Medical Microbiology and Immunology (2022) 211:153–160
https://doi.org/10.1007/s00430-022-00739-4

ORIGINAL INVESTIGATION

Genetic variability in minor capsid protein (L2 gene) of human papillomavirus type 16 among Indian women

Arati Mane[1] · Sanket Limaye[2] · Linata Patil[1] · Urmila Kulkarni-Kale[2]

Received: 17 January 2022 / Accepted: 21 April 2022 / Published online: 13 May 2022

Abstract

Human papillomavirus type 16 (HPV-16) is the predominant genotype worldwide associated with invasive cervical cancer and hence remains as the focus for diagnostic development and vaccine research. L2, the minor capsid protein forms the packaging unit for the HPV genome along with the L1 protein and is primarily associated with transport of genomic DNA to the nucleus. Unlike L1, L2 is known to elicit cross-neutralizing antibodies and thus becomes a suitable candidate for pan-HPV prophylactic vaccine development. In the present study, a total of 148 cervical HPV-16 isolates from Indian women were analyzed by PCR-directed sequencing, phylogenetic analysis and in silico immunoinformatics tools to determine the L2 variations that may impact the immune response and oncogenesis. Ninety-one SNPs translating to 35 non-synonymous amino acid substitutions were observed, of these 16 substitutions are reported in the Indian isolates for the first time. T245A, L266F, S378V and S384A substitutions were significantly associated with high-grade cervical neoplastic status. Multiple substitutions were observed in samples from high-grade cervical neoplastic status as compared to those from normal cervical status ($p = 0.027$), specifically from the D3 sub-lineage. It was observed that substitution T85A was part of both, B and T cell epitopes recognized by MHC-I molecules; T245A was common to B and T cell epitopes recognized by MHC-II molecules and S122P/A was common to the region recognized by both MHC-I and MHC-II molecules. These findings reporting L2 protein substitutions have implications on cervical oncogenesis and design of next-generation L2-based HPV vaccines.

Keywords Human papillomavirus 16 · L2 variations · Phylogeny · Virus bioinformatics · Immunoinformatics · Cervical cancer

Introduction

Cervical cancer is the fourth common cancer occurring in women worldwide, with an estimated 570,000 new cases and 250,000 deaths occurring annually [1, 2]. Infection with high-risk or oncogenic human papillomavirus (HPV) types plays a critical role in the development of cervical cancer. HPV type 16 (HPV-16) remains the most prevalent high-risk type globally and in India [2, 3]. The ~8 kb genome of HPV-16 is packaged in the viral capsid which is composed of 360 molecules of the major capsid protein L1 and up to 72 molecules of the minor capsid protein L2, which largely lies buried inside L1 [4, 5]. The capsid proteins can assemble into virus-like particles (VLPs) and induce highly neutralizing antibodies that have shown promise in the prevention of papillomavirus-associated cancers. The currently available HPV vaccines, Gardasil and Cervarix are based on the L1, major capsid protein. However, targeting the L1 protein alone, prevents infection by limited HPV types because of lack of cross-reactive epitopes toward other HPV types. The HPV L2 protein on the other hand offers a distinct advantage, as it not only elicits neutralizing antibodies but also cross-neutralizing antibodies for different HPV types, especially against its 'N' terminus [4, 6, 7]. Therefore, development of a pan-HPV prophylactic vaccine derived from L2 regions seems feasible and needs to be explored.

Arati Mane and Sanket Limaye have contributed equally to this work.

Edited by: Ulrike Wieland.

✉ Arati Mane
amane@nariindia.org

[1] ICMR - National AIDS Research Institute, '73' G Block, MIDC, Bhosari, Pune 411026, India

[2] Savitribai Phule Pune University, Ganeshkhind Road, Pune 411007, India

For the development of an efficient L2 vaccine construct, recognition of appropriate epitope sequence is imperative to generate an efficient immune response [8, 9]. Implementation of informatics tools and online databases can facilitate this and reduce the time and experimental costs involved [10]. Determination of genetic variability, pertaining to epitopes, is important, as substitution of one or more amino acid(s) in the L2 protein could impact the conformation of epitopes relevant for viral neutralization. L2 protein also plays an important role during HPV infection, by binding to a secondary viral receptor to facilitate exit from the endosomes and delivery of the viral genome to appropriate domains within the nucleus [11]. Variations in L2 gene might lead to amino acid substitutions and the associated conformational changes can impact viral assembly and clinical outcome of infection. Ideally, vaccine constructs and diagnostics need to be developed locally and hence it is important to understand the geographical variations.

