

Supplementary information for:

A lineage-specific protein network at the trypanosome nuclear envelope

Erin R. Butterfield¹, Samson O. Obado^{2†}, Simon R. Scutts^{3‡}, Wenzhu Zhang⁴, Brian T. Chait⁴, Michael P. Rout² and Mark C. Field^{1,5*}

¹School of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, UK,

²The Rockefeller University, 1230 York Avenue, New York, NY10065, USA,

³Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1EF, UK, ⁴Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York, USA, ⁵Biology Centre, Czech Academy of Sciences, Institute of Parasitology, České Budějovice, Czech Republic.

†Current address: Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, USA

‡Current address: Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Melbourne, VIC 3052, Australia

*Corresponding author: Telephone: +44 (0)751-550-7880, email: mfield@mac.com, mfield@dundee.ac.uk

Supplementary Results

Identification of LAP orthologs: LAP orthologs were primarily identified through best reciprocal BLAST¹ (BRB) and iterative HMMER² searching of the predicted proteomes. However, as LAP sequences were incomplete or absent from many Kinetoplastid organisms we performed additional searches against genome sequences, alternative strains and assemblies and analysed raw read data.

The *L. donovani* LAP71 ortholog is missing the N-terminal region due to a mispredicted start site, which is rectified by an upstream in-frame methionine. We were unable to identify a *S. culicis* ortholog from the predicted proteome, however,

tBLASTn searches against the *S. culicis* TCC012E assembly (GCA_000482145)³, identified a full-length LAP102 ORF. No *A. deanei* LAP102 ortholog was initially identified in the predicted proteome, however, batch_brb⁴ searching using the *E. monterogei* LAP102 ortholog as query and relaxed coverage (20%) against the updated *A. deanei* assembly (GCA_903995115)⁵ identified a highly divergent putative LAP102 ortholog.

We were unable to detect a *T. cruzi* LAP73 ortholog in the Dm28c strain, however, a full-length ortholog is present in the TCC strain (GCA_003177095.1)⁶. The predicted protein sequence for the *Phytomonas* sp. Hart1 LAP73 ortholog is missing the N-terminal, this is due to a misprediction of the start site, using an upstream in-frame methionine results in a full-length ORF and a Nup53-type RRM can be detected. We were unable to identify a LAP73 *Phytomonas* sp. EM1 ortholog from the predicted proteome, searches against the genome identify an orthologous region although the C-terminal of the protein appears to be missing likely due to an unsequenced region located directly downstream. As there were multiple possibilities for the *Phytomonas* sp. EM1 ORF start site we predicted the start based on conservation with the *Phytomonas* sp. Hart1 start site.

The *B. ayala* LAP92 ortholog is missing the N-terminal due to a misprediction of the start site, utilizing an upstream in-frame methionine rectifies this. We were unable to identify an *E. monterogei* LAP92 ortholog from the predicted proteome, tBLASTn searches against the genome identifies an orthologous region in the genome producing a single full-length ORF.

The *S. culicis* LAP102 ortholog is missing the C-terminal and contains an unsequenced region, however, a full-length LAP102 ORF was identified in the *S. culicis* TCC012E assembly (GCA_000482145)³. The *E. monterogei* LAP102 is encoded by two adjacent genes. Analysis of the RNA-seq reads from Albanaz et al.⁷ identifies an insertion in the *E. monterogei* genome (52142 C), correction of this results in a single LAP102 ORF. As the *P. confusum* LAP102 candidate (PCON_0077700) contains large additional N and C-terminal regions (relative to the other kinetoplastid orthologs) and a large unsequenced region, we processed and assembled the polyA RNA-seq reads from Skalický et al.⁸. Searching the assembled transcriptome using batch_brb we were able to identify a single LAP102 candidate. Translation of this transcript identified seven candidate starting methionines, with four

possessing a candidate upstream splice leader site⁹. We therefore predicted the start site based on conservation with other kinetoplastid sequences.

A second *T. brucei* sequence is predicted for LAP333, but as it contains only the N-terminal of the protein and is in the genome bin it is likely due to sequencing/assembly errors. Two orthologs are predicted in *T. cruzi* Dm28c for LAP333, these genes are adjacent in the assembly and comparison of these to the TbLAP333 sequence shows one is the N-terminal and the other is the C-terminal of the protein. A single gene is predicted in an updated *T. cruzi* Dm28c genome assembly (GCA_002219105; BCY84_22033)¹⁰⁻¹² and a single gene is predicted in the *T. cruzi* CL Brener Esmeraldo-like strain. ((GCA_00209065.1)^{13,14}, suggesting the fragmented prediction is due to sequencing/assembly errors. In *Phytomonas* sp. *Hart1* the N-terminal is missing from the predicted LAP333 ortholog. tBLASTn searches of the genome identified the N-terminal in a different reading frame separated from the remainder of the gene by an unsequenced region. No *Phytomonas* sp. *EM1* ortholog could be identified in the predicted proteome, tBLASTn searches of the genome identified regions of homology across different reading frames which are separated by unsequenced regions. The *S. culicis* predicted ortholog is missing the C-terminal. While searches of the *S. culicis* genome identified more of the sequence, the contig ends before the C-terminal of the sequence. We were instead able to identify a full-length sequence in the TCC012E *S. culicis* assembly (GCA_000482145)³. Initial searches against the predicted proteome from *A. deanei* identified a partial LAP333 ortholog. tBLASTn searches against the updated *A. deanei* assembly (GCA_903995115)⁵ showed a region of homology to LAP333 which is predicted as four partial adjacent ORFs and no predicted ORF covering the N-terminal region. The absence of an N-terminal ORF is due to a misprediction of the start site for ADEAN_000926000, while analysis of the Illumina genome reads from Davey et al.⁵ identified several indels (LR877166: 368562, (370304,5), 370880, 370227, 370235, 370238, (370248,9), (370256,7)) leading to the fragmented gene prediction. Correction of these produces a single ORF which is supported by the identification of a full-length LAP333 ortholog in an alternate *A. deanei* assembly (GCA_001659865¹⁵; Scaffold3003_contig_10:15021-24341 (LXWQ01000889)). In *B. saltans* LAP333 is predicted to be fragmented into two adjacent genes. Analysing the RNA-seq reads from Jackson et al.¹⁶ we identified multiple regions of poor mapping against the genome. We reassembled the transcriptome with Trinity¹⁷ and analysed with

batch _brb⁴, identifying three transcripts encoding full-length LAP333 orthologs in *B. saltans*. Alignment of these to the genome identifies several indels in the genomic sequence (Supplementary Figure S27).

To confirm kinetoplastid restriction of LAP333 we performed additional BRB searches against the EukProt TCS database¹⁸ (excluding *Nonionella stella*) (Figure 5, Supplementary Figure S16). We found the *T. brucei*, *T. cruzi* and *L. major* hits previously identified in addition to two partial *T. cruzi* LAP333 orthologs (P017989 and P008050) reflecting the Non-Esmeraldo haplotype¹³ N and C terminals of LAP333. We were unable to identify orthologs in the early-branching *Neobodo designis* or any of the other Discoba organisms. We identified partial orthologs in TSAR organisms *Telonema* sp. *P-2*, *Colponema vietnamica* and *Colponemidia* sp. *Colp-10* – orthologous to the C-terminal of LAP333 and containing at least one C-terminal transmembrane domain. Using *Telonema* sp. *P-2* as query for BRB, further putative orthologs were identified in *T. subtile*, and *Telonema* sp. *P-2*. However, as these organisms were cocultured in the presence of kinetoplastid organisms (*Procryptobia sorokini* for Telonemia¹⁹ and *Parabodo caudatus* for Colponemidae²⁰) the detection of partial LAP333 orthologs likely represents contamination. Furthermore, these results suggest the presence of a partial LAP333 in Parabodonida although further work will be required to confirm this is a full-length gene.

A second LAP59 sequence is predicted in *T. brucei*, however, as the N-terminal is missing and the sequence is located in the genome bin, it is likely this is due to sequencing/ assembly errors. We were unable to identify a full-length *S. culicis* ortholog due to unsequenced regions and the end of the contig in two respective *S. culicis* assemblies (GCA_000442495²¹ and GCA_000482145³), however, a full-length sequence is present in the related *Strigomonas galati* (GCA_000482125.1)³ which contains two N-terminal transmembrane domains and a C-terminal cytochrome B561 domain. The *A. deanei* and *E. monterogei* predicted proteins (EPY21985 and EMOLV88_060005900 respectively) are missing the N-terminal trans-membrane domains, due to a misprediction of the start site. Correction to an upstream in-frame Methionine identifies the N-terminal trans-membrane domains (two in *A. deanei* and one in *E. monterogei*). We were unable to identify an Alveolate LAP59 ortholog by sequence searching, however, two putative orthologs were identified in the ciliate *Stentor coeruleus* (A0A1R2AVT9 and A0A1R2CN18) by searching the InterPro database²² for Alveolate proteins containing N-terminal transmembrane domains and

a C-terminal cytochrome B561-related domain (IPR019176). These sequences identified LAP59 orthologs in reverse BLAST and the C-terminal domains of the AlphaFold^{23,24} structures align to the *T. brucei* LAP59 C-terminal domain (RMSD 1.31 and 1.28 respectively).

Due to the presence of a Sac3 domain in LAP173, additional filtering of short sequences (< 1000 aa) was performed to prevent misidentification of Sac3 domain containing proteins as LAP173 orthologs. We were, therefore, unable to initially identify orthologs in *T. cruzi*, *S. culicis* or *A. deanei*. However, additional searching of alternative strains did identify full-length sequences in the *S. culicis* TCC012E (GCA_000482145)³ and *T. cruzi* CL Brener Esmeraldo-like (GCA_00209065.1)^{13,14} strains. For *A. deanei* tBLASTn searches against the updated assembly (GCA_903995115)⁵ identified a region of homology partially covered by two predicted genes. Analysis of the genomic reads from Davey et al.⁵ identified several indels (LRR877162: 97187, (98387, 98388), 98642, 100294) in the genomic sequence, correction of which resulted in a full-length gene prediction. No orthologs were identified in *Phytomonas* sp. despite additional tBLASTn searches against their genomes.

References

1. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped BLAST And PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997; 25:3389–402.
2. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* 2013; 41:e121.
3. Alves JMP, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, Vasconcelos ATR, Sagot M-F, Teixeira MMG, Motta MCM, et al. Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evol Biol* 2013; 13:190.
4. Butterfield ER, Abbott JC, Field MC. Automated phylogenetic analysis using best reciprocal BLAST. In: de Pablos, LM, Sotillo, J, editors. *Parasite Genomics: Methods in Molecular Biology*. New York: Humana; 2021. page 41–63.
5. Davey JW, Catta-Preta CMC, James S, Forrester S, Motta MCM, Ashton PD, Mottram JC. Chromosomal assembly of the nuclear genome of the endosymbiont-bearing trypanosomatid *Angomonas deanei*. *G3: Genes, Genomes, Genetics* 2021; 11:jkaa018.
6. Berná L, Rodriguez M, Chiribao ML, Parodi-Talice A, Pita S, Rijo G, Alvarez-Valin F, Robello C. Expanding an expanded genome: long-read sequencing of *Trypanosoma cruzi*. *Microb Genom* 2018; 4:e000177.
7. Albanaz ATS, Gerasimov ES, Shaw JJ, Sádlová J, Lukeš J, Volf P, Opperdoes FR, Kostygov AY, Butenko A, Yurchenko V. Genome analysis of *Endotrypanum* and *Porcisia* spp., closest phylogenetic relatives of *Leishmania*, highlights the role of amastins in shaping pathogenicity. *Genes (Basel)* 2021; 12:444.
8. Skalický T, Dobáková E, Wheeler RJ, Tesařová M, Flegontov P, Jirsová D, Votýpková J, Yurchenko V, Ayala FJ, Lukeš J. Extensive flagellar remodeling during the complex life cycle of *Paratrypanosoma*, an early-branching trypanosomatid. *Proc Natl Acad Sci U S A* 2017; 114:11757–62.

9. Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, Roditi I, Ochsenreiter T. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog* 2010; 6:e1001037.
10. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* 2009; 38:D457–62.
11. TriTrypDB. BCY84_22033 [Internet]. [cited 2019 May 14]; Available from: https://tritrypdb.org/tritrypdb/app/record/gene/BCY84_22033
12. TriTrypDB. Data Set: Genome Sequence and Annotation [Internet]. [cited 2019 May 14]; Available from: https://tritrypdb.org/tritrypdb/app/record/dataset/DS_4c287690ee
13. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, Ghedin E, Worthey EA, Delcher AL, Blandin G, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas Disease. *Science* 2005; 309:409–15.
14. Weatherly DB, Boehlke C, Tarleton RL. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics* 2009; 10:255.
15. Morales J, Kokkori S, Weidauer D, Chapman J, Goltsman E, Rokhsar D, Grossman AR, Nowack ECM. Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid *Angomonas deanei*. *BMC Evol Biol* 2016; 16:247.
16. Jackson AP, Otto TD, Aslett M, Armstrong SD, Bringaud F, Schlacht A, Hartley C, Sanders M, Wastling JM, Dacks JB, et al. Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Current Biology* 2016; 26:161–72.
17. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011; 29:644–52.
18. Richter DJ, Berney C, Strassert JFH, Poh Y-P, Herman EK, Muñoz-Gómez SA, Wideman JG, Burki F, de Vargas C. EukProt: A database of genome-scale

predicted proteins across the diversity of eukaryotes. Peer Community Journal 2022; 2:e56.

19. Strassert JFH, Jamy M, Mylnikov AP, Tikhonenkov D V., Burki F. New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. *Mol Biol Evol* 2019; 36:757–65.
20. Tikhonenkov D V., Strassert JFH, Janouškovec J, Mylnikov AP, Aleoshin V V., Burki F, Keeling PJ. Predatory colponemids are the sister group to all other alveolates. *Mol Phylogenet Evol* 2020; 149:106839.
21. Motta MCM, Martins AC de A, de Souza SSA, Catta-Preta CMC, Silva R, Klein CC, de Almeida LGP, de Lima Cunha O, Ciapina LP, Brocchi M, et al. Predicting the proteins of Angomonas deanei, Strigomonas culicis and their respective endosymbionts reveals new aspects of the Trypanosomatidae family. *PLoS One* 2013; 8:e60209.
22. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 2021; 49:D344—D354.
23. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021; 596:583–9.
24. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022; 50:D439–44.
25. DeepMind. AlphaFold Colab [Internet]. Available from: <https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>
26. Linding R, Russell RB, Nedvva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 2003; 31:3701–8.
27. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods* 2022; 19:679–82.

28. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2022;
29. Archer SK, Inchaustegui D, Queiroz R, Clayton C. The cell cycle regulated transcriptome of *Trypanosoma brucei*. *PLoS One* 2011; 6:e18425.
30. Holden JM, Koreny L, Obado SO, Ratushny A V., Chen W-M, Chiang J-H, Kelly S, Chait BT, Aitchison JD, Rout MP, et al. Nuclear pore complex evolution: A trypanosome Mlp analogue functions in chromosomal segregation but lacks transcriptional barrier activity. *Mol Biol Cell* 2014; 25:1421–36.
31. Marques CA, Ridgway M, Tinti M, Cassidy A, Horn D. Genome-scale RNA interference profiling of *Trypanosoma brucei* cell cycle progression defects. *Nat Commun* 2022; 13:5326.
32. DuBois KN, Alsford S, Holden JM, Buisson J, Swiderski M, Bart J-M, Ratushny A V., Wan Y, Bastin P, Barry JD, et al. NUP-1 is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. *PLoS Biol* 2012; 10:e1001287.
33. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012; 61:539–42.
34. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010; 59:307–21.
35. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; 32:1792–7.
36. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009; 25:1189–91.
37. Bolte S, Cordelières FP. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* 2006; 224:213–32.
38. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. Fiji: An open-source platform for biological-image analysis. *Nat Methods* 2012; 9:676–82.

39. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 2.5.2. 2015;
40. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 2001; 98:10037–41.
41. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 2007; 35:W522–5.
42. PyMOLWiki. Color h [Internet]. 2019 [cited 2023 Jun 26]; Available from: https://pymolwiki.org/index.php/Color_h
43. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* 1984; 179:125–42.
44. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2019; 20:1160–6.
45. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 2017; 45:D200–3.
46. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 2004; 338:1027–36.
47. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res* 2007; 35:429–32.
48. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a Hidden Markov Model: Application to complete genomes. *J Mol Biol* 2001; 305:567–80.

Supplementary figure legends

Supplementary figure S1: LAP71 and LAP92 AlphaFold predictions. LAP71 (**A**) and LAP92 (**B**) structures were predicted using the DeepMind monomer Colab notebook^{23–25} with default settings. Structures are coloured by pLDDT where dark blue is very confident (pLDDT > 90), cyan is confident (90 > pLDDT > 70), yellow is low confidence (70 > pLDDT > 50) and orange is very low confidence (pLDDT < 50). pLDDT and predicted aligned error plots are shown for each structure.

Supplementary figure S2: LAP333 globular regions used for structural predictions with AlphaFold DeepMind monomer models. The LAP333 protein sequence was analysed using the default settings of GlobPlot²⁶ to determine the globular regions to use these for AlphaFold^{23,24} predictions without the relaxation stage. Uppercase letters indicate globular regions predicted by GlobPlot²⁶. Coloured boxes indicate sequences used for AlphaFold^{23,24} DeepMind monomer predictions²⁵.

Supplementary figure S3: LAP333 AlphaFold DeepMind monomer models. AlphaFold DeepMind monomer model^{23–25} was used to predict the structure of LAP333 fragments (Supplementary Figure S2). Structures are coloured as Supplementary figure S1. pLDDT and predicted aligned error plots are shown for each structure.

Supplementary figure S4: LAP333 globular regions used for structural predictions with AlphaFold ColabFold monomer models. The LAP333 protein sequence was analysed using the default settings of GlobPlot²⁶ to determine the globular regions to use these for AlphaFold^{23,24} predictions without the relaxation stage. Uppercase letters indicate globular regions predicted by GlobPlot²⁶. Coloured boxes or black outlines indicate sequences used for AlphaFold^{23,24} ColabFold monomer predictions²⁷.

Supplementary figure S5: LAP333 ColabFold monomer models. AlphaFold ColabFold monomer model^{23,24,27} was used to predict the structure of LAP333 fragments (Supplementary Figure S4). Structures are coloured by pLDDT as

Supplementary figure S1. pLDDT and predicted aligned error plots are shown for each structure.

Supplementary figure S6: LAP333 DeepMind and ColabFold multimer models.

AlphaFold^{23,24} DeepMind²⁵ and ColabFold²⁷ multimer²⁸ models were used to predict the structure of LAP333. Structures are coloured by pLDDT as per Supplementary figure S1. pLDDT and predicted aligned error plots are shown for each structure.

Supplementary figures S7, S8: LAP333 DeepMind and ColabFold multimer PAE plots with domains respectively. Identified LAP333 domains were plotted onto the AlphaFold^{23,24} DeepMind²⁵ and ColabFold²⁷ multimer²⁸ models predicted aligned error plots. Red lines indicate region boundaries. Coloured boxes along the scales indicate the type of domain identified. Orange: signal peptide, yellow: Ig-like or α/β fold, green: Ig-like fold, black: possible Ig-like fold, blue: coiled-coil, pink: region containing three *trans*-membrane domains. Nuclear localization signal is not shown. Graphs show the low confidence of interdomain interactions.

