The iron-sulfur scaffold protein HCF101 unveils the complexity of organellar evolution
in SAR, Haptista and Cryptista

Jan Pyrihä, Vojtěch Žárský, Justin D. Fellows, Christopher Grosche, Dorota Wloga, Boris Striepen, Uwe G. Maier, Jan Tachezy.

a Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Průmyslová 595, 252 50 Vestec, Czech Republic.

b Department of Cellular Biology, University of Georgia, Athens, Georgia, USA.

c Laboratory for Cell Biology, Philipps University Marburg, Karl-von-Frisch-Str. 8, 35032, Marburg, Germany.

d LOEWE Center for Synthetic Microbiology (Synmikro), Hans-Meerwein-Str. 6, 35032 Marburg, Germany

e Laboratory of Cytoskeleton and Cilia Biology, Nencki Institute of Experimental Biology of Polish Academy of Sciences, 3 Pasteur Street, 02-093 Warsaw, Poland.

f To whom correspondence should be addressed: Jan Tachezy, Department of Parasitology, Faculty of Science, Charles University BIOCEV, Průmyslová 595, 252 50 Vestec, Czech Republic.

E-mail: tachezy@natur.cuni.cz
Abstract

**Background:** Nbp35-like proteins (Nbp35, Cfd1, HCF101, Ind1, and AbpC) are P-loop NTPases that serve as components of iron-sulfur cluster (FeS) assembly machineries. In eukaryotes, Ind1 is present in mitochondria, and its function is associated with the assembly of FeS clusters in subunits of respiratory Complex I, Nbp35 and Cfd1 are the components of the cytosolic FeS assembly (CIA) pathway, and HCF101 is involved in FeS assembly of photosystem I in plastids of plants (chHCF101). The AbpC protein operates in Bacteria and Archaea. To date, the cellular distribution of these proteins is considered to be highly conserved with only a few exceptions.

**Results:** We searched for the genes of all members of the Nbp35-like protein family and analyzed their targeting sequences. Nbp35 and Cfd1 were predicted to reside in the cytoplasm with some exceptions of Nbp35 localization to the mitochondria; Ind1 was found in the mitochondria, and HCF101 was predicted to reside in plastids (chHCF101) of all photosynthetically active eukaryotes. Surprisingly, we found a second HCF101 paralog in all members of Cryptista, Haptista, and SAR that was predicted to predominantly target mitochondria (mHCF101), whereas Ind1 appeared to be absent in these organisms. We also identified a few exceptions, as apicomplexans possess mHCF101 predicted to localize in the cytosol and Nbp35 in the mitochondria. Our predictions were experimentally confirmed in selected representatives of Apicomplexa (*Toxoplasma gondii*), Stramenopila (*Phaeodactylum tricornutum, Thalassiosira pseudonana*), and Ciliophora (*Tetrahymena thermophila*) by tagging proteins with an transgenic reporter. Phylogenetic analysis suggested that chHCF101 and mHCF101 evolved from a common ancestral HCF101 independently of the Nbp35/Cfd1 and Ind1 proteins. Interestingly, phylogenetic analysis supports rather a lateral gene transfer of ancestral HCF101 from bacteria than its acquisition being associated with either α-proteobacterial or cyanobacterial endosymbionts.
Conclusion: Our searches for Nbp35-like proteins across eukaryotic lineages revealed that SAR, Haptista, and Cryptista possess mitochondrial HCF-101. Because plastid localization of HCF101 was only known thus far, the discovery of its mitochondrial paralog explains confusion regarding the presence of HCF101 in organisms that possibly lost secondary plastids (e.g., ciliates, Cryptosporidium) or possess reduced nonphotosynthetic plastids (apicomplexans).

Keywords

HCF101, Ind1, iron-sulfur cluster, mitochondrion, plastid, evolution

Background

Iron-sulfur (FeS) cluster assembly pathways are essential for all three domains of life: Bacteria, Archaea, and Eukarya. In eukaryotes, there are three main pathways, which are localized in distinct cellular compartments: mitochondria, plastids, and the cytosol. The organellar pathways were acquired through endosymbiosis of proteobacteria and cyanobacteria that evolved into mitochondria and plastids, respectively [1, 2]. The mitochondrial FeS cluster assembly (ISC) machinery operates in nearly all forms of mitochondria including anaerobic hydrogenosomes [3] and highly reduced mitosomes [4]. The pathway in plastids is called the sulfur utilization factor (SUF) system, which is present in primary [5] as well as secondary plastids [6–8]. The ISC machinery is functionally linked to the third system, the cytosolic FeS cluster assembly (CIA) machinery. Phylogenetic analysis suggested that the CIA pathway was present in the last eukaryotic common ancestor (LECA) and that its components are predominantly of bacterial origin [9, 10]. There are few
known exceptions to the highly conserved setup of FeS assembly machineries, and all these exceptions concern protists adapted to anaerobic or microaerobic conditions with modified mitochondria. Archamoebae replaced the ISC pathway with two components of a nitrogen-fixing (NIF) machinery that were acquired by lateral gene transfer (LGT) from ε-proteobacteria[11]. The NIF system operates in the cytosol of *Entamoeba histolytica* or in the cytosol and hydrogenosomes of *Mastigamoeba balamuthi* [12]. Similarly, the breviate *Pygsuia biforma* apparently replaced the ISC system with an archaean SUF system [13, 14]. Finally, three SUF components (SufC, SufB, and fused protein SufDSU) of bacterial origin were found in the cytosol of the oxymonad *Monocercomonoides sp.*, which lost its mitochondria [15].

The only proteins that are common to the CIA, ISC, and SUF pathways are P-loop NTPases with the ParA domain: Nbp35/Cfd1, Ind1, and high chlorophyll fluorescence 101 (HCF101), respectively (hereafter Nbp35-like proteins). In CIA, Nbp35/Cfd proteins serve in the initial phase of FeS assembly as a [4Fe-4S] scaffold using sulfur and iron that are exported from mitochondria [16]. The FeS cluster is then transferred via Nar1 and the CIA/Cia2/MMS19 targeting complex to apo-proteins. Ind1 serves as a scaffold in later stages of FeS assembly to deliver [4Fe-4S] clusters specifically to the apo-subunits of mitochondrial respiratory complex I, and thus, the presence of Ind1 closely matches the complex I distribution [17]. Its necessity for complex I maturation underlines the presence of Ind1 in hydrogenosomes, in which complex I is reduced to only two FeS subunits [14, 18]. HCF101 was shown to transport [4Fe-4S] clusters to photosystem I subunits and heterodimeric ferredoxin-thioredoxin reductase complexes in plastids of *Arabidopsis thaliana* [19, 20].

It is believed that the canonical distribution of FeS cluster assembly machineries and thus that of machinery-specific Nbp35-like proteins is highly conserved in eukaryotes, including protists with primary or complex plastids. The latter organelles evolved in
eukaryotic hosts from eukaryotic symbionts with green (Euglenozoa and Chlorarachniophyceae) or red (Stramenopila, Alveolata, Haptophytes, and Cryptophytes) plastids [21, 22]. These complex plastids are surrounded by three or more membranes and characterized by the presence of a periplastidal compartment, the extremely reduced cytosol of the endosymbiont and, in the case of cryptophytes and chlorarachniophytes, of a remnant nucleus (nucleomorph). Interestingly, the presence of nucleomorph, which is likely dependent on activities of FeS proteins, correlates with the presence of the endosymbiotic CIA, including Nbp35 that is retained in the periplastidial compartment [7] in addition to CIA in the host cytosol. This curious finding further exemplifies the conserved topology of Nbp35 and other CIA components.

