A horizontally transferred $tRNA^{Cys}$ gene in the sugar beet mitochondrial genome: evidence that the gene is present in diverse angiosperms and its transcript is aminoacylated

Kazuyoshi Kitazaki, Tomohiko Kubo, Hiroyo Kagami, Takuma Matsumoto, Asami Fujita, Hiroaki Matsuhira, Muneyuki Matsunaga and Tetsuo Mikami

Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan

Corresponding author: Tomohiko Kubo
Laboratory of Genetic Engineering, Research Faculty of Agriculture
Hokkaido University
N-9, W-9, Kita-ku, Sapporo 060-8589, Japan.
Tel and Fax: +81-11-706-2484
E-mail: gelab@abs.agr.hokudai.ac.jp

Running head: Horizontally transferred tRNA gene in angiosperm

Key words: *horizontal gene transfer, plant mitochondria, angiosperm, tRNA, dual targeting, aminoacyl-tRNA synthetase*

Accession numbers of nucleotide sequences: AB434712, AB434715, AB434716, AB602769-AB602798, AB604166-AB604174, AB606424-AB606430

Word counts: summary, 222; introduction, 517; results, 2642; discussion, 1070; experimental procedures, 990; acknowledgements, 138; table and figure legends, 354.

Present addresses: K. Kitazaki, Central Research Institute of Electric Power Industry, Abiko, Chiba, 270-1194, Japan

H. Matsuhira, National Agricultural Research Center for Hokkaido Region, 1 Hitsujigaoka, Toyohira-ku, Sapporo, Hokkaido, 062-8555, Japan

Kazuyoshi Kitazaki and Tomohiko Kubo contributed equally to this article.

SUMMARY

Of the two $tRNA^{Cys}$(GCA) genes, *trnC1*-GCA and *trnC2*-GCA, previously identified in mitochondrial genome of sugar beet, the former is a native gene and probably a pseudo copy, whereas the latter, of origin unknown, is transcribed into a tRNA [$tRNA^{Cys2}$(GCA)]. In this study, the *trnC2*-GCA sequence was mined from various public databases. To evaluate whether or not the *trnC2*-GCA sequence is located in the mitochondrial genome, the relative copy number of its sequence was assessed in a number of angiosperm species, using a quantitative real time PCR assay. The *trnC2*-GCA sequence was found to exist sporadically in the mitochondrial genomes of a wide range of angiosperms. The mitochondrial $tRNA^{Cys2}$(GCA) species from sugar beet, spinach and cucumber were found to be aminoacylated, indicating that they may participate in translation. We also identified a sugar beet nuclear gene that encodes cysteinyl-tRNA synthetase, which is dual-targeted to mitochondria and plastids, and may aminoacylate $tRNA^{Cys2}$(GCA). What is of particular interest is that *trnC1*-GCA and *trnC2*-GCA co-exist in the mitochondrial genomes of eight diverse angiosperms, including spinach, and that the spinach $tRNA^{Cys1}$(GCA) is also aminoacylated. Taken together, our observations lead us to surmise that *trnC2*-GCA might have been horizontally transferred to a common ancestor of eudicots, followed by co-existence and dual-expression of *trnC1*-GCA and *trnC2*-GCA in mitochondria with occasional loss or inactivation of either *trnC*-GCA gene during evolution.

INTRODUCTION

Plant mitochondria rely upon three different categories of tRNAs for protein synthesis (Marechal-Drouard *et al.,* 1993). The classification of these tRNAs is based upon the subcellular location and the evolutionary origin of their genes. Native tRNAs are encoded by mitochondrial genes derived from the α-proteobacterial-type ancestor, plastid-like tRNAs transcribed from plastid DNA insertions in the mitochondrial genome, and nucleus-encoded tRNAs imported from the cytosol (Marechal-Drouard *et al.,* 1993). The number and/or identity of tRNA species in the different categories vary among plant species (Kubo and Mikami, 2007).

The differing patterns of mitochondrial tRNAs are correlated with the characteristics of their corresponding aminoacyl-tRNA synthetases (aaRSs), which catalyze the charging of one of the 20 amino acids to the cognate tRNAs with high fidelity (McClain, 1993) and are believed to co-evolve with their tRNAs (Lipman and Hou, 1998). This inference was substantiated by a systematic analysis of organellar aaRSs in *Arabidopsis thaliana* (L.) Heynh., which identified 24 organellar aaRSs, 17 dual-targeted to mitochondria and plastids, and 5 targeted to both the cytosol and mitochondria (Duchene *et al.,* 2005). Fourteen of the 17 aaRSs shared by both organelles are of eubacterial origin and, for the most part, aminoacylate organelle-encoded tRNAs of eubacterial origin. All five of the aaRSs shared between the cytosol and mitochondria, are eukaryotic in nature, and correspond to mitochondrially-imported cytosolic tRNAs.

We previously determined the complete nucleotide sequence of the mitochondrial genomes from normal (fertile) and male-sterility-inducing cytoplasms of sugar beet (*Beta vulgaris* L., Caryophyllales) (Kubo *et al.,* 2000; Satoh *et al.,* 2004). Twenty-one different tRNA genes for 16 amino acids were identified by homology to their respective counterparts in *A. thaliana* and other higher plants, and/or by their standard cloverleaf structures (Kubo *et al.,* 2000). Of these, 12 were native resident genes, 8 originated in plastids, and one, the $tRNA^{Cys}$ gene (designated *trnC2*-GCA), was of unknown origin.

Interestingly, this novel *trnC2*-GCA gene is transcribed whereas the native $tRNA^{Cys}$ gene (*trnC1*-GCA), present in the sugar beet mitochondrial genome, is most likely a pseudo gene copy (Kubo *et al.,* 2000). This raises the intriguing question as to whether the *trnC2*-GCA gene originated via a horizontal transfer from an unrelated organism. If this were the case, one would expect that sequences homologous to *trnC2*-GCA would exhibit a phylogenetically disjunctive distribution and that some of them would be expressed as they are in sugar beet. In the case of the group I intron of *cox1*, a well-known example of a horizontally-transferred sequence in plant mitochondria, distribution of the intronic sequence is observed in a wide range of species but in a patchy manner (*i.e.*, both intron-bearing and intron-missing species co-exist in a taxon) (Cho *et al.,* 1998).

To address the question as to whether *trnC2*-GCA arose from a horizontal transfer, we chose to study the evolution of the *trnC2*-GCA gene. We show that *trnC2*-GCA does exist in the mitochondria of diverse plant species, but with capricious taxonomic distribution and that the gene is expressed in some taxa. We also characterize the sugar beet nuclear gene coding for cysteinyl-tRNA synthetase (*cysRS*), which most likely recognizes and aminoacylates the *trnC2*-GCA transcript [$tRNA^{Cys2}$(GCA)].

RESULTS

trnC2-GCA sequences in databases

We began this study by searching DDBJ/EMBL/GenBank databases for the *trnC2*-GCA sequence using BLAST. When the 72-bp coding region of sugar beet *trnC2*-GCA was used as a query, three entries were retrieved. The first two were the entire nucleotide sequence of watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai, Cucurbitaceae) mitochondrial (mt) DNA (GQ856147, Alverson *et al.*, 2010) and mung bean (*Vigna radiata* (L.) R. Wilczek, Fabaceae) mtDNA (HM367685, Alverson et al. 2011) (Figure 1

and Figure S1). The other entry was an expressed sequence tag (EST) of groundnut (*Arachis hypogaea* L., Fabaceae) (EZ742316, 942 bp), which contains the entire *trnC2*-GCA (Figure S2). No additional genic sequences (protein-coding sequences and open reading frames longer than 90 bp) were detected within EZ742316. A BLAST search against the DDBJ/EMBL/GenBank databases using EZ742316 as a query revealed that a 287-bp stretch containing *trnC2*-GCA displayed the highest similarity and was best matched to a portion of mung bean mtDNA (Figure S2). We also found a 175 bp sequence homologous to part of tobacco (*Nicotiana tabacum* L.) mtDNA, outside the 287-bp stretch (Figure S2).

Identification of the mitochondrial *trnC2*-GCA in watermelon and mung bean prompted us to assess the presence or absence of *trnC2*-GCA in 27 additional plant mitochondrial genomes whose entire sequences were available. In these mitochondrial genome sequences, native *trnC1*-GCA and/or plastid-like *trnC*-GCA occurred but no *trnC2*-GCA was seen (Table 1). Database searches were further performed using 9 public services other than DDBJ/EMBL/GenBank, leading to the discovery of the *trnC2*-GCA sequence in cucumber (*Cucumis sativus* L., Cucurbitaceae), soybean (*Glycine max* (L.) Merr., Fabaceae) and *Aquilegia coerulea* E.James (Ranunculaceae) (Figure S1).

We retrieved two cucumber scaffolds (the assemblages of short sequence reads), scaffold_repeat_001243 (3264 bp) and scaffold_04063 (196081 bp), from the Cucurbit Genomics Database (CuGenDB) and Phytozome, respectively. Because our sequence analysis revealed that scaffold_repeat_001243 was entirely included within scaffold_04063 and that the overlapped region was nearly identical, only scaffold_04063 was analyzed further. The scaffold_04063 actually contains an intact *trnC2*-GCA (Figure S1). In the Phytozome, 17 transcripts are mapped on scaffold_04063. We searched the sequences with homology to the 17 transcripts against the DDBJ/EMBL/GenBank databases by using BLAST. As shown in Table S1, 12 out of 17 transcripts have high homology either with mitochondrial genes such as *atp1* and *atp6* or with mtDNA sequences, whereas one is homologous to plastid DNA and four have no homology to any

known sequences. This strongly suggests that scaffold_04063 contains the authentic cucumber mitochondrial sequences which are registered in the DDBJ/EMBL/GenBank databases. We tested 120 further entries of authentic cucumber mtDNA sequences (listed in Table S2) for homology to scaffold_04063. Through a BLAST2 search, 81 entries showed homology with scaffold_04063 (expected value < 1×10^{-10}) (Table S2).

From Soybase and Phytozome we retrieved soybean scaffold_682 (7503 bp), which has not yet been assigned to any of the 20 chromosomal linkage groups of soybean. Sequence analysis revealed that *trnC2*-GCA was included within a 2467 bp region that exhibits high homology to mung bean mtDNA (Figure S3). Although additional segments with significant homology to plant mtDNA sequences are scattered throughout scaffold_682, there are no mitochondrial genes other than *trnC2*-GCA in this scaffold.

Scaffold_23 (4097486 bp) was found to contain two copies of *Aquilegia coerulea* *trnC2*-GCA, arranged in a tail-to-tail manner at a distance of 63702 bp. Twenty nine transcripts mapped near the two *trnC2*-GCA copies were found to be associated with mitochondrial genes or mtDNA (Table S3).

A detailed comparison of the *trnC2*-GCA sequences detected is given in a later section.

Detection of *trnC2*-GCA in four additional angiosperm species

To examine whether the *trnC2*-GCA sequence is present in other angiosperm species, total cellular DNA was isolated from 26 species, including sugar beet and cucumber (Figure 2) and subjected to PCR using primers 1 and 2 (Table S4) targeting the *trnC2*-GCA coding region. The assay resulted in a product of ~72 bp (expected amplicon size) from templates of *Daphniphyllum macropodum* Miq. (Daphniphyllaceae), *Hibbertia pedunculata* R.Br. ex DC. (Dilleniaceae) and *Basella rubra* L. (Basellaceae) (Figure 2). PCR products were sequenced to clearly demonstrate their derivation from the

trnC2-GCA gene (Figure S1). A PCR-generated probe covering the sugar beet *trnC2*-GCA (primers 3 and 4) was allowed to hybridize with DNA gel blots of total cellular DNA from *Hibbertia* and *Daphniphyllum*. In both cases, a strong signal band appeared, confirming the presence of *trnC2*-GCA in *Hibbertia* and *Daphniphyllum* (Figure 3).

Since it was difficult to prepare sufficient DNA or RNA from *Daphniphyllum, Hibbertia* and *Basella* for further analysis and we had found spinach (*Spinacea oleracea* L. , Amaranthaceae) to also carry the *trnC2*-GCA sequence (Figure 3) we considered the latter species suitable for more detailed study. Two clones (2.7-kbp *Eco*RI and 3.2-kbp *Pst*I fragments) that hybridized to the *trnC2*-GCA probe were selected from genomic libraries derived from mitochondria-enriched DNA of spinach. Nucleotide sequences of the two clones were assembled into a single continuous segment of 5504 bp, which clearly contained *trnC2*-GCA (Figure S4). No mitochondrial genes other than *trnC2*-GCA were found in the 5504 bp sequence, though 8 mtDNA-associated regions (115 to 711 bp in length) were present in the segment.

Estimation of relative copy number of *trnC2*-GCA

We sought to determine whether the *trnC2*-GCA gene was located in the mitochondrion or the nucleus. Resting on the assumption that mitochondrial genes are in high copy number relative to single or low-copy-number nuclear genes, we employed a quantitative real time PCR (qPCR) assay to compared the copy number of the *trnC2*-GCA gene to that of a low copy nuclear gene. Sugar beet was chosen to test the feasibility of this qPCR assay. As a reference, we selected a gene encoding granule-bound starch synthase (*bvWX*) from a list of low-copy nuclear genes (Sang, 2002; Duarte et al., 2010). The *cob* gene was selected as a mitochondrial control given its ubiquitousness in angiosperm mitochondrial genomes (Adams et al., 2002). Experimental conditions are detailed in

EXPERIMENTAL PROCEDURES. The Ct (threshold cycle estimated by real-time PCR) value for the reference nuclear gene (*bvWX*) was significantly greater than that of either of the mitochondrial genes examined, the differences being 6.0 to 6.14 against *cob* and 5.32 to 5.46 against *trnC2*-GCA. The amplification efficiency was estimated to be ~1 (Table S5), making it likely that the observed difference in Ct (ΔCt) reflected a difference in copy-number of the target sequences in the sample DNA. Hereafter we used $2^{\Delta Ct}$ as an index of the relative copy-number of the investigated gene to the control nuclear gene. The $2^{\Delta Ct}$ values for *bvWX,* sugar beet *cob* and sugar beet *trnC2*-GCA were 1 (2^0), 61.13 and 43.44, respectively (Table 2). These values are comparable to the mtDNA copy number reported in *Arabidopsis thaliana*, *Antirrhinum majus* L.*,* tobacco and *Pelargonium zonale* (L.) L'Hér. ex Ait. (Wang *et al.*, 2010). Variation in the copy number of different mitochondrial genes has been observed in some angiosperms*,* and is seemingly associated with the complex organization of angiosperm mitochondrial genomes (Woloszynska *et al.*, 2006; Preuten *et al.*, 2010).

With the aim of further verifying the validity of our quantification method, qPCR was performed to estimate the $2^{\Delta Ct}$ values of spinach-, cucumber, and soybean *trnC2*-GCA genes. Using a spinach orthologue to *Arabidopsis* at2g32520 as a reference nuclear gene, $2^{\Delta Ct}$ values of spinach *cob* and *trnC2*-GCA were calculated to be 22.71 and 14.27, respectively (Table 2). Likewise, the $2^{\Delta Ct}$ of cucumber and soybean *trnC2*-GCA genes were 27 and 34 times higher, respectively, than that of their respective reference nuclear genes (Table 2). The soybean *cob* exhibited a high $2^{\Delta Ct}$ value (68.6), although the soybean homologue of *cob* has been mapped to chromosome 07 (Glyma07g12630 in Phytozome). We think it likely that the soybean *cob* sequence has recently migrated into the nuclear genome and a mitochondrial copy still exists. No *trnC2*-GCA sequence is found in the plastid genomes of cucumber, spinach or soybean (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid).

Since the qPCR assay yielded reliable results for the quantification of *trnC2*GCA copy abundance, it was applied to seven further species from six families. Given that only

a limited amount of DNA could be isolated from the supplementary species — *Chenopodium album* L. (Amaranthaceae), *Celosia cristata* L. (Amaranthaceae), *Basella rubra*, *Mirabilis jalapa* L. (Nyctaginaceae), *Myrtillocactus geometrizans* (Mart.) Console (Cactaceae), *Daphniphyllum macropodum*, and *Rumex obtusifolius* L. (Polygonaceae) —. nucleotide sequences of *cob* and reference nuclear genes from these species were obtained either through a database search or PCR amplification (Table S5). Given $2^{\Delta Ct}$ values ranging from 6.61 to 37.02 in *Chenopodium*, *Celosia*, *Basella*, *Myrtillocactus*, and *Daphniphyllum,* the qPCR assay strongly suggests the presence of the mitochondrial *trnC2*-GCA sequence in these species (Table 2). The $2^{\Delta Ct}$ of *Chenopodium* mitochondrial genes was substantially lower than those of the other mitochondrial genes examined, possibly as a result of the higher ploidy level of the nuclear genome in this species (Bhargava *et al.*, 2009). Efforts to PCR-amplify the entire *trnC2*-GCA from *Mirabilis* and *Rumex* failed, indicating the lack of an intact gene in these species.

Sequence comparison of *trnC2*-GCA genes

After the detection of mitochondrial *trnC2*-GCA genes in a variety of angiosperms, the next aim was to compare their nucleotide sequences. This was undertaken for eight mitochondrial *trnC2*-GCA (from sugar beet, spinach, watermelon, cucumber, soybean, mung bean, *Daphniphyllum* and *Basella*) and four *trnC2*-GCA [from groundnut, *Aquilegia* (2 copies) and *Hibbertia*] whose subcellular localization remained uncertain. Figure 1 shows a spinach *trnC2*-GCA gene that bears the normal pattern of invariant and semi-invariant residues. Comparatively, the sugar beet *trnC2*-GCA sequence has a nucleotide substitution (T60-to-C) in the T loop and a nucleotide deletion (C20) in the D loop. We also found the spinach sequence to differ by two nucleotides (A38 and C56) from that of soybean and mung bean, by one nucleotide (G28) from that of cucumber and watermelon, and by one nucleotide (A64) from that of groundnut.

Aminoacylation of the tRNA^{Cys2}(GCA) molecule

We investigated the *in vivo* aminoacylation of the tRNA species transcribed from the sugar beet *trnC2*-GCA [tRNA^{Cys2}(GCA)]. Total RNA was isolated under both basic and acidic conditions: the former condition was expected to cause hydrolysis of the ester bond between tRNAs and amino acids, whereas, under the latter condition, tRNAs were expected to mostly remain aminoacylated. Different forms of tRNA^{Cys2}(GCA) were detected by RNA gel blot hybridization using the *trnC2*-GCA probe (Figure 4). When total RNA isolated under basic condition was examined, only one faster-migrating band was observed, which might correspond to uncharged tRNA^{Cys2}(GCA). In contrast, RNA gel blot analysis of total RNA extracted under acidic condition revealed a slower-migrating band as well as a faster-migrating one, the former of which most likely represents aminoacylated tRNA^{Cys2}(GCA).

Similar results were also obtained when the mitochondrial tRNA^{Cys2}(GCA) species from spinach and cucumber was subjected to a similar investigation of *in vivo* aminoacylation (Figure 4). On the other hand, the *trnC2*-GCA probe failed to hybridize with total RNA isolated from watermelon and soybean, indicating that the mitochondrial *trnC2*-GCA genes found in these two species are not transcribed at a detectable level (Figure 4).

Isolation and characterization of a mitochondrial cysteinyl-tRNA synthetase gene from sugar beet

We sought to clone a mitochondrial *cysRS* gene from sugar beet. Since no reports of plant organelle *cysRS* had been published when we initiated this study, we searched databases

for *Arabidopsis* genomic sequences with similarities to the *E. coli cysRS* gene (accession number X56234). The search revealed a *cysRS* gene (AC006593) that consisted of eight exons encoding 563 amino acids and was located on chromosome 2. We amplified a DNA fragment containing the three exons (4, 5, and 6) of *Arabidopsis cysRS* (*atcysRS*), which was subsequently used to screen a sugar beet cDNA library. Eventually, two incomplete cDNA clones were obtained and sequenced. Their 100% identical overlap sequences indicate that they are derived from the same mRNA but they lack the 5' portion of the coding region. 5' RACE (rapid amplification of cDNA ends) was carried out to generate a full-length cDNA fragment (*bvcysRS*) that contained an ORF capable of encoding 582 amino acid residues (Figure S5).

The 582-residue polypeptide was 68.5% identical to the atcysRS polypeptide and possesses two conserved motifs (His-Ile-Gly-His and Lys-Met-Ser-Lys-Ser) characteristic of class I aaRS (McClain 1993). Alignment of the *bvcysRS* sequence with that of *E. coli cysRS* revealed the presence of an NH_2-terminal extension in the sugar beet protein, suggesting that an organellar targeting sequence may be encoded in this NH_2-terminal domain (Figure S6).

DNA gel blot analysis using the sequences encompassing exons 2, 3 and 4 (PCR amplified with primers F and A2) as a probe showed that *bvcysRS* occurs as a single-copy gene (Figure 5). Transcription of *bvcysRS* was examined by qPCR using the primers A1 and Bv1. The analysis resulted in the amplification of a cDNA fragment corresponding to *bvcysRS* as expected (Figure 5), confirming that the gene in question is expressed in flower buds and leaves. To determine the exon/intron structure of *bvcysRS*, the corresponding genomic DNA fragment was PCR-amplified, cloned and sequenced . As seen in Figure S5, we found that an additional intron (88 nt in size) is present in Ala505 and that the positions of the other seven introns are identical to those observed for the introns in *atcysRS* (see Figure S6).

