Supplemental Methods

A novel fragmented mitochondrial genome in the protist pathogen *Toxoplasma gondii* and related tissue coccidia

Sivaranjani Namasivayam[a,b], Rodrigo P. Baptista[b,c], Wenyuan Xiao[a,b], Erica M. Hall[a], Joseph S. Doggett[d], Karin Troell[e] and Jessica C. Kissinger[a,b,c,1]

[a] Department of Genetics, University of Georgia, Athens, GA, 30602, USA
[b] Center for Tropical and Emerging Global Diseases, University of Georgia, 30602, USA
[c] Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA
[d] Division of Infectious Diseases, Oregon Health Sciences University, OR, 97239, USA
[e] Department of Microbiology, National Veterinary Institute, SE-75189 Uppsala, Sweden

Corresponding Author: Jessica C Kissinger, University of Georgia, 500 D.W. Brooks Drive, Coverdell Center, Athens, GA, 30602, USA

jkissing@uga.edu, +1-706-542-6562
ORCID: 0000-0002-6413-1101

SUPPLEMENTAL METHODS

Multiple Sequence Alignment. Predicted cytochrome sequences (nucleotide and amino acid) we compared to other apicomplexan mitochondrial proteins using MUSCLE (Edgar 2004) (Supplemental Fig. S7). The GenBank accession numbers are: *P. falciparum* (AAC63390.2, AAC63389.2, AAC63391.1) and *E. tenella* (BAJ25753.1, BAJ25754.1, BAJ25752.1) and their associated CDSs.

DNA Electrophoresis and Southern Analysis. Restriction digestion of *T. gondii* RH genomic DNA was performed with XhoI (New England BioLabs, Ipswich, MA) at 37°C and run on a 0.8% agarose (Sigma, St. Louis, Mo) gel with a 1 kb ladder (NEB, Ipswich MA) and positive control. The gel was then soaked in 0.25M HCl for 15min to depurinate and then denatured with 0.5M NaOH, 1M NaCl for 30min and neutralized with 1.5M Tris HCL, 3MNaCl, pH 7.4, twice, 15 min each, followed by over-night transfer to

a nylon membrane and UV fixation. Pre- and post- staining with ethidium bromide was used to examine successful transfer of DNA to the membrane. Blots were pre-hybridized in 30 ml hybridization solution (50% deionized formamide, 5X Denhardt's solution, 5X SSC, 0.1% SDS, 0.1 mg/ml boiled salmon sperm DNA). A radioactive *coxI* probe (367 bp sequence block C) was boiled for 5min and immediately chilled on ice prior to hybridization at 43°C or 58°C. The membranes were washed several times, each time for 15 min with increasing stringency and finally in 0.1X SSC, 0.1% SDS at 65°C.

Enriched Mitochondrial Fractions. *T. gondii* tachyzoites (~ 2-4 x 10^9) were purified as in (Miranda et al. 2010). The purified suspension was centrifuged and the pellet resuspended in lysis buffer followed by centrifugation at 700 g. The wet pellet was mixed with silicon carbide and this mix was transferred to a mortar and manually ground for 3-4 15 sec periods. This mix was resuspended in ~20 ml of lysis buffer, decanted and clarified by three centrifugations at 50 X g (5 min) and 150 X g (twice) (10 min each). The supernatant was centrifuged at 1000 X g or 1200 X g for 10 min to collect the mitochondrial-enriched pellet.

Nucleic acid Extraction and Sequencing. DNA for Illumina (San Diego, CA, USA) and/or Oxford Nanopore (Oxford, United Kingdom) sequencing was extracted from *T. gondii* ME49 and *N. caninum* Nc-1 parasites using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's cultured cells protocol. The DNA was eluted from the spin columns using 50 µl Buffer AE. DNA was concentrated to a final concentration of 20 µg/ml using a speedvac and measured on a Qubit. PE-150 Illumina sequencing with a peak insert size range of 450-500 bp was performed at the Georgia Genomics and Bioinformatics Core on a MiSeq. MinION sequencing was performed using the Rapid Sequencing Kit, SQK-RAD004, and 400 ng of genomic DNA on a Nanopore flowcell 9.4.1 according to manufacturer instructions. MinKNOW version 18.03.1 was used to call bases.

PCR

The High-fidelity Platinum Taq polymerase (Invitrogen, Waltham MA, USA) was used for PCR with: initial denaturation at 95˚C for 3 minutes; followed by 35 cycles at 95˚C for 60 s; 54˚C for 45 s; and 72˚C for 60 s. The PCR products were visualized with agarose gel electrophoresis and if not pure, bands were gel extracted prior to sequencing using a PCR purification kit (Qiagen, Hilden, Germany). If the yield was low, extracted DNA was used for a second PCR amplification.

