

Figure S1. N-terminal 450 characters of a multiple sequence alignment of a data set of eukaryotic RPA2 proteins. A multiple sequence alignment was generated from a data set of 16 eukaryotic RPA2 proteins. The trypanosomatid sequences contain more than 200 amino acids more in their N-termini than other RPA2 sequences. Residues are displayed with a blue or cyan background colour to indicate either identity or similarity to the majority consensus. Yellow flags on the *T. brucei* sequence indicate the amino acids tagged in this study. Blue bars mark stretches rich in basic amino acids in the trypanosome proteins. The red bar marks a region of limited similarity, the green bar marks RPA2 domain A. Species abbreviations: AraTh: *Arabidopsis thaliana*, PhyPa: *Physcomitrella patens*, CaeEl: *Caenorhabditis elegans*, DicDi: *Dictyostelium discoideum*, HomSa: *Homo sapiens*, SacCe: *Saccharomyces cerevisiae*, SchPo: *Schizosaccharomyces pombe*, PhaTr: *Phaeodactylum tricornutum*, PhySo: *Phytophthora sojae*, ParTe: *Paramecium tetraurelia*, TetTh: *Tetrahymena thermophila*, DroMe: *Drosophila melanogaster*, EntHi: *Entamoeba histolytica*, TriVa: *Trichomonas vaginalis*, CryPa: *Cryptosporidium parvum*, ToxGo: *Toxoplasma gondii*, TheAn: *Theileria annulata*, CyaMe: *Cyanidioschyzon merolae*, GiaLa: *Giardia lamblia*, PlaFa: *Plasmodium falciparum*, LeiMa: *Leishmania major*, TryBr: *Trypanosoma brucei*.

