Mol Gen Genet (1996) 252:195–206

ORIGINAL PAPER

Sarah E. Goulding · Richard G. Olmstead
Clifford W. Morden · Kenneth H. Wolfe

Ebb and flow of the chloroplast inverted repeat

Received: 1 February 1996 / Accepted: 9 May 1996

Abstract The endpoints of the large inverted repeat (IR) of chloroplast DNA in flowering plants differ by small amounts between species. To quantify the extent of this movement and define a possible mechanism for IR expansion, DNA sequences across the IR–large single-copy (IR–LSC) junctions were compared among 13 *Nicotiana* species and other dicots. In most *Nicotiana* species the IR terminates just upstream of, or somewhere within, the 5′ portion of the *rps19* gene. The truncated copy of this gene, *rps19′*, varies in length even between closely related species but is of constant size within a single species. In *Nicotiana*, six different *rps19′* structures were found. A phylogenetic tree of *Nicotiana* species based on restriction site data shows that the IR has both expanded and contracted during the evolution of this genus. Gene conversion is proposed to account for these small and apparently random IR expansions. A large IR expansion of over 12 kb has occurred in *Nicotiana acuminata*. The new IR–LSC junction in this species lies within intron 1 of the *clpP* gene. This rearrangement occurred via a double-strand DNA break and recombination between poly (A) tracts in *clpP* intron 1 and upstream of *rps19*. *Nicotiana acuminata* chloroplast DNA contains a 'molecular fossil' of the IR–LSC junction that existed prior to this dramatic rearrangement.

Communicated by R. Hagemann

S. E. Goulding · K. H. Wolfe (✉)
Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland

R. G. Olmstead
Department of Botany, University of Washington, Seattle, WA 98195, USA

C. W. Morden
Department of Botany/H.E.B.P., University of Hawaii, Honolulu, HI 96822, USA

Key words Inverted repeat · Large single copy region · Chloroplast DNA · Gene conversion · Double-stranded DNA break

Introduction

A large inverted repeat (IR) is a nearly universal feature of chloroplast genomes in land plants. In typical angiosperms such as tobacco (Shinozaki et al. 1986), the IR is approximately 25 kb in length and the two copies divide the rest of the circular chloroplast genome into a large single copy (LSC) region of about 87 kb and a small single copy (SSC) region of around 18 kb. This organisation is found in the vast majority of flowering plants (Palmer 1991; Downie and Palmer 1992) but a few species have IRs that are either greatly expanded [for example, 76 kb in *Pelargonium hortorum* (Palmer et al. 1987a); 37 kb in *Nicotiana acuminata* (Shen et al. 1982); 37 kb in *Berberis* and *Mahonia* (Kim and Jansen 1994)] or greatly reduced (e.g. the IR in *Coriandrum sativum* is less than half the normal size; Palmer 1985). Examples of smaller changes include a 4 kb expansion in *Anemone* and related species in the Ranunculaceae (Hoot and Palmer 1994), and a probable 7 kb contraction in *Cuscuta* (Bömmer et al. 1993). One copy of the IR has apparently been completely lost at least three times in angiosperms: it is not present in legumes such as pea and broad bean, in *Conopholis americana* (Orobanchaceae), or in *Erodium* and *Sarcocaulon* (Geraniaceae) (Palmer et al. 1987b; Wolfe 1988; Herdenberger et al. 1990; Downie and Palmer 1992). The IR is also essentially absent from the gymnosperm black pine (Wakasugi et al. 1994). This indicates that the presence of an IR is not essential for chloroplast genome function.

The IR has several properties which indicate that it is subject to the continuous operation of a molecular gene conversion and copy correction mechanism. First, the two copies of the IR appear to be identical and can

undergo flip-flop recombination (Palmer 1983): chloroplast DNA (cpDNA) isolated from a plant consists of an equimolar mixture of two inversion isomers in which the LSC and SSC are in different relative orientations. Second, the rate of neutral nucleotide substitution is lower within the IR than in the single-copy regions (Wolfe et al. 1987; Birky and Walsh 1992). Third, it can expand and contract slightly, as discussed here.

We have focussed on the junctions (termed J_{LA} and J_{LB} by Shinozaki et al. 1986) between the IR and the LSC region of cpDNA, because more sequence data were available in the literature for these junctions than for the IR/SSC junctions (J_{SA} and J_{SB}). In angiosperms the *S10* operon of ribosomal proteins (*rpl23–rpl2–rps19–rpl22–rps3–rpl16–rpl14–rps8–infA–rpl36–rps11–rpo*A) initiates within the IR and is transcribed across J_{LB} as a single polycistronic transcript (Tonkyn and Gruissem 1993). The 3′ part of this operon (*rpl22* and genes downstream) is located in the LSC adjacent to J_{LB}. However, at J_{LA} the duplicated 5′ part of the *S10* operon abuts on a second operon, *psbA–trnH*, transcribed in the opposite direction towards the IR (Fig. 1). In many angiosperms, such as spinach (Fig. 1a), J_{LA} lies within *rps19* with the result that the 5′ end of this gene is duplicated. The truncated copy, referred to as *rps19′*, could potentially encode a shortened version of ribosomal protein S19 but there is no evidence that such a protein is actually translated (Zurawski et al. 1984; Harris et al. 1994) or that *rps19′* is even transcribed (Tonkyn and Gruissem 1993).

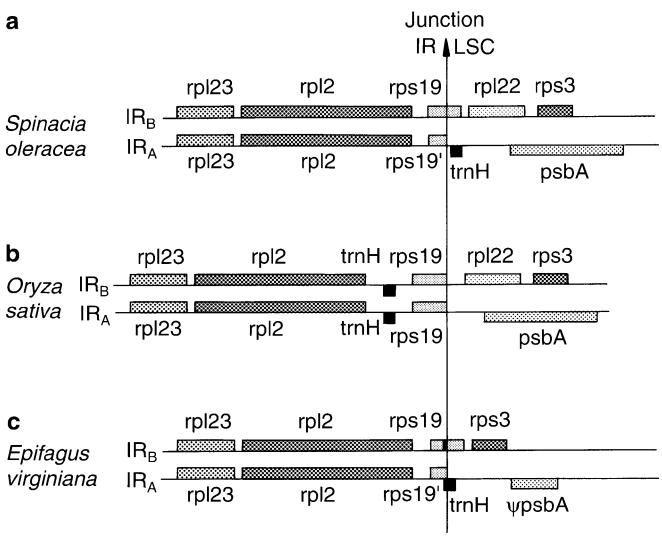


Fig. 1a–c Comparison of the structures of IR/LSC junction regions in three angiosperm species. Coding sequences are represented by *shaded boxes* and are drawn approximately to scale. Genes drawn *above* the horizontal are transcribed from left to right. *Vertical arrows* indicate the junction between inverted repeat (IR) and large single copy (LSC) sequences. Data are from Zurawski et al. (1984), Hiratsuka et al. (1989) and Wolfe et al. (1992b)

Although the IR in many angiosperms is approximately constant in size (~ 25 kb), sequence analysis of its endpoints has shown that small differences exist between species in the extent of the IR (Zurawski et al. 1984; Maier et al. 1990) (Fig. 1). For example, in *N. debneyi* the 5′ end of *rps19* is located in the IR, whereas in its congener *N. tabacum* this gene is entirely within the LSC (Zurawski et al. 1984; Shinozaki et al. 1986). In other angiosperms more complex arrangements are seen in these regions. The cpDNAs of rice and maize (Poaceae) have a fully duplicated *rps19* gene but in these species the *trnH* gene (encoding a histidine tRNA) has moved and is now located within the IR between *rpl2* and *rps19* (Fig. 1b). In *Epifagus virginiana* the final six nucleotides of the IR are shared by the genes *rps19* and *trnH* (Fig. 1c), in opposite transcriptional orientations. These six bases are identical to the 3′ ends of *trnH* genes from other species, and cause substitution of at least one amino acid residue in the *Epifagus rps19* protein as compared to other S19 proteins. A similar overlap (of 20 bp) occurs between $trnL_{UAG}$ and ORF1738 at the IR/SSC junctions in *Epifagus* (Wolfe et al. 1992a).

