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.

Molecular evolution of the ASP/RON1/SUSD protein family across Apicomplexa

Ahmet Efe Köseoğlu

ahmetefekoseoglu@gmail.com

COMED Therapeutics

Research Article

Keywords: Apicomplexa, ASP/RON1/SUSD protein family, CCP/Sushi domain, molecular evolution, phylogenetics

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11885739/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

Purpose

Apical sushi protein (ASP) is a rhoptry neck-associated invasion protein that has been primarily investigated in *Plasmodium* species. Although ASP has been implicated in host-cell invasion, its evolutionary distribution outside malaria parasites and its relationship to homologous proteins described in other apicomplexan lineages remain poorly understood. Here, the molecular evolution of the ASP/RON1/SUSD protein family across representative members of the phylum Apicomplexa was investigated.

Methods

ASP sequences were retrieved from public databases and used to identify homologous proteins in non-plasmodial apicomplexans through BLASTp searches. Comparative sequence analyses, optimized multiple sequence alignments, conserved domain characterization, AlphaFold3 structural modelling, and Maximum Likelihood phylogenetic reconstruction were performed using representative homologues selected to capture major taxonomic diversity and sequence architectures.

Results

ASP orthologues exhibited extensive sequence conservation throughout the genus *Plasmodium* despite their occurrence in parasites infecting diverse vertebrate hosts. Codon-based evolutionary analyses further demonstrated that *Plasmodium* ASP orthologues evolve predominantly under strong purifying selection, indicating substantial functional constraint throughout their evolutionary history. Homologous proteins were additionally identified in *Toxoplasma*, *Neospora*, *Cyclospora*, *Cystoisospora*, *Besnoitia*, and *Cryptosporidium*. Notably, these proteins were annotated under different names, including ASP, rhoptry neck protein 1 (RON1), and sushi-domain-containing proteins (SUSDs). Conserved domain analyses revealed lineage-specific remodeling of domain architecture, including differential distribution of CCP/Sushi domains among major apicomplexan lineages, whereas refined sequence alignments, phylogenetic reconstruction, and structural modelling supported evolutionary continuity despite extensive sequence divergence.

Conclusions

Collectively, the results support the hypothesis that ASP, RON1, and SUSD proteins represent divergent descendants of an ancient apicomplexan protein family. These findings establish a unified evolutionary framework for understanding ASP/RON1/SUSD family molecular evolution and provide new insights into protein family diversification across Apicomplexa.

Introduction

The phylum Apicomplexa comprises predominantly parasitic protists responsible for numerous diseases affecting humans, livestock, companion animals, and wildlife [1–3]. Members of this lineage are characterized by an apical complex, a specialized assemblage of cytoskeletal and secretory structures that mediates host-cell attachment, invasion, and establishment within host tissues [1, 4, 5]. The phylum includes medically and veterinary important genera such as *Plasmodium*, *Toxoplasma*, *Cryptosporidium*, *Neospora*, *Cyclospora*, *Cystoisospora*, and *Besnoitia*, which collectively impose substantial global health and economic burdens [2, 3, 6–8].

The evolutionary success of apicomplexan parasites is closely linked to their ability to invade host cells. Comparative genomic studies have shown that proteins operating at the parasite-host interface are frequently subjected to strong selective pressures associated with host adaptation, immune evasion, and invasion requirements [9]. Nevertheless, some invasion-associated protein families remain detectable across broad taxonomic distances despite extensive sequence divergence, suggesting that certain molecular components of host-cell invasion have been retained throughout apicomplexan diversification [9, 10]. Investigating the evolutionary history of such proteins can therefore provide insights into conserved invasion mechanisms and the diversification of protein families across evolutionarily distant parasite lineages.

Host-cell invasion is mediated by specialized secretory organelles, including micronemes, rhoptries, and dense granules [11]. Among these, rhoptries play central roles by delivering proteins involved in parasite entry and host-cell manipulation. Rhoptry neck proteins, including members of the rhoptry neck (RON) family, contribute to moving-junction formation and mediate critical interactions between parasite and host-cell membranes during invasion [12, 13]. Consequently, rhoptry-associated proteins provide valuable systems for investigating the molecular evolution of invasion mechanisms and invasion-associated protein families across Apicomplexa.

Among proteins associated with the apical invasion machinery, apical sushi protein (ASP) remains comparatively understudied. ASP was originally identified in *Plasmodium falciparum* as a conserved protein encoded by the PFD0295c locus and proposed to participate in erythrocyte invasion [14]. Initial analyses revealed that ASP contains a sushi domain, also known as a complement control protein (CCP) or short consensus repeat (SCR) domain, a conserved protein-interaction module involved in adhesion, recognition, and extracellular signalling processes [14, 15]. ASP was subsequently shown to be conserved among multiple *Plasmodium* species and expressed during invasive parasite stages, supporting a role in host-cell invasion and parasite development [14, 16, 17].

Subsequent studies refined the understanding of ASP localization and function. Immunolocalization analyses demonstrated that ASP is associated with the rhoptry neck rather than micronemes [18], while proteomic investigations identified ASP among proteins released during erythrocyte invasion and implicated it in invasion-associated processes at the parasite-host interface [19]. ASP has since been incorporated into the repertoire of rhoptry neck-associated proteins involved in host-cell invasion [12, 13].

More recently, studies in *Plasmodium berghei* showed that ASP/RON1 is expressed and localized in both merozoites and sporozoites, suggesting that its role may extend beyond merozoite invasion [17].

Interestingly, proteins showing similarity to *Plasmodium* ASP have been annotated under different names across Apicomplexa. In several coccidian parasites, related proteins are described as rhoptry neck protein 1 (RON1), whereas cryptosporidian homologues are often annotated as sushi-domain-containing proteins (SUSDs). Despite these differing annotations, it remains unclear whether ASP, RON1, and SUSDs represent distinct protein families or highly divergent members of a common ancestral family. This uncertainty complicates comparative evolutionary analyses and may obscure the recognition of homologous proteins across different apicomplexan research communities.

Despite growing interest in ASP and related proteins, most studies have focused almost exclusively on malaria parasites [10, 14, 18, 19]. Consequently, little is known about the phylogenetic distribution, conservation, and diversification of ASP-related proteins across Apicomplexa. The increasing availability of genomic resources now enables comparative investigations of protein-family evolution across diverse apicomplexan lineages [20]. Such approaches are particularly informative for invasion-associated proteins because they can reveal evolutionary conservation despite substantial sequence divergence, taxonomic diversification, and domain architectural remodeling [9].

In the present study, the molecular evolution of the ASP/RON1/SUSD protein family was investigated across representative apicomplexan taxa. Building upon previous observations that ASP-like proteins have been annotated differently across apicomplexan lineages, the study aimed to evaluate the evolutionary relationships among these homologues and to determine whether lineage-specific variation in sequence composition, domain architecture, and predicted structural organization remains consistent with homology. ASP conservation within *Plasmodium* species was first assessed, followed by the identification of homologous proteins in non-plasmodial apicomplexans. Comparative sequence analyses, conserved domain characterization, phylogenetic reconstruction, and structural modelling were employed to evaluate patterns of conservation and diversification across the identified homologues. Together, these analyses provide a comparative framework for understanding the conservation, diversification, and evolutionary relationships of this understudied protein family across Apicomplexa.

Methods

Retrieval of Apical Sushi Protein Sequences

Apical sushi protein (ASP) sequences were retrieved from the NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/) using the search term “apical sushi protein”. Searches were restricted to Protists and filtered to include only RefSeq entries in order to obtain curated full-length protein sequences. Sequence accession numbers and annotations were recorded for downstream analyses. To identify ASP homologues outside the genus *Plasmodium*, the ASP sequence of *Plasmodium falciparum* 3D7 (XP_001351372.1) was selected as the reference sequence and used as the query for subsequent similarity searches (Fig. 1).

Multiple Sequence Alignment of Plasmodium ASP Sequences

ASP sequences retrieved from *Plasmodium* species were aligned using COBALT (Constraint-Based Multiple Alignment Tool) available through the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/tools/cobalt/). Default alignment parameters were used. The resulting alignment was examined for sequence conservation, variable regions, and insertion/deletion patterns among ASP orthologues within the genus *Plasmodium* (Fig. 1).

Codon-Based Selection Analysis of Plasmodium ASP Orthologues

To investigate selective pressures acting on apical sushi protein (ASP) within the genus *Plasmodium*, coding sequences corresponding to the ASP orthologues were retrieved from NCBI nucleotide records. Codon-aware multiple sequence alignment was performed in Geneious Prime 2026.1.2 using the Translation Align function. Sequences were translated using the standard genetic code and translation frame 1, and amino acid alignments were generated using the MAFFT alignment algorithm with the E-INS-i iterative refinement strategy, the BLOSUM62 substitution matrix, and a gap opening penalty of 3.0. The resulting protein-guided codon alignment was manually inspected to confirm preservation of the reading frame and the absence of internal stop codons.

Site-specific selection analyses were performed using the Fixed Effects Likelihood (FEL) method implemented in the Datamonkey Adaptive Evolution Server (http://www.datamonkey.org). FEL estimates synonymous (dS) and nonsynonymous (dN) substitution rates at individual codon sites using a maximum-likelihood framework and identifies codons evolving under pervasive purifying or diversifying selection. Analyses were performed using the universal genetic code with site-to-site synonymous rate variation enabled. Because the objective was to characterize selective pressures acting across the evolutionary history of *Plasmodium* ASP, all branches of the inferred phylogeny were included in the analysis. Codon sites with p values ≤ 0.1 were considered to exhibit statistically significant evidence of positive or purifying selection (Fig. 1).