Organellar targeting of bvcysRS

Web-based prediction of subcellular localization of the putative organellar bvcysRS replied secretory pathway (TargetP) and possibly mitochondrial (Predotar). To further analyze the subcellular localization of the putative organellar bvcysRS, the first 106 amino acids of this enzyme were fused to the GFP reporter protein and tested by transient expression in the epidermal cells of onion bulbs. Two RFP-protein fusion constructs, Mt-RFP and Pt-RFP, were used to determine a mitochondrial and plastidic pattern: the former construct contains a mitochondrial targeting presequence of *Arabidopsis* ATPase delta subunit cDNA (Arimura and Tsutsumi, 2002) and the latter contains a plastid targeting presequence of *Arabidopsis* RuBisco activase. The green fluorescence of bvcysRS::GFP apparently overlapped with both Mt-RFP and Pt-RFP red fluorescence, indicating dual targeting of bvcysRS to mitochondria and plastids (Figure 6a-f). When GFP without any presequences was expressed, no such fluorescent pattern was observed (Figure 6g).

tRNA^{Cys}(GCA) genes of different origins co-exist in some angiosperm mitochondria

Finally, we focused on the distribution in plant mitochondrial genomes of the *trnC1*-GCA sequence, descendant from the genome of an ancestral endosymbiotic bacterium (Gray, 1999). As described above, a search of completely sequenced mitochondrial genomes from plants confirmed the presence of *trnC1*-GCA sequence in 20 out of 30 mitochondrial genomes. The remaining 10 mitochondrial genomes, all of which are from the members of the grass family (Poaceae), are devoid of both *trnC1*-GCA and *trnC2*-GCA. Instead, the Poaceae mitochondrial genomes possess a plastid-like *trnC*-GCA. In Cucurbitaceae, watermelon contains both *trnC1*-GCA and *trnC2*-GCA, whereas zucchini contains *trnC1*-GCA and plastid-like *trnC*-GCA. Mung bean contains

both *trnC1*-GCA and *trnC2*-GCA.

Since the entire nucleotide sequence of cucumber mtDNA is unavailable yet, DNA gel blot analysis and qPCR assay were carried out to determine whether *trnC1*-GCA is present in the cucumber mitochondrial genome. Both analysis failed to reveal the presence of *trnC1*-GCA in this taxon (Figure 3 and Table 2).

Two additional observations merit comment. The *trnC1*-GCA probe was found to hybridize to spinach total cellular DNA (Figure 3). Based on the qPCR assay (Table S5), the $2^{\Delta Ct}$ value of spinach *trnC1*-GCA was comparable to that of *trnC2*-GCA and *cob* (Table 1), which demonstrates that *trnC1*-GCA as well as *trnC2*-GCA resides in the spinach mitochondrial genome. Furthermore, expression and *in vivo* aminoacylation of spinach $tRNA^{Cys1}$(GCA) were confirmed by the fact that a slower-migrating band was detected when total RNA isolated under acidic condition was allowed to be hybridized with the *trnC1*-GCA probe (Figure 4).

The qPCR assay was conducted to determine whether the *trnC1*-GCA sequence exists in *Chenopodium*, *Celosia*, *Basella*, *Myrtillocactus*, *Daphniphyllum*, *Mirabilis* and *Rumex* (Table S5). The $2^{\Delta Ct}$ value of *trnC1*-GCA was similar to that of *cob* but much greater than that of nuclear gene control (Table 2), indicating the presence of this gene in their mtDNA.

Discussion

The present study was designed to follow-up the initial observation that while mitochondrial $tRNA^{Cys}$(GCA) is encoded by either a native gene or a plastid-like gene in many angiosperms, a novel *trnC2*-GCA gene of unknown origin is the only functional mitochondrial $tRNA^{Cys}$(GCA) gene in sugar beet (Kubo *et al*., 2000). First we wanted to examine the distribution of *trnC2*-GCA in a broad range of plant species. For this purpose, various public databases were mined for the presence or absence of *trnC2*-GCA. The taxa

surveyed include 30 plant species for which the entire mitochondrial genome sequence is available. To investigate the sub-cellular location of the *trnC2*-GCA sequence, we also applied a qPCR assay to estimate the relative copy number of this sequence in eight angiosperm species. These analyses, together with a PCR survey show that *trnC2*-GCA evidently resides in the mitochondrial genomes of an additional 10 species: watermelon and cucumber in the Cucurbitaceae; soybean and mung bean in the Fabaceae; *Daphniphyllum* in the Daphniphyllaceae; *Basella* in the Basellaceae; spinach, *Chenopodium* and *Celosia* in the Amaranthaceae; and *Myrtillocactus* in the Cactaceae. Our survey also indicates the presence of *trnC2*-GCA in groundnut, *Aquilegia* and *Hibbertia*, but it remains uncertain whether or not these *trnC2*-GCA sequences are mitochondrially located. On the other hand, no *trnC2*-GCA was found in a number of plant species, including the taxa (*e.g.*, *Arabidopsis*, rice and grapevine) whose genomic informations are almost entirely accessible.

The taxa judged to carry the mitochondrial *trnC2*-GCA represent a diverse assemblage of plants and have been assigned to four sub-class (viz. Hamamelidae, Caryophyllidae, Dilleniidae and Rosidae) of dicots. Interestingly, within the family Cucurbitaceae, watermelon and cucumber retain the mitochondrial *trnC2*-GCA, while zucchini lacks the gene (see Table 1). The order Caryophyllales contains mitochondrial *trnC2*-GCA-lacking taxa (e. g. *Silene*, *Mirabilis* and *Rumex*) as well as a number of mitochondrial *trnC2*-GCA-carrying taxa (e. g. sugar beet, spinach, *Chenopodium*, *Celosia*, *Basella* and *Myrtillocactus*). These results remind us of the case of the *cox1* intron, the only reported group I intron in the mitochondrial genomes of angiosperms (Cho *et al.*, 1998). The intron is considered to have been acquired by horizontal transfer from a fungal source in the common ancestor of angiosperms. From a survey of hundreds of diverse plant species, Sanchez-Puerta *et al.* (2008) and Cusimano *et al.* (2008) drew an inference that such fungal donations might have been followed by many plant-to-plant lateral transfers of as well as occasional losses of the intron, resulting in a sporadic distribution of the intron in angiosperm mitochondria. Analogous reasoning can be

applied to the patchy phylogenetic distribution of the origin-unknown *trnC2*-GCA, though we cannot rule out the possibility that horizontal gene transfer from some unrelated organism occurred independently in different lineages of angiosperms.

Little is known about the origin of *trnC2*-GCA sequence. The best match between the plant *trnC2*-GCA and sequences in DDBJ/EMBL/GenBank databases is *trnC*-GCA of *Sphaerobacter thermophilus* (accession number CP001823), a Gram-positive, non-spore-forming bacterium isolated from an aerobic thermophilic sludge (Hensel *et al.*, 1989; Hugenholtz and Stackebrandt, 2004). However, the *trnC2*-GCA coding region is short (72-73 bp), making it difficult to draw conclusions or to conduct a robust phylogenetic analysis. Identification of the donor organism is the future challenge.

In this study, we demonstrated the sugar beet $tRNA^{Cys2}$(GCA) molecules to be aminoacylated *in vivo*, which supports the idea that they may participate in mitochondrial translation. Considering that no nuclear encoded $tRNA^{Cys}$(GCA) imported from the cytosol has been documented in angiosperm mitochondria to date (Duchene *et al.*, 2009), it seems most likely that functional $tRNA^{Cys}$(GCA) is encoded solely by the *trnC2*-GCA gene in sugar beet mitochondria. Once *trnC2*-GCA was captured and activated, a state of dual intact and transcribed *trnC*-GCA genes [*viz*. the novel and native (or plastid-like)] copies] must have been established. As far as we examined, such a state persists in only spinach where both *trnC2*-GCA and *trnC1*-GCA are actually aminoacylated *in vivo* and not mutually exclusive. Inactivation (silencing) of either *trnC2*-GCA or *trnC1*-GCA (or plastid-like *trnC*-GCA) might have followed the stage of dual expression. The pseudogenization of *trnC1*-GCA is exemplified by the sugar beet copy (Kubo *et al*., 2000) whereas silencing of *trnC2*-GCA is represented by the watermelon copy. Although the reason for these tRNA replacements is not obvious, there seems no selective advantage for mitochondrial $tRNA^{Cys}$(GCA) to be coded for differently by *trnC2*-GCA versus *trnC1*-GCA in angiosperms such as sugar beet and watermelon.

For the novel $tRNA^{Cys2}$(GCA) species to function, it must be correctly recognized by its cognate cysteinyl-tRNA synthetase. In higher plants, mitochondrial aaRSs are

encoded by the nuclear genome and are post-translationally addressed to mitochondria (Duchene *et al.,* 2005). Some of these aaRSs were reported to have the same genetic origin as their substrate tRNAs. This is the case for $tRNA^{Cys}$(GCA) and its cognate *cysRS* in *Arabidopsis,* both of which are considered to be of genuine mitochondrial origin (Peeters *et al.,* 2000; Duchene *et al.,* 2005). In this study, we isolated a sugar beet cDNA encoding the cysRS enzyme (bvcysRS). The sugar beet gene appears to be an orthologue of *Arabidopsis cysRS*, implying that *bvcysRS* is also mitochondrial in origin. Homologous *cysRS* genes are found in some angiosperm species such as cucumber and rice, in which origin of mitochondrial $tRNA^{Cys}$(GCA) genes differs (Figure S6). As discussed above, the sugar beet *trnC2*-GCA may have been acquired by horizontal gene transfer from an unrelated organism. The use of the $tRNA^{Cys2}$(GCA) species in mitochondrial translation might have been made possible because the cysRS enzyme(s) of mitochondrial origin and $tRNA^{Cys2}$(GCA) happened to work efficiently together in aminoacylation.

This speculation is not unreasonable, considering the identity elements that have previously been defined for $tRNA^{Cys}$. Lipman and Hou (1998) stated that the U73 in the acceptor stem, the GCA anticodon, and the 15-48 base pair are the important features for aminoacylation of *E. coli* $tRNA^{Cys}$ by the *E. coli* cysRS. The sugar beet $tRNA^{Cys2}$(GCA) was shown to carry U73 and the GCA anticodon (Kubo *et al.,* 2000). It is also worth noting that the sugar beet and *E. coli* tRNAs share the same flanking sequences (from U33 to A38) of the GCA anticodon (Kubo et al. 2000; see also Figure. 1). Given that sugar beet cysRS and *E. coli* cysRS share the same tRNA recognition specificity because of their amino-acid sequence similarity, one can expect that sugar beet cysRS is able to charge the $tRNA^{Cys2}$(GCA). However, the presence of another cysRS in sugar beet mitochondria cannot be entirely ruled out.

EXPERIMENTAL PROCEDURES

Database search and sequence analysis

The URLs of web sites consulted in this study are summarized in Table S6. Nucleotide sequence data of the scaffolds mentioned in this study were used under permission of their sources. Sequencher 4.5 (Gene Codes, Ann Arbor, MI, USA) and Genetyx-mac (Genetyx, Tokyo, Japan) were also used for sequence analysis.

Plant materials

Of a total of 33 plant species were used in this study, 26 are listed in Figure 2. Soybean (cv. 'Kurobe'), spinach (*Spinacea oleracea*), *Chenopodium album, Celosia cristata* (cv. 'Yachiyo-keitou*,' Myrtillocactus geometrizans,* watermelon (cv. 'Kinzan'), and *Rumex obtusifolius* were also used. Sugar beet (cv. 'TK81-O'), cucumber (cv. 'Kinuhikari'), soybean, *Celosia* and watermelon were grown in a greenhouse from seeds. Other materials were collected from the Botanical Garden of Hokkaido University, the Experiment Farm of Hokkaido University, or purchased from a local market.

Isolation of DNA

The isolation of total cellular DNA employed the method of Doyle and Doyle (1990). Mitochondria-enriched DNA of spinach was isolated from green leaves as described in Mikami *et al*. (1985). When necessary, sample DNA was purified by centrifugation in continuous CsCl density gradients.

DNA gel blot hybridization

Digested with a restriction endonuclease (Takara Bio, Ohtsu, Japan) total cellular DNA (5 μg) was then electrophoresed in a 1.0% agarose gel. The separated nucleic acid fragments were transferred to a Hybond N+ membrane (GE Healthcare UK, Amersham Place, England), and hybridization was carried out according to the instruction manual. Hybridization probes were labeled using the AlkPhos Direct DNA labeling system (GE Healthcare UK).

PCR

PCR primers used in this study are listed in Table S4. The *trnC2*-GCA sequences were amplified with GoTaq (Promega, Madison, WI, USA) using 10-15 ng of total cellular DNA as template. RT-PCR was performed according to the method of Singer-Sam *et al.* (1990) using primers A1 and Bv1. Genomic DNA fragments of *bvcysRS* were amplified by using LA-Taq (Takara Bio) with primer combinations of cys1-cys2, cys1-cys4, cys3-cys6, and cys5-cys7. DNA fragments for hybridization probe experiments were amplified using rTaq (Takara Bio) or GoTaq. All the amplifications were done using the Gene Amp PCR System 9700 (Applied Biosystems, Drive Foster City, CA, USA) and Veriti thermal cycler (Applied Biosystems).

Quantitative real time PCR

DNA samples were assayed using DNA-Engine PTC-200, Chromo 4 and Opticon Monitor v. 3.1 (Bio-Rad Laboratories, Hercules, CA, USA) with 2 x SYBR GreenER

QPCR universal (Invitrogen, Carlsbad, CA, USA). Nucleotide sequences of targets were determined prior to the assay when the available sequence data were insufficient. PCR primers for the target capture were; *cob,* cob-seq-FW and cob-seq-RV; at2g32520-orthologues, 32520-FW and 32520-RV; EMB2765-orthologue, EMB-FW and EMB-RV; RP-ERS1, ERS1-FW and ERS1-RV. The conditions of each assay are summarized in Table S5. Reactions were carried out in a total volume of 25 μl with a final concentration of 0.4-μM primers. The quantification protocol began at 95˚C (3min), followed by 40 cycles of 95˚C (10sec) and 58.8-61.8 (depending on primer combination; see Table S5) (1min) with single data acquisition. After the quantification, all the reactants were heated to 95˚C (1min) and then cooled to 50˚C. A melting-curve was then drawn (53 to 87˚C, data acquisition every 0.5˚C) to verify that there was a single amplicon. This was confirmed by agarose gel electrophoresis after standard PCR of the same template/primer set combination. Amplification efficiency was examined by the quantification of five serial dilution of sample DNA (1:10, 1:20, 1:40, 1:80 and 1:160) for each primer set (Table S5). Quantification to calculate $2^{\Delta Ct}$ was carried out using a 1:50 dilution sample, with at least two replicates.

Acid-PAGE

Total cellular RNA was extracted from leaves ground in liquid nitrogen under acidic (pH 4.5) conditions (Chomczynski and Sacchi,1987). Nucleic acids were dissolved in 0.3 M sodium acetate (either pH 4.5 or pH 9.0). Samples (10-15 μg) were electrophoresed in 10-15% polyacrylamide gel containing 8 M urea and 0.1 M sodium acetate (pH 4.5) (Varshney *et al.,* 1991) and electroblotted onto a Hybond N+ membrane. Procedures for hybridization are as described above.

Molecular Cloning

Poly $(A)^+$ RNA from flower buds of sugar beet line 'TK81-O' were used to construct a cDNA library in a lambda gt10 vector using a TimeSaver cDNA Synthesis Kit (GE Healthcare UK, Amersham Place, England). A total of 4 x 10^5 recombinant phages were subjected to screening. 5' RACE (rapid amplification of cDNA ends) was carried out using a 5'-Full RACE Core Set (Takara Bio, Ohtsu, Shiga, Japan). Flower bud RNA was reverse-transcribed with a RT1 primer. The resultant cDNA was circularized with T4 RNA ligase and amplified by nested PCR using two sets of primers (primers A1 and S1, and primers A2 and S2). For genomic cloning, mitochondria-enriched DNA was digested with restriction enzymes and ligated into a pUC119 vector. Colony hybridization was carried out to isolate the clone to be analyzed (Sambrook et al., 1989). Cloning of PCR products were done by using a TOPO PCR cloning kit (Invitrogen). Nucleotide sequencing was carried out using a Li-COR 4000L system (Li-COR, Lincoln NE, USA).

GFP Expression Vector and Construct

A portion of a sugar beet cDNA clone was amplified by PCR using the primers Sa and Nc. The PCR product was then digested with *Sal*I and *Nco*I. The resultant DNA fragment was ligated into a pTH2 vector that contained the sGFP-TYG gene (Chiu *et al.,* 1996) driven by a cauliflower mosaic virus 35S promoter. Mt-RFP, which contains DsRedII (Clontech, Palo Alto, CA, USA), was a gift from Dr. S-i Arimura (University of Tokyo). For Pt-RFP construction, the mitochondria-targeting presequence of the Mt-RFP was replaced with the presequence (1-58 amino acid residues) of *Arabidopsis* RuBisco activase (DDBJ/EMBL/GenBank accession number X14212) that was amplified with primers

Sa-rubisco-FW and Nc-rubisco-RV from *Arabidopsis* cDNA. The plasmid was introduced into the epidermal cells of onion bulb using the GIE-III IDERA (Tanaka, Sapporo, Japan). MitoTracker Orange (Molecular Probes, Eugene, OR, USA) was used to visualize mitochondria. Fluorescent signals were captured by a BX50 microscopic system combined with DP70 digital camera (Olympus, Tokyo, Japan).

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr. T. Azuma, Dr. T. Hirata, Dr. Y. Kishima, Dr. T. Uchiyama, Dr. T. Yamada, and Dr. S. Niikura for providing plant materials or DNA samples, Dr. S. Huang and Dr. M. Goodstein for permission of using scaffold sequences, Dr. S.-i. Arimura for providing a plasmid, and Dr. L. Marechal-Drouard for her helpful comments and continuous encouragements. This work was done in part at the Research Center for Molecular Genetics, Hokkaido University, and was supported in part by Grants in Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, a Grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences, Japan, Shorai Foundation For Science and Technology, Kuribayashi Scholarship Foundation, and the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN).

Supplementary materials

Table S1 List of the best matched entries to the transcripts mapped onto scaffold_04063

Table S2 List of some authentic cucumber mtDNA sequences used in this study

Table S3 List of best matched entries to the transcripts that have been mapped near two *trnC2*-GCA copies in scaffold_23

Table S4 List of primers used in this study

Table S5 Summary of qPCR

Table S6 Summary of web URLs consulted in this study

Figure S1 Comparison of *trnC2*-GCA sequences

Figure S2 Nucleotide sequence of EZ742316

Figure S3 Nucleotide sequence of scaffold_682

Figure S4 Nucleotide sequence of the 5504-bp sequence containing spinach *trnC2*-GCA

Figure S5 Nucleotide sequence of sugar-beet gene coding for cysteinyl-tRNA synthetase (*bvcysRS*)

Figure S6 Multiple alignments of the amino acid sequences from cysteinyl-tRNA synthetases

REFERENCES

Adams, K.L., Qiu, Y.L., Stoutemyer, M. and Palmer, J.D. (2002) Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci. USA* 99, 9905-9912.

Alverson, A.J., Wei, X.X., Rice, D.W., Stern, D.B., Barry, K. and Palmer, J.D. (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* 27, 1436-1448.

Alverson, A.J., Zhuo, S., Rice, D.W., Sloan, D.B. and Palmer, J.D. (2011) The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS One* 6, e16404.

Arimura, S. and Tsutsumi, N. (2002) A dynamin-like protein (ADL2b), rather than FtsZ, is involved in Arabidopsis mitochondrial division. *Proc. Natl. Acad. Sci. USA* 99,

5727-5731.

Bhargava, A., Shukla, S., Kumar, R. and Ohri, D. (2009) Metroglyph analysis of morphological variation in *Chenopodium* spp. *World J. Agr. Sci.* 5, 117-120.

Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J. (1996) Engineering GFP as a vital reporter in plants. *Curr. Biol.* 6, 325-330.

Cho, Y., Qiu, Y.L., Kuhlman, P. and Palmer, J.D. (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. USA* 95, 14244-14249.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guaniduim thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156-159.

Cusimano, N., Zhang, L.B. and Renner, S.S. (2008) Reevaluation of the *cox1* group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Mol. Biol. Evol.* 25, 265-276.

Doyle, J.J. and Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus* 12, 13-15.

Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., Leebens-Mack, J. and dePamphilis, C.W. (2010) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus, Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.* 10, 61.

Duchene, A.M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N.M., Zaepfel, M., Marechal-Drouard, L. and Small, I. (2005) Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 102, 16484-16489.

Duchene, A.M., Pujol, C. and Marechal-Drouard, L. (2009) Import of tRNAs and

aminoacyl-tRNA synthetases into mitochondria. *Curr. Genet.* 55, 1-18.

Gray, M.W. (1999) Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* 9, 678-687.

Hensel, R., Demharter, W. and Hilpert, R. (1989) The microflora involved in aerobic-thermophilic sludge stabilization. *Syst. Appl. Microbiol.* 11, 312-319.

Hugenholtz, P. and Stackebrandt, E. (2004) Reclassification of Sphaerobacter thermophilus from the subclass Sphaerobacteridae in the phylum Actinobacteria to the class Thermomicrobia (emended description) in the phylum Chloroflexi (emended description). *Int. J. Syst. Evol. Microbiol.* 54, 2049-2051.

Kubo, T., and Mikami, T. (2007) Organization and variation of angiosperm mitochondrial genome. *Physiol. Plant.* 129, 6-13.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A. and Mikami, T. (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$(GCA). *Nucl. Acids Res.* 28, 2571-2576.

Lipman, R.S. and Hou, Y.-M. (1998) Aminoacylation of tRNA in the evolution of an aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci. USA.* 95, 13495-13500.

Marechal-Drouard, L., Weil, J.-H. and Dietrich, A. (1993) Transfer RNAs and transfer RNA genes in plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 44, 13-32.

McClain, W.H. (1993) Rules that govern tRNA identity in protein synthesis. *J. Mol. Biol.* 234, 257-280.

Mikami, T., Kishima, Y., Sugiura, M. and Kinoshita, T. (1985) Organelle genome diversity in sugar-beet with normal and different sources of male sterile cytoplasms. *Theor. Appl. Genet.* 71, 166-171.

Peeters, N.M., Chapron, A., Giritch, G., Grandjean, O., Lancelin, D., Lhomme, T., Vivrel, A. and Small, I. (2000) Duplication and quadruplication of *Arabidopsis thaliana* cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. *J. Mol. Evol.* 50, 413-423.

Preuten, T., Cincu, E., Fuchs, J., Zoschke, R., Liere, K. and Borner, T. (2010) Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *Plant J.,* 64, 948-959.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

Sanchez-Puerta, M.V., Cho, Y., Mower, J.P., Alverson, A.J. and Palmer, J.D. (2008) Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. *Mol. Biol. Evol.* 25, 1762-1777.

Sang, T. (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit. Rev. Biochem. Mol. Biol.* 37, 121-147.

