ACS OMEGA

Open Access



http://pubs.acs.org/journal/acsodf

Article

# Excavation of Biomarker Candidates for the Diagnosis of *Talaromyces marneffei* Infection via Genome-Wide Prediction and Functional Annotation of Secreted Proteins

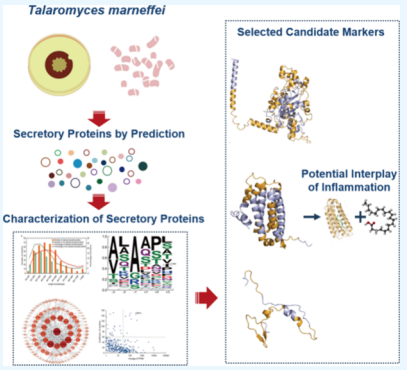
Jing Dan,[#] Wudi Wei,[#] Weijie Ou, Guangshi Gao, Wanjun Song, Li Ye, Hao Liang, Xuzhen Guo,* Lei Tan,* and Junjun Jiang*

Cite This: *ACS Omega* 2024, 9, 27093−27103

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** *Talaromyces marneffei* is the third most common infectious pathogen in AIDS patients and leads to the highest death rate in Guangxi, China. The lack of reliable biomarkers is one of the major obstacles in current clinical diagnosis, which largely contributes to this high mortality. Here, we present a study that aimed at identifying diagnostic biomarker candidates through genome-wide prediction and functional annotation of *Talaromyces marneffei* secreted proteins. A total of 584 secreted proteins then emerged, including 382 classical and 202 nonclassical ones. Among them, there were 87 newly obtained functional annotations in this study. The annotated proteins were further evaluated by combining RNA profiling and a homology comparison. Three proteins were ultimately highlighted as biomarker candidates with robust expression and remarkable specificity. The predicted phosphoinositide phospholipase C and the galactomannoprotein were suggested to play an interactive immune game through metabolism of arachidonic acid. Therefore, they hold promise in developing new tools for clinical diagnosis of *Talaromyces marneffei* and also possibly serve as molecular targets for future therapy.




## INTRODUCTION

Opportunistic infection associated with HIV is one of the leading causes of death in AIDS populations due to the destruction and suppression of $CD4^+$ T cells. The most frequently detected pathogens of HIV opportunistic infections include *Talaromyces marneffei* (*T. marneffei*), *Cryptococcus*, *Pneumocystis*, *Histoplasma*, *Mycobacterium tuberculosis* (*M. tuberculosis*)[1−5] et al. *T. marneffei* is ranked top three among these pathogens in Thailand, Vietnam, and Southern China, with infection rate ranging from 0.13 to 26.76%,[6] and ∼50,000 AIDS patients get infected globally per year.[3] Further, in Guangxi Province, China, the mortality rate of *T. marneffei* infection ranges from 8 to 40%,[6] which is the highest among all pathogens listed above.

To our knowledge, inhalation of pathogenic conidia is the key step of its infection route,[7] which then causes a systemic fatal infection in immune-compromised patients.[7,8] The typical features of *T. marneffei* infection include fever, unintentional weight loss, hepatosplenomegaly, lymphadenopathy, and papules with central necrosis in 60−70% infected AIDS patients.[9,10] The mortality rate was estimated up to 91% among individuals without antifungal therapy. However, symptomatic treatments in time significantly cut it down to 30%.[11] Therefore, an effective early diagnosis is severely demanded for rescuing infected ones.

Unfortunately, current *T. marneffei* diagnoses are still ineffective, including mycological culture, immunodetection, and nucleic acid test. Mycological culture, the golden standard with great accuracy, generally takes 2−4 weeks and shows low sensitivity (76% for HIV-positive and 47.1% for HIV-negative) and pitifully leads to the loss of therapeutic time windows.[12] The sensitivity of nucleic acid test is merely 70.4 and 52.2% pre- and postantifungal treatment, respectively.[13] In terms of immune detection, the outcomes are largely determined by the target molecules. For instance, $\beta$-D glucans and galactomannoprotein can only present fungal infections but show little distinction of *T. marneffei* from other species.[14−16] Mp1p, the only biomarker commercially available, luckily provides a sensitivity up to 86.3%, a specificity nearly 99%, and a test period of 2−3 h.[17,18] In one word, application of Mp1p notably enhanced the *T. marneffei* clinical diagnosis, while the accuracy is still constrained in samples of early infectious stage. For this challenge, excavation of new biomarkers is emergently awaiting to reinforce immunodetections.

Secreted proteins have been regarded as a treasurable pool for biomarker discovery because of their critical roles in

Received: January 17, 2024
Revised: May 28, 2024
Accepted: June 5, 2024
Published: June 13, 2024

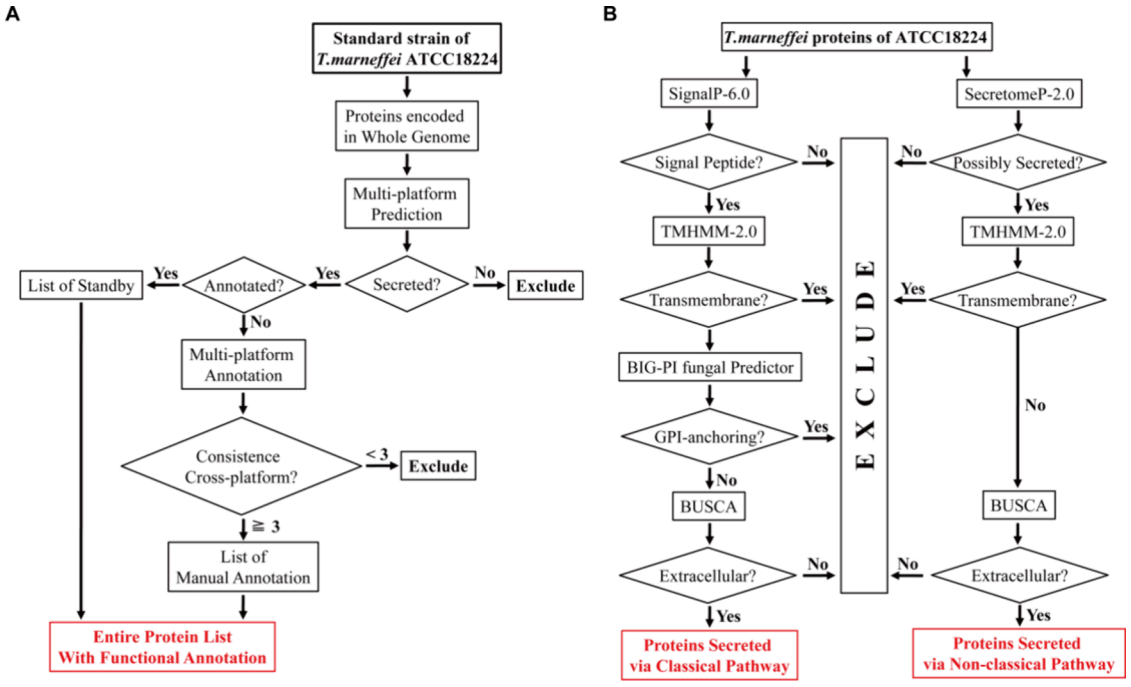



**Figure 1.** Prediction and functional annotation of secreted proteins. (A) Entire procedure of genome-wide prediction and annotation. (B) Prediction procedure for proteins secreted via both classical and nonclassical pathways.

