Heliyon 10 (2024) e24972

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Discovering conserved epitopes of Monkeypox: Novel immunoinformatic and machine learning approaches

Mohammad Izadi [a,*], Fatemeh Mirzaei [a], Mohammad Aref Bagherzadeh [b], Shamim Ghiabi [c], Alireza Khalifeh [d]

[a] Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran
[b] Student Research Committee, Jahrom University of Medical Sciences, Jahrom, Iran
[c] Department of Medical Chemistry, Faculty of Pharmacy, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran
[d] Department of Pathology, Faculty of Dentistry, Shiraz Branch, Islamic Azad of University, Shiraz, Iran

ARTICLE INFO

Keywords:
Monkeypox
Conserved epitopes
Machine learning
Immunoinformatic

ABSTRACT

The Monkeypox virus, an Orthopoxvirus with zoonotic origins, has been responsible for a growing number of human infections reminiscent of smallpox since May 2022, as reported by the World Health Organization. As of now, there are no established medical treatments for managing Monkeypox infections. In this study, we used machine learning to select conserved epitopes. Proteins were determined using Reverse Vaccinology and Gene Ontology subcellular localization, and their epitopes were predicted. NextClade was used to calculate the number of mutations in each amino acid position using 2433 Monkeypox sequences. The Unsupervised Nearest Neighbor machine learning algorithm and ideal matrix [0 0] were used to calculate the conservancy score of epitopes. Six proteins were determined for epitope prediction. Finally, 47 MHC-I epitopes, 5 MHC-II epitopes, and 10 Linear B cell epitopes were discovered. Our method can select epitopes for vaccine design to prevent viruses with accelerated evolution and high mutation rate.

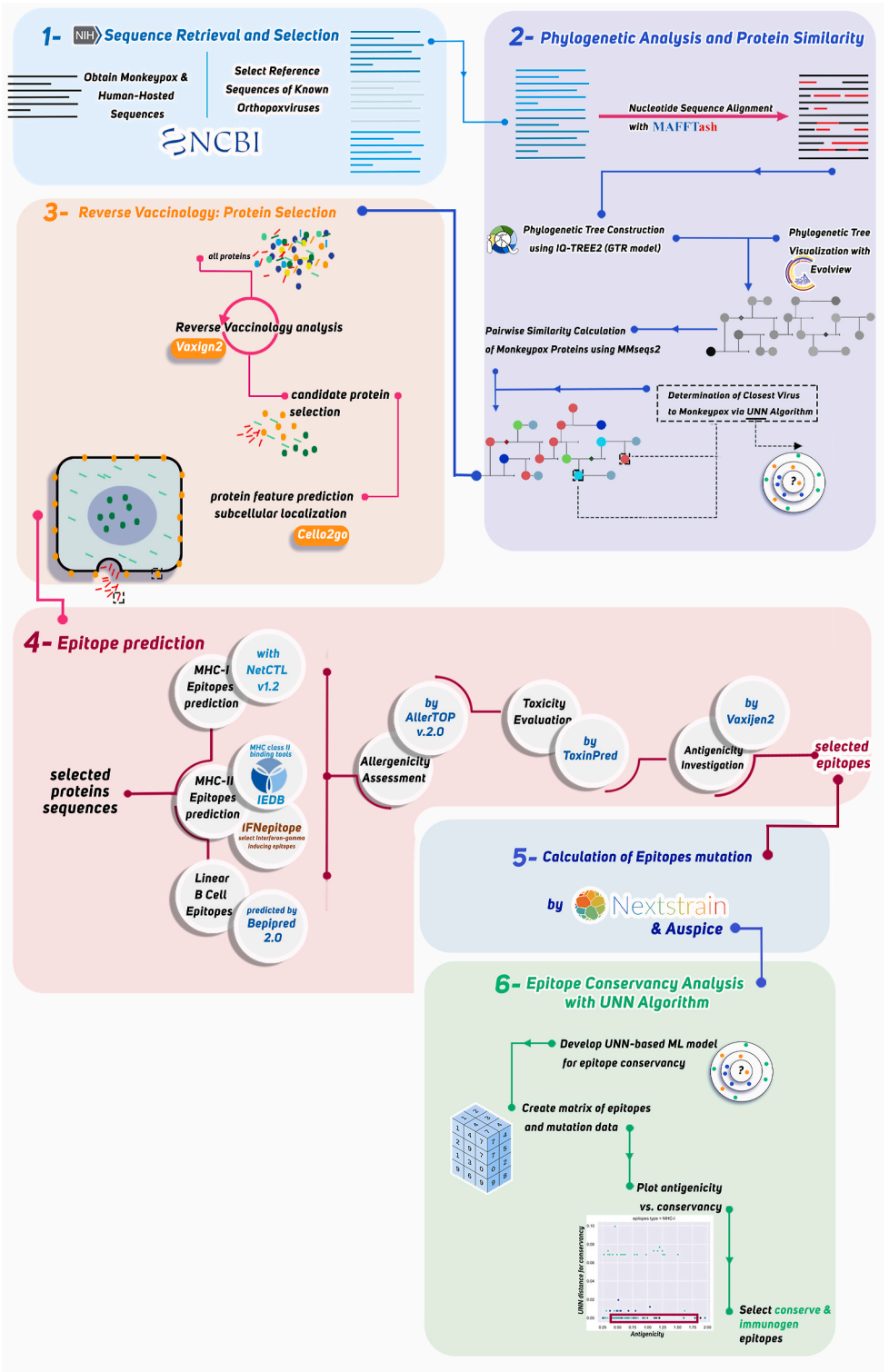
1. Introduction

Monkeypox is a DNA virus belonging to the family Poxviridae, subfamily Chordopoxvirinae, and genus Orthopoxvirus. It leads to a disease known as Mpox, which exhibits similar features to Smallpox [1]. Monkeypox virus is more prevalent in the central and western regions of Africa compared to other areas. It has two clades, namely the Central African (Clade I) and West African (Clade II) variants, with the Clade I variant being more severe and having a higher mortality rate [2]. In recent years, there have been reports of confirmed and suspected Mpox cases in several European and American countries and non-endemic regions. Notably, many of these patients had no history of traveling to endemic areas, highlighting the importance of studying this zoonotic disease [3].