Satoh, M., Kubo, T., Nishizawa, S., Estiati, A., Itchoda, N. and Mikami, T. (2004) The cytoplasmic male-sterile type and normal type mitochondrial genomes of sugar beet share the same complement of genes of known function but differ in the content of expressed ORFs. *Mol. Genet. Genom.* 272, 247-256.

Singer-Sam, J., Robinson, M.O., Bellve, A.R., Simon, M.I. and Riggs, A.D. (1990) Measurement by quantitative PCR of changes in *HPRT*, *PGK-1*, *PGK-2*, *APRT*, *MTase*, and *Zfy* gene transcripts during mouse spermatogenesis. *Nucl. Acids Res.* 18, 1255-1259.

Varshney, U., Lee, C.-P. and Rajbhandary, U.L. (1991) Direct analysis of aminoacylation of tRNAs *in vivo*. *J. Biol. Chem.* 266, 24712-24718.

Wang, D.-Y., Zhang, Q., Liu, Y., Lin, Z.-F., Zhang, S.-X., Sun, M.-X. and Sodmergen (2010) The levels of male gametic mitochondrial DNA are highly regulated in angiosperms with regard to mitochondrial inheritance. *Plant Cell* 22, 2402-2416.

Woloszynska, M., Kmiec, B., Mackiewicz, P. and Janska, H. (2006) Copy number of bean mitochondrial genes estimated by real-time PCR does not correlate with the number of gene loci and transcript levels. *Plant Mol. Biol.* 61, 1-12.

Table 1 Mitochondrial genome sequences of some land plants and their tRNA^{Cys}(GCA) genes

Division	Scientific name	Common name	Accession no.	trnC1-GCA	trnC2-GCA	pt-trnC-GCA[1]
Magnoliophyta (Angiosperms)	*Arabidopsis thaliana*		Y08501	+	-	-
	Beta vulgaris	Sugar beet	BA000009	ψ[2]	+	-
	Brassica napus	Rapeseed	AP006444	+	-	-
	Carica papaya	Papaya	EU431224	+	-	-
	Citrullus lanatus	Watermelon	GQ856147	+	ψ	-
	Cucurbita pepo	Zucchini	GQ856148	+	-	+
	Nicotiana tabacum	Tobacco	BA000042	+	-	-
	Oryza rufipogon		AP011076	-	-	+
	Oryza sativa Indica Group	Rice	DQ167399	-	-	+
	Oryza sativa Japonica Group	Rice	BA000029	-	-	+
	Silene latifolia		HM562727	+	-	-
	Sorghum bicolor		DQ984518	-	-	+
	Tripsacum dactyloides	Gamagrass	DQ984517	-	-	+
	Triticum aestivum	Wheat	AP008982	-	-	+
	Vigna radiata	Mung bean	HM367685	+	+	-
	Vitis vinifera	Grapevine	FM179380	+	-	-
	Zea luxurians		DQ645537	-	-	+
	Zea mays subsp. mays	Maize	AY506529	-	-	+
	Zea mays subsp. parviglumis		DQ645539	-	-	+
	Zea perennis		DQ645538	-	-	+
Cycadophyta	*Cycas taitungensis*		AP009381	+	-	-

Marchantiophyta	*Marchantia polymorpha*	M68929	+	-	-
	Pleurozia purpurea	FJ999996	+	-	-
Bryophyta	*Physcomitrella patens*	AB251495	+	-	-
Anthocerotophyta	*Phaeoceros laevis*	GQ376531	+	-	-
	Megaceros aenigmaticus	EU660574	+	-	-
Charophyta	*Chaetosphaeridium globosum*	AF494279	+	-	-
	Chara vulgaris	AY267353	+	-	-
	Chlorokybus atmophyticus	EF463011	+	-	-
Streptophyta	*Mesostigma viride*	AF353999	+	-	-

[1]Plastid-like trnC-GCA; [2]Pseudo gene

Table 2 $2^{\Delta Ct}$ values of *cob*, *trnC1*-GCA and *trnC2*-GCA in some angiosperms estimated by qPCR assay

		$2^{\Delta Ct}$					
		cob		trnC1		trnC2	
Scientific name	Common name	Mean	SD[1]	Mean	SD	Mean	SD
Beta vulgaris	Sugar beet	61.130	6.281	40.132	2.752	43.443	2.341
Spinacea oleracea	Spinach	22.706	0.111	19.429	0.381	14.271	0.070
Chenopodium album		16.574	0.812	9.353	0.321	6.612	0.097
Celosia cristata		70.524	0.691	35.302	2.422	27.578	0.946
Basella rubra		80.837	5.938	61.820	0.000	31.922	2.033
Mirabilis jalapa		42.518	0.208	58.892	0.289	-[2]	
Myrtillocactus geometrizans		66.949	0.000	47.015	0.461	31.451	0.308
Daphniphyllum macropodum		54.192	0.000	35.767	1.402	31.342	0.307
Rumex obtusifolius		91.458	0.897	51.628	0.759	-	
Cucumis sativus	Cucumber	35.384	0.347	-		27.097	0.398
Glycine max	Soybean	68.597	1.009	30.804	0.302	33.943	0.333

[1]Standard deviation; [2]Under detectable level

Figure legends

Figure 1 The cloverleaf structure deduced from the *trnC2*-GCA gene sequence in spinach. Nucleotide numbering follows that of McClain (1993). The differing nucleotides from spinach are boxed (See also Figure S1).

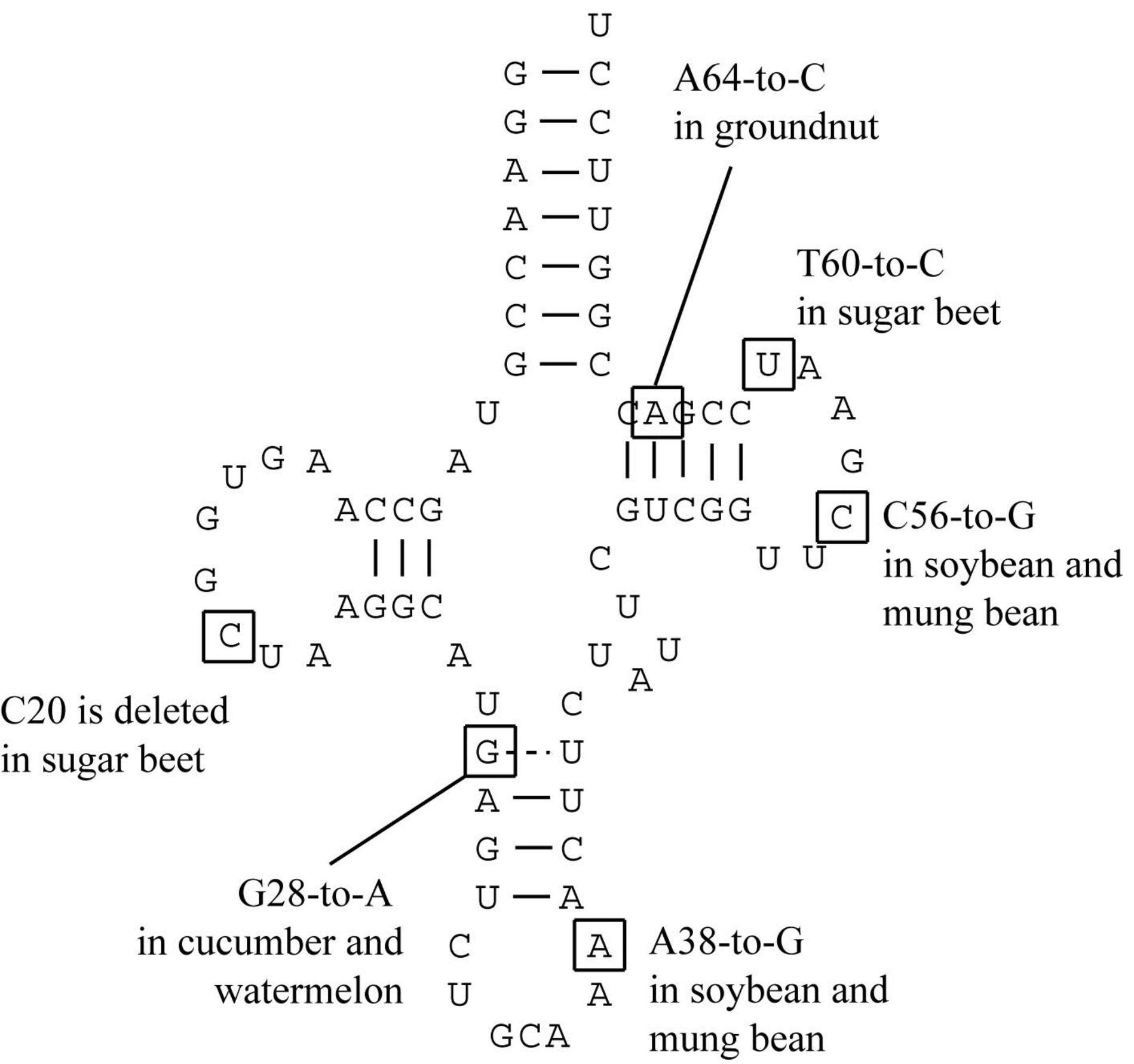
Figure 2 An image of the gel electrophoresis performed after PCR amplification of 26 plants. The amplicons were electrophoresed in 10% LongRanger gel. The gel was stained with ethidium bromide.

Figure 3 DNA gel blot analysis of the total cellular DNAs of *Hibbertia pedunculata* (Hp), spinach (So), *Daphniphyllum macropodum* (Dm) and cucumber (Cs). DNA samples were digested with *Hin*dIII enzyme. Blots were probed with the *trnC1*-GCA or the *trnC2*-GCA sequences (see text for probe information). Size markers are shown on the right (kbp).

Figure 4 Acid PAGE analysis of sugar beet (Bv), soybean (Gm), watermelon (Cl), cucumber (Cs) and spinach (So) total cellular RNA isolated under pH 4.5 or pH 9.5. Blots were probed with the *trnC2*-GCA or *trnC1*-GCA probes.

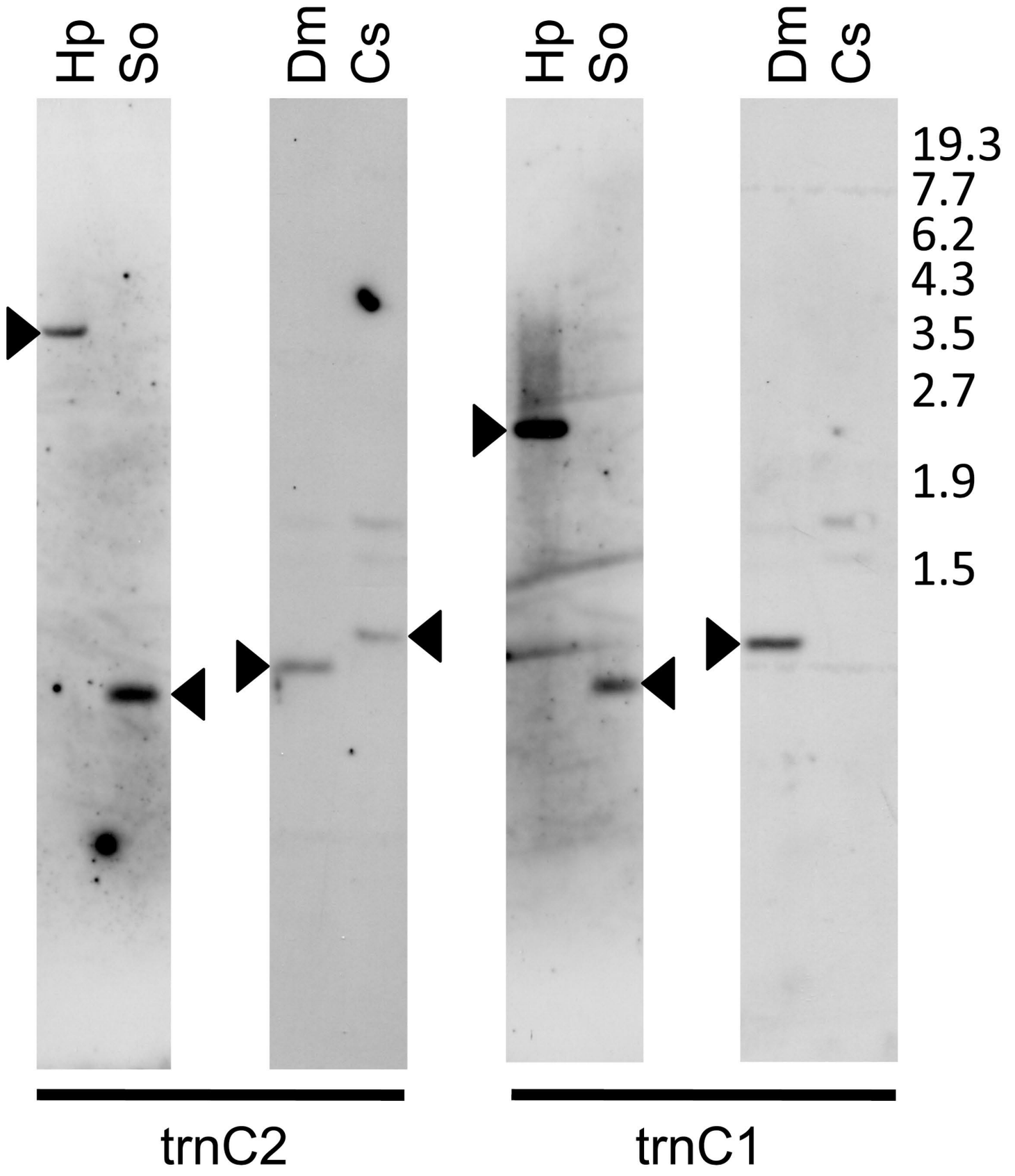
Figure 5 Organization and expression analysis of *bvcysRS*. **a** Image of DNA gel blot analysis of sugar beet total DNA with a *bvcysRS* probe. Abbreviations of restriction endonucleases are, V for *Eco*RV, H for *Hin*dIII, and S for *Spe*I. Size markers are shown on the right. **b** Image of 2% agarose gel electrophoresis after RT-PCR of total cellular RNA from the flower buds (Fl) and leaves (Le) of sugar beet. Total DNA was used in the control experiment. Complementary DNA derived amplicons (0.3 kb), which can be distinguished from genomic DNA derived amplicons (0.5 kb, see Figure. S5), were obtained from flower buds and leaves.

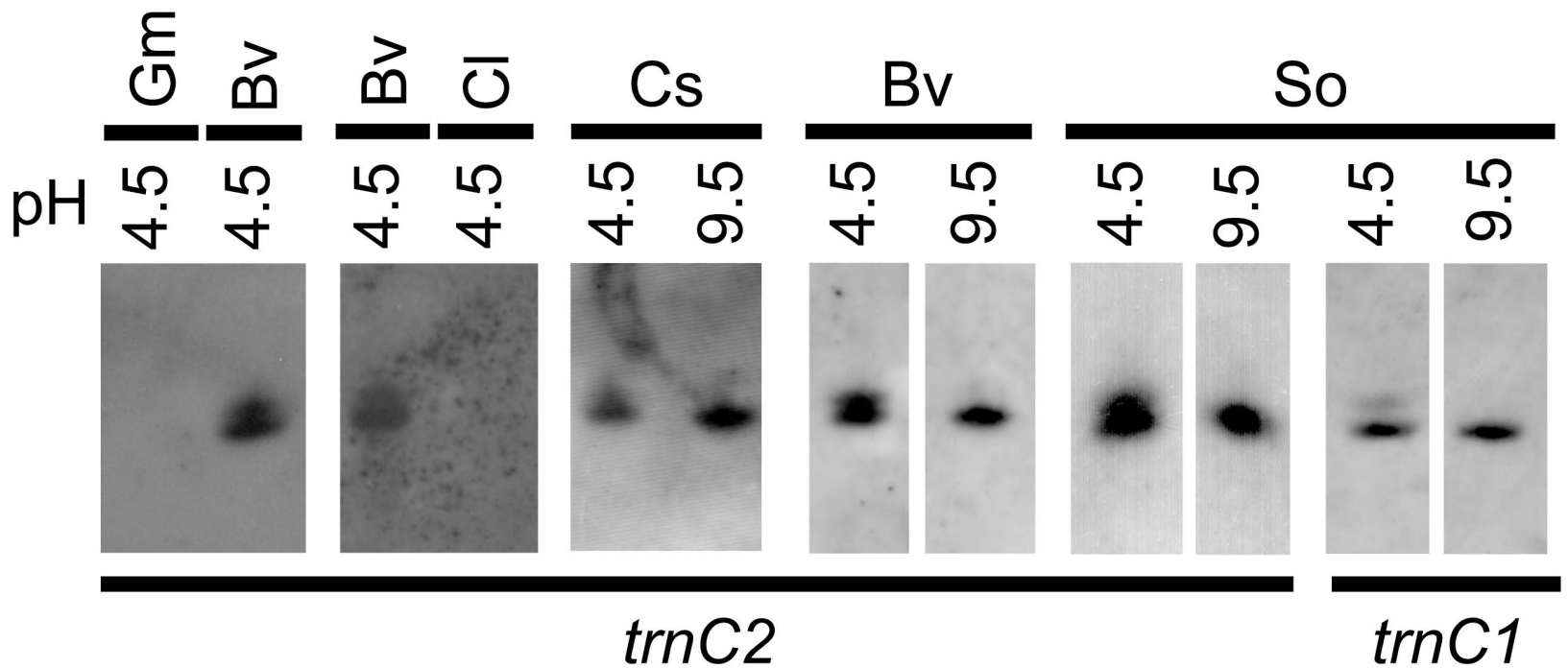
Figure 6 Transient expression of GFP or RFP fusion proteins in the epidermal cells of an onion bulb. Images of bvcysRS::GFP (panels a and d), MitoTracker-Orange staining (b), Pt-RFP (e) and pTH2 (g) are shown. Panels c and f are the merged images of a and b, and d and e, respectively. Scale bar is 10 μm in panels a to f, and 50 μm in panel g.

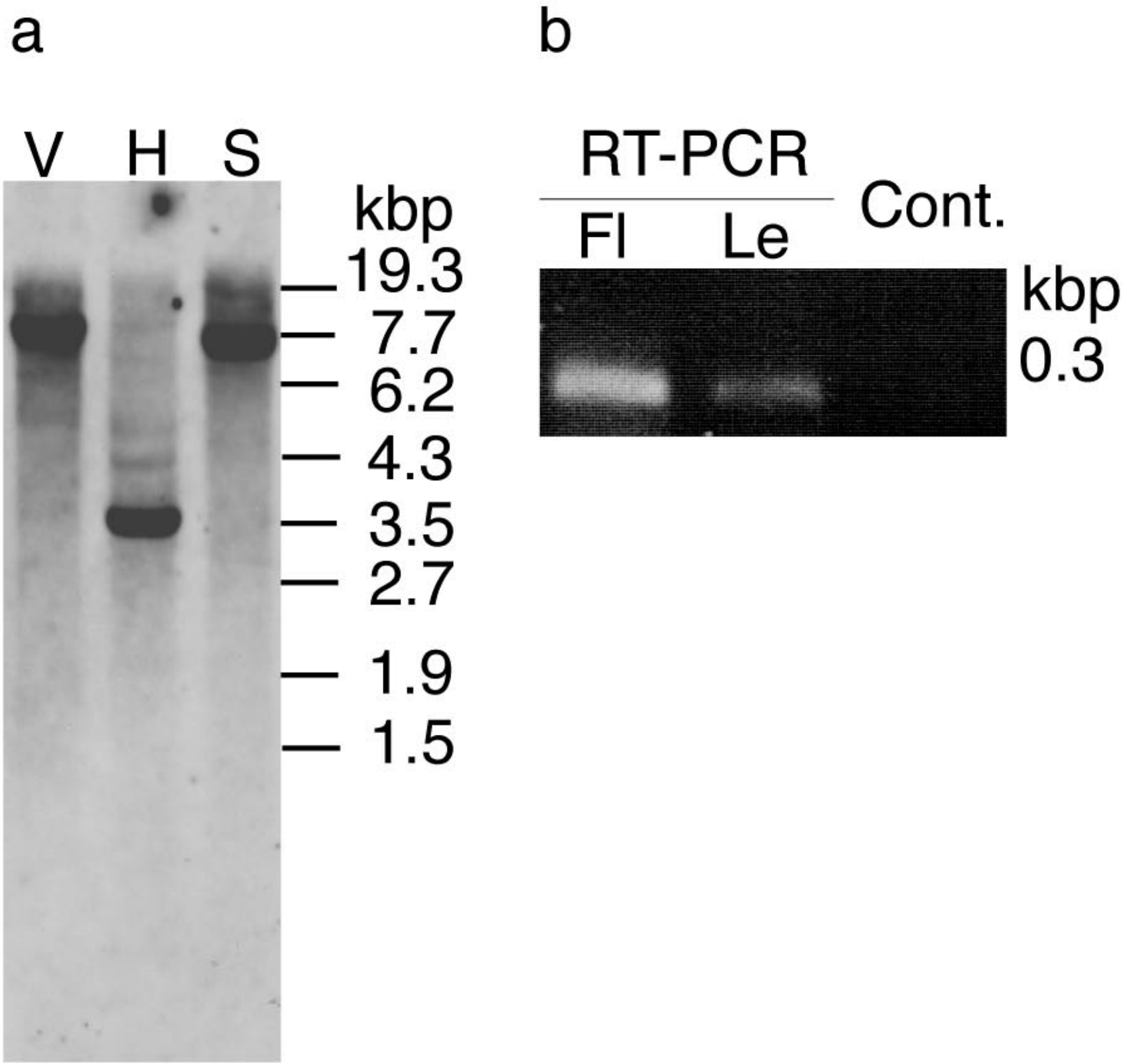

A64-to-C
in groundnut
T60-to-C
in sugar beet
C56-to-G
in soybean and
mung bean
C20 is deleted
in sugar beet
G28-to-A
in cucumber and
watermelon
A38-to-G
in soybean and
mung bean

trnC2-GCA (~73 bp)