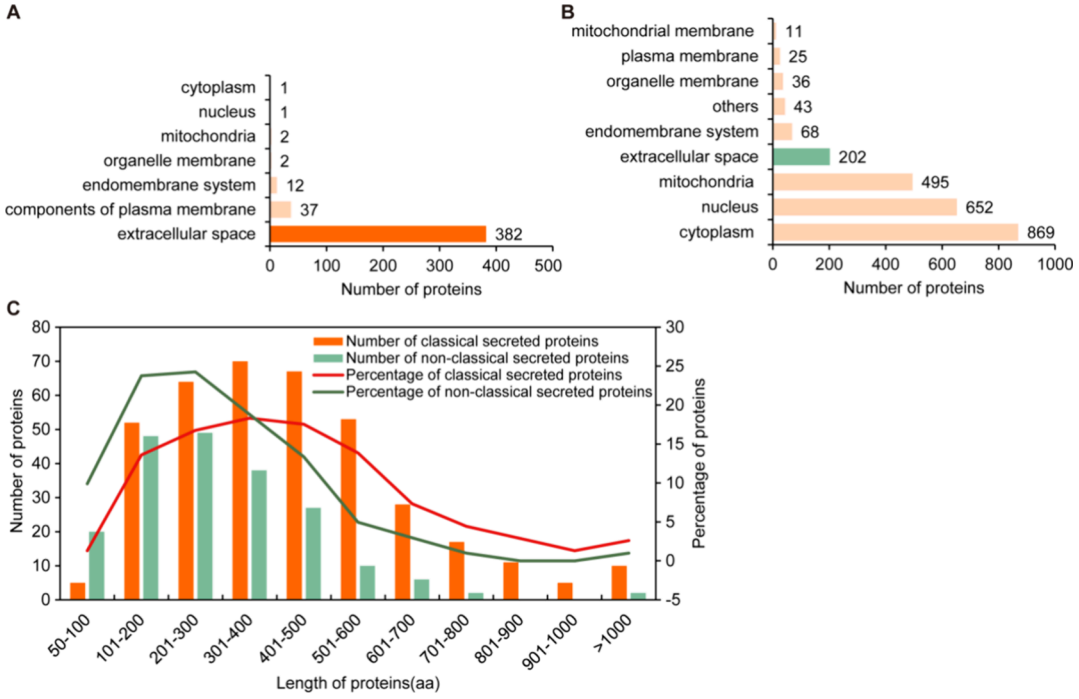



**Figure 2.** Analysis of *T. marneffei* secreted proteins. (A, B) Final highlights of secreted proteins with BUSCA. (A) 382 proteins secreted via the classical pathway were predicted out of 437 candidate proteins absent of GPI anchoring site and (B) 202 ones secreted via the nonclassical pathway were predicted out of 2401 candidates without transmembrane domain. (C) Polypeptide length distribution of secreted proteins; histogram represents the protein number, and the solid line represents percentage in each group.

multiple biological and pathological processes.[19−21] The extracellular secretome of *T. marneffei* is still mysterious since the function and localization remain unclear for >70% proteins encoded by standard genome ATCC18224 (10,707 proteins in total).[22] With growing amounts of whole-genome sequencing data, in-depth functional investigations of proteins have been conducted in various microorganisms,[23−25] which brought up massive proteins waiting for annotation, which were termed as hypothetical or uncharacterized proteins (HPs or UPs).[26] Encouragingly, a panel of bioinformatic platforms in rapid development thus provided the capability of discovery, prediction, and annotation for protein function and localization.


https://doi.org/10.1021/acsomega.4c00571
ACS Omega 2024, 9, 27093−27103

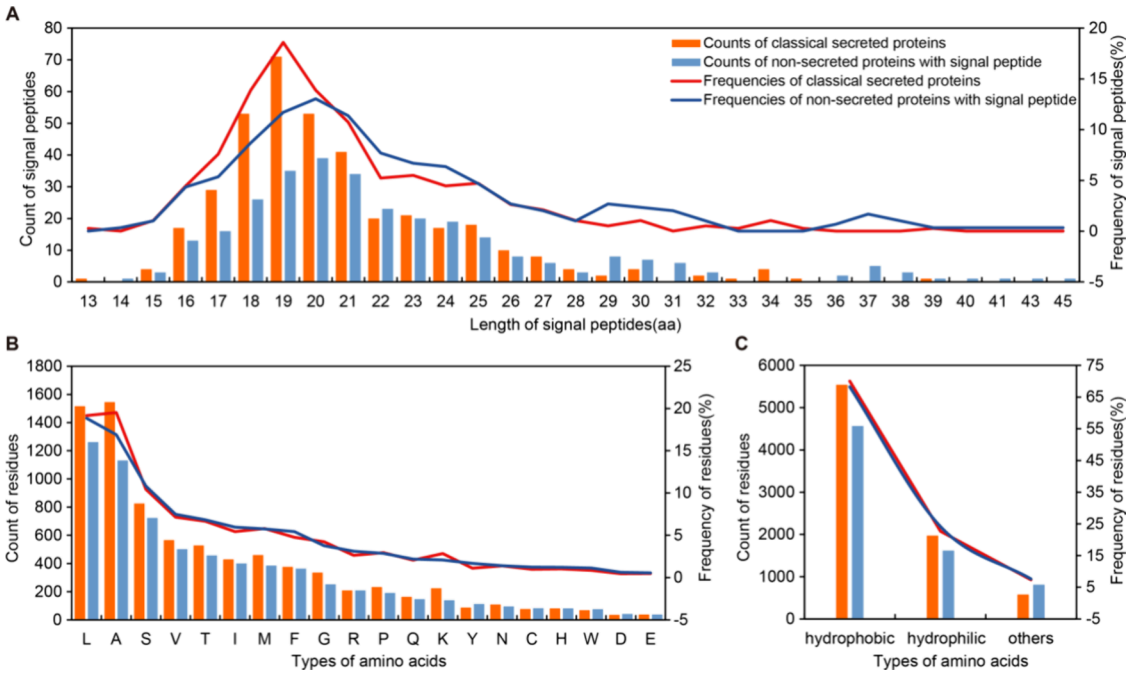



**Figure 3.** Analysis of signal peptide length and amino acid composition. (A) Length range of signal peptide in classical secreted and nonsecreted proteins. (B) Composition of 20 amino acids in signal peptides. (C) Composition of three classes of amino acids in signal peptides (hydrophobic, hydrophilic, and other). Hydrophobic amino acids include A, F, G, I, L, M, P, V, and W; hydrophilic amino acids include C, N, Q, S, T, and Y; and other amino acids include K, R, H, E, and D. For the entire figure, frequency and corresponding count were presented with the solid line and histogram, respectively. The color used for classical secreted proteins is orange and nonsecreted ones with blue.

Herein, we carried out a step-by-step prediction on *T. marneffei* proteins secreted via classical and nonclassical pathways, among which HPs and UPs were annotated. A comprehensive analysis was then conducted by the combination of RNA sequencing and homology comparison. Potential biomarkers arose with robust abundance and clarified function. Further evaluation was performed on their specificity, structure, and antigenicity. With the investigations above, three proteins were eventually highlighted as promising candidates and provide opportunities for the innovation of *T. marneffei* immune diagnosis.

## RESULTS

Secreted proteins share structural features that are devoid of transmembrane domains and intracellular localization signals.[25] Proteins are secreted via two pathways, classical and nonclassical, depending on the presence or absence of signal peptides. In the classical pathway, proteins are recognized with signal peptides, transported from endoplasmic reticulum to plasma membrane, and extracellularly released due to the absence of GPI anchoring sites. The proteins in a nonclassical way are secreted without signal peptides but through alternative strategies, including Golgi bypass, vesicles, transporters, etc.[27,28]

Starting from the entire set of *T. marneffei* proteins, genome-wide prediction and functional annotation were carried out sequentially on the standard genome ATCC18224 (Figure 1A). Due to the presence/absence of signal peptides, the procedure of prediction was then organized accordingly for proteins secreted via classical and nonclassical pathways (Figure 1B).

**Prediction of Classical and Nonclassical Secreted Proteins.** 681 proteins with N-terminal signal peptides arose from 10,707 input proteins with SignalP-6.0, representing 6.36% of the entire set of *T. marneffei* proteins. Subsequently, 186 proteins were excluded since they contained at least one transmembrane domain, potentially serving as ion channels, membrane receptors, or membrane-anchoring proteins. From the remaining 495 proteins, 437 ones lacking GPI anchoring sites were reserved. Finally, 382 candidates were identified as extracellular secretion type via the classical pathway (Figure 2A), which are all listed in Table S1. On the contrary, 3040 out of the total 10,707 proteins were revealed as potential nonclassical secreted proteins, comprising 28.39% of the entire *T. marneffei* proteome. Afterward, 2401 proteins lacking transmembrane domains were identified. In the next step, 2199 proteins were excluded due to their intracellular presence primarily (Figure 2B). Consequently, 202 proteins were highlighted at the last step as secreted via the nonclassical pathways (listed in Table S2). In total, 584 secreted proteins were predicted in the *T. marneffei* genome, corresponding to 5.45% of the whole standard genome.

Polypeptide length was subsequently analyzed for both classical and nonclassical secreted proteins (Figure 2C). For classical ones, the polypeptide length ranges from 57 aa to 1618 aa, with 435.16 aa as the average and 401 aa as the median. Over 80% proteins in this group fall within 100−600 aa. In the nonclassical group, the length range is between 53 aa and 1115 aa, with an average of 291.19 aa and a median of 258 aa. The range from 100 aa to 600 aa also occupy >80% nonclassical ones. It implied smaller proteins are more likely to be secreted via nonclassical pathways.

