Proliferation of Direct Repeats near the *Oenothera* Chloroplast DNA Origin of Replication

Barbara B. Sears, Lara L. Stoike, Wan-Ling Chiu¹

Department of Botany and Plant Pathology, Michigan State University

The spacer between the 16S and 23S tRNA genes of the chloroplast DNA has been implicated as an origin of replication in several species of plants. In the evening primrose, *Oenothera*, this site was found to vary greatly in size, with plastid genomes (plastomes) being readily distinguished. To determine whether plastome “strength” in transmission could be correlated with variation at oriB, the 16S tRNA-trnL spacer was sequenced from five plastomes. The size variation was found to be due to differential amplification (and deletion) of combinations of sequences belonging to seven families of direct repeats. From these comparisons, one short series of direct repeats and one region capable of forming a hairpin structure were identified as candidates for the factor that could be responsible for the differences between strong and weak plastome types.

Ample sequence variation allowed phylogenetic inferences to be made about the relationships among the plastomes. Phylogenetic trees also could be constructed for most of the families of direct repeats. The amplifications and deletions of repeats that account for the size variation atoriB are proposed to have occurred through extensive replication slippage at this site.

Introduction

In both chloroplast and mitochondrial DNAs, sites at which displacement loops (D-loops) have been mapped by electron microscopy are thought to contain the origins of replication for those molecules (Wolstenholme, Goddard, and Fauron 1979; Tewari 1988). In mitochondrial DNA (mtDNA) of different animals, the D-loop “control region” has been mapped to different sites, but in all cases, the D-loop is located in noncoding DNA adjacent to rRNA and/or tRNA genes (e.g., Lewis et al. 1994; Stewart and Baker 1994; Berg, Mourn, and Johansen 1995; Miracle and Campton 1995; Zardoya et al. 1995). In the control region, DNA length heterogeneity has been observed as a result of amplification and deletion of tandem direct repeats, occurring either on an evolutionary time scale (Lewis et al. 1994; Stewart and Baker 1994), or with a frequency that results in heteroplasmyn within an individual (Arnason and Rand 1992; Berg, Mourn, and Johansen 1995; Miracle and Campton 1995). The DNA variation at the control region has provided a useful genetic marker for population genetics and biosystematics, but no study has investigated whether the length mutations have a functional consequence for replication.

In angiosperms, chloroplast DNA (cpDNA) D-loops also seem to be located near rRNA or tRNA genes. Electron microscopic observations have indicated that one or two pairs of D-loops exist on the cpDNA, depending on whether the DNA molecule contains a single rRNA operon, as does pea (Meeker, Nielson, and Tewari 1988), or two rRNA operons located in opposite inverted repeats, as in tobacco (Takeda, Hirokawa, and Nagata 1992) and the evening primrose, *Oenothera* (Chiu and Sears 1992). In maize, petunia, soybean, and pea, two-dimensional electrophoresis and procedures of in vitro replication have been applied to map the origins of cpDNA replication (de Haas et al. 1986, 1987; Gold et al. 1987; Carrillo and Borgord 1988; Hedrick et al. 1993; Reddy et al. 1994). The various procedures often have not implicated the same sites, but all of the potential sites for the origins of replication, the spacer between the 16S rRNA gene and the 23S rRNA is most consistently indicated (Meeker, Nielson, and Tewari 1988; Chiu and Sears 1992; Takeda, Hirokawa, and Nagata 1992; Nielsen, Lu, and Tewari 1993; Reddy et al. 1994; Staub and Maligna 1994). Furthermore, in spinach chloroplasts, this site has been found to be bound preferentially to the thylakoid membrane, with a possible role in replication or organization and segregation of the nucleoid (Liu and Rose 1992).

This investigation was undertaken to determine if sequence differences at the origins of replication of the plastid genome (plastome) of *Oenothera* (Chiu and Sears 1992) might be responsible for the differential rates of plastid multiplication that have been observed for plastome types I–IV (reviewed by Chiu, Stubbe, and Sears 1988). In plants belonging to the genus *Oenothera*, biparental inheritance of the plastids is observed, with transmission through both the pollen and egg. Investigations using the electron microscope (Schötz 1954; Diers 1963; Meyer and Stubbe 1974) and DAPI-staining (Corriveau and Coleman 1990) have indicated that egg cells that carry different plastome types have approximately the same number of plastids, and pollen generative cells that contain different plastome types also have similar numbers of plastids and nucleoids. But, when one parent is kept constant, and different plastids are brought into the zygote from the other parent, the progeny will consistently show few plastids from the paternal parent (<1%), intermediate paternal transmission (15%), or high levels of paternal transmission (50% or more). Since their numerical input is similar, the relative abundance of the maternal and paternal plastids within the progeny is a measure of their com-

¹ Present address: Department of Biological Sciences, Florida International University.

Key words: *Oenothera*, replication slippage, direct repeats, origin of replication, cpDNA evolution, indels.

