COMMENTARY

YCF1: A Green TIC: Response to the de Vries et al. Commentary

Masato Nakai
Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

This response to a recent Commentary article by de Vries et al. highlights critical errors in the annotation and identification of Ycf1 homologs in the sequenced chloroplast genomes. Contrary to what is reported by de Vries et al., the majority of chloroplast genomes sequenced to date appear to have retained a typical Ycf1 sequence (i.e., including the N-terminal 6TM domain and a variable hydrophilic C-terminal domain) as my group previously reported. Our evidence continues to support the model that Ycf1 forms an essential component of a “green TIC” that is largely conserved among the Chlorophyta and land plants. Since the establishment of this green TIC with Tic20 as the core component, some cases of loss of Ycf1 during the evolution of the green lineages might be regarded as modifications or alterations of the complex. Here, I discuss our working model that the presence of an alternative “nonphotosynthetic-type” or “ancestral-type” TIC might explain other (or specific) cases of the lack of Ycf1, not only in early lineages, including Glaucophyta and Rhodophyta, but also in the grasses.

Virtually all chloroplasts/plastids in today’s photosynthetic, plastid-containing eukaryotes derive from one successful primary endosymbiotic event with a cyanobacterium-like ancestor. During evolution, massive transfer of genes from the endosymbiont to the host’s nuclear genome occurred concomitant with the establishment of a protein transport system that allows these nucleus-encoded proteins back into the endosymbiotic organelle. Two successive protein translocons at the outer and inner envelope membranes of chloroplast, termed TOC and TIC, respectively, are responsible for this protein transport. My group recently identified a novel TIC complex consisting of Tic20, Tic56, Tic100, and Tic214 in Arabidopsis thaliana (Kikuchi et al., 2013). We found that Tic214 is encoded by the previously enigmatic essential chloroplast gene ycf1 (Boudreau et al., 1997; Drescher et al., 2000). This discovery led us to an extensively revised view of the molecular mechanisms of chloroplast protein import (Nakai, 2015). We found this TIC complex to be absolutely required for photosynthetic protein import in green tissues and, thus, essential for plant viability in Arabidopsis (Hirabayashi et al., 2011; Kikuchi et al., 2013). Experimental evidence for the presence of a similar TIC complex in pea (Pisum sativum) (Kikuchi et al., 2009) supports the idea that this complex is a general feature of the TIC apparatus. The TIC complex and Ycf1 appear to have evolved dramatically along with the evolution of the green lineage, including green algae and land plants; thus, they can be regarded as a “general green TIC.”

In their commentary, de Vries et al. (2015) repeated a phylogenetic analysis of Ycf1 and asked why 30 years of TOC/TIC research has missed this green TIC complex. Our response is that many things have been missed over the last 30 years. Several “old” Tic proteins, such as Tic40 and Tic110, were mistakenly identified very early (probably because of their high abundance in the inner envelope membrane and/or of the use of inadequate methodology) and have long been believed to be true translocon components without further concrete evidence for their direct involvement in protein transport ( Bölter et al., 2015; Paila et al., 2015). It has become clear that it is time to reevaluate the functions of these historic Tic proteins as well as of the so-called redox regulators, namely, Tic32, Tic55, and Tic62 (Nakai, 2015). To address this, we developed a highly specific method to purify and identify functional TIC complex components (Supplemental Figure 1; Kikuchi et al., 2013; Nakai, 2015). For now, until the historic TIC components are demonstrated to function directly in protein transport, we have removed them from our working model (Figure 1).

While we need to recognize the limitations of predicting the essentiality of a gene/protein based solely on its presence or absence in available genome sequence and annotation databases, de Vries et al. (2015) raise several important questions about Ycf1 function and the evolution of the TIC system. Principal among these are: (1) What is the true evolutionary history of Ycf1, and (2) how can the absences of this “essential” function be explained in the grasses and certain other lineages? Here, we highlight what we believe to be critical errors presented by de Vries et al. (2015) in the annotation and identification of Ycf1 homologs in the sequenced chloroplast genomes and summarize hypotheses that offer reasonable answers to both of these questions.

