International Journal of Multidisciplinary and Current Research ISSN: 2321-3124
Research Article Available at: http://ijmcr.com

Immunoinformatics Approach for Designing Epitope-Based Peptides Vaccine of L1 Major Capsid Protein against HPV Type 16

Sahar Suliman Mohamed*, Marwa Mohamed Osman*, Samah Mahmoud Sami, Hanaa Abdalla Elgailany, Maysaa Khalid Bushara, Hadeel Abdelrahman Hassan, Ghidaa Faisal Mustafa, Amna Idris Ali, Afra Abd Elhamid Fadl Alla, Alaa Abusufian Elkabashi & Mohamed Ahmed Salih

Department of Biotechnology, Africa city of Technology- Khartoum, Sudan. Corresponding author: Marwa Mohamed Osman: +249912260109
* Sahar Suliman Mohamed and Marwa Mohamed Osman contributed equally

Accepted 10 Sept 2016, Available online 19 Sept 2016, **Vol.4 (Sept/Oct 2016 issue)**

Abstract

Human papilloma viruses (HPV) are small DNA, non-enveloped, double-stranded and closed circular viruses. There are more than 150 HPV identified types. Genital HPV types are categorized according to their epidemiologic association with cervical cancer to high and low risk types. The high risk type HPV 16 is the most common in the world. we aimed to design a universal peptide based vaccine against HPV type 16 virus using Immunoinformatics Approach through prediction of highly conserved T and B-cell epitopes from the most abundant and highly immunogenic protein (L1 major capsid protein) derived from HPV type 16 strains all over the world. All sequences of the L1 major capsid protein were retrieved from NCBI database. Potentially continuous B and T cell epitopes were predicted using tools from immune epitope data base analysis resource (IEDB-AR). The Allergenicity of predicted epitopes was analyzed by AllerTOP Tool and the coverage was determined throughout the worlds. The B cell epitope ^{243}***KSEV***^{246} *and T cell epitopes* ^{363}***TRSTNMSLC***^{371}, ^{470}***YTFWEVNLK***^{478}, ^{260}***YIKMVSEPY***^{268} *and* ^{236}***FTTLQANKS***^{244} *were suggested to become universal peptide based vaccine against HPV type 16. We hope to confirm our findings by adding complementary steps of both in vitro and in vivo studies to support this new universal predicted vaccine.*

Keywords: *Human papilloma viruses (HPV), Epitope, vaccine & In Silico.*

Introduction

Human papilloma viruses (HPV) are small DNA viruses, non-enveloped, double-stranded and closed circular. The genome is approximately 8,000 base pairs, encoding six early proteins [E1-E6] and two late proteins [L1 and L2] with a strict tropism for human epithelial cells. It is responsible for benign lesions of the skin and mucous membranes. HPV is also involved in the development of various mucocutaneous tumors: Bowen's disease, non-melanoma skin cancers and genital carcinomas. They may be divided into cutaneous and mucosal types depending on the type of epithelium they infect [1-5].

There are more than 150 HPV identified types, including approximately 40 that infect the genital area. Genital HPV types are categorized according to their epidemiologic association with cervical cancer to high and low risk types [3,6,7]. The high risk HPV types (HPV.16, .18, .31, .33, and .45) have been linked to several epithelial cancers in human (uterine cervix, head, neck and esophageal cancer) [8]. HPV 16 is the most common in the world, except Indonesia and Algeria, where HPV 18 is the most common. HPV 45 shows high frequency in West Africa. Types 33, 39, and 59 are concentrated in Central and South America [9]. The epidemiological studies have established that the most common high risk HPV type is type 16. It is present in approximately 50% of the >500,000 cases of cervical cancer diagnosed annually worldwide. This cervical cancer is the second most frequent cancer in women worldwide with 250.000 deaths yearly [1,5,6]. Transmission of HPV occurs primarily by skin to skin contact. Basal cells of stratified squamous epithelium may be infected by HPV. Other cell types appear to be relatively resistant [10,11] .

HPV L1 and L2 late gene expression is required for virus production and normally occurs in terminally differentiated cells at the very top of the epithelium [1]. The L1 ORF is the most conserved gene within the genome and has therefore been used for the identification of new PV types over the past 15 years [4]. L1 protein is the major structural protein (360 copies) and assembled into 72 pentameric capsomeres that are arranged in an icosahedral array. It represents 80% of the viral capsid proteins, being the most abundant protein and highly immunogenic. Whereas the L2 protein, the

minor capsid protein (72 copies), is not necessary to form viral particles [6,12,13] .

HPV vaccine was first introduced in the United States; the vaccination is directed to prevent HPV infection [14]. There were two prophylactic vaccines have been developed and tested in large multicentric trials [15-17]. Quadrivalent HPV vaccine (Gardasil, produced by Merck and Co, Inc., Whitehouse Station, New Jersey) , licensed for use in females and males aged 9 through 26 years, and Bivalent HPV vaccine (Cervarix, produced by GlaxoSmithKline, Rixensart, Belgium) ,licensed for use in females aged 9 through 25 years. Both of them are composed of type-specific HPV L1 protein, the major capsid protein [14]. The HPV VLPs contain no DNA and hence are noninfectious and it elicits a strong and sustained type-specific response [14]. The limitations of these available vaccines are: (a) these vaccines do not protect against all high-risk HPV types; (b) they do not treat existing HPV infections; (c) the long-term duration of protection and the required length of protection to prevent cancer are unknown; detect evidence of waning immunity over 5 years [14-16] .

A recent approach known as vaccinomics integrating immunogenetics and immunogenomics with bioinformatics has been used for the development of new vaccines [18-20]. For this reason, the rapid in silico informatics-based approach became much popular with the recent advancement in the sequencing of many pathogen genomes and protein sequence databases [20].
In this study we aim to design a universal peptide based vaccine against HPV type 16 virus using computational method through prediction of highly conserved T and B-cell epitopes from the most abundant and highly immunogenic protein (L1 major capsid protein) derived from HPV type 16 strains all over the world.

Martials and Methods

Protein sequence retrieval

The sequences of the late major capsid protein (L1) of HPV type 16 were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) in May 2016, and then all the sequences were stored as a FASTA format for further analysis. These sequences were isolated from different geographical areas (Japan, China, Mexico, Iran, India, Netherlands, Canada, Germany, Brazil, Sweden, Croatia, Thailand and Pakistan) from 2001-2014. The retrieved strains are listed in Table (A) in supplementary data

Retrieved Strains Phylogeny

The relationships of all retrieved strains were studied using phylogeny.fr online software (http:// phylogeny.lirmm.fr/ phylo_cgi/index.cgi) [21].

Conserved region identification

BioEdit sequence alignment editor (v7.0.9) was used to identify the conserved regions with ClustalW Multiple alignment compared to L1 major capsid protein of HPV type 16 reference sequence under gene bank accession number NP_041332.1 [22]

Identification of the B cell epitope

Potentially continuous B cell epitopes were predicted using tools from immune epitope data base analysis resource (IEDB-AR) (http://tools.iedb.org/bcell/) by Bepipred linear epitope prediction analysis [23] after submitting the reference sequence. IEDB analysis resource was also used for the analysis of the epitope conservancy (http://tools. iedb.org/tools/ conservancy/ iedb_input) [24]. Only 100% conserved epitopes were selected.Then IEDB tools were used to identify the B cell antigenicity including the Kolaskar and Tongaonkar antigenicity scale [25], Emini surface accessibility prediction [26] and Parker Hydrophilicity Prediction [27] for the selected epitopes with thresholds of 1.033, 1.00 and 1.359 respectively. Epitopes which pass these tests were predicted as B cell epitope.

Identification of the T cell epitope

IEDB was used to predict T cell epitopes that bind major histocompatibility complex (MHC) class 1 and 2. For MHC class 1 (T helper epitope-THL) ; the reference sequence was submitted for peptide prediction using MHC-I Binding Predictions (http://tools.iedb.org/mhci/), the stabilized matrix base method (SMM) was used to calculate the half-maximal inhibitory concentration (IC_{50}) values of peptide binding to MHC-I molecules. For the binding analysis, all the alleles were selected, and the length was set at 9.0 before prediction was done and percentile rank cutoff was set below 1.00 [28,29]. As in B cell epitope prediction, only 100% conserved epitopes were determined by IEDB epitope conservancy tool and selected for further investigations.

Then epitopes were classified according to their IC50 into high affinity epitopes (IC_{50}<50), moderate affinity epitopes (IC_{50}<500) and low affinity epitopes (IC_{50} <500), only high and moderate affinity epitopes with their corresponding alleles were subjected for population coverage analysis.

In MHC class 2 (cytotoxic T cell epitope CTL) : repeating the same steps as MHC class 1 but setting the percentile cutoff at 10 using MHC-II Binding Predictions tool (http://tools.iedb.org/mhcii/) [30].

Assessment of Epitope Allergenicity

AllerTOP (http://www.pharmfac.net/allertop) was used to analyze the predicted epitopes (especially B and CTL epitopes). AllerTOP uses a model based on amino acid z-descriptors, ACC protein transformation and k nearest neighbors clustering. It defines the most probable route of exposure of tested proteins predicted as an allergen: food, inhalant or toxin [31]. The predicted epitopes were analyzed as: 'Probable Allergen' or 'Probable Non-allergen'

Calculation of Population Coverage

Population coverage of the whole world for epitope was assessed by the IEDB population coverage calculation tool (http://tools.iedb.org/tools/population/iedb_input). This tool calculates the fraction of individuals predicted to respond to a given set of epitopes with known MHC restrictions. This calculation is made on the basis of HLA genotypic frequencies assuming non-linkage disequilibrium between HLA loci [32]. Here the allelic frequencies of the interacting HLA alleles were used for the prediction of the population coverage for the corresponding epitope. Epitopes with the highest frequencies were selected for modeling.

