Endosymbiotic Evolution: Transfer of Plastid DNA to the Nucleus and its Stability Following Integration

Anna Erica Sheppard

A thesis submitted for the degree of Doctor of Philosophy

Discipline of Genetics

School of Molecular and Biomedical Science

The University of Adelaide

September 2009

Contents

Amendments to Thesis

Page 18, last line should read:

synthesised using random hexamer primers followed by PCR amplification of the *neo* cDNA

Page 19, 3rd paragraph should read:

For technical reasons, the two editing regions were amplified separately from cDNA samples synthesised using different *neo* primers. In contrast to the results obtained using random hexamer cDNA, some editing was observed at all three editing sites, but the majority of transcript remained unedited (Figure 5). The proportion of edited transcript was slightly higher in *tp*neoACG8-2 compared to *tp*neoACG5-3, which may or may not have a biological basis. To confirm that the presence of detectable edited transcript was the result of using a different primer for reverse transcription, rather than different PCR primers, the experiment using random hexamer cDNA was repeated with the same PCR primers as those used for the *neo*-specific cDNA experiment. Again this revealed no detectable editing (Figure 4B).

Chapter 3, Figure 4 legend should read:

Arrows indicate primers that were used for PCR prior to sequencing. In **A**, PCR was performed with primers spanning the entire *neo* gene and in **B**, PCRs were performed separately for the two regions containing editing sites. Primers indicated are (left to right) psbF+Xho1 and neoR2(AL) in **A** and psbF+Xho1, TneoR1, NPTF2 and neoR2(AL) in **B**. See Table 2 for primer sequences.

Abstract

In eukaryotes, cytoplasmic organellar genomes have contributed to a wide variety of nuclear genes. Therefore DNA transfer from plastids and mitochondria to the nucleus has been an important driving force in eukaryotic evolution. Recent advances have enabled this transfer process to be demonstrated experimentally for the plastid in *Nicotiana tabacum*. Investigation of the frequencies, mechanisms and consequences of plastid-to-nucleus DNA transfer will shed further light on the complexities of endosymbiotic evolution.

To determine the frequencies of plastid-to-nucleus DNA transfer in the male and female germlines of *N. tabacum*, reciprocal crosses were performed using a transplastomic line (*tp*7) as either the male or the female parent. In its plastid genome *tp*7 contained a kanamycin resistance gene (*neo*) under the control of a nuclear promoter, so any resistant progeny were expected to be the result of *neo* transfer to the nucleus. In the cross where *tp*7 was the male parent, 13 resistant plants were obtained from a screen of 146,000 progeny indicating a transfer frequency in the male germline of approximately 1 event for every 11,000 pollen grains. Screening of 273,000 progeny from the cross where *tp*7 was the female parent revealed only 1 resistant plant, demonstrating a much lower transfer frequency in the female germline. The programmed degradation of plastids during pollen development may release DNA fragments that integrate into the nucleus, thus explaining the high transfer frequency in the male germline.

To investigate plastid-to-nucleus DNA transfer in somatic cells, a transplastomic line (*tp*GUS) was generated containing *gus* under the control of a nuclear promoter in the plastid genome. Staining of various somatic tissues revealed discrete foci of *gus* expression. The frequency of *gus* transfer to the nucleus was estimated to be 1 event for every 200,000 mature leaf cells. Self-fertilised progeny of *tp*GUS were used to detect stable transfer of *gus* to the nucleus, revealing a similar germline transfer frequency to that obtained with *tp*7.

To determine whether an RNA intermediate is involved in sequence transfer from the plastid to the nucleus, a transplastomic line (*tp*neoACG) was generated containing, in its plastid genome, a nuclear promoter-driven *neo* gene with a start codon that required plastid RNA editing. Screening revealed a number of kanamycin-resistant progeny plants. Surprisingly, *neo* was unedited in these plants, indicating that *neo* was active in the absence of RNA editing. Analysis of *tp*neoACG revealed that only a low proportion of *neo* transcripts were edited, thus precluding unequivocal conclusions regarding the importance of RNA in plastid-to-nucleus transfer.

The stability of newly integrated plastid DNA sequences in the nucleus was analysed using kanamycin-resistant progeny plants of *tp*7. Around half of the plants showed non-Mendelian

segregation, with fewer resistant progeny than expected. Loss of the *neo* gene was shown to be responsible for this instability. Therefore, plastid DNA integration into the nuclear genome and its subsequent deletion both occur, with the latter process varying from very frequent at some nuclear loci to undetectable at others. These may be important evolutionary processes in the generation of novel nuclear genes.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Anna Sheppard and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed .....*Date*.....

List of Publications

Sheppard A.E., Ayliffe M.A., Blatch L., Day A., Delaney S.K., Khairul-Fahmy N., Li Y., Madesis P., Pryor A.J. & Timmis J.N. (2008) Transfer of Plastid DNA to the Nucleus Is Elevated during Male Gametogenesis in Tobacco. *Plant Physiology* **148**, 328-36.

Sheppard A.E. & Timmis J.N. (2009) Instability of Plastid DNA in the Nuclear Genome. *PLoS Genetics* **5**(1), e1000323. doi:10.1371/journal.pgen.1000323

Acknowledgements

I would like to express my sincere gratitude to the following people:

- Jeremy Timmis, for supervision and guidance, and for giving me the opportunity and support to do this PhD.
- Joan Kelly, my co-supervisor, for always being available to give advice about anything and everything.
- Anil Day and his lab members in Manchester, UK, for performing plastid transformation.
- Yuan Li, for technical assistance with seed plating.
- Mum, for proofreading this thesis.
- Members of the Timmis lab, past and present, for making it such an enjoyable place to work.
- Everyone in the Genetics Discipline who has helped me along the way.
- My family and friends, for their endless support and encouragement.

Chapter 1: Introduction

Endosymbiotic Evolution and Gene Relocation to the Nucleus

An endosymbiotic relationship between two organisms exists when one organism lives inside the other. There are many endosymbiotic relationships in nature, but by far the best studied examples are mitochondria and plastids which reside in eukaryotic cells. These cytoplasmic organelles are descended from free-living prokaryotic ancestors (an α-proteobacterium in the case of mitochondria and a cyanobacterium in the case of plastids), but they have now lost their autonomy and are dependent on host cells for survival. In this way endosymbiotic coevolution has led to the generation of modern eukaryotic cells, which are made up of compartments that can no longer function independently.

Consistent with their endosymbiotic origin, mitochondria and plastids have their own genetic material and transcription and translation machinery, but their genomes contain far fewer genes than those of their prokaryotic relatives (Adams & Palmer 2003; Timmis *et al.* 2004; Keeling & Palmer 2008). Mitochondrial genomes that have been examined contain between 3 and 67 protein-coding genes. Plastid genomes generally have a slightly greater coding capacity with around 80 protein-coding genes present in the plastid genomes of many land plants and as many as 200 in some algae. However in all cases this is a massive reduction from the thousands of genes present in α-proteobacterial and cyanobacterial genomes, implying that the vast majority of genes that were originally present in the endosymbionts have been lost from the organellar genomes during endosymbiotic coevolution. One reason for this is that some genes were probably no longer required in the new intracellular environment and were therefore lost altogether. Another reason for organelle genome size reduction is that many genes have relocated to the nuclear genome. These relocated genes, when they are functional, are transcribed and translated in the nucleus/cytoplasm genetic compartment. The resulting protein products are then imported into the appropriate organelle. In this way mitochondria and plastids contain many more proteins than are encoded in their own genomes.

Over time, this process of functional gene relocation to the nucleus has resulted in modern eukaryotic nuclear genomes containing many genes that code for organellar proteins. In some cases this gene relocation appears to have ceased, such as in animals where the set of genes encoded in the mitochondrial genome is highly conserved (Boore 1999). However, in other cases functional gene relocation is still occurring with a remarkably high incidence. For example, Millen *et al.* (2001) inferred 24 independent losses of *infA* from the plastid genome of 309 angiosperms surveyed. These losses presumably represent plastid-to-nucleus gene transfer since *infA* is

thought to be an essential gene (Millen *et al.* 2001). As functional gene transfer to the nucleus must precede loss from the plastid, independent losses do not necessarily mean independent transfer events. Analysis of nuclear *infA* sequences from 4 separate loss lineages suggested 4 independent transfer events, indicating that transfer of *infA* to the nucleus may have occurred as many as 24 times among the 309 species surveyed. A similar situation has been described for the mitochondrial gene *rps10*, for which 26 separate losses from the mitochondrial genome were found among 277 angiosperms surveyed (Adams *et al.* 2000). Similarly to the work with *infA*, a number of nuclear sequences were analysed, again suggesting multiple independent transfers to the nucleus.

While many genes that control organelle biogenesis and function have relocated to the nucleus, the process of functional gene relocation is quite complex. Firstly, there must be DNA transfer from the organelle to the nucleus. Once integrated into the nuclear genome, an organellar gene must somehow acquire nuclear regulatory sequences so that it can be expressed in an appropriate manner. If the protein product is to be targeted to the organelle from which the gene was derived, then an organelle targeting sequence may also be required. These acquisitions could occur fortuitously as a consequence of the site of integration in the nuclear genome or as a result of post-insertional rearrangements. One way these acquisitions may occur is by duplication of regulatory and/or targeting sequences of previously transferred genes. For example, a number of recently transferred mitochondrial genes have presequences homologous to the anciently transferred gene *hsp70*, including *rps10* in *Fuchsia* (Adams *et al.* 2000) and *sdh3* in several plant species (Liu *et al.* 2009). Alternatively, duplication may not be required, as in the case of the *rps14* gene in rice which has integrated into an intron of *sdhB* (Kubo *et al.* 1999). Alternative splicing results in the generation of separate *rps14* and *sdhB* transcripts, but they share the same 5' region which encodes a mitochondrial targeting sequence. These sequences could also be acquired *de novo*, or in some cases they may already be present prior to transfer (Stegemann & Bock 2006; Ueda *et al.* 2008; Liu *et al.* 2009).

There are also many nuclear genes that are derived from organellar genomes but which are involved in non-organellar functions. Martin *et al.* (2002) estimated that 18% of *Arabidopsis* nuclear genes are derived from the plastid ancestor on the basis of sequence comparisons with 3 cyanobacterial genomes, 16 other prokaryotic reference genomes and yeast. The genes for which homologues could only be detected in cyanobacteria or which branched with cyanobacteria in phylogenetic analysis were inferred to be derived from the plastid ancestor. These genes fell into a variety of functional categories and only around half were predicted to be targeted to the plastid. A more extensive analysis of *Arabidopsis*, rice, *Chlamydomonas reinhardtii* and a red alga, by comparison with 9 cyanobacterial genomes, 215 other reference prokaryotic genomes and 13

reference eukaryotic genomes, suggested a cyanobacterial origin for around 14% of genes in each of these four species (Deusch *et al.* 2008). Again, a wide variety of functional categories were represented by these genes which were by no means confined to plastid functions. Therefore, instead of replacing the gene in the organelle, an organellar gene that has integrated into a nuclear chromosome also has the possibility of evolving to perform a new function.

In addition to whole gene transfers from organelles to the nucleus, smaller fragments of organelle DNA can also migrate to the nucleus and become functionally important. Analysis of human, *Arabidopsis*, rice and yeast nuclear genomes revealed 45 organelle DNA insertions that contributed sequences to protein-coding exons, excluding insertions that represented entire organelle genes (Noutsos *et al.* 2007). In the majority of cases, either the sequence came from a non-protein-coding region of the organelle genome, or the organelle reading frame was not maintained, such that the organelle DNA insertions coded for entirely novel protein sequences. The number of exonic organelle DNA insertions revealed in this study is likely to be an underestimate, since highly diverged or very short insertions would not have been detectable (Noutsos *et al.* 2007).

Because of the large number of nuclear genes that have a cytoplasmic organellar origin (in whole or in part), the process of DNA transfer from mitochondria and plastids to the nucleus has been important in shaping eukaryotic nuclear genomes. Therefore, knowledge of the mechanisms and consequences of this DNA transfer will provide a better understanding of eukaryotic genome evolution.

Nuclear Integrants of Organellar DNA

Organellar DNA sequences have been found in the nuclear genomes of a wide variety of eukaryotes. These usually non-functional sequences have been designated numts (nuclear integrants of mitochondrial DNA) and nupts (nuclear integrants of plastid DNA). Although the first of these sequences were discovered over 25 years ago (Farrelly & Butow 1983; Jacobs *et al.* 1983; Timmis & Scott 1983), large-scale characterisation has only been possible with the advent of whole genome sequencing.

Analyses of human nuclear and mitochondrial genomes have revealed the presence of approximately 300 numts (Mourier *et al.* 2001; Tourmen *et al.* 2002; Woischnik & Moraes 2002; Bensasson *et al.* 2003; Mishmar *et al.* 2004; Ricchetti *et al.* 2004). These range in size from very small up to 14,654 bp and all regions of the mitochondrial genome are represented (Mourier *et al.* 2001). Phylogenetic analysis and the identification of human-specific numts indicate that transfer

is a continuous, ongoing process (Mourier *et al.* 2001; Woischnik & Moraes 2002; Bensasson *et al.* 2003; Ricchetti *et al.* 2004).

Analysis of numts in 13 species with sequenced mitochondrial and nuclear genomes revealed enormous variation (Richly & Leister 2004a). *Anopheles gambiae* was found to have no numts, *Caenorhabditis elegans*, *Plasmodium falciparum*, *Drosophila melanogaster* and *Fugu rubripes* each had less than 10, while hundreds were found in human, rice and *Arabidopsis* genomes. The proportion of the nuclear genome represented by numts was highest in the two plant species analysed, with numts comprising 0.1% and 0.17% of the nuclear genomes of rice and *Arabidopsis* respectively. The variation in numt abundance between species could not be explained by differences in mitochondrial or nuclear genome sizes, or nuclear gene density. A similar study analysed nupts in four species with sequenced plastid and nuclear genomes (Richly & Leister 2004b). Here, *P. falciparum* and *C. reinhardtii* were found to have relatively few nupts, while *Arabidopsis* had about 300 and rice over 2000.

Various features of numts and nupts in *Arabidopsis* and rice genomes have been examined. In these species, numts and nupts are often (but not always) found as clusters, where multiple fragments of mitochondrial and/or plastid DNA are present at the same locus (Richly & Leister 2004b; Noutsos *et al.* 2005). This could be due to multiple fragments inserting at the same time, multiple fragments inserting sequentially at the same locus or decay of large fragments following insertion. While bioinformatic studies are unable to distinguish unequivocally between these possibilities, the high level of sequence identity of fragments with organellar genomes and the presence of both mitochondrial and chloroplast DNA at the same locus argue in favour of the first possibility (Richly & Leister 2004b; Noutsos *et al.* 2005). The age of an organelle DNA insertion in the nucleus can be estimated based on comparison with the organellar genome. In rice, nupt ages have been determined in this way revealing a huge bias towards young insertions, with most nupts estimated to have integrated within the past one million years (Matsuo *et al.* 2005). This suggests that many nupts are rapidly eliminated from the nuclear genome.

Bioinformatic studies have provided a great deal of information regarding the distributions and features of numts and nupts in different species. However, they are somewhat limited with respect to determining details of the transfer process because they require making inferences about transfer events that have occurred in the past. Another limitation of bioinformatic studies is the accuracy of the data on which they are based. For example, the honeybee genome has been shown to be very rich in numts (Behura 2007; Pamilo *et al.* 2007), but analysis of an early version of the genome sequence suggested no numts were present (Pereira & Baker 2004). To provide an alternative way of studying the process of organelle-to-nucleus DNA transfer, a number of groups have sought to detect novel transfer events in the laboratory.

Experimental Detection of Organelle-to-Nucleus DNA Transfer

In the yeast *Saccharomyces cerevisiae*, the frequency of mitochondrion-to-nucleus DNA transfer was measured using a strain with *TRP1* (a nuclear gene necessary for tryptophan biosynthesis) and a nuclear origin of replication *ARS1* inserted into the mitochondrial genome (Thorsness & Fox 1993). This strain also had a deletion of the nuclear copy of *TRP1*, so it was auxotrophic for tryptophan. However, it was able to give rise to tryptophan prototrophs as a result of DNA transfer of *TRP1* from the mitochondrion to the nucleus. This occurred at a rate of approximately 5×10^{-6} events per cell per generation (Thorsness & Fox 1993). Due to the presence of *ARS1* it was possible for the mitochondrial DNA to be maintained in the nucleus as a plasmid. In fact, less than 1% of transfer events involved integration into a nuclear chromosome (Thorsness & Weber 1996). Therefore, the frequency of stable mitochondrion-to-nucleus transfer in *S. cerevisiae* is much less than 5×10^{-6} events per cell per generation.

To measure the frequency of plastid-to-nucleus DNA transfer in a higher plant system, *N. tabacum* (tobacco) was used. Tobacco is the only higher plant system for which plastid transformation is routine (Maliga 2004; Wang *et al.* 2009), thus making it the most feasible system for detecting plastid-to-nucleus DNA transfer experimentally. Large-scale screening is also possible since a single plant can produce hundreds of thousands of progeny. Although the tobacco nuclear genome has not been sequenced and therefore genome-wide characterisation of tobacco nupts has not been possible, hybridisation studies have indicated that the tobacco genome is rich in nupts (Ayliffe & Timmis 1992a, b; Ayliffe *et al.* 1998).

To detect plastid-to-nucleus DNA transfer, two identical transplastomic lines (*tp*7 and *tp*17) were generated in which a kanamycin resistance gene (*neo*) under the control of nuclear regulatory sequences was inserted into the inverted repeat region of the plastid genome (Huang *et al.* 2003). *Neo* was under the control of the cauliflower mosaic virus 35S promoter (which drives nearly ubiquitous expression from the nuclear genome in higher plants) and it contained a nuclear intron derived from the potato *ST-LS1* gene. Therefore, it was expected that *tp*7/*tp*17 would not be kanamycin-resistant, as *neo* should not be expressed from the plastid genome, but if *neo* was able to get into the nucleus, this would result in kanamycin resistance. Seeds derived from crosses between a wild type female parent and a *tp*7/*tp*17 male parent were grown on medium containing kanamycin to select for nuclear transposition of *neo* in the male germline. Screening of approximately 250,000 seedlings gave rise to 16 independently derived plants with heritable nuclear insertions of *neo*. This corresponds to a frequency of approximately one transfer event per 16,000 pollen grains. Since this experiment only detected transfer of *neo*, rather than any other part of the plastid genome, and only events where the entire *neo* gene was inserted into a

transcriptionally active region of the nuclear genome, it must have underestimated the frequency of plastid-to-nucleus DNA transfer. Molecular analysis indicated that the integrants were generally large, with many being >20 kb (Huang *et al.* 2004). Therefore, a large piece of plastid DNA enters the nuclear genome in at least one in 16,000 pollen grains, representing an enormous influx of foreign DNA into the nuclear genome.

