Curr Genet (1997) 31: 276–279

SHORT COMMUNICATION

Jotaro Aii · Yuji Kishima · Tetsuo Mikami
Taiji Adachi

Expansion of the IR in the chloroplast genomes of buckwheat species is due to incorporation of an SSC sequence that could be mediated by an inversion

Received: 6 May / 1 November 1996

Abstract The chloroplast genomes in buckwheat species contain large inverted repeats which are at least 4 kbp longer than the majority of those in land plants. The length of the buckwheat inverted repeats was attributable to an additional region located adjacent to the borders of the small single-copy region. We have cloned and sequenced a 5.2-kbp *Sma*I fragment corresponding to this extra region in the inverted repeats. A homology search revealed that the sequence of the *Sma*I fragment is highly homologous to one side of the small single-copy region of the inverted repeats in dicot chloroplast DNAs such as tobacco and beechdrops. Interestingly, a 3.7-kbp segment in the middle of the *Sma*I fragment is inserted in the opposite orientation relative to those of the other dicot species, and 17-bp direct repeats are found located at both the ends of the additional region. These results suggest that expansion of the inverted repeats in buckwheat chloroplast DNA might have been associated with an inversion.

Key words Buckwheat · Chloroplast DNA · Inversion · Inverted repeats

Introduction

Large inverted repeats (IR) are a characteristic feature of chloroplast genomes in most plants, except for some legume species and all conifers (reviewed by Palmer 1991). Previous studies have suggested that the IR is important for structural stabilization of cpDNA (Palmer and Thompson 1982) and conservation of the essential chloroplast genes (Palmer and Stein 1986). Most species have an IR of about 25 kbp (Palmer 1991), although the IR can vary from 10 to 76 kbp among land plants. The size of IR in land plants apparently tends to be larger (Palmer and Stein 1986), and smaller IRs (less than 15 kbp) have been found mainly in lower land plants such as *Marchantia polymorpha* (Ohyama et al. 1986) and some fern species (Stein et al. 1992). The cpDNA of geranium has an unusually large IR (76 kbp) and is the only example to have been analyzed in terms of its expansion process (Palmer et al. 1987). It has been postulated that the enlargement of the IR in geranium has been accompanied by several inversions.

We have analyzed the chloroplast genomes in buckwheat (*Fagopyrum*) species and found that the IR is at least 4 kbp longer than the average for land plants (Kishima et al. 1995). In the present paper we provide sequence data associated with the expansion of the IR, document an inversion, and discuss the expansion process.

J. Aii · T. Adachi
Applied Genetics and Biotechnology Division,
Faculty of Agriculture, Miyazaki University,
Miyazaki 889-21, Japan

Y. Kishima (✉) · T. Mikami
Laboratory of Genetic Engineering, Faculty of Agriculture,
Hokkaido University, Sapporo 060, Japan

Communicated by K. J. Newton

Materials and methods

Buckwheat cpDNA was prepared as described by Kishima et al. (1990) with a slightly modified chloroplast isolation procedure in which a 20–45% stepwise sucrose gradient was applied. A 5.2-kbp *Sma*I fragment (*Sma*-8) located as shown on the map (see Fig. 1), was eluted from an 0.7% agarose gel and cloned into Bluescript SK+ (Stratagene). For sequencing, *Sma*-8 was subcloned by the deletion method using the Kiro Sequencing Deletion Kit (Takara). Sequencing was performed by dideoxy chain-termination using an ALF DNA sequencer (Pharmacia). A homology search of the sequence was conducted through Internet Database programs BLAST provided by GenomeNet in Kyoto, Japan. The *Sma*-8 sequence reported in this paper has been deposited in the DDBJ database (accession no. D82050).

Results and discussion

We previously reported that the chloroplast genomes in buckwheat species have a relatively long IR of at least