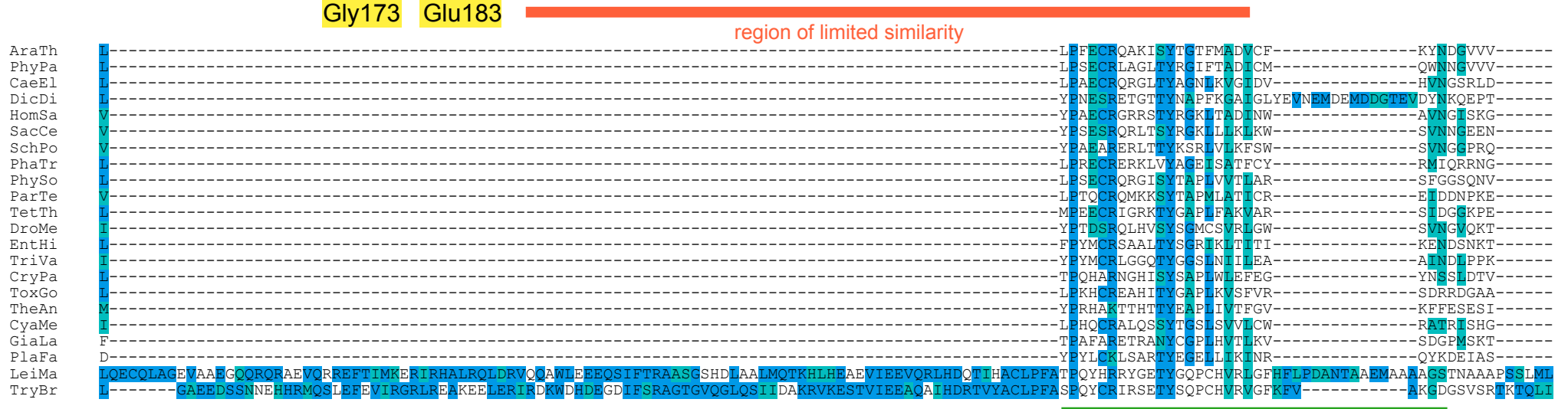
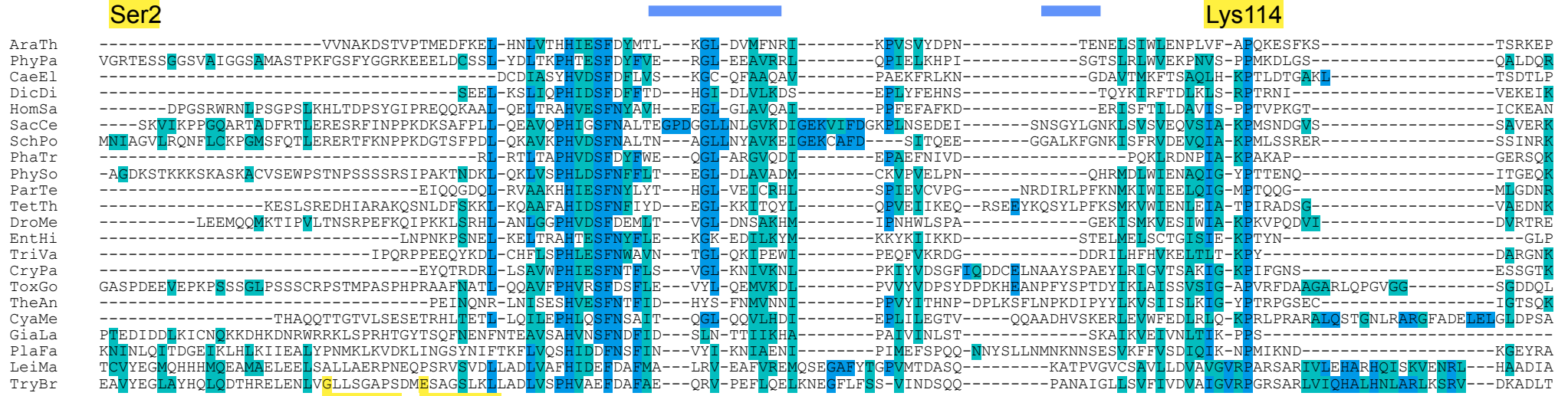
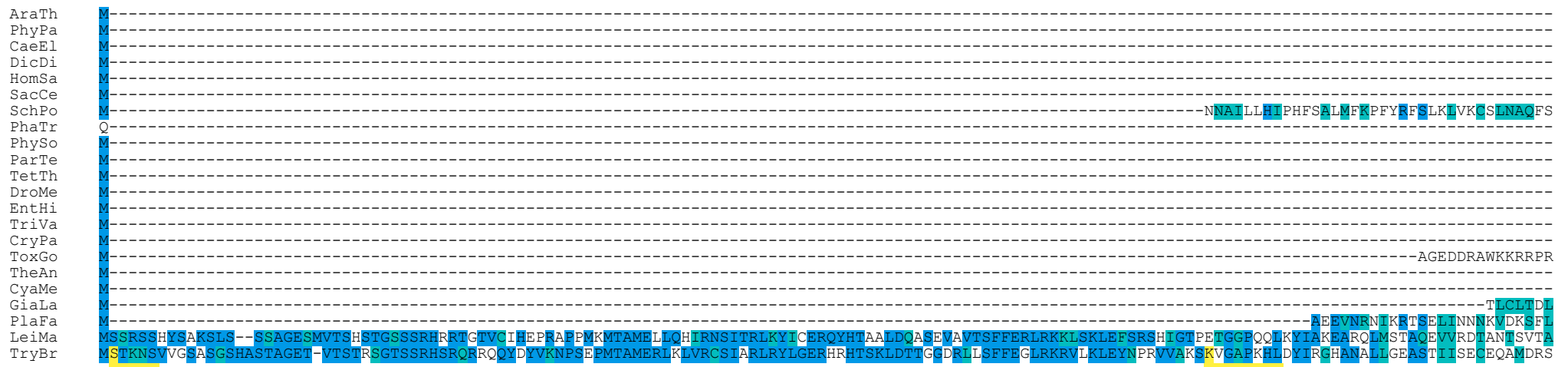
Figure S2. Multiple sequence alignment of trypanosomatid RPA2 N-termini. RPA2 N-termini from available trypanosome genomes. Note that most of the basic residues in the basic stretches in the *T. brucei* sequence are not conserved. Residues were given a blue or cyan background colour to indicate either identity or similarity to the majority consensus. Blue bars mark stretches rich in basic amino acids in *T. brucei*. The green bar marks RPA2 domain A.

Figure S3. Genomic PCR products confirming the correct integration of the tagging constructs in bloodstream-form (A) and procyclic (B) cells. Expected PCR product size for control PCR (primers c) was 428 bp, for PCR from *YFP* to *RPA2* (y) was 956 bp for RPA2, 610 bp for RPA2 Δ_{1-113} and 443 bp for RPA2 Δ_{1-172} . DNA ladders are NEB 100 bp and 1 kb ladders.

