



Contents lists available at ScienceDirect

Mitochondrion

journal homepage: [www.elsevier.com/locate/mito](http://www.elsevier.com/locate/mito)

## The mitochondrial respiratory chain of the secondary green alga *Euglena gracilis* shares many additional subunits with parasitic Trypanosomatidae

Emilie Perez<sup>a,b</sup>, Marie Lapaille<sup>a</sup>, Hervé Degand<sup>d</sup>, Laura Cilibrasi<sup>a</sup>, Alexa Villavicencio-Queijeiro<sup>e</sup>, Pierre Morsomme<sup>d</sup>, Diego González-Halphen<sup>e</sup>, Mark C. Field<sup>f</sup>, Claire Remacle<sup>a,c</sup>, Denis Baurain<sup>b,c,\*</sup>, Pierre Cardol<sup>a,c,\*\*</sup>

<sup>a</sup> Genetics and Physiology of Microalgae, Department of Life Sciences, University of Liège, B-4000 Liège, Belgium

<sup>b</sup> Eukaryotic Phylogenomics, Department of Life Sciences, University of Liège, B-4000 Liège, Belgium

<sup>c</sup> PhytoSYSTEMS, University of Liège, B-4000 Liège, Belgium

<sup>d</sup> Institut des Sciences de la Vie, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

<sup>e</sup> Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico

<sup>f</sup> Division of Biological Chemistry and Drug Discovery, University of Dundee, Dundee, Scotland DD1 5EH, United Kingdom

### ARTICLE INFO

#### Article history:

Received 9 December 2013

Received in revised form 6 February 2014

Accepted 7 February 2014

Available online xxxx

#### Keywords:

Euglenozoa

Trypanosomatidae

OXPHOS

Proteomics

Sequence database mining

Large-scale phylogenetics

### ABSTRACT

The mitochondrion is an essential organelle for the production of cellular ATP in most eukaryotic cells. It is extensively studied, including in parasitic organisms such as trypanosomes, as a potential therapeutic target. Recently, numerous additional subunits of the respiratory-chain complexes have been described in *Trypanosoma brucei* and *Trypanosoma cruzi*. Since these subunits had apparently no counterparts in other organisms, they were interpreted as potentially associated with the parasitic trypanosome lifestyle. Here we used two complementary approaches to characterise the subunit composition of respiratory complexes in *Euglena gracilis*, a non-parasitic secondary green alga related to trypanosomes. First, we developed a phylogenetic pipeline aimed at mining sequence databases for identifying homologues to known respiratory-complex subunits with high confidence. Second, we used MS/MS proteomics after two-dimensional separation of the respiratory complexes by Blue Native- and SDS-PAGE both to confirm *in silico* predictions and to identify further additional subunits. Altogether, we identified 41 subunits that are restricted to *E. gracilis*, *T. brucei* and *T. cruzi*, along with 48 classical subunits described in other eukaryotes (*i.e.* plants, mammals and fungi). This moreover demonstrates that at least half of the subunits recently reported in *T. brucei* and *T. cruzi* are actually not specific to Trypanosomatidae, but extend at least to other Euglenozoa, and that their origin and function are thus not specifically associated with the parasitic lifestyle. Furthermore, preliminary biochemical analyses suggest that some of these additional subunits underlie the peculiarities of the respiratory chain observed in Euglenozoa.

© 2014 Elsevier B.V. and Mitochondria Research Society. All rights reserved.

### 1. Introduction

The mitochondrion is a eukaryotic organelle acquired through endosymbiosis of an  $\alpha$ -proteobacterium by either an ancestral eukaryote (Dyall et al., 2004; Gray et al., 1999) or an archaeon (Martin and Muller, 1998). Subsequently the vast majority of genes encoding mitochondrial proteins were transferred from the mitochondrial genome to the nuclear genome, and the residual mitochondrial genome

mainly encodes products involved in the synthesis of mitochondrial proteins (tRNA and rRNA) and some subunits of the oxidative phosphorylation complexes (Gabaldon and Huynen, 2005; Saccone et al., 2006). The mitochondrion plays a central role in energy production as the site of the latter stages of respiration, *i.e.*, the Krebs cycle in the matrix and oxidative phosphorylation in the inner membrane. The respiratory chain consists of four classical multi-protein complexes: complex I (CI or NADH:ubiquinone oxidoreductase, EC 1.5.6.3), complex II (CII or succinate:ubiquinone oxidoreductase, EC 1.3.5.1), complex III (CIII or ubiquinol:cytochrome *c* oxidoreductase, EC 1.10.2.2) and complex IV (CIV or cytochrome *c* oxidase, EC 1.9.3.1). Except for CII, these complexes couple electron transfer with translocation of protons from the matrix to the intermembrane space, generating a proton gradient used by a fifth complex (CV or F<sub>1</sub>F<sub>0</sub>-ATP synthase, EC 3.6.3.14) for ATP synthesis. Many organisms however have alternative pathways that transfer

\* Correspondence to: D. Baurain, Bvd du Rectorat, 27, B22, Institute of Botany, Dept of Life Sciences, University of Liège, 4000 Liège, Belgium. Tel.: +32 4 3663864.

\*\* Correspondence to: P. Cardol, Bvd du Rectorat, 27, B22, Institute of Botany, Dept of Life Sciences, University of Liège, 4000 Liège, Belgium. Tel.: +32 4 3663840.

E-mail addresses: [denis.baurain@ulg.ac.be](mailto:denis.baurain@ulg.ac.be) (D. Baurain), [Pierre.cardol@ulg.ac.be](mailto:Pierre.cardol@ulg.ac.be) (P. Cardol).

electrons without concomitant proton translocation, such as type-II NAD(P)H dehydrogenases (from NAD(P)H to ubiquinone) and alternative oxidases (AOX; from ubiquinol to molecular oxygen) (Michalecka et al., 2003; Van Aken et al., 2009).

Phylogenetic relationships amongst eukaryotes are notoriously difficult to resolve and thus highly disputed. In two recent syntheses of the phylogenetic literature, eukaryotes are considered as composed of about 6–9 major lineages, of which well-studied green plants, fungi and animals only correspond to a very small fraction (Adl et al., 2012; Walker et al., 2011). Amongst these major lineages excavates are a putative assemblage of (often heterotrophic) flagellates, proposed on the basis of shared morphological characters (e.g., the ventral feeding groove and associated cytoskeletal structures) and molecular data. Excavates are subdivided into two subgroups, Metamonada and Discoba, the latter consistently recovered in phylogenomic analyses (Hampl et al., 2009; Zhao et al., 2012). Within Metamonada, most lineages lack typical mitochondria and instead possess hydrogenosomes (e.g., *Trichomonas* in parabasalids) or mitosomes (e.g., *Giardia* in diplomonads) (Adl et al., 2012; Walker et al., 2011). In contrast, the three lineages comprising Discoba are mostly mitochondriate: Heterolobosea have a regular mitochondrion with a very gene-rich genome (e.g., *Naegleria gruberi*) (Gray et al., 2004), Jakobids harbour mitochondria retaining one of the highest gene complements known to date (e.g., *Reclinomonas* and *Andalucia*) (Burger et al., 2013; Lang et al., 1997), and, finally, Euglenozoa are a very diverse collection of protists, some of which are parasitic (e.g., Trypanosomatidae in kinetoplastids), others photosynthetic (e.g., euglenophytes in euglenids) or free-living heterotrophs (e.g., diplomemids). Yet, all of them present very complex mitochondria with discoidal cristae (Simpson, 1997).

Amongst the studied Trypanosomatidae are many pathogens, including *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*, respectively causing sleeping sickness, Chagas' disease and leishmaniasis in humans. Extensively studied, these parasites are characterised by a complex life cycle and an energetic metabolism possessing many unique features. On the one hand, an alternative electron transfer chain, composed of a glyceraldehyde-3-phosphate dehydrogenase (G3PD) and an alternative oxidase (TAO or AOX), replaces the classical complexes I–IV during the bloodstream (or mammalian-infective) stage (Chaudhuri et al., 2006). On the other hand, respiratory complexes I–IV and the ATP synthase (complex V) found in the procyclic (or insect-infective) stage of trypanosomes possess many apparently-specific additional subunits (Acestor et al., 2011; Morales et al., 2009; Panigrahi et al., 2008; Zikova et al., 2008, 2009).

*Euglena gracilis* is a secondary green alga, stemming from a secondary endosymbiosis between a prasinophyte green alga and a eukaryotic phagotroph belonging to the euglenids (Gibbs, 1981; Turmel et al., 2009). The *E. gracilis* mitochondrion harbours all of the classical respiratory-chain complexes. However, besides CIII, that is sensitive to antimycin A but insensitive to myxothiazol (Moreno-Sánchez et al., 2000), an antimycin A-resistant alternative pathway with CIII-like activity (including the proton translocation) carries a minor fraction of the electron flux in the presence of antimycin A (Sharpless and Butow, 1970b). *E. gracilis* also has an AOX that is active at pH 6.5 and at temperatures below 20 °C (Castro-Guerrero et al., 2004), as well as a cytochrome *c* oxidase activity partially insensitive to cyanide in the presence of L-lactate (Moreno-Sánchez et al., 2000). Its CIV, meanwhile, has a very unusual subunit composition (Bronstrup and Hachtel, 1989).

In this work, we analysed the subunit composition of the classical respiratory chain complexes (CI–V) in *E. gracilis* by combining targeted proteomics, complete proteome database mining and thorough phylogenetic analyses. We show that beyond the subunits found in all mitochondriate eukaryotes studied so far, *E. gracilis* mitochondria also contain a minimum of 40 subunits previously identified only in parasitic Trypanosomatidae (kinetoplastids). Their presence in this photosynthetic alga belonging to a group of non-parasitic euglenids and, for some also retained in the colourless *Euglena longa* and/or in the free-

living *Diplonema papillatum* (diplomemids), extends their distribution to at least Euglenozoa as a whole. These data also strongly suggest that these additional subunits are not associated with a parasitic lifestyle, but rather have a fundamental role in the Euglenozoa.

## 2. Materials and methods

### 2.1. Strain and growth conditions

The *E. gracilis* strain used in this work (SAG 1224-5/25) was obtained from the University of Göttingen (Sammlung von Algenkulturen, Germany). Cells were grown in the dark (DK) or under low (50  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ ; LL) or medium (200  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ ; ML) light at 25 °C, in liquid mineral Tris-minimum-phosphate (TMP) at pH 7.0 supplemented with a mix of vitamins (biotin 10<sup>-7</sup>%, B12 vitamin 10<sup>-7</sup>% and B1 vitamin 2  $\times$  10<sup>-5</sup>% (w/v)) and eventually with an organic carbon source (acetate 17 mM or 60 mM, ethanol 20 mM or 200 mM) (Harris, 1989; Mego and Farb, 1974; Sharpless and Butow, 1970a). In all experiments, cells in exponential phase (1–2  $\times$  10<sup>6</sup> cells/mL) were harvested by centrifugation for 10 min at 500–1500 g.

### 2.2. Oxygen consumption

Dark respiration rates were measured using a Clark Electrode (Hansatech Instruments, King's Lynn, England) as described in Duby and Matagne (1999). Rotenone (CI, 1–200  $\mu\text{M}$ ), antimycin A (CIII, 0.5–20  $\mu\text{M}$ ), SHAM (AOX, 0.2–2 mM), myxothiazol (CIII, 1–20  $\mu\text{M}$ ), cyanide (CIV, 0.02–2 mM) and oligomycin (CV, 0.5–40  $\mu\text{M}$ ) were added to cells grown in the presence of 60 mM acetate under low light.

### 2.3. Protein extracts

Crude total membrane fractions were prepared according to Remacle et al. (2001). The crude mitochondrial fraction was obtained as described by Moreno-Sánchez and Raya (1987). At the end of this protocol, the supernatant was centrifuged for 10 min at 10,000 g at 4 °C and the pellet containing all organelles was suspended in 250  $\mu\text{L}$  of buffer (Remacle et al., 2001) and loaded onto a mannitol/Percoll gradient (Percoll 250 mM mannitol 13%, 21% and 45% (v/v)). The gradient and sample were centrifuged for 50 min at 40,000 g at 4 °C. Mitochondria were collected at the interface between the 21% and 45% layers and then washed twice by centrifugation for 10 min at 11,000 g at 4 °C. The mitochondrial pellet was finally recovered in 300  $\mu\text{L}$  of buffer. Protein concentrations were determined by the method of Bradford (1976).

