Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

In silico genomic, phylogenetic, and functional characterization of candidate DNase-like proteins in Paracoccidioides brasiliensis and Paracoccidioides lutzii: implications for neutrophil extracellular traps

Olatunbosun Joseph Agunbiade
São Paulo State University (UNESP)
Olabisi Flora Davies-Bolorunduro
Floret Center for Advanced Genomics and Bioinformatics Research
Luciane Alarcão Dias-Melicio
dias.melicio@unesp.br

São Paulo State University (UNESP)

Research Article

Keywords: Paracoccidioides brasiliensis, Paracoccidioides lutzii, TatD, DNase, neutrophil extracellular traps, NETs, fungal infection

Posted Date: September 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10842272/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Paracoccidioides brasiliensis and *Paracoccidioides lutzii* are thermodimorphic fungal pathogens causing paracoccidioidomycosis, a systemic mycosis endemic to Latin America. Here, we investigated PADG_08285 and PADG_11161, two hypothetical proteins from *Paracoccidioides* associated with extracellular DNase activity and potential involvement in neutrophil extracellular trap (NET) degradation. Proteins were retrieved using Biopython Entrez, and *P. lutzii* orthologs were identified by local BLAST+ searches. Multiple sequence alignment and maximum-likelihood phylogenetic reconstruction were performed using MAFFT and IQ-TREE, respectively. Functional domains, physicochemical properties, localization, signal peptides, and structural models were assessed using InterPro, Biopython ProteinAnalysis, DeepLoc, SignalP 6.0, and AlphaFold. PADG_08285 and PADG_11161 were identified as one-to-one orthologs of A0A0A2UZ79 and C1H8L0, respectively. Phylogenetic analysis indicated that short- and long-form proteins diverged before *Paracoccidioides* speciation. Short-form proteins were classified as TatD family deoxyribonucleases and showed conservation of the predicted catalytic tetrad Glu121, His158, His184, and Asp232, whereas long-form proteins were assigned to the Exonuclease V family. Neither candidate possessed a canonical signal peptide. Our findings identify the short-form TatD protein as the stronger candidate for the reported extracellular DNase activity and support further experimental investigation of its potential role in NET degradation.

1. Introduction

Paracoccidioides brasiliensis and *Paracoccidioides lutzii* are thermodimorphic fungal pathogens that cause paracoccidioidomycosis (PCM), a systemic mycosis endemic to Latin America (Peçanha et al. 2022a; de Oliveira et al. 2023). The disease is particularly associated with Brazil, where most reported cases occur, although infections have also been described in other countries of South and Central America (Agrizzi et al. 2026). PCM is closely associated with rural and agricultural environments, reflecting the ecological distribution of *Paracoccidioides* and the frequent exposure of individuals involved in activities that bring them into contact with soil (Zeminian de Oliveira et al. 2021). The disease predominantly affects the adults and can present with pulmonary involvement followed by dissemination to other tissues, resulting in chronic disease and substantial morbidity in affected individuals (Shikanai-Yasuda et al. 2017; Peçanha et al. 2022b; Agrizzi et al. 2026). The ability of *Paracoccidioides* to survive and adapt within the mammalian host is closely related to its thermally regulated transition between an environmental mycelial phase and a parasitic yeast phase (Castilho et al. 2014; Chaves et al. 2021). This transition is triggered primarily by the increase in temperature encountered during host infection and represents an important component of fungal adaptation and pathogenicity (Venturini et al. 2026).

The establishment of infection requires the fungus to overcome several components of the host immune response. Neutrophils are among the first immune cells recruited to sites of infection and contribute to fungal control through phagocytosis, production of reactive molecules, release of antimicrobial components, and formation of neutrophil extracellular traps (NETs) (Desai and Lionakis 2018; Urban and Nett 2019; Klassert et al. 2023). NETs consist of decondensed chromatin, histones, and antimicrobial proteins. Because the DNA scaffold is central to their structural integrity, it represents a primary target for microbial nucleases facilitating immune evasion.

The ability to modify or degrade NETs can therefore contribute to microbial survival during infection. Microbial nucleases that degrade extracellular DNA have been described as mechanisms of immune evasion in several pathogenic organisms. By breaking down the DNA framework of NETs, extracellular nucleases can reduce the stability of these structures and potentially facilitate microbial escape (Zonta et al. 2021). Fungi are also capable of interacting with NETs, and extracellular DNase activity capable of degrading NET DNA has been reported in *Paracoccidioides* (Della Coletta et al. 2015; Urban and Nett 2019; Zonta et al. 2021). Despite this observation, the protein or proteins responsible for this activity have not been clearly established. The absence of a defined molecular basis for this phenotype leaves how Paracoccidioides interacts with neutrophil-mediated host defense poorly understood.

PADG_08285 and PADG_11161 are two proteins from *P. brasiliensis* that have been associated with extracellular DNase activity (Zonta et al. 2021). Both remain insufficiently characterized, and their annotation as hypothetical or DNase-like proteins does not establish their enzymatic identity or mechanism of action. The two proteins are also markedly different in size. PADG_08285 contains 332 amino acids, whereas PADG_11161 contains 742 amino acids, suggesting that they may represent different protein families rather than closely related forms of the same nuclease. This distinction is important because the identification of a nuclease family can provide information about the likely catalytic mechanism and substrate characteristics of a protein. Establishing whether these proteins belong to distinct nuclease families may therefore help determine which candidate has stronger molecular support for the reported DNase activity.

Comparison between *P. brasiliensis* and *P. lutzii* allows us to examine the conservation of these candidate proteins. Although both species are associated with PCM, they represent distinct members of the genus and contribute to the epidemiological and biological diversity of the disease (Teixeira et al. 2014; Aranteset al. 2015). Conserved proteins involved in host adaptation or fungal survival are expected to retain sequence and structural characteristics across related species. Identification of corresponding proteins in *P. lutzii* and assessment of their evolutionary relationships can therefore help determine whether the candidate nuclease features are conserved within *Paracoccidioides*. Conservation of a protein family together with residues associated with catalytic activity would provide stronger support for a biologically relevant function than annotation based only on sequence similarity. The distinction between the two candidate proteins is particularly important in this context. A protein associated with DNase activity may belong to a nuclease family with a well-defined catalytic architecture, while a second protein associated with the same phenotype may have a different function or contribute through another mechanism. Domain and family analysis can therefore provide a more informative basis for interpreting the original DNase-like annotation.

