RESEARCH ARTICLE

Reticulate evolution in eukaryotes: Origin and evolution of the nitrate assimilation pathway

Eduard Ocaña-Pallarès, Sebastián R. Najle, Claudio Scazzocchio, Iñaki Ruiz-Trillo

1 Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Catalonia, Spain, 2 Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET) and Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Ocampo y Esmeralda s/n, Rosario S2000FHQ, Argentina, 3 Department of Microbiology, Imperial College, London, United Kingdom, 4 Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France, 5 Departamento de Genética, Microbiología e Estadística, Facultad de Biología, Universitat de Barcelona (UB), Barcelona, Catalonia, Spain, 6 ICREA, Barcelona, Catalonia, Spain

* eduard.ocana@ibe.upf-csic.es (EOP); inaki.ruiz@ibe.upf-csic.es (IRT)

Abstract

Genes and genomes can evolve through interchanging genetic material, this leading to reticulate evolutionary patterns. However, the importance of reticulate evolution in eukaryotes, and in particular of horizontal gene transfer (HGT), remains controversial. Given that metabolic pathways with taxonomically-patchy distributions can be indicative of HGT events, the eukaryotic nitrate assimilation pathway is an ideal object of investigation, as previous results revealed a patchy distribution and suggested that the nitrate assimilation cluster of dikaryotic fungi (Opisthokonta) could have been originated and transferred from a lineage leading to Oomycota (Stramenopiles). We studied the origin and evolution of this pathway through both multi-scale bioinformatic and experimental approaches. Our taxon-rich genomic screening shows that nitrate assimilation is present in more lineages than previously reported, although being restricted to autotrophs and osmotrophs. The phylogenies indicate a pervasive role of HGT, with three bacterial transfers contributing to the pathway origin, and at least seven well-supported transfers between eukaryotes. In particular, we propose a distinct and more complex HGT path between Opisthokonta and Stramenopiles than the one previously suggested, involving at least two transfers of a nitrate assimilation gene cluster. We also found that gene fusion played an essential role in this evolutionary history, underlying the origin of the canonical eukaryotic nitrate reductase, and of a chimeric nitrate reductase in Ichthyosporea (Opisthokonta). We show that the ichthyosporean pathway, including this novel nitrate reductase, is physiologically active and transcriptionally co-regulated, responding to different nitrogen sources; similarly to distant eukaryotes with independent HGT-acquisitions of the pathway. This indicates that this pattern of transcriptional control evolved convergently in eukaryotes, favoring the proper integration of the pathway in the metabolic landscape. Our results highlight the importance of reticulate evolution in eukaryotes, by showing the crucial contribution of HGT and gene fusion in the evolutionary history of the nitrate assimilation pathway.
Author summary

One of the most relevant findings in evolution was that lineages, either genes or genomes, can evolve through interchanging genetic material. For example, exon shuffling can lead to genes with complete novel functions, and genomes can acquire novel functionalities by means of horizontal gene transfer (HGT). Whereas HGT is known to be an important driver of metabolic remodelling and ecological adaptations in Bacteria, its importance and prevalence in eukaryotes remains controversial. We show that HGT played a major role in the origin and evolution of the eukaryotic nitrate assimilation pathway, with several bacteria-to-eukaryote and eukaryote-to-eukaryote transfers promoting the acquisition of this ecologically-relevant pathway to autotrophs and to distinct groups of osmotrophs. Moreover, we also show that gene fusion was important in this evolutionary history, underlying the origin of the canonical eukaryotic nitrate reductase, but also of a non-canonical nitrate reductase that we describe in Ichthyosporea, a poorly-characterized eukaryotic group that includes many parasitic species. In conclusion, our results highlight the importance of reticulate evolution in eukaryotes, by showing the contribution of HGT and gene fusion in the evolutionary history of the nitrate assimilation pathway.

Introduction

One of the most significant advances in evolution was the realization that lineages, either genes or genomes, can also evolve through interchanging genetic material, this leading to reticulate evolutionary patterns [1,2]. Reticulate evolution, and in particular horizontal gene transfer (HGT), is widely accepted as an important mechanism in prokaryotes [3]. However, its occurrence is still subject to controversy in eukaryotes, and its prevalence and mechanistic basis are active areas of study [4,5]. The finding of homologous genes in distantly related lineages may suggest the occurrence of HGT events [6]. However, taxonomically-patchy distributed genes can also be the result of secondary losses. Hence, the most accurate methodology for HGT detection consists of finding topological incongruences between the reconstructed phylogenetic trees and the species phylogeny [7].

Adaptation to new environments requires metabolic remodeling, and HGT of metabolic genes between prokaryotes occurs at a higher rate than that of informational genes [8]; which may facilitate the recipients’ rapid adaptation [9]. Numerous metabolic pathways in eukaryotes are of bacterial origin [6], transferred from endosymbionts [10]; and many proposed HGTs between eukaryotes also involve metabolic genes [11–13]. Hence, patchily distributed metabolic pathways make good candidate subjects for investigating possible HGT events in eukaryotes.

The nitrate assimilation pathway was shown to be patchily distributed in eukaryotes [14]. The ability to use nitrate as a nitrogen source is not essential, but valuable in nitrate-rich environments [15,16]. In order to reduce nitrate to ammonium, a specific pathway is required, involving, at least, a nitrate transporter, a nitrate reductase and a nitrite reductase (Nitrate Assimilation Proteins, NAPs) [17]. In eukaryotes, NAPs were first characterized in plants and fungi and later identified in other eukaryotes, including green and red alga, diatoms and Oomycota [14,18]. A study published a decade ago proposed that the nitrate assimilation pathway characteristic of many fungal species originated in a stramenopiles lineage leading to Oomycota, from where it would have been transferred to the root of Dikarya (Fungi). The authors also hypothesized that the acquisition of this metabolic pathway might have been an
important innovation for the colonization of dry land by this fungal group [14]. However, the absence of genomic data from many eukaryotic groups left uncertainty surrounding this proposed HGT event as well as the degree to which HGT influenced the evolutionary history of this pathway in eukaryotes. We therefore performed an extensive survey of NAPs and NAP clusters in order to understand the origins and the evolution of the eukaryotic nitrate assimilation pathway.

Our updated taxon sampling extends the presence of this ecologically-relevant pathway to many previously unsampled lineages, showing a patchy distribution that overlaps with the distribution of autotrophy and osmotrophy in the eukaryotic tree. The reconstructed history indicates a pervasive role of HGT underlying this patchy distribution, with three independent bacterial transfers contributing to the origins of the pathway and at least seven well-supported transfers of NAPs and NAP clusters between eukaryotes. Gene fusion was also crucial in the evolution of this pathway, underlying the origin of the canonical eukaryotic nitrate reductase, as well as a nitrate reductase of chimeric origin found in the NAP clusters of two ichthyosporeans. Finally, we demonstrate that this cluster is functional in the ichthyosporean *Sphaeroforma arctica*, with NAPs showing a strong co-regulation in response to environmental nitrogen sources. The similarities of this transcriptional control with that shown for many lineages with distinct horizontal acquisitions of the pathway indicate that this regulatory response has convergently evolved multiple times in eukaryotes.

**Results**

**NAP genes in eukaryotes**

The minimal metabolic pathway required to incorporate nitrate into the cell and reduce it into ammonium includes a nitrate transporter, a nitrate reductase and a nitrite reductase (Fig 1) [18]. The nitrate transporter NRT2 and the nitrate reductase EUKNR are involved in the first two steps of the pathway in all the eukaryotes in which this metabolism has been studied. For the third and last step of the pathway, two nitrite reductases have been characterized in eukaryotes: a chloroplastic ferredoxin-dependent enzyme (Fd-NIR, characterized in land plants and green algae); and a cytoplasmic NAD(P)H dependent cytosolic enzyme (NAD(P)H-NIR, characterized in fungi).

We screened NAP genes in a taxon sampling designed to cover the broadest possible eukaryotic diversity (Fig 2). Among the 60 taxa with at least one NAP gene detected, 47 have the complete pathway (i.e. the transporter, the nitrate reductase and one of the two nitrite reductases; see S1 Fig and Table A in S1 Supporting information). The distribution of NAP genes across eukaryotes is highly correlated, as expected for genes involved in the same pathway (S2A Fig). However, considering only taxa with at least one NAP gene, the two nitrite reductases, NAD(P)H-nir and Fd-nir, show an almost completely anti-correlated distribution (S2B Fig). Fd-nir is restricted to autotrophic lineages (including facultative autotrophs), as expected for a chloroplast enzyme [19]. In contrast, NAD(P)H-nir is mostly distributed along heterotrophs, although it is also present in the myzozoans *Symbiodinium minutum* and *Vitrella brassicaformis*, in which the Fd-nir is absent; and in four Ochrophyta species, in which both nitrite reductases are present.

The widespread and patchy distribution of NAP genes is correlated with the distribution of different nutrient acquisition strategies within the eukaryotic tree (S2C Fig). We found NAP genes in all the sampled autotrophs (see green circles in Fig 2). This includes taxa from groups whose plastid originated from a cyanobacterial endosymbiont (primary plastids): Glaucophyta, Rhodophyta and Chloroplastida; as well as algal groups whose plastid originated from an eukaryotic endosymbiont (complex plastids): Haptophyta, Cryptophyta, Chlorarachniophyta,
Among heterotrophs, we found the complete pathway in Fungi and Oomycota, as reported in previous studies, but also in Teretospora and Labyrinthulea. These groups are phylogenetically distant but share many analogous cellular and ecological features related to their proposed convergent evolution towards an osmotrophic lifestyle [21] (Fungal-like osmotrophs, see brown circles in Fig 2). We did not find the entire nitrate assimilation pathway in any of the phagotrophic lineages sampled (S2C Fig).

The distinct origins of NAP genes in eukaryotes

Previous studies proposed a bacterial origin for the transporter and the two nitrite reductases [14]. However, which particular bacterial group(s) were the possible donors of these three NAP genes was not determined. We investigated the origin of Fd-nir, NAD(P)H-nir and nrt2 in eukaryotes using a comprehensive and taxonomically representative prokaryotic dataset (Fig 3).

The bacterial donors of Fd-nir, NAD(P)H-nir and nrt2. The reconstructed phylogenies show in all cases a well-supported monophyletic clade that includes all eukaryotic sequences (see dark green and light blue clades in Fig 3), suggesting that each eukaryotic NAP descends from a single acquisition from prokaryotes. In particular, the three NAPs would have been transferred from Bacteria, given the distal branching of archaeal sequences (dark purple clades) respect to eukaryotes.

Our phylogeny supports that the eukaryotic nitrite reductase Fd-NIR descends from Cyanobacteria (light green clades), as was previously suggested based on sequence-similarity analyses [22] (100% UFBoot; Fig 3 and S3 Fig). Because of its cyanobacterial origin and given that Fd-NIR activity has been located in the chloroplast, it is tempting to propose that Fd-nir was transferred to eukaryotes from the cyanobacterial endosymbiont from which all the primary
Fig 2. Distribution of NAPs among 172 sampled eukaryotic genomes. The evolutionary relationships between the sampled species, represented in a cladogram, were constructed from recent bibliographical references (see Materials and methods section). Species names were colored according to the taxonomic groups to which they belong. The presence of each NAP in each taxon is shown with symbols. Black symbols indicate reticulate evolution and the eukaryotic nitrate assimilation pathway.
plastids originated. However, not all the proteins of plastidic activity originated from this organelle [23], so it remains unclear whether Fd-nir originated from the endosymbiont or not. If Fd-nir is of plastidic origin, we would then expect a similar phylogenetic position of the eukaryotic Fd-NIR in relation to Cyanobacteria as that found in the phylogenies of the photosystem II subunit III and the ribosomal protein L1; two genes of bona fide plastidic origin (encoded in the plastid genome of Cyanophora paradoxa [24]). The branching pattern of eukaryotic sequences in Fd-NIR and in the phylogenies of these two plastidic genes suggest an early-branching cyanobacterial lineage as the donor in all cases (S3 Fig, S4 Fig and S5 Fig). Notwithstanding whether plastids originated from an early or a deep cyanobacterial lineage [24,25], we interpret the similar phylogenetic relationships between eukaryotes and Cyanobacteria in all our phylogenies as moderate support for a plastidic origin for Fd-nir. In all the sampled taxa we found Fd-nir in genomic sequences corresponding to the nuclear genome. This indicates that Fd-nir would have been transferred to the nucleus before the divergence of all primary algal lineages, as indeed occurred with a substantial fraction of the plastid proteomes [10].

A cyanobacterial origin is unlikely for NAD(P)H-nir and nrt2 (Fig 3). In the NAD(P)H-NIR phylogeny (Fig 3 and S6 Fig), the sister-group position of Planctomycetes (light purple clade) to all eukaryotes suggest that this cytoplasmic nitrite reductase originated in eukaryotes through a HGT from this marine bacterial group. Finally, the phylogeny of NRT2 does not support any particular bacterial lineage as the donor of this nitrate transporter to eukaryotes (Fig 3 and S7 Fig).

EUKNR originated by gene fusion. In contrast to Fd-nir, NAD(P)H-nir and nrt2, euknr is restricted to eukaryotes. The specific arrangement of protein domains shown by this nitrate reductase (Fig 4B) suggests a chimeric origin involving the fusion of different proteins. Hence, we used a sequence similarity network-based approach [2] to investigate which ancestral protein families were involved in EUKNR origins. A first network between EUKNR and similar eukaryotic and prokaryotic sequences was constructed (Fig 4A; see Materials and methods section for details about the network construction process). The topology of the network shows five different clusters, each one representing a specific protein family, namely, bacterial sulfite oxidases (SUOX), eukaryotic SUOX with a Cytochrome b5 domain (Cyt-b5), eukaryotic SUOX without a Cyt-b5 domain, EUKNR, and NADH reductases (Fig 4A). The pattern connecting the EUKNR with the eukaryotic SUOX and NADH reductase clusters is characteristic of composite genes [26]; in which two unrelated gene families are connected in the network through an intermediate gene family. This suggests that EUKNR shares homology with both eukaryotic SUOX and NADH reductases [2]. Hence, a gene fusion between eukaryotic SUOX and NADH reductases would account for the origin of respectively the N-terminal and C-terminal EUKNR domains.

