論文の内容の要旨

生産・環境生物学専攻
平成２０年度博士課程　入学
氏名　シュラッダ　ロイ
指導教員名　堤　伸浩

論文題目 **A study of chloroplast gene transfer and substitution by comparative analysis among flowering plants**

（被子植物のゲノム比較による葉緑体遺伝子の転移と置換に関する研究）

It is universally accepted that chloroplasts (cps) evolved via endosymbiosis of a cyanobacteria. After endosymbiosis, massive gene losses that were redundant between the protochloroplast derived from the cyanobacterial endosymbiont and primitive eukaryote cell had occurred, and transfer of essential genes from the ancestor to the nucleus decisively shaped the present plant nuclear and cp genome during evolution. Alternatively, nuclear genes with different organellar origins were replaced with cp-encoded genes, resulting in the present structures of the cp genome. Although such gene transfers and substitutions are important genetic events in evolution, little is known about this process.

It is considered that mitochondria also evolved via endosymbiosis of a α-proteobacteria in eukaryote. More genes (especially ribosomal-protein-encoding genes) are encoded in the mitochondrial genome of angiosperms than in those of vertebrates and fungi. Furthermore, the number of genes encoded in the mitochondrial genome varies among plant species. These clues

suggest that gene transfer is still ongoing in angiosperms. Thus, the mitochondrial genome of angiosperms is a good tool for the study of gene transfer events from the mitochondria to the nucleus and provides a way of understanding the mechanism of gene transfer in eukaryote. Compared with angiosperm mitochondrial genomes, the genome structure and gene content of the cp genome are highly conserved in evolutionary distant lineages. Hence, little is known regarding the events involved in the transfer of genes from the cp to the nucleus. However, several cases of gene loss from the cp genome have been found, because of the increasing number of completely sequenced flowering plant cp genomes. This will allow us to analyze the process of gene transfer from the cp genome to the nucleus.

The present thesis is aimed at understanding the mechanisms underlying evolutionary gene substitution and transfer via the comparative analysis of two genes that encode the cp ribosomal proteins S16 (RPS16) and L32 (RPL32).

I. *rps16* in the cp genome displays a variable status during evolution among *Arabidopsis* and its closely related species

rps16 is generally encoded by two exons (separated by one group II intron) that are located in the cp genomes of flowering plants. However, it has been reported that, in mono- and dicotyledonous plants, the cp-encoded RPS16 protein was replaced by the product of the nuclear-encoded *rps16*, which was transferred from the mitochondria to the nucleus before the early divergence of angiosperms. It is suggested that the present status of *rps16* gene substitution in most angiosperm cp genomes is the intermediate stage.

In this study, I have identified the different functional statuses of *rps16* in several cp genomes in the genus *Arabidopsis* and its close relatives. Eleven complete Brassicaceae cp genomes (*Aethionema grandiflorum*, *Arabis hirsuta*, *Barbarea verna*, *Brassica rapa* subsp pekinensis, *Capsella bursa-pastoris*, *Crucihimalaya wallichii*, *Draba nemorosa*, *Lepidium virginicum*, *Lobularia maritima*, *Nasturtium officinale*, and *Olimarabidopsis pumila*) and the partial chloroplast genome including *rps16*, of *Sinapis alba* are available from current databases. Sequence comparison revealed that the cp-encoded *rps16* genes of four Brassicaceae species

have become pseudogenes, as these genes contain a deletion within their coding sequence in *Arabis hirsuta*, a nonsense mutation in *Aethionema grandiflorum*, and the complete loss of the second exon in *D. nemorosa* and *L. maritima.*

The fact that *Arabis* is phylogenetically close to *Arabidopsis thaliana* raised the possibility that the pseudogenization of cp-encoded *rps16* observed in *Arabis hirsuta* might also occur in the *Arabidopsis* lineage. Further analysis of *Arabidopsis thaliana* and its close relatives (*Arabidopsis arenosa, Arabidopsis lyrata, O. pumila*, and *Crucihimalaya lasiocarpa*) has shown that pseudogenization occurred via the loss of the splicing capacity of the group II intron. The 5′ splice site of the group II intron changed from GUGYG to GUACG in *Arabidopsis thaliana.* However, the splice site consensus sequences were conserved in other closely related species. RT–PCR was conducted to confirm the splicing activity of the group II intron among *Arabidopsis thaliana* and its close relatives. The results revealed that only the primary transcript was amplified in *Arabidopsis thaliana, Arabis hirsuta*, and *O. pumila*, suggesting the loss of the splicing of the intron in these plants. This raised the possibility of the widespread pseudogenization of *rps16* in the angiosperm cp genomes via the loss of its splicing capacity, even when the *rps16* encoded in the cp genome is transcriptionally active.

The estimated time of the divergence of *Arabidopsis thaliana* from all other *Arabidopsis* species is 3.0–5.8 million years ago (mya), and the time of the divergence of the *Arabidopsis* and *Olimarabidopsis* (*Crucihimalaya*) species is estimated at 10–14 mya. This suggests that the independent pseudogenization of the cp-encoded *rps16* in *Arabidopsis thaliana* and *O. pumila* via dysfunctional splicing occurred within the last 5.8 and 14 myr, respectively. The onset of cp-encoded *rps16* gene substitution was minimally estimated in previous works at 140–150 mya. Considering the time of divergence of *Arabidopsis thaliana* from *O. pumila*, the nuclear genome gained an *rps16* copy ~140 mya and the cp and nuclear copies have coexisted (perhaps redundantly) since then. However, in the last 5.8–14 myr, the cp-encoded *rps16* copies have become recognizable pseudogenes in *Arabidopsis thaliana* and *O. pumila.* This suggests that the process of complete gene substitution of cp-encoded *rps16* lasted for over 126 myr in *Arabidopsis thaliana* and *O. pumila.*

Why does the loss of *rps16* from the cp genome seem to have accelerated in evolutionarily recent times in the *Arabidopsis* lineage? It was predicted that the level of inbreeding is positively associated with the level of functional transfer (and loss) of organellar genes. Interestingly, this study revealed that self-compatible plants tend to lose *rps16* from their cp genomes, whereas self-incompatible plants tend to retain *rps16* in their cp genomes. Self-compatibility may be one of the explanations for the acceleration of *rps16* gene loss from the cp genome in Brassicaceae, although the underlying mechanism remains completely unknown.