Identification of ASP Homologues in Non-Plasmodial Apicomplexans

To identify ASP homologues in other apicomplexan lineages, the *Plasmodium falciparum* ASP sequence (XP_001351372.1) was used as a query in BLASTp searches against the NCBI non-redundant (nr) protein database (https://blast.ncbi.nlm.nih.gov/). Searches were restricted to the phylum Apicomplexa while excluding members of the genus *Plasmodium* (taxid:5820). Homologous proteins were selected based on sequence similarity, E-value significance, protein name (description), and taxonomic representation. Representative homologues from major apicomplexan groups, including coccidians and cryptosporidians, were selected for comparative analyses (Fig. 1).

Comparative Sequence Alignment, Representative Sequence Selection, and Domain Analysis

Protein sequences identified through BLASTp searches were imported into the NCBI Multiple Sequence Alignment (MSA) Viewer for comparative analysis. Short partial sequences and highly truncated proteins were excluded to improve comparability among taxa. The resulting alignment was visualized using the built-in Z-to-A sequence ordering option of the NCBI Multiple Sequence Alignment Viewer. Conserved regions, insertion/deletion patterns, and sequence divergence were examined across representative apicomplexan homologues.

Conserved domain architectures were characterized using the NCBI Batch Web CD-Search Tool (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) against the Conserved Domain Database (CDD). Domain assignments, superfamily classifications, E-values, bit scores, and positional coordinates were recorded for each representative protein. Particular attention was given to the distribution and positional coordinates of sushi domains, complement control protein (CCP) modules, and short consensus repeat (SCR) domains across different apicomplexan taxa (Fig. 1).

Alignment Optimization via Iterative Refinement

To improve the alignment of highly divergent homologues, representative protein sequences were re-aligned using the MAFFT multiple aligner Version 1.5.0 plugin implemented within Geneious Prime 2026.1.2 software (https://www.geneious.com). The E-INS-i iterative refinement strategy was employed using the BLOSUM62 scoring matrix. A gap opening penalty of 3.0 was applied during alignment optimization. The resulting alignment was used for subsequent comparative, phylogenetic, and structural analyses (Fig. 1).

Phylogenetic Analysis

Phylogenetic relationships among the 13 representative apicomplexan homologues were reconstructed in Geneious Prime 2026.1.2 using the optimized amino acid sequence alignment generated by MAFFT (E-INS-i). Representative taxa were selected to maximize phylogenetic breadth across major apicomplexan lineages while minimizing redundancy among highly conserved lineage-specific homologues. Maximum Likelihood (ML) trees were inferred using the PhyML plugin version 3.3.20180621. Analyses were performed under the WAG amino acid substitution model with Gamma-distributed rate variation among sites (G, 4 categories). The WAG + G model was selected as a widely used empirical amino acid substitution model that has been extensively applied in protein phylogenetic analyses [21–23].

All alignment sites were retained in the analysis without site masking or gap exclusion. Branch support was assessed using the Shimodaira-Hasegawa-like (SH-like) approximate likelihood-ratio test (aLRT). SH-like aLRT was selected because it provides a fast and robust likelihood-based estimate of local branch support and has been shown to perform as a powerful alternative to conventional nonparametric bootstrap methods, particularly in maximum-likelihood phylogenetic analyses of divergent datasets [24]. The resulting phylogenetic tree was visualized and rooted using the *Cryptosporidium* lineage as an outgroup (Fig. 1).

Three-Dimensional Protein Structure Prediction

Three-dimensional structures of representative ASP/RON1/SUSD homologues were predicted using AlphaFold3 (https://alphafoldserver.com) [25]. Representative proteins were selected to capture the major structural and domain architectural diversity observed across the ASP/RON1/SUSD protein family. Predicted models were evaluated using confidence metrics, including predicted local distance difference test (pLDDT) scores. Regions with low-confidence predictions were interpreted cautiously, particularly within potentially disordered or highly variable regions (Fig. 1).

Results

Identification of ASP Sequences in Plasmodium

Application of the search term “apical sushi protein complete” with Protists and RefSeq filters identified 13 full-length ASP sequences from members of the genus *Plasmodium*. The retrieved sequences represented parasites infecting humans, non-human primates, rodents, and birds and ranged from 672 to 767 amino acids in length (Fig. 2) (Table 1).

Table 1
Apical sushi protein (ASP) sequences retrieved from representative *Plasmodium* species and used for comparative analyses.

NCBI accession no	Species	Protein name	Amino acid length
XP_018643288.1	*Plasmodium gaboni*	apical sushi protein	738
XP_067075245.1	*Plasmodium brasilianum*	apical sushi protein	705
XP_001351372.1	**Plasmodium falciparum 3D7**	**apical sushi protein**	731
XP_028859379.1	*Plasmodium malariae*	apical sushi protein	708
XP_028541828.1	*Plasmodium gonderi*	apical sushi protein	762
XP_028536696.1	*Plasmodium adleri*	apical sushi protein	767
XP_019970771.1	*Plasmodium reichenowi*	apical sushi protein	743
XP_002261019.1	*Plasmodium knowlesi* strain H	apical sushi protein, putative	726
XP_028531557.1	**Plasmodium relictum**	**apical sushi protein, putative**	672
XP_028530928.1	*Plasmodium gallinaceum*	apical sushi protein, putative	675
XP_016655489.1	*Plasmodium chabaudi chabaudi*	apical sushi protein, putative	725
XP_037490347.1	*Plasmodium vinckei vinckei*	apical sushi protein, putative	712
XP_034421836.1	**Plasmodium berghei ANKA**	**apical sushi protein, putative**	720

Sequences selected for downstream comparative analyses are indicated in bold and correspond to the representative *Plasmodium* species included in the multiple sequence alignment shown in Fig. 2. Because ASP sequences exhibited extensive conservation across *Plasmodium* species, representative taxa infecting primate (*Plasmodium falciparum*), rodent (*Plasmodium berghei*), and avian (*Plasmodium relictum*) hosts were selected to maximize host-range representation while minimizing redundancy in downstream analyses.

ASP is Conserved across Plasmodium Species

Multiple sequence alignment generated using COBALT revealed extensive sequence conservation among ASP proteins across *Plasmodium* species (Fig. 2). Conserved regions were distributed throughout most of the protein sequence, whereas sequence variation was concentrated within several internal regions of the alignment. Despite variation in protein length, no major lineage-specific deletions were observed among the analysed *Plasmodium* sequences.

Strong Purifying Selection Constrains ASP Evolution in Plasmodium

Codon-based evolutionary analysis of *Plasmodium* ASP orthologues using the Fixed Effects Likelihood (FEL) framework demonstrated that ASP evolution is dominated by pervasive purifying selection. Analysis of 867 codon positions identified 223 sites evolving under significant negative selection and only five sites under significant diversifying positive selection ($p \leq 0.1$). The predominance and broad distribution of negatively selected codons across the ASP coding sequence are consistent with the extensive sequence conservation observed among ASP orthologues and indicate strong functional constraint throughout the evolutionary history of the protein. In contrast, the limited number of positively selected codons suggests that adaptive diversification has affected only a small subset of ASP residues during *Plasmodium* evolution (Supplementary Figs. 1 and 2; Table 2). Detailed site-specific FEL estimates are provided in Supplementary Table 1.

Table 2
Summary of codon-based selection analysis of *Plasmodium* ASP orthologues using the Fixed Effects Likelihood (FEL) method implemented in Datamonkey.

FEL analysis parameter	Result
Number of sequences	13
Codon positions analyzed	867
Non-invariant codons tested	773
Median branches tested per site	23
Codons under purifying selection ($p \leq 0.1$)	223
Codons under positive selection ($p \leq 0.1$)	5

A total of 867 codon positions from 13 ASP orthologues were analyzed. The results indicate a predominance of purifying selection across the ASP coding sequence, with only a limited number of codons exhibiting evidence of diversifying positive selection.

Identification of ASP Homologues in Non-Plasmodium Apicomplexans

BLASTp analysis using the *Plasmodium falciparum* ASP sequence (XP_001351372.1) identified homologous proteins in multiple non-plasmodial apicomplexan taxa (Table 3). Homologues were detected in coccidian parasites including *Toxoplasma gondii*, *Neospora caninum*, *Cyclospora cayetanensis*, *Cystoisospora suis*, and *Besnoitia besnoiti*, as well as in multiple *Cryptosporidium* species. Although sequence identities were generally low (between 20–40%), significant matches were consistently recovered across these lineages. Notably, homologous proteins were annotated under different names,

including rhoptry neck protein 1 (RON1), sushi-domain-containing proteins (SUSDs), and hypothetical proteins, despite displaying detectable similarity to *Plasmodium* ASP.

Table 3
ASP, rhoptry neck protein 1 (RON1), and sushi-domain-containing protein (SUSD) homologues identified in non-plasmodial apicomplexans.