In the network shown in Fig 4A, only eukaryotic SUOX without a Cyt-b5 domain are connected to EUKNR. This suggests that EUKNR are more related to SUOX without a Cyt-b5, a result in agreement with standard phylogenetic methods (EUKNR sequences branched closer to SUOX without Cyt-b5; see S8 Fig). To determine the origin of the Cyt-b5 region, we used the Cyt-b5 domain of the two EUKNR reference sequences to construct a second network including those eukaryotic and prokaryotic proteins that aligned to this specific EUKNR
The two Cyt-b5 EUKNR regions connected with a lower E-value with Cyt-b5 monodomain proteins than with proteins whose architectures contain other domains in region (Fig 4C). The two Cyt-b5 EUKNR regions connected with a lower E-value with Cyt-b5 monodomain proteins than with proteins whose architectures contain other domains in
This strongly suggests that the Cyt-b5 region of EUKNR was not acquired from SUOX but rather originated from a Cyt-b5 monodomain protein. We thus propose that EUKNR has a chimeric origin, evolving from a fusion of genes belonging to three distinct families: eukaryotic SUOX (without Cyt-b5), Cyt-b5 monodomain proteins, and NADH reductases.

**Evaluating the impact of HGT in NAPs evolution**

Some of the topologies shown in the phylogenies of NAPs (Fig 5) strongly disagree with the eukaryotic species tree (Fig 2), and hence would require a large number of ancestral paralogues and differential parologue losses to be accounted by a strictly vertical inheritance scenario. In general (and with the exception of *nrt2*, see below), there is usually one copy of NAP genes per genome (see Table A in S1 Supporting information). Therefore, we did not find any *a priori* reasons to hypothesize that the number of copies could have been greater in the ancestral
To evaluate potential cases of HGT, we performed AU tests [27] (see all tested topologies and AU-test results in Table C in S1 Supporting information), as well as additional phylogenetic inferences excluding conflicting taxa and increasing the taxon sampling by incorporating orthologues from the taxon-rich Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) dataset [28] (MMETSP trees, see Materials and methods section).

**Fd-NIR.** This chloroplast nitrite reductase is restricted to photosynthetic groups, including all the primary algal groups (Glaucophyta, Rhodophyta, Chloroplastida), which belong to the Archaeplastida supergroup, as well as most of the sampled complex plastid algal groups, with the exception of the two photosynthetic myzozoans sampled (Alveolata, SAR). These complex plastid algal groups include Haptophyta, Ochrophyta (Stramenopiles, SAR), Bigellwieniella natans (Chlorarachniophyta, Rhizaria, SAR), and Guillardia theta (Cryptophyta, Archaeplastida) (Figs 2 and 5).
In the inferred phylogenetic tree, *Galdieria sulphuraria* (Rhodophyta) Fd-NIR is the earliest branch within the eukaryotic clade (Fig 5, S9 Fig), in disagreement with the accepted eukaryotic tree (Fig 2). However, the low nodal support and the fact that it branches with other Rhodophyta sequences in the MMETSP tree suggest that this position is artefactual (S10 Fig). Surprisingly, sequences from three Chloroplastida branch together with sequences from *Chondr us crispus* and *Pyropia yezoensis* (Rhodophyta), and are hence separated from the rest of Chloroplastida (we rejected the monophyly of Chlorophyta) (p-AU 0.0019). This unexpected topology is also observed and well supported in the MMETSP tree, suggesting that it is unlikely to represent a phylogenetic artefact. Because we showed that all eukaryotic Fd-nir descend from a unique transfer from Cyanobacteria (Fig 3), this conflicting topology could only be explained either by ancestral HGT within Archaeplastida or by ancestral duplication and differential paralogue loss.

All Fd-NIR sequences from Ochrophyta, Chlorarachniophyta and Haptophyta form a monophyletic clade that branch within Archaeplastida, with strong support in the MMETSP tree (99% UFBoot). We rejected a topology constraining the monophyly of all Archaeplastida sequences (p-AU 0.0009). These results, together with Fd-NIR being a plastidic enzyme, supports a common origin of both Fd-nir and plastids in these complex plastid algal groups. The phylogenetic position of Haptophyta sequences within Ochrophyta is in agreement with recent studies suggesting an Ochrophyta origin of the Haptophyta plastid [29,30] (we rejected the monophyly of Ochrophyta sequences) (p-AU 0.0029). The position of Chlorarachniophyta Fd-NIR within Ochrophyta, however, is more difficult to explain. One could argue that this is due to a low phylogenetic signal, given that Chlorarachniophyta (Rhizaria, SAR) is the closest group to Ochrophyta (Stramenopiles, SAR) among the taxa with Fd-NIR. In fact, we could not reject an alternative topology representing a vertical inheritance of Fd-nir in these two groups from a SAR common ancestor (*B. natans* as sister-group to the Ochrophyta + Haptophyta clade) (p-AU 0.2318). However, we consider an HGT from Ochrophyta to Chlorarachniophyta more parsimonious since the same topology was recovered in both the MMETSP tree and NRT2 tree (Fig 5, see below).

**NRT2.** This nitrate transporter is widely distributed among eukaryotes with nitrate and nitrite reductase genes (S2 Fig), with numerous species harboring lineage-specific duplications (see blue dots in S13 Fig). In particular, we found 66 species-specific duplications among the 56 species in which we found nrt2, with 36 duplication events occurring in Streptophyta (Chloroplastida). Again, to reconcile the recovered topology with the eukaryotic tree (Fig 2), a strictly vertical inheritance scenario would require a large number of ancestral duplications and differential paralogue losses. While obvious nrt2 paralogues are observed (S13 Fig), these correspond to recent duplications given that sequences from the same species branch close to each other. Therefore, as with other NAP genes, a strictly vertical inheritance scenario would be poorly supported given the absence of evident ancestral paralogues.

Sequences from Archaeplastida groups with primary plastids appear as the earliest clades of the tree (Fig 5), together with other eukaryotes (we rejected the monophyly of all Archaeplastida sequences) (p-AU 0.0014). The earliest-branching eukaryotic clade comprises only sequences from Glaucophyta (UFBoot 100%). The other eukaryotic NRT2 sequences branch in two clades that are strongly supported also in the taxon-rich MMETSP tree (S14 Fig). The first of these two clades includes all the Chloroplastida and Haptophyta sequences. It is unclear whether Haptophyta are more related to Chloroplastida [31] or to the SAR supergroup [32] at species level. If Haptophyta were more related to SAR, its position in this tree could be interpreted as a support for a horizontal origin of nrt2 from Chloroplastida. Indeed, a previous study suggested that Haptophyta could have received genes of non-plastidic function from the green-plastid lineage [33]. The second clade includes Rhodophyta and Cryptophyta sequences.
branching as sister-group to a SAR + Opisthokonta clade. Even though Cryptophyta presumably belongs to the Archaeplastida supergroup, its position as sister-group to Rhodophyta is unexpected [31,32] and could represent an ancestral Archaeplastida paragologue conserved in both groups. However, the plastid proteomes of Cryptophyta show clear signatures of a Rhodophyta contribution [20,34], and hence \textit{nrt2} could have been transferred to Cryptophyta from a red algal endosymbiont. Since a red algal signal has also been found in some SAR plastid proteomes [20,34], we also propose a second transfer from Rhodophyta to a SAR common ancestor; although we cannot discard an alternative scenario involving a first transfer from Rhodophyta to Cryptophyta and then from Cryptophyta to a SAR common ancestor (p-AU of 0.2851). As with Fd-NIR (Fig 5), sequences from \textit{B. natans} (Chlorarachniophyta, Rhizaria, SAR) branch within a clade including Ochrophyta sequences (Stramenopiles, SAR). Given that additional Chlorarachniophyta sequences robustly branch as sister-groups to and within Ochrophyta in the NRT2 and Fd-NIR MMETSP trees (S14 Fig and S10 Fig, respectively), we propose that these two NAP genes were co-transferred from Ochrophyta to Chlorarachniosphyta. Indeed, while all Chlorarachniophyta plastids presumably descend from a green algal endosymbiont [20], the chimeric signal of their plastid proteome suggests that other algal lineages could have contributed to the gene repertoire of this mixotrophic algal group [35].

\textbf{NRT2 and NAD(P)H-NIR.} The topology of NRT2 within the SAR + Opisthokonta clade resembles that of the NAD(P)H-NIR tree, except for Chlorarachniophyta, absent in NAD(P) H-NIR tree (Fig 5). Myzozoan sequences appear as the earliest-branching clade (Alveolata, SAR), with a clade including Ochrophyta sequences (Stramenopiles, SAR) branching as sister-group to a clade including the sequences from Oomycota and Labyrinthulea (Stramenopiles) and Teretosporea and Fungi (Opisthokonta). Despite the topologies within the Stramenopiles + Opisthokonta clade are not identical for NRT2 and NAD(P)H-NIR, the fact that clusters of both genes are found in many lineages of both taxonomic groups (Fig 2) strongly suggests that the whole pathway would have followed the same evolutionary path in Stramenopiles and Opisthokonta, with NAPs possibly transferred together as a cluster.

Our test of alternative topologies rejected the monophilies of Opisthokonta, Stramenopiles and Teretosporea for both NAD(P)H-NIR (p-AU of 0.0024, 0.0000 and 0.0051, respectively) and NRT2 phylogenies (p-AU 0.0245, 0.0024 and 0.0116, respectively). This strongly supports HGTs involving these groups. There are two reasons in favor of at least two HGTs between Stramenopiles and Opisthokonta. Firstly, as with NAD(P)H-NIR, the earliest branching positions of sequences from other SAR lineages to the Stramenopiles + Opisthokonta clade suggests at least one transfer from Stramenopiles to Opisthokonta. Indeed, we rejected a NRT2 topology compatible with a vertical inheritance scenario of this gene in Opisthokonta from a common ancestor of all eukaryotes (the Stramenopiles + Opisthokonta clade as sister-group to other eukaryotes, which would suggest an HGT origin of \textit{nrt2} in Labyrinthulea and Oomycota from Opisthokonta) (p-AU 0.0015). Secondly, an HGT specifically involving Oomycota and Ichthyosporea is strongly supported in both phylogenies (100% UFBoot). We next evaluated the following hypothetical HGT scenarios (represented in S15 Fig):

The first three scenarios consider that 2 HGT events occurred: (H1) proposes a late HGT from Ichthyosporea to Oomycota, and also that all opisthokont sequences descend from a single HGT from Stramenopiles. In such a case, we would propose an ancestral stramenopiles lineage leading to Labyrinthulea as the donor given the sister-group position of Labyrinthulea to Opisthokonta + Oomycota clade in NAD(P)H-NIR (Fig 5). We consider an ancestral lineage leading to Labyrinthulea (hereafter referred to Labyrinthulea) rather than a common ancestor of Stramenopiles because otherwise we would expect Labyrinthulea + Ochrophyta as sister-group to the Opisthokonta + Oomycota clade. (H2) In contrast to H1, H2 considers an HGT from Oomycota to Ichthyosporea, and hence ichthyosporean sequences would not...
descend from the ancestral labyrinthulean transfer to Opisthokonta but from a more recent HGT. (H3) Oomycota would have been the donor to a common ancestor of Opisthokonta instead of Labyrinthulae, and a transfer from *C. limacisporum* to Labyrinthulae would have also occurred. The following two scenarios consider one additional HGT: (H4) Ichthyosporea would be the donor to Oomycota. At least one labyrinthulean transfer would have occurred either (i) to Fungi or (ii) to Teretosporea. If (i), NAPs would have originated in Teretosporea from Labyrinthulae or Fungi. If (ii), NAPs would have originated in Fungi from Labyrinthulae or Teretosporea. (H5) In contrast to H4, Oomycota would have been the donor to Ichthyosporea. Hence, in this scenario, the lineage leading to *C. limacisporum* is the receptor/donor of the transfers involving Teretosporea in H4. Finally, these three additional scenarios (H6-H8) assume that the earliest transfer to Opisthokonta was from a common ancestor of Stramenopiles and not from a lineage leading to Labyrinthula: (H6) a first transfer from Stramenopiles to Fungi and a second transfer from Oomycota to Ichthyosporea. A third transfer to *C. limacisporum* from (i) Fungi or (ii) Labyrinthulae. (H7) In contrast to H6, the transfer from an ancestral stramenopile would have been to a lineage leading to *C. limacisporum* instead of to Fungi. The third transfer would have been to Fungi from (i) *C. limacisporum* or (ii) Labyrinthulae. (H8) A first transfer from Stramenopiles to a common ancestor of Opisthokonta, and a second transfer from Ichthyosporea to Oomycota.

We tested whether topologies representing these potential scenarios are statistically rejected by the phylogenetic signal of NRT2 and NAD(P)H-NIR (see Table C in S1 Supporting information for the results of the tests of topologies, as well as for the constrained topologies used). Whereas only H3 and H7(ii) were rejected in NRT2 (p-AU < 0.05), we rejected all scenarios except H1, H4 and H8 in NAD(P)H-NIR. The common feature of the scenarios rejected in NAD(P)H-NIR (H2, H3 and H5-H7) is that Oomycota would be the donor to Ichthyosporea, whereas H1, H4 and H8 hypothesise a transfer from Ichthyosporea to Oomycota.

If the transfer was from Ichthyosporea to Oomycota, in the absence of ichthyosporean sequences, Oomycota should be more related to *C. limacisporum* than to Ochrophyta (the closest lineages to Ichthyosporea and Oomycota, respectively, at the species level in both phylogenies); and the same for ichthyosporean sequences in the absence of Oomycota. In contrast, a closer relationship to Ochrophyta would be expected if the transfer was from Oomycota to Ichthyosporea. We thus evaluated if the exclusion of either Ichthyosporea or Oomycota could clarify the direction of the transfer by performing topological tests on two reduced datasets, each one excluding sequences from one of the two groups. In the dataset without Ichthyosporea, tests included the best topology and two alternative topologies constraining (i) the monophyly of Oomycota + Ochrophyta and (ii) the monophyly of Oomycota + *C. limacisporum*. For the second reduced dataset the same procedure was carried out but with Ichthyosporea instead of Oomycota.