The data summarised in Fig. 1 suggests that the IR can expand and contract somewhat during evolution. To analyse this process in more detail we have determined the structures of the IR/LSC junctions in several closely related species and cultivars in the genus *Nicotiana* and family Solanaceae. By combining data on the IR endpoints with a cpDNA-derived phylogeny of the species, it is possible to determine the extent of IR junction mobility during their evolution. We present evidence for both expansion and contraction of the IR over a short distance (< 100 bp) in different *Nicotiana* lineages and develop a model for short IR expansions by means of gene conversion. We have also analysed a very recent dramatic expansion of the IR in *N. acuminata* and propose that this rearrangement occurred via a mechanism involving double-strand DNA breaks and subsequent repair.

Materials and methods

Plant material and DNA preparation

The species studied and sources of plant material, DNA or sequences are listed in Table 1. Seeds of *N. tabacum* cultivars were obtained from the USDA Tobacco Research Laboratory (Oxford, North Carolina, USA; courtesy of V. Sisson), the Solanaceae Seed Collection at the University of Birmingham (UK; courtesy of R. Lester), and the Beal Botanical Garden (Michigan State University, East Lansing, Michigan, USA). These, and *Digitalis purpurea* and *Solanum lycopersicum* obtained locally, were grown in the laboratory and total cellular DNA was extracted by the method of Guidet (1994) or Doyle and Doyle (1987). DNA was extracted by the method of Guidet (1994) from two cigars, one Dutch 'cahibo' purchased in 1994 and donated by L. Skrabanek and one Cuban 'cahibo' presented to Prof. D. McConnell by President Fidel Castro in 1983. Other DNAs were generously provided by Dr. T. Kavanagh

Table 1 Sources of plant materials, DNA and sequences

Species	Source/reference	Accession number
New sequences determined in this study		
Family Solanaceae		
Nicotiana plumbaginifolia	T. Kavanagh	Z71240, Z71241
Nicotiana sylvestris	BIRM S.0567	Z71233
Nicotiana nudicaulis	BIRM S.0451	Z71232
Nicotiana glauca	BIRM S.0024	Z71244
Nicotiana velutina	BIRM S.1024	Z71243
Nicotiana glutinosa	BIRM S.1002	Z71238
Nicotiana bigelovii	BIRM S.0903	Z71225, Z71226
Nicotiana palmeri	BIRM S.1027	Z71234, Z71235
Nicotiana alata	Beal Botanical Garden	Z71239
Nicotiana attenuata	BIRM S.1001	Z71242
Nicotiana acuminata	BIRM S.0372	Z71253, Z71254
Nicotiana tabacum cultivars:		
cv. Samsun NN	T. Kavanagh	Z71236, Z71237
cv. Havana 38	USDA TRL ref. PI 552432	Z71228
cv. Petit Havana	USDA TRL ref. PI 552516	Z71229
cv. Xanthi nc	USDA TRL ref. PI 552488	Z71227
cv. unknown	Dutch cahibo cigar	Z71231
cv. unknown	Cuban cahibo cigar	Z71230
Solanum nigrum	T. Kavanagh	Z71249, Z71250
Solanum tuberosum	T. Kavanagh	Z71247, Z71248
Solanum lycopersicum cv. Moneymaker	Mackey's Seeds, Dublin	Z71245, Z71246
Digitalis purpurea (Scrophulariaceae)	Mackey's Seeds, Dublin	Z71251, Z71252
Sequences obtained from the literature:		
Family Solanaceae		
Nicotiana debneyi	Zurawski et al. (1984)	X00796, X00798
Nicotiana tabacum cv. Bright Yellow 4	Shinozaki et al. (1986)	Z00044
Petunia hybrida	Aldrich et al. (1988)	M35955, M37322
Family Brassicaceae		
Arabidopsis thaliana	Liere et al. (1995)	X79898
	Newman et al. (1994)	R65293
	Höfte et al. (1993)	Z26532
Sinapis alba	Nickelsen and Link (1990)	X17331
Epifagus virginiana (Orobanchaceae)	Wolfe et al. (1992b)	M81884
Glycine max (Leguminosae)	Spielmann et al. (1988)	X06429
	Spielmann and Stutz (1983)	K01756
Spinacia oleracea (Chenopodiaceae)	Zurawski et al. (1984)	X00797
Helianthus annuus (Compositae)	Ambrosini et al. (1992)	X60428

(Trinity College Dublin). Extracted DNAs were used directly for PCR amplification without further purification.

PCR amplification

Primers for PCR amplification were designed from multiple species alignments of the genes flanking the IR/LSC junctions. These were as follows: RPL2, 5′-GATAATTTGATTCTTCGTCGCC-3′; RPL22, 5′-ACTCTTCGTGCTTTGTCAGC-3′; TRNH, 5′-CGGATGTAGCCAAGTGGATC-3′. These primers correspond to the 3′ end of *rpl2* and the 5′ ends of *rpl22* and *trnH*. J_{LA} regions were amplified with primer pair RPL2–TRNH for all species listed in Table 1 with the exception of *N. acuminata*. The products varied between 150 and 310 bp in different species. J_{LB} products were amplified with primer pair RPL2–RPL22 for only eight DNAs (see Fig. 2) and products were about 480 bp in size.

The *N. acuminata* junctions were amplified using primers designed from *N. tabacum psbB* and *clpP* sequences: PSBB (5′ end) 5′-GATACCAAGGCAAACCC-3′; CLPP (intron 1), 5′-AAGATCCGCCCGATTTG-3′. Amplification with oligonucleotide pairs PSBB–TRNH (spanning J_{LA}) and PSBB–CLPP (spanning J_{LB}) gave PCR products from *N. acuminata* of 1065 bp and 983 bp, respectively. Products were cloned into the pCRII vector (Invitrogen) according to supplier's instructions. Plasmid DNA was isolated for sequencing as described by Murphy and Kavanagh (1988). Sequencing was performed on an ABI automated sequencer using double-stranded DNA as templates with dye-labelled primers (Genpak, Sussex, UK).

For each species two independent PCR amplifications were carried out and one cloned product was sequenced on one strand from each. Some sequence conflicts or ambiguities were resolved by sequencing a third clone; in other cases ambiguities remain and are shown in Fig. 2. Sequence analysis was carried out on a Sun workstation using the Staden package; sequence alignments were made by eye.