Supplementary figure 9: Cell cycle analysis of LAPs. **(A)** Proyclic transcription data for the LAPs from Archer et al.²⁹, showing that expression of LAPs is fairly consistent across the cell cycle. LAP92 – a basket Nup which has moonlighting involvement with the mitotic spindle³⁰ – is included for reference. **(B)** Bloodstream RNA interference target sequencing (RIT-seq) cell cycle data for the LAPs from Marques et al.³¹ showing that LAPs don't show major cell cycle defects. NUP-1 (a known disruptor of the cell cycle)³², NUP-2 and Nup92 are included for reference.

Supplementary figures S10, S11, S13, S17: Phylogenetic analysis of the LAPs. Predicted orthologs were phylogenetically analysed using MrBayes³³ and PhyML³⁴. Resulting trees were combined into a single tree. Node values show the MrBayes³³ posterior probabilities and PhyML Bootstrap. – indicates differing topology between the two methods.

Supplementary figure S12: Alignment of LAP73 orthologs with *S. culicis* homolog. Predicted LAP73 orthologs and homologs were aligned with MUSCLE³⁵

and visualized in Jalview³⁶ with the Clustal colour scheme. Box shows the predicted *T. brucei* LAP73 Nup53-type RNA recognition motif. Alignment shows limited conservation of the *S. culicis* sequence, making its assignment as a LAP73 ortholog unclear.

Supplementary figure S14: LAP102 phylogeny. Predicted orthologs were phylogenetically analysed using MrBayes³³ and PhyML³⁴. Resulting trees were combined into a single tree. Node values show the MrBayes³³ posterior probabilities and PhyML Bootstrap. *E. monterogei* is represented by the illumina corrected genome sequence. *P. confusum* is represented by the ORF predicted from the Trinity¹⁷ assembled transcript TRINITY_DN2138_c0_g1_i2 using a Kinetoplastid conserved start site. Due to the high levels of conservation the editing frequency was decreased (gaps allowed in 75% of sequences) to improve phylogenetic signal.

Supplementary figure S15: Alignment of LAP102 orthologs showing *P. confusum* insertions. Predicted LAP102 orthologs and homologs were aligned with MUSCLE³⁵ and visualized in Jalview³⁶ with the Clustal colour scheme. *E. monterogei* is represented by the illumina corrected genome sequence. *P. confusum* is represented by the ORF predicted from the Trinity¹⁷ assembled transcript TRINITY_DN2138_c0_g1_i2 using a Kinetoplastid conserved start site. Alignment shows large *P. confusum* insertions relative to other Kinetoplastid sequences.

Supplementary figure S16: LAP333 phylogeny. Predicted orthologs were phylogenetically analysed using MrBayes³³ and PhyML³⁴. Resulting trees were combined into a single tree. Node values show the MrBayes³³ posterior probabilities and PhyML Bootstrap. *B. saltans* is represented by the Trinity¹⁷ assembled transcript TRINITY_DN4963_c0_g1_i6, *A. deanei* is represented by the illumina corrected genome sequence. *Phytomonas sp. EM1* and *Hart1* were excluded from phylogenetic analysis due to the presence of large unsequenced regions.

Supplementary figure S18: LAP173 phylogeny. Predicted orthologs were phylogenetically analysed using MrBayes³³ and PhyML³⁴. Resulting trees were combined into a single tree. Node values show the MrBayes³³ posterior probabilities

and PhyML Bootstrap. *A. deanei* is represented by the illumina corrected genome sequence.

Supplementary figure S19: LAP173 alignment. Predicted ortholog sequences were aligned using MUSCLE³⁵ and the alignment was visualized using Jalview³⁶. Colours are the Clustal colour scheme. *A. deanei* is represented by the illumina corrected genome sequence. The alignment shows conservation of the Sac3 domain and C-terminal region across the eukaryotes supporting the identification of the kinetoplastid sequences as possible Sac3 orthologs.

Supplementary figure S20: Colocalization of the LAP protein cohort with the NPC. Pearson's correlation coefficients for the colocalization of selected LAPs with FG nups of the NPC were calculated for regions of interest surrounding the nucleus using the BIOP JACoP plugin³⁷ in FIJI³⁸. Values greater than 0.6 (dashed line) or below -0.6 (dashed line) indicate positive and negative correlations respectively (and thereby colocalization or exclusion respectively). Table indicates the number of cells counted for each stage of the cell cycle. For post mitotic cells both nuclei were assessed resulting in the graph showing double the cell number (indicated as a “*” in the table). The data indicates some colocalization of LAPs 71 and 102 with FG Nups while no colocalization is seen between LAP73 and the FG Nups.

Supplementary figure S21: Colocalization of selected LAPs with NUP-1. Pearson's correlation coefficients for the colocalization of the LAPs with NUP-1 were calculated for regions of interest surrounding the nucleus using the BIOP JACoP plugin³⁷ in FIJI³⁸. Values greater than 0.6 (dashed line) or below -0.6 (dashed line) indicate positive and negative correlations respectively (and thereby colocalization or exclusion respectively). Table indicates the number of cells counted for each stage of the cell cycle. For post mitotic cells both nuclei were assessed resulting in the graph showing double the cell number (indicated by a “*” in the table). The data indicates there is some colocalization of the LAPs with NUP-1 although this is lost for LAP102 during post mitosis and only present for LAP73 during interphase. This is consistent with the previously described relocalisation of NUP-1 during mitosis to the distal regions of the separating daughter nuclei³².

Supplementary figures S22 – 25: AlphaFold multimer modelling of LAP333 DeepMind fragments with LAP59. AlphaFold DeepMind multimer^{25,28} was used to predict the interactions between LAP333 fragments and LAP59. LAP333 fragment sequences used were the LAP333 DeepMind monomer fragment sequences. **(A)** LAP333 fragment complexed with LAP59 with colours indicating protein (brown: LAP333, blue: LAP59). **(B)** LAP333 fragment complexed with LAP59 with colours indicating pLDDT as per Supplementary figure 1. For (A) and (B) ‘ indicates LAP59. **(C)** pLDDT plots for LAP333 fragment and LAP59. **(D)** predicted aligned error plot where the red lines indicate the end of the LAP333 fragment.

Supplementary figure S26: Electrostatic and hydropathy properties for the LAP333 F2B and LAP59 interaction site. **(A)** Surface electrostatic charge for the LAP333 F2B and LAP59 interaction site was calculated using the PyMOL³⁹ APBS^{40,41} plugin. LAP333 F2B model shown in brown, LAP59 model shown in blue. Images show the LAP333 F2B interaction site is primarily negative while the LAP59 interaction site is primarily positive. **(B)** Surface hydropathy for the LAP333 F2B and LAP59 interaction site was calculated using the color_h PyMOL script^{42,43}. LAP333 F2B and LAP59 models coloured as per (A). Images show the interaction site for both molecules are primarily hydrophobic.

Supplementary figure S27: Alignment of *B. saltans* LAP333 transcripts against the genome. The three LAP333 *B. saltans* transcripts identified by batch_brb⁴ searching the Trinity¹⁷ assembled *B. saltans* transcriptome were aligned to the *B. saltans* genome (CYKH01000671:26230-35710) using MAFFT⁴⁴ and visualized in Jalview³⁶ with the Nucleotide colour scheme. Alignment shows several indels in the genomic sequence relative to the Trinity assembled transcripts.

Supplementary table legends

Supplementary table S1: List of organisms and sources used for the phylogenetic analysis of the LAPs.

Supplementary tables S2, S3: ESI LC-MS results for the analysis of total sample for NUP-1 and NUP-2 co-immunoprecipitation experiments respectively.

Supplementary tables S4: NCBI CDD search⁴⁵, Phobius^{46,47} and TMHMM⁴⁸ were used to predict the domains for LAP59.

Supplementary table S5: List of LAP orthologs identified.

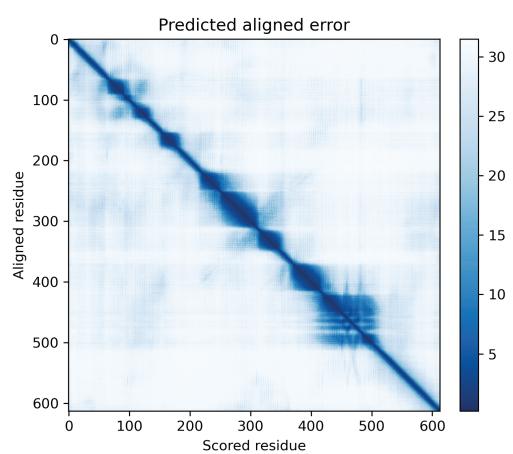
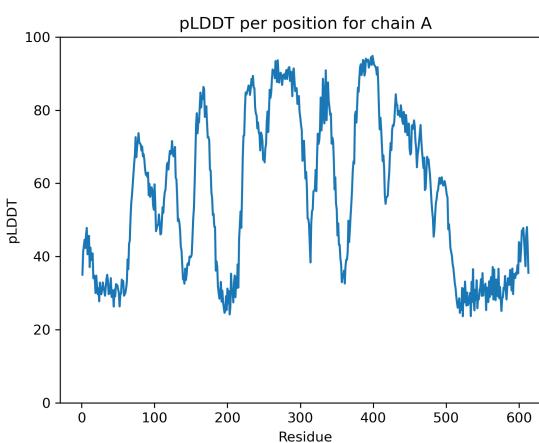
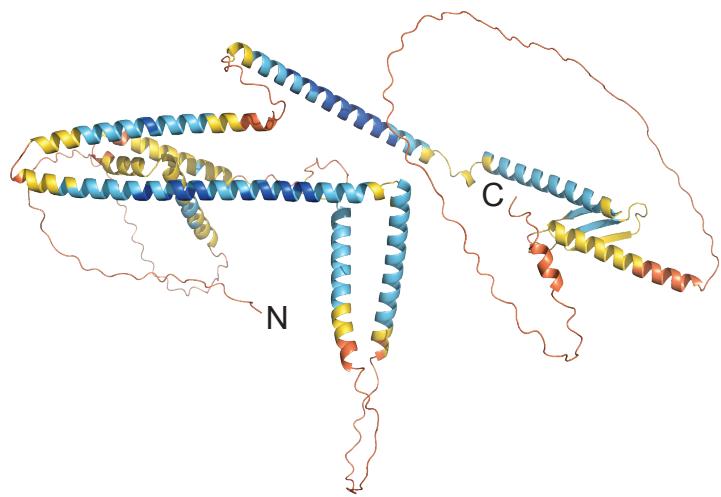
Supplementary table S6: Protein sequences of LAP orthologs.

Supplementary tables S7: ESI LC-MS results for the analysis of total sample for LAP102.

Figure S1

A

LAP71



B

LAP92

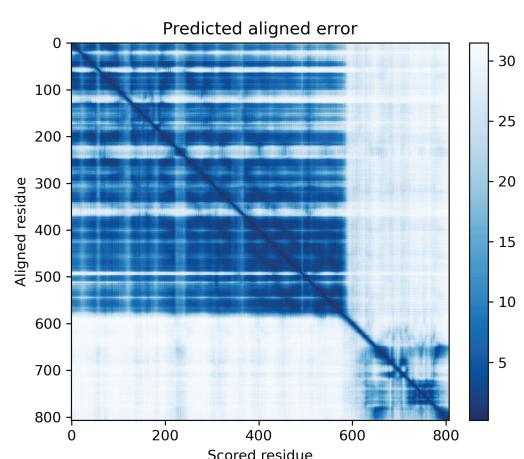
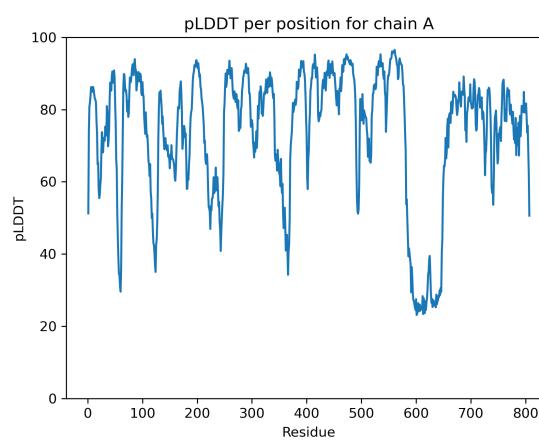
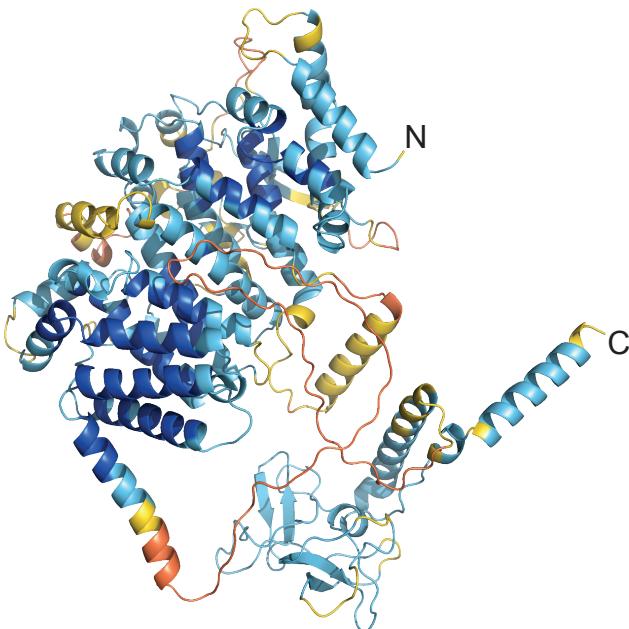


Figure S2

LAP333 sequences used for Fragmented DeepMind AlphaFold structural predictions

>LAP333_GlobDoms 76-201, 249-485, 504-612, 691-947, 1127-1338, 1382-1663, 1683-2319, 2334-2481, 2548-2880, 2899-3031
tmtrvgdgv lsrppfaflt vsviaalat hafveasvea tphnlselne qlfksseflv pnvssspcsv hhgcpHASMV WATGTFTAPD VLTPASFLVS ANLTQLNLC CGYKFNGRLL REQSRYDTFL YDLRSTPFH
TTLGEKRRM FPRDRVVIFD TQRQPPHYRE SSHVPSAEM QRGNQDGDFV IVLVTGAVGP Grrrvplqqp pttalnlpl sglrggfdmp lgsrtgnssv gsgsavdgAV GTSADIQLRV SVAIKCKCP HTLLVMPVLH
ARPDTLPPIS IVGLGSHCGV GQKLRSVTPE MDVKGKVLRQ PVIQLANGW AASLTLPQAL LPATLVHII TDTTAVTGA NAVLNVSLYT ALREEVEGSG MVQPQITTSV YFSVIPHHTY NVMTVSRGA ATQTLSGLFV
LRLPLTLEGK AAQFSNTVQP HFGVGRRLRL SRITREREEL KQAFAMLPDE GRPSHTGYWI RTQTSdevrg sdgspphph npfSEAFWY PLTVGRNVLS YYVRDMDEV C EPVLLEMQEL WAVQALLLEP VTSPVTVCDD
RISLEAFELR GLLETMRSRQ RMLIRLAGDL PWEDLARSGG TQEVEGHFK CTesgvpcr iatgmggshf gvlqmlkrgv aelmwvvrv e vdtstylqhie qlrqrqgsdv vdddndnegrp fggfldpsls RVISSDGRLT
VIEWRSPSVL IVRMSALPR RVDVKLPME RRLTVRPPHP VIGQHEESLD ERDKDTDIQI TGRWVNNPI SLISVDETYA AELNTSAVPN IRQVLPEGTL QFIAEDPGEA VQDKLLAAPY TSRCPPRVVE YIVSLVAVPA
PPVAPEFVPV YTEDGCVALS PPREPQKDEE VRWDIRPCQE ASAPPYFSKE KQKVDTLKGY NGSADALVVV CGVEVDRWCE AMWSARNLVG SVTQTFSLRR CQRPHLLfh kfsgdrrgs gspidtlalp svevldggit
tgdesqhtvt rsgdnsnttd etmvhheaqp ryfarrvipi agyaavhyt fnysgsqgdy evdfayehda ahqlhvtash fartglavsd drsgksdepn tftpdssaa attagplllr rfvrpttvv sgshehgsng
kdnneg IFSD EAAYWEMERH HRDYNIDREC DARFSHNNSQ TGQYMQHLHG SNSSNYPFQY NVRLSLAAVA AAAELPSRKG AAAEFARRGM CLSRPGNISF LALNPRLV D VIGVRDIAFC ESTFNFSRHP WLHTKYHYDF
TWTCENTATG RSSTDNTAD GVMMVEKRSP TFSILYQGSA GATPCKLTVY NALTETLESE TFMLKRVHPA PVTLNFSVls dnwapgnpm rrsgrqsnsf gsqrsttagk eggeplgkhq sSAAGMRTHW IVGKRAVLHV
DAPADGRTSQ WRAVTVPAAE TTSVGTQRLL DSIQIPLRQ PHPRAILRGA GGATVLLNNI RVPGMYTLSH QHLDPVHPNV CPNVTVTETL HVTHARVLEN ELTVCGDTAV LTAALPQAI NCSFTTRWEV VSVQTADNIT
LTDETNFSGI KFEDVSSPHT VVKGLPGIV TVRWSVYASA RTTDDERRQ KGQPGILVDY DTVRLLVLT NLRTTRLVLT SSSVEMLITG STLGWVLVEP ADRRASALLS ARDPTLRVGR VYKnttsgj ggggnnsgrs
edAPEETALF VGNLPVGTT LQYELHASPR LFGAMLRKRC PFVQQRLLTI ERVSAFLSVN TTLDHECDAY SGRGRFSICL HAEGIKPASS SNAVAAQEK P SSSNLGVHLV APGDFENSIK SSAERLTASG CSLPWVART
TRWSDIKVDN SRQCISFTVI ATEHHLQSKE VSHHLFLAQE HLSMSQPCHH TGDFVPSEV IEGASTWSVG GSAEDSSWFR FRSRPAAEAS DGLSKFLPFM NPSTLVREDR AMRPLLQLG VADIVASELD ISERTYMHWE
RSNATDRGAG MKNGSRSDV VFHLDRSQPG RMASFLHATS SHPYAGQKGE GSHDHHQERE VMESFLAKCV RGTLRGSEVL TYWTEVNHPD GAAGGTAPEW EYVVRHVGGA EVTCEPLFAS LEQLLRLPDD TESTLRSRRA
SFVPSKEQKF VSQEVVEQQI KALREEFRRT REGIDVTGLR PEVQQALAR RITKSQLD EKARREKA EK DRQHAQQQQ LVLTAATAAM TAEERFVAAA RLYEGVTGEG VYAIRVKLPY YASFSPHNL FLRTALHKDL
ICGTNISVTS TAPASLVLFS RVEVMDEPPA VGASPPAIKQ CEVEGRVPTI FINVSGTRFV SGGAFPPQNV LLEELQNPDV ieegggpegs tagVGLASCL AQMRAGMDVS VDSATRKHLR LQLSPCPSFQ MLSPIEYNEQ
QESQQLLKRL QLPRLLRITM RDALVQDSHT QGRTTISFVV EVRPTLALLH LRLSLPSQGDS LPVVAATGR KRPLLLCEAD MRRKGVELGV ELIGDTWNEH Vgedeegaan pgdygyssdd ydadds snap ivegsggwav
vnagsslnt rlrpgqsegr sragvgkkr SPHSVALLNS FSIVEHHSFD SVIHLFLQQG GVYRSHFVSY FFSEVLKDGP NGQQQRLIYR YNDTFIGVRI TSERNMQKDT LSSGTHDKYY MGASLVVDDG TNLLPPIRPF
STIEEILFAV PGIATECQGC LLADSTFFVA GDLTGLWSSP SMLQTRYLQS SQPARSTFSA PLRIPFYDAA VVPREWRFTT MCPTAALVAL RSLWHSWNQ GVQPFVLMEL VVYSLRDMLG GEVPLARRPA AILTSVSPV
GHLLKQTHVK TAQGEEMIPE IMFNVNTSAA ARAYAMMAAD FVRYGRAHV SVHDYVRDEP QSGGALAVNG SVFLKVCVPA iphsstrgdg tpsqvnldPQ CSVVSFTHSL LLMEEPYAGA PYASWYRILW GLIYVPLLL
HHVVTAVVGY AALQQLPVR LLLLRSQPQL HTASIGLTTM LFVALLFLCP LVEWLWVSLV LSLWAFVLYH ERRLLATAIVM IIHGLLWTFI L