Address for correspondence and reprints: Barbara B. Sears, Department of Botany and Plant Pathology, Michigan State University, East Lansing, Michigan 48824. E-mail: sears@pilot.msu.edu.


© 1996 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

850
petitive multiplication abilities. Some plastids transmit well in most nuclear backgrounds, and are referred to as having “strong” plastomes (I, III), while others are transmitted to fewer progeny, and hence are considered to have “medium” (II) or “weak” (IV) plastomes (Schütz 1974, 1975; Chiu, Stubbe, and Sears 1988). The weak plastomes are thought to have persisted because they are the only plastome types that are compatible with the nuclear genotypes of certain species of *Oenotheca*, with which they have co-evolved (Stubbe 1964). All of the *Oenotheca* plastome types have restriction fragment length polymorphisms (RFLPs) relative to each other and scattered throughout the cpDNA (Gordon et al. 1982).

Since transmission strength is intrinsic to the plastome, a feature of the cpDNA or a plastome gene product must be responsible for the differences among plastome types. Studies of the *iojap* strain of maize (Walbot and Coe 1979; Han, Coe, and Martienssen 1992) and the *albosintras* mutant of barley (Börner, Schumann, and Hagemann 1976; Hess et al. 1992) point to the likelihood that a noncoding region may be responsible for controlling plastome multiplication. In these mutants, the plastids lack ribosomes, and are thus incapable of translating cpDNA-encoded mRNAs. Nevertheless, the cpDNA and the plastids themselves are perpetuated and segregated to daughter cells. These data suggest that the essential components for plastid multiplication are encoded by nuclear genes, and indicate that a noncoding component of the plastome, the origin of replication, is perhaps the most likely candidate for the locus responsible for modulating the efficiency of plastid multiplication.

To test the hypothesis that differences at the origin of replication might be responsible for the intrinsic differences in multiplication of the different plastome types, we first located the positions of D-loops on the *Oenotheca* cpDNA molecule of a weak (IV) and strong (I-D) plastome type. The D-loops were mapped by identifying and measuring their locations on cpDNA restriction fragments that had been spread, shadowed, and visualized by electron microscopy (Chiu and Sears 1992). Two pairs of D-loops were observed, as indicated in figure 1A, and their location and frequency of occurrence were identical for the two plastome types. Although physical mapping of D-loops is one of the more rigorous ways to identify an origin of replication, the localization is not very precise, since the spreading procedure can stretch linear and high molecular weight DNA more than the small, circular DNA that is included as a control. Thus, our subsequent sequence comparisons at the sites tentatively mapped as the replication origins were intended to search for sequences that would definitively indicate the presence of an origin, and to assess the possibility that local DNA sequence variation could be responsible for differences in the efficiency of initiation of cpDNA replication among the strong and weak plastome types.

### Materials and Methods

#### Plant Material

The *Hookeri* and *Johansen* strains each carry their native plastomes, designated I-HK or I-D, respectively. “D” is used as the designation for this particular *Johansen* line because it was maintained at the University of Dusseldorf by Prof. Wilfried Stubbe. Other plastomes were transferred into the *Johansen* nuclear background by Prof. Stubbe from *O. atroviolens* (plastome IV), *O. suaveolens* (plastome II), or *O. lamarkiana* (plastome III). The plant materials, provided as seed by Prof. W. Stubbe, were maintained by self-pollination.

#### DNA Analysis

CpDNA was isolated according to Sears, Chiu, and Wolffson (1995). Enzymes for restriction analysis and cloning were obtained from Gibco-BRL. The third-largest *Sal* I fragment from each plastome was isolated by phenol-chloroform extraction of the band from a low-melting agarose gel following electrophoresis in TAE buffer (Sambrook, Fritch, and Maniatis 1989). These 18-kb DNA fragments were cloned into the *Sal* I site of pRL498 (Elhai and Wolk 1988). From those clones, the largest *Eco* RI fragments (approximately 5 kb) were isolated and subcloned into the *Eco* RI site of pBR328. Subsequent electrophoresis of restricted DNA was performed in TBE buffer (Sambrook, Fritch, and Maniatis 1989).

A smaller subclone was produced from the I-D clone by digesting it with *Eco* RI and *Bss* HII, filling in the ends with the Klenow fragment of DNA Polymerase I, and then cutting it with *Eco* RI. The fragments thus obtained were ligated with pIC20H (Marsh, Erfle, and Wykes 1987) that had been digested with *Eco* RI and *Sma* I. A plasmid containing the smaller *Eco* RI-*Bss* HII subfragment was identified by the size of its insert and was named pOjD-BEI. Deletion derivatives of pOjD-BEI were obtained using exonuclease III (Henikoff 1984) after digesting with *Sst* I and *Eco* RI. The ends of the resulting fragments were made blunt with mungbean nuclease and rejoined. Other deletion derivatives were obtained by cutting the original plasmid with either *Hinc* II or *Bam* HII and religating. The inserts in these subclones were sequenced as double-stranded template using the Sequenase DNA sequencing kit (United States Biochemical Corporation) with the denaturation procedure of Hsiao (1991). The sequence data thus generated was used to construct primers to sequence the same region from clones of the *Eco* RI fragments from the other plastome types. Both strands were sequenced from all of the clones.

