Supplemental Figures

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

Sivaranjani Namasivayam[1,2], Rodrigo P. Baptista[2,3], Wenyuan Xiao[1,2], Erica M. Hall[1], Joseph S. Doggett[4,5], Karin Troell[6] and Jessica C. Kissinger[1,2,3]

[1] Department of Genetics, University of Georgia, Athens, GA, 30602, USA
[2] Center for Tropical and Emerging Global Diseases, University of Georgia, 30602, USA
[3] Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA
[4] Division of Infectious Diseases, Oregon Health Sciences University, OR, 97239, USA
[5] Division of Infectious Diseases, Veterans Affairs Portland Health Care System, Portland, Oregon 97239, USA
[6] 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

TABLE OF CONTENTS

Fig. S1 – Annotated *Toxoplasma gondii* RH sequenced mtDNA PCR products. Fasta format sequences for these products can be found in Dataset_S1.

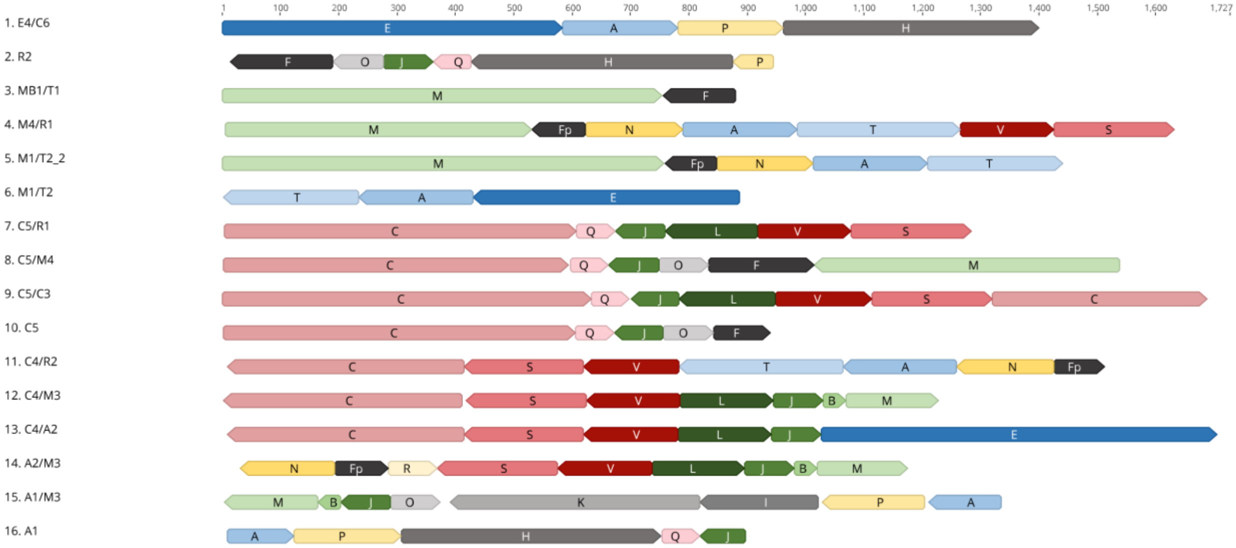
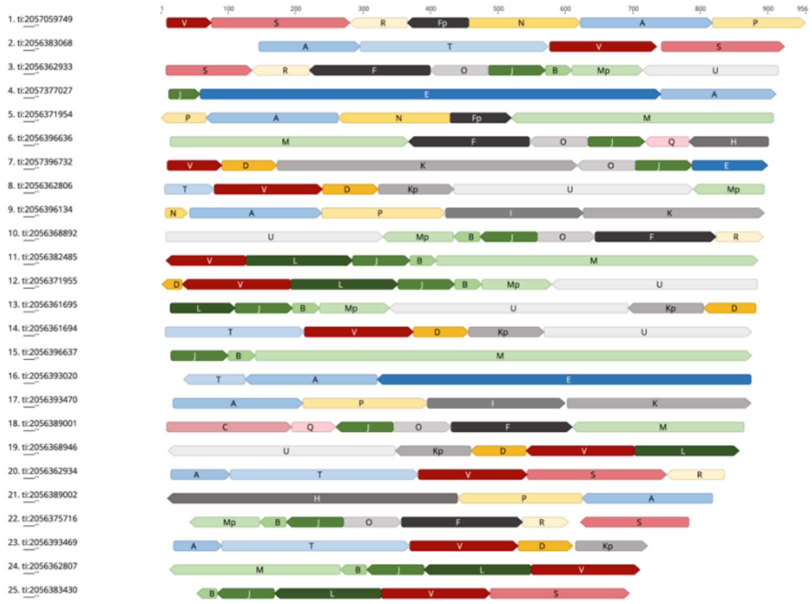
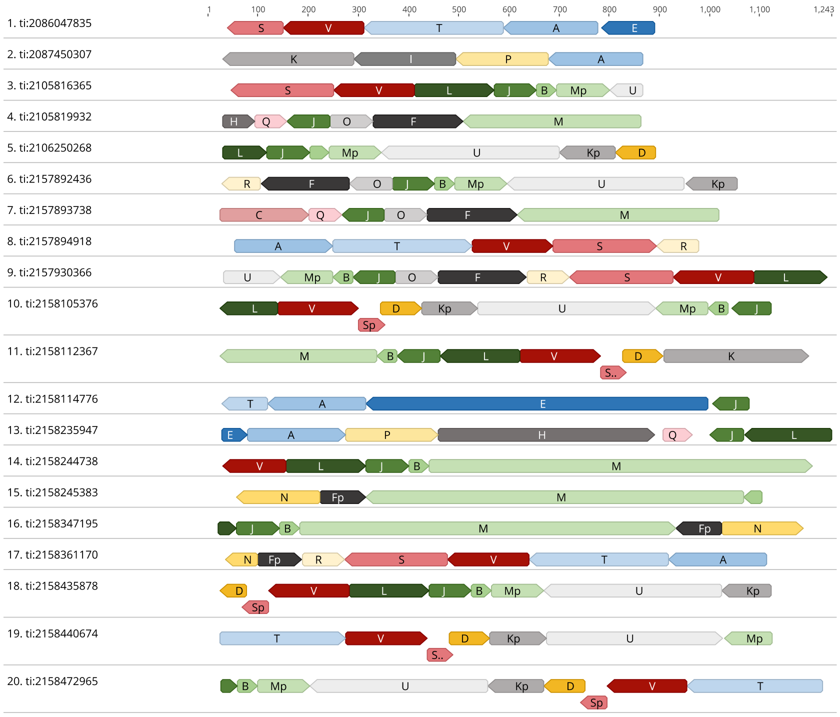


Fig. S2 - Annotation of select A) *T. gondii* and B) *N. caninum* genomic Sanger reads from NCBI SRA database using the 21 sequence blocks from the respective organism

A) *Toxoplasma gondii*



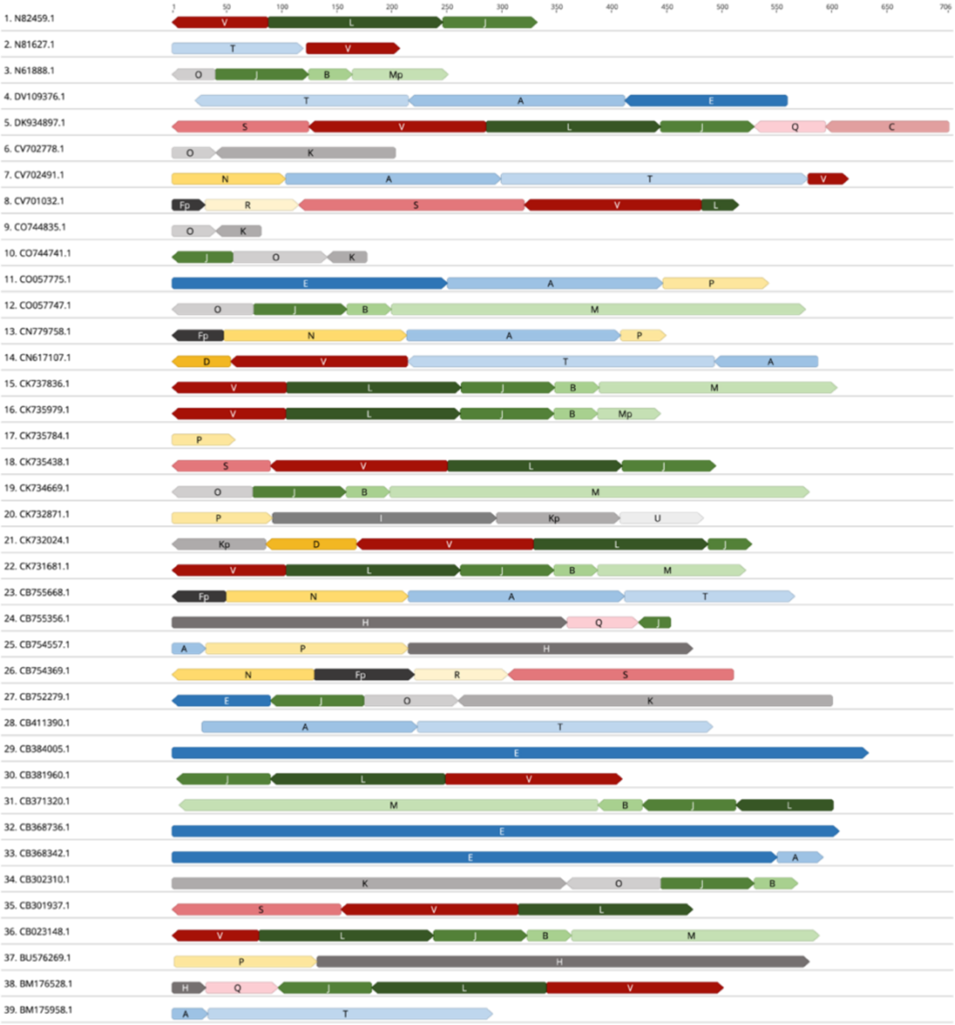
B) *Neospora caninum*



*When block R and F/Fp occur next to each other in *Neospora canin*um, the last 4 bp of F/Fp are missing and instead this junction contains the 2 nt AC/GT depending on orientation

Fig. S3 - Annotation of select A) *T. gondii* and B) *N. caninum* Sanger EST reads from the NCBI EST database using the 21 sequence blocks from the respective organism.

