bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Original Article

Design of multi epitope-based peptide vaccine against E protein of human COVID-19: An immunoinformatics approach

Miyssa I. Abdelmageed[1], Abdelrahman H. Abdelmoneim[2], Mujahed I. Mustafa[3*], Nafisa M. Elfadol[4], Naseem S. Murshed[5], Shaza W. Shantier[6] and Abdelrafie M. Makhawi[3]

[1]Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan.
[2]Faculty of Medicine, Alneelain University, Khartoum, Sudan.
[3]Department of Biotechnology, University of Bahri, Khartoum, Sudan.
[4]Department of Molecular biology, National University Biomedical Research Institute, National University, Khartoum, Sudan.
[5]Department of Microbiology, International University of Africa, Khartoum, Sudan.
[6]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan.

*Corresponding author: Mujahed I. Mustafa; mujahedibrahim44@gmail.com

Abstract

Background: New endemic disease has been spread across Wuhan City, China on December 2019. Within few weeks, the World Health Organization (WHO) announced a novel coronavirus designated as coronavirus disease 2019 (COVID-19). In late January 2020, WHO declared the outbreak of a "public-health emergency of international concern" due to the rapid and increasing spread of the disease worldwide. Currently, there is no vaccine or approved treatment for this emerging infection; thus the objective of this study is to design a multi epitope peptide vaccine against COVID-19 using immunoinformatics approach.
Method: Several techniques facilitating the combination of immunoinformatics approach and comparative genomic approach were used in order to determine the potential peptides for designing the T cell epitopes-based peptide vaccine using the envelope protein of 2019-nCoV as a target.
Results: Extensive mutations, insertion and deletion were discovered with comparative sequencing in COVID-19 strain. Additionally, ten peptides binding to MHC class I and MHC class II were found to be promising candidates for vaccine design with adequate world population coverage of 88.5% and 99.99%, respectively.
Conclusion: T cell epitopes-based peptide vaccine was designed for COVID-19 using envelope protein as an immunogenic target. Nevertheless, the proposed vaccine is rapidly needed to be validated clinically in order to ensure its safety, immunogenic profile and to help on stopping this epidemic before it leads to devastating global outbreaks.

Keywords: Novel coronavirus (COVID-19), Envelope protein, Emerging infection, Epitope-based vaccine, Immunoinformatics approach.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1. Introduction:

Coronaviruses (CoV) are a large family of zoonotic viruses that cause illness ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS-CoV) and severe Acute Respiratory Syndrome (SARS-CoV). In the last decades, six strains of coronaviruses were identified, however in December 2019; a new strain has been spread across Wuhan City, China [1, 2]. It was designated as coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO) [3]. In late January 2020, WHO declared the outbreak a global pandemic with cases in more than 45 countries where the COVID-19 spreading fast outside China, most significantly in South Korea, Italy and Iran with over 2,924 deaths and 85,212 cases confirmed while 39,537 recovered at 29 February 2020, 06:05 AM (GMT).

COVID-19 is a positive-sense single stranded RNA virus (+ssRNA). Its RNA sequence is approximately 30,000 bases in length [4]. It belongs to the subgenus *Sarbecovirus,* Genus *Betacoronavirus* within the family *Coronaviridae.* The corona envelope (E) protein is a small, integral membrane protein involved in several aspects of the virus' life cycle, such as pathogenesis, envelope formation, assembly and budding; alongside with its interactions with both other CoVs proteins (M, N & S) and host cell proteins (release of infectious particles after budding) [5-9].

The infected person characterized with fever, upper or lower respiratory tract symptoms, or diarrhoea, lymphopenia, thrombocytopenia, and increased C-reactive protein and lactate dehydrogenase levels or combination of all these within 3-6 days after exposure. Further molecular diagnosis can be made by Real Time-PCR for genes encoding the internal RNA-dependent RNA polymerase and Spike's receptor binding domain, which can be confirmed by Sanger sequencing and full genome analysis by NGS, multiplex nucleic acid amplification and microarray-based assays [10-14]

A phylogenetic tree of the mutation history of a family of viruses is possible to reconstruct with a sufficient number of sequenced genomes. The phylogenetic analysis indicates that COVID-19 is likely originated from bats [15]. It also showed that is highly related with at most seven mutations relative to a common ancestor [16].

The sequence of COVID-19 RBD, together with its RBM that contacts receptor angiotensin-converting enzyme 2 (ACE2), was found similar to that of SARS coronavirus. On January 2020, a group of scientists demonstrated that ACE2 could act as the receptor for COVID-19 [17-21].

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

However, COVID-19 differs from other previous strains in having several critical residues at 2019-nCoV receptor-binding motif (particularly Gln493) which provide advantageous interactions with human ACE2 [15]. This difference in affinity possibly explains why the novel coronavirus is more contagious than those other viruses.

At present, there is no vaccine or approved treatment for humans, but Chinese traditional medicine, such ShuFengJieDu Capsules and Lianhuaqingwen Capsule, could be possible treatments for COVID-19. However, there are no clinical trials approving the safety and efficacy for these drugs [22].

The main concept within all the immunizations is the ability of the vaccine to initiate an immune response in a faster mode than the pathogen itself. Although traditional vaccines, which depend on biochemical trials, induced potent neutralizing and protective responses in the immunized animals but they can be costly, allergenic, time consuming and require in vitro culture of pathogenic viruses leading to serious concern of safety [23, 24]. Thus the need for safe and efficacious vaccines is highly recommended.

Peptide-based vaccines do not need in vitro culture making them biologically safe, and their selectivity allows accurate activation of immune responses [25, 26]. The core mechanism of the peptide vaccines is built on the chemical method to synthesize the recognized B-cell and T-cell epitopes that are immunodominant and can induce specific immune responses. B-cell epitope of a target molecule can be linked with a T-cell epitope to make it immunogenic. The T-cell epitopes are short peptide fragments (8-20 amino acids), whereas the B-cell epitopes can be proteins [27, 28]. Therefore, in this study, we aimed to design a peptide-based vaccine to predict epitopes from corona envelope (E) protein using immunoinformatics analysis [29-34]. While rapid further studies are recommended to prove the efficiency of the predicted epitopes as a peptide vaccine against this emerging infection.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2. Materials and Methods

Workflow summarizing the procedures for the epitope-based peptide vaccine prediction is shown in **Fig. 1**

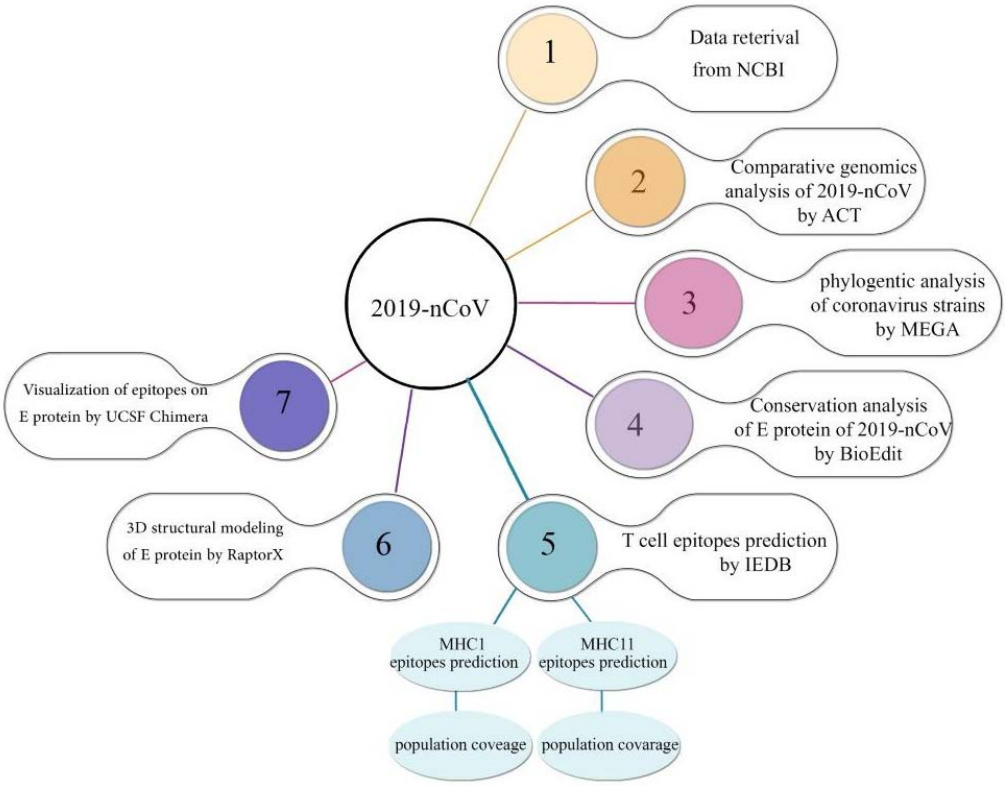

Figure 1: Descriptive workflow for the epitope-based peptide vaccine prediction.