#### Computer Programs Used for Sequence Analysis

EDITBASE, provided by Niels Nielsen (Purdue University), and the REPEAT feature of the GCG Program of the Wisconsin Computer Group were used to search for repeated sequences. Many direct repeats were identified by visual inspection of the DNA sequence data. Regions capable of forming secondary structures were found using STEMLOOP of the GCG Program, with the specifications of a minimum stem size of 6,
minimum bonds/stem of 12, and maximum loop size of 20. Relationships of repeat family sequences were assessed by PAUP 3.1 (Swofford and Begle 1993) using MacClade (Maddison and Maddison 1992) for data entry with a Macintosh IIsi. The branch-and-bound algorithm of the bootstrap program was used, with 100 replicates.

**Results**

**One D-Loop Region is Proximal to a Highly Variable Region of the *Oenothera* Plastome**

The third largest *Sal* I fragment from *Oenothera* plastome types I and IV contains two D-loop sites (Chiu and Sears 1992; fig. 1A). This fragment was cloned from...
the 16S rRNA gene and the trnl gene (figs. 2, 3, and 4). The alignments of figure 3 show data from the five Oenothera plastomes and the orthologous region from tobacco cpDNA (Shinozaki et al. 1986). Gaps were introduced to improve the alignments, with asterisks at one border of most gaps indicating ambiguity in the position of the gap. The smallest 16S rRNA-trnl spacer is found in plastome III, and is twice the size of the equivalent region in tobacco. All of the size variability is due to the presence of varying numbers and combinations of sequences that can be subdivided into seven families of repeats (fig. 4).

Each repeat family contains repeat variants, which are the smallest units that have been independently deleted or amplified during the evolution of this region of cpDNA. Larger repeat units can be discerned because the repeats generally are found in a recurring context, as shown in the overview provided by figure 5. Thus, a member of the F repeat family tends to follow a D-repeat; E-repeats always follow F-repeats, and the G-repeats almost always follow one or more E-repeats.

Several of the indels that differentiate members of the same repeat family resulted from the addition of one or more bases in a simple repetitive sequence, such as the stretch of cytosines in repeats A6 and F5 or the stretch of thymines in repeats A6 and A7. Other expansions resulted from small duplications, such as the AGC duplication at the left end of the F2 repeat and the ATTAG duplication at the left end of the G7 repeat. In repeat families A, B, D, F, and G, one or two variants appear to have resulted from deletion events.

Evolutionary Relationships of the Plastomes Inferred from Parsimony Analysis

The entire variable region was analyzed by PAUP with the Nicotiana tabacum sequence as the outgroup. When each position was counted as a character and gaps were counted as missing characters, no resolution was obtained; rather, 21 equally parsimonious trees were produced. However, when each gap was counted as a single character, the number of characters was reduced to 268, and an exhaustive search produced a single most-parsimonious tree, in which the length of the shortest tree was 84. The bootstrap values are shown in figure 6A. Plastomes III and IV branch away from the other plastome types early; the two plastome I types are closely related; and the intermediate position of plastome II between plastome I and IV is well supported. The early divergence of plastome IV would have been predicted from the existence of a number of unique repeat variants, A6, A7, C1, F3, F5, G1, G4, and G5, in its oriB region, even though all but two of these were excluded from the PAUP analysis because they exist at positions where gaps occur in the alignment with the other plastome types. In this treatment, the consensus position of plastome III is only supported in 17% of the bootstrap trees, so its phylogenetic position remains uncertain. The apparent basal divergence of plastome III (fig. 6A) is probably a consequence of overcounting the number of gaps relative to the other Oenothera plastomes. For the PAUP analysis of figure 6A, each gap in every plas-

cpDNA isolated from five different Oenothera plastomes (II, III, IV, and the Hookeri and Johansen strains of plastome I); the positions of the D-loops (oriA and oriB) are shown in figure 1B. Because restriction analysis indicated that the internal EcoRI fragment containing oriB was highly variable in size, it was subcloned and mapped in more detail. As shown in figure 2, the variable region is contained on a discrete subfragment.