Data with regard to molecular analysis of L2 protein of HPV-16 from India is negligible. The present study was thus carried out to determine the genetic variability in the L2 protein of HPV-16 and to analyze the association of the identified variations with cervical disease status. We further predicted the immunogenic B and T cell epitopes of L2 using in silico immunoinformatics tools and identified the substitutions that may impact these epitopes.

Methods

The study was approved by Institutional Ethics Committee of ICMR-National AIDS Research Institute (ICMR-NARI), Pune, India [NARI/EC/approval/17-18/31].

Clinical specimens

A total of 148 cervical specimens positive for HPV-16 infection collected from women undergoing cervical screening from different geographical locales from India and stored at −70 °C as part of the sample repository in the Microbiology laboratory, ICMR-NARI were included in the study. The samples were classified as belonging to normal ($n=41$), low-grade ($n=45$) or high-grade ($n=62$) cervical status as described previously [12]. HPV detection and genotyping for the samples was done using the Linear Array HPV genotyping test (Roche Molecular Systems, USA).

Nucleic acid extraction and sequencing

DNA was extracted from cervical samples using the QIAamp DNA mini kit (Qiagen, USA) and analyzed for L2 (3373–4794 bp) using two pairs of gene specific primers [HPV-16 L2 F1- TTACTTAACAATGCGACACA, HPV-16 L2 R1-TTATCCACATCTATACCTTCA, HPV-16 L2 F2-CCCTGCTTTTGTAACCACTC and HPV-16 R2-CGTGCAACATATTCATCCGT [13]. The amplified PCR products were purified with QIAquick PCR purification kit (Qiagen, USA) and sequenced using Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). The sequence data was curated using SeqScape (V.2.6) and aligned in SEED 2 software [14]. MEGA7 [15] was used to identify site specific nucleotide variations and their corresponding amino acid substitutions using the reference entry, Refseq_ID: NC_001526.4 [16].

Phylogenetic analysis

Phylogenetic analysis of the 148 isolates of L2 gene sequences generated in this study was carried out using both alignment-based Maximum Likelihood (ML) method (for 1000 bootstrap runs) implemented in IQTREE server (http://iqtree.cibiv.univie.ac.at/) [17] and Return Time Distribution (RTD) (http://bioinfo.unipune.ac.in/RTD/home.htm) developed in-house which is an alignment-free method [18] to analyze their clustering pattern. A dataset consisting of total 211 entries of L2 was used which includes 63 entries from GenBank with known lineages (reference dataset; https://www.ncbi.nlm.nih.gov/nucleotide/) and 148 sequences generated in this study. Typing of the HPV-16 Indian isolates was done on basis of their L1 gene [12] and these lineage and sub-lineages were assigned to the L2 sequences of corresponding isolates. The OTU (operational taxonomic unit) labels in the tree were generated as 'lineage/sub-lineage_isolate-ID_gene'. The OTU labels of isolates for which lineage/sub-lineage could not be assigned due to non-availability of their corresponding L1 sequences were generated as 'XX_isolate-ID_gene'. Phylogenetic trees were visualized using the iTOL server (https://itol.embl.de/) [19]. Recombination detection was carried out using the RDP4 software [20], while selection pressure analysis was carried out using Datamonkey server (http://www.datamonkey.org/) and a stringent cut-off of 10^{-5} was applied.

B cell and T cell epitope prediction

Experimentally characterized B and T cell epitopes of L2, were extracted from Immune Epitope Database (IEDB, http://www.iedb.org/home_v3.php). We further predicted linear B cell epitopes for L2 protein using Bepipred Linear Epitope Prediction 2.0 [21], Chou & Fasman Beta-Turn Prediction [22], Emini Surface Accessibility Prediction [23], Karplus & Schulz Flexibility Prediction [24], Kolaskar & Tongaonkar Antigenicity [25] and Parker Hydrophilicity Prediction [26] methods available on IEDB resource. Consensus peptides predicted using all the six prediction methods were taken and overlapping antigenic regions were

concatenated. T cell epitopes having potential to recognize both, MHC-I and MHC-II were predicted. MHC-I binding epitopes were predicted using MHC-I binding consensus [27] and MHC-I processing (Proteosome, TAP) methods [28]. The predicted epitopes were filtered using percentile rank, predicted IC50 < 50 nM, total score (combined score of proteasome, TAP and cleavage scores) and checked for consensus. MHC-II binding epitopes were predicted using MHC-II binding consensus method [29, 30] and filtered using both, percentile rank and predicted IC50 (<50 nM). The overlapping MHC-I and MHC-II peptides were concatenated to yield non-redundant antigenic regions.