2.1. Data retrieval:

Full genebank files of the complete genomes and annotation of COVID-19(NC_04551) SARS-CoV(FJ211859), MESA-CoV(NC_019843), HCoV-HKU1(AY884001), (HCoV-OC43 (KF923903), HCoV-NL63 (NC_005831) and HCoV-229E (KY983587) were retrieved from the National Center of Biotechnology Information (NCBI); while The FASTA format of envelope (E) protein (YP_009724392.1) , spike (S)protein (YP_009724390.1), nucleocapsid (N) protein (YP_009724397.2) and membrane (M) protein (YP_009724393.1) of 2019-nCoV and the envelope (E) protein of two Chinese and two American sequences (YP009724392.1, QHQ71975.1, QHO60596.1 and QHN73797.1) were obtained from the NCBI. (https://www.ncbi.nlm.nih.gov/)

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2.2. The Artemis Comparison Tool (ACT):

ACT is an in silico analysis software for visualization of comparisons between complete genome sequences and associated annotations [35]. It is also applied to identify regions of similarity, rearrangements and insertions at any level from base-pair differences to the whole genome. (https://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act).

2.3. VaxiJen server:

It is the first server for alignment-independent prediction of protective antigens. It allows antigen classification solely based on the physicochemical properties of proteins without recourse to sequence alignment. It predicts the probability of the antigenicity one or multiple of protein based on auto cross covariance (ACC) transformation of protein sequence. Structural CoV-2019 protein (N,S,E and M) was analyzed by VaxiJen with threshold of 0.4 [36]. (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html)

2.4. BioEdit:

It is a software package proposed to stream a distinct program that can run nearly any sequence operation as well as a few basic alignment investigations. The sequences of E protein were retrieved from UniProt were run in BioEdit to determine the conserved sites through ClustalW in the application settings [37].

2.5. The Molecular Evolutionary Genetics Analysis (MEGA):

MEGA (version 10.1.6) is software for the comparative analysis of molecular sequences. It is used for pairwise and multiple sequences alignment alongside construction and analysis of phylogenetic trees and evolutionary relationships. The gap penalty was 15 for opening and 6.66 for extending the gap for both pairwise and multiple sequences alignment. Bootstrapping of 300 was used in construction of maximum like hood phylogenetic tree [38, 39]. (https://www.megasoftware.net).

2.6. Prediction of T-cell epitopes:

IEDB tools were used to predict the conserved sequences (10-mersequence) from HLA class I and class II T-cell epitopes by using artificial neural network (ANN) approach [40-42]. Artificial Neural Network (ANN) version 2.2 was chosen as Prediction method as it depends on the median inhibitory concentration (IC50) [40, 43-45]. For the binding analysis, all the alleles were carefully chosen, and the length was set at 10 before prediction was done. Analysis of epitopes binding to MHC class I and II molecules was assessed by the IEDB MHC prediction server at (http://tools.iedb.org/mhci/) and (http://tools.iedb.org/mhcii/), respectively. All conserved immunodominant peptides binding to MHC I and II molecules at score equal or less than 100 median inhibitory concentrations (IC50) and 1000, respectively were selected for further analysis while epitopes with IC50 greater than 100 were eliminated [46].

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2.7. Population coverage analysis:

Population coverage for each epitope was carefully determined by the IEDB population coverage calculation tool. Due to the diverse binding sites of epitopes with different HLA allele, the most promising epitope candidates were calculated for population coverage against the whole world, China and Europe population to get and ensure a universal vaccine [47, 48]. (http://tools.iedb.org/population/)

2.8. Tertiary structure (3D) Modeling:

The reference sequence of E protein that has been retrieved from gene bank was used as an input in RaptorX to predict the 3D structure of E protein [49, 50], the visualization of the obtained 3D protein structure was performed in UCSF Chimera (version1.8) [51].

2.9. In silico Molecular Docking:

2.9.1. Ligand Preparation:

In order to estimate the binding affinities between the epitopes and molecular structure of MHC I and MHC II, in silico molecular docking were used. Sequences of proposed epitopes were selected from COVID-19 reference sequence using UCSF Chimera 1.10 and saved as a PDB file. The obtained files were then optimized and energy minimized. The HLA-A*02:01 was selected as the macromolecule for docking. Its crystal structure (4UQ3) was downloaded from the RCSB Protein Data Bank (http://www.rcsb.org/pdb/home/home.do), which was in a complex with an azobenzene-containing peptide [52].

All water molecules and heteroatoms in the retrieved target file 4UQ3 were then removed. Target structure was further optimized and energy minimized using Swiss PDB viewer V.4.1.0 software[53].

2.9.2. Molecular docking

Molecular docking was performed using Autodock 4.0 software, based on Lamarckian Genetic Algorithm; which combines energy evaluation through grids of affinity potential to find the suitable binding position for a ligand on a given protein[54,55]Polar hydrogen atoms were added to the protein targets and Kollman united atomic charges were computed. The target's grid map was calculated and set to 60×60×60 points with grid spacing of 0.375 Å. The grid box was then allocated properly in the target to include the active residue in the center. The genetic algorithm and its run were set to 100.The docking algorithms were set to default. Finally, results were

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

retrieved as binding energies and poses that showed lowest binding energies were visualized using UCSF Chimera.

3. Results:

3.1. The Artemis Comparison Tool:

The reference sequence of envelope protein was aligned with HCov-HKU1 reference protein using artemis comparison tool as illustrated in (Fig. 2).

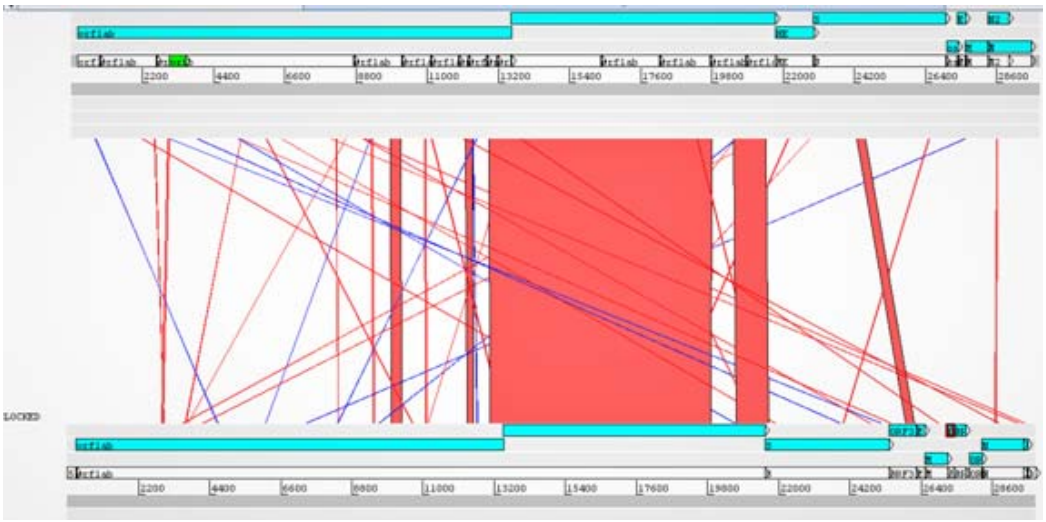

Figure 2: Artemis analysis of envelope protein displaying 3 windows, the upper window represents HCov- HKU1 reference sequence and its genes are highlighted in blue starting from *orflab* gene and ending with *N* gene. The middle window describes the similarities and the difference between the two genomes. Red lines indicate match between genes from the two genomes blue lines indicates inversion which represents same sequences in the two genomes but they are organized in the opposite direction, and the lower windows represents COVID-19and its genes started from *ARFLAB* and ends with *N* genes.

