Rapid Genetic Code Evolution in Green Algal Mitochondrial Genomes

Emmanuel Noutahi,[1] Virginie Calderon,[2] Mathieu Blanchette,[3,4] Nadia El-Mabrouk,[1] and Bernd Franz Lang*,[5]

[1]Département d'Informatique et de Recherche opérationnelle (DIRO), Université de Montréal, CP 6128 succursale Centre-Ville, Montreal, QC, Canada

[2]Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada

[3]School of Computer Science, McGill University, McConnell Engineering Bldg., Montréal, QC H3A 0E9, Canada

[4]McGill Centre for Bioinformatics, McGill University, Montréal, QC, Canada

[5]Département de Biochimie, Centre Robert Cedergren, Université de Montréal, CP 6128 succursale Centre-Ville, Montreal, QC, Canada

*Corresponding author: E-mail: franz.lang@umontreal.ca.

Associate editor: Amanda Marie Larracuente

Abstract

Genetic code deviations involving stop codons have been previously reported in mitochondrial genomes of several green plants (Viridiplantae), most notably chlorophyte algae (Chlorophyta). However, as changes in codon recognition from one amino acid to another are more difficult to infer, such changes might have gone unnoticed in particular lineages with high evolutionary rates that are otherwise prone to codon reassignments. To gain further insight into the evolution of the mitochondrial genetic code in green plants, we have conducted an in-depth study across mtDNAs from 51 green plants (32 chlorophytes and 19 streptophytes). Besides confirming known stop-to-sense reassignments, our study documents the first cases of sense-to-sense codon reassignments in Chlorophyta mtDNAs. In several Sphaeropleales, we report the decoding of AGG codons (normally arginine) as alanine, by tRNA(CCU) of various origins that carry the recognition signature for alanine tRNA synthetase. In Chromochloris, we identify tRNA variants decoding AGG as methionine and the synonymous codon CGG as leucine. Finally, we find strong evidence supporting the decoding of AUA codons (normally isoleucine) as methionine in Pycnococcus. Our results rely on a recently developed conceptual framework (CoreTracker) that predicts codon reassignments based on the disparity between DNA sequence (codons) and the derived protein sequence. These predictions are then validated by an evaluation of tRNA phylogeny, to identify the evolution of new tRNAs via gene duplication and loss, and structural modifications that lead to the assignment of new tRNA identities and a change in the genetic code.

Key words: Chlorophyta, mitochondria, Sphaeropleales, tRNA evolution, codon reassignment, genetic code.

Introduction

Green plants (Viridiplantae) constitute a monophyletic group divided into two major lineages: the Streptophyta which comprises Charophyta plus land plants, and the Chlorophyta (Chlorophyceae, Ulvophyceae, Trebouxiophyceae classes forming the core Chlorophyta, and Prasinophyceae, a paraphyletic assemblage of lineages believed to have diverged early within Chlorophyta) (Lewis and McCourt 2004) (see fig. 1 and supplementary table S1, Supplementary Material online). In the following, we will use this taxonomic classification, and avoid the term "green algae" (Charophyta plus Chlorophyta, a paraphyletic grouping) that was previously common in the literature.

In recent years, the mitochondrial genomes (mtDNA) of green plants have gained interest due to their complex genome organization and pronounced structural diversity across clades, correlating with major differences in evolutionary rates (Palmer et al. 2000; Burger and Nedelcu 2012; Rodríguez-Salinas et al. 2012a). In particular, compared with the hyperinflated, sometimes multipartite mitochondrial genomes of land plants, those of Chlorophyta tend to have a more compact genome organization, with a reduced gene and intron count. To date, several distinct patterns of evolution have been described among the Chlorophyta, ranging from highly reduced/derived to ancestral types (Turmel et al. 1999; Nedelcu et al. 2000). In several Chlamydomonadales (Chlorophyceae class) and in Pedinomonas, a member of Pedinophyceae that usually groups with the core Chlorophyta, the mtDNAs feature both a greatly reduced gene content and a highly accelerated rate of sequence evolution. These mitochondrial genomes are characterized by an absence of ribosomal protein-coding genes, a severely reduced tRNA gene repertoire, and the presence of fragmented rRNA genes (Buchheim et al. 1996; Denovan-Wright et al. 1998; Kroymann and Zetsche 1998; Turmel et al. 1999; Smith and Lee 2008). By contrast, the much larger mtDNAs of several Prasinophyceae (Nephroselmis, Ostreococcus, Micromonas) and Prototheca (class Trebouxiophyceae) have retained more ancestral, prokaryotic features, characterized by the presence of genes encoding ribosomal proteins, a
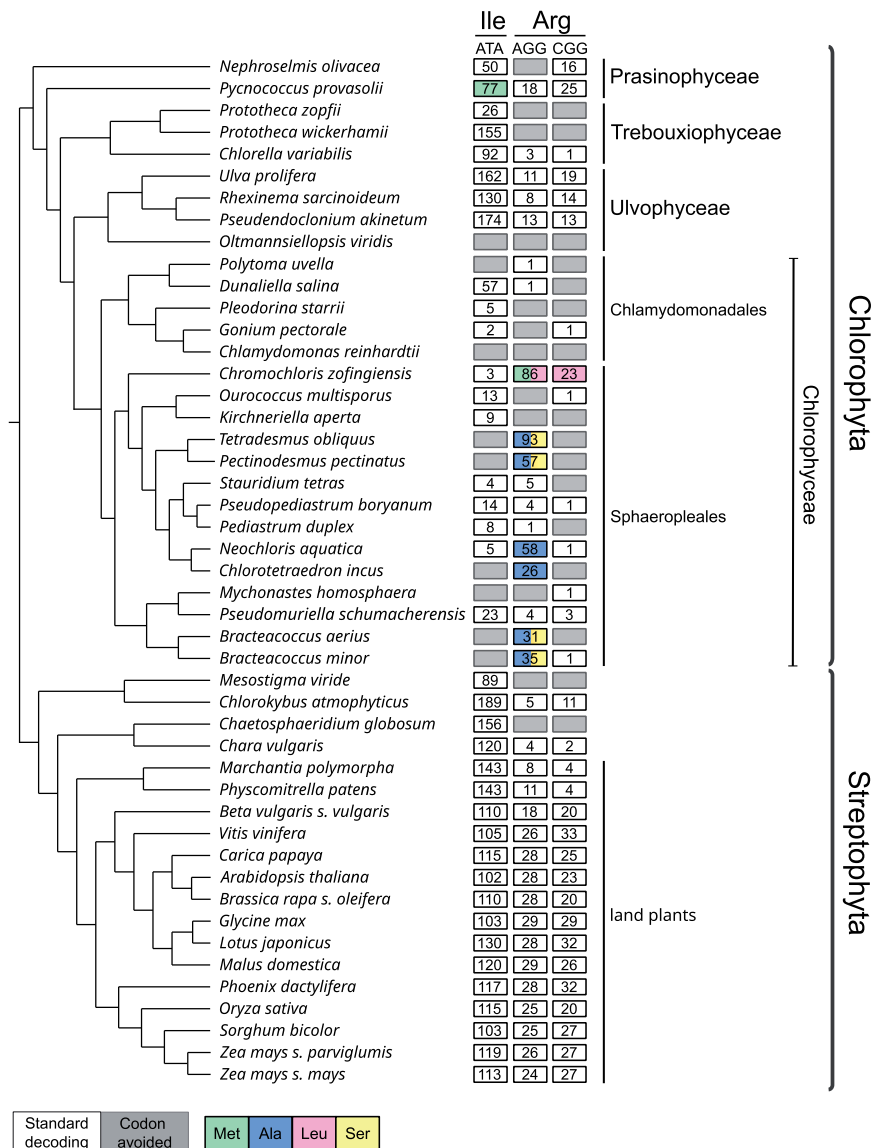
FIG. 1. Sense-to-sense codon reassignments predicted by CoreTracker in Viridiplantae. The species tree on the left is obtained using the combined information from multiple sources (see Materials and Methods). Numbers in rectangles indicate, for each mitochondrial genome, the use of the corresponding codon in protein-coding genes, ignoring intronic regions. Empty gray rectangles indicate that the codon is entirely avoided in the corresponding genome and colored rectangles indicate the decoding of the codon.

nearly complete set of tRNA genes, and the occasional presence of genes for 5S rRNA and RNase P RNA (Wolff et al. 1994; Turmel et al. 1999; Burger and Nedelcu 2012; Rodríguez-Salinas et al. 2012a). Mitochondrial genomes of an intermediate type between derived and ancestral have also been described. In most Sphaeropleales (Chlorophyceae class) and in *Pycnococcus provasolii* (Prasinophyceae), mtDNAs display features from both the reduced and ancestral types (Kück et al. 2000; Nedelcu et al. 2000; Turmel et al. 2010; Fučíková, Lewis, Gonzalez-Halphen, et al. 2014). Their size and gene content are not as reduced as that of the Chlamydomonadales, but they lack genes for RNase P RNA, 5S rRNA, and a few tRNAs. In addition, rRNA genes are fragmented and gene pieces are scrambled across the genome, as in Chlamydomonadales (Boer and Gray 1988). The intermediate type has since been suggested to represent a transitional stage of the mitochondrial genome streamlining in chlorophytes (Nedelcu et al. 2000).

The extensive changes observed in the reduced and intermediate mtDNA type in Chlorophyta have resulted in considerable modifications inside their translation machinery. Deviations of the genetic code have been noted in some of them (involving various stop codon reassignments), contrasting with their ancestral counterparts and with land plants where the standard translation code is maintained. For example, in several Sphaeropleales, the serine UCA and UCG

codons have been reported as stop codons (Kück et al. 2000; Nedelcu et al. 2000; Fučíková, Lewis, Gonzalez-Halphen, et al. 2014). There is also evidence of UAG (stop) codons being decoded as either alanine or leucine in some chlorophycean taxa (Hayashi-Ishimaru et al. 1996; Kück et al. 2000; Fučíková, Lewis, Gonzalez-Halphen, et al. 2014). Finally, in *Pedinomonas minor* and *Pycnococcus provasolii* mitochondrial genomes, UGA stop codons are reassigned to tryptophan, with the latter genome using in turn UUG and UUA leucine codons for translation termination (Turmel et al. 1999, 2010). We hypothesize that the list of reported deviations from the standard translation code may be incomplete, because all described cases only involve easily identifiable stop codon reassignments. In other words, transitions from one sense codon to another might have gone unnoticed, especially in the context of the remarkably high sequence evolution rates. Therefore, we have decided to apply CoreTracker, our recently developed method for the identification of codon reassignments (Noutahi et al. 2017), to green plant mtDNAs. CoreTracker has been accurately predicting codon reassignments, including known CUN(Leu → Thr or Ala) in yeast mitochondria (Li and Tzagoloff 1979; Sibler et al. 1981; Su et al. 2011; Ling et al. 2014) and AGG(Arg → Ser or Gly) in metazoan mitochondria (Knight et al. 2001; Sengupta et al. 2007). Expanding the repertoire of known codon reassignments will help with understanding the underlying evolutionary scenarios and biochemical constraints, including tRNA structures, tRNA synthetase activities, duplication, and neofunctionalization of tRNAs and tRNA synthetases, and biases in mitochondrial codon usage patterns.

The discussion about the genetic and evolutionary mechanisms that explain codon evolution started with the discovery of an alternative code (UGA → Trp) in human and yeast mitochondria (Barrell et al. 1979; Fox 1979), followed by a variety of additional reassignments principally identified in animal and fungal mtDNAs (e.g., AGR(Arg → Ser) in most invertebrates, CUN(Leu → Thr or Ala) in yeast) (for a review, see Lang et al. 2012). Current evolutionary scenarios have been reconciled under three main theories. The *codon capture* hypothesis states that codon reassignment arises from the disappearance of a set of codons, followed by the loss of the respective cognate tRNA, and evolution of a new tRNA (usually from a gene duplicate) that reads the codon differently (Osawa and Jukes 1989; Osawa et al. 1992). In contrast, the *ambiguous intermediate* hypothesis postulates an intermediate state in which the codon is distinctly decoded into more than one amino acid (Schultz and Yarus 1994, 1996). The *genome streamlining* hypothesis states that evolutionary pressure on genome size reduction leads to minimization of the translational machinery, potentially resulting in codon reassignments (Andersson and Kurland 1995, 1998). Finally, a *tRNA-loss-driven codon reassignment* mechanism was recently proposed (Mühlhausen and Kollmar 2014a; Kollmar and Mühlhausen 2017). It posits that codon reassignment is primarily driven by mutation or loss of a tRNA or a release factor, such that decoding of the cognate codon is halted before a new tRNA intervenes to decode the codon under a new identity.

In the following, we have undertaken an in-depth study of codon reassignment across 51 Viridiplantae mtDNAs using a framework based on the recently developed CoreTracker. This framework accounts for codon reassignments (causing a disparity between DNA and protein sequences) as well as for tRNA evolution through duplications, losses, remodeling, and structural change. Its application has led us to uncover a complex evolution of the genetic code in the reduced and intermediate mtDNA types in Chlorophyta, with several independent and distinct scenarios leading to sense and stop codon reassignments.