Analysis of Oxford Nanopore Reads. Two strategies were used to identify mtDNA reads. First, the Nanopore reads were screened for the 21 mtDNA SBs using BLASTN (E-value $1e^{-10}$, nucleotide identity ≥ 60% and ≤ 10% alignment length mismatch). Second, the reads were aligned to the 21 mtDNA SBs using Exonerate v2.4.0 (Slater and Birney 2005). The best 2000 alignments with an alignment ≥ 60% were classified as putative mtDNA. To remove NUMT-containing reads, the putative mtDNA reads were aligned to *T. gondii* ME49 nuclear sequences masked for NUMTs using BWA v0.7.17 (Li and Durbin 2009). Non-aligning sequences were considered putative mitochondrial reads. Finally, error-correction was performed using proovread (Hackl et al. 2014) with *T. gondii* ME49 Illumina reads that mapped to the 21 mtDNA SBs (see below). These error-corrected reads (Supplemental Data S6) were annotated with mtDNA SBs using BLASTN as described. All unannotated sequence bits in each Nanopore read, which typically consisted of approximately 50-100 bp at the beginning of a read were examined and compared to each other for any repeating patterns to identify any additional mtDNA sequence.

The main difference between *T. gondii* and *N. caninum* SB's is due to the acquired presence of block Sp in *N. caninum* between blocks D and V altered version of SBs (NA) and (F/FpR) junctions Closer examination of the SBs Sp and D in *N. caninum* revealed an 8 bp overlap between the 3′ end of Sp and the 5′ end of D suggesting microhomology as a possible mechanism for the creation of this arrangement.

Dot Plot data set details. The 21 full-length SBs (excluding partials) were compared to each other in Fig. 4A-B and Supplemental Fig S11 A-B. The sequences were sorted by size from largest to smallest, yielding the following SB order on each axis [C M E H K U T S I A P F N V L O R J D Q B]. The dot plots of the ONT reads utilized the 25 longest *N. caninum* ONT mtDNA reads and 48 longest *T. gondii* ME49 ONT mtDNA reads in the exact order represented below. Fig. 4C-E represent *T. gondii* ME49 ONT mtDNA reads plotted against themselves with different window sizes. Fig. 4F is a plot of 25 *N. caninum* ONT mtDNA reads on the vertical axis vs the same 48 *T. gondii* ME49 ONT mtDNA reads used in the other ONT panels. Data on the libraries that yielded these reads are located in Supplemental Table S3.

Name	Nanopore read name	Length	Name	Nanopore read name	Length
NcNano_16	a0171521-cbdd-4035-82cd-345f072b52f7	15781	TgNano_27	80e16e15-e446-4e7a-8f3d-6bf9804672ae	23356
NcNano_17	a9a80eee-fcb1-4987-bb5a-e8ed53a6fcaa	8658	TgNano_3	09d6f4a1-4c78-4aa5-aa9d-bcbfc0bfeddf	15258
NcNano_8	57b046b1-da7d-4790-935b-09aa7ce1c4ce	8643	TgNano_17	551bcaa6-78a6-4b11-8cc1-d28b1072aea4	14268
NcNano_1	01b89a37-be91-4e28-9781-52a0a10b6c75	7846	TgNano_13	4a67917f-40b9-4069-9a10-a3d0b14177fd	14034
NcNano_20	c07fb06c-9bdd-4bb4-9490-85e66184f693	7620	TgNano_4	188f54cf-e93c-47f4-beb9-79cf2224e25c	11658
NcNano_4	2b86c0e1-cde5-4131-a213-d3c308f0de49	6878	TgNano_23	6b30cf05-c49e-409d-8415-724b469c9fef	11059
NcNano_15	95e710eb-ec54-46a8-9e6b-c41cefb9243b	6845	TgNano_11	445ac7c0-953a-4003-aa0c-9ff0a24ec9bb	10009
NcNano_21	cce5a084-529d-4fc9-8211-5177e41487c0	6752	TgNano_15	50a5090a-60db-4dda-8740-9ff7097a9e83	8691
NcNano_12	829a9244-4c81-41f7-a799-7c54eb5794b5	6480	TgNano_35	cb691aab-c97f-4d85-862c-80a9a7c1c05f	8683
NcNano_11	73e6d7fe-dd77-4cc4-af78-cbb17e21d594	6470	TgNano_28	9150e675-890e-40ab-b24d-db9db35a7a20	8402
NcNano_24	f974937c-c166-4dab-ada3-240bd5588e37	6305	TgNano_41	e5e52891-9720-47ec-9eb4-451adfe536c4	8338
NcNano_7	34fce961-3df0-47fd-8d16-7d12c7fef7c0	6016	TgNano_26	7876e473-7148-414d-8ff1-e6fb270b1582	7932
NcNano_22	de36d323-28ab-48d0-980d-229aaffe847d	5776	TgNano_7	22690990-a058-4283-b7c7-31fcd318a8bb	7612
NcNano_6	317cd209-e09b-41b2-a76a-96e5afb0fee1	5216	TgNano_42	e613ab53-3b6b-4c22-8d5a-057801b31fd1	7504
NcNano_14	88daffdf-5eaf-47bf-98f8-c3cca524352f	4705	TgNano_30	9fade8e8-9bc2-41bb-8af4-0449ee8d2fc0	7463
NcNano_10	71e21190-3867-410d-9a34-e8ff3cf41625	4449	TgNano_32	b077641f-e402-45f5-851e-acb894ac1e70	7352
NcNano_5	2cca988f-4aaf-4911-991f-d20a5ba97bde	4278	TgNano_24	6fe0cf4c-f443-44c7-9ea9-908ea3a4d51a	6674
NcNano_25	f9f9c531-a1f8-4163-88db-6f95141c96ed	4196	TgNano_33	bfdeae82-dff5-43b4-b161-534f47e8ab97	6344
NcNano_23	f2230b9b-c622-4082-ab40-472a1a68a1e1	3319	TgNano_44	e98b32e1-8494-454d-afc0-d178476e50c1	6224
NcNano_3	0b9e1642-08f8-4e4c-838e-25929c5ad6b6	2961	TgNano_47	fc2f9aaf-fef5-4d97-aef4-dd91dc3e343f	6217
NcNano_2	01fba50a-71a0-4da0-bd8a-360f82565958	2959	TgNano_43	e65ceec9-469e-4c98-b797-923342f3f5ab	6051
NcNano_18	b163171e-727d-4633-be8e-d68e5e118ada	2622	TgNano_40	e32f2dad-4608-49ff-b62e-379af84de712	5828