3.2. VaxiJen server:

The mutated proteins were tested for antigenicity using VaxiJen software, where the envelope protein found as the best immunogenic target in Table 1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 1: VaxiJen overall prediction of probable COVID-19 antigen:

Protein	Result	VaxiJen prediction
Protein E	0.6025	Probable antigen
Protein M	0.5102	Probable antigen
Protein S	0.4646	Probable antigen
Protein N	0.5059	Probable antigen

3.3. BioEdit:

Sequence alignment of COVID-19envelope protein was done using BioEdit software which shows total conservation across four sequences which were retrieved from China and USA (Fig. 3).

Figure 3: Sequence alignment of envelope protein of COVID-19 using BioEdit software (total conservation through the 4 strains: 2 form China and 2 from USA)

3.4. The Molecular Evolutionary Genetics Analysis:

To study the evolutionary relationship between all the seven strains of coronavirus, a multiple sequence alignment (MSA) was performed using ClustalW by MEGA software. This alignment was used to construct maximum likelihood phylogenetic tree as seen in **Fig. 4**.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

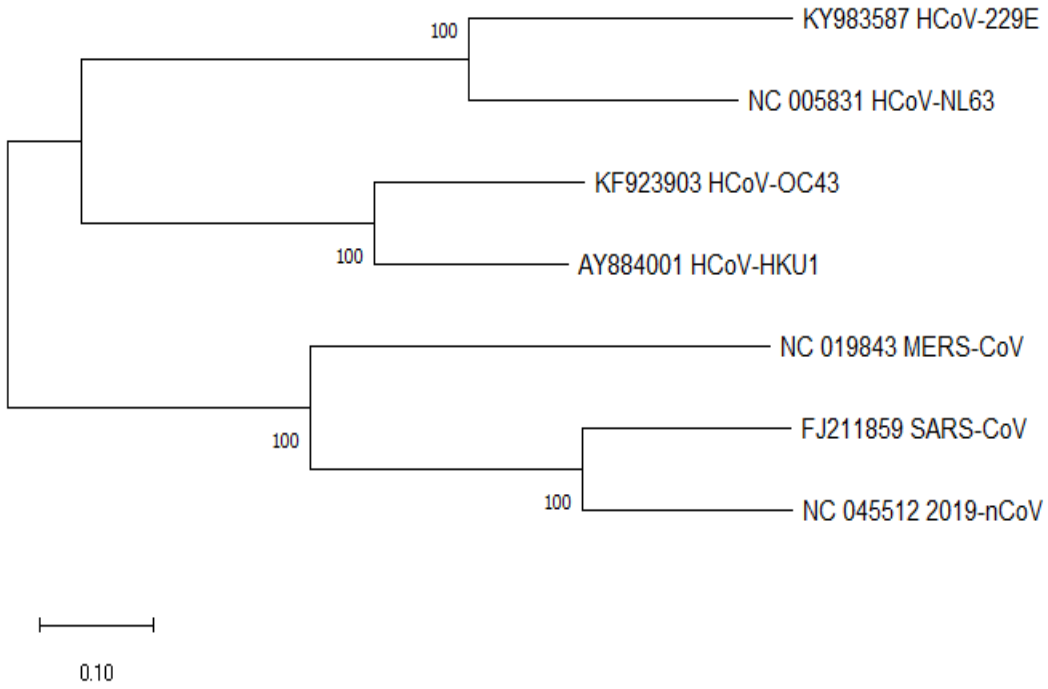

Figure 4: Maximum likelihood phylogenetic tree which describes the evolutionary relationship between the seven strains of coronavirus.

3.5. Prediction of T-cell epitopes and population coverage:

IEDB website was used to analyze 2019-nCoV envelope protein for T cell related peptides. Results show ten MHC class I and II associated peptides with high population coverage (Tables 2 and 3; Fig. 5). The most promising peptides were visualized using UCSF Chimera software (Fig.6a and b).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2: The most promised MHC class I related peptides in envelope protein based vaccine of COVID-19 along with the predicted World, China, Europe and East Asia coverage:

Peptide	Alleles	Coverage	Combined coverage of 10 peptide
YVYSRVKNL	HLA-C*14:02,HLA-C*12:03,HLA-C*07:01,HLA-C*03:03,HLA-C*06:02	50.02%	World: 88.5%
SLVKPSFYV	HLA-A*02:06,HLA-A*02:01,HLA-A*68:02	42.53%	China: 78.17%
SVLLFLAFV	HLA-A*02:06,HLA-A*68:02,HLA-A*02:01	42.53%	Europe: 92.94%
FLAFVVFLL	HLA-A*02:01,HLA-A*02:06	40.60%	East Asia: 80.78%
VLLFLAFVV	HLA-A*02:01	39.08%	
RLCAYCCNI	HLA-A*02:01	39.08%	
FVSEETGTL	HLA-C*03:03,HLA-C*12:03,HLA-A*02:06,HLA-A*68:02,HLA-B*35:01	28.22%	
LTALRLCAY	HLA-A*01:01,HLA-A*30:02,HLA-B*15:01	26.34%	
LVKPSFYVY	HLA-B*15:01,HLA-A*29:02,HLA-A*30:02,HLA-B*35:01	21.72%	
NIVNVSLVK	HLA-A*68:01,HLA-A*11:01	20.88%	

Table 3: The most promised MHC class II related peptides in envelope protein based vaccine of COVID-19 along with the predicted World, China, Europe and East Asia coverage:

Peptide Sequence	Alleles	World coverage	Coverage/10 peptide
KPSFYVYSRVKNLNS	HLA-DPA1*01:03,HLA-DPB1*02:01,HLA-DPB1*03:01,HLA-DPB1*04:01,HLA-DPA1*02:01,HLA-DPB1*05:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DPB1*06:01,HLA-DPB1*14:01,HLA-DPB1*01:01,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQA1*06:01,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQA1*02:01,HLA-DRB1*01:01,HLA-DRB1*07:01,HLA-DRB1*08:01,HLA-DRB1*09:01,HLA-DRB1*11:01,HLA-DRB4*01:03,HLA-DRB1*04:01,HLA-DRB1*10:01,HLA-DRB1*04:05,HLA-DRB1*13:01,HLA-DRB1*08:02,HLA-DRB1*16:02,HLA-DRB1*15:01,HLA-DRB3*03:01,HLA-DRB5*01:01,HLA-DRB3*02:02,HLA-DRB1*04:04,HLA-DRB1*13:02	99.93%	World : 99.99%

