Local repeat sequence organization of an intergenic spacer in the chloroplast genome of *Chlamydomonas reinhardtii* leads to DNA expansion and sequence scrambling: a complex mode of "copy-choice replication"?

MAHENDRA D WAGLE*, SUBHOJIT SEN and BASUTHKAR J RAO†

Department of Biological Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India

**Present address: Institute of Molecular Agrobiology, 1, Research Link, National University of Singapore, Singapore 117 604*

†*Corresponding author (Fax, 91-22-2152110; Email, bjrao@tifr.res.in).*

Parent-specific, randomly amplified polymorphic DNA (RAPD) markers were obtained from total genomic DNA of *Chlamydomonas reinhardtii*. Such parent-specific RAPD bands (genomic fingerprints) segregated uniparentally (through mt^+) in a cross between a pair of polymorphic interfertile strains of *Chlamydomonas* (*C. reinhardtii* and *C. minnesotti*), suggesting that they originated from the chloroplast genome. Southern analysis mapped the RAPD-markers to the chloroplast genome. One of the RAPD-markers, "P2" (1·6 kb) was cloned, sequenced and was fine mapped to the 3 kb region encompassing 3′ end of 23S, full 5S and intergenic region between 5S and psbA. This region seems divergent enough between the two parents, such that a specific PCR designed for a parental specific chloroplast sequence within this region, amplified a marker in that parent only and not in the other, indicating the utility of RAPD-scan for locating the genomic regions of sequence divergence. Remarkably, the RAPD-product, "P2" seems to have originated from a PCR-amplification of a much smaller (about 600 bp), but highly repeat-rich (direct and inverted) domain of the 3 kb region in a manner that yielded no linear sequence alignment with its own template sequence. The amplification yielded the same uniquely "sequence-scrambled" product, whether the template used for PCR was total cellular DNA, chloroplast DNA or a plasmid clone DNA corresponding to that region. The PCR product, a "unique" new sequence, had lost the repetitive organization of the template genome where it had originated from and perhaps represented a "complex path" of copy-choice replication.

1. Introduction

The discovery of uniparentally inherited streptomycin resistant mutants founded the field of chloroplast genetics in *Chlamydomonas reinhardtii* (Sager 1954). Using density labelling and equilibrium centrifugation studies, Sager and Lane (1972) provided the first evidence for uniparental transmission of chloroplast DNA in *C. reinhardtii*. Subsequent molecular genetic and classical genetic work further established this phenomenon. These studies also revealed some interesting aspects of genetic regulation associated with uniparental (through mt^+) transmission of chloroplast genomes in *Chlamydomonas* (Armbrust *et al* 1993; Sears and VanWinkle-Swift 1994). Optically tweezed single zygote capture of *Chlamydomonas* was elegantly used to demonstrate the temporally and spatially regulated digestion of chloroplast genome of the other parent, namely mt^-, during selective transmission of mt^+ chloroplast genome (Nishimura *et al* 1999). These results also suggested that the two chloroplasts, from the opposite mating types, remained physically separated during zygote formation that accompanies selective degradation of the mt^- chloroplast genome. However, reciprocal crosses between *Chlamydomonas eugametos*

Keywords. *Chlamydomonas*; copy-choice replication; DNA repeats; sequence organization; RAPD

and *Chlamydomonas moewusii* seem to follow a more complex pattern where specific alleles from either parents are transmitted by unidirectional gene conversions which implies that chloroplast genomes from two different gametes recombine efficiently before complete degradation ensues (Bussieres *et al* 1996). Unlike meiotic recombination in eukaryotes, chloroplast recombination resembles yeast, mitochondrial and phage recombination in that it is typically nonreciprocal and involves multiple rounds of exchange events among a population of parental genomes (Boynton *et al* 1992). Chloroplast genome exchanges in a cross between *C. reinhardtii* and *C. smithii* revealed non-random distribution of events revealing hot-spots and adjacent cold-spots of recombinational activity (Newman *et al* 1992). Most of the reported studies relied on genetic markers with attendant phenotypes to monitor genetic exchanges that were subsequently characterized at molecular level. In order to map the patterning of hot and cold nature of recombinational events across the chloroplast genome *in toto*, one needs to develop parental specific DNA markers along the entire length of the genome, which can then be easily monitored in each meiotic progeny quantitatively. In addition, chloroplast DNA recombination is very active in vegetative cells of *C. reinhardtii*, as indicated by the dispersive labelling of plastid DNA in density transfer experiments (Turmel *et al* 1981).

In this study, we used randomly amplified polymorphic DNA (RAPD) analysis to scan regions of sequence divergence between a pair of interfertile but highly polymorphic strains, *C. reinhardtii* and *C. minnesott*i, in order to design parental specific PCR markers. Using A/T-rich arbitrary primers, we selectively amplified and mapped PCR markers to the chloroplast genome. Upon more careful analysis of one of the markers, we uncovered a highly repeat rich domain between 3′ of 23S, a full 5S and intergenic region between 5S and psbA in the chloroplast genome from where a uniquely scrambled PCR product emanated that had no linear sequence correspondence with its own template. We propose that a complex, yet unique nonrandom pattern of copy choice replication generated such a product, whether the template chosen was total cellular genomic DNA, purified chloroplast DNA or a plasmid clone harbouring the requisite template.

2. Materials and methods

Strains such as CC125, CC124 and CC1952 were provided by Dr Elizabeth Harris (Chlamydomonas Stock Center). All the media used to grow the cultures are prepared as described by Harris (1989). Chemicals used for media preparation were obtained from USB-Amersham, Sigma, Merck, SRL and Glaxo Laboratories (India) Ltd. Preparation of whole cellular genomic DNA, chloroplast DNA was done as described (Harris 1989). Mating reaction followed by tetrad analyses were done as described (Dutcher 1995).

2.1 *RAPD PCR reaction (two-step protocol)*

In a 50 µl reaction, genomic DNA (10 ng) was amplified using W68 primer (5′-AATGGAGAAGGTAGAGGAT-TTAGAAAAAATTTA-3′, 10 µM nucleotide concentration), all four dNTPs (200 µM, each), Taq Polymerase (Amersham, 1 unit) in 26 mM Tris-HCl (pH 8·4), 4 mM $MgCl_2$ at the following conditions: 5 cycles at 94°C (denaturation) for 1 min, 35°C (annealing) for 1·5 min and 72°C (extension) for 2 min, followed by 30 cycles of 94°C for 1 min, 45°C for 1·5 min and 72°C for 2 min. After the end of the last cycle, tubes were kept at 72°C for 5 min. Five µl of this reaction mixture from such a PCR reaction was further amplified in the second round using the same reaction conditions. The amplification conditions were: 35 cycles of 94°C for 1 min, 52°C for 1·5 min, and 72°C for 2 min. At the end of the last cycle, the tubes were kept at 72°C for 5 min and stored at 4°C till further gel analysis.

3. Results

3.1 *Uniparental segregation of RAPD-markers*

Random primer mediated polymerase chain reactions are known to yield amplification products which behave as authentic genetic markers (Williams *et al* 1990; Haring *et al* 1996; Saal and Wrickle 1999). Naturally occurring strains of *Chlamydomonas*, although polymorphic across entire genomes, are still interfertile in a laboratory cross (Lemieux *et al* 1980; Mets 1980; Gross *et al* 1988). Such strains are very useful for generating strain-specific RAPD markers. Using this approach we generated markers across the *Chlamydomonas* genome, which we plan to use for studying recombinational exchanges in chloroplast and nuclear compartments.

