國立臺灣大學生命科學院生態學與演化生物學研究所

博士論文

Institute of Ecology and Evolutionary Biology
College of Life Science
National Taiwan University

Doctoral Dissertation

全寄生植物菱形奴草質體基因組演化之研究

Evolutionary Studies of Plastid Genome of Holoparasitic *Mitrastemon kanehirai*

徐馨怡

Shin-Yi Shyu

指導教授：胡哲明 博士

Advisor: Jer-Ming Hu, Ph.D.

中華民國 102 年 8 月

August, 2013

謝　辭

終於，這段漫長的博士生生涯可以畫下句點了。

首先謝謝我的指導教授胡哲明老師，我想老師在心裡應該也想著「終於！」並且也鬆了一口氣吧！謝謝老師在這段長長的時間裡的指導與照顧，老師不只引領我進入這個研究領域，也擴展了我的生活視野。感謝口試委員邱少婷老師、蕭淑娟老師、鍾國芳老師以及劉少倫老師，撥冗在匆促的時間內審閱論文並提供建議與匡正。

謝謝一路走來與我互相扶持的實驗室夥伴們。一起討論實驗、一起出外採集、一起遭遇實驗的撞牆期，沒有你們的協助我無法順利完成我的實驗。我的玩樂好友們，謝謝你們聽我吐苦水、照顧我這個窮學生，這期間我們一起經歷了許多難忘的體驗。其他對我實驗提供協助的學長姐、一起帶實驗課的助教還有許多未曾謀面卻大方提供資訊的朋友，謝謝你們給與的溫暖。

最後謝謝我的家人們，雖然不太清楚我在做啥，也搞不懂為什麼一個博士班要念這麼久，你們還是盡其所能的協助我以及給與無限的包容與支持。

感謝這一路上所有幫助過我的人～謝謝！

摘　要

非光合作用植物只存在著退化的質體，且其質體基因組已高度的縮減。菱形奴草（*Mitrastemon kanehirai*）為台灣特有的全寄生植物，屬於嚴重瀕臨絕滅的珍貴保育物種。目前菱形奴草只有一條質體序列被發表，而此序列 pt16S rDNA 顯示了演化速度加快的現象。

本論文首先比較了六種 DNA 萃取方法在兩種非光合作用植物日本蛇菰（*Balanophora japonica*）和菱形奴草上的成效。而利用 Barnwell 等人(1998)針對富含黏液的多肉植物發展出的方法所抽出的蛇菰及奴草 DNA，其純度皆能夠進行限制酶切割反應。此外並成功的在菱形奴草運用了 Milligan 在 1989 年所敘述利用高鹽溶液去除細胞核 DNA 的方法，有效的提高其質體 DNA 的含量。

研究顯示除了質體外，異營性植物的細胞核以及粒腺體小次單元 rDNA 也有演化速度加快的現象。為了釐清異營性植物細胞內三種 SSU rDNA 間演化速度的相對關係，本研究檢視了九種異營性植物（包括一種半寄生、五種全寄生以及三種真菌異營性的植物）的粒線體 19S、質體 16S 以及核 18S rDNA；藉由相對速率分析和譜系分析兩種方法來估算這些 rDNA 序列的演化速度。分析顯示與其他被子植物相比較，非光合作用植物的 pt16S 及 nr18S rDNA 序列，其取代率明顯的增高，而 mt19S rDNA 則無此現象。九種被檢視的植物中，日本蛇菰和菱形奴草的 pt16S 和 nr18S rDNA 的變異最大，並且伴隨著序列中 GC 比例的降低。

此外本論文利用了次世代定序技術，成功的獲得了非光合作用植物菱形奴草質體基因組的完整序列。與一般陸生植物的葉綠體基因組相比，菱形奴草的質體基因組喪失了所有與光合作用相關的基因，並欠缺反向重覆區域，而保留下來的基因大多與轉譯作用相關。這個保有 4 個 rRNA、4 個 tRNA 及 18 個蛋白質編碼(protein-coding)基因、大小只有 25,740 bp 的質體基因組為目前已知的最小質體基因組。

關鍵字：非光合作用植物、異營性植物、小次單元 rDNA、演化速度、質體基因組、菱形奴草、質體基因組

ABSTRACT

Non-photosynthetic plants only retain remnant plastids and their plastome is highly reduced. *Mitrastemon kanehirai*, a root holoparasite, is endemic to Taiwan and considered an endangered species. *Mitrastemon kanehirai* has only one plastid sequence has been reported, and the sequence, pt16S rDNA, shows increased substitution rate.

In this dissertation, the performance of six DNA extraction procedures for two non-photosynthetic plants, *Balanophora japonica* and *M. kanehirai*, were compared. All six procedures yielded DNA of sufficient quality for PCR, and the method described by Barnwell et al. (1998) performed well in isolating DNA from both species for restriction enzyme digestion. Meanwhile, enrichment of *M. kanehirai* plastid DNA content was achieved by using the 'high salt' methods based on protocol presented by Milligan (1989).

High rate of nucleotide substitution in three subcellular SSU rDNAs have been reported in heterotrophic plants, and the rate heterogeneity among these sequences are presented in this dissertation. Mt19S, pt16S and nr18S rDNA sequences from nine heterotrophic plants, including one hemiparasitic, five holoparasitic and three mycoheterotrophic plants, were examined. Rate heterogeneity among various rDNA sequences was evaluated by relative rate tests and phylogenetic analysis. Both pt16S and nr18S rDNA sequences of non-photosynthetic species show significant increases of substitution rate, but the phenomenon was not found in mt19S rDNA. The extreme divergent pt16S and nr18S rDNA sequences were found in *B. japonica* and *M. kanehirai*, and accompanied by a decrease in GC content of the sequences.

Mitrastemon kanehieai plastome was sequenced by using next generation sequencing technology. The genome is smallest plastome that have been described with size of 25,740 bp. Only 26 genes were retained in the plastome, which include 4 rRNAs, 4 tRNAs and 18 protein-coding genes. These retained genes are mostly involved in translation machinery. All photosynthesis-related genes were lost, and the inverted repeat region is absent. Despite the enormous reduction, the *M. kanehirai* plastome is a functional gene expression system. DNA transfer from plastid to nucleus and horizontal transfer from the host to the parasite were also observed in *M. kanehirai*.

Keywords: heterotrophic plant, *Mitrastemon kanehirai*, non-photosynthetic plant, plastome, small-subunit rDNA, substitution rate.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. Overview

The heterotrophic plants

The heterotrophic plants are those having all or some of resources required to support their vital processes and can be classified as either mycoheterotrophs or haustorial parasites (Leake, 1994; Nickrent, 2002; Irving and Cameron, 2009; Selosse and Cameron, 2010; Těšitel et al., 2011). Mycoheterotrophs obtain their nutrition from the associated plant via a mycorrhizal fungus whereas haustorial parasites feed directly on another plant (the host) via haustoria (Kuijt, 1969; Nickrent, 2010; Selosse and Cameron, 2010). Noted that not all of the haustorial parasites are non-photosynthetic, and they can be categorized into two groups, hemiparasites and holoparasites, based on the ability to perform photosynthesis. Hemiparasites are chlorophyllous and photosynthetic during at least one stage of their life cycle, and they obtain water and nutrients from the host xylem. Some advanced hemiparasites (e.g. dwarf mistletoes) also obtain photosynthates from the host phloem. Holoparasites, on the other hands, are totally achlorophyllous (or nearly so), non-photosynthetic, and must rely on their host for water and nutrients from the host xylem and photosynthates from the host phloem (Nickrent, 1997; Nickrent et al., 2000; Nickrent, 2002; Heide-Jørgensen, 2008; Irving and Cameron, 2009; Nickrent, 2010). The term ‘non-photosynthetic plants’, refer to both mycoheterotrophs and holoparasitic plants that completely lost photosynthetic ability.

It has been reported that the parasitic lifestyle has evolved at least 11 times in flowering plants (Barkman et al., 2007). There are about 4,400 parasitic species of flowering plants, which consist about 1% of angiosperms (Heide-Jørgensen, 2008; Nickrent, 2010). Classifications of heterotrophic plants based on morphological characters have long been difficult because of the reduction of their morphological features and frequent convergence on character evolution. In addition, phylogenetic reconstruction of heterotrophic plants faces great challenges with their highly divergent ptDNA sequences. Furthermore, non-photosynthetic plants lost their photosynthetic ability and genes related to photosynthesis were either lost or pseudogenized. Therefore, it makes even more difficult to select suitable molecular markers for non-photosynthetic plants.

Phylogenetic study in heterotrophic plants

Two plastid genes, *rps2* and *rbcL*, have been used to examine the phylogeny of some parasitic taxa (dePamphilis et al., 1997; Nickrent et al., 2000). Since these two genes might be lost in non-photosynthetic plants, *rps2* and *rbcL* are not proper molecular markers for non-photosynthetic plants. The plastid 16S and nuclear 18S rDNA sequences also have been utilized in phylogenetic studies of parasitic plant and the results show evolutionary rates of the two rDNAs are increased in some parasitic plants (Nickrent and Starr, 1994; Nickrent and Duff, 1996; Nickrent et al., 1997; Nickrent et al., 2000; Lemaire

et al., 2010; Su and Hu, 2012). The mitochondrial sequences have been shown as suitable markers for reconstructing the phylogeny of parasitic plants (Barkman et al., 2004; Nickrent et al., 2004). However, massive mitochondrial DNA transfers in parasitic plants also have been reported (Xi et al., 2012; Xi et al., 2013), and therefore the usage of mitochondrial sequences as molecular markers should be very careful in these taxa.

Cuscuta is one of the most intensely studied genera of parasitic plants, and several *Cuscuta* plastomes have be completely sequenced (Revill et al., 2005; Stefanović and Olmstead, 2005; Funk et al., 2007; McNeal et al., 2007; Braukmann et al., 2013). The size of the *Cuscuta* plastome appears to be correlated with the ability of photosynthesis. The genus *Cuscuta* represents an evolution process from hemiparasites to holoparasites, with some of the *Cuscuta* species still retain partial photosynthetic ability. In addition, several heterotrophic plant plastomes, including some mycoheterotrophic orchids, have been described recently. All these plastome data enable us to compare the plastid sequences, gene content and gene order in heterotrophic plants. By analyzing and comparing the plastome sequence, it will give us more insight into the evolution of plastid genome in heterotrophic plants.

Mitrastemon are root endoparasites distributed in Central America, East and Southeast Asia (Yamamoto, 1936; Meijer and Veldkamp, 1993). Previous studies were mostly focused on morphological observation. *Mitrastemon* has long been thought to

belong to the Rafflesiaceae (Hayata, 1913; Kuijt, 1969; Meijer and Veldkamp, 1993; Bouman and Meijer, 1994; Mabberley, 1997). However, recent studies have placed *Mitrastemon* (Mitrastemonaceae) in Ericales, but not Rafflesiaceae (in Malpighiales) according to their molecular phylogenetic analyses based on nuclear 18S rDNA and mitochondrial *mat*R sequences (Fig. 1-1) (Barkman et al., 2004; Nickrent et al., 2004). There is only one plastid sequence of *Mitrastemon* (16S rDNA) available in GenBank. The 16S sequence showed higher accelerated substitution rate than the pt16S of *Epifagus virginiana*, which is a holoparasite with a reduced plastome of size 70 kb (Nickrent et al., 1997; Nickrent et al., 2000).

Two *Mitrastemon* species, *M. kanehirai* and *M. kawasasakii* were found in Taiwan, both were reported endemic to Taiwan (Yang and Lu, 1996). In the two species, *M. kanehirai* is considered an endangered species and only found in few limited localities with few individuals within populations.

This dissertation focuses on the plastome of *Mitrastemon kanehirai* (Fig. 1-2), a non-photosynthetic plant that is not close to any other heterotrophic plants with complete plastome sequence. By analyzing and comparing the *M. kanehirai* plastome sequence with other plants, we hope it will help us further understand the evolution of plastid genome in non-photosynthetic plants.

In order to obtain total genomic DNA and plastid DNA with high quality from *M.*

kanehirai, I have tested several DNA extraction and plastid isolation procedures. Only some of the DNA extraction procedures successfully isolated DNA from *M. kanehirai*. Although the attempt of isolating pure and intact *M. kanehirai* plastid failed, I indeed succeeded to obtain enriched plastid DNA from *M. kanehirai* DNA preparation. The comparison of different DNA extraction procedures and the application of ptDNA enrichment protocols are presented in Chapter 2.

Previous studies have documented accelerated evolutionary rates in heterotrophic plants for nuclear 18S rRNA genes (Nickrent and Starr, 1994; Nickrent and Duff, 1996; Lemaire et al., 2010) and plastid 16S rRNA genes (Nickrent et al., 1997; Nickrent et al., 2000); however, high evolutionary rate is not a ubiquitous phenomenon in small-subunit rDNAs of heterotrophic plants. Therefore in Chapter 3, I extended my survey examining mt19S, pt16S and nr18S rDNA sequences from nine heterotrophic plants in order to examine whether there is any correlation of evolutionary rate patterns among the three subcellular SSU rDNAs in heterotrophic plants.

In Chapter 4, the theme of this dissertation, we sequenced the complete plastid genome of *M. kanehirai* by using next generation sequencing technology. The plastome sequence was analyzing and compared with other plastomes to improve our understanding of plastome evolution in non-photosynthetic plants.

Lastly, a brief summary based on the results from each chapter is given in Chapter 5.

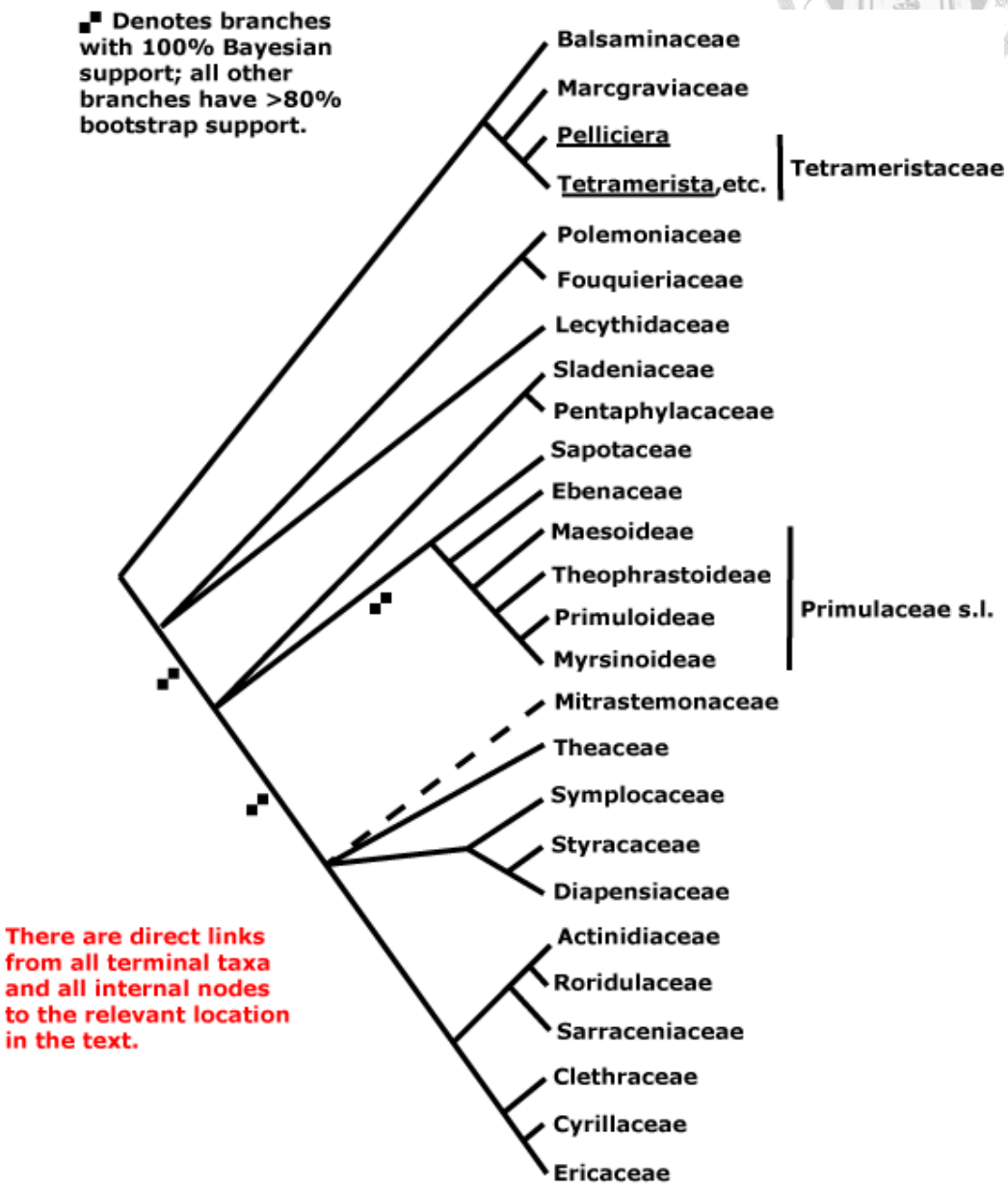


Figure 1-1. The phylogenetic tree of Ericales according to the APG III system (Stevens, 2001 onwards).

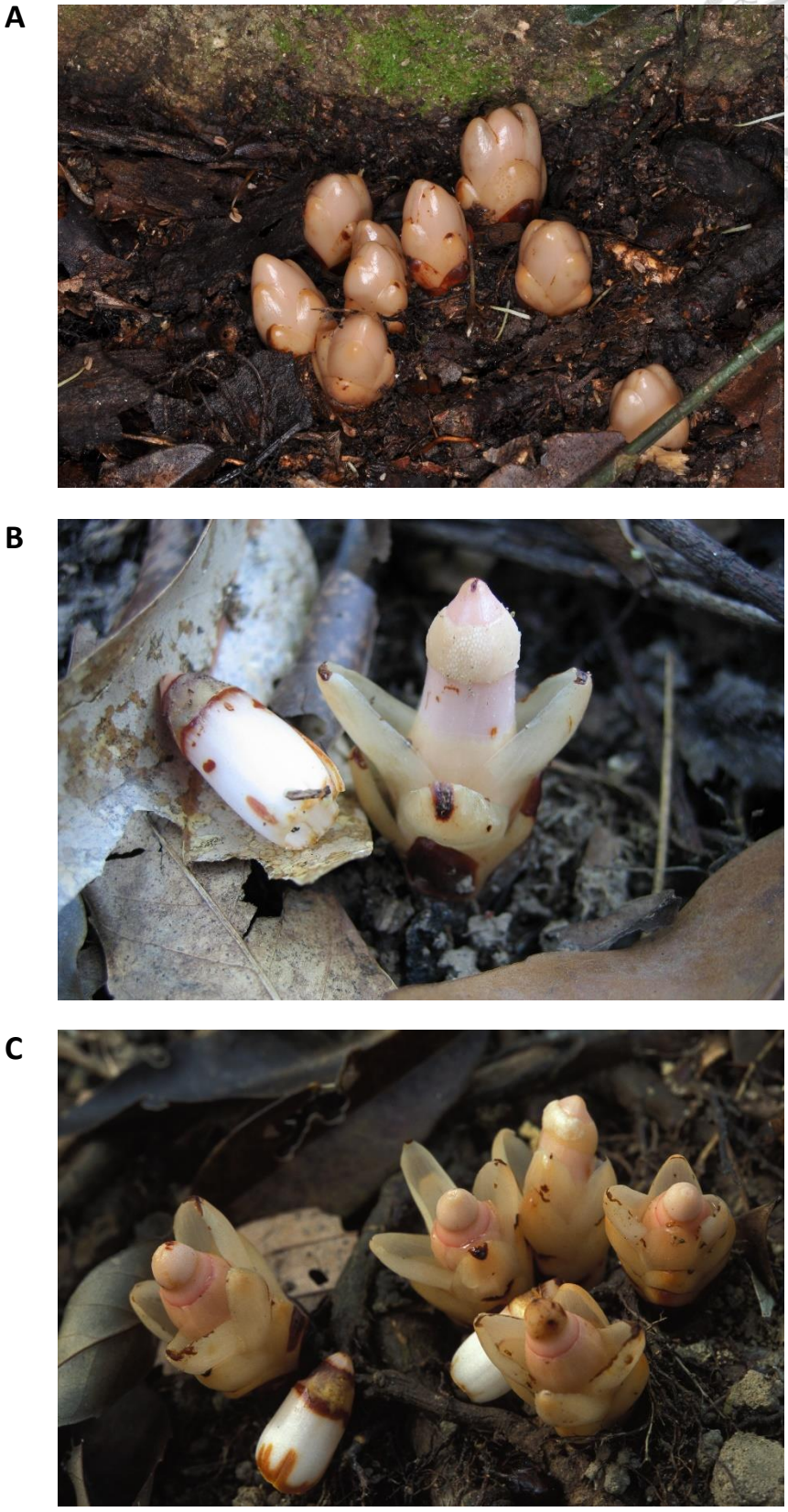
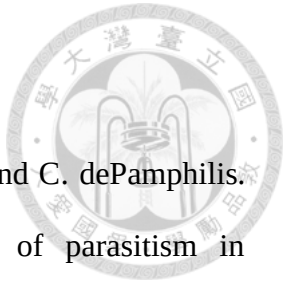


Figure 1-2. Photos of *Mitrastemon kanehirai*. A young stage, B male stage and C female stage. Photos were taken by Jer-Ming Hu.



References

Barkman, T., J. McNeal, S.H. Lim, G. Coat, H. Croom, N. Young and C. dePamphilis. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol Biol. 7: 248.

Barkman, T.J., S.H. Lim, K.M. Salleh and J. Nais. 2004. Mitochondrial DNA sequences reveal the photosynthetic relatives of *Rafflesia*, the world's largest flower. Proc Nat Acad Sci USA. 101: 787-792.

Bouman, F. and W. Meijer. 1994. Comparative structure of ovules and seeds in Rafflesiaceae. Plant Syst Evol. 193: 187-212.

Braukmann, T., M. Kuzmina and S. Stefanovic. 2013. Plastid genome evolution across the genus *Cuscuta* (Convolvulaceae): two clades within subgenus *Grammica* exhibit extensive gene loss. J Exp Bot. 64: 977-89.

dePamphilis, C.W., N.D. Young and A.D. Wolfe. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: many losses of photosynthesis and complex patterns of rate variation. Proc Nat Acad Sci USA. 94: 7367-7372.

Funk, H., S. Berg, K. Krupinska, U. Maier and K. Krause. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. BMC Plant Biol. 7: 45.

Hayata, B. 1913. On the systematic position of *Mitrastemon*, as a genus representing a special tribus of the Rafflesiaceae. Icones Plantarum Formosanarum. 3: 199-213.

Heide-Jørgensen, H.S. 2008. Parasitic Flowering Plants, Koninklijke Brill NV, Leiden, The Netherlands.

Irving, L.J. and D.D. Cameron. 2009. You are what you eat: interactions between root parasitic plants and their hosts. *In* K. Jean-Claude, and D. Michel, (eds.).

Advances in Botanical Research. Academic Press, pp. 87-138.

Kuijt, J. 1969. Rafflesiaceae, Hydnoraceae, and Balanophoraceae. The Biology of Parasitic Flowering Plants. University of California Press, pp. 118-133.

Leake, J.R. 1994. The biology of myco-heterotrophic ('saprophytic') plants. New Phytol. 127: 171-216.

Lemaire, B., S. Huysmans, E. Smets and V. Merckx. 2010. Rate accelerations in nuclear 18S rDNA of mycoheterotrophic and parasitic angiosperms. J Plant Res. 124: 561-576.

Mabberley, D.J. 1997. The plant-book, 2nd. Cambridge University Press, Cambridge.

McNeal, J., J. Kuehl, J. Boore and C. de Pamphilis. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. BMC Plant Biol. 7: 57.

Meijer, W. and J.F. Veldkamp. 1993. A revision of *Mitrastema* (Rafflesiaceae). Blumea. 38: 221-229.

Nickrent, D.L. 1997. The parasitic plant connection. http://www.parasiticplants.siu.edu/.

Nickrent, D.L. 2002. Parasitic plants of the world. In J.A. López-Sáez, P. Catalán and L. Sáez (eds.), Parasitic plants of the Iberian Peninsula and Balearic Islands. Mundi-Prensa Libros, S. A., Madrid, Spain, pp. 7-27.

Nickrent, D.L., A. Blarer, Y.L. Qiu, R. Vidal-Russell and F.E. Anderson. 2004. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. BMC Evol Biol. 4: 40-56.

Nickrent, D.L. and R.J. Duff. 1996. Molecular studies of parasitic plants using ribosomal RNA. *In* M.T. Moreno, J.I. Cubero, D. Berner, D. Joel, L.J. Musselman, and C. Parker, (eds.). Advances in Parasitic Plant Research,. pp. 28-52.

Nickrent, D.L., R.J. Duff, A.E. Colwell, A.D. Wolfe, N.D. Young, K.E. Steiner and C.W. dePamphilis. 2000. Molecular phylogenetic and evolutionary studies of parasitic

plants. *In* D.E. Soltis, P.S. Soltis, and J.J. Doyle, (eds.). Molecular Systematics of Plants II: DNA sequencing. Kluwer Academic Publishers, Norwell, Massachusetts, USA, pp. 211-241.

Nickrent, D.L., R.J. Duff and D.A.M. Konings. 1997. Structural analyses of plastid-derived 16S rRNAs in holoparasitic angiosperms. Plant Mol Biol. 34: 731-743.

Nickrent, D.L., Musselman L. J. 2010. Introduction to parasitic flowering plants. http://www.apsnet.org/edcenter/intropp/PathogenGroups/Pages/ParasiticPlants.aspx.

Nickrent, D.L. and E.M. Starr. 1994. High-rates of nucleotide substitution in nuclear small-subunit (18S) rDNA from holoparasitic flowering plants. J Mol Evol. 39: 62-70.

Revill, M.J.W., S. Stanley and J.M. Hibberd. 2005. Plastid genome structure and loss of photosynthetic ability in the parasitic genus *Cuscuta*. J Exp Bot. 56: 2477-2486.

Selosse, M.A. and D.D. Cameron. 2010. Introduction to a virtual special issue on mycoheterotrophy: New Phytologist sheds light on non-green plants. New Phytol. 185: 591-593.

Stefanović, S. and R. Olmstead. 2005. Down the slippery slope: plastid genome evolution in Convolvulaceae. J Mol Evol. 61: 292-305.

Stevens, P.F. 2001 onwards. Angiosperm Phylogeny Website. Version 12, July 2012. http://www.mobot.org/MOBOT/research/APweb/.

Su, H.J. and J.M. Hu. 2012. Rate heterogeneity in six protein-coding genes from the holoparasite *Balanophora* (Balanophoraceae) and other taxa of Santalales. Ann Bot. 110: 1137-1147.

Těšitel, J., J. Lepš, M. Vráblová and D.D. Cameron. 2011. The role of heterotrophic carbon acquisition by the hemiparasitic plant *Rhinanthus alectorolophus* in seedling establishment in natural communities: a physiological perspective. New

Phytol. 192: 188-199.

Xi, Z., R. Bradley, K. Wurdack, K. Wong, M. Sugumaran, K. Bomblies, J. Rest and C. Davis. 2012. Horizontal transfer of expressed genes in a parasitic flowering plant. BMC Genomics. 13: 227.

Xi, Z., Y. Wang, R.K. Bradley, M. Sugumaran, C.J. Marx, J.S. Rest and C.C. Davis. 2013. Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genet. 9: e1003265.

Yamamoto, Y. 1936. Species nova Mitrastemonacearum (Rafflesiacearum) ex Mexico. Botanical Magazine (Tokyo). 50: 539-541.

Yang, Y.P. and S.Y. Lu. 1996. Rafflesiaceae. *In* Editorial Committee of the Flora of Taiwan, (eds.). Flora of Taiwan, second edition. Editorial Committee of the Flora of Taiwan, Taipei, Taiwan, pp. 652-655.

Chapter 2.

Formatted for submission to *Taiwania* (Accepted July 03, 2013)

Comparison of six DNA extraction procedures and the application of plastid DNA enrichment methods in selected non-photosynthetic plants

Shin-Yi Shyu[1] and Jer-Ming Hu[1]

[1]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taiwan

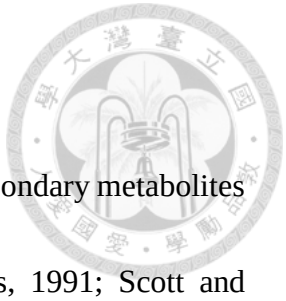
Authors' contributions:

Shin-Yi Shyu designed and performed the experiments, analyzed data and wrote the manuscript. Jer-Ming Hu administered the experiments and edited the manuscript.

Abstract

Genomic DNA was isolated using three DNA extraction commercial kits and three CTAB-based methods for two non-photosynthetic plants, *Balanophora japonica* and *Mitrastemon kanehirai*. The quality of the isolated DNA was evaluated and subjected to following restriction enzyme digestions. All six procedures yielded DNA of sufficient quality for PCR, and the method described by Barnwell et al. (1998) performed well in isolating DNA from both species for restriction enzyme digestion. In addition, we succeeded to enrich plastid DNA content by using the methods depending on a high salt buffer to deplete nuclear material. The 'high salt' methods based on protocol presented by Milligan (1989) were able to increase plastid DNA effectively and significantly reduce nuclear DNA from *M. kanehirai*. The plastid DNA enrichment protocols are inexpensive and not time-consuming, and may be applicable to other non-photosynthetic plants.

Keywords: CTAB, DNA isolation, heterotrophic plants, plastid DNA, polysaccharide.



Introduction

Many heterotrophic plants contain polysaccharides and other secondary metabolites that interfere with DNA isolations (Hayata, 1913; Do and Adams, 1991; Scott and Playford, 1996; Nickrent et al., 2000; Tsai et al., 2008; Yu et al., 2010; Wang et al., 2012). These compounds sometimes prevent enzymes to access DNA, and therefore inhibiting follow-up experiments such as polymerase chain reaction (PCR) or restriction enzyme digestions. Many DNA extraction methods, including most commercial kits, are generally designed for cultivated species, which contain much less interfering compounds for isolating DNAs, and therefore might be inapplicable for heterotrophic plants (Do and Adams, 1991; Scott and Playford, 1996). Previous studies (Nickrent et al., 1997a; Nickrent et al., 1997b) showed that the CTAB-based method described by Nickrent (1994) can successfully extract DNA from non-photosynthetic plants with quality good enough for PCR, but is insufficient for enzyme digestions.

We have applied the common DNA extraction methods (Doyle and Doyle, 1987) to the non-photosynthetic parasitic *Balanophora* species, and it is generally fine to obtain good quality DNA for PCR amplification (Su and Hu, 2012; Su et al., 2012). However, the *Balanophora* DNA from such methods sometimes failed to perform well in enzyme digestions in our preliminary surveys. Furthermore, ordinary DNA extraction methods cannot guarantee to have enough plastid DNA since very few plastids are present in the

cells of the non-photosynthetic plants (dePamphilis and Palmer, 1990; Nickrent et al., 1997b; Nickrent et al., 2000). All available methods for plastid DNA isolation are designed for isolating ordinary non-reduced chloroplasts in green plants, for example, the gradient-based methods for plastid isolation (Kolodner and Tewari, 1975; Palmer, 1986), and others to enrich organelles (e.g. Herrmann, 1982; Palmer, 1982; Bookjans et al., 1984; Milligan, 1989; Triboush et al., 1998; Kausch et al., 1999). The isolation methods with DNAase-I treatment postulated by Herrmann (1982) require a large amount of plant samples, which is also unpractical in our study, since the plant materials are usually limited.

However, some methods (e.g. Milligan, 1989; Triboush et al., 1998) have been used without the requirement on the color of materials and just need a small amount of tissue. The Milligan's procedure (1989) depends on a high salt buffer to solubilize nuclear material in order to obtain a chloroplast fraction, and Triboush *et al*. (1989) isolated organelles by differential centrifugation. These methods are thus have potentials for plastid isolation in non-photosynthetic plants.

In our preliminary survey on 12 DNA extraction methods for two non-photosynthetic plants, *Balanophora japonica* Makino and *Mitrastemon kanehirai* Yamamoto, six of them showed promising results, while the others performed badly, with low or no yield of DNA (see supplementary data Table S2-1). In this report, we compared

the performance of the six procedures that successfully isolated DNA from these two non-photosynthetic plants and evaluated with the following PCR and restriction enzyme digestion. The procedures include three DNA extraction commercial kits, the two methods mentioned above and another CTAB-based method that Barnwell et al. (1998) developed for the highly mucilaginous succulent plants. At the same time, we attempted to enrich plastid DNA content during extraction of *M. kanehirai* DNA for studying its plastid genome by using and modifying the Milligan and Triboush methods. The results were compared with the proportion of plastid vs. nuclear DNA content among the three different plastid enrichment methods.

Materials and methods

Plant Materials

Balanophora japonica and *M. kanehirai* are both non-photosynthetic plants native to Taiwan. The materials (*B. japonica*: Mt. Datong, Taipei County, Sep. 29, 2005, *Hu1567*; *M. kanehirai*: Lienhuachih, Nantou County, Oct. 12, 2010, *Hu1810*) were freshly collected and stored in -20℃. Frozen tissues were used for each DNA extraction method.