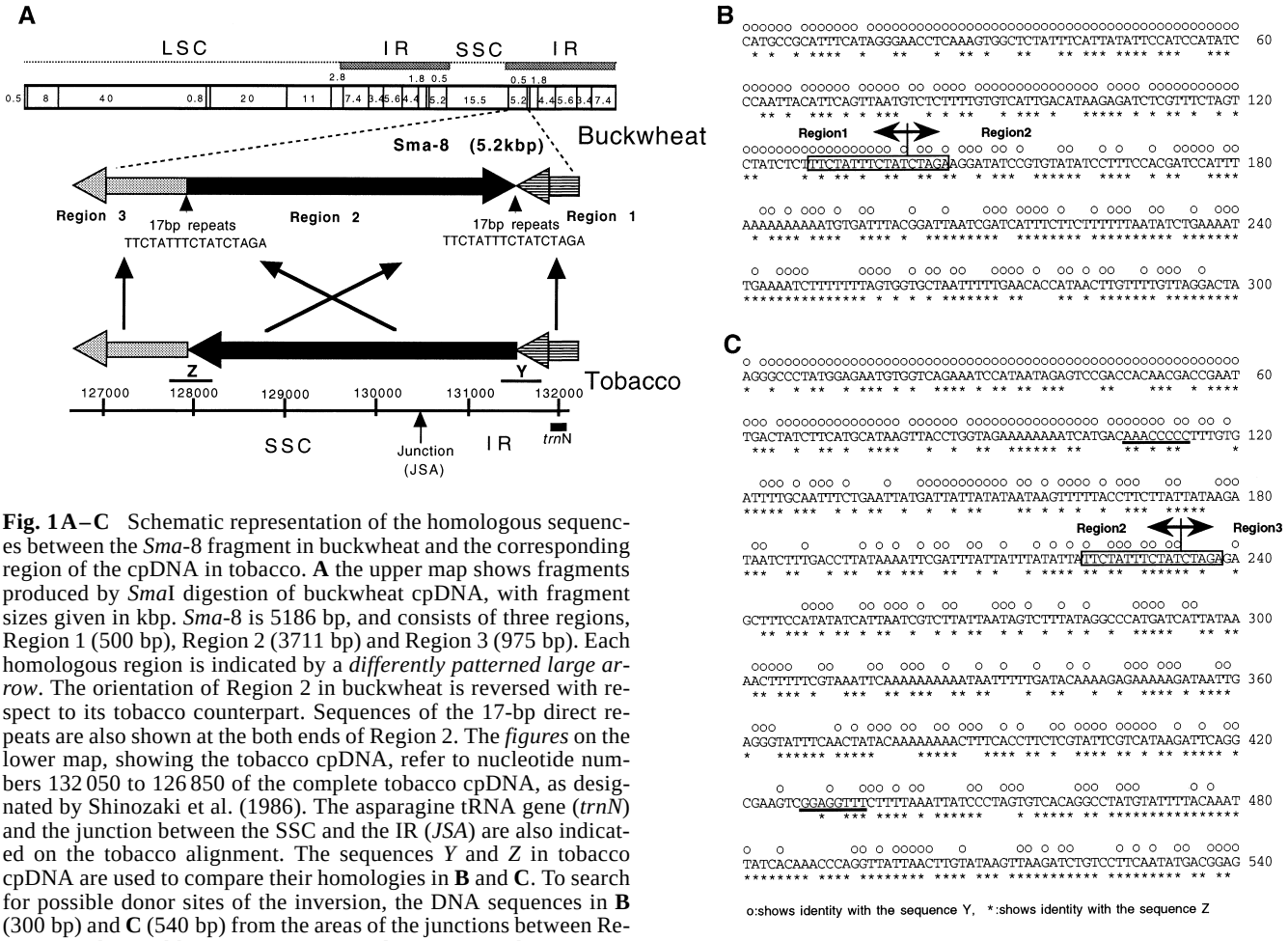

Fig. 1A–C Schematic representation of the homologous sequences between the *Sma*-8 fragment in buckwheat and the corresponding region of the cpDNA in tobacco. **A** the upper map shows fragments produced by *Sma*I digestion of buckwheat cpDNA, with fragment sizes given in kbp. *Sma*-8 is 5186 bp, and consists of three regions, Region 1 (500 bp), Region 2 (3711 bp) and Region 3 (975 bp). Each homologous region is indicated by a *differently patterned large arrow*. The orientation of Region 2 in buckwheat is reversed with respect to its tobacco counterpart. Sequences of the 17-bp direct repeats are also shown at the both ends of Region 2. The *figures* on the lower map, showing the tobacco cpDNA, refer to nucleotide numbers 132 050 to 126 850 of the complete tobacco cpDNA, as designated by Shinozaki et al. (1986). The asparagine tRNA gene (*trnN*) and the junction between the SSC and the IR (*JSA*) are also indicated on the tobacco alignment. The sequences *Y* and *Z* in tobacco cpDNA are used to compare their homologies in **B** and **C**. To search for possible donor sites of the inversion, the DNA sequences in **B** (300 bp) and **C** (540 bp) from the areas of the junctions between Region 1 and 2, and between Region 2 and 3, respectively, are compared with both Y and Z. *Open circles* above nucleotides indicate nucleotide matches with those of sequence Y (**A**) from the tobacco cpDNA. *Asterisks* below show nucleotide matches with sequence Z (**A**). The sequences Y and Z correspond to nucleotides nos. 131 300–131 700 and nos.127 800–128 200, respectively, in tobacco cpDNA (Shinozaki et al. 1986). The *boxed sequences* indicate the 17-bp direct repeats, and the *underlined* sequences are the imperfect 8-bp inverted repeats

