Mitochondrial genes of Euglenozoa (Kinetoplastida, Diplomea, and Euglenida) are notorious for being barely recognizable, raising the question of whether such divergent genes actually code for functional proteins. Here we demonstrate the translation and identify the function of five previously unassigned y genes encoded by mitochondrial DNA (mtDNA) of diplonemids. As is the rule in diplomimid mitochondria, y genes are fragmented, with gene pieces transcribed separately and then trans-spliced to form contiguous mRNAs. Further, y transcripts undergo massive RNA editing, including uridine insertions that generate up to 16-residue-long phenylalanine tracts, a feature otherwise absent from conserved mitochondrial proteins. By protein sequence analyses, MS, and enzymatic assays in Diplonema papillatum, we show that these y genes encode the subunits Nad2, -3, -4L, -6, and -9 of the respiratory chain Complex I (CI; NADH:ubiquinone oxidoreductase), Complex II (CII; ubiquinol:cytochrome c oxidoreductase), Complex III (CIII; ubiquinone:cytochrome c oxidoreductase), and Complex IV (CIV; cytochrome c oxidase). Complex V (CV; ATP synthase) exploits the resulting transmembrane H⁺ gradient for regenerating ATP from ADP and phosphate (oxidative phosphorylation). The complexes interact in the mitochondrial inner membrane, forming higher-order structures called supercomplexes (SC), such as the respirasome, which includes a single CI and CIV, as well as a CIII homodimer (SCI,SCI,III,IV,1; reviewed by Milenkovic et al. (5) and Letts et al. (6)). It should be noted that mitochondria from several organisms (Saccharomyces cerevisiae, Schizosaccharomyces pombe, land plants such as mistletoe, and various protists like Plasmodium and Chromera) lack conventional Complex I (also referred to as type I NADH dehydrogenase). In these latter taxa, electron transport to ubiquinone is catalyzed by a nonproton-pumping enzyme (type II NAD(P)H dehydrogenase).

Our knowledge about the molecular mechanics underlying energy transfer is partial. The rotary motor-like behavior of the ATP synthase has been examined in great detail (7). Further, the mechanism of coupling electron transport to proton translocation has been elucidated for two respiratory complexes. Complex III utilizes as a shuttle the pool of ubiquinone, a molecule that is freely mobile in the lipid bilayer (8). Complex IV, in contrast, traps protons and channels them across the membrane (9). Much less is known about the energy

3 The abbreviations used are: CI, Complex I; CIII, Complex III; CV, Complex IV; CV, Complex V; BN-PAGE, blue native PAGE; CBB, Coomassie Brilliant Blue; DDM, n-dodecyl-β-D-maltoside; HMM, Hidden Markov model; iBAQ, intensity-based absolute quantification; mtDNA, mitochondrial DNA; MTP, mitochondrial targeting peptide; SC, supercomplex(es); SOD, superoxide dismutases; TMH, trans-membrane helix; γCA, γ-carboxy anhydrase; 3D, three-dimensional; PAI, protein abundance index; nt, nucleotide(s); aa, amino acid(s).
transfer of Complex I. Three-dimensional (3D) reconstructions, bioenergetics analyses, and computational modeling suggest that the proton-motive force is generated through electrostatic coupling between membrane and matrix components of the complex (10–12); however, the exact mechanism is still uncertain.

Mitochondrial Complex I (type I NADH dehydrogenase) is the largest of the respiratory-chain complexes. It is the major entry point of electrons into the respiratory chain by oxidizing NADH and transferring electrons via ubiquinone to cytochrome c. Fourteen subunits of the complex originate from the bacterial ancestor of the organelle. These subunits constitute the catalytic core (i.e. the minimum set for NADH:ubiquinone oxidoreduction and proton pumping) (4). Among the genes specifying core subunits (“core genes”), some have remained in the mitochondrial DNA (mtDNA), whereas others have migrated to the nucleus over evolutionary time. The number of mtDNA-encoded subunits varies from five (e.g. in the green alga *Chlamydomonas*) to a maximum of 12 in jakobids (13, 14), with the remaining core genes residing in the nucleus.

An additional set of ~30 proteins form the so-called accessory (supernumerary) subunits of mitochondrial CI. These components do not directly participate in the catalytic activity, but rather in assembly, structure, and stabilization of the complex (15, 16). The accessory proteins can be categorized into conserved and more narrowly distributed, lineage-specific subunits. Homologs of ~20 subunits were biochemically identified in CI across various eukaryotic lineages (17–23). The same studies also identified a class of CI subunits occurring only in certain lineages. The number of these clade-specific subunits is typically in the order of 2–5 in fungi, animals, and plants (17, 19, 23) to almost 3 dozen postulated in trypanosomatids (20).

One example of a broadly (but not universally) distributed subunit is the γ-carboxic anhydrase (γCA), which is part of CI in a large number of eukaryotes, including plants, green algae, discobans, amoebzoans, and possibly stramenopiles, but not in animals and fungi (21, 24). The physiological role of γCAs in CI has remained unclear, but at least in plants, a heterotramer of γCAs and γCA-like proteins is required for the proper assembly of the membrane arm of Complex I (25–27) and remains associated with the mature respiratory complex (28). Examples of clade-specific CI subunits are NDUFV3 (10 kDa) and NDUFA10 (42 kDa), which so far have been identified only in animals, as well as NUSM, NUTM, and NUUM in yeasts and P1–P3 in land plants (17, 19, 23).

The most well-conserved proteins of Complex I are the core components, which in general can be readily identified by sequence similarity. However, in divergent eukaryotes and taxa without closely related model organisms, even otherwise highly conserved proteins can be difficult to recognize (21). This applies in particular to mitochondrion-encoded genes (e.g. see Burger et al. (29) and Kannan and Burger (30)).

A group of eukaryotes renowned for extremely rapidly evolving mitochondrial genes are the diplonemids, which, together with kinetoplastids and euglenids, form Euglenozoa (Fig. 1). The latter clade, in turn, is part of the supergroup Discoba, an assemblage that diverged very early from the branches leading to the well-studied plants, animals, and fungi. Diplonemids have been traditionally considered a species-poor and insignificant group. Recently, this view has drastically changed. Global environmental surveys have revealed that diplonemids are, in fact, one of the most abundant and diverse marine eukaryotes on Earth (31, 32).

In addition to rapidly evolving mitochondrial gene sequences, the mitochondrial genome structure of diplonemids is most eccentric (33, 34). mtDNA of the type species Diplonema papillatum consists of ~80 distinct circular chromosomes either 6 or 7 kilobase pairs long, and its mitochondrial genes are split into up to 11 pieces (modules) of 43–534 nt. Each such piece is encoded individually on one of the DNA circles, transcribed separately into RNA precursors, end-processed to contain exclusively coding sequence, and finally joined into contiguous RNAs (35, 36). The molecular mechanism by which this unique trans-splicing proceeds has yet to be unraveled.

A further remarkable post-transcriptional process in diplonemid mitochondria is RNA editing, which affects more than three-fourths of mtDNA-encoded genes. One type of RNA editing observed in this group somewhat resembles U-insertion editing in kinetoplastid mitochondria, as up to 50 uridines are added at 3′-ends of certain gene module transcripts. This “U-appendage” RNA editing takes place before trans-splicing. The other RNA editing types are C-to-U and A-to-I substitutions, which are tightly clustered (e.g. 29 sites within a stretch of 55 nt in the nad4 pre-mRNA of *D. papillatum*). RNA editing in diplomemid mitochondria is function-critical, restoring reading frames and compensating for lost portions of gene sequences (37).

The gene complement identified so far in diplomemid mtDNAs comprises 12 typical mitochondrial genes, encoding the large and small subunit mitribosomal RNAs, as well as protein components of the respiratory chain and oxidative phosphorylation. The corresponding genes are *nad1*, *-4*, *-5*, *-7*, and *-8* (CI subunits); *cob* (CIII); *cox1*, *-2*, and *-3* (CIV); and *atp6* (CV). The function of six additional genes termed *y1–y6* has remained obscure, although five of the six are conserved across diplomemids (38).