Stegemann *et al.* (2003) used a similar approach to investigate plastid-to-nucleus DNA transfer in somatic cells of tobacco. Again, *neo* was inserted into the plastid genome under the control of the 35S promoter, but in this experiment no intron was used. Explants of 20 young leaves from transplastomic plants were subjected to selection on medium containing kanamycin, and 12 independent kanamycin-resistant plants were regenerated. Based on the number of cells present in each tobacco leaf, the authors estimated the transfer frequency to be approximately one event per 5 million cells. However, this may be an underestimate as the calculations did not take into account that not all cells are able to be regenerated (Stegemann *et al.* 2003). Taken together, these results suggest that the frequency of plastid-to-nucleus DNA transfer is much higher in the male germline than in somatic cells, although technical reasons for the large difference in transfer frequencies measured cannot be ruled out.

The alga *C. reinhardtii* has also been used to test the frequency of plastid-to-nucleus DNA transfer. Here, a zeomycin resistance gene under the control of a nuclear promoter was inserted into the plastid genome, both with and without a nuclear intron (Lister *et al.* 2003). Several billion haploid cells were tested for zeomycin resistance, but no transfer was detected. This implies that the transfer frequency is much lower in *C. reinhardtii* than in tobacco, consistent with the finding that its nuclear genome contains only a few small nupts (Richly & Leister 2004b).

Mechanisms of Organelle-to-Nucleus DNA Transfer

While systems are now in place to detect organelle-to-nucleus sequence transfer, a number of questions remain as to how this transfer occurs. For example, there has been some debate over whether organelle-to-nucleus transfer occurs by direct transfer of DNA or by transfer of reverse-transcribed RNA. Some studies support a direct DNA mechanism because numts and nupts can be very long and contain non-coding regions. For example, a 620 kb numt and a 131 kb nupt have been found in *Arabidopsis* and rice respectively, which are strongly suggestive of direct DNA transfer (Stupar *et al.* 2001; Yu *et al.* 2003). Whole genome studies have also indicated that any part of an organelle genome can be transferred to the nucleus and there is no evidence for overrepresentation of highly transcribed regions in numts and nupts (Woischnik & Moraes 2002;

Richly & Leister 2004a, b; Matsuo *et al.* 2005). However, there are a number of examples of functional mitochondrial gene relocation to the nucleus where splicing and RNA editing appear to have occurred prior to transfer, suggesting the involvement of an RNA intermediate (Nugent & Palmer 1991; Grohmann *et al.* 1992; Adams *et al.* 2000). Experimental evidence in support of a direct DNA mechanism has been found in yeast, where it has been shown that in a mutant strain with a high rate of mitochondrion-to-nucleus transfer, transfer occurs independently of an RNA intermediate (Shafer *et al.* 1999). However, it is possible that the mechanism of transfer is variable, so RNA-mediated transfer may also have an important role to play. Determining the mechanism(s) of transfer in other systems will help to shed light on this.

There are other aspects of the mechanism of transfer that have yet to be elucidated, such as the cellular processes involved in organellar nucleic acid migration to the nucleus. One possibility is that transfer occurs following organelle lysis, where DNA or RNA could be released and enter the nucleus. Several lines of evidence support this idea. *C. reinhardtii* appears to have a much lower frequency of plastid-to-nucleus DNA transfer than many other species, since it has very few nupts (Richly & Leister 2004b) and transfer of a marker gene from the plastid to the nucleus could not be detected among several billion somatic cells (Lister *et al.* 2003). This may be because *C. reinhardtii* has only one plastid per cell, such that plastid lysis is expected to be lethal during vegetative growth. In tobacco, transfer of a marker gene from the plastid to the nucleus has been detected at a frequency of 1 in 16,000 pollen grains (Huang *et al.* 2003) or 1 in 5 million somatic cells (Stegemann *et al.* 2003). If this difference in transfer frequency is really due to the different tissues analysed, rather than other experimental differences, a possible explanation relates to the mechanisms underlying uniparental inheritance of plastids. In tobacco, plastids are inherited maternally and programmed plastid degradation occurs during pollen formation. It is possible that the high transfer frequency seen in the male germline is due to liberation of plastid DNA during this degradation process (Timmis *et al.* 2004). If this is the case, it would be expected that the transfer frequency would be much lower in somatic cells, where plastid lysis is likely to occur only occasionally.

Instability of Plastid DNA in the Nucleus

When Huang *et al.* (2003) examined DNA transfer from the plastid to the nucleus in tobacco, 18 kanamycin-resistant plants were obtained from crosses between a wildtype female parent and a transplastomic male parent carrying a nuclear *neo* gene in the plastid. These plants were designated kr2-kr19. Since *neo* had relocated to the nuclear genome, Mendelian segregation ratios would be expected for the kanamycin resistance phenotype in self-fertilised and

backcrossed progeny of kr2-kr19, assuming they were hemizygous at a single locus. In fact, only 12 plants gave rise to the expected genetic ratios. Of the remaining six plants, four showed a reduction in the proportion of resistant progeny from self-fertilisation, and two (kr2 and kr15) did not transmit the resistance phenotype to any progeny. Therefore, for some nuclear integrants the kanamycin resistance phenotype is unstable, with fewer than expected progeny receiving an expressed copy of *neo*. It was not determined whether this instability was variable within a single plant, in which case seed capsules derived from different areas of the plant may give different results with respect to the instability. If this is the case, the three outcomes described above (Mendelian segregation, reduction in the proportion of resistant progeny from self-fertilisation and no resistant progeny) may be a consequence of the particular seed capsules that were tested, rather than representing biologically meaningful groupings. In order to resolve this, it would be necessary to determine the segregation of kanamycin resistance for a number of individual seed capsules for each kr plant.

There are several possible explanations for the instability of the kanamycin resistance phenotype. Epigenetic silencing of transgenes is relatively common in plants (for a review, see Matzke & Matzke 1998), so it could be that *neo* is being silenced. Alternatively, it is possible that deletion of *neo* is responsible for the instability. In support of this, Southern analysis revealed that *neo* was not present in progeny of kr2 and kr15 (Huang *et al.* 2003). If this is the case, it could explain why the high frequency of plastid DNA integration into the nuclear genome is not necessarily detrimental, as without a high frequency of balancing deletion the nucleus would perpetually increase in size.

Aims

In tobacco, plastid-to-nucleus DNA transfer occurs at a remarkably high frequency which can be measured in the laboratory. Using the tobacco system to detect *de novo* transfer events, questions relating to the mechanisms and consequences of plastid DNA transfer to the nucleus can now be tested experimentally.

Crosses using a transplastomic tobacco plant (*tp*7/*tp*17) as the male parent revealed transfer of the *neo* gene to the nucleus in around 1 in 16,000 progeny (Huang *et al.* 2003). Experiments using a different transplastomic line estimated transfer of *neo* to the nucleus in 1 in 5 million somatic cells (Stegemann *et al.* 2003). A possible explanation for this large difference in transfer frequencies is that the programmed degradation of plastids during pollen development releases plastid DNA that is able to integrate into the nucleus, thus accounting for the high transfer

frequency that has been observed in the male germline. If this is the case, it would be expected that the transfer frequency in the female germline would be much lower than that in the male germline. Chapter 2 describes experiments to test this, where reciprocal crosses of tobacco were performed using *tp*7 as either the male or the female parent followed by selection of the progeny to detect transfer of *neo* to the nucleus.

Chapter 2 also describes experiments using a transplastomic line with a visual marker (*gus*) for detecting transfer to the nucleus, as opposed to a selectable marker. This transplastomic line (*tp*GUS) was used to detect somatic transfer, as well as stable germline transfer. As *gus* was present in a single copy region of the plastid genome, while in *tp*7, *neo* was present in the inverted repeat region of the plastid genome, these results were used to investigate the possible effects of copy number and location in the plastid genome on transfer frequency.

While the majority of evidence supports a direct DNA mechanism for plastid-to-nucleus DNA transfer, some evidence suggests the involvement of an RNA intermediate. In chapter 3, a transplastomic line (*tp*neoACG) was generated in which *neo* contained several plastid RNA editing sites, so that any nuclear integrants arising from RNA-mediated transfer should retain a signature of this editing. Progeny of *tp*neoACG were then screened for transfer events in order to directly test the mode of transfer.

In some plants with a newly integrated copy of *neo* in the nuclear genome (which was derived from the plastid genome), kanamycin resistance is present in fewer progeny than would be expected from Mendelian predictions (Huang *et al.* 2003). In chapter 4, the stability of kanamycin resistance was characterised in detail for nine independent lines containing *neo* in the nuclear genome and the molecular mechanism for the loss of kanamycin resistance was determined.

Chapter 2

Transfer of Plastid DNA to the Nucleus Is Elevated during Male Gametogenesis in Tobacco

Anna E. Sheppard[1], Michael A. Ayliffe[2], Laura Blatch[3], Anil Day[3], Sven K. Delaney[1], Norfarhana Khairul-Fahmy[3], Yuan Li[1], Panagiotis Madesis[3], Anthony J. Pryor[2] and Jeremy N. Timmis[1]

[1] School of Molecular and Biomedical Science, University of Adelaide, South Australia 5005, Australia

[2] CSIRO Plant Industry, Australian Capital Territory 2601, Australia

[3] Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, United Kingdom

Plant Physiology, September 2008, Vol. 148, pp. 328–336

STATEMENT OF AUTHORSHIP

Transfer of Plastid DNA to the Nucleus Is Elevated during Male Gametogenesis in Tobacco

Plant Physiology, September 2008, Vol. 148, pp. 328–336

Anna Sheppard (Candidate)

Performed all experiments except as detailed below and assisted in writing the manuscript.

I hereby certify that the statement of contribution is accurate

Signed .....*Date*.....

Jeremy Timmis

Supervised development of work and assisted in writing the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for inclusion of the paper in the thesis

Signed .....*Date*.....

Yuan Li

Screened *tp*GUS progeny and verified staining in positive plants, as shown in Figure 2 J-L.

I hereby certify that the statement of contribution is accurate and I give permission for inclusion of the paper in the thesis

Signed .....*Date*.....

Sven Delaney

Performed experiments presented in Figure 2 H-I, performed quantitative real-time PCR and GUS assays on gametophytic tissues and assisted in writing the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for inclusion of the paper in the thesis

Signed .....*Date*.....

Michael Ayliffe

Supervised development of work, performed Southern blotting and assisted in writing the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for inclusion of the paper in the thesis

Signed .....*Date*.....

Anthony Pryor

Performed reciprocal crosses of tobacco using *tp*7 and NtBAR/GUS as parents.

I hereby certify that the statement of contribution is accurate and I give permission for inclusion of the paper in the thesis

Signed .....*Date*.....

Laura Blatch

Performed plastid transformation and isolation of antibiotic resistant *tp*GUS calli.

Norfarhana Khairul-Fahmy

Performed quantitative analysis on the number of sectors in cotyledons and first true leaves (Figure 5) and sectors shown in Figure 4 A-B.

Panagiotis Madesis

Performed isolation and screening of antibiotic resistant *tp*GUS calli, DNA blot analysis and collection of seeds, fixation and GUS staining of leaf pieces and microscopic analysis of slides for Figure 4 C-H.

Anil Day

Supervised development of work by L. Blatch, N. Khairul-Fahmy and P. Madesis, analysed data presented in Figures 4 and 5 and wrote associated text, excluding analysis of the 18 cm leaf, and commented on the manuscript.

I hereby certify that the statement of contribution is accurate for L. Blatch, N. Khairul-Fahmy, P. Madesis and myself and permission is given for inclusion of the paper in the thesis

Signed .....*Date*.....

Transfer of Plastid DNA to the Nucleus Is Elevated during Male Gametogenesis in Tobacco[1][OA]

Anna E. Sheppard, Michael A. Ayliffe, Laura Blatch, Anil Day, Sven K. Delaney, Norfarhana Khairul-Fahmy, Yuan Li, Panagiotis Madesis, Anthony J. Pryor, and Jeremy N. Timmis*

School of Molecular and Biomedical Science, University of Adelaide, South Australia 5005, Australia (A.E.S., S.K.D., Y.L., J.N.T.); CSIRO Plant Industry, Australian Capital Territory 2601, Australia (M.A.A., A.J.P.); and Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, United Kingdom (L.B., A.D., N.K.-F., P.M.)

In eukaryotes, many genes were transferred to the nucleus from prokaryotic ancestors of the cytoplasmic organelles during endosymbiotic evolution. In plants, the transfer of genetic material from the plastid (chloroplast) and mitochondrion to the nucleus is a continuing process. The cellular location of a kanamycin resistance gene tailored for nuclear expression (35S*neoSTLS*2) was monitored in the progeny of reciprocal crosses of tobacco (*Nicotiana tabacum*) in which, at the start of the experiments, the reporter gene was confined either to the male or the female parental plastid genome. Among 146,000 progeny from crosses where the transplastomic parent was male, 13 transposition events were identified, whereas only one atypical transposition was identified in a screen of 273,000 transplastomic ovules. In a second experiment, a transplastomic β-glucuronidase reporter gene, tailored to be expressed only in the nucleus, showed frequent stochastic expression that was confined to the cytoplasm in the somatic cells of several plant tissues. This gene was stably transferred in two out of 98,000 seedlings derived from a male transplastomic line crossed with a female wild type. These data demonstrate relocation of plastid DNA to the nucleus in both somatic and gametophytic tissue and reveal a large elevation of the frequency of transposition in the male germline. The results suggest a new explanation for the occurrence of uniparental inheritance in eukaryotes.

The plastid (chloroplast) genome of higher plants has been reduced to approximately 130 genes, while its cyanobacterial ancestor is estimated to have contained more than 3,000 genes (Timmis et al., 2004). Most erstwhile prokaryotic genes of the ancestor were either lost or transferred to the nucleus during more than a billion years of endosymbiotic evolution (Timmis et al., 2004) such that extant plastid biogenesis is heavily dependent upon nuclear genes. Thousands of functional nuclear genes in Arabidopsis (*Arabidopsis thaliana*) are derived from the endosymbiont genome (Martin et al., 2002), and, while many of these make products that enter the plastids, many others support cellular processes that are unrelated to plastid biogenesis and biochemistry. A similar scenario underlies mitochondrial evolution (Timmis et al., 2004). Gene transfer is initiated by the transposition of endosymbiont nucleic acid sequences to the nucleus where they may be retained in large numbers as nonfunctional *numts* and *nupts* (nuclear integrants of mitochondrial and plastid DNA [ptDNA], respectively) or, much more rarely, they evolve into novel nuclear genes concerned with organelle biogenesis or with new cellular or extracellular functions (Martin et al., 2002).

Two independent experimental estimates (Huang et al., 2003a; Stegemann et al., 2003) of the frequency of de novo *nupt* formation in tobacco (*Nicotiana tabacum*) differ by more than two orders of magnitude. Both used a similar experimental approach involving transformation of the plastid genome with a selectable marker gene, *neo* (encoding neomycin phosphotransferase), that required nuclear relocation to confer kanamycin resistance. A significant difference between these reports is that the first experiments (Huang et al., 2003a) measured transposition in a male transplastomic parent by screening the progeny of crosses to wild-type females for kanamycin resistance and did not involve selection pressure at the time of transposition. The frequency of transposition was estimated as the proportion of kanamycin-resistant progeny, and one in approximately 16,000 pollen grains carried an active nuclear copy of *neo*. The second experiments (Stegemann et al., 2003) regenerated kanamycin-resistant plants from somatic cells of transplastomic plants that were screened under antibiotic selection. The frequency of transposition was calculated using an estimate of the total number of cells in the screened leaf explants and the number of kanamycin-resistant plants recovered. A transposition event was estimated to occur once in every five million cells screened by this procedure.

[1] This work was supported by the Australian Research Council (grant no. DP0557496).

* Corresponding author; e-mail jeremy.timmis@adelaide.edu.au.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Jeremy N. Timmis (jeremy.timmis@adelaide.edu.au).

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.119107

These two experimental measurements were derived from different tissues, raising the possibility of tissue-specific rates of ptDNA relocation. Male gametes of most angiosperm species undergo a programmed elimination of plastids during pollen development (Yu and Russell, 1994; Nagata et al., 1999), a process that underpins maternal inheritance of organellar genes (Mogensen, 1996). Here, we show that transfer of ptDNA to the nucleus takes place far more frequently in the male germline than in the female germline. We also demonstrate a high frequency of transfer in somatic tissues.

RESULTS

The Frequency of Plastid-to-Nucleus DNA Transposition Is Lower in the Female Germline Than in the Male Germline

In a previous study (Huang et al., 2003a), the frequency of ptDNA transfer to the tobacco nucleus was determined by analysis of progeny derived from fertilization of wild-type plants with pollen from a transplastomic line (*tp7*). Contained within the *tp7* plastid genome was a 35S*neoSTLS2* transgene tailored specifically for nuclear expression (Fig. 1, A and C). This transgene consisted of a region encoding neomycin phosphotransferase under the regulatory control of the cauliflower mosaic virus 35S promoter and terminator (Benfey and Chua, 1990), which allowed selection of nuclear integration events by screening for kanamycin resistance during early seedling growth. A nuclear intron (*STLS2*), designed to prevent translation of a functional protein even if unexpected transcription occurred in the plastid, was inserted into the *neo* reading frame (Fig. 1A). When *tp7* was used as a male parent in a cross to wild-type tobacco, one in approximately 16,000 progeny were resistant to kanamycin, indicative of multiple independent transfers of this plastid transgene to the nucleus and subsequent nuclear expression (Huang et al., 2003a). The selected kanamycin-resistant plants each carried a copy, or sometimes several copies, of 35S*neoSTLS2* integrated into chromosomal DNA that, with some exceptions (Huang et al., 2003a), behaved as single Mendelian loci in genetic studies.

