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

Identification of DEDD- and PHP-Superfamilies of Proofreading Exonucleases in the Acidic Protein Subunit PA of RNA Polymerase of Human Influenza Viruses

Peramachi Palanivelu *

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

World Journal of Advanced Research and Reviews, 2022, 16(02), 804–824

Publication history: Received on 12 October 2022; revised on 19 November 2022; accepted on 21 November 2022

Article DOI: https://doi.org/10.30574/wjarr.2022.16.2.1260

Abstract

RNA polymerase from human influenza viruses A, B and C is a heterotrimric enzyme, made up of three different subunits. It performs the crucial function of both genome replication as well as transcription. One of the RNA polymerase subunits, the polymerase acidic protein subunit (PA), is suggested to function as an endonuclease in a 'cap-snatching' mechanism, unique to influenza viruses. However, by multiple sequence alignment (MSA) analysis, it was found that the PA subunits of the polymerases do harbour typical proofreading (PR) DEDD-superfamily of exonuclease active site in all three viruses. However, in human influenza A virus, an additional putative PR exonuclease active site amino acids belonging to Polymerase-Histidinol Phosphatase (PHP)-superfamily are identified. The identified active site amino acids data are in close agreement with the similar DEDD- and PHP-superfamilies of PR exonucleases, already reported from both DNA-dependent RNA polymerases (DdRps) and RNA-dependent RNA polymerases (RdRps) from prokaryotes, eukaryotes and RNA viruses. The putative PHP–family PR exonuclease active site, identified by MSA analysis, is also in close agreement to the already reported PHP–family of PR exonuclease active sites from DdDps of replicases and pol X polymerases of the bacterial kingdom. Only in the pandemic causing human influenza A virus, the putative PHP-family PR exonuclease domain is found along with the DEDD-family PR exonuclease domain.

Keywords: Human Influenza Viruses; RNA Polymerase; Polymerase Acidic Protein Subunit; Proofreading Exonucleases; DEDD-superfamily Exonucleases; PHP-superfamily Exonucleases; Active site structures

1. Introduction

Maintaining genome stability is crucial to all forms of life. It is achieved by the exceptional fidelity of the replicating enzymes and the closely associated 3'→5' PR exonucleases. The DNA and RNA replicases exhibit strong discrimination between dNTPs and NTPs and rarely insert a wrong nucleotide during the genome replication and hence, the error rate in DNA or RNA replication is very, very minimal and is usually in the order of $\sim 10^{-6}$-10^{-9} and$\sim 10^{-4}$ to 10^{-6}, respectively [1]. However, it should be noted that even one error in critical areas of a protein maybe detrimental to the very survival of organisms. Therefore, the replicases invariably possess PR mechanism(s) to correct any error that might occur during the genome replication processes. When a mismatch is encountered by the DNA or RNA polymerases during replication, the polymerases stall/pause, which in-turn activates the PR function. The PR exonucleases promptly excise the mismatched base. Following the excision of the wrong base, the correct base is inserted and the replication proceeds. Therefore, these replicases are invariably associated with PR enzyme(s) to correct any insertion error(s) during genome replication and thus, make an exact copy of the original genome.

There are mainly two types of PR exonucleases: either they are part of the DNA/RNA polymerases forming multifunctional enzymes (e.g.), DNA pol I, or as a tightly associated enzyme of a multienzyme complex (e.g.), DNA pol III, ExoNs (NSP14) in the SARS-Coronaviruses and in influenza viruses [1, 2]. The mutation rate of influenza A virus is

* Corresponding author: Peramachi Palanivelu
Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai – 625 021, India.

$\sim 1.5 \times 10^{-5}$ mutations/ nucleotide/ replication cycle. It is interesting to note that in human influenza A virus, up to 50% of the nucleotide changes result in amino acid changes whereas amino acid changes are less frequent in influenza B and C viruses and in avian influenza virus A [2]. Thus, these PR exonucleases are an important class of exonucleases and are ubiquitous in biological systems. Based on the active site structures, they are classified into two families, viz. DEDD- and PHP-superfamilies. The DEDD-superfamily consists of again two subfamilies, viz. DEDDy and DEDDh, depending upon whether they employ an invariant Y or a H as the proton acceptor during catalysis [3]. These two superfamilies are invariably found/associated with the DNA/RNA replicases to repair any error during the replications processes [3, 4]. Whereas the DEDD-superfamily is reported from all the kingdoms, the PHP-superfamily has been reported only from the bacterial kingdom. For example, the PHP-superfamily is found in the bacterial replicative DNA polymerases III, in their α-catalytic subunits, (DNA pols III belong to C-family polymerases), and in some members of bacterial X polymerases. (The A- and B-families of the replicative polymerase are found both in prokaryotes and eukaryotes, but the C-family is found only in bacterial kingdom. All the three families are involved in genome replication and repair. Whereas the A–family polymerase possesses two exonuclease domains (3→5' exonuclease and 5'→3' exonuclease), the B-family polymerase possesses only one exonuclease domain, i.e., 3'→5' PR exonuclease domain) [5].

Influenza, commonly known as "the flu", is an infectious disease caused by influenza viruses. Influenza viral infection is an airborne, highly contagious that generally causes acute respiratory illness resulting in variable degrees of systemic symptoms from mild fatigue to respiratory failure and death. Therefore, influenza viral infections are a major public health concern throughout the world.

The influenza viruses are enveloped viruses and belong to the family, Orthomyxoviridae. Out of the 4 genera of the influenza viruses (A-D), only the influenza A, B and C viruses infect humans and the D infects mainly cattle. Among the influenza viruses that infect humans, the infections by influenza A virus is most common and is the main cause of the flu pandemics. Influenza A and B viruses cause seasonal epidemics and affect each year approximately 5–10% of the adult and 20–30% of the pediatric population. According to the World Health Organization, influenza epidemics lead to 3–5 million cases of severe illness and 2,90,000–6,50,000 respiratory deaths each year (This estimate does not take into account of deaths from other diseases such as cardiovascular disease, which can be influenza-related). While the influenza A virus shows animal to human transmissions, the influenza B and C viruses show very limited host range and appear predominantly in humans [6]. This is mainly because the influenza B and C viruses do not have animal reservoirs like the influenza A virus [7]. Out of the three most widely known human influenza viruses, the influenza A and B viruses cause substantial morbidity and mortality in humans and a considerable financial burden worldwide, whereas the influenza C virus causes sporadic outbreaks in humans causing only mild upper respiratory infections [8]. Their genome organization, surface proteins, serotypes, host-cell receptors, transcription and replication processes are already described by Palanivelu [9].

1.1. Structural Features of the Viral RNA Polymerase

As the influenza viruses are negative-strand RNA viruses, they are replicated in the host nucleus, unlike SARS-CoVs, where the lifecycle of the viruses is completed in the cytoplasm itself. The viral RNA genome (vRNA) is transcribed into mRNAs and replicated through a complementary RNA (cRNA) intermediate to produce a large quantity of progeny vRNAs. However, both the RNAs are synthesized by the same enzyme, i.e., the viral RNA-dependent RNA polymerase. Thus, the viral RNA polymerase is the crucial enzyme that is involved in the viral multiplication processes in the host cells. The viral RNA polymerase is a heterotrimeric protein, made up of three different subunits; two basic protein subunits, viz. PB1, PB2 and an acidic protein subunit, PA. Each of them performs a defined function in viral transcription as well as replication processes. For example, the PB1 subunit functions as a polymerase, the PB2 subunit performs the cap-snatching function [9] and the PA subunit has been mainly shown to perform the PR exonuclease function [This work]. The X-ray crystallography structures of the complete polymerase heterotrimer from the influenza A virus has been reported by Pflug et al. [10] and Fan et al. [11]. The X-ray structures revealed that the polymerase forms a compact particle with PB1 at its centre, capped on one face by PB2 and clamped between the two globular domains of PA (Fig. 1). Like other DNA and RNA polymerases, the polymerase subunit PB1 showed the canonical right-hand-like fold, possessing fingers, palm, and thumb subdomains [10, 11].

The influenza viral RNA polymerase (EC 2.7.7.48) is a primer-dependent enzyme, i.e., the enzyme cannot copy the (-) strand RNA template without a small piece of RNA that aligns with the template RNA and provides a starting point for mRNA synthesis. The PB2 subunit of the polymerase generates the primer by a unique *'cap-snatching'* mechanism.

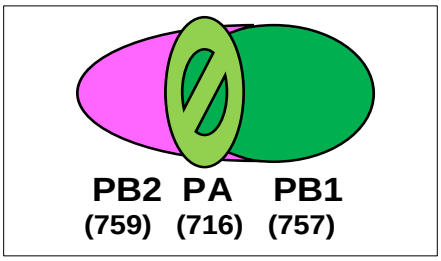


Figure 1 A schematic diagramme showing various subunits of the RNA polymerase from the influenza A virus. Adapted from Palanivelu [9]. (The number of amino acids is given in brackets)

1.2. Role(s) of PA Subunit of the Polymerase in Viral RNA Synthesis

The functions of the three polymerase subunits in transcription and replication have been partially elucidated, yet the precise nature of these functions and their efficiencies in viral RNA transcription and replication depend upon after the final assembly of the heterotrimer polymerase complex. One of the subunits, the PA subunit, has been implicated in various functions in viral RNA synthesis [1 and the references therein]. However, the following results substantiate that the main function of the PA subunit is in the replication of the viral genome. Nakagawa et al. [1996 [12] have shown that the PB1 subunit alone can catalyze cRNA synthesis, and the PA subunit in addition to the PB1 subunit is needed for viral genome replication. By subunit reconstitution experiments, Honda et al. [13] have shown the PB2-PB1 binary complexes showed essentially the same catalytic properties as PB2-PB1-PA heterotrimeric complexes, whereas PB1-PA binary complexes catalyzed *de novo* initiation of vRNA synthesis in the absence of primers *in vitro*, suggesting again its role in the replication process. Further insights on its role in the replication were provided by Kawaguchi et al. [14] using a temperature sensitive (ts) PA subunit mutant. By primer extension assays, they have shown that the tsPA mutant (ts53) virus was defective in synthesis of vRNA and cRNA at the nonpermissive temperature, whereas mRNA synthesis was not affected. Furthermore, ts53 polymerases, including the tsPA, which have been once assembled at the permissive temperature did not inhibit the viral mRNA synthesis even at the nonpermissive temperature. These results suggest that the PA subunit is required mainly for the synthesis of the viral genome and hence, could involve in the PR function to synthesize error-free genomic vRNAs. In this communication, the PA subunits from all the three human influenza viruses are analyzed and the presence of a DEDD- and a PHP-families of PR exonucleases are identified and reported.

2. Material and methods

The protein sequence data of the polymerase acidic protein subunit, PA, of human influenza viruses, A, B and C 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 from the data already available from biochemical, SDM and X-ray crystallographic analyses of the DEDD- and PHP-superfamily exonucleases from various sources were used to arrive at the possible amino acids that make the exonuclease active sites of the PA subunits. Expasy tool was used for pI calculations of the PA subunits of the polymerase.

3. Results and discussion

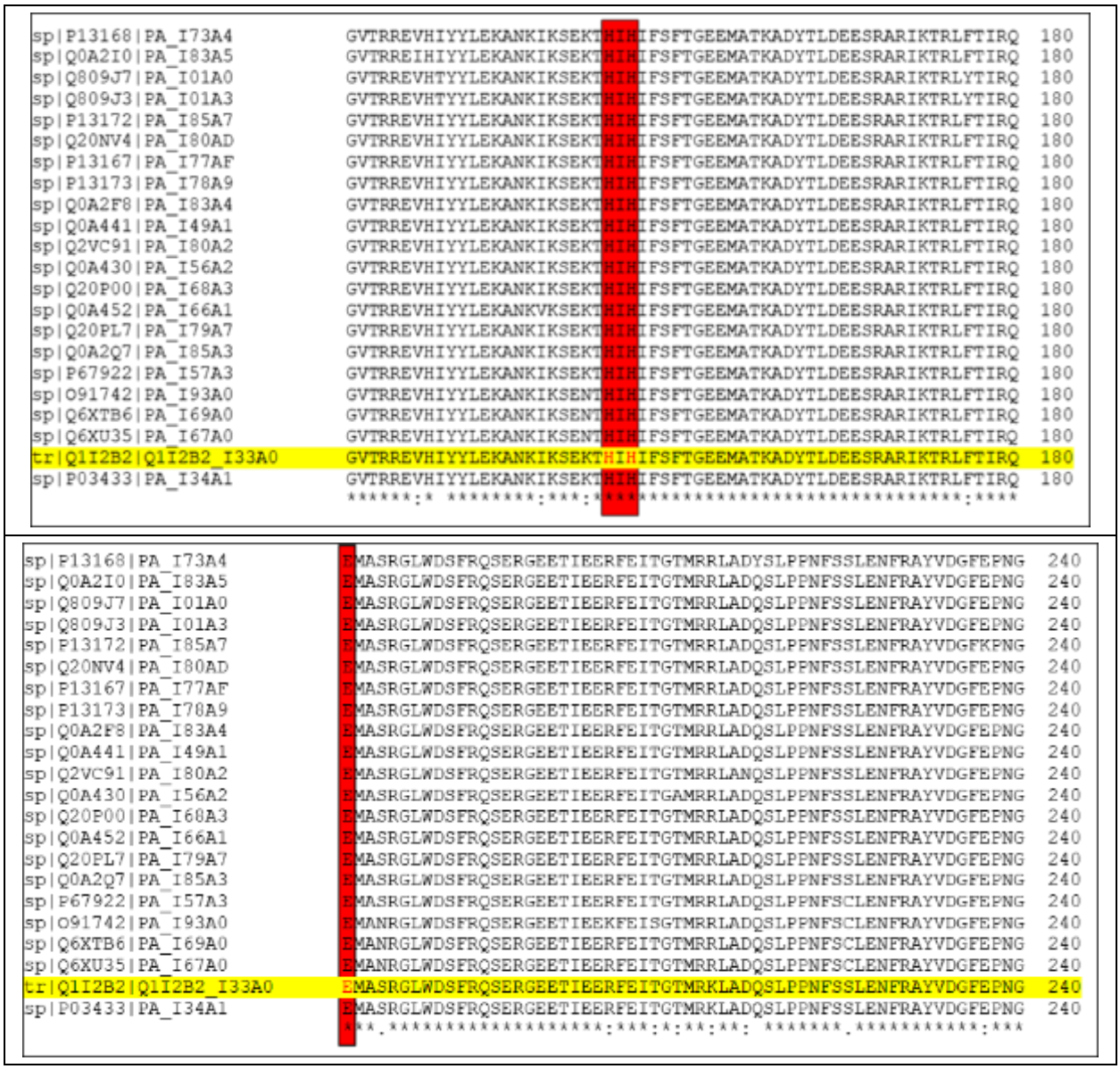
3.1. MSA Analysis of the Polymerase PA subunit of the Influenza Virus A