Fragment 1



Fragment 2A



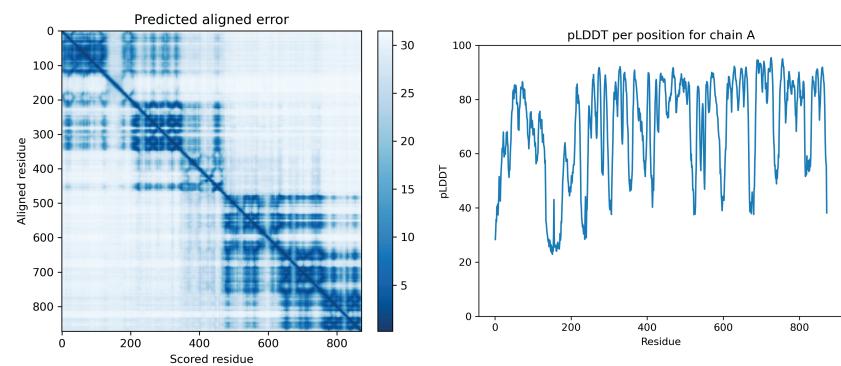
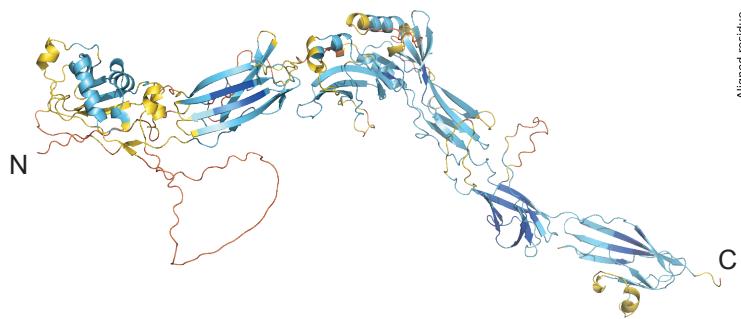
Fragment 2B



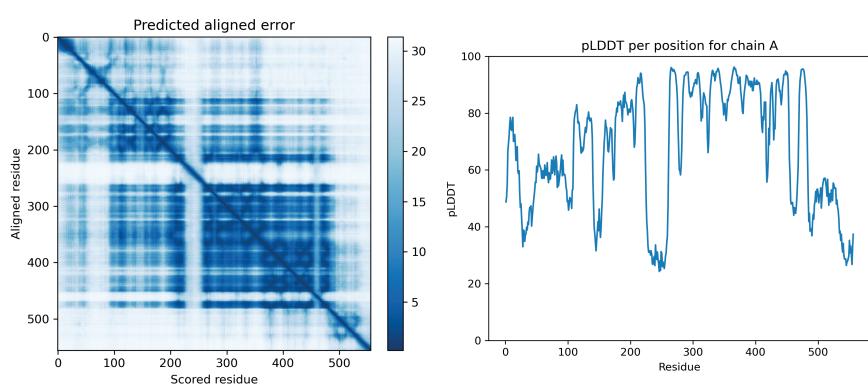
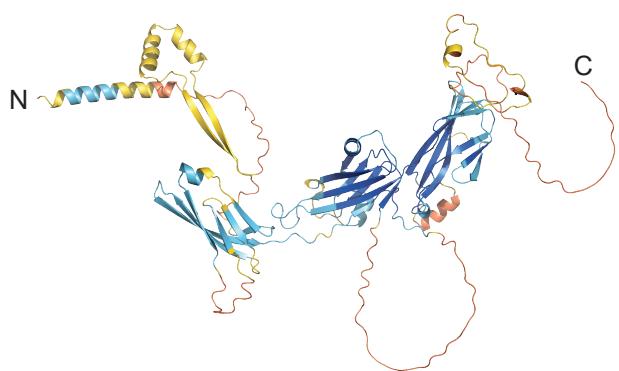
Fragment 3

Figure S3

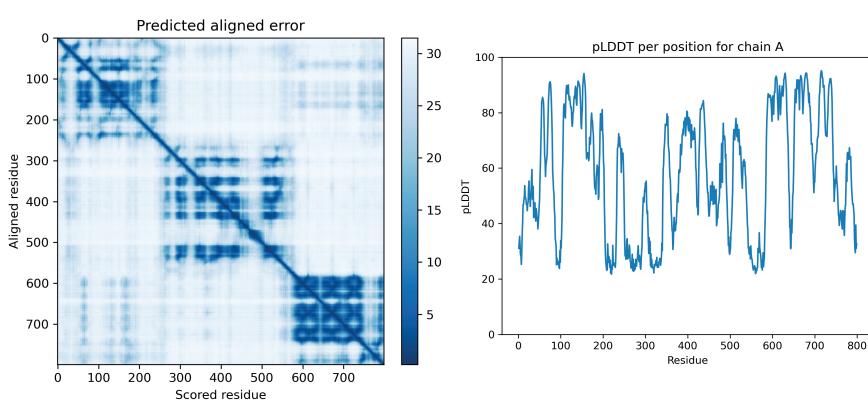
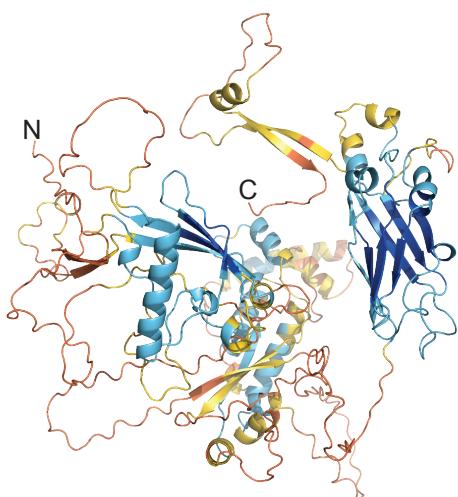
LAP333 Fragment 1 DeepMind



LAP333 Fragment 2A DeepMind



LAP333 Fragment 2B DeepMind



LAP333 Fragment 3 DeepMind

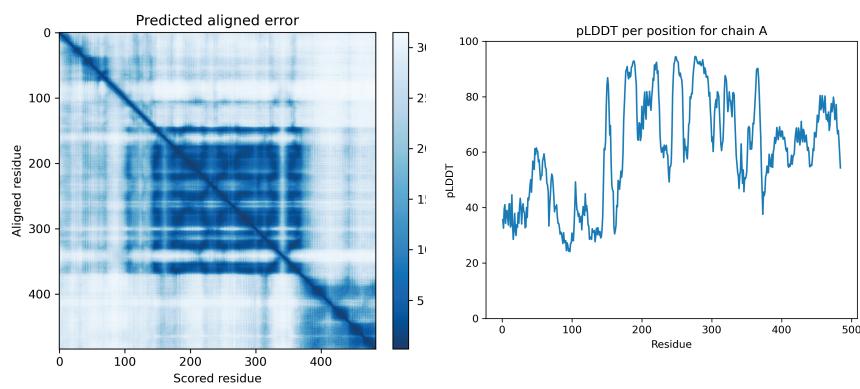
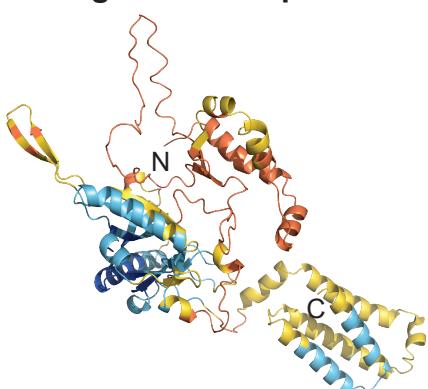


Figure S4

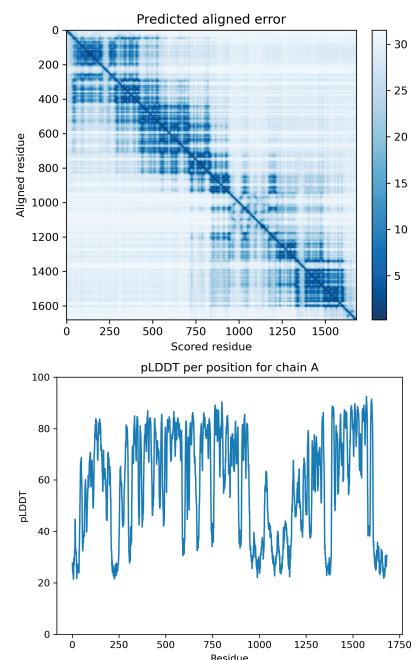
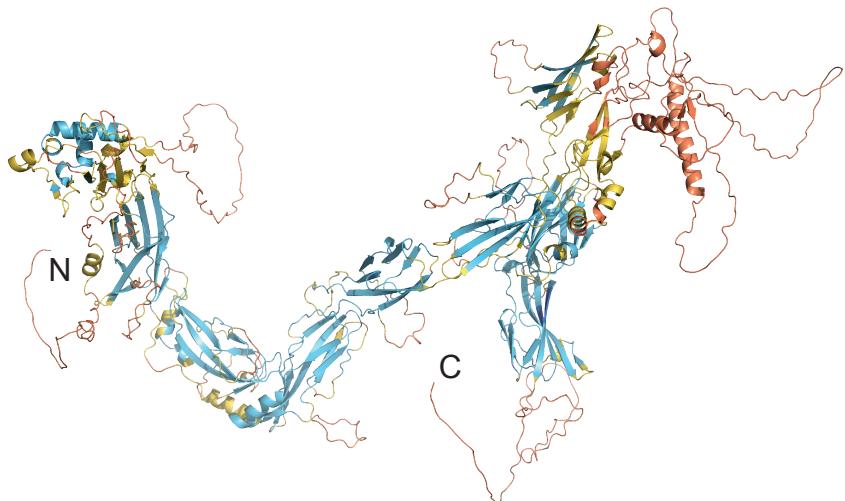
LAP333 sequences used for Fragmented ColabFold AlphaFold structural predictions

>LAP333_GlobDoms 76-201, 249-485, 504-612, 691-947, 1127-1338, 1382-1663, 1683-2319, 2334-2481, 2548-2880, 2899-3031
tmtrvgdgv lsrppfaflt vsviaalat hafveasvea tphnlsele qlfksseflv pnvssspcsv hhgcpHASMV WATGFTTAPD VLTPASFLVS ANLTQLNLNC CGYKFNGRLL REQSRYDTFL YDLLRSTPFH
TTLGEKRRM FPRDRVVIFD TQRQPPHRYE SSHVPSAEM QRGNNDGDDV IVLVTGAVGP Girrvplqqp pttalnlpl sglrrggfdmp lgsrtgnssv gsgsavdgAV GTSADIQLRV SNVAIKCKCP HTLLVMVPLH
ARPDTLPPIS IVGLGSHCGV GQKLRLSVPTE MDVKGVLRQ PVIQLANGW AASLTLPAL LPATLVHWII TDTTAVTGAA NAVLNVSLYT ALREEVEGSG MVQPQITTSQ YFSVIPHHTY NVMVTVSRGA ATQLSGLFV
LRLPLTLEGK AAQFSNVTQP HFGVGRRRLR SRITREREEL KQAFAMLPDE GRPSHTGYWI RTQTSdevrg sdgspphp npfSEAFVWY PLTVGRNVLS YYVRDMDEV C EPVLLEMQEL WAVQALLLEP VTSPVTVCDD
RISLEAFELR GLLETMRSRQ RMLIRLAGDL PWEDLARSQG TQEVEGHFK CTesvgvpqr iatgmggshf gvlqmlkrgv aelmwvvrv e vdttsylqhie qlrqrqgsdv vdddndnepg fggfldpsls RVISSDGRLT
VIEVRSPSVL IVERRMSALPR RVDVKLPMHE RRLTVRPPHP VIGQHEESLD ERDKDTDIQT TGRWVNPPSI SLISVDETYA AELNTSAVPN IRQVLPEGTL QFIAEDPGEA VQDKLLAAPY TSRCPPRVVE YIVSLVAVPA
PPVAPEFVVPV YTETDGCVALS PPREPQKDEE VRWDIRPCQE ASAPPYFSKE KQKVDLTKGY NGSADALVVV CGEVDRWCE AMWSARNLVG SVTQTFSLRR CQRPPHllfh kfsgddrrgs gspidtlalp svewlldgit
tgdesqhtvt rsgdnsnttd etmvhheadp rvfarryipi agvaaavhyt fnvsgsgqdy evdfayehda ahglhvash fartglavsd drsgksdepn tftdpddsaa attagplllr rfvrpptvgv sshehgsgn
kdnnes IFSD EAAYWEMERH HRDYNIDREC DARFSHNNSQ TGQYMQLHGL SNSSNYPFQY NVRLSLAAVA AAAELPSRKG AAAEFARRGM CLSRPGNISF LALNPRGLVD VIGVRDIAFC ESTFNFSRHP WLHTKYHYDF
TWTCENTAG RSSTDNTAD GVMMVEKRSP TFSILYQGSA GATPCKLTVY NALTETLESE TFMLKRVHPA PVTLNFSVls dnwapgnpcm rrsgrqsnsf gsqsrttagk eggeplghq sSAAGMRTHW IVGKRAVLHV
DAPADGRTSQ WRAVTVVPAE TTSVGTQRLR DSIQIIPLRQ PHPRAILRGA GGATVLLNNI RVPGMYTLSH QHLDPVHPNV CPNVTVTETL HVTHARVLEN ELTVCGDTAV LTAAPLPQAI NCSFTTRWEV VSVQTADNIT
LTDETNFSGI KFEDVSSPHT VVKGLPFGIV TVRWSVYASA RTTDDDERRQ KGQPGILVVDY DTVRLLVLT E NLRTTRLVTL SSSVEMLITG STLGWVLVEP ADRRASALLS ARDPTLRVGR VYKnqttsgj ggggnhsgsrs
edAPEETALF VGNLPVGTTT LQYELHASPR LFGAMLRKRC PFVQQRLLTI ERVSAFLSVN TTLDHECDAY SGRGRFSICL HAEGIKPASS SNAVAAQEK P SSSNLGVHLV APGDFENS IK SSAERLTASG CSLPWVART
TRWSDIKVDN SRQCISFTVI ATEHHLQSKE VSHHFLQA E HLSMSQPCHH TGGDFVPSEV IEGASTWSVG GSAEDSSWFR FRSRPAAEAS DGLSKFLPFM NPSTLVREDR AMRPLLQLG VADIVASELD ISERTYMHWE
RSNATDRGAG MKNGRSRSVDV VFHLDRSQPG RMASFHLATS SHPYAGQKGE GHSDHHQERE VMESFLAKCV RGTLRGSEVL TYWTEVNHPD GAAGGTAPEW EYVVRHVGGA EVTCEPLFAS LEQLLRLPDD TESTLRSRRRA
SFVPSKEQKF VSQEVEVQQK KALREEFRLT REGIDVTGLR PEVQQALARR RITKSQLDEI EKARREKAEK DRQHAQQQQQ LVLTAATAAM TAEERFVAAA RLYEGVTGEG VYAIRVKLPY YASFSVPHNL FLRTALHKDL
ICGTNISVTS TAPASLVLFS RVEVMDEPPA VGASPPAIK CEVEGRVPTI FINVSGTRFV SGGAFPPQNV LLEELQNPDV ieegggpegs tagVGLASCL AQMRAGMDVS VDSATRKHLR LQLSPCPSFQ MLSPIEYNEQ
QESGQLLKRL QLPRLLLRTM RDALVQDSHT QGRTTISFVV EVRPTLALLH LRSLPSQGDS LPVVAATGR KRPLLC EAD MRRKGVELGV ELIGDTWNEH V jgedeegaan pgdygyssdd ydadds snap ivesggwaw
vnagsslnt rlrpgqsegr sravvgkRKR SPHSVALLNS FSIVEHHFSD SVHHLFLQQG GVYRSHFVSY FFSEVLKD PG NGQQQR LIYR YN DFIGVRI T SERNMQKDT LSSGTHDKYY MGASLVVDDG TNLLPPIRPF
STIEEILFAV PGIAATECQGC LLADSTFFVA GDLTGLWSSP SMLQTRYLQS SQPARSTFSA PLRIPFYDAA VVPREWRFTT MCPTAALVAL RSLWHSWNNQ GVQPFLVMEL VVYSLRDMLG GEVPLARRPA AI LTSVSVPL
GHLLKQTHVK TAQGEEMIPE IMFNVNTSAA ARAYAMMAAD FVRYGRAHVH SVHDYVRDEP QSGGALAVNG SVFLKVCVPA iphsstrgdg tpsqvnldPQ CSVVFTHSL LLMEEPYAGA PYASWYRILW GLIYYPLLLL
HHVVTAVVGY AALQQLPVR LLRLRSPQDL HTASIGLTTM LFVALLFLCP LVEWLWVSLV LSLWAFVLYH ERRLATAIVM IIHGLLWTFI L

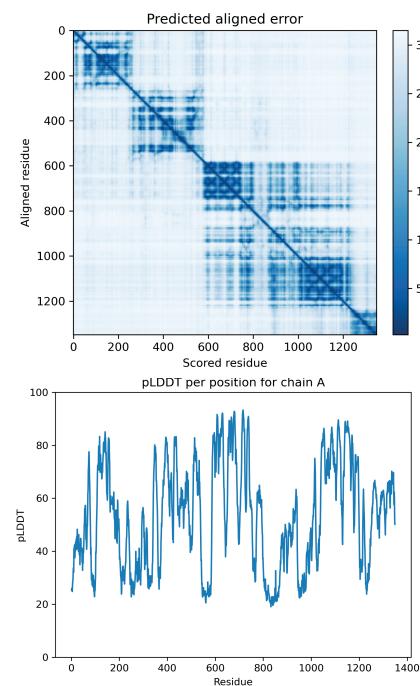
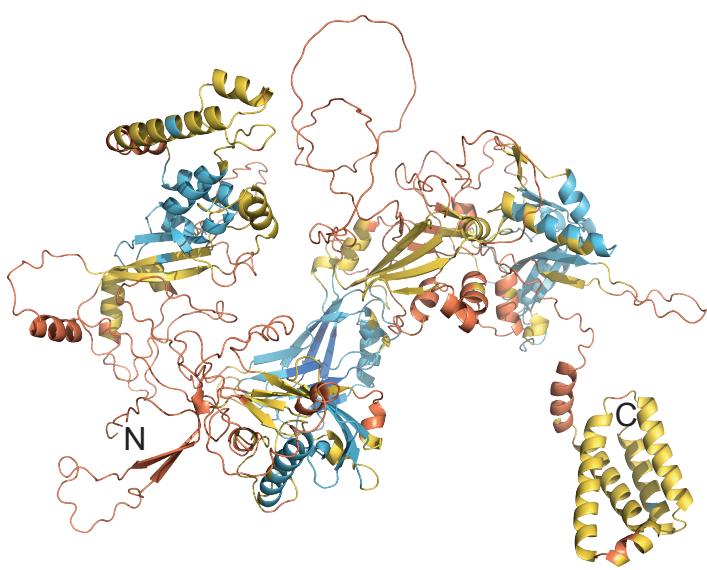


