Supplementary Materials

Genetic evidence that an endosymbiont-derived ERAD system functions in import of apicoplast proteins, *Swati Agrawal, Giel G. van Dooren, Wandy L. Beatty, and Boris Striepen*

1. Supplementary Figures:

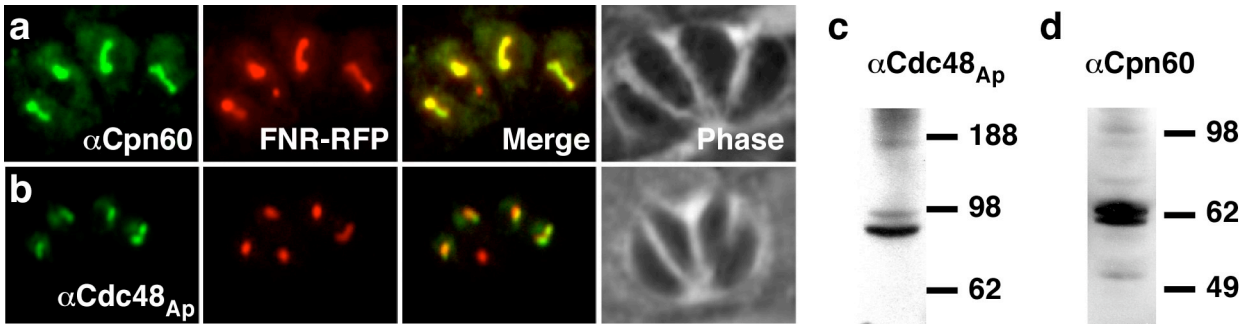

Figure S1 | Antibodies raised against recombinant $Cdc48_{AP}$ and Cpn60. Fragments of the $Cdc48_{Ap}$ and Cpn60 genes were expressed as His-tagged recombinant fusion proteins in *E. coli*. The proteins were purified by affinity chromatography and used to immunize rabbits. The resulting anti-sera were tested against *T. gondii* by **a**,**b**, Immunofluorescence and **c**, **d** Western blot. Both sera (αCpn60, $\alpha Cdc48_{Ap}$) detect protein products of the expected molecular weight and show co-labeling with the apicoplast marker FNR-RFP.

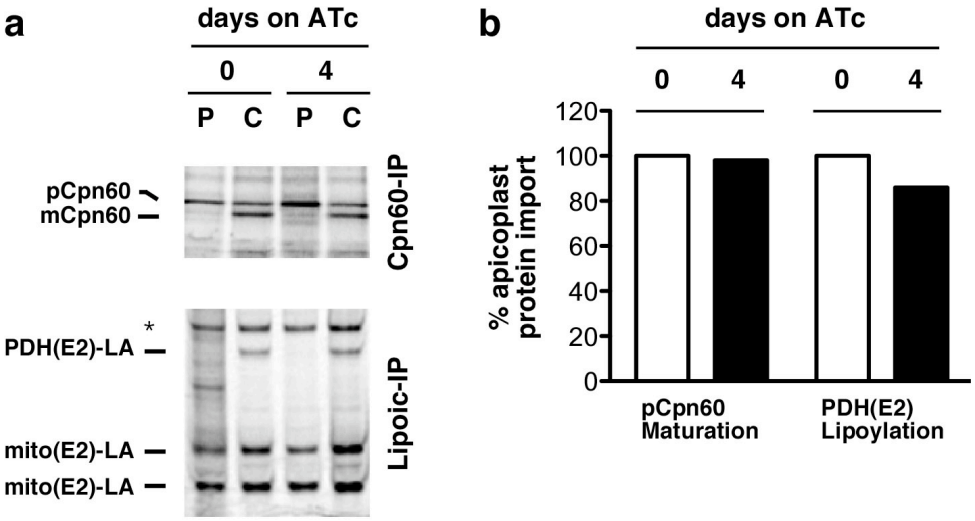

Figure S2 | **ATc treatment alone has no significant effect on apicoplast protein import.** As additional control for experiments shown in Fig. 5 we performed pulse-chase analysis using the parental $Der1_{Ap}$ line, which still carries the native copy of $Der1_{Ap}$. (**a**) Post-translational

