

Sequence analysis and structural modelling of major capsid protein L1 of avian papillomavirus from African Grey Parrot

Zacharia Kadiayeno Egbunu^a, Yong Zi Yap^b, Nurulhuda Najihah^c, Abdul Razak Mariatulqabtiah^{a,c*}

^aFaculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

^bFaculty of Science, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Johor, Malaysia

^cInstitute of Bioscience, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

Received 11th October 2024 / Accepted 24th December 2024 / Published 31st December 2024

Abstract. Papillomaviruses are non-enveloped, and icosahedral in structure with a double-stranded circular DNA. They are responsible for inducing regressing papillomas (warts) on mucosal or keratinized epithelia across a diverse range of species including mammals, reptiles, birds and fish. Unlike human and bovine papillomaviruses, avian papillomaviruses (AvPV) received little attention in terms of sequence analysis and protein structure repository. This may be due to the less severity of morbidity and mortality compared to papillomavirus disease manifestations in human. The African Grey Parrot (*Psittacus erithacus*) is the first avian species to have a complete AvPV genome sequenced (PePV). Nonetheless, sequence analysis of its genes is limited with no three-dimensional structure reported in Protein Data Bank. Therefore, the aims of this study are to analyse the sequence of major capsid protein L1 of PePV, to assess its physicochemical properties, to generate its secondary and three-dimensional structures and to elucidate the quality of the generated L1 structural models. The PePV L1 capsid protein was analysed using online bioinformatics tools namely NCBI GenBank, PaVE, MUSCLE, ProtParam, PSIPRED and SOPMA. SWISS-MODEL, RaptorX and C-I-TASSER were implemented for high-quality structural modelling prior to comparison using PyMOL and Molprobit. Results demonstrated that the PePV L1 capsid protein was slightly acidic and thermally-stable. Additionally, the PePV's host specificity is closed related to *Fringilla coelebs* papillomavirus (FcPV1) and *Serinus canaria* papillomavirus (ScPV1), both of which infect birds from the Austravales clade. Structure predictions reveal slight structural difference and similarities albeit the SWIS-MODEL and C-I-TASSER showed relatively high-quality models which were considered as the basis for structural comparison and reliability. Further research on avian major capsid protein L1 of PePV is anticipated to improve the current knowledge on AvPV's structure-function relationship thus control the viral transmission in endangered birds.

Keywords: *Psittacus erithacus*, avian papillomavirus, L1 capsid protein, African Grey Parrot

INTRODUCTION

Papillomaviruses belonging to the family Papillomaviridae are small, non-enveloped, icosahedral, double-stranded circular DNA viruses approximately 60 nm in size, responsible for inducing regressing papillomas (warts) on mucosal or keratinized epithelia across a diverse range of species (Najihah *et al.*, 2023). An array of findings demonstrated their capacity to infect

many animals, including mammals, reptiles, fish, and birds (Van Doorslaer *et al.*, 2018), nonetheless, the great majority have been identified in mammals. The quantity of papillomaviruses identified in avian species is increasing, particularly due to the emergence of metagenomic studies (Truchado *et al.*, 2018), which has led to their detection from various sample types, including mucosal surfaces of the oral cavity, pharynx, choanal slit and cloaca which

* Author for correspondence: Abdul Razak Mariatulqabtiah, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
Email – mariatulqabtiah@upm.edu.my

are common in psittacine birds (Jones *et al.*, 2020). According to the International Committee on Taxonomy of Viruses (ICTV) classification, the virus is classified based on the host infected. Hence papillomavirus infecting humans and birds are referred to as human papillomavirus (HPV) and avian papillomavirus (AvPV), respectively. Most papillomaviruses infect humans, and more than 200 HPV have been extensively investigated because of their role, particularly in cervical cancer (Milano *et al.*, 2023); only about 112 animal papillomaviruses have been identified up to 2013 (D'arc *et al.*, 2020).

AvPV has been isolated from African Grey Parrot, *Psittacus erithacus* (Gaynor *et al.*, 2015; Canuti *et al.*, 2019). The African Grey Parrot is among the most intelligent and fascinating avian pets, indigenous to the rainforests of West and Central Africa, distinguished by its stunning grey feathers and vibrant red tail, and known for its beautiful appearance and remarkable cognitive abilities, problem-solving and communication skills. However, the bird was enlisted as an endangered species on the International Union for Conservation of Nature (IUCN) Red List (BirdLife International, 2021).. This may be due to wildlife trade (Atoussi *et al.*, 2020), habitat loss, and susceptibility to diseases such as AvPV (Greenacre, 2005) causing lesions, particularly in the cloacal and digestive tracts. A study from sampled captive psittacine birds in Seri Kembangan, Malaysia, reported a 33% incidence of AvPV but none for beak and feather disease virus and avian polyomavirus (Padzil *et al.*, 2021). This suggests that AvPV possesses a significant threat to psittacine birds including African Grey Parrot.

The L1 capsid protein of *Psittacus erithacus* AvPV (PePV) is a structural protein that gives the icosahedral architecture to papillomavirus; it possesses the ability to self-assemble independently into an immunogenic structure (virus-like particle, VLP) that mimics the true external structure of the virus even in the absence of chaperones, facilitating viral attachment and entry into host cells (Schiller & Lowy, 2012). The immunogenicity of this protein is well established due to its conserved gene which is also useful for classification and construction of the phylogenetic tree (Bernard *et al.*, 2010; Buck *et al.*, 2013,), demonstrating its potential to elicit high titre-

neutralizing immune responses important for vaccine development (Kirnbauer *et al.*, 1992; Lowy & Schiller, 2006). Nonetheless, appreciable information on the phylogeny and structure of the L1 and other PePV proteins is relatively limited despite being the first complete AvPV genome extracted from this species (Tachezy *et al.*, 2002). There are only 52 records of AvPV in public database UniProt (<https://www.uniprot.org/>), in which 9 of them are on L1 capsid protein, while none was found in the Protein Data Bank (PDB, <https://www.rcsb.org/>). As a comparison, bovine papillomavirus generated 1095 records in UniProt and 13 structure predictions on PDB as at 12 October, 2024.

