



OPEN ACCESS

EDITED BY

Tarek A. Ahmad,
Bibliotheca Alexandrina, Egypt

REVIEWED BY

Sudha Chaturvedi,
Wadsworth Center, United States
Hamid Morovati,
Shiraz University of Medical Sciences,
Iran

*CORRESPONDENCE

Mugdha Srivastava
mugdha.srivastava@uni-wuerzburg.de
Thomas Dandekar
dandekar@
biozentrum.uni-wuerzburg.de

†These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Infectious Diseases – Surveillance,
Prevention, and Treatment,
a section of the journal
Frontiers in Medicine

RECEIVED 31 July 2022

ACCEPTED 17 October 2022

PUBLISHED 03 November 2022

CITATION

Gupta SK, Osmanoglu Ö, Minocha R,
Bandi SR, Bencurova E, Srivastava M
and Dandekar T (2022) Genome-wide
scan for potential CD4+ T-cell
vaccine candidates in *Candida auris*
by exploiting reverse vaccinology
and evolutionary information.
Front. Med. 9:1008527.
doi: 10.3389/fmed.2022.1008527

COPYRIGHT

© 2022 Gupta, Osmanoglu, Minocha,
Bandi, Bencurova, Srivastava and
Dandekar. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Genome-wide scan for potential CD4+ T-cell vaccine candidates in *Candida auris* by exploiting reverse vaccinology and evolutionary information

Shishir K. Gupta^{1,2†}, Özge Osmanoglu^{1†}, Rashmi Minocha^{3†},
Sourish Reddy Bandi^{1,4}, Elena Bencurova¹,
Mugdha Srivastava^{1,5*} and Thomas Dandekar^{1,6*}

¹Department of Bioinformatics, Biocenter, Functional Genomics and Systems Biology Group, University of Würzburg, Würzburg, Germany, ²Evolutionary Genomics Group, Center for Computational and Theoretical Biology, University of Würzburg, Würzburg, Germany, ³Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India, ⁴Institute of Experimental Biomedicine, University Hospital Würzburg, Würzburg, Germany, ⁵Core Unit Systems Medicine, University of Würzburg, Würzburg, Germany, ⁶BioComputing Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Candida auris is a globally emerging fungal pathogen responsible for causing nosocomial outbreaks in healthcare associated settings. It is known to cause infection in all age groups and exhibits multi-drug resistance with high potential for horizontal transmission. Because of this reason combined with limited therapeutic choices available, *C. auris* infection has been acknowledged as a potential risk for causing a future pandemic, and thus seeking a promising strategy for its treatment is imperative. Here, we combined evolutionary information with reverse vaccinology approach to identify novel epitopes for vaccine design that could elicit CD4+ T-cell responses against *C. auris*. To this end, we extensively scanned the family of proteins encoded by *C. auris* genome. In addition, a pathogen may acquire substitutions in epitopes over a period of time which could cause its escape from the immune response thus rendering the vaccine ineffective. To lower this possibility in our design, we eliminated all rapidly evolving genes of *C. auris* with positive selection. We further employed highly conserved regions of multiple *C. auris* strains and identified two immunogenic and antigenic T-cell epitopes that could generate the most effective immune response against *C. auris*. The antigenicity scores of our predicted vaccine candidates were calculated as 0.85 and 1.88 where 0.5 is the threshold for prediction of fungal antigenic sequences. Based on our results, we conclude that our vaccine candidates have the potential to be successfully employed for the treatment of *C. auris* infection. However, *in vivo* experiments are imperative to further demonstrate the efficacy of our design.

KEYWORDS

Candida auris, T-cell epitope, epitope prediction, positive selection, evolution, immune-informatics

Introduction

In recent years, life-threatening fungal diseases have increased, and new infections have emerged. Selection pressure of climate change has significantly contributed to the emergence of *Candida auris* as a pathogen (1). *C. auris*, a multidrug-resistant ascomycete, was first isolated in 2009 in Japan. Up to date, it has been detected in 32 countries over six continents (2). *C. auris* infects various tissues and organs, including central nervous system (3), cardiovascular system (4), respiratory tract (5), bones and joints (6), and possesses the potential to cause nosocomial infections (7). The diagnosis of *C. auris* is difficult due to non-availability of specific laboratory techniques which could rapidly and accurately detect it in clinical samples. Furthermore, treatment of infections caused by *C. auris* poses a real challenge due to its high multidrug-resistivity pattern. The reduced susceptibility to conventional antifungal drugs such as azoles and amphotericin B leads to high mortality rate, rising up to 60% (8, 9). Furthermore, *C. auris* can persist on abiotic surfaces such as healthcare instruments for several weeks, which can facilitate prolonged pathogen persistence and high transmissibility (10). This also explains why transmission of *C. auris* to hosts frequently occurs in hospitals. Its persistence in the environment is a trait that differentiates it from other *Candida* species.

Development of anti-fungal vaccines which can trigger the host immune response to generate immunological memory against fungi and their spores remains challenging. Nevertheless, three approaches of immunization are commonly used in development of anti-fungal vaccines; (i) vaccination with live-attenuated strains, which can be hazardous due to possible disease development in immunocompromised patients (11), (ii) immunization with recombinant proteins containing the immunogenic sequences (epitopes), or (iii) immunization with polysaccharides, which are present only in fungal cell wall (12).

To date, there is no human vaccine approved against *C. auris* infections, however, several research groups have made significant progress in its development. Currently, most studies are focused on the Als3, a member of agglutinin-like sequence (Als) family of proteins with adhesive and invasive properties (13). Researchers have proposed Als3 as a promising therapeutic target for *Candida albicans* vaccine development (14–17). In a recent study, Singh et al. (18) identified three adhesin/invasin proteins in *C. auris*, that shared sequence and structural homology to Als3 protein of *C. albicans*. The NDV-3A vaccine proposed by Singh et al. (18) which was based on the N-terminus of Als3-protein sequence significantly blocked biofilm production ability of *C. auris in vitro*. These authors also found that when combined with the antifungal drug micafungin, the NDV-3A vaccine augmented the protective efficacy of this drug against *C. auris* infection in neutropenia mice. It further induced regulatory CD4+ T helper (TH) cells in infected mice,

which comprised of Th1, Th2, and Th17 subcellular population (18). In another study, immunizing mice with rAls3p-N vaccine induced T-cell mediated protection in *C. albicans*, which further signifies the crucial role of CD4+ T lymphocytes as well as associated cytokines such as IFN- γ in acquired immune response against the pathogen (19, 20).

The objective of the current study was to identify novel putative CD4+ T-cell epitopes with vaccine potential against *C. auris* infection by harnessing the evolutionary information combined with reverse vaccinology approach. To this end, we extensively scanned the family of proteins coded by *C. auris* genome and identified the potential vaccine candidates. In search of epitope-based vaccine candidates, one could expect that the pathogen may acquire substitutions in epitopes targeted by immune memory over the years. For instance, hemagglutinin protein of influenza A evolves under strong selection from antibodies (21). To avoid such possibilities, we computed non-synonymous and synonymous distances and then tested for sites with statistical evidence where the accumulation of non-synonymous substitutions exceeds that of the synonymous substitutions. We thus discarded all the rapidly evolved genes of *C. auris* with positively selected sites. We further used the highly conserved regions of protein sequence alignment of multiple *C. auris* strains, so as to identify major candidates for designing the new potential vaccine. This subsequently decreased the possibility of epitopes to rapidly evolve and escape the immune recognition. Moreover, using the stringent *in silico* analysis, we finally identified two conserved immunogenic and antigenic CD4+ T-cell epitopes that could be used for efficient immune memory generation against *C. auris*.

Materials and methods

Genome sequences and quality control

The protein coding genes and translated transcriptomes of five *C. auris* strains (*C. auris* 6,684 from India, *C. auris* B8441 from Pakistan, *C. auris* B11220 from Japan, *C. auris* B11221 from South Africa, and *C. auris* B11243 from Venezuela) were retrieved from NCBI GenBank release 231.0 (22). The annotation completeness of the genome assemblies was accessed with BUSCO v3 (23) using Fungi odb9 database that contains single-copy orthologs (SCOs) selected from OrthoDB v9 (24, 25). **Supplementary Table 1** lists the version and details of the genomic assembly.

Identification of orthologs

Clustering of orthologous genes was performed using Orthofinder v2.3.3 (26) with Diamond v0.9.24.125 (27) (under default settings) based on translated proteomes of selected

five strains. The orthologous clusters that did not contain orthologous genes from all the selected five strains were discarded to only keep genes conserved in all strains. Among the genes in each conserved orthologous clusters, genes that belong to *C. auris* B11220 (4,860 genes in total) were used as reference for further analysis due to higher quality and annotation completeness of its assembly.

Filtering of candidate genes

Conserved *C. auris* B11220 genes were filtered to eliminate the unlikely vaccine candidates. **Table 1** list the tools used for filtering. The first round of elimination was performed based on their secretion or cellular localization. Secretion of the proteins was predicted by checking whether they have a signal peptide or a glycosylphosphatidylinositol (GPI)-anchor or a transmembrane (TM) domain. Signal peptides were identified by using TargetP (28), SignalP (29), Phobius (30), and FunsecKB2 (31). GPI-anchor were identified using PredGPI (32). To identify TM domain, TMHMM (33) and Phobius (30) were used.