Modeling of B & T epitopes

The structure of the L1 major capsid protein of human papillomavirus type 16 was predicted in Protein Data Bank (PDB) format, using the web portal RaptorX (http://raptorx.uchicago.edu/) [33]. UCSF Chimera 1.8 visualization software [34] was used for visualization of the selected epitopes.

Results

Phylogenetic Analysis of Retrieved Strains

The relationships of all retrieved strains of L1 major capsid protein of HPV type 16 are illustrated in Figure (1) below.

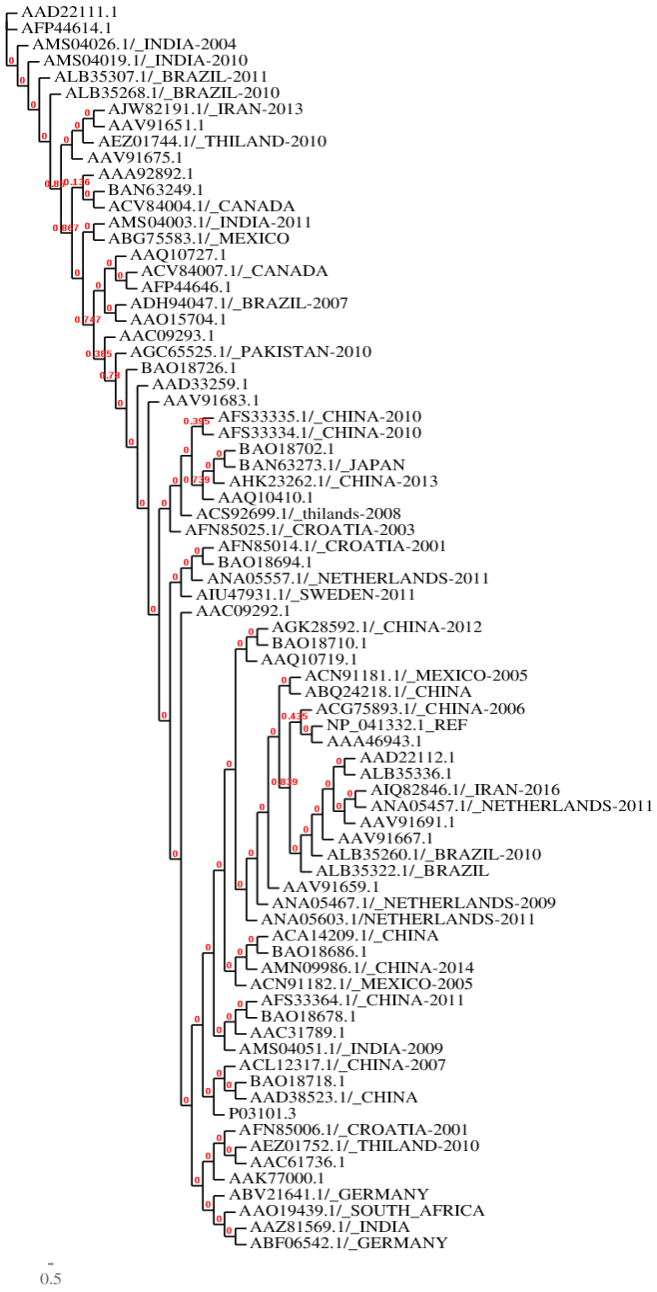

Figure (1): Phylogenetic tree of the retrieved sequences of L1 major capsid protein of HPV type 16. (The branch length is proportional to the number of substitutions per site)

Table (1): list of B- cell epitopes predicted by different scales of L1 major capsid protein in HPV type 16

Bepipred Epitope/Threshold 0.35	Position	Length	Emini Surface Score/Threshold 1.000	Antigenicity Score/Threshold 1.033	Hydrophilicity Score/Threshold 1.359	Beta Turn Score/Threshold 1.010
VYLPPVP	37-43	7	0.432	1.196	-1.743	1.041
PIKKPNNN	77-84	8	3.559	0.933	3.575	1.276
PDTSFYNP	112-119	8	2.151	0.993	2.725	1.274
VGRGQP	133-138	6	0.806	1.014	3.333	1.178
PPIGEHWGKGS	189-198	11	0.63	0.972	2.309	1.207
ANKSEV	241-246	6	1.184	1.003	4.233	0.983
*KSEV	243-246	4	1.217	1.044	4.075	0.92
EPYG	266-269	4	1.467	0.988	3.425	1.24
YIKGSG	302-307	6	0.61	1	2.283	1.195
YFPTPSGSM	317-325	9	0.813	1.001	1.422	1.196
QGHN	343-346	4	1.325	0.942	5.2	1.263
TTRS	362-365	4	1.931	0.926	5.275	1.075
TYKNT	380-384	5	2.869	0.937	4.24	1.126
FGLQPPPGGTLE	430-441	12	0.408	1.015	1.233	1.142
GLQPPP	431-436	6	1.105	1.055	1.467	1.282
LQPPPG	432-437	6	1.105	1.055	1.467	1.282
KHTPPAPKED	455-464	10	6.486	0.985	4.49	1.135
ATPTT	514-518	5	1.283	0.971	3.96	1.012

*proposed epitope

Prediction of B-cell epitope

HPV Type 16 L1 major capsid protein was subjected to Bepipred linear epitope prediction, Emini surface accessibility, Kolaskar and Tongaonkar antigenicity, Parker hydrophobicity and Chou and Fasman beta turn prediction methods in IEDB with their default thresholds setting. Only Three epitopes were found to have cutoff prediction scores above threshold scores, namely **KSEV** from 243 to 246 and **GLQPPP** from 431 to 436 or **LQPPPG** from 432 to 437, Figure (2).

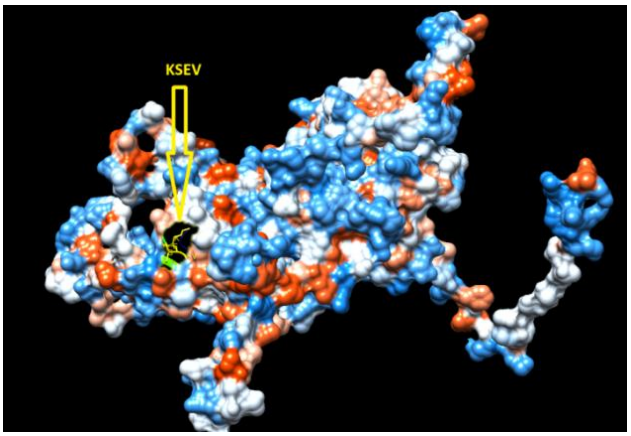

Figure (2): proposed B-Cell Epitope of L1 major capsid protein

The other epitopes were not satisfied the threshold value of Kolaskar and Tongaonkar antigenicity, the result of all conserved predicted B cell epitopes are listed in Table (1)

Prediction of Cytotoxic T-lymphocyte epitope and interaction with Mouse MHC Class I and Modeling

L1 major capsid protein was analyzed using IEDB MHC-1 binding prediction tool to predict T cell epitope interacted with different types of selected Human MHC Class I alleles. Based on Consensus (SMM) with percentile rank ≤1 and IC_{50} <500; 16 conserved peptides were predicted to interact with different Human MHC-1 alleles. **YTFWEVNLK** epitope from 470 to 478 had had higher affinity to interact with (HLA-A*68:01 and HLA-A*11:01) alleles with IC_{50}= 8.27 and 23.35, respectively and intermediate affinity to bind with HLA-A*03:01 (IC_{50}=235.4 nm). Then **NTNFKEYLR** from 383 to 391 which had higher affinity to interact with (HLA-A*68:01 and HLA-A*31:01) alleles with scores IC_{50}= 10.53 and 22.15 nm. Followed by **TRSTNMSLC** epitope from 363 to 371 had higher affinity to interact with tow alleles (HLA-C*07:01, HLA-C*06:02) IC_{50}=6.36 and 43.63nm, respectively, Figure (3). Other CTL epitopes **FVTVVDTTR, RLVWACVGV, FQMSLWLPS** and **SGLQYRVFR** had affinity to interact with one allele by higher affinity. The CTL epitopes and their corresponding human MHC-1 alleles are shown in Table (2).

