A three-step model for the rearrangement of the chloroplast *trnK-psbA* region of the gymnosperm *Pinus contorta*

Jonas Lidholm* and Petter Gustafsson
Department of Plant Physiology, University of Umeå, S-901 87 Umeå, Sweden

Received March 18, 1991; Revised and Accepted May 14, 1991 EMBL accession nos X57097 – X57099 (incl.)

ABSTRACT

A region of the *Pinus contorta* chloroplast genome which contains a duplication of the *psbA* gene was characterized. From previous experiments it was known that the two copies of the *psbA* gene were located approximately 3.3 kilobase pairs (kbp) apart, that they had the same orientation and that one endpoint of the duplication was 19 base pairs (bp) downstream of the *psbA* stop codon. In order to determine the size and additional genetic content of the duplicated segment, both copies as well as the intervening DNA were sequenced completely. It was found that the duplicated segment was 1969 bp long, that the two copies were completely identical and were separated by 2431 bp. The duplicated segment carried, in addition to *psbA*, the 3′ exon of the *trnK* gene, which was partially included in a 124 bp direct repeat. The translocated copy of the duplicated segment was found to be inserted upstream of the *trnK*(UUU) gene and was immediately followed by a repeated 41 bp stretch from the *psbA* coding region. The *trnK* gene was split by a 2509 bp intron which contained an open reading frame of 515 codons. Sequence comparisons of the duplicated segment and its flanking DNA to the corresponding regions of *P. sylvestris*, a species which lacks the rearrangements found in *P. contorta*, made it possible to identify 3 – 9 bp homologies within which recombinations had occurred. A model was derived which would accommodate the conversion of a *trnK-psbA* locus of the ancestral, *P. sylvestris*-like organization into the rearranged structure found in *P. contorta*.

INTRODUCTION

The chloroplasts of green plants contain multiple copies of a 120 – 180 kbp circular genome. The highly conserved genetic content of chloroplast genomes comprises genes for 4 different rRNAs, 30 different tRNAs, some 45 polypeptides of known function and approximately 40 open reading frames (ORFs), from which gene products have not yet been detected or functionally assigned (1, 2, 3).

The chloroplast genomes of most plants are organized into two single-copy regions which are separated by two long inverted repeat segments. Whereas many chloroplast genomes are nearly or entirely colinear, those of some lineages, such as grasses (3), conifers (4, 5, 6) and certain legumes (7), are extensively rearranged, as compared to the dicot consensus structure. Despite numerous reports describing these and other alterations on a gross scale, relatively few have addressed the molecular mechanisms underlying their occurrence.

A variety of evidence has established that plastids are proficient in homologous recombination. However, this process may not be the most important for the generation of rearrangements, since chloroplast genomes are generally devoid of long repeated sequences other than the two large inverted repeat segments (7). Instead, most genome rearrangements are presumably the result of some kind of illegitimate recombination. It has been shown that an intermolecular recombination between two tRNA genes and a subsequent deletion could account for one of three inversions characteristic of the grass family (3), whereas the other two might have occurred through recombination between 11 – 16 bp direct repeats found in association with the inversion endpoints (8, 9). In a study where length mutations in the chloroplast genomes of three wheat species were examined, 5 and 9 bp direct repeats were found at the ends of two overlapping segments that were present in one of the species but alternatively lost in the others, an observation which strongly suggested that recombination between these repeats had caused the deletions (10).

The recombination events whereby rearrangements are formed constitute the molecular basis of large-scale genome evolution. Thus, to better understand the mechanisms of this process, it is important to identify and characterize sequences involved in recombination. In bacteria, this has been done by sequence alignments of novel joints in e.g. plasmid/phage recombinants with their parental DNAs (11, 12, 13). Since such a procedure is not feasible in the study of natural chloroplast DNA rearrangements, the analysis must rely on comparisons to other, unrearranged genomes. However, since most rearrangements which are retained do not disrupt genes or cotranscribed gene clusters (5, 14, 15), their endpoints lie in intergenic regions where the sequence evolution is unconstrained. Chloroplast DNA rearrangements which are characteristic of major plant groups

* To whom correspondence should be addressed

and therefore date far back in time, may for this reason not allow unambigous identification of recombination sites. This problem was realized when attempts were made to find inversion endpoints in rice by aligning rice and tobacco sequences (3, 9). More suitable for the purpose are genome alterations which occur among closely related taxa and thus are likely to be of recent origin. In such instances appropriate interspecific sequence alignments are easier to make and recombination sites can be distinguished at a high level of accuracy and resolution. The analysis of length mutations among wheat species (10) is an example of such a study.

We have previously shown that a chloroplast DNA segment carrying the *psbA* gene is translocationally duplicated in two closely related pine species (16). Here we report the results of a comparative sequence analysis of one of these pine species and another, which lacks the duplication. The aim of the study was to characterize the rearrangement and to identify the sequence elements which were involved in its formation.

MATERIALS AND METHODS

Chloroplast DNA from *P. sylvestris*, prepared as described (17), was generously provided by Dr. A. E. Szmidt. The clones of chloroplast DNA from *P. contorta* used in this study, pPCB121, pPCH157 and pPCB932, have been described previously (16). Templates for sequencing were either subcloned restriction or sonication fragments, or deletion clones generated by exonuclease *Bal*31. Cloning and sequencing were performed using standard procedures as described (16). Computer analysis of sequence data was carried out using the UWGCG software package (18).