Next, proteins with any of the features predicting their sorting *via* classical secretion pathway were further filtered based on the presence of endoplasmic reticulum (ER) retention signal that may restrict their secretion to extracellular space or cell membrane. This step of filtering through the detection of PROSITE pattern PS00014 was performed by ScanProsite (34). Last step of the first round of elimination was done by sorting the proteins by the presence of a signal peptide or a TM domain or a GPI anchor. Then the cellular localization of the proteins was assessed by Deeploc (35). Accordingly, the proteins with a signal peptide were filtered further by their cellular localization being cell membrane, while the proteins with TM domain or GPI anchor were reduced to the ones with extracellular localization.

Second round of filtering was based on fungal effectors and adhesins which possess a higher probability of being a vaccine target (36). The filtering was performed using prediction tools EffectorP (37), FungalRV (38), and FaaPred (39) on both sets of proteins that were subject to classical secretion pathway either to the cell membrane or extracellular space.

Positive selection analysis

Proteins identified from the previous step were aligned with their orthologs from other *C. auris* strains under study. T-Coffee (40) was used for the multiple sequence alignment (MSA) which combines the output of different aligners to enhance the MSA accuracy. The aligned amino acid sequences together with the corresponding nucleotide sequences of each ortholog group were converted into nucleotide alignments at the codon level using the program PAL2NAL (41). Since removal

of unreliable regions increases the power to detect positive selection, we used stringent Gblocks filtering (type = codons; minimum length of a block = 4; no gaps allowed) to remove gap-rich regions from the alignments (42). We used site model implemented in codeml program from the PAML 4.2b package (43) to detect the sites in the alignment under positive selection. The Bayes empirical Bayes approach was employed to estimate the probabilities of positive selection for specific codons under the likelihood framework (44). FEL (45) and MEME (46) were used to evaluate for positive selection among filtered genes. The cut off parameter was set to $p < 0.01$. If the positive selection was detected by any of mentioned three programs, the genes were removed for the subsequent analysis.

Highly conserved regions of ortholog clusters

Conservation of amino acid residues in the aligned pathogen sequences was estimated by Shannon entropy function (47) using the Protein Variability Server.¹ We selected the Shannon entropy variability threshold of 1.0 to extract the highly conserved consensus subsequences of length >9 mers. The Shannon entropy (S) for every position in the sequence alignment was calculated as

$$S = - \sum_{i=0}^N F_i \log_2 F_i$$

where F_i is the fraction of residues of amino acid type i , and N is the number of amino acid types (48).

CD4+ T-cell binding peptide prediction

The peptides that trigger major histocompatibility complex (MHC) class-II responses are often longer than class-I peptides. These MHC binders are <13–15 amino acids longer with a core sequence of about 9 mers, usually three anchor residues, and their ends extend beyond the peptide-binding groove (49, 50). The MHC class-II binding 13–15 mer peptides for alleles were predicted using Immune Epitope Database (IEDB) (51) recommended prediction method and on full HLA reference set with percentile rank.

Host similarities and antigenicity of MHC class-II binders

To assess the similarity of the predicted MHC class-II binders with human sequences, Blast similarity search

¹ <http://imed.med.ucm.es/PVS>

TABLE 1 Web-servers and databases used for sequence filtering.

Name	Description	Method	References
DeepLoc 2.0	Eukaryotic protein subcellular localization	Machine learning-deep neural networks	97
EffectorP 2.0	Fungal effector proteins	Machine learning	37
PredGPI	GPI-anchor prediction	Support vector machine (SVM) and hidden markov model (HMM)	32
TargetP	Eukaryotic protein subcellular localization	Neural network	98
FungalRV	Fungal adhesin prediction	SVM	38
Phobius	Transmembrane topology and signal peptide prediction	Hidden markov model (HMM)	99
SignalP 6.0	Signal peptide prediction	Deep neural network	100
TMHMM 2.0	Transmembrane helices prediction	Hidden markov model	33
FAAPred	Fungal adhesins and adhesin-like protein prediction	Support vector machine (SVM)	39
FunSecKB2	A fungal protein subcellular location knowledgebase		101
ScanProsite	PROSITE signature detection		34

(52) was performed. The Blast parameters were tailored appropriately for this analysis, in accordance with the short size of the peptides (such as matrix PAM30; word-size 2). We first screen the binders against human assembly from ENSEMBL (assembly version GRCh38.p13) and then against NCBI non-redundant human sequence database. Highly similar sequences were discarded from further evaluation. We further performed Blast analysis against a customized sequence database of experimentally verified autoimmune class-II epitopes downloaded from IEDB database (51) to avoid the induction of potential autoimmune reactions. Only the sequence with less than 35% sequence identity over 80% query coverage were evaluated for the antigenicity using antigenicity prediction server² and Vaxijen server (53). VaxiJen is based on auto cross covariance (ACC) transformation (53) of protein sequences into uniform vectors of principal amino acid properties, with a threshold value 0.5 for prediction of fungal antigenic sequences. Only consensus predictions from the antigenicity prediction server were considered.

Population coverage

Before population coverage analysis the filtered epitopes binding affinity for HLA class II alleles was accessed using MHC-II prediction tool at IEDB using the neural align method 2.3 (NetMHCII 2.3) and epitope length 11–15 amino acids. Resulting alleles were sorted according to their IC₅₀ (inhibition concentration 50) values and only alleles with IC₅₀ equal to or less than 500 nM were selected for further analysis. For the calculation of the population coverage, selected alleles were examined by IEDB Population coverage tool. This analysis was performed for 16 different geographical regions and global population.

² <http://imed.med.ucm.es/Tools/antigenic.pl>

Results

With the aim to identify novel putative CD4+ T-cell epitopes against *C. auris* infection, a pipeline was developed including several steps based on orthology, cellular localization, positive selection, MSAs, T-cell binding potential, antigenicity, and population coverage.

Candida auris B11220 as representative of five strains

Genome assemblies of *C. auris* strains originating from different geographical locations were used for the first step of assessing possible common genes between different strains to be adopted as potential vaccine targets. All the used strains had almost similar sizes in the range of 12.1–12.7 Mb with GC-content of ~45%. These genomes were originally assembled at scaffold level with scaffold N50 ranging between 60 and 2400×10^3 bp (Supplementary Table 1). Next, the Benchmarking Universal Single-Copy Orthologs (BUSCO) (23) software was used for assessment of the quality and completeness of assembly and annotation of the used *C. auris* strains. Employing hidden Markov model (HMM) (54) profiles for a match between given proteome and the OrthoDBv9 (24, 25) SCO dataset for fungi, BUSCO determines the fraction of complete, fragmented, and missing genes for each genome. In other words, it assesses the quality of the genomic data based on expected gene content, which is determined/guided by the fungal SCO dataset. As mentioned above, low quality caused by technical issues in genome assemblies can be detected by metrics like N50 for contigs or scaffolds. However, quality of the genome resource that can be affected by, for instance, contamination, is also influential for the comparative downstream analyses. Therefore, to detect any quality issues with the genomic resource, BUSCO uses an expected gene set of SCO as marker genes for genome/assembly completeness (55).

Genes that were marked as complete were further classified in single-copy and duplicated and quantification of all fractions was thereby assessed by BUSCO. The *C. auris* 6,684 strain assembly showed the lowest completeness level with 33 missing genes and five fragmented genes, whereas the highest quality of assembly and annotation was observed for *C. auris* strains B11220 and B11243 with 285 out of 290 complete genes (Figure 1). We identified a total number of 5,331 orthologous groups (OG) among five strains of *C. auris*. In our analysis, 4,752 out of 5,331 OGs were found to encompass all five strains and 4,616 OG were found to be single copy orthologs clusters. B12221 strain of *C. auris* was determined to have the highest number of duplication events with 49 duplications and the highest number of genes with 118 genes to have more than one copy. *C. auris* 6,684 strain was found to be missing in 405 OG, missing in highest number of OG among five strains.

Based on our orthology analysis, assembly quality, and genome completeness, *C. auris* strain B11220 was selected for further analysis and steps of filtering of orthologous gene clusters were followed as shown in Figure 2.