The localization of HCF101 has not been experimentally studied in most eukaryotic lineages. Moreover, because HCF101 is essential for photosystem I and consequently photosynthesis, it could be particularly interesting to investigate its presence and cellular localization in organisms that possess non-photosynthesizing plastids such as the apicoplast in apicomplexans. The genes for HCF101 have been noticed in several apicomplexan genomes such as Toxoplasma gondii and Plasmodium falciparum, and their possible localization in the apicoplast has been suggested [23, 24]. However, these HCF101 homologs lack the targeting signals one would expect for proteins localized to the apicoplast [23, 24]. Even more puzzling is the identification of HCF101 in the genome of Cryptosporidium parvum, which has lost its plastid [24]. Therefore, we decided to search for Nbp35-like genes across eukaryotic genomes and to predict their cellular localization based on their organellar targeting presequences. In selected protists, we verified the localization of Nbp35-like proteins experimentally. The most surprising result is the identification of the mitochondrial form of HCF101 in protists with complex plastids.
**Results**

**Distribution of Nbp35-like proteins in eukaryotes**

We searched for Nbp35, Cfd1, Ind1, and HCF101 in genomes and transcriptomes across the main eukaryotic lineages, and for each protein we predicted its putative cellular localization (Table 1, Table S1). While Nbp35 was found ubiquitously in all lineages as reported previously [10], Cfd1 was generally present in Ophistokonta, Amoebozoa, Cryptista, Glaucophyta, and Excavata (Metamonada, Discoba) supergroups but absent in the remaining groups of Archaeplastida, SAR, and Haptista. Diplomonads such as *Giardia intestinalis* and *Spironucleus salmonicida* represent the only exception within excavates as they lack Cfd1 (Table 1)[25]. Nbp35 homologs were not identified in only four organisms, most likely due to the incompleteness of the available sequencing data (Table 1). Curiously, in *Mastigamoeba balamuthi*, there are three Nbp35 paralogs, of which one was predicted to possess N-terminal hydrogenosomal targeting presequences (Table 1). Furthermore, we also predicted a mitochondrial targeting signal for Nbp35 proteins in Apicomplexa and Chromerids.

As expected, Ind1 was predicted to be present in the mitochondria of Ophistokonta, Amoebozoa, Archaeplastida, and Excavata groups except for organisms that lack complex I (Table 1). Interestingly, we did not identify Ind1 in any organism with complex plastids. While this is not surprising for apicomplexans that lack complex I such as *Toxoplasma gondii* and *Plasmodium falciparum* and evolutionarily related chromerids *Chromera vellia* and *Vitrella brassica*, Ind1 was also absent in all other lineages of the SAR, Haptista and Cryptista groups, despite the presence of genes for the FeS subunits of Complex I in these organisms [17].

Finally, we searched for genes encoding the HCF101 protein. This protein could be easily distinguished from other Nbp35-like proteins based on the presence of two extra
domains, an N-terminal FeS assembly P domain (FSCA, previously domain of unknown function DUF59) and a C-terminal DUF971 [20]. Surprisingly, the distribution of HCF101 was limited not only to lineages harboring primary plastids (Viridiplantae, Rhodophyta, Glauco phyta) or complex plastids of red (SAR, Cryptophyta, Haptophyta) or green (chlororachi niophytes, euglenids and some dinoflagellates) origin, but the gene was also present in the remaining nonphotosynthetic members of SAR, Haptista, and Cryptista. Every photosynthetically active eukaryote possesses a gene that encodes HCF101 with either an N-terminal primary plastid targeting signal (Archaeplastida) or a bipartite signal, which targets the protein to complex plastids (chHCF101) (Fig. 1). Strikingly, in all members of SAR, Haptista, and Cryptista (formerly referred to as Chromalveolata), we found a second HCF101 paralog, with predicted mitochondrial localization (mHCF101). The only unexpected variation of this cellular localization was found in Alveolata. In apicomplexans that harbor a nonphotosynthetic apicoplast, HCF101 was predicted to reside in the cytosol, while Nbp35 possesses an N-terminal extension, which may target the protein to the mitochondria. The same cytosolic distribution of HCF101 and possibly mitochondrial Nbp35 we predicted also for evolutionarily related Chromerids that possess photosynthetic plastids, and therefore also chHCF101. In other Alveolates, such as *Perkinus marinus*, that possesses cryptic nonphotosynthetic plastids and in ciliates that lack plastids, we predicted standard cytosolic localization for Nbp35 and mitochondrial localization of putative mHCF101, whereas chHCF101 is absent. The distribution of Nbp35, mHCF101, and chHCF101 in dinoflagellates is likely similar to that in Stramenopila and Rhizaria; however, predictions of protein localization in some dinoflagellates had low confidence.

**Experimental localization of selected Nbp35-like proteins**
The identification of mHCF101 and the unexpected localization predicted for Nbp35 in apicomplexans prompted us to select three protists that are amenable for cell transformation and investigate the localization of Nbp35-like proteins using protein tagging. First, we tested genes from the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* that possess secondary plastids. *P. tricornutum* cells were transformed to express homologous eGFP-tagged mHCF101, chHCF101, and Nbp35 as well as heterologous genes from *T. pseudonana*. Fluorescence microscopy revealed that both mHCF101 proteins labeled structures corresponding to mitochondria as indicated by colabeling with MitoTracker. These structures were clearly distinct from plastids, in which we observed labeling with chHCF101 (Fig. 2). As expected, Nbp35 labeling corresponded to the cytosol. Next, we tested localization of putative mHCF101 and Nbp35 in the ciliate *Tetrahymena thermophila*, which lacks plastids. HA-tagged mHCF101 appeared in numerous round mitochondria organized in longitudinal arrays that were again also labeled with MitoTracker (Fig. 3). Nbp35 appeared as a diffuse signal within the cell corresponding to the cytosol. Finally, we expressed HCF101 and Nbp35 in *T. gondii* (Fig. 3). This organism lacks mitochondrial Ind1 and possesses a reduced nonphotosynthetic plastid, the apicoplast. Nbp35 clearly colocalized in tubular structures with the mitochondrial marker F1-ATPase. Putative HCF101 appeared within the cell as a cytosolic protein. No localization of HCF101 to the apicoplast was observed using the antibody against plastidial CPN60. These experimental data confirmed the predicted localization of mHCF101 in diatoms and the ciliate and mitochondrial localization of Nbp35 in *Toxoplasma*.

Phylogeny of HCF101
To learn about the evolutionary history of chHCF101 and mHCF101 and to obtain further support for predictions of their cellular localization, we performed phylogenetic analysis. In the first step, we were interested in the relationship between HCF101 and other members of the Nbp35-like protein family. We analyzed a large dataset of 8440 amino acid sequences including mHCF101, chHCF101, Nbp35, Cfd1, and Ind1 as well as prokaryotic homologs of ApbC proteins with the ParA domain. We expected that chHCF101 originated from a cyanobacterial endosymbiont that evolved to a plastid, similar to Ind1, which was acquired with the α-proteobacterial ancestor of mitochondria [9]. However, chHCF101 and mHCF101 formed a common clade with various lineages of bacteria that appeared at the base of the HCF101 subtree, including proteobacteria, the PVC group, and Bacteroidetes. There is no obvious support for the cyanobacterial ancestry of HCF101 and thus for endosymbiotic gene transfer (EGT), although the overall resolution of the tree is low. As expected, Ind1 formed a clade with the majority of eukaryotic sequences and α-proteobacteria at a basal position that is consistent with EGT origin of the protein (Fig. 4). Interestingly, Ind1 of kinetoplastids appeared at a separate position in the Ind1/α-proteobacterial subtree than the rest of the eukaryotic sequences, suggesting its specific phylogenetic history. It is possible that hand in hand with the presence of an atypical Complex I in kinetoplastids, Ind1 protein also underwent dramatic evolutonal reshaping [26]. Finally, Nbp35/Cfd1 clustered together with various eubacterial and archaeabacterial sequences as observed previously [9, 10].

Because the statistical support was moderate throughout the phylogenetic tree, we also performed prediction of protein domains with a focus on the presence of the HCF101 marker domains FSCA and DUF951 to obtain more information to estimate a possible HCF101 origin (Additional file 1, Table S2). This analysis showed that the majority of bacterial sequences have a FSCA-ParA structure (3420), or contain the ParA domain only (2762), and there are also various other domain combinations. Of note, 18 sequences obtained from
proteobacteria, PVC group members, and *Bacteroidetes* clustered with eukaryotic HCF101
and shared the characteristic FSCA-ParA-DUF951 domains structure of HCF101.