Table (2): list of the CTL epitopes which had high and intermediate binding affinity with the selected Human MHC Class I alleles

Epitope	allele	start	end	length	percentile rank	IC_{50} (SMM)
*TRSTNMSLC	HLA-C*07:01	363	371	9	0.9	6.36
	HLA-C*06:02				0.85	43.63
*YTFWEVNLK	HLA-A*68:01	470	478	9	0.15	8.27
	HLA-A*11:01				0.25	23.35
	HLA-A*03:01				0.85	235.4
FVTVVDTTR	HLA-A*68:01	356	364	9	0.3	16.58
RLVWACVGV	HLA-A*02:06	123	131	9	0.7	19.39
FQMSLWLPS	HLA-A*02:06	25	33	9	0.6	8.96
SGLQYRVFR	HLA-A*31:01	92	100	9	0.8	40.32
NTNFKEYLR	HLA-A*68:01	383	391	9	0.2	10.53
	HLA-A*31:01				0.55	22.15
*YIKMVSEPY	HLA-B*15:01	260	268	9	0.3	23.28
	HLA-A*25:01				0.75	241.59
FYLRREQMF	HLA-A*23:01	274	282	9	0.35	58
	HLA-B*35:03				0.8	66.56
	HLA-A*24:02				0.35	87.63
WEVNLKEKF	HLA-B*18:01	473	481	9	0.25	58.08
VPLDICTSI	HLA-B*51:01	246	254	9	0.2	154.39
	HLA-B*53:01				1	277.26
FFYLRREQM	HLA-B*14:02	273	281	9	0.4	244.37
DICTSICKY	HLA-A*25:01	249	257	9	0.25	67.31
KYTFWEVNL	HLA-A*23:01	469	477	9	0.5	196.08
IYILVITCY	HLA-A*29:02	6	14	9	0.95	93.7
IFFQMSLWL	HLA-A*23:01	23	31	9	0.85	370.19

IC_{50}<50: High affinity, Ic_{50}<500: Intermediate affinity *proposed epitope

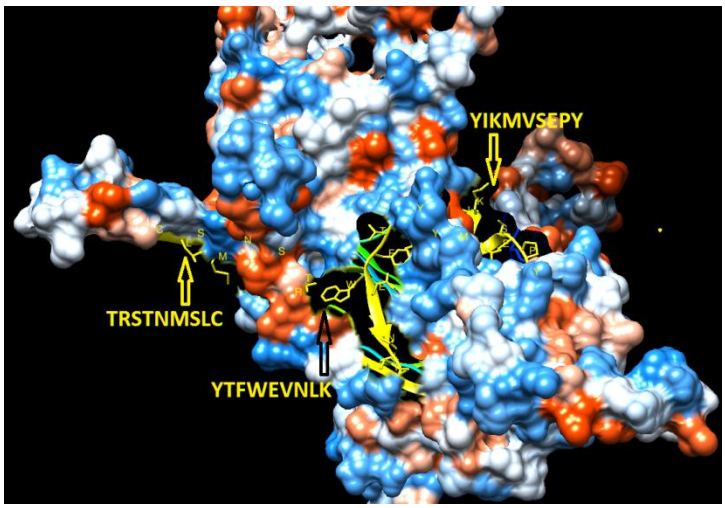

Figure (3): proposed CTL Epitopes of L1 major capsid protein

Prediction of T helper cell epitope and interaction with Mouse MHC Class II and modeling

T-cell epitopes from L1 major capsid protein were predicted using MHC-II binding prediction method; based on Consensus (SMM) with percentile rank ≤10 and IC_{50} <5000.There were 37 predicted conserved HTL epitopes found to interact with Human MHC-II alleles by high and intermediate affinity. The 9-mer peptide (core) **YIKMVSEPY** had high affinity to interact with (HLA-DRB1*04:05) allele with IC_{50}= 37 nm and intermediate affinity to interact with (HLA-DRB1*04:01, HLA-DRB5*01:01, HLA-DRB1*09:01, HLA-DRB1*15:01, HLA-DRB1*12:01 and HLA-DRB1*11:01) alleles (IC50= 75, 88, 205, 429 and 216, respectively). While **FTTLQANKS** had high affinity to interact with (HLA-DRB1*01:01) allele with IC_{50}= 8 nm and intermediate affinity to interact with (HLA-DRB1*11:01, HLA-DRB1*04:01, HLA-DRB5*01:01, HLA-DRB1*04:05 and HLA-DRB1*09:01) alleles (IC50= 73, 97, 57, 429 and 182, respectively), Figure (4). The other predicted HTL Epitope had high and intermediate affinity with selected Human alleles and the result is listed in Table (3) below.

There were several overlapping between MHC Class I epitopes and MHC Class II epitopes. These overlapping are illustrated in Table (4).

Table (3): list of the HTL epitopes which had high and intermediate binding affinity with the Human MHC Class II alleles

Epitope (core)	allele	peptide	start	end	percentile rank	IC_{50} (SMM)
*FTTLQANKS	HLA-DRB1*01:01	GAMDFTTLQANKSEV	232	246	0.96	8
	HLA-DRB1*11:01				5.07	73
	HLA-DRB1*04:01				0.83	97
	HLA-DRB5*01:01				4.69	57
	HLA-DRB1*04:05				6.31	429
	HLA-DRB1*09:01	FGAMDFTTLQANKSE	231	245	6.48	182
AMDFTTLQA	HLA-DRB1*01:01	GFGAMDFTTLQANKS	230	244	3.45	8
IKKPNNNKI	HLA-DRB1*13:02	PYFPIKKPNNNKILV	74	88	3.54	15
FFYLRREQM	HLA-DRB1*11:01	YGDSLFFYLRREQMF	268	282	1.82	17
	HLA-DPA1*02:01/DPB1*01:01				0.32	37
	HLA-DPA1*01:03/DPB1*02:01				3.97	86
	HLA-DPA1*03:01/DPB1*04:02				7.85	138
	HLA-DPA1*01/DPB1*04:01				2.88	272
	HLA-DRB5*01:01	DSLFFYLRREQMFVR	270	284	2.13	263
	HLA-DRB4*01:01				2.4	332
*YIKMVSEPY	HLA-DRB1*04:05	PDYIKMVSEPYGDSL	258	272	0.94	37
	HLA-DRB1*04:01				1.06	75
	HLA-DRB5*01:01				2.24	88
	HLA-DRB1*09:01				1.19	205
	HLA-DRB1*15:01				8.05	247
	HLA-DRB1*12:01				4.08	429
	HLA-DRB1*11:01	YPDYIKMVSEPYGDS	257	271	5.07	216
ICKYPDYIK	HLA-DRB1*04:05	ICKYPDYIKMVSEPY	254	268	0.94	38
	HLA-DRB5*01:01				2.24	83
	HLA-DRB1*09:01				2.86	374
	HLA-DRB1*15:01				7.9	410
FVTVVDTTR	HLA-DRB5*01:01	QLFVTVVDTTRSTNM	354	368	0.62	61
	HLA-DRB1*07:01				8.37	106
	HLA-DRB1*04:05				2.22	157
	HLA-DRB1*04:01				2.01	199
	HLA-DRB1*08:02	WGNQLFVTVVDTTRS	351	365	3.84	444
CWGNQLFVT	HLA-DRB5*01:01	CWGNQLFVTVVDTTR	350	364	0.62	62
LKKYTFWEV	HLA-DPA1*02:01/DPB1*01:01	DDPLKKYTFWEVNLK	464	478	3.43	77
	HLA-DPA1*01:03/DPB1*02:01				2.59	88
	HLA-DPA1*01/DPB1*04:01				1.73	216
YPDYIKMVS	HLA-DRB1*04:01	ICKYPDYIKMVSEPY	254	268	1.74	80
	HLA-DRB1*11:01	ICKYPDYIKMVSEPY	254	268	5.07	233
VDTTRSTNM	HLA-DRB1*07:01	VTVVDTTRSTNMSLC	357	371	4.28	81
DYKQTQLCL	HLA-DPA1*02:01/DPB1*01:01	ECISMDYKQTQLCLI	171	185	3.55	93
	HLA-DPA1*01/DPB1*04:01				6.85	268
YKQTQLCLI	HLA-DPA1*02:01/DPB1*01:01	SMDYKQTQLCLIGCK	174	188	2.68	98
	HLA-DPA1*03:01/DPB1*04:02				6.77	174
	HLA-DPA1*01/DPB1*04:01				5.74	281
	HLA-DRB1*07:01				4.92	362
PLKKYTFWE	HLA-DPA1*01:03/DPB1*02:01	PKEDDPLKKYTFWEV	461	475	5.54	100
	HLA-DPA1*01/DPB1*04:01				5.77	243
STNMSLCAA	HLA-DQA1*01:02/DQB1*06:02	TTRSTNMSLCAAIST	362	376	1.96	130
YKNTNFKEY	HLA-DPA1*01:03/DPB1*02:01	SETTYKNTNFKEYLR	377	391	2.03	131
	HLA-DPA1*01/DPB1*04:01				4.2	321
	HLA-DPA1*02:01/DPB1*05:01				6.2	365
NQLFVTVVD	HLA-DRB1*04:05	CWGNQLFVTVVDTTR	350	364	7.09	158
TTRSTNMSL	HLA-DRB1*07:01	VVDTTRSTNMSLCAA	359	373	7.84	158
	HLA-DQA1*01:02/DQB1*06:02	VVDTTRSTNMSLCAA	359	373	2.76	162
YFPTPSGSM	HLA-DRB1*07:01	ASSNYFPTPSGSMVT	313	327	7.74	191
NMSLCAAIS	HLA-DRB1*04:05	RSTNMSLCAAISTSE	364	378	3.87	193
LFVTVVDTT	HLA-DRB1*04:01	CWGNQLFVTVVDTTR	350	364	2.87	231
YLRREQMFV	HLA-DPA1*02:01/DPB1*01:01	YLRREQMFVRHLFNR	275	289	6.17	263
ISMDYKQTQ	HLA-DRB1*03:01	ECISMDYKQTQLCLI	171	185	0.31	254
VSGLQYRVF	HLA-DPA1*01:03/DPB1*02:01	VPKVSGLQYRVFRIH	88	102	4.27	256
	HLA-DRB1*11:01	PKVSGLQYRVFRIHL	89	103	8.82	372
WGNQLFVTV	HLA-DPA1*01:03/DPB1*02:01	NNGICWGNQLFVTVV	346	360	7.31	288
	HLA-DPA1*03:01/DPB1*04:02				9.6	490
NKILVPKVS	HLA-DRB1*11:01	PNNNKILVPKVSGLQ	81	95	4.13	296
ICWGNQLFV	HLA-DPA1*01:03/DPB1*02:01	HNNGICWGNQLFVTV	345	359	8.3	297
	HLA-DRB1*09:01				3.71	448
WEVNLKEKF	HLA-DRB1*11:01	TFWEVNLKEKFSADL	471	485	3.72	301

