The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution†

Stephan Greiner1, Xi Wang1,2, Uwe Rauwolf1, Martina V. Silber1, Klaus Mayer2, Jörg Meurer1, Georg Haberer2 and Reinhold G. Herrmann1,*

1Department Biologie I, Bereich Botanik, Ludwig-Maximilians-Universität, Menzinger Strasse 67, 80 638 Munich and 2MIPS/IBI Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health GmbH, Ingolstädter Landstrasse 1, 85 764 Neuherberg, Germany

Received November 30, 2007; Revised February 1, 2008; Accepted February 8, 2008

ABSTRACT

The flowering plant genus *Oenothera* is uniquely suited for studying molecular mechanisms of speciation. It assembles an intriguing combination of genetic features, including permanent translocation heterozygosity, biparental transmission of plastids, and a general interfertility of well-defined species. This allows an exchange of plastids and nuclei between species often resulting in plastome–genome incompatibility. For evaluation of its molecular determinants we present the complete nucleotide sequences of the five basic, genetically distinguishable plastid chromosomes of subsection *Oenothera* (=*Euoenothera*) of the genus, which are associated in distinct combinations with six basic genomes. Sizes of the chromosomes range from 163 365 bp (plastome IV) to 165 728 bp (plastome I), display between 96.3% and 98.6% sequence similarity and encode a total of 113 unique genes. Plastome diversification is caused by an abundance of nucleotide substitutions, small insertions, deletions and repetitions. The five plastomes deviate from the general ancestral design of plastid chromosomes of vascular plants by a subsection-specific 56 kb inversion within the large single-copy segment. This inversion disrupted operon structures and predated the divergence of the subsection presumably 1 My ago. Phylogenetic relationships suggest plastomes I–III in one clade, while plastome IV appears to be closest to the common ancestor.

INTRODUCTION

Plastid chromosomes have received considerable attention to examine phylogenetic relationships among plants as well as to probe into eukaryotic genome evolution. Their endosymbiotic ancestry, limited coding potential, relatively conserved organization and well-defined structure provide a unique source of information to address a wide range of fundamental questions.

At present, sequences of entire plastid chromosomes are available from more than 100, generally distantly related taxa including protist, thallophytic, bryophytic and vascular plants (www.ncbi.nlm.nih.gov). This information has clarified overall evolutionary relationships for most groups of photoautotrophs as well as the outlines of the evolution of the eukaryotic cell with its compartmentalized genome. It emerged that three basic processes shaped that evolution: (i) the restructuring of the genetic potentials of the symbiotic partners (1); this conversion occurred predominantly at the unicellular level (2); (ii) the generation of novel gene sets and genetic programs with the advent of multicellular organisms and (iii) a co-evolution of the cellular genetic compartments that together constitute an integrated genetic system (3,4).

The impact of this latter process that operates in relatively short evolutionary time scales becomes apparent after interspecific organelle exchanges, which even between related species can frequently cause serious disturbances...
in the development of the resulting hybrids or cybrids (5,6). Compartmental co-evolution generates Dobzhansky–
Muller incompatibilities responsible for reproduction barriers and consequently represents an important, often
neglected, factor in speciation processes (7,8).

Combined with cybrid technology and transplastomic
approaches, sequence data of entire plastid chromosomes
from closely related species could provide access to
molecular mechanisms underlying plant genome evolution
(6,9). However, such data are still scarce. Some information
is available from the Solanaceae family (10–13), from
a few related cereal species, notably maize, rice, wheat,
barley and sugar cane (14–18), and from a few legumes
(19,20). These data have led to refined taxon topologies in
some cases, and a recent study of the interspecific
plastome–genome cybrids between tobacco and
Arropa
has allowed identification of a first molecular mechanism
that causes plastome–genome incompatibility and could
contribute to speciation processes (6,9). Ultimately, such
approaches are meaningful only in a context in which the
relationships of the plant groups being considered are well
understood and other strategies, notably formal nuclear
and organelle genetics, can be incorporated to overcome
limitations of traditional phylogenetic analyses.