Therefore, this study examines the L1 protein of the AvPV identified from African Grey Parrot, utilizing computational techniques to elucidate its phylogenetic relationships and structural characteristics. The research advances the scientific understanding of papillomaviruses originated from birds and aids conservation efforts for endangered species by addressing viral diseases that may affect their population.

MATERIALS AND METHODS

Sequence data information and retrieval

The whole genome nucleotide sequence of PePV was obtained from NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) with accession number NC_003973.1. The major capsid L1 protein sequence was also obtained from the same database with the accession number NP_647590.1.

Multiple sequence alignment

BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) search server was used to obtain similar protein sequences. The PaVE database (<https://pave.niaid.nih.gov/>) was also recruited to obtain other AvPV that were not listed in the BLASTp search results. The sequences were screened and selected to reduce the bias on gull papillomavirus and kittiwake papillomavirus. Sixteen protein sequences, including PePV1, were chosen for multiple sequence alignment by MUSCLE (Edgar, 2004) using sums-of-pair scores and k-tuples to calculate

the distances between sequences. A default parameter was used in the alignment, followed by manual adjustment. Duplicates were eliminated by computing the pairwise identity using a number of difference method, and the quality was assessed by “Pairwise amino acid identity” and “GUIDANCE2” with the amino acids substitution model and p-distance model parameters (Sela *et al.*, 2015). “Compute Overall Mean Distance” (COMD) was also employed (Blackburne & Whelan, 2012) to measure the average distance between sequences and assess divergence and identity and a confidence score was obtained from the aligned amino acids sequences.

Phylogenetic tree reconstruction

The phylogenetic tree of PePV with other papillomavirus species was reconstructed by utilizing the earlier multiple sequence alignment obtained. The phylogenetic analysis used 16 amino acid sequences of the L1 major capsid protein of papillomavirus with only complete sequences, which were extracted from NCBI GenBank. The maximum likelihood method (Felsenstein, 1981) and best fitting substitution model (Le & Gascuel, 2008; Zou *et al.*, 2024) were employed with an optimal evolutionary model and evaluated by the built-in model selection feature of MEGAX. For further evaluation of the reliability of the model, bootstraps of 100 replicates were performed.

Additionally, four non-AvPV sequences were extracted from the GenBank and included to further elucidate the evolutionary relatedness of AvPV and non-AvPV and the methods were repeated to reconstruct the phylogenetic tree using MEGAX.

Physicochemical properties and secondary structure prediction

The physicochemical property predictions of the L1 protein of PePV were analysed using an online web server, ProtParam (<https://web.expasy.org/protparam/>), to assess the protein function, stability and structure prediction. The secondary structure was predicted using PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) (McGuffin *et al.*, 2000; Buchan & Jones, 2019) and SOPMA (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma

[html](#)) to analyse the composition of the alpha helix, strands and coil (Sawal *et al.*, 2023).

Three-dimensional (3D) protein structure prediction and validation

The 3D prediction of the protein structure was modelled by three web servers, SWISS-MODEL (Waterhouse *et al.*, 2018), RaptorX (Wang *et al.*, 2016), and C-I-TASSER (Yang *et al.*, 2015). The sequence of the L1 major capsid protein of PePV was submitted separately to the tools, and the structures were predicted and compared for high-quality prediction. The models generated from the three servers were compared using PyMOL (DeLano, 2002) “Superimpose” function. The differences were analysed and validated by using Molprobit (Williams *et al.*, 2018) and “SAVES 6.0” (integrated server containing ERRAT, Verify3D, PROCHECK, WHATCHECK, and PROVE). The best structure was chosen based on the Ramachandran distribution of the residues in the most favoured and generally allowed region of the plot with consideration of individual ERRAT and VERIFY3D scores.

RESULTS AND DISCUSSION

Multiple sequence alignment

The BLAST nucleotide search reveals three results with 100% of query coverage and identity to PePV with accession numbers NC_003973.1, AF420235.1 and AF502599.1. The NC_003973.1 was chosen for protein sequence analysis (accession number NP_647590.1). To obtain similar protein sequences, BLASTp is performed with default parameter and arranged according to E value. The generated 100% query coverage and percentage of identity indicate same organization in the sequence and high similarities and relationship between the query and the sequences. Sixteen amino acid sequences were selected and aligned using MUSCLE method in MEGAX (Table 1). PePV-1 is of the genus Thetapapillomavirus, and the order of host is Psittaciformes. It shares 58.23 % identity with FIPV-1 from francolin, and 62.24 % identity with AplaPV-1 from duck. The assessment of the multiple sequence alignment revealed a reliable

alignment (Figure 1) with overall mean distance (p-distance) of 0.4. A p-distance less than 0.7 value reveals reliability of alignment (Shen *et al.*, 2022). GUIDANCE2 evaluated the multiple sequence alignment score as 96.0852. An alignment score more than 0.93 was considered acceptable (Penn *et al.*, 2010).

Phylogenetic tree reconstruction of L1 major capsid proteins of AvPV

This phylogenetic analysis uses 16 amino acid sequences of L1 major capsid proteins of PV with only the complete sequence, which was extracted from NCBI GenBank. Based on the maximum likelihood tree, there are 11 clades that separate the AvPV. In this phylogenetic analysis, PePV1 is

forming a single clade and is equally closely related to FcPV1 and ScPV1, which have 52.93% and 55.27% amino acid sequence identity, respectively. Both FcPV-1 and ScPV-1 form a monophyletic clade with a confidence of 100 bootstrap score. Cross-species transmission was also observed in the same order of host (Charadriiformes), such as LsmiPV-3 (Gull papillomavirus 3) is more closely related to RtriPV-1,-2,-3 (Kittiwake papillomavirus) than LsmiPV-1 and LmPV-2 (Gull papillomavirus 1 and 2). FgPV-1 position also revealed distant relations with other AvPV (Figure 2a) but closely related to reptile papillomavirus (CcPV-1 and CmpV-2) (Figure 2b).