For NAD(P)H-NIR, the Oomycota + Ochrophyta topology was rejected in the dataset excluding Ichthyosporea (p-AU 0.0163). However, the equivalent topology in the dataset without Oomycota (i.e. Ichthyosporea + Ochrophyta) was not rejected (p-AU 0.0721). In the case of the NRT2 reduced datasets, none of the topologies were rejected (see Table C in S1 Supporting information). In conclusion, the results, specifically from NAD(P)H-NIR, altogether provide more support for the scenarios proposing a transfer from Ichthyosporea to Oomycota (H1, H4 and H8). However, the fact that other scenarios were not rejected in NRT2 and the lack of concordance between the best topologies inferred for the reduced datasets (*S11* and *S12* Figs and *S16* and *S17* Figs) indicate limitations in the phylogenetic signal to provide conclusive support for any of the proposed scenarios.

**EUKNR.** The distribution of *euknr* was found to be very similar to that of *nrt2* (*S2* Fig). *Euknr* is present in most photosynthetic and non-photosynthetic organisms for which we...
inferred the capability to assimilate nitrate (Figs 2 and 5). Interestingly, we found euknr (but no other NAP genes) also in Chromosphaera perkinsii (Ichthyosporea, Opisthokonta). The presence of euknr in an additional ichthyosporean besides Creolimax fragrantissima and Sphaeroforma arctica may be an indicator that their NAP genes were vertically inherited from an opisthokont ancestor (see S15 Fig). Alternatively, if all ichthyosporean NAP genes descend from a transfer from Oomycota, the presence of euknr in C. perkinsii would indicate that the transfer was to an ancestral ichthyosporean. In both cases, the pathway would have been lost in the other ichthyosporeans, many of which have been described as parasitic species [36]. Unfortunately, the phylogenetic signal does not allow to confidently infer the evolutionary history of euknr, including the eukaryotic lineage in which this nitrate reductase would have had originated (Fig 5 and S18 Fig).

Notwithstanding the overall weak support of the phylogeny, we found three unexpected and well-supported relationships between distantly related taxa. Firstly, Oomycota sequences branch as the sister-group to the ichthyosporeans C. fragrantissima and S. arctica, as in the NRT2 and NAD(P)H-NIR trees (UFBoot 95%). This strongly indicates that a transfer of the whole pathway occurred between Oomycota and Ichthyosporea. Secondly, there is a clade that comprises several distantly related fungal sequences as well as a sequence from Acanthamoeba castellanii (Amoebozoa). However, sequences from these fungal taxa are also found in another clade that includes the A. nidulans sequence of bona fide nitrate reductase activity (named Anid_NaR in the euk_db dataset) [37][38]. Moreover, there is indirect experimental evidence suggesting that the A. nidulans euknr paralogue could not function as a nitrate reductase [39]. Thus, we propose that a fungal paralogue of euknr, of uncertain function, was transferred from an ancestral fungus to a lineage leading to A. castellanii. In fact, the finding of a gene transfer in A. castellanii is not surprising considering the extensive signatures of HGT found in this early amoebozoan lineage [40]. Thirdly, B. natans branches in-between Chlorophyta, in agreement with its plastid being originated from a green algal endosymbiont [20].

Origin and evolution of NAP clusters

We then inquired the importance of NAP clustering in shaping the evolution of this pathway. To this end, we analyzed the distribution of the clusters within the eukaryotic tree (Fig 2). We found clusters in >56% of the sampled eukaryotes with at least two NAP genes in the genome (Table A in S1 Supporting information). Clusters were patchily distributed in Fungi, Teretospora, Oomycota, Ochrophyta, Labyrinthulea, Myzozoa, Chlorarachniophyta, Chlorophyta, Rhodophyta and Haptophyta. The patchy distribution of NAP clusters within these groups suggests that many de-clustering events had occurred, assuming that de-clustering events are more parsimonious than de novo clustering events. NAP genes are always found unclustered in Cryptophyta, Glaucophyta and Streptophyta. While Cryptophyta and Glaucophyta are poorly represented in our dataset, the absence of clusters in Streptophyta (includes land plants) is remarkable since NAPs are found in the 9 sampled genomes of this group (Table A in S1 Supporting information).

We found that >64% of the detected clusters include the whole pathway, nrt2, euknr and either NAD(P)H-nir or Fd-nir. In Ochrophyta, the only eukaryotic group with taxa showing both nitrite reductases in the same genome, we found clusters comprising nrt2 or euknr and either Fd-nir or NAD(P)H-nir, but never both (Fig 2). In agreement with the gene distribution, clusters with Fd-nir are only found in autotrophs. While clusters with NAD(P)H-nir are also found in autotrophs, in particular in two Ochrophyta (Stramenopiles, SAR) and one Myzozoa (Alveolata, SAR) species; they are mostly distributed along osmotrophic taxa from Oomycota and Labyrinthulea (Stramenopiles, SAR) and from Fungi and Teretospora (Opisthokonta)
(NAP clusters with NAD(P)H-nir hereafter abbreviated as hNAPc, for "heterotrophic NAP clusters").

The presence in two of the three primary algal groups of NAP clusters including Fd-nir leads to two potential scenarios. First, we can envision a unique origin of the cluster in an archaeplastidan ancestor. If Glaucophyta, where the NAPs are not clustered (Fig 2), was an earlier lineage than Rhodophyta and Chloroplastida [24], cluster formation could have occurred in the last common ancestor of Rhodophyta and Chloroplastida. If so, at least two de-clusterization events would have occurred: one in the lineage leading to C. crispus and P. yeozenis (Rhodophyta) and the other in the lineage leading to Streptophyta (Chloroplastida) (Fig 2). A second scenario would imply at least two independent clustering events, in the lineages leading to Cyanidioschyzon merolae and G. sulphuraria (Rhodophyta) and to Chlorophyta (Chloroplastida).

The tendency of NAP genes to be clustered in green and red algae lineages may have facilitated the transfer of the entire pathway during the endosymbiotic events involving these algal groups [20]. However, the phylogenetic signal of NAPs suggests that not all the clusters in complex plastid algae would have been acquired from a single endosymbiont, with at least two independent clustering events occurring in the lineages leading to Chrysochromulina sp. (Haptophyta) and B. natans (Chlorarachniophyta). In Chrysochromulina sp., the cluster would have had originated after the acquisition of nrt2 and Fd-nir from Chloroplastida and Ochrophyta, respectively (Fig 5). In B. natans, the cluster would have had originated after the acquisition of euknr and Fd-nir from Chloroplastida and Ochrophyta, respectively.

For hNAPc, we propose that this cluster could have been transferred between heterotrophs given that sequences from taxa bearing the cluster (Fig 2) branch close to each other in the NAP trees (Fig 5). This would have allowed transfers of the entire metabolic pathway, which we consider more parsimonious than individual transfers of the genes followed by multiple clusterization events. Thus, in agreement with NRT2 and NAD(P)H-NIR phylogenies, we propose that hNAPc would have been originated in a common ancestor of Alveolata and Stramenopiles and later transferred between Stramenopiles and Opisthokonta. As shown in the previous section, the phylogenetic signal is not conclusive with the number and direction of hNAPc transfers that had occurred between Stramenopiles and Opisthokonta.

A tetrapyrrole methylase and the origin of NAPs in Opisthokonta

Given the uncertainty of the phylogenetic signal, we searched for additional features that could help clarify which of the proposed hypotheses for the origin of hNAPc in Opisthokonta is more parsimonious (S15 Fig). We checked intron positions, but we found them to be poorly conserved and not useful to clarify phylogenetic relationships. We found, however, in the genomes of C. limacisporum (Teretosporea, Opisthokonta), C. fragrantissima (Ichthyosporea, Teretosporea) and Aplanochytrium kerguelense (Labyrinthulea, Stramenopiles) an additional protein annotated with a TP_methylase Pfam domain clustering with NAP genes. The three proteins showed the highest similarity to Uroporphyrinogen-III C-methyltransferases, a class of tetrapyrrole methylases involved in the biosynthesis of siroheme (which works as a prosthetic group for many enzymes, NAD(P)H-NIR included) [41]. A phylogenetic tree of this protein family showed a clade that includes the three proteins clustered with NAP genes as well as other eukaryotic proteins branching within a bacterial clade (S19 Fig and S20 Fig). Therefore, we consider this group a subfamily of eukaryotic tetrapyrrole methylases, hereafter referred to as "TPmet".

TPmet is patchily distributed in eukaryotes (S21 Fig). As hNAPc, it is mostly restricted to Opisthokonta and Stramenopiles, which may well suggest that tpmet originated in one of the
two groups through a transfer from the other. In such a case, given that hNAPc most likely originated in Opisthokonta from Stramenopiles (see previous Results sections), and given the presence of tpmet+hNAPc in both groups; hypothesizing that hNAPc and tpmet co-originated in Opisthokonta through a tpmet+hNAPc transfer minimizes the number of HGTs and clustering events required to explain both tpmet and tpmet+hNAPc distributions. Moreover, given the presence of tpmet in many early opisthokont lineages (Choanoflagellatea, Ichthyosporea, Nuclearia, Fungi; see S15 and S21 Figs), the receptor of that hypothetical transfer would have been a common ancestor of Opisthokonta, as proposed by H1 and H8 scenarios. Indeed, both scenarios also propose that ichthyosporeans vertically inherited NAPs from Teretosporea rather than receiving them from Oomycota, which is supported by the presence of tpmet +hNAPc in C. fragrantissima and C. limacisporum but not in Oomycota. Under both scenarios, NAPs would have been secondarily lost multiple times in Opisthokonta, whereas tpmet would have been co-opted by other metabolic pathways (S15 Fig). Alternatively, H4, the other scenario that was not rejected by AU-tests, proposes independent hNAPc transfers to Fungi and Teretosporea (S15 Fig). This scenario is more parsimonious with NAP losses but requires an additional hNAPc transfer to Opisthokonta and possibly also a tpmet transfer. Unfortunately, the TPmet phylogeny has low UFBoot values (S20 Fig) and does not allow either to confirm or to reject any of the proposed scenarios (S15 Fig).

A novel chimeric nitrate reductase in ichthyosporean NAP clusters

In C. fragrantissima and S. arctica, rather than the canonical nitrate reductase, we identified a gene clustered with nrt2 and NAD(P)H-nir that has a chimeric domain architecture consisting of (i) the first three Pfam domains of the EUKNR in the N-terminal region; and (ii) the first two Pfam domains of the NAD(P)H-NIR in the C-terminal region (Fig 6A). A domain architecture analysis of proteins from euk_db and prok_db (see Materials and methods section) showed this unexpected domain architecture to be restricted to these two ichthyosporeans. Phylogenetic analyses showed that the region containing the Oxidoreductase molybdopterin binding, Mo-co oxidoreductase dimerisation and Cytochrome b5-like Heme/Steroid binding Pfam domains corresponds to the EUKNR family (Fig 5 and S18 Fig), which includes the nitrate reducing module characteristic of this nitrate reductase [42]. In contrast, the C-terminal region, corresponding to the Pfam domains Pyridine nucleotide-disulphide oxidoreductase and BFD-like [2Fe-2S] binding domain, branched within the NAD(P)H-NIR clade in a tree including all the eukaryotic and prokaryotic proteins containing this pair of domains (Fig 6B and S22 Fig). In the latter tree, the two sequences branched as the sister-group to C. fragrantissima and S. arctica NAD(P)H-NIR proteins. Therefore, we propose that this chimeric gene originated by the replacement of the canonical C-terminal EUKNR region with the N-terminal region of the NAD(P)H-NIR in a common ancestor of these two ichthyosporeans (hereafter we refer to this gene as C. fragrantissima and S. arctica putative nitrate reductase, abbreviated as CS-pNR). This event must have occurred after the HGT event involving Ichthyosporea and Oomycota, since the nitrate reductase of Oomycota comprises the canonical domain architecture of EUKNR (Fig 6A).

S. arctica has a NAP cluster functional for nitrate assimilation

We sought experimental evidence of nitrate assimilation in S. arctica, as a representative of an ichthyosporean NAP cluster including the putative uncharacterized nitrate reductase (CS-pNR). We developed a minimal growth medium in which the nitrogen source (N source) can be controlled (‘modified L1 medium–mL1’, see Materials and methods). We then tested the growth of S. arctica in mL1 minimal medium with different N sources (Fig 7A).
minimal medium conditions (mL1 + different N sources), cells were smaller than in Marine Broth, used as the positive control. We observed a slight growth in the negative control (‘mL1’) after 168 hours, which we hypothesize may be due to the use of cell reserves, or the utilization of vitamins from the medium as N source. We detected a clearly stronger growth in mL1 supplemented with either NaNO$_3$, (NH$_4$)$_2$SO$_4$ or urea, compared to mL1 without any N source.

The growth observed in mL1 + NaNO$_3$ shows that *S. arctica* is able to assimilate nitrate.

The finding that *S. arctica* can grow using nitrate as the sole N source implies that this organism must have a nitrate reductase activity, and the CS-pNR is indeed a strong candidate

---

**Fig 6. NAP clusters in Ichthyosporeae and the origins of a putative novel nitrate reductase.** (A) Cluster organization and protein domain architecture of NAP clusters from the ichthyosporeans *Creolimax fragrantissima*, *Sphaeroforma arctica* and *Phytophthora infestans* (Oomycota). Within each cluster each box represents a gene, with the arrowhead indicating its orientation. The Pfam domains predicted for the corresponding protein sequences are represented inside each box (see panel). (B) Schemes showing a simplified representation of the maximum likelihood phylogenetic trees inferred for the N-terminal and C-terminal regions of the putative nitrate reductase identified in *C. fragrantissima* and *S. arctica* (CS-pNR, see Results section). For an entire representation of the phylogenetic trees, see S18 Fig and S22 Fig for N-terminal and C-terminal regions, respectively. (C) Schematic representation of the evolutionary origin of CS-pnr, inferred from the phylogenies shown in (B).

![Diagram of Pfam protein domains](https://doi.org/10.1371/journal.pgen.1007986.g006)
to carry out this enzymatic activity, in line with the bioinformatics evidence (Fig 6). In general, NAP genes from different eukaryotic species had been shown to be co-regulated in response to environmental N sources [19,43–46]. Hence, a co-regulated expression of CS-pnr with nrt2 and NAD(P)H-nir would be consistent with their proposed role in nitrate assimilation. We
thus measured the levels of expression of the three genes in *S. arctica*, in the presence of different N sources (Fig 7B). The three *S. arctica* genes were up-regulated either in mL1 without any N source as well as in mL1 + NaNO$_3$. In contrast, we observed that the three genes were poorly expressed in mL1 + (NH$_4$)$_2$SO$_4$ and in mL1 + urea. These results show that the cluster is functional in *S. arctica* and also that its expression is regulated in response to different N sources.