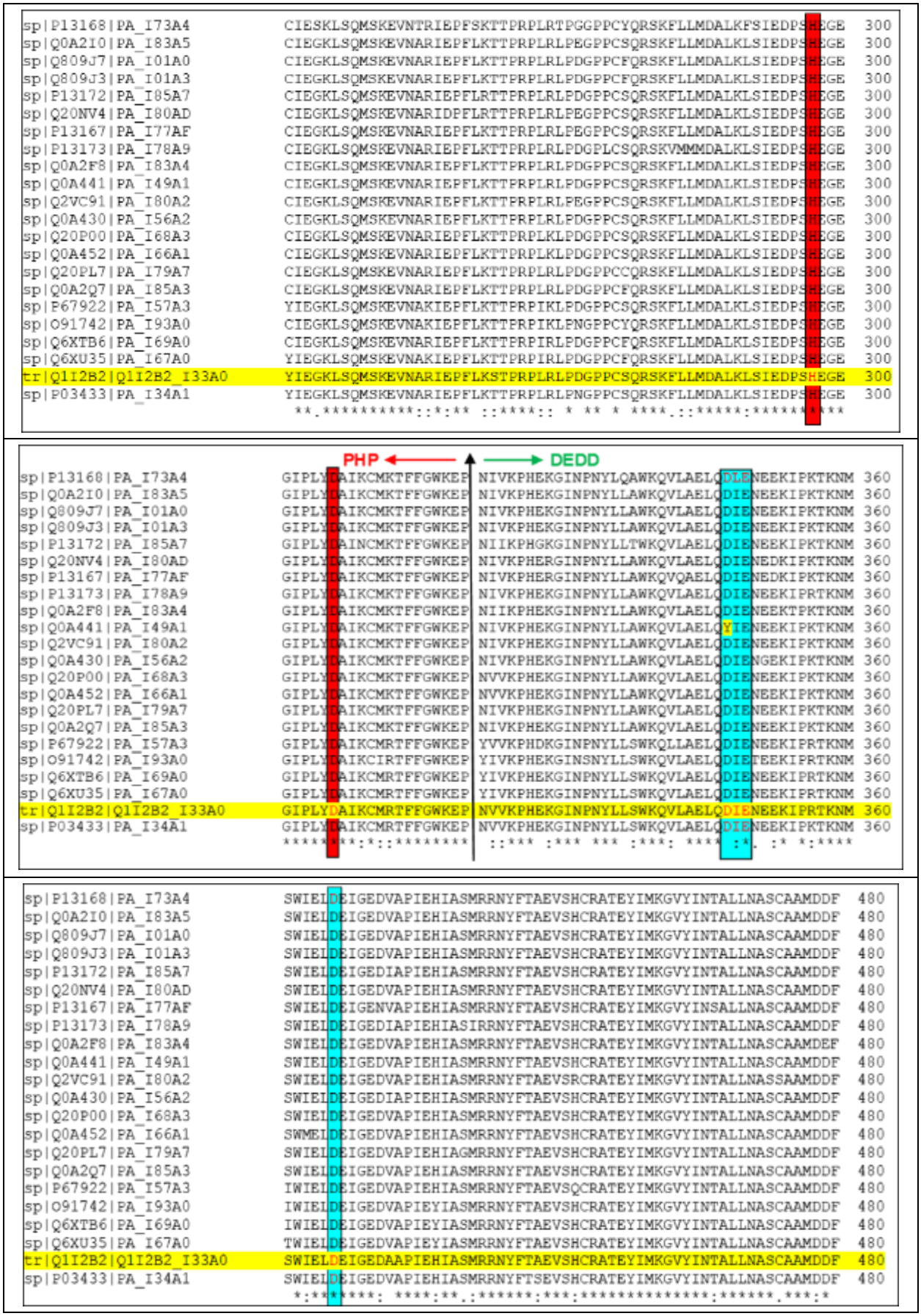
Figure 2 shows the MSA of the PA subunits of the polymerase from several subtypes of the influenza A virus (only the required regions for the discussions are shown here). The pandemic causing influenza A virus, H1N1, is highlighted in yellow, and it showed a theoretical pI of ~5.5. It is clear from Fig. 2 that the PA subunits of different subtypes of the influenza A virus are highly conserved with only a small number of variations. A typical DEDD-family of PR exonuclease active site amino acids (highlighted in light blue) with the pattern DxE---D-----**H**---D, is identified, where the invariant **H** could act as the general base in catalysis. It is interesting to note that almost all the DEDD-superfamily use H, except 3 of them, where they use Y as the proton acceptor during catalysis (Table 1) [1].

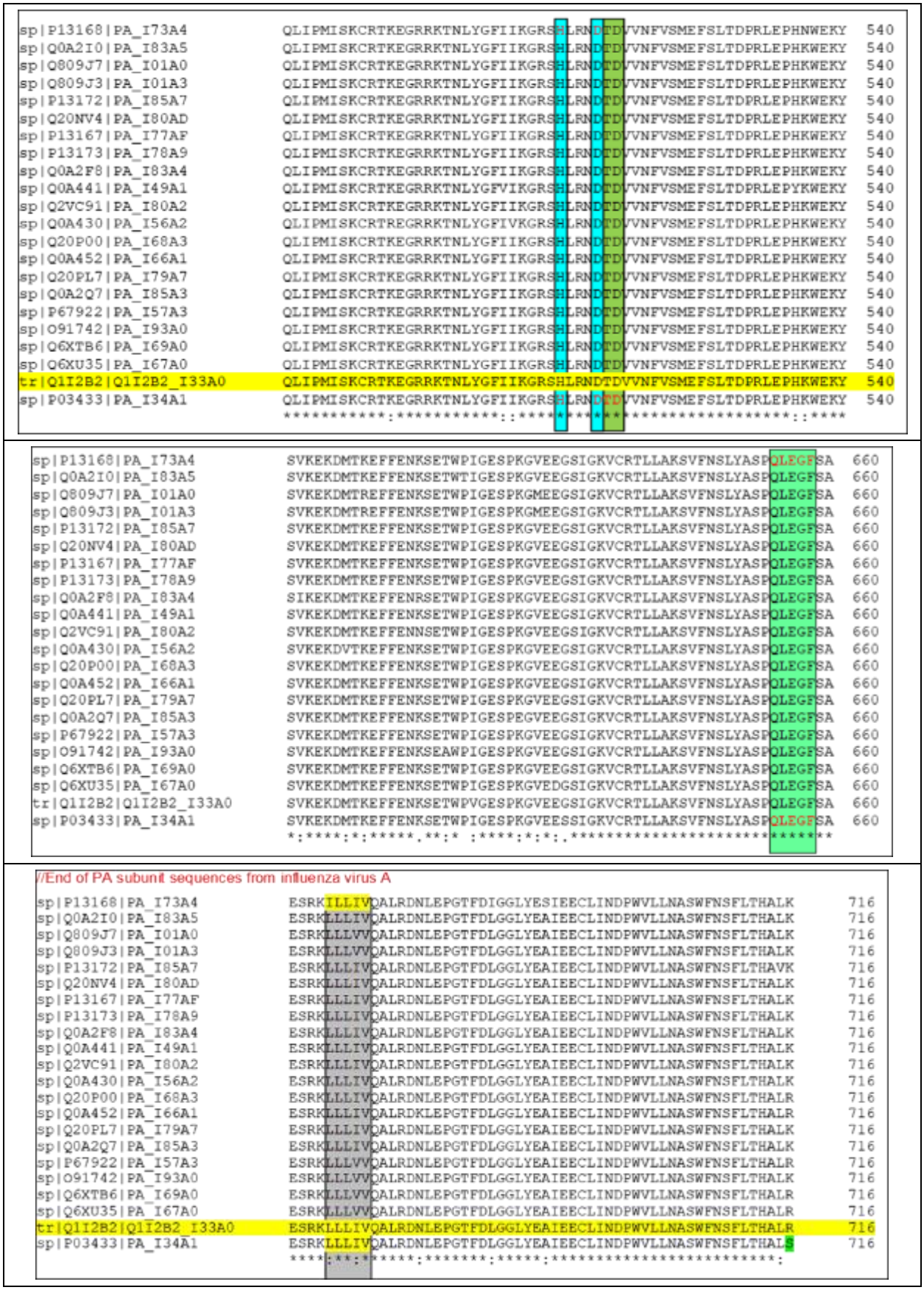
Interestingly, another family of PR exonuclease domain is also identified in the PA subunits of different subtypes of the influenza A virus. This second putative PR exonuclease domain belongs to the PHP-superfamily at the N-terminal region of all the subtypes of the influenza A virus (highlighted in red). The PHP-family showed the typical conserved amino acid pattern -HxH----E---D---**H**- as reported in other PHP-family PR exonucleases (Table 2). It is interesting to note, that the PHP-family is found mainly in the bacterial replicases (in DNA pols III-catalytic α-subunits) and is implicated in metal-dependent proofreading function during replication [1, 15]. The PHP-family exonucleases are also shown to be

involved in base excision repair (BER), which is a major pathway involved in the removal of damaged bases *in vivo*. The BER mechanism is conserved from bacteria to eukaryotes (BER enzymes remove the damaged base-containing nucleotide, resulting a single nucleotide gap that is filled by a DNA polymerase and sealed by a DNA ligase). The presence of the PHP exonuclease domain in the PA subunits of the viral RNA polymerase, which are reported only from the bacterial kingdom, is intriguing. In other words, the PA subunit of the polymerase from influenza A virus could use both the families of PR exonucleases to safeguard their genome (Fig. 2). These two different PR exonucleases maybe exclusive to each other in repairing the genome during the replication processes. In addition, a –DxD- type metal-binding motif was also found in the DEDD exonuclease region (highlighted in green). The C-terminal end is completely conserved with a highly basic amino acid residue K/R in all except in H1N1 (Puerto Rico, 1933), where it ends in a Ser residue. The invariant penta-peptide found in all the human viruses is highlighted in light green. An invariant penta-peptide of branched chain amino acids is also found in the N-terminal region.

