The Apicomplexa-specific glucosamine-6-phosphate N-acetyltransferase gene family encodes a key enzyme for glycoconjugate synthesis with potential as therapeutic target

Marta Cova1, Borja López-Gutiérrez2, Sara Artigas-Jerónimo1, Aida González-Díaz1, Giulia Bandini2, Steven Maere1,3,4,5, Lorenzo Carretero-Paulet3,4,5 & Luis Izquierdo1

Apicomplexa form a phylum of obligate parasitic protozoa of great clinical and veterinary importance. These parasites synthesize glycoconjugates for their survival and infectivity, but the enzymatic steps required to generate the glycosylation precursors are not completely characterized. In particular, glucosamine-phosphate N-acetyltransferase (GNA1) activity, needed to produce the essential UDP-N-acetylglucosamine (UDP-GlcNAc) donor, has not been identified in any Apicomplexa. We scanned the genomes of Plasmodium falciparum and representatives from six additional main lineages of the phylum for proteins containing the Gcn5-related N-acetyltransferase (GNAT) domain. One family of GNAT-domain containing proteins, composed by a P. falciparum sequence and its six apicomplexan orthologs, rescued the growth of a yeast temperature-sensitive GNA1 mutant. Heterologous expression and in vitro assays confirmed the GNA1 enzymatic activity in all lineages. Sequence, phylogenetic and synteny analyses suggest an independent origin of the Apicomplexa-specific GNA1 family, parallel to the evolution of a different GNA1 family in other eukaryotes. The inability to disrupt an otherwise modifiable gene target suggests that the enzyme is essential for P. falciparum growth. The relevance of UDP-GlcNAc for parasite viability, together with the independent evolution and unique sequence features of Apicomplexa GNA1, highlights the potential of this enzyme as a selective therapeutic target against apicomplexans.

Apicomplexa form one of the largest and most diverse phyla of obligate intracellular parasites in the protist kingdom. They exhibit a fascinating biology, featuring (i) a specialized set of microtubules at the apical end of the cell, (ii) plastid-derived organelle called apicoplast, lost in some lineages, and (iii) a wide variety of morphologies and complex life cycles, consistent with the ability to infect almost every kind of animal from mollusks to mammals1. Although it has been estimated that between 1.2 and 10 million species exist2, only about 5,000–6,000 have been identified to date1, including species of clinical importance. Plasmodium species, for example, cause malaria, a disease responsible for more than 200 million new cases and 445,000 deaths in 20163. Toxoplasmosis, caused by the ubiquitous human pathogen Toxoplasma gondii, contributes to congenital disease and opportunistic infections in immunocompromised persons4. Cryptosporidium parvum is responsible for cryptosporidiosis.

1ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clinic – Universitat de Barcelona, Barcelona, Spain.
2Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, Boston, USA.
3Ghent University, Department of Plant Biotechnology and Bioinformatics, B-9052, Ghent, Belgium. 4VIB Center for Plant Systems Biology, B-9052, Ghent, Belgium. 5Bioinformatics Institute Ghent, Ghent University, B-9052, Ghent, Belgium. Borja López-Gutiérrez and Sara Artigas-Jerónimo contributed equally to this work. Correspondence and requests for materials should be addressed to L.C.-P. (email: locar@psb.ugent.be) or L.I. (email: luis.izquierdo@isglobal.org)
such as the Acyl-CoA in these clades matched INTERPRO protein functional domains generically associated to the GNAT domain, architecture within each of these 9 clades appeared relatively well-conserved. Many of the non-GNAT domains

Despite the overall high degree of divergence in terms of sequence composition and length, the protein domain may lead to unexpected clustering in the tree37.

T. gondii orthogroups, clustering 23 out of 71 sequences, were not properly resolved in our phylogenetic analyses. A possible reason for this may be related to the unusually long branches recovered for some P. falciparum orthogroups plus one singleton in the EupathDB database (http://eupathdb.org/eupathdb/, Supplementary Tables S1 and S2). In order to reconstruct the evolutionary relationships between these sequences, we performed phylogenetic analyses using three alternative methods (see Methods). 9 out of the 18 orthogroups, encompassing 48 out of the 71 sequences, were retrieved by all three methods as well-supported clades in the tree (Fig. 2).

Tables S1 and S2). In order to reconstruct the evolutionary relationships between these sequences, we performed phylogenetic analyses using three alternative methods (see Methods). 9 out of the 18 orthogroups, encompassing 48 out of the 71 sequences, were retrieved by all three methods as well-supported clades in the tree (Fig. 2).

The independent evolution and unique sequence features of this enzyme, which we show is likely essential for parasite growth, highlight GNA1 as a potential pan-apicomplexan drug target amenable to selective inhibition.

In summary, the identification of GNA1 in the genome of T. gondii, or any other apicomplexan, has remained elusive29 despite the fact that the presence of UDP-GlcNAc was verified in different parasite stages 34–36. However, the identification of GNA1 in the genome of P. falciparum, or any other apicomplexan, has remained elusive29 despite the fact that the presence of UDP-GlcNAc was verified in different parasite stages 34–36. Here we identify GNA1 in P. falciparum and six other species representing main lineages of the phylum Apicomplexa. GNA1 forms a specific gene family with an evolutionary origin at least as ancient as the phylum. The independent evolution and unique sequence features of this enzyme, which we show is likely essential for parasite growth, highlight GNA1 as a potential pan-apicomplexan drug target amenable to selective inhibition.