modification of Cpn60 and PHD-E2 were analyzed after 0 or 4 days of ATc treatment as detailed in Fig. 5 by pulse-chase labeling, immunoprecipitation, SDS-PAGE and autoradiography (* human PDH-E2). (**b**) Maturation of Cpn60 and lipoylation of PDH-E2 were quantified by phosphorimager analysis of the respective bands in the autoradiographs shown in (**a**) followed by calculation of the ratio of mature to precursor protein (Cpn60) and ration of PDH-E2 to mitochondrial E2 (lipoylation; untreated control ratios were arbitrarily set to 100%).

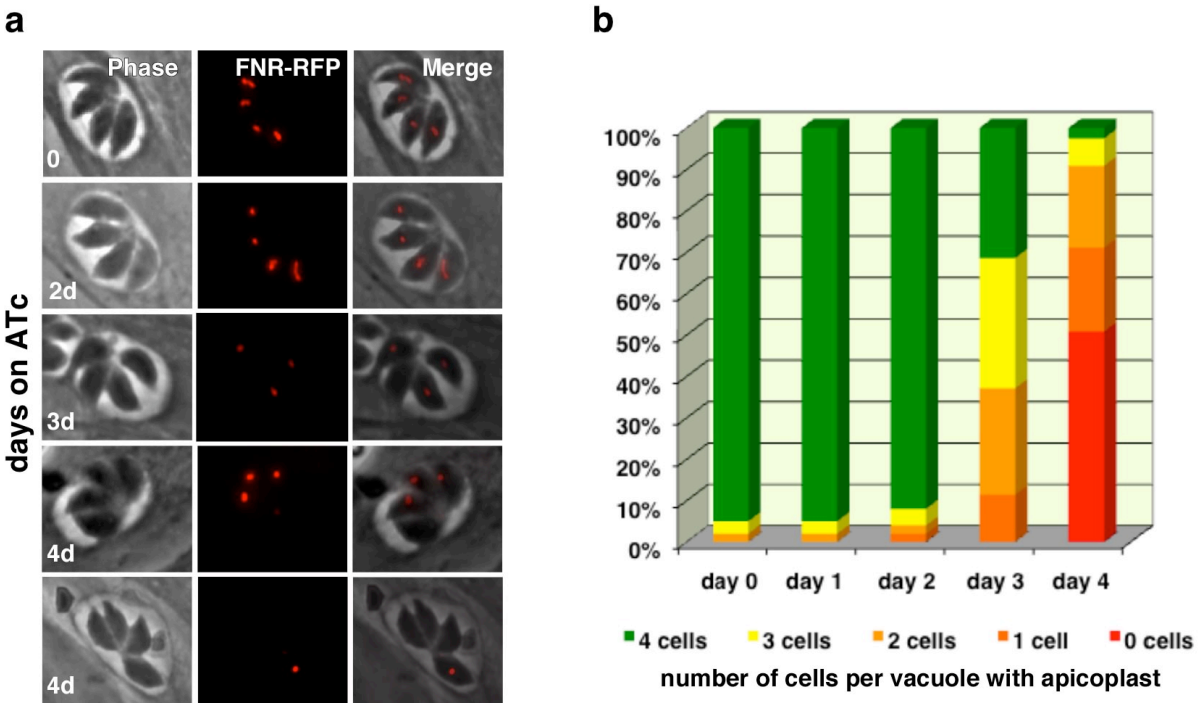

Figure S3 | Loss of Der1$_{Ap}$ results in apicoplast biogenesis defects. Der1$_{Ap}$ mutant parasites expressing the apicoplast marker FNR-RFP were grown in the presence of ATc for 0-4 days. The presence of apicoplast was scored by life cell microscopy (1) in 100 vacuoles (each containing four parasites (see **a**). We note plastid counts indistinguishable from controls for the first two days with a dramatic decline on day 3 and 4. Panel **a,** shows representative micrographs and **b**, shows quantification of plastid numbers for 100 4-cell vacuoles for each sample. Note that plastid biogenesis is not affected by ATc treatment in wild type parasites (1,2).