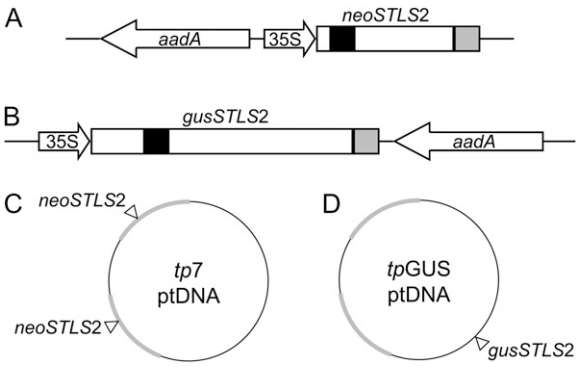

Figure 1. Generation of *tp7* and *tp*GUS transplastomic lines. A and B, Structure of the *tp7* (A) and *tp*GUS (B) transplastomic cassettes. *neoSTLS2* and *gusSTLS2* are under the control of the 35S promoter (arrows) and 35S terminator sequences (gray boxes). The *STLS2* intron is indicated by a black box. The *aadA* spectinomycin resistance gene is under the control of plastid regulatory sequences (Huang et al., 2003a; Zubko et al., 2004). C and D, Location of *neoSTLS2* (C) and *gusSTLS2* (D) in the tobacco plastid genomes of *tp7* and *tp*GUS, respectively. In line *tp7*, *neoSTLS2* is integrated into the inverted repeats (shown in gray) and is therefore present as two copies per plastome. In *tp*GUS, *gusSTLS2* is integrated into the large single-copy region of the plastid genome.

To determine whether the frequency of maternal plastid-to-nucleus DNA transfer was similar to that in the male germline, crosses were performed using the *tp7* transplastomic as the female parent (*tp7* ♀ × NtBAR/GUS ♂), i.e. the reverse direction of that used by Huang et al. (2003a). The male parent used in this reciprocal cross was homozygous for a *BAR/GUS* nuclear transgene, which acted as a nuclear marker to confirm that any emerging kanamycin-resistant progeny were derived from the cross. A complication in screening progeny from the cross in which *tp7* was used as the female parent was that 35S*neoSTLS2* in the *tp7* transplastome conferred partial kanamycin resistance (Huang et al., 2003a; Stegemann et al., 2003), and this was maternally inherited by all the progeny, thus precluding an identical screen to that used previously (Huang et al., 2003a). To overcome this background resistance, the concentration of kanamycin used in screening the reciprocal cross progeny was increased from 150 μg mL^{-1} to 300 μg mL^{-1} (Stegemann et al., 2003). To validate effective selection at this antibiotic concentration, emasculated *tp7* flowers were fertilized with pollen from plants hemizygous for newly transposed 35S*neoSTLS2* in their nuclear genomes (kr18 described previously [Huang et al., 2003a], and kr2.2, kr2.3, kr2.7, kr2.9, and kr2.10 described below; kr, kanamycin resistant). In each cross, all progeny contained the maternal *tp7* transplastome, while one-half of the progeny were expected to also contain a nuclear 35S*neoSTLS2* copy inherited from the hemizygous paternal kanamycin-resistant parent. For each of these crosses, 13 seeds were plated at defined positions among a high density of *tp7* seeds on medium containing 300 μg mL^{-1} kanamycin. After 3 months, progeny containing both plastid and nuclear copies of 35S*neoSTLS2* from all genotypes except kr2.3 were clearly distinguished at this antibiotic concentration from seedlings exhibiting the background resistance conferred by the transplastome alone (Fig. 2A). While it is possible that the kr2.3 nuclear genotype is not detectable above the background resistance of the transplastome, a more likely explanation is that none of the progeny contained a nuclear copy of 35S*neoSTLS2*, because this genotype is known to be highly unstable (A.E. Sheppard and J.N. Timmis, unpublished data). Nevertheless, it can be concluded that this screen is very close in sensitivity to that used previously for crosses where the transplastomic parent was male.

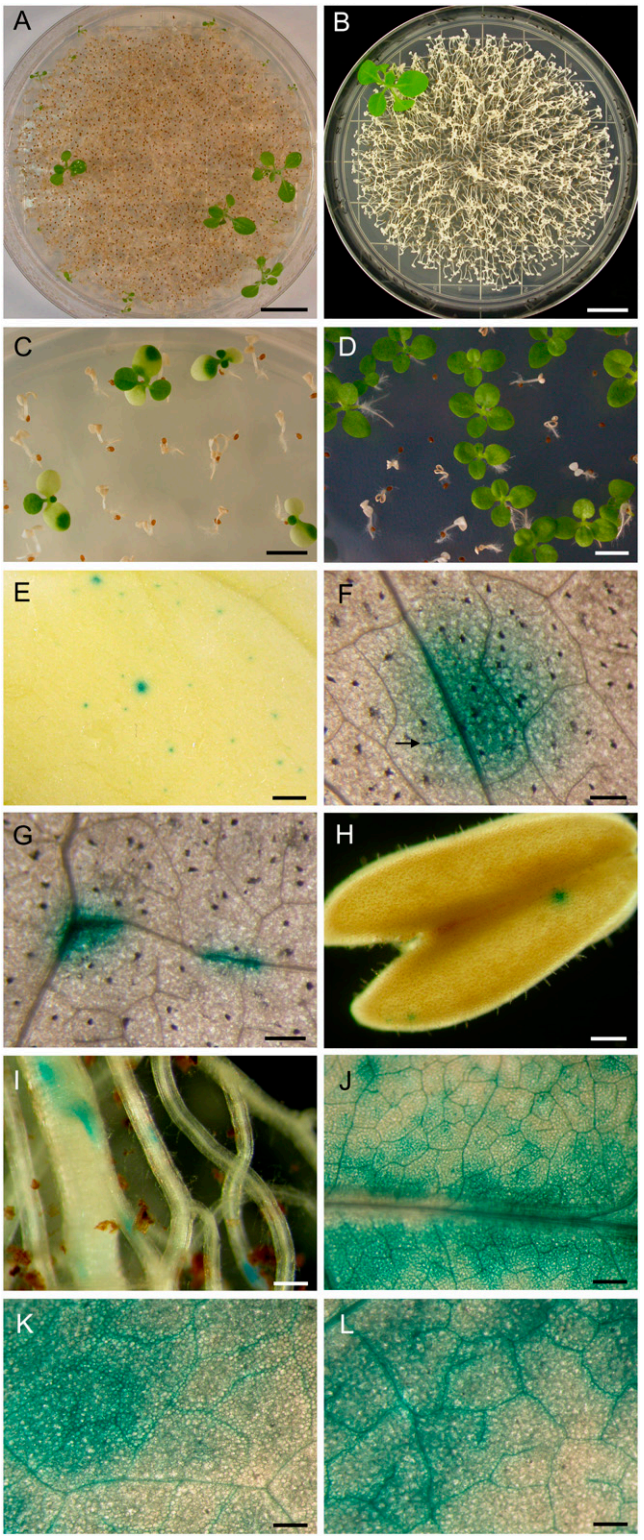

Figure 2. Detection and analysis of ptDNA transfer to the nucleus. A, Four kanamycin-resistant plants from the segregating progeny of *tp7* ♀ × kr2.7 ♂ among 2,000 progeny of self-fertilized *tp7* grown for 3 months on medium containing 300 μg mL^{-1} kanamycin. B, One kanamycin-resistant plant (kr2.4) among 2,000 progeny of NtBAR/GUS ♀ × *tp7* ♂ grown for 11 weeks on medium containing 150 μg mL^{-1} kanamycin. C, Progeny of wild-type ♀ × kr3.1 ♂ grown for 6 weeks on medium containing 150 μg mL^{-1} kanamycin. D, Progeny of wild-type ♀ × kr4.1 ♂ grown for 3 weeks on medium containing 150 μg mL^{-1} kanamycin. E to I, Leaf (E–G), anther (H), and root (I) tissues from *tp*GUS plants that have been histochemically stained for GUS expression. The arrow in F indicates a stained trichome. J, A 3.5-cm leaf of NtBAR/GUS that has been stained for GUS expression. K, A 6-cm leaf of gs1.1 that has been stained for GUS expression. L, An 8-cm leaf of gs1.2 that has been stained for GUS expression. Scale bars = 20 mm (A and B), 5 mm (C and D), 2 mm (E), 0.3 mm (F and G), and 0.5 mm (H–L).

Having established the ability of the selection regime to recognize transposition events efficiently, approximately 273,000 seeds from *tp7* ♀ × NtBAR/GUS ♂ crosses were screened under these conditions. After 3 months, no kanamycin-resistant seedlings similar to the positive controls were observed. However, in contrast to the progeny of the reverse cross (Huang et al., 2003a), some of the plants were not killed by the antibiotic, though they showed severe growth restriction (Fig. 2A). After a further 6 to 9 months, a number of seedlings were still alive, but these lacked the clear phenotype associated with a nuclear 35S*neoSTLS*2 in the transplastomic background (Fig. 2A). The most promising candidates for true kanamycin resistance (green seedlings that were clearly larger than other seedlings on the same plate) were tested by backcrossing to female wild type to remove plastid-localized 35S*neoSTLS*2 genes. Of these, only a single plant (kr3.1) produced a proportion of kanamycin-resistant progeny, and these appeared to be chimeric for kanamycin resistance, containing both resistant and susceptible sectors on their cotyledons (Fig. 2C), suggesting unstable expression of *neo* in the nucleus. Consequently, it was not possible to determine unequivocally the proportion of kanamycin-resistant progeny, because seedlings with only very small resistant sectors could not be clearly distinguished from fully sensitive seedlings. Positive GUS expression in progeny of kr3.1 confirmed that it had resulted from a cross with the homozygous NtBAR/GUS parent (data not shown). PCR analysis confirmed the presence of *neo* in kanamycin-resistant plants (data not shown), indicating that kr3.1 arose from transposition of the plastid-encoded *neo* to the nucleus. Nevertheless, the phenotype of the progeny of this plant was atypical compared with those reported earlier (Huang et al., 2003a), and the antibiotic selection time required for its identification was much greater than that necessary for the positive controls (Fig. 2A).

Given the remarkably low frequency of transplastomic 35S*neoSTLS*2 transposition in the female germline (one atypically resistant plant in 273,000), screening of the original cross undertaken by Huang et al. (2003a) was repeated using the parental lines described above to confirm elevated transfer rates in pollen. That is, pollen from the *tp7* transplastomic line was used to fertilize NtBAR/GUS plants, and seedlings derived from this cross were screened for kanamycin resistance using 150 μg mL^{-1} kanamycin. Ten

kanamycin-resistant seedlings (kr2.1–kr2.10) were isolated from 126,000 seedlings (Fig. 2B), and all except kr2.8 survived to maturity after transplanting to soil. To ensure that the elevated levels of kanamycin (300 μg mL^{-1}) used for screening progeny from the reciprocal crosses did not affect the assay, a further 20,000 seedlings from progeny of the NtBAR/GUS ♀ × *tp*7 ♂ crosses were screened at the higher kanamycin concentration, resulting in the isolation of three additional kanamycin resistant plants.

Positive GUS expression in kr2.1 to kr2.10 (except kr2.8, which did not survive to maturity and was therefore not tested) confirmed that each had resulted from a cross with the homozygous NtBAR/GUS parent (data not shown). The independent origin of these kanamycin-resistant plants was demonstrated by DNA blotting of total cellular DNA restricted with *Xba*I (Fig. 3). DNA from each line showed a unique combination of restriction patterns when hybridized with *aadA*- and *neo*-specific probes. The NtBAR/GUS transgenotype DNA showed weak cross hybridization at high molecular size to the *aadA* probe. DNA of *tp*7 showed the expected hybridization at 10.9 kb to the *neo* probe and at 11.4 kb and 18.5 kb to the *aadA* probe (Huang et al., 2003a). An unexpected band was also seen at approximately 2.5 kb with the *aadA* probe, which is most likely due to a rearrangement of the transplastome arising from recombination between the native *psbA* gene and the plastid-derived *psbA* terminator sequence that regulates *aadA*.

An overall frequency of one kanamycin resistance event in approximately 11,000 pollen grains was obtained from the two experiments that used seedlings derived from NtBAR/GUS ♀ × *tp*7 ♂ crosses. This is consistent with the previously reported transposition frequency of one event in 16,000 pollen grains (Huang et al., 2003a). In marked contrast, only a single atypical transposition event could be detected among 273,000 progeny of the reciprocal cross. These data demonstrate an approximately 25-fold greater frequency of transposition and stable integration of the plastid-encoded 35S*neoSTLS*2 gene into the plant nuclear genome in the male germline compared with the female germline. Even if we include only those kanamycin-resistant lines where transposition was confirmed at the molecular level, which gives a more conservative estimate of nine events out of 126,000 progeny from the cross where the transplastomic parent was male, the reciprocal difference described here is highly significant ($P < 0.001$; Fisher's Exact Test). As a final confirmation of the reliability of the selection regime used for screening *tp*7 ♀ × NtBAR/GUS ♂ crosses, 40,000 seedlings (sufficient to give >97% chance of recovering an event assuming a male germline transposition frequency of one in 11,000) of self-fertilized *tp*7 were screened at 300 μg mL^{-1} kanamycin. Similarly to the *tp*7 ♀ × NtBAR/GUS ♂ cross described above, no resistant seedlings could be observed after 3 months, but some were still alive after a further 3.5 months. The most promising candidates for true kanamycin resistance were tested by backcrossing to female wild type, and a single plant (kr4.1) was identified that produced Mendelian ratios of kanamycin-resistant progeny (Fig. 2D).

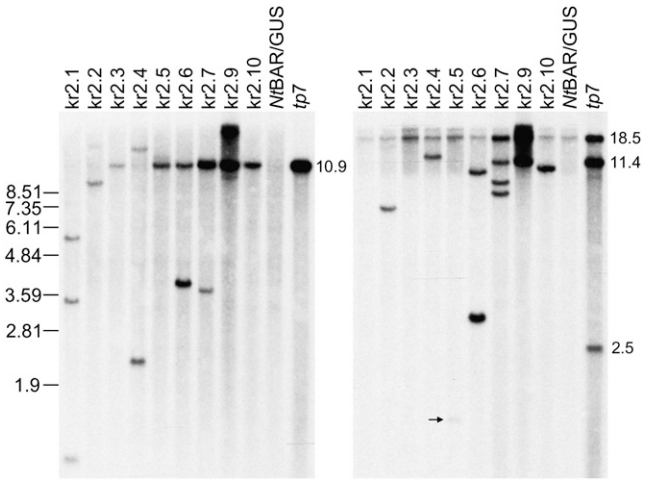

Figure 3. DNA-blot analysis of *Xba*I-digested DNA from nine kanamycin-resistant plants hybridized with *neo* (left) or *aadA* (right) probes. Each lane contained 10 μg of DNA except for *tp*7, which contained 50 ng of DNA. Size markers and approximate sizes for *tp*7 bands are shown in kb. The arrow indicates a faint band for kr2.5.

Monitoring Plastid-to-Nucleus DNA Transposition Using a *gus* Reporter Gene

To investigate the timing and frequency of ptDNA transposition in somatic cells, a second transplastomic line (*tp*GUS) analogous to *tp*7 but containing a *gus* reporter gene in place of *neoSTLS*2 was generated (Fig. 1B). The *tp*GUS line is homoplastomic (data not shown) for a single 35S*gusSTLS*2 and *aadA* cassette inserted into the plastid genome near *rbcL* (which encodes the large subunit of ribulose bisphosphate carboxylase-oxygenase; Fig. 1D). Due to the presence of a nuclear promoter (35S) and intron (*STLS*2), the *gus* gene was expected to be expressed only upon transfer to the nucleus.

Histochemical staining of *tp*GUS plants identified sectors of GUS-positive tissue in leaves, cotyledons, roots, vasculature, anther walls, and trichomes. Staining appeared as small, discrete foci of expression in these tissues surrounded by areas where staining was not detectable (Fig. 2, E–I). To verify that these sectors were the result of transfer of 35S*gusSTLS*2 to the nucleus rather than activation of the gene in the plastid genome, the cellular localization of the GUS protein was examined. After transfer of 35S*gusSTLS*2 from the plastid to the nucleus, the resulting GUS enzyme, which does not contain any organelle targeting signals, would be expected to accumulate in the cytosol where the *gus* mRNA is translated. Nonlocalized and uniform staining of cells is diagnostic of GUS located in the cytosol, and this was observed in all the blue sectors examined in *tp*GUS plants. Figure 4 shows

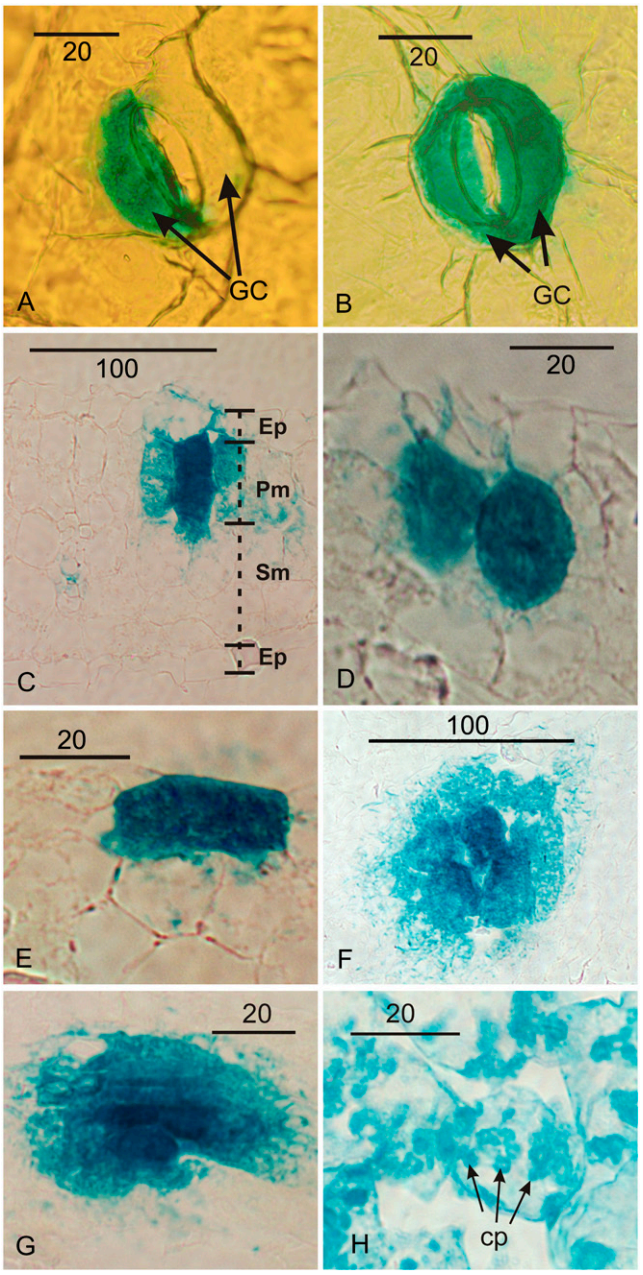

Figure 4. Microscopic analysis of GUS expression in *tp*GUS transplastomic plants. A and B, Single (A) and double (B) guard cells (labeled GC) expressing GUS. C, Transverse section showing a pallisade cell expressing GUS. Labels indicate epidermis (Ep), pallisade mesophyll (Pm), and spongy mesophyll (Sm). D, Two mesophyll cells expressing GUS. E, Epidermal cell expressing GUS. F and G, Multiple cells expressing GUS. H, Mesophyll cells in a previously isolated pUM79 transplastomic plant (see Kode et al., 2006) showing GUS localized to chloroplasts (cp). Scale bars = 20 or 100 microns.

examples of sectors composed of one (Fig. 4, A, C and E), two (Fig. 4, B and D), or more (Fig. 4, F and G) blue cells in *tp*GUS plants. Leakage of cytosolic contents results in patchy staining of adjacent cells and intercellular spaces (Fig. 4, C–G). These results are clearly distinguishable from expression in the chloroplasts of a control transplastomic line containing the *gus* gene driven by a plastid promoter (pUM79; Kode et al., 2006) in which GUS activity is confined to chloroplasts (Fig. 4H).