Plant species	Family	Order
Lysichiton camtschatcense Schott	Araceae	Alismatales
Colocasia esculenta (L.) Schott	Araceae	Alismatales
Petasites japonicus (sieb. et Zucc.) Maxim.	Asteraceae	Asterales
Mentha spicata L.	Lamiaceae	Lamiales
Forsythia suspensa (Thunb.) Vahl	Oleaceae	Lamiales
Antirrhinum majus L.	Scrophulariaceae	Lamiales
Ipomoea batatas L.	Convolvulaceae	Solanales
Capsicum annuum L.	Solanaceae	Solanales
Solanum tuberosum L.	Solanaceae	Solanales
Cucumis sativus L.	Cucurbitaceae	Cucurbitales
Daphniphyllum macropodum Miq.	Daphniphyllaceae	Saxifragales
Dioscorea japonica Thumb.	Dioscoreaceae	Liliales
Allium cepa L.	Liliaceae	Liliales
Allium fistulosum L.	Liliaceae	Liliales
Magnolia kobus DC.	Magnoliaceae	Magnoliales
Nelumbo nucifera Gaertn.	Nelumbonaceae	Nymphaeales
Hibbertia pedunculata R.Br. ex DC.	Dilleniaceae	Dilleniales
Cryptotaenia japonica Hassk.	Apiaceae	Apiales
Daucus carota L.	Apiaceae	Apiales
Petroselinum crispum (Mill.) Nyman ex A. W. Hill	Apiaceae	Apiales
Medicago sativa L.	Fabaceae	Fabales
Trifolium pratense L.	Fabaceae	Fabales
Skimmia japonica Thurb.	Rutaceae	Sapindales
Malus domestica Borkh.	Rosaceae	Rosales
Beta vulgaris L.	Amaranthaceae	Caryophyllales
Basella rubra L.	Basellaceae	Caryophyllales


Hp
So
Dm
Cs
Hp
So
Dm
Cs
19.3
7.7
6.2
4.3
3.5
2.7
1.9
1.5
trnC2
trnC1


Gm
Bv
Bv
Cl
Cs
Bv
So
pH
4.5
4.5
4.5
4.5
4.5
9.5
4.5
9.5
4.5
9.5
4.5
9.5
trnC2
trnC1


a
V
H
S
kbp
19.3
7.7
6.2
4.3
3.5
2.7
1.9
1.5
b
RT-PCR
Fl
Le
Cont.
kbp
0.3

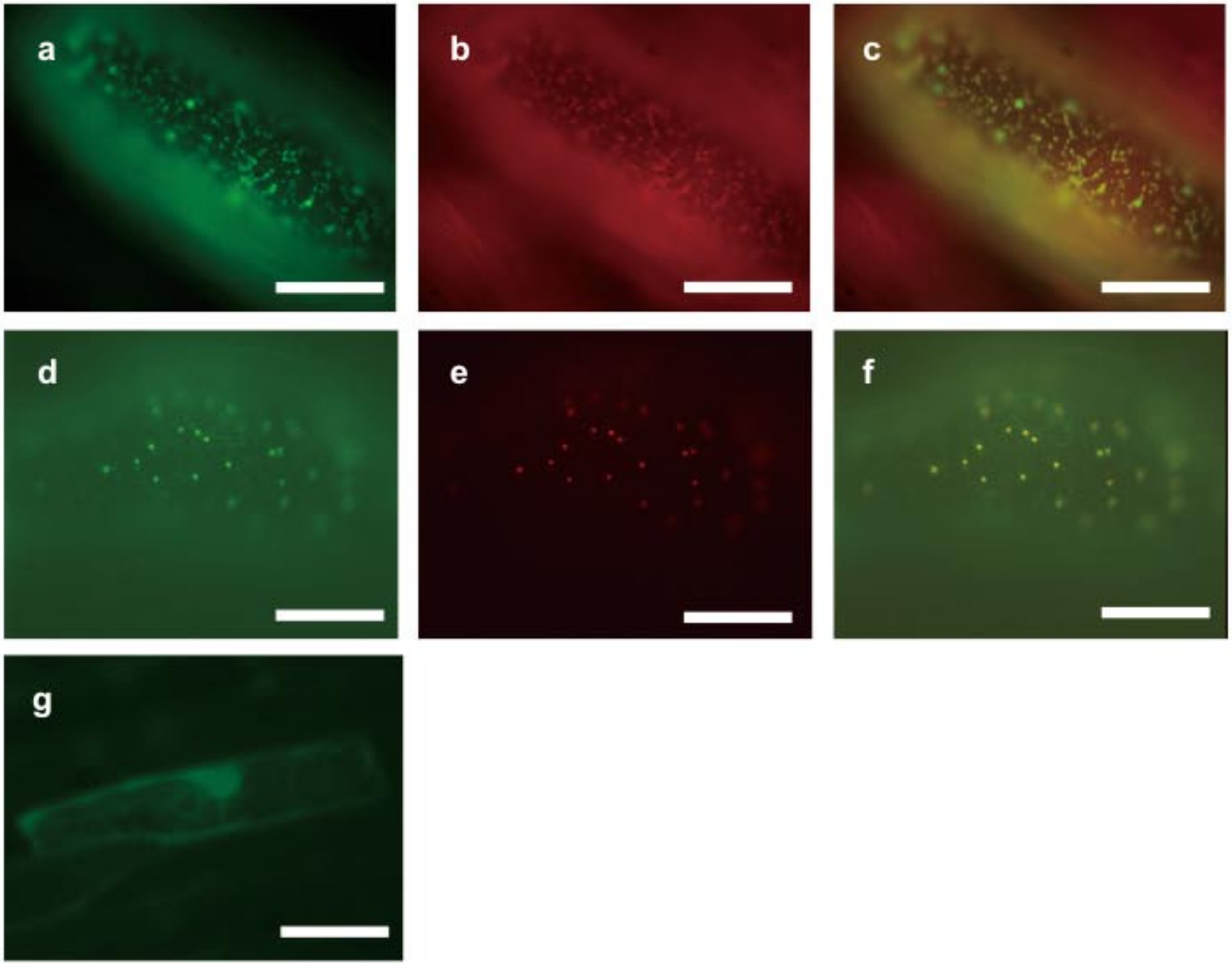

a
b
c
d
e
f
g

Table S1 List of the best matched entries to the transcripts mapped onto scaffold_04063

Table S2 List of some authentic cucumber mtDNA sequences used in this study

Table S3 List of best matched entries to the transcripts that have been mapped near two *trnC2*-GCA copies in scaffold_23

Table S4 List of primers used in this study

Table S5 Summary of qPCR

Table S6 Summary of web URLs consulted in this study

Figure S1 Comparison of *trnC2*-GCA sequences of sugar beet (DDBJ/EMBL/GenBank accession no. BA000009), spinach (this study), cucumber (Phytozome, scaffold_04063), watermelon (DDBJ/EMBL/GenBank, GQ856147), two copies of *Aquilegia coerulea* (Phytozome, scaffold_23), soybean (Phytozome, scaffold_682), groundnut (DDBJ/EMBL/GenBank,EZ742316) and mung bean (DDBJ/EMBL/GenBank, HM367685). Dashes are incorporated for the maximum matching. Nucleotide sequence of partial *trnC2*-GCA-coding regions of *Basella rubra* (this study), *Daphniphyllum macropodum* (this study) and *Hibbertia pedunculata* (this study) are also aligned. Nucleotide residues are coordinated to the source entries.

Figure S2 Nucleotide sequence of EZ742316, an EST entry of ground nut. *trnC2*-GCA sequence is underlined. Homologous sequences to plant mitochondrial DNA are boxed.

Figure S3 Nucleotide sequence of scaffold_682, a genomic sequence taken from

Phytozome. *trnC2*-GCA sequence is underlined. Homologous sequences to plant mitochondrial DNA or other entries are boxed.

Figure S4 Nucleotide sequence of the 5504-bp sequence containing spinach *trnC2*-GCA. *trnC2*-GCA sequence is underlined. Homologous sequences to plant mitochondrial DNA are boxed.

Figure S5 Nucleotide sequence of sugar-beet gene coding for cysteinyl-tRNA synthetase (*bvcysRS*). Exons and introns are indicated by upper and lower cases, respectively. Putative translation product is shown below in single letter code. Positions of oligonucleotides used for 5' RACE and RT-PCR are underlined.

Figure S6 Multiple alignments of the amino acid sequences from cysteinyl-tRNA synthetases of sugar beet (bv), *Arabidopsis* (at) (AC006593, also annotated as at2g31170), cucumber (Csa002320 in CuGenDB), rice (Os09g0556500 in RAP) and *E. coli* (ec) (X56234) (**a**). Dashes are incorporated for maximum matching. Two conserved motifs (His-Ile-Gly-His and Lys-Met-Ser-Lys-Ser) that are characteristic of class I aaRS are doubly and singly underlined, respectively. The positions of introns are indicated by slash. Length of the introns are summarized below (in bp) (**b**).

Table S1 List of the best matched entries to the transcripts mapped onto scaffold_04063

Name of transcripts	Best matched entries				
	Accession no.	Description	Nucleotide coordinate	Note	Expect
Cucsa.392380	GQ856147	Citrullus lanatus mitochondrion, complete genome	113506-112522	atp1	0.0
Cucsa.392390	GQ856147	Citrullus lanatus mitochondrion, complete genome	189302-189135	nad1-intron 4	1e-71
Cucsa.392400	AY258277	Cucumis sativus Calypso NADH dehydrogenase subunit 5 (nad5) gene, partial cds; mitochondrial gene for mitochondrial product	6487-6561	nad5-intron 1	3e-21
Cucsa.392410	-[1]	-	-	-	-
Cucsa.392420	GQ220326	Vitis vinifera strain PN40024 mitochondrion, partial genome	798-655	Intergenic region	2e-42
Cucsa.392430	AF288043	Cucumis sativus cultivar Calypso atp9 and atp6 genes, complete sequence; mitochondrial genes for mitochondrial products	4511-3797	atp6	0.0
Cucsa.392440	AF288043	Cucumis sativus cultivar Calypso atp9 and atp6 genes, complete sequence; mitochondrial genes for mitochondrial products	8643-8369	Intergenic region	8e-129
Cucsa.392450	GQ856147	Citrullus lanatus mitochondrion, complete genome	121121-121440	trnC2-GCA	1e-100
Cucsa.392460	-	-	-	-	-
Cucsa.392470	GQ856147	Citrullus lanatus mitochondrion, complete genome	331394-331208	Intergenic region	7e-56
Cucsa.392480	FJ007641	Cucumis sativus NADH dehydrogenase subunit 1 (nad1) gene, exons 2, 3 and partial cds; mitochondrial	1530-1778	nad1-intron 1	2e-123
Cucsa.392490	GQ856147	Citrullus lanatus mitochondrion, complete genome	334756-334304	nad1-intron 3	2e-170
Cucsa.392500	AF290299	Cucumis sativus clone V69 mitochondrial genomic sequence	111-242	This transcript also contains plastid-like sequence which is homologous to cemA.	3e-58
Cucsa.392510	AB586273	Coccinia sp. SH-2010 chloroplast gene for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds, isolate: T059	955-1305	Plastid-like sequence which is homologous to rbcL	8e-156
Cucsa.392520	-	-	-	-	-
Cucsa.392530	-	-	-	-	-
Cucsa.392540	AF290296	Cucumis sativus clone U50F mitochondrial genomic sequence	244-187	Intergenic region	2e-19

[1]No entry was hit under the threshold value (1e−10).

Table S2 List of some authentic cucumber mtDNA sequences used in this study

Accession no.	Homology to scaffold_04063	Description
AF282389	+[1]	Cucumis sativus clone A10 mitochondrial genomic sequence
AF282390	+	Cucumis sativus clone A43 mitochondrial genomic sequence
AF282391	+	Cucumis sativus clone B99 mitochondrial genomic sequence
AF282392	+	Cucumis sativus clone C114 mitochondrial genomic sequence
AF282393	+	Cucumis sativus clone F16 mitochondrial genomic sequence
AF282394	+	Cucumis sativus clone G102 mitochondrial genomic sequence
AF282395	+	Cucumis sativus clone I51 mitochondrial genomic sequence
AF282396	+	Cucumis sativus clone J7 mitochondrial genomic sequence
AF282397	+	Cucumis sativus clone K34 mitochondrial genomic sequence
AF282398	+	Cucumis sativus clone M102 mitochondrial genomic sequence
AF282399	+	Cucumis sativus clone R64 mitochondrial genomic sequence
AF282400	+	Cucumis sativus clone S100 mitochondrial genomic sequence
AF282401	+	Cucumis sativus clone T106 mitochondrial genomic sequence
AF282402	+	Cucumis sativus clone U38 mitochondrial genomic sequence
AF282403	+	Cucumis sativus clone U63 mitochondrial genomic sequence
AF288043	+	Cucumis sativus cultivar Calypso atp9 and atp6 genes, complete sequence; mitochondrial genes for mitochondrial products
AF288044	+	Cucumis sativus cultivar Calypso apocytochrome b (cob) gene, complete cds; and tRNA-His (trnH) and tRNA-Thr (trnT) genes, complete sequence; mitochondrial genes for mitochondrial products
AF290215	+	Cucumis sativus clone B105F mitochondrial genomic sequence
AF290216	+	Cucumis sativus clone B105R mitochondrial genomic sequence
AF290217	-[2]	Cucumis sativus clone B10F mitochondrial genomic sequence
AF290218	-	Cucumis sativus clone B10R mitochondrial genomic sequence
AF290219	-	Cucumis sativus clone B68F mitochondrial genomic sequence
AF290220	+	Cucumis sativus clone B68R mitochondrial genomic sequence
AF290221	+	Cucumis sativus clone B80F mitochondrial genomic sequence
AF290222	+	Cucumis sativus clone B86F mitochondrial genomic sequence
AF290223	-	Cucumis sativus clone B86R mitochondrial genomic sequence
AF290224	-	Cucumis sativus clone C107 mitochondrial genomic sequence
AF290225	+	Cucumis sativus mitochondrial genomic sequence
AF290226	-	Cucumis sativus clone G49 mitochondrial genomic sequence
AF290227	+	Cucumis sativus clone I107F mitochondrial genomic sequence
AF290228	+	Cucumis sativus clone I3R mitochondrial genomic sequence
AF290229	+	Cucumis sativus clone I72F mitochondrial genomic sequence
AF290230	+	Cucumis sativus clone L1-1F20R mitochondrial genomic sequence
AF290231	-	Cucumis sativus clone L1-1F5F mitochondrial genomic sequence
AF290232	+	Cucumis sativus clone L1-1G23F mitochondrial genomic sequence
AF290233	+	Cucumis sativus clone L1-1G23R mitochondrial genomic sequence
AF290234	+	Cucumis sativus clone L1-1J6F mitochondrial genomic sequence
AF290235	-	Cucumis sativus clone L1-1J6R mitochondrial genomic sequence
AF290236	+	Cucumis sativus clone L1-1K8F mitochondrial genomic sequence
AF290237	-	Cucumis sativus clone L1-1K8R mitochondrial genomic sequence
AF290238	+	Cucumis sativus clone L1-1M17F mitochondrial genomic sequence
AF290239	+	Cucumis sativus clone L1-1M17R mitochondrial genomic sequence
AF290240	-	Cucumis sativus clone L1-2C16F mitochondrial genomic sequence
AF290241	-	Cucumis sativus clone L1-2C16R mitochondrial genomic sequence
AF290242	-	Cucumis sativus clone L1-2D5F mitochondrial genomic sequence
AF290243	+	Cucumis sativus clone L1-2D5R mitochondrial genomic sequence
AF290244	-	Cucumis sativus clone L1-2D8R mitochondrial genomic sequence
AF290245	-	Cucumis sativus clone L1-2E14R mitochondrial genomic sequence
AF290246	+	Cucumis sativus clone L1-2F3F mitochondrial genomic sequence
AF290247	-	Cucumis sativus clone L1-2F3R mitochondrial genomic sequence

AF290248	-	Cucumis sativus clone L1-2F6F mitochondrial genomic sequence
AF290249	-	Cucumis sativus clone L1-2F6R mitochondrial genomic sequence
AF290250	-	Cucumis sativus clone L1-2K6R mitochondrial genomic sequence
AF290251	+	Cucumis sativus clone L2-1A9F mitochondrial genomic sequence
AF290252	-	Cucumis sativus clone L2-1A9R mitochondrial genomic sequence
AF290253	+	Cucumis sativus clone L2-1C8F mitochondrial genomic sequence
AF290254	-	Cucumis sativus clone L2-1C8R mitochondrial genomic sequence
AF290255	-	Cucumis sativus clone L2-1K18F mitochondrial genomic sequence
AF290256	-	Cucumis sativus clone L2-1K18R mitochondrial genomic sequence
AF290258	-	Cucumis sativus clone L2-1L11R mitochondrial genomic sequence
AF290259	-	Cucumis sativus clone L2-1M17R mitochondrial genomic sequence
AF290260	-	Cucumis sativus mitochondrial genomic sequence
AF290261	-	Cucumis sativus clone L2-2D8F mitochondrial genomic sequence
AF290262	+	Cucumis sativus clone L2-2E14F mitochondrial genomic sequence
AF290263	+	Cucumis sativus clone L2-2G11F mitochondrial genomic sequence
AF290264	-	Cucumis sativus clone L2-2J8F mitochondrial genomic sequence
AF290265	+	Cucumis sativus clone L2-2J8R mitochondrial genomic sequence
AF290266	+	Cucumis sativus clone L2-2K6F mitochondrial genomic sequence
AF290267	-	Cucumis sativus clone L2-2K6R mitochondrial genomic sequence
AF290268	+	Cucumis sativus clone L2-2N10F mitochondrial genomic sequence
AF290269	+	Cucumis sativus clone L2-2N10R mitochondrial genomic sequence
AF290270	+	Cucumis sativus clone L2-3O16R mitochondrial genomic sequence
AF290271	-	Cucumis sativus clone L2-3C14F mitochondrial genomic sequence
AF290272	+	Cucumis sativus clone L2-3C14R mitochondrial genomic sequence
AF290273	+	Cucumis sativus clone L2-3C23F mitochondrial genomic sequence
AF290274	+	Cucumis sativus clone L2-3C6F mitochondrial genomic sequence
AF290275	-	Cucumis sativus clone L2-3C6R mitochondrial genomic sequence
AF290276	-	Cucumis sativus clone L2-3C7F mitochondrial genomic sequence
AF290277	+	Cucumis sativus clone L2-3G19F mitochondrial genomic sequence
AF290278	-	Cucumis sativus clone L2-3G19R mitochondrial genomic sequence
AF290279	+	Cucumis sativus clone L2-3G6F mitochondrial genomic sequence
AF290280	+	Cucumis sativus clone L2-3G6R mitochondrial genomic sequence
AF290281	-	Cucumis sativus clone L2-3I17F mitochondrial genomic sequence
AF290282	-	Cucumis sativus clone L2-3I17R mitochondrial genomic sequence
AF290283	-	Cucumis sativus clone L2-3O16F mitochondrial genomic sequence
AF290284	-	Cucumis sativus clone N43F mitochondrial genomic sequence
AF290285	+	Cucumis sativus clone O69F mitochondrial genomic sequence
AF290286	+	Cucumis sativus clone O69R mitochondrial genomic sequence
AF290287	+	Cucumis sativus clone O71F mitochondrial genomic sequence
AF290288	+	Cucumis sativus clone O71R mitochondrial genomic sequence
AF290289	-	Cucumis sativus clone P14F mitochondrial genomic sequence
AF290290	+	Cucumis sativus clone P1FR mitochondrial genomic sequence
AF290291	+	Cucumis sativus clone R36F mitochondrial genomic sequence
AF290292	+	Cucumis sativus clone R36R mitochondrial genomic sequence
AF290293	+	Cucumis sativus clone T112F mitochondrial genomic sequence
AF290294	+	Cucumis sativus mitochondrial genomic sequence
AF290295	+	Cucumis sativus clone T71F mitochondrial genomic sequence
AF290296	+	Cucumis sativus clone U50F mitochondrial genomic sequence
AF290297	+	Cucumis sativus clone U50R mitochondrial genomic sequence
AF290298	-	Cucumis sativus clone V46F mitochondrial genomic sequence
AF290299	+	Cucumis sativus clone V69 mitochondrial genomic sequence
AF290300	+	Cucumis sativus clone V94R mitochondrial genomic sequence
AF291430	+	Cucumis sativus clone B21F mitochondrial genomic sequence
AF291431	+	Cucumis sativus clone B21R mitochondrial genomic sequence
AF291432	-	Cucumis sativus clone G9F mitochondrial genomic sequence

AF291433	+	Cucumis sativus clone G9R mitochondrial genomic sequence
AF291434	+	Cucumis sativus clone I7F mitochondrial genomic sequence
AF291435	+	Cucumis sativus clone I7R mitochondrial genomic sequence
AY258270	+	Cucumis sativus MSC16 mitochondrial ribosomal protein L5 (rpl5) gene, complete cds; and JLV5 deletion front junction sequence
AY258271	+	Cucumis sativus MSC16 rearranged 18S ribosomal RNA (rrn18) and 5S ribosomal RNA (rrn5) genes, complete sequence; and NADH dehydrogenase subunit 5 (nad5) gene, partial cds; mitochondrial genes for mitochondrial products
AY258272	+	Cucumis sativus MSC16 mitochondrial tRNA-Thr (trnT) gene, complete sequence; and JLV5 deletion rear junction sequence
AY258273	+	Cucumis sativus MSC16 ribosomal protein L5 (rpl5) gene, complete cds; mitochondrial gene for mitochondrial product
AY258274	+	Cucumis sativus Calypso ribosomal protein L5 (rpl5) and ribosomal protein S14 (rps14) pseudogenes, complete sequence
AY258275	+	Cucumis sativus Calypso ribosomal protein L5 (rpl5) gene, complete cds; mitochondrial gene for mitochondrial product
AY258277	+	Cucumis sativus Calypso NADH dehydrogenase subunit 5 (nad5) gene, partial cds; mitochondrial gene for mitochondrial product
AY258278	+	Cucumis sativus Calypso 18S ribosomal RNA (rrn18) and 5S ribosomal RNA (rrn5) genes, complete sequence; mitochondrial genes for mitochondrial products
AY357206	+	Cucumis sativus var. sikkimensis mitochondrial 5S ribosomal RNA (rrn5) gene, partial sequence; rrn5-rrn18 intergenic spacer, complete sequence; and 18S ribosomal RNA (rrn18) gene, partial sequence
FJ007641	+	Cucumis sativus NADH dehydrogenase subunit 1 (nad1) gene, exons 2, 3 and partial cds; mitochondrial
FJ007642	+	Cucumis sativus NADH dehydrogenase subunit 4 (nad4) gene, exons 3, 4 and partial cds; mitochondrial
FJ007643	+	Cucumis sativus NADH dehydrogenase subunit 7 (nad7) gene, exons 1 through 5 and partial cds; mitochondrial

[1]Expexted value is less than 1e-10.