To assess the purity of mitochondrial fractions, we first determined that chlorophyll content was below the detection threshold ( $<10^{-3}$   $\mu\text{g}$  chlorophyll per  $\mu\text{g}$  of protein). We also performed a preliminary analysis of the total protein content of mitochondrial fractions by LC-MS/MS analysis. Amongst the ~400 proteins that matched known proteins in GenBank (nr) (Benson et al., 2013) (E-value  $<10^{-10}$ ), >80% corresponded to known mitochondrial components (data not shown).

### 2.4. Enzyme activities

Enzyme activity analyses were performed on crude total membrane fractions (50–150  $\mu\text{g}$  of protein) from cells grown in the presence of 17 mM acetate under low light. CI (rotenone-sensitive NADH: duroquinone oxidoreductase), CII + CIII (succinate:cytochrome *c* oxidoreductase) and CIV (cyanide-sensitive cytochrome *c* oxidase) activities were measured as described in Cardol et al. (2002) and Remacle et al. (2001). CV (oligomycin-sensitive ATP synthase) activity was measured as described in Villavicencio-Queijeiro et al. (2009) based on the method from Pullman et al. (1960). The specific activities were reported as the activity inhibited by maximal concentration of inhibitors used *in vivo* (see Results).

## 2.5. Protein electrophoresis

To conduct Blue Native polyacrylamide gel electrophoresis (BN-PAGE) analyses (Schagger and von Jagow, 1991), protein complexes were first solubilised in the presence of either N-dodecyl- $\beta$ -D-maltoside (0.25–1.5%), digitonin (0.5–3.5%) (w/v), 750 mM epsilon-amino-n-caproic acid, 0.5 mM EDTA and 50 mM Bis-Tris pH 7.0, and centrifuged for 15 min at 15,000 g at 4 °C to remove insoluble matter. 0.5% (w/v) sodium taurodeoxycholate was then added to the supernatant prior to separation by electrophoresis on a 4–10% or 4–12% (w/v) polyacrylamide gradient BN gel.

Complex I activity was detected by incubating the gel in 100 mM HEPES–KOH pH 8.0 or Tris–HCl pH 7.5, containing 0.2 mM NADH and 0.1% (w/v) Nitroretazolium Blue chloride (NBT). ATP synthase activity was detected by incubating the gel in 50 mM HEPES pH 8.0, containing 10 mM ATP and 30 mM CaCl<sub>2</sub>.

BN-gels were also electroblotted according to standard protocols onto polyvinylidene fluoride membranes (Amersham GE Healthcare). Detection was performed using a BM Chemiluminescence Western blotting kit (Roche, Basel, Switzerland) with anti-rabbit peroxidase-conjugated antibodies. We used rabbit sera obtained against *Polytomella* sp. Pringsheim 198.80  $\beta$  subunit (1:200,000).

Coomassie-Blue staining and the second dimensional Tricine-sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) procedure were performed as previously described (Cardol et al., 2004).

## 2.6. Mass spectrometry (MS) analyses

Stained proteins associated with spots or bands of interest were manually excised and analysed by mass spectrometry as described in Lapaille et al. (2010). MS and MS/MS queries were performed either against the GenBank protein database (nr) (Benson et al., 2013) or against a database of conceptually translated *E. gracilis* sequences assembled from contigs of 454 genomic reads and from contigs of public ESTs (see below). Precursor tolerance of 150 ppm for MS spectra and 0.1 Da fragment tolerance for MS/MS spectra were allowed. A charge state of +1 was selected. A single trypsin miscleavage and variable modifications consisting of methionine oxidation and acrylamide-modified cysteine were allowed. For protein direct identification with MASCOT, protein scores greater than 60 were considered as significant ( $P < 0.05$ ).

## 2.7. 454 sequencing of *E. gracilis* genome

*E. gracilis* DNA sequence data were obtained from strain Z1 (kind gift from William Martin, University of Düsseldorf). Cells were maintained in Huttner's medium at ~20 °C in Erlenmeyer flasks, and with ambient illumination. Cells were harvested by centrifugation for 10 min at 1000 g, washed twice with ice cold PBS and snap frozen on dry ice. Pellets were thawed and DNA extracted using a Qiagen DNAeasy extraction kit for total DNA (Qiagen UK, Manchester), and quality controlled using a NanoDrop spectrophotometer and by 1% agarose gel electrophoresis to determine that DNA was of high molecular weight. DNA was sequenced using a Titanium Roche 454 instrument and standard protocols. Approximately 1.7 million reads, corresponding to ~500 million bases that passed quality control were obtained and used in the subsequent analysis. Assembly of reads was performed using Newbler (<http://www.454.com/products/analysis-software/>).

## 2.8. In silico analyses

The initial dataset was composed of the complete protein databases of 21 organisms downloaded from various sources: *Escherichia coli* K-12 MG1655, *Mesorhizobium loti* MAFF303099, *Rickettsia prowazekii* Rp22, *Nostoc* sp. PCC 7120, *Prochlorococcus marinus* MIT9303, *Microcoleus chthonoplastes* sp. PCC 7420, *Komagataella pastoris* (NCBI RefSeq; <ftp://ftp.ncbi.nih.gov>), *Homo sapiens*, *Bos taurus*, *Saccharomyces cerevisiae* (Ensembl 68; <ftp://ftp.ensembl.org>), *Chlamydomonas reinhardtii*, *Volvox carterii*, *Micromonas* sp. RCC299, *N. gruberi*, *Phaeodactylum tricornutum*, *Dictyostelium purpureum* (JGI Genome Portal; <http://genome.jgi.doe.gov>), *T. brucei*, *L. major* (TriTrypDB; <http://tritrypdb.org>), *Arabidopsis thaliana* (TAIR; <ftp://ftp.arabidopsis.org>), *Tetrahymena thermophila* (TGD; <http://www.ciliate.org>) and *Cyanidioschyzon merolae* (University of Tokyo; <http://merolae.biol.s.u-tokyo.ac.jp>).

To these complete proteomes, we added our database of translated *E. gracilis* sequences used in the MS analyses (see above). Briefly, the 23,372 ESTs of *E. gracilis* available at the NCBI (as of May 24, 2012) were contiged with CAP3 (Huang and Madan, 1999) and then translated in the reading frame yielding the longest protein fragment (using BioPerl *longorf.pl* script) (Stajich et al., 2002), whereas ~250-Mbp of 454 reads was assembled with the Velvet *de novo* assembler [version 1.2.03] (Zerbino and Birney, 2008) and then translated in the same way as EST contigs.

OrthoMCL [version 2.0.3] (Li et al., 2003) was used according to the protocol of Fischer et al. (2011) with modified settings: percent-match cutoff = 20 and E-value-exponent cutoff = 1. Five values were tested for the inflation parameter ( $I = 1.1–1.5$ ), but except for the tree in Fig. 3A (see below), the analyses reported in this article are all based on the MCL-1.2 run. Orthologous groups and singletons were annotated by BLASTP [version 2.2.25+] using a list of handpicked reference protein sequences (classical subunits of the complexes of the respiratory chain and those described in trypanosomes), which allowed us to identify the groups corresponding to the respiratory complex subunits. The sequences in each annotated orthologous group were aligned with Clustal Omega [version 1.1.0] (Sievers et al., 2011) and most alignments automatically cleared of the ambiguously aligned positions and of the sequences that were too incomplete to be accurately positioned in the trees (see below for details). The final alignments were then submitted to phylogenetic analysis with PhyML [version 3.0] (Guindon et al., 2010) using the LG + F +  $\Gamma_4$  model (Le and Gascuel, 2008; Yang, 1993). The starting tree for the heuristic search was computed by parsimony and the search included both NNI (nearest-neighbour interchange) and SPR (subtree pruning and regrafting) topological moves. The resulting trees were used to select the monophyletic subgroup of interest whenever there were several paralogues included in the same OrthoMCL orthologous group (i.e., plastidial or cytosolic in addition to mitochondrial sequences). Each selected group or subgroup was then used to build a HMM profile with HMMER [version 3.0] (<http://hmm.janelia.org>). Using a customised version of HaMSTR [version 9] (Ebersberger et al., 2009), these HMM profiles allowed us to search for additional orthologous sequences in our untranslated *E. gracilis* database and in the (largely incomplete) EST databases of *E. longa* and *D. papillatum* available at the NCBI.

In order to analyse the evolution of QCR1, QCR2 and QCRTB2, we considered an additional batch of 47 broadly sampled organisms, for which we downloaded the complete protein databases: *Azospirillum brasilense*, *Batrachochytrium dendrobatidis*, *Rhizopus delemar*, *Acanthamoeba castelanii*, *Polysphondylium pallidum*, *Leishmania infantum*, *Leishmania donovani*, *Leishmania mexicana*, *Leishmania braziliensis*, *T. cruzi*, *Angomonas deanei*, *Strigomonas culicis*, *Trichomonas vaginalis*, *Galdieria sulphuraria*, *Ectocarpus siliculosus*, *Nannochloropsis gaditana*, *Cryptosporidium muris*, *P. marinus*, and *Oxytricha trifallax* (NCBI RefSeq); *Monosiga brevicollis*, *Lottia gigantea*, *Guillardia theta*, *Emiliania huxleyi*, *Aurantiochytrium limacinum*, *Aureococcus anophagefferens*, *Bigelowiella natans*, *Chlorella variabilis*, and *Ostreococcus tauri* (JGI Genome Portal); *Puccinia graminis*, *Ustilago maydis*, *Amphimedon queenslandica*, *Nematostella vectensis*, *Dictyostelium discoideum*, *Entamoeba histolytica*, *Giardia lamblia*, *Hyaloperonospora arabidopsidis*, *Pythium ultimum*, *Phytophthora infestans*, *Thalassiosira pseudonana*, *Plasmodium falciparum*, *Toxoplasma gondii* (Ensembl Genomes 17; <http://ensemblgenomes.org>), *Selaginella moellendorffii*, *Physcomitrella patens*, and *Oryza sativa* (Phytozome 9.0;

<http://www.phytozome.net>); *Chondrus crispus* (ENA; <http://www.ebi.ac.uk/ena>); *Pyropia yezoensis* (NRIFS; [http://nrifs.fra.affrc.go.jp/ResearchCenter/5\\_AG/genomes/nori](http://nrifs.fra.affrc.go.jp/ResearchCenter/5_AG/genomes/nori)); and *Cyanophora paradoxa* (Rutgers University; <http://cyanophora.rutgers.edu/cyanophora>). Finally, we also used the conceptual translation of the transcriptome of *Porphyridium purpureum* downloaded from the NCBI.

The HMM profile of the subgroup QCR1/QCR2/QCRTB2, derived from an orthologous group obtained in the MCL-1.1 run, was used to search for homologous sequences in a combined database of our 68 organisms with HMMER. To select QCR-like sequences, we drew a negative logarithmic plot of HMMER E-values using R (<http://www.R-project.org>) and set the E-value cutoff to  $10^{-107}$ , i.e., just before the first undisputable drop in  $-\log_{10}(\text{E-value})$ . The 178 retained sequences were aligned with MUSCLE [version 3.8.31] (Edgar, 2004), which resulted in a better alignment than Clustal Omega on this particular protein family. We then enriched this alignment with the four partial orthologous sequences from *E. gracilis* and *E. longa* identified by HaMStR.

Before phylogenetic analysis, the latter alignment and the MUSCLE alignment of the MCL-1.1 orthologous group were both filtered to eliminate poorly aligned positions and mostly incomplete sequences using the Bio-MUST-Core software package (D.B., unpublished). Briefly, positions due to insertions in less than 50% of the sequences were discarded. Gblocks 0.91b (Castresana, 2000) was then used with loose parameters to further filter the least reliably aligned positions. Finally, the sequences having more than 90% missing characters with respect to the longest sequence were discarded. The two pre-processed alignments were submitted to phylogenetic inference using PhyML as above, and statistical support was estimated through the analysis of 100 bootstrap pseudo-replicates (Felsenstein, 1985). Tree rooting was done manually in Seaview [version 4.4.2] (Gouy et al., 2010), while ladderisation and taxonomic colouring were automatically conducted using Bio-MUST-Core. Final trees were displayed and arranged in FigTree [version 1.4.0] (<http://tree.bio.ed.ac.uk/software/figtree>). It is worth mentioning that a similar procedure was used to ensure that apparently Euglenozoa-specific subunits were indeed restricted to the latter group.