The cellular location of these proteins is also relevant to their proposed involvement in NET degradation (Della Coletta et al. 2015). Classical secretion of many fungal proteins involves an N-terminal signal peptide, making the presence or absence of such a sequence relevant when assessing the extracellular DNase hypothesis. However, the absence of a canonical signal peptide does not by itself demonstrate that a protein cannot occur outside the cell. Protein export can occur through mechanisms that do not depend on classical signal peptides, and computational localization predictions alone cannot establish

the actual cellular distribution of a protein. For this reason, localization predictions and signal peptide analysis need to be considered together with the protein's evolutionary and functional characteristics. In the present study, these features were examined for both candidate proteins and their *P. lutzii* counterparts. Structural characterization further supports evaluation of the candidate proteins. Identification of conserved catalytic residues from sequence and domain analyses does not establish whether these residues occupy an appropriate structural arrangement. Predicted three-dimensional structures can provide additional information about the organization of conserved residues and allow comparison of the candidate proteins across species. For the short-form protein, structural examination of the predicted TatD-related region can therefore complement the sequence-based evidence and provide a basis for subsequent experimental investigation (Zhang et al. 2023). This study provides a comparative characterization of PADG_08285 and PADG_11161 from *P. brasiliensis* and their corresponding proteins in *P. lutzii*. Here, we characterized the evolutionary, structural, and physicochemical properties of these candidates to clarify their molecular identity and prioritize targets for experimental validation. Particular attention was given to the distinction between the short and long forms and to the extent to which their predicted features support a role in DNase activity. By bringing these findings together, the study provides a molecular basis for prioritizing candidate proteins for experimental investigation of extracellular nuclease activity and their possible contribution to NET-associated host defense during *Paracoccidioides* infection.

2. Materials and Methods

2.1. Computational environment and software

All computational analyses, except subcellular localization and signal peptide prediction, were performed in Python 3.11.15 using the latest available versions of the respective software and Python libraries at the time of the study. Biopython was used for sequence retrieval, sequence processing, and physicochemical analysis (Cock et al. 2009). BLAST + was used for ortholog identification (Camacho et al. 2009), MAFFT for multiple sequence alignment (Katoh 2002), IQ-TREE for maximum-likelihood phylogenetic reconstruction using the Q.MAMMAL+G4 amino acid substitution model as determined by ModelFinder based on Bayesian Information Criterion (Minh et al. 2020), and InterPro-based analyses for protein domain and family characterization (Blum et al. 2025). Three-dimensional protein models were obtained from the AlphaFold Protein Structure Database for structural analysis (Varadi et al. 2022). Subcellular localization was predicted using DeepLoc (Almagro Armenteros et al. 2017; Thumuluri et al. 2022), while signal peptide prediction was performed using SignalP 6.0 (Teufel et al. 2022; Lisacek 2024; Nielsen 2025). These two analyses were conducted outside the Visual Studio Code Jupyter environment. The complete computational workflow was implemented through Jupyter Notebooks to facilitate reproducibility.

2.2. Sequence retrieval and ortholog identification

The amino acid sequence of the candidate proteins from *Paracoccidioides brasiliensis* were retrieved programmatically from the National Center for Biotechnology Information (NCBI) protein database using the Entrez module of Biopython (Sayers et al. 2021). The retrieved proteins were obtained in FASTA format and used as input for subsequent analyses.

Orthologous proteins in *P. lutzii* were identified using BLAST+ protein searches against the *P. lutzii* Pb01 reference proteome (strain ATCC MYA-826). Each *P. brasiliensis* candidate protein was used independently as a query sequence. Candidate matches were evaluated using percentage sequence identity, alignment length, E-value, and bit score. Reciprocal best-hit relationships were used to support one-to-one orthology assignments (Tatusov et al. 1997; Moreno-Hagelsieb and Latimer 2008).

To confirm that the identified paralogs represent fixed evolutionary traits rather than strain-specific artifacts, expanded BLASTp searches were conducted against the broader *Paracoccidioides* genus (taxid: 12086). This confirmed strict intra- and inter-species conservation across multiple sequenced isolates and cryptic species, including *P. brasiliensis* (Pb18), *P. lutzii* (Pb01), and *P. americana* (Pb03). Furthermore, to establish an evolutionary anchor, homologous sequences were identified through an expanded BLASTp search against the *Histoplasma* genus. Representative outgroup proteins (*Histoplasma capsulatum* strain G186AR) were selected based on appropriate evolutionary divergence while maintaining strict conservation of the core functional domains to confirm the ancestral origin and deep conservation of these distinct paralogs.

2.3. Multiple sequence alignment and phylogenetic reconstruction

Multiple sequence alignment was performed using MAFFT. The candidate proteins and their identified orthologous proteins were included in the alignment. The resulting alignment was used for maximum-likelihood phylogenetic reconstruction with IQ-TREE. The best-fitting amino acid substitution model was selected according to the Bayesian information criterion before tree reconstruction. Branch support was assessed using 1,000 ultrafast bootstrap replicates (Hoang et al. 2018). The resulting phylogenetic topology was used to examine the evolutionary relationships among the candidate proteins and their orthologs.

