Curr Genet (1988) 14:287–292

Current Genetics

Variation in copy number of a 24-base pair tandem repeat in the chloroplast DNA of *Oenothera hookeri* strain Johansen*

Kimberly Blasko, Sara A. Kaplan, Kelly G. Higgins, Ruth Wolfson, and Barbara B. Sears

Department of Botany and Plant Pathology, Michigan State University, East Lansing, MI 48824, USA

Summary. A highly variable region of chloroplast DNA has been analyzed from three isolates of *Oenothera hookeri* strain Johansen. The variability results from the presence of two, four or seven copies of a discrete 24-base pair tandem repeat in a segment of the chloroplast DNA within the inverted repeat. Alignment of this DNA region with the published tobacco cpDNA sequence shows that in *Oenothera*, the repeats are insertions within a large unidentified reading frame, with each repeat unit specifying an eight amino acid in-frame addition. A model to explain the frequent alterations in the copy number of this 24-bp unit is proposed: imprecise alignment and recombination between the two large inverted repeats followed by copy correction could result in an amplification or deletion of the 24-bp segments.

Key words: Chloroplast DNA – Recombination – *Oenothera*

Introduction

Chloroplast DNA (cpDNA) is considered to be highly conserved in evolutionary terms (Sears 1983; Curtis and Clegg 1984; Palmer 1985a, 1985b; Zurawski and Clegg 1987). Few differences have been noted in intraspecific comparisons using restriction endonuclease digestion patterns (e.g. Scowcroft 1979; Timothy et al. 1979; Metzlaff et al. 1981; Palmer and Zamir 1982; Clegg et al. 1984; Banks and Birky 1985). DNA sequence conservation is high, even when broad evolutionary comparisons are made (e.g., Shinozaki et al. 1983; Curtis and Clegg 1984; Willey et al. 1984; Zurawski et al. 1984; Zurawski and Clegg 1987). Comparing cpDNAs between species within the same genus (Gordon et al. 1982; Salts et al. 1984; Palmer et al. 1985; Schmitz and Kowallik 1986; Doebley et al. 1987) has indicated that frequent changes occur by deletion/insertion events. Our investigations have focused on *Oenothera hookeri* strain Johansen, which contains plastome type I according to the classification scheme of Stubbe (reviewed by Kutzelnigg and Stubbe 1974). Initially, we found variations indicative of insertion/deletion events in several discrete regions of cpDNA from plants within a *plastome mutator* isolate of this plant line (Johnson et al. 1988). Because the DNA variations did not correlate with the occurrence of plastome mutations, we concluded that the insertion/deletion events were not causally related to the mutations. Furthermore, when we extended these characterizations to a wild-type (non-*plastome mutator*) line of the same species, we found additional variability in the size of the same fragments. The experiments reported here have focused on one of these sites which is located within the inverted repeat. DNA sequencing shows that variation in the copy number of a 24-base pair tandem repeat is responsible for the variability.

Materials and methods

Plant material. Two plant lines were established from the *plastome mutator* isolate from the Cornell University line of *Oenothera hookeri* str. Johansen (Epp 1973). As described by Johnson et al. (1988), these lines are called "Cornell-1" and "Cornell-2". In this paper, we abbreviate these as lines C_1 and C_2. Line D represents a cultivar of the same original Cleland strain of *Oenothera hookeri* str. Johansen which was maintained separately by Prof. W. Stubbe at the University of Duesseldorf for several decades.

* This paper is dedicated in memory of Barry K. Chelm

Offprint requests to: B. Sears

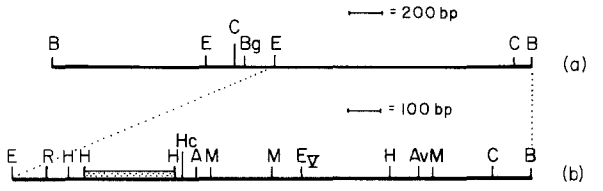

Fig. 1a, b. Fine structure map of the Bam12 fragment. The entire 3.0-kb fragment is shown in (a), while the 1.6-kb subcloned fragment is enlarged in (b). Restriction enzyme designations are as follows: B = *BamHI*, E = *EcoRI*, C = *ClaI*, Bg = *BglII*, R = *RsaI*, H = *HinfI*, Hc = *HincII*, M = *MaeIII*, A = *AluI*, Ev = *EcoRV*, Av = *AvaI*