All *in silico* analyses were carried out on a desktop workstation running Bio-Linux 6 (<http://nebc.nerc.ac.uk/tools/bio-linux>) (Field et al., 2006).

### 3. Results

#### 3.1. *In silico* analysis of the subunit composition of respiratory-chain complexes in Euglenozoa

Additional, putatively taxon-specific, subunits of respiratory-chain complexes often share two attributes that make them difficult to be reliably identified in the genomes of non-model organisms. First, they are poorly conserved [e.g., NDUFC2/B14.5b, NDUFA7/B14.5a, NDUFA3/B9 (Cardol, 2011; Huynen et al., 2009)], which rules out simple BLAST searches and calls for iterative (e.g., PSI-BLAST) or HMM-based (e.g., HMMER) approaches. Second, they are regularly recruited from large multi-gene families with many paralogous groups, the evolution of which cannot be untangled without broadly sampled phylogenetic trees (Gabaldon and Koonin, 2013; Koski and Golding, 2001). For example, in CI, NDUFA11 is a TIM17/22-like protein, CAG9 is a gamma carbonic anhydrase, NDUFA2 features a mitochondrial ribosomal protein L51/S25/CI-B8 domain, NDUFA1 is an acyl-carrier protein, and NDUFB9 belongs to the LYR protein family while in CIII, QCR1/2 both belong to the peptidase family M16. Consequently, when searching for the orthologues of known additional subunits in a target genome, it may be hard to ensure that a given gene, identified with a weak PSI-BLAST hit, actually corresponds to the reference subunit used as the query for the search, and not to a paralogous gene of the same family, potentially fulfilling a different function in a different cell compartment. In this work, we used a carefully designed phylogenetic pipeline to ensure

maximal specificity when identifying orthologous mitochondrial proteins in five Euglenozoa (*E. gracilis*, *E. longa*, *T. brucei*, *L. major* and *D. papillatum*) and in a heterolobosean outgroup (*N. gruberi*) (Fig. 1). This higher reliability is at the expense of lower sensitivity, which means that we probably failed to identify genuine orthologous genes, some of them previously reported, e.g., in *N. gruberi* (Cardol, 2011). However, we argue our approach is sound in the context of a study aimed at expanding the taxonomic distribution of reportedly taxon-specific subunits.

To this end, we compiled an inventory of 99 subunits constituting the traditional complexes of the respiratory chain in model eukaryotes, using *S. cerevisiae*, *H. sapiens* (both Opisthokonta) and *A. thaliana* (Viridiplantae = green plants) as reference organisms. In addition, we also included the 84 apparently specific subunits described in *T. brucei* and/or *T. cruzi* (Acestor et al., 2011; Morales et al., 2009; Panigrahi et al., 2008; Zikova et al., 2008, 2009) (Supplemental Table 1). In the following, TC and TB notations refer to subunits described in *T. cruzi* and *T. brucei*, respectively. We then collected the complete proteomes of six bacteria (*E. coli* K-12 MG1655, *M. loti* MAFF303099, *R. prowazekii* Rp22, *Nostoc* sp. PCC 7120, *P. marinus* MIT9303, *M. chthonoplastes* sp. PCC 7420) and 15 broadly sampled eukaryotes (*H. sapiens*, *B. taurus*, *S. cerevisiae*, *K. pastoris*, *D. purpureum*, *C. merolae*, *C. reinhardtii*, *V. carteri*, *Micromonas* sp., *A. thaliana*, *P. tricornutum*, *T. thermophila*, *N. gruberi*, *T. brucei* and *L. major*). To this database, we added private 454 genomic contigs and public EST contigs of *E. gracilis*, all translated in the frame that yielded the longest predicted protein fragment (see Materials and methods for details).

In order to confidently cluster orthologous respiratory complex subunits in our complete proteomes, we first carried out an all-vs-all comparison of our database sequences using BLASTP. Application of the OrthoMCL pipeline to the resulting BLAST reports allowed us to generate between 35,800 and 49,522 orthologous groups, by varying the inflation parameter (I) from 1.1 to 1.5. These groups only contained orthologous and in-paralogous genes (due to terminal duplications), with out-paralogous genes (due to ancestral duplications) automatically sorted out into different groups. This unbiased global approach spared us the need to choose 99 specific E-value thresholds for separating orthologues from paralogues in each individual case, as done in more conventional database mining studies. For each of these five MCL runs, we annotated the orthologous groups by comparison with our inventory of reference proteins (again using BLASTP). This revealed that the value of “I” leading to the best delineation of the different subunits was 1.2. Amongst the 605 annotated orthologous groups of the MCL-1.2 run, we carefully selected one group for each subunit, generally the one including both the reference protein(s) and the largest sample of species. These 99 groups were then aligned, cleared of ambiguously aligned positions and of incomplete sequences (because short sequences hinder analyses), and finally used to compute phylogenetic trees (Supplemental Figs. 1 and 2; see also Fig. 3 below). In each of these trees, we identified and analysed the subtree corresponding to the mitochondrial paralogue (by opposition to the plastid paralogue) by its proximity to the two  $\alpha$ -proteobacteria and/or by its inclusion of the reference protein(s). However, as the sequences in our *E. gracilis* database had only been translated in a single frame and were sometimes very short, each mitochondrial subgroup was re-analysed with HaMStR (based on HMM profiles and applying the best-reciprocal hit (BRH) orthology criterion) to find additional orthologous sequences in our *E. gracilis* nucleotide database (using six-frame translation) and in the incomplete public EST databases of the secondarily non-photosynthetic *E. longa* and of the free-living diplomonid *D. papillatum*. Altogether, this allowed us to identify a total of 84 mitochondrial subunits in euglenids (*E. gracilis* and *E. longa*), 8 in *D. papillatum* and 118 in kinetoplastids (*T. brucei* and *L. major*). Note that all the subunits specifically described in trypanosomes were also found in *L. major*. This is why, in the following, we refer to them as “apparently kinetoplastid-specific subunits”. Furthermore, 60 subunits were found for the heterolobosean *N. gruberi*, which we selected as outgroup (Fig. 2). Accession numbers and

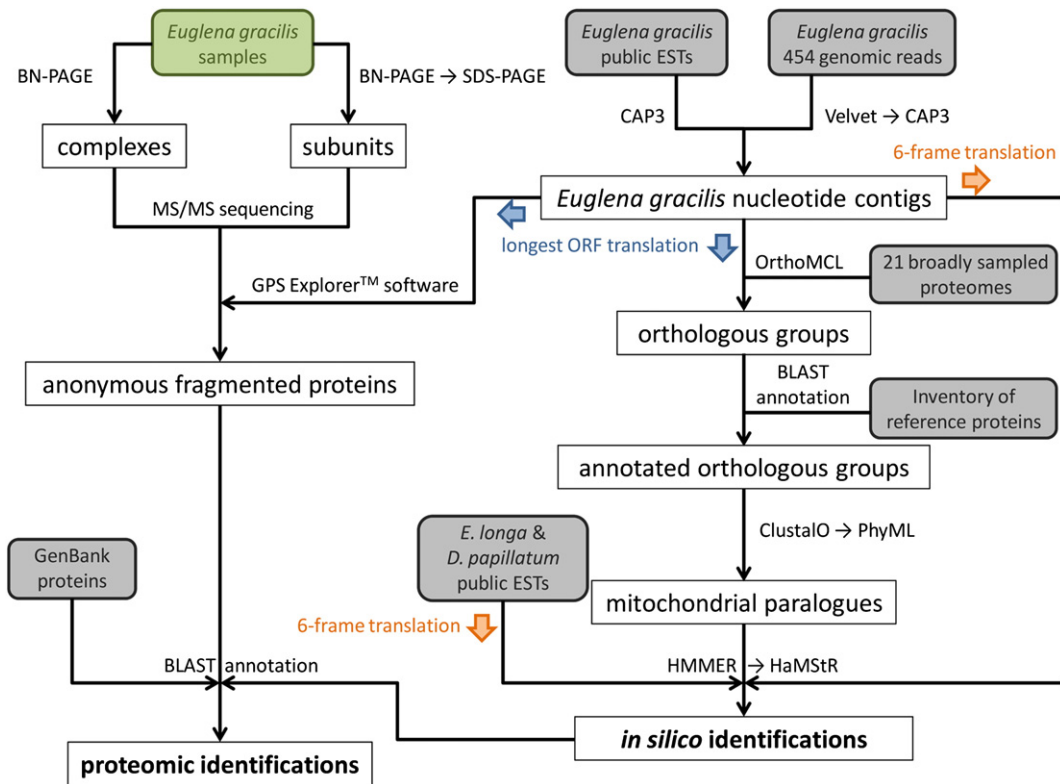


Fig. 1. Scheme of the computational pipeline for *in silico* and proteomic identifications.

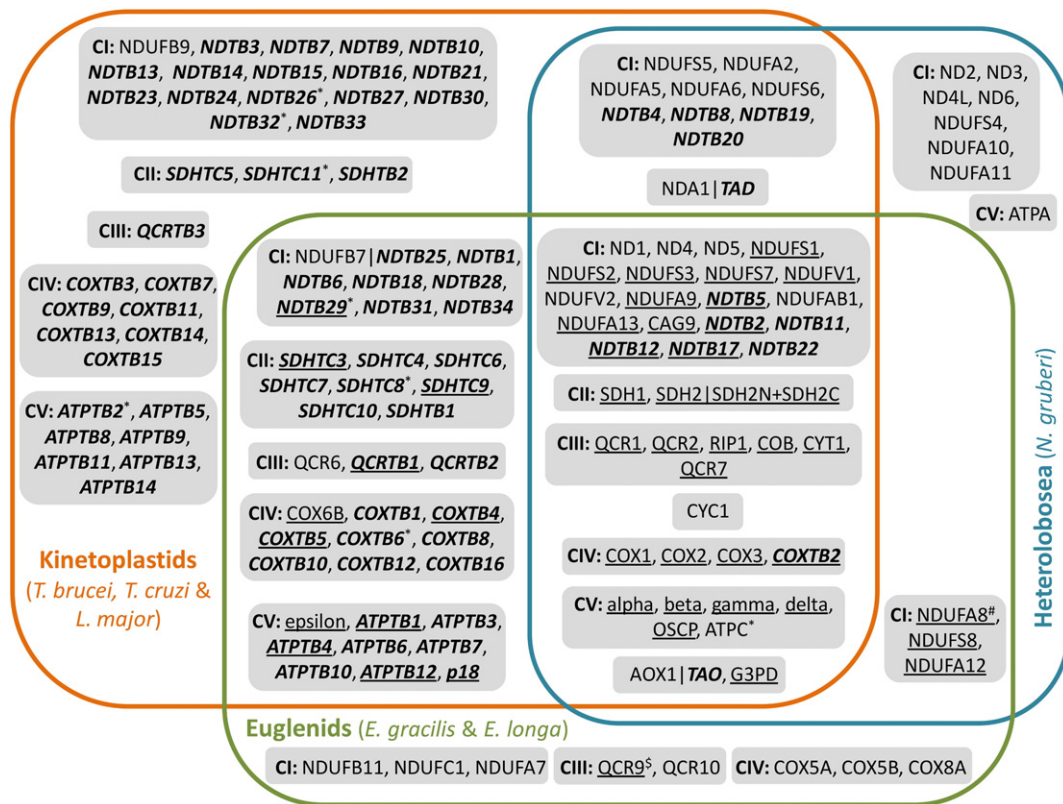


Fig. 2. Subunits of respiratory-chain complexes and proteins of alternative pathways found within euglenids (*E. gracilis* and *E. longa*), kinetoplastids (*T. brucei*, *T. cruzi* and *L. major*) and Heterolobosea (*N. gruberi*). <sup>#,5</sup>At the exception of *Euglena* *NDUFA8* (see text for details and Fig. 6) and *QCR9*, all proteins were identified by bioinformatic analyses (see Supplemental Files 1 and 5). Underlined proteins were found in *E. gracilis* by proteomic analyses (see Supplemental Table 3 and Supplemental File 8). Proteins marked in bold and italics are specific to Euglenozoa. For each sector of the diagram, the proteins are grouped by respiratory complex. Proteins marked with \* were also found in diplomonids (*D. papillatum*; see Supplemental File 6). <sup>5</sup>*Euglena* *QCR9* protein is not homologous to the canonical *QCR9* subunit found in mammals and fungi (see text for details).

identified sequences in excavate representatives are given in Supplemental Files 1–7.