**Characteristics of Signal Peptides Contained in Classical Secreted Proteins.** Signal peptides are vital for transportation of proteins localized either extra- or intracellular.[29] Some differences have been revealed on the signal peptides between classical secreted and nonsecreted proteins. In secreted proteins, the length varied from 13 to 39 aa, with


https://doi.org/10.1021/acsomega.4c00571
ACS Omega 2024, 9, 27093−27103

an average of 20.72 amino acids and a median of 20 amino acids (Figure 3A). Meanwhile, the length in nonsecreted proteins ranged from 14 aa to 45 aa, with 22.39 aa as average and 21 aa as median. The frequencies were then calculated for signal peptides at each given length through dividing their numbers by the number of signal peptide containing proteins in the same class. Among the 382 classical secreted ones, the most frequently used signal peptide length is 19 aa, accounting for 18.59% (71 proteins). However, 20 aa signal peptides are most frequently observed in nonsecreted proteins (13.04%, 39 proteins).

Similarly, slight differences about the usage of amino acids in signal peptides were also quantified between classical secreted and nonsecreted proteins through dividing the times of employing a certain amino acid in signal peptides by the total residues number of signal peptides in same class of protein. For both types of signal peptides, A, L, and S are hired as the top three residues but ranked differently. In secreted proteins, they are utilized as 19.52% A, 19.15% L, and 10.43% S, while 18.85% L, 16.89% A, and 10.81% S are used in nonsecreted proteins (Figure 3B). With a similar calculation, utilization of hydrophobic, hydrophilic, and other amino acids in signal peptides was showed as 70, 22.66, and 7.34% for classical secreted proteins, respectively. In nonsecreted proteins, the percentage is 68.19% hydrophobic, 24.18% hydrophilic, and 7.63% other amino acids (Figure 3C). For both types of proteins, hydrophobic amino acids were preferred to be incorporated in the signal peptide, especially alanine and leucine.

**Analysis of Signal Peptide Cleavage Motif in Classical Secreted Proteins.** The cleavage motif is typically defined as a 6-amino acid sequence flanking the peptide-bond breaking site, comprising 3 aa on the N-side and another 3 aa on the C-side (numbered as −3, −2, −1, + 1, + 2, + 3).[24] The most privileged cleavage motif counted among classically secreted proteins was A-L-A-A-P-L (Figure 4A). The cleavage motif (A-L-A-A-S-T) was the most favorable for nonsecreted proteins (Figure 4B). An obvious difference lies in the +2 and +3 sites.

For both groups of proteins (Figure 4A,B), the cleavage motif contains highly dominant sites, including the −1 site with A and the −3 site for A/V. Strong diversity can be visualized in the other four sites (−2, + 1, + 2, and +3) (Figure 4A,B). At the −2 site, L, A, S, Q and T are employed as top five residues, whereas a ∼ 4% difference can be seen for both S and Q between the two groups (Figure 4C). Residue E at the +2 site also shows a similar level of difference (Figure 4E). The more considerable divergence emerges on residues A and I at the +1 site and P at the +2 site (Figure 4D,E). Counts and residue usage frequencies at each site on cleavage motif are listed in Table S3.

As a brief summary, These cleavage motifs belong to A-X-A type, which also exists in other fungi[24,30] and is generally proceeded by a serine protease termed classical signal peptidase I.[31]

**Functional Annotation of HPs and UPs.** Out of 584 secreted proteins predicted in this study, functions of 143 HPs and 106 UPs remained unclear (listed in Table S4), while 335 ones contain records of function in NCBI. 87 proteins obtained functional information with high confidence, which were agreed across no less than 3 independent resources (listed in Table S5). 87 newly annotated proteins showed >90% homology in comparison, with which the accuracy of annotations was ensured (listed in Table S6). Hydrolase was

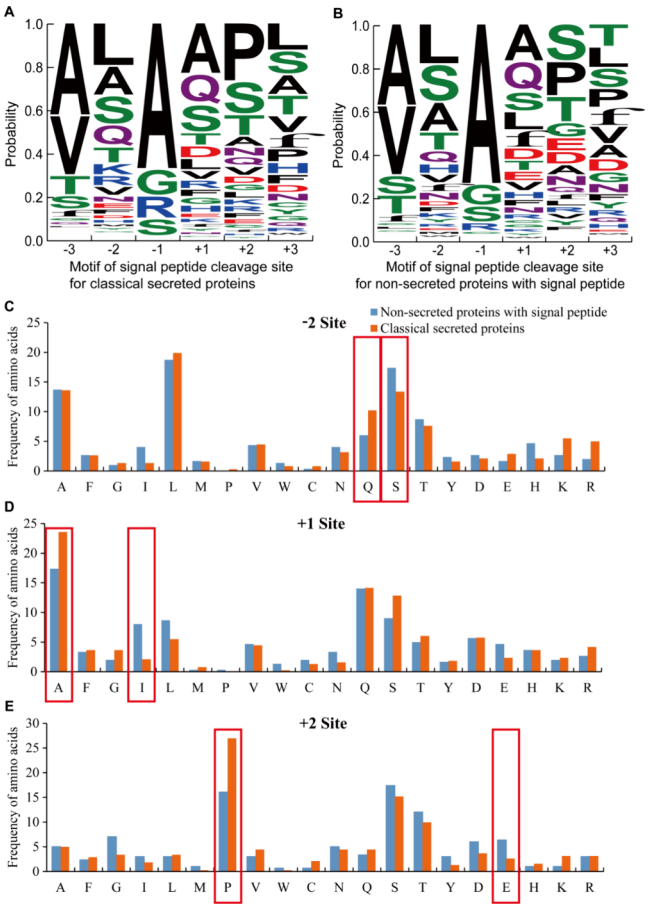



**Figure 4.** Amino acid residue composition of signal peptide cleavage motifs. (A) Distribution of 20 amino acids on each site of cleavage motifs in classical secreted proteins. B) Same distribution in nonsecreted proteins. (C−E) Frequency of 20 amino acids at (C) −2, (D) + 1, and (E) +2 sites for both classical secreted proteins (orange) and nonsecreted proteins (blue). Red circles indicated the amino acids showed the top two remarkable difference between the two groups.

indicated as a major type in proteins annotated in this work, including 42 proteins and accounting for 48%. Further analysis of the protein−protein interaction (PPI) was conducted on earlier annotated proteins (Figure 5). Intriguingly, the number of hydrolases reached 195 with a percentage of 58.2%, including the proteins with the top 3 interaction scores in this PPI network. Taken together, the importance of secreted hydrolases is suggested based on both NCBI records and our results.

**RNA Profiling.** A total of 10,314 genes were acquired from the genome-wide profiling results, which covers >96% of the entire collection of *T. marneffei* ATCC18224 proteins. Since Mp1p protein has been widely utilized in immune diagnosis as the solitary biomarker for *T. marneffei* infection,[32] it was hired in this study as the reference for collection of proteins with robustness. The average FPKM value of its encoding gene is 311.49, while the coefficient of variation is as high as 150.46%. The threshold was then set up as FPKM ≥ 30 and the coefficient of variation ≤60%, which means the gene expression level is close to or higher than 10% of Mp1p but with much better stability.

