Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron

J. Stephen Gantt, Sandra L. Baldauf¹,
Patrick J. Calie², Norman F. Weeden³ and
Jeffrey D. Palmer¹

Department of Plant Biology, University of Minnesota, St Paul, MN 55108, ¹Department of Biology, Indiana University, Bloomington, IN 47405, ²Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109 and ³Department of Horticultural Sciences, Cornell University, Geneva, NY 14456, USA

Communicated by J.H. Weil

Most chloroplast and mitochondrial proteins are encoded by nuclear genes that once resided in the organellar genomes. Transfer of most of these genes appears to have occurred soon after the endosymbiotic origin of organelles, and so little is known about the process. Our efforts to understand how chloroplast genes are functionally transferred to the nuclear genome have led us to discover the most recent evolutionary gene transfer yet described. The gene *rpl22*, encoding chloroplast ribosomal protein CL22, is present in the chloroplast genome of all plants examined except legumes, while a functional copy of *rpl22* is located in the nucleus of the legume pea. The nuclear *rpl22* gene has acquired two additional domains relative to its chloroplast ancestor: an exon encoding a putative N-terminal transit peptide, followed by an intron which separates this first exon from the evolutionarily conserved, chloroplast-derived portion of the gene. This gene structure suggests that the transferred region may have acquired its transit peptide by a form of exon shuffling. Surprisingly, phylogenetic analysis shows that *rpl22* was transferred to the nucleus in a common ancestor of all flowering plants, at least 100 million years preceding its loss from the legume chloroplast lineage.

Key words: chloroplast/evolutionary gene transfer/gene evolution/intron/ribosomal protein

Introduction

The structure and gene content of the plastid genome are well conserved among divergent land plants (Palmer, 1985; Palmer and Stein, 1986; Shinozaki *et al.*, 1986; Hiratsuka *et al.*, 1989; Fukuzawa *et al.*, 1988; Palmer, 1991). Typically, this genome is 140–160 kb in length, contains a large (~25 kb) inverted repeat, and encodes ~100 proteins, only a small proportion of those needed for chloroplast function. The majority of chloroplast proteins are encoded by nuclear genes and imported post-translationally into the plastid. Because the gene content of the plastid genome is so similar among widely diverse plants, most genes are thought to have been transferred from the ancestral endosymbiont to the nucleus early in plant evolution (Palmer, 1985). A single clear example exists of a 'recent' evolutionary transfer of a chloroplast gene to the nucleus.

The *tufA* gene, encoding chloroplast elongation factor Tu, appears to have been transferred from the chloroplast to the nucleus in the algal lineage that gave rise to land plants, ~500 million years ago (Palmer, 1991; Baldauf and Palmer, 1990). Little is known about the events involved in a functional relocation of a chloroplast gene to the nucleus; therefore we have screened for additional, even more recent cases of transfer.

The presence in chloroplast DNA (cpDNA) of a homolog (termed *rpl22*) of the *Escherichia coli* gene encoding ribosomal protein L22 was first shown by DNA sequencing of the tobacco chloroplast genome (Tanaka *et al.*, 1986). This gene was subsequently found in cpDNAs of one other dicot (Zhou *et al.*, 1989), two monocots (Hiratsuka *et al.*, 1989; McLaughlin and Larrinua, 1987), a non-vascular land plant (Fukuzawa *et al.*, 1988), and three highly diverse algae (Christopher *et al.*, 1988; Kao and Wu, 1990; Evrard *et al.*, 1990; Michalowski *et al.*, 1990). However, *rpl22*, which is part of a large cluster of ribosomal protein genes in rice, tobacco, and liverwort, is not found in its normal location in soybean cpDNA (Spielmann *et al.*, 1988) and is absent from subclover (*Trifolium subterraneum*) cpDNA (Milligan *et al.*, 1989). Here we report that the plastid genomes of apparently all legumes lack *rpl22*, raising the question of whether this gene has been entirely lost from legumes or was transferred to the nucleus. We show that in pea, *rpl22* is located in the nucleus and describe features of the gene that provide insight into the timing and mechanism of gene transfer.

Results

Distribution of rpl22 in cpDNAs of monocots and dicots

In the course of our studies on cpDNA rearrangements in legumes, we noticed that a tobacco *rpl22* gene probe consistently failed to hybridize to legume cpDNAs. To investigate this further, we used Southern blot hybridization to survey cpDNAs from 17 diverse flowering plants, representing 14 families and eight of 11 subclasses (Cronquist, 1988) of angiosperms. A probe containing the 5' 176 bp of the tobacco chloroplast *rpl22* gene hybridized strongly to all angiosperm cpDNAs tested except the four legumes, which represent three diverse tribes (Figure 1). Identical results were obtained with a non-overlapping, 111 bp fragment containing the central portion of tobacco *rpl22* (data not shown). DNA sequencing of the region of the chloroplast genome containing one or both ribosomal protein genes (*rps19* and *rps3*) that normally flank *rpl22* has failed to reveal any trace of the gene in cpDNAs of soybean and subclover (Spielmann *et al.*, 1988; Milligan *et al.*, 1989).