### 3.1.1. Complex I

We found 19 of the 45 conventional subunits of the eukaryotic CI in *E. gracilis*. The majority of these subunits were either core subunits (ND1/4/5, NDUFS1–3/7/8 and NDUFV1/2) that trace back to the  $\alpha$ -proteobacterial ancestor of the mitochondrion or amongst the additional subunits whose primary structure is the best conserved of CI (NDUFA7/9/12/13, NDUFAB1, NDUFB7/11, NDUFC1, and CAG9). The CAG9 group corresponds to gamma carbonic anhydrases, two of which are present in *E. gracilis* (CAG1/2). Most of these 19 proteins were also identified in *N. gruberi*. Out of the 34 additional CI subunits described in *T. brucei* as potentially kinetoplastid-specific (Acestor et al., 2011; Panigrahi et al., 2008), we found 14 subunits in *E. gracilis* (NDTB1/2/5/6/11/12/17/18/22/25/28/29/31/34) (Fig. 2). Three of these subunits were also identified in *D. papillatum* (NDTB26/29/32) and 10 in *N. gruberi* (NDTB2/4/8/11/12/17–20/22). Finally, we concluded that NDTB25 is actually homologous to a conventional subunit (NDUFB7), that NDTB5 is paralogous to a conventional subunit (NDUFA9), and that NDTB2 and NDTB3 are paralogous to each other.

### 3.1.2. Complex II

CII is typically described as a four-subunit complex in mammal/fungal species. In contrast, *T. cruzi* presents a set of 11 subunits, of which only three are identical to the conventional subunits SDH1 and the two fragments of SDH2 (SDH2N and SDH2C) (Gawryluk and Gray, 2009; Morales et al., 2009). Four of the *T. cruzi*-specific subunits (SDHTC3/6/8/9) and two other additional subunits (SDHTB1/2) have been identified in *T. brucei* (Acestor et al., 2011). Similarly, we found the canonical SDH1 and SDH2 subunits in *E. gracilis* (the latter also split in two parts, SDH2N and SDH2C), as well as eight apparently kinetoplastid-specific subunits (SDHTC3/4/6–10 and SDHTB1) (Fig. 2). In addition, SDHTC8 and SDHTC11 were identified in *D. papillatum*, while only SDH1 and SDH2 were found in *N. gruberi*.

### 3.1.3. Complex III

Amongst the 10 classical CIII subunits, eight were found in *E. gracilis* (QCR1, QCR2, RIP1, COB, CYT1, QCR6, QCR7 and QCR10) and six in *N. gruberi* (QCR1, QCR2, RIP1, COB, CYT1 and QCR7) (Fig. 2). In *T. brucei*, three additional apparently specific subunits (QCRTB1, QCRTB2 and QCRTB3) have been described (Acestor et al., 2011), two of which were found in *E. gracilis* (QCRTB1/2). Interestingly, we noticed that in the MCL-1.1 run, QCRTB2 fell in an orthologous group of M16 peptidases also containing QCR1 and QCR2. Thus, to confirm that QCRTB2 was indeed homologous to QCR1/QCR2, we aligned all the sequences of the MCL group and used the resulting alignment for inferring a phylogenetic tree (Fig. 3A). This tree allowed us to separate the sequences belonging to the subgroup QCR1/QCR2/QCRTB2 from the rest of the M16 peptidase family. To determine the evolutionary relationships between QCR1, QCR2 and QCRTB2, we built a HMM profile of this subgroup and used it to search a larger database enriched of 47 broadly sampled eukaryotic proteomes (in addition to the 21 used above; see Materials and methods for a complete list of organisms). QCR-like sequences matching the HMM profile were then aligned, supplemented with four short sequences of *E. gracilis* and *E. longa* identified by HaMStR, and finally submitted to phylogenetic analysis. From the obtained tree (Fig. 3B), we concluded that QCRTB2 was likely the result of a duplication of the QCR1 gene having taken place in the common ancestor of Euglenozoa.

### 3.1.4. Complex IV

Amongst the 13 mammalian CIV subunits (Kadenbach et al., 1983), we identified seven subunits in *E. gracilis* (COX1–3/5A/5B/6B/8A), and only three in *N. gruberi*, all encoded in the mitochondrial genome (Fig. 2). Amongst the 15 apparently kinetoplastid-specific subunits described in *T. brucei* (Zikova et al., 2008), nine were found in *E. gracilis*

(COXTB1/2/4–6/8/10/12/16), one in the partial database of *D. papillatum* (COXTB6) and one in the complete proteome of *N. gruberi* (COXTB2).

### 3.1.5. ATP synthase

Eukaryotic F<sub>1</sub>F<sub>0</sub>-ATP synthases of mammal, fungal and land plant model organisms (Collinson et al., 1994; Heazlewood et al., 2003; Velours and Arselin, 2000) usually consist of ~20 subunits, seven of which were found in *E. gracilis* ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , OSCP, c), one in *D. papillatum* (subunit c) and seven in *N. gruberi* ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , OSCP, a, c). Amongst the 15 apparently specific subunits described in *T. brucei* (Zikova et al., 2009), we identified eight subunits in *E. gracilis* (ATPTB1/3/4/6/7/10/12 and p18) and one in *D. papillatum* (ATPTB2) (Fig. 2).

### 3.1.6. Alternative pathways

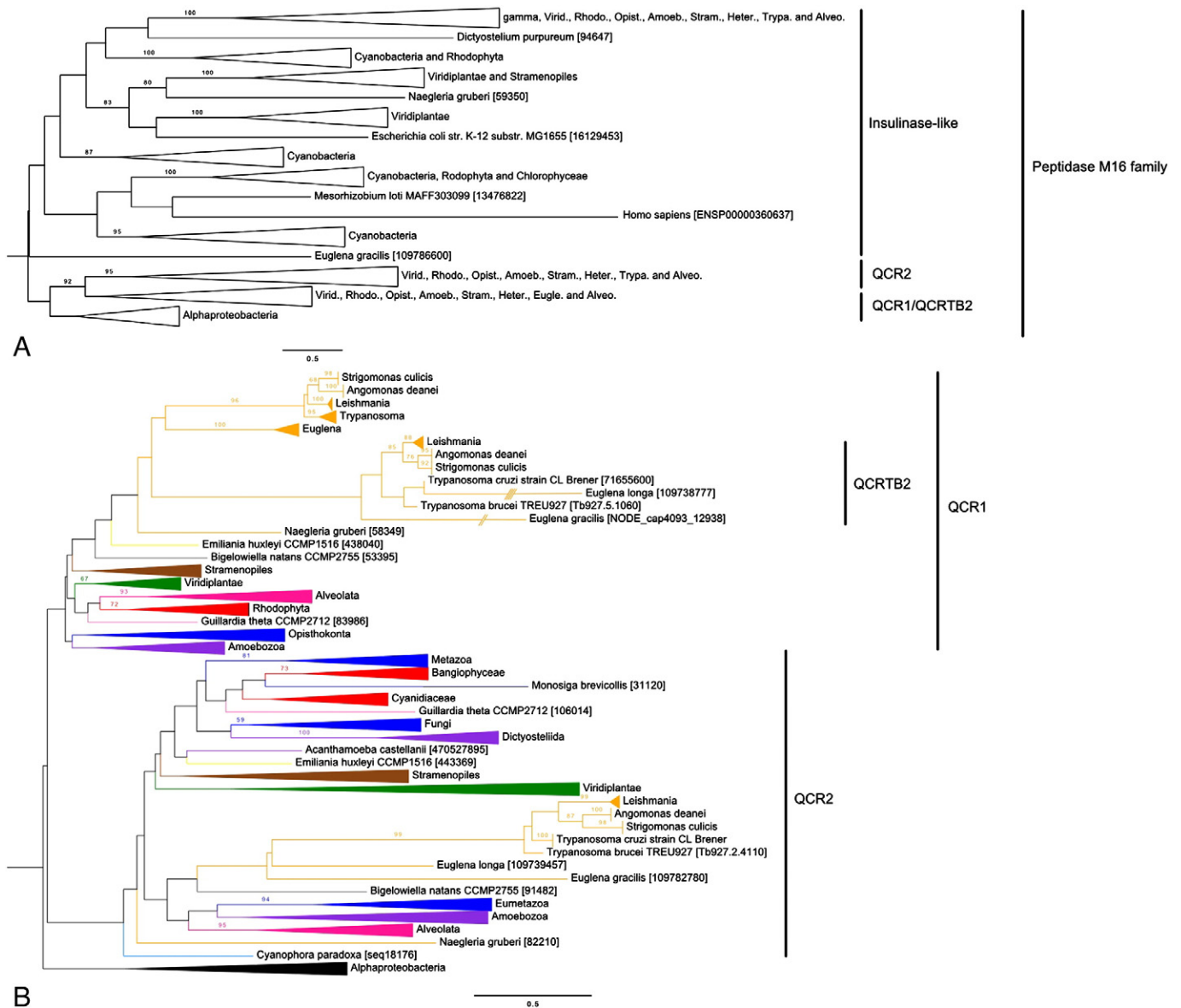
Alternative pathways described in trypanosomes are similar to the alternative pathways described in other protists, green plants and fungi: an alternative oxidase (AOX or TAO), a type-II NADH dehydrogenase (NDA or TAD) (Acestor et al., 2011; Fang and Beattie, 2002) and a glyceraldehyde-3-phosphate dehydrogenase (G3PD), the activity of which compensates for the absence of oxidative phosphorylation during the bloodstream stage (Chaudhuri et al., 2006). Our analyses demonstrated that trypanosomal alternative oxidase and type-II NADH dehydrogenase are homologous to those of the other organisms. While we recovered an AOX and a G3PD in *E. gracilis*, we did not find a NDA (Fig. 2). All three of these enzymes were found in *N. gruberi*.

## 3.2. Proteomic analysis of the subunit composition of the respiratory-chain complexes in *E. gracilis*

We sought to robustly validate our *in silico* predictions by determining whether the proteins identified by data mining were *bona-fide* components of the respiratory-chain complexes in *E. gracilis*. To identify growth conditions where *E. gracilis* accumulated more mitochondria and respiratory complexes, we compared various combinations of light intensity (darkness, low light or medium light) and availability of an exogenous carbon source [absence or presence of acetate (17 mM or 60 mM) or ethanol (20 mM or 200 mM)]. For each culture, dark *in vivo* respiratory rates were then estimated (Supplemental Table 2). Oxygen consumption per cell was greater in the presence of high concentrations of acetate or ethanol, independent of light availability. The maximal value (~1.0–1.5  $\mu\text{mol O}_2 \text{ h}^{-1} 10^{-6}$  cells) was in good accordance with a previous report (Buetow, 1961). Lowest respiration rates (~0.1  $\mu\text{mol O}_2 \text{ h}^{-1} 10^{-6}$  cells) were observed in the absence of exogenous carbon sources.