CLUSTAL O (1.2.4) MSA of Acidic Protein (PA) subunit from different subtypes of influenza virus A




PHP
DEDD


sp|P13168|PA_I73A4 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHNWEKY 540
sp|Q0A2I0|PA_I83A5 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q809J7|PA_I01A0 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q809J3|PA_I01A3 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|P13172|PA_I85A7 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q20NV4|PA_I80AD QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|P13167|PA_I77AF QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|P13173|PA_I78A9 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q0A2F8|PA_I83A4 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q0A441|PA_I49A1 QLIPMISKCRTKEGRRKTNLYGFVIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPYKWEKY 540
sp|Q2VC91|PA_I80A2 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q0A430|PA_I56A2 QLIPMISKCRTREGRRKTNLYGFIVKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q20P00|PA_I68A3 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q0A452|PA_I66A1 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q20PL7|PA_I79A7 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q0A2Q7|PA_I85A3 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|P67922|PA_I57A3 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|O91742|PA_I93A0 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q6XTB6|PA_I69A0 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|Q6XU35|PA_I67A0 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
tr|Q1I2B2|Q1I2B2_I33A0 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|P03433|PA_I34A1 QLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKY 540
sp|P13168|PA_I73A4 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q0A2I0|PA_I83A5 SVKEKDMTKEFFENKSETWTIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q809J7|PA_I01A0 SVKEKDMTREFFENKSETWPIGESPKGMEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q809J3|PA_I01A3 SVKEKDMTREFFENKSETWPIGESPKGMEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|P13172|PA_I85A7 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q20NV4|PA_I80AD SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|P13167|PA_I77AF SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|P13173|PA_I78A9 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q0A2F8|PA_I83A4 SIKEKDMTKEFFENRSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q0A441|PA_I49A1 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q2VC91|PA_I80A2 SVKEKDMTKEFFENNSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q0A430|PA_I56A2 SVKEKDVTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q20P00|PA_I68A3 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q0A452|PA_I66A1 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q20PL7|PA_I79A7 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q0A2Q7|PA_I85A3 SVKEKDMTKEFFENKSETWPIGESPEGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|P67922|PA_I57A3 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|O91742|PA_I93A0 SVKEKDMTKEFFENKSEAWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q6XTB6|PA_I69A0 SVKEKDMTKEFFENKSETWPIGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|Q6XU35|PA_I67A0 SVKEKDMTKEFFENKSETWPIGESPKGVEDGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
tr|Q1I2B2|Q1I2B2_I33A0 SVKEKDMTKEFFENKSETWPVGESPKGVEEGSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
sp|P03433|PA_I34A1 SVKEKDMTKEFFENKSETWPIGESPKGVEESSIGKVCRTLLAKSVFNSLYASPQLEGFSA 660
//End of PA subunit sequences from influenza virus A
sp|P13168|PA_I73A4 ESRKILLIVQALRDNLEPGTFDIGGLYESIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q0A2I0|PA_I83A5 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q809J7|PA_I01A0 ESRKLLLVVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q809J3|PA_I01A3 ESRKLLLVVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|P13172|PA_I85A7 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHAVK 716
sp|Q20NV4|PA_I80AD ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|P13167|PA_I77AF ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|P13173|PA_I78A9 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q0A2F8|PA_I83A4 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q0A441|PA_I49A1 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q2VC91|PA_I80A2 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q0A430|PA_I56A2 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q20P00|PA_I68A3 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR 716
sp|Q0A452|PA_I66A1 ESRKLLLIVQALRDKLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR 716
sp|Q20PL7|PA_I79A7 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q0A2Q7|PA_I85A3 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|P67922|PA_I57A3 ESRKLLLVVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR 716
sp|O91742|PA_I93A0 ESRKLLLVVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK 716
sp|Q6XTB6|PA_I69A0 ESRKLLLVVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR 716
sp|Q6XU35|PA_I67A0 ESRKLLLVVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR 716
tr|Q1I2B2|Q1I2B2_I33A0 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR 716
sp|P03433|PA_I34A1 ESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALS 716

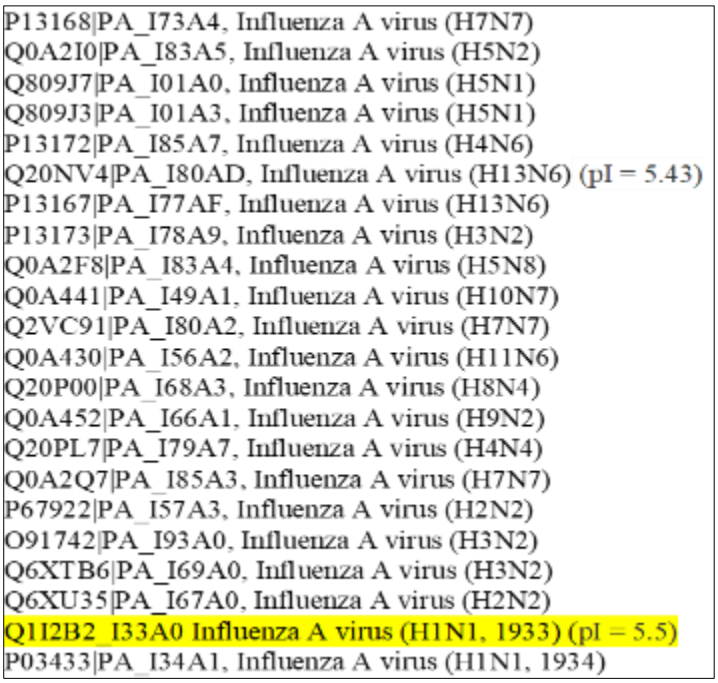


Figure 2 MSA of polymerase acidic proteins (PA) from different strains of influenza virus A

Even though some reports implicated the PA subunit in the cap-snatching function [1 and references therein], but by using temperature-sensitive mutants with defects in the PA gene, it was shown that the principal role of the PA subunit is not in viral mRNA synthesis, but in viral RNA replication [16, 17]. This finding was further corroborated by reverse genetics experiments of the PA subunit by Hao et al. [18]. By using the H9N2 subtype of Avian Influenza Virus (AIV), they found that the PA played a key role in replication and airborne transmission. (Currently, H9N2 is the most prevalent subtype AIV in many countries and regions). The above findings are further confirmed by the presence of putative active sites of the two major PR exonuclease families in the PA subunits of the polymerase. Fig. 3 shows the organization of the two PR superfamilies on the PA subunit of the influenza A virus. The HNH is found in the N-terminal region as found in the bacterial replicases and the DEDD(H) family in the C-terminal region of the PA subunit.

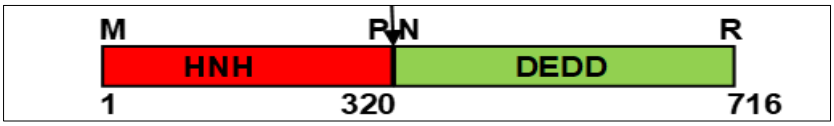


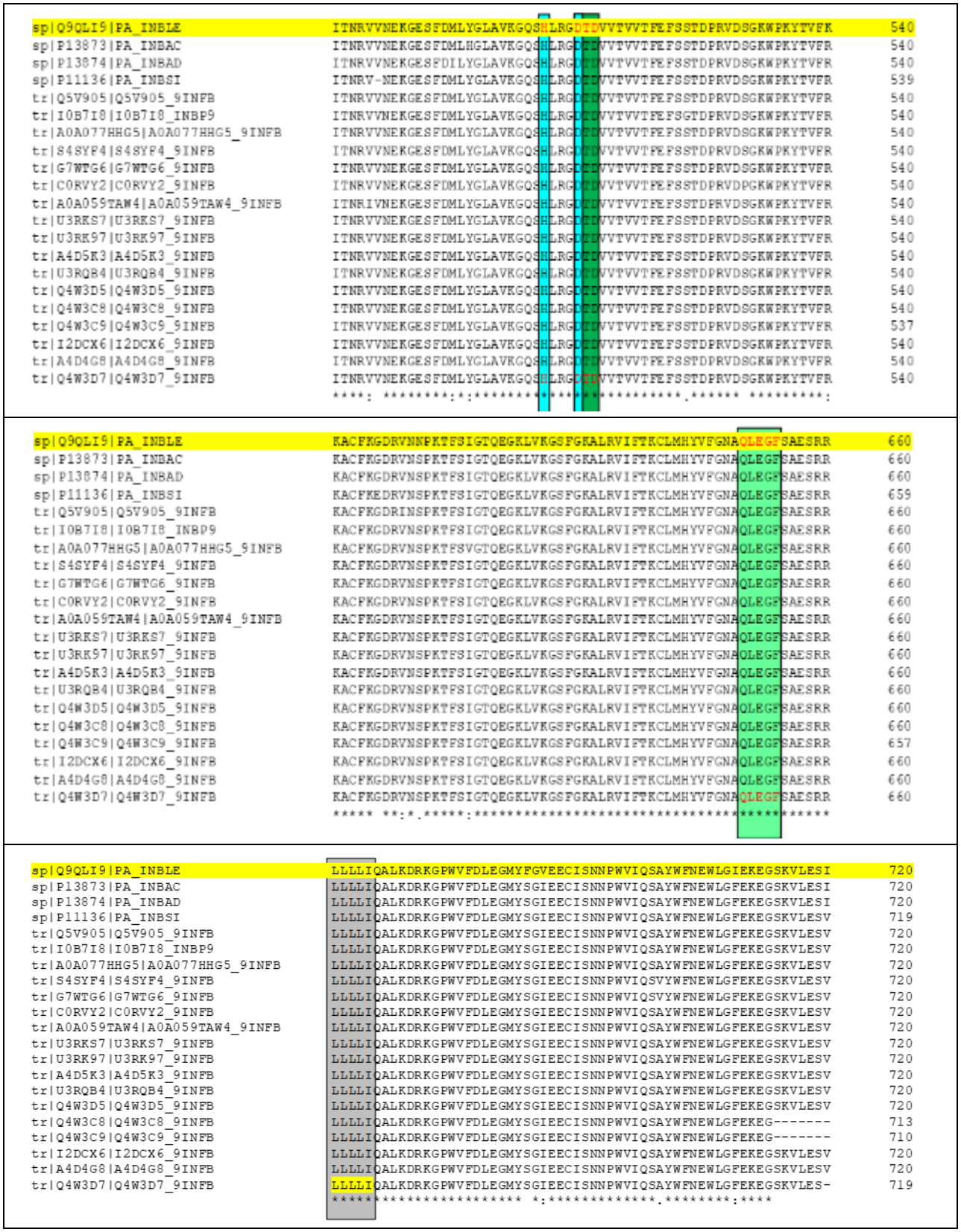
Figure 3 Schematic organization of the two distinct PR exonuclease domains in the PA subunits of Influenza A virus (numbering from the influenza A virus, H1N1 subtype)

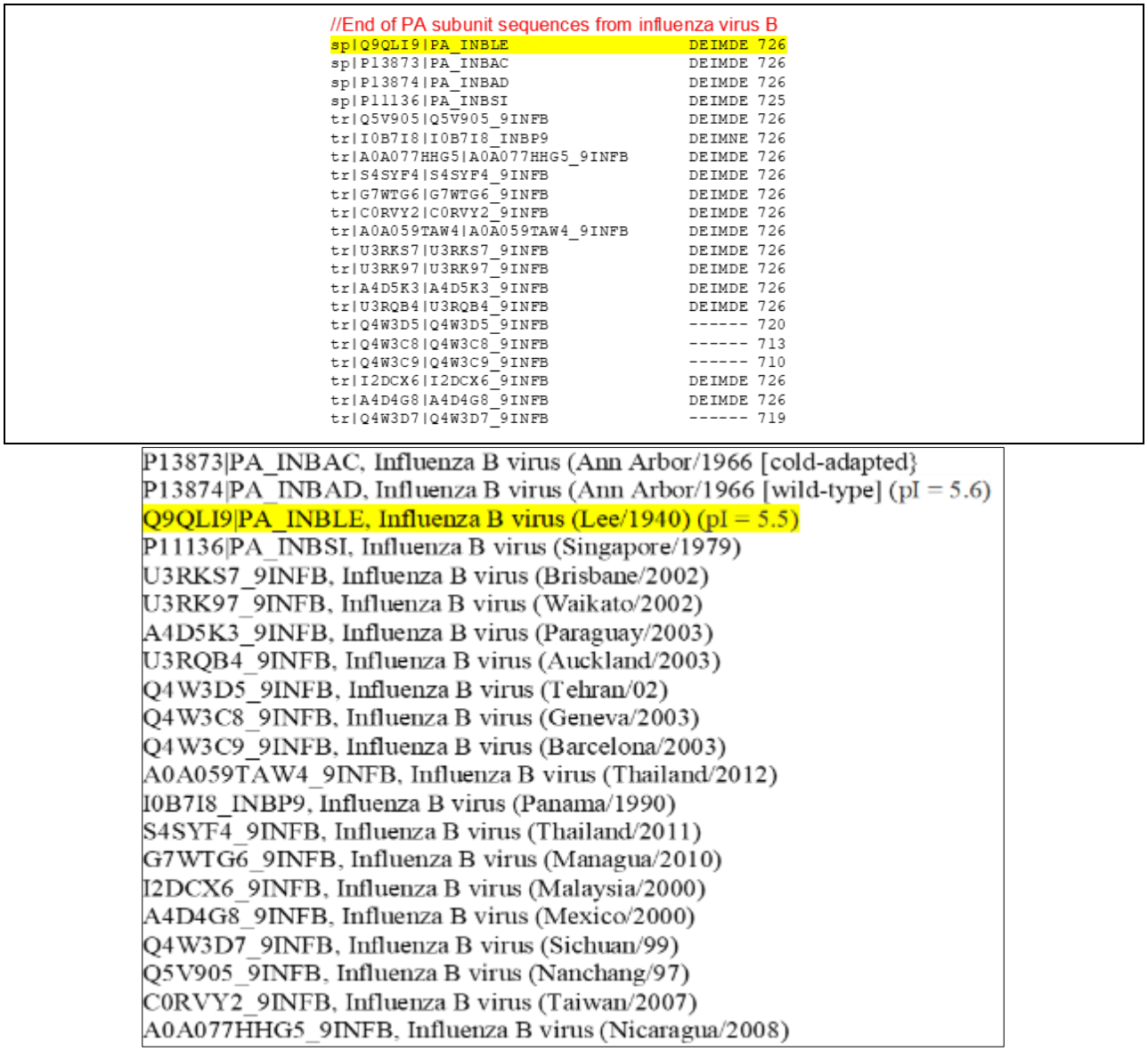
3.2. MSA Analysis of the Polymerase PA Subunit of Influenza B virus

Figure 4 shows the MSA of the PA subunits of the polymerase from the influenza B viral strains, characterized from different regions (only the required regions for the discussions are shown here). The influenza B viral strain, 1940, is highlighted in yellow and it showed a theoretical pI of ~5.5, same as that of the PA subunits of the influenza A virus. It is interesting to note that the PA subunits of different influenza B viral strains are almost completely conserved for the entire sequence as compared to influenza A virus. In influenza B virus also a typical DEDD(H)-family of conserved amino acids (highlighted in light blue) with the pattern DxE---D-----**H**---D, is identified. In a striking contrast to influenza A virus, the putative PHP-family PR exonuclease active site is not identified. In other words, the PA subunit of the polymerase from influenza B virus, possibly uses only the DEDD(H)-family of PR exonuclease. The nuclear localization signal (NLS) is highlighted in orange. A penta-peptide of branched chain amino acids (LLLLI) is observed here also, as in influenza A virus (LLLIV). In contrast to A virus, which ends in a basic amino acid R/K, the C-terminal ends in an acidic amino acid, E, in most of the B viral strains. The invariant penta-peptide found in all the human viruses is highlighted in light green.