Results and Discussion

We have analyzed a data set comprising 19 streptophyte and 32 chlorophyte mitochondrial genomes. Genetic code alterations involving stop codons have already been reported in 17 of the analyzed chlorophytes (including the 15 selected Sphaeropleales), suggesting that additional sense-to-sense codon reassignments might also exist in these genomes. In contrast, no genetic code alteration is known in streptophytes but their phylogenetic proximity to chlorophytes makes them suitable for use as reference genomes.

UGA Stop Codon Is Used as Tryptophan in *Pedinomonas* and *Pycnococcus*

In *P. minor*, Turmel et al. (1999) have shown that UGA codons are decoded as tryptophan by a Trp-tRNA(UCA). They later reported the use of UGA for tryptophan also in the mtDNA of *P. provasolii* (Turmel et al. 2010). Due to the absence of a tRNA(UCA), it was suggested that UGA is read instead by the canonical mtDNA-encoded Trp-tRNA(CCA). From the analysis of multiple sequence alignments of standard mitochondrial proteins, we have recovered these two stop codon reassignments (see supplementary fig. S1A, Supplementary Material online, for an illustration of the sequence analysis on the Nad1 alignment). The decoding of UGA as tryptophan in *P. minor* was further confirmed by analysis of its tRNA(UCA), which exhibits known Trp-tRNA identity determinants (table 1 and supplementary fig. S1C, Supplementary Material online).

UAG Stop Codon Is Decoded as Leucine in Scenedesmaceae but Is Read as Alanine in *Neochloris* and Hydrodictyaceae

Previous analyses in certain Chlorophyceae have uncovered that UAG stop is used as a sense codon. By analyzing a multiple sequence alignment of *cox1* genes, Hayashi-Ishimaru et al. (1996) predicted that UAG codes for alanine in Hydrodictyaceae, whereas it is decoded as leucine in Scenedesmaceae. Decoding of UAG as leucine in Scenedesmaceae was later corroborated by a thorough analysis of the mitochondrial genome of *Tetradesmus obliquus* (formerly known as *Scenedesmus obliquus*) (Kück et al. 2000;

Table 1. Predicted Codon Reassignments in Sphaeropleales and Arguments Supporting the Predictions.

Reassignments	mtDNAs	Prob.	Evidence from tRNA Analysis
AGG(Arg→Ala)	*Tetradesmus obliquus*	0.99	tRNA(CCU) with major Ala-tRNA identity determinants[a]
	Pectinodesmus pectinatus	0.99	• G3:U70 invariant wobble pair in conserved G^{1}GGC^4
	Chlorotetraedron incus	0.68	• A73 at the discriminator position
	Neochloris aquatica	0.99	TFAM classification[b]
	Bracteacoccus minor	0.61	
	B. aerius	0.98	
AGG(Arg→Ser)	*T. obliquus*	0.94	No support
	P. pectinatus	0.89	
	B. aerius	0.94	
	B. minor	0.79	
AGG(Arg→Leu)	*Chromochloris zofingiensis*	0.96	No support
AGG(Arg→Met)	*C. zofingiensis*	0.98	tRNA(CCU) highly similar to Met-tRNAs.
			TFAM classification
CGG(Arg→Leu)	*C. zofingiensis*	0.99	tRNA(CCG) highly similar to Leu-tRNAs.
			Presence of multiple Leu-tRNA identity determinants[c]
			• long variable arm
			• A14 in a Reverse-Hoogsteen interaction with U8
			• A73 at the discriminator position
			TFAM classification
ATA(Ile→Met)	*Pycnococcus provasolii*	0.99	Divergent Met-tRNA(CAU) with potential posttranscriptional modification of C34 to U34
UGA(Stop→Trp)	*P. provasolii*	—	Divergent Trp-tRNA(CCA) with potential posttranscriptional modification of C34 to U34
UAG(Stop→Leu)	*T. obliquus*	—	tRNA(CUA) highly similar to Leu-tRNAs
			Presence of multiple Leu-tRNA identity determinants
	P. pectinatus		TFAM classification
UAG(Stop→Ala)	*N. aquatica*	—	tRNA(CUA) with major Ala-tRNA determinants
	Stauridium tetras		TFAM classification
	Pediastrum duplex		
	Pseudopediastrum boryanum		
UGA(Stop→Trp)	*Pedinomonas minor*	—	tRNA(UCA) with Trp-tRNA identity determinants[d]
			• A1:U72, G2:C71 base pairs
			• G73 base at the discriminator position

[a]Giegé and Eriani (2014), Saks et al. (1994), Giegé et al. (1998), McClain and Foss (1988), and Musier-Forsyth et al. (1991).
[b]TFAM, respectively, predicted a Trp-tRNA and a Met-tRNA identity for tRNA(CCU) in *B. minor* and *T. obliquus*, see supplementary table S2 of the Supplementary Material online.
[c]Tukalo et al. (2013), Asahara et al. (1993), Giegé et al. (1998), Sohm et al. (2003).
[d]Pak et al. (1992), Himeno et al. (1991).

Nedelcu et al. 2000). In a more recent analysis of Sphaeropleales mtDNAs, Fučíková, Lewis, Gonzalez-Halphen, et al. (2014), have suggested instead that UAG is decoded as leucine and not alanine, in *Neochloris aquatica* and in Hydrodictyaceae (Fučíková, Lewis, Gonzalez-Halphen, et al. 2014; Farwagi et al. 2015).

We have re-examined the decoding of UAG in two Scenedesmaceae (*Tetradesmus obliquus, Pectinodesmus pectinatus*), three Hydrodictyaceae (*Pediastrum duplex, Stauridium tetras, Pseudopediastrum boryanum*) and in *Neochloris*.

In both *Tetradesmus* and *Pectinodesmus*, we have inferred a UAG(Stop → Leu). As shown in table 1, the respective tRNAs(CUA) share several characteristics with Leu-tRNAs, most importantly, known identity determinants (table 1 and supplementary fig. S2, Supplementary Material online), and a classification as Leu-tRNA by TFAM (supplementary table S2, Supplementary Material online). On the other hand, sequence comparisons and tRNA analyses unambiguously suggest that UAG is decoded as alanine in Hydrodictyaceae and in *Neochloris* (see table 1), contrasting with the results reported in (Fučíková, Lewis, Gonzalez-Halphen, et al. 2014).

As illustrated by the Nad1 multiple sequence alignment in supplementary figure S1A, Supplementary Material online, UAG is used almost exclusively by Hydrodictyaceae and *Neochloris*, in positions where alanine is predominant in other Sphaeropleales. Furthermore, the predicted mt-tRNAs(CUA) in those genomes were classified as Ala-tRNA by TFAM (supplementary table S2, Supplementary Material online), carry Ala-tRNA identity determinants, and lack distinctive Leu-tRNA characteristics such as an elongated variable loop (supplementary fig. S3, Supplementary Material online). Finally, by analyzing publicly available RNA-seq data, we have confirmed that the predicted mt-tRNA(CUA) genes are expressed in both *T. obliquus* and *P. boryanum* (see section 1.2 of the Supplementary Material online).

Sense-to-Sense Mitochondrial Codon Reassignments in Chlorophyta

In addition to confirming and rectifying previously described stop-to-sense codon reassignments, several unknown sense-to-sense codon reassignments have also been identified. From the analysis of the 13 standard mitochondrial genes in the
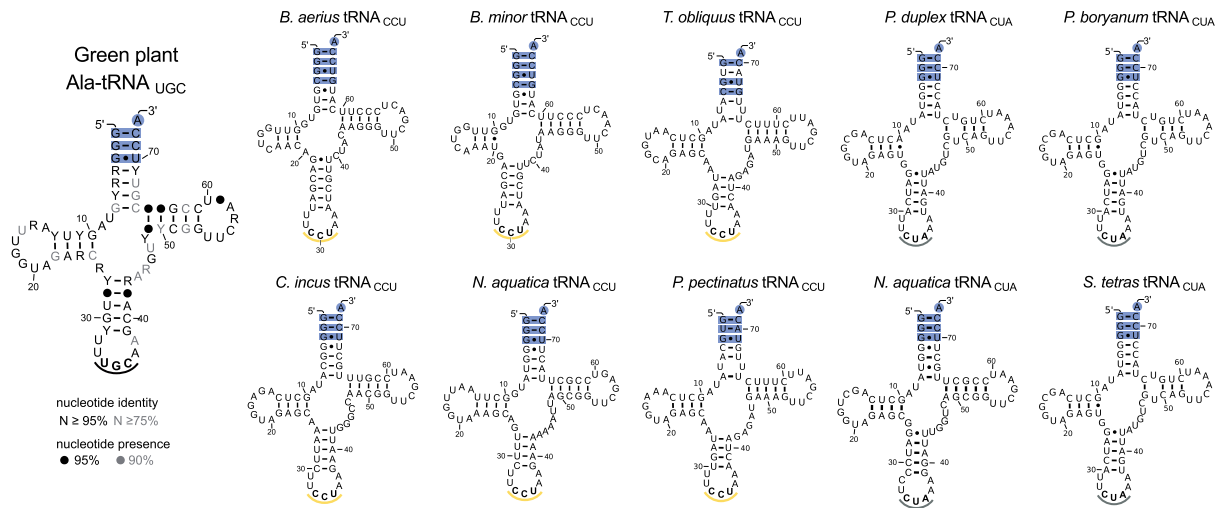
FIG. 2. The consensus secondary structure of chlorophyte mitochondrion-encoded Ala-tRNA(UGC) is shown on the left. tRNA identity determinants for alanine are indicated by purple boxes and circle. On the right side, it is shown that the mtDNA-encoded tRNA(CCU) in Sphaeropleales and tRNA(CUA) in *Neochloris aquatica* and Hydrodictyaceae display multiple Ala-tRNA identity determinants, namely the G3:U70 invariant base pair, part of the usually conserved G^1GGC^4 sequence, and A73 at the discriminator position. These results are consistent with the predicted AGG(Arg → Ala) in these Sphaeropleales and with UAG(Stop → Ala) in *N. aquatica* and Hydrodictyaceae. All tRNAs(CCU) are also missing the extra arm that characterizes Ser-tRNA (Bilokapić et al. 2009; Tukalo et al. 2013), confirming their Ala-tRNA identity.

selected 51 green plants, CoreTracker predicted a total of 14 sense-to-sense codon reassignments (fig. 1), including six AGG(Arg → Ala), four AGG(Arg → Ser), one AGG(Arg → Leu), one AGG(Arg → Met), one CGG(Arg → Leu), and one AUA(Ile → Met). It is noteworthy that all the predicted reassignments are in Chlorophyta, and only in Chlorophyta mitochondrial genomes that have been classified as intermediately derived. Whereas AUA(Ile → Met) is specific to *Pycnococcus provasolii*, reassignments involving arginine codons (AGG and CGG) are restricted to Sphaeropleales. The predicted reassignments and the supporting evidence are summarized in table 1. More detailed analyses are described as follows.

AGG Is Either Reassigned or Avoided in Sphaeropleales

In Sphaeropleales without AGG codon reassignment, the codon is usually avoided (fig. 1), suggesting that AGG codon reassignment was introduced *via* codon capture, at the emergence of Sphaeropleales. Predicted AGG reassignments in Sphaeropleales are correlated with the presence of cognate tRNAs that have a CCU anticodon, and no apparent sequence or structural similarity with typical Arg-tRNAs. We have confirmed that the predicted mt-tRNA(CCU) genes in *Chromochloris zofingiensis* and *Tetradesmus obliquus* are effectively expressed (see Supplementary Material online). In *Bracteacoccus*, the predicted mt-tRNAs(CCU) display an atypical secondary structure with a reduced D-arm (3 nucleotide pairs in the stem and four to five nucleotides in the D-loop; fig. 2). Due to the shortened D-arm, the reverse-Watson–Crick (WC) base pair 15:48 required in tertiary interactions is seemingly translocated to G14:C43 in *B. aerius* and to G14:U43 in *B. minor*.

In the two *Bracteacoccus*, in Scenedesmaceae (*Tetradesmus*, *Pectinodesmus*) and in Neochloridaceae (*Chlorotetraedron*, *Neochloris*), but surprisingly not in Hydrodictyaceae, CoreTracker predicted the decoding of AGG arginine codons as alanine. The codon was also predicted as serine in Scenedesmaceae and *Bracteacoccus*, but support for AGG(Arg → Ala) is stronger (see table 1). Furthermore, both TFAM classification and secondary structure analysis of tRNAs(CCU) favor only an Ala-tRNA identity (table 1 and fig. 2).

In stark contrast, AGG codons in *Chromochloris zofingiensis* are predicted to be reassigned to either methionine or leucine, with a high probability for both predictions ($P = 0.981$ and $P = 0.969$, respectively). However, only AGG(Arg → Met) is supported by TFAM classification (supplementary table S2, Supplementary Material online) and tRNA(CCU) is not only highly similar to mtDNA-encoded chlorophyte Met-tRNA(CAU) but also lacks distinctive Leu-tRNA identity elements including the elongated variable loop (see table 1 and supplementary fig. S4, Supplementary Material online). The combined evidence is therefore in favor of an AGG(Arg → Met) codon reassignment. Analysis of mitochondrial protein sequence alignments reveals that *C. zofingiensis* uses AUG methionine codons in ten sequence positions that in other species are predominantly leucine, implying the absence of strong selection against substitution of leucine by methionine in its genome. This suggests that *C. zofingiensis* might have also been using AGG codons as methionine in other leucine-predominant positions.