Results

Classification of the superfamily of GNAT-domain containing proteins in Apicomplexa. To identify apicomplexan proteins presenting a GNAT domain, we searched the genomes of P. falciparum and 6 additional species, representing the main lineages of the phylum, for sequences containing the INTERPRO protein functional domain (GNAT domain). 71 protein sequences were detected, belonging to 18 orthogroups plus one singleton in the EupathDB database (http://eupathdb.org/eupathdb/, Supplementary Tables S1 and S2). In order to reconstruct the evolutionary relationships between these sequences, we performed phylogenetic analyses using three alternative methods (see Methods). 9 out of the 18 orthogroups, encompassing 48 out of the 71 sequences, were retrieved by all three methods as well-supported clades in the tree (Fig. 2). Despite the overall high degree of divergence in terms of sequence composition and length, the protein domain architecture within each of these 9 clades appeared relatively well-conserved. Many of the non-GNAT domains in these clades matched INTERPRO protein functional domains generically associated to the GNAT domain, such as the Acyl-CoA N-acetyltransferase domain, or associated to specific families of GNAT proteins, such as N-myristoyltransferases (NMT) or histone acetyltransferases (HAT) (Supplementary Table S3). The remaining 9 orthogroups, clustering 23 out of 71 sequences, were not properly resolved in our phylogenetic analyses. A possible reason for this may be related to the unusually long branches recovered for some T. gondii sequences, which may lead to unexpected clustering in the tree37. T. gondii showed 15 GNAT-domain containing proteins compared
to 9–11 in the remaining species, likely resulting from lineage-specific gene duplications (Fig. 2; Supplementary Table S1).

**Identification of the GNA1 family in Apicomplexa.** In yeast and other eukaryotic organisms, the glucosamine-6-phosphate acetyltransferase activity is known to be essential and encoded by GNA1, a specific family of GNAT proteins. The evolutionarily conserved GNA1 gene family typically encodes proteins of 150–200 aa length and only displays the GNAT and/or Acyl-CoA N-acyltransferase domains. Out of the ten *P. falciparum* GNAT sequences, five were discarded as candidate GNA1s as they also contained protein domains corresponding to specific families of GNAT proteins other than GNA1. The remaining five were selected as candidate sequences and assayed for their putative GNA1 activity.

The five selected *P. falciparum* candidate sequences and *Trypanosoma brucei* GNA1 (*TbGNA1*) were cloned into the pRS421 pt plasmid and used to transform the *S. cerevisiae* GNA1 thermosensitive (ScGNA1-ts) mutant. Only PF3D7_0629000 and *TbGNA1* could support yeast growth in the absence of endogenous GNA1 at 37 °C (Fig. 3A), and we consequently renamed PF3D7_0629000 as *PfGNA1*. Our phylogenetic analyses (Fig. 2) clustered *PfGNA1* together with six additional sequences within a robustly supported clade corresponding to

---

**Figure 2.** Phylogeny, protein domain architecture and taxonomic distribution of 71 GNAT proteins from seven apicomplexan species. The maximum-likelihood unrooted phylogenetic tree is drawn to scale, with branch lengths proportional to evolutionary distances between nodes. Branches leading to extant nodes in the tree are colored according to the species. Indicated in bold are sequences examined for GNA1 activity. Values next to the orthogroup-defining nodes indicate statistical support from maximum-likelihood, Bayesian and neighbor-joining phylogenetic analysis, respectively. For EupathDB orthogroups retrieved as well-supported clades in the tree, colored strips indicate their taxonomic distribution (top right legend: Api, Apicomplexa; Euk, Eukaryota; Eug, Euglenozoa; Ubi, ubiquitous; Bab, Babesia; The, Theileria; Fun, Fungi; Tox, Toxoplasma; Eim, Eimeria). The apicomplexan-specific GNA1 family is enclosed within a continuous line box, while the HAT and the NMT families are enclosed within a dashed line box. A schematic protein architecture delineating the occurrence of conserved motifs detected using MEME is shown next to each protein (see Table S2). The distance between vertical lines on the right is 200 amino acids.
orthogroup OG5_147324. The predicted orthologs to PfGNA1 from Babesia bovis, C. parvum, Eimeria tenella, T. gondii, Theileria annulata, and G. niphandrodes were also cloned in pRS421 pt and functional complementation assays were performed. All orthologous sequences were able to rescue the ScGNA1-ts mutant at 37 °C (Fig. 3B).

Independent evolutionary origin of the Apicomplexa-specific GNA1 family. According to the OrthoMCLDB database, of all orthogroups grouping Apicomplexa GNAT proteins, only OG5_144531 and OG5_147324 (Apicomplexa GNA1 family) showed a taxonomic distribution restricted to apicomplexan organisms (Supplementary Table S2), suggesting their independent evolutionary origin within the phylum. In order to examine whether this represents an unusual feature of the Apicomplexa GNA1 family, we used the information provided by OrthoMCLDB to look at the taxonomic distribution of 30 P. falciparum enzymes involved in sugar nucleotide biosynthesis and related pathways across 150 species representing 12 eukaryote and prokaryote lineages (Fig. 4A, Supplementary Table S4). For each associated orthogroup, we plotted as a heat map the fraction of species within each taxonomic lineage represented by at least one orthologous sequence. The orthogroup containing Apicomplexa GNA1s clusters together with the presumed subunit H of phosphatidylinositol N-acetylglucosaminyltransferase (PIG-A subunit H, PF3D7_1141400), involved in GPI-anchor biosynthesis. These two were the only orthogroups with no potential orthologs outside of Apicomplexa, including T. brucei, the sister phylum to Apicomplexa.