Figure S5

LAP333 N terminal ColabFold



LAP333 C terminal ColabFold



LAP333 middle ColabFold

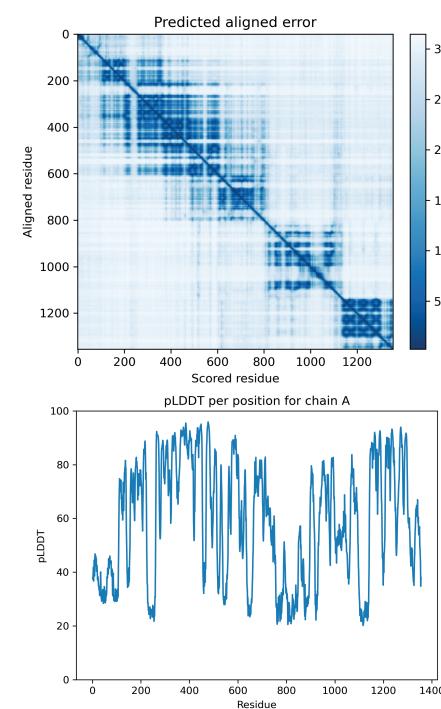
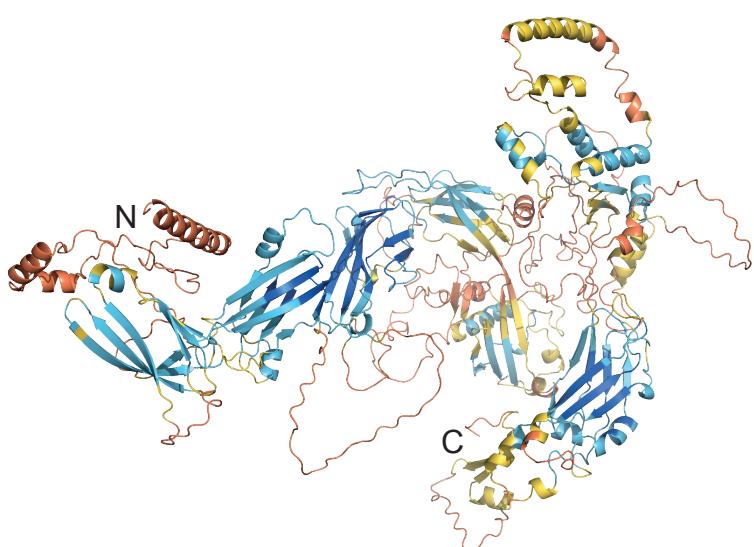
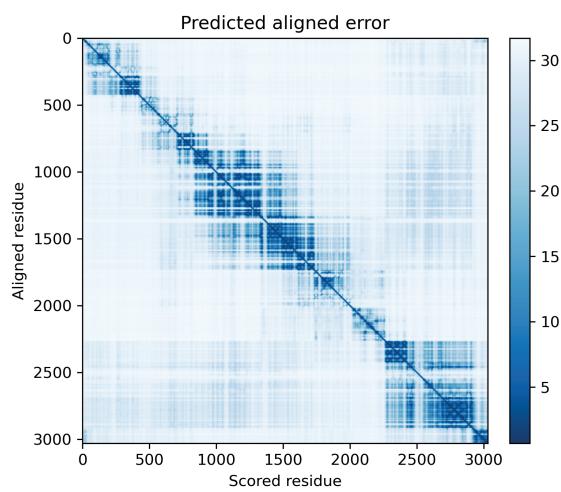
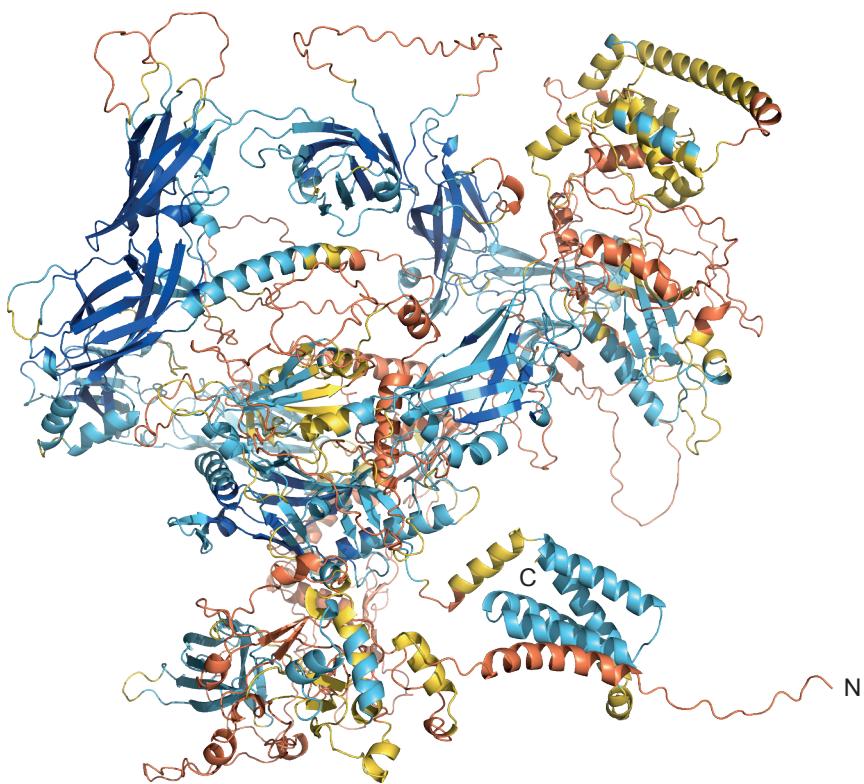


Figure S6

LAP333 full-length DeepMind



LAP333 full-length ColabFold

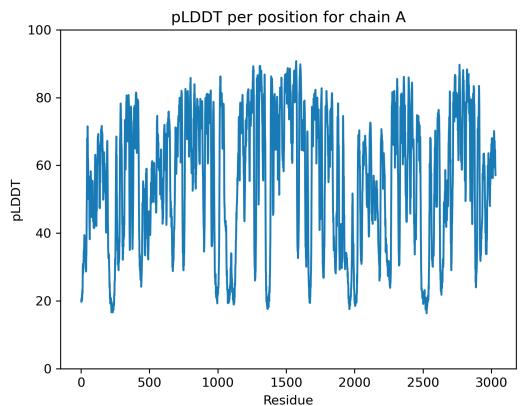
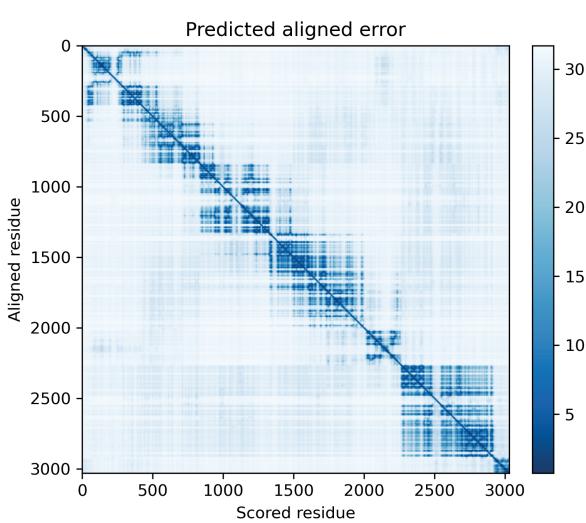
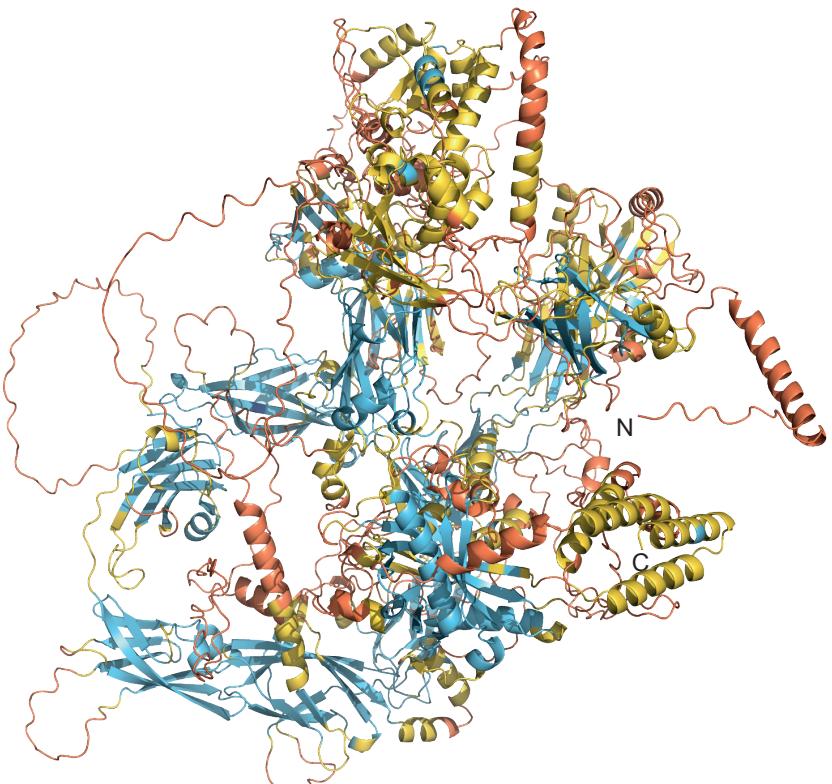