NCBI accession no	Species	Protein name (description)	Length	BLAST hit	E-value	Identities
KFG40781.1	**Toxoplasma gondii p89**	rhoptry neck protein RON1, partial	415	2	1e-11	37/138(27%)
					0.005	22/59(37%)
CEL77302.1	*Toxoplasma gondii* VEG	TPA: sushi domain-containing protein / SCR repeat-containing protein	1131	2	3e-15	43/138(31%)
					3e-15	39/156(25%)
KFH02422.1	*Toxoplasma gondii* VAND	rhoptry neck protein RON1, partial	315	2	6e-15	51/166(31%)
					3e-14	37/132(28%)
KYK70095.1	*Toxoplasma gondii* TgCatPRC2	rhoptry neck protein RON1	1158	2	3e-15	50/166(30%)
					3e-15	39/156(25%)
KFG65751.1	*Toxoplasma gondii* RUB	rhoptry neck protein RON1, partial	315	2	8e-15	50/166(30%)
					3e-14	37/132(28%)
XP_002364247.2	**Toxoplasma gondii ME49**	rhoptry neck protein RON1	1158	2	3e-15	39/156(25%)
					3e-15	50/166(30%)
KFH13080.1	*Toxoplasma gondii* MAS	rhoptry neck protein RON1, partial	315	2	5e-15	51/166(31%)
					2e-14	37/132(28%)
EPR60880.1	*Toxoplasma gondii* GT1	rhoptry neck protein RON1	1158	2	3e-15	50/166(30%)
					3e-15	39/156(25%)
KFG44603.1	**Toxoplasma gondii GAB2-2007-GAL-DOM2**	rhoptry neck protein RON1	277	2	2e-14	37/132(28%)
					8e-07	41/156(26%)
KFG55774.1	*Toxoplasma gondii* FOU	rhoptry neck protein RON1, partial	426	2	7e-12	37/138(27%)
					0.005	21/59(36%)
PIL95883.1	*Toxoplasma gondii* COUG	rhoptry neck protein RON1, partial	472	2	2e-12	36/132(27%)
					0.002	21/59(36%)
RQX72368.1	*Toxoplasma gondii* CAST	rhoptry neck protein RON1	1122	2	3e-15	50/166(30%)
					4e-15	39/156(25%)

NCBI accession no	Species	Protein name (description)	Length	BLAST hit	E-value	Identities
KYF47557.1	*Toxoplasma gondii* ARI	rhoptry neck protein RON1, partial	503	2	2e-12	37/138(27%)
					0.003	21/59(36%)
KAF4639179.1	*Toxoplasma gondii*	rhoptry neck protein RON1	1158	2	3e-15	50/166(30%)
					3e-15	39/156(25%)
CEL69708.1	**Neospora caninum**	TPA: sushi domain-containing protein / SCR repeat-containing protein, putative	1133	2	3e-15	39/127(31%)
					9e-12	31/97(32%)
XP_067917733.1	**Cystoisospora suis**	rhoptry neck protein	1000	2	2e-14	40/131(31%)
					3e-06	30/148(20%)
OEH76101.1	**Cyclospora cayetanensis**	rhoptry neck protein RON1	1009	2	4e-15	36/145(25%)
					3e-04	34/103(33%)
KAK6590179.1	**Cryptosporidium xiaoi**	hypothetical protein RS030_172642	889	2	7e-07	35/112(31%)
					0.004	40/126(32%)
XP_028874935.1	*Cryptosporidium ubiquitum*	sushi domain-containing protein	1006	2	1e-06	33/103(32%)
					2e-04	28/79(35%)
TRY49946.1	*Cryptosporidium tyzzeri*	Sushi/SCR/CCP domain containing protein	996	2	2e-06	32/103(31%)
					1e-04	32/80(40%)
XP_067201224.1	*Cryptosporidium sp.* chipmunk genotype I	sushi-domain containing secreted with a signal peptide	1003	2	5e-07	34/103(33%)
					1e-04	32/80(40%)
XP_627153.1	**Cryptosporidium parvum Iowa II**	sushi-domain containing secreted protein; with a signal peptide, low complexity region followed by a sushi domain	1006	2	6e-06	31/103(30%)
					1e-04	32/80(40%)

NCBI accession no	Species	Protein name (description)	Length	BLAST hit	E-value	Identities
KAL5366988.1	*Cryptosporidium parvum*	sushi-domain containing secreted protein	998	2	6e-06	31/103(30%)
					1e-04	32/80(40%)
POM84918.1	*Cryptosporidium meleagridis*	Sushi domain (SCR repeat) family protein	997	2	2e-06	33/103(32%)
					3e-04	28/79(35%)
KAK9170937.1	*Cryptosporidium meleagridis*	Sushi/SCR/CCP domain containing protein	997	2	2e-06	33/103(32%)
					3e-04	28/79(35%)
XP_667945.1	**Cryptosporidium hominis TU502**	Sushi domain	972	2	2e-06	32/103(31%)
					1e-04	32/80(40%)
KAL3126476.1	*Cryptosporidium hominis*	sushi-domain containing secreted protein	971	2	2e-06	32/103(31%)
					1e-04	32/80(40%)
PPS98050.1	*Cryptosporidium hominis*	Sushi/SCR/CCP domain containing protein	971	2	2e-06	32/103(31%)
					1e-04	32/80(40%)
KAJ1606308.1	*Cryptosporidium canis*	sushi-domain containing secreted protein	945	2	3e-05	33/112(29%)
					0.001	30/87(34%)
XP_029218004.1	**Besnoitia besnoiti**	rhoptry neck protein RON1	991	2	4e-15	45/171(26%)
					4e-14	40/124(32%)

Sequences selected for downstream comparative analyses are indicated in bold and correspond to the representative homologues included in the multiple sequence alignment shown in Fig. 3.

Comparative Alignment Reveals Lineage-Specific Divergence Patterns

Multiple sequence alignment of representative ASP homologues across Apicomplexa generated using NCBI MSA Viewer revealed several conserved sequence blocks shared among all analysed taxa (Fig. 3). Relative to *Plasmodium* ASP sequences, homologues from *Toxoplasma gondii* exhibited extensive gap-rich regions within the alignment. Similar patterns were observed in *Neospora caninum*, *Cyclospora cayetanensis*, and *Cystoisospora suis*, whereas additional lineage-specific regions of divergence were present among certain *Toxoplasma gondii* homologues. *Cryptosporidium* homologues displayed the most extensive alignment discontinuities relative to the *Plasmodium* reference sequences. Despite substantial

sequence divergence, several conserved blocks remained detectable across all representative homologues (Fig. 3).

Conserved Domain Architecture and Boundary Mapping

Conserved domain analysis of the 13 representative apicomplexan homologues revealed substantial lineage-specific variation in domain architecture (Supplementary Fig. 3–5). In the haemosporidian lineage, represented by *Plasmodium falciparum*, *Plasmodium relictum*, and *Plasmodium berghei*, CCP/Sushi domains were consistently detected near the C-terminus (Supplementary Fig. 3).

In contrast, no CCP/Sushi domains were detected in the representative coccidian homologues analysed, including *Toxoplasma gondii*, *Neospora caninum*, *Cystoisospora suis*, *Cyclospora cayetanensis*, and *Besnoitia besnoiti*. Instead, these proteins contained alternative domain annotations and superfamily assignments, including the 2A1904 superfamily in *Toxoplasma gondii* (XP_002364247.2) and *Besnoitia besnoiti* (XP_029218004.1), and the PHA03169 domain in *N. caninum* (CEL69708.1) and *Besnoitia besnoiti* (XP_029218004.1) (Supplementary Fig. 3–5). Considerable variation in domain composition was also observed among the analysed *Toxoplasma gondii* homologues.

Within the cryptosporidian lineage, CCP domains were detected in *Cryptosporidium parvum* and *Cryptosporidium hominis*, where they were located near the C-terminus (Supplementary Fig. 5). These proteins also contained an upstream MSCRAMM_ClfA domain. However, this arrangement was not conserved across all cryptosporidian homologues, as *Cryptosporidium xiaoi* lacked a detectable CCP domain and instead exhibited a distinct domain configuration comprising PRK08581 and PTZ00108 superfamily annotations (Supplementary Fig. 5).

Refined Multiple Sequence Alignment and Conserved Sequence Blocks

Representative homologues were re-aligned using the MAFFT E-INS-i algorithm to improve alignment across highly divergent sequences (Fig. 4). Compared with the preliminary MSA Viewer alignment, the refined alignment exhibited a more coherent arrangement of conserved sequence blocks across the 13 representative taxa.

Several conserved blocks remained detectable throughout the alignment despite substantial differences in sequence length, domain architecture, and taxonomic affiliation. Although extensive gap-rich regions were still present in coccidian and cryptosporidian homologues relative to the *Plasmodium* ASP sequences, the refined alignment retained recognizable regions of similarity across all analysed taxa. Many of the large discontinuities observed in the preliminary alignment were reduced following iterative refinement, resulting in improved continuity of conserved sequence blocks throughout the dataset (Fig. 4).