Chloroplast DNA isolation and cloning. cpDNA was isolated as described by Johnson et al. (1988). The appropriate clones for line D were obtained by screening a shotgun library prepared from BamHI-digested cpDNA cloned into pBR322 (Maniatis et al. 1982). From lines C_1 and C_2, the desired fragments of cpDNA were removed from a preparative agarose gel and purified using the freeze and crush method (Smith 1980), followed by phenol extraction. The purified fragments were cloned into the BamHI site of pBR322 after alkaline phosphatase (Calbiochem) treatment of the vector (Maniatis et al. 1982).

A 1.6-kb EcoRI-BamHI fragment (Fig. 1b) was subcloned from each of the Bam12 clones by gel purification and insertion into pBR322 after removal of a 375-bp EcoRI-BamHI fragment from the vector. The remaining 4.0-kb DNA of the vector was gel-purified by electroelution onto a DEAE cellulose membrane (Schleicher and Schuell NA-45), following the protocol described in Schleicher and Schuell bulletin no. 364. The membrane was rinsed with water and then incubated for 4 h with 400 µl 1.5 M NaCl, 1 mM EDTA, 10 mM Tris pH 7.5 at 65 °C in a microfuge tube. The membrane was then discarded and the DNA was ethanol-precipitated.

M13 cloning and sequencing. Replicative form (RF) DNAs from M13mp18 and M13mp19 were prepared according to the triton lysis procedure described by Moore (1987). The 520-bp HincII-EcoRI fragment (Fig. 1b) from each of the three lines was cloned into the SmaI and EcoRI sites of M13mp18. To sequence the second strand of the variable region, the RsaI-BamHI fragment from line C_2 (Fig. 1b) was cloned into the SmaI and BamHI sites of the m13mp19 vector, while the 1.6-kb EcoRI/BamHI fragment from the lines C_1 and D was transferred into m13mp19 using sites for the same enzymes in the polylinker.

The preparation of single strand M13 DNA (Messing 1983) and dideoxy DNA sequencing (Sanger et al. 1977) were performed as outlined in the Bethesda Research Laboratories (Gaithersburg, MD) instruction manual on "M13 Cloning/ Dideoxy Sequencing", using the 17-bp M13 primer. For sequencing the larger inserts from lines C_1 and D which had been cloned into m13mp19, we used the extended Klenow technique described in the 1987 BRL publication, Focus 9(3).

Results

Detailed restriction mapping of the Bam12 fragment

One of the two variable regions in cpDNA of the three *Oenothera hookeri* lines was contained within the twelfth largest BamHI fragment (a 3.0-kb fragment referred to as Bam12) which mapped to the inverted repeat (Kaplan 1987; Johnson et al. 1988). We cloned this fragment from lines D, C_1 and C_2 and characterized it using restriction enzymes (Fig. 1a). The variation in fragment size was mapped to a discrete region, and our subsequent analyses utilized a 1.6-kb subclone representing the right half of each Bam12 fragment, defined by the sites for EcoRI and BamHI. Fine mapping of the subcloned fragments with a number of restriction enzymes narrowed the size variability to a region shown as a dotted box in Fig. 1b. Using the HincII endonuclease, the size variation between subclones from the three closely related plant lines can be seen clearly in Fig. 2. From digestions of the DNA with HinfI, the variable segment could be localized in a discrete region of about 220-bp in line D, 150-bp in line C_1, and 100-bp in line C_2.