VNLKEKFSA	HLA-DRB3*01:01	VNLKEKFSADLDQFP	475	489	0.73	315
SMDYKQTQL	HLA-DRB1*07:01	ECISMDYKQTQLCLI	171	185	3.78	327
IKMVSEPYG	HLA-DRB1*09:01	DYIKMVSEPYGDSLF	259	273	2.63	352
	HLA-DRB1*15:01				8.57	420
FWEVNLKEK	HLA-DPA1*02:01/DPB1*01:01	KKYTFWEVNLKEKFS	468	482	4.39	403
VTVVDTTRS	HLA-DRB1*08:02	QLFVTVVDTTRSTNM	354	368	0.72	406
MDFTTLQAN	HLA-DQA1*01:02/DQB1*06:02	FGAMDFTTLQANKSE	231	245	9.72	409
FYLRREQMF	HLA-DRB1*07:01	SLFFYLRREQMFVRH	271	285	7.61	463
LVPKVSGLQ	HLA-DRB5*01:01	NNNKILVPKVSGLQY	82	96	7.03	471
KILVPKVSG	HLA-DRB1*11:01	NNKILVPKVSGLQYR	83	97	4.63	475

IC_{50}<50: High affinity, IC_{50}<500: Intermediate affinity *proposed epitope

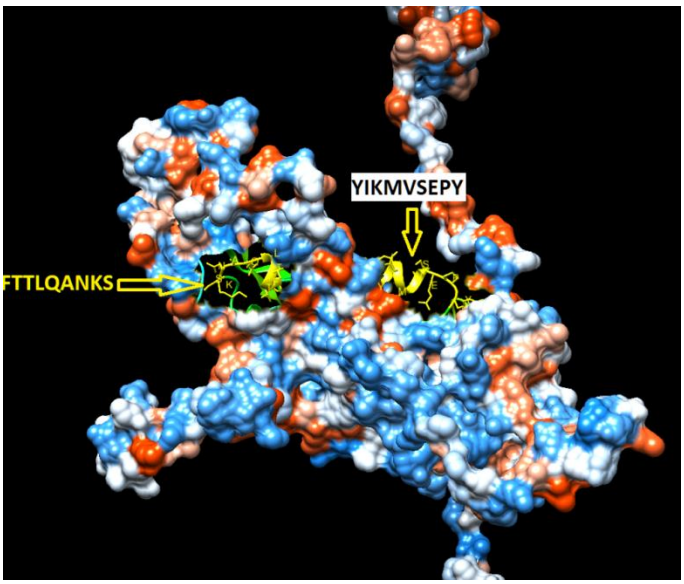

Figure (4): proposed HTL Epitopes of L1 major capsid protein

Table (4): Overlapping between MHC class I and II T cell epitopes

THL EPITOPE	peptide	start	end	CTL EPITOPE
FFYLRREQM	YGDSLF**FYLRREQMF**	268	282	FYLRREQMF
	DSLF**FYLRREQMF**VR	270	284	
ICKYPDYIK	ICKYPD**YIKMVSEPY**	254	268	YIKMVSEPY
YPDYIKMVS	ICKYPD**YIKMVSEPY**	254	268	
IKMVSEPYG	D**YIKMVSEPY**GDSLF	259	273	
ICKYPDYIK	WGNQL**FVTVVDTTR**S	351	365	FVTVVDTTR
CWGNQLFVT	CWGNQL**FVTVVDTTR**	350	364	
VTVVDTTRS	QL**FVTVVDTTR**STNM	354	368	
NQLFVTVVD	CWGNQL**FVTVVDTTR**	350	364	
LFVTVVDTT	CWGNQL**FVTVVDTTR**	350	364	
FWEVNLKEK	KK**YTFWEVNLK**EKFS	468	482	YTFWEVNLK
LKKYTFWEV	DDPLKK**YTFWEVNLK**	464	478	
	DDPLK**KYTFWEVNL**K	464	478	KYTFWEVNL
FWEVNLKEK	K**KYTFWEVNL**KEKFS	468	482	
	KKYTF**WEVNLKEKF**S	468	482	WEVNLKEKF
VDTTRSTNM	VTVVDT**TRSTNMSLC**	357	371	TRSTNMSLC
STNMSLCAA	T**TRSTNMSLC**AAIST	362	376	
TTRSTNMSL	VVDT**TRSTNMSLC**AA	359	373	
VSGLQYRVF	VPKV**SGLQYRVFR**IH	88	102	SGLQYRVFR
	PKV**SGLQYRVFR**IHL	89	103	
FYLRREQMF	SL**FFYLRREQM**FVRH	271	285	FFYLRREQM
YKNTNFKEY	SETTYK**NTNFKEYLR**	377	391	NTNFKEYLR

The underlined and highlighted residues are the 9-mer MHC class I T cell epitopes overlapping the 15-mer MHC class II T cell epitopes

Allergenicity Test

The predicted B cell epitopes and HTL epitopes that bind with different set of MHC Class II alleles by binding affinity < 500 were subjected to AllerTOP 1.0 software to avoid production of IgE antibodies as possible. The results are listed in Table (5).

Table (5): Result of Allergenicity Test of predicted B and all HTL epitopes using AllerTOP 1.0 software

HTL EPITOPES	Result	HTL EPITOPES	Result	B CELL EPITOPES	Result
*FTTLQANKS	Probable Non-Allergen	ISMDYKQTQ	Probable Non-Allergen	KSEV	Probable Non-Allergen
*AMDFTTLQA	Probable Non-Allergen	VSGLQYRVF	Probable Non-Allergen	GLQPPP	Probable Allergen
*IKKPNNNKI	Probable Allergen	WGNQLFVTV	Probable Non-Allergen	LQPPPG	Probable Allergen
*FFYLRREQM	Probable Non-Allergen	NKILVPKVS	Probable Allergen		
*YIKMVSEPY	Probable Non-Allergen	ICWGNQLFV	Probable Allergen		
*ICKYPDYIK	Probable Allergen	WEVNLKEKF	Probable Non-Allergen		
FVTVVDTTR	Probable Allergen	VNLKEKFSA	Probable Non-Allergen		
CWGNQLFVT	Probable Allergen	SMDYKQTQL	Probable Non-Allergen		
LKKYTFWEV	Probable Non-Allergen	IKMVSEPYG	Probable Non-Allergen		
YPDYIKMVS	Probable Allergen	FWEVNLKEK	Probable Allergen		
VDTTRSTNM	Probable Non-Allergen	VTVVDTTRS	Probable Non-Allergen		
DYKQTQLCL	Probable Non-Allergen	MDFTTLQAN	Probable Allergen		
YKQTQLCLI	Probable Allergen	FYLRREQMF	Probable Allergen		
PLKKYTFWE	Probable Non-Allergen	LVPKVSGLQ	Probable Allergen		
STNMSLCAA	Probable Non-Allergen	KILVPKVSG	Probable Non-Allergen		
YKNTNFKEY	Probable Allergen	YFPTPSGSM	Probable Non-Allergen		
NQLFVTVVD	Probable Allergen	NMSLCAAIS	Probable Non-Allergen		
TTRSTNMSL	Probable Non-Allergen	LFVTVVDTT	Probable Non-Allergen		
YLRREQMFV	Probable Allergen				

*peptide bind with different alleles by high affinity

Table (6): population coverage in different geographical areas

Population / Area	Class I Coverage	Population / Area	Class II Coverage
East Asia	88.09%	East Asia	72.75%
Northeast Asia	80.29%	Northeast Asia	48.67%
South Asia	82.00%	South Asia	45.30%
Southeast Asia	79.63%	Southeast Asia	46.40%
Southwest Asia	73.75%	Southwest Asia	26.21%
Europe	87.73%	Europe	61.85%
East Africa	76.48%	East Africa	53.90%
West Africa	79.52%	West Africa	52.25%
Central Africa	76.42%	Central Africa	44.25%
South Africa	85.83%	South Africa	7.65%
North America	83.04%	North America	66.78%
Central America	5.10%	Central America	15.30%
South America	75.01%	South America	29.02%
Oceania	84.36%	Oceania	56.79%
Australia	75.82%	Australia	22.76%

Table (7): The coverage in countries from which retrieved strains isolated from

Country	Class I Coverage	Country	Class II Coverage
Thailand	81.33%	Thailand	46.16%
United States	83.16%	United States	66.96%
Japan	88.83%	Japan	70.33%
China	80.03%	China	48.67%
Mexico	83.19%	Mexico	24.70%
Iran	82.29%	Iran	39.65%
India	80.57%	India	45.21%
Netherlands	0.00%	Netherlands	63.13%
Canada	0.00%	Canada	25.37%
Germany	88.62%	Germany	72.18%
Brazil	80.57%	Brazil	27.64%
Sweden	88.92%	Sweden	78.01%
Croatia	87.43%	Croatia	50.33%
Pakistan	80.74%	Pakistan	1.18%

Table (8): population coverage of each epitopes (CTL and HTL) in the world

EPITOPE	COVERAGE CLASS I	NO. OF ALLELES	EPITOPE	COVERAGE CLASS II	NO. OF ALLELES
YTFWEVNLK	35.75%	3	YIKMVSEPY	48.01%	7
TRSTNMSLC	33.31%	2	FTTLQANKS	39.02%	6
FYLRREQMF	28.67%	3	FVTVVDTTR	32.82%	5
YIKMVSEPY	11.52%	2	ICKYPDYIK	26.80%	4
NTNFKEYLR	11.03%	2	IKMVSEPYG	24.18%	2
VPLDICTSI	9.87%	2	YPDYIKMVS	21.13%	2
WEVNLKEKF	7.32%	1	VDTTRSTNM	18.23%	1
FVTVVDTTR	5.83%	1	SMDYKQTQL	18.23%	1
KYTFWEVNL	5.43%	1	FYLRREQMF	18.23%	1
IFFQMSLWL	5.43%	1	TTRSTNMSL	18.23%	2
SGLQYRVFR	5.36%	1	YFPTPSGSM	18.23%	1
IYILVITCY	3.89%	1	ISMDYKQTQ	17.84%	1
DICTSICKY	3.36%	1	AMDFTTLQA	11.53%	1
FFYLRREQM	2.88%	1	LFVTVVDTT	11.21%	1
FQMSLWLPS	1.95%	1	KILVPKVSG	10.54%	1
RLVWACVGV	1.95%	1	WEVNLKEKF	10.54%	1
Epitope set	**83.87%**		FFYLRREQM	10.54%	7
			NKILVPKVS	10.54%	1
			IKKPNNNKI	6.69%	1
			ICWGNQLFV	6.40%	2
			NQLFVTVVD	3.02%	1
			NMSLCAAIS	3.02%	1
			VTVVDTTRS	2.33%	1
			Epitope set	**81.81%**	

Analysis of the Population Coverage

HLA distribution of alleles varies among different geographic regions around the world. Thus, population coverage must be taken into a different set of alleles to cover all regions as possible and to obtain effective vaccine. Different MHC-I and II alleles that interacted with predicted CTL and THL epitopes with high and intermediate binding affinity were selected to analysis the population coverage. The results of population coverage of all epitopes in the world are listed in Table (6).