CGG Is Decoded as Leucine in *Chromochloris*

In addition to AGG(Arg → Met), the mitochondrial genetic code of *C. zofingiensis* also differs from other Sphaeropleales

mtDNA by the presence of CGG(Arg → Leu), making it one of the rare case in which two previously synonymous codons are reassigned to different amino acids. The predicted decoding of CGG as leucine is confirmed by the presence of a Leu-tRNA(CCG) transcript whose identity is confidently supported by TFAM prediction, as well as the analysis of tRNA sequence and secondary structure (see table 1 and supplementary fig. S2, Supplementary Material online).

AUA Is Used as Methionine in Pycnococcus

In *P. provasolii*, AUA codons are predicted to be reassigned from isoleucine to methionine with high support ($P = 0.988$). A majority (55.4%) of *P. provasolii*'s AUA codons were found in methionine-predominant positions, versus only 5.3% in isoleucine-predominant positions. Consequently, translation of AUA codons as methionine leads to a significant improvement of the overall protein sequence alignment (P value $= 1.23e$-06 in a Wilcoxon signed-rank test). Whereas sequence analyses strongly suggest that AUA is read as methionine in *P. provasolii* mtDNA, identification of a corresponding tRNA is less straightforward.

In the standard code, AUA codons are recognized as isoleucine by an Ile-tRNA(k^2CAU) with the cytidine at the wobble position modified to lysidine (k^2C) (Muramatsu et al. 1988; Weber et al. 1990). This modification allows Ile-tRNA(k^2CAU) to pair with AUA, unlike the standard Met-tRNA(CAU) which only pairs with AUG. In all reported genomes in which AUA is reassigned to methionine, the canonical Ile-tRNA(k^2CAU) is absent. Instead, a Met-tRNA(CAU) deciphering both AUA and AUG, often through some posttranscriptional modifications, is found (Moriya et al. 1994; Tomita et al. 1999). To understand the decoding of AUA in *Pycnococcus*, we have undertaken a systematic analysis of its tRNA(CAU).

All three tRNA(CAU) with distinct identity (fMet, Met, and Ile) that are characteristic for bacteria are usually (but not always) also found in mitochondrial genomes. Yet, only two are predicted in *Pycnococcus*. To determine the correct annotation of each mt-tRNA(CAU) in *P. provasolii*, we built a phylogeny of all green plant tRNAs with CAU anticodon (excluding identical copies), then mapped each tRNA to its predicted identity by TFAM. The genes fell into three distinct phylogenetic groups that were mostly consistent with TFAM annotations (supplementary fig. S5, Supplementary Material online). Accordingly, one tRNA(CAU) was identified as a Met-tRNA initiator (fMet). However, doubts persist in the case of a second tRNA(CAU) identity. TFAM reports it as a Met-tRNA elongator, and there is high bootstrap support (0.914) for its grouping with both Ile-tRNA(CAU) and Met-tRNA(CAU) elongator, without further resolution. We suggest that in *P. provasolii*, this peculiar tRNA(CAU) may be partially modified at the wobble position, and that further unidentified structural features enable it to decode both AUG and AUA codons as methionine. Note that in all other Chlorophyta mtDNAs, except *Pseudomuriella schumacherensis* where Ile-tRNA(CAU) is absent, AUA codons are simply avoided,

indicating that a standard Met-tRNA(CAU) is usually unable to recognize this codon.

Origin of Reassigned tRNAs in Chlorophyta

To infer the evolutionary history of reassigned tRNAs, we have performed phylogenetic analyses using both a phylogenetic tree and split networks (Huson and Bryant 2006). The phylogenetic network (fig. 3) provides a clear distinction between tRNA families, even when considering alternative splits, but is limited to only seven families of interest (see Materials and Methods). On the other hand, the phylogenetic tree (supplementary figs. S6–S9, Supplementary Material online) provides a better resolution inside tRNA groups, which is useful for the identification of recent duplications.

As anticipated, the reassigned tRNAs do not group with their expected isoacceptors under the standard decoding. They also appear to have various distinct origins (see table 2).

Distinct Origins of Ala-tRNAs(CCU) in Sphaeropleales

According to our phylogenetic analyses (fig. 3 and supplementary fig. S6, Supplementary Material online), the two *Bracteacoccus* Ala-tRNAs(CCU) cluster with the Trp-tRNA(CCA) group. Interestingly, an analysis of gene order in *Bracteacoccus* is consistent with their emergence from an ancestral tandem duplication of Trp-tRNA(CCA) (supplementary fig. S7, Supplementary Material online), followed by alloacceptor tRNA remodeling. The Trp-tRNA origin hypothesis is further supported by the TFAM classification, which predicts *B. minor*'s tRNA(CCU) as Trp-tRNA and *B. aerius*'s tRNA(CCU) as Ala-tRNA, which is coherent given their inferred ancestral origin and current identities.

On the other hand, the predicted Ala-tRNA(CCU) in *C. incus* and *N. aquatica* cluster with the Ala-tRNAs(UGC) (fig. 3 and supplementary fig. S8, Supplementary Material online), suggesting that the reassigned Ala-tRNAs in Neochloridaceae originated from an ancestral duplication of Ala-tRNA(UGC). Yet, there is no evidence for a tandem duplication scenario of Ala-tRNA(UGC). It is possible, either that the duplication event was not in tandem, or it was followed by subsequent mitochondrial genomic rearrangements, which are frequent in Sphaeropleales, especially in Neochloridaceae Fučíková, Lewis, Gonzalez-Halphen, et al. (2014). After duplication, the new copy did not diverge much from its precursor, with most sequence changes located in the anticodon arm, while the acceptor stem was left almost intact (see fig. 2 and supplementary fig. S3, Supplementary Material online). As alanine identity determinants remain conserved, the new tRNA(CCU) is able to decode AGG codons as alanine in both *N. aquatica* and *C. incus*.

Finally, the Ala-tRNA(CCU) identified in the two Scenedesmaceae were either found inside a Trp-tRNA(CCA) or a Met-tRNA(CAU) (see fig. 3 and supplementary fig. S9, Supplementary Material online). In fact, Trp-tRNA(CCA) and Met-tRNA(CAU) share a common ancestry according to the tRNA phylogeny. Sequence similarity analysis between *T. obliquus* mt-tRNAs indicates that tRNA(CCU) is more similar to Met-tRNA(CAU) than to Trp-tRNA(CCA)
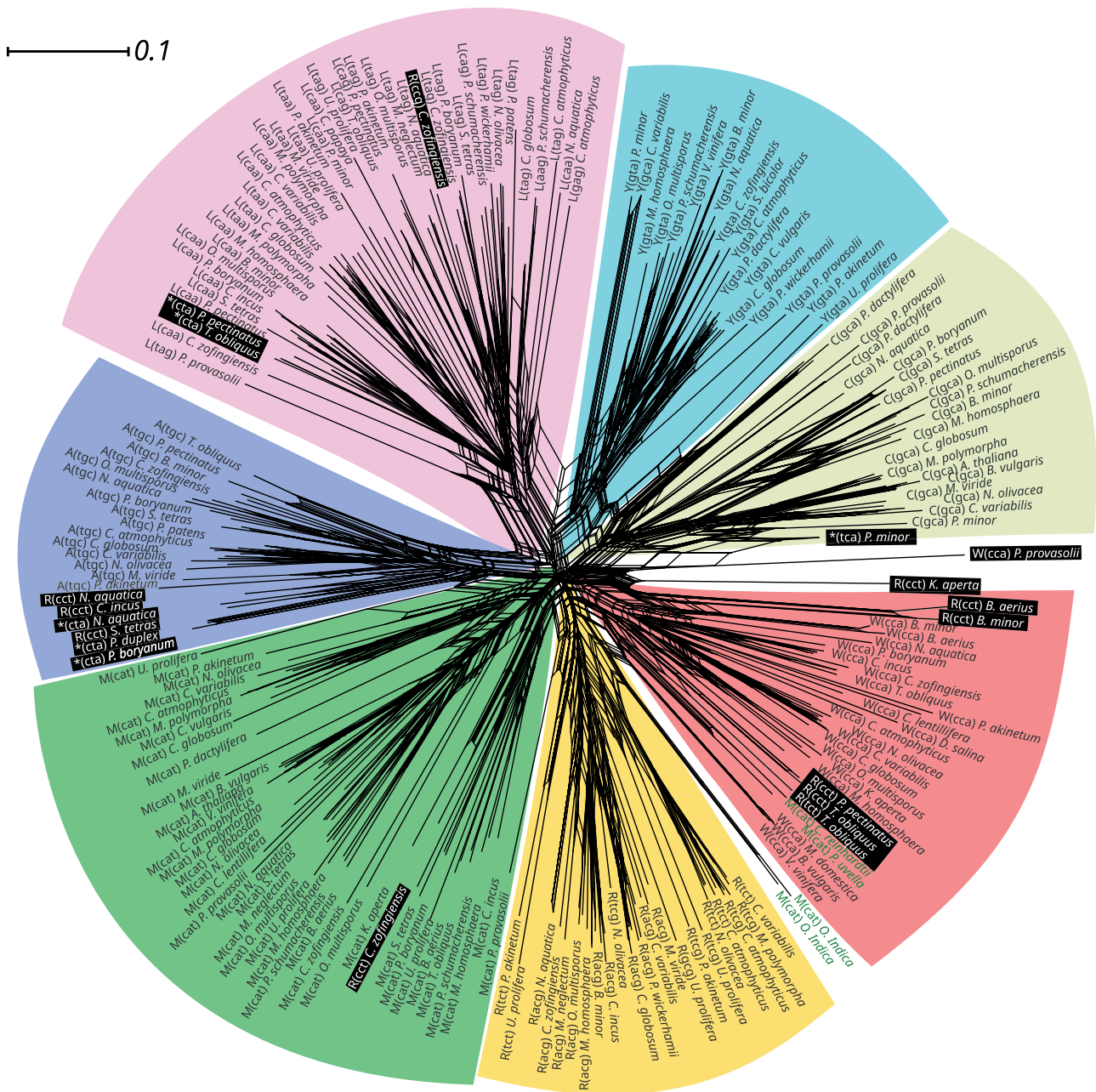
Fig. 3. Unrooted phylogenetic network of 321 tRNAs from seven groups (Met, Arg, Tyr, Cys, Ala, Leu, and Trp) constructed using the neighbor-net method as implemented in SplitsTree v4.14.6 (see Materials and Methods). Each group of tRNAs is shown with a different color and sequences that do not cluster with their group are displayed using the group color. For clarity, the label of some leaves that cluster with their respective tRNA isoacceptors is not shown. Mutant tRNAs that could potentially be involved in codon reassignments are shown in white, with a black background. Note that the identity of tRNAs under the standard decoding is preserved here to highlight their unusual affiliation.

(62.5% vs. 57.8% identity). A similar result was obtained for *P. pectinatus*. Furthermore, TFAM predicts a Met-tRNA identity for the two tRNAs(CCU), although with weak support (supplementary table S2, Supplementary Material online). This observation suggests that Ala-tRNA(CCU) in Scenedesmaceae originates from an alloacceptor remodeling after duplication of Met-tRNA(CAU). The inferred Met-tRNA(CAU) origin of Ala-tRNA(CCU) and Trp-tRNA(CCA) in Scenedesmaceae is also concordant with the fact that only one nucleotide change is needed to switch the anticodon from one tRNA to another.

Codon Reassignments in *Chromochloris* Caused by tRNA Isoacceptor Remodeling

In *C. zofingiensis*, besides a Met-tRNA(CCU) decoding AGG, a Leu-tRNA(CCG) able to decode CGG codons has been identified. These two tRNAs were, respectively, grouped with their corresponding tRNA isoacceptors (fig. 3 and supplementary figs. S9 and S10, Supplementary Material online). Furthermore, sequence analyses show a high sequence and structure similarity between the corresponding pairs (supplementary figs. S2 and S4, Supplementary Material online). As tRNAs with such similar sequences and with the same

Table 2. Evolutionary Origin of the Mutant mt-tRNAs Decoding Codons under A New Identity in Chlorophytes.

tRNA	mtDNA	Evolutionary Origin
Ala-tRNA(CCU)	*Bracteaccocus*	Tandem duplication of Trp-tRNA(CCA) at early divergence of *Bracteaccocus* + alloacceptor remolding
	Neochloris	Duplication of Ala-tRNA(UGC) in an ancestral mtDNA of Neochloridaceae + isoacceptor remolding
	Chlorotetraedron	
	Tetradesmus	Alloacceptor remolding of possible Met-tRNA copy
	Pectinodesmus	
Met-tRNA(CCU)	*Chromochloris*	Recent duplication of Met-tRNA(CAU) + isoacceptor remolding
Leu-tRNA(CCG)	*Chromochloris*	Recent duplication of Leu-tRNA + isoacceptor remolding
Ala-tRNA(CUA)	*Neochloris*	Duplication of Ala-tRNA(UGC) + isoacceptor remolding
	Hydrodictyaceae	
Leu-tRNA(CUA)	*Tetradesmus*	Recent duplication of Leu-tRNA(CAA) + isoacceptor remolding
	Pectinodesmus	
Trp-tRNA(UCA)	*Pedinomonas*	Recent tandem duplication of Cys-tRNA(UCC) + alloacceptor remolding
Trp-tRNA(CCA)	*Pycnococcus*	Divergent tRNA(CCA) of unknown origin
Met-tRNA(CAU)	*Pycnococcus*	Divergent tRNA(CAU) of unknown origin

anticodon are not present in the mtDNAs of close species, Leu-tRNA(CCG) and Met-tRNA(CCU) very likely originated from recent duplications in *C. zofingiensis* mitochondria. More precisely, Leu-tRNA(CCG) arises from a tandem duplication of Leu-tRNA(UAG) (see supplementary figs. S2 and S10, Supplementary Material online), followed by mutations in the anticodon. Likewise, Met-tRNA(CCU) emerges from a recent isoacceptor remodeling of a Met-tRNA(CAU) copy and reads AGG codons as Met (supplementary fig. S4, Supplementary Material online).