DISCREPANCIES AND ERRORS IN ANNOTATION AND INTERPRETATION OF YCF1 SEQUENCES

Decades of chloroplast genome research have revealed that chloroplast-encoded genes are frequently divided into exons and introns and, in some cases, may contain a RNA (transcript) editing site or a trans-splicing site where two different transcripts derived from two distinct regions of a chloroplast genome are spliced and fused to form a full-length mRNA (Glanz and Kücek, 2009; Stern et al., 2010; Wicke et al., 2011; Takenaka et al., 2013). Ribosomal frameshifting also occurs in chloroplasts (Kohl and Bock, 2009). In addition, incorrect annotation or misannotation of coding sequences is common in the genome sequences deposited in the public databases (Schnoes et al., 2009). Moreover, recent high-throughput but short-read sequencing technologies
Substrate specificities and redundancies of distinct TOC and TIC complexes still remain to be further elucidated. See text for details. (Adapted from Nakai [2015], Figure 3, with permission from Elsevier.)

Although de Vries et al. (2015) performed a tBLASTn search (as shown in their Supplemental Data Set 2 and Supplemental Figure 1), they apparently disregarded the resulting data to arrive at their main conclusion; Figures 1 and 3 in de Vries et al. (2015) were prepared solely based on their initial data obtained from simple HMM homology searches performed only against publically annotated protein sequences. The reasons provided were that the “majority of additionally detected sequences are very fragmented and likely do not stem from functional proteins” and that they “do not wish to speculate on their possible functionality or correct annotation, and, more importantly, the difference is irrelevant for our conclusion.” However, I believe that this omission led to potentially serious errors in the subsequent analysis. For example, their initial HMM searches failed to identify ycf1 homologs in Ginkgo biloba, Erycina pusilla, and Lactuca sativa; yet these three genomes clearly contain continuous open reading frames (i.e., containing no introns) encoding Ycf1 proteins of 1749, 1784, and 1726 amino acids, respectively, each carrying typical N-terminal 6TMs (as shown in the tBLASTn data in Supplemental Data Set 2 of de Vries et al. [2015]). These findings prompted me to survey their reported tBLASTn data in further detail, and I found additional errors as described below.

In their Figure 3 (sequence ii; de Vries et al., 2015), the presence of the N-terminal 6TMs in the Ycf1 of Picea abies is questioned, and this appears to be supported by their tBLASTn search, which likewise failed to identify the N-terminal 6TM region (Supplemental Data Set 2 in de Vries et al., 2015). However, conducting a simple tBLASTn search via the NCBI site, I was easily able to identify a continuous reading frame in the 5’ upstream region of the annotated P. abies ycf1 gene that encodes the typical 6TM segments (as shown in Supplemental Data Set 1 of this article). de Vries et al. (2015) wrote that “A region encoding TMs is present upstream of the annotated genes... However, the largest variation is found within the region downstream of the N-terminal TM domains” and further emphasized that “the spruces Picea morsomicola and Picea abies are interesting. Each spruce encodes a single YCF1 of ~1800 amino acids, but the N-terminal regions encoding the TM domains are missing.” However, as shown in my Supplemental Data Set 1, both P. morsomicola and P. abies in fact contain in-frame reading frames encoding this important 6TM domain.

I performed tBLASTn searches one by one on 33 chloroplast genomes (not including the Poaceae, to be discussed below) included in Supplemental Data Set 2 of de Vries et al. (2015), in which no typical Ycf1 had been identified. As shown in my Supplemental Data Set 1, four of these 33 chloroplast genomes appear to be not completely sequenced or contain uncertain sequence in the TM regions (as shown in my Supplemental Data Set 1, four of these 33 chloroplast genomes appear to be not completely sequenced or contain uncertain sequence in the Ycf1-coding region, while one of them, Picea sitchensis most likely does encode a typical Ycf1, like P. abies. Among the remaining 29 chloroplast genomes, 15 genomes surprisingly encode a typical Ycf1 if U-to-C RNA editing occurs at one nucleotide position (Takenaka et al., 2013). Three other genomes seem to encode relatively shorter Ycf1 proteins containing only a 6TM domain. However, based on the presence of downstream reading frames encoding the C-terminal domain in each, it is

![Figure 1. Revised Working Model for the Coordinated Function of the Photosynthetic-Type Major TIC Complex and the Alternative Nonphotosynthetic-Type TIC Complex in Substrate-Specific or Tissue-Specific Protein Import in Concert with Different Types of TOC Complexes in Most Land Plants.](image-url)
possible that sequencing/assembling errors underlie the apparent reduction in length of these proteins. Thus, only 10 organisms of this list of 33 (not including Poaceae) potentially lack Ycf1, among which five organisms belong to related Geraniaeae and three to Asteraceae. It should be noted that ycf1 regions of 5 of those 10 organisms might have been misassembled because of their locations around the inverted repeat regions on the chloroplast genomes. Therefore, of the chloroplast genomes that have been determined so far, besides members of the Poaceae, the majority of them appear to have retained a typical Ycf1 (i.e., including the N-terminal 6TM domain and a variable hydrophilic C-terminal domain) as we previously described (Kikuchi et al., 2013).