Vaccination plays a crucial role in preventing future Mpox outbreaks. Vaccination against smallpox, which can induce relative immunity against the Monkeypox virus through cross-reaction, has been a key preventive measure [4]. However, due to the eradication of smallpox and the discontinuation of general vaccination, immunity against the Monkeypox virus has declined, possibly contributing to the recent increase in infections [5]. Additionally, no specific Mpox vaccine has been developed to date. The currently approved vaccines for smallpox and Mpox are based on attenuated viruses or vaccinia replicators. However, these types of vaccines can sometimes lead to severe complications such as myocarditis, underscoring the need for the development of new-generation vaccines

* Corresponding author. School of Medicine, Shiraz University of Medical Sciences, Karim Khan Zand Blvd, Shiraz, Iran.
E-mail address: mohammadizadi@sums.ac.ir (M. Izadi).

https://doi.org/10.1016/j.heliyon.2024.e24972
Received 23 November 2022; Received in revised form 12 December 2023; Accepted 17 January 2024
Available online 23 January 2024



(caption on next page)

Fig. 1. illustrates the step-by-step workflow employed in our study. We began by collecting a dataset consisting of 2442 Monkeypox virus sequences and reference sequences of Monkeypox virus, along with 9 other Orthopoxviruses from the National Center for Biotechnology Information database (Step 1). Next, we aligned these 10 Orthopoxvirus sequences using MAFFT and constructed a phylogenetic tree using IQ-Tree2. This allowed us to determine the pairwise similarity between the target proteins of these viruses and identify the closest Orthopoxviruses to Monkeypox virus (Step 2). In Step 3, we conducted reverse vaccinology analysis on the reference sequence proteins of Monkeypox virus using Vaxign2 and Cello2go. This analysis helped us identify specific target proteins for epitope prediction. Moving on to Step 4, we predicted the MHC-I, MHC-II, and Linear B cell epitopes of the selected target proteins using various tools (details provided in Supplementary Table 1). Step5, the amino acid mutations of selected proteins calculated by NextClade (provided by NextStrain). Number of mutation and site of mutation calculated for each predicted epitope. In Step 6, we constructed a matrix to represent the predicted epitopes. This matrix indicated the number of mutations and number of site of mutations that were affected for each epitope. To assess the conservancy of the predicted epitopes, we employed the Unsupervised Nearest Neighbor algorithm to measure the distance of each epitope to the ideal matrix [0 0]. This distance value served as a conservancy score. Finally, we used a Scatter plot with antigenicity on the x-axis and normalized Unsupervised Nearest Neighbor distance on the y-axis to aid in the selection of antigenic and conserved epitopes.

with fewer side effects against Mpox [6].

Conserved epitope prediction for new-platform vaccine design can be facilitated by employing machine learning, bioinformatics and immunoinformatics approaches. To address the critical need of predicting conserved epitopes for the development of effective vaccines, one study employed bioinformatics techniques to systematically analyze the Mpox 2022 outbreak. This analysis resulted in the identification of novel mutations and highlighted ten proteins that are prone to mutation [3]. Utilizing immunoinformatics, some studies devised a multi-epitope vaccine against Mpox by accurately predicting stable interactions with human receptors TLR5 and TLR4. They ensured antigenicity by selecting immunodominant B-cell and T-cell epitopes [7–9]. Furthermore, these studies assessed the conservation of the predicted epitopes by aligning them and determining the percentage of pairwise similarity [8,9]. However, this approach usually employed conservation of predicted epitopes, and machine learning can potentially assist in this task. Machine learning enables the understanding of the virus's structure, selection of target proteins for epitope prediction, monitoring of virus mutations over time, and identification of conserved epitopes. These methods are essential in the design of immunogenic vaccines against Monkeypox [10–12].

In this study, we employed immunoinformatics and machine learning approaches to predict and select conserved epitopes, considering the accelerated evolution of Monkeypox virus [13]. Our aim was to identify reliable epitopes that are likely to remain protected against future strains, probably serving as ideal targets for designing immunogenic vaccines.

2. Materials and methods

2.1. Data source

We obtained the reference sequence of Monkeypox virus (NC_063383.1) and 2442 human-hosted Monkeypox virus sequences from 1996 to October 2022 from the National Center for Biotechnology Information (NCBI) Virus database. The complete list of these sequences is available in Supplementary Information: sup file 1. We also selected one reference sequences for each known Orthopoxviruses, including Camelpox (NC_003391.1), Cowpox (NC_003663.2), Ectromelia (NC_004105.1), Raccoonpox (NC_027213.1), Skunkpox (NC_031038.1), Taterapox (NC_008291.1), Vaccinia (NC_006998.1), Variola (NC_001611.1), and Volepox (NC_031033.1), from NCBI (Supplementary Table 1). The workflow of our study is presented in Fig. 1 and Supplementary Fig. 1.

