Supplementary Material

Table S1. Primers used to link genomic contigs.

Name	Primer sequence (5′ – 3′)
B1	GAGGTGCCCTCTTCCTTTCT
B2	TGGTAACTTCCAAATCACCTCT
B3	TTTACGTTTGTTAGAGCGAAGTC
B4	GGGACTTGAACCGCAACTTC
B5	CGTTCATGCAATGAAAACAGAAGC
B6	CGGCTGATTACACCTCCGTG
P1	AGTCTCTCTAGCCGGTATTAGT
P2	TGCCAAGACGAATACCTGACT
P3	GTGACGTACGAAGTATGTTCCCT
P4	GTACTCCAGTGGCTCCTACG
P5	GGACATCATAGAAACATAATCTAACCTATC
P6	CATCAAATCTAAATCACTATACTCATATG

Table S2. GenBank accession numbers of protein sequences used to build hidden Markov models

Species	ccmC	ccmF	TatC	Rpl10	Rps4
Andalucia godoyi	YP_007890513.1	YP_007890514.1 AGH24008.1	YP_007890501.1	YP_007890474.1	YP_007890498.1
Histiona aroides	YP_007890585.1 AGH24079.1	YP_007890584.1 AGH24078.1	YP_007890573.1	YP_007890543.1	YP_007890570.1
'Jakoba bahamiensis'	YP_007890628.1	YP_007890629.1 AGH24123.1	YP_007890642.1	YP_007890649.1	YP_007890639.1 YP_007890639.1
Jakoba libera	YP_007890700.1	YP_007890699.1 AGH24193.1	YP_007890756.1	YP_007890732.1	YP_007890684.1
Naegleria fowleri	AOS85673.1 AOS85627.1 YP_007890053.1 AFP72326.1	AOS85658.1 AOS85612.1 YP_007890060.1 AFP72333.1	N/A	N/A	AOS85659.1 AOS85613.1 YP_007890062.1 AFP72335.1
Naegleria gruberi	AAG17810.1 NP_066532.1	AAG17817.1 NP_066539.1	AAG17818.1	N/A	AAG17819.1
Reclinomonas americana	NP_044790.1 AAD11905.1 AGH24367.1 AGH24434.1 AGH24300.1	NP_044789.1 AAD11904.1 AGH24366.1 AGH24433.1 AGH24299.1	AAD11893.1 NP_044778.1 AGH24355.1 AGH24422.1 AGH24288.1 O21266.1	NP_044748.1 AAD11863.1 AGH24325.1 AGH24392.1 AGH24258.1	NP_044775.1 AAD11890.1 O21263.1 AGH24352.1 AGH24419.1 AGH24285.1
'Seculamonas ecuadoriensis'	YP_007890768.1	YP_007890781.1 AGH24476.1	YP_007890761.1 AGH24456.1	YP_007890790.1	YP_007890767.1
Tsukubamonas globosa	N/A	N/A	N/A	N/A	YP_009004138.1

Table S3. Primers used to confirm RNA editing sites in amoeba BB2

Name	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)	Position on mtDNA
B1	GCCAGGAAGTGGTGGTATTCT	CCGCTATTCGACGCAAAAACA	30511-31294
Br5	AGTGCGACTAGTGCTGAACC	GATGCGATTGTTGAAGGCCC	33489-34977
Br7	CGGCTACCGTCTAAACGAGG	GCAACTTCTTTGCAGTTTCCTG	35051-35692
Br9	GTTCTCAAGCGTTGGACCGT	AAGTGAAAATGCAAAATCGAAGTCA	36580-37216
Br13	TCTTTATGTTCATTGCGCTATTTGT	ACGTTTTCCGTTCTTAAAAGTCCA	38703-40109
Br15	TGGACTTTTAAGAACGGAAAACG	TCAAACATTCATTACACACAGCTT	40086-41253
Br21	AGGTTTGCCAAAGACGGTGA	TTCCACATACGTGGGTTCGG	55272-55645
Br23	GTCCATGTAGGTACGCCGAA	AACGCCCTTCGCTATAAGCA	61595-62301

Table S4. Number of editing sites for each gene by nucleotide type.