2.4. Protein domain and family characterization

Protein domain architecture and family assignments were assessed using InterPro-based annotation. The analysis incorporated the available member databases within the InterPro framework, including Pfam, PANTHER, PROSITE, Conserved Domain Database, CATH-Gene3D, PIRSF, and SUPERFAMILY (Sigrist et al. 2010; Marchler-Bauer et al. 2017; Bitard-Feildel and Callebaut 2017; Sayers et al. 2021; Thomas et al. 2022; Blum et al. 2025; Paysan-Lafosse et al. 2025). The resulting annotations were used to identify predicted protein families, domains, and conserved functional sites. Conserved residues associated with predicted nuclease functions were examined by comparing the corresponding positions among the candidate proteins and their orthologs. Conserved functional sites identified through the domain analyses were recorded for subsequent comparative interpretation.

2.5. Physicochemical characterization

Physicochemical properties were calculated from the amino acid proteins using the ProteinAnalysis module of Biopython (Cock et al. 2009). The parameters evaluated included molecular weight, isoelectric point (pI), instability index, and grand average of hydropathicity (GRAVY) (Kyte and Doolittle 1982). The calculated properties were used for comparative characterization of the candidate proteins and their orthologs.

2.6. Subcellular localization and signal peptide prediction

Subcellular localization was predicted using DeepLoc (Almagro Armenteros et al. 2017; Thumuluri et al. 2022). Localization probabilities were obtained for the cellular compartments included in the prediction framework, including the cytoplasm, nucleus, extracellular space, cell membrane, mitochondrion, plastid, endoplasmic reticulum, lysosome or vacuole, Golgi apparatus, and peroxisome. The predicted localization profiles were recorded for comparative analysis. The presence of a canonical N-terminal signal peptide was assessed using SignalP 6.0 (Teufel et al. 2022). Signal peptide and non-signal-peptide probabilities were recorded for each protein, and the resulting predictions were subsequently compared among the candidate proteins and their orthologs.

2.7. Structural analysis

Three-dimensional protein models were obtained from the AlphaFold Protein Structure Database (Varadi et al. 2022). The predicted structures were examined to assess the overall structural organization of the candidate proteins and the spatial arrangement of conserved residues identified during the sequence and domain analyses. Structural models were compared between the candidate proteins and their corresponding orthologs to assess conservation of relevant structural features.

3. Results

3.1. Ortholog Identification and Evolutionary Topology

Reciprocal BLAST+ searches against the reference proteome of *Paracoccidioides lutzii* (strain Pb01 / ATCC MYA-826) identified distinct one-to-one orthology relationships for both *Paracoccidioides brasiliensis* candidate nucleases. The short-form *P. brasiliensis* sequence, PADG_08285, comprises 332 amino acid residues and aligned directly with the uncharacterized *P. lutzii* target A0A0A2UZ79, which spans 315 amino acids (Table 1). This local alignment showed 97.78% sequence identity across a continuous 315-residue alignment block. The statistical significance of this match is established by an expectation value of 0 and a bit score of 639 (Table 2). Concurrently, the long-form *P. brasiliensis* query locus, PADG_11161, encompassing 742 amino acid residues, paired with the *P. lutzii* ortholog C1H8L0 (Table 1), is classified under the database description of Exonuclease V. This sequence alignment spans a length of 740 residues, achieving a sequence identity score of 94.47% over a 742-residue alignment block. The structural homology is supported by an expectation value of 0 and an alignment bit score of 1285 (Table 2). Evolutionary relationships inferred via maximum-likelihood phylogenetic reconstruction resolved the divergence history of these candidate proteins. Phylogenetic analysis separated the candidate proteins into two distinct evolutionary clades. The primary node separating the short-form group, comprising PADG_08285 and A0A0A2UZ79, from the long-form group, comprising PADG_11161 and C1H8L0, is supported by an absolute ultrafast bootstrap value of 100 (Fig. 1), confirming the branch separation.

Table 1
Primary sequence characteristics of candidate DNase-like proteins and their identified *P. lutzii* orthologs.

Sequence ID	Description	Sequence Length
PADG_08285	PADG_08285 P. brasiliensis candidate DNase-like protein	332
PADG_11161	PADG_11161 P. brasiliensis candidate DNase-like protein	742
C1H8L0	P. lutzii ortholog	740
A0A0A2UZ79	P. lutzii ortholog	315

Table 2
Local BLAST+ metrics supporting one-to-one orthologous relationships between P. brasiliensis and P. lutzii candidates.

Query Gene	Subject ID	Identity(%)	Align Length	E-value	Bit Score	Protein Description
PADG_08285	tr|A0A0A2UZ79|A0A0A2UZ79_PARBA	97.778	315	0	639	tr|A0A0A2UZ79|A0A0A2UZ79_PARBA Uncharacterized protein OS=Paracoccidioides lutzii (strain ATCC MYA-826 / Pb01) OX = 502779 GN=PAAG_12429 PE = 3 SV = 1
PADG_11161	tr|C1H8L0|C1H8L0_PARBA	94.474	742	0	1285	tr|C1H8L0|C1H8L0_PARBA Exonuclease V OS=Paracoccidioides lutzii (strain ATCC MYA-826 / Pb01) OX = 502779 GN=PAAG_07101 PE = 3 SV = 1

3.2. Nuclease Domain Classification and Catalytic Architecture

InterPro annotation revealed distinct enzymatic domains for the two protein families. The short-form orthologs, PADG_08285 and A0A0A2UZ79, mapped to the TatD-related deoxyribonuclease family (Kabsch et al. 1990; Chen et al. 2014). Relevant database identifiers include InterPro IPR001130 and IPR050891, Pfam PF01026, and Conserved Domain Database cd01310. This family forms a structural clade within the metal-dependent hydrolase superfamily, noted by CATH-Gene3D G3DSA:3.20.20.140 and SUPERFAMILY SSF51556 (Table 3). Active-site residue profiling demonstrated complete conservation of the core catalytic machinery across both species. In *P. brasiliensis* PADG_08285, the active site consists of a conserved catalytic tetrad: Glutamic Acid at position 121,

Histidine at position 158, Histidine at position 184, and Aspartic Acid at position 232. This configuration is replicated identically in the *P. lutzii* short-form ortholog A0A0A2UZ79 at positions Glu121, His158, His184, and Asp232 (Table 4).