Table 1. L1 amino acid sequence of avian papillomavirus extracted from NCBI Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>)

Type of Virus	Genus of Virus	Host	Order of host	Accession number
PePV-1	Thetapapillomavirus 1	<i>Psittacus erithacus timneh</i> (African Grey Parrot)	Psittaciformes	NP_647590.1
AplaPV-1	Unclassified	<i>Anas platyrhynchos</i> (Duck/Mallard)	Anseriformes	QBR99472.1
FcPV-1	Etapapillomavirus 1	<i>Fringilla coelebs</i>	Passeriformes	NP_663767.1
FgPV-1	Treiszetapapillomavirus	<i>Fulmarus glacialis</i>	Procellariiformes	YP_009041476.1
FIPV-1	Dyoepsilonpapillomavirus 1	<i>Francolinus leucoscepus</i>	Galliformes	YP_003104804.1
PaPV-1	Treisepsilonpapillomavirus 1	<i>Pygoscelis adeliae</i>	Sphenisciformes	YP_009022077.1
PaPV-2	Treisepsilonpapillomavirus	<i>Pygoscelis adeliae</i>	Sphenisciformes	ATL23484.1
ScPV-1	Etapapillomavirus	<i>Serinus canaria</i>	Passeriformes	YP_009551921.1
FarcPV-1	Unclassified	<i>Fratercula arctica</i> (Atlantic puffin)	Charadriiformes	QBR99465.1
LsmiPV-3 strain NL15_H1356	Unclassified	<i>Larus smithsonianus</i> (American Herring Gull)	Charadriiformes	QBR99520.1
LsmiPV-1 strain NL15_H1392	Unclassified	<i>Larus smithsonianus</i> (American Herring Gull)	Charadriiformes	QBR99501.1
LmPV-2 strain NL15_G1519	Unclassified	<i>Larus marinus</i> (Great Black-backed Gull)	Charadriiformes	QBR99511.1
RtriPV-6	Unclassified	<i>Rissa tridactyla</i> (Kittiwake)	Charadriiformes	QBR99529.1
RtriPV-2	Unclassified	<i>Rissa tridactyla</i> (Kittiwake)	Charadriiformes	QBR99486.1
RtriPV-3	Unclassified	<i>Rissa tridactyla</i> (Kittiwake)	Charadriiformes	QBR99512.1
RtriPV-1	Unclassified	<i>Rissa tridactyla</i> (Kittiwake)	Charadriiformes	QBR99521.1

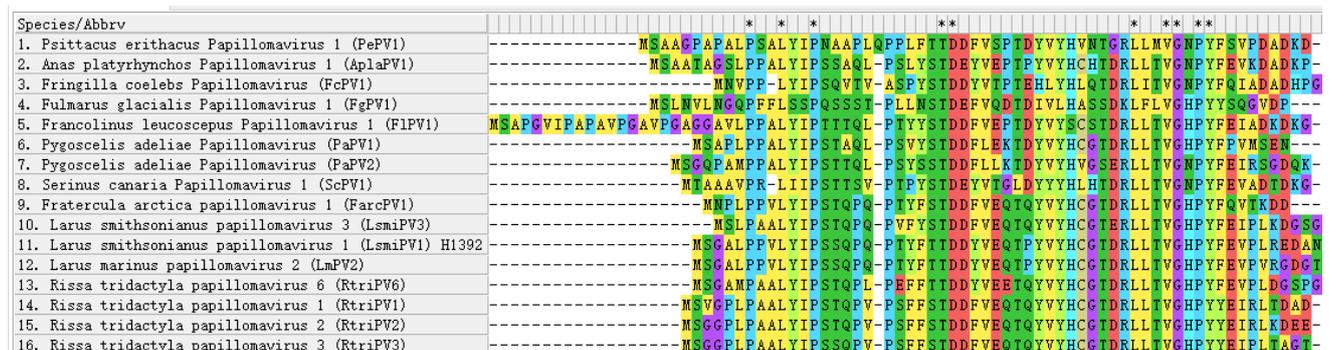


Figure 1. Multiple sequence alignment for 16 papillomaviruses using MUSCLE in MEGAX.