II. Comparative genomic analysis of gene transfer of chloroplast *rpl32* in *Malpighiales* demonstrates its parallel retention and progressive pseudogenization

rpl32, which was first characterized for its location between *ndhF* and *trnL* on the small single-copy region of the cp genome of *Tobacco*, is generally encoded by the cp genome in flowering plants. However, previous studies in *Malpighiales* revealed that this gene was functionally transferred to the nucleus in *Bruguiera gymnorrhiza* and in the genus *Populus* (*P. alba* and *P. trichocarpa*). In *B. gymnorrhiza*, cp *rpl32* is encoded via alternative splicing at the seventh intron of the cp Cu–Zn superoxide dismutase gene (*sod-1*), encoded in the nucleus. On the other hand, the *sod-1* sequence containing cp *rpl32* is duplicated and subfunctionalized in the *Populus* genus. It has been strongly suggested that cp *rpl32* has acquired the sequence that encodes the transit peptide from the *sod-1* before the divergence of *Malpighiales*.

To confirm the status of cp- and nuclear-encoded cp *rpl32*, comparative genomic analysis of cp *rpl32* was conducted in eight species (*Passiflora citrina*, *Euphorbia sieboldiana*, *Calophyllum inophyllum*, *Acalypha hispida*, *Hypericum erectum*, *Viola mandshurica*, *Manihot esculenta*, and *Ochna serrulata*) from six distinct families of *Malpighiales*.

Genomic PCR was conducted to determine whether *rpl32* was lost from the cp genome of seven species (*O. serrulata*, *C. inophyllum*, *H. erectum*, *E. sieboldiana*, *M. esculenta*, *V. mandshurica*, and *P. citrina*). *rpl32* was found in the cp genomes of *C. inophyllum* and *M. esculenta* and their expression was detected using RT–PCR, suggesting that the active *rpl32* was retained in these two species. However, cp *rpl32* was inactivated in the five remaining species.

Three species (*O. serrulata*, *H. erectum*, and *E. sieboldiana*) lost *rpl32* completely from their cp genomes. Pseudo-*rpl32* was found in the cp genomes of *V. mandshurica* and *P. citrina*. These observations suggest that cp *rpl32* may have been transferred to the nucleus in *P. alba* and in most *Malpighiales* species.

Previous works suggest that the integration of cp *rpl32* into the seventh intron of *sod-1* occurred in the nuclear genome of *Malpighiales*. Genomic PCR was conducted to determine whether cp *rpl32* was integrated into the seventh intron position of *sod-1* in eight species using the primer pairs designed from the conserved *sod-1* and nuclear-encoded cp *rpl32* sequences among *Malpighiales* species. Cp *rpl32* was encoded in the seventh intron position of *sod-1* in six species (*C. inophyllum*, *A. hispida*, *E. sieboldiana*, *M. esculenta*, *V. mandshurica*, and *P. citrina*). RT–PCR was conducted to confirm the expression of nuclear-encoded cp *rpl32* in *C. inophyllum*, *E. sieboldiana*, *M. esculenta*, and *V. mandshurica*. Direct sequencing of each RT–PCR product revealed that all products were amplified from the transferred nuclear cp *rpl32* integrated into the seventh intron position of *sod*-1. This result suggests that *rpl32* gene transfer to the seventh intron of *sod-1* occurred widely in *Malpighiales*, as predicted.

In *M. esculenta*, RT–PCR analysis revealed that the entire sequence of the seventh intron of the *rpl32* mRNA was not spliced out and that five amino acids deduced from the RT–PCR product were deleted in the conserved RPL32 domain. A sequence identical to that of the RT–PCR product of nuclear-encoded cp *rpl32* was detected in the whole-genome sequence of *M. esculenta*, which is available from the Joint Genome Initiative (http://www.jgi.doe.gov/CSP/), and no other cp *rpl32* genes were detected in the nuclear genome. These results suggest that nuclear-encoded cp *rpl32* is inactivated in *M. esculenta*. To the best of my knowledge, this is the first example of the inactivation of a gene transferred from the cp to the nucleus.

The dual expression of cp *rpl32* encoded in the cp and nuclear genomes was observed in *C. inophyllum*. Hence, cp *rpl32* gene transfer was identified as the intermediate stage in *C. inophyllum*. The order *Malpighiales* comprises around 700 genera and over 16,000 species in 30 families. Comparative analysis strongly suggests that the gene transfer of cp *rpl32* to the seventh intron of *sod-1* in the nucleus occurred in the common ancestor of *Malpighiales*. The estimated

time of diversification of *Malpighiales* is around 114 mya. Therefore, the intermediate stage of cp *rpl32* has been ongoing over the past 114 myr in *C. inophyllum*.

The study of two cp ribosomal genes, *rps16* and *rpl32*, allowed the estimation of the period of gene transfer. This study provides novel evidences for the processes involved in gene transfer and substitution from the cp to the nucleus and the requirement of an extra-long period for the successful completion of these processes.