2.2. Phylogenetic analysis of orthopoxviruses and pairwise similarity

We aligned the 10 nucleotide sequences of the selected Orthopoxviruses using MAFFT (https://mafft.cbrc.jp/alignment/software/). Subsequently, we constructed a phylogenetic tree using IQ-TREE2 (http://www.iqtree.org/) with maximum likelihood to the GTR model. The resulting phylogenetic tree was visualized using Evolview v3 (Evolview:Tree View (evolgenius.info)). Additionally, we calculated the pairwise similarity of Monkeypox target proteins with other Orthopoxviruses using MMseqs2 easy-search (https://github.com/soedinglab/MMseqs2). We determined the closest virus to Monkeypox virus based on protein similarities and employing the Unsupervised Nearest Neighbor (UNN) algorithm. By percent of pairwise similarity that calculated by MMseqs2, we calculate distance of each virus to Monkeypox virus by UNN algorithm. This algorithm was implemented in Python using the NearestNeighbors package from scikit-learn. Subsequently, distances were normalized by dividing all values by the maximum distance. The mean percentage of identity for each protein across all known Orthopoxviruses was calculated and employed to identify proteins with the highest similarity among all known Orthopoxviruses. These proteins are considered potential targets for a cross-protective vaccine.

2.3. Reverse vaccinology analysis

We utilized the Vaxign2 web-based tool for Reverse Vaccinology (RV) analysis [12]. This tool helps in selecting protein targets for vaccine design. Furthermore, we employed the Cello2go web-based tool, which predicts various target protein features based on Gene Ontology Annotation, including subcellular localization, molecular function, biological process, and cellular component [14]. We selected candidate proteins for epitope prediction based on Vaxign2 with probability $\geq$0.51 and Vaxign-ML score $\geq$90.0. Additionally, we used Cello2go with Eukaryote virus databases and an E-value of 0.001 for subcellular localization. Membrane and extracellular

proteins were chosen due to their availability to immune system cells. Furthermore, we considered relevant literature to select other target proteins.

2.4. Epitope prediction

For MHC-I epitope prediction, we utilized the NetCTL v1.2 server (https://services.healthtech.dtu.dk/services/NetCTL-1.2/1-Submission.php). We predicted 9-mer MHC-I epitopes for supertypes A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62 with a score ≥0.75. We checked the immunogenicity of the epitopes using the IEDB Class I Immunogenicity tool (http://tools.iedb.org/immunogenicity/), and selected epitopes with scores ≥0.

For MHC-II epitope prediction, we used the Peptide binding to MHC class II molecules tool (http://tools.iedb.org/mhcii/) with the model labaled as "IEDB recommended 2.22" for the full HLA reference set. We predicted MHC-II epitopes with a length of 15 amino acids, selecting peptides with a percentile rank ≤1.

We employed the Bepipred2.0 tool (http://tools.iedb.org/bcell/) available in IEDB for predicting linear B cell epitopes, considering a default threshold and selecting only epitopes with a length ≥6.