Decoding of UAG as Sense Codon by tRNA Isoacceptor Remodeling in Scenedesmaceae, Hydrodictyaceae, and *Neochloris*

The Leu-tRNA(CUA) identified in Scenedesmaceae and the Ala-tRNA(CUA) identified in *Neochloris* and Hydrodictyaceae, respectively, share several features with their corresponding isoacceptors (supplementary figs. S2 and S3, Supplementary Material online) and group with them in phylogenetic clustering (fig. 3 and supplementary figs. S8 and S10, Supplementary Material online). The Leu-tRNA(CUA) in Scenedesmaceae originates from a tandem duplication of Leu-tRNA(CAA) followed by mutations of the anticodon to recognize UAG codons (supplementary fig. S7, Supplementary Material online). Similarly, Ala-tRNA(CUA) in Hydrodictyaceae and *Neochloris* most likely emerge from a duplication of Ala-tRNA(UGC) that accumulates mutations in its anticodon to allow UGA recognition, in a similar way as Ala-tRNA(CCU) in Neochloridaceae (supplementary figs. S3 and S7, Supplementary Material online).

Trp-tRNA(UCA) Originates from Cys-tRNA(GCA) in *Pedinomonas*

In *P. minor*, a Trp-tRNA(UCA) decoding the UGA stop codons as tryptophan has been predicted previously. To accommodate effective decoding of UGA codons, the standard tRNA(CCA) has just to be modified by a single change from C to U in the anticodon. Unexpectedly, the corresponding *P. minor* tRNA groups with the Cys-tRNA(GCA) and the Tyr-tRNA(GUA) of the same genome (see supplementary fig. S1B, Supplementary Material online), pointing to their common

ancestry *via* recent duplications. Interestingly, its mtDNA is extremely reduced, missing several tRNA genes, but encoding two Tyr-tRNA(GUA) both adjacent to Cys-tRNA(GCA) and Trp-tRNA(UCA) (see supplementary fig. S1D, Supplementary Material online). This observation, taken together with their observed grouping in the phylogeny, suggests that both Tyr-tRNA(UCA) and Trp-tRNA(GCA) originate from recent tandem duplications of Cys-tRNA(GCA), after the genome had already undergone its genetic streamlining. Due to the high similarity that Trp-tRNA(UCA) shares with Cys-tRNA(GCA), TFAM even mistakenly classified it as a Cys-tRNA, but its identity is clearly Trp-tRNA as it exhibits several Trp-tRNA identity determinants (see table 1) but lacks major Cys-tRNA elements like the universally conserved U73 at the discriminator position (Pallanck et al. 1992; Hamann and Hou 1995; Mallick et al. 2005).

Divergent tRNAs Explain Codon Reassignments in *Pycnococcus*

In *P. provasolii* mtDNA, sequence analysis predicts AUA(Ile → Met) and UGA(Stop → Trp) codon reassignments. The decoding of these two codons is difficult to assess since tRNAs with matching anticodons were not found in the genome. However, the prominent occurrence of these two particular codons in coding regions suggests that they are read, perhaps less effectively, by other tRNAs.

In the mitochondrial genome of *P. provasolii*, we have identified one fMet-tRNA(CAU) and a second tRNA(CAU) of uncertain identity. As stated earlier, this second divergent tRNA(CAU) of unknown origin might be able to decode AUA codons as methionine.

The mtDNA genome of *P. provasolii* also exhibits a divergent Trp-tRNA(CCA), currently of unknown origin, that does not properly cluster with other mtDNA-encoded chlorophyte Trp-tRNAs(CCA). This tRNA(CCA) must be able to read both UGG and UGA codons, via some sequence or structural modification. Indeed, a single base mutation of "G24" to "A24" in the D-arm of *E. coli* Trp-tRNA(CCA) has been linked to insertion of tryptophan when reading UGA codons Hirsh (1971); Cochella and Green (2005); Schmeing et al. (2011). Because both UGA and UGG codons are almost equally used

as tryptophan codons (45% vs. 55%; Turmel et al. 2010), an efficient decoding of both codons is required. In fact, a C-to-U editing event at the wobble position of a tRNA(CCA) has been reported as an efficient way of allowing such decoding in *Leishmania tarentolae* (Alfonzo et al. 1999; Lang et al. 2012). However, we have not found any evidence supporting such modification in the tRNA(CCA) transcript.

Genetic Code Alterations and Chlorophyta Phylogeny

Despite the increasing number of sequenced chlorophyte genomes, their evolution remains poorly understood. Recent studies have struggled to resolve the phylogenetic relationships among deep, fast-evolving lineages (Fučíková, Lewis, Gonzalez-Halphen, et al. 2014; Fučíková, Lewis, Lewis, et al. 2014; Lemieux et al. 2014a; Fučíková et al. 2016; Sun et al. 2016; Fang et al. 2017). In particular, the evolution of the morphologically simple and similar, yet genetically divergent, Sphaeropleales is still not well understood (Fučíková and Lewis 2012; Tippery et al. 2012; Fučíková, Lewis, Gonzalez-Halphen, et al. 2014; Fučíková, Lewis, Lewis, et al. 2014; Fučíková et al. 2016). Previous attempts at inferring a phylogenetic tree for Sphaeropleales have revealed conflicting signals between mitochondrial, nuclear, and chloroplast data, with the mitochondrial protein-based tree displaying stronger inconsistencies toward the other two (Fučíková, Lewis, Gonzalez-Halphen, et al. 2014; Fučíková et al. 2016). Similar erroneous phylogenetic placements, as a result of long-branch attraction, have also been observed for other chlorophytes when concatenated mitochondrial data are used (Turmel et al. 1999; Nedelcu et al. 2000; Pombert et al. 2004). These errors have mainly been attributed to systematic errors of phylogenetic reconstructions and to the fast substitution rate of sequence evolution in Chlorophyta mtDNAs (Fang et al. 2017).

Although codon and amino acid usage bias have been shown to influence the accuracy of phylogenetic reconstruction in green plants (Turmel et al. 1999; Cox et al. 2014), the potential effect of genetic code alterations remains to be assessed. Hence, we have inferred a new Bayesian tree from the accurately translated mtDNA-encoded protein sequences (fig. 4). As expected, some topological differences, mainly regarding the position of genera considered as *incertae sedis* within Sphaeropleales (Bracteacoccaceae, Mychonastaceae, Pseudomuriellaceae, Chromochloridaceae), were observed.

In this new phylogeny, most chlorophyte groups were successfully recovered as monophyletic. Outside of Sphaeropleales, the observed branching pattern is nearly identical to the one reported in previous mitochondrial trees, with the exception of *Pycnococcus* now affiliating with other prasinophytes as observed in phylogenies based on nuclear and chloroplast data (Lemieux et al. 2014b; Leliaert et al. 2016; Fang et al. 2017; Satjarak et al. 2017).

Regarding the branching inside Sphaeropleales, we have recovered the Selenastraceae genera (*Monoraphidium*, *Ourococcus*, and *Kirchneriella*) as monophyletic and the sister relationship to Scenedesmaceae + Neochloridaceae + Hydrodictyaceae, consistent with previous inferences based on the nuclear 18S rRNA and chloroplast genes (Krienitz et al.

2011; Tippery et al. 2012; Fučíková et al. 2016). Our Bayesian tree subdivides Sphaeropleales into two main lineages: (*Mychonastes*, *Bracteaccoccus*) and the group containing (*Pseudomuriella*, *Chromochloris*) + (Selenastraceae, Scenedesmaceae, Neochloridaceae), which received weak support (pp = 0.57). The alternative and preferred topology in the maximum likelihood tree, affiliates (*Pseudomuriella*, *Chromochloris*) with (*Mychonastes*, *Bracteaccoccus*) instead. This alternative grouping is within expectations (Fučíková and Lewis 2012; Fučíková et al. 2013), because with the exception of *Mychonastes*, all the above-named species display the same cellular organization and are morphologically almost indistinguishable (Fučíková and Lewis 2012). It was hypothesized that their shared characteristics could be monophyletic within Sphaeropleales (Fučíková and Lewis 2012; Fučíková et al. 2013), but phylogenetic inferences often fail to recover it (Fučíková, Lewis, Gonzalez-Halphen, et al. 2014; Farwagi et al. 2015). Our new phylogeny using corrected protein sequences is not in disagreement with that hypothesis, and further suggests that the morphologically distinct *Mychonastes* might also be included within the clade. Previously, placement of the latter has been challenging, with various unsupported and conflicting positions inferred (Fučíková et al. 2016).

In light of our new results, topology incongruence in chlorophyte phylogenies could have also stemmed from undetected sense-to-sense codon reassignments introducing biases in the inferences. Indeed, the phylogenetic tree inferred using correctly translated mitochondrial proteins seems in better agreement with previous inferences based on nuclear and chloroplast data. It is likely that the site-heterogeneous CAT-GTR model we used with PhyloBayes has been useful in coping with long-branch attraction (Lartillot et al. 2007) in these fast-evolving species, yet using the correct genetic codes has been likely as important.

Evolution of the mtDNA Genetic Code in Chlorophyta Is Facilitated by Genome Minimization

One of the outcomes of our study is the observation that changes in the genetic code only occur in mitochondrial genomes that have undergone reduction (Sphaeropleales, *Pycnococcus* and *Pedinomonas*). Yet, the severely reduced mitochondrial genomes of Chlamydomonodales do not feature any genetic code alteration, suggesting that the reduced gene count is not a sufficient condition for codon reassignment. The same conclusion applies to yeast species (such as *S. cerevisiae* and *Ashbya* species) that have lost all *nad* genes in their mtDNA Sibler et al. (1981); Ling et al. (2014); Su et al. (2011); Freel et al. (2015). At a more general level, Massey and Garey (2007) uncovered a negative correlation between the mtDNA-encoded proteome size and the number of observed genetic code alterations. They proposed that reduced "proteomic constraints" due to genome minimization allow changes in the genetic code to be more tolerated, and less likely to be lethal. This hypothesis differs from the genome streamlining model of Kurland and colleagues (Andersson and Kurland 1995, 1998), in which shrinkage of the tRNA repertoire is the driving force for codon reassignments.
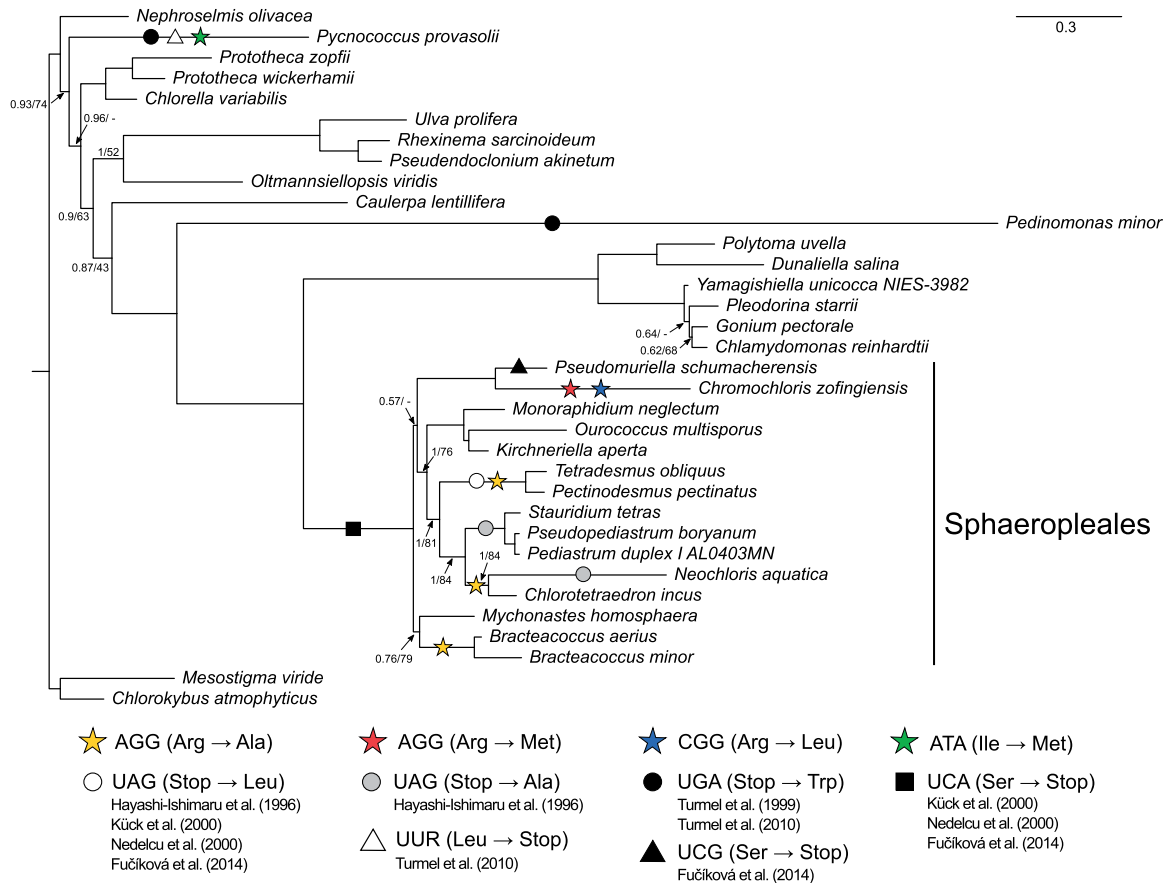
FIG. 4. Phylogeny of 30 core chlorophytes inferred using a supermatrix built from the properly translated sequences of the 13 standard mitochondrial proteins. The tree presented here is the majority-rule posterior consensus tree inferred with PhyloBayes under the CATGTR + GAMMA model. Both posterior probability (pp) from the Bayesian analyses (left value) and bootstrap support from RAxML LG + GAMMA maximum likelihood reconstruction (right value) are shown. A dash (−) is used to indicate when a branch is not present in the ML tree. Internal nodes without any label have a parental branch with a PP value of 1.0 and a bootstrap support >95%. Branch lengths are shown and correspond to the estimated number of amino acid substitutions per site. The tree is rooted using *Mesostigma* and *Chlorokybus* as outgroups. Known deviations of the mitochondrial code are also indicated, with the standard genetic code assumed to be used at the root of the phylogeny.

