACGAAC- is the CS in SARS-CoV-2). Thus, the genomic and subgenomic RNAs have the same common leader sequence and TRS at their 5'-ends. Because TRS-B is a signal for RdRp to switch templates, it is possible that recombination events are more likely to occur at or near TRS-B sites. As mentioned elsewhere, all the subgenomic mRNAs are created bv discontinuous transcription. The discontinuous transcription requires base-pairing between cisacting body TRSs, with the leader sequence located at the 5'-end of the viral genome. Because of such discontinuous extension of minus strands, all these subgenome-length minus strands carry the complement of the leader sequence at their 3'-ends [29, 30]. During replication mode, the Replication-Transcription Complex (RTC) ignores the transcription signals.



Fig. 20 A proposed simplified model for the replication, transcription and translation processes in SARS-CoVs

Palanivelu; IJBCRR, 30(3): 33-62, 2021; Article no.IJBCRR.70674



Fig. 21 Steps (1-4) involved in the proposed mechanism of action of ExoN PR Exonuclease of SARS-CoV-2

Steps 1 and 2: The wrongly added base is excised by the PR exonuclease by the mismatch induced activation of the water molecule bound to Zn^{2+} , initiating proton transfer with the simultaneous nucleophilic attack on the susceptible phosphodiester bond by the highly reactive Zn-hydroxide. **Step 3 and 4:** The wrongly added base is excised, and the polymerase resume synthesis with the right

Step 3 and 4: The wrongly added base is excised, and the polymerase resume synthesis with the right nucleotide

function for excising a mismatched nucleotide

The RTC plays three different roles during replication and transcription processes, viz. i) synthesis of the (-) RNA strand using genomic (+) strand as template, ii) production of large number of the (+) genomic RNAs from the (-) RNA strand as template and iii) transcription of genomic and subgenomic mRNAs from the (-) RNA strand. The mutations that occur during the (-) strand synthesis will be reflected in the (+) strand genomic RNAs and in the genomic and subgenomic mRNAs and hence, in the nonstructural, structural and accessory proteins (NSPs, SPs, APs). Effecting a large deletion, the 5" leader sequence is fused to each subgenomic RNAs, which ensures discontinuous transcription of subgenomic mRNAs. Fig. 20 shows a proposed model for the replication, transcription and translation processes in SARS-CoVs.

6. GENERAL MECHANISM PROPOSED FOR PR EXONUCLEASES IN DNA/RNA SYNTHESIS AND MODIFICATIONS

A two-metal ion mechanism is proposed for the PR exonucleases from both the superfamilies. A water-bound Zn²⁺ is proposed as the primary metal ion which initiates the catalysis and the Mg²⁺ which is also making the active site, is proposed as the supporting metal ion. The mismatch at the active site of the polymerase, signals the PR exonuclease to move-in and excise the mismatch. Zinc is placed as the primary metal ion as it is one of the most ubiquitous cofactors found in a large number of enzymes and proteins. For example, >450 enzymes and proteins use Zn^{2+} as cofactor. In these enzymes, the zinc atoms are known to play both the structural and catalytic roles. For example, zinc based catalysis is established in many enzymes like carboxypeptidases-A, carbonic anhydrases, thermolysin, alkaline phosphatases, metallo *B*-lactamases, PR exonucleases of DNA and RNA polymerases, RNA modifying enzymes, etc. [31 and references therein]. One common theme proposed for many of these enzymes is the activation of a water molecule coordinated to the Zn²⁺ for a nucleophilic attack on the carbonyl carbon of a peptide bond, or the phosphorus atom of a phosphoester bond. Furthermore, it is proposed that several of the zinc enzymes' action is facilitated by the formation of a zinc-hydroxide [1 and references therein]. During instances of mismatch of nucleotides, the RdRp stalls/pause, allowing the PR domain to excise the mismatched nucleotide. Fig. 21 shows a proposed mechanism for PR exonuclease

during RNA replication in SARS-CoV-2. In addition to the catalytic zinc, a second active

