Current Plant Biology 1 (2014) 68–72

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

Expression of a transferred nuclear gene in a mitochondrial genome

Yichun Qiu, Samuel J. Filipenko, Aude Darracq, Keith L. Adams*

Department of Botany, University of British Columbia, 6270 University Blvd, Vancouver, BC, Canada V6T 1Z4

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form 22 August 2014
Accepted 23 August 2014

Keywords:
Mitochondria
Intracellular gene transfer
Genome evolution
Brassicaceae

ABSTRACT

Transfer of mitochondrial genes to the nucleus, and subsequent gain of regulatory elements for expression, is an ongoing evolutionary process in plants. Many examples have been characterized, which in some cases have revealed sources of mitochondrial targeting sequences and cis-regulatory elements. In contrast, there have been no reports of a nuclear gene that has undergone intracellular transfer to the mitochondrial genome and become expressed. Here we show that the *orf164* gene in the mitochondrial genome of several Brassicaceae species, including *Arabidopsis*, is derived from the nuclear *ARF17* gene that codes for an auxin responsive protein and is present across flowering plants. *Orf164* corresponds to a portion of *ARF17*, and the nucleotide and amino acid sequences are 79% and 81% identical, respectively. *Orf164* is transcribed in several organ types of *Arabidopsis thaliana*, as detected by RT-PCR. In addition, *orf164* is transcribed in five other Brassicaceae within the tribes Camelineae, Erysimeae and Cardamineae, but the gene is not present in *Brassica* or *Raphanus*. This study shows that nuclear genes can be transferred to the mitochondrial genome and become expressed, providing a new perspective on the movement of genes between the genomes of subcellular compartments.

1. Introduction

Since the origins of mitochondria and plastids by endosymbiosis, three genomes have been coexisting in plant cells. There has been a tendency for DNA from the organelle genomes to be transferred to the nuclear genome, creating many nuclear mitochondrial (numt) sequences and nuclear plastid (nupt) sequences. Numerous pseudogenes of mitochondrial or chloroplast origin are present in nuclear genomes of a wide variety of eukaryotes (reviewed in Refs. [1,2]). In some cases, large regions of mitochondrial and chloroplast DNA have been transferred to the nuclear genome (e.g., [3–5]). Some mitochondrial and plastid genes were transferred to nuclear genome and then became expressed by acquiring existing nuclear *cis*-regulatory elements, as well as mitochondrial or chloroplast targeting sequences, then often replacing the functions of their counterparts in the organellar genomes (reviewed in Refs. [6–8]).

Angiosperm mitochondrial genomes contain DNA derived from the nuclear genome, although amounts vary among species. A large amount of the nuclear-derived DNA is from transposable elements, although sequences derived from exons of nuclear genes also are present in some mitochondrial genomes [9–13,37]. It has been inferred that the sequences derived from nuclear genes in mitochondrial genomes are pseudogenes. No nuclear-derived sequences have yet been reported as expressed. Here we show a case of a mitochondrial gene transferred from the nuclear genome that has become expressed.

2. Methods and materials

Sequences of *orf164* and *ARF17* from *Arabidopsis thaliana* were obtained from TAIR (v.10). Sequences of *orf164* and *ARF17* from *Arabidopsis lyrata* were obtained from the PLAZA v3.0 Dicots database (http://bioinformatics.psb.ugent.be/plaza/; [14]). BLAST searches of GenBank were used to search for sequences homologous to *orf164* and *ARF17* in other species. The nucleotide and amino acid alignments were generated by MUSCLE and followed by manual adjustments [15].

To analyze sequence rate evolution of *orf164*, sequences of *ARF17* were obtained from several eurosid species including *Carica papaya*, *Citrus sinensis*, *Eucalyptus grandis*, *Populus trichocarpa* and *Prunus persica* from PLAZA v3.0 Dicots [14], and *Tarenaya hassleriana* from GenBank's wgs database (gb|AOUI01012032.1), and aligned with *orf164* and *ARF17* using MUSCLE with the default settings [15]. The dN/dS ratio along each branch was determined using a phylogeny-based free-ratio test using Codeml in PAML [16].

