Supplementary Information

for

Evolutionary history of mitochondrial genomes in Discoba, including the extreme halophile *Pleurostomum flabellatum* (Heterolobosea)

Khaoula Ettahi\(^a\), Duck Hyun Lhee\(^a\), Ji Yeon Sung\(^b\), Alastair G. B. Simpson\(^{c,d}\), Jong Soo Park\(^{b,e,*}\),

Hwan Su Yoon\(^{a,*}\)

\(^a\)Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, South Korea
\(^b\)Department of Oceanography, Kyungpook Institute of Oceanography, School of Earth System Sciences, Kyungpook National University, Daegu 41566, South Korea
\(^c\)Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada
\(^d\)Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada
\(^e\)Research Institute for Dok-do and Ulleung-do Island, Kyungpook National University, Daegu 41566, South Korea

\(^*\)Author for Correspondence: Hwan Su Yoon, Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, South Korea, Telephone number: +82-31-290-5915; FAX: +82-31-290-7015; e-mail: hsyoon2011@skku.edu,

Jong Soo Park, Department of Oceanography, School of Earth System Sciences, Kyungpook National University, Daegu 41566, South Korea, Telephone number: +82-53-950-5391; FAX: +82-53-950-5397; e-mail: jongsoopark@knu.ac.kr
Supplementary Information Legends

**Supplementary Table S1.** General Features and Comparison of the Heteroloboseans, Jakobids and *Tsukubamonas* Mitogenomes.

**Supplementary Table S2.** Comparison of Protein-Coding Gene Sets in Heterolobosea and *Tsukubamonas*.

**Supplementary Table S3.** Comparison of Structural RNAs Gene Sets in Heterolobosea and *Tsukubamonas*.

**Supplementary Table S4.** Relative synonymous codon usage (RSCU) of each amino acid in the mitogenome of *Pleurostomum flabellatum, Naegleria gruberi/Naegleria fowleri, Pharyngomonas kirbyi, Acrasis kona, Stachyamoeba lipophora, Heteroloboseid sp. BB2 and Tsukubamonas globosa*. More frequently used codons (RSCU>1) are in bold.

**Supplementary Fig. S1. Comparison of the three available sequences of orf145.** (A) The position of the conserved orf145 in the three heterolobosean species: *Pleurostomum flabellatum, Naegleria gruberi* and *Naegleria fowleri*. The step size is 200 bp in *P. flabellatum* and 100 bp in *Naegleria*. (B) Sequence logo plots of aligned orf145 amino acids generated using the Geneious 8.1.2 multiple sequence alignment tool.

**Supplementary Fig. S2. Relative synonymous codon usage (RSCU) in heterolobosean mitogenomes.** The 22 codon families consisting of a total of 61 two- and four-fold degenerate synonymous codons are plotted on the x-axis. The RSCU values are shown on the y-axis. The most used synonymous codon in each family is at the bottom. Red-colored codons are not present in the mitogenome.
Supplementary Fig. S3. Distribution of the synonymous codons on the first and second axes of the correspondence analysis. Representation of the first two axes of the correspondence analysis performed on the codon frequency of heterolobosean mitochondrial-encoded proteins.

Supplementary Fig. S4. Bayesian phylogenetic tree of the Discoba based on mitogenome data under the CAT + GTR model. Contrary to ML tree, Heteroloboseid sp. BB2 and Pharyngomonas are paraphyletic in the Bayesian tree inference.

Supplementary Fig. S5. Mitochondrial genome synteny of Pleurostomum flabellatum and Pharyngomonas kirbyi. Images were generated using the Mauve genome alignment tool with default settings (Darling et al. 2010). Color-coded syntenic blocks indicate conserved segments. Sequence similarity is shown within each syntenic block. Regions with no color indicate no detectable homology between the two mitogenomes.

Supplementary Fig. S6. Maximum likelihood phylogenetic tree of ccmF protein sequences. The ccmF gene was found in Pleurostomum, Naegleria and Stachyamoeba, previously misannotated by orf457 (YP_009118159.1). The ccmF gene is encoded in a wide array of other eukaryotic lineages, namely, Viridiplantae, Rhodophyta, Alveolata, Malawimonadidae, Metazoa, Cryptophyta and Jakobida. Our ccmF phylogenetic analysis places the P. flabellatum sequence as more closely related to Rhodophyta and Malawimonadidae mitochondrial ccmFs than to either Tetramitia (St. lipophora and Naegleria) or Jakobida ccmFs, probably owing to the rapid divergence of Heterolobosean species.

Supplementary Fig. S7. Maximum likelihood phylogenetic tree of rps4 protein sequences. The gene rps4 gene was found uniquely in Pleurostomum and Naegleria among Heterolobosea. The gene rps4 is encoded in a wide array of other eukaryotic lineages, namely, Viridiplantae, Rhodophyta, Alveolata, Malawimonadidae, Metazoa, Cryptophyta and Jakobida. Reconstruction from rps4
clustered *P. flabellatum* and *Naegleria rps4s* firmly in one clade (BS=96%). This clade was more closely connected to mitochondrial *rps4s* of Viridiplantae than to either Jakobida or *Tsukubamonas rps4s*.

**Supplementary Fig. S8. Maximum likelihood phylogenetic tree of ccmC protein sequences.**

Reconstruction from *ccmC* of *Naegleria* genus, are well resolved as a monophyletic cluster with mitochondrial *ccmC* of Rhodophyta (BS=95%) and distantly clustering with mitochondrial *ccmC* of Jakobida and Viridiplantae (BS=83%).

**Supplementary Fig. S9. Maximum likelihood phylogenetic tree of tatC protein sequences.** The phylogenetic analysis of the TAT machinery inducer in most of TAT-containing eukaryotes *tatC* (Petrů et al. 2018) strongly clustered *Naegleria* into one clade with Rhodophyta mitochondrial *tatC* (BS=90%) but somewhat distantly clustered with Jakobida *tatC*.

**Supplementary Fig. S10. Maximum likelihood phylogenetic tree of ccmA protein sequences.** The *ccmA* gene of *Stachyamoeba lipophora*, is resolved as sister-taxa to *ccmA* of several Proteobacteria with low bootstrap value (below 50%) and distantly related to mitochondrial *ccmA* of Jakobida align with other eukaryotes (mostly Viridiplantae) (BS=89%).