In this report, we present experimental and *in silico* evidence that the enigmatic mitochondrial *y* genes of diplomemids in fact code for CI subunits. We further establish the catalogue of nucleus-encoded CI proteins for *D. papillatum*. These *y* genes illustrate the extent to which fast-evolving

---

**Figure 1. Phylogenetic placement of diplomemids in the eukaryotic tree.** Diplonemids (blue) belong to Euglenozoa (shaded background) together with kinetoplastids and euglenids. The phylogenetic tree was constructed using 10 mitochondrion-encoded Nad proteins from representative species of major eukaryotic groups. Note the high evolutionary rates in euglenozoans. Scale bar, number of per-site substitutions.
mtDNAs such as those from diplonemids are exploring the functional limits of sequence space. In addition, the example of y genes demarcates the limitations of computational function assignment based on protein sequence alone, with up to 30% of yet unassigned protein-coding genes even in genomes of model organisms such as yeast, *Escherichia coli*, *Arabidopsis*, and humans (39, 40).

**Results**

**Gene structure and post-transcriptional maturation of y genes in *D. papillatum* mtDNA**

We reported earlier six unidentified genes (*y1*–*y6*) in mtDNA of *D. papillatum*. These genes consist of two to five modules that are transcribed separately followed by trans-splicing of module transcripts to contiguous polyadenylated mRNAs (Table 1). Preliminary tandem MS data supported the notion that these y genes code for proteins rather than structural RNAs (37). Compared with the assigned mitochondrion-encoded genes of *D. papillatum*, RNA editing of y genes is much more prevalent. Post-transcriptional substitutions are observed in more than half of the y genes, but only in 20% of the assigned genes. Further, U-appendage RNA editing affects all y genes, but only 60% of the assigned genes, with the numbers of both sites per gene and uridines added per site being considerably higher in the y genes (Table 1).

**Table 1**

Protein-coding genes specified by diplonemid mtDNA

| Gene | Gene product (complex) | Number of modules | Number of RNA editing sites |
|------|------------------------|-------------------|-----------------------------|
|      |                        |                   | Substitution | U-appendage |
|      |                        |                   | (number of uridines) | (number of uridines) |
| *nad1*<sup>a</sup> | NADH dehydrogenase subunit 1 (CI) | 5 | 0 | 1 (16) |
| *nad4*<sup>a</sup> | NADH dehydrogenase subunit 4 (CI) | 8 | 29 | 1 (2) |
| *nad5*<sup>a</sup> | NADH dehydrogenase subunit 5 (CI) | 11 | 0 | 0 |
| *nad7*<sup>a</sup> | NADH dehydrogenase subunit 7 (CI) | 9 | 1 | 0 |
| *nad8*<sup>a</sup> | NADH dehydrogenase subunit 8 (CI) | 3 | 0 | 0 |
| *col1*<sup>a</sup> | Apocynochrome b (CIII) | 6 | 0 | 1 (3) |
| *cox1*<sup>a</sup> | Cytochrome oxidase subunit 1 (CIV) | 9 | 0 | 1 (6) |
| *cox2*<sup>a</sup> | Cytochrome oxidase subunit 2 (CIV) | 4 | 0 | 1 (3) |
| *cox3*<sup>a</sup> | Cytochrome oxidase subunit 3 (CIV) | 3 | 0 | 1 (1) |
| *atp6*<sup>b</sup> | ATP synthase subunit 6 (CV) | 3 | 0 | 0 |
| *y1* (nad3)<sup>b</sup> | NADH dehydrogenase subunit 3 (CI) | 2 | 11 | 1 (4) |
| *y3* (nad2)<sup>b</sup> | NADH dehydrogenase subunit 2 (CI) | 5 | 0 | 0 |
| *y2* (nad9)<sup>b</sup> | NADH dehydrogenase subunit 9 (CI) | 4 | 0 | 0 |
| *y4* | Protein of unknown function | 2 | 0 | 0 |
| *y5* (nad6)<sup>b</sup> | NADH dehydrogenase subunit 6 (CI) | 3 | 18 | 1 (50) |
| *y6* (nad4L)<sup>b</sup> | NADH dehydrogenase subunit 4L (CI) | 2 | 0 | 1 (6) |

<sup>a</sup> Homologs are present in euglenids (*E. gracilis*) and kinetoplastids (*T. brucei*).  
<sup>b</sup> Homologs have not been detected in euglenids but are present in kinetoplastids (*T. brucei*): *murf1* (nad2), *cr3* (nad4L), *cr4* (nad6), and *cr5* (nad3).

Most y genes are conserved across diplonemids and have remote sequence similarity to nad genes

Sequence similarity searches by BLAST indicated that mtDNAs from *Diplonema ambulator*, *Rhynchopus euleelides*, and *Flectonema neradi* (previously *Diplonema* sp. 2) potentially code for proteins marginally resembling *y1*, *y2*, *y3*, *y5*, and *y6* of *D. papillatum*. No homologs could be detected for *y4* (38).

To conduct more sensitive searches of homologs beyond diplonemids, we built for each Y protein multiple alignments of sequences from the four taxa and constructed profile hidden Markov models (HMMs), with which we searched for homologs among GenBank<sup>TM</sup> RefSeq mitochondrial proteins from diverse eukaryotes. *Y1* was identified as a potential divergent Nad3 (*E*-value $-10^{-4}$), but no significant hits (if any) were found for other Y proteins. More convincing results were obtained with profile HMM–profile HMM comparisons, which are more sensitive than profile–sequence comparison methods at detecting remote similarity (41). For that, profile HMMs were built from a taxonomically broad sampling for each of the mitochondrion-encoded CI subunits not previously known to be encoded by diplonemid mtDNA (*nad2*, -3, -4L, -6, and -9). Comparison of these Nad profile HMMs with Y-protein profile HMMs yielded top hits with *E*-values ranging from $8.1 \times 10^{-4}$ (Nad9) to $5.4 \times 10^{-2}$ (Nad6), assigning Y proteins to Nad2 (*Y3*), Nad3 (*Y1*), Nad4L (*Y6*), Nad6 (*Y5*), and Nad9 (*Y2*), all subunits whose genes previously seemed missing from *D. papillatum* mtDNA (Table 1, Fig. 2 (A–E), and Figs. S1 and S2).

The assignment of *Y5* to Nad6 was the most challenging, because the sequence similarity is only marginal. Still, two protein features support this assignment. First, polar and apolar residues in Y5 homologs are distributed across the protein in a similar fashion as in Nad6 (although suggesting four instead of five transmembrane helices (TMHs) in Y5; Fig. S3). Second, all diplomids share a conserved hydrophobic patch that in other Nad6 homologs contains the tentatively functionally important and conserved Tyr residue of the third TMH (Ytr/Phe in diplomids; Fig. 2D) (42). More details on the evolutionary remodeling of diplonemid Nad6(Y5) are presented in the supporting Results and Figs. S1–S3.

With the five newly assigned genes, the set of mitochondrion-encoded *nad* genes in diplomids has grown to 10, encoding Nad1–9 and Nad4L, as seen in kinetoplastids and many other eukaryotic groups (43). (For the assignment of kinetoplastid *murf1*, *cr3*, *cr4*, and *cr5* to nad2, nad4L, nad6, and nad3, respectively, see supporting Results (30, 44).)