We analysed the PCR patterns from total genomic DNA of few sets of complete tetrads obtained from a cross between *C. reinhartii* (CC125, mt^+) and *C. minnesoti* (CC1952, mt^-), a pair of polymorphic interfertile strains. Based on their unique segregational patterns, we assessed whether parent-specific RAPD-bands segregate as nuclear, chloroplast or mitochondrial markers in a meiotic cross between CC125 and CC1952. Since chloroplast genome in *C. reinhardtii* is considerably richer in A + T content than that of nuclear genome (Sager and Ishida 1963), we tested several A + T rich primers for random primer amplification. Primers with arbitrary sequences with high

A + T content (> 60%) and in the size-range of 20–30-mers were tested using total genomic DNA. Most of the primers yielded RAPD-band patterns that segregated in a parent-specific manner, in spite of using total genomic DNA as template for the amplification reiterating the rationale of design. A representative pattern is shown with one of the primers (W68) on a complete tetrad, with each spore DNA analysed in duplicate (figure 1). Segregation of nuclear genetic markers was also examined in such complete tetrads using phenotypic characters of each strain. CC125 is mutant for "y1" whereas CC1952 is wild type for the same. Conversely, CC1952 is "agg1$^-$" whereas CC125 is wild type for the same (Harris 1989). In the tetrads analysed for RAPD-bands, these two nuclear markers always segregated in 2 : 2 ratio (data not shown). The bands that were mt$^+$ parent specific (P2 and P4, figure 1) were recovered in the tetrads whereas those from mt$^-$ were not. Interestingly enough, RAPD-patterns were identical for a sample whether the template used for PCR was total genomic DNA or purified chloroplast DNA (data not shown). The conditions of PCR and the A/T-rich nature of the primers used seem to selectively sample the chloroplast genome rather than the nuclear one, for successful amplifications. We selected one each of the amplified RAPD-bands from CC125 (P2-band) and CC1952 (P5-band) that appeared to be parent-specific (figure 2A) and probed the PCR-gel patterns of both parents (figure 2B, C). Southern analysis confirmed the specificity of the P2 and P5 bands to parents CC125 and CC1952 respectively. The cross hybridization with the other parents was marginal which suggested that such PCR-products represented unique molecular markers of the chloroplast genome. To substantiate the chloroplast location of the markers, we probed restriction enzyme digested purified chloroplast genomic DNA with the RAPD-bands separately. Each RAPD-marker hybridized to different regions in the chloroplast genomic digest. As an example, we present the data obtained for RAPD-band P2 (figure 2D). P2 hybridized to a 10·6 kb band which corresponds to either fragments 3–4 or 5 on chloroplast genome restriction map with *Eco*R1 (Harris 1989).

3.2 *Mapping of RAPD-marker P2 on the chloroplast genome*

The entire chloroplast genome is available in the form of a plasmid library (Chlamydomonas Genetic Stock Centre, chlamy@acpub.duke.edu). Plasmid clones P124 (*Bam*HI fragments 14/15, 3·0 kb insert) and P59 (*Bam*HI fragments

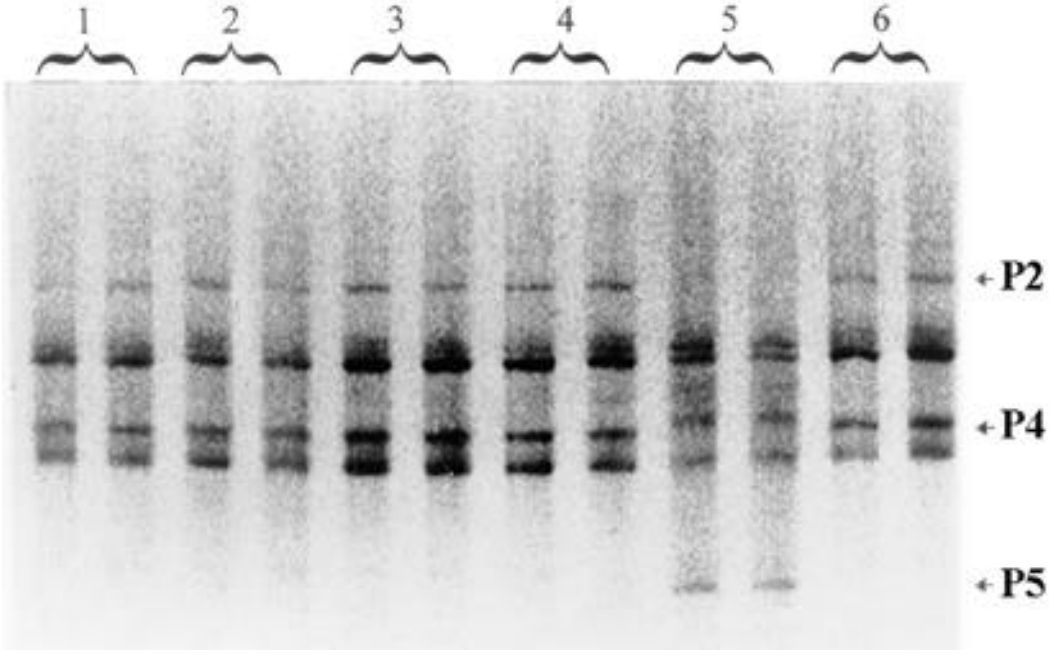

Figure 1. Segregation of RAPD markers in a cross between CC125 and CC1952.

Tetrads obtained from a cross between CC125 and CC1952 were dissected, spore colonies grown in R-medium. RAPD-PCR was performed on genomic DNA prepared from these spore-colonies and analysed on 1·2% agarose gel pre-stained with ethidium bromide. A sample of RAPD-PCR obtained with primer W68 (5′-AATGGAGAAGGTAGAGGATTTAGAAAAAATTTA-3′) (see §2) is represented. Lanes 1-4, RAPD-PCR of a set of tetrads, in duplicates; lane 5; CC1952; lane 6; CC125. Parental specific bands are indicated as P2, P4 and P5, on the right.

11/12, 7 kb insert) cover most of *Eco*R1 fragments 3–4 (10·6 kb) whereas clone P62 (*Eco*RI fragment 5, insert 10·5 kb) covers the entire *Eco*R1 fragment 5 (Harris 1989). A set of plasmid PCRs were performed using the same arbitrary primer that yielded the RAPD-bands, P2, P4 and P5. A set of plasmid clones containing regions other than *Eco*R1 fragments 3–4 or 5 was used as negative controls. Arbitrary primer PCR generated a product of the right size, only with plasmid clone P124 and not with either P59 or P62 (figure 3A, lanes 1, 2 and 4 respectively). Irrelevant plasmid clones yielded no PCR-products (lanes 3, 5 and 6). The RAPD-PCR product obtained with P124 matched in size with the original P2-band (about 1·6 kb, cloned) (lane 7). Using specific controls, we confirmed that the RAPD-PCR product from P124 and specific PCR product from P2-clone originated spuriously from their respective insert-regions rather than from vector-regions. Purified insert-DNA's from P124 and P2 revealed a unique PCR-product whose size was identical to each other (figure 3B, lanes 1 and 5 respectively). Entire plasmid controls, namely P124 and P2, yielded PCR-products identical to each other as well as to those from the purified inserts (lanes 2 and 6). This analysis demonstrated that RAPD-band P2 originated from the chloroplast genomic region represented by *Bam*HI fragments 14/15 of 3 kb size which in turn was a part of *Eco*RI fragments 3–4 in the chloroplast genomic map (Harris 1989).

