



Minireview

History of chloroplast genomics

Masahiro Sugiura

Graduate School of Natural Sciences, Nagoya City University, Yamanohata, Mizuho, Nagoya 467-8501, Japan
(e-mail: sugiura@nsc.nagoya-cu.ac.jp; fax: +81-52-8726021)

Received 14 September 2002; accepted in revised form 9 December 2002

Key words: Erwin Baur, Lawrence Bogorad, chloroplast DNA, complete nucleotide sequence, Masahiro Ishida, physical map, rRNA gene, Ruth Sager, Masahiro Sugiura

Abstract

The presence of chloroplast DNA was established in 1963. With the development of recombinant DNA technologies, chloroplast DNA was selected as one of the first candidates for genome sequencing. The first physical map was reported for maize chloroplasts in 1976. As tobacco has been popular as an experimental system, tobacco chloroplast DNA has been extensively analyzed and the complete nucleotide sequence was established in 1986. This sequencing and the availability of tobacco chloroplast transformation techniques and of *in vitro* expression systems have formed the basis of an ongoing functional genomics program.

Introduction

Soon after the rediscovery of Mendelian genetics at the beginning of the last century, the first reports of non-Mendelian inheritance were published, based on studies of variegation in the higher plants *Pelargonium* and *Mirabilis*, respectively (Baur 1909; Correns 1909). The authors observed that the variegation of these leaves was inherited in a pattern that could not be explained under the framework of Mendelian genetics. Further analysis revealed that the genetic determinants for these characters were associated with chloroplasts. Recent analysis of their articles, however, leads to the conclusion that Erwin Baur alone (Baur 1909) deserves credit for the theory of chloroplast inheritance (Hagemann 2000). Further developments in chloroplast-associated inheritance were carried out mainly with the angiosperm *Oenothera* (reviewed in Harte 1994) and with the unicellular alga *Chlamydomonas* (reviewed in Gillham 1978a).

Discovery of chloroplast DNA

In 1951, it was first suggested that chloroplasts from the moss *Selaginella* and from two flowering plant

species contained DNA (Chiba 1951), and numerous related reports appeared in the late 1950s and early 1960s (reviewed in Sager and Schlanger 1976; Gillham 1978b; Bogorad 1998). Ris and Plaut (1962) demonstrated cytologically the presence of DNA in chloroplasts of *Chlamydomonas*. The following year, Ruth Sager and Masahiro R. Ishida (1963) (Figure 1; see Pardee 2001) provided strong evidence for the presence of unique DNA in chloroplasts from *Chlamydomonas*. They developed a method which allowed for the isolation of relatively intact chloroplasts from the alga and subsequently found an enriched satellite DNA with a buoyant density of 1,702 gm/cm³ and GC content of 39.3%, which could be distinguished from nuclear DNA (Figure 2). Similar results were reported from other laboratories during that same year (Chun et al. 1963; Kirk 1963), and thus by the end of 1963 the presence of chloroplast DNA was widely acknowledged. The Miescher–Ishida Prize was established by the International Society of Endocytobiology to recognize outstanding scientists who discovered or gave unique findings in the field of endocytobiology and to those who have contributed to the development of this science. The establishment of the Prize was suggested by Dr Masahiro R. Ishida in 1986



Figure 1. Ruth Sager (1918–1997) (left) and Masahiro R. Ishida (1926–2000) (right). Photos were obtained from Prof. Hiroshi Sano and Mrs. Miyo Ishida.

and sponsored originally with gold-plated medals, 'In commemoration of the discoveries of nuclear and extranuclear DNA.' The Prize is named in recognition for the DNA isolation by Dr Friedrich Miescher in Tübingen in 1869, and by Dr Masahiro R. Ishida, who first extracted chloroplast DNA together with Dr R. Sager in 1963 (cited from the Society Letters 'Endocytobiosis and Cell Research' 13: 138, 1999).

Following the initial discovery of chloroplast DNA, physical studies of the DNA structure were carried out using electron microscopy. Manning et al. (1971) observed the first circular DNA molecules in lysates of *Euglena* chloroplasts. Large circular DNA molecules have since been obtained from chloroplasts of several higher plant species (Kolodner and Tewari 1975; Herrmann et al. 1975).