**Supplementary Figure S11. Multiple sequence alignment of Mic60 homologs.** (A) Mic60 sequences of *Naegleria gruberi* (Naegru_XP_002683319.1), and two putative Mic60 homologs from the *Pleurostomum flabellatum* genome assembly (Pleulab_MW019460 and Pleulab_MW019459) were aligned. The alignment of translated transcript sequences of the coding regions (Pleulab_transcr_seq01 and Pleulab_transcr_seq02) was also added. The mitofilin domains are marked in black frames. (B) 5’ and 3’ UTR regions of transcript sequences are highlighted.
Table S1. General features and comparison of the Heteroloboseans, Jakobids and *Tsukubamonas* mitogenomes.

| Species                        | Size (bp) | Coding (%) | A%  | C%  | G%  | T%  | A+T % | G+C % | AT Skew | GC Skew |
|--------------------------------|-----------|------------|-----|-----|-----|-----|-------|-------|--------|---------|
| **Heterolobosea**              |           |            |     |     |     |     |       |       |        |         |
| *Pleurostomum flabellatum*     | 57,829    | 86.1       | 32.2| 11.4| 17.6| 38.9| 71.1  | 29    | -0.094 | 0.214   |
| *Naegleria gruberi*            | 49,843    | 93.1       | 35.5| 9.2 | 13  | 42.2| 77.7  | 22.2  | -0.086 | 0.171   |
| *N. fowleri*                   | 49,531    | 91.6       | 34.4| 10.1| 15.1| 40.3| 74.7  | 25.2  | -0.079 | 0.198   |
| *Acrasis kona*                 | 51,458    | 93.2       | 45.9| 5.4 | 11.3| 37.4| 83.3  | 16.7  | 0.102  | 0.353   |
| *Stachymoeba lipophora*        | 49,742    | 91.2       | 31.7| 11.3| 15.9| 41.1| 72.8  | 27.2  | -0.129 | 0.169   |
| *Pharyngomonas kirbyi*         | 75,717    | 92.0       | 35.2| 15.1| 14.7| 35.1| 70.3  | 29.8  | 0.001  | -0.013  |
| Heteroloboseid sp. BB2         | 119,312   | 81.0       | 34.3| 15.4| 15.2| 35  | 69.3  | 30.6  | -0.010 | -0.007  |
| **Tsukubamonadida**            |           |            |     |     |     |     |       |       |        |         |
| *Tsukubamonas globosa*         | 48,463    | 90.2       | 28.6| 15.2| 18.6| 37.6| 66.2  | 33.8  | -0.136 | 0.101   |
| **Jakobida**                   |           |            |     |     |     |     |       |       |        |         |
| *Jakoba bahamiensis*           | 65,327    | 93.0       | 30.4| 13.9| 18.3| 37.4| 67.8  | 32.2  | -0.103 | 0.137   |
| *J. libera*                    | 100,252   | 78.5       | 35.4| 16.3| 15.7| 32.6| 68.0  | 32.0  | 0.041  | -0.019  |
| *Reclimonas americana*         | 69,586    | 90.7       | 36.8| 11.4| 14.8| 37.1| 73.2  | 26.1  | -0.004 | 0.130   |
| *Histiona aroides*             | 70,176    | 90.8       | 32.7| 16.9| 18.5| 31.9| 64.6  | 35.4  | 0.012  | 0.045   |
| *Seculamonas ecuadoriensis*    | 69,158    | 88.4       | 34.7| 16.1| 15.8| 33.5| 68.1  | 31.9  | 0.018  | -0.009  |
| *Andalucia godoyi*             | 67,656    | 89.8       | 31.5| 17.9| 18.4| 32.2| 67.3  | 36.3  | -0.011 | 0.014   |
| *Ophirina amphinema*           | 59,094    | 94.1       | 27.6| 16.4| 22.7| 33.3| 60.9  | 39.1  | -0.094 | 0.161   |

Lin*Linear mitogenome.
Table S2. Comparison of protein-coding gene sets in Heterolobosea and Tsukubamonas.

| Biological Process                  | Gene | P. flabellatum | N. gruberi/fowleri | A. kona | S. lipophora | P. kirbyi | Heter. sp. BB2 | T. globosa |
|-------------------------------------|------|----------------|-------------------|--------|--------------|----------|---------------|------------|
| NADH dehydrogenase (CI)             | nad1 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad2 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad3 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad4 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad4L| ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad5 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad6 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad7 | ●              | ●                 | ○      | ●            | ●        | ●             | ●          |
|                                     | nad8 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | nad9 | ●              | ●                 | ○      | ●            | ●        | ●             | ●          |
|                                     | nad10| ○              | ○                 | ○      | ○            | ○        | ○             | ○          |
|                                     | nad11| ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
| Succinate dehydrogenase (CII)       | sdh2 | ●              | ●                 | ○      | ●            | ●        | ○             | ●          |
|                                     | sdh3 | ○              | ○                 | ○      | ○            | ○        | ○             | ●          |
|                                     | sdh4 | ○              | ○                 | ○      | ○            | ○        | ○             | ●          |
| Cytochrome bc complex (CIII)        | cytB | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
| Cytochrome c oxidase (CIV)          | cox1 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | cox2 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | cox3 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
| ATP synthase (CV)                   | atp1 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | atp3 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | atp4 | ○              | ○                 | ○      | ○            | ○        | ○             | ●          |
|                                     | atp6 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | atp8 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | atp9 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
| SSU ribosomal                       | rps1 | ○              | ○                 | ○      | ○            | ○        | ○             | ○          |
|                                     | rps2 | ●              | ●                 | ○      | ●            | ●        | ●             | ●          |
|                                     | rps3 | ●              | ●                 | ○      | ●            | ●        | ●             | ●          |
|                                     | rps4 | ●              | ●                 | ○      | ●            | ●        | ●             | ●          |
|                                     | rps7 | ●              | ●                 | ●      | ○            | ●        | ●             | ●          |
|                                     | rps8 | ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | rps10| ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | rps11| ●              | ●                 | ●      | ●            | ●        | ●             | ●          |
|                                     | rps12| ●              | ●                 | ○      | ●            | ●        | ●             | ●          |
|                   | rps13 | rps14 | rps19 |
|-------------------|-------|-------|-------|
| LSU ribosomal     |       |       |       |
| rpl1              | ●     |       |       |
| rpl2              | ●     | ●     |       |
| rpl5              | ●     | ●     | ●     |
| rpl6              | ●     | ●     | ●     |
| rpl10             |       | ○     | ●     |
| rpl11             | ●     | ●     | ●     |
| rpl14             | ●     | ●     | ●     |
| rpl16             | ●     | ●     |       |
| rpl18             | ○     | ○     | ○     |
| rpl19             | ○     | ○     | ○     |
| rpl20             | ○     | ○     | ○     |
| rpl27             | ○     | ○     | ○     |
| rpl31             | ○     | ○     | ○     |
| rpl32             | ○     | ○     | ●     |
| rpl34             | ○     | ○     | ○     |
| rpl35             | ○     | ○     | ○     |
| Elongation factor |       |       |       |
| tufA              | ○     | ○     | ○     |
| Core RNA polymerase|      |       |       |
| rpoA              | ○     | ○     | ○     |
| rpoB              | ○     | ○     | ○     |
| rpoC              | ○     | ○     | ○     |
| Sigma-like factor |       |       |       |
| rpoD              | ○     | ○     | ○     |
| ABC transporter   |       |       |       |
| ccmA              | ○     | ○     | ●     |
| ccmB              | ○     | ○     | ○     |
| Heme delivery     |       |       |       |
| ccmC              | ○     | ●     | ○     |
| SecY-independent transporters |     |       |       |
| tatA              | ○     | ○     | ○     |
| tatC              | ○     | ●     | ○     |
| SecY-type transporter |     |       |       |
| secY              | ○     | ○     | ○     |
| Cytochrome c oxidase assembly |      |       |       |
| cox11             | ●     | ●     | ○     |
| cox15             | ○     | ○     | ○     |
| Heme c maturation |       |       |       |
| ccmF              | ●     | ●     | ○     |
Table S3. Comparison of structural RNAs gene sets in Heterolobosea and *Tsukubamonas.*