After that, genes encoding 375 of 584 secreted proteins were plotted for FPKM value ≥1, and then 97 proteins were


https://doi.org/10.1021/acsomega.4c00571
ACS Omega 2024, 9, 27093−27103

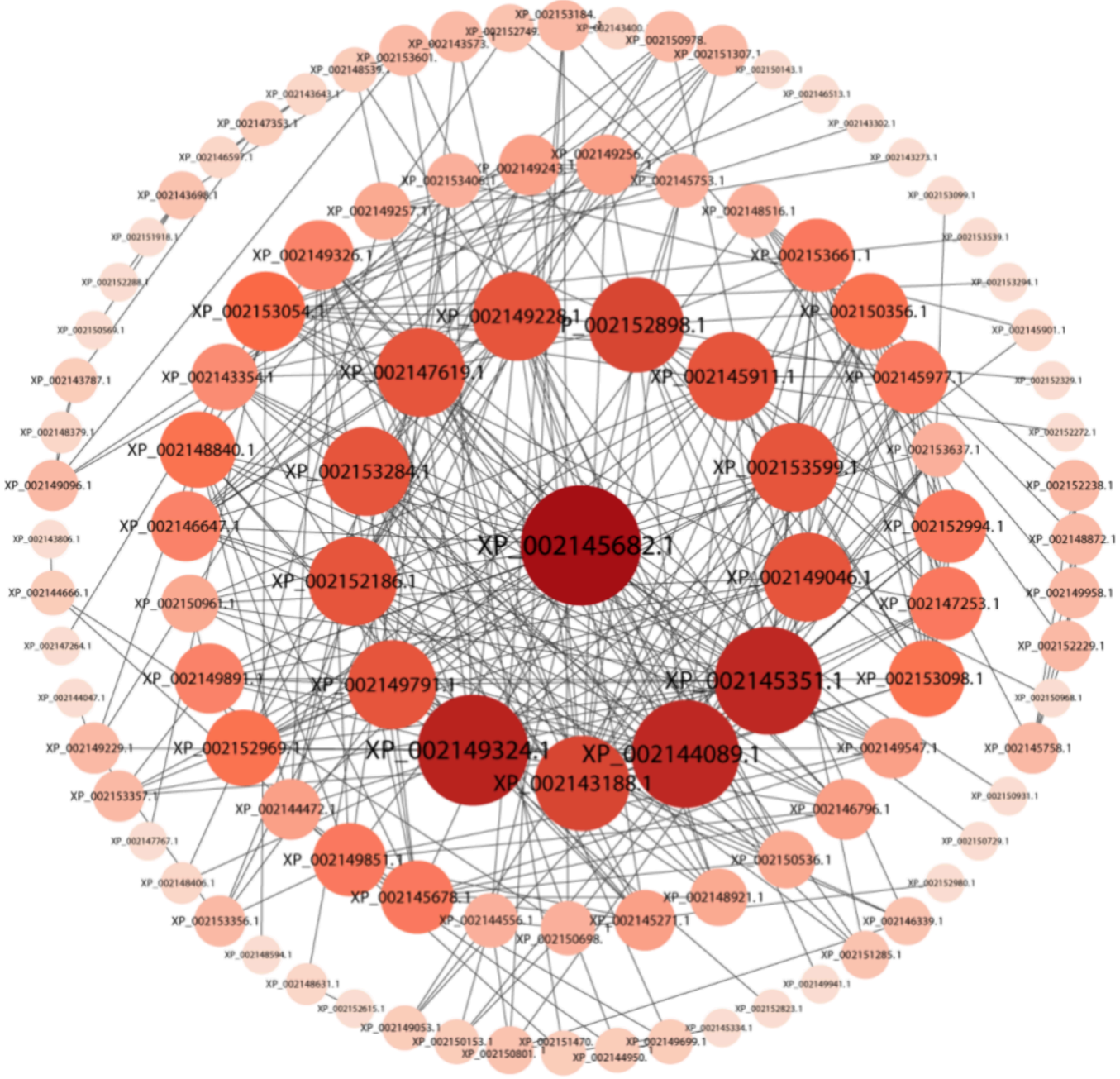



**Figure 5.** PPI network of secreted proteins with earlier NCBI records of function. Larger sizes and darker colors indicate a higher score of interactions.

obtained with the threshold above (Figure 6A, lower right). This set of 97 proteins contains 14 annotated in this study and 59 annotated with NCBI records (Figure 6B,C). Hence, 73 proteins were harvested in total and demonstrated with robust expression and functional information (listed in Table S7).

**Selection and Characterization of Biomarkers.** To set the standard of specificity for potential biomarkers, protein homology comparison was first applied on the Mp1p protein.[33] No significant similarity was found with any proteins encoded by other five most dominant pathogens in HIV coinfections, consistent with the clinical records of its exceptional specificity (the results are listed in Table S8).[17] Three proteins (XP_002147788.1, XP_002148731.1, and XP_002149805.1) were then highlighted with specificity as remarkable as Mp1p through the same procedure from the 73 proteins harvested above (listed in Table S8).

Among these three highlighted biomarkers, XP_002147788.1 and XP_002148731.1 are classical secreted proteins and XP_002149805.1 is a nonclassical one. For the functional category, XP_002147788.1 was classified into phosphoinositide phospholipase C, a subfamily of phospholipase C, which exists widely across bacteria, fungi, plants, and mammals (Figure 7A).[34] XP_002148731.1 was annotated as a potential galactomannoprotein since it is rich in serine and threonine residues and composed of 5$\alpha$-helixes (Access ID:AF-B6QK32-F1) (Figure 7B).[35] Similar structural properties were also observed in Mp1p(PDB: 6J6F).[36,37] XP_002149805.1 was predicted to be a certain type of hydrolase but not ensured in oxoprolinase or phospholipase family (Figure 7C). The former


https://doi.org/10.1021/acsomega.4c00571
ACS Omega 2024, 9, 27093−27103

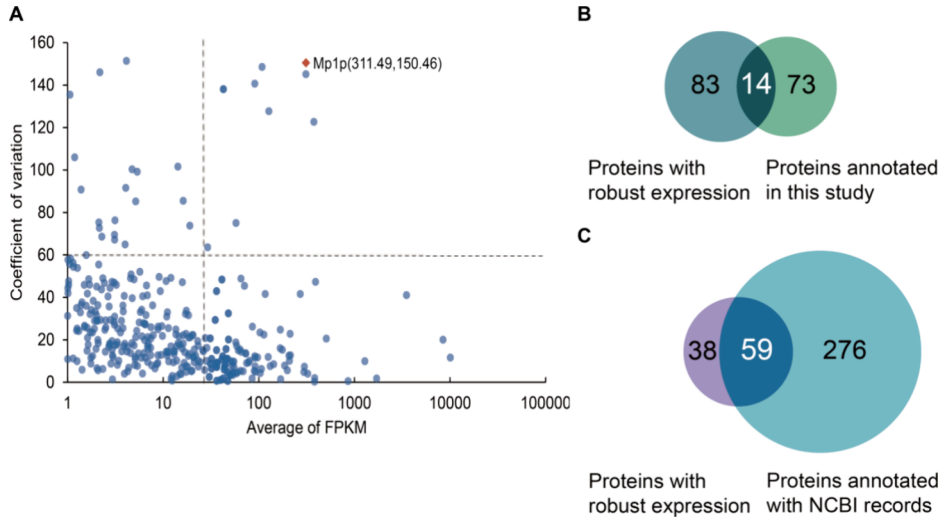



**Figure 6.** Expression and functional state of secreted proteins. (A) Plot of 375 genes encoded secreted proteins with average value of FPKM ≥ 1. The horizontal coordinates are shown in a logarithmic form. Vertical dashed line indicates FPKM value = 30, and the horizontal dashed line indicates coefficient of variation = 60%. (B) Venn diagrams between proteins with robust mRNA level and proteins annotated in the current study. (C) Venn diagrams between proteins with robust mRNA level and proteins annotated with NCBI records.