CLUSTAL O (1.2.4) MSA of Acidic Protein (PA) Subunit from different strains of influenza virus B


sp|Q9QLI9|PA_INBLE VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSDESSLDEEGKGRVLSRLTELQAE 180
sp|P13873|PA_INBAC VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
sp|P13874|PA_INBAD VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNEHSLDEEGKGRVLSRLTELQAE 180
sp|P11136|PA_INBSI VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|Q5V905|Q5V905_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|I0B7I8|I0B7I8_INBP9 VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|A0A077HHG5|A0A077HHG5_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|S4SYF4|S4SYF4_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|G7WTG6|G7WTG6_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|C0RVY2|C0RVY2_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|A0A059TAW4|A0A059TAW4_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|U3RKS7|U3RKS7_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|U3RK97|U3RK97_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|A4D5K3|A4D5K3_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|U3RQB4|U3RQB4_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|Q4W3D5|Q4W3D5_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|Q4W3C8|Q4W3C8_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|Q4W3C9|Q4W3C9_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 177
tr|I2DCX6|I2DCX6_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|A4D4G8|A4D4G8_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
tr|Q4W3D7|Q4W3D7_9INFB VGITKGLADDYFWKKKEKLGNSMELMIFSYNQDYSLSNESSLDEEGKGRVLSRLTELQAE 180
NLS
sp|Q9QLI9|PA_INBLE LSLKNLWQVLIGEEEIEKGIDFKLGQTISKLRNISVPAGFSNFEGMRSYIDNIDPKGAIE 240
sp|P13873|PA_INBAC LSLKNLWQVLIGEEDIEKGIDFKLGQTISKLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
sp|P13874|PA_INBAD LSLKNLWQVLIGEEDIEKGIDFKLGQTISKLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
sp|P11136|PA_INBSI LSLKNLWQVLIGEEDIEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|Q5V905|Q5V905_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|I0B7I8|I0B7I8_INBP9 LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|A0A077HHG5|A0A077HHG5_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|S4SYF4|S4SYF4_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|G7WTG6|G7WTG6_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNMDPKGAIE 240
tr|C0RVY2|C0RVY2_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|A0A059TAW4|A0A059TAW4_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|U3RKS7|U3RKS7_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|U3RK97|U3RK97_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|A4D5K3|A4D5K3_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|U3RQB4|U3RQB4_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|Q4W3D5|Q4W3D5_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|Q4W3C8|Q4W3C8_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|Q4W3C9|Q4W3C9_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 237
tr|I2DCX6|I2DCX6_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|A4D4G8|A4D4G8_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
tr|Q4W3D7|Q4W3D7_9INFB LSLKNLWQVLIGEEDVEKGIDFKLGQTISRLRDISVPAGFSNFEGMRSYIDNIDPKGAIE 240
sp|Q9QLI9|PA_INBLE YAKWATGDGLTYQKIMKEVAIDDETMYQEEPKIPNKCRVAAWVQAEMNLLSTLTSKRALD 420
sp|P13873|PA_INBAC YAKWATGDGLTYQKIMKEVAIDDETMYQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
sp|P13874|PA_INBAD YAKWATGDGLTYQKIMKEVAIDDETMYQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
sp|P11136|PA_INBSI YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|Q5V905|Q5V905_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|I0B7I8|I0B7I8_INBP9 YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|A0A077HHG5|A0A077HHG5_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|S4SYF4|S4SYF4_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|G7WTG6|G7WTG6_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|C0RVY2|C0RVY2_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|A0A059TAW4|A0A059TAW4_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|U3RKS7|U3RKS7_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|U3RK97|U3RK97_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|A4D5K3|A4D5K3_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|U3RQB4|U3RQB4_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|Q4W3D5|Q4W3D5_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|Q4W3C8|Q4W3C8_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|Q4W3C9|Q4W3C9_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 417
tr|I2DCX6|I2DCX6_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|A4D4G8|A4D4G8_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420
tr|Q4W3D7|Q4W3D7_9INFB YAKWATGDGLTYQKIMKEVAIDDETMCQEEPKIPNKCRVAAWVQTEMNLLSTLTSKRALD 420


sp|Q9QLI9|PA_INBLE ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFK 540
sp|P13873|PA_INBAC ITNRVVNEKGESFDMLHGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
sp|P13874|PA_INBAD ITNRVVNEKGESFDILYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
sp|P11136|PA_INBSI ITNRV-NEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 539
tr|Q5V905|Q5V905_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|I0B7I8|I0B7I8_INBP9 ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSGTDPRVDSGKWPKYTVFR 540
tr|A0A077HHG5|A0A077HHG5_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|S4SYF4|S4SYF4_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|G7WTG6|G7WTG6_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|C0RVY2|C0RVY2_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDPGKWPKYTVFR 540
tr|A0A059TAW4|A0A059TAW4_9INFB ITNRIVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|U3RKS7|U3RKS7_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|U3RK97|U3RK97_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|A4D5K3|A4D5K3_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|U3RQB4|U3RQB4_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|Q4W3D5|Q4W3D5_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|Q4W3C8|Q4W3C8_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|Q4W3C9|Q4W3C9_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 537
tr|I2DCX6|I2DCX6_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|A4D4G8|A4D4G8_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
tr|Q4W3D7|Q4W3D7_9INFB ITNRVVNEKGESFDMLYGLAVKGQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFR 540
sp|Q9QLI9|PA_INBLE KACFKGDRVNNPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
sp|P13873|PA_INBAC KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
sp|P13874|PA_INBAD KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
sp|P11136|PA_INBSI KACFKEDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 659
tr|Q5V905|Q5V905_9INFB KACFKGDRINSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|I0B7I8|I0B7I8_INBP9 KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|A0A077HHG5|A0A077HHG5_9INFB KACFKGDRVNSPKTFSVGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|S4SYF4|S4SYF4_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|G7WTG6|G7WTG6_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|C0RVY2|C0RVY2_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|A0A059TAW4|A0A059TAW4_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|U3RKS7|U3RKS7_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|U3RK97|U3RK97_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|A4D5K3|A4D5K3_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|U3RQB4|U3RQB4_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|Q4W3D5|Q4W3D5_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|Q4W3C8|Q4W3C8_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|Q4W3C9|Q4W3C9_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 657
tr|I2DCX6|I2DCX6_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|A4D4G8|A4D4G8_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
tr|Q4W3D7|Q4W3D7_9INFB KACFKGDRVNSPKTFSIGTQEGKLVKGSFGKALRVIFTKCLMHYVFGNAQLEGFSAESRR 660
sp|Q9QLI9|PA_INBLE LLLLIQALKDRKGPWVFDLEGMYFGVEECISNNPWVIQSAYWFNEWLGIEKEGSKVLESI 720
sp|P13873|PA_INBAC LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESI 720
sp|P13874|PA_INBAD LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESI 720
sp|P11136|PA_INBSI LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 719
tr|Q5V905|Q5V905_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|I0B7I8|I0B7I8_INBP9 LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|A0A077HHG5|A0A077HHG5_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|S4SYF4|S4SYF4_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSVYWFNEWLGFEKEGSKVLESV 720
tr|G7WTG6|G7WTG6_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSVYWFNEWLGFEKEGSKVLESV 720
tr|C0RVY2|C0RVY2_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|A0A059TAW4|A0A059TAW4_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|U3RKS7|U3RKS7_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|U3RK97|U3RK97_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|A4D5K3|A4D5K3_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|U3RQB4|U3RQB4_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|Q4W3D5|Q4W3D5_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|Q4W3C8|Q4W3C8_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEG------- 713
tr|Q4W3C9|Q4W3C9_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEG------- 710
tr|I2DCX6|I2DCX6_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|A4D4G8|A4D4G8_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLESV 720
tr|Q4W3D7|Q4W3D7_9INFB LLLLIQALKDRKGPWVFDLEGMYSGIEECISNNPWVIQSAYWFNEWLGFEKEGSKVLES- 719



```
//End of PA subunit sequences from influenza virus B
sp|Q9QLI9|PA_INBLE                      DEIMDE 726
sp|P13873|PA_INBAC                      DEIMDE 726
sp|P13874|PA_INBAD                      DEIMDE 726
sp|P11136|PA_INBSI                      DEIMDE 725
tr|Q5V905|Q5V905_9INFB                  DEIMDE 726
tr|I0B7I8|I0B7I8_INBP9                  DEIMNE 726
tr|A0A077HHG5|A0A077HHG5_9INFB          DEIMDE 726
tr|S4SYF4|S4SYF4_9INFB                  DEIMDE 726
tr|G7WTG6|G7WTG6_9INFB                  DEIMDE 726
tr|C0RVY2|C0RVY2_9INFB                  DEIMDE 726
tr|A0A059TAW4|A0A059TAW4_9INFB          DEIMDE 726
tr|U3RKS7|U3RKS7_9INFB                  DEIMDE 726
tr|U3RK97|U3RK97_9INFB                  DEIMDE 726
tr|A4D5K3|A4D5K3_9INFB                  DEIMDE 726
tr|U3RQB4|U3RQB4_9INFB                  DEIMDE 726
tr|Q4W3D5|Q4W3D5_9INFB                  ------ 720
tr|Q4W3C8|Q4W3C8_9INFB                  ------ 713
tr|Q4W3C9|Q4W3C9_9INFB                  ------ 710
tr|I2DCX6|I2DCX6_9INFB                  DEIMDE 726
tr|A4D4G8|A4D4G8_9INFB                  DEIMDE 726
tr|Q4W3D7|Q4W3D7_9INFB                  ------ 719
```

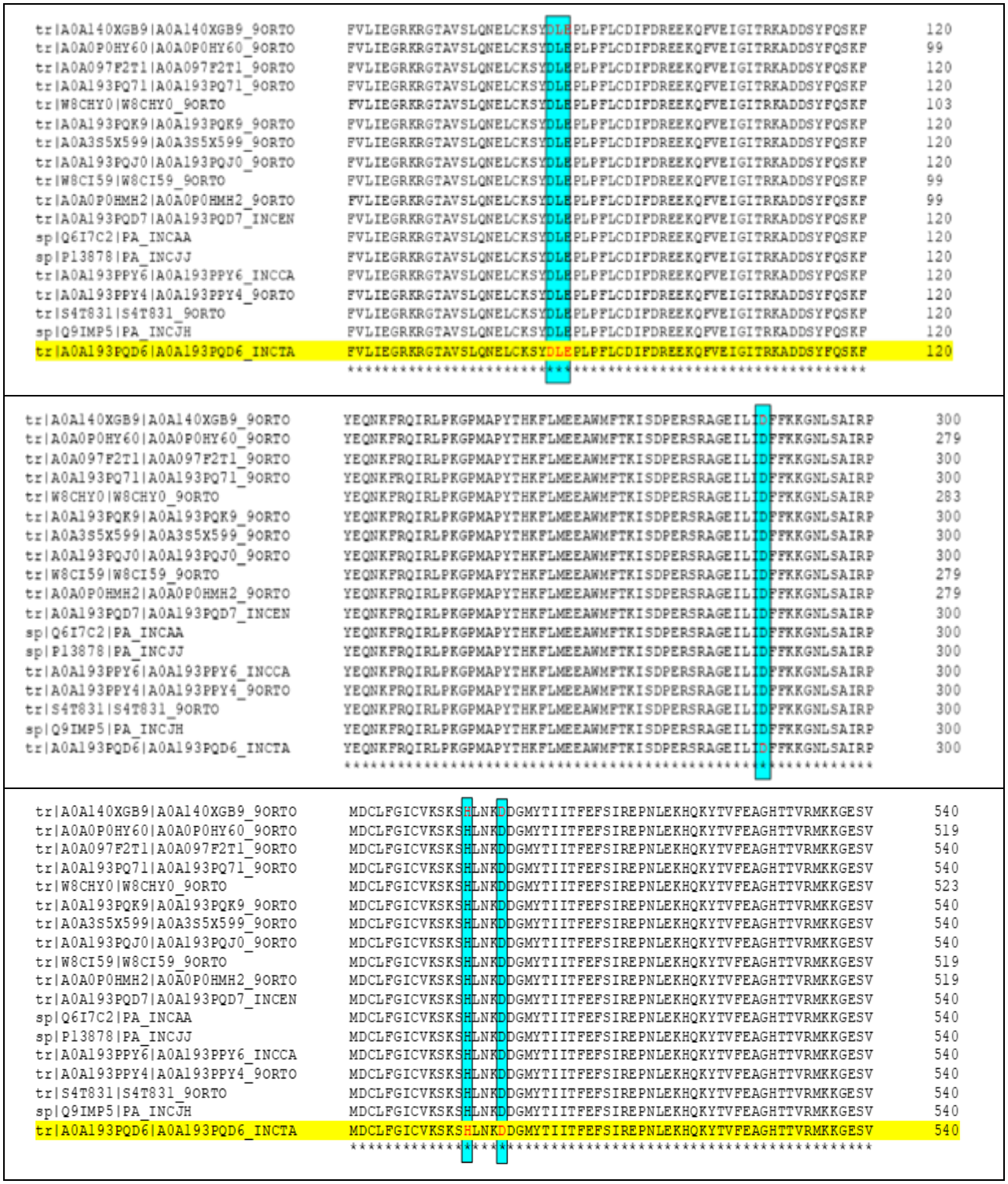
P13873|PA_INBAC, Influenza B virus (Ann Arbor/1966 [cold-adapted]
P13874|PA_INBAD, Influenza B virus (Ann Arbor/1966 [wild-type] (pI = 5.6)
Q9QLI9|PA_INBLE, Influenza B virus (Lee/1940) (pI = 5.5)
P11136|PA_INBSI, Influenza B virus (Singapore/1979)
U3RKS7_9INFB, Influenza B virus (Brisbane/2002)
U3RK97_9INFB, Influenza B virus (Waikato/2002)
A4D5K3_9INFB, Influenza B virus (Paraguay/2003)
U3RQB4_9INFB, Influenza B virus (Auckland/2003)
Q4W3D5_9INFB, Influenza B virus (Tehran/02)
Q4W3C8_9INFB, Influenza B virus (Geneva/2003)
Q4W3C9_9INFB, Influenza B virus (Barcelona/2003)
A0A059TAW4_9INFB, Influenza B virus (Thailand/2012)
I0B7I8_INBP9, Influenza B virus (Panama/1990)
S4SYF4_9INFB, Influenza B virus (Thailand/2011)
G7WTG6_9INFB, Influenza B virus (Managua/2010)
I2DCX6_9INFB, Influenza B virus (Malaysia/2000)
A4D4G8_9INFB, Influenza B virus (Mexico/2000)
Q4W3D7_9INFB, Influenza B virus (Sichuan/99)
Q5V905_9INFB, Influenza B virus (Nanchang/97)
C0RVY2_9INFB, Influenza B virus (Taiwan/2007)
A0A077HHG5_9INFB, Influenza B virus (Nicaragua/2008)