Physical maps of chloroplast DNAs

The demonstration of a unique DNA species in chloroplasts led to intensive studies of both the structure of chloroplast DNA and its expression. Novel techniques developed in the mid-1970s, such as gene cloning and DNA sequencing, have helped to accelerate these studies. John R. Bedbrook and Lawrence Bogorad (1976) reported the first physical map of maize chloroplast DNA using digestion with several class II restriction endonucleases (Figure 3). This approach led to convincing evidence for the homogeneity and circularity of chloroplast DNA molecules (reviewed in Bedbrook and Kolodner 1979). At that time, restriction site maps of chloroplast DNA from other species were also reported (Vedel et al. 1976; Herrmann et al. 1976; Gray and Hallick 1976). Bogorad's group then cloned the first chloroplast gene from maize in 1977 (Bedbrook

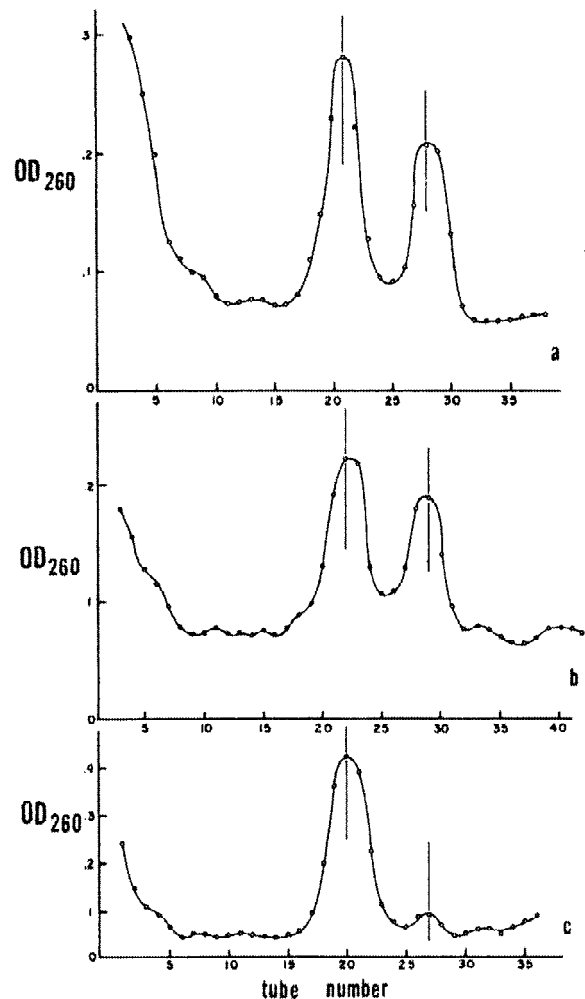


Figure 2. Banding of DNA from *Chlamydomonas reinhardtii* chloroplasts (a and b) and whole cells (c) in CsCl density gradients. The satellite band (peak on the right) was enriched in chloroplast preparations, representing chloroplast DNA. From Sager and Ishida (1963).

et al. 1977). Using cloned DNA fragments they reported the first fine structure map of the 16S-23S-5S rRNA genes in chloroplast DNA. The overall arrangement of the two copies of the rRNA genes in maize chloroplast DNA was determined by examining R-loops between rRNA and total chloroplast DNA in the electron microscope, and by mapping restriction sites on chloroplast DNA. These studies and others established a new field, 'chloroplast molecular biology,' and the organization and expression of chloroplast genomes were among the most extensively studied fields in plant molecular biology in the late 1970s and early 1980s (see Bogorad, this issue).

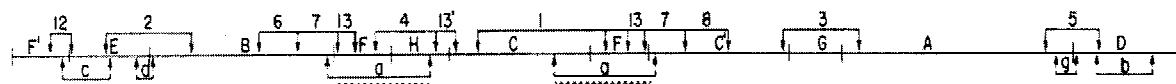


Figure 3. Physical map of the maize chloroplast DNA. *Sal* I fragments are shown in the central horizontal line by vertical lines. The *Bam*H I fragments are indicated on the upper horizontal line by arrows with solid tails, and the *Eco* RI fragments are shown on the lower line by arrows with tails. The 16 and 23S rRNA genes are known to lie within the region of the squiggled horizontal line. The circular DNA is shown in linear form. From Bedbrook and Bogorad (1976).