Phylogenetic Reconstruction of ASP/RON1/SUSD Homologues

Maximum Likelihood phylogenetic analysis performed under the WAG + G model recovered three major clades corresponding to cryptosporidian, haemosporidian, and coccidian lineages (Fig. 5). The *Cryptosporidium* homologues formed the basal lineage, with *Cryptosporidium parvum* and *Cryptosporidium hominis* recovered as sister taxa and *Cryptosporidium xiaoi* occupying a more divergent position (SH-like = 0.991). The haemosporidian clade comprised the three *Plasmodium* ASP sequences and was strongly supported (SH-like = 1.00). Within this clade, *Plasmodium relictum* was recovered as a sister to the *Plasmodium berghei-Plasmodium falciparum* lineage. All representative coccidian homologues formed a well-supported monophyletic clade (SH-like = 0.995). Within this lineage, *Cyclospora cayetanensis* and *Cystoisospora suis* occupied the earliest branching positions, whereas *Besnoitia besnoiti*, *Neospora caninum*, and the *Toxoplasma gondii* homologues formed a more closely related subgroup (SH-like = 0.926). Overall, the phylogenetic relationships were broadly consistent with the sequence and domain architecture patterns identified in the comparative analyses.

Structural Conservation and Divergence of ASP Homologues

AlphaFold3 models were generated for representative ASP/RON1/SUSD homologues across Apicomplexa to evaluate patterns of structural conservation and divergence (Fig. 6). Predicted protein lengths ranged from 277 amino acids in a truncated *Toxoplasma gondii* RON1 homologue to 1158 amino acids in the full-length *Toxoplasma gondii* RON1 protein (Table 4). Predicted template modelling (pTM) scores were uniformly low (0.23–0.39), indicating limited confidence in global domain organization (Table 4). In contrast, several local structural elements, including α-helical and β-sheet-rich regions, were predicted with moderate to high confidence (Fig. 6).

Predicted disorder content varied substantially among homologues, ranging from 31% to 71% of residues (Table 4). The highest disorder fractions were observed in *Cryptosporidium parvum* (71%), *Cryptosporidium hominis* (70%), and *Plasmodium berghei* (68%), whereas the truncated *Toxoplasma gondii* homologue exhibited the lowest value (31%). Predicted aligned error (PAE) maps revealed multiple discrete regions of high local confidence separated by extensive regions of elevated uncertainty (Supplementary Fig. 6). Similar structural organization patterns were observed across ASP, RON1, and SUSD proteins despite substantial differences in sequence length and primary sequence composition (Fig. 6 and Supplementary Fig. 6).

Table 4. AlphaFold3 prediction statistics for representative ASP, RON1, and SUSD homologues across Apicomplexa.

Discussion

Molecular evolutionary analyses have become indispensable for understanding how protein families originate, diversify, and acquire new biological functions across deep evolutionary timescales [26–29]. Studies spanning diverse biological systems have demonstrated that proteins often retain conserved structural frameworks or functional cores despite extensive sequence divergence, lineage-specific adaptations, and substantial remodeling of domain architecture [26, 27, 29].

By integrating comparative sequence analyses, phylogenetic reconstruction, conserved domain characterization, and structural modelling, it becomes possible to distinguish ancestral molecular features from derived innovations and to reconstruct the evolutionary processes that shape protein family diversification [27–29]. These approaches provide valuable insights not only into evolutionary history but also into the mechanisms by which protein families diversify, adapt, and acquire lineage-specific functions. Importantly, they enable the recognition of deep evolutionary relationships that may remain obscured when sequence similarity alone is considered, particularly in protein families that have undergone substantial divergence, domain reorganization, and lineage-specific evolutionary remodeling [26, 28, 29].

Such approaches are particularly informative in Apicomplexa, where proteins involved in host-cell invasion are exposed to competing pressures of functional conservation, host adaptation, immune evasion, and ecological specialization [9]. Comparative genomic studies have further shown that many apicomplexan protein families have undergone extensive lineage-specific diversification and innovation, particularly among proteins associated with host invasion and parasite survival processes, highlighting the dynamic evolutionary history of host-parasite interaction systems [30]. These evolutionary dynamics can obscure homology relationships among deeply diverged proteins and complicate efforts to reconstruct the origins and diversification of invasion-associated protein families. Consequently, despite the central importance of the apical invasion machinery for parasite survival and infectivity, the evolutionary histories of many of its constituent proteins remain incompletely understood.

The present study represents the first broad comparative evolutionary investigation of the ASP/RON1/SUSD protein family across major apicomplexan lineages. Previous studies have focused primarily on ASP in *Plasmodium* species, where experimental evidence has linked the protein to apical organelle localization, rhoptry neck function, and host-cell invasion [14, 18, 19]. Moreover, ASP/RON1-related proteins have been detected in multiple invasive stages of *Plasmodium*, supporting their broader importance within the apicomplexan invasion machinery [17]. In contrast, homologous proteins identified in coccidian parasites have generally been annotated as RON1, whereas related proteins in *Cryptosporidium* species are frequently described as SUSDs or hypothetical proteins (Table 3), obscuring their broader evolutionary relationships. The combined evidence from BLASTp-based homology searches, optimized multiple sequence alignment, conserved domain analyses, phylogenetic reconstruction, and structural modelling suggests that these differently annotated proteins are unlikely to represent independent protein families. Instead, they are best interpreted as highly divergent members of a single ancestral protein family that predates the diversification of major apicomplexan lineages and subsequently underwent lineage-specific sequence divergence, domain remodeling, and annotation divergence while retaining detectable evolutionary continuity across extant apicomplexan taxa.

A notable finding of the present study was the extensive conservation observed among ASP homologues within the genus *Plasmodium* (Fig. 2) (Table 1). Despite representing parasites infecting humans, non-human primates, rodents, and birds, ASP sequences remained highly conserved throughout most of their length. Such conservation is consistent with previous studies implicating ASP in invasion-related processes and documenting its expression in multiple invasive parasite stages, suggesting an important functional role during host-cell recognition and invasion [14, 17]. The maintenance of this degree of

conservation across evolutionarily distant *Plasmodium* species indicates the presence of strong functional constraints acting on the protein. This observation is particularly noteworthy because proteins operating at the parasite-host interface are often among the most rapidly diverging components of parasite genomes due to host adaptation and immune-mediated diversification [9]. The conservation observed here therefore suggests that key ASP-associated functions have been maintained under persistent functional constraints throughout *Plasmodium* evolution, consistent with long-term purifying selection acting on biologically important features of the protein.

Furthermore, codon-based evolutionary analyses performed exclusively on *Plasmodium* ASP orthologues provided independent evidence supporting the strong evolutionary conservation observed at the sequence level. The Fixed Effects Likelihood (FEL) analysis identified pervasive purifying selection acting across the majority of codon positions, with 223 codons evolving under significant negative selection and only five codons exhibiting evidence of positive selection (Supplementary Fig. 1 and 2) (Table 2) (Supplementary Table 1). The predominance of negatively selected sites indicates that substantial functional constraints have operated throughout the evolutionary history of ASP within the genus *Plasmodium*. Such patterns are consistent with proteins that perform essential biological functions and tolerate only limited amino acid variation. The relatively small number of positively selected codons further suggests that adaptive diversification has played only a minor role in ASP evolution, reinforcing the interpretation that key structural and functional features of the protein have been maintained by long-term purifying selection.

The evolutionary significance of ASP becomes more apparent when homologues identified outside *Plasmodium* species are considered. Homologous proteins were detected in representatives of *Toxoplasma*, *Neospora*, *Cyclospora*, *Cystoisospora*, *Besnoitia*, and *Cryptosporidium* (Fig. 4–6) (Table 3), substantially expanding the known taxonomic distribution of this protein family. These findings indicate that ASP-related proteins are not restricted to malaria parasites and instead likely originated prior to the diversification of several major extant apicomplexan lineages. Such an interpretation is consistent with previous comparative genomic studies demonstrating that key components of host-cell invasion pathways can remain conserved across substantial evolutionary distances despite profound differences in host range, tissue tropism, life-cycle complexity, and transmission strategies [9, 10].

Similar patterns of deep conservation coupled with lineage-specific diversification have been reported across numerous apicomplexan protein families, where homologous proteins retain detectable evolutionary relationships despite substantial sequence divergence and differences in annotation among major taxonomic groups [30]. Consistent with these observations, phylogenetic reconstruction, optimized sequence alignments, and structural modelling recovered congruent signals supporting evolutionary continuity among hemosporidian *Plasmodium* ASP, coccidian RON1, and cryptosporidian SUSD proteins despite extensive sequence divergence. Collectively, these findings support the interpretation that these proteins represent divergent descendants of a common ancestral protein family that has undergone substantial lineage-specific diversification during apicomplexan evolution.

Initial comparative sequence analyses, based on the pre-alignment by NCBI MSA Viewer, suggested extensive insertion/deletion events among coccidian and cryptosporidian homologues relative to

Plasmodium ASP proteins (Fig. 3). However, following refinement of the representative sequence dataset using the MAFFT E-INS-i algorithm substantially altered this interpretation (Fig. 4). The optimized alignment retained coherent sequence block architecture across all analysed taxa despite considerable divergence in sequence length, domain composition, and annotation history.

This observation highlights an important methodological consideration when investigating deeply divergent protein families. Alignment strategies that do not adequately account for large insertions, deletions, or divergent domain architectures may exaggerate the extent of apparent sequence fragmentation among homologous proteins. The refined alignment therefore suggests that at least some of the insertion/deletion patterns inferred from preliminary analyses likely reflect alignment-dependent artefacts rather than extensive sequence loss. Such findings underscore the importance of using alignment methods optimized for highly divergent proteins when reconstructing the evolutionary history of ancient protein families. More broadly, they illustrate how methodological choices can influence evolutionary interpretations, particularly in protein families that have experienced substantial sequence divergence and lineage-specific structural remodeling.