[2]Expected value is more than 1e-10.

Table S3 List of best matched entries to the transcripts that have been mapped near two *trnC2*-GCA copies

Name of transcripts	Best matched entries				
	Accession no.	Description	Nucleotide coordinate	Note	Expect
AcoGoldSmith_v1.017308m.g	GQ220324	Vitis vinifera strain PN40024 mitochondrion, partial genome	60189-59722	Intergenic region	0.0
AcoGoldSmith_v1.023702m.g	EF470527	Brassica juncea var. tumida clone pVT-mTM-9 unknown gene; mitochondrial	500-820	Intergenic region	1e-61
AcoGoldSmith_v1.014537m.g	FM179380	Vitis vinifera complete mitochondrial genome, cultivar Pinot noir clone ENTAV115	252094-252318	Intergenic region	2e-67
AcoGoldSmith_v1.026523m.g	GQ220323	Vitis vinifera strain PN40024 mitochondrion, partial genome	289869-290328	nad4-exon 3	0.0
AcoGoldSmith_v1.019244m.g	GQ220323	Vitis vinifera strain PN40024 mitochondrion, partial genome	284871-285455	nad4-exon 2	0.0
AcoGoldSmith_v1.017727m.g	GQ220323	Vitis vinifera strain PN40024 mitochondrion, partial genome	283049-283310	nad4-exon 1	3e-119
AcoGoldSmith_v1.012284m.g	GQ220323	Vitis vinifera strain PN40024 mitochondrion, partial genome	281509-282782	Intergenic region	0.0
AcoGoldSmith_v1.024181m.g	DQ381455	Beta vulgaris subsp. vulgaris NADH dehydrogenase subunit 2 (nad2) mRNA, complete cds; mitochondrial	554-1468	nad2-exon 3, 4, 5	0.0
AcoGoldSmith_v1.025383m.g	BA000042	Nicotiana tabacum mitochondrial DNA, complete genome	228569-228745	Overlapping with an ORF annotated as orf138c	2e-78
AcoGoldSmith_v1.017450m.g	_[1]	-	-	-	-
AcoGoldSmith_v1.018496m.g	GQ220325	Vitis vinifera strain PN40024 mitochondrion, partial genome	93539-92903	tatC	0.0
AcoGoldSmith_v1.026004m.g	AC007729	Arabidopsis thaliana chromosome 2 BAC T18C6 genomic sequence, complete sequence	47499-47377	The matched sequence is integrated Arabidopsis mtDNA copy.	7e-48
AcoGoldSmith_v1.013489m.g	AJ965437	Helianthus annuus mitochondrial nad9 gene, tRNA-Pro gene and tRNA-Trp gene, clone pstI-14.5	2393-3004	nad9	0.0
AcoGoldSmith_v1.023258m.g	XM_0022795	PREDICTED: Vitis vinifera hypothetical protein LOC100252457 (LOC100252457), mRNA	1-1268	O-fucosyltransferase	0.0
AcoGoldSmith_v1.013662m.g	GQ220326	Vitis vinifera strain PN40024 mitochondrion, partial genome	30640-30816	Intergenic region	2e-18
AcoGoldSmith_v1.012204m.g	GQ856148	Cucurbita pepo mitochondrion, complete genome	721361-720741	Intergenic region	8e-144
AcoGoldSmith_v1.025384m.g	AY832036	Eschscholzia californica cytochrome c biogenesis ccmB (ccmB) gene, partial cds; mitochondrial	37-454	ccmB	0.0
AcoGoldSmith_v1.027805m.g	AF319171	Magnolia x soulangeana ribosomal protein S19 (rps19) and ribosomal protein S3 (rps3) genes, complete cds; and ribosomal protein L16 (rpl16) gene, partial cds; mitochondrial genes for mitochondrial products	3204-3774	rps3	0.0
AcoGoldSmith_v1.014263m.g	EU431224	Carica papaya mitochondrion, complete genome	121059-120257	atp6	0.0
AcoGoldSmith_v1.025975m.g	GQ856148	Cucurbita pepo mitochondrion, complete genome	140509-139469	Intergenic region	0.0
AcoGoldSmith_v1.027445m.g	BA000042	Nicotiana tabacum mitochondrial DNA, complete genome	377993-377310	nad7-exon 2	0.0
AcoGoldSmith_v1.020333m.g	GQ220324	Vitis vinifera strain PN40024 mitochondrion, partial genome	95936-96226	nad7-exon 1	2e-144
AcoGoldSmith_v1.026461m.g	GQ856147	Citrullus lanatus mitochondrion, complete genome	279455-279901	rps7	0.0
AcoGoldSmith_v1.024016m.g	GQ856147	Citrullus lanatus mitochondrion, complete genome	363569-362931	ccmC	0.0
AcoGoldSmith_v1.014359m.g	AP011077	Oryza sativa Indica Group mitochondrial DNA, complete genome, cultivar: Lead rice	276675-276816	atp1	1e-62
AcoGoldSmith_v1.024461m.g	GQ220326	Vitis vinifera strain PN40024 mitochondrion, partial genome	9139-10720	cox1	0.0
AcoGoldSmith_v1.021082m.g	GQ220323	Vitis vinifera strain PN40024 mitochondrion, partial genome	18240-18719	atp8	0.0
AcoGoldSmith_v1.025501m.g	Z68127	M.grandiflora mitochondrial cox3 gene	258-798	cox3	0.0
AcoGoldSmith_v1.012221m.g	EU431224	Carica papaya mitochondrion, complete genome	321619-321884	rpl5	2e-110

[1]No entry is hit under the threshold value (1e−10).

Table S4 List of primers used in this study

Name of primers	Nucleotid sequence
1	5'-GGAACCGTAGCCAAGTGG-3'
2	5'-AGGAACCGGTCGGGTTCGAAC-3'
3	5'-CAGGTAGGGAGAACTCT-3'
4	5'-TTTGAACCTTTGATCCG-3'
5	5'-TCTTCTCTGGTAACCCG-3'
6	5'-TTGGAAGTTCCAGAATG-3'
RT1	5'-ATGCACCCAGTAGC-3'
A1	5'-CTTTCTCGAATCAATAGCTACC-3'
A2	5'-CTCACCAGCTCGATTGTCCTC-3'
Bv1	5'-TGACAAGATAATTGCCAGAG-3'
S1	5'-GGCATATTGAATGCAGTGCTATG-3'
S2	5'-CAGAGCTGTGCTGCATGTTGTG-3'
F	5'-GACGTCTGTTATGTTCG-3'
cys1	5'-ATTTTGCATCTAAAGATTATCATCC-3'
cys2	5'-CTTTCTCGAATCAATAGCTACC-3'
cys3	5'-TCTCCAATAAATTACTCTGA-3'
cys4	5'-GTATGAAGATCATCTGACATCG-3'
cys5	5'-CTCTACATCAGCTTAGAGAG-3'
cys6	5'-TCCATCAGAGCAATTCCCAC-3'
cys7	5'-CTCCAACCAATTCAAATATC-3'
akaz-nuc-FW	5'-TAATAGCACCTGCCCAGTCC-3'
akaz-nuc-RV	5'-CGTGGAAAGGTGGGTTTAGA-3'
beet-nuc-Fw	5'-CTTGTAAGGAGCAAAGGGTGCC-3'
beet-nuc-RV	5'-TCAACTGTCGCACTGGTATCCC-3'
cuc-nuc-FW	5'-GCCATTGTTGTGCTTCAAGA-3'
cuc-nuc-RV	5'-TGAAACCAGATCCAAGCTCA-3'
ezo-nuc-FW	5'-AAACTGCTTTTCGTGGTTGG-3'
ezo-nuc-RV	5'-TTCGGCTGCTTTACCTCAGT-3'
gish-nuc-FW	5'-ATTTGTGTTGTTGGCGATGA-3'
gish-nuc-RV	5'-AAATCTTGCCCACCTTTGTG-3'
keit-nuc-FW	5'-GGCAACAGAAGAAAGCAAGC-3'
keit-nuc-RV	5'-GAAGGTGTGCCCTTGATTGT-3'
oshi-nuc-FW	5'-TGCCCGAAAAATATCTGACC-3'
oshi-nuc-RV	5'-ATGTGGTTTTTGGTGGTGGT-3'
sabo-nuc-FW	5'-CAGGATACTGCATGGGAGGT-3'
sabo-nuc-RV	5'-AACACCAGGGACAAGGACAG-3'
soy-nuc-FW	5'-GCCCAAAAACATTCACCAAC-3'
soy-nuc-RV	5'-GAACACCGCGTTTGAGATTT-3'
spin-nuc-FW	5'-TGAAGCTCAGCATTTGATGG-3'

spin-nuc-RV	5'-GTGAGCCATTTTCCTTGAGC-3'
turu-nuc-FW	5'-TGGGAAGGTGGGTTTAGATG-3'
turu-nuc-RV	5'-ACAGCACCCTTCCAATCAAG-3'
cob-FW	5'-GTGCTATTGCGGCTACACCT-3'
cob-RV	5'-CCACCCCGCCTCATATTGTG-3'
trnC1-FW	5'-AATGGAAATGTATCGGACTGCAAA-3'
trnC1-RV	5'-AGGCCAAGGACGGGGTCG-3'
trnC2-FW	5'-AGTGGCTAAGGCATGAGTCTGCAAA-3'
trnC2-RV	5'-AGGAACCGGTCGGATTCGAA-3'
trnC2-bv-FW	5'-AGTGGTAAGGCATGAGTCTGCAAA-3'
trnC2-bv-RV	5'-AGGAACCGGTCGGGTTCGAA-3'
trnC2-cs-FW	5'-AGTGGCTAAGGCATAAGTCTGCAAA-3'
trnC2-gm-FW	5'-AGTGGCTAAGGCATGAGTCTGCAAG-3'
trnC2-gm-RV	5'-AGGAACCGGTCGGATTCCAA-3'
cob-seq-FW	5'-GTTGGTTGGGTAGCTTTTGC-3'
cob-seq-RV	5'-TAACAAATGGTGCCTCCACA-3'
32520-FW	5'-GTGCTTCAAGAATGGTGGGG-3'
32520-RV	5'-CTGCCACATCTGAAAACCCAGCAA-3'
EMB-FW	5'-CAGATTATGTGGGATGAAAGCC-3'
EMB-RV	5'-CATAGCCCAAAAATATGTTGTGC-3'
ERS1-FW	5'-GCTGTCGCTCTTTCTCATGC-3'
ERS1-RV	5'-AGAACTGGCCTTTACTTACGG-3'
Sa	5'-AAGGTCGACATGGTTGTGTTGGCAAGG-3'
Nc	5'-TCACCATGGATCCCACTTTTCCATCTAC-3'
Sa-rubisco-FW	5'-CCGTCGACATGGCCGCCGCAGTTTCCAC-3'
Nc-rubisco-RV	5'-CCCCATGGCGTACTGCCTAAGGCCTTGGC-3'

Table S5 Summary of qPCR

Scientific name	Common name	Target genes	Expected residence	Source of target sequence	Primers		Amplification efficiency
					Name	Temperature of annealing/reaction	
Beta vulgaris	Sugar beet	cob	Mitochondrion	DDBJ/EMBL/GenBank BA000009	cob-FW cob-RV	60.8	0.95
		trnC1-GCA	Mitochondrion	DDBJ/EMBL/GenBank BA000009	trnC1-FW trnC1-RV	58.8	1.00
		trnC2-GCA	Mitochondrion	DDBJ/EMBL/GenBank BA000009	trnC2-bv-FW trnC2-bv-RV	61.8	1.00
		Granule bound starch synthase	Nucleus	DDBJ/EMBL/GenBank BQ588728	beet-nuc-Fw beet-nuc-RV	60.8	1.01
Spinacea oleracea	Spinach	cob	Mitochondrion	This study	cob-FW cob-RV	60.8	0.96
		trnC1-GCA	Mitochondrion	-[1]	trnC1-FW trnC1-RV	58.8	1.01
		trnC2-GCA	Mitochondrion	This study	trnC2-FW trnC2-RV	58.8	1.04
		Orthologue of Arabidopsis at2g32520	Nucleus	This study	spin-nuc-FW spin-nuc-RV	60.8	1.02
Chenopodium album		cob	Mitochondrion	This study	cob-FW cob-RV	60.8	1.01
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	0.96
		trnC2-GCA	Mitochondrion	-	trnC2-FW trnC2-RV	58.8	0.94
		Orthologue of Arabidopsis at2g32520	Nucleus	This study	akaz-nuc-FW akaz-nuc-RV	60.8	0.97
Celosia cristata		cob	Mitochondrion	This study	cob-FW cob-RV	60.8	0.98
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	0.98
		trnC2-GCA	Mitochondrion	-	trnC2-FW trnC2-RV	58.8	0.93
		Orthologue of Arabidopsis at2g32520	Nucleus	This study	keit-nuc-FW keit-nuc-RV	60.8	0.92
Basella rubra		cob	Mitochondrion	This study	cob-FW cob-RV	60.8	0.96
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	0.97
		trnC2-GCA	Mitochondrion	-	trnC2-FW trnC2-RV	58.8	0.98
		Orthologue of Arabidopsis at2g32520	Nucleus	This study	turu-nuc-FW turu-nuc-RV	60.8	0.99
Mirabilis jalapa		cob	Mitochondrion	This study	cob-FW cob-RV	60.8	1.00
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	1.04

		trnC2-GCA	Mitochondrion	-	trnC2-FW trnC2-RV	58.8	NA[2]
		Ubiquitin ligase (XB3)	Nucleus	DDBJ/EMBL/GenBank EF470291	oshi-nuc-FW oshi-nuc-RV	60.8	0.99
Myrtillocactus geometrizans		cob	Mitochondrion	This study	cob-FW cob-RV	60.8	0.99
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	0.96
		trnC2-GCA	Mitochondrion	-	trnC2-FW trnC2-RV	58.8	0.98
		Orthologue of Arabidopsis at2g32520	Nucleus	This study	sabo-nuc-FW sabo-nuc-RV	60.8	0.98
Daphniphyllum macropodum		cob	Mitochondrion	This study	cob-FW cob-RV	60.8	1.02
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	0.98
		trnC2-GCA	Mitochondrion	-	trnC2-FW trnC2-RV	58.8	0.94
		Orthologue of EMB2765 (DDBJ/EMBL/GenBank FJ669873)	Nucleus	This study (resequenced using Japanese sample)	ezo-nuc-FW ezo-nuc-RV	60.8	1.03
Rumex obtusifolius		cob	Mitochondrion	This study	cob-FW cob-RV	60.8	0.98
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	0.94
		trnC2-GCA	Mitochondrion	-	trnC2-FW trnC2-RV	58.8	NA
		Orthologue of Ethylene response sensor (RP-ERS1) (DDBJ/EMBL/GenBank U63291)	Nucleus	This study (resequenced using Japanese sample)	gish-nuc-FW gish-nuc-RV	60.8	0.996
Cucumis sativus	Cucumber	cob	Mitochondrion	CuGenDB CU4910 and CU4436	cob-FW cob-RV	60.8	1.04
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	NA
		trnC2-GCA	Mitochondrion	Phytozome scaffold_04063	trnC2-cs-FW trnC2-RV	61.8	1.05
		Carboxymethylenebutenolidase	Nucleus	CuGenDB CU6306	cuc-nuc-FW cuc-nuc-RV	60.8	1.13
Glycine max	Soybean	cob	Mitochondrion	Phytozome Glyma07g12630	cob-FW cob-RV	60.8	0.93
		trnC1-GCA	Mitochondrion	-	trnC1-FW trnC1-RV	58.8	0.90
		trnC2-GCA	Mitochondrion	Phytozome scaffold_682	trnC2-gm-FW trnC2-gm-RV	61.8	0.95
		Seed specific protein	Nucleus	DDBJ/EMBL/GenBank FJ572966	soy-nuc-FW soy-nuc-RV	60.8	0.991

[1]No sequence data.

[2]Not applicable.

Table S6 Summary of web URLs consulted in this study[1]

Programs	Sites	URLs
Blastn	NCBI	http://blast.ncbi.nlm.nih.gov/Blast.cgi
	Phytozome	http://www.phytozome.net/search.php
	GRAMENE	http://www.gramene.org/Multi/blastview
	TAIR	http://www.arabidopsis.org/Blast/index.jsp
	BrachyBase	http://www.brachybase.org/blast/
	Cucurbit Genomics Database	http://www.icugi.org/cgi-bin/ICuGI/genome/blast.cgi
	Soybase and Soybean Breeder's Toolbox	http://soybase.org/GlycineBlastPages/
	miyakogusa.jp	http://www.kazusa.or.jp/lotus/blast.html
	Medicago truncatula hapmap project	http://www.medicagohapmap.org/advanced_search_page.php?seq
	sol genomics network	http://solgenomics.net/tools/blast/index.pl
Blast2	NCBI	http://blast.ncbi.nlm.nih.gov/Blast.cgi
Prediction of subcellular localization	TargetP	http://www.cbs.dtu.dk/services/TargetP/
	Predotar	http://urgi.versailles.inra.fr/predotar/predotar.html

[1]All programs were run with default parameters.

```
Sugar beet     GGAACCGTAGCCAAGTGG-TAAGGCATGAGTCTGCAAAA   354386
Spinach        GGAACCGTAGCCAAGTGGCTAAGGCATGAGTCTGCAAAA     3184
Cucumber       GGAACCGTAGCCAAGTGGCTAAGGCATAAGTCTGCAAAA    99396
Watermelon     GGAACCGTAGCCAAGTGGCTAAGGCATAAGTCTGCAAAA   257936
Aquilegia1     GGAACCGTAGCCAAGTGGCTAAGGCATGAGTCTGCAAAA  2437495
Aquilegia2     GGAACCGTAGCCAAGTGGCTAAGGCATGAGTCTGCAAAA  2501192
Soybean        GGAACCGTAGCCAAGTGGCTAAGGCATGAGTCTGCAAGA     1309
Groundnut      GGAACCGTAGCCAAGTGGCTAAGGCATGAGTCTGCAAAA      131
Mung bean      GGAACCGTAGCCAAGTGGCTAAGGCATGAGTCTGCAAGA   274537
Basella                       CTAAGGCATGAGTCTGCAAAA       21
Daphniphyllum                 CTAAGGCATGAGTCTGCAAAA       21
Hibbertia                     CTAAGGCATGAGTCTGCAAAA       21


Sugar beet     CTTCTATTCGTCGGTTCGAACCCGACCGGTTCCT        354420
Spinach        CTTCTATTCGTCGGTTCGAATCCGACCGGTTCCT          3218
Cucumber       CTTCTATTCGTCGGTTCGAATCCGACCGGTTCCT         99362
Watermelon     CTTCTATTCGTCGGTTCGAATCCGACCGGTTCCT        257970
Aquilegia1     CTTCTATTCGTCGGTTCGAATCCGACCGGTTCCT       2437529
Aquilegia2     CTTCTATTCGTCGGTTCGAATCCGACCGGTTCCT       2501158
Soybean        CTTCTATTCGTCGGTTGGAATCCGACCGGTTCCT          1275
Groundnut      CTTCTATTCGTCGGTTCGAATCCGCCCGGTTCCT           165
Mung bean      CTTCTATTCGTCGGTTGGAATCCGACCGGTTCCT        274571
Basella        CTTCTATTCGTCG                                34
Daphniphyllum CTTCTATTCGTCG                                34
Hibbertia      CTTCTATTCGTCG                                34
```

Figure S1 Comparison of *trnC2*-GCA sequences of sugar beet (DDBJ/EMBL/GenBank accession no. BA000009), spinach (this study), cucumber (Phytozome, scaffold_04063), watermelon (DDBJ/EMBL/GenBank, GQ856147), two copies of *Aquilegia coerulea* (Phytozome, scaffold_23), soybean (Phytozome, scaffold_682), groundnut (DDBJ/EMBL/GenBank,EZ742316) and mung bean (DDBJ/EMBL/GenBank, HM367685). Dashes are incorporated for the maximum matching. Nucleotide sequence of partial *trnC2*-GCA-coding regions of *Basella rubra* (this study), *Daphniphyllum macropodum* (this study) and *Hibbertia pedunculata* (this study) are also aligned. Nucleotide residues are coordinated to the source entries.