Results

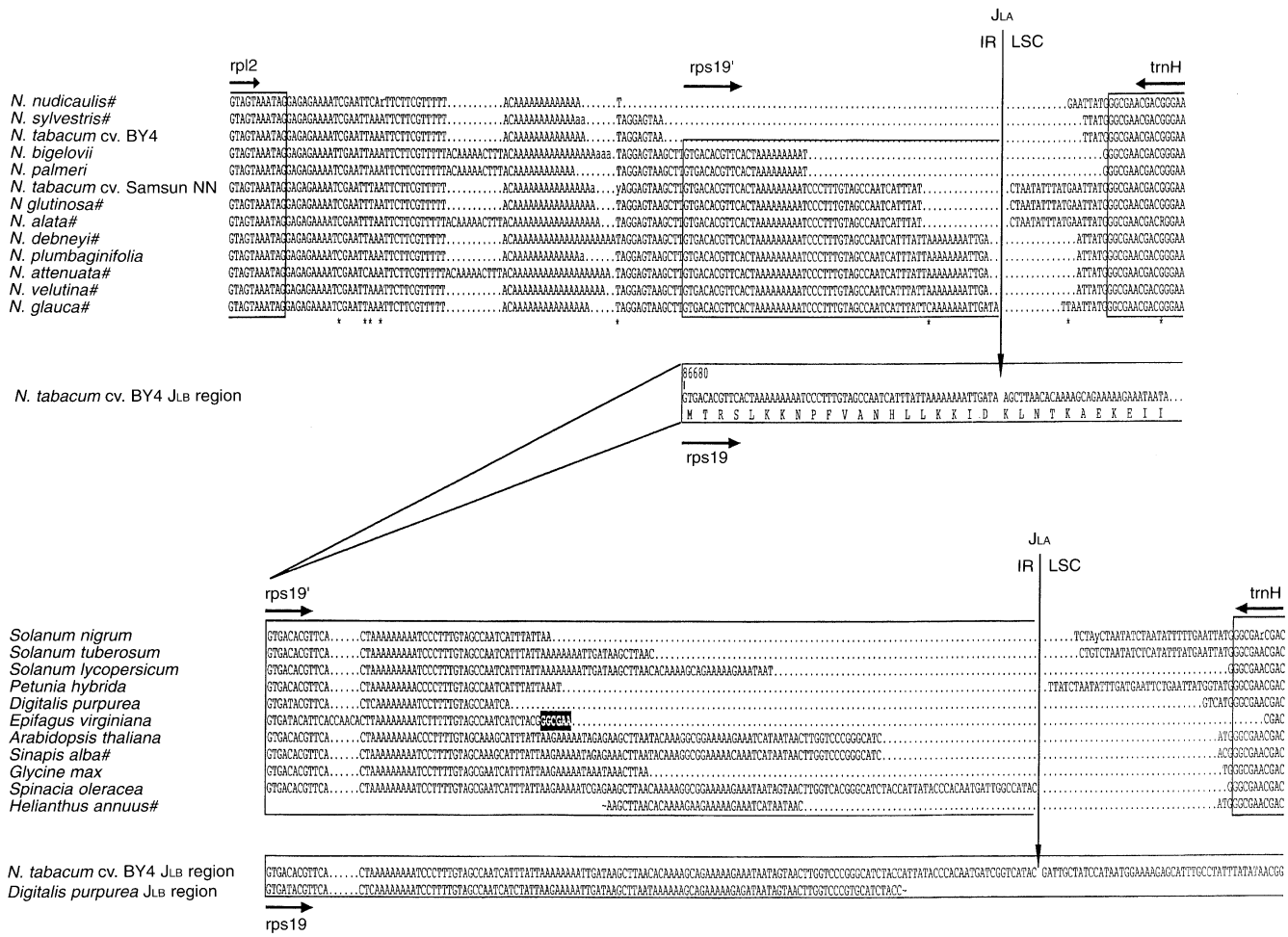
Definition of IR/LSC junctions (J_{LA} and J_{LB}) in *Nicotiana* species

We determined the sequence of the region spanning J_{LA} (between the 3′ end of *rpl2* and the 3′ end of *trn*H)

from 11 species in the genus *Nicotiana*, including *N. tabacum* cv. Samsun NN. These sequences were compared to published data for *N. tabacum* cv. Bright Yellow 4 (BY4) (Shinozaki et al. 1986) and *N. debneyi* (Zurawski et al. 1984) (Fig. 2, upper). To determine the exact location of the IR/LSC junction we also sequenced the region around J_{LB} (between the 3′ end of *rpl2* and the 5′ end of *rpl22*, containing a complete *rps19* gene) from four species (*N. tabacum* cv. Samsun NN, *N. plumbaginifolia*, *N. bigelovii* and *N. palmeri*). In each case the *rps19* sequence was identical to the published sequence for *N. tabacum* cv. BY4, so we did not consider it necessary to sequence the J_{LB} region from the other *Nicotiana* species but instead inferred the junction location by comparison with the *N. tabacum* *rps19* sequence.

Alignment of the *Nicotiana* J_{LA} region sequences is straightforward, with almost all of the differences between species being due to length and not point mutations (Fig. 2, upper). The 13 *Nicotiana* junction sequences fall into six groups, with different extents of *rps19* duplication and different amounts of 'spacer' DNA between J_{LA} and the 3′ end of *trnH*. In *N. nudicaulis*, *N. sylvestris* and *N. tabacum* cv. BY4 J_{LA} is located upstream of *rps19*, whereas in other sequences

Fig. 2 Alignments of J_{LA} region sequences from the genus *Nicotiana* (*upper panel*) and other dicot species (*lower panel*). Dots represent gaps inserted for alignment. *Vertical arrows* show the position of J_{LA}. *Asterisks* indicate nucleotide substitutions among the Nicotiana species. *Hash symbols* (#) next to species names indicate those for which the corresponding J_{LB} sequence is not known. *Tilde symbols* (~) indicate incomplete sequence data. The six nucleotides in the *E. virginiana* IR that are shared by *rps19* and *trnH* are *highlighted*. IUPAC nucleotide ambiguity codes (R and Y) written in *upper case* indicate positions that were unclear in sequencing reactions from individual clones. *Lower case* lettering indicates positions where conflicting sequences were obtained from independent PCR clones from a single species. The *N. tabacum* cv. BY4 sequence is representative of *N. tabacum* cultivars Petit Havana, Havana 38, Xanthi nc and the Dutch and Cuban cigars, the only difference being in the length of poly(A) tract upstream of the start codon of the *rps19* gene (A_{14}, A_{13}, A_{14}, A_{15} and A_{13} respectively). The *N. glauca* sequence contains a silent point substitution (A to C) in the coding sequence of *rps*19′ relative to the other species; its J_{LB} region has not been sequenced so we do not know whether this mutation is also present in the complete *N. glauca* *rps19* gene or if it reflects contraction of the IR/LSC junction to this point. The *A. thaliana* sequence shown is a composite of data from Liere et al. (1995), Newman et al. (1994) and Höfte et al. (1993); an apparent frameshift mutation in *Arabidopsis* *rps19* just upstream of the IR/LSC junction in the sequence of Liere et al. (1995) is not present in sequences of the J_{LA} and J_{LB} regions determined (in the course of mass cDNA sequencing projects) by Newman et al. (1994) and Höfte et al. (1993) and is not shown here. An apparent frameshift in the Sinapis *rps19* sequence (Nickelsen and Link 1990) at a site identical to that in *Arabidopsis* was also corrected



between 24 and 61 nucleotides of *rps19* lie within the IR. The J_{LA}–trnH spacer ranges in length between one and 19 nucleotides.

In the spacer between *rpl2* and *rps19* two other polymorphic regions are present. An 11-nucleotide sequence, ACAAAAACTTT, is present in *N. bigelovii*, *N. palmeri*, *N. alata* and *N. attenuata* but not other Nicotiana species. Since this sequence is also present in other species such as *Solanum* and *Petunia* (data not shown), the length difference is due to a deletion in some *Nicotiana* lineages. However, the species in which the 11-nucleotide block has been deleted do not form a monophyletic group (see Fig. 3); at least two independent deletions of this sequence must be proposed, one in *N. glutinosa* and a second in the common ancestor of the other seven species in which it is absent. The deletion has occurred between two copies of a repeated heptamer (ACAAAAA), which makes parallel independent deletions quite plausible.

The second region of variation in the *rpl2–rps19* spacer is a tract of adenine residues, varying in length from A_{12} in *N. palmeri* to A_{20} in *N. debneyi*. Other dicot species also contain poly(A) tracts of varying lengths at this position (Zurawski and Clegg 1987). We observed length differences in this region between cloned PCR products from some individual species as indicated in Fig. 2, but we cannot say whether this represents genuine intraspecific polymorphism in cpDNA of the type recently reported by Powell et al. (1995a,b), or whether it is due to PCR artefacts.

No natural intraspecies polymorphism for junction site exists in *N. tabacum*

The IR/LSC junctions occur at different sites in two cultivars of *N. tabacum* (Fig. 2). In cultivar Samsun NN, 46 nucleotides of *rps19* are duplicated, and the spacer between J_{LA} and *trnH* is 19 nucleotides long; we will represent this arrangement by the notation 46/19. In cultivar BY4 (Shinozaki et al. 1986) the junction has the structure −4/5. This suggested that the J_{LA} junction site might be polymorphic within individual species. To investigate this further we sequenced the J_{LA} region from three other *N. tabacum* cultivars (Petit Havana, Xanthi nc and Havana 38). We also extracted DNA from two cigars (see Table 1), assumed to be *N. tabacum*, and sequenced the J_{LA} region from these. In all five cases the sequence obtained had the arrangement −4/5 as in cultivar BY4.