In a single 18-cm leaf, 228 GUS-stained sectors were observed. Using the same estimation of cell number (Hannam, 1968) as Stegemann et al. (2003), this implies transfer and expression of 35S*gusSTLS*2 in at least one in 200,000 somatic cells, if each sector is assumed to result from a single transfer event. This frequency is approximately 25 times higher than that measured by Stegemann et al. (2003) where the observed transposition events required stable integration and subsequent cell division and plant regeneration.

Due to the significant discrepancy between transfer frequencies, we sought to provide a more detailed analysis of the DNA transfer frequency. Seeds were germinated in vitro and GUS sectors scored when cotyledons and the first true leaves had reached a length of 3 to 4 mm (Fig. 5). Large variations in sector numbers were found in different cotyledons and leaves, which is reflected in large SDs. No sectors were found in the cotyledons and leaves of a transplastomic line lacking the *gus* gene (negative control). All four *tp*GUS transplastomic lines (*tp*GUS5.3, 5.6, 8.4, and 9.4), which were derived from independent transformation events, gave a similar range of sector numbers in cotyledons and leaves. This consistency between transplastomic lines supports plastid-to-nucleus transfer of 35S*gusSTLS*2 as an explanation for the origin of sectors. The possibility that the sectors arise from activation of a silenced *gus* gene, inadvertently introduced into the nucleus during transformation, is unlikely given the comparable sector frequencies in the four transplastomic lines. An average of between five and six sectors per cotyledon or leaf was found when the results from all four transplastomic *tp*GUS lines were combined (111 cotyledons and 89 leaves). The total number of cells per cotyledon or leaf was estimated to be 100,000 (see "Materials and Methods"), which corresponds to a DNA transfer frequency of approximately one event per 18,000 cells, if each sector is assumed to result from an independent transfer event. The correspondence in sector frequencies between cotyledons and leaves probably reflects the similar numbers of cells present in the organs when sectors were scored and a similar frequency of transfer.

GUS-expressing sectors were found in all types of leaf cells, and their appearance appeared to be random. The variations in sector sizes probably reflect the timing of plastid-to-nucleus DNA transfer during leaf development. Among 225 sectors examined, we found 121 single-cell sectors, 60 two-cell sectors, 23 three-cell sectors, and 12 four-cell sectors. The remaining nine sectors contained five to 10 cells. Guard cells that are derived from a common guard mother cell (Pillitteri et al., 2007) provide a clear example of DNA transfer in progenitor and terminally differentiated cells. DNA transfer in a terminally differentiated cell gives rise to a single GUS-positive guard cell (Fig. 4A),

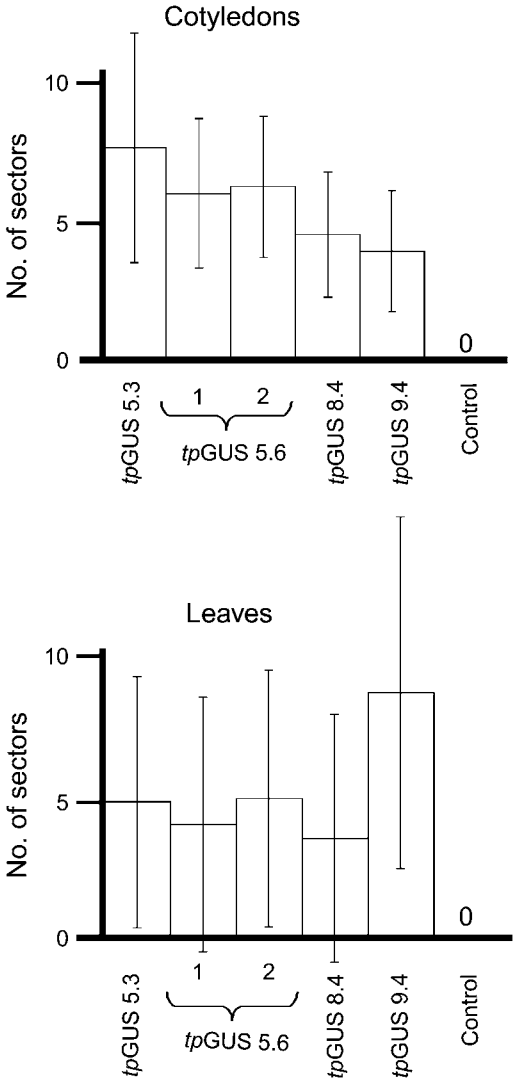

Figure 5. Analysis of sector frequency in cotyledons (3–4 mm) and first true leaves (3–4 mm) in *tp*GUS lines and a control transplastomic line lacking the *gus* gene. SDs are shown. $n = 24$ cotyledons for *tp*GUS lines and 16 cotyledons for the control line. $n = 15$ to 21 leaves for *tp*GUS lines and 12 leaves for the control line.

whereas two GUS-positive sister guard cells probably result from DNA transfer in their shared mother cell (Fig. 4B).

The Effect of Gene Copy Number and Location in the Transplastome on Transposition Frequency

Potentially, both copy number and location of a reporter gene within the plastid genome could affect the frequency of its transposition to the nucleus. In *tp*7, the 35S*neoSTLS*2 transgene was inserted into the inverted repeat region of the plastid genome between the *16S* rRNA and *rps*7/12 genes and is therefore present as two gene copies per plastid genome (Fig. 1C). In contrast, the 35S*gusSTLS*2 reporter gene is located adjacent to *rbcL* in the large single-copy region of the plastid genome (Fig. 1D). Therefore, these different transplastomic types may show different frequencies of plastid-to-nucleus DNA transfer because of the copy number difference of the reporter gene in the transplastome or different transposition frequencies existing among plastid sequences.

This hypothesis was tested by performing histochemical GUS staining to identify progeny with widespread GUS-positive staining from self-fertilized *tp*GUS plants. Two GUS-positive plants, gs1.1 and gs1.2, were identified in a nondestructive screen of 98,000 seedlings, giving a frequency of one recoverable transposition event in approximately 49,000 progeny. The histochemical staining phenotypes of these two lines were similar to nuclear 35S*gus*-positive control plants (Fig. 2, J–L). Spliced *gus* mRNA, lacking the *STLS*2 intron, was amplified from gs1.1 and gs1.2, verifying that these two lines were the result of nuclear transposition of 35S*gusSTLS*2 (Fig. 6). In contrast, the plastid-localized 35S*gusSTLS*2 gene remained unspliced. Furthermore, after backcrossing gs1.1 to female wild type, GUS staining was present in approximately one-half of the progeny (data not shown), consistent with the expected segregation pattern for a nuclear gene. Therefore, prima facie, the frequency of plastid-to-nucleus DNA transfer observed in *tp*GUS appears to be approximately 4-fold lower than that observed for 35S*neoSTLS*2 in *tp*7. This relatively small difference in transposition frequencies could be due to sampling error arising from the rarity of the events or could be due to kanamycin selection being more efficient than vital GUS staining in identifying seedlings resulting from transfer events in pollen. However, from these experiments, it may be concluded that the insertion of transgenes in these two very different transplastomic locations with two quite different reporter genes does not appear to have a large effect on transposition frequency.

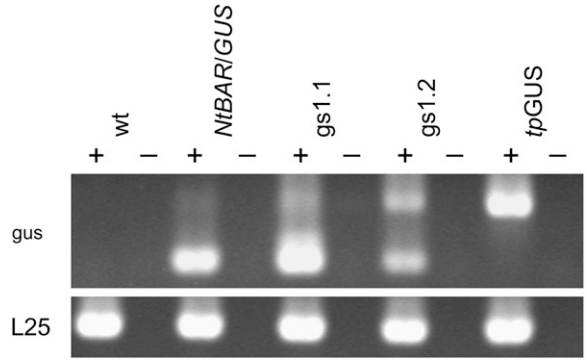

Figure 6. RT-PCR analysis. RT-PCR was performed using *gus* primers spanning the *STLS*2 intron. The higher band (450 bp) represents unspliced transcript, and the lower band (250 bp) represents spliced transcript. In the NtBAR/GUS control, the *gus* gene does not contain an intron. Control RT-PCRs with ribosomal protein L25 primers are also shown, giving a single 370-bp band. Lanes marked − indicate no RT.

DISCUSSION

The results from reciprocal crosses demonstrate a large difference between the male and female germlines in the frequency of DNA transposition from the plastid to the nuclear genome. Crosses where the male parent was transplastomic gave a transposition frequency of one stable event per 11,000 pollen grains, while crosses where the female parent was transplastomic showed one stable transposition in 273,000 ovules, but the phenotype of the plant recovered from this cross was atypical compared with all others isolated. Hence, the frequency of a newly transposed fragment of ptDNA in male gametes of tobacco is at least an order of magnitude higher than in those of the female.

The elevated frequency of transposition observed within the male germline may be associated with mechanisms that prevent paternal inheritance of the plastid genome, because degradation of the plastid genome in male gametes may result in DNA fragments that could enter and transform the nucleus. We attempted to detect this directly in microspores, pollen grains, and growing pollen tubes by two methods. First, we used quantitative real-time PCR to target spliced *neo* transcripts in *tp*7, and, second, we used histochemical and quantitative GUS assays in *tp*GUS (data not shown). Neither of these approaches were successful in detecting elevated ptDNA transfer in any particular cell type, which is consistent with other reports indicating that the 35S promoter has very poor expression in these tissues (Wilkinson et al., 1997; Custers et al., 1999). Therefore, with the transplastomic lines available, current methods appear to be inappropriate for the identification of those cells that are prone to integrate ptDNA in their nuclei and then transmit this insertion to subsequent generations. A pollen-specific promoter could be used to provide more efficient transcription than the 35S promoter in the male germline, but this would still be confounded by the possibility that transposition is elevated in a cell type that is not part of the expression repertoire of the chosen promoter.

Histochemical analysis of the transplastomic line *tp*GUS indicates that plastid nucleic acid enters the nucleus in a variety of somatic tissues at a high frequency. These data are the first demonstration of somatic transposition of a plastid transgene to the nucleus in the absence of antibiotic selection and plant regeneration. Analysis of a single 18-cm leaf indicated a transfer frequency of approximately one event per 200,000 cells. A more thorough analysis of 3- to 4-mm cotyledons and leaves indicated a transfer frequency of one event per 18,000 cells. There are several experimental differences that could explain this discrepancy in calculated transposition frequencies. First, the tissues analyzed were at different stages of development; second, the former experiment was performed with a soil-grown plant while the latter experiments used seedlings grown in vitro; and finally, different methods were used for the calculation of cell number. The frequency of transposition we observed in leaves is 25 to 300 times higher than a previous estimate based on regenerative selection (kanamycin) of somatic cells containing stable integrants of ptDNA in the nucleus (Stegemann et al., 2003). Potentially, the histochemical staining of *tp*GUS tissues identified both stable and transient expression, which could account for the difference. Therefore, it is likely that ptDNA enters the nucleus and is expressed far more frequently than it integrates into a chromosome. In addition, not all kanamycin-resistant cells would be expected to survive selection and be capable of regenerating into plants.

The frequency of transposition varied greatly between replicate plants and tissues, implying that it is not tightly regulated and is mainly the result of chance plastid degradation and nucleic acid escape. A rare stochastic process such as this might be expected to give rise to a predominance of small sectors due to the larger number of cells present at later stages of leaf development compared with those present in the leaf initials. Indeed, large sectors of GUS-stained tissue were not found in older leaves, reflecting the rarity of transfer events early in leaf development and also suggesting that any stable integration events that occur at earlier stages of leaf development very rarely involve cells that proceed to further divisions. Therefore, because the 35S*gusSTLS*2 and 35S*neoSTLS*2 systems are likely to be comparable, the regeneration-based selection procedures applied by Stegemann et al. (2003) must have induced some cells to propagate when their fate would have been nonproliferative in normal leaf development.

The reciprocal difference between the frequencies of plastid-to-nucleus DNA transfer in the male and female germlines may have arisen under the influence of the selective pressures that maintain uniparental organelle inheritance. It is clear that transfer of plastid genes to the nuclear genome (either to replace the original gene or to take on a new function) has been selected for, because this has been such a widespread phenomenon throughout eukaryotic evolution (Martin et al., 2002, 2003; Timmis et al., 2004). In fact, recent results indicate that smaller fragments of noncoding organellar DNA can also be incorporated as functional exons in the nucleus (Noutsos et al., 2007). Therefore, transposition of ptDNA to the nucleus must confer a selective advantage at the species level, even though most ptDNA transpositions are likely to be nonfunctional or detrimental. However, inheritance and replication of a functional chloroplast genome is essential for survival. Processes that allow transposition in the male germline, while suppressing transposition in the female germline, would maximize the benefits associated with transposition while maintaining the essential functions of the plastid genome, and hence would become a characteristic of successful species. There is ample evidence that both loss of plastids and degradation of ptDNA occur during pollen development, particularly in species that show uniparental inheri-

tance of plastid genes (for review, see Mogensen, 1996), thus providing a mechanism by which transposition may occur in the male germline. The less frequent circumstance of biparental cytoplasmic inheritance (Mogensen, 1996) appears to argue against this postulate, but it is clear that the mode of inheritance of organellar genes is a variable character that has oscillated during evolution (Birky, 2001). Therefore, uniparental inheritance can be viewed as a means by which transposition of organellar DNA to the nucleus takes place without compromising organelle function in the zygote and thus it has evolved as a consequence of these two selective pressures. If true, this is a new explanation for the widespread incidence of uniparental inheritance in eukaryotes.

Some plant biotechnologists have advocated the placement of transgenes in the plastid genome to ensure their containment in the maternal parent and prevent their escape through pollen dispersal. This study demonstrates that the frequency of plastid transgene relocation to the nucleus in the male germline is an order of magnitude higher than in the female germline. Hence, plastid transgenesis alone does not provide complete transgene containment in tobacco, and additional safeguards will be necessary to eliminate all possibility of transgene escape.

MATERIALS AND METHODS

Plant Growth Conditions

Tobacco (*Nicotiana tabacum*) plants grown in soil were kept in a controlled environment chamber with a 14-h-light/10-h-dark and 25°C-day/18°C-night growth regime.

Analysis of Kanamycin Resistance

Kanamycin selection was performed using 0.5× Murashige and Skoog salt medium (Murashige and Skoog, 1962) containing 150 or 300 μg mL^{-1} kanamycin for plants with wild-type plastids or 300 μg mL^{-1} kanamycin for plants with *tp7* plastids. Screens were performed by plating surface-sterilized seeds on 150-mm plates containing 80 mL of the above medium at a density of 2,000 seeds/plate. Progeny testing of kanamycin-resistant plants was performed by plating surface-sterilized seeds on 90-mm plates containing 20 mL of the above medium. All plates were incubated at 25°C with 16 h light/8 h dark.

Analysis of GUS Activity

For histochemical GUS assays, tissues were fixed by vacuum infiltration in 100 mM sodium phosphate buffer, pH 7.0, 0.12% formaldehyde, 0.1% β-mercaptoethanol, 0.1% Triton X-100 for 10 min, washed three times with 100 mM sodium phosphate buffer, pH 7.0, and stained in 45 mM sodium phosphate buffer, pH 7.0, 0.45 mM potassium ferricyanide, 0.45 mM potassium ferrocyanide, 0.1% Triton X-100, 0.05% chloramphenicol, 0.1% β-mercaptoethanol, 10% dimethyl sulfoxide, 0.1% X-Gluc overnight at 37°C. After staining, tissues were cleared in 70% ethanol.

Viable GUS staining was performed in tissue culture as described (Martin et al., 1992).

For the analysis of GUS sectors in seedlings, seeds were germinated on Murashige and Skoog medium as described (Kode et al., 2006). Whole seedlings were incubated in buffer containing X-Gluc at 37°C, then chlorophyll was removed with ethanol before mounting on slides to examine sectors using a Leica S8APO Stereo Zoom microscope. For sections, leaf pieces were fixed (0.3% formaldehyde, 30 min) and stained in X-Gluc buffer as described (Klösgen and Weil, 1991). Leaf pieces were embedded in wax (Thermoshandon Histocentre 3), sectioned into 5-μm slices (Microm HM330), cleared (Histo-Clear; National Diagnostics), and mounted on slides for microscopy (Leica DMR microscope). The average number of cells per leaf or cotyledon was estimated to be 100,000 cells based on counting the average number of cells in a 100-μm × 100-μm square (20 cells), the average cotyledon area (9.2 mm^2), or leaf area (9.6 mm^2) and the observation of six cell layers in the cotyledons and leaves examined (e.g. see Fig. 4C).