Figure S7

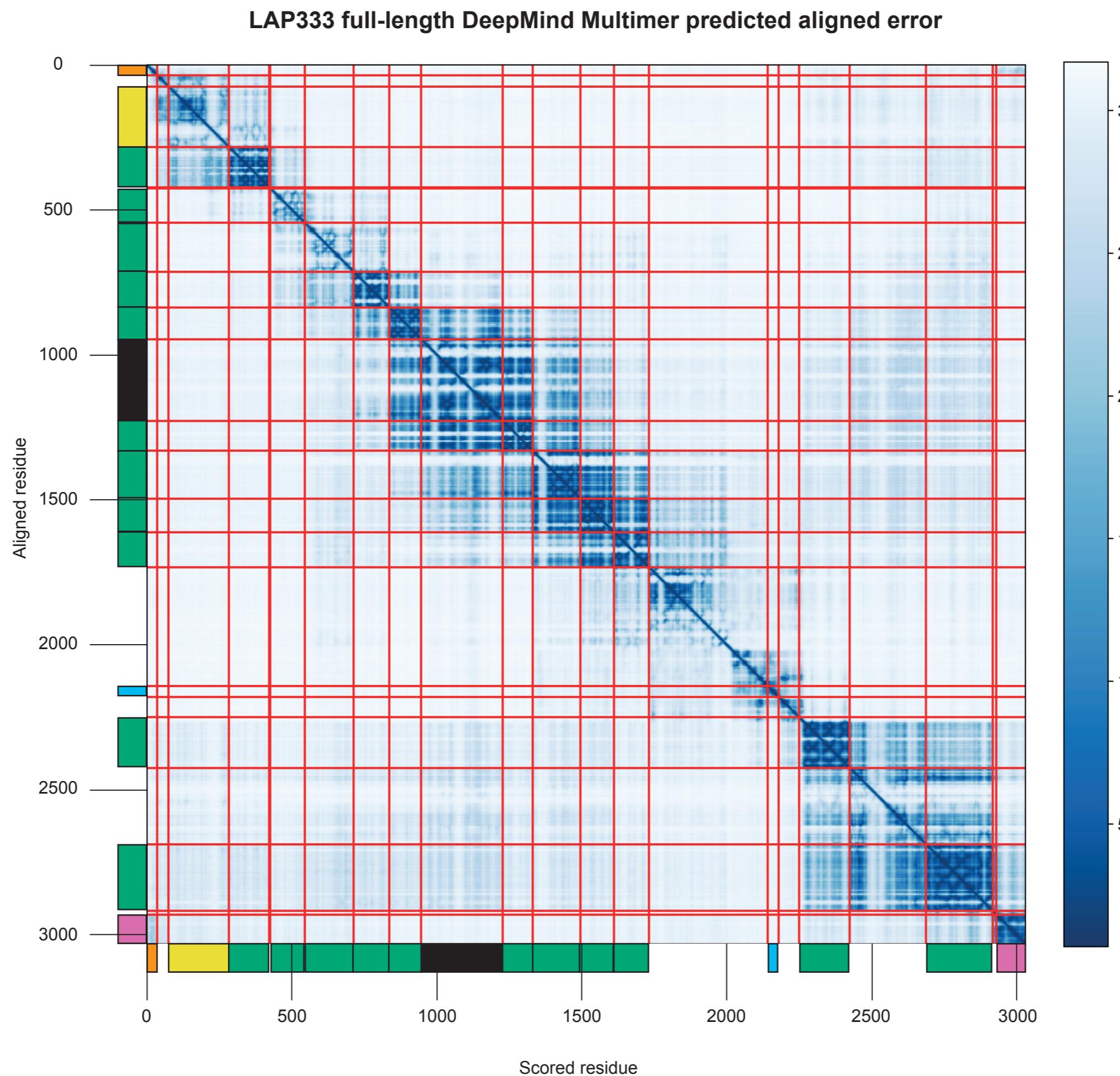


Figure S8

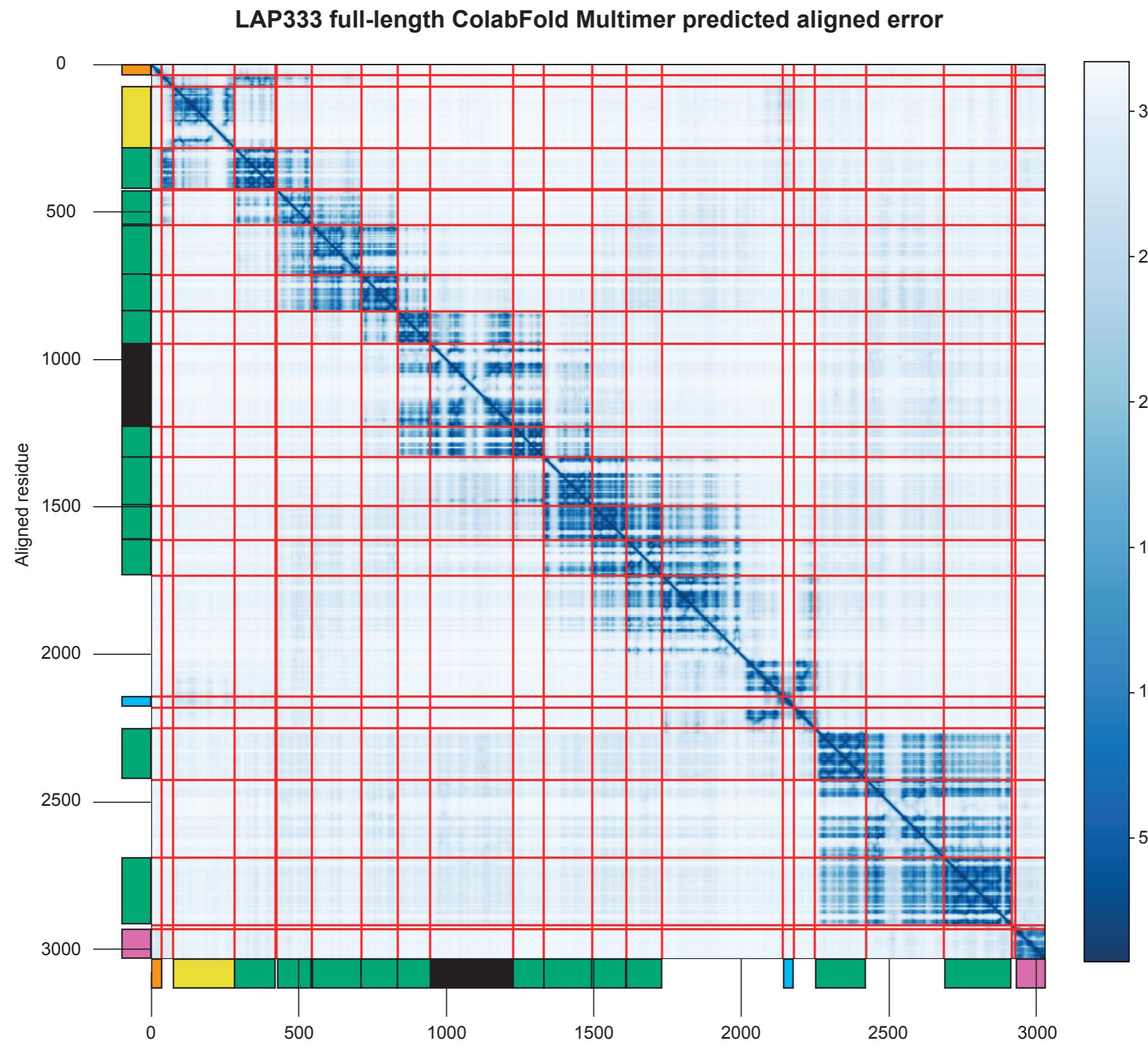


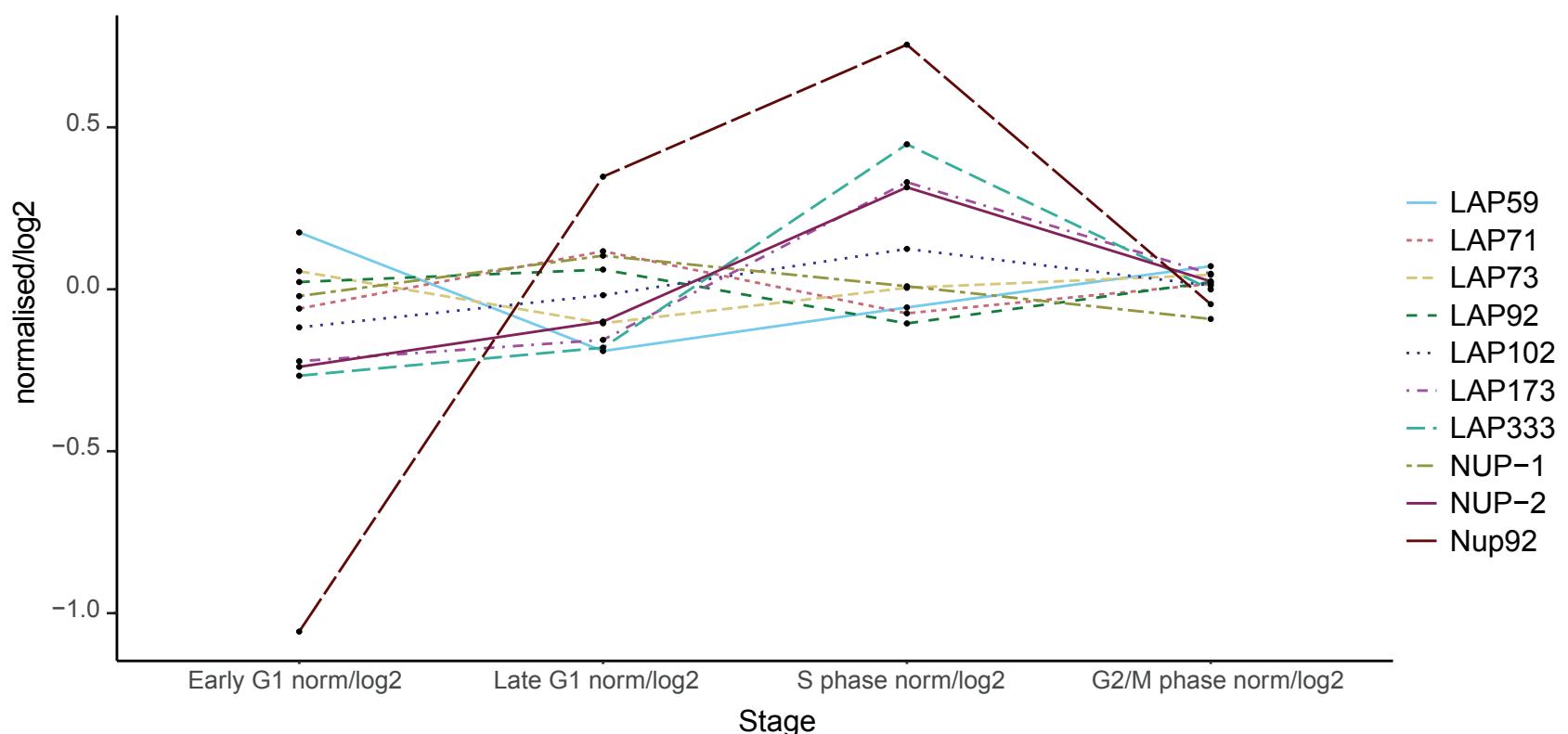
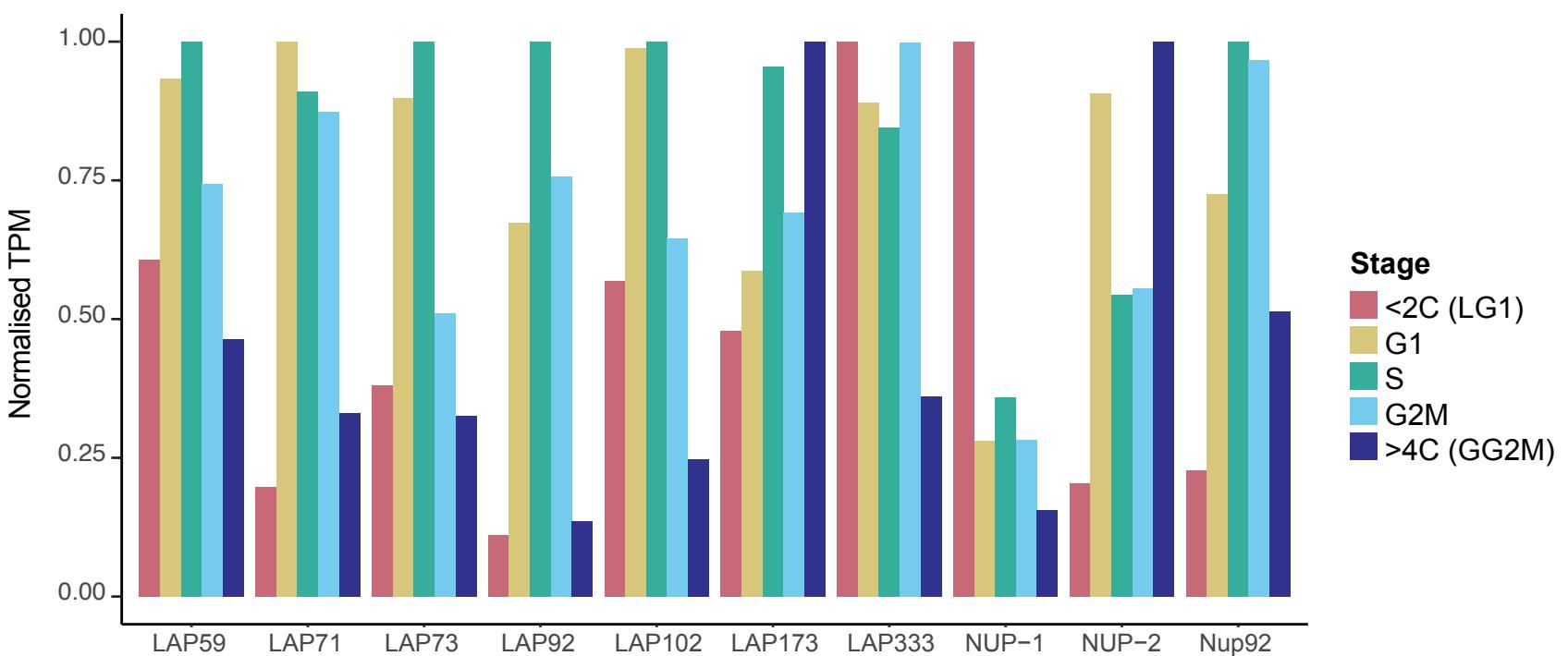
Figure S9**A****B**