Statistical analysis

Statistical analysis was done using SPSS (V.15.0). HPV-16 L2 variations and cervical lesion grade were compared using Mann–Whitney U test as described earlier [12]. Odds ratio and 95% confidence intervals were computed to determine the association between normal and high-grade cervical disease status. Differences were considered to be statistically significant if $p < 0.05$.

Results

HPV-16 L2 variations

A total of 148 HPV-16 positive cervical samples were analyzed to study the extent of variations in L2 gene that resulted in detection of 91 nucleotide polymorphisms, of which, 43 (47.3%) were non-synonymous and 48 (52.7%) were synonymous nucleotide variations, translating to 35 (39.8%) non-synonymous and 53 (60.2%) synonymous amino acid substitutions, with no indels.

A total of 16 amino acid substitutions L75F, T85A, T94A, S122A, S134R, S270N, D272N, N273S, D334T, Q342L, T352P, H354Q, T377S, L390F, I418M, S426A in L2 protein are reported in the Indian HPV-16 isolates for the first time. Amino acid substitutions L330F (75.6%), S269P (28.4%) and D334N (24.3%) were most frequently observed.

Evolutionary analysis of HPV-16 L2

Phylogenetic trees depict the evolution of L2 gene of Indian isolates (Fig. 1). Similar clustering patterns were observed for both, alignment-based ML and alignment-free RTD methods (Fig. 1a, b) barring relative order of isolates as part of lineage assignments. Of the 148 Indian isolates, 134 isolates were assigned lineage on the basis of L1 gene, whereas 14 isolates could not be assigned lineage due to lack of corresponding L1 sequences. As can be seen in Fig. 1, 124/134 (92.5%) isolates clustered in accordance with lineage assignment based on L1. Thus, 107/124 (86.3%) and 17/124 (13.7%) isolates clustered with members of lineage A and lineage D, respectively. Of the 10/134 (7.5%) isolates, that did not cluster in accordance with their respective assigned lineages, 7 and 2 isolates with assigned lineage A clustered with members of lineage D and C respectively. The remaining 1 isolate with assigned lineage D clustered with members of lineage C. The 14 isolates without assigned lineage, clustered into A and D lineage clusters with 7 isolates in each. No evidence of recombination and positive selection was observed.