NcNano_9	6c5c7174-5cc4-4cf3-a49a-f8add8374521	2479	TgNano_45	f3772ab8-ad3c-4a45-90f4-41f8baf081ca	5775
NcNano_13	84abb7a8-e0b2-4e83-b2f2-33e1664fa20b	1956	TgNano_25	734acb5d-d7a7-4299-a6db-1032095a98d6	5635
NcNano_19	b189d5ef-6a1a-44c5-b81d-539f27986de1	1951	TgNano_20	57b25780-5da3-489c-a075-0a1cbe13b8af	5629
			TgNano_21	6248daa1-ea90-4b18-ae2c-682f8e67b080	5525
			TgNano_8	2334281a-0fd8-4c87-9478-cdbdc8366b3c	5401
			TgNano_29	968eeb3d-bb74-4ff7-8c45-553b2e6c9529	5206
			TgNano_16	53ee132e-5707-4d09-97a3-a7171b0a1f81	5186
			TgNano_34	c61269fa-fa08-4679-8983-f1a01a1635e4	4821
			TgNano_36	cd0c4600-e5f6-4666-803b-dda5975e0db2	4748
			TgNano_9	2b941d62-dffb-41f4-bd3a-74d872421b0f	4362
			TgNano_19	574ee7ee-519b-4d5d-a7b3-e1a83f9781c4	4302
			TgNano_14	4f9137f3-d610-44a1-ad50-da462668413d	3844
			TgNano_6	215a3b8a-85c4-4ee8-9c1d-b2578589eb2c	3827
			TgNano_5	1f2d5cfc-32c1-4a5c-8bfa-d9e8a7abbc7c	3805
			TgNano_31	ad9d243a-deef-4fa1-bef3-3734df6b0b3a	3739
			TgNano_37	cebf1af9-05be-4956-a6c1-6d9b6f5abb1e	3659
			TgNano_39	e0f8b1c5-1311-4f1b-9cc0-6f2d74d3eaeb	3527
			TgNano_48	0d0d37ef-be92-43f4-83f6-f7099585641c	3451
			TgNano_10	3e490cf6-dd1c-44d2-b39c-d2c7eed3a18e	3126
			TgNano_22	64830f72-ed3f-48d6-b856-f25b6f751c83	3116
			TgNano_46	fadfacb2-e251-4149-bc9f-c297168a76b2	2888
			TgNano_12	45d118dc-a990-4d53-8b47-612d9c0c7a89	2880
			TgNano_18	5570a5f7-480a-4c6d-a3e4-74e8b0e2893c	2715
			TgNano_1	0102d442-3d3e-4b9b-aa5a-279a3a2d3d4b	2651
			TgNano_2	08029a7a-3b57-4343-9250-b2e4475f00a2	2238
			TgNano_38	de78e2ed-3a35-4183-92f6-b8aaedaf6e9e	2089

Circularization of Nanopore mtDNA reads. To check if any of the Nanopore mtDNA reads circularize, Circlator v1.5.3 with minimus2 was used (Hunt et al. 2015).

SUPPLEMENTAL REFERENCES

Edgar R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**: 1792-1797.

Hackl T, Hedrich R, Schultz J, Forster F. 2014. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**: 3004-3011.

Hikosaka K, Nakai Y, Watanabe Y, Tachibana S, Arisue N, Palacpac NM, Toyama T, Honma H, Horii T, Kita K et al. 2011. Concatenated mitochondrial DNA of the coccidian parasite Eimeria tenella. *Mitochondrion* **11**: 273-278.

Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* **16**: 294.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Miranda K, Pace DA, Cintron R, Rodrigues JC, Fang J, Smith A, Rohloff P, Coelho E, de Haas F, de Souza W et al. 2010. Characterization of a novel organelle in Toxoplasma gondii with similar composition and function to the plant vacuole. *Mol Microbiol* **76**: 1358-1375.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24-26.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31.