We next tested the effect of increasing concentrations of classical potent respiratory-complex inhibitors on the *in vivo* respiratory rate. Maximal inhibition ( $I_M$ ; percentage of total respiration) and half maximal inhibitory concentration ( $IC_{50}$ ,  $\mu\text{M}$ ) were determined (Supplemental Fig. 3). Rotenone (CI,  $I_M = 61 \pm 8\%$ ;  $IC_{50} = \sim 20 \mu\text{M}$ ), antimycin A (CIII,  $I_M = 71 \pm 7\%$ ;  $IC_{50} < 1 \mu\text{M}$ ), cyanide (CIV,  $I_M = 89 \pm 4\%$ ;  $IC_{50} = \sim 20 \mu\text{M}$ ) and oligomycin (CV,  $I_M = 62 \pm 5\%$ ;  $IC_{50} = \sim 8 \mu\text{M}$ ) inhibited respiration, suggesting that the corresponding complexes might participate in respiration. In contrast, SHAM and myxothiazol, which classically inhibit AOX and alternative CIII, respectively, did not significantly inhibit respiration. Specific enzyme activities were also measured on crude total membrane fractions using the highest inhibitor concentrations tested *in vivo*. These analyses confirmed that rotenone-sensitive NADH:duroquinone oxidoreductase (CI,  $24 \pm 6 \text{ nmol NADH min}^{-1} \text{ mg}^{-1} \text{ prot}$ ), antimycin A-sensitive succinate:cytochrome c oxidoreductase (CII + III,  $9 \pm 5 \text{ nmol cyt c min}^{-1} \text{ mg}^{-1} \text{ prot}$ ), cyanide-sensitive cytochrome c oxidase (CIV,  $58 \pm 12 \text{ nmol cyt c min}^{-1} \text{ mg}^{-1} \text{ prot}$ ) and oligomycin-sensitive ATP synthase (CV,  $90 \pm 40 \text{ nmol ATP min}^{-1} \text{ mg}^{-1} \text{ prot}$ ) were all present in *E. gracilis*.

To analyse the composition of respiratory-chain complexes, we solubilised membrane complexes in their native form and separated them by BN-PAGE. The effect of increasing concentrations of two mild

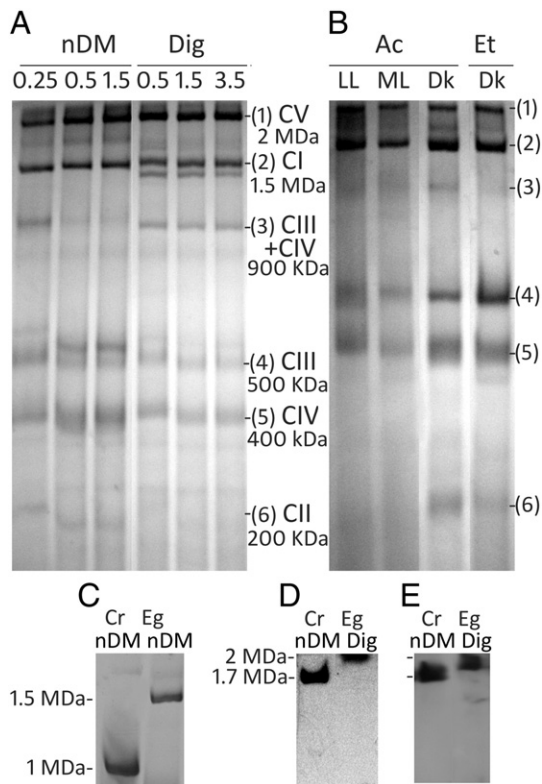


**Fig. 3.** Phylogenetic relationships between QCR1, QCR2 and QCR2B2. A. Maximum-likelihood tree (LG + F +  $\Gamma_4$  model) of the MCL group obtained with  $I = 1.1$  for the QCR1, QCR2 and QCR2B2 reference sequences (169 sequences  $\times$  207 amino acid positions). Gamma:  $\gamma$ -proteobacteria, Virid.: Viridiplantae, Rhodo.: Rhodophyta, Opist.: Opisthokonta, Amoeb.: Amoebozoa, Stram.: Stramenopiles, Heter.: Heterolobosea, Trypa.: Trypanosomatidae, Eugle.: Euglenozoa, and Alveo.: Alveolata. B. Phylogenetic tree showing the paralogy between QCR1, QCR2 and QCR2B2. Black:  $\alpha$ -proteobacteria, green: Viridiplantae, red: Rhodophyta, blue: Opisthokonta, violet: Amoebozoa, brown: Stramenopiles, orange: Excavates, deep pink: Alveolata, yellow: Haptophyceae, grey: Rhizaria, azure: Glaucocystophyceae and pink: Cryptophyta. Branches marked by // and //// were reduced to one half and one quarter of their length, respectively. Bootstrap support values  $\geq 50\%$  are shown above the corresponding nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

non-ionic detergents (n-dodecyl maltoside and digitonin) on protein solubilisation was evaluated. For cells grown in the dark in the presence of high concentrations of acetate, six major bands ranging from ~200 kDa to ~2 MDa could be visualised (Fig. 4A). Bands 2 (~1.5 MDa) and 4 (~500 kDa) sometimes appeared as doublets. While band 3 (~900 kDa) was recovered in higher abundance with digitonin, bands 4, 5 (~400 kDa) and 6 (~200 kDa) were more abundant with n-dodecyl maltoside. We then compared the distribution of these bands in mitochondrial extracts of cells cultivated in the growth conditions yielding the highest respiratory rates (*i.e.*, high concentrations of acetate or ethanol in the dark or in the light). However, light + ethanol conditions were avoided since they led to the well-known phenomenon of cell bleaching due to catabolite repression of chloroplast development in *E. gracilis* (Monroy and Schwartzbach, 1984). Bands 1, 2, 4, and 5 were present in every condition, while all six bands were more abundant in extracts of dark-grown cells (Fig. 4B). Further, in-gel staining aimed at

highlighting ATPase and NADH dehydrogenase activities allowed us to hypothesise that band 1 comprised ATP synthase (CV) and band 2, CI (Fig. 4C, D). Immunodetection on western blot with antibodies against the  $\beta$  subunit confirmed the presence of this CV subunit in band 1 (Fig. 4E).

The subunits of each complex separated by BN-PAGE were then tentatively resolved by SDS-PAGE and stained with Coomassie Blue. The starting material was a purified mitochondrial fraction from cells grown in the dark in the presence of 200 mM ethanol. As shown in Fig. 5A, at least 18, 22, 8, and 10 protein bands could be visualised for bands 1, 2, 4 and 5, respectively. The same pattern of spots was also obtained for a crude membrane extract of cells grown under low light in the presence of 60 mM acetate (data not shown). The most prominent spots were excised out of the gel and analysed by tandem mass spectrometry (MS/MS). The anonymous fragmented proteins of *E. gracilis* obtained in MS analyses were annotated using BLASTP queries against



**Fig. 4.** Coomassie-Blue stained images of respiratory-chain complexes from *Euglena gracilis* separated by BN-PAGE. A. 100  $\mu$ g of protein from crude membrane fraction of cells grown in the dark in the presence of 60 mM acetate was solubilised by treatment with n-dodecyl-maltoside (nDM) or digitonin (Dig). Detergent concentrations are given in % (w/v) above the lane. B. Crude mitochondrial fraction solubilised by treatment with 1% (w/v) nDM. Ac, 60 mM acetate; Et: 200 mM ethanol; DK: dark; LL: low light; ML: medium light. The different complexes are identified by coloration and/or mass spectrometry analysis of their constituents. C, D, and E. Proteins from crude membrane fraction of *Chlamydomonas reinhardtii* (Cr) and *Euglena gracilis* (Eg) cells grown in low light in the presence of 60 mM acetate. 1.5% (w/v) detergent. Upper part of BN gels stained for NADH dehydrogenase activity using NBT as an electron acceptor (C), for ATPase activity (D) or immunoblotted with ATP synthase  $\beta$  subunit antibody (E).

GenBank (nr) and our own protein predictions (Fig. 1). Forty-seven of the 58 analysed protein spots matched a protein in our database and 30 corresponded to an annotated protein in GenBank (Fig. 5A, Supplemental Table 3). In band 1, six of the classical eukaryotic ATP synthase subunits were found ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  and OSCP), in addition to four of the subunits described in *T. brucei* (ATPTB1/4/12 and p18). In band 2, we found 5 canonical CI subunits (NDUFS2/3, NDUFV1, NDUFA12/13, and 2 gamma carbonic anhydrases CAG1/2) and 3 subunits described in *T. brucei* CI (NDTB2/12/17). Two additional proteins, glyceraldehyde-3-phosphate dehydrogenase (G3PD) and chaperone protein HSP40 (DNAJ), were also identified. Band 4 corresponded to CIII with 4 identified subunits (QCR1/7, RIP1, QCRTB1) and band 5 to CIV with 4 known subunits (COX3/6B, COXTB4/5). Band 6 could unfortunately not be resolved into well-defined spots. A partial subunit pattern (7 bands) of band 3 was obtained from a different experiment (Fig. 5B). Interestingly, this corresponded to a mixture of CIII (500 kDa) and CIV (400 kDa), which are probably associated *in vivo* in a supercomplex (900 kDa).

In a second approach, bands 2–6 from BN-PAGE were directly excised and each band was analysed by liquid chromatography-mass spectrometry (LC–MS) to identify possible additional subunits (see Supplemental File 8). Additional subunits could be identified for CI (band 2; NDUFS1/7/8, NDUFA8/9 and NDTB5/29), CIII (bands 3 and 4; CYT1, COB, QCR9), CIV (bands 3 and 5; COX1/2, COXTB2) and CII (band 6; SDH1 and SDH2N/C, SDHTC3 and SDHTC9). Surprisingly, the NDUFA8 subunit was found by proteomics but not by *in silico* analysis. This is explained by the observation that the orthologous group that annotated the protein

fragment was composed only of euglenozoan sequences, for which the first BLASTP hit was the NDUFA8 subunit of *Ciona intestinalis*, but with a high (thus not significant) E-value (0.053). In accordance with this observation, a multiple alignment built from reference NDUFA8 sequences (from *A. thaliana*, *H. sapiens* and *C. intestinalis*) showed that the NDUFA8 subunits of Euglenozoa were poorly conserved, while retaining the majority of the characteristic cysteine residues (Fig. 6). As for the NDUFA8 subunit, the QCR9 subunit was only found by proteomics. This protein was previously identified by N-terminal sequencing as a peculiar component of CIII in *E. gracilis* (Cui et al., 1994). We could not identify any similar sequences in other eukaryotes. Thus, we propose that the *Euglena* QCR9 does not exhibit any conserved domain and does not correspond to the canonical QCR9 subunit described in other species.

A summary of the proteins identified in this study is given in Fig. 2.