In the second step, we focused on more detailed phylogenetic analysis of chHCF101
and mHCF101 (Fig. 5). The phylogenetic tree revealed that chHCF101 and mHCF101 are
paralogs that evolved from a common HCF101 ancestor, possibly by duplication events.
ChHCF101 and mHCF101 formed two monophyletic groups albeit with low support. The
chHCF101 tree is by-and-large consistent with the current concept of eukaryotic phylogeny.
There are three well-supported clades of Archaeplastida with primary plastids for
Viridiplantae, Glaucophyta, and Rhodophyta together with protists that harbor corresponding
secondary plastids. Thus, chHCF101 in Viridiplantae clusters together with Euglenozoa that
possesses secondary plastids of green origin. ChHCF101 of Rhodophyta is at the base of
Stramenopila, Chromerida, Cryptophytes, and Haptophytes, which contain Rhodophyta-
derived red secondary plastids. However, there are some exceptions. Some dinoflagellates
such *Alexandrium* and *Symbiodinium* that have secondary plastids of red origin yet seem to
possess chHCF101 related to the green plastid lineage. Conversely, although
chlorarachniophytes acquired secondary plastids of green ancestry, their chHCF101 clustered
within orthologs of red plastids. Interestingly, this single gene phylogeny supports the close
relationship between plastids of Haptophytes and Cryptophytes. Branching of the main groups
was further supported by a comparison of six conserved amino acid residues (AA 461-466
according to *Arabidopsis thaliana*) in the DUF971 domain. The common motive for
Viridiplantae, Euglenozoa and Dinophyta was D[K,R,Q,T][G,S]Ax[G,S], chHCF101 of
Glaucophyta, Rhodophyta, Stramenophila, and Chromerida possess the highly conserved
motif C[R,S]CAxC, and Chlorarachniophyta, Cryptophytes and Haptophytes possess
CRSP[A,T,S]N.
The observed branching order of mHCF101 is poorly supported (Fig. 5); nevertheless, separation of the chHCF101 and mHCF101 groups provides a tool for our prediction of cell localization, as several sequences included in Table 1 were incomplete and thus preclude confident predictions based on the identification of N-terminal targeting motifs. For example, in dinoflagellates, we found complete sequences of two HCF101 paralogs only for *A. tamarense*. Sequences of all other dinoflagellates were incomplete; however, phylogenetic analysis clearly separated group mHCF101 including *A. tamarense* HCF101 with mitochondrial targeting presequence and formed a subtree of dinoflagellates with high statistical support. The other HCF101 dinoflagellate paralogs appeared within the chloroplast group. We were particularly interested in the origin of HCF101 of apicomplexans that lack N-terminal targeting sequences, and in *T. gondii*, we demonstrated its cytosolic localization. HCF101 proteins of *T. gondii* and other related apicomplexans including *Cystoisospora suis* clearly appeared within the mHCF101 group, at the base of a well-supported subtree of apicomplexans, chromerids, and dinoflagellates. Therefore, apicomplexan HCF101s seem not to be derived from plastids (apicoplast), contrary to previous assumptions [23, 24]. Another interesting question was the origin of HCF101 in ciliates which lack plastids. Phylogeny of HCF101 showed that HCF101 in ciliates is not related to chHCF101 but clustered within mHCF101s.

Several members of the PVC group such as *Kiritimatiellaceae bacterium* and *Verrucomicrobia bacterium* are at the base of the HCF101 tree. Although this tree is poorly resolved, it is noteworthy that the bacterial conserved motif of DUF971 C[A,R,N,H]CA[A,L]C is similar to the motifs in chHCF101 as well as most mHCF101 (Fig. 5). Interestingly, the verrucomicrobial HCF101 clustered with the orthologs of the glaucophyta group, which is considered to possess the most primitive plastid. Thus, based on the phylogeny analysis, the presence of bacterial HCF101-like proteins with specific domain
structures, and the conserved DUF971 motif, the subset of PVC group members represents the best candidates for the origin of eukaryotic HCF101.

Discussion

In this work, we screened Nbp35-like homologs across eukaryotes and predicted their cellular localization. This analysis discovered the existence of a mitochondrial HCF101 homolog that is common to all tested members of SAR, Haptophytes, and Cryptophytes. Localization of mHCF101 was predicted based on the identification of N-terminal mitochondrial targeting sequences and supported by a phylogenetic analysis that separated mHCF101 from the chHCF101 paralog. Moreover, mitochondrial localization of mHCF101 was experimentally verified for mHCF101 encoded in the genomes of two diatoms (T. pseudonana, P. tricornutum) and the ciliate T. thermophila. Curiously, but consistently with the in silico predictions, we found mHCF101 in the cytosol of T. gondii, while Nbp35 was localized to the mitochondrion. Evolutionary analysis of HCF101 proteins and their specific distribution suggested that HCF101 was gained potentionally via LGT from bacteria of the PVC lineage either by a common ancestor of Archaoplastida to serve in the chloroplast (plastid-first hypothesis) or by a common ancestor of Archeaplantida SAR, Haptista and Cryptista to serve first in mitochondria.

The presence of mHCF101 is coincident with the absence of Ind1, which is involved in the maturation of complex I FeS subunits. This specific distribution suggests that mHCF101 may act as a functional homolog of Ind1. Both proteins share conserved nucleotide-binding domain characteristics of the Mrp (MetG-related protein)/Nbp35 subclass of ParA P-loop NTPases [27], which includes the conserved CxxC motif. This motif is essential to bind the transient [4Fe4S] cluster that is transferred to the target FeS proteins [20,
It is evident that chHCF101 in chloroplasts and Ind1 in mitochondria transfers labile FeS clusters to different targets. However, both proteins are able to deliver the labile cluster to the S. cerevisiae model [4Fe4S] acceptor protein, isopropyl malate isomerase, in vitro [20, 28]. Thus, the function of HCF101 proteins and Ind1 might be interchangeable. The major difference between HCF101 and Ind1 is the presence of N- and C-terminal domains in the former protein. The FSCA domain is present at the N-terminus of HCF101 (just after the N-terminal targeting sequence) and in a few other eukaryotic proteins involved in FeS assembly such as Cia2 of CIA machinery [29] and asymmetric leaves1/2 enhancer7 (AE7), which is a Cia2 homolog in A. thaliana [30]. The FSCA domain in combination with ParA was identified in a large number of bacterial and some archael FeS cluster carrier proteins (this work). Importantly, in Staphylococcus aureus, the FSCA domain is composed solely of the SufT subunit of SUF machinery and acts as an auxiliary FeS cluster maturation factor [31]. Therefore, the fusion of FSCA and Nbp35-like protein might be beneficial for more efficient transfer of FeS centers to target proteins. The function of C-terminal DUF971 of HCF101 is currently elusive. However, we noticed that DUF971 present at the C-termini of most chHCF101 and mHCF101 proteins contains a highly conserved CxCxxC motif that may have the capacity to bind divalent metals [32]. Further studies are required to clarify a function of mHCF101 and DUF971 in particular.

The evolutionary journey taken by mHCF101 to arrive in the mitochondria of SAR, Haptista, and Cryptista is a puzzle, but multiple evolutionary scenarios could potentially explain the origin of this gene. Our phylogenetic and domain analysis of HCF101 proteins together with their distribution in eukaryotes suggested that ancestral HCF101 was not acquired via EGT from cyanobacteria that possess simple ParA domain-containing proteins without FCSA and DUF971. Rather, it was gained via LGT from bacteria of the PVC lineage that possessed an HCF101-like protein of the FSCA-ParA-DUF971 domain structure and
cluster (although with weak support) with chHCF101 of glaucophytes. The key question is whether HCF101 was first targeted to chloroplasts or to mitochondria. Considering the chloroplast-first scenario (Fig. 6A), we can hypothesize that HCF101 was acquired by a common ancestor of Archaeplastida, targeted to chloroplasts, and evolved independently in glaucophytes, green algae/land plants, and red algae. Then, HCF101 was transferred via secondary endosymbiosis of green plastids to Euglenozoa and by transfer of red plastids to a putative common ancestor of SAR, Haptista and Cryptista. In this hypothetical ancestor, the HCF101 was duplicated, and one of the paralogs was targeted to mitochondria (mHCF101), where it functionally replaced Ind1. Alternatively (mitochondria-first), we can hypothesize that HCF101 was first present in the mitochondria of a common ancestor of Archeplastida, SAR, Haptistae and Cryptista [33, 34] and functioned in parallel with Ind1 (Fig. 6B). HCF101 in Archeplastida was then retargeted from mitochondria to the plastid (chHCF101), while Ind1 was lost at least twice independently in a common ancestor of Cryptophytes and of Haptophytes plus SAR.

The proposed plastid-first scenario for HCF101 evolution is consistent with the “chromalveolate” hypothesis that is based on the idea that all lineages with a red secondary plastid are monophyletic [35]. In support of this hypothesis, it has been proposed that all members of chromalveolates share unique SELMA (symbiont-derived ERAD-like machinery) to target proteins into secondary plastids via the endoplasmic reticulum [36, 37]. Furthermore, remnant plastids of some seemingly aplastidal-like members of chromalveolates such as Perkinus marinus were discovered [38]. In ciliates that lack plastid, several proteins of algal origin were previously identified including a MinD-like hypothetical protein in T. thermophila [39]. In our analysis, we identify this protein as mHCF101, and its mitochondrial localization was experimentally confirmed in Tetrahymena. Thus, in addition to SELMA, the
presence of mHCF101 in mitochondria together with the absence of Ind1 is another feature that is common to chromalveolates.