HPV-16 L2 variations and cervical disease status

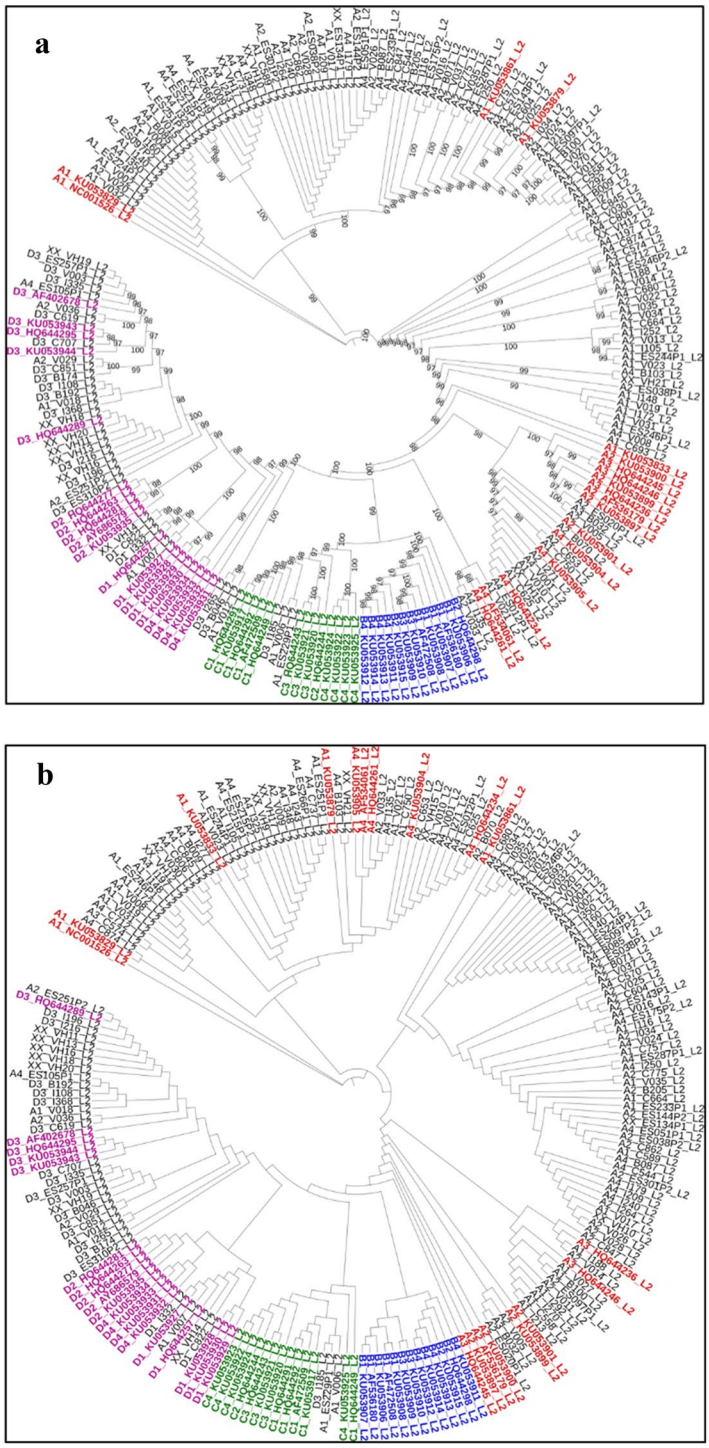
The association of amino acid substitutions with cervical lesion grade is shown in Table 1. Significantly higher occurrence of multiple substitutions in samples with high-grade cervical status as compared to normal cervical status ($p = 0.027$) was observed. T245A, L266F, S378V and S384A substitutions showed significant association with high-grade cervical status as compared to normal cervical status. These substitutions were predominantly observed in the D3 sub-lineage. The T245A substitution was observed in 19 samples (A1:2, A2:4, D1:1, D3:6 and unassigned sub-lineage:6), L266F was observed in 20 samples (A1:2, A2:4, D1:1, D3:7 and unassigned sub-lineage:6), S378V was observed in 20 samples (A1:2, A2:4, D1:1, D3:7 and unassigned sub-lineage:6) and S384A was observed in 20 samples (A1:2, A2:4, D1:1, D3:7 and unassigned sub-lineage:6). These 4 substitutions were found to be co-mutating in 19 samples.

Epitope prediction

Six experimentally validated non-overlapping B cell epitopes were obtained from IEDB. A total of 7 B cell epitopes were predicted. Amino acid substitutions L75F, T85A, T94A and T245A were part of these B cell epitopes (Table 2). One experimentally validated T cell epitope (MHC-I) was obtained from IEDB. Filtering of predicted MHC-I epitopes resulted into 20 overlapping peptides that were concatenated to yield 10 unique non-redundant regions. A total of 42 overlapping epitopes were predicted for MHC-II that resulted in 8 discrete regions. Amino acid substitutions T85A, S122P/A, S134R, T245A, L266F, S269P, S270N and I306L were part of predicted T cell epitopes (Table 3).

Discussion

HPV-16 is the predominant genotype worldwide, associated with invasive cervical cancers and hence remains as the focus for HPV diagnostic development and vaccine research. L2 protein of HPV is a major component required



Footnote: Reference dataset with known lineages is color coded as: Lineage A (red), B (blue), C (green), D (magenta)

Fig. 1 Phylogenetic tree of HPV-16 L2 gene: **a** Alignment-based Maximum Likelihood tree and **b** Alignment-free RTD-based tree

for virus assembly along with L1 and helps in transport of viral particles to the host cell nucleus. L2 protein also helps in mediating the increased efficiency of formation of VLPs by binding with L1 [31, 32]. The C-terminal amino acid residues 396–439 of HPV11 L2 are shown to form the L1 binding domain [11, 33]. This L2 region of HPV11 corresponds

Table 1 *Human papillomavirus* 16 amino acid substitutions in L2 by cervical lesion grade