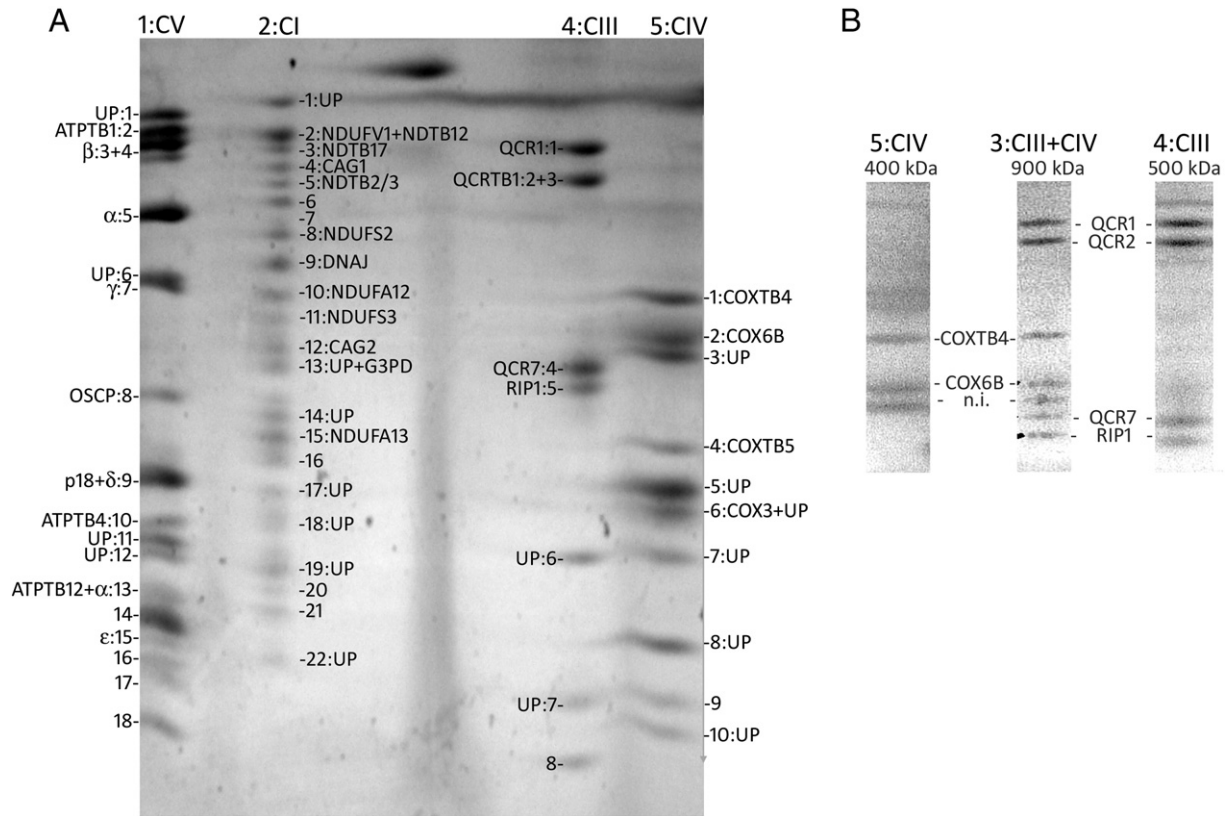
#### 4. Discussion

##### 4.1. Similar subunit composition of respiratory-chain complexes in kinetoplastids and euglenids

Respiratory-chain complexes have a dual genetic origin, most subunits being encoded in the nucleus, while a few subunits of prokaryotic origin remain encoded by the mitochondrial genome (Gabaldon and Huynen, 2005; Saccone et al., 2006). Compared with their prokaryotic counterparts, mitochondrial respiratory-chain complexes have acquired additional subunits, although most of these are not directly involved in the catalytic activity of the enzyme complexes. During the last decade, many studies have highlighted the peculiarities of respiratory-chain complexes in trypanosomes (Acestor et al., 2011; Morales et al., 2009; Panigrahi et al., 2008; Zikova et al., 2008, 2009), in which the majority of the identified subunits are apparently unrelated to those described in green plants, mammals or fungi. Trypanosomes are part of Euglenozoa, a eukaryotic phylum first described three decades ago and grouping kinetoplastids and euglenids (Cavalier-Smith, 1981). At the same time, a secondary symbiotic origin of euglenids chloroplast was proposed (Gibbs, 1981). More recently, Euglenozoa were extended to also include diplomonids (Simpson, 1997). The mitochondrial respiratory chain of *E. gracilis* is well studied (Bronstrup and Hachtel, 1989; Buetow, 1961; Castro-Guerrero et al., 2004, 2005; Moreno-Sánchez et al., 2000; Sharpless and Butow, 1970a,b; Wilson and Danforth, 1958) but the protein composition of its complexes has only been investigated in a couple of studies (Bronstrup and Hachtel, 1989; Cui et al., 1994). In the present work, we addressed this gap by combining a conservative phylogenetic pipeline for *in silico* data mining of complete proteomes with proteomic analyses performed under multiple conditions. It is worth noting that some low molecular mass subunits may probably not be visualised by the Coomassie-Blue staining, and that some hydrophobic subunits may have been lost during the second electrophoresis step. Moreover, such proteins contain only few tryptic cleavage sites, thus generating a limited number of peptides in the mass range suitable for MS analysis. In particular, none of the seven hydrophobic core subunits, which are usually mitochondrially encoded (ND1–6, ND4L), could be detected, and which suggests that despite the multiple and exhaustive analysis presented here that the complete composition of the *E. gracilis* respiratory apparatus remains to be achieved. Nonetheless, at this stage, the respiratory-chain complexes I–V of *E. gracilis* consist of at least 92 different proteins, 79 of which are shared with trypanosomes and 34 of which, according to our analyses, appear to be restricted to Euglenozoa.

From these data, we can draw three main conclusions: (i) the peculiar subunit composition of the respiratory-chain complexes described in trypanosomes originated in the common ancestor that they share with euglenids and, for a few subunits, in the common ancestor of Discoba, since some were also identified in the heterolobosean *N. gruberi*; (ii) these data reinforce the idea that the additional subunits found in trypanosomes are genuine components of respiratory complexes of





**Fig. 5.** Two-dimensional resolution of the mitochondrial protein complexes from *Euglena gracilis*. A. BN gel lane loaded with 500 µg of protein was cut out and placed horizontally for subsequent resolution of the protein complexes into their respective components on Tricine-SDS-PAGE. The main complexes on the first dimension BN-PAGE (Fig. 4A, B) are indicated at the top of the gel. Coomassie Blue-stained image. The numbered spots correspond to polypeptides that were subject to MS/MS analysis. The corresponding sequences are given in Supplemental File 8 and Supplemental Table 3. UP, database match to an unknown protein; n.i., no annotation due to the lack of a significant score in MS analysis. B. Comparison of partial pattern of CIII (band 4), CIV (band 5) and supercomplex III + IV (band 3).

these parasites, and; (iii) contrarily to recent suggestions (Panigrahi et al., 2008), the additional subunits shared with euglenids are unlikely to play any specific role in the parasitic lifestyle of trypanosomes. Thus, besides canonical subunits shared by all eukaryotic groups, respiratory complexes I–V in Euglenozoa also possess specific subunits that either have diverged beyond recognition or have been recruited early in the evolution of the lineage. Two scenarios may account for the recruitment of new subunits: (i) acquisition of xenologous genes by lateral gene transfer, either of

bacterial or eukaryotic origin, or (ii) retargeting of cytosolic proteins to the mitochondria, possibly after duplication of the corresponding genes. In both cases, these novel subunits would have assembled to form a new scaffold around the canonical core of respiratory complexes. From our *in silico* analyses, it is not possible to differentiate between these two models, as we generally failed to identify homologues (neither paralogues or orthologues) for Euglenozoa-specific genes, even after thorough HMM searches against the 68-species database built for

<i>Euglena gracilis</i>	71	PNEQTKKEHEGIFGMGIPENVLFSFGHWYRMSQPLEKAREEYLVCKHNA PKRAA PTECLE	130
<i>Trypanosoma brucei</i>	1	-MDQFTKPLEGIFRDGIPAPVLRFAFAPLYQALPSLQDAVSA SRDCYY--WRANPLKQVE	56
<i>Leishmania major</i>	1	-MEQFVQPLEGIFRDGIPAPVLRFAFAPLFQALPSLQERVEAS RDCYY--WRANPMKCLE	56
<i>Ciona intestinalis</i>	35	---GILKDIKRGDLGVPSHV LKASAFQYAN--ECNEVNKEFMLCREEEM---DPRKCLK	85
<i>Homo sapiens</i>	8	---PTLEELKVDEVKISSAVLKA AAHHYGA--QCDKPNKEFMLGRWEEK---DPRRCL	58
<i>Arabidopsis thaliana</i>	1	---MSSAVDATGNPIPTSAVLTASAKHIGM--RCMPENVAFLCKKKNPD---NPEKCLD	51
		* * *	
<i>Euglena gracilis</i>	131	EAKNMFNLYLHMSEIPFRTPKQOSADYTYCVETQGRHKPGMDYGARSF TSFVRYCLPEQE	190
<i>Trypanosoma brucei</i>	57	EDVQTVTSFMQASEASFR LCPQQSATLLKCHMTEPA-----RALFFCRDEEW	103
<i>Leishmania major</i>	57	EDIDTVTGFMQACEASFR LCPQQSATLLKCHMTEPA-----RAVYFCRDEEW	103
<i>Ciona intestinalis</i>	86	YNIKVSDCAENFFRKMTTACADEIVAFGKCLERDHK-----RSFVYCRDEEQV	132
<i>Homo sapiens</i>	59	EGKLVNKCALDFFRQIKRHC AEPFTEYWTCIDYTGQ-----QLFRHCRKQQA	105
<i>Arabidopsis thaliana</i>	52	KGRDVTRCV LGLLKD LHKQCKQEMDDYVGC MYYYT-----NEFDLCKRKEQE	97
		* * *	
<i>Euglena gracilis</i>	191	AFEECLGKTYGCKFPAPAHGQF---YQRSA-----RFTNLPWDNIYA----	230
<i>Trypanosoma brucei</i>	104	EWRTCLMDQTGIRFWPYANAPIG-APWSNGGQTEDFHLEDRFFYEN-FSWRRRKAAMLAV	161
<i>Leishmania major</i>	104	EWRSCLMDQTGIRFWPYANAPIG-APWSNGGQTEDFHLEDRFFYEN-FSFWRRRGAMLAV	161
<i>Ciona intestinalis</i>	133	KFDRCMFEKLGIDKKYNAVELDQTVKTRDPAPKPNFKLKNYDHPSP LMPDWNRLPKIED	192
<i>Homo sapiens</i>	106	KFDECVLDKLGWVRPDLGELSKG-----	128
<i>Arabidopsis thaliana</i>	98	AFEKVCPLK-----	106

**Fig. 6.** Clustal-Omega alignment between NDUFA8 sequences of *Arabidopsis thaliana*, *Homo sapiens* and *Ciona intestinalis* and the potential NDUFA8 sequences of *Trypanosoma brucei*, *Leishmania major* and *Euglena gracilis*. \*: conserved characteristic cysteines of NDUFA8. Alignment formatted with Jalview [version 2.7] (Waterhouse et al., 2009).

studying the evolution of QCR1/QCR2/QCRTB2, as well as against another database of about 400 complete proteomes representative of prokaryotic diversity (data not shown). The exceptions here may be NDTB2/3/4/8/10/11/17/18/22/29, SDHTB2, QCRTB1/2 and COXTB6/16, all of which possess conserved domains (but not full-length architectures) found in other eukaryotic groups. Thus, whatever their ultimate origin, Euglenozoa-specific genes are likely to have undergone extensive evolutionary divergence.

#### 4.2. Peculiar features of respiratory-chain complexes in *E. gracilis*

Our study also highlights some interesting features at the level of individual respiratory-chain complexes in *E. gracilis*.

##### 4.2.1. Complex I

The approximate molecular mass of CI in *E. gracilis* is ~1.5 MDa. In contrast, in all eukaryotes investigated so far, including *T. brucei* (Acestor et al., 2011), the mammal *B. taurus* (Carroll et al., 2006), the yeast *Yarrowia lipolytica* (Angerer et al., 2011), the green alga *C. reinhardtii* (Cardol et al., 2004) and the amoeba *Acanthamoeba castellanii* (Gawryluk et al., 2012), CI has a molecular mass of ~900–1000 kDa. In green plants, mammals and fungi, CI has however been found in association with dimeric CIII and/or CIV, leading to supercomplexes of >1.5 MDa (Cardol et al., 2008; Dudkina et al., 2005; Schagger and Pfeiffer, 2000). However, in the present study, no evidence for the presence of CIII or CIV subunits could be obtained in the band corresponding to CI in *E. gracilis*. Compared with CI in fungi and mammals, CI in green plants (Cardol et al., 2004; Sunderhaus et al., 2006), amoeba, and presumably other eukaryotes (Gawryluk and Gray, 2010) comprises additional subunits belonging to the  $\gamma$ -carbonic anhydrase (CAG) protein family. Two CAG proteins have also been specifically identified in CI of *E. gracilis*. In green plants, it has been shown that these proteins form a matrix-exposed domain (Sunderhaus et al., 2006), the presence of which does not substantially increase the molecular mass of CI. G3PD and DNAJ enzymes have also been found in association with CI in *E. gracilis*. Interestingly, a DNAJ homologue was recently identified in CI of *T. brucei* (Acestor et al., 2011). DNAJ/HSP40 (heat shock protein 40) primarily stimulates ATPase activity of HSP70 chaperones (Qiu et al., 2006) and these proteins are broadly distributed amongst eukaryotic lineages. Although we cannot completely rule out the possibility that DNAJ/HSP40 and G3PD are non-mitochondrial contaminations of the mitochondrial fraction (see Protein extracts), our study brings evidence of a potential interaction between this cohort of proteins and respiratory CI. We recently proposed that some proteins might use CI as an anchoring point and that these interactions might be rather specific to limited groups or species (Cardol, 2011): acyl carrier protein (NDUFAB1/ACPM) and deoxyribonucleoside kinase-like subunit (NDUFA10/42 kDa) in mammals, galactono-lactone dehydrogenase in land plants (Klodmann et al., 2010), or rhodanese in *Y. lipolytica* (Angerer et al., 2011). Similarly, G3PD and DNAJ could be specifically associated with CI in *E. gracilis*. Thus, the apparent higher molecular mass of *E. gracilis* CI probably does not result from an abnormal molecular mass of the canonical subunits or from association with other respiratory complexes, but might rather be the consequence of interactions with proteins bearing functions that are not directly related to CI activity.