**Discussion**

**Nitrate assimilation is restricted to autotrophs and fungal-like osmotrophs**

Our screening of NAP genes provides an updated and comprehensive picture of the distribution of the nitrate assimilation pathway in eukaryotes (Fig 2). Besides the taxa included in previous studies [14,47], we describe the presence of the complete pathway in Haptophyta, Cryptophyta, Chlorarachniophyta, Myzozoa, Labyrinthulea and Teretosporea (S1 Fig). While all autotrophs analyzed have NAP genes, this is not the case for heterotrophs, where we only found NAP genes in taxa from those groups that have convergently evolved to a fungus-like osmotrophic lifestyle [21], that is Fungi, Ichthyosporea, Oomycota and Labyrinthulea (Fig 2 and S2 Fig). The absence of this pathway in phagotrophs is probably due to the fact that this nutrient acquisition strategy provides access to organic nitrogen sources, whose incorporation is energetically less demanding than nitrate. Thus, NAP genes would be less likely to be acquired by phagotrophic lineages in the absence of positive selection favoring it. At the same time, a lineage with a phagotrophic lifestyle that may have vertically inherited the pathway from less specialized phagotrophic ancestors will be more prone to lose it, as presumably occurred with genes involved in the synthesis of certain amino acids in Metazoa [48].

**HGT and the evolutionary history of NAPs in eukaryotes**

The patchy distribution of this metabolic pathway (Fig 2) and the large number of non-vertical relations observed in our phylogenies (Fig 5) are not consistent with a scenario considering only vertical transmission and gene loss. We consider that some of the unexpected topologies found represent indeed *bona fide* gene transfers, because we consistently recovered them in more than one NAP phylogeny and/or because they fit with endosymbiotic events proposed for the acquisition of complex plastids [10,34]. Here we detail our proposed evolutionary scenario to account for the distribution and the phylogenetic signal of NAPs (Fig 8):

**Origin of the pathway.** Three of the four NAP genes originated in eukaryotes through independent transfers from Bacteria. The nitrite reductases NAD(P)H-nir and Fd-nir were most likely transferred from Planctomycetes and Cyanobacteria (Fig 3), respectively; with Fd-nir showing signatures of a plastidic origin specially if, as proposed by some authors, this organelle originated from an early-cyanobacterial lineage [25,49]. The particular bacterial donor of the nitrate transporter nrt2 remains unclear (Fig 3). The nitrate reductase euknr originated through the fusion of three eukaryotic genes: a sulfite oxidase, a Cyt-b5 monodomain and a FAD/NAD reductase (Fig 4). Furthermore, we hypothesize that the pathway, including nrt2, Fd-nir and euknr, was established in an early-Archaeplastida ancestor, as discussed below. First, the three pathway activities most likely originated in the same eukaryotic ancestor. If this is the case, and the plastidic nitrite reductase Fd-NIR, and not NAD(P)H-NIR, was present in the original nitrate assimilation toolkit, this would imply an Archaeplastida origin of the pathway, given that it is well established that plastids originated in this group. The phylogenies of Fd-NIR and NRT2 are consistent with this hypothesis, as they show Archaeplastida sequences in the earliest branches within eukaryotes (Fig 5). The fact that the same topology is not observed in the EUKNR tree does not contradict our argument, since the EUKNR tree showed low statistical nodal support (S18 Fig). The alternative NAD(P)H-nir-early scenario,
Fig 8. A hypothetical scenario for the evolution of the nitrate assimilation pathway in eukaryotes, based on the current phylogenetic data and in the taxonomic distribution of NAPs. For each transfer proposed, donor and receptor lineages are indicated; as well as if the transfer is related to the origin of the pathway (1), or if autotrophic (2) or osmotrophic (3) lineages were involved (see the Discussion section ‘HGT and the evolutionary history of NAPs in eukaryotes’). Transfers of NAP genes in clusters are represented with the corresponding NAP symbols surrounded by a square. Branches in red are those where loss of the entire pathway would have occurred, which were parsimoniously inferred from the reconstructed evolutionary history. For the sake of simplicity, some species were collapsed into clades representing higher taxonomic categories. For each species/clade, NAP gene presence/absence (NAP symbols) and their cluster status (symbols colored in black for those NAPs found in the same gene cluster) are indicated. For those clades in which not all the represented species have the same NAP content and cluster status (labeled with ‘•’), the most prevalent ones are shown (see Table A in S1 Supporting information for a complete representation of the NAP content and cluster status). Distinct scenarios considering transfers between Stramenopiles and Opisthokonta were evaluated (see S15 Fig and ‘NRT2 and NAD(P)H-NIR’ Results subsection). H1 and H8 scenarios are both the most parsimonious with the number of HGT events among those not rejected with the AU-test. Both scenarios propose a transfer from an ancestral Stramenopile (either from the lineage leading to Labyrinthulea or from a common ancestor of Stramenopiles) to a common ancestor of Opisthokonta.

https://doi.org/10.1371/journal.pgen.1007986.g008
while still possible, is less parsimonious because it disagrees with the NRT2 phylogeny and requires additional secondary losses of this gene.

**Transfers between autotrophs.** We propose that NAP genes were transferred from Archaeplastida to other eukaryotic groups during the endosymbiotic events that led to the origin of complex plastids, as it has been shown for numerous genes not necessarily related to plastidic functions [10]. Consistent with this, our phylogenies suggest multiple NAP transfers between algal lineages (Fig 8), with sequences from the complex plastid algal groups branching as sister-groups to the early-branching Archaeplastida sequences in the Fd-NIR and NRT2 trees (Fig 5). For some transfers, the donor and the receptor lineages coincide with proposed endosymbiotic events. This is the case of *euknr* from Chlorophyta to Chlorarachniophyta [20], *Fd-nir* from Ochrophyta to Haptophyta [30] and *nrt2* from Rhodophyta to Cryptophyta and to SAR [20]. Even though we also found some unexpected transfers between algal lineages, we also consider them as potential endosymbiotic gene transfers. The reason is that the origin of complex plastids is not clearly elucidated, partly due to the heterogeneous phylogenetic signal shown by the plastid proteomes [34]. Based on this heterogeneity, the target-ratchet model proposes that complex plastids resulted from a long-term serial association with different transient endosymbionts, all of which could have contributed in shaping the proteome of the host lineage [20]. Consistent with this model, we found NAP genes in Haptophyta and Chlorarachniophyta that would have been transferred from different potential algal endosymbionts (Fig 8).

**Transfers between osmotrophs.** From the gene distribution and the phylogenies (Fig 5), we parsimoniously propose that NAD(P)H-nir was transferred from Planctomycetes to a common ancestor of Alveolata and Stramenopiles. The advent of this cytoplasmic nitrite reductase would have resulted in a eukaryotic nitrate assimilation pathway independent from Fd-NIR, and hence independent from the chloroplast. We found NAP sequences from distinct osmotrophic lineages from Stramenopiles and Opisthokonta branching together in the trees (Fig 5), strongly suggesting HGTs involving these groups. Based in these phylogenies (Fig 5) but also in an analysis of the distribution and the gene composition of the clusters (Figs 2 and 6), at least two transfers would have occurred: at least one transfer of a NAP cluster from an ancestral stramenopiles (possibly from a lineage leading to Labyrinthula) to Opisthokonta; and a more recent HGT involving Ichthyosporea (Teretosporea, Opisthokonta) and Oomycota (Stramenopiles) (see all the scenarios evaluated in S15 Fig).

Among all the scenarios evaluated (S15 Fig) that were not rejected by AU-tests (H1, H4 and H8; see Table C in S1 Supporting information), both H1 and H8 minimize the number of HGTs required to explain the distributions of NAPs and NAP clusters (hNAPc), *tpmet* and NAP clusters with *tpmet* (*tpmet*-hNAPc) (see the Results section “A tetrapyrrole methylase and the origin of NAPs in Opisthokonta”). Both scenarios assume a *tpmet*-hNAPc transfer from an ancestral stramenopiles to a common ancestor of Opisthokonta, followed by multiple secondarily losses of NAPs and a more recent hNAPc transfer from Ichthyosporea to Oomycota. Under these scenarios, the ancestor of opisthokonts, which has been suggested to be a phagotroph [50], would have had the faculty to assimilate nitrate. This may be seen problematic given the strong anticorrelation shown between phagotrophy and nitrate assimilation (Fig 2, S2 Fig). Thus, the mentioned anticorrelation may be considered as an argument in favor of H4 (S15 Fig), which was also not rejected by AU-tests. Compared to H1 and H8, H4 considers that Teretosporea and Fungi would not have vertically inherited NAPs from a common ancestor, and hence H4 is less parsimonious considering the number of proposed NAP HGTs (3 instead of 2) but requires less secondary NAP losses. H4 also uncouples the origin of *tpmet* in Opisthokonta from the origin of NAPs, thus possibly implying an additional HGT of *tpmet* to a common ancestor of Opisthokonta (S15 Fig). If, notwithstanding, H1 or H8 were correct,
this may suggest that the common ancestor of Opisthokonta already had some pathways valuable for an osmotrophic lifestyle. Interestingly, this hypothetical potential for osmotrophy may have favored the transitions towards this lifestyle that occurred in Fungi and Teretosporea [50]. Indeed, the role of HGT in shaping the gene toolkits for osmotrophic functions is well documented in Oomycota and Fungi [12,51]. Our finding of HGT events involving taxa from these two groups but also from Teretosporea and Labyrinthula extends the potential scope and importance of this mechanism in the acquisition of metabolic features associated to an osmotrophic lifestyle [21].

These hypothetical scenarios proposed for the origin of nitrate assimilation in Opisthokonta disagrees with the evolutionary scenario hypothesized from previous results [14]. In particular, it was proposed that a stramenopiles lineage leading to Oomycota could have transferred a NAP cluster to Dikarya (Fungi), while in our trees with an updated taxon-richer dataset, Oomycota branches as sister-group to Ichthyosporea within the Opisthokonta + Stramenopiles clade (Fig 5). To evaluate whether the discrepancies with previous studies are due to differences in the taxon sampling, we constructed trees excluding all sequences from Teretosporea and Labyrinthula, which were not available in previous analyses. Interestingly, in the absence of these two groups, we recovered the Oomycota (Stramenopiles) sequences branching as sister-group to Fungi for both NRT2 and NAD(P)H-NIR phylogenies with a reduced dataset (S23 Fig and S24 Fig, respectively). In agreement with the H1 and H8 or H4 scenarios, we propose that in the absence of Labyrinthula and Teretosporea, Oomycota (Stramenopiles) branched as more related to Fungi (Opisthokonta) than to Ochrophyta (Stramenopiles) because Oomycota would have received the NAP genes from Opisthokonta, in particular from Ichthyosporea. Notwithstanding, the support for the proposed scenarios is susceptible to change with the addition of further taxa, given the dependence of HGT inference to the taxon sampling used [7].

HGT of NAPs could be favored by the metabolic, genomic and ecological landscapes

While HGT in eukaryotes has been the subject of controversy, there is an increasing number of gene families where HGT has been shown to play a role [11,52,53]. The results presented here show the evolutionary history of the nitrate assimilation pathway as a striking example of the importance that gene transfer between eukaryotes may have in the evolution of a certain metabolic pathway. Among the transfers proposed by the most parsimonious scenario (Fig 8), we consider at least the following ones as bona fide transfers because of being well supported by the data (see Results section): 1) At least one transfer of a NAP cluster from an ancestral stramenopiles to Opisthokonta; 2) a NAP cluster transfer between Ichthyosporea and Oomycota; 3) a nrt2 transfer between Haptophyta and Chlorophyta; 4 and 5) a Fd-nir transfer from primary algae to SAR, and from Ochrophyta to Haptophyta; 6) a euknr transfer from Chlorophyta to Chlorarachniophyta; and 7) a transfer of a euknr paralogue of unknown function from Fungi to a lineage leading to A. castellanii.

We argue that NAP genes may be particularly prone to be successfully transferred. At a metabolic level, pathways downstream to nitrate assimilation and the enzymes involved in the synthesis of the molybdenum cofactor (required for the activity of a number of enzymes, EUKNR included) are widespread in eukaryotes [42,48,54]. This would facilitate the functional coupling of the newly transferred pathway to the metabolic network. At a genomic level, NAP genes are frequently organized in gene clusters in eukaryotic genomes (Fig 2). This would allow the acquisition of the whole pathway in a single HGT event [55], which is also more likely to be positively selected than separate transfers of individual components of the pathway.
Moreover, the presence of the whole pathway in the same genomic region could also favor the evolution of a co-regulated transcriptional control after the HGT acquisition [56]. There are various reported examples of other metabolic gene clusters transferred between eukaryotes [57]. At an ecological level, nitrate concentrations have been fluctuating in the course of evolutionary history [58], and are still highly dependent on regional and seasonal changes [16]. Thus, in some circumstances NAP genes could be dispensable while in other circumstances their acquisition through HGT would be favored. This dynamic evolutionary fitness could imply that even one given eukaryotic lineage could have acquired and lost the faculty of nitrate assimilation more than once in the course of its evolution.

Nitrate assimilation in Ichthyosporea: A putative novel nitrate reductase

The presence of NAP genes in Ichthyosporea, described as animal symbionts or parasites [36] and phylogenetically related to Metazoa [59], was not previously reported. In the NAP clusters of *C. fragrantiissima* and *S. arctica*, we found a putative nitrate reductase gene that originated in a common ancestor of these two ichthyosporeans (*CS-pnr*) from the fusion of the N-terminal region of the EUKNR with the C-terminal region of the NAD(P)H-NIR (Fig 6). The presence of the nitrate reducing module characteristic of EUKNR [42] strongly suggests that the clustered *CS-pNR* is a functional nitrate reductase. The growth on nitrate as sole nitrogen source of *S. arctica* (mL1 + NaNO₃, Fig 7A), in the absence of any other candidate enzyme in the genome, constitutes almost a definitive proof for this function. This is further supported by the strong transcriptional co-regulation of *CS-pnr* with *nrt2* and NAD(P)H-nir in response to the availability of different nitrogen sources. In particular, these genes are poorly expressed on easily assimilable nitrogen sources (urea and ammonium) and highly expressed in a nitrogen-free medium as well as in the presence of nitrate (Fig 7B).