Gene	Length (bp)	A	U	G	C	Total
atp1	2442	5	0	13	1	19
atp3	861	4	0	6	1	11
atp6	756	1	0	4	1	6
atp8	381	0	0	1	0	1
atp9	231	0	0	2	0	2
cob	1203	0	1	9	0	10
cox1	2016	1	1	23	2	27
cox11	411	0	0	1	0	1
cox2	744	0	0	7	1	8
cox3	1455	0	0	11	0	11
nad1	1281	0	0	11	1	12
nad11	2112	0	0	14	3	17
nad2	1533	0	3	8	0	11
nad3	465	0	0	3	0	3
nad4	1407	0	0	12	0	12
nad4L	345	0	0	3	0	3
nad5	1986	1	0	16	1	18
nad6	642	0	0	6	1	7
nad7	1203	0	0	12	2	14
nad8	483	1	0	4	1	6
nad9	669	2	0	2	0	4
rpl10	582	0	1	4	0	5
rpl11	516	1	0	2	0	3
rpl14	369	1	0	4	0	5
rpl16	396	0	0	3	0	3
rpl2	807	0	0	11	0	11
rpl32	177	0	0	0	0	0
rpl5	525	0	0	5	0	5
rpl6	687	0	0	7	0	7
rps10	648	0	0	2	0	2
rps11	2139	0	0	14	0	14
rps12	363	1	0	3	0	4
rps13	444	0	0	3	0	3
rps14	306	1	0	3	0	4
rps19	282	0	0	3	0	3
rps2	993	0	0	5	0	5
rps3	3030	1	0	13	1	15
rps7	930	0	0	7	0	7
rps8	396	0	1	2	0	3
sdh2	768	0	0	6	1	7

Gene	Length (bp)	A	U	G	C	Total
orf164	495	0	0	4	0	4
orf242	729	1	1	3	1	6
orf3630	10893	9	5	71	3	88
orf925	2778	2	1	12	0	15
rnl	2840	0	0	20	2	22
rns	1654	0	1	15	2	18
trnA(ugc)	73	0	0	0	0	0
trnC(gca)	71	0	0	1	0	1
trnD(guc)	74	0	0	1	0	1
trnE(uuc)	73	0	0	0	0	0
trnF(gaa)_1	72	0	0	1	0	1
trnF(gaa)_2	75	0	0	0	0	0
trnG(ucc)	71	0	0	1	0	1
trnH(gug)	74	0	0	0	0	0
trnI(gau)	73	0	0	1	0	1
trnK(uuu)	72	0	0	1	0	1
trnL(caa)	82	0	0	0	0	0
trnL(gag)	74	0	0	0	0	0
trnL(uaa)	84	0	0	1	0	1
trnL(uag)	74	0	0	0	0	0
trnM(cau)_1	72	0	1	0	0	1
trnM(cau)_2	73	0	0	0	0	0
trnN(guu)	73	0	0	0	0	0
trnP(ugg)	74	0	0	1	0	1
trnQ(uug)	72	0	0	0	0	0
trnR(ucu)	74	0	0	1	0	1
trnS(gcu)	85	1	0	0	0	1
trnS(gga)	85	0	0	0	0	0
trnS(uga)	82	0	0	0	0	0
trnW(cca)	74	0	0	0	0	0
trnY(gua)	83	0	0	1	0	1
intergenic region	83	1	0	0	0	1
Total		**34**	**16**	**400**	**25**	**475**