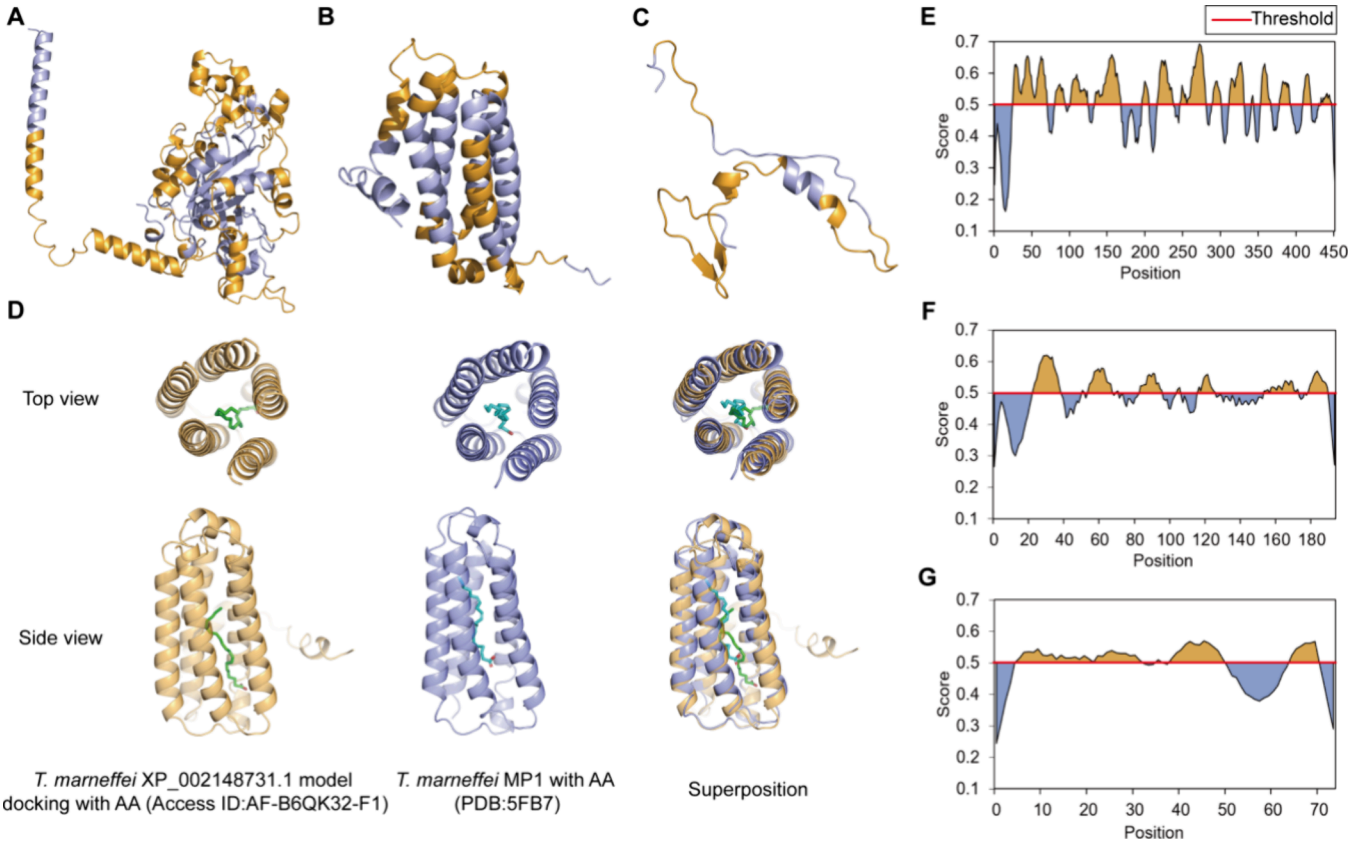



**Figure 7.** Prediction of structures and potential B-cell epitopes of the biomarker candidates. (A) Structure of XP_002147788.1. (B) Structure ofXP_002148731.1. C) Structure of XP_002149805.1. Tertiary structures of three proteins were downloaded from the Alphafold database (Access id: AF-B6QFP2−1, AF-B6QK32-F1, and AF-B6QH20-F1).[35,41] (D) Structural docking of TM galactomannoproteins with an AA molecule. (E) Potential B-cell epitopes of XP_002147788.1. (F) Potential B-cell epitopes of XP_002148731.1. (G) Potential B-cell epitopes of XP_002149805.1. Threshold was set as 0.5 for the prediction of potential epitopes. Regions above the threshold are indicated with yellow. The structure of *T. marneffei* MP1(PDB: 5FB7) was adapted from "Sze KH, Lam WH, Zhang H, et al. *Talaromyces marneffei* Mp1p Is a Virulence Factor that Binds and Sequesters a Key Proinflammatory Lipid to Dampen Host Innate Immune Response. Cell Chem Biol. 2017;24(2):182−194."

result was suggested by InterPro, and the latter one was from NCBI-CD SEARCH and SUPERFAMILY2.0.

After the comparison of XP_002148731.1 and Mp1p-like protein 10 (MPLP10), we strongly prefer that they are the same protein based on their similarities in protein sequence,



https://doi.org/10.1021/acsomega.4c00571
*ACS Omega* 2024, 9, 27093−27103

gene sequence, and genomic circumstance (Figure S1). Molecular docking was carried out with arachidonic acid (AA), an inflammatory-inducing lipid molecule. As revealed by the docking result (Figure 7D), AA was enclosed within the hydrophobic cavity comprising of its five $\alpha$-helixes, which is similar to the interaction between Mp1p and AA. Meanwhile, binding simulation was carried out between AA and XP_002148731.1 or 5FB7 in Autodock Vina[38,39] (Figure S2). Only very slight difference was showed in the $\Delta G$ value of two complexes, −7.85 kcal/mol for AA-XP_002148731.1 and −7.88 kcal/mol for AA-5FB7. Furthermore, an RMSD value of 1.117 Å was confirmed with structural comparison between two binding complexes.

B-cell epitopes included in these three proteins were further identified in silico for further design of the detection strategies that target the antibodies in serum/plasma samples of infected patients. With the IEBD web server,[40] 13 epitopes showed in XP_002147788.1, the potential phospholipase C. These epitopes are located at 24−70, 80−94, 101−124, 133−167, 196−204, 216−237, 246−280, 289−300, 311−331, 353−366, 378−392, 409−419, and 433−447, with the score of antigenicity reaching 0.7 (Figure 7E). In XP_002148731.1, the candidate of galactomannoprotein, six fragments were predicted as B-cell epitopes, which are ranged as 22−38, 54−66, 83−95, 116−125, 159−172, and 178−189 with antigenicity score >0.6 (Figure 7F). Similarly, XP_002149805.1 was estimated containing three epitopes for B cells, positioned as 4−31, 38−49, and 64−70 and scored between 0.5 and 0.6 (Figure 7G).

## DISCUSSION

The ineffectiveness of current diagnostic methods for *T. marneffei* infection occasionally leads to delayed treatments and increases the mortality rate in the case of opportunistic infections related to HIV. One of the critical challenges is the lack of qualified biomarkers. The reliability of a diagnostic biomarker shall be evaluated multifacetedly, including distinctiveness, robustness, accessibility, and clear biological/pathological functions. Great value lies in secreted proteins since they are released out of microorganisms generally and are, therefore, possibly easy to be accessed in the body fluid of hosts. With our prediction, a list of 584 secreted proteins emerged from *T. marneffei* genome, comprising 382 classical and 202 nonclassical proteins. 87 HPs and UPs among them received their functional information thereafter.

For these 584 secreted proteins, slight but clear differences were found in their molecular size and signal peptide when comparing the classical secreted proteins, nonclassical proteins, and nonsecreted proteins containing signal peptide. The molecular size, or represented as the distribution of polypeptide length, looked larger in the classical group rather than in the nonclassical group. For the signal peptides, they serve the protein transportation to endoplasmic reticulum for both classical secreted group and the nonsecreted group. After arrival, they are cleaved off, and other motifs on the proteins will be exploited in subsequent sorting steps, including acidic sorting signals, a folded epitope, and other unknown signals.[29] In one word, signal peptides are removed much earlier than making the decision to secrete a protein or not. As for the cleavage motif of signal peptides, in both groups, the same type (A-X-A) was adopted and proceeded with the same class of protease (classical signal peptidase I). Even though differential incorporation was discovered with certain amino acids at each site of the motif, the biological principles lying behind are still waiting for future studies.

In the lateral stage of current research, three candidate biomarkers were luckily highlighted through combination of RNA sequencing and homology comparison. Their promising value in diagnosis is strongly supported by the impressive specificity from other pathogens and the robust level of mRNA expression. It is worth mentioning that these biomarker candidates were contributed from newly annotated proteins in this investigation rather than former NCBI records. Their capability of antibody induction was underlined with the prediction of B-cell epitopes inside since the existence of corresponding antibodies in the host body fluid can provide the historical information on *T. marneffei* infection and will also be adopted as diagnostic markers.

XP_002147788.1 and XP_002148731.1, potentially a phosphoinositide phospholipase C and a galactomannoprotein, possibly are involved in a certain interactive game of immune mediating with AA, an omega-6 polyunsaturated fatty acid serving as a crucial inducer of proinflammatory signaling.

AA is commonly preserved abundantly in the phospholipids of the human cell membrane and lipid droplets in immune cells under normal conditions.[42] Upon inflammation or microbial infection, the AA molecule will be released from the cell membrane phospholipids under the catalysis of phospholipase A2.[43] An escalation of plasma concentration will then occur pathologically up to 500 $\mu$M from the physiological range of 0.1−50 $\mu$M.[43] Members of phospholipase C family have been reported previously involved in this immune reaction through catalyzing the generation of DAG (diacylglycerol), precursor of AA biosynthesis.[34,44]

On the other side, AA is the only kind of fatty acid capable of binding to Mp1p protein in *T. marneffei* infected macrophages. One or two molecules of AA can be accommodated inside of each ligand binding domain of Mp1p (LBD1/2). Based on the five $\alpha$-helixes from each LBD, the hydrophobic cavity is formed which traps the AA molecule tightly. In this cavity, the hydrocarbon tail of AA is encircled with massive hydrophobic interactions, while its carboxyl head is held by a few hydrogen bonds.[37,45] Trapping of free AA potentially enhances the anti-inflammatory response thereafter and facilitates the immune evasion of *T. marneffei* ultimately.[37,45] Interestingly, a similar cavity with strong hydrophobicity is also clearly seen in the tertiary structure of the XP_002148731.1 protein, which has been predicted to play similar hydrophobic interactions with AA. Hence, functional similarities are latent behind these two galactomannoproteins.