**Y proteins are components of the respiratory chain Complex I**

To verify whether Y proteins are indeed part of CI, we separated mitochondrial multiprotein complexes of *D. papillatum* by blue native PAGE (BN-PAGE). Complex I was identified by in-gel staining of NADH-dehydrogenase activity, revealing

---

*Respiratory chain Complex I of diplonemids*

---

J. Biol. Chem. (2018) 293(41) 16043–16056
three distinct bands migrating at 2.1–2.5 MDa, 1.3–1.5 MDa, and 1.1–1.3 MDa (Fig. 3 and Fig. S4). In the following, we will refer to these bands as large, medium, and small, respectively. The ratio of NADH-dehydrogenase activity (relative to protein quantity) across these bands was ~2.3:3 according to a time series of activity staining.
Respiratory chain Complex I of diplonemids

Composition of Complex I from D. papillatum

We investigated which of the proteins detected in MS experiments have significant sequence similarity with CI subunits from model organisms. Specifically, we performed sequence similarity searches of all experimentally confirmed NADH:ubiquinone oxidoreductase subunits from model organisms against the D. papillatum proteome. Inversely, we searched for homologs of all proteins, for which peptides were detected by MS analysis of the three complexes, against GenBankTM. When an otherwise conserved CI component was not found by BLAST (e.g. NDUF2, NDUF8, and NDUF6), the corresponding Pfam or in-house–generated profile HMM was used to search against the entire genome-inferred Diplonema proteome.

By employing this strategy, we identified among the experimentally determined Diplonema proteins the expected four nucleus-encoded core subunits and 16 of the ~25 accessory subunits that were broadly distributed across eukaryotes (Table 2 and Fig. 4). In addition, all three examined CI-containing complexes from Diplonema possess six of ~35 accessory proteins that were reported to be associated specifically with CI from Trypanosoma brucei (Table 2 and Fig. 4) (20, 47). Thus, the count of nucleus-encoded CI subunits having homologs in other lineages (hereafter termed “conventional”) is 26. It should be noted that D. papillatum appears to lack certain otherwise widely distributed subunits, such as NDUF4, NDUF1, NDUF11, and NDUFB1–NDUF5 (Fig. 4 and Table S3) because their homologs are absent from the inferred proteome.

Next, we asked whether the 26 conventional nucleus-encoded components indeed co-purify with CI. For that, we performed a quantitative analysis of the MS-identified proteins contained in the three BN-PAGE bands. Two metrics were employed: the abundance determined by normalized peptide intensities (iBAQ values) and the enrichment in bands compared with the mitochondrial lysate. In the three bands, all but one nucleus-encoded CI subunits were highly abundant and enriched (Fig. 5 and Table S2). The only exception was Diplonema’s NDUFAB1 homolog, which was less enriched probably because it is also part of other mitochondrial complexes (for details, see “Discussion”).

Each of the BN-PAGE bands also contained 15 additional nucleus-encoded proteins with a similar enrichment profile as bona fide CI subunits but without significant sequence similarity to known subunits (Fig. 5 and Fig. S6). Four of these proteins contain domains known to be involved in oxidoreductive processes, such as FAD and NAD binding, superoxide dismutase, and adrenodoxin, and five other proteins are predicted to have transmembrane helices (Table 3). These 15 novel proteins, which are conserved across the four diplonemid species, probably represent diplonemid-specific CI subunits.

The set of conventional nucleus-encoded CI subunits together with the 10 mtDNA-encoded Nad proteins, and the putative 15 novel components (in total ~425 kDa) add up to a cumulative size of ~1.6 MDa, which is compatible with the migration behavior of the medium BN-PAGE band displaying NADH-dehydrogenase activity (Fig. 3).

All three BN-PAGE bands were analyzed by MS (Table S1). For that, the bands were digested with trypsin or with trypsin and chymotrypsin combined, the latter to increase the number of peptides in the observable size range for mtDNA-encoded proteins. Trypsin digestion produced peptides of a subset of assigned mtDNA-encoded proteins and four of the six Y proteins (Y2, Y3, Y4, and Y5), whereas combined digestion yielded peptides of all assigned mtDNA-encoded Nad proteins, as well as five Y proteins (Y1, Y2, Y3, Y4, and Y6; Table 2 and Fig. S5).

To determine which components belong to the same complex, we performed protein quantification based on peptide intensities (intensity-based absolute quantification (iBAQ) (45)), as well as spectral counts (protein abundance index (PAI) (46)) versus protein enrichment in a given CI band relative to total mitochondrial lysate. The results of this analysis (Fig. 5 and Table S2) strongly indicated that Y1, Y2, Y3, and Y6 are indeed part of the same complex as the previously assigned Nad proteins (and nucleus-encoded CI subunits; see below). Due to technical challenges, Y5 could not be reliably quantified (see also supporting Results).

The only remaining unassigned protein is Y4. It is predicted to have two TMHs and was detected in all three CI bands (with the highest abundance in the large band; see Table S2). Because no homologs were found in the inferred proteomes of other diplonemids (38), we speculate that Y4 could be a novel, fast-evolving accessory protein of the diplonemid respiratory chain.
Respiratory chain Complex I of diplonemids

Table 2
MS identifications of 36 assigned Complex I subunits from D. papillatum

| Subunit | Type | Length | Predicted size | MTP cleavage (length) |
|---------|------|--------|----------------|----------------------|
| Dp ID   |       |        |                |                      |
| Nad1    | ND1  | C      | 307            | 33.3                 | NA                   |
| Nad2 (Y3) | ND2  | C      | 327            | 35.4                 | NA                   |
| Nad3 (Y1) | ND3  | C      | 115            | 12.6                 | NA                   |
| Nad4    | ND4  | C      | 434            | 47.2                 | NA                   |
| Nad4L (Y6) | ND4L | C      | 80             | 8.7                  | NA                   |
| Nad5    | ND5  | C      | 559            | 60.4                 | NA                   |
| Nad6 (Y5) | ND6  | C      | 136            | 15.4                 | NA                   |
| m.9489  | NDUF51 | AB     | 292            | 32.5                 | (17)                 |
| m.9707  | NDUF52 | AB     | 377            | 40.5                 | NA                   |
| m.11119 | NDUF53 | AB     | 169            | 18.6                 | NA                   |
| m.919   | NDUF54 | AB     | 166            | 19.1                 |                      |
| m.3933  | NDUF55 | AB     | 216            | 24.3                 | (10)                 |
| m.20953 | NDUF56 | AB     | 209            | 23.4                 |                      |
| m.7247  | NDUF57 | AB     | 122            | 13.1                 | NA                   |
| m.1531  | NDUF58 | AB     | 502            | 55.5                 | (14)*                |
| m.18000 | NDUF59 | AB     | 263            | 28.9                 | (24)*                |
| m.44    | NDUF60 | AB     | 192            | 21.2                 |                      |
| m.9838  | NDUF61 | AB     | 390            | 44.2                 | (9)*                 |
| m.24276 | NDUF62 | AB     | 607            | 69.4                 | (10)                 |
| m.8672  | NDUF63 | AB     | 238            | 28.0                 |                      |
| m.20953 | NDUF64 | AB     | 458            | 51.2                 | (10)                 |
| m.1088  | NDUF65 | AB     | 365            | 42.0                 | (16)                 |
| m.4241  | NDUF66 | AB     | 216            | 24.3                 |                      |
| m.4695  | NDUF67 | AB     | 105            | 12.1                 | (20)                 |
| m.2862  | NDUF68 | AB     | 255            | 29.4                 |                      |
| m.22777 | NDUF69 | AB     | 303            | 34.8                 | (15)*                |
| m.20813 | NDUF70 | AB     | 224            | 26.9                 |                      |
| m.3369  | NDUF71 | AB     | 155            | 18.1                 | (20)*                |
| m.27345 | NDUF72 | AK     | 272            | 29.4                 |                      |
| m.18432 | NDUF73 | AK     | 399            | 44.9                 |                      |
| m.3527  | NDUF74 | AK     | 54             | 61.6                 |                      |
| m.52599 | NDUF75 | AK     | 319            | 33.1                 |                      |
| m.3383  | NDUF76 | AK     | 369            | 41.3                 |                      |
| m.60044 | NDUF77 | AK     | 479            | 51.7                 | (11)                 |
| m.60040 | NDUF78 | AK     | 139            | 16.4                 |                      |