DNA Sequence Analysis Shows that Variability Near oriB is Due to the Proliferation of Direct Repeats

Manual sequencing revealed the existence of a complex array of direct repeats in the spacer between

**Fig. 2**—Restriction digestion patterns of the cloned EcoRI fragment from several plastomes. Plasmids containing DNA from plastomes I-D, IV, I-H, II, and III were digested with EcoRI, Pvu II, and Xho I. Molecular weight markers (M) consisted of lambda DNA digested with EcoRI and HindIII. The fragment marked v is the pBR328 vector; fragments marked a, b, and c are indicated in the diagram in figure 1, panel B.
Fig. 3.—Sequence alignment of the cpDNA 16S rRNA-rnl spacer from Oenothera plastomes II, III, IV, I-Düsseldorf strain (I-D), and I-Hookeri strain (I-Hk) with the sequence from Nicotiana tabacum (Nta) (Shinozaki et al. 1986). Base substitutions are denoted with lower case letters; gaps in the alignments are indicated with hyphens (-); asterisks (*) indicate border sites at one end of a gap, where a perfect alignment would also be possible for the number of starred nucleotides from the sequence positioned at the other end of the gap. The presence of repeats is noted above the appropriate sequence, with the letters A-G and a number. The number of nucleotides in each line was varied in order to keep repeat family sequences intact. The sequences have been assigned the following GenBank accession numbers: U41045 (IV), U41043 (II), U41042 (I-D), U41046 (I-Hk), U41044 (III).

Evolutionary Relationships Within the Repeat Families Inferred from Parsimony Analysis

In the A, B, C, and G repeat families, a single member of the repeat family is present in all of the Oenothera plastomes. According to principles of parsimony, a shared trait such as this probably reflects an ancestral character (Maddison and Maddison 1992). A similar conclusion would be reached from a cursory comparison of the outgroup (tobacco) sequence with each family of repeats: in each case, the ubiquitous repeat (AI, BI, C, E, FI, G) is the one that most closely resembles the tobacco sequence (fig. 4). To test this hy-
For repeat families A, B, C, and G, relationships could be inferred, as shown in figure 6B (top), 6C, 6D, and 6E (top), respectively. Tree lengths were obtained using the exhaustive search algorithm: the length of the shortest tree for repeat family A was 17; for repeat fam-

Figure 3.—Continued.

thesis, PAUP was applied to each repeat family, as it has been for mammalian SINES (Zietkiewicz and La-

buda 1996), with each repeat unit entered as a taxon, each nucleotide defining a character, and each gap of two or more bases also defining a single character.
**Fig. 4.**—Repeat families located between the 16S rRNA gene and the rml gene on the Oenothera (Oen) cpDNA molecule. In the analogous region of cpDNA from *Pisum sativum* (Psa), *Glycine max* (Gma), *Spinacea oleracea* (Sol), *Daucus carota* (Dca), and *Nicotiana tabacum* (Nta), a single copy of each type of sequence is present. For the shorthand designation, each core sequence in tobacco has been assigned a capital letter A-G; in other species, subscripts indicate a sequence that is slightly different; in *Oenothera*, numbers following the capital letter designate different members of each repeat family. Variations from the first sequence listed for *Oenothera* are given as lower case letters representing the nucleotides A, T, G, or C. The order of the listings within each repeat family roughly reflects similarity of sequence. The sequences from the other dicots were obtained from GenBank through the GCG fetch program: *P. sativum* (accession numbers M98273 and M37430), *G. max* (accession number M37149), *S. oleracea* (accession number M21453), *D. carota* (accession number X73670), and *N. tabacum* (accession number Z00044).
**A. Abbreviated sequence:**

| Psa | Ap | Bp | CpDp |
| Gma | Ag | Bg | CpDg |
| Sol | At | Bt | CpDx |
| Dca | Ad | Bc | CdDd |
| Nta | A | B | C D |
| IV | A1 | B1 | B2 | C | D | F2 | E | G |
| I-D | A1 | B1 | B2 | C | D | F2 | E | G |
| I-Hk | A1 | B1 | C | E | G |
| III | A1 | B1 | C | B3 | C3 |

**B. Extremely abbreviated sequence:**

| Nta | AB | CD | FEG |
| IV | ABCD | EFG | A | FE | GEGAFEG |
| II | ABCD | EFG | GAF | GEGAFEG |
| I-D | ABCD | EFG | GAF | GEGAFEG |
| I-Hk | AB | C | CD | EFG | GAF | GEGAFEG |
| III | AB | C | BCC | EFG | GAF | GEGAFEG |

**FIG. 5.—Shorthand sequence of cpDNA region between 16S rRNA gene and trnl intron. (A) Alignments of cpDNA sequences from five Oenothera plastomes and other dicots, using the code and abbreviations of figure 4. (B) The sequences in panel A are further condensed by listing only the repeat family letters. The last line lists the basal sequence that probably existed in the ancestor of the Oenothera plastomes.**
FIG. 6.—Phylogenetic relationships inferred from oriB sequences. Unless otherwise indicated, bootstrapping was performed, using the branch-and-bound search algorithm, with *N. tabacum* designated as the outgroup. Fifty percent majority-rule consensus trees are shown; numbers indicate the frequency of occurrence of that branch in 100 replicates. Abbreviations are as in figure 4. (A) Phylogenetic relationships among the five plastome types, based on the entire oriB sequence, with each gap counted as a single character. (B) Relationships among members of repeat family A, based on a comparison of all of the family members (top), or only those present in plastome I-D (middle), or only those present in plastome IV (bottom). (C) Relationships among members of repeat family B. (D) Relationships among members of repeat family C. (E) Relationships among members of repeat family G. Top tree is a consensus of three trees retained in a branch-and-bound search (tree length = 18) conducted by PAUP to assess relationships among all members of repeat family G. The position of repeat G7 was the only variable feature among the trees. Bootstrapping was used to assess the relationships among the G-repeats present in plastome I-D (middle), or those present in plastome IV (bottom).