A) Annotated *Toxoplasma gondii* EST reads



B) Annotated *Neospora caninum* Sanger EST reads

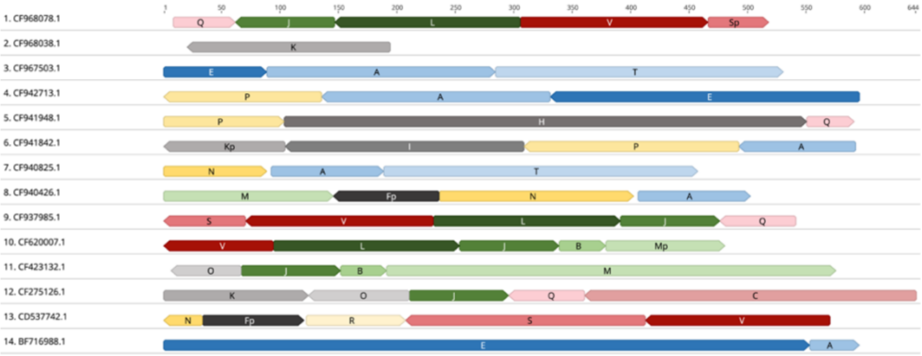
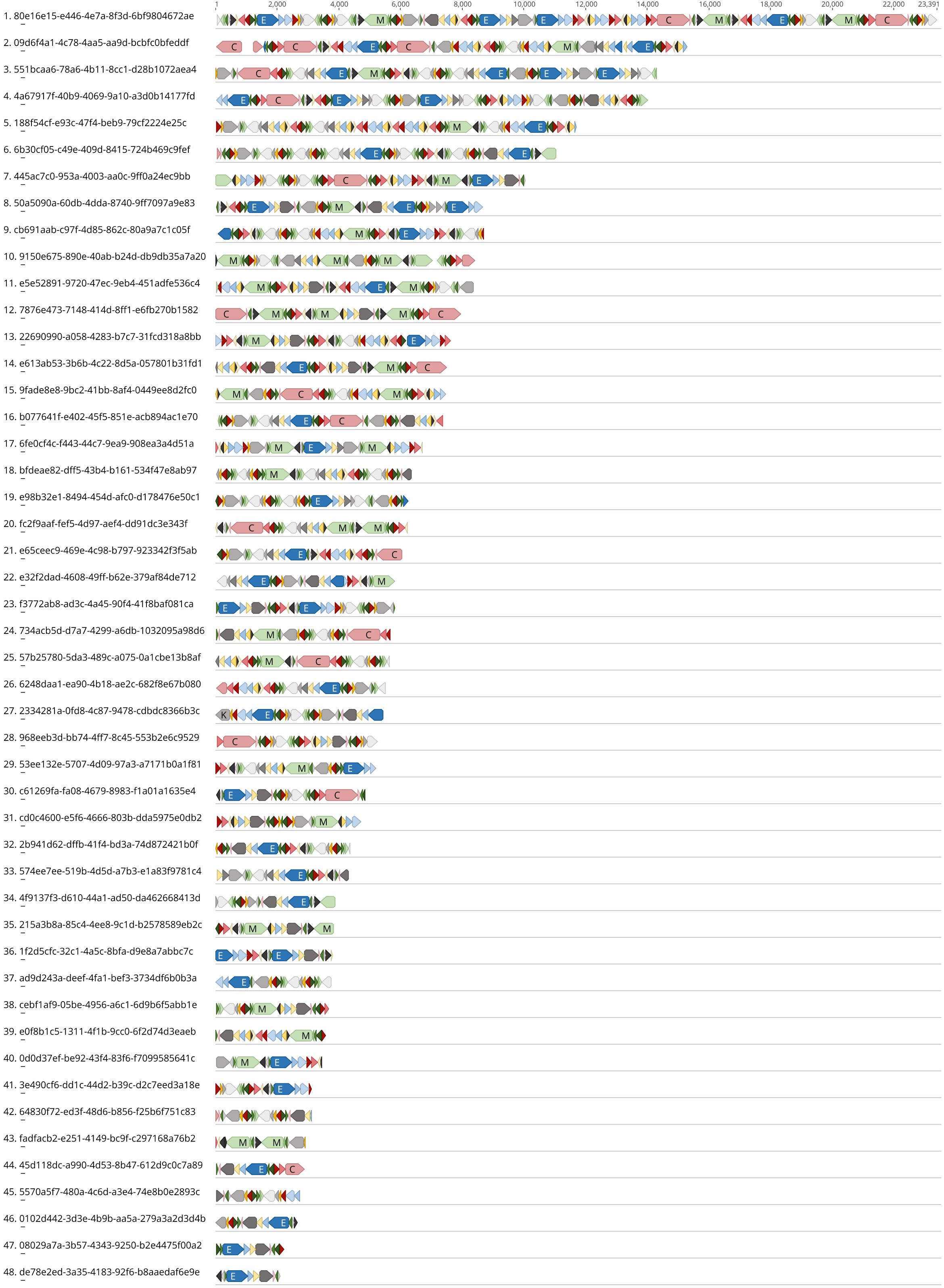
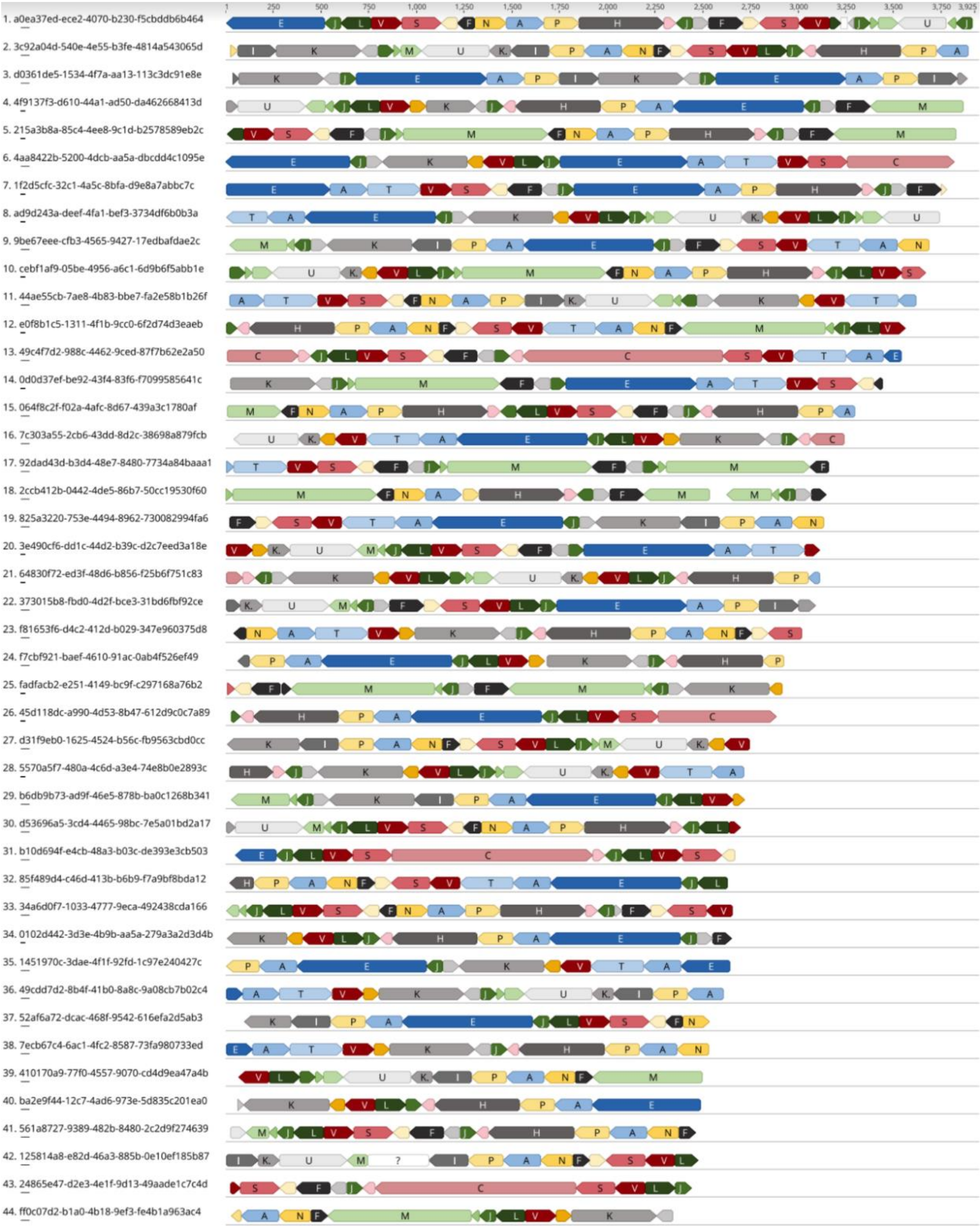


Fig. S4 – Annotated *Toxoplasma gondii* ME49 mtDNA Nanopore reads arranged by length.

a) *Toxoplasma gondii* ME49 Nanopore reads 4-23 kb

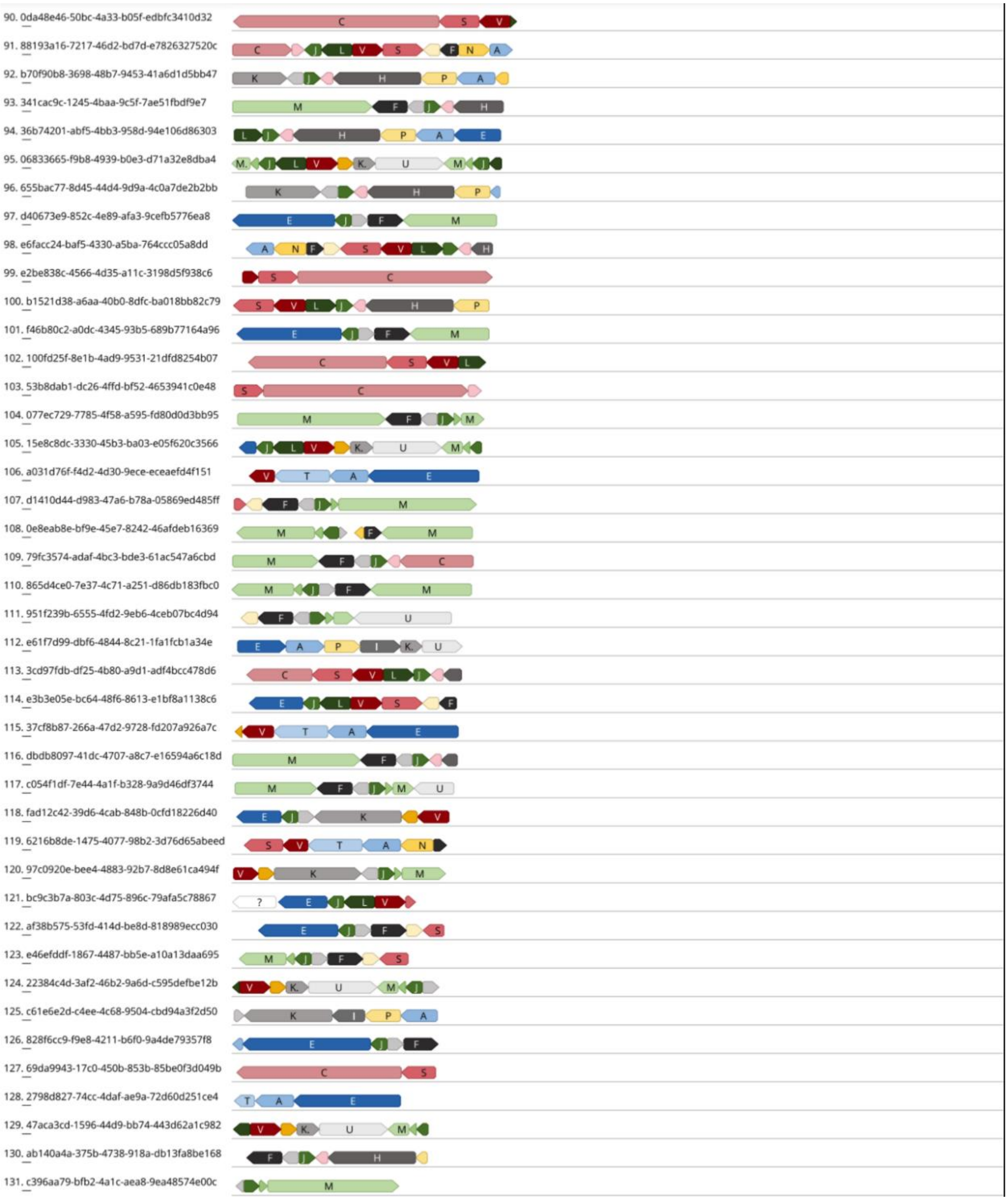


b) *Toxoplasma gondii* ME49 Nanopore reads 1-4 kb



Toxoplasma gondii ME49 Nanopore reads 1-4 kb (continued)

Toxoplasma gondii ME49 Nanopore reads 1-4 kb (continued)



T. gondii	*Th. parva*	*P. falciparum*	mtrRNA segment	*T. gondii* SB	*Th. parva*# Z23263	*P. falciparum*# M76611.1
🟦	🟦	🟦	SSUA	U-204-314	3666-3561	2023-1916
🟦	🟦	🟦	SSUB	K-268-150	4182-4074	505-390
🟦	🟦	🟦	SSUD	M-635-675	3800-3733	5446-5379
🟦	🟦	🟦	SSUE	M-727-754	2095-2057	1638-1680*
🟦	🟦	🟦	SSUF	I-98-156	1840-1900	5507-5447
🟩	🟩	🟩	LSUA	U-37-202	4073-3906	5201-5026
🟩	🟩	🟩	LSUB	W-13-37	3667-3691	4618*-4586
☐	🟩	🟩	LSUC		1824-1839	225*-204
🟩	🟩	🟩	LSUD	H-331-247	3114-3196	5854-5772
🟩	🟩	🟩	LSUE	H-54-244	3199-3377	5771-5577
🟩	🟩	🟩	LSUF	M-366-459	3507-3395	1516-1630
🟩	🟩	🟩	LSUG	M-485-591	5667-5775	389-283
☐	🟧	🟧	RNA1		3827-3905	593-506
☐	🟧	🟧	RNA2		3098-3032	1698-1763
☐	🟧	🟧	RNA3		5658-5584	1830-1910
☐	☐	🟧	RNA4			4625-4696
☐	☐	🟧	RNA5			4716-4802
☐	🟧	🟧	RNA6			4808-4865
☐	🟧	🟧	RNA7		1643-1584	5283-5202
🟧	🟧	🟧	RNA8	F-104-11	1678-1758	5954-5855
☐	🟧	🟧	RNA9		3031-2940	72-125
🟧	🟧	🟧	RNA10	K-445-405;O-17-86	4211-4183	724-625
☐	🟧	🟧	RNA11		4359-4301	5340-5284
☐	🟧	🟧	RNA12		2300-2240	4887-4945
☐	🟧	🟧	RNA13		mtR_52b/35	5025-4996
☐	🟧	🟧	RNA14		4401-4373	5546-5508
☐	🟧	🟧	RNA15		2039-1997	624-594
☐	☐	🟧	RNA16		1759-1790	Mar-33
☐	🟧	🟧	RNA17			126-165
☐	🟧	🟧	RNA18		2192-2154	4964-4988
☐	☐	🟧	RNA19		1518-1494	5576-5547
☐	☐	🟧	RNA20			34-71
☐	☐	🟧	RNA21			1807*-1829
☐	☐	🟧	RNA22			5378-5341

Fig. S5 – Comparative representation of identified mtrRNA segments in the Apicomplexa.