Filtering pipeline identifies candidate genes with possible extracellular or membrane localization

First, filtering was performed for identification of genes with possible signaling routes by controlling following features: the existence of a signal peptide with TargetP (28), SignalP (29), Phobius (30), and FunsecKB2 (31), the existence of a GPI anchor with PredGPI (32), and the existence of a TM domain with TMHMM (33) and Phobius (30). Elimination of genes with no probability of being signaled into or through membranes led to a reduction to 2,381 genes that were further eliminated in case of the presence of ER retention signal. Further classification of these signaling proteins was performed through assessment of location by testing the presence of a TM domain or a GPI-anchor (30, 32, 33). Almost half (1,180) of 2,366 possibly signaled proteins were found to possess TM domains or a GPI-anchor. These 1,180 proteins detected here were then further filtered by their location determined by DeepLoc (35). A total of 280 proteins were detected with no cell membrane, and possibly secreted *via* the classical secretion pathway. On the other hand, the other half without TM domains or a GPI-anchor (1,186 proteins) were filtered down to 91 proteins by their location being extracellular. Next step identified 23 proteins for each class as effector or adhesion proteins that may have a higher probability to be a good vaccine target. Out of 23 membrane effector or adhesion proteins, 4 were filtered due to high similarity to human proteins while for extracellular effector or adhesion proteins only 3 were eliminated in this step. This last step ended in 19 membrane and 20 extracellular proteins with

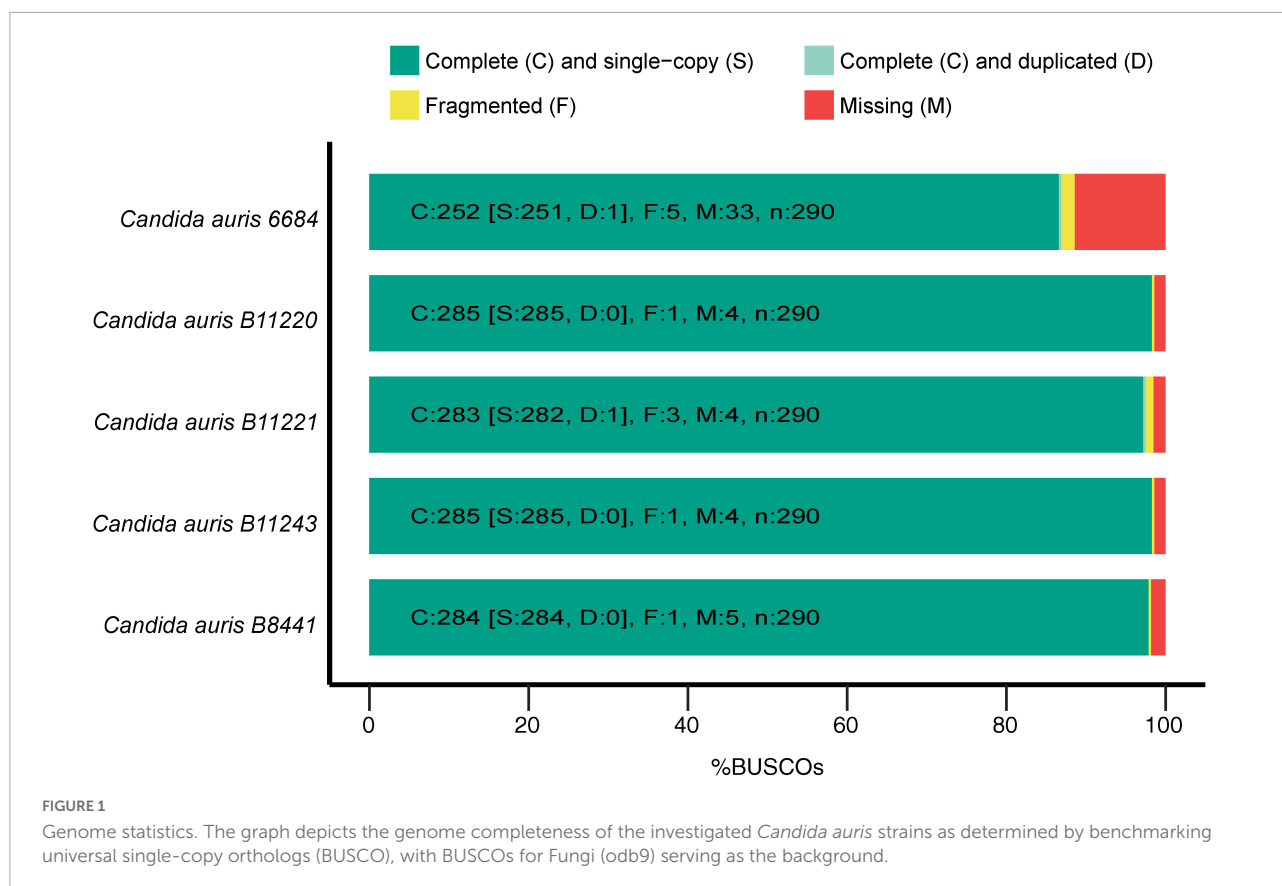
low or no similarity to human proteins that reduces the risk of cross-reactivity with host's self-proteins.

Next, the 39 proteins clustered in 38 orthogroups were checked for signs of positive selection. Out of 38 gene alignments, 5 were identified by our pipeline with sites under positive selection (Supplementary Dataset 1). PAML (43) identified sites under selection in orthogroups OG0003318 and OG0003714. Selection in OG0000005 was identified by both FEL (45) and MEME (46). In addition, MEME also identified selection in OG0000033 and OG0004411. These genes were removed for the subsequent analysis. PAML identifies the sites that could be under continuous changes while MEME identifies the episodic positive selection which implies that site changes are kept in clade to provide the advantage in the new environment. As a consequence of selection pressure from the host environment, *C. auris* like other pathogens can undergo evolutionary changes for enhanced survival.

Conserved regions reveal immunogenic and antigenic peptides

For the proteins coded from the genes with no sign of positive selection, we then focused on highly conserved regions of their MSA. This is particularly important to increase the coverage of vaccine candidates as the regions of low Shannon entropy would theoretically remain conserved even upon adding the new *C. auris* strains in the MSA. Therefore, we selected the candidate proteins from conserved regions using Shannon entropy as previously used (56–58). Shannon entropy reflects the degree of variability of protein sequence fragments and supports their evolutionary stability inferences. Stable peptides are characterized by low entropy and an entropy value of 2.0 indicates conserved fragments. With increasing variability of a site, entropy increases and is influenced by both the number of variants at that site and their respective frequency. The estimated average conservancy less than 1.0 postulated that the proposed epitopes would be highly conserved among used *C. auris* strains. We used these conserved fragments originated from 38 alignments and in total identified 7,149 CD4+ T-cell binders.

If the predicted peptides share a higher identity with host genomic regions, they can be imitated as self-molecules. The ability to prevent immunological responses against self-antigens is advantageous which would liberate the vaccine from the risk of inducing autoimmunity. Therefore, to avoid the possibility of causing autoimmunity because of the homology of predicted MHC class-II binders with humans, we performed the blast similarity search analysis (see Materials and methods). For searching against IEDB auto IEDB we used stringent cutoff and only selected the MHC-binders that have less than 35% identity over 80% sequence coverage. This removed >99% of predicted MHC-binders and only 6 MHC-binders could cross



this threshold (**Supplementary Table 2**). Such reduction was not surprising because of the close evolutionary distance between the eukaryotic pathogen *C. auris* and humans (1,105 Mya) compared to bacterial (such as *Mycobacterium tuberculosis*; 4,290 Mya) pathogens as estimated with TimeTree (59).

Furthermore, only the ability to bind to the MHC receptor does not guarantee that the binder peptide is antigenic. Out of six, only two predicted MHC class-II binders were found to be antigenic. VaxiJen antigenicity score of QTTCFQTEYYDPYIS and FVDPKCCCDPKMIK was 0.85 and 1.88, respectively with probable fungal antigen annotation at 0.5 threshold (**Figure 3A**). The 3D structure of these epitopes is shown in **Figure 3B**.

High population coverage is required for the recognition of epitopes

The MHC class II molecules are known to be highly polymorphic and are critically involved in recognition and defense against pathogens. The distribution of MHC alleles differs among various ethnic groups worldwide. Therefore, a peptide that acts as a T-cell epitope in majority of a population with a particular MHC alleles distribution may not work in a population with a different MHC alleles distribution. Thus,

allelic distribution is crucial also for the adaptation of the population to environmental changes and it varies among different geographical regions (60). Both identified epitopes were therefore submitted to IEDB server as described in methodology section. Only MHC class II binders with strong to intermediate binding affinity (IC_{50} less or equal to 500 nM) were selected to calculate population coverage (**Figure 3C** and **Table 2**). The analysis of both epitopes revealed high coverage for the global population (99.09%). Therefore, we assume that the identified epitopes can be recognized by sufficient HLA alleles and cover most of the world's population. However, we noted significantly lower binding abilities in Central American (57.2%) and South African (45.98%) regions (**Figure 4**), thus only approximately half of the population can recognize selected epitopes. The low binding ability is due to the small number of the epitopes (only 2 epitopes were examined) and low IC_{50} value. The IC_{50} value ≤ 500 was selected to obtain more accurate results when even lower load of fungi can be recognized by immune cells, which is crucial mainly for the immunocompromised patients. However, selecting more benevolent values, the number of the binding alleles would increase and thus also the probability for the better coverage in Central American and South African population, where candidemia caused by *C. auris* accounts for 10% of all cases (61).