RESULTS

Determination of the endpoints and genetic content of the duplicated segment

We recently reported that the *psbA* gene in *Pinus contorta* resides on a duplicated segment of the chloroplast DNA (16). It was shown that one endpoint of the duplication was located 19 bp downstream of the *psbA* stop codon and that the duplication extended to somewhere between 0.7 and 1.2 kb upstream of the start codon. In order to determine the upstream endpoint and additional genetic content of the duplicated segment we sequenced the region between the two *psbA* copies, along with that upstream of *psbAI* (Figures 1, 2A). These sequences were merged with those of *psbAI* and *psbAII*, which had previously been determined, to form a continuous, 7411 bp sequence of the region covered by clones pPCB121, pPCH157 and pPCB932 (Figure 1, 2A). It was found that the sequences upstream of the two gene copies were identical to a position 888 bp upstream of the start codon, beyond which there was no homology. Hence, the total size of the duplicated segment, including the *psbA* gene and the 19 bp of downstream sequence, was 1969 bp (Figure 2A). The 35 bp sequence between positions 207 and 241 of the duplicated segment was identified as the 3′ exon of the *trnK* gene (Figures 1 and 2A). The location of this gene upstream of *psbA* was the same as in other land plants, in which *trnK* is also a split gene. In *P. contorta*, the 23 last nucleotides of the *trnK* 3′ exon were found to constitute the first part of a 124 bp tandem repeat (Figures 2A and 3-II), i.e. a short tandem repeat was present within the large duplicated segment.

Genetic identity of the integration site

In our previous paper we showed that *psbAII* was located at the ancestral location upstream of the *trnH* gene, whereas *psbAI* was translocated to a novel position. To determine the genetic identity of the integration site for this translocation, we analyzed the sequence flanking the *psbAI*-containing copy of the duplicated segment. At a distance of 91 bp downstream of this segment, the 5′ exon of the split *trnK*(UUU) gene was found (Figures 1 and 2A). The intron separating the 5′ exon of *trnK* from the 3′ exon, located upstream of *psbAII*, was found to be 2509 bp long and to encompass an open reading frame of 515 codons (ORF515; Figures 1 and 2A). In total, the two copies of the duplicated segment were separated by 2431 bp. The structure of the pine *trnK* gene was very similar to that of other land plants (1, 3, 19, 20), e.g. mustard which has a 2574 bp long intron containing an open reading frame of 524 codons (20). The amino acid sequence deduced from ORF515 was 42–45% identical to the corresponding sequences from other plants.

Sequence elements involved in the rearrangements

In order to identify the recombination sites in the *trnK-psbA* region in *P. contorta*, we decided to analyze the corresponding region of another pine species, *Pinus sylvestris* (Scots pine), which had been shown to lack the *psbA* duplication (16). Relevant chloroplast DNA fragments were cloned (Figure 1) and the regions equivalent to the upstream part of the large duplicated segment and to the integration site upstream of the *trnK* gene were sequenced (Figure 2B).

The tandem duplication of the trnK-containing segment. The second copy of the tandemly duplicated 124 bp segment in *P. contorta* was immediately followed by a third copy of its first 7 nucleotides, CGGGTTC. Hence, the duplication appeared as two 117 bp segments demarcated, and separated, by 7 bp direct repeat units (the three copies are designated **A**, **A/A′** and **A′**, respectively, in Figure 3-II). In *P. sylvestris*, the corresponding segment, which was 122 bp long, was present as a single-copy element, flanked by the same 7 bp direct repeats as in *P. contorta* (designated **a** and **a′**; Figure 3-II). This suggests that the tandem duplication in *P. contorta* was created by intermolecular recombination between the **A** and **A′**.

The left endpoint of the large psbA-containing segment. An alignment of the integration site sequences of the two pine species revealed that a 226 bp sequence was present in *P. sylvestris* in place of the large duplicated segment in *P. contorta* (Figure 2, 3-I). Thus, it appears that the integration of the duplicated segment was not a simple insertion but rather a replacement event.

The positions of two copies of an 8 bp sequence in *P. sylvestris* (AATAGAAA, designated **b'** upstream of the *trnK* 5′ exon and **b** upstream of the *trnK* 3′ exon; Figure 3-I, II), were found to coincide precisely with the virtually identical first 9 bp of the duplicated segment in *P. contorta* (AATAGGAAA, designated **B′** and **B**, respectively; Figure 3-I, II). The high level of sequence identity between the two pine species suggests that both of the **B** and **B′** elements in *P. contorta* were present at the same positions before the duplication occurred and that the recombination at the left border of the large duplicated segment took place within these sequences.