Moreover, using simple reciprocal best-hit BLAST searches between P. falciparum and human (E < 10^-5), no potential human orthologs could be unambiguously defined for only two of the 30 P. falciparum enzymes, including PfGNA1 and PF3D7_0517500 (Supplementary Table S4). All together these analyses supported a rather exceptional evolutionary history of the Apicomplexa GNA1 family, compared to that of other enzyme families involved in glycoconjugate synthesis.

To further substantiate the orthologous relationships of genes belonging to the Apicomplexa GNA1 family, we examined their syntenic arrangements. A pairwise synteny analysis was performed by comparing the genomic region containing PfGNA1 in P. falciparum and 10 additional Plasmodium species, revealing series of collinear genes between the two regions (Supplementary Fig. S1). Similarly, synteny was examined both within and between species belonging to Aconoidasida (P. falciparum and B. bovis) and Conoidasida (C. parvum, T. gondii and T. annulata), the two main classes in the phylum (Supplementary Fig. S2). Despite the deep evolutionary divergence between the two classes within the phylum (~817 MYA), synteny could also be observed in some pairwise comparisons along hundreds of Mb of genomic regions, further supporting the single and independent origin of the GNA1 family in Apicomplexa.

We also used the alignment of the GNAT protein domain conserved regions (resulting after removing poorly aligned or highly diverged regions) from the seven Apicomplexa GNA1 family sequences in Fig. 2 as a seed to build a Hidden Markov Model (HMM) profile. The resulting HMM profile was in turn used as a query in iterative searches against the UNIPROT database, with the HMM profile being iteratively rebuilt on the basis of the retrieved hits, ultimately resulting in the identification of a set of 50 significant protein hits exclusive to
50 strains from 30 apicomplexan species. We selected a representative strain from each species, including the seven for which GNA1 activity was confirmed experimentally, to perform phylogenetic and sequence analyses (Supplementary Tables S5 and S6). Extensive diversification could be observed among lineages at the level of sequence length and architecture of protein motifs, reflecting the old age of the Apicomplexa clade\(^4^0\) (Fig. 4B). However, the phylogenetic relationships among sequences in the tree generally correspond well to the accepted taxonomic relationships among the species represented\(^4^1\). This observed congruence between the gene and the species trees is also compatible with a single evolutionary origin of the gene family. Comparative sequence analysis between apicomplexan and non-apicomplexan GNA1 families further supports their independent evolutionary origin. The large superfamily of GNAT-domain containing proteins groups different families of enzymes that use acetyl coenzyme A (AcCoA) to transfer an acetyl group to a substrate\(^2^6^,3^2^,3^3\). Subsequent evolutionary diversification of the superfamily would have been shaped by extensive sequence divergence and lineage-specific domain gain and loss. Although highly divergent at the sequence level, GNAT domains are well-conserved in structure and catalytic mechanism\(^2^6^,3^2^,3^3^, suggesting a common evolutionary origin. In order to search for signatures of remote amino acid sequence homology between GNAT domains of the Apicomplexa GNA1 and non-apicomplexan eukaryote GNA1 enzyme families, we compared the multiple sequence alignment of the conserved region of the GNAT domain in a dataset of 30 apicomplexan GNA1 sequences to the conserved regions of the GNAT domain from 30 GNA1 from non-apicomplexan eukaryote organisms, ranging from yeast to human (Supplementary Table S7). The conserved regions of the GNAT domain were extracted by removing poorly aligned or highly diverged regions from their respective alignments. The average percentage of pairwise sequence identity within the ca. 90 amino acid length alignment of apicomplexan GNA1s was 50.46% (similarity 67.71%), close to the ones observed within non-apicomplexan GNA1s (49.74%; similarity: 66.97%). In contrast, when apicomplexan and non-apicomplexan GNA1 sequences were aligned together (Supplementary Fig. S3), the average percentage of pairwise sequence identity between the two sets at the level of the conserved region of the GNAT domain drops to 18.06% (similarity 26.49%). The alignment
in vitro expressed and assayed metabolism, driven by a SSAT activity, although to our knowledge this enzyme has never been heterologously expressed, and purified the corresponding proteins. The enzymes used GlcN-6P as substrate and AcCoA as donor to produce GlcNAc-6P, as observed by HPLC-MS/MS (Fig. 6A). Although PfGNA1 and TgGNA1 were deficiently expressed, they also showed residual activity as detected by HPLC-MS/MS (Supplementary Fig. S5). In addition, the remaining recombinant proteins demonstrated clear GNA1 activity in colorimetric assays (Supplementary Fig. S6).

Since CpgGNA1 (cgd4_4000) has been previously described as a spermidine/spermine N-acetyltransferase (SSAT), involved in the back-conversion of polyamine spermine to spermidine and putrescine in C. parvum, we decided to further characterize its enzymatic activity. Kinetic analyses showed that CpgGNA1 used GlcN-6P as substrate to generate GlcNac-6P displaying classical Michaelis-Menten kinetics, with a K_m for AcCoA of 241.7 ± 35.3 μM and 719.1 ± 119.0 μM for GlcN-6P (Fig. S7A). In addition, in our hands CpgGNA1 was not able to acetylate spermine or spermidine and SSAT activity could not be detected either in colorimetric (Supplementary Fig. S7B) or HPLC-MS/MS based assays (Fig. 6B). Furthermore, on the basis of the relative levels of GlcNac-6P generated, CpgGNA1 activity was not inhibited by the presence of increasing concentrations of spermine or spermidine (Supplementary Fig. S7C). Hence, our data clearly indicates that the primary biochemical function of CpgGNA1 is to produce GlcNAc-6P, as observed by HPLC-MS/MS (Fig. 6A). Although T. gondii also shows an active polyamine retroconversion metabolism, driven by a SSAT activity, although to our knowledge this enzyme has never been heterologously expressed and assayed in vitro. Nevertheless, as in the case of CpgGNA1, TgGNA1 distinctly acetylated GlcN-6P,
and this activity was neither inhibited by spermine nor spermidine, suggesting that the function of TgGNA1 is indeed to acetylate GlcN-6P (Supplementary Fig. S7C).