site metal ion like Mg^{2+}/Zn^{2+} is also known to be essential for assisting the catalysis by SDM experiments. The secondary metal ion is suggested to function to stabilize the transient pentacovalent species and/or to facilitate the leaving of the 3' oxyanion from an axial position, whereas the primary metal ion facilitate the formation of an attacking hydroxide ion. Involvement of both the metal ions has been unequivocally proved by X-ray crystallographic data and SDM experiments. The PHP superfamily of PR exonucleases invariably use Zn²⁺ as the primary metal ion, as other metal ions could not fit into its place whereas the DEDD superfamily of enzymes which use Zn²⁺ as the primary metal ion, but other divalent metal ions like Mg²⁺/Mn²⁺ could also fit into the primary Zn²⁺ site as it is essentially made up of three acidic amino acids [32]. In the DEDYD subfamily of exonucleases, Tyr is used as the proton acceptor. In fact, Tyr serves as the nucleophile in the active site of topoisomerases, which also makes phosphodiester bond breaks during catalysis [33].

7. CONCLUSIONS

The PR functions form an important component of the replicative polymerases in biological systems. This study shows there are at least two different types of PR exonucleases performing this function in prokaryotes and viruses (CoVs). They belong to either DEDD or PHP superfamilies. The DEDD superfamily of exonucleases essentially uses the four acidic amino acids. DEDD in their catalytic site with additional Y or H as the proton acceptor (-DEDDY/H-). The PHP superfamily essentially uses -HxH-E-H-D- pattern at the active site with an invariant H as the proton acceptor. The PHP exonucleases are found to be mainly a Zn²⁺dependent enzymes. From the similarities of active site amino acids/motifs, it may be concluded that the DEDD and PHP superfamilies of PR exonucleases should have evolved from a common ancestor but diverged very long ago. The structural features of the PR enzymes from the CoVs suggest that CoVs may have acquired the exonuclease function, possibly from a prokaryote. However, the presence of two zincbinding sites in PR active site of SARS, SARSrelated CoVs and HCoVs sets it apart from their homologues.

ACKNOWLEDGMENTS

The author wishes to thank Dr. H. Shakila, Professor and Head, Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai, for useful suggestions on the manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

- Palanivelu P.DNA polymerases An 1. insight into their active sites and mechanism of action, In: Recent Advances in Biological Research. SCIENCEDOMAIN International Book Publishers, UK, Vol 1, pp 1-39, ISBN: 9788193422441. DOI: 10.9734/bpi/rabr/v1; 2019.
- 2. Zuo Y, Deutscher MP. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res. 2001;29:1017–1026.
- Nagpal S, Nair DT. The PHP domain of PolX from *Staphylococcus aureus* aids high fidelity DNA synthesis through the removal of misincorporated deoxyribo-, ribo- and oxidized nucleotides. Nat Sci Rep. 2021;11:4178.
- Stano NM, Chen J, McHenry CS. A coproofreading Zn²⁺-dependent exonuclease within a bacterial replicase. Nat Struct Mol Biol. 2006;13:458–459.
- Teplyakov A, Obmolova G, Khil PP, Howard AJ, Camerini-Otero RD, Gilliland GL. Crystal Structure of the *Escherichia coli* YcdX Protein Reveals a Trinuclear Zinc Active Site. Proteins: Struct Funct Genet. 2003;51:315–318.
- Derbyshire V, Freemont PS, Sanderson MR, Beese L, Friedman JM, Joyce CM, Steitz TA. Genetic and crystallographic studies of the 3,5'-exonucleolyitc site of DNA polymerase I. Science. 1988;240: 199-201.
- Derbyshire V, Grindley ND, Joyce CM. The 3'-5' exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction. EMBO J. 1991;10:17-24.
- Joyce CM, Steitz TA. DNA polymerase I: from crystal structure to function via genetics. Trends Biochem Sci. 1987;12: 288-292.