Figure 4 MSA of polymerase acidic proteins (PA) from different strains of influenza virus B

3.3. MSA Analysis of the Polymerase PA Subunit of Influenza Virus C

Figure 5 shows the MSA of the PA subunits of the RNA polymerase from the influenza virus C from different regions (only the required regions for the discussions are shown here). The influenza C viral strain, 1947, is highlighted in yellow and it showed a much higher theoretical pI of ~6.8. It is interesting to note that the theoretical pIs of the PA subunits in influenza A and B viruses are in the acidic range (~5.5), but in C virus, it is in the neutral range ~7.0. Except for small sequence variations in the N- and C-terminal regions, the PA subunits of different influenza C viral strains are almost completely conserved for the entire sequence as compared to influenza A virus. In influenza C virus also a typical DEDD-family of conserved amino acids (highlighted in light blue) with the pattern DxE---D-----**H**---D, is identified, as in influenza A and B viruses. However, the putative PHP-family PR exonuclease active site is not identified in the polymerase PA subunits of influenza C viral strains, which is in close agreement with B virus, but not with A virus. In other words, the PA subunit of the polymerase from influenza virus C possibly uses only DEDD(H)-family of PR exonuclease. A –DxD- type of metal-binding motif is also found (data not shown) and the C-terminal ends in D, similar to the B virus. The invariant penta-peptide found in all the human viruses is highlighted in light green.

CLUSTAL O (1.2.4) MSA of Acidic Protein subunit (PA) from different strains of influenza virus C


tr|A0A140XGB9|A0A140XGB9_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A0P0HY60|A0A0P0HY60_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 99
tr|A0A097F2T1|A0A097F2T1_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A193PQ71|A0A193PQ71_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|W8CHY0|W8CHY0_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 103
tr|A0A193PQK9|A0A193PQK9_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A3S5X599|A0A3S5X599_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A193PQJ0|A0A193PQJ0_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|W8CI59|W8CI59_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 99
tr|A0A0P0HMH2|A0A0P0HMH2_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 99
tr|A0A193PQD7|A0A193PQD7_INCEN FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
sp|Q6I7C2|PA_INCAA FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
sp|P13878|PA_INCJJ FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A193PPY6|A0A193PPY6_INCCA FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A193PPY4|A0A193PPY4_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|S4T831|S4T831_9ORTO FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
sp|Q9IMP5|PA_INCJH FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A193PQD6|A0A193PQD6_INCTA FVLIEGRKRGTAVSLQNELCKSYDLEPLPFLCDIFDREEKQFVEIGITRKADDSYFQSKF 120
tr|A0A140XGB9|A0A140XGB9_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A0P0HY60|A0A0P0HY60_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 279
tr|A0A097F2T1|A0A097F2T1_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A193PQ71|A0A193PQ71_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|W8CHY0|W8CHY0_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 283
tr|A0A193PQK9|A0A193PQK9_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A3S5X599|A0A3S5X599_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A193PQJ0|A0A193PQJ0_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|W8CI59|W8CI59_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 279
tr|A0A0P0HMH2|A0A0P0HMH2_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 279
tr|A0A193PQD7|A0A193PQD7_INCEN YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
sp|Q6I7C2|PA_INCAA YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
sp|P13878|PA_INCJJ YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A193PPY6|A0A193PPY6_INCCA YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A193PPY4|A0A193PPY4_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|S4T831|S4T831_9ORTO YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
sp|Q9IMP5|PA_INCJH YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A193PQD6|A0A193PQD6_INCTA YEQNKFRQIRLPKGPMAPYTHKFLMEEAWMFTKISDPERSRAGEILIDFFKKGNLSAIRP 300
tr|A0A140XGB9|A0A140XGB9_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|A0A0P0HY60|A0A0P0HY60_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 519
tr|A0A097F2T1|A0A097F2T1_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|A0A193PQ71|A0A193PQ71_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|W8CHY0|W8CHY0_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 523
tr|A0A193PQK9|A0A193PQK9_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|A0A3S5X599|A0A3S5X599_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|A0A193PQJ0|A0A193PQJ0_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|W8CI59|W8CI59_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 519
tr|A0A0P0HMH2|A0A0P0HMH2_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 519
tr|A0A193PQD7|A0A193PQD7_INCEN MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
sp|Q6I7C2|PA_INCAA MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
sp|P13878|PA_INCJJ MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|A0A193PPY6|A0A193PPY6_INCCA MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|A0A193PPY4|A0A193PPY4_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|S4T831|S4T831_9ORTO MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
sp|Q9IMP5|PA_INCJH MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540
tr|A0A193PQD6|A0A193PQD6_INCTA MDCLFGICVKSKSHLNKDDGMYTIITFEFSIREPNLEKHQKYTVFEAGHTTVRMKKGESV 540

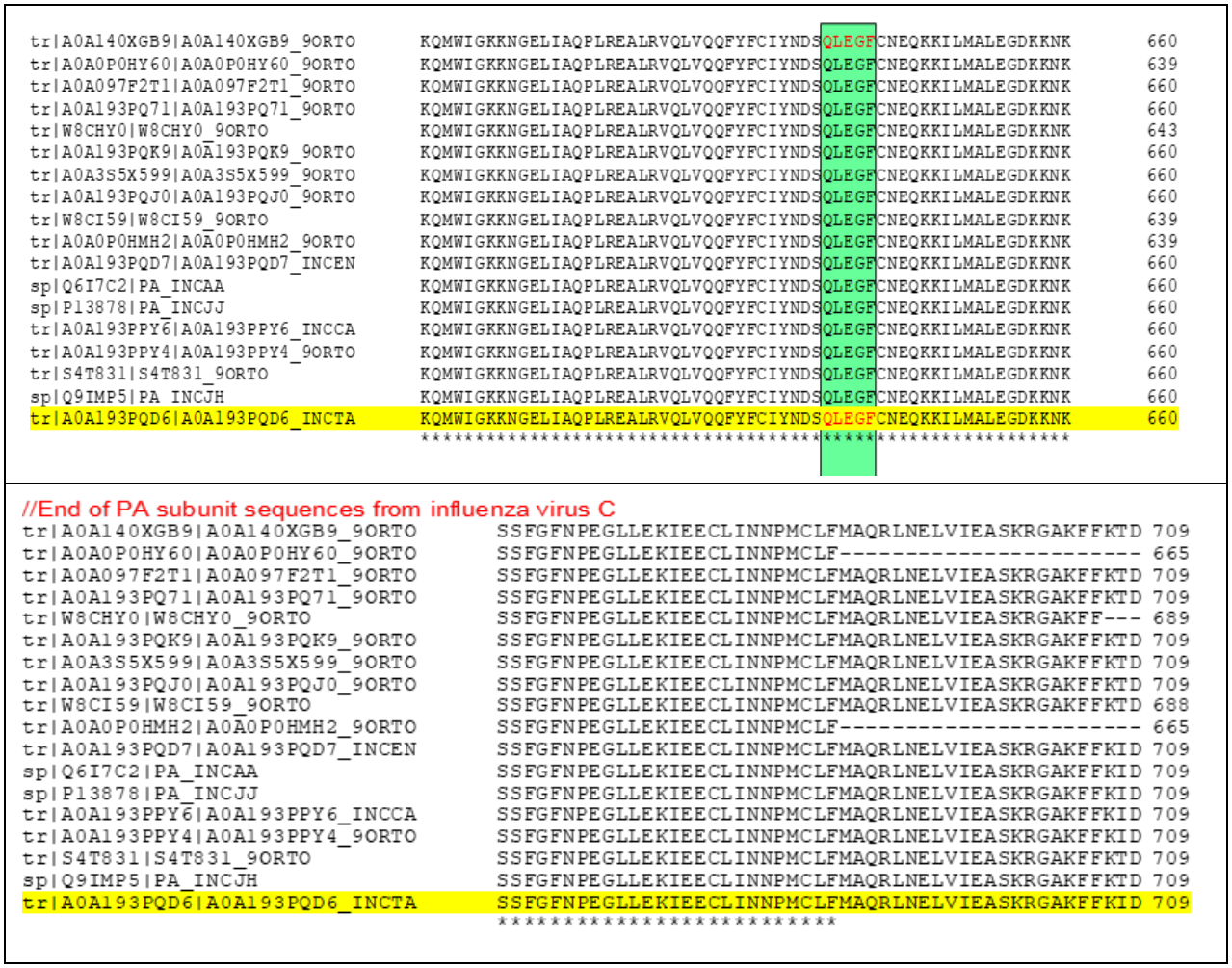


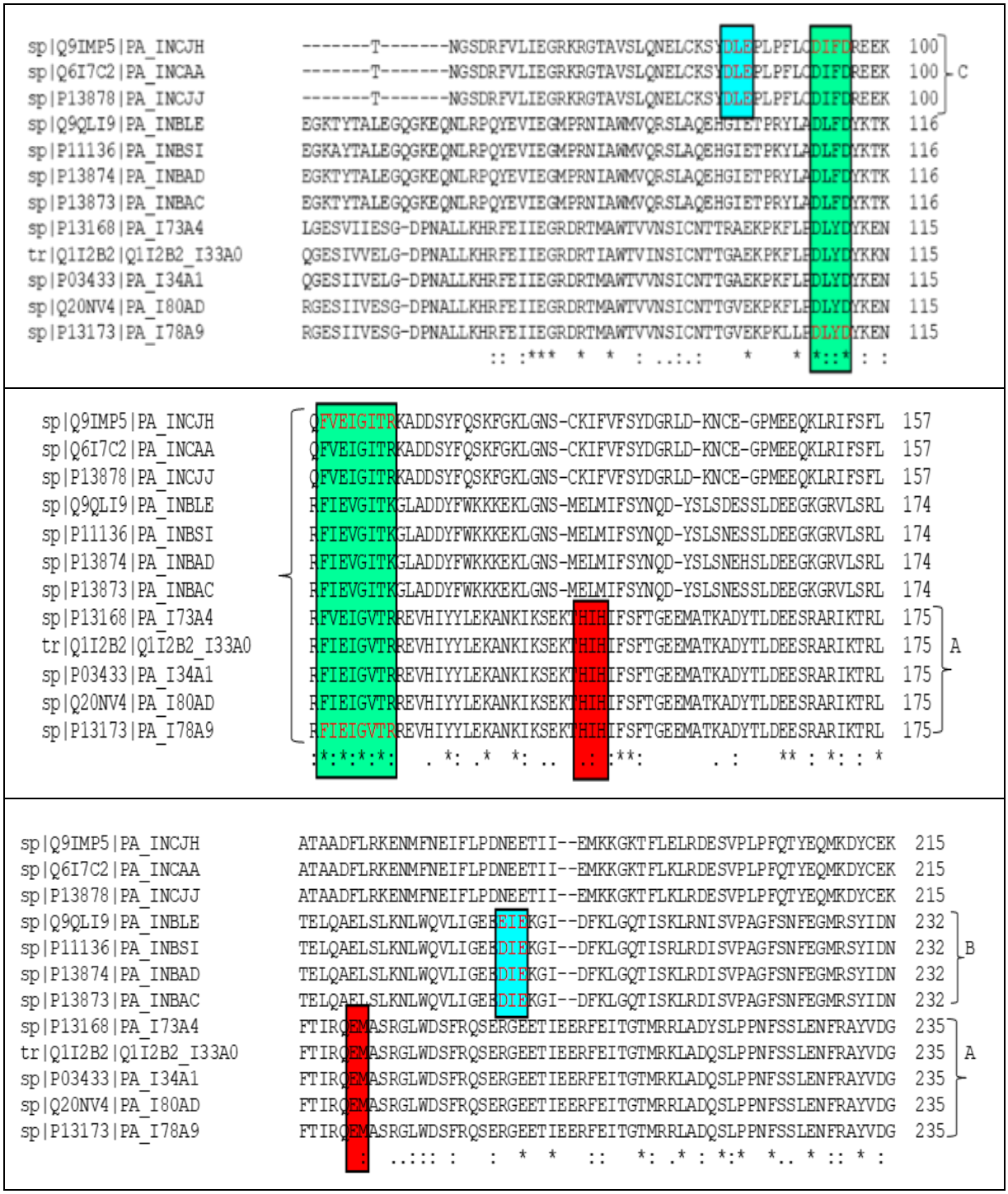
Figure 5 MSA of polymerase acidic proteins (PA) from different strains of influenza virus C