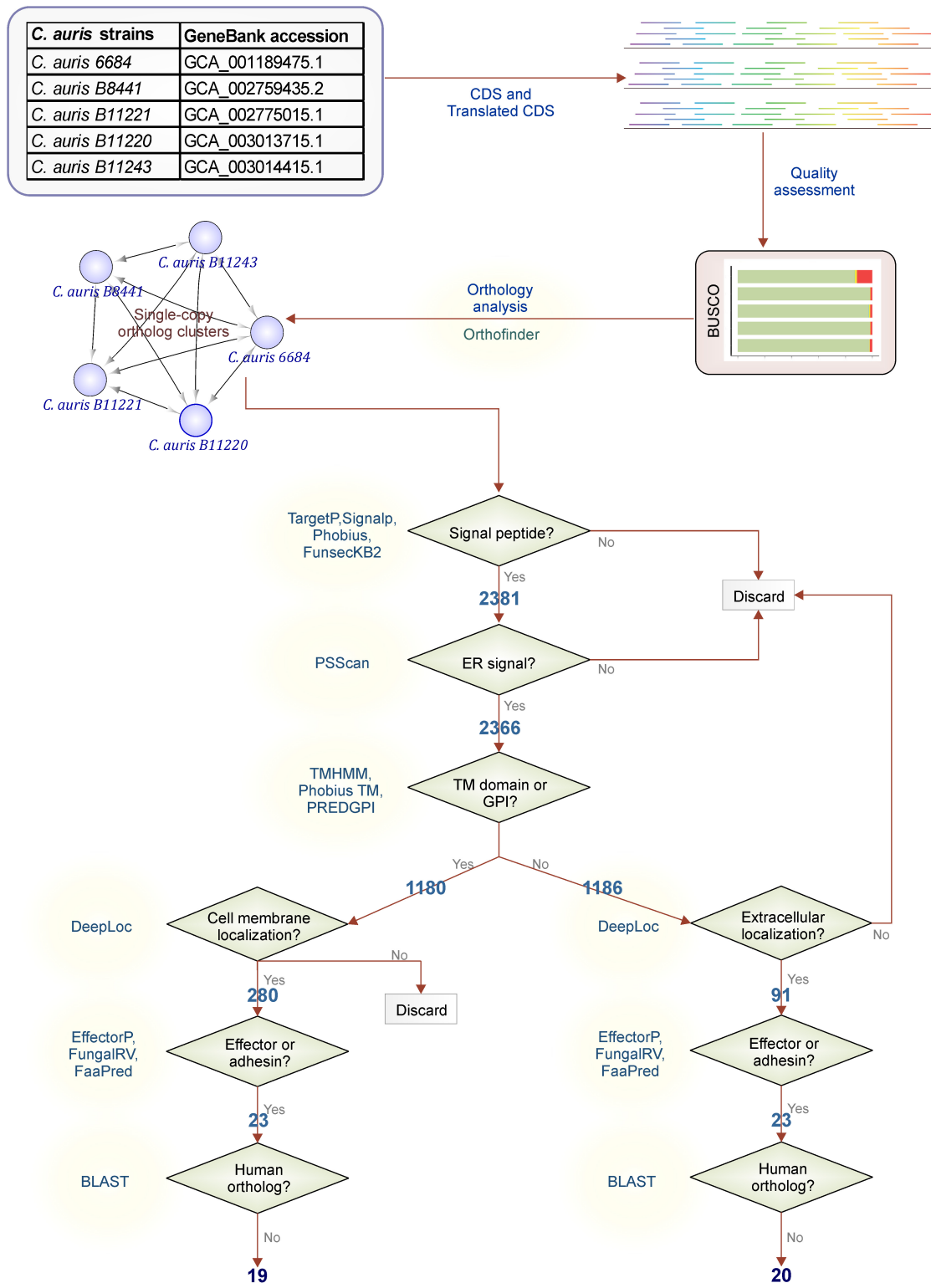


FIGURE 2

Filtering workflow. Steps used for *C. auris* protein filtering adapted from Vivek-Ananth et al. (79) The *C. auris* proteins with orthologs in all strains were filtered to obtain membrane proteins or extracellular secreted proteins. Secretion to the extracellular space or to the membrane was determined by existence of a signal peptide or glycosylphosphatidylinositol (GPI) anchor or a transmembrane (TM) domain. The proteins with an endoplasmic reticulum (ER) signal peptides were filtered out to ensure the localization of the protein to the membrane or extracellular space. Further steps were taken to predict the effector or adhesins and to filter out proteins that may invoke cross-reactivity.

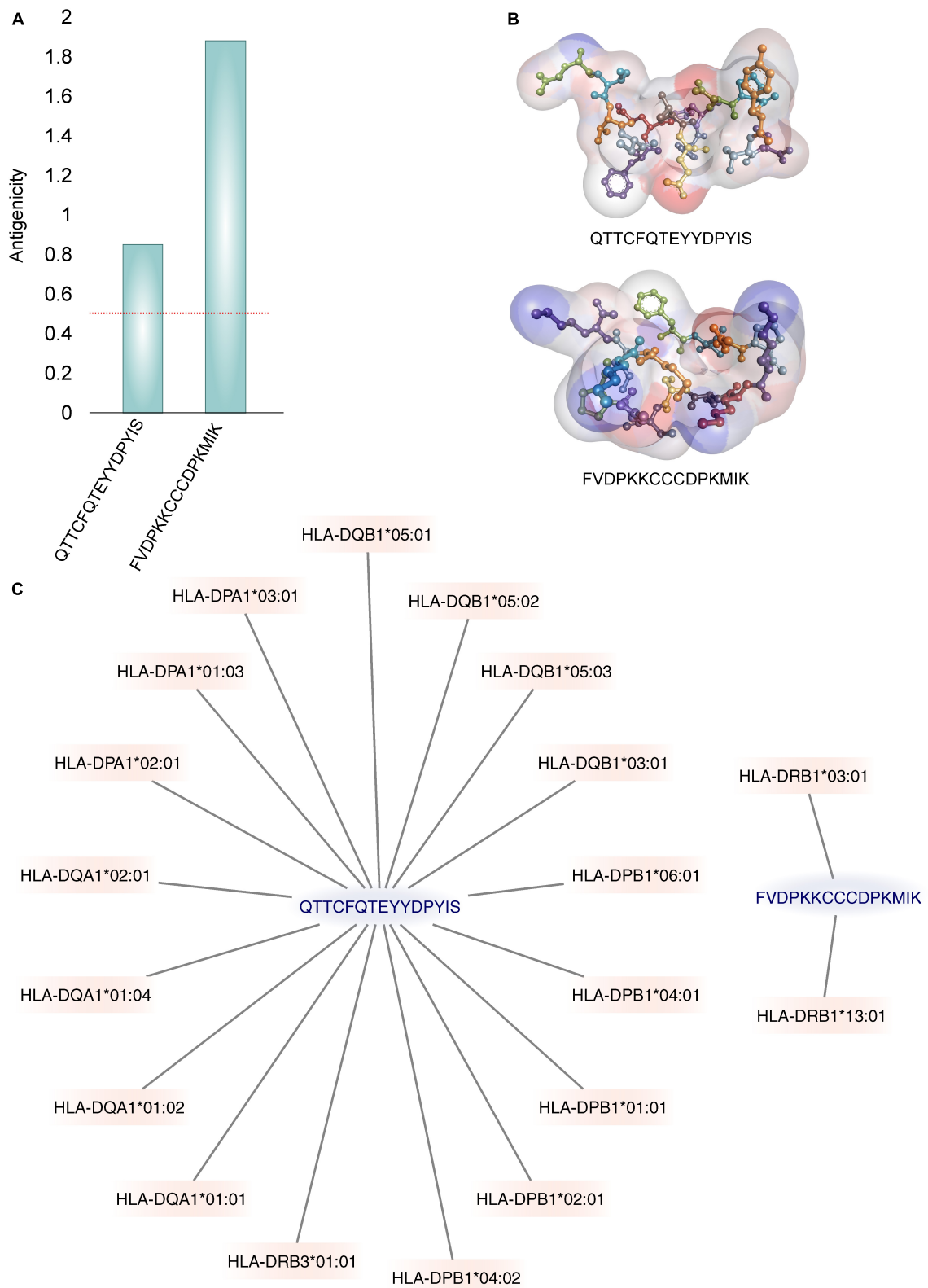
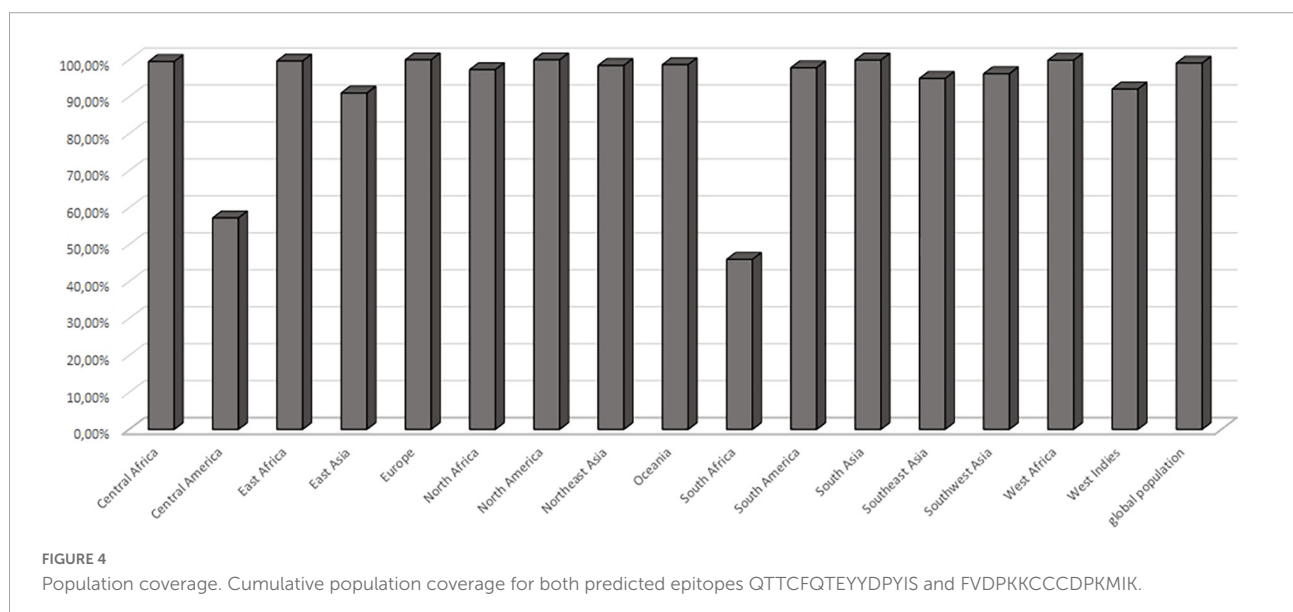