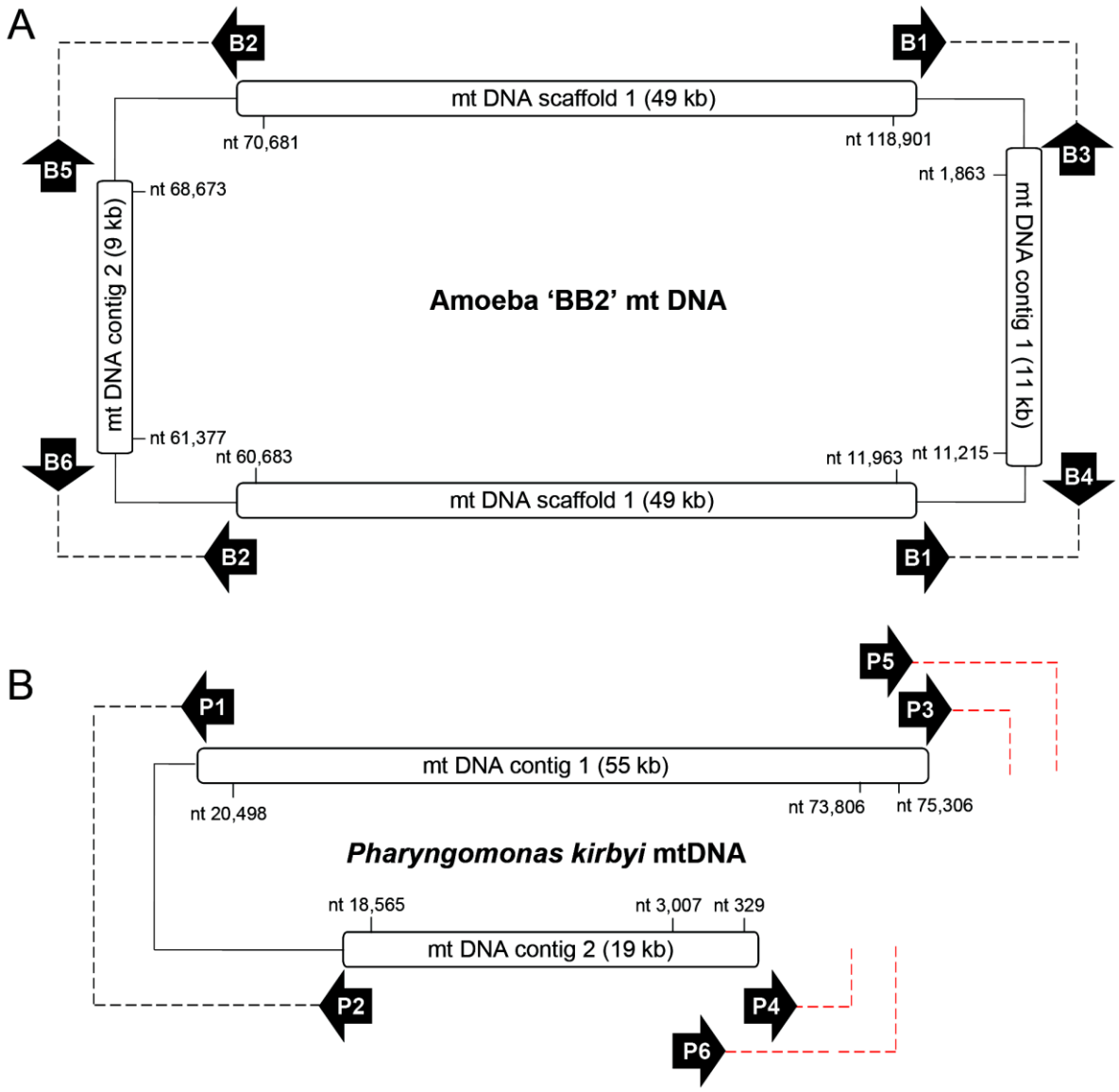

Figure S1. Schematic maps demonstrating how gaps in (A) amoeba 'BB2' and (B) *P. kirbyi* mitochondrial genomes were closed after sequencing of long-range PCR amplicons amplified with primers (black arrows, nucleotide numbers indicate the positions of binding sites) listed in Table S1. The set of PCR reactions that did not yield products is highlighted in red dashed lines.