Figure S4. Cellular localization of N-terminal RPA2 truncations in procyclic cells. While TY-YFP-RPA2 localizes to the nucleus, TY-YFP-rpa2 Δ_{1-113} and -rpa2 Δ_{1-172} are excluded. NLS-TY-YFP-RPA2 has a different sub-nuclear localization to NLS-TY-YFP-rpa2 Δ_{1-172} . YFP fluorescence is pseudo-coloured in green, DAPI in blue. Bars: 1 μ m.

Figure S5. Cellular localisation of NLS-TY-YFP-RPA2 and NLS-TY-YFP-rpa2 Δ_{1-172} in bloodstream-form cells. A strong, punctate extra-nucleolar YFP signal was observed (arrowheads) in many cells indicating that NLS-TY-YFP-RPA2 is found in the ESB, as for TY-YFP-RPA2. This region also corresponds to a region of low DAPI stain intensity. YFP fluorescence was pseudocolored in green, DAPI in blue. Bar: 1 μ m.

Figure S1



RPA2 domain A

Figure S2

L. major MSSRS--YSAKLSSSAGESMVTSHSTGSSSRHRTGTVCIHFRAPPKMTAMELLOHIRNSITRLKYICERQYHTAALDQASEAVTSFFERLRKKLSKLEFSRSH
L. infantum MSSRS--YSAKLSSSAGESMVTSHSTGSSSRHRTGTAGIHFRAPPKMTAMERLQHIRNSITRLKYICERQYHTAALDQASEAVTSFFERLRKKLSKLEFSRSH
L. braziliensis MSSRS--YSAKLSSSAGESMVTSHSTGSSSRHRTGTGIIHFRAPPKMTAAERLQHVNSITRLKYICERQHHTAALDQAAEASVMTFFERLRKKLPRIEFSRSH
T. cruzi MSKNP--RSSHNSVGENYTSRHS--SVSSRRRRC--EKRRVLLPPANMTFFERLQLLRNALTROMYIGERHLSAMLDTAGGKVLFFEKLRRLPKLEYNPRK
T. vivax MSVKTS--TSSSHSSSAGENYTSRHS--CHSSRCRE--QHHGFSMLPPIKMTAERLKHVNSITRLRSTIGERHSHQELDTAGGNELEFFDGLRRLPKLEYNPRR
T. brucei MSTKNSVVGASGSHSTAGETVTSRHS--GTSSRHSRCRRQYDYVKNPSEPMTAMERLKLVRCSLARLRYLGERHRSKLDTTGGDRLLSFFEGRLRRLPKLEYNPRV
T. congolense MEFVGEC--SLGSRSTCEINTSGRS--GSSRRKPLARRRRERVAPPEPKAMEKQLVRCSLTRRLRYLCERYRHMTLDTDNELQNFERSLSDHMTLHHPRA

L. major IGTPEGGPQOLKYLANEARQLMSTACEVVRDTANISVTA TCVEGMOHHMQEAMAELEELSALLAERPNEQPSRVSDLLADLVAFHIDEFDAPMALRVEAFVREMOS
L. infantum IGTPEGGPQOLKYMANESRQLMSTACEVVRDTANISATATCVEGMOHHMQEAMAELEELSALLAERPNEQPSRVNADLLADLVAFHIDEFDAPMSLRVEAFAREMOS
L. braziliensis IGTPEGGPRLQYILKEVROLMSTACELVDRDTANASATATVEGMOHHMQEAMAELEELSVLLAEHPNEQPSRVNIDLLADLVAFHIDEFDAPMALRVEAFVREMOS
T. cruzi LGCPIGNGPKLOXYIREQTLIQDAANVVLMSCEAOKKCKDAVYEGLSYHYLQEVLRLEETIQHLLSHAPSEMPSPRVNTDOLLADLVSCFINEFDAPAEORVEAFVSEAKN
T. vivax LGAPEGGPKHLOXYITRAQNERLSEANASTLPTCEQLDLSAVAVYQCLAYHLOEAREHEHLVYLVSNPPSRMDSLENTLISDLVTPHITTEPDAPAEORVDAFIREMKN
T. brucei VAKSKVGA PKHLDYIRGHANALLGEASTITISECEQAMDRSEAVYEGLAYHLODTHRELENLVLLSGAPSDMESAGSKLLADLVSPHVAEFDAPAEORVPEFIQELKN
T. congolense ICKSIGKPRHLLHMNLVKSLLNEATTLIPRCEFRMRSLSESTYKGLGYHLODQVCRELEGLQVLLSRAPSEMSTFRSTELLFNLVSIHVDDFISLSNQRLLQSFVREVCN