The retrieved strains used in this study were isolated from different countries. The higher coverage in these countries was shown in Germany (88.62 /72.18%), Sweden (88.92/ 78.01%) and Japan (88.83/70.33%) in MHC Class I and Class II, respectively. The coverage for other regions is shown in Table (7) and the coverage of each epitopes (CTL and HTL) in the world is shown in Table (8).

Discussion

The recent approach to develop vaccines is the peptide vaccine strategy. This strategy depends on the usage of short peptide fragments(epitopes) contained within single protein of the microbes to induce positive, desirable T-cell and B-cell mediated immune responses. In addition Peptide vaccines have the advantage of the exclusion of unnecessary antigenic load, not only participate little to the immune response, but may make situation worse by participating in induction of allergenic and /or reactogenic responses [35].

The interaction between HPV capsid proteins (L1 and L2) and surface molecules of human epithelial cells occurs during early stages of the HPV infection to gain the entry for the viral DNA; so they are ideal targets for a prophylactic vaccine [36]. We used L1 protein as a target site for designing of our vaccine against HPV type 16.

As we all Know; B-cell epitopes typically belong to one of two classes: linear (continuous or sequential) epitopes or conformational (discontinuous) epitopes. Linear epitopes are short peptides that correspond to a contiguous amino acid sequence fragment of a protein [37]. Based on this fact; we chose our predicted B cell epitopes to be linear (continuous) .These predicted epitopes scores were above thresholds in following Bepipred linear epitope prediction, Emini surface accessibility, Parker hydrophilicity, Kolaskar and Tongaonkar antigenicity and Chou and Fasman beta turn prediction methods in IEDB. This role is important in determining a potential and effective peptide antigen for B cell. As the result shown in table (1), we found only three epitopes had cutoff prediction scores above threshold scores, ^{243}**KSEV**^{246} and ^{431}**GLQPPP**^{436} or ^{432}**LQPPPG**^{437} but only ^{243}**KSEV**^{246} was free from allergenicity.

Shuchi Kaushik *et al.* (2013) found at position 7-10 in conserved region the epitope (**TTRS**), the same epitope was found in this study revealed in Table (1) but with low Antigenicity [38].

Promiscuous T-cell epitopes that can be presented by multiple human leukocyte antigens (HLAs) are prime targets for vaccine and immunotherapy development because they are effective in a high proportion of the human population [39]. We chose the most common HLA-A and HLA-B alleles for prediction of CTL epitopes. So, according to Table (2), we found that ^{470}**YTFWEVNLK**^{478} epitope had high binding affinity in interaction with (HLA-

A*68:01 and HLA-A*11:01) alleles and intermediate affinity to bind with HLA-A*03:01 and predicted Probable Non Allergen by Allergenicity Test while [383]**NTNFKEYLR**[391] had high affinity to interact with (HLA-A*68:01 and HLA-A*31:01) alleles but predicted as Probable Allergen. Other CTL epitopes **FVTVVDTTR**, **RLVWACVGV**, **FQMSLWLPS** and **SGLQYRVFR** had ability to interact with one allele by high affinity and predicted Probable Non Allergen except **FVTVVDTTR** epitope which was considered HTL epitope also according to Table (3). Our analysis has shown the proposed CTL epitopes after performing Allrgenisity Test were **TRSTNMSLC, YTFWEVNLK and YIKMVSEPY.**

As in Table (3), we found that HTL epitope [260]**YIKMVSEPY**[268] and [236]**FTTLQANKS**[244] had high and intermediate affinities to interact with different MHC II alleles were predicted in Allergenicity Test as Probable Non-Allergen as shown in Table (5). [260]**YIKMVSEPY**[268] epitope was a common epitope in both MHC class I and MHC class II epitopes so it is probably the best predicted epitope. Although [254]**ICKYPDYIK**[262] had high affinity to interact with (HLA-DRB1*04:05) allele and intermediate affinity to interact with 4 alleles but could be an Allergen according to allergenicity prediction test.

An allergic reaction occurs when a susceptible organism is re-exposed to a specific allergen. The allergen-specific HTL drive the B cells to produce IgE, which binds to mast cells, basophils and activated eosinophils [31] .Thus; we subjected all predicted B and HTL cells to allergenicity test. We represented the results in table (5). Also, we tested all predicted CTL cells for more confirmation to exclude the allergenicity if possible but we did not include the results in above mentioned table.

Population coverage analysis is conducted in order to develop an effective vaccine based on the fact that HLA genes are highly polymorphic; they have many alleles. Those alleles vary among different geographic regions around the world. So, according to Table (7) and among 16 geographical areas we found the higher population coverage in MHC class I in East Asia (88.09%), followed by Europe (87.73%) then South Africa (85.83%),Oceania (84.36%), North America (83.04%), South Asia (82.00%) and Northeast Asia (80.29%), while the lower coverage was found in Central America (5.10%). According to Sanjosé S.D *et. al.* (2007); eastern Africa registered the highest adjusted HPV prevalence (31·6%, 29·5–33·8) [40]. In our study we found the population coverage in East Africa was very high in MHC class I (76.48%).thus the results agree.

In MHC class II, we found the maximum coverage in East Asia (72.75%), North America (66.78%) and Europe (61.85%) while the minimum was in South Africa (7.65%). We applied the coverage in countries from which retrieved strains were isolated and we found the higher coverage among these countries in Germany (88.62 /72.18%), Sweden (88.92/ 78.01%) and Japan (88.83/70.33%) in MHC Class I/Class II, respectively.

We observed several overlaps between MHC class I and II in T cell epitopes of M polyprotein and our results illustrated in Table (4). These overlaps could increase the possibility of antigen presentation to immune cells via both MHC class I and II pathways [41].

To sum up our findings the most interesting epitopes that supposed to be used as prophylactic peptide vaccines against HPV are for B cell [243]**KSEV**[246], MHC-I [363]**TRSTNMSLC**[371], [470]**YTFWEVNLK**[478], [260]**YIKMVSEPY**[268] and MHC-II [236]**FTTLQANKS**[244] and [260]**YIKMVSEPY**[268] which was probably the best epitope predicted.

Conclusion

This present study involved the usage of immunoinformatics in vaccine prediction. We used these approaches for prediction of antigenic determinants in the protein sequence of L1 major capsid protein of HPV virus genotype 16 without using their cultures. These approaches of computational immunology may now drastically reduce the time for the identification promiscuous antigenic peptides. Our Predicted B and T cell epitopes were based on the predictive and analytic tool (IEDB-AR). These epitopes could serve as a useful diagnostic reagent for evaluating T-cell responses in the context of natural infection and also might be helpful for designing a subunit vaccine against HPV virus. We can confirm our findings by adding complementary steps of both in vitro and in vivo studies to support this universal predicted vaccine for this type of HPV.

Acknowledgments

Authors would like to thanks African City of Technology members for their assistance and help.

Competing Interests

The authors declare that they have no competing interests.

References

[1] Kajitani N., Schwartz S. RNA binding proteins that control human papillomavirus gene expression. Biomolecules 2015; 5(2):758-74. DOI: 10.3390/biom5020758.

[2] Leto Md, Santos Júnior GF, Porro AM, Tomimori J. Human papillomavirus infection: etiopathogenesis, molecular biology and clinical manifestations. An Bras Dermatol 2012; 86(2):306-17. [PMID: 21603814]

[3] Markowitz LE, Dunne EF, Saraiya M, Chesson HW, Curtis CR, Gee J, *et al*. Human papillomavirus vaccination recommendations of the advisory committee on immunization practices. CDC,MMWR 2014; 63(RR05):1-30. Available at: http:// www.cdc.gov/ mmwr/preview/ mmwrhtml/rr6305a1.htm

[4] De Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology* 2004; 324(1):17-27. [PMID: 15183049].

[5] Tzu-Yu Liu, Waleed M. Hussein, Istvan Toth, Mariusz Skwarczynski. Advances in peptide-based human papillomavirus therapeutic vaccines. *Current Topics in Medicinal Chemistry* 2012;12(14),1581-1592. DOI: 10.2174/156802612802652402.