The genus Oenothera is uniquely suited for studying
diversification of populations and speciation processes
(5,21–25), because it possesses a unique combination of
genetic features, and a comprehensive taxonomy, cytoge-
netics, ecology and classical genetics has been developed
over more than a century. The features include the
possibility of wide interspecific crossing, biparental
transmission of organelles, generation of fertile interspe-
cific plastome–genome cybrids or hybrids, as well as
relatively frequent reciprocal translocations of chromo-
some arms that cause structural heterozygotic genomes.
Collectively, these features allow the exchange of plastids
and nuclei between species as well as the exchange of
individual (or more) chromosome pairs or entire haploid
nuclei between species as well as the exchange of
nuclear genome combinations in approximately a
dozenspecies, predominantly over the North American
continent (22,23).

To gain causal access to compartmental co-evolution of
plant genomes and speciation processes, we have begun to
develop cell and molecular biological approaches for the
Onagraceae genus. In preceding communications we have reported on the complete nucleotide sequence of the plastid chromosome of Oenothera elata subsp. hookeri
strain johansen, representing the Oenothera plastome I
(30), on an EST library derived from the Oenothera A
genome (31), and on cell, tissue culture and transformation
protocols of that material (32–34). In this study we
present and compare the complete nucleotide sequences of
representatives of the five basic Oenothera plastomes and
deduce their evolutionary relationship and temporal
appearance. In a forthcoming study the phylogenetic
relevance of the changes observed and potential plastome-
encoded candidates responsible for interspecific plastome–
genome incompatibility will be evaluated.

MATERIALS AND METHODS
Plastome sequencing
Fully developed leaves of Oenothera elata subsp. hookeri
strain johansen (plastome I; 35), Oe. biennis strain
stauveolens Grado (plastome II; 36), Oe. glazioviana strain
rr-lamarckiana Sweden (ex genetica: Oe. lamarckiana,
plastome III; 37), Oe. parviflora strain atrovirens (syn:
Oe. cruciata, plastome IV; 38) and Oe. argilleola strain
douthat 1 (plastome V; 39) were collected from 5-
to-6-week-old greenhouse grown plants [for detailed taxon-
omy of the subsection Oenothera see (23)]. DNA from
plastome I was prepared from intact chloroplasts essen-
tially as described in (40). Molecular cloning, sequencing
and sequence analysis were performed as given in (11,30).
Total DNA was isolated from plants carrying plastomes I
to V, respectively, using the DNeasy™ Plant Mini Kit
(Qiagen, Hilden, Germany). Since a number of errors were
noted in the original sequence of plastome I (30), especially
in the IR as also noted by others (41), its DNA was
re-sequenced completely. PCR-derived fragments or frag-
ments subcloned from plastid chromosomes serving as
sequencing templates, spanning both DNA strands and
overlapping a minimum of 200 bp, were generated over the
entire chromosome of these plastomes. Almost the same,
approximately 450, oligonucleotide primers (MWG
Biotech, Ebersberg, Germany) employed for plastome I
were used. Nucleotide sequences were determined by the
dideoxy chain termination method (42) using energy-
transfer fluorochrome-labeled dideoxynucleotides (43)
with an ABI 377 robot (Applied Biosystems, Darmstadt,
Germany). Sequence data were first subjected to the
BLAST algorithm (44,45) provided by the National Centre for Biotechnology Information (Bethesda, MD, USA). Assembly and evaluation of sequences were performed with the SeqMan 6.1 program (DNASTAR Inc., Madison, WI, USA) using plastome I (accession no. AJ271079) as template for the other plastomes. All plastome sequences were aligned using the MegAlin 6.1 program (DNASTAR Inc., Madison, WI, USA) and the program BioEdit 5.0.9 (North Carolina State University) as alignment editor (46). Gene annotation of plastomes II to V was guided by the (North Carolina State University) as alignment editor (46).
RESULTS

The plastid chromosomes of the five Oenothera species that represent the genetically distinguishable basic plastomes of subsection Oenothera were completely sequenced on both strands using a primer-based strategy. The sequences are deposited in GeneBank under the accession nos. AJ271079.3, EU262887, EU262889, EU26890 and EU262891. They include a revised version of plastome I of Oe. elata subsp. hookeri strain johansen (see ‘Materials and Methods’ section) and confirm the corrected ycf2 sequence reported recently (41). A corrigeendum of the previous plastome I data will appear (30).

Size, gene content and design of the Oenothera plastid chromosomes

The individual plastome sequences assemble as circular structures of 165,728 bp (plastome I), 164,807 bp (plastome II), 165,225 bp (plastome III), 163,365 bp (plastome IV) and 165,055 bp (plastome V) in size including 56.6% coding regions and 43.4% non-coding regions (spacers and introns). The overall G+C content is 39.1%, about 41.7% in coding and 36.5% in non-coding sequence intervals (Supplementary Table 1), and is thus somewhat higher than found usually (36.8–39.0%) (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/plastids_tax.html). As expected, the chromosomes display the typical quadripartite anatomy (40,58) characteristic for many plastid chromosomes, i.e. a pair of large inverted repetitive regions and are highly variable between all species.

Sequences divergence and sequence repetition in Oenothera plastomes

The five plastomes are perfectly syntenic. This allowed a comparative analysis of indels and sequence repetition, which are usually difficult to analyze between distant species. Insertions, deletions and repetitions are relatively frequent. Relative to plastome IV, an additional 1456 nt insertions and 3819 nt deletions are found in plastome I. The corresponding nt numbers for the plastomes II, III and V are 1156/2598, 1701/3561, 864/2557, respectively.
Figure 2. Gene map of the *Oenothera* plastid chromosomes. Arrows mark the inversion breakpoints. Genes drawn on the outside are transcribed clockwise, on the inside counterclockwise.
Indels occur less often in genes and are present in only relatively few polypeptide genes, notably in accD, clpP, ndhD, ndhF, rps18, rpl22, ycf1 and ycf2. These changes in coding potential will be discussed separately (see later and Supplementary Figures 1–4 and Supplementary Tables 4 and 5).

Overall distribution and types of repeats are highly similar between all five plastomes. On average 61 tandem repeats from 55 in plastome IV up to 70 in plastome I were detected, with a mean copy number of 4.5 copies per tandem and an average size of 41 bp per copy. The largest tandem repeat regions which span more than 1 kb and consist of variants of the AAG/TTC trinucleotide sequence are found in all plastomes at the two ycf2 genes in the IR. Expanded tandem repeats are also frequent and overlapping in accD and ycf1 contributing to the substantial sequence divergence of these genes. Rpl32, ndhF and the tandemly repeated trnMCAU1 and trnMCAU2 are associated with small identical tandem repeats in all five plastomes. More interesting are tandem repeats within coding regions, which are specific for particular plastomes, generally of moderate size and low copy number. Such repeats that change coding potential are found in clpP, ndhF and rps18 (Supplementary Figures 2 and 3). The plastome-specific repeat differences associated with ccsA, rpl22, rpl32, rps19 and trnSGCU are located outside the respective coding sequences.

On average approximately 70 palindromic repeats with a maximal gap size of 3 kb or less were detected within the five plastome sequences. Palindrome sizes, however, are smaller and far less variable than tandem sizes, ranging from 32 bp (detection limit given by the threshold applied) up to 56 bp. They were detected in accD, ccsA, matK, ndhD, ndhF, ndhJ, petD, psaA, psaB, psbH, rpl32, rpoA, rpoB, rps18, ycf1, ycf2 and ycf4, but no notable changes in coding sequences are caused by palindrome repeats among the plastome types. Exceptions are the highly polymorphic genes accD and ycf2 as well as ndhD in which a palindromic repeat associated with a polyA-stretch shows a frame changing insertion of a single nucleotide (Supplementary Figure 2). Presence and absence of repetitive elements and divergent regions correlate well, but repeat content and high divergence are not strongly linked (Figure 3). The large inversion

Except for the 56 kb inversion mentioned earlier, no other inversion has been detected in the five plastome sequences. The junctions of the inversion in the intergenic regions between trnQUGU/accD and rps16/rbcL are highly divergent and contain palindromes as well as tandem repeats (Figure 3). Overall the trnQUGU/accD spacer contains more tandem repeats (five to eight, depending on the plastome) than the rps16/rbcL spacer, with a maximum of

Figure 3. Overall distribution of sequence divergence and repeat content in the Oenothera plastid chromosomes. Percentage repeat content and divergence on the y-axis were calculated using sliding sequence intervals of 1 kb with a step size of 10 bp along a multiple chromosome alignment. Whole chromosome alignments have been deduced from concatenated multiple alignments of intergenic regions and genes (see ‘Materials and Methods’ section). Repetitive regions frequently overlap with areas of high divergence.
two tandem repeats in plastome IV and none in plastome V. Three palindromic repeats were detected at the rps16/rbcL junction of plastomes I and II. In the trnQUUG/accD spacer region between two and no palindromes were identified. Of particular note are the elements of clustered palindromes split among both spacer regions which were detected in all five plastomes. Their number varies enormously, between 1 (plastome V) and 17 (plastome III). Also, the same sequence motifs can be present in tandem and palindromic arrangements. The repetition patterns preclude an accurate demarcation of the insertion breakpoints and also affect the highly variable N-terminal region of AccD (Supplementary Figure 1 and see later).

In an attempt to better understand the inversion breakpoints and underlying processes, the corresponding regions in *Oe. villaricae* strain berteriana Schwemmle, a member of the closely related sister subsection Munzia, were sequenced (rps16/trnQUUG, accession no. EU255777 and rbcL/accD, accession no. EU255778). Subsection Munzia lacks this inversion (62). The berteriana regions corresponding to the *Oenothera* breakpoints between rps16/trnQUUG and rbcL/accD do not display pronounced divergence. The spacer region between rps16 and trnQ in that plastid chromosome lacks tandem and palindromic repeats. Nevertheless, the entire region is conserved and present in two parts in all five *Oenothera* plastomes, separated by an interspersed *Oenothera*-specific sequence interval. The berteriana region between rbcL and accD, in turn, lacks palindromes, but contains two tandem repeats. Approximately 1.5 kb of the berteriana rbcL/accD spacer are unique to the berteriana plastome. Conversely, trnQ/accD spacer sequences between 1.5 and 2.5 kb depending on the *Oenothera* plastome have no equivalent in the berteriana plastome. The same holds true for the *Oenothera* rps16/rbcL spacer, in which the number of unique nucleotides differs between approximately 50 and 500. The repeat structure in the rps16/rbcL spacer of the *Oenothera* plastomes appears to be linked to the inversed arrangement of rbcL and the result of duplication and relocation during the inversion process, as the berteriana equivalent is missing. As mentioned, palindromic repeat copies split among both spacers are not rare in the *Oenothera* plastomes, but no such cases were detected in berteriana with the selected threshold.

Differences in coding regions among *Oenothera* plastomes I–V

Three distinct sequence classes, protein-coding loci, tRNA and rRNA genes, have been evaluated for sequence conservation and changes between the five plastomes. Among the 78 protein-coding genes identified (including ycf1 to ycf4), 31 are identical in their nucleotide sequence, while 47 are variable between at least one plastome pair (Supplementary Tables 4 and 5). All carry substitutions of individual nucleotides and eight additionally indels. Sixteen genes, namely atpE, cemA, ndhI, ndhL, petA, petB, psaB, psaC, psbC, rpl16, rpl33, rpoC1, rps4, rps11, rps14 and rps19 contain only synonymous substitutions which do not change the coding context. Only non-synonymous substitutions causing amino acid changes were detected in 11 genes: ndhB, ndhC, ndhH, psbB, rpl32, rpoA, rpoB, rpoC2, rps2, rps8 and rps15. Rpl22 contains synonymous substitutions and a single base insertion/deletion, ndhF synonymous substitutions and a multiple base pair indel, rps18 multiple indels and a non-synonymous substitution. Both types of substitutions are found in ndhD, in plastome I together with a single base pair insertion, and a multiple base pair insertion in plastome V. Both, synonymous and non-synonymous substitutions are present in the remaining 16 genes (accD, atpA, atpB, atpF, ccsA, clpP, matK, ndhA, ndhE, petD, psaA, psbA, rps3, ycf1, ycf3 and ycf4). Five of them, clpP, accD, ndhD, ycf1 and ycf2, differ by multiple indels among plastomes. Of the 31 tRNA genes, 30 are identical and 1, trnM<sub>CAL</sub>II, is variable in 1 nt. The mutation is not part of the anticodon, should have no or only a negligible effect on the folding of the tRNA, and hence not influence function, last not least since a second, not mutated copy, trnM<sub>CAL</sub>I, is present in the *Oenothera* plastomes. The four rrr genes are identical in all five *Oenothera* plastomes.

Protein-coding genes with length polymorphisms can be operationally grouped into three categories, those with reading frame shifts (ndhD, rpl22 and rps18), those without reading frame shifts (ycf1, ycf2, accD, clpP and ndhF) and a third class, including atpA and psbB, which are known to differ in electrophoretic mobility of their products in plastome III and plastomes IV and V, respectively, independent of the genotype with which the plastid types are associated with (29). As most of the differences described reside in polypeptide regions known to be highly variable in plastid chromosomes in general, often at the very C-terminus of a polypeptide, it is unlikely that they are of functional and/or evolutionary relevance. Therefore, details of the changes in these 10 loci including verification by SNP, CAPS, Western (cemA) analysis and comparison with corresponding loci of usually 4–8 reference species are presented in Supplementary Figures 1–4 and Supplementary Table 5.

Differences in intergenic regions among plastomes I–V

Intergenic sequence intervals, which house various functional elements, such as promoters, terminators, processing signals or ribosome binding sites, are less conserved among plastomes (76 variable, 37 identical). On average, for each of the 113 intergenic regions, 24.1 indels (2743 for 110) and the result of duplication and relocation during the inversion process, as the berteriana equivalent is missing. As mentioned, palindromicrepeat copies split among both spacers are not rare in the *Oenothera* plastomes, but no such cases were detected in berteriana with the selected threshold.

Intergenic sequence intervals, which house various functional elements, such as promoters, terminators, processing signals or ribosome binding sites, are less conserved among plastomes (76 variable, 37 identical). On average, for each of the 113 intergenic regions, 24.1 indels (2743 for 110) and the result of duplication and relocation during the inversion process, as the berteriana equivalent is missing. As mentioned, palindromicrepeat copies split among both spacers are not rare in the *Oenothera* plastomes, but no such cases were detected in berteriana with the selected threshold.
five plastomes, respectively. Differences found comprise a relatively wide spectrum of changes, including additional or lacking predicted binding sites as well as single point mutations and spacing differences. Their functional relevance as that of transcriptional start sites remain to be verified.

Regions coding for ribosomal binding sites were detected in 30 genes within the first 23 bp upstream of the initial start codon by their match to the 3' end of the 16S rRNA. The sequences show no significant differences among the plastomes.

Evolutionary analysis, plastome pedigree

Phylogenetic trees were generated by NJ, ML and MP using the 47 variable genes as well as 76 variable intergenic regions, and rooted by orthologous genes or sequence intervals, respectively, of the Lotus japonicus plastome sequence (19) as outgroup. All three methods grouped plastomes I, II and III in one clade with plastomes I and II being close relatives (bootstrap values 100% for NJ and ML) (Figure 4). ML supported a common ancestor for clade I–III and a clade consisting of IV and V (bootstrap 99%). MP resulted in one maximal parsimonious tree, placing plastome V in one clade with plastomes I–III, while plastome IV is located in a separate branch. This tree is supported both by NJ analysis of the concatenated alignments (bootstrap 96.7% for separation of plastomes V and IV) as well as by a species/consensus tree of the 47 individual NJ gene trees. In the consensus tree, only 28.9% result in a branching pattern identical to the ML topology, while 40% of all gene trees support a separate branch for plastome IV. Hence, we propose that the latter tree topology reflects the phylogenetic relationship between the five Oenothera plastomes (see ‘Discussion’ section).

Average substitution rates of protein-coding genes between the five plastomes were investigated to derive an estimate of their divergence time. Average Ks values between the five plastomes varied over more than one order of magnitude, from $3 \times 10^{-4}$ for plastomes I and II up to $3.6 \times 10^{-3}$ for plastomes I/II and V. Applying a calibration derived for dicotyledonous chloroplast genes (70), divergence for plastomes ranged generally between 416,000 (II and III) and 830,000 years. The most distant pair (I and V) has diverged approximately 1 million years ago, while divergence of plastomes I and II is estimated to be as recent as 83,000 years ago (Supplementary Table 6).

All data obtained not shown in the manuscript are available upon request.

DISCUSSION

In this study we present the first complete sequences of plastid chromosomes from closely related, morphologically distinct and still interbreeding species, for which an organelle and nuclear genetics including cybrid technology is available. The populations of the North American subsection Oenothera are well characterized taxonomically, geographically and ecologically for more than a century (5,21–24). Interspecific cybrids of the five genetically distinguishable basic plastome types and the six basic genomes of that subsection can exist in all natural or artificial combinations (5,26 and Figure 1). The latter combinations are frequently developmentally impaired, revealing that co-evolution of nuclear and organellar genomes is an important element not only in speciation, but also in pre-speciation processes. In the clade diverging populations become apparent in changing chromosome formulas within species (22) and in intraspecific restriction fragment polymorphism of plastomes (29). The integrated compartmentalized genetic system of the plant cell, as of eukaryotic cells in general, evolves in its entirety generating additional constraints to evolution not found in prokaryotes. A natural or an artificially introduced mutation in any one of the genetic compartments of the cell necessitates a compensatory event or has to fulfil the requirement that the entire system remains genetically balanced, fully functional and phylogenetically fit (28).

The coding potentials of the five plastomes, 113 unique genes, are nearly identical and comparable to those of plastid chromosomes of vascular plants in size, organization, gene clustering and conservation. They deviate from the ‘ancestral form’ of plastid chromosomes by a single kilobase-magnitude inversion in the large single-copy segment. Their sizes are moderately larger than those
found generally for higher plants and belong to the largest plastid chromosomes known from vascular plants that usually range between 130 and 160 kb (reviewed in 71). Genes are well-definable, except of the N-terminus of cemA (ycf10) encoding an inner envelope polypeptide involved in CO₂ uptake, even after comparison with cemA loci from 50 reference species in which N-terms are not consistent (Supplementary Figure 1). Judged from PCR and Western analysis, the locus does not appear to be a pseudogene; also no evidence was found for a nuclear copy of this gene. This point remains to be settled in general.

Despite their gross conservation and close relationship, Oenothera plastid genomes are remarkably diverged, compared, for instance, to those of the Solanacean genera Atropa belladonna and tobacco (11). All four kinds of mutations that have been reported from plastid chromosomes, point mutations including changes in restriction sites, repetitions, insertions/deletions and inversions occurred. The former are scattered all over the circular chromosome, but found with different frequency in the individual sequence classes (Figure 3). Somewhat less mutations are expectedly found in the IR, probably due to copy-correction of the repeated segments (e.g. 72). Insertions/deletions and repetitions in direct or inverse orientation and restricted to repeat sizes of 16/30 bp or larger with ≥90% sequence identity are relatively frequent (30,73). Repetitive and divergent regions are correlated suggesting a role of repeats in generating divergent regions, e.g. by illegitimate recombination. Slipped mispairing may cause or contribute to changes in a number of instances as well (73–75). Coding sequences are generally well-conserved among the plastomes. Eight genes, almost exclusively non-photosynthetic, differ by indels, nearly a dozen by repeats including plastome-specific changes. Although non-protein and non-RNA coding regions, such as intragenic sequences and introns (together amounting nearly 45% of the chromosomes) evolve faster than genes, they still are quite well-conserved commensurate with the close relationship of plastomes and indicative that functional elements within these sequence intervals are conserved and compactly bundled. The junctions of the IR regions that often cause size differences between spermatophyte plastid chromosomes (e.g. 76) are almost invariant.

Inversions of different sizes in plastid chromosomes have been reported from various taxa including Bryophyta, gymnosperms, grasses, Fabaceae, Ranunculaceae, Caryophyllaceae, Geraniaceae, Asteraceae, Campanulaceae and Onagraceae (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/plastids_tax.html). Three points of general interest are noteworthy. (i) Features that have been proposed to mediate inversions in plastid chromosomes by homologous or illegitimate, inter- or intramolecular recombination include the presence of dispersed repeated sequences and/or tRNA genes which frequently lie at or adjacent to endpoints (18,77–80). However, none of the elements discussed appears to extend to all evolutionary rearrangements and no satisfactory general model or explanation which would account for their involvement is available. In Oenothera, only one of the junctions lies adjacent to a tRNA gene. The exact breakpoints delimiting the inversion between trnQ and rps16 on the one end and aceD and rbcL on the other (63) cannot be assigned by sequence comparison with the non-rearranged berteriana plastome, because of the complex pattern of repetitive sequences, of a lack of readily discernible vestiges of sequences involved bordering inversions, and because causes and mechanisms of the structural alteration that might reveal footprints of the event are not really understood. Single copies of sequences that are repetitive, scrambled and arranged in a relatively complex way in Oenothera are detectable in the Munzia plastome. Whether the structural change is in any way directly related to these elements remains to be shown.

(ii) Mapping of polymorphic restriction fragments in Oenothera plastomes has provided evidence of five preferred sites of mutation (40) which coincide well with sequence data and of which two may be recombino-genetically relevant, since they reside at or near the inversion points (Figure 3). An intriguing number of inversions in different materials occurred at nearly identical positions, including these sites e.g. (15,18–20,78,79,81–83). For instance, a 50 kb inversion reported from legumes (19,20,83) resembles, but is not identical, to that found in Oenothera since it includes rps16 besides trnQ at one of the junctions. Two of the three inversions in Adonis and polymorphism in some grasses include the same region (15,79). These mutational ‘hot spots’ may be shared with the non-rearranged plastid chromosomes of tobacco species (84) and restriction analysis indicates that at least some of them contribute to intraplasmatic variation in Oenothera, which has been noted from all five plastomes (29 and unpublished data). Intraplasmatic variation is not rare (80,85), genetically virtually neutral, but the relevance of individual events in speciation, remains to be verified. In Oenothera, it would be particularly interesting to see whether the patterns reflect solely mechanistic imprecision with no phylogenetic relevance or correlate in part with geographic trends and genetic drifts of subpopulations in species differing in their chromosome formulas (22,24). Whatever causes and relation to recombinational events, it is unlikely that the shared sites reflect random effects. The underlying processes are probably comparable and of general nature, and, at least in some instances, may involve further features, such as transcription, some sequence specificity, elements of secondary structure or sites resembling att-lambda (86,87). Obviously, structural changes in plastid chromosomes are more complex than presently anticipated. An understanding of the requirements and modes of recombination could presumably contribute to clarify functional and evolutionary processes of plastid chromosomes, possibly also some way of transposition of sequence intervals to other organelles, but this can probably not be achieved by mere sequence study and will require more penetrating approaches, such as transformation strategies to determine recombino-genic activity of distinct constructs in vivo before general principles become apparent.

(iii) The Oenothera inversion disrupts known transcriptional linkage. Disruption of operons in plastid DNA of vascular plants by structural rearrangements is considered
to be rare; only a few cases have been postulated from legumes and Campanulaceae (79,88–90). However, at present this inference is not fully conclusive due to the recent demonstration of a second plastid located RNA polymerase of nuclear origin (NEP, 67) which has not yet been considered in rearrangement analysis. As the ancestral eubacterial RNA polymerase type PEP, the phage-type NEP polymerase reads the entire plastid chromosome (47) though from different, often multiple and even operon-internal promoters (91). The Oenothera inversion separates accD from rbcL which in non-rearranged plastid chromosomes are members of a larger gene cluster, including rbcL-accD-psaA-ycf4-ycf10-petA. In tobacco, spinach, Arabidopsis and maize, this cluster is co-transcribed from a PEP promoter upstream of rbcL into a giant RNA species of ~9 kb that is subsequently processed, often rapidly (Figure 5; 92). Removal of rbcL from the operon either required the replacement of the rbcL promoter by that of trnQ, use of the operon-internal NEP promoter in front of accD (93), acquisition of a new PEP promoter 5' to accD (or trnQ) or use of the NEP promoter upstream of rpoB (c.f. Figure 2). The latter alternative is not likely, because of the general weakness of that promoter. On the other hand, the translocated rbcL promoter may serve the rbcL-rps16-trnKU_\text{UUU}-\text{matK-}(psbA-trnH\text{UGG}) gene cluster, since an RNA species of corresponding size can be detected in Oenothera (Figure 5). Obviously, the promoter patterns of both regions need to be functionally defined to settle the transcriptional consequences of the inversion.

At the generic level the large inversion shared among all Oenothera plastomes provides a robust phylogenetic marker. The almost identical endpoints and its absence in the closely related South American subsection Munzia, in subsection Raimannia (62) or in the sister group Epilobium (94) suggest that this inversion is not caused by one of the proposed rare parallel inversions (79–81), but has arisen monophyletically within the Oenothera clade and late in the history of the Onagracean complex. It predates the divergence of the subsection, occurred either in the basic plastome IV or in the common ancestor of the subsection, and hence marks a recent split in the history of the genus. Even if the absolute calibration with less than 1 million years based on the molecular clock may require correction, the more than 10-fold difference of divergence times of the five plastomes will remain (Supplementary Table 6). The plastome cladograms determined separately from protein-coding and intergenic regions are nearly congruent and confirm the pedigree deduced from formal genetic approaches on the basis of both compartmental compatibility relations to genomes and differences in multiplication rates of plastids which are independent traits (5,73,95). The genetic determinants for the latter feature reside predominantly in the plastome (96). The pedigree is also consistent with changes in the thylakoid proteome mentioned earlier, notably the psbB variance of plastomes IV and V and that of atpA of plastome III (29 and Supplementary Figure 4), with a 2bp IR extension shared by plastomes IV and V as well as with various other characters.

Sequences of entire plastid chromosomes have become a powerful research tool, not only to probe into chloroplast biogenesis, function and engineering, but also into eukaryotic genome evolution. The genus Oenothera provides unique opportunities to develop phylogenetic analyses based on mere sequence data into a "functional molecular phylogeny" by applying other kinds of approaches, such as formal genetics combined with those of molecular and cell biology or physiology, to study diversification of populations and speciation processes in wider perspective. The availability of complete Oenothera plastome sequences should facilitate access to unveil molecular mechanisms that contribute to plant genome evolution, specifically to compartmental co-evolution, leading to divergence of populations or new species.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by Grants from the Deutsche Forschungsgemeinschaft (SFB-TR1) to R.G.H. and J.M., and by the Hanns-Seidel-Stiftung (Bundesministerium für Bildung und Forschung) to S.G. We highly appreciate the participation of Rainer Maier who passed away in the early phase of the project. We thank Elli Gerick and Ingrid Duschank for their excellent technical assistance, Peter Poltnigg and Elena Funk for help in sequence evaluation and Holger Grüne for the spinach Northern blot. The CemA antiserum was kindly provided by Jürgen Soll. Funding to pay the Open Access publication charges for this article was provided by the Deutsche Forschungsgemeinschaft (SFB-TR1).

Conflict of interest statement. None declared.

REFERENCES