have a monophyletic origin, whereas repeat sequences that are found only in one plastome type are likely to be less related to the other derived repeats than they are to the original, ancestral repeat. Consideration of these subsets had the potential of revealing homoplasies in the inclusive analysis. Plastome types I-D and IV were chosen for this analysis because they contain the largest number of different repeat units within each family. For repeat families A and G, PAUP produced well-supported branches on the plastome I and plastome IV repeat trees; for repeat family F, simultaneous divergence of the repeats was once again indicated. For the A-repeat family, the position of A2 relative to the other repeats differs between the two trees (fig. 6B, middle and bottom), but we do not believe that the A2 repeat sequence is a homoplasy, because it occurs in exactly the same sequence context *(G6-A2-B1-C)* in both plastomes I and IV. In our opinion, the inclusive top tree in figure 6B seems to best reflect the phylogeny of the A-repeats, since similarities in position suggest a common origin for repeats A2 and A4, and shared sequence variations in the A3, A4, A5, and A6 repeats, and the A6 and A7 repeats also support a common origin.

For the G-repeat family, the branching of the G6 and G7 repeats differs between the plastome I tree and the plastome IV tree. We interpret their early branching
in the plastome I tree to indicate that duplications of the ubiquitous G6 sequence occurred in that lineage, with one duplication either coincident with or followed by an event that duplicated the first five bases to give repeat G7, and the other duplication coincident with or followed by the insert shared by repeats G8 and G9, which diverged after another duplication. The G6 and G7 repeats arose before the divergence of plastome IV, but in that plastome they were not duplicated further. Rather, the G4 and G5 repeats arose from duplication and divergence of the ancestral repeat, most likely represented by the ubiquitous G repeat or by the plastome IV-specific G1.

The 16S rRNA-trnl spacer of plastome III is quite different from the other plastome types: it is the smallest, and it contains only a single copy of a repeat from family F. The small size could indicate that plastome III experienced the fewest repeat amplifications, but we believe that the diminutive size more likely resulted from deletions during the evolution of this plastome type. The complete absence of sequence D, which is found in other dicots, indicates that one or more deletions occurred in the evolution of plastome III. Furthermore, two repeats that are unique to plastome III, B3 and G5, are clearly deletion variants in their respective repeat families. The B3 repeat probably arose during the deletion of the D1-F2-E-E-G6-A2 segment and the left half of B1. Similarly, the C3 repeat may have been created by the deletion of adjacent repeats D2-F2, leaving only the first two bases of D2 as a footprint, attached to the right side of C1. Thus the ancestral sequence that we propose at the bottom of figure 5 contains several segments that are shared among all of the other plastomes and absent from plastome III.

Structure-Function Relationships: Correlation Between Sequences and Plastome Strength

Although substantially different in size, the most similar plastomes in terms of sequence are from the Johansen and Hookeri strains of O. elata. Plastomes from both strains (I-D and I-Hk) are classified as plastome type I (Stubbe 1989), and both are strong in terms of their transmission in crosses (Chiu, Stubbe, and Sears 1988). Although the Hookeri plastome is about 210 bp smaller than is the Johansen plastome in this region, they share a number of complex sequence arrangements, and the unique repeat family variants A5 and C2 (fig. 4). The absence of repeat B2 and the repeat series D1-F2-E-E-G6-A2-B1 in the Hookeri plastome is probably a consequence of deletions, since those segments are present in the highly divergent plastome IV. Two other variants, A3 and G2, are present in the plastomes of Johansen, Hookeri, and also plastome II from O. suaveolens, which is considered to be a fairly closely related plastome type, according to its plastome-genome compatibility traits (Stubbe 1989). The sequence from plastome III was compared with those from Hookeri and Johansen, because it also is a strong plastome type. From this comparison, only one continuous segment in the primary sequence is a candidate for enhancing the multiplication of those plastids: the repeat series E1-E-

![Figure 7](https://example.com/fig7.png)

**Fig. 7.—Patterns of inverted repeats in the 16S rRNA-trnl spacer.** The position of sequences that can form stem loops are indicated by symmetrical points on either side of the diagonal that represent the control region sequence from each of the five plastomes. Since a maximum loop size of 20 was specified, all inverted repeats lie close to the diagonal of the comparison of forward and reverse sequences from the five Oenothera plastomes. The classes of stem loops of highest stability are outlined with different shapes, as noted in the key on the figure.

G4-E-G6-E is shared by those three plastomes in the approximate middle of the spacer (fig. 5).