Overall, the stepwise prediction and annotation provided us three promising biomarker candidates for *T. marneffei* immunodetection. These molecules will promote a new tool of diagnosis based on antigen−antibody interaction. Moreover, they may even serve as potential targets to break the immune evasion of *T. marneffei*.

## CONCLUSIONS

In this study, we identified 382 classical and 202 nonclassical secreted proteins encoded in *T. marneffei* standard. Functions were newly assigned to 87 unannotated ones. Three predicted proteins (XP_002147788.1, XP_002148731.1, and XP_002149805.1) were strongly highlighted with remarkable specificity, robust expression, distinct function, and clear antigenicity. Based on the structural similarity to phosphoinositide phospholipase C and galactomannoprotein Mp1p,


https://doi.org/10.1021/acsomega.4c00571
ACS Omega 2024, 9, 27093−27103

respectively, XP_002147788.1 and XP_002148731.1 were suggested involving in the game of immune and inflammation through biosynthesis and capture of AA.

Ultimately, this study proposed a small set of proteins, serving as promising candidate biomarkers for *T. marneffei* diagnosis and also potential therapeutic targets.

## MATERIALS AND METHODS

**Protein Encoding Information on Standard *T. marneffei* Strain.** The entire set of protein sequences encoded in *T. marneffei* standard strain, ATCC18224, was downloaded from the National Center for Biotechnology Information (NCBI) database, with accession number GCA_000001985.1.

**Prediction of Classical Secreted Proteins.** According to Figure 1B, SignalP-6.0 was employed in the initial step to predict the presence of N-terminal signal peptide for the proteins secreted via the classical pathway.[46] Outcome proteins were then applied in the prediction of transmembrane structure with TMHMM-2.0.[47] For the proteins free of transmembrane helix, the third step was then carried out with BIG-PI Fungal Predictor to exclude the lipid-anchored proteins.[48] BUSCA was utilized in the last step to evaluate the extracellular localized proteins from the lipid-anchoring free ones.[49] Web addresses of online platforms used in this study are listed in Table S9.

**Prediction of Nonclassical Secreted Proteins.** For the nonclassical secreted proteins, which are defined due to the absence of N-terminal signal peptide, SecretomeP-2.0[50] was hired to initiate the prediction following the design in Figure 1B. The proteins showed no signal peptide and with a score >0.6 were loaded subsequently to TMHMM-2.0 and BUSCA to obtain the proteins lacking either transmembrane domain or intracellular localization signal.

**Functional Annotation of Hypothetical and Uncharacterized Proteins.** NCBI retrieving was performed to clarify the status of functional annotation for both classical and nonclassical secreted proteins. Cross-platform annotation was then conducted for HPs or UPs as follows.

First, protein−protein interaction was analyzed with STRING.[51] The confident level was set as 0.4 as the medium. Integrative analysis on the protein family and structural domain was conducted via five online tools, including InterPro, Pfam, CATH, NCBI-CD SEARCH, and SUPERFAMILY2.0.[52−56] New annotations were then recognized once results showed agreements or similarities from at least three out of five independent tools. For the newly annotated proteins, homology comparison was performed within the NCBI nonredundant protein database against related fungus in *Talaromyces*. Threshold for identity was set as ≥90%.[57−59]

**RNA Profiling.** Standard strain of *T. marneffei* ATCC18224 was obtained from American Type Culture Collection (ATCC) and cultured under yeast condition. In brief, the fungus was expanded at 37 °C for 14 days, with 200−220 rpm rotation in RPMI 1640 medium plus 10% FBS (fetal bovine serum). The fungus was then precipitated with centrifuge of 14000 g for 10 min.[60] RNA samples were then extracted and sequenced in GeneChem (China). Three independent biological replicates were set in parallel from culture to profiling. Reads were calculated and normalized for each unique transcript. The raw data of RNA sequencing had been uploaded to the NCBI database, which can be retrieved from NCBI with accession number GSE266975. Average and standard deviation of the FPKM (fragments per kilobase of transcript per million mapped reads) value were calculated for genes expressed in the pathogenic yeast condition. To highlight the ones with robust expression, the threshold was set accordingly as average ≥30 (close to 10% of Mp1p abundance) and coefficient of variation ≤60% (ratio of standard deviation to average).

**Prediction of B-Cell Epitopes on Biomarker Candidates.** The online platform IEDB was utilized to predict the B-cell epitopes of biomarker candidates.[40]

**Molecular Docking.** Structure model of the Mp1p-like domain (XP_002148731.1) was obtained from the AlphaFold2 database (access ID: AF-B6QK32-F139).[35,41] Structure data file of AA was acquired from PubChem (access ID: 444899341).[61] Protein−ligand docking was performed using the ROISE server and the Ligand Docking protocol,[62−64] with default parameter settings. A total of 200 ligand conformers were generated using BCL,[65] and the docking process began with the original coordinates from the ligand structure data file. Low-resolution sampling involved 500 steps of Monte Carlo sampling with a maximum search radius of 5.0 Å, and the initial position of the ligand was randomized within a 3 Å radius and restricted in a 15.0 Å radius grid. The translation size was set to 0.1 Å, and the rotation size was set to 5.0° in the low-resolution Monte Carlo step. Six cycles of high-resolution docking and three cycles of repacking were performed. Rosetta docking output score data are included in Table S10, and the top 10 docking models with the lowest interface score/total score ratio were selected (Figure S3). The top-ranked docking model was further superimposed using Wincoot 0.9.7 with AF-B6QK32-model as the reference coordinate.[66] To illustrate the ligand binding pocket of the docking model, top and side views were created using Open-Source PyMOL.[67]

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c00571.

Comparison between XP_002148731.1 and MPLP10; structural docking of TM galactomannoproteins with AA molecule using Autodock Vina; and molecular docking scores (PDF)

Classical secreted proteins; nonclassical secreted proteins; counts and residue usage frequency at each site on cleavage motif of classical secreted and nonsecreted proteins; functional status of secreted proteins; annotated results of HPs and UPs from five independent resources; results of homology comparison of 87 newly annotated proteins; 73 proteins with robust expression and functional information; results of homology comparison of 73 proteins with robust expression and functional information; online platforms used in this study; and molecular docking ranking (XLSX)

## AUTHOR INFORMATION

### Corresponding Authors

**Xuzhen Guo** − *Center for Energy Metabolism and Reproduction, Institute of Biomedicine and Biotechnology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*; Email: guoxuzhen14@163.com



https://doi.org/10.1021/acsomega.4c00571
*ACS Omega* 2024, 9, 27093−27103

**Lei Tan** − *Center for Energy Metabolism and Reproduction, Institute of Biomedicine and Biotechnology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China; Department of Cardiology, Shenzhen Guangming District People's Hospital, Shenzhen 518055, China*; Email: lei.tan@siat.ac.cn

**Junjun Jiang** − *Collaborative Innovation Centre of Regenerative Medicine and Medical BioResource Development and Application Co-constructed by the Province and Ministry and Guangxi Key Laboratory of AIDS Prevention and Treatment & Biosafety III Laboratory, Guangxi Medical University, Nanning, Guangxi 530021, China*; Email: jiangjunjun@gxmu.edu.cn

### Authors

**Jing Dan** − *Collaborative Innovation Centre of Regenerative Medicine and Medical BioResource Development and Application Co-constructed by the Province and Ministry and Guangxi Key Laboratory of AIDS Prevention and Treatment & Biosafety III Laboratory, Guangxi Medical University, Nanning, Guangxi 530021, China; Center for Energy Metabolism and Reproduction, Institute of Biomedicine and Biotechnology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*; orcid.org/0009-0004-1474-354X

**Wudi Wei** − *Guangxi Key Laboratory of AIDS Prevention and Treatment & Biosafety III Laboratory, Guangxi Medical University, Nanning, Guangxi 530021, China*

**Weijie Ou** − *Center for Energy Metabolism and Reproduction, Institute of Biomedicine and Biotechnology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*