The −4/5 arrangement is also found in *N. sylvestris*, which is thought to be the maternal parent of the tetraploid species *N. tabacum* (Goodspeed 1954; Gray et al. 1974). Thus, *N. tabacum* cv. BY4 and the other cultivars with the −4/5 arrangement have retained maternally inherited *N. sylvestris* cytoplasm. The *N* gene, conferring resistance to tobacco mosaic virus, was introduced into *N. tabacum* cv. Samsun NN by a crossbreeding strategy that involved *N. glutinosa* as the maternal grandmother (Clausen and Goodspeed 1925; Holmes 1938; Dunigan et al. 1987). Thus *N. tabacum* cv. Samsun NN contains an *N. glutinosa*-derived cytoplasm, and both of these have the arrangement 46/19 at J_{LA} (as does *N. alata*). Crossbreeding experiments during cultivation, not natural sequence evolution, are responsible for the junction site difference seen between *N. tabacum* cultivars.

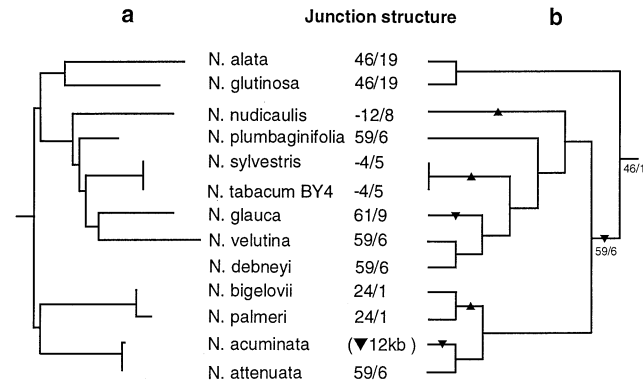


Fig. 3a, b Phylogeny of *Nicotiana* species studied. **a** Part of the most parsimonious tree obtained from cpDNA restriction site data from more than 100 Solanaceae species (R. G. Olmstead unpublished), drawn to scale. **b** Schematic representation of the tree in **a**, but with the root moved to a position more compatible with the similarity between the J_{LA}–*trnH* spacer sequences in *N. alata* and *N. glutinosa* and those in other Solanaceae (see text). *N. debneyi* was arbitrarily placed beside *N. velutina* (both are of Australian origin). The numbers of nucleotides of *rps19* within the IR, and in the J_{LA}–*trnH* spacer region, are indicated for each species (left and right of the slash mark, respectively). Inferred expansions (▼) and contractions (▲) are indicated

IR/LSC junctions in other dicot species

We sequenced the IR/LSC junctions (both J_{LA} and J_{LB}) from four other species (*Solanum nigrum*, *S. tuberosum*, *S. lycopersicum*, *Digitalis purpurea*) and compiled data from the literature for seven other dicots (Fig. 2, lower; Table 1). The extent of *rps19* duplication in these species ranges from 41 to 143 bp, with *trnH* spacer sizes ranging from −6 bp (in *Epifagus*) and +1 bp in spinach to 36 bp in *Petunia*.

There is some evidence from Fig. 2 that the IR endpoints tend to cluster in closely related species. In *Nicotiana* the amount of *rps19* that is duplicated ranges from −12 to +61 bp, whereas in the family Solanaceae it ranges up to 92 bp; in two members of the Scrophulariales (*Digitalis* and *Epifagus*) it is 41 and 47 bp (the *Epifagus* IR also contains 6 bp of duplicated *trnH*, highlighted in Fig. 2); and the J_{LA} regions in *Arabidopsis thaliana* and *Sinapis alba* (family Brassicaceae) are essentially identical. Allowing for a greater sample size from the Solanaceae, this clustering suggests that the IR endpoint moves randomly by small distances (<100 bp).

Inferring IR expansion and contraction in *Nicotiana*

Restriction maps have been constructed for the chloroplast genomes of the *Nicotiana* species included in this study (with the exception of *N. debneyi*), as part of a larger phylogenetic survey of the Solanaceae based on restriction fragment length polymorphism (RFLP) data (Olmstead and Palmer 1991; R. G. Olmstead, unpublished). Figure 3a shows part of the most parsimonious tree obtained from RFLP data on over 100 Solanaceae species. Figure 3b shows an alternative tree which we consider to be more likely, even though it is one step longer when all Solanaceae are considered; for the *Nicotiana* taxa alone, the two trees are of equal length.

The reason for favouring the tree in Fig. 3b is that the sequence of the 19-bp J_{LA}–*trnH* spacer in *N. glutinosa* and *N. alata* almost perfectly matches parts of the spacer sequences from *S. nigrum* and *S. tuberosum* (though not *S. lycopersicum*, whose spacer is only 1 bp long) (Fig. 2). This suggests that this 19-bp sequence was ancestrally present in *Nicotiana*, so we have chosen a phylogenetic tree (Fig. 3b) that places *N. glutinosa* and *N. alata* at the base of the genus *Nicotiana*. This tree implies a single loss of the spacer sequence in one *Nicotiana* lineage (as well as one in tomato), whereas the most parsimonious RFLP tree (Fig. 3a) required either two parallel losses of this sequence in *Nicotiana* or a highly unlikely convergent gain in the *glutinosa/alata* lineage. The argument that this 19-bp spacer is ancestral to *Nicotiana* is also central to the proposed mechanism of IR expansion through gene conversion (see below).

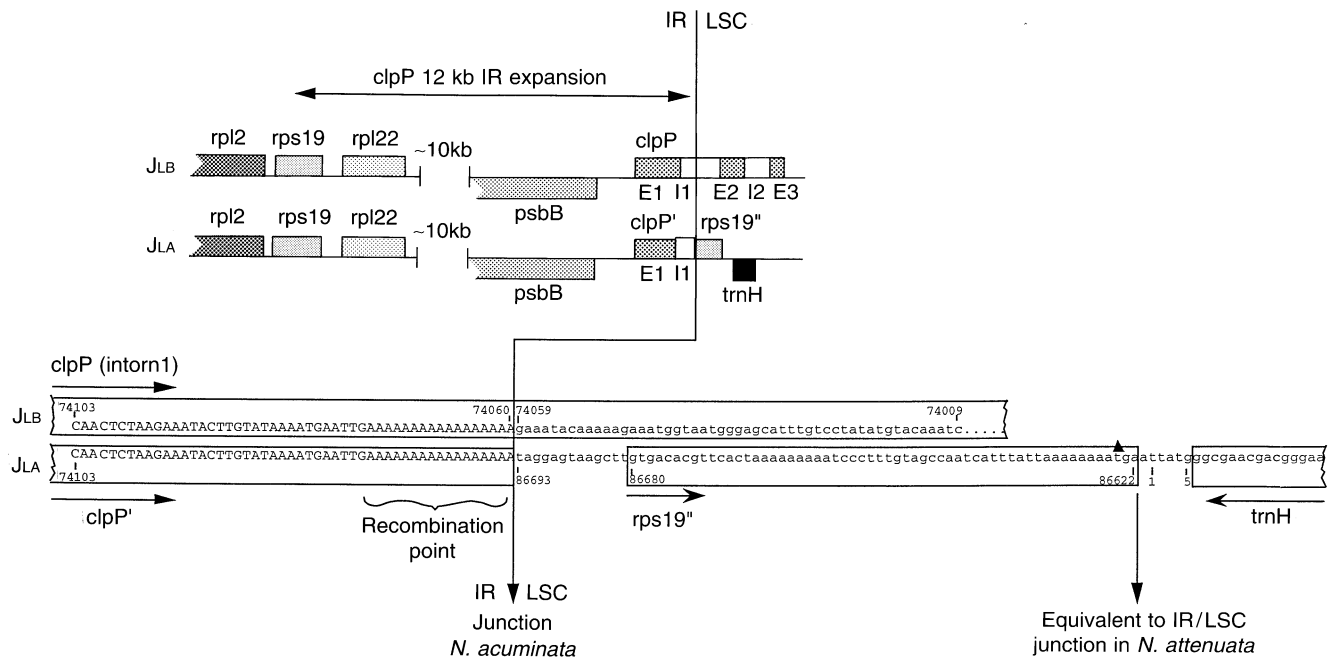
From the phylogenetic tree (Fig. 3b) two small expansions and three contractions on the IR can be inferred within *Nicotiana*, in addition to the dramatic expansion that has occurred in *N. acuminata* (Shen et al. 1982).