- Beese LS, Steitz TA. Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. 1991;10:25–33.
- Wang F, Yang W. Structural insight into translesion synthesis by DNA Pol II. Cell. 2009;139:1279–1289.
- Fijalkowska IJ, Schaaper RM. Mutants in the Exo I motif of *Escherichia coli* dnaQ: Defective proofreading and inviability due to error catastrophe. Proc Natl Acad Sci. (USA). 1996;93:2856-2861.
- 12. Cisneros GA, Perera L, Schaaper RM, Pedersen LC, London RE, Pedersen LG, Darden TA. Reaction mechanism of the ε subunit of *E. coli* DNA polymerase III: Insights into active site metal coordination and catalytically significant residues. J Am Chem Soc. 2009;131:1550–1556.
- Hamdan S, Carr PD, Brown SE, Ollis DL, Dixon NE. Structural Basis for Proofreading during Replication of the *Escherichia coli* Chromosome. Structure. 2002;10:535–546.
- Blanco L, Bernad, A, Salas M. (1992). Evidence favouring the hypothesis of a conserved 3'-5' exonuclease active site in DNA-dependent DNA polymerases. Gene. 1992;112:139-44.
- Zuo Y, Zheng H, Wang Y, Chruszcz M, Cymborowski M, Skarina T, Savchenko A, Malhotra A,Minor W. Crystal Structure of RNase T, an exoribonuclease involved in tRNA maturation and end-turnover. Structure. 2007;15:417–428.
- Ramadan K, Shevelev I, Hübscher U. The DNA-polymerase-X family: controllers of DNA quality?. Nat. Rev. Mol. Cell Biol. 2004;5:1038–1043.
- Uchiyama Y, Takeuchi R, Kodera H, Sakaguchi K. Distribution and roles of Xfamily DNA polymerases in eukaryotes. Biochimie. 2009;91:165–170.
- Banos B, Lazaro JM, Villar L, Salas M, de Vega M.Editing of misaligned 3'-termini by an intrinsic 3'-5' exonuclease activity residing in the PHP domain of a family X DNA polymerase. Nucleic Acids Res. 2008;36:5736-5749.
- 19. Nakane S, Nakagawa N, Kuramitsu S, Masui R. Characterization of DNA polymerase X from *Thermus thermophilus* HB8 reveals the POLXc and PHP domains are both required for 3'–5' exonuclease activity. Nucleic Acids Res. 2009;37:2037–2052.

- 20. Palanivelu P. Multiple sequence analysis of Polygalacturonases and Invertases and Phase Shift in Conserved Motifs. Ind J Biotechnol. 2007;6:24-30.
- 21. Poltronieria P, Sunb B, Mallardoc M. RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. Current Genomics, 2015;16:327-335.
- 22. Palanivelu P. Analyses of the Spike Proteins of Severe Acute Respiratory Syndrome-Related Coronaviruses. Microbiol Res J Int. 2020;30:32-50.
- 23. Palanivelu P.RNA-Dependent RNA Polymerases of Severe Acute Respiratory Syndrome-Related Coronaviruses- An Insight into their Active Sites and Mechanism of Action. International J Biochem Res Rev. 2021;29:29-52.
- 24. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ (2006) Nidovirales: Evolving the largest RNA virus genome. Virus Res. 2006;117:17–37.
- Lauber C, Goeman JJ, Parquet MDC, Nga PT, Snijder EJ, Morita K, Gorbalenya AE. 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. PLoS Pathog Available:9:e1003500.https://doi.org/10.13 71/journal.ppat.1003500.
- 26. Ogando NS, Ferron F, Decroly E, Canard B, Posthuma CC, Snijder EJ. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. Front Microbiol. 10:1813.

DOI: 10.3389/fmicb.2019.01813.

- Ogando NS, Zevenhoven-Dobbe JC, Meer YVD, Bredenbeek PJ, Posthuma CC, Snijder EJ. The Enzymatic Activity of the nsp14 Exoribonuclease Is Criticalfor Replication of MERS-CoV and SARS-CoV-2. J Virol. 2020;94:e01246-20.
- Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. Proc Natl Acad Sci (USA). 2015;112:9436 –9441.
- Minskaia E, Hertzig T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, Ziebuhr J. Discovery of an RNA virus 3'-5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc Natl Acad Sci (USA). 2006;103:5108– 5113.
- Yang Y, Yan W, Hall AB, Jiang X. Characterizing Transcriptional Regulatory Sequences in Coronaviruses and Their Role in Recombination. Mol Biol Evol. 2021;38:1241–1248.
- 31. Coleman JE. Zinc enzymes. Curr Opin Chem Biol. 1998;2:222–234.
- 32. Freemont PS, Friedman JM, Beese LS, Sanderson MR, Steitz TA. Proc Natl Acad Sci. (USA). 1988;85:8924-8928.
- Champoux JJ. DNA Topoisomerases: Structure, Function, and Mechanism. Annu Rev Biochem. 2001;70:369–413.

© 2021 Palanivelu; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history: The peer review history for this paper can be accessed here: https://www.sdiarticle4.com/review-history/70674