FIGURE 3
 The best two identified epitopes. **(A)** The Vaxijen server's calculated antigenic potential (96) is displayed. A horizontal red line indicates the antigenicity threshold. **(B)** Illustrations of the 3D structures of epitopes. The residues are shown under a soft surface and are colored by atom charges. **(C)** Graphs are used to show how binding with MHC-II alleles is displayed. Only the alleles with binding IC50 values ≤ 500 nM are displayed.

TABLE 2 List of selected alleles used for the population coverage analysis.

Epitope	Allele	IC50	Percentile rank	Allele locus	Prediction method
QTTCFQTEYYDPYIS	HLA-DRB3*01:01	205.6	6.2	HLA-DR	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*01:01/DQB1*05:01	17.2	0.02	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*01:02/DQB1*05:02	278.5	1.8	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*01:04/DQB1*05:03	290.9	0.31	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DQA1*02:01/DQB1*03:01	359.7	34	HLA-DQ	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*01:03/DPB1*06:01	17.40	2.2	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*01:03/DPB1*04:01	35.10	0.51	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*02:01/DPB1*01:01	198.60	2	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*01:03/DPB1*02:01	256.00	6.9	HLA-DP	NN-align 2.3
QTTCFQTEYYDPYIS	HLA-DPA1*03:01/DPB1*04:02	370.4	6.5	HLA-DP	NN-align 2.3
FVDPKCCCCDPKMIK	HLA-DRB1*13:01	368.1	50	HLA-DR	NN-align 2.3
FVDPKCCCCDPKMIK	HLA-DRB1*03:01	430.5	6.3	HLA-DR	NN-align 2.3



Discussion

The best strategy to control an infectious disease in a given population and one of the most efficient, quickest, and affordable ways to promote public health is through vaccination. Given the fact that *C. auris* can be resistant to nearly all available antifungal medications, the most effective way to fight this evolving and fatal pathogen is potentially by using a vaccine targeted against it. Notably, epitope-based vaccines are known to generate a stronger immune response against a pathogen as compared to the whole protein vaccines (62). Previously, vaccine development was solely dependent on experimental techniques which required extensive resources and time. However, with recent advancements in the fields of bioinformatics and reverse vaccinology, as pioneered by the work of Rino Rappouli, (63), we have efficient *in silico* approaches to screen the genome for best epitopes

which can thus be employed to design novel vaccines in a cost and time effective manner. These approaches have been already successfully employed by many researchers for identifying potent vaccine candidates (56, 64–67), which further endorses the potential of *in silico* approaches in vaccinology. Reverse vaccinology is now especially poised to produce more vaccines after the promising discovery of the first accepted multicomponent meningococcal serogroup B (MenB) vaccine (4CMenB; Bexsero®) which was proposed from computational approaches and later received huge success in laboratory experiments (63, 68).

Th1 cells (a subtype of CD4+ TH cells) are known to play a central role in providing immunity against fungi and effective fungal vaccines. They also induce generation of pro-inflammatory signature cytokines IFN- γ and TNF- α and play a role in production of opsonizing antibodies which promotes enhanced phagocytosis at the sites of infection (69). In addition,

emerging evidence states that Th17 cells (another subtype of CD4+ TH cells) have also been implicated in the generation of specialized immune response against fungal infection and usually a balance between the Th1 and Th17 associated responses is desired (69, 70). In memory pool of humans, anti-fungal Th17 memory cells have been found and they have also been known to play a role in induced vaccine protection in mice (71). Furthermore, Th17 cells act majorly at mucosal surfaces and thus, Th17 cells inducing vaccines are adequate to protect against deadly lung infection caused by *Blastomyces dermatitidis* in mice and also against three major systemic mycoses in North America (72). Altogether, it is highly suggested to target Th17 induction in the potential vaccine designs against systemic fungal infections (73).

We have previously developed and used different strategies to increase the ability of computer-aided vaccine designing approaches to identify vaccine candidates such as (1) epitope selection from conserved regions (74), (2) profile similarity method to analyze biased-ness of predicted epitopes toward profiles of experimentally validated epitopes (74), (3) introducing cleavage sites for targeted cleavage of multi-epitopic vaccine (56, 75), (4) B-cell epitope prediction by docking (74), (5) structure-based epitope identification (57), (6) CpG optimization (76), (7) sequence homology search against host (77), (8) adding adjuvant such as IL-12 (78), and (9) population coverage analysis (58, 77). In this study, we integrated the evolutionary analysis with our reverse vaccinology, i.e., bioinformatical genome and epitope search pipeline (74, 77) to identify vaccine candidates. The approach we used can be particularly helpful for designing vaccine against any emerging pathogens. Here we used this for epitope identification for potential vaccine against *C. auris*.

Currently, only the NDV-3A vaccine has shown potential so far in immunizing mice against *C. auris* infection (18). We extensively scanned the whole *C. auris* genome coded candidate proteins for their potential as vaccine candidates. We focused on secreted proteins and cell membrane proteins as they are particularly important for fungal host-pathogen interactions. During pathogenesis, these proteins play a role in the interaction with the host immune cells. Moreover, they are good candidates for targeting because of easier access and low expected-resistance mechanisms (79). Therefore, we followed the workflow in [Figure 2](#) [adapted from Vivek-Ananth et al. (79)] to filter *C. auris* proteins based on existence of a signal peptide, a GPI anchor or a TM domain. The filtered proteins were then classified in two separate groups of secreted and membrane proteins.

Five major clades of *C. auris* have been identified up to date with tens to hundreds of thousands of single nucleotide polymorphisms, which can be potentially linked to various infection strategies and outbreaks (80, 81). It had been shown that highly related isolates play important role in local and ongoing transmission, however, undergone clonal expansion

was detected in each clade (82). Moreover, as fungus has evolved as a pathogen under strong selection pressure imposed by climate change, it is important to recognize the rapidly evolved genes of *C. auris* (83). This knowledge is critical not only to understand the transmission dynamics, but also to design the suitable vaccine. In comparison of other genes, rapidly evolved genes have higher possibility to mutate to facilitate the pathogen survival in critical environment. We identified the positive selection signals and removed the proteins coded by such genes. Studies have shown that mutations in viral epitopes can promote hindrance in viral recognition by CD4+ and CD8+ T-cell receptors leading to escape of viruses from immune surveillance, thus escalating pathologic conditions (84). For development of a successful *C. auris* vaccine, it is crucial to have a memory pool of T-cells which can recognize *C. auris* and provide immunity despite its continuous evolution. To maintain high affinity, T-cell memory pool epitopes which are highly conserved and resistant to mutations should be considered. Thus, we focused on the conserved regions of MSAs after elimination of ortholog clusters with signs of positive selection.

Finally, we determined candidates from the conserved regions that are highly immunogenic, antigenic, and that show no significant similarity with host proteins. Furthermore, we also analyzed if the identified MHC class-II epitopes showed any similarity to other pathogenic organisms including fungi (taxid4751), bacteria (taxid2) and parasites from clade of *Protostomia* (taxid33317) that include parasitic worms *Nematoda* and *Platyhelminthes* and other parasites like *Trypanosomatida* (txid2704949) and *Plasmodium* (txid5820). We performed our Blast search with the parameters adjusted for short sequence length (see Host similarities and antigenicity of MHC class-II binders). Our search against bacteria and chosen parasite sequences showed no similar hits to both epitopes (e -value < 0.01). On the other hand, we have found seven fungal hits for epitope 1 with the e -value lower than 0.01. All hits showed 85.71% identity and 93.3% query coverage ([Supplementary Figure 1](#)). Six of seven hits are from *Hanseniaspora* species, a yeast genus while 1 is from *Brettanomyces*, a non-spore forming yeast genus. Thus, we also performed more specific Blast search for more possible hits in yeast first against the *Saccharomycetes* (txid4891) class and then directly with the *Candida* genus (txid5475, excluding *C. auris* from the search) of the same class. We found no hits with e -value < 0.01 in other *Candida* species for any of the epitopes. However, we found that 42 yeast proteins show similarity to epitope 1 (e -values < 0.01) and further investigated these hits. Top seven hits showed query coverage higher than 90% and identity score higher than 85%. When we compared these hits to the previous fungal proteins, we found a 100% overlap. The 17 of the remaining hits also had identity score over 78% and query coverage over 85% for epitope 1. Finally, the 18 remaining hits only showed identity scores higher than 70% and query

coverage higher than 60% and thus were not determined to be similar to epitope 1.