L. major EGAFTYGPVMTDASQKATPVGVCSAVLLDVAVGVRPARSARIVLEHARHQISKVENRHAADIALQECQLAGEVAEAGQORQRAEVORREFTIMKERTRHAIQLDRVQQ
L. infantum EGAFTYGPVMTDASQKATPVGVCSAVLLDVAVGVRPARSARIVLEHARHQINKVENRHAADIALQECQAGEVAADGQORQRAEVORREFTIMKERTRHAIQLDRVQQ
L. braziliensis EGAFTYGPVMTDASQKATRIGVCSAVLLDVAVGVRPGHSAIVLDHARHNKIKVENRHAADIALQDQCPVGEAAVQQRQAEVORREFTIMKERTRHAIQLDRVQQ
T. cruzi ECTLFTAVVNELSQRAPSTGLVSI FVLDVAIGVRPGRCARIVIEHAQHNLTKLENRLHTAGSVLS-----AMT-SEKDLFTTORRELAILKDRVAGSIAEILRLKK
T. vivax EGVAYTAVVNEIHRPPKPVGLVSVFLLDVAIGVRPGCAKIVIEHAKHNLTKIENRIGLASQSLN-----VKESERNLQORREYTLIQERLTDVNLDDKVKI
T. brucei EGFLEFSVINDSQPFANAIGLLSVFIVDVAIGVRPGRSARLVIQHALHNLARLKS RVDKADITIG-----REESSNNEHHRMCSLEHEVIRGRLREKKEELERIRD
T. congolense EGALPHYAVKN-TEQSEVAVGLVSVFPEVAIGVFPKGSARIVIQAHHNLHKIESRLKALQVLE-----GTS-VPDDHAWRORRELKLLQDRVKESIGDLDRIKS

L. major AWLEEEQSIFTRAASGSHDLAALMOTKHLHEAEVIEEVRRLHDQTIHACLPPFATPQVHRRYGETYGPCHVRLGFHFLPDANTAAMAAAGS TNAAAFSSLMIRTLSTE
L. infantum AWLEEEQSIFTRAASGSHDLAALMOTKHLREAEVIEEVRRLHDQTIHACLPPFATPQVHRRYGETYGPCHVRLGFHFLPDANTAAMAAAGS ANAAAFSSLMIRTLSTE
L. braziliensis TWREEEQSIFTRA TSGSHDLVLMOTKHLREAEVIEEVLHDQTIHACLPPFATPQVHRRYGETYGPCHIRLGFHFLPDNTAAELAAASGTSAAAFSSLIIRTLSSSE
T. cruzi KWEEEEELTTERAASGQHVLLSLVMKRLKEAEIIEEAQLIHDRAMHSCILHSAPOYRRYS ETYAQPVYQIGKFKL-----CGEQVESQPKSMITLYALVPE
T. vivax EMTAEEEDTFRALSNKQSI PSLIRMKRLKESVVIHKVQNIHDCAVYSCLPFATPQVSRHSETYGPCHVRLGLKES-----TEERVDCQORSRTLYLMP
T. brucei KKDHDGDFISRACTVQGGQSIIIDAKRVKESVIEEAQAIHDRVYACLPFASPOVCHIRSETYSQPCHVRVGFKEV-----AKGDCSRSRTKTLTYLNLMP
T. congolense RWTADDEDLFRKASCKEQSLLILMETQARELAVVEEVQLHDHTVRACILFASPOVCRWHGETYSQPCHVRVIGYKFI-----GDDCGSQRVYQKISVLMPE