Isolation and characterization of a pea rpl22 cDNA

The tobacco cpDNA *rpl22* probe was used to screen a pea cDNA library, and the insert from one hybridizing

Fig. 1. *rpl22* is present in cpDNA in all angiosperms except legumes. **(Top)** Electrophoresis in a 0.9% agarose gel of 2 μ g of cpDNA from *Zea mays* (lane 1, *Eco*RI fragments), *Narcissus tazetta* (2, *Sac*I–*Pvu*II), *Aristolochia durior* (3, *Sac*I–*Pvu*II), *Delphinium grandiflorum* (4, *Sac*I–*Pvu*II), *Eschscholzia californica* (5, *Sac*I–*Pvu*II), *Pilea microphylla* (6, *Sac*I–*Pvu*II), *Spinacia oleracea* (7, *Sac*I–*Pst*I), *Rumex obtusifolius* (8, *Sac*I–*Pvu*II), *Glycine max* (9, *Sac*I–*Pvu*II), *Medicago sativa* (10, *Hind*III), *Trifolium subterraneum* (11, *Eco*RI), *Pisum sativum* (12, *Sac*I–*Pst*I), *Aesculus californica* (13, *Sac*I–*Pvu*II), *Pelargonium hortorum* (14, *Sac*I–*Pvu*II), *Brassica campestris* (15, *Sac*I–*Pst*I), *Nicotiana tabacum* (16, *Hind*III), and *Lactuca sativa* (17, *Hind*III). **(Bottom)** Hybridization to a filter blot of the above gel with a nick-translated, 32 P-labeled 209 bp *Pst*I–*Xba*I fragment of the tobacco chloroplast *rpl22* gene (Shinozaki *et al.*, 1986). This fragment contains 33 bp of 5' noncoding sequence and the first 176 bp of *rpl22* coding sequence (Palmer, 1986). Size markers at right are in kilobases.

recombinant was subcloned and sequenced (Figure 2A). This cDNA (PsCL22-6) contained a single long uninterrupted reading frame of 204 codons, corresponding to a 22 791 dalton protein, 53 bp of 5' and 193 bp of 3' non-coding sequence, and a 14 bp poly(A)-tract at the 3' end. When the deduced amino acid sequence of this clone was used to search a protein sequence database, significant similarities were found only with chloroplast and eubacterial L22 proteins. Alignment of pea CL22 with several of these proteins shows similarity over nearly the entire overlapping region (Figure 2B). The most striking difference among these proteins is an amino terminal extension in pea CL22 of ~70–80 amino acids that is not found in the other proteins. This N-terminal sequence has many of the

characteristics, such as a net positive charge (+5) and a high (19%) serine/threonine content (von Heijne *et al.*, 1989), that are common to chloroplast transit peptides, which function in targeting cytoplasmically-synthesized proteins to the chloroplast. The N-terminal sequence also lacks several characteristics of mitochondrial targeting sequences. In all but one of 132 imported mitochondrial proteins examined by von Heijne (1986), a strongly amphiphilic α -helical structure was predicted in the targeting sequence before the polypeptide's first acid amino acid. The presence in the CL22 sequence of aspartic acid (position 14) before a predicted amphiphilic helix (von Heijne, 1986) would therefore be very unusual for a mitochondrial targeting sequence.

Northern blot analysis confirmed that the pea *rpl22* gene is actively expressed and, as expected for a cytoplasmic mRNA, its transcript is polyadenylated. A single *rpl22* transcript of ~1.0 kb was enriched in poly(A)⁺ RNA and depleted in poly(A)[–] RNA (Figure 3). In control hybridizations, a similar pattern was obtained for *cab4*, a well characterized nuclear gene encoding a chloroplast protein (Tobin and Silverthorne, 1985).

Pea rpl22 is located in the nucleus

To confirm that the PsCL22-6 cDNA clone was indeed derived from the product of a nuclear gene, we hybridized it to a Southern blot containing total DNA from two parental pea lines and 41 lines derived from their F₂ offspring. The clone hybridized to a single band whose *Eco*RV polymorphism (Figure 4) segregated in a ratio of 1:2:1 ($\chi^2 = 0.46$, $P = 0.79$). The joint segregation analysis showed a recombination distance of 15% between *rpl22* and *Acp-4*, a gene encoding β -acid phosphatase that was previously shown to be located on chromosome 6 of pea (Weeden and Wolko, 1990).