However, an increasing number of phylogenetic studies favor multiple secondary (or serial) endosymbioses in these lineages [33, 40][41]. They refute the chromalveolate hypothesis by placing Cryptophytes within Archeplastida and through the discovery of novel groups such as katablepharids (Cryptista) [42] and centrohelids (Haptista), in which so far no evolutionary traces of plastids have been found. Thus, their lack of plastids could reflect the primary absence of plastids rather than secondary loss [33]. Interestingly, even these lineages contain mHCF101 instead of Ind1, supporting the idea of multiple independent losses of Ind1.

Tertiary endosymbiosis is another facet that complicates tracing HCF101 evolution, particularly in dinoflagellates. Our phylogenetic analyses revealed that chHCF101 of Karenia clustered within the Haptophytes subtree. This is fully consistent with previous inferences that Karenia and related genera of dinoflagellates with the fucoxanthin-containing plastids [43, 44] lost the ancestral secondary plastid, which was replaced by a new plastid from Haptophytes via tertiary endosymbiosis [45–48]. Interestingly, another group of dinoflagellates including Alexandrium and Symbiodinium with peridin-containing plastids of red origin appeared at the base of Viridiplantae in the chHCF101 subtree, which may suggest experience with a green plastid before acquiring the red plastid, as suggested in several studies [40, 48–50]. In contrast to chHCF101 phylogenies, the monophyletic origin of mHCF101 was observed for both groups of dinoflagellates regardless of the multiendosymbiotic events, which clearly reflected different evolutionary histories for mHCF101 and chHCF101.

Another example of the complex evolution of chHCF101 is found in Chlorarachniophytes, which possess a complex plastid of green origin [51]. Perplexingly, the
phylogeny of chHCF101 suggested a red origin for this protein, which clustered with
Haptophytes and Cryptophytes and shared the unique CRSP[T,A,S]N motif of DUF971.
However, this finding may not be so surprising. Previous analyses of the chlorarachniophyte
Bigelowiella natans classified several genes of likely algal origin to be potentially acquired
from the red lineage [52]. These ‘red’ genes are rather puzzling, but might have originated
from cryptic endosymbioses involving red algae prior to the more recent acquisition of a
green lineage endosymbiont [53, 54].

Based on our and previous analyses [10], Nbp35 seems to be the only essential FeS
cluster assembling P-loop ATPase present in all eukaryotic cells. Typically Nbp35 is a
cytosolic member of the CIA machinery; however, there are multiple examples of
mitochondrial localization. In this work, we demonstrated targeting of a single Nbp35 to the
T. gondii mitochondrion, and we similarly predicted mitochondrial localization for other
apicomplexans and chromerids based on their targeting signals. Three Nbp35 genes were
observed in the unrelated free-living archamoebae M. balamuthi, from which a single Nbp35
paralog possesses the mitochondrial/hydrogenosomal targeting sequence, and its
hydrogenosomal localization was supported by previous proteomic data [55]. Dual
mitosomal/cytoplasmic localization of two out of three Nbp35 paralogs was observed in
metamonad G. intestinalis [25]. A common property shared by these organisms with
mitochondrion-associated Nbp35 is that they lack Complex I and Ind1. It is tempting to
speculate that mitochondrial Nbp35 replaces Ind1 and serves in the delivery of [4Fe4S]
clusters to proteins other than Complex I subunits. However, Ind1 is highly specific for
Complex I, and its involvement in the maturation of other FeS proteins was not observed [17,
28].

Conclusions
The searches for Nbp35-like proteins across eukaryotic lineages revealed mitochondrial HCF-101 homologs that are present exclusively in SAR, Haptista, and Cryptista. Thus, the presence of mHCF101 and lack of Ind1 are the first nonplastidial common features of these lineages formerly grouped under chromalveolates. Phylogeny of the HCF101 protein suggested that both mHCF101 and chHCF101 are paralogs and that an ancestral HCF101 more likely was gained by LGT from bacteria than via EGT.

**Methods**

*Toxoplasma gondii* cultivation, genetic manipulation, and microscopy.

Tachyzoites of *T. gondii* derived from strain RH were cultivated and genetically manipulated as described previously [56]. HCF101 (TGME49_318590) and Nbp35 (TGME49_280730) coding sequences were amplified from *T. gondii* cDNA and cloned in frame with a triple hemagglutinin (HA) epitope tag at the 3’ end into plasmid pDt7s4HA. The constructs were transiently transfected into the *T. gondii Δku80/TATi* strain [57] using a BTX ECM 630 electroporator (Harward Apparatus). Confluent human foreskin fibroblasts (HFF) were infected with transfected parasites and fixed after 24 hours of infection with 4% formaldehyde and permeabilized with 0.2% Triton X-100. Immunofluorescence microscopy was performed using the primary antibodies anti-HA (Roche), mouse anti-*T. gondii* mitochondrial F1-ATPase [58], and rabbit anti-apicoplast HSP60 [59]. Secondary antibodies used were goat anti-rat Alexa Fluor 488, goat anti-mouse Alexa Fluor 546, and goat anti-rabbit Alexa Fluor 546. Images were obtained on an Applied Precision Delta Vision microscope and were deconvolved and adjusted using Softworx software (GE Healthcare).

*Tetrahymena thermophila* cultivation, genetic manipulation, and microscopy.
T. thermophila CU428 strain was cultivated axenically in SPP medium (1% proteose-peptone, 0.2% glucose, 0.1% yeast extract, and 0.003% ferric-sodium: EDTA supplied with an antibiotic-antimycotic mix (Invitrogen, Carlsbad, CA)[60]. The insertion of transgenes into the T. thermophila macronucleus was performed as described previously [61]. Genes coding for Nbp35 (XP_001033404, TThERM 0312220) and mHCF101 (XP_001007903, TThERM 00538790) were amplified from genomic DNA and inserted into the pFAP44-3HA vector [62], which allows the expression of C-terminal-3HA tagged protein under its native promoter [63]. Transfected cells were selected under an increasing concentration of paromomycin (100 µg-1000 µg per ml) and decreasing concentration of CdCl₂.

Living cells of T. thermophila were stained by Mitotracker Red CMXRos (Molecular Probes, Invitrogen) following the manufacturer’s protocol. Then, the cells were spread on polylysine-coated slides and immediately fixed using methanol, permeabilized with acetone, and immunostained by a α-HA tag rat monoclonal antibody (Roche) and Alexa Fluor 488 (green) donkey α-rat antibody (Invitrogen). Nuclei were stained with 4’,6-diamidino-2-fenylindol (DAPI). The slides were examined using an Olympus IX81 microscope equipped with an MT20 illumination system.

Phaeodactylum tricornutum and Thalassiosira pseudonana cultivation, genetic manipulation, and microscopy

P. tricornutum (Bohlin, University of Texas Culture Collection, strain 646) and Thalassiosira pseudonana Hasle et Heimdal CCMP1335 were axenically grown in artificial seawater medium, made by dissolving “Tropic marine” salt (Wartenberg, Germany) to obtain 35 units of practical salinity and enriched by Guillard’s (F/2) Marine Water Enrichment Solution. The cells were cultivated at 22°C under continuous illumination (80 mmol photons
per m² per s) with agitation (150 rpm) in 250 mL Erlenmeyer flasks to a density of approximately 7×10⁶ cells/ml.

_P. tricornutum_ genes for Nbp35 (XP_002179311), mHCF101 (Joint Genome Institute, JGI portal ID 49356), and chHCF101 (JGI portal ID 1865) and _T. pseudonana_ genes for Nbp35 (XP_002289427), mHCF101 (XP_002290238), and chHCF101 (XP_002293925) were amplified from corresponding cDNA and cloned for expression in _P. tricornutum_ with C-terminal e-GFP in vector pPHA-NR4 [36]. Biolistic transfection was carried out as described previously [64] using M10 tungsten particles and 1350 psi rupture discs together with the Bio-Rad Biolistic PDS-1000/He particle delivery system. Transfected cells were grown at 22°C under continuous illumination (80 mmol photons per m² per s) on plates containing solid f/2-medium with 1.3% agar, 1.5 mM NH₄⁺ as the sole nitrogen source and 75 µg/ml Zeocin™ as a selection marker. Protein expression under the control of the nitrate reductase promoter (pPha-NR vector) was induced by cultivation on 0.9 mM NO₃ for 2 days.