The distribution of conserved domains across the ASP/RON1/SUSD family revealed a complex pattern of evolutionary remodeling (Supplementary Fig. 3–5). Canonical CCP/Sushi domains were consistently identified in *Plasmodium* ASP proteins and were also detected in selected cryptosporidian homologues. In contrast, equivalent domains were not detected in the analysed coccidian RON1 proteins despite their clear phylogenetic affinity to the broader ASP/RON1/SUSD family (Fig. 5; Supplementary Fig. 3–5). Therefore, this pattern may parsimoniously suggest lineage-specific loss of ancestral CCP/Sushi modules during coccidian evolution. However, alternative explanations like independently acquiring the same domains in plasmodial and cryptosporidial lineages must also be considered.

Importantly, the phylogenetic placement of these proteins within a strongly supported coccidian clade argues that the absence of detectable CCP/Sushi domains alone is insufficient to reject homology with the broader ASP/RON1/SUSD family. Furthermore, CCP/Sushi-containing proteins have previously been reported from coccidian parasites [9, 31], indicating that the pattern observed here is specific to the ASP/RON1/SUSD lineage rather than reflecting a general absence of CCP/Sushi domains from coccidian genomes. The inability of current domain prediction algorithms to identify CCP/Sushi domains within coccidian homologues may therefore reflect extensive sequence divergence beyond recognition thresholds, lineage-specific domain modification, or evolutionary loss.

Regardless of the precise mechanism, the combined evidence from domain architecture analyses, phylogenetic reconstruction, sequence conservation patterns, and structural modelling (Fig. 4–6; Supplementary Fig. 3–6) suggests that the ASP/RON1/SUSD family has experienced substantial lineage-specific remodeling during apicomplexan evolution, consistent with broader patterns of protein family diversification reported across the phylum [20, 30]. The differential distribution of CCP/Sushi domains among *Plasmodium*, coccidian, and cryptosporidian homologues therefore provides evidence that the evolutionary history of the ASP/RON1/SUSD family involved significant reorganization of domain architecture following the diversification of major apicomplexan lineages.

Additional evidence for architectural diversification was provided by the emergence of alternative domain architectures among non-haemosporidian homologues (Supplementary Fig. 3–5). Several coccidian proteins displayed superfamily assignments distinct from those observed in *Plasmodium* ASP proteins, whereas cryptosporidian homologues exhibited unique combinations of CCP and adhesion-associated (MSCRAMM_ClfA) domains. The occurrence of a bacterial adhesion-associated MSCRAMM_ClfA signature in cryptosporidian homologues is particularly intriguing and may reflect ancient horizontal acquisition, extensive lineage-specific domain remodeling, or convergent structural evolution, although additional phylogenetic analyses will be required to distinguish among these possibilities [32]. These observations suggest that the ASP/RON1/SUSD family has undergone repeated evolutionary modification following the diversification of major apicomplexan lineages while retaining sufficient sequence conservation to preserve detectable homology.

Such remodeling may reflect adaptation to the diverse invasion strategies, host associations, and life-history traits characteristic of different apicomplexan taxa [9, 30]. Although the functional consequences of these architectural differences remain unknown, the persistence of homologous proteins across evolutionarily distant lineages suggests that diversification occurred within the context of broadly conserved host-interaction and invasion-associated biological processes.

Phylogenetic reconstruction provided an independent line of evidence supporting the evolutionary history of the ASP/RON1/SUSD family (Fig. 5). Maximum Likelihood analyses recovered strongly supported clades corresponding closely to established apicomplexan taxonomy. *Plasmodium* ASP proteins formed a coherent haemosporidian cluster, coccidian RON1 homologues grouped together within a distinct monophyletic lineage, and cryptosporidian SUSDs occupied the deepest branching position within the analysed dataset. The recovery of *Cyclospora*, *Cystoisospora*, *Besnoitia*, *Neospora*, and *Toxoplasma* within a coherent coccidian lineage further supports the evolutionary continuity of these proteins despite their considerable biological, ecological, and annotation differences. Importantly, these phylogenetic relationships broadly mirrored the major sequence conservation and domain architecture patterns observed throughout the study.

The congruence between phylogenetic relationships, taxonomic classification, sequence conservation patterns, and domain architecture strongly supports the interpretation that ASP, RON1, and SUSDs represent homologous members of a single protein family that diversified following the evolutionary radiation of major apicomplexan lineages. Notably, the recovered topology is also broadly consistent with current apicomplexan taxonomy, providing additional support that the selected homologues capture genuine evolutionary relationships rather than artefacts arising from sequence similarity searches alone.

The AlphaFold3 predictions revealed a consistent architectural pattern across all analysed homologues, characterised by compact folded domains embedded within extensive regions of low-confidence structure that are consistent with intrinsic disorder (Fig. 6; Supplementary Fig. 6) (Table 4). Although global model confidence remained relatively low for several proteins, reflecting uncertainty in the relative positioning of distant regions, local structural confidence was consistently higher within the folded domains. Accordingly, predicted structures and associated confidence metrics were interpreted primarily as complementary

evidence supporting higher-order structural conservation rather than as definitive structural determinations. Such a combination of conserved structural cores and flexible low-confidence segments suggests that ASP/RON1/SUSD family proteins possess a modular structural organisation that may facilitate conformational flexibility and interaction versatility. Intrinsic disorder is frequently associated with transient protein-protein interactions, conformational adaptability, and multi-partner binding, properties that are highly compatible with evolutionary roles at the parasite-host interface and in complex invasion-associated biological processes [33, 34].

The recurrence of this overall architectural pattern across evolutionarily distant apicomplexan lineages despite substantial sequence divergence (Fig. 4 and 6; Supplementary Fig. 6) is consistent with the broader observation that protein families can retain conserved higher-order structural and organisational features even as primary sequences undergo extensive evolutionary diversification [26, 27, 29].

Taken together, the combined evidence derived from sequence conservation (Fig. 2), comparative alignments (Figures 3 and 4), conserved domain analyses (Supplementary Fig. 3–5), phylogenetic reconstruction (Fig. 5), and structural modelling (Fig. 6; Supplementary Fig. 6) supports a model in which ASP, RON1, and SUSD proteins represent divergent descendants of an ancient apicomplexan protein family. Accordingly, evolutionary inference was based on the convergence of multiple independent lines of evidence rather than on pairwise sequence identity alone. Although substantial lineage-specific sequence divergence, domain remodeling, and architectural diversification have occurred during evolution, particularly among coccidian and cryptosporidian lineages, multiple independent lines of evidence support a shared evolutionary origin.

The observed distribution of CCP/Sushi domains further suggests that the evolutionary history of this family involved differential domain retention, modification, and/or loss across major apicomplexan lineages. Nevertheless, experimental characterization of non-plasmodial homologues will be required to determine whether evolutionary conservation is accompanied by conservation of biological function.

Nevertheless, codon-based selection analyses could only be performed reliably for *Plasmodium* ASP orthologues because their high sequence conservation permitted the generation of robust codon alignments. Equivalent dN/dS-based analyses were not extended to the broader ASP/RON1/SUSD family because the extensive sequence divergence, lineage-specific insertion/deletion patterns, and substantial domain architectural remodeling observed across apicomplexan lineages precluded reliable codon-based evolutionary inference.

By integrating evidence from sequence, domain, phylogenetic, and structural analyses, the present study provides a unified evolutionary framework for understanding ASP/RON1/SUSD family molecular evolution and establishes a foundation for future investigations into the evolutionary diversification, domain remodeling, and functional evolution of this ancient apicomplexan protein family.

Conclusion

This study provides the first comprehensive evolutionary analysis of the ASP/RON1/SUSD protein family across major apicomplexan lineages and demonstrates that *Plasmodium* ASP, coccidian RON1, and cryptosporidian SUSD proteins are best interpreted as divergent descendants of a single ancient protein family. Despite extensive sequence divergence, differences in annotation, and lineage-specific remodeling of domain architecture, multiple independent lines of evidence from sequence, domain, phylogenetic, and structural analyses support a shared evolutionary origin. Furthermore, codon-based selection analyses demonstrated that *Plasmodium* ASP orthologues evolve predominantly under strong purifying selection, indicating substantial functional constraint throughout their evolutionary history. The differential distribution of CCP/Sushi domains among major apicomplexan lineages further suggests a complex history of domain retention, modification, and/or loss during family diversification. Collectively, these findings establish a unified evolutionary framework for understanding ASP/RON1/SUSD family molecular evolution and provide a foundation for future studies investigating the functional and evolutionary significance of this ancient invasion-associated protein family across Apicomplexa.

Declarations

Author Contributions A.E.K.: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original draft, Writing – review & editing, Supervision.

Conflict of Interest The authors declare no competing interests.

Acknowledgements The graphical abstract was created using items from BioRender (https://www.biorender.com).

Funding This research was not funded.

Data Availability Supplementary Table 1. Site-by-site results of the Fixed Effects Likelihood (FEL) analysis of selective pressures acting on *Plasmodium* ASP orthologues. For each codon position, synonymous substitution rate (α), nonsynonymous substitution rate (β), approximate likelihood-ratio test (aLRT) statistic, p-value, total branch length, and inferred selection class are reported. Supplementary Fig. 1–6. Supplementary figures supporting the codon-based selection, conserved domain, and structural analyses of ASP/RON1/SUSD homologues across Apicomplexa.