Figure S10

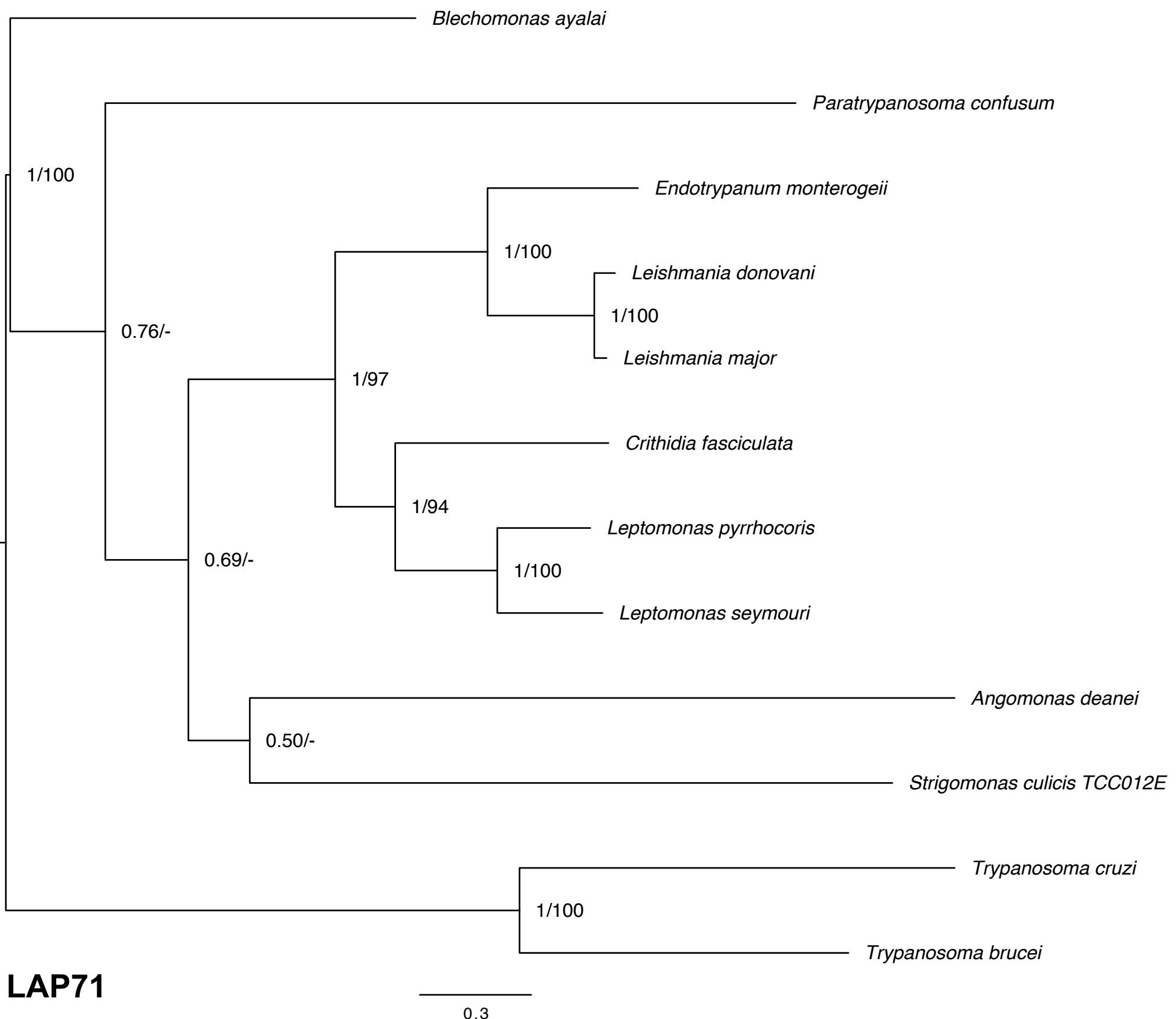


Figure S11

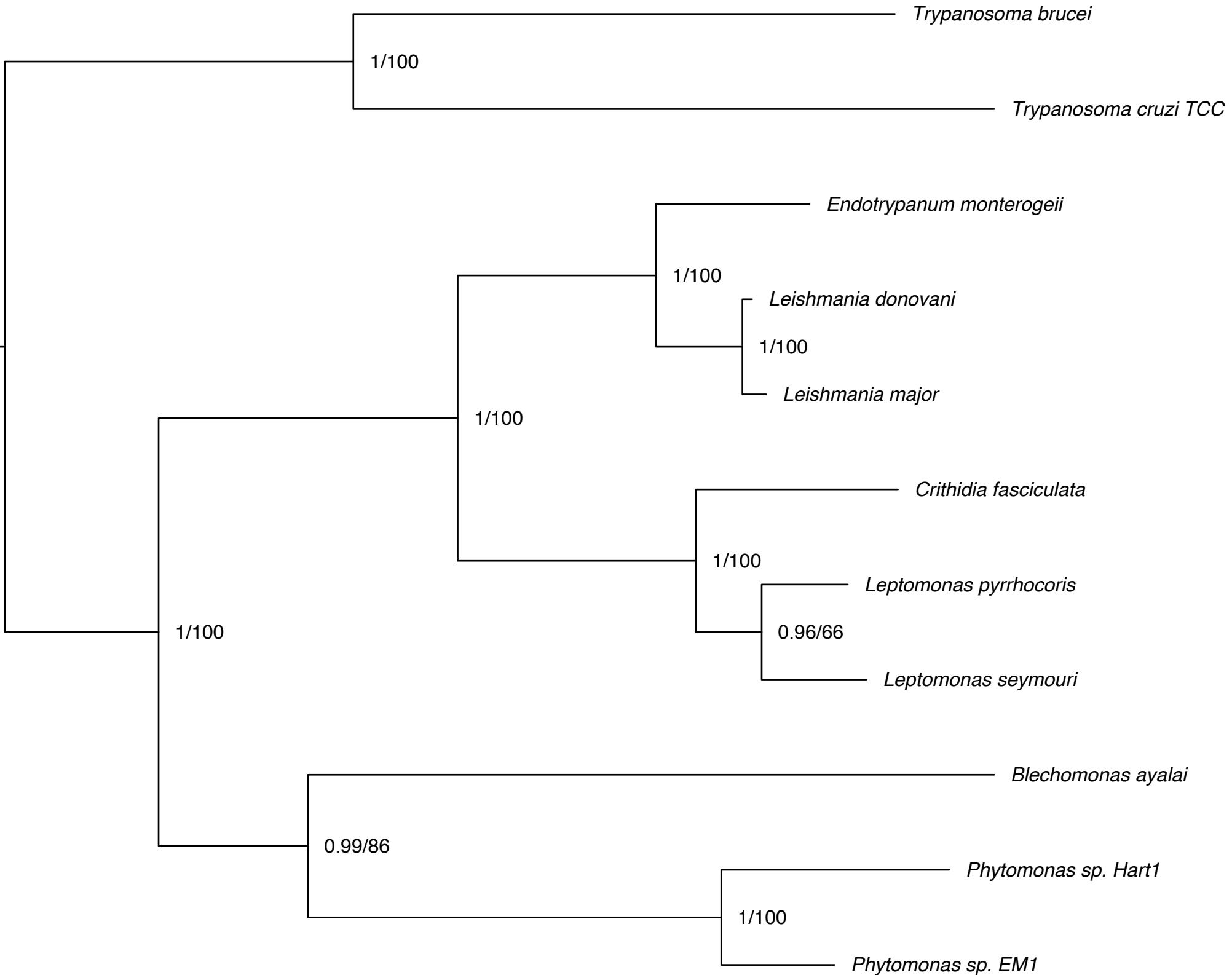


Figure S12

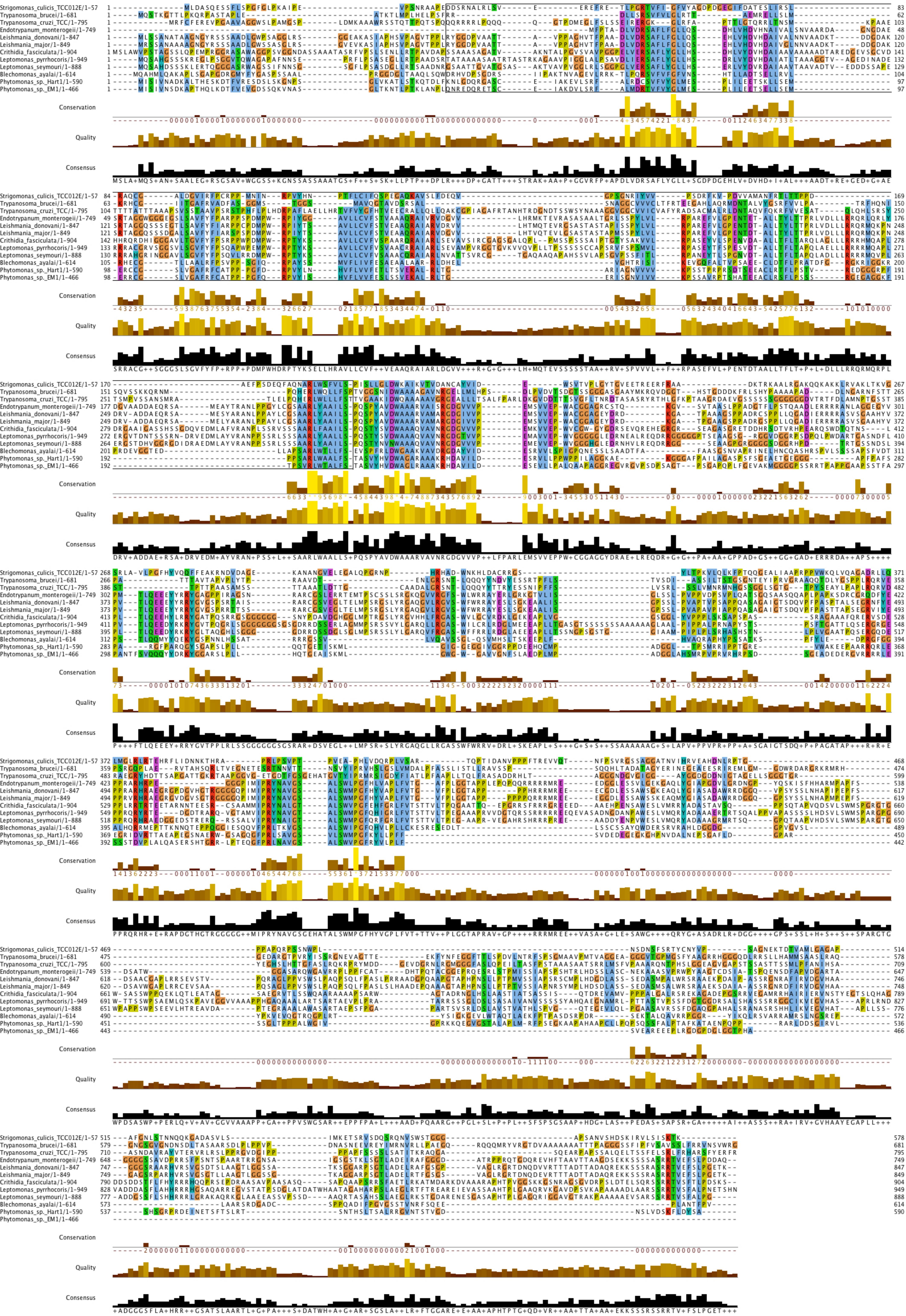


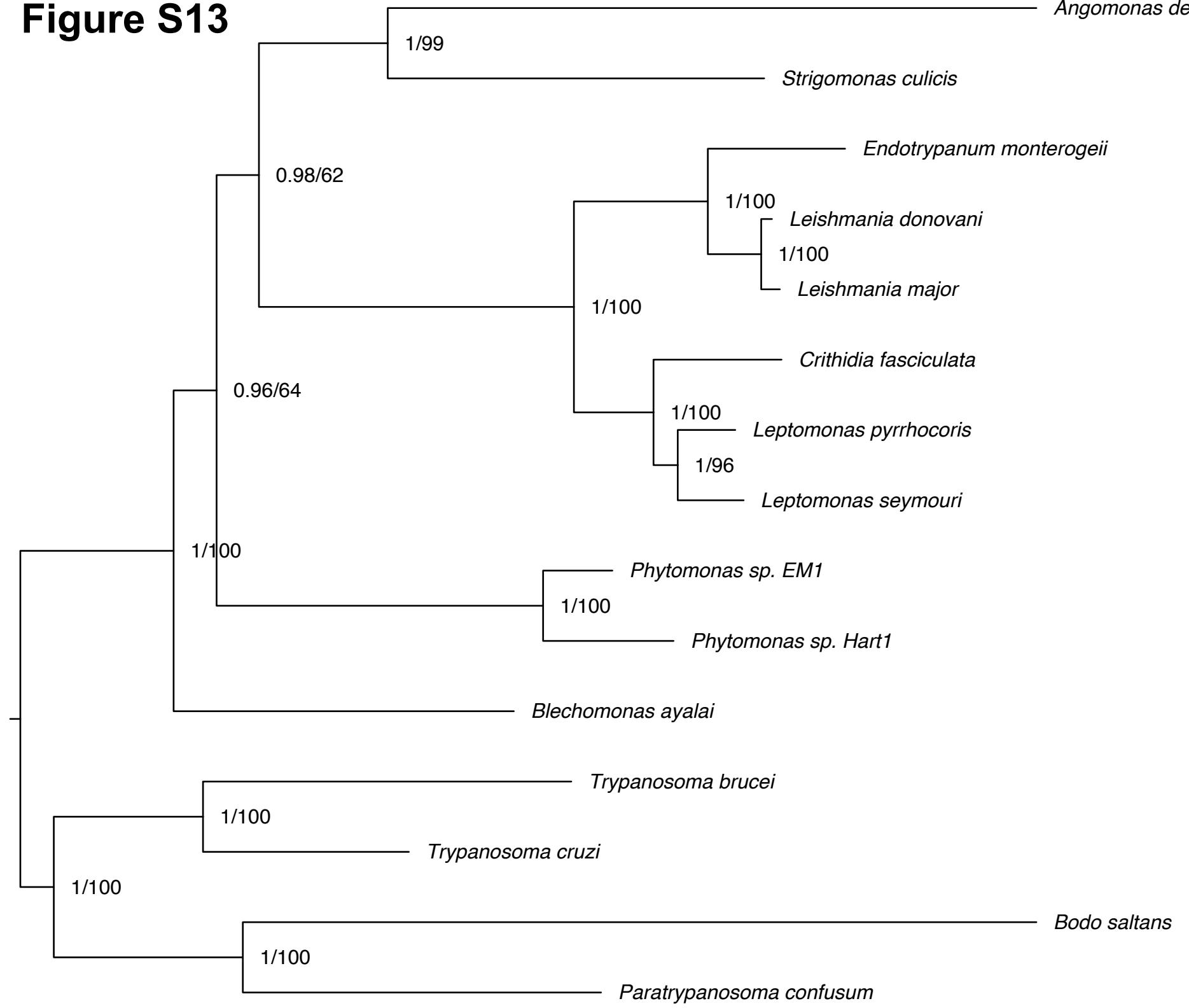
Figure S13

Figure S14

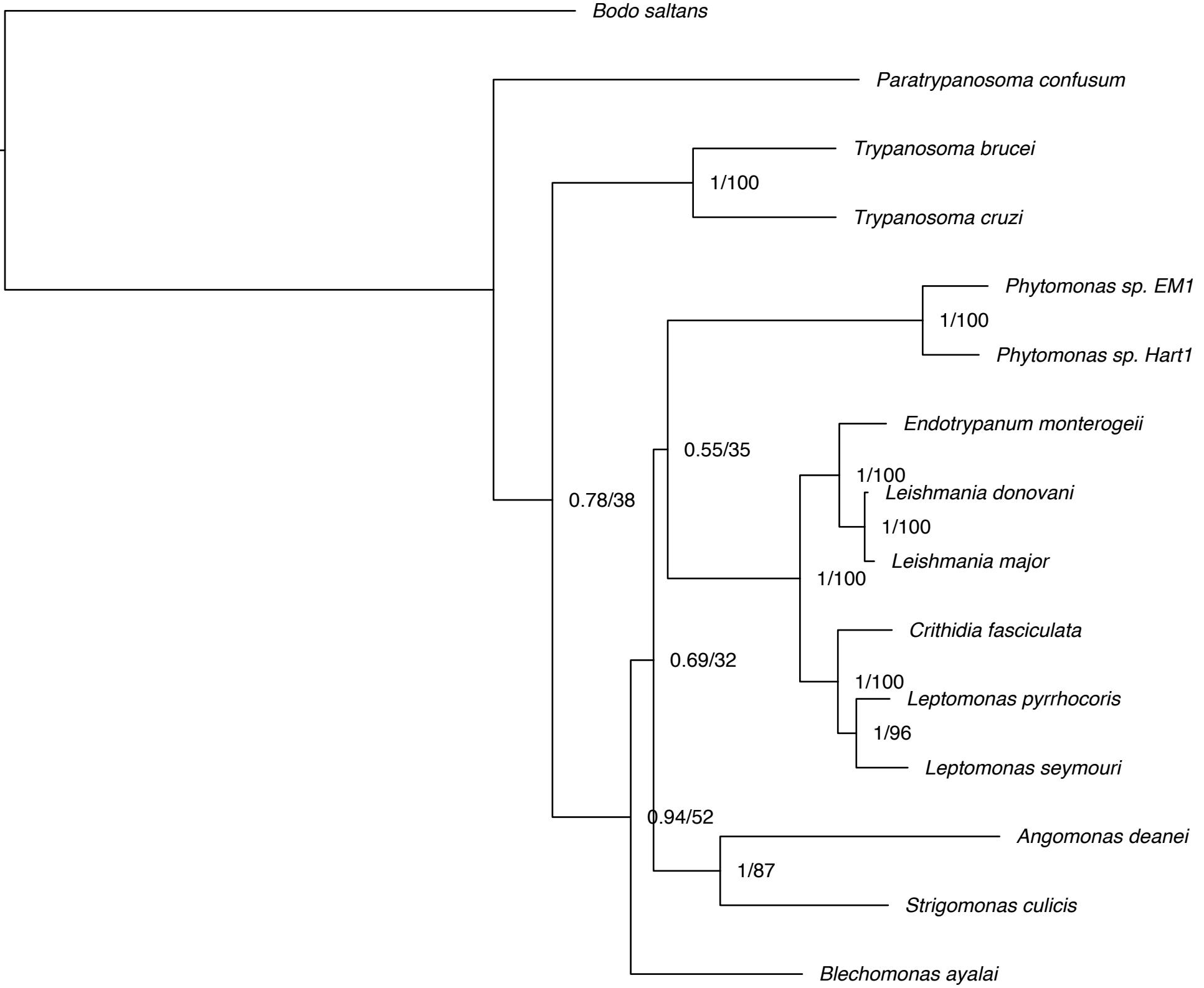


Figure S15




Figure S16

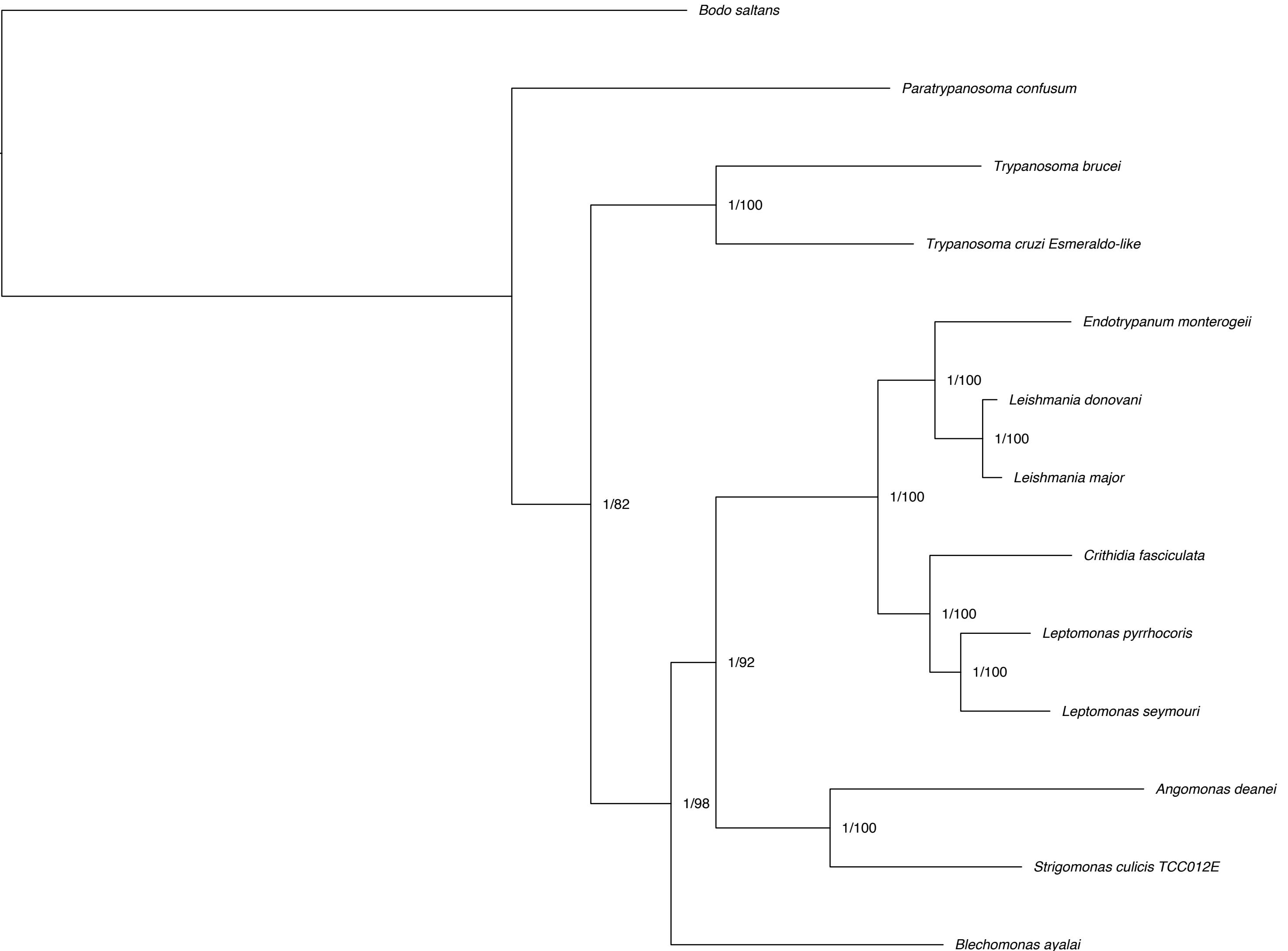


Figure S17

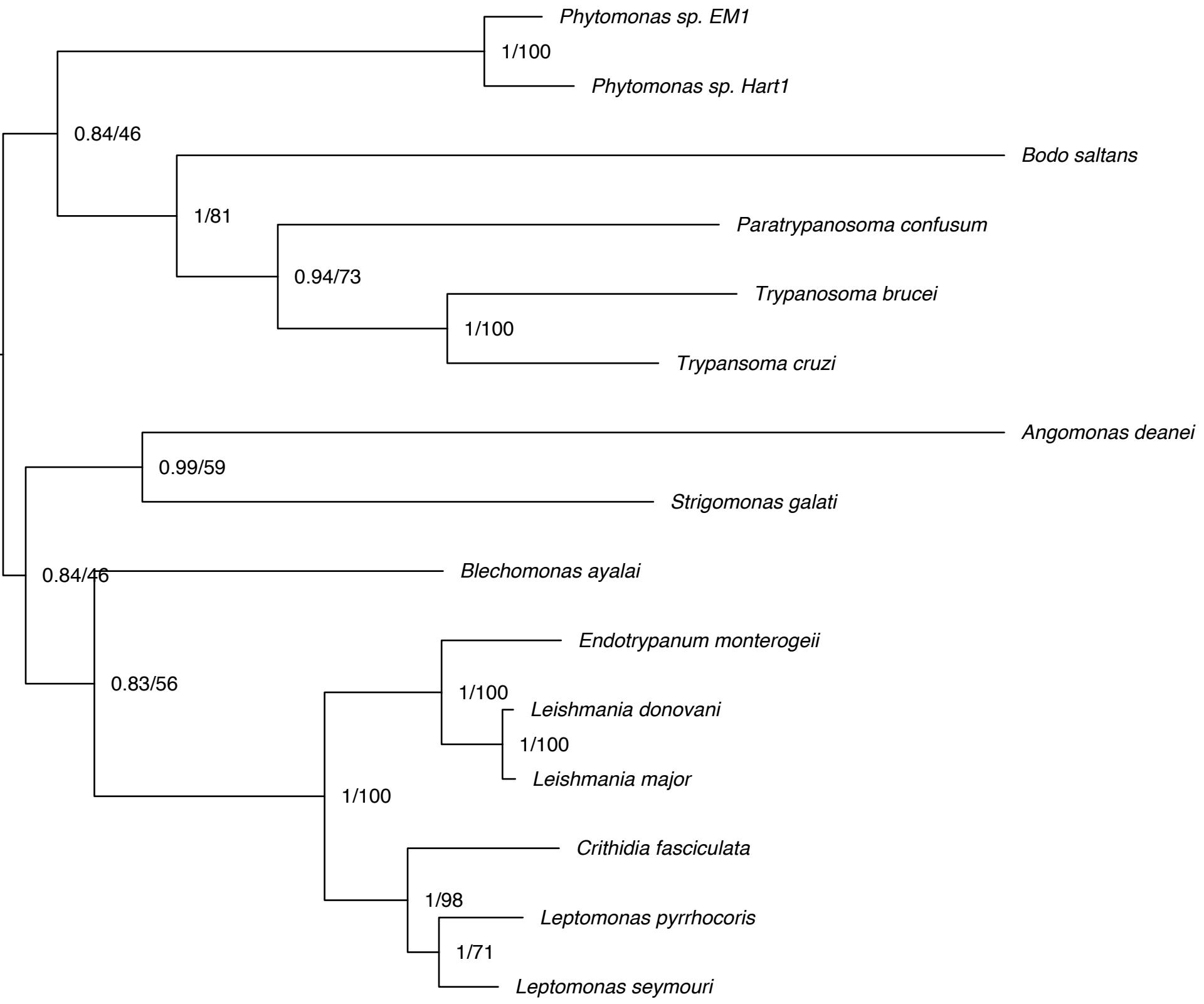


Figure S18

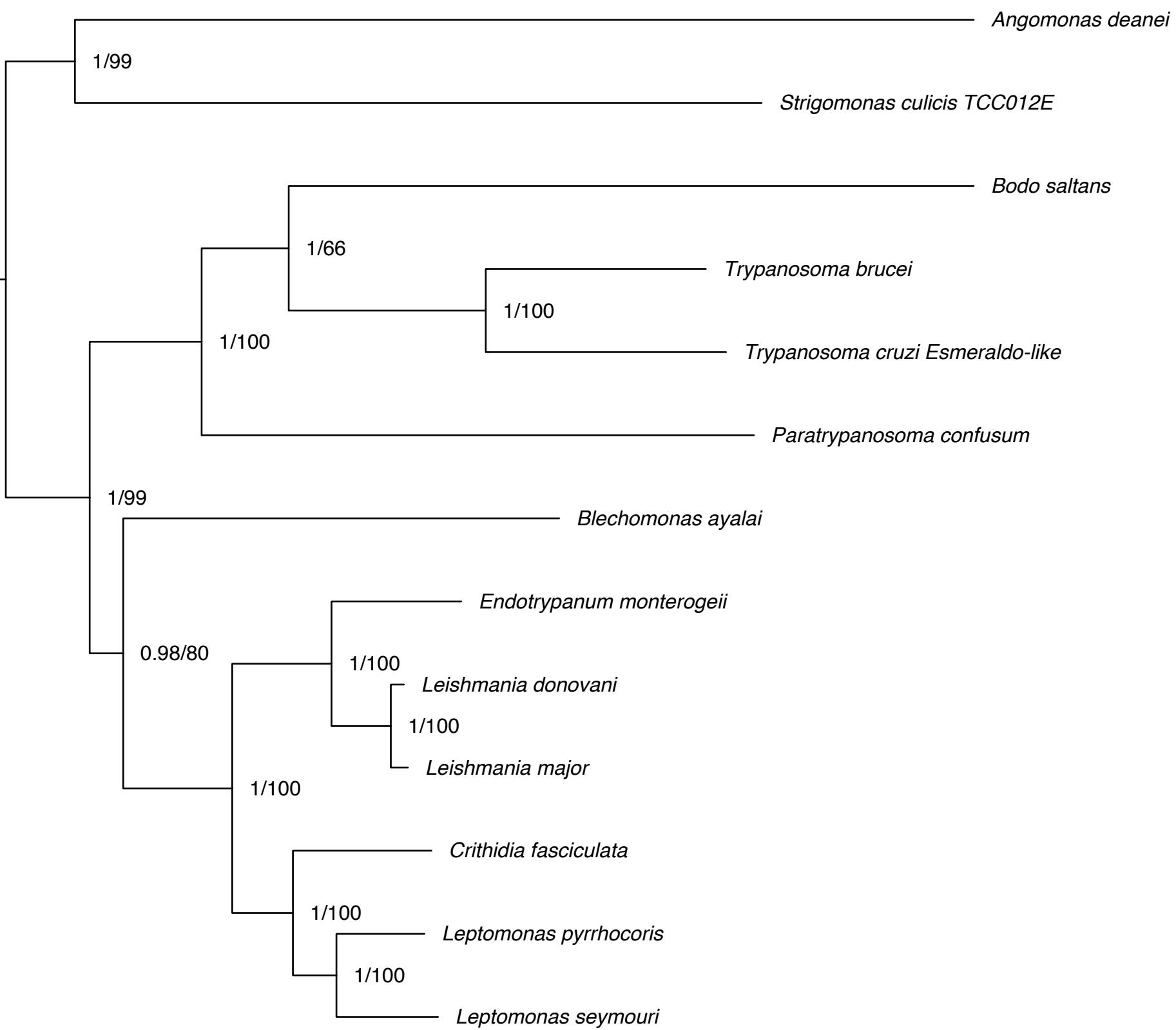


Figure S19

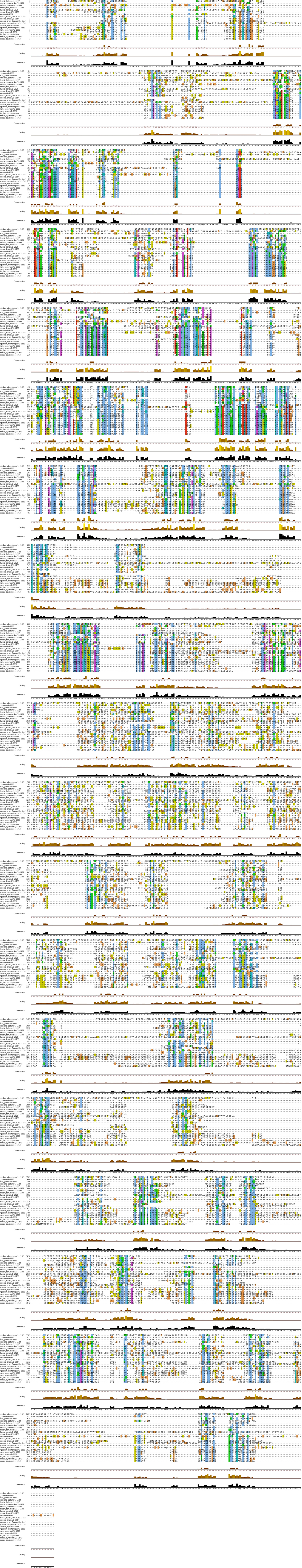
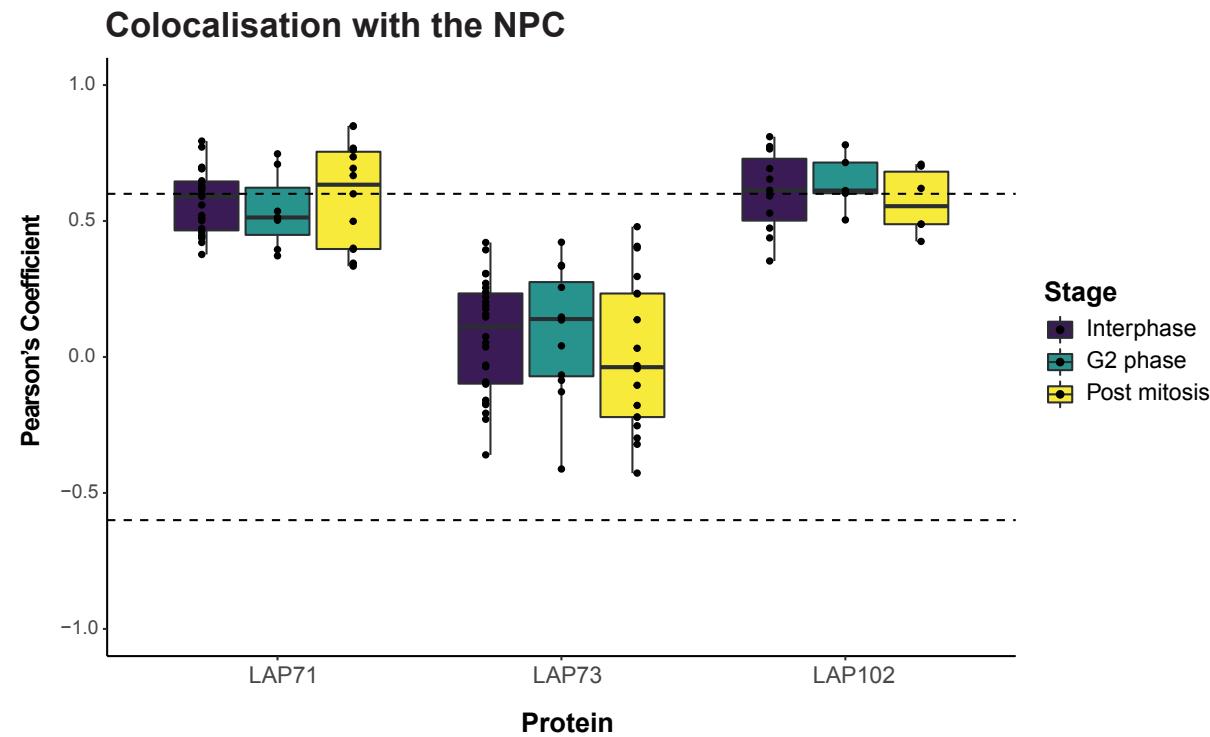