28.3 kbp (Kishima et al. 1995). Other angiosperm species usually have IRs of 22–26 kbp (Palmer 1991). It is thought that the large size of the IR in buckwheat is due to the inclusion of an additional sequence near the ends of the IR, and we have therefore cloned and sequenced *Sma*-8 fragment consisting of 5186 bp which lies at the ends of the IR beside the small single-copy region (SSC). Despite the low level of the hybridization with tobacco probes (Kishima et al. 1995), *Sma*-8 and tobacco cpDNA were found to be over 70% similar in 70% of the whole *Sma*-8 sequence. This homology occurs at one of the border regions of the IR and the SSC in tobacco. This region corresponds to nucleotide nos. 126 850–132 050 on the complete sequence of tobacco cpDNA (Shinozaki et al. 1986). However, this homologous region is interrupted by an inversion within *Sma*-8.

The schema depicted in Fig. 1 A represents the *Sma*-8 in buckwheat and its homologous region in tobacco. The *Sma*-8 fragment can be divided into three parts. Region 1 (0.5 kbp) is homologous to the tobacco IR sequence adjacent to the SSC (nucleotide nos. 132 050–131 600). This region includes a gene for asparagine tRNA, *trnN*. In Region 2, the next 3.7 kbp reverses the tobacco homologous sequence representing the junction between the SSC and the IR (nucleotide nos. 128 000–131 600). This junction in tobacco cpDNA was designated as JSA (Shinozaki et al. 1986). We found direct repeats of 17 bp (TTCTATTTC-TATCTAGA) at both ends of Region 2. Comparison with tobacco cpDNA indicates that Region 3 (1.0 kbp) corresponds to one side of the SSC (nucleotide nos. 128 000–126 850).

Similar results were also obtained in beechdrops (*Epifagus virginiana*) which has a minimal chloroplast genome (70 kbp) (Wolfe et al. 1992). In beechdrops, the length of the SSC was reduced to 26% of that of the tobacco, but the *Sma*-8 homologous region was preserved. However, monocot cpDNAs such as rice and maize (Hiratsuka et al. 1989; Maier et al. 1990) do not show similarity to the *Sma*-8 sequence.

To identify the exact endpoints of the inversion, the sequences around the termini of Region 2 were examined. Figure 1 B, C shows the nucleotide identities corresponding to the sequence Y and Z in tobacco cpDNA (see

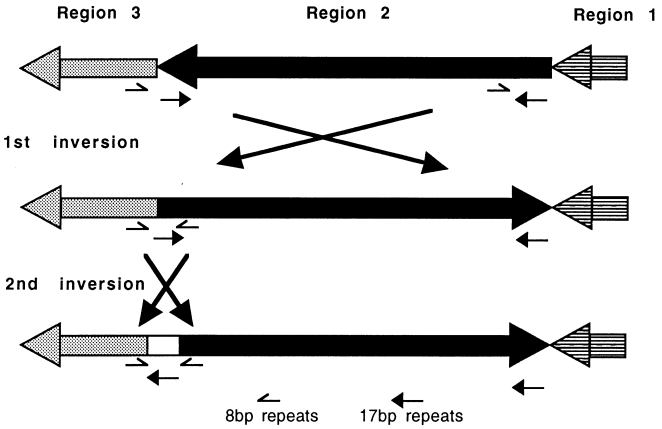

Fig. 2 Postulated event resulting in two inversions within the *Sma*-8 segment. Recombination between the 17-bp inverted repeats could have resulted in the 1st inversion. Recombination between 8-bp repeats could have resulted in the 2nd inversion. The large, *differently patterned arrows* correspond to each region in *Sma*-8, while the *white box* shows the sequence between the 8-bp repeats that has little similarity to the Y and Z tobacco cpDNA sequences (Fig. 1 C). It is not clear how the sequence between the 8-bp repeats lost its identity to its counterparts