Table 3
Functional domain signatures and protein superfamily assignments based on InterPro database mapping.

Protein ID	Database	Domain Accession	Domain Name	Domain Type
C1H8L0	interpro	IPR019190	Exonuclease V	family
C1H8L0	pfam	PF09810	Exonuclease V - a 5' deoxyribonuclease	Family
C1H8L0	panther	PTHR14464	EXONUCLEASE V	Family
A0A0A2UZ79	cdd	cd01310	TatD DNAse	Domain
A0A0A2UZ79	cathgene3d	G3DSA:3.20.20.140	Metal-dependent hydrolases	Homologous superfamily
A0A0A2UZ79	interpro	IPR001130	3'-5' ssDNA/RNA exonuclease TatD-like	family
A0A0A2UZ79	interpro	IPR018228	Deoxyribonuclease, TatD-related, conserved site	Conserved site
A0A0A2UZ79	interpro	IPR032466	Metal-dependent hydrolase	Homologous superfamily
A0A0A2UZ79	interpro	IPR050891	TatD-type Metallo-dependent Hydrolase	family
A0A0A2UZ79	pfam	PF01026	TatD related DNase	domain
A0A0A2UZ79	pirsf	PIRSF005902	Mg-dependent DNase, TatD type	family
A0A0A2UZ79	prosite	PS01090	TatD deoxyribonuclease family signature 2	Conserved site
A0A0A2UZ79	panther	PTHR10060	TatD-type Metallo-dependent Hydrolase	family
A0A0A2UZ79	ssf	SSF51556	Metallo-dependent hydrolases	Homologous superfamily

Table 4
Putative active-site residue locations and interspecies conservation.

Protein	UniProt ID	Feature Type	Residue Position	Description
PADG_08285	C1GLP4	Binding site	121–121	Glutamic Acid
PADG_08285	C1GLP4	Binding site	158–158	Histidine
PADG_08285	C1GLP4	Binding site	184–184	Histidine
PADG_08285	C1GLP4	Binding site	232–232	Aspartic Acid
P lutzii (Short)	A0A0A2UZ79	Binding site	121–121	Glutamic Acid
P lutzii (Short)	A0A0A2UZ79	Binding site	158–158	Histidine
P lutzii (Short)	A0A0A2UZ79	Binding site	184–184	Histidine
P lutzii (Short)	A0A0A2UZ79	Binding site	232–232	Aspartic Acid
PADG_11161	A0A0A0HU65	Binding site	248–248	Aspartic Acid
PADG_11161	A0A0A0HU65	Binding site	389–389	Histidine
PADG_11161	A0A0A0HU65	Binding site	470–470	Serine
PADG_11161	A0A0A0HU65	Binding site	586–586	Phenylalanine
P lutzii (Long)	C1H8L0	Binding site	246–246	Aspartic Acid
P lutzii (Long)	C1H8L0	Binding site	387–387	Histidine
P lutzii (Long)	C1H8L0	Binding site	467–467	Serine
P lutzii (Long)	C1H8L0	Binding site	588–588	Asparagine

The long-form proteins, PADG_11161 and C1H8L0, were assigned to the Exonuclease V family, corresponding to InterPro IPR019190, Pfam PF09810, and PANTHER PTHR14464 (Table 3). Comparative alignment of the putative binding positions within these long-form candidates revealed structural variations between the two species. The initial binding coordinates show positional conservation, featuring Aspartic Acid at Asp248 in PADG_11161 and Asp246 in C1H8L0, Histidine at His389 in PADG_11161 and His387 in C1H8L0, and Serine at Ser470 in PADG_11161 and Ser467 in C1H8L0 (Table 4). However, a non-conservative substitution occurs at the fourth position. At this location, the *P. brasiliensis* sequence features a hydrophobic aromatic Phenylalanine residue at position 586, whereas the *P. lutzii* counterpart contains a polar Asparagine residue at position 588.

3.3. Physicochemical Properties and Biophysical Stability Profiles

Primary sequence analysis revealed biophysical differences between the short-form TatD and long-form Exonuclease V lineages, particularly concerning molecular mass and predicted stability. The instability index (threshold < 40) demonstrated that the two lineages generated opposing profiles (Guruprasad et al. 1990). The short-form TatD nucleases generated instability index values of 29.17 for PADG_08285 and 33.37 for A0A0A2UZ79. The long-form Exonuclease V targets generated instability scores above the stability boundary, with PADG_11161 reaching 52.51 and C1H8L0 rising to 57.40.

Distinct trends were also observed in overall mass and hydrophobic profiles (Fig. 2). The short-form variants possess smaller molecular masses, calculated at 37.08 kDa for PADG_08285 and 34.90 kDa for A0A0A2UZ79. The isoelectric points are centered at 6.09 and 5.55, respectively. The long-form variants exhibit larger molecular masses, determined at 81.46 kDa for PADG_11161 and 81.42 kDa for C1H8L0. Their isoelectric points are calculated at 5.50 and 5.66, respectively. All four candidate proteins displayed negative Grand Average of Hydropathicity scores (Kyte and Doolittle 1982). The short forms scored − 0.319 for PADG_08285 and − 0.214 for A0A0A2UZ79, while the long forms scored − 0.511 for PADG_11161 and − 0.542 for C1H8L0 (Table 5). These negative parameters indicate a hydrophilic surface architecture across both groups. The high sequence identity (97.78%) and conservation of the predicted catalytic residues *between P. brasiliensis PADG_08285 and P. lutzii A0A0A2UZ79* support the conservation of the short-form TatD-like protein within the two species examined. However, the present data do not establish that this protein is a genus-wide virulence factor or *that P. lutzii* uses a non-classical secretory pathway to counter NETs. These hypotheses require direct experimental evidence, including demonstration of extracellular localization and nuclease activity.

Table 5
Predicted physicochemical parameters and biophysical stability markers.