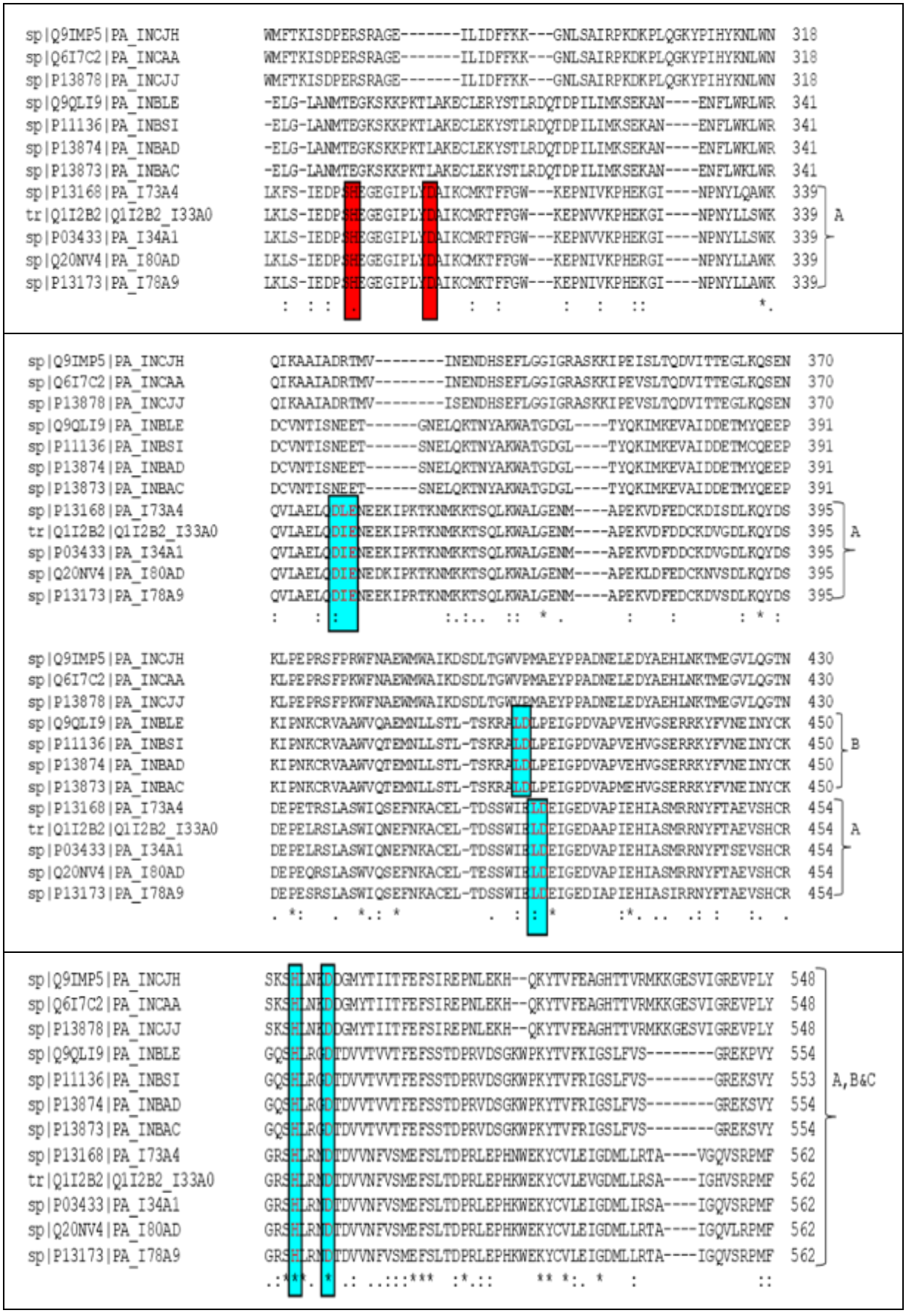
3.4. Mix and Match Analysis of the PA Subunits of Human Influenza Viruses A, B and C

Figure 6 shows the mix and match analysis of the PA subunits from all three human influenza viruses (only the required regions for the discussions are shown here). But for the last two invariant amino acids of the DED**D(H)**-superfamily, viz.

as -**H---D**-, the other conserved amino acid motifs are scattered. The PHP-exonuclease domain (highlighted in red) is found only in the influenza virus A. It is interesting to note that five peptide regions ranging from tetra- to octa-peptide are highly conserved in all three human influenza viruses (highlighted in light green). Interestingly, the penta-peptide -**QLEGF**- is completely conserved in all the three human influenza viruses, but the 3 of the downstream amino acids are different, but conserved for each virus (Figs. 2, 4, 5 and 6). It is interesting to note that the 'Leu-repeat' is found only in the pandemic/endemic causing influenza A and B viruses (Fig. 6).

CLUSTAL O (1.2.4) MSA of the PA subunits of the polymerases of influenza viruses A, B and C




sp|Q9IMP5|PA_INCJH WMFTKISDPERSRAGE-------ILIDFFKK---GNLSAIRPKDKPLQGKYPIHYKNLWN 318
sp|Q6I7C2|PA_INCAA WMFTKISDPERSRAGE-------ILIDFFKK---GNLSAIRPKDKPLQGKYPIHYKNLWN 318
sp|P13878|PA_INCJJ WMFTKISDPERSRAGE-------ILIDFFKK---GNLSAIRPKDKPLQGKYPIHYKNLWN 318
sp|Q9QLI9|PA_INBLE -ELG-LANMTEGKSKKPKTLAKECLERYSTLRDQTDPILIMKSEKAN----ENFLWRLWR 341
sp|P11136|PA_INBSI -ELG-LANMTEGKSKKPKTLAKECLEKYSTLRDQTDPILIMKSEKAN----ENFLWKLWR 341
sp|P13874|PA_INBAD -ELG-LANMTEGKSKKPKTLAKECLEKYSTLRDQTDPILIMKSEKAN----ENFLWKLWR 341
sp|P13873|PA_INBAC -ELG-LANMTEGKSKKPKTLAKECLEKYSTLRDQTDPILIMKSEKAN----ENFLWKLWR 341
sp|P13168|PA_I73A4 LKFS-IEDPSHEGEGIPLYDAIKCMKTFFGW---KEPNIVKPHEKGI----NPNYLQAWK 339
tr|Q1I2B2|Q1I2B2_I33A0 LKLS-IEDPSHEGEGIPLYDAIKCMRTFFGW---KEPNVVKPHEKGI----NPNYLLSWK 339
sp|P03433|PA_I34A1 LKLS-IEDPSHEGEGIPLYDAIKCMRTFFGW---KEPNVVKPHEKGI----NPNYLLSWK 339
sp|Q20NV4|PA_I80AD LKLS-IEDPSHEGEGIPLYDAIKCMKTFFGW---KEPNIVKPHERGI----NPNYLLAWK 339
sp|P13173|PA_I78A9 LKLS-IEDPSHEGEGIPLYDAIKCMKTFFGW---KEPNIVKPHEKGI----NPNYLLAWK 339
A
sp|Q9IMP5|PA_INCJH QIKAAIADRTMV--------INENDHSEFLGGIGRASKKIPEISLTQDVITTEGLKQSEN 370
sp|Q6I7C2|PA_INCAA QIKAAIADRTMV--------INENDHSEFLGGIGRASKKIPEVSLTQDVITTEGLKQSEN 370
sp|P13878|PA_INCJJ QIKAAIADRTMV--------ISENDHSEFLGGIGRASKKIPEVSLTQDVITTEGLKQSEN 370
sp|Q9QLI9|PA_INBLE DCVNTISNEET------GNELQKTNYAKWATGDGL----TYQKIMKEVAIDDETMYQEEP 391
sp|P11136|PA_INBSI DCVNTISNEET------SNELQKTNYAKWATGDGL----TYQKIMKEVAIDDETMCQEEP 391
sp|P13874|PA_INBAD DCVNTISNEET------SNELQKTNYAKWATGDGL----TYQKIMKEVAIDDETMYQEEP 391
sp|P13873|PA_INBAC DCVNTISNEET------SNELQKTNYAKWATGDGL----TYQKIMKEVAIDDETMYQEEP 391
sp|P13168|PA_I73A4 QVLAELQDLENEEKIPKTKNMKKTSQLKWALGENM----APEKVDFEDCKDISDLKQYDS 395
tr|Q1I2B2|Q1I2B2_I33A0 QVLAELQDIENEEKIPRTKNMKKTSQLKWALGENM----APEKVDFDDCKDVGDLKQYDS 395
sp|P03433|PA_I34A1 QVLAELQDIENEEKIPKTKNMKKTSQLKWALGENM----APEKVDFDDCKDVGDLKQYDS 395
sp|Q20NV4|PA_I80AD QVLAELQDIENEDKIPKTKNMKKTSQLKWALGENM----APEKLDFEDCKNVSDLKQYDS 395
sp|P13173|PA_I78A9 QVLAELQDIENEEKIPRTKNMKKTSQLKWALGENM----APEKVDFEDCKDVSDLKQYDS 395
A
sp|Q9IMP5|PA_INCJH KLPEPRSFPRWFNAEWMWAIKDSDLTGWVPMAEYPPADNELEDYAEHLNKTMEGVLQGTN 430
sp|Q6I7C2|PA_INCAA KLPEPRSFPKWFNAEWMWAIKDSDLTGWVPMAEYPPADNELEDYAEHLNKTMEGVLQGTN 430
sp|P13878|PA_INCJJ KLPEPRSFPKWFNAEWMWAIKDSDLTGWVPMAEYPPADNELEDYAEHLNKTMEGVLQGTN 430
sp|Q9QLI9|PA_INBLE KIPNKCRVAAWVQAEMNLLSTL-TSKRALDLPEIGPDVAPVEHVGSERRKYFVNEINYCK 450
sp|P11136|PA_INBSI KIPNKCRVAAWVQTEMNLLSTL-TSKRALDLPEIGPDVAPVEHVGSERRKYFVNEINYCK 450
sp|P13874|PA_INBAD KIPNKCRVAAWVQTEMNLLSTL-TSKRALDLPEIGPDVAPVEHVGSERRKYFVNEINYCK 450
sp|P13873|PA_INBAC KIPNKCRVAAWVQTEMNLLSTL-TSKRALDLPEIGPDVAPMEHVGSERRKYFVNEINYCK 450
sp|P13168|PA_I73A4 DEPETRSLASWIQSEFNKACEL-TDSSWIELDEIGEDVAPIEHIASMRRNYFTAEVSHCR 454
tr|Q1I2B2|Q1I2B2_I33A0 DEPELRSLASWIQNEFNKACEL-TDSSWIELDEIGEDAAPIEHIASMRRNYFTAEVSHCR 454
sp|P03433|PA_I34A1 DEPELRSLASWIQNEFNKACEL-TDSSWIELDEIGEDVAPIEHIASMRRNYFTSEVSHCR 454
sp|Q20NV4|PA_I80AD DEPEQRSLASWVQSEFNKACEL-TESSWIELDEIGEDVAPIEHIASMRRNYFTAEVSHCR 454
sp|P13173|PA_I78A9 DEPESRSLASWIQSEFNKACEL-TDSSWIELDEIGEDIAPIEHIASIRRNYFTAEVSHCR 454
B
sp|Q9IMP5|PA_INCJH SKSHLNEDDGMYTIITFEFSIREPNLEKH--QKYTVFEAGHTTVRMKKGESVIGREVPLY 548
sp|Q6I7C2|PA_INCAA SKSHLNEDDGMYTIITFEFSIREPNLEKH--QKYTVFEAGHTTVRMKKGESVIGREVPLY 548
sp|P13878|PA_INCJJ SKSHLNEDDGMYTIITFEFSIREPNLEKH--QKYTVFEAGHTTVRMKKGESVIGREVPLY 548
sp|Q9QLI9|PA_INBLE GQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFKIGSLFVS--------GREKPVY 554
sp|P11136|PA_INBSI GQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFRIGSLFVS--------GREKSVY 553
sp|P13874|PA_INBAD GQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFRIGSLFVS--------GREKSVY 554
sp|P13873|PA_INBAC GQSHLRGDTDVVTVVTFEFSSTDPRVDSGKWPKYTVFRIGSLFVS--------GREKSVY 554
sp|P13168|PA_I73A4 GRSHLRNDTDVVNFVSMEFSLTDPRLEPHNWEKYCVLEIGDMLLRTA----VGQVSRPMF 562
tr|Q1I2B2|Q1I2B2_I33A0 GRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKYCVLEVGDMLLRSA----IGHVSRPMF 562
sp|P03433|PA_I34A1 GRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKYCVLEIGDMLIRSA----IGQVSRPMF 562
sp|Q20NV4|PA_I80AD GRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKYCVLEIGDMLLRTA----IGQVLRPMF 562
sp|P13173|PA_I78A9 GRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKYCVLEIGDMLLRTA----IGQVSRPMF 562
A,B&C