| Number of peptides | Trypsin | sequence coverage | Trypsin + chymotrypsin | sequence coverage |
|--------------------|---------|-------------------|------------------------|-------------------|
| Nad1               | 3       | 10.4%             | 10                     | 26.1%             |
| m.919              | 18      | 76.3%             | 19                     | 52.6%             |
| m.3933             | 10      | 42.3%             | 7                      | 23.1%             |
| m.7247             | 52      | 85.8%             | 39                     | 55.3%             |
| m.1531             | 14      | 56.1%             | 27                     | 64.5%             |
| m.18000            | 22      | 94.8%             | 21                     | 62.3%             |
| m.9838             | 54      | 90.0%             | 40                     | 69.2%             |
| m.24276            | 60      | 87.8%             | 65                     | 62.0%             |
| m.8672             | 51      | 88.4%             | 43                     | 69.4%             |
| m.20953            | 57      | 84.3%             | 29                     | 51.4%             |
| m.1088             | 14      | 73.0%             | 8                      | 38.1%             |
| m.4241             | 2       | 26.0%             | 18                     | 61.5%             |
| m.4695             | 24      | 87.8%             | 15                     | 37.4%             |
| m.2862             | 23      | 60.9%             | 24                     | 49.3%             |
| m.22777            | 28      | 79.7%             | 13                     | 42.2%             |
| m.20813            | 12      | 40.9%             | 7                      | 35.1%             |
| m.3369             | 34      | 100%              | 47                     | 84.1%             |
| m.27345            | 47      | 85.9%             | 29                     | 53.5%             |
| m.18432            | 56      | 93.7%             | 42                     | 64.9%             |
| m.3527             | 30      | 67.9%             | 30                     | 61.3%             |
| m.52599            | 37      | 82.6%             | 36                     | 59.5%             |
| m.3383             | 57      | 78.6%             | 59                     | 72.0%             |
| m.60044            | 45      | 80.3%             | 49                     | 69.7%             |

Discussion

Diplomem mitochondrial Y genes code for highly derived Complex I subunits

Combined experimental and computational approaches revealed that the Y proteins of diplonemids are mitochondrion-encoded CI subunits. Several of these subunits (e.g. Nad9 and Nad2) in diplonemids are even more derived than their kinetoplastid counterparts, which existed unassigned in the literature for more than 20 years under the acronyms MURF and CR (e.g. see Duarte et al. (44) and Opperdoes et al. (48)).

The detection of Y (and other Nad) proteins by MS further demonstrates that trans-spliced and edited mRNAs are indeed translated, because certain observed peptides cover module junctions and post-transcriptionally added U-tracts (Fig. S5). Particularly remarkable is the faithful translation of >10-nt-long U tracts. Although homopolymeric stretches in mRNAs are known to induce nucleotide skipping or back-slapping of the translational machinery (see references in Burger et al. (49)), we
did not observe any indication of such ribosomal frameshifting in *D. papillatum*.

**Conspicuous low-complexity regions in the newly assigned Nad proteins of diplonemids**

The most extreme case of U-appendage RNA editing occurs in *nad6* (y5). In all examined diplomids, the *nad6* mRNA contains 50 post-transcriptionally added uridines specifying a 16-residue-long Phe tract. Single-amino acid (homopeptide) repeats such as these Phe tracts form low-complexity regions in the corresponding protein. Reported low-complexity region proteins, which are quite common across eukaryotes and bacteria, are made up predominantly of Ser, Gly, Ala, Asn, and Gln residues, whereas Phe repeats are among the rarest (50, 51). Due to their bulky and hydrophobic nature, Phe residues (and also Leu, Ile, and Val) occur mostly in TMHs (52, 53). Accordingly, computational protein-structure analyses predict that the Phe tracts of diplonemid Nad2 (Y3) and Nad6 (Y5) (as well as Nad1 and Nad5 (38)) are part of TMHs that are superimposable with TMHs from homologs whose 3D structure is known (Fig. 2 and Fig. S1). The only exception is Nad9 (Y2), which in all diplomids contains six consecutive Phe residues that are not predicted to form a transmembrane helix. Structural alignments to Nad9 counterparts with resolved 3D structures indicate that this Phe tract corresponds to a segment that is buried deep inside the protein bundle of the Q module.

**Structural consequences of divergent mitochondrion-encoded subunits**

Several mitochondrion-encoded components of diplonemid CI lack TMHs that are otherwise present in proteins from most other eukaryotic groups, including kinetoplastids. For example, instead of the canonical 14 TMHs, Nad2 (Y3) of diplomids is predicted to contain only 11; the three N-terminal helices are absent, as in mammalian Nad2 (Figs. S1 and S2). Because in the 3D structures of mammalian CI, these missing segments are not structurally compensated for by other proteins (54, 55), they appear to be dispensable.

Structural deviations are also observed in diplomemid Nad4L (Y6) and Nad6 (Y5), in which the otherwise highly conserved most N-terminal TMH is missing or severely truncated. (The same appears to apply to their kinetoplastid counterparts (Fig. S1)). In all determined CI structures, the N-terminal TMHs of Nad4L and Nad6 are in physical contact with one another.
Respiratory chain Complex I of diplonemids

Figure 5. Quantification of Complex I components in Diplonema. Enrichment–abundance plots of proteins detected in MS analyses of BN-PAGE bands displaying NADH-dehydrogenase activity. A, proteins detected in the trypsin digestion of the large band and in the trypsin + chymotrypsin digestions of the large, medium, and small bands. Highlighted are Complex I (CI) subunits with homologs in other species (dark blue) and putative diplonemid-specific components (light blue), CIII and CIV subunits (dark red and magenta, respectively), and other proteins observed as significantly enriched and abundant in the large band (orange). CV subunits shown as dark gray dots are proteins not observed as being enriched in any of the three analyzed BN-PAGE bands. Dotted lines indicate significance thresholds inferred from nucleus-encoded conventional CI subunits. The lower of the two vertical enrichment thresholds in the trypsin-digested large band indicates the levels of the multifunctional NDUFAB1 subunit (mitochondrial acyl-carrier protein). Inset, color key. For details, see Fig. S6 and Table S2.

B, enrichment–abundance plots highlighting only protein classes of interest observed in the trypsin digestion of the large band. Note that the CII and CIV outliers are the mtDNA-encoded proteins Cob (CIII) and Cox1 and Cox2 (CIV); as all three are within the thresholds in the trypsin + chymotrypsin digestion of the large band (see also Table S2), their levels in the trypsin digestion are an artifact, likely due to their physicochemical properties.

Table 3

MS identifications of 15 putative new Complex I subunits of D. papillatum

| Dp IDa | Subunitb | Domain(s)c | Length | Predicted mass | MTP cleavage (length)d | Trypsin | Trypsin + chymotrypsin |
|--------|-----------|------------|--------|----------------|------------------------|---------|-----------------------|
|        |           |            |        | AaKd           |                        | Number of peptides | Sequence coverage |
|        |           |            |        |                |                        | Number of peptides | Sequence coverage |
| m.1717 NDUDP1 | PF00970, PF00175 | 322 | 36.3 | + (8)* | 28 | 66.8 | 25 | 50.0 |
| m.11128 NDUDP2 | PF00970, PF00175 | 297 | 32.5 | ? | 29 | 87.5 | 38 | 65.3 |
| m.2267 NDUDP3 | PF02777 | 383 | 43.8 | + (17) | 45 | 82.0 | 37 | 54.3 |
| m.9618 NDUDP4 | PLN2593 | 319 | 36.9 | + (20) | 28 | 68.3 | 22 | 44.5 |
| m.5632 NDUDP5 | 2 TMHs | 157 | 18.8 | ? | 16 | 56.1 | 10 | 35.0 |
| m.7243 NDUDP6 | TMH | 243 | 28.5 | + (24) | 19 | 56.0 | 15 | 48.1 |
| m.20688 NDUDP7 | TMH | 284 | 32.8 | ? | 20 | 52.1 | 18 | 46.5 |
| m.30841 NDUDP8 | TMH | 161 | 18.8 | ? | 11 | 47.2 | 9 | 36.6 |
| m.16247 NDUDP9 | TMH | 108 | 12.0 | ? | 7 | 57.4 | 8 | 56.5 |
| m.10625 NDUDP10 | TMH | 269 | 31.3 | ? | 25 | 82.5 | 15 | 51.6 |
| m.19023 NDUDP11 | TMH | 406 | 46.0 | ? | 51 | 81.8 | 31 | 50.5 |
| m.38012 NDUDP12 | TMH | 176 | 20.4 | + (9) | 30 | 92.6 | 18 | 67.0 |
| m.5431 NDUDP13 | 276 | 31.9 | + (13)* | 25 | 69.7 | 16 | 36.0 |
| m.15440 NDUDP14 | 222 | 26.5 | + (14)* | 13 | 54.1 | 13 | 37.4 |
| m.15811 NDUDP15 | 101 | 11.3 | ? | 4 | 29.7 | 9 | 48.5 |