Although we agree that circumstances provided by the reduced mtDNA-encoded proteome size increase the likelihood of codon reassignments, we argue that a combination of several additional factors (e.g., codon usage patterns, GC fluctuation, genetic drift, gain/loss of tRNAs) are also at play. The importance of such factors are discussed below.
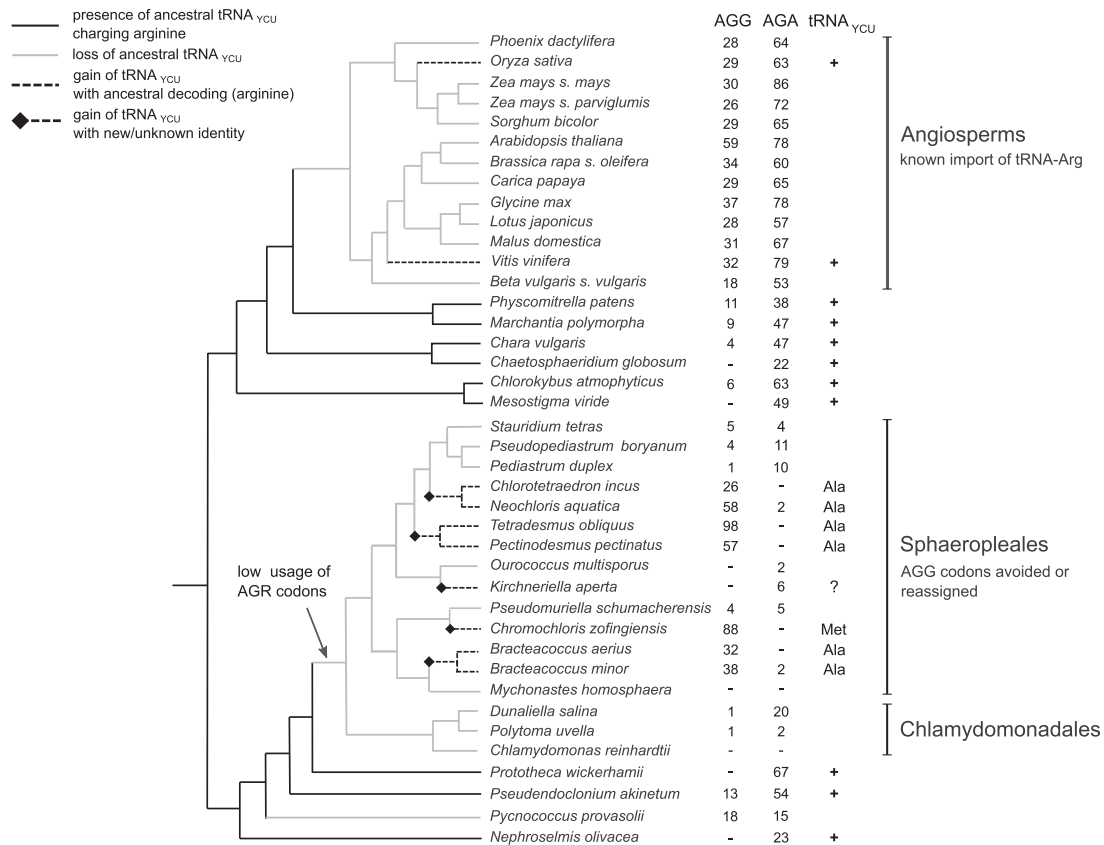
Polyphyly of AGG Decoding Is Driven by Loss of Ancestral tRNA(UCU) and Codon Disappearance

Our analyses suggest repeated, reassignment of AGG codons during the mitochondrial genome evolution of Sphaeropleales. We have inferred a distinct evolutionary origin for Ala-tRNA(CCU) in Scenedesmaceae, Neochloridaceae, and *Bracteacoccus*, and uncovered an entirely different decoding of the same codon as methionine in *Chromochloris*. Since the clade comprising Selenastraceae + Scenedesmaceae + Hydrodictyaceae + Neochloridaceae was recovered with high statistical support (pp = 1), and no alternative decoding of AGG codons was found in neither Selenastraceae nor Hydrodictyaceae, our hypothesis of a

distinct origin for tRNA(CCU) in Sphaeropleales seems valid. A similar polyphyletic decoding was reported in yeast nuclear genomes where CUG (leucine) codons are reassigned to serine or alanine in some species (Mühlhausen and Kollmar 2014a; Krassowski et al. 2018). The alternative decoding of CUG codons in these nuclear genomes was linked to the loss of the Leu-tRNA(CAG) that normally recognizes this codon (Mühlhausen et al. 2016; Kollmar and Mühlhausen 2017).

The reassignment of AGG codons in Sphaeropleales appears to be correlated with the loss of the ancestral Arg-tRNA(UCU) in all Chlorophyceae (Sphaeropleales and Chlamydomonadales; fig. 5). However, the loss of tRNA(UCU) was not the sole prerequisite of the reassignment. In angiosperm mtDNAs, tRNA(UCU) is also lost (see fig. 5), but rather than reassigning the cognate codons, a new Arg-tRNA is imported from the nucleus (Marechal-Drouard et al. 1995; Glover et al. 2001).

Obviously, AGG reassignment in Sphaeropleales cannot be explained by the ambiguous intermediate mechanism. Under this model, the codon would need to be simultaneously assigned to two or more tRNAs in the common ancestor

Fig. 5 legend:
— presence of ancestral tRNA(YCU) charging arginine
— loss of ancestral tRNA(YCU)
- - - gain of tRNA(YCU) with ancestral decoding (arginine)
◆- - - gain of tRNA(YCU) with new/unknown identity

	AGG	AGA	tRNA(YCU)	
Phoenix dactylifera	28	64		
Oryza sativa	29	63	+	
Zea mays s. mays	30	86		
Zea mays s. parviglumis	26	72		
Sorghum bicolor	29	65		
Arabidopsis thaliana	59	78		Angiosperms
Brassica rapa s. oleifera	34	60		known import of tRNA-Arg
Carica papaya	29	65		
Glycine max	37	78		
Lotus japonicus	28	57		
Malus domestica	31	67		
Vitis vinifera	32	79	+	
Beta vulgaris s. vulgaris	18	53		
Physcomitrella patens	11	38	+	
Marchantia polymorpha	9	47	+	
Chara vulgaris	4	47	+	
Chaetosphaeridium globosum	-	22	+	
Chlorokybus atmophyticus	6	63	+	
Mesostigma viride	-	49	+	
Stauridium tetras	5	4		
Pseudopediastrum boryanum	4	11		
Pediastrum duplex	1	10		
Chlorotetraedron incus	26	-	Ala	
Neochloris aquatica	58	2	Ala	
Tetradesmus obliquus	98	-	Ala	
Pectinodesmus pectinatus	57	-	Ala	Sphaeropleales
Ourococcus multisporus	-	2		AGG codons avoided or
Kirchneriella aperta	-	6	?	reassigned
Pseudomuriella schumacherensis	4	5		
Chromochloris zofingiensis	88	-	Met	
Bracteacoccus aerius	32	-	Ala	
Bracteacoccus minor	38	2	Ala	
Mychonastes homosphaera	-	-		
Dunaliella salina	1	20		
Polytoma uvella	1	2		Chlamydomonadales
Chlamydomonas reinhardtii	-	-		
Prototheca wickerhamii	-	67	+	
Pseudendoclonium akinetum	13	54	+	
Pycnococcus provasolii	18	15		
Nephroselmis olivacea	-	23	+	

low usage of AGR codons

Fig. 5. Evolutionary history of tRNA(YCU) in green plants. The usage of AGR codons is shown in the first two columns, with a dash used to indicate when the codon is missing in the coding regions of the considered genome. The presence/absence of tRNA(YCU) in each genome is also shown (third column). Moreover, the gain and loss history of tRNA(YCU) in each lineage during their mitochondrial evolution is indicated on the phylogeny.

of Chlorophyceae, requiring the independent loss of tRNA(UCU) in Sphaeropleales and Chlamydomonadales. However, because AGG codons are decoded by several tRNAs of distinct origins, one has to assume either the presence of all these tRNAs at once, or transitions through successive ambiguous decoding states. Neither of these scenarios seems probable, as they require several consecutive tRNA losses in each lineage.

In contrast, under the codon capture hypothesis, the reassignment would need to be preceded by the disappearance of AGG codons in all Sphaeropleales. The AT-rich (53–70%) mitochondrial genomes of the core Chlorophyta actually exhibit a preference for AGA over the synonymous AGG codon. In most chlorophycean mtDNAs where a reassignment of AGG was not predicted, the codon is either missing or avoided. The additional absence of AGG in Prototheca suggests that the disappearance of the codon could have predated the loss of tRNA(UCU), as required by the codon capture model. However, tRNA(UCU) translates the AGR block, and if it were deleted before AGA disappearance, there would likely be no alternative tRNAs to translate AGA. To avoid disrupting mRNA translation, the codon capture hypothesis would require liberation of the complete AGR codon family before the reassignment. Severely reduced usage of AGA codons in Sphareropleales does indicate that AGA could have ultimately vanished before AGG reassignment

(fig. 5), but the more precise timing of that disappearance relative to the loss of tRNA(UCU) remains to be established.

We favor the hypothesis that the mitochondrial genome minimization process in Chlorophyceae might have caused not only the loss of tRNA(UCU) but also contributed to AGR usage reduction. For example, more than a third of AGA codons in coding regions of Prototheca and Pseudendoclonium are found in genes that are missing in Sphaeropleales mtDNAs.

After the loss of Arg-tRNA(UCU) and the disappearance of AGG codons, new mutant tRNAs(CCU) with high specificity for AGG codons emerged in some Sphaeropleales, via tRNA duplication and remodeling. These new tRNAs were able to decode AGG under a new identity when the codon was reintroduced. One major characteristic of the mutant tRNAs(CCU) is that they are recognized by aminoacyl-tRNA synthetases (aaRS) that either do not use the anticodon as part of their identity determinants or are nondiscriminative against it. Since such aaRS are infrequent Giegé et al. (1998), the new identity of the mutant tRNAs(CCU) was mainly restricted to alanine, resulting in a remarkable example of evolutionary convergence in Sphaeropleales.

Our hypothesis of a gain of the new tRNAs before reappearance of the AGG codon is supported by the respective presence of divergent tRNA(CCU) and tRNA(UCU) in K. aperta and T. obliquus (see fig. 3), although the cognate

AGG and AGA codons are, respectively, missing in these genomes (fig. 5).

Gain of a New tRNA Caused CGG Codon Reassignments in *Chromochloris*

In *C. zofingiensis*, CGG codons are decoded as leucine by a Leu-tRNA(CCG) that originates from the isoacceptor remodeling of a Leu-tRNA. This reassignment follows a mechanism different from AGG reassignment.