DNA extraction methods

The methods used for comparison are listed in Table 2-1, including three commercial kits (*Method 1-3*) and three CTAB-based methods (*Method 4-6*).

Total genomic DNAs were extracted by *Method 1-3* according to the corresponding manufacturer's protocols. *Method 1*, the DNeasy Plant Mini Kit (QIAGEN), uses the QIAshredder spin column to remove initial precipitates and cell debris, and a DNeasy column to capture DNA. *Method 2*, the Tri-Plant Genomic DNA Reagent Kit (Geneaid), and *Method 3*, the TRI Reagent (Molecular Research Center), both use their own particular reagents to lyse plant samples and then follow by an isopropanol or ethanol precipitation.

Method 4 denoted for the standard CTAB method described by Doyle and Doyle (1987). Plant materials were ground in liquid nitrogen and then incubated in 10 volumes of preheated 2× CTAB buffer (100 mM Tris-HCl, 1.4 M NaCl, 20 mM EDTA, 2% w/v CTAB, 2% w/v PVP40 and added 0.2% β-mercaptoethanol just before use.) at 65℃ for 1 h with occasional swirling. The solution was mixed with 10 mL of Chloroform:isoamyl alcohol (24:1, v/v) and was blended thoroughly. This was followed by a centrifugation at 9,000 g for 10 min, and the aqueous phase was transferred to a new centrifuge tube. The DNA was precipitated by adding 0.7 volume of isopropanol and incubated at -20℃ for 30 min. The DNA pellet was collected by a centrifugation for 10 min at 10,000 g and washed with cold 75% ethanol. The pellet was resuspended in 2 mL of TE buffer, and then RNase digestion was performed.

Method 5 denoted for the "delayed hot CTAB" method described by Nickrent (1994).

The sample was cut into small pieces and homogenized with hot 95℃ CTAB buffer (about 25 mL for every 2-3 g of plant tissue). The modified 2× CTAB buffer is composed of 100 mM Tris-HCl, 1.4 M NaCl, 30 mM EDTA, 2% w/v CTAB, 5 mM ascorbic acid, 4 mM diethyldithiocarbamic acid and 2% w/v PVP40, the latter two ingredients were added just before use. The extract was strained through cheesecloth into 50-mL centrifuge tube and then incubated at 70-80℃ for 30 min with occasional swirling. The sample was briefly centrifuged without pausing, and the supernatant was transferred to a new tube. Chloroform:isoamyl alcohol (24:1) (0.7 volume) was added and the solution was mixed for 5 min. This was then centrifuged at 9,000 g for 15 min, and the aqueous phase was transferred to a new centrifuge tube. The DNA was precipitated by adding 0.7 volume of ice-cold isopropanol and incubated at -20℃ for at least 1 h. The DNA pellet was collected by a centrifugation for 20 min at 10,000 g. Then the pellet was resuspended in 3 mL of TE and 2 mL of 4 M NH_4OAc. This was followed by an extraction with an equal volume of phenol:chloroform (1:1), and 2 volumes of ethanol were added to the aqueous phase. The content was incubated at -20℃ for at least 30 min, and the DNA pellet was collected by a centrifugation, then proceeded an RNase treatment.

Method 6 denoted for the extraction procedure developed by Barnwell et al. (1998) for the highly mucilaginous succulent plants. The frozen plant tissue was ground to powder, 5 volumes of extraction buffer (100 mM Tris-HCl, 1.4 M NaCl, 20 mM Na_2

EDTA, 2% w/v CTAB, 1% w/v PVP40) were added and mixed. The homogenate was then incubated at 65℃ for 30 min with occasional shaking followed by a centrifugation at 3,000 g for 5 min. The supernatant was mixed with 1.25 volumes of 10% CTAB (w/v, in 0.7 M NaCl), and the mixture was vortexed for 10 s and centrifuged at 3,000 g for 5 min. The supernatant was thoroughly mixed with 3 volumes of precipitation buffer (50 mM Tris-HCl, 10 mM Na_2 EDTA, 1% w/v CTAB). The mixture was incubated at room temperature for 30 min and then centrifuged at 5,000 g for 15 min. The pellet was dissolved in high salt TE buffer (10 mM Tris-HCl, 1.0 M NaCl, 1 mM Na_2 EDTA), and 2 volumes of ice-cold ethanol were added followed by incubation at -20℃ for 1 h. The DNA was pelleted by a centrifugation and washed twice with 70% ethanol.

Plastid DNA enrichment methods

Three plastid enrichment methods were analyzed in this study. *Method PE1* denoted for the procedure developed by Milligan (1989) that incorporates several other methods of extracting chloroplast DNA. The procedure is summarized below. The tissue was ground with 6 volumes of ice-cold isolation buffer (50 mM Tris-HCl, 1.25 M NaCl, 5 mM EDTA, 0.1% w/v BSA, 0.1% β-mercaptoethanol), and the homogenate was filtered through 4 layers of cheesecloth. The plastids were pelleted by a centrifugation at 3,000 g for 10 min and then resuspended in 10 mL of cold isolation buffer. The centrifugation and resuspension were repeated once, and 0.1 volume of 10% CTAB was added to lyse the

plastids. The extract was then incubated at 60℃ for 1 h and followed by extraction with Chloroform:isoamyl alcohol (24:1). The DNA was precipitated by adding 0.7 volume of cold isopropanol to the aqueous phase and incubated at -20℃ for at least 30 min. During the isolation of plastids, the materials should be kept at 4℃.

Method PE2 denoted for a method combined with Milligan's protocol (1989) and the Tri-Plant Genomic DNA Reagent Kit. Plastid pellet was isolated by centrifugation at 6,000 g for 20 min instead of 3,000 g for 10 min following Milligan's protocol. The DNA was then extracted from pellet by using Tri-Plant Genomic DNA Reagent Kit.

Method PE3 denoted for the method mainly based on Triboush's DNA extraction method (Triboush et al., 1998), combined with Milligan's protocol (1989), as described below. All the operations of isolating plastids were conducted in ice. The sample was homogenized with 6 volumes of STE buffer (50 mM Tris-HCl, 400 mM sucrose, 20 mM Na_2EDTA, 0.2% w/v BSA, 0.2% β-mercaptoethanol). The homogenate was filtered through 4 layers of cheesecloth and centrifuged at 200 g for 20 min. The supernatant was centrifuged at 3,700 g for 20 min, and the pellet was resuspened in 20 mL of isolation buffer (based on Milligan's protocol). It was then repeated the centrifugation and resuspension once, and then the DNA was obtained by following Milligan's protocol.

Real-time PCR

Real-time PCR and data analysis were performed in the CFX96^{TM} Real-Time PCR

Detection Systems (Bio-Rad Laboratories). Sequences of nuclear and plastid SSU fragments were amplified from *M. kanehirai* DNA extracted by different methods. The primers SSU1594F: 5'-CTACGTCCCTGCCCTTTGTA-3' and SSU1703R: 5'-GGACTTCTCGCGGCATCACGAG-3' were used to amplify a nuclear 18S rDNA fragment; the primers 16S298F: 5'-GGAAACAGCCCAGATCATCA-3' and 16S436R: 5'-GCCGACATTCTCACTTCTGC-3' were used to amplify the plastid 16S rDNA. The primers were designed based on *M. kanehirai* sequences in our preliminary survey. The PCR mixture (20 μL) contained 10 μL KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems), 50 nM (nr18S rDNA) or 100 nM (pt16S rDNA) of each primer and 20 ng of extracted DNA was used as template. The amplification program initiated at 95℃ for 3 min, followed by 40 cycles at 95℃ for 10 s and 64℃ for 30 s, and finally 95℃ for 10 s. Melting curve analysis was carried out after amplification. All experiments were performed in triplicate.

Results and discussion

Performance of different isolation procedures

Balanophora japonica and *M. kanehirai* both lost their photosynthetic ability completely and plants are very rich in polysaccharides and other secondary metabolites (Hayata, 1913; Wang et al., 2012). Many of the commercial kits and methods failed to

extract high quality DNA from heterotrophic plants (listed in supplementary data Table S2-1). Here we show the six methods that can successfully extract the DNAs from these two non-photosynthetic plants.

The ratios of the absorbance at 260 and 280 nm ($A_{260/280}$) of DNA isolated from two species with different procedures are in the range of 1.28-2.11 (Table 2-2). Two of the three CATB-based methods (*Method 4* and *6*) yielded better results than all the others, including the commercial kits, for *B. japonica*. In comparison, *Method 2* and *6* worked best for *M. kanehirai*.

Only *Method 6* (Barnwell et al., 1998) performed well on both *B. japonica* and *M. kanehirai*, with acceptable $A_{260/280}$ ratio. However, *Method 2* and *4* showed inconsistent results between *B. japonica* and *M. kanehirai*. Nonetheless, the quality of all DNA isolated by different procedures were good enough for the following PCR (data not shown).

To further examine the DNA quality, we proceeded with restriction enzyme digestion on the obtained DNA extracts. The result shows that the *Method 6* (Barnwell protocol) performed better than the other procedures in extracting DNA from *B. japonica* (Fig. 2-1A), and *Method 2* (the Tri-Plant Genomic DNA Reagent Kit) performed best in *M. kanehirai* (Fig. 2-1B). In general, all procedures performed better in *M. kanehirai* which is likely because it contains less polysaccharides than *B. japonica*. However, *Method 3*

(the TRI Reagent) and *Method 4* (the Doyle and Doyle protocol) could not produce DNA bands for *M. kanehirai* (M3 and M4 in Fig. 2-1), compared to *B. japonica*. From all of the commercial kits tested, we found that most kits extracted DNA by using columns did not perform well. It is probably because these kits were unable to eradicate polysaccharides that prevent the elution of DNA from columns and result in the low yields (Fleischmann and Heubl, 2009). Among the three CTAB-based DNA isolation methods, it seems that shortening the initial incubation time and increasing the incubation temperature could improve DNA purity, since the *Method 5* (Nickrent protocol) performed better than *Method 4*. However, *Method 6* that Barnwell et al. (1998) developed increases CTAB concentration in a step-wise manner in order to avoid co-precipitation of polysaccharides with DNA, which was the most effective procedure to obtain DNA with high quality. As for yields, all the commercial kits produced more amount of DNA than the CTAB-based methods. The *Method 6* produced least but the purest DNA among all procedures, which might result from the protocol's additional purification steps and less efficient precipitation buffer.

Plastid DNA enrichment

The Milligan's 'high salt buffer' method (1989) and the Triboush's differential centrifugation method (1998) were used and modified to enrich plastid DNA content for extracting *M. kanehirai* DNA. The performance of these procedures was evaluated by

relative quantification of nr18S and pt16S rDNA in the DNAs extracted by different methods with Barnwell protocol as the reference. DNA isolated by all methods (included the Barnwell protocol) could amplify nr18S and pt16S rDNA fragments by real-time PCR (Fig. 2-2). However, *Method PE3* (Triboush-based method) not only failed to enrich plastid DNA in our tests but also yielded the lowest quality DNA that the *Cq* (quantification cycle) values of *PE3* were the largest among all methods in both nr18S and pt16S rDNA. The two Milligan-based protocols (*Method PE1* and *PE2*) were capable of increasing plastid DNA content more than 1.5 times and meanwhile reduced the proportion of nuclear DNA effectively (Fig. 2-3). The pt16S/nr18S rDNA ratios of extracted DNA by using these two procedures are significantly higher than *Method 6* (Fig. 2-4). The best performed method was *Method PE2*, the Milligan protocol combined with the Tri-Plant Genomic DNA Reagent Kit, which yielded pt16S/nr18S rDNA ratio almost 15 times higher than *Method 6* with good quality of DNA from *M. kanehirai*. It suggests that increasing centrifugal speed in the beginning step of collecting plastids could enrich plastid DNA further since *Method PE3* performed better than *Method PE2* (Fig. 2-3 and Fig. 2-4).

The result shows that the procedure based on differential centrifugation failed to apply to our studying materials. Although the high salt buffer-based protocols could not eliminate nuclear DNA completely, they still could enrich plastid DNA content

significantly (*Method PE2 and PE3*). These protocols are inexpensive and not time-consuming, and may be applicable to other non-photosynthetic plants, which will be useful in studying the plastid genome of heterotrophic plants.

Overall, the Barnwell protocol performed best among all examined methods, but it is inapplicable for small amount of plant samples. However, our results suggest that there is no DNA isolation protocol can be applied to all plants because of the presence of various compounds in the plant tissue and it cannot have DNA with both the highest quality and quantity from the same protocols. Additional effort to find out or modify isolation procedures is necessary in order to obtain high quality DNA for non-photosynthetic plants.

Table 2-1. The methods used in this study.

Method		Reference
DNA isolation		
Method 1	Column-based commercial kit	DNeasy Plant Mini Kit, QIAGEN, Manchester, UK
Method 2	Particular reagent- based commercial kit	Tri-Plant Genomic DNA Reagent Kit, Geneaid, New Taipei City, Taiwan
Method 3	Particular reagent- based commercial kit	TRI Reagent – RNA, DNA, protein isolation reagent, Molecular Research Center, Cincinnati, OH, USA
Method 4	The standard CTAB method	Doyle and Doyle, 1987
Method 5	The delayed hot CTAB method	Nickrent, 1994
Method 6	The increased CTAB method	Barnwell et al., 1998
Plastid DNA enrichment		
PE1	Depleted nuclear material by using a high salt buffer	Milligan, 1989
PE2	Combined *PE1* with *Method 2*	Milligan, 1989
PE3	Isolated organelles by differential centrifugation	Milligan, 1989; Triboush *et al*., 1998

Table 2-2. $A_{260/280}$ ratios of DNA extracted by commercial kits and CTAB-based methods.

Procedure	$A_{260/280}$	
	B. japonica	*M. hanehirai*
Kit		
Method 1	1.75	1.76
Method 2	1.28	2.08
Method 3	1.91	1.48
CTAB-based Method		
Method 4	2.06	1.73
Method 5	1.83	1.81
Method 6	2.10	2.11

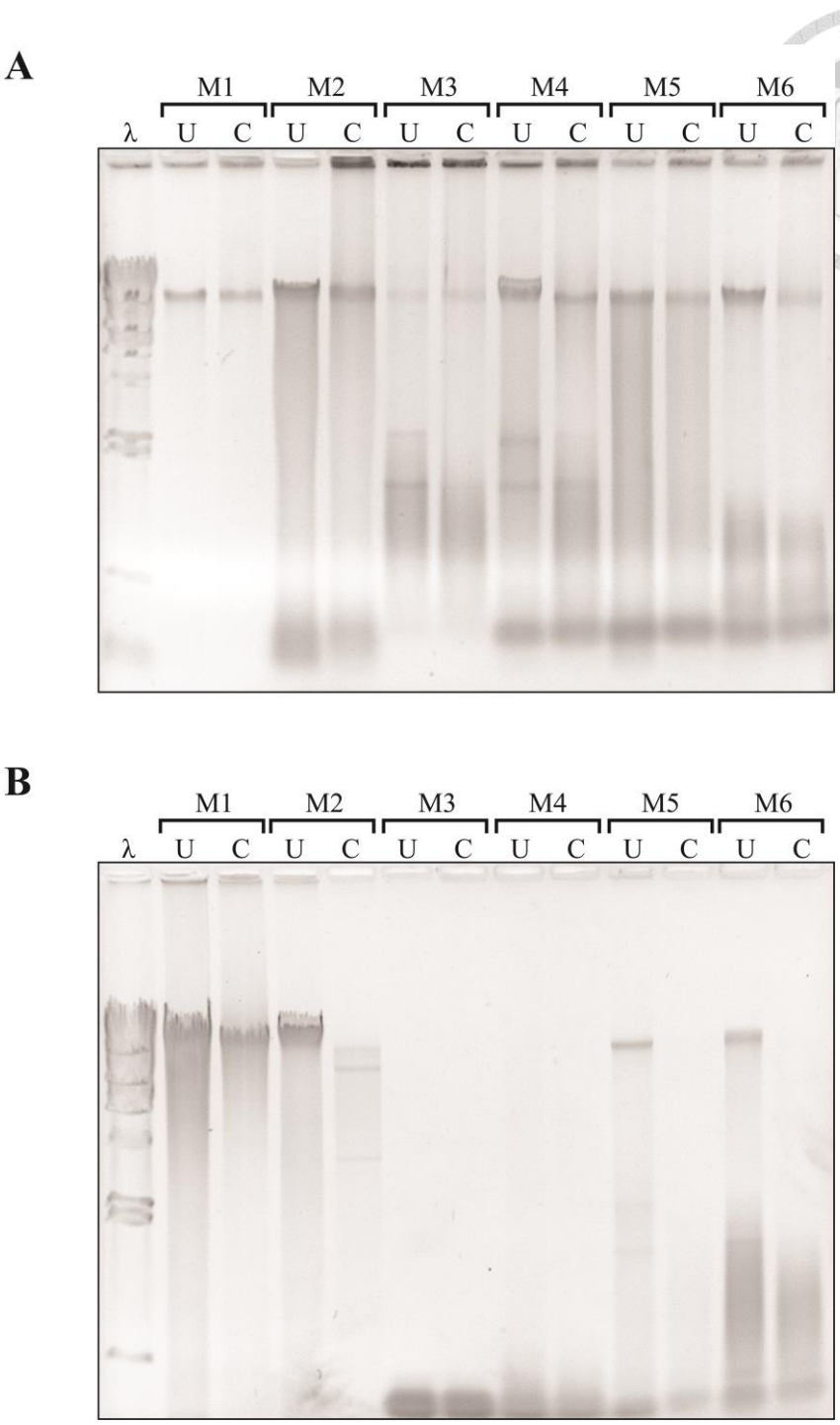


Figure 2-1. The results of restriction enzyme digestion on the isolated DNAs. A: *Balanophora japonica*. B: *Mitrastemon kanehirai*. 4 μg DNA was digested with 4 U *Eco*RI/μg DNA for 1 h at 37°C, and then were separated on a 0.8% TAE agarose gel. U, uncut DNA; C, cut DNA; λ, lambda DNA/*Hind*III marker; M1, QIAGEN DNeasy Plant Mini Kit; M2, Geneaid Tri-Plant Genomic DNA Reagent Kit; M3, Molecular Research Center TRI Reagent; M4, the Doyle and Doyle protocol; M5, the Nickrent protocol; M6, the Barnwell protocol.

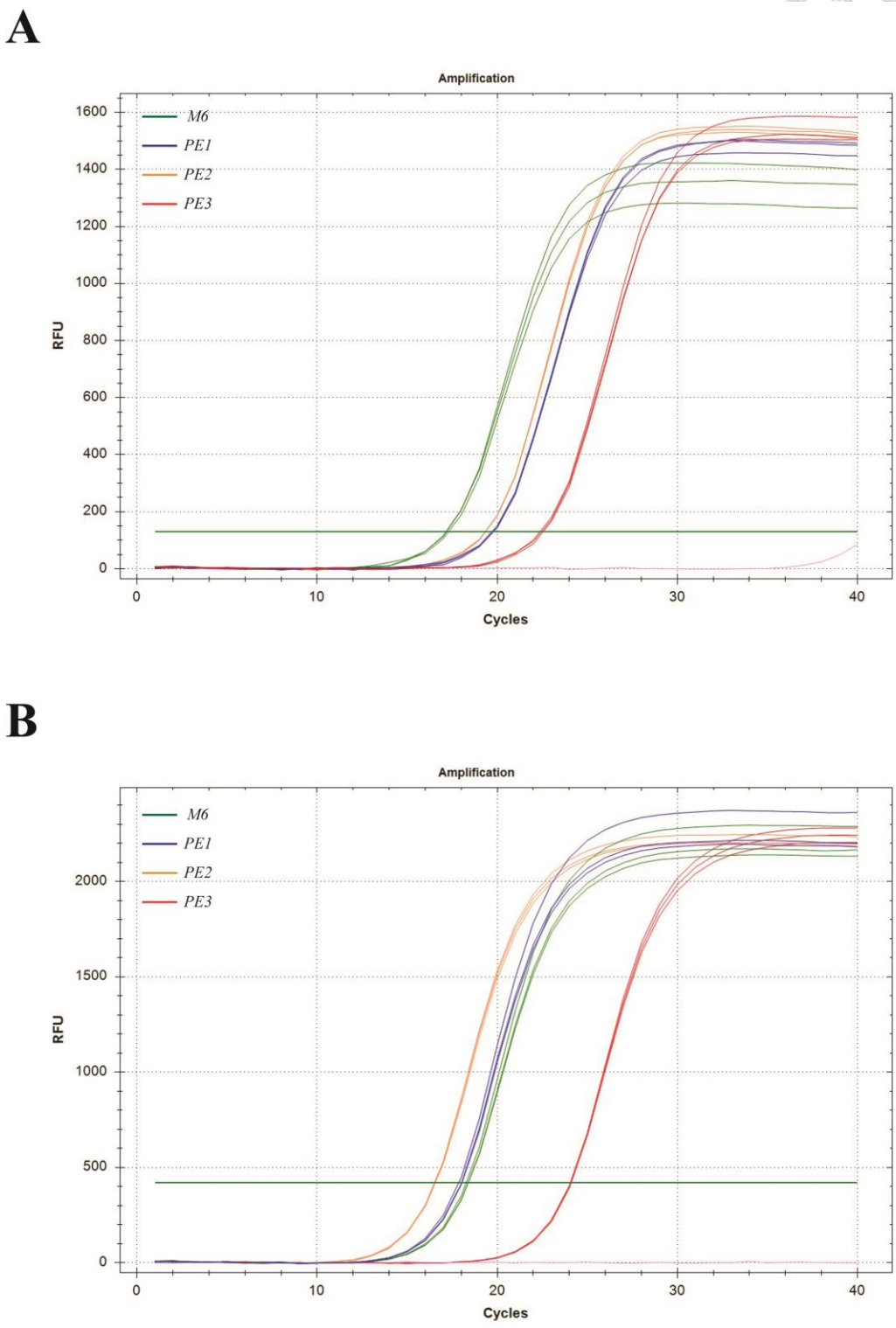


Figure 2-2. Real-time PCR amplification plot of *Mitrastemon kanehirai* DNA. A: nr18S rDNA. B: pt16S rDNA. M6, the Barnwell protocol; PE1, the Milligan protocol; PE2, the Milligan protocol combined with Geneaid Tri-Plant Genomic DNA Reagent Kit; PE3, the Triboush method combined with Milligan protocol.

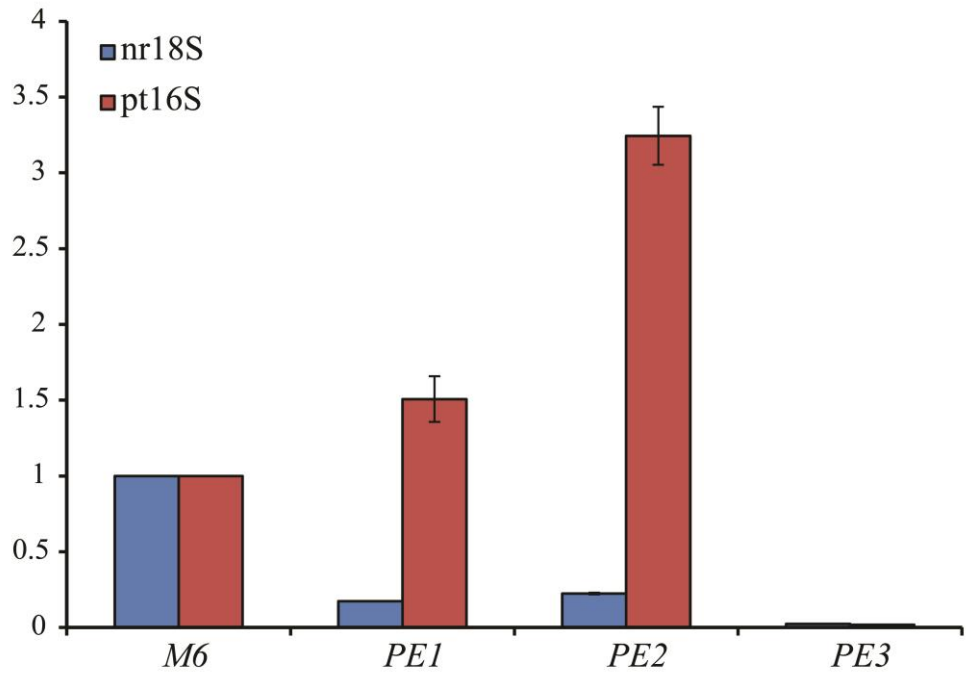


Figure 2-3. Relative quantification of nr18S and pt16S rDNA. *Mitrastemon kanehirai* DNAs extracted by three methods were compared with Barnwell protocol as the reference. M6, the Barnwell protocol; PE1, the Milligan protocol; PE2, the Milligan protocol combined with Geneaid Tri-Plant Genomic DNA Reagent Kit; PE3, the Triboush method combined with Milligan protocol.

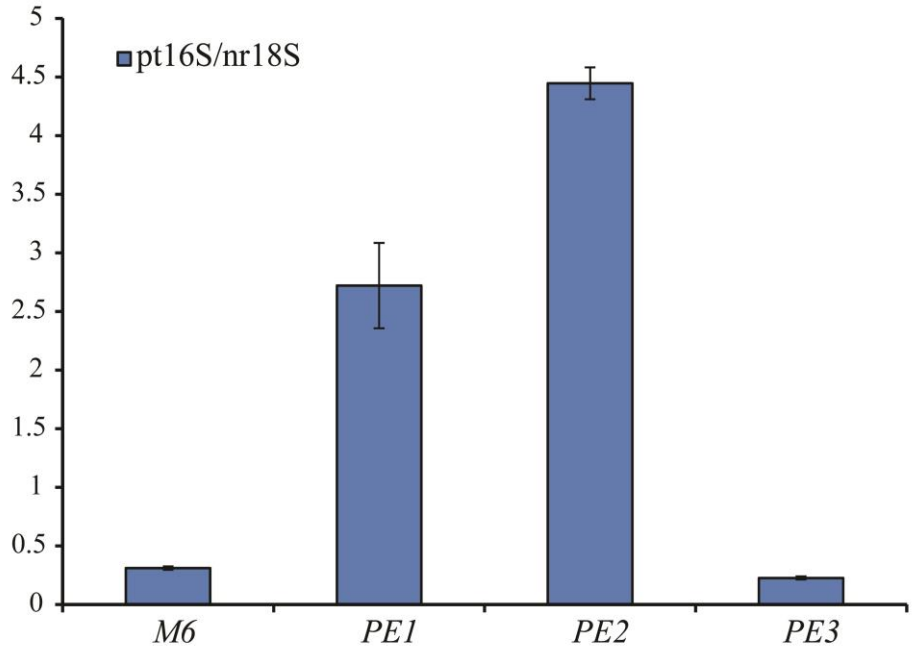
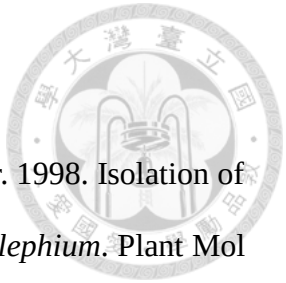


Figure 2-4. The ratio of pt16S/nr18S rDNA in extracted *Mitrastemon kanehirai* DNAs. The pt16S rDNA content was compared with nr18S rDNA in the same DNA samples extracted by four different procedures. M6, the Barnwell protocol; PE1, the Milligan protocol; PE2, the Milligan protocol combined with Geneaid Tri-Plant Genomic DNA Reagent Kit; PE3, the Triboush method combined with Milligan protocol.



References

Barnwell, P., A.N. Blanchard, J.A. Bryant, N. Smirnoff and A.F. Weir. 1998. Isolation of DNA from the highly mucilagenous succulent plant *Sedum telephium*. Plant Mol Biol Rep. 16: 133-138.

Bookjans, G., B.M. Stummann and K.W. Henningsen. 1984. Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. Anal Biochem. 141: 244-247.

Croy, E.J., T. Ikemura, A. Shirsat and R.R.D. Croy. 1993. Plant nucleic acids. *In* R.R.D. Croy, (ed.). Plant molecular biology labfax. Oxford, pp. 21-48.

dePamphilis, C.W. and J.D. Palmer. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature. 348: 337-339.

Do, N. and R.P. Adams. 1991. A simple technique for removing plant polysaccharide contaminants from DNA. Biotechniques. 10: 162, 164, 166.

Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19: 11-15.

Fleischmann, A. and G. Heubl. 2009. Overcoming DNA extraction problems from carnivorous plants. Anal Jardin Bot Mad. 66: 209-215.

Hayata, B. 1913. On the systematic position of *Mitrastemon* as a genus representing a special tribus of the Rafflesiaceae. Icones Plantarum Formosanarum. pp. 199-213.

Herrmann, R.G. 1982. The preparation of circular DNA from plastids. *In* M. Edelman, R.B. Hallick, and N.H. Chua, (eds.). Methods in chloroplast molecular biology. Elsevier Biomedical Press, Amsterdam, pp. 259-280.

Kausch, A.P., T.P. Owen, Jr., S. Narayanswami and B.D. Bruce. 1999. Organelle isolation by magnetic immunoabsorption. Biotechniques. 26: 336-343.

Kolodner, R. and K.K. Tewari. 1975. The molecular size and conformation of the chloroplast DNA from higher plants. BBA-Nucleic Acids and Protein Synthesis. 402: 372-390.

Milligan, B. 1989. Purification of chloroplast DNA using hexadecyltrimethylammonium bromide. Plant Mol Biol Rep. 7: 144-149.

Nickrent, D., R.J. Duff and D.A.M. Konings. 1997a. Structural analyses of plastid-derived 16S rRNAs in holoparasitic angiosperms. Plant Mol Biol. 34: 731-743.

Nickrent, D.L. 1994. From field to film - rapid sequencing methods for field-collected plant species. Biotechniques. 16: 470-475.

Nickrent, D.L., R.J. Duff, A.E. Colwell, A.D. Wolfe, N.D. Young, K.E. Steiner and C.W. dePamphilis. 2000. Molecular phylogenetic and evolutionary studies of parasitic plants. *In* D.E. Soltis, P.S. Soltis, and J.J. Doyle, (eds.). Molecular Systematics of Plants II: DNA sequencing. Kluwer Academic Publishers, Norwell, Massachusetts, USA, pp. 211-241.

Nickrent, D.L., O.Y. Yan, R.J. Duff and C.W. dePamphilis. 1997b. Do nonasterid holoparasitic flowering plants have plastid genomes? Plant Mol Biol. 34: 717-729.

Palmer, J.D. 1982. Physical and gene mapping of chloroplast DNA from *Atriplex triangularis* and *Cucumis sativa*. Nucleic Acids Res. 10: 1593-1605.

Palmer, J.D. 1986. Isolation and structural analysis of chloroplast DNA. *In* H.W. Arthur Weissbach, (ed.). Methods in Enzymology. Academic Press, pp. 167-186.

Scott, K.D. and J. Playford. 1996. DNA extraction technique for PCR in rain forest plant species. Biotechniques. 20: 974, 977, 979.

Su, H.J. and J.M. Hu. 2012. Rate heterogeneity in six protein-coding genes from the holoparasite *Balanophora* (Balanophoraceae) and other taxa of Santalales. Ann Bot. 110: 1137-1147.

Su, H.J., J. Murata and J.M. Hu. 2012. Morphology and phylogenetics of two

holoparasitic plants, *Balanophora japonica* and *Balanophora yakushimensis* (Balanophoraceae), and their hosts in Taiwan and Japan. J Plant Res. 125: 317-326.

Sytsma, K.J. 1994. DNA extraction from recalcitrant plants: long, pure, and simple? *In* R.P. Adams, J.S. Miller, E.M. Golenberg, and J.E. Adams, (eds.). Conservation of plant genes II. Missouri Botanical Garden, Missouri, pp. 69-81.

Triboush, S.O., N.G. Danilenko and O.G. Davydenko. 1998. A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. Plant Mol Biol Rep. 16: 183-189.

Tsai, T.H., G.J. Wang and L.C. Lin. 2008. Vasorelaxing alkaloids and flavonoids from *Cassytha filiformis*. J Nat Prod. 71: 289-291.

Wang, X., Z. Liu, W. Qiao, R. Cheng, B. Liu and G. She. 2012. Phytochemicals and biological studies of plants from the genus *Balanophora*. Chem Cent J. 6: 79.

Yu, F.R., Y. Liu, Y.Z. Cui, E.Q. Chan, M.R. Xie, P.P. McGuire and F.H. Yu. 2010. Effects of a flavonoid extract from *Cynomorium songaricum* on the swimming endurance of rats. Am J Chin Med. 38: 65-73.

Supplementary data

Table S2-1. Other methods have been tested in this study.

Method	Reference
DNA isolation	
Column-based commercial kit	Fast ID Genomic DNA Extraction Kit, Genetic ID, Fairfield, IA, USA[a]
Column-based commercial kit	Plant Genomic DNA Mini Kit, BIOMAN, New Taipei City, Taiwan[a]
Column-based commercial kit	Plant Genomic DNA Purification Kit, GeneMark, Taichung City, Taiwan[a]
Particular reagent- based commercial kit	TRIzol® Reagent, Invitrogen, Carlsbad, CA, USA[b]
The rainforest method	Scott and Playford, 1996[c]
Modified CTAB methods	Doyle and Doyle, 1987; Croy et al., 1993; Sytsma, 1994[c]
Plastid DNA enrichment	
The sunflower method	Triboush et al., 1998[b]

[a] The yields of DNA from these methods were very low, and $A_{260/280}$ ratios of DNAs were below 1.2.