In order to align the sequence of P2 with that of P124, we sequenced P2-clone from either end. We retrieved about 700 bases from either ends of the insert which encompassed almost the entire length of the insert (figure 4A). At either end, the 5′-sequences revealed a complete match with the vector sequences followed by the entire stretch of arbitrary primer sequence. Down stream to the arbitrary primer sequence followed a sequence that failed to linearly align with any stretch in the 3 kb insert of P124 or with the chloroplast genome. About 2 kb, 5′-end sequence of P124 insert was available in the database which belonged to 3′-end of 23S and whole of 5S RNA genes. The remaining 1 kb, belonging to the intergenic region between 5S and psbA, was sequenced by

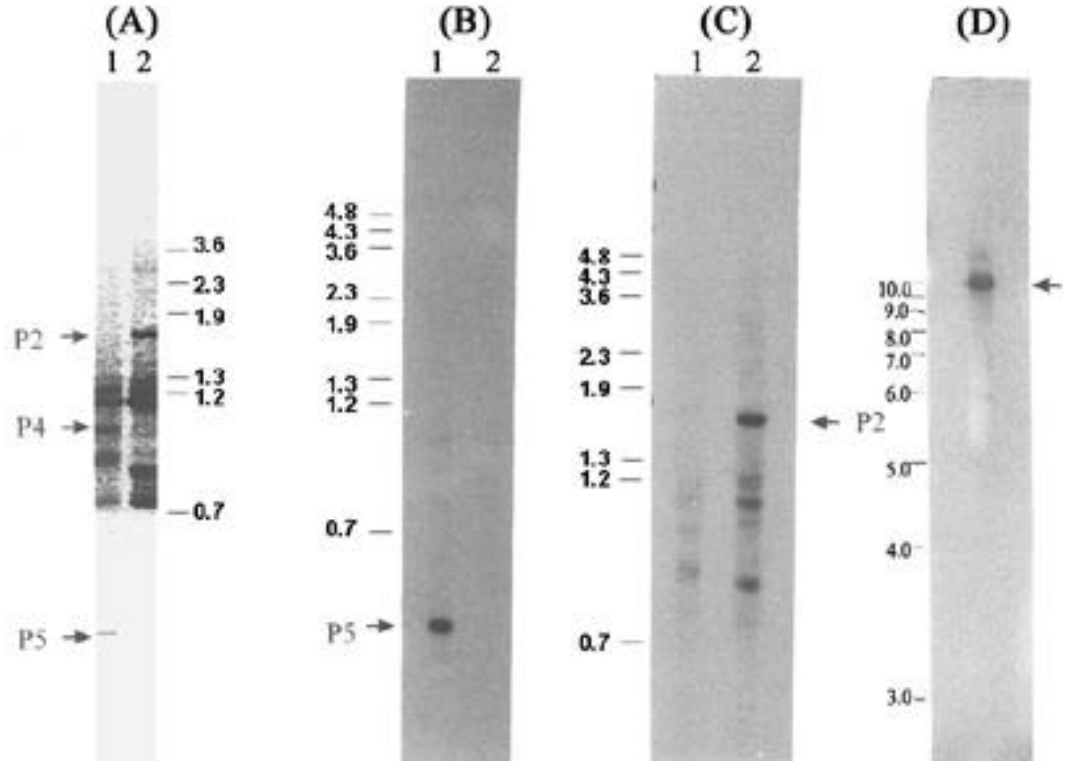

Figure 2. Parental specificity and localization of RAPD-PCR markers to chloroplast genome by Southern analyses. (**A**) RAPD-PCR was performed on total genomic DNA extracted from CC1952 (lane 1) and CC125 (lane 2) followed by analysis in an agarose gel containing ethidium bromide. P2, P4 and P5 bands are marked and the positions of standard markers are specified. Southern analysis of RAPD profile shown in (**A**) using either P5-DNA (**B**) or P2-DNA (**C**) as a probe. Chloroplast DNA from CC125 was purified by NaI gradient centrifugation (Harris 1989) followed by complete digestion with *Eco*RI and resolution on a 0·8% agarose gel, where all expected *Eco*RI fragments were resolved as revealed by ethidium bromide staining (not shown). The fragments were transferred to nylon membrane and probed with RAPD-marker P2 (**D**).

us. Sequences retrieved at either end of the P2-clone did not match with any sequence in P124 insert. We verified again the authenticity of P2-sequences by repeating the cloning and partial sequencing of P2-band obtained from the arbitrary primer PCR of total genomic DNA, purified chloroplast genomic DNA as well as P124 plasmid DNA. All the three templates yielded P2-clones that had identical sequences adjoining the vector and arbitrary primer sequences and matched with the original P2-sequence (figure 4A). Therefore, it was clear that P2 represented a unique product of arbitrarily primed PCR from chloroplast sequences of P124 region.

After failing to align P2 sequences with that of P124, we reconfirmed most of the sequence in P124 which matched perfectly well with that reported earlier. Thereby, we ruled out any untoward sequencing artifacts in our experiments. Lack of any significant sequence alignment of P2 with its own template, namely P124, was very surprising, as well as striking. The P2 sequences retrieved from SP6 and T7 ends of the cloned insert (in pBluescript) scored only 61% and 56% interrupted alignments with P124 template respectively. The ALIGN program revealed that even these low scores were not significant because the alignments included several penalties, such as contiguous mismatches and deletions encompassing 1–6 nucleotides, but no sequence additions. In order to rationalize this discrepancy, we suspected that P124 target sequences might have been unusually rich in stable secondary structures that could have led to such a PCR-amplification product. We are not aware of any such example in the literature where the PCR-product reveals such gross misalignment with the template.

3.3 *An unusual "repeat rich" organization in P124*

We analysed P124-insert for unusual patterns of direct and invert repeat motifs by employing the program, Large Dot Plots, that basically plots a correlation dot matrix of sequence identities and/or complementarities across any two given sequences. The algorithm uses a word size of eight and remaps the matches onto a 500 × 500 grid, essentially splitting a sequence into 500 octanucleotide units (http://alces.med.umn.edu/rawdot.html). Such a search conducted on 2579 bp stretch of P124-insert is plotted as a correlation dot-matrix between two identical strands (referred as *cis*-strand complementarity, where recurrence of an octanucleotide repeat is marked as a dot). A diagonal and the symmetric distribution of dots across the diagonal in such a plot is a consequence of the comparison made between two identical strands. The dots that are widely scattered across the diagonal indicate a relatively random occurrence of direct repeats. Closely spaced repeats appear as dots closer to the diagonal. Moreover, a cluster of dots along the diagonal suggests a higher density of closely spaced direct repeats. The correlation-matrix obtained on 2579 bp of P124 revealed that there was a dense cluster of direct repeats at the interval of 1800–2400 nucleotide positions, whereas the rest of the region was sparsely populated (figure 4B).