**Guangshi Gao** − *Geekgene Technology Co. Ltd., Beijing 100091, China*

**Wanjun Song** − *Geekgene Technology Co. Ltd., Beijing 100091, China*

**Li Ye** − *Guangxi Key Laboratory of AIDS Prevention and Treatment & Biosafety III Laboratory, Guangxi Medical University, Nanning, Guangxi 530021, China*

**Hao Liang** − *Collaborative Innovation Centre of Regenerative Medicine and Medical BioResource Development and Application Co-constructed by the Province and Ministry and Guangxi Key Laboratory of AIDS Prevention and Treatment & Biosafety III Laboratory, Guangxi Medical University, Nanning, Guangxi 530021, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c00571

### Author Contributions

#J.D. and W.W. contributed equally to this work. L.T., J.J., and X.G. conceived the project. J.D. and W.O. conducted bioinformatic predictions. W.W. performed *T. marneffei* culture and RNA sequencing. G.G. and W.S. performed RNA sequencing analysis. J.D. charged the entire process of data analysis. J.D., W.W., L.T., X.G., and J.J. prepared the manuscript.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We are grateful for the National Natural Science Foundation of China (NSFC; 82302550) and One Hundred Talented Youth Program of the Chinese Academy of Sciences to Dr. Lei Tan.

## REFERENCES

(1) Rajasingham, R.; Smith, R. M.; Park, B. J.; et al. Global Burden of Disease of HIV-Associated Cryptococcal Meningitis: An Updated Analysis. *Lancet Infect Dis* **2017**, *17* (8), 873−881.

(2) Friedland, G.; Churchyard, G. J.; Nardell, E. Tuberculosis and HIV Coinfection: Current State of Knowledge and Research Priorities. *J. Infect. Dis.* **2007**, *196*, S1−S3.

(3) Hu, Y.; Zhang, J.; Li, X.; et al. *Penicillium marneffei* Infection: An Emerging Disease in Mainland China. *Mycopathologia* **2013**, *175* (1−2), 57−67.

(4) Thomas, C. F., Jr.; Limper, A. H. Current Insights into the Biology and Pathogenesis of *Pneumocystis* Pneumonia. *Nat. Rev. Microbiol* **2007**, *5* (4), 298−308.

(5) Cáceres, D. H.; Gómez, B. L.; Tobón, Á. M.; et al. Tackling Histoplasmosis Infection in People Living with HIV from Latin America: From Diagnostic Strategy to Public Health Solutions. *J. Fungi* **2023**, *9* (5), 558 DOI: 10.3390/jof9050558.

(6) Qin, Y.; Huang, X.; Chen, H.; et al. Burden of *Talaromyces marneffei* Infection in People Living with HIV/AIDS in Asia During Art Era: A Systematic Review and Meta-Analysis. *BMC Infect Dis* **2020**, *20* (1), 551.

(7) Pruksaphon, K.; Nosanchuk, J. D.; Ratanabanangkoon, K.; et al. *Talaromyces marneffei* Infection: Virulence, Intracellular Lifestyle and Host Defense Mechanisms. *J. Fungi* **2022**, *8* (2), 200 DOI: 10.3390/jof8020200.

(8) Tsang, C. C.; Lau, S. K. P.; Woo, P. C. Y. Sixty Years from Segretain's Description: What Have We Learned and Should Learn About the Basic Mycology of *Talaromyces marneffei*? *Mycopathologia* **2019**, *184* (6), 721−729.

(9) Cao, C.; Xi, L.; Chaturvedi, V. Talaromycosis (Penicilliosis) Due to *Talaromyces* (*Penicillium*) *marneffei*: Insights into the Clinical Trends of a Major Fungal Disease 60 years after the Discovery of the Pathogen. *Mycopathologia* **2019**, *184* (6), 709−720.

(10) Limper, A. H.; Adenis, A.; Le, T.; et al. Fungal Infections in HIV/AIDS. *Lancet Infect Dis* **2017**, *17* (11), e334−e343.

(11) Son, V. T.; Khue, P. M.; Strobel, M. Penicilliosis and AIDS in Haiphong, Vietnam: Evolution and Predictive Factors of Death. *Med. Mal Infect* **2014**, *44* (11−12), 495−501.

(12) Kawila, R.; Chaiwarith, R.; Supparatpinyo, K. Clinical and Laboratory Characteristics of Penicilliosis marneffei among Patients with and without HIV Infection in Northern Thailand: A Retrospective Study. *BMC Infect Dis* **2013**, *13*, 464.

(13) Hien, H. T. A.; Thanh, T. T.; Thu, N. T. M.; et al. Development and Evaluation of a Real-Time Polymerase Chain Reaction Assay for the Rapid Detection of *Talaromyces marneffei* Mp1 Gene in Human Plasma. *Mycoses* **2016**, *59* (12), 773−780.

(14) Huang, Y. T.; Hung, C. C.; Liao, C. H.; et al. Detection of Circulating Galactomannan in Serum Samples for Diagnosis of *Penicillium marneffei* Infection and Cryptococcosis among Patients Infected with Human Immunodeficiency Virus. *J. Clin Microbiol* **2007**, *45* (9), 2858−62.

(15) Xue, X.; Zou, J.; Fang, W.; et al. Characteristics and Prognosis of *Talaromyces marneffei* Infection in HIV-Positive Children in Southern China. *Mycopathologia* **2022**, *187* (2−3), 169−180.

(16) Yoshimura, Y.; Sakamoto, Y.; Lee, K.; et al. *Penicillium marneffei* Infection with β-D-Glucan Elevation: A Case Report and Literature Review. *Intern Med.* **2016**, *55* (17), 2503−6.

(17) Thu, N. T. M.; Chan, J. F. W.; Ly, V. T.; et al. Superiority of a Novel Mp1p Antigen Detection Enzyme Immunoassay Compared to Standard Bactec Blood Culture in the Diagnosis of Talaromycosis. *Clin Infect Dis* **2021**, *73* (2), e330−e336.

(18) Wang, Y. F.; Cai, J. P.; Wang, Y. D.; et al. Immunoassays Based on *Penicillium marneffei* Mp1p Derived from *Pichia Pastoris*