In our work, we used 5 *C. auris* strains spanning 4 clades namely *C. auris* 6,684 Clade I from India, *C. auris* B8441 Clade I from Pakistan, *C. auris* B11220 Clade II from Japan, *C. auris* B11221 Clade III from South Africa, and *C. auris* B11243 Clade IV from Venezuela. For further analysis to determine the clade coverage of our epitopes we also performed a protein sequence comparison against *Candida auris* (taxid:498019). This is especially important, since our analysis did not include the Iran strain IFRC2087, belonging to the recently identified clade V. We found 5 hits that showed 100% identity and coverage to epitope 1 from strains 6,684 (India, Clade I), and JCM 15448 (Japan, Clade II), B11221 (South Africa, Clade III), B11243 (Venezuela, Clade IV), and IFRC2087 (Iran, Clade V). For epitope 2, we also found 5 hits from all clades, respectively from strains 6,684 (India, Clade I), B11220 (Japan, Clade II), B11221 (South Africa, Clade III), B11243 (Venezuela, Clade IV), and IFRC2087 (Iran, Clade V) (**Supplementary Table 3**). Furthermore, since many isolates from different strains only have their genomes sequenced and no protein annotation yet, we have also performed a blast search in the nucleotide database with tblastn algorithm. Our analysis revealed 43 hits for each epitope (with 100% coverage and identity). Moreover, we found these hits to be from the same strains for both epitopes. Among these strains, 32 belong to clade I (28 from Lebanon: Beirut, 1 from China: Beijing, 1 from India, 1 from Italy: Genoa, and 1 from United Arab Emirates), 5 belong to clade II (1 from each Canada: Alberta, China: Shenyang, Japan, South Korea, and USA: New York), 4 belong to clade III (1 from each Canada: Quebec, South Africa, USA: California, USA: Indiana), 2 belong to clade IV (Colombia: Cartagena, and Venezuela) and 1 belongs to clade V from Iran (**Supplementary Figure 2**).

Taking all these results and comparison into account, we further decided to combine these two epitopes using the linker KFERQ (56, 85) so as to generate a more potent immune response from our vaccine construct. We thus performed *in silico* simulations for possible immune responses using the C-IMMSIM server (86, 87). Two injections were given with 4 weeks interval and immune response over a 6 months-long period was predicted. We have used the alleles that we determined to cover larger portion of the global population (namely HLA-DRB1*13:01, HLA-DRB1*03:01, and HLA-DRB3*01:01 since only these three were included in the database). Predicted response for all three combinations of the HLA alleles showed high innate response after the first injection and even higher IgM counts after the second injection. This demonstrates long-term immune response which is still retained upto 6 months (**Supplementary Figure 3**). We also observed an increase in the memory CD4+ T-cells' number to as high as 1,600 cells per mm³ after the second injection and still a population of 100 cells per mm³ even after 6 months (**Supplementary Figure 4**).

Since 2019, SARS-CoV-2 belongs to one of the most prevalent viral pathogens worldwide. It is well documented that COVID-19 patients often suffer from various infectious diseases, which also include fungal infections (88–91). This is due to the weakened immune system of the patients, as well as the treatment used to cure COVID-19 infection. A study from 2020 showed that 91.8% of COVID-19 patients were reported to suffer from secondary infections caused by bacterial agents, while 23.3% suffered from fungal co-infections (92). Coronavirus-associated *Candida* infection (CACa) is the third most prevalent fungal disease associated with SARS-CoV-2, with a high mortality rate of 67.849% (91). However, it is necessary to state that the majority of these patients also had other complications or diseases, such as diabetes mellitus, hypertension, or obesity (93). Therefore, due to lack of documentation, it is unclear whether *C. auris* infection alone was responsible for the higher mortality rate or other risk factors significantly contributed toward it. Moreover, a recent meta-analysis showed that men are more frequently affected by CACa as compared to women, with a 3.7 times greater risk of co-infection. Another important fact is that COVID-19 infections may significantly alter host immune response, as a result of which a fungal pathogen can resist its disposal with usual antifungal drugs (94). However, there may also be the possibility that it is a consequence of drug-drug interactions, which are administrated against the COVID-19 and co-infections.

Overall, this study presents a refined strategy to improve the current computer-aided vaccine design. The major limitation of our study is the current lack of the experimental validation, including functional analysis to reveal if our epitopes can really evoke the immune responses in the cell lines and in appropriate animal models. This process is though very challenging, however, we hope that with the right research collaboration, we can switch our work from theoretical study to the experimental work. Our results are very promising despite getting only two epitopes from our conservative approach. Furthermore, we achieved good immunogenicity, antigenicity, and high populations coverage in our predicted vaccine candidates with least possibility of epitopes to get evolved or mutated.

Conclusion

Vaccine design against invasive fungal pathogens has been challenging (95). In recent years, epitope-based vaccines have proven to provoke a potent and efficient defense response of the immune system in a targeted manner. They are synthesized at a lower cost, have good stability and relative safety, and have no limitations in target diseases. Thus, robust computational methods that have considerable predictive and analytical information combined with reverse vaccinology can encourage the prediction of novel epitopes that can serve as powerful vaccine candidates. Here, combining evolutionary

information and employing an immune-informatics approach, we identified two potent vaccine candidates for designing a *C. auris* vaccine. This strategy provides a new approach to identify highly conserved putative T-cell epitopes for other emerging pathogens. We screened the whole *C. auris* genome and rejected the genes showing adaptive evolution. Using the highly conserved regions from candidate proteins, we further enabled the identification of two MHC class-II epitopes QTTCFQTEYYDPYIS and FVDPKCCCDPKMIK, which besides showing high strain level conservancy, showed good immunogenicity, antigenicity, and no significant similarity with humans. In the follow up study, the results need to be experimentally confirmed by the peptide vaccine formulation in the laboratory followed by clinical trials.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary material](#).

Author contributions

SG and MS designed the project and had the main conceptual ideas and made the first manuscript draft. ÖO, RM, SB, and EB performed the data analysis. SG, ÖO, and MS wrote the manuscript with feedback and the manuscript edits from all authors. TD provided the critical feedback, supervision, and helped to shape the research. All authors contributed to the article and approved the submitted version.

References

- Casadevall A, Kontoyiannis DP, Robert V. On the emergence of *Candida auris*: climate change, azoles, swamps, and birds. *mBio*. (2019) 10:e1397–1319. doi: 10.1128/mBio.01397-19
- Lone SA, Ahmad A. *Candida auris*-the growing menace to global health. *Mycoses*. (2019) 62:620–37. doi: 10.1111/myc.12904
- Singhal T, Kumar A, Borade P, Shah S, Soman R. Successful treatment of *C. auris* shunt infection with intraventricular caspofungin. *Med Mycol Case Rep*. (2018) 22:35–7. doi: 10.1016/j.mmcr.2018.08.005
- Schelenz S, Hagen F, Rhodes JL, Abdolrasouli A, Chowdhary A, Hall A, et al. First hospital outbreak of the globally emerging *Candida auris* in a European hospital. *Antimicrob Resist Infect Control*. (2016) 5:35. doi: 10.1186/s13756-016-0132-5
- Azar MM, Turbett SE, Fishman JA, Pierce VM. Donor-derived transmission of *Candida auris* during lung transplantation. *Clin Infect Dis*. (2017) 65:1040–2. doi: 10.1093/cid/cix460
- Roberts SC, Zembower TR, Bolon MK, Kadakia AR, Gilley JH, Ko JH, et al. Successful treatment of a *Candida auris* intra-articular infection. *Emerg Microbes Infect*. (2019) 8:866–8. doi: 10.1080/22221751.2019.1625287
- Sherry L, Ramage G, Kean R, Borman A, Johnson EM, Richardson MD, et al. Biofilm-forming capability of highly virulent, multidrug-resistant *Candida auris*. *Emerg Infect Dis*. (2017) 23:328–31. doi: 10.3201/eid2302.161320
- Chakrabarti A, Sood P, Rudramurthy SM, Chen S, Kaur H, Capoor M, et al. Incidence, characteristics and outcome of ICU-acquired candidemia in India. *Intensive Care Med*. (2015) 41:285–95. doi: 10.1007/s00134-014-3603-2
- Calvo B, Melo AS, Perozo-Mena A, Hernandez M, Francisco EC, Hagen F, et al. First report of *Candida auris* in America: clinical and microbiological aspects of 18 episodes of candidemia. *J Infect*. (2016) 73:369–74. doi: 10.1016/j.jinf.2016.07.008
- Welsh RM, Bentz ML, Shams A, Houston H, Lyons A, Rose LJ, et al. Survival, persistence, and isolation of the emerging multidrug-resistant pathogenic yeast *Candida auris* on a plastic healthcare surface. *J Clin Microbiol*. (2017) 55:2996–3005. doi: 10.1128/JCM.00921-17
- Cassone A. Fungal vaccines: real progress from real challenges. *Lancet Infect Dis*. (2008) 8:114–24. doi: 10.1016/S1473-3099(08)70016-1
- Tso GHW, Reales-Calderon JA, Pavelka N. The elusive anti-*Candida* vaccine: lessons from the past and opportunities for the future. *Front Immunol*. (2018) 9:897. doi: 10.3389/fimmu.2018.00897
- Liu Y, Filler SG. *Candida albicans* Als3, a multifunctional adhesin and invasins. *Eukaryot Cell*. (2011) 10:168–73. doi: 10.1128/EC.00279-10