The direct repeats are the predominant feature in the DNA of this region, but the STEMLOOP program of GCG indicated that many inverted repeats exist in the 16S rRNA-trnl spacer. The FOLDRNA program of GCG was also used to analyze the region. It revealed complex secondary structures for all of the plastomes; however, no structure was unique to the strong or weak plastomes. Furthermore, it is not likely that the complex secondary structure would exist in vivo, since such a large segment of DNA is unlikely to be single-stranded. In contrast, stem loops are composed of adjacent inverted repeats, representing secondary structures that could form if short regions of DNA were melted or transiently single-stranded. An overview of the locations of potential stem loops is provided in figure 7, and the hairpin sequences are shown in detail in figure 8, with the strength of the hairpins noted in the figure legend. The number of stem loops possible in each Oenothera plastome was generally proportional to the size of the oriB region, ranging from 46 for plastome III to 148 for plastome I-D. The strongest potential hairpins involve the B repeats at the left side of the region (fig. 8A), the F-E-G repeat at the right end (fig. 8E), and the F-repeats (fig. 8B) that are scattered throughout the region.

Although plastomes III and I-Hk lack repeat B2, which is a part of the three strongest secondary structures in the other plastomes, figure 8A shows that a hairpin that is almost as strong could form between sequences of the A1 B1 C repeats present in those plastomes at
the same position. The B-hairpins and many of the F-hairpins are shared by the strongest and weakest plastomes, I-D and IV, and, thus, most of the secondary structures are not likely to be responsible for multiplication differences. However, we note that the weak hairpin formed by the repeat series G-E-G is uniquely absent from plastome IV, and varies in number among the plastomes, with the medium strong plastome II having only two copies. Thus, this particular secondary structure is a candidate for the factor that causes differential plastome multiplication.

Discussion

The presence of tandem direct repeats in the oriB region of five Oenothera plastomes and the RFLPs that result from the amplification or deletion of these repeats are reminiscent of the animal mitochondrial DNA control regions. The existence of an extensive array of both direct and inverted repeats may be critical to the function of this region as a replication origin and/or membrane attachment point of the Oenothera cpDNA. Conceivably, the proliferation of the repeats and the associated secondary structures enhanced the function of this site in the Oenotheras, whereas in maize and soybean, other regions are thought to serve the role of cpDNA origin (Carrillo and Bogorad 1988; Hedrick et al. 1993).

Initially, this study was undertaken to collect data relevant to the hypothesis that differences at the plastome replication origins affect the initiation of replication of the plastome and, consequently, the speed or efficiency of cpDNA replication and the ability of the chloroplasts to multiply. Although the oriB region was found to be quite polymorphic, the size of this region does not correlate to relative transmission strength of the plastome types since the strongest plastomes (I-D, I-Hk, and III) have large, medium, and small spacers, and plastome IV, which is the weakest plastome type, has the second largest spacer. In the primary sequence (fig. 3), only a centrally located segment, containing the

\[ \Delta G = -19.5 \text{ kcal/mol}. \]

(B) Stem loops possible in the F-family of repeats. The F1 repeat is found in all of the plastomes \[ \Delta G = -17.7 \text{ kcal/mol} \]; F2 occurs in plastomes I-D, I-Hk, II, and IV \[ \Delta G = -15.8 \text{ kcal/mol} \]; F3 and F5 are found only in plastome IV \[ \Delta G = -9.5 \text{ and } -18.0 \text{ kcal/mol respectively} \]. (C) A highly conserved stem loop \[ \Delta G = -7.4 \text{ kcal/mol} \] shared by cpDNAs located about 90 bases from the end of the 16S rRNA gene, to the left of the region of repeat proliferation (the pre box in fig. 7). Species abbreviations as in figure 4. (D) The G6-E-G6 stem loop. This hairpin \[ \Delta G = -2.8 \text{ kcal/mol} \] could form in plastomes I-D, I-Hk, II, and III. Hairpins of less stability could form in those sites where a G7, G8, or G9 repeat is present instead of one of the G6 repeats. (E) The F-E-G stem loop at the right end of the oriB spacer in plants abbreviated in panel C. Carrot (Dca) and tobacco (Nta) have the identical sequence, except for the absence of a base (shown as a -) in carrot cpDNA. The Oenothera stem-loop has a \[ \Delta G = -16.3 \text{ kcal/mol} \] short stretch of repeat G that was obtained only from plastome I-D (refer to fig. 3). This sequence is likely to be shared by all of the Oenothera plastomes because it contains exon 1 of the trnL gene; additionally, an oligonucleotide at the right end of G was used as a primer for sequencing in the reverse direction, and the smallest bands on the sequencing gel had the same mobility from the various plastomes.
prior to the next slippage event, resulting in families of related repeats (figs. 3 and 5), with different repeats being differentially amplified. Slippage of the mother strand during replication could have caused the deletions that eliminated large segments of plastome III and plastome I of the *Hookeri* strain, such as the second large gap in the sequence alignment that is conserved in all of the other plastome types. During the evolution of the subsection *Oenothera*, a huge number of slippage events and mutations must have occurred to result in the divergent sequences observed here. Our findings at the oriB site contrast dramatically with our previous comparisons of these same plastome types at the rpl16-rpl14 and rpl14-rps8 intergenic spacers, where only six informative sites were found in 472 bp, three of which were base substitutions (Wolfson, Higgins, and Sears 1991).