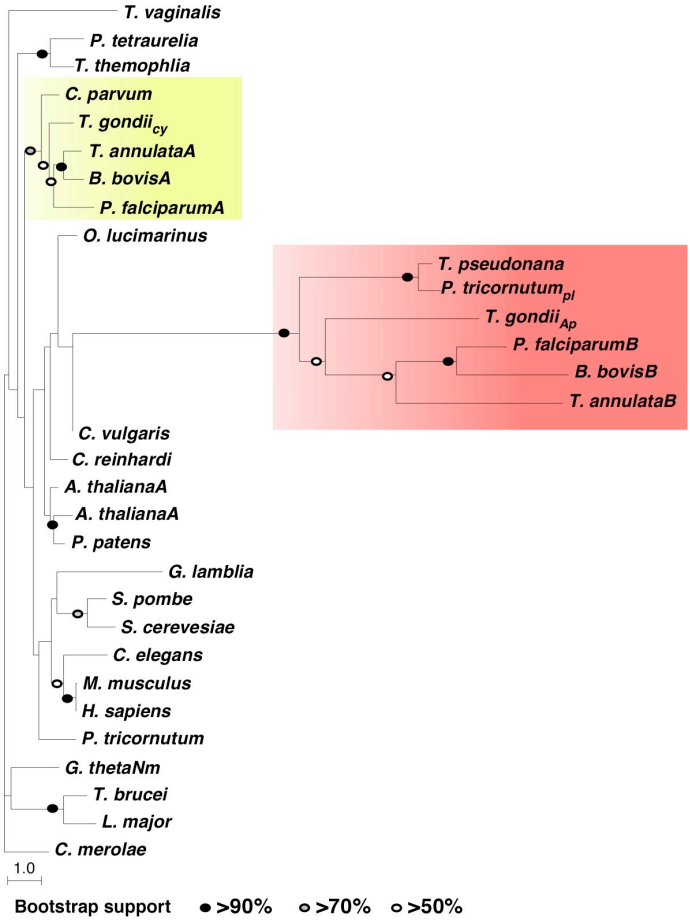

Figure S4 | Divergent origins of *T. gondii* Ufd proteins. RAxML maximum likelihood tree derived from alignments of Ufd proteins from 25 taxa (480 aligned amino acid characters were used after manual inspection, accession numbers are provided in the supplement). Bootstrap analyses were conducted using 100 replicates. Nm, nucleomorph (remnant endosymbiont nucleus), Pl, plastid, Ap, apicoplast, Cy, cytoplasm. Der1 protein sequences were to divergent to construct multiple sequence alignments.

2. Supplementary Methods

Phylogenetic analyses

Cdc48 homologues included in the analyses were ([genbank accession number], [*joint genome institute accession numbers]) from human [AAI21795], *Mus musculus* [NP_033529], *Caenorhabditis elegans* [NP_496273], *Drosophila melanogaster* [NP_477369), *Saccharomyces cerevisiae* [NP_010157], *Schizosaccharomyces pombe* [NP_593287], *Dictyostelium discoideum* [XP_636910], *Ostreococcus lucimarinus* [XP_001415566], *Chlorella vulgaris* [42408*], *Chlamydomonas reinhardtii* [XP_001696503], *Physcomitrella patens* [XP_001777213], *Arabidopsis thaliana* (A. thalianaA [P54609], A. thalianaB [NP_190891], (A. thalianaC [NP_568114]), *Emiliania huxleyi* [557511*], *Hyaloperonospora* parasitica [AAY58902], *Toxoplasma gondii* (TgCdc48$_{Cy}$ [1207706] and TgCdc48$_{Ap}$ [FJ976519]), *Plasmodium falciparum* (P. falciparumA [CAG25009] and P. falciparumB [XP_001349023]), *Theileria annulata*