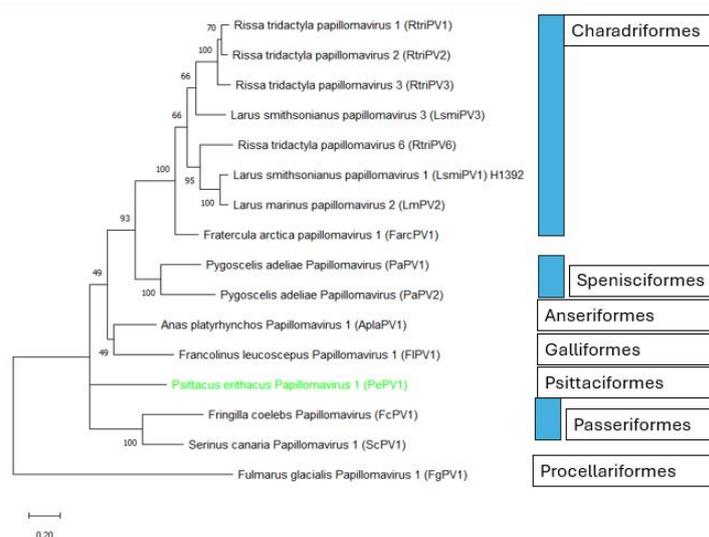


Figure 2a. Phylogenetic Analysis of PePV1 using L1 protein alignment of 16 AvPV with GenBank accession numbers in categories: Kittiwake *Rissa tridactyla* papillomavirus (RtriPV) 1 (QBR99521.1), RtriPV2 (QBR99486.1), (RtriPV) 3 (QBR99512.1), RtriPV6 (QBR99529.1), *Larus smithsonianus* (Lsmi) PV1 (QBR99501.1), LsmiPV3 (QBR99520.), LmPV2 (QBR99511.1), *Fratercula arctica* papillomavirus (FarcPV) 1 (QBR99465.1), *Pygoscelis adeliae* Papillomavirus (PaPV) 1 (YP_009022077.1), PaPV2 (ATL23484.1), *Anas platyrhynchos* PV1 (QBR99472.1), *Francolinus leucoscepus* PV1 (YP_003104804.1), PePV1 (NP_647590.1), *Fringilla coelebs* PV1 (NP_663767.1), *Serinus canaria* PV1 (YP_009551921.1), *Fulmarus glacialis* PV1 (YP_009041476.1).

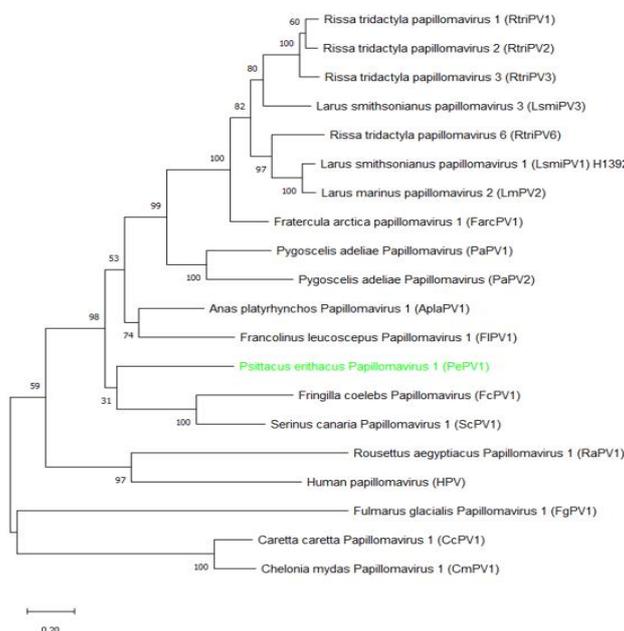


Figure 2b. Phylogenetic Analysis of PePV1 using L1 protein alignment of 16 AvPV and non-avian (mammalian and reptilian) papillomaviruses with GenBank accession numbers in categories; Kittiwake papillomavirus RtriPV1 (QBR99521.1), RtriPV2 (QBR99486.1), RtriPV3 (QBR99512.1), RtriPV6 (QBR99529.1), Gull papillomavirus LsmiPV1 (QBR99501.1), LsmiPV3 (QBR99520.), LmPV2 (QBR99511.1), FarcPV1 (QBR99465.1), PaPV1 (YP_009022077.1), PaPV2 (ATL23484.1), AplaPV1 (QBR99472.1), FIPV1 (YP_003104804.1), PePV1 (NP_647590.1), FcPV1 (NP_663767.1), ScPV1 (YP_009551921.1), *Fulmarus glacialis* papillomavirus FgPV1 (YP_009041476.1) and the non-avian RaPV1 (YP_717913.1), HPV (AYA94151.1), CcPV1 (YP_002308363.1) and CmPV1 (ACD39811.1).

Physicochemical properties and secondary structures

The ExPASy-ProtParam server and Psipred and SOPMA revealed that the physicochemical properties of the L1 protein of PePV is 57263.90 kg mol⁻¹ in molecular weight and 5.88 in theoretical pI (isoelectric point) (Table 2). The pI value is less than seven and this indicates the protein is slightly acidic in nature. The extinction coefficient values for L1 PePV at 280 nm range from 80790 to 81665 M⁻¹cm⁻¹. The smaller value is with the assumption of no cystine. The result of the primary analysis demonstrated that the L1 protein of PePV is slightly hydrophilic in nature with a negative value of GRAVY (-0.327) due to the higher content of non-polar residues (Chang & Yang, 2013). It also composed of 51.2% of non-polar residues and 48.9% of polar residues. The amino acid residue with the highest percentage in the protein was proline (9.6%), followed by alanine (8.5%) and leucine (8.3%), which indicates that it is slightly unstable in vitro

with an instability index of 40.06 against the cut-off value of 40. However, it is thermally stable due to the high content of aliphatic side chains with an aliphatic index of 73.35.

The secondary structure of L1 PePV predicted by PSIPRED server shows 14.62% of helix, 16.54% of beta sheet and 68.85% of coil. There are 8.08% disordered protein binding region and 2.69% disordered region in L1 PePV secondary structure. Similarly, in SOPMA secondary structure analysis, the highest component is coil with 52.5% followed by strand (22.50%) and alpha helix (19.23%) (Table 3). The existence of disordered regions of protein increased the difficulties in protein structure prediction, as they lack fixed structures and have flexible conformations during their native state (Forman-Kay & Mittag, 2013; Babu, 2016). However, this could be advantageous in protein binding and posttranslational modifications (Liu & Huang, 2014).

Table 2. Physicochemical properties of L1 PePV as predicted by ProtParam server.

Physicochemical Properties	Value
Molecular Weight	57263.90kg mol ⁻¹
Theoretical pI	5.88
Total number of negatively charged residues (Asp + Glu)	57
Total number of positively charged residues (Arg + Lys)	52
Extinction coefficient (assuming that all cysteine residues appear as half cystines)	81665 M ⁻¹ cm ⁻¹
Extinction coefficient (assuming that no cysteine appears as half cystine)	80790 M ⁻¹ cm ⁻¹
Instability index	40.06
Aliphatic index	73.35
Grand average of hydropathicity (GRAVY)	- 0.327

Table 3. Comparison of secondary analysis results from PSIPRED and SOPMA.