Expansion of the IR by 12 kb in *N. acuminata*

Early electron microscopy and restriction mapping studies revealed that *N. acuminata* had an expanded IR as compared to its closest relatives (Shen et al. 1982). Its chloroplast genome is enlarged to 171 kb and each copy of the IR represents 37 kb of this. From our restriction map of the *N. acuminata* genome (R.G.Olmstead unpublished) we estimated that the IR had expanded by approximately 12 550 bp, which would place the new IR/LSC junction near the *clpP* gene, at the equivalent of *N. tabacum* genome coordinate 74 100. From restriction site data published by Shen et al. (1982) we obtained an independent estimate of coordinate 73 900.

On this basis PCR amplification of the J_{LA} region in *N. acuminata* was carried out using primers corresponding to the 5′ end of *psbB* and the 3′ end of *trnH* (see Fig. 4). A PCR product of 1.1 kb was obtained and sequenced completely. From this data we confirmed that the *N. acuminata* IR/LSC junction occurs within intron 1 of *clpP* (at a position equivalent to *N. tabacum*

Fig. 4 Structure and sequence of IR/LSC junctions in *N. acuminata*. Coding sequences are indicated by *shaded boxes*, intronic sequences in *clpP* by *open boxes*. The *vertical arrow* delineates the junction between inverted repeat (IR) and large single copy sequence (LSC). In the lower panel IR sequence is shown in *upper case*, LSC sequence in *lower case*. Equivalent *N. tabacum* genome coordinates (Shinozaki et al. 1986) are shown for some positions. The poly(A) tract where recombination is inferred to have occurred between the *clpP* and *rps19* regions is marked by a brace (recombination point). The junction between *N. acuminata rps19″* and *trnH*, corresponding to J_{LA} in *N. attenuata*, is marked by an arrow. The ▲ symbol indicates the site of a single-nucleotide deletion in *N. acuminata rps19″* (T in *N. acuminata*; TT in *N. attenuata* and other species; equivalent to *N. tabacum* nucleotides 86624–86625)



coordinate 74 060) and used *N. tabacum* sequence data to design an oligonucleotide from the 3′ section of *clp*P intron 1 that permitted PCR amplification of the J_{LB} region of *N. acuminata* (within *clp*P intron 1). Figure 4 illustrates schematically and at the sequence level the structure of these two regions. *N. acuminata clp*P intron 1 contains a poly(A) tract which has expanded relative to *N. tabacum* from A_{11} to A_{17}. The IR/LSC junction lies at the end of this tract.

The structure of the LSC adjacent to J_{LA} in *N. acuminata* is more complicated than expected. Rather than the partially duplicated *clpP* gene (*clpP′*) meeting the single-copy sequence downstream of *trnH*, a truncated copy of *rps19* (termed *rps19″*) was found. The sequence downstream of J_{LA} in *N. acuminata* corresponds to the J_{LA} region in *N. attenuata*, its closest relative, with a truncated (59 bp) copy of *rps19* followed by a 6-bp spacer and *trnH*. The identity to the *rps19* region includes 13 bp upstream of the *rps19* start codon (*N. tabacum* coordinates 86 693–86 681; Fig. 4). Further upstream of this, the A_{17} tract in *N. acuminata* can be aligned with either the poly(A) tract upstream of *rps19* in other *Nicotiana* species (Fig. 2) or the poly(A) tract in *clpP* intron 1. This poly(A) tract at the IR/LSC boundary in *N. acuminata* therefore contains the point of recombination between the *clpP* and *rps19* regions (Fig. 4). *N. tabacum* has an A_{11} tract in *clpP* intron 1 and an A_{14} tract upstream of *rps19*, whereas *N. acuminata* has a recombinant A_{17} tract. The resultant *N. acuminata* genome now has 2.2 copies of *rps19* – two contained within the IR and a third truncated pseudogene copy (*rps19″*) in the LSC. The sequence across the junction between *N. acuminata rps*19″ and *trnH* is identical to the J_{LA} sequence in *N. attenuata* (both having the arrangement 59/6) except that the *N. acuminata* sequence has sustained a single nucleotide deletion within *rps19″* close to the junction (Fig. 4). Therefore this point in the *N. acuminata* plastid genome was probably its J_{LA} prior to the rearrangement that caused the 12-kb IR expansion.

Discussion

IR movement is random and generally conservative

The data presented in this paper demonstrate that the IR/LSC boundaries are not static but rather are subject to a dynamic and random process that allows, for the most part, conservative expansions and contractions of the IR. Among the 13 *Nicotiana* species studied here, six J_{LA} locations were identified which include two independent expansion and three separate contraction events. In the wider analysis of other dicots only two species (*A. thaliana* and *S. alba*) had coincident J_{LA} points (Fig. 2, lower). More limited data available for the regions surrounding the IR/SSC junctions indicate similar situations to those described here (Maier et al. 1990; Wolfe et al. 1992a). Although no differences in J_{LA} were observed within a single species, other studies have revealed population polymorphism in other areas of the genome, specifically in lengths of mononucleotide repeat tracts (Powell et al. 1995a, b).

The IR may terminate within genes, giving rise to nonfunctional and potentially disruptive 5′ or 3′ truncated genes, their translated equivalents perhaps having deleterious effects on essential processes. The conservative nature of IR movement generally avoids disruptive genomic rearrangement but it is interesting to note that in none of the species studied to date does the IR junction occur close to the 3′ end of a ribosomal protein gene; in dicots up to 50% of *rps19* is duplicated, whereas in grasses it lies entirely within the IR. Almost complete ribosomal proteins could probably be assembled into ribosomes but might fail to function.

Tonkyn and Gruissem (1993) reported that, in spinach, expression of the truncated copy of the *S10* operon ($S10_A$), which includes the *rps19′* sequence, is repressed relative to the complete copy ($S10_B$). They suggested that this may be due to transcriptional interference by the highly and convergently transcribed *psbA–trnH* operon. The chloroplast gene encoding ribosomal protein S4 in *Chlamydomonas reinhardtii* spans the IR–SSC junction such that the 3′ end of the gene is duplicated (Randolph-Anderson et al. 1995). The truncated 3′ part of *rps4* present in the IR lacks a promoter and the authors found no evidence of its transcription. It is notable that although the C-terminus of this protein is not highly conserved, the *C. reinhardtii* S4 protein has lost a conserved sequence motif present in this region in many other species and has gained a sequence encoded by the IR at the C-terminus, not found in other species. These differences may have occurred as a result of movement of the IR junction (Randolph-Anderson et al. 1995).

Evidence for gene conversion in cpDNAs

Gene conversion has been implicated as a potentially important mechanism of cpDNA evolution (Bowman et al. 1988; Morton and Clegg 1993). There is both evolutionary and biochemical evidence to support the existence of such a process.

The chloroplast genomes of some land plants contain small repeats in addition to the large IR (Bowman and Dyer 1986; Palmer et al. 1987a; Shimada and Sugiura 1989). Such sequences provide the potential for cpDNA rearrangement by homologous recombination, either within a single molecule or between two molecules. In the cases of rice (Shimada and Sugiura 1989) and wheat (Bowman and Dyer 1986) many of the pseudogenes present in cpDNA derive from tRNA and ribosomal protein genes encoded within the IR. Three-quarters of repeated sequences detected in wheat

cpDNA contain a copy within the IR as well as in the single-copy region (Bowman and Dyer 1986). In the grass family Poaceae cpDNAs contain a pseudogene of *rpl23* that is diverging more slowly than the surrounding regions (Bowman et al. 1988). Results based on divergence estimates, transition/transversion biases and sequence composition provide evidence to suggest that *ψrpl23* is maintained by gene conversion with its functional homologue which lies in the IR (Morton and Clegg 1993; Bowman et al. 1988). The chloroplast genome of *Pinus thunbergii* (Wakasugi et al. 1994) contains numerous short repeated sequences, including a 495-bp IR which is probably related to the 25-kb IR of angiosperms (Tsudzuki et al. 1992), a slightly imperfect IR of 829–835 bp containing *psaM* and *trnS*, another IR of 188 bp, and direct repeats of 362, 280 and 89 bp. Whether any of the *Pinus* IRs are subject to flip-flop recombination or a reduced evolutionary rate has not been established.