| Gene                | P. flabellatum | N. gruberi | N. fowleri | Ac. kona | St. lipophora | Ph. kirbyi | Heter. sp. BB2 | T. globosa |
|---------------------|----------------|------------|------------|----------|---------------|------------|----------------|-----------|
| rnl                 | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| rns                 | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| rnr5                | ○              | ○          | ○          | ○        | ○             | ○          | ●              | ●         |
| rnpB                | ○              | ○          | ○          | ○        | ○             | ○          | ○              | ○         |
| ssrA                | ○              | ○          | ○          | ○        | ○             | ○          | ○              | ○         |
| trnA(ugc)           | ○              | ○          | ○          | ○        | ○             | ●          | ●              | ●         |
| trnC(gca)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnD(guc)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnE(uuc)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnF(gaa)           | ●              | ●          | ●          | ○        | ●             | ●          | ●              | ●         |
| trnG(gcc)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnG(ucc)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnH(gug)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnI(cau)           | ○              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnI(gau)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnK(uuu)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnL(caa)           | ●              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnL(gag)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnL(uaa)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnL(uag)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnM(cau)e          | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnM(cau)f          | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnN(gau)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnP(ugg)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnQ(uug)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnR(acg)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnR(ucc)           | ○              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnS(gca)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnS(gga)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnS(uga)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnT(ugu)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnV(gac)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnV(uac)           | ○              | ○          | ○          | ○        | ●             | ●          | ●              | ●         |
| trnW(cca)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |
| trnY(gua)           | ●              | ●          | ●          | ●        | ●             | ●          | ●              | ●         |

²Two copies of the specific gene.
Table S4. Relative synonymous codon usage (RSCU) of each amino acid in the mitogenome of *Pleurostomum flabellatum*, *Naegleria gruberi/Naegleria fowleri*, *Pharyngomonas kirbyi*, *Acrasis kona*, *Stachyamoeba lipophore*, *Heteroloboseid* sp. BB2 and *Tsukubamonas globosa*. More frequently used codons (RSCU > 1) are in bold.

| Amino Acid | Codon | Count | RSCU | % | Amino Acid | Codon | Count | RSCU | % |
|-----------|-------|-------|------|---|-----------|-------|-------|------|---|
| Ala       | GCU(A) | 184   | 2.07 | 1.29 | Lys       | AAG(K) | 263   | 0.52 | 1.85 |
| Ala       | GCC(A) | 39    | 0.44 | 0.27 | Lys       | AAA(K) | 752   | 1.48 | 5.28 |
| Ala       | GCA(A) | 108   | 1.21 | 0.76 | Met       | AUG(M) | 365   | 1    | 2.57 |
| Arg       | CGU(R) | 131   | 1.17 | 0.92 | Phe       | UUC(F) | 1453  | 1.79 | 10.21 |
| Arg       | CGG(R) | 33    | 0.3  | 0.23 | Pro       | CCU(P) | 174   | 0.21 | 1.22 |
| Arg       | CGC(R) | 28    | 0.25 | 0.20 | Pro       | CG(C(P) | 57    | 0.6  | 0.40 |
| Arg       | CGA(R) | 79    | 0.71 | 0.56 | Pro       | CCC(P) | 22    | 0.23 | 0.15 |
| Asn       | AAU(N) | 674   | 1.59 | 4.74 | Ser1      | AGA(S1) | 298   | 2.67 | 2.09 |
| Asn       | AAC(N) | 175   | 0.41 | 1.23 | Ser1      | ACG(S1) | 68    | 0.36 | 0.48 |
| Asp       | GAU(D) | 433   | 1.65 | 3.04 | Ser1      | AGU(S1) | 254   | 1.35 | 1.79 |
| Asp       | GAC(D) | 91    | 0.35 | 0.64 | Ser1      | AGG(S1) | 100   | 0.9  | 0.70 |
| Cys       | UGU(C) | 195   | 1.63 | 1.37 | Ser1      | AG(S1) | 129   | 1.37 | 0.91 |
| Cys       | UGC(C) | 44    | 0.37 | 0.31 | Ser1      | AGA(S1) | 298   | 2.67 | 2.09 |
| Cys       | UAA(N) | 283   | 1.53 | 1.99 | Ser2      | UCG(S2) | 97    | 0.51 | 0.68 |
| Gln       | CAA(Q) | 87    | 0.47 | 0.61 | Ser2      | UCC(S2) | 73    | 0.39 | 0.51 |
| Glu       | GAG(E) | 153   | 0.61 | 1.08 | Ser2      | UCA(S2) | 254   | 1.35 | 1.79 |
| Glu       | GAA(E) | 351   | 1.39 | 2.47 | Ser2      | ACU(T) | 200   | 1.6  | 1.41 |
| Gly       | GGU(G) | 201   | 1.18 | 1.41 | Ter       | UAG(*) | 11    | 0.58 | 0.08 |
| Gly       | GGG(G) | 155   | 0.91 | 1.09 | Ter       | UAA(*) | 33    | 1.74 | 0.23 |
| Gly       | GGC(G) | 42    | 0.25 | 0.30 | Thr       | ACC(T) | 43    | 0.34 | 0.30 |
| Gly       | GGA(G) | 281   | 1.66 | 1.97 | Thr       | ACG(T) | 79    | 0.63 | 0.56 |
| His       | CAU(H) | 232   | 1.6  | 1.63 | Thr       | ACC(T) | 43    | 0.34 | 0.30 |
| His       | CAC(H) | 58    | 0.4  | 0.41 | Thr       | ACA(T) | 179   | 1.43 | 1.26 |
| Ile       | AUU(I) | 664   | 1.6  | 4.67 | Trp       | UGG(W) | 165   | 1    | 1.16 |
| Ile       | AUC(I) | 102   | 0.25 | 0.72 | Trp       | UGA(W) | 13    | 0.68 | 0.09 |
| Leu1      | CU(L)  | 186   | 0.71 | 1.31 | Tyr       | UAU(Y) | 684   | 1.52 | 4.81 |
| Leu1      | CUG(L) | 64    | 0.24 | 0.45 | Tyr       | UAC(Y) | 217   | 0.48 | 1.53 |
| Leu1      | CUC(L) | 25    | 0.1  | 0.18 | Val       | GUU(V) | 343   | 1.72 | 2.41 |
| Leu1      | CUA(L) | 110   | 0.42 | 0.77 | Val       | GUG(V) | 128   | 0.64 | 0.90 |
| Leu2      | UUG(L) | 476   | 1.82 | 3.35 | Val       | GUC(V) | 43    | 0.22 | 0.30 |
| Leu2      | UUA(L) | 709   | 2.71 | 4.98 | Val       | GUA(V) | 282   | 1.42 | 1.98 |
| Total     |       | 14229 | 64   | 100  |            |        |       |      |    |
### Naegleria gruberi/ Naegleria fowleri