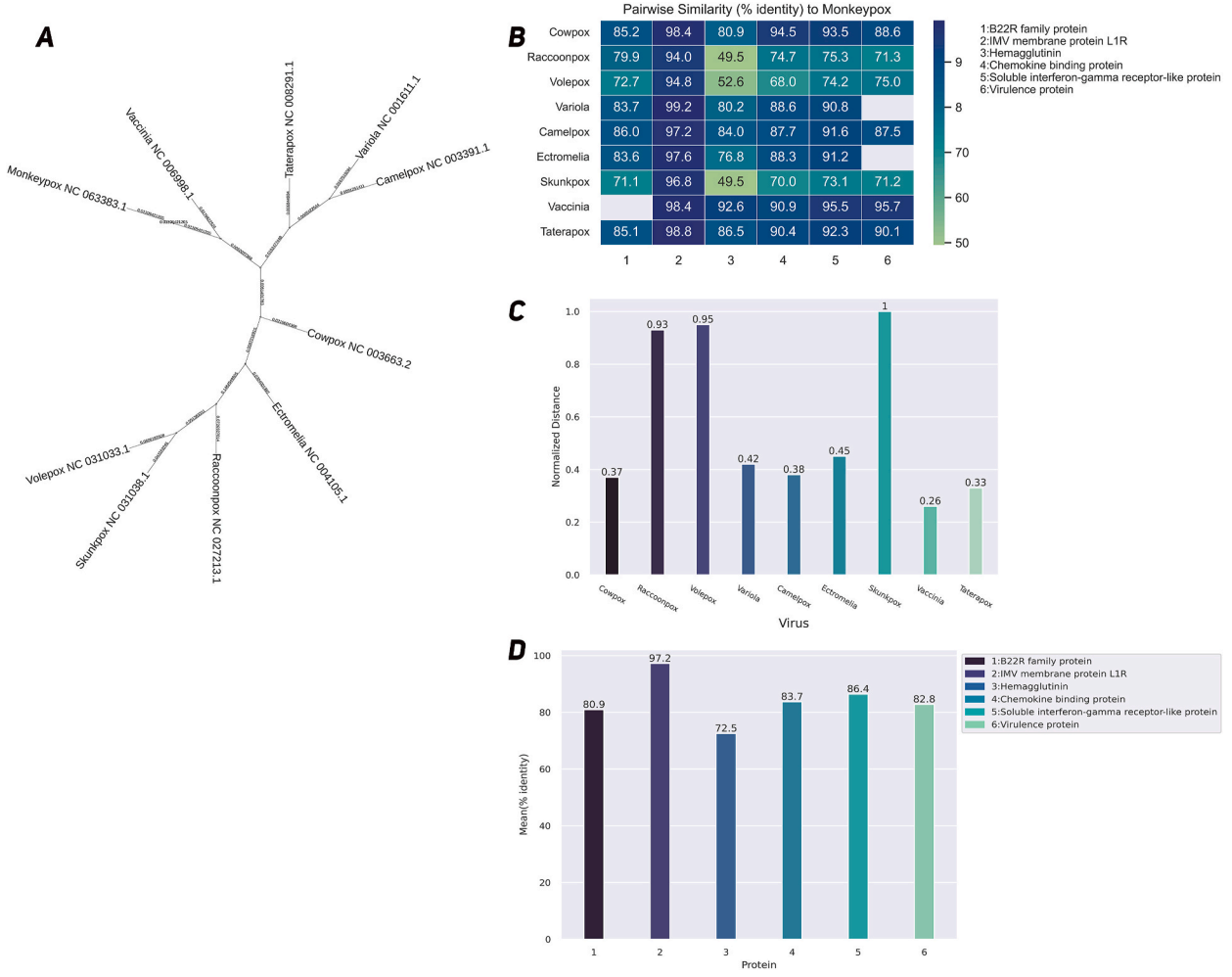


Fig. 2. Presents a comprehensive phylogenetic analysis of Orthopoxviruses. (A), the phylogeny tree is depicted, showcasing the relationship between 10 National Center for Biotechnology Information reference sequences of known Orthopoxvirus members. The tree was constructed using IQ-Tree2 and visualized using Evolview v3, with the root set to Monkeypox virus. Interestingly, the analysis reveals Vaccinia to be the closest relative to Monkeypox virus. (B), a heatmap illustrates the percentage of pairwise similarity of 6 target proteins among the 9 other known Orthopoxvirus members in comparison to Monkeypox virus.(C), the plot displays the Normalized Unsupervised Nearest Neighbor distance of each Orthopoxvirus to Monkeypox virus based on the 6 target proteins. To ensure accuracy, this normalization involves dividing each distance by the maximum distance. As confirmed by the phylogeny tree, the analysis highlights Vaccinia as the Orthopoxvirus with the lowest normalized distance to Monkeypox virus, indicating a closer relationship between the two. (D) presents a plot showcasing the mean percentage of identity for 9 others known Orthopoxviruses compared to Monkeypox virus proteins. Interestingly, it shows that the IMV membrane Protein L1R exhibits the highest mean identity among Orthopoxviruses. This finding suggests its potential utility as a candidate for developing cross-reactive vaccines.

To assess the allergenicity of the epitopes, we used AllerTOP v.2.0 (https://www.ddg-pharmfac.net/AllerTOP/), selecting non-allergenic epitopes. We evaluated the toxicity of the epitopes using ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/algo.php) with a threshold set at 0. Furthermore, we investigated the antigenicity of the epitopes using Vaxijen2 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with a threshold of 0.25. For MHC-II epitopes, we used the IFNepitope server (https://webs.iiitd.edu.in/raghava/ifnepitope/predict.php) to select epitopes that induce Interferon-gamma. The features of the epitope prediction tool, along with their corresponding thresholds, are listed in Supplementary Table 2.

2.5. Amino acid mutations

We utilized NextClade CLI, a powerful tool for analyzing viral genetic sequences, to identify nucleotide and amino acid mutations, perform mutation calling, clade assignment, sequence quality check, and genome alignment. All sequences extracted from NCBI were analyzed using the Human Monkeypox (hMPXV) dataset of NextClade CLI. We performed a quality check, including assessing missing nucleotides, mixed sites, private mutations, mutation clusters, stop codons, and frameshifts [15]. Sequences that did not pass the quality check were excluded from the analysis. We visualized the results obtained from NextClade CLI, including the phylogeny tree, nucleotide mutations, and amino acid mutations using Auspice (auspice.us). Based on the amino acid mutations obtained, we calculated the number of mutations and the number of mutation sites for each epitope.

2.6. Machine learning approach to epitope conservation

We developed a machine learning model using the UNN algorithm to calculate epitope conservancy. The UNN algorithm operates on unlabeled datasets and relies on proximity measures to identify clusters, patterns and distance. To implement the UNN algorithm, we utilized the scikit-learn NearestNeighbors package in Python, which provided efficient and scalable nearest neighbor search capabilities. This allowed us to effectively calculate distance of each epitope to ideal matrix and identify epitope conservancy. We created a matrix where the rows represented epitopes, and the columns indicated the number of mutations and the number of mutation sites. UNN can calculate distance of each epitope based on the number of mutations and the number of mutation sites features to ideal matrix. Each column was weighted by Shannon entropy. The model calculated the distance of each epitope to the ideal matrix [0 0]. Subsequently, we normalized the obtained distances using Min-Max normalization (by deviding each distance to max distance). The normalized distance represents the conservancy score of each epitope. We plotted a scatter plot with the x-axis representing antigenicity and the y-axis representing the conservancy score. Based on this plot, we selected conserved and antigenic epitopes (Antigenicity threshold = 0.4).