In addition to the obscure depiction of the N-terminal 6TMs of *P. abies* Ycf1 as mentioned above, Figure 3 presented by de Vries et al. (2015) contains other misleading information. Their tBLASTn search identified only a short N-terminal part of *Brassaiopsis hainla* Ycf1 (as shown in their Supplemental Data Set 2 and Figure 3, iii, and also as emphasized in their text); however, it contains a fairly long C-terminal conserved domain that was clearly depicted in the original article on the determination of chloroplast genome sequences of five related species in the Araliaceae (Li et al., 2013). Moreover, although *Marchantia* Ycf1 and *Schizomeris* Ycf1 were depicted to contain only four TMs (de Vries et al., 2015; Figure 3, iv and vii), this is based on the use of a single TMHMM program to predict TM segments; other programs, such as Sosui or Phobius, predict the more typical 6TMs (Hirokawa et al., 1998; Käll et al., 2004). In addition, de Vries et al. discuss a very short form of *Arabidopsis* ATCG01000, which occurs because the corresponding region is located on one of the two inverted repeats (Sato et al., 1999). This kind of truncation frequently occurs in chloroplast genomes, not only for ycf1 genes, but also for other essential genes encoding, e.g., the ribosomal protein Rps19 (Li et al., 2013) or the photosystem component PsbA (Lin et al., 2012). In most cases, an intact counterpart is retained near the border of the other inverted repeat region.

HOW CAN WE EXPLAIN THE EVOLUTIONARY HISTORY OF YCF1, INCLUDING ITS DISAPPEARANCE IN THE GRASSES?

The TIC system has evolved dramatically since the initial endosymbiotic event, through modification of both the nuclear and the chloroplast genomes (as also described in Kikuchi et al., 2013; Löffelhardt, 2014; Sommer and Schleiff, 2014). Tic20 orthologs are well conserved among virtually all plastid-containing lineages, likely reflecting a central role for it in protein translocation (Töpel and Jarvis, 2011). By contrast, some lineages, namely, Glaucophyta and Rhodophyta, encode neither Tic56 nor Tic100, both of which are indispensable in Arabidopsis and seemingly also lack the chloroplast-encoded Tic214/Ycf1. Chlorophyta encode a shorter version of Tic56 in the nuclear genome and Ycf1 in the chloroplast genome, but apparently lack the Tic100 component. Because Ycf1 (as well as Tic56 and Tic100) can be regarded as a kind of appended component to a preexisting TIC complex carrying Tic20 as a central core, it is entirely possible that modifications might have occurred in other components to compensate for the lack of Ycf1 in some species of green lineages. It is not unreasonable to suppose that the loss of ycf1 and the development of modifications to compensate for it have occurred multiple times in evolutionary history, as reported for a variety of other biological processes. For example, C4 photosynthesis has evolved independently up to 66 times within the angiosperms (Sage et al., 2012; Schulze et al., 2013).

Compared with these anticipated modifications that might have occurred in the green TIC for the above-mentioned species lacking YCF1, the situation of grasses seems more complicated. Grass species for which complete genomic sequences are available have Tic20 orthologs but lack Tic56, Tic100, or YCF1 (Tic214) in their nuclear or chloroplast genomes (Kikuchi et al., 2013). A possible explanation for this enigmatic evolutionary history of Ycf1 and the TIC transport system was proposed by Kikuchi et al. (2013) and further described in a more recent review article (Nakai, 2015). As illustrated in Figure 1, in addition to the major photosynthetic-type TIC (i.e., the green TIC containing Ycf1 that predominantly functions in green tissues), there appears to be an alternative nonphotosynthetic-type minor TIC import system for a certain subset of proteins. This system includes the Tic20 paralog Tic20-IV in Arabidopsis but lacks Tic56, Tic100, and Tic214 (Hirabayashi et al., 2011; Kikuchi et al., 2013; Parker et al., 2014; Nakai, 2015). Experimental data shown in recent work by Köhler et al. (2015) provide support for this model of more than one TIC complex. They found that the major TIC complex containing Tic56 shows very strong physical interaction with the major TOC complex containing the Toc159 receptor. In the absence of Tic56 (i.e., the absence of the major TIC complex), plants exhibit severe albinism consistent with being unable to import most photosynthetic proteins into the chloroplasts. It should be noted, however, that the plants still can import various nonphotosynthetic or housekeeping proteins into the plastids (Hirabayashi et al., 2011; Kikuchi et al., 2013), suggesting that an alternative TIC complex supports their import.