##### 4.2.2. Complex II

CII (succinate dehydrogenase) was found to be more abundant in heterotrophic conditions compared to mixotrophic conditions. This is in good agreement with a previous report (Brown and Preston, 1975). CII usually comprises four subunits (SDH1–4): a flavoprotein (SDH1), an iron–sulphur (Fe–S) protein (SDH2) and two membrane subunits (SDH3/SDH4) providing ligands to heme *b* and a reduction site for ubiquinone. As previously pointed out (Morales et al., 2009), membrane anchor subunits are highly divergent between bacteria, mammals and other eukaryotes, and thus are difficult to identify with standard

BLAST searches (Acestor et al., 2011). Candidates for *T. cruzi* and *T. brucei* SDH3 and SDH4 were proposed based on the presence of quinone/heme-binding motifs, predicted *trans*-membrane domains and similar protein sizes, but a direct comparison with these proteins did not reveal any obvious predicted structural or motif similarity (Acestor et al., 2011; Morales et al., 2009). This further illustrates the difficulty to assess orthology for small hydrophobic membrane proteins [also see Discussion in (Cardol, 2011)]. Additional subunits were also found in trypanosomes (Acestor et al., 2011; Morales et al., 2009), some of them also present in *E. gracilis*. However, recruitment of taxon-specific subunits is not limited to Euglenozoa, since unrelated additional subunits have also been described in *A. thaliana* (Millar et al., 2004).

##### 4.2.3. Complex III

Unlike the four other complexes involved in the electron transfer chain, CIII of trypanosomes has a very small set of additional subunits (QCRTB1, QCRTB2 and QCRTB3), one of which (QCRTB2) is a paralogue of the QCR1 core subunit. In *E. gracilis*, seven of the ten classical CIII subunits were found (QCR1/2/6/7, RIP1, COB, and CYT1), along with two of the three subunits described in trypanosomes (QCRTB1/2). It is tempting to speculate that the switch from antimycin A-sensitive CIII activity to myxothiazol-sensitive CIII activity (the so-called *bc1*-bypass) (Moreno-Sánchez et al., 2000; Sharpless and Butow, 1970b) could be due to the differential expression of core 1 paralogues in *E. gracilis* whose presence would modify the affinity of cytochrome *b* Qo and Qi sites for myxothiazol and antimycin A, respectively.

##### 4.2.4. Complex IV

A preliminary analysis of CIV subunit composition suggested an atypical subunit composition of the enzyme, when compared to mammalian cytochrome *c* oxidase (Bronstrup and Hachtel, 1989). In the present work, the presence of four classical subunits (COX1/2/3/6B) was confirmed, along with at least three subunits described in *T. brucei* (COXTB2/4/5). These additional subunits probably constitute a different scaffold compared to the one described in mammals (Tsukihara et al., 1996); this may explain why *E. gracilis* cytochrome *c* oxidase activity was low with heterologous bovine cytochrome *c550* and 35-fold higher with the homologous *E. gracilis* cytochrome *c558* (Bronstrup and Hachtel, 1989). Euglena CIII and CIV were also found to form a supercomplex of ~900 kDa that is stabilised in the presence of the mild non-ionic detergent digitonin. This association of CIII and CIV has been previously described in *S. cerevisiae* and *B. taurus* (Cruciat et al., 2000; Schagger and Pfeiffer, 2000).

##### 4.2.5. Complex V

ATP synthase of *E. gracilis* also exhibited an unusually large molecular mass (>2 MDa). The dimeric nature of ATP synthase is now largely acknowledged in many organisms, and in mammals and yeasts, the dimer has a molecular mass of ~1.2 MDa and is thought also to be responsible for shaping mitochondrial cristae (Davies et al., 2012; Velours and Arselin, 2000; Walker et al., 1991). Beyond mammals, fungi and flowering plants, the first organism where an unusual structure and subunit composition of the mitochondrial ATP synthase was found was the green alga *C. reinhardtii* and its colourless relative *Polytomella* sp. (~1.6–1.7 MDa) (Dudkina et al., 2006; Vazquez-Acevedo et al., 2006). More recently, unusual subunit composition for the ATP synthase was also reported in the ciliate (alveolates) *T. thermophila* (Balabaskaran Nina et al., 2010), and in *T. brucei* (Zikova et al., 2009). In these species, as in *E. gracilis*, only canonical subunits involved in the  $F_1$  catalytic head ( $\alpha$ ,  $\beta$ , OSCP), the rotor (*a*, *c* ring) and the central axis ( $\delta$ ,  $\epsilon$ ,  $\gamma$ ) have been found. In contrast, subunits involved in the peripheral stator (*b*/ATP4, *d*, *e*, *f*, *g*, etc.) are missing and have been replaced by new sets of subunits that have no counterparts in other lineages. In *E. gracilis*, we identified seven of the 14 new subunits discovered in *T. brucei*, for which no homologues could be identified outside Euglenozoa. This strongly suggests

that, as we earlier proposed in the case of mitochondrial ATP synthase from chlorophycean green algae (see Discussion in Lapaille et al., 2010), the recruitment of new subunits might be concomitant to the loss of mitochondrial genes for ATP synthase proteins (including subunit b/ATP4 gene) in the course of mitochondrial gene relocation into the nucleus.

In conclusion, our study of the mitochondrial respiratory chain in *E. gracilis* provides evidence that the additional subunits of the different complexes described in trypanosomes are not specific to kinetoplastids, but rather are at least shared with other Euglenozoa. Consequently, their presence cannot be explained by the parasitic lifestyle of Trypanosomatidae, as many Euglenozoa are non-parasitic (e.g. photosynthetic euglenids and free-living diplomonads). However, since many of these subunits are not found beyond Euglenozoa (or Discoba), these differences may explain the biochemical peculiarities observed for the respiratory-chain complexes of kinetoplastids and euglenids.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.mito.2014.02.001>.

## Acknowledgements

We thank M. Radoux and G. Gain for their technical help. This work was supported by the University of Liège (SFRD-11/05 to P.C. and SFRD-12/04 to D.B.), the Fonds National de la Recherche Scientifique (an Incentive Grant for Scientific Research MIS F.4520; FRFC 2.4597.11; FRFC 2.4567.11; CDR J.0138.13) and the FRS-FNRS/CONAcYt B330/123/11 (Belgium–Mexico). E.P. is supported by the Belgian FRIA FRS-FNRS. P.C. is a research associate of FRS-FNRS.

## References

- Acestor, N., Zikova, A., Dalley, R.A., Anupama, A., Panigrahi, A.K., Stuart, K.D., 2011. *Trypanosoma brucei* mitochondrial respiratome: composition and organization in procyclic form. *Mol. Cell. Proteomics* 10 (M110), 006908.
- Adl, S.M., Simpson, A.G., Lane, C.E., Lukes, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenerath, M., Lara, E., Le Gall, L., Lynn, D.H., McManus, H., Mitchell, E.A., Mozley-Stanridge, S.E., Parfrey, L.W., Pawlowski, J., Rueckert, S., Shadwick, R.S., Schoch, C.L., Smirnov, A., Spiegel, F.W., 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* 59, 429–493.
- Angerer, H., Zwicker, K., Wumaier, Z., Sokolova, L., Heide, H., Steger, M., Kaiser, S., Nubel, E., Brutschy, B., Radermacher, M., Brandt, U., Zickermann, V., 2011. A scaffold of accessory subunits links the peripheral arm and the distal proton-pumping module of mitochondrial complex I. *Biochem. J.* 437, 279–288.
- Balabaskaran Nina, P., Dudkina, N.V., Kane, L.A., van Eyk, J.E., Boekema, E.J., Mather, M.W., Vaidya, A.B., 2010. Highly divergent mitochondrial ATP synthase complexes in *Tetrahymena thermophila*. *PLoS Biol.* 8, e1000418.
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. *Nucleic Acids Res.* 41, D36–D42.
- Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.
- Bronstrup, U., Hachtel, W., 1989. Cytochrome *c* oxidase of *Euglena gracilis*: purification, characterization, and identification of mitochondrially synthesized subunits. *J. Bioenerg. Biomembr.* 21, 359–373.
- Brown, G.E., Preston, J.F., 1975. Changes in mitochondrial density and succinic dehydrogenase activity in *Euglena gracilis* as a function of the dependency on light for growth. *Arch. Microbiol.* 104, 233–236.
- Buetow, D.E., 1961. Ethanol stimulation of oxidative metabolism in *Euglena gracilis*. *Nature* 190, 1196–1196.
- Burger, G., Gray, M.W., Forget, L., Lang, B.F., 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout Jakobid protists. *Genome Biol. Evol.* 5, 418–438.
- 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.
- Cardol, P., Matagne, R.F., Remacle, C., 2002. Impact of mutations affecting ND mitochondria-encoded subunits on the activity and assembly of complex I in *Chlamydomonas*. Implication for the structural organization of the enzyme. *J. Mol. Biol.* 319, 1211–1221.
- Cardol, P., Vanrobaeys, F., Devreese, B., Van Beeumen, J., Matagne, R.F., 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.
- Cardol, P., Boutaffala, L., Memmi, S., Devreese, B., Matagne, R.F., Remacle, C., 2008. In *Chlamydomonas*, the loss of ND5 subunit prevents the assembly of whole mitochondrial complex I and leads to the formation of a low abundant 700 kDa subcomplex. *Biochim. Biophys. Acta* 1777, 388–396.
- Carroll, J., Fearnley, I.M., Skehel, J.M., Shannon, R.J., Hirst, J., Walker, J.E., 2006. Bovine complex I is a complex of 45 different subunits. *J. Biol. Chem.* 281, 32724–32727.
- Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.
- Castro-Guerrero, N.A., Krab, K., Moreno-Sánchez, R., 2004. The alternative respiratory pathway of *Euglena* mitochondria. *J. Bioenerg. Biomembr.* 36, 459–469.
- Castro-Guerrero, N.A., Jasso-Chavez, R., Moreno-Sánchez, R., 2005. Physiological role of rhodoquinone in *Euglena gracilis* mitochondria. *Biochim. Biophys. Acta* 1710, 113–121.
- Cavalier-Smith, T., 1981. Eukaryote kingdoms: seven or nine? *Biosystems* 14, 461–481.
- Chaudhuri, M., Ott, R.D., Hill, G.C., 2006. Trypanosome alternative oxidase: from molecule to function. *Trends Parasitol.* 22, 484–491.
- Collinson, I.R., Runswick, M.J., Buchanan, S.K., Fearnley, I.M., Skehel, J.M., van Raaij, M.J., Griffiths, D.E., Walker, J.E., 1994. Fo membrane domain of ATP synthase from bovine heart mitochondria: purification, subunit composition, and reconstitution with F1-ATPase. *Biochemistry* 33, 7971–7978.
- Cruciat, C.M., Brunner, S., Baumann, F., Neupert, W., Stuart, R.A., 2000. The cytochrome *bc1* and cytochrome *c* oxidase complexes associate to form a single supercomplex in yeast mitochondria. *J. Biol. Chem.* 275, 18093–18098.
- Cui, J.Y., Mukai, K., Saeki, K., Matsubara, H., 1994. Molecular cloning and nucleotide sequences of cDNAs encoding subunits I, II, and IX of *Euglena gracilis* mitochondrial complex III. *J. Biochem.* 115, 98–107.
- Davies, K.M., Anselmi, C., Wittig, I., Faraldo-Gomez, J.D., Kuhlbrandt, W., 2012. Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. *Proc. Natl. Acad. Sci. U. S. A.* 109, 13602–13607.
- Duby, F., Matagne, R.F., 1999. Alteration of dark respiration and reduction of phototrophic growth in a mitochondrial DNA deletion mutant of *Chlamydomonas* lacking cob, nd4, and the 3' end of nd5. *Plant Cell* 11, 115–125.
- Dudkina, N.V., Eubel, H., Keegstra, W., Boekema, E.J., Braun, H.P., 2005. Structure of a mitochondrial supercomplex formed by respiratory-chain complexes I and III. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3225–3229.
- Dudkina, N.V., Sunderhaus, S., Braun, H.P., Boekema, E.J., 2006. Characterization of dimeric ATP synthase and cristae membrane ultrastructure from *Saccharomyces* and *Polytomella* mitochondria. *FEBS Lett.* 580, 3427–3432.
- Dyall, S.D., Brown, M.T., Johnson, P.J., 2004. Ancient invasions: from endosymbionts to organelles. *Science* 304, 253–257.
- Ebersberger, I., Strauss, S., von Haeseler, A., 2009. HaMSTR: profile hidden Markov model based search for orthologs in ESTs. *BMC Evol. Biol.* 9, 157.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Fang, J., Beattie, D.S., 2002. Novel FMN-containing rotenone-insensitive NADH dehydrogenase from *Trypanosoma brucei* mitochondria: isolation and characterization. *Biochemistry* 41, 3065–3072.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N., Thurston, M., 2006. Open software for biologists: from famine to feast. *Nat. Biotechnol.* 24, 801–803.
- Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., Stoecker Jr., C.J., 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr. Protoc. Bioinformatics* 12, 11–19 (Chapter 6, Unit 6).
- Gabalton, T., Huynen, M.A., 2005. Lineage-specific gene loss following mitochondrial endosymbiosis and its potential for function prediction in eukaryotes. *Bioinformatics* 21 (Suppl. 2), ii144–ii150.
- Gabalton, T., Koonin, E.V., 2013. Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* 14, 360–366.
- Gawryluk, R.M., Gray, M.W., 2009. A split and rearranged nuclear gene encoding the iron-sulfur subunit of mitochondrial succinate dehydrogenase in Euglenozoa. *BMC Res. Notes* 2, 16.
- Gawryluk, R.M., Gray, M.W., 2010. Evidence for an early evolutionary emergence of gamma-type carbonic anhydrases as components of mitochondrial respiratory complex I. *BMC Evol. Biol.* 10, 176.
- Gawryluk, R.M., Chisholm, K.A., Pinto, D.M., Gray, M.W., 2012. Composition of the mitochondrial electron transport chain in *Acanthamoeba castellanii*: structural and evolutionary insights. *Biochim. Biophys. Acta* 1817, 2027–2037.
- Gibbs, S.P., 1981. The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann. N. Y. Acad. Sci.* 361, 193–208.
- Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224.
- Gray, M.W., Burger, G., Lang, B.F., 1999. Mitochondrial evolution. *Science* 283, 1476–1481.
- Gray, M.W., Lang, B.F., Burger, G., 2004. Mitochondria of protists. *Annu. Rev. Genet.* 38, 477–524.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G., Roger, A.J., 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc. Natl. Acad. Sci. U. S. A.* 106, 3859–3864.
- Harris, E.H., 1989. *The Chlamydomonas Sourcebook*. Academic Press, San Diego.
- Heazlewood, J.L., Whelan, J., Millar, A.H., 2003. The products of the mitochondrial orf25 and orfB genes are FO components in the plant F1FO ATP synthase. *FEBS Lett.* 540, 201–205.
- Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9, 868–877.
- Huynen, M.A., de Hollander, M., Szklarczyk, R., 2009. Mitochondrial proteome evolution and genetic disease. *Biochim. Biophys. Acta* 1792, 1122–1129.