**Acknowledgments**

We thank Dr. Helmut Bertrand for the use of his laboratory facilities for some of the sequencing and for helpful discussions, and Gabi Schewe for technical assistance. We also thank Drs. José Panero, Tao Sang, and two anonymous reviewers for their critical reading of the manuscript, and Dr. David Jarrell for assistance with PAUP. This work was supported by grant DCB-9019488 from the National Science Foundation.

Note added in proof: After the completion of this work, a size variant of oriB, which is equivalent in size to the plastome IV locus, was found to be more typical for plastome II.

**Literature Cited**


Amylase Promoter Structures in Primates

The Structure and Sequence of the New World Monkey Amylase Promoter

We determined the nucleotide sequence of the squirrel monkey NWM-0.3 fragment to verify the structure predicted from its size and hybridization characteristics (fig. 3). As predicted, this promoter did not contain the γ-actin pseudogene (fig. 4). Comparison of the amylase promoter sequences downstream of the pseudogene insert showed that the squirrel monkey amylase promoter is 93% identical to the human AMY2B promoter and 73% identical to the mouse pancreatic amylase promoter (fig. 4C). The mouse pancreatic amylase promoter also does not contain retroviral inserts (Osborn et al. 1987). The structural similarity between the New World monkey and murine pancreatic amylase promoters suggests that an amylase gene without a γ-actin pseudogene represents the mammalian precursor structure.

Identification of Primate Amylase Genes Associated with Retroviral Inserts

To test for the presence of the amylase-associated retrovirus, the gag/exa2 primers were used to amplify a 1.6-kb fragment indicative of an AMY1-like structure. The AMY1-1.6 fragment was amplified from human genomic DNA and the structure was verified by Southern hybridization with AMY, ACT, and LTR probes (fig. 5). When the nonhuman primate species were tested, the AMY1-1.6 fragment was amplified from three ape species (chimpanzee, gorilla, and orangutan), but was not amplified from Old World monkeys (seven species) or New World monkeys (three species). The AMY1-1.6 fragment includes proviral 5' LTR and gag sequences. To test for the presence of the 3' end of the retrovirus, we utilized a primer from the env region paired with a primer upstream of the inserts (a2a). The a2a/env primer pair amplified a 0.7-kb fragment from human and ape genomic DNA, but not from the Old World and New World monkeys (table 2; data not shown). This result is consistent with the previous pattern of amplification observed with the gag/exa2 primer pair, and indicates that the apes, but not the more distantly related monkey species, contain an AMY1-like gene that is associated with a complete endogenous retrovirus.

An AMY2A-like Gene is Present in the Apes

To check for the presence of an AMY2A-like gene, we paired the a2a flanking primer with the amylase exonic primer exa2 to amplify a 1.4-kb fragment that includes the γ-actin pseudogene and retroviral LTR characteristic of this gene structure. The AMY2A-1.4 fragment was amplified from human DNA (fig. 6, lane 1) and its structure was verified by hybridization with AMY, ACT, and LTR probes (data not shown). The pattern of amplification of the AMY2A-1.4 fragment followed the results of amplification of the other fragments containing retroviral sequences: the fragment was detected in the genomes of apes (chimpanzee, gorilla, and orangutan), but was not observed in Old World monkeys (five species) or New World monkeys (three species) (fig. 6). Thus, the results of amplification with the gag/exa2, a2a/env, and a2a/exa2 primer pairs were consistent with the interpretation that insertion of the endogenous retrovirus into the primate amylase cluster occurred sometime after the divergence of the New and Old World monkeys and before the divergence of the apes.

The Structure and Nucleotide Sequence of a Novel Amylase Promoter in Old World Monkeys

In place of the expected AMY2A-1.4 fragment, the a2a/exa2 primer pair amplified a 1.0-kb fragment from the five Old World monkey species (fig. 6, lanes 5–9). Southern blot analysis demonstrated that this OWM-1.0 fragment hybridized to the AMY and ACT probes, but not to the LTR probe (data not shown). Based on its size and hybridization characteristics, this fragment was hypothesized to represent a novel amylase gene structure containing a truncated γ-actin pseudogene and no retroviral insert. The nucleotide sequence of the Japanese
sequences abbreviated as $E1-E7-E6-E$ (figs. 4 and 5), is shared among the strong plastome types (I-D, I-Hk, and III) and absent from the weaker plastome types (II and IV). Assessment of additional plastomes with different transmission traits will allow this correlation to be examined further.