Structure of pea rpl22

We examined the structure of pea *rpl22* by sequencing a portion of cloned genomic DNA. The sequence of *rpl22* was identical to that of the cDNA except that, as expected, the gene lacked the 3' poly(A)-tract found in the cDNA and, unexpectedly, contained a 305 bp intron near the 5' end (Figure 2A). This intron separates the gene into two exons (Figure 2B) each having different functions and probably different origins. The 5' exon encodes most, if not all, of the putative transit peptide and the 3' exon encodes all of the evolutionarily conserved chloroplast-derived ribosomal protein. The 522 bp of genomic sequence upstream of the *rpl22* coding region is extremely (86%) AT-rich. AT-rich regions of DNA are known to bind a variety of plant DNA-binding proteins and are thought to be associated with the nuclear scaffold (Jacobsen *et al.*, 1990; Mirkovitch *et al.*, 1984).

Phylogenetic analysis of rpl22

The evolutionary relationship between the pea nuclear *rpl22* and the chloroplast and bacterial *rpl22* genes was investigated by parsimony analysis. We analyzed amino acids rather than nucleotides because the latter are likely to be saturated at silent positions in these comparisons and because nuclear and chloroplast genes are known to have different codon usage patterns and rates of silent substitutions (Wolfe *et al.*, 1987; Campbell and Gowri, 1990; Oliver *et al.*, 1990).

A

```

25      50      75      100
aatcatattatatactataattttt tttttaaatcaaaacaacaggttc atatttagttatagatgatgacta ataattgataaattagtgaagtgt
ttgataaatttaataatttttattaat tgataaataataaataacataaaaa atatttaatacattattgttttatt ttaaatataaataaataataatga 200
caaagtgaattttaataaaacatgt aaagataaaaattttaaaaaatta ataaattataaaataaaaacatta tttgaaatatcgctaaaaaaataac
taaactaagaaaaagtttttagtaag agtatttaaaaattaaatttaaaat aaatttcattcttaatactttttgt tttgggtgtttaaataatataaaaa 400
gaattaattgaaaagtaataaaaag aacagaaaaaaaattattttcaaat aagtttacttaatttaaaagtaaat aaaagaacagaaaaataagaagag
ggataagttgatatgttttcatTTG TTATCCATCCAGTCTTATCTTAGGT TTTTCATTCACTCATCATCATCGGA ATGGCTCTCTCTTTAGCCATTACG 600
TTCCTCTGTTTCGCGATAACCTGCT TCGACCCCAACCATTTCAATCCAA TCCGTCCAAACCTACTCAAATTCC CGATACCCCTCCGCCCGAATTCC
ATGTTCTCTCTCTCTCTTTCAAT GACATTTCCCTCAAAACGGTTACCC CCGACAACAAAACCTCTTTCGTTTG CCGCTTTAACGCTACCCAGCTCGAA 800
Ggtactctttctctctcatctctt atcttctgtatctttctttcttga tttcatccactgcttttttttacc tatcttctatcccttagaaaacccc
caaaaataacttctcaacactggcg agtatttgttggatgtgaccccaat tatactgatctgcaatgtttatatt tttatcttcattcatattgttatct 1000
gttttagctttttcaattttcagata tttcagatattttcatgatgcatttg atatatccttaggctagacagttaa catacatttaaaatttaagtctcac
tgcagTTCAGGAGACCAACCAACCA TACGCAGAAACCTATGCTGTAGGGA GACACATTGCAATGTCCGCTGACAA AGCTCGAAGAGTAGTCGATCAGATT 1200
CGCGGGCGTCCGTACGAGGCATCAC TCATGGTATTGGAACCTCATGCCTTA TCGTGCTGTGAAGCCATTATTAAG ATTGTGTTTTCTGCTGGAGCAAATG
CAAGCCATAACTTGGGTCTGAGCAA ATCAAGTTTGTATTAGTAAAGCA GAGGTTAATGAAGGCAAGACTTTGA AAAGAGTAAGAGCTCGAGCTCAGG 1400
GCGGGCTAACCAATTCTAAAGCGT ACTTGTGCATATTACCATTCGGTGA GAGGTTTACCTGATGAGTCTGACAA GGAAGAAAATTCCTCTTGATCCTGT
TTGAAGACAGCAAAATATGTATCC CTGTTATGAATCTTCTCCAAGCATT ACATTGATGGAACCTTCTGCATT TATATTATGTAGTTTATTAGTGA 1600
CTTCTACATGAGTGTGAGAATCCT ATAAATGTAATAATTTCAACTTAA TACTAGAATGAAATGTTTACCATT GATATTGATATGAATttattttgt
tgtactagacttaggttttagttt aaaaatcttgcaatgcgttgaaacat agcctattagcatatattgaatt gaatatgttgatgcacaagggtgat 1800
aaagatgaataatgaacaagagaga aaagagagaataatacaaaactttc attattcttgaatggttaccacaaat ggttgacagaaaaattgtctcacac
actgcaaaaggaataaatgtataat ttgttgacaccactcactgtcttaa atacaagtgaggacttaggagaaat actaaaataccccttaacaaacttt 2000
gtctacaagtttctacaaaatga attaattctaaccactcccttaat tcatattcagcttaatacaaatgaac aacactaatttcatccctcaagcat
agaatcgatcagctcttgatgcct tcattagaacatctgtcagctgctt ctgggtgctacaatgcataacttct agca