3. Results

3.1. Phylogenetic analysis identifies monkeypox and vaccinia relationship and potential vaccine target

The phylogenetic analysis of Orthopoxviruses revealed that Monkeypox is most closely related to Vaccinia (Fig. 2A). A pairwise similarity analysis demonstrated that Vaccinia exhibits the highest similarity to Monkeypox virus, particularly in the confirmation of the phylogenetic tree (Fig. 2B and C). Among the proteins analyzed, the IMV membrane protein L1R showed the greatest similarity and

Table 1
The table provided herein presents a compilation of proteins meticulously chosen for epitope prediction. The initial five proteins were selected based on their Vaxign Machine Learning score and Adhesion probability, determined through Vaxign2's reverse vaccinology approach. For these proteins, their subcellular localization was determined using the Cello2go server, as Vaxign2 couldn't predict their localization, leading to an "Unknown" value in the Vaxign2 Localization column of the table. In contrast, the B22R family protein (B21R) was chosen after an exhaustive literature review due to its remarkable immunogenicity. The subcellular localization of the B22R family protein (B21R) was not determined by the Cello2go server because Vaxign2 was capable of predicting its localization.

Protein Accession	Protein Name	Vaxign-ML Score	Vaxign2 Localization (Probability)	Adhesion Probability	*Trans*-membrane Helices	Similar Human Protein	Cello2go Localization
YP_010377077.1	IMV membrane protein L1R	90.9	Unknown (Prob. = 0.20)	0.679	1	No	Plasma membrane
YP_010377148.1	Chemokine binding protein	90.9	Unknown (Prob. = 0.20)	0.583	0	No	Extracellular
YP_010377158.1	Hemagglutinin	90.9	Unknown (Prob. = 0.20)	0.789	1	No	Plasma membrane
YP_010377165.1	Soluble interferon-gamma receptor-like protein	90.9	Unknown (Prob. = 0.25)	0.676	0	No	Extracellular
YP_010377175.1	Virulence protein	90.9	Unknown (Prob. = 0.20)	0.797	1	No	Plasma membrane
YP_010377176.1	B22R family protein (B21R)	96.9	Cytoplasmic Membrane (Prob. = 0.95)	0.097	1	No	–

could serve as a target for the development of cross-protective vaccines (Fig. 2D).

3.2. Identification of key candidate protein for epitope prediction

To select target proteins for further analysis, Vaxign2 RV analysis was employed, excluding 37 out of 179 Monkeypox proteins due to their similarity with human proteins. Only five proteins with a Vaxign-ML score ≥90.0 and probability ≥0.51 were identified. However, the subcellular localization of these proteins was unknown. To address this, Cello2go was utilized, revealing that three of the selected proteins (60 %) are located in the membrane, while the remaining two (40 %) are secreted proteins. The five selected proteins are as follows: IMV membrane protein L1R (YP_010377077.1), Hemagglutinin (YP_010377158.1), Virulence protein (YP_010377175.1), Chemokine binding protein (YP_010377148.1), and Soluble interferon-gamma receptor-like protein (YP_010377165.1). The literature review indicated that the B22R family protein (B21R) is a promising choice for epitope prediction due to its possession of multiple immunodominant regions and highly immunogenic epitopes [16]. The Vaxign-ML score of this protein was 96.9, higher than that of other proteins, thus making it a key candidate for further analysis (Table 1).

3.3. Prediction and selection of MHC-I, MHC-II and linear B cell epitopes

For MHC-I epitope prediction, a total of 720 epitopes were predicted using NetCTL v1.2, out of which 351 were found to be immunogenic. Following assessment of allergenicity, toxicity, and antigenicity, 136 epitopes remained. Among these, 72 epitopes were associated with the B22R family protein (B21R). Additionally, the IEDB MHC-II tool predicted 116 epitopes, of which 40 were allergens. No toxicity was observed in the remaining epitopes. Among them, 49 epitopes were identified as interferon gamma inducers, resulting in the final selection of 35 antigenic epitopes for MHC-II. Furthermore, Bepipred2.0 predicted 107 linear B cell epitopes, of which 83 had a length of at least 6 amino acids and 43 were non-allergenic. Out of the non-allergenic epitopes, only two were found to be toxic. After evaluating antigenicity, a total of 28 epitopes were chosen for further analysis (Fig. 3A and. B, Supplementary information: sup file 3, sup file 4, sup file 5).

3.4. Mutation analysis reveals low conservation in B22R family protein (B21R)

NextClade CLI was employed to analyze 2433 sequences, eliminating 123 sequences with poor quality control. This analysis identified 2337 nucleotide positions and 3989 amino acid positions that exhibited mutations. Notably, the B22R family protein (B21R) showed 2377 mutations, indicating low conservation. The total number of amino acid mutations and the number of mutation sites for each epitope were calculated (Supplementary information: sup file 6).

3.5. Machine learning model assists in epitope selection based on antigenicity and conservancy scores

A machine learning model was trained using the UNN algorithm to calculate the distance of each epitope from the ideal matrix [0 0]. Consequently, a scatter plot was generated, with the antigenicity of the epitopes on the x-axis and the conservancy score (normalized UNN distance) on the y-axis. Based on this plot, epitopes were selected based on Normalized UNN distance and antigenicity (threshold = 0.4), resulting in the identification of 47 MHC-I epitopes, 5 MHC-II epitopes, and 10 linear B cell epitopes (Fig. 4A and. B, Supplementary information: sup file 7, sup file 8).