Expression System for Diagnosis of Penicilliosis. *PLoS One* **2011**, *6* (12), No. e28796.
(19) Park, B. C.; Reese, M.; Tagliabracci, V. S. Thinking Outside of the Cell: Secreted Protein Kinases in Bacteria, Parasites, and Mammals. *IUBMB Life* **2019**, *71* (6), 749−759.
(20) Wang, X.; Chung, K. P.; Lin, W.; et al. Protein Secretion in Plants: Conventional and Unconventional Pathways and New Techniques. *J. Exp Bot* **2018**, *69* (1), 21−37.
(21) Zhao, G.; Zhu, X.; Zhang, H.; et al. Novel Secreted Protein of *Mycoplasma bovis* Mbovp280 Induces Macrophage Apoptosis through Cryab. *Front Immunol* **2021**, *12*, No. 619362.
(22) National Center for Biotechnology Information. Https://Www.Ncbi.Nlm.Nih.Gov/Taxonomy/Browser/Wwwtax.Cgi?Id=441960&Mode=Info. (accessed September 20, 2023).
(23) Liu, X. Y.; X. Ruan, M.; F. Cheng, S.; et al. Functional Prediction and Analysis of *Agaricus bisporus* Secretome. *Acta Edulis Fungi* **2017**, *24*, 3.
(24) Ning, Z. J.; H. Jiang, B.; Liu, Q.; et al. Genome-Wide Prediction and Analysis of Secreted Proteins of *Enterocytozoon hepatopenaei*. *Prog. Fish. Sci.* **2020**, *41* (6), 165−173.
(25) Zeng, R.; Gao, S.; Xu, L.; et al. Prediction of Pathogenesis-Related Secreted Proteins from *Stemphylium lycopersici*. *BMC Microbiol* **2018**, *18* (1), 191.
(26) Mazumder, L.; Hasan, M. R.; Fatema, K.; et al. Structural and Functional Annotation and Molecular Docking Analysis of a Hypothetical Protein from *Neisseria gonorrhoeae*: An in-Silico Approach. *Biomed. Res. Int.* **2022**, *2022*, No. 4302625.
(27) Cohen, M. J.; Chirico, W. J.; Lipke, P. N. Through the Back Door: Unconventional Protein Secretion. *Cell Surf.* **2020**, *6*, No. 100045.
(28) Viotti, C. Er to Golgi-Dependent Protein Secretion: The Conventional Pathway. *Methods Mol. Biol.* **2016**, *1459*, 3−29.
(29) Barlowe, C. K.; Miller, E. A. Secretory Protein Biogenesis and Traffic in the Early Secretory Pathway. *Genetics* **2013**, *193* (2), 383−410.
(30) Xing, Q. K.; L. Li, X.; Cao, Y.; et al. Prediction and Analysis of Candidate Secreted Proteins from the Genome of *Lasiodiplodia theobromae*. *Sci. Agric. Sin.* **2020**, *53*, 5027−5038.
(31) Zalucki, Y. M.; Jennings, M. P. Signal Peptidase I Processed Secretory Signal Sequences: Selection for and against Specific Amino Acids at the Second Position of Mature Protein. *Biochem. Biophys. Res. Commun.* **2017**, *483* (3), 972−977.
(32) Thompson, G. R., 3rd; Le, T.; Chindamporn, A.; et al. Global Guideline for the Diagnosis and Management of the Endemic Mycoses: An Initiative of the European Confederation of Medical Mycology in Cooperation with the International Society for Human and Animal Mycology. *Lancet Infect Dis* **2021**, *21* (12), e364−e374.
(33) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; et al. Gapped Blast and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389−402.
(34) Cocco, L.; Follo, M. Y.; Manzoli, L.; et al. Phosphoinositide-Specific Phospholipase C in Health and Disease. *J. Lipid Res.* **2015**, *56* (10), 1853−60.
(35) Varadi, M.; Anyango, S.; Deshpande, M.; et al. Alphafold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50* (D1), D439−d444.
(36) Liao, S.; Tung, E. T.; Zheng, W.; et al. Crystal Structure of the Mp1p Ligand Binding Domain 2 Reveals Its Function as a Fatty Acid-Binding Protein. *J. Biol. Chem.* **2010**, *285* (12), 9211−20.
(37) Lam, W. H.; Sze, K. H.; Ke, Y.; et al. *Talaromyces marneffei* Mp1 Protein, a Novel Virulence Factor, Carries Two Arachidonic Acid-Binding Domains to Suppress Inflammatory Responses in Hosts. *Infect. Immun.* **2019**, *87*, 4.
(38) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; et al. Autodock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf Model* **2021**, *61* (8), 3891−3898.
(39) Forli, S.; Huey, R.; Pique, M. E.; et al. Computational Protein-Ligand Docking and Virtual Drug Screening with the Autodock Suite. *Nat. Protoc* **2016**, *11* (5), 905−19.
(40) Vita, R.; Zarebski, L.; Greenbaum, J. A.; et al. The Immune Epitope Database 2.0. *Nucleic Acids Res.* **2010**, *38*, D854−D862.
(41) Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly Accurate Protein Structure Prediction with Alphafold. *Nature* **2021**, *596* (7873), 583−589.
(42) Hanna, V. S.; Hafez, E. A. A. Synopsis of Arachidonic Acid Metabolism: A Review. *J. Adv. Res.* **2018**, *11*, 23−32.
(43) Sonnweber, T.; Pizzini, A.; Nairz, M.; et al. Arachidonic Acid Metabolites in Cardiovascular and Metabolic Diseases. *Int. J. Mol. Sci.* **2018**, *19*, 3285.
(44) Katan, M. Families of Phosphoinositide-Specific Phospholipase C: Structure and Function. *Biochim. Biophys. Acta* **1998**, *1436* (1−2), 5−17.
(45) Sze, K. H.; Lam, W. H.; Zhang, H.; et al. *Talaromyces marneffei* Mp1p Is a Virulence Factor That Binds and Sequesters a Key Proinflammatory Lipid to Dampen Host Innate Immune Response. *Cell Chem. Biol.* **2017**, *24* (2), 182−194.
(46) Petersen, T. N.; Brunak, S.; von Heijne, G.; et al. Signalp 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat. Methods* **2011**, *8* (10), 785−6.
(47) Krogh, A.; Larsson, B.; von Heijne, G.; et al. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **2001**, *305* (3), 567−80.
(48) Eisenhaber, B.; Schneider, G.; Wildpaner, M.; et al. A Sensitive Predictor for Potential GPI Lipid Modification Sites in Fungal Protein Sequences and Its Application to Genome-Wide Studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J. Mol. Biol.* **2004**, *337* (2), 243−53.
(49) Savojardo, C.; Martelli, P. L.; Fariselli, P.; et al. BUSCA: An Integrative Web Server to Predict Subcellular Localization of Proteins. *Nucleic Acids Res.* **2018**, *46* (W1), W459−w466.
(50) Bendtsen, J. D.; Jensen, L. J.; Blom, N.; et al. Feature-Based Prediction of Non-Classical and Leaderless Protein Secretion. *Protein Eng. Des Sel* **2004**, *17* (4), 349−56.
(51) Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; et al. The STRING Database in 2023: Protein-Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest. *Nucleic Acids Res.* **2023**, *51* (D1), D638−d646.
(52) Blum, M.; Chang, H. Y.; Chuguransky, S.; et al. The Interpro Protein Families and Domains Database: 20 Years On. *Nucleic Acids Res.* **2021**, *49* (D1), D344−d354.
(53) Lu, S.; Wang, J.; Chitsaz, F.; et al. Cdd/Sparcle: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48* (D1), D265−d268.
(54) Mistry, J.; Chuguransky, S.; Williams, L.; et al. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* **2021**, *49* (D1), D412−d419.
(55) Pandurangan, A. P.; Stahlhacke, J.; Oates, M. E.; et al. The Superfamily 2.0 Database: A Significant Proteome Update and a New Webserver. *Nucleic Acids Res.* **2019**, *47* (D1), D490−d494.
(56) Sillitoe, I.; Lewis, T. E.; Cuff, A.; et al. Cath: Comprehensive Structural and Functional Annotations for Genome Sequences. *Nucleic Acids Res.* **2015**, *43*, D376−D381.
(57) da Costa, W. L. O.; Araújo, C. L. A.; Dias, L. M.; et al. Functional Annotation of Hypothetical Proteins from the *Exiguobacterium antarcticum* Strain B7 Reveals Proteins Involved in Adaptation to Extreme Environments, Including High Arsenic Resistance. *PLoS One* **2018**, *13* (6), No. e0198965.
(58) Dey, S.; Shahrear, S.; Afroj Zinnia, M.; et al. Functional Annotation of Hypothetical Proteins from the *Enterobacter cloacae* B13 Strain and Its Association with Pathogenicity. *Bioinf. Biol. Insights* **2022**, *16*, No. 11779322221115535.
(59) Johnson, M.; Zaretskaya, I.; Raytselis, Y.; et al. Ncbi Blast: A Better Web Interface. *Nucleic Acids Res.* **2008**, *36*, W5−W9.


https://doi.org/10.1021/acsomega.4c00571
*ACS Omega* 2024, 9, 27093−27103

(60) Wei, W.; Wang, G.; Zhang, H.; et al. *Talaromyces marneffei* Suppresses Macrophage Inflammation by Regulating Host Alternative Splicing. *Commun. Biol.* **2023**, *6* (1), 1046.
(61) Kim, S.; Chen, J.; Cheng, T.; et al. Pubchem 2023 Update. *Nucleic Acids Res.* **2023**, *51* (D1), D1373−d1380.
(62) Combs, S. A.; Deluca, S. L.; Deluca, S. H.; et al. Small-Molecule Ligand Docking into Comparative Models with Rosetta. *Nat. Protoc* **2013**, *8* (7), 1277−98.
(63) DeLuca, S.; Khar, K.; Meiler, J. Fully Flexible Docking of Medium Sized Ligand Libraries with Rosettaligand. *PLoS One* **2015**, *10* (7), No. e0132508.
(64) Lyskov, S.; Chou, F. C.; Conchúir, S.; et al. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (Rosie). *PLoS One* **2013**, *8* (5), No. e63906.
(65) Kothiwale, S.; Mendenhall, J. L.; Meiler, J. Bcl::Conf: Small Molecule Conformational Sampling Using a Knowledge Based Rotamer Library. *J. Cheminform* **2015**, *7*, 47.
(66) Emsley, P.; Lohkamp, B.; Scott, W. G.; et al. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (Pt 4), 486−501.
(67) Schrodinger, L. *The Pymol Molecular Graphics System, Version 1.8*, 2015.



https://doi.org/10.1021/acsomega.4c00571
*ACS Omega* 2024, 9, 27093−27103