- data from (Feagin et al., 2012). *-inferred coordinates as in Feagin et al., 2012. The *T. gondii* boundaries are inferred. The lack of computationally identified mtrRNAs in *T. gondii* does not mean they are absent. GenBank accession numbers for the mtDNA genome coordinates are as indicated for *Theileria parva* and *Plasmodium falciparum*. SB = sequence block. Blue squares mtSSU segment identified; Green mtLSU segment identified; Orange rRNA segment identified; Blank square, segment not identified (but may be present).

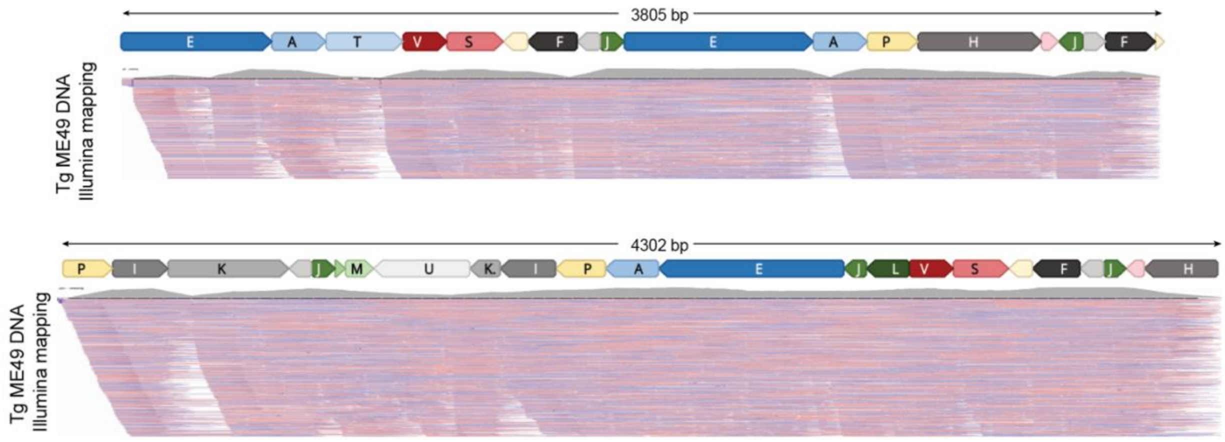
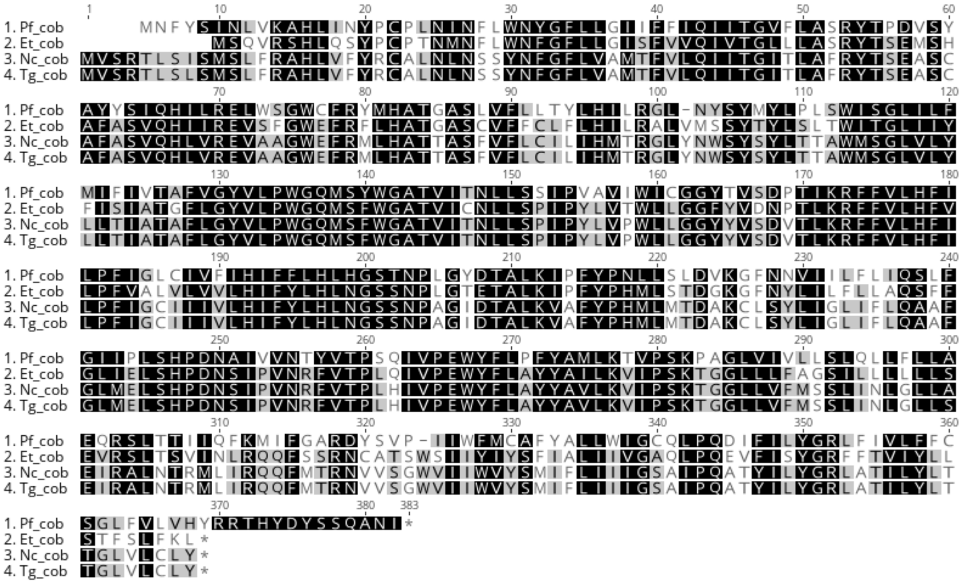


Fig. S6 - Annotated *Toxoplasma gondii* ME49 ONT mtDNA reads with Illumina mapping. Each panel represents the annotation of a single Nanopore read with SBs. There are no intervening nucleotides between SBs. The length of the read is indicated above the annotation. MtDNA-specific paired-end Illumina reads generated from *T. gondii* ME49 DNA (SRR9200762) were mapped to the Nanopore sequences and the alignment was visualized using the Integrated Genomics Viewer (IGV). Both Illumina read ends were required to map at 100% identity. A histogram of read abundance is shown in grey just below each annotated Nanopore read. Red and blue lines below each histogram indicate the mapped paired-end reads. Not all mapped reads are shown.

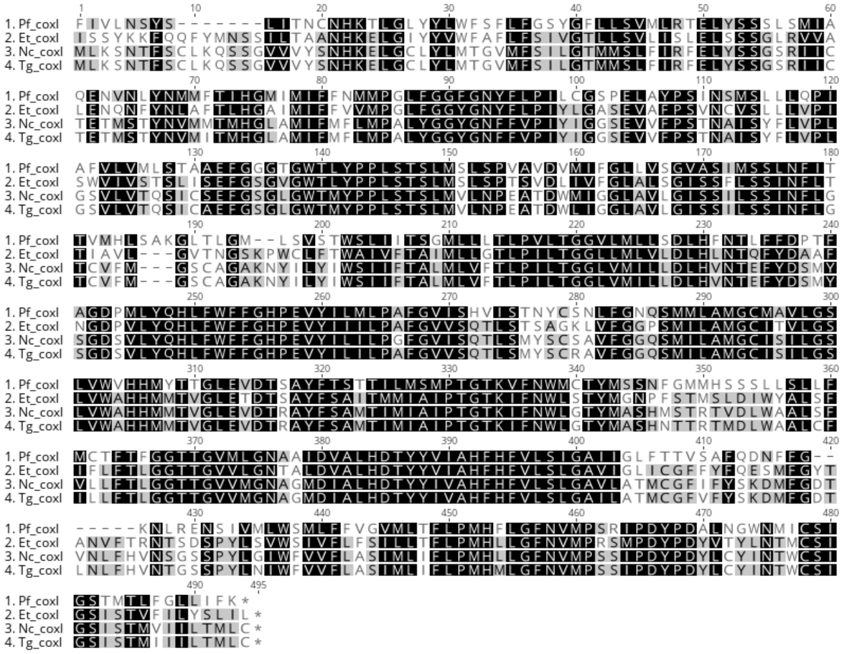
Fig. S7 - Multiple sequence alignments of mitochondrial cytochrome proteins. Multiple sequence alignment of COXI, COXIII and COB proteins (A-C) and genes (D-F) from the apicomplexan parasites *P. falciparum*, *E. tenella*, *T. gondii* and *N. caninum*. *T. gondii* and *N. caninum* cytochrome sequences were identified as described in the methods and results. Codon 'ATA' was annotated as the start Methionine for the COXI protein in *T. gondii* and *N. caninum*. Amino acid and corresponding CDS nucleotide sequences were downloaded from NCBI for *P. falciparum* (AAC63390.2, AAC63389.2, AAC63391.1) and *E. tenella* (BAJ25753.1, BAJ25754.1, BAJ25752.1).

A) Cytochrome B

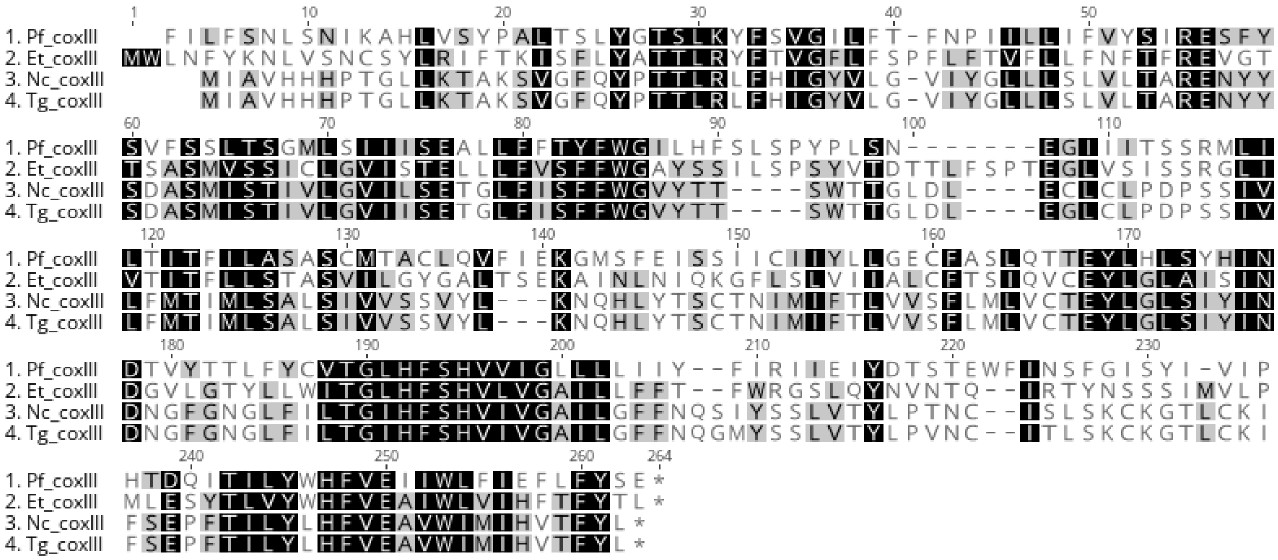


```
1. Pf_cob        MNFYSINLVKAHLINYPCPLNINFLWNYGFLLGIIFFIQIITGVFLASRYTPDVSY
2. Et_cob            MSQVRSHLQSYPCPTNMNFLWNFGFLLGISFVVQIVTGLLLASRYTSEMSH
3. Nc_cob    MVSRTLSISMSLFRAHLVFYRCALNLNSSYNFGFLVAMTFVLQIITGITLAFRYTSEASC
4. Tg_cob    MVSRTLSLSMSLFRAHLVFYRCALNLNSSYNFGFLVAMTFVLQIITGITLAFRYTSEASC

1. Pf_cob    AYYSIQHILRELWSGWCFRYMHATGASLVFLLTYLHILRGL-NYSYMYLPLSWISGLILF
2. Et_cob    AFASVQHIIREVSFGWEFRFLHATGASCVFFCLFLHILRALVMSSYTYLSLTWITGLIIY
3. Nc_cob    AFASVQHLVREVAAGWEFRMLHATTASFVFLCILIHMTRGLYNWSYSYLTTAWMSGLVLY
4. Tg_cob    AFASVQHLVREVAAGWEFRMLHATTASFVFLCILIHMTRGLYNWSYSYLTTAWMSGLVLY

1. Pf_cob    MIFIVTAFVGYVLPWGQMSYWGATVITNLLSSIPVAVIWICGGYTVSDPTIKRFFVLHFI
2. Et_cob    FISIATGFLGYVLPWGQMSFWGATVICNLLSPIPYLVTWLLGGFYVDNPTLKRFFVLHFV
3. Nc_cob    LLTIATAFLGYVLPWGQMSFWGATVITNLLSPIPYLVPWLLGGYYVSDVTLKRFFVLHFI
4. Tg_cob    LLTIATAFLGYVLPWGQMSFWGATVITNLLSPIPYLVPWLLGGYYVSDVTLKRFFVLHFI

1. Pf_cob    LPFIGLCIVFIHIFFLHLHGSTNPLGYDTALKIPFYPNLLSLDVKGFNNVIILFLIQSLF
2. Et_cob    LPFVALVLVVLHIFYLHLNGSSNPLGTETALKIPFYPHMLSTDGKGFNYLILFLLAQSFF
3. Nc_cob    LPFIGCIIIVLHIFYLHLNGSSNPAGIDTALKVAFYPHMLMTDAKCLSYLIGLIFLQAAF
4. Tg_cob    LPFIGCIIIVLHIFYLHLNGSSNPAGIDTALKVAFYPHMLMTDAKCLSYLIGLIFLQAAF

1. Pf_cob    GIIPLSHPDNAIVVNTYVTPSQIVPEWYFLPFYAMLKTVPSKPAGLVIVLLSLQLLFLLA
2. Et_cob    GLIELSHPDNSIPVNRFVTPLQIVPEWYFLAYYAILKVIPSKTGGLLLFAGSILLLLLLS
3. Nc_cob    GLMELSHPDNSIPVNRFVTPLHIVPEWYFLAYYAVLKVIPSKTGGLLVFMSSLINLGLLA
4. Tg_cob    GLMELSHPDNSIPVNRFVTPLHIVPEWYFLAYYAVLKVIPSKTGGLLVFMSSLINLGLLS

1. Pf_cob    EQRSLTTIIQFKMIFGARDYSVP-IIWFMCAFYALLWIGCQLPQDIFILYGRLFIVLFFC
2. Et_cob    EVRSLTSVINLRQQFSSRNCATSWSIIYIYSFIALIIVGAQLPQEVFISYGRFFTVIYLL
3. Nc_cob    EIRALNTRMLIRQQFMTRNVVSGWVIIWVYSMIFLIIIGSAIPQATYILYGRLATILYLT
4. Tg_cob    EIRALNTRMLIRQQFMTRNVVSGWVIIWVYSMIFLIIIGSAIPQATYILYGRLATILYLT

1. Pf_cob    SGLFVLVHYRRTHYDYSSQANI*
2. Et_cob    STFSLFKL*
3. Nc_cob    TGLVLCLY*
4. Tg_cob    TGLVLCLY*
```