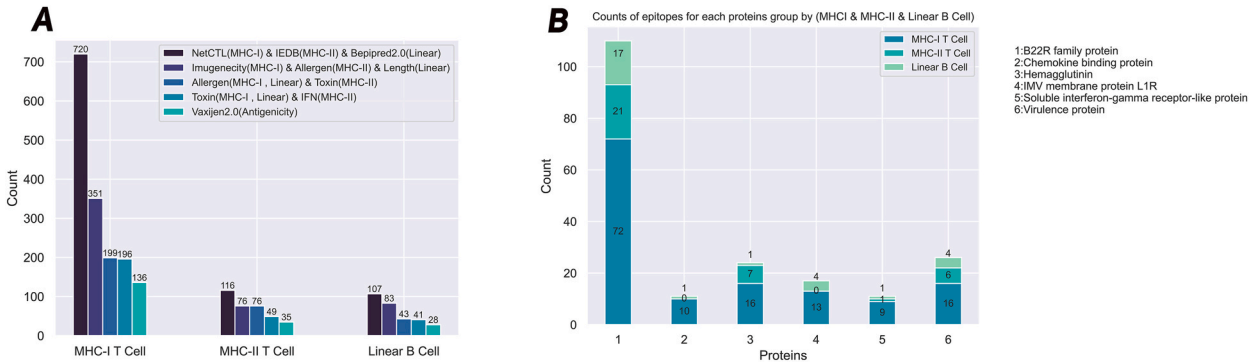


Fig. 3. Epitope prediction analysis. (A) the plot illustrates the predicted epitope counts at each stage of prediction for MHC-I, MHC-II, and Linear B cell epitopes. The legend indicates the corresponding tools used for each step in the prediction process, with the epitope counts displayed atop each column. (B) the plot showcases the final predicted epitope counts separately for each target protein. The numbers within each protein column represent the respective counts of MHC-I, MHC-II, and Linear B cell epitopes, from bottom to top.

4. Discussion

Due to the emergence of Monkeypox virus in regions where it is not commonly found, concerns have been raised about the potential for human-to-human transmission in these areas [3]. Recent studies indicate that the evolution of Monkeypox virus has been happening at an accelerated pace [13]. Our analysis of the genetic relationships reveals that the closest strain to Monkeypox virus is Vaccinia, and it is suggested that Vaccinia-based vaccines used for smallpox could be effective in preventing Mpox outbreaks, as long as the evolution of Monkeypox virus does not significantly reduce the vaccine's effectiveness [17]. Before examining the impact of evolution on the vaccine, it is important to develop new vaccine platforms specifically tailored to Monkeypox virus mutations in order to enhance prevention measures against its evolution.

In this study, we employed Reverse vaccinology and conducted a literature review to select proteins for epitope prediction. The chosen proteins include IMV membrane protein L1R (YP_010377077.1), which is an envelope protein with myristoylation [18]; Hemagglutinin (YP_010377158.1), a protein found in infected cells' membranes and viral particles' envelope [19]; Virulence protein (YP_010377175.1), a hypothetical protein similar to Vaccinia B20R proteins [20]; Chemokine binding protein (YP_010377148.1), a secreted protein from infected cells that plays a role in modulating the immune system by binding to chemokines [21]; and soluble interferon-gamma receptor-like protein (YP_010377165.1), an extracellular and secreted protein that hinders the binding of interferon-gamma to cellular receptors, thereby preventing cellular immunity [22]. According to literature findings, the B22R family protein (B21R) is a surface glycoprotein with highly immunogenic epitopes, and all individuals infected with Monkeypox virus respond to at least three epitopes of this protein [16]. employed the B22R family protein (B21R) and soluble interferon-gamma receptor-like protein [8]. Similarly, the studies conducted by Akhtar et al. utilized the IMV membrane protein L1R and Hemagglutinin for their epitope prediction [9]. Aziz et al. focused on a virulence protein [23], while Hayat et al. employed a chemokine binding protein in their vaccine design [24].

Similar to our previous study [25], we performed epitope prediction, calculated the number of mutations using 2433 sequences, and determined the total number of mutations as well as the mutation sites for each epitope. Until June 2022, a total of 1121 Single Nucleotide Polymorphisms (SNPs) had been reported. However, as of October 2022, we identified 2337 SNPs, further confirming the accelerated evolution of Monkeypox virus [3]. In comparison to other studies, our epitope prediction for the B22R family protein (B21R) identified only the IAYRNDTSF CTL epitope as common [8]. However, this particular epitope is not considered conserved in