All large chloroplast IRs so far known minimally encode the rRNA operon. Some algal IRs are of limited size such that they contain only the rRNA operon (Yamada 1991). The photosynthetic protist *Euglena gracilis* has no IR but maintains a tandem array of one partial and three complete rRNA operons (Hallick et al. 1993). Eubacteria, and thus presumably the cyanobacterial endosymbiotic progenitor of cpDNA, contain multiple copies of this operon dispersed throughout their genomes and maintain their identity by gene conversion (e.g. Fleischmann et al. 1995). Loss of DNA during streamlining of the chloroplast genome included loss of copies of this operon, and the process is reflected in the various arrangements of the rRNA operons observed in cpDNAs (Delp and Kössel 1991). A relaxation of specificity of the rRNA operon gene conversion mechanism may have allowed subsequent expansion of the IR, now homogenising a larger region than before.

As predicted by the endosymbiont theory, recombination mechanisms in the chloroplast may be related to their eubacterial counterparts (Palmer 1992; Cerutti et al. 1995). The RecA protein of *Escherichia coli* (reviewed by Kowalczykowski and Eggleston 1994) may act as a DNA-dependent ATPase that promotes homologous pairing of DNA molecules followed by extensive strand exchange and the production of heteroduplex DNA (a recombination substrate). Evidence for the existence of similar proteins in chloroplasts is accumulating: a protein immunologically related to *E. coli* RecA, inducible by DNA-damaging agents, has been localized in the stroma of pea chloroplasts (Cerutti et al. 1992, 1993); a plastid protein that is 53% identical to *E. coli* RecA is encoded by the nuclear genome of *A. thaliana* (Cerutti et al. 1992) and two additional related cDNAs from this organism have been identified (Pang et al. 1992). More recently, dominant negative mutants of the RecA protein of *E. coli* have been shown to inhibit cpDNA recombination and repair, presumably by interference with its chloroplast homologue (Cerutti et al. 1995).

Short IR expansions may occur via gene conversion

Contraction of the IR can occur by deletion of DNA from one copy of the IR, either within the IR or across one of the IR/single-copy boundaries. It is harder to imagine how IR expansion can occur. Close analysis of the J_{LA} arrangement in *Nicotiana* species suggests that expansion of the IR may be explained most simply by a gene conversion mechanism. The spacer region between J_{LA} and *trnH* in *Nicotiana* is apparently derived from a longer ancestral sequence present in two of the *Solanum* species (Fig. 2). On this basis the *N. glutinosa*/*N. alata* lineage (J_{LA} arrangement 46/19) was suggested to be the basal branch of the phylogenetic tree depicted in Fig. 3b. The J_{LA} ratio of 59/6 which is ancestral to all other *Nicotiana* lineages, arose through the gain of an extra 13 bp of *rps19* sequence and concomitant loss of 13 bp from the spacer region (Fig. 2).

A model to explain this expansion event by gene conversion is set out in Fig. 5. Recombination between the two copies of the IR is assumed to occur continually. Branch migration of a Holliday junction across an IR/LSC junction results in the formation of heteroduplex before this process stalls. Resolution of heteroduplex may then proceed by sequence correction against either strand; in one direction this yields the parental situation but the other causes J_{LA} to move (Fig. 5). This process is constrained by the necessity to retain functional copies of *rps19* and *trnH*, so, in general, only non-essential spacer DNA may be overwritten.

This model can explain how the tobacco J_{LA} arrangement 59/6 can arise from a 46/19 ancestor (Fig. 5a) but it must be noted that it does not adequately explain the expansion event that occurred in *N. glauca*. Here, the J_{LA} ratio is 61/9, so both *rps19′* and the J_{LA}–*trnH* spacer have grown relative to a presumed 59/6 ancestor. This situation could possibly have arisen by insertion of the sequence TATTA precisely at J_{LA} (see Fig. 2), but this seems unlikely. However, the proposed gene conversion mechanism is also consistent with the unusual structure of J_{LA} in *E. virginiana* (Fig. 5b): gene conversion has overwritten at least six bases of *rps19* sequence with *trnH*-derived DNA, causing these two genes to share sequences at this point. By comparison between *Epifagus* and *Digitalis* sequences we suggest that the *Epifagus* arrangement has arisen from a *Digitalis*-like ancestor, with a 12-bp region of heteroduplex being formed (6 bp of spacer and 6 bp of *trnH* sequence), which was then repaired with the first 4 bp reverting to *rps19* and the last 8 bp to *trnH*.

For the IR to expand by gene conversion, either *rps19* or *trnH* sequences must overwrite dispensable LSC DNA at the other IR/LSC junction. Where does this DNA come from? In the case of *Epifagus*, a small

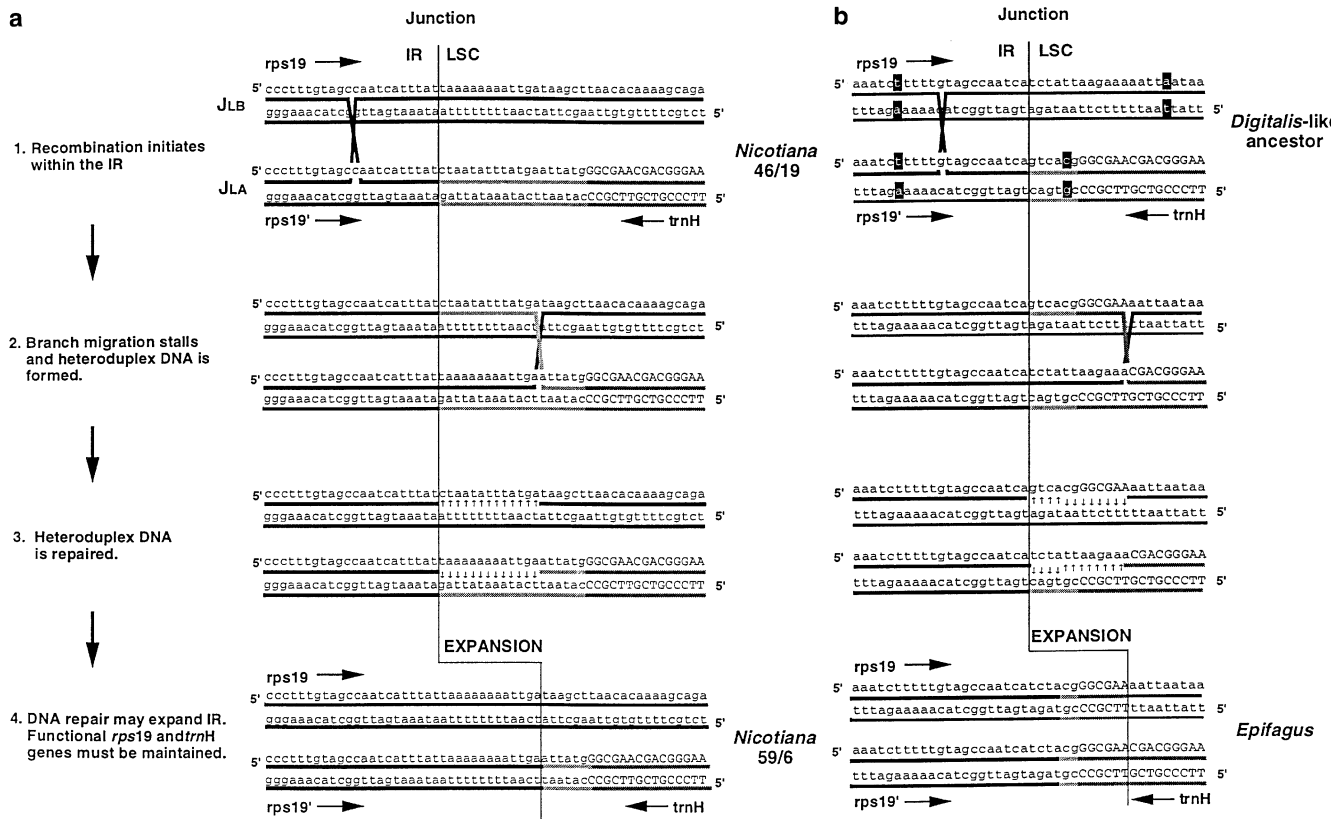


Fig. 5a, b Gene conversion mechanism for small IR movements. *Black underlines* indicate *rps19* coding sequence; *dark grey lines* indicate *trnH*; *pale grey lines* indicate intergenic spacer. *Vertical lines* show the position of IR/LSC junctions. *Small arrows* represent the direction of repair of heteroduplex DNA. **a** Stepwise model with sequence detail showing the expansion of the IR from a 46/19 to a 59/6 J_{LA} arrangement in *Nicotiana*. **b** Model for the origin of the *Epifagus* J_{LA} and J_{LB} structures from a *Digitalis*-like ancestor. The *Epifagus* and *Digitalis* sequences differ by 3 nucleotide substitutions, in addition to the IR junction movement, so for clarity these substitutions (*highlighted* in the top panel) are shown as having occurred in a hypothetical *Digitalis*-like ancestral sequence