```

B

```

1      10      20      30      40      50      60      70      *      80      90      100
Ps  MALSLAINVPPVRDNLRRPQPFQSQFRPNLLKFPISPSSARIRCSSSSSFNDISLKTVPDNKNSFVCRFNATQLEVQETNQPYAETAYVGRHIRMSADKA
Ph                                     MIKKNNKKR.T.VCRL.KY.G..TH..
Nt                                     ML..KKT.V..L.E..S.....
Mp                                     MQ..TSNKKIR..AK..H..PH.V
Ec                                     M..I.KH..A.S..Q.V

110     120     130     140     150     160     170     180     190     200
Ps  RRVVDQIRGRPYEASLMVLMPYRACEAI1KIVFSAGANASHNLGLSKSSLFISKAENVNEGKTLKRVRARQAQGRANQILKRTCHITITVRGLPDESDEK
Ph  ...I.....S..ET..I.....YP.L.LIS.VA..GI..MNFDET..I.....P.A.KFKP..K.QSYP.KRS.....VL.DISLNEKY.
Nt  ...IN.....S..ET..I.....YP.L.LIY..A...Y.M.SSEAN.V.....G.T.V.KLKP..R..SFP.KRS.....VMKDISLDEYV
Mp  ...S.....S..QA..I..F.....NP.LQLLS..A..N..F.....TN....EIQ..K.TFF..FQP.....GYP.H.P.....VLNI..K
Ec  .L.A.L...KKVSQA.DI.TYTNKK.AVLVK.VLE..I...E..D.ADIDD.KVT.IF.D..PSM..IMP..K...DR.....S...VV.SDR

Ps  ENSS
Ph  .YI.ENHGPS
Nt  .MY.LKKTRWKKSTAMPYRDMYNSGGLWOKK

```

Fig. 2. Pea *rpl22* contains an intron that splits the gene into an exon encoding a putative transit sequence and an exon encoding the evolutionarily conserved ribosomal protein. (A) Nucleotide sequence of pea *rpl22*. Upper case letters refer to nucleotides found in both the genomic DNA and in the cDNA clone PsCL22-6. An intron extending from position 802 to 1105 is denoted by boldface lower case letters. The assignment of the first ATG (underlined) in the cDNA as the site of translational initiation is supported by an in-frame TAG stop codon 30 bp upstream of the ATG codon. The underlined TGA codon beginning at nucleotide 1492 terminates translation. Nucleotide 1687 is followed by a poly(A) tract in the cDNA. (B) Alignment of pea CL22 (Ps) amino acid sequence with the amino acid sequences of *Pelargonium hortorum* (Ph), *Nicotiana tabacum* (Nt), *Marchantia polymorpha* (Mp), and *E. coli* (Ec) ribosomal proteins (Shinozaki *et al.*, 1986; Fukuzawa *et al.*, 1988; Zurawski and Zurawski, 1985). Numbering is from the predicted amino terminal amino acid of pea CL22. Identical residues are denoted by dots (.). The asterisk marks the location of the intron in pea *rpl22*, which splits the Val76 codon (GTT) between G and T, relative to the amino acid sequence of pea CL22.

In order to include a chloroplast sequence from a plant of reasonably close relationship to legumes, we sequenced *rpl22* from cpDNA of *Pelargonium hortorum* (Figure 2B), a member of the same subclass, Rosidae, as legumes (Cronquist, 1988). A molecular phylogeny (unpublished data) based on sequences of the chloroplast *rbcL* gene agrees with the traditional classification (Cronquist, 1988) in placing *Pisum* and *Pelargonium* as closer relatives to each other than to either of the other dicots (*Nicotiana* and *Spinacia*) included in the phylogenetic analysis of Figure 5. If the putative transfer of *rpl22* to the nucleus of legumes had occurred more or less simultaneously with its loss from the chloroplast

genome, then the chloroplast and nuclear *rpl22* genes would be orthologous to one another and the closest relative of the pea nuclear gene should be the *Pelargonium* chloroplast gene. However, bootstrap analysis places the pea nuclear gene outside the entire lineage of angiosperm chloroplast genes, both dicot and monocot (Figure 5). This tree also corresponds to one of three equal length shortest trees found by maximum parsimony analysis. In all three trees, the pea nuclear gene was placed outside the entire clade of angiosperm chloroplast genes; this result is strongly supported by bootstrap analysis (94% replication). All alternative branching patterns occurring in the other two