The results from the RT-qPCR experiments can be most easily rationalized by a straightforward repression process. However, specific induction by nitrate cannot be excluded. In the nicotinate assimilation pathway of *A. nidulans*, we see both specific induction and high expression under nitrogen starvation conditions, mediated by the same transcription factor [60]. It is possible that in this latter instance the intracellular inducer is generated by degradation of intracellular metabolites. Similarly, in the absence of any added nitrogen source, a high-affinity nitrate transporter may scavenge residual nitrate present in the nitrate-free culture medium, as it has been specifically shown for *A. nidulans* [61], *Hansenula polymorpha* [62] and *C. reinhardii* [63]. In agreement with this, RNAseq data show that in *A. nidulans*, an organism where the nitrate-responsive transcription factor has been thoroughly studied [61], nitrate starvation results in high expression of the three genes in the NAP cluster [64].

The transcriptional regulation of NAPs has been characterized in land plants [43], Chlorophyta [19], Rhodophyta [44], Fungi [45,46]; and now also in the ichthyosporean *S. arctica* (Fig 7). The independent origins of NAP genes in some of these groups (Fig 8), together with the shown lineage-specific differences at the regulatory elements [19,65–67] suggests that natural selection promoted the evolution of analogous regulatory responses, favoring the integration of this pathway into the metabolic landscape after its acquisition through HGT.

Materials and methods

Phylogenetic screening of NAPs

An updated database of 174 eukaryotic proteomes (euk_db) was constructed (January 2017), using predicted protein sequences from publicly available genomic or transcriptomic projects [68–70]. The complete list of species, with the corresponding abbreviations, is available in Table A in S1 Supporting information. The phylogenetic relationships between all the sampled
Eukaryotes were constructed from recent bibliographical references [71–82]. Protein domain architectures from all euk_db sequences were obtained with PfamScan (a Hidden Markov Model [HMM] search-based tool) using Pfam A version 29 [83]. A database of prokaryotic protein sequences (prok_db) was constructed from Uniprot bacterial and archaeal reference proteomes (Release 2016_02) [70] with the aim of detecting potential prokaryotic contamination in euk_db as well as to investigate the prokaryotic origins of eukaryotic NAP genes.

For each NAP, we followed a multi-step procedure in order to maximize both sensitivity and specificity in the orthology assignation process (see S2 Supporting information for detailed information about the particular strategy followed for each NAP; available in [doi.org/10.6084/m9.figshare.6462311.v1]). The overall strategy consisted first in identifying potential NAP family members in euk_db with BLAST (version 2.3.0+) [84] and HMMER (version 3.1b1) [85]. For BLASTP searches, we queried the databases using the NAP sequences from Chlamydomonas reinhardtii and A. nidulans [18], downloaded from Phytozome 11 [86] and NCBI protein databases [68], respectively [BLASTP: -evalue 1e-5, only non-default software parameters specified]. For HMMER searches [hmmsearch], we used the HMM Pfam domains MFS_1 for NRT2, Oxidored_molyb and Mo-co_dimer for EUKNR, and NIR_SIR for NAD(P)H-NIR and Fd-NIR. The candidate sequences retrieved from the BLASTP and hmmsearch analyses were submitted to cdhit (version 4.6) [87] [-c 0.99] to remove repeated/very recent paralogues (i.e. redundant sequences). We used the non-redundant candidate sequences to detect potential prokaryotic homologues in prok_db [-evalue 1e-5], to use them as outgroups to eukaryotic sequences and/or to detect potential euk_db contaminant sequences during the phylogenetic analyses. The non-redundant candidate sequences and the captured prokaryotic homologs were submitted to an iterative process in which we recursively performed phylogenetic inferences with the sequences non-discarded in the previous steps until we reached a set of bona fide NAP family members. The criteria to discard sequences in each step was mainly phylogenetic, but also assisted with functional information of each candidate sequence, predicted from their Pfam domain architecture and from their best-scoring BLASTP hit [-evalue 1e-3] against the SwissProt database [70] (downloaded on July 2016). Those potential eukaryotic NAPs that branched separately from other eukaryotic sequences within a prokaryotic clade in the phylogenies were considered as contaminants if they correspond to a euk_db proteome generated from transcriptomic data obtained from cultures with bacterial contamination, or if they are encoded in potentially contaminant genomic scaffolds. Previous to all phylogenetic inferences, sequences were aligned with MAFFT (version v7.123b) [88] [mafft-einsi] and alignments were trimmed with trimAl (version v1.4.rev15) [89] using the -gappyout option. Maximum likelihood phylogenetic inference was done using RAxML (version 8.2.4) [90] with rapid bootstrap analysis (100 replicates) and using the best model according to BIC criteria in ProtTest analyses (version 3.2) [91].

In Table A in S1 Supporting information, for each species, the columns corresponding to NAPs are colored in blue when at least 1 bona fide member has been identified. They are colored in light brown when all members identified are likely to correspond to bacterial contamination, and in red when no NAPs were identified. The sequence names of all the bona fide and contaminant NAPs are also indicated. All the NAP sequences used in the phylogenetic inferences carried out in this study, as well as alignments of the NAP phylogenies in eukaryotes (Fig 5), are available in S3 Supporting information (doi.org/10.6084/m9.figshare.6462311.v1).

Re-annotation of NAPs using TBLASTN

For those eukaryotes in which we detected an incomplete presence of the pathway (i.e. having only genes coding for some but not the three required steps, see Fig 1), we performed
additional searches in the genomic sequences of the corresponding organism using the reference NAP protein sequences [TBLASTN: -evalue 1e-5]. This additional search allowed us to re-annotate two putative NAPs absent from euk_db (Fd-NIR in Aureococcus anophagefferens and NRT2 in Ostreococcus tauri) that were later incorporated in the phylogenies.

We also searched for potentially transferred prokaryotic nitrate and nitrite reductases, whose presence would suggest a replacement of their eukaryotic counterparts. While a putative ‘Copper containing nitrite reductase’ was found in the amoebozoan Acanthamoeba castellanii, we considered this sequence as an uncharacterized copper oxidase not necessarily involved in nitrite reduction. We were based in the fact that (1) the most similar sequences in euk_db and prok_db correspond to few distantly related eukaryotes without any NAPs or with already the complete eukaryotic pathway predicted such as C. reinhardtii and (2) the absence of the characteristic InterPro Nitrite reductase, copper-type signature.

Correlation between NAPs distribution and feeding strategies

We constructed phylogenetic profiles for each NAP gene family: vectors with presence/absence information (coded in “1” or “0”, respectively), with every position of the vectors corresponding to a certain species sampled in our euk_db dataset. These vectors were then used to quantify the correlation between the distributions of the different NAPs by computing the inverse of the Hamming distance between each pair of phylogenetic profiles. We also quantified the correlation between the distribution of the different NAPs and the distribution of the different nutrient acquisition strategies in eukaryotes. For that, we classified eukaryotes into ‘Autotrophs/Mixotrophs’ (i.e. strictly and facultative autotrophs) and ‘Non-autotrophs’. ‘Non-autotrophs’ (i.e. strictly heterotrophs) were further subclassified into ‘Phagotrophs’, ‘Fungal-like osmotrophs’ and ‘Others’ (Table A in S1 Supporting information). The category ‘Phagotrophs’ include all heterotrophs that feed by phagotrophy. The category ‘Fungal-like osmotrophs’ include all heterotrophs that belong to eukaryotic groups with cellular and physiological features characteristic of a fungal-like osmotrophic lifestyle [21]. These include Fungi, Teretosporea, Oomycota and Labyrinthulea. The category ‘Others’ include all the heterotrophs not classified in any of these categories, all of them belonging to eukaryotic groups with a parasitic lifestyle.

Evolution of NAP genes

We used the bona fide eukaryotic NAP sequences identified to reconstruct the evolutionary history of the NAP gene families in eukaryotes. We excluded all the sequences with less than half of the median length of the corresponding NAP family in order to remove fragmented sequences that could mislead the alignment and the phylogenetic inference processes. Sequences were aligned and trimmed with MAFFT [mafft-einsi] and trimAl [-gappyout]. For the phylogenies, we used IQ-TREE (version 1.5.3) [92] instead of RAxML given that an approximately unbiased (AU) test can be performed in IQ-TREE [27]. AU test was used to evaluate whether the robustness of those branches indicating potential gene transfer events are significantly higher than other alternative topologies (10000 replicates; see all the alternative topologies tested and AU test results in Table C in S1 Supporting information). Trees representing the alternatives topologies were constructed also with IQ-TREE, using the same alignments and evolutionary models and constraining the topologies with Newick guide tree files [-g option]. For bootstrap support assessment, we used the ultrafast bootstrap option (1000 replicates) because it was shown to be faster and less biased than standard methods [93,94]. For model selection, we used ModelFinder, already implemented in IQ-TREE [95].
Moreover, and given that some eukaryotic groups were poorly represented due to the lack of genomic data, we constructed additional NAP trees incorporating orthologues from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) dataset [28]. We queried the reference NAP protein sequences against all the MMETSP transcripts [BLASTP: -evalue 1e-3]. MMETSP NAP orthologues were identified from the aligned sequences by means of Reciprocal Best Hits (RBH) [96] and best-scoring BLASTP hit against SwissProt database [-evalue 1e-3]. Whereas some contamination may be expected from the MMETSP dataset, this does not influence the deduced presence of NAP genes in the eukaryotic groups (Fig 1), as to establish the latter we only employed the curated and comprehensive euk_db dataset.

**Comprehensive screening of NAPs in prokaryotes**

We used the *bona fide* eukaryotic NAP sequences to capture potential prokaryotic orthologues of *nrt2*, *NAD(P)H-nir* and *Fd-nir*. First, we queried those sequences against prok_db with BLASTP [-max_target_seqs 100, -evalue 1e-5]. Protein domain architectures were annotated with PfamScan, and those with clearly divergent architectures were discarded. The remaining prokaryotic sequences were aligned with eukaryotic NAPs using MAFFT [mafft-einsi]. The alignments were trimmed with trimAl [-gappyout] and the phylogenetic inferences were done with IQ-TREE [ultrafast bootstrap 1000 replicates, best model selected with ModelFinder]. Prokaryotic sequences were taxonomically characterized by aligning them against a local NCBI nr protein database (downloaded on November 2016), and only hits with more than 99% of identity and query coverage were considered [BLASTP: -task blastp-fast].

To ensure that the taxonomic representation of prok_db allow to detect signatures of genes likely to have been transferred from Alphaproteobacteria and Cyanobacteria (the putative donors of the mitochondria and the plastid, respectively [97]), we constructed control phylogenies using in each case two genes with a known plastidic ('Photosystem II subunit III' and 'ribosomal protein L1' [24]) (S4 Fig and S5 Fig, respectively) and mitochondrial origin ('Cytochrome c oxidase subunit III' and 'Cytochrome b' [98]) S25 Fig and S26 Fig, respectively). For the mitochondrial and plastid control genes, the eukaryotic sequences used to query prok_db were retrieved from a subset of proteomes from plastid-bearing eukaryotes. For the detection of potential orthologues in prok_db, alignment and phylogenetic inference; we used the same procedure, software and parameters as with the NAP trees (see above).

**Construction of sequence similarity networks**

**Sequence similarity network of full length EUKNR.** The EUKNR protein sequences were aligned against a database including euk_db and prok_db [BLASTP: -max_target_seqs 10000, -evalue 1e-3]. Aligned sequences were concatenated with the EUKNR and redundant sequences were removed with cdhit before being aligned all-against-all with BLASTP [-max_target_seqs 10000, -evalue 1e-3]. We used Cytoscape (version v.321) [87] to construct a sequence similarity network from BLAST results, represented using the organic layout option. In the network, each aligned protein correspond to a node. Nodes are connected through edges if the corresponding sequences aligned with a lower E-value than the threshold value. A relaxed E-value threshold would lead to an over-connected network, with edges connecting very divergent proteins. On the other hand, a strong threshold would lead to an under-connected network, having only connections between strongly similar proteins. After exploring different thresholds, we chose an E-value of 1e-82 because it allows to represent only the most similar protein families to the N-terminal and C-terminal regions of EUKNR. We performed as well the following modifications in order to remove redundant and non-informative
connections and to facilitate the analysis and interpretation of the network: (i) we removed self-loops and duplicate edges; (ii) we removed those nodes that were not connected to the EUKNR cluster or that were connected with a distance of more than two nodes; (iii) non-EUKNR sequence names were modified to include information of their protein domain architectures [PfamScan]; (iv) we removed nodes and edges corresponding to proteins with strong evidence of corresponding to miss-predicted proteins (e.g. spurious domain architectures). Nodes representing proteins that only connected with mis-predicted sequences were also removed (information about the list of proteins, their domain architecture and the particular reasons for their exclusion is available in Table B in S1 Supporting information). The BLAST output file used to construct the network as well as the Cytoscape file corresponding to the final network are available in S4 Supporting information (doi.org/10.6084/m9.figshare.6462311.v1).

To validate whether EUKNR are more phylogenetically related to non-Cyt-b5 sulfite oxidases than to Cyt-b5 sulfite oxidases (see the corresponding Results section), we constructed a phylogenetic tree with the identified EUKNR sequences and the sulfite oxidases detected during the network construction process. MAFFT [mafft-einsi], trimAl [-gappyout] and IQ-TREE [ultrafast bootstrap 1000 replicates, best model selected with ModelFinder] were used for phylogenetic inference.

**Sequence similarity network of EUKNR Cyt-b5 domain.** The regions of the *A. nidulans* and *C. reinhardtii* EUKNR corresponding to the Cyt-b5 Pfam domain were aligned against a database including euk_db and prok_db [BLASTP: -max_target_seqs 10000, -evalue 1e-3]. Non-redundant and non-EUKNR sequences were concatenated with the two EUKNR Cyt-b5 sequences and an all-against-all alignment was performed [BLASTP: -max_target_seqs 10000, -evalue 1e-3]. Sequence names were modified to include information of their protein Pfam domain architectures [PfamScan]. A sequence similarity network was constructed with Cytoscape and represented with the organic layout option (as with full length EUKNR), removing self-loops and duplicate edges and using an E-value threshold of 1e-17. We also removed those nodes that were not connected to *A. nidulans* or *C. reinhardtii* EUKNR Cyt-b5 regions or that were connected with a distance of more than two nodes. The BLAST output file used to construct the network as well as the Cytoscape file corresponding to the final network are available in S4 Supporting information (doi.org/10.6084/m9.figshare.6462311.v1).