a Identifier of D. papillatum proteins.
b Nomenclature based on that of nucleus-encoded CI subunits from other organisms (e.g. T. brucei (NDUTR)).
c Computationally predicted TMHs or protein domains (PF00970: FAD binding; PF00175: NAD binding; PF02777: Fe/Mn-superoxide dismutase; PLN2593: adrenodoxin-like).
d Experimental evidence for MTP cleavage and MTP length. For details, see Table 2.
Therefore, the loss/truncation of both helices in diplonemids may be the result of protein co-evolution. It remains to be determined how or if at all in diplonemids this intersubunit interaction has been replaced (see also Supporting Discussion).

Functional consequences of divergent mitochondrion-encoded subunits

The majority of diplonemid Y proteins are components of the Pp module (i.e. the membrane part of CI close to the ubiquinone-binding pocket). Nad3 (Y1), Nad4L (Y6), and Nad6 (Y5), together with Nad1, are thought to constitute the proton E-channel (10, 55), whereas Nad2 (Y3) is an antiporter-type subunit forming a proton channel on its own.

In Nad2 (Y3) of diplonemids, TMH5 contains a mid-helix tyrosine (Tyr-TMH5) instead of the canonical Lys-TMH5 (TMH8) that has been proposed to play a crucial role in the hydration of the proton channel (56). (The helix numbering refers to the mammalian Nad2 protein with bacterial/canonical numbering in parentheses.) Apparently, any polar residue is suited for this hydration, allowing the substitution of the conserved lysine by tyrosine, not only in diplonemids, but also in kinetoplastids and ciliates (Fig. 2A). Further, diplonemid Nad2 lacks the highly conserved mid-helix Glu-TMH2 (TMH5), a residue that is considered to transmit the electrostatic signal during the proton-pumping cycle (see Di Luca et al. (56) and references therein). We propose that a diplonemid-specific and invariable Asp in TMH3 functionally replaced the canonical Glu-TMH2, because the former occupies a position that is just opposite of the latter in the 3D structure model of Nad2 (Fig. S7).

Conventional mitochondrial Nad3 is characterized by the quasi-universally conserved motif (Y/F)ECGF in the disordered loop that is located between TMH1 and TMH2, whereas in diplonemids, the corresponding motif is Y(D/E)AG(I/V) (Fig. 2B). The central Cys residue (which, in mitochondria, is only rarely substituted by Ser as, for example, in many bacteria), has been implicated in the switching of Complex I between the active and de-active state and was proposed to be a site of redox regulation (57, 58). CI of diplonemids might have lost this capacity, or, alternatively, employs a different redox sensor.

Diplonemid Nad6 (Y5), the least conserved of all Y proteins, was recognized by a Nad6-like distribution of polar and apolar residues and a segment in TMH3 that usually contains a mid-helix Tyr. This residue is considered to be involved in the proton E-channel hydration (10, 55) or to participate in the conformational changes of CI during its catalytic cycle (42). In D. papillatum and F. neradi, a Phe residue occupies this position. The replacement of the polar aromatic Tyr by the nonpolar Phe likely does not affect the hydration of the E-channel as a whole, given that most other functionally important residues of this channel (4) are well-conserved across diplonemids. Another possibility is that a patch of hydrophilic residues occurring in the preceding helix of diplonemid Nad6 compensates for the absence of Tyr-TMH3 (Fig. S2). Alternatively, it may be the aromatic character of the residue that is more critical due to its role in conformational changes (42).

Composition of Complex I from D. papillatum

As compiled in Table 2, CI from D. papillatum (and by inference also that of the other examined diplonemids) consists of 36 assigned components: the universal set of 14 core subunits, 17 broadly distributed eukaryotic accessory subunits, four accessory subunits apparently restricted to Euglenozoa and Heterolobosea, and one other accessory subunit thus far identified only in CI of Trypanosoma (i.e. kinetoplastid + diplomimid–specific; Fig. 4). In addition, we identified a set of 15 proteins that are putative diplonemid-specific CI components (Table 3). Some of these novel proteins may functionally substitute for accessory subunits missing from Diplonema but otherwise broadly conserved, notably those that are transmembrane proteins of the Pp and Pd modules of CI (Fig. 4 and Table S3).

Given these 51 proteins, the predicted mass of CI from Diplonema is ~1.6 MDa, which is similar to that from other euglenozoans, but much larger than that from animals, plants, and fungi (Table 4). The latter taxa not only contain fewer lineage-specific components (Fig. 4), but many of their conventional subunits are much shorter (Table 5).

Several proteins associated with diplomemid CI are worth discussing in more detail. The first two are m.3369 (CA) and m.2734 (CAL; Table 2), members of the yCA family. In plants

Table 4

Complex I masses across eukaryotes

| Organism                | Number of subunits | Complex I mass |
|-------------------------|--------------------|----------------|
|                         |                    | Sum of subunit sizes | Sizes of BN-PAGE bands | References |
|                         |                    | MDa              | ND              |            |
| Box taurus              | 44                 | 0.96            | 1 ( + 1.5 + 1.7)* | 17          |
| Komagataea (Pichia) pastoris | 41               | 0.97            | ND*             | 19          |
| Yarrowia lipolytica     | 40                 | 0.95            | ~1              | 102         |
| Acanthamoeba castellani | 43                 | 0.94            | ~1              | 22          |
| Arabidopsis thaliana    | 44–48              | 1.1             | 1.1 ( + 1.5)*   | 23, 93      |
| Chlamydomonas reinhardtii | 43               | 0.97            | ~1              | 17          |
| Trypanosoma brucei      | >55                | 2.6             | >2              | 20, 47      |
| Euglena gracilis        | >33                | ND              | ~1.5            | 60          |
| Diplonema papillatum    | 36–51              | 1.2, 1.6*       | ~1.2 + ~1.5 + ~2.4* | This work |

* Supercomplex I,II, and/or SCI,IIIIV, (respirasome) displaying NADH-dehydrogenase activity, which have been analyzed by MS and/or structural analyses. These have been observed in digitonin-solubilized mitochondria of, for example, Arabidopsis (66, 93) and various mammals (103, 104).
* ND, not determined.
* Inferred size based on the 36 assigned conventional subunits versus these 36 plus 15 diplonemid-specific components.
* Average size estimates.
Respiratory chain Complex I of diplonemids

Table 5
Complex I subunits from B. taurus and Diplonema having the most pronounced size differences

| Subunit     | Diplonema | Bovine* | Δ | Size difference |
|-------------|-----------|---------|---|-----------------|
| NDUFS1      | 30.5°     | 77.0    | -46.5 | -60.4°*          |
| NDUFS5      | 19.1°     | 12.5°   | 6.6  | 52.6            |
| NDUFS6      | 23.1°     | 10.5°   | 12.6 | 120.0           |
| NDUFA3      | 21.2°     | 11°     | 10.2 | 92.8            |
| NDUFA5      | 43.1°     | 13.2°   | 29.9 | 226.3           |
| NDUFA6      | 68.2°     | 15°     | 53.2 | 354.9           |
| NDUFA8      | 28°       | 20°     | 8.0  | 40.2            |
| NDUFA9      | 49.8°     | 39.1°   | 10.7 | 37.5            |
| NDUFA12     | 40.3°     | 17.1°   | 23.2 | 135.9           |
| NDUFA13     | 24.6°     | 16.6°   | 8.0  | 48.4            |
| NDUFB7      | 29.3°     | 16.5°   | 12.8 | 78.0            |
| NDUFB9      | 33.1°     | 21.7°   | 11.4 | 52.5            |

* Size of the mature protein. Asterisks indicate subunits known to lack a cleavable MTP (17).
* Experimentally confirmed MTP subtracted.
* NDUF51 of all euglenozoans is half the size compared with the counterpart from other eukaryotes (20, 60).
* No evidence for cleaved MTP.