B) Cytochrome oxidase I



C) Cytochrome oxidase III

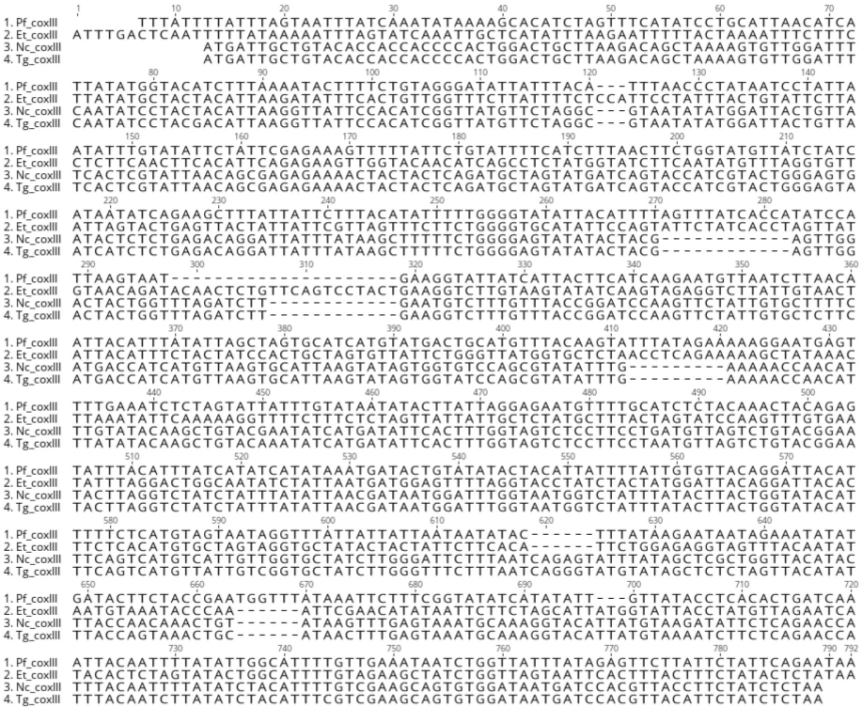


D) Cytochrome B

```
          1        10        20        30        40        50        60        70        80        90        100       110
1. Pf_cob             ATGAACTTTTACTCTATTAATTTAGTTAAAGCACACTTAATAAATTACCCATGTCCATTGAACATAAACTTTTTATGGAATTACGGATTCCTTTTAGGAATAA
2. Et_cob                                           ATGTCTCAAGTGAGATCTCACCTACAATCATATCCATGTCCAACCAATATGAACTTCCTTTGGAATTTTGGTTTCTTATTAGGAATTT
3. Nc_cob ATGGTTTCGAGAACACTCAGTATATCCATGAGTCTATTCCGGGCACACCTTGTCTTTTATCGGTGTGCTCTAAATCTAAATTCATCTTATAACTTTGGTTTTTTAGTTGCAATGA
4. Tg_cob ATGGTTTCGAGAACACTCAGTCTATCTATGAGTCTATTCCGGGCACACCTTGTCTTTTATCGGTGTGCTCTAAATCTAAATTCATCTTATAATTTTGGTTTCTTAGTTGCAATGA
              120       130       140       150       160       170       180       190       200       210       220       230
1. Pf_cob TATTTTTTATTCAAATTATAACAGGTGTATTTTTAGCAAGTCGATATACACCAGATGTTTCATATGCATATTATAGTATACAACACATTTTAAGAGAATTATGGAGTGGATGGTG
2. Et_cob CTTTTGTTGTCCAAATTGTAACAGGATTATTATTAGCATCTAGATATACTAGTGAAATGTCACATGCCTTTGCTAGTGTACAACATATTATTCGAGAGGTGTCTTTCGGTTGGGA
3. Nc_cob CCTTTGTACTCCAAATAATTACAGGTATCACTCTAGCATTTAGATATACTTCAGAAGCATCTTGTGCATTTGCTAGCGTTCAACATCTAGTTAGAGAAGTAGCAGCAGGATGGGA
4. Tg_cob CCTTTGTACTCCAAATAATTACAGGTATCACTCTAGCGTTTAGATATACTTCTGAAGCATCTTGTGCATTTGCTAGCGTTCAACATCTAGTTAGAGAAGTAGCAGCAGGATGGGA
                   240       250       260       270       280       290       300       310       320       330       340
1. Pf_cob TTTTAGATACATGCACGCAACAGGTGCTTCTCTTGTATTTTTATTAACATATCTTCATATTTTAAGAGGATTA---AATTACTCATATATGTATTTACCATTATCATGGATATCT
2. Et_cob ATTTAGATTCCTACATGCAACTGGAGCATCCTGCGTATTCTTCTGTCTATTTCTTCATATTCTTCGAGCTCTAGTAATGAGTAGTTATACCTATCTAAGTCTGACATGGATTACT
3. Nc_cob ATTTAGGATGTTGCATGCTACAACTGCCTCTTTCGTATTCTTGTGTATTTTAATTCACATGACTCGAGGACTGTATAACTGGAGTTATAGTTATTTAACTACCGCTTGGATGTCT
4. Tg_cob ATTTAGGATGTTGCATGCTACAACAGCCTCTTTTGTATTCTTGTGTATTTTAATTCACATGACTCGAGGATTGTATAACTGGAGTTATAGTTATTTAACTACCGCTTGGATGTCT
              350       360       370       380       390       400       410       420       430       440       450       460
1. Pf_cob GGATTGATTTTATTTATGATATTTATTGTAACTGCTTTCGTTGGTTATGTCTTACCATGGGGTCAAATGAGTTATTGGGGTGCAACTGTAATTACTAACTTGTTATCCTCTATTC
2. Et_cob GGACTTATTATCTACTTTATTTCTATCGCAACAGGATTCTTAGGTTATGTACTACCATGGGGTCAAATGAGTTTCTGGGGTGCTACTGTAATTTGTAACCTACTCTCACCAATTC
3. Nc_cob GGTTTAGTTTTATATCTACTTACTATAGCCACTGCCTTCCTTGGATATGTACTACCATGGGGACAAATGAGTTTCTGGGGTGCTACAGTGATTACAAACCTCCTTTCTCCAATAC
4. Tg_cob GGTTTAGTTTTATATTTACTTACTATAGCCACTGCCTTCCTCGGATATGTACTACCATGGGGACAAATGAGTTTCTGGGGTGCTACAGTCATTACAAACCTCCTTTCTCCAATAC
                   470       480       490       500       510       520       530       540       550       560       570
1. Pf_cob CAGTAGCAGTAATTTGGATATGTGGAGGATATACTGTGAGTGATCCTACAATAAAACGATTTTTTGTACTACATTTTATCTTACCATTTATTGGATTATGTATTGTATTTATACA
2. Et_cob CATACCTTGTAACTTGGTTACTAGGAGGTTTCTATGTGGATAATCCTACCTTAAAAAGATTCTTTGTATTACATTTCGTACTTCCATTTGTAGCTCTTGTACTTGTAGTTCTACA
3. Nc_cob CATATTTGGTACCTTGGCTACTTGGAGGATACTACGTATCTGATGTAACATTAAAACGATTCTTTGTATTACATTTTATATTGCCTTTTATAGGTTGTATTATAATAGTTTTACA
4. Tg_cob CATATTTGGTACCTTGGCTACTTGGAGGTTACTACGTATCTGATGTAACATTAAAACGATTCTTTGTATTACATTTTATCTTGCCTTTTATTGGTTGTATTATAATAGTTTTACA
              580       590       600       610       620       630       640       650       660       670       680       690
1. Pf_cob TATATTTTTCTTACATTTACATGGTAGCACAAATCCTTTAGGGTATGATACAGCATTAAAAATACCCTTTTATCCAAATCTATTAAGTCTTGATGTTAAAGGATTTAATAATGTT
2. Et_cob TATTTTCTATCTACATCTAAACGGATCTAGTAACCCACTGGGTACAGAAACTGCTTTAAAAATACCATTCTATCCTCATATGTTAAGTACAGATGGAAAAGGGTTTAACTACTTA
3. Nc_cob TATCTTCTACTTACATTTAAATGGTTCTAGTAACCCTGCAGGTATAGATACCGCGCTTAAAGTTGCCTTTTATCCTCATATGTTAATGACGGATGCTAAATGTCTATCATACTTA
4. Tg_cob TATCTTCTATTTACATTTAAATGGTTCTAGCAATCCTGCAGGTATTGATACCGCGCTTAAAGTTGCCTTTTATCCTCATATGTTAATGACAGATGCTAAATGTCTATCCTATTTA
                   700       710       720       730       740       750       760       770       780       790       800
1. Pf_cob ATAATTTTATTTCTAATACAAAGTTTATTTGGAATTATACCTTTATCACATCCTGATAATGCTATCGTAGTAAATACATATGTTACTCCATCTCAAATTGTACCTGAATGGTACT
2. Et_cob ATCTTATTCCTATTAGCTCAATCATTCTTCGGTCTAATTGAATTATCACATCCAGATAACAGTATTCCTGTAAATAGATTTGTAACACCACTACAAATTGTACCAGAGTGGTACT
3. Nc_cob ATTGGATTAATCTTCTTACAAGCGGCTTTCGGTTTGATGGAACTTTCACACCCAGATAACTCCATACCAGTCAACCGGTTCGTAACACCACTTCATATCGTACCAGAATGGTATT
4. Tg_cob ATTGGATTAATCTTCTTACAAGCGGCTTTTGGTTTGATGGAACTATCACATCCAGATAACTCAATACCAGTCAACCGGTTCGTAACACCACTTCATATCGTACCTGAATGGTATT
              810       820       830       840       850       860       870       880       890       900       910       920
1. Pf_cob TTCTACCATTTTATGCAATGTTAAAAACTGTTCCAAGTAAACCAGCTGGTTTAGTAATTGTATTATTATCATTACAATTATTATTCTTATTAGCAGAACAAAGAAGTTTAACAAC
2. Et_cob TCTTAGCATATTATGCTATCTTAAAAGTTATTCCAAGTAAAACTGGAGGTCTATTACTTTTCGCGGGTAGCATTCTATTATTATTACTTCTAAGTGAGGTTCGATCACTTACTAG
3. Nc_cob TTCTAGCATATTATGCGGTGTTAAAAGTAATACCATCCAAAACCGGTGGTTTGTTAGTATTTATGTCATCCCTTATTAATCTAGGTTTATTAGCAGAGATTCGAGCTTTAAATAC
4. Tg_cob TTCTAGCATATTATGCGGTGTTAAAAGTAATTCCATCCAAAACCGGTGGTTTGTTAGTATTTATGTCATCTCTGATTAATTTAGGTTTACTTTCTGAGATTCGAGCTTTAAATAC
                   930       940       950       960       970       980       990      1,000     1,010     1,020     1,030
1. Pf_cob TATAATTCAATTTAAAATGATTTTTGGTGCTAGAGATTATTCTGTTCCT---ATTATATGGTTTATGTGTGCATTCTATGCTTTATTATGGATTGGATGTCAATTACCACAAGAT
2. Et_cob TGTAATAAACTTACGACAACAATTCTCTTCAAGAAATTGTGCAACATCTTGGAGTATTATCTATATCTACTCATTTATTGCTCTTATTATTGTTGGAGCTCAATTACCTCAAGAA
3. Nc_cob CCGAATGTTGATTCGTCAACAATTTATGACTCGAAATGTAGTCAGTGGATGGGTTATTATTTGGGTATATAGTATGATCTTCTTGATTATTATAGGTAGTGCTATTCCACAAGCG
4. Tg_cob CCGAATGTTAATTCGTCAACAGTTCATGACTCGAAATGTAGTCAGTGGATGGGTGATTATTTGGGTATATAGTATGATCTTCTTGATTATTATAGGTAGTGCTATTCCACAAGCA
             1,040     1,050     1,060     1,070     1,080     1,090     1,100     1,110     1,120     1,130     1,140   1,149
1. Pf_cob ATATTCATTTTATATGGTCGATTATTTATTGTATTATTTTTCTGTAGTGGTTTATTTGTACTTGTTCATTATAGACGAACACATTATGATTACAGCTCCCAAGCAAACATATAA
2. Et_cob GTATTTATTTCATATGGTAGATTCTTTACCGTAATCTATCTTTTAAGTACATTTAGTTTATTCAAACTGTAA
3. Nc_cob ACTTACATCTTATATGGTAGACTAGCTACTATCTTATATCTTACTACCGGATTGGTACTATGCTTATACTAA
4. Tg_cob ACATACATCTTATATGGTAGATTAGCTACTATCTTATACCTTACTACCGGATTGGTTCTATGCTTATACTAA
```

