World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Replication of Genomes in Eukaryotes: Analyses of the Structure and Functions of the Replicative ε Polymerases from Yeasts, Plants and Animals

Peramachi Palanivelu *

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

World Journal of Advanced Research and Reviews, 2025, 26(03), 545-578

Publication history: Received on 28 April 2025; revised on 31 May 2025; accepted on 03 June 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.3.2222

Abstract

In all living organisms, genome replication is an indispensable activity in their life cycle. As compared to prokaryotes, genome replication in eukaryotes is much more complex. In eukaryotes, the genome replication is accomplished by a more specialized, multi-protein complex known as replisome. Therefore, for a complete understanding of the replication process in eukaryotes, the structure-function relationship of their replisome complex is important. The eukaryotic replisome complex is mainly composed of 3 replicative polymerases (pols), viz. α, ε and δ. The initiation of genome replication starts with the synthesis of primers by the α pol (a primase), which is followed by the synthesis of leading- and lagging-strands by the next two replicative pols, viz. ε and δ, respectively. Among them, the pol ε is an essential enzyme that links the DNA replication machinery to S-phase checkpoint-control in eukaryotes. The active sites of pol and proofreading (PR) domains of α and δ were analyzed from yeasts, animal and plant sources and reported by this author already [1, 2]. The third replicative enzyme, viz. the pol ε is analyzed and reported here. Interestingly, the yeast, animal and plant ε pols possess almost identical pol, PR and carboxyl terminal domains (CTD) with the same active site amino acids. Their catalytic cores use the same template-binding pair (–YG-), the catalytic amino acid (K), the nucleotide selection amino acid (Q) and similar catalytic metal-binding motifs in the pol domain as reported for the other replicative pols. However, they exhibit poor identities among themselves, e.g., the human pol ε showed only 46.69% and 43.19% identities to the plant (*Arabidopsis thaliana*) and yeast (*Saccharomyces cerevisiae*) pol ε sequences, respectively. Furthermore, these eukaryotic ε pols use the same active site amino acids in the PR exonuclease domain, and belong to the DEDD(Y)-superfamily of PR exonucleases. Moreover, similar to the other two replicative enzymes, the ε pols also use similar regulatory Zn^{2+}-binding motifs (ZBMs) in their CTD. In addition, it harbours a unique ZBM in their pol domain. Interestingly, the two invariant motifs, viz –SLYPS- and -YGDTD- which are the characteristic motifs found in the other replicative α and δ pols and B-family of DNA pols, are not found in ε pols. Many specialized, conserved sequence motifs are also identified in the ε pols and discussed.

Keywords: Eukaryotic genome replication; DNA polymerase δ; DNA polymerase ε; DNA polymerase ε active sites; Proofreading exonuclease active sites; *Saccharomyces cerevisiae*; *Arabidopsis thaliana*; *Homo sapiens*

1. Introduction

In all living organisms, DNA pols play a central role in DNA replication to ensure faithful transmission of genetic information from one generation to another. Therefore, to accomplish this important task in their lifecycle, organisms have evolved and possess a high-fidelity genome replication machinery to preserve and maintain the blueprint of life. The replication machinery possesses several types of DNA pols and repair mechanisms to ensure the duplication of an exact copy of the original genome. To date, five different DNA pols have been characterized in *Escherichia coli*, eight in *S. cerevisiae*, and as many as 16 in humans [3]. However, only three of these DNA pols, viz. α, ε and δ are involved in the duplication of the nuclear genome in all eukaryotes [4]. Even though more than a dozen different DNA pols have also

* Corresponding author: Peramachi Palanivelu

been reported to perform various replicative and repair functions in plant cells, yet the detailed properties of them are still very limited. Whereas the α pol (the primase) is involved in the synthesis of the DNA primers to initiate the DNA duplication process, the other two replicative pols, viz. δ and ε, use these DNA primers and replicate the whole genome and make an exact copy of the original genome. All three replicative pols belong to B-family pols which are not only involved in replication, but also in repair any error that may occur during the replication process [5]. The B-family DNA pols are reported to function in both prokaryotes and eukaryotes, e.g., pol II, pol B, pol α, pol ε, pol δ, pol ζ. They are bifunctional enzymes, i.e., in addition to polymerization, they also exhibit 3'→5' PR exonuclease (PR exo) repair function [6]. Even though the replicative pols have an efficient PR function, mutation(s) that rarely escape the replication and repair processes generate the most important new genetic variants in plants and animals. The structural and functional aspects of the replicative pols, viz. α and δ, from yeasts, higher fungi and animals have been extensively analyzed and reported by this author [1, 2]. In this article, the third replicative pol ε from yeasts, higher fungi, plants and animals is analyzed and reported.

The DNA replication process is highly conserved in all domains of life [7]. Though genome replication in eukaryotes is known to be incredibly more complex, but it is successfully accomplished by a highly sophisticated and coordinated series of molecular events. The replication process, apart from the essential roles played by the three replicative pols, also depends on the participation of other enzymes and regulatory proteins like, primases, cell-cycle kinases, replicative helicases, single-strand binding proteins (SSBs), additional repair enzymes, ligases, etc. Given the importance of accurate DNA replication, proper functioning of all these enzymes and the accessory proteins is critical to maintain the genome integrity.

The first step in genome replication in both prokaryotes and eukaryotes is the initiation. This is accomplished by a multi-structural enzyme and protein complex known as a primosome. The primosome essentially consists of an origin-of-replication initiator protein, a replicative helicase, a helicase loader, SSBs, a primase, a topoisomerase, etc. [8]. After successful initiation of replication, the replication process is taken over by a next multi-protein complex known as the replisome, which extends the replication process and completes it. In both prokaryotes and eukaryotes, the initiation step is performed by the enzyme, RNA primase, which synthesizes the required RNA primer (short segment of SS-RNA typically of 10-20 nucleotides) to initiate the replication process. However, for eukaryotic replications, in addition to the RNA primer, an additional primer, viz. a DNA primer is also required for further extension, which is provided by the enzyme, DNA pol α. The DNA pol α extends the RNA primers by synthesizing short stretches of DNA, which is used by the pol ε and pol δ for further extension of the DNA synthesis in a processive manner and complete the replication process. During replication, the leading-strand is synthesized by pol ε (pol2), and the lagging-strand is synthesized by pol δ (pol3). During replication the leading-strand is synthesized continuously, whereas the lagging-strand is synthesized discontinuously, in fragments of ~200 nucleotides long, known as the Okazaki fragments (OFs), named after the discoverer. All these OFs are connected together in the next step to form a complete complementary strand by a DNA ligase [9].

1.1. Eukaryotic ε DNA Polymerases

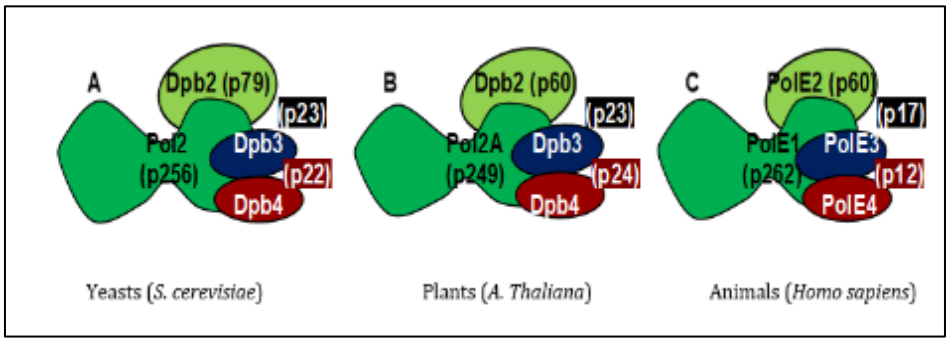
From the above discussions, it is clear that the bulk of the DNA synthesis in eukaryotes is performed by only two of the three replicative polymerases, viz. pols ε and δ. DNA pol ε (also known as Pol2 in yeasts and plants and PolE in humans) completes the leading- and lagging-strand syntheses, respectively. Pol ε was first isolated and characterized from yeast cells as early as 1970 as a nuclear polymerase [10]. It was initially considered to be a form of pol δ that was processive in the absence of proliferating cell nuclear antigen (PCNA) and replication factor C (RFC) [11] (The function of RFC is to load PCNA, the processivity factor of the eukaryotic DNA polymerases δ and ε, onto primed DNA templates). Later, it was found that pol ε is an independent replicative enzyme in all eukaryotes. Nowadays, its catalytic subunit is isolated and characterized and its gene is cloned and sequenced from several eukaryotes [11 and references therein]. This enzyme, along with the other two eukaryotic replicative (pols α and δ), are grouped under the B-family polymerases [12]. A great deal of information is available now on the structure and functions of the DNA pols ε from a large number of eukaryotes like yeasts, higher fungi, plants and animals. The plant enzyme from *A. thaliana* has been characterized in sufficient detail and used for further analysis and comparison.

1.1.1. Subunit structure

Interestingly, all the pol ε enzymes from eukaryotes are composed of 4 subunits exhibiting similar subunit structures. Figures 1A, B and C show the subunit structures of the ε pols from yeasts, plants and animals. For example, the yeast enzyme's 4 subunits are named as, Pol2, Dpb2, Dpb3 and Dpb4 (with the molecular masses of 256, 79, 23 and 22 kDa, respectively) and the plant enzyme also follows more or less the same nomenclature as the yeast enzyme as, Pol2A, Dpb2, Dpb3 and Dpb4 (with the molecular masses of 249, 60, 23 and 24 kDa, respectively), but the human enzyme subunits are named as, PolE1-PolE4 (with the molecular masses of 262, 60, 17 and 12 kDa, respectively). The 4 subunits

are assembled with a stoichiometry of 1:1:1:1. Thus, the DNA pol ε is a heterotetrameric enzyme, made up of the main catalytic subunit (Pol2/Pol2A/PolE1) which assumes a biloped structure, where the NTD and CTD lopes are connected by a linker region and the non-catalytic subunit (Dpb2/PolE2). The Dpb2/PolE2 subunit mediates the homodimerization of the enzyme to its active form and therefore, is essential for the enzyme activity and cell viability. Genetic and biochemical studies indicate that the putative zinc-finger regions in the catalytic subunit (Pol2/Pol2A/PolE1) are important for its interactions with the Dpb2/PolE2 essential subunit. The other two smaller subunits (Dpb3/PolE3 and Dpb4/PolE4) are found to be nonessential for the catalytic activity of the enzyme, but are found to play important regulatory roles [11]. For example, each of the smaller subunits, viz. PolE3 and PolE4 possess H2A/H2B type histone folds and form a tight PolE3-PolE4 dimer. The dimer interacts with the larger catalytic subunit of the enzyme. (For example, PolE3 interacts with subunits PolE1 and PolE4; and the PolE4 interacts with subunits PolE1 and PolE3 and with the subunit of chromatin remodelling complex, (CHRAC), a multi-protein assembly that regulate gene expression and chromatin structure. Though the non-catalytic subunits are highly conserved from yeasts to humans, they largely differ in their sizes in humans [11 and references therein].

Among the plant ε pols (Pol2), the pol ε from *A. thaliana* has been the most studied one. The pol ε complex in *A. thaliana* has been shown to play diverse regulatory roles. For example, it is implicated in DNA replication, DNA repair (nucleotide and base excision repairs), DNA recombination and also in epigenetic gene-silencing activities [11]. Unlike the yeast and human enzymes, the *Arabidopsis* enzyme possesses two isoenzymes, viz. Pol2A and Pol2B. The *Pol2A* gene is ~16 kb long (49 exons) and encodes a protein of 2,161 amino acids (pI/Mw: 6.19/2,48,897), whereas the ~13 kb long *Pol2B* gene with same number of exons *(*49 exons) encodes a protein of 2,138 amino acids which shows 78.5% identity to Pol2A [13]. Importantly, both the isozymes contain all the motifs that are necessary for a functional Pol2 catalytic subunit. However, the expression of *Pol2B* has been detected mostly under adverse environmental conditions and loss of function of *Pol2B* has resulted in no visible phenotype, suggesting that the Pol2A forms the main enzyme [13]. Furthermore, the knockout mutants of the (Pol2A) are found to be lethal, suggesting its importance [14, 12]. The non-enzymatic CTD plays a crucial role and it not only mediates the interaction of the main catalytic subunit with the other three smaller regulatory subunits, but also links the replication machinery to the S-phase checkpoint [15, 16]. However, its overall subunit composition and various domains of its catalytic subunit's structure are mostly understood from the data available from yeast and human enzymes.



Adapted from [11, 12].

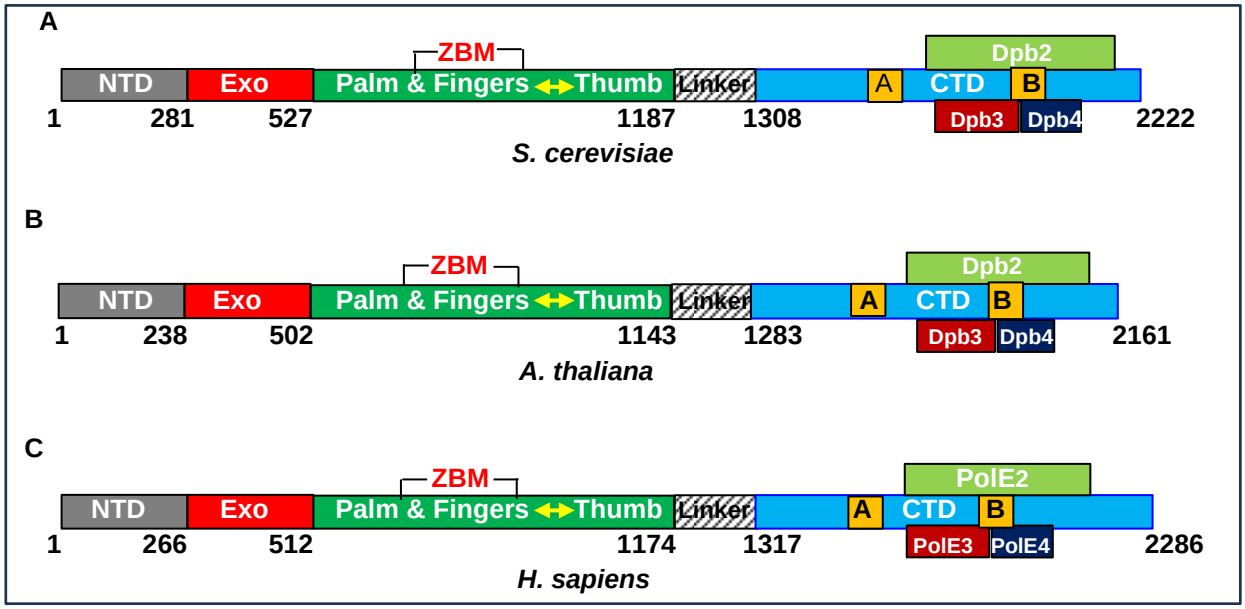
The pol ε subunit p17 has been identified as an integral subunit of the chromatin-remodelling factor, CHRAC (Chromatin Accessibility Complex). CHRAC is an ISWI (Imitation Switch), i.e., a chromatin remodelling nuclear-ATPase containing complex that regulates chromatin accessibility and nucleosome spacing. In CHRAC, the p17 forms a dimer with p12 subunit that can interact with histones as well as DNA [17].

Figures 1A, B and C. Subunit structures of the eukaryotic DNA pol ε (*S. cerevisiae*, *A. thaliana* and *Homo sapiens*) (the numbers in brackets indicate molecular masses of the subunits in kDa)

1.1.2. Domain organization of the pol ε catalytic subunit

The domain organization of the pol ε's (pol2) catalytic subunit from the yeast, *S. cerevisiae*, has been extensively studied and characterized. It contains two pol modules, one in the N-terminal domain (NTD) and the other in the CTD. The NTD pol ε module possesses both the pol and PR exonuclease activities, whereas the CTD mediates the crucial interactions between the three smaller regulatory subunits with the catalytic subunit and also links the replication machinery to the S-phase checkpoint [11]. The pol ε exhibits the highest processivity as compared to the other two replicative pols, viz. pols α, and δ. This high processivity was found to be increased even further by the nonessential subunits, Dpb3 and Dpb4, through their interactions with the replication clamp, the PCNA. Furthermore, the genetic and biochemical studies indicate that the putative ZFMs in the main catalytic subunit of the enzyme are important for their interactions with the other subunits and for the catalytic efficiency of the enzyme [18].

The main catalytic subunit of the yeast pol ε holoenzyme is dissected into different functional domains as, NTD (31–281), PR exonuclease domain (282–527) and the palm (528–950), fingers (769–833), and thumb (951–1186) subdomains of the pol domain (that are organized into a toroid) [19], and a linker domain which connects the pol to the CTD. Figures 2A, 2B and 2C show the various domains and the interactions of the subunits with the CTD of the pols ε from different eukaryotic sources.



ZBM, Zinc-binding motif; A & B are CysA and CysB, respectively

Figures 2A, B and C. Tentative domain organization the main catalytic subunit of the ε pols from eukaryotes (*S. cerevisiae, A. thaliana* and *Homo sapiens*) and their interacting accessory subunits

2. Materials and Methods

The protein sequence data of ε DNA pols from yeasts, higher fungi, plants and animals 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, and the data already available from biochemical, SDM, cryogenic-Electron Microscopy (cryo-EM) and X-ray crystallographic analyses on the ε pols are used to confirm the possible amino acids at the active sites of the pol, PR exonuclease domains and regulatory motifs.

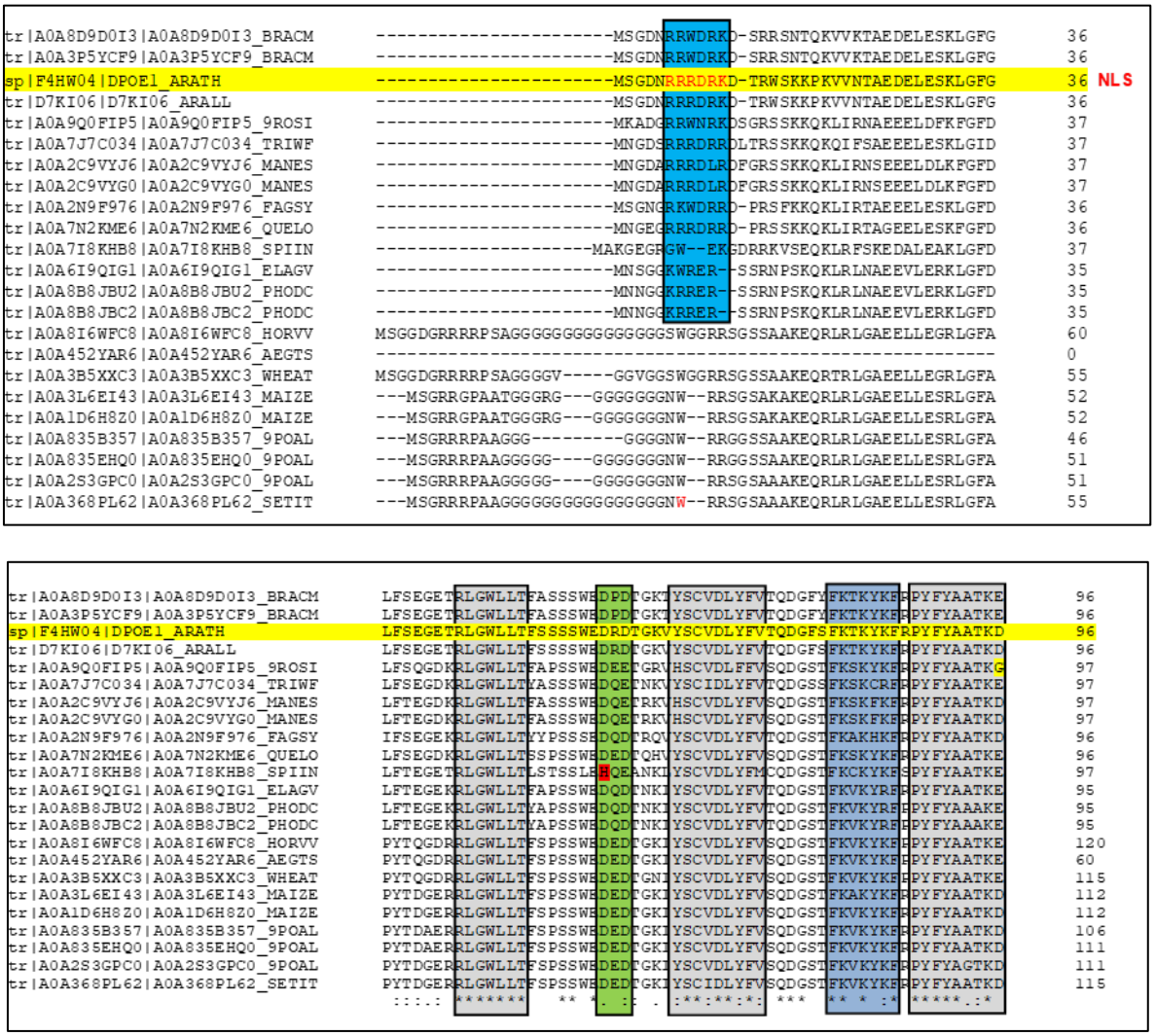
3. Results and Discussion

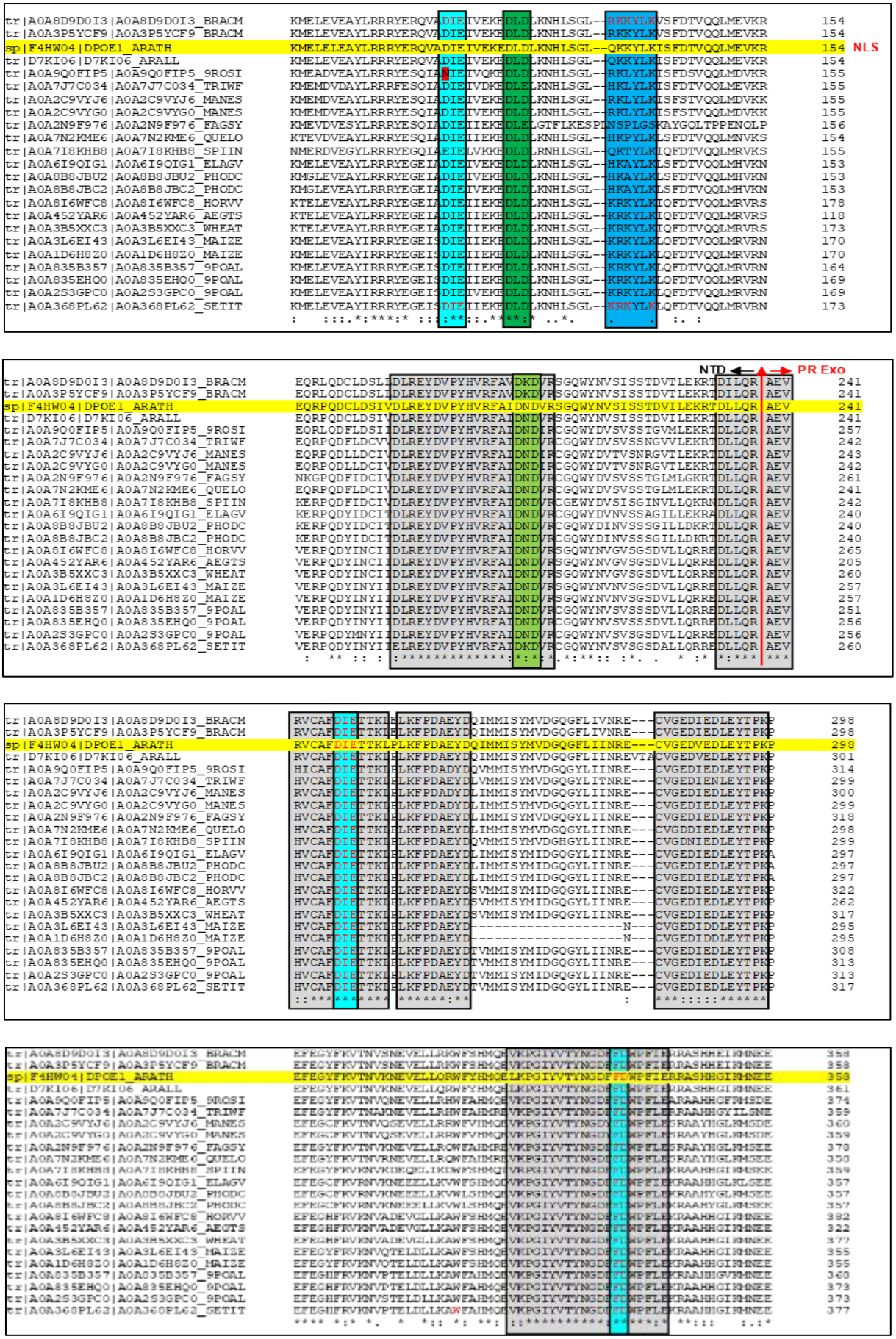
3.1. MSA analysis of DNA polymerase ε from various plant sources

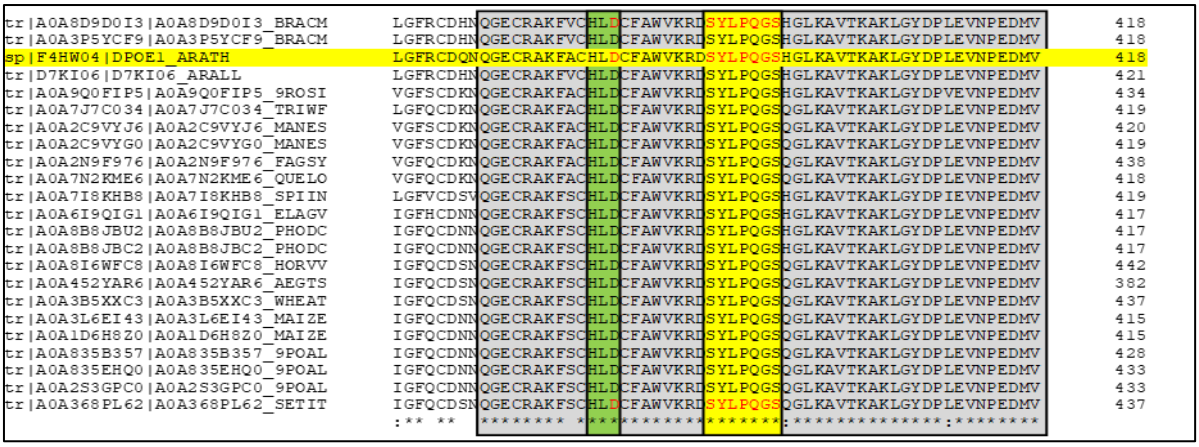
The DNA pol ε was analyzed from yeasts, higher fungi, plants and animal sources. The structural and functional aspects of this enzyme from yeasts and higher fungal sources were already reported by this author [8]. Fig. 3 shows the MSA of pol ε from various plant sources (only the required regions for the discussions are shown). The model plant, *A. thaliana*'s sequence is used as the standard and highlighted. All the major domains (NTD, PR Exo, Pol, Linker and CTD) are indicated with arrow marks. Large numbers of highly conserved small and large peptides are found from NTD to CTD. It is interesting to note that the highly conserved active site amino acids are usually embedded within these conserved peptides (highlighted). The NTD is conserved in all, but aligned with some gaps. It harbours 3 conserved -DxD- type metal-binding motifs and 3 polybasic peptides (highlighted) and some are implicated as nuclear localization signal (NLS). The NTD is followed by the PR exonuclease and pol domains. The proposed PR exonuclease active site amino acids are highly conserved in all the enzymes (highlighted) and it belongs to the DEDD-superfamily of exonucleases and uses an invariant Y as proton acceptor during catalysis. This is in complete agreement with the other DNA-dependent DNA pols (DdDps) from prokaryotes and eukaryotes [1, 6]. An invariant motif, –**SYLP**QG**S**- (highlighted in yellow) is found within the PR exo region in the plant ε pols as found in yeasts, higher fungi and animals [8]. This sequence is somewhat similar to the characteristic –**SLYPS**- motif found in the other two eukaryotic replicative DNA pols, α and δ. In both the cases, this motif is implicated in dNTP-binding. In the former, it is found within the proposed PR exo region, whereas in the latter two pols, it is located in the pol region itself. The PR exo domain is followed by the pol domain and

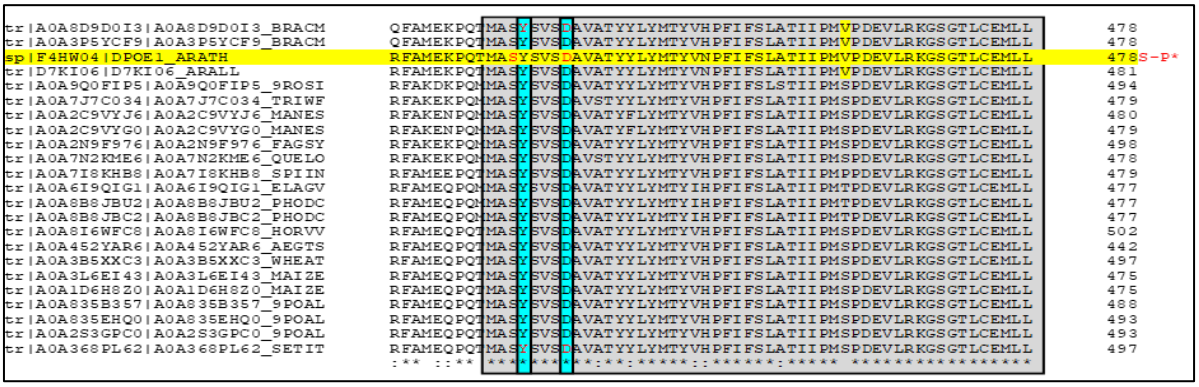
is also highly conserved in all. The proposed template-binding and catalytic pairs, and metal-binding sites are highlighted in yellow and green, respectively. The template-binding pair (-YG-), catalytic amino acid (K) and the nucleotide selection amino acid (Q) are in close agreement with other DdDps and DdRps [6]. Unlike other two eukaryotic replicative pols, the pol ε domain contains a ZBM (highlighted) within. The pol domain is followed by a linker region which connects the pol domain to the CTD. Interestingly, a second putative pol active site amino acids are identified in the CTD and highlighted in yellow. The CTD contains the two typical, highly conserved ZBMs and one of them (CysB) is known to bind the 4Fe-4S cluster (highlighted in orange). It is interesting to note that these 2 ZBMs are found invariably in all three replicative pols at their CTDs. In addition to the Cs which make these ZBMs, there are invariant Cs found throughout the sequence (data not shown) which may be involved in disulphide bond formation. The CTD also contains a –DxD- type metal-binding motif within the CysA. It is interesting to note that the NTD, pol and linker regions contain conserved polybasic peptides within them (highlighted) and are implicated as NLSs. (A NLS is a short amino acid sequence that tags proteins for transport into the nucleus of a cell. NLSs are made up of positively charged amino acids, such as Ks and Rs, that are exposed on the protein's surface, with a consensus sequence of -**KK/R**x**K/R**- e. g., -P**KKKRK**V- is the first discovered NLS from the SV40 large antigen. Interestingly, the deletion of an NLS usually disrupts the nuclear import process. In the same way, fusing a non-nuclear protein to an NLS often results in the protein being imported into the nucleus). A large number of invariant Ws are found throughout the sequence.