RPA2 domain A

Figure S3

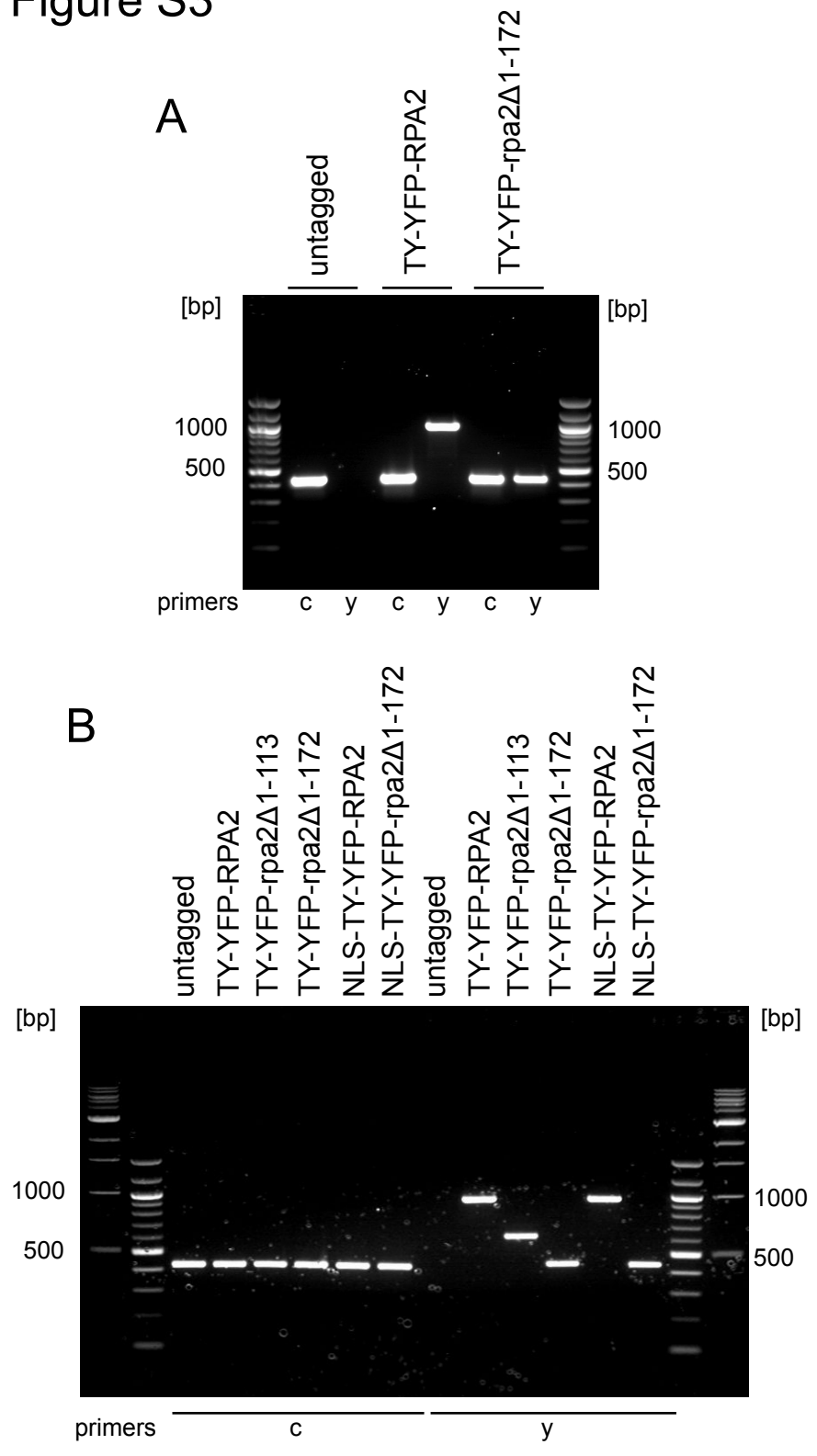


Figure S4

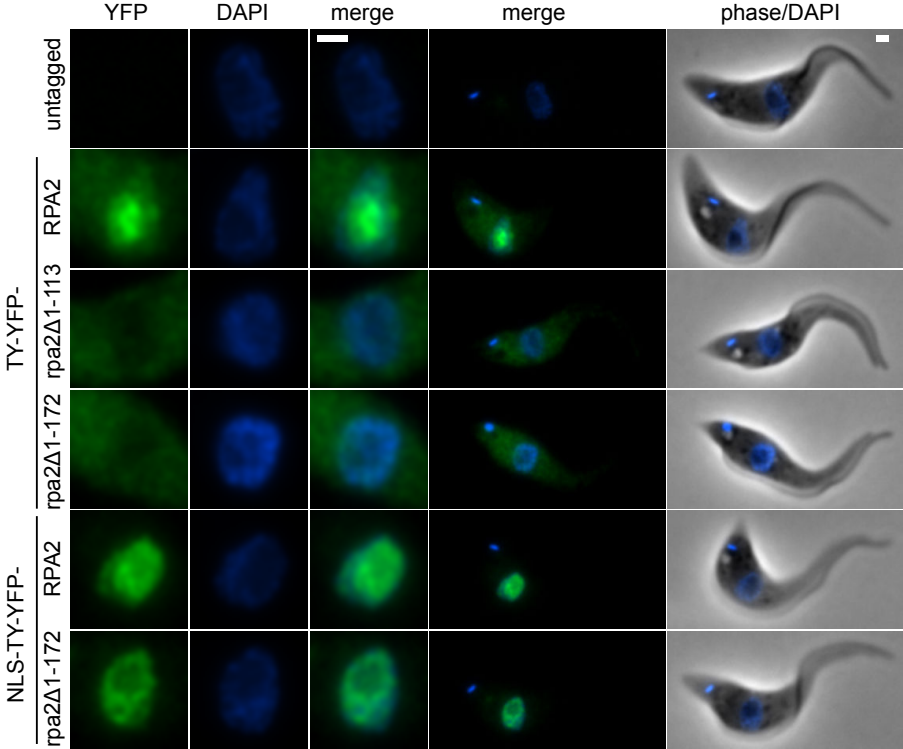