section of *rps19* was replaced, presumably without destroying protein function. In the *Nicotiana* species, *rps19* sequences overwrote spacer DNA beside *trnH*. This spacer DNA could arise by either of two mechanisms. (i) Point mutation within one copy of the IR by definition contracts the IR junctions back to the last nucleotide of identity. The resulting single-copy sequence may now be freed from copy-correction and can diverge to become J_{LA}–*trnH* spacer. (ii) Alternatively, intergenic spacer may expand by tandem duplication of minirepeats. There is some evidence for such imperfect repeat sequences in the short J_{LA}–*trnH* spacer (e.g. CTAATAT, TTATG) of the Solanaceae, which are similar to expanding repeats seen in other plastid genomes (e.g. Wolfson et al. 1991). Either mechanism would allow for expansion of the LSC neighbouring the IR junctions, and subsequent IR expansion.

More dramatic IR junction rearrangements can occur via double-strand DNA breaks

Large IR expansion events, such as those observed in *Pelargonium* (Palmer et al. 1987a) and *N. acuminata* (Shen et al. 1982; this study), involve a considerable amount of sequence duplication that is unlikely to be attributable to gene conversion. Sequencing of the IR/LSC regions in *N. acuminata* not only defined the junctions but also provided a clue as to how the IR may have expanded (Fig. 4). The IR of this species now contains two full copies of the *S10* operon and duplicated sequences that include *psbB* and exon 1 of *clpP*, encompassing approximately 12 562 bp more DNA than in its closest relative, *N. attenuata* (Fig. 4). The expansion in *N. acuminata* is remarkably similar in extent to an independent IR expansion of 11.5 kb in *Berberis* and *Mahonia*, which also encompasses the whole *S10* operon and *psbB* (Kim and Jansen 1994).

Yamada (1991) proposed a mechanism involving double reciprocal recombination between IR segments during replication to account for cpDNA IR expansion among algal species, which are subject to processes similar to those described here (Boudreau and Turmel 1995). However, this model and a similar model proposed by Palmer et al. (1985) both rely on the presence of dispersed sequence repeats, rarely found in land plant cpDNAs, and are only applicable if such repeats are located precisely at IR–single copy endpoints.

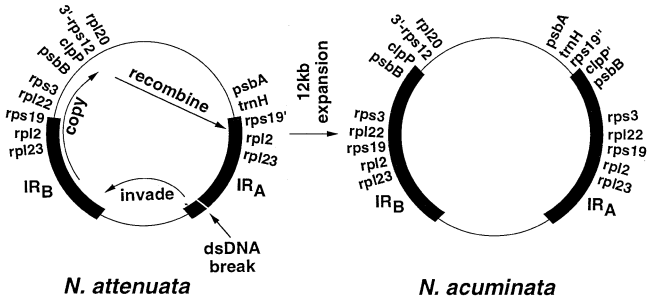


Fig. 6 Proposed mechanism for the IR expansion in *N. acuminata* from a *N. attenuata*-like progenitor, beginning with a double-strand DNA break in IR_A, followed by strand invasion, copying, and subsequent recombination and recircularisation. *Thickened* regions indicate the IR

We instead propose a model (Fig. 6) to account for the novel situation in *N. acuminata*. A double-strand break initiating within IR_A would leave the 3′ end of one strand free; this could be repaired against the complementary strand of IR_B, and the second strand subsequently repaired. The copy-repair might proceed beyond IR_B, incorporating in excess of 12 kb of previously LSC sequence into the IR. Illegitimate recombination between the poly(A) tracts of *clpP* intron 1 and the *rpl2–rps19′* spacer in IR_A restores a circular cpDNA molecule, with the fingerprint of the old J_{LA} being maintained in the LSC downstream of *trnH*. The identity of the *N. acuminata rps*19″*–trnH* spacer to the J_{LA} region of *N. attenuata* implies that the IR expansion in *N. acuminata* is a recent evolutionary event that has occurred since the divergence of these two species. This concurs with the inference from cpDNA RFLP data, which identified only a single restriction site difference for 11 enzymes that were mapped (Olmstead and Palmer 1991).

This model also can account for the situation in rice and maize where the *trnH* gene has been relocated between two genes of the *S10* operon (Fig. 1b). Here, the double-strand break must have initiated within IR_B and used IR_A as a template, allowing inclusion of *trnH* in the IR without disruption of functional genes. A second event would be required to account for the IR expansion encompassing the entire *rps19* gene within it.

This study provides sequence evidence for the operation of two distinct modes of IR junction evolution. (i) a random, continuous, conservative gene conversion process involving small stretches of DNA; and (ii) a mechanism for incorporating extensive stretches of single-copy cpDNA into the IR via double strand breaks but operating only very rarely. The former model is more likely to be related to the regular maintenance of the IR structure as whole; the second may perhaps occur as a by-product of DNA repair.

Acknowledgements We thank T. Kavanagh, V. Sisson and R. Lester for plant materials, and Peter Medgyesy and Jeff Palmer for discussion. We also thank Amanda Lohan for general discussion and technical help. This study was supported by the Trinity College Dublin Academic Research Fund and by Forbairt

References

Aldrich J, Cherney BW, Williams C, Merlin E (1988) Sequence analysis of the junction of the large single copy region and the large inverted repeat in the petunia chloroplast genome. Curr Genet 14:487–492

Ambrosini M, Ceci LR, Fiorella S, Gallerani R (1992) Comparison of regions coding for tRNA(His) genes of mitochondrial and chloroplast DNA in sunflower: a proposal concerning the classification of 'CP-like' tRNA genes. Plant Mol Biol 20:1–4

Birky CW, Walsh JB (1992) Biased gene conversion, copy number and apparent mutation rate differences within chloroplast and bacterial genomes. Genetics 130:677–683

Bömmer D, Haberhausen G, Zetsche K (1993) A large deletion in the plastid DNA of the holoparasitic flowering plant *Cuscuta reflexa* concerning two ribosomal proteins (*rpl2*, *rpl23*), one transfer RNA (*trnI*) and an ORF2280 homologue. Curr Genet 24:171–176

Boudreau E, Turmel M (1995) Gene rearrangements in Chlamydomonas chloroplast DNAs are accounted for by inversions and by the expansion/contraction of the inverted repeat. Plant Mol Biol 27:351–364

Bowman CM, Dyer TA (1986) The location and possible evolutionary significance of small dispersed repeats in wheat ctDNA. Curr Genet 10:931–941

Bowman CM, Barker RF, Dyer TA (1988) In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination. Curr Genet 14:127–136

Cerutti H, Osman M, Grandoni P, Jagendorf AT (1992) A homolog of *Escherichia coli* RecA protein in plastids of higher plants. Proc atl Acad Sci USA 89:8068–8072

Cerutti H, Ibrahim Z, Jagendorf AT (1993) Treatment of pea (*Pisum sativum* L.) with DNA-damaging agents induces a 39-kilodalton chloroplast protein immunologically related to *Escherichia coli* RecA. Plant Physiol 102:155–163

Cerutti H, Johnson AM, Boynton JE, Gillham NW (1995) Inhibition of chloroplast DNA recombination and repair by dominant negative mutants of *Escherichia coli* RecA. Mol Cell Biol 15:3003–3011

Clausen RE, Goodspeed TH (1925) Interspecific hybridisation in *Nicotiana* II. A tetraploid *glutinosa–tabacum* hybrid, an experimental verification of Winge's hypothesis. Genetics 10:278–284