(but not opisthokonts (21, 24)), these proteins are important and integral CI components (28, 59), and they have also been detected in this complex of algae, kinetoplastids, euglenids, and amoebozoans (Fig. 4) (18, 20, 22, 60). In plants, γCA5s have been proposed to interconvert CO2, water, and their bicarbonate ion and to play a role in the transport of carbon dioxide and bicarbonate across membranes. However, a specific functional connection between this enzyme class and CI is unclear. Moreover, in many eukaryotes, most of the residues proposed to be essential for the γCA catalysis have been substituted (24); this is also the case for the two diplonemid proteins. While in Diplonema CI bands we detected two γCA family members, the nuclear genome encodes in total five paralogs, each with a predicted mitochondrial targeting peptide (data not shown). As suggested for the five γCA and γCA-like proteins in Arabidopsis (26), the paralogs from Diplonema may form heterooligomers of different composition under various growth conditions or life cycle stages.

Second, the NDUFAB1 homolog (m.4241; mitochondrial acyl-carrier protein) of Diplonema is probably a multifunctional protein with copies occurring in several different protein complexes, as observed in other eukaryotes. Specifically, the counterpart in yeast and humans not only serves as a structural scaffold in CI, where it binds the LYR family subunits NDUFA6 and NDUFB9 (61–63), but also is part of the assembly machinery of other respiratory complexes and further acts as the multifunctional mitochondrial acyl carrier protein in mitochondrial fatty acid synthesis (64). The multiple locations of NDUFAB1 could explain the here-reported moderate enrichment of this protein in Diplonema CI preparations compared with whole mitochondrial lysate (Fig. 5).

NDUTB2/3 and NDUTB17, which occur in CI only in trypanosomes (20) and diplonemids (Table 2 and Fig. 4), have significant sequence similarity to 2-enoyl-thioester reductase and acyl-CoA synthetase, respectively. These three integral complex subunits could assure a physical connection between CI and the fatty acid degradation pathway by channeling electrons harvested from NADPH to CI.

Finally, one of the putative diplonemid-specific accessory CI components (m.2267) is an iron-dependent superoxide dismutase homolog (Fe/Mn-SOD-type; Table 3). SOD enzymes transform reactive oxygen species, generated by mitochondrial electron transport, to hydrogen peroxide. Although SOD is present in nearly all mitochondria, its specific localization is not known except in Caenorhabditis elegans, where SOD proteins have been shown to be associated with the respiratory chain supercomplex III:IV (65). Interestingly, CI of Trypanosoma includes two paralogs of an unrelated enzyme with the same catalytic function (Fe-SOD; NDUTB10 and -11) (20).

Organization of Diplonema’s respiratory complexes into a respirasome

The largest BN-PAGE band displaying NADH-dehydrogenase activity that we isolated from D. papillatum (Fig. 3) contained not only subunits of CI, but also of CIII and CIV, present in abundance and enrichment comparable with those of CI (Fig. 5, Fig. S6, and Table S2). This result indicates that the large band represents Diplonema’s respirasome (i.e. the supercomplex III:IV). The supercomplex from Diplonema likely has the identical complex ratio of 1:2:1 as the predominant form in animals and fungi (28), in contrast to the plant supercomplex, which consists only of CI and CIII (SCI, IIIII) (28, 66).

Notably, Diplonema’s SCI:III:IV is unusually stable, even in preparations in which relatively strong detergents have been employed, n-dodecyl-β-d-maltoside (DDM) or Triton X-100 instead of digitonin, the latter being required for supercomplex isolation in most model organisms (e.g. see Schägger and Pfeifer (67); Fig. 3). Therefore, the respirasome of Diplonema is a promising candidate for 3D structure determination. Specifically, it would be interesting to validate the above proposed structural deviations of CI subunits from D. papillatum, in particular the loss of a TMH in both Nad6 and Nad4L, the resulting unconventional helix packing of these components, and the remodeled protein–protein interaction between Nad6 and Nad4L.

Conclusions

Because the bacterial CI seems to work well with the 14 core subunits alone, a debate has ensued about the role of the numerous accessory subunits in the mitochondrial complex. Hypotheses range from catalysis optimization (15) to stability (68) to assembly (63). In fact, the issue of the raison-d’être of supernumerary components applies to all molecular machineries of the eukaryotic cell that are of bacterial origin. The shortcoming of the hypotheses noted above is their presumption that the only evolutionary force is natural selection, ignoring the possibility that random genetic drift (e.g. constructive neutral evolution (69–71)) may equally well lead to complexity.

A connected question is why CI contains enzymes whose functions are unrelated to electron transport and proton pumping. Again, CI is not the only respiratory complex with functionally unconnected protein associates. One among many examples is Complex III, to which the mitochondrial protein peptidase is tethered (e.g. see Mach et al. (72) and references therein). In our view, the question is not necessarily appropriate, because (i) some of the so-called supernumerary proteins
may not be additions to a particular complex but rather to the respirasome, which serves as a physical anchor, and (ii) these proteins may just happen to co-purify with one or the other respiratory complex, depending on the strength of particular protein–protein interactions within the respirasome.

**Experimental procedures**

**Strains and culture conditions**

D. papillatum (ATCC 50162), D. ambulator (ATCC 50223), Diplonema sp. 2 (ATCC 50224; recently designated formally as Flectonema neradi (73)), and R. euleeides (ATCC 50226) were cultivated axenically as described earlier (74, 75).

**Isolation of mitochondria**

Cells were grown until late exponential phase; harvested by centrifugation; resuspended in a buffer containing 1.2 M sorbitol, 20 mM HEPES, pH 7.5, 2.5 mM EDTA, pH 8.0, and 1× qOmplete EDTA-free protease inhibitors (Roche Applied Science); and then lysed in a nitrogen cavitation chamber (Parr Instrument Co.) under 30-bar nitrogen pressure. The cell lysate was ultracentrifuged on a two-step sucrose gradient (36 and 60%), and the fraction enriched in mitochondria was collected from the 36/60% sucrose interface. A detailed protocol is available at https://doi.org/10.17504/protocols.io.fqbmwmw.4

**Preparation of mitochondrial lysates, native PAGE, and in-gel activity staining**

Liquid nitrogen–cooled mitochondrial pellets were pulverized using TissueLyserII (Qiagen) in the lysis buffer (20 mM Tris-HCl, pH 7.5, 40 mM KCl, 3 mM MgCl₂, 2.5 mM DTT, and 1× qOmplete EDTA-Free), and the grindate was stored at −80 °C. To obtain a membrane-enriched fraction, a lysate aliquot corresponding to 800 μg of proteins was centrifuged, and the pellet membranes were solubilized with DDM in the lysis buffer or in the loading buffer (50 mM BisTris, pH 7.0, 50 mM NaCl, 2 mM 6-aminohexanoic acid, 1× qOmplete EDTA-Free). Cleared detergent-solubilized lysates were mixed with Coolmassie Brilliant Blue (CBB) G-250 for BN-PAGE, and the electrophoretic separation of protein complexes was performed according to published protocols (76–78) with minor modifications (see supporting Experimental procedures). Complex I was detected within the polyacrylamide gel by adding NADH and staining with nitro blue tetrazolium, and Complex V by adding ATP and Pb(NO₃)₂, essentially as described (79).