Total RNA was extracted from multiple organ types of *A. thaliana* (ecotype col-0) and from seedlings of *A. arenosa* using the Ambion RNAqueous Kit following the manufacturer's protocol.

* Corresponding author. Tel.: +1 604 822 2355.
E-mail address: keitha@mail.ubc.ca (K.L. Adams).

http://dx.doi.org/10.1016/j.cpb.2014.08.002

Leaves from *Capsella bursa-pastoris*, *Turritis glabra*, *Erysimum pulchellum*, *Cardamine flexuosa* and *Armoracia rusticana* were used for RNA extraction as above. Nucleic acid concentration and purity were determined using a spectrophotometer and quality was visualized through gel electrophoresis on 2% agarose gel. RNA was treated with DNase-I (New England Biolabs) as outlined by the manufacturer's instructions. Reverse transcription was carried out using M-MLV reverse transcriptase (Invitrogen) following the manufacturer's instructions along with random hexamer primers (IDT). Then PCR reactions were performed with cDNAs as templates. Two pairs of *orf164*-specific primers were: forward-1, 5′-ATTGACGGCTGAAGCTGTCTCTGA-3′; reverse-1, 5′-ACGCCATGGACCAGTTTCCTGATA-3′; forward-2, 5′-TGTAGTTATTATCAGAGCAATGGAGGCG-3′; reverse-2, 5′-ATAGTGAAGGGGATCTTATACCTGAAGC-3′. Primers for other genes included: *orfX* forward (5′-TGGAGAACAAAGGACGAAATACA-3′) and reverse (5′-TATCCGGAGGTGTGGAAAGA-3′); *ccb203* forward (5′-GACCACTACTTCGCCTCTTTG-3′) and reverse (5′-CTATGAACGGGAGCTAGCAATC-3′); *matR* forward (5′-TTAAGGACAGGTCGTCGTATTG-3′) and reverse (5′-GGTCTCTCATGGCCCAATTAT-3′); *cox2* forward (5′-CGATGAGCAGTCACTCACTTT-3′) and reverse (5′-ATTGGATACCCGAGAACCATAATC-3′). The PCR cycling program for *orf164* amplification was 94° for 3 min; 20–35 cycles of 94° for 30 s, 55° for 30 s, 72° for 30 s; and 72° for 7 min. PCR cycling conditions for the other genes were the same except that 52° was used as the annealing temperature. PCR products were visualized on 1.2% agarose gels, the bands were cut out of the gels, DNA was eluted and then sequenced to confirm that the amplified sequences were the correct targets.

To identify other mitochondrial open reading frames of nuclear origin, all sequences of nuclear genes in *A. thaliana* were obtained from TAIR (v.10) and aligned against the *A. thaliana* mitochondrial genome (GI:26556996) using YASS software [17] with default parameters. We identified genes having an *e*-value <1.0E−10 and not located in the chr2:3247243-3509307 region (corresponding to the mitochondrial genome insertion into chromosome 2). The resulting list was filtered to remove transposable element-related sequences, mitochondrial sequences transferred to the nucleus, short open reading frames (less than 300 bp), nuclear intron-derived sequences, and mitochondrial-nuclear sequence pairs with less than 60% identity (Table S1).

Supplementary Table S1 related to this article can be found, in the online version, at doi:10.1016/j.cpb.2014.08.002.

3. Results and discussion

3.1. Orf164 in the mitochondrial genome of A. thaliana

Orf164 is a predicted gene in the *A. thaliana* mitochondrial genome [36], located between the tRNA gene *trnQ* (tRNA-Gln) and a pseudo-tRNA gene *ψtrnW* for tRNA-Trp (Fig. 1). *Orf164*, which has a locus number ATMG00940, contains an intact open reading frame of 495 nucleotides corresponding to 164 amino acids, according to TAIR (v.10) database (http://www.arabidopsis.org/). However, when we analyzed the genomic DNA and cDNA of *orf164* by Sanger sequencing following PCR, we detected a sequencing error close to the 3′ end of the predicted coding sequence, where an additional A should be present after the 446th nucleotide A, causing subsequent frame shift, and introducing a new stop codon that ends the coding region earlier. We also checked the recently sequenced and assembled complete mitochondrial genomes from three different *A. thaliana* ecotypes (C24, L*er* and Col-0) from Davila et al. [18] and we found the same additional A. The corrected *orf164* open reading frame should be 462 nucleotides and 153 amino acids.