| AA    | Codon | Count | RSCU  | %    | AA    | Codon | Count | RSCU | %    |
|-------|-------|-------|-------|------|-------|-------|-------|------|------|
| Ala   | GCU(A) | 206/152 | 2.03/1.55 | 1.52/1.15 | Lys   | AAA(K) | 987/791 | 1.86/1.64 | 7.29/6.00 |
| Ala   | GCA(A) | 193/193 | 1.9/1.97 | 1.43/1.46 | Lys   | AAG(K) | 72/791  | 0.14/0.36 | 0.53/1.32 |
| Ala   | GCG(A) | 5/32   | 0.05/0.33 | 0.04/0.24 | Met   | AUG(M) | 265/297 | 1/1   | 1.96/2.25 |
| Ala   | GCC(A) | 2/14   | 0.02/0.14 | 0.01/0.11 | Met   | AUA(M) | 381/595 | 0.86/1.33 | 2.81/4.51 |
| Arg   | CGU(R) | 139/104 | 1.99/1.31 | 1.03/0.79 | Phe   | UUU(F) | 1475/1265 | 1.87/1.78 | 10.90/9.59 |
| Arg   | CGA(R) | 11/22  | 0.16/0.28 | 0.08/0.17 | Phe   | UUC(F) | 99/154  | 0.13/0.22 | 0.73/1.17 |
| Arg   | CGC(R) | 3/7    | 0.04/0.09 | 0.02/0.05 | Pro   | CCC(P) | 178/143 | 2.12/1.79 | 1.32/1.08 |
| Arg   | CGG(R) | 1/9    | 0.01/0.11 | 0.01/0.07 | Pro   | CCG(P) | 145/137 | 1.73/1.71 | 1.07/1.04 |
| Asn   | AAU(N) | 873/771 | 1.81/1.78 | 6.45/5.85 | Pro   | CCG(P) | 10/29   | 0.12/0.36 | 0.07/0.22 |
| Asn   | AAC(N) | 89/94  | 0.19/0.22 | 0.66/0.71 | Pro   | CCC(P) | 3/11    | 0.04/0.14 | 0.02/0.08 |
| Asp   | GUA(D) | 378/324 | 1.84/1.73 | 2.79/2.46 | Ser1  | AGA(S1) | 258/261 | 3.69/3.29 | 1.91/1.98 |
| Asp   | GAC(D) | 32/51  | 0.16/0.27 | 0.24/0.39 | Ser1  | AGU(S1) | 309/317 | 1.82/1.91 | 2.28/2.40 |
| Cys   | UGU(C) | 151/178 | 1.83/1.78 | 1.12/1.35 | Ser1  | AGC(S1) | 30/51   | 0.18/0.31 | 0.22/0.39 |
| Cys   | UGC(C) | 14/22  | 0.17/0.22 | 0.10/0.17 | Ser1  | AGG(S1) | 7/73    | 0.1/0.92  | 0.05/0.55 |
| Gln   | CAA(Q) | 245/183 | 1.83/1.63 | 1.81/1.39 | Ser2  | UCU(S2) | 406/327 | 2.39/1.97 | 3.00/2.48 |
| Gln   | CAG(Q) | 23/41  | 0.17/0.37 | 0.17/0.31 | Ser2  | UCA(S2) | 233/178 | 1.37/1.07 | 1.72/1.35 |
| Glu   | GAA(E) | 385/355 | 1.79/1.61 | 2.84/2.69 | Ser2  | UCG(S2) | 28/84   | 0.16/0.51 | 0.21/0.64 |
| Glu   | GAG(E) | 46/86  | 0.21/0.39 | 0.34/0.65 | Ser2  | UCC(S2) | 13/38   | 0.08/0.23 | 0.10/0.29 |
| Gly   | GGU(G) | 278/265 | 2.09/1.97 | 2.05/2.01 | Thr   | UAA(*)  | 37/26   | 2.41/1.7  | 0.27/0.20 |
| Gly   | GGA(G) | 225/196 | 1.69/1.46 | 1.66/1.49 | Ter   | UAG(*)  | 9/17    | 0.59/1.11 | 0.07/0.13 |
| Gly   | GGG(G) | 26/57  | 0.2/0.42  | 0.19/0.43 | Thr   | ACA(T)  | 261/239 | 2.25/1.89 | 1.93/1.81 |
| Gly   | GGC(G) | 3/20   | 0.02/0.15 | 0.02/0.15 | Thr   | ACU(T)  | 197/190 | 1.7/1.5    | 1.46/1.44 |
| His   | CAU(H) | 213/188 | 1.8/1.69  | 1.57/1.43 | Thr   | ACG(T)  | 3/53    | 0.03/0.42 | 0.02/0.40 |
| His   | CAC(H) | 24/34  | 0.2/0.31  | 0.18/0.26 | Thr   | ACC(T)  | 2/24    | 0.02/0.19 | 0.01/0.18 |
| Ile   | AUA(I) | 887/657 | 2/1.47    | 6.55/4.98 | Trp   | UGG(W)  | 140/137 | 1/1    | 1.03/0.04 |
| Ile   | AUC(I) | 61/91  | 0.14/0.2  | 0.45/0.69 | Trp   | UGA(W)  | 0/3     | 0/0.2  | 0.00/0.02 |
| Leu   | CUU(L) | 95/147 | 0.34/0.52 | 0.70/1.11 | Tyr   | UAU(Y)  | 946/878 | 1.84/1.74 | 6.99/6.66 |
| Leu   | CUA(L) | 84/181 | 0.3/0.65  | 0.62/1.37 | Tyr   | UAC(Y)  | 85/132  | 0.16/0.26 | 0.63/1.00 |
| Leu   | CUG(L) | 8/37   | 0.03/0.13 | 0.06/0.28 | Val   | GUA(V)  | 397/391 | 2.12/2.13 | 2.93/2.97 |
| Leu   | CUC(L) | 0/15   | 0/0.05   | 0/0.11   | Val   | GGU(V)  | 328/254 | 1.75/1.38 | 2.42/1.93 |
| Leu2  | UUA(L) | 1427/1081 | 5.05/3.86 | 10.54/8.20 | Val   | GUG(V)  | 18/57   | 0.1/0.31 | 0.13/0.43 |
| Leu2  | UUG(L) | 80/220 | 0.28/0.79 | 0.59/1.67 | Val   | GUC(V)  | 7/32    | 0.04/0.17 | 0.05/0.24 |