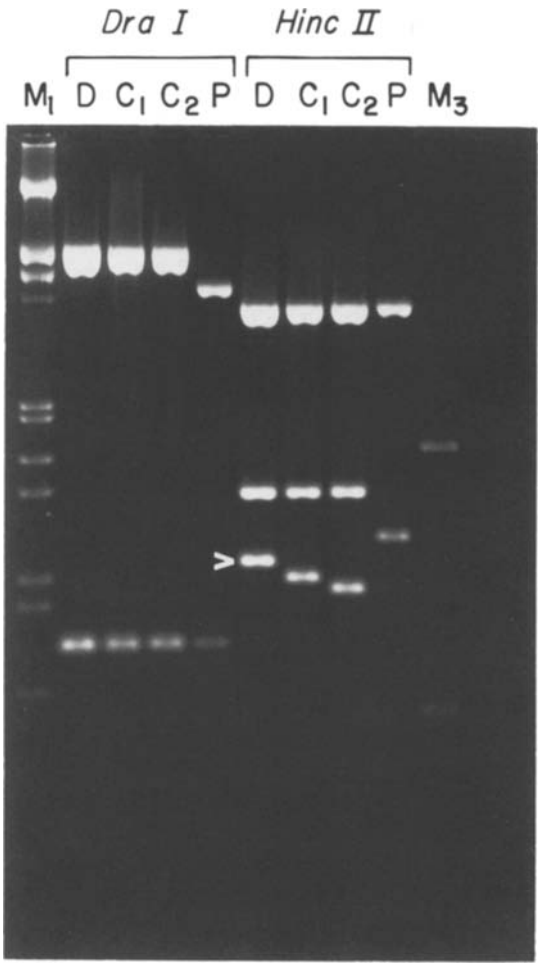

Fig. 2. Restriction digest patterns of plasmid DNAs carrying the 1.6-kb subcloned fragment from lines D, C_1 and C_2. Plasmid DNAs were digested with DraI or HincII and were separated by electrophoresis on a 0.8% agarose gel with DNA size standards (M_1 = lambda DNA digested with HindIII and EcoRI, M_3 = pBR322 plasmid digested with HinfI; P = pBR322 plasmid digested with either DraI or HincII as indicated). The *arrowhead* indicates the position of the variable fragment in line D

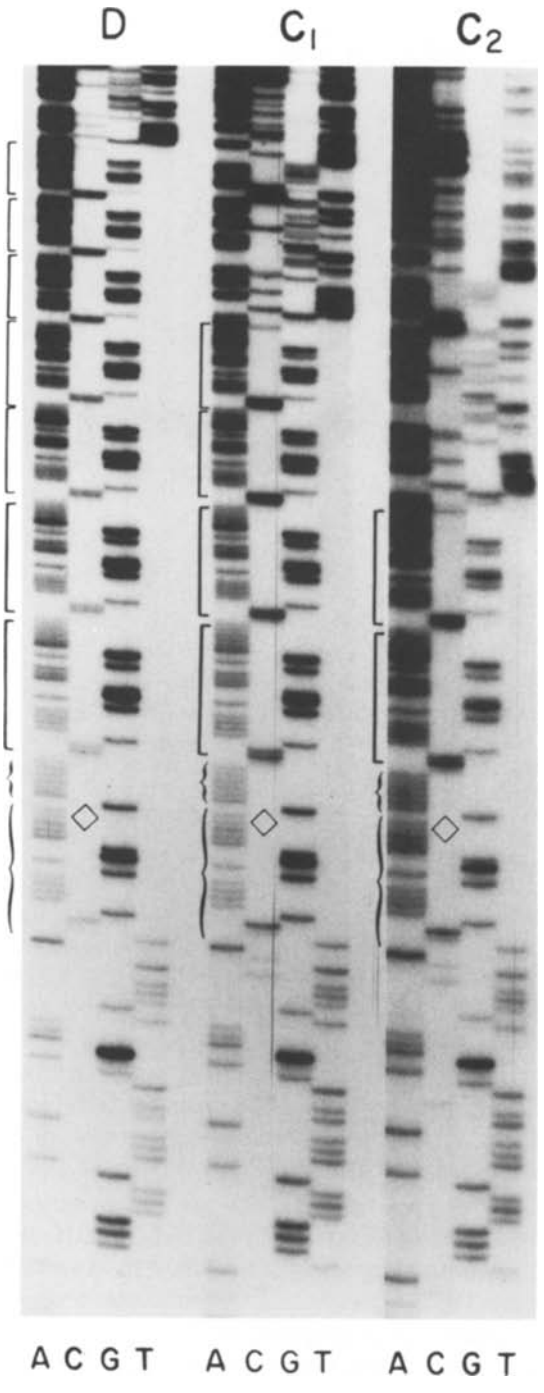

Fig. 3. DNA sequence of the variable region from lines D, C_1 and C_2. The 24-bp repeat segments are indicated with square brackets, while the 5′ near-repeat is indicated with two curved brackets. The left and right corners of a diamond symbol indicate the site of a one base-pair difference (presence of an A rather than a G) in the near-repeat which distinguishes it from the tandem repeats