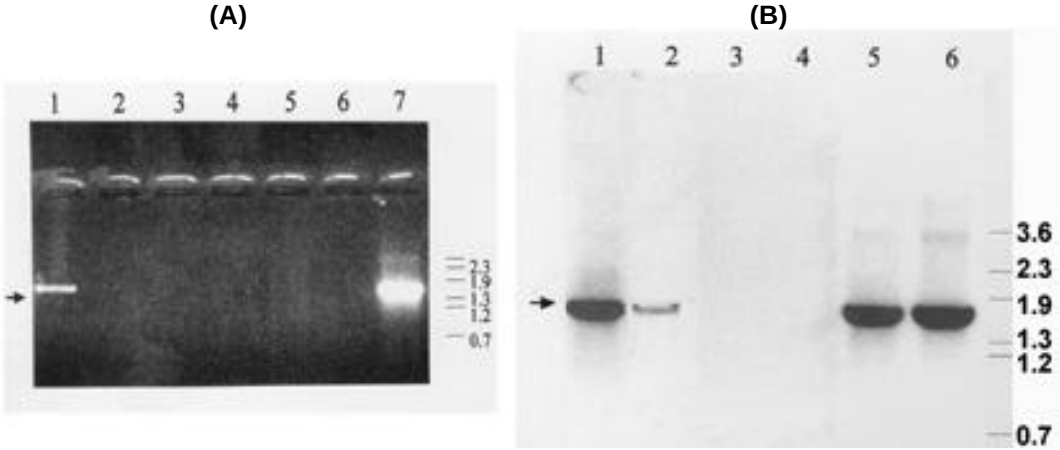

Figure 3. RAPD-PCR analyses of plasmid clones carrying various fragments of chloroplast genome. Various plasmid clones were subjected to RAPD-PCR analyses, followed by analyses on agarose gels. The numbering of plasmid clones is as specified in Chlamydomonas Culture Stock Center and the fragment numbering is as described in the source book (Harris 1989). (**A**) Lane 1; P124 clone carrying *Bam*HI fragments 14/15; lane 2, P59 clone carrying *Bam*HI fragments 11/12; lane 3, P50 clone carrying *Bam*HI fragment 4; lane 4, P62 clone carrying *Eco*RI fragment 5; lane 5, P279 clone carrying *Eco*RI fragment 25; lane 6, P19 clone carrying *Bam*HI fragment 8; lane 7 (a positive control): plasmid clone where RAPD-marker P2 was cloned. (**B**) Lane 1, chloroplast genomic insert purified from plasmid clone P124; lane 2, entire P124 clone; lane 3, P62 clone carrying *Eco*RI fragment 5; lane 4, P279 clone carrying *Eco*RI fragment 25; lane 5, insert purified from plasmid clone into which P2- marker was cloned; lane 6, plasmid clone containing P2-marker (last two lanes represent positive controls).

Such a cluster of direct repeats was rather striking and indicates the unique nature of the sequence pattern in the 600 bp region corresponding to the intergenic spacer between *5S* and *psbA* genes.

Similar analysis was done to locate the distribution of inverted repeats. This was achieved by comparing the recurrence of octanucleotide units in 2579 bp of P124 (*cis*-strand) and its exact complement (*trans*-strand). Such a recurrence of sequence identity in a *trans*-strand is plotted (referred as *trans*-strand complementarity) against its location on the original *cis*-strand (figure 4C). Such an analysis revealed a correlation-matrix where the dots were symmetrically scattered across the median of a weak diagonal, the diagonal itself being a result of closely spaced inverted repeats along the sequence. The dots scattered far away from the diagonal indicated distantly separated inverted repeats that were somewhat uniformly distributed. Surprisingly enough, there was a strong cluster of dots again at the 1800–2400 nucleotide positions of *cis*-strand (or at 200–800 nucleotide positions of *trans*-strand), the same location where previous analysis had shown a cluster of direct repeats (figure 4B, C). A simple interpretation of this analyses is that the sequence between the 1800–2400 nucleotides is strikingly richer in short units of direct and inverted repeats, the upper size-limit of the repeat unit being at least an octanucleotide, the primary search-unit in the program employed.

To be able to assess the PCR-ability of this unique template, we analysed how the putative binding sites of the arbitrary primer (W68) used in the current PCR-reactions are distributed in P124 *vis-a-vis* the domain that is rich with the repeats described here. We aligned the arbitrary primer sequence on the top-strand of P124 (*cis*-strand) using the ALIGN program. Out of the four best alignments (rightward arrowheads) that matched the primer with the template at its 3′-ends, three were from the repeat-domain region (figure 4D). In all the cases, the alignments seemed sufficiently strong to function during the low stringency PCR employed here. In the same

P2-sequence, (1576 bases)

```
   1  AATGGAGAAG GTAGAGGATT TAGAAAAAAT TTAgcaattg ctacattata
  51  agttaatgta ccattttcaa ttgaagcgaa tttaatgatt tgaatttgta
 101  ctttatctcc taatttagca cctaaattag gttttacaaa taaaatagta
 151  tttgttggta ctgtagttaa attgcctgct gaaccatttt gattattaat
 201  acgtgtatct aaacctaccg ctccttttt aaataaagtt tgttcaaaac
 251  catttacact taaattttgt gctgttgctg ttgttacttt aaatctagat
 301  gttttaaata acgttttgca ttttttggca atacaacagt aaattttgtg
 351  cctttaaagc tgttttgtgt taagcttgtt gaaacttgat ttaaacgttg
 401  agttgaaact gttgctccca aatgcatacc atttttnccg tgtacaagnc
 451  tgaacttttc accatgtang caattttaca tagttgtact attgatatac
 501  cnaatanatc tgtngttgtc nattgangnc ttacgncacg ggtatgtgta
 551  tatagtctgt nttgttgncc ttatatacat nattttntcn attanctact
 601  acgtacattg gtgcnnctat caggatntac cttattaatt nntngtggtc
 651  cntggggnan ngancannat ttanncnntg gccaccgntt taaaannttc
 701  cgatattggn acggtgttnt ttnggttnna antactagat aaaaggnttg
 751  ttaanaaaat tagccgcagc ctttannngnt aaacnttang ncctngccan
 801  taaaaaattg ncccccttccc ccttngggga aatttaaanc cantaattna
 851  tggnnccnnn ngccnnnaaa aanttaaaat tcccggtaaa ggggaaggtc
 901  cntccttttnc cccttntggg gannatttaa nanactcccn aaatgaaggc
 951  caaggtaaan ttttaggnag gtaaggtaaa ccctgcctaa nngcaagcna
1001  aataaatttt atttcttaag ttttacttgg cccngaaggg aaggacgtcc
1051  agtggcagtt gcgaagtatt aatattgnan attaatatct tgccaactgc
1101  cacttaaaat ttattatccc gaagggacgt cntaatataa atattgaatg
1151  ttnacatact ccgaaggagg acgccagtgg cagtggtacc accactgcct
1201  agtatgtaaa cctgctagcg cagtaaataa atgttattct aagttgactt
1251  gcccgaagtg gttagccgat aggcgaggca gccaataaat gttatcccgt
1301  aaagggatgc cagttaaaat ttatttacct atcaggcgat atccttcgca
1351  gaatattaat ctgggagcgc agcctttaac gagttagtag acttcccttt
1401  cggggctgcc acataacaaa gacattttat gccataaagg ggttgtctct
1451  aacttattat gtgtctgaag cagaaaagtt tgtaaggcat tcaatttcta
1501  ttcaaaatcc actattgatt tttatatttt tcttttctat taaTAAATTT
1551  TTTCTAAATC CTCTACCTTC TCCATT
```

Figure 4A. For caption, see p. 591.

alignment, we wanted to check the status of primer binding sites on the complementary (bottom) strand of P124 (*trans*-strand). We did this by searching for the sequences that are complementary to the arbitrary primer (W68 complement) on the top-strand of P124. The best alignments that scored well at the 5′-ends of complementary primer (that would correspond to the 3′-ends of the arbitrary primer), are depicted (leftward arrowheads, figure 4D). Again, all the best alignments happened selectively in the repeat-domain. Therefore, it was clear that almost all putative primer-binding sites were concentrated on either strands of the repeat-domain. If the putative primer binding sites were real, the PCR product P2 should have originated in an unknown manner from the repeat-domain itself (see §4). To test this, we repeated the sequence alignment between P2 and P124-insert. In this analysis, the P2 sequence was split into several non-overlapping segments of 200 bp each and aligned with the entire insert of P124 at various stringencies. The alignment at the highest stringency (80%–100%) ensued essentially in the region encompassed by the repeat domain described earlier (homology scale panels of figure 4B, C). Only upon lowering the stringency of alignment did the rest of the region in P124 reveal a match with P2 sequences.

We confirmed the results of sequence alignments by fine-mapping P2 sequences on P124 insert by Southern analysis. Purified insert from P124 clone was digested by various restriction enzymes followed by probing with labelled P2 sequences (figure 5A, B). About 2 kb sequence of P124 insert available in the data-bank and the remaining 1 kb that was sequenced by us enabled the construction of a restriction map for the enzymes used in this experiment (data not shown). Sizes of the restriction fragments obtained were consistent with those predicted from such a restriction map. Southern analysis revealed that P2 sequences were entirely concentrated towards the end where repeat-domain was located in P124 insert. In fact, the fine location of P2 homology coincided precisely with those fragments of *Dra*I (lane 6) and *Alu*I (lane 7) digests that contained the repeat-domain. The Southern mapping of P2 sequences in P124 corroborated the

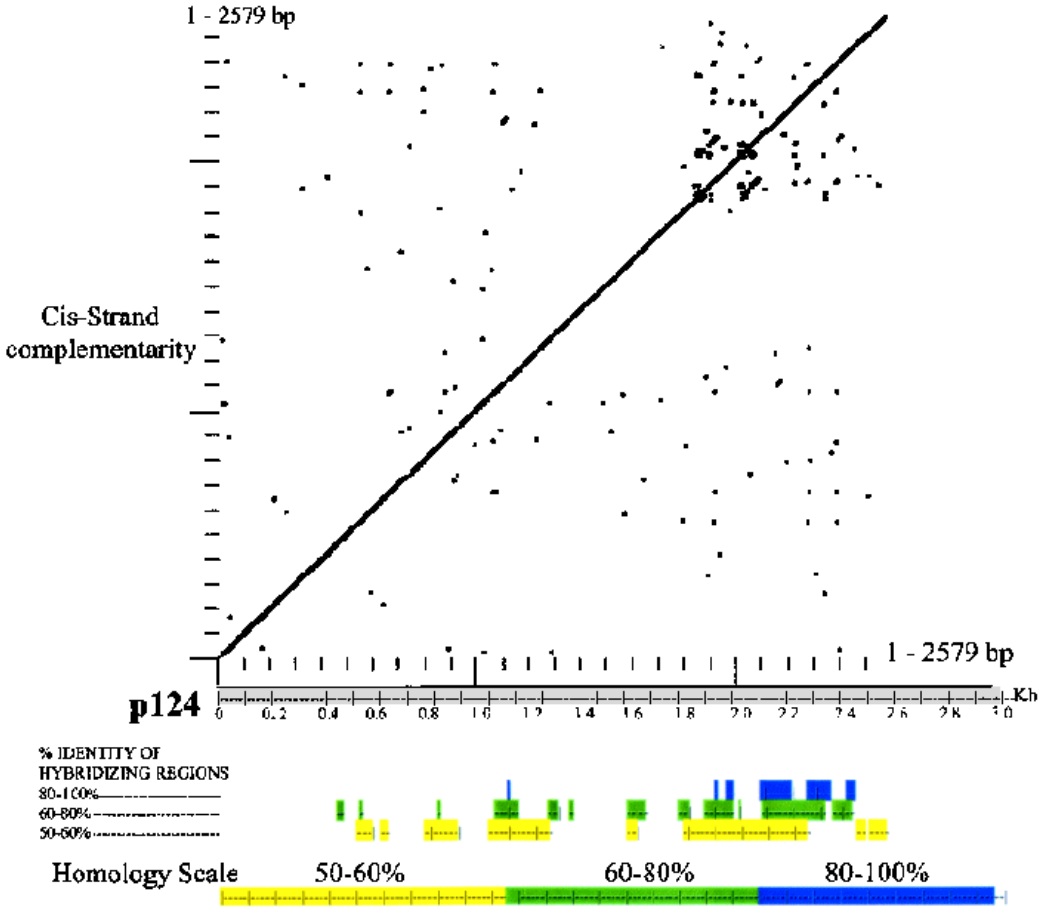

Figure 4B. For caption, see p. 591.

alignment results obtained at the highest stringency (80%–100%) (figure 4B, C) and demonstrated the localization of P2 sequences essentially in the repeat domain region at the 1800–2400 nt in P124. These analyses suggest that the putative repeat domain at the 1800–2400 nt location is the prime source of template for generating P2 product by arbitrary primer PCR. To test the fate of such a repeat sequence organization, following PCR reaction, we analysed sequence organization of the P2 product by the Dot Plot program, as described earlier. The correlation-matrix for direct as well as inverted repeat sequences revealed that the P2 sequence had lost the repeat-cluster of the type observed in P124 at 1800–2400 nt locations (data not shown). The matrix showed a pattern where direct and invert repeats were uniformly distributed all along the stretch without the typical clustering that was seen earlier in P124. The results pointed out to a PCR reaction which presumably ensued in a repeat domain, but in the process had either "scrambled" up the organization of the domain or deleted the repeat motifs by a process reminiscent of "copy-choice" replications (see §4).

3.4 *Design of parental specific PCR-marker*

PCR product P2 was mt^+ specific as shown by Southern analysis of RAPD-gel (figure 2C). One of our main goals in this project was to design specific PCR-markers for chloroplast genome for either parent. RAPD analyses was intended to scan the areas of chloroplast genome with sufficient sequence divergence so that the classical two-primer PCR reactions that are parental specific could be set-up after retrieving partial sequence information from parent-specific RAPD-markers. This could be done even if no prior sequence information of chloroplast genome is available for a given parent, the case for CC1952 parent in the present study. As has been described the P2 seems to have originated from an intergenic repeat-domain and it is very likely that mt^+ specificity of P2 was due to the

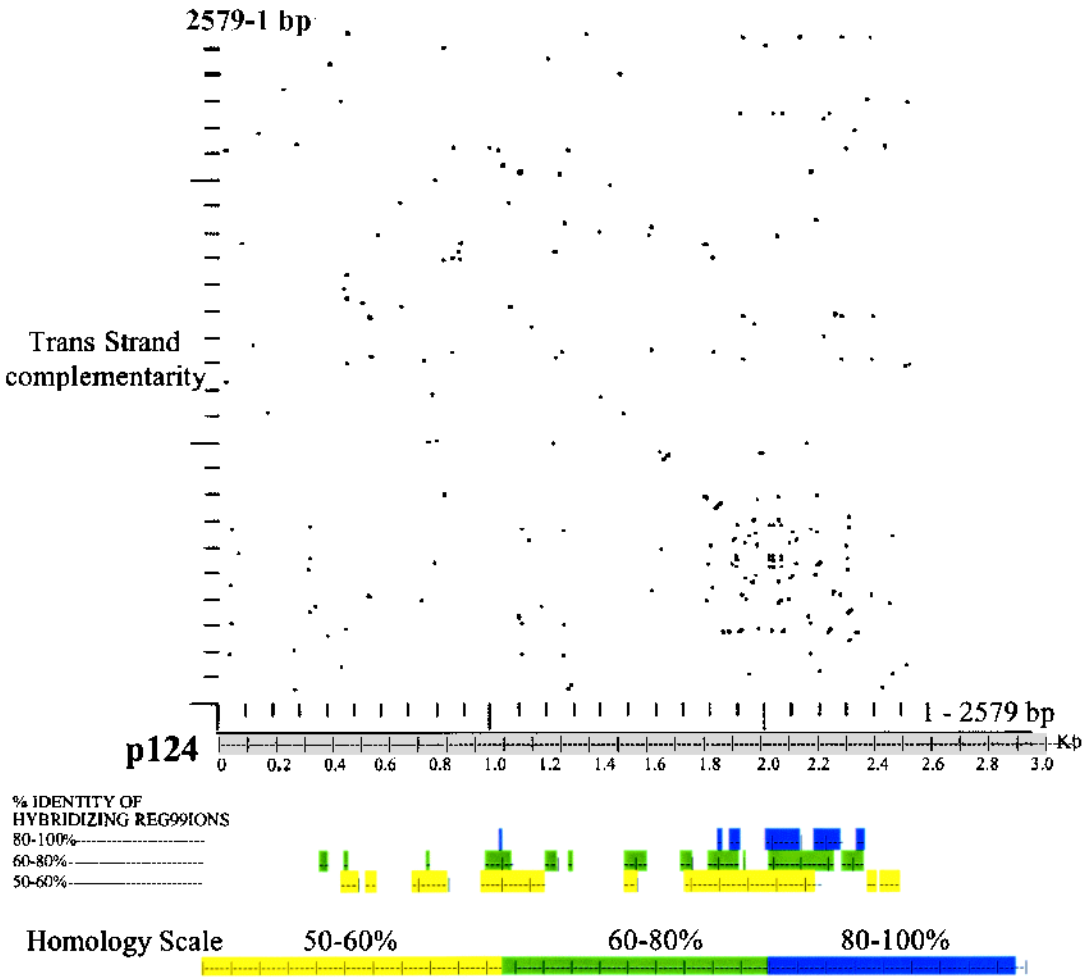

Figure 4C. For caption, see p. 591.

presence of repeat domain specifically in mt^+ parent. This possibility was tested by the following specific PCR reaction. If the repeat-domain is mt^+ specific, a primer sequence selected from within the repeat should amplify a specific product either in conjunction with another primer from the repeat-domain itself or its vicinity, only in mt^+ parent. Given the complex nature of the PCR outcome observed in P2 product, we surmised that a combination of two primers, one from the repeat and the other from its vicinity, is a better design to obtain an authentic two primer-PCR marker that is specific for mt^+. To decide the primer from the vicinity, we chose a region that was relatively poor in the distribution of closely spaced direct repeats. We selected one such region between 1·0–1·8 kb of P124 sequence. This part of the matrix was remarkably denuded of the correlation-dots, indicating the paucity of direct repeats (see correlation-matrix, figure 4B). A PCR reaction between a primer from this region (1550 nt location) and another one from the repeat-domain (2300 nt location) was compared for CC125 and CC1952. The PCR reaction was completely specific for CC125, the parent that housed the repeat-domain (figure 6). CC1952 did not show any PCR-signal, either with single primers or with both primers. The PCR product recovered from CC125 was of the expected size, matched well with that from the P124 control (figure 6), mapped to the expected location in P124-insert (data not shown) and therefore can serve as authentic PCR-marker for mt^+-chloroplast genome.

4. Discussion

The paper describes a method to locate regions of sequence divergence between a set of genomes, without *a priori* sequence information, and thereby enabling us to generate parental specific PCR-markers targeted to those regions. The methodology has been described for chloroplast genomes of two parental strains that are known to be polymorphic across their genomes and yet are interfertile in mating. Such polymorphic strains are likely to enhance the success rate of generating parental specific PCR-markers. A quick scan for regions of sequence divergence between two genomes is achieved by obtaining random primer mediated PCR-amplifications that are not only parent-specific, but also segregate as true genetic markers (figure 1). A large number of such markers can be generated using a few random primer PCR reactions. By using the total genomic DNA from a few complete tetrads, we could quickly establish the segregation patterns of such PCR-bands. In this study, we directed the RAPD to the chloroplast genome by employing several AT-rich random primers and obtained markers, several of whom segregated as mt^+ markers. We deliberately chose A/T-rich primers so as to selectively target chloroplast genomes in total genomic DNA preparations. Expectedly, the success rate of generating chloroplast genome specific markers from total genomic preparations turned out to be high (about 70%–80%) in our PCR-trials employing about 20 such primers (data not shown). We followed up the molecular details of one such

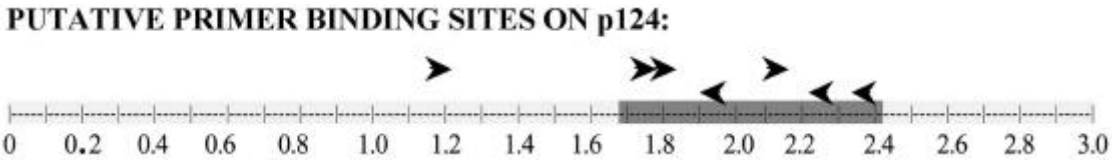

Figure 4A–D. (**A**) Sequence of P2-marker in 5′ to 3′ direction. Sequences in bold correspond to that of W68 primer used in RAPD-PCR. (**B, C**) {(http://alces.med.umn.edu/rawdot.html) This web-page accesses a very fast dot plot algorithm designed for comparing large DNA sequences with itself or its reverse complement. The algorithm uses a word size of eight and re-maps the matches onto a 500 × 500 grid. As the query becomes large, it becomes necessary to set a threshold for plotting, below which the score is cut off. The default output is currently a postscript file that will be displayed externally to the WWW browser. This program is derived from an X-windows based interactive version that saves the internal mapping so that changes in the cutoff score can be displayed rapidly (Stew Scherer: stew@mail.ahc.umn.edu). The program runs on a Sun Sparc 5/85. It makes hash tables of octamer usage and remaps all matches on a 500 × 500 grid. The program counts all dots in each grid square. Envision the result as a 3-dimensional surface with the number of dots in each square as the Z-coordinate. The cutoff is an XY-plane slicing the surface. The current version can handle sequences of several MB but only large nearly identical regions can be detected above the noise.} The plots represent comparison of P124 strand with either itself (*cis*-strand complementarity in **B**) or with its complement (*trans*-strand complementarity, **C**). The former locates direct repeat clusters, whereas the latter the invert repeat clusters (see §3).

At the bottom, the homologies between P2 and P124 are compared along the entire length of the P124 DNA sequence (the X-axis of matrix-plot), based on BLAST comparisons (using default parameters) (www.ncbi.nlm.nih.gov/gorf/bl2.html, at the NCBI server) (Tatusova and Madden 1999). Regions showing different percentages of homology are differently shaded, as indicated. (**D**) P124 (3 kb) is represented as a thin line on which the region exhibiting direct and invert repeat cluster (**B**, **C**) is depicted as a shaded region. RAPD-primer W68 sequence was mapped to the chloroplast insert of P124. Rightward arrows represent a match with W68, whereas the leftward arrows represent a match with W68-complement (TTACCTCTTCCATCTCCTAAATCTTTTTTAAAT).

PCR-reaction to establish our original premise that starting from RAPD-markers, one could lead to specific two-primer PCR reaction that is selective for one parent.

The method is based on the assumption that an RAPD-band which is specific to a single parent in a Southern analysis (figure 2A, B) marks a sequence divergence between the two parents at that genomic location with respect to the given primer sequence. If that is so, cloning followed by partial sequence determination of such RAPD-bands from either parent should lead us to sufficient sequence information in the vicinity of putative primer binding sites based on which one could design specific and classic two-primer PCR markers. We tested this for a marker P2 that was specific for CC125 (mt^+ strain) (figure 2). This eventually lead us to the desired goal where, based on P2, we could design a CC125 specific PCR-marker using two specific primers (figure 6).

In the process, we found that, although P2 was a genuine marker with respect to genetic segregation, it would not align linearly to its original parental sequence from the chloroplast genome. The sequence was significantly "scrambled up" during the PCR reaction. The sequence scrambling was so severe that we could not align the sequence linearly to the target region (encompassing 3′-end of 23S, full 5S and intergenic region between 5S and psbA). Several Southern probing and PCR experiments led us to conclude that indeed P2 could have originated from the target mentioned above (figures 2D, 3A, B, 5A, B). Sequence alignments, by a program, essentially led us to the same target region (figure 4B, C, bottom panels). Several studies, reported in the past, have documented PCR mediated sequence rearrangements (Odelberg *et al* 1995; Loewen and Switala 1995; Judo *et al* 1998; Viswanathan *et al* 1999). Such rearrangements essentially involved either complex translocations of the polymerase from one strand to the other (switching in *trans*) or from one region of the same strand to another (switching in *cis*) or a combination of both. It is also proposed that stable secondary structures formed by inverted repeats are the primary source of switching by the polymerase. The propensity of such a switch in *cis* is accentuated further when invert repeats are flanked by direct repeats to bracket the strong stem-loop structures (Ji *et al* 1994). Classically, it is this design that results in "copy choice replication" (Canceill and Ehrlich 1996). Chloroplast genomes are unique in being able to efficiently cross-talk where, within a single chloroplast, a few hundred circular DNA molecules share their genetic information with one another physically. This process leads to efficient "homoplasmy". It is believed that

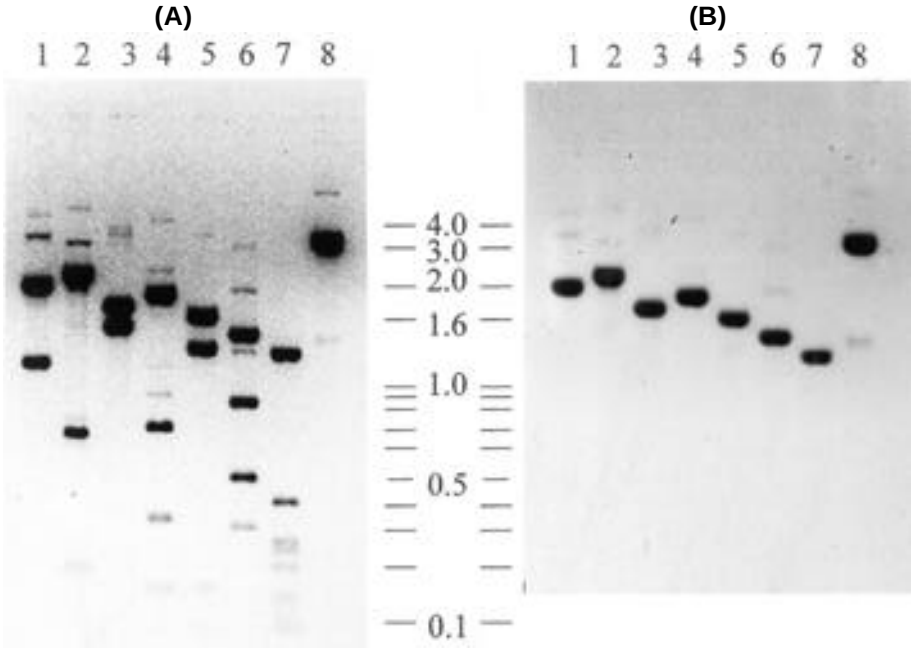

Figure 5. Southern mapping of P2 sequences on p124 genomic insert. The purified chloroplast genomic insert from P124 clone was digested with various enzymes and revealed by ethidium bromide staining of an agarose gel (**A**) or Southern analysis of the same using P2 as a probe (**B**). *Xho*I (lane 1), *Sau3*A (lane 2), *Sac*II (lane 3), *Hpa*II (lane 4), *Hae*III (lane 5), *Dra*I (lane 6), *Alu*I (lane 7), undigested purified insert control (lane 8). The sizes corresponding to the standard markers are indicated.

the robust machinery that promotes the high rate of homologous recombination, coupled with high frequency of "copy choice" replications between DNA molecules lead to such efficient "homoplasmy" (Boynton *et al* 1992). As mentioned earlier, "copy-choice" replication models envisage that a target DNA rich with direct and invert repeats has a higher propensity to generate stable secondary structures that are more amenable for polymerase switching (Canceill and Ehrlich 1996).

Fine mapping (Southern hybridization and high stringency sequence alignment) of P2 sequences in P124 insert revealed that most of the P2 sequences originated from a portion of P124 (figure 5A, B) that houses a cluster of direct as well as inverted repeats (figure 4B, C). The repeat domain encompassed a stretch of 600 nts between 1800–2400 bp. Correlation maps revealed a similar density of direct and invert repeats within this zone. At a DNA sequence level, this region belonged to an intergenic spacer between *5S* and *psbA* genes (Harris 1989). We conducted a similar correlation analyses on few other chloroplast and nuclear genes whose sequences were available in the database (data not shown). Sequences in *atpA* gene showed hardly any clustering of either direct or invert repeats whereas that in rDNA regions showed a high density of both direct as well as invert repeats. However, even in the rDNA region, the clustering of direct and invert repeats was essentially non-overlapping and distributed more uniformly across the entire region (data not shown). Similarly, *arg* locus in the nuclear genome showed several direct as well as invert repeat rich regions that were essentially non-overlapping and distributive in pattern. Amongst the sequences we surveyed, only P124 sequences exhibited this unique feature where both direct and invert repeat rich regions overlapped with each other, thereby revealing that the same portion of target was intimately involved in a secondary structural organization that involved both types of repeat elements. Here we can draw a parallel between the requirement of "copy-choice" replication competent target versus the repeat sequence features of P124. Both are endowed with a combination of direct as well as invert repeats that facilitate polymerase jumping from one site to the other. A unique feature of such "copy-choice" replication is that the products have been denuded of direct as well as invert repeats as a result of polymerase jumping from one point to an exact equivalent point in a direct repeat, bypassing an intervening invert repeat through a stem-loop structure (Canceill and Ehrlich 1996). Indeed, when we performed correlation analyses for the repeats in the product of PCR replication (namely P2), we found that P2 had lost both direct as well as invert repeat clusters. The sequence pattern in P2 was unremarkable, with no indications of any repeat richness. The product was consistent with a model of "copy-choice" replication. The examples available in the literature for "copy-choice" replication are limited and represent systems where a small number of polymerase slippages are involved (Ji *et al* 1994; Zhang *et al* 1994; Canceill and Ehrlich 1996; Zaphiropoulos 1998; Viswanathan *et al* 1999). Perhaps, the product described in P2 sequence represents a first case where a target sequence rich with appropriate repeat motifs (P124) gets "scrambled" extensively, but very reproducibly, through repetitive steps of polymerase slippage that result into a complex product of "copy-choice" replication. The PCR product is reminiscent of a unique "Hamiltonian Path" chosen during amplification events (Adleman 1994; Lee *et al* 1999). The scrambling of sequence information through such a domain via the proposed mechanism involving copy-choice replications is compounded further when the repeat-rich domain houses several putative sites for primer binding (figure 4D). In principle, many of these sites might initiate priming of replication on either strand independently, proceed through copy-choice mode, during which different amplicons might anneal with each other due to the repeat-rich nature of the region, generating a bigger product than the repeat domain itself. In the process, it is possible to visualize the emergence of an amplicon that has accumulated at such high concentrations that it competes with other amplicons in an exponential amplification, utilizing the same arbitrary primer at either end. P2 might represent one such unique product. In fact, it has been observed that *Chlamydomonas* chloroplast genome contains domains of repeat sequences at several locations that are rich in short sequences of highly conserved multiple direct and inverted repeat sequences

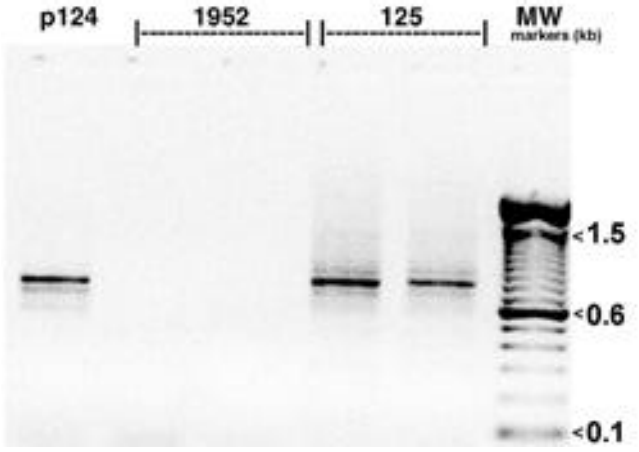

Figure 6. A PCR reaction between a primer from the upstream region of the repeat-domain (1550 nt location) (5′-GGTGTA-CAGTCAGCAATGGC-3′) and from the repeat-domain (2300 nt location) (5′-TTATTTGCTGCGCTAGCAGG-3′) was compared for CC125 and CC1952 genomic DNA (in duplicates). A positive control was done with P124 clone. Standard size marker positions are indicated.

(Leu 1998). The repeat domain discussed in this paper might well belong to such a family. The evolutionary and physiological significance of such extraordinary organizational elements in chloroplast genomes are not clear presently.

Acknowledgements

We would like to thank Mr Jatin and Mr Vijay of the Photography Section of TIFR for their help and the members of the laboratory for their valuable input during the course of this work. We would also like to thank TIFR for funding this research endeavour.

References

Adleman L M 1994 Molecular computation of solutions to combinatorial problems; *Science* **266** 1021–1024

Armbrust E V, Ferris P J and Goodenough U W 1993 A mating type-linked gene cluster expressed in *Chlamydomonas* zygotes participates in the uniparental inheritance of the chloroplast genome; *Cell* **74** 801–811

Boynton J E, Gillham N W, Newman S M and Harris E H 1992 Organelle genetics and transformation of *Chlamydomonas*; in *Cell organelles* (ed.) R G Hermann (New York: Springer-Verlag) pp 3–64

Bussieres J, Lemieux C, Lee R W and Turmel M 1996 Optional elements in the chloroplast DNAs of *Chlamydomonas eugametos* and *C. moewusii*: unidirectional gene conversion and co-conversion of adjacent markers in high-viability crossses; *Curr. Genet.* **30** 356–365

Canceill D and Ehrlich S D 1996 Copy-choice recombination mediated by DNA polymerase III holoenzyme from *Escherichia coli*; *Proc. Natl. Acad. Sci. USA* **93** 6647–6652

Dutcher S K 1995 Mating and tetrad analysis in *Chlamydomonas reinhardtii*; *Methods Cell Biol.* **47** 531–540

Gross C H, Ranum L P and Lefebvre P A 1988 Extensive restriction fragment length polymorphisms in a new isolate of *Chlamydomonas reinhardtii*; *Curr. Genet.* **13** 503–508

Haring M A, Schuring J, Urbans A, Musgrave H. van den Ende and Farris P J 1996 Random amplified polymorphic DNA markers (RAPDs) as a tool for gene mapping in *Chlamydomonas eugametos*; *J. Phycol.* **32** 1043–1048

Harris E H 1989 *The Chlamydomonas sourcebook* (San Diego: Academic press)

Ji W, Zhang X Y, Warshamana G S, Qu G Z and Ehrlich M 1994 Effect of internal direct and inverted *Alu* repeat sequences on PCR; *PCR Methods Appl.* **4** 109–116

Judo M S B, Wedel A B and Wilson C 1998 Stimulation and suppression of PCR-mediated recombination; *Nucleic Acids Res.* **26** 1819–1825

Lee C M, Kim S W, Kim S M and Sohn U 1999 DNA computing the Hamiltonian path problem; *Mol. Cells* **9** 464–469

Lemieux C, Turmel M and Lee R W 1980 Characterization of Chloroplast DNA in *Chlamydomonas eugametos* and *C. moewusii* and its inheritance in hybrid progeny; *Curr. Genet.* **2** 139–147

Leu S 1998 Extraordinary features in the *Chlamydomonas reinhardtii* chloroplast genome: (1) rps2 as part of a large open reading frame; (2) A *C. reinhardtii* specific repeat sequence; *Biochim. Biophys. Acta* **1365** 541–544

Loewen P C and Switala J 1995 Template secondary structure can increase the error frequency of the DNA polymerase from *Thermus aquaticus*; *Gene* **164** 59–63

Mets L 1980 Uniparental inheritance of chloroplast DNA sequences in interspecific hybrids of *Chlamydomonas*; *Curr. Genet.* **2** 131–138

Newman S M, Harris E H, Johnson A M, Boynton J E and Gillham N W 1992 Nonrandom distribution of chloroplast recombination events in *Chlamydomonas reinhardtii*: evidence for a hotspot and an adjacent cold region; *Genetics* **132** 413–429

Nishimura Y, Misumi O, Matsunaga S, Higashiyama T, Yokota A and Kuroiwa T 1999 The active digestion of uniparental chloroplast DNA in a single zygote of *Chlamydomonas reinhardtii* is revealed by using the optical tweezer; *Proc. Natl. Acad. Sci. USA* **96** 12577–12582

Odelberg S J, Weiss R B, Hata A and White R 1995 Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I; *Nucleic Acids Res.* **23** 2049–2057

Saal B and Wrickle G 1999 Development of simple sequence repeat markers in rye (*Secale cereale* L.); *Genome* **42** 964–972

Sager R 1954 Mendelian and non-Mendelian inheritance of streptomycin resistance in *Chlamydomonas reinhardtii*; *Proc. Natl. Acad. Sci. USA* **75** 356–363

Sager R and Lane D 1972 Molecular basis of maternal inheritance; *Proc. Natl. Acad. Sci. USA* **69** 2410–2413

Sager R and Ishida M R 1963 Chloroplast DNA in *Chlamydomonas*; *Proc. Natl. Acad. Sci. USA* **50** 725–730

Sears B B and VanWinkle-Swift K 1994 The salvage/turnover/repair (STOR) model for uniparental inheritance in *Chlamydomonas*: DNA as a source of sustenance; *J. Hered.* **85** 366–376

Tatusova T A and Madden T L 1999 Blast 2 sequences – a new tool for comparing protein and nucleotide sequences; *FEMS Microbiol. Lett.* **174** 247–250

Turmel M, Limieux C and Lee R W 1981 Dispersive labeling of *Chlamydomonas* chloroplast DNA in ^{15}N-^{14}N density transfer experiments; *Curr. Genet.* **4** 91–97

Viswanathan V K, Krcmarik K and Cianciotto N P 1999 Template secondary structure promotes polymerase jumping during PCR amplification; *Biotechniques* **27** 508–511

Williams J G, Kubelik A R, Livak K J, Rafalski J A and Tingey S V 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers; *Nucleic Acids Res.* **18** 6531–6535

Zaphiropoulos P G 1998 Non-homologous recombination mediated by *Thermus aquaticus* DNA polymerase I. Evidence supporting a copy choice mechanism; *Nucleic Acids Res.* **26** 2843–2848

Zhang Z, Huang K, Zhang Y, Liu N and Yang K 1994 Interference by complex structures of target DNA with specific PCR amplification; *Appl. Biochem. Biotechnol.* **44** 15–20

MS received 27 July 2001; accepted 10 September 2001

Corresponding editor: MAN MOHAN JOHRI