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	HLA-DPA1*01:03,HLA-DPB1*02:01,HLA-		
VKPSFYVYSRVKNLN	DPB1*04:01,HLA-DPB1*03:01,HLA-DPA1*02:01,HLA-DPB1*05:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DPB1*06:01,HLA-DPB1*01:01,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQA1*06:01,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQA1*02:01,HLA-DRB1*07:01,HLA-DRB1*08:01,HLA-DRB1*01:01,HLA-DRB1*09:01,HLA-DRB1*11:01,HLA-DRB4*01:03,HLA-DRB1*15:01,HLA-DRB1*13:01,HLA-DRB3*03:01,HLA-DRB1*10:01,HLA-DRB1*16:02,HLA-DRB1*08:02,HLA-DRB1*04:05,HLA-DRB5*01:01,HLA-DRB1*13:02,HLA-DRB3*02:02,HLA-DRB1*04:01,HLA-DRB1*04:04	99.92%	China: 99.96%
LVKPSFYVYSRVKNL	HLA-DPA1*01:03,HLA-DPB1*02:01,HLA-DPB1*04:01,HLA-DPA1*02:01,HLA-DPB1*05:01,HLA-DPB1*06:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DPA1*02:01,HLA-DPB1*01:01,HLA-DQA1*06:01,HLA-DQB1*04:02,HLA-DQA1*05:01,HLA-DQA1*02:01,HLA-DQA1*01:04,HLA-DQB1*05:03,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DRB1*07:01,HLA-DRB1*08:01,HLA-DRB1*09:01,HLA-DRB1*11:01,HLA-DRB4*01:03,HLA-DRB3*03:01,HLA-DRB1*01:01,HLA-DRB1*15:01,HLA-DRB1*16:02,HLA-DRB1*13:01,HLA-DRB1*10:01,HLA-DRB1*08:02,HLA-DRB5*01:01,HLA-DRB1*13:02,HLA-DRB1*04:05,HLA-DRB3*02:02,HLA-DRB1*04:01	99.90%	Europe : 100.0%
PSFYVYSRVKNLNSS	HLA-DPA1*01:03,HLA-DPB1*02:01,HLA-DPB1*03:01,HLA-DPB1*04:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DPA1*02:01,HLA-DPB1*05:01,HLA-DPB1*06:01,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQA1*06:01,HLA-DRB1*01:01,HLA-DRB1*08:01,HLA-	99.86%	East Asia:99.91%

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	DRB1*04:01,HLA-DRB1*11:01,HLA-DRB1*09:01,HLA-DRB1*07:01,HLA-DRB4*01:03,HLA-DRB1*04:05,HLA-DRB1*10:01,HLA-DRB1*13:01,HLA-DRB1*08:02,HLA-DRB1*16:02,HLA-DRB1*15:01,HLA-DRB3*03:01,HLA-DRB3*02:02,HLA-DRB1*04:04,HLA-DRB5*01:01,HLA-DRB1*13:02	
NIVNVSLVKPSFYVY	HLA-DPA1*01:03,HLA-DPB1*02:01,HLA-DPB1*04:01,HLA-DPB1*06:01,HLA-DPA1*02:01,HLA-DPB1*01:01,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQA1*02:01,HLA-DQB1*03:01,HLA-DQB1*03:03,HLA-DQB1*03:03,HLA-DQB1*04:02,HLA-DRB1*12:01,HLA-DRB1*01:01,HLA-DRB5*01:01,HLA-DRB3*03:01,HLA-DRB1*13:01,HLA-DRB1*07:01,HLA-DRB1*15:01,HLA-DRB4*01:03,HLA-DRB1*04:04,HLA-DRB1*08:02,HLA-DRB1*09:01,HLA-DRB1*13:02,HLA-DRB1*11:01,HLA-DRB1*04:05,HLA-DRB1*10:01	99.77%
LLVTLAILTALRLCA	HLA-DPA1*01:03,HLA-DPB1*02:01,HLA-DPB1*06:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQA1*02:01,HLA-DQB1*03:01,HLA-DQB1*03:03,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQA1*06:01,HLA-DQA1*01:03,HLA-DQB1*06:03,HLA-DRB4*01:03,HLA-DRB1*01:01,HLA-DRB1*13:01,HLA-DRB1*04:04,HLA-DRB5*01:01,HLA-DRB3*03:01,HLA-DRB1*10:01,HLA-DRB1*15:01,HLA-DRB1*07:01,HLA-DRB1*11:01,HLA-DRB1*08:01,HLA-DRB1*12:01,HLA-DRB1*03:01,HLA-DRB4*01:01,HLA-DRB1*16:02,HLA-DRB1*08:02	99.72%
SFYVYSRVKNLNSSR	HLA-DPA1*01:03,HLA-DPB1*03:01,HLA-DPB1*02:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DPA1*02:01,HLA-DPB1*05:01,HLA-DPB1*06:01,HLA-	99.72%

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	DQA1*05:01,HLA-DQB1*04:02,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DRB1*04:01,HLA-DRB1*01:01,HLA-DRB1*11:01,HLA-DRB1*08:01,HLA-DRB1*07:01,HLA-DRB1*09:01,HLA-DRB4*01:03,HLA-DRB1*10:01,HLA-DRB1*13:01,HLA-DRB1*04:05,HLA-DRB1*08:02,HLA-DRB1*16:02,HLA-DRB3*02:02,HLA-DRB3*03:01,HLA-DRB1*04:04,HLA-DRB1*15:01,HLA-DRB5*01:01,HLA-DRB1*13:02		
LVTLAILTALRLCAY	HLA-DPA1*01:03,HLA-DPB1*02:01,HLA-DPB1*06:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQA1*02:01,HLA-DQB1*03:01,HLA-DQB1*03:03,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQB1*06:02,HLA-DQA1*02:01,HLA-DQA1*06:01,HLA-DRB4*01:03,HLA-DRB1*01:01,HLA-DRB1*13:01,HLA-DRB1*04:04,HLA-DRB1*12:01,HLA-DRB1*10:01,HLA-DRB5*01:01,HLA-DRB1*15:01,HLA-DRB1*11:01,HLA-DRB3*03:01,HLA-DRB1*03:01,HLA-DRB1*08:01,HLA-DRB1*07:01,HLA-DRB4*01:01,HLA-DRB1*16:02,HLA-DRB1*04:02,HLA-DRB1*08:02	99.69%	
VTLAILTALRLCAYC	HLA-DPA1*01:03,HLA-DPB1*06:01,HLA-DPA1*03:01,HLA-DPB1*04:02,HLA-DQA1*02:01,HLA-DQB1*03:01,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQB1*06:02,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQB1*03:03,HLA-DQA1*06:01,HLA-DQB1*04:02,HLA-DRB1*01:01,HLA-DRB4*01:03,HLA-DRB1*13:01,HLA-DRB1*04:04,HLA-DRB1*12:01,HLA-DRB1*10:01,HLA-DRB5*01:01,HLA-DRB1*15:01,HLA-DRB1*11:01,HLA-DRB1*03:01,HLA-DRB3*03:01,HLA-DRB1*08:01,HLA-DRB1*07:01,HLA-DRB4*01:01,HLA-DRB1*04:02		99.56%

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

CNIVNVSLVKPSFYV	HLA-DPA1*01:03,HLA-DPB1*06:01,HLA-DPB1*04:02,HLA-DQA1*01:02,HLA-DQB1*05:01,HLA-DQA1*05:01,HLA-DQB1*04:02,HLA-DQA1*02:01,HLA-DQB1*03:01,HLA-DQB1*03:03,HLA-DQA1*01:03,HLA-DQB1*06:03,HLA-DRB3*03:01,HLA-DRB1*12:01,HLA-DRB5*01:01,HLA-DRB1*01:01,HLA-DRB1*07:01,HLA-DRB4*01:03,HLA-DRB1*13:01,HLA-DRB1*15:01,HLA-DRB1*08:02,HLA-DRB1*04:04,HLA-DRB1*09:01,HLA-DRB1*13:02,HLA-DRB1*11:01,HLA-DRB1*04:05,HLA-DRB4*01:01,HLA-DRB1*10:01	99.53%