In all Sphaeropleales mtDNAs, CGG codons are naturally avoided. The observed low use of the codon can be attributed both to its reduced specificity toward the Arg-tRNA(ACG) decoding the CGN family and to the high AT-pressure in Sphaeropleales mtDNAs leading to general avoidance of GC-rich codons. Following the disappearance of the codon, the codon capture model would also require the loss of the ancestral Arg-tRNA(ACG), which remains present in all Sphaeropleales. We propose, instead, that the reassignment occurred due to the gain of a new Leu-tRNA(CCG) with higher anticodon specificity for CGG codons. During a transitional period of decoding ambiguity, the codon could have been read by both Leu-tRNA(CCG) and Arg-tRNA(ACG). However, this decoding ambiguity would have minimal impact due to the low use of CGG at that time point. Relaxation of the mutation bias would then allow increased use of the codons that were fully captured by the new Leu-tRNA(CCG).

Mutant tRNAs Decoding AUA and UGA in *Pycnococcus*

In *Pycnococcus*, we predicted an AUA(Ile → Met) and confirmed the UGA(Stop → Trp) previously identified by (Turmel et al. 2010). These two reassignments are the most frequently observed genetic code modifications in mtDNA. AUA(Ile → Met) has been described in the mitochondrial genome of yeast, metazoa, and Xanthophyceae (Ehara et al. 1997; Yokobori et al. 2001; Sengupta et al. 2007). Likewise, UGA(Stop → Trp) was reported in many mitochondrial genomes and some nonmitochondrial systems (Massey and Garey 2007).

In contrast with its closest relatives (*Prototheca* and *Nephroselmis*) where all three Met-tRNA(CAU), fMet-tRNA(CAU) and Ile-tRNA(CAU) are found, only the fMet-tRNA(CAU) and a divergent tRNA(CAU), evidently acquired recently, are present in *Pycnococcus*. In support of the AUA(Ile → Met) transition, it can be argued that the ancestral mtDNA-encoded Ile-tRNA(CAU) was either lost or possibly transferred to the nucleus during the mitochondrial genome streamlining process in *Pycnococcus* (Turmel et al. 2010). While Turmel et al. (2010) suggested that AUA is decoded by a nucleus- or chloroplast-encoded Ile-tRNA, we conjecture instead that the codon is decoded conjointly with AUG as methionine by the mutant tRNA(CAU). Given the high AT-content of *Pycnococcus* mtDNA and the frequent usage of AUA codons in its relatives, it is somewhat unlikely that AUA vanished before the loss of Ile-tRNA(CAU). Therefore, the codon capture mechanism cannot explain this reassignment. Some have suggested that AUA(Ile → Met) is usually initiated by the loss of Ile-tRNA(CAU)

(Sengupta et al. 2007) and not by the disappearance of the codon. We favor this hypothesis for *Pycnococcus*. After the loss of Ile-tRNA(CAU), it is possible that translation of AUA codons was supported by Ile-tRNA(GAU), albeit inefficiently (Yokobori et al. 2001), before the mutant Met-tRNA(CAU) was acquired.

In contrast with AUA reassignment, UGA (Stop Trp) is entirely compatible with the codon capture mechanism. UGA codons are effectively missing from the mtDNA of *Nephroselmis*, *Ostreococcus*, *Prototheca*, where the AT-rich UAA stop codon is preferred. We suggest an evolutionary scenario in which UGA first disappears from the genome due to AT-pressure, followed by loss of the release factor's ability to recognize it. In parallel, a mutant Trp-tRNA(CCA) able to translate both UGA and UGG codons was acquired. Upon reappearance of UGA in the genome, the codon was then decoded as tryptophan. Reintroduction of UGA could have been facilitated by synonymous mutations of UGG to UGA, due to the sustained AT-pressure that caused UGA disappearance.

It remains unknown at which point during evolution the mutant tRNA(CAU) and tRNA(CCA) were acquired and by what mechanisms they decode noncognate codons. We suspect that both mutant tRNAs were simultaneously acquired from the same source. Evidence of gene acquisition by gene transfer has previously been reported in some Chlorophyta (Brouard et al. 2008) and the same could have occurred in *Pycnococcus*. In fact, principal component analysis and clustering of codon usage pattern in coding regions reveals a strong mitochondrial codon bias in *Pycnococcus*, not resembling any other chlorophyte mtDNAs. An alternative scenario would assume a posttranscriptional editing of the two tRNAs. Because, neither results reported by Turmel et al. (2010), nor our transcriptome analysis (see Supplementary Material online) support such hypothesis, only a modification similar to C: 34 to f^5C: 34 (5-formylcytidine) as observed in mtDNA-encoded Met-tRNA(CAU) of animals (Moriya et al. 1994; Tomita et al. 1999; Nakano et al. 2016) seems possible. Indeed, in *Pycnococcus* mtDNA, tRNA(CCA) and tRNA(CAU) are the only tRNAs with a cytidine at the wobble position of their anticodon, and therefore could both be modified by a posttranscriptional mechanism targeting the C: 34 nucleotide. Admittedly, these explanations remain highly speculative. To understand how these tRNAs are able to decode noncanonical codons, RNA sequencing, including all potential modifications, of mt-tRNAs and in vivo characterization of aaRS activities will be required.

Stop Codon Capture in Chlorophyta

Similar to other known stop codon reassignments, UAG reassignments in *N. aquatica*, Hydrodictyaceae, and Scenedesmaceae most likely follow a codon capture mechanism (Sengupta et al. 2007). Since stop codons are initially rare, their disappearance is conceivable. In particular, besides UAA, all standard stop codons, namely UGA and UAG, are missing in Sphaeropleales, with UCA and UCC acting as translation termination signals instead (Nedelcu et al. 2000; Fučíková, Lewis, Gonzalez-Halphen, et al. 2014; Farwagi

et al. 2015). With the independent emergence, via isoacceptor remodeling, of new tRNAs with anticodon CUA, the codon was fully captured at its reintroduction, by Leu-tRNA(CUA) in Scenedesmaceae and by Ala-tRNA(CUA) in Neochloris and Hydrodictyaceae.

UGA reassignment to tryptophan in Pedinomonas likely follows a similar mechanism, with a few modifications. Because the standard Trp-tRNA(CCA) is missing, it is likely that the new Trp-tRNA(UCA) can decode both UGA and UGG codons as tryptophan, in accordance with the "U:R wobble" rule (e.g., Martin et al. 1980; Sibler et al. 1980). However, UGG codons only account for 3% of Trp positions in Pedinomonas, contrasting with Pynococcus where both codons are used in equal frequency. Nevertheless, the high AT mutation pressure in the genome and the absence of UGA in its close relatives (Turmel et al. 1999) are clear indicators of a codon capture mechanism. Although disappearance of UGA codons was propitious for the reassignment, we believe that the critical factor here was the loss of several genes, including Trp-tRNA(CCA), during the severe mtDNA reduction in Pedinomonas. To compensate for this loss, and ensure continued translation of genes, chloroplast- or nucleus-encoded tRNAs could have been imported. Since tRNA import is not believed to be a dynamic process that could automatically adapt to new changes in the mitochondrial genome (Salinas et al. 2012), it is possible that a Trp-tRNA was not efficiently imported for translation of UGG codons, progressively driving them out of the genome. This hypothesis is supported by the sporadic usage in Pedinomonas of other amino acids at conserved Trp positions (supplementary fig. S1A, Supplementary Material online). The subsequent gain of a new Trp-tRNA(UCA), through tRNA remodeling of a Cys-tRNA(GCA) copy, would later allow decoding of UGA as tryptophan upon reintroduction of the codon in the genome.

Why Are Sense-to-Sense Codon Reassignments Only Located in Chlorophytes with an Intermediate Type of mtDNA?

An important characteristic of the predicted deviations of sense codon decoding in chlorophytes is that they are only located in mtDNAs thought to represent an intermediate stage, during the streamlining process toward a reduced mitochondrial genome (see fig. 4). The observed decoding disparity between Sphaeropleales and Chlamydomonadales, their sister group that has conserved the standard decoding, is astounding. It suggests a link between the "intermediate state" of mtDNA in Sphaeropleales and their genetic code evolution.

The mechanisms responsible for mtDNA streamlining in chlorophytes are still not very well understood (Nedelcu et al. 2000). It has been suggested that streamlining occurred because of competition among mtDNA molecules for faster replication time and that it promotes transfer of genes to the nucleus (Selosse et al. 2001). Recent genome analyses have revealed abundant evidence for transfer of mitochondrial genes into nuclear chromosomes, allowing them to escape

the higher rate of mildly deleterious mutations in the mitochondrial compartment (Lynch and Blanchard 1998; Martin and Herrmann 1998). Migration of mitochondrial genes to the nuclear genome was reported in Chlamydomonadales (Pérez-Martínez et al. 2001; Pérez-Martínez et al. 2000; Funes et al. 2002; Cardol et al. 2006) and to a certain extent in Tetradesmus (Pérez-Martínez et al. 2001; Adams and Palmer 2003). These observations have led to the current assumption that the mtDNA streamlining process started in the chlorophycean ancestor of Sphaeropleales and Chlamydomonadales, but was later halted, for some reason(s), in Sphaeropleales (Nedelcu et al. 2000; Rodríguez-Salinas et al. 2012b). In light of our results, we propose that genetic code deviations in the mtDNA of Sphaeropleales could have contributed to their unusual genome organization. Indeed, genes transferred to the nuclear genome are expected to be decoded by nucleus-encoded tRNAs, then imported back into the mitochondria where they will be functionally active. Therefore, decoding differences between the two compartments would result in nonfunctional protein products, which may be lethal. It is likely that the first genetic code alteration in the Sphaeropleales lineage was the reassignment of UCA (serine) codons to stop, as this reassignment appears to be shared by all Sphaeropleales (Fučíková, Lewis, Gonzalez-Halphen, et al. 2014). Mitochondrial gene migration to the nucleus would then have been forcibly stopped, due to the genetic code modification preventing correct translation of transferred genes to the nucleus. The resulting "intermediate-type mtDNA" with reduced proteome size and already altered tRNA repertoire (e.g., loss of the Arg-tRNA(UCU)) would have facilitated further changes in the genetic code, leading to the set of deviations currently observed. A similar scenario could also explain the predicted code alterations in Pycnococcus. However, analysis of additional chlorophyte mitochondrial genomes is required for a full understanding of the link between streamlining and genetic code alterations.

Concluding Remarks

The evolution of chlorophyte mtDNAs is characterized by the presence of multiple independently occurring codon reassignments in distinct lineages. Our results unveil a significant variation in their mitochondrial genetic code, even when comparing closely related species (fig. 4). Similar demonstrations of the genetic code evolving differently, and in parallel, in close genomes were previously reported in yeast (Ling et al. 2014) and metazoan (Abascal, Posada, et al. 2006) mitochondria. These results highlight why, despite the rarity of codon reassignments, it is crucial to determine the proper decoding pattern in fast-evolving genomes and not simply adopt a code used in close relatives.

Our study demonstrates an unexpected tolerance toward numerous codon deviations in chlorophyte mitochondrial genomes, hinting at the existence of a tremendous diversity of mtDNAs in green plants and eukaryotes in general. Although the present study provides new information for understanding mitochondrial genome evolution in chlorophytes, exploration of mtDNA evolution in additional

chlorophyte species, especially the early diverging members, will be interesting in order to gain a more general understanding of the early steps in codon evolution.

Materials and Methods

Outline of the Framework

We have developed a new framework for studying genetic code evolution, extending CoreTracker (Noutahi et al. 2017), a method for inferring codon reassignments. CoreTracker evaluates statistically significant differences between nucleotide sequences and expected amino acids in the derived protein sequences, taking conservation at each position into consideration. It differs from similar approaches (Abascal, Zardoya, et al. 2006; Dutilh et al. 2011; Mühlhausen and Kollmar 2014b) by the fact that it simultaneously handles a set of related genomes in a phylogenetic context and integrates a validation step ensuring high precision. This new approach was proven more accurate and more flexible than known alternatives, as it allows prediction of codon reassignments, without restriction to any specific phyla or genome type.

The extended framework consists of four complementary modules allowing for both predictions of genetic code alterations and inference of underlying evolutionary scenarios. We present a summary of the framework below. A more detailed discussion of each module, with an illustration on the UGA(Stop → Trp) codon reassignment in *P. minor*, is provided in section 1.1 of the Supplementary Material online.

In the first module, candidate codon reassignments are predicted for a set of phylogenetically related genomes, based on comparative sequence analysis of the aligned proteins versus corresponding nucleotide sequences. Since genetic code alteration is tightly linked to changes in tRNAs, we investigate the evolutionary history of tRNAs in a second step. The main objective is to identify tRNAs with anticodons capable of decoding candidate codons for reassignment and to infer their evolutionary history. Evidence of codon reassignment through tRNA identity switch includes: 1) histories of multiple tRNA duplications and losses, 2) structural tRNA remodeling, and 3) mutations in the anticodon loop or acceptor stem that are the main aminoacyl-tRNA synthetase recognition elements.