Tobacco chloroplast DNA

Tobacco plants are a well-established model experimental system; the presence of many mutants and interspecific hybrids has been of great use in classical genetic and evolutionary studies. Furthermore, the relative ease by which tobacco cells can be manipulated and cultured, for example in cell and protoplast cultures, protoplast fusion and regeneration, naturally lend them towards studying somatic cell genetics. This popularity has earned tobacco the nickname ‘the *E. coli* of the plant world’ (Kung 1984). Given that tobacco is one of the most widely used experimental plants, especially as a host of transgenic analyses, it should come as no surprise that it has been widely chosen for the study of the chloroplast genome by many laboratories. Physical maps of tobacco chloroplast DNA have been constructed (Jurgenson and Bourque 1980; Seyear et al. 1981; Fluhr and Edelman 1981; Hildebrand et al. 1985). As it has been difficult to isolate large quantities of chloroplast DNA in pure form, a clone bank of tobacco chloroplast DNA has been constructed (Fluhr et al. 1983).

In the summer of 1976, I stayed in the Basel Institute for Immunology, Switzerland, and had the opportunity to attend a series of evening seminars in the EMBO Course on ‘DNA restriction endonucleases: reactions and applications’ held at the Biozentrum, the University of Basel, organized by Werner Arber. This occasion was so exciting that I decided to change my discipline from the prokaryote (*Escherchia coli*) to the eukaryote. I favored plants rather than animals probably because my major was plant physiology during my undergraduate and graduate courses, and I was interested in chloroplast DNAs due to their small sizes. After three months stay in Basel, I returned to the National Institute of Genetics, Mishima, Japan. I consulted with plant geneticists in the Institute, which plant species I should use, and started preliminary work, such as preparation of plasmids, and restriction endonucleases and chloroplast DNA from tobacco (Sugiura 1998).



Figure 4. Analyses of tobacco chloroplast DNA was started in 1976 by Mie Suzuki, a technician (left), and the author, Masahiro Sugiura (right). Then Jun Kusuda, a post-doctoral fellow, joined us and later Mie and Jun were married.

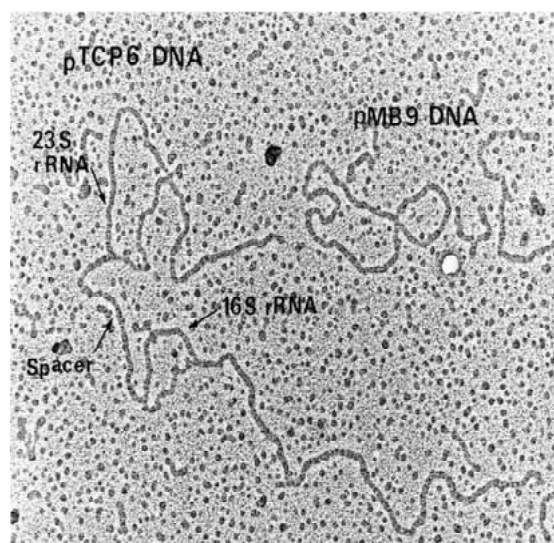


Figure 5. Detection of tobacco chloroplast 16S and 23S rRNA genes by the R-loop method. Electron micrograph of the R-loop formed between linearized pTCP6 DNA (consisting of 16S/23rDNA and pMB9, 8.7 Md and 16S/23S rRNAs). Circular pMB9 DNA (3.5 Md) is the plasmid vector used. For details, see Kusuda et al. (1980). This photo was taken by Kazuo Shinozaki in the late 1970s. From Sugiura (1987) with modifications.

Chloroplast DNA from leaves of tobacco (*Nicotiana tabacum* var. Bright Yellow 4) was prepared by our research group in 1977 and digested with *EcoRI* (Figure 4). The *EcoRI* DNA fragments that hybridized with tobacco chloroplast rRNAs were cloned into a plasmid pMB9 (the first practical cloning vehicle used in the 1970s) (Figure 5), and the result was published as our first report on chloroplast genome research (Sugiura and Kusuda 1979). The first nucleotide sequence was determined for the 4.5S rRNA and its gene by the 'wandering spot' and the Maxam-Gilbert methods, respectively (Takaiwa and Sugiura 1980a, 1980b). Since then, we have analyzed the structure and expression of a number of individual genes from tobacco chloroplasts (reviewed in Sugiura 1987, 1998).