(T.annulataA [XP_953837] and T.annulataB [XP_955188]), *Babesia bovis* (B. bovisA [XP_001610758] and B. bovisB [XP_001609335]) and *Cryptosporidium hominis* [XP_667275], *Thalassiosira pseudonana* (T.pseudonanaA [XP_002286617] and T.pseudonanaB [XP_002289352] and *Phaeodactylum tricornutum* (P. tricornutumA [XP_002185883] and P. tricornutumB [50978*]), *Guillardia theta* [XP_001713564], Cyanidioschyzon *merolae* (http://merolae.biol.s.u-tokyo.ac.jp/ CmCdc48A [c12f0001], *Tetrahymena thermophila* [XP_001007447] and *Paramecium tetraurelia* [XP_001456132]), *Trichomonas vaginalis* [XP_001317755], *Giardia Lamblia* [XP_001704687], *Entamoeba histolytica* [AAF74998]), *Trypanosoma brucei* (TbCdc48 [AAC02215]) and *Leishmania major* [XP_001686709]).

Ufd protein sequences from 25 eukaryote taxa were used to generate a multi-sequence alignment using ClustalX version 2.0. Upon manual inspection 480 aligned amino acid characters were selected for further analysis (alignments are available on request from the authors). We note that Ufd1 is not sufficiently conserved to resolve the tree of live, however the divergent origin of the cytoplasmic and apicoplast proteins are evident. The data set was subjected to maximum likelihood-based phylogenetic analysis using RAxML (3) version 7.0.4 available at http://8ball.sdsc.edu:8889/cipres-web. A phylogenetic tree was constructed employing the GAMMA+P-Invar evolutionary model and the model parameters used were alpha: 1.636618, invar: 0.000117 and Tree-Length: 21.567051. Bootstrap analyses were conducted using 100 replicates as described previously(4). Ufd homologues included in the analysis are from human [NP_005650], *Mus musculus* [AAH06630], *Caenorhabditis elegans* [NP_502348], *Saccharomyces cerevisiae* [NP_011562], *Schizosaccharomyces pombe* [NP_596780], *Ostreococcus lucimarinus* [XP_001419494], *Chlorella vulgaris* [28219*], *Chlamydomonas reinhardtii* [XP_001699038], *Physcomitrella patens* [XP_001778779], *Arabidopsis thaliana* (A. thalianaA [NP_001077933], *A. thalianaB* [NP_973557], *Toxoplasma gondii* (T. gondiiCy [FJ976516] and *T. gondii*Ap [FJ976517]), *Plasmodium falciparum* (*P. falciparumA* [XP_001348351] and *P. falciparum*B [CAX64217]), *Theileria annulata* (*T. annulataA* [XP_952190] and *T. annulataB* [XP_953638]), *Babesia bovis* (*B. bovisA* [XP_001610110] and *B.bovis*B [XP_001611093]) and *Cryptosporidium parvum* [XP_625717], the *Thalassiosira pseudonana* (*T. pseudonana* [XP_002289664]) and *Phaeodactylum tricornutum* (*P. tricornutum*A [XP_002178490] and *P. tricornutum*B [49319*]), *Guillardia theta* nuleomorph (*G. theta*Nm [AAF24006.1], *Cyanidioschyzon merolae* (http://merolae.biol.s.u-tokyo.ac.jp/ CmCdc48A [11f0001], the *Tetrahymena thermophila* [XP_001010411] and *Paramecium tetraurelia* [XP_001437627]), *Trichomonas vaginalis* [XP_001321792], *Giardia Lamblia* [XP_001706512] *Trypanosoma brucei* (TbCdc48 [XP_823013]) and *Leishmania major* [XP_001687240]).

Expression of recombinant $Cdc48_{Ap}$ and Cpn60 and antibody production

T.gondii $Cdc48_{Ap}$ and Cpn60 sequences (primers used for amplification are listed in ST3) were amplified and introduced into the bacterial expression plasmid pAVA421(5) by ligation independent cloning as described previously (6) to generate a fusion protein carrying a six histidine tag at the N-terminus. Plasmids were introduced into the *E. coli* BL21 strain. Recombinant fusion proteins were purified by affinity chromatography on Ni2+NTA resin in presence of 6M urea according to the manufacturer's protocol (Qiagen). A polyclonal anti-serum was generated by rabbit immunization (Cocalico Biologicals, Inc. PA, USA) following a standard protocol. The antisera were tested by Western blot and IFA against *T. gondii* (see Fig. S1).