**Detection of NAP clusters**

For the detection of clusters of NAP genes, we scanned the genomes of those sampled eukaryotes with more than 1 NAP gene identified. We aligned the NAPs of each organism against the corresponding genomes using TBLASTN [-evalue 1e-3]. The genomic location of each NAP was annotated based on the TBLASTN hit with the highest score. Then, we looked for genomic fragments with more than 1 NAP genes annotated, and the genes were considered to be in a cluster when they were proximally located in that fragment. In the case of *Corallochytrium limacisporum*, the two NAP genes detected were found in terminal positions of two separate fragments of the genome assembly (nrt2 in scaffold99_len85036_cov0 and NAD(P)H-nir in scaffold79_len158446_cov0). To figure out whether these two genes are in different scaffolds because of an assembly artifact, we designed primers directed to the terminal regions of both fragments (ClimH_R73C and ClimH_F72C, see all the primers used in this work in Table D in S1 Supporting information). These primers were used to check, by PCR, whether the two scaffolds are contiguous on the same chromosome. We obtained a PCR fragment of ~500 bp that was cloned into pCR2.1 vector (Invitrogen) and Sanger sequenced. BLAST analysis of the sequenced products (available in Table D in S1 Supporting information) showed the presence
of regions from both scaffolds in the extremes of the PCR fragment, confirming that nrt2 and NAD(P)H-nir are clustered in this species.

Furthermore, we investigated the genomic regions flanking the clusters of C. fragrantissima, S. arctica, C. limacisporum, Phytophthora infestans and A. kerguelense in order to find additional genes in the NAP clusters of Opisthokonta and SAR. Because we found a TP_methylase Pfam domain protein (TP_methylase) clustered with NAP genes in three of these genomes, we scanned the remaining SAR and Opisthokonta for the presence of additional clusters of NAP genes with a TP_methylase. To do that, we retrieved all the TP_methylase of each organism and aligned them against the corresponding genome [TBLASTN: -evalue 1e-3]. As with NAP genes, the genomic location of each TP_methylase was annotated considering the BLAST hit with the highest score.

**Phylogenetic analyses of tetrapyrrole methylase proteins**

All the TP_methylase in euk_db were retrieved and used to detect similar sequences in prok_db [BLASTP: 1e-3]. Among the aligned sequences from prok_db, only those with a detected TP_methylase Pfam domain were kept [PfamScan]. TP_methylase sequences from euk_db and prok_db were aligned with MAFFT and trimmed with trimAl [-gappyout]. Because there were >1000 sequences in the alignment, we used FastTreeMP (version 2.1.9) [99] for the construction of the phylogenetic tree instead of IQ-TREE. We kept for further analyses the sequences in the blue clade because it included the three TP_methylase proteins found in cluster with NAP genes (sequences pointed by arrows in S19 Fig). Because in that blue clade eukaryotic sequences were monophyletic and branched within a bacterial clade, we considered all the eukaryotic sequences of this clade as a particular eukaryotic TP_methylase protein family (TPmet). We used all TPmet sequences to capture potential prokaryotic homologs of this specific family in prok_db [BLASTP: -evalue 1e-3], which were incorporated to TPmet sequences for a second phylogenetic tree (S27 Fig). To that end, sequences were aligned with MAFFT [mafft-einsi], trimmed with trimAl [-gappyout], and the tree was inferred with IQ-TREE [ultrafast bootstrap 1000 replicates, best model selected with ModelFinder]. To get a higher phylogenetic resolution of TPmet and their prokaryotic relatives, a third and last phylogenetic inference (S20 Fig) was done with sequences labeled in blue in S27 Fig (for phylogenetic inference, we used the same procedure as for the second tree).

We also constructed a Venn diagram to evaluate the coincidence between the phylogenetic distributions of TPmet and NAD(P)H-nir families along eukaryotes. In this analysis, we excluded the TPmet sequence belonging to N. vectensis (Nvec_XP_001617771) because it is located in a genomic fragment (NW_001825282.1) that most likely represents a contaminant scaffold. In particular, the phylogenetic tree revealed that this protein is identical to a region of the TPmet found in the choanoflagellate Monosiga brevicollis (Mbre_XP_001745780). We found that this M. brevicollis protein, as well as the TPmet protein found for the choanoflagellate Salpingoeca rosetta (Sros_PTSG_11107), are encoded in large genomic fragments (1259938 bp in the case of M. brevicollis), while the N. vectensis protein is found in a small genomic fragment (1325 bp). Moreover, this N. vectensis fragment entirely aligned without mismatches with the M. brevicollis fragment (CH991551, between the 280127–281451 positions), indicating that this most likely represents a contamination from the M. brevicollis genome.

**Phylogenetic analyses of the C-terminal region of CS-pNR**

Sequences from euk_db and prok_db as well as from MMETSP and Microbial Dark Matter database (MDM_db) [100] (downloaded in January 2017) were scanned for the co-presence
Pyr_redox_2 and Fer2_BFD Pfam domains [hmmssearch]. Sequences with these pair of domains were retrieved and aligned with MAFFT [mafft-einsi]. We only kept the region of the alignment that correspond to Pyr_redox_2 and Fer2_BFD Pfam domains. The alignment was further trimmed with trimAl [-gappyout], and IQ-TREE was used for the phylogenetic inference [ultrafast bootstrap 1000 replicates, best model selected with ModelFinder].

**Cells and growth conditions**

*S. arctica* JP610 was grown axenically at 12°C in 25 cm² or 75 cm² culture flasks (Corning) filled, respectively, with 5 mL or 20 mL of Marine Broth (Difco). For nitrogen limitation experiments, cells were incubated in modified L1 medium (mL1) [101] (dx.doi.org/10.17504/protocols.io.wydfs6), of the following composition (per liter): 35 g marine salts (Instant Ocean), 10 g dextrose, 5 mg NaH₂PO₄·H₂O, 1.17 x 10⁻⁵ M Na₂EDTA·2H₂O, 1.17 x 10⁻⁸ M FeCl₃·6H₂O, 9.09 x 10⁻⁷ M MnCl₂·4H₂O, 8.00 x 10⁻⁸ M ZnSO₄·7H₂O, 5.00 x 10⁻⁸ M CoCl₂·6H₂O, 1 x 10⁻⁸ M CuSO₄·5H₂O, 8.22 x 10⁻⁹ M Na₂MoO₄·2H₂O, 1 x 10⁻⁸ M H₂SeO₃, 1 x 10⁻⁸ M NiSO₄·6H₂O, 1 x 10⁻⁸ M Na₃VO₄, 1 x 10⁻⁸ M K₂Cr₂O₇, 2.96 x 10⁻⁸ M thiamine·HCl, 2.05 x 10⁻¹⁰ M biotin, 3.69 x 10⁻¹¹ M cyanocobalamin. For nitrogen supplementation experiments, mL1 medium (dx.doi.org/10.17504/protocols.io.wydfs6) was supplemented with either 100 mM NaNO₃, 100 mM (NH₄)₂SO₄ or 100 mM urea as nitrogen source, as specified in the text. Photomicrographies were taken with a Nikon Eclipse TS100 equipped with a DS-L3 camera control unit (Nikon). Images were processed with image.

**RNA isolation, cDNA synthesis and real-time PCR analyses**

The expression levels of *S. arctica* NAP genes in cultures with different nitrogen sources were analyzed using real-time PCR. *S. arctica* cells were grown for 10 days in 75 cm² cell culture flasks (Corning) with 20 mL Marine Broth (Difco). Cells were scraped and collected by centrifugation at 4500 x g for 5 min at 12°C in 50 mL Falcon tubes (Corning). Supernatant was discarded and pellets were washed twice by resuspension with 20 mL of mL1 medium to wash out any trace of Marine Broth. An aliquot of the washed cells was collected as time 0. Cells were finally resuspended in mL1 medium, distributed equally into four 25 cm² culture flasks, and supplemented with different nitrogen sources. Aliquots were collected at 6, 12 and 24 hours. At each time-point, cells were pelleted in 15 mL Falcon tubes (Corning), supernatant was discarded and the pellets were resuspended in 1 mL Trizol reagent (Invitrogen) and transferred to 1.5 mL microfuge tubes with safe lock (Eppendorf). Tubes were subjected to two cycles of freezing in liquid nitrogen and thawing at 50°C for 5 min. After this treatment, samples were kept at -20°C until further processing. To eliminate any trace of genomic DNA, total RNA was treated with Amplification Grade DNAse I (Roche) and precipitated with ethanol in the presence of LiCl. The absence of genomic DNA was confirmed using a control without RNA was treated with Amplification Grade DNAse I (Roche) and precipitated with ethanol in the presence of LiCl. The absence of genomic DNA was confirmed using a control without...
Supporting information
S1 Supporting information. Includes all the supplementary tables.
(XLSX)
S2 Supporting information. Supplementary methods for the ’Phylogenetic screening of NAPs’ section.
(PDF)
S3 Supporting information. Includes all the sequences and the alignments used in the NAP phylogenies.
(ZIP)
S4 Supporting information. Input and output files of the sequence similarity networks.
(PDF)
S1 Fig. Completeness of the nitrate assimilation pathway in the 172 sampled eukaryotic genomes. The evolutionary relationships between the sampled species, represented in a cladogram, were constructed from recent bibliographical references (see Materials and methods section). Species names were colored according to the taxonomic groups to which they belong. The presence of each NAP in each taxon is shown with symbols. Black symbols indicate genes that are found within genome clusters of NAP genes. For illustration purposes, some clades of species (e.g. Metazoa) were collapsed into a single terminal leaf. For detailed information about the taxonomic categories and the NAP profiles and NAP cluster status of each species, see Table A in S1 Supporting information. Species are labelled as to whether they include a complete (dark blue circle) or partial pathway (light blue circle). The presence of the pathway was considered complete when the transporter and the two reductase activities (i.e. NRT2, EUKNR and at least 1 of the two NIRs) were detected in the genome.
(PDF)
S2 Fig. Correlation measures of NAPs distribution. (A) Correlation (from 0 to 1) between the distributions of the four NAPs in the entire eukaryotic dataset and (B) in eukaryotes from which at least one NAP was identified. (C) Correlation between the presence of NAPs with the nutrient acquisition strategies within the entire eukaryotic dataset (from 0 to 1) (see Materials and methods section).
(PDF)
S3 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic and prokaryotic Fd-NIR amino acid sequences. The tree was rooted in the branch that separates the eukaryotic clade from the rest of the tree. Statistical support values (1000-replicates UFBoot) are shown in all nodes. Prokaryotic sequences were colored according to the corresponding phylum or class, while eukaryotes were colored according to whether they contain or not a plastid/plastid-related organelle (see panel).
(PDF)
S4 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic and prokaryotic 'Photosystem II subunit III' amino acid sequences (plastidic protein). Statistical support values (1000-replicates UFBoot) are shown for all nodes. Prokaryotic sequences were colored according to the corresponding phylum or class, while eukaryotes were colored according to whether they contain or not a plastid/plastid-related organelle (see panel).
(PDF)
S5 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic and prokaryotic 'Ribosomal protein L1' amino acid sequences (plastidic protein). Statistical support values (1000-replicates UFBoot) are shown for all nodes. Non-informative clades were collapsed. Prokaryotic sequences were colored according to the corresponding phylum or class, while eukaryotes were colored according to whether they contain or not a plastid/plastid-related organelle (see panel).

(S5 Fig. PDF)

S6 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic and prokaryotic NAD(P)H-NIR amino acid sequences. The tree was rooted in the branch that separates the eukaryotic clade from the rest of the tree. Statistical support values (1000-replicates UFBoot) are shown in all nodes. Prokaryotic sequences were colored according to the corresponding phylum or class, while eukaryotes were colored according to whether they contain or not a plastid/plastid-related organelle (see panel).

(S6 Fig. PDF)

S7 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic and prokaryotic NRT2 amino acid sequences. The tree was rooted in the branch that separates the eukaryotic clade from the rest of the tree. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Prokaryotic sequences were colored according to the corresponding phylum or class, while eukaryotes were colored according to whether they contain or not a plastid/plastid-related organelle (see panel).

(S7 Fig. PDF)

S8 Fig. Schematic representation of a maximum likelihood phylogenetic tree (IQ-TREE) including the identified EUKNR and SUOX sequences. The sulfite oxidases (SUOX) sequences were detected during the EUKNR sequence-similarity network reconstruction process. The topology suggests that EUKNR sequences are more related to SUOX without a Cyt-b5 domain, in agreement with the network results.

(S8 Fig. PDF)

S9 Fig. Maximum likelihood phylogenetic tree inferred from eukaryotic Fd-NIR amino acid sequences, with some prokaryotic sequences used as outgroup (see Materials and methods section). The tree was rooted in the branch that separates the eukaryotic clade from the bacterial. Statistical support values (1000-replicates UFBoot) are shown in all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences.

(S9 Fig. PDF)

S10 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic Fd-NIR amino acid sequences, with some prokaryotic sequences used as outgroup and including sequences from the MMETSP dataset (see Materials and methods section). The tree was rooted in the branch that separates the eukaryotic clade from the prokaryotic sequences. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Eukaryotic sequence names from euk_db are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). Sequences from MMETSP are colored in black. All sequences starting with 'UP-' correspond to prokaryotic sequences.

(S10 Fig. PDF)
S11 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NAD (P)H-NIR, with some prokaryotic sequences used as outgroup and excluding Creolimax fragrantissima and Sphaeroforma arctica sequences. The tree was rooted in the branch that separates the eukaryotic clade from the bacterial sequences, with nodes. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with ‘UP-’ correspond to prokaryotic sequences.

(PDF)

S12 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NAD (P)H-NIR, with some prokaryotic sequences used as outgroup and excluding Oomycota sequences. The tree was rooted in the branch that separates the eukaryotic clade from the bacterial sequences, with nodes. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with ‘UP-’ correspond to prokaryotic sequences.

(PDF)

S13 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NRT2, with some prokaryotic sequences used as outgroup. The tree was rooted in the branch that separates the eukaryotic clade from the bacterial sequences. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with ‘UP-’ correspond to prokaryotic sequences. Nodes with blue circles correspond to species-specific duplication events.