CRISPR-Cas9-based gene disruption strongly suggests PfGNA1 is essential for parasite growth. Finally, to gain further insight into the biological function of PfGNA1, and to evaluate its potential as a drug target, we assessed whether it was required for parasite growth using CRISPR-Cas9-based techniques. In a first set of experiments, we attempted to disrupt PfGNA1 by CRISPR-Cas9-assisted gene truncation. Whereas controls using analogous plasmids targeting other regions generated viable parasites, no parasites harbouring a truncated version of PfGNA1 could be obtained (Fig. 7A). This suggests that the PfGNA1 gene plays an essential role for the survival of the parasite. In a second set of experiments, we were also unable to introduce nonsense mutations in the N-terminus of the GNAT domain conserved region. In contrast, in three independent biological replicates carried out in parallel, we successfully managed to obtain viable parasites harbouring synonymous nucleotide substitutions in the same region of the GNAT conserved domain (Fig. 7C). Therefore, nucleotides in the catalytic GNAT domain of PfGNA1 are genetically modifiable as long as the open reading frame is not altered. These data strongly suggest that an unaltered version of the PfGNA1 protein is required for parasite viability, at least during *P. falciparum* asexual blood stages.

Figure 6. Enzymatic activities of recombinant Apicomplexa GNA1 proteins. (A) The purified recombinant GNA1 enzymes were assayed in the presence of GlcN-6P and AcCoA (blue line, 25 µg/mL GNA1; red line, 5 µg/mL GNA1; green line, no GNA1). (B) The purified recombinant CpGNA1 enzyme was assayed in the presence of spermine and AcCoA. Chromatograms (left panels) show the detection of an N-acetyl spermine peak by HPLC-MS/MS in samples containing spermine (pink); AcCoA (black); AcCoA and spermine (blue); AcCoA, spermine and 250 µg/mL of CpGNA1 (red); AcCoA, spermine and 500 µg/mL of CpGNA1 (green); and AcCoA, spermine and 1000 µg/mL of CpGNA1 (grey). Nonenzymatic acetylation is observed in all the assays containing spermine and AcCoA. In these reactions, the generation of N-acetyl spermine is not dependent on the concentration of CpGNA1, as it is shown in superimposed chromatograms (right panel).
Discussion

Although GNA1 enzymes have been identified and characterized throughout the eukaryote kingdom, sequence similarity-based approaches failed to identify GNA1 genes in Apicomplexa. This difficulty was likely due to the extensive divergence at the sequence level featuring the different GNAT domain-containing enzyme families with specific acetyl acceptor substrates and taxonomic coverages. Here, we performed a systematic classification of the large superfamily of GNAT domain-containing proteins in seven Apicomplexa species with fully sequenced genomes. Our phylogenetic analysis revealed 9 well-supported gene families displaying conserved architecture of protein motifs. Among them, we identified protein families containing NMT or HAT domains, including representatives of the latter family from *P. falciparum* (PF3D7_0823300) and *T. gondii* (TGGT1_254555 and TGGT1_243440) that had been previously characterized. We tested five *P. falciparum* sequences, showing no evident correspondence with other GNAT domain-containing gene families, by complementation assays using a yeast thermosensitive strain defective in GNA1. As a result, we identified *PfGNA1* as able to support yeast growth at restrictive temperature. *PfGNA1* belongs to a well-defined family containing putative orthologs in all six other Apicomplexa species, all of which rescued the mutant under non-permissive conditions. The GNA1 family in Apicomplexa appears to have had a single origin early during diversification of the phylum, an observation supported by synteny-based analysis. Furthermore, using a HMM profile based on Apicomplexa GNA1 sequences as a query to scan the UNIPROT database, only sequences belonging to the phylum resulted in significant hits. Consistently with a separate evolutionary origin of the family, the phylogenetic tree of putative GNA1 from 30 species of Apicomplexa essentially recapitulated the accepted evolutionary relationships among the species, based on both ultrastructural and developmental characterizations as well as on molecular phylogenies. Altogether, our results support an independent origin of the GNA1 family in Apicomplexa, which can be traced back to between 817 MYA (estimated time of divergence between the two main classes of the phylum) and 1344 MYA (estimated time of divergence between Apicomplexa and its sister group of ciliates represented by *T. termophila*).

The comparison of the GNAT domains of apicomplexan and non-apicomplexan GNA1s showed that domain conservation was essentially restricted to a few specific amino acid positions around the GNAT six-amino acid consensus motif. Furthermore, the conserved residues were, with very few exceptions, distinct from the ones reported as critical for the catalytic activity of human GNA1. This observation suggests that, rather than evolving from an ancestral eukaryote GNA1 family, apicomplexan GNA1s might have evolved in parallel from a separate GNAT domain-containing gene lineage. Furthermore, considering the evolutionary time since divergence of the phylum from other eukaryotic lineages, Apicomplexa-specific GNA1s may have evolved functionally divergent catalytic residues exploitable for the design of selective small-molecule inhibitors.