References

1. Rueckert S, Betts EL, Tsaousis AD (2019) The symbiotic spectrum: where do the gregarines fit? Trends Parasitol 35:687–694. https://doi.org/10.1016/j.pt.2019.06.013
2. Can H, Erkunt Alak S, Köseoğlu AE et al (2020) Do *Toxoplasma gondii* apicoplast proteins have antigenic potential? An *in silico* study. Comput Biol Chem 84:107158. https://doi.org/10.1016/j.compbiolchem.2019.107158
3. Köseoğlu AE, Can H, Karakavuk M et al (2022) Molecular prevalence and subtyping of *Cryptosporidium* spp. in fecal samples collected from stray cats in İzmir, Turkey. BMC Vet Res 18:89.

https://doi.org/10.1186/s12917-022-03190-y

4. Cavalier-Smith T (1993) Kingdom protozoa and its 18 phyla. Microbiol Rev 57:953–994. https://doi.org/10.1128/mr.57.4.953-994.1993
5. Adl SM, Leander BS, Simpson AG et al (2007) Diversity, nomenclature, and taxonomy of protists. Syst Biol 56:684–689. https://doi.org/10.1080/10635150701494127
6. Feix AS, Cruz-Bustos T, Ruttkowski B et al (2020) Characterization of *Cystoisospora suis* sexual stages *in vitro*. Parasit Vectors 13:143. https://doi.org/10.1186/s13071-020-04014-4
7. Dubey JP (2021) Outbreaks of clinical toxoplasmosis in humans: five decades of personal experience, perspectives and lessons learned. Parasit Vectors 14:263. https://doi.org/10.1186/s13071-021-04769-4
8. Morrissette N, Abbaali I, Ramakrishnan C et al (2023) The tubulin superfamily in apicomplexan parasites. Microorganisms 11:706. https://doi.org/10.3390/microorganisms11030706
9. Templeton TJ (2007) Whole-genome natural histories of apicomplexan surface proteins. Trends Parasitol 23:205–212. https://doi.org/10.1016/j.pt.2007.03.001
10. Templeton TJ, Iyer LM, Anantharaman V et al (2004) Comparative analysis of apicomplexa and genomic diversity in eukaryotes. Genome Res 14:1686–1695. https://doi.org/10.1101/gr.2615304
11. Gubbels MJ, Duraisingh MT (2012) Evolution of apicomplexan secretory organelles. Int J Parasitol 42:1071–1081. https://doi.org/10.1016/j.ijpara.2012.09.009
12. Counihan NA, Kalanon M, Coppel RL et al (2013) *Plasmodium* rhoptry proteins: why order is important. Trends Parasitol 29:228–236. https://doi.org/10.1016/j.pt.2013.03.003
13. Liffner B, Balbin JM, Wichers JS et al (2021) The ins and outs of *Plasmodium* rhoptries, focusing on the cytosolic side. Trends Parasitol 37:638–650. https://doi.org/10.1016/j.pt.2021.03.006
14. O'Keeffe AH, Green JL, Grainger M et al (2005) A novel Sushi domain-containing protein of *Plasmodium falciparum*. Mol Biochem Parasitol 140:61–68. https://doi.org/10.1016/j.molbiopara.2004.12.003
15. Muta T, Miyata T, Misumi Y et al (1991) *Limulus* factor C. An endotoxin-sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like, and lectin-like domains. J Biol Chem 266:6554–6561
16. Vanegas M, Bermúdez A, Guerrero YA et al (2014) Protecting capacity against malaria of chemically defined tetramer forms based on the *Plasmodium falciparum* apical sushi protein as potential vaccine components. Biochem Biophys Res Commun 451:15–23. https://doi.org/10.1016/j.bbrc.2014.06.143
17. Tokunaga N, Nozaki M, Tachibana M et al (2019) Expression and localization profiles of rhoptry proteins in *Plasmodium berghei* sporozoites. Front Cell Infect Microbiol 9:316. https://doi.org/10.3389/fcimb.2019.00316
18. Srivastava A, Singh S, Dhawan S et al (2010) Localization of apical sushi protein in *Plasmodium falciparum* merozoites. Mol Biochem Parasitol 174:66–69. https://doi.org/10.1016/j.molbiopara.2010.06.003

19. Zuccala ES, Gout AM, Dekiwadia C et al (2012) Subcompartmentalisation of proteins in the rhoptries correlates with ordered events of erythrocyte invasion by the blood stage malaria parasite. PLoS ONE 7:e46160. https://doi.org/10.1371/journal.pone.0046160
20. Swapna LS, Parkinson J (2017) Genomics of apicomplexan parasites. Crit Rev Biochem Mol Biol 52:254–273. https://doi.org/10.1080/10409238.2017.1290043
21. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851
22. Arenas M, Sánchez-Cobos A, Bastolla U (2015) Maximum-likelihood phylogenetic inference with selection on protein folding stability. Mol Biol Evol 32:2195–2207. https://doi.org/10.1093/molbev/msv085
23. Del Amparo R, Arenas M (2023) Influence of substitution model selection on protein phylogenetic tree reconstruction. Gene 865:147336. https://doi.org/10.1016/j.gene.2023.147336
24. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55:539–552. https://doi.org/10.1080/10635150600755453
25. Abramson J, Adler J, Dunger J et al (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630:493–500. https://doi.org/10.1038/s41586-024-07487-w
26. Mukherjee K, Moroz LL (2024) Evolution of g-type lysozymes in metazoa: insights into immunity and digestive adaptations. Front Cell Dev Biol 12:1487920. https://doi.org/10.3389/fcell.2024.1487920
27. Köseoğlu AE, Deniz Köseoğlu G, Özgültekin B et al (2025) Tracing the molecular evolution of Foxp2 proteins in vertebrates from fish to tetrapods: Insights into poly-Q tract variation, structural changes, and interaction networks. Comput Biol Chem 121:108840. https://doi.org/10.1016/j.compbiolchem.2025.108840
28. Ito T, Guenther C, Ishikawa E et al (2025) Phylogenetic and structural insights into the origin of C-type lectin Mincle in vertebrates. Immunogenetics 77:18. https://doi.org/10.1007/s00251-025-01375-x
29. Chatterjee A, Bhattacharya S, Venkatasubramanian A et al (2026) Tracing the molecular evolution of essential GTPase Obg CgtA reveals a direct association of its C-terminal domain with the diversity of bacterial phyla. Comput Biol Chem 123:109038. https://doi.org/10.1016/j.compbiolchem.2026.109038
30. Wasmuth J, Daub J, Peregrín-Alvarez JM et al (2009) The origins of apicomplexan sequence innovation. Genome Res 19:1202–1213. https://doi.org/10.1101/gr.083386.108
31. Templeton TJ, Pain A (2016) Diversity of extracellular proteins during the transition from the 'proto-apicomplexan'alveolates to the apicomplexan obligate parasites. Parasitology 143:1–17. https://doi.org/10.1017/S0031182015001213
32. Ganesh VK, Rivera JJ, Smeds E et al (2008) A structural model of the *Staphylococcus aureus* ClfA–fibrinogen interaction opens new avenues for the design of anti-staphylococcal therapeutics. PLoS Pathog 4:e1000226. https://doi.org/10.1371/journal.ppat.1000226
33. Blundell TL, Gupta MN, Hasnain SE (2020) Intrinsic disorder in proteins: Relevance to protein assemblies, drug design and host-pathogen interactions. Prog Biophys Mol Biol 156:34–42.

https://doi.org/10.1016/j.pbiomolbio.2020.06.004
34. Trivedi R, Nagarajaram HA (2022) Intrinsically disordered proteins: an overview. Int J Mol Sci 23:14050. https://doi.org/10.3390/ijms232214050

Table 4

Table 4 is available in the Supplementary Files section.