Transformants were analyzed with a Leica TCSSP2 confocal laser scanning microscope. Mitochondrial localization was verified with MitoTracker® Orange CMTMRos (Life Technologies). The fluorescence of enhanced green fluorescent protein (eGFP) and chlorophyll was excited with an argon laser (65 mW) at 488 nm and detected with two photomultiplier tubes at bandwidths of 500 to 520 nm and 625 to 720 nm for eGFP and chlorophyll fluorescence, respectively. MitoTracker® Orange CMTMRos was excited with a HeNe(1.2 mW) laser at 543 nm, and emission was detected at 560-590 nm. Pictures were assembled in ImageJ (http://imagej.nih.gov/ij/index.html) using the Loci Bio-Formats plug-in (http://www.openmicroscopy.org/site/products/bio-formats).

Searches for protein sequences and targeting predictions
Homologs of Ind1, Nbp35, Cfd1, and Hcf101 proteins were retrieved using the BLAST algorithm [65] from the NCBI nr database (https://www.ncbi.nlm.nih.gov), JGI genome (https://genome.jgi.doe.gov/portal/), iMicrobe (https://www.imicrobe.us/), and Uniprot (https://www.uniprot.org/). Genes for Cyanophora paradoxa were obtained from the database at http://cyanophora.rutgers.edu/cyanophora/home.php. Protein sequences for Eutreptiella gymnastica and Pyramimonas parkeae were kindly provided by Vladimír Hampl (Charles University, Prague, Czech Republic), Euglena gracilis sequences by Marek Eliáš (University of Ostrava, Czech Republic), and Chromera velia and Vitrella brassicaformis by Miroslav Oborník (Biology center AS, Czech Budweis, Czech Republic). For each retrieved protein sequence, a given database, dataset, and gene number is indicated in supplementary Table S1.

Protein sequences of four Nbp35-like protein categories were aligned using the MUSCLE algorithm [66] in Geneious® 11.1.5 software with default settings. Protein sequences with incomplete N-terminal parts were excluded from further protein localization analysis. In a minority of cases, when N-terminal methionine was absent, but we identified methionine within the first 10 amino acids of the N-terminus, we shortened the sequence, and localization prediction was carried with lower confidence as indicated in Table 1. Subcellular targeting of proteins was predicted using TargetP-1.1 ([67], http://www.cbs.dtu.dk/services/TargetP-1.1/index.php); TargetP- 2.0 ([68], http://www.cbs.dtu.dk/services/TargetP/); DeepLoc-1.0 ([69], http://www.cbs.dtu.dk/services/DeepLoc/, accurate Profiles protein model); MitoFates ([70], http://mitf.cbrc.jp/MitoFates/cgi-bin/top.cgi); MitoProt ([71], https://ihg.gsf.de/ihg/mitoprot.html); SignalP 4.1 ([72], http://www.cbs.dtu.dk/services/SignalP-4.1/); SignalP 5 ([73], http://www.cbs.dtu.dk/services/SignalP/); Phobius ([74], http://phobius.sbc.su.se/); PSORT II ([75], https://psort.hgc.jp/form2.html); ChloroP ([76], https://psort.hgc.jp/form2.html).
Multiloc ([78], https://omictools.com/multiloc-tool); and PlasmoAP ([79]; https://plasmodb.org/plasmo/plasmoap.jsp). Furthermore, proteins with detected signal peptides were shortened according to the predicted cleavage site of signal peptidase (SignalP 5, HECTAR, and TargetP 2 programs), and the presence of subsequent putative transit peptide was detected with the MitoFates, TargetP2, and ChloroP algorithms. A search for the motif of transit peptide cleavage by stromal processing peptidase was carried as described [80–82].

Phylogenetic analysis

For the initial analysis of Nbp-35-like proteins (Fig. 4), homologs of the ParA domain (PF10609) from the Pfam database were searched in the Uniprot database using HMMER (version 3). A total of 22328 sequences with e-values below the 1e-50 cutoff were selected. Selected sequences were grouped into groups that share 90% sequence identity using CD-HIT, and for each such group, one sequence was selected to reduce redundancy, resulting in a dataset of 9139 sequences. Sequences were then aligned using MAFFT [83] with default settings, and the multiple sequence alignment was trimmed using BMGE [84] with the BLOSUM30 matrix and a block size of one, resulting in an alignment with 189 aligned amino acid positions. Sequences that were aligned at less than 126 positions (more than 63 gaps) were removed from the dataset, resulting in 8440 sequences. These were again realigned and trimmed resulting in an alignment with 185 aligned amino acid positions. A phylogenetic tree was then inferred using FastTree [85] with default settings.

For detailed HCF101 analysis (Fig. 5), a dataset of 107 HCF101 proteins and their homologs was manually assembled. The sequences were aligned using MAFFT [83] with “–
maxiterate 1000” and “–local pair” parameters. The alignment was trimmed using BMGE [84] with the BLOSUM30 matrix and a block size of one, which resulted in 311 aligned amino acid positions. A maximum likelihood phylogenetic tree was inferred using IQ-Tree (version 1.6) [86] with the best selected mixture model LG+C60+G, and the topology was tested using 10 000 ultrafast bootstraps. A Bayesian phylogenetic tree was inferred using PhyloBayes (ver. 3) [87] and the CAT-Poisson model, running two chains for 20 000 generations. The first 2 000 generations were discarded (burnin), and every tenth generation was sampled. The chains converged, with the maxdiff value 0.076.

Domain searches
Conserved protein domains were detected by searching sequences against the Pfam database (ver. 32) using HMMER (ver. 3)[88]. Hits with e-values below 1e-5 were considered.