[b] The methods failed to extract DNA from *B. japonica* and *M. kanehirai*.

[c] The methods failed to improve DNA quality compared with *Method 4*.

Chapter 3.

Formatted for submission to *Botanical Studies* (Accepted June 05, 2013)

Heterogeneous evolutionary rates of nuclear and organelle small-subunit rDNAs in heterotrophic plants

Shin-Yi Shyu[1] and Jer-Ming Hu[1]

[1]Institute of Ecology and Evolutionary Biology, National Taiwan University, Taiwan

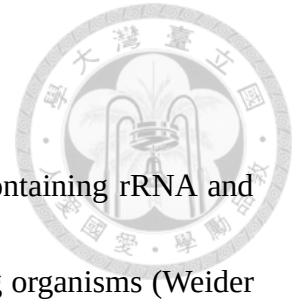
Authors' contributions:

Shin-Yi Shyu designed and performed the experiments, analyzed data and drafted the manuscript. Jer-Ming Hu administered the experiments and edited the manuscript.

Abstract

High rates of nucleotide substitution within small-subunit rRNA genes from all three genomes have been reported in various angiosperms, particularly for the lineages with a heterotrophic mode of life. However, substitution rates of rDNA sequences vary among subcellular genomes and among taxa. In order to elucidate patterns of evolutionary rates among the three subcellular SSU rDNAs in heterotrophic plants, we examined mitochondrial 19S, plastid 16S and nuclear 18S rDNA sequences from one hemiparasitic, five holoparasitic and three mycoheterotrophic plants. Among these nine heterotrophic plants, six of them are non-photosynthetic, while others retain partial photosynthetic ability. Rate heterogeneity was estimated with relative rate tests and phylogenetic analyses. Our results show that both pt16S and nr18S rDNA sequences of non-photosynthetic species have significantly increased substitution rates in comparison with their autotrophic relatives. However, this phenomenon was not found in mt19S rDNA. The pronounced divergent pt16S and nr18S rDNA sequences were only found in *Balanophora japonica* and *Mitrastemon kanehirai*, and accompanied by a decrease in GC contents of the rDNA sequences. In contrast, *Aeginetia indica*, *Cassytha filiformis*, *Cheilotheca humilis*, *Cheilotheca macrocarpa*, *Cuscuta australis*, *Galeola lindleyana* and *Orobanche coerulescens*, do not exhibit consistent patterns between pt16S and nr18S rDNA substitution rates, indicating that the accelerated evolutionary rates are not synchronized among the three subcellular SSU.

Keywords: heterotrophic plant, non-photosynthetic plant, small-subunit rDNA, substitution rate.



Introduction

Ribosomes, each is composed of a large and small subunit containing rRNA and proteins, are the sites of protein synthesis and essential for all living organisms (Weider et al., 2005). The structural conservation and functional constrains of ribosome also provide the common uses of the sequences of rDNA as a molecular marker for phylogenetic reconstruction among organisms (Hillis and Dixon, 1991). The nuclear 18S (nr18S) rDNA and the lesser used plastid 16S (pt16S) mitochondrial 19S (mt19S) sequences have been widely used to examine higher-level phylogenetic relationships of the land plants (Duff and Nickrent, 1999; Soltis et al., 1999; Soltis and Soltis, 2000a).

In eukaryotes, nuclear genes encoding 18S, 5.8S and 28S rRNA are generally in tandem arrays of repeat units (rDNA) separated by intergenic spacers (IGS) (Kupriyanova, 2000). The sequences among the repeats usually show high degree of homogeneity through a process of concerted evolution (Richard et al., 2008), although polymorphism within individuals does occur occasionally, especially in the spacer regions or non-functional rDNA copies (Bailey et al., 2003). Nevertheless, the rDNA sequences are still the most conserved DNA segment in all living organisms and widely used for higher level phylogenetic reconstruction (Soltis et al., 1999; Soltis & Soltis, 2000a, 2000b; Weisburg et al., 1991).

Although rDNA sequences generally show quite low variations, several studies

documented the accelerated evolutionary rates of nr18S in lineages of heterotrophic plants, including parasitic and mycoheterotrophic plants (Nickrent and Starr, 1994; Nickrent and Duff, 1996; Lemaire et al., 2010; Su and Hu, 2012). Similarly, substitution rate acceleration is found in some heterotrophic plants for pt16S genes (Nickrent et al., 1997a). However, these increases of substitution rates are not corresponding to any particular nutrient-uptake mode because not all achlorophyllus plants display such increased substitution rates (Lemaire et al., 2010; Young & dePamphilis, 2005). The holoparasitic plants *Cynomorium coccineum* (Cynomoriaceae) and *Orobanche fasciculata* (Orobanchaceae), for examples, do not show significantly accelerated substitution in nr18S rDNA sequence (dePamphilis et al., 1997; Lemaire et al., 2010).

Most of the sequence variation in those with accelerated evolutionary rates are not in the functionally or structurally important regions of rDNA (Lemaire et al., 2010), suggesting that the elevated substitution rates in rDNA might reflect an overall increases of mutations in the genome. Several hypotheses have been proposed to explain the rate acceleration in certain plants, like the presence of defective DNA repair efficiency, shorter generation time, higher speciation rates and smaller effective population size (Nickrent and Starr, 1994; Nickrent et al., 1998; Lemaire et al., 2010). Such scenarios would predict similar rate acceleration patterns in the same genome, and also for the genes with similar selection pressures under the same functional constraints. The nuclear and organelle

ribosomal genes thus have similar substitution rate patterns under this speculation. It has been shown that in nonasterid holoparasites the rate acceleration are coincident among mt19S, pt16S and nr18S rDNA sequences (Duff and Nickrent, 1997). However, such pattern does not hold in our preliminary survey for the holoparasitic plant *Balanophora* in that it has a relatively slow substitution rate in mt19S, but extremely high for pt16S and nr18S rDNA sequences. It suggested a more comprehensive survey is needed to evaluate the patterns of substitution rates in heterotrophic plants.

In this study we examined whether there is any correlation of evolutionary rate patterns among the three subcellular SSU rDNAs in heterotrophic plants. The mt19S, pt16S and nr18S rDNAs were amplified and evaluated, including one hemiparasite *Cassytha filiformis* (Lauraceae), five holoparasites: *Aeginetia indica*, *Orobanche coerulescens* (both Orobanchaceae), *Balanophora japonica* (Balanophoraceae), *Cuscuta australis* (Convolulaceae, retains partial photosynthetic ability), *Mitrastemon kanehirai* (Mitrastemonaceae) and three mycoheterotrophic plants: *Cheilotheca humilis*, *Cheilotheca macrocarpa* (both Ericaceae-Monotropoideae), *Galeola lindleyana* (Orchidaceae, retains partial photosynthetic ability). Relative rate tests were performed for all three groups of SSU rDNA-sequences among selected heterotrophic taxa and related autotrophic lineages in order to examine their evolutionary rate variation.

Materials and Methods

Plant materials and DNA extraction

The lifestyle and GenBank accession numbers of all selected heterotrophic lineages for the three genes is listed in Table 3-1. Fresh samples of the nine species, including one hemiparasite, five holoparasities and three mycoheterotrophic plants, were collected from Taiwan and stored at -20℃ for DNA extraction and further experiments. Total genomic DNA of *B. japonica* and *M. kanehirai* were isolated using a modified CTAB methods (Barnwell et al., 1998), while the remaining samples were isolated following a standard CTAB method (Doyle and Doyle, 1987). The PCR products for the different SSU rDNA regions were amplified by using different primers listed in Table 3-2. Because some of the PCR produced more than one product, the PCR products with the corrected size were then cloned into the pGEM-T Vector (Promega, Madison, WI, USA). The sequences were determined by an automatic DNA analyzer at Academia Sinica, and both strands of sequences were further examined by Sequencher 4.5 (Gene Code Corp., USA).

Evolutionary rate analyses

For all mt19S, pt16S and nr18S rDNA sequences the closest related taxa to the nine heterotrophic plants in this study were selected and downloaded from NCBI GenBank (accession numbers see Appendix Table A1). Three data matrices were constructed, and the alignments were conducted by ClustalX 1.83 (Thompson et al., 1997) and visually

confirmed using MacClade 4.06 (Maddison and Maddison, 2000). The V1 and V7 regions of mt19S rDNA, corresponded to positions 71-240 and 1214-1447 on the *Phaeoceros* structure model (Duff and Nickrent, 1999), were excluded since they are too variable in the sequence and in length. Other regions that vary in length among the selected land plants did not interfere with the alignment, and therefore were included in the analysis. Evolutionary divergence, including nucleotide composition and genetic distance, among the rDNA sequences was determined using MEGA version 5 (Tamura et al., 2011). The nucleotide substitution rates of three rDNAs were estimated simply by the number of substitutions per site compared with *Glycine max* sequences in order to compare with previous studies. Relative rate tests were conducted using RRTree under Phylemon 2.0 (Sánchez et al., 2011) with Kimura 2-parameter model (Kimura, 1980), to compare the substitution rates between the heterotrophic species and the corresponding autotrophic relatives.

Patterns of nucleotide substitution in SSU rDNAs

Conserved and variable nucleotides in the SSU rDNA datasets were identified using the CHART option of MacClade 4.06. The maximum parsimony trees were constructed based on three rDNA datasets with *Amborella* as outgroup. The topology of the trees was adjusted according to APG III system (Chase and Reveal, 2009), and then the trees were used as the backbone for evaluation. The patterns were examined within heterotrophic

(Table 3-1) and autotrophic species (Table 3-4) separately, and both groups included six taxa. The heterotrophic species that retain partial photosynthetic ability were eliminated from the analyses. The difference between two groups was calculated and values of steps per nucleotide site were depicted on a histograms. Positive and negative values represent character substitutions contributed by heterotrophic and autotrophic species, respectively.

Results

SSU rDNAs in the selected heterotrophic plants

All nuclear and organelle rDNA sequences were successfully amplified and identified for the nine selected heterotrophic plants (Table 3-1). The SSU rDNA sequences all show conserved and variable regions roughly corresponding to the ones described by previous studies (Duff and Nickrent, 1997; Nickrent et al., 1997a; Soltis and Soltis, 2000b). For example, the mt19S rDNA of heterotrophic species have a conserved 'core' region and the V1-V7 variable regions similar to their autotrophic relatives (Fig. 3-1). In contrast, the nr18S and pt16S rDNAs in heterotrophic plants are more variable than their autotrophic counterparts (Fig. 3-1). All nine heterotrophic species are prone to small insertion/deletion (indels) in their pt16S rDNA sequences, and large indels are only found in *A. indica* (14 bp insert), *B. japonica* (17 bp deletion) and *M. kanehirai* (20 bp deletion) (Supplementary data Fig. S3-2).

Sequences composition bias and substitution

The GC contents of mt19S, pt16S and nr18S rDNA sequences in the heterotrophic plants are 51-56%, 24-57% and 42-50%, respectively. The lowest GC% of mt19S rDNA we sampled is in *Cu. australis* (50.5%). In comparison, *B. japonica* has the lowest pt16S (GC=24.1%) and nr18S (GC=41.5%) in plastid and nuclear rDNA, respectively (Table 3-3). Overall the GC% of mt19S rDNA in heterotrophic plants is close to the GC content of mt19S sequences in angiosperms (53.7±1%) (Duff and Nickrent, 1997) and the GC contents of pt16S and nr18S rDNAs are both lower than the averages of green plants (Table 3-3).

In order to examine the rate variation patterns of the SSU rDNA sequences in the selected heterotrophic plants, we estimated the nucleotide substitutions of the three rDNAs from these plants by comparing with mitochondrial, plastid and nuclear rDNA sequences of *Glycine max* (Table 3-3). Four of the nine identified pt16S rDNAs in heterotrophic plants show much higher substitution rate than those in green plants, i.e. *B. japonica* (33.73%), *Ch. humilis* (6.56%), *Ch. macrocarpa* (6.43%) and *M. kanehirai* (14.49%). In comparison, a different set of the taxa show the elevated rate in nr18S rDNAs, i.e. *B. japonica* (12.02%), *Cu. australis* (6.17%), *G. lindleyana* (5.57%) and *M. kanehirai* (7.26%). Nevertheless, the divergences of pt16S and nr18S sequences of these taxa are both higher than the average of sequence variation in the green plants (2-3% and

3.6%, respectively) (Nickrent and Starr, 1994; Nickrent et al., 1997a). Moreover, only *G. lindleyana* shows distinct variation among the mt19S rDNA sequences we sampled (Table 3-3). Based on the comparison of the ratios of substitution rates among three rDNAs, it shows that the rate substitution patterns varied among species, i.e. they are not synchronized within species. The divergence could be higher in plastid rDNAs in some species, e.g. *Ch. humilis* and *Ch. macrocarpa*, but higher in nuclear rDNAs in the other heterotrophic plants like *Ca. filiformis* or *A. indica* (Table 3-3).

The transition/transversion (T_S/T_V) ratios of mt19S in the heterotrophic plants are within the range of photosynthetic plants, while several species have much higher T_S/T_V ratio in their nr18S or pt16S. The highest T_S/T_V ratios of pt16S and nr18S are found in *Ca. filiformis* (4.571) and *Ch. macrocarpa* (3.923) respectively. Transition biases present in plant nuclear and chloroplast genomes have been described, and transition bias in rDNA stem regions may help maintain secondary structure (Soltis and Soltis, 2000b). However, the T_S/T_V ratio does not show any correlation among the three SSU rDNAs and the corresponding trophic modes. This result shows that the pt16S rDNA of *B. japonica* has the most extreme divergence among all the identified sequences (Table 3-3 and Supplementary data Fig. S3-2). In addition, there is an inverse correlation between GC content and substitution in pt16S rDNAs and is not shown in the other two genes (mt19S and nr18S).

Relative rate tests

The results of relative rate tests for the pt16S and nr18S rDNA datasets show similar pattern indicating significant higher rates for all heterotrophic plants, except for *Ca. filiformis* (Table 3-4). In mt19S rDNAs, however, eight out of the twelve heterotrophic taxa show rates significantly diverged from their autotrophic counterparts, but the substitution rates of heterotrophic taxa are not always higher than autotrophic taxa. For examples, *B. japonica*, *M. kanehirai*, *O. coerulescens*, *Ch. humilis* and *Ch. macrocarpa* all show significant lower substitution rates than their autotrophic relatives (Table 3-4).

Discussion

The identified SSU rDNAs from heterotrophic plants show various degrees of sequence divergence among the nuclear and organelle genomes. However, there is no clear pattern of the rate heterogeneity associated with the trophic form, i.e. how much photosynthetic ability the heterotrophic plants retained. In general, mitochondrial 19S sequences are most conserved, with substitution rate only up to 3% (*G. lindleyana*), compared with nuclear and plastid SSU rDNAs. In comparison, the non-photosynthetic parasite *B. japonica* shows extreme substitutions in both of the nr18S and pt16S rDNAs among the studied plants.

Although nearly all the heterotrophic plants show significantly elevated substitution

rate in their pt16S and nr18S rDNAs, for most of them the rates do not increase dramatically like previous studies described for non-photosynthetic plants like *Corynaea*, *Cytinus*, or *Hydnora* (Nickrent and Starr, 1994; Nickrent et al., 1997a). The magnitudes of pt16S rDNA rate variations in *A. indica* and *O. coerulescens*, that both are achlorophyllous, fall into the range of most angiosperms, whereas *B. japonica* nr18S rDNA is the only one that evolved more than three times faster compared to the nonparasitic plants (Nickrent and Starr, 1994; Nickrent et al., 2000). The result is congruent with previous studies using another *Balanophora* species, *B. fungosa* for the analysis (Su and Hu, 2012). Although previous study indicated that nonasterid holoparasites show significantly increased substitution rates in their core mt19S rDNA sequences (2.3~7.6%) (Duff and Nickrent, 1997), in our study mt19S rDNA exhibits little divergence among the nine heterotrophic plants.

Several holoparasites have been reported that all three SSU rDNAs show increased substitution rates (Nickrent and Starr, 1994; Duff and Nickrent, 1997; Nickrent et al., 1997a; Nickrent et al., 2000), and an acceleration of plastid genomes parallel to high mitochondrial divergence is also described in some autotrophic plants (Soria-Hernanz et al., 2008; Sloan et al., 2012). Since our results are somewhat different from those studies, we re-evaluated the substitution and relative rates of three holoparasites (*Corynaea, Cytinus, Hydnora*) by using the same methods in this study (Tables 3-3, -4). Our results

show the three holoparasites indeed revealed much higher substitution in their three SSU rDNAs; however, the rate increases of *Cytinus* mt19S rDNA is not significant compared with their autotrophic counterparts in the RRTree test. The mt19S rDNA sequences of all selected taxa (excepted *G. lindleyana*) are not prone to high-elevated substitution rates, even in *B. japonica* where pt16S and nr18S rDNAs are more divergent than *Corynaea*. Therefore, our observations suggest that rate heterogeneity among genomes is distinct for selected taxa. Similar unequal rate accelerations between mitochondrial and plastid sequences have been reported in *Silene vulgaris*, but for the protein coding genes, e.g. comparatively high variation in mitochondrial *atp1* and *atp9* than others mt genes (Houliston and Olson, 2006).

Previous studies have found that the relative rate of synonymous substitutions of mitochondrial, plastid and nuclear genes of angiosperms is 1:3:16 and that the ratio can go up to 1:16:75 in *Arabidopsis* for protein-coding genes (Drouin et al., 2008; Huang et al., 2012). If the rate acceleration prevails across the three genomes, the ratio in heterotrophic plants should be similar to the value of angiosperms. However, the substitution ratios of the three SSU rDNAs in heterotrophic plants are different from the ratio patterns of other angiosperms. One of the reasons that none of the nine heterotrophic plants show such pattern in our study could be due to the differences between substitution models between rDNA and protein-coding genes. The synonymous substitutions per sites

(K_s) were obtained with Li's methods (Li et al., 1985; Li, 1993), whereas the substitution rates of rDNAs were simply compared with the sequence of *Glycine* in our study. Therefore, we re-estimated the ratio of three SSU rDNAs from previous studies (Nickrent and Starr, 1994; Duff and Nickrent, 1997; Nickrent et al., 1997a) and compared them with *Glycine* sequences, and found the substitution ratio is 1:2.6:3.8 in autotrophic plants. This ratio is closer to our results. In addition, from the ratio we found seven out of the twelve heterotrophic plants have higher substitutions in their 16S rDNA than in 18S, and all taxa show their pt16S rDNA evolved relatively faster than nr18S in RRTree test (K1/K2 value in Table 3-4). It indicates that the plastome of heterotrophic plants might evolve faster than other two subcellular genomes. But there is no clear pattern on substitution and relative rates among genomes, and an accelerated rate in one SSU rDNA does not imply increasing rates for rDNAs of the other genomes. These results suggest that the three subcellular genomes are under independent evolutionary trajectories in both autotrophic and heterotrophic plants.

For the nucleotide composition, significant decreases in GC contents of pt16S rDNA are found in *B. japonica*, *M. kanehirai*, *Ch. Humilis* and *Ch. macrocarpa*. In plastid genome, rRNA genes have the highest GC content of any coding regions; however, GC content of *B. japonica* pt16S rDNA (24.1%) is even lower than whole chloroplast genome in green plants (35.4-39.6%) (Jansen and Ruhlman, 2012). The decreased of GC content

of SSU rDNAs is accompanied by high substitution rates. The inverse correlation between GC content and substitution rate of pt16S rDNA in these four species is congruent with the pattern in other holoparasitic angiosperms (Nickrent et al., 1997a). However, in the other rDNAs, this phenomenon is only found in *B. japonica* nr18S rDNA.

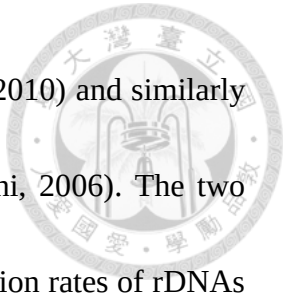
We also found that there are frequent indels in pt16S rDNA sequences of selected heterotrophic plants (Supplementary data Fig. S3-2), which have been found in other holoparasitic plants (Nickrent et al., 1997a) and even more pronounced in non-photosynthetic green algae (Nedelcu, 2001). In comparison, angiosperm nr18S rDNA has fewer indels and most indels are only one or two nucleotides in length (Supplementary data Fig. S3-3) (Soltis and Soltis, 2000b). Whether or not high substitution rate and the larger indels of the sequences affect the structure and function of pt16S rDNA, it requires further structure analysis of sequences and RNA expression experiment.

Transition biases have been observed in pt16S and nr18S rDNA sequences for both autotrophic and heterotrophic plants, and vary in magnitude (Nickrent et al., 1997a; Soltis and Soltis, 2000b). The nr18S and pt16S rDNAs of all species examined reveal a strong bias towards transitional mutations. On the contrary, mt19S rDNA does not show the same base composition bias.

Relaxation of selection pressure as a consequence of loss of photosynthetic ability might be the main factor for increased substitution rates found in plastid, but the mutations

in mt19S, nr18S rDNAs of heterotrophic plants is not clear at present. It is generally assumed that rate increases in nr18S rDNA sequences are due to an overall elevated mutation rate in the genome, by mechanisms like defective DNA repair efficiency or rapid generation time (Nickrent and Starr, 1994; Lemaire et al., 2010). It should be noted that such mechanisms that previous speculated for accelerated substitution rates in a particular organism mostly have the effects applied to whole genome. Changes in the substitution rate pattern among the genomes (mt, pt, and nr) of the same species would violate these hypotheses. In addition, the organelle DNA replications are generally assumed to be controlled by nucleus since almost all genes for proteins involved in the replication machinery are located in the nucleus (Heinhorst and Cannon, 1993; Nielsen et al., 2010), and so may applied to DNA repairing in organelles (Kimura and Sakaguchi, 2006). Therefore, it seems that the substitution rates of rDNAs should be in congruence among nuclear and organelle genomes. Our recent study also shows that substitution rate heterogeneity within the nuclear genome (i.e. 18S rDNA) of *Balanophora*, is not associated with rate increases in other nuclear protein coding genes (Su and Hu, 2012). Therefore, the rate heterogeneity among the three genomes for heterotrophic plants could involve different mechanisms working on the three rDNA genes.

The organelle DNA replications are generally assumed to be controlled by nuclear genome since almost all genes for proteins involved in the replication machinery are



located in the nucleus (Heinhorst and Cannon, 1993; Nielsen et al., 2010) and similarly for proteins of DNA repairing in organelles (Kimura and Sakaguchi, 2006). The two organelles are prokaryote origin; therefore, it seems that the substitution rates of rDNAs should be in congruence between organelle genomes. However, the actual controls for organelle replication initiation, replication and copy number are still not well understood (Nielsen et al., 2010). Interestingly, there are indeed specialized organelle nuclei found in plants, hinted complex dynamics of interactions between the three genomes (nr, pt and mt) (Sakai et al., 2004). Our results clearly show rate heterogeneity in the substitution rate pattern among the nuclear, plastid and mitochondrial genomes of selected heterotrophic plants. Much is still unknown for the function and biogenesis of the plastids in these heterotrophic plants, thus quite difficult to speculate the driving forces on the rate heterogeneity. With more understanding of the variations on the DNA replication/repair machineries among genome compartments and among diverse species could thus elucidate the discrepancy among rate acceleration of three subcellular genomes

Table 3-1. The information for source materials of SSU rDNA sequences.

Species	Family	Trophic mode[a]	Accession no.		
			nr18S	pt16S	mt19S
Cassytha filiformis[b]	Lauraceae	HE	KC588400	KC588391	KC588409
Balanophora japonica	Balanophoraceae	HO	KC588399	KC588390	KC588408
Cuscuta australis[b]	Convolvulaceae	HO	KC588403	KC588394	KC588412
Mitrastemon kanehirai	Mitrastemonaceae	HO	KC588405	KC588396	KC588414
Aeginetia indica	Orobanchaceae	HO	KC588398	KC588389	KC588407
Orobanche coerulescens	Orobanchaceae	HO	KC588406	KC588397	KC588415
Cheilotheca humilis	Ericaceae	MY	KC588401	KC588392	KC588410
Cheilotheca macrocarpa	Ericaceae	MY	KC588402	KC588393	KC588411
Galeola lindleyana[b]	Orchidaceae	MY	KC588404	KC588395	KC588413

[a] Trophic mode: HE, hemiparasitic; HO, holoparasitic; MY, mycoheterotrophic

[b] Species retain some photosynthetic ability, and was excluded from the analyses of nucleotide pattern.

Table 3-2. Primers used in this study.

Position	Primer name	Sequence from 5' to 3'	Reference
Plastid	16S F[ade]	AACAAGGAAGCTATAAGTAATGCAA	
	16S 8F[bc]	GGAGAGTTCGATCCTGGCTCAG	Nickrent et al., 1997b
	16S 734F[c]	TGGGATTAGAGACCCCAGTA	
	16S 878R[c]	GCCCCCGYCAATTCCT	
	16S 1461R[c]	GGTATTCTAGCCACACTTTCCAG	
	16S 1508R[a]	ACCAAAATACCCAACAAGCA	
	16S 1508R1[e]	CCCAAAAAACCCAACAAGCA	
	16S R[d]	ACATGGGGACGTAAAACAGG	
	23S 459R[b]	CTT TCC CTC ACG GTA	
Mitochondrion	m19S-9F[f]	GAGTTTGATCCTGGCTCAGA	Duff and Nickrent, 1997, 1999
	m19S-434F[c]	GCCGCTTGTAAAGCTC	Duff and Nickrent, 1997, 1999
	m19S-950R[c]	AAGGTTTTGCGCGTTGTATC	
	m19S-1949R[f]	GCCACAGGTTCCCCTACGGCT	Duff and Nickrent, 1997, 1999
Nucleus	SSU4F[abcd]	TTGGTTGATCCTGCCAGTAG	
	SSU12F[e]	TCCTGCCAGTASTCATATGC	Malécot and Nickrent, 2008
	SSU1769R[f]	CACCTACGGAAACCTTGTT	Nickrent and Starr, 1994

[a] Primers used in *A. indica*, *Ch. humilis*, *Cu. australis*, *G. lindleyana* and *O. coerulescens*.

[b] Primers used in *M. kanehirai*.

[c] Primers used in *B. japonica*.

[d] Primers used in *Ca. filiformis*.

[e] Primers used in *Ch. macrocarpa*.

[f] Primers used in all species.

Table 3-3. Features of mt19S, pt16S and nr18S rDNAs from heterotrophic plants.

Taxa	GC%			Substitution rate (%)[a]				T_S/T_V[a]		
	19S	16S	18S	19S[b]	16S	18S	Ratio[c]	19S[b]	16S	18S
Green plants	53.7±1[e]	55.6[e]	49.4±2[e]	0.6~1.4[e]	2~3[f]	3.6[g]	1:3:16[h]	0.10~1.14[e]	2[e]	2[e]
Ca. filiformis	52.6	56.1	50.3	1.43	2.44	4.42	1:1.7:3.1	0.374	4.672	1.748
A. indica	52.4	54.6	48.3	1.37	2.81	4.81	1:2.1:3.5	1.343	2.039	2.612
B. japonica	53.6	24.1	41.5	0.85	33.73	12.02	1:39.9:14.2	0.860	1.704	2.952
Cu. australis	50.5	56.6	47.2	1.17	2.25	6.17	1:1.9:5.3	1.257	1.795	3.349
M. kanehirai	54.3	44.6	48.6	1.37	14.49	7.26	1:10.6:5.3	0.753	2.861	3.088
O. coerulescens	53.6	56.6	49.5	1.24	1.81	3.06	1:1.5:2.5	0.904	1.923	1.988
Ch. humilis	54.3	52.7	48.3	1.37	6.56	3.88	1:4.8:2.8	0.753	2.511	3.277
Ch. macrocarpa	54.4	52.3	48.2	1.43	6.43	3.50	1:4.5:2.4	0.694	3.124	4.041
G. lindleyana	56.2	55.7	49.6	2.73	3.12	5.57	1:1.1:2.0	0.398	3.249	1.603
Corynaea[d]	53.3	26.2	46.4	3.39	31.41	6.92	1:9.3:2.0	0.973	2.328	2.178
Cytinus[d]	54.7	49.6	47.2	2.49	7.88	6.32	1:3.2:2.5	0.566	2.290	3.272
Hydnora[d]	54.5	42.3	47.1	3.25	19.36	7.03	1:6.0:2.2	0.422	3.147	2.002

[a] Substitution rates and T_S/T_V of SSU rDNAs were compared with *Glycine max*.

[b] Calculation of mitochondrial 19S rDNA sequences excluded the V1 and V7 regions.

[c] Ratios of substitution rate between mt19S and pt16S, nr18S rDNAs.

[d] Sequences of SSU rDNAs were obtained from GenBank.

[e] Duff and Nickrent, 1997

[f] Nickrent et al., 1997a

[g] Nickrent and Starr, 1994

[h] Drouin et al., 2008

Table 3-4. Results of relative rate tests for comparing SSU rDNA substitution rates between heterotrophic lineages (Lineage 1) and their autotrophic relatives (Lineage 2).

Lineage 1	Lineage 2	K1-K2[a]	K1/K2[a]	*P* value[b]
mt19S				
Ca. filiformis	*Cinnamomum, Laurus*	0.006	1.34	6.4×10^{-2}
A. indica	*Digitalis*[c], *Veronica*[c]	-0.005	0.82	1.4×10^{-1}
B. japonica	*Lepidoceras*[c], *Vitis*	-0.005	0.74	$\mathbf{1.0 \times 10^{-2}}$
Cu. australis	*Nicotiana*	-0.001	0.97	7.6×10^{-1}
M. kanehirai	*Beta*[c], *Silene latifolia*[c], *Silene vulgaris*[c]	-0.014	0.75	$\mathbf{1.7 \times 10^{-2}}$
O. coerulescens	*Digitalis*[c], *Veronica*[c]	-0.008	0.66	$\mathbf{1.8 \times 10^{-3}}$
Ch. humilis	*Beta*[c], *Silene latifolia*[c], *Silene vulgaris*[c]	-0.014	0.76	$\mathbf{2.0 \times 10^{-2}}$
Ch. macrocarpa	*Beta*[c], *Silene latifolia*[c], *Silene vulgaris*[c]	-0.014	0.76	$\mathbf{2.0 \times 10^{-2}}$
G. lindleyana	*Asparagus, Iris*	0.015	1.62	$\mathbf{1.0 \times 10^{-3}}$
Corynaea	*Lepidoceras, Vitis*	0.027	2.46	$\mathbf{5.5 \times 10^{-7}}$
Cytinus	*Brassica, Raphanus*	0.011	0.76	1.0×10^{-1}
Hydnora	*Aristolochia, Dr. winteri, Saururus*	0.019	2.03	$\mathbf{3.4 \times 10^{-5}}$
pt16S				
Ca. filiformis	*Cinnamomum*	0.003	1.25	2.5×10^{-1}
A. indica	*Antirrhinum*[c], *Olea*, *Sesamum*[c]	0.020	3.67	$\mathbf{2.3 \times 10^{-7}}$
B. japonica	*Heisteria*[c], *Ximenia*	0.552	62.18	$\mathbf{1.0 \times 10^{-7}}$
Cu. australis	*Ipomoea*, *Nicotiana*, *Solanum*	0.014	2.77	$\mathbf{1.5 \times 10^{-5}}$
M. kanehirai	*Camellia*[c], *Stewartia*[c], *Symplocos*[c]	0.180	27.79	$\mathbf{1.0 \times 10^{-7}}$
O. coerulescens	*Antirrhinum*[c], *Olea*, *Sesamum*[c]	0.008	2.17	$\mathbf{7.0 \times 10^{-4}}$
Ch. humilis	*Camellia*[c], *Stewartia*[c], *Symplocos*[c]	0.070	10.67	$\mathbf{1.0 \times 10^{-7}}$
Ch. macrocarpa	*Camellia*[c], *Stewartia*[c], *Symplocos*[c]	0.068	10.48	$\mathbf{1.0 \times 10^{-7}}$
G. lindleyana	*Apostasia, Oncidium, Phalaenopsis*	0.017	3.12	$\mathbf{3.7 \times 10^{-6}}$
Corynaea	*Heisteria*, *Ximenia*	0.516	53.85	$\mathbf{1.0 \times 10^{-7}}$
Cytinus	*Arabidopsis, Carica*	0.087	11.59	$\mathbf{1.0 \times 10^{-7}}$
Hydnora	*Aristolochia, Dr. granadensis, Saururus*	0.264	52.54	$\mathbf{1.0 \times 10^{-7}}$
nr18S				
Ca. filiformis	*Cinnamomum, Laurus, Sassafras*	0.006	1.16	1.1×10^{-1}
A. indica	*Antirrhinum*[c], *Olea*, *Sesamum*[c]	0.020	1.51	$\mathbf{1.0 \times 10^{-5}}$
B. japonica	*Heisteria*[c], *Santalum* , *Ximenia*	0.111	3.88	$\mathbf{1.0 \times 10^{-7}}$
Cu. australis	*Convolvulus, Ipomoea*	0.036	1.83	$\mathbf{1.0 \times 10^{-7}}$
M. kanehirai	*Clethra*[c], *Pyrola*[c], *Symplocos*[c]	0.041	2.03	$\mathbf{1.0 \times 10^{-7}}$
O. coerulescens	*Antirrhinum*[c], *Olea*, *Sesamum*[c]	0.007	1.17	$\mathbf{2.2 \times 10^{-2}}$
Ch. humilis	*Clethra*[c], *Pyrola*[c], *Symplocos*[c]	0.009	1.22	$\mathbf{2.2 \times 10^{-2}}$
Ch. macrocarpa	*Clethra*[c], *Pyrola*[c], *Symplocos*[c]	0.007	1.17	$\mathbf{4.9 \times 10^{-2}}$
G. lindleyana	*Apostasia, Cymbidium, Oncidium*	0.021	1.55	$\mathbf{3.9 \times 10^{-5}}$
Corynaea	*Heisteria*, *Santalum* , *Ximenia*	0.048	2.20	$\mathbf{1.0 \times 10^{-7}}$
Cytinus	*Arabidopsis, Carica*	0.038	1.90	$\mathbf{1.0 \times 10^{-7}}$
Hydnora	*Aristolochia, Dr. winteri, Saururus*	0.046	2.23	$\mathbf{1.0 \times 10^{-7}}$

[a] *Amborella* is used as outgroup for all of SSU rDNAs; relative rate tests of mt19S rDNA sequences exclude V1 and V7 regions.