[6] Bang HB, Lee YH, Lee YJ, Jeong KJ. High-level production of human papillomavirus (HPV) type 16 L1 in Escherichia coli. *J Microbiol*. *Biotechnol*. 2016; 26(2):356-363. [PMID: 26608168].
[7] Moody CA, Laimins LA. Human papillomavirus oncoprotein: pathways to transformation. *Nat Rev Cancer.* 2010 10(8):550-60. [PMID: 20592731].
[8] Patel MC, Patkar KK, Basu A, Mohandas KM, Mukhopadhyaya R. Production of immunogenic human papillomavirus-16 major capsid protein derived virus like particles. *Indian J Med Res.* 2009;130(3): 213-218. [PMID: 19901429].
[9] A. Alba , M. Cararach, C. Rodriguez-Cerdeira. The human papillomavirus (HPV) in human pathology: Description, pathogenesis, oncogenic role, epidemiology and detection techniques. *The Open Dermatology Journal* 2009; 3: 90-102. DOI: 10.2174/1874372200903010090.
[10] Daniel Tena Gomez and Juana Lopez Santos. Human papillomavirus infection and cervical cancer: pathogenesis and epideminology. *Communicating Current Research and Educational Topics and Trends in Applied Microbiology* 2007;1: 680-88.
[11] Eileen M. Burd. Human papillomavirus and cervical cancer. *Clin Microbiol Rev.* 2003 ; 16(1): 1–17. DOI: 10.1128/CMR.16.1.1-17.2003. [PMCID: PMC145302].
[12] Meloni A, Pilia R, Campagna M, Usai A, Masia G, Caredda V, Coppola RC. Prevalence and molecular epidemiology of human papillomavirus infection in Italian women with cervical cytological abnormalities. *J Public Health Res*. 2014; 3(1):157. [PMID: 25170506].
[13] Dunne EF, Unger ER, Sternberg M, McQuillan G, Swan DC, Patel SS, Markowitz LE. Prevalence of HPV infection among females in the United States. *JAMA.* 2007;297(8):813-9. [PMID: 17327523].
[14] Markowitz L.E., Dunne E.F., Saraiya M., Chesson H.W., Curtis C.R., Gee J. *et. al.* Human Papillomavirus Vaccination Recommendations of the Advisory Committee on Immunization Practices (ACIP). Recommendations and Reports, CDC 2014;63 **:**(RR05);1-30. Available at: http://www. cdc.gov/mmwr/ preview/mmwrhtml/rr6305a1.htm
[15] Villa LL, Costa RL, Petta CA, Petta CA, Andrade RP, Ault KA, Giuliano AR, Wheeler CM *et al.* Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncol* 2005;6(5):271–8. PMID: 15863374 DOI: 10.1016/S1470-2045(05)70101-7
[16] Harper DM, Franco EL, Wheeler CM, Moscicki A.B.,Romanowski B.,Roteli-Martins C.M. *et al.* Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial. *Lancet* 2006;367:1247–55. DOI:10.1016/S0140-6736(06)68439-0.
[17] Dr. J. Langley, Dr. B. Warshawsky, Dr. S. Ismail , Dr. N. Crowcroft , Ms. A. Hanrahan, Dr. B. Henry , Dr. D Kumar, Dr. A. McGeer, Dr. S. McNeil, Dr. B. Seifert, Dr. C. Quach-Thanh, Dr. D. Skowronski, Dr. B. Tan, Dr. C. Cooper. Update on Human Papillomavirus (HPV) Vaccines. *CCDR*. 2012; 38: 1481-8531.
[18] Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. *Pharmacogenomics* 2009;10(5):837-52. [PMID: 19450131].
[19] Sirskyj D, Diaz-Mitoma F, Golshani A, Kumar A, Azizi A. innovative bioinformatic approaches for developing peptide-based vaccines against hypervariable viruses. *Immunol Cell Biol.* 2011;89(1):81-9. [PMID: 20458336].
[20] Oany AR, Emran AA, and Jyoti TP. Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach; *Drug Des Devel Ther.*2014; 8: 1139-49. DOI: 10.2147/DDDT.S67861.
[21] Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol*. 2010, May;59(3):307-21. [PMID: 20525638].
[22] Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser.* 1999;41:95-98.
[23] Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006; 2:2. [PMID: 16635264].
[24] Bui H. H,Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2007; 8(1):361. DOI: 10.1186/1471-2105-8-361.
[25] Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 1990;276(1-2):172-4. [PMID: 1702393].
[26] Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol.* 1985; 55:836-39. [PMID: 2991600].
[27] Parker JM, Guo D, Hodges RS. 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.* 1986;25:5425-32. [PMID: 2430611].
[28] Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 2005; 6:132. [PMID: 15927070].
[29] Lundegaard C, Lund O, and Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics*. 2008; 24:1397-98. [PMID: 18413329].
[30] Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics*. 2007;8:238. [PMID: 17608956].
[31] Ivan Dimitrov, Lyudmila Naneva, Irini Doytchinova, Ivan Bangov. AllergenFP: allergenicity prediction by descriptor fingerprints. *BIOINFORMATICS* 2014; (30)6: 846–851 DOI:10.1093/bioinformatics/btt619
[32] Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics.* 2006 17;7:153. [PMID: 16545123].
[33] Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, *et al.* Template-based protein structure modeling using the RaptorX web server. Nature Protocols. 2012;7(8):1511-22. DOI:10.1038/nprot.2012.085.
[34] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25(13):1605-12. [PMID: 15264254].
[35] Weidang Li , Medha D. Joshi , Smita Singhania , Kyle H. Ramsey and Ashlesh K. Murthy , Peptide Vaccine: Progress and Challenges, *Vaccines* 2014, *2*(3), 515-536; DOI:10.3390/vaccines2030515.
[36] Tomar A., Kushwah A. Advances in human papilloma virus vaccines:a review. *Int J Basic Clin Pharmacol*. 2014; 3(1): 37-43. DOI:10.5455/2319-2003.ijbcp20140237.
[37] EL-Manzalawy Y. & Honavar V. Recent advances in B-cell epitope prediction methods. *Immunome Research* 2010, 6(Suppl 2):S2. DOI: 10.1186/1745-7580-6-S2-S2.
[38] Kaushik S., Shrivastav V.K. , Shrivastav A., Jana A.M.. Silico vaccine design against the target l1 binding protein of human papillomavirus, an etiological agent of cervical cancer, using bioinformatics tools. *JPSR*. 2013; 4(12): 4758-4762.
[39] Wang S., Guo L., Liu D., Liu W.. HLAsupE: an integrated database of HLA supertype-specific epitopes to aid in the development of vaccines with broad coverage of the human population. *BMC Immunol*. 2016; 17: 17. [PMCID: PMC4910211].
[40] Sanjosé S.D., Diaz M., Castellsagué X., Clifford G., Bruni L., Muñoz N., Bosch F.X. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. *Lancet Infect Dis.* 2007; 7 (7): 453–59. DOI: http://dx.doi.org/10.1016/S1473-3099(07)70158-5.
[41] Osman MM, ElAmin EE, Al-Nour MY, Alam SS, Adam RS, Ahmed AA, *et al.* In Silico Design of Epitope Based Peptide Vaccine against Virulent Strains of (HN)-Newcastle Disease Virus (NDV) in Poultry Species. *IJMCR* 2016;4 (Sept/Oct 2016 issue):868-78. Available at: http://ijmcr.com/category/ijmcr/vol-4-sept-oct-2016/

Supplementary Data

Table (A): Retrived sequences of L1 major capsid protien of HPV type 16

Accession No.	Country	Collection date
*NP_041332.1	unknown	unknown
AJW82191.1	Iran	2013
AJW82216.1	Iran	2013
AJW82215.1	Iran	2013
AJW82214.1	Iran	2013
AJW82212.1	Iran	2013
AJW82211.1	Iran	2013
AJW82213.1	Iran	2013
AJW82210.1	Iran	2013
AJW82209.1	Iran	2013
AJW82208.1	Iran	2013
AJW82207.1	Iran	2013
AJW82206.1	Iran	2013
AJW82205.1	Iran	2013
AJW82203.1	Iran	2013
AJW82202.1	Iran	2013
AJW82201.1	Iran	2013
AJW82200.1	Iran	2013
AJW82199.1	Iran	2013
AJW82204.1	Iran	2013
AJW82198.1	Iran	2013
AJW82197.1	Iran	2013
AJW82196.1	Iran	2013
AJW82195.1	Iran	2013
AJW82194.1	Iran	2013
AJW82193.1	Iran	2013
AJW82192.1	Iran	2013
AJW82190.1	Iran	2013
BAN63249.1	unknown	unknown
AEZ01744.1	Thailand	2010
AEZ01736.1	Thailand	2010
AEZ01721.1	Thailand	2010
BAO18726.1	unknown	unknown
BAO18718.1	unknown	unknown
BAO18710.1	unknown	unknown
BAO18702.1	unknown	unknown
BAO18694.1	unknown	unknown
BAO18686.1	unknown	unknown
BAO18678.1	unknown	unknown
AFP44614.1	USA	unknown
AFP44606.1	USA	unknown
AFP44590.1	USA	unknown
AFP44582.1	USA	unknown
AFP44566.1	USA	unknown
AFP44558.1	USA	unknown
AFP44534.1	USA	unknown
AFP44478.1	USA	unknown
AFP44462.1	USA	unknown
AFP44438.1	USA	unknown
AFP44406.1	USA	unknown
AFP44374.1	USA	unknown
AFP44294.1	USA	unknown
AFP44278.1	USA	unknown
AFP44270.1	USA	unknown
AFP44254.1	USA	unknown
AFP44246.1	USA	unknown
AFP44230.1	USA	unknown
AFP44206.1	USA	unknown
AFP44166.1	USA	unknown
AFP44150.1	USA	unknown
BAN63273.1	Japan	unknown
BAN63265.1	Japan	unknown