(PDF)

S14 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NRT2 amino acid sequences, with some prokaryotic sequences used as outgroup and including sequences from the MMETSP dataset (see Materials and methods section). The tree was rooted at the branch that separates the eukaryotic clade from the bacterial sequences. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Eukaryotic sequence names from euk_db are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). Sequences from MMETSP are colored in black. All sequences starting with ‘UP-’ correspond to prokaryotic sequences.

(PDF)

S15 Fig. Eight hypothetical scenarios evaluated for the origin and evolution of NAPs and NAP clusters in Stramenopiles and Opisthokonta (see ‘NRT2 and NAD(P)H-NIR’ Results subsection). Apart of NAPs, tpmet is also considered, as we found this gene in three NAP clusters (see ‘A tetrapyrrole methylase and the origin of NAPs in Opisthokonta’ Results section). For each scenario, we indicate the branches in which gene transfer, clustering, de-clustering and gene loss events are proposed to have occurred in the evolution of Stramenopiles and Opisthokonta. The proposed donors of the transfers are also indicated. With the exception of Sphaeroforma arctica, Creolimax fragrantissima and Corallochytrium limacisporum, the other species were grouped and the clades were named according to (i) the more inclusive taxonomical category of the taxa represented or (ii) with the four-letter code of the taxa represented (see Table A in S1 Supporting information). For each clade, a symbol of any of the four inspected genes is represented if we detected them in at least one taxa of that clade. Similarly, the largest
cluster of TPmet + NAP genes found in each clade is indicated.

(S16 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NRT2, with some prokaryotic sequences used as outgroup and excluding Creolimax fragrantissima and Sphaeroforma arctica sequences. The tree was rooted at the branch that separates the eukaryotic clade from the bacterial sequences. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences.

(PDF)

(S17 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NRT2, with some prokaryotic sequences used as outgroup and excluding sequences from Oomycota. The tree was rooted at the branch that separates the eukaryotic clade from the bacterial sequences. Statistical support values (1000-replicates UFBoot) are shown in all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences.

(PDF)

(S18 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic EUKNR, with some sulfite oxidase sequences used as outgroup. The tree was rooted in the branch that separates the EUKNR clade from the three sulfite oxidase sequences. Statistical support values (1000-replicates UFBoot) are shown in all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences.

(PDF)

(S19 Fig. Maximum likelihood phylogenetic tree (FastTree) of TP_methylase Pfam domain proteins from euk_db and prok_db (see Materials and methods section). Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences. A second phylogenetic tree (S27 Fig) was constructed using sequences from the blue clade (named as TPmet proteins, see Materials and methods section). The three sequences found in cluster with NAP genes are indicated with arrows.

(PDF)

(S20 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) of TP_methylase Pfam domain proteins (see Materials and methods section), including all the sequences from the blue clades in S27 Fig. All the eukaryotic sequences of the tree are considered to belong to a subset of tetrapyrrrole methylase proteins named TPmet family. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences. The three sequences found in cluster with NAP genes are indicated with arrows.

(PDF)
S21 Fig. Distribution of TPmet in eukaryotes. Venn diagram representing the quantitative distribution of the sampled eukaryotes (euk_db) recording the presence/absence of the NAD(P)H-nir and the TPmet genes. A ranking of the taxonomic groups that have at least one representative species with the TPmet but without the NAD(P)H-nir is also represented.

S22 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) of the regions of all the euk_db, prok_db, MMETSP and MDM_db proteins with the Pyr_redox_2 and Fer2_BFD Pfam domains (see Materials and methods section). Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences. Sequences from MMETSP are colored in black. Blue and orange clades represent the sequences corresponding to the Creolimax fragrantissima and Sphaeroforma arctica EUKNR and NAD(P)H-NIR, respectively.

S23 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NRT2, with some prokaryotic sequences used as outgroup and excluding sequences from Labyrinthula and Teretosporea. The tree was rooted at the branch that separates the eukaryotic clade from the bacterial sequences. Statistical support values (1000-replicates UFBoot) are shown in all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences.

S24 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic NAD(P)H-NIR, with some prokaryotic sequences used as outgroup and excluding sequences from Labyrinthula and Teretosporea. The tree was rooted at the branch that separates the eukaryotic clade from the bacterial sequences, with nodes. Statistical support values (1000-replicates UFBoot) are shown for all nodes. Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with 'UP-' correspond to prokaryotic sequences.

S25 Fig. Unrooted representation of a maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic (mitochondrial protein) and prokaryotic 'Cytochrome c oxidase subunit III' amino acid sequences. Prokaryotic sequences are colored according to the corresponding phylum or class, while eukaryotes are colored according to whether they contain or not a plastid/plastid-related organelle (see panel). As expected, Alphaproteobacteria is the sister group to eukaryotes, suggesting that the taxonomic representation of prok_db allow to detect proteins with signatures of Alphaproteobacteria, and hence of putative mitochondrial origin. The process of phylogenetic inference and taxonomic assignation is explained in Materials and methods section.

S26 Fig. Unrooted representation of a maximum likelihood phylogenetic tree (IQ-TREE) inferred from eukaryotic (mitochondrial protein) and prokaryotic 'Cytochrome b' amino acid sequences. Prokaryotic sequences are colored according to the corresponding phylum or class, while eukaryotes are colored according to whether they contain or not a plastid/plastid-related organelle (see panel). As expected, Alphaproteobacteria is the sister group to
eukaryotes, suggesting that the taxonomic representation of prok_db allow to detect proteins with signatures of Alphaproteobacteria, and hence of putative mitochondrial origin. The process of phylogenetic inference and taxonomic assignation is explained in Materials and methods section.

S27 Fig. Maximum likelihood phylogenetic tree (IQ-TREE) of TPmet proteins (selected from the blue clade in S19 Fig), with some prokaryotic sequences used as outgroup (see Materials and methods section). Eukaryotic sequence names are abbreviated with the four-letter code (see Table A in S1 Supporting information) and colored according to their major taxonomic group (see panel). All sequences starting with ‘UP-’ correspond to prokaryotic sequences. A third and last phylogenetic tree was constructed using sequences from the blue clades (see S20 Fig).

Acknowledgments
We thank José Luís Maestro for his help in qPCR design and qPCR results interpretation, Meritxell Antó for technical assistance, and Michelle Leger for their valuable comments on the manuscript. EOP also warmly acknowledge the MCG C. Committee for their insightful advice and final approval of figures design.

Author Contributions
Conceptualization: Eduard Ocaña-Pallarès, Sebastián R. Najle, Claudio Scazzocchio, Iñaki Ruiz-Trillo.
Data curation: Eduard Ocaña-Pallarès.
Formal analysis: Eduard Ocaña-Pallarès, Sebastián R. Najle, Claudio Scazzocchio.
Funding acquisition: Iñaki Ruiz-Trillo.
Supervision: Claudio Scazzocchio, Iñaki Ruiz-Trillo.
Writing – original draft: Eduard Ocaña-Pallarès.
Writing – review & editing: Sebastián R. Najle, Claudio Scazzocchio, Iñaki Ruiz-Trillo.

References
1. Arnold ML, Fogarty ND. Reticulate evolution and marine organisms: The final frontier? Int J Mol Sci. 2009; 10: 3836–3860. https://doi.org/10.3390/ijms10093836 PMID: 19865522
2. Corel E, Lopez P, Méheust R, Bapteste E. Network-Thinking: Graphs to Analyze Microbial Complexity and Evolution. Trends Microbiol. 2016; 24: 224–237. https://doi.org/10.1016/j.tim.2015.12.003 PMID: 26774999
3. Treangen TJ, Rocha EPC. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet. 2011; 7: e1001284. https://doi.org/10.1371/journal.pgen.1001284 PMID: 21298028
4. Martin WF. Too Much Eukaryote LGT. BioEssays. 2017; 39: 1700115. https://doi.org/10.1002/bies.201700115
5. Leger MM, Eme L, Stairs CW, Roger AJ. Demystifying Eukaryote Lateral Gene Transfer. BioEssays. 2017; 39: 1700242. https://doi.org/10.1002/bies.201700115
6. Andersson JO, Hirt RP, Foster PG, Roger AJ. Evolution of four gene families with patchy phylogenetic distributions: Influx of genes into prokrist genomes. BMC Evol Biol. 2006; 6. https://doi.org/10.1186/1471-2148-6-27 PMID: 16551952
7. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 2008; 9: 605–618. https://doi.org/10.1038/nrg2386 PMID: 18591983

8. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: The complexity hypothesis. Proc Natl Acad Sci. 1999; 96: 3801–3806. https://doi.org/10.1073/pnas.96.7.3801 PMID: 10097118

9. Chu HY, Sprouffske K, Wagner A. Assessing the benefits of horizontal gene transfer by laboratory evolution and genome sequencing. BMC Evolutionary Biology. 2018; 18. https://doi.org/10.1186/s12862-018-1164-7 PMID: 2973327

10. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004; 5: 123–135. https://doi.org/10.1038/nrg1271 PMID: 14735123

11. Naranjo-Ortíz MA, Brock M, Brunke S, Hube B, Marcit-Houben M, Gabaldón T. Widespread inter- and intra-domain horizontal gene transfer of D-amino acid metabolism enzymes in eukaryotes. Front Microbiol. 2016; 7. https://doi.org/10.3389/fmicb.2016.00007

12. Richards TA, Leonard G, Soanes DM, Talbot NJ. Gene transfer into the fungi. Fungal Biol Rev. 2011; 25: 98–110. https://doi.org/10.1016/j.fbr.2011.04.003

13. Soanes D, Richards T. Horizontal gene transfer in eukaryotic plant pathogens. Annu Rev Phytopathol. 2014; 52: 583–614. https://doi.org/10.1146/annurev-phyto-102313-050127 PMID: 25090479

14. Slot JC, Hibbett DS. Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: A phylogenetic study. PLoS ONE. 2007; 2: e1097. https://doi.org/10.1371/journal.pone.0001097 PMID: 17971860

15. Canfield DE, Glazer AN, Falkowski PG. The Evolution and Future of Earth’s Nitrogen Cycle. Science 2010; 330: 192–196. https://doi.org/10.1126/science.1186120 PMID: 20929768

16. Crawford N M., Glass D.M. A. Molecular and physiological aspects of nitrate uptake in plants. Trends Plant Sci. 1998; 3: 389–395. https://doi.org/10.1016/S1360-1385(98)01311-9

17. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. Nat Rev Microbiol. 2018; 16: 263–276. https://doi.org/10.1038/nrmicro.2018.9 PMID: 29398704

18. Crawford NM. The molecular genetics of nitrate assimilation in fungi and plants. Annu Rev Genet. 1993; 27: 115–146. https://doi.org/10.1146/annurev-genet.27.120193.000555 PMID: 8122899

19. Sanz-Luque E, Chamizo-Ampudia A, Llamas A, Galvan A, Fernandez E. Understanding nitrate assimilation and its regulation in microalgae. Front Plant Sci. 2015; 6: 899. https://doi.org/10.3389/fpls.2015.00899 PMID: 26579149

20. Keeling PJ. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu Rev Plant Biol. 2013; 64: 583–607. https://doi.org/10.1146/annurev-plant-050312-120144 PMID: 23451781

21. Richards TA, Jones MDM, Leonard G, Bass D. Marine Fungi: Their Ecology and Molecular Diversity. Ann Rev Mar Sci. 2012; 4: 495–522. https://doi.org/10.1146/annurev-marine-050312-120144 PMID: 22457985

22. Luque I, Flores E, Herrero A. Nitrite reductase gene from Synechococcus sp. PCC 7942: homology between cyanobacterial and higher-plant nitrite reductases. Plant Mol Biol. 1993; 21: 1201–1205. PMID: 8490140

23. Sato N, Awai K. “Prokaryotic pathway” is not prokaryotic: Noncyanobacterial originof the chloroplast lipid biosynthetic pathway revealed by comprehensive phylogenomic analysis. Genome Biol Evol. 2017; 9: 3162–3178. https://doi.org/10.1093/gbe/evx236 PMID: 29145666

24. Ponce-Toledo RL, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. Curr Biol. 2017; 27: 386–391. https://doi.org/10.1016/j.cub.2016.11.056 PMID: 28132810

25. Criscuolo A, Gribaldo S. Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. Mol Biol Evol. 2011; 28: 3019–3032. https://doi.org/10.1093/molbev/msr108 PMID: 21652613

26. Patmanathan JS, Lopez P, Lapointe F-J, Baptiste E. CompositeSearch: A Generalized Network Approach for Composite Gene Families Detection. Mol Biol Evol. 2017; 35: 252–255. https://doi.org/10.1093/molbev/msx283 PMID: 29092069

27. Shimodaira H. An Approximately Unbiased Test of Phylogenetic Tree Selection. Syst Biol. 2002; 51: 492–508. https://doi.org/10.1086/363655 PMID: 12079646

28. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biol. 2014; 12: e1001889. https://doi.org/10.1371/journal.pbio.1001889 PMID: 24959919
29. Dorrell RG, Gile GH, McCallum G, Baptiste EP, Klinger CM, Freeman KD, et al. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. eLife. 2017; 6: e23717. https://doi.org/10.7554/eLife.23717 PMID: 28498102

30. 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: 5764. https://doi.org/10.1038/ncomms6764 PMID: 25493338

31. Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petřželková R, Eme L, et al. A Eukaryote without a Mitochondrial Organelle. Curr Biol. 2016; 26: 1274–1284. https://doi.org/10.1016/j.cub.2016.03.053 PMID: 27185558

32. Burki F, Kaplan M, Tikhonenkov D V., Zlatogursky V, Minh BQ, Radaykina L V., et al. 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: 20152802. https://doi.org/10.1098/rspb.2015.2802 PMID: 26817772

33. Miller JJ, Delwiche CF. Phylogenomic analysis of Emiliania huxleyi provides evidence for haptophyte-stramenopile association and a chimeric haptophyte nuclear genome. Mar Genomics. 2015; 21: 31–42. https://doi.org/10.1016/j.margen.2015.02.008 PMID: 25746767

34. Burki F. The Convoluted Evolution of Eukaryotes With Complex Plastids. Adv Bot Res. 2017; 84: 1–30. https://doi.org/10.1016/bs.abr.2017.06.001

35. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature. 2012; 492: 59–65. https://doi.org/10.1038/nature11681 PMID: 23201678