[b] Significance of the *P* values <0.05.

[c] Autotrophic species included in the analyses of nucleotide pattern.

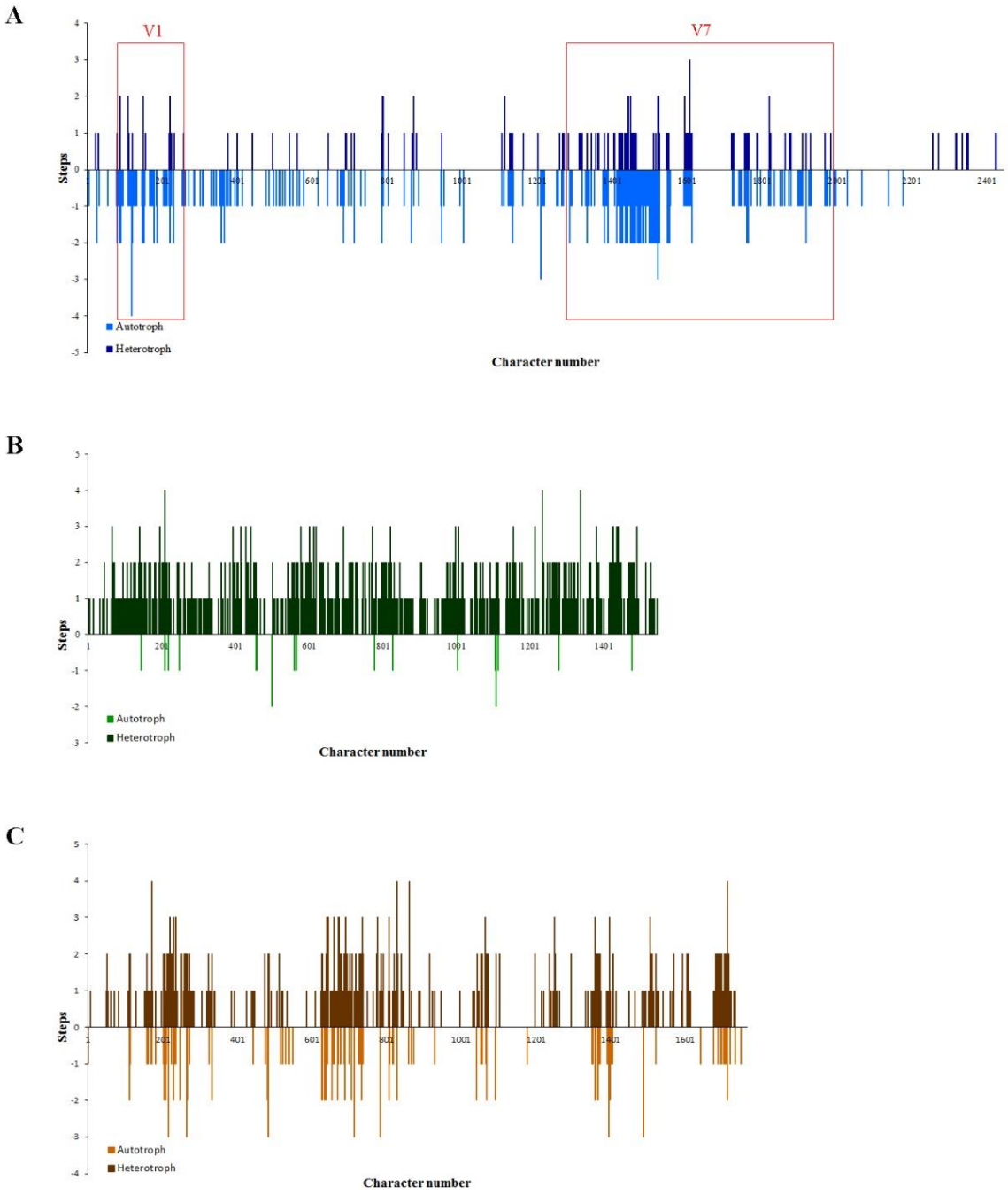


Figure 3-1. Patterns of nucleotide substitution in A: mt19S, B: pt16S and C: nr18S rDNAs across selected taxa. The histograms above the x-axis are patterns of the heterotrophic plants that completely lost their photosynthetic ability, and the histograms below are patterns of their autotrophic counterparts.

References

Bailey, C.D., T.G. Carr, S.A. Harris, and C.E. Hughes. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol Phylogenet Evol. 29: 435-455.

Barnwell, P., A.N. Blanchard, J.A. Bryant, N. Smirnoff and A.F. Weir. 1998. Isolation of DNA from the highly mucilagenous succulent plant *Sedum telephium*. Plant Mol Biol Rep. 16: 133-138.

Chase, M.W. and J.L. Reveal. 2009. A phylogenetic classification of the land plants to accompany APG III. Bot J Linn Soc. 161: 122-127.

dePamphilis, C.W., N.D. Young and A.D. Wolfe. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: Many losses of photosynthesis and complex patterns of rate variation. Proc Nat Acad Sci USA. 94: 7367-7372.

Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19: 11-15.

Drouin, G., H. Daoud and J. Xia. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 49: 827-831.

Duff, R.J. and D.L. Nickrent. 1997. Characterization of mitochondrial small-subunit ribosomal RNAs from holoparasitic plants. J Mol Evol. 45: 631-639.

Duff, R.J. and D.L. Nickrent. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. Am J Bot. 86: 372-386.

Heinhorst, S. and G.C. Cannon. 1993. DNA replication in chloroplasts. J. Cell Sci. 104: 1-9.

Hillis, D.M. and M.T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quart Rev Biol. 66: 411-453.

Houliston, G.J. and M.S. Olson. 2006. Nonneutral evolution of organelle genes in *Silene vulgaris*. Genetics. 174: 1983-1994.

Huang, C.C., K.H. Hung, W.K. Wang, C.W. Ho, C.L. Huang, T.W. Hsu, N. Osada, C.C. Hwang and T.Y. Chiang. 2012. Evolutionary rates of commonly used nuclear and organelle markers of *Arabidopsis* relatives (Brassicaceae). Gene. 499: 194-201.

Jansen, R.K. and T.A. Ruhlman. 2012. Plastid genomes of seed plants. In R. Bock and V. Knoop (eds.), Genomics of chloroplasts and mitochondria. Springer Science Business Media, Dordrecht, Heidelberg, New York, London, pp. 103-126.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16: 111-120.

Kimura, S. and K. Sakaguchi. 2006. DNA repair in plants. Chem Rev. 106: 753-766.

Kupriyanova, N.S. 2000. Conservation and variation of ribosomal DNA in eukaryotes. Mol Biol. 34: 753-765.

Lemaire, B., S. Huysmans, E. Smets and V. Merckx. 2010. Rate accelerations in nuclear 18S rDNA of mycoheterotrophic and parasitic angiosperms. J Plant Res. 124: 561-576.

Li, W.H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol. 36: 96-99.

Li, W.H., C.I. Wu and C.C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol. 2: 150-174.

Maddison, D.R. and W.P. Maddison. 2000. MacClade 4: Analysis of phylogeny and character evolution, Version 4. Sinauer Associates, Sunderland, MA, USA.

Malécot, V. and D.L. Nickrent. 2008. Molecular phylogenetic relationships of Olacaceae and related Santalales. Syst Bot. 33: 97-106.

Nedelcu, A.M. 2001. Complex patterns of plastid 16S rRNA gene evolution in nonphotosynthetic green algae. J Mol Evol. 53: 670-679.

Nickrent, D.L. 2002. Parasitic plants of the world. In J.A. López-Sáez, P. Catalán and L. Sáez (eds.), Parasitic plants of the Iberian Peninsula and Balearic Islands. Mundi-Prensa Libros, S. A., Madrid, Spain, pp. 7-27.

Nickrent, D.L. and R.J. Duff. 1996. Molecular studies of parasitic plants using ribosomal RNA. In M.T. Moreno, J.I. Cubero, D. Berner, D. Joel, L.J. Musselman, and C. Parker (eds.), Advances in Parasitic Plant Research. Dirección General de Investigación Agraria, Servicio de Publicaciones y Divulgación, Cordoba, Spain, pp. 28-52.

Nickrent, D.L., R.J. Duff, A.E. Colwell, A.D. Wolfe, N.D. Young, K.E. Steiner and C.W. dePamphilis. 2000. Molecular phylogenetic and evolutionary studies of parasitic plants. In D.E. Soltis, P.S. Soltis, and J.J. Doyle (eds.), Molecular Systematics of Plants II: DNA sequencing. Kluwer Academic Publishers, Norwell, Massachusetts, USA, pp. 211-241.

Nickrent, D.L., R.J. Duff and D.A.M. Konings. 1997a. Structural analyses of plastid-derived 16S rRNAs in holoparasitic angiosperms. Plant Mol Biol. 34: 731-743.

Nickrent, D.L. and E.M. Starr. 1994. High-rates of nucleotide substitution in nuclear small-subunit (18S) rDNA from holoparasitic flowering plants. J Mol Evol 39: 62-70.

Nickrent, D.L., O.Y. Yan, R.J. Duff and C.W. dePamphilis. 1997b. Do nonasterid holoparasitic flowering plants have plastid genomes? Plant Mol Biol. 34: 717-729.

Nielsen, B.L., J.D. Cupp and J. Brammer. 2010. Mechanisms for maintenance, replication, and repair of the chloroplast genome in plants. J Exp Bot. 61: 2535-2537.

Richard, G.F., A. Kerrest, and B. Dujon. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev. 72: 686-727.

Sánchez, R., F. Serra, J. Tárraga, I. Medina, J. Carbonell, L. Pulido, A. de María, S. Capella-Gutiérrez, J. Huerta-Cepas, T. Gabaldón, J. Dopazo and H. Dopazo. 2011. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res. 39: W470-W474.

Sakai, A., H. Takano and T. Kuroiwa. 2004. Organelle nuclei in higher plants: structure, composition, function, and evolution. Int Rev Cytol. 238: 59-118.

Sloan, D. and D. Taylor. 2012. Evolutionary rate variation in organelle genomes: The role of mutational processes. In C.E. Bullerwell (ed.), Organelle genetics. Springer, Berlin, Heidelberg, pp. 123-146.

Sloan, D.B., A.J. Alverson, M. Wu, J.D. Palmer and D.R. Taylor. 2012. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol Evol. 4: 294-306.

Soltis, D.E. and P.S. Soltis. 2000a. Choosing an approach and an appropriate gene for phylogenetic analysis. In D.E. Soltis, P.S. Soltis, and J.J. Doyle (eds.), Molecular Systematics of Plants II: DNA sequencing. Kluwer Academic Publishers, Norwell, Massachusetts, USA, pp. 1-42.

Soltis, P.S. and D.E. Soltis. 2000b. Molecular evolution of 18S rDNA in angiosperms: implications for character weighting in phylogenetic analysis. In D.E. Soltis, P.S. Soltis, and J.J. Doyle (eds.), Molecular Systematics of Plants II: DNA sequencing. Kluwer Academic Publishers, Norwell, Massachusetts, USA, pp. 188-210.

Soltis, P.S., D.E. Soltis, P.G. Wolf, D.L. Nickrent, S.M. Chaw and R.L. Chapman. 1999. The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? Mol Biol Evol. 16: 1774-1784.

Soria-Hernanz, D.F., J.M. Braverman and M.B. Hamilton. 2008. Parallel rate

heterogeneity in chloroplast and mitochondrial genomes of Brazil nut trees (Lecythidaceae) is consistent with lineage effects. Mol Biol Evol. 25: 1282-1296.

Su, H.J. and J.M. Hu. 2012. Rate heterogeneity in six protein-coding genes from the holoparasite *Balanophora* (Balanophoraceae) and other taxa of Santalales. Ann Bot. 110: 1137-1147.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28: 2731-2739.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882.

Weider, L.J., J.J. Elser, T.J. Crease, M. Mateos, J.B. Cotner, and T.A. Markow. 2005. The functional significance of ribosomal (r)DNA variation: impacts on the evolutionary ecology of organisms. Ann Rev Ecol Evol. 36: 219-242.

Weisburg, W.G., S.M. Barns, D.A. Pelletier, D.J. Lane. 1991. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 173: 697-703.

Wolfe, K.H., D.S. Katz-Downie, C.W. Morden and J.D. Palmer. 1992. Evolution of the plastid ribosomal RNA operon in a nongreen parasitic plant: Accelerated sequence evolution, altered promoter structure, and tRNA pseudogenes. Plant Mol Biol. 18: 1037-1048.

Young, N.D. and C.W. dePamphilis. 2005. Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. BMC Evol Biol. 5: 16-25.

Supplementary Data

Figure S3-1. The alignment of mt19S rDNA sequences in the nine heterotrophic plants. The V1 and V7 regions of m19S rDNAs were excluded.

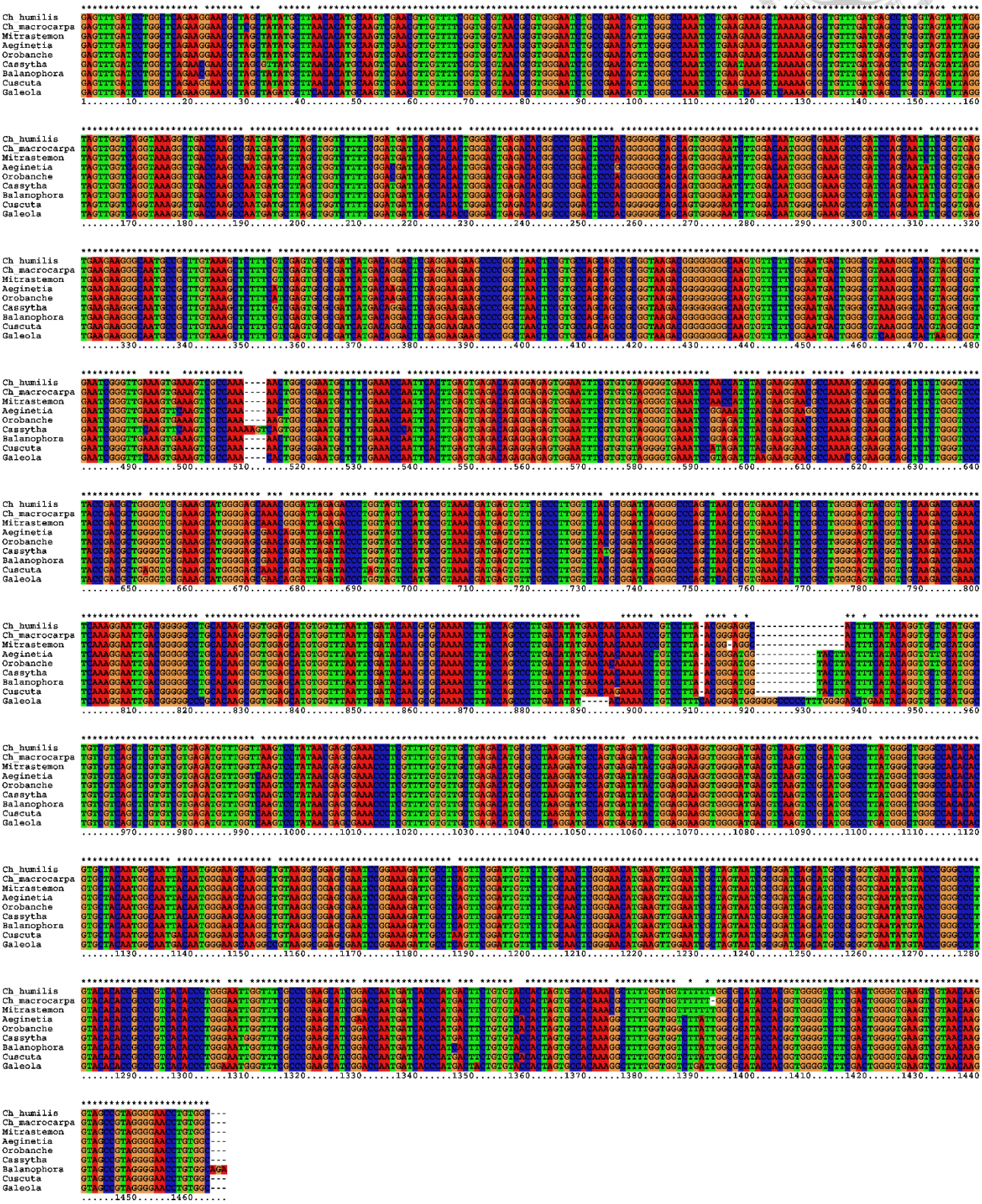


Figure S3-2. The alignment of pt16S rDNA sequences in the nine heterotrophic plants.

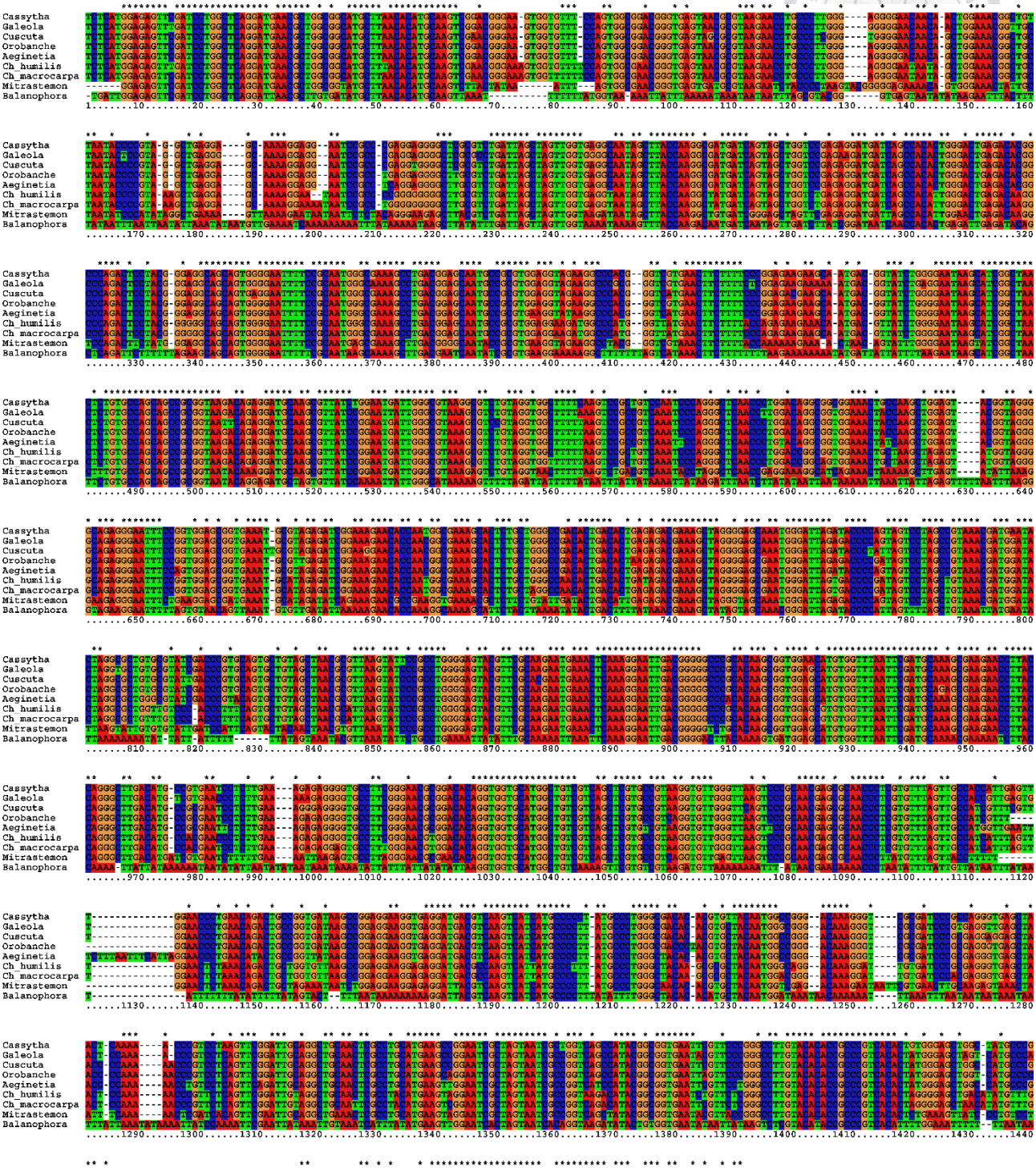
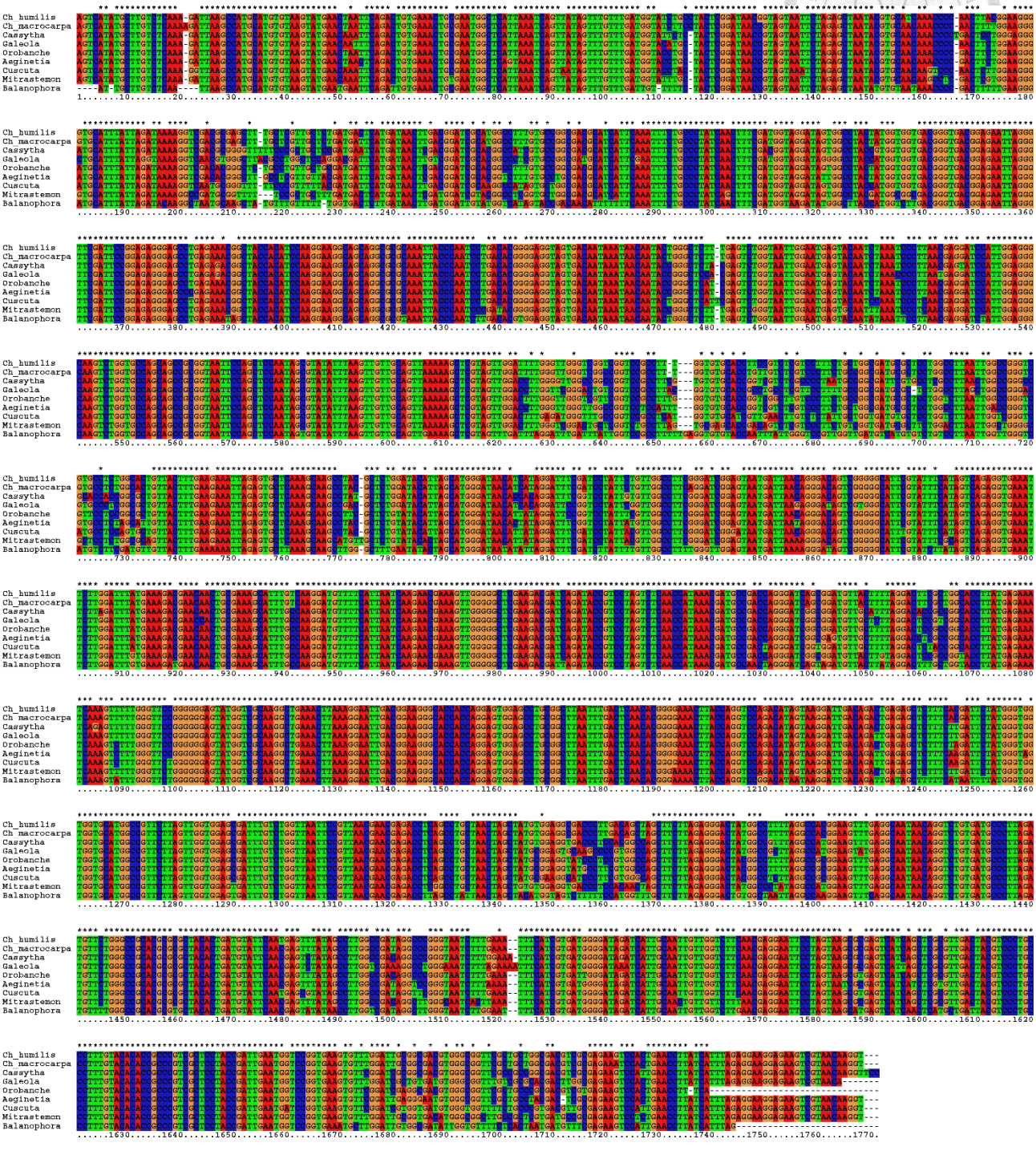


Figure S3-3. The alignment of nr18S rDNA sequences in the nine heterotrophic plants.



Chapter 4

Complete plastid genome sequence of the non-photosynthetic plant *Mitrastemon kanehirai*

Abstract

Plastids in most green plants are where photosynthesis takes place, and it is believed that plastid is derived from an ancient endosymbiosis of cyanobacteria. The size of plastome ranges from 100 to 200 kb in most land plants, and the gene content and organization of the genome are highly conserved within land plants. Previous studies have shown that the non-photosynthetic plants only retain a reduced plastome and lost most of the photosynthesis related genes. In this study, I sequenced the complete plastid genome of a non-photosynthetic plant, *Mitrastemon kanehirai*, by using next generation sequencing technology. *Mitrastemon kanehirai* possesses a plastome of 25,740 bp, which is smallest plastome that have been described in land plants. All genes related to photosynthesis are lost and the inverted repeat region is absent. Only 26 genes were retained, including 4 rRNAs, 4 tRNAs and 18 protein-coding genes. Most of these genes are involved in translation machinery. Despite the enormous reduction, the *M. kanehirai* plastome still retain a functional gene expression system. DNA transfer from the plastid to the nucleus and horizontal transfer from the host to the parasite were also observed in *M. kanehirai*. In addition, we found that the plastome size of non-photosynthetic plants is inversely related to their pt16S rDNA substitution rate. Based on this observation, we speculated that *M. kanehirai* has almost reached a minimum limit of plastid genome size.

Keywords: heterotrophic plant, *Mitrastemon kanehirai*, non-photosynthetic plant, plastid gene, plastome.

Introduction

Plastid genome

Plastids are the eukaryotic organelles responsible for photosynthesis and derived from an initial endosymbiosis of a cyanobacteria about 1-1.5 billion years ago (Martin and Kowallik, 1999; Butterfield, 2000; McFadden, 2001; Douzery et al., 2004; Keeling, 2004; Yoon et al., 2004; Waters and Langdale, 2009). Plastomes (i.e., plastid genomes) have experienced a process of severe genome reduction during the course of endosymbiosis and plant evolution. The ptDNAs are circular molecules ranging in size from 100 to 200 kb and can be divided into four parts with two inverted repeats (IR, about 25 kb) separating large single copy region (LSC) and small single copy region (SSC) of land plant plastomes (Palmer and Delwiche, 2000; Raubeson and Jansen, 2005). Plastid genomes encode only about 5~10% as many genes as the free-living cyanobacteria, indicating that many genes were either lost or transferred to the nucleus during the process of genome reduction. It should be noted that plastids contain just about as many proteins as their cyanobacterial relatives, and previous studies suggested that about 1,000 to 5,000 proteins in higher plants are targeted to plastids (Abdallah et al., 2000; Cavalier-Smith, 2000; Rujan and Martin, 2001; Martin et al., 2002). The genes for these proteins are likely re-located in the nuclear genome and it has been estimated that about 4,500 of *Arabidopsis* protein-coding genes (about 18% of the total) were obtained from the cyanobacterial

ancestor of plastids (Martin et al., 2002).

The transfer of plastid genes to the nucleus mostly occurred at the initiation stage of endosymbiosis; however, several studies have shown that the plastid-to-nucleus DNA transfer is a prevailing, continuing and natural process at unpredictably high frequency (Timmis et al., 2004). The core gene sets of plastids remain mostly the same, although the gene relocation process happened massively and in parallel during the early evolution of algal diversification. It suggests that the relocation of plastid genomes was under similar selective pressures but not reduced randomly.

The genes that retained in plastids can be divided into three categories: (1) transcription and translation related genes; (2) photosynthetic genes; and (3) other biosynthetic genes (Lohan and Wolfe, 1998; Martin et al., 2002; Odintsova and Yurina, 2003). Two main theories have been proposed to explain why these genes retained in plastids: the 'hydrophobicity' and 'redox control' (Race et al., 1999; Timmis et al., 2004). The former one suggests that hydrophobic proteins are difficult to import into organelles, and so must be retained in organelle genomes, rather than transferred to the nuclear genome. The 'redox control' theory assumes that it is necessary for organelles to retain genes involved in their electron-transfer chain and gene expression machinery in response to redox state, and hence they can maintain redox balance avoiding the production of highly toxic reactive oxygen species (Timmis et al., 2004).

In addition, a ‘co-inheritance’ theory describes that gene co-inheritance probably is essential to successful functional gene transfer and thus restricts the transfer of organellar genes to the nucleus (Brandvain et al., 2007). However, these theories address primarily the need to keep photosynthetic genes in plastids. Other hypotheses, such as the ‘essential tRNAs’ and ‘limited transfer window’ hypothesis, which account for the retention of a plastome in non-photosynthetic species, have been proposed as well (Barbrook et al., 2006). For example, the ‘essential tRNAs’ hypothesis speculated that plastid $tRNA^{Glu}$ is essential for heme biosynthesis in plants and algae, so the gene *trnE* must be retained in the plastome. Similarly in apicomplexan parasites, plastid $tRNA^{fMet}$ is essential for mitochondrial protein synthesis, and therefore the *trnfM* gene is retained. The ‘limited transfer window’ hypothesis argues that the opportunity for DNA transfer is greatly reduced in protists such as the apicomplexans and *Chlamydomonas* because they possess only a single plastid per cell, and such transfers mostly would be lethal to the cell. Therefore, the retention of genes in apicoplast genomes might just be incapable to get them out.

The plastomes of non-photosynthetic plants

Non-photosynthetic plants lost photosynthetic ability and relied on other sources for carbon hydrates. These heterotrophic plants can be either parasitic or mycoheterotrophic, depends on the mode of nutrient uptake and if they have direct contact with host plants.

Non-photosynthetic plants retain remnant plastids and their plastome is usually highly reduced (Barbrook et al., 2006). Until now, there are only four complete plastid sequences available for non-photosynthetic plants, *Cistanche deserticola* (102,657 bp, Orobanchaceae) (Li et al., 2013), *Epifagus virginiana* (70,028 bp, Orobanchaceae) (Wolfe et al., 1992), *Rhizanthella gardneri* (59,190 bp, Orchidaceae) (Delannoy et al., 2011) and *Neottia nidus-avis* (92,060 bp, Orchidaceae) (Logacheva et al., 2011). Among these, *C. deserticola* and *E. virginiana* are holoparasitic, while the others are mycoheterotrophic plants. In addition, it has been estimated that *Conopholis americana* has the plastome of 43 kb (Colwell, 1994) and *Cytinus ruber* plastome is approximately 20 kb which is the smallest documented for angiosperms so far (Nickrent et al., 1997b).

All of the plastid genomes of these non-photosynthetic plants are smaller than half the size of *Nicotiana tabacum* plastome (156 kb) or other green seed plants. Among these, *C. deserticola*, *E. virginiana* and *N. nidus-avis* plastomes all retain nearly full-sized inverted repeat regions (22 to 24 kb), but the IRs of *R. gardneri* plastome are less than 10 kb, which is much shorter than the other non-photosynthetic plants. In the four plastomes with complete sequence, nearly all of their photosynthesis related genes were either lost or became pseudogenes. Despite the size variation, the four plastomes share a very similar gene content and gene order structure. Furthermore, several studies have revealed that plastid genes of the non-photosynthetic show accelerated evolutionary rate in some non-

photosynthetic plants (dePamphilis et al., 1997; Nickrent et al., 1997a). Given that the unique characteristics described above, non-photosynthetic plants can be seen as the perfect natural mutants to study the functional aspects on evolution of plastomes. However, little is known about the genome structure of the plastome in non-photosynthetic plants. To date, the analysis of complete plastome from non-photosynthetic plants has been restricted to two families, Orobanchaceae and Orchidaceae. Here, I presented the entire plastome of a holoparasitic plant, *Mitrastemon kanehirai* Yamamoto, to increase the understanding of parallel genome reduction in plastid evolution.

