Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA

Received: 17 July 2001 / Accepted: 15 October 2001 / Published online: 22 November 2001
© Springer-Verlag 2001

Abstract Structural features of the wheat plastome were clarified by comparison of the complete sequence of wheat chloroplast DNA with those of rice and maize chloroplast genomes. The wheat plastome consists of a 134,545-bp circular molecule with 20,703-bp inverted repeats and the same gene content as the rice and maize plastomes. However, some structural divergence was found even in the coding regions of genes. These alterations are due to illegitimate recombination between two short direct repeats and/or replication slippage. Overall comparison of chloroplast DNAs among the three cereals indicated the presence of some hot-spot regions for length mutations. Whereas the region with clustered tRNA genes and that downstream of rbcL showed divergence in a species-specific manner, the deletion patterns of ORFs in the inverted-repeat regions and the borders between the inverted repeats and the small single-copy region support the notion that wheat and rice are related more closely to each other than to maize.

Keywords Chinese Spring wheat · Chloroplast DNA · Complete sequence · Hypervariable region · Structure analysis

Introduction

A characteristic feature of plastomes is the occurrence of long inverted repeats, ranging in size from 10 to 85 kb (Palmer 1991), that separate the rest of the molecule into large and small single-copy regions. It is well known that the structure and gene content of plastomes are conserved among diverse plants. The complete sequencing of chloroplast DNAs from various plants, such as tobacco (Shinozaki et al. 1986), rice (Hiratsuka et al. 1989), maize (Maier et al. 1995), Arabidopsis (Sato et al. 1999), Oenothera (Hupfer et al. 2000), spinach (Schmitz-Linneweber et al. 2001), black pine (Wakasugi et al. 1994), liverwort (Ohguma et al. 1986), and the algae Chlorella (Wakasugi et al. 1997), and Euglena (Hallick et al. 1993), has confirmed the conservative nature of plastome evolution. Although the overall structure of the plastome is thought to be conserved by the stabilizing action of the long inverted repeats, structural alterations in plastomes, such as inversions (Howe et al. 1988; Hiratsuka et al. 1989), translocations (Ogi...
1988), and deletions (Palmer 1991), have been found among angiosperms. Furthermore, sequence analysis of chloroplast DNAs in related plants has allowed the identification of hot spots for length mutations (Ogi- 
hara et al. 1988), and this comparative approach is expected to further our understanding of the mecha-
nism(s) of, and better define the traits affected by, plastome evolution.

Wheat is one of the major crops in the Gramineae, 
and is the most important member of the Triticeae. 
The genome constitution of each species belonging to 
the *Triticum-Aegilops* complex and the phylogenetic 
relationships among them are well defined (Kihara 
1954). In addition to the nuclear genome analysis, 
nucleus-cytoplasm (NC) hybrids have been established 
by combining tester nuclei from common wheat with 
the cytoplasm from all other *Triticum-Aegilops* species 
(Tsunewaki 1993). By cultivating these NC hybrids, 
the biological effects of alien plasmons can be ana-
yzed precisely. Accordingly, the molecular basis of 
nucleus-cytoplasm interaction in wheat species is now open for clarification.

In order to analyze the overall structure of the 
wheat plastome, we have recently completed the se-
queencing of the wheat chloroplast DNA, which re-
resents the third such plastome to be fully sequenced 
among the grasses (Ogihara et al. 2000). In that 
report, we listed the plasmid clones which cover the 
entire wheat plastome, and the gene content of each. 
By comparison of the entire sequence of the plastomes 
of three cereals, namely, wheat, rice and maize, we 
have now clarified the structural features of the hy-
pervariable regions in the plastomes of these grass 
plants and the alterations in gene structure between 
them. The results of this analysis are presented here.

---

**Materials and methods**

The entire sequence of wheat chloroplast DNA was determined as previously reported (Ogihara et al. 2000). Nucleotide 
sequences were retrieved from the DDBJ, EMBL and NCBI 
nucleotide data banks. Assembly and manipulation of sequences 
were performed with the GENETYX program (Software 
Development Co., Tokyo,). Identity searches were carried out 
with the FASTA (Pearson and Lipman 1988) and BLAST 
(Altschul et al. 1990) programs. The entire sequences of chlo-
roplast genomes from wheat, rice, maize and tobacco were 
compared with each other by using the harr plot program 
(Sonnhammer and Durbin 1995).

---

**Results and discussion**

**Size and genetic organization of the wheat 
chloroplast genome**

The circular wheat chloroplast DNA is 134,545 bp long, 
and each of the inverted repeats (IRs) is 20,703 bp in 
length. The IR regions divide the rest of the sequence 
into segments of 80,349 bp [the large single-copy (LSC) 
region] and 12,790 bp [the small single-copy (SSC) 
region], as shown in Fig. 1. Although the molecular size of 
wheat chloroplast DNA was previously reported to be 
134,540 bp (Ogihara et al. 2000), further examination of 
sequence traces have led to the revised size of 
134,545 bp. The arrangement and locations of the 
plastome genes are also given in Fig. 1. The number and 
content of functional wheat chloroplast genes, so far 
identified, are identical to those of rice and maize 
(Hiratsuka et al. 1989; Maier et al. 1995).

However, some structural alterations in the chol-
roplast genes have been recognized even among grass 
plastomes. These alterations are the result of two 
processes: replication slippage or intramolecular re-
combination mediated by simple direct repeats, and 
simple nucleotide loss during replication (Ogihara et al. 
1991). As regards the former, an 81-bp deletion is 
found in the middle of the wheat *rpoC2* gene, but not 
in that of rice or maize. Since a duplicated 5-bp oli-
gonucleotide (CTTTT) is located at each end of the 
deleted portion in the chloroplast DNAs of rice and 
maize, this 81-bp deletion is assumed to have been 
caused by recombination via the short direct repeats 
(Ohnishi et al. 1999). A second example is an 18-bp 
deletion in wheat *infA*. A tandem duplication of an 18-
bp unit is located near the 3' terminus of the *infA* 
genes from rice and maize. One of these tandem 
repeats was lost in the wheat gene. As for the latter 
process, a 19-bp deletion at the 3'-end of the wheat 
*rpl22* gene, a 6-bp deletion near the 5'-region of rice 
*ndhI*, a 3-bp deletion near the 5'-region of wheat 
*ndhK*, a 2-bp deletion at the 3'-end of wheat 
*atpA*, and a 1-bp deletion at the 3'-end of rice 
*atpF* and wheat *ycf5* were detected by comparison of the entire sequences of the 
chloroplast DNAs of the three grasses. These nucleo-
tide eliminations resulted in deletions of one amino 
acid (*rpl22* and *ndhK* of wheat), two amino acids 
(wheat *atpA* and rice *ndhI*), or three amino acids (rice 
*atpF*), and the addition of one amino acid in one case 
(wheat *ycf5*). These structural alterations are not lo-
cated in conserved regions of these genes, and there-
fore, these changes probably do not affect protein 
function.

**Sequence comparison of the whole plastome among 
grass plants**

The wheat chloroplast genome represents the third to be 
completely sequenced in grass plants, and also in 
monocots. Because rice (which provided the first com-
plete sequence of a plastome from monocots; Hiratsuka 
et al. 1989), maize (second candidate; Maier et al. 1995) 
and wheat have diverged almost equally among grasses 
(e.g. Chase et al. 1993), the availability of the complete 
plastome sequences of these three grass plants should 
help to clarify the variability of the plastomes during the 
evolution of grass plants.