Secondary Structure	PSIPRED	SOPMA
Alpha Helix	76 (14.62%)	100 (19.23%)
Strand	86 (16.54%)	117 (22.50%)
Beta Turn	-	30 (5.77%)
Coil	358 (68.85%)	273 (52.5%)
Disordered protein binding	42 (8.08%)	-
Disordered	14 (2.69%)	-

Three-dimensional (3D) protein structure prediction

The three-dimensional structure prediction by the three prediction modelling tools, the SWISS homology model, RaptorX and C-I-TASSER, revealed relatively good quality L1 protein structures. SWISS-MODEL builds its model with a template 3iyjA with a sequence identity of 42.31% between the template and the query sequence, with a similarity of 0.42. This indicates the relationship between the template and query sequence is not very close. However, the structure is reliable as protein sequence identity with more than 35% indicates a similar protein structure between protein pairs homology (Krissinel, 2007). RaptorX generated 5 model structures and root-mean-square deviation (RMSD) scores, which range between 7.8485 to 9.1270. Consequently, to assess the quality of the prediction for the best structure, the least modelled structure with a smaller RMSD score was tentatively chosen, indicating high similarity for further validations (Reva *et al.*, 1998). However, larger proteins with more than 200 amino acid residues could have a bigger value of RMSD, which makes this indicator relatively unreliable (Sargsyan *et al.*, 2017). C-I-TASSER generates the top 10 predicted structures threading templates chosen according to their normalized Z-score (Norm. Z-score)

using the LOMETS2 threading programs such as SPARKS-K. A higher Z-score indicates better quality; the highest score in the top 10 templates is 23.60 (generated by 3iyjA) for chain A of L1 capsid protein of bovine papillomavirus type 1, which is similar to the SWISS-MODEL used template. The closest template with the highest score was 6bt3I (chain I of L1 capsid protein of HPV type 16).

The C-I-TASSER also provides the C-score of the top 5 best models for confidence level. C-score range between -5 to 2 (Zhang, 2008), giving credence to the structure quality for the templates 3iyjA and 6bt3I for bovine and human PV and relatively similar to our query sequence for PePV which was not listed in the PDB (Table 4). In addition, the TM-score of 0.89 ± 0.07 was obtained for our query sequence (TM ranged from 0 to 1), which also indicated the similarity of the structure with the native structure (template), which may be indirectly connected to its gene ontology and biological function, such as endocytosis and virion attachment for icosahedral capsid virus. This result supports the function of the L1 major capsid protein of PePV as helping the binding of the virus to the hosts, guiding the entry of the viral genome into the host nucleus (DiGiuseppe *et al.*, 2017; Van Doorslaer *et al.*, 2018).

Table 4. Top ten threading templates for L1 PePV protein structure prediction.

No	PDB hit	ID1	ID2	Cov	Norm. Zscore	C-score	Threading program
1	7kzff	0.43	0.40	11.62	4.86	1.25	SPARKS-K
2	3iyjF	0.41	0.36	10.65	5.14	0.25	FFAS-3D
3	predicted	0.42	0.38	11.13	23.60*	-1.03	HHpred
4	3iyjA	0.40	0.37	10.88	3.72	-2.76	MUSTER
5	6bt3I	0.44	0.39	11.32	9.49*	-3.17	CNFpred
6	3iyjA	0.42	0.38	11.24	3.70	ND	HHsearch-2
7	3iyjA	0.40	0.37	10.78	6.51	ND	Neff-PPAS
8	3iyjA	0.41	0.37	10.93	5.93	ND	HHsearch
9	1dzla	0.44	0.38	11.25	4.27	ND	PROSPECTOR2
10	3iyjA	0.40	0.36	10.61	8.64	ND	SAM

Key: ID1 and ID2 indicate identity between template and query sequence. Cov represents coverage. (*) represent closely related template structure to L1 PePV.

Structure evaluation and comparison of L1 major capsid of PePV models

The structural validation of the L1 major capsid of PePV predicted template's structure used by both SWISS-MODEL and C-I-TASSER was evaluated by web servers using SAVES 6.0, which consists of ERRAT, VERIFY3D, PROVE, PROCHECK and WHATCHECK. The results indicated relatively high quality and close agreement with the SWISS homology model and C-I-TASSER in this study with 85.8% and 70.5% of the residue distribution in the most favoured region of the Ramachandran plot, 85.75 and 82.49 ERRAT scores respectively; hence were considered as the basis for comparison and reliability (Table 5) but with slight structural differences and similarities (Figure 3). In the Molprobity evaluation, Chain A of the SWISS-MODEL revealed the best among the three models in terms of Molprobity score. The molprobity score of lower values and higher percentiles, indicates the high stereochemical quality of protein structure, combines all-atom contacts (clash score) and protein geometry (rotamers and Ramachandran) into a single score (Davis *et al.*, 2007).

This aspect of the study therefore establishes a connection between structural findings and biological implications of L1's role in host-virus interactions, particularly for its possible functions in virion attachment and immune evasion. A corresponding study on HPV have also focused on the L1 main capsid protein forming the icosahedral surface of the virion, that recombinant L1 proteins can autonomously self-assemble into highly immunogenic structures that closely resemble the natural surface of the native HPV virions giving a baseline for preventive vaccine targeting HPV.