The *CpGNA1* sequence (cgd4_4000) had been previously characterized to encode a SSAT, involved in the reverse polyamine biosynthetic pathway in *C. parvum*. Here, whereas a robust GNA1 activity could be detected...
with as low as 0.2 μg/mL of purified C. parvum protein, we did not detect a SSAT activity with up to 1000 μg/mL, both through colorimetric and HPLC-MS/MS based in vitro assays. In addition, neither CpGNA1 nor TgGNA143 activities were inhibited by the presence of spermine or spermidine. Considering that the rest of the apicomplexan proteins tested showed evident GNA1 activities, our results suggest that the primary function of this family of enzymes is the acetylation of GlcN6P. Thus, the SSAT activity previously described for CpGNA1 would be residual, secondary, and/or just restricted to specific taxonomic lineages, life cycle stages or physiological conditions.

Our multiple efforts to ablate PjGNA1 using CRISPR-Cas9-based techniques were unsuccessful, despite evidence that the locus was genetically modifiable. This strongly suggests that the gene is essential for the growth of the malaria parasite, in agreement with UDP-GlcNAc being a bottleneck metabolite46, likely due to its involvement in the biosynthesis of important glycoconjugates, such as GPI anchors. Nevertheless, the relevance of GNA1 in other apicomplexan organisms is at present unknown. Notably, in a recent work, a loss-of-function genome-wide screen was performed in T. gondii tachyzoites to assess the contribution of targeted genes to cell fitness 47. With the exception of TgGNA1 (TGGT1_243600), all the genes putatively encoding for enzymatic activities involved in UDP-GlcNAc biosynthesis (i.e. glucosamine:fructose-6-P-amidotransferase, TGGT1_231350; and N-acetylgalactosamine-phosphomutase, TGGT1_246450) or UDP-GlcNac utilization (i.e. Phosphatidylinositol N-acetylgalacosaminyltransferase, TGGT1_241860; and N-acetylgalacosaminyl phosphate transferase, TGGT1_244520) contributed greatly to T. gondii survival. This underlines the importance of the amino sugar biosynthetic route, whose presence is confirmed by labeling the glycoproteins of the parasite with [3H]GlcNAc48. The striking lack of contribution of TgGNA1 to T. gondii fitness47 might be suggesting the existence of another gene in the parasite's genome encoding for a redundant GNA1 activity, although it might also be highlighting certain limitations of the aforementioned study, such as the effect on proteins with slow turnover or the time selected for the screening readout 47. Indeed, considering the comprehensive analysis of GNA1-containing sequences in different Apicomplexa presented here, and the functional complementation assays performed with additional P. falciparum sequences, no other suitable candidate sequence encoding for a redundant GNA1 activity could be identified. The generation of conditional GNA1 mutants, still a challenging task in apicomplexan organisms, will contribute to comprehensively define whether GNA1 biochemical activity is essential in different species and to confirm its suitability as a potential therapeutic target for drug development.

Methods

Sequence and phylogenetic analysis. Phylogenetic analyses were performed on the basis of multiple alignments of amino acid sequences obtained using MUSCLE50. Bayesian and maximum-likelihood analyses were carried out using the Blosum62 + G (eight categories, shape parameter: 1.87) + F protein evolution model51, selected by ProtTest v3.2 as the best-fitting to the data52. Bayesian analysis was implemented in MrBayes 3.2.553. Searches were run with four Markov chains for one million generations sampling every 100th tree. After the stationary phase was reached, determined by the average standard deviation of split sequences approaching 0 (<0.05), the first 2,500 trees were discarded as burn-in. A consensus tree was then constructed to evaluate posterior probabilities on clades. Maximum-likelihood trees were constructed using PhyML v3.1, with tree topology searching optimized using the subtree pruning and regrafting option54. The statistical support of the retrieved topology was assessed using the Shimodaira-Hasegawa-like approximate likelihood ratio test55. Neighbor-joining phylogenetic analyses were conducted in Seaview v4.5.4, with statistical support on clades assessed using a bootstrap analysis with 1,000 replicates. Trees were represented and edited using iTOL v3.3.2. Profile hidden Markov models (HMMs) were used to identify conserved motifs shared among proteins56. Profile hidden Markov models (HMMs) were generated and calibrated using HMMER v3.057, on the basis of MUSCLE protein alignments58 further edited with Gblocks59. Logo representations of multiple protein sequence alignments were obtained using Skyline60. Protein secondary structures for apicomplexan GNA1 sequences were predicted using SABLE62.

Cloning of apicomplexan GNA1 candidate sequences. pRS421 vector (ATCC 87475), containing ampicillin and MET1563, was modified by cloning S. cerevisiae GPD promoter and CYC1 terminator sequences using SacII/BamHI and XhoI/KpnI restriction sites. The new vector generated, pRS421 pt was used to clone different GNA1 candidate sequences. Genomic DNAs of C. parvum, E. tenella, T. gondii, and T. brucei were used as templates for PCR amplification with Platinum DNA Polymerase High Fidelity (Thermo Scientific), using specific primers including BamHI and XhoI restriction sites (Supplementary Table S8). The remaining sequences cloned in pRS421 pt were codon-optimized synthetic genes (GenScript or Integrated DNA Technologies) based on the predicted amino acid sequences.