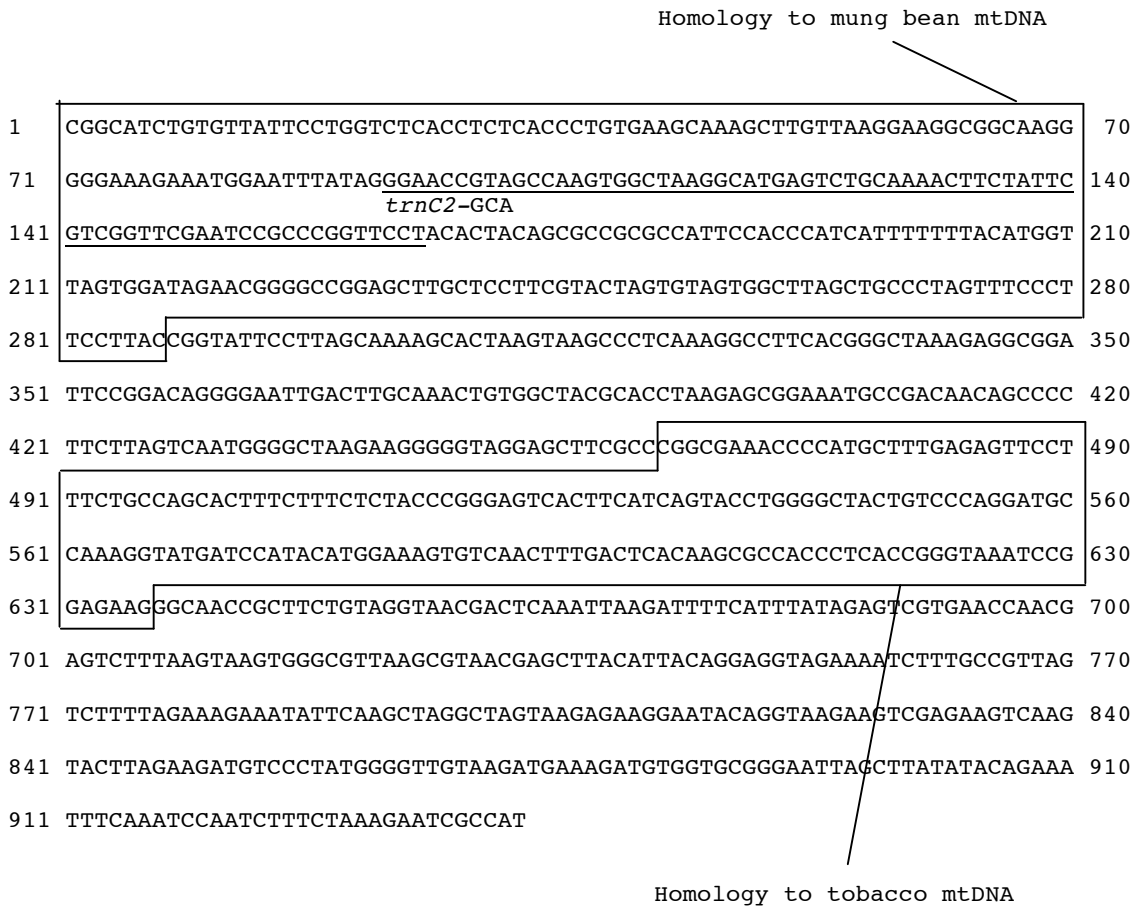
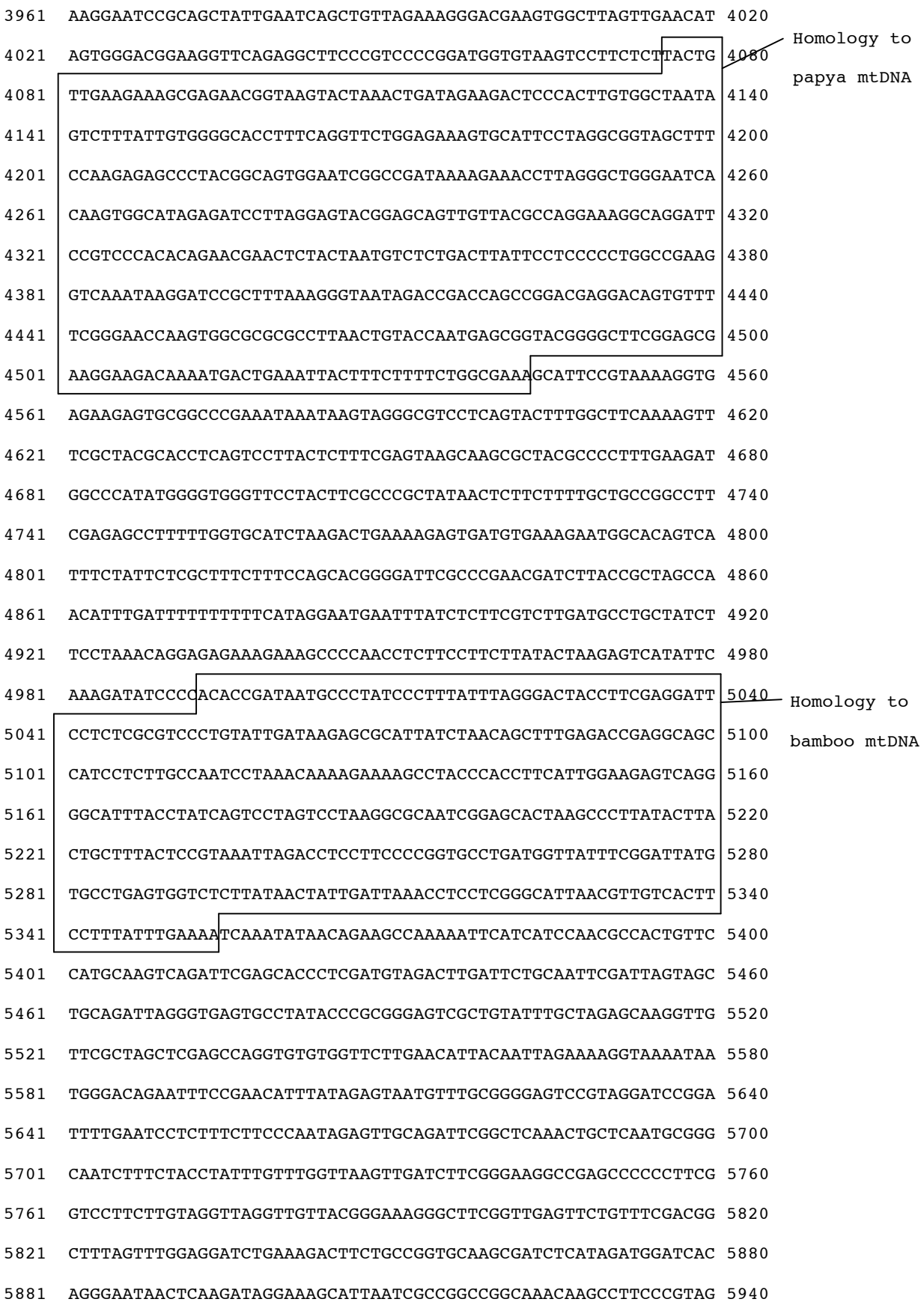


Figure S2 Nucleotide sequence of EZ742316, an EST entry of ground nut. *trnC2*-GCA sequence is underlined. Homologous sequences to plant mitochondrial DNA are boxed.


Homology to
Mungbean
mtDNA
trnC2-GCA (complementary)

```
1981  CAGAAAAACTCCCATATTCATATTTTTAAGTAACCATCGCATAATAATAATATCTGATGT  2040
2041  AGTATCAGCAAGAGCCCTTTACTCGGCTTCTGAAACAAGAGAACGGACCGACTCTAACTA  2100
2101  ATCCACTCCTAATAACAGGAAGGCTAAAGGCAGACCTCGTCCTACTAAAGGAAAGGGCTC  2160
2161  TTTCTATTCCGAGTTGGCCTGCACCGGTCCTATCCTATTCGATTAGGAAACTGATGATTC  2220
2221  TAGAGCAAGTGACATCCATATCAAATCCTCCAGAAGTTATGCTCAACAGGAAAGTCATTA  2280
2281  AGAAGTGCCACAGCACTATTGCTATTGATCGGACATTCACAATTGCATTACCTGAATGGA  2340
2341  ATGGCAGCAGTCTTCTTAGTCCCAATCCCTGCGCATTCAAGAACAAGCCCATCTAGGAAT  2400
2401  ATCTAATCCCTTGGGTACGAACTAGCATTCGATTCTGTGTACGTGGTAAACTCTTGATCC  2460
2461  TTTAAAGAGAAAGAGGGCATCAACAATTTAACTTCCTTGCCGCTGGCTTTTCAACTAACT  2520
2521  GGCCAGGAGAGTTTACTTATGACTTCCTTAGGACTGCAATAGGGCTTGCTTAGTCACCTC  2580
2581  CTTAACTCATAGAATGCTTGAAAACTATCTTAATTGTGACTTGCTTTCCAAAACCAAAAC  2640
2641  AGGATAACTCAAAGGCTTGATCACGGGATTTGACTTCCTGCTAATAATCCTGCTGTATAA  2700
2701  CCCCTCCGCAAGGTAACCTATCACTCCGGGCTTACCTTCAAAAAGTGTAACGACCTAGAA  2760
2761  GGTCTAGAAAATGACTATTAGTTAGACGGAATAGTCATTTCATTGAAACTATTACGAGGA  2820
2821  CTGTACTAGAACAGGCTCCCATGGAACACAACTACACCTCGGCTTGCCTCTCAGTTGAAA  2880
2881  GAGGAAGATGCCATGAGACTGATCTGGATATTCCCTAAGTACAAGGAAAGATCTCCATTA  2940
2941  TAACCCAACCCTGTACGTACCCTTTTGCCATCGTATCCCGTAAGCCCTTCTTTCTTTTAG  3000
3001  CCAACTCCTTCTACTTCAAATAGCTCTATCCACTGGGACGGACTTTAAGCCTTTCTCTTC  3060
3061  TTTACTTAACCTTGCTCCGCAACTGAAAGGGGTCGGTAGACTGACTTGCTTGCGGGACTA  3120
3121  ATGCTAACAAATTGCTCAACTACTCGCTTCAAGTCTCCCGCTCTCTTATCTTGAATCAGT  3180
3181  GTGCACCGAGGGAGTCAAGATTTATAGAGTCAGTGTGCAGTGGCCTATTTTTCATTCAAC  3240
3241  TAATTCATAGTTCACTTAGCCCTCTAGCTAAAAGCTATCAGTCAAGTTAGGAGGAGGGTT  3300
3301  GAATCTGTACTTCCTTTCCTTCCTCCAACTTCTTAGCCCCGATCTTTGGAATTTCCTAGT  3360
3361  GAGGGTCCCGATAGGGGCAATCTCAAGGATTCGTCCTACTTCTAGATTGAACCACCTGAA  3420
3421  AAAGGCTATGGTGACGCTTTCTCACTCATTGATCAGGATTTTTTGTCGTGTTATTCGCAT  3480
3481  TCTGTTTTGGCTACTCACGAAACTCTTTATCCTTAATAGGAAGAACTCCCTTTAGAGGAA  3540
3541  GGATACAAAGACCAAGCCTGTCTTTCTCTCTGAAAAAGGGCATATGGGGCAGACAGACTA  3600
3601  TTTTTGGGGACTAATCTGCTAATGAATGAAGGAAGGAGAAGGCCTTAATTGTCTATCCAA  3660
3661  AGGTCCGCCTTGGGTGGAAATAGGCGAATCAAGTGGGTGGGAATGATCCGAGCCCCCTAT  3720
3721  GTCAATTTCAGTAAGGTAACTAAGGTTTACAATAATTAATTTCAGTAGTGCGTTACCTTG  3780
3781  CTGCTCTATCTATCGGGTAATCAGTCTCTTGGTCAAAAGCCGGTACCTAATTCCCCTGCT  3840
3841  GGTGGGCCGATCCTCGTCTAGCGGTGGGAACCGGCTGGTATGGATCAATTGCTTGCCCTG  3900
3901  TCTCTCTTCTGTTCAACCATCGAGAATCCGCAGTTCTCTTTTTGGGTGGAATAATAGGCG  3960
```


3961 AAGGAATCCGCAGCTATTGAATCAGCTGTTAGAAAGGGACGAAGTGGCTTAGTTGAACAT 4020
4021 AGTGGGACGGAAGGTTCAGAGGCTTCCCGTCCCCGGATGGTGTAAGTCCTTCTCTTACTG 4080
4081 TTGAAGAAAGCGAGAACGGTAAGTACTAAACTGATAGAAGACTCCCACTTGTGGCTAATA 4140
4141 GTCTTTATTGTGGGGCACCTTTCAGGTTCTGGAGAAAGTGCATTCCTAGGCGGTAGCTTT 4200
4201 CCAAGAGAGCCCTACGGCAGTGGAATCGGCCGATAAAAGAAACCTTAGGGCTGGGAATCA 4260
4261 CAAGTGGCATAGAGATCCTTAGGAGTACGGAGCAGTTGTTACGCCAGGAAAGGCAGGATT 4320
4321 CCGTCCCACACAGAACGAACTCTACTAATGTCTCTGACTTATTCCTCCCCCTGGCCGAAG 4380
4381 GTCAAATAAGGATCCGCTTTAAAGGGTAATAGACCGACCAGCCGGACGAGGACAGTGTTT 4440
4441 TCGGGAACCAAGTGGCGCGCGCCTTAACTGTACCAATGAGCGGTACGGGGCTTCGGAGCG 4500
4501 AAGGAAGACAAAATGACTGAAATTACTTTCTTTTCTGGCGAAAGCATTCCGTAAAAGGTG 4560
4561 AGAAGAGTGCGGCCCGAAATAAATAAGTAGGGCGTCCTCAGTACTTTGGCTTCAAAAGTT 4620
4621 TCGCTACGCACCTCAGTCCTTACTCTTTCGAGTAAGCAAGCGCTACGCCCCTTTGAAGAT 4680
4681 GGCCCATATGGGGTGGGTTCCTACTTCGCCCGCTATAACTCTTCTTTTGCTGCCGGCCTT 4740
4741 CGAGAGCCTTTTTGGTGCATCTAAGACTGAAAAGAGTGATGTGAAAGAATGGCACAGTCA 4800
4801 TTTCTATTCTCGCTTTCTTTCCAGCACGGGGATTCGCCCGAACGATCTTACCGCTAGCCA 4860
4861 ACATTTGATTTTTTTTTTCATAGGAATGAATTTATCTCTTCGTCTTGATGCCTGCTATCT 4920
4921 TCCTAAACAGGAGAGAAAGAAAGCCCCAACCTCTTCCTTCTTATACTAAGAGTCATATTC 4980
4981 AAAGATATCCCCACACCGATAATGCCCTATCCCTTTATTTAGGGACTACCTTCGAGGATT 5040
5041 CCTCTCGCGTCCCTGTATTGATAAGAGCGCATTATCTAACAGCTTTGAGACCGAGGCAGC 5100
5101 CATCCTCTTGCCAATCCTAAACAAAAGAAAAGCCTACCCACCTTCATTGGAAGAGTCAGG 5160
5161 GGCATTTACCTATCAGTCCTAGTCCTAAGGCGCAATCGGAGCACTAAGCCCTTATACTTA 5220
5221 CTGCTTTACTCCGTAAATTAGACCTCCTTCCCCGGTGCCTGATGGTTATTTCGGATTATG 5280
5281 TGCCTGAGTGGTCTCTTATAACTATTGATTAAACCTCCTCGGGCATTAACGTTGTCACTT 5340
5341 CCTTTATTTGAAAATCAAATATAACAGAAGCCAAAAATTCATCATCCAACGCCACTGTTC 5400
5401 CATGCAAGTCAGATTCGAGCACCCTCGATGTAGACTTGATTCTGCAATTCGATTAGTAGC 5460
5461 TGCAGATTAGGGTGAGTGCCTATACCCGCGGGAGTCGCTGTATTTGCTAGAGCAAGGTTG 5520
5521 TTCGCTAGCTCGAGCCAGGTGTGTGGTTCTTGAACATTACAATTAGAAAAGGTAAAATAA 5580
5581 TGGGACAGAATTTCCGAACATTTATAGAGTAATGTTTGCGGGGAGTCCGTAGGATCCGGA 5640
5641 TTTTGAATCCTCTTTCTTCCCAATAGAGTTGCAGATTCGGCTCAAACTGCTCAATGCGGG 5700
5701 CAATCTTTCTACCTATTTGTTTGGTTAAGTTGATCTTCGGGAAGGCCGAGCCCCCCTTCG 5760
5761 GTCCTTCTTGTAGGTTAGGTTGTTACGGGAAAGGGCTTCGGTTGAGTTCTGTTTCGACGG 5820
5821 CTTTAGTTTGGAGGATCTGAAAGACTTCTGCCGGTGCAAGCGATCTCATAGATGGATCAC 5880
5881 AGGGAATAACTCAAGATAGGAAAGCATTAATCGCCGGCCGGCAAACAAGCCTTCCCGTAG 5940
Homology to papya mtDNA
Homology to bamboo mtDNA

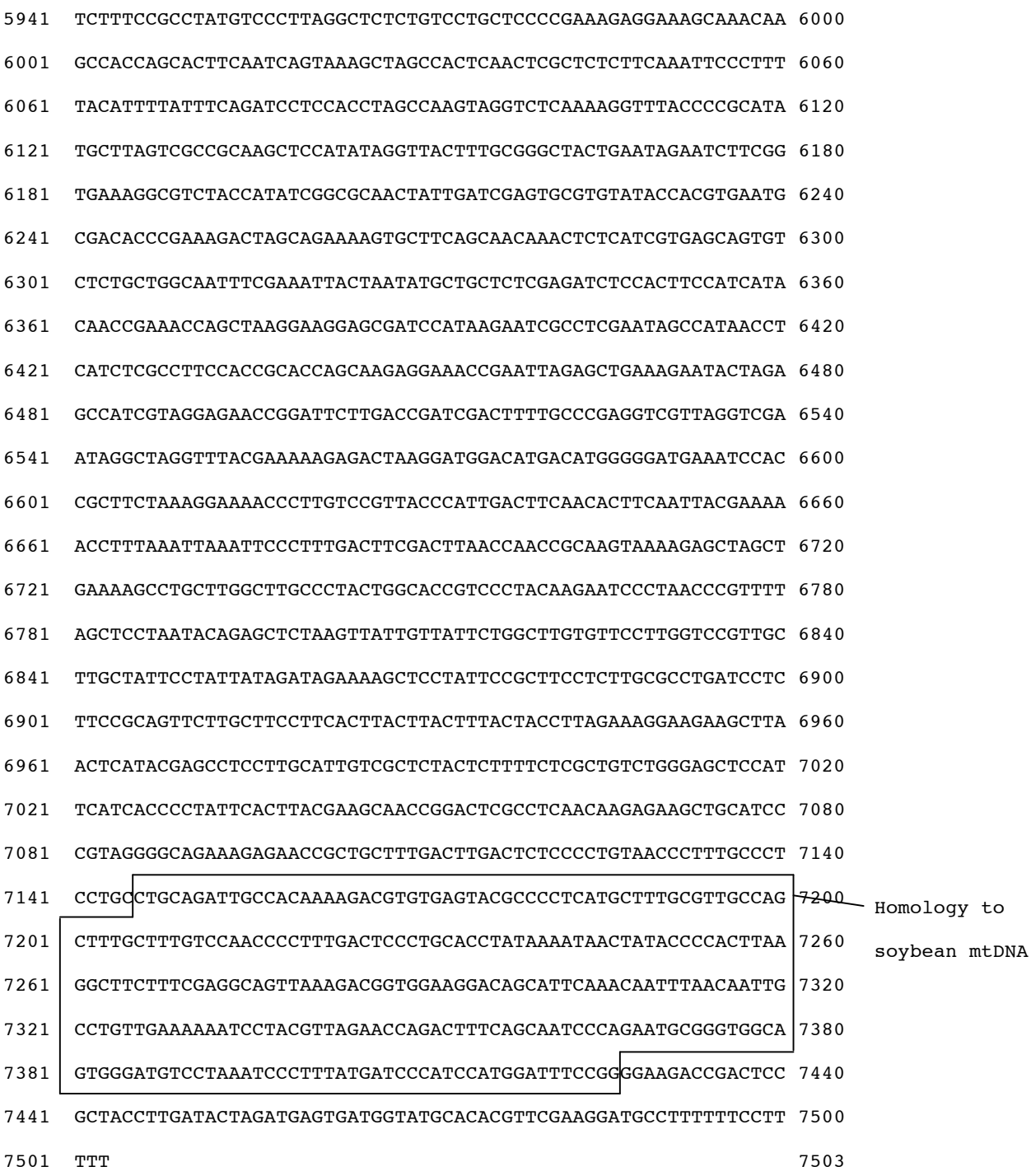
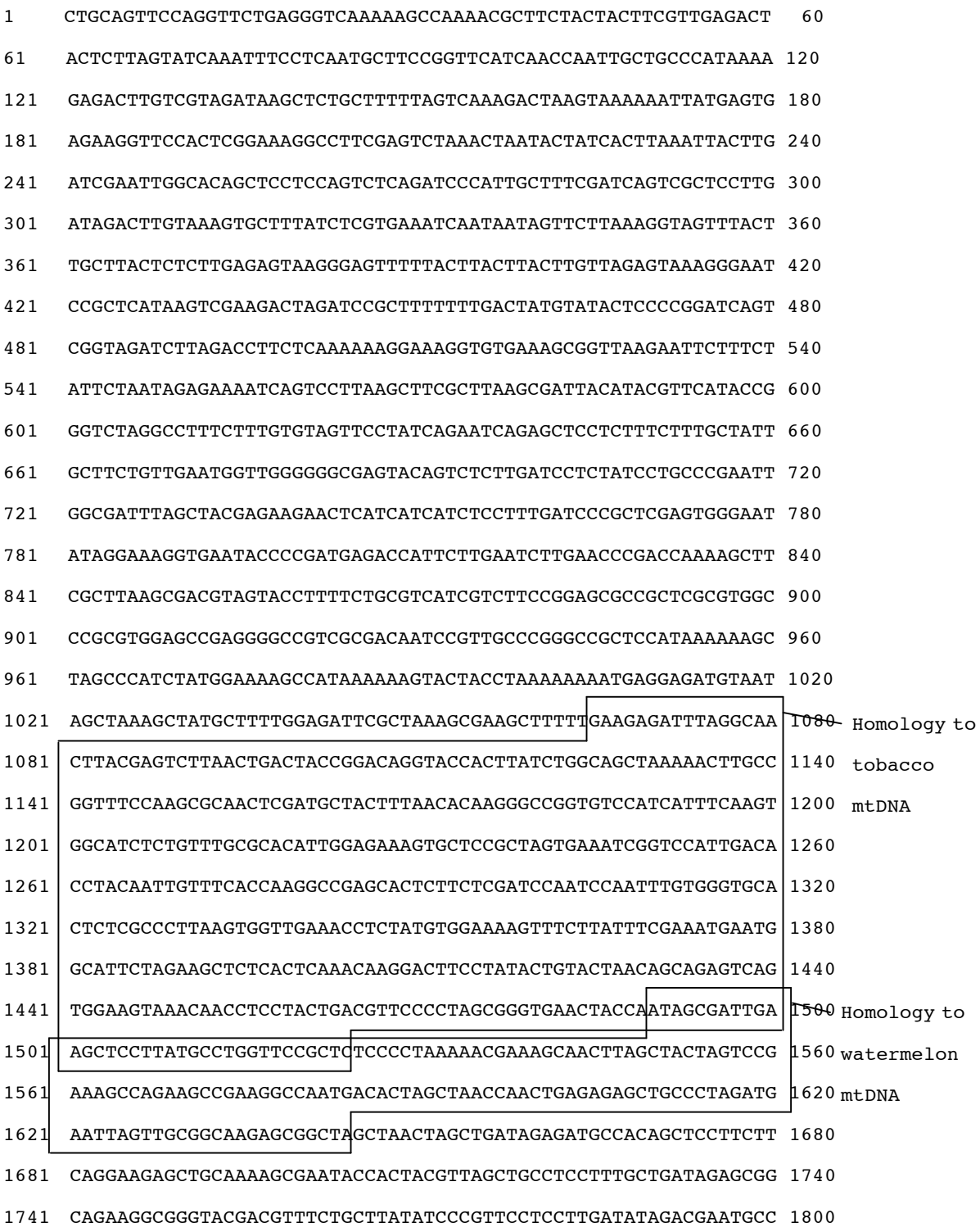
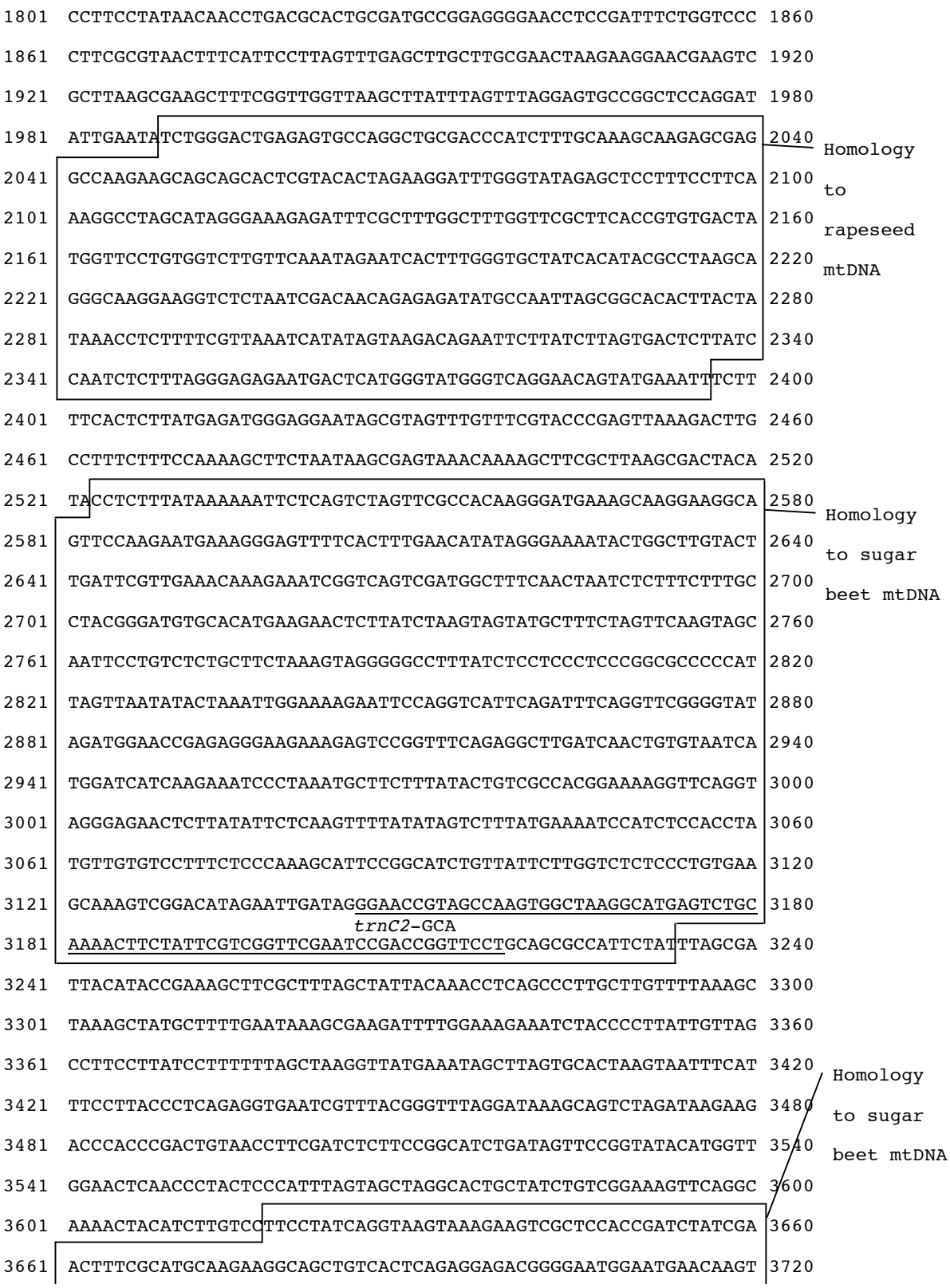