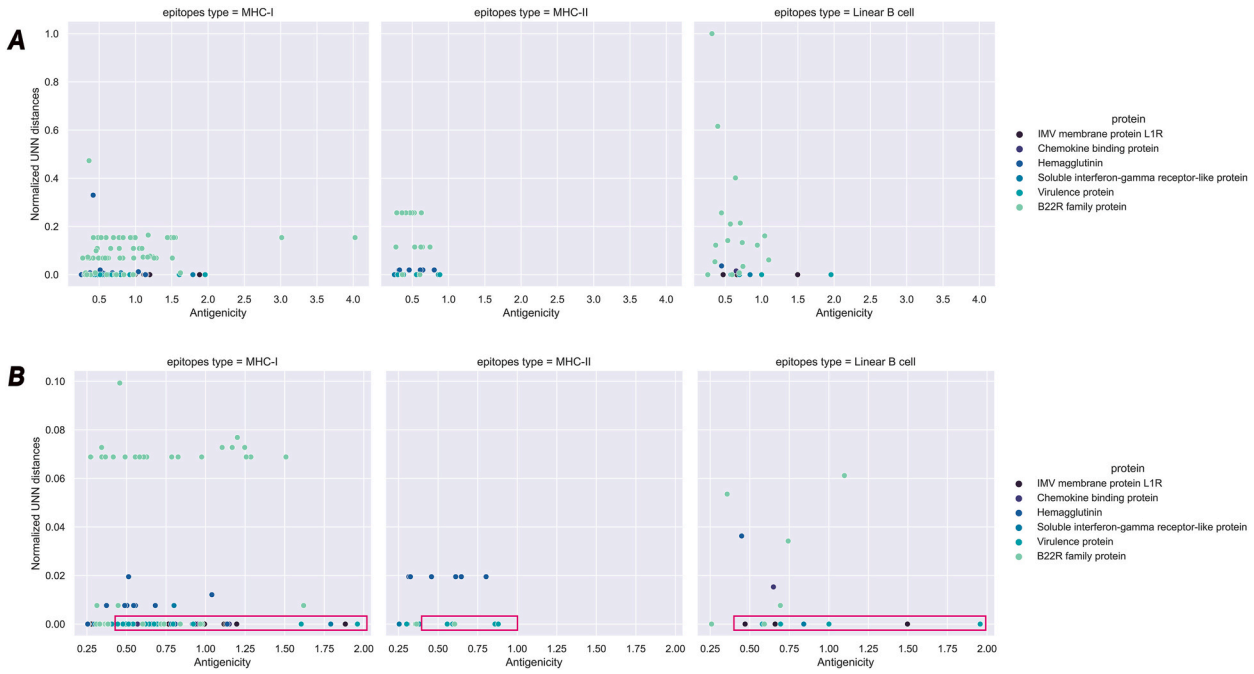


Fig. 4. Select Epitopes based on antigenicity and normalized Unsupervised Nearest Neighbor distance (A) presents a scatter plot that illustrates the antigenicity (x-axis) and normalized Unsupervised Nearest Neighbor distance of each predicted epitope (y-axis). The antigenicity of each epitope was determined using the Vaxijen2 server. The matrix contains rows of predicted epitopes, with two columns representing the number of mutations and the number of mutation sites, which were used to calculate the distance of each epitope from the ideal [0 0] matrix. The Unsupervised Nearest Neighbor distance of each epitope was normalized by dividing it by the maximum distance, resulting in a conservancy score for each epitope. This scatter plot was utilized to select conserved and antigenic epitopes. (B) an enlarged version of Fig. 4A's scatter plot is presented. Using this scaled-up scatter plot, specific epitopes were chosen based on the criteria of having an antigenicity value greater than or equal to 0.4 and a Normalized Unsupervised Nearest Neighbor distance of 0. Subsequently, a total of 47 MHC-I epitopes, 5 MHC-II epitopes, and 10 Linear B cell epitopes were identified and discovered (red rectangle shows selected epitopes and antigenicity threshold = 0.4).

our analysis due to the presence of 10 mutations across 9 amino acid positions (UNN distance = 13.4536 and Normalized UNN distance = 0.0727), indicating a low level of conservation. In contrast, the previous articles report a conservancy of 66 % for this epitope using alignment and pairwise similarity methods, and they utilize it as a conserved epitope for vaccine design [8]. This discrepancy might be attributed to their use of alternative methods and their analysis limited to only six MPXV isolates, whereas our study encompasses 2433 Monkeypox virus sequences. Furthermore, when considering other articles that investigate similar target proteins, we do not identify any common epitopes. This disparity could be attributed to differences in the approaches and tools employed, or the omission of certain epitopes in the main body of those articles.

Most other studies commonly employ IEDB Conservancy tools or BLASTp to calculate the conservancy of predicted epitopes. However, when compared to our method, these studies often face limitations in calculating epitope conservancy due to the restricted number of protein sequences considered. Our approach, on the other hand, utilizes NextClade CLI and machine learning, enabling us to encompass a larger number of proteins and achieve more accurate conservancy calculations. Nonetheless, there is a drawback to our method in that it determines the antigenicity of conserved epitopes based on the reference sequence among all strains, neglecting the importance of epitope frequency across sequences. If an epitope has a low frequency among all strains, its efficacy may be diminished. For instance, our method may indicate high conservancy for an epitope due to a single mutation, but if it has a low frequency across strains, the antigenicity may be lower. To address this concern, we propose incorporating the determination of strain development in specific areas where vaccine design is intended, in combination with the frequency of epitopes using the Shannon entropy method. Furthermore, considering the antigenicity of more frequently occurring epitopes could potentially resolve this issue within our method.But in general the selection of conserved epitopes based on the UNN algorithm represents a novel approach that can be employed for organisms exhibiting a high number of mutations and rapid evolution. We anticipate that the identified epitopes will contribute to the design of an effective vaccine that can help prevent the spread of Mpox. Our methodology holds promise for epitope selection and vaccine design for viruses characterized by high mutation rates.

5. Conclusion

In this study, immunoinformatics and machine learning approaches were employed to predict and select conserved epitopes for the design of immunogenic vaccines against Mpox. The phylogenetic analysis revealed that Monkeypox virus is closely related to Vaccinia, suggesting the potential cross-protective effect of Vaccinia-based vaccines against Mpox. Reverse vaccinology analysis and literature review identified several target proteins, including the B22R family protein (B21R), which is known to have highly immunogenic epitopes. Epitope prediction was performed using various tools, and a total of 47 MHC-I epitopes, 5 MHC-II epitopes, and 10 linear B cell epitopes were selected. Machine learning was utilized to calculate epitope conservancy scores, allowing the identification of epitopes with high antigenicity and conservation. These findings provide valuable insights for the development of new-generation vaccines against Mpox.

5.1. Limitations

The study's limitations encompass several key aspects. Firstly, the reliance on publicly available Monkeypox virus sequences might not fully encompass the diverse strains, impacting epitope prediction accuracy. The utilization of existing epitope prediction tools and databases, subject to ongoing evolution, could affect the reliability of predictions. Experimental validation of predicted epitopes is essential, as the study primarily focuses on in silico analysis. Furthermore, the potential emergence of novel strains and antigenic variations necessitates continuous epitope prediction updates. While the study provides a foundation for vaccine design, challenges such as formulation, safety, and regulatory considerations remain. Overall, the study offers valuable insights into Mpox vaccine development, but further research, validation, and adaptation to evolving strains are crucial for advancement.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

All data generated or analyzed during this study are included in this published article.