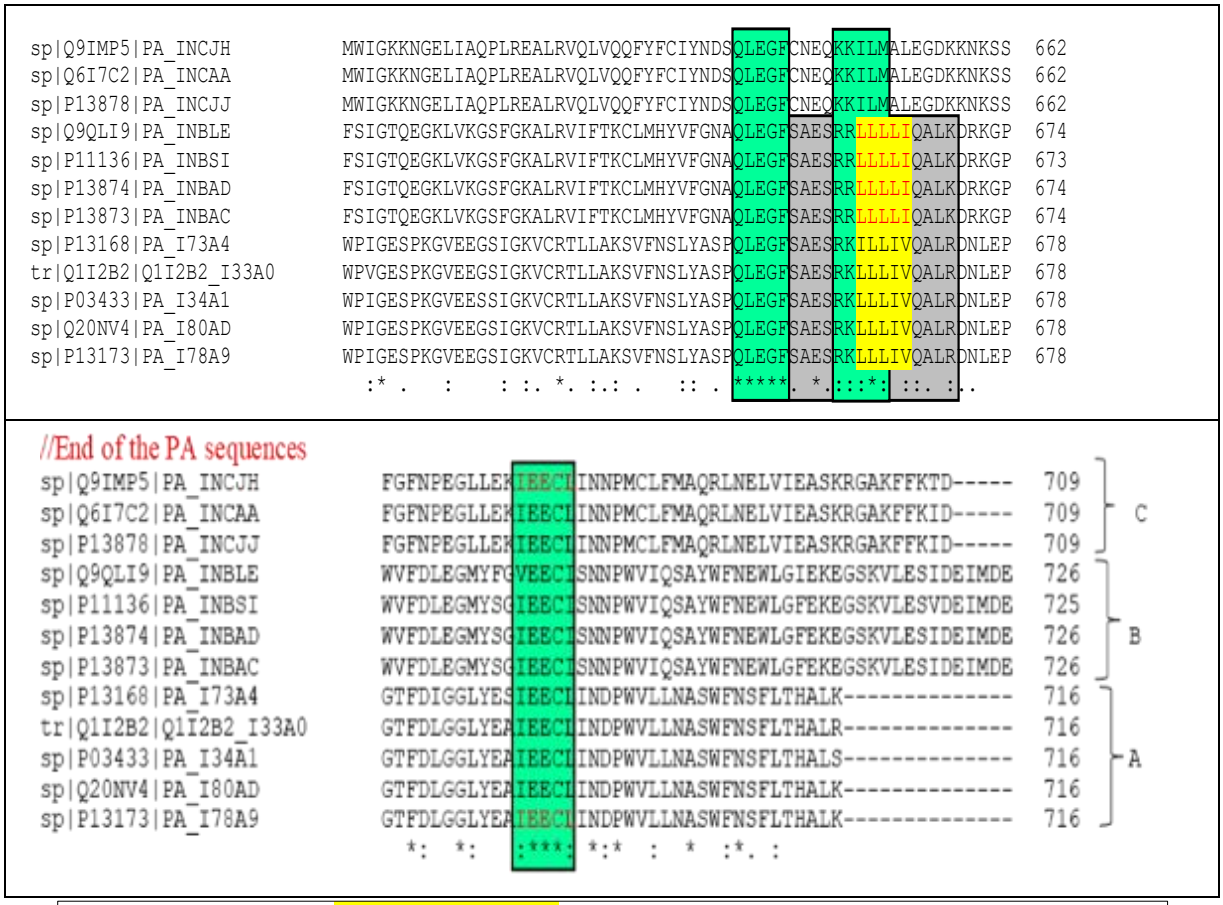


Figure 6 Mix and Match alignment of Influenza A, B and C viruses

3.5. Active Site Analyses of the PR Exonucleases in the PA subunit of Human Influenza Viral Polymerases

The active site amino acids are arrived at from sequence similarity, is in close agreement with the already well-established from DEDD- and PHP-superfamilies from other life forms. Furthermore, the involvement of the identified active site amino acids in the above two superfamilies of PR exonucleases has been confirmed by SDM and X-ray crystallographic techniques from different sources, as substantiated below.

3.5.1. Active Site Amino Acid Analyses of the DEDD(Y/H)-superfamily PR exonuclease

The PR exonuclease of DNA polymerase I (*E. coli*) belongs to -DEDD(Y)-family and the active site amino acids were confirmed both by SDM and X-ray crystallographic analyses by different investigators [19, 20, 21]. The PR exonuclease of DNA polymerase II (*E. coli*) also belongs to -DEDD(Y)-family and the active site amino acids were confirmed both by SDM and X-ray crystallographic analyses by Wang and Yang [22]. They have shown that the amino acids 147 to 367

comprise the 3'-5' PR exonuclease domain and 368 to 783 are involved in polymerase function. As the D^{335}→N mutant lost its exonuclease activity, suggesting its involvement in the catalysis. The active site amino acids were further corroborated by the following (D^{156}→N, D^{229}→N and D^{335}→N) exo$^-$ mutants [22].

The RNase D (EC 3.1.13.5), one of the seven exoribonucleases, which involves in the 3'-maturation of several stable RNAs like tRNA, 5S rRNA, and other small structured RNAs, was also shown to belong to the -DEDD(Y)-family [3].

Another interesting PR type 3'→5' exonuclease was found not in polymerases, but in the tRNA processing enzyme, RNase T (EC 3.1.13.-). In *E. coli*, the RNase T with at least five amino acid residues, viz. Asp^{23}, Glu^{25}, Asp^{125}, His^{181} and Asp^{186} are found in the active site. 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. [23] which was found to belong to DEDD(H)-family.

The DEDD(H)-family PR exonuclease in the ε-subunit of the bacterial replicase multienzyme complexes (DNA pols III) belongs to the DnaQ-H-family with the four active site carboxylates (Asp^{12}, Glu^{14}, Asp^{103}, and Asp^{167}) with the invariant His^{162}, which acts as the general base in catalysis [24]. Fijalkowska and Schaaper [24] also found that modification of the two conserved amino acid residues, viz. Asp^{12}→Ala and Glu^{14}→Ala, in the ε-subunit by SDM experiments resulted in the loss of the exonuclease function and hence, suggested playing a role in the coordination of the catalytic metal ion. These observations were further confirmed by X–ray crystallographic analysis of the ε-186 by Hamdan et al. [25] (Table 1). It is interesting to note that a similar DEDD(H)-superfamily is identified in the PA subunits of the RNA polymerases of all the three human influenza viruses (Table-1).

3.5.2. Active Site Amino Acid Analyses of PHP-superfamily PR exonuclease

The second type of PR exonuclease superfamily is the PHP-superfamily. The PHP-superfamily was found in pol III and pol X DNA polymerases and are mostly reported from the bacterial kingdom. They use a different initial triad (-HxH-) as compared to the DEDD family (-DxE-). The PHP-active site amino acids of DNA polymerases X from *Bacillus subtilis* have been extensively studied by Nagpal and Nair [4]. The PHP domain has been shown to possess the 3′–5′ exonuclease activity and performed the PR function in the DNA polymerase X [26]. Further, SDM experiments showed that the polymerase and PR activities were independent to each other. For example, in a double mutant where both the Ds are replaced by Ala (^{193}**D**L**D**→**A**L**A**), it abolished the polymerase activity [26], whereas in a similar experiment where both the Hs are replaced by Ala (^{339}**H**M**H**→**A**M**A**), it abolished the exonuclease activity. The PHP-PR exonuclease activity was further confirmed by analysis of deletion mutants too. For example, in a deletion mutant 316–570 (mutant ΔPHP), the exonuclease activity was completely abolished [26].

Another PHP-family was discovered in a phosphoesterase (YcdX) from *E. coli*. The X-ray crystallographic studies of *E. coli* YcdX have shown that the catalytic site of YcdX consisted of three Zn atoms and were very similar to those enzymes which hydrolyze phosphoester bonds [27].

Furthermore, a novel, intrinsic PHP-superfamily of PR exonuclease activity was reported in the α-catalytic subunit of DNA polymerase III itself from the thermophilic bacterium, *Thermus thermophilus* by Stano et al. [28]. They found that the PHP-type PR exonuclease domain was in the same polypeptide as the polymerase domain and was located in the N-terminal region of the polymerase as in influenza virus A (Tables 1 and 2). It should be noted that the PHP-superfamily not only involve in PR function but also in other error-correcting functions like BER and non-homologous end joining (NHEJ) to maintain the stability of prokaryotic and eukaryotic genomes [27 and references therein].

3.5.3. Active Site Amino Acids of DEDD(H)-Superfamily in SARS-CoVs

As in human influenza viruses, in SARS-CoVs the PR function is also performed by an associated subunit, known as ExoN, which is encoded by the non-structural protein 14 (NSP14). In the ExoN also, the PR function is performed by the DEDD-family of PR exonucleases [1]. Interestingly, the DEDD(H)-family's active site amino acids (highlighted in dark blue) in SARS-CoVs have been further confirmed by SDM experiments. For example, the MERS-CoV and SARS-CoV-2 ExoNs were analyzed by SDM experiments by Ogando et al. [29]. They subjected all five predicted active-site residues of the MERS-CoV's ExoN domain (D^{90}, E^{92}, E^{191}, H^{268} and D^{273}) by replacing them with Ala as well as with more conservative amino acids (D→E or Q; E→D or Q). This SDM experiment yielded a total of 14 ExoN active-site mutants, including the double mutant of D^{90}→A/E^{92}→A in motif I, which was frequently used as a prototypic viable ExoN knockout mutant in SARS-CoV-1 studies. They found the following SDM experiments of the MERS-CoV ExoN yielded non-viable phenotypes: D^{90}→A/Q/E; E^{92}→A/D/Q; E^{191}→A/Q; D^{273}→A/E/Q; H^{268}→A, suggesting their involvement in the catalysis of the ExoN PR exonuclease activity (Tables 1 and 2). Further, to analyze the replication and viability of MERS-CoV, the double mutant of D^{90}→A/E^{92}→A in motif I was used. Surprisingly, they could not rescue any viable progeny in which the two key

residues of the ExoN active site amino acids were mutated. Interestingly, the same set of amino acids that make the PR exonuclease domain of the ExoNs of SARS-CoV-1, MERS-CoV and SARS-CoV-2 is also found in human influenza viruses A, B and C (Fig. 7A, Tables 1 and 2).

The DEDD-family exonucleases also involve in PR function in eukaryotic organisms also. Recently the involvement of the DEDD-family in PR exonuclease in the eukaryotic replicase, DNA polymerase ε, from *Saccharomyces cerevisiae* has been reported. They found that the double mutant $D^{290}\rightarrow A/E^{292}\rightarrow A$ (from the invariant first triad –DxE- of the DEDD–family), was exonuclease deficient [30]. The replicative DNA polymerase ε exhibits both polymerase and exonuclease activities [31]. These results suggest that the DEDD-superfamily of PR exonucleases are ubiquitous and are found in the viral, bacterial, fungal, plant and animal kingdoms.

Table 1 shows the summary of the identified DEDD- and PHP-families' exonuclease active site amino acids from all the three influenza viruses, A, B and C and compared with the same families reported from different bacterial and other viral sources.

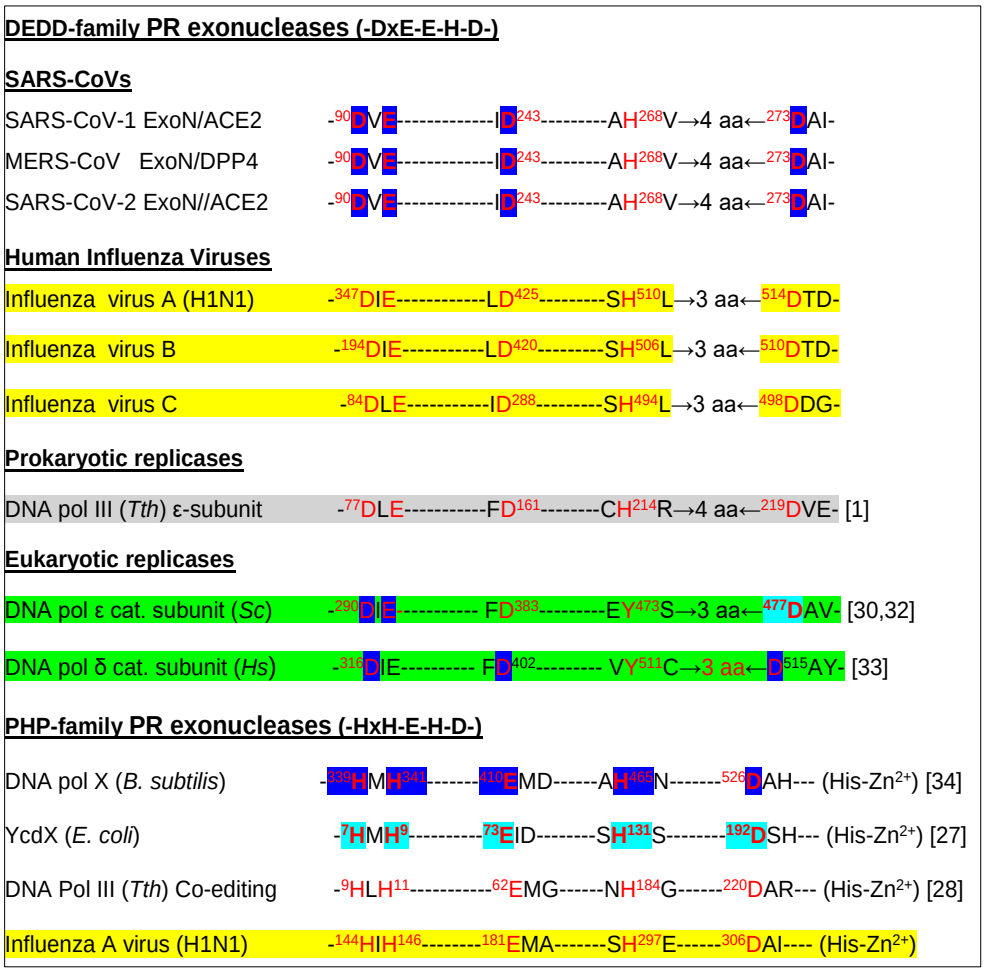
Table 1 Summary of the active site amino acids of DEDD- and PHP-superfamilies