Figure S3 Nucleotide sequence of scaffold_682, a genomic sequence taken from Phytozome. *trnC2*-GCA sequence is underlined. Homologous sequences to plant mitochondrial DNA or other entries are boxed.


1 CTGCAGTTCCAGGTTCTGAGGGTCAAAAAGCCAAAACGCTTCTACTACTTCGTTGAGACT 60
61 ACTCTTAGTATCAAATTTCCTCAATGCTTCCGGTTCATCAACCAATTGCTGCCCATAAAA 120
121 GAGACTTGTCGTAGATAAGCTCTGCTTTTTAGTCAAAGACTAAGTAAAAAATTATGAGTG 180
181 AGAAGGTTCCACTCGGAAAGGCCTTCGAGTCTAAACTAATACTATCACTTAAATTACTTG 240
241 ATCGAATTGGCACAGCTCCTCCAGTCTCAGATCCCATTGCTTTCGATCAGTCGCTCCTTG 300
301 ATAGACTTGTAAAGTGCTTTATCTCGTGAAATCAATAATAGTTCTTAAAGGTAGTTTACT 360
361 TGCTTACTCTCTTGAGAGTAAGGGAGTTTTTACTTACTTACTTGTTAGAGTAAAGGGAAT 420
421 CCGCTCATAAGTCGAAGACTAGATCCGCTTTTTTTGACTATGTATACTCCCCGGATCAGT 480
481 CGGTAGATCTTAGACCTTCTCAAAAAAGGAAAGGTGTGAAAGCGGTTAAGAATTCTTTCT 540
541 ATTCTAATAGAGAAAATCAGTCCTTAAGCTTCGCTTAAGCGATTACATACGTTCATACCG 600
601 GGTCTAGGCCTTTCTTTGTGTAGTTCCTATCAGAATCAGAGCTCCTCTTTCTTTGCTATT 660
661 GCTTCTGTTGAATGGTTGGGGGGCGAGTACAGTCTCTTGATCCTCTATCCTGCCCGAATT 720
721 GGCGATTTAGCTACGAGAAGAACTCATCATCATCTCCTTTGATCCCGCTCGAGTGGGAAT 780
781 ATAGGAAAGGTGAATACCCCGATGAGACCATTCTTGAATCTTGAACCCGACCAAAAGCTT 840
841 CGCTTAAGCGACGTAGTACCTTTTCTGCGTCATCGTCTTCCGGAGCGCCGCTCGCGTGGC 900
901 CCGCGTGGAGCCGAGGGGCCGTCGCGACAATCCGTTGCCCGGGCCGCTCCATAAAAAAGC 960
961 TAGCCCATCTATGGAAAAGCCATAAAAAAGTACTACCTAAAAAAAATGAGGAGATGTAAT 1020
1021 AGCTAAAGCTATGCTTTTGGAGATTCGCTAAAGCGAAGCTTTTTGAAGAGATTTAGGCAA 1080
1081 CTTACGAGTCTTAACTGACTACCGGACAGGTACCACTTATCTGGCAGCTAAAAACTTGCC 1140
1141 GGTTTCCAAGCGCAACTCGATGCTACTTTAACACAAGGGCCGGTGTCCATCATTTCAAGT 1200
1201 GGCATCTCTGTTTGCGCACATTGGAGAAAGTGCTCCGCTAGTGAAATCGGTCCATTGACA 1260
1261 CCTACAATTGTTTCACCAAGGCCGAGCACTCTTCTCGATCCAATCCAATTTGTGGGTGCA 1320
1321 CTCTCGCCCTTAAGTGGTTGAAACCTCTATGTGGAAAAGTTTCTTATTTCGAAATGAATG 1380
1381 GCATTCTAGAAGCTCTCACTCAAACAAGGACTTCCTATACTGTACTAACAGCAGAGTCAG 1440
1441 TGGAAGTAAACAACCTCCTACTGACGTTCCCCTAGCGGGTGAACTACCAATAGCGATTGA 1500
1501 AGCTCCTTATGCCTGGTTCCGCTCTCCCCTAAAAACGAAAGCAACTTAGCTACTAGTCCG 1560
1561 AAAGCCAGAAGCCGAAGGCCAATGACACTAGCTAACCAACTGAGAGAGCTGCCCTAGATG 1620
1621 AATTAGTTGCGGCAAGAGCGGCTAGCTAACTAGCTGATAGAGATGCCACAGCTCCTTCTT 1680
1681 CAGGAAGAGCTGCAAAAGCGAATACCACTACGTTAGCTGCCTCCTTTGCTGATAGAGCGG 1740
1741 CAGAAGGCGGGTACGACGTTTCTGCTTATATCCCGTTCCTCCTTGATATAGACGAATGCC 1800
Homology to
tobacco
mtDNA
Homology to
watermelon
mtDNA


1801 CCTTCCTATAACAACCTGACGCACTGCGATGCCGGAGGGGAACCTCCGATTTCTGGTCCC 1860
1861 CTTCGCGTAACTTTCATTCCTTAGTTTGAGCTTGCTTGCGAACTAAGAAGGAACGAAGTC 1920
1921 GCTTAAGCGAAGCTTTCGGTTGGTTAAGCTTATTTAGTTTAGGAGTGCCGGCTCCAGGAT 1980
1981 ATTGAATATCTGGGACTGAGAGTGCCAGGCTGCGACCCATCTTTGCAAAGCAAGAGCGAG 2040
2041 GCCAAGAAGCAGCAGCACTCGTACACTAGAAGGATTTGGGTATAGAGCTCCTTTCCTTCA 2100
2101 AAGGCCTAGCATAGGGAAAGAGATTTCGCTTTGGCTTTGGTTCGCTTCACCGTGTGACTA 2160
2161 TGGTTCCTGTGGTCTTGTTCAAATAGAATCACTTTGGGTGCTATCACATACGCCTAAGCA 2220
2221 GGGCAAGGAAGGTCTCTAATCGACAACAGAGAGATATGCCAATTAGCGGCACACTTACTA 2280
2281 TAAACCTCTTTTCGTTAAATCATATAGTAAGACAGAATTCTTATCTTAGTGACTCTTATC 2340
2341 CAATCTCTTTAGGGAGAGAATGACTCATGGGTATGGGTCAGGAACAGTATGAAATTTCTT 2400
2401 TTCACTCTTATGAGATGGGAGGAATAGCGTAGTTTGTTTCGTACCCGAGTTAAAGACTTG 2460
2461 CCTTTCTTTCCAAAAGCTTCTAATAAGCGAGTAAACAAAAGCTTCGCTTAAGCGACTACA 2520
2521 TACCTCTTTATAAAAAATTCTCAGTCTAGTTCGCCACAAGGGATGAAAGCAAGGAAGGCA 2580
2581 GTTCCAAGAATGAAAGGGAGTTTTCACTTTGAACATATAGGGAAAATACTGGCTTGTACT 2640
2641 TGATTCGTTGAAACAAAGAAATCGGTCAGTCGATGGCTTTCAACTAATCTCTTTCTTTGC 2700
2701 CTACGGGATGTGCACATGAAGAACTCTTATCTAAGTAGTATGCTTTCTAGTTCAAGTAGC 2760
2761 AATTCCTGTCTCTGCTTCTAAAGTAGGGGGCCTTTATCTCCTCCCTCCCGGCGCCCCCAT 2820
2821 TAGTTAATATACTAAATTGGAAAAGAATTCCAGGTCATTCAGATTTCAGGTTCGGGGTAT 2880
2881 AGATGGAACCGAGAGGGAAGAAAGAGTCCGGTTTCAGAGGCTTGATCAACTGTGTAATCA 2940
2941 TGGATCATCAAGAAATCCCTAAATGCTTCTTTATACTGTCGCCACGGAAAAGGTTCAGGT 3000
3001 AGGGAGAACTCTTATATTCTCAAGTTTTATATAGTCTTTATGAAAATCCATCTCCACCTA 3060
3061 TGTTGTGTCCTTTCTCCCAAAGCATTCCGGCATCTGTTATTCTTGGTCTCTCCCTGTGAA 3120
3121 GCAAAGTCGGACATAGAATTGATAGGGAACCGTAGCCAAGTGGCTAAGGCATGAGTCTGC 3180
3181 AAAACTTCTATTCGTCGGTTCGAATCCGACCGGTTCCTGCAGCGCCATTCTATTTAGCGA 3240
3241 TTACATACCGAAAGCTTCGCTTTAGCTATTACAAACCTCAGCCCTTGCTTGTTTTAAAGC 3300
3301 TAAAGCTATGCTTTTGAATAAAGCGAAGATTTTGGAAAGAAATCTACCCCTTATTGTTAG 3360
3361 CCTTCCTTATCCTTTTTTAGCTAAGGTTATGAAATAGCTTAGTGCACTAAGTAATTTCAT 3420
3421 TTCCTTACCCTCAGAGGTGAATCGTTTACGGGTTTAGGATAAAGCAGTCTAGATAAGAAG 3480
3481 ACCCACCCGACTGTAACCTTCGATCTCTTCCGGCATCTGATAGTTCCGGTATACATGGTT 3540
3541 GGAACTCAACCCTACTCCCATTTAGTAGCTAGGCACTGCTATCTGTCGGAAAGTTCAGGC 3600
3601 AAAACTACATCTTGTCCTTCCTATCAGGTAAGTAAAGAAGTCGCTCCACCGATCTATCGA 3660
3661 ACTTTCGCATGCAAGAAGGCAGCTGTCACTCAGAGGAGACGGGGAATGGAATGAACAAGT 3720
Homology to rapeseed mtDNA
Homology to sugar beet mtDNA
Homology to sugar beet mtDNA
trnC2-GCA

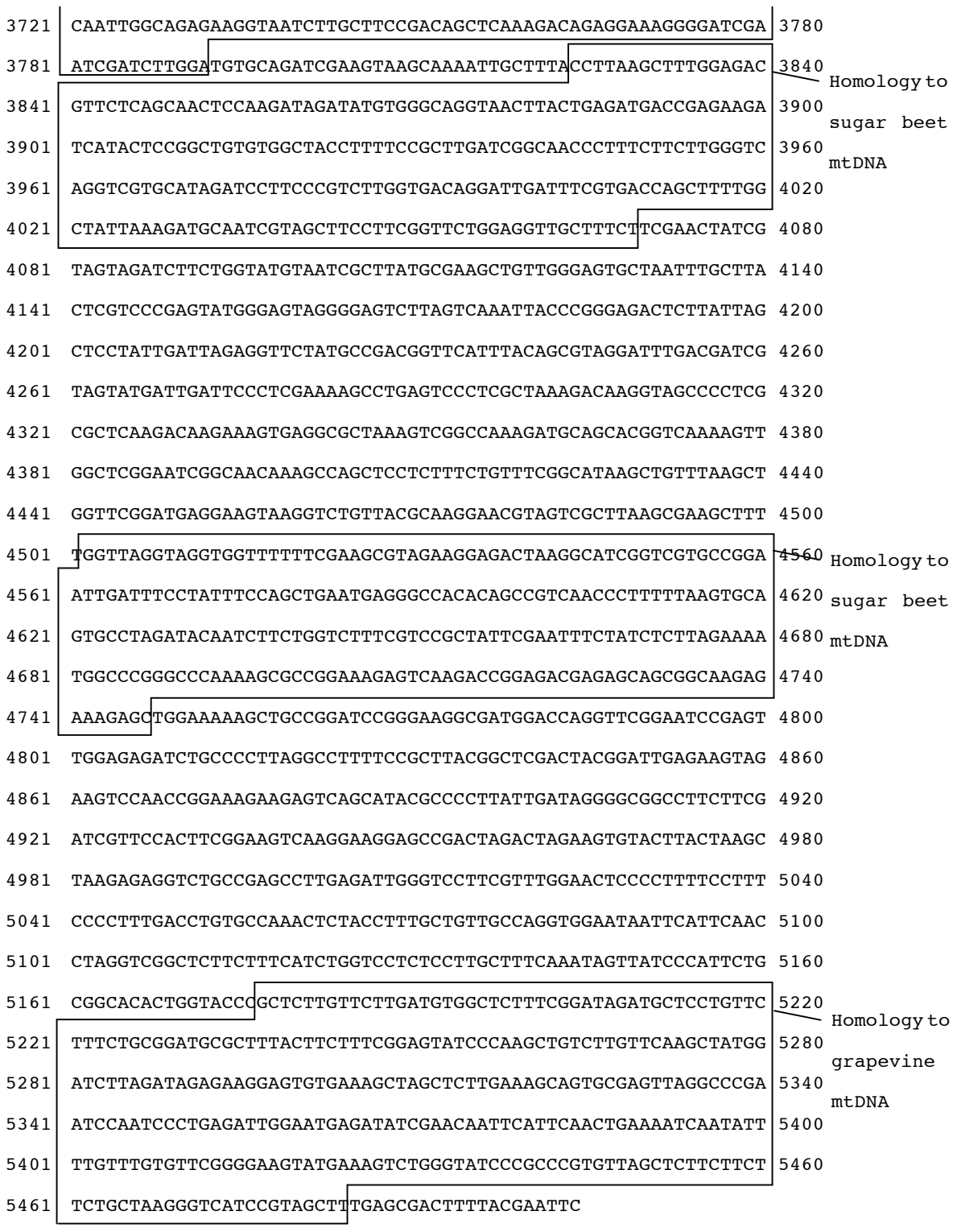

3721 CAATTGGCAGAGAAGGTAATCTTGCTTCCGACAGCTCAAAGACAGAGGAAAGGGGATCGA 3780
3781 ATCGATCTTGGATGTGCAGATCGAAGTAAGCAAAATTGCTTTACCTTAAGCTTTGGAGAC 3840
Homology to
3841 GTTCTCAGCAACTCCAAGATAGATATGTGGGCAGGTAACTTACTGAGATGACCGAGAAGA 3900
sugar beet
3901 TCATACTCCGGCTGTGTGGCTACCTTTTCCGCTTGATCGGCAACCCTTTCTTCTTGGGTC 3960
mtDNA
3961 AGGTCGTGCATAGATCCTTCCCGTCTTGGTGACAGGATTGATTTCGTGACCAGCTTTTGG 4020
4021 CTATTAAAGATGCAATCGTAGCTTCCTTCGGTTCTGGAGGTTGCTTTCTTCGAACTATCG 4080
4081 TAGTAGATCTTCTGGTATGTAATCGCTTATGCGAAGCTGTTGGGAGTGCTAATTTGCTTA 4140
4141 CTCGTCCCGAGTATGGGAGTAGGGGAGTCTTAGTCAAATTACCCGGGAGACTCTTATTAG 4200
4201 CTCCTATTGATTAGAGGTTCTATGCCGACGGTTCATTTACAGCGTAGGATTTGACGATCG 4260
4261 TAGTATGATTGATTCCCTCGAAAAGCCTGAGTCCCTCGCTAAAGACAAGGTAGCCCCTCG 4320
4321 CGCTCAAGACAAGAAAGTGAGGCGCTAAAGTCGGCCAAAGATGCAGCACGGTCAAAAGTT 4380
4381 GGCTCGGAATCGGCAACAAAGCCAGCTCCTCTTTCTGTTTCGGCATAAGCTGTTTAAGCT 4440
4441 GGTTCGGATGAGGAAGTAAGGTCTGTTACGCAAGGAACGTAGTCGCTTAAGCGAAGCTTT 4500
4501 TGGTTAGGTAGGTGGTTTTTTCGAAGCGTAGAAGGAGACTAAGGCATCGGTCGTGCCGGA 4560
Homology to
4561 ATTGATTTCCTATTTCCAGCTGAATGAGGGCCACACAGCCGTCAACCCTTTTTAAGTGCA 4620
sugar beet
4621 GTGCCTAGATACAATCTTCTGGTCTTTCGTCCGCTATTCGAATTTCTATCTCTTAGAAAA 4680
mtDNA
4681 TGGCCCGGGCCCAAAAGCGCCGGAAAGAGTCAAGACCGGAGACGAGAGCAGCGGCAAGAG 4740
4741 AAAGAGCTGGAAAAAGCTGCCGGATCCGGGAAGGCGATGGACCAGGTTCGGAATCCGAGT 4800
4801 TGGAGAGATCTGCCCCTTAGGCCTTTTCCGCTTACGGCTCGACTACGGATTGAGAAGTAG 4860
4861 AAGTCCAACCGGAAAGAAGAGTCAGCATACGCCCCTTATTGATAGGGGCGGCCTTCTTCG 4920
4921 ATCGTTCCACTTCGGAAGTCAAGGAAGGAGCCGACTAGACTAGAAGTGTACTTACTAAGC 4980
4981 TAAGAGAGGTCTGCCGAGCCTTGAGATTGGGTCCTTCGTTTGGAACTCCCCTTTTCCTTT 5040
5041 CCCCTTTGACCTGTGCCAAACTCTACCTTTGCTGTTGCCAGGTGGAATAATTCATTCAAC 5100
5101 CTAGGTCGGCTCTTCTTTCATCTGGTCCTCTCCTTGCTTTCAAATAGTTATCCCATTCTG 5160
5161 CGGCACACTGGTACCGCTCTTGTTCTTGATGTGGCTCTTTCGGATAGATGCTCCTGTTC 5220
Homology to
5221 TTTCTGCGGATGCGCTTTACTTCTTTCGGAGTATCCCAAGCTGTCTTGTTCAAGCTATGG 5280
grapevine
5281 ATCTTAGATAGAGAAGGAGTGTGAAAGCTAGCTCTTGAAAGCAGTGCGAGTTAGGCCCGA 5340
mtDNA
5341 ATCCAATCCCTGAGATTGGAATGAGATATCGAACAATTCATTCAACTGAAAATCAATATT 5400
5401 TTGTTTGTGTTCGGGGAAGTATGAAAGTCTGGGTATCCCGCCCGTGTTAGCTCTTCTTCT 5460
5461 TCTGCTAAGGGTCATCCGTAGCTTTGAGCGACTTTTACGAATTC

Figure S4 Nucleotide sequence of the 5504-bp sequence containing spinach *trnC2*-GCA. *trnC2*-GCA sequence is underlined. Homologous sequences to plant mitochondrial DNA are boxed.