**Total** 13536/13185 | 64/64 | 100/100
| AA  | Codon | Count | RSCU | %  | AA  | Codon | Count | RSCU | %  |
|-----|-------|-------|------|----|-----|-------|-------|------|----|
| Ala | GCU(A) | 306   | 2.04 | 1.55 | Lys | AAA(K) | 1463  | 1.58 | 7.40 |
| Ala | GCA(A) | 181   | 1.2  | 0.92 | Lys | AAG(K) | 388   | 0.42 | 1.96 |
| Ala | GCG(A) | 45    | 0.3  | 0.23 | Met | AUA(M) | 978   | 1.61 | 4.95 |
| Ala | GCC(A) | 69    | 0.46 | 0.35 | Met | AUG(M) | 482   | 1    | 2.44 |
| Arg | CGU(R) | 139   | 0.86 | 0.70 | Phe | UU(F)  | 1194  | 1.52 | 6.04 |
| Arg | CGA(R) | 57    | 0.35 | 0.29 | Phe | UUC(F) | 377   | 0.48 | 1.91 |
| Arg | CGG( R) | 29 | 0.18 | 0.15 | Pro | CCC(P) | 340   | 2.24 | 1.72 |
| Arg | CGC( R) | 27 | 0.17 | 0.14 | Pro | CCA(P) | 159   | 1.05 | 0.80 |
| Asn | AAC(N) | 399   | 0.59 | 2.02 | Ser1| AGA(S1) | 552   | 3.43 | 2.79 |
| Asn | AAU(N) | 958   | 1.44 | 2.50 | Ser1| AGU(S1) | 384   | 1.49 | 1.94 |
| Cys | UGU(C) | 119   | 1.43 | 0.60 | Ser1| AGG(S1) | 161   | 1    | 0.81 |
| Cys | UGC(C) | 48    | 0.57 | 0.24 | Ser1| AGC(S1) | 162   | 0.63 | 0.82 |
| Gln | CAA(Q) | 327   | 1.42 | 1.65 | Ser2| UCA(S2) | 312   | 1.21 | 1.58 |
| Gln | CAG(Q) | 132   | 0.58 | 0.67 | Ser2| UCC(S2) | 112   | 0.43 | 0.57 |
| Glu | GAG(E) | 158   | 0.47 | 0.80 | Ser2| UCG(S2) | 60    | 0.23 | 0.30 |
| Glu | GAA(E) | 520   | 1.53 | 2.63 | Ter | UAU(Y)  | 923   | 1.4  | 4.67 |
| Gly | GGA(G) | 360   | 1.83 | 1.82 | Thr | ACA(T)  | 269   | 1.2  | 1.36 |
| Gly | GGU(G) | 233   | 1.19 | 1.18 | Thr | ACC(T)  | 111   | 0.5  | 0.56 |
| Gly | GGG(G) | 117   | 0.6  | 0.59 | Thr | ACU(T)  | 428   | 1.91 | 2.17 |
| Gly | GGC(G) | 75    | 0.38 | 0.38 | Thr | ACC(T)  | 269   | 1.2  | 1.36 |
| His | CAC(H) | 135   | 0.57 | 0.68 | Thr | ACG(T)  | 87    | 0.39 | 0.44 |
| His | CAU(H) | 341   | 1.43 | 1.73 | Thr | AGC(T)  | 87    | 0.39 | 0.44 |
| Ile | AAU(I) | 663   | 1.09 | 3.35 | Trp | UGG(W)  | 200   | 1    | 1.01 |
| Ile | AUC(I) | 187   | 0.31 | 0.95 | Trp | UGA(W)  | 12    | 0.82 | 0.06 |
| Leu1 | CUU(L) | 386   | 1.04 | 1.95 | Tyr | UAU(Y)  | 923   | 1.4  | 4.67 |
| Leu1 | CUA(L) | 268   | 0.72 | 1.36 | Tyr | UAC(Y)  | 396   | 0.6  | 2.00 |
| Leu1 | CUC(L) | 122   | 0.33 | 0.62 | Val | GUA(V)  | 425   | 1.68 | 2.15 |
| Leu1 | CUG(L) | 114   | 0.31 | 0.58 | Val | GUC(V)  | 337   | 1.34 | 1.70 |
| Leu2 | UUA(L) | 1006  | 2.7  | 5.09 | Val | GUG(V)  | 153   | 0.61 | 0.77 |
| Leu2 | UUG(L) | 341   | 0.91 | 1.73 | Val | GUC(V)  | 94    | 0.37 | 0.48 |
| **Total** | | 19768 | 64 | 100 |
| Acrasis kona |
|-------------|
| AA | Codon | Count | RSCU | % | AA | Codon | Count | RSCU | % |
| Ala | GCA(A) | 173 | 2.58 | 1.22 | Lys | AAG(K) | 174 | 0.19 | 1.23 |
| Ala | GCU(A) | 89 | 1.33 | 0.63 | Lys | AAA(K) | 1643 | 1.81 | 11.61 |
| Ala | GCG(A) | 3 | 0.04 | 0.02 | Met | AUA(M) | 1188 | 2.1 | 8.40 |
| Ala | GCC(A) | 3 | 0.04 | 0.02 | Met | AUG(M) | 351 | 1 | 2.48 |
| Arg | CGU(R) | 27 | 0.36 | 0.19 | Phe | UUU(F) | 1016 | 1.97 | 7.18 |
| Arg | CGA(R) | 3 | 0.04 | 0.02 | Phe | UUC(F) | 13 | 0.03 | 0.09 |
| Arg | CGC(R) | 1 | 0.01 | 0.01 | Pro | CCA(P) | 118 | 2.57 | 0.83 |
| Arg | CGG(R) | 0 | 0 | 0.00 | Pro | CCC(P) | 1 | 0.02 | 0.01 |
| Asn | AAU(N) | 1690 | 1.89 | 11.94 | Pro | CCG(P) | 5 | 0.11 | 0.04 |
| Asn | AAC(N) | 103 | 0.11 | 0.73 | Pro | CCC(P) | 1 | 0.02 | 0.01 |
| Asp | GAU(D) | 433 | 1.92 | 3.06 | Ser1 | AGA(S1) | 386 | 5.1 | 2.73 |
| Asp | GAC(D) | 18 | 0.08 | 0.13 | Ser1 | AGU(S1) | 379 | 2.95 | 2.68 |
| Cys | UGU(C) | 128 | 1.92 | 0.90 | Ser1 | AGG(S1) | 37 | 0.49 | 0.26 |
| Cys | UGC(C) | 5 | 0.08 | 0.04 | Ser1 | AGC(S1) | 16 | 0.12 | 0.11 |
| Gln | CAA(Q) | 145 | 1.88 | 1.02 | Ser2 | UCU(S2) | 138 | 1.07 | 0.98 |
| Gln | CAQ(Q) | 9 | 0.12 | 0.06 | Ser2 | UCC(S2) | 230 | 1.79 | 1.63 |
| Glu | GAA(E) | 521 | 1.83 | 3.68 | Ser2 | UCG(S2) | 4 | 0.03 | 0.03 |
| Glu | GAG(E) | 48 | 0.17 | 0.34 | Ser2 | UCC(S2) | 5 | 0.04 | 0.04 |
| Gly | GGU(G) | 197 | 2.06 | 1.39 | Ter | UAG(*) | 2 | 0.17 | 0.01 |
| Gly | GGA(G) | 170 | 1.78 | 1.20 | Ter | UAA(*) | 29 | 2.49 | 0.20 |
| Gly | GGG(G) | 14 | 0.15 | 0.10 | Thr | ACU(T) | 143 | 1.27 | 1.01 |
| Gly | GCC(G) | 1 | 0.01 | 0.01 | Thr | ACA(T) | 297 | 2.63 | 2.10 |
| His | CAU(H) | 175 | 1.92 | 1.24 | Thr | ACG(T) | 4 | 0.04 | 0.03 |
| His | CAC(H) | 7 | 0.08 | 0.05 | Thr | ACC(T) | 7 | 0.06 | 0.05 |
| Ile | AUU(I) | 502 | 0.89 | 3.55 | Trp | UGG(W) | 103 | 1 | 0.73 |
| Ile | AUC(I) | 5 | 0.01 | 0.04 | Trp | UGA(W) | 4 | 0.34 | 0.03 |
| Leu1 | CUU(L) | 53 | 0.18 | 0.37 | Tyr | UAU(Y) | 962 | 1.91 | 6.80 |
| Leu1 | CUA(L) | 16 | 0.06 | 0.11 | Tyr | UAC(Y) | 44 | 0.09 | 0.31 |
| Leu1 | CUC(L) | 0 | 0 | 0.00 | Val | GUA(V) | 339 | 2.35 | 2.40 |
| Leu1 | CUG(L) | 1 | 0 | 0.01 | Val | GUU(V) | 164 | 1.14 | 1.16 |
| Leu2 | UUA(L) | 1569 | 5.4 | 11.09 | Val | GUG(V) | 73 | 0.51 | 0.52 |
| Leu2 | UUG(L) | 105 | 0.36 | 0.74 | Val | GUC(V) | 0 | 0 | 0 |
| **Total** | | | **14149** | **64** | **100** |
| AA     | Codon  | Count | RSCU | %   | AA     | Codon  | Count | RSCU | %   |
|--------|--------|-------|------|-----|--------|--------|-------|------|-----|
| Ala    | GCU(A) | 188   | 1.7  | 1.43| Lys    | AAA(K) | 709   | 1.6  | 5.40|
| Ala    | GCC(A) | 28    | 0.25 | 0.21| Met    | AUG(M) | 313   | 1    | 2.38|
| Ala    | GCA(A) | 200   | 1.81 | 1.52| Lys    | AAG(K) | 180   | 0.4  | 1.37|
| Arg    | CGG( R)| 7     | 0.09 | 0.05| Pro    | CCA(P) | 170   | 2    | 1.29|
| Arg    | CGU(R) | 136   | 1.7  | 1.04| Phe    | UUU(F) | 1203  | 1.85| 9.16|
| Arg    | CGC( R)| 14    | 0.17 | 0.11| Pro    | CCC(P) | 7     | 0.08 | 0.05|
| Asn    | AAU(N)| 668   | 1.68 | 5.09| Ser1   | AGA(S1)| 195   | 2.44| 1.49|
| Asn    | AAC(N)| 129   | 0.32 | 0.98| Pro    | CCG(P) | 23    | 0.27 | 0.18|
| Asp    | GAU(D)| 354   | 1.71 | 2.70| Ser1   | AGU(S1)| 273   | 1.43| 2.08|
| Asp    | GAC(D)| 61    | 0.29 | 0.46| Ser1   | AGU(S1)| 273   | 1.43| 2.08|
| Cys    | UGU(C)| 164   | 1.6  | 1.25| Ser1   | AGG(S1)| 65    | 0.81| 0.50|
| Cys    | UGC(C)| 41    | 0.4  | 0.31| Ser1   | AGC(S1)| 71    | 0.37| 0.54|
| Gln    | CAA(Q)| 201   | 1.47 | 1.53| Ser2   | UCA(S2)| 256   | 1.34| 1.95|
| Gln    | CAG(Q)| 73    | 0.53 | 0.56| Ser2   | UCC(S2)| 60    | 0.31| 0.46|