The variability in fragment size is due to different numbers of copies of a 24-bp tandem repeat

The EcoRI/HincII fragment from the left end of the 1.6-kb subclones (Fig. 1b) including the variable region was subcloned into M13mp18 for sequencing from the HincII site towards the left, while the RsaI/BamHI or the EcoRI/BamHI fragment was cloned into M13mp19 for sequencing from the RsaI or EcoRI site towards the right. Comparison of the DNAs cloned from the three plant lines shows that they are identical in this region except that line C_2 has two copies of a 24-bp tandem repeat, line C_1 has four copies, and line D has seven copies (shown by square brackets in Fig. 3, and in Fig. 4, indicated as the boxed sequence labelled as 2). An additional nearly exact copy of the repeat having 23/24 identical nucleotides is found at the 5′ end of each repeat series (rounded brackets in Fig. 3, and in Fig. 4, indicated as the boxed sequence labelled as 1). As clearly illustrated in Fig. 3, the 100-200-bp stretch of DNA which contains the repeats has a very asymmetrical distribution of A-T base pairs. Although each repeat is 60% A-T, *all* of the adenines are on one strand and *all* of the thymines are on the complementary strand.

The 24-bp repeat segment and the surrounding sequences were compared with the analogous region of cpDNA from tobacco (Shinozaki et al. 1986) and spinach (Zhou et al. 1988). Although the 24-bp repeat does not exist in these other dicot cpDNAs, the surrounding area is highly conserved and an exact copy of the first 16-bp of the near-repeat is present in tobacco (with a single base pair change in spinach), in precise alignment with the sequence in *Oenothera* (Fig. 4). Figure 5 summarizes the sequence alignments in a diagrammatic format. A computer search of the entire tobacco cpDNA indicated that the 16-bp sequence shown in Figs. 4 and 5 contains the highest homology (14/14-bp) to the 24-bp segment. This region is within a large unidentified open reading frame (ORF) called ORF-1708 in tobacco and ORF-2131 in spinach. A 12/13-bp match of the same part occurs later in ORF-1708 (at tobacco cpDNA position 93,340), and the other end of the 24-bp segment has an exact match (10/10) in the spacer DNA between ORF-73 and ORF-74B (position 72,750) of tobacco.