BAN63257.1	Japan	unknown
BAN63241.1	Japan	unknown
BAN63233.1	Japan	unknown
BAN63225.1	Japan	unknown
AEZ01752.1	Thailand	2010
AEZ01729.1	Thailand	2010
AEZ01713.1	Thailand	2010
AEZ01697.1	Thailand	2010
AEZ01705.1	Thailand	2010
ACS92699.1	Thailand	2008
ACS92683.1	Thailand	2008
ACS92691.1	Thailand	2008
ACS92675.1	Thailand	2008
ACS92667.1	Thailand	2008
ACS92659.1	Thailand	2008
ACS92651.1	Thailand	2008
ACA14209.1	china	unknown
ACA14201.1	china	unknown
ACA14193.1	china	unknown
AGK28592.1	china	2012
AFP44646.1	unknown	unknown
AFP44638.1	unknown	unknown
AFP44630.1	unknown	unknown
AFP44622.1	unknown	unknown
AFP44598.1	unknown	unknown
AFP44574.1	unknown	unknown
AFP44550.1	unknown	unknown
AFP44542.1	unknown	unknown
AFP44526.1	unknown	unknown
AFP44518.1	unknown	unknown
AFP44510.1	unknown	unknown
AFP44502.1	unknown	unknown
AFP44430.1	unknown	unknown
AFP44422.1	unknown	unknown
AFP44414.1	unknown	unknown
AFP44398.1	unknown	unknown
AFP44390.1	unknown	unknown
AFP44382.1	unknown	unknown
AFP44366.1	unknown	unknown
AFP44358.1	unknown	unknown
AFP44350.1	unknown	unknown
AFP44342.1	unknown	unknown
AFP44334.1	unknown	unknown
AFP44326.1	unknown	unknown
AFP44318.1	unknown	unknown
AFP44310.1	unknown	unknown
AFP44302.1	unknown	unknown
AFP44286.1	unknown	unknown
AFP44262.1	unknown	unknown
AFP44238.1	unknown	unknown
AFP44222.1	unknown	unknown
AFP44214.1	unknown	unknown
AFP44198.1	unknown	unknown
AFP44190.1	unknown	unknown
AFP44182.1	unknown	unknown
AFP44174.1	unknown	unknown
AFP44158.1	unknown	unknown
AFP44142.1	unknown	unknown
AFP44134.1	unknown	unknown
AFP44126.1	unknown	unknown
ACG75893.1	china	2006
AAO85415.1	unknown	unknown
ACN91181.1	Mexico	2005
ACN91180.1	Mexico	2005
ACN91179.1	Mexico	2005
ACN91176.1	Mexico	2005
ACN91175.1	Mexico	2005
ACN91171.1	Mexico	2005
ACN91170.1	Mexico	2005
ACN91168.1	Mexico	2005
ACN91161.1	Mexico	2005
ACN91160.1	Mexico	2005
ACN91158.1	Mexico	2005
AAD33259.1	unknown	unknown
AIQ82846.1	Iran	2013
AIQ82845.1	Iran	2013
AIQ82844.1	Iran	2013
AIQ82843.1	Iran	2013
AIQ82842.1	Iran	2013
AIQ82841.1	Iran	2013
AIQ82840.1	Iran	2013
AIQ82839.1	Iran	2013
AIQ82838.1	Iran	2013

AIQ82837.1	Iran	2013
AIQ82836.1	Iran	2013
AIQ82835.1	Iran	2013
AIQ82834.1	Iran	2013
AIQ82833.1	Iran	2013
AIQ82832.1	Iran	2013
AIQ82831.1	Iran	2013
AIQ82830.1	Iran	2013
AIQ82829.1	Iran	2013
AIQ82828.1	Iran	2013
AIQ82827.1	Iran	2013
AIQ82826.1	Iran	2013
AIQ82825.1	Iran	2013
AIQ82824.1	Iran	2013
AIQ82823.1	Iran	2013
AIQ82822.1	Iran	2013
AIQ82821.1	Iran	2013
AIQ82820.1	Iran	2013
AIQ82819.1	Iran	2013
AIQ82818.1	Iran	2013
AIQ82817.1	Iran	2013
AIQ82816.1	Iran	2013
ACN91182.1	Mexico	2005
ACN91178.1	Mexico	2005
ACN91177.1	Mexico	2005
ACN91174.1	Mexico	2005
ACN91173.1	Mexico	2005
ACN91172.1	Mexico	2005
ACN91167.1	Mexico	2005
ACN91166.1	Mexico	2005
ACN91165.1	Mexico	2005
ACN91164.1	Mexico	2005
ACN91163.1	Mexico	2005
ACN91162.1	Mexico	2005
ACN91159.1	Mexico	2005
ACN91157.1	Mexico	2005
AHK23262.1	China	2013
ACL12317.1	china	2007
AMS04051.1	India	2009
AMS04043.1	India	2009
AMS04035.1	India	40138
AMS04026.1	India	2004
AMS04019.1	India	2010
AMS04003.1	India	2011
AML83910.1	India	2004
ANA05557.1	Netherlands	2011
ANA05556.1	Netherlands	2011
ANA05555.1	Netherlands	2011
ANA05554.1	Netherlands	2011
ANA05553.1	Netherlands	2011
ANA05552.1	Netherlands	2011
ANA05551.1	Netherlands	2011
ANA05550.1	Netherlands	2011
ANA05549.1	Netherlands	2011
ANA05548.1	Netherlands	2011
ANA05547.1	Netherlands	2011
ANA05546.1	Netherlands	2011
ANA05545.1	Netherlands	2011
ANA05544.1	Netherlands	2011
ANA05543.1	Netherlands	2011
ANA05542.1	Netherlands	2011
ANA05541.1	Netherlands	2011
ANA05540.1	Netherlands	2011
ANA05539.1	Netherlands	2011
ANA05538.1	Netherlands	2011
ANA05537.1	Netherlands	2011
ANA05536.1	Netherlands	2011
ANA05535.1	Netherlands	2011
ANA05534.1	Netherlands	2011
ANA05533.1	Netherlands	2011
ANA05532.1	Netherlands	2011
ANA05531.1	Netherlands	2011
ANA05530.1	Netherlands	2011
ANA05529.1	Netherlands	2011
ANA05528.1	Netherlands	2011
ANA05527.1	Netherlands	2011
ANA05526.1	Netherlands	2011
ANA05525.1	Netherlands	2011
ANA05524.1	Netherlands	2011
ANA05523.1	Netherlands	2011
ANA05522.1	Netherlands	2011
ANA05521.1	Netherlands	2011
ANA05520.1	Netherlands	2011

ANA05519.1	Netherlands	2011
ANA05518.1	Netherlands	2011
ANA05517.1	Netherlands	2011
ANA05516.1	Netherlands	2011
ANA05515.1	Netherlands	2011
ANA05514.1	Netherlands	2011
ANA05513.1	Netherlands	2011
ANA05512.1	Netherlands	2011
ANA05511.1	Netherlands	2011
ANA05510.1	Netherlands	2011
ANA05509.1	Netherlands	2011
ANA05508.1	Netherlands	2011
ANA05507.1	Netherlands	2011
ANA05506.1	Netherlands	2011
ANA05505.1	Netherlands	2011
ANA05504.1	Netherlands	2011
ANA05503.1	Netherlands	2011
ANA05502.1	Netherlands	2011
ANA05501.1	Netherlands	2011
ANA05500.1	Netherlands	2011
ANA05499.1	Netherlands	2011
ANA05498.1	Netherlands	2011
ANA05497.1	Netherlands	2011
ANA05496.1	Netherlands	2011
ANA05495.1	Netherlands	2011
ANA05494.1	Netherlands	2011
ANA05493.1	Netherlands	2011
ANA05492.1	Netherlands	2011
ANA05491.1	Netherlands	2011
ANA05490.1	Netherlands	2011
ANA05489.1	Netherlands	2011
ANA05488.1	Netherlands	2011
ANA05487.1	Netherlands	2011
ANA05486.1	Netherlands	2011
ANA05485.1	Netherlands	2011
ANA05484.1	Netherlands	2011
ANA05483.1	Netherlands	2011
ANA05482.1	Netherlands	2011
ANA05481.1	Netherlands	2011
ANA05480.1	Netherlands	2011
ANA05479.1	Netherlands	2011
ANA05478.1	Netherlands	2011
ANA05477.1	Netherlands	2011
ANA05476.1	Netherlands	2011
ANA05475.1	Netherlands	2011
ANA05474.1	Netherlands	2011
ANA05473.1	Netherlands	2011
ANA05472.1	Netherlands	2011
ANA05471.1	Netherlands	2011
ANA05470.1	Netherlands	2011
ANA05469.1	Netherlands	2011
ANA05468.1	Netherlands	2011
ANA05467.1	Netherlands	2009
ANA05466.1	Netherlands	2009
ANA05465.1	Netherlands	2009
ANA05464.1	Netherlands	2009
ANA05463.1	Netherlands	2009
ANA05462.1	Netherlands	2009
ANA05461.1	Netherlands	2009
ANA05460.1	Netherlands	2009
ANA05459.1	Netherlands	2009
ANA05458.1	Netherlands	2009
ANA05457.1	Netherlands	2011
ANA05456.1	Netherlands	2011
ANA05455.1	Netherlands	2011
ANA05454.1	Netherlands	2011
ANA05453.1	Netherlands	2011
ANA05452.1	Netherlands	2011
ANA05451.1	Netherlands	2011
ANA05450.1	Netherlands	2011
ANA05449.1	Netherlands	2011
ANA05448.1	Netherlands	2011
ANA05447.1	Netherlands	2009
ANA05446.1	Netherlands	2009
ANA05445.1	Netherlands	2009
ANA05444.1	Netherlands	2009
ANA05443.1	Netherlands	2009
ANA05442.1	Netherlands	2009
ANA05441.1	Netherlands	2009
ANA05440.1	Netherlands	2009
ANA05439.1	Netherlands	2009
ANA05438.1	Netherlands	2009
ANA05437.1	Netherlands	2009