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

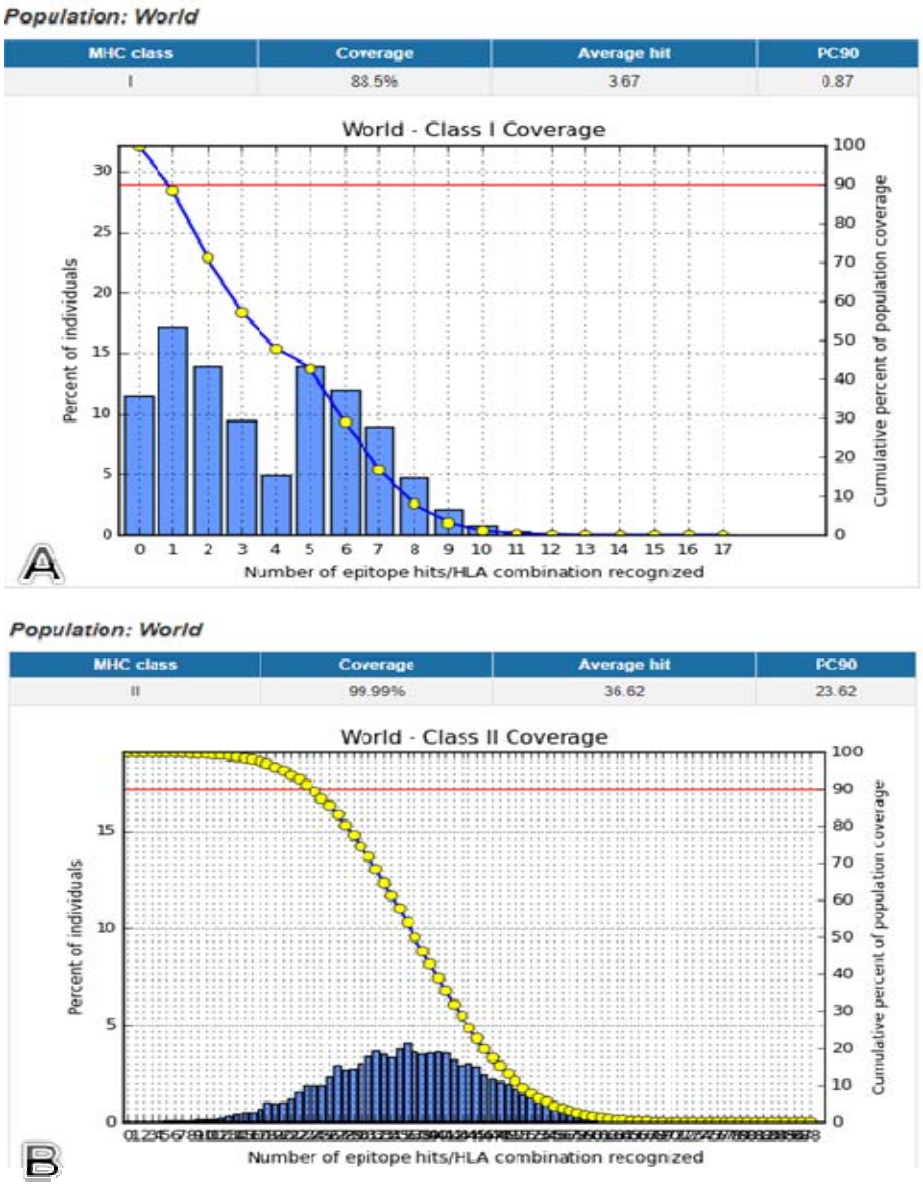

Figure 5: Schematic diagrams (**A**) and (**B**) showing world population coverage of envelope protein of COVID-19 binding to MHC class I and MHC class II molecule respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

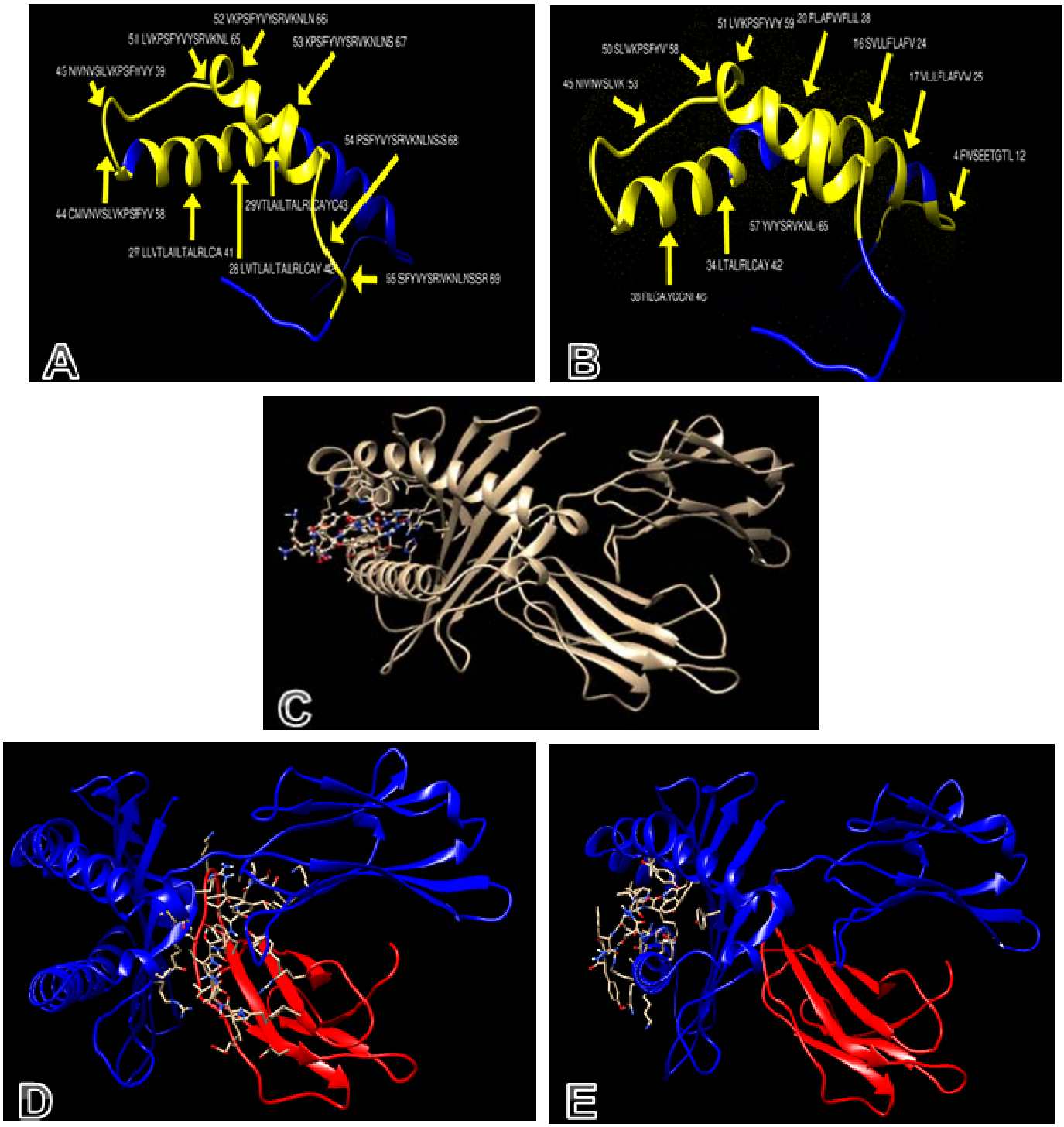

Figure 6: 3D structures visualized by UCSF Chimera: (**A**) and (**B**) showing the most promising peptides in envelope protein of COVID-19 (yellow colored) binding to MHC class I and MHC class II respectively, while (**C**, **D**, and **E**) showing the molecular docking of YVYSRVKNL, LAILTALRL and SLVKPSFYV peptides of coronavirus docked in HLA-A*02:01 respectively

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

4. Discussion:

Designing of a novel vaccine is very crucial to defending the rapid endless of global burden of disease [56-59]. In the last few decades, biotechnology has advanced rapidly; alongside with the understanding of immunology which assisted the rise of new approaches towards rational vaccines design [60]. Peptide-based vaccines are designed to elicit immunity particular pathogens by selectively stimulating antigen specific for B and T cells [61].Applying the advanced bioinformatics tools and databases, various peptide-based vaccines could be designed where the peptides act as ligands [62-64]. This approach has been used frequently in Saint Louis encephalitis virus [65], dengue virus [66], chikungunya virus [67] proposing promising peptides for designing vaccines.

The COVID-19 is an RNA virus which tends to mutate more commonly than the DNA viruses [68]. These mutations lied on the surface of the protein, which make COVID-19 more superior than other previous strains by inducing its sustainability leaving the immune system in blind spot [69].