Table 1. Identification of Nbp35-like proteins in selected representatives of eukaryotic lineages and prediction of their cellular localization.
| Archaeplastida | Viridiplantae | Arabidopsis thaliana | Nbp35 | Ind1 | chHCF101 |
|----------------|--------------|---------------------|-------|------|----------|
| Chlorella variabilis | Nbp35 | Ind1 | chHCF101 |
| Coccomyxa subellipsoides | Nbp35, chHCF101 | Ind1# |
| Micromonas pusilla | Nbp35 | Ind1 | chHCF101 |
| Oryza sativa | Nbp35, chHCF101 | Ind1 |
| Ostreococcus tauri | chHCF101 | Ind1 |
| Physcomitrella patens | Nbp35 | Ind1 | chHCF101 |
| Pyramimonas parkeae | Nbp35 | Ind1* | chHCF101 |
| Selaginella moellendorffii | Nbp35 | Ind1# | chHCF101# |
| Glaucophyta | Glacocystis sp. | Nbp35 | Ind1 | chHCF101* |
| | Gloeochaete wittrockiana | Nbp35*, Cfd1* | Ind1* | chHCF101 |
| Rhodophyta | Chondrus crispus | Nbp35 | na | chHCF101 |
| | Cyanidioschyzon merolae | Nbp35 | Ind1 | chHCF101 |
| | Erythrolobus madagascarensis | Nbp35* | Ind1* | chHCF101* |
| | Galderia sulphuraria | Nbp35 | Ind1 | chHCF101 |
| | Madagascaria erythrocladiodes | Nbp35* | na | chHCF101 |
| | Porphyridium aerugineum | Nbp35 | Ind1 | chHCF101* |
| | Rhodosorus marinus | Nbp35* | Ind1* | chHCF101* |
| | Timspurkia oligopyrenoides | na | Ind1 | chHCF101* |
| SARCH | | | | | |
| Alveolata | Apicomplexa | Babesia bovis | mHCF101* | Nbp35# |
| | | Cryptosporidium muris | mHCF101 | Nbp35# |
| | | Toxoplasma gondii | mHCF101 | Nbp35 |
| | | Plasmodium falciparum | mHCF101 | Nbp35# |
| | | Plasmodium yoelii | mHCF101 | Nbp35# |
| | Chromerida | Chromera vellia | mHCF101 | Nbp35# |
| | | Vitrella brassica | mHCF101 | Nbp35# |
| | | Perkinsus marinus | Nbp35# | mHCF101# |
| Dinoflagellata | | Alexandrium monilatum | Nbp35* | mHCF101* | chHCF101* |
| | | Alexandrium tamarense | Nbp35* | mHCF101 | chHCF101 |
| | | Durinska baltica | Nbp35 | na | chHCF101* |
| | | Dinophysis acuminata | Nbp35* | mHCF101* | chHCF101* |
| | | Karenia brevis | Nbp35* | mHCF101 | chHCF101 |
| | | Oxyrrhis marina | Nbp35* | mHCF101* | na |
| | | Symbiodinium sp. | Nbp35# | mHCF101* | chHCF101 |
| Ciliata | | Tetrahymena thermophila | Nbp35 | mHCF101 |
| Organism                          | Tag1   | Tag2               |
|----------------------------------|--------|--------------------|
| *Oxytricha trifallax*            | Nbp35  | mHCF101            |
| *Paramecium tetraurelia*         | Nbp35  | mHCF101            |
| *Stylonychia sp.*                | Nbp35  | mHCF101            |
| Rhizaria                         |        |                    |
| *Ammonia sp.*                    | Nbp35  | mHCF101*           |
| *Elphidium margaritaceum*        | Nbp35  | mHCF101*           |
| *Reticulomyxa filosa*            | Nbp35  | mHCF101            |
| *Paulinella chromatophora*       | na     | na                |
| Chlorarachniophyta               |        | Bacterial HCF101-like |
| *Bigelovia longifilifera*        | Nbp35  | mHCF101*           |
| *Bigelovia natans*               | Nbp35* | mHCF101*           |
| *Lotharella globosa*             | Nbp35* | mHCF101*           |
| Stramenopila                     |        |                    |
| *Oomyctota*                      |        |                    |
| *Albugo candidi*                 | Nbp35  | mHCF101#           |
| *Albugo laibochii*               | Nbp35  | mHCF101#           |
| *Aphanomyces astaci*             | Nbp35  | mHCF101            |
| *Aphanomyces invadans*           | Nbp35  | mHCF101            |
| *Phytophthora infestans*         | Nbp35  | mHCF101*           |
| *Phytophthora parasitica*        | Nbp35  | mHCF101            |
| *Saprolegnia diclina*            | Nbp35  | mHCF101            |
| Ochrophyta                        |        |                    |
| *Aureococcus anophagefferens*    | Nbp35  | mHCF101*           |
| *Ectocarpus siliculosus*         | Nbp35# | mHCF101            |
| *Phaeodactylum tricornutum*       | Nbp35  | mHCF101            |
| *Nannochloropsis gaditana*       | Nbp35  | mHCF101            |
| *Schizochytrium aggregatum*      | Nbp35  | mHCF101            |
| *Thalassiosira oceanica*         | Nbp35  | mHCF101*           |
| *Thalassiosira pseudonana*       | Nbp35  | mHCF101*           |
| *Blastocystis hominis*           | Nbp35  | mHCF101            |
| Haptista                         |        |                    |
| *Chrysochromulina polyplepis*    | Nbp35* | mHCF101*           |
| *Emiliania huxleyi*              | Nbp35  | mHCF101            |
| *Exanthemachrysis gayraliae*      | na     | mHCF101*           |
| *Gephyrocapsa oceanica*          | Nbp35  | mHCF101*           |
| *Isochrysis galbana*             | Nbp35  | mHCF101*           |
| *Pleurochrysis carterae*         | Nbp35  | mHCF101*           |
| *Prymnesium parvum*              | Nbp35* | mHCF101*           |
| *Centrohelida*                   | Nbp35* | mHCF101*           |
| Cryptista                        |        |                    |
| *Cryptophyta*                    |        |                    |
| *Chroomonas mesostigma*          | Nbp35*, Cfd1* | mHCF101* |
| *Cryptomonas curvata*            | na     | mHCF101            |
| *Cryptomonas paramecium*         | Nbp35*, Cfd1* | mHCF101 |

24
| Organism                        | Gene Localization | Cellular Localization |
|--------------------------------|-------------------|-----------------------|
| Geminigera cryophila           | Nbp35*, Cfd1      | mHCF101               |
| Guillardia theta               | Nbp35/Cfd1        | mHCF101*              |
| Hanusia phi                    | Nbp35*, Cfd1*     | mHCF101               |
| Hemiselmis rufescens           | Nbp35, Cfd1*      | mHCF101*              |
| Proteomonas sulcata            | Nbp35*, Cfd1*     | mHCF101*              |
| Rhodomonas sp.                 | Nbp35, Cfd1*      | na                    |
| Katablepharida                 | Nbp35*, Cfd1*     | mHCF101*              |
| Roombia truncata               | Nbp35, Cfd1*      | mHCF101               |
| Goniomonas                     |                    |                       |
| Goniomonas avonlea             | Nbp35*, Cfd1*     | mHCF101               |
| Goniomonas pacifica            | Nbp35, Cfd1*      | mHCF101*              |
| Proteins with experimentally verified localization are in bold. Taxons in red indicate available genome sequence, taxons in red indicate available transcriptome. * Incomplete sequence of the gene, cellular localization is predicted based on the phylogenetic analysis (Fig. 5); # prediction with low confidence; na, gene was not identified in available transcriptome. |

References

1. Tachezy J, Sánchez LB, Müller M. Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. Mol Biol Evol. 2001;18:1919–28.

2. Lill R. Function and biogenesis of iron-sulphur proteins. Nature. 2009;460:831–8.

3. Sutak R, Dolezal P, Fiumera HL, Hrdy I, Dancist A, Delgadillo-Correa M, et al. Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate
eukaryote *Trichomonas vaginalis*. Proc Natl Acad Sci U S A. 2004;101:10368–73.

4. Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, Van Der Giezen M, et al. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature. 2003;426:172–6.

5. Takahashi Y, Tokumoto U. A third bacterial system for the assembly of iron-sulfur clusters with homologs in Archaea and plastids. J Biol Chem. 2002;277:28380–3.

6. Novák Vanclová AMG, Zoltner M, Kelly S, Soukal P, Záhonová K, Füssy Z, et al. Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. New Phytol. 2020;225:1578–92. doi:10.1111/nph.16237.

7. Grosche C, Diehl A, Rensing SA, Maier UG. Iron-sulfur cluster biosynthesis in algae with complex plastids. Genome Biol Evol. 2018;10:2061–71. doi:10.1093/gbe/evy156.

8. Füssy Z, Oborník M. Complex endosymbioses I: From primary to complex plastids, multiple independent events. In: Methods in Molecular Biology. Humana Press Inc.; 2018. p. 17–35. doi:10.1007/978-1-4939-8654-5_2.

9. Freibert SA, Goldberg A V., Hacker C, Molik S, Dean P, Williams TA, et al. Evolutionary conservation and in vitro reconstitution of microsporidian iron-sulfur cluster biosynthesis. Nat Commun. 2017;8:13932. doi:10.1038/ncomms13932.

10. Tsaousis AD, Gentekaki E, Eme L, Gaston D, Roger AJ. Evolution of the cytosolic iron-sulfur cluster assembly machinery in Blastocystis species and other microbial eukaryotes. Eukaryot Cell. 2014;13:143–53. doi:10.1128/EC.00158-13.

11. Gill EE, Diaz-Trivino S, Barbera MJ, Silberman JD, Stechmann A, Gaston D, et al. Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba balamuthi*. Mol
12. Nývtová E, Šuták R, Harant K, Šedínová M, Hrdý I, Pačes J, et al. NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of Mastigamoeba balamuthi. Proc Natl Acad Sci U S A. 2013;110:7371–6.

13. Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, et al. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist Pygsuia. Curr Biol. 2014;24:1176–86. doi:10.1016/j.cub.2014.04.033.

14. Leger MM, Eme L, Hug LA, Roger AJ. Novel hydrogenosomes in the microaerophilic jakobid Stygiella incarcerata. Mol Biol Evol. 2016;33:2318–36. doi:10.1093/molbev/msw103.

15. Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, et al. A eukaryote without a mitochondrial organelle. Curr Biol. 2016;26:1274–84.

16. Pandey AK, Pain J, Dancis A, Pain D. Mitochondria export iron-sulfur and sulfur intermediates to the cytoplasm for iron-sulfur cluster assembly and tRNA thiolation in yeast. J Biol Chem. 2019;294:9489–502.

17. Bych K, Kerscher S, Netz DJ, Pierik AJ, Zwicker K, Huynen MA, et al. The iron-sulphur protein Ind1 is required for effective complex I assembly. EMBO J. 2008;27:1736–46.

18. Hrdý I, Hirt RP, Dolezal P, Bardonová L, Foster PG, Tachezy J, et al. Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature. 2004;432:618–22.