ANA05436.1	Netherlands	2009
ANA05435.1	Netherlands	2009
ANA05434.1	Netherlands	2009
ANA05433.1	Netherlands	2009
ANA05432.1	Netherlands	2009
ANA05431.1	Netherlands	2009
ANA05430.1	Netherlands	2009
ANA05429.1	Netherlands	2009
ANA05428.1	Netherlands	2009
ANA05427.1	Netherlands	2009
ANA05426.1	Netherlands	2009
ANA05425.1	Netherlands	2009
ANA05424.1	Netherlands	2009
ANA05423.1	Netherlands	2009
ANA05422.1	Netherlands	2009
ANA05421.1	Netherlands	2009
ANA05420.1	Netherlands	2011
ANA05419.1	Netherlands	2011
ANA05418.1	Netherlands	2009
ANA05417.1	Netherlands	2009
ANA05416.1	Netherlands	2009
ANA05415.1	Netherlands	2009
ANA05414.1	Netherlands	2009
ANA05413.1	Netherlands	2009
ANA05412.1	Netherlands	2009
ANA05411.1	Netherlands	2009
ANA05410.1	Netherlands	2009
ANA05409.1	Netherlands	2009
ANA05408.1	Netherlands	2009
ANA05407.1	Netherlands	2009
ANA05406.1	Netherlands	2009
ANA05405.1	Netherlands	2009
ANA05404.1	Netherlands	2009
ANA05403.1	Netherlands	2009
ANA05402.1	Netherlands	2009
ANA05401.1	Netherlands	2009
ANA05400.1	Netherlands	2009
ANA05399.1	Netherlands	2009
ANA05398.1	Netherlands	2009
ANA05397.1	Netherlands	2009
ANA05396.1	Netherlands	2009
ANA05395.1	Netherlands	2009
ANA05394.1	Netherlands	2009
ANA05393.1	Netherlands	2009
ANA05392.1	Netherlands	2009
ANA05391.1	Netherlands	2009
ANA05390.1	Netherlands	2009
ANA05389.1	Netherlands	2009
ANA05388.1	Netherlands	2009
ANA05387.1	Netherlands	2009
ANA05386.1	Netherlands	2009
ANA05385.1	Netherlands	2009
ANA05384.1	Netherlands	2009
ANA05383.1	Netherlands	2009
ANA05382.1	Netherlands	2009
ANA05381.1	Netherlands	2009
ANA05380.1	Netherlands	2009
ANA05379.1	Netherlands	2009
ANA05378.1	Netherlands	2009
ANA05377.1	Netherlands	2009
ANA05376.1	Netherlands	2009
ANA05375.1	Netherlands	2009
ANA05374.1	Netherlands	2009
ANA05373.1	Netherlands	2009
ANA05372.1	Netherlands	2009
ANA05371.1	Netherlands	2009
ANA05370.1	Netherlands	2009
ANA05369.1	Netherlands	2009
ANA05368.1	Netherlands	2009
ANA05367.1	Netherlands	2009
ANA05366.1	Netherlands	2009
ANA05365.1	Netherlands	2011
ANA05364.1	Netherlands	2011
ANA05363.1	Netherlands	2009
AAQ10410.1	unknown	unknown
ACV84007.1	Canada	unknown
ACV84006.1	Canada	unknown
ACV84005.1	Canada	unknown
ACV84003.1	Canada	unknown
ACV53981.1	Canada	unknown
ACV53980.1	Canada	unknown
ACV53979.1	Canada	unknown
ACV53977.1	Canada	unknown

ACV53976.1	Canada	unknown
ACV53975.1	Canada	unknown
ACV53974.1	Canada	unknown
ACV53973.1	Canada	unknown
ACV53972.1	Canada	unknown
ACV53971.1	Canada	unknown
ACV53970.1	Canada	unknown
ACV53969.1	Canada	unknown
ACV53968.1	Canada	unknown
ACV53966.1	Canada	unknown
ACV53967.1	Canada	unknown
ACV53965.1	Canada	unknown
ACV53964.1	Canada	unknown
ACV53963.1	Canada	unknown
ACV53962.1	Canada	unknown
ACV53961.1	Canada	unknown
ACV53960.1	Canada	unknown
ACV53959.1	Canada	unknown
ACV53958.1	Canada	unknown
ACV53957.1	Canada	unknown
ACV53956.1	Canada	unknown
AAQ10727.1	unknown	unknown
AAQ10719.1	unknown	unknown
P03101.3	unknown	unknown
AAD22111.1	unknown	unknown
AAD22113.1	unknown	unknown
AAC09293.1	unknown	unknown
AAD22112.1	unknown	unknown
AAA46943.1	unknown	unknown
ABG75583.1	unknown	unknown
AAC09292.1	unknown	unknown
AAC61736.1	unknown	unknown
AAA92892.1	unknown	unknown
AAK77000.1	unknown	unknown
ABQ24218.1	China	unknown
AGC65525.1	Pakistan	2010
AMN09986.1	China	2014
AMN09985.1	China	2014
AMN09984.1	China	2014
AAC31789.1	unknown	unknown
AFN85025.1	Croatia	2003
AFN85014.1	Croatia	2001
AFN85006.1	Croatia	2001
AAV91675.1	unkown	unkown
AAV91651.1	unkown	unkown
AAD38523.1	China	unkown
AAV91691.1	unknown	unknown
AAV91683.1	unknown	unknown
AAV91667.1	unknown	unknown
AAV91659.1	unknown	unknown
ABV21641.1	Germany	unknown
ALB35307.1	Brazil	2011
ALB35300.1	Brazil	2011
ALB35284.1	Brazil	2011
ALB35268.1	Brazil	2010
AAO19439.1	South Africa	2007
AFS33335.1	China	2010
ADH94047.1	Brazil	2005
ALB35315.1	Brazil	unknown
ALB35292.1	Brazil	2011
ALB35260.1	Brazil	2010
ALB35252.1	Brazil	2010
ALB35244.1	Brazil	2010
ALB35236.1	Brazil	2010
ABF06542.1	Germany	unknown
ALT54634.1	Brazil	unknown
ALT54602.1	Brazil	unknown
ALB35322.1	Brazil	unknown
AIU47931.1	Sweden	2011
AFS33364.1	China	2011
AFS33363.1	China	2011
AFS33362.1	China	2011
AFS33361.1	China	2011
AFS33360.1	China	2011
AFS33359.1	China	2011
AFS33358.1	China	2011
AFS33357.1	China	2011
AFS33356.1	China	2011
AFS33355.1	China	2011
AFS33354.1	China	2011
AFS33353.1	China	2011
AFS33353.1	China	2011
AFS33351.1	China	2011

AFS33350.1	China	2011
AFS33349.1	China	2011
AFS33348.1	China	2011
AFS33347.1	China	2011
AFS33346.1	China	2011
AFS33345.1	China	2011
AFS33344.1	China	2011
AFS33343.1	China	2011
AFS33342.1	China	2011
AFS33341.1	China	2011
AFS33340.1	China	2010
AFS33339.1	China	2010
AFS33338.1	China	2010
AFS33337.1	China	2010
AFS33336.1	China	2010
AFS33334.1	China	2010
ALT54618.1	Brazil	unknown
ALB35336.1	Brazil	unknown
ALB35329.1	Brazil	unknown
AAZ81569.1	India	unknown
ACV84004.1	Canada	unknown
ACV84002.1	Canada	unknown
ACV53978.1	Canada	unknown
AAO15704	unknown	unknown
ANA05603.1	Netherlands	2011
ANA05602.1	Netherlands	2011
ANA05601.1	Netherlands	2011
ANA05600.1	Netherlands	2011
ANA05599.1	Netherlands	2011
ANA05598.1	Netherlands	2011
ANA05597.1	Netherlands	2011
ANA05596.1	Netherlands	2011
ANA05595.1	Netherlands	2011
ANA05594.1	Netherlands	2011
ANA05593.1	Netherlands	2011
ANA05592.1	Netherlands	2011
ANA05591.1|	Netherlands	2011
ANA05590.1	Netherlands	2011
ANA05589.1	Netherlands	2011
ANA05588.1	Netherlands	2011
ANA05587.1	Netherlands	2011
ANA05586.1	Netherlands	2011
ANA05585.1	Netherlands	2011
ANA05584.1	Netherlands	2011
ANA05583.1	Netherlands	2011
ANA05582.1	Netherlands	2011
ANA05581.1	Netherlands	2011
ANA05580.1	Netherlands	2011
ANA05579.1	Netherlands	2011
ANA05578.1	Netherlands	2011
ANA05577.1	Netherlands	2011
ANA05576.1	Netherlands	2011
ANA05575.1	Netherlands	2011
ANA05574.1	Netherlands	2011
ANA05573.1	Netherlands	2011
ANA05572.1	Netherlands	2011
ANA05571.1	Netherlands	2011
ANA05570.1	Netherlands	2011
ANA05562.1	Netherlands	2011
ANA05565.1	Netherlands	2011
ANA05564.1	Netherlands	2011
ANA05563.1	Netherlands	2011
ANA05561.1	Netherlands	2011
ANA05566.1	Netherlands	2011
ANA05567.1	Netherlands	2011
ANA05568.1	Netherlands	2011
ANA05569.1	Netherlands	2011
ANA05560.1	Netherlands	2011
ANA05559.1	Netherlands	2011

**Ref*. strain