E) Cytochrome oxidase I

F) Cytochrome oxidase III



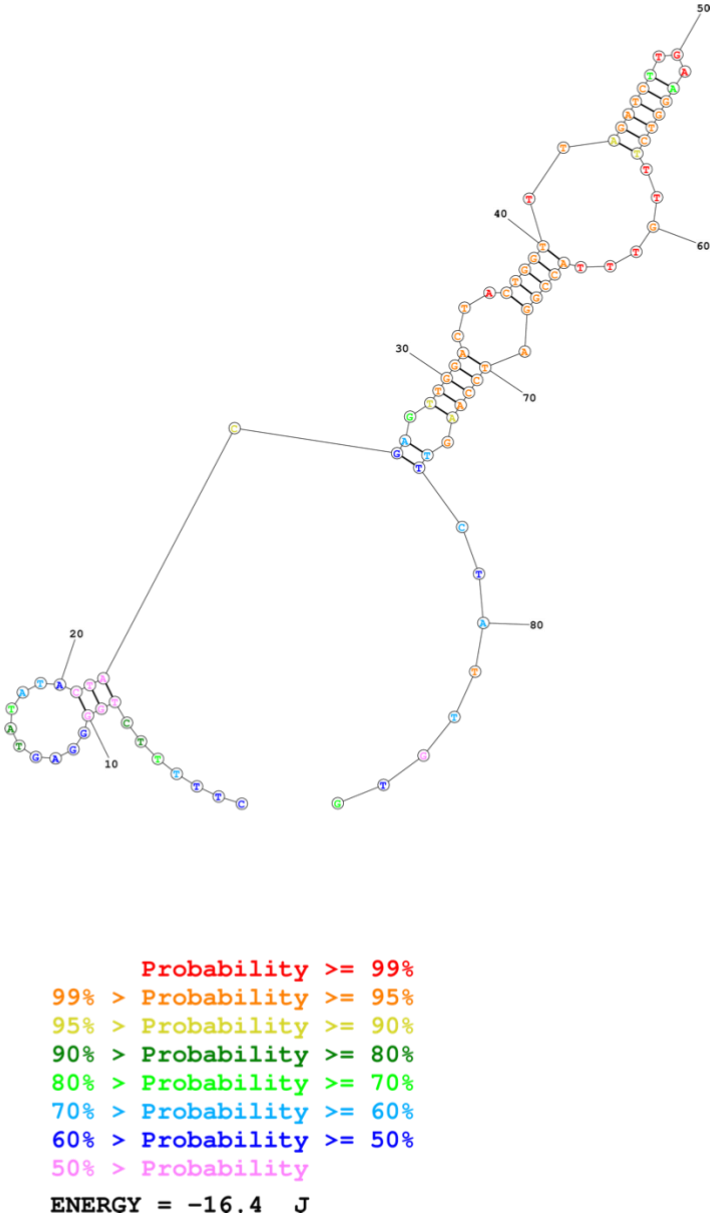
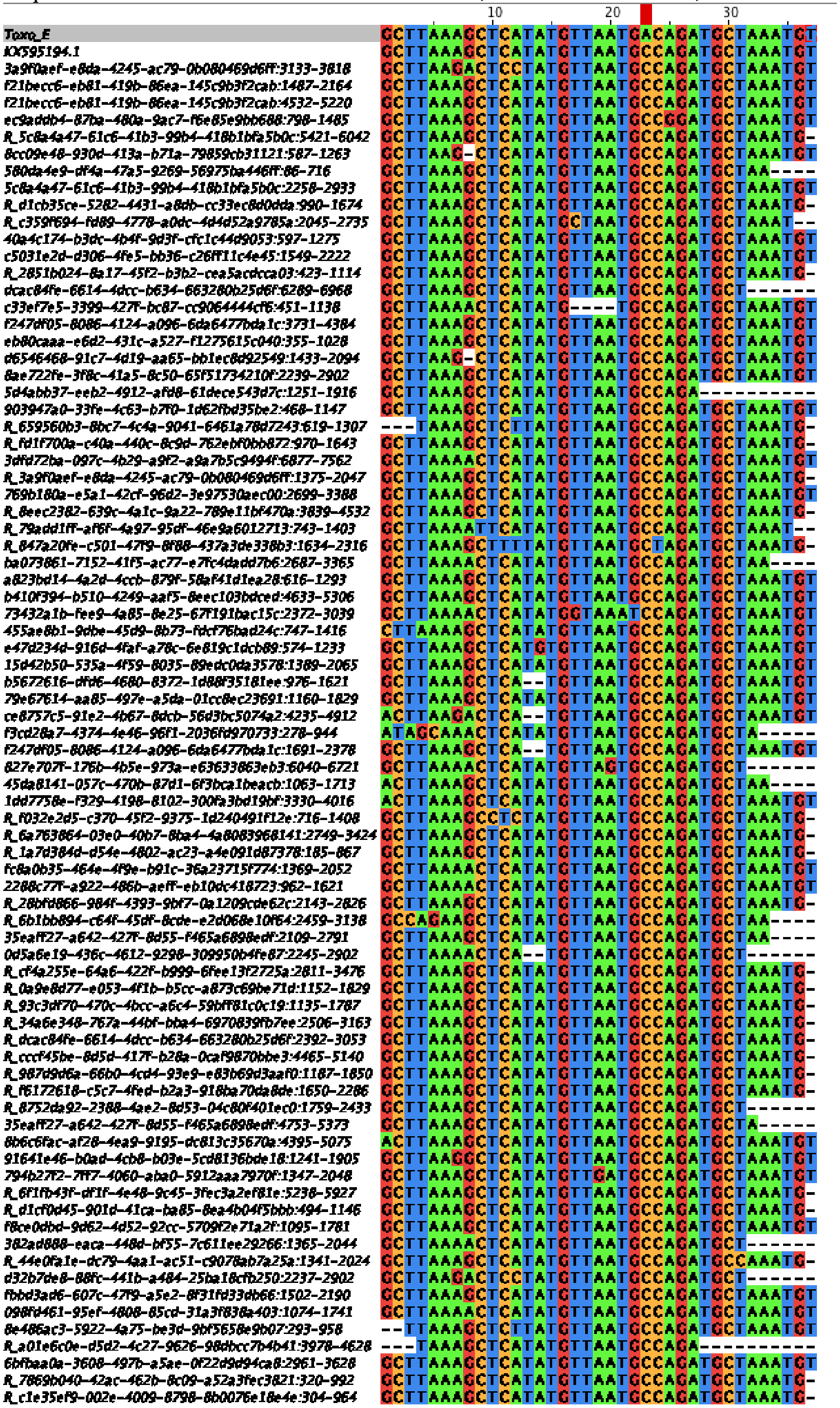


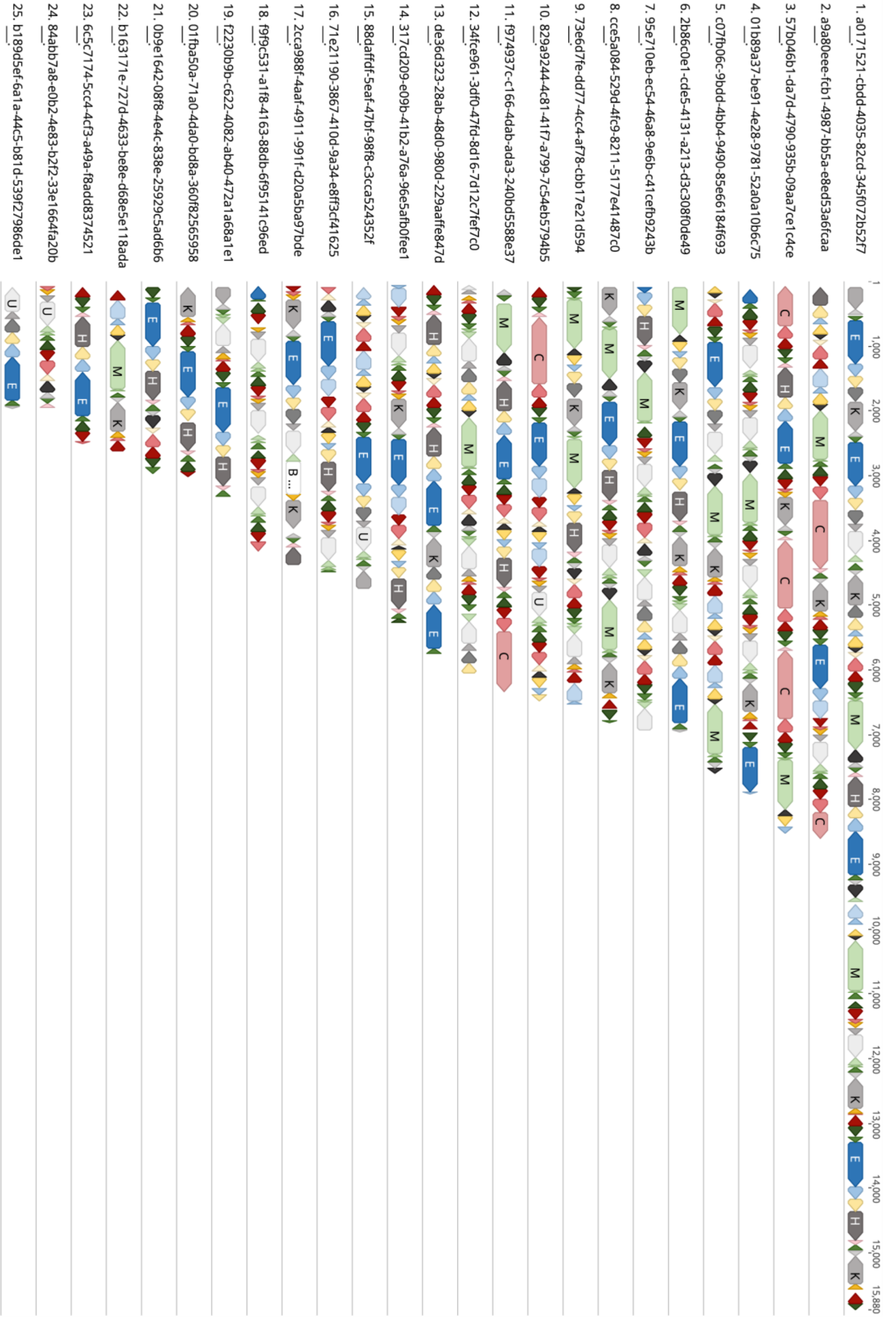
Fig. S8 - Secondary structure of *T. gondii* sequence block J. Structure analysis was performed with RNA structure Web Servers for RNA Secondary Structure Prediction. The Mathews Group. https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html
Settings: MaxExpect partition.pfs MaxExpect.ct --gamma 1 --percent 10 --structures 20 --window 3