Mitrastemon kanehirai is a root holoparasitic plant parasitized on roots of Fagaceae species, is endemic to Taiwan (Yang and Lu, 1996). The genus *Mitrastemon* contains only 2-4 species, and *M. kanehirai* sometimes was treated as a variety or as synonym of *Mitrastemon yamamotoi* (Matuda, 1947; Meijer and Veldkamp, 1993). There is only one plastid sequence (16S rDNA) of *Mitrastemon* that was available on GenBank. The 16S sequence showed higher accelerated substitution rate than the pt16S of *E. virginiana*, which is a holoparasite with a reduced plastome of size 70 kb (Nickrent et al., 1997a; Nickrent et al., 2000). In this study, I sequenced the complete plastid genome of *M. kanehirai* by using next generation sequencing technology. By analyzing and comparing the *M. kanehirai* plastome sequence with other plants, it provides a more comprehensive

insight into the evolution of plastid genome in non-photosynthetic plants.

Materials and methods

Plant material and DNA extraction methods

Mitrastemon kanehirai is a non-photosynthetic root endoparasite native to Taiwan. The material used in this study (*M. kanehirai*: Lienhuachih, Nantou County, Oct. 12, 2010, *Hu1810*) was collected in fresh and immediately immersed in liquid nitrogen, and subsequently stored in -20℃. Frozen tissues were used for the following DNA and RNA extraction methods.

Total genomic DNA was isolated using a modified CTAB method developed by Barnwell et al. (1998) for the highly mucilaginous succulent plant. Plastid DNA was enriched by combining Milligan's protocol (Milligan, 1989) with the Tri-Plant Genomic DNA Reagent Kit (Geneaid). Plastid pellet was isolated by centrifugation at 6,000 g for 20 min instead of 3,000 g for 10 min in Milligan's protocol, and then the DNA was extracted from pellet by using Tri-Plant Genomic DNA Reagent Kit. Details of the extraction method can been seen in Chapter 2.

Long PCR and reverse transcription-PCR

Long PCR was performed by using BD Advantage^{TM} 2 Polymerase mix (Clontech) according to manufacturer's instructions, and the primers used are listed in supplementary

data Table S4-1. The PCR products were then cloned into the pCR® -XL-TOPO® vector (Invitrogen) following user manual. One*Taq*™ DNA Polymerase (New England Biolabs) was used to amplify the DNA fragments of *M. kanehirai* plastome to further confirm its sequence.

Total RNAs were extracted using Plant Concert Reagent (Invitrogen). The cDNA was synthesized using the SuperScript™ III RNase H- Reverse Transcriptase Kit (Invitrogen), and then used as a template in the following PCR.

Real-time quantitative PCR and Southern blot analyses

Real-time PCR and data analysis were performed in the CFX96™ Real-Time PCR Detection Systems (Bio-Rad Laboratories). The PCR mixture (20 μL) contained 10 μL KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems), 50 nM or 100 nM of each primers (Supplementary data Table S4-1) and 20 ng of extracted DNA was used as template. The amplification program initiated at 95℃ for 3 min, followed by 40 cycles at 95℃ for 10 s and 64℃ for 30 s, and finally 95℃ for 10 s. Melting curve analysis was carried out after PCR amplification. All experiments were performed in triplicate.

For Southern blot analyses, the genomic DNA of *M. kanehirai* was digested with *Bam*HI, *Eco*RI and *Hind*III separately (10 μg/reaction; 2 U enzyme/μg DNA) for 1 h at 37℃, and then were separated on a 0.7% TAE agarose gel. The hybridization probes were labeled with DIG by containing DIG-11-dUTP (Roche Applied Science) in the PCR

reaction mixture. CDP-*Star* (Roche Applied Science) was used for detecting the probe signal on blots. The procedures of DIG labeling, blot transfer, hybridization (at 65℃) and detection all followed DIG Application Manual for Filter Hybridization (Roche Applied Science).

Next generation sequencing and assembly

The next generation sequencing was provided by the sequencing service of National Yang-Ming University VYM Genome Research Center (VYMGC). A paired-end library was prepared from the plastid DNA-enriched sample and sequenced using the GAII platform (Illumina). The *de novo* assembly was carried out using CLC Genomics Workbench 4.0.2 (CLC Bio) by VYMGC. The plastid genome was also assembled using Velvet 1.2.07 (Zerbino and Birney, 2008) with the assembly parameter was set to k = 89.

Annotation and plastome map drawing

The initial annotation of *M. kanehirai* plastome was produced using online automatic annotator DOGMA (Wyman et al., 2004). The gene annotation was further surveyed by the orf prediction of web server WebMGA (Wu et al., 2011) and NCBI Conserved Domain-Search (CD-Search); and the identification of tRNAs was done by tRNAscan-SE 1.21 (Schattner et al., 2005). The RNAs were then verified by NCBI BLAST search for final adjustment of gene annotation. The map of the plastome was drawn using OGDRAW online tool (Lohse et al., 2013). We selected four plastomes for

comparison: *N. tabacum* (the reference, NC_001879), *Ardisia polysticta* (the closest relative in Ericales, NC_021121), and two other non-photosynthetic plants, *E. virginiana* (NC_001568) and *R. gardneri* (NC_014874). The comparison of plastomes was performed using Progressive Mauve (Darling et al., 2010) with default parameters and minor manual modification.

Results

Plastome sequence

Initially, I obtained a large fragment (Fragment A, 4,585 bp) of *M. kanehirai* plastid sequence by performing long PCR with universal primers which locate at 16S rDNA and 5S rDNA (Supplementary data Table S4-1). Part of the obtained sequence is highly similar to *M. yamamotoi* pt16S rDNA sequence that has been reported on GenBank (Nickrent et al., 1997a). Most NGS contigs larger than 10 kb assembled by both software are mitochondrial sequences. Nonetheless, there are contigs that their sequences overlapping with Fragment A from both assemblies. The largest contig that contains *M. kanehirai* plastid sequence yield from CLC Genomics Workbench is 16,032 bp (coverage = 3127.77X) at size, and 15,116 bp (coverage = 386.20X) from Velvet. The two contigs could be further assembled to form a circular genome of 25,740 bp (Fig. 4-1). Primers (Supplementary data Table S4-1) designed based on this plastome sequence were used to

perform PCR in order to further confirm the sequence.

The *M. kanehirai* plastome is smaller than the plastome of the non-photosynthetic orchid *R. gardneri* (Delannoy et al., 2011), which makes this genome the smallest sequenced plastid genome of land plants. The IR was absent in the plastome, but the regions corresponding to the large single-copy region and the small single-copy region can still be recognized. All genes related to photosynthesis as well as the transcription-related genes were lost. The *M. kanehirai* plastome contains only 26 genes encoding 18 proteins, 4 rRNAs and 4 tRNAs (Table 4-1). The overall GC content of the plastome is 22.5%, and rRNA genes has the highest GC content of 39.2% (Fig. 4-1 and Table 4-2). All four rRNA genes are present in the genome, but their sequences are much more divergent from other green plants (Table 3-3). There are only four tRNA genes retained: *trnfM-cau*, *trnC-gca*, *trnE-uuc* and *trnI-cau*. Fourteen of the *M. kanehirai* plastid genes encode proteins of the translation machinery, including 3 *rpl* genes, 10 *rps* genes and an initiation factor (*infA*). The other four protein-coding genes, *accD*, *clpP*, *ycf1* and *ycf2*, are also retained in other non-photosynthetic plants (Table 4-1). Despite size variation and sequence divergence compared with other green plants, the CD-Search results of most of these deduced proteins met the criteria of specific hit suggests that they still maintain their functions. It is less certain with *rps8* and *rps18* since CD-Search found only non-specific hit for these two putative proteins. The functionality of *ycf2* remained uncertain

because of its highly divergent sequence and lacking intact functional domain.

I examined the RNA expression of several genes to further explore the functionality of *M. kanehirai* plastid genes. The expression of five translation-related genes which include *infA*, *rrn16* and 3 *rps* genes (*rps3, 7, 14*) were detected, as well as the other four genes — *accD*, *clpP*, *ycf1* and *ycf2*. (Fig. 4-2). It indicates that the two large ORFs, *ycf1* and *ycf2*, are very likely functional. In addition, detection of the correctly spliced cDNA of *clpP* shows that RNA splicing is also occurring in *M. kanehirai* plastids.

Other large DNA fragments

In my preliminary experiment, long PCR was used to amplify *M. kanehirai* plastid sequence with universal primers from the conserved regions. In addition to the sequence of Fragment A matches the *M. yamamotoi* pt16S rDNA (U67742), I also found other ptDNA-like sequences by using primers based on the IR region (Dhingra and Folta, 2005) (Fig. 4-3). One of these DNA fragments (Fragment B, amplified by primers 16S977F and IRB27R), its sequence is highly similar to the plastid IR segment ranging from *rrn16* to *ycf1* in green plants (Supplementary data Table S4-2). Fragment A and B both contain four rDNA sequences. Southern blots were performed with hybridization probes specific to these two fragments (Probe A to Frag. A, Probe B to Frag. B, Supplementary data Table S4-1). The result of Southern bolts showed that signals of Probe B were much weaker than Probe A from all three restriction enzymes (Fig. 4-4, compared with the marker

signals), it suggests that Frag. A and B might be not located in the same genome, i.e. Frag. B might be from nuclear genome, or other organelle genome. The blots showed hybridization to a fragment about 8 kb for *Eco*RI digestion and about 1 kb for *Hind*III (Fig. 4-4A), which is corresponding well with the cutting map of *M. kanehirai* plastome (Fig. 4-1). In Fig. 4-4, there is no signal yielded for *Bam*HI digestion, it might be due to the enzyme produced only one full-size fragment which was too large that its transfer from the gel to membrane was inefficient.

In addition, real-time quantitative PCR was performed to further examine the two fragments by comparing their content with nr18S rDNA in the total genomic DNA (Fig. 4-5). The signal of nr18S rDNA emerged first, followed by Frag. A and Frag. B was the last. Although the *Cq* (quantification cycle) values of Frag. A and nr18S rDNA were close, nr18S rDNA is a multiple-copy nuclear gene; it indicates Fragment A isn't located in the nuclear genome (Fig. 4-6A). And the *Cq* value difference between Frag. A and B confirms that the two fragment are indeed not located in the same genome. The content of Frag. B is lowest since *Cq* inversely correlated with starting copies (Fig. 4-6B). Also, numbers of mitochondrial genome per cell is high in meristematic and reproductive tissues where the total DNA extracted from, so Frag. B is unlikely to reside in mitochondrial genomes. It is reasonable to assume Frag. B locates in the *M. kanehirai* nuclear genome.

Discussion

Plastid genes retained in M. kanehirai

Mitrastemon kanehirai plastome contains only 26 genes and is the smallest confirmed plastid genome, being smaller than the apicoplastat genome of the malaria parasite *Plasmodium vivax* (30 kb, NC_017932). All the genes required for photosynthesis were lost, as well as five genes related to RNA metabolism. The loss of the *matK* and *rpo* genes was not surprised since they were lost or turned into pseudogenes in some other non-photosynthetic plants (Table 4-1). Despite the loss of *matK*, *M. kanehirai* plastome retains four group II introns, one of which was shown to be correctly spliced (*clpP*). The loss of *matK* has been observed in *Cuscuta* species from subgenus *Grammica*, which have also lost group IIa introns (McNeal et al., 2009). In addition, it has been reported that *matK* became a pseudogene in many orchids and is completely lacked in *R. gardneri* (Kores et al., 2001; Delannoy et al., 2011; Logacheva et al., 2011). The result suggests that the role of *matK* for splicing in *M. kanehirai* is likely non-essential as in these mycoheterotrophic orchids.

All four rRNA genes are present in *M. kanehirai*, but unlike in *R. gardneri* they do not share high similarity with their orthologs from other Ericales species. The same phenomenon was also observed in ribosomal protein genes**.** *Mitrastemon kanehirai* plastome encodes 14 ribosomal protein genes, three less *rpl* genes than *R. gardneri* plastome (*rpl14, 20, 23*). Although these genes maintain their functional domain in the

predicted amino acid sequences, their nucleotide sequences are highly divergent from their Ericales orthologs. There are only four tRNA genes retained in *M. kanehirai*: *trnC*, *trnE*, *trnfM* and *trnI*. The later three are preserved in all examined non-photosynthetic plants (Table 4-1). Plastid $tRNA^{Glu}$, encoded by the *trnE* gene, is essential for protein and tetrapyrrole synthesis (Barbrook et al., 2006). In all plants, it is required in heme biosynthesis for mitochondrial respiratory complexes and other essential proteins. Because plastid $tRNA^{Glu}$ has to interact with glutamyl-tRNA reductase, glutamyl-tRNA synthetase and elongation factor EF-Tu, the cytosolic counterpart couldn't easily replace this tRNA (Barbrook et al., 2006). In prokaryotic systems, the initiator tRNA is encoded by *trnfM*, which is different from eukaryotic systems. Organellar $tRNA^{Ile}$ with a CAU anticodon in which C carries a lysidine modification, hence organellar isoleucyl-tRNA synthetase wouldn't recognized its cytosolic counterpart. Therefore, it would be difficult to replace these tRNAs with importing cytosolic tRNAs.

Two genes, *accD* and *clpP*, involved in plastid metabolism are retained, as well as two large ORF, *ycf1* and *ycf2*, which their functions are still unknown. The *accD* gene encodes a carboxylase (ACCase), which provides malonyl-CoA for the biosynthesis of fatty acid, and *clpP* codes for a catalytic subunit of a multimeric protease. The genes *accD*, *ycf1* and *ycf2* are conserved in almost all land plants, and *clpP* is the only protein-coding gene presented in all land plants. The *ycf1* and *ycf2* sequence similarity among land plants

is extraordinarily low compared to other plastid genes, but they are functional and essential for cell survival in plants (Drescher et al., 2000). The nucleotide sequence of *ycf2* in *M. kanehieai* is highly divergent and it is the only protein-coding gene which sequence did not possess intact functional domain in the result of NCBI Protein BLAST search. Despite the enormous reduction, my results indicated that *M. kanehirai* plastome is a functional gene expression system, but its sequence is too divergent to be suitable molecular markers for phylogenetic analysis.

Comparison of the plastome with other non-photosynthetic plants

I compared the plastomes of three non-photosynthetic plants with two green plants. The result shows different degree of plastome reduction among non-photosynthetic plants (Fig. 4-7). *Epifagus virginiana* lost most of its photosynthetic genes in the LSC and retains nearly full-sized IR, while the IR were further contracted in *R. gardneri*, and then *M. kanehirai* plastome lacks the IR. These plastomes represent three stages in the progress of the plastome reduction in non-photosynthetic plants. Also, non-photosynthetic plant plastomes show the existence of a gene core set, and the gene order among them is well conserved. It suggests that the evolution of plastomes in different non-photosynthetic plants is under similar constraints. The reduction of *M. kanehirai* plastome occurred majorly in the LSC and the IR (Fig. 4-7). *Mitrastemon kanehirai* plastome lacks the IR, i.e. lost the *ycf2* part of the IRa and the rDNA part of the IRb, compared with other green

plants. The absence of the IR and more compact genome make *M. kanehirai* plastome the smallest plastid genome that have been described.

However, the plastome size might not reach the minimum yet. I analyzed two factors, the rDNA sequence variation and the genome size, in heterotrophic plants which plastome sizes were available in GenBank. The result shows that the substitution rate is inversely related to plastome size (Fig. 4-8). According to the correlation between the pt16S divergence and the plastome size, the possible candidates with small plastid genomes are *Balanophora japonica*, *Corymaea* and *Hydnora*. If the trends can apply to all other plants, the plastome size of *Cytinus* might not as small as the previous study suggested (Nickrent and Duff, 1996) (Table. 3-3).

Apart from photosynthesis related genes, the pattern of plastid gene lost in non-photosynthetic plants is inconsistent. The *rpo* genes which encode the plastid RNA polymerase were lost or became pseudogenes in *M. kanehirai*. The shortening of the IR did not show preference for the IRa or the IRb, *R. gardneri* and *M. kanehirai* retained the rDNA operon from the different copies of the IR.

In the tRNA genes among the examined non-photosynthetic plants, the patterning is not always consistent. The 'essential tRNAs' hypothesis which was introduced by Barbrook *et al* (2006) that describes plastid $tRNA^{Glu}$ is essential for heme biosynthesis, so retention of the gene *trnE* in the plastome is necessary to non-photosynthetic plants.

However, two other tRNA genes, *trnY* and *trnQ*, which was described difficult to be replaced by their cytosolic counterparts (Delannoy et al., 2011), were absent in *M. kanehirai*, although retained in *R. gardneri* (Table 4-1).

The *accD*, *ycf1* and *ycf2* genes have been reported lost in some land plant lineages, as well as evidences for transferred of ribosomal protein genes from plastid to nucleus (Xiong et al., 2009; Fleischmann et al., 2011). All these studies and non-photosynthetic plant plastome sequences support the 'essential tRNAs' hypothesis that all the plastid genes of non-photosynthetic plant eventually would be all lost, with only *trnE* maintained in the plastid in a replicating DNA minicircle transcribed by the imported nuclear-encoded plastid RNA polymerase.

The plastome GC content of non-photosynthetic plants

Except for *M. kanehirai*, the plastome GC content of non-photosynthetic plants ranges from 34.2% to 36.8%, which is within the range of 34-40% GC content among most seed plant plastomes (Table 4-2) (Jansen and Ruhlman, 2012). The GC content of their rRNA genes is close to the average value of 52.9% among 150 plastomes (Smith, 2009). The GC content is correlated to the length in the single copy regions of plastid genomes, especially in the SSC (Table 4-2). In general, GC content is higher in coding regions than in non-coding regions (Cai et al., 2006; Jansen and Ruhlman, 2012). In comparison, *N. nidus-avis* has 27 pseudogenes in its plastome of 92 kb, and *E. virginiana*

plastome has 15 of 70 kb (Li et al., 2013). *Neottia nidus-avis* plastome contains much more non-coding regions than *E. virginiana* plastome, which is probably the reason that *N. nidus-avis* has lower plastome GC content than *E. virginiana* (Burger and Lang, 2003).

The GC content is usually highest in the IR regions and lowest in the SSC of plastid genome. The higher GC content in the IR is due to the presence of rRNA genes that have the highest GC content of any coding regions (Cai et al., 2006; Jansen and Ruhlman, 2012). Therefore, if a plastome lacks the IR regions, i.e., having only one copy of rRNA genes, would lower its GC content (Smith et al., 2011). For example, an IR-lost legume - *Medicago truncatula*, has one of the lowest GC content plastome among seed plants (Cai et al., 2006; Smith et al., 2011). In addtion, the unusual high GC content of the SSC region in *R. gardneri* is due to the residence of rRNA genes in its SSC. Both *M. kanehirai* and *R. gardneri* plastome have only one copy of rRNA genes; however, *M. kanehirai* has much lower plastome GC content than *R. gardneri*. Although there is no clear correlation between GC content and genome size of plastids, plastids with small genome size indeed tend to have lower GC content (Smith, 2009).

One of the common feature of the plastome is its low overall GC content (Howe et al., 2003). The factors causing the low GC content in plastomes are poorly understood and probably differ both among and within lineages. Several hypotheses have been proposed. For instance, some argue that AT mutation pressure coupled with inefficient

plastid DNA repair process caused the low GC content of ptDNA (Howe et al., 2003; Kusumi and Tachida, 2005; Khakhlova and Bock, 2006; Lynch, 2007). Other studies invoke selection for translational efficiency to explain the high levels of AT in plastomes (Morton, 1993, 1998; Lynch, 2007). However, since the plastomes from non-photosynthetic plants are no longer exposed to high levels of reactive oxygen species and often have a reduced tRNA set, it suggests that there are a diversity of factors biasing plastomes towards A and T (Smith et al., 2011).

In addition, low GC content of the sequence would interfere with the amplification of PCR, which caused difficulties to obtain complete plastid sequence by using ordinary PCR-based amplification. It also thus difficult to verify NGS data by proceeding PCR, until the One*Taq*™ DNA Polymerase (New England Biolabs) was used to overcome the difficulties in PCR amplification.

Plastid DNA transfers

DNA transfer from the plastid to the nucleus has occurred during the course of endosymbiosis and the plant evolution. In *Arabidopsis*, it has been estimated that about 18% of nuclear genes are derived from plastid (Martin et al., 2002). DNA transfer from organellar genomes to the nucleus has been thought to be an important driving force in eukaryotic evolution (Martin et al., 2002; Huang et al., 2003; Reyes-Prieto et al., 2006; Sheppard and Timmis, 2009). Both functional gene and non-functional DNA transfer

from the plastid to the nucleus have been reported and the process continues at high frequency (Martin, 2003; Stegemann et al., 2003; Sheppard et al., 2008; Sheppard and Timmis, 2009; Xiong et al., 2009; Wicke et al., 2011; Wang et al., 2012). The transfer of plastid DNA to the nucleus could be in forms as 'bulk DNA' or through 'cDNA intermediates' (Adams et al., 2000; Adams and Palmer, 2003; Huang et al., 2003; Martin, 2003; Stegemann et al., 2003; Timmis et al., 2004). My results suggest the transfer of a contiguous piece of bulk ptDNA was also present in *M. kanehirai*. I obtained several large DNA fragments (> 4.5 kb) in my attempt to amplifying and sequencing *M. kanehirai* plastome (Supplementary data Table S4-2, Fig. 4-3). Three fragments were found that their sequences are highly similar to ptDNA of green plants, and the closest sequences of two fragments are from Ericales species. There are other fragments of sequences show only a short region (400-700 bp) similar to ptDNA.

One of these fragments carries a short ptDNA-like sequence which is most similar to ptDNA of Fagales, the order *M. kanehirai*'s host belongs to (Fig. 4-3). Massive horizontal gene transfer (HGT) in mitochondrial DNA has been observed (Goremykin et al., 2009; Keeling, 2010; Xi et al., 2012; Xi et al., 2013). In contrast, HGT was much less described in plastid DNA (Park et al., 2007; Stegemann et al., 2012; Li et al., 2013). And HGT appears to be facilitated by the intimate physical association between the parasitic plants and their hosts. It seems that *M. kanehirai* obtained ptDNA from its host via HGT.

Among the three fragments which whole sequences are similar to green plants, Fragment B was shown to reside in the nuclear genome based on the results of Southern blots and real-time quantitative PCR experiments (Fig. 4-4, -5 and -6). The other two fragments are likely to locate in the nucleus; however, we cannot exclude the possibility that they are resident in mitochondria as has been found in several species (Timmis et al., 2004; Goremykin et al., 2009; Iorizzo et al., 2012). And from their sequence similarity with green plants, these transfer events is not occurred in recent times. As for the fragments that contain short ptDNA sequences, their location are uncertain. They could reside either in the nucleus or mitochondria. In addition, the short ptDNA sequences of these fragments contain both coding and non-coding regions. It suggests that they were also transferred as the form 'bulk DNA' whether the ptDNAs were obtained from the host or not.

There are several complete non-photosynthetic plant plastome sequences have been described, and among them the smallest plastome was found in *R. gardneri* with size less than 60 kb. In this study, I present that *M. kanehirai* plastome of 25,740 bp smaller than *R. gardneri* plastome, which makes it the smallest sequenced plastid genome. There are only 26 genes, encoding 4 rRNAs, 4 tRNAs and 18 proteins, retained in *M. kanehirai* plastome, and also, the IR was absent in this plastome. Despite the enormous size reduction, *M. kanehirai* plastome shares the same core gene set with other non-

photosynthetic plants and has the gene order corresponding well to *N. tabacum*. Based on the result of plastid gene expression, *M. kanehirai* plastome is a functional gene expression system. In addition, ptDNA fragments which reside in genomes other than the plastome were found, and one of these fragments carries a short ptDNA sequence which is most similar to Fagales ptDNA. It implies that transfers of plastid DNA to other genomes and HGT from the host to the parasite has occurred in *M. kanehirai*.

Table 4-1. Gene contents (without photosynthesis related genes) of plastomes in 5 non-photosynthetic plants.

	Protein Synthesis						Transfer RNA Genes				
	Cd	Nn	Ev	Rg	Mk		Cd	Nn	Ev	Rg	Mk
infA	**+**	**+**	**+**	**+**	**+**	*trnA-ugc*	+	+	φ	–	–
rpl2	**+**	**+**	**+**	**+**	**+**	*trnC-gca*	+	+	φ	+	+
rpl14	+	+	φ	+	–	*trnD-guc*	+	+	+	+	–
rpl16	**+**	**+**	**+**	**+**	**+**	*trnE-uuc*	**+**	**+**	**+**	**+**	**+**
rpl20	+	+	+	+	–	*trnF-gaa*	+	+	+	+	–
rpl22	+	+	–	–	–	*trnfM-cau*	**+**	**+**	**+**	**+**	**+**
rpl23	φ	+	φ	+	–	*trnG-gcc*	+	+	–	–	–
rpl32	+	+	–	–	–	*trnG-ucc*	+	+	–	–	–
rpl33	+	+	+	φ	–	*trnH-gug*	+	+	+	–	–
rpl36	**+**	**+**	**+**	**+**	**+**	*trnI-cau*	**+**	**+**	**+**	**+**	**+**
rps2	**+**	**+**	**+**	**+**	**+**	*trnI-gau*	+	–	φ	–	–
rps3	**+**	**+**	**+**	**+**	**+**	*trnK-uuu*	+	–	–	–	–
rps4	**+**	**+**	**+**	**+**	**+**	*trnL-caa*	+	+	+	–	–
rps7	**+**	**+**	**+**	**+**	**+**	*trnL-uaa*	+	+	–	φ	–
rps8	**+**	**+**	**+**	**+**	**+**	*trnL-uag*	+	+	+	–	–
rps11	**+**	**+**	**+**	**+**	**+**	*trnM-cau*	+	+	+	–	–
rps12	+	+	+	φ	+	*trnN-guu*	+	+	+	–	–
rps14	**+**	**+**	**+**	**+**	**+**	*trnP-ugg*	+	φ	+	–	–
rps15	+	+	–	–	–	*trnQ-uug*	+	+	+	+	–
rps16	+	+	–	–	–	*trnR-acg*	+	+	+	–	–
rps18	+	φ	+	+	+	*trnR-ucu*	+	+	φ	–	–
rps19	**+**	**+**	**+**	**+**	**+**	*trnS-gcu*	+	+	+	–	–
						trnS-gga	+	+	φ	–	–
	RNA Metabolism					*trnS-uga*	+	+	+	–	–
	Cd	Nn	Ev	Rg	Mk	*trnT-ggu*	+	+	–	–	–
matK	+	φ	+	–	–	*trnT-ugu*	+	+	–	–	–
rpoA	φ	–	φ	–	–	*trnV-gac*	+	+	–	–	–
rpoB	φ	φ	–	–	–	*trnV-uac*	+	φ	–	–	–
rpoC1	–	–	–	–	–	*trnW-cca*	+	+	+	+	–
rpoC2	φ	φ	–	–	–	*trnY-gua*	+	+	+	+	–
	Essential Genes						Ribosomal RNA Genes				
	Cd	Nn	Ev	Rg	Mk		Cd	Nn	Ev	Rg	Mk
clpP	**+**	**+**	**+**	**+**	**+**	rrn16	**+**	**+**	**+**	**+**	**+**
accD	φ	+	+	+	+	*rrn23*	**+**	**+**	**+**	**+**	**+**
ycf1	φ	+	+	+	+	*rrn4.5*	**+**	**+**	**+**	**+**	**+**
ycf2	**+**	**+**	**+**	**+**	**+**	*rrn5*	**+**	**+**	**+**	**+**	**+**

* Cd, *Cistanche deserticola*; Nn, *Neottia nidus-avis*; Ev, *Epifagus virginiana*; Rg, *Rhizanthella gardneri*; Mk, *Mitrastemon kanehirai*; φ, pseudogene; +, present; –, missing. Genes present in all 5 plastomes are indicated in bold.

Table 4-2. Length and GC content of plastid regions in 2 photosynthetic and 5 non-photosynthetic plants.

Taxa	Length (kb)/GC content (%)				
	Total	LSC	SSC	IR	rDNA[a]
Photosynthetic plant					
Nicotiana tabacum[b]	155.9/37.8	86.7/36.0	18.6/32.1	25.3/43.2	55.4
Medicago truncatula[c]	124.0/34.0	-	-	-	54.3
Non-photosynthetic plant					
Cistanche deserticola[d]	102.7/36.8	49.1/32.8	8.8/27.5	22.4/43.0	55.0
Neottia nidus-avis[e]	92.1/34.4	36.4/29.2	7.8/25.3	23.9/39.9	54.4
Epifagus virginiana[f]	70.0/36.0	19.8/29.2	4.8/22.7	22.7/40.3	54.2
Rhizanthella gardneri[g]	59.2/34.2	26.4/29.3	13.3/37.0	9.8/38.7	51.8
Mitrastemon kanehirai	25.7/22.5	-	-	-	40.4

* LSC, large single copy region; SSC, small single copy region; IR, inverted repeats.

[a] GC content of rRNA genes.

[b] NC_001879 (Kunnimalaiyaan and Nielsen, 1997).

[c] NC_003119, an IR-lost legume.

[d] NC_021111 (Li et al., 2013).

[e] NC_016471 (Logacheva et al., 2011).

[f] NC_001568 (Wolfe et al., 1992).

[g] NC_014874 (Delannoy et al., 2011).

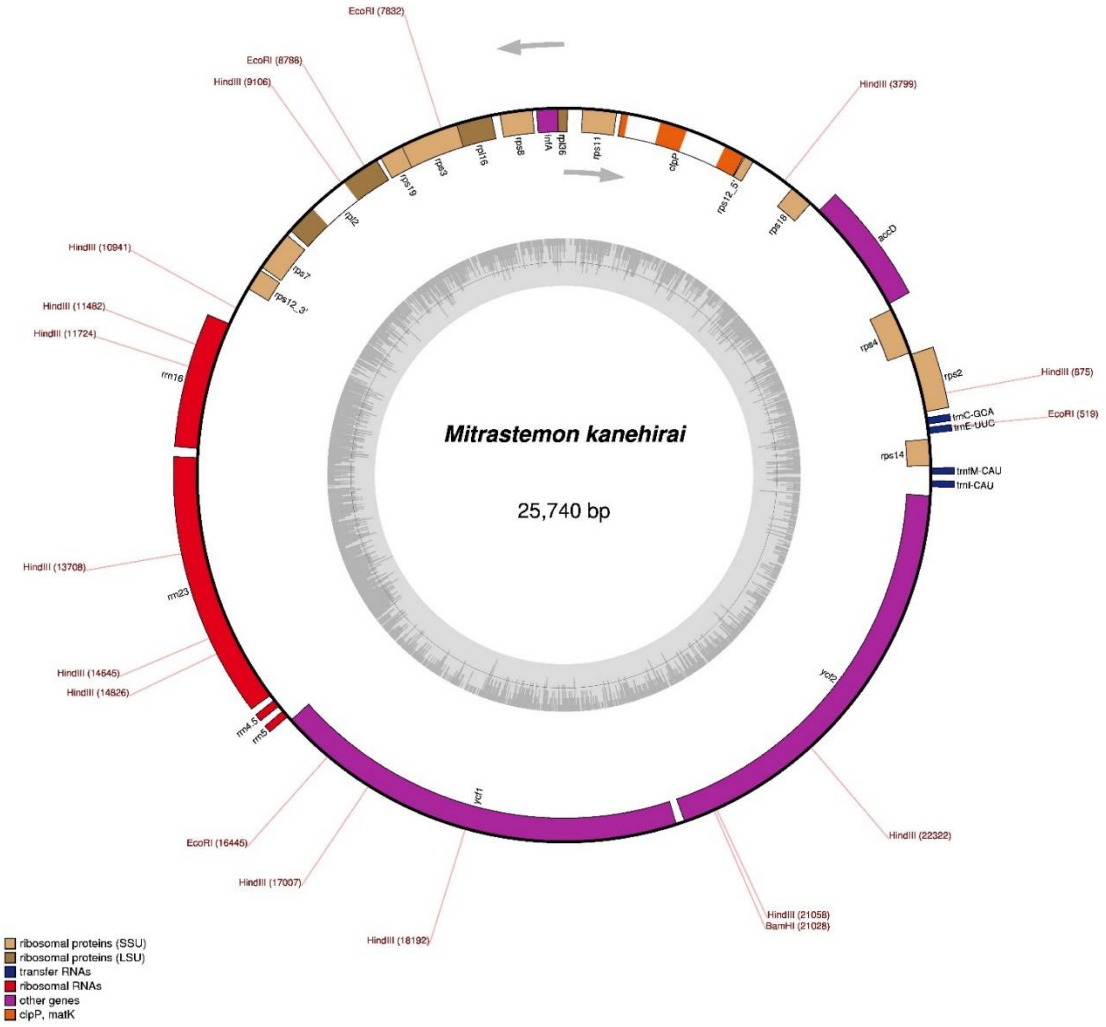


Figure 4-1. Circular map of *Mitrastemon kanehirai* plastome. Genes shown inside the circle are transcribed clockwise, those outside the circle are transcribed counterclockwise. The GC content is indicated in the inner circle. The restriction enzyme, *Bam*HI, *Eco*RI and *Hind*III, cutting sites are also shown in the map.