Protein ID	Sequence Length	Molecular Weight (kDa)	Isoelectric Point (pI)	Instability Index	GRAVY Score
PADG_08285	332	37.08	6.09	29.17	-0.319
PADG_11161	742	81.46	5.5	52.51	-0.511
C1H8L0	740	81.42	5.66	57.4	-0.542
A0A0A2UZ79	315	34.9	5.55	33.37	-0.214

3.4. Subcellular Localization Mapping and the Secretion Model

DeepLoc compartmentalization analysis predicted that all four candidate proteins reside within the intracellular matrix, returning preferential distribution probabilities for the cytoplasm and the nucleus(Almagro Armenteros et al. 2017; Thumuluri et al. 2022). DeepLoc analysis indicated high cytoplasmic and nuclear probability scores for all candidates, with negligible baseline probabilities for standard extracellular localization (Table 6; Fig. 4). This prediction of an intracellular destination is supported by the results of N-terminal sorting signal mapping. SignalP 6.0 analysis established the absence of a canonical N-terminal signal peptide across all four proteins (Teufel et al. 2022). The algorithms assigned probability scores to the non-secretory classification category, returning a value of 0.999999 for both short-form targets, PADG_08285 and A0A0A2UZ79, and a value of 1.000000 for both long-form targets, PADG_11161 and C1H8L0 (Table 7).

Table 6
DeepLoc subcellular localization probability distributions across ten eukaryotic cellular compartments.

Protein ID	Localizations	Signals	Cytoplasm	Nucleus	Extracellular	Cell membrane	Mitochondrion	Plastid	Endoplasmic reticulum	Lysos
PADG_08285	Cytoplasm|Nucleus	Nuclear localization signal	0.6993	0.6003	0.0314	0.0539	0.3273	0.0524	0.1196	0.139
PADG_11161	Cytoplasm|Nucleus	Nuclear localization signal	0.4869	0.5961	0.0461	0.0797	0.1647	0.0063	0.1071	0.047
C1H8L0	Nucleus	Nuclear localization signal	0.4735	0.5986	0.0625	0.0848	0.1714	0.0069	0.0951	0.053
A0A0A2UZ79	Cytoplasm|Nucleus	Nuclear localization signal	0.7097	0.5516	0.0186	0.0766	0.2623	0.0321	0.1010	0.173

Table 7
SignalP 6.0 probability distributions for canonical N-terminal signal peptides.

Protein ID	SignalP Prob (OTHER)	SignalP Prob (SIGNAL PEPTIDE)	SignalP Prediction
PADG_08285	0.999999	0	OTHER
PADG_11161	1	0	OTHER
C1H8L0	1	0	OTHER
A0A0A2UZ79	0.999999	0	OTHER

4. Discussion

Disclosure statement

No potential conflict of interest was reported by the authors.

Ethics, Consent to Participate, and Consent to Publish declarations

Not applicable.

Declaration of generative AI use

During the preparation of this manuscript, the authors used Google Gemini to assist with text formatting, language editing, and code troubleshooting. The tool was not used to write the manuscript, generate scientific data, perform analyses, interpret results, or draw scientific conclusions. All scientific content, analyses, interpretations, and code were prepared, reviewed, and verified by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES; Finance Code 001, master fellowship awarded to Olatunbosun Joseph Agunbiade).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions

CRediT: Olatunbosun Joseph Agunbiade: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Visualization; Olabisi Flora Davies-Bolorunduro: Methodology, Validation, Writing - review & editing; Luciane Alarcão Dias-Melicio: Conceptualization, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declarations

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES; Finance Code 001, master fellowship awarded to Olatunbosun Joseph Agunbiade).

Author Contribution

Olatunbosun Joseph Agunbiade: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft, Visualization; Olabisi Flora Davies-Bolorunduro: Methodology, Validation, Writing - review & editing; Luciane Alarcão Dias-Melicio: Conceptualization, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Agrizzi LC, Kruschewsky WLL, Maifrede SB, Nogarol MAU, de Camargo ZP, Rodrigues AM, Grão-Velloso TR, Peçanha PM, Falqueto A, Gonçalves SS. Rare clinical presentations of paracoccidioidomycosis: Epidemiological, clinical, and microbiological insights from a Brazilian reference center. Braz J Infect Dis [Internet] [accessed 2026 Aug 16]. 2026;30(4). https://doi.org/10.1016/j.bjid.2026.105825.

2. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017. DeepLoc: prediction of protein subcellular localization using deep learning.Hancock J, editor. Bioinformatics. 33(21):3387–3395. https://doi.org/10.1093/bioinformatics/btx431
3. Amundsen SK, Smith GR. RecBCD enzyme: mechanistic insights from mutants of a complex helicase-nuclease. Microbiol Mol Biol Rev. 2023;87(4):e00041–23. https://doi.org/10.1128/mmbr.00041-23.
4. ARANTES TD BAGAGLIE, NIÑO-VEGA G, SAN-BLAS G, THEODORO RC. Paracoccidioides brasiliensis AND Paracoccidioides lutzii, A SECRET LOVE AFFAIR. Rev Inst Med Trop São Paulo. 2015;57(Suppl 19):25–30. https://doi.org/10.1590/S0036-46652015000700006.
5. Arnold DA, Kowalczykowski SC. Rec BCD Helicase/Nuclease. Encycl Life Sci [Internet]. 1st ed. Wiley; 2001. [accessed 2026 Aug 16]. https://doi.org/10.1038/npg.els.0000586. [place unknown].
6. Bitard-Feildel T, Callebaut I. Exploring the dark foldable proteome by considering hydrophobic amino acids topology. Sci Rep. 2017;7(1):41425. https://doi.org/10.1038/srep41425.
7. Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, Pinto BL, Orr A, Paysan-Lafosse T, Ponamareva I, et al. InterPro: the protein sequence classification resource in 2025. Nucleic Acids Res. 2025;53(D1):D444–56. https://doi.org/10.1093/nar/gkae1082.
8. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421.
9. Castilho DG, Chaves AFA, Xander P, Zelanis A, Kitano ES, Serrano SMT, Tashima AK, Batista WL. Exploring Potential Virulence Regulators in Paracoccidioides brasiliensis Isolates of Varying Virulence through Quantitative Proteomics. J Proteome Res. 2014;13(10):4259–71. https://doi.org/10.1021/pr5002274.
10. Chaves AFA, Navarro MV, de Barros YN, Silva RS, Xander P, Batista WL. Updates in Paracoccidioides Biology and Genetic Advances in Fungus Manipulation. J Fungi. 2021;7(2):116. https://doi.org/10.3390/jof7020116.
11. Chen W-J, Lai P-J, Lai Y-S, Huang P-T, Lin C-C, Liao T-H. Probing the catalytic mechanism of bovine pancreatic deoxyribonuclease I by chemical rescue. Biochem Biophys Res Commun. 2007;352(3):689–96. https://doi.org/10.1016/j.bbrc.2006.11.078.
12. Chen Y-C, Li C-L, Hsiao Y-Y, Duh Y, Yuan HS. Structure and function of TatD exonuclease in DNA repair. Nucleic Acids Res. 2014;42(16):10776–85. https://doi.org/10.1093/nar/gku732.
13. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25(11):1422–3. https://doi.org/10.1093/bioinformatics/btp163.
14. Coletta AMD, Bachiega TF, Silva JC, de Q ÂMV, de Faveri C, Marques JD, Marques SA, Ximenes MEA, Dias-Melicio VF. LA. 2015. Neutrophil Extracellular Traps Identification in Tegumentary Lesions of Patients with Paracoccidioidomycosis and Different Patterns of NETs Generation In Vitro. PLoS Negl Trop Dis. 9(9):e0004037. https://doi.org/10.1371/journal.pntd.0004037
15. Desai JV, Lionakis MS. The role of neutrophils in host defense against invasive fungal infections. Curr Clin Microbiol Rep. 2018;5(3):181–9. https://doi.org/10.1007/s40588-018-0098-6.
16. Gurunathan S, Kang M-H, Qasim M, Khan K, Kim J-H. Biogenesis, Membrane Trafficking, Functions, and Next Generation Nanotherapeutics Medicine of Extracellular Vesicles. Int J Nanomed. 2021;16:3357–83. https://doi.org/10.2147/IJN.S310357.
17. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35(2):518–22. https://doi.org/10.1093/molbev/msx281.
18. Jones SJ, Worrall AF, Connolly BA. Site-directed Mutagenesis of the Catalytic Residues of Bovine Pancreatic Deoxyribonuclease I. J Mol Biol. 1996;264(5):1154–63. https://doi.org/10.1006/jmbi.1996.0703.
19. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66. https://doi.org/10.1093/nar/gkf436.
20. Klassert TE, Hölzer M, Zubiria-Barrera C, Bethge J, Klaile E, Müller MM, Marz M, Slevogt H. Differential Transcriptional Responses of Human Granulocytes to Fungal Infection with Candida albicans and Aspergillus fumigatus. J Fungi. 2023;9(10):1014. https://doi.org/10.3390/jof9101014.
21. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32. https://doi.org/10.1016/0022-2836(82)90515-0.
22. Lisacek F, editor. 2024. Protein Bioinformatics [Internet]. New York, NY: Springer US; [accessed 2026 Aug 16]. https://doi.org/10.1007/978-1-0716-4007-4
23. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45(Database issue):D200–3. https://doi.org/10.1093/nar/gkw1129.
24. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.Teeling E, editor. Mol Biol Evol. 37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015
25. Miura N, Ueda M. Evaluation of Unconventional Protein Secretion by Saccharomyces cerevisiae and other Fungi. Cells. 2018;7(9):128. https://doi.org/10.3390/cells7090128.
26. Moreno-Hagelsieb G, Latimer K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics. 2008;24(3):319–24. https://doi.org/10.1093/bioinformatics/btm585.
27. Nielsen H. Practical Applications of Language Models in Protein Sorting Prediction: SignalP 6.0, DeepLoc 2.1, and DeepLocPro 1.0. In: Kc DB, editor. Large Lang Models LLMs Protein Bioinforma [Internet]. Volume 2941. New York, NY: Springer US; ; 2025. pp. 153–75. [accessed 2026 Aug 16]. https://doi.org/10.1007/978-1-0716-4623-6_10.