Figure S20



| Protein | Stage | Mean | Median | Standard deviation | No. of Cells |
|---------|--------------|------------|---------|--------------------|--------------|
| LAP71 | Interphase | 0.57121739 | 0.591 | 0.11651569 | 23 |
| | G2 phase | 0.53928571 | 0.513 | 0.14304395 | 7 |
| | Post mitosis | 0.58857143 | 0.6335 | 0.19656736 | 7* |
| LAP73 | Interphase | 0.07273077 | 0.111 | 0.21174939 | 26 |
| | G2 phase | 0.09366667 | 0.1395 | 0.23809446 | 12 |
| | Post mitosis | 0.00666667 | -0.0375 | 0.28204276 | 9* |
| LAP102 | Interphase | 0.60863636 | 0.613 | 0.14857609 | 11 |
| | G2 phase | 0.6426 | 0.612 | 0.10713916 | 5 |
| | Post mitosis | 0.57216667 | 0.5545 | 0.12126569 | 3* |

Figure S21

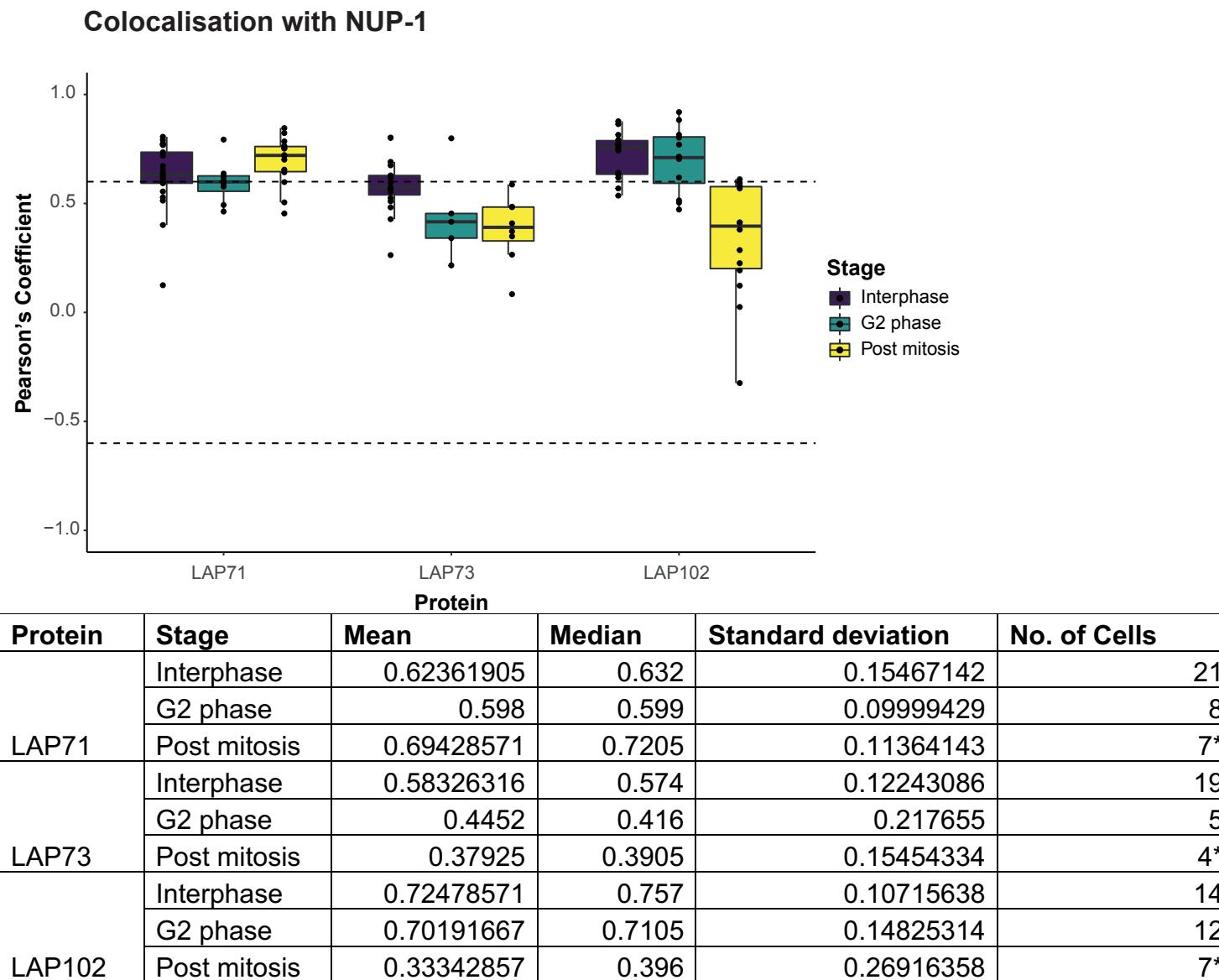
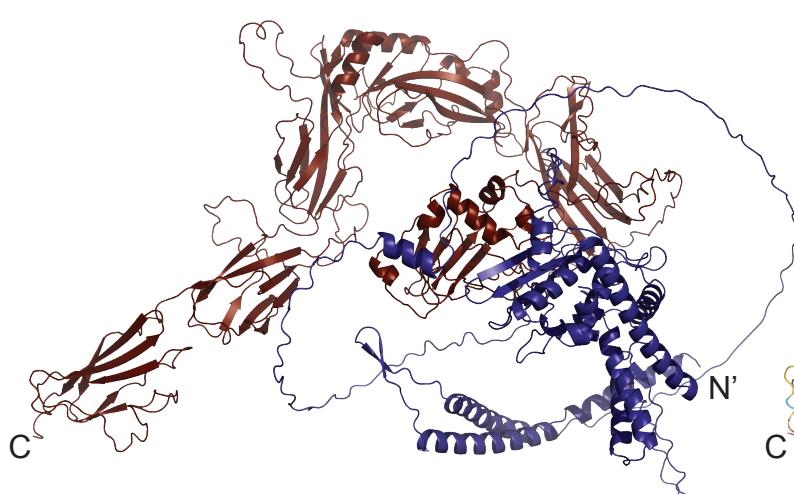
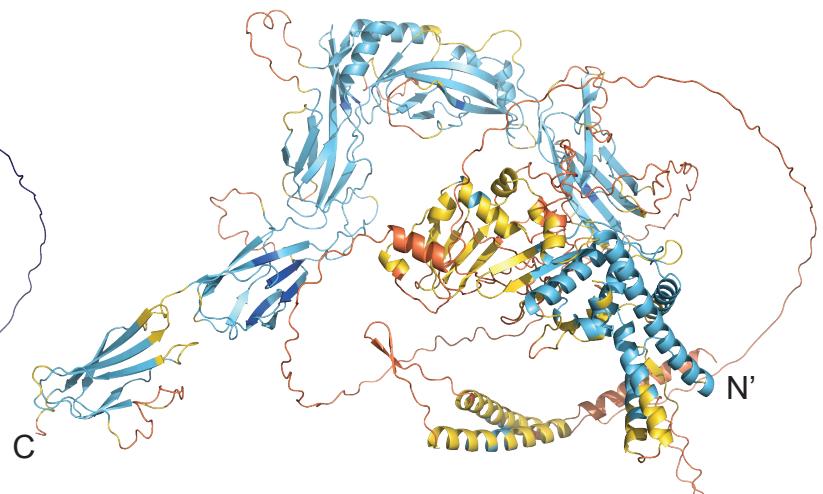


Figure S22

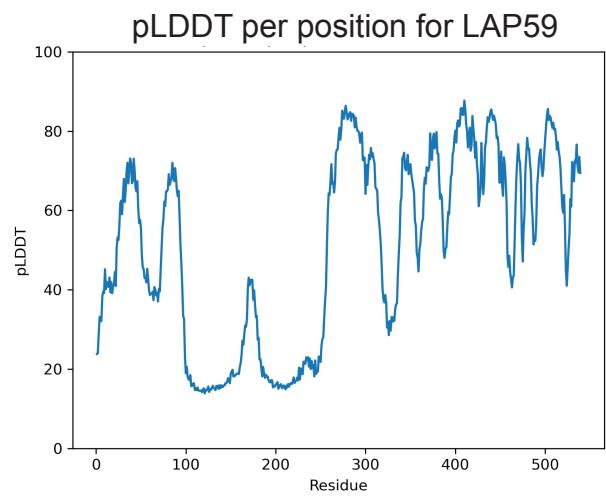
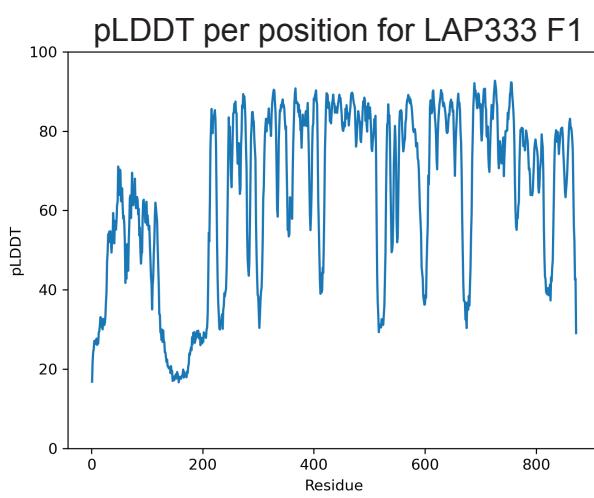
A



B



C



D

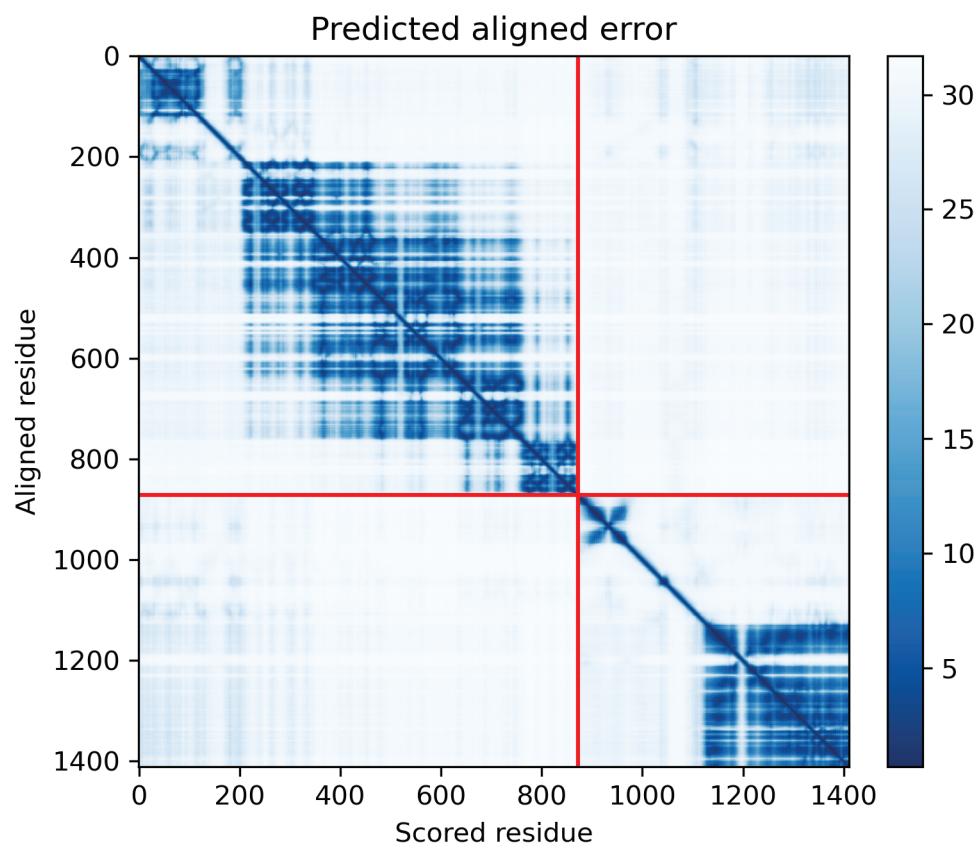


Figure S23

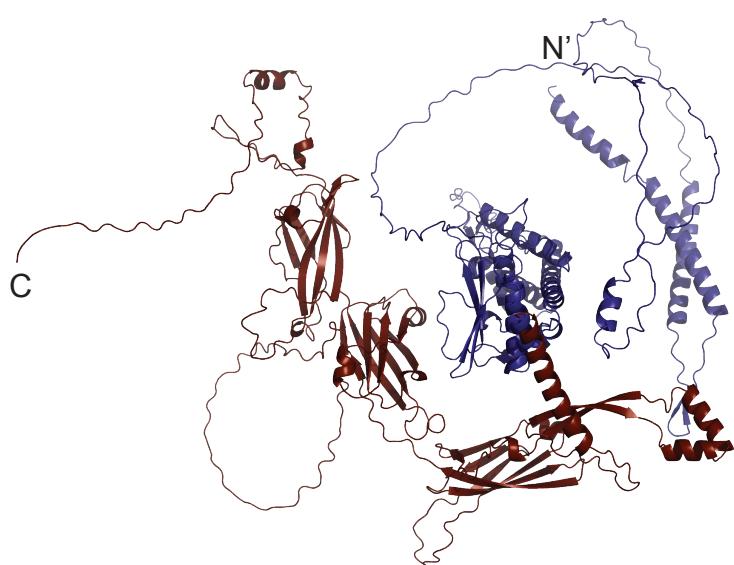
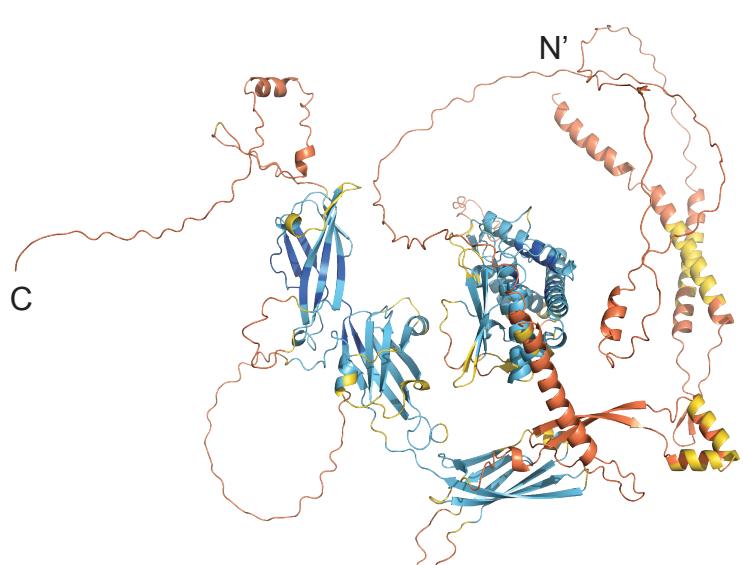
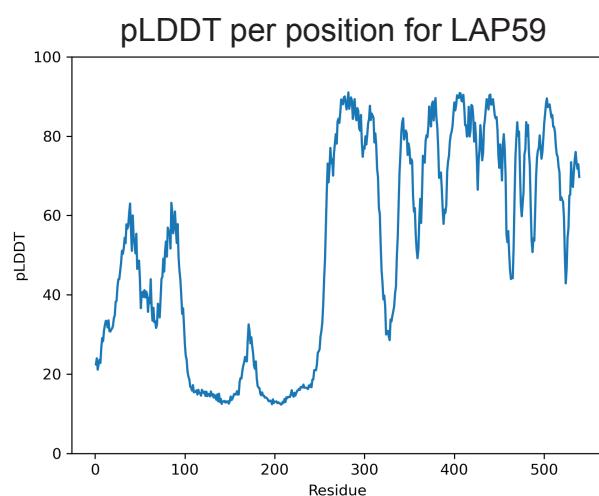
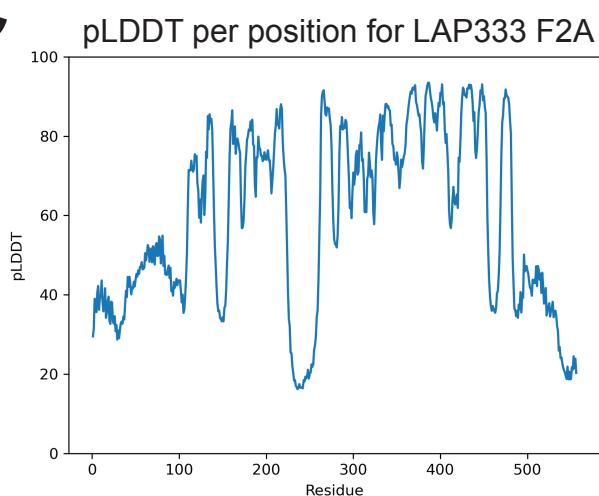
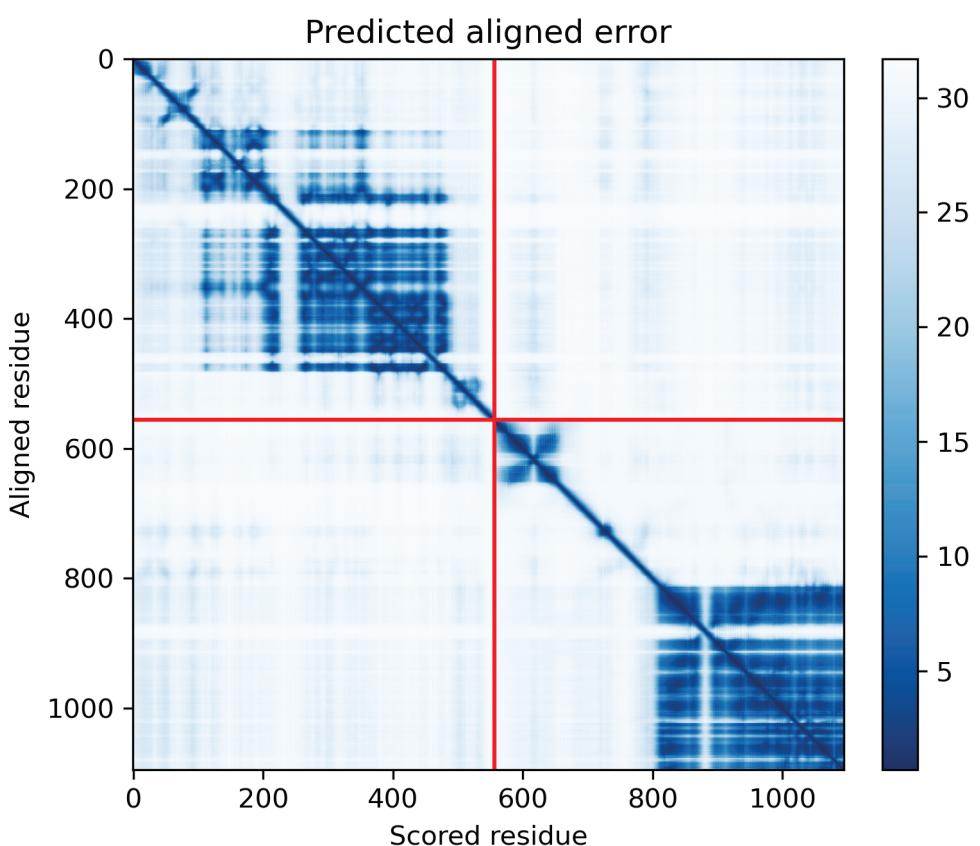
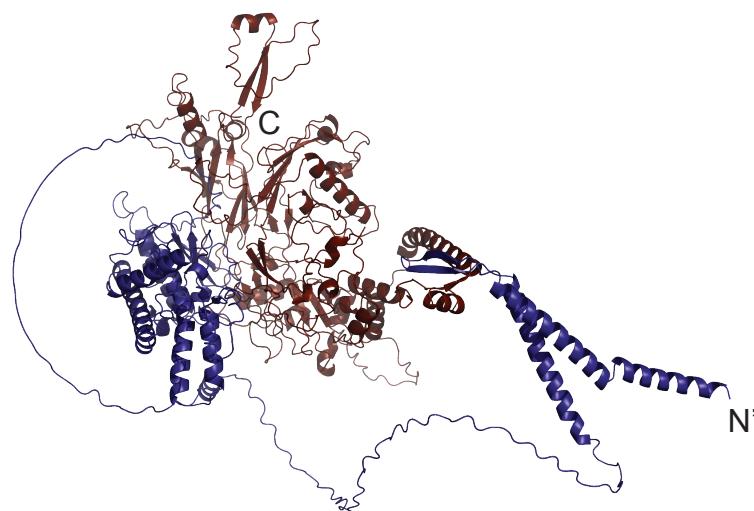
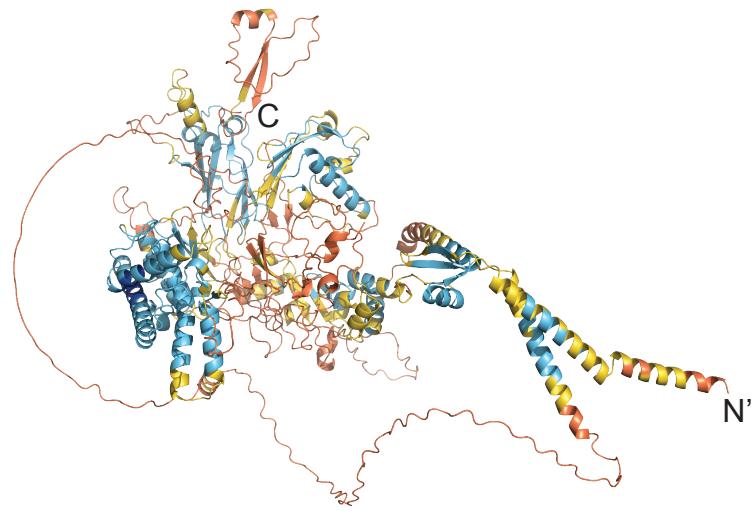
A**B****C****D**