Fig. 3. Pea *rpl22* is transcribed into poly(A)⁺ RNA. Total (T), poly(A)⁺ (+), and poly(A)⁻ (-) RNAs from light-grown pea seedlings were fractionated on a 1.5% agarose/formaldehyde gel, transferred to a nylon filter, and sequentially hybridized with a pea *rpl22* probe and a tomato *cab4* gene probe provided by E. Pichersky. Procedures for RNA extraction and analysis are as previously described (Baldauf and Palmer, 1990).

Fig. 4. *rpl22* is a single copy nuclear gene in pea. Ten micrograms each of 12 preparations of total pea DNA were cut with *EcoRV*, electrophoresed in a 0.9% agarose gel, transferred to a nylon membrane, and hybridized with ³²P-labeled PsCL22-6 cDNA. The left two lanes show the restriction fragment length polymorphism phenotypes of two parental pea lines (Slow and JI 1794), and the remaining lanes show the phenotypes from ten F₃ families, each derived from a different F₂ plant.

shortest trees showed <30% bootstrap support. An additional four steps are required to produce a parsimony tree in which the pea gene is internal to the monocot lineage as a sister group to the three dicot chloroplast genes, and a total of six extra steps are required to place pea as the sister taxon to *Pelargonium* (bootstrap values of 1.6% and 0.5%, respectively). In contrast, only one additional step is required to produce a tree that places pea *rpl22* outside the entire lineage of land plant chloroplast genes, raising the possibility (supported by a bootstrap value of 14%) that the transfer event may have occurred even earlier. Phylogenies identical to the one generated by bootstrap parsimony analysis of amino acid sequences (Figure 5) were also obtained by the PROTPARS algorithm of Phylip 3.1 (provided by J. Felsenstein), by two different distance methods [neighbor joining (Saitou and Nei, 1987) and Fitch–Margoliash (Fitch

Fig. 5. *rpl22* was transferred to the nucleus in a common ancestor of flowering plants. The tree shown is the 50% majority rule tree derived by bootstrap analysis of deduced amino acid sequences. Numbers correspond to bootstrap percentages. The tree is rooted using eubacteria as outgroups and horizontal branches are drawn proportional to length. This tree has a total length of 248 steps, and its consistency index is 0.81, excluding autapomorphies. See, Hiratsuka et al. (1989), Fukuzawa et al. (1988), Tanaka et al. (1986), Zhou et al. (1989), McLaughlin and Larrinua (1987), Christopher et al. (1988), Kao and Wu (1990), Evrard et al. (1990), Zurawski and Zurawski (1985) and Ohkubo et al. (1987), for sources of *rpl22* sequences besides the two reported here.

and Margoliash, 1967) methods], and by parsimony analysis of nucleotide sequences (data not shown).

Discussion

We have shown that *rpl22* sequences are present in cpDNA of all land plants examined with the exception of legumes (Figure 1) and that in pea, *rpl22* is located in the nucleus. Pea *rpl22* contains two exons, which appear to encode functionally distinct domains of the *rpl22* polypeptide and to be of different evolutionary origins. The 5' exon encodes a putative chloroplast targeting peptide and the 3' exon encodes an evolutionarily conserved ribosomal protein. The 3' exon of the pea nuclear gene is consistently placed by phylogenetic analyses within the clade of chloroplast genes (Figure 5) and is extremely distantly related to its plant cytosolic homolog (Madsen et al., 1991). Therefore, the 3' exon is clearly of chloroplast origin, rather than being derived by substitution of some ancestrally mitochondrial or nuclear *rpl22*-like gene. Sequences corresponding to the pea *rpl22* 5' exon and intron have not been found in any characterized chloroplast or bacterial *rpl22* gene (Shinozaki et al., 1986; McLaughlin and Larrinua, 1987; Christopher et al., 1988; Fukuzawa et al., 1988; Hiratsuka et al., 1989; Zhou et al., 1989; Evrard et al., 1990; Kao and Wu, 1990) and are likely of nuclear origin. Similar gene structures, that is, separate exons that encode targeting and mature protein domains, have been observed in nuclear genes encoding the small subunit of ribulose biphosphate carboxylase, a chloroplast protein, and several mitochondrial proteins (Takiguchi et al., 1987; Tsuzuki et al., 1987; Takeshima et al., 1988; Lagace et al., 1987; Wolter et al., 1988; Haas et al., 1989; Dyer et al., 1989).