**Sample preparation for MS and protein identification and quantification**

Samples of whole mitochondria for tandem mass spectrometry (MS/MS) were prepared by two approaches (for details, see supporting Experimental procedures). Briefly, in the first approach, thawed mitochondria were heat-denatured in the presence of 2% SDS and concentrated by electrophoresis in a Tris-glycine SDS-polyacrylamide stacking gel. The band was cut out, and proteins were fixed by methanol and acetic acid before submission for MS. In the second approach, the mito-ochondrial grindate was denatured essentially as above, but to remove the detergent, the sample was then diluted to 0.05% SDS in 6 M urea buffered with 100 mM Tris, pH 8.5, and concentrated using an Amicon Ultra 10,000 molecular weight cutoff device, essentially following the FASP protocol (80). For the analyses of protein complexes, samples were electrophoretically separated and then excised from CBB-stained gels.

**Proteomics sample processing and analyses, including enzymatic digestion (trypsin or trypsin and chymotrypsin), MS, and peptide searches using Mascot in Proteome Discoverer (Matrix Science), were outsourced to the Proteomics Discovery Platform at the Institut de Recherches Cliniques de Montréal (IRCM) and to the Center for Advanced Proteomics Analyses at the Institut de Recherche en Immunologie et en Cancérologie (IRIC) in Montreal. Spectra were also searched in-house with MaxQuant version 1.6.1.0 following published protocols (81, 82). We used a custom database of D. papillatum proteins (based on the ORFs predicted from mitochondrial and nuclear transcripts). Depending on the particular protein sample (see Table S1), either trypsin or trypsin and chymotrypsin were specified as the digestion enzyme(s), allowing up to two and three missed cleavage sites, respectively.

Proteins were quantified by calculating PAI and iBAQ, essentially as devised by Rappsilber et al. (46) and Schwahnhäuser et al. (45), respectively. To identify proteins enriched in BN-PAGE bands displaying NADH-dehydrogenase activity, we compared the normalized protein abundances in bands with those in mitochondrial lysates. For further details on protein identification and quantification, see supporting Experimental procedures.

**Protein sequence analyses**

Assignment of Y proteins was performed by searching for their homologs among the mtDNA-encoded CI proteins, previously unidentified in diplonemids, from a wide variety of eukaryotes. Published protein sequences were first clustered using CD-HIT (83), and from among the cluster representatives, ~100 reliable sequences per protein were manually selected. For each protein, selected RefSeq sequences were aligned using MUSCLE (84) and MAFFT (85), and HHsuite (86) was then used to generate profile HMMs and to compare each profile with profile HMMs of Y proteins. The best scoring hit for each Y protein was selected for multiple-sequence alignment. For additional assignment criteria used, see supporting Results and supporting Experimental procedures. A phylogenetic tree was constructed with PhyloBayes version 4.1c (87), essentially as described previously (38).

Transmembrane helices were predicted using TMHMM2.0 (webservice at http://www.cbs.dtu.dk/services/TMHMM/) and PolyPhobius (http://phobius.sbc-su.se/poly.html) (88). Protein 3D models were built based on alignments to their mitochondrial homologs with known 3D structures downloaded from the Protein Data Bank in Europe using Phyre2 (89) and SWISS-MODEL (90). Resulting models were visually examined in UCSF Chimera (91).

For the identification of nucleus-encoded subunits, the custom database of nucleus-encoded D. papillatum proteins was searched by BLAST (92) with a collection of previously vali-
Respiratory chain Complex I of diplomemens

dated CI components from various organisms (18–20, 22, 23, 44, 60, 93, 94). When no homolog was identified by BLAST, a Pfam (95) or an in-house—generated profile HMM was used to search the Diplomena collection with HMMPR (96).

All proteins detected as abundant and enriched by MS of BN-PAGE band samples were searched against GenBank™RefSeq proteins. If no annotated homolog was found, domain composition was analyzed by SMART (97), Pfam (95), and NCBI CDD (98). Mitochondrial protein-targeting peptides were predicted by MitoProtII (99), TargetP (100), and MitoFates (101). For more details on sequence analyses, see supporting Experimental procedures.

Author contributions—M. V. and G. B. conceptualization; M. V. data curation; M. V. and A. L.-K. formal analysis; M. V., A. L.-K., and M. W. G. investigation; M. V. and A. L.-K. methodology; M. V. and G. B. writing—original draft; M. V. and G. B. writing-review and editing; M. B. supervision; M. G. funding acquisition.

Acknowledgments—We thank M. Aoulad Aissa for excellent technical assistance and S. Truche for the gift of porcine heart (both from Université de Montréal, Montreal, Canada). We also thank R. Salavati and E. Shoubridge (McGill University, Montreal, Canada) for the gift of protein complex markers, as well as M. Cool (Université de Montréal), I. Škodová-Sveráková, and A. Horváth (Comenius University, Bratislava, Slovakia) for advice and discussions regarding in-gel assays.