36. Glockling SL, Marshall WL, Gleason FH. Phylogenetic interpretations and ecological potentials of the Mesomycetozoea (Ichthyosporea). Fungal Ecol. 2013; 6: 237–247. https://doi.org/10.1016/j.funeco.2013.03.005

37. MacDonald DW, Cove DJ. Studies on temperature-sensitive mutants affecting the assimilatory nitrate reductase of Aspergillus nidulans. Eur J Biochem. 1974; 47: 107–110. PMID: 4611763

38. Cove DJ. Genetic studies of nitrate assimilation in Aspergillus nidulans. Biol Rev Camb Philos Soc. 1979; 54: 291–327. PMID: 389305

39. Tomsett AB, Cove DJ. Deletion mapping of the niaA niaD region of Aspergillus nidulans. Genet Res. 1979; 34: 19–32. PMID: 389740

40. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, et al. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biology. 2013; 14: R11. https://doi.org/10.1186/gb-2013-14-2-r11 PMID: 23375108

41. Vega JM, Garrett RH. Siroheme: a prosthetic group of the Neurospora crassa assimilatory nitrite reductase. J Biol Chem. 1975; 250: 7980–7989. PMID: 126995

42. Fischer K, Barbier GG, Hecht H, Mendel RR, Campbell WH. Structural Basis of Eukaryotic Nitrate Reduction: Crystal Structures of the Nitrate Reductase Active Site. Plant Cell. 2005; 17: 1167–1179. https://doi.org/10.1105/tpc.104.029694 PMID: 15772287

43. Baloff S, Kavoshi G, Khodabarin B. Nitrate reductase, nitrite reductase, glutamine synthetase, and glutamate synthase expression and activity in response to different nitrogen sources in nitrogen-starved wheat seedlings. Biotechnol Appl Biochem. 2016; 63: 220–229. https://doi.org/10.1002/bab.1362 PMID: 25676153

44. Imamura S, Terashita M, Ohnuma M, Maruyama S, Minoda A, Weber APM, et al. Nitrate assimilatory genes and their transcriptional regulation in a unicellular red alga cyanidioschyzon merolae: Genetic evidence for nitrite reduction by a sulfite reductase-like enzyme. Plant Cell Physiol. 2010; 51: 707–717. https://doi.org/10.1093/pcp/pcp043 PMID: 20375110

45. Punt PJ, Strauss J, Smit R, Kinghorn JR, van den Hondel CA, Scacciozziio C. The intergenic region between the divergently transcribed niiA and niaD genes of Aspergillus nidulans contains multiple NirA binding sites which act bidirectionally. Mol Cell Biol. 1995; 15: 5688–5699. PMID: 7565720

46. Unkles SE, Hawker KL, Grieve C, Campbell EI, Montague P, Kinghorn JR. crnA encodes a nitrate transporter in Aspergillus nidulans. Proc Natl Acad Sci U S A. 1991; 88: 204–208. PMID: 1986367

47. Slot JC, Hallström KN, Matheny PB, Hibbett DS. Diversification of NRT2 and the origin of its fungal homolog. Mol Biol Evol. 2007; 24: 1731–1743. https://doi.org/10.1093/molbev/msm098 PMID: 17513882

48. Guedes R, Prosdocimi F, Fernandes G, Moura L, Ribeiro H, Ortega J. Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryote evolution. BMC Genomics. 2011; 12: S2. https://doi.org/10.1186/1471-2164-12-S4-S2 PMID: 22369087
49. Li B, Lopes JS, Foster PG, Embley TM, Cox CJ. Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. Mol Biol Evol. 2014; 31: 1697–1709. https://doi.org/10.1093/molbev/msu105 PMID: 24795089

50. Torruella G, de Mendoza A, Grau-Bové X, Ruiz-Trillo I. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. Curr Biol. 2015; 25: 2404–2410. https://doi.org/10.1016/j.cub.2015.07.053 PMID: 26365255

51. Savory F, Leonard G, Richards TA. The Role of Horizontal Gene Transfer in the Evolution of the Oomycetes. PLoS Pathog. 2015; 11: e1004805. https://doi.org/10.1371/journal.ppat.1004805 PMID: 26020232

52. Gluck-Thaler E, Slot JC. Dimensions of Horizontal Gene Transfer in Eukaryotic Microbial Pathogens. PLoS Pathog. 2015; 11: e1005156. https://doi.org/10.1371/journal.ppat.1005156 PMID: 26513155

53. Takishita K, Chikaraishi Y, Leger MM, Kim E, Yabuki A, Ohkouchi N, et al. Lateral transfer of tetrahymanol-synthesizing genes has allowed multiple diverse eukaryote lineages to independently adapt to environments without oxygen. Biol Direct. 2012; 7: 5. https://doi.org/10.1186/1745-6150-7-5 PMID: 22296756

54. Schwarz G, Mendel RR. Molybdenum cofactor biosynthesis and molybdenum enzymes. Annu Rev Plant Biol. 2006; 57: 623–647. https://doi.org/10.1146/annurev.arplant.57.032905.105437 PMID: 16669776

55. Slot JC. Fungal Gene Cluster Diversity and Evolution. Adv Genet. 2017; 100: 141–178. https://doi.org/10.1016/bs.adgen.2017.09.005 PMID: 29153399

56. Hurst LD, Pál C, Lercher MJ. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 2004; 5: 299–310. https://doi.org/10.1038/nrg1319 PMID: 15131653

57. Wisecaver JH, Rokas A. Fungal metabolic gene clusters-caravans traveling across genomes and environments. Front Microbiol. 2015; 6: 161. https://doi.org/10.3389/fmicb.2015.00161 PMID: 25784900

58. Stüeken EE, Kipp MA, Koehler MC, Buick R. The evolution of Earth’s biogeochemical nitrogen cycle. Earth-Science Rev. 2016; 160: 220–239. https://doi.org/10.1016/j.earscirev.2016.07.007

59. Grau-Bóveda X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA, et al. Dynamics of genomic innovation in the unicellular ancestry of animals. eLife. 2017; 6: e26036. https://doi.org/10.7554/eLife.26036 PMID: 28726632

60. Ámon J, Fernández-Martín R, Bokor E, Cultrone A, Kelly JM, Filippi M, et al. A eukaryotic nicotinate-inducible gene cluster: convergent evolution in fungi and bacteria. Open Biol. 2017; 7: 170199. https://doi.org/10.1098/rsob.170199 PMID: 29212709

61. Schinko T, Gallmetzer A, Amillis S, Strauss J. Pseudo-constitutivity of nitrate-responsive genes in nitrate reductase mutants. Fungal Genet Biol. 2013; 54: 34–41. https://doi.org/10.1016/j.fgb.2013.02.003 PMID: 23454548

62. Navarro FJ, Perdomo G, Tejera P, Medina B, Machín F, Guillén RM, et al. The role of nitrate reductase in the regulation of the nitrate assimilation pathway in the yeast Hansenula polymorpha. FEMS Yeast Res. 2003; 4: 149–55. https://doi.org/10.1016/S1567-1356(03)00163-6 PMID: 14613879

63. Llamas A, Igerño ML, Galván A, Fernández E. Nitrate signalling on the nitrate reductase gene promoter depends directly on the activity of the nitrate transport systems in Chlamydomonas. Plant J. 2002; 30: 261–271. PMID: 12000675

64. Sibthorp C, Wu H, Cowley G, Wong PWH, Palaima P, Morozov IY, et al. Transcriptome analysis of the filamentous fungus Aspergillus nidulans directed to the global identification of promoters. BMC Genomics. 2013; 14: 847. https://doi.org/10.1186/1471-2164-14-847 PMID: 24299161

65. Konishi M, Yanagisawa S. Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. Nat Commun. 2013; 4: 1617. https://doi.org/10.1038/ncomms2621 PMID: 23511481

66. Goh LK, Sorkin A, Blythe J. R2R3-type MYB transcription factor, CmMYB1, is a central nitrogen assimilation regulator in Cyanidioschyzon merolae. Proc Natl Acad Sci. 2009; 106: 12548–12553. https://doi.org/10.1073/pnas.0902790106 PMID: 19592510

67. Berger H, Pachlinger R, Morozov I, Goller S, Narendja F, Caddick M, et al. The GATA factor AreA regulates localization and in vivo binding site occupancy of the nitrate activator NirA. Mol Microbiol. 2006; 59: 433–446. https://doi.org/10.1111/j.1365-2958.2005.04957.x PMID: 16390440

68. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2017; 45: D12–D17. https://doi.org/10.1093/nar/gkw1071 PMID: 27899561

69. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. 2014; 42: D26–D31. https://doi.org/10.1093/nar/gkt1069 PMID: 24225321
70. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004; 32: D115–D119. https://doi.org/10.1093/nar/gkh131 PMID: 14681372

71. Kang S, Tice AK, Spiegel FW, Silberman JD, Pánek T, Čepička I, et al. Between a Pod and a Hard Test: The Deep Evolution of Amoebae. Mol Biol Evol. 2017; 34: 2258–2270. https://doi.org/10.1093/molbev/msx162 PMID: 28505375

72. Brown MW, Heiss AA, Kamikawa R, Inagaki Y, Yabuki A, Tice AK, et al. Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. Genome Biol Evol. 2018; 10: 427–433. https://doi.org/10.1093/gbe/evy014 PMID: 29360967

73. Muñoz-Gómez SA, Mejía-Franco FG, Durnin K, Colp M, Grisdale CJ, Archibald JM, et al. The New Red Algal Subphyllum Proteorhodophytina Comprises the Largest and Most Divergent Plastid Genomes Known. Curr Biol. 2017; 27: 1677–1684. https://doi.org/10.1016/j.cub.2017.04.054 PMID: 28528908

74. Ruhfel BR, Gitzendanner MA, Solitis PS, Solitis DE, Burleigh J. From algae to angiosperms—infering the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol. 2014; 14: 23. https://doi.org/10.1186/1471-2148-14-23 PMID: 24533922

75. Kurtzman CP, mateo RQ, Kolecka A, Theelen B, Robert V, Boekhout T. Advances in yeast systematics and phylogeny and their use as predictors of biotechnologically important metabolic pathways. FEMS Yeast Res. 2015; 15: https://doi.org/10.1093/femsyr/fov050 PMID: 26136514

76. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature. 2006; 443: 818–822. https://doi.org/10.1038/nature05110 PMID: 28528908

77. Janouškovec J, Tikhonenkov D V., Burki F, Howe AT, Rohwer FL, Mylnikov AP, et al. A New Lineage of Eukaryotes Illuminates Early Mitochondrial Genome Reduction. Curr Biol. 2017; 27: 3717–3724. https://doi.org/10.1016/j.cub.2017.10.051 PMID: 29174886

78. Sierra R, Cañas-Duarte SJ, Burki F, Schwelm A, Fogełqvist J, Dixelius C, et al. Evolutionary origins of rhizarian parasites. Mol Biol Evol. 2015; 33: 980–983. https://doi.org/10.1093/molbev/msv340 PMID: 26681153

79. Derelle R, López-García P, Timpano H, Moreira D. A phylogenomic framework to study the diversity and evolution of stramenopiles (= heterokonts). Mol Biol Evol. 2016; 33: 2890–2898. https://doi.org/10.1093/molbev/msw168 PMID: 27512113

80. McCarthy CGP, Fitzpatrick DA. Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes. mSphere. 2017; 2: e00095–17. https://doi.org/10.1128/mSphere.00095-17 PMID: 28435885

81. He D, Sierra R, Pawlowski J, Baldauf SL. Reducing long-branch effects in multi-protein data uncovers a close relationship between Alveolata and Rhizaria. Mol Phylogenet Evol. 2016; 101: 1–7. https://doi.org/10.1016/j.ympev.2016.04.033 PMID: 27123217

82. Derelle R, Torruella G, Klimes V, Brinkmann H, Kim E, Vlček Č, et al. Bacterial proteins pinpoint a single eukaryotic root. Proc Natl Acad Sci U S A. 2015; 112: E693–E699. https://doi.org/10.1073/pnas.1420657112 PMID: 25646484

83. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. Nucleic Acids Res. 2014; 42: D222–D230. https://doi.org/10.1093/nar/gkt1223 PMID: 24288371

84. Aittschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 22317112

85. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011; 7: e1002195. https://doi.org/10.1371/journal.pcbi.1002195 PMID: 22039361

86. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012; 40: D1178–D1186. https://doi.org/10.1093/nar/gkr944 PMID: 22110026

87. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Proc Natl Acad Sci U S A. 2003; 10: 2498–2504. https://doi.org/10.1073/pnas.1239303103 PMID: 14597658

88. Kato K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30: 3059–3066. https://doi.org/10.1093/nar/gkf436 PMID: 12136088

89. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25: 1972–1973. https://doi.org/10.1093/bioinformatics/btp348 PMID: 19505945
90. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–1213. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

91. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011; 27: 1164–1165. https://doi.org/10.1093/bioinformatics/btr088 PMID: 21335321

92. 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: 268–274. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

93. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018; 35: 518–522. https://doi.org/10.1093/molbev/msx281 PMID: 29077904

94. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast Approximation for Phylogenetic Bootstrap. Mol Biol Evol. 2013; 30: 1188–1195. https://doi.org/10.1093/molbev/mst024 PMID: 23418397

95. Kalyaanamorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14: 587–589. https://doi.org/10.1038/nmeth.4285 PMID: 28481363

96. Ward N, Moreno-Hagelsieb G. Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss? PLoS ONE. 2014; 9: e101850. https://doi.org/10.1371/journal.pone.0101850 PMID: 25013894

97. Ku C, Nelson-sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, et al. Endosymbiotic origin and differential loss of eukaryotic genes. Nature. 2015; 524: 427–432. https://doi.org/10.1038/nature14963 PMID: 26287458

98. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The Origin and Diversification of Mitochondria. Curr Biol. 2017; 27: R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015 PMID: 29112874

99. Price MN, Dehul PS, Arkin AP. FastTree 2—Approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010; 5: e9490. https://doi.org/10.1371/journal.pone.0009490 PMID: 20224823

100. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013; 499: 431–437. https://doi.org/10.1038/nature12352 PMID: 23851394

101. Guillard RRL, Hargraves PE. Stichochrysis immobilis is a diatom, not a chrysophyte. Phycologia. 1993; 32: 234–236. https://doi.org/10.2216/0031-8884-32-3-234.1