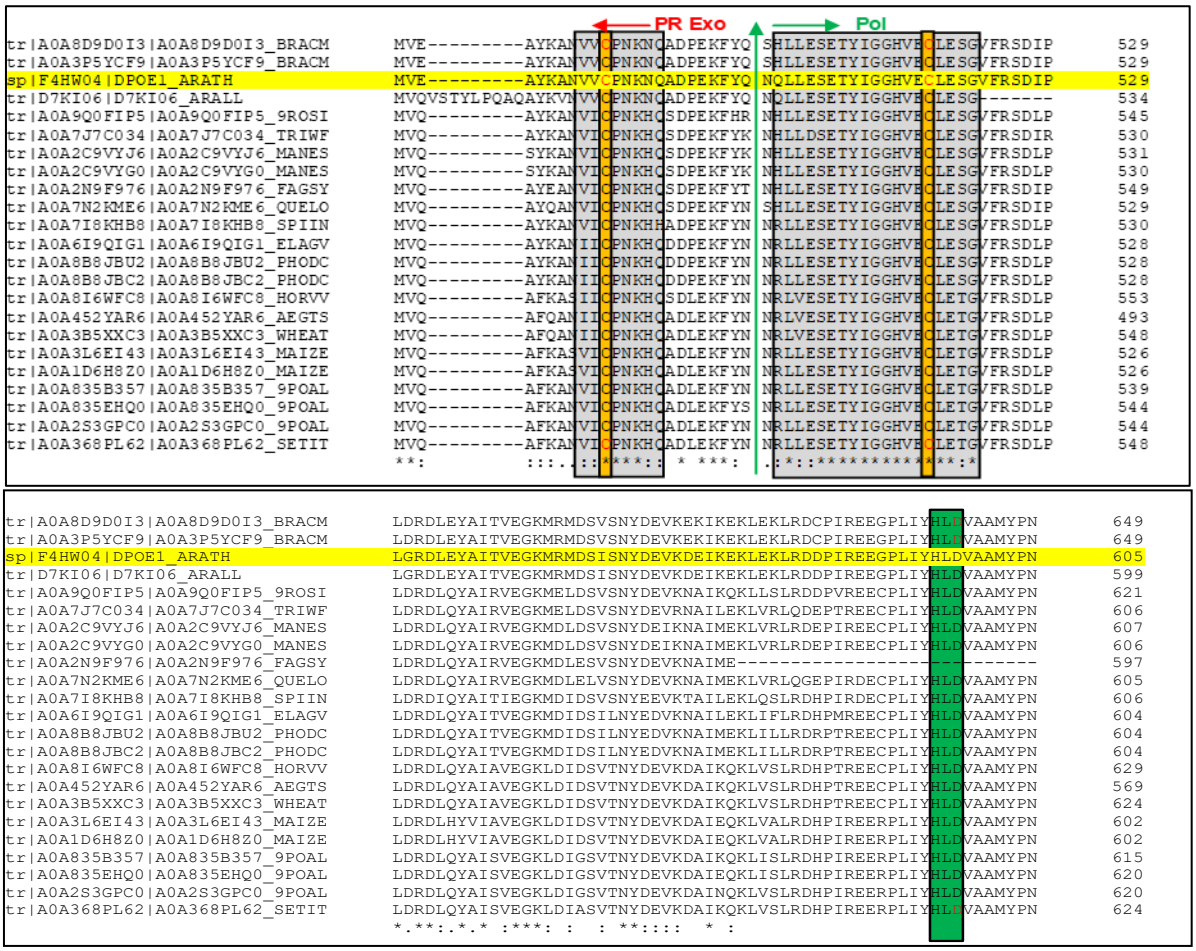
CLUSTAL O (1.2.4) MSA of the ε DNA polymerases from various plant sources

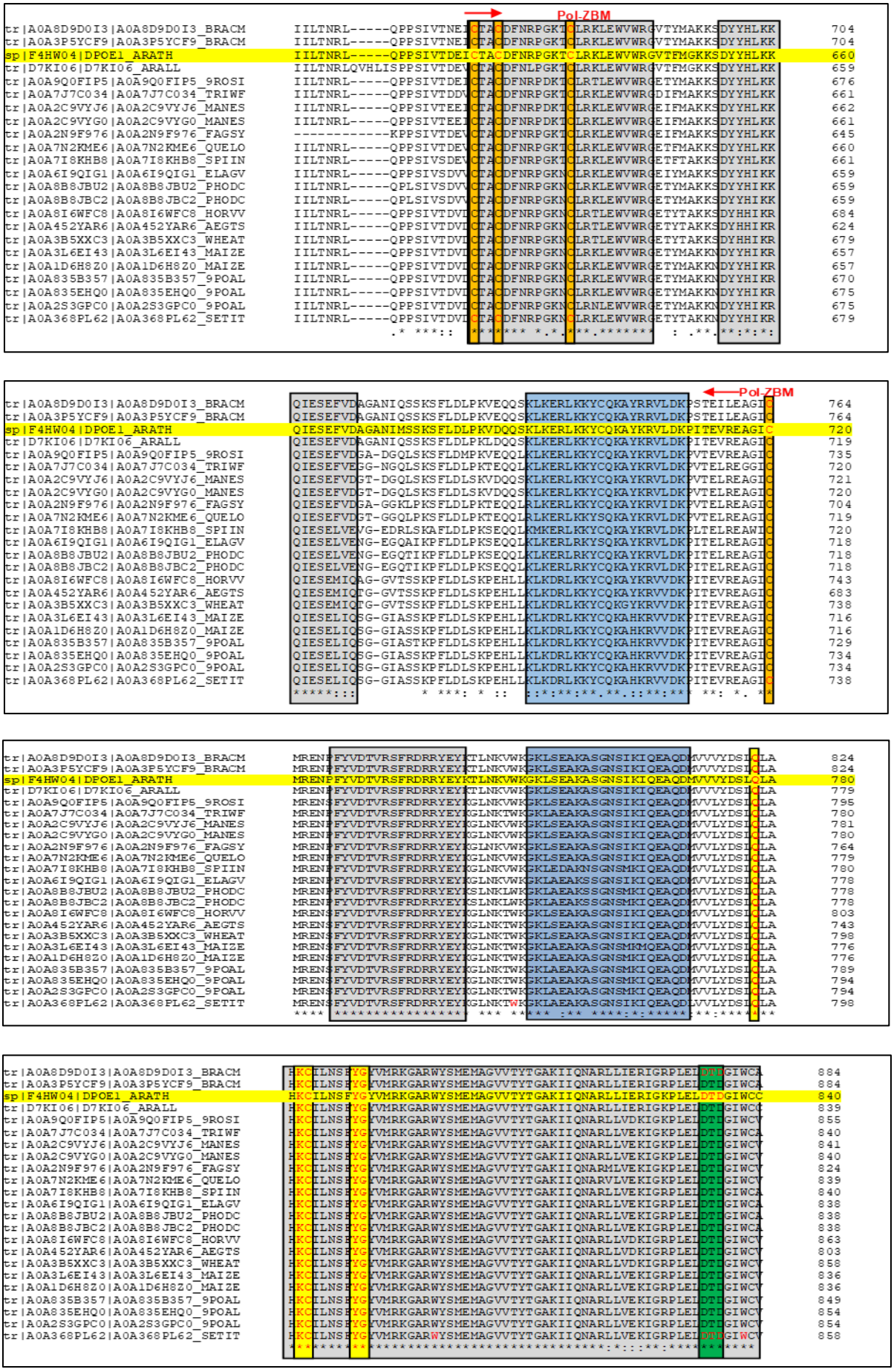


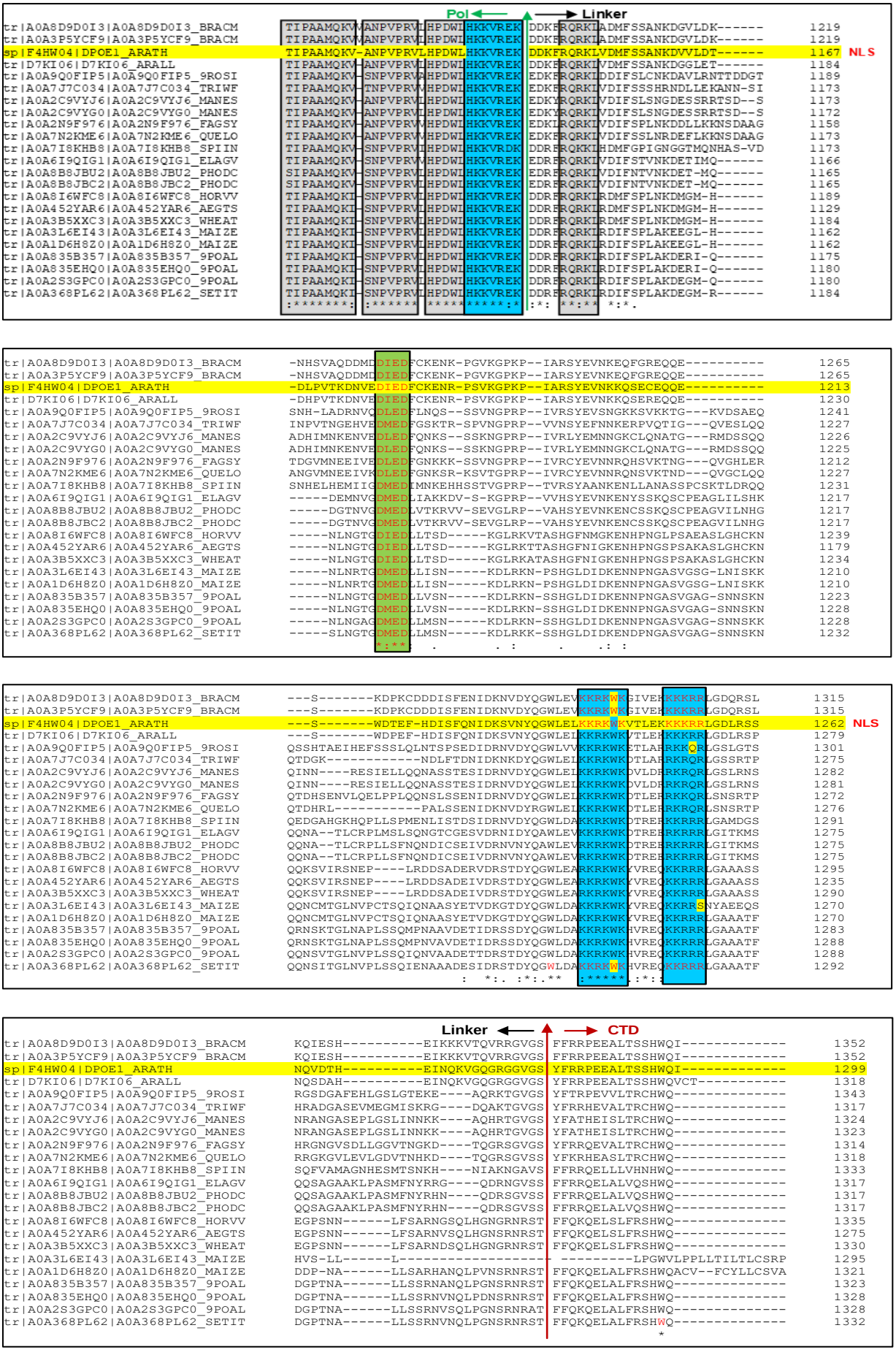

NLS
NTD
PR Exo

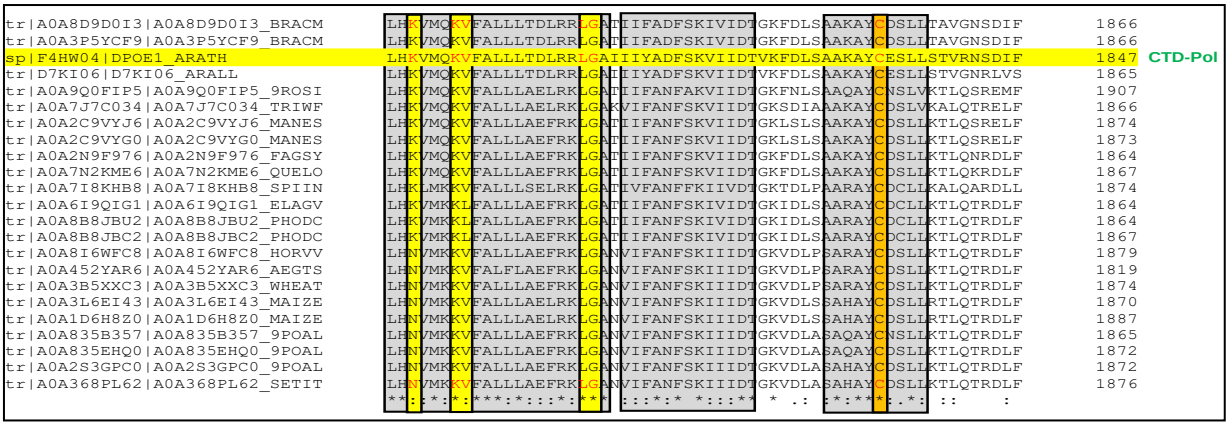


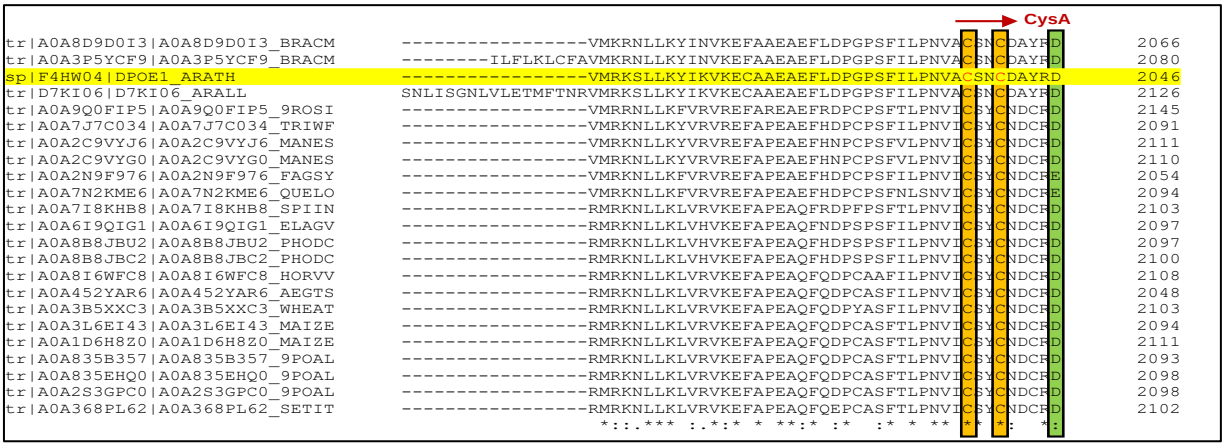


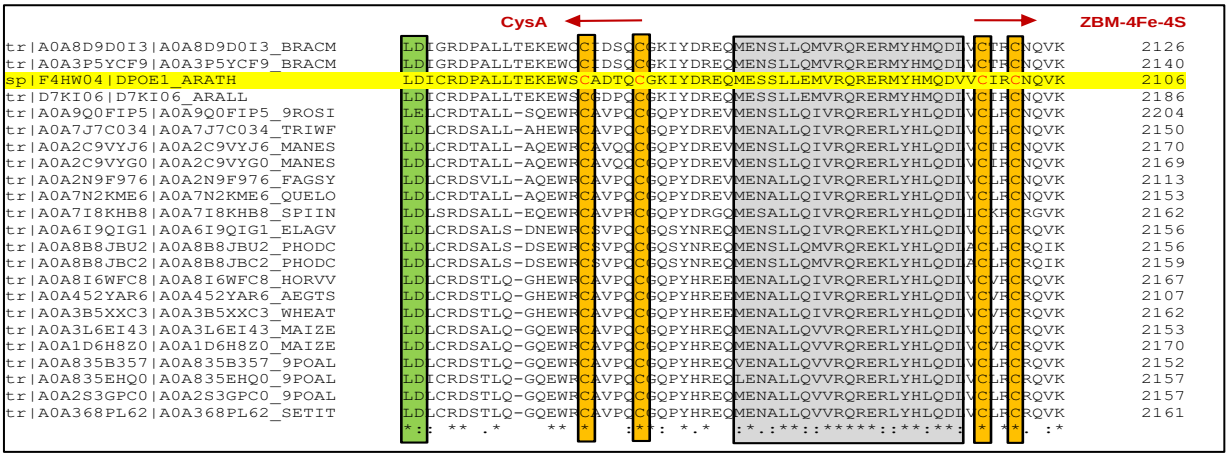

PR Exo
Pol

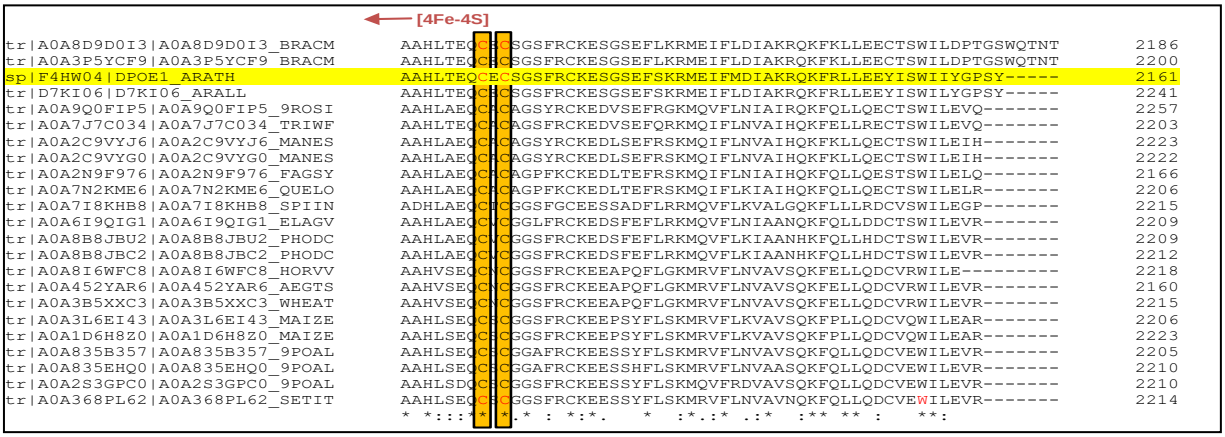

Pol-ZBM
Pol-ZBM


Pol
Linker
NLS
NLS
Linker
CTD


CTD-Pol


CysA


CysA
ZBM-4Fe-4S


[4Fe-4S]

```
//End of the plant pol ε sequences
tr|A0A8D9D0I3|A0A8D9D0I3_BRACM      GRHHS  2191
tr|A0A3P5YCF9|A0A3P5YCF9_BRACM      GRHHS  2205
sp|F4HW04|DPOE1_ARATH               -----  2161
tr|D7KI06|D7KI06_ARALL              -----  2241
tr|A0A9Q0FIP5|A0A9Q0FIP5_9ROSI      -----  2257
tr|A0A7J7C034|A0A7J7C034_TRIWF      -----  2203
tr|A0A2C9VYJ6|A0A2C9VYJ6_MANES      -----  2223
tr|A0A2C9VYG0|A0A2C9VYG0_MANES      -----  2222
tr|A0A2N9F976|A0A2N9F976_FAGSY      -----  2166
tr|A0A7N2KME6|A0A7N2KME6_QUELO      -----  2206
tr|A0A7I8KHB8|A0A7I8KHB8_SPIIN      -----  2215
tr|A0A6I9QIG1|A0A6I9QIG1_ELAGV      -----  2209
tr|A0A8B8JBU2|A0A8B8JBU2_PHODC      -----  2209
tr|A0A8B8JBC2|A0A8B8JBC2_PHODC      -----  2212
tr|A0A8I6WFC8|A0A8I6WFC8_HORVV      -----  2218
tr|A0A452YAR6|A0A452YAR6_AEGTS      -----  2160
tr|A0A3B5XXC3|A0A3B5XXC3_WHEAT      -----  2215
tr|A0A3L6EI43|A0A3L6EI43_MAIZE      -----  2206
tr|A0A1D6H8Z0|A0A1D6H8Z0_MAIZE      -----  2223
tr|A0A835B357|A0A835B357_9POAL      -----  2205
tr|A0A835EHQ0|A0A835EHQ0_9POAL      -----  2210
tr|A0A2S3GPC0|A0A2S3GPC0_9POAL      -----  2210
tr|A0A368PL62|A0A368PL62_SETIT      -----  2214
```

NLS, Nuclear Localization Signal

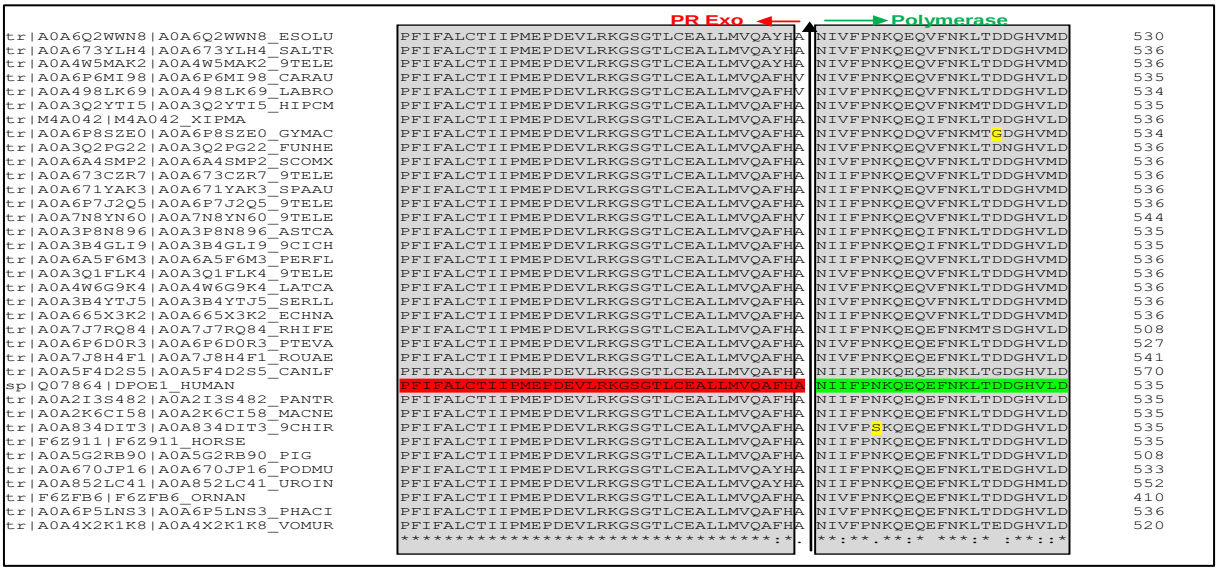
A0A8D9D0I3_BRACM, *Brassica campestris*
A0A3P5YCF9_BRACM, *Brassica campestris*
F4HW04|DPOE1_ARATH, *Arabidopsis thaliana*
D7KI06_ARALL, *Arabidopsis lyrata* subsp. *lyrata*
A0A9Q0FIP5_9ROSI, *Turnerasubulata*
A0A7J7C034_TRIWF, *Tripterygium wilfordii*
A0A2C9VYJ6_MANES, *Manihot esculenta*
A0A2C9VYG0_MANES, *Manihot esculenta*
A0A2N9F976_FAGSY, *Fagus sylvatica*
A0A7N2KME6_QUELO, *Quercus lobata*
A0A7I8KHB8_SPIIN, *Spirodela intermedia*
A0A6I9QIG1_ELAGV, *Elaeisguineensis* var. *tenera*
A0A8B8JBC2_PHODC, *Phoenix dactylifera*
A0A8B8JBU2_PHODC, *Phoenix dactylifera*
A0A8I6WFC8_HORVV, *Hordeum vulgare*
A0A452YAR6_AEGTS, *Aegilops tauschii* subsp. *strangulata*
A0A3B5XXC3_WHEAT, *Triticum aestivum*
A0A3L6EI43_MAIZE, *Zea mays*
A0A1D6H8Z0_MAIZE, *Zea mays*
A0A835B357_9POAL, *Digitaria exilis*
A0A835EHQ0_9POAL, *Digitaria exilis*
A0A2S3GPC0_9POAL, *Panicum hallii*
A0A368PL62_SETIT, *Setariaitalica*

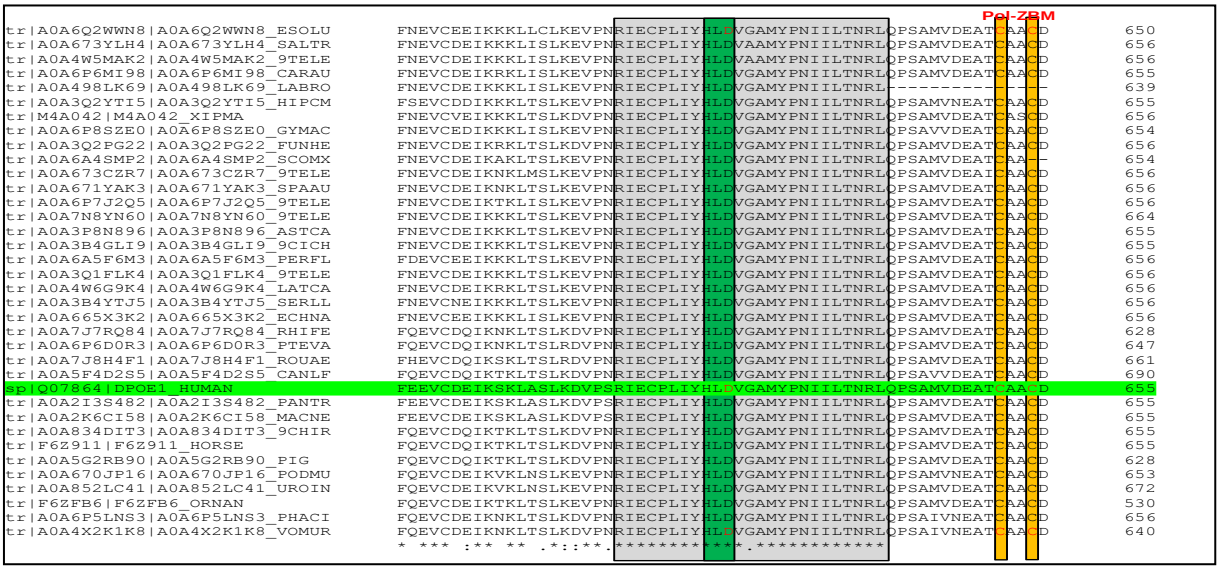
Figure 3 MSA of the catalytic subunit of the DNA pol ε (Pol2A) from various plant sources

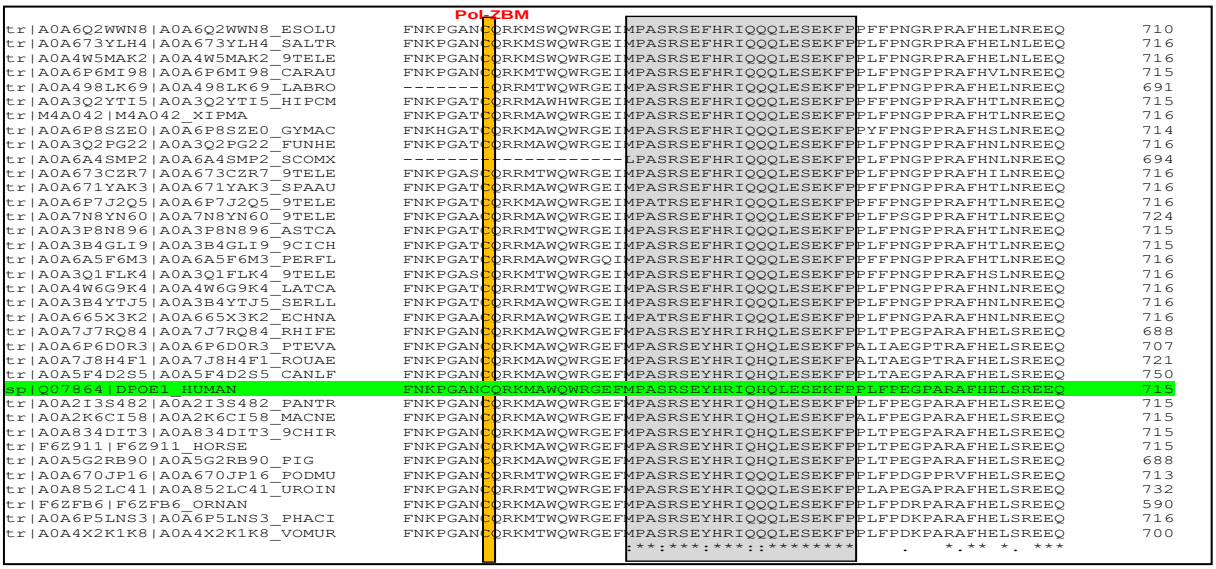
Figure 4 shows the MSA of the DNA pol ε from various animal sources (only the required regions for the discussions are shown here). The human sequence is used as the standard and highlighted. The PR exo and pol regions in the human sequence are highlighted in red and green, respectively, and the PR active site amino acids are highlighted in light blue and the pol active site amino acids are highlighted in yellow. The typical, completely conserved –DEDD- superfamily of PR exonuclease active site amino acids are found with an invariant Y as proton acceptor in the NTD and is in close agreement with the other RdRps/DdDps [8]. The proposed template-binding pair (-YG-), catalytic amino acid (K) and the dNTP selection amino acid (Q) are found in the pol domain and are in close agreement as reported in other DdDps and DdRps [8,16] and also with the plant enzymes (Fig.3). Unlike other two eukaryotic replicative pols, the pol ε harbours a unique ZBM within the pol domain, similar to the pol ε from the plant sources (highlighted). A dNTP selecting **–SYLP**VG**S**- motif is found within the first PR exo region as found in plant enzymes, **–SYLP**QG**S**-. However, a second PR exonuclease active site and a putative pol catalytic core are also identified in the CTD (highlighted). Similar to the **–SYLPXXS-** at the NTD, a **–SYLEP**G**S-** motif (highlighted in yellow) is found within the second PR exo region in the CTD. A large NTD deletion analysis has implicated a possible second pol region in the CTD, e.g., a large N-terminal deletion (Δ176-1134 amino acids) of the yeast *pol ε* gene (which deletes both the pol and PR exo domains), but retains the C-terminal region intact, showed that the yeast was viable, but grew slowly, suggesting a possible second pol active site in the CTD as discussed elsewhere [20]. The CTD contains the two regular ZBMs and one of them (CysB) binds the 4Fe-4S cluster (highlighted in orange) as found in all the three eukaryotic replicative pols. The putative NLSs are highlighted. Many conserved Ws are observed in the CTD, suggesting their possible involvement in interactions with the other subunits.