Although the complete set of mitochondrion-specific tRNA identity rules is largely unknown, major identity determinants and antideterminants in several bacterial tRNAs are well characterized (Saks et al. 1994; Giegé et al. 1998) and seemingly valid for mitochondria (Lang et al. 2012; Salinas-Giegé et al. 2015). The presence (or lack thereof) of these (anti)-determinants in the secondary structure of considered tRNAs is evaluated in a third step to infer their identity. Further verification by using computational methods specific to the problem such as TFAM (Ardell and Andersson 2006) is also performed. Finally, results of the three steps, together with information on gene order and ancestral tRNA gene content, are jointly used to infer the evolutionary history of predicted mitochondrial genetic code changes.

Mitochondrial Genome Data Set

The data set considered in this study involves 51 complete green plant mitochondrial genomes taken from NCBI (see supplementary table S1, Supplementary Material online), including 32 Chlorophyta, and among them, 14 recently sequenced Sphaeropleales mtDNAs. The mitochondrial protein-coding gene data set was constructed directly from the GenBank genome annotations, except for *Kirchneriella aperta* and the three Hydrodictyaceae, which required reannotation using MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl, last accessed February 18, 2019), followed by manual verification of completeness/correctness of the automated annotations. Translated coding sequences were obtained by considering already proposed stop codon reassignments. We further eliminated the disparity between genomic coding sequences and corresponding protein sequences in land plants, caused by C-to-U RNA editing. For this purpose, edited sites were obtained from the REDIdb RNA editing database (Picardi et al. 2007) and NCBI annotations.

Phylogenetic Species Tree

As accurate clade structure is important for correctly inferring the evolutionary history of code alterations, the phylogenetic species tree topology is based on the current view of green plant evolution (Soltis et al. 2011; Leliaert et al. 2012; Fučíková, Lewis, Gonzalez-Halphen, et al. 2014; Ruhfel et al. 2014; Fang et al. 2017). For comparison and discussion purposes, a Bayesian tree for Chlorophyta was also constructed from the concatenation of the 13 standard mtDNA-encoded proteins (Cob, Cox1–2–3, Atp6–9, and Nad1–2–3–4–4L-5–6) with PhyloBayes (Lartillot et al. 2009), while taking into account genetic code alterations. For PhyloBayes analyses, we removed constant sites with the -dc parameter and used the CAT-GTR model with six discrete categories and four independent chains. The chains were run for 10,000 cycles, with the first 7,000 used as burn-in, ensuring a maximum discrepancy of 0.05. We have also performed maximum likelihood phylogenetic analyses with RAxML v8.2.11 (Stamatakis 2014) under an unpartitioned scheme, using a LG+GAMMA model and 100 bootstraps.

Codon Reassignment Prediction

Prediction of sense-to-sense codon reassignment was performed with CoreTracker (Noutahi et al. 2017) using an HMM alignment refinement and the following parameters: −id 0.5 −ic 0.3 −gap 0.4, which correspond, respectively, to the minimum amino acid identity, the minimum information content and the maximum gap proportion accepted in each column of the alignment. The input submitted to CoreTracker is the phylogenetic species tree, the nucleotide sequences of conserved mitochondrial protein-coding genes, namely those of the respiratory chain complex (Cob, Cox1–2–3, Atp1–4–6–8–9, Nad1–2–3–4–4L-5–6–7–9, and Sdh3), as well as their corresponding translated amino acid sequences. Note that by using the option "−gap 0.4," any gene missing in >40% of the genomes will be automatically removed during analysis.

Stop-to-sense and sense-to-stop codon reassignments were obtained from genome annotations and confirmed by comparing the length of annotated proteins to their homologs, to detect missing C-terminal domains and stop codon read-through.

Transfer RNA Analysis

The 1,114 tRNAs from the 51 green plant mitochondrial genomes were annotated with RNAfinder (http://megasun. bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl; last accessed February 18, 2019) and confirmed with tRNAscan-SE (Schattner et al. 2005) using parameter -O for organellar tRNAs. Among the predicted tRNA genes, five contained intronic regions, which we removed before aligning the sequences. Multiple sequence alignment of tRNAs was done with LocARNA (Will et al. 2007) and an in-house script that uses the consensus secondary structure returned by RNAfinder as structural constraints. After manually editing the alignment to remove the hypervariable region, a maximum likelihood (ML) tree was inferred with FastTree v2.1.7 (Price et al. 2010) under the GTR+GAMMA model. Bootstrapping was performed for the ML tree with 1,000 replicates. Note that the phylogenetic trees constructed from tRNA sequences rely only on few informative positions and therefore do not necessarily exhibit sufficient and accurate resolution of branching order. Nevertheless, they provide correct tRNA grouping in most instances, which is the information required for our purpose of inferring the evolution of the genetic code. Inside the phylogenetic tree, tRNAs were consistently grouped into large and distinct clades of isoacceptors (tRNAs charging the same amino acid), with only a few exceptions. These clades often received high bootstrap, but the relationship between some isoacceptors within them could not be confidently determined with the corresponding branches having low bootstrap support (<0.2).

To predict the identity class of some tRNAs of interest, namely those with either a questionable identity or an evolutionary history compatible with codon reassignments, we used TFAM (Ardell and Andersson 2006). For this, we first compiled, from the chosen species, sets of mitochondrial tRNA isoacceptors with unambiguous identity for the following seven amino acids: methionine, leucine, tryptophan, alanine, cysteine, arginine, and tyrosine. These groups correspond to tRNA isoacceptors either sharing the same decoding with a tRNA that has a questionable identity or located in close vicinity inside the phylogenetic tree. Using TFAM, we first constructed a profile for each of the seven groups, then computed a score for every tRNA with a dubious identity against these profiles (see supplementary table S2, Supplementary Material online).

To confidently assess the phylogenetic placement of the tRNAs of interest, relative to the seven groups, while considering the eventual presence of ambiguous phylogenetic signals, we built a split network. The network was generated with SplitsTree v.4.14.6 (Huson and Bryant 2006) using the neighbor-net method. To reduce its complexity, sequence redundancy between tRNAs with trustworthy identity was removed using the CD-HIT suite (Li and Godzik 2006),

generating a data set of 199 representative sequences with <97% identity.

Finally, we built an alignment for each tRNA group and created a covariance model using cmbuild and cmcalibrate from the Infernal 1.1.1 package (Nawrocki and Eddy 2013). Transfer RNAs with dubious identity were aligned against the appropriate covariance models using cmsearch with the "–A" switch. For visualization and editing of sequence alignments, we used Jalview (Waterhouse et al. 2009) and Inkscape v0.48 (https://inkscape.org; last accessed February 18, 2019). Transfer RNA secondary structure diagrams were created using R2R (Weinberg and Breaker 2011) and further edited using Inkscape v0.48.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

Abascal F, Posada D, Knight RD, Zardoya R. 2006. Parallel evolution of the genetic code in arthropod mitochondrial genomes. *PLoS Biol.* 4(5):e127.

Abascal F, Zardoya R, Posada D. 2006. GenDecoder: genetic code prediction for metazoan mitochondria. *Nucleic Acids Res.* 34(Web Server):W389–W393.

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol.* 29(3):380–395.

Alfonzo JD, Blanc V, Estévez AM, Rubio MAT, Simpson L. 1999. C to U editing of the anticodon of imported mitochondrial tRNATrp allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J.* 18(24):7056–7062.

Andersson SG, Kurland CG. 1995. Genomic evolution drives the evolution of the translation system. *Biochem Cell Biol.* 73(11–12):775–787.

Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. *Trends Microbiol.* 6(7):263–268.

Ardell DH, Andersson SG. 2006. TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Res.* 34(3):893–904.

Asahara H, Himeno H, Tamura K, Hasegawa T, Watanabe K, Shimizu M. 1993. Recognition nucleotides of *Escherichia coli* tRNA-Leu and its elements facilitating discrimination from tRNA-Ser and tRNA-Tyr. *J Mol Biol.* 231(2):219–229.

Barrell B, Bankier A, Drouin J. 1979. A different genetic code in human mitochondria. *Nature* 282(5735):189.

Bilokapić S, Ban N, Weygand-Đurašević I. 2009. Seryl-tRNA Synthetases: enzymes with Multiple Personalities. *Croat Chem Acta.* 82:493–501.

Boer PH, Gray MW. 1988. Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55(3):399–411.

Brouard JS, Otis C, Lemieux C, Turmel M. 2008. Chloroplast DNA sequence of the green alga *Oedogonium cardiacum* (Chlorophyceae): unique genome architecture, derived characters shared with the Chaetophorales and novel genes acquired through horizontal transfer. *BMC Genomics* 9:290.

Buchheim MA, Lemieux C, Otis C, Gutell RR, Chapman RL, Turmel M. 1996. Phylogeny of the Chlamydomonadales (Chlorophyceae): a comparison of ribosomal RNA gene sequences from the nucleus and the chloroplast. *Mol Phylogenet Evol.* 5(2):391–402.

Burger G, Nedelcu AM. 2012. Mitochondrial genomes of algae. In: Bock, Ralph, Knoop, Volker (eds,) Genomics of chloroplasts and mitochondria. The Netherlands: Springer. p. 127–157.

Cardol P, Lapaille M, Minet P, Franck F, Matagne RF, Remacle C. 2006. ND3 and ND4L subunits of mitochondrial complex I, both nucleus encoded in *Chlamydomonas reinhardtii*, are required for activity and assembly of the enzyme. *Eukaryot Cell.* 5(9):1460–1467.

Cochella L, Green R. 2005. An active role for tRNA in decoding beyond codon: anticodon pairing. *Science* 308(5725):1178–1180.

Cox CJ, Li B, Foster PG, Embley TM, Civán P. 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Syst Biol.* 63(2):272–279.

Denovan-Wright EM, Nedelcu AM, Lee RW. 1998. Complete sequence of the mitochondrial DNA of *Chlamydomonas eugametos*. *Plant Mol Biol.* 36(2):285–295.

Dutilh BE, Jurgelenaite R, Szklarczyk R, van Hijum SAFT, Harhangi HR, Schmid M, de Wild B, Françoijs K-J, Stunnenberg HG, Strous M, et al. 2011. FACIL: fast and accurate genetic code inference and logo. *Bioinformatics* 27(14):1929–1933.

Ehara M, Hayashi-Ishimaru Y, Inagaki Y, Ohama T. 1997. Use of a deviant mitochondrial genetic code in yellow-green algae as a landmark for segregating members within the phylum. *J Mol Evol.* 45(2):119–124.

Fang L, Leliaert F, Zhang ZH, Penny D, Zhong BJ. 2017. Evolution of the Chlorophyta: insights from chloroplast phylogenomic analyses. *J Syst Evol.* 55(4):322–332.

Farwagi AA, Fučíková K, McManus HA. 2015. Phylogenetic patterns of gene rearrangements in four mitochondrial genomes from the green algal family Hydrodictyaceae (Sphaeropleales, Chlorophyceae). *BMC Genomics* 16:826.

Fox TD. 1979. Five TGA 'stop' codons occur within the translated sequence of the yeast mitochondrial gene for cytochrome c oxidase subunit II. *Proc Natl Acad Sci U S A.* 76(12):6534–6538.

Freel KC, Friedrich A, Schacherer J. 2015. Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res.* 15(4):fov023.

Fučíková K, Flechtner VR, Lewis LA. 2013. Revision of the genus *Bracteacoccus* Tereg (Chlorophyceae, Chlorophyta) based on a phylogenetic approach. *Nova Hedwigia* 96:15–59.

Fučíková K, Lewis LA. 2012. Intersection of *Chlorella*, *Muriella* and *Bracteacoccus*: resurrecting the genus *Chromochloris* KOL et CHODAT (Chlorophyceae, Chlorophyta). *Fottea* 12:83–93.

Fučíková K, Lewis PO, Gonzalez-Halphen D, Lewis LA. 2014. Gene arrangement convergence, diverse intron content, and genetic code modifications in mitochondrial genomes of Sphaeropleales (Chlorophyta). *Genome Biol Evol.* 6(8):2170–2180.

Fučíková K, Lewis PO, Lewis LA. 2014. Putting *incertae sedis* taxa in their place: a proposal for ten new families and three new genera in Sphaeropleales (Chlorophyceae, Chlorophyta). *J Phycol.* 50(1):14–25.

Fučíková K, Lewis PO, Lewis LA. 2016. Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution. *Mol Phylogenet Evol.* 98:176–183.

Funes S, Davidson E, Claros MG, Van Lis R, Pérez-Martínez X, Vázquez-Acevedo M, King MP, González-Halphen D. 2002. The typically mitochondrial DNA-encoded ATP6 subunit of the F1F0-ATPase is encoded by a nuclear gene in *Chlamydomonas reinhardtii*. *J Biol Chem.* 277(8):6051–6058.

Giegé R, Eriani G. 2014. Transfer RNA Recognition and Aminoacylation by Synthetases. In eLS, John Wiley & Sons, Ltd (Ed.). doi:10.1002/9780470015902.a0000531.pub3.

Giegé R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* 26(22):5017–5035.

Glover KE, Spencer DF, Gray MW. 2001. Identification and structural characterization of nucleus-encoded transfer RNAs imported into wheat mitochondria. *J Biol Chem.* 276(1):639–648.