References

1. Falkowski, P. G., Fenchel, T., and Delong, E. F. (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320, 1034–1039 CrossRef Medline
2. Sano, A. A. (2015) A giant molecular proton pump: structure and mechanism of respiratory complex I. Nat. Rev. Mol. Cell Biol. 16, 375–388 CrossRef Medline
3. Milenkovic, D., Blaza, J. N., Larsson, N.-G., and Hirst, J. (2017) The enigma of the respiratory chain supercomplex. Cell Metab. 25, 765–776 CrossRef Medline
4. Letts, J. A., and Sazanov, L. A. (2017) Clarifying the supercomplex: the higher—order organization of the mitochondrial electron transport chain. Nat. Struct. Mol. Biol. 24, 800–808 CrossRef Medline
5. Junge, W., and Nelson, N. (2015) ATP synthase. Annu. Rev. Biochem. 84, 631–657 CrossRef Medline
6. Xia, D., Esler, L., Tang, W.-K., Zhou, F., Zhou, Y., Yu, L., and Yu, C.-A. (2013) Structural analysis of cytochrome-bc1 complexes: implications to the mechanism of function. Biochim. Biophys. Acta 1827, 1278–1294 CrossRef Medline
7. Rich, P. R. (2017) Mitochondrial cytochrome c oxidase: catalysis, coupling and controversies. Biochem. Soc. Trans. 45, 813–829 CrossRef Medline
8. Baradaran, R., Berrisford, J. M., Minhas, G. S., and Sazanov, L. A. (2013) Crystal structure of the entire respiratory complex I. Nature 494, 443–448 CrossRef Medline
9. Verkhovsky, M., and Bloch, D. A. (2013) Energy—converting respiratory Complex I: on the way to the molecular mechanism of the proton pump. Int. J. Biochem. Cell Biol. 45, 491–511 CrossRef Medline
10. Hirst, J. (2013) Mitochondrial complex I. Annu. Rev. Biochem. 82, 551–575 CrossRef Medline
11. Lang, B. F., Gray, M. W., and Burger, G. (1999) Mitochondrial genome evolution and the origin of eukaryotes. Annu. Rev. Genet. 33, 351–397 CrossRef Medline
12. Burger, G., Gray, M. W., Forget, L., and Lang, B. F. (2013) Strikingly bacteria—like and gene—rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. 5, 418–438 CrossRef Medline
13. Verkhovskaya, M., and Bloch, D. A. (2013) Energy—converting respiration and controversies. J. Biol. Chem. 288, 16043–16056 CrossRef Medline
14. Xia, D., Esser, L., Tang, W.-K., Zhou, F., Zhou, Y., Yu, L., and Yu, C.-A. (2013) Structural analysis of cytochrome bc complex: implications to the mechanism of function. Cell Metab. 18, 2027–2037 CrossRef Medline
15. Meyer, E. H. (2012) Proteomic investigations of complex I composition: how to define a subunit? Front. Plant Sci. 3, 106 Medline
16. Cardol, P., Vanrobaeys, F., Devreebe, B., Van Beeumen, J., Matagne, R. F., and Remacle, C. (2004) Higher plant—like subunit composition of mitochondrial complex I from Chlamydomonas reinhardtii: 31 conserved components among eukaryotes. Biochim. Biophys. Acta 1658, 212–224 CrossRef Medline
17. Bridges, H. R., Fearnley, I. M., and Hirst, J. (2010) The subunit composition of mitochondrial NADH:ubiquinone oxidoreductase (complex I) from Pichia pastoris. Mol. Cell. Proteomics 9, 2318–2326 CrossRef Medline
18. Cardol, P., Vanrobaeys, F., Devreebe, B., Van Beeumen, J., Matagne, R. F., and Remacle, C. (2004) Higher plant—like subunit composition of mitochondrial complex I from Chlamydomonas reinhardtii: 31 conserved components among eukaryotes. Biochim. Biophys. Acta 1658, 212–224 CrossRef Medline
19. Cardol, P. (2011) Mitochondrial NADH:ubiquinone oxidoreductase (complex I) in eukaryotes: a highly conserved subunit composition highlighted by mining of protein databases. Biochim. Biophys. Acta 1807, 1390–1397 CrossRef Medline
20. Gavrulyk, R. M. R., Chisholm, K. A., Pinto, D. M., and Gray, M. W. (2012) Composition of the mitochondrial electron transport chain in Acanthamoeba castellanii: Structural and evolutionary insights. 1817, 2027–2037 CrossRef Medline
21. Gavrulyk, R. M. R., and Gray, M. W. (2010) Evidence for an early evolutionary emergence of γ-type carbonic anhydrases as components of mitochondrial respiratory complex I. BMC Evol. Biol. 10, 176 CrossRef Medline
22. Meyer, E. H., Solheim, C., Tanz, S. K., Bonnard, G., and Millar, A. H. (2011) Insights into the composition and assembly of the membrane arm of plant complex I through analysis of subcomplexes in Arabidopsis mutant lines. J. Biol. Chem. 286, 26081–26092 CrossRef Medline
23. Córdoba, J. P., Marchetti, F., Soto, D., Martin, M. V., Pagnussat, G. C., and Zabaleta, E. (2016) The CA domain of the respiratory complex I is required for normal embryogenesis in Arabidopsis thaliana. J. Exp. Bot. 67, 1589–1603 CrossRef Medline
24. Fromm, S., Braun, H.-P., and Peterhansel, C. (2016) Mitochondrial γ carbonic anhydrases are required for complex I assembly and plant re—productive development. New Phytol. 211, 194–207 CrossRef Medline
25. Davies, K. M., Blum, T. B., and Kühlbrandt, W. (2018) Conserved in situ arrangement of complex I and II2 in mitochondrial respiratory chain supercomplexes of mammals, yeast, and plants. Proc. Natl. Acad. Sci. U.S.A. 115, 3024–3029 CrossRef Medline
26. Burger, G., Lang, B. F., Braun, H.-P., and Marx, S. (2003) The enigmatic mitochondrial ORF ymf39 codes for ATP synthase b chain. Nucleic Acids Res. 31, 2353–2360 CrossRef Medline
27. Knectan, S., and Burger, G. (2008) Unassigned MURF1 of kinetoplastids codes for NADH dehydrogenase subunit 2. BMC Genomics 9, 455 CrossRef Medline
28. de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Remacle, C., Aury, J.-M., Colin, S., Aury, J.-M., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348, 1261605–1261605 CrossRef Medline
Respiratory chain Complex I of diplonemids

73. Tashyrev, D., Prokopchuk, G., Yahuki, A., Kaur, B., Faktorová, D., Votýpka, J., Kusaka, C., Fujikura, K., Shiratori, T., Ishida, K.-I., Horák, A., and Lukeš, J. (2018) Phylogeny and morphology of new diplonemids from Japan. *Protist* 169, 158–179 CrossRef Medline

74. Kiethega, G. N., Turcotte, M., and Burger, G. (2011) Evolutionarily conserved *cox1* trans-splicing without cis-motifs. *Mol. Biol. Evol.* 28, 2425–2428 CrossRef Medline

75. Valach, M., Moreira, S., Kiethega, G. N., and Burger, G. (2014) Trans-splicing and RNA editing of LSU rRNA in *Diplonema* mitochondria. *Nucleic Acids Res.* 42, 2660–2672 CrossRef Medline

76. Schägger, H., Cramer, W. A., and von Jagow, G. (1994) Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal. Biochem.* 217, 220–230 CrossRef Medline

77. Wittig, I., Braun, H.-P., and Schägger, H. (2006) Blue native PAGE. *Nat. Protoc.* 1, 418–428 CrossRef Medline

78. Wittig, I., Karas, M., and Schägger, H. (2007) High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol. Cell. Proteomics* 6, 1215–1225 CrossRef Medline

79. Zerbetto, E., Vergani, L., and Dabbeni-Sala, F. (1997) Quantification of muscle mitochondrial oxidative phosphorylation enzymes via histochemical staining of blue native polyacrylamide gels. *Electrophoresis* 18, 2059–2064 CrossRef Medline

80. Winiewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362 CrossRef Medline

81. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372 CrossRef Medline

82. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319 CrossRef Medline

83. Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659 CrossRef Medline

84. Edgar, R. C. (2004) *MUSCLE*: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797 CrossRef Medline

85. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780 CrossRef Medline

86. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* 9, 173–175 CrossRef Medline

87. Larillot, N., Lepage, T., and Blanquart, S. (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288 CrossRef Medline

88. Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2005) An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* 21, Suppl. 1, i251–i257 CrossRef Medline

89. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858 CrossRef Medline

90. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kieber, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014) *SWISS-MODEL* : modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258 CrossRef Medline

91. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 CrossRef Medline

92. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402 CrossRef Medline

93. Senkler, J., Senkler, M., Eubel, H., Hildebrandt, T., Lengw ensus, C., Schertl, P., Schwarzländer, M., Wagner, S., Wittig, I., and Braun, H.-P. (2017) The mitochondrial complexome of *Arabidopsis thaliana*. *Plant J.* 89, 1079–1092 CrossRef Medline

94. Smith, D. G. S., Gweryluk, R. M. R., Spencer, D. F., Pearlman, R. E., Sia, K. W. M., and Gray, M. W. (2007) Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. *J. Mol. Biol.* 374, 837–863 CrossRef Medline

95. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285 CrossRef Medline

96. Eddy, S. R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, e1002195 CrossRef Medline

97. Letunic, I., and Bork, P. (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496 CrossRef Medline

98. Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Cui, T., Parkhill, J., Liu, F., Marchler, G. H., Song, J. S., Thanki, N., et al. (2015) CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* 43, D222–D226 CrossRef Medline

99. Claros, M. G., and Vincens, P. (1996) Computational method to predict mitochondrial imported proteins and their targeting sequences. *Eur. J. Biochem.* 241, 779–786 CrossRef Medline

100. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016 CrossRef Medline

101. Fukasawa, Y., Tsuiji, J., Fu, S.-C., Tomii, K., Horton, P., and Imai, K. (2015) MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteomics* 14, 1113–1126 CrossRef Medline

102. Morgner, N., Zickermann, V., Kerscher, S., Wittig, I., Abdurakhmanova, A., Barth, H.-D., Brutschy, B., and Brandt, U. (2008) Subunit mass fingerprint of mitochondrial complex I. *Cell* 138, 1384–1391 CrossRef Medline

103. Wu, M., Gu, J., Guo, R., Huang, Y., and Yang, M. (2016) Structure of mammalian respiratory supercomplex 11III2IV1. *Cell* 167, 1598–1609.e10 CrossRef Medline

104. Letts, J. A., Fiedorczuk, K., and Sazanov, L. A. (2016) The architecture of respiratory supercomplexes. *Nature* 537, 644–648 CrossRef Medline