A secondary, overlapping rearrangement. The 41 bp stretch of DNA that immediately followed the duplicated segment in *P.*

contorta was found not to match the *P. sylvestris* sequence at the integration site (Figure 3-I) but was instead identical to an internal part of *psbA* (Figure 3-IV). This 41 bp stretch was flanked by identical short sequences at its two locations; by TAT at its left side (designated **D** within *psbA* and **D′** downstream of *psbAI*; Figure 3-I, IV) and by CCTC at its right side (designated **E** and **E′**, respectively; Figure 3-I, IV). Including these flanking sequences the length of the short duplicated segment was 48 bp. The TAT sequence (**D′**) appeared as the last three bp of the large duplicated segment (Figure 3-III) and the CCTC sequence (**E′**) as the first four bp of homology between *P. contorta* and *P. sylvestris* at the downstream side of the integration site (Figure 3-I). Thus, it appears that a translocational duplication of a short segment from the *psbA* coding region occurred subsequent to the insertion of the large segment, as a result of recombination across the flanking **D/D′** and **E/E′** sequences. This secondary rearrangement apparently spanned the original right border of the large duplication and hence removed that recombination site.

The right endpoint of the large psbA-containing segment. We wanted to further examine the possibility that the large duplication originally included DNA downstream of the observed right endpoint, as indicated above. To do this we searched for homologies between the region downstream of *psbAII* in *P. contorta* and the 226 bp sequence upstream of the *trnK* 5′ exon in *P. sylvestris*, corresponding to the segment lost from the integration site in *P. contorta*. Such homologies, if found, could be considered as possible downstream recombination sites in a primary insertion of the *psbAI*-containing segment. The search resulted in identification of three alternative, potentially relevant homologies, designated $\mathbf{C_1}$, $\mathbf{C_2}$ and $\mathbf{C_3}$ in *P. contorta* (Figure 3-III) and $\mathbf{c_1}'$, $\mathbf{c_2}'$ and $\mathbf{c_3}'$ in *P. sylvestris* (Figure 3-I). The $\mathbf{C_1}/\mathbf{c_1}'$ sequences would form a 12 bp hybrid with a bulge of one nucleotide in the $\mathbf{C_1}$ strand, the $\mathbf{C_2}/\mathbf{c_2}'$ pair would form an 11 bp hybrid with one mismatch and $\mathbf{C_3}/\mathbf{c_3}'$ would form a 9 bp perfect match. However, since the sequences making up the pairs were collected from two different species, the resolution of the comparison to the level of single nucleotides could be misleading. We therefore consider the three homologies equally probable as primary recombination sites.

A three-step model for the rearrangement of the *trnK-psbA* locus

The anomalous structure of the *trnK-psbA* region in *P. contorta* was found to be the result of at least three separate rearrangements: the translocational duplication of the large segment carrying the *psbA* gene, the tandem duplication of a 124 bp stretch containing a part of the 3′ exon of *trnK*, and a replacement of the primary downstream border of the large duplication by a short segment of *psbA* coding sequence.

The sequence elements that appear to have been engaged in the recombination events have been discussed in detail above. The sequential order of the rearrangements can be deduced from the present organization of the region. Since the tandem duplication of the *trnK*-containing 124 bp stretch was found in both copies of the large *psbA*-containing segment, this event was probably the first to take place, whereas the last must have been the one which placed the short *psbA* coding sequence downstream of *psbAI*. Because the two segment replacements both appeared as non-reciprocal, i.e. the donor site was left intact, they, as well as the 124 bp tandem repeat, were most likely generated by intermolecular recombinations, followed by elimination of the donor molecules from the population.

These findings and conclusions are summarized in a model (Figure 4) describing three consecutive, intermolecular recombination events which together would accommodate the conversion of a *trnK-psbA* locus such as that of *P. sylvestris* into the rearranged structure found in *P. contorta*.