The comparison of the cpDNA sequences in Fig. 4 indicates that upstream of the tandem repeats, base-pair substitutions and minor deletions have occurred with a low but equal frequency in both the evening primrose and spinach, as compared to the apparently more conservative cpDNA of tobacco. However, in the region downstream of the repeats, the spinach and tobacco sequences are 100% conserved, while the *Oenothera* sequence differs at 17% of the residues, due to base substitutions. None of the various changes between *Oenothera* and the other two dicots, including the base-pair substi-

```
S.o. 5'-T A T C T G A T T C   A G A A - - - A A G   A A C T T G C A T C
N.t. 5'-T A T C T G A T T C   A G A A G G G A A G   A A C T T G C A T C
O.j. 5'-T A T C T G A T T C   A G A A G G G A A G   A A C T T G t A T C

S.o.   A G T A T C T C A A   T T T C A A T T C A   A A C A T G G G T T
N.t.   A G T A T C T C A A   T T T C A A T T C A   A A C A T G G G T T
O.j.   A G T A T C T C A A   T T T C A A T T C A   A A C A T G G G T T

S.o.   T t A T T C A C A C   T C C A T G T T C T   G A G A A A T A T T
N.t.   T G A T T C A C A C   T C C A T G T T C T   G A G A A A g A T T
O.j.   T G A T T - - - - -   - - - - T a T T C T   G A G A A A T g T T

                ①
S.o.   T A c t A T C g G A   A A A G A G G A A A   A A A - - - - - - -
N.t.   T A T C A T C C G A   A A A G A G G A A A   A A A - - - - - - -
O.j.   T c T C A T C C G A   A A A G A G G A A A   A A g A A A A A A C
  ②
S.o.   - - - - - - - - - -   - - - - - - - - - -   - - - C G G A G T C
N.t.   - - - - - - - - - -   - - - - - - - - - -   - - - C G G A G T C
O.j. [C C G A A A A G A G   G A A A G A G A A A   A A A C]n a G A G T C

S.o.   T T T G T C T A A A   G A A A T G C G T T   G A G A A A G G G C
N.t.   T T T G T C T A A A   G A A A T G C G T T   G A G A A A G G G C
O.j.   T T T a T C T A A A   G c A A T G g G T T   G A G A A A G t G C

S.o.   A G A T G T A T A G   A A C C T T T C A A   C G A G A T A G T G-3'
N.t.   A G A T G T A T A G   A A C C T T T C A A   C G A G A T A G T G-3'
O.j.   A G A T G g A T A G   A g C C T T g C A A   g G A G A g A G g G-3'
```

Fig. 4. Sequence of cpDNA from *Nicotiana tabacum* (N.t.) beginning at base-pair 92,363 of Shinozaki et al. (1986), compared to the analogous sequence from spinach (S.o.) of Zhou et al. (1988) and *Oenothera hookeri* strain Johansen (O.j.). Sites of base substitution are indicated by small letters; regions of deletion are shown by dashes. In the *Oenothera* sequence, a 24-bp tandem repeat is enclosed in a *box* and is designated with the number 2. Depending on the plant line, two, four or seven exact copies of the 24-bp segment are present. Immediately 5' to the repeat is a near-repeat whic is also *boxed* and marked with a 1

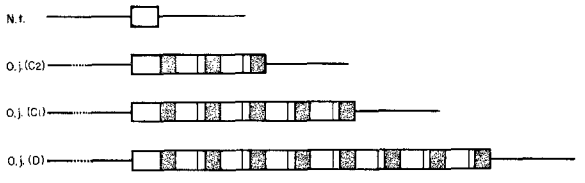

Fig. 5. Diagrammatic summary of a sequence comparison of the hypervariable region from the three *Oenothera* lines (C_2, C_1 and D), aligned with the analogous region from *Nicotiana tabacum* (N.t.). The *white box* indicates the first 16-bp; the remaining 8-bp are shown as a *stippled box*. The *bar* indicates the only nucleotide difference between the near-repeat and the tandem repeats

tutions, a 9-bp deletion, a 6-bp addition, and multiples of the 24-bp tandem repeat, interrupt or cause a frame-shift in the reading frame of ORF-1708. Each 24-bp repeat adds codons for the following amino acids to the reading frame: pro-glu-lys-arg-lys-glu-lys-lys.

Discussion

Recently, vom Stein and Hachtel (1988) compared cpDNAs between two species within the subsection *Munzia* of the genus *Oenothera*, and found two copies of a 24-bp tandem repeat in cpDNA from *O. odorata* in comparison to *O. berteriana* which had one copy. Although these repeat segments were also within the large inverted repeat, they mapped to a non-coding region within an intron of a gene which has been tentatively identified as coding for a subunit of an NADH dehydrogenase. This 24-bp sequence has no homology to the one which we have characterized. In *Chlamydomonas*, a region within the inverted repeat and purported to contain short, tandem repeats has been found to be highly variable in size in chloroplast transformation experiments using a particle gun (Boynton et al. 1988).

The 24-bp repeat segment that we have described from *Oenothera hookeri* strain Johansen is also found within the inverted repeat of the cpDNA. In the model proposed in Fig. 6, we show how recombination between tandem repeats present in opposite copies of the inverted repeat can result in variation in the copy number of the tandem repeat segments. As shown in Fig. 6a, an exact pairing and recombination between the inverted repeats would not affect the copy number of the 24-bp segment, but an imprecise alignment of these segments (Fig. 6b) followed by recombination could change the copy number in the two inverted repeats (Fig. 6c). If followed by a copy correction process, the number of 24-bp repeat segments within the inverted repeat could be altered (Fig. 6d). The near-repeat could also function in this capacity.