19. Lezhneva L, Amann K, Meurer J. The universally conserved HCF101 protein is involved in assembly of [4Fe-4S]-cluster-containing complexes in Arabidopsis thaliana chloroplasts.
20. Schwenkert S, Netz DJA, Frazzon J, Pierik AJ, Bill E, Gross J, et al. Chloroplast HCF101 is a scaffold protein for [4Fe-4S] cluster assembly. Biochem J. 2009;425:207–14. doi:10.1042/BJ20091290.

21. Keeling PJ. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci. 2010;365:729–48. doi:10.1098/rstb.2009.0103.

22. Archibald JM. The puzzle of plastid evolution. Curr Biol. 2009;19:R81–8.

23. Pala ZR, Saxena V, Saggu GS, Garg S. Recent advances in the [Fe–S] cluster biogenesis (SUF) pathway functional in the apicoplast of Plasmodium. Trends Parasitol. 2018;34:800–9. doi:10.1016/j.pt.2018.05.010.

24. Seeber F, Soldati-Favre D. Metabolic pathways in the apicoplast of Apicomplexa. Int Rev Cell Mol Biol. 2010;281:161–228. doi:10.1016/S1937-6448(10)81005-6.

25. Pyrih J, Pyrihová E, Kolísko M, Stojanovová D, Basu S, Harant K, et al. Minimal cytosolic iron-sulfur cluster assembly machinery of Giardia intestinalis is partially associated with mitosomes. Mol Microbiol. 2016;102:701–14.

26. Opperdoes FR, Michels PAM. Complex I of Trypanosomatidae: does it exist? Trends Parasitol. 2008;24:310–7. doi:10.1016/j.pt.2008.03.013.

27. Leipe DD, Wolf YI, Koonin E V., Aravind L. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol. 2002;317:41–72.

28. Sheftel AD, Stehling O, Pierik AJ, Netz DJA, Kerscher S, Elsässer H-P, et al. Human Ind1, an iron-sulfur cluster assembly factor for respiratory complex I. Mol Cell Biol. 2009;29:6059–73.
Human CIA2A-FAM96A and CIA2B-FAM96B integrate iron homeostasis and maturation of different subsets of cytosolic-nuclear iron-sulfur proteins. Cell Metab. 2013;18:187–98.

The DUF59 family gene AE7 acts in the cytosolic iron-sulfur cluster assembly pathway to maintain nuclear genome integrity in Arabidopsis. Plant Cell. 2012;24:4135–48.

The DUF59 containing protein SufT is involved in the maturation of iron-sulfur (FeS) proteins during conditions of high FeS cofactor demand in Staphylococcus aureus. PLOS Genet. 2016;12:e1006233. doi:10.1371/journal.pgen.1006233.

Short oligopeptides with three cysteine residues as models of sulphur-rich Cu(i)- and Hg(ii)-binding sites in proteins. Metallomics. 2018;10:1232–44.

Untangling the early diversification of eukaryotes: A phylogenomic study of the evolutionary origins of centrohelida, haptophyta and cryptista. Proc R Soc B Biol Sci. 2016;283.

The New Tree of Eukaryotes. Trends Ecol Evol. 2020;35:43–55.

Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Eukaryot Microbiol. 1999;46:347–66.

Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of...
30. ubiquitin-dependent translocation from degradation. Eukaryot Cell. 2012;11:1472–81.
doi:10.1128/EC.00183-12.

37. Felsner G, Sommer MS, Gruenheit N, Hempel F, Moog D, Zauner S, et al. ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. Genome Biol Evol. 2011;3:140–50.
doi:10.1093/gbe/evq074.

38. Sakamoto H, Suzuki S, Nagamune K, Kita K, Matsuzaki M. Investigation into the physiological significance of the phytohormone abscisic acid in Perkinsus marinus, an oyster parasite harboring a nonphotosynthetic plastid. J Eukaryot Microbiol. 2017;64:440–6.
doi:10.1111/jeu.12379.

39. Reyes-Prieto A, Moustafa A, Bhattacharya D. Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. Curr Biol. 2008;18:956–62.

40. Keeling PJ. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu Rev Plant Biol. 2013;64:583–607.

41. Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. The evolution of photosynthesis in chromist algae through serial endosymbioses. Nat Commun. 2014;5.

42. Burki F, Okamoto N, Pombert JF, Keeling PJ. The evolutionary history of haptophytes and cryptophytes: Phylogenomic evidence for separate origins. Proc R Soc B Biol Sci. 2012;279:2246–54. doi:10.1098/rspb.2011.2301.

43. Kite GC, Dodge JD. Structural organization of plastid DNA in two anomalously pigmented dinoflagellates. J Phycol. 1985;21:50–6. doi:10.1111/j.0022-3646.1985.00050.x.

44. Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, et al.
Phylogenetic analyses indicate flint the 19’hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. Mol Biol Evol. 2000;17:718–29. doi:10.1093/oxfordjournals.molbev.a026350.

45. Burki F, Imanian B, Hehenberger E, Hirakawa Y, Maruyama S, Keeling PJ. Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. Eukaryot Cell. 2014;13:246–55.

46. Kamikawa R, Yazaki E, Tahara M, Sakura T, Matsuo E, Nagamune K, et al. Fates of evolutionarily distinct, plastid-type glyceraldehyde 3-phosphate dehydrogenase genes in kareniacean dinoflagellates. J Eukaryot Microbiol. 2018;65:669–78.

47. Hwan SY, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. Mol Biol Evol. 2005;22:1299–308. doi:10.1093/molbev/msi118.

48. Dorrell RG, Howe CJ. Integration of plastids with their hosts: Lessons learned from dinoflagellates. Proc Natl Acad Sci U S A. 2015;112:10247–54. doi:10.1073/pnas.1421380112.

49. Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, et al. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr Biol. 2004;14:213–8.

50. Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, et al. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. Mol Biol Evol. 2008;25:2653–67. doi:10.1093/molbev/msn206.

51. Ishida K, Green BR, Cavalier-Smith T. Diversification of a chimaeric algal group, the chlorarachniophytes: Phylogeny of nuclear and nucleomorph small-subunit rRNA genes. Mol
52. Curtis BA, Tanifuji G, Maruyama S, Gile GH, Hopkins JF, Eveleigh RJM, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature. 2012;492:59–65. doi:10.1038/nature11681.

53. Ponce-Toledo RI, Moreira D, López-García P, Deschamps P. Secondary plastids of euglenids and chlorarachniophytes function with a mix of genes of red and green algal ancestry. Mol Biol Evol. 2018;35:2198–204. doi:10.1093/molbev/msy121.

54. Archibald JM, Rogers MB, Toop M, Ishida K ichiro, Keeling PJ. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga Bigelowiella natans. Proc Natl Acad Sci U S A. 2003;100:7678–83. doi:10.1073/pnas.1230951100.

55. Le T, Žárský V, Nývltová E, Rada P, Harant K, Vancová M, et al. Anaerobic peroxisomes in Mastigamoeba balamuthi. Proc Natl Acad Sci. 2020;117:2065-75. doi:10.1073/pnas.1909755117.

56. Striepen B, Soldati D. Genetic manipulation of Toxoplasma gondii. In: Toxoplasma gondii. Academic Press; 2007. p. 391–418.

57. Sheiner L, Demerly JL, Poulsen N, Beatty WL, Lucas O, Behnke MS, et al. A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. PLoS Pathog. 2011;7.

58. Chen AL, Moon AS, Bell HN, Huang AS, Vashisht AA, Toh JY, et al. Novel insights into the composition and function of the Toxoplasma IMC sutures. Cell Microbiol. 2017;19. doi:10.1111/cmi.12678.

59. Agrawal S, van Dooren GG, Beatty WL, Striepen B. Genetic evidence that an
endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins. J Biol Chem. 2009;284:33683–91.

60. Gorovsky MA, Yao MC, Keevert JB, Pleger GL. Chapter 16 Isolation of micro- and macronuclei of *Tetrahymena pyriformis*. Methods Cell Biol. 1975;9 C:311–27. doi:10.1016/S0091-679X(08)60080-1.

61. Wloga D, Camba A, Rogowski K, Manning G, Jerka-Dziadosz M, Gaertig J. Members of the NIMA-related kinase family promote disassembly of cilia by multiple mechanisms. Mol Biol Cell. 2006;17:2799–810.