Funding

We gratefully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) project number 210879364 and the CRC/Transregio124 FungiNet (Project B1 – SG and TD). MS would like to thank for financial support by the Gleichstellungsbüro, University of Würzburg, Germany.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.1008527/full#supplementary-material>

14. Spellberg BJ, Ibrahim AS, Avanesian V, Fu Y, Myers C, Phan QT, et al. Efficacy of the anti-*Candida* rAls3p-N or rAls1p-N vaccines against disseminated and mucosal candidiasis. *J Infect Dis.* (2006) 194:256–60. doi: 10.1086/504691
15. Brena S, Omaetxebarria MJ, Elguetzabal N, Cabezas J, Moragues MD, Ponton J. Fungicidal monoclonal antibody C7 binds to *Candida albicans* Als3. *Infect Immun.* (2007) 75:3680–2. doi: 10.1128/IAI.01840-06
16. Schmidt CS, White CJ, Ibrahim AS, Filler SG, Fu Y, Yeaman MR, et al. NDV-3, a recombinant alum-adjuvanted vaccine for *Candida* and *Staphylococcus aureus*, is safe and immunogenic in healthy adults. *Vaccine.* (2012) 30:7594–600. doi: 10.1016/j.vaccine.2012.10.038
17. Edwards JE Jr., Schwartz MM, Schmidt CS, Sobel JD, Nyirjesy P, Schodel F, et al. A fungal immunotherapeutic vaccine (NDV-3A) for treatment of recurrent vulvovaginal candidiasis—A phase 2 randomized, double-blind, placebo-controlled trial. *Clin Infect Dis.* (2018) 66:1928–36. doi: 10.1093/cid/ciy185
18. Singh S, Uppuluri P, Mamouei Z, Alqarihi A, Elhassan H, French S, et al. The NDV-3A vaccine protects mice from multidrug resistant *Candida auris* infection. *PLoS Pathog.* (2019) 15:e1007460. doi: 10.1371/journal.ppat.1007460
19. Spellberg B, Ibrahim AS, Lin L, Avanesian V, Fu Y, Lipke P, et al. Antibody titer threshold predicts anti-candidal vaccine efficacy even though the mechanism of protection is induction of cell-mediated immunity. *J Infect Dis.* (2008) 197:967–71. doi: 10.1086/529204
20. Spellberg B, Ibrahim AS, Yeaman MR, Lin L, Fu Y, Avanesian V, et al. The antifungal vaccine derived from the recombinant N terminus of Als3p protects mice against the bacterium *Staphylococcus aureus*. *Infect Immun.* (2008) 76:4574–80. doi: 10.1128/IAI.00700-08
21. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, et al. Mapping the antigenic and genetic evolution of influenza virus. *Science.* (2004) 305:371–6. doi: 10.1126/science.1097211
22. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* (2007) 35:D21–5. doi: 10.1093/nar/gkl986
23. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* (2015) 31:3210–2. doi: 10.1093/bioinformatics/btv351
24. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, et al. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* (2015) 43:D250–6. doi: 10.1093/nar/gku1220
25. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* (2017) 45:D744–9. doi: 10.1093/nar/gkw1119
26. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* (2015) 16:157. doi: 10.1186/s13059-015-0721-2
27. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* (2015) 12:59–60. doi: 10.1038/nmeth.3176
28. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* (2000) 300:1005–16. doi: 10.1006/jmbi.2000.3903
29. Armenteros JJA, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* (2019) 37:420–3. doi: 10.1038/s41587-019-0036-z
30. Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol.* (2004) 338:1027–36. doi: 10.1016/j.jmb.2004.03.016
31. Meinken J, Asch DK, Neizer-Ashun KA, Chang G-H, Cooper CR JR., Min XJ. FunSecKB2: a fungal protein subcellular location knowledgebase. *Comput Mol Biol.* (2014) 4:1–17.
32. Pierleoni A, Martelli PL, Casadio R. PredGPI: a GPI-anchor predictor. *BMC Bioinformatics.* (2008) 9:392. doi: 10.1186/1471-2105-9-392
33. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* (2001) 305:567–80. doi: 10.1006/jmbi.2000.4315
34. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* (2006) 34:W362–5. doi: 10.1093/nar/gkl124
35. Almagro Armenteros JJ, Sonderby CK, Sonderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics.* (2017) 33:3387–95. doi: 10.1093/bioinformatics/btx431
36. Nami S, Mohammadi R, Vakili M, Khezripour K, Mirzaei H, Morovati H. Fungal vaccines, mechanism of actions and immunology: a comprehensive review. *Biomed Pharmacother.* (2019) 109:333–44. doi: 10.1016/j.biopha.2018.10.075
37. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol.* (2018) 19:2094–110. doi: 10.1111/mpp.12682
38. Chaudhuri R, Ansari FA, Raghunandan MV, Ramachandran S. FungalRV: adhesin prediction and immunoinformatics portal for human fungal pathogens. *BMC Genomics.* (2011) 12:192. doi: 10.1186/1471-2164-12-192
39. Ramana J, Gupta D. FaaPred: a SVM-based prediction method for fungal adhesins and adhesin-like proteins. *PLoS One.* (2010) 5:e9695. doi: 10.1371/journal.pone.0009695
40. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* (2000) 302:205–17. doi: 10.1006/jmbi.2000.4042
41. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* (2006) 34:W609–12. doi: 10.1093/nar/gkl315
42. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* (2007) 56:564–77. doi: 10.1080/10635150701472164
43. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* (2007) 24:1586–91. doi: 10.1093/molbev/msm088
44. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* (2005) 22:1107–18. doi: 10.1093/molbev/msi097
45. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* (2005) 22:1208–22. doi: 10.1093/molbev/msi105
46. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* (2012) 8:e1002764. doi: 10.1371/journal.pgen.1002764
47. Oliveira L, Paiva PB, Paiva AC, Vriend G. Identification of functionally conserved residues with the use of entropy-variability plots. *Proteins.* (2003) 52:544–52. doi: 10.1002/prot.10490
48. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* (1948) 27:379–423.
49. Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol.* (1995) 13:587–622. doi: 10.1146/annurev.iy.13.040195.003103
50. Sinigaglia F, Hammer J. Motifs and supermotifs for MHC class II binding peptides. *J Exp Med.* (1995) 181:449–51. doi: 10.1084/jem.181.2.449
51. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* (2019) 47:D339–43. doi: 10.1093/nar/gky1006
52. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* (2009) 10:421. doi: 10.1186/1471-2105-10-421
53. Doytchinova IA, Flower DR. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics.* (2007) 8:4. doi: 10.1186/1471-2105-8-4
54. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* (2011) 7:e1002195. doi: 10.1371/journal.pcbi.1002195
55. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc.* (2021) 1:e323. doi: 10.1002/cpz1.323
56. Gupta SK, Singh A, Srivastava M, Gupta SK, Akhooon BA. In silico DNA vaccine designing against human papillomavirus (HPV) causing cervical cancer. *Vaccine.* (2009) 28:120–31. doi: 10.1016/j.vaccine.2009.09.095
57. Gupta SK, Srivastava M, Akhooon BA, Gupta SK, Grabe N. In silico accelerated identification of structurally conserved CD8+ and CD4+ T-cell epitopes in high-risk HPV types. *Infect Genet Evol.* (2012) 12:1513–8. doi: 10.1016/j.meegid.2012.02.022
58. Singh KP, Verma N, Akhooon BA, Bhatt V, Gupta SK, Gupta SK, et al. Sequence-based approach for rapid identification of cross-clade CD8+ T-cell vaccine candidates from all high-risk HPV strains. *3 Biotech.* (2016) 6:39. doi: 10.1007/s13205-015-0352-z
59. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* (2017) 34:1812–9. doi: 10.1093/molbev/msx116

60. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool.* (2005) 2:16. doi: 10.1186/1742-9994-2-16
61. Govender NP, Magobo RE, Mpenbe R, Mhlanga M, Matlapeng P, Corcoran C, et al. *Candida auris* in South Africa, 2012–2016. *Emerg Infect Dis.* (2018) 24:2036.
62. De Groot AS, McMurry J, Marcon L, Franco J, Rivera D, Kutzler M, et al. Developing an epitope-driven tuberculosis (TB) vaccine. *Vaccine.* (2005) 23:2121–31. doi: 10.1016/j.vaccine.2005.01.059
63. Gorringe AR, Pajon R. Bexsero: a multicomponent vaccine for prevention of meningococcal disease. *Hum Vaccin Immunother.* (2012) 8:174–83. doi: 10.4161/hv.18500
64. Chaitra MG, Hariharaputran S, Chandra NR, Shaila MS, Nayak R. Defining putative T-cell epitopes from PE and PPE families of proteins of *Mycobacterium tuberculosis* with vaccine potential. *Vaccine.* (2005) 23:1265–72. doi: 10.1016/j.vaccine.2004.08.046
65. Bencurova E, Gupta SK, Oskoueian E, Bhide M, Dandekar T. Omics and bioinformatics applied to vaccine development against *Borrelia*. *Mol Omics.* (2018) 14:330–40. doi: 10.1039/c8mo00130h
66. Jain R, Jain A, Verma SK. Prediction of epitope based peptides for vaccine development from complete proteome of novel corona virus (SARS-COV-2) using immunoinformatics. *Int J Pept Res Ther.* (2021) 27:1729–40. doi: 10.1007/s10989-021-10205-z
67. Sethi G, Sethi S, Krishna R. Multi-epitope based vaccine design against *Staphylococcus epidermidis*: a subtractive proteomics and immunoinformatics approach. *Microb Pathog.* (2022) 165:105484. doi: 10.1016/j.micpath.2022.105484
68. Martin NG, Snape MD. A multicomponent serogroup B meningococcal vaccine is licensed for use in Europe: what do we know, and what are we yet to learn? *Expert Rev Vaccin.* (2013) 12:837–58. doi: 10.1586/14760584.2013.814862
69. Lin L, Ibrahim AS, Xu X, Farber JM, Avanesian V, Baquir B, et al. Th1-Th17 cells mediate protective adaptive immunity against *Staphylococcus aureus* and *Candida albicans* infection in mice. *PLoS Pathog.* (2009) 5:e1000703. doi: 10.1371/journal.ppat.1000703
70. Bartemes KR, Kita H. Innate and adaptive immune responses to fungi in the airway. *J Allergy Clin Immunol.* (2018) 142:353–63. doi: 10.1016/j.jaci.2018.06.015
71. Romani L. Immunity to fungal infections. *Nat Rev Immunol.* (2011) 11:275–88. doi: 10.1038/nri2939
72. Wuthrich M, Gern B, Hung CY, Ersland K, Rocco N, Pick-Jacobs J, et al. Vaccine-induced protection against 3 systemic mycoses endemic to North America requires Th17 cells in mice. *J Clin Invest.* (2016) 126:795. doi: 10.1172/JCI85788
73. Wuthrich M, Gern B, Hung CY, Ersland K, Rocco N, Pick-Jacobs J, et al. Vaccine-induced protection against 3 systemic mycoses endemic to North America requires Th17 cells in mice. *J Clin Invest.* (2011) 121:554–68. doi: 10.1172/JCI43984
74. Gupta SK, Srivastava M, Akhooon BA, Smita S, Schmitz U, Wolkenhauer O, et al. Identification of immunogenic consensus T-cell epitopes in globally distributed influenza-A H1N1 neuraminidase. *Infect Genet Evol.* (2011) 11:308–19. doi: 10.1016/j.meegid.2010.10.013
75. Ranjbar MM, Gupta SK, Ghorban K, Nabian S, Sazmand A, Taheri M, et al. Designing and modeling of complex DNA vaccine based on tropomyosin protein of *Boophilus* genus tick. *Appl Biochem Biotechnol.* (2015) 175:323–39. doi: 10.1007/s12010-014-1245-z
76. Baloria U, Akhooon BA, Gupta SK, Sharma S, Verma V. In silico proteomic characterization of human epidermal growth factor receptor 2 (HER-2) for the mapping of high affinity antigenic determinants against breast cancer. *Amino Acids.* (2012) 42:1349–60. doi: 10.1007/s00726-010-0830-x
77. Gupta SK, Smita S, Sarangi AN, Srivastava M, Akhooon BA, Rahman Q, et al. In silico CD4+ T-cell epitope prediction and HLA distribution analysis for the potential proteins of *Neisseria meningitidis* Serogroup B—a clue for vaccine development. *Vaccine.* (2010) 28:7092–7. doi: 10.1016/j.vaccine.2010.08.005
78. Akhooon BA, Slathia PS, Sharma P, Gupta SK, Verma V. In silico identification of novel protective VSG antigens expressed by *Trypanosoma brucei* and an effort for designing a highly immunogenic DNA vaccine using IL-12 as adjuvant. *Microb Pathog.* (2011) 51:77–87. doi: 10.1016/j.micpath.2011.01.011
79. Vivek-Ananth RP, Mohanraj K, Vandanashree M, Jhingran A, Craig JP, Samal A. Comparative systems analysis of the secretome of the opportunistic pathogen *Aspergillus fumigatus* and other *Aspergillus* species. *Sci Rep.* (2018) 8:6617. doi: 10.1038/s41598-018-25016-4
80. Muñoz JF, Gade L, Chow NA, Loparev VN, Juieng P, Berkow EL, et al. Genomic insights into multidrug-resistance, mating and virulence in *Candida auris* and related emerging species. *Nat Commun.* (2018) 9:5346. doi: 10.1038/s41467-018-07779-6
81. Chow NA, de Groot T, Badali H, Abastabar M, Chiller TM, Meis JF. Potential fifth clade of *Candida auris*, Iran, 2018. *Emerg Infect Dis.* (2019) 25:1780–1. doi: 10.3201/eid2509.190686
82. Chow NA, Muñoz JF, Gade L, Berkow EL, Li X, Welsh RM, et al. Tracing the evolutionary history and global expansion of *Candida auris* using population genomic analyses. *mBio.* (2020) 11:e3364–3319. doi: 10.1128/mBio.03364-19
83. Jackson BR, Chow N, Forsberg K, Litvintseva AP, Lockhart SR, Welsh R, et al. On the origins of a species: what might explain the rise of *Candida auris*? *J Fungi.* (2019) 5:58. doi: 10.3390/jof5030058
84. Ciurea A, Hunziker L, Martinić MMA, Oxenius A, Hengartner H, Zinkernagel RM. CD4(+) T-cell-epitope escape mutant virus selected *in vivo*. *Nat Med.* (2001) 7:795–800.
85. Chiang HL, Terlecky SR, Plant CP, Dice JF. A role for a 70-kilodalton heat shock protein in lysosomal degradation of intracellular proteins. *Science.* (1989) 246:382–5. doi: 10.1126/science.2799391
86. Castiglione F, Deb D, Srivastava AP, Liò P, Liso A. From infection to immunity: understanding the response to SARS-CoV2 through in-silico modeling. *Front Immunol.* (2021) 12:646972. doi: 10.3389/fimmu.2021.646972
87. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One.* (2010) 5:e9862. doi: 10.1371/journal.pone.0009862
88. Ezeokoli OT, Gcilitshana O, Pohl CH. Risk factors for fungal co-infections in critically ill COVID-19 patients, with a focus on immunosuppressants. *J Fungi.* (2021) 7:545. doi: 10.3390/jof7070545
89. Hoenigl M, Seidel D, Sprute R, Cunha C, Oliverio M, Goldman GH, et al. COVID-19-associated fungal infections. *Nat Microbiol.* (2022) 7:1127–40. doi: 10.1038/s41564-022-01172-2
90. Song G, Liang G, Liu W. Fungal co-infections associated with global COVID-19 pandemic: a clinical and diagnostic perspective from China. *Mycopathologia.* (2020) 185:599–606. doi: 10.1007/s11046-020-00462-9
91. Vaseghi N, Sharifisooraki J, Khodadadi H, Nami S, Safari F, Ahangarkani F, et al. Global prevalence and subgroup analyses of Coronavirus disease (COVID-19) associated *Candida auris* infections (CACA): a systematic review and meta-analysis. *Mycoses.* (2022) 65:683–703. doi: 10.1111/myc.13471
92. Zhu X, Ge Y, Wu T, Zhao K, Chen Y, Wu B, et al. Co-infection with respiratory pathogens among COVID-2019 cases. *Virus Res.* (2020) 285:198005. doi: 10.1016/j.virusres.2020.198005
93. Vinayagamorthy K, Pentapati KC, Prakash H. Prevalence, risk factors, treatment and outcome of multidrug resistance *Candida auris* infections in Coronavirus disease (COVID-19) patients: a systematic review. *Mycoses.* (2022) 65:613–24. doi: 10.1111/myc.13447
94. Arastehfar A, Carvalho A, van de Veerdonk FL, Jenks JD, Koehler P, Krause R, et al. COVID-19 associated pulmonary aspergillosis (CAPA)—from immunology to treatment. *J Fungi.* (2020) 6:91. doi: 10.3390/jof6020091
95. Medici NP, Del Poeta M. New insights on the development of fungal vaccines: from immunity to recent challenges. *Memor Inst Oswaldo Cruz.* (2015) 110:966–73.
96. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* (2007) 8:4. doi: 10.1186/1471-2105-8-4
97. Thummuluri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* (2022) 50:W228–34. doi: 10.1093/nar/gkac278
98. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance.* (2019) 2:e201900429. doi: 10.26508/lsa.201900429
99. Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* (2007) 35:W429–32. doi: 10.1093/nar/gkm256
100. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol.* (2022) 40:1023–5. doi: 10.1038/s41587-021-01156-3
101. Lum G, Min XJ. FunSecKB: the fungal secretome knowledgebase. *Database (Oxford).* (2011) 2011:bar001. doi: 10.1093/database/bar001