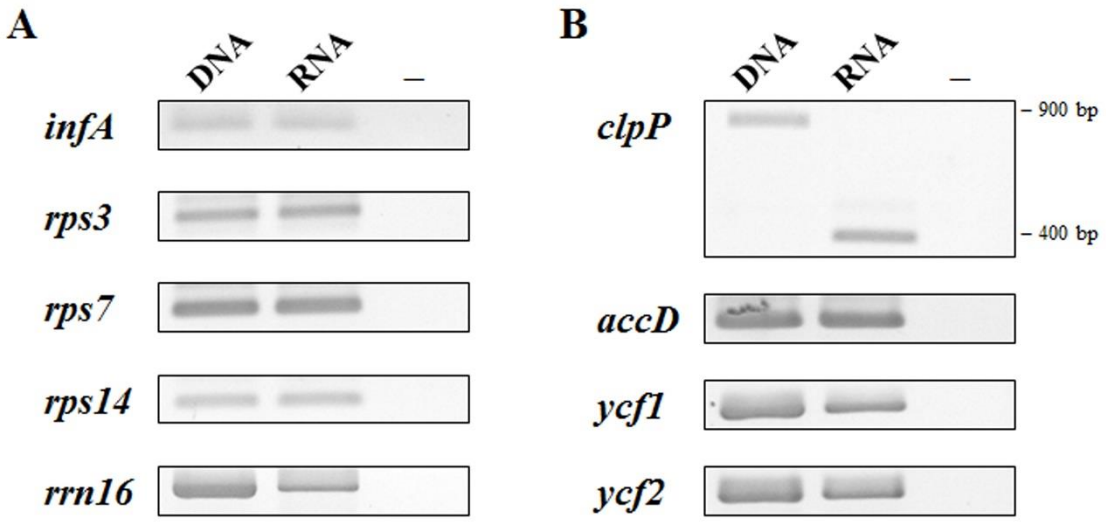


Figure 4-2. Expression of the plastid genes in *Miteastemon kanehirai*. A. Expression of five translation-related genes include one rRNA gene. B. Expression of the four essential genes. DNA, total genomic DNA was used as template in PCR; RNA, cDNA was synthesized from RNA by using gene specific primers, and then was used as template in PCR; –, negative control.

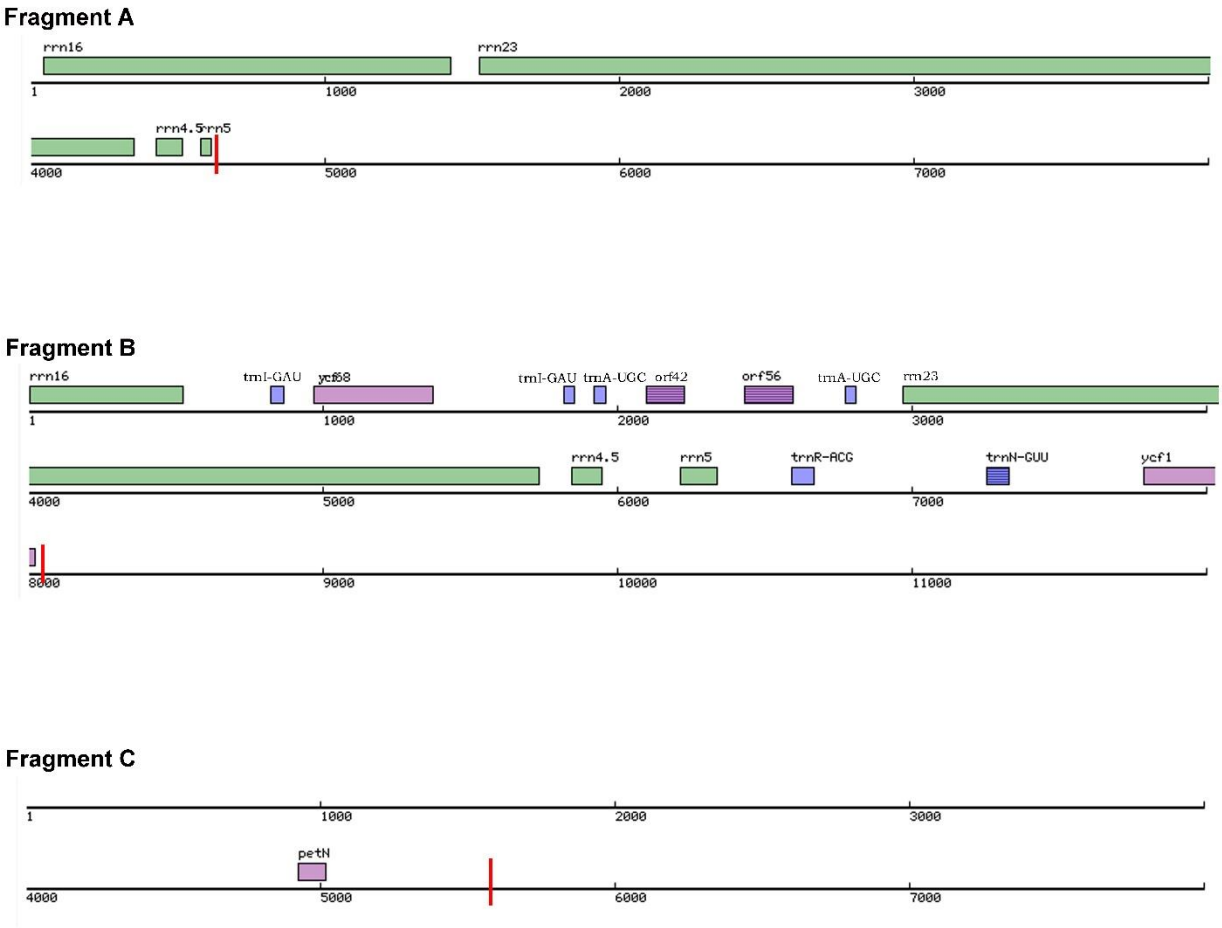


Figure 4-3. Maps of three large DNA fragments. Fragment A: amplified by primers 16S34F and IRB27R; Fragment B: amplified by primers 16S977F and IRB27R; Fragment C: amplified by primers IRB2F and IRB27R. These maps were produced by online automatic annotator DOGMA with minor modification.

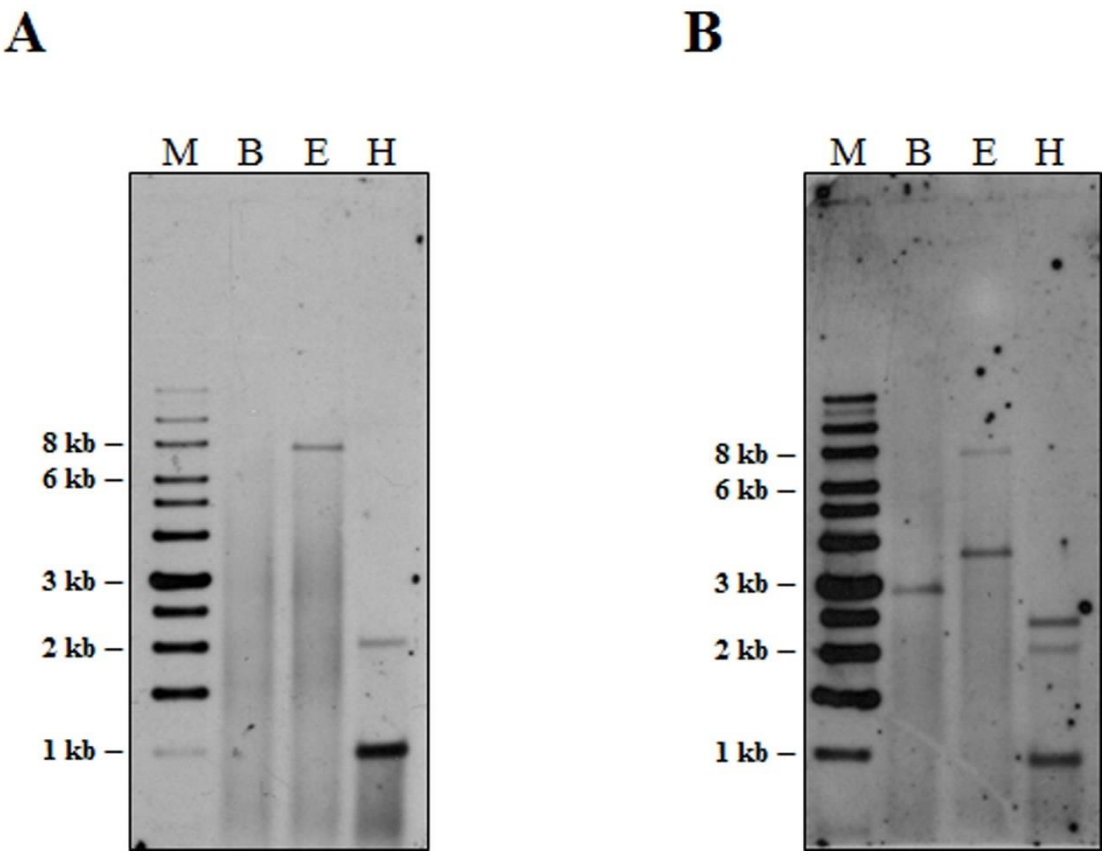


Figure 4-4. Results of Southern blot analyses of *Mitrastemon kanehirai*. A. Blot using a probe specific to a *M. kanehirai* fragment (Fragment A) (3-min exposure). B. Blot using a probe specific to a ptDNA-like fragment (Fragment B) (12-min exposure). M, DNA ladder; B, *Bam*HI-digested DNA; E, *Eco*RI-digested DNA; H, *Hind*III-digested DNA.

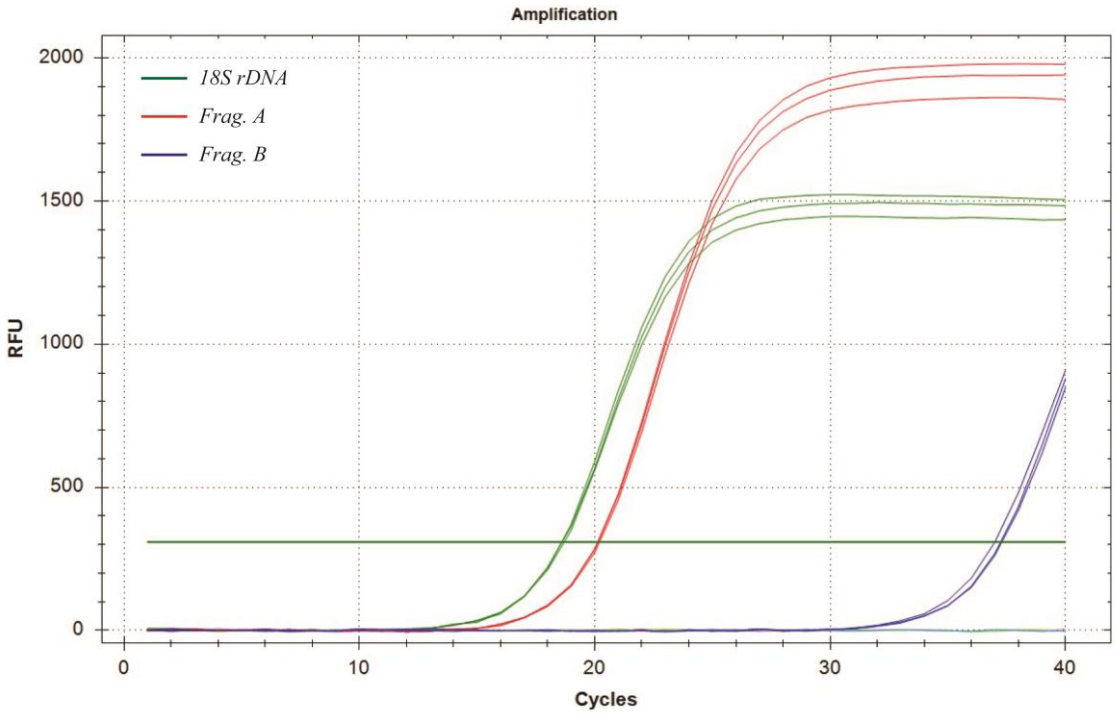


Figure 4-5. The real-time PCR amplification plot of three *Mitrastemon kanehirai* DNA fragments. The primers were designed specific to the three fragments: 18S rDNA, nuclear 18S rDNA; Frag. A, the plastid fragment similar to *M. yamamotoi* 16S rDNA sequence; Frag. B, the fragment which sequence is highly similar to plastid IR of green plants.

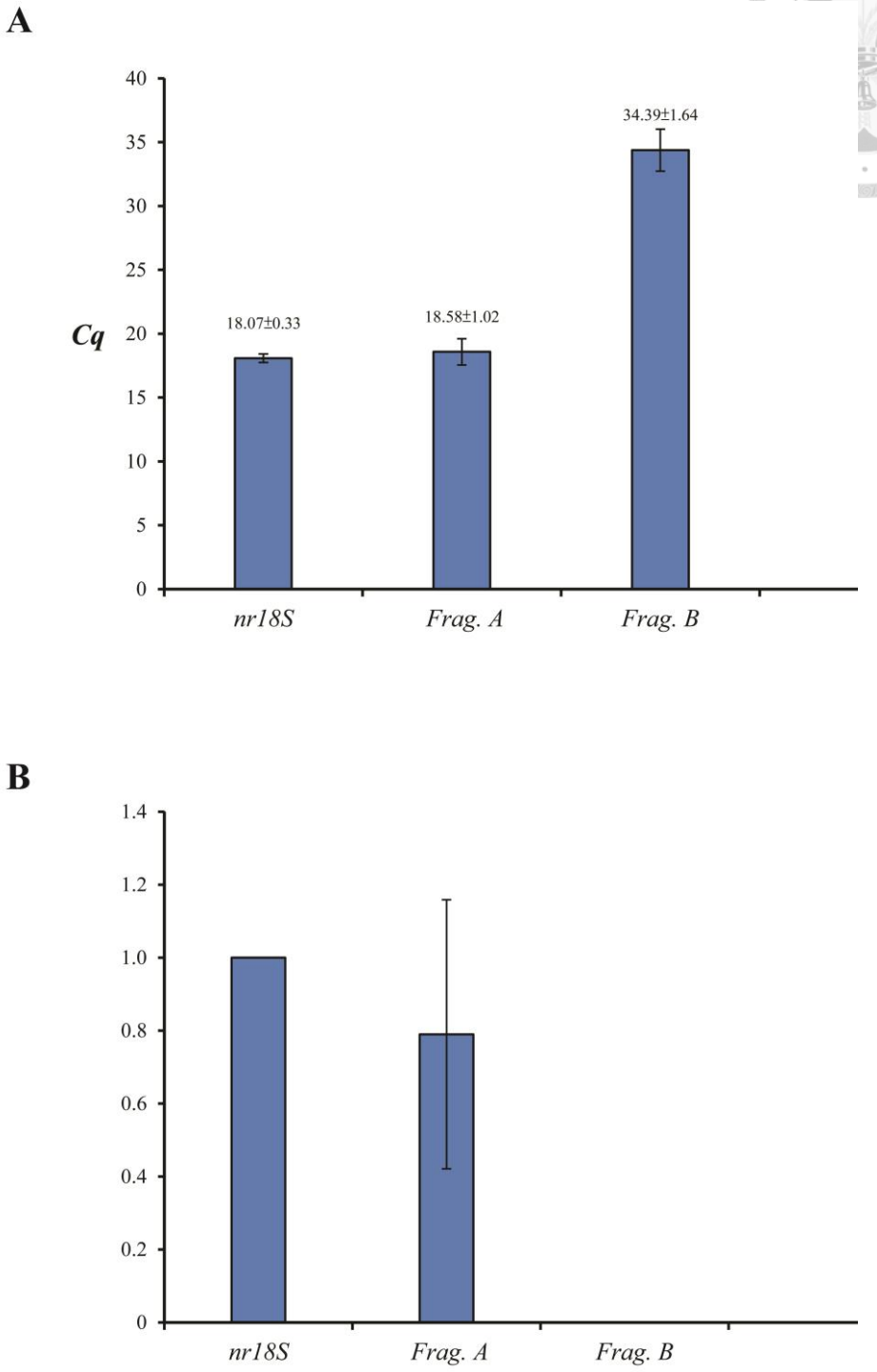


Figure 4-6. The histogram of real-time PCR amplification of three *Mitrastemon kanehirai* DNA fragments. A. The *Cq* values of three fragments. B. Relative quantification of Fragment A and B. The content of Fragment A and B was compared with nr18S rDNA content.

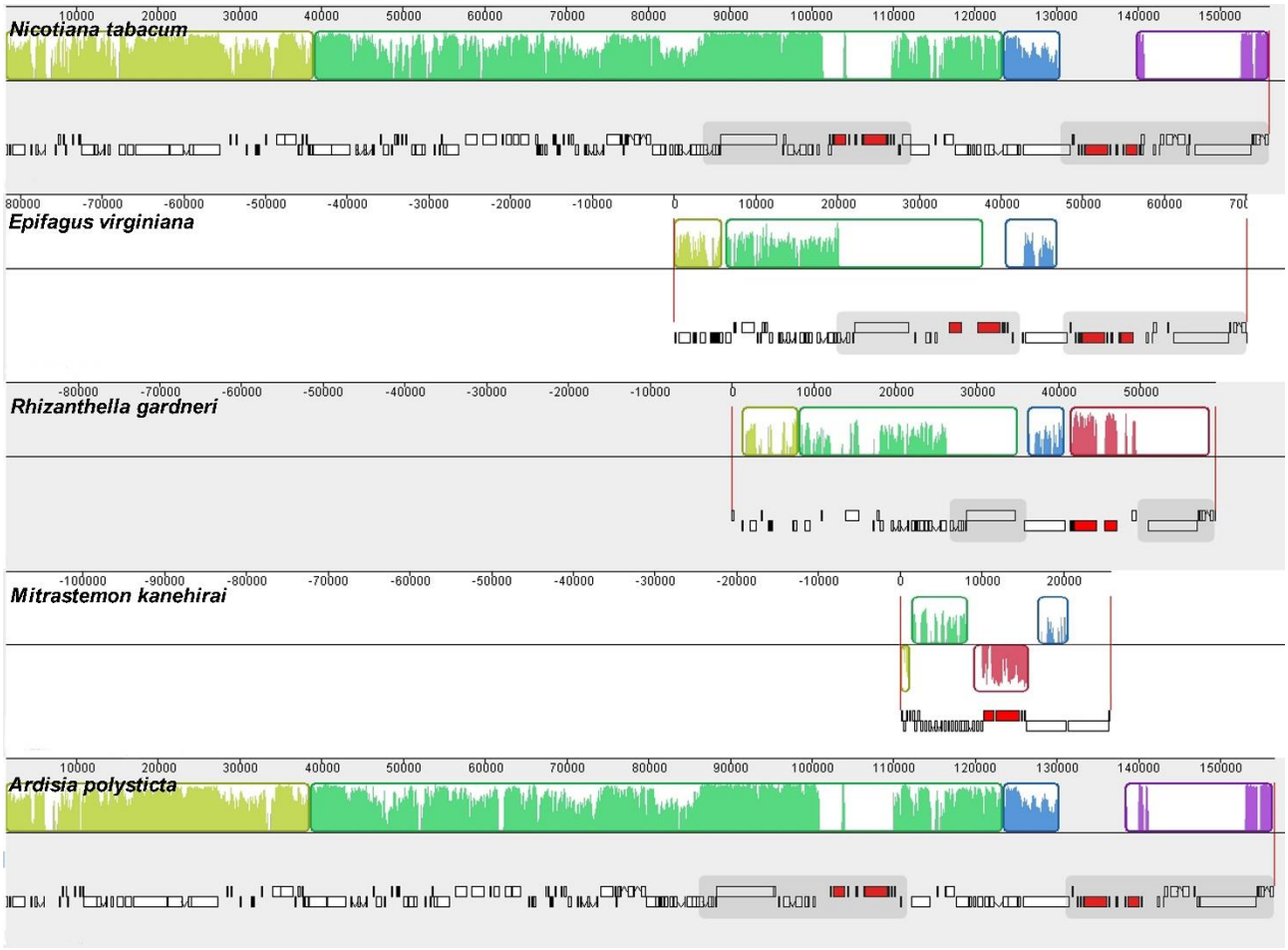


Figure 4-7. The comparison of three non-photosynthetic and two green plant plastomes. The homologous regions are depicted in the same color blocks at the upper part of each plastome, and annotated genes are shown as white boxes, rDNA as red at the lower part. The gray bars at lower part indicate the IRs.

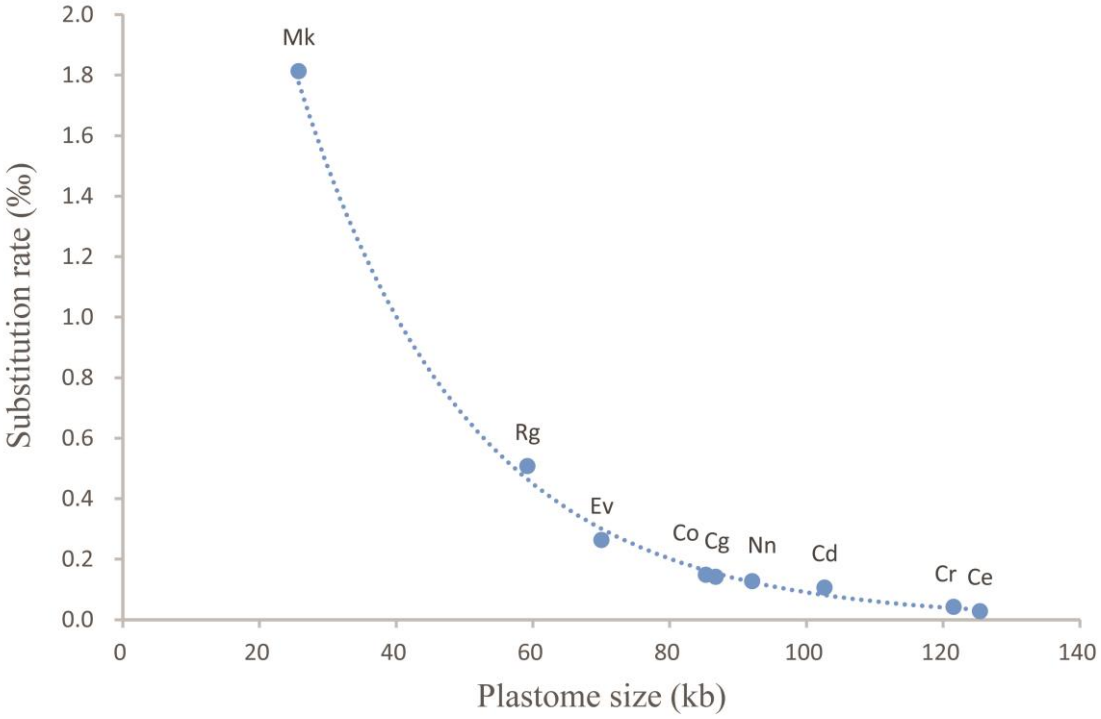
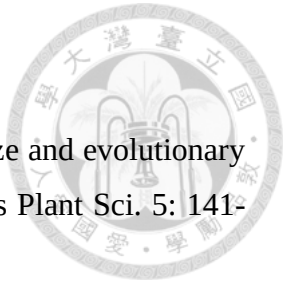


Figure 4-8. Relationship between plastome size and pt16S rDNA substitution rate among heterotrophic plants. Cd, *Cistanche deserticola*; Ce, *Cuscuta exaltata*; Cg, *Cuscuta gronovii*; Co, *Cuscuta obtusiflora*; Cr, *Cuscuta reflexa*; Ev, *Epifagus virginiana*; Mk, *Mitrastemon kanehirai*; Nn, *Neottia nidus-avis*; Rg, *Rhizanthella gardneri*.



References

Abdallah, F., F. Salamini and D. Leister. 2000. A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. Trends Plant Sci. 5: 141-142.

Adams, K.L., D.O. Daley, Y.L. Qiu, J. Whelan and J.D. Palmer. 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature. 408: 354-357.

Adams, K.L. and J.D. Palmer. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 29: 380-395.

Barbrook, A.C., C.J. Howe and S. Purton. 2006. Why are plastid genomes retained in non-photosynthetic organisms? Trends Plant Sci. 11: 101-108.

Brandvain, Y., M.S. Barker and M.J. Wade. 2007. Gene co-inheritance and gene transfer. Science. 315: 1685.

Burger, G. and B.F. Lang. 2003. Parallels in genome evolution in mitochondria and bacterial symbionts. IUBMB Life. 55: 205-212.

Butterfield, N.J. 2000. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. Paleobiology. 26: 386-404.

Cai, Z., C. Penaflor, J. Kuehl, J. Leebens-Mack, J. Carlson, C. dePamphilis, J. Boore and R. Jansen. 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. BMC Evol Biol. 6: 77.

Cavalier-Smith, T. 2000. Membrane heredity and early chloroplast evolution. Trends Plant Sci. 5: 174-182.

Colwell, A.E. 1994. Genome evolution in a non-photosynthetic plant, *Conopholis americana*. Ph.D. dissertation. Washington University, St. Louis, MO.

Darling, A.E., B. Mau and N.T. Perna. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 5: e11147.

Delannoy, E., S. Fujii, C. Colas des Francs-Small, M. Brundrett and I. Small. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. Mol Biol Evol. 28: 2077-2086.

dePamphilis, C.W., N.D. Young and A.D. Wolfe. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: many losses of photosynthesis

and complex patterns of rate variation. Proc Nat Acad Sci USA. 94: 7367-7372.

Dhingra, A. and K.M. Folta. 2005. ASAP: amplification, sequencing & annotation of plastomes. BMC Genomics. 6: 176-188.

Douzery, E.J.P., E.A. Snell, E. Bapteste, F. Delsuc and H. Philippe. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc Nat Acad Sci USA. 101: 15386-15391.

Drescher, A., S. Ruf, T. Calsa, H. Carrer and R. Bock. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. Plant J. 22: 97-104.

Fleischmann, T.T., L.B. Scharff, S. Alkatib, S. Hasdorf, M.A. Schöttler and R. Bock. 2011. Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. Plant Cell. 23: 3137-3155.

Goremykin, V.V., F. Salamini, R. Velasco and R. Viola. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol. 26: 99-110.

Howe, C.J., A.C. Barbrook, V.L. Koumandou, R.E.R. Nisbet, H.A. Symington and T.F. Wightman. 2003. Evolution of the chloroplast genome. Philos Trans R Soc B-Biol Sci. 358: 99-107.

Huang, C.Y., M.A. Ayliffe and J.N. Timmis. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature. 422: 72-76.

Iorizzo, M., D. Senalik, M. Szklarczyk, D. Grzebelus, D. Spooner and P. Simon. 2012. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 12: 61-77.

Jansen, R.K. and T.A. Ruhlman. 2012. Plastid genomes of seed plants. *In* R. Bock, and V. Knoop, (eds.). Genomics of Chloroplasts and Mitochondria. Springer Science+Business Media, Dordrecht Heidelberg New York London, pp. 103-126.

Keeling, P.J. 2004. Diversity and evolutionary history of plastids and their hosts. Am J Bot. 91: 1481-1493.

Keeling, P.J. 2010. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc B-Biol Sci. 365: 729-748.

Khakhlova, O. and R. Bock. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46: 85-94.

Kores, P.J., M. Molvray, P.H. Weston, S.D. Hopper, A.P. Brown, K.M. Cameron and M.W. Chase. 2001. A phylogenetic analysis of Diurideae (Orchidaceae) based on plastid DNA sequence data. Am J Bot. 88: 1903-1914.

Kunnimalaiyaan, M. and B.L. Nielsen. 1997. Fine mapping of replication origins (*ori*A and *ori*B) in *Nicotiana tabacum* chloroplast DNA. Nucleic Acids Res. 25: 3681-3686.

Kusumi, J. and H. Tachida. 2005. Compositional properties of green-plant plastid genomes. J Mol Evol. 60: 417-425.

Li, X., T.C. Zhang, Q. Qiao, Z.M. Ren, J.Y. Zhao, T. Yonezawa, M. Hasegawa, M.J.C. Crabbe, J.Q. Li and Y. Zhong. 2013. Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). PLoS ONE. 8.

Logacheva, M.D., M.I. Schelkunov and A.A. Penin. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. Genome Biol Evol. 3: 1296-1303.

Lohan, A.J. and K.H. Wolfe. 1998. A subset of conserved tRNA genes in plastid DNA of nongreen plants. Genetics. 150: 425-433.

Lohse, M., O. Drechsel, S. Kahlau and R. Bock. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41: W575-581.

Lynch, M. 2007. The Origins of Genome Architecture, Sinauer Associates Inc.

Martin, W. 2003. Gene transfer from organelles to the nucleus: Frequent and in big chunks. Proc Nat Acad Sci USA. 100: 8612-8614.

Martin, W. and K. Kowallik. 1999. Annotated English translation of Mereschkowsky's 1905 paper ‘Über Natur und Ursprung der Chromatophoren imPflanzenreiche’. Eur J Phycol. 34: 287-295.

Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa and D. Penny. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Nat Acad Sci USA. 99: 12246-12251.

Matuda, E. 1947. On the genus *Mitrastemon*. Bulletin of the Torrey Botanical Club. 74: 133-141.

McFadden, G.I. 2001. Primary and secondary endosymbiosis and the origin of plastids. J

Phycol. 37: 951-959.

McNeal, J.R., J.V. Kuehl, J.L. Boore, J. Leebens-Mack and C.W. dePamphilis. 2009. Parallel loss of plastid introns and their maturase in the Genus *Cuscuta*. PLoS ONE. 4: e5982.

Meijer, W. and J.F. Veldkamp. 1993. A revision of *Mitrastema* (Rafflesiaceae). Blumea. 38: 221-229.

Milligan, B. 1989. Purification of chloroplast DNA using hexadecyltrimethylammonium bromide. Plant Mol Biol Rep. 7: 144-149.

Morton, B. 1993. Chloroplast DNA codon use: evidence for selection at the *psb* A locus based on tRNA availability. J Mol Evol. 37: 273-280.

Morton, B. 1998. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. Mol Evol. 46: 449-459.

Nickrent, D.L. and R.J. Duff. 1996. Molecular studies of parasitic plants using ribosomal RNA. *In* M.T. Moreno, J.I. Cubero, D. Berner, D. Joel, L.J. Musselman, and C. Parker, (eds.). Advances in Parasitic Plant Research,. pp. 28-52.

Nickrent, D.L., R.J. Duff and D.A.M. Konings. 1997a. Structural analyses of plastid-derived 16S rRNAs in holoparasitic angiosperms. Plant Mol Biol. 34: 731-743.

Nickrent, D.L., O.Y. Yan, R.J. Duff and C.W. dePamphilis. 1997b. Do nonasterid holoparasitic flowering plants have plastid genomes? Plant Mol Biol. 34: 717-729.

Odintsova, M.S. and N.P. Yurina. 2003. Plastid genomes of higher plants and algae: structure and functions. Mol Biol. 37: 649-662.

Palmer, J.D. and C.F. Delwiche. 2000. The origin and evolution of plastids and their genomes. *In* D.E. Soltis, P.S. Soltis, and J.J. Doyle, (eds.). Molecular Systematics of Plants II: DNA sequencing. Kluwer Academic Publishers, Norwell, Massachusetts, USA, pp. 375-409.

Park, J.M., J.F. Manen and G.M. Schneeweiss. 2007. Horizontal gene transfer of a plastid gene in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). Mol Phylogenet Evol. 43: 974-985.

Race, H.L., R.G. Herrmann and W. Martin. 1999. Why have organelles retained genomes? Trends Genet. 15: 364-370.

Raubeson, L.A. and R.K. Jansen. 2005. Chloroplast genomes of plants. *In* R.J. Henry, (eds.). Plant Diversity and Evolution: genotypic and phenotypic variation in higher plants. CABI Publishing, Cambridge, Mass., pp. 45-68.

Reyes-Prieto, A., J.D. Hackett, M.B. Soares, M.F. Bonaldo and D. Bhattacharya. 2006.

Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr Biol. 16: 2320-2325.

Rujan, T. and W. Martin. 2001. How many genes in *Arabidopsis* come from cyanobacteria? An estimate from 386 protein phylogenies. Trends Genet. 17: 113-120.

Schattner, P., A.N. Brooks and T.M. Lowe. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33: W686-W689.

Sheppard, A.E., M.A. Ayliffe, L. Blatch, A. Day, S.K. Delaney, N. Khairul-Fahmy, Y. Li, P. Madesis, A.J. Pryor and J.N. Timmis. 2008. Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. Plant Physiol. 148: 328-336.

Sheppard, A.E. and J.N. Timmis. 2009. Instability of plastid DNA in the nuclear genome. PLoS Genet. 5: e1000323.

Smith, D. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. Plant Mol Biol. 71: 627-639.

Smith, D.R., F. Burki, T. Yamada, J. Grimwood, I.V. Grigoriev, J.L. Van Etten and P.J. Keeling. 2011. The GC-rich mitochondrial and plastid genomes of the green alga *Coccomyxa* give insight into the evolution of organelle DNA nucleotide landscape. PLoS ONE. 6: e23624.

Stegemann, S., S. Hartmann, S. Ruf and R. Bock. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. Proc Nat Acad Sci USA. 100: 8828-8833.

Stegemann, S., M. Keuthe, S. Greiner and R. Bock. 2012. Horizontal transfer of chloroplast genomes between plant species. Proc Nat Acad Sci USA. 109: 2434-2438.