62. Urbanska P, Joachimiak E, Bazan R, Fu G, Poprzeczko M, Fabczak H, et al. Ciliary proteins Fap43 and Fap44 interact with each other and are essential for proper cilia and flagella beating. Cell Mol Life Sci. 2018;75:4479–93. doi:10.1007/s00018-018-2819-7.

63. Dave D, Wloga D, Gaertig J. Manipulating ciliary protein-encoding genes in *Tetrahymena thermophila*. Methods Cell Biol. 2009;93:1–20. doi:10.1016/S0091-679X(08)93001-6.

64. Hempel F, Bozarth AS, Lindenkamp N, Klingl A, Zauner S, Linne U, et al. Microalgae as bioreactors for bioplastic production. Microb Cell Fact. 2011;10.

65. Altschul SF, Madden TL, Schaffer AA, Zhang JP, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. isi:A1997XU7930002.

66. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. isi:000220487200025.

67. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2:953–71.
68. Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance. 2019;2. doi:10.26508/lsa.201900429.

69. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics (Oxford, England). 2017;33:3387–95. doi:10.1093/bioinformatics/btx431.

70. Fukasawa Y, Tsuji J, Fu S-C, Tomii K, Horton P, Imai K. MitoFates: Improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol Cell Proteomics. 2015;14:1113–26.

71. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem. 1996;241:779–86.

72. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nature Methods. 2011;8:785–6. doi:10.1038/nmeth.1701.

73. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–3. doi:10.1038/s41587-019-0036-z.

74. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338:1027–36.

75. Nakai K, Horton P. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci. 1999;24:34–5. doi:10.1016/S0968-0004(98)01336-X.
76. Emanuelsson O, Nielsen H, Heijne G Von. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 1999;8:978–84. doi:10.1110/ps.8.5.978.

77. Gschloessl B, Guermeur Y, Cock JM. HECTAR: A method to predict subcellular targeting in heterokonts. BMC Bioinformatics. 2008;9:393. doi:10.1186/1471-2105-9-393.

78. Höglund A, Dönnes P, Blum T, Adolph HW, Kohlbacher O. MultiLoc: Prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. Bioinformatics. 2006;22:1158–65. doi:10.1093/bioinformatics/btl002.

79. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, et al. Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. Science. 2003;299:705–8. doi:10.1126/science.1078599.

80. Apt KE, Zaslavkaia L, Lippmeier JC, Lang M, Kilian O, Wetherbee R, et al. In vivo characterization of diatom multipartite plastid targeting signals. J Cell Sci. 2002;115:4061–9. doi:10.1242/jcs.00092.

81. Woehle C, Dagan T, Martin WF, Gould SB. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related Chromera velia. Genome Biol Evol. 2011;3:1220–30. doi:10.1093/gbe/evr100.

82. Huesgen PF, Alami M, Lange PF, Foster LJ, Schröder WP, Overall CM, et al. Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. PLoS One. 2013;8:e74483. doi:10.1371/journal.pone.0074483.

83. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
84. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010;10:210. doi:10.1186/1471-2148-10-210.

85. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One. 2010.

86. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. doi:10.1093/molbev/msu300.

87. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. 2009;25:2286–8.

88. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 Update. Nucleic Acids Res. 2015;43:W30–8.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials
All data generated or analysed during this study are included in this published article and its supplementary information files.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

Laboratory of JT was supported by NPU II (LQ1604) provided by the Ministry of Education, Youth and Sport (MEYS) of the Czech Republic, CePaViP (CZ.02.1.01/0.0/0.0/16_019/0000759) provided by ERD Funds, and MICOBION funded from EU H2020 (No 810224). CG and UGM were supported by the LOEWE Center for Synthetic Microbiology (Synmikro).

**Authors' contributions**

JP and JT conceived the study. JP and VŽ performed bioinformatic analyses, JDF, BS, CG. DW, UGM performed cell localization studies, JP and JT wrote the paper. All authors read and approved the final manuscript.

**Legends to figures**

Figure 1. Predictions of N-terminal targeting sequences for chloroplast (chHCF101) and mitochondrial versions of HCF101 (mHCF101) in selected alveolates and stramenopiles. Red color highlights the mitochondrial leader sequence with cleavage sites predicted with Mitofates (star), TargetP 2 (circle), and PSORT II (triangle). Yellow color indicates a signal peptide, which is cleaved right before the phenylalanine residue, a typical feature for signal peptide cleavage in diatoms. This cleavage site was predicted with high support by the TargetP, Signal P, and HECTAR algorithms. Green represents a predicted transit peptide with
two possible cleavage sites for stromal processing peptidase, manually predicted based on previous studies [82]. Gray color highlights conserved residues.

Figure 2. Localization of Nbp35-like proteins in diatoms. Genes for chHCF101, mHCF101, and Nbp35 from *P. tricornutum* and *T. pseudonana* were expressed in *P. tricornutum* with C-terminal e-GFP tag (green). PAF, plastid autofluorescence (red); MitoT, MitoTracker Orange (blue); DIC, differential interference contrast. Scale bar 10µm.

Figure 3. Localization of Nbp35 and mHCF101 in *T. thermophila* and *T. gondii*. Nbp35 and mHCF101 were expressed under the control of their respective native promoter with a C-terminal HA tag (green). Specific polyclonal antibodies against F1-ATPase (red), and HSP60 (red) were used as mitochondrial and apicoplastidal markers, respectively. MitoT, Mitotracker Red, (red); DIC, differential interference contrast. Scale bar 10 µm (*T. thermophila*) and 5 µm (*T. gondii*).

Figure 4. A phylogenetic tree of the Nbp35-like proteins (the ParA family). The tree was built using FastTree, and nodes with support below 0.9 were collapsed (see materials and methods). The leaves are color coded to indicate taxonomy (8840 sequences), with annotations of selected sequences. The scale bar represents the estimated number of amino acid substitutions per site.

Figure 5. Phylogeny of HC101 proteins. The tree was built using the PhyloBayes CAT-Poisson mixture model. Numbers at nodes of the tree indicate statistical support in the form of posterior probability of the PhyloBayes analysis and an ultrafast bootstrap of the IQ-Tree analysis (see materials and methods). The scale bar represents the estimated number of amino acid substitutions per site. The conserved cysteine motif in the DUF971 domain is displayed for each protein sequence when available.
Figure 6. Scheme of HCF101 evolution. A. HCF101 distribution explained via the Chromalveolate hypothesis. Upon the acquisition of HCF101-like protein via LGT from bacteria to the ancestor of Archeplastida, chHCF101 was established in the plastid. Euglenozoa gained HCF101 via secondary endosymbiosis with a donor containing green plastid. A common ancestor of Chromalveolata gained red plastid via secondary endosymbiosis, the gene for chHCF101 was duplicated and one copy was targeted to mitochondria (mHCF101), where it replaced the Ind1 gene. mHCF101 is common to all chromalveolates, while several lineages lost chHCF101 together with loss of the secondary plastid (Cryptosporidium, Ciliates, Oomycota, centrohelids, catablepharids). B. HCF101 distribution explained via ‘multiple secondary endosymbiosis’. Mitochondrially localized HCF101 together with Ind1 was present in a common ancestor of Archeplastida, SAR, Cryptista, and Haptista. Then, in (i) Cryptista and in (ii) a common ancestor of Haptista and SAR, the Ind1 gene was lost, whereas the Archeplastida gene for the Ind1 protein remained in the mitochondria, and mHCF101 was retargeted to the plastid. Then, chHCF101 was introduced to Cryptophytes, Haptophytes, and certain SAR groups via multiple secondary endosymbiosis. chHCF101 is absent in lineages that did not experience secondary endosymbiosis.

PR, protein retargeting; GL, gene loss; GD, gene duplication; PL, plastid loss; PES, primary endosymbiosis; SES, secondary endosymbiosis; TES, tertiary endosymbiosis.

Additional files

Additional file 1, Table S1.xls

List of retrieved protein sequences that were used for cellular localization predictions. Sequence, Nbp35-like protein category, organism, its taxonomy, and source of the sequence in various databases are highlighted in columns A to L. Completeness of the sequence is
indicated in Column M. In the case that the protein sequence was manually corrected, the
trimmed sequence is displayed in column N. Columns O to P indicate final protein
localization prediction. Columns R to Z, AA to AJ, and AK to AO indicate support for the
presence of signal peptide, mitochondrial leader sequence, or plastid transit peptide,
respectively.

Additional file 2, Table S2.xls

Predictions of domain structure in Nbp35 homologs included in phylogenetic analysis (Fig.
4).