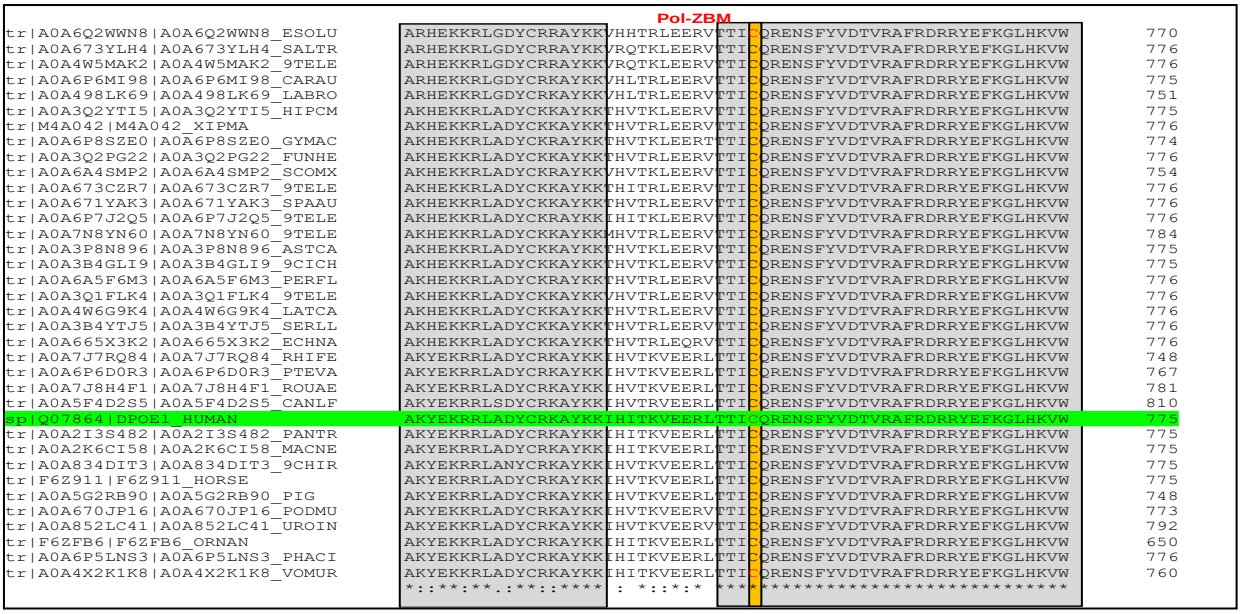
CLUSTAL O (1.2.4) MSA of ε DNA pols from various animal sources