Importantly, such an alternative TIC system probably has a direct evolutionary relationship with those functioning in the Poaceae as well as in the Glaucophyta and Rhodophyta, which may have retained a simpler ancestral-type TIC system (Figure 1; Kikuchi et al., 2013; Nakai, 2015). This hypothesis was first presented by Kikuchi et al. (2013). To address these issues, we are currently working on the identification of all components of the grass-type TIC complex as well as of the red algal TIC complex and confirming their involvement in protein translocation across the inner envelope membrane of respective chloroplasts.

An important question then arises: Why might grasses have adopted an alternative nonphotosynthetic-type TIC system in place of the major TIC system found in other land plants? While there are various possibilities, including an occurrence of crucial modifications of this alternative system, the answer might be related to the characteristic chloroplast biogenesis of grasses, recently highlighted in an excellent review article by Pogson et al. (2015). Compared with most other plants, including dicots whose apical meristems are generally located at the shoot
tip, in grasses a unique shoot meristem called the intercalary meristem is located at the base of the leaf blade just above the ground level or near the node. Grass leaves extend in length through formation of new cells at the base of each leaf blade with subsequent cell expansion. Thus, chloroplast development from the nonphotosynthetic proplastids to photosynthetic chloroplasts can be observed as a gradient along the leaf blade. Given such continuous chloroplast biogenesis in early grasses, there might have been some advantage in utilizing a nonphotosynthetic-type TIC system throughout leaf development with certain modifications.

The elucidation of an alternative TIC system may offer insight into the origin of the ycf1 gene, which has long been believed to have appeared suddenly in the chloroplast genomes of Chlorophyta (Wicke et al., 2011). It can be reasonably hypothesized that a simpler ancestral-type TIC system, retained in Glaucophyta as well as Rhodophyta, also served as a prototype of the major photosynthetic-type green TIC complex containing Ycf1. I hypothesize that Ycf1 may have arisen from the chloroplast genome-encoded Tic20, at least in part. Although they lack ycf1, the known chloroplast genomes of Rhodophyta do encode a conserved Tic20 homolog.

Generally speaking, Ycf1 consists of the N-terminal 6TM domain and the hydrophilic C-terminal domain, the former of which may contribute to the formation of a protein-conducting channel together with Tic20, which itself contains four transmembrane helices. The amino acid sequences of both domains of Ycf1 appear not to be strictly conserved (Kikuchi et al., 2013). It is possible that the most important evolutionary constraint on Ycf1 was to avoid specific strong interactions with any segment of the thousands of different incoming translocating preproteins. The major photosynthetic-type green TIC changed markedly with the emergence of Chlorophyta and land plants, including the addition of two novel nuclear-encoded components, Tic56 and Tic100, as well as the rapid size expansion of Ycf1. The evolutionary history of Ycf1 and the TIC system might be related to high demand for efficient transport of particular subsets of proteins across the inner envelope membrane of chloroplasts, a demand that might have begun when “green” Chlorophyta emerged.

Supplemental Data

Supplemental Figure 1. Updated method to purify and identify TIC complex using translocation intermediates.

Supplemental Data Set 1. One-by-one tBLASTn results for reinvestigation of 33 plastid genomes for which de Vries et al. (2015) had been unable to detect a canonical ycf1 gene.

ACKNOWLEDGMENTS

I thank members of the Nakai laboratory, especially Shingo Kikuchi, Yukari Asakura, and Yoshiko Kotani, for research assistance and valuable discussions. I also thank all my past and current collaborators as well as Yasuyuki Hashiguchi, Yumi Nakai (Osaka Medical College), Takayuki Kohchi (Kyoto University), and Olivier Vallon (CNRS/Université Pierre et Marie Curie) for evolutionary considerations.

AUTHOR CONTRIBUTIONS

The author is solely responsible for the content of this article.

Received April 28, 2015; revised June 1, 2015; accepted June 2, 2015; published June 12, 2015.

REFERENCES


COMMENTARY