In support of this model, we note that intramolecular recombination has been shown to occur between the

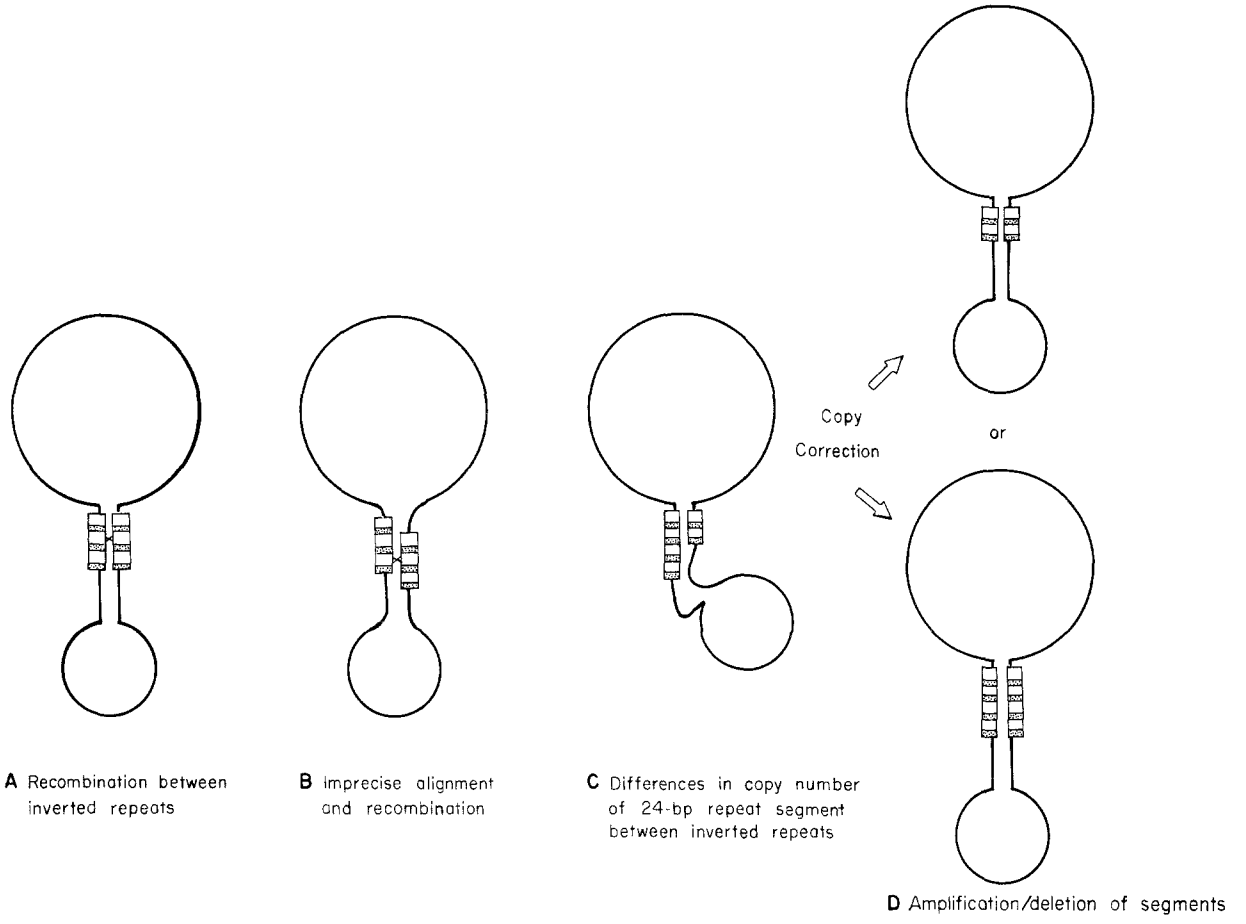

Fig. 6. Model for alterations in copy number of the 24-bp tandem repeat within the large inverted repeat of the cpDNA molecule