NTD
PR Exo
Exo1

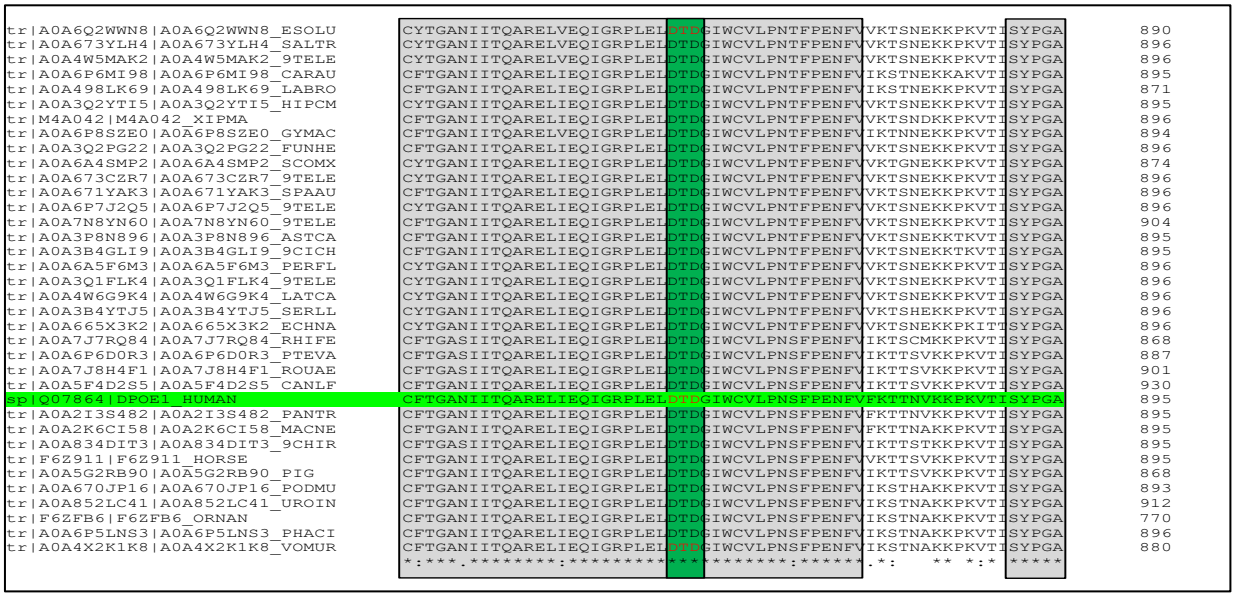

PR Exo
Polymerase

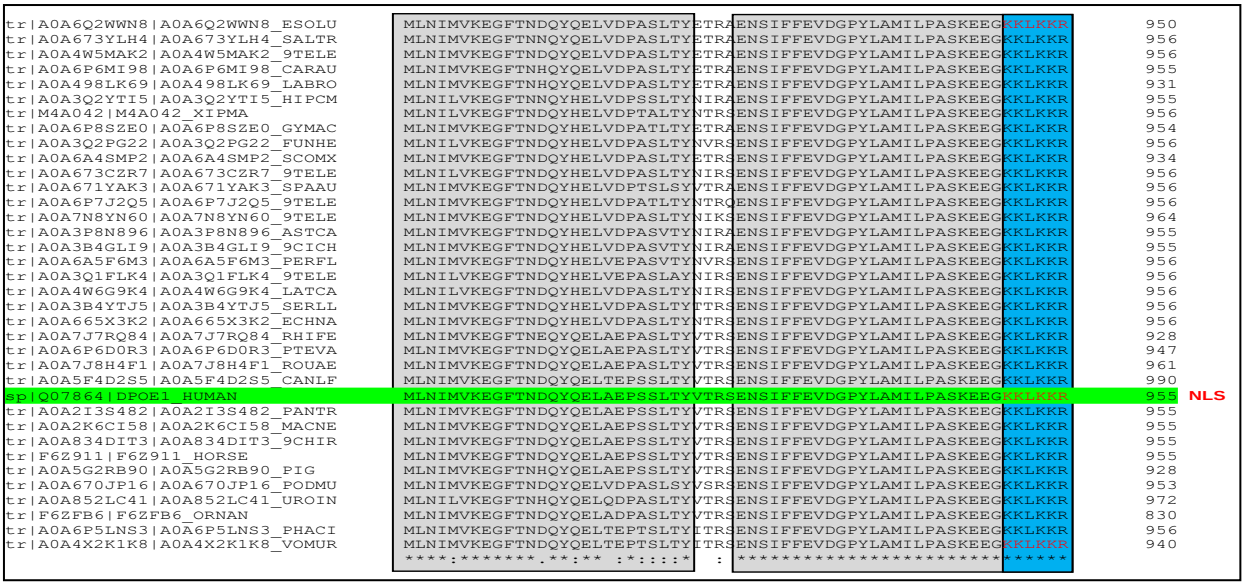

Pol-ZBM

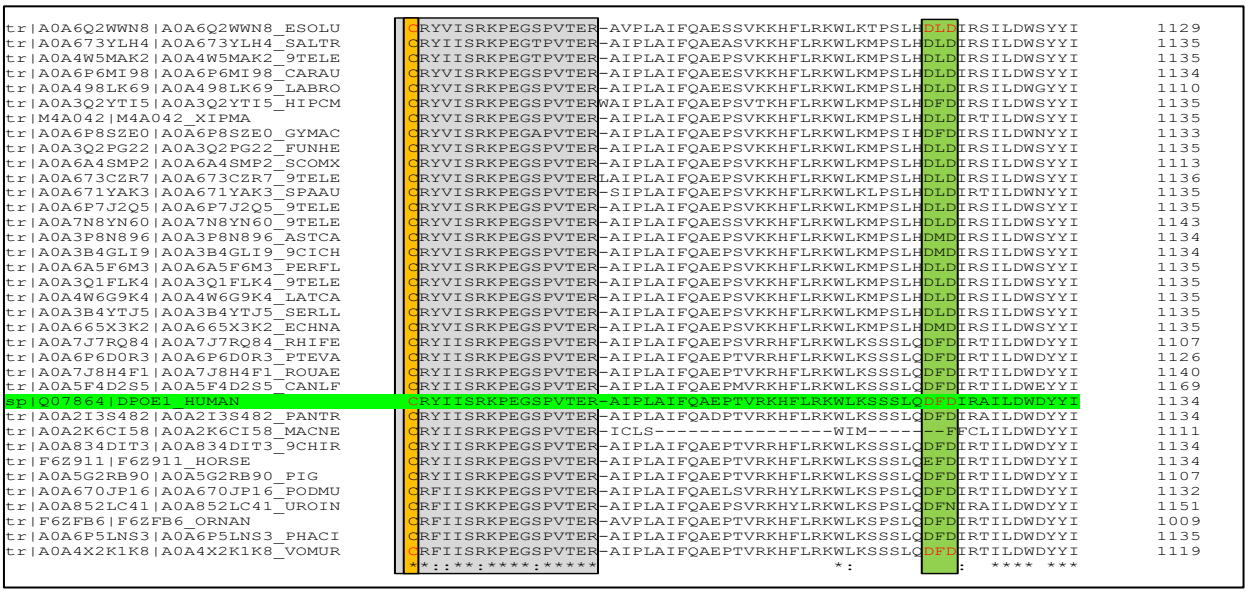

Pol-ZBM

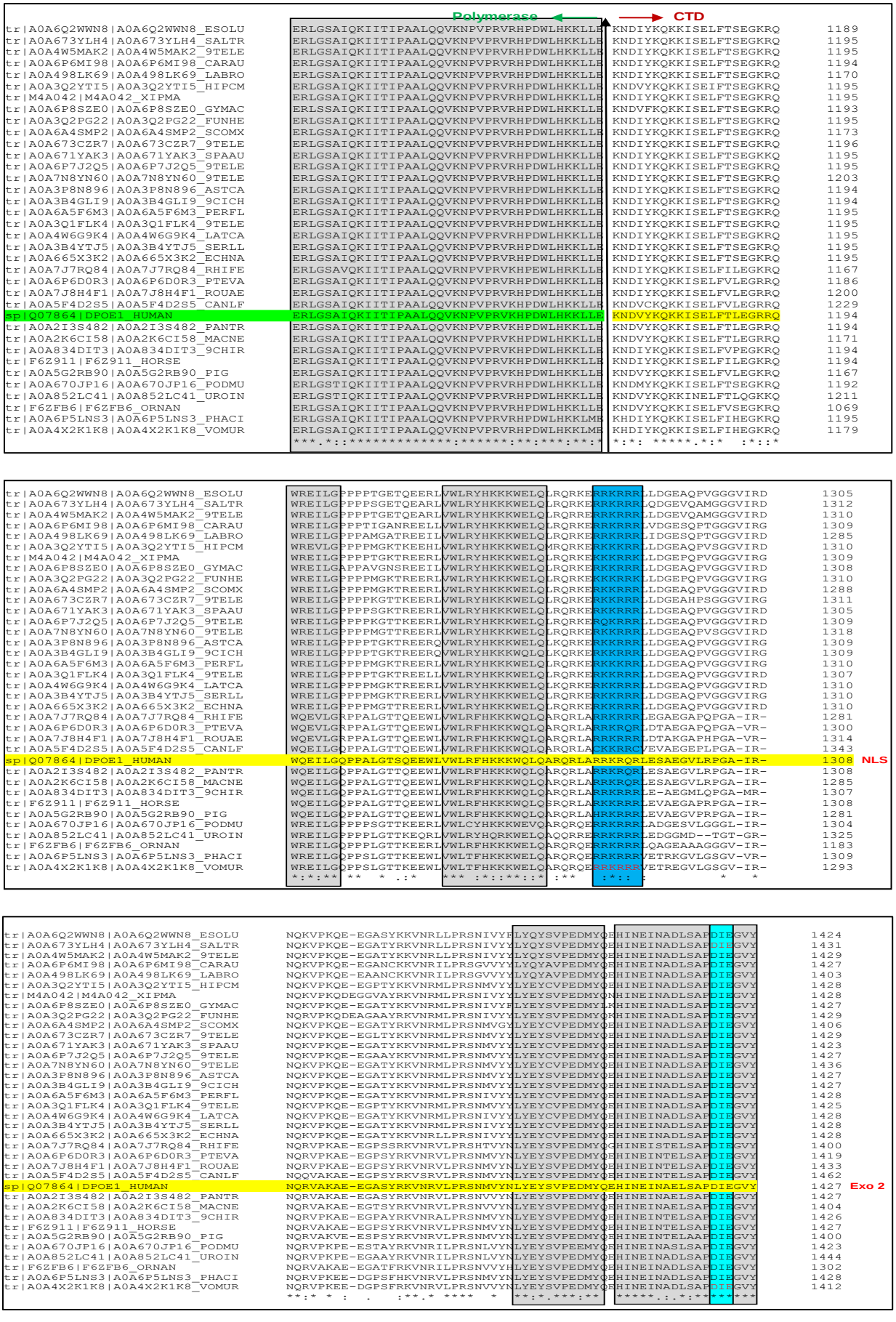

Pol-ZBM

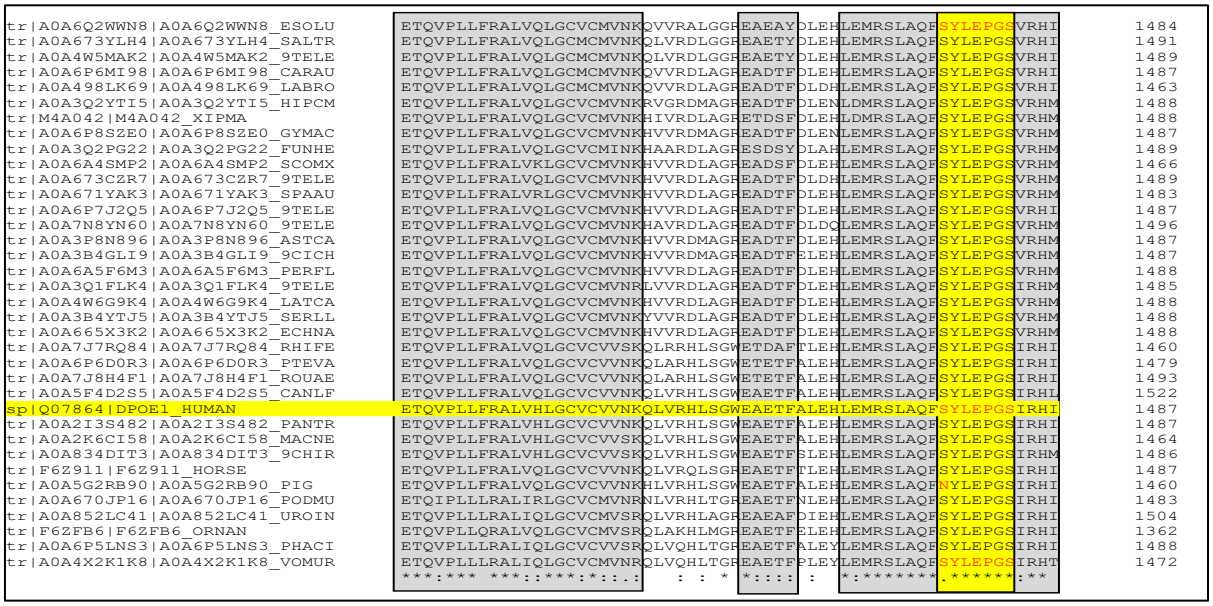

Pol-NTD

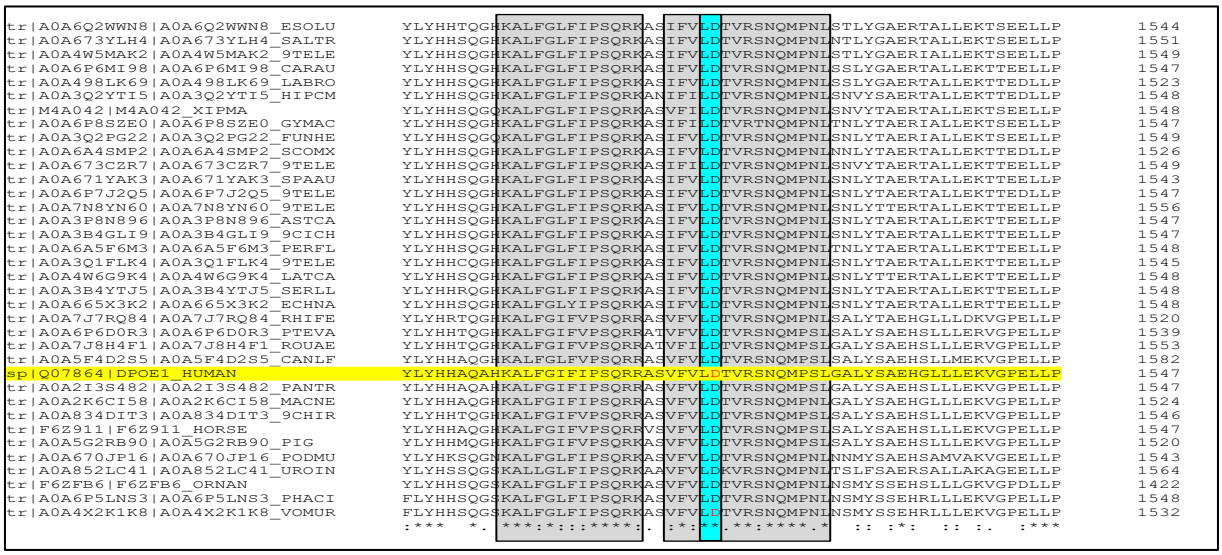


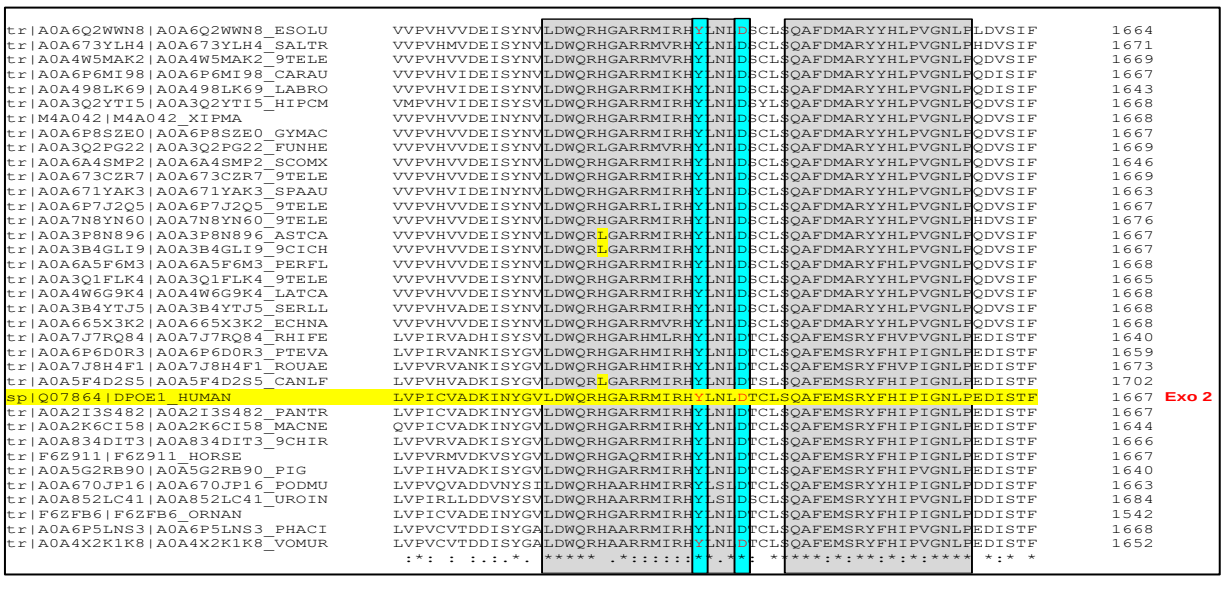

NLS

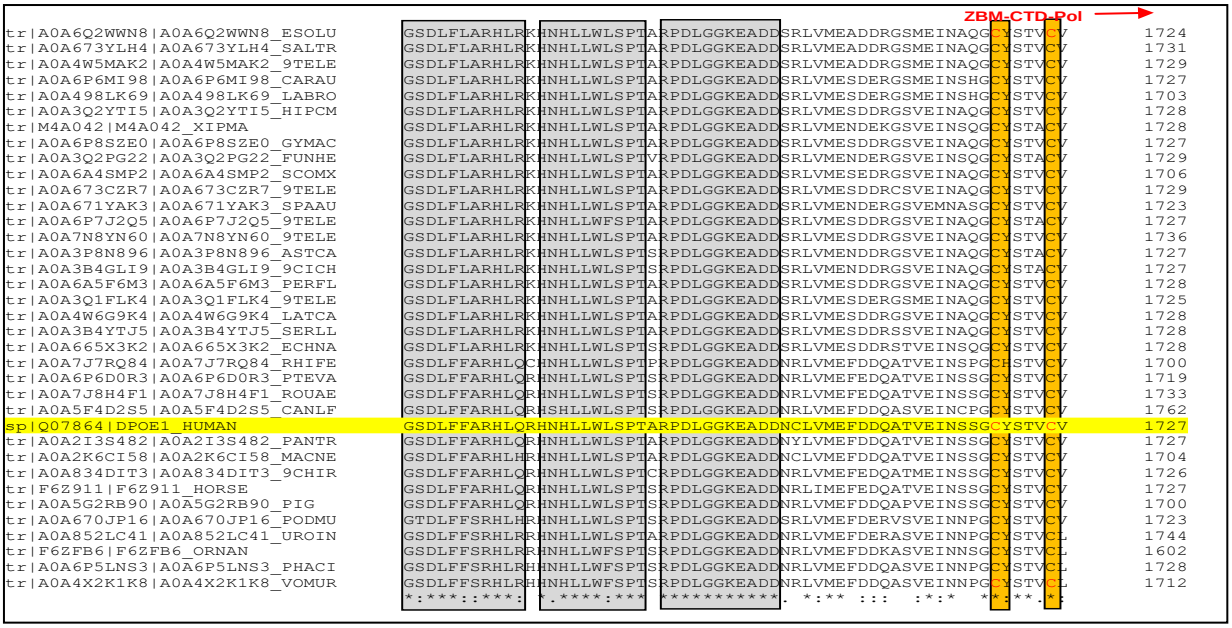


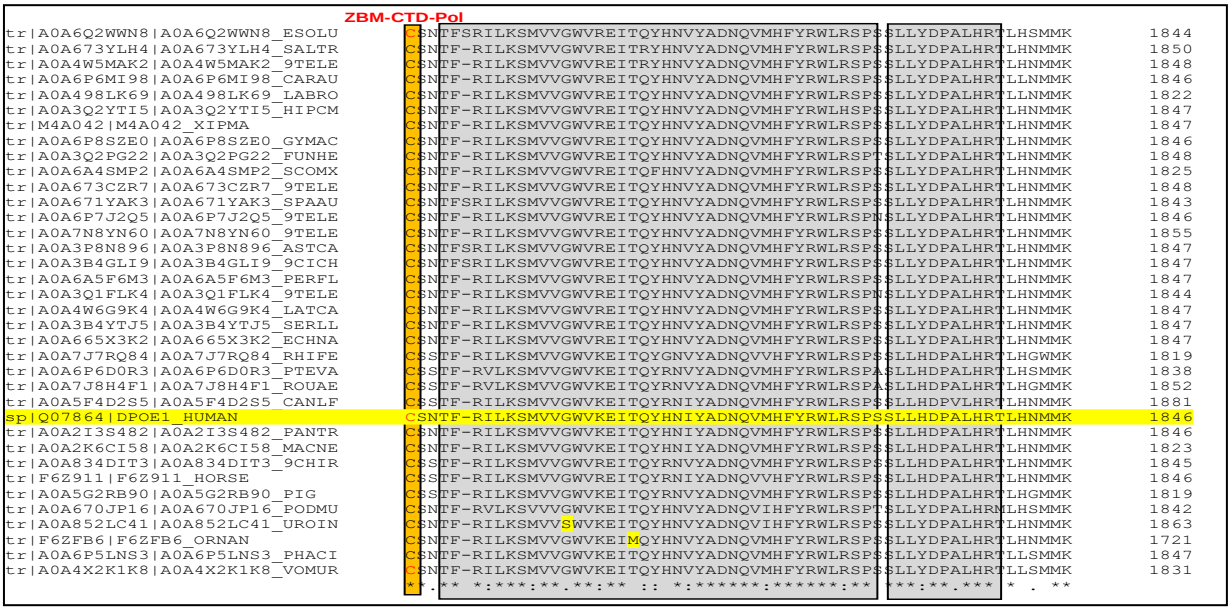

Polymerase
CTD
NLS
Exo 2

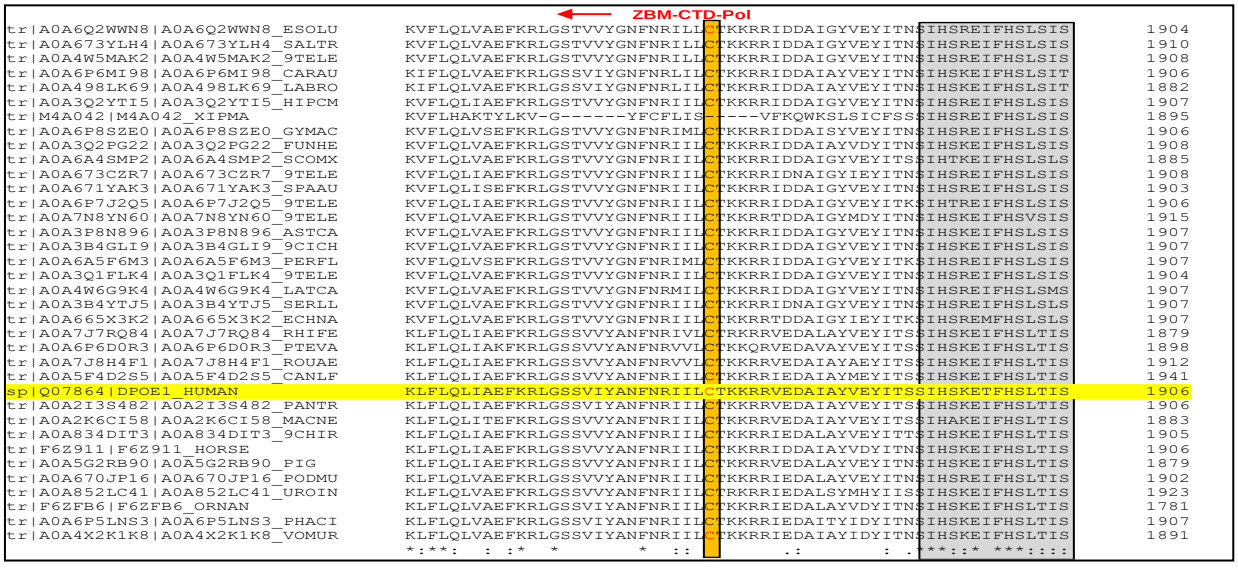


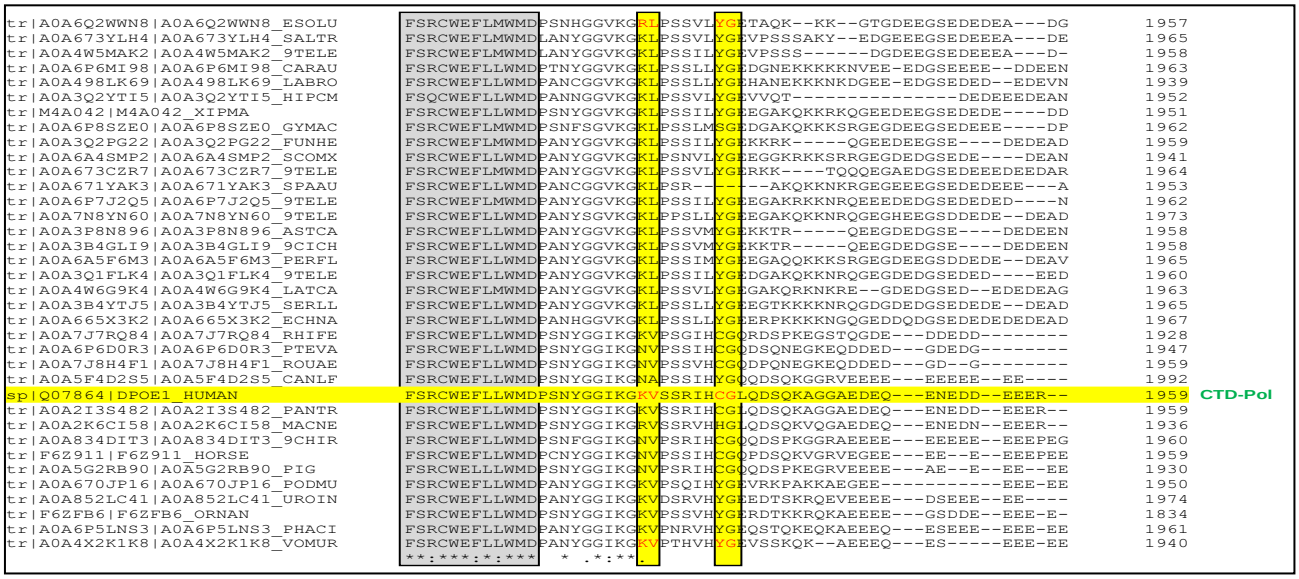


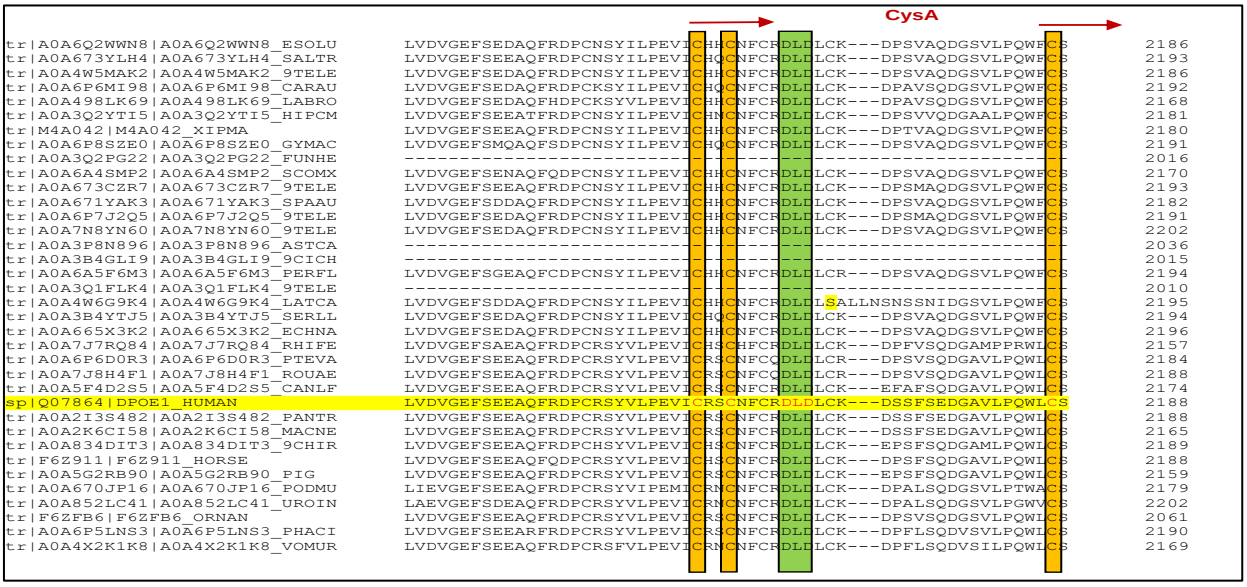

Exo 2

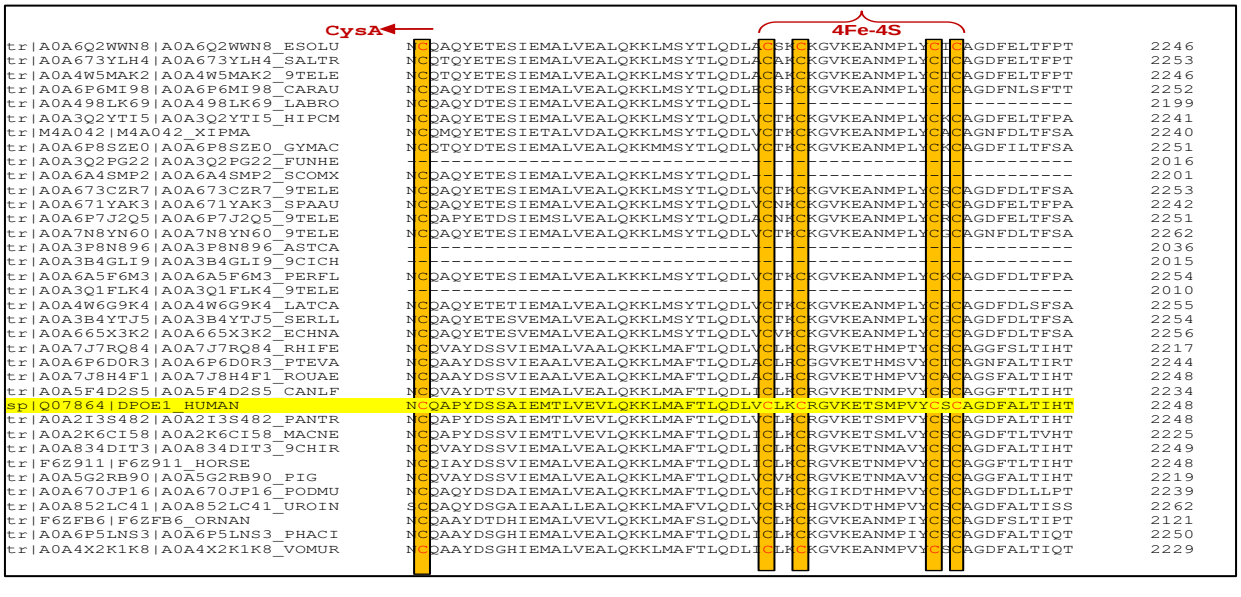

ZBM-CTD-Pol


ZBM-CTD-Pol


ZBM-CTD-Pol


CTD-Pol


CysA


CysA
4Fe-4S

```
//End of ε pol sequences from animal sources
tr|A0A6Q2WWN8|A0A6Q2WWN8_ESOLU      AAHYNMNFLEETIDWLLVMSPQISQAARH-   2288
tr|A0A673YLH4|A0A673YLH4_SALTR      AAHYNMNFLEETIDWLLVMSPQINQSR---   2293
tr|A0A4W5MAK2|A0A4W5MAK2_9TELE      AAHYNMNFLEETIDWLLVMSPQISQSR---   2286
tr|A0A6P6MI98|A0A6P6MI98_CARAU      AAHYNMNFLLETIDWVLSMNA---------   2286
tr|A0A498LK69|A0A498LK69_LABRO      AAHYNMNFLLETIDWVLSMNA---------   2232
tr|A0A3Q2YTI5|A0A3Q2YTI5_HIPCM      ASHYNMSFLKETIDWLMVMSPQISQSAS--   2282
tr|M4A042|M4A042_XIPMA              ASHYNMRFLGETIDWLLEMSPQIGQSRH--   2281
tr|A0A6P8SZE0|A0A6P8SZE0_GYMAC      ASHYSMSYLGETIDWLLVMSPQISQSAL--   2292
tr|A0A3Q2PG22|A0A3Q2PG22_FUNHE      -----------------------------   2016
tr|A0A6A4SMP2|A0A6A4SMP2_SCOMX      ASHYNMSFLEETIDWLLVMSPHVGQSIH--   2236
tr|A0A673CZR7|A0A673CZR7_9TELE      ASHYNMSFLEETIEWLRGMSPQINQSTHKI   2296
tr|A0A671YAK3|A0A671YAK3_SPAAU      ASHYSMSFLEETIDWLVVMSPHINQSAR--   2283
tr|A0A6P7J2Q5|A0A6P7J2Q5_9TELE      ASHYNMTFLEETIDWLLLMSPQISQSAR--   2292
tr|A0A7N8YN60|A0A7N8YN60_9TELE      ASHYNMRFLEETIDWLLVMSPQISQSAH--   2303
tr|A0A3P8N896|A0A3P8N896_ASTCA      -----------------------------   2036
tr|A0A3B4GLI9|A0A3B4GLI9_9CICH      -----------------------------   2015
tr|A0A6A5F6M3|A0A6A5F6M3_PERFL      ASHYKMSFLDETLDWLLVMSPQISQNAH--   2295
tr|A0A3Q1FLK4|A0A3Q1FLK4_9TELE      -----------------------------   2010
tr|A0A4W6G9K4|A0A4W6G9K4_LATCA      ASHYSMSFLEETIDWLLVMSPQISQSAR--   2296
tr|A0A3B4YTJ5|A0A3B4YTJ5_SERLL      ASHYNMSFLEETIDWLLAMSPQISQSAH--   2295
tr|A0A665X3K2|A0A665X3K2_ECHNA      ASHYNMSFLEETIDLLLAMSPQISQTAH--   2297
tr|A0A7J7RQ84|A0A7J7RQ84_RHIFE      AQHYGMSYLMETLEWLLEKNPQLGH-----   2255
tr|A0A6P6D0R3|A0A6P6D0R3_PTEVA      AQHCGMSYLMETLEWLLQRNPQLGH-----   2282
tr|A0A7J8H4F1|A0A7J8H4F1_ROUAE      AQHCGMSYLMETLEWLLQKNPQLGQ-----   2286
tr|A0A5F4D2S5|A0A5F4D2S5_CANLF      AQHYGMSYLTETLEWLLQKNPQLGH-----   2319
sp|Q07864|DPOE1_HUMAN               AQHYGMSYLLETLEWLLQKNPQLGH-----   2286
tr|A0A2I3S482|A0A2I3S482_PANTR      AQHYSMSYLLETLEWLLQKNPQLGR-----   2286
tr|A0A2K6CI58|A0A2K6CI58_MACNE      AQHYGMSYLLETLEWLLQKNPQRGR-----   2263
tr|A0A834DIT3|A0A834DIT3_9CHIR      AQHYGMSYLMETLEWLLQKNPQLSH-----   2287
tr|F6Z911|F6Z911_HORSE              AQHYGMSYLMESLEWLLQKNPQLGH-----   2286
tr|A0A5G2RB90|A0A5G2RB90_PIG        AQHYGMSYLTETLEWLLQQNPQLGH-----   2257
tr|A0A670JP16|A0A670JP16_PODMU      ARHYNMDYLLENIKWLLQMNPQLLL-----   2277
tr|A0A852LC41|A0A852LC41_UROIN      ARHYSMAHLLETIEWLLSTNPQLRP-----   2300
tr|F6ZFB6|F6ZFB6_ORNAN              AQHYAMSYLLETIEWLLQMNPRLPR-----   2159
tr|A0A6P5LNS3|A0A6P5LNS3_PHACI      ARHYGMTYLTEMIEWLLQMNPQLAN-----   2288
tr|A0A4X2K1K8|A0A4X2K1K8_VOMUR      ARHYGMTYLTEMIEWLLQMNPQLAN-----   2267
```

A0A6Q2WWN8_ESOLU, Esox lucius
A0A673YLH4_SALTR, Salmo trutta
A0A4W5MAK2_9TELE, Hucho hucho
A0A6P6MI98_CARAU, Carassius auratus
A0A498LK69_LABRO, Labeo rohita
A0A3Q2YTI5_HIPCM, Hippocampus comes
M4A042_XIPMA, Xiphophorus maculates
A0A6P8SZE0_GYMAC, Gymnodraco acuticeps
A0A3Q2PG22_FUNHE, Fundulus heteroclitus
A0A6A4SMP2_SCOMX, Scophthalmus maximus
A0A673CZR7_9TELE, Sphaeramia orbicularis
A0A671YAK3_SPAAU, Sparus aurata
A0A6P7J2Q5_9TELE, Parambassis ranga
A0A7N8YN60_9TELE, Mastacembelus armatus
A0A3P8N896_ASTCA, Astatotilapia calliptera
A0A3B4GLI9_9CICH, Pundamilia nyererei
A0A6A5F6M3_PERFL, Perca fluviatilis
A0A3Q1FLK4_9TELE, Acanthochromis polyacanthus
A0A4W6G9K4_LATCA, Lates calcarifer
A0A3B4YTJ5_SERLL, Seriola lalandi dorsalis
A0A665X3K2_ECHNA, Echeneis naucrates
A0A7J7RQ84_RHIFE, Rhinolophus ferrumequinum
A0A6P6D0R3_PTEVA, Pteropus vampyrus
A0A7J8H4F1_ROUAE, Rousettus aegyptiacus
A0A5F4D2S5_CANLF, Canis lupus familiaris
Q07864|DPOE1_HUMAN, Homo sapiens
A0A2I3S482_PANTR, Pan troglodytes
A0A2K6CI58_MACNE, Macaca nemestrina
A0A834DIT3_9CHIR, Phyllostomus discolor
F6Z911_HORSE, Equus caballus
A0A5G2RB90_PIG, Sus scrofa
A0A670JP16_PODMU, Podarcis muralis
A0A852LC41_UROIN, Urocolius indicus
F6ZFB6_ORNAN, Ornithorhynchus anatinus
A0A6P5LNS3_PHACI, Phascolarctos cinereus
A0A4X2K1K8_VOMUR, Vombatus ursinus

Figure 4 MSA of ε DNA polymerases from various animal sources

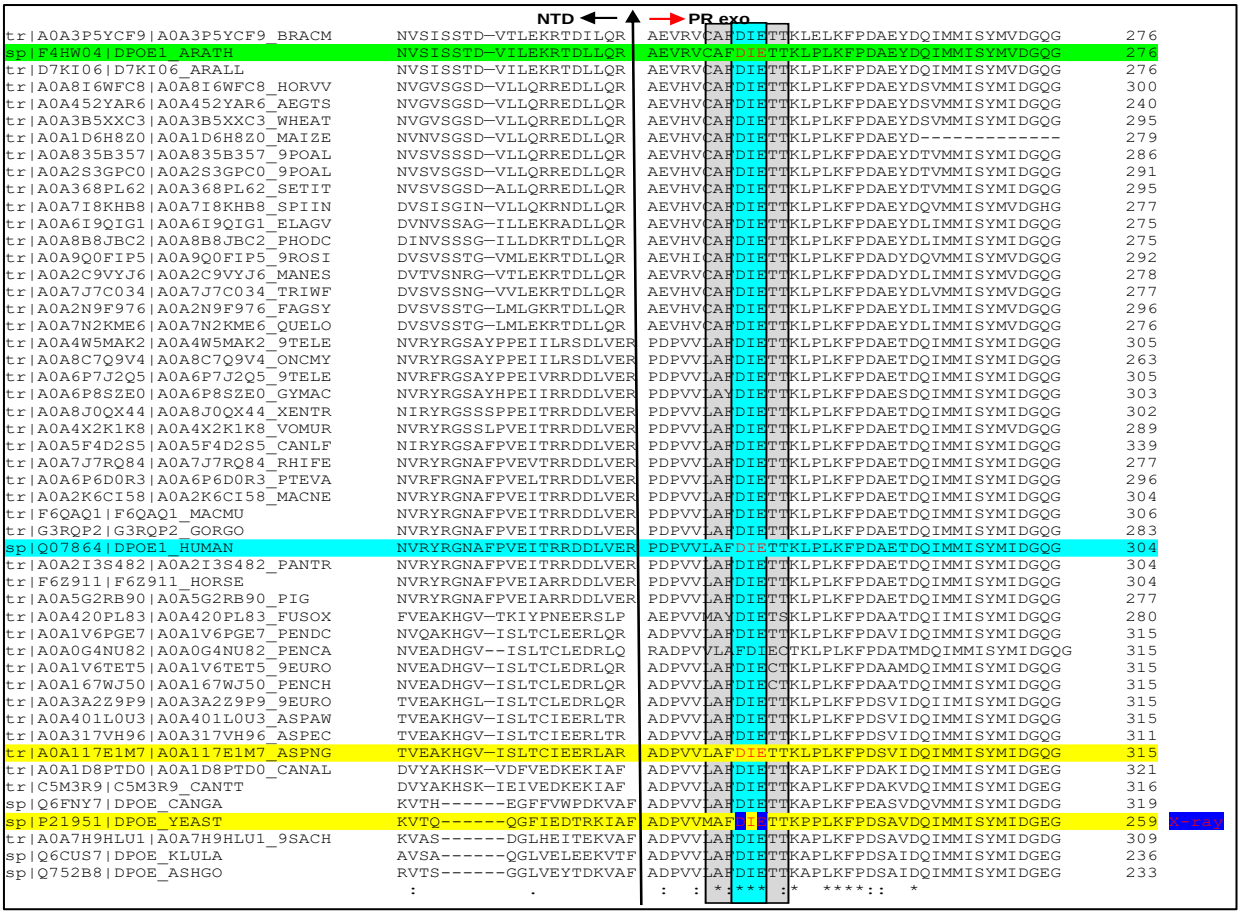
3.2. 'Mix and Match' MSA analysis of the DNA pols ε from yeasts, higher fungi, plants and animals

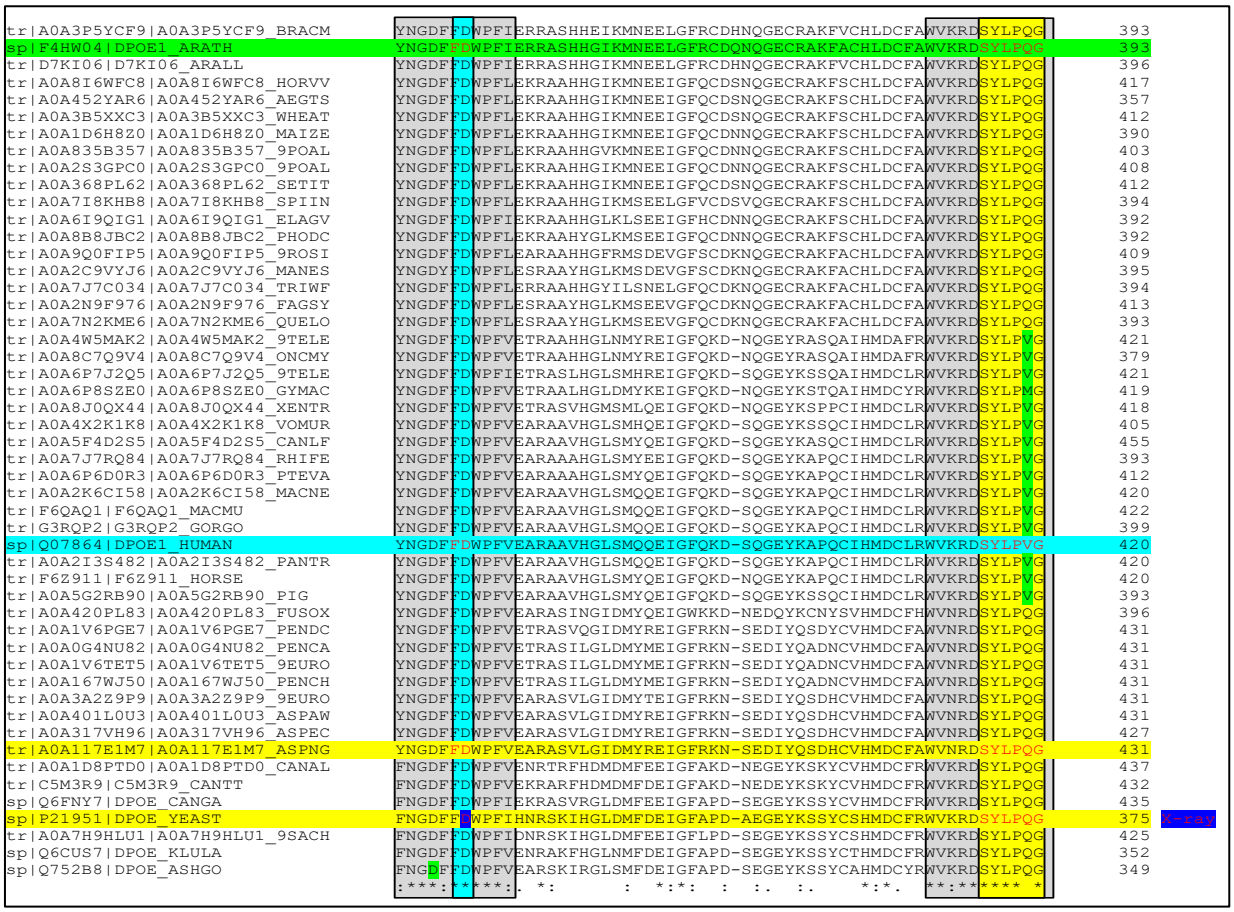
Figure 5 shows the 'Mix and Match' MSA of the ε pols from all the four eukaryotic sources, viz. yeasts, higher fungi, plants and animals (only the required regions for the discussions are shown). *S. cerevisiae*, *Aspergillus niger*, *A. thaliana* and human sequences are used as the standards and highlighted. All the major domains (NTD, PR Exo, Pol, Linker and CTD) are indicated tentatively with arrow marks. The NTDs show many gaps in the alignment, but with a small number of conserved peptides. Strikingly, a highly conserved dodecapeptide is found in this region. The NTD is followed by a highly conserved PR exo domain. The proposed PR exonuclease active site amino acids are completely conserved in all (highlighted) and it belongs to the DEDD-superfamily with an invariant Y as a proton acceptor. The presence of a typical –DEDD-superfamily exonuclease active site in all ε pols is in close agreement with the other DdDps [8]. An invariant, characteristic motif –**SYLP**Q/VGS- (highlighted in yellow) is found within the PR exo domain in the ε pols of all eukaryotes. This sequence is somewhat similar to the characteristic –**SLYP**S- motif reported in the other two replicative pols, α and δ [2] and, in general, in all the B-family pols. In all these pols, viz. ***ε,*** δ and α, this motif is implicated in dNTP-binding.

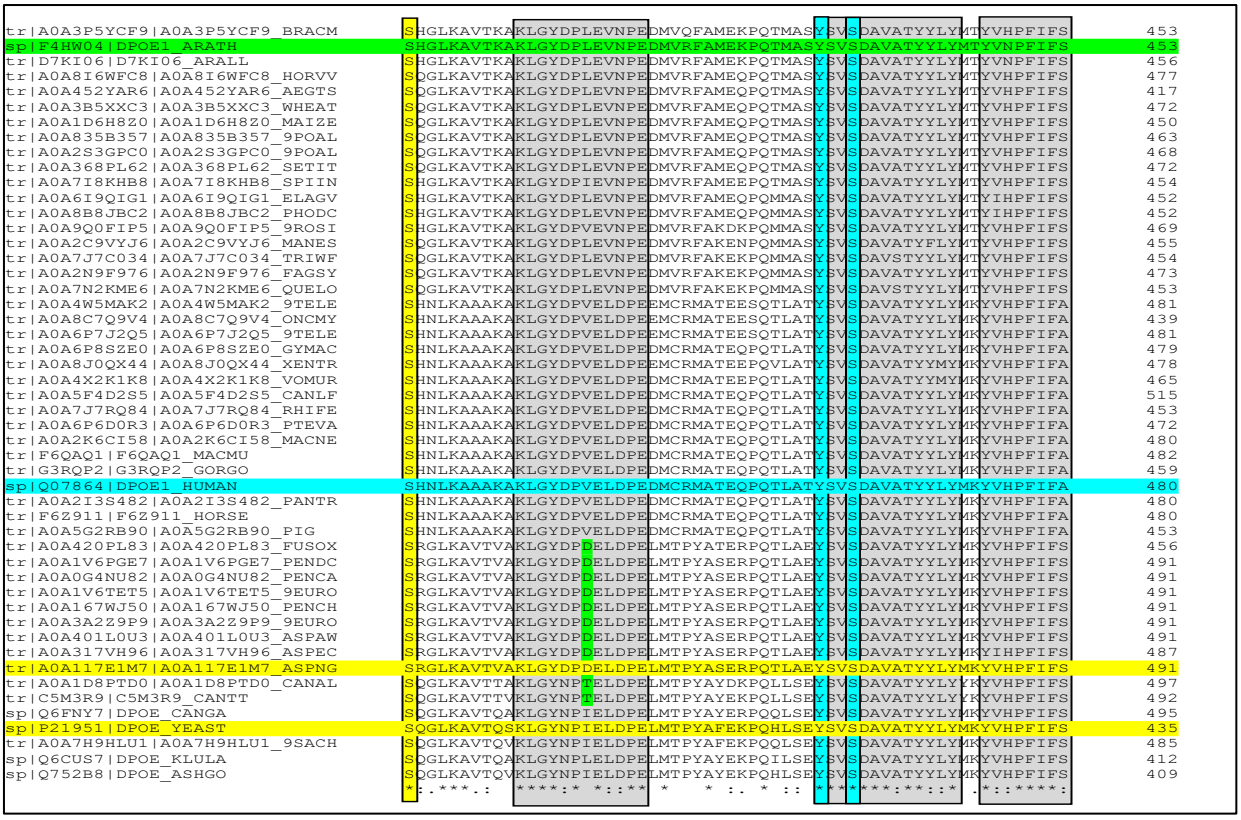
The PR exo domain is followed by the pol domain and is also highly conserved in all. Interestingly, the template-binding pair (-YG-), catalytic amino acid (K) and dNTP selection amino acid (Q) are completely conserved in all (highlighted in yellow) and are in close agreement with other DdDps and DdRps [6, 21]. The ε pol's active site amino acids are found within highly conserved large peptides. The pol domain also contains the completely conserved–DxD- type catalytic Mg^{2+} site and is in close agreement with other DNA pols. A unique ZBM that is conserved within the pol domain in all ε pols (highlighted) is suggested to play a role in linking redox status of cells to replication. The pol domain is followed by a linker region which connects the pol to the CTD. The linker region contains a highly conserved NLS-like motif in all pols ε (highlighted). The linker region is not conserved much. The CTD possesses two invariant ZBMs and one of them binds to the 4Fe-4S cluster (highlighted in orange) and a –DxD- type metal-binding motif is found within the first ZBM. The second pol and PR exo active sites do not align in all. Interestingly, the catalytic amino acid K, pairs with –C- as -KC- in plant and animal enzymes, but with -V- as–KV- in yeasts and higher fungal enzymes.