>exon 1

```
ATTTTGCATCTAAAGATTATCATCCCCGAGGGCTGAGAATGGTTGTGTTGGCAAGGTGCT   60
                                          M  V  V  L  A  R  C
GCAAACCCTTTCTCCCGCCATATCTATCCCCTCTTTCTCTCTCCGAAATCCCACTTCAAT  120
C  K  P  F  L  P  P  Y  L  S  P  L  S  L  S  E  I  P  L  Q
TCTATTCTCCATTCTCAAACAAAATCAGAAACTATGCAATCAACAAAAGGGTTCGTTTCT  180
F  Y  S  P  F  S  N  K  I  R  N  Y  A  I  N  K  R  V  R  F
TTTGTTCAATGGGTTCTTCTCAAGCCTCAATTAATGGTAACAATGGCAATAATAATTGGG  240
F  C  S  M  G  S  S  Q  A  S  I  N  G  N  N  G  N  N  N  W
TTGGTAGTTCCAATAATCAGCAATTAGAGGGTGCAATCAAGAAAGATTTGTGGGTGTATA  300
V  G  S  S  N  N  Q  Q  L  E  G  A  V  K  K  D  L  W  V  Y
ATACAATGAGTAGGAAAAAAGAGATTTTTAAGCCTAAAGTAGATGGAAAAGTGGGAATGT  360
N  T  M  S  R  K  K  E  I  F  K  P  K  V  D  G  K  V  G  M
ATGTTTGTGGAGTTACTGCTTACGATTTCAGCCATATTGGTCATGCTAGAGCTTATGTTA  420
Y  V  C  G  V  T  A  Y  D  F  S  H  I  G  H  A  R  A  Y  V
                  >intron 1
CATTTGATATTCTTTACAGgtccaaagtccacccctttttctttctctctctacaagtttc  480
T  F  D  I  L  Y  R
taatgttgtaaaattttgttttggatacttatggtttgtgtatgattgttaatttttgaa  540
atttagatgccctttttacttgatttttgttgggtggagattagaatttgctagttgttta  600
ttatcagttgaggataaactattttcagatgttgtgaatatgaaaagatgaaaactttgt  660
gttaatttgccgagaaatttggattctgttgaatgggtgagttgcttgattttgatttat  720
aagcaaaagttaagttggttctaataaggaaatgttggtataatagttaccagtttaacc  780
agtggtcagaatttcatgtggtcgtttaattgatatgcctccctcaatcatatcgacttt  840
ctaggagattaggcaaaaatgatttgttcttggtctttttttcactctcttacatctaatg  900
atgtgaacaaatgaagaaccgtattaaaatatggaatgcagctgaaaatttgtctgtctt  960
tctgtataatcatttgattcctatatgctgtgaagttcaagcaaatttcgaaaggatagt 1020
taaaagagggtgagctagggagtttaattaccatagtaatggccattagtggtagacctg 1080
agcatatattttcgtactttctagcaagatgaaggttggtgggttaatattccgtagata 1140
atttatctacttctccttgggtattgtactagttattgtgggtgttcaatgtacatgtac 1200
                                          >exon 2
ctgctgattagtgctgtttatttgctctgataagttgacagGTATCTCAAGTATCTGGGC 1260
                                            Y  L  K  Y  L  G
    F                                  Bv1    >intron 2
TATGACGTCTGTTATGTTCGAAACTTCACTGATGTTGATGACAAGgttaggtaacgaatt 1320
 Y  D  V  C  Y  V  R  N  F  T  D  V  D  D  K
gaactgtataatcacgtgcttatgaatttagttaaatctcttccttttggttgtcattct 1380
tcctaactcatttgctctatgtcttccttagtacatagcatctgttcacaatgtgtttgt 1440
ggttgtttgagttatctcgatgtttacattttattatattaagaaatcaacagttttgtg 1500
cttcaatgtcattatgcttttgccttaatttatgaagtaccgactttcttatatgtgtct 1560
```

```
                                        >exon 3  Bv1
gacaggatacctgggtgtggaattttcgaattctgtagATAATTGCCAGAGCCAACGAAG 1620
                                          I  I  A  R  A  N  E
TGGGGGAGAaTCCAATATCTCTTAGCAGACGTTTTTGTGATGAATTTCATCAAGACATGT 1680
V  G  E  N  P  I  S  L  S  R  R  F  C  D  E  F  H  Q  D  M
CATATCTTCATTGTTTACCTCCTTCTGTAGAACCTCGTGTCTCGGATCACATGCCGCAAA 1740
S  Y  L  H  C  L  P  P  S  V  E  P  R  V  S  D  H  M  P  Q
                    >intron 3
TTATAGACATGATTAAGCAGgtaatttttttctttcctatcaagcgatgtgtttttcct 1800
I  I  D  M  I  K  Q
aaagtcatggtttgatgctattcatattctgttggtggacgaaattcttttagcataatt 1860
gttttacttcgaagatctgcattgcacaaacattgaaatttttaattctacttgctgtct 1920
aaattcaacttttgttagaatcagctccactaatatcttgttttccatctttttctggca 1980
 >exon 4
gATTGTTGATAATGGCTTTGCCTATGGTGTGGATGGGGATGTCTATTTCTCTGTGGAAAA 2040
  I  V  D  N  G  F  A  Y  G  V  D  G  D  V  Y  F  S  V  E  K
                                                        A2
GTTTCCTGATTATGGACGATTGTCTGGACGAAAATTGGAGGACAATCGAGCTGGTGAGCG 2100
  F  P  D  Y  G  R  L  S  G  R  K  L  E  D  N  R  A  G  E  R
                                                  >intron 4
                    A1
GGTAGCTATTGATTCGAGAAAGCGCAATCCAGCTGATTTTGCATTATGGAAAgtaagtaa 2160
  V  A  I  D  S  R  K  R  N  P  A  D  F  A  L  W  K
taatttatatttttttctaaaatatcatatacatgtttatggaaattgtttgaatggga 2220
taagtcatctttgctttaatcaaggctgttaacatgcaactatattcagtttctggtaga 2280
cacggagagtggggatagtcagcagtaggtagtgtttattacaatgctttcttctcttgc 2340
atgttctgctaagattatatgttagcctgagaaatatttggtgcaatgcttaagcacctg 2400
aaatcctgaattttgtatatagaatgatgagtgtaaacttttaacttagaccatgccaat 2460
ttttttgagaatagctaagcaattgttttggtgatgcctcatggaaagccttttatgcc 2520
aatgtttttctgttccctgttttgcttcccactgctcactttgaaaagcatgaaaaaag 2580
gatagatttttgttttctgaagcatgcttcagtgctgcatgcacccagttggtgtttgtt 2640
gtggtaatatttcttttaatttgctgcactatttagctgcttttgatatctgacttctta 2700
                     >exon 5
caaaatatgcttctactacagTCTGCCAAAGAAGGAGAGCCCTTTTGGGAGAGTCCATGG 2760
                       S  A  K  E  G  E  P  F  W  E  S  P  W
                    S1
GGTCCTGGAAGACCTGGTTGGCATATTGAATGCAGTGCTATGAGCTCTGTATATTTGGGT 2820
 G  P  G  R  P  G  W  H  I  E  C  S  A  M  S  S  V  Y  L  G
TACTCTTTTGACATACATGGTGGTGGGATGGACCTTATGTTTCCTCACCATGAAAACGAA 2880
 Y  S  F  D  I  H  G  G  G  M  D  L  M  F  P  H  H  E  N  E
```

```
      S2                                                  RT1
ATTGTTCAGAGCTGTGCTGCATGTTGTGAAAGTAATATACGCTACTGGGTGCATAATGGA 2940
 I  V  Q  S  C  A  A  C  C  E  S  N  I  R  Y  W  V  H  N  G
TTTGTCACAATTAATTCAGAAAAAATGTCCAAATCTCTTGGAAACTTCTTCACAATTCGG 3000
 F  V  T  I  N  S  E  K  M  S  K  S  L  G  N  F  F  T  I  R
   >intron 5
CAGgtaagaggacttagctatttcaggtttgcagtcccaagtattgtgatcattgctttg 3060
 Q
cttcatgttcatagttctgttgttctttttatgtcctgatctagtaagccttggtcttt 3120
    >exon 6
ccagGTAATAGAACTTTATCACCCGCTTGCTTTGAGACTCTTTCTCATGGGAACCCACTA 3180
     V  I  E  L  Y  H  P  L  A  L  R  L  F  L  M  G  T  H  Y
TCGATCTCCAATCAATTACTCTGATGTATTATTAGAGAGTGCGTCAGAACGCATTTTCTA 3240
  R  S  P  I  N  Y  S  D  V  L  L  E  S  A  S  E  R  I  F  Y
         >intron 6
CATATACCAGgtatgtctcattattgtattctacgcgaaaatggtggacaaaacttaggg 3300
  I  Y  Q
agcattaaaacctttctatattattaaaattttgtgccttcctaacttagacttgggtct 3360
gtcctctttttgcctataaattctaatgaaactgctagttattatggtatgaccattgga 3420
tgtagattttgtgacatgaatcctactgcctaaaagtttttctgcgtctcttctcaagtt 3480
cctttctcttctctttgtggagtaatatttgtttgtgatatgggcagaacctagatatcg 3540
                                                    >exon 7
tgttgtactaaatctcctactgctccttttactttatgcaatctatatgcagACATTACA 3600
                                                      T  L  Q
AGACAGTCAAGCTCTTCTGAGCGAGTATGGGAGAGCTAATTTAAAGGACTCAATACCCCA 3660
  D  S  Q  A  L  L  S  E  Y  G  R  A  N  L  K  D  S  I  P  Q
GGATATTGTTAGTTGCATCAATGATTTGCAAAATAGTTTTGTGATTTCGATGTCAGATGA 3720
  D  I  V  S  C  I  N  D  L  Q  N  S  F  V  I  S  M  S  D  D
TCTTCATACACCTGTTGTTCTTGCTGCACTTTCTGACCCATTGAAGACTATCAATGATCT 3780
  L  H  T  P  V  V  L  A  A  L  S  D  P  L  K  T  I  N  D  L
               >intron 7
TCTTCATACTCGCAAGgtattcaactttctggctgatccccaattcatcttttctggtcc 3840
  L  H  T  R  K
atatacaaggatattatgtgatctttgtgggatgaatttttgactgtgtcgcgcaccag 3900
ttttgcaaagttttattcactcagaatatccatacctttctgttccttatctctagcatc 3960
actcttttgggagagggtttgcatgttatgcaagtatgaacttattttgctccaagctat 4020
ctggctcactatgcgtgttctataagcaagggacagaaataaaaatcattgtctatatcg 4080
cttgattaaactaaatccaacttgcagaagttttattaaccaatgtctgctagtgtgtga 4140
aacaacgatcacattatcacttgaaatatatgataaacttcatccaactctagaacaatg 4200
ataaaatcatgaaaactatgaaagaaagcagataggatgtataacggtactattggattg 4260
```

```
atcctgggtgatgaaattcccgaaatttatgatttgttctttaaaaatctttgtgacgca 4320
tttttaaggagaataacggatgctctttcatgtggttttcgttgtagtcgcttcttttaa 4380
ccatcctatataccgttttctaaactccattgaggggagaatggctttagaatttggtat 4440
cttaatgggaaaaactcaaatcttcaaatacttctgtaaaacctgcaacatgttgacttt 4500
cacacatgtcaatacactactttggctcttaatatctttgattatcgattagtaaaactt 4560
ataaaagattgatagtttgaagatacattttgataccaatccaacaatatattatatgat 4620
agctattgtatttatatatcagctttgaaatgcgtaaaagtttagtattttaatagcaat 4680
                                                      >exon 8
tttgcatgtatcagaaacatagttagtataattttgacacatggcgctacagGGAAAGAA 4740
                                                        G  K  K
GCGTGAGCTTCGAGTAGAATCACTTGCAGCTTTAGAGAAGACAATCAAGCATGTGCTGAC 4800
  R  E  L  R  V  E  S  L  A  A  L  E  K  T  I  K  H  V  L  T
                                    >intron 8
TGCTCTTGGGTTGATGCCATCAAGTTATTCAGAGgttaaaatattttgatattttcagat 4860
  A  L  G  L  M  P  S  S  Y  S  E
ttaatgcatctaacattcaaatattcaataactttgacttttgttaaattgtgatgttgc 4920
  >exon 9
agGCTCTACATCAGCTTAGAGAGAAGGCTTTGAAACGCGTGAAGTTTTCAGAAACTGAAG 4980
   A  L  H  Q  L  R  E  K  A  L  K  R  V  K  F  S  E  T  E
TTCAGCAAAAGATTGAAGAGAGAGACATGGCAAGGAAGTACAAAGAATATGAGAAGTCTG 5040
V  Q  Q  K  I  E  E  R  D  M  A  R  K  N  K  E  Y  E  K  S
ATGCAATTAGGAAAGAGTTGGCTGCTGTGGGAATTGCTCTGATGGACAGCCCAGAAGGGA 5100
D  A  I  R  K  E  L  A  A  V  G  I  A  L  M  D  S  P  E  G
CGACCTGGAGACCTGTGGTTCCTACTGCACTACAGCAAGAGCCAGTTATGGCTAATTGAG 5160
T  T  W  R  P  V  V  P  T  A  L  Q  Q  E  P  V  M  A  N  *
TCAATGTATAACGTGATTTTTTTATACCATATTAGTTGATCAACAATTACTATCCAAGGA 5220
GACAAAGGATTCAGAATATTGGGAAATCCGCTTTATGATCAAATCAGTCTTCTTGTGAGG 5280
AGACCTGGGGATAATCATAGGATGTGGAAATTAGCACTATTGTGACGATGATTATATTTT 5340
TTCCCCCTACATTATGGATTTTCAAGAGCTTATGGTGTGTATTTGAGTGAATAAAGCAAC 5400
CATCATTTTTACCATTTTACTGCTGTGTCTGTGACACCCTTATTGCGATTCTTTAATGTT 5460
GAAGCCAGGCTGAGAATATGTGCCATTAAAGTTGAGGAAACAGAAACAGTTTTAAGAGGG 5520
GGAAAATTCACTTTGATATTTGAATTGGTTGGAG                           5554
```

Figure S5 Nucleotide sequence of sugar-beet gene coding for cysteinyl-tRNA synthetase (*bvcysRS*). Exons and introns are indicated by upper and lower cases, respectively. Putative translation product is shown below in single letter code. Positions of oligonucleotides used for 5' RACE and RT-PCR are underlined.

a

```
bv  MVV----LARCCKPFLPPYLSPLSLSEIPLQFYSPFSN-KIRNYAINKRVRFFCSMGSSQASINGNNGNNNWVGSSNNQQLEGAVKKDLWVYNTMSRKKE   95
at  MASSVLNLFKSCRPFTPIRFSSLPKSQFRIQFPLRPGK-ETQLRRCFTTL---SSLTDGGAPISG--------------------GKELWLHNTMSRKKE   76
cs  MAS----LLKFYNPLTLTRFTPVSLSQAFLRRTLRNSH-FSFFNSATTFARFTSSPTSSQLPVSAPNLKE------INALRHDSTSSELWLHNTMSRKKE   90
os  MAA-----ARRAAGLLPLLLSSPSRARLPHRQALALTPPLLRPHRLYSHSPKPSSSAAFSAFASASNGAP------------AGRARELHLYNTKSRRKE   83
ec                                                                                    MLKIFNTLTRQKE   13

                                  / Intron 1                    / Intron 2
bv  IFKPKVDG-KVGMYVCGVTAYDFSHIGHARAYVTFDILYRYLKYLGYDVCYVRNFTDVDDKIIARANEVGENPISLSRRFCDEFHQDMSYLHCLPPSVEP  194
at  LFKPKVEG-KVGMYVCGVTAYDLSHIGHARVYVTFDVLLRYLKHLGYEVSYVRNFTDVDDKIIARAKELEEDPISLSRRFCEEFNRDMEQLQCLDPSVQP  175
cs  VFKPKVEG-KVGMYVCGVTAYDLSHIGHARVYVTFDVLYRYLRHLGYEVLYVRNFTDVDDKIIARANELGEDPLNLSRRYCEEFRRDMMYLHCLPPSVEP  189
os  LFQPRVPGGEVGMYVCGVTPYDDSHIGHARAYVAFDVLYRYLRYLDHKVRYVRNFTDIDDKIIARANQLGEDPFSLSKRYSDDFLSDMANLHCLPPSVEP  183
ec  EFKPIHAG-EVGMYVCGITVYDLCHIGHGRTFVAFDVVARYLRFLGYKLKYVRNITDIDDKIIKRANENGESFVAMVDRMIAEMHKDFDALNILRPDMEP  112

                  / Intron 3                                                / Intron 4
bv  RVSDHMPQIIDMIKQIVDNGFAYGVD-GDVYFSVEKFPDYGRLSGRKLEDNRAGERVAIDSRKRNPADFALWKSAKEGEPFWESPWGPGRPGWHIECSAM  293
at  RVSDHIPQIIDLIKQILDNGYAYKVD-GDIYFSVDKFPTYGKLSGRKLEDNRAGERVAVDTRKKHPADFALWKAAKEGEPFWESPWGRGRPGWHIECSAM  274
cs  QVSDHMPQIIDMIKQILDNGYAYSVD-GDVYFNVDKFPEYGQLSGRKLEDNRAGERVSVDSRKKNPADFALWKSAKEGEPFWESPWGPGRPGWHIECSAM  288
os  RVSDHIDQIINMIKQIIDNDCAYAIG-GDVYFSVENFPEYGDLSGRKLDDNRAGERVAVDERKKNPADFALWKAAKDGEPSWDSPWGPGRPGWHIECSAM  282
ec  RATHHIAEIIELTEQLIAKGHAYVADNGDVMFDVPTDPTYGVLSRQDLDQLQAGARVDVVDDKRNPMDFVLWKMSKEGEPSWPSPWGAGRPGWHIECSAM  212

                                                                  / Intron 5
bv  SSVYLGYSFDIHGGGMDLMFPHHENEIVQSCAACCESNIRYWVHNGFVTINSEKMSKSLGNFFTIRQVIELYHPLALRLFLMGTHYRSPINYSDVLLES  392
at  SAAYLGYSFDIHGGGMDLVFPHHENEIAQSCAACDSSNISYWIHNGFVTVDSEKMSKSLGNFFTIRQVIDLYHPLALRLFLMGTHYRSPINYSDFLLES  373
cs  SASYLGYSFDIHGGGMDLVFPHHENEIAQSCAACRTSNVSYWVHNGFVTIDSEKMSKSLGNFFTIRQVIDLYHPLALRLFLLGTHYRSPINYSDLLLES  387
os  SAHYLGHSFDIHGGGEDLIFPHHENEIAQSRAACCDSSINYWIHNGFVNVNSQKMSKSLGNFVTIRKVTELYHPLALRMFLLGTHYRSPINYTIEQLNV  381
ec  NCKQLGNHFDIHGGGSDLMFPHHENEIAQSTCAHDGQYVNYWMHSGMVMVDREKMSKSLGNFFTVRDVLKYYDAETVRYFLMSGHYRSQLNYSEENLKQ  311
```

```
                / Intron 6                                                    / Intron 7
bv  ASERIFYIYQTLQDSQALLSEYGRANLKDSIPQDIVSCINDLQNSFVISMSDDLHTPVVLAALSDPLKTINDLLHTRKGKKRELRVESLAALEKTIKHVL  492
at  ASERIFYIYQTLHDCESALGEKDSTFENGSVPSDTLTSINTFRTEFVASMSDDLLTPVTLAAMSEPLKTINDLIHTRKGKKQARREESLKALETTIRDVL  473
cs  ASDRIFYIYQTLDDCRTVISQEDESSFKGPIAPSLVEEINKFSNVFLTSMSDDIHTPVVLAALSDPLKIINDLLHTRKGKKQEFRMESLAALEKIIGNVL  487
os  ASDRLYYTYQTLQDCEESCQQH-QSKAGDPLPVNTTNCIQKLHDEFETSMSDDLHTSVALAAISEPLKVMNDLLHTRKGKKQEKRLESLSAMEEKIRMVL  481
ec  ARAAVERLYTALRG-----TDKTVAPAGG----------EAFEARFIEAMDDDFNTPEAYSVLFDMAREVN--------RLKAEDMAAANAMASHLRKLS  388

                  / Intron 8
bv  TALGLMPSSYSEALHQLREKALKRVKFSETEVQQKIEERDMARKNKEYEKSDAIRKELAAVGIALMDSPEGTTWRPVVPTALQQEPVMAN*  582
at  TILGLMPTSYSEVLEQLKEKALKRAGLKEEDVLQRVQERTDARKNKEYERSDAIRKDLAKVGIALMDSPEGTTWRPAIPLALQEPVTTTP*  563
cs  SILGLMPASYSEALQQLKEKALTRAKMTNDQVLQKIEERNAARKNKEYEKSDSIRTDLAAVGISLMDGPNGTTWRPTVPLALQEHQASST*  577
os  SVLGLLPSSYYEALQQLREKALRRASMTEEQVLQKIEERTSARKAKQYEKSDEIRKELAAVGIALMDGPDGTTWRPSVPLSEQGVVAST*  570
ec  AVLGLLEQE-PEAFLQSGAQADDSE---VAEIEALIQQRLDARKAKDWAAADAARDRLNEMGIVLEDGPQGTTWRRK*               461
```

b

	Intron 1	Intron 2	Intron 3	Intron 4	Intron 5	Intron 6	Intron 7	Intron 8
bv	80	294	222	570	12	343	939	88
at	102	774	84	86	7	171	77	None
cs	101	1089	231	303	167	387	298	281
os	668	404	449	145	74	889	133	809

Figure S6 Multiple alignments of the amino acid sequences from cysteinyl-tRNA synthetases of sugar beet (bv), *Arabidopsis* (at) (AC006593, also annotated as at2g31170), cucumber (Csa002320 in CuGenDB), rice (Os09g0556500 in RAP) and *E. coli* (ec) (X56234) (**a**). Dashes are incorporated for maximum matching. Two conserved motifs (His-Ile-Gly-His and Lys-Met-Ser-Lys-Ser) that are characteristic of class I aaRS are doubly and singly underlined, respectively. The positions of introns are indicated by slash. Length of the introns are summarized below (in bp) (**b**).