Figures

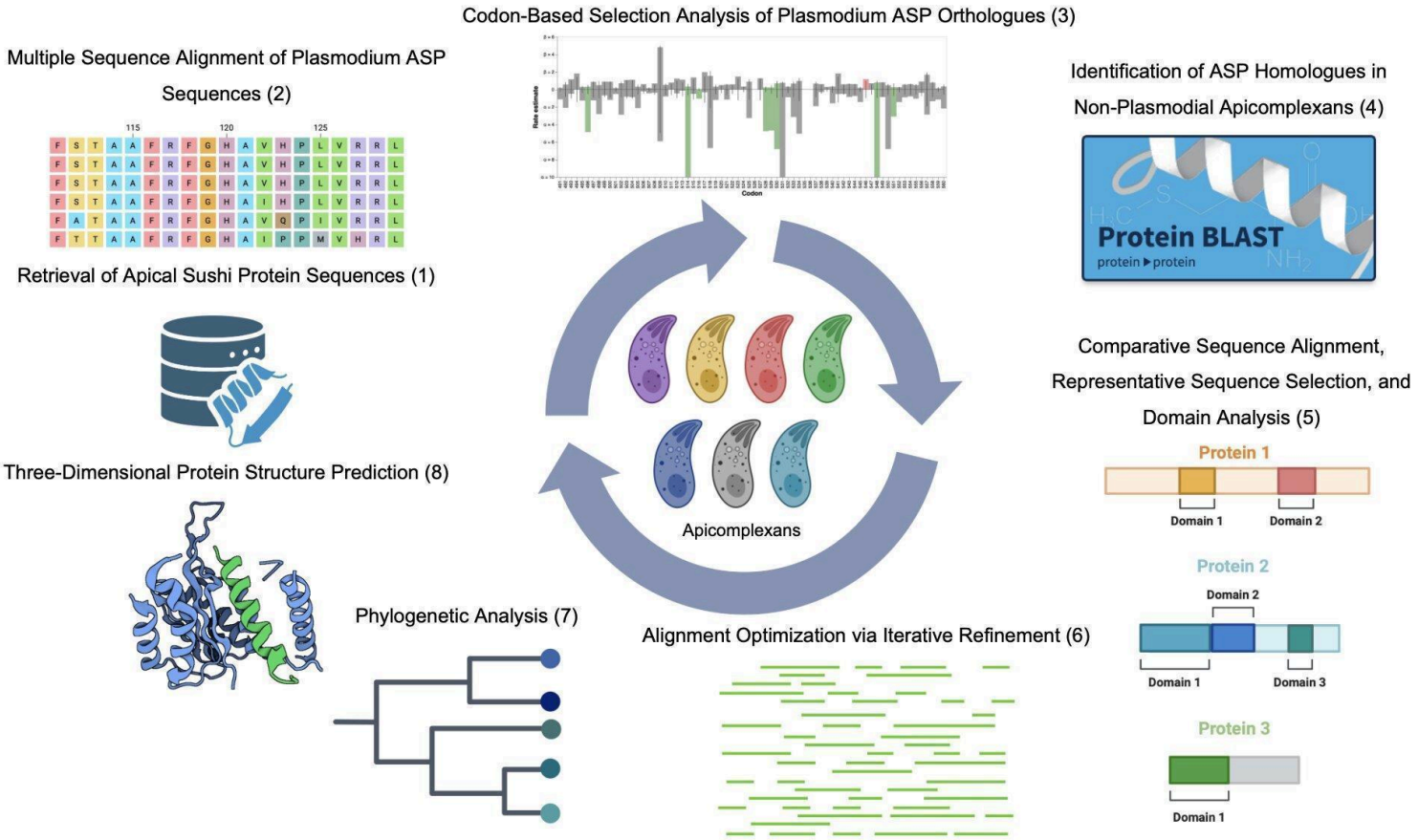


Figure 1

Analytical workflow used to investigate the evolutionary relationships and structural diversification of the ASP/RON1/SUSD protein family across Apicomplexa. (1) Apical sushi protein (ASP) sequences were retrieved from the NCBI Protein database using curated RefSeq records. (2) ASP orthologues from *Plasmodium* species were aligned to identify conserved sequence features and establish a reference framework. (3) Codon-based selection analysis was performed on *Plasmodium* ASP orthologues to characterize site-specific selective pressures. (4) Homologues in non-plasmodial apicomplexans were identified by BLASTp searches using the *Plasmodiumfalciparum* ASP sequence as the query. (5) Representative homologues were selected for comparative sequence and conserved domain analyses. (6) Multiple sequence alignments were optimized using iterative refinement implemented in MAFFT (E-INS-i)

to improve homology assessment among highly divergent sequences. (7) The optimized alignment was used for maximum-likelihood phylogenetic reconstruction. (8) Representative proteins were subjected to AlphaFold3-based structural prediction to evaluate the conservation and diversification of protein architecture across the ASP/RON1/SUSD family. The figure was created using items from BioRender (https://www.biorender.com).

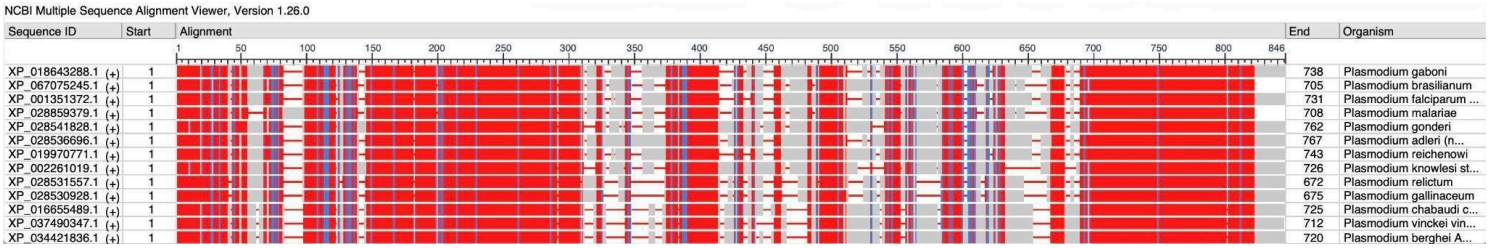


Figure 2

Conservation of apical sushi protein (ASP) orthologues across *Plasmodium* species. COBALT multiple sequence alignment of 13 full-length ASP proteins retrieved from the NCBI RefSeq database. Extensive sequence conservation is observed across most of the protein despite the inclusion of parasites infecting humans, non-human primates, rodents, and birds. Sequence variation is largely restricted to several internal regions, whereas no major lineage-specific deletion patterns are evident. Protein lengths range from 672 to 767 amino acids.

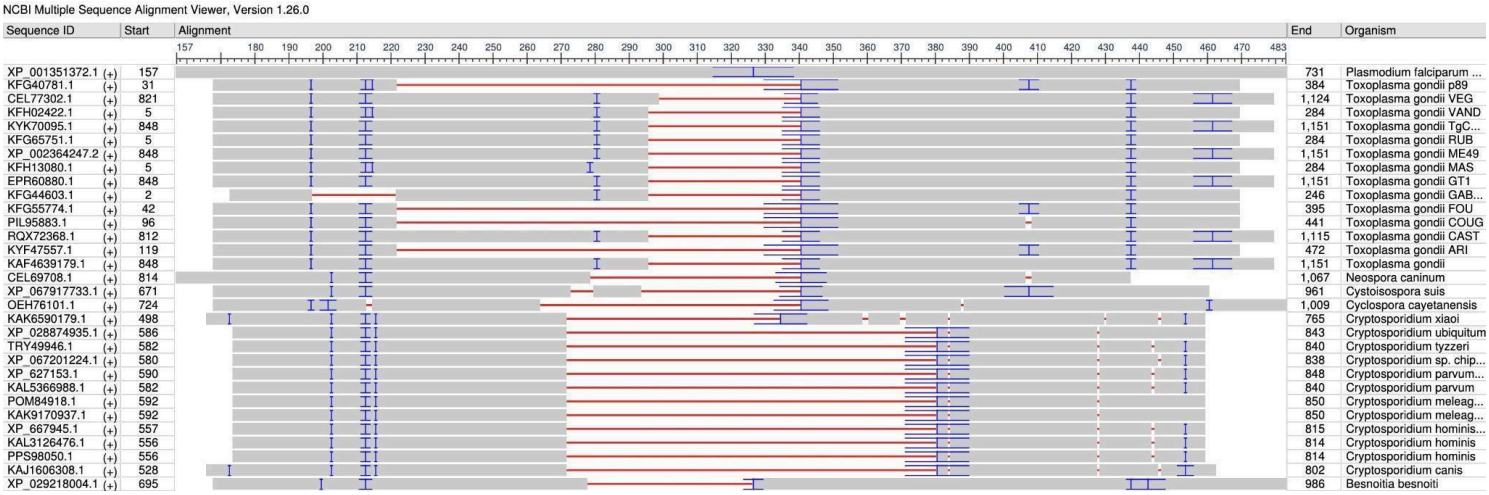


Figure 3

Comparative alignment of ASP/RON1/SUSD proteins across Apicomplexa. Multiple sequence alignment of representative ASP-related homologues identified using BLASTp searches with *Plasmodium falciparum* ASP as the query sequence. Proteins included representatives annotated as ASP, RON1, and SUSDs. Conserved sequence blocks remain detectable across all analysed apicomplexan lineages despite extensive lineage-specific insertion/deletion events. The largest reductions are observed among cryptosporidian homologues, whereas intermediate deletion patterns characterize coccidian taxa.

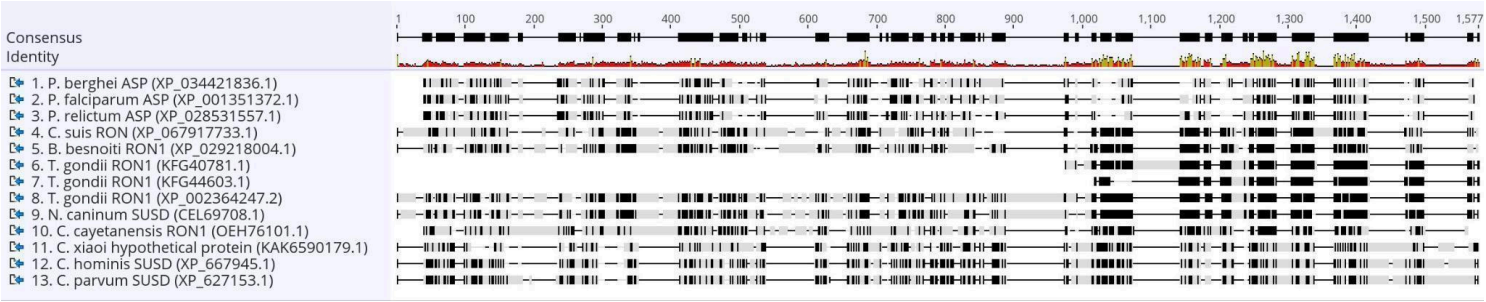


Figure 4

Multiple sequence alignment of representative ASP, RON1, and SUSD homologues across Apicomplexa. Conserved and variable regions across the aligned sequences are displayed using the NCBI Multiple Sequence Alignment Viewer. The alignment reveals conserved sequence blocks interspersed with lineage-specific insertions, deletions, and low-complexity regions. The consensus identity profile is shown above the alignment.