| Gly    | GAA(E)| 292   | 1.46 | 2.22| Ser2   | UCC(S2)| 35    | 0.18| 0.27|
| Gly    | GAG(E)| 107   | 0.54 | 0.81| Ser2   | UCC(S2)| 35    | 0.18| 0.27|
| Gly    | GGA(G)| 207   | 1.41 | 1.58| Ter    | UAA(*)| 34    | 2.22| 0.26|
| Gly    | GGG(G)| 49    | 0.33 | 0.37| Thr    | ACU(T)| 249   | 1.75| 1.90|
| Gly    | GCC(G)| 35    | 0.24 | 0.27| Thr    | ACA(T)| 237   | 1.66| 1.81|
| His    | CAU(H)| 249   | 1.77 | 1.90| Thr    | ACG(T)| 52    | 0.36| 0.40|
| His    | CAC(H)| 32    | 0.23 | 0.24| Thr    | ACC(T)| 32    | 0.22| 0.24|
| Ile    | AUU(I)| 760   | 1.78 | 5.79| Trp    | UGG(W)| 157   | 1    | 1.20|
| Ile    | AUC(I)| 85    | 0.2  | 0.65| Trp    | UGA(W)| 1     | 0.07| 0.01|
| Leu1   | CUG(L)| 13    | 0.05 | 0.10| Val    | GUU(V)| 393   | 2.05| 2.99|
| Leu1   | CUU(L)| 293   | 1.12 | 2.23| Tyr    | UAU(Y)| 731   | 1.68| 5.57|
| Leu1   | CUC(L)| 16    | 0.06 | 0.12| Val    | GUA(V)| 257   | 1.34| 1.96|
| Leu1   | CUA(L)| 103   | 0.39 | 0.78| Tyr    | UAC(Y)| 140   | 0.32| 1.07|
| Leu2   | UUA(L)| 878   | 3.35 | 6.69| Val    | GUG(V)| 81    | 0.42| 0.62|
| Leu2   | UUG(L)| 269   | 1.03 | 2.05| Val    | GUC(V)| 34    | 0.18| 0.26|
| **Total** |     | 13129 | 63.97| 100 |
| AA   | Codon | Count | RSCU | %   | AA   | Codon | Count | RSCU | %   |
|------|-------|-------|------|-----|------|-------|-------|------|-----|
| Ala  | GCU(A) | 172   | 1.3  | 1.00| Lys  | AAA(K)| 932   | 1.48 | 5.40|
| Ala  | GCC(A) | 65    | 0.49 | 0.38| Met  | AUG(M)| 384   | 1    | 2.22|
| Ala  | GCG(A) | 79    | 0.6  | 0.46| Met  | AUA(M)| 511   | 0.98 | 2.96|
| Ala  | GCA(A) | 213   | 1.61 | 1.23| Lys  | AAG(K)| 328   | 0.52 | 1.90|
| Arg  | CGC( R)| 71    | 0.47 | 0.41| Pro  | CCA(P)| 171   | 1.61 | 0.99|
| Arg  | CGU(R) | 216   | 1.43 | 1.25| Phe  | UUU(F)| 1201  | 1.53 | 6.96|
| Arg  | CGG( R)| 58    | 0.38 | 0.34| Pro  | CCC(P)| 145   | 1.36 | 0.84|
| Arg  | CGA( R)| 140   | 0.93 | 0.81| Phe  | UUC(F)| 365   | 0.47 | 2.11|
| Asn  | AAU(N)| 646   | 1.46 | 3.74| Ser1 | AGA(S1)| 272   | 1.8  | 1.58|
| Asn  | AAC(N)| 240   | 0.54 | 1.39| Ser1 | AGC(S1)| 129   | 0.62 | 0.75|
| Asp  | GAU(D)| 279   | 1.38 | 1.62| Ser1 | AGG(S1)| 148   | 0.98 | 0.86|
| Asp  | GAC(D)| 126   | 0.62 | 0.73| Ser1 | ACC(S1)| 129   | 0.62 | 0.75|
| Cys  | UGU(C)| 275   | 1.19 | 1.59| Ser1 | AGG(S1)| 148   | 0.98 | 0.86|
| Cys  | UGC(C)| 186   | 0.81 | 1.08| Ser1 | AGC(S1)| 129   | 0.62 | 0.75|
| Gln  | CAA(Q)| 378   | 1.44 | 2.19| Ser2 | UCU(S2)| 323   | 1.56 | 1.87|
| Gln  | CAG(Q)| 146   | 0.56 | 0.85| Ser2 | UCA(S2)| 271   | 1.31 | 1.57|
| Glu  | GAG(E)| 124   | 0.51 | 0.72| Ser2 | UCC(S2)| 153   | 0.74 | 0.89|
| Glu  | GAA(E)| 367   | 1.49 | 2.13| Ser2 | UCG(S2)| 140   | 0.68 | 0.81|
| Gly  | GGA(G)| 146   | 1.22 | 0.85| Ter  | UAA(*)| 430   | 1.54 | 2.49|
| Gly  | GGC(G)| 82    | 0.69 | 0.48| Thr  | ACA(T)| 254   | 1.45 | 1.47|
| Gly  | GGG(G)| 62    | 0.52 | 0.36| Thr  | ACU(T)| 215   | 1.22 | 1.25|
| Gly  | GGU(G)| 188   | 1.57 | 1.09| Ter  | UAG(*)| 176   | 0.63 | 1.02|
| His  | CAU(H)| 357   | 1.51 | 2.07| Thr  | ACG(T)| 136   | 0.77 | 0.79|
| His  | CAC(H)| 116   | 0.49 | 0.67| Thr  | ACC(T)| 98    | 0.56 | 0.57|
| Ile  | AUU(I)| 822   | 1.57 | 4.76| Trp  | UGG(W)| 221   | 1    | 1.28|
| Ile  | AUC(I)| 235   | 0.45 | 1.36| Trp  | UGA(W)| 232   | 0.83 | 1.34|
| Leu1 | CUG(L)| 104   | 0.3  | 0.60| Val  | GUU(W)| 367   | 1.66 | 2.13|
| Leu1 | CUU(L)| 382   | 1.12 | 2.21| Tyr  | UAU(Y)| 711   | 1.48 | 4.12|
| Leu1 | CUC(L)| 124   | 0.36 | 0.72| Val  | GUA(V)| 263   | 1.19 | 1.52|
| Leu1 | CUA(L)| 234   | 0.68 | 1.36| Tyr  | UAC(Y)| 252   | 0.52 | 1.46|
| Leu2 | UUG(L)| 431   | 1.26 | 2.50| Val  | GUC(V)| 132   | 0.6  | 0.76|
| Leu2 | UUA(L)| 779   | 2.28 | 4.51| Val  | GUG(V)| 124   | 0.56 | 0.72|
| Total|       | 17261 | 64   | 100 |
| AA  | Codon | Count | RSCU | %  | AA  | Codon | Count | RSCU | %  |
|-----|-------|-------|------|----|-----|-------|-------|------|----|
| Ala | GCU(A)| 330   | 1.95 | 2.66 | Lys | AAA(K)| 662   | 1.76 | 5.34|
| Ala | GCC(A)| 54    | 0.32 | 0.44 | Met | AUG(M)| 343   | 1    | 2.76|
| Ala | GCG(A)| 73    | 0.43 | 0.59 | Met | AUA(M)| 191   | 0.6  | 1.54|
| Ala | GCA(A)| 219   | 1.3  | 1.77 | Lys | AAG(K)| 92    | 0.24 | 0.74|
| Arg | CGC(R)| 22    | 0.25 | 0.18 | Pro | CCA(P)| 215   | 1.95 | 1.73|
| Arg | CGU(R)| 297   | 3.41 | 2.39 | Phe | UUU(F)| 790   | 1.73 | 6.37|
| Arg | CGG(R)| 34    | 0.39 | 0.27 | Pro | CCC(P)| 215   | 1.36 | 1.21|
| Arg | CGA(R)| 135   | 1.55 | 1.09 | Phe | UUC(F)| 124   | 0.27 | 1.00|
| Asn | AAU(N)| 393   | 1.54 | 3.17 | Pro | CCG(P)| 57    | 0.52 | 0.46|
| Asn | AAC(N)| 118   | 0.46 | 0.95 | Pro | CCC(P)| 20    | 0.18 | 0.16|
| Asp | GAC(D)| 66    | 0.32 | 0.53 | Ser1| AGU(S1)| 214  | 1.35 | 1.72|
| Asp | GAU(D)| 353   | 1.68 | 2.85 | Ser1| AGA(S1)| 32   | 0.37 | 0.26|
| Cys | UGC(C)| 23    | 0.26 | 0.19 | Ser1| AGC(S1)| 30   | 0.19 | 0.24|
| Cys | UGU(C)| 156   | 1.74 | 1.26 | Ser1| AGG(S1)| 3    | 0.03 | 0.02|
| Gln | CAA(Q)| 293   | 1.48 | 2.36 | Ser2| UCU(S2)| 412  | 2.6  | 3.32|
| Gln | CAG(Q)| 104   | 0.52 | 0.84 | Ser2| UCA(S2)| 106  | 0.67 | 0.85|
| Glu | GAG(E)| 155   | 0.71 | 1.25 | Ser2| UCC(S2)| 127  | 0.8  | 1.02|
| Glu | GAA(E)| 282   | 1.29 | 2.27 | Ser2| UCG(S2)| 63   | 0.4  | 0.51|
| Gly | GGA(G)| 142   | 0.82 | 1.14 | Ter | UAA(*)| 35   | 2.02 | 0.28|
| Gly | GGG(G)| 39    | 0.23 | 0.31 | Thr | ACU(T)| 311  | 1.69 | 2.51|
| Gly | GGC(G)| 42    | 0.24 | 0.34 | Thr | ACA(T)| 261  | 1.42 | 2.10|
| Gly | GGU(G)| 469   | 2.71 | 3.78 | Ter | UAG(*)| 14   | 0.81 | 0.11|
| His | CAC(H)| 46    | 0.34 | 0.37 | Thr | ACC(T)| 85   | 0.46 | 0.69|
| His | CAU(H)| 225   | 1.66 | 1.81 | Thr | ACG(T)| 78   | 0.42 | 0.63|
| Ile | AUU(I)| 633   | 1.99 | 5.10 | Trp | UGG(W)| 159  | 1    | 1.28|
| Ile | AUC(I)| 130   | 0.41 | 1.05 | Trp | UGA(W)| 3    | 0.17 | 0.02|
| Leu1| CUG(L)| 29    | 0.11 | 0.23 | Val | GUU(V)| 473  | 2.05 | 3.81|
| Leu1| CUU(L)| 191   | 0.76 | 1.54 | Tyr | UAU(Y)| 485  | 1.73 | 3.91|
| Leu1| CUC(L)| 22    | 0.09 | 0.18 | Val | GUA(V)| 286  | 1.24 | 2.51|
| Leu1| CUA(L)| 78    | 0.31 | 0.63 | Tyr | UAC(Y)| 76   | 0.27 | 0.61|
| Leu2| UUA(L)| 862   | 3.42 | 6.95 | Val | GUG(V)| 141  | 0.61 | 1.14|
| Leu2| UUG(L)| 332   | 1.32 | 2.68 | Val | GUC(V)| 22   | 0.1  | 0.18|