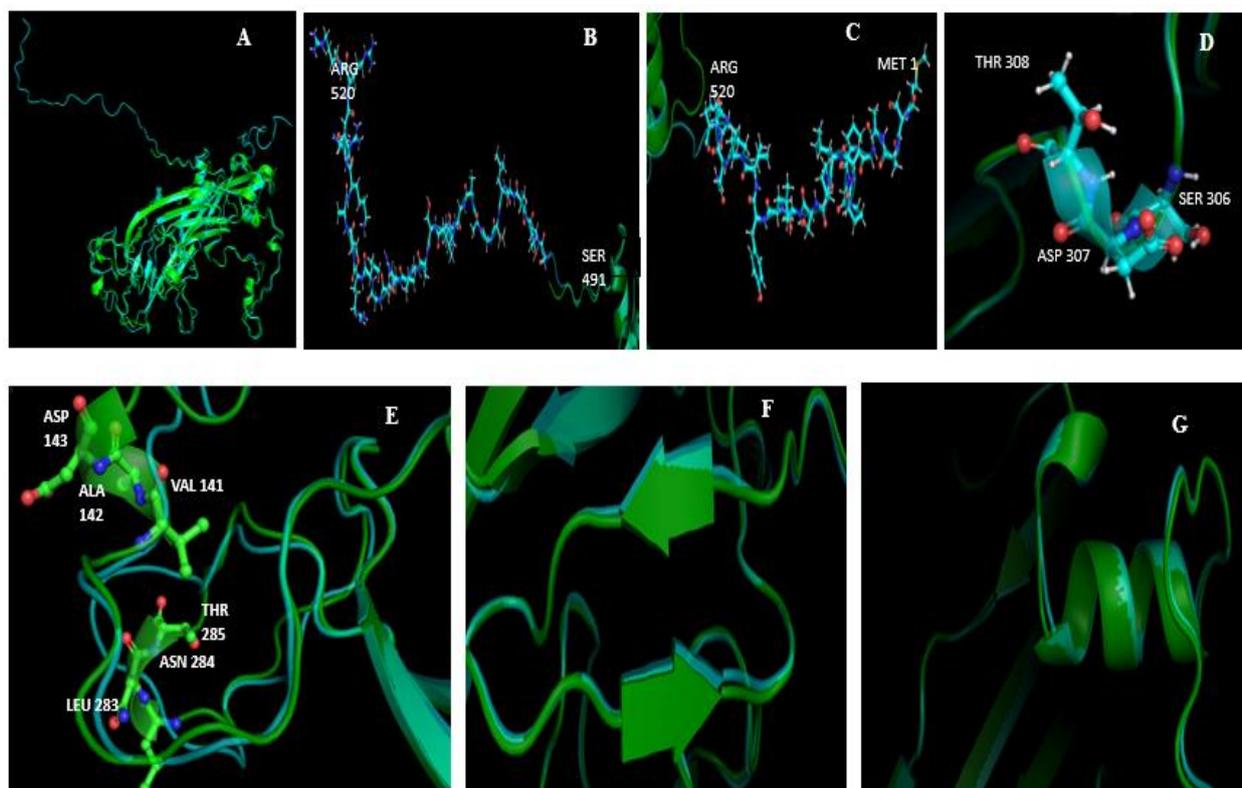
CONCLUSION

The evolutionary analysis of the L1 protein sequence suggests that virus-host codivergence plays a role in the pathophysiology of avian papillomavirus. PePV-1 is closely linked to FcPV-1 and ScPV-1, indicating that the virus infects

hosts that are phylogenetically similar, specifically Psittaciformes and Passeriformes, which belong to the same clade in avian phylogenetic studies. Nonetheless, the limited availability of AvPV genome sequences in the database may hinder the tree's ability to accurately represent the complete ecology and evolution of AvPV. The physicochemical properties of this protein examined by ProtParam, however, reveal slight instability *in vitro*, yet demonstrate thermal stability owing to a large concentration of aliphatic side chains. The secondary structure of this protein examined with both SOPMA and PSIPRED web servers indicated that coils are the predominant secondary structural configuration in the protein structure. This protein is anticipated to possess a greater number of coils, hence complicating protein structure prediction. Consequently, three web servers (SWISS-MODEL, RaptorX, C-I-TASSER) were selected to acquire the optimal and high-quality tertiary structure representing the L1 main capsid protein of PePV. As of now, no AvPV has been submitted to the PDB. The template adopted is bovine papillomavirus, identified as the optimal model examined by SWISS-MODEL. The percentage identity obtained by this tool indicated that homology modelling remains a viable method for predicting the protein structure of L1 PePV as it offers high-quality and accurate structural predictions, as the template is sourced from the PDB database. The models' quality by SAVES 6.0 and Molprobity, with SWISS-MODEL, also demonstrated superior performance in Ramachandran analysis, exhibiting fewer conflicts and higher residue accuracy compared to other servers. The final structure is often submitted with all parameters in SAVES 6.0 and Molprobity and regarded as the best protein model relative to others, demonstrating significant enhancement in Ramachandran analysis and ERRAT score. X-ray crystallography and nuclear magnetic resonance studies could be employed to validate the precision of this protein structure prediction. Conversely, C-I-TASSER effectively identifies the role of the L1 main capsid protein in the Gene Ontology term as facilitating the binding process and augmenting the entry of the viral DNA into the host nucleus.

Table 5. Stereochemical evaluation of template structure by SWISS-MODEL, C-I-TASSER, and RaptorX web servers.

Model	ERRAT (%)	VERIFY (%)	PROVE (%) of buried protein	Molprobitry (%)	Rama Favoured (%)
SWISS-MODEL					
Homo-hexamer	81.0551	68.01	0	90.4	85.6
Chain A	85.7494	74.21	8.8	91.30	85.8
C-I-TASSER Model					
Model 1	82.4945	63.08	7.9	77.41	70.5
RaptorX Models					
Model 1	76.6454	71.15	18.4	85.52	74.5

**Figure 3.** Computationally generated structures of the AvPV L1 gene. Superimpose of SWISS-MODEL Chain A (green colour) with C-I-TASSER (cyan colour) in cartoon view (A), the C-I-TASSER amino acid from 491 (SER) - 520 (ARG) that is different from SWISS-MODEL Chain A in ball and stick model (B), the different folding of loops from amino acid 1-21 of C-I-TASSER (C), comparison of secondary structure at SDT 306-307-308 amino acids from C-I-TASSER (helix) and SWISS-MODEL (loop) tools (D), comparison of secondary structures in VAD (141-142-143) and LNT (283-284-285), in which SWISS-MODEL represents in helices while C-I-TASSER represents in loops (E), and similar sheets (F) and helices (G) structures between SWISS-MODEL and C-I-TASSER.