In cpDNA of *Chlamydomonas* and pea, potential stem loops have been noted in regions to which the origins of replication have been mapped (Wu et al. 1986; Nielsen, Lu, and Tewari 1993). As indicated in figures 6 and 7, in the *Oenothera* plastomes, many short inverted repeats exist in this region and have the potential of forming hairpins, and several of the sequences contain repeats of similar sequence to those that could form hairpins in pea cpDNA (Nielsen, Lu, and Tewari 1993). The hairpin shown in figure 8C is identical in all of the *Oenothera* plastomes examined, and is relatively conserved in other land plants. Thus, it is likely to have functional importance, either for *oriB* or for the processing of the transcript. Since the sequence is identical in all of the *Oenothera* plastomes, it is not a candidate for the factor that is responsible for differential multiplication of the evening primrose chloroplast.

Several of the stem loops shown in figure 8 are shared by all of the *Oenothera* plastomes. None of the hairpins is unique to the strong plastomes (I-Hk, I-D, and III) and absent from the medium-strong and weak plastomes. However, the hairpin formed by repeats G-E-G (fig. 8D) is completely absent from plastome IV, while it is present in two to five copies in the other plastomes (fig. 7). This potential secondary structure and the short primary sequence mentioned previously are thus candidates for conferring a difference in transmission abilities to the plastomes.

Given the imprecision of D-loop mapping (due to the differential stretching of linear DNA molecules and the smaller circular size standards), the possibility also exists that the replication origins are located adjacent to, but not at, the sites indicated in figure 1B. In that case, the highly variable region between the 16S rRNA and trnl genes probably be located between the two origins. Such a location could render it highly susceptible to replication slippage since it would be single-stranded during the initial stages of replication. If that is the case, the presence of complex repeats in the region may not be functionally meaningful in terms of controlling cpDNA replication. Rather, their occurrence could indicate that this region lacks functional constraints and thus is highly mutable with no deleterious consequence to the cpDNA.

Regardless of the function of the region we have designated as *oriB*, the sequence data proved useful for the analysis of phylogenetic relationships among the five plastomes analyzed (fig. 6A). The early branching of plastome type IV could be taken to support Stubbe's theory that plastome IV is "primitive" based on genetic data (Stubbe 1964). Because it contains many unique sequences within the *oriB* site, plastome IV cannot be considered to be "ancestral" to the *Oenothera* plastomes, but it may retain some primitive traits due to its early divergence from the other plastome types. In fact, repeat sequences C1 and G1, which are both unique to plastome IV, may represent such primitive characters (figs. 6D and 6E).

Alignments of the *Oenothera* sequences with the 16S rRNA-trnl region from cpDNA of other plants show that this region has expanded in the *Oenothera* plastome by the amplification of several families of repeats (figs. 3 and 4). Since other dicots have only a single copy of a sequence from families A-G, amplifications of various combinations of the repeats appear to have occurred many times during the evolutionary divergence of species in *Oenothera* subsection *Oenothera*. As Hipkins et al. (1995) found when they examined a different intergenic region in cpDNA of Douglas fir, base substitutions or small insertion/deletion events occurring after the duplications allow some of the events to be ordered. Parsimony analysis clarified the relationships in most of the repeat families, with a single best tree being found in the analysis of repeats A, B, and C (figures 6B, 6C, and 6D).

The amplifications of the repeats seem likely to have occurred through replication slippage, a process that has been characterized extensively in bacteria, and hypothesized to occur in the plastid genetic system (e.g., Palmer 1985b, 1991; Zurawski and Clegg 1987; Wolfson, Higgins, and Sears 1991; Cummings, King, and Kellogg 1994; Sears, Chiu, and Wolfson 1995). None of the previously reported cpDNA segments contains as extensive or complex an array of direct repeats as those reported here. Several features of the sequence between the plastome 16S rRNA and the trnl genes probably contribute toward its tendency to undergo slipped strand mispairing during replication. The region contains A-T-rich stretches, which are frequently sites of replication slippage in bacteria (Levinson and Gutman 1987). Furthermore, replication slippage is abetted by hairpin stabilization of the slipped, mispaired intermediates (Albertini et al. 1982; Glickman and Ripley 1984; Trinh and Sinden 1991); and, as figure 8 shows, a number of hairpins are possible in this region. Finally, if the cpDNA spacer indeed lies in or between origins of replication, it would exist for some time as single-stranded DNA, and that would provide another preferred substrate for replication slippage (Allgood and Silhavy 1988), because the secondary structures could readily form. In fact, all of the hairpins of figure 8 are located adjacent to indels deduced from the sequence alignments of section 3. Heteroplasmy at the plastome control region was never observed in any wild-type lines using PCR and Southern blotting (data not shown), but the hypervariability of the rRNA operon spacer in the *Oenothera* plastome suggests that it is extremely susceptible to change in an evolutionary time frame. Examination of more plastomes may indicate if the region is under any functional constraints in terms of its sequence content or size.

In the cpDNA spacer, direct repeats would have been generated initially by slippage of the daughter strand, mispairing, and continued replication. Therefore, further slippage would have amplified various subsets of the direct repeats. Some of the repeats suffered mutation


BARBARA A. SCHAAL, reviewing editor

Accepted March 31, 1996