Fig. 1A). Sequences Y and Z should have similarities in putative donor sites for the inversion. If so, homologies of those sequences to *Sma*-8 would cross at certain points near the ends of Region 2. In Fig. 1 B, a high level of the identities, 92% and 75%, was obtained in the right part of Y vs the left end of Region 1 and the right part of Z vs the right end of Region 2, respectively. This illustrates that the 17 bp between Regions 1 and 2 should be an endpoint of the inversion. However, further exchanges between Y and Z are unlikely to occur at the other 17-bp area, between Regions 2 and 3 (Fig. 1 C). Instead, two imperfect 8-bp inverted repeat sequences, AAACCCC and GGAGGTTT, interrupt 94% and 80% matchings with the sequences from the left part of Y and the left part of Z, respectively. These imperfect 8-bp inverted repeats might have induced a smaller inversion. We can postulate that two inversions took place in the *Sma*-8 fragment, as shown in Fig. 2; the first inversion occurring in the 3.7 kbp between the 17-bp repeats which used to be in inverted orientation, and the second inversion arising between the 8-bp imperfect repeats on both sides of the 17 bp at the junction of Regions 2 and 3. As a result of the second inversion, the two 17-bp sequences have the same orientation.

It is generally difficult to judge whether inversion or expansion of the IR occurred first. In the case of buckwheat, however, we can assume that the inversion occurred before expansion of the IR, because this inversion had to connect the junction between the IR and SSC which existed in the ancestral species. In addition, the region incorporated into the IR is on the border where the inversion took place.

Considering this circumstantial evidence, we can hypothesize about the expansion processes which occurred in the IR of the buckwheat progenitor. After the inversion, direct repeats of about 1 kbp would have been constructed with a part of Region 2 which used to be the end of the IR. Such long direct repeats could cause a deletion due to a loop-out-type recombination as is often seen in mitochondrial genomes (Lonsdale et al. 1988). Since the long direct repeats might have been brought to both sides of the SSC region, the cpDNA with the inversion had to abolish the deletion of the SSC or the rest of the cpDNA. To achieve this, it may be that the region containing the inversion was integrated into the IR. Palmer (1985) proposed a mechanism for the incorporation of part of the SSC into the IR, resulting in an expansion of the IR. Short inverted repeats located outside of the IR would be paired at the same time that the IR itself is paired and subsequent copy correction would lead to duplication of the intervening single-copy region. In the case of buckwheat, according to this hypothesis, the copy correctional duplication might have taken place through the pairing of short repeats located in the SSC beyond *Sma*-8 and the IR. Afterwards, the cpDNA with the long IR including *Sma*-8 would have been selected in the buckwheat progenitor. We are currently investigating further downstream from the *Sma*-8 fragment as well as the IR structures in other species in the Polygonaceae of which buckwheat is a member.

Acknowledgements We thank Mr. T. Kageyama for excellent technical assistance. We also thank Drs. T. Kubo and Y. Yanai for their helpful comments and discussions on the manuscript. A part of this work was done at the Isotope Center, Miyazaki University, and the Research Center for Molecular Genetics, Hokkaido University. Contributions of the Laboratory of Plant Breeding, Faculty of Agriculture, Miyazaki Univ., No. 95.

References

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 K, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Kishima Y, Yabuya T, Adachi T (1990) Isolation and macromolecular analysis of chloroplast DNA in common buckwheat. Fagopyrum 10:65–68

Kishima Y, Ogura K, Mizukami K, Mikami T, Adachi T (1995) Chloroplast DNA analysis in buckwheat species: phylogenetic relationships, origin of the reproductive systems and extended inverted repeats. Plant Sci 108:173–179

Lonsdale DM, Brears T, Hodge TP, Melville SE, Rottmann WH (1988) The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. Phil Trans R Soc Lond 319:149–163

Maier RM, Doery I, Igloi G, Koessel H (1990) The *ndhH* genes of graminean plastomes are linked with the junctions between small single-copy and inverted repeat regions. Curr Genet 18:245–250

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (eds) Molecular biology of plastids and the photosynthetic apparatus. Academic Press, New York, pp 5–53

Palmer JD, Stein DB (1986) Conservation of chloroplast genome structure among vascular plants. Curr Genet 10:823–833

Palmer JD, Thompson W (1982) Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29:537–550

Palmer JD, Nugent JM, Herbon LA (1987) Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc Natl Acad Sci USA 84:769–773

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 B-Y, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Stein DB, Conant DS, Ahearn ME, Jordan ET, Kirch SA, Hasebe M, Iwatsuki K, Tan MK, Thomson JA (1992) Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. Proc Natl Acad Sci USA 89:1856–1860

Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. Proc Natl Acad Sci USA 89:10648–10652