HPV-16 L2 Amino acid substitutions	Cervical disease status			Total ($n=148$)	OR[a] (95% CI)	p-value
	Normal ($n=41$)	Low-grade ($n=45$)	High-grade ($n=62$)			
L75F	2	0	0	2	0.73 (0.29, 1.78)	0.512
T85A	2	0	0	2	–	–
T94A	0	2	0	2	–	–
S122P	8	5	20	36	0.51 (0.19, 1.30)	0.180
S122A	0	3	0	3	–	–
S134R	1	0	0	1	–	–
T245A	2	3	19	24	0.12 (0.02, 0.53)	**<0.01**
L266F	2	4	20	26	0.10 (0.02, 0.49)	**<0.01**
S269P	10	9	23	42	0.55 (0.23, 1.32)	0.201
S270N	0	3	0	3	–	–
D272N	0	3	0	3	–	–
N273S	0	2	1	3	–	–
I306L	0	2	1	3	–	–
L330F	30	40	42	112	1.96 (0.77, 5.01)	0.180
T332S	1	0	9	10	0.15 (0.02, 1.21)	0.084
D334N	8	10	19	37	0.47 (0.18, 1.21)	0.124
D334H	0	0	1	1	–	–
D334T	0	0	1	1	–	–
E338D	0	5	3	8	–	–
Q342L	2	0	0	2	–	–
T352P	3	0	0	3	–	–
H354Q	0	0	2	2	–	–
T377S	0	0	1	1	–	–
S378V	5	5	20	30	0.29 (0.09, 0.85)	**0.033**
S378F	8	8	20	36	0.51 (0.19, 1.30)	0.180
S384A	5	5	20	30	0.29 (0.09, 0.85)	**0.033**
V385I	8	5	22	35	0.44 (0.17, 1.11)	0.120
L390F	0	0	1	1	–	–
I418M	0	0	1	1	–	–
I420T	8	5	20	33	0.51 (0.19, 1.30)	0.180
Q423H	0	3	1	4	–	–
A424T	8	5	20	33	0.51 (0.19, 1.30)	0.180
S426A	2	0	0	2	–	–
I428L	2	5	5	12	0.58 (0.11, 3.16)	0.699
A443G	7	5	20	32	0.43 (0.16, 1.14)	0.110

OR Odds ratio, *95% CI* 95% confidence interval

[a]OR with 95% CI are presented for normal versus high-grade cervical status. The p values of significant associations are marked in boldface

to 412–455 amino acid residues of HPV-16. Seven non-synonymous amino acid substitutions were observed in this region, of which I420T and A443G substitutions were also reported at the corresponding positions in L2 of HPV11 [11]. These substitutions might affect the binding of L2 with respective L1 molecule and thereby impact formation of VLPs. It is also noteworthy that these substitutions (I420T and A443G) were predominantly observed in samples from high-grade cervical disease and may have a role in cervical oncogenesis as well.

The phylogenetic tree constructed from L2 gene sequences showed that majority of the isolates (92.5%) clustered in accordance with the lineages assigned on the basis of L1 as reported previously [12]. However, changes in the sub-lineage clustering pattern were observed, which indicates the differential evolution of L1 and L2.

Table 2 B cell epitopes (the amino acid and the substitution is highlighted in red and the substitution is listed in parentheses)

Description	Start	End
Experimentally validated B cell epitopes from IEDB		
QLYKTCKQAGTCPPDIIPKV	17	36
RTGYIPL(F)GTRPPT	69	81
LVEETSFIDAGAP	108	120
STHNYEEIPMDTFIVSTNPNTVTSSTPI	189	216
SGYIPANTTIPF	391	402
YMLRKRRKRLPYFF	453	466
Predicted B cell epitopes		
HKRSAKRTKRASATQLYKTC	3	22
GTRPPTATDT(A)LAPVRPPLT(A)VDPV	76	98
TTHNNPTFTDPSVLQPP	156	172
VARLGLYSRTTQQVKVVDPAFVTT(A)P	222	246
LDIVALHRPALTSRRTG	284	300
GAKVHYYYDLS	321	331
FYLHPSYYMLRKRRKR	446	461

Table 3 T cell epitopes (the amino acid and the substitution is highlighted in red and the substitution is listed in parentheses)

Description	Start	End
Experimentally validated T cell epitopes from IEDB		
AILDINNTV	144	152
Predicted MHC-I epitope		
ILQYGSMGVFF	45	55
TATDT(A)LAPV	81	89
SLVEETSFIDAGAPTS(P/A)V	107	123
AILDINNTV	144	152
AETGGHFTL	175	183
GLYSRTTQQVKVVDPAF	226	242
KLITYDNPAY	248	257
DFLDIVALHR	282	291
RI(L)GNKQTLRTRSGKSIGAKVHYY	305	327
SYYMLRKRRKRLPYFFSDV	451	469
Predicted MHC-II epitope		
GSMGVFFGGLGIGTGSGT	49	66
EETSFIDAGAPTS(P/A)VPSIP	110	127
PDVSGFS(R)ITTSTDTPAILDINNTVTTVT	128	156
IPMDTFIVSTNPNTVTSST	196	214
KVVDPAFVTT(A)PTKLITYDNP	236	255
AYEGIDVDNTL(F)YFS(P)S(N)N	256	271
IAPDPDFLDIVALHRPALTSR	277	297
FYLHPSYYMLRKRRKRLPYFFSDVSLAA	446	473