In our present work, different peptides were proposed for the designing of vaccine against COVID-19 (Fig. 1). In the beginning, the whole genome of COVID-19 was analyzed by comparative genomic approach to determine the potential antigenic target [70]. Artemis Comparative Tool (ACT) was used to analyze human Coronavirus (HCov-HKU1) reference sequence vs. Wuhan-Hu-1 COVID-19. Results obtained (Fig.2) revealed extensive mutation among the tested genomes. New genes; *ORF8* and *ORF6*; were found inserted in COVID-19 which were absent in HCov-HKU1 that might be acquired by the horizontal gene transmission [71]. High rate of mutation between the two genomes were observed in the region from 20000 bp to the end of the sequence. This region encodes the four major structural proteins in coronavirus which are envelope (E) protein, nucleocapsid (N) protein, membrane (M) protein, and spike (S) protein, all of which are required to produce a structurally complete virus [72, 73].

These conserved antigenic sites were revealed in previous studies through sequence alignment between MERS-CoV and Bat-coronavirus [74] and analyzed in SARS-CoV [75].

The four proteins were then analyzed by Vaxigen software to test the probability of antigenic proteins. Protein E was found to be the most antigenic gene with the highest probability as shown in Table 1. Literature survey confirmed this result in which protein E was investigated in severe acute respiratory syndrome (SARS) in 2003 and, more recently, Middle-East respiratory

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

syndrome (MERS) [72]. Furthermore, the conservation of this protein against the seven strains was tested and confirmed through the use of BioEdit package tool (Fig. 3).

As phylogenetic analysis is very powerful tool for determining the evolutionary relationship between strains. Multiple sequence alignment (MSA) was performed using ClustalW for the seven strains of coronavirus, which are COVID-19(NC_04551), SARS-CoV (FJ211859), MESA-CoV (NC_019843), HCoV-HKU1 (AY884001), HCoV-OC43 (KF923903), HCoV-NL63 (NC_005831) and HCoV-229E (KY983587). The maximum likelihood phylogenetic tree revealed that COVID-19 is found in the same clade of SARS-CoV, thus the two strains are highly related to each other (Fig. 4).

The immune response of T cell is considered as a long lasting response compared to B cell, where the antigen can easily escape the antibody memory response [76]. Vaccines that effectively generate cell-mediated response are needed to provide protection against the invading pathogen. Moreover the CD8+ T and CD4+ T cell responses play a major role in antiviral immunity [77]. Thus designing vaccine against T cell is much more important.

Choosing protein E as the antigenic site, the binding affinity to MHC molecules was then evaluated. The protein reference sequence was submitted to IEDB MHC predication tool. 21 peptides were found to bind MHC class I with different affinities (Table 1), from which ten peptides were selected for vaccine design based on the number of alleles and world population percentage (Table 2; Fig. 5).Analysis in IEDB MHC II binding prediction tool resulted in prediction of 61 peptides (Table 2), from which ten peptides were selected for vaccine design based on the number of alleles and world population percentage (Table 3; Fig.5). Unfortunately, IEDB did not give any result for B cell epitopes, this might be due to the length of the COVID-19 (75 amino acids).

It is well known that peptides recognized with high number of HLA molecules are potentially inducing immune response. Based on the aforementioned results and taking into consideration the high binding affinity to both MHC classes I and II, conservancy and population coverage, three peptides are strongly proposed to formulate a new vaccine against COVID-19.

These findings were further confirmed by the results obtained for the molecular docking of the proposed peptides and HLA-A*02:01. The formed complex between MHC molecule and the three peptides (YVYSRVKNL, SLVKPSFYV and LAILTALRL) have shown peptide amino- and carboxyl-termini forming one and three hydrogen bonds, respectively at the two ends of a binding

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

groove with MHC residues with least binding energy -13.2 kcal/mol, -11 kcal/mol and -11.3 kcal/mol, respectively (Fig.6C, D and E).

Although both flu and anti-HIV drugs are used currently in China for treatment of COVID-19, and chloroquine phosphate, an old drug for treatment of malaria, has recently found to have apparent efficacy and acceptable safety against COVID-19 [78,79]; nevertheless more studies are required to standardize these therapies. In addition, there has been some success in the development of mouse models of MERS-CoV and SARS-CoV infection, and candidate vaccines where the envelope (E) protein is mutated or deleted have been described [80-86]. To best of our knowledge, this is the first study to identify certain peptides in envelope (E) protein as candidates for COVID-19. Accordingly, these epitopes were strongly recommended as promising epitopes vaccine candidate against T cell.

5. Conclusion:

Extensive mutations, insertion and deletion were discovered in COVID-19 strain using the comparative sequencing. In addition, a number of MHC class I and II related peptides were found promising candidates. Among which the peptides YVYSRVKNL, SLVKPSFYV and LAILTALRL show high potentiality for vaccine design with adequate world population coverage. T cell epitope-based peptide vaccine was designed for COVID-19 using envelope protein as an immunogenic target; nevertheless, the proposed vaccine rapidly needs to be validated clinically ensuring its safety and immunogenic profile to help on stopping this epidemic before it leads to devastating global outbreaks.

Acknowledgement:

The authors acknowledge the Deanship of Scientific Research at University of Bahri for the supportive cooperation.

Data Availability:

All data underlying the results are available as part of the article, and no additional source data are required.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Authors' contributions:

MIA: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing– original draft. AHA: Formal analysis, Investigation, Methodology. MIM: Methodology, Writing– original draft, Writing – review and editing. NME: Formal analysis, Methodology, Visualization. NSM: Conceptualization, Resources, Writing –review & editing. SWS: Visualization, Validation, Writing – review and editing, AMM: Data curation, Conceptualization, Project administration, Supervision, Writing – review and editing. All authors have read and approved the final manuscript.

Abbreviations:

WHO: World Health Organization; CoV: Coronaviruses; MERS-CoV: Middle East Respiratory Syndrome; SARS-CoV: Acute Respiratory Syndrome; COVID-19: novel coronavirus; HCoV-HKU1: Human coronavirus HKU1; HCoV-OC43: Human coronavirus OC43; HCoV-NL63: Human coronavirus NL63; HCoV-229E: Human coronavirus 229E; ACE2: angiotensin converting enzyme 2; RBM: receptor-binding motif; RBM: receptor-binding domain; vs.: versus; ACT: Artemis Comparison Tool; ACC: auto cross covariance; MEGA: Molecular Evolutionary Genetics Analysis; ANN: Artificial Neural Network; IEDB: Immune Epitope Database; IC50: median inhibitory concentrations; MHC I: Major Histocompatibility complex class I; MHC II: Major Histocompatibility complex class II PDB: Protein database; MSA: multiple sequence alignment.

References:

1. H. Lu, C. W. Stratton, and Y. W. Tang, "Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle," *J Med Virol*, Jan 16 2020.
2. D. S. Hui, I. A. E, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, *et al.*, "The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China," *Int J Infect Dis*, vol. 91, pp. 264-266, Jan 14 2020.
3. D. Benvenuto, M. Giovannetti, A. Ciccozzi, S. Spoto, S. Angeletti, and M. Ciccozzi, "The 2019-new coronavirus epidemic: evidence for virus evolution," *J Med Virol*, Jan 29 2020.
4. *"CoV2020". GISAID EpifluDB. Archivedfrom the original on 12 January 2020. Retrieved February 2020*