Exo-Family Binding	Consensus A Site Pattern	Proton Acceptor	Catalytic Metal ion*	No. of Zn-binding site(s)
DEDD(Y/H)-family				
DNA Pol I	-D^{355}xE^{357}-D^{424}-Y^{497}-D^{501}-	Tyr	Zn^{2+}	1
DNA Pol II	-D^{156}xE-D^{229}-Y^{331}-D^{335}-	Tyr	Zn^{2+}	1
RNase D	-D^{28}xE^{30}-D^{85}-Y^{151}-D^{155}-	Tyr	Zn^{2+}	1
DNA Pol III, ε-subunit	-D^{12}xE^{14}-D^{103}-H^{162}-D^{167}-	His	Zn^{2+}	1
RNase T	-D^{23}xE^{25}-D^{125}-H^{181}-D^{186}-	His	Zn^{2+}	1
SARS-CoV-1 ExoN/ACE2^	-DxE-D-H-D-	His	Zn^{2+}	2
MERS-CoV ExoN/DPP4	-DxE-D-H-D-	His	Zn^{2+}	2
SARS-CoV-2 ExoN//ACE2	-DxE-D-H-D-	His	Zn^{2+}	2
Influenza virus A	-DxE-D-H-D-	His	Zn^{2+}	1
Influenza virus B	-DxE-D-H-D-	His	Zn^{2+}	1
Influenza virus C	-DxE-D-H-D-	His	Zn^{2+}	1
PHP-family				
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
Influenza virus A	-HxH-E-H-D-	His	Zn^{2+}	1

Adapted from Palanivelu [2021 [1] A Site, Active site; *Water-bound Zn^{2+} Active site amino acids, confirmed by SDM, are highlighted in dark blue. Influenza viral PR exonuclease active site amino acids, identified by sequence similarity, are highlighted in yellow. ^Similar active sites and structural features are found in other SARS-related CoVs and Human CoVs. *ACE2, Angiotensin-Converting Enzyme 2; DPP4, Dipeptidyl peptidase 4.*

Table 2 shows the DEDD- and PHP superfamily PR exonuclease active site amino acids and their distance conservations from the pandemic causing human viruses, and prokaryotic and eukaryotic sources.

Table 2 Active site amino acids of DEDD- and PHP-superfamilies and their distance conservation



Adapted from Palanivelu [2021 [1]

Active site amino acids, confirmed by SDM analysis, are highlighted in dark blue and from X-ray data are highlighted in light blue.

Influenza viral PR exonuclease active site amino acids, identified by sequence similarity, are highlighted in yellow.

The proposed active site structures are shown in Figs. 7A and 7B. The proposed PR catalytic mechanism involves proton abstraction by the active site His from the His-bound water molecule, which results in a highly reactive metal-hydroxide. The metal-hydroxide now attacks the electrophilic centre on the susceptible phosphate on the mismatched base, resulting in the cleavage of the phosphodiester bond and subsequently, the mismatched base is removed. The proton abstracted by the active site His in the first step is now transferred to 3'-Oxygen anion, generating 3'-OH at the growing end. Now the correct base is inserted and the polymerase resumes its reaction. In addition to the catalytic zinc, a second active site metal ion like Mg^{2+}/Zn^{2+} is also shown to be essential by SDM experiments for assisting the catalysis.

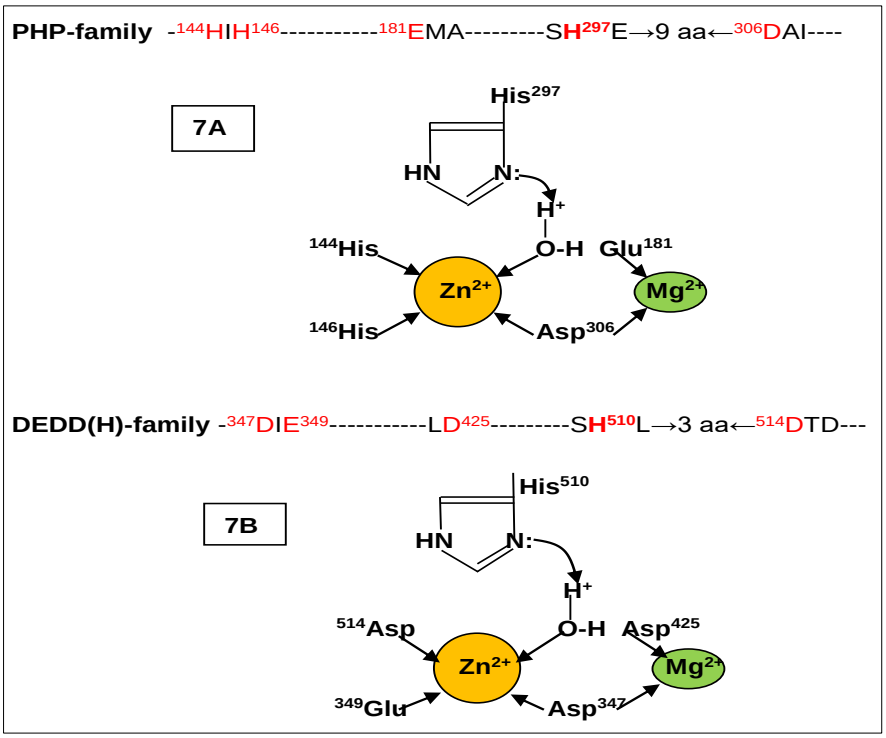


Figure 7A and 7B Proposed PR active site amino acids of the of human influenza A virus

(**Fig. 7A** DEDD(H)-superfamily and **Fig. 7B** PHP-superfamily; the amino acid numberings are from the PA subunit of the influenza A virus)

4. Conclusion

The vRNAs make mRNAs (for transcription) and cRNAs (for replication) using the same enzyme, viz. the viral RdRp. The cRNAs are perfect copies of the vRNAs from end to end and need an efficient PR machinery to make error-free viral genomes that are transferred to the progeny viruses. All the DNA and RNA replicases invariably habour PR exonuclease(s) to correct any mismatch during the replication processes. MSA analyses have shown that the PA subunit of the RNA polymerase from all the three human influenza viruses possess a typical PR exonuclease domain belonging to the DEDD-superfamily. However, the PA subunit of the pandemic causing influenza A virus possesses an additional putative PR exonuclease active site domain belonging to the PHP-superfamily. The PHP-superfamily's PR exonuclease domain is reported so far only in the bacterial kingdom. The identification of a putative PHP-family PR exonuclease in the human influenza A virus suggests its possible mobility from the bacterial kingdom to the viral kingdom during the evolutionary process. Furthermore, in bacterial kingdom PHP-family PR exonuclease domain is commonly found along with the DNA polymerase domain on the same polypeptide, but in the human influenza A virus it is found in the closely associated PA subunit, but for an RNA polymerase PR. These results will facilitate designing potential new anti-influenza drugs targeting the PA subunit of the viral RNA polymerase for treatments of flu in the future.

Compliance with ethical standards

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.

Disclosure of conflict of interest

The author has declared that no competing interests exist.

References

[1] Palanivelu P. An overview of the proofreading functions in bacteria and SARS-Coronaviruses. Int J Biochem Res Rev. 2021; 30: 33-62.

[2] Granoff A, Webster R (Editors) Encyclopedia of Virology (Second Edition) Elsevier, 1999, eBook ISBN: 9780080547978.

[3] Zuo Y, Deutscher MP. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res. 2001;29:1017–1026.

[4] 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.

[5] Timinskas K, Balvociute M, Timinskas A, Venclovas C. Comprehensive analysis of DNA polymerase III a subunits and their homologs in bacterial genomes. Nucleic Acids Res. 2014; 42: 1393–1413.

[6] Dangi T, Jain A. Influenza Virus: A Brief Overview. Proc. Natl. Acad. Sci. USA. 2012; 82:111–121. DOI 10.1007/s40011-011-0009-6.

[7] Su S, Fu X, Li G, Kerlin F, Veit M. Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics, Virulence. 2017; 8:1580-1591.

[8] Guo Y, Jin F, Wang P, Wang M, Zhu JM. Isolation of Influenza virus C from pigs and experimental infection of pigs with influenza C virus. J Gen Virol. 1983; 64:177-182.

[9] Palanivelu P. Identification of HNH Endonuclease Domain in the Basic Protein 2 Subunit of the Polymerase of Human Influenza Viruses, World J of Rec Res Rev. 2022; 15: 425–442.

[10] Pflug A, Guilligay D, Reich S, Cusack S. Structure of influenza A polymerase bound to the viral RNA promoter. Nature. 2014;516:355–360.

[11] Fan H, Walker AP, Carrique L, Keown JR, Martin IS, Karia D, Sharps J, Hengrung N, Pardon E, Steyaert J, Grimes JM, Fodor E. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. Nature. 2019; 573:287-295.

[12] Nakagawa, Y., K. Oda, and S. Nakada. 1996. The PB1 subunit alone can catalyze cRNA synthesis, and the PA subunit in addition to the PB1 subunit is required for viral RNA synthesis in replication of the influenza virus genome. J. Virol. 70:6390–6394.

[13] Honda, A., K. Mizumoto, and A. Ishihama. Minimum molecular architectures for transcription and replication of the influenza virus. Proc Natl Acad Sci. (USA). 2002; 99:13166-13171.

[14] Kawaguchi A, Naito T, and Nagata K. Involvement of Influenza Virus PA Subunit in Assembly of Functional RNA Polymerase Complexes. J Virol. 2005; 79: 732–744.

[15] Barros T, Joel Guenther1, Brian Kelch, Jordan Anaya, Arjun Prabhakar, Mike O'Donnell, John Kuriyan, Meindert H Lamers. A structural role for the PHP domain in E. coli DNA polymerase III. BMC Struct Biol. 2013; 13:8 www.biomedcentral.com/1472-6807/13/8.

[16] Krug RM, Ueda M, Palese P. Temperature-sensitive mutants of influenza WSN virus defective in virus- specific RNA synthesis. J virol. 1975; 64: 790-796.

[17] Mahy BWJ, Barret T, Nichol ST, Penn CR, Wolstenholme AJ. Analysiis of the functions of influenza virus genome RNA segments by use of temperature-sensitive mutants of fowl plague virus. In "Replication of negative strand viruses" (DHL Bishop and RW Compans, Eds). PP 379-387, Elsevier/North Holland, New York, 1981.

[18] Hao M, Han S, Meng D, Li R, Lin J, Wang M, Zhou T, Chai T. The PA Subunit of the Influenza Virus Polymerase Complex Affects Replication and Airborne Transmission of the H9N2 Subtype Avian Influenza Virus. Viruses. 2019; 11:40. https://doi.org/10.3390/v11010040.

[19] 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.

[20] 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.

[21] Joyce CM, Steitz TA. DNA polymerase I: from crystal structure to function via genetics. Trends Biochem Sci. 1987;12: 288-292.

[22] Wang F, Yang W. Structural insight into translesion synthesis by DNA Pol II. Cell. 2009;139:1279–1289.

[23] 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.

[24] Fijalkowska IJ, Schaaper RM. Mutants in the Exo I motif of Escherichia coli dnaQ: Defective proofreading and in viability due to error catastrophe. Proc Natl Acad Sci. (USA). 1996;93:2856-2861.

[25] 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.

[26] 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.

[27] 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.

[28] Stano NM, Chen J, McHenry CS. A coproofreading Zn2+-dependent exonuclease within a bacterial replicase. Nat Struct Mol Biol. 2006;13:458–459.

[29] Ogando NS, Zevenhoven-Dobbe JC, Meer YVD, Bredenbeek PJ, Posthuma CC, Snijder EJ. The Enzymatic Activity of the nsp14 Exoribonuclease Is Critical for Replication of MERS-CoV and SARS-CoV-2. J Virol. 2020; 94: e01246-20.

[30] Pinto MN, ter Beek J, Ekanger LA, Erik Johansson E, Barton JK. The [4Fe4S] Cluster of Yeast DNA Polymerase ε Is Redox Active and Can Undergo DNA-Mediated Signaling. J Am Chem Soc.: https://doi.org/10.1021/jacs.1c07150

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

[32] 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

[33] 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.

[34] Rodríguez, Martín MT, de Vega M. An array of basic residues is essential for the nucleolytic activity of the PHP domain of bacterial/archaeal PolX DNA polymerases. Sci Rep. 2019; 9:9947 | https://doi.org/10.1038/s41598-019-46349-8