Timmis, J.N., M.A. Ayliffe, C.Y. Huang and W. Martin. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5: 123-135.

Wang, D., A.H. Lloyd and J.N. Timmis. 2012. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. Proc Nat Acad Sci USA. 109: 2444-2448.

Waters, M.T. and J.A. Langdale. 2009. The making of a chloroplast. EMBO J. 28: 2861-2873.

Wicke, S., G. Schneeweiss, C. dePamphilis, K. Müller and D. Quandt. 2011. The

evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76: 273-297.

Wolfe, K.H., C.W. Morden and J.D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Nat Acad Sci USA. 89: 10648-10652.

Wu, S., Z. Zhu, L. Fu, B. Niu and W. Li. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 12: 444.

Xi, Z., R. Bradley, K. Wurdack, K. Wong, M. Sugumaran, K. Bomblies, J. Rest and C. Davis. 2012. Horizontal transfer of expressed genes in a parasitic flowering plant. BMC Genomics. 13: 227.

Xi, Z., Y. Wang, R.K. Bradley, M. Sugumaran, C.J. Marx, J.S. Rest and C.C. Davis. 2013. Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genet. 9: e1003265.

Xiong, A.S., R.H. Peng, J. Zhuang, F. Gao, B. Zhu, X.Y. Fu, Y. Xue, X.F. Jin, Y.S. Tian, W. Zhao and Q.H. Yao. 2009. Gene duplication, transfer, and evolution in the chloroplast genome. Biotechnol Adv. 27: 340-347.

Yang, Y.P. and S.Y. Lu. 1996. Rafflesiaceae. *In* Editorial Committee of the Flora of Taiwan, (eds.). Flora of Taiwan, second edition. Editorial Committee of the Flora of Taiwan, Taipei, Taiwan, pp. 652-655.

Yoon, H.S., J.D. Hackett, C. Ciniglia, G. Pinto and D. Bhattacharya. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol. 21: 809-818.

Supplementary Data

Table S4-1. Primers list.

Primer (F)	Sequence from 5' to 3'	Primer (R)	Sequence from 5' to 3'
Real-time quantitative PCR & Southern blots[a]			
SSU1594F	CTACGTCCCTGCCCTTTGTA	SSU1703R	GGACTTCTCGCGGCATCACGAG
ProbeA-F	GGAAACAGCCCAGATCATCA	ProbeA-R	GCCGACATTCTCACTTCTGC
ProbeB-F	TGCCATGGTAAGGAAGAAGG	ProbeB-R	TGCATGAAGGAGTCAGATGC
Reverse transcription-PCR			
16S8F	GGAGAGTTCGATCCTGGCTCAG	16S1462R	GGTTGATCCAGCTACACCTTCCAG
infA-F	TTTTAGGTTATGTTTCTGGGAAAA	infA-R	TCCTTTGTCGGGATCGTATT
rps3-90F	TCAAAATTTTCAAGAAGATCAAAAA	rps3-554R	ACTTGTCCCTTTCGAATCCA
rps7-22F	AAAAACAAAATTATAAAACCAGATCC	rps7-399R	AGCTTCACCCCTTCCCTTAG
rps14-52F	TGCAAATATAAAAATATTCGTAAGTCA	rps14-296R	TTTGGTTTTGTCACTCCTGGT
clpP-F	TGAAATTTCTAATCAACTTACTGGTC	clpP-R	TTCCATGTCTTCAGATAAAATCCA
accD-F	GCATGCCATGGATGAAAATC	accD-R	GCTCGATAACCCTTTTACCG
ycf1-F	TTTCTTTATTTTGCGGATCACA	ycf1-R	CAAATTTTGTGAATTTTCTTCTTCTG
ycf2-F	TTTTCCGGATTTATAAGTCATACAAG	ycf2-R	AAAGATCTTGTGTATGGAAAACCA
Long PCR[b]			
16S34F	CTGGCGGCTGTATGCTTAAC	5S58R	TAACCACCAAGTTCGGGATG
16S977F	AGTGCCTTAGGGAACGCGAACACAG	IRB27R	CCAATGCTAGATGCAGAGGCGCATA
IRB2F	CATCTGGCTTATGTTCTTCATGTAGC	IRB6R	TGAATATGTTAGATACCTGTGACTCG
IRB6F	TTCTCGAACCGAGAGATCCA	IRB12R	ATGCTTGCGTATTCGTCCAT
IRB12F	GGTCTGTCCCGGTATGGAAT	16S1307R	ACCGGCGATTACTAGCGATT
Plastome sequence confirmation			
Mit1F	CCATTTCGAACTTGGTGGTT	Mit1R	AACCTTGTAATTTTGATTTATCTTACG
Mit2F	TTATAAAATACGTATGCTTGTGTTACG	Mit2R	AATAGCCGAAGCTTGGAATG
Mit3F	CAAATTTTGTGAATTTTCTTCTTCTG	Mit3R	TTTCTTTATTTTGCGGATCACA
Mit4F	TTAAATTGAAAGTTTCTTGTTCTTGA	Mit4R	AAAATGGAAAAATTTGACACATAAA
Mit5F	GGAAAAGGCATTCTACCTAAATTAAA	Mit5R	ATCCCGCGGATCCTTTTT
Mit6F	TTTGGGTGTATATTTTTGTTGATAGG	Mit6R	TTGATCCTATATTAATAGCTCCAACA
Mit7F	AAAGATCTTGTGTATGGAAAACCA	Mit7R	TTTTCCGGATTTATAAGTCATACAAG
Mit8F	ATTCCGGGTTGTTCTATATTTTT	Mit8R	CCAGAAATGGGATTATTTTCTGA
Mit9F	TCCGCTAAAGGTAAGAAACCA	Mit9R	AAAATTTTATCATATCCCGAAAAA
Mit10F	AAGGGGTTATCAAACGTTATTTTT	Mit10R	GTTATGAGCCTTGCGAGCTT
Mit11F	ATGCGCAGGTTCAATTCCTA	Mit11R	TGGTTAACATACCACCAATCCA
Mit12F	GCTTGTCATTTAGTTTTTGATGC	Mit12R	TGGATTATTGCCAGGTCTCA
Mit13F	AACCCGTCGAACCTTTAAAAT	Mit13R	TCCCATACTACCGCCCATAA
Mit14F	GCATGCCATGGATGAAAATC	Mit14R	CATTTAGAAATCTCCGGTTATAGAAAA
Mit15F	TTCAGGCTTTAGCTACTATAACTTTCC	Mit15R	GTTTTGCATTCCCTCATGCT
Mit16F	AAATTCCCCAGCCTGTGTTT	Mit16R	GAAGGGGAAGAACCCTCTTG
Mit17F	CCAGGTATTAAAAGCCCTCCA	Mit17R	AATACGATCCCGACAAAGGA
Mit18F	TCGCCGAGGTGAAGTTTTTA	Mit18R	GAATGGATTCGAAAGGGACA
Mit19F	TTGATATTCCAGATATTCTGCCTCT	Mit19R	CCGCAGGTACCTTAGCAAAA
Mit20F	GGCCGTTTTCCTAACCATCT	Mit20R	AGCTAAGGGAAGGGGTGAAG
Mit21F	GCCAGAGTTCTACTTGATTCTGC	Mit21R	CCCCCGTACTTAGGGGTAGA

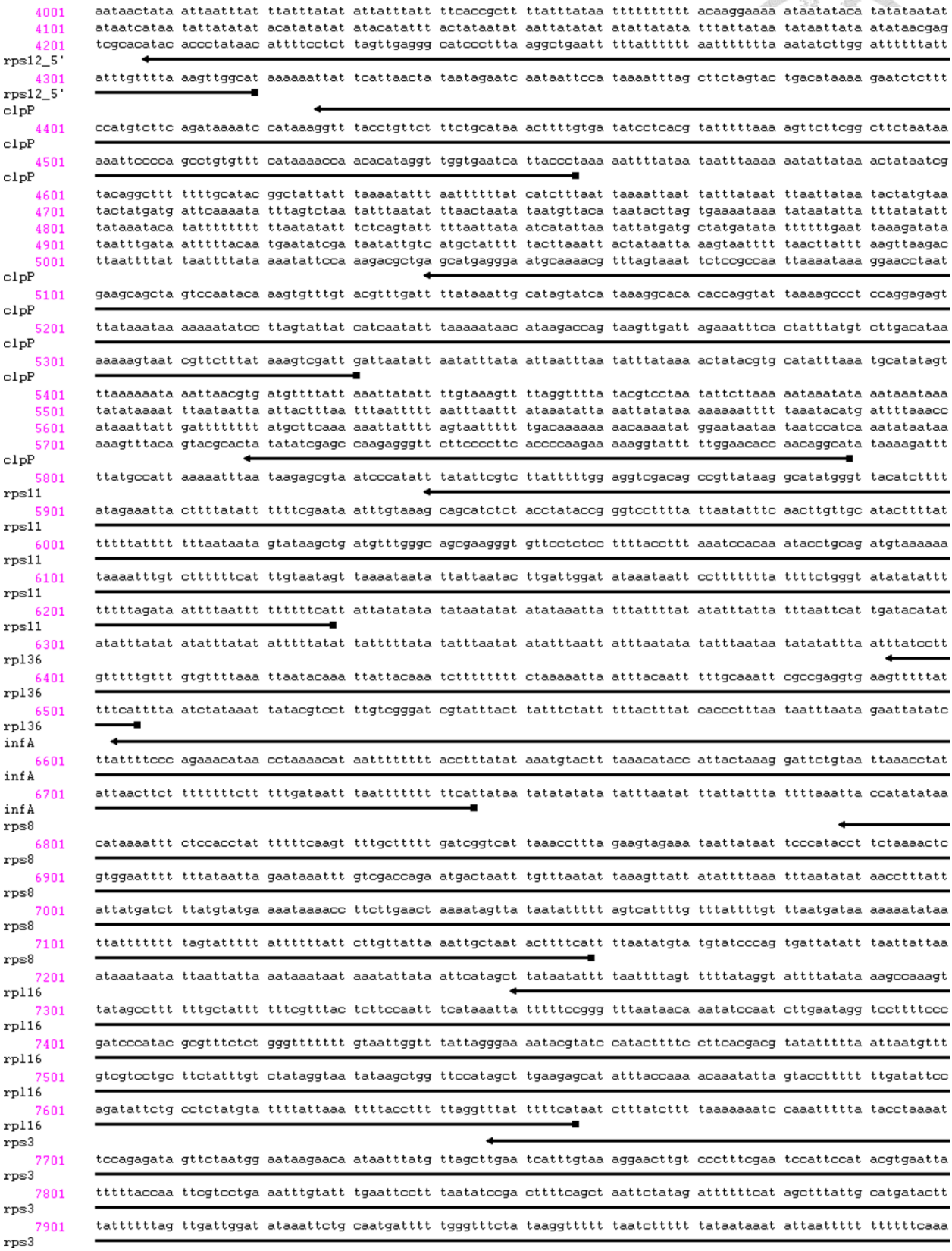
[a] Primers were used for RT qPCR and probe syntheses of Southern blot. SSU1594F and SSU1703R were primers for amplifying nr18S rDNA fragment.

[b] Primers were used in order for obtaining Frag. A, B, C, D and E (Table S4-2).

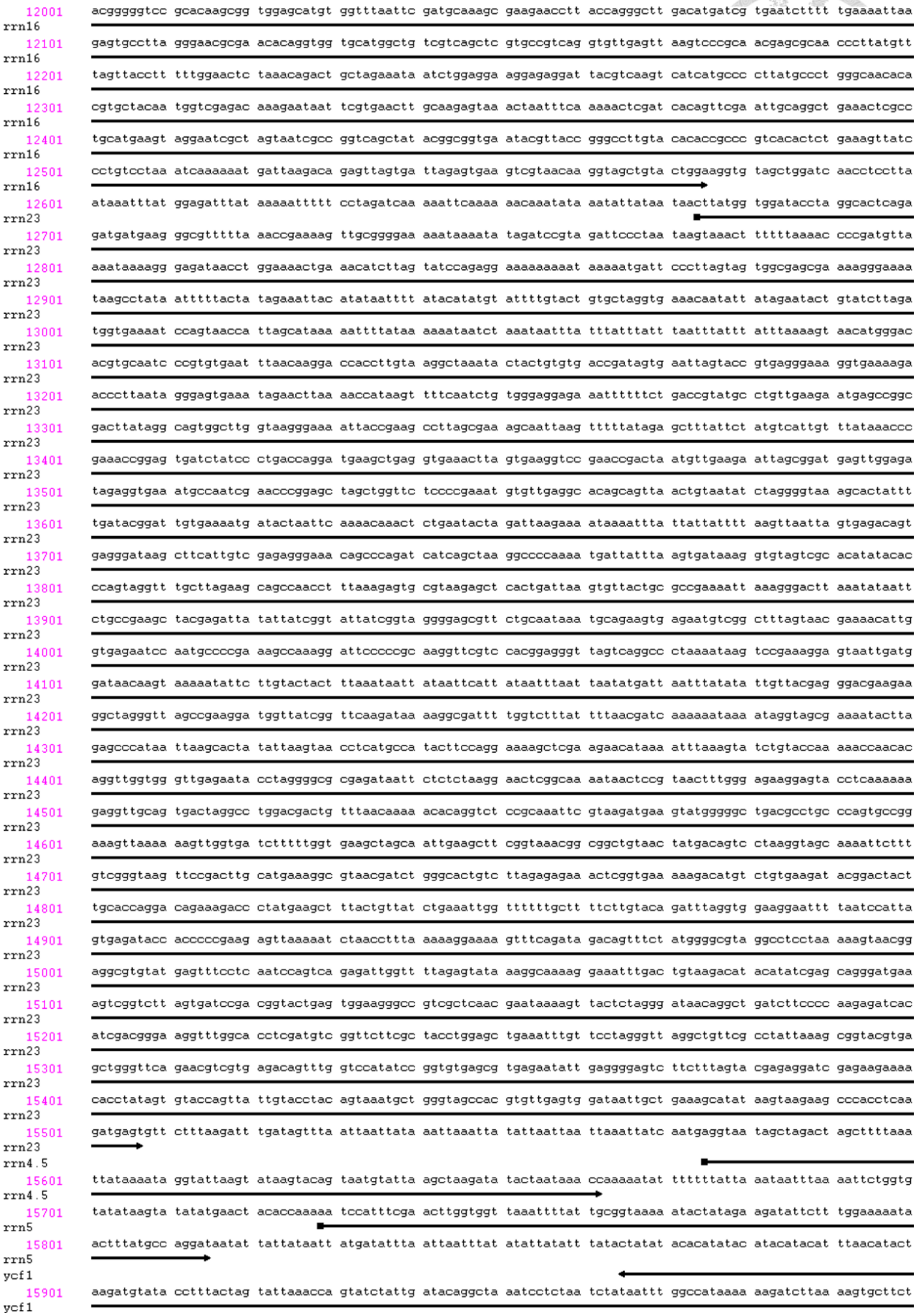
Table S4-2. Sequence information of large fragments obtained in this study.

	Size (bp)	Sequence information
Obtained from Long PCR		
Fragment A	4,585	Contained *Mitrastemon* 16S sequence
Fragment B	8,019	Highly similar to IRs of Ericales species
Fragment C	5,603	Contained a short ptDNA fragment
Fragment D	6,227	Highly similar to IRs of green plants
Fragment E	7,675	Highly similar to IRs of Ericales species
Obtained from NGS		
Contig A	15,116	Contained Fragment A
Contig B	16,032	Contained Fragment A
Contig C	11,882	Contained a short ptDNA fragment
Contig D	18,673	Contained a short Fagales-like ptDNA fragment
Assembly A	25,740	The *M. kanehirai* plastome, assembled from Contig A and B
Assembly B	15,448	Assembled from Frag. B and E
Assembly C	17,050	Assembled from Frag. C and Contig C

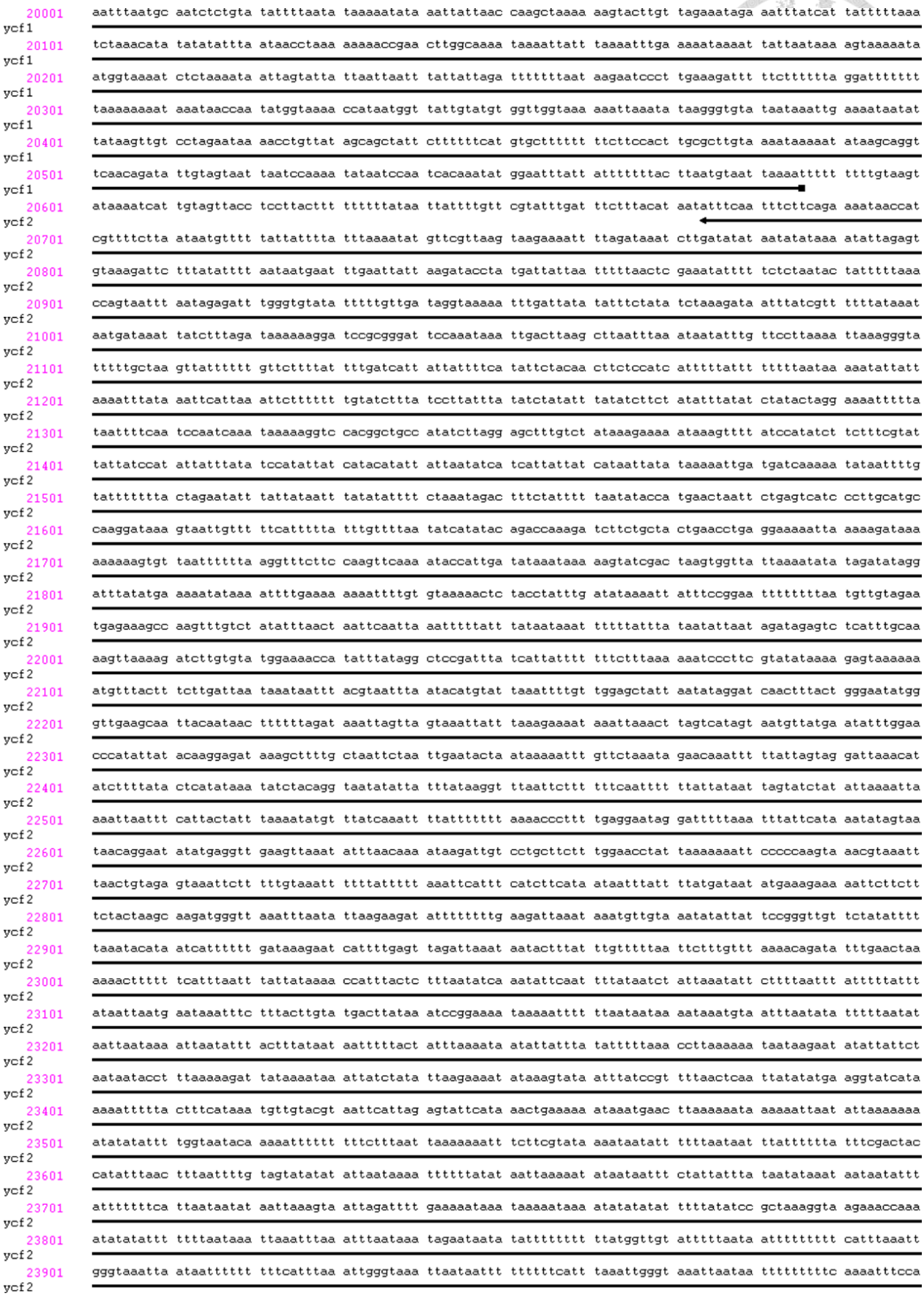
Figure S4-1. Complete plastome sequence of *Mitrastemon kanehirai*.

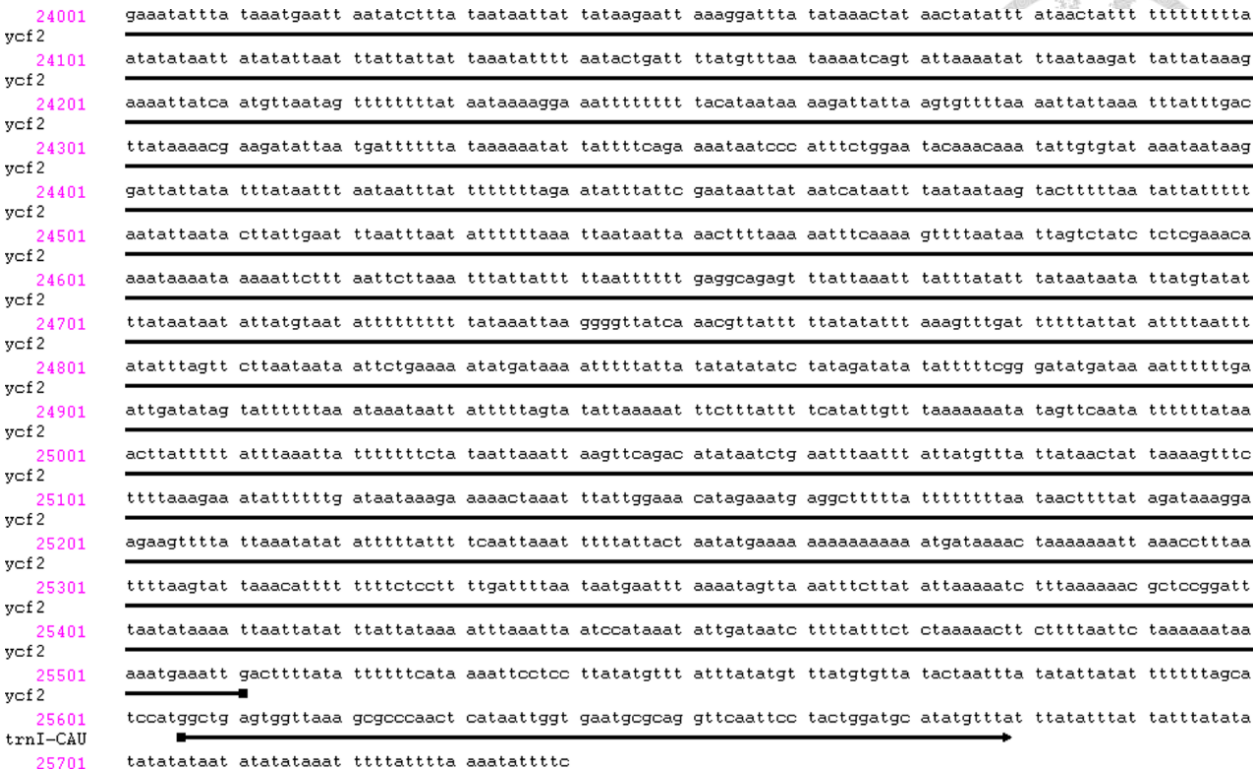

4001 aataactata attaatttat ttatttatat attatttatt ttcaccgctt ttatttataa ttttttttt acaaggaaaa ataatataca tatataatat
4101 ataatcataa tattatatat acatatatat atacatattt actataatat aattatatat atattatata tttattataa tataattata atataacgag
4201 tcgcacatac accctataac attttcctct tagttgaggg catcccttta aggctgaatt tttatttttt aatttttta aatatcttgg atttttatt
rps12_5'
4301 atttgtttta aagttggcat aaaaaattat tcattaacta taatagaatc aataattcca taaaatttag cttctagtac tgacataaaa gaatctcttt
rps12_5'
clpP
4401 ccatgtcttc agataaaatc cataaaggtt tacctgttct ttctgcataa acttttgtga tatcctcacg tatttttaaa agttcttcgg cttctaataa
clpP
4501 aaattcccca gcctgtgttt cataaaacca acacataggt tggtgaatca ttaccctaaa aattttataa taatttaaaa aatattataa actataatcg
clpP
4601 tacaggcttt ttttgcatac ggctattatt taaaatattt aatttttat catctttaat taaaattaat tatttataat ttaattataa tactatgtaa
4701 tactatgatg attcaaaata tttagtctaa tatttaatat ttaactaata taatgttaca taatacttag tgaaaataaa tataatatta tttatatatt
4801 tataaatata tatttttttt ttaatatatt tctcagtatt tttaattata atcatattaa tattatgatg ctatgatata ttttttgaat taaagatata
4901 taatttgata atttttacaa tgaatatcga taatattgtc atgctatttt tacttaaatt actataatta aagtaatttt taacttattt aagttaagac
5001 ttaattttat taattttata aaatattcca aagacgctga gcatgaggga atgcaaaacg tttagtaaat tctccgccaa ttaaaataaa ggaacctaat
clpP
5101 gaagcagcta gtccaataca aagtgtttgt acgtttgatt ttataaattg catagtatca taaaggcaca caccaggtat taaaagccct ccaggagagt
clpP
5201 ttataaataa aaaaatatcc ttagtattat catcaatatt taaaaataac ataagaccag taagttgatt agaaatttca ctatttatgt cttgacataa
clpP
5301 aaaaagtaat cgttctttat aaagtcgatt gattaatatt aatatttata attaatttaa tatttataaa actatacgtg catatttaaa tgcatatagt
clpP
5401 ttaaaaaata aattaacgtg atgttttatt aaattatatt ttgtaaagtt ttaggtttta tacgtcctaa tattcttaaa aataaatata aataaataaa
5501 tatataaaat ttaataatta attactttaa tttaattttt aatttaattt ataaatatta aattatataa aaaaaatttt taaatacatg attttaaacc
5601 ataaattatt gatttttttt atgcttcaaa aaattatttt agtaattttt tgacaaaaaa aacaaaatat ggaataataa taatccatca aatataataa
5701 aaagtttaca gtacgcacta tatatcgagc caagagggtt cttccccttc accccaagaa aaaggtattt ttggaacacc aacaggcata taaaagattt
clpP
5801 ttatgccatt aaaaatttaa taagagcgta atcccatatt tatattcgtc ttatttttgg aggtcgacag ccgttataag gcatatgggt tacatctttt
rps11
5901 atagaaatta cttttatatt ttttcgaata atttgtaaag cagcatctct acctataccg ggtcctttta ttaatatttc aacttgttgc atacttttat
rps11
6001 tttttatttt tttaataata gtataagctg atgtttgggc agcgaagggt gttcctctcc ttttaccttt aaatccacaa atacctgcag atgtaaaaaa
rps11
6101 taaaatttgt ctttttttcat ttgtaatagt taaaataata ttattaatac ttgattggat ataaataatt ccttttttta ttttctgggt atatatattt
rps11
6201 tttttagata attttaattt ttttttcatt attatatata tataatatat atataaatta tttatttttat atatttatta tttaattcat tgatacatat
rps11
6301 atatttatat atatttatat atttttatat tatttttata tatttaatat atatttaatt atttaatata tatttaataa tatatattta atttatcctt
rpl36
6401 gtttttgttt gtgttttaaa ttaatacaaa ttattacaaa tctttttttt ctaaaaatta atttacaatt tttgcaaatt cgccgaggtg aagtttttat
rpl36
6501 tttcatttta atctataaat tatacgtcct ttgtcgggat cgtatttact tatttctatt tttactttat caccctttaa taatttaata gaattatatc
rpl36
infA
6601 ttattttccc agaaacataa cctaaaacat aatttttttt acctttatat aaatgtactt taaacatacc attactaaag gattctgtaa ttaaacctat
infA
6701 attaacttct ttttttttctt tttgataatt taatttttttt ttcattataa tatatatata tatttaatat ttattattta ttttaaatta ccatatataa
infA
rps8
6801 cataaaattt ctccacctat tttttcaagt tttgcttttt gatcggtcat taaaccttta gaagtagaaa taattataat tcccatacct tctaaaactc
rps8
6901 gtggaatttt tttataatta gaataaattt gtcgaccaga atgactaatt tgtttaatat taaagttatt atattttaaa tttaatatat aacctttatt
rps8
7001 attatgatct ttatgtatga aaataaaacc ttcttgaact aaaatagtta taatattttt agtcattttg tttattttgt ttaatgataa aaaaatataa
rps8
7101 ttatttttttt tagtattttt atttttttatt cttgttatta aattgctaat acttttcatt ttaatatgta tgtatcccag tgattatatt taattattaa
rps8
7201 ataaataata ttaattatta aataaataat aaatattata attcatagct tataatattt taattttagt ttttataggt attttatata aagccaaagt
rpl16
7301 tatagccttt tttgctattt tttcgtttac tcttccaatt tcataaatta tttttccggg tttaataaca aatatccaat cttgaatagg tccttttccc
rpl16
7401 gatcccatac gcgtttctct gggtttttttt gtaattggtt tattagggaa aatacgtatc catactttttc cttcacgacg tatattttta attaatgttt
rpl16
7501 gtcgtcctgc ttctatttgt ctataggtaa tataagctgg ttccatagct tgaagagcat atttaccaaa acaaatatta gtacctttttt ttgatattcc
rpl16
7601 agatattctg cctctatgta ttttattaaa ttttaccttt ttaggtttat ttttcataat ctttatcttt taaaaaaatc caaattttta tacctaaaat
rpl16
rps3
7701 tccagagata gttctaatgg aataagaaca ataatttatg ttagcttgaa tcatttgtaa aggaacttgt ccctttcgaa tccattccat acgtgaatta
rps3
7801 tttttaccaa ttcgtcctga aatttgtatt tgaattcctt taatatccga cttttcagct aattctatag atttttttcat agctttattg catgatactt
rps3
7901 tatttttttag ttgattggat ataaattctg caatgatttt tgggtttcta taaggttttt taatcttttt tataataaat attaattttt ttttttcaaa
rps3


8001 agttaataat aaaattaatc ttttttttaa ttcctgtggt gaatttttt ttaataattt agtaaaacct ataaataata ttatttaat aaaattaatt
rps3
8101 ttttttttaa tttttatacg tgtaattcct tcaattaggt taatatcata agaaatttta taaatttgtt gtttactaaa aaaaaaattt ataaaatttc
rps3
8201 ttatttttg atcttcttga aaattttgaa aataattttt cttcttagag aaccaaaaag aatcataatt tattgttata ccaagtctat agattaatgg
rps3
8301 attaattttt tgtcccataa aatttttat gactagattt attatcacct tttaattttg tatgccctct aaaatttata gtaggtgaaa attctcctaa
rps3
rps19
8401 tttgtgacct attaaataac tagttataaa aagaggtata tgctctctac cattatatac agcaatagtt tgtccgatca tcgtaggtat aatattcgat
rps19
8501 gatcgagacc aagtttttat gaatttttt ttttttttta agagtttatg tgaaacaaaa ggatttttt ttaaagaacg tgccataact taatattatt
rps19
8601 ttatattatt aatctttgct attaattata tttatttttt aatgcgtcta tgaataataa aattattact atatttttt ttttttctgc ttttttccc
rpl2
8701 caacgtggga aaaccccagg gagttacggg ttttttttcta cctattgaag atttaccttc accgccccca tgtggatgat ccacagaatt catagcgagt
rpl2
8801 cctcttactt ttggccgttt tcctaaccat cttttagatc ctgctttaaa aaaatgttta ttttttttt ctatgtttcc tacttgacct atagttgctg
rpl2
8901 tacagctttt aaatataaac ctaaaatttc ctgaaggtaa ttttaccata acataattat tatttttga tattaatttt gctaaggtac ctgcggatct
rpl2
9001 aactaattga ccgccctttc cgagtgttat ttctatatta tgaataaatg ttcctaaagg aatattggat gaaatagatt taaaaaatct tttcccaaac
rpl2
9101 tgtacaagct tttttacaaa acatacagct ttccagatat atttcatata taaatatatt gtataataaa ttataataat taacatccta aaattattgt
9201 ttcatcatca tctggcttaa tagcatacag attgaatgac tccatggaat cacgctaaat gataacttag gggatatttt taatatattt tatataagtt
9301 tttaaattgc ctttgaccat caaattaaaa ttatttatat attttattgt taatttcccc atattatagt tattatagta atattatagt attaaataaa
9401 taaaataaat attaaattaa tcacttgctt atatatttat atatttttat ttttaaatta agtgataatt ttataagttt tatcgtaaac ttaattggta
9501 taaatttcca attacgtgta aataacatag ttcaaaccgc actcaaaggt aaaacatttc ctatttttaa agaaacttct gtaccagata taataatatc
rpl2
9601 tcctattgta aaacctctag tatgtaaaat atatctttta tcaccatctt cataatatac tagacaaata tatgaatttc tattagggtc atattcaata
rpl2
9701 gttttaattt taccatataa atttatattt tctcgtttaa aatctatttt acgatataaa cgtttatggc ccccacctct atgtcttgta gaaattgttc
rpl2
9801 ctttattatt tcgaccatta ctttttccat actgtccaga agttaaattc ttatatggat ttaaagtttt taaattacta atcataaaat aatcataaaa
rpl2
9901 tttttaatgt attgaatttg atagtattta aattataaga ttatttgaaa ttagccagag ttctacttga ttctgcaatt ttatgaatta attctttttt
rps7
10001 tcgtatagct tcacccttc ccttagctgc atctattaat tccgaactaa attttaaaga catatttttt ccaaatcttt tttttgatgc agttaataac
rps7
10101 catcgtatag caagtatttt tccttttgta gattttattt caataggaat tttaaaaatt gatccacgtt tatgttttgt ttgtactgat aacttggggg
rps7
10201 ttactctatt tattgcctta tttaaaataa aaagtggttt ttttttttt ttacaaatat tatttagggt attatatata ataaaataag ctaatgattt
rps7
10301 ttttccattt ttcataatat gattaattaa tatattaatg gatttattat aataaatagg atctggtttt ataattttgt tttttaaaag tctatgtcgt
rps7
10401 gacataataa ttaataattt taaaaagatt ttatattaaa ttttttaact ccatacttag aacgactctt tttacgatct tgtactccga tagcatcaag
rps7
rps12_3'
10501 ggctcctctt ataacgtgat atcttacgcc tggtaaatct ttaaccctcc ctccccttat taaaactaca gaatgttctt ttaaagtatg tcctatgccg
rps12_3'
10601 ggtatataag cagtaacttc aaatccagag gttaatttta ctctaacaac tttacgtaaa gcagaatttg gttttttagg ggtaatagtg gaaaagttga
rps12_3'
10701 ctaacaatat ttatataaaa tattataatt aaatttttata ttttgtaagc cacttctaca gaaacgtaca tgaaattttc atttcatacg gctcctcagc
10801 ttaatcatat ccatagatta tttagtattt taatatatat aaatataaat tactaaatta ctaaatgtag tattaattta aattattaaa taattaaata
10901 ataatataat ataaatattt atgaataatt gagttaacat aagcttatcc tgataattat atatataaat tatatatata aataagatta tatataatat
11001 acatattaaa tattatatta atatattata tattattata tattattatt atatattatt ttgatgcttt tatgctttaa aaaaaattta aaatttaacg
11101 tagctataac ataatataac tataattact aaggagagtt tgatcctagc tcaggatgaa cgctggcggt atgcttaaca catgcaagtc ttactataaa
rrn16
11201 tttagtggcg aacgggtgag tgatgcgtaa gaatctaccc ctaagtacgg gggagaaaac agtgggaaac tattgctaat atcccatata ggctgaaaag
rrn16
11301 ttaaaagaat aataattctc tacagggaag agcttacgtc tgattagcta gttggtaaga taatagctta ccaaggctgt gatcgggagc tagttcgaga
rrn16
11401 ggatgattag ccacattgga actgagacaa ggtccagact tctatgggag gcagcagtgg ggaatttcc gcaatgagcg aaagcttgac ggggcaatac
rrn16
11501 cgcgtgaagg tagaaggcct acgggtcgta aacttctttt accaaaaaag aaaactaaca gtatttgggg aataagtatc ggctaacttt gtgccagcag
rrn16
11601 ccgcggtaat acaaaggatg caagcgttat ccggaatgat tgggcgtaaa gagtctgtag gtaacttttt aagtctgacg tcaaatacct aggctcaacc
rrn16
11701 gaggaaaggc atcagaaact aaaaagcttg agtatattaa agggaagagg gaatttctga aggagcgatg aaatgcatag atatcagaaa gaacgccgaa
rrn16
11801 ggtgaaaacg ctcttctgta ttgatactga cattgagaga cgaaagctag ggtagcaaat gggattagag accccagtag tcctagctgt aaacgatgga
rrn16
11901 tattaagtat tgtgtgtatt gatccattca gtactacaac taacgtgtta aatatcccgc ctggggagta cgttcgcaag aatgaaactc aaaggaattg
rrn16