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

5. P. Venkatagopalan, S. M. Daskalova, L. A. Lopez, K. A. Dolezal, and B. G. Hogue, "Coronavirus envelope (E) protein remains at the site of assembly," *Virology*, vol. 478, pp. 75-85, Apr 2015.
6. J. L. Nieto-Torres, M. L. Dediego, E. Alvarez, J. M. Jimenez-Guardeno, J. A. Regla-Nava, M. Llorente, *et al.*, "Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein," *Virology*, vol. 415, pp. 69-82, Jul 5 2011.
7. K. M. Curtis, B. Yount, and R. S. Baric, "Heterologous gene expression from transmissible gastroenteritis virus replicon particles," *J Virol*, vol. 76, pp. 1422-34, Feb 2002.
8. J. Ortego, D. Escors, H. Laude, and L. Enjuanes, "Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome," *J Virol*, vol. 76, pp. 11518-29, Nov 2002.
9. T. R. Ruch and C. E. Machamer, "A single polar residue and distinct membrane topologies impact the function of the infectious bronchitis coronavirus E protein," *PLoS Pathog*, vol. 8, p. e1002674, 2012.
10. J. F. Chan, S. Yuan, K. H. Kok, K. K. To, H. Chu, J. Yang, *et al.*, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *Lancet*, Jan 24 2020.
11. N. Zhang, L. Wang, X. Deng, R. Liang, M. Su, C. He, *et al.*, "Recent advances in the detection of respiratory virus infection in humans," *J Med Virol*, Jan 15 2020.
12. J. B. Mahony, A. Petrich, and M. Smieja, "Molecular diagnosis of respiratory virus infections," *Crit Rev Clin Lab Sci*, vol. 48, pp. 217-49, Sep-Dec 2011.
13. V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, *et al.*, "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR," *Euro Surveill*, vol. 25, Jan 2020.
14. Y. Chen, Q. Liu, and D. Guo, "Emerging coronaviruses: genome structure, replication, and pathogenesis," *J Med Virol*, Jan 22 2020
15. Y. Wan, J. Shang, R. Graham, R. S. Baric, and F. Li, "Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS, J Virol, Jan 29 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

16. "Genomic epidemiology of novel coronavirus (nCoV)". nextstrain.org. Retrieved 11 February 2020.
17. Letko M, Munster V (January 2020). "Functional assessment of cell entry and receptor usage for lineage B β-coronaviruses, including 2019-nCoV". bioRxiv 2020.01.22.915660 Check |biorxiv= value (help).
18. Letko, M, Marzi, et al. (February 2020). "Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses". Nature Microbiology. doi:10.1038/s41564-020-0688-y. PMID 32094589.
19. El Sahly HM. "Genomic Characterization of the 2019 Novel Coronavirus". New England Journal of Medicine. Retrieved 9 February 2020.
20. Gralinski LE, Menachery VD (January 2020). "Return of the Coronavirus: 2019-nCoV". Viruses. **12** (2): 135. doi:10.3390/v12020135. PMID 31991541.
21. Lu R, Zhao X, Li J, et al. (January 2020). "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding". Lancet. **395**(10224):565–574. doi:10.1016/S0140-6736(20)30251 8. PMID 32007145.
22. H. Lu, "Drug treatment options for the 2019-new coronavirus (2019-nCoV)," *Biosci Trends*, Jan 28 2020.
23. Y. T. Lo, T. W. Pai, W. K. Wu, and H. T. Chang, "Prediction of conformational epitopes with the use of a knowledge-based energy function and geometrically related neighboring residue characteristics," *BMC Bioinformatics*, vol. 14 Suppl 4, p. S3, 2013.
24. W. Li, M. D. Joshi, S. Singhania, K. H. Ramsey, and A. K. Murthy, "Peptide Vaccine: Progress and Challenges," *Vaccines (Basel)*, vol. 2, pp. 515-36, Jul 2 2014.
25. A. W. Purcell, J. McCluskey, and J. Rossjohn, "More than one reason to rethink the use of peptides in vaccine design," *Nat Rev Drug Discov*, vol. 6, pp. 404-14, May 2007.
26. N. L. Dudek, P. Perlmutter, M. I. Aguilar, N. P. Croft, and A. W. Purcell, "Epitope discovery and their use in peptide based vaccines," *Curr Pharm Des*, vol. 16, pp. 3149-57, 2010.
27. S. Dermime, D. E. Gilham, D. M. Shaw, E. J. Davidson, K. Meziane el, A. Armstrong, *et al.*, "Vaccine and antibody-directed T cell tumour immunotherapy," *Biochim Biophys Acta*, vol. 1704, pp. 11-35, Jul 6 2004.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

28. R. H. Meloen, J. P. Langeveld, W. M. Schaaper, and J. W. Slootstra, "Synthetic peptide vaccines: unexpected fulfillment of discarded hope?," *Biologicals,* vol. 29, pp. 233-6, Sep-Dec 2001.
29. V. Brusic and N. Petrovsky, "Immunoinformatics and its relevance to understanding human immune disease," *Expert Rev Clin Immunol,* vol. 1, pp. 145-57, May 2005.
30. N. Tomar and R. K. De, "Immunoinformatics: a brief review," *Methods Mol Biol,* vol. 1184, pp. 23-55, 2014.
31. S. Khalili, A. Jahangiri, H. Borna, K. Ahmadi Zanoos, and J. Amani, "Computational vaccinology and epitope vaccine design by immunoinformatics," *Acta Microbiol Immunol Hung,* vol. 61, pp. 285-307, Sep 2014.
32. N. Tomar and R. K. De, "Immunoinformatics: an integrated scenario," *Immunology,* vol. 131, pp. 153-68, Oct 2010.
33. N. R. Hegde, S. Gauthami, H. M. Sampath Kumar, and J. Bayry, "The use of databases, data mining and immunoinformatics in vaccinology: where are we?," *Expert Opin Drug Discov,* vol. 13, pp. 117-130, Feb 2018.
34. A. A. Bahrami, Z. Payandeh, S. Khalili, A. Zakeri, and M. Bandehpour, "Immunoinformatics: In Silico Approaches and Computational Design of a Multi-epitope, Immunogenic Protein," *Int Rev Immunol,* vol. 38, pp. 307-322, 2019.
35. T. J. Carver, K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill, "ACT: the Artemis Comparison Tool," *Bioinformatics,* vol. 21, pp. 3422-3, Aug 15 2005.
36. I. A. Doytchinova and D. R. Flower, "VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines," *BMC Bioinformatics,* vol. 8, p. 4, Jan 5 2007.
37. O. O. Oni, A. A. Owoade, and C. A. O. Adeyefa, "Design and evaluation of primer pairs for efficient detection of avian rotavirus," *Trop Anim Health Prod,* vol. 50, pp. 267-273, Feb 2018.
38. G. Stecher, K. Tamura, and S. Kumar, "Molecular Evolutionary Genetics Analysis (MEGA) for macOS," *Mol Biol Evol,* Jan 6 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

39. K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0," *Mol Biol Evol*, vol. 24, pp. 1596-9, Aug 2007.
40. M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemoller, K. Lamberth, S. Buus, *et al.*, "Reliable prediction of T-cell epitopes using neural networks with novel sequence representations," *Protein Sci*, vol. 12, pp. 1007-17, May 2003.
41. C. Lundegaard, K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen, "NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11," *Nucleic Acids Res*, vol. 36, pp. W509-12, Jul 1 2008.
42. M. Andreatta and M. Nielsen, "Gapped sequence alignment using artificial neural networks: application to the MHC class I system," *Bioinformatics*, vol. 32, pp. 511-7, Feb 15 2016.
43. S. Buus, S. L. Lauemoller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, *et al.*, "Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach," *Tissue Antigens*, vol. 62, pp. 378-84, Nov 2003.
44. C. Lundegaard, O. Lund, and M. Nielsen, "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*, vol. 24, pp. 1397-8, Jun 1 2008.
45. C. Lundegaard, M. Nielsen, and O. Lund, "The validity of predicted T-cell epitopes," *Trends Biotechnol*, vol. 24, pp. 537-8, Dec 2006.
46. S. Southwood, J. Sidney, A. Kondo, M. F. del Guercio, E. Appella, S. Hoffman, *et al.*, "Several common HLA-DR types share largely overlapping peptide binding repertoires," *J Immunol*, vol. 160, pp. 3363-73, Apr 1 1998.
47. A. Patronov and I. Doytchinova, "T-cell epitope vaccine design by immunoinformatics," *Open Biol*, vol. 3, p. 120139, Jan 8 2013.
48. H. H. Bui, J. Sidney, K. Dinh, S. Southwood, M. J. Newman, and A. Sette, "Predicting population coverage of T-cell epitope-based diagnostics and vaccines," *BMC Bioinformatics*, vol. 7, p. 153, Mar 17 2006.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