28. Octaviano-Azevedo CE, Beraldo KRF, Leitão-Júnior NP, de Souza CM, da Silva CP, Sinigaglia RC, Garcés EAM, Duarte EL, Tashima AK, Juliano MA, Puccia R. Virulence and Stress-Related Proteins Are Differentially Enriched and N-Terminally Acetylated in Extracellular Vesicles from Virulent Paracoccidioides brasiliensis. J Fungi. 2025;11(10):751. https://doi.org/10.3390/jof11100751.
29. de Oliveira LBZ, Della Coletta AM, Gardizani TP, Garces HG, Bagagli E, Trilles L, Barrozo LV, Marques S, de Faveri A, Dias-Melicio JD LA. Molecular Phylogenetic Analysis of Paracoccidioides Species Complex Present in Paracoccidioidomycosis Patient Tissue Samples. Microorganisms. 2023;11(3):562. https://doi.org/10.3390/microorganisms11030562.
30. Paysan-Lafosse T, Andreeva A, Blum M, Chuguransky SR, Grego T, Pinto BL, Salazar GA, Bileschi ML, Llinares-López F, Meng-Papaxanthos L, et al. The Pfam protein families database: embracing AI/ML. Nucleic Acids Res. 2025;53(D1):D523–34. https://doi.org/10.1093/nar/gkae997.
31. Peçanha PM, Peçanha-Pietrobom PM, Grão-Velloso TR, Rosa Júnior M, Falqueto A, Gonçalves SS. Paracoccidioidomycosis: What We Know and What Is New in Epidemiology, Diagnosis, and Treatment. J Fungi. 2022a;8(10):1098. https://doi.org/10.3390/jof8101098.
32. Peçanha PM, Peçanha-Pietrobom PM, Grão-Velloso TR, Rosa Júnior M, Falqueto A, Gonçalves SS. Paracoccidioidomycosis: What We Know and What Is New in Epidemiology, Diagnosis, and Treatment. J Fungi. 2022b;8(10):1098. https://doi.org/10.3390/jof8101098.
33. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021;49(D1):D10–7. https://doi.org/10.1093/nar/gkaa892.
34. Shikanai-Yasuda MA, Mendes RP, Colombo AL, de Queiroz-Telles F, Kono ASG, Paniago AMM, Nathan A, Valle ACF do, Bagagli E, Benard G et al. 2017. Brazilian guidelines for the clinical management of paracoccidioidomycosis. Rev Soc Bras Med Trop. 50:715–740. https://doi.org/https://doi.org/10.1590/0037-8682-0230-2017
35. Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res. 2010;38(Database issue):D161–6. https://doi.org/10.1093/nar/gkp885.
36. Tatusov RL, Koonin EV, Lipman DJ. A Genomic Perspective on Protein Families. Science. 1997;278(5338):631–7. https://doi.org/10.1126/science.278.5338.631.
37. Teixeira M, de Theodoro M, Oliveira RC, de Machado FFM, Hahn GC, Bagagli RC, San-Blas E, Soares Felipe G MS. Paracoccidioides lutzii sp. nov.: biological and clinical implications. Med Mycol. 2014;52(1):19–28. https://doi.org/10.3109/13693786.2013.794311.
38. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, Von Heijne G, Nielsen H. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023–5. https://doi.org/10.1038/s41587-021-01156-3.
39. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L-P, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci. 2022;31(1):8–22. https://doi.org/10.1002/pro.4218.
40. Thumuluri 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(W1):W228–34. https://doi.org/10.1093/nar/gkac278.
41. Urban CF, Nett JE. Neutrophil extracellular traps in fungal infection. Semin Cell Dev Biol. 2019;89:47–57. https://doi.org/10.1016/j.semcdb.2018.03.020.
42. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, 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):D439–44. https://doi.org/10.1093/nar/gkab1061.
43. Venturini J, Fernandez NB, de Macedo PM, Cavalcante R, de Caceres S, Borges DH, Rodrigues CL, Gasparoto AM, Giusiano ALDB, Fava G WS, et al. Paracoccidioidomycosis in the 21st century: Challenges and milestones. PLoS Negl Trop Dis. 2026;20(1):e0013819. https://doi.org/10.1371/journal.pntd.0013819.
44. Zeminian de Oliveira LB, Della Coletta AM, Gardizani TP, Barrozo LV, Miot HA, De Faveri J, Dias-Melicio LA. Paracoccidioimycosis and white individuals: Susceptibility and biogeographic aspects in an important endemic area in Brazil. PLoS Negl Trop Dis. 2021;15(2):e0009086. https://doi.org/10.1371/journal.pntd.0009086.
45. Zhang S, Zhang T, Fu Y. Proteome-wide structural analysis quantifies structural conservation across distant species. Genome Res. 2023;33(11):1975–93. https://doi.org/10.1101/gr.277771.123.
46. Zonta YR, Dezen ALO, Della Coletta AM, Yu KST, Carvalho L, Santos LA dos, Deprá I, de Kratofil C, Willson RM, Zbytnuik ME et al. L,. 2021. Paracoccidioides brasiliensis Releases a DNase-Like Protein That Degrades NETs and Allows for Fungal Escape. Front Cell Infect Microbiol [Internet]. [accessed 2026 Aug 16] 10. https://doi.org/10.3389/fcimb.2020.592022

Figures

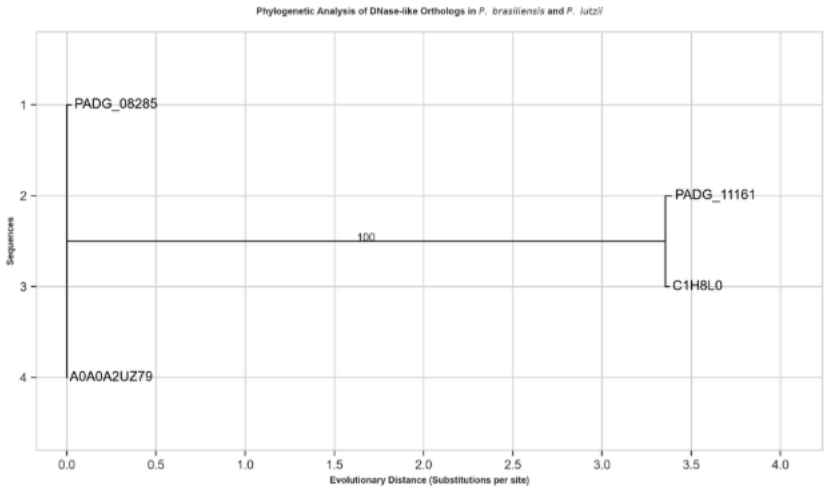


Figure 1

Maximum-likelihood phylogenetic tree of candidate DNase-like orthologs in *P. brasiliensis*and *P. lutzii*. The topology, generated using IQ-TREE, shows a clear separation between the short-form TatD lineage and the long-form Exonuclease V clade, indicating a gene duplication event prior to speciation. Node values represent ultrafast bootstrap support (1,000 pseudoreplicates). The scale bar represents evolutionary distance in substitutions per site.

Alt Text:A horizontal phylogenetic tree displaying two distinct main branches. The first branch clusters the short-form proteins PADG_08285 and A0A0A2UZ79, while the second branch clusters the long-form proteins PADG_11161 and C1H8L0. The deep node separating the two clades shows a bootstrap support value of 100.