Fig. S9 - Multiple sequence alignment of a subset of uncorrected Nanopore reads from the ENU mutant strain ELQ-316. The wt *T. gondii* RH sequence is on the first row. The mutant sequence is on the second row. The mutation, which is an A -> C, is located at position 23.



```
Toxo_E                                              GCTTAAAGCTCATATGTTAATGACAGATGCTAAATGT
KX595194.1                                          GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
3a9f0aef-e8da-4245-ac79-0b080469d6ff:3133-3818      GCTTAAGACTCCTATGTTAATGCCAGATGCTAAATGT
f21becc6-eb81-419b-86ea-145c9b3f2cab:1487-2164      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
f21becc6-eb81-419b-86ea-145c9b3f2cab:4532-5220      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
ec9addb4-87ba-480a-9ac7-f6e85e9bb688:798-1485       GCTTAAAGCTCATATGTTAATGCCGGATGCTAAATGT
R_5c8a4a47-61c6-41b3-99b4-418b1bfa5b0c:5421-6042    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
8cc09e48-930d-413a-b71a-79859cb31121:587-1263       GCTTAAG-CTCATATGTTAATGCCAGATGCTAAATGT
580da4e9-df4a-47a5-9269-56975ba446ff:86-716         GCTTAAAGCTCATATGTTAATGCCAGATGCTAA----
5c8a4a47-61c6-41b3-99b4-418b1bfa5b0c:2258-2933      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
R_d1cb35ce-5282-4431-a8db-cc33ec8d0dda:990-1674     GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_c359f694-fd89-4778-a0dc-4d4d52a9785a:2045-2735    GCTTAAAGCTCATATGCTAATGCCAGATGCTAAAT--
40a4c174-b3dc-4b4f-9d3f-cfc1c44d9053:597-1275       GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
c5031e2d-d306-4fe5-bb36-c26ff11c4e45:1549-2222      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
R_2851b024-8a17-45f2-b3b2-cea5acdcca03:423-1114     GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
dcac84fe-6614-4dcc-b634-663280b25d6f:6289-6968      GCTTAAAGCTCATATGTTAATGCCAGATGCT------
c33ef7e5-3399-427f-bc87-cc9064444cf6:451-1138       GCTTAAAACTCATATG----TGCCAGATGCTAAATGT
f247df05-8086-4124-a096-6da6477bda1c:3731-4384      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
eb80caaa-e6d2-431c-a527-f1275615c040:355-1028       GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
d6546468-91c7-4d19-aa65-bb1ec8d92549:1433-2094      GCTTAAG-CTCATATGTTAATGCCAGATGCTAAATGT
8ae722fe-3f8c-41a5-8c50-65f51734210f:2239-2902      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
5d4abb37-eeb2-4912-afd8-61dece543d7c:1251-1916      GCTTAAAGCTCATATGTTAATGCCAGA----------
903947a0-33fe-4c63-b7f0-1d62fbd35be2:468-1147       GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
R_659560b3-8bc7-4c4a-9041-6461a78d7243:619-1307     ---TAAAGCTCTTATGTTAATGCCAGATGCTAAATG-
R_fd1f700a-c40a-440c-8c9d-762ebf0bb872:970-1643     GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
3dfd72ba-097c-4b29-a9f2-a9a7b5c9494f:6877-7562      GCTTAAAACTCATATGTTAATGCCAGATGCTAAATGT
R_3a9f0aef-e8da-4245-ac79-0b080469d6ff:1375-2047    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
769b180a-e5a1-42cf-96d2-3e97530aec00:2699-3388      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
R_8eec2382-639c-4a1c-9a22-789e11bf470a:3839-4532    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_79add1ff-af6f-4a97-95df-46e9a6012713:743-1403     GCTTAAAATTCATATGTTAATGCCAGATGCTAAAT--
R_847a20fe-c501-47f9-8f88-437a3de338b3:1634-2316    GCTTAAAGCTTTTATGTTAATGCTAGATGCTAAATG-
ba073861-7152-41f5-ac77-e7fc4dadd7b6:2687-3365      GCTTAAAACTCATATGTTAATGCCAGATGCTAA----
a823bd14-4a2d-4ccb-879f-58af41d1ea28:616-1293       GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
b410f394-b510-4249-aaf5-8eec103bdced:4633-5306      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
73432a1b-fee9-4a85-8e25-67f191bac15c:2372-3039      GCTTAAAACTCATATGGTAAATCCAGATGCTAAATGT
455ae8b1-9dbe-45d9-8b73-fdcf76bad24c:747-1416       CTTAAAAGCTCATATGTTAATGCCAGATGCTAAATGT
e47d234d-916d-4faf-a78c-6e819c1dcb89:574-1233       GCTTAAAGCTCATGTGTTAATGCCAGATGCTAAATGT
15d42b50-535a-4f59-8035-89edc0da3578:1389-2065      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
b5672616-dfd6-4680-8372-1d88f35181ee:976-1621       GCTTAAAGCTCA--TGTTAATGCCAGATGCTAAATGT
79e67614-aa85-497e-a5da-01cc8ec23691:1160-1829      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
ce8757c5-91e2-4b67-8dcb-56d3bc5074a2:4235-4912      ACTTAAGACTCA--TGTTAATGCCAGATGCTAAATGT
f3cd28a7-4374-4e46-96f1-2036fd970733:278-944        ATAGCAAACTCATATGTTAATGCCAGATGCTA-----
f247df05-8086-4124-a096-6da6477bda1c:1691-2378      GCTTAAAGCTCA--TGTTAATGCCAGATGCTAAATGT
827e707f-176b-4b5e-973a-e63633863eb3:6040-6721      GCTTAAAACTCATATGTTAGTGCCAGATGCT------
45da8141-057c-470b-87d1-6f3bca1beacb:1063-1713      ACTTAAAGCTCATATGTTAATGCCAGATGCTAA----
1dd7758e-f329-4198-8102-300fa3bd19bf:3330-4016      ACTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
R_f032e2d5-c370-45f2-9375-1d240491f12e:716-1408     GCTTAAAGCCTCTATGTTAATGCCAGATGCTAAATG-
R_6a763864-03e0-40b7-8ba4-4a8083968141:2749-3424    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_1a7d384d-d54e-4802-ac23-a4e091d87378:185-867      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
fc8a0b35-464e-4f9e-b91c-36a23715f774:1369-2052      GCTTAAAACTCATATGTTAATGCCAGATGCTAAATGT
2288c77f-a922-486b-aeff-eb10dc418723:962-1621       GCTTAAAACTCATATGTTAATGCCAGATGCTAAATGT
R_28bfd866-984f-4393-9bf7-0a1209cde62c:2143-2826    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_6b1bb894-c64f-45df-8cde-e2d068e10f64:2459-3138    GCCAGAAGCTCATATGTTAATGCCAGATGCTAA----
35eaff27-a642-427f-8d55-f465a6898edf:2109-2791      GCTTAAAGCTCATATGTTAATGCCAGATGCTAA----
0d5a6e19-436c-4612-9298-309950b4fe87:2245-2902      GCTTAAAACTCA--TGTTAATGCCAGATGCT------
R_cf4a255e-64a6-422f-b999-6fee13f2725a:2811-3476    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_0a9e8d77-e053-4f1b-b5cc-a873c69be71d:1152-1829    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_93c3df70-470c-4bcc-a6c4-59bff81c0c19:1135-1787    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_34a6e348-767a-44bf-bba4-6970839fb7ee:2506-3163    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_dcac84fe-6614-4dcc-b634-663280b25d6f:2392-3053    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_cccf45be-8d5d-417f-b28a-0caf9870bbe3:4465-5140    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_987d9d6a-66b0-4cd4-93e9-e83b69d3aaf0:1187-1850    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_f6172618-c5c7-4fed-b2a3-918ba70da8de:1650-2286    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_8752da92-2388-4ae2-8d53-04c80f401ec0:1759-2433    GCTTAAAGCTCATATGTTAATGCCAGATGCT------
35eaff27-a642-427f-8d55-f465a6898edf:4753-5373      GCTTAAAGCTCATATGTTAATGCCAGATGCTA-----
8b6c6fac-af28-4ea9-9195-dc813c35670a:4395-5075      ACTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
91641e46-b0ad-4cb8-b03e-5cd8136bde18:1241-1905      GCTTAAGGCTCATATGTTAATGCCAGATGCTAAATGT
794b27f2-7ff7-4060-aba0-5912aaa7970f:1347-2048      GCTTAAAGCTCATATGTTGATGCCAGATGCTAAATGT
R_6f1fb43f-df1f-4e48-9c45-3fec3a2ef81e:5238-5927    GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_d1cf0d45-901d-41ca-ba85-8ea4b04f5bbb:494-1146     GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
f8ce0dbd-9d62-4d52-92cc-5709f2e71a2f:1095-1781      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
382ad888-eaca-448d-bf55-7c611ee29266:1365-2044      GCTTAAAACTCATATGTTAATGCCAGATGCT------
R_44e0fa1e-dc79-4aa1-ac51-c9078ab7a25a:1341-2024    GCTTAAAGCTCATATGTTAATGCCAGATGCCAAATG-
d32b7de8-88fc-441b-a484-25ba18cfb250:2237-2902      GCTTAAGACTCCTATGTTAATGCCAGATGCT------
fbbd3ad6-607c-47f9-a5e2-8f31fd33db66:1502-2190      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
098fd461-95ef-4808-85cd-31a3f838a403:1074-1741      GCTTAAAACTCATATGTTAATGCCAGATGCTAAATGT
8e486ac3-5922-4a75-be3d-9bf5658e9b07:293-958        --TTAAAGCTCTTATGTTAATGCCAGATGCTAAATGT
R_a01e6c0e-d5d2-4c27-9626-98dbcc7b4b41:3978-4628    ---TAAAGCTCATATGTTAATGCCAGA----------
6bfbaa0a-3608-497b-a5ae-0f22d9d94ca8:2961-3628      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATGT
R_7869b040-42ac-462b-8c09-a52a3fec3821:320-992      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
R_c1e35ef9-002e-4009-8798-8b0076e18e4e:304-964      GCTTAAAGCTCATATGTTAATGCCAGATGCTAAATG-
```

Fig. S10 – Annotated *Neospora caninum* mtDNA Nanopore reads (2-15 kb) arranged by length