3. Supplementary Tables

Supplementary Table ST1: Identification of the full coding sequences for *T. gondii* ERAD components. For seven genes we experimentally evaluated the gene model derived from the automatic annotation process (toxoDB.org). We identified additional exons and/or divergent 5' and 3' termini by RACE and RT-PCR, subcloning and sequencing. We provide the ToxoDB genome identifiers below for reference but note that further work should be based on the validated coding sequences available from genbank using the indicated accession numbers.

Name	Genbank-accession number	ToxoDB identifier	Primers for 5'RACE	Primers for 3'RACE
Der1-1$_{ER}$	FJ976521	TGME49_094290	EST's covering 5' end were present	EST's covering 3' end were present
Der1-2$_{ER}$	FJ976522	TGME49_017160	5'- TGAAAAAGTACGAACCGACGCCAAAGATATGCGACAGTCC	5'-TTTGGCGTCGGTTCGTACTTTTTCAGCGG
Cdc48$_{cyt}$	FJ976518	TGME49_073090	5'- GGCAAGTAAATCAGGTCCCCCACGGTCAG Nested 5'-AAAGTGGCAAGTTCCTTCCTCCGCAACGAA	5'-TCCCGAGAGCGCACCAGCGAAGAAGGAAGAG Nested 5'- ACTCGGGAGCCGTGGAAGAACCGGATCAAAA
Ufd$_{cyt}$	FJ976516	TGME49_070530	5'-GGCAAGTAAATCAGGTCCCCCACGGTCAG nested 5'-AAAGTGGCAAGTTCCTTCCTCCGCAACGAA	5'TCCCGAGAGCGCACCAGCGAAGAAGGAAGAG nested 5'-ACTCGGGAGCCGTGGAAGAACCGGATCAAAA
Der1$_{Ap}$	FJ976520	TGME49_081940	5'- CGACGAAGAAAAAGACGACAGGAGCGACAGTG nested 5'- CTTTGTCAGCTTCGGCGTCGCCTTCCACT	5'-CACTGTCGCTCCTGTCGTCTTTTTCTTCGTCG nested 5'-CCTGCTTTCTTCGCTATTCTTCCTCGGCCCGTT
Cdc48$_{Ap}$	FJ976519	TGME49_121640	5'ATCCGCCGCAAATTCGCCTCGCTCTCCCCAGCCAACT nested 5'-GTGAAGCAATACGCCCCGCGGTGTCTGT	5'-CGCAAACCTATCTTCGTCATCGGAGCCACCAATCG nested 5'-CGTGGACATTGAAGACATGGCTCGGAGACTCGAAGGCTTTTC
Ufd$_{Ap}$	FJ976517	TGME49_085700	RACE unsuccessful	5'-CTTCCTCCCCTTATGGGTCATGAAGGCACTCGACTTGCG nested 5'-AATGGGAGCGCCTTCCGCTGGCGGGTCACGT

Supplementary Table ST2: Construction of tagged genes for subsequent localization of their protein products.