| Total|       | 12407 | 64   | 100 |
Figure S1. Comparison of the three available sequences of *orf145*. (A) The position of the conserved *orf145* in the three heterolobosean species: *Pleurostomum flabellatum*, *Naegleria gruberi* and *Naegleria fowleri*. The step size is 200 bp in *P. flabellatum* and 100 bp in *Naegleria*. (B) Sequence logo plots of aligned *orf145* amino acids generated using the Geneious 8.1.2 multiple sequence alignment tool.
Figure S2. Relative synonymous codon usage (RSCU) in heterolobosean mitogenomes. The 22 codon families consisting of a total of 61 two- and four-fold degenerate synonymous codons are plotted on the x-axis. The RSCU values are shown on the y-axis. The most used synonymous codon in each family is in the bottom. Red-colored codon, codon is not present in the mitogenome.
Figure S3. Distribution of the synonymous codons on the first two axes of the correspondence analysis. Representation of the first two axes of the correspondence analysis performed on the codon frequency of heterolobosean mitochondrial-encoded proteins.
Figure S4. Bayesian phylogenetic tree of the Discoba based on mitogenome data under the CAT + GTR model. Contrary to ML tree inference, Heteroloboseid sp. BB2 and Pharyngomonas are paraphyletic in Bayesian tree inference.
Figure S5. Mitochondrial genome synteny of *Pleurostomum flabellatum* and *Pharyngomonas kirbyi*. Images were generated using the Mauve genome alignment tool with default settings (Darling et al., 2010). Color-coded syntenic blocks indicate conserved segments. Sequence similarity is shown within each syntenic block. Regions with no color indicate no detectable homology between the two mitogenomes.
Figure S6. Maximum likelihood phylogenetic tree of ccmF protein sequences. The ccmF gene was found in Pleurostomum, Naegleria and Stachyamoeba, previously misannotated by orf457 (YP_009118159.1). The ccmF gene is encoded in a wide array of other eukaryotic lineages, namely, Viridiplantae, Rhodophyta, Alveolata, Malawimonadidae, Metazoa, Cryptophyta and Jakobida. Our ccmF phylogenetic analysis places the Pl. flabellatum sequence as more closely related to Rhodophyta and Malawimonadidae mitochondrial ccmFs than to either Tetramitia (St. lipophora and Naegleria) or Jakobida ccmFs, probably owing to the rapid divergence of Heterolobosean species.
Figure S7. Maximum likelihood phylogenetic tree of rps4 protein sequences. The gene rps4 was found uniquely in *Pleurostomum* and *Naegleria* among Heterolobosea. The gene rps4 is encoded in a wide array of other eukaryotic lineages, namely, Viridiplantae, Rhodophyta, Alveolata, Malawimonadidae, Metazoa, Cryptophyta and Jakobida.