The L2 residues 1-MRHKRSAKRTKR-12 and 456-RKRRKR-461 constitute the nuclear localization signal (NLS) and the region 296-SRRTGIRYSRIGNKQTL-RTRS-316 constitutes the arginine rich nuclear retention signal (NRS). Substitution of arginine residues within NRS lead to reduction in L2 concentration in nucleus even though both the NLS are retained [34]. The only substitution observed in the NRS was I306L that occurs immediately after arginine at 305th position and might impact L2 concentration in nucleus, which needs to be validated experimentally.

Nucleotide variations and the resultant amino acid substitutions can alter the L2 protein properties which can impact the carcinogenic potential. We observed that substitutions T245A, L266F, S378V and S384A were significantly associated with the high-grade cervical disease. These substitutions need further functional validation to decipher their role as molecular marker(s) of cervical carcinogenesis. Contrary to a previous report we did not find difference in the prevalence of non-synonymous variations between samples from normal and high-grade cervical disease, though the occurrence of multiple non-synonymous variations differed significantly [13].

In the present study, the potential impact of substitutions on antigenicity and immunogenicity was evaluated. The amino acid residues 69–81 and 108–120 of L2 protein are highly conserved and play an important role in inducing neutralizing antibodies [35, 36]. Two monoclonal antibodies (mAbs) KIL2 and MAb6 recognize L2 regions, 73–79 and 65–81 respectively [37, 38]. The L75F substitution is part of these antibody binding sites and may impact the production of neutralizing antibodies. In addition to experimentally validated epitopes reported in IEDB, we predicted, both B and T cell epitopes using computational methods. L75F, the only experimentally validated substitution in B cell epitope was rarely observed in the 148 samples studied. Recently, a number of HPV vaccine constructs based on L2 covering different epitope regions are reported [39–42]. The substitution T85A observed in our study is part of the L2 region of the proposed therapeutic and prophylactic vaccine construct based on L2, E6 and E7 regions [41]. Therefore, T85A substitution might impact vaccine efficacy being common to both B and MHC-I epitopes.

The substitution S122P/A is common to the epitope region recognized by both MHC-I and MHC-II. The substitution T245A is part of both B and T cell epitopes (recognized by MHC-II), while the substitution L266F is part of epitope recognized by MHC-II. These substitutions were observed to be associated with high-grade cervical lesions. Thus, the regions containing substitutions (T245A and L266F) need to be given consideration for disease severity status, immune response based on antigenic diversity, validating existing vaccines and designing novel vaccine candidates.

To conclude, the findings reported in this study would help to understand the impact of L2 protein substitutions on cervical oncogenesis as well as in considerations for design

of next-generation L2-based HPV vaccine, subject to experimental validations.

Acknowledgements Bioinformatics Centre, Savitribai Phule Pune University is supported by Centre of Excellence (CoE) grant by the Department of Biotechnology, Government of India.

Author contributions AM and UK-K devised the project. LP and SL carried out the laboratory and bioinformatics work, respectively. Manuscript writing was done by AM, UK-K and SL. All authors contributed to final draft.

Funding The study is supported by the Science and Engineering Research Board, Department of Science and Technology, Government of India (EMR/2016/002498) to NARI, Pune, India and Savitribai Phule Pune University, Pune, India.

Data availability HPV-16 L2 sequences of Indian isolates are deposited in GenBank with accession codes MT013256-MT013284.