Because phylogenetic analysis places pea *rpl22* entirely outside the clade of angiosperm chloroplast genes, we postulate that the gene was transferred to the nucleus well before its subsequent loss from the chloroplast genome of a legume ancestor. Based on the best phylogenetic tree, nuclear transfer probably preceded chloroplast loss by at least 100 million years, the difference between the age of legumes

[50–100 million years (Crepet and Taylor, 1985)] and the date of the monocot–dicot split [~200 million years (Wolfe *et al.*, 1989a)]. The transfer might be even more ancient, by at least 200 million years, if one considers the possibility, suggested by trees of one step longer than the best tree, that the gene was transferred prior to the radiation of land plants.

We expect that the transfer of a gene to the nucleus and its loss from the chloroplast would be temporally uncoupled events, since gene loss cannot occur until the nuclear gene becomes functional, producing a chloroplast-targeted protein. Prior to the chloroplast's loss of a gene, the transferred gene must acquire sequences conferring active and properly regulated nuclear transcription, correct transcript processing, and product targeting to the chloroplast. Acquisition of the proper combination of all of these sequences is likely to be rare, and the precise history of acquisitions is likely to be different for each case of successful gene transfer. However, out of the myriad of activation routes imaginable, two seem particularly appealing given the intron/exon structure of the transferred *rpl22* gene. First, the transferred DNA may have inserted into a copy of an active nuclear gene encoding a chloroplast protein, specifically within an intron separating its transit peptide domain from the rest of the gene. In this case the transferred gene could have readily cannibalized many of the elements necessary for expression and targeting ('the lucky insertion' scenario). Second, the transferred gene may have inserted into an inert nuclear locus, but have acquired transit peptide and 5' expression sequences via a form of exon-shuffling (Gilbert, 1987; Traut, 1988) mediated by recombination between the 'outtron' sequences upstream of the inserted gene and the intron sequences of an active nuclear gene of the type described above ('the lucky shuffling' scenario).

Regardless of how *rpl22* became functional in the nucleus, we would expect the period of potential gene activation to be necessarily limited, to perhaps 10–20 million years, because of the disabling occurrences of deletion, nonsense, and mis-sense mutations that would render the transferred gene permanently inactive. In contrast, we have estimated that transfer of *rpl22* to the nucleus preceded loss of the chloroplast *rpl22* gene by at least 100 million years. Such a time period seems far too long for a gene to exist in the nuclear genome in the absence of selection without becoming a pseudogene, given the 0.6 substitutions per nucleotide position estimated to occur during this period (Wolfe *et al.*, 1989b). These considerations lead us to conclude that the transferred DNA must have become activated fairly soon after its transfer to the nucleus and that a nuclear *rpl22* homolog has been under selection in plants during most or all of angiosperm evolution. The product of such a nuclear gene might have functioned as a chloroplast ribosomal protein during this entire period, thereby providing a duplicate function to the one provided by the chloroplast *rpl22*. Alternatively, in most angiosperm lineages, the nuclear *rpl22* product might function elsewhere (in the cytoplasm or mitochondrion), either as a ribosomal protein or perhaps in some other capacity. According to this general scenario, the advent of targeting the protein to the chloroplast in a legume-specific ancestor could have led to relaxed selection on and the loss of the legume chloroplast *rpl22* gene. We would predict that other angiosperms besides legumes should contain a nuclear *rpl22* gene.

The presence in both the nucleus and organelle of duplicate copies of an ancestrally organellar gene has also been

proposed for the chloroplast gene *tufA* (Baldauf *et al.*, 1990) and the mitochondrial gene *atp9* (van den Boogaart *et al.*, 1982). In the former case, functional divergence and specialization of the duplicate genes has been postulated to be part of the driving force leading to retention of both genes in certain algal lineages. In other lineages, the chloroplast *tufA* gene has been lost, apparently, as in the case of *rpl22*, after a long period of coexistence of nuclear and chloroplast duplicates. Whatever the precise history of these evolutionary gene transfers, it is clear that they may be much more complicated than initially suspected and may allow for surprising and perhaps biologically important modes of molecular evolutionary divergence.

Materials and methods

CpDNAs were isolated from sucrose gradient-purified chloroplasts (Palmer, 1986). For the inheritance analysis, total DNA was prepared (Doyle and Doyle, 1987) from pea parental lines Slow, a marker line derived from cultivated material, and JI 1794, a line obtained from the John Innes Institute, Norwich, UK, and which is probably a representative of subspecies *humile*. The segregating F₂ population consisted of 41 plants that had been scored for 45 other segregating markers besides the *rpl22* polymorphism. Linkage between markers was estimated from F₂ segregation data using the method of maximum likelihood as applied by the LINKAGE-1 program (Suiter *et al.*, 1983).