Figure S5

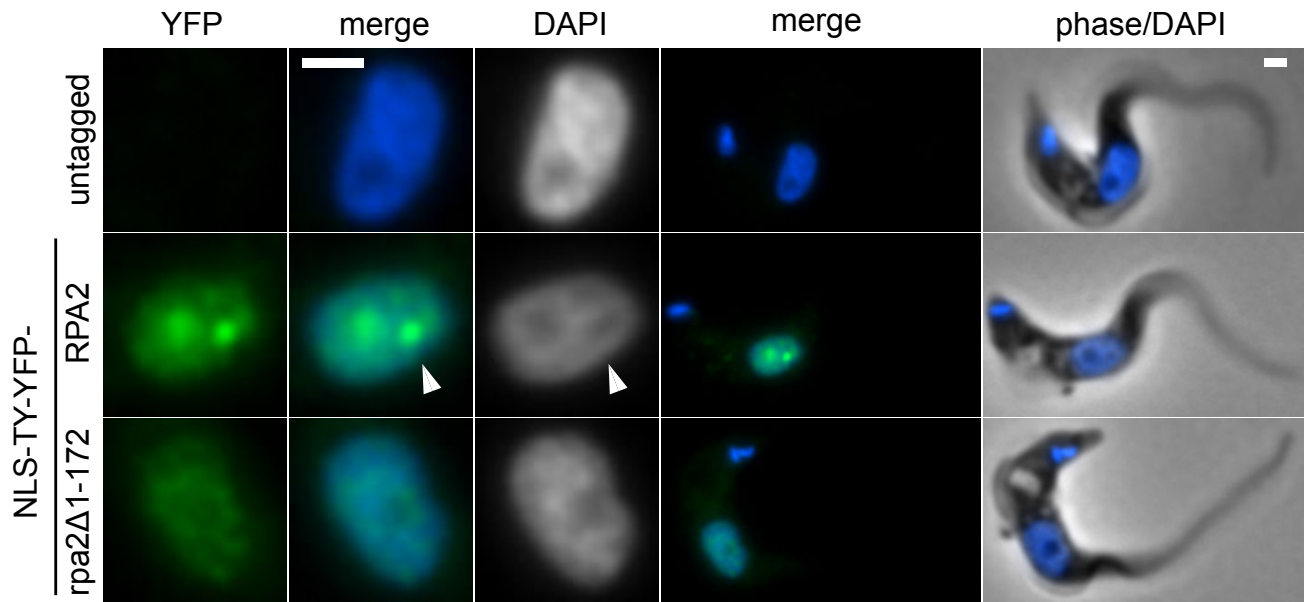


Table S1. NCBI and genome references of RPA2 protein data set.