Delp G, Kössel H (1991) rRNAs and rRNA genes of plastids. In: Bogorad L, Vasil IK (eds) The molecular biology of plastids, Cell culture and somatic cell genetics of plants (vol 7A). Academic Press, San Diego, pp 139–167

Downie SR, Palmer JD (1992) Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 14–35

Doyle JJ, Doyle JL (1987) A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Dunigan DD, Golemboski DB, Zaitlin M (1987) Plant resistance to viruses. Ciba Foundation Symposia 133:120–135. Wiley, Chicester

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields CA, Gocayne JD, Scott JD, Shirley R, Liu LI, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL,

McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269 : 496–512

Gray JC, Kung SD, Wildman SG (1974) Origin of *Nicotiana tabacum* L. detected by polypeptide composition of Fraction I protein. Nature 252 : 226–227

Goodspeed TH (1954) The genus Nicotiana: origins, relationships and evolution of its species in the light of their distribution, morphology and cytogenetics. Chronica Botanica Company

Guidet F (1994) A powerful new technique to quickly prepare hundreds of plant extracts for PCR and RAPD analyses. Nucleic Acids Res 22:1772–1773

Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E (1993) Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res 21 : 3537–3544

Harris EH, Boynton JE, Gillham NW (1994) Chloroplast ribosomes and protein synthesis. Microbiol Rev 58 : 700–754

Herdenberger F, Pillay DTN, Steinmetz A (1990) Sequence of the *trnH* gene and the inverted repeat structure deletion site of the broad bean chloroplast genome. Nucleic Acids Res 18 : 1287

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete nucleotide sequence of rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217 : 185–194

Höfte H, Desprez T, Amelsem J, Chiapello H, Caboche M, Miosan A, Juorjon MF, Charpenteau JL, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu DY, Mache R, Raynal M, Cooke R, Grellet F, Delseny M, Parmentier Y, Marcillac G, Gigot C, Fleck J, Phillips G, Axelos M, Bardet C, Tremousatgue D, Lescure B (1993) An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. Plant J 4 : 1051–1061

Holmes FO (1938) Inheritance of resistance to tobacco-mosaic virus in tobacco. Phytopathol 28 : 553–561

Hoot SB, Palmer JD (1984) Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. J Mol Evol 38 : 274–281

Kim YD, Jansen RK (1994) Characterization and phylogenetic distribution of chloroplast DNA rearrangement in the Berberidaceae. Plant Syst Evol 193 : 107–114

Kowalczykowski SC, Eggleston AK (1994) Homologous pairing and DNA strand exchange proteins. Annu Rev Biochem 63 : 991–1043

Liere K, Kestermann M, Müller U, Link G (1995) Identification and characterisation of the *Arabidopsis thaliana* chloroplast DNA region containing the genes *psbA*, *trnH* and *rps19′*. Curr Genet 28 : 128–130

Maier RM, Döry I, Kössel H (1990) The *ndhH* genes of graminean plastomes are linked with the junctions between small single copy and inverted repeat regions. Curr Genet 18 : 245–250

Morton BR, Clegg MT (1993) A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). Curr Genet 24 : 357–365

Murphy G, Kavanagh T (1988) Speeding-up the sequencing of double-stranded DNA. Nucleic Acids Res 16 : 5198

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, and Somerville C, (1994) Genes galore : a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones. Plant Physiol 106 : 1241–1255

Nickelsen J, Link G (1990) Nucleotide sequence of the mustard chloroplast genes *trnH* and *rps19′*. Nucleic Acids Res. 18 : 1051

Olmstead RG, Palmer JD (1991) Chloroplast DNA and systematics of the Solanaceae. In: Hawkes JG, Lester RN, Nee M, Estrada N (eds) Solanaceae III. Taxonomy, chemistry, evolution. Royal Botanic Gardens, Richmond, UK, pp 161–168

Palmer JD (1983) Chloroplast DNA exists in two orientations. Nature 301 : 92–93

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19 : 325–354

Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (eds) The molecular biology of plastids. Cell culture and somatic cell genetics of plants (vol 7A). Academic Press, San Diego, pp 5–53

Palmer JD (1992) Comparison of chloroplast and mitochondrial genome evolution in plants. In: RG Herrmann (ed) Cell organelles. Springer-Verlag, Vienna pp 99–133

Palmer JD, Boynton JE, Gillham NW, Harris EH (1985) Evolution and recombination of the large inverted repeat in Chlamydomonas chloroplast DNA. In: Steinback KE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Laboratory Press, New York, pp 269–278

Palmer JD, Nugent JM, Herbon LA (1987a) Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions and two repeat families. Proc Natl Acad Sci USA 84 : 769–773

Palmer JD, Osorio B, Aldrich J, Thompson, WF (1987b) Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr Genet 11 : 275–286

Pang Q, Hays JB, Rajagopal I (1993) Two cDNAs from the plant *Arabidopsis thaliana* that partially restore recombination proficiency and DNA-damage resistance to *E. coli* mutants lacking recombination-intermediate-resolution activities. Nucleic Acids Res 21 : 1647–1653

Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995a) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. Curr Biol 5 : 1023–1029

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA. (1995b) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92 : 7759–7763

Randolph-Anderson BL, Boynton JE, Gillham NW, Huang C, Liu XQ (1995) The chloroplast gene encoding ribosomal protein S4 in *Chlamydomonas reinhardtii* spans an inverted repeat – unique sequence junction and can be mutated to suppress a streptomycin dependence mutation in ribosomal protein S12. Mol Gen Genet 247 : 295–305

Shen GF, Chen K, Wu M, Kung SD (1982) Nicotiana chloroplast genome IV. *N. accuminata* has larger inverted repeats and genome size. Mol Gen Genet 187 : 12–18

Shimada H, Sugiura M (1989) Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr Genet 16 : 293–301

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5 : 2043–2049

Spielmann A, Stutz E (1983) Nucleotide sequence of soybean chloroplast DNA regions which contain the *psbA* and *trnH* genes and cover the ends of the large single copy region and one end of the inverted repeats. Nucleic Acids Res 11 : 7157–7167

Spielmann A, Roux E, Von Allmen JM, Stutz E (1988) The soybean chloroplast genome: complete sequence of the *rps19* gene, including flanking parts containing exon 2 of *rpl2* (upstream), but lacking *rpl22* (downstream). Nucleic Acids Res 16 : 1199

Tonkyn JC, Gruissem W (1993) Differential expression of the partially duplicated chloroplast S10 ribosomal protein operon. Mol Gen Genet 241 : 141–152

Tsudzuki J, Ito S, Tsudzuki T, Wakasugi T, Sugiura M (1992) Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. Mol Gen Genet 232 : 206–214

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91 : 9794–9798

Wolfe KH (1988) The site of deletion of the inverted repeat in pea chloroplast DNA contains duplicated gene fragments. Curr Genet 13 : 97–99

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondria, chloroplast and nuclear DNAs. Proc Natl Acad Sci USA 84 : 9054–9058

Wolfe KH, Morden CW, Palmer JD (1992a) Small single-copy region of plastid DNA in the non-photosynthetic angiosperm *Epifagus virginiana* contains only two genes. J Mol Biol 223 : 95–104

Wolfe KH, Morden CW, Palmer JD (1992b) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci USA 89 : 10648–10652

Wolfson R, Higgins KG, Sears BB (1991) Evidence for replication slippage in the evolution of Oenothera chloroplast DNA. Mol Biol Evol 8 : 709–720

Yamada T (1991) Repetitive sequence-mediated rearrangements in *Chlorella ellipsoidea* chloroplast DNA: completion of nucleotide sequence of the large inverted repeat. Curr Genet 19 : 139–147

Zurawski G, Clegg MT (1987) Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure–function and phylogenetic studies. Annu Rev Plant Physiol 38 : 391–418

Zurawski G, Bottomley W, Whitfeld PR (1984) Junctions of the large single copy region and the inverted repeats in *Spinacia oleracea* and *Nicotiana debneyi* chloroplast DNA: sequence of the genes for $tRNA^{His}$ and the ribosomal proteins S19 and L2. Nucleic Acids Res 12 : 6547–6558