A lambda gt10 cDNA library was prepared from poly(A)⁺ RNA extracted from 10–12-day-old pea (*Pisum sativum* var. Little Marvel) seedlings (Gubler and Hoffman, 1983). This library was screened with a ³²P-labeled 209 bp *Pst*I–*Xba*I fragment containing the first 176 bp of the coding region of the tobacco chloroplast *rpl22* gene and 33 bp of 5' non-coding sequence (Palmer, 1986). The hybridizing recombinant bacteriophage that contained the largest cDNA insert was subcloned to generate a plasmid clone (named PsCL22-6) containing a 875 bp cDNA insert within the vector pUC119.

A pea (var. Little Marvel) genomic DNA library was constructed in Lambda Dash (Stratagene). Genomic DNA was partially digested with *Sau*3AI and fractionated by centrifugation through a sodium chloride gradient. DNA ranging in size from 17–25 kb was collected, ligated into *Bam*HI-digested vector, and packaged *in vitro*. A total of 4 × 10⁵ recombinants was screened by plaque hybridization using ³²P-labeled PsCL22-6 cDNA. One hybridizing clone was isolated and a 7.0 kb *Hind*III restriction fragment containing the hybridizing region was subcloned into pUC119.

The nucleotide sequences of both strands of the entire cDNA and a portion of the genomic DNA were determined from nested deletion (Henikoff, 1987) fragments by the dideoxy termination method using Sequenase (US Biochemicals). *rpl22* was mapped to a cloned, 4.0 kb *Pst*I fragment from *Pelargonium hortorum* cpDNA and both strands of the gene were sequenced using synthetic oligonucleotides as primers.

Phylogenetic analyses used all unambiguously aligned positions, which correspond to a 93 amino acid internal segment (Figure 2B, positions 96–188). Amino acids were scored as equally weighted, unordered, multistate characters. The branch and bound algorithm of PAUP (version 3.0k–; David Swofford, Illinois Natural History Survey) was used to find the shortest trees. Bootstrap analysis consisted of 500 replications using the asis addition sequence and holding four trees at each step. The *Zea rpl22* sequence (McLaughlin and Larrinua, 1987) used in Figure 5 was modified by the introduction of a single nucleotide deletion in a string of seven adenosine residues corresponding to amino acids 117–118. As published, the *Zea* sequence following position 118 consists of eight amino acids with no similarity to any known CL22, followed by a stop codon. The deletion extends the *Zea* open reading frame by over 30 amino acids, 27 of which are identical to the *Oryza* CL22. We suggest that either the *Zea* gene has been subject to a very recent insertion mutation or the published sequence in this region is the result of a sequencing or cloning artefact. In either case, use of the extended form of the sequence is valid for phylogenetic analysis.

Acknowledgements

We thank Ingrid Apel, Deborah Katz-Downie, and Janet Hampton for technical assistance; Hans Bohnert and Masahiro Sugiura for providing *rpl22* sequences from *Cyanophora* and rice, respectively, prior to publication; David Swofford and Joseph Felsenstein for providing programs for phylogenetic analysis; John Doebley and David Swofford for helpful

discussions; and Jocelyn Shaw and Ken Wolfe for critical reading of the manuscript. This work was supported by NIH R29 GM-38769 to J.S.G., NIH NRSA GM-11948 to P.J.C., and NIH RO1 GM-35087 to J.D.P.