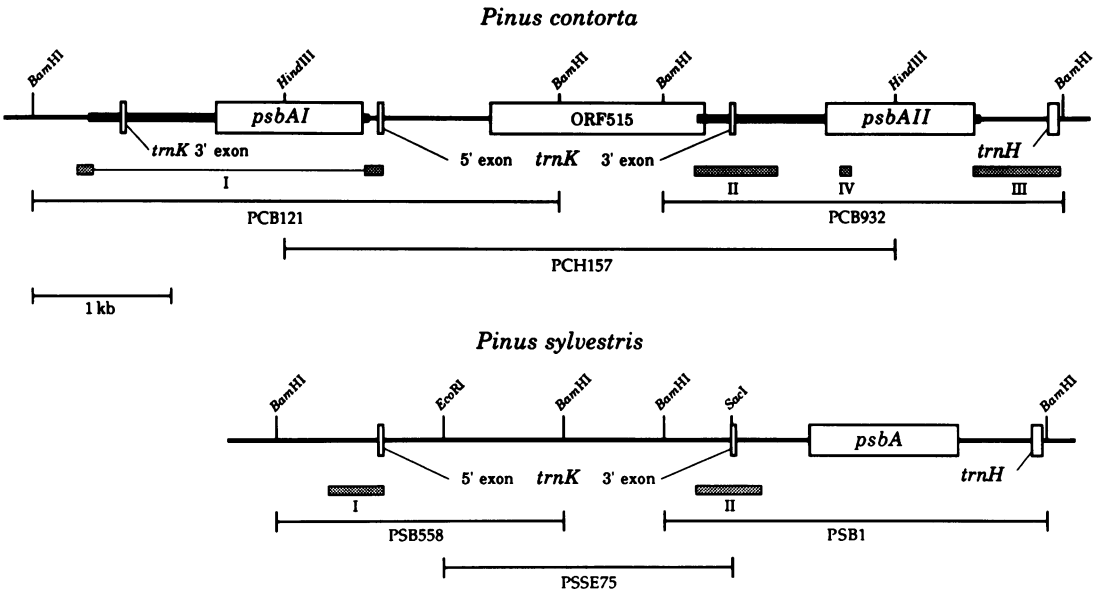

Figure 1. Maps of the chloroplast DNA region of *Pinus contorta* which carries the 2.0 kb duplication and the equivalent region of the *P. sylvestris* chloroplast genome. The inserts of clones pPCB121, pPCH157 and pPCB932 from *P. contorta* and pPSB558, pPSSE75 and pPSB1 from *P. sylvestris* are shown below the genomic representations. The two copies of the large duplicated segment containing the *psbA* gene are indicated by heavy lines in the *P. contorta* map. The *trnK* intron of *P. sylvestris* was not completely sequenced and the putative ORF is therefore not indicated. The direction of transcription of the *psbA* and *trnK* genes is from the left to the right. The filled boxes and roman numerals below the maps indicate the origin of the sequences used in Figure 3.