Southern Blotting

DNA blot analyses were carried out as described (Ayliffe and Timmis, 1992; Huang et al., 2003b).

Generation of *tp*GUS

A *gus* gene containing the second intron of the potato (*Solanum tuberosum*) *STLS-1* gene inserted into the open reading frame, with 35S promoter and terminator sequences, was amplified from p35S GUS INT (Vancanneyt et al., 1990) using the following primers (*Not*I and *Sac*II sites underlined): 5′-ATCGTA<u>GCGGCCGC</u>AACATGGTGGAGCACGACACTCTCGTCTAC-3′ and 5′-TGACTA<u>CCGCGGC</u>ATGCCTGCAGGTCACTGGATTTTGGTTTTAGG-3′.

The resulting PCR product was cloned into pGEM-T Easy (Promega). The *Sac*II/*Not*I fragment of this vector containing *gus* and the *Apa*I/*Sac*II fragment of pUM35 containing *aadA* flanked by the *Brassica napus 16S rrn* promoter and *psbC* terminator regions (Zubko et al., 2004) were then ligated into *Apa*I/*Not*I-digested pATB27-link (Zubko et al., 2004) to generate the transformation vector. Transplastomic plants were isolated by selecting for dark-green spectinomycin-resistant shoots following particle bombardment of pale-green Δ*rbc*L leaves as described (Kode et al., 2006).

Reverse Transcription-PCR Analysis

RNA extraction was performed using an RNeasy Plant Mini kit (Qiagen) and genomic DNA contamination removed using a TURBO DNA-*free* kit (Ambion). Reverse transcription (RT) was then performed using an Advantage RT-for-PCR kit (CLONTECH) with oligo(dT) primer. All kits were used in accordance with the manufacturers' instructions. PCR amplification was performed using *Taq* DNA Polymerase (New England Biolabs) according to standard protocols. Primers used were 5′-TCATTACGGCAAAGTGTGGGTC-3′ and 5′-GTAGAGCATTACGCTGCGATGG-3′ for *gus* PCRs and 5′-AAAATCT-GACCCCAAGGCAC-3′ and 5′-GCTTTCTTCGTCCCATCAGG-3′ for L25 PCRs.

ACKNOWLEDGMENTS

We thank Tracy Miller and Eun-Lee Jeong for technical assistance. We thank Anne Warhurst in the Histology Facility (Life Sciences, Manchester) for preparing tissue sections.

Received March 13, 2008; accepted July 20, 2008; published July 25, 2008.

LITERATURE CITED

Ayliffe MA, Timmis JN (1992) Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. Theor Appl Genet **85:** 229–238

Benfey PN, Chua NH (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science **250:** 959–966

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annu Rev Genet **35:** 125–148

Custers JBM, Snepvangers S, Jansen HJ, Zhang L, Campagne MMV (1999) The 35S-CaMV promoter is silent during early embryogenesis but activated during nonembryogenic sporophytic development in microspore culture. Protoplasma **208:** 257–264

Hannam RV (1968) Leaf growth and development in the young tobacco plant. Aust J Biol Sci **21:** 855–870

Huang CY, Ayliffe MA, Timmis JN (2003a) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature **422:** 72–76

Huang CY, Ayliffe MA, Timmis JN (2003b) Organelle evolution meets biotechnology. Nat Biotechnol **21:** 489–490

Klösgen RB, Weil JH (1991) Subcellular location and expression level of a chimeric protein consisting of the maize waxy transit peptide and the beta-glucuronidase of *Escherichia coli* in transgenic potato plants. Mol Gen Genet **225:** 297–304

Kode V, Mudd EA, Iamtham S, Day A (2006) Isolation of precise plastid deletion mutants by homology-based excision: a resource for site-directed mutagenesis, multi-gene changes and high-throughput plastid transformation. Plant J **46:** 901–909

Martin T, Schmidt R, Altmann T, Frommer WB (1992) Non-destructive assay systems for detection of β-glucuronidase activity in higher plants. Plant Mol Biol Rep **10:** 37–46

Martin W (2003) Gene transfer from organelles to the nucleus: frequent and in big chunks. Proc Natl Acad Sci USA **100:** 8612–8614

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA **99:** 12246–12251

Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. Am J Bot **83:** 383–404

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant **15:** 473–497

Nagata N, Saito C, Sakai A, Kuroiwa H, Kuroiwa T (1999) The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. Planta **209:** 53–65

Noutsos C, Kleine T, Armbruster U, DalCorso G, Leister D (2007) Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. Trends Genet **23:** 597–601

Pillitteri LJ, Sloan DB, Bogenschutz NL, Torii KU (2007) Termination of asymmetric cell division and differentiation of stomata. Nature **445:** 501–505

Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. Proc Natl Acad Sci USA **100:** 8828–8833

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet **5:** 123–135

Vancanneyt G, Schmidt R, O'Connor-Sanchez A, Willmitzer L, Rocha-Sosa M (1990) Construction of an intron-containing marker gene: splicing of the intron in transgenic plants and its use in monitoring early events in *Agrobacterium*-mediated plant transformation. Mol Gen Genet **220:** 245–250

Wilkinson JE, Twell D, Lindsey K (1997) Activities of CaMV 35S and *nos* promoters in pollen: Implications for field release of transgenic plants. J Exp Bot **48:** 265–275

Yu HS, Russell SD (1994) Populations of plastids and mitochondria during male reproductive cell maturation in *Nicotiana tabacum* L.: a cytological basis for occasional biparental inheritance. Planta **193:** 115–122

Zubko MK, Zubko EI, van Zuilen K, Meyer P, Day A (2004) Stable transformation of petunia plastids. Transgenic Res **13:** 523–530

Chapter 3: Is an RNA Intermediate Involved in Plastid-to-Nucleus Transfer?

Introduction

Sequence transfer from the plastid to the nucleus in tobacco has been detected experimentally using transplastomic lines (*tp*7 and *tp*17) which contain, in their plastid genomes, a kanamycin resistance gene (*neo*) under the control of a nuclear promoter. Kanamycin selection of backcrossed *tp*7/*tp*17 progeny revealed a transfer frequency in the male germline of approximately 1 event for every 11,000 – 16,000 pollen grains (Huang *et al.* 2003; Sheppard *et al.* 2008). Many of the kanamycin-resistant plants obtained from these experiments were found to have very large regions of plastid DNA integrated into the nuclear genome along with *neo*, with the size of integrants often exceeding 20 kb (Huang *et al.* 2004; Sheppard *et al.* 2008). Due to the large size of integrants, it seems likely that transfer occurred by a direct DNA mechanism. However, there is some evidence for organelle-to-nucleus transfer involving an RNA intermediate (Nugent & Palmer 1991; Grohmann *et al.* 1992; Adams *et al.* 2000). Northern analysis of *tp*7 and *tp*17 revealed abundant *neo*-hybridising transcripts of 3.3 kb and 2.2 kb (both larger than the *neo* gene) which were assumed to be the result of transcriptional read-through from nearby promoter(s) (Huang *et al.* 2003, 2004), so smaller integrants could have resulted from RNA-mediated transfer. It is also possible that longer transcripts are present at low levels that were not detected by Northern blotting, in which case the large integrants could also have resulted from RNA-mediated transfer.

To test the mode of transfer directly, experiments were designed to detect plastid-to-nucleus transfer only if it occurred via an RNA intermediate. These experiments made use of plastid RNA editing. In higher plants, some plastid RNAs undergo C to U editing at specific sites where this specificity is conferred by the flanking sequence (reviewed by Bock 2000). It has been shown that sequence encompassing an editing site can be fused to a transgene with editing occurring at high efficiency (Chaudhuri *et al.* 1995; Bock *et al.* 1996).

The experiments to detect RNA-mediated plastid-to-nucleus transfer were designed to be as similar as possible to previous experiments to detect plastid-to-nucleus transfer (Huang *et al.* 2003; Sheppard *et al.* 2008). A new transplastomic line was generated (*tp*neoACG) in which the *neo* gene had been modified to incorporate several plastid RNA editing sites, one of which was required for the functionality of *neo*, but in all other ways *tp*neoACG was identical to *tp*7/*tp*17. The experiments were designed in this way so that the results could be compared directly to

previous experiments where the frequency of total plastid-to-nucleus transfer (both DNA and RNA-mediated) was determined. In this way the current experiments would not only determine whether RNA-mediated transfer occurs at an appreciable frequency, but also if it does, what proportion of transfer events involve an RNA intermediate. Given the abundant *neo*-hybridising transcripts in *tp*7/*tp*17 that were revealed by Northern blotting (Huang *et al.* 2003, 2004), it was expected that if plastid-to-nucleus transfer involving an RNA-intermediate occurs at a reasonable frequency, then it should be detectable with the *tp*neoACG system.

In order to generate *tp*neoACG, a new transformation vector was generated (pPRV111A::neoSTLS2ACG) which was derived from the transformation vector used to generate *tp*7/*tp*17 (pPRV111A::neoSTLS2). In both vectors, *neo* is under the control of the 35S promoter and contains the nuclear STLS2 intron, such that kanamycin selection can be used to detect the transfer of *neo* to the nucleus (Figure 1). Both vectors also contain identical regions of homology to the plastid genome. When the vectors are used for generating transplastomic lines, these regions of homology mediate homologous recombination with the plastid genome and thus determine the site of integration of *neo*. Therefore *neo* will integrate into the same region of the plastid genome in each case. Adjacent to *neo* in both transformation cassettes is *aadA*, which is used for selection of transplastomic lines.

The crucial modifications in the new vector (pPRV111A::neoSTLS2ACG) compared with pPRV111A::neoSTLS2 were two modifications to the *neo* gene (Figure 1, B and C). Firstly, the start codon of *neo* was removed and replaced with a 101 bp region derived from tobacco *psbF* and *psbL* genes which contained a single plastid C to U RNA editing site. This editing site was part of an ACG codon in frame with the *neo* open reading frame, so that editing would create an AUG start codon. The 101 bp region containing the editing site had previously been inserted into the tobacco plastid genome as a fusion with a kanamycin resistance gene under the control of a plastid promoter and the editing site was shown to be efficiently edited (~70%) in this sequence context (Chaudhuri *et al.* 1995). In the absence of plastid RNA editing, *neo* mRNA lacks a start codon and is therefore not expected to be expressed. Editing of the RNA editing site creates a start codon. Therefore, if transfer of *neo* to the nucleus occurs via an edited RNA intermediate, then the nuclear integrant will have a functional version of *neo* and the plant should be kanamycin-resistant. On the other hand, if transfer occurs by a direct DNA mechanism, then the nuclear integrant will contain *neo* without a start codon and the plant should be kanamycin-sensitive. In this way, kanamycin selection can be used to detect only those events where transfer of *neo* to the nucleus occurs via an RNA intermediate.

It is possible that a kanamycin-resistant plant could be obtained by a C to T mutation of the editing site prior to, during or after transfer to the nucleus, rather than by plastid RNA editing. To

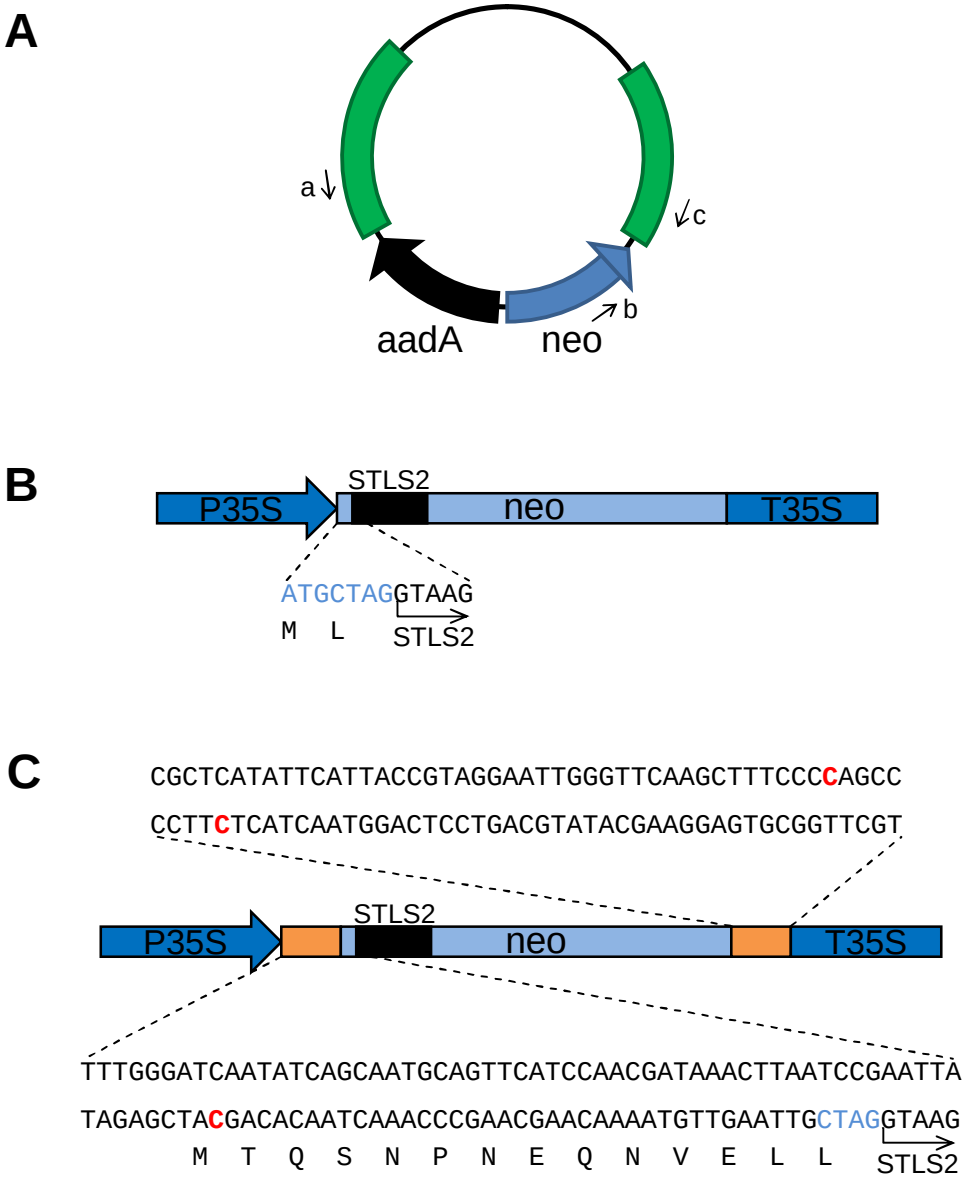

Figure 1. Experimental design. A. Structure of pPRV111A::neoSTLS2 and pPRV111A::neoSTLS2ACG. The *neo* gene differs between these two vectors (shown in B and C), but otherwise they are identical. Regions of homology to the plastid genome are shown in green. Small arrows indicate primers (a, b and c) that were used for homoplasmy testing (see Figure 2). **B.** Structure of *neo* in pPRV111A::neoSTLS2. *Neo* is under the control of 35S promoter and terminator sequences and contains the STLS2 intron. The sequence of the first 12 bases of *neo*, beginning from the start codon, is shown along with the protein translation. **C.** Structure of *neo* in pPRV111A::neoSTLS2ACG. *Neo* is under the control of 35S promoter and terminator sequences and contains the STLS2 intron. Orange regions are derived from plastid genes (see text for details) and contain three C to U editing sites (shown in red). *Neo* sequence is shown in blue. The protein translation is also shown (after C to U editing).

distinguish between these possibilities, a 94 bp region of *ndhB* containing two plastid RNA editing sites was introduced into the 3' UTR of *neo* (Figure 1C). This 94 bp region had previously been inserted into the tobacco plastid genome in the 3' UTR of an *aadA* gene under the control of a plastid promoter and the editing sites were efficiently edited (~95%) in this sequence context (Bock *et al.* 1996). If a kanamycin-resistant plant obtained from screening really was the result of RNA-mediated transfer then these two additional sites would be expected to be edited also. The presence of three edited sites at a novel nuclear locus would rule out point mutation as an origin and unequivocally implicate an RNA intermediate.

Results

Two independent transplastomic lines were generated using pPRV111A::neoSTLS2ACG, named *tp*neoACG5-3 and *tp*neoACG8-2. While homoplasmy is not essential for the experiments involving *tp*neoACG, it is desirable as it ensures stability of the transplastomic lines. Furthermore, if the results obtained from screening *tp*neoACG for RNA-mediated transfer are to be compared to the results obtained from *tp*7/*tp*17, then the transplastomic lines should be identical in this respect. Therefore, PCR analysis was used to test *tp*neoACG5-3 and *tp*neoACG8-2 for homoplasmy (Figure 2). *Tp*7 was used as a control since it had previously been shown to be homoplastomic by Southern analysis (Huang *et al.* 2003). PCR was performed using primers flanking the site of integration in the plastid genome so that a product would be amplified if there were any wildtype plastid genomes remaining (Figures 1A and 2A). As shown in Figure 2A, a strong band of 600 bp was obtained when using wildtype DNA as template. All three transplastomic lines gave very faint bands of the same size (barely visible in Figure 2A), which could have arisen from a small number of remaining native plastomes, or alternatively they could be amplification products of nupts. These lines also gave products corresponding to the expected sizes for the transformed plastid genomes (3510 bp for *tp*7 and 3704 bp for *tp*neoACG). Primers designed to amplify only from the transplastomes gave products of the expected sizes (770 bp for *tp*7 and 874 bp for *tp*neoACG), confirming that integration occurred correctly (Figures 1A and 2B). Figure 2C shows control PCRs using primers directed against a different region of the plastid genome, so amplification of a 731 bp product was expected in all cases. From these results it was concluded that very few, if any, wildtype plastid genomes remained in *tp*neoACG5-3 and *tp*neoACG8-2.