References

- Baldauf, S.L. and Palmer, J.D. (1990) *Nature*, **344**, 262–265.
- Baldauf, S.L., Manhart, J.R. and Palmer, J.D. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5317–5321.
- Campbell, W.H. and Gowri, G. (1990) *Plant Physiol.*, **92**, 1–11.
- Christopher, D.A., Cushman, J.C., Price, C.A. and Hallick, R.B. (1988) *Curr. Genet.*, **14**, 275–286.
- Crepet, W.L. and Taylor, D.W. (1985) *Science*, **228**, 1087–1089.
- Cronquist, A. (1988) *The Evolution and Classification of Flowering Plants*. New York Botanical Garden, NY.
- Doyle, J.J. and Doyle, J.L. (1987) *Phytochem. Bull.*, **19**, 11–15.
- Dyer, M.R., Gay, N.J., Powell, S.J., Walker, J.E. (1989) *Biochemistry*, **28**, 3670–3680.
- Evrard, J.L., Kuntz, M. and Weil, J.-H. (1990) *J. Mol. Evol.*, **30**, 16–25.
- Fitch, W.M. and Margoliash, E. (1967) *Science*, **155**, 279–284.
- Fukuzawa, H., Kohchi, T., Sano, T., Shirai, H., Umesono, K., Inokuchi, H., Ozeki, H. and Ohyama, K. (1988) *J. Mol. Biol.*, **203**, 333–351.
- Gilbert, W. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 901–905.
- Gubler, U. and Hoffman, B.J. (1983) *Gene*, **25**, 263–269.
- Haas, R.C., Korenfeld, C., Zhang, Z., Perryman, B., Roman, D. and Strauss, A.W. (1989) *J. Biol. Chem.*, **264**, 2890–2897.
- Henikoff, S. (1987) *Methods Enzymol.*, **155**, 156–165.
- 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.
- Jacobsen, K., Laursen, N.B., Jensen, E.O., Marcker, A., Poulsen, C. and Marcker, K.A. (1990) *Plant Cell*, **2**, 85–94.
- Kao, J. and Wu, M. (1990) *Nucleic Acids Res.*, **18**, 3067.
- Lagace, M., Howell, B.W., Burak, R., Lusty, C.J. and Shore, G.C. (1987) *J. Biol. Chem.*, **262**, 10415–10418.
- Madsen, L.H., Dreiberg, J.D. and Gausing, K. (1991) *Curr. Genet.*, **19**, 417–422.
- McLaughlin, W.E. and Larrinua, I.M. (1987) *Nucleic Acids Res.*, **15**, 4356.
- Michalowski, C.B., Pranzagl, B., Löffelhardt, W. and Bohnert, H.J. (1990) *Mol. Gen. Genet.*, **224**, 222–231.
- Milligan, B.G., Hampton, J.N. and Palmer, J.D. (1989) *Mol. Biol. Evol.*, **6**, 355–368.
- Mirkovitch, J., Mirault, M.-E. and Laemmli, U.K. (1984) *Cell*, **39**, 223–232.
- Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F. and Osawa, S. (1987) *Mol. Gen. Genet.*, **210**, 314–322.
- Oliver, J.L., Marin, A. and Martinez-Zapater, J.M. (1990) *Nucleic Acids Res.*, **18**, 65–73.
- Palmer, J.D. (1985) *Annu. Rev. Genet.*, **19**, 325–354.
- Palmer, J.D. (1986) *Methods Enzymol.*, **118**, 167–185.
- Palmer, J.D. and Stein, D.B. (1986) *Curr. Genet.*, **10**, 823–833.
- Palmer, J.D. (1991) In Bogorad, L. and Vasil, I.K. (eds) *The Molecular Biology of Plastids*. Academic Press, Orlando, FL, Vol. 7A, pp. 5–53.
- Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol.*, **4**, 406–425.
- 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, 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.
- Spielmann, A., Roux, E., von Allmen, J.-M. and Stutz, E. (1988) *Nucleic Acids Res.*, **16**, 1199.
- Suiter, K.A., Wendel, J.F. and Case, J.S. (1983) *J. Hered.*, **74**, 203–204.
- Takeshima, H., Joh, T., Tsuzuki, T., Shimada, K. and Matsukado, Y. (1988) *J. Mol. Biol.*, **200**, 1–11.
- Takiguchi, M., Murakami, T., Miura, S. and Mori, M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6136–6140.
- Tanaka, M., Wakasugi, T., Sugita, M., Shinozaki, K. and Sugiura, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6030–6034.
- Tobin, E.M. and Silverthorne, J.A. (1985) *Annu. Rev. Plant Physiol.*, **36**, 569–593.
- Traut, T.W. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2944–2948.
- Tsuzuki, T., Obaru, K., Setoyama, C. and Shimada, K. (1987) *J. Mol. Biol.*, **198**, 21–31.
- van den Boogaart, P., Samallo, J. and Agsteribbe, E. (1982) *Nature*, **298**, 187–189.
- von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) *Eur. J. Biochem.*, **180**, 535–545.
- von Heijne, G. (1986) *EMBO J.*, **5**, 1335–1342.
- Weeden, N.F. and Wolko, B. (1990) In O'Brien, S.J. (ed.), *Genetic Maps*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 5th edition, pp. 6.106–6.112.
- Wolfe, K.H., Li, W.-H. and Sharp, P.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 9054–9058.
- Wolfe, K.H., Gouy, M., Yang, Y.-W., Sharp, P.M. and Li, W.-H. (1989a) *Proc. Natl. Acad. Sci. USA*, **86**, 6201–6205.
- Wolfe, K.H., Sharp, P.M. and Li, W.-H. (1989b) *J. Mol. Evol.*, **29**, 208–211.
- Wolter, F.P., Fritz, C.C., Willmitzer, L., Schell, J. and Schreier, P.H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 846–850.
- Zhou, D.-X., Quigley, F., Massenet, O. and Mache, R. (1989) *Mol. Gen. Genet.*, **216**, 439–445.
- Zurawski, G. and Zurawski, S.M. (1985) *Nucleic Acids Res.*, **13**, 4521–4526.

Received on June 6, 1991