Hamann CS, Hou YM. 1995. Enzymic aminoacylation of tRNA acceptor stem helixes with cysteine is dependent on a single nucleotide. *Biochemistry* 34(19):6527–6532.

Hayashi-Ishimaru Y, Ohama T, Kawatsu Y, Nakamura K, Osawa S. 1996. UAG is a sense codon in several chlorophycean mitochondria. *Curr Genet.* 30(1):29–33.

Himeno H, Hasegawa T, Asahara H, Tamura K, Shimizu M. 1991. Identity determinants of *E. coli* tryptophan tRNA. *Nucleic Acids Res.* 19(23):6379–6382.

Hirsh D. 1971. Tryptophan transfer RNA as the UGA suppressor. *J Mol Biol.* 58(2):439–458.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23(2):254–267.

Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet.* 2(1):49–58.

Kollmar M, Mühlhausen S. 2017. Nuclear codon reassignments in the genomics era and mechanisms behind their evolution. *BioEssays* 39(5):1600221.

Krassowski T, Coughlan AY, Shen X-X, Zhou X, Kominek J, Opulente DA, Riley R, Grigoriev IV, Maheshwari N, Shields DC, et al. 2018. Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. *Nat Commun.* 9(1):1887.

Krienitz L, Bock C, Dadheech P, Pröschold T. 2011. Taxonomic reassessment of the genus Mychonastes (Chlorophyceae, Chlorophyta) including the description of eight new species. *Phycologia* 50(1):89–106.

Kroymann J, Zetsche K. 1998. The mitochondrial genome of *Chlorogonium elongatum* inferred from the complete sequence. *J Mol Evol.* 47(4):431–440.

Kück U, Jekosch K, Holzamer P. 2000. DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a nonsense codon. *Gene* 253(1):13–18.

Lang BF, Lavrov D, Beck N, Steinberg SV. 2012. Mitochondrial tRNA structure, identity, and evolution of the genetic code. In: Bullerwell, Charles E. (eds) Organelle Genetics: Evolution of Organelle Genomes and Gene Expression. Berlin, Heidelberg: Springer. p. 431–474.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol.* 7(Suppl 1):S4.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.

Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De Clerck O. 2012. Phylogeny and molecular evolution of the green algae. *Crit Rev Plant Sci.* 31(1):1–46.

Leliaert F, Tronholm A, Lemieux C, Turmel M, DePriest MS, Bhattacharya D, Karol KG, Fredericq S, Zechman FW, Lopez-Bautista JM. 2016. Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the chlorophyta, palmophyllophyceae class. nov. *Sci Rep.* 6:25367.

Lemieux C, Otis C, Turmel M. 2014a. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol Biol.* 14:211.

Lemieux C, Otis C, Turmel M. 2014b. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics* 15(1):857.

Lewis LA, McCourt RM. 2004. Green algae and the origin of land plants. *Am J Bot.* 91(10):1535–1556.

Li M, Tzagoloff A. 1979. Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene. *Cell* 18(1):47–53.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.

Ling J, Daoud R, Lajoie MJ, Church GM, Söll D, Lang BF. 2014. Natural reassignment of CUU and CUA sense codons to alanine in *Ashbya* mitochondria. *Nucleic Acids Res.* 42(1):499–508.

Lynch M, Blanchard JL. 1998. Deleterious mutation accumulation in organelle genomes. In: Woodruff, Ronny C. and Thompson, James N. (eds) Mutation and evolution. The Netherlands: Springer, p. 29–39.

Mallick B, Chakrabarti J, Sahoo S, Ghosh Z, Das S. 2005. Identity elements of archaeal tRNA. *DNA Res.* 12(4):235–246.

Marechal-Drouard L, Small I, Weil JH, Dietrich A. 1995. Transfer RNA import into plant mitochondria. *Meth. Enzymol.* 260:310–327.

Martin NC, Pham HD, Underbrink-Lyon K, Miller DL, Donelson JE. 1980. Yeast mitochondrial tRNATrp can recognize the nonsense codon UGA. *Nature* 285(5766):579.

Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* 118(1):9–17.

Massey SE, Garey JR. 2007. A comparative genomics analysis of codon reassignments reveals a link with mitochondrial proteome size and a mechanism of genetic code change via suppressor tRNAs. *J Mol Evol.* 64(4):399–410.

McClain WH, Foss K. 1988. Changing the identity of a tRNA by introducing a GU wobble pair near the 3′ acceptor end. *Science* 240(4853):793–796.

Moriya J, Yokogawa T, Wakita K, Ueda T, Nishikawa K, Crain PF, Hashizume T, Pomerantz SC, McCloskey JA, Kawai G. 1994. A novel modified nucleoside found at the first position of the anticodon of methionine tRNA from bovine liver mitochondria. *Biochemistry* 33(8):2234–2239.

Mühlhausen S, Findeisen P, Plessmann U, Urlaub H, Kollmar M. 2016. A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Res.* 26(7):945–955.

Mühlhausen S, Kollmar M. 2014a. Molecular phylogeny of sequenced Saccharomycetes reveals polyphyly of the alternative yeast codon usage. *Genome Biol Evol.* 6(12):3222–3237.

Mühlhausen S, Kollmar M. 2014b. Predicting the fungal CUG codon translation with Bagheera. *BMC Genomics* 15:411.

Muramatsu T, Yokoyama S, Horie N, Matsuda A, Ueda T, Yamaizumi Z, Kuchino Y, Nishimura S, Miyazawa T. 1988. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. *J Biol Chem.* 263(19):9261–9267.

Musier-Forsyth K, Usman N, Scaringe S, Doudna J, Green R, Schimmel P. 1991. Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. *Science* 253(5021):784–786.

Nakano S, Suzuki T, Kawarada L, Iwata H, Asano K, Suzuki T. 2016. NSun3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNA Met. *Nat Chem Biol.* 12(7):546.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29(22):2933–2935.

Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res.* 10(6):819–831.

Noutahi E, Calderon V, Blanchette M, Lang BF, El-Mabrouk N. 2017. CoreTracker: accurate codon reassignment prediction, applied to mitochondrial genomes. *Bioinformatics* 33(21):3331–3339.

Osawa S, Jukes T, Watanabe K, Muto A. 1992. Recent evidence for evolution of the genetic code. *Microbiol Rev.* 56(1):229–264.

Osawa S, Jukes TH. 1989. Codon reassignment (codon capture) in evolution. *J Mol Evol.* 28(4):271–278.

Pak M, Pallanck L, Schulman LH. 1992. Conversion of a methionine initiator tRNA into a tryptophan-inserting elongator tRNA *in vivo*. *Biochemistry* 31(13):3303–3309.

Pallanck L, Li S, Schulman L. 1992. The anticodon and discriminator base are major determinants of cysteine tRNA identity *in vivo*. *J Biol Chem.* 267(11):7221–7223.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci U S A.* 97(13):6960–6966.

Perez-Martínez X, Antaramian A, Vázquez-Acevedo M, Funes S, Tolkunova E, d'Alayer J, Claros MG, Davidson E, King MP, González-Halphen D. 2001. Subunit II of cytochrome c oxidase in Chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem.* 276(14):11302–11309.

Pérez-Martínez X, Vázquez-Acevedo M, Tolkunova E, Funes S, Claros MG, Davidson E, King MP, González-Halphen D. 2000. Unusual location of a mitochondrial gene subunit IIi of cytochrome c oxidase is encoded in the nucleus of chlamydomonad algae. *J Biol Chem.* 275(39):30144–30152.

Picardi E, Regina TMR, Brennicke A, Quagliariello C. 2007. REDIdb: the RNA editing database. *Nucleic Acids Res.* 35(Database issue):D173–D177.

Pombert JF, Otis C, Lemieux C, Turmel M. 2004. The complete mitochondrial DNA sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) highlights distinctive evolutionary trends in the Chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae. *Mol Biol Evol.* 21(5):922–935.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.

Rodríguez-Salinas E, Remacle C, González-Halphen D. 2012a. Green Algae Genomics: a Mitochondrial Perspective. In: L. Maréchal-Drouard (Ed.), Mitochondrial Genome Evolution, Advances in botanical research. Vol. 63. San Diego: Elsevier, p. 187–214.

Rodríguez-Salinas E, Riveros-Rosas H, Li Z, Fučíková K, Brand JJ, Lewis LA, González-Halphen D. 2012b. Lineage-specific fragmentation and nuclear relocation of the mitochondrial cox2 gene in chlorophycean green algae (Chlorophyta). *Mol Phylogenet Evol.* 64(1):166–176.

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms – inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol Biol.* 14:23.

Saks ME, Sampson JR, Abelson JN. 1994. The transfer RNA identity problem: a search for rules. *Science* 263(5144):191–197.

Salinas T, Duby F, Larosa V, Coosemans N, Bonnefoy N, Motte P, Maréchal-Drouard L, Remacle C. 2012. Co-evolution of mitochondrial tRNA import and codon usage determines translational efficiency in the green alga *Chlamydomonas*. *PLoS Genet.* 8(9):e1002946.

Salinas-Giegé T, Giegé R, Giegé P. 2015. tRNA biology in mitochondria. *Int J Mol Sci.* 16(3):4518–4559.

Satjarak A, Burns JA, Kim E, Graham LE. 2017. Complete mitochondrial genomes of prasinophyte algae *Pyramimonas parkeae* and Cymbomonas tetramitiformis. *J Phycol.* 53(3):601–615.

Schattner P, Brooks AN, Lowe TM. 2005. The trnascan-se, snoscan and snogps web servers for the detection of tRNAs and snornas. *Nucleic Acids Res.* 33(Web Server issue):W686–W689.

Schmeing TM, Voorhees RM, Kelley AC, Ramakrishnan V. 2011. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol.* 18(4):432.

Schultz DW, Yarus M. 1994. Transfer RNA mutation and the malleability of the genetic code. *J Mol Biol.* 235(5):1377–1380.

Schultz DW, Yarus M. 1996. On malleability in the genetic code. *J Mol Evol.* 42(5):597–601.

Selosse MA, Albert B, Godelle B. 2001. Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends Ecol Evol.* 16(3):135–141.

Sengupta S, Yang X, Higgs PG. 2007. The mechanisms of codon reassignments in mitochondrial genetic codes. *J Mol Evol.* 64(6):662–688.

Sibler A, Bordonne R, Dirheimer G, Martin R. 1980. Primary structure of yeast mitochondrial tryptophan-tRNA capable of translating the termination UGA codon. *C R Seances Acad Sci D.* 290:695–698.

Sibler AP, Dirheimer G, Martin RP. 1981. Nucleotide sequence of a yeast mitochondrial threonine-tRNA able to decode the CUN leucine codons. *FEBS Lett.* 132(2):344–348.

Smith DR, Lee RW. 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. *Mol Biol Evol.* 25(3):487–496.

Sohm B, Frugier M, Brule H, Olszak K, Przykorska A, Florentz C. 2003. Towards understanding human mitochondrial leucine aminoacylation identity. *J Mol Biol.* 328(5):995–1010.

Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot.* 98(4):704–730.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Su D, Lieberman A, Lang BF, Simonović M, Söll D, Ling J. 2011. An unusual tRNAThr derived from tRNAHis reassigns in yeast mitochondria the CUN codons to threonine. *Nucleic Acids Res.* 39(11):4866–4874.

Sun L, Fang L, Zhang Z, Chang X, Penny D, Zhong B. 2016. Chloroplast phylogenomic inference of green algae relationships. *Sci Rep.* 6:20528.

Tippery NP, Fučíková K, Lewis PO, Lewis LA. 2012. Probing the monophyly of the Sphaeropleales (Chlorophyceae) using data from five genes. *J Phycol.* 48(6):1482–1493.

Tomita K, Ueda T, Ishiwa S, Crain PF, McCloskey JA, Watanabe K. 1999. Codon reading patterns in *Drosophila melanogaster* mitochondria based on their tRNA sequences: a unique wobble rule in animal mitochondria. *Nucleic Acids Res.* 27(21):4291–4297.

Tukalo MA, Yaremchuk G, Kovalenko OP, Kriklivyi IA, Gudzera O. 2013. Recognition of tRNAs with a long variable arm by aminoacyl-tRNA synthetases. *Biopolym Cell.* 29(4):311–323.

Turmel M, Lemieux C, Burger G, Lang BF, Otis C, Plante I, Gray MW. 1999. The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: two radically different evolutionary patterns within green algae. *Plant Cell* 11(9):1717–1729.

Turmel M, Otis C, Lemieux C. 2010. A deviant genetic code in the reduced mitochondrial genome of the picoplanktonic green alga *Pycnococcus provasolii*. *J Mol Evol.* 70(2):203–214.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191.

Weber F, Dietrich A, Weil JH, Maréchal-Drouard L. 1990. A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. *Nucleic Acids Res.* 18(17):5027–5030.

Weinberg Z, Breaker RR. 2011. R2R-software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* 12:3.

Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R. 2007. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol.* 3(4):e65.

Wolff G, Plante I, Lang BF, Kück U, Burger G. 1994. Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*: gene content and genome organization. *J Mol Biol.* 237(1):75–86.

Yokobori S, Suzuki T, Watanabe K. 2001. Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. *J Mol Evol.* 53(4–5):314–326.