Name	Primers used for cloning	Vector used for cloning	Method of localization	Localization
Der1-1$_{ER}$	for 5'- AGATCTAAAATGGGGTTCCCTTCA GCTCTCT rev5'- CCTAGGGTACGAGTCTGACGGC TTCTCGCCGAT	pCTH	Cloned and introduced as minigene	ER
Der1-2$_{ER}$	for5'- CTAGAGATCTAAAATGGCGCAGG TGGACTTGTTCTTC rev5'-CTAGCCTAGGCTCTCTTGGGGTGA GGCGCGACCTC	pCTH	Cloned and introduced as minigene	ER
Cdc48$_{cyt}$	for 5'-CTAGGGATCCAAAATGGCCGGCG GCATTCGCAG rev 5'-CTAGCCTAGGCGAGTAGAGGTCAT CGTCATCCGCG	pCTM3	Cloned and introduced as minigene	Cytosol
Ufd$_{cyt}$	for5'-CTAGAGATCTAAAATGTTCAGTCG CCACGTAGCGAATCTG rev 5'- CTAGCCTAGGCTCGCAGGTATTG CCTTTTCCAAAG	pCTH	Cloned and introduced as minigene	Cytosol
Der1$_{Ap}$	for 5'-AGATCTAAAATGGAAAGAGGGGA TTTTTTCTCACT rev 5'- CTAGTCCGTTTCCAACGGCGTC CTCGTTTAA	pDT7S4HA	Cloned and introduced as minigene	Apicoplast
Cdc48$_{Ap}$	for 5'-CTAGCAAAATGGGGACTGCGTGG TGCCCTCTCG rev 5'- CTAGCCTAGGCTTTGTTTCCTTC GCCGTCTCCGT	pCTM3	Cloned and introduced as minigene (antiserum was raised against recomb. protein)	Apicoplast
Ufd$_{Ap}$	for 5'-AACAAGAGAAACAGAGAGAGAA GAAGAGAACAGAAGACACGAAACAC GCAAGGTACCCGTACGACGTCCCGGACT AC rev 5'-GCTCTTGTTGACTCTTCTCTTCCC GGATGTCATGCGAGTCTCTCCTTCGC ATACGACTCACTATAGGGCGAATTGG	Cosmid TOXOx83	Tagged 3'end of gene in cosmid	Apicoplast

Supplementary Table ST3: Construction of Der1$_{Ap}$ KO construct and Cdc48$_{Ap}$ and Cpn60 antibody expression vectors

Name	Primers used for cloning	Vector used
Der1 $_{Ap}$ 5' flank for KO vector	For 5'- CTAGACTAGTGGCTGTTCCTTCCCCACT GTATTAT Rev 5'- CTAGGACGTCGCTGGTGAACAGGAAGCA CGACCTT	pTCY *Spe*I and *Aat*II sites
Der1 $_{Ap}$ 3' flank for KO vector	for 5'- CTAGgtcgacACGCTTCGATGTCTTTATCG GCATC and rev 5'- CTAGGGTACCCGAGCATTGCACACGCTC GTCCTCC	pTCY *Sal*I and *Kpn*I sites
Cdc48$_{Ap}$ Antibody	Primers for 5'- GGGTCCTGGTTCGATGGATCCTTCAGTG GTTTTCCTCTCGCC and rev 5'- CTTGTTCGTGCTGTTTATTAGTCTTCCCA GCGAACGTCTGGCACTT	pAVA421, Cloned by LIC
Cpn60 Antibody	For 5'- gggtcctggttcgatgAAAGATCGCACGTCGATT CTTACAAGG rev5'- cttgttcgtgctgtttattaTGCCATTGGCATGTCTG GTACATC.	pAVA421, Cloned by LIC

Supplementary References

1. Mazumdar, J., Wilson, E., Masarek, K., Hunter, C., and Striepen, B. (2006) *Proc Natl Acad Sci U S A* **103**, 13192-13197
2. Van Dooren, G. G., Tomova, C., Agrawal, S., Humbel, B. M., and Striepen, B. (2008) *Proc Natl Acad Sci U S A* **105**, 13574-13579
3. Stamatakis, A. (2006) *Bioinformatics* **22**, 2688-2690
4. Stamatakis, A., Hoover, P., and Rougemont, J. (2008) *Syst Biol* **57**, 758-771
5. Alexandrov, A., Vignali, M., LaCount, D. J., Quartley, E., de Vries, C., De Rosa, D., Babulski, J., Mitchell, S. F., Schoenfeld, L. W., Fields, S., Hol, W. G., Dumont, M. E., Phizicky, E. M., and Grayhack, E. J. (2004) *Mol Cell Proteomics* **3**, 934-938
6. Aslanidis, C., and de Jong, P. J. (1990) *Nucleic Acids Res* **18**, 6069-6074