Reconstruction from rps4 clustered *Pl. flabellatum* and *Naegleria rps4s* firmly in one clade (BS=96%). This clade was more closely connected to mitochondrial rps4s of Viridiplantae than to either Jakobida or *Tsukubamonas rps4s*.
Figure S8. Maximum likelihood phylogenetic tree of ccmC protein sequences. Reconstruction from ccmC of Naegleria genus, are well resolved as a monophyletic cluster with mitochondrial ccmCs of Rhodophyta (BS=95%) and distantly clustering with mitochondrial ccmCs of Jakobida and Viridiplantae (BS=83%).
Figure S9. Maximum likelihood phylogenetic tree of tatC protein sequences. The phylogenetic analysis of the TAT machinery inducer in most of TAT-containing eukaryotes tatC (Petru et al. 2018) strongly clustered Naegleria into one clade with Rhodophyta mitochondrial tatCs (BS=90%) but somewhat distantly clustered with Jakobida tatCs.
Figure S10. Maximum likelihood phylogenetic tree of ccmA protein sequences. The ccmA gene of St. lipophora, is resolved as sister-ta to ccmAs of several Proteobacteria with low bootstrap value (below 50%; supplementary fig. S9) and distantly related to mitochondrial ccmAs of Jakobida align with other eukaryotes (mostly Viridiplantae) (BS=89%).
Figure S11. Multiple sequence alignment of Mic60 homologs. (A) *Naegleria gruberi* Mic60 sequence (Naegru_XP_002683319.1), and two putative Mic60 homologs from the *Pleurostomum flabellatum* genome assembly (Pleuflab_MW019459 and Pleuflab_MW019460) were aligned. The alignment of translated transcript sequences of the coding regions (Pleuflab_transcr_seq01 and Pleuflab_transcr_seq02) was also added. The mitofilin domains are marked in black frames. (B) 5’ and 3’ UTR regions of transcript sequences are highlighted.
Figure S11. Multiple sequence alignment result of Mic60 homologs. (A) *Naegleria gruberi* Mic60 sequence (Naegru_XP_002683319.1), and two putative Mic60 homologs from the *Pleurostomum flabellatum* genome assembly (Pleuflab_MW019459 and Pleuflab_MW019459) were aligned. The alignment of translated transcript sequences of the coding regions (Pleuflab_transcr_seq01 and Pleuflab_transcr_seq02) was also added. The mitofilin domains are marked in black frames. (B) 5' and 3' UTR regions of transcript sequences are highlighted.