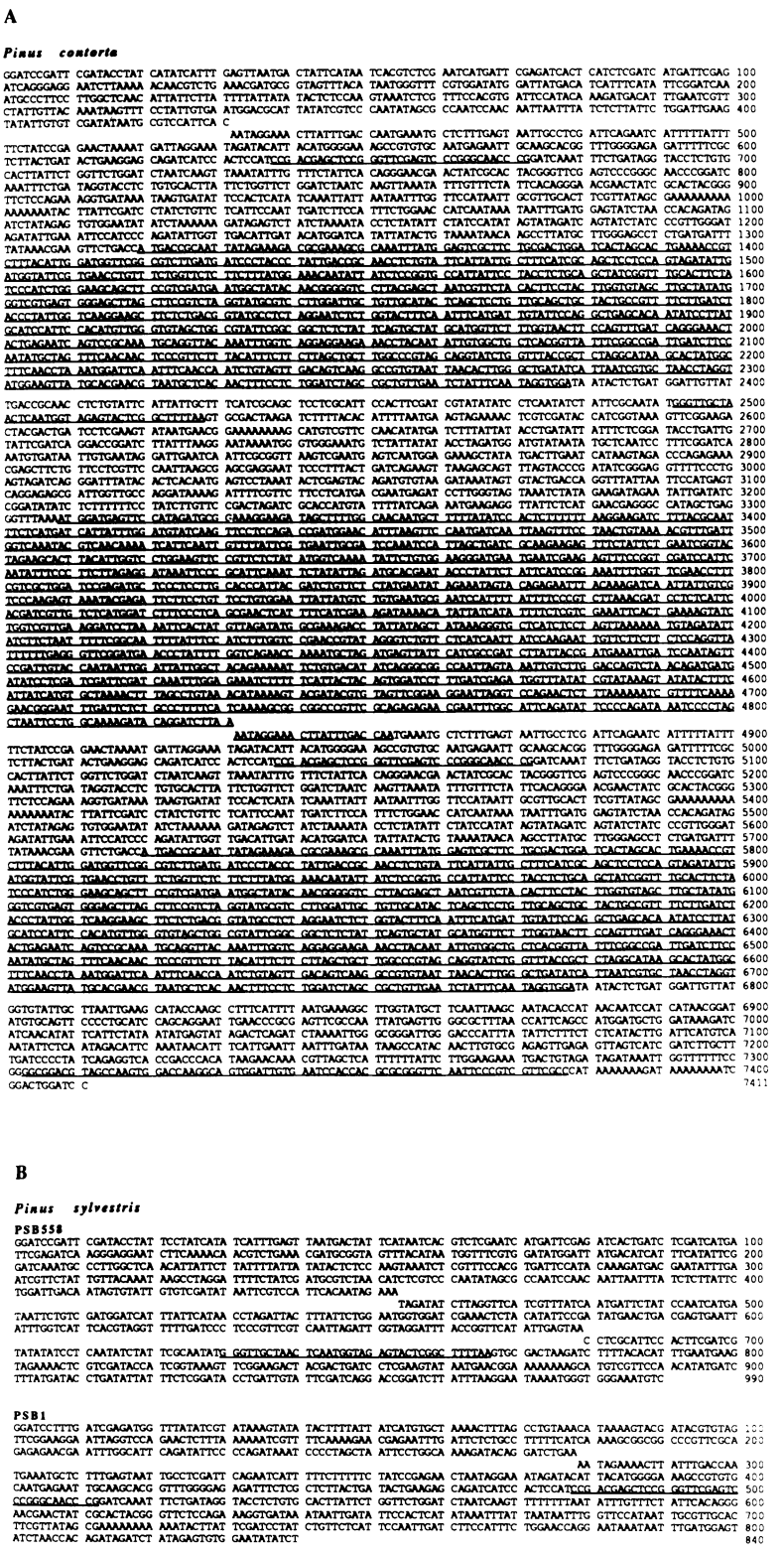

Figure 2. A: Complete DNA sequence from the region covered by the *P. contorta* clones pPCB121, pPCH157 and pPCB932. Breaks in the sequence indicate the large duplicated segment (positions 432−2400 and 4832−6800). The underlining marks the translocated *trnK* 3′ exon (position 638−672), *psbAI* (position 1320−2378), the *trnK*(UUU) 5′ exon (position 2492−2528), ORF515 (position 3309−4853), the *trnK* 3′ exon (position 5038−5072), *psbAII* (position 5720−6778) and *trnH*(GUG) (position 7303−7377). The *P. contorta* sequence between positions 813 and 2518, and from position 5213 to the end was part of a previous report (11) but is included here to give a better overview and to facilitate and correlation to Figure 3. B: Partial DNA sequences from the *P. sylvestris* clones pPSB558 and pPSB1, starting at the left *Bam*HI site of each clone. The breaks in the sequence of PSB558 indicate the 226 bp segment (position 454−679) missing in *P. contorta*. The underlining marks the *trnK*(UUU) 5′ exon (position 730−766). In the PSB1 sequence, the break after position 278 indicates the position equivalent to the left endpoint of the large duplicated segment in *P. contorta*. The underlined sequence shows the *trnK* 3′ exon (position 478−512).