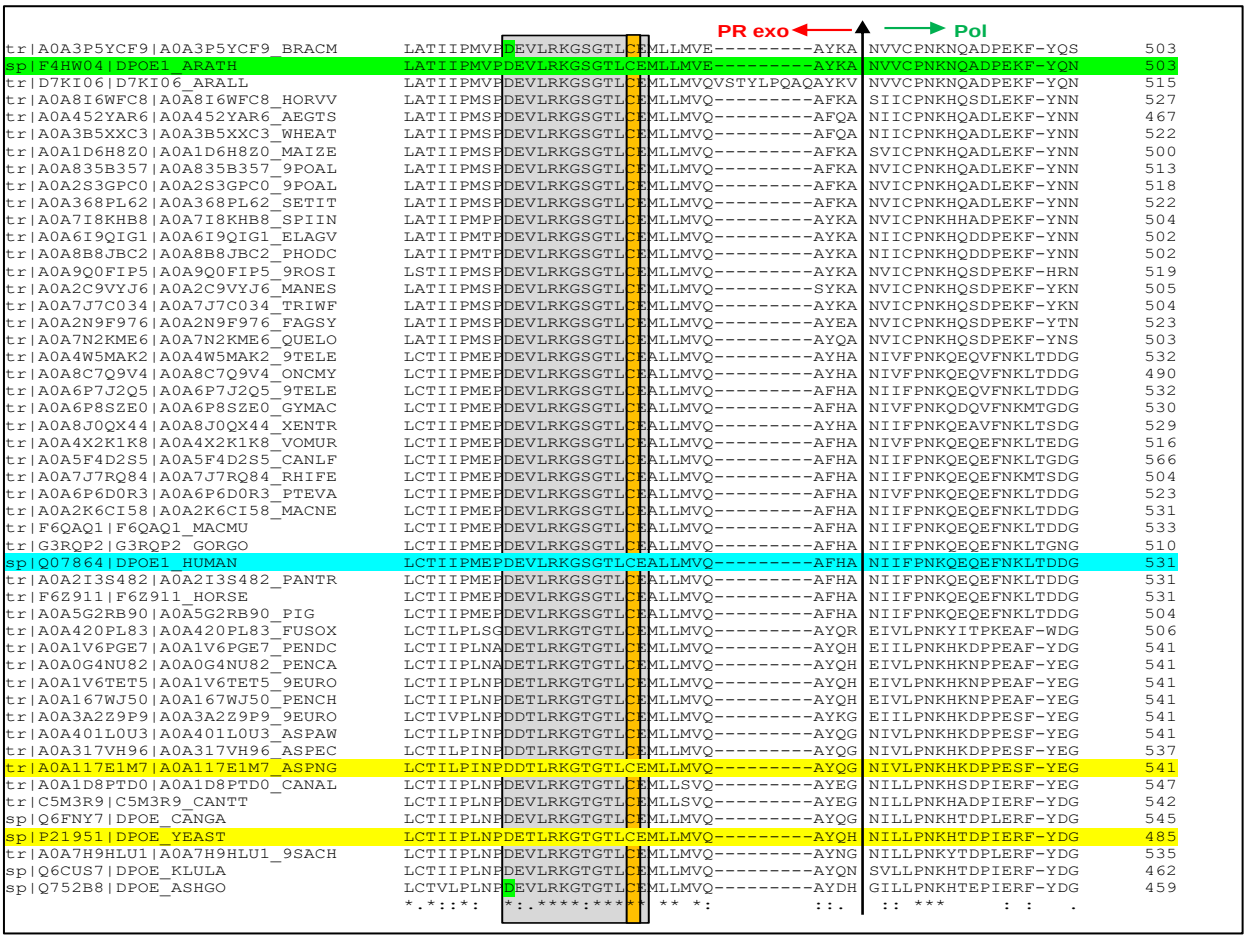
CLUSTAL O (1.2.4) 'Mix and Match' MSA of all the four eukaryotic ε DNA pols (from yeasts, higher fungi, plants and animals)

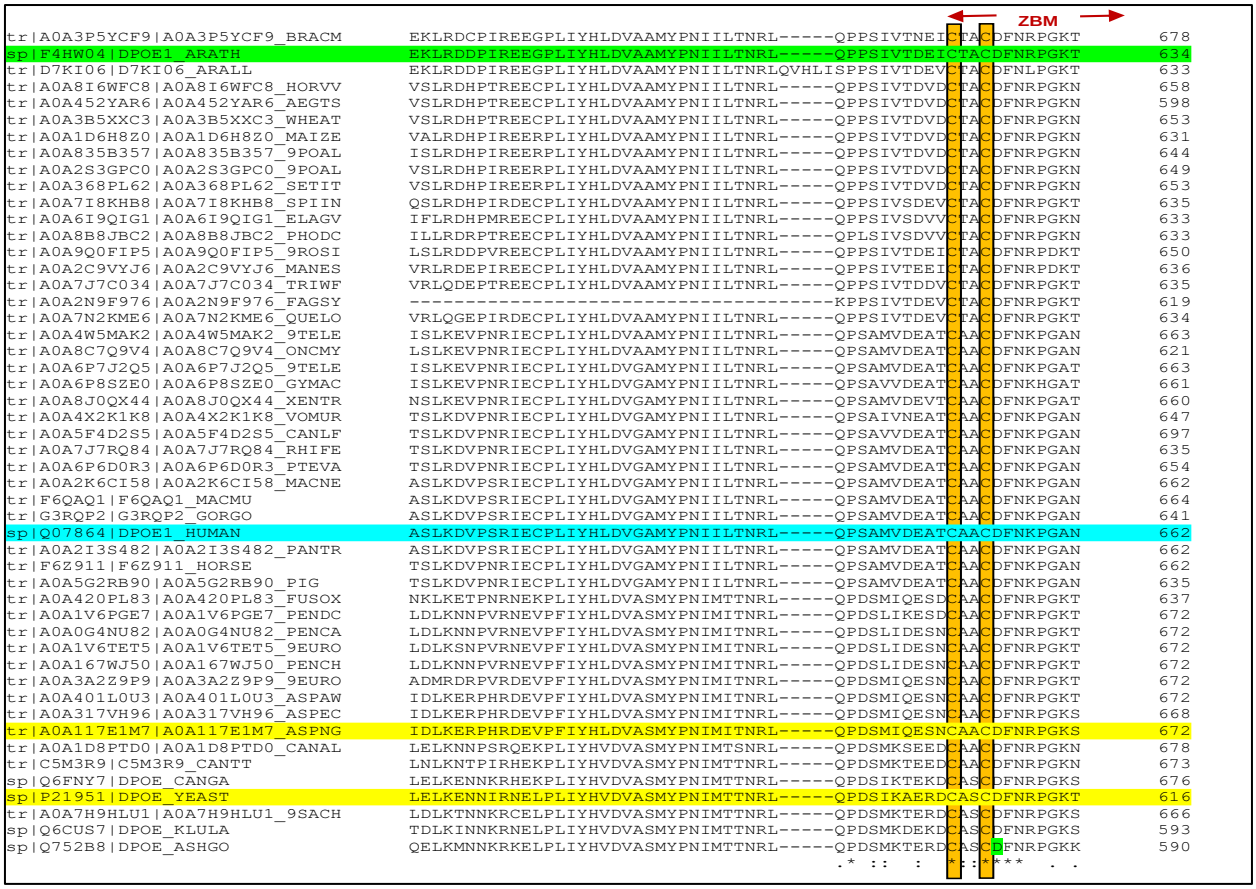
```
tr|A0A3P5YCF9|A0A3P5YCF9_BRACM      ILAGK-----------------REQRLQDCLDSLLDLREYDVPYHVRFAVDKDVRSGQWY      218
sp|F4HW04|DPOE1_ARATH               ILAGK-----------------REQRPQDCLDSIVDLREYDVPYHVRFAIDNDVRSGQWY      218
tr|D7KI06|D7KI06_ARALL              ILAGK-----------------REQRPQDCLDSIVDLREYDVPYHVRFAIDNDVRSGQWY      218
tr|A0A8I6WFC8|A0A8I6WFC8_HORVV      IYGVK-----------------RVERPQDYINCIIDLREYDVPYHVRFAIDNDVRSGQWY      242
tr|A0A452YAR6|A0A452YAR6_AEGTS      IYGVK-----------------RVERPQDYINCIIDLREYDVPYHVRFAIDNDVRSGQWY      182
tr|A0A3B5XXC3|A0A3B5XXC3_WHEAT      IYGVK-----------------RVERPQDYINCIIDLREYDVPYHVRFAIDNDVRSGQWY      237
tr|A0A1D6H8Z0|A0A1D6H8Z0_MAIZE      IHGVK-----------------RVERPQDYINYIIDLREYDVPYHVRFAIDNDVRSGQWY      234
tr|A0A835B357|A0A835B357_9POAL      IHGVK-----------------RVERPQDYINYIIDLREYDVPYHVRFAIDNDVRCGQWY      228
tr|A0A2S3GPC0|A0A2S3GPC0_9POAL      IHGVK-----------------RVERPQDYMNYIIDLREYDVPYHVRFAIDNDVRCGQWY      233
tr|A0A368PL62|A0A368PL62_SETIT      IHGVK-----------------RVERPQDYINYIIELREYDVPYHVRFAIDKDVRCGQWY      237
tr|A0A7I8KHB8|A0A7I8KHB8_SPIIN      IFTGK-----------------SKERPQDFIDCIVDLREYDVPYHVRFAIDNDVRCGEWY      219
tr|A0A6I9QIG1|A0A6I9QIG1_ELAGV      IYNGK-----------------SKERPQDYIDCITDLREYDVPYHVRFAIDNDVRCGQWY      217
tr|A0A8B8JBC2|A0A8B8JBC2_PHODC      MYNGK-----------------SKERPQDYIDCITDLREYDVPYHVRFAIDNDVRCGQWY      217
tr|A0A9Q0FIP5|A0A9Q0FIP5_9ROSI      ILSGKR--QVFHDHMHELTHYFLEQRLQDFLDSIIDLREYDVPYHVRFAIDNDIRCGQWY      234
tr|A0A2C9VYJ6|A0A2C9VYJ6_MANES      LLTGKS-----------------REQRPQDLLDCIVDLREYDVPYHVRFAIDNDIRCGQWY      220
tr|A0A7J7C034|A0A7J7C034_TRIWF      ILTGK-----------------REQRPQDFLDCVVDLREYDVPYHVRFAIDNDVRCGQWY      219
tr|A0A2N9F976|A0A2N9F976_FAGSY      FDISV-----------------ENKGPQDFIDCIVDLREYDVPYHVRFAIDNDVRCGQWY      238
tr|A0A7N2KME6|A0A7N2KME6_QUELO      IIAGK-----------------REQRPQDFLDCIVDLREYDVPYHVRFAIDNDVRCGQWY      218
tr|A0A4W5MAK2|A0A4W5MAK2_9TELE      MLSSALAGGSV--SAA--DEDGMSKKTSDQMDNIVDMREYDVPYHVRLSIDLKIHVAHWY      245
tr|A0A8C7Q9V4|A0A8C7Q9V4_ONCMY      MLSSALAGGSV--SAA--DEDGLSKKTSDQMDNIVDMREYDVPYHVRLSIDLKIHVAHWY      203
tr|A0A6P7J2Q5|A0A6P7J2Q5_9TELE      MLSSALSGGTV--ASA--DEDSISKSISDQLDNIVDMREYDVPYHVRVSIDLKIHVAHWY      245
tr|A0A6P8SZE0|A0A6P8SZE0_GYMAC      MLSSALSGGNV--NTA--EEDGPSKRISDQLDNIVDMREYDVPYHMRLSIDLKIHVAHWY      243
tr|A0A8J0QX44|A0A8J0QX44_XENTR      MLSSALTGNNM--GA---EEEGPSKKISDQMENIVDMREYDVPYHVRVSIDLKIHVAHWY      242
tr|A0A4X2K1K8|A0A4X2K1K8_VOMUR      MLSSALMGGSV--TT---DEEGSSKKIADQLDNVVDMREYDVPYHIRLSIDLKIHVAHWY      229
tr|A0A5F4D2S5|A0A5F4D2S5_CANLF      MLSSVLQGGSV--VI---DEEEASKKVTDQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      279
tr|A0A7J7RQ84|A0A7J7RQ84_RHIFE      MLSSVLQGGSV--IL---DEEEASKKTADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      217
tr|A0A6P6D0R3|A0A6P6D0R3_PTEVA      MLSSVLQGGGA--II---DEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      236
tr|A0A2K6CI58|A0A2K6CI58_MACNE      LLSSVLQRGGV--IT---DEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      244
tr|F6QAQ1|F6QAQ1_MACMU             LLSSVLQRGGV--IT---DEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      246
tr|G3RQP2|G3RQP2_GORGO             LLSSVLQRGGV--IT---DEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      223
sp|Q07864|DPOE1_HUMAN              LLSSVLQRGGV--IT---DEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      244
tr|A0A2I3S482|A0A2I3S482_PANTR      LLSSVLQRGGV--IT---DEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      244
tr|F6Z911|F6Z911_HORSE             MLSSVLQGGSV--II---DEEETSKKMADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      244
tr|A0A5G2RB90|A0A5G2RB90_PIG       MLSSVLQGGSV--VM---DEEETSKKIADQLDNIVDMREYDVPYHIRLSIDLKIHVAHWY      217
tr|A0A420PL83|A0A420PL83_FUSOX      ----NGDFDLFDDSRD--DDRNMNTALSDASDFIVDIREYDVPYHVRVMIDLDIRAGKWY      222
tr|A0A1V6PGE7|A0A1V6PGE7_PENDC      MSSATASFDLFDDEII--NEQRPN-GNTQASDFIIDIREYDVPYHVRVCIDKDIRVGKWY      257
tr|A0A0G4NU82|A0A0G4NU82_PENCA      MTSATAGIDLFDDEII--NEQRPN-GNLQASDFIVDIREYDVPYHVRVCIDKDIRIGKWY      257
tr|A0A1V6TET5|A0A1V6TET5_9EURO      MTSATAGIDLFDDEII--NEQRPN-GNLQASDFIVDIREYDVPYHVRVCIDKDIRIGKWY      257
tr|A0A167WJ50|A0A167WJ50_PENCH      MTSATAGIDLFDDEII--NEQRPN-GNLQASDFIVDIREYDVPYHVRVCIDKDIRIGKWY      257
tr|A0A3A2Z9P9|A0A3A2Z9P9_9EURO      ISSANAGFDLFDDELI--NESRTN-NSMNASDFIVDIREYDVPYHVRVAIDKDIRIGKWY      257
tr|A0A401L0U3|A0A401L0U3_ASPAW      ISSANAGFDLFDDELI--NESRSN-ATMNASDFIIDIREYDVPYHVRVSIDKDIRIGKWY      257
tr|A0A317VH96|A0A317VH96_ASPEC      ISSANAGFDLFDDELI--NESRSN-ATMNASDFIIDIREYDVPYHVRVSIDK----GKWY      253
tr|A0A117E1M7|A0A117E1M7_ASPNG      ISSANAGFDLFDDELI--NESRSN-ATMNASDFIIDIREYDVPYHVRVSIDKDIRIGKWY      257
tr|A0A1D8PTD0|A0A1D8PTD0_CANAL      KEI---------------------ETIMDPSVYIDDIREYDVPYHVRVSIDRNVRVGKWY      263
tr|C5M3R9|C5M3R9_CANTT             KDV---------------------DSSSDPSTYIDDIREYDVPYHVRVSIDKQLRVGKWY      258
sp|Q6FNY7|DPOE_CANGA               SNISNGNGMLSSNVNR--FASSSVQDKKDAKQYIEDIREYDVPYHVRVSIDKDIRVGKWY      265
sp|P21951|DPOE_YEAST               AAN--------------------GSEKVDAKHLIEDIREYDVPYHVRVSIDKDIRVGKWY      205
tr|A0A7H9HLU1|A0A7H9HLU1_9SACH      QGA--------------------GNSFRDAKSLIEEVREYDVPYHVRVSIDKNIRVGKWY      255
sp|Q6CUS7|DPOE_KLULA               GSD---------------------YSTRDVKTLIEDIREYDVPYHVRVSIDKNIRVGKWY      182
sp|Q752B8|DPOE_ASHGO               NDT---------------------TNRRDAKMLIDDIREYDVPYHVRVCIDKEIRVGKWY      179
                                                             :    : ::*********:*. :*     ..**
```