Future studies using a more advanced structure prediction tool such as AlphaFold which is often at near-experimental resolution, especially for single-domain proteins like VLPs, could further investigate the function of PePV L1 protein. Additionally, *in vitro* experiments including crystallographic, cryo-EM or bioassays can be done to empirically validate its structural accuracy and biological characteristics for a more robust analysis. Therefore, further research on avian major capsid protein L1 of PePV is anticipated to improve the current knowledge on AvPV's structure-function relationship thus control the viral transmission in endangered birds.

ACKNOWLEDGEMENTS

This study was supported by the Ministry of Higher Education Malaysia (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2018/STG05/UPM/02/31) and Geran Inisiatif Putra Siswazah (GP-IPS/2023/9767400) from the Universiti Putra Malaysia (UPM). We thank TETFUND from the Federal Government of Nigeria and MOHE for financially supporting Z. K. E. and N. N. studentships, respectively.

CONFLICT OF INTEREST

The authors have declared that no conflict of interest exists.

REFERENCES

- Atoussi, S., Bergin, D., Razkallah, I., Nijman, V., Bara, M., Bouslama, Z., & Houhamdi, M. 2020. The trade in the endangered African Grey Parrot *Psittacus erithacus* and the Timneh Parrot *Psittacus timneh* in Algeria. *Ostrich* 91(3): 214–220. <https://doi.org/10.2989/00306525.2020.1763492>
- Babu, M. M. 2016. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions* 44(5): 1185–1200. <https://doi.org/10.1042/BST20160172>
- Bernard, H. U., Burk, R. D., Chen, Z., van Doorslaer, K., zur Hausen, H., de Villiers, E. M. 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 25;401(1):70-9. doi: 10.1016/j.virol.2010.02.002.
- BirdLife International. 2021. *Psittacus erithacus*. *The IUCN Red List of Threatened Species* 2021: e.T22724813A154428817. <https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22724813A154428817.en>. Accessed on 08 November 2024.
- Blackburne, B. P., & Whelan, S. 2012. Measuring the distance between multiple sequence alignments. *Bioinformatics* 28(4): 495–502. <https://doi.org/10.1093/bioinformatics/btr701>
- Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research* 47(W1): W402–W407. <https://doi.org/10.1093/nar/gkz297>
- Buck, C. B., Day, P. M., & Trus, B. L. 2013. The papillomavirus major capsid protein L1. *Virology* 445(1–2): 169–174. <https://doi.org/10.1016/j.virol.2013.05.038>
- Canuti, M., Munro, H. J., Robertson, G. J., Kroyer, A. N. K., Roul, S., Ojic, D., Whitney, H. G., & Lang, A. S. 2019. New insight into avian papillomavirus ecology and evolution from characterization of novel wild bird papillomaviruses. *Frontiers in Microbiology*, 10: 701. <https://doi.org/10.3389/fmicb.2019.00701>
- Chang, K. Y., Yang, J. R. 2013. Analysis and Prediction of Highly Effective Antiviral Peptides Based on Random Forests. *PLoS ONE* 8(8): e70166. <https://doi.org/10.1371/journal.pone.0070166>
- D'arc, M., Moreira, F. R. R., Dias, C. A., Souza, A. R., Seuánez, H. N., Soares, M. A., Tavares, M. C. H., & Santos, A. F. A. 2020. The characterization of two novel neotropical primate papillomaviruses supports the ancient within-species diversity model. *Virus Evolution* 6(1): veaa036. <https://doi.org/10.1093/ve/veaa036>
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., & Richardson, D. C. 2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35(Web Server): W375–W383. <https://doi.org/10.1093/nar/gkm216>
- De Villiers, E.-M., Fauquet, C., Broker, T. R., Bernard, H.-U., & Zur Hausen, H. 2004. Classification of papillomaviruses. *Virology* 324(1): 17–27. <https://doi.org/10.1016/j.virol.2004.03.033>
- DeLano, W. L. 2002. PyMOL: An open-source molecular graphics tool. *Newsletter on Protein Crystallography* 40(1): 82–92.
- DiGiuseppe, S., Bienkowska-Haba, M., Guion, L. G. M., Keiffer, T. R., & Sapp, M. 2017. Human papillomavirus major capsid protein L1 remains associated with the incoming viral genome throughout the entry process. *Journal of Virology* 91(16): e00537-17. <https://doi.org/10.1128/JVI.00537-17>
- Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5): 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17(6): 368–376. <https://doi.org/10.1007/BF01734359>
- Forman-Kay, J. D., & Mittag, T. 2013. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* 21(9): 1492–1499. <https://doi.org/10.1016/j.str.2013.08.001>
- Gaynor, A. M., Fish, S., Duerr, R. S., Cruz, F. N. D., & Pesavento, P. A. 2015. Identification of a novel papillomavirus in a northern fulmar (*Fulmarus glacialis*) with viral production in cartilage. *Veterinary Pathology* 52(3): 553–561. <https://doi.org/10.1177/0300985814542812>
- Greenacre, C. B. 2005. Viral diseases of companion birds. *Veterinary Clinics of North America: Exotic Animal Practice* 8(1): 85–105. <https://doi.org/10.1016/j.cvex.2004.09.005>
- Jones, A. L., Suárez-Bonnet, A., Mitchell, J. A., Ramirez, G. A., Stidworthy, M. F., & Priestnall, S. L. 2020. Avian papilloma and squamous cell carcinoma: A histopathological,