In parallel with this work, tobacco chloroplast DNA was digested either completely or partially with several restriction enzymes. The resulting DNA fragments were cloned into several plasmid vectors including pBR322 (the most widely used vector in the late 1970s to early 1980s). A cosmid vector pHc79 (the first vector that could efficiently ligate DNA fragments 10 times larger than plasmid vectors) was used to separately clone the right and left segments of the large inverted repeat. Cloned DNA fragments were overlapping to eliminate the possibility that very short restriction fragments would be overlooked. This allowed for the construction of a clone bank of the entire tobacco chloroplast DNA as a set of overlapping restriction endonuclease fragments, which in turn would facilitate the sequencing of the entire chloroplast genome (Sugiura et al. 1986). Using this clone bank, we assembled individual gene sequences, filled gaps and finished the compilation of the entire tobacco chloroplast DNA sequence (Shinozaki et al. 1986). Kanji Ohyama came to see me in the National Institute of Genetics in the early 1980s and then sent his student for a short period to work together on their chloroplast DNA. We have a joint publication (Ohyama et al. 1983). In the same year, the entire sequence of *Marchantia polymorpha* chloroplast DNA was also reported (Ohyama et al. 1986). Hence, the genome projects of two chloroplast DNAs had finished before the term 'genome project' was acknowledged.

The gene order of tobacco is found in most other angiosperms and also represents the ancestral gene order of vascular plants (Palmer 1991). Therefore, tobacco DNA has often served as a reference for chloroplast genomes. Since 1986, many new genes have been identified and sequencing errors have also been

identified. Therefore, the revised version of tobacco chloroplast DNA sequence and gene map was published (Wakasugi et al. 1998). The most recent update of the tobacco chloroplast gene map was reported by Wakasugi et al. (2001). Many coworkers in my laboratory devoted time to sequencing and re-sequencing the DNA (Figure 6).

Concluding remarks

'Genomics' research can be divided into two stages. The first stage was the genome sequencing project. The chloroplast genome was selected as the first target in plants, and the sequencing project of tobacco started in the mid 1970s and completed in 1986. The complete nucleotide sequence of two chloroplast genomes has provided scientists with an enormous amount of fundamental information. The second stage is the post-genomic project or functional genomics stage, which in chloroplasts was already underway in 1986. The first focus was to identify novel genes from the complete nucleotide sequence data, this initially led to the identification of eighty-two different genes from the tobacco chloroplast genome. During the past 15 years, 31 additional genes have been identified, only two genes per year (Wakasugi et al. 2001). This illustrates clearly that the identification of novel genes is an extremely arduous task, requiring creative experimental approaches. In addition, ten long regions (1 kb or over) have yet to be assigned either to any gene or to any significant Open Reading Frames (ORFs). A gene encoding a small RNA of 218 nt was found in such a region and there is speculation that further small RNA species might be encoded there. Furthermore, RNA editing might create new protein-coding frames after transcription (Tsudzuki et al. 2001). Even with these recent developments, there is still much ambiguity about the full coding potential of the chloroplast genome.

The mechanism of chloroplast gene expression has been a most interesting target to deal with in the functional genomics era. Recent studies have shown that chloroplast gene expression is much more complex than previously thought; there are multiple classes of promoters, RNA polymerases, multiple RNA processing steps (RNA cleavage/trimming, *cis/trans* splicing, RNA editing, and RNA stability), and multiple mechanisms for translational initiation (reviewed in Sugiura et al. 1998). *In vitro* systems supporting accurate transcription (Kapoor and Sugiura 1999), trans-



葉緑体に感謝する会 杉浦昌弘先生中日文化賞受賞記念会 1997.6.14 於 ホテルブルブラ玉山

Figure 6. Photograph of laboratory members who held a meeting named 'We Thank Chloroplasts' to congratulate Sugiura (the author) on his receiving the Central Japan Culture Prize. The photo was taken on June 14, 1997, in Nagoya. *Front row (left to right):* Sanjay Kapoor, Meenu Kapoor, S. W. Gawronski, Anelia Vassileva, Masahiro Sugiura, Noriko Sugiura, Keiko Nakashima, Nobuaki Hayashida, Hiroaki Shimada, Mamoru Sugita. *In back of Sugiura:* Tetsuro Hirose (left) and Tatsuya Wakasugi (right).

lation (Hirose and Sugiura 1996, 1997) and RNA editing (Hirose and Sugiura 2001) are now available from tobacco chloroplasts. These systems together with tobacco chloroplast transformation techniques (Svab and Maliga 1993) will provide powerful tools to elucidate further genome structures, gene expression processes, and nuclear factors responsible for chloroplast genome expression.

Acknowledgments

This manuscript was edited by Tom Beatty and Govindjee.

References

- Baur E (1909) Das wesen und die Erblichkeitsverhältnisse der 'Varietates albomarginatae hort.' von *Pelargonium zonale*. Z Indukt Abstamm Vererbungsl 1: 330–351
- Bedbrook JR and Bogorad L (1976) Endonuclease recognition sites mapped on *Zea mays* chloroplast DNA. Proc Natl Acad Sci USA 73: 4309–4313
- Bedbrook JR and Kolodner R (1979) The structure of chloroplast DNA. Annu Rev Plant Physiol 30: 593–620
- Bedbrook JR, Kolodner R and Bogorad L (1977) *Zea mays* chloroplast ribosomal RNA genes are part of a 22,000 base pair inverted repeat. Cell 11: 739–749
- Bogorad L (1998) Discovery of chloroplast DNA, genomes and genes. In: Kung SD and Yang SF (eds) Discoveries in Plant Biology, Vol II, pp 15–43. World Scientific, Singapore
- Bogorad L (2003) Photosynthesis research: advances through molecular biology – the beginnings, 1975–1980s and on..... Photosynth Res 76: 13–33 (this issue)
- Chiba Y (1951) Cytochemical studies on chloroplasts. I. Cytologic demonstration of nucleic acids in chloroplasts. Cytologia (Tokyo) 16: 259–264
- Chun EHL, Vaughan MH Jr. and Rich A (1963) The isolation and characterization of DNA associated with chloroplast preparations. J Mol Biol 7: 130–141
- Correns C (1909) Vererbungsversuche mit blaß(gelb) grünen und buntblättrigen sippen bei *Mirabilis jalapa*, *Urtica pilulifera* und *Lunaria annua*. Z Indukt Abstamm Vererbungsl 1: 291–329
- Fluhr R and Edelman M (1981) Physical mapping of *Nicotiana tabacum* chloroplast DNA. Mol Gen Genet 181: 484–490
- Fluhr R, Fromm H and Edelman M (1983) Clone bank of *Nicotiana tabacum* chloroplast DNA: mapping of the alpha, beta and epsilon subunits of the ATPase coupling factor, the large subunit of ribulosebiphosphate carboxylase, and the 32-kDal membrane protein. Gene 25: 271–280
- Gillham NW (1978a) Chloroplast genetics of *Chlamydomonas*. In: Organelle Heredity, pp 347–438. Raven Press, New York
- Gillham NW (1978b) Genetic material of chloroplasts and mitochondria. In: Organelle Heredity, pp 39–79. Raven Press, New York
- Gray PW and Hallick R (1976) Restriction endonuclease map of chloroplast DNA from *Euglena gracilis*. In: Bucher Th, Neupert W, Sebald W and Werner S (eds) Genetics and Biogenesis of Chloroplasts and Mitochondria, pp 347–350. North-Holland Publishing, Amsterdam
- Hagemann R (2000) Erwin Baur or Carl Correns: who really created the theory of plastid inheritance? J Hered 91: 435–440
- Harte C (1994) *Oenothera* – Contribution of a Plant to Biology. Springer-Verlag, Berlin
- Herrmann RG, Bohnert HJ, Kowallik KV and Schmitt JM (1975) Size, conformation and purity of chloroplast DNA of some higher plants. Biochim Biophys Acta 378: 305–317
- Herrmann RG, Bohnert HJ, Driesel A and Hobom G (1976) The location of rRNA genes on the restriction endonuclease map of the *Spinacia oleracea* chloroplast DNA. In: Bucher T, Neupert W, Sebald W and Werner S (eds) Genetics and Biogenesis of Chloroplasts and Mitochondria, pp 351–359. North-Holland Publishing, Amsterdam
- Hildebrand M, Jurgenson JE, Ramage RT and Bourque DP (1985) Derivation of a physical map of chloroplast DNA from *Nicotiana tabacum* by two-dimensional gel and computer-aided restriction analysis. Plasmid 14: 64–79
- Hirose T and Sugiura M (1996) *Cis*-acting elements and *trans*-acting factors for accurate translation of chloroplast *psbA* mRNAs: development of an *in vitro* translation system from tobacco chloroplasts. EMBO J 15: 1687–1695
- Hirose T and Sugiura M (1997) Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast *ndhD* mRNA: a possible regulatory mechanism for expression of a chloroplast operon consisting of functionally unrelated genes. EMBO J 16: 6804–6811
- Hirose T and Sugiura M (2001) Involvement of a site-specific *trans*-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast *in vitro* RNA editing system. EMBO J 20: 1144–1152
- Jurgenson JE and Bourque DP (1980) Mapping of rRNA genes in an inverted repeat in *Nicotiana tabacum* chloroplast DNA. Nucleic Acids Res 8: 3505–3516
- Kapoor S and Sugiura M (1999) Identification of two essential sequence elements in the nonconsensus type II *PatpB-290* plastid promoter by using plastid transcription extracts from cultured tobacco BY-2 cells. Plant Cell 11: 1799–1810
- Kirk JTO (1963) The deoxyribonucleic acid of broad bean chloroplasts. Biochim Biophys Acta 76: 417–424
- Kolodner R and Tewari KK (1975) The molecular size and conformation of the chloroplast DNA from higher plants. Biochim Biophys Acta 402: 372–390
- Kung SD (1984) Genetic analysis of tobacco RuBPCASE. Critical Rev Plant Sci 1: 227–267
- Kusuda J, Shinozaki K, Takaiwa F and Sugiura M (1980) Characterization of the cloned ribosomal DNA of tobacco chloroplasts. Mol Gen Genet 178: 1–7
- Manning JE, Wolstenholme DR, Ryan RS, Hunter JA and Richards OC (1971) Circular chloroplast DNA from *Euglena gracilis*. Proc Natl Acad Sci USA 68: 1169–1173
- Ohyama K, Yamano Y, Fukuzawa H, Komano T, Yamagishi H, Fujimoto S and Sugiura M (1983) Physical mapping of chloroplast DNA from liverwort *Marchantia polymorpha* L. cell suspension cultures. Mol Gen Genet 189: 1–9
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umeshima K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H and Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322: 572–574
- Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L and Vasil IK (eds) The Molecular Biology of Plastids, pp 5–53. Academic Press, San Diego, California

- Pardee AB (2001) Ruth Sager, Feb. 7, 1918–March 29, 1997. Biographical Memoirs, Vol 80, pp 276–289. National Academy Press, Washington, DC
- Ris H and Plaut W (1962) Ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas*. *J Cell Biol* 13: 383–391
- Sager R and Ishida MR (1963) Chloroplast DNA in *Chlamydomonas*. *Proc Natl Acad Sci USA* 50: 725–730
- Sager R and Schlanger G (1976) Chloroplast DNA: Physical and genetic studies. In: King RC (ed) *Handbook of Genetics*, pp 371–423. Plenum Press, New York
- Seyer P, Kowallik KV and Herrmann RG (1981) A physical map of *Nicotiana tabacum* plastid DNA including the location of structural genes for ribosomal RNAs and the large subunit of ribulose biphosphate carboxylase/oxygenase. *Curr Genet* 3: 189–204
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H and Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5: 2043–2049
- Sugiura M (1987) Structure and function of the tobacco chloroplast genome. *Bot Mag Tokyo* 100: 407–436
- Sugiura M (1998) The discovery of the complete sequence of tobacco and rice chloroplast genomes. In: Kung SD and Yang SF (eds) *Discoveries in Plant Biology*, Vol II, pp 45–60. World Scientific, Singapore
- Sugiura M and Kusuda J (1979) Molecular cloning of tobacco chloroplast ribosomal RNA genes. *Mol Gen Genet* 172: 137–141
- Sugiura M, Shinozaki K, Zaita N, Kusuda M and Kumano M (1986) Clone bank of the tobacco (*Nicotiana tabacum*) chloroplast genome as a set of overlapping restriction endonuclease fragments: mapping of eleven ribosomal protein genes. *Plant Sci* 44: 211–216
- Svab Z and Maliga P (1993) High frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc Natl Acad Sci USA* 90: 913–917
- Takaiwa F and Sugiura M (1980a) The nucleotide sequence of 4.5S ribosomal RNA from tobacco chloroplasts. *Nucleic Acids Res* 8: 4125–4130
- Takaiwa F and Sugiura M (1980b) Nucleotide sequences of the 4.5S and 5S ribosomal RNA genes from tobacco chloroplasts. *Mol Gen Genet* 180: 1–4
- Tsudzuki T, Wakasugi T and Sugiura M (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol* 53: 327–332
- Vedel F, Quetier F and Bayen M (1976) Specific cleavage of chloroplast DNA from higher plants by EcoRI restriction nuclease. *Nature* 263: 440–442
- Wakasugi T, Sugita M, Tsudzuki T and Sugiura M (1998) Updated gene map of tobacco chloroplast DNA. *Plant Mol Biol Rep* 16: 231–241
- Wakasugi T, Tsudzuki T and Sugiura M (2001) The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. *Photosynth Res* 70: 107–118