Organism	Abbreviation	Protein ID (RPA2)
<i>Arabidopsis thaliana</i>	AraTh	NP_564341.2
<i>Caenorhabditis elegans</i>	CaeEl	NP_492476.1
<i>Cryptosporidium parvum</i>	CryPa	XP_628100.1
<i>Cyanidioschyzon merolae</i>	CyaMe	CMS114C
<i>Dictyostelium discoideum</i>	DicDi	XP_629008.1
<i>Drosophila melanogaster</i>	DroMe	NP_476708.1
<i>Entamoeba histolytica</i>	EntHi	XP_652347.1
<i>Giardia lamblia</i>	GiaLa	XP_001708931.1
<i>Homo sapiens</i>	HomSa	NP_061887.2
<i>Leishmania major</i>	LeiMa	XP_001683815.1
<i>Monosiga brevicollis</i>	MonBr	XP_001750532.1
<i>Naegleria gruberi</i>	NaeGr	55898
<i>Paramecium tetraurelia</i>	ParTe	XP_001456821.1
<i>Paramecium tetraurelia</i>	ParTe	XP_001430889.1
<i>Phaeodactylum tricornutum</i>	PhaTr	XP_002177384.1
<i>Physcomitrella patens</i>	PhyPa	XP_001773955.1
<i>Phytophthora ramorum</i>	PhyRa	42703
<i>Phytophthora sojae</i>	PhySo	108266
<i>Plasmodium falciparum</i>	PlaFa	XP_001348029.1
<i>Saccharomyces cerevisiae</i>	SacCe	NP_015335.1
<i>Schizosaccharomyces pombe</i>	SchPo	NP_595819.1
<i>Takifugu rubripes</i>	TakFu	SINFRUP00000142122
<i>Tetrahymena thermophila</i>	TetTh	XP_001030557.1
<i>Thalassiosira pseudonana</i>	ThaPs	XP_002293007.1
<i>Theileria annulata</i>	TheAn	XP_953002.1
<i>Toxoplasma gondii</i>	ToxGo	EEB04043.1
<i>Trichomonas vaginalis</i>	TriVa	XP_001318134.1
<i>Trypanosoma cruzi</i>	TryCr	Tc00.1047053506925.40
<i>Trypanosoma brucei</i>	TryBr	XP_828155.1

Table S2. Oligonucleotides used in this study

oligo name	sequence
SiMP1	
Ups_F	ctctogagctCTTTATCCGCCGAAAATTGGT
Ups_R	acatagatctGGGGAGCGGTGCCCTGTTT
Downs_F	cacatgatgcGTTCTCGTGATCGCTGCAC
Downs_R	acgagctcTCAGGTTCAGGTTCAAGGGTC
EP1-prom_F	ttaaagcttGTTGTCAATGGGGTTAAGCG
EP1-prom_R	ttggatccGTGCAGCGCATCACGAGAAC
Tet-Neo_F	CCGGTAATACGTCTCACTATA
Tet-Neo_R	ACGGGGCTACAATTAATACAT
linker_F	CAGATCTAAGCTTGTAAACCTCGA
linker_R	GATCTCGAGGTTAACCAAGCTTAGATCTGCATG
endogenous tagging	
RPA2Ser2_F	CCttaga TCAACAAAAGAAATTCGGTTGT
RPA2Ser2_R	AActcgagAA TAA TGGTCTGATGCCTCTC
RPA2Lys114_F	AAtctagaAAAAGTCGGTGCACCAAAG
RPA2Lys114_R	AActcgagGGCAGGTTGTTGAGAATCAT
RPA2Gly173_F	TTtctagaGGGTTGCTATCTGGTGC
RPA2Gly173_R	GGctcgagGATCGGC TTTGTCAACAC
RPA2Glu183_F	GatctagaGAGAGCGCTGGATCTTTG
RPA2Glu183_R	GActcgagTCTTAAACGGCCACCGAAT
RPA25UTR_F	TTctcgagCCCAGTTGTGAGGTTGTTG
RPA25UTR_R	TTggatccTTTGCACGC TTCACGC
NLS cloning	
LaNLSTYoligo_F	agcttATGCGCGGTCAACAAGCGCTCCCGCTCCCGCGA
LaNLSTYoligo_R	ctagfGTCAAGTGGATCTTGGTTAGTAGTGGACCTCCTC
ectopic expression	
XbaRPA2Nt_F	ATtctagaTCAACAAAAGAAATTCGGTTGTA
BamTAARPA2Nt_R	ATggatccTTATACTAAGTTCTCCAACCTCGC
genomic sequencing	
gS2_F	CTGCTGGAGTTCGTGAC
gG173_R	CCTGTGATGTTCCGTTATTGCTA

Table S3. GeneDB references for additional trypanosome RPA2 proteins.

Organism	Abbreviation	Protein ID (RPA2)
<i>Leishmania infantum</i>	LeIn	LinJ25_V3.0630
<i>Leishmania braziliensis</i>	LeiBr	LbrM25_V2.2130*
<i>Trypanosoma vivax</i>	TryVi	Tviv811a10.p1k_4
<i>Trypanosoma congolense</i>	TryCo	congo1390g02.q1k_4

*) We modified the start codon of this protein from the annotated to an in-frame start codon further upstream. This resulted in the full-length protein displayed in Fig. S1.

Table S4. Sources of predicted proteomes used in this study.