Yeast complementation assays. A S. cerevisiae GNA1 thermosensitive mutant, ScGNA1-ts38, was used for GNA1 complementation assays. The mutant contains a temperature-sensitive copy of GNA1 (YFL017C), which encodes a GNA1 enzymatic activity essential for yeast survival. The thermosensitive allele is marked by URA3 and the mutant was not able to grow above 34°C. ScGNA1-ts contains also a mutated MET15 gene that allows auxotrophic complementation with pRS421 pt vector in minimal media without uracil, methionine and cysteine (-ura, -met, -cys). After transformation with pRS421 pt based constructs, yeast cells were grown in serial dilution at permissive (23°C) and restrictive (37°C) temperatures.

Apicomplexan GNA1 protein expression and purification. Codon-optimized versions of apicomplexan GNA1 (PjGNA1, BbGNA1, CpGNA1, EtGNA1, TgGNA1, TaGNA1, and GnGNA1) native sequences were cloned in a pGEX 6P-1 vector containing an N-terminal glutathione S-transferase (GST) tag and transformed in E. coli BL21 (DE3). Cultures were grown for 16 h at 30°C after induction, lysed and supernatants containing the glutathione S-transferase (GST)-tagged GNA1 protein were filtered and purified through GSTrap HP 1-ml or GST SpinTrap columns (GE Healthcare). The proteins of interest were then dialyzed (Thermo Scientific
GNA1 and SSAT in vitro assays. GNA1 activity was assayed using 500 µM of AcCoA, 500 µM of GlcN-6P and different concentrations of apicomplexan GNA1 in 50 µl of 25 mM Tris–HCl–150 mM NaCl, pH 7.2 solution. To determine SSAT activity the same protocol was used but using as substrate 500 µM of spermine and 3 different CpgNA1 concentrations (250 µg/ml, 500 µg/ml and 1000 µg/ml). For SSAT colorimetric assays 500 µM of spermine or spermidine were used as substrate and 1, 20 or 500 µg/ml of CpgNA1. Reactions were allowed to proceed for 30 min at room temperature before being stopped by boiling for 10 min.

Liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-MS/MS). LC-MS/MS analyses were carried out on an UPLC – Acquity system (Waters) coupled by electrospray ionization to an API3000 triple quadrupole LC-MS/MS mass spectrometer (Perkin-Elmer Sciex), using a Kinetex® 2.6 µm Hilisc 100 Å column (150 × 4.6 mm, Phenomenex) for Glc-6P/GlcNAc-6P detection, and a an XBridge Hilisc 5.0µm, 130 Å (150 × 4.6 mm, Waters) for spermine/N-acetylserine detection (see Supplementary Methods).

CRISPR-Cas9 disruption and PFGNA1-editing constructs. Parasites were cultured and transfected either by electroporating ring-stage parasites or by nucleofection of schizont stages, as previously described64. All the methods were carried out in accordance with relevant guidelines and regulations and human erythrocytes and serum were purchased from the Banc de Sang i Teixits (Catalonia, Spain), after ethical approval from the Comité Étic Investigació Clínica Hospital Clinic de Barcelona. A single guide RNA (sgRNA) targeting the PFGNA1 consensus motif [((R/Q)-X-X-Q-X-G] was chosen using the Eukaryotic Pathogen CRISPR gRNA Design Tool64. For PFGNA1-disruption homology regions (HR) 1 and 2 were amplified from 3D7 P. falciparum genomic DNA using primers P1/P2 and P3/P4 and cloned in plasmid pL7 using Sphi/AllI and EcoRI/NcoI restriction sites, respectively64. sgRNA was integrated replacing pL7 BtgZI-adaptor (Fig. 7A). For PFGNA1 edition, sgRNA and Cas9-expressing construct (pDC2-Cas9-hDHFRyFCU)66 and linearized pUC19 plasmids were used as backbone. The sgRNA sequence was cloned in the pDC2-Cas9-sgRNA using primers P5/P6. PFGNA1 coding sequences with a shield mutation in the protospacer-adjacent motif (PAM)64 and silent or nonsense mutations were cloned in pUC19 to generate pUC19-silent PFGNA1 or pUC19-nonsense PFGNA1 (Fig. 7B). All primers used are described in Supplementary Table S8.