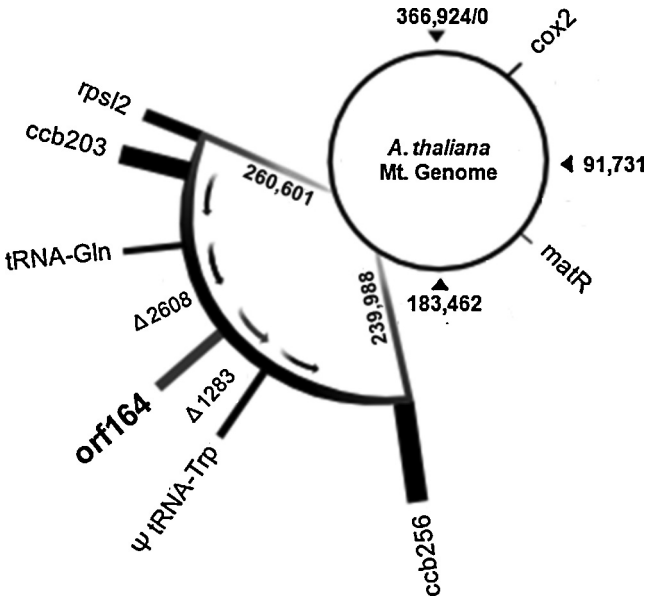

Fig. 1. Diagram of the *Arabidopsis thaliana* mitochondrial genome, with the region around *orf164* shown in detail. Arrows indicate the direction of transcription of the genes. Triangles followed by numbers indicate the sizes of the intergenic regions upstream and downstream of *orf164*.

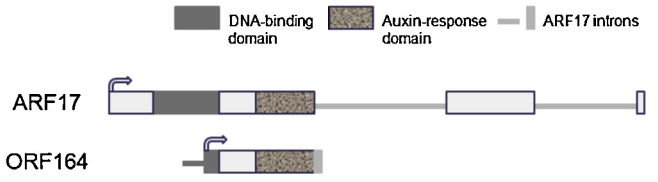

Fig. 2. Structures of *orf164* and *ARF17*. Arrows indicate the transcription start sites. Boxes indicate exons and bars indicate introns.

3.2. Orf164 is similar to nuclear ARF17 and derived from nuclear to mitochondrial intracellular gene transfer

Orf164 has high sequence similarity to a nuclear gene, *ARF17* (*AUXIN RESPONSE FACTOR 17*, AT1G77850). Comparing the sequences, *orf164* and *ARF17* share 79% nucleotide sequence identity and 81% amino acid identity. *ARF17* has two exons, and the first exon contains a DNA-binding domain and a domain regulating auxin-response gene expression (Fig. 2). *Orf164* starts at the position corresponding to the 206th codon within exon 1 of *ARF17*, using an ATG start codon that corresponds to an internal methionine codon in *ARF17*. At the 3′end of the *orf164* coding region, there are eight out of nine consecutive nucleotides that are identical to the intron at the exon–intron junction within *ARF17* (Fig. 3). The nucleotide in this region that is not identical to *ARF17* was a mutation that created the *orf164* stop codon. Eighty-four bp of the 5′UTR of *orf164* is derived from *ARF17* (Fig. 3). A mitochondrial sequence with similarity to *ARF17* was noticed by Hagen and Guilfoyle [19] and Liscum and Reed [20] in articles on *ARF* genes, but neither report identified the mitochondrial sequence as being *orf164* nor did any further characterization.

Using BLAST searches, we found many *ARF17* orthologous genes in a variety of angiosperm species. However, *orf164* has no homologous sequence in any sequenced mitochondrial genomes other than in *Arabidopsis*. We found a sequence from *A. lyrata*, AL3G32400, which is almost identical to *orf164* but annotated as a nuclear gene. However, a block of ten thousand base pairs surrounding AL3G32400 is about 99% identical to the *A. thaliana* mitochondrial genome, indicating that the sequence in *A. lyrata* is actually