CRediT authorship contribution statement

Mohammad Izadi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Fatemeh Mirzaei:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Mohammad Aref Bagherzadeh:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Shamim Ghiabi:** Writing – original draft, Formal analysis. **Alireza Khalifeh:** Writing – original draft.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT3.5 / OpenAI in order to revised and quality of language of the

article. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

There is no funding for the current study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24972.

References

[1] E. Alakunle, et al., Monkeypox virus in Nigeria: infection biology, epidemiology, and evolution, Viruses 12 (11) (2020).
[2] N. Chen, et al., Virulence differences between monkeypox virus isolates from West Africa and the Congo basin, Virology 340 (1) (2005) 46–63.
[3] L. Wang, et al., Genomic annotation and molecular evolution of monkeypox virus outbreak in 2022, J. Med. Virol. 95 (1) (2023) e28036.
[4] S.N. Shchelkunov, S.S. Marennikova, R.W. Moyer, Orthopoxviruses Pathogenic for Humans, Springer Science & Business Media, 2006.
[5] P.-Y. Nguyen, et al., Reemergence of human monkeypox and declining population Immunity in the context of urbanization, Nigeria, 2017–2020, Emerg. Infect. Dis. 27 (4) (2021) 1007.
[6] M.F. Gruber, Current status of monkeypox vaccines, npj Vaccines 7 (1) (2022) 94.
[7] S. Sanami, et al., In silico design and immunoinformatics analysis of a universal multi-epitope vaccine against monkeypox virus, PLoS One 18 (5) (2023) e0286224.
[8] M. Waqas, et al., Immunoinformatics design of multivalent epitope vaccine against monkeypox virus and its variants using membrane-bound, enveloped, and extracellular proteins as targets, Front. Immunol. 14 (2023).
[9] N. Akhtar, et al., Immunoinformatics-aided design of a peptide based multiepitope vaccine targeting glycoproteins and membrane proteins against monkeypox virus, Viruses 14 (11) (2022).
[10] A.N. Oli, et al., Immunoinformatics and vaccine development: an overview, ImmunoTargets Ther. 9 (2020) 13.
[11] E. Mohanty, A. Mohanty, Role of artificial intelligence in peptide vaccine design against RNA viruses, Inform. Med. Unlocked 26 (2021) 100768.
[12] E. Ong, et al., Vaxign2: the second generation of the first Web-based vaccine design program using reverse vaccinology and machine learning, Nucleic Acids Res. 49 (W1) (2021) W671–w678.
[13] J. Isidro, et al., Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus, Nat. Med. 28 (8) (2022) 1569–1572.
[14] C.S. Yu, et al., CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation, PLoS One 9 (6) (2014) e99368.
[15] I. Aksamentov, et al., Nextclade: clade assignment, mutation calling and quality control for viral genomes, J. Open Source Softw. 6 (67) (2021) 3773.
[16] E. Hammarlund, et al., Multiple diagnostic techniques identify previously vaccinated individuals with protective immunity against monkeypox, Nat. Med. 11 (9) (2005) 1005–1011.
[17] A.K. Rao, et al., Use of JYNNEOS (smallpox and monkeypox vaccine, live, nonreplicating) for preexposure vaccination of persons at risk for occupational exposure to Orthopoxviruses: recommendations of the advisory committee on immunization practices - United States, 2022, MMWR Morb. Mortal. Wkly. Rep. 71 (22) (2022) 734–742.
[18] H.-P. Su, et al., The 1.51-Å structure of the poxvirus L1 protein, a target of potent neutralizing antibodies, Proc. Natl. Acad. Sci. USA 102 (12) (2005) 4240–4245.
[19] B. Moss, Poxvirus entry and membrane fusion, Virology 344 (1) (2006) 48–54.
[20] M.R. Mauldin, et al., Exportation of monkeypox virus from the african continent, J. Infect. Dis. 225 (8) (2022) 1367–1376.
[21] J.M. Jones, et al., Monkeypox virus viral chemokine inhibitor (MPV vCCI), a potent inhibitor of rhesus macrophage inflammatory protein-1, Cytokine 43 (2) (2008) 220–228.
[22] A. Alcamí, G.L. Smith, Soluble interferon-gamma receptors encoded by poxviruses, Comp. Immunol. Microbiol. Infect. Dis. 19 (4) (1996) 305–317.
[23] S. Aziz, et al., Contriving multi-epitope vaccine ensemble for monkeypox disease using an immunoinformatics approach, Front. Immunol. 13 (2022) 1004804.
[24] C. Hayat, et al., Design of a novel multiple epitope-based vaccine: an immunoinformatics approach to combat monkeypox, J. Biomol. Struct. Dyn. (2022) 1–12.
[25] M.A. Bagherzadeh, et al., Considering epitopes conservity in targeting SARS-CoV-2 mutations in variants: a novel immunoinformatics approach to vaccine design, Sci. Rep. 12 (1) (2022) 14017.

Abbreviations

NCBI -: National Center for Biotechnology Information
RV -: Reverse Vaccinology
MHC-I -: Major Histocompatibility Complex class I
MHC-II -: Major Histocompatibility Complex class II
IEDB -: Immune Epitope Database and Analysis Resources
SNP -: Single Nucleotide Polymorphism
UNN –: Unsupervised Nearest Neighbor
NextClade CLI: NextClade Command Line Interface