49. D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, *et al.*, "GenBank," *Nucleic Acids Res,* vol. 45, pp. D37-d42, Jan 4 2017.
50. S. Wang, W. Li, S. Liu, and J. Xu, "RaptorX-Property: a web server for protein structure property prediction," *Nucleic Acids Res,* vol. 44, pp. W430-5, Jul 8 2016.
51. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, *et al.*, "UCSF Chimera--a visualization system for exploratory research and analysis," *J Comput Chem,* vol. 25, pp. 1605-12, Oct 2004.
52. Choo JA, Thong SY, Yap J, van Esch WJ, Raida M, Meijers R, et al. Bioorthogonal cleavage and exchange of major histocompatibility complex ligands by employing azobenzene-containing peptides. Angewandte Chemie (International ed in English). 2014;53(49):13390-4.
53. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997;18(15):2714-23.
54. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry. 2009;30(16):2785-91.
55. Morris G, S. Goodsell D, S. Halliday R, Huey R, Hart W, Belew R, et al. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function1998. 1639-62 p
56. S. J. Marshall, "Developing countries face double burden of disease," *Bull World Health Organ,* vol. 82, p. 556, Jul 2004.
57. A. S. De Groot and R. Rappuoli, "Genome-derived vaccines," *Expert Rev Vaccines,* vol. 3, pp. 59-76, Feb 2004.
58. B. Korber, M. LaBute, and K. Yusim, "Immunoinformatics comes of age," *PLoS Comput Biol,* vol. 2, p. e71, Jun 30 2006.
59. A. S. Fauci, "Emerging and re-emerging infectious diseases: influenza as a prototype of the host-pathogen balancing act," *Cell,* vol. 124, pp. 665-70, Feb 24 2006.
60. S. Chaudhury, E. H. Duncan, T. Atre, S. Dutta, M. D. Spring, W. W. Leitner, *et al.*, "Combining immunoprofiling with machine learning to assess the effects of adjuvant formulation on human vaccine-induced immunity," *Hum Vaccin Immunother,* pp. 1-12, Oct 7 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

61. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* 2007; 6 (5): 404-14. PMID: 17473845.
62. S. S. Usmani, R. Kumar, S. Bhalla, V. Kumar, and G. P. S. Raghava, "In Silico Tools and Databases for Designing Peptide-Based Vaccine and Drugs," *Adv Protein Chem Struct Biol*, vol. 112, pp. 221-263, 2018.
63. S. K. Dhanda, S. S. Usmani, P. Agrawal, G. Nagpal, A. Gautam, and G. P. S. Raghava, "Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics," *Brief Bioinform*, vol. 18, pp. 467-478, May 1 2017.
64. Y. Khazaei-Poul, S. Farhadi, S. Ghani, S. A. Ahmadizad, and J. Ranjbari, "Monocyclic Peptides: Types, Synthesis and applications," *Curr Pharm Biotechnol*, Jan 20 2020.
65. M. A. Hasan, M. Hossain, and M. J. Alam, "A computational assay to design an epitope-based Peptide vaccine against saint louis encephalitis virus," *Bioinform Biol Insights*, vol. 7, pp. 347-55, 2013.
66. S. Chakraborty, R. Chakravorty, M. Ahmed, A. Rahman, T. M. Waise, F. Hassan, *et al.*, "A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions," *In Silico Biol*, vol. 10, pp. 235-46, 2010.
67. M. Tahir Ul Qamar, A. Bari, M. M. Adeel, A. Maryam, U. A. Ashfaq, X. Du, *et al.*, "Peptide vaccine against chikungunya virus: immuno-informatics combined with molecular docking approach," *J Transl Med*, vol. 16, p. 298, Oct 27 2018.
68. S. S. Twiddy, E. C. Holmes, and A. Rambaut, "Inferring the rate and time-scale of dengue virus evolution," *Mol Biol Evol*, vol. 20, pp. 122-9, Jan 2003.
69. J. M. Christie, H. Chapel, R. W. Chapman, and W. M. Rosenberg, "Immune selection and genetic sequence variation in core and envelope regions of hepatitis C virus," *Hepatology*, vol. 30, pp. 1037-44, Oct 1999.
70. R. C. Hardison, "Comparative genomics," *PLoS Biol*, vol. 1, p. E58, Nov 2003.
71. C. Gilbert, A. Chateigner, L. Ernenwein, V. Barbe, A. Bezier, E. A. Herniou, *et al.*, "Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons," *Nat Commun*, vol. 5, p. 3348, 2014.
72. D. Schoeman and B. C. Fielding, "Coronavirus envelope protein: current knowledge," *Virol J*, vol. 16, p. 69, May 27 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

73. E. Mortola and P. Roy, "Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system," *FEBS Lett*, vol. 576, pp. 174-8, Oct 8 2004.
74. R. Sharmin and A. B. Islam, "Conserved antigenic sites between MERS-CoV and Bat-coronavirus are revealed through sequence analysis," *Source Code Biol Med*, vol. 11, p. 3, 2016.
75. J. Xu, J. Hu, J. Wang, Y. Han, Y. Hu, J. Wen, *et al.*, "Genome organization of the SARS-CoV," *Genomics Proteomics Bioinformatics*, vol. 1, pp. 226-35, Aug 2003.
76. Black M, Trent A, Tirrell M, Olive C. Advances in the design and delivery of peptide subunit vaccines with a focus on toll-like receptor agonists. *Expert Rev Vaccines* 2010; 9: 157-173.
77. Sesardic D. Synthetic peptide vaccines. *J Med Microbiol* 1993; 39: 241-242.
78. Breakthrough: Chloroquine phosphate has shown apparentefficacy in treatment of COVID-19 associated pneumonia inclinical studies.
79. Potential Inhibitors Against Papain-like Protease of Novel Coronavirus (COVID-19) from FDA Approved Drugs.
80. J. W. Westerbeck and C. E. Machamer, "A Coronavirus E Protein Is Present in Two Distinct Pools with Different Effects on Assembly and the Secretory Pathway," *J Virol*, vol. 89, pp. 9313-23, Sep 2015.
81. M. L. Dediego, L. Pewe, E. Alvarez, M. T. Rejas, S. Perlman, and L. Enjuanes, "Pathogenicity of severe acute respiratory coronavirus deletion mutants in hACE-2 transgenic mice," *Virology*, vol. 376, pp. 379-89, Jul 5 2008.
82. J. Netland, M. L. DeDiego, J. Zhao, C. Fett, E. Alvarez, J. L. Nieto-Torres, *et al.*, "Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease," *Virology*, vol. 399, pp. 120-128, Mar 30 2010.
83. P. B. McCray, Jr., L. Pewe, C. Wohlford-Lenane, M. Hickey, L. Manzel, L. Shi, *et al.*, "Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus," *J Virol*, vol. 81, pp. 813-21, Jan 2007.
84. J. A. Regla-Nava, J. L. Nieto-Torres, J. M. Jimenez-Guardeno, R. Fernandez-Delgado, C. Fett, C. Castano-Rodriguez, *et al.*, "Severe acute respiratory syndrome coronaviruses

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.04.934232; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

with mutations in the E protein are attenuated and promising vaccine candidates," *J Virol*, vol. 89, pp. 3870-87, Apr 2015.

85. J. Zhao, K. Li, C. Wohlford-Lenane, S. S. Agnihothram, C. Fett, J. Zhao, *et al.*, "Rapid generation of a mouse model for Middle East respiratory syndrome," *Proc Natl Acad Sci U S A*, vol. 111, pp. 4970-5, Apr 1 2014.
86. X. Guo, Y. Deng, H. Chen, J. Lan, W. Wang, X. Zou, *et al.*, "Systemic and mucosal immunity in mice elicited by a single immunization with human adenovirus type 5 or 41 vector-based vaccines carrying the spike protein of Middle East respiratory syndrome coronavirus," *Immunology*, vol. 145, pp. 476-84, Aug 2015.