To detect RNA-mediated transfer, *tp*neoACG was backcrossed to female wildtype and 328,000 progeny screened for kanamycin resistance (Table 1). In the initial stages of screening there were several plants that appeared to be resistant but were still somewhat sickly. Therefore any plants that appeared resistant were transferred to non-selective medium in an attempt to recover as

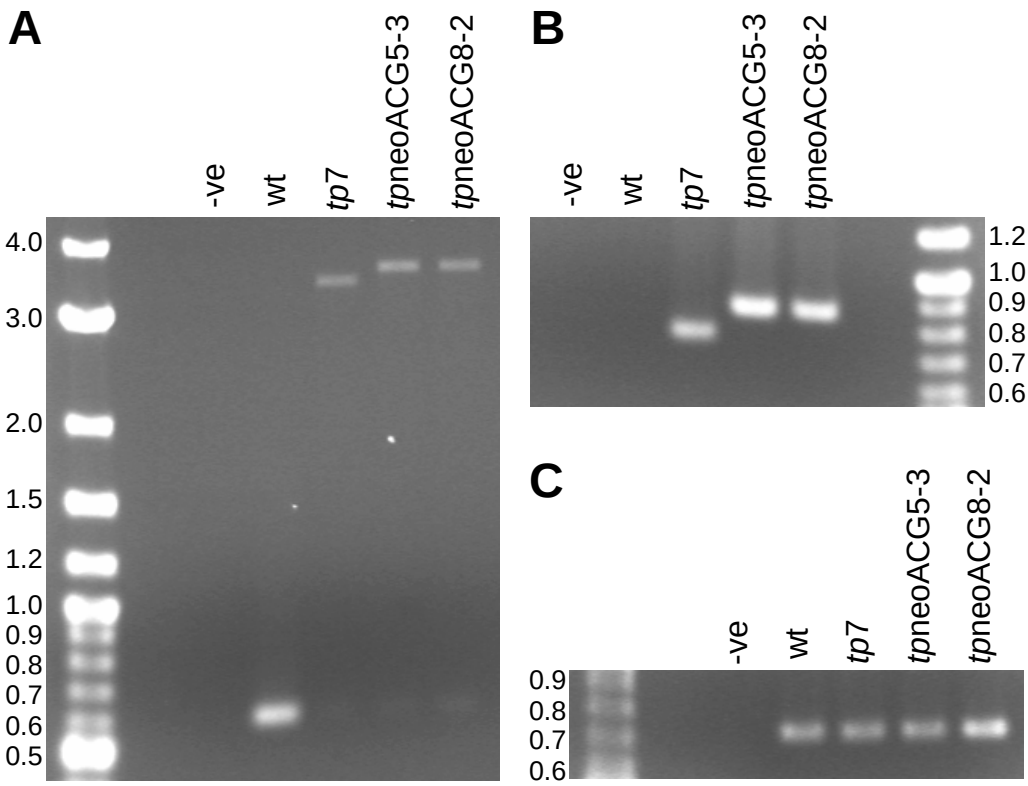

Figure 2. Homoplasmy testing. A. PCRs using primers cp140111F and cp140710R (indicated by 'a' and 'c' in Figure 1A). Expected band sizes are: 600 bp for wt, 3510 bp for *tp*7 and 3704 bp for *tp*neoACG. **B.** PCRs using primers neoF3 and cp140710R (indicated by 'b' and 'c' in Figure 1A). Expected band sizes are: 770 bp for *tp*7 and 874 bp for *tp*neoACG. **C.** PCRs using primers cp145418F(CH) and cp146142R. The expected product size is 731 bp. The sizes of molecular weight markers are given in kilobases. -ve no template.

many of these conceivably resistant plants as possible. Of the 328,000 progeny screened, 10 putative resistant plants were obtained. These were named kr5.1-kr5.10. After removal from selective medium, only 6 plants survived (kr5.1, kr5.2, kr5.3, kr5.4, kr5.6 and kr5.7). Presumably the other 4 plants died as a result of kanamycin exposure, so they may have been false positives or they may have been only partially kanamycin-resistant (see Discussion).

	Screened	**Putative Positives**	**Confirmed Positives**
*tp*neoACG5-3	188,000	5	1 (kr5.6)
*tp*neoACG8-2	140,000	5	5 (kr5.1-kr.5.4, kr5.7)
Total	328,000	10	6

Table 1. Results of screening wt ♀ x *tp*neoACG ♂ progeny for kanamycin resistance.

Kr5.1-kr5.7 were grown to maturity and DNA was prepared from leaf tissue, followed by PCR and sequencing to determine the sequences at the editing sites. For each plant, all three editing sites were C's (Figure 3), indicating that transfer had not involved an edited RNA intermediate.

This result indicated that the *neo* gene was somehow active in the absence of the intended start codon. One possibility for this could be translation initiation from a downstream AUG, but the next in-frame AUG is almost halfway through the coding sequence so it is unlikely that a functional protein would be produced. Alternatively, translation may be initiated from a non-AUG codon (Kobayashi *et al.* 2002; Depeiges *et al.* 2006; Wamboldt *et al.* 2009). Since the plants obtained in this experiment were less vigorous on kanamycin medium than those from previous screens involving the unmodified *neo* gene (e.g. Sheppard *et al.* 2008), it is likely that translation occurs only at low efficiency.

Sequencing of *neo* from kr5.1-kr5.7 (Figure 3, see above) indicated that *neo* transfer to the nucleus did not occur via an edited RNA intermediate. Therefore transfer occurred either by a direct DNA mechanism or via an unedited RNA intermediate. To attempt to distinguish between these possibilities, the efficiency of RNA editing was analysed in *tp*neoACG. If the efficiency of editing was close to 100%, this would effectively rule out transfer via unedited RNA as a possibility, thus showing that transfer occurred by a direct DNA mechanism.

To determine what proportion of *neo* RNA molecules were edited in *tp*neoACG, cDNA was synthesised using random hexamer primers followed by PCR amplification of the *neo* cDNA

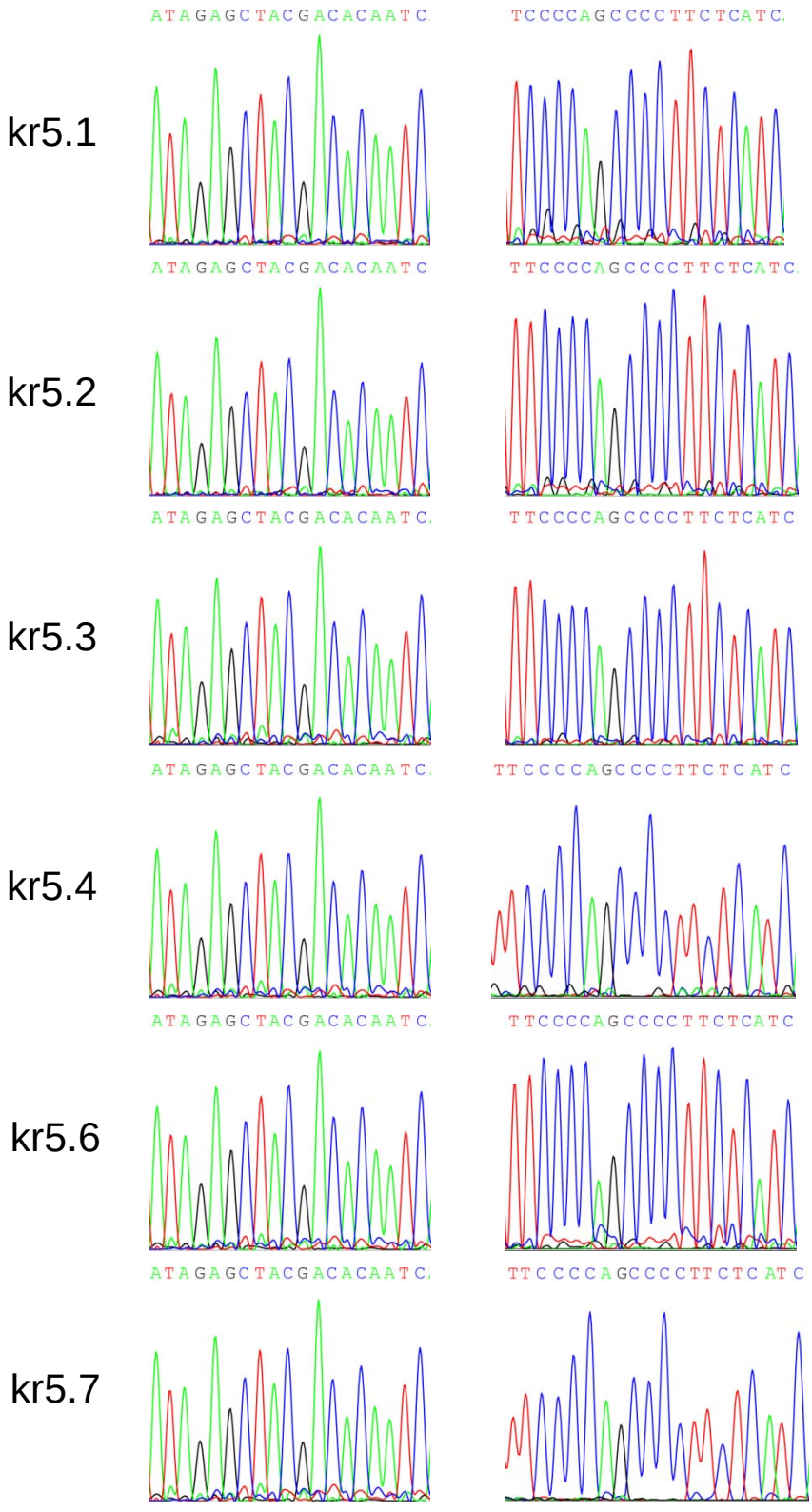

Figure 3. Sequencing of editing sites in kanamycin-resistant plants. The editing sites are boxed. For kr5.4 and kr5.7, sequencing of the two *ndhB* editing sites was performed using a reverse primer, so the sequences shown have been reverse-complemented.

including all three editing sites. This PCR product was then sequenced, revealing no clear evidence for editing at any of the editing sites (Figure 4A). This indicated that it was not possible to distinguish between DNA and RNA-mediated transfer, since the vast majority of transcripts were unedited.

The lack of detectable edited transcript in *tp*neoACG could be due to a low editing efficiency. However, this is not expected since the editing sites each have adequate flanking sequence which has previously been shown to be sufficient for efficient editing to occur (Chaudhuri *et al.* 1995; Bock *et al.* 1996). The other possibility is that a large proportion of antisense transcript is present, which would not be edited. Northern analysis of *tp*7/*tp*17 (Huang *et al.* 2003, 2004) used a double-stranded DNA probe, so it is not known whether the *neo*-hybridising bands represented sense or antisense transcript. To attempt to distinguish between the two possibilities for the lack of edited transcript (low editing efficiency or high proportion of antisense transcript), cDNA was synthesised using a *neo* primer complementary to the sense strand, so that theoretically only sense transcript would be represented in the cDNA population. If this revealed a high proportion of edited transcript (as opposed to the results obtained with cDNA synthesised using random hexamer primers which do not distinguish between the two strands) then it could be concluded that highly abundant antisense transcripts were the main reason for the low representation of edited RNA species.

For technical reasons, the two editing regions were amplified separately from cDNA samples synthesised using different *neo* primers. In contrast to the results obtained using random hexamer cDNA, some editing was observed at all three editing sites, but the majority of transcript remained unedited (Figure 5). The proportion of edited transcript was slightly higher in *tp*neoACG8-2 compared to *tp*neoACG5-3, which may or may not have a biological basis. To confirm that the presence of detectable edited transcript was the result of using a different primer for reverse transcription, rather than different PCR primers, the experiment using random hexamer cDNA was repeated with the same PCR primers as those used for the *neo*-specific cDNA experiment. Again this revealed no detectable editing (Figure 4B).

When a *neo* primer was used for cDNA synthesis, *neo* sense transcript should have been reverse-transcribed, but *neo* antisense transcript should not. However, since reverse transcription is performed at a relatively low temperature (42°C), it is possible that the primer was able to bind fortuitously to a few bases of homology on the antisense transcript. If this is the case, then cDNA synthesis using a *neo* primer will enrich for sense transcript (compared to using random hexamer primers), but not eliminate antisense transcript.

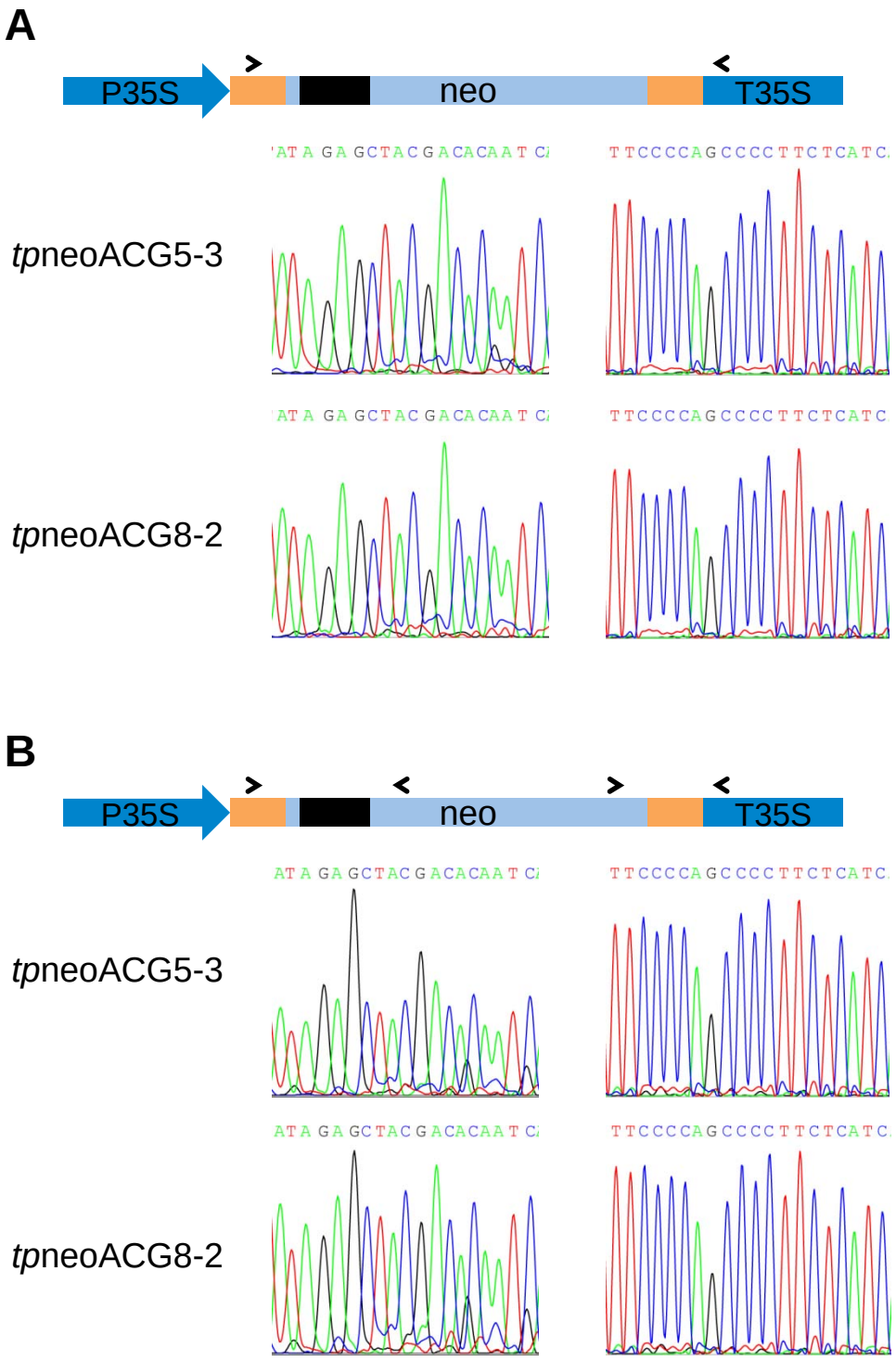

Figure 4. Sequencing of editing sites from *tp*neoACG cDNA synthesised with random hexamer primers. Arrows indicate primers that were used for PCR prior to sequencing. In **A**, PCR was performed with primers spanning the entire *neo* gene and in **B**, PCRs were performed separately for the two regions containing editing sites. Primers indicated are (left to right) psbF+Xho1 and neoR2(AL) in **A** and psbF+Xho1, TneoR1, NPTF2 and neoR2(AL) in **B**. See Table 2 for primer sequences.

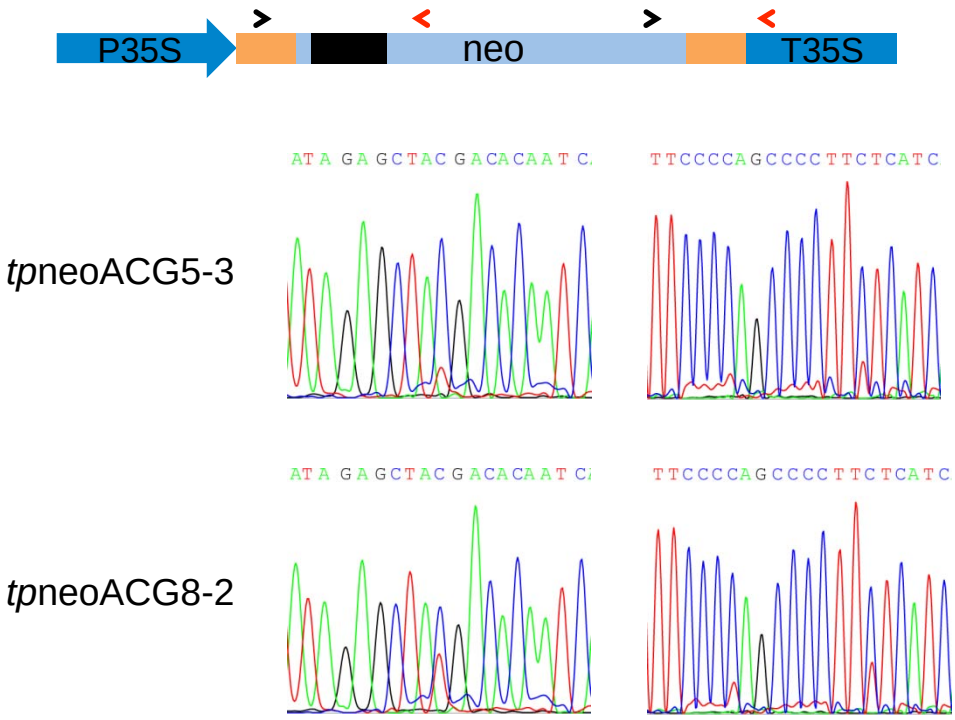

Figure 5. Sequencing of editing sites from *tp*neoACG cDNA synthesised with *neo* primers. Red arrows indicate primers that were used for reverse transcription. PCR primers were identical to those used in figure 4B and are shown here in red and black.