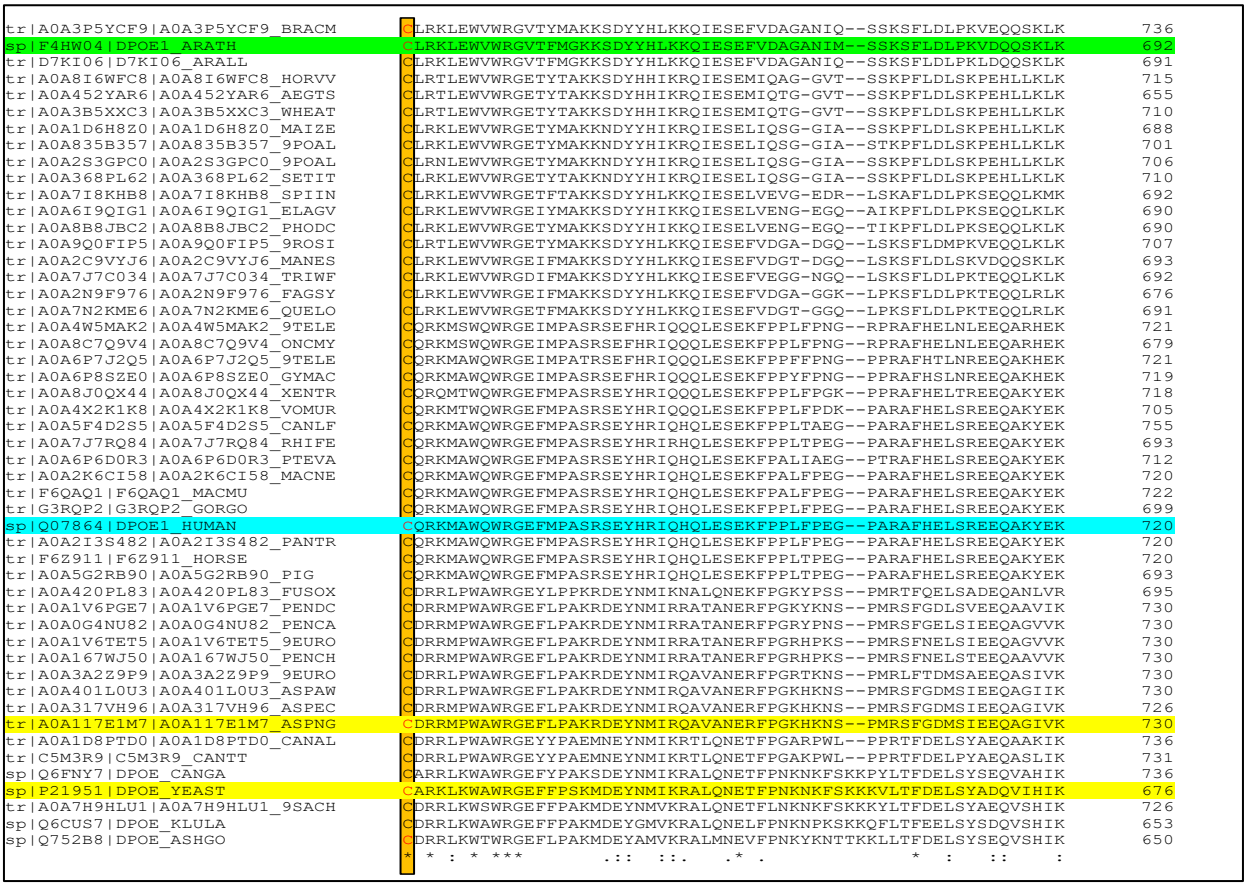

NTD
PR exo
X-ray

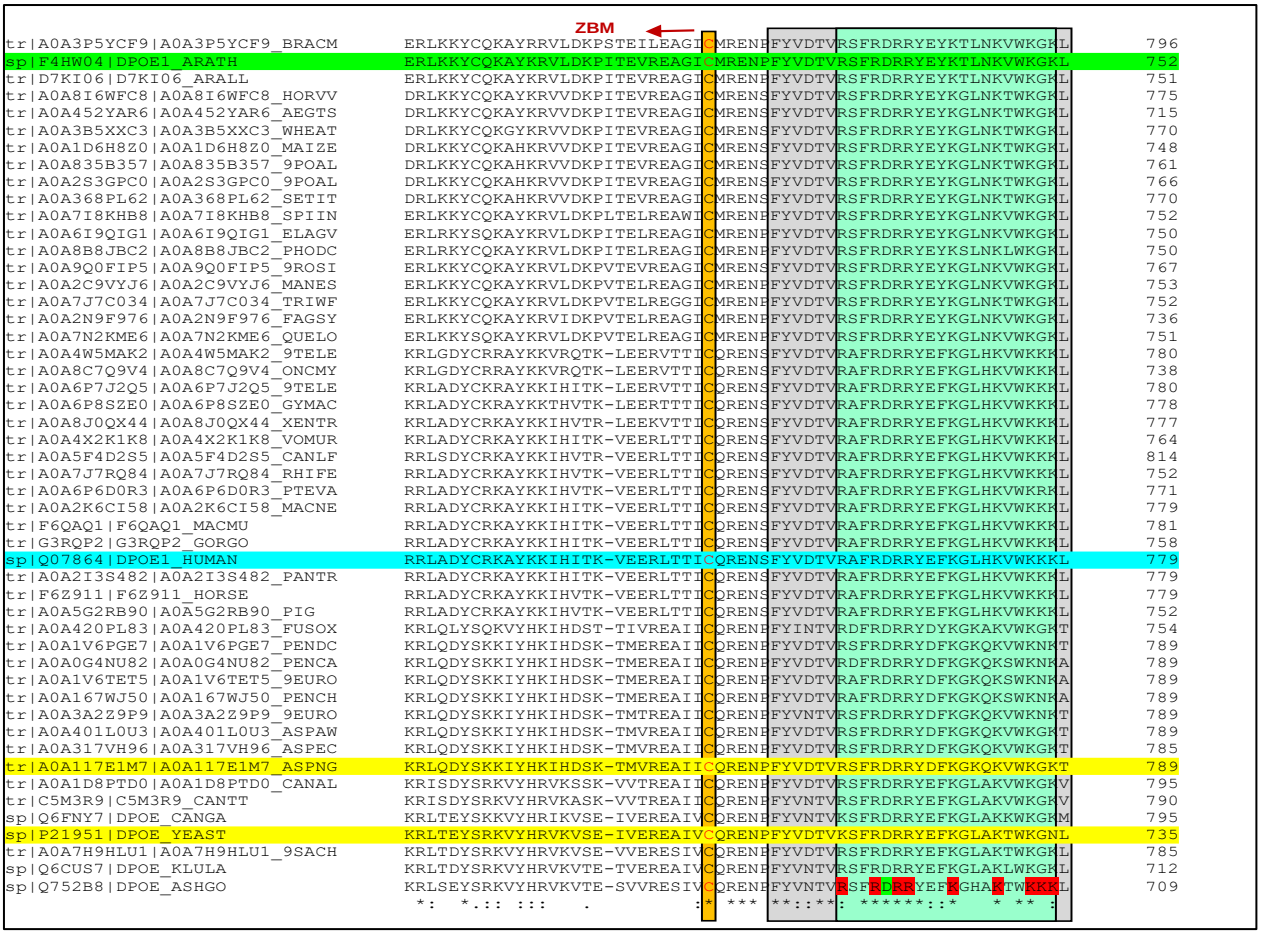

X-ray

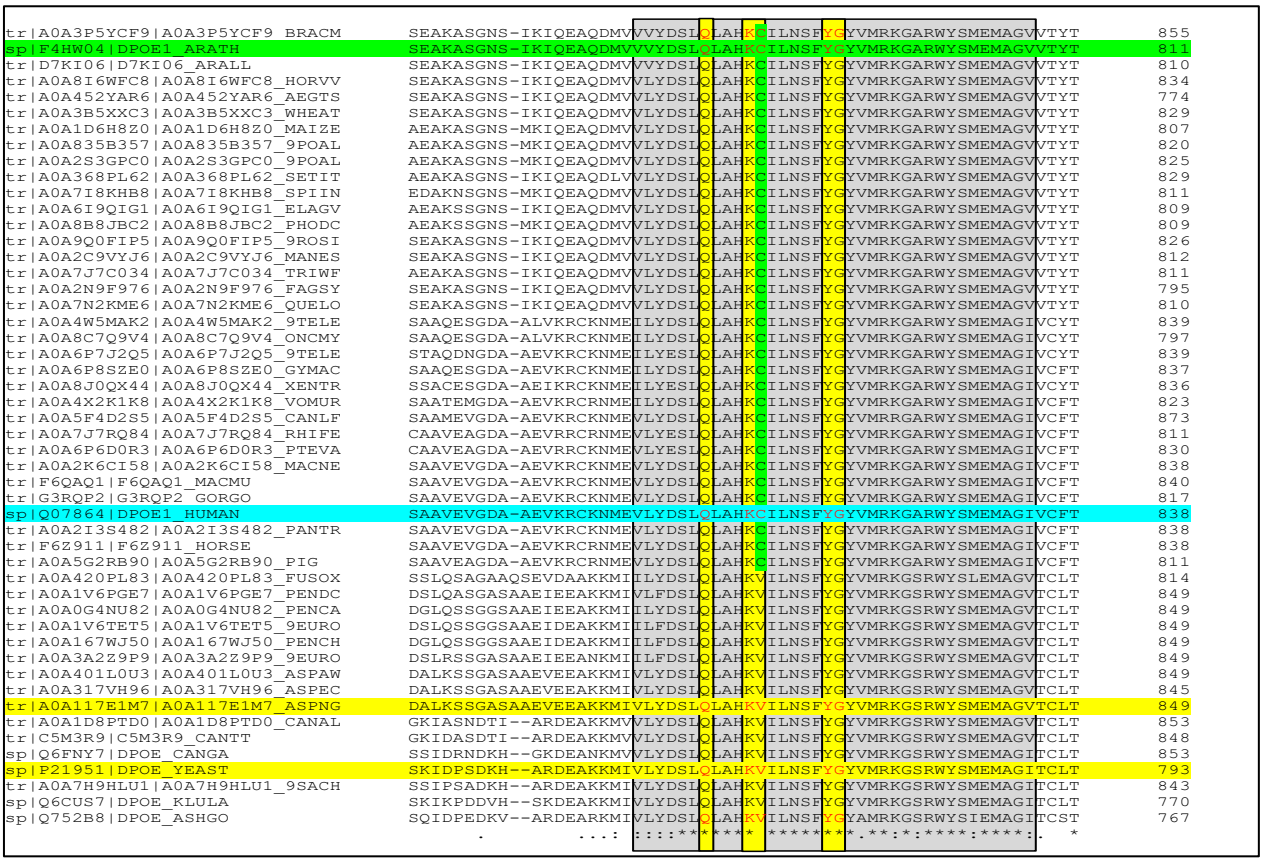


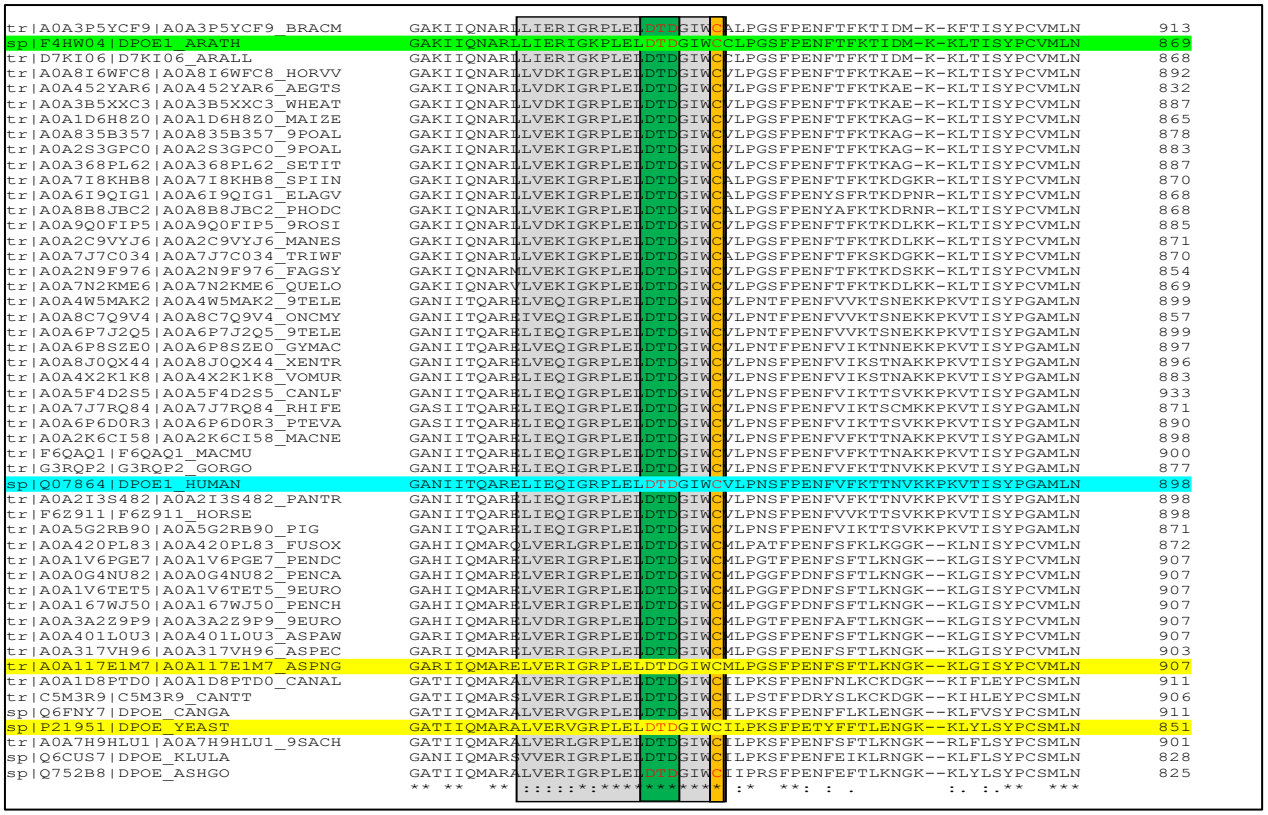

PR exo
Pol

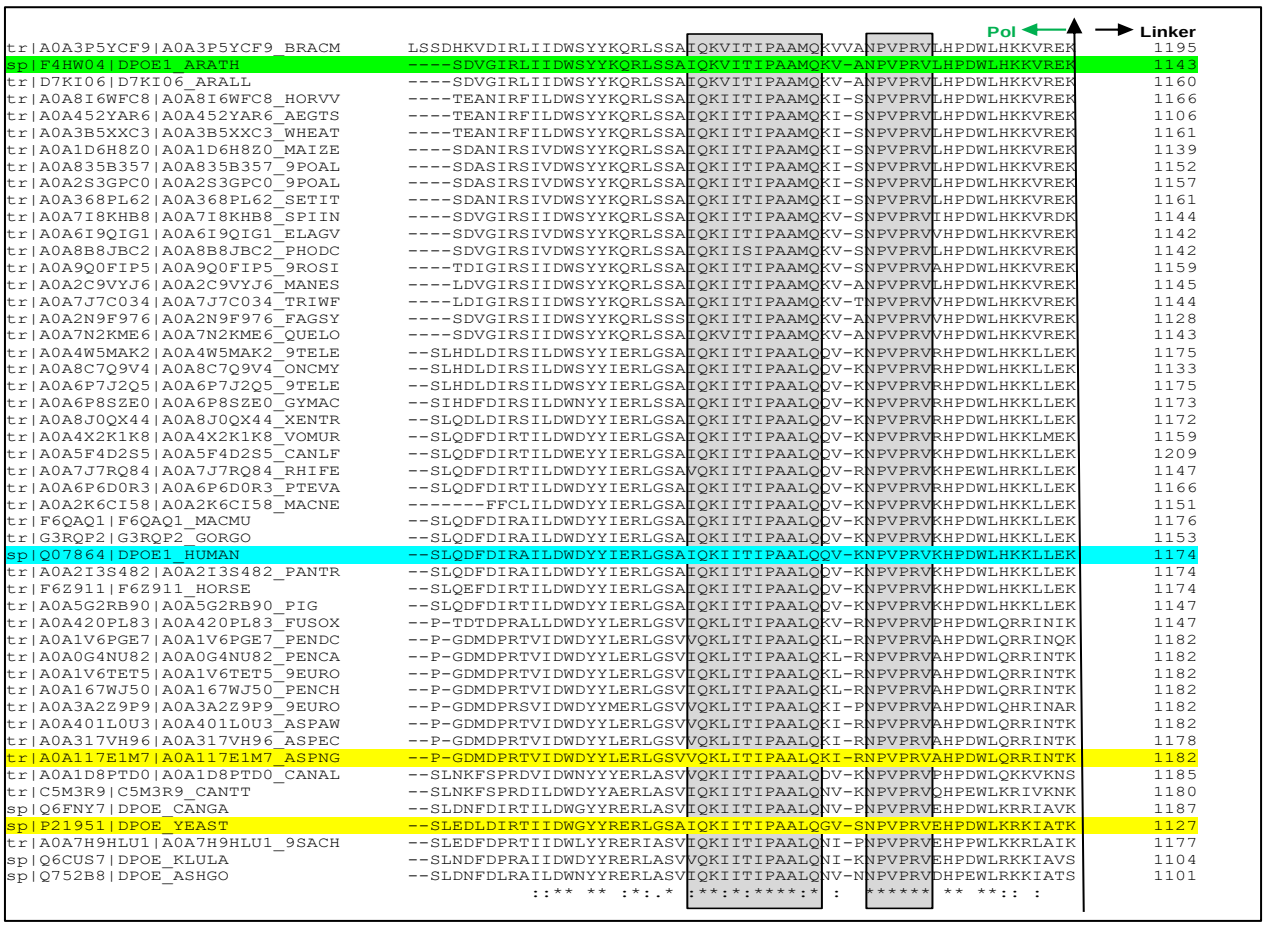

ZBM

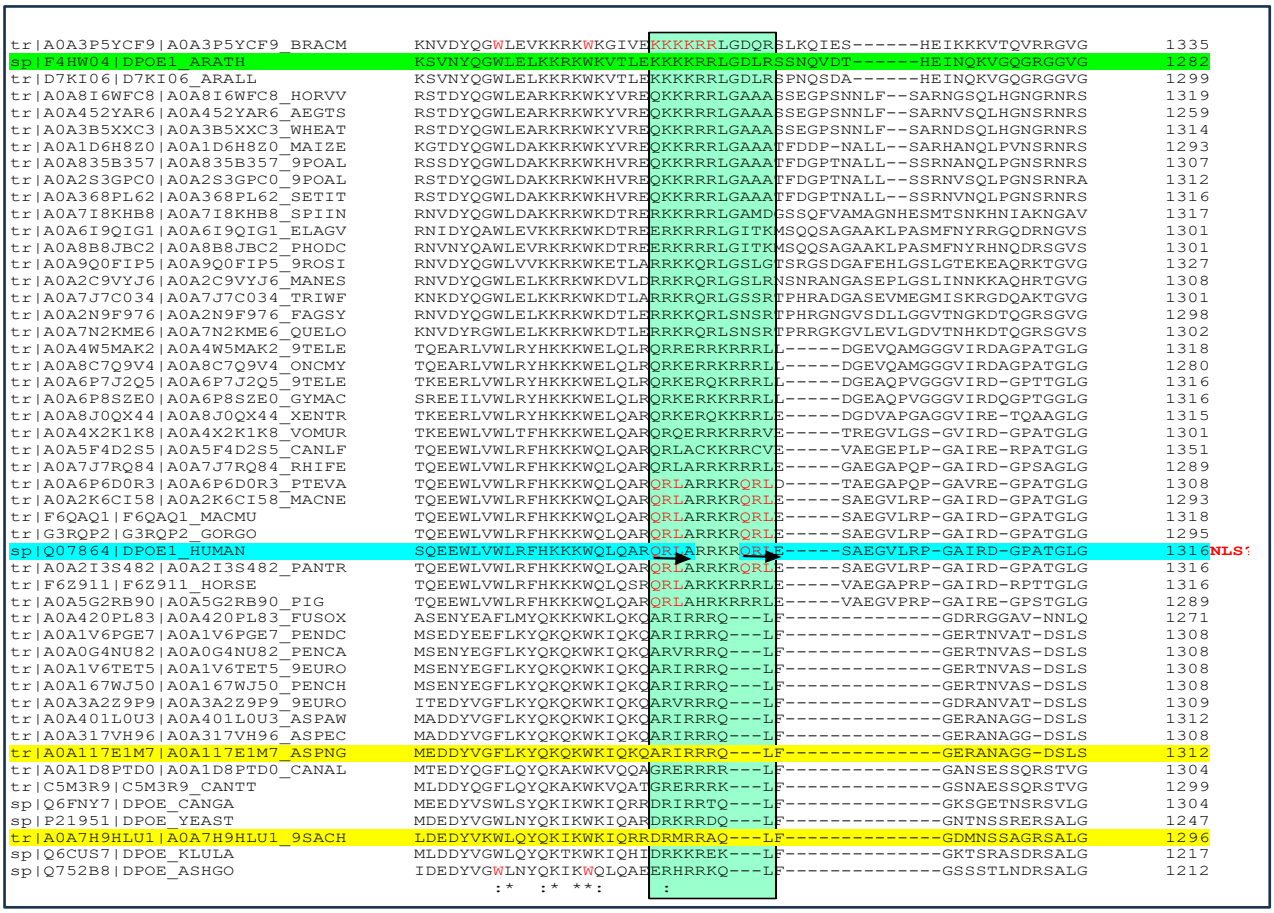


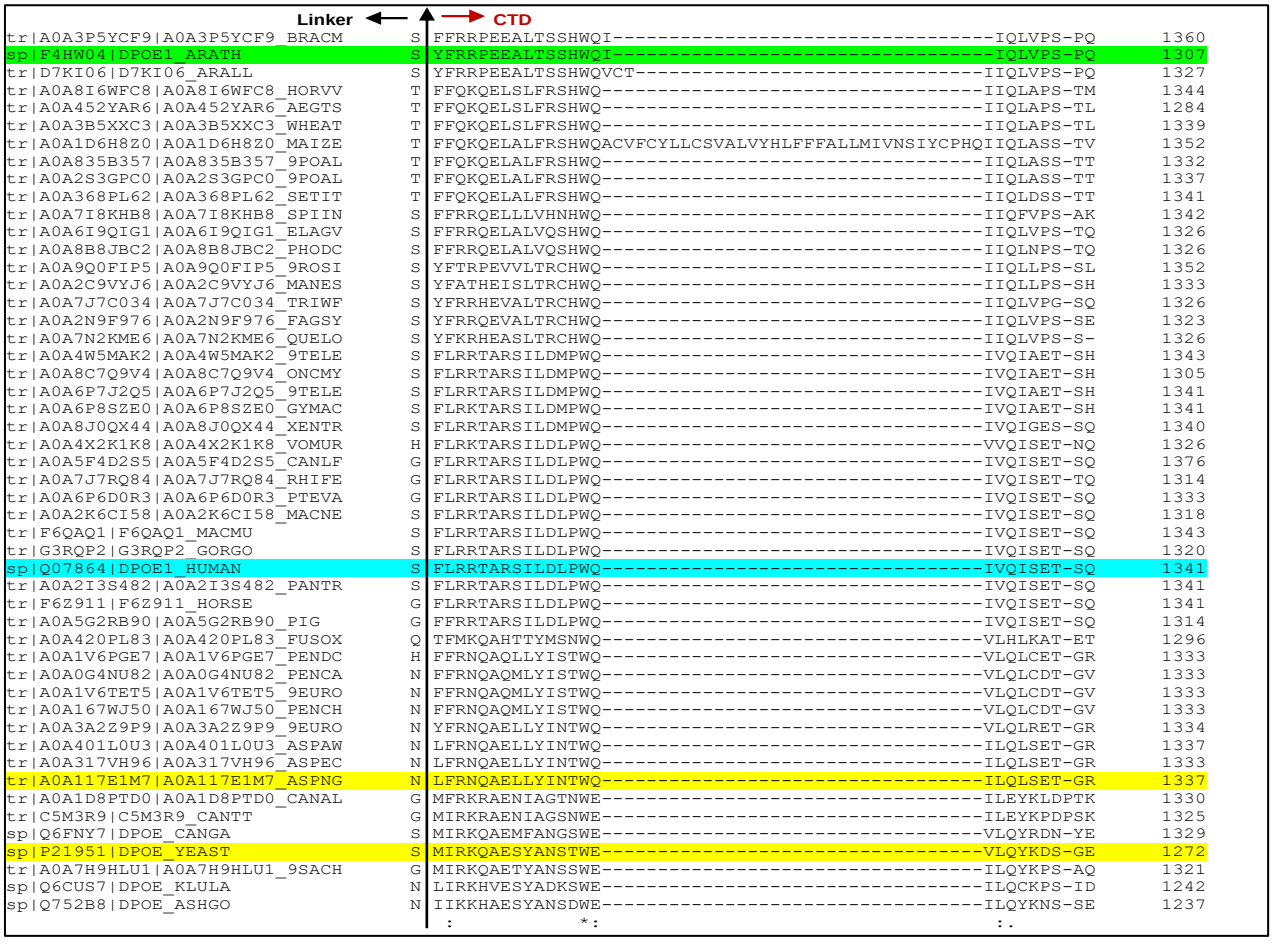

ZBM
tr|A0A3P5YCF9|A0A3P5YCF9_BRACM ERLKKYCQKAYRRVLDKPSTEILEAGICMRENPFYVDTVRSFRDRRYEYKTLNKVWKGKL 796
sp|F4HW04|DPOE1_ARATH ERLKKYCQKAYKRVLDKPITEVREAGICMRENPFYVDTVRSFRDRRYEYKTLNKVWKGKL 752
tr|D7KI06|D7KI06_ARALL ERLKKYCQKAYKRVLDKPITEVREAGICMRENPFYVDTVRSFRDRRYEYKTLNKVWKGKL 751
tr|A0A8I6WFC8|A0A8I6WFC8_HORVV DRLKKYCQKAYKRVVDKPITEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKTWKGKL 775
tr|A0A452YAR6|A0A452YAR6_AEGTS DRLKKYCQKAYKRVVDKPITEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKTWKGKL 715
tr|A0A3B5XXC3|A0A3B5XXC3_WHEAT DRLKKYCQKGYKRVVDKPITEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKTWKGKL 770
tr|A0A1D6H8Z0|A0A1D6H8Z0_MAIZE DRLKKYCQKAHKRVVDKPITEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKTWKGKL 748
tr|A0A835B357|A0A835B357_9POAL DRLKKYCQKAHKRVVDKPITEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKTWKGKL 761
tr|A0A2S3GPC0|A0A2S3GPC0_9POAL DRLKKYCQKAHKRVVDKPITEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKTWKGKL 766
tr|A0A368PL62|A0A368PL62_SETIT DRLKKYCQKAHKRVVDKPITEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKTWKGKL 770
tr|A0A7I8KHB8|A0A7I8KHB8_SPIIN ERLKKYCQKAYKRVLDKPLTELREAWICMRENPFYVDTVRSFRDRRYEYKGLNKVWKGKL 752
tr|A0A6I9QIG1|A0A6I9QIG1_ELAGV ERLRKYSQKAYKRVLDKPITELREAGICMRENPFYVDTVRSFRDRRYEYKGLNKVWKGKL 750
tr|A0A8B8JBC2|A0A8B8JBC2_PHODC ERLRKYCQKAYKRVLDKPITELREAGICMRENPFYVDTVRSFRDRRYEYKSLNKLWKGKL 750
tr|A0A9Q0FIP5|A0A9Q0FIP5_9ROSI ERLKKYCQKAYKRVLDKPVTEVREAGICMRENSFYVDTVRSFRDRRYEYKGLNKVWKGKL 767
tr|A0A2C9VYJ6|A0A2C9VYJ6_MANES ERLKKYCQKAYKRVLDKPVTELREAGICMRENPFYVDTVRSFRDRRYEYKGLNKVWKGKL 753
tr|A0A7J7C034|A0A7J7C034_TRIWF ERLKKYCQKAYKRVLDKPVTELREGGICMRENPFYVDTVRSFRDRRYEYKGLNKTWKGKL 752
tr|A0A2N9F976|A0A2N9F976_FAGSY ERLKKYCQKAYKRVIDKPVTELREAGICMRENPFYVDTVRSFRDRRYEYKGLNKVWKGKL 736
tr|A0A7N2KME6|A0A7N2KME6_QUELO ERLKKYSQKAYKRVLDKPVTELREAGICMRENPFYVDTVRSFRDRRYEYKGLNKVWKGKL 751
tr|A0A4W5MAK2|A0A4W5MAK2_9TELE KRLGDYCRRAYKKVRQTK-LEERVTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 780
tr|A0A8C7Q9V4|A0A8C7Q9V4_ONCMY KRLGDYCRRAYKKVRQTK-LEERVTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 738
tr|A0A6P7J2Q5|A0A6P7J2Q5_9TELE KRLADYCRRAYKKIHITK-LEERVTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 780
tr|A0A6P8SZE0|A0A6P8SZE0_GYMAC KRLADYCRRAYKKTHVTK-LEERTTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 778
tr|A0A8J0QX44|A0A8J0QX44_XENTR KRLADYCRKAYKKIHVTR-LEEKVTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 777
tr|A0A4X2K1K8|A0A4X2K1K8_VOMUR KRLADYCRKAYKKIHITK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 764
tr|A0A5F4D2S5|A0A5F4D2S5_CANLF RRLSDYCRKAYKKIHVTR-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 814
tr|A0A7J7RQ84|A0A7J7RQ84_RHIFE RRLADYCRKAYKKIHVTK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 752
tr|A0A6P6D0R3|A0A6P6D0R3_PTEVA RRLADYCRKAYKKIHVTK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 771
tr|A0A2K6CI58|A0A2K6CI58_MACNE RRLADYCRKAYKKIHITK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 779
tr|F6QAQ1|F6QAQ1_MACMU RRLADYCRKAYKKIHITK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 781
tr|G3RQP2|G3RQP2_GORGO RRLADYCRKAYKKIHITK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 758
sp|Q07864|DPOE1_HUMAN RRLADYCRKAYKKIHITK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 779
tr|A0A2I3S482|A0A2I3S482_PANTR RRLADYCRKAYKKIHITK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 779
tr|F6Z911|F6Z911_HORSE RRLADYCRKAYKKIHVTK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 779
tr|A0A5G2RB90|A0A5G2RB90_PIG RRLADYCRKAYKKIHVTK-VEERLTTICQRENSFYVDTVRAFRDRRYEFKGLHKVWKKKL 752
tr|A0A420PL83|A0A420PL83_FUSOX KRLQLYSQKVYHKIHDST-TIVREAIICQRENPFYINTVRDFRDRRYDYKGKAKVWKGKT 754
tr|A0A1V6PGE7|A0A1V6PGE7_PENDC KRLQDYSKKIYHKIHDSK-TMREAIICQRENPFYVDTVRSFRDRRYDFKGKQKVWKNKT 789
tr|A0A0G4NU82|A0A0G4NU82_PENCA KRLQDYSKKIYHKIHDSK-TMREAIICQRENPFYVDTVRDFRDRRYDFKGKQKSWKNKA 789
tr|A0A1V6TET5|A0A1V6TET5_9EURO KRLQDYSKKIYHKIHDSK-TMREAIICQRENPFYVDTVRAFRDRRYDFKGKQKSWKNKA 789
tr|A0A167WJ50|A0A167WJ50_PENCH KRLQDYSKKIYHKIHDSK-TMREAIICQRENPFYVDTVRAFRDRRYDFKGKQKSWKNKA 789
tr|A0A3A2Z9P9|A0A3A2Z9P9_9EURO KRLQDYSKKIYHKIHDSK-TMTREAIICQRENPFYVNTVRSFRDRRYDFKGKQKVWKNKT 789
tr|A0A401L0U3|A0A401L0U3_ASPAW KRLQDYSKKIYHKIHDSK-TMVREAIICQRENPFYVDTVRSFRDRRYDFKGKQKVWKGKT 789
tr|A0A317VH96|A0A317VH96_ASPEC KRLQDYSKKIYHKIHDSK-TMVREAIICQRENPFYVDTVRSFRDRRYDFKGKQKVWKGKT 785
tr|A0A117E1M7|A0A117E1M7_ASPNG KRLQDYSKKIYHKIHDSK-TMVREAIICQRENPFYVDTVRSFRDRRYDFKGKQKVWKGKT 789
tr|A0A1D8PTD0|A0A1D8PTD0_CANAL KRISDYSRKVYHRVKSSK-VVTREAIICQRENPFYVDTVRSFRDRRYEFKGLAKVWKGKV 795
tr|C5M3R9|C5M3R9_CANTT KRISDYSRKVYHRVKASK-VVTREAIICQRENPFYVNTVRSFRDRRYEFKGLAKVWKGKV 790
sp|Q6FNY7|DPOE_CANGA KRLTEYSKKVYHRIKVSE-IVEREAIVCQRENPFYVNTVKSFRDRRYEFKGLAKKWKGKM 795
sp|P21951|DPOE_YEAST KRLTEYSRKVYHRVKVSE-IVEREAIVCQRENPFYVDTVKSFRDRRYEFKGLAKTWKGNL 735
tr|A0A7H9HLU1|A0A7H9HLU1_9SACH KRLTDYSRKVYHRVKVSE-VVERESIVCQRENPFYVDTVRSFRDRRYEFKGLAKTWKGKL 785
sp|Q6CUS7|DPOE_KLULA KRLTDYSRKVYHRVKVTE-TVEREAIVCQRENPFYVNTVRSFRDRRYEFKGLAKLWKGKL 712
sp|Q752B8|DPOE_ASHGO KRLSEYSRKVYHRVKVTE-SVVRESIVCQRENPFYVNTVKSFKDRRYEFKGHAKTWKKKL 709

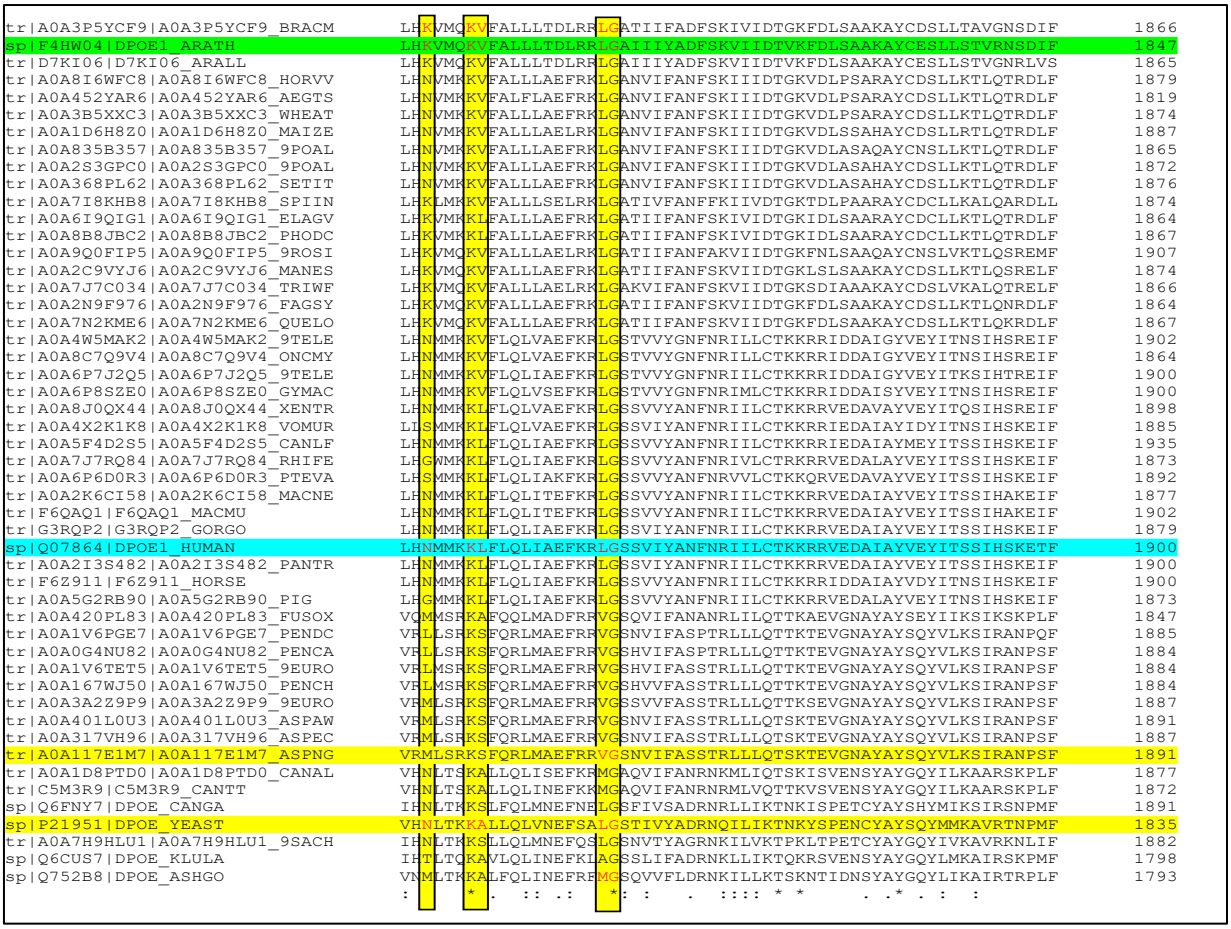

tr|A0A3P5YCF9|A0A3P5YCF9_BRACM SEAKASGNS-IKIQEAQDMVVYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 855
sp|F4HW04|DPOE1_ARATH SEAKASGNS-IKIQEAQDMVVYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 811
tr|D7KI06|D7KI06_ARALL SEAKASGNS-IKIQEAQDMVVYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 810
tr|A0A8I6WFC8|A0A8I6WFC8_HORVV SEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 834
tr|A0A452YAR6|A0A452YAR6_AEGTS SEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 774
tr|A0A3B5XXC3|A0A3B5XXC3_WHEAT SEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 829
tr|A0A1D6H8Z0|A0A1D6H8Z0_MAIZE AEAKASGNS-MKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 807
tr|A0A835B357|A0A835B357_9POAL AEAKASGNS-MKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 820
tr|A0A2S3GPC0|A0A2S3GPC0_9POAL AEAKASGNS-MKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 825
tr|A0A368PL62|A0A368PL62_SETIT AEAKASGNS-IKIQEAQDLVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 829
tr|A0A7I8KHB8|A0A7I8KHB8_SPIIN EDAKNSGNS-MKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 811
tr|A0A6I9QIG1|A0A6I9QIG1_ELAGV AEAKSSGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 809
tr|A0A8B8JBC2|A0A8B8JBC2_PHODC AEAKSSGNS-MKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 809
tr|A0A9Q0FIP5|A0A9Q0FIP5_9ROSI SEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 826
tr|A0A2C9VYJ6|A0A2C9VYJ6_MANES SEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 812
tr|A0A7J7C034|A0A7J7C034_TRIWF AEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 811
tr|A0A2N9F976|A0A2N9F976_FAGSY SEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 795
tr|A0A7N2KME6|A0A7N2KME6_QUELO SEAKASGNS-IKIQEAQDMVVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGVVTYT 810
tr|A0A4W5MAK2|A0A4W5MAK2_9TELE SAAQESGDA-ALVKRCKNMEILYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCYT 839
tr|A0A8C7Q9V4|A0A8C7Q9V4_ONCMY SAAQESGDA-ALVKRCKNMEILYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCYT 797
tr|A0A6P7J2Q5|A0A6P7J2Q5_9TELE STAQDNGDA-AEVKRCKNMEILYESLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCYT 839
tr|A0A6P8SZE0|A0A6P8SZE0_GYMAC SAAQESGDA-AEVKRCKNMEILYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 837
tr|A0A8J0QX44|A0A8J0QX44_XENTR SSACESGDA-AEIKRCKNMEILYESLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCYT 836
tr|A0A4X2K1K8|A0A4X2K1K8_VOMUR SAATEMGDA-AEVKRCRNMEILYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 823
tr|A0A5F4D2S5|A0A5F4D2S5_CANLF SAAMEVGDA-AEVKRCKNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 873
tr|A0A7J7RQ84|A0A7J7RQ84_RHIFE CAAVEAGDA-AEVRRCRNMEVLYESLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 811
tr|A0A6P6D0R3|A0A6P6D0R3_PTEVA CAAVEAGDA-AEVRRCKNMEVLYESLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 830
tr|A0A2K6CI58|A0A2K6CI58_MACNE SAAVEVGDA-AEVKRCKNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 838
tr|F6QAQ1|F6QAQ1_MACMU SAAVEVGDA-AEVKRCKNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 840
tr|G3RQP2|G3RQP2_GORGO SAAVEVGDA-AEVKRCKNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 817
sp|Q07864|DPOE1_HUMAN SAAVEVGDA-AEVKRCKNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 838
tr|A0A2I3S482|A0A2I3S482_PANTR SAAVEVGDA-AEVKRCKNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 838
tr|F6Z911|F6Z911_HORSE SAAVEVGDA-AEVKRCRNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 838
tr|A0A5G2RB90|A0A5G2RB90_PIG SAAVEAGDA-AEVKRCRNMEVLYDSLQLAHKCILNSFYGYVMRKGARWYSMEMAGIVCFT 811
tr|A0A420PL83|A0A420PL83_FUSOX SSLQSAGAAQSEVDAAKKMILYDSLQLAHKVILNSFYGYVMRKGSRWYSLEMAGVTCLT 814
tr|A0A1V6PGE7|A0A1V6PGE7_PENDC DSLQASGASAAEIEEAKKMILVLFDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 849
tr|A0A0G4NU82|A0A0G4NU82_PENCA DGLQSSGGSAAEIEEAKKMILYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 849
tr|A0A1V6TET5|A0A1V6TET5_9EURO DSLQSSGGSAAEIDEAKKMILFDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 849
tr|A0A167WJ50|A0A167WJ50_PENCH DGLQSSGGSAAEIDEAKKMILFDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 849
tr|A0A3A2Z9P9|A0A3A2Z9P9_9EURO DSLRSSGASAAEIEEANKMILFDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 849
tr|A0A401L0U3|A0A401L0U3_ASPAW DALKSSGASAAEVEEAKKMIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 849
tr|A0A317VH96|A0A317VH96_ASPEC DALKSSGASAAEVEEAKKMIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 845
tr|A0A117E1M7|A0A117E1M7_ASPNG DALKSSGASAAEVEEAKKMIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 849
tr|A0A1D8PTD0|A0A1D8PTD0_CANAL GKIASNDTI--ARDEAKKMVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 853
tr|C5M3R9|C5M3R9_CANTT GKIDASDTI--ARDEAKKMVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGVTCLT 848
sp|Q6FNY7|DPOE_CANGA SSIDRNDKH--GKDEANKMVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLT 853
sp|P21951|DPOE_YEAST SKIDPSDKH--ARDEAKKMIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLT 793
tr|A0A7H9HLU1|A0A7H9HLU1_9SACH SSIPSADKH--ARDEAKKMIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLT 843
sp|Q6CUS7|DPOE_KLULA SKIKPDDVH--SKDEAKKMIVLYDSLQLAHKVILNSFYGYVMRKGSRWYSMEMAGITCLT 770
sp|Q752B8|DPOE_ASHGO SQIDPEDKV--ARDEARKMIVLYDSLQLAHKVILNSFYGYAMRKGSRWYSIEMAGITCST 767