Figure S24

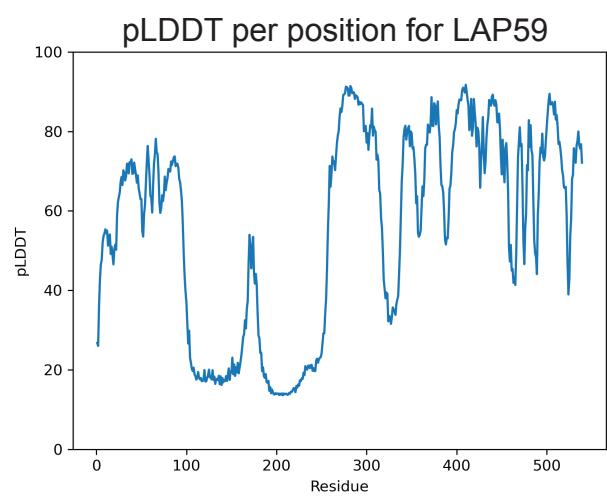
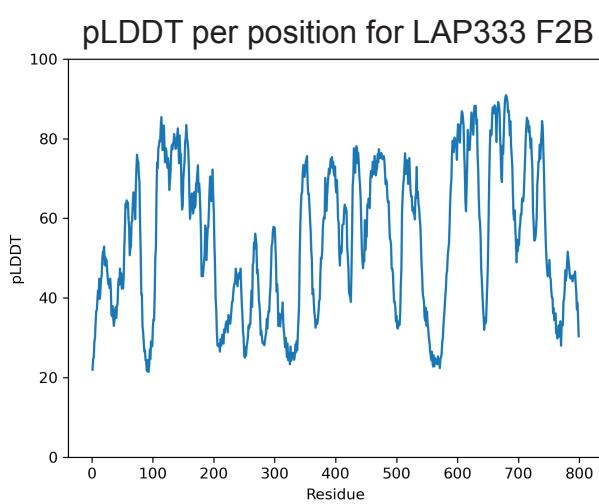
A



B



C



D

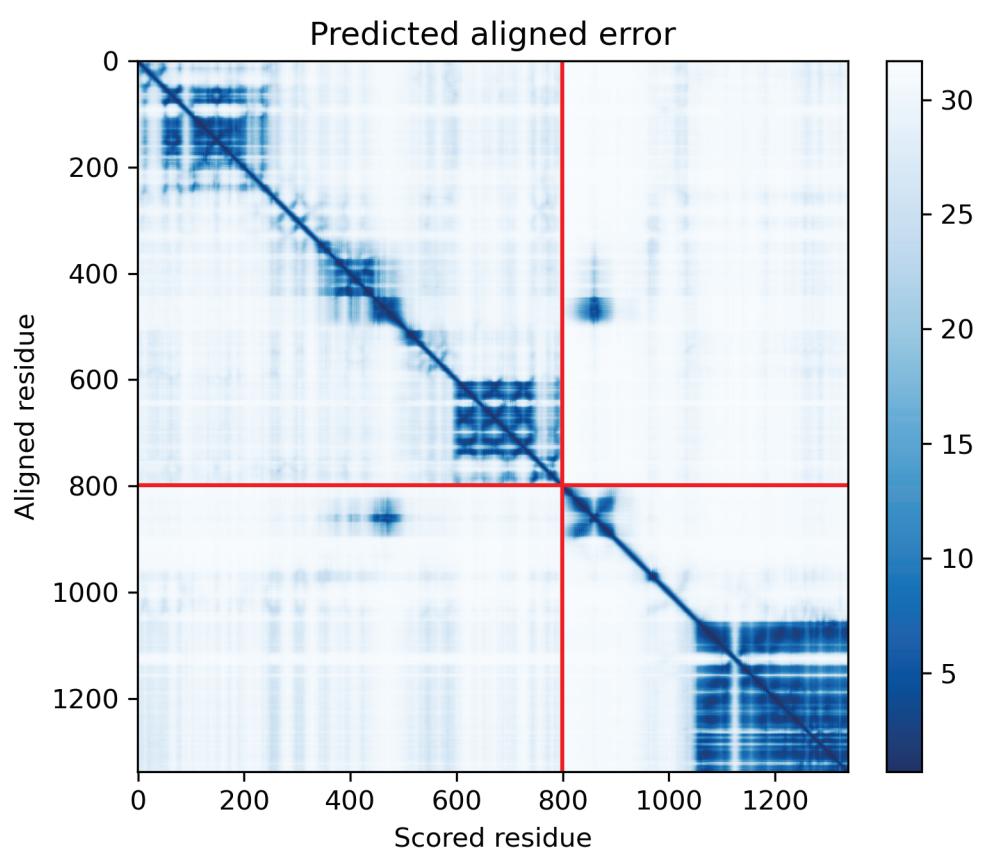
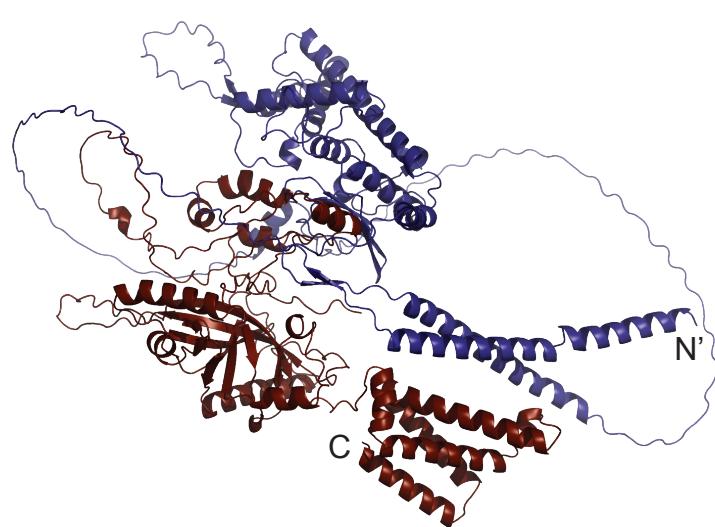
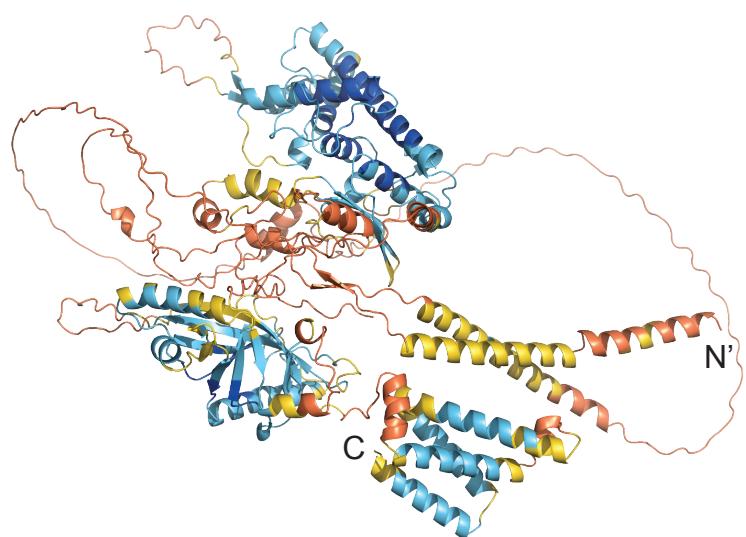


Figure S25

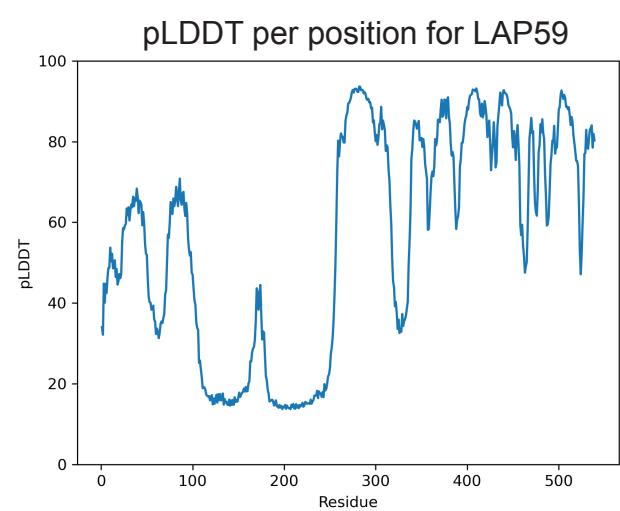
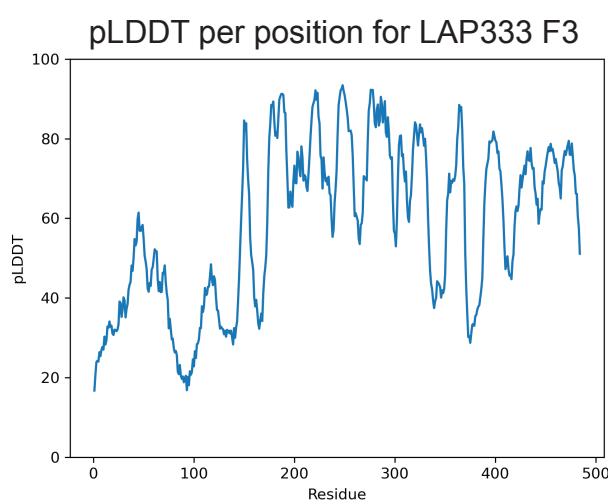
A



B



C



D

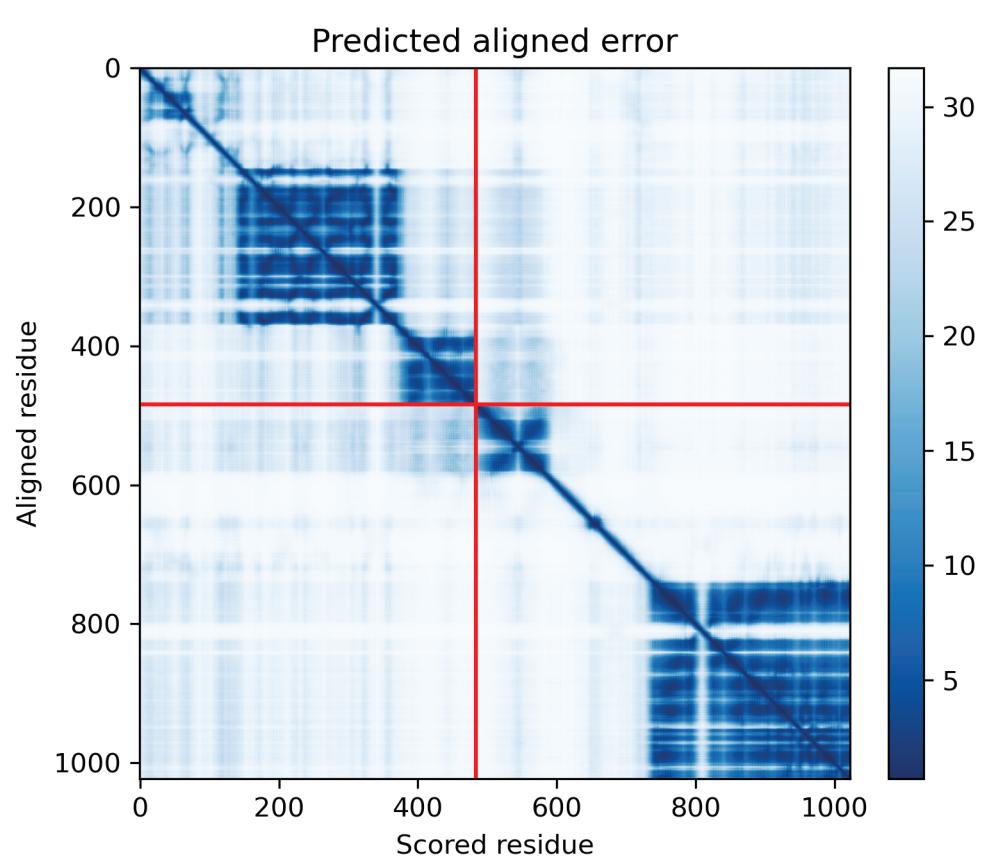
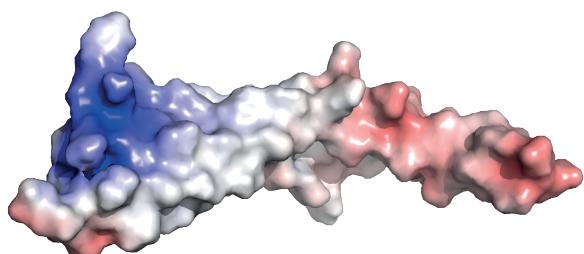


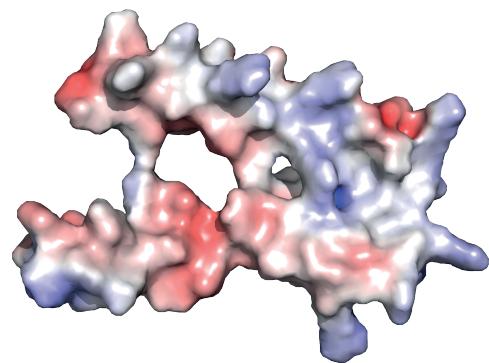
Figure S26

A

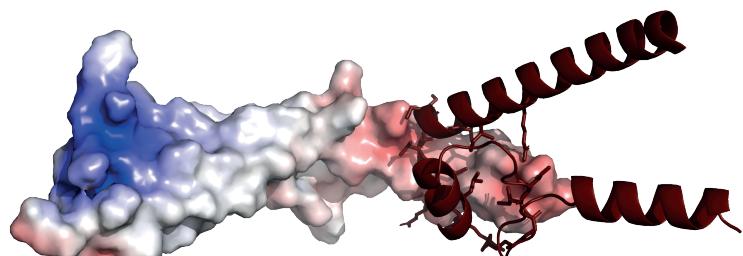
LAP59



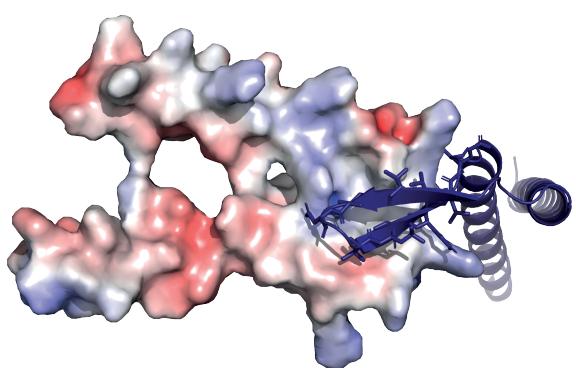
LAP333



LAP59 with LAP333

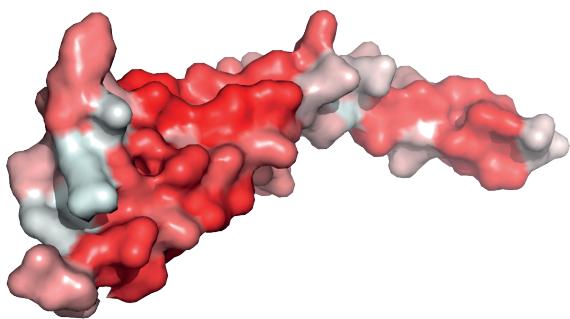


LAP333 with LAP59

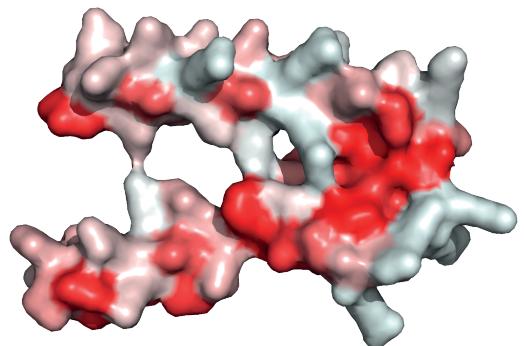


B

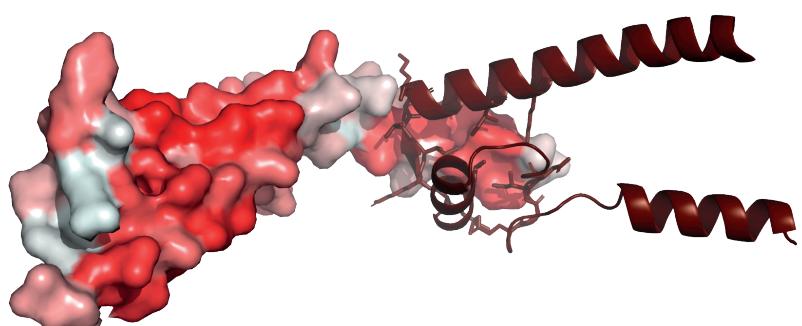
LAP59



LAP333



LAP59 with LAP333



LAP333 with LAP59

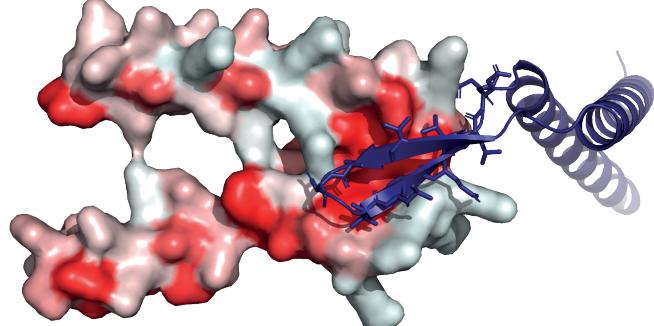


Figure S27