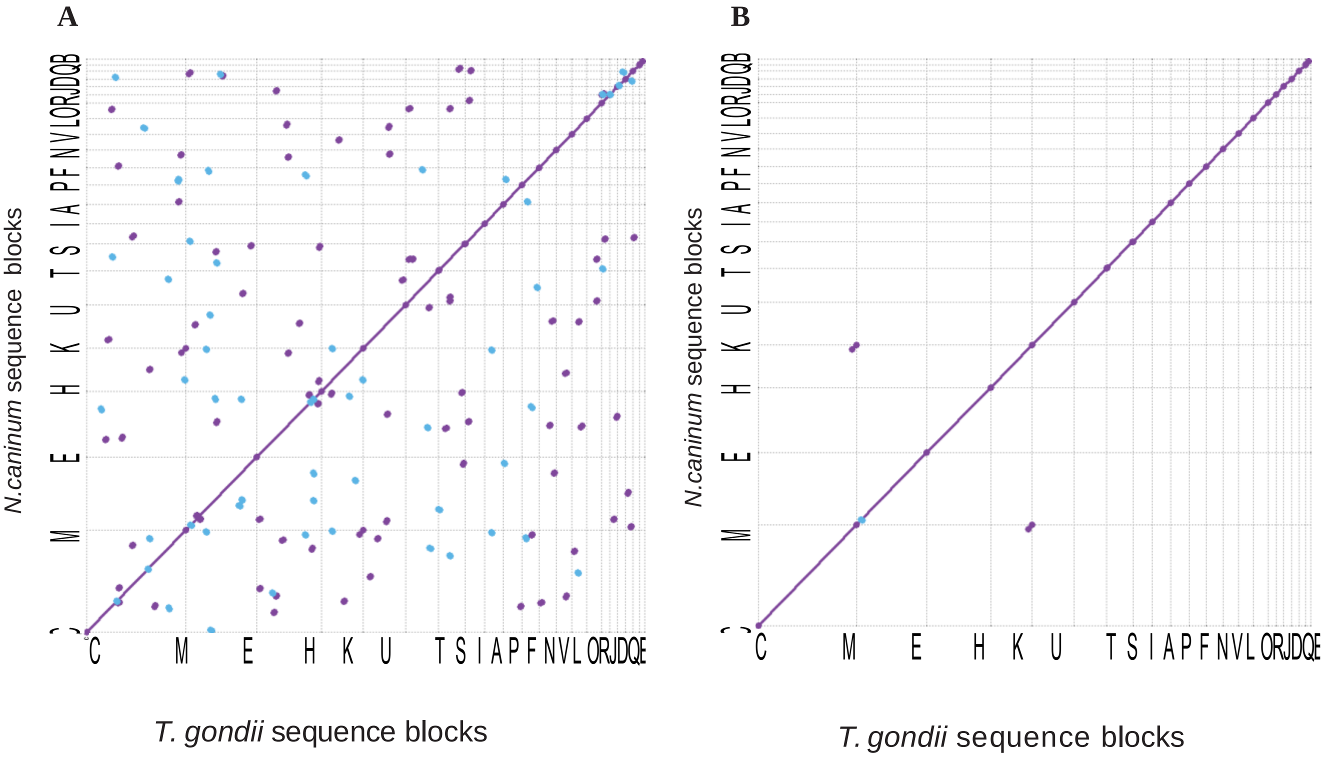
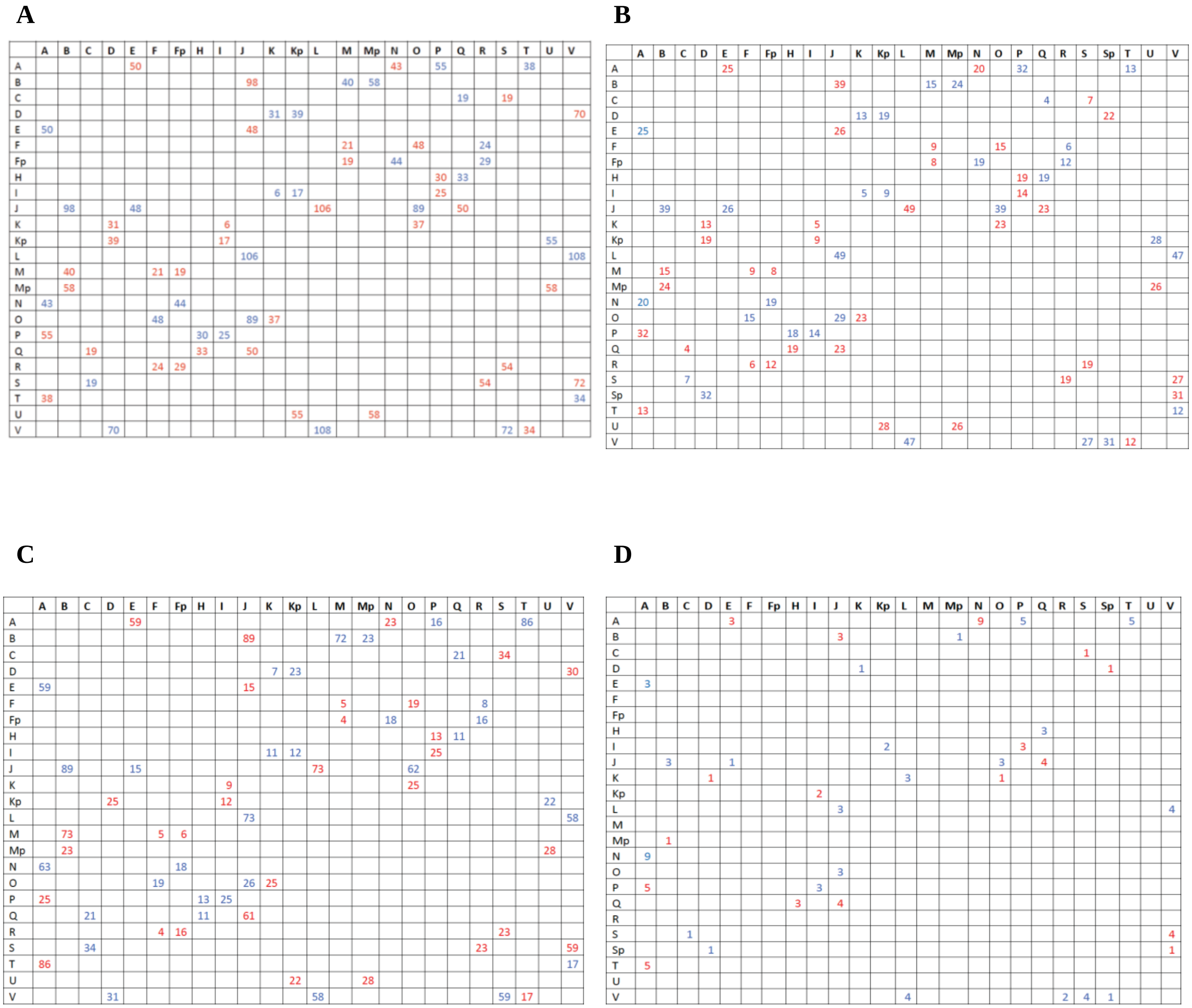


Fig. S11 - Dot plot comparisons of *T. gondii* ME49 and *N. caninum* Nc-1 sequence blocks to detect microhomology. (A-B) The 21 *T. gondii* mtDNA SBs were aligned to the 21 mtDNA blocks from *N. caninum*. Matches detected with window sizes of 10 bp in (A,) and 15 bp in (B) are displayed in blue or purple for forward and reverse matches respectively.



A

	A	B	C	D	E	F	Fp	H	I	J	K	Kp	L	M	Mp	N	O	P	Q	R	S	T	U	V
A					50											43		55				38		
B										98				40	58									
C																			19		19			
D											31	39												70
E	50									48														
F														21			48			24				
Fp														19		44				29				
H																		30	33					
I											6	17						25						
J		98			48								106				89		50					
K				31					6								37							
Kp				39					17														55	
L										106														108
M		40				21	19																	
Mp		58																					58	
N	43						44																	
O						48				89	37													
P	55							30	25															
Q			19					33		50														
R						24	29														54			
S			19																	54				72
T	38																							34
U												55			58									
V				70									108								72	34		

B

	A	B	C	D	E	F	Fp	H	I	J	K	Kp	L	M	Mp	N	O	P	Q	R	S	Sp	T	U	V
A					25											20		32					13		
B										39				15	24										
C																			4		7				
D											13	19										22			
E	25									26															
F														9			15			6					
Fp														8		19				12					
H																		19	19						
I											5	9						14							
J		39			26								49				39		23						
K				13					5								23								
Kp				19					9															28	
L										49															47
M		15				9	8																		
Mp		24																						26	
N	20						19																		
O						15				29	23														
P	32							18	14																
Q			4					19		23															
R						6	12														19				
S			7																	19					27
Sp				32																					31
T	13																								12
U												28			26										
V													47								27	31	12		

C

	A	B	C	D	E	F	Fp	H	I	J	K	Kp	L	M	Mp	N	O	P	Q	R	S	T	U	V
A					59											23		16				86		
B										89				72	23									
C																			21		34			
D											7	23												30
E	59									15														
F														5			19			8				
Fp														4		18				16				
H																		13	11					
I											11	12						25						
J		89			15								73				62							
K									9								25							
Kp				25					12														22	
L										73														58
M		73				5	6																	
Mp		23																					28	
N	63						18																	
O						19				26	25													
P	25							13	25															
Q			21					11		61														
R						4	16														23			
S			34																	23				59
T	86																							17
U												22			28									
V				31									58								59	17		

D

	A	B	C	D	E	F	Fp	H	I	J	K	Kp	L	M	Mp	N	O	P	Q	R	S	Sp	T	U	V
A					3											9		5					5		
B										3					1										
C																					1				
D											1											1			
E	3																								
F																									
Fp																									
H																			3						
I												2						3							
J		3			1												3		4						
K				1									3				1								
Kp									2																
L										3															4
M																									
Mp		1																							
N	9																								
O										3															
P	5								3																
Q								3		4															
R																									
S			1																						4
Sp				1																					1
T	5																								
U																									
V													4								2	4	1		

Fig. S12 - Upstream and downstream sequence blocks in *T. gondii* ME49 and *N. caninum* Nanopore reads. The identity of blocks located immediately at the 5-prime and 3-prime end of each of the 21 sequences blocks and 3 partials were counted, and the counts are shown as a matrix. The number of times a block on the y-axis is present at the 5-prime or 3-prime end in reference to the block on the x-axis is indicated in blue or red respectively. If two blocks were not found next to each other that corresponding square is blank. Only upstream or downstream block occurrences appearing >3 times were counted to avoid anomalies. (A) Data derived from *T. gondii* ME49 Nanopore DNA reads Supplemental Fig. S4; Dataset S3. (B) *N. caninum* Nanopore DNA reads, Supplemental Fig. S10; Dataset S8 (C) *T. gondii* Nanopore RNA reads, Dataset S6 and (D) *N. caninum* Sanger EST reads. Please note the absence of a block relationships in (D) is likely due to missing data since the data set was sparse.

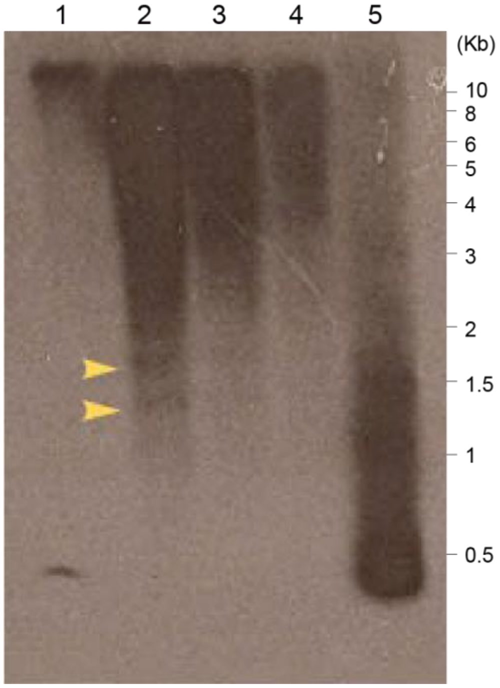


Fig. S13 - *T. gondii* mtDNA is variable in size and does not exist as a tandem repeat unit. Southern blot analysis. Total DNA from the *T. gondii* RH strain was hybridized to a 350 bp *coxI* probe under very stringent conditions. Lane 1: Undigested 2 µg genomic DNA; 2: *XhoI* Digested 15 µg genomic DNA; 3: *XhoI* Digested 5 µg genomic DNA; 4: *XhoI* Digested 2 µg genomic DNA; 5: positive control 350 bp probe fragment. *XhoI* cuts only once in the mitochondrial SBs, it cuts inside sequence block E of the *cob* gene. The yellow arrows point to distinct bands (~1.65 and 1.35 kb) on the blot. Size markers are indicated.