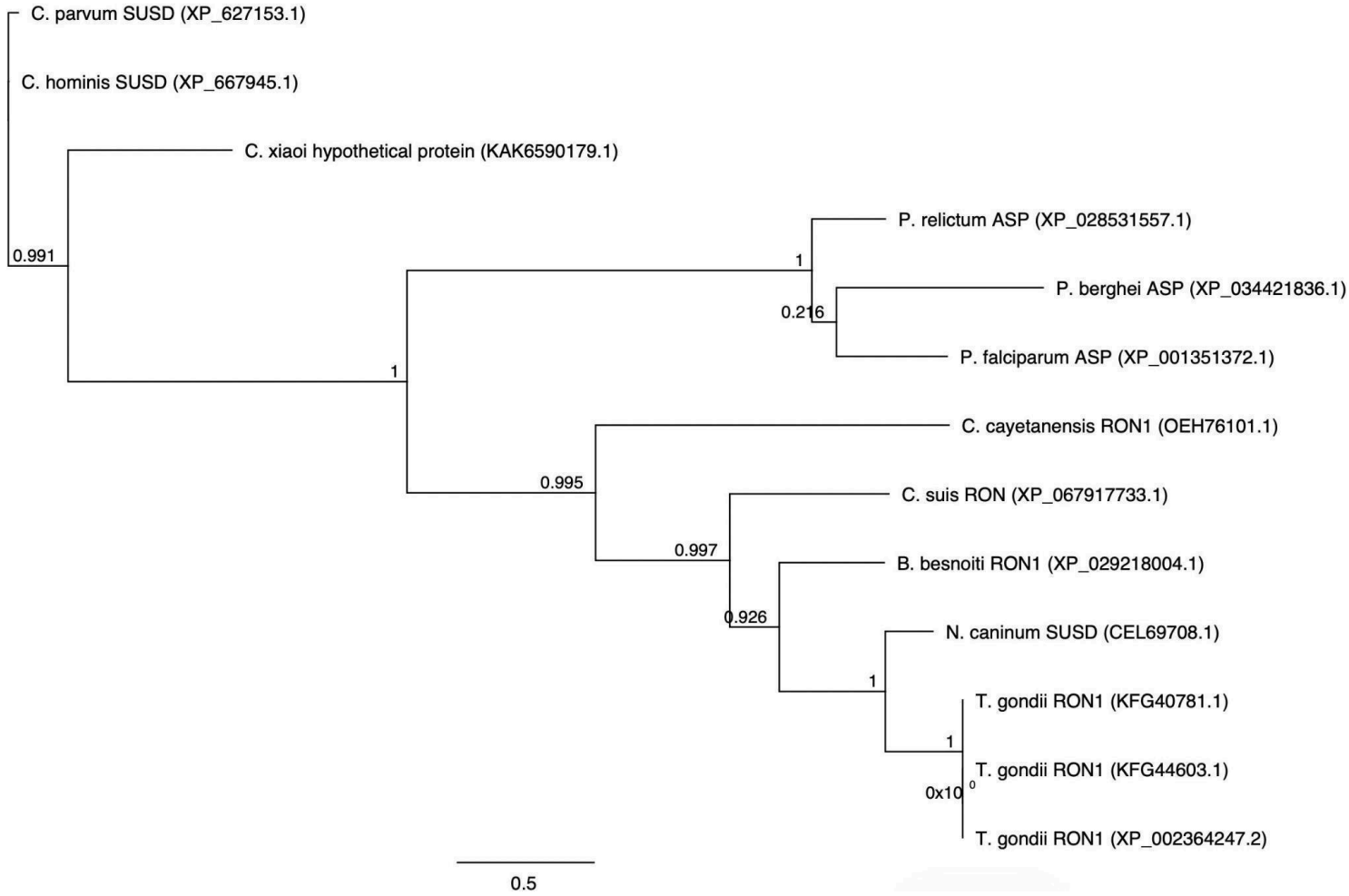


Figure 5

Maximum Likelihood phylogenetic relationships among representative ASP, RON1, and SUSD homologues from Apicomplexa. The phylogenetic tree was reconstructed using PhyML under the WAG+G amino acid

substitution model. Node values indicate Shimodaira-Hasegawa-like approximate likelihood-ratio test (SH-like aLRT) support. Three major clades corresponding to *Plasmodium* ASP proteins, coccidian RON1 homologues, and cryptosporidian SUSDs were recovered. The tree was rooted using the *Cryptosporidium*lineage as an outgroup. Scale bar indicates amino acid substitutions per site.

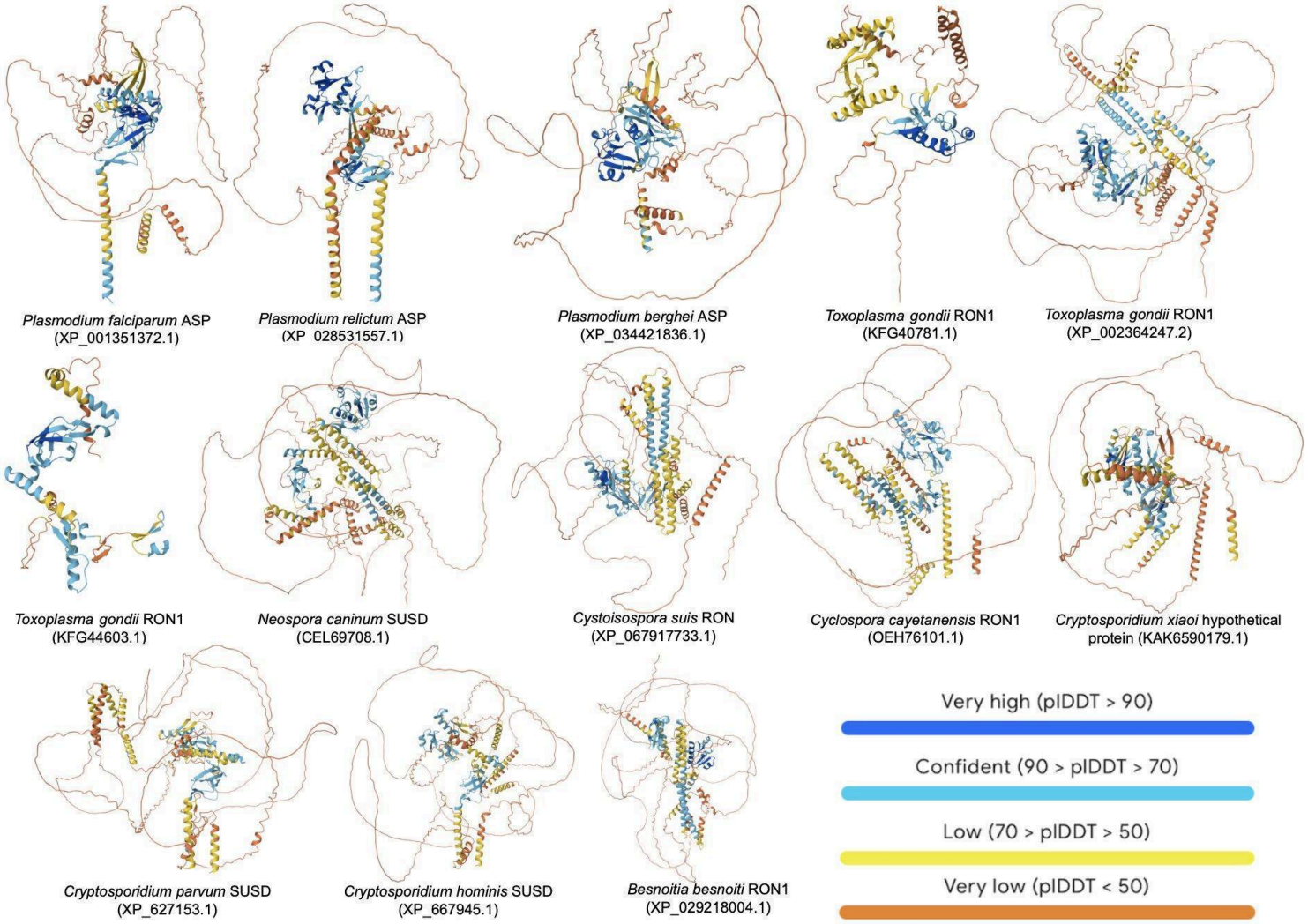


Figure 6

AlphaFold3 structural models of ASP/RON1/SUSD homologues identified across Apicomplexa. Predicted three-dimensional structures were generated using AlphaFold3 and visualized using the per-residue confidence colouring scheme (blue, high confidence; cyan/yellow, intermediate confidence; orange, low confidence). Across all homologues, compact folded regions composed of α-helices and β-sheets were observed together with extensive low-confidence regions corresponding to predicted intrinsically disordered segments. Similar overall structural organization patterns were observed despite substantial sequence divergence and variation in protein length.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryFig.16.docx
- Table4.docx
- GraphicalAbstract.jpg