When cDNA was synthesised using a *neo* primer, the majority of transcript was unedited, so it is possible that the editing efficiency is relatively low. Alternatively, the editing efficiency may be high, with this result simply being a consequence of antisense transcript still being highly represented in the cDNA population synthesised using a *neo* primer. Therefore it is not possible to make a conclusion about the editing efficiency on the basis of this result. However, it can be concluded that antisense *neo* transcript is present in *tp*neoACG, since some edited transcript could be detected after enriching for sense transcript, but it could not be detected when using random hexamer primers for reverse transcription. Therefore the presence of antisense *neo* transcript contributed to the overall lack of edited transcript, but it may not have been the sole cause.

Discussion

In an attempt to distinguish between two possible modes of plastid-to-nucleus transfer (direct DNA transfer and RNA-mediated transfer), transplastomic lines were generated which contained, in their plastid genomes, a *neo* gene designed for nuclear expression but with several plastid RNA editing sites. One of these editing sites was designed so that editing would be required to create a translational start codon for *neo*. Surprisingly, *neo* was found to be active without this site being edited. This meant that the screen was able to detect not only transfer events involving an edited RNA intermediate, but also direct DNA transfer events and events involving an unedited RNA intermediate. Therefore sequencing was required to determine whether an edited RNA intermediate was involved for each kanamycin-resistant plant obtained from screening. A total of 328,000 progeny of *tp*neoACG were screened, revealing 6 kanamycin-resistant plants. Sequencing indicated that none of these arose from transfer involving an edited RNA intermediate. However, sequencing of *tp*neoACG cDNA indicated that only a very low proportion of *neo* transcripts were edited, thus making it impossible to distinguish between direct DNA transfer and RNA-mediated transfer.

Previous screens that were not designed to distinguish between direct DNA transfer and RNA-mediated transfer measured a transfer frequency of 1 event for every 11,000 – 16,000 male gametes (Huang *et al.* 2003; Sheppard *et al.* 2008). However, the current screen revealed only 6 resistant plants out of 328,000 tested, corresponding to a frequency of 1 in 55,000. A possible reason for the lower proportion of resistant plants in this screen is that even though the *neo* gene is able to confer some level of resistance in the absence of editing, the level of resistance may not always be enough to overcome the selection regime used. In this case the 6 resistant plants obtained may have had multiple nuclear insertions of *neo*, or *neo* may have inserted into a region

of the genome that allowed higher expression than other integrants. In this way the 4 putative kanamycin-resistant plants that died could have had a nuclear insertion of *neo* that made them less sensitive to kanamycin than wildtype, but not sufficiently resistant to survive. Presumably some other plants had nuclear insertions of *neo*, but were not distinguishable at all.

In these experiments, it was not possible to distinguish between direct DNA transfer and RNA-mediated transfer because only a very low proportion of *neo* transcripts were edited in *tp*neoACG. This was due at least in part to the presence of a large amount of antisense *neo* transcript, which was presumably the result of transcriptional read-through from adjacent plastid promoter(s). Editing efficiency may also have contributed, but it is not clear why this would have been an issue since the sequences used here had been used successfully previously, with around 70% editing of the *psbL* editing site and around 95% editing of the *ndhB* editing sites (Chaudhuri *et al.* 1995; Bock *et al.* 1996). It should be noted that *neo* transcript was analysed in leaf tissue of *tp*neoACG, whereas screening for *neo* transfer to the nucleus was performed with progeny of *tp*neoACG backcrossed to female wildtype. Therefore transfer presumably occurred sometime during pollen development, but experiments have so far been unable to determine the specific stage when this occurs (Sheppard *et al.* 2008). It is possible that the proportion of edited *neo* transcript differs between leaf tissue and the tissue where transfer occurs in the male germline.

The overall aim of these experiments was to determine whether plastid-to-nucleus sequence transfer occurs by a direct DNA mechanism or whether an RNA intermediate is involved. A high proportion of *neo* antisense transcript was found in *tp*neoACG, thus making it impossible to distinguish between the two modes of transfer using this transplastomic line. One possibility for improving the experimental design could be to place *neo* (with the 35S promoter) downstream of a strong plastid promoter. This would ensure high levels of sense *neo* transcript, so that even if some antisense transcript was produced it would only be a small proportion of the total *neo* transcript. Having a large amount of *neo* transcript would also presumably maximise the chances of RNA-mediated transfer occurring. Another possibility for minimising antisense transcript that could be used in conjunction with this would be to insert *neo* into a different region of the plastid genome (or insert it in the opposite orientation to that used here) in order to reduce any transcriptional read-through from adjacent plastid promoters that generates antisense transcript. One problem with using a plastid promoter to increase the levels of *neo* sense transcript is that the promoter should be active in the tissue where plastid-to-nucleus transfer occurs. As described above, the precise timing of transfer is unknown, so the correct choice of an appropriate promoter is currently not possible. An alternative approach that would eliminate this problem would be to screen for transfer in somatic tissues, although it is likely that this would be less relevant than germline transfer in evolutionary terms.

Materials and Methods

Generation of pPRV111A::neoSTLS2ACG

Firstly, pPRV111A::neoSTLS2 was modified in two ways: the G in the *neo* start codon was changed to a C to introduce an AvrII restriction site and 4 bases (ATCT) were introduced downstream of the *neo* stop codon to introduce a BglII site. In order to achieve this, *neo* was amplified from pPRV111A::neoSTLS2 using the primers neoF+Avr2 and neoR+Bgl2 which contained the intended modifications. An A-tailing reaction was performed with this PCR product which was then cloned into pGemT Easy (Promega) according to the manufacturer's instructions to generate pGemTEasy+neo. The 1 kb fragment resulting from fully digesting pGemTEasy+neo with XhoI followed by partial digestion with PpuMI was purified and cloned into pPRV111A::neoSTLS2 which had been digested with XhoI and PpuMI to generate pPRV111A::neoSTLS2+RS.

A 94 bp region of *ndhB* was amplified from tobacco genomic DNA using the primers ndhBF+Bgl2 and ndhBR+Bgl2, digested with BglII and cloned into pPRV111A::neoSTLS2+RS which had been digested with BglII to generate pPRV111A::neoSTLS2+ndhB.

A 101 bp region of *psbF* and *psbL* was amplified from tobacco genomic DNA using the primers psbF+Xho1 and psbR+Nhe1, digested with XhoI and NheI and cloned into pPRV111A::neoSTLS2+ndhB which had been digested with XhoI and AvrII to generate pPRV111A::neoSTLS2ACG.

Plastid transformation

*Tp*neoACG was generated by particle bombardment of tobacco leaf cells using pPRV111A::neoSTLS2ACG, which was performed in the laboratory of Dr Anil Day of The University of Manchester, UK.

Plant growth conditions and kanamycin selection

Tobacco plants grown in soil were kept in a controlled environment chamber with a 14 hr light/10 hr dark and 25°C day/18°C night growth regime.

Screening of wt ♀ x *tp*neoACG ♂ progeny was performed by plating surface-sterilised seeds on 150 mm plates containing 80 ml of 0.5x MS salt medium (Murashige & Skoog 1962) containing 150 μg ml^{-1} kanamycin at a density of approximately 2,000 seeds per plate. After around 2 to 4 weeks any plants that appeared resistant were transferred to 0.5x MS salt medium without kanamycin. Surviving plants were later transferred to soil.

DNA preparation

DNA extraction from tobacco leaf tissue was performed using a DNeasy Plant Mini kit (Qiagen) according to the manufacturer's instructions.

RNA preparation and cDNA synthesis

RNA extraction from tobacco leaf tissue was performed using an RNeasy Plant Mini kit (Qiagen) and genomic DNA contamination removed using a TURBO DNA-*free* kit (Ambion). Reverse transcription was then performed using an Advantage RT-for-PCR kit (Clontech) with 20 pmol primer (random hexamer or *neo*-specific) per reaction. All kits were used in accordance with the manufacturers' instructions. The *neo* primer used for analysis of the *psbL* editing site was TneoR1 and the *neo* primer used for analysis of the *ndhB* editing sites was neoR2(AL).

PCR and Sequencing

PCR of pPRV111A::neoSTLS2 using the primers neoF+Avr2 and neoR+Bgl2 was performed with *PfuTurbo* DNA Polymerase (Stratagene) according to the manufacturer's instructions. All other PCRs were performed using *Taq* DNA Polymerase (New England Biolabs or Roche) according to the manufacturers' instructions. Gel electrophoresis of PCR products was performed using 1% agarose gels in 1x TAE buffer and 2-Log DNA Ladder (New England Biolabs) was used for size comparison.

Prior to sequencing, PCR products were purified using either a PCR Purification Kit (Qiagen) or a Gel Extraction Kit (Qiagen) according to the manufacturer's instructions.

Sequencing was performed using BigDye Terminator v3.1 (Applied Biosystems). Each 20 µL reaction contained 1 µL Ready Reaction Mix, 3 µL Sequencing Buffer, 5 pmol primer and template DNA. Thermal cycling was performed with an initial denaturation step at 96°C for 2 min followed by 26 cycles of 96°C for 30 sec, 50°C for 15 sec and 60°C for 4 min. Extension products were purified by isopropanol precipitation. Each 20 µL reaction was mixed with 80 µL 75% isopropanol, incubated for 15 min and centrifuged at 16,000 x g for 20 min. The supernatant was removed and a further 250 µL 75% isopropanol added. Centrifugation was then performed at 16,000 x g for 5 min, the supernatant removed and the pellet dried. Purified extension products were analysed using a 3730 DNA Analyzer (Applied Biosystems) by the Institute of Medical and Veterinary Science (Adelaide, South Australia).

For kr5.1-kr5.7, the *psbL* editing site was amplified using the primers A35SF1 and npt2allR and sequencing performed with the A35SF1 primer. The *ndhB* editing sites were amplified using either neoF3 or NPTF2 with neoR2(AL) and sequencing was performed with either neoF3 or neoR2(AL).

For *tp*neoACG, the primers used to amplify the entire *neo* gene were psbF+Xho1 and neoR2(AL). The primers used to amplify the *psbL* editing site on its own were psbF+Xho1 and TneoR1 and the primers used to amplify the *ndhB* editing sites on their own were NPTF2 and neoR2(AL). In all cases psbF+Xho1 was used for sequencing the *psbL* editing site and neoF3 was used for sequencing the *ndhB* editing sites.

Primer sequences are given in Table 2.

Primer Name	**Sequence (5' – 3')**
neoF+Avr2	TGGACCTCGAGTGGCCACCATCCTAGGTAAGTTTCTGCTTCTAC
neoR+Bgl2	TTTGCGGGTCCCGAGATCTCAGAAGAACTCGTCAAGAAGG
psbF+Xho1	GTGTGCCTCGAGTTTGGGATCAATATCAGCAATGCAGGTC
psbR+Nhe1	CTGTCGGCTAGCAATTCAACATTTTGTTCGTTCGGGTTTG
ndhBF+Bgl2	ACCTTCAGATCTCGCTCATATTCATTACCGTAGGAATTGGG
ndhBR+Bgl2	CACAACAGATCTACGAACCGCACTCCTTCGTATACGTCAG
cp140111F	TCGAACTGATGACTTCCACCAC
cp140710R	TACCATGAGAGAAGCAAGGAGG
neoF3	GCGTTGGCTACCCGTGATATTGCTG
cp145418F(CH)	GACTGGACGAAACCAAGAAA
cp146142R	GTCTACCATTTCACCACCAAGG
A35SF1	AACATGGTGGAGCACGACACTC
npt2allR	CCAATAGCAGCCAGTCCCTTC
neoR2(AL)	GATAGATTTGTAGAGAGAGACTGGTG
NPTF2	GCTATCAGGACATAGCGTTG
TneoR1	GGACAGGTCGGTCTTGACAAAAAGAACC

Table 2. Primer Sequences. Restriction sites are underlined.

Chapter 4

Instability of Plastid DNA in the Nuclear Genome

Anna E. Sheppard and Jeremy N. Timmis

School of Molecular and Biomedical Science, University of Adelaide, South Australia 5005, Australia

PLoS Genetics, January 2009, Vol. 5, e1000323

STATEMENT OF AUTHORSHIP

Instability of Plastid DNA in the Nuclear Genome

PLoS Genetics, January 2009, Vol. 5, e1000323

Anna Sheppard (Candidate)

Performed experiments, analysed the data, wrote the manuscript and acted as corresponding author.

I hereby certify that the statement of contribution is accurate

Signed .....*Date*.....

Jeremy Timmis

Supervised development of work and assisted in writing the manuscript.

I hereby certify that the statement of contribution is accurate and I give permission for inclusion of the paper in the thesis

Signed .....*Date*.....

OPEN ACCESS Freely available online

PLoS GENETICS

Instability of Plastid DNA in the Nuclear Genome

Anna E. Sheppard*, Jeremy N. Timmis

School of Molecular and Biomedical Science, The University of Adelaide, South Australia, Australia

Abstract

Functional gene transfer from the plastid (chloroplast) and mitochondrial genomes to the nucleus has been an important driving force in eukaryotic evolution. Non-functional DNA transfer is far more frequent, and the frequency of such transfers from the plastid to the nucleus has been determined experimentally in tobacco using transplastomic lines containing, in their plastid genome, a kanamycin resistance gene (*neo*) readymade for nuclear expression. Contrary to expectations, non-Mendelian segregation of the kanamycin resistance phenotype is seen in progeny of some lines in which *neo* has been transferred to the nuclear genome. Here, we provide a detailed analysis of the instability of kanamycin resistance in nine of these lines, and we show that it is due to deletion of *neo*. Four lines showed instability with variation between progeny derived from different areas of the same plant, suggesting a loss of *neo* during somatic cell division. One line showed a consistent reduction in the proportion of kanamycin-resistant progeny, suggesting a loss of *neo* during meiosis, and the remaining four lines were relatively stable. To avoid genomic enlargement, the high frequency of plastid DNA integration into the nuclear genome necessitates a counterbalancing removal process. This is the first demonstration of such loss involving a high proportion of recent nuclear integrants. We propose that insertion, deletion, and rearrangement of plastid sequences in the nuclear genome are important evolutionary processes in the generation of novel nuclear genes. This work is also relevant in the context of transgenic plant research and crop production, because similar processes to those described here may be involved in the loss of plant transgenes.

Citation: Sheppard AE, Timmis JN (2009) Instability of Plastid DNA in the Nuclear Genome. PLoS Genet 5(1): e1000323. doi:10.1371/journal.pgen.1000323

Editor: Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, United States of America

Received July 24, 2008; **Accepted** December 2, 2008; **Published** January 2, 2009

Funding: This research was supported under the Australian Research Council's Discovery Projects funding scheme (project numbers DP0557496 and DP0667006).

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: anna.sheppard@adelaide.edu.au

Introduction

In eukaryotes, plastids and mitochondria are derived from once free living cyanobacteria and α-proteobacteria respectively [1,2]. Over evolutionary time, many of their genes have been relocated to the nuclear genome and in many cases this is an ongoing process [3–5]. Such functional gene transfer is not a trivial process and is dependent on several steps. The DNA sequence encoding the gene must not only integrate into the nuclear genome, but also it must acquire appropriate regulatory sequences for expression in the nucleus. Although an organellar sequence may occasionally integrate directly into a fortuitous location in the nuclear genome and become immediately functional, it is likely that most functional gene transfer events involve postinsertional rearrangements that bring the organellar gene into the context of a nuclear promoter [6]. In many cases these transfers involve gene products that retain their original function and are targeted back to the appropriate organelle and such genes must also acquire a transit peptide-encoding sequence. However, the original organellar function is not always maintained. For example, in Arabidopsis it has been estimated that approximately 18% (4,500) of nuclear genes are plastid-derived, and a large proportion of their products are not targeted to the plastid [7]. In algae this is also the case, although a lower proportion of ancestral cyanobacterial genes appear to have assumed non-plastid functions [8]. Therefore, organellar genomes have been a significant source of new genes in eukaryotic evolution.

While functional gene transfers from the plastid to the nuclear genome are relatively rare, non-functional sequence transfer occurs much more frequently and many nuclear genomes are riddled with such sequences, designated *nupts* (nuclear integrants of plastid DNA) [9]. The frequency of *nupt* formation has been measured experimentally in *Nicotiana tabacum* using transplastomic lines containing in their plastid genome a kanamycin resistance gene (*neo*) under the control of nuclear regulatory sequences, so that kanamycin selection can be used to detect transfer of *neo* to the nuclear genome. From these experiments it has been estimated that the frequency of transfer in the male germline is approximately 1 event per 11,000 to 16,000 pollen grains [10,11], while the frequencies of transfer in the female germline and in somatic cells appear to be much lower [11,12]. A number of the kanamycin resistant (kr) lines derived from the former experiments have been partially characterised at the molecular level and their causative experimental *nupts* are characteristically tens of kilobases in size [13]. The high frequency of plastid DNA (ptDNA) integration into the nuclear genome, together with the typically large size of the integrants, suggests the occurrence of counterbalancing removal events that would prevent a progressive increase in nuclear genome size. In fact, genome-wide analyses have revealed that decay of plastid sequences in the nuclear genome occurs relatively quickly in evolutionary terms [9]. With the experimental kr lines now available we have new tools with which to analyse any loss or decay that may occur within one or a few generations. Some of these kr lines were previously found to be unstable with respect to the kanamycin resistance phenotype in that there was a deficiency of kanamycin resistant progeny compared with Mendelian expectations [10]. Here we provide a detailed analysis of this instability in nine new kr lines [11] and we show that it is due to deletion of *neo*.