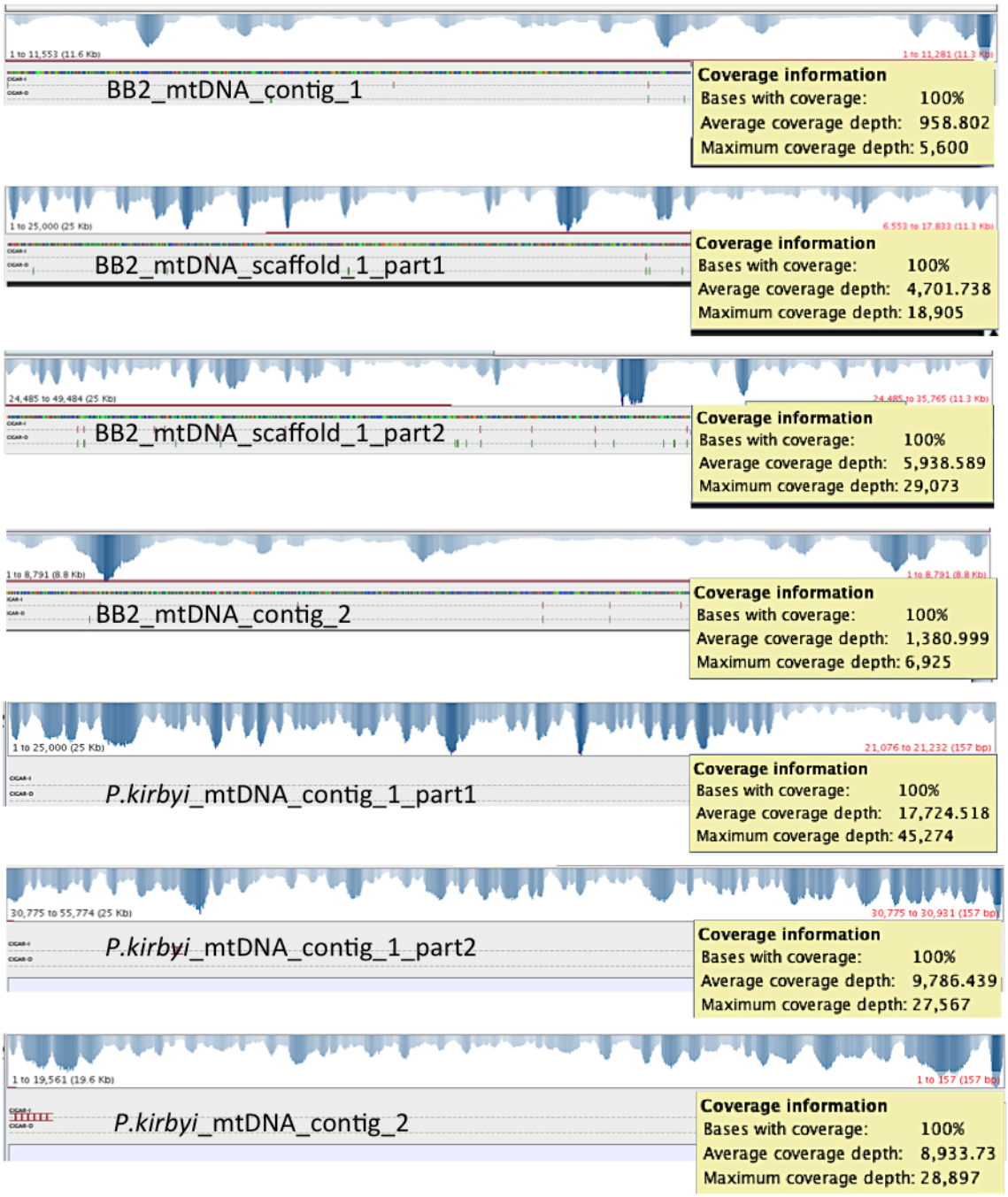

Figure S2. Assembly coverage plots for the mitochondrial genome of BB2 and *P. kirbyi* by contigs (shown in Figure S1).

	S10												Spc
Rickettsia	*rps10*	– – –	*rpl3*	*rpl4*	*rpl23*	*rpl2*	*rps19*	*rpl22*	– – –	*rps3*	*rpl16*	*rpl29*	*rps17*	*rpl14*
Andalucia	*rps10*	– – –	– – –	– – –	– – –	*rpl2*	*rps19*	– – –	– – –	*rps3*	*rpl16*	– – –	– – –	*rpl14*
Tsukubamonas	– – –	– – –	– – –	– – –	– – –	*rpl2*	*rps19*	– – –	– – –	*rps3*	*rpl16*	– – –	– – –	*rpl14*
Amoeba BB2	*rps10*	– – –	– – –	– – –	– – –	*rpl2*	*rps19*	– – –	*urf219*	*rps3*	*rpl16*	– – –	– – –	*rpl14*
Pharyngomonas	*rps10*	– – –	– – –	– – –	– – –	*rpl2*	*rps19*	– – –	– – –	*rps3*	*rpl16*	– – –	– – –	*rpl14*
Naegleria	*rps10*	*rpl11*	– – –	– – –	– – –	*rpl2*	*rps19*	– – –	– – –	*rps3*	*rpl16*	– – –	– – –	*rpl14*