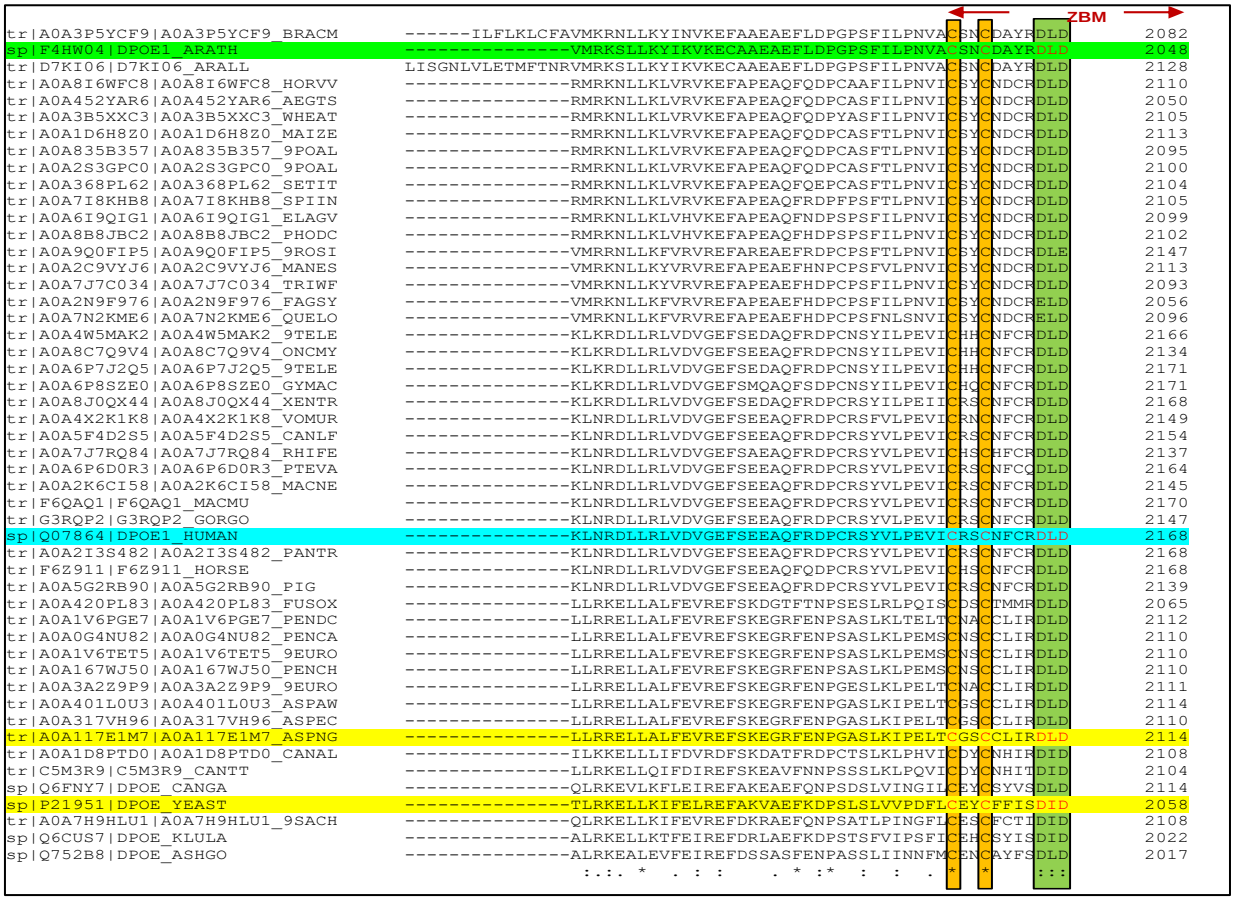


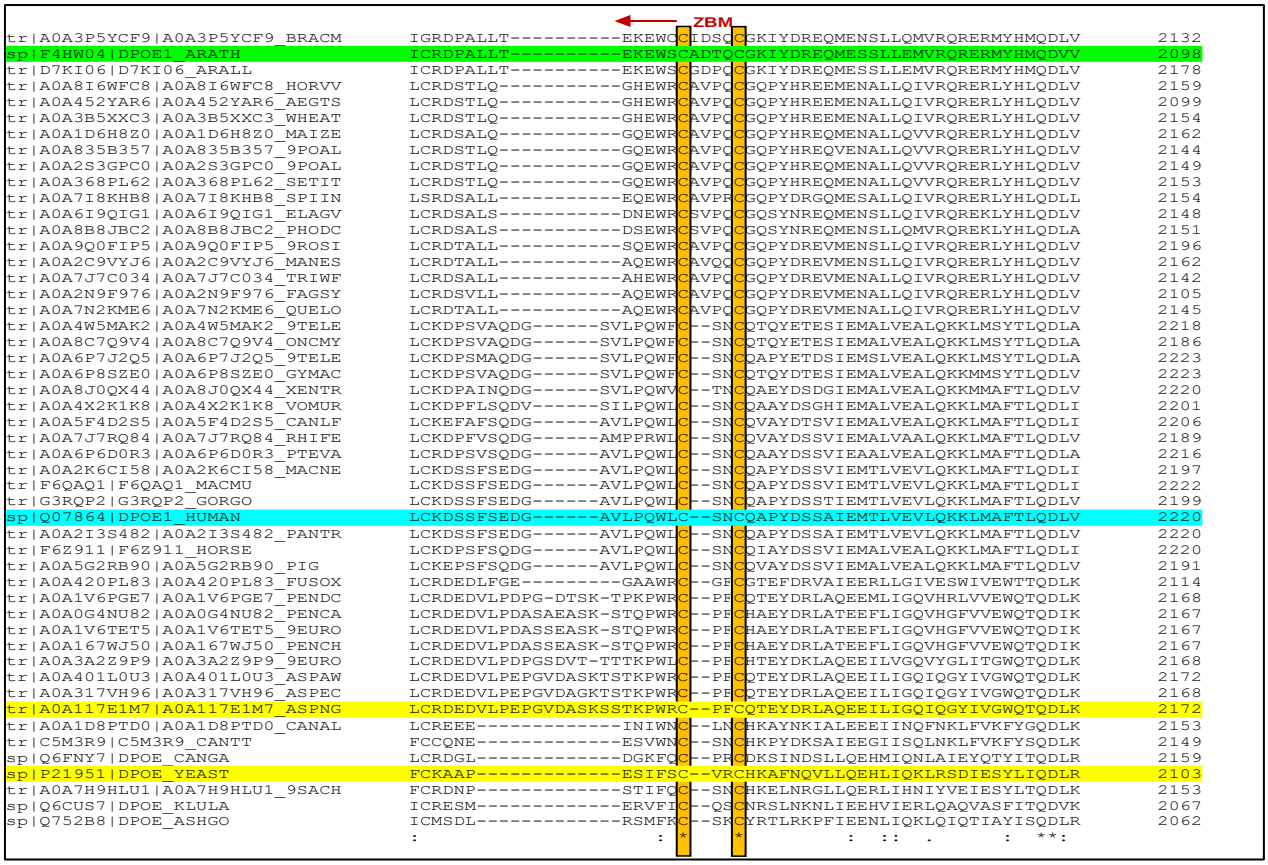

Pol
Linker

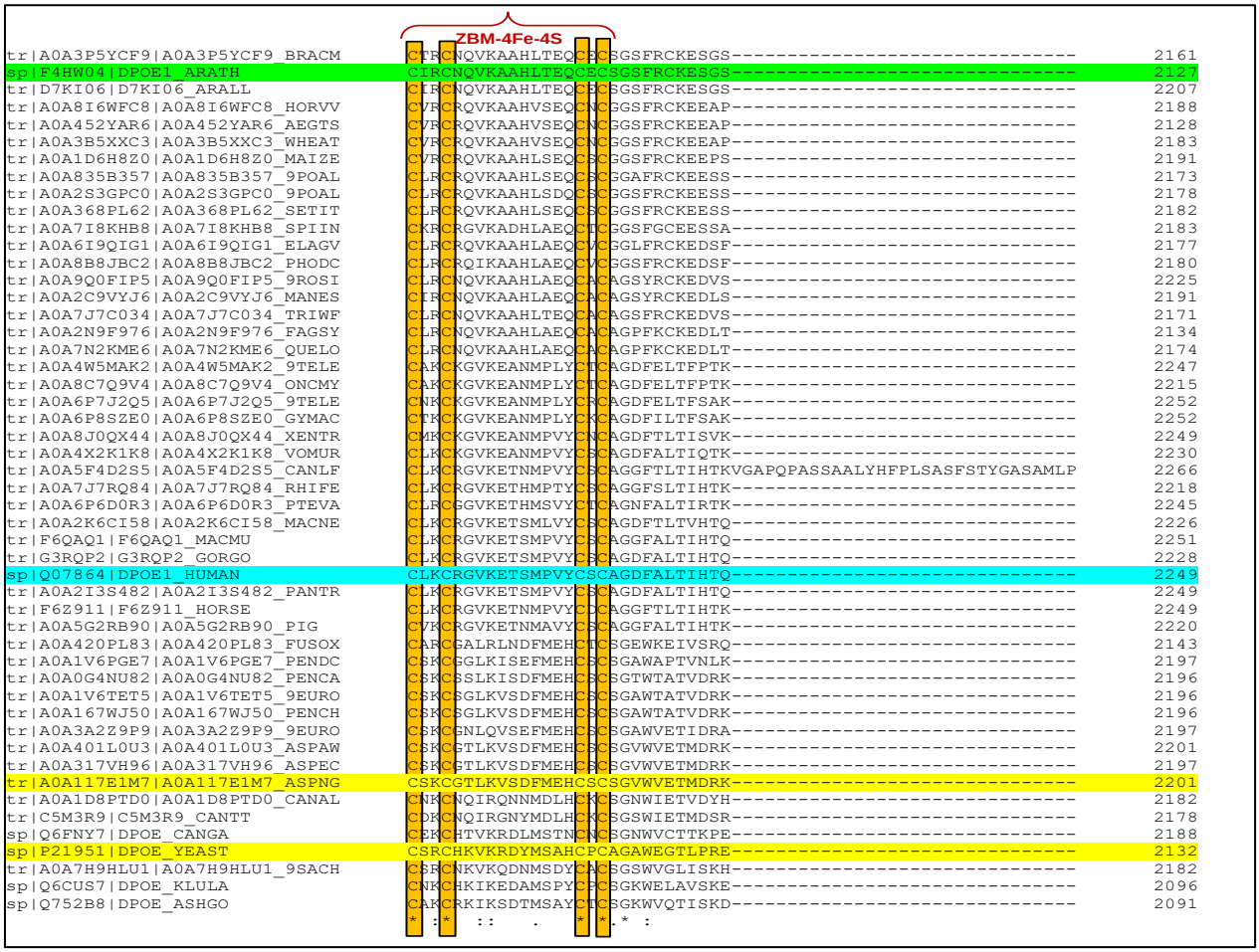

tr|A0A3P5YCF9|A0A3P5YCF9_BRACM KNVDYQGWLEVKKRKWKGIVEKKKKRRLGDQRSLKQIES------HEIKKKVTQVRRGVG 1335
sp|F4HW04|DPOE1_ARATH KSVNYQGWLELKKRKWKVTLEKKKKRRLGDLRSSNQVDT------HEINQKVGQGRGGVG 1282
tr|D7KI06|D7KI06_ARALL KSVNYQGWLELKKRKWKVTLEKKKKRRLGDLRSPNQSDA------HEINQKVGQGRGGVG 1299
tr|A0A8I6WFC8|A0A8I6WFC8_HORVV RSTDYQGWLEARKRKWKYVREQKKRRRLGAAASSEGPSNNLF--SARNGSQLHGNGRNRS 1319
tr|A0A452YAR6|A0A452YAR6_AEGTS RSTDYQGWLEARKRKWKYVREQKKRRRLGAAASSEGPSNNLF--SARNVSQLHGNSRNRS 1259
tr|A0A3B5XXC3|A0A3B5XXC3_WHEAT RSTDYQGWLEARKRKWKYVREQKKRRRLGAAASSEGPSNNLF--SARNDSQLHGNGRNRS 1314
tr|A0A1D6H8Z0|A0A1D6H8Z0_MAIZE KGTDYQGWLDAKKRKWKYVREQKKRRRLGAAATFDDP-NALL--SARHANQLPVNSRNRS 1293
tr|A0A835B357|A0A835B357_9POAL RSSDYQGWLDAKKRKWKHVREQKKRRRLGAAATFDGPTNALL--SSRNANQLPGNSRNRS 1307
tr|A0A2S3GPC0|A0A2S3GPC0_9POAL RSTDYQGWLDAKKRKWKHVREQKKRRRLGAAATFDGPTNALL--SSRNVSQLPGNSRNRA 1312
tr|A0A368PL62|A0A368PL62_SETIT RSTDYQGWLDAKKRKWKHVREQKKRRRLGAAATFDGPTNALL--SSRNVNQLPGNSRNRS 1316
tr|A0A7I8KHB8|A0A7I8KHB8_SPIIN RNVDYQGWLDAKKRKWKDTREKKKRRRLGAMDGSSQFVAMAGNHESMTSNKHNIAKNGAV 1317
tr|A0A6I9QIG1|A0A6I9QIG1_ELAGV RNIDYQAWLEVKKRKWKDTREERKRRRLGITKMSQQSAGAAKLPASMFNYRRGQDRNGVS 1301
tr|A0A8B8JBC2|A0A8B8JBC2_PHODC RNVNYQAWLEVRKRKWKDTREERKRRRLGITKMSQQSAGAAKLPASMFNYRHNQDRSGVS 1301
tr|A0A9Q0FIP5|A0A9Q0FIP5_9ROSI RNVDYQGWLVVKKRKWKETLARRKKQRLGSLGTSRGSDGAFEHLGSLGTEKEAQRKTGVG 1327
tr|A0A2C9VYJ6|A0A2C9VYJ6_MANES RNVDYQGWLELKKRKWKDVLDRRKRQRLGSLRNSNRANGASEPLGSLINNKKAQHRTGVG 1308
tr|A0A7J7C034|A0A7J7C034_TRIWF KNKDYQGWLELKKRKWKDTLARRKRQRLGSSRTPHRADGASEVMEGMISKRGDQAKTGVG 1301
tr|A0A2N9F976|A0A2N9F976_FAGSY RNVDYQGWLELKKRKWKDTLERRKKQRLSNSRTPHRGNGVSDLLGGVTNGKDTQGRSGVG 1298
tr|A0A7N2KME6|A0A7N2KME6_QUELO KNVDYRGWLELKKRKWKDTLERRKRQRLSNSRTPRRGKGVLEVLGDVTNHKDTQGRSGVS 1302
tr|A0A4W5MAK2|A0A4W5MAK2_9TELE TQEARLVWLRYHKKKWELQLRQRERRKRRRLL-----DGEVQAMGGGVIRDAGPATGLG 1318
tr|A0A8C7Q9V4|A0A8C7Q9V4_ONCMY TQEARLVWLRYHKKKWELQLRQRKERRKRRRLL-----DGEVQAMGGGVIRDAGPATGLG 1280
tr|A0A6P7J2Q5|A0A6P7J2Q5_9TELE TKEERLVWLRYHKKKWELQLRQRKERQKRRRLL-----DGEAQPVGGGVIRD-GPTTGLG 1316
tr|A0A6P8SZE0|A0A6P8SZE0_GYMAC SREEILVWLRYHKKKWELQLRQRKERKKRRRLL-----DGEAQPVGGGVIRDQGPTGGLG 1316
tr|A0A8J0QX44|A0A8J0QX44_XENTR TKEERLVWLRYHKKKWELQARQRKERQKRRLE-----DGDVAPGAGGVIRE-TQAAGLG 1315
tr|A0A4X2K1K8|A0A4X2K1K8_VOMUR TKEEWLVWLTFHKKKWELQARQRQERRKRRRVE-----TREGVLGS-GVIRD-GPATGLG 1301
tr|A0A5F4D2S5|A0A5F4D2S5_CANLF TQEEWLVWLRFHKKKWQLQARQRLACKRRCVE-----VAEGEPLP-GAIRE-RPATGLG 1351
tr|A0A7J7RQ84|A0A7J7RQ84_RHIFE TQEEWLVWLRFHKKKWQLQARQRLARRKRRRLE-----GAEGAPQP-GAIRD-GPSAGLG 1289
tr|A0A6P6D0R3|A0A6P6D0R3_PTEVA TQEEWLVWLRFHKKKWQLQARQRLARRKRQRLD-----TAEGAPQP-GAVRE-GPATGLG 1308
tr|A0A2K6CI58|A0A2K6CI58_MACNE TQEEWLVWLRFHKKKWQLQARQRLARRKRQRLE-----SAEGVLRP-GAIRD-GPATGLG 1293
tr|F6QAQ1|F6QAQ1_MACMU TQEEWLVWLRFHKKKWQLQARQRLARRKRQRLE-----SAEGVLRP-GAIRD-GPATGLG 1318
tr|G3RQP2|G3RQP2_GORGO TQEEWLVWLRFHKKKWQLQARQRLARRKRQRLE-----SAEGVLRP-GAIRD-GPATGLG 1295
sp|Q07864|DPOE1_HUMAN SQEEWLVWLRFHKKKWQLQARQRLARRKRQRLE-----SAEGVLRP-GAIRD-GPATGLG 1316
tr|A0A2I3S482|A0A2I3S482_PANTR TQEEWLVWLRFHKKKWQLQARQRLARRKRQRLE-----SAEGVLRP-GAIRD-GPATGLG 1316
tr|F6Z911|F6Z911_HORSE TQEEWLVWLRFHKKKWQLQSRQRLARKKRRRLE-----VAEGAPRP-GAIRE-RPTTGLG 1316
tr|A0A5G2RB90|A0A5G2RB90_PIG TQEEWLVWLRFHKKKWQLQARQRLAHRKRRRLE-----VAEGVPRP-GAIRE-GPSTGLG 1289
tr|A0A420PL83|A0A420PL83_FUSOX ASENYEAFLMYQKKKWKLQKQARIRRRQ---LF----------GDRRGGAV-NNLQ 1271
tr|A0A1V6PGE7|A0A1V6PGE7_PENDC MSEDYEEFLKYQKQKWKIQKQARIRRRQ---LF----------GERTNVAT-DSLS 1308
tr|A0A0G4NU82|A0A0G4NU82_PENCA MSENYEGFLKYQKQKWKIQKQARVRRRQ---LF----------GERTNVAS-DSLS 1308
tr|A0A1V6TET5|A0A1V6TET5_9EURO MSENYEGFLKYQKQKWKIQKQARIRRRQ---LF----------GERTNVAS-DSLS 1308
tr|A0A167WJ50|A0A167WJ50_PENCH MSENYEGFLKYQKQKWKIQKQARIRRRQ---LF----------GERTNVAS-DSLS 1308
tr|A0A3A2Z9P9|A0A3A2Z9P9_9EURO ITEDYVGFLKYQKQKWKIQKQARVRRRQ---LF----------GDRANVAT-DSLS 1309
tr|A0A401L0U3|A0A401L0U3_ASPAW MADDYVGFLKYQKQKWKIQKQARIRRRQ---LF----------GERANAGG-DSLS 1312
tr|A0A317VH96|A0A317VH96_ASPEC MADDYVGFLKYQKQKWKIQKQARIRRRQ---LF----------GERANAGG-DSLS 1308
tr|A0A117E1M7|A0A117E1M7_ASPNG MEDDYVGFLKYQKQKWKIQKQARIRRRQ---LF----------GERANAGG-DSLS 1312
tr|A0A1D8PTD0|A0A1D8PTD0_CANAL MTEDYQGFLQYQKAKWKVQQAGRERRRR---LF----------GANSESSQRSTVG 1304
tr|C5M3R9|C5M3R9_CANTT MLDDYQGFLQYQKAKWKVQATGRERRRK---LF----------GSNAESSQRSTVG 1299
sp|Q6FNY7|DPOE_CANGA MEEDYVSWLSYQKIKWKIQRRDRIRRTQ---LF----------GKSGETNSRSVLG 1304
sp|P21951|DPOE_YEAST MDEDYVGWLNYQKIKWKIQARDRKRRDQ---LF----------GNTNSSRERSALG 1247
tr|A0A7H9HLU1|A0A7H9HLU1_9SACH LDEDYVKWLQYQKIKWKIQRRDRMRRAQ---LF----------GDMNSSAGRSALG 1296
sp|Q6CUS7|DPOE_KLULA MLDDYVGWLQYQKTKWKIQHIDRKKREK---LF----------GKTSRASDRSALG 1217
sp|Q752B8|DPOE_ASHGO IDEDYVGWLNYQKIKWQLQAEERHRRKQ---LF----------GSSSTLNDRSALG 1212
NLS


Linker
CTD
tr|A0A3P5YCF9|A0A3P5YCF9_BRACM S FFRRPEEALTSSHWQI----------IQLVPS-PQ 1360
sp|F4HW04|DPOE1_ARATH S YFRRPEEALTSSHWQI----------IQLVPS-PQ 1307
tr|D7KI06|D7KI06_ARALL S YFRRPEEALTSSHWQVCT----------IIQLVPS-PQ 1327
tr|A0A8I6WFC8|A0A8I6WFC8_HORVV T FFQKQELSLFRSHWQ----------IIQLAPS-TM 1344
tr|A0A452YAR6|A0A452YAR6_AEGTS T FFQKQELSLFRSHWQ----------IIQLAPS-TL 1284
tr|A0A3B5XXC3|A0A3B5XXC3_WHEAT T FFQKQELSLFRSHWQ----------IIQLAPS-TL 1339
tr|A0A1D6H8Z0|A0A1D6H8Z0_MAIZE T FFQKQELALFRSHWQACVFCYLLCSVALVYHLFFFALLMIVNSIYCPHQIIQLASS-TV 1352
tr|A0A835B357|A0A835B357_9POAL T FFQKQELALFRSHWQ----------IIQLASS-TT 1332
tr|A0A2S3GPC0|A0A2S3GPC0_9POAL T FFQKQELALFRSHWQ----------IIQLASS-TT 1337
tr|A0A368PL62|A0A368PL62_SETIT T FFQKQELALFRSHWQ----------IIQLDSS-TT 1341
tr|A0A7I8KHB8|A0A7I8KHB8_SPIIN S FFRRQELLLVHNHWQ----------IIQFVPS-AK 1342
tr|A0A6I9QIG1|A0A6I9QIG1_ELAGV S FFRRQELALVQSHWQ----------IIQLVPS-TQ 1326
tr|A0A8B8JBC2|A0A8B8JBC2_PHODC S FFRRQELALVQSHWQ----------IIQLNPS-TQ 1326
tr|A0A9Q0FIP5|A0A9Q0FIP5_9ROSI S YFTRPEVVLTRCHWQ----------IIQLLPS-SL 1352
tr|A0A2C9VYJ6|A0A2C9VYJ6_MANES S YFATHEISLTRCHWQ----------IIQLLPS-SH 1333
tr|A0A7J7C034|A0A7J7C034_TRIWF S YFRRHEVALTRCHWQ----------IIQLVPG-SQ 1326
tr|A0A2N9F976|A0A2N9F976_FAGSY S YFRRQEVALTRCHWQ----------IIQLVPS-SE 1323
tr|A0A7N2KME6|A0A7N2KME6_QUELO S YFKRHEASLTRCHWQ----------IIQLVPS-S- 1326
tr|A0A4W5MAK2|A0A4W5MAK2_9TELE S FLRRTARSILDMPWQ----------IVQIAET-SH 1343
tr|A0A8C7Q9V4|A0A8C7Q9V4_ONCMY S FLRRTARSILDMPWQ----------IVQIAET-SH 1305
tr|A0A6P7J2Q5|A0A6P7J2Q5_9TELE S FLRRTARSILDMPWQ----------IVQIAET-SH 1341
tr|A0A6P8SZE0|A0A6P8SZE0_GYMAC S FLRKTARSILDMPWQ----------IVQIAET-SH 1341
tr|A0A8J0QX44|A0A8J0QX44_XENTR S FLRRTARSILDMPWQ----------IVQIGES-SQ 1340
tr|A0A4X2K1K8|A0A4X2K1K8_VOMUR H FLRKTARSILDLPWQ----------VVQISET-NQ 1326
tr|A0A5F4D2S5|A0A5F4D2S5_CANLF G FLRRTARSILDLPWQ----------IVQISET-SQ 1376
tr|A0A7J7RQ84|A0A7J7RQ84_RHIFE G FLRRTARSILDLPWQ----------IVQISET-TQ 1314
tr|A0A6P6D0R3|A0A6P6D0R3_PTEVA G FLRRTARSILDLPWQ----------IVQISET-SQ 1333
tr|A0A2K6CI58|A0A2K6CI58_MACNE S FLRRTARSILDLPWQ----------IVQISET-SQ 1318
tr|F6QAQ1|F6QAQ1_MACMU S FLRRTARSILDLPWQ----------IVQISET-SQ 1343
tr|G3RQP2|G3RQP2_GORGO S FLRRTARSILDLPWQ----------IVQISET-SQ 1320
sp|Q07864|DPOE1_HUMAN S FLRRTARSILDLPWQ----------IVQISET-SQ 1341
tr|A0A2I3S482|A0A2I3S482_PANTR S FLRRTARSILDLPWQ----------IVQISET-SQ 1341
tr|F6Z911|F6Z911_HORSE G FLRRTARSILDLPWQ----------IVQISET-SQ 1341
tr|A0A5G2RB90|A0A5G2RB90_PIG G FFRRTARSILDLPWQ----------IVQISET-SQ 1314
tr|A0A420PL83|A0A420PL83_FUSOX Q TFMKQAHTTYMSNWQ----------VLHLKAT-ET 1296
tr|A0A1V6PGE7|A0A1V6PGE7_PENDC H FFRNQAQLLYISTWQ----------VLQLCET-GR 1333
tr|A0A0G4NU82|A0A0G4NU82_PENCA N FFRNQAQMLYISTWQ----------VLQLCDT-GV 1333
tr|A0A1V6TET5|A0A1V6TET5_9EURO N FFRNQAQMLYISTWQ----------VLQLCDT-GV 1333
tr|A0A167WJ50|A0A167WJ50_PENCH N FFRNQAQMLYISTWQ----------VLQLCDT-GV 1333
tr|A0A3A2Z9P9|A0A3A2Z9P9_9EURO N YFRNQAELLYINTWQ----------VLQLRET-GR 1334
tr|A0A401L0U3|A0A401L0U3_ASPAW N LFRNQAELLYINTWQ----------ILQLSET-GR 1337
tr|A0A317VH96|A0A317VH96_ASPEC N LFRNQAELLYINTWQ----------ILQLSET-GR 1333
tr|A0A117E1M7|A0A117E1M7_ASPNG N LFRNQAELLYINTWQ----------ILQLSET-GR 1337
tr|A0A1D8PTD0|A0A1D8PTD0_CANAL G MFRKRAENIAGTNWE----------ILEYKLDPTK 1330
tr|C5M3R9|C5M3R9_CANTT G MIRKRAENIAGSNWE----------ILEYKPDPSK 1325
sp|Q6FNY7|DPOE_CANGA S MIRKQAEMFANGSWE----------VLQYRDN-YE 1329
sp|P21951|DPOE_YEAST S MIRKQAESYANSTWE----------VLQYKDS-GE 1272
tr|A0A7H9HLU1|A0A7H9HLU1_9SACH G MIRKQAETYANSSWE----------ILQYKPS-AQ 1321
sp|Q6CUS7|DPOE_KLULA N LIRKHVESYADKSWE----------ILQCKPS-ID 1242
sp|Q752B8|DPOE_ASHGO N IIKKHAESYANSDWE----------ILQYKNS-SE 1237