Author Summary

In eukaryotes, mitochondria and plastids are the descendents of once free-living prokaryotic ancestors. Over time, these organelles have donated a great deal of genetic material to the nuclear genome. Although usually non-functional, these DNA transfer events have, over evolutionary time, resulted in a large pool of functional nuclear genes and therefore the process of DNA transfer has been an important driving force in eukaryotic evolution. Previous studies showed that DNA transfer of a specific marker gene (*neo*) from the plastid to the nucleus occurred in one in every 11,000 to 16,000 male gametes. Because of this high frequency of transfer and the large size of integrants, this process would be expected to result in a cumulative increase in genome size, unless there are counterbalancing deletion events. In this study, we analysed the stability of the *neo* gene after integration into the nuclear genome. We found that the gene is highly unstable, with deletion often occurring within a single generation. These results indicate that plastid DNA insertion into and removal from the nuclear genome are in dynamic equilibrium, thus providing a mechanism by which the chances of functional DNA insertion are maximised without compromising the nuclear genome as a whole.

Results

Phenotypic Analysis of Instability

To investigate the genetic stability of kanamycin resistance seen in newly transposed *nupts* containing the *neo* gene, we analysed nine kr plants (kr2.1–2.7, kr2.9 and kr2.10), each of which resulted from an independent transposition of ptDNA to the nucleus [11]. Preliminary work indicated that different seed capsules from the same plant sometimes gave rise to different ratios of kanamycin resistant to kanamycin sensitive (kr:ks) progeny. This suggested that the results presented by Huang *et al.* [10] provided an incomplete picture of the nature of instability. Therefore we determined the proportion of kanamycin resistant progeny arising from a large number of individual self-fertilised seed capsules for each of the nine independent kr plants.

For four of these plants (kr2.1, kr2.2, kr2.4 and kr2.6), all seed capsules gave the expected 3:1 Mendelian ratio of kr:ks progeny (Table 1), indicating that kanamycin resistance in these lines was relatively stable. Four plants (kr2.3, kr2.5, kr2.7 and kr2.10) showed variability between seed capsules of the same plant (Table 1 and Figure 1), with seed from some capsules showing the expected 3:1 ratio and others showing a significant reduction in the proportion of resistant progeny. The severity of this reduction was variable between seed capsules, ranging from the statistical threshold of detection to a complete loss of kanamycin resistance. There appeared to be a loose clustering of seed capsules with aberrant ratios (Figure 1), suggesting that, at least in some cases, kanamycin resistance was lost somatically in the cell lineages leading to those seed capsules. The ninth plant, kr2.9, showed a significant reduction in the proportion of kanamycin resistant progeny in all seed capsules tested, with ratios approximating 1:1 instead of the expected 3:1 kr:ks (Table 1).

When the seed capsules that deviated significantly from 3:1 kr:ks were excluded and the data for the remaining seed capsules of each plant combined to give larger numbers for analysis, the progeny of three of the four unstable plants (kr2.3, kr2.5 and kr2.7) showed a statistically significant deviation from 3:1 kr:ks (Table 2). In all cases this was caused by a reduction in the proportion of kanamycin resistant progeny, indicating that instability was also playing a role in some of the seed capsules which initially appeared to be segregating normally. A paucity of resistant progeny was also revealed for the pooled data of kr2.10, though the reduction was not significant. This is surprising as the proportion of capsules showing significant deviation from 3:1 kr:ks was highest in this line. Although analyses of individual capsules of kr2.1 suggested normal Mendelian inheritance, when the data from all its seed capsules were combined, a significant deviation from 3:1 kr:ks was revealed, indicating that this line shows some instability. In contrast, no significant instability could be detected by pooling data for kr2.2, kr2.4 or kr2.6.

Table 1. Segregation of kanamycin resistance in individual self-fertilised seed capsules of kr2.1-2.10.

	Seed Capsules	Highest %R (*P*)	Lowest %R (*P*)
kr2.1	31	78% (NS)	69% (NS)
kr2.2	26	80% (NS)	70% (**)[a]
kr2.3	22	78% (NS)	0% (***)
kr2.4	25	82% (NS)	68% (NS)
kr2.5	27	77% (NS)	0% (***)
kr2.6	18	78% (NS)	71% (NS)
kr2.7	24	77% (NS)	57% (***)
kr2.9	39	60% (***)	44% (***)
kr2.10	31	78% (NS)	0% (***)

The percentage of kanamycin resistant progeny from the seed capsules which gave the highest and lowest kr:ks ratios for each of kr2.1-2.10 are shown. Approximately 200–250 seeds from each capsule were tested for kanamycin resistance. *P* values correspond to deviation from 3:1 kr:ks.
NS $P>0.01$, ** $P<0.01$, *** $P<0.001$.
[a]This is the only seed capsule of kr2.2 which deviated significantly from 75% kanamycin resistance, so given the level of significance it is likely to represent random variation rather than a biological effect.
doi:10.1371/journal.pgen.1000323.t001

Homozygous lines of kr2.3, kr2.7 and kr2.10 were obtained by growing self-fertilised progeny of the original hemizygous plants and selecting those which gave 100% kanamycin resistant progeny after self-fertilisation. Individual plants from these 100% resistant progeny (one for each kr line) were then grown to maturity and self-fertilised and backcrossed progeny were tested for kanamycin resistance. The homozygous kr2.3 plant behaved in a similar way to the original hemizygous plant; some crosses segregated according to Mendelian expectations for a hemizygote, while others showed a reduced proportion of kanamycin resistant progeny (Figure 2A). None showed a significant increase in the proportion of kanamycin resistant progeny compared with expectations for a hemizygote. Therefore, it is most likely that one copy of *neo* was lost early in the development of this plant, causing an initially homozygous plant to become hemizygous. All progeny from the homozygous kr2.7 plant were resistant to kanamycin, so instability could not be detected in this case (Figure 2B). The homozygous kr2.10 plant gave some kanamycin sensitive progeny from most crosses, but kanamycin resistant percentages were close to 100% (Figure 2C). Therefore, in this case we can be very confident that the instability of kanamycin resistance was occurring in a homozygote. All these data indicate that the instability is maintained in descendents of the original kr plants and that it is not affected by homozygosity. Furthermore, the instability does not appear to be affected by the type of cross (self-fertilisation or backcrossing) or by the direction of backcross-

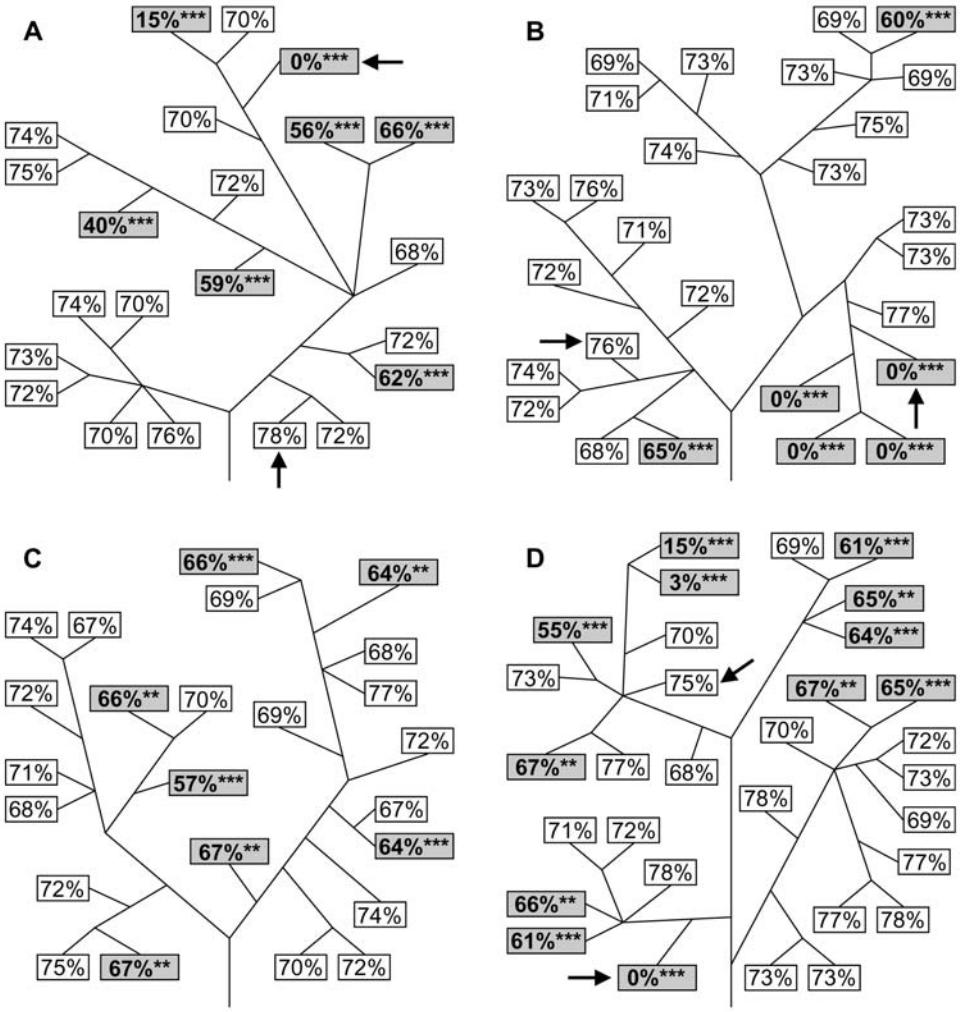

Figure 1. Instability of kanamycin resistance. The percentages of kanamycin resistant progeny from a number of self-fertilised seed capsules are shown for kr2.3 (A), kr2.5 (B), kr2.7 (C) and kr2.10 (D). Each box represents a seed capsule, with the percentage of kanamycin resistant progeny from that capsule shown. Approximately 200–250 seeds from each capsule were tested for kanamycin resistance. Lines represent branches (not to scale) and are included to show the branching pattern of the plants from which individual seed capsules were progeny tested. Seed capsules that deviate significantly ($P<0.01$) from the expected 75% kanamycin resistant progeny are highlighted with grey shading. Arrows indicate seed capsules which were used for PCR and DNA blot analysis (see Figures 3 and 4). ** $P<0.01$, *** $P<0.001$.
doi:10.1371/journal.pgen.1000323.g001

ing (Figure 2). Consistent with results for the initial kr plants (Figure 1), the kr2.7 line appears to be more stable than kr2.3 or kr2.10.

The Molecular Basis of Instability

Loss of kanamycin resistance could be caused by a genetic change in the *neo* gene (deletion or sequence decay), or by silencing of the gene through epigenetic mechanisms, a phenomenon which is commonly observed with plant transgenes [14]. To distinguish between these possibilities, progeny from seed capsules of kr2.3, kr2.5 and kr2.10 which had completely lost kanamycin resistance (Figure 1) were analysed for the presence of *neo* by PCR (Figure 3A–C). In all three cases, *neo* could not be amplified using primers designed to amplify most of the gene, suggesting that at least part of it had been deleted. In contrast the primers were able to amplify the target sequences from sibling plants derived from normally segregating seed capsules. Further analyses with alternative primers designed to amplify smaller fragments also suggested that extensive deletion was involved (data not shown). Therefore, hybridisation experiments were performed, which confirmed that the *neo* gene had been lost (Figure 4A). In addition, probing was performed with *aadA*, a gene which was initially used for selection of transplastomic lines [10] and which was co-transferred in whole or part to the nucleus in kr2.3, kr2.5 and kr2.10 [11]. In progeny of kr2.3 and kr2.10 *aadA* was also lost, suggesting deletions of at least 2.4 kb encompassing both genes (Figure 4B). In the kr2.5 integrant, only a small fragment of *aadA* is present (Figure 5; see below) and consequently it was not detected by the hybridisation experiments (Figure 4B). Therefore it could not be determined whether the partial gene was lost in this case.

Progeny of kr2.9, the plant that consistently gave ~50% resistant progeny from self-fertilisation, were also analysed for the presence of *neo* by PCR. Since there were no seed capsules where kanamycin resistance had been completely lost, 36 seeds from a typical, previously tested, capsule were grown in the absence of selection and assayed individually by PCR (Figure 3D). Thirteen of these gave a positive result, which deviates significantly ($P<0.0001$) from the 75% that would be expected if epigenetic

Table 2. Overall segregation of kanamycin resistance in self-fertilised progeny of kr2.1-2.10.

	Resistant	Sensitive	*P*
kr2.1	6479	2327	**
kr2.2[a]	5342	1840	NS
kr2.3	2974	1127	***
kr2.4	3925	1366	NS
kr2.5	3770	1429	***
kr2.6	3336	1149	NS
kr2.7	2601	1050	***
kr2.10	3090	1124	NS

The total number of kanamycin resistant and sensitive seedlings from all seed capsules of kr2.1-2.10, excluding those which deviated significantly ($P<0.01$) from 3:1 kr:ks, are shown. *P* values correspond to deviation from 3:1 kr:ks. NS $P>0.01$, ** $P<0.01$, *** $P<0.001$.

[a]In this case the one seed capsule which deviated significantly from 3:1 kr:ks was included in the analysis (see Table 1).

doi:10.1371/journal.pgen.1000323.t002

silencing were the mechanism responsible. Therefore, the instability in kr2.9 is also due to the loss of *neo*, even though it behaves in quite a different way to the other unstable kr plants (Table 1).

Sequence Analysis of Integration Sites

Based on the above results, the loss of *neo* must occur by deletion of some or all of the *nupt* or by such a high level of sequence decay that specific probes and primers are no longer able to recognise the remaining sequences. Clearly, it would be useful to know what sequences are involved in plastid sequence integration in the nuclear genome and subsequent deletion and / or decay. Ideally we would like to know the sequence of the pre-insertion site (i.e. the sequence present prior to insertion), the sequence of the whole integrant and what remains following deletion. However, determining these sequences is not a trivial process. Even the first step (of determining the pre-insertion site) is a challenging task. The reason for this is that the integrants are often very large, with *neo* and *aadA* being flanked by many kilobases of cointegrated ptDNA [10,11,13]. Techniques that are generally used for determining flanking sequence in transgenic lines, such as genome walking and thermal asymmetric interlaced (TAIL-) PCR, have two limitations that prevent their use in this context (except in special cases, see below). They require the use of a primer that binds uniquely to the target sequence (i.e. the primer binding site must not be present elsewhere in the genome) and the amount of sequence information that can be obtained is generally only up to a few kilobases from the primer binding site. Since ptDNA sequences are present in high copy number in the nuclear genome as pre-existing *nupts* [9,15], primers that bind within ptDNA are useless for genome walking and TAIL-PCR. Therefore, these approaches can only utilise primers that bind within *neo* or *aadA*. As the amount of ptDNA flanking these marker genes exceeds the amount of sequence information that can be obtained using these techniques in the vast majority of cases, these approaches are not useful for determining junction sequences beyond the ptDNA immediately flanking *neo* and *aadA*. Therefore, they cannot be used in determining the pre-insertion site unless one of the marker genes is very close to the integrant boundary.

In kr2.5, PCR results indicated that *aadA* was truncated (data not shown), so in this special case it was possible to determine the adjacent sequence using TAIL-PCR. This sequence is composed of

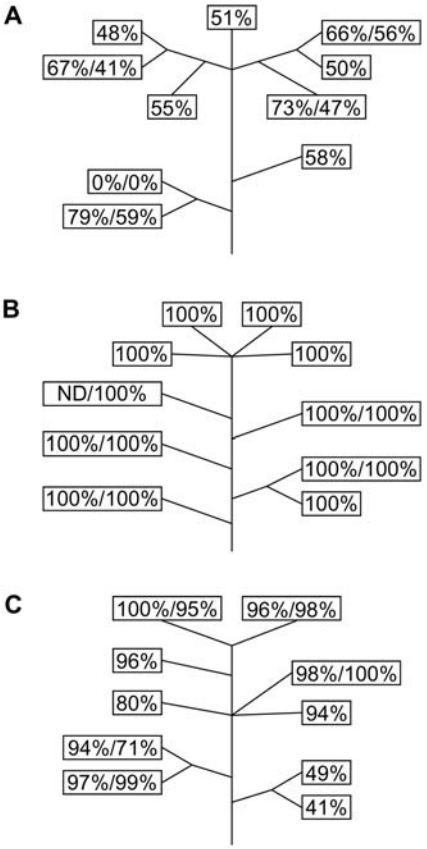

Figure 2. Analysis of instability in homozygous descendents of kr2.3 (A), kr2.7 (B) and kr2.10 (C). Each box represents a seed capsule. Where one number is shown, it indicates the percentage of kanamycin resistant progeny from a backcross to male wildtype. Where two numbers are shown, that flower was used for both self-fertilisation and backcrossing to female wildtype. The numbers indicate the percentage of kanamycin resistant progeny from self-fertilisation and backcrossing respectively. Each number represents approximately 100–150 seeds which were tested for kanamycin resistance. Lines represent branches (not to scale) and are included to show the branching pattern of the plants from which individual seed capsules were progeny tested. ND not determined.

doi:10.1371/journal.pgen.1000323.g002

41 bp of ptDNA, 25 bp of unidentified sequence, presumed to be derived from the nuclear genome, and >1.1 kb of continuous ptDNA with the first 41 bp being identical to that described above (Figure 5A). Standard PCR using a more distant ptDNA primer indicated that this ptDNA sequence continues for at least 2.6 kb. Interestingly, the ptDNA sequence is from a region at least 13 kb away from *aadA* in the transplastome (Figure 5B). Another feature of this sequence is a 3 bp region of microhomology at the *aadA* / ptDNA junction (Figure 5A). Microhomology is a characteristic of illegitimate recombination and is often found at *nupt* junctions [13]. Because the sequence adjacent to *aadA* is ptDNA, albeit distant from *aadA* in the transplastome, TAIL-PCR cannot be used to obtain any further information about the sequence of the kr2.5 integrant or its pre-insertion site, for the same reasons as described above.

An alternative method that can yield sequence information more distant from the reporter genes is inverse PCR. However, this method also has its limitations as at least one primer binding site must be within *neo* or *aadA*. Therefore obtaining sequence data for larger integrants can pose serious technical difficulties. Another barrier to determining pre-insertion sites is that the integrants are