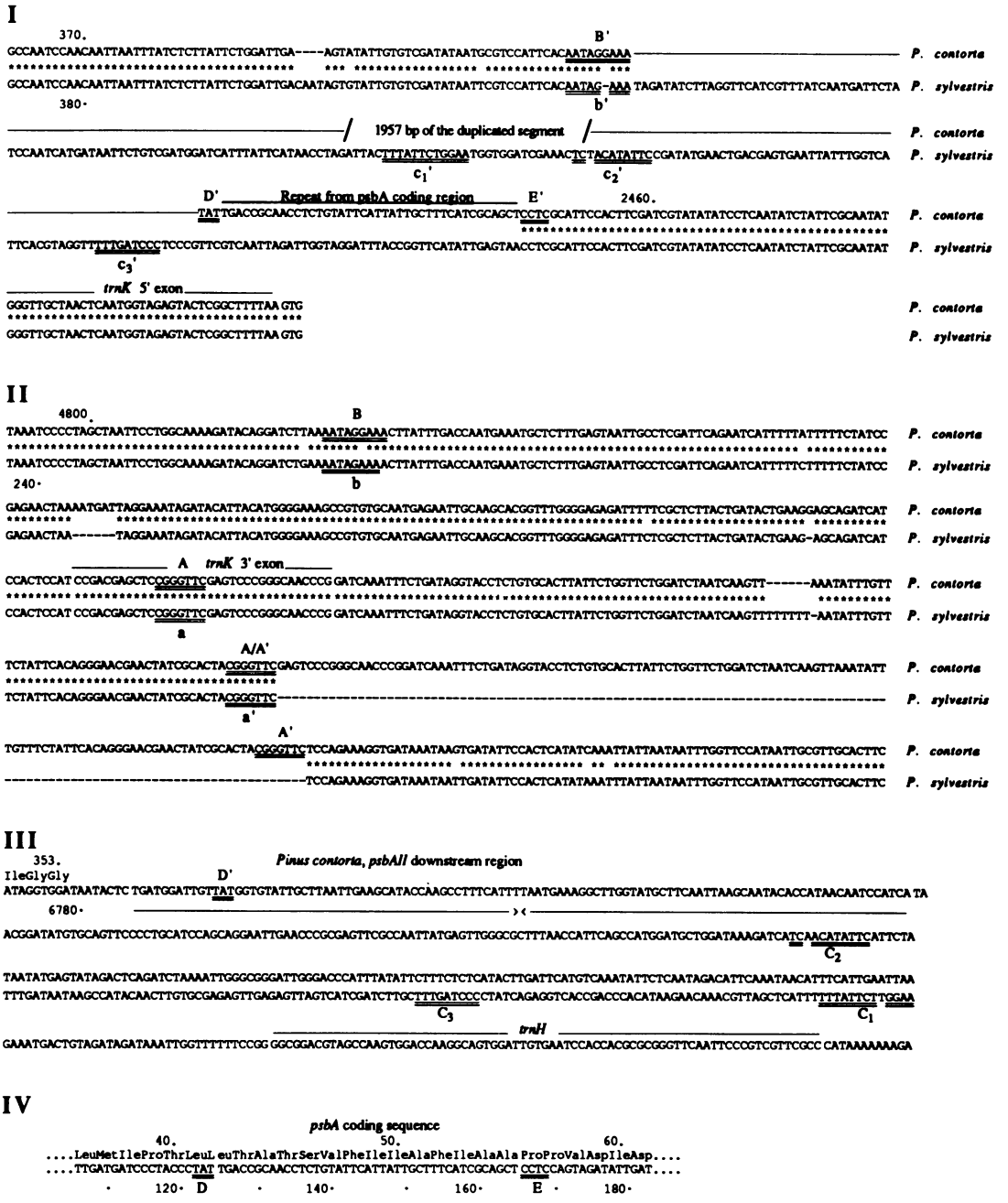

Figure 3. Extractions and alignments of sequences from Figure 2 on which the model presented in Figure 4 is based. **I**: Alignment of the sequence of the insertion site for the translocational duplication in *P. contorta* with the sequence of the equivalent region of *P. sylvestris*. **B'** indicates the first nine nucleotides of the duplication whereas **b'** is an almost identical sequence in *P. sylvestris*. An identical copy of **b'** (marked **b** in alignment II) is present upstream of the *trnK* 3' exon, at the position which corresponds to the left extreme of the duplicated segment in *P. contorta*. The whole of the *psbAI*-containing copy of the duplicated segment in *P. contorta*, except for the first nine and the last three nucleotides, is shown as horizontal lines. The *P. sylvestris* homologues of three possible alternatives for **C'**-sites (see part III) are indicated. The corresponding *P. contorta* sequence containing these sites was, according to our model, removed by a rearrangement that occurred subsequent to the large duplication. The left and right endpoints of this secondary duplication/replacement are marked by **D'** and **E'**, respectively. Hyphens (-) indicate 'missing' nucleotides ; asterisks (*) indicate positions where the aligned sequences are identical. The nucleotide numbering refers to that in Figure 2. **II**: Alignment of the sequences upstream of *psbA* of *P. contorta* (*psbAII*) and *P. sylvestris*. The element marked **B** constitutes the first 9 bp of the duplicated segment in *P. contorta* and may have been the substrate for recombination at the left border of the duplication. The equivalent *P. sylvestris* element is marked **b**. **A** and **A'** are identical 7 bp stretches which border the 124 bp tandem repeat in *P. contorta* and which may have served as recombination substrates in this small duplication. The *P. sylvestris* elements equivalent to **A** and **A'** are marked **a** and **a'**, respectively. Hyphens (-) indicate 'missing' nucleotides; asterisks (*) indicate identical nucleotides. The nucleotide numbering refers to that of Figure 2. **III**: The *P. contorta* sequence of the region downstream of *psbAII* including the last three amino acid residues of the D1 polypeptide. The short **D'** sequence, which constitutes the last three nucleotides of the large duplication and which appears to have been the left recombination site for the secondary rearrangement downstream of *psbAI* is indicated, as well as three possible recombination sites for the primary right endpoint of the large duplication (C_1-C_3). The head-to-head arrows show a dyad symmetry, potentially conferring an RNA stem-loop structure. The nucleotide numbering refers to that of Figure 2. **IV**: A part of the *psbA* coding sequence which includes the stretch that was duplicated in the secondary rearrangement downstream of *psbAI*. According to our model, this event concommitantly removed the primary right border of the large duplicated segment. The endpoints of this duplicated/translocated stretch are indicated by **D** and **E**. The nucleotide numbering refers to the *psbA* coding sequence, starting at the initiator ATG.

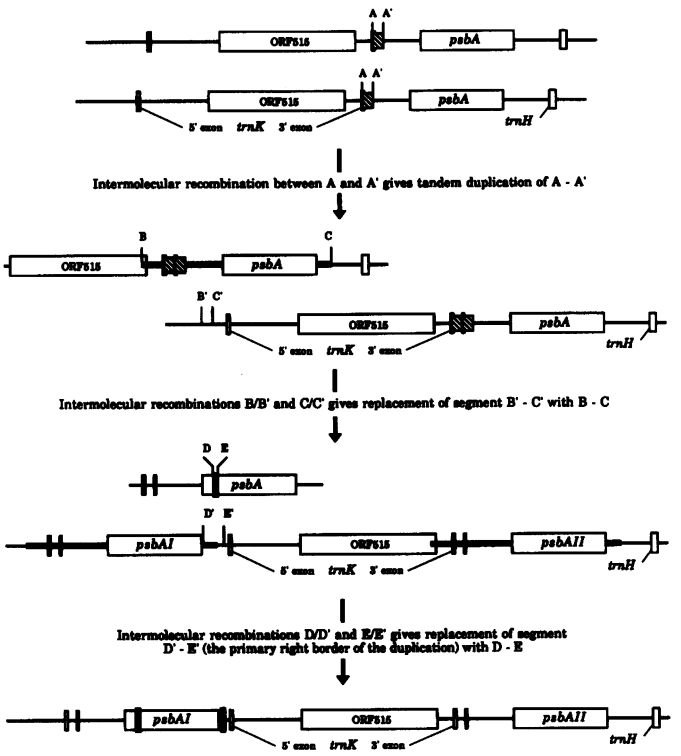

Figure 4. A proposed model for three consecutive rearrangement events which would account for the conversion of a *trnK-psbA* region of *P. sylvestris*-like structure to that of *P. contorta*. At each stage, the segment to be involved in the next rearrangement is shown as a filled box bordered by vertical bars which indicate the specific sequence elements that could act as substrates for the subsequent recombination process. The alphabetical designations of these sequences refer to those used in Figure 3. In all, the rearrangements involve the duplication of segments A-A', B-C and D-E and deletion of segments B'-C' and D'-E'.