12001 acgggggtcc gcacaagcgg tggagcatgt ggtttaattc gatgcaaagc gaagaacctt accagggctt gacatgatcg tgaatctttt tgaaaattaa
rrn16
12101 gagtgcctta gggaacgcga acacaggtgg tgcatggctg tcgtcagctc gtgccgtcag gtgttgagtt aagtcccgca acgagcgcaa cccttatgtt
rrn16
12201 tagttacctt tttggaactc taaacagact gctagaaata atctggagga aggagaggat tacgtcaagt catcatgccc cttatgccct gggcaacaca
rrn16
12301 cgtgctacaa tggtcgagac aaagaataat tcgtgaactt gcaagagtaa actaatttca aaaactcgat cacagttcga attgcaggct gaaactcgcc
rrn16
12401 tgcatgaagt aggaatcgct agtaatcgcc ggtcagctat acggcggtga atacgttacc gggccttgta cacaccgccc gtcacactct gaaagttatc
rrn16
12501 cctgtcctaa atcaaaaaat gattaagaca gagttagtga ttagagtgaa gtcgtaacaa ggtagctgta ctggaaggtg tagctggatc aacctcctta
rrn16
12601 ataaatttat ggagatttat aaaaattttt cctagatcaa aaattcaaaa aacaaatata aatattataa taacttatgg tggataccta ggcactcaga
rrn23
12701 gatgatgaag ggcgttttta aaccgaaaag ttgcggggaa aaataaaata tagatccgta gattccctaa taagtaaact tttttaaaac cccgatgtta
rrn23
12801 aaataaaagg gagataacct ggaaaactga aacatcttag tatccagagg aaaaaaaaat aaaaatgatt cccttagtag tggcgagcga aaagggaaaa
rrn23
12901 taagcctata atttttacta tagaaattac atataatttt atacatatgt attttgtact gtgctaggtg aaacaatatt atagaatact gtatcttaga
rrn23
13001 tggtgaaaat ccagtaacca ttagcataaa aattttataa aaaataatct aaataattta tttatttatt taatttattt atttaaaagt aacatgggac
rrn23
13101 acgtgcaatc ccgtgtgaat ttaacaagga ccaccttgta aggctaaata ctactgtgtg accgatagtg aattagtacc gtgagggaaa ggtgaaaaga
rrn23
13201 acccttaata gggagtgaaa tagaacttaa aaccataagt tttcaatctg tgggaggaga aatttttttct gaccgtatgc ctgttgaaga atgagccggc
rrn23
13301 gacttatagg cagtggcttg gtaagggaaa attaccgaag ccttagcgaa agcaattaag tttttataga gctttattct atgtcattgt ttataaaccc
rrn23
13401 gaaaccggag tgatctatcc ctgaccagga tgaagctgag gtgaaactta gtgaaggtcc gaaccgacta atgttgaaga attagcggat gagttggaga
rrn23
13501 tagaggtgaa atgccaatcg aacccggagc tagctggttc tccccgaaat gtgttgaggc acagcagtta actgtaatat ctaggggtaa agcactattt
rrn23
13601 tgatacggat tgtgaaaatg atactaattc aaaacaaact ctgaatacta gattaagaaa ataaaattta ttattatttt aagttaatta gtgagacagt
rrn23
13701 gagggataag cttcattgtc gagagggaaa cagcccagat catcagctaa ggccccaaaa tgattattta agtgataaag gtgtagtcgc acatatacac
rrn23
13801 ccagtaggtt tgcttagaag cagccaacct ttaaagagtg cgtaagagct cactgattaa gtgttactgc gccgaaaatt aaagggactt aaatataatt
rrn23
13901 ctgccgaagc tacgagatta tattatcggt attatcggta ggggagcgtt ctgcaataaa tgcagaagtg agaatgtcgg ctttagtaac gaaaacattg
rrn23
14001 gtgagaatcc aatgccccga aagccaaagg attcccccgc aaggttcgtc cacggagggt tagtcaggcc ctaaaataag tccgaaagga gtaattgatg
rrn23
14101 gataacaagt aaaaatattc ttgtactact ttaaataatt ataattcatt ataatttaat taatatgatt aatttatata ttgttacgag ggacgaagaa
rrn23
14201 ggctagggtt agccgaagga tggttatcgg ttcaagataa aaggcgattt tggtctttat tttaacgatc aaaaaataaa ataggtagcg aaaatactta
rrn23
14301 gagcccataa ttaagcacta tattaagtaa cctcatgcca tacttccagg aaaagctcga agaacataaa atttaaagta tctgtaccaa aaaccaacac
rrn23
14401 aggttggtgg gttgagaata cctaggggcg cgagataatt ctctctaagg aactcggcaa aataactccg taactttggg agaaggagta cctcaaaaaa
rrn23
14501 gaggttgcag tgactaggcc tggacgactg tttaacaaaa acacaggtct ccgcaaattc gtaagatgaa gtatgggggc tgacgcctgc ccagtgccgg
rrn23
14601 aaagttaaaa aagttggtga tctttttggt gaagctagca attgaagctt cggtaaacgg cggctgtaac tatgacagtc ctaaggtagc aaaattcttt
rrn23
14701 gtcgggtaag ttccgacttg catgaaaggc gtaacgatct gggcactgtc ttagagagaa actcggtgaa aaagacatgt ctgtgaagat acggactact
rrn23
14801 tgcaccagga cagaaagacc ctatgaagct ttactgttat ctgaaattgg ttttttgctt ttcttgtaca gatttaggtg gaaggaattt taatccatta
rrn23
14901 gtgagatacc acccccgaag agttaaaaat ctaaccttta aaaaggaaaa gtttcagata gacagtttct atggggcgta ggcctcctaa aaagtaacgg
rrn23
15001 aggcgtgtat gagtttcctc aatccagtca gagattggtt ttagagtata aaggcaaaag gaaatttgac tgtaagacat acatatcgag cagggatgaa
rrn23
15101 agtcggtctt agtgatccga cggtactgag tggaagggcc gtcgctcaac gaataaaagt tactctaggg ataacaggct gatcttcccc aagagatcac
rrn23
15201 atcgacggga aggtttggca cctcgatgtc ggttcttcgc tacctggagc tgaaatttgt tcctagggtt aggctgttcg cctattaaag cggtacgtga
rrn23
15301 gctgggttca gaacgtcgtg agacagtttg gtccatatcc ggtgtgagcg tgagaatatt gaggggagtc ttctttagta cgagaggatc gagaagaaaa
rrn23
15401 cacctatagt gtaccagtta ttgtacctac agtaaatgct gggtagccac gtgttgagtg gataattgct gaaagcatat aagtaagaag cccacctcaa
rrn23
15501 gatgagtgtt ctttaagatt tgatagttta attaattata aattaaatta tattaattaa ttaaattatc aatgaggtaa tagctagact agcttttaaa
rrn23
rrn4.5
15601 ttataaaata ggtattaagt ataagtacag taatgtatta agctaagata tactaataaa ccaaaaatat ttttttatta aataatttaa aattctggtg
rrn4.5
15701 tatataagta tatatgaact acaccaaaaa tccatttcga acttggtggt taaattttat tgcggtaaaa atactataga agatattctt tggaaaaata
rrn5
15801 actttatgcc aggataatat tattataatt atgatattta attaatttat atattatatt tatactatat acacatatac atacatacat ttaacatact
rrn5
ycf1
15901 aagatgtata cctttactag tattaaacca gtatctattg atacaggcta aatcctctaa tctataattt ggccataaaa aagatcttaa aagtgcttct
ycf1

16001 ttttatcat aatcattatt ataattataa ttaatataat tattaaaatt attatttttt ttctttaatg taaaagtaaa acaacttaaa attctaaatt
ycf1
16101 ctctacggcg attatatagt aaaatatttt caataattaa agaaaaaaaa taatttttac ttttttttt taaatataaa gatattttta cagtttgata
ycf1
16201 cataatacta tatccacttt cctgtaaata tttactaaaa ttatctatat taactattcc cctttctata agttcaattt ttagtacatt ttctccactt
ycf1
16301 cctattatat ttaaattaaa ttcatctcta agaattgatg aaagaataat ttttttagta ttttgtcttg attgaaaaat taaactatat atttttaaat
ycf1
16401 ttttaaaaat tttttcatat attttaagtt tatattgatt aaaagaattc gataatttac tttctttttc ggtatcttcg gtttttacat taattttaat
ycf1
16501 ttctttttca taaaaaaatt tcatattcgg tataaaatat ggttttaatt catatatttt tttctttaat atcaagtctg atagtatatt ataatcataa
ycf1
16601 ttactattaa tttgtttttt taataaatta ttatttttat aattttcaaa attaatatat ttatagcata gaaaattata tctataactt ttttttatct
ycf1
16701 taatttgttt ttttaataaa tttaatttta ataacgaata tttattaaat ttacttttta attttttttat ttgatataag ttgtattgat aataaaactt
ycf1
16801 taatttttt tttattgaat ttttaatata aaaagatgtt cctttatatt ttaaaataaa tcttaattcg tttttagata tattattaat ttctgtttgt
ycf1
16901 gctaatttat aaaatacgta tgcttgtgtt acgtaagata aatcaaaatt acaaggttta tatttatttt tattattaac ttttgaaatt ttaatataaa
ycf1
17001 ttttacaagc tttaataata aaaataaata tatttttat tattaatttt tttattaatt taaaatttcg aataaatcta gttttttttt ttaaagtttt
ycf1
17101 tatattaaat tttatatata atttaatatt ttgataaaaa aaaaatattt tatttttaa tttaaaatca aaaataaatt taataaattt atcaaaataa
ycf1
17201 aatttatatt tttttttctt tttttttatt ttataggttc taagtttaaa attgatattt ttgattttag aaatatcctt aattacgtca tccgttacgt
ycf1
17301 catcgattct tcttcttttt tgttttatta taattacatt tttatttttа cttacttttа ctttacttat attatatata atcacatata tattattaaa
ycf1
17401 tatatttaat aaatttaaat atatactttt aaaatatata atatatatat ttatatatat attatatatt aatttttta tatttttat atatttatat
ycf1
17501 ttatgtaggt ttactttaca attaataata tttaagtaaa aaaatatata tatttttta cttaatatat ttatttttaa atatattata tttcttaata
ycf1
17601 tatatataca atttttaaaa gatattatta tatttatatt tattatatct ttataatttt tatgtaaata tgtatttaaa tatgtattta ataatttatt
ycf1
17701 aataatataa ttaaaataat ttttttтcaa ataaaaaaat atatatatat tatatatata tttatctttt ttacttttta gttttttaaa aatgggttga
ycf1
17801 aaaatagatt taaaagtaga ttgacgctta cggttggggc caaaaggatg ttttgcttct cctccccaaa ctgttaaaaa aacatattga aatttttttt
ycf1
17901 tttgtcctat ttttaattgg aaaggggata aaatttttat ttgaaatcct tttattaaca aattttgtga attttcttct tctgataatg gaataccttc
ycf1
18001 ataagtacat tttatatata tttctttttt taaacgtttt aaatcttcat accattcggt agattcaaaa aaaaatatac ggataatatt tttgattatt
ycf1
18101 ataaaaaaat aaagtttaat atattttctt aaaaataatt gagttattaa cataactccc cttactattt gaccaaaaat aataccattc caagcttcgg
ycf1
18201 ctatttttaa tcttgatttt tctttatata atttttтaaa tttttcgtag tgtaattttt ttttataaat attatttttt ttttttaagg attttatgtt
ycf1
18301 aacaaagaaa tttatatact ttttaaaata attatatatt atatcaaaaa atctaaaaag aggccatttt cttattctat ctaaaaaaat aggagaatgt
ycf1
18401 acatttattt gaatattttt aaaagtaaca gtttttcgtc ttttaacccg tatagaacct tttattaatc ctcgtctaaa atctgatttt tcaaaataac
ycf1
18501 gtggagtaat aatagtcgtc tgttcactat tttcactact attactagta ttttcattat caatatcaga cttaccggta aaccctatta cacttttтaa
ycf1
18601 ttttcttgta cgtagttgac gtagaaataa atgtctagct tctctttcat ctttttctat taagttcatg tctatgtctt cgatcaattt atatggccat
ycf1
18701 agaggaattt tttttttтac tttttctatt tttatctttt cagttaattc tattgcagga aaattaataa ttagcattaa gatttttтgc ataattattt
ycf1
18801 ttaaattgaa agtttcttgt tcttgataaa gtttttтata tattttattg aatttattaa attttttttt ataaatatta tttgaataat aatataattc
ycf1
18901 tttagaaacg ggataaacat cggtaattat tctacgatgt gatccgcaaa ataaagaaaa taaagggtca taaattatgg gtaaatattt ttttttтctt
ycf1
19001 ttatttttac ataattgttt ttttttttct agtattatta ttagtattgt ttctaataat tttaaatttt gtggaaattt atttaatttt aataataata
ycf1
19101 aactttcatc caatttttct atttttattt taattttagc tttaaaaatt tcaaatttat tattattttt taaaatccac tctgtatcat cttgtgtaaa
ycf1
19201 gtctatttta aaatcaaaaa cttttactat tatttgatta aatttttтta aacttggtgg aaaggtgaaa gatattcttt tttttccata actctggcaa
ycf1
19301 gtattaaaaa aatactgtga aacttccttt tttatagcat ttgtttgctt atttatttca tcatctaaaa tatttatgta tcgtaaaggt tgattccatc
ycf1
19401 tattataatt aaatagtagt attacaagag gtttttcaaa ccaataaggt cttattctat ttagtaaata aatcgatagt gaagggttgg gtattttata
ycf1
19501 aaaaaatata ctttttttтt cgttaaatag tttacgtaaa tattcttctc tatcattgtt taaattaatt ttttttttta aataatttтt aataaattct
ycf1
19601 agaatagatt catccaaagt taaagatatt ttttcagagt ttaaattata tatattttct attttttttt ctttttggg tattgttaat ataaattttc
ycf1
19701 tagtaagaat aggaaaaggc attctaccta aattaaaaag acataaaata aatagaataa tattaaaaat ttgagtaact gacattttta aatttattat
ycf1
19801 aaaatattta tgtgtcaaat ttttccattt taatttttтt aatctattaa acaatatagg atttaattta attaatccaa cttttactaa atatttaaaa
ycf1
19901 aaaaaatttc taattttaaa aaaaaaatat ttaaataaat ttaatataaa tttactactt atcttattaa tattactaat attttttttt ttaaatataa
ycf1


20001 aatttaatgc aatctctgta tattttaata taaaaatata aattattaac caagctaaaa aagtacttgt tagaaataga aatttatcat tatttttaaa
ycf1
20101 tctaaacata tatatattta ataacctaaa aaaaaccgaa cttggcaaaa taaaattatt taaaatttga aaaataaaat tattaataaa agtaaaaata
ycf1
20201 atggtaaaat ctctaaaata attagtatta ttaattaatt tattattaga ttttttttaat aagaatccct tgaaagattt ttctttttta ggatttttt
ycf1
20301 taaaaaaaat aaataaccaa tatggtaaaa ccataatggt tattgtatgt ggttggtaaa aaattaaata taagggtgta taataaattg aaaataatat
ycf1
20401 tataagttgt cctagaataa aacctgttat agcagctatt ctttttttcat gtgctttttt ttcttccact tgcgcttgta aaataaaaat ataagcaggt
ycf1
20501 tcaacagata ttgtagtaat taatccaaaa tataatccaa tcacaaatat ggaatttatt atttttttac ttaatgtaat taaaattttt ttttgtaagt
ycf1
20601 ataaaatcat tgtagttacc tccttacttt ttttttataa ttattttgtt cgtatttgat ttctttacat aatatttcaa tttcttcaga aaataaccat
ycf2
20701 cgttttctta ataatgtttt tattattta tttaaaatat gttcgttaag taagaaaatt ttagataaat cttgatatat aatatataaa atattagagt
ycf2
20801 gtaaagattc tttatatttt aataatgaat ttgaattatt aagataccta tgattattaa tttttaactc gaaatatttt tctctaatac tatttttaaa
ycf2
20901 ccagtaattt aatagagatt tgggtgtata tttttgttga taggtaaaaa tttgattata tatttctata tctaaagata atttatcgtt ttttataaat
ycf2
21001 aatgataaat tatctttaga taaaaaagga tccgcgggat tccaaataaa ttgacttaag cttaatttaa ataatatttg ttccttaaaa ttaaagggta
ycf2
21101 tttttgctaa gttatttttt gttcttttat tttgatcatt attattttca tattctacaa cttctccatc atttttattt tttttaataa aaatattatt
ycf2
21201 aaaatttata aattcattaa attctttttt tgtatcttta tccttattta tatctatatt tatatcttct atatttatat ctatactagg aaaatttttа
ycf2
21301 taattttcaa tccaatcaaa taaaaaggtc cacggctgcc atatcttagg agctttgtct ataaagaaaa ataaagtttt atccatatct tctttcgtat
ycf2
21401 tattatccat attatttata tccatattat catacatatt attaatatca tcattattat cataattata taaaaattga tgatcaaaaa tataattttg
ycf2
21501 tatttttttа ctagaatatt tattataatt tatatatttt ctaaatagac tttctatttt taatatacca tgaactaatt ctgagtcatc ccttgcatgc
ycf2
21601 caaggataaa gtaattgttt ttcattttta tttgttttaa tatcatatac agaccaaaga tcttctgcta ctgaacctga ggaaaaatta aaaagataaa
ycf2
21701 aaaaaagtgt taatttttta aggtttcttc caagttcaaa ataccattga tataaataaa aagtatcgac taagtggtta ttaaaatata tagatatagg
ycf2
21801 atttatatga aaaatataaa attttgaaaa aaaattttgt gtaaaaactc tacctatttg atataaaatt atttccggaa ttttttttaa tgttgtagaa
ycf2
21901 tgagaaagcc aagtttgtct atatttaact aattcaatta aatttttatt tataataaat tttttattta taatattaat agatagagtc tcatttgcaa
ycf2
22001 aagttaaaag atcttgtgta tggaaaacca tatttatagg ctccgattta tcattatttt tttctttaaa aaatcccttc gtatataaaa gagtaaaaaa
ycf2
22101 atgtttactt tcttgattaa taaataattt acgtaattta atacatgtat taaattttgt tggagctatt aatataggat caactttact gggaatatgg
ycf2
22201 gttgaagcaa ttacaataac ttttttagat aaattagtta gtaaattatt taaagaaaat aaattaaact tagtcatagt aatgttatga atatttggaa
ycf2
22301 cccatattat acaaggagat aaagcttttg ctaattctaa ttgaatacta ataaaaattt gttctaaata gaacaaattt ttattagtag gattaaacat
ycf2
22401 atcttttata ctcatataaa tatctacagg taatatatta tttataaggt ttaattcttt tttcaatttt ttattataat tagtatctat attaaaatta
ycf2
22501 aaattaattt cattactatt taaaatatgt ttatcaaatt ttatttttttt aaaacccttt tgaggaatag gatttttaaa tttattcata aatatagtaa
ycf2
22601 taacaggaat atatgaggtt gaagttaaat atttaacaaa ataagattgt cctgcttctt tggaacctat taaaaaaatt cccccaagta aacgtaaatt
ycf2
22701 taactgtaga gtaaattctt tttgtaaatt ttttatttttt aaattcattt catcttcata ataatttatt ttatgataat atgaaagaaa aattcttctt
ycf2
22801 tctactaagc aagatgggtt aaatttaata ttaagaagat attttttttg aagattaaat aaatgttgta aatatattat tccgggttgt tctatatttt
ycf2
22901 taaatacata atcatttttt gataaagaat cattttgagt tagattaaat aatactttat ttgtttttaa ttctttgttt aaaacagata tttgaactaa
ycf2
23001 aaaacttttt tcatttaatt tattataaaa ccatttactc tttaatatca aatattcaat tttataatct attaaatatt cttttaattt atttttattt
ycf2
23101 ataattaatg aataaatttc tttacttgta tgacttataa atccggaaaa taaaaatttt ttaataataa aataaatgta atttaatata tttttaatat
ycf2
23201 aattaataaa attaatattt actttataat aatttttact atttaaaata atattattta tatttttaaa ccttaaaaaa taataagaat atattattct
ycf2
23301 aataatacct ttaaaaagat tataaaataa attatctata ttaagaaaat ataaagtata atttatccgt tttaactcaa ttatatatga aggtatcata
ycf2
23401 aaaatttttа ctttcataaa tgttgtacgt aattcattag agtattcata aactgaaaaa ataaatgaac ttaaaaaata aaaaattaat attaaaaaaa
ycf2
23501 atatatattt tggtaataca aaaatttttt tttctttaat taaaaaaatt tcttcgtata aaataatatt ttttaataat ttatttttta tttcgactac
ycf2
23601 catatttaac tttaattttg tagtatatat attaataaaa ttttttatat aattaaaaat ataataattt ctattattta taatataaat aataatattt
ycf2
23701 atttttttca ttaataatat aattaaagta attagatttt gaaaaataaa taaaaataaa atatatatat ttttatatcc gctaaaggta agaaaccaaa
ycf2
23801 atatatattt ttttaataaa ttaaatttaa atttaataaa tagaataata tatttttttt ttatggttgt atttttaata attttttttt catttaaatt
ycf2
23901 gggtaaatta ataatttttt tttcatttaa attgggtaaa ttaataattt ttttttcatt taaattgggt aaattaataa tttttttttc aaaatttcca
ycf2


24001 gaaatattta taaatgaatt aatatcttta taataattat tataagaatt aaaggattta tataaactat aactatattt ataactattt ttttttttta
ycf2
24101 atatataatt atatattaat ttattattat taaatatttt aatactgatt ttatgtttaa taaaatcagt attaaaatat ttaataagat tattataaag
ycf2
24201 aaaattatca atgttaatag ttttttttat aataaaagga aatttttttt tacataataa aagattatta agtgttttaa aattattaaa tttatttgac
ycf2
24301 ttataaaacg aagatattaa tgatttttta taaaaaatat tattttcaga aaataatccc atttctggaa tacaaacaaa tattgtgtat aaataataag
ycf2
24401 gattattata tttataattt aataatttat ttttttaga atatttattc gaataattat aatcataatt taataataag tacttttaa tattatttt
ycf2
24501 aatattaata cttattgaat ttaatttaat atttttaaa ttaataatta aacttttaaa aatttcaaaa gttttaataa ttagtctatc tctcgaaaca
ycf2
24601 aaataaaata aaaattcttt aattcttaaa tttattattt ttaatttttt gaggcagagt ttattaaatt tatttatatt tataataata ttatgtatat
ycf2
24701 ttataataat attatgtaat attttttttt tataaattaa ggggttatca aacgttattt ttatatattt aaagtttgat tttttattat attttaattt
ycf2
24801 atatttagtt cttaataata attctgaaaa atatgataaa atttttatta tatatatatc tatagatata tatttttcgg gatatgataa aatttttga
ycf2
24901 attgatatag tatttttaa ataaataatt atttttagta tattaaaaat ttctttattt tcatattgtt taaaaaaata tagttcaata ttttttataa
ycf2
25001 acttattttt atttaaatta tttttttcta taattaaatt aagttcagac atataatctg aatttaattt attatgttta ttataactat taaaagtttc
ycf2
25101 ttttaaagaa atatttttg ataataaaga aaaactaaat ttattggaaa catagaaatg aggcttttta ttttttttaa taacttttat agataaagga
ycf2
25201 agaagtttta ttaaatatat atttttattt tcaattaaat ttttattact aatatgaaaa aaaaaaaaaa atgataaaac taaaaaaatt aaacctttaa
ycf2
25301 ttttaagtat taaacatttt ttttctcctt ttgattttaa taatgaattt aaaatagtta aatttcttat attaaaaatc tttaaaaaac gctccggatt
ycf2
25401 taatataaaa ttaattatat ttattataaa atttaaatta atccataaat attgataatc ttttatttct ctaaaaactt cttttaattc taaaaaataa
ycf2
25501 aaatgaaatt gacttttata ttttttcata aaattcctcc ttatatgttt atttatatgt ttatgtgtta tactaattta tatattatat ttttttagca
ycf2
25601 tccatggctg agtggttaaa gcgcccaact cataattggt gaatgcgcag gttcaattcc tactggatgc atatgtttat ttatatttat tatttatata
trnI-CAU
25701 tatatataat atatataaat ttttatttta aaatattttc

Chapter 5. Summary

In this dissertation, a successful application of ptDNA enrichment protocols in *Mitrastemon kanehirai* was reported. These protocols are inexpensive, not time-consuming and do not need large amount of plant materials. These protocols can be very likely applied to other non-photosynthetic plants, to facilitate the studies of the plastid genome of heterotrophic plants.

By examining mt19S, pt16S and nr18S rDNA sequences from several heterotrophic plants, I found there is no correlation of evolutionary rate patterns among the three subcellular SSU rDNAs in heterotrophic plants. Although this research did not cover all of heterotrophic plants, it is the first study to analyze substitution rates of all three SSU rDNAs in heterotrophic plants. The result indicates that the accelerated evolutionary rate is not synchronized among the three subcellular SSU rDNAs for the same species, and the phenomenon is not ubiquitous in heterotrophic plants. It seems that the non-photosynthetic lifestyle has less impact on nucleotide substitutions in mitochondrial genome compared to the nuclear and plastid genomes. However, many factors could affect the evolutionary rate of plant genomes. I am looking forward having one non-photosynthetic plant with its three genome sequenced in the near future to give us a comprehensive view of subcellular genomes interaction in non-photosynthetic plants.

I reported the smallest plant plastome from *M. kanehirai*, with size of 25,740 bp, which encodes 4 rRNAs, 4 tRNAs and 18 protein-coding genes. The inverted repeat region is absent in *M. kanehirai* plastome, and *ycf2* and *rrn* genes included in IR were retained. This research gives us a broader view and deeper insight into the process of plastome evolution in heterotrophic plants, and provides more evidence to the 'enssential tRNAs' hypothesis. In addition, I addressed the challenge in sequencing non-photosynthetic plant plastomes. The sequence divergence, as well as the high AT content would be problematic in PCR-based amplification and the following sequencing. Also, I pointed out the correlation between the pt16S substitution rate and the plastome size in heterotrophic plants. The variation of pt16S sequence will serve as a good indicator for plastome size, helping researchers to evaluate appropriate targets for studying heterotrophic plant plastomes.

This dissertation described a practical application of plastid enrichment protocols in heterotrophic plants and evolutionary rate heterogeneity among SSU rDNAs in heterotrophic plants. It provides very useful experimental information of studying heterotrophic plant plastomes and the results confirmed that there is no pattern of the rate heterogeneity associated with plant life form. More important, this dissertation reported the smallest plastid genome that have been described and it helps us further understand the process of plastome evolution in heterotrophic plants.

Appendix

Table A1. GenBank accession numbers of the sequences used for relative rate test.

Taxa	Species	mt19S	pt16S	nr18S
Basal angiosperm				
Amborellales				
Amborellaceae	*Amborella trichopoda*	AF193987	NC005086	U42497
Piperales				
Aristolochiaceae	*Aristolochia macrophylla*	DQ008674	DQ629461	AF206855
Hydnoraceae	*Hydnora africana*	U82637	U67745	L25681
Saururaceae	*Saururus cernuus*	DQ008732	HQ664635	U42805
Winteraceae	*Drimys granadensis*		NC_008456	
Winteraceae	*Drimys winteri*	AF197162		U42823
Laurales				
Lauraceae	*Cinnamomum camphora*	DQ008711	HQ664631	AF206888
Lauraceae	*Laurus nobilis*	AF193990		AF197580
Lauraceae	*Sassafras albidum*			U52031
Monocots				
Asparagales				
Asparagaceae	*Asparagus officinalis*	DQ008678		
Iridaceae	*Iris* sp.	AF161087		
Orchidaceae	*Apostasia wallichii*		HQ183491	HM640780
Orchidaceae	*Cymbidium goeringii*			AJ271248
Orchidaceae	*Oncidium ensatum*			HM640779
Orchidaceae	*Oncidium Gower Ramsey*		NC014056	
Orchidaceae	*Phalaenopsis equestris*		NC017609	
Eudicots				
Brassicales				
Brassicaceae	*Arabidopsis thaliana*		NC_000932	X16077
Brassicaceae	*Brassica juncea*	NC_016123		
Brassicaceae	*Raphanus sativus*	AB694743		
Caricaceae	*Carica papaya*		NC_010323	U42514
Caryophyllales				
Amaranthaceae	*Beta vulgaris*	FQ014226		

Caryophyllaceae	*Silene latifolia*	NC014487	NC016730	
Caryophyllaceae	*Silene vulgaris*	HM562728		
Ericales				
Clethraceae	*Clethra alnifolia*			AF419793
Monotropoideae	*Pyrola picta*			U59936
Symplocaceae	*Symplocos paniculata*		HM164068	SPU43297
Theaceae	*Camellia obtusifolia*		HM164060	
Theaceae	*Stewartia malacodendron*		HM164067	
Lamiales				
Oleaceae	*Olea europaea*			L49289
Oleaceae	*Olea woodiana*		NC015608	
Pedaliaceae	*Sesamum indicum*		NC016433	AJ236041
Plantaginaceae	*Antirrhinum majus*		GQ997041	AJ236047
Plantaginaceae	*Digitalis purpurea*	AF193999		
Plantaginaceae	*Veronica agrestis*	AY818950		
Malvales				
Cytinaceae	*Cytinus ruber*	U82639	U47845	L24085
Solanales				
Convolvulaceae	*Convolvulus arvensis*			AJ236013
Convolvulaceae	*Ipomoea hederacea*			U38310
Convolvulaceae	*Ipomoea purpurea*		NC009808	
Solanaceae	*Nicotiana tabacum*	BA000042	NC001879	
Solanaceae	*Solanum tuberosum*		NC008096	
Santalales				
Balanophoraceae	*Corynaea crassa*	U82636	U67744	L24400
Erythropalaceae	*Heisteria concinna*		HQ664616	L24146
Santalaceae	*Lepidoceras chilense*	U82641		
Santalaceae	*Santalum album*			L24416
Ximeniaceae	*Ximenia americana*		GQ997924	L24428
Vitales				
Vitaceae	*Vitis vinifera*	NC012119		