- immunohistochemical and virological study. *Journal of Comparative Pathology* 175: 13–23. <https://doi.org/10.1016/j.jcpa.2019.11.007>
- Kirnbauer, R., Booy, F., Cheng, N., Lowy, D. R., & Schiller, J. T. 1992. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. *Proceedings of the National Academy of Sciences*, 89(24), 12180–12184.
- Krissinel, E. 2007. On the relationship between sequence and structure similarities in proteomics. *Bioinformatics* 23(6): 717–723. <https://doi.org/10.1093/bioinformatics/btm006>
- Le, S., & Gascuel, O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* 25(7): 1307–1320. <https://doi.org/10.1093/molbev/msn067>
- Liu, Z., & Huang, Y. 2014. Advantages of proteins being disordered. *Protein Science* 23(5): 539–550. <https://doi.org/10.1002/pro.2443>
- Lowy, D. R & Schiller, J. T. 2006. Prophylactic human papillomavirus vaccines. *J Clin Invest*. 116(5):1167-73. doi: 10.1172/JCI28607
- McGuffin, L. J., Bryson, K., & Jones, D. T. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16(4): 404–405. <https://doi.org/10.1093/bioinformatics/16.4.404>
- Milano, G., Guarducci, G., Nante, N., Montomoli, E., & Manini, I. 2023. Human papillomavirus epidemiology and prevention: Is there still a gender gap? *Vaccines* 11(6): 1060. <https://doi.org/10.3390/vaccines11061060>
- Najihah, N., Nurul Najian, A. B., Syahir, A., Abu, J., Ho, K. L., Siang Tan, W., & Mariatulqabtiyah, A. R. 2023. Evaluation of avian papillomavirus occurrences and effective sampling materials for screening purposes in bird species through systematic review and meta-analysis. *Pertanika Journal of Tropical Agricultural Science* 46(2): 671–685. <https://doi.org/10.47836/pjtas.46.2.17>
- Padzil, M. F. M., Halim, N. S. A., Najihah, N., Nurul Najian, A. B., Abu, J., Isa, N. M., Lau, H. Y., & Mariatulqabtiyah, A. R. 2021. Evaluation of beak and feather disease virus, avian polyomavirus and avian papillomavirus of captives psittacine birds in Seri Kembangan, Selangor, Malaysia. *Malaysian Journal of Microbiology* 17(3): 338-344. <http://dx.doi.org/10.21161/mjm.201062>
- Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., & Pupko, T. 2010. GUIDANCE: A web server for assessing alignment confidence scores. *Nucleic Acids Research* 38(Web Server): W23–W28. <https://doi.org/10.1093/nar/gkq443>
- Reva, B. A., Finkelstein, A. V., & Skolnick, J. 1998. What is the probability of a chance prediction of a protein structure with an RMSD of 6 Å? *Folding and Design* 3(2): 141–147. [https://doi.org/10.1016/S1359-0278\(98\)00019-4](https://doi.org/10.1016/S1359-0278(98)00019-4)
- Sargsyan, K., Grauffel, C., & Lim, C. 2017. How molecular size impacts RMSD applications in molecular dynamics simulations. *Journal of Chemical Theory and Computation* 13(4): 1518–1524. <https://doi.org/10.1021/acs.jctc.7b00028>
- Sawal, H. A., Nighat, S., Safdar, T., & Anees, L. 2023. Comparative in silico analysis and functional characterization of TANK-binding kinase 1-binding protein 1. *Bioinformatics and Biology Insights* 17: 117793222311648. <https://doi.org/10.1177/11779322231164828>
- Schiller, J. T. & Lowy, D. R. 2012. Understanding and learning from the success of prophylactic human papillomavirus vaccines. *Nat. Rev. Microbiol.*; 10(10): 681-92. doi: 10.1038/nrmicro2872
- Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. 2015. GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research* 43(W1): W7–W14. <https://doi.org/10.1093/nar/gkv318>
- Shen, C., Zaharias, P., & Warnow, T. 2022. MAGUS+eHMMs: Improved multiple sequence alignment accuracy for fragmentary sequences. *Bioinformatics* 38(4): 918–924. <https://doi.org/10.1093/bioinformatics/btab788>
- Tachezy, R., Rector, A., Havelkova, M., Wollants, E., Fiten, P., Opendakker, G., Jenson, B., Sundberg, J., & Van Ranst, M. 2002. Avian papillomaviruses: The parrot *Psittacus erithacus* papillomavirus (PePV) genome has a unique organization of the early protein region and is phylogenetically related to the chaffinch papillomavirus. *BMC Microbiology* 2: 19. <https://doi.org/10.1186/1471-2180-2-19>
- Truchado, D. A., Williams, R. A. J., & Benítez, L. 2018. Natural history of avian papillomaviruses. *Virus Research* 252: 58–67. <https://doi.org/10.1016/j.virusres.2018.05.014>
- Van Doorslaer, K., Chen, Z., Bernard, H.-U., Chan, P. K. S., DeSalle, R., Dillner, J., Forslund, O., Haga, T., McBride, A. A., Villa, L. L., Burk, R. D., & ICTV Report Consortium. 2018. ICTV virus taxonomy profile: Papillomaviridae. *Journal of General Virology* 99(8): 989–990. <https://doi.org/10.1099/jgv.0.001105>
- Wang, S., Li, W., Liu, S., & Xu, J. 2016. RaptorX-Property: A web server for protein structure property prediction. *Nucleic Acids Research* 44(W1): W430–W435. <https://doi.org/10.1093/nar/gkw306>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* 46(W1): W296–W303. <https://doi.org/10.1093/nar/gky427>
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., & Richardson, D. C. 2018. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* 27(1): 293-315. <https://doi.org/10.1002/pro.3330>
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. 2015. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods* 12(1), 7–8: <https://doi.org/10.1038/nmeth.3213>
- Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9(1): 40. <https://doi.org/10.1186/1471-2105-9-40>
- Zou, Y., Zhang, Z., Zeng, Y., Hu, H., Hao, Y., Huang, S., & Li, B. 2024. Common methods for phylogenetic tree construction and their implementation in R. *Bioengineering* 11(5): 480. <https://doi.org/10.3390/bioengineering11050480>