	Spc										Alpha			
Rickettsia	*rpl24*	*rpl5*	*rps14*	*rps8*	*rpl6*	*rpl18*	*rps5*	*rpl30*	*rpl15*	*secY*	*adk*	*rps13*	*rps11*	*rpoA*
Andalucia	– – –	*rpl5*	*rps14*	*rps8*	*rpl6*	*rpl18*	– – –	– – –	– – –	– – –	– – –	*rps13*	*rps11*	*rpoA*
Tsukubamonas	– – –	*rpl5*	*rps14*	*rps8*	*rpl6*	– – –	– – –	– – –	– – –	– – –	– – –	*rps13*	*rps11*	– – –
Amoeba BB2	– – –	*rpl5*	*rps14*	*rps8*	*rpl6*	– – –	– – –	– – –	– – –	– – –	– – –	*rps13*	*rps11*	– – –
Pharyngomonas	– – –	*rpl5*	*rps14*	*rps8*	*rpl6*	– – –	– – –	– – –	– – –	– – –	– – –	*rps13*	*rps11*	– – –
Naegleria	– – –	*rpl5*	*rps14*	*rps8*	*rpl6*	– – –	– – –	– – –	– – –	– – –	– – –	*rps11*	*rps13*	– – –

Figure S3. Gene order comparison of Discoba mtDNA and α-proteobacteria *Rickettsia prowazekii* in the three contiguous ribosomal protein operons (S10-Spc-Alpha). Genes in reversed order are highlighted in green.