References
1. Cavalier-Smith, T. Kingdom protozoa and its 18 phyla. Microbiol. Rev. 57, 953–94 (1993).
2. Adl, S. M. et al. Diversity, Nomenclature, and Taxonomy of Protists. Syst. Biol. 56, 684–689 (2007).
3. World Health Organization. World Malaria Report, 2017 (2017).
4. Martens, C. & Vandepoele, K. & Van de Peer, Y. Whole-genome analysis reveals molecular innovations and evolutionary transitions in chromalveolate species. Proc. Natl. Acad. Sci. USA 105, 3427–32 (2008).
5. Coppenis, I. Targeting lipid biosynthesis and salvage in apicomplexan parasites for improved chemotherapies. Nat. Rev. Microbiol. 11, 823–35 (2013).
6. Rodriguez, J. A. et al. Parasite Glycobiology: A Bittersweet Symphony. Plos Pathog. 11, e1005169 (2015).
7. McConville, M. J. & Ferguson, M. A. The structure, biosynthesis and function of glycosylated phosphatidylinositoles in the parasitic protozoa and higher eukaryotes. Biochem. J. 394(Pt 2), 303–24 (1993).
8. Stiepen, B. et al. Molecular structure of the ‘low molecular weight antigen’ of Toxoplasma gondii: a glucose (1,4)-N-acetylgalactosamine makes free glycol-phosphatidylinositolosides highly immunogenic. J. Mol. Biol. 266, 797–813 (1997).
9. Naik, R. S. et al. Glycosylphosphatidylinositol Anchors of Plasmodium falciparum: Molecular Characterization and Naturally Elicited Antibody Response That May Provide Immunity to Malaria Pathogenesis. J. Exp. Med. 192, 1563–1576 (2000).
10. Hoering, T. L., Masterson, W. J., Englund, P. T. & Hart, G. W. Biosynthesis of the glycosyl phosphatidylinositol membrane anchor of the trypanosomainsurface glycophosphoprotein. Origin of the non-acetylated glucosamine. J. Biol. Chem. 261, 1168–73 (1986).
11. Bushkin, G. G. et al. Suggestive evidence for Darwinian Selection against asparagine-linked glycans of Plasmodium falciparum and Toxoplasma gondii. Eukaryot. Cell 9, 228–41 (2010).
12. Haserick, J. R., Leon, D. R., Samuelson, J. & Costello, C. E. Asparagine-linked Glycans of Cryptosporidium parvum Contain a Single Long Arm, Are Barely Processed in the ER or Golgi, and Show a Strong Bias for Sites with Threonine. Mol. Cell. Proteomics 16(4 suppl 1), S42–S53 (2017).
13. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. Biochim. Biophys. Acta 1426, 239–57 (1999).
14. Samuelson, J. & Robbins, P. W. Effects of N-glycan precursor length diversity on quality control of protein folding and on protein glycosylation. Semin. Cell Dev. Biol. 41, 121–128 (2015).
15. Izquierdo, L. et al. Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharidyltransferases. EMBO J. 28, 2650–2661 (2009).
16. Parodi, A. J. N-glycosylation in trypanosomatid protozoa. Glycobiology 3, 193–199 (1993).
17. Banerjee, S., Robbins, P. W. & Samuelson, J. Molecular characterization of nucleocytoplasmic O-GlcNac transferases of Giardia lamblia and Cryptosporidium parvum. Glyobiology 19, 331–336 (2008).
18. Cevallos, A. M. et al. Mediation of Cryptosporidium parvum infection in vitro by mucin-like glycoproteins defined by a neutralizing monoclonal antibody. Infect. Immun. 68, 5167–75 (2000).
19. Tomita, T. et al. Making Home Sweet and Sturdy: Toxoplasma gondii ppGalNAc-Ts Glycosylate in Hierarchical Order and Confer Cyst Wall Rigidity. MBio 8, e02048–16 (2017).
25. Riegler, H. et al. Crystal structure and functional characterization of a glucosamine-6-phosphate N-acetyltransferase from Arabidopsis thaliana. Biochem. J. 443, 427–37 (2012).
26. Maríno, K. et al. Characterization, localization, essentiality, and high-resolution crystal structure of glucosamine 6-phosphate N-acetyltransferase from Trypanosoma brucei. Eukaryot. Cell 10, 985–97 (2011).
27. Mio, T., Yamada-Okabe, T., Arisawa, M. & Yamada-Okabe, H. Saccharomyces cerevisiae GNA1, an essential gene encoding a novel acetyltransferase involved in UDP-N-acetylglucosamine synthesis. J. Biol. Chem. 274, 424–9 (1999).
28. Stokes, M. J. et al. The synthesis of UDP-N-acetylglucosamine is essential for bloodstream form trypanosoma brucei in vitro and in vivo and UDP-N-acetylglucosamine starvation reveals a hierarchy in parasite protein glycosylation. J Biol Chem 283, 16147–16161 (2008).
29. Cova, M., Rodrigues, J. A., Smith, T. K. & Izquierdo, L. Sugar activation and glycosylation in Plasmodium. Malar. J. 14, 427 (2015).
30. Dyda, F., Klein, D. C. & Hickman, A. B. GCN5-Related N-Acetyltransferases: A Structural Overview. Annu. Rev. Biophys. Biomol. Struct. 29, 81–103 (2000).
31. Vetting, M. W. et al. Structure and functions of the GNAT superfamily of acetyltransferases. Arch. Biochem. Biophys. 433, 212–226 (2005).
32. Wang, J., Liu, X., Liang, Y.-H., Li, L.-F. & Su, X.-D. Acceptor substrate binding revealed by crystal structure of human glucosamine-6-phosphate N-acetyltransferase 1. FEBS Lett. 382, 2973–2978 (2008).
33. Hurtado-Guerrero, R. et al. Structural and kinetic differences between human and Aspergillus fumigatus D-glucosamine-6-phosphate N-acetyltransferase. Biochem. J. 415, 217–223 (2008).
34. Sanz, S. et al. Biosynthesis of GDP-fucose and other sugar nucleotides in the stages of Plasmodium falciparum. J. Biol. Chem. 288, 16506–17 (2013).
35. Sanz, S. et al. The disruption of GDP-fucose de novo biosynthesis suggests the presence of a novel fucose-containing glycoconjugate in Plasmodium asexual blood stages. Sci. Rep. 6, 37230 (2016).
36. López-Gutiérrez, B., Śliszgala, R. R. & Izquierdo, L. Sugar nucleotide quantification by liquid chromatography tandem mass spectrometry reveals a distinct profile in Plasmodium falciparum sexual stage parasites. Biochem. J. 474, 897–905 (2017).
37. Felsenstein, J. Inferring phylogenies. (Sinauer Associates, 2004).
38. Ben-Aroya, S. et al. Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. Mol. Cell 30, 248–58 (2008).
39. Koo, C.-H., Wars, J. P. & Kissinger, J. C. The Apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. Mol. Biol. Evol. 25, 2689–98 (2008).
40. Yarlett, N. et al. Cryptosporidium parvum spermidine/spermine N1-acetyltransferase exhibits different characteristics from the host enzyme. Mol. Biochem. Parasitol. 152, 170–180 (2007).
41. Cook, T. et al. Divergent polyamine metabolism in the Apicomplexa. Microbiology 153, 1123–1130 (2007).
42. Fan, Q., An, L. & Cui, L. Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologe involved in chromatin remodeling. Eukaryot. Cell 3, 264–76 (2004).
43. Bhati, M. M., Livingstone, M., Mullapudi, N. & Sullivan, W. J. Pair of unusual GCN5 histone acetyltransferase and ADA2 homologues in the protozoan parasite Toxoplasma gondii. Eukaryot. Cell 5, 62–76 (2006).
44. Chiappinno-Pepe, A., Tymoshenko, S., Ataman, M., Soldati-Favre, D. & Hatzimanikatis, V. Bioenergetics-based modeling of Plasmodium falciparum metabolism reveals its essential genes, nutritional requirements, and thermodynamic bottlenecks. PLOS Comput. Biol. 13, e1005397 (2017).
45. Siddiqi, S. M. et al. A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. Cell 166, 1423–1435.e12 (2016).
46. Zinecker, C. F. et al. Two glycoforms are present in the GPI-membrane anchor of the surface antigen 1 (P30) of Toxoplasma gondii. Mol. Biochem. Parasitol. 116, 127–35 (2001).
47. Tomavo, S., Schwarz, R. T. & Dubremetz, J. F. Evidence for glycosyl-phosphatidylglycerol anchoring of Toxoplasma gondii major surface antigens. Mol. Cell. Biol. 9, 4576–80 (1989).
48. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughpout. Nucleic Acids Res. 32, 1792–7 (2004).
49. Henkoff, S. & Henkoff, J. G. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89, 19615–9 (1992).
50. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–5 (2011).
51. Ronquist, F. & Huelsenbeck, J. P. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–4 (2003).
52. Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. USA 108, 13624–9 (2011).
53. Kuo, C.-H., Wars, J. P. & Kissinger, J. C. The Apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. Mol. Biol. Evol. 25, 2689–98 (2008).
54. Yarlett, N. et al. Cryptosporidium parvum spermidine/spermine N1-acetyltransferase exhibits different characteristics from the host enzyme. Mol. Biochem. Parasitol. 152, 170–180 (2007).
55. Cook, T. et al. Divergent polyamine metabolism in the Apicomplexa. Microbiology 153, 1123–1130 (2007).
56. Fan, Q., An, L. & Cui, L. Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologe involved in chromatin remodeling. Eukaryot. Cell 3, 264–76 (2004).
57. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245 (2016).
58. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–8 (2009).
59. Eddy, S. R. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7, e1002195 (2011).
60. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. 17, 540–552 (2000).
61. Wheeler, T. J., Clements, J. & Finn, R. D. Skyline: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics 15, 7 (2014).
62. Adamczak, R., Porollo, A. & Meller, J. SABLE protein structure prediction server at, http://sable.ccbmc.org/ (2003).
63. Baker Brachmann, C. et al. Designer Deletion Strains derived from Saccharomyces cerevisiae S288C: A Useful set of Strains and Plasmids for PCR-mediated Gene Disruption and Other Applications. Yeast 14, 115–132 (1998).
64. Ghobrial, M. et al. Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. Nat. Biotechnol. 32, 819–821 (2014).
65. Peng, D. & Tarleton, R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. Microb. genomics 1, e000033 (2015).
66. Ng, C. L. et al. CRISPR-Cas9-modified p/fmdr1 protects Plasmodium falciparum asexual blood stages and gametocytes against a class of piperazine-containing compounds but potentiates artesinin-based combination therapy partner drugs. Mol. Microbiol. 101, 381–393 (2016).
67. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc. 58, 236–244 (1963).
68. Wallace, H. M. & Evans, D. M. In Polyamine Protocols 59–68 (Humana Press, 1998).
Acknowledgements

ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya. LI and MC are members of the GlycoPar-EU FP7 funded Marie Curie Initial Training Network (GA 608295). This work was supported by SAF2016-76080-R grant from the Spanish Ministry of Economy (AEI/FEDER, UE). Work performed at Boston University was supported by NIH grant R01 AI110638. LC-P was supported by a non-contributory unemployment subside from the Spanish Government, and later by a postdoctoral fellowship from the VIB Research Foundation. We are grateful to M. Ramirez, CCITUB (Scientific and Technological Centers Universitat de Barcelona), N. S. Berrow and R. Garcia for technical support. The authors would like to especially thank Philip Hieter and colleagues for providing us with the S. cerevisiae GNA1-ts strain. DNAs from Tg, Cp and Em were a generous gift from John Samuelson. TbGNA1 was kindly provided by Michael A. J. Ferguson’s lab. We would also like to thank Ellen Knupefer for the CRISPR/Cas9 vector pDC2-Cas9-hDHFRyFCU and cloning instructions. We are also indebted to Hernando A. del Portillo, Elsy M. Ngwa, Markus Aebi and Francoise H. Routier for helpful ideas, discussions and support.

Author Contributions

M.C., L.C.-P. and L.I. conceived the work. L.C.-P. performed bioinformatics analyses. M.C., S.A.-J., A.G.-D. and G.G. were responsible for yeast complementation and GNA1 heterologous expression and purification. M.C. and B.L.-G. conducted GNA1 in vitro assays. M.C. was responsible of PfGNA1 genetic manipulation. M.C., L.C.-P., S.M. and L.I. interpreted the results. M.C., L.C.-P. and L.I. outlined the draft. All authors contributed to the writing and review of this manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-22441-3.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018