inverted repeats of cpDNA (Palmer 1983; Palmer 1985b). Furthermore, some type of copy correction mechanism or gene conversion must occur between the inverted repeats since point mutations (Erickson et al. 1986) and large deletions (Myers et al. 1982) are found to occur symmetrically in both copies of the inverted repeat. In the three plant lines that we have examined, several lines of evidence indicate that each is homoplasmic for one size of the Bam12 cpDNA fragment (and thus numbers of copies of the tandem repeats): 1) ethidium bromide-stained gels consistently show a double-molar fragment at this position (Johnson et al. 1988); 2) Southern blotting of cpDNA reveals only one type of banding pattern for each line, even when enzymes are used which cut the Bam12 fragment several times (Kaplan 1987; Johnson 1988); and 3) in our cloning experiments, we have recovered only a single type of Bam12 fragment from each plant line.

The three plant lines studied here are all descended from the Cleland collection of *Oenothera johansen* (Cleland 1972) which is now considered to be a subspecies or strain of *O. hookeri* (personal communication, W. Stubbe and W. Dietrich, University of Duesseldorf). Line D was maintained by Prof. W. Stubbe in Germany for about twenty years, while the cornell lines (C_1 and C_2) can be traced back through just a couple of generations to the same plant (Johnson et al. 1988). Thus, the DNA variability that we have found in the comparison of these three lines represents a very rapid alteration in the cpDNA molecule. These results suggest that cpDNA changes which appear to be due to insertion/deletion events may not be useful for establishing phylogenetic relationships.

In contrast, an assessment of base-pair substitutions in the entire sequenced region shows no changes among the three *Oenothera* lines. In the sequences upstream of the tandem repeats, approximately equal numbers of changes are found in the comparison of the evening primrose and spinach sequences with that of tobacco. However, downstream of the repeats, the cpDNA of the evening primrose is much more variable than are the cpDNAs from spinach and tobacco. Conceivably, the processes which initially created the repeats involved DNA rearrangements and/or other types of instability of the cpDNA molecule which preferentially affected the downstream region. Even within the repeat segment, an asymmetry of conservation can be seen. The left side of the repeat has a 16-bp exact homology with the analogous

region of cpDNA from tobacco, whereas the right side is totally dissimilar.

In most coding regions of cpDNA, sequence changes are infrequent (reviewed by Curtis and Clegg 1984; Zurawski and Clegg 1987), yet the variability that we have studied occurs in the middle of a very large open reading frame deduced from sequence comparisons with tobacco and spinach cpDNA (Shinozaki et al. 1986; Zhou et al. 1988). In contrast to tobacco and spinach, this region of *Oenothera* cpDNA contains three separate insertion/deletion events (a 9-bp deletion, a 6-bp addition, and multiple additions of the 24-bp tandem repeat), all of which maintain the reading frame. This observation supports the notion that the large open reading frame encodes a functional protein, although no mRNA has yet been detected (Zhou et al. 1988). Conceivably, the ORF encodes a protein which is less abundant than the proteins required for photosynthesis, which could render both the polypeptide and its message difficult to detect.

Although we do not yet know the role of the protein encoded by this ORF, we assume that the insertions occur in a domain of the primary amino acid sequence which is unessential for function of the protein, since no phenotypic differences have been noted between the plant lines. Supporting this idea is the fact that this domain is totally absent from the analogous ORF of *Marchantia* (Ohyama et al. 1986). Alternatively, the gene may not be expressed, or may be expressed only at a specific developmental stage. Identification and comparison of the protein products of the large open reading frame in the three *Oenothera* lines described here may help in the characterization of the product encoded by this gene.

Acknowledgements. The authors wish to acknowledge Dr. Barry Chelm for help with the DNA sequence comparisons and for stimulating discussions. The authors also thank Ellen Johnson for her assistance with many aspects of this project. K. B. and K. G. H. received partial support from the M.S.U. McKnight Photosynthesis Training Grant. The research program was supported initially by an M.S.U. Biomedical Research Grant, and subsequently by N.S.F. grant DCB-8502849. M.S.U. Agricultural Experiment Station Journal No. 12631.