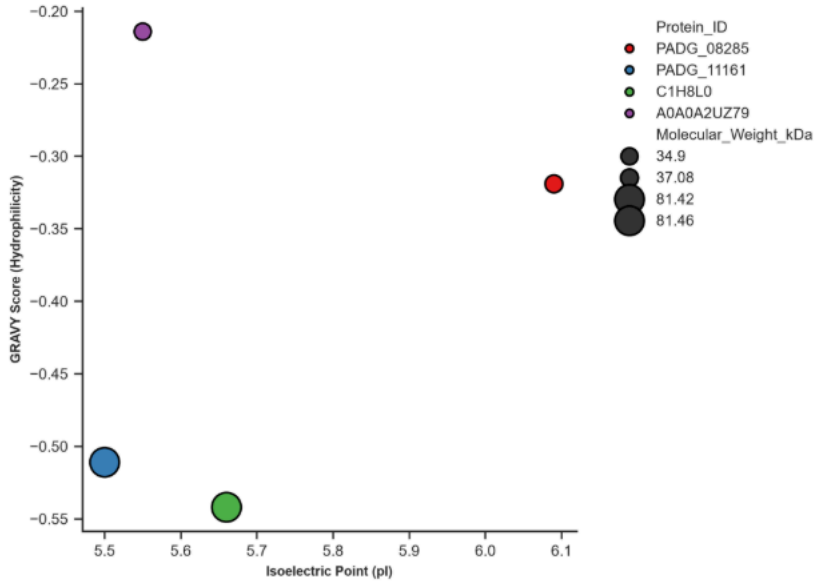


Figure 2

Physicochemical properties of the candidate nucleases. Isoelectric point (pI) is plotted against the Grand Average of Hydropathicity (GRAVY) score for each sequence. Marker size represents calculated molecular mass (kDa), highlighting the distinct physical clustering of the short- and long-form protein lineages.

Alt Text:A scatter plot with isoelectric point on the x-axis and GRAVY score on the y-axis. Four data points represent the candidate proteins. The two short-form proteins group closely together with smaller marker sizes representing lower masses, while the two long-form proteins group together with larger marker sizes representing higher masses. All four points fall in the negative range on the y-axis, indicating hydrophilicity.

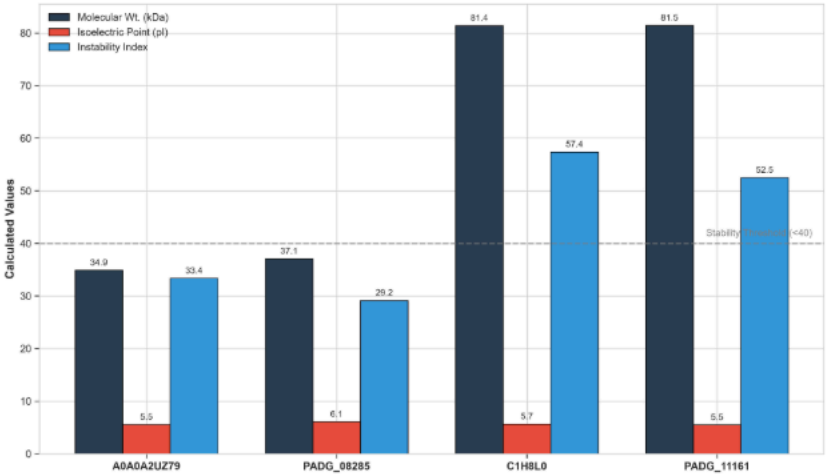


Figure 3

Calculated biophysical parameters of the candidate proteins. The bar chart compares molecular weight (kDa), isoelectric point (pI), and instability index across the four targets. The horizontal dashed line represents the structural instability threshold of 40; only the short-form TatD variants fall below this boundary, indicating stable globular structures.

Alt Text:A grouped bar chart comparing molecular weight, isoelectric point, and instability index for four proteins. The instability index bars for the short-form proteins A0A0A2UZ79 and PADG_08285 are below a dashed horizontal threshold line at 40. The instability index bars for the long-form proteins C1H8L0 and PADG_11161 extend well above the threshold line. Molecular weight bars are significantly taller for the long-form proteins.

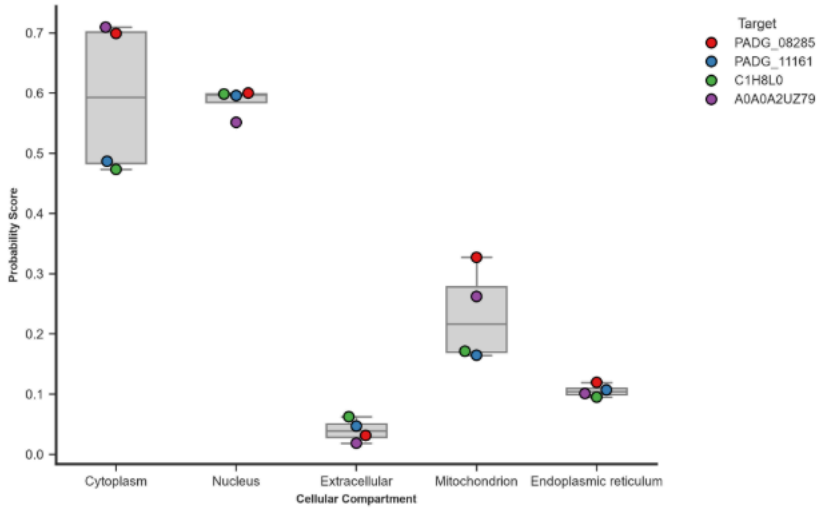


Figure 4

Predicted subcellular localization probabilities. DeepLoc modeling indicates that all candidate nucleases predominantly localize to the cytoplasm and nucleus, with negligible probabilities for the extracellular space and conventional secretory pathways.

Alt Text:A boxplot displaying probability scores for protein localization across different cellular compartments. The cytoplasm and nucleus categories display the highest probability scores, grouping tightly between 0.47 and 0.71. Conversely, the extracellular compartment category shows near-zero probabilities for all four candidate proteins.