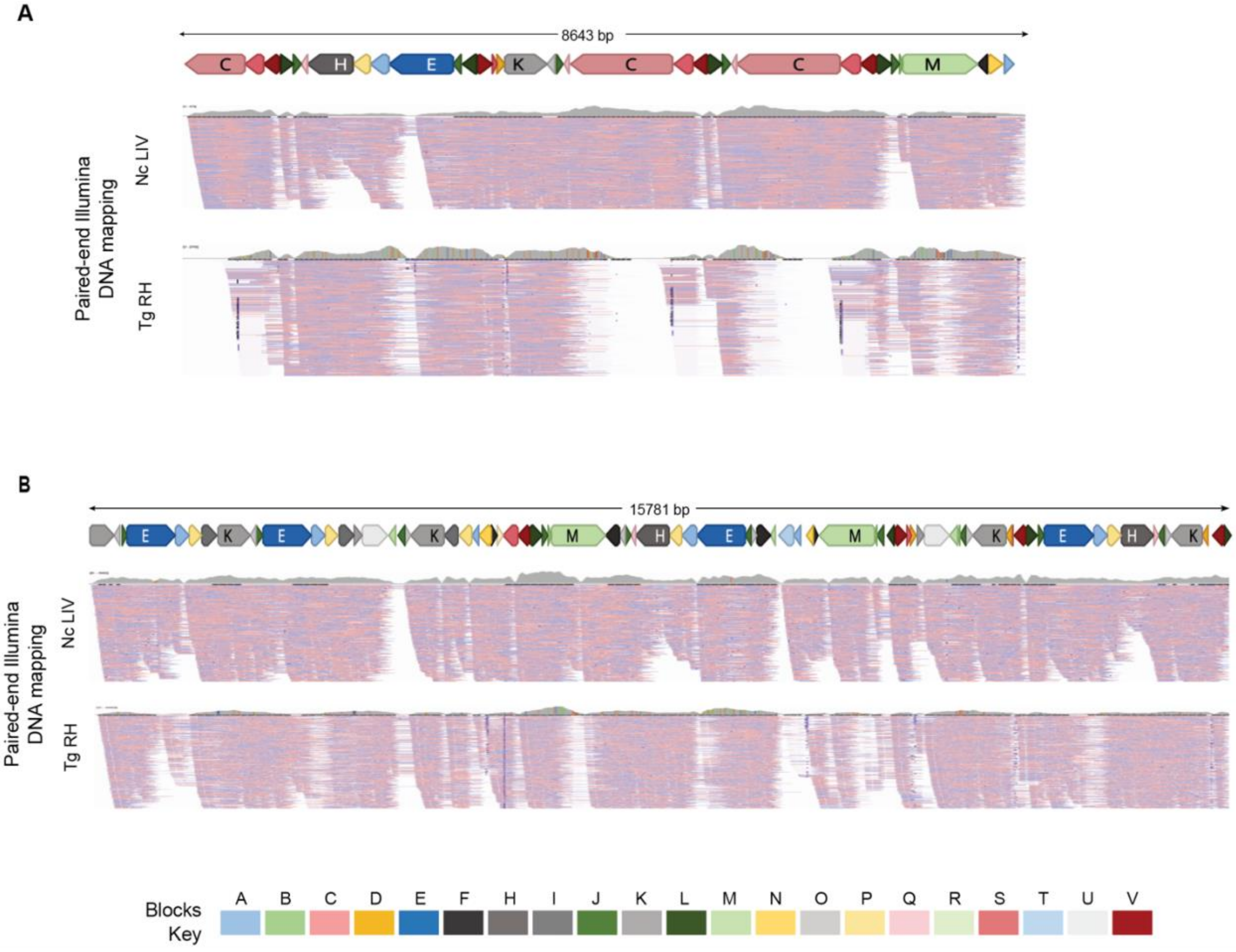


Fig. S14 - *N. caninum* nanopore sequence comparison and degeneration. (A-B) SBs were annotated on two *N. caninum* Nc-1 nanopore sequences reads at scale. The blocks are colored as shown in the key at the bottom of the figure. The two 'Paired-end Illumina DNA mapping' tracks show paired-end read mapping of *N. caninum* (Nc) LIV (ERR012900) and *T. gondii* (Tg) RH (SRR521957) mtDNA-specific reads. Mapping of Nc LIV reads required 100% nucleotide identity whereas 1% mismatch was allowed for mapping *T. gondii* RH reads. Reads were independently mapped to each of the Nanopore mtDNA reads and visualized using IGV. Red and blue lines below each read indicate the mapped Illumina paired-ends.

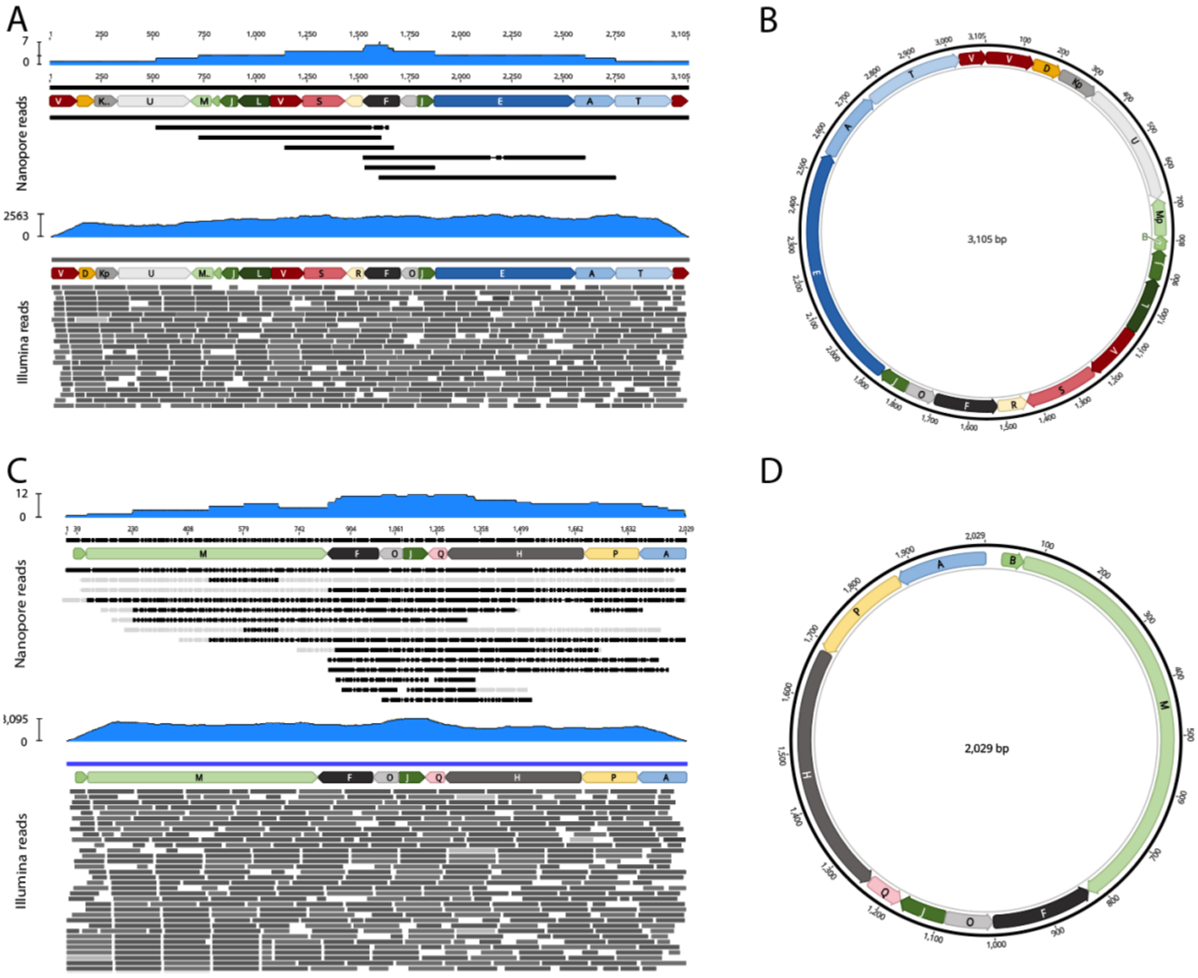


Fig. S15 - Nanopore and Illumina support for possible circular structures. (A, C) Mapped tracks of TgME49 Nanopore (top panels) and TgME49 Illumina reads (bottom panels) to single Nanopore reads identified as possibly existing as circular molecules according to Circlator predictions (B, D). Note the presence of the *cob* gene blocks EAT in panel B. Mapping of single and paired-end Illumina reads was permitted. Not all mapped reads are shown. The depth of read mapping is indicated above each panel in blue. The portions of Nanopore reads that are greyed out in panel C do not map to the annotated Nanopore read above. Additional predictions some of which encode the remaining cytochrome genes are provided in Supplemental Fig. S16.

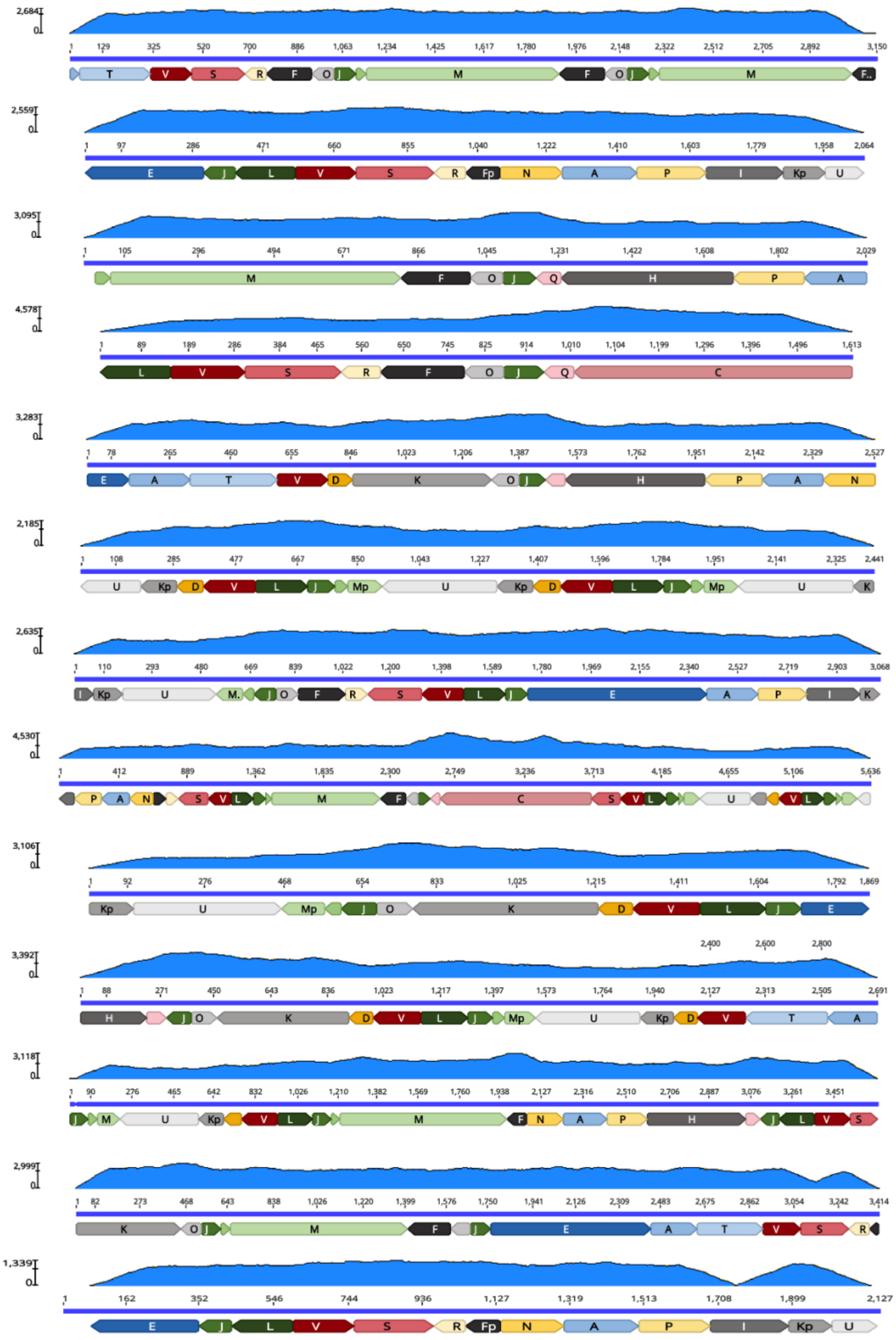


Fig. S16 - Nanopore reads predicted to circularize. *T. gondii* ME49 Illumina mapping frequency support for predicted circular sequences. Mapped read counts are displayed above each Nanopore read in blue. Some reads show fairly uniform coverage and others do not. Scale is as indicated for each read. Full-length cytochrome genes are observed.