References

Banks JA, Birky CW Jr (1985) Proc Natl Acad Sci USA 82:6950–6954

Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, Sanford JC (1988) Science 240:1534–1538

Clegg MT, Brown AHD, Whitfeld PR (1984) Genet Res 43:339–343

Cleland RE (1972) Oenothera: cytogenetics and evolution. Academic Press, New York, p 370

Curtis SE, Clegg MT (1984) Mol Biol Evol 1:291–301

Doebley JF, Ma DP, Renfroe WT (1987) Curr Genet 11:617–624

Epp M (1973) Genetics 75:465–483

Erickson JM, Rahire M, Rochaix JD, Mets L (1986) Science 228:204–207

Gordon KHJ, Crouse EJ, Bohnert HJ, Herrmann RG (1982) Theor Appl Genet 61:373–384

Johnson EM, Kaplan SA, Blasko K, Chin W-L, Sokalski M, Wolfson R, Sears BB (1988) Mol Gen Genet, in press

Johnson EM (1988) Physiological and molecular characterizations of plastids from a plastome mutator-induced mutant of *Oenothera hookeri* strain Johansen. PhD Thesis, Michigan State University, East Lansing, MI

Kaplan SA (1987) Characterization of chloroplast DNA from wildtype and mutant plastids of *Oenothera hookeri* str. Johansen. MSc Thesis, Michigan State University, East Lansing, MI

Kutzelnigg H, Stubbe W (1974) Sub-Cell Biochem 3:73–89

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning, a lab manual. Cold Spring Harbor Press, Cold Spring Harbor, New York, pp 545

Messing J (1983) Methods Enzymol 101:20–78

Metzlaff M, Boerner T, Hagemann R (1981) Theor Appl Genet 60:37–41

Moore D (1987) In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) Current protocols in molecular biology. John Wiley § Sons, New York, pp 1.7.1–1.7.7

Myers AM, Grant DM, Rabert DK, Harris EH, Boynton JE. Gillham NW (1982) Plasmid 9:321–324

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Cheng Z, Aota S-I, Inokuchi H, Ozeki H (1986) Plant Mol Biol Rep 4:146–175

Palmer JD (1983) Nature 301:92–93

Palmer JD (1985a) Annu Rev Genet 19:325–354

Palmer JD (1985b) In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 131–239

Palmer JD, Zamir D (1982) Proc Natl Acad Sci USA 79:5006–5010

Palmer JD, Jorgensen RA, Thompson WF (1985) Genetics 109:195–213

Salts Y, Herrmann RG, Peleg N, Lavi U, Izhar S, Frankel R, Beckmann JS (1984) Theor Appl Genet 69:1–14

Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74:5463–5467

Schmitz UK, Kowallik KV (1986) Plant Mol Biol 7:115–127

Scowcroft WR (1979) Theor Appl Genet 55:133–137

Sears BB (1983) In: Stadler Symposium 15:119–139

Shinozaki K, Yamada C, Takahata N, Sugiura M (1983) Proc Natl Acad Sci USA 80:4050–4054

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayshida N, Matsubayasha N, 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, Kata A, Tohdoh N, Shimada H, Sugiura M (1986) EMBO J 5:2043–2049

Smith, HO (1980) Methods Enzymol 65:371–380

Stein J vom, Hachtel W (1988) Curr Genet 13:191–197

Timothy DH, Levings CS III, Pring DR, Conde MF, Kermicle JL (1979) Proc Natl Acad Sci USA 76:4220–4224

Willey DL, Howe CJ, Auffret AD, Bowman CM, Dyer TA, Gray JC (1984) Mol Gen Genet 194:416–422

Zhou DX, Massenet O, Quigley F, Marion MJ, Moneger F, Huber P, Mache R (1988) Curr Genet 13:433–439

Zurawski G, Clegg MT (1987) Annu Rev Plant Physiol 38:391–418

Zurawski G, Clegg MT, Brown AHD (1984) Genetics 106:735–749

Communicated by R. W. Lee

Received May 10, 1988