- Kadenbach, B., Jarausch, J., Hartmann, R., Merle, P., 1983. Separation of mammalian cytochrome *c* oxidase into 13 polypeptides by a sodium dodecyl sulfate-gel electrophoretic procedure. *Anal. Biochem.* 129, 517–521.
- Klodmann, J., Sunderhaus, S., Nimitz, M., Jansch, L., Braun, H.P., 2010. Internal architecture of mitochondrial complex I from *Arabidopsis thaliana*. *Plant Cell* 22, 797–810.
- Koski, L.B., Golding, G.B., 2001. The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* 52, 540–542.
- Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Gray, M.W., 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387, 493–497.
- Lapaille, M., Escobar-Ramirez, A., Degand, H., Baurain, D., Rodriguez-Salinas, E., Coosemans, N., Boutry, M., Gonzalez-Halphen, D., Remacle, C., Cardol, P., 2010. Atypical subunit composition of the chlorophycean mitochondrial F1F0-ATP synthase and role of Asa7 protein in stability and oligomycin resistance of the enzyme. *Mol. Biol. Evol.* 27, 1630–1644.
- Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320.
- Li, L., Stoeckert Jr., C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189.
- Martin, W., Muller, M., 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37–41.
- Mego, J.L., Farb, R.M., 1974. Alcohol dehydrogenases of *Euglena gracilis*, strain Z. *Biochim. Biophys. Acta* 350, 237–239.
- Michalecka, A.M., Svensson, A.S., Johansson, F.I., Agius, S.C., Johanson, U., Brennicke, A., Binder, S., Rasmussen, A.G., 2003. *Arabidopsis* genes encoding mitochondrial type II NAD(P)H dehydrogenases have different evolutionary origin and show distinct responses to light. *Plant Physiol.* 133, 642–652.
- Millar, A.H., Eubel, H., Jansch, L., Kruff, V., Heazlewood, J.L., Braun, H.P., 2004. Mitochondrial cytochrome *c* oxidase and succinate dehydrogenase complexes contain plant specific subunits. *Plant Mol. Biol.* 56, 77–90.
- Monroy, A.F., Schwartzbach, S.D., 1984. Catabolite repression of chloroplast development in *Euglena*. *Proc. Natl. Acad. Sci. U. S. A.* 81, 2786–2790.
- Morales, J., Mogi, T., Mineki, S., Takashima, E., Mineki, R., Hirawake, H., Sakamoto, K., Omura, S., Kita, K., 2009. Novel mitochondrial complex II isolated from *Trypanosoma cruzi* is composed of 12 peptides including a heterodimeric Ip subunit. *J. Biol. Chem.* 284, 7255–7263.
- Moreno-Sánchez, R., Raya, J.C., 1987. Preparation of coupled mitochondria from *Euglena* by sonication. *Plant Sci.* 48, 151–157.
- Moreno-Sánchez, R., Covian, R., Jasso-Chavez, R., Rodriguez-Enriquez, S., Pacheco-Moises, F., Torres-Marquez, M.E., 2000. Oxidative phosphorylation supported by an alternative respiratory pathway in mitochondria from *Euglena*. *Biochim. Biophys. Acta* 1457, 200–210.
- Panigrahi, A.K., Zikova, A., Dalley, R.A., Acestor, N., Ogata, Y., Anupama, A., Myler, P.J., Stuart, K.D., 2008. Mitochondrial complexes in *Trypanosoma brucei*: a novel complex and a unique oxidoreductase complex. *Mol. Cell. Proteomics* 7, 534–545.
- Pullman, M.E., Penefsky, H.S., Datta, A., Racker, E., 1960. Partial resolution of the enzymes catalyzing oxidative phosphorylation. I. Purification and properties of soluble dinitrophenol-stimulated adenosine triphosphatase. *J. Biol. Chem.* 235, 3322–3329.
- Qiu, X.B., Shao, Y.M., Miao, S., Wang, L., 2006. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell. Mol. Life Sci.* 63, 2560–2570.
- Remacle, C., Baurain, D., Cardol, P., Matagne, R.F., 2001. Mutants of *Chlamydomonas reinhardtii* deficient in mitochondrial complex I: characterization of two mutations affecting the nd1 coding sequence. *Genetics* 158, 1051–1060.
- Saccone, C., Lanave, C., De Grassi, A., 2006. Metazoan OXPHOS gene families: evolutionary forces at the level of mitochondrial and nuclear genomes. *Biochim. Biophys. Acta* 1757, 1171–1178.
- Schagger, H., Pfeiffer, K., 2000. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J.* 19, 1777–1783.
- Schagger, H., von Jagow, G., 1991. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* 199, 223–231.
- Sharpless, T.K., Butow, R.A., 1970a. An inducible alternate terminal oxidase in *Euglena gracilis* mitochondria. *J. Biol. Chem.* 245, 58–70.
- Sharpless, T.K., Butow, R.A., 1970b. Phosphorylation sites, cytochrome complement, and alternate pathways of coupled electron transport in *Euglena gracilis* mitochondria. *J. Biol. Chem.* 245, 50–57.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- Simpson, A.G.B., 1997. The identity and composition of the Euglenozoa. *Arch. Protistenkd.* 148, 10.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.J., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D., Birney, E., 2002. The BioPerl toolkit: Perl modules for the life sciences. *Genome Res.* 12, 1611–1618.
- Sunderhaus, S., Dudkina, N.V., Jansch, L., Klodmann, J., Heinemeyer, J., Perales, M., Zabaleta, E., Boekema, E.J., Braun, H.P., 2006. Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. *J. Biol. Chem.* 281, 6482–6488.
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yoshikawa, S., 1996. The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science* 272, 1136–1144.
- Turmel, M., Gagnon, M.C., O'Kelly, C.J., Otis, C., Lemieux, C., 2009. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.* 26, 631–648.
- Van Aken, O., Giraud, E., Clifton, R., Whelan, J., 2009. Alternative oxidase: a target and regulator of stress responses. *Physiol. Plant.* 137, 354–361.
- Vazquez-Acevedo, M., Cardol, P., Cano-Estrada, A., Lapaille, M., Remacle, C., Gonzalez-Halphen, D., 2006. The mitochondrial ATP synthase of chlorophycean algae contains eight subunits of unknown origin involved in the formation of an atypical stator-stalk and in the dimerization of the complex. *J. Bioenerg. Biomembr.* 38, 271–282.
- Velours, J., Arselin, G., 2000. The *Saccharomyces cerevisiae* ATP synthase. *J. Bioenerg. Biomembr.* 32, 383–390.
- Villavicencio-Queijeiro, A., Vazquez-Acevedo, M., Cano-Estrada, A., Zarco-Zavala, M., Tuena de Gomez, M., Mignaco, J.A., Freire, M.M., Scofano, H.M., Foguel, D., Cardol, P., Remacle, C., Gonzalez-Halphen, D., 2009. The fully-active and structurally-stable form of the mitochondrial ATP synthase of *Polytomella* sp. is dimeric. *J. Bioenerg. Biomembr.* 41, 1–13.
- Walker, J.E., Lutter, R., Dupuis, A., Runswick, M.J., 1991. Identification of the subunits of F1F0-ATPase from bovine heart mitochondria. *Biochemistry* 30, 5369–5378.
- Walker, G., Dorrell, R.G., Schlacht, A., Dacks, J.B., 2011. Eukaryotic systematics: a user's guide for cell biologists and parasitologists. *Parasitology* 138, 1638–1663.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
- Wilson, B.W., Danforth, W.F., 1958. The extent of acetate and ethanol oxidation by *Euglena gracilis*. *J. Gen. Microbiol.* 18, 535–542.
- Yang, Z., 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10, 1396–1401.
- Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829.
- Zhao, S., Burki, F., Brate, J., Keeling, P.J., Klaveness, D., Shalchian-Tabrizi, K., 2012. Collodictyon—an ancient lineage in the tree of eukaryotes. *Mol. Biol. Evol.* 29, 1557–1568.
- Zikova, A., Panigrahi, A.K., Uboldi, A.D., Dalley, R.A., Handman, E., Stuart, K., 2008. Structural and functional association of *Trypanosoma brucei* MIX protein with cytochrome *c* oxidase complex. *Eukaryot. Cell* 7, 1994–2003.
- Zikova, A., Schnauffer, A., Dalley, R.A., Panigrahi, A.K., Stuart, K.D., 2009. The F(0)F(1)-ATP synthase complex contains novel subunits and is essential for procyclic *Trypanosoma brucei*. *PLoS Pathog.* 5, e1000436.