ZBM


ZBM


ZBM-4Fe-4S

```
//End of pol ε sequences from all the four eukaryotes
tr|A0A3P5YCF9|A0A3P5YCF9_BRACM       --      2205
sp|F4HW04|DPOE1_ARATH                --      2161
tr|D7KI06|D7KI06_ARALL               --      2241
tr|A0A8I6WFC8|A0A8I6WFC8_HORVV       --      2224
tr|A0A452YAR6|A0A452YAR6_AEGTS       --      2160
tr|A0A3B5XXC3|A0A3B5XXC3_WHEAT       --      2215
tr|A0A1D6H8Z0|A0A1D6H8Z0_MAIZE       --      2223
tr|A0A835B357|A0A835B357_9POAL       --      2205
tr|A0A2S3GPC0|A0A2S3GPC0_9POAL       --      2210
tr|A0A368PL62|A0A368PL62_SETIT       --      2214
tr|A0A7I8KHB8|A0A7I8KHB8_SPIIN       --      2215
tr|A0A6I9QIG1|A0A6I9QIG1_ELAGV       --      2209
tr|A0A8B8JBC2|A0A8B8JBC2_PHODC       --      2212
tr|A0A9Q0FIP5|A0A9Q0FIP5_9ROSI       --      2257
tr|A0A2C9VYJ6|A0A2C9VYJ6_MANES       --      2223
tr|A0A7J7C034|A0A7J7C034_TRIWF       --      2203
tr|A0A2N9F976|A0A2N9F976_FAGSY       --      2166
tr|A0A7N2KME6|A0A7N2KME6_QUELO       --      2206
tr|A0A4W5MAK2|A0A4W5MAK2_9TELE       --      2286
tr|A0A8C7Q9V4|A0A8C7Q9V4_ONCMY       --      2254
tr|A0A6P7J2Q5|A0A6P7J2Q5_9TELE       --      2292
tr|A0A6P8SZE0|A0A6P8SZE0_GYMAC       TS      2298
tr|A0A8J0QX44|A0A8J0QX44_XENTR       --      2286
tr|A0A4X2K1K8|A0A4X2K1K8_VOMUR       --      2267
tr|A0A5F4D2S5|A0A5F4D2S5_CANLF       --      2319
tr|A0A7J7RQ84|A0A7J7RQ84_RHIFE       --      2255
tr|A0A6P6D0R3|A0A6P6D0R3_PTEVA       --      2282
tr|A0A2K6CI58|A0A2K6CI58_MACNE       --      2263
tr|F6QAQ1|F6QAQ1_MACMU               --      2288
tr|G3RQP2|G3RQP2_GORGO               --      2265
sp|Q07864|DPOE1_HUMAN                --      2286
tr|A0A2I3S482|A0A2I3S482_PANTR       --      2286
tr|F6Z911|F6Z911_HORSE               --      2286
tr|A0A5G2RB90|A0A5G2RB90_PIG         --      2257
tr|A0A420PL83|A0A420PL83_FUSOX       --      2175
tr|A0A1V6PGE7|A0A1V6PGE7_PENDC       --      2229
tr|A0A0G4NU82|A0A0G4NU82_PENCA       --      2228
tr|A0A1V6TET5|A0A1V6TET5_9EURO       --      2228
tr|A0A167WJ50|A0A167WJ50_PENCH       --      2228
tr|A0A3A2Z9P9|A0A3A2Z9P9_9EURO       --      2229
tr|A0A401L0U3|A0A401L0U3_ASPAW       --      2233
tr|A0A317VH96|A0A317VH96_ASPEC       --      2229
tr|A0A117E1M7|A0A117E1M7_ASPNG       --      2233
tr|A0A1D8PTD0|A0A1D8PTD0_CANAL       --      2211
tr|C5M3R9|C5M3R9_CANTT               --      2207
sp|Q6FNY7|DPOE_CANGA                 --      2217
sp|P21951|DPOE_YEAST                 --      2162
tr|A0A7H9HLU1|A0A7H9HLU1_9SACH       --      2212
sp|Q6CUS7|DPOE_KLULA                 --      2125
sp|Q752B8|DPOE_ASHGO                 --      2120
```

From PLANTS

A0A3P5YCF9_BRACM, *Brassica campestris*
D7KI06_ARALL, *Arabidopsis lyrata*
A0A452YAR6_AEGTS, *Aegilops tauschii*
A0A1D6H8Z0_MAIZE, *Zea mays*
A0A2S3GPC0_9POAL, *Panicum hallii*
A0A7I8KHB8_SPIIN, *Spirodela intermedia*
A0A8B8JBU2_PHODC, *Phoenix dactylifera*
A0A2C9VYJ6_MANES, *Manihot esculenta*
A0A2N9F976_FAGSY, *Fagus sylvatica*

F4HW04|DPOE1_ARATH, *Arabidopsis thaliana*
A0A8I6WFC8_HORVV, *Hordeum vulgare* subsp. *Vulgare*
A0A3B5XXC3_WHEAT, *Triticum aestivum*
A0A835B357_9POAL, *Digitaria exilis*
A0A368PL62_SETIT, *Setaria italic*
A0A6I9QIG1_ELAGV, *Elaeis guineensis* var. *tenera*
A0A9Q0FIP5_9ROSI, *Turnera subulata*
A0A7J7C034_TRIWF, *Tripterygium wilfordii*
A0A7N2KME6_QUELO, *Quercus lobata*

From ANIMALS

A0A6P7J2Q5_9TELE, *Parambassis ranga*
A0A6P7J2Q5_9TELE, *Hucho hucho*
A0A8J0QX44_XENTR, *Xenopus tropicalis*
A0A5F4D2S5_CANLF, *Canis lupus familiaris*
A0A6P6D0R3_PTEVA, *Pteropus vampyrus*
F6QAQ1_MACMU, *Macaca mulatta*
Q07864|DPOE1_HUMAN, *Homo sapiens*
F6Z911_HORSE, *Equus caballus*

A0A8C7Q9V4_ONCMY, *Oncorhynchus mykiss*
A0A6P8SZE0_GYMAC, *Gymnodraco acuticeps*
A0A4X2K1K8_VOMUR, *Vombatus ursinus*
A0A7J7RQ84_RHIFE, *Rhinolophus ferrumequinum*
A0A2K6CI58_MACNE, *Macaca nemestrina*
G3RQP2_GORGO *Gorilla gorilla*
A0A2I3S482_PANTR, *Pan troglodytes*
A0A5G2RB90_PIG, *Sus scrofa*

From YEASTS and HIGHER FUNGI

A0A420PL83_FUSOX, *Fusarium oxysporum*
A0A0G4NU82_PENCA, *Penicillium camemberti*
A0A167WJ50_PENCH, *Penicillium chrysogenum*
A0A401L0U3_ASPAW, *Aspergillus awamori*
A0A117E1M7_ASPNG, *Aspergillus niger*
C5M3R9_CANTT, *Candida tropicalis*
P21951|DPOE_YEAST, *Saccharomyces cerevisiae*
Q6CUS7|DPOE_KLULA, *Kluyveromyces lactis*

A0A1V6PGE7_PENDC, *Penicillium decumbens*
A0A1V6TET5_9EURO, *Penicillium flavigenum*
0A3A2Z9P9_9EURO, *Aspergillus sclerotialis*
A0A317VH96_ASPEC, *Aspergillus eucalypticola*
A0A1D8PTD0_CANAL, *Candida albicans*
Q6FNY7|DPOE_CANGA, *Candida glabrata*
A0A7H9HLU1_9SACH, *Torulaspora* sp.
Q752B8|DPOE_ASHGO, *Ashbya gossypii*

Figure 5 Mix and Match MSA of all the four eukaryotic ε DNA polymerase catalytic subunits

3.3. ZBMs and their role(s) in eukaryotic replicative polymerases

All the three eukaryotic replicative DNA pols (α, ε and δ) invariably possess the two completely conserved ZBMs at the CTDs of their catalytic subunits, and are known as CysA and CysB, suggesting an important role(s) in eukaryotic genome replication and regulation. In pol ε, the CysA (-$Cx_2Cx_nCx_{2/4}C$-) site is the regular ZBM, where a Zn^{2+} binds, but a [4Fe-

4S]$^{2+}$ binds to the CysB (-$Cx_2Cx_{11}CxC$-) site. The [4Fe-4S]$^{2+}$ cluster is implicated in the redox signaling and regulation of the replication process in eukaryotes. In fact, the CysB motif is present in all the four yeast's B-family DNA pols which include the DNA mutagenesis enzyme pol ζ [22]. Interestingly, the ε pols contain an additional ZBM within the pol domain which is also shown to bind an [4Fe-4S]$^{2+}$ cluster. Based on spectroscopic and other data, Jain *et al*. [23] suggested that the three highly conserved C residues (C^{665}, C^{668}, C^{677} and C^{763}) in yeast ε pol's pol domain are involved in binding to [4Fe-4S] cluster (Fig. 3). This was based on their observation that the wild-type yeast pol ε's catalytic core was yellowish-brown, but a mutant in which all the three residues (C^{665}, C^{677} and C^{763}) were mutated, the mutant enzyme was colourless and its UV-Vis spectrum lacked the broad peak centered at 400 nm. Besides, they also found that this triple mutant was deficient only in the DNA pol activity, but not in the PR exonuclease activity, suggesting a link between the pol activity and the [4Fe-4S] cluster suggesting its possible involvement in the eukaryotic genome replication and regulation. Thus, the additional ZBM found exclusively in the ε pols suggests that ε pols are more sensitive to oxidative stress than the other two replicative pols [23]. Furthermore, Pinto *et al*. [24] have shown that the [4Fe-4S]$^{2+}$ cluster of yeast DNA polymerase ε is redox active and can undergo DNA-mediated signaling. A double cysteine-to-serine mutant (C^{665}→S and C^{668}→S) of $Pol2_{CORE}exo^-$, which lacked the [4Fe-4S] cluster, showed no redox signal upon oxidation as the wildtype enzyme. Significantly, protein oxidation yielded a sharp decrease in polymerization, while rereduction restored the activity almost to the level of the untreated enzyme.

It is interesting to note that pol δ also possesses two ZBMs at its CTD, viz. CysA (-Cx_2CnCx_2C-) and CysB (-$Cx_2Cx_9Cx_4C$-) [2]. Similar SDM experiments on pol δ by Netz *et al*. [22] have shown that the CysB motif was bound to [4Fe-4S] cluster and involved in the formation of the pol active complexes. For example, the loss of [4Fe-4S] cluster binding by Cys-ligand mutagenesis of pol δ, destabilized the CTD and abolished interactions between the polD1 and polD2 subunits. They found that the conversion of the C residues in the CysB site, viz. C^{1056}, C^{1059}, C^{1069}, C^{1074} to **A** residues abolished the [4Fe-4S] binding [22]. Furthermore, a lethal double mutation **C^{1059}→S/C^{1074}→S** in CysB site of pol δ disrupted its binding to both polD1 and polD2. However, in marked contrast, a lethal double mutation in CysA site (**C^{1012}→S/C^{1027}→S**) did not alter the subunit composition of the purified pol δ complex, and also did not affect the pol δ interactions with the subunits as evidenced from a yeast two hybrid analysis. However, the double mutation severely decreased PCNA-dependent replication processivity, suggesting the utmost importance of this metal centre for polymerase function. Thus, the CysA (PCNA-binding) and CysB (subunits interactions) motifs play distinct roles in the eukaryotic replicative pols.

4. Active site analyses of the DNA polymerase ε

The pol domain's active site amino acids of the pol ε from plant and animal sources is mostly arrived at from the data available on the yeast enzyme. The active sites of the yeast enzyme were analyzed by SDM, deletion mutagenesis and crystallographic studies [16, 19, 25, 26]. Pavlov *et al*. [16] found that in the yeast enzyme, the **Y^{831}→A** replacement by SDM reduces replication fidelity and its participation in chromosomal replication suggesting its involvement in the pol active site. Interestingly, the present study shows that the **Y^{831}** (highlighted in dark blue in Table 1) is in the proposed and highly conserved template-binding **–Y^{831}G**- pair of the yeast pol ε.

Kesti *et al*. [20] found that a yeast strain with an N-terminal deletion, which deletes the pol domain of the *pol* ε (*pol2*) gene, but retaining the C-terminal region intact, grew slowly and was viable. Their in-frame deletion experiment lacked the sequence, encoding amino acids 176–1134 of the 2222–amino-acid pol ε protein containing all of the conserved motifs associated with the DNA pol and PR exonuclease functions. They found that such a large N-terminal deletion of the yeast enzyme consisting of all the known pol and PR exonuclease domains was dispensable for DNA replication, repair, and cell viability. Furthermore, such deletion of the PR and pol catalytic domains of pol ε from the NTD was not lethal and did not block chromosomal DNA replication [20]. Besides, they found that the C-terminal portion of the enzyme was both necessary and sufficient for viability and hence, suggested that another polymerase could possibly substitute for the polymerization function of pol ε in their N-terminal deletion mutant. Interestingly, the present study has shown that the presence of one more putative pol active site in the CTD (highlighted in yellow) could have possibly compensated for the pol function in their N-terminal deletion mutants. In another deletion analysis, Ohya *et al*. [27] found a *pol* ε deletion mutant of yeast was temperature-sensitive, exhibited severe defects in chromosomal DNA replication at the permissive temperature, and underwent premature senescence, suggesting the role of PR function in such deletion mutants as the CTD does not possess a functional PR active site in the yeast pol ε (Table 1).

The crystal structure of yeast DNA pol ε catalytic subunit has been solved by Jain *et al*. [25]. The yeast enzyme adopts the universal 'right-hand' DNA pol fold with an active site formed by a 'palm' (holding the catalytic residues for dNTP addition), a 'thumb' (that binds the template-primer duplex) and 'fingers' (interacting with incoming nucleotides) and an exonuclease domain. The enzyme's palm subdomain was found to be larger and more elaborate than that is found in pols α and δ. In the pol ε, the pol and exo active sites are separated by ~41 Å in a direction roughly perpendicular to the

DNA axis. Pol ε's fidelity to a nascent Watson-Crick base-pairing was determined primarily by residues V^{825}, N^{828}, S^{829}, **Y^{831}**and **G^{832}** from the fingers' domain (all the five amino acids are in the proposed catalytic core of the enzyme) and by Y^{645} from the palm domain. As discussed elsewhere, the **Y^{831}→A** replacement in pol ε reduced replication fidelity and its participation in chromosomal replication, but without eliminating an additional function that is essential for viability [25]. It is interesting to note that the same invariant **$Y^{831}G^{832}$** pair is proposed here as the template-binding pair by sequence similarity and is in close agreement with other DdDps and DdRps. The pol active site is characterized by the catalytic metal-binding residues, D^{640} and D^{877}. Both (–HV**D^{640}**- and –DT**D^{877}**-) are found to be completely conserved in the eukaryotic ε pols. The PR exonucleases active site amino acids are further confirmed by SDM and X-ray crystallographic data. When the D^{290} and E^{292} of the first triad in the PR exonuclease were mutated to A (highlighted in dark blue in Table 1), the exonuclease activity was abolished, confirming their direct involvement in PR function [15]. The x-ray crystallographic data further confirmed that the PR exo domain was located on the opposite side of the DNA of the thumb domain and contained the catalytic residues (D^{290}, E^{292} and D^{477} and D^{275}, E^{277} (numberings are based on the yeast and human enzymes, respectively) for PR activity [25]. A catalytic metal ion was found to be coordinated by D^{290} (D^{275} in human enzyme) (Fig. 5A). These active site amino acids are in complete agreement with the present MSA analysis (highlighted in light blue).

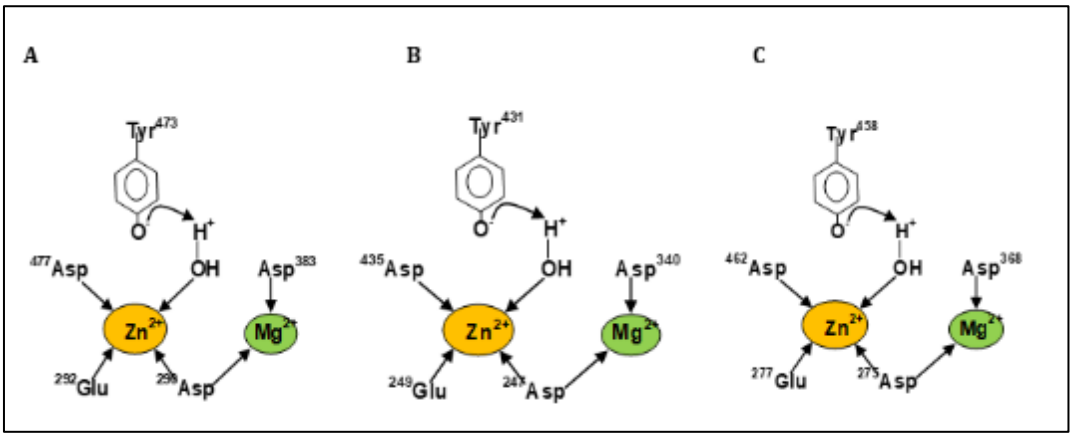
Table 1 shows a comparative account of the active site amino acids in the PR exonuclease and pol domains of the eukaryotic DNA pol ε from different sources from yeasts to humans. The PR exonuclease and pol active sites are almost identical in all eukaryotes from lower (yeasts and higher fungi) to higher (plants and animals) except for a small difference where the plant and animal enzymes use a –KC- pair at the proton abstraction site and the yeast and higher fungal enzymes use a –KV- pair without altering the invariant K in all. The putative second pol active site amino acids are also identified in the CTD. However, the second PR exonuclease active site in the CTD is not complete in all ε pols (Table 1).

Table 1 Proposed active site amino acids of PR exonuclease and pol domains of eukaryotic ε pols

	S. cerevisiae	*A. thaliana*	*Homo sapiens*
PR Exo-1 (NTD)	-DIE^{292}—-FD^{383}—Y^{473}—D^{477}-	-DIE^{249}—FD^{340}—Y^{431}—D^{435}-	-D^{275}IE^{277}—F^{367}D^{368}—Y^{458}S^{459}–D^{462}-
Pol-1(NTD)	-^{816}YDSLQ^{4}LAHK^{1}VILNSFY^{8}GYV^{834}-	-^{774}YDSLQ^{4}LAHK^{1}CILNSFY^{8}GYV^{792}-	-^{801}YDSLQ^{4}LAHK^{1}CILNSFY^{8}GYV^{819}-
PR Exo-2 (CTD)	-DIE^{1216}———?———?—?-	-DIE^{1180}———?———?—?-	-DIE^{1424}———LD^{1516}———Y^{1636}—D-^{1640}
Pol-2 (CTD)	-^{1837}HNLTKK^{1}ALLQLVNEFSAL^{13}GS-	-^{1789}HKVMQK^{1}VFALLLTDLRRL^{13}GA-	-^{1842}HNMMKK^{1}LFLQLIAEFKRL^{13}GS-

Adapted from Palanivelu [8]; Active site amino acids in dark blue are confirmed by SDM analysis;
*Mutations in pol ε PR exo active site region (highlighted in dark blue) lead to several types of cancers, e.g., **D^{275}** leads to endometrial, breast, glioblastoma, colorectal, lung cancers; **F^{367}** leads to endometrial, colon cancers; **S^{459}** leads to colon, endometrial, glioblastoma, duodenal cancers.*

The 3'→5' PR exonucleolytic activity is intrinsic to the replicative DNA pols ε and δ and is essential for the faithful replication of the genomic DNA. Figs. 6A, B and C show the proposed active site amino acids at the PR exonuclease active sites of yeast, plant and human ε pols, respectively, which is based on the crystallographic, SDM and MSA data.



Figures 6A, B and C. Proposed amino acids at the PR exonuclease active sites of DNA pols ε of *S. cerevisiae* (A), *A thaliana* (B) and humans (C)

The proposed two-metal ion in the active site in the PR exonucleases is based on Beese and Steitz's observation of similar active site structure in the 3'→5' PR exonuclease active site in *E. coli* pol I [28] and is also further confirmed from a large number DdRps [6].

5. Conclusions

Understanding the functions of the replicative pols, viz. α, ε and δ, which make the replisome complex, is a prerequisite for understanding the mechanism of genome replication and maintenance of genome integrity in eukaryotes. The present study unravels striking similarities of one of the main replicative pols ε from lower to higher eukaryotes at the molecular level. The MSA analysis has shown that the ε pols in all eukaryotes contain the same template-binding pairs, catalytic and nucleotide selection amino acids and also the same catalytic metal-binding motifs. Furthermore, the two completely conserved ZBMs in the CTDs of all the three replicative pols suggest their essential roles in genome replication and regulation. A unique $[4Fe\text{-}4S]^{2+}$ cluster-binding motif in the ε pol domain is implicated in the redox signaling and the regulation of the replication process in eukaryotes as this enzyme links the DNA replication machinery to S-phase checkpoint-control in eukaryotes. Moreover, the PR exonuclease active site amino acids are identical in all eukaryotic enzymes and belong to the DEDD(Y) superfamily of PR exonucleases. These findings establish that the pol, PR exonuclease and CTD domains of the ε pols and their roles in genome replication, maintenance of genome integrity and regulation are highly conserved across all eukaryotes, from yeasts to higher animals. The highly conserved pol, PR exonuclease and CTD domains strongly suggest a possible common evolutionary origin of the ε pols in eukaryotes.

Compliance with ethical standards

Acknowledgments

The author wishes to thank Dr. P. Sathiyamoorthy, Chief Scientist, GEM Research Foundation, Chennai, for critical evaluation of the manuscript.

Disclosure of conflict of interest

The author has declared that no competing interests exist.

References

[1] Palanivelu P. Analyses of priming reactions and proofreading functions during initiation of replication of prokaryotic and eukaryotic genomes. Br J Pharm Med Res. 2022; 7:3790-7828.

[2] Palanivelu P. Replicative δ polymerases from plants and animals possess very similar polymerase, proofreading and regulatory domains. World J Adv Res Rev. 2024, 23:2239–2260.

[3] Jain R, Aggarwal AK, Rechkoblit O. Eukaryotic DNA polymerases. Curr Opin Struct Biol. 2018; 53:77–87.

[4] Johansson E, Dixon N. Replicative DNA Polymerases. Cold Spring Harb Perspect Biol. 2013; 5: a012799, doi: 10.1101/cshperspect. a012799.

[5] Pavlov YI, Shcherbakova PV, Rogozin IB. Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. Int Rev Cytol. 2006; 255:41–132.

[6] Palanivelu P. DNA polymerases – An insight into their active sites and mechanism of action, In: Recent Advances in Biological Research, Vol 1, Chapter 2, pp 1-39, SCIENCEDOMAIN International Book Publishers, UK, ISBN: 9788193422441, DOI: 10.9734/bpi/rabr/v1; (2019).

[7] DePamphilis ML, Bell SD. Genome duplication (Concepts, mechanisms, evolution and disease). London and New York: Garland Science; 2010.

[8] Palanivelu P. An Insight into the Mechanism of Genome Duplication in Eukaryotes: Polymerase and Proofreading Functions by Eukaryotic DNA Replicases. Int J Recent Sci Res. 2022; 13:2076-2116.

[9] Burgers PMJ, Kunkel TA. Eukaryotic DNA replication fork. Annu Rev Biochem. 2017; 86: 417–438.

[10] Wintersberger U, Wintersberger E. Studies on deoxyribonucleic acid polymerases from yeast. Partial purification and properties of two DNA polymerases from mitochondria-free cell extracts. Eur J Biochem. 1970; 13: 11–19.

[11] Pospiech H, Syväoja JE. DNA Polymerase epsilon - More Than a Polymerase. Sci World J. 2003; 3:87-104.

[12] Garcia J-A P, De Veylder L, Raynaud C. Plant DNA polymerases. Int J Mol Sci. 2019; 20: 4814.

[13] Ronceret A, Guilleminot J, Lincker F, Gadea-Vacas J, Delorme V, Bechtold N, *et al*. Genetic analysis of two Arabidopsis DNA polymerase epsilon subunits during early embryogenesis. Plant J. 2005; 44:223–236.

[14] Jenik PD, Jurkuta RE, Barton MK. Interactions between the cell cycle and embryonic patterning in *Arabidopsis* uncovered by a mutation in DNA polymerase epsilon. Plant Cell. 2005; 17: 3362–77.

[15] Pavlov YI, Maki S, Maki H, Kunkel TA. Evidence for interplay among yeast replicative DNA polymerases alpha, delta and epsilon from studies of exonuclease and polymerase active site mutations. BMC Biol, 2004; 2:1-13.

[16] Pavlov YI, Shcherbakova PV, Kunkel TA: *In vivo* consequences of putative active site mutations in yeast DNA polymerases, alpha, epsilon, delta, and zeta. Genetics. 2001; 159:47-64.

[17] Poot RA, Dellaire G, Hulsmann BB, Grimaldi MA, Corona DFV, Becker PB, *et al*. HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. EMBO J. 2000; 19: 3377–3387.

[18] Dua R, Levy DL, Campbell JL. Role of the putative zinc finger domain of *Saccharomyces cerevisiae* DNA polymerase ε in DNA replication and the S/M checkpoint pathway. J Biol Chem. 1998; 273:30046–30055.

[19] Yuan Z, Georgescu R, Schauer GD, O'Donnell ME, Li H. Structure of the polymerase ε holoenzyme and atomic model of the leading strand replisome. Nat commun. 2020; 11:1-10.

[20] Kesti T, Flick K, Keranen S, Syvaoja JE, Wittenberg C: DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. Mol Cell. 1999; 3:679-685.

[21] Palanivelu P. Eukaryotic Multi-subunit DNA dependent RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism. In: Advances and Trends in Biotechnology and Genetics., Vol. 1. pp 1-38, SCIENCEDOMAIN International Book Publishers, UK, Print ISBN: 978-93-89246-59-9, DOI:10.9734/bpi/atbg/v1;2020.

[22] Netz DJ, Stith CM, Stumpfig M, Kopf G, Vogel D, Genau HM, *et al*. Eukaryotic DNA polymerases require an iron-sulfur cluster for the formation of active complexes. Nat Chem Biol. 2012; 8:125–132.

[23] Jain R, Vanamee ES, Dzikovski BG, Buku A, Johnson RE, *et al*. (2014) An Iron-Sulfur Cluster in the Polymerase Domain of Yeast DNA Polymerase epsilon. J Mol Biol. 2014; 426:301–308.

[24] Pinto MN, Beek JT, Ekanger LA, Johansson E, Barton JK. The [4Fe-4S] Cluster of yeast DNA polymerase ε is redox active and can undergo DNA-mediated signaling. J Am Chem Soc. 2021: DOI: 10.1021/jacs.1c07150.

[25] Jain R, Rajashankar KR. Buku A, Johnson RE, Prakash L, Prakash S, *et al*. Crystal structure of yeast DNA polymerase epsilon catalytic domain. PLoS ONE. 2014; 9:e94835. doi: 10.1371/journal.pone.0094835.

[26] Jain R, Rice WJ, Malik R, Johnson RE, Prakash L, Prakash S, *et al*. Cryo-EM structure and dynamics of eukaryotic DNA polymerase δ holoenzyme. Nat Struct Mol Biol. 2019; 26:955–962.

[27] Ohya T, Kawasaki Y, Hiraga S, Kanbara S, Nakajo K, Nakashima N, Suzuki A, Sugino A. The DNA Polymerase Domain of pol ε Is Required for Rapid, Efficient, and Highly Accurate Chromosomal DNA Replication, Telomere Length Maintenance, and Normal Cell Senescence in *Saccharomyces cerevisiae.* J Biol Chem. 2002; 277:28099–28108

[28] 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.