Organism	Source	Version	Web reference	Download filename
<i>Arabidopsis thaliana</i>	TAIR	TAIR7	www.arabidopsis.org/	TAIR7_pep_20070425
<i>Caenorhabditis elegans</i>	WormBase	WS179	www.sanger.ac.uk/Projects/C_elegans/	wormpep170.tar.gz
<i>Chlamydomonas reinhardtii</i>	JGI	v3.0	www.jgi.doe.gov/	proteins.Chlre3.fasta.gz
<i>Cryptosporidium parvum</i>	CryptoDB	v3.4	www.cryptodb.org/	CparvumAnnotatedProtein.fsa.zip
<i>Cyanidioschyzon merolae</i>	C.merolae genome project	?	merolae.biol.s.u-tokyo.ac.jp/	cds.fasta
<i>Dictyostelium discoideum</i>	dictyBase	05-20-2007	dictybase.org/	dicty_predictions_protein.gz
<i>Drosophila melanogaster</i>	ENSEMBL	BDGP4.3, 46.43b	www.ensembl.org/info/data/	Drosophila_melanogaster.BDGP4.3.46.pep.all.fa.gz
<i>Entamoeba histolytica</i>	geneDB	17102005	www.genedb.org/	GeneDB_Ehistolytica_Proteins_17102005
<i>Giardia lamblia</i>	GiardiaDB	April 17, 2007	www.mbl.edu/Giardia	orfs_aa.fas.gz
<i>Homo sapiens</i>	VEGA	Jun2006	vega.sanger.ac.uk/	Homo_sapiens.VEGAJun.pep.tot.fa.gz
<i>Leishmania major</i>	geneDB	v5.2	www.genedb.org/	LmjFwholegenome_20060901_V5.2.pep
<i>Monosiga brevicollis</i>	JGI	v1.0	www.jgi.doe.gov/	Monbr1_best_proteins.fasta.gz
<i>Naegleria gruberi</i>	JGI	v1.0	www.jgi.doe.gov/	Naegr1_best_proteins.fasta.gz
<i>Paramecium tetraurelia</i>	ParameciumDB	v1.11	paramecium.cgm.cnrs-gif.fr/	Ptetraurelia_peptides_v1.04.fasta
<i>Phaeodactylum tricomutum</i>	JGI	v2.0	www.jgi.doe.gov/	Phatr2_geneModels_FilteredModels_aa.fasta.gz
<i>Physcomitrella patens</i>	JGI	v1.1	www.jgi.doe.gov/	proteins.Phypta1_FilteredModels3.fasta.gz
<i>Phytophthora sojae</i>	JGI	v1.1	www.jgi.doe.gov/	proteins.FM_Physo1_1.fasta.gz
<i>Phytophthora ramorum</i>	JGI	v1.1	www.jgi.doe.gov/	proteins.FM_Phyra1_1.fasta.gz
<i>Plasmodium falciparum</i>	PlasmoDB	v5.3	www.plasmodb.org/	PfalciparumAnnotatedProteins_plasmoDB-5.3.fasta
<i>Saccharomyces cerevisiae</i>	ENSEMBL	SGD1.01.46	www.ensembl.org/info/data/	Saccharomyces_cerevisiae.SGD1.01.46.pep.all.fa
<i>Schizosaccharomyces pombe</i>	Sanger	v19?	www.sanger.ac.uk/Projects/S_pombe	pompep
<i>Takifugu rubripes</i>	ENSEMBL	Assembly4, 46	www.ensembl.org/info/data/	Takifugu_rubripes.FUGU4.46.pep.all.fa.gz
<i>Tetrahymena thermophila</i>	TIGR	10/24/2006	www.tigr.org/tdb/e2k1/ta1/	TTA1.pep
<i>Thalassiosira pseudonana</i>	JGI	v3.0	www.jgi.doe.gov/	Thaps3_geneModels_FilteredModels_aa.fasta.gz
<i>Theileria annulata</i>	GeneDB	v4	www.genedb.org/	TANN.GeneDB.pep
<i>Toxoplasma gondii</i>	ToxoDB	v4.1	www.toxodb.org/	TgondiiAnnotatedProteins_toxoDb-4.1.fasta
<i>Trichomonas vaginalis</i>	TIGR	20050331	www.tigr.org/tdb/e2k1/tvg/	02152007_all_model_aa.fasta
<i>Trypanosoma brucei</i>	geneDB	v4	www.genedb.org/	Tb927_Proteins_May06_v4.fas.gz
<i>Trypanosoma cruzi</i>	T cruziDB	v5.0	www.tcruzidb.org	TcruziAnnotatedProtein.fas