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, Bray F (2020) Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Health 8(2):e191–e203. https://doi.org/10.1016/S2214-109X(19)30482-6
2. World Health Organization. Human papillomavirus (HPV) and cervical cancer. November 2020. https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer. Accessed November 2020
3. Human Papillomavirus and Related Diseases Report 2019. Version posted at www.hpvcentre.net on 17 June 2019. https://hpvcentre.net/statistics/reports/IND.pdf. Accessed 17 June 2019
4. Buck CB, Trus BL (2012) The papillomavirus virion: a machine built to hide molecular Achilles' heels. Adv Exp Med Biol 726:403–422. https://doi.org/10.1007/978-1-4614-0980-9_18
5. Popa A, Zhang W, Harrison MS, Goodner K, Kazakov T, Goodwin EC, Lipovsky A, Burd CG, DiMaio D (2015) Direct binding of retromer to human papillomavirus type 16 minor capsid protein L2 mediates endosome exit during viral infection. PLoS Pathog 11(2):e1004699. https://doi.org/10.1371/journal.ppat.1004699
6. Yue Y, Yang H, Wu K, Yang L, Chen J, Huang X, Pan Y, Ruan Y, Zhao Y, Shi X, Sun Q, Li Q (2013) Genetic variability in L1 and L2 genes of HPV-16 and HPV-58 in Southwest China. PLoS One 8(1):e55204. https://doi.org/10.1371/journal.pone.0055204
7. Olczak P, Roden RBS (2020) Progress in L2-based prophylactic vaccine development for protection against diverse Human Papillomavirus genotypes and associated diseases. Vaccines (Basel) 8(4):568. https://doi.org/10.3390/vaccines8040568
8. Bahmani B, Amini-Bayat Z, Ranjbar MM, Bakhtiari N, Zarnani AH (2020) HPV16-E7 protein T cell epitope prediction and global therapeutic peptide vaccine design based on human leukocyte antigen frequency: an in-silico study. Int J Pept Res Ther 27(1):365–378. https://doi.org/10.1007/s10989-020-10089-5
9. Kumar A, Hussain S, Sharma G, Mehrotra R, Gissmann L, Das BC, Bharadwaj M (2015) Identification and validation of immunogenic potential of India specific HPV-16 variant constructs: In-silico & in-vivo insight to vaccine development. Sci Rep 5:15751. https://doi.org/10.1038/srep15751
10. Kaliamurthi S, Selvaraj G, Kaushik AC, Gu KR, Wei DQ (2018) Designing of $CD8^+$ and $CD8^+$-overlapped $CD4^+$ epitope vaccine by targeting late and early proteins of human papillomavirus. Biologics 12:107–125. https://doi.org/10.2147/BTT.S177901
11. Wang JW, Roden RB (2013) L2, the minor capsid protein of papillomavirus. Virology 445(1–2):175–186. https://doi.org/10.1016/j.virol.2013.04.017
12. Mane A, Patil L, Limaye S, Nirmalkar A, Kulkarni-Kale U (2020) Characterization of major capsid protein (L1) variants of Human papillomavirus type 16 by cervical neoplastic status in Indian women: Phylogenetic and functional analysis. J Med Virol 92(8):1303–1308. https://doi.org/10.1002/jmv.25675
13. Bhattacharjee B, Mandal NR, Roy S, Sengupta S (2008) Characterization of sequence variations within HPV16 isolates among Indian women: prediction of causal role of rare non-synonymous variations within intact isolates in cervical cancer pathogenesis. Virology 377(1):143–150. https://doi.org/10.1016/j.virol.2008.04.007
14. Větrovský T, Baldrian P, Morais D (2018) SEED 2: a user-friendly platform for amplicon high throughput sequencing data analyses. Bioinformatics 34:2292–2294. https://doi.org/10.1093/bioinformatics/bty071
15. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054
16. Parton A (1990) Nucleotide sequence of the HPV16 L1 open reading frame. Nucleic Acids Res 18:3631. https://doi.org/10.1093/nar/18.12.3631
17. Trifinopoulos J, Nguyen LT, von Haeseler A et al (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 4:W232–W235. https://doi.org/10.1093/nar/gkw256
18. Kolekar P, Kale M, Kulkarni-Kale U (2012) Alignment-free distance measure based on return time distribution for sequence analysis: Applications to clustering, molecular phylogeny and subtyping. Mol Phylogenet Evol 65:510–522. https://doi.org/10.1016/j.ympev.2012.07.003
19. Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44(W1):W242–W245. https://doi.org/10.1093/nar/gkw290
20. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1(1):vev003. https://doi.org/10.1093/ve/vev003
21. Jespersen MC, Peters B, Nielsen M, Marcatili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 45(W1):W24–W29. https://doi.org/10.1093/nar/gkx346
22. Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol 47:45–148. https://doi.org/10.1002/9780470122921
23. Emini EA, Hughes JV, Perlow DS, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55(3):836–839. https://doi.org/10.1128/JVI.55.3.836-839.1985
24. Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins. Naturwissenschaften 72:212–213. https://doi.org/10.1007/BF01195768
25. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS

Lett 276(1–2):172–174. https://doi.org/10.1016/0014-5793(90)80535-q
26. Parker JM, Guo D, Hodges RS (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 25(19):5425–5432. https://doi.org/10.1021/bi00367a013
27. Moutaftsi M, Peters B, Pasquetto V, Tscharke DC, Sidney J, Bui HH, Grey H, Sette A (2006) A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. Nat Biotechnol 24(7):817–819. https://doi.org/10.1038/nbt1215
28. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Røder G, Peters B, Sette A, Lund O, Buus S (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE 2(8):e796. https://doi.org/10.1371/journal.pone.0000796
29. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4):e1000048. https://doi.org/10.1371/journal.pcbi.1000048
30. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinform 11:568. https://doi.org/10.1186/1471-2105-11-568
31. Kirnbauer R, Taub J, Greenstone H, Roden R, Dürst M, Gissmann L, Lowy DR, Schiller JT (1993) Efficient self-assembly of human papillomavirus type 16 L1 and L1–L2 into virus-like particles. J Virol 67(12):6929–6936. https://doi.org/10.1128/JVI.67.12.6929-6936.1993
32. Zhou J, Sun XY, Stenzel DJ, Frazer IH (1991) Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. Virology 185(1):251–257. https://doi.org/10.1016/0042-6822(91)90772-4
33. Finnen RL, Erickson KD, Chen XS, Garcea RL (2003) Interactions between papillomavirus L1 and L2 capsid proteins. J Virol 77(8):4818–4826. https://doi.org/10.1128/jvi.77.8.4818-4826.2003
34. Mamoor S, Onder Z, Karanam B, Kwak K, Bordeaux J, Crosby L, Roden RB, Moroianu J (2012) The high risk HPV16 L2 minor capsid protein has multiple transport signals that mediate its nucleocytoplasmic traffic. Virology 422(2):413–424. https://doi.org/10.1016/j.virol.2011.11.007
35. Slupetzky K, Gambhira R, Culp TD, Shafti-Keramat S, Schellenbacher C, Christensen ND, Roden RB, Kirnbauer R (2007) A papillomavirus-like particle (VLP) vaccine displaying HPV16 L2 epitopes induces cross-neutralizing antibodies to HPV11. Vaccine 25(11):2001–2010. https://doi.org/10.1016/j.vaccine.2006.11.049
36. Kawana K, Matsumoto K, Yoshikawa H, Taketani Y, Kawana T, Yoshiike K, Kanda T (1998) A surface immunodeterminant of human papillomavirus type 16 minor capsid protein L2. Virology 245(2):353–359. https://doi.org/10.1006/viro.1998.9168
37. Rubio I, Seitz H, Canali E, Sehr P, Bolchi A, Tommasino M, Ottonello S, Müller M (2011) The N-terminal region of the human papillomavirus L2 protein contains overlapping binding sites for neutralizing, cross-neutralizing and non-neutralizing antibodies. Virology 409(2):348–359. https://doi.org/10.1016/j.virol.2010.10.017
38. Kawana K, Yoshikawa H, Taketani Y, Yoshiike K, Kanda T (1999) Common neutralization epitope in minor capsid protein L2 of human papillomavirus types 16 and 6. J Virol 73(7):6188–6190. https://doi.org/10.1128/JVI.73.7.6188-6190.1999
39. Pouyanfard S, Spagnoli G, Bulli L, Balz K, Yang F, Odenwald C, Seitz H, Mariz FC, Bolchi A, Ottonello S, Müller M (2018) Minor capsid protein L2 polytope induces broad protection against oncogenic and mucosal Human Papillomaviruses. J Virol 92(4):e01930-e2017. https://doi.org/10.1128/JVI.01930-17
40. Namvar A, Bolhassani A, Javadi G, Noormohammadi Z (2019) In silico/In vivo analysis of high-risk papillomavirus L1 and L2 conserved sequences for development of cross-subtype prophylactic vaccine. Sci Rep 9(1):15225. https://doi.org/10.1038/s41598-019-51679-8
41. Bagheri A, Nezafat N, Eslami M, Ghasemi Y, Negahdaripour M (2021) Designing a therapeutic and prophylactic candidate vaccine against human papillomavirus through vaccinomics approaches. Infect Genet Evol 95:105084. https://doi.org/10.1016/j.meegid.2021.105084
42. Ray S, Wirth DM, Ortega-Rivera OA, Steinmetz NF, Pokorski JK (2022) Dissolving microneedle delivery of a prophylactic HPV vaccine. Biomacromol 23(3):903–912. https://doi.org/10.1021/acs.biomac.1c01345