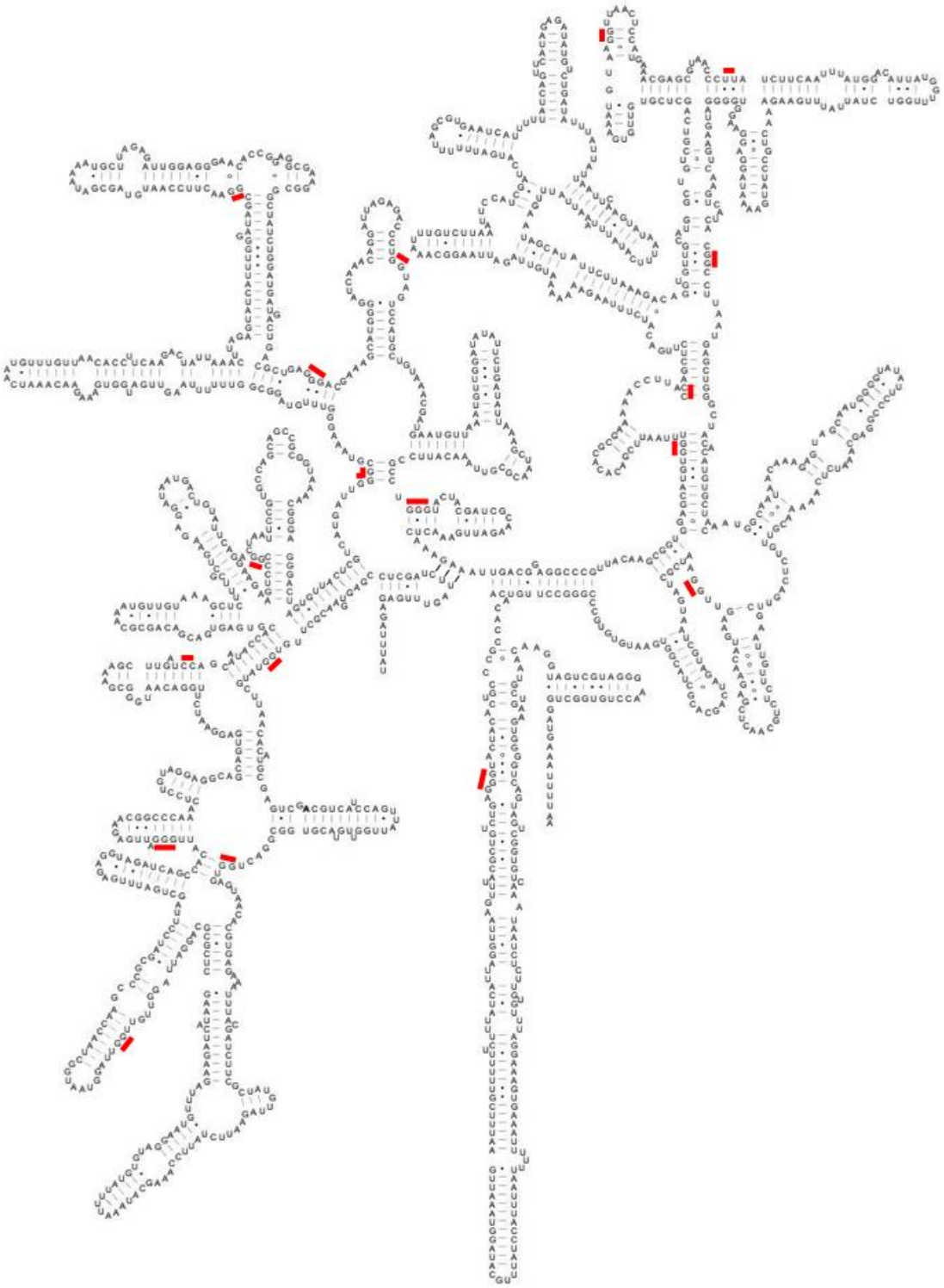

Figure S4. Predicted secondary structure of BB2 mitochondrial SSU rRNA. Regions including an editing site are indicated with red lines. For each highlighted region, only one nucleotide is inserted (the exact insertion sites are unknown since nucleotides are inserted next to one or more encoded nucleotides with the same identity).

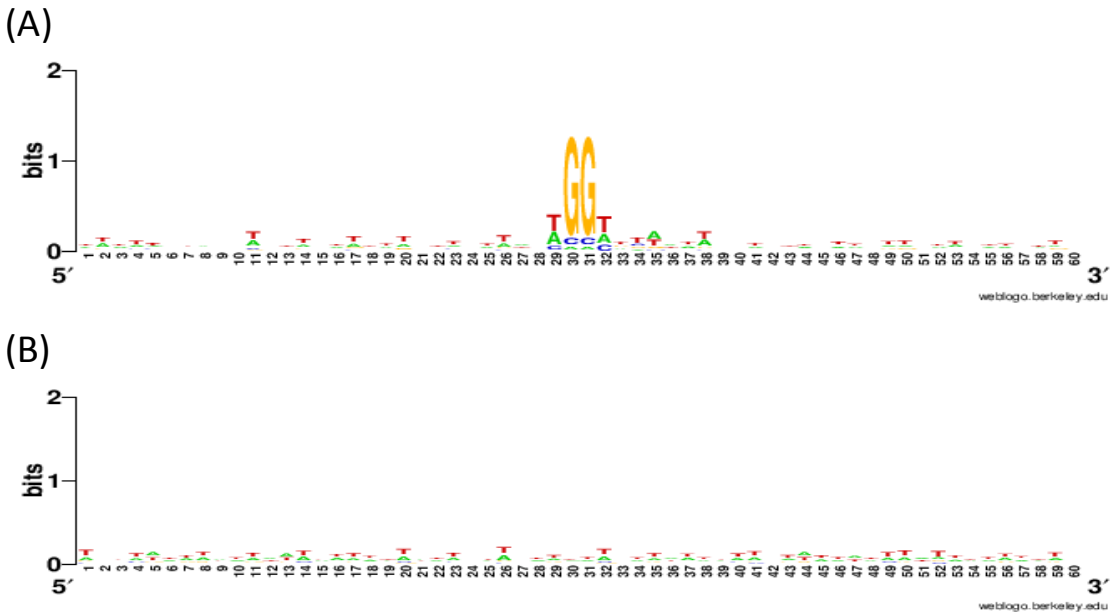

Figure S5. Sequence LOGO of (A) 311 60-nucleotide-long sequences near editing sites and (B) 328 random 60-nt-long sequences.