DISCUSSION

In a recent report we showed that the *psbA* gene is translocationally duplicated in two closely related pine species but present as a single-copy gene in six others (16). In this rearrangement, we saw an opportunity for a recombination site analysis which would not be hampered by the problem of mutational obliteration and for which an unrearranged genome, suitable for sequence comparison, was available.

DNA sequences from two pine chloroplast genomes were analyzed in the study; that of *Pinus contorta* which contains the gene duplication and that of *P. sylvestris* which contains a single-copy *psbA* gene. The analysis revealed that the *psbA* region of *P. contorta* actually held three rearrangements, none of which was present in *P. sylvestris*. A copy of a 2 kbp segment, containing the *psbA* gene and the downstream part of the split *trnK* gene, was found to be inserted 2.4 kbp upstream of its original position, in front of the *trnK*(UUU) gene. Within the large duplicated segment, a small tandem repeat was present and at the downstream border of the translocated copy, an overlapping segment replacement appeared to have removed its original endpoint. The temporal order of the rearrangements was evident from the structure itself. By various comparisons within the *P. contorta* sequence and to the equivalent regions of *P. sylvestris*, we could unambigously identify four of five inferred recombination sites, while one had been lost due to the secondary, overlapping rearrangement. However, based on appropriate comparisons to the corresponding *P. sylvestris* sequence, three alternative homologies were distinguished one of which may represent the subsequently lost recombination site.

The recombination sites unambigously identified in this study were 3 to 9 bp direct repeats. No conspicuous sequence similarity was observed among these repeats or to any of the previously reported repeats in Douglas-fir (21), liverwort (22), *Oenothera* (23, 24), wheat (10, 25, 26) and rice (9), which have been found in association with endpoints of dispersed repeats or rearrangements in the chloroplast DNA of these species.

None of the repeats between which recombinations have occurred seem to be of sufficient length to direct classical homologous recombination, a process which in *E. coli* requires 40–50 bp of homology (27). In bacteria, non-homologous recombination can occur by a number of different mechanisms (13). Some of these are mediated by dedicated enzymes and represent specific processes such as phage integration and excision, transposon movement and 'programmed' inversions (13). This type of recombination normally uses specific sites and is then categorized as legitimate. Recombination can also result from incidental events in other processes which involve breakage and joining of DNA strands, and in DNA replication. This is referred to as illegitimate recombination and occurs usually, but not necessarily, between short homologous sequences, without any distinct site preference. The fact that there was no sequence similarity among the recombination sites found in this study, or to any of the repeats from other species, suggests that the rearrangements described here were created by non-specific, illegitimate recombination events, such as those mediated by DNA gyrase activity in *E. coli* (11, 12). In contrast, a striking resemblance of several of the previously reported repeats to the bacteriophage lambda attachment site in *E. coli* has been observed (see compilation in ref. 21). This might reflect the existence of other, sequence-specific mechanisms of recombination in plastids.

The rearrangements in the *trnK-psbA* region of *P. contorta* appear to result from both single and double cross-overs. In the structure generated by the first of the rearrangements, the 124 bp tandem duplication, the second copy of the segment was immediately followed by a third copy of its first 7 bp. Similar situations have been found in *Vicia faba* (28) and in another part of the *P. contorta* genome (Lidholm, unpublished observation). In *P. sylvestris*, the corresponding segment was present as a single-copy element flanked by 7 bp direct repeats. A tandem duplication of this segment could occur by a single cross-over between the upstream repeat of one genome molecule and the downstream repeat of another, followed by an intramolecular homologous recombination at an arbitrary site to resolve the resulting dimer of the genome. Alternatively, the segment could first be excised as a circular molecule by intramolecular recombination between the direct repeats and subsequently integrated into an intact copy of the genome by homologous recombination. The two other rearrangements both seem to result from intermolecular, double cross-overs between short sequences of homology on either side of the replaced segments (i.e. unequal segment exchange).

Among the rearrangements that are found in chloroplast genomes, duplications of large segments, in the size range of one kbp or longer, seem to be particularly rare. In the case of tandem duplications, the reason for this is probably that they are readily eliminated by intramolecular homologous recombination. However, when the copies of a duplicated segment have the same orientation but are not juxtaposed, homologous recombination

between the copies will delete the intervening region of the genome. In *P. contorta*, a single recombination between the copies of the large duplicated segment would cause the loss of *trnK*/ORF515. Lethality of such a mutation may explain the apparent stability of the rearrangement.

In the analysis of the integration site region of the two pine species, an observation was made which raises the possibility that the duplicated segment in *P. contorta* in fact performs an active function at its new position. The only canonical promoter structure (TTGACA-N_{17}-TATAAT) for the *trnK* gene in *P. sylvestris* was found 297 bp upstream of its 5′ exon, 20 bp upstream of the 226 bp sequence which is replaced by the large duplicated segment in *P. contorta*. In the equivalent promoter structure of *P. contorta* four bp were found to be deleted, starting at the C-residue of the putative −35 box. The resulting structure, TTGAAG-N_{13}-TATAAT, is probably not a functional promoter. Furthermore, several potential stem-loop structures following the *trnK* 3′ exon, located on the duplicated segment, might act as transcription terminators (20). However, 1.2 kbp upstream of the *trnK* 5′ exon, the intact *psbAI* promoter is present, and since there is no other promoter-like structure between this and *trnK*, transcription of the *trnK* gene might be driven by the *psbAI* promoter. We intend to pursue a transcription analysis of the *trnK-psbA* regions of *P. contorta* and *P. sylvestris* to examine this possibility.

In this context, it is interesting to note that whichever of the three alternative **C** elements may have been the original right endpoint of the duplication, the extensive dyad symmetry downstream of *psbA* (Figure 3-III) would have been included in the duplicated segment. The consequence of this would have been that a strong mRNA hairpin structure was present between *trnK* and the *psbAI* promoter, on which it possibly relied for its transcription. Thus, transcriptional dependence of *trnK* on the *psbAI* promoter may have been a functional rationale for the last rearrangement, which removed the original downstream part of the duplicated segment.

ACKNOWLEDGEMENTS

The skillful technical assistance of Cathrine Persson in parts of the sequencing work is gratefully acknowledged. We thank Dr. A.E.Szmidt and colleagues for the gift of *P. sylvestris* chloroplast DNA. This work was supported by grants from the Swedish Council for Forestry and Agricultural Research, The Swedish Natural Science Research Council and The Bo Rydin Foundation for Scientific Research.

REFERENCES

1. Shinozaki, K., Ohme, M., Tanaka, M., Wakazugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) *EMBO J.*, **5**, 2043−2049.
2. Ohyama, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1988) *J. Mol. Biol.*, **203**, 281−298.
3. Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) *Mol. Gen. Genet.*, **217**, 185−194.
4. Lidholm, J., Szmidt, A.E., Hällgren, J.-E. and Gustafsson, P. (1988) *Mol. Gen. Genet.*, **212**, 6−10.
5. Strauss, S.H., Palmer, J.D., Howe, G.T. and Doerksen, A.H. (1988) *Proc. Natl. Acad. Sci. USA.*, **85**, 3898−3902.
6. White, E.E. (1990) *Theor. Appl. Genet.*, **79**, 119−124 .
7. Palmer, J.D. (1985) In MacIntyre R.J. (ed.), Monographs in Evolutionary Biology: Molecular Evolutionary Genetics. Plenum Press, New York, pp131−240.
8. Howe, C.J., Barker, R.F., Bowman, C.M. and Dyer, T.A. (1988) *Curr. Genet.*, **13**, 343−349.
9. Shimada, H. and Sugiura, M. (1989) *Curr. Genet.*, **16**, 293−301.
10. Ogihara, Y., Terachi, T. and Sasakuma, T. (1988) *Proc. Natl. Acad. Sci. USA.*, **85**, 8573−8577.
11. Marvo, S.L., King, S.R. and Jaskunas, S.R. (1983) *Proc. Natl. Acad. Sci. USA.*, **80**, 2452−2456.
12. Naito, A., Naito, S. and Ikeda, H. (1984) *Mol. Gen. Genet.*, **193**, 238−243.
13. Ehrlich, S.D. (1989) In Berg, D.E. and Howe, M.M. (eds.), Mobile DNA. Americam Society for Microbiology, Washington, D.C., pp. 799−832.
14. Michalowski, C., Breunig, K.D. and Bohnert, H.J. (1987) *Curr. Genet.*, **11**, 265−274.
15. Woodbury, N.W., Roberts, L.L., Palmer, J.D. and Thompson, W.F. (1988) *Curr. Genet.*, **14**, 75−89.
16. Lidholm, J., Szmidt, A.E. and Gustafsson, P. (1991) *Mol. Gen. Genet.*, in press.
17. Szmidt, A.E., Lidholm, J. and Hällgren, J.-E. (1986) In Lindgren, D. (ed.), Provenances and Forest Tree Breeding for High Latitudes. Proceedings of the Frans Kempe symposium, Umeå, pp. 269−280.
18. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucl. Acids Res.*, **12**, 387−395.
19. Umesono, K., Inokuchi, H., Shiki, Y., Takeuchi, M., Chang, Z., Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K. and Ozeki, H. (1988) *J. Mol. Biol.*, **203**, 299−331.
20. Neuhaus, H. and Link, G. (1987) *Curr. Genet.*, **11**, 251−257.
21. Tsai, C.-H. and Strauss, S.H. (1989) *Curr. Genet.*, **16**, 211−218.
22. Zhou, D.X., Massenet, O., Quigley, F., Marion, M.J., Monéger, F., Huber, P. and Mache, R. (1988) *Curr. Genet.*, **13**, 433−439.
23. wom Stein, J. and Hachtel, W. (1988) *Curr. Genet.*, **13**, 191−197.
24. wom Stein, J. and Hachtel, W. (1988) *Mol. Gen. Genet.*, **213**, 513−518.
25. Howe, C.J. (1985) *Curr. Genet.*, **10**, 139−145.
26. Quigley, F. and Weil, J.H. (1985) *Curr. Genet.*, **9**, 495−503.
27. Smith, G.R. (1988) *Microbiol. Rev.*, **52**, 1−28.
28. Bonnard, G., Weil, J.-H. and Steinmetz, A. (1985) *Curr. Genet.*, **9**, 417−422.