

## **The entire gene organization of rice chloroplast DNA --- 2001 update**

J Tsudzuki<sup>1</sup>, T Tsudzuki<sup>2</sup>, T Wakasugi<sup>3</sup>, K Kinoshita<sup>4</sup>, T Kondo<sup>4</sup>, M Sugiura<sup>5</sup>

<sup>1</sup>*School of Life Studies, Sugiyama-Jogakuen University, Nagoya 464-8662, Japan.*

<sup>2</sup>*Department of Information and Policy Studies, Aichi-Gakuin University, Nisshin 470-0195, Japan.*

<sup>3</sup>*Department of Biology, Toyama University, Toyama 930-8555, Japan.*

<sup>4</sup>*CTI Co., Nagoya, 450-0003, Japan.*

<sup>5</sup>*Graduate School of Natural Sciences, Nagoya City University, Nagoya 467-8501, Japan.*

*Key words: chloroplast genome, Oryza sativa, gene organization*

### **Introduction**

The entire nucleotide sequence of the chloroplast genome (134,525 bp) from rice, *Oryza sativa* cv. *Nipponbare*, was determined and published in 1989 (Hiratsuka et al., 1989, Accession No. X15901). The rice chloroplast genes were identified by homology to their tobacco counterparts (Hiratsuka et al., 1989, Shimada et al., 1990, Shimada and Sugiura 1991). Since then, new genes in chloroplast have been identified with several methods (Stoebe et al., 1998, Hager et al., 1999, Swiatek et al., 2001) and update for the tobacco counterpart was summarized lately (Wakasugi et al., 1998). Here we summarize the newly identified genes in rice chloroplast genome during 12 years, and comparison made among the three completely sequenced chloroplast DNAs from monocot plants.

### **Result and Discussion**

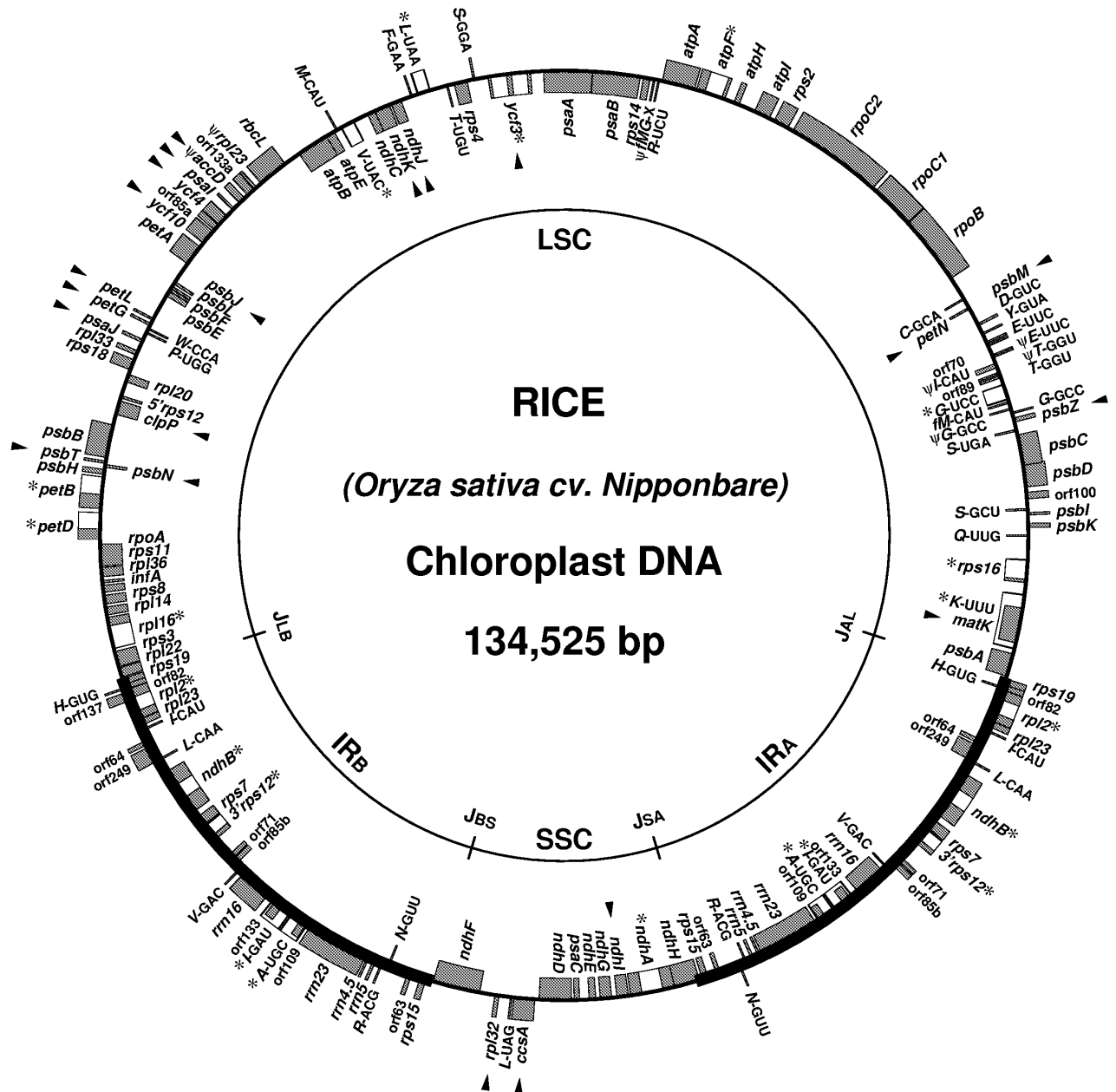
#### *Newly identified genes*

Nineteen new protein-coding genes have been identified on the genome since the first publication. The former *psbG* gene was renamed to *ndhK*. The *accD* is present as a pseudogene. Positions of these genes are shown on circle map of the genome illustrated in Figure 1 and listed in Table 1 according to those functions. There are 34 RNA genes, 74 protein-coding genes (including *ψaccD*), 3 hypothetical chloroplast open reading frames (*ycfs*), altogether 111 genes and *ycfs* in the genome. All these are listed in Table 2. The *sprA* found in tobacco (Vera and Sugiura, 1994) is not present in rice. The *rnpB* found in algae (Baum, et al., 1996) and supposed in maize (Collins, et al., 2000) is not confirmed yet in rice although partial homology is detected.

#### *Comparison among monocot chloroplast DNAs*

All genes and *ycfs* except *ψaccD* (total 110 genes) in the genome are shared with other entirely sequenced chloroplast genome from monocot plants, maize (140,387 bp, Maier et al., 1995, Accession No. X86563) and wheat (134,545 bp, Ogihara et al., 2000, Accession No. AB042240). Arrangements of those genes on three genomes are essentially the same. Three genomes also share several pseudogenes which located at inversion breakpoints (Shimada and Sugiura, 1989), but the *ψaccD* is found in neither maize nor wheat. Seventeen ORFs (60

codons or more in rice) are conserved among three monocot genomes, these include lately proposed 9 *ycfs* (*ycf68* -- *ycf76*) (Stoebe et al., 1998).



**Figure 1.** Genes and *ycfs* shown inside the circle are transcribed clockwise, and those on the outside are transcribed counter-clockwise. Newly identified genes (including *ψaccD* and renamed) and *ycfs* are indicated with triangle. ORFs (60 codons or more) are also illustrated. Asterisks indicate intron-containing genes.

**Table 1.** Newly Identified Rice Chloroplast Genes and *ycfs* during the Past 12 Years

Function	Gene	Position		Former name
		From	To	
Photosystem I	<i>psaI</i>	57,222	57,332	ORF36
	<i>psaJ</i>	64,622	64,756	ORF44
Photosystem II	<i>psbJ</i>	61,687	61,565	ORF40
	<i>psbM</i>	16,685	16,789	ORF34
	<i>psbN</i>	70,777	70,646	ORF43
	<i>psbT</i>	70,490	70,597	<i>ycf8</i> (ORF35)
	<i>psbZ</i>	11,937	12,125	<i>ycf9</i> (ORF62)
Cytochrome b/f complex	<i>petG</i>	63,799	63,912	ORF37
	<i>petL</i>	63,531	63,626	<i>ycf7</i> (ORF31)
	<i>petN</i>	17,645	17,556	<i>ycf6</i> (ORF29)
Ribosomal protein	<i>rpl32</i>	104,352	104,543	ORF63
NADH dehydrogenase	<i>ndhI</i>	110,536	110,000	ORF178
	<i>ndhJ</i>	48,471	47,988	ORF159
	<i>ndhK</i>	49,309	48,569	<i>psbG</i>
Other protein gene	<i>ccsA</i>	105,236	106,201	<i>ycf5</i> (ORF321)
	<i>clpP</i>	68,288	67,638	ORF216
	<i>matK</i>	3,296	1,668	ORF542
	<i>ψaccD</i>	56,553	56,873	pseudogene, ORF106
<i>ycfs</i>	<i>ycf3</i>	43,837	41,851	IRF170
	<i>ycf4</i>	57,702	58,259	ORF185
	<i>ycf10</i>	58,677	59,369	<i>cemA</i> , HBP(ORF230)

**Table 2.** Genes Contained in the Rice Chloroplast Genome

<b>Protein genes</b>	
<b>Photosynthesis</b>	
Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
Cytochrome b/f complex	<i>petA, *petB, *petD, petG, petL, petN,</i>
ATP synthase	<i>atpA, atpB, atpE, *atpF, atpH, atpI</i>
Rubisco	<i>rbcL</i>
<b>Ribosomal proteins</b>	
Large subunits	<i>*rpl2, rpl14, *rpl16, rpl20 rpl22, rpl23, rpl32, rpl33, rpl36</i>
Small subunits	<i>rps2, rps3, rps4, rps7, rps8, rps11, *rps12, rps14, rps15, *rps16, rps18, rps19</i>
<b>Transcription/translation</b>	
RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
Translation factor	<i>infA</i>
NADH-dehydrogenase	<i>*ndhA, *ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
Other protein genes	<i>ccsA, clpP, matK, <math>\psi accD</math></i>
<i>ycfs</i>	<i>*ycf3, ycf4, ycf10</i>
<b>number of protein genes: 77 (including pseudogene)</b>	

Table 2. Continued

RNA genes	
Ribosomal RNAs	<i>rrn23, rrn16, rrn5, rrn4.5</i>
Transfer RNAs	<i>*trnA(UGC), trnC(GCA), trnD(GUC), trnE(UUC), trnF(GAA), trnG(GCC), *trnG(UCC), trnH(GUG), trnI(CAU), *trnI(GAU), *trnK(UUU), trnL(CAA), *trnL(UAA), trnL(UAG), trnM(CAU), trnfM(CAU), trnN(GUU), trnP(UGG), trnQ(UUG), trnR(ACG), trnR(UCU), trnS(GCU), trnS(GGA), trnS(UGA), trnT(GGU), trnT(UGU), trnV(GAC), *trnV(UAC), trnW(CCA), trnY(GUA)</i>
number of RNA genes: 34	
Total gene number: 111 (including pseudogene)	

\* indicates intron-containing genes.

## References

- Baum, M., Cordier, A. and Schön, A. (1996) *Journal of Molecular Biology* **257**, 43-52
- Collins, L.J., Moulton, V. and Penny, D. (2000) *Journal of Molecular Evolution* **51**, 194-204
- Hager, M., Biehler, K., Illerhaus, J, Ruf, S. and Bock, R. (1999) *EMBO Journal* **18**, 5834-5842
- Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) *Molecular and General Genetics* **217**, 185-194
- Maier, R.M., Neckermann, K., Igloi, G.L. and Kössel, H. (1995) *Journal of Molecular Biology* **251**, 614-628
- Ogihara, Y., Isono, K., Kojima, T., Endo, A., Hanaoka, M., Shiina, T., Terachi, T., Utsugi, S., Murata, M., Mori, N., Takumi, S., Ikeo, K., Gojobori, T., Murai, R. Matsuoka Y, Ohnishi Y, Tajiri H and Tsunewaki, K. (2000) *Plant Molecular Biology Reporter* **18**, 243-253
- Shimada, H., Fukuta, M., Ishikawa, M. and Sugiura, M.(1990) *Molecular and General Genetics* **221**, 395-402
- Shimada, H. and Sugiura, M. (1989) *Current Genetics* **16**, 293-301
- Shimada, H. and Sugiura, M. (1991) *Nucleic Acids Research* **19**, 983-995
- Stoebe, B., Martin, W. and Kowallik, K. (1998) *Plant Molecular Biology Reporter* **16**, 243-255
- Swiatek, M., Kuras, R., Sokolenko, A., Higgs, D., Olive, J., Cinque, G., Müller, B., Eichacker, L.A., Stern, D.B. Bassi, R., Herrmann, R.G. and Wollman, F-A. (2001) *Plant Cell* **13**, 1347-1368
- Vera, A. and Sugiura, M. (1994) *EMBO Journal* **13**, 2211-2217
- Wakasugi, T., Sugita, M., Tsudzuki, T. and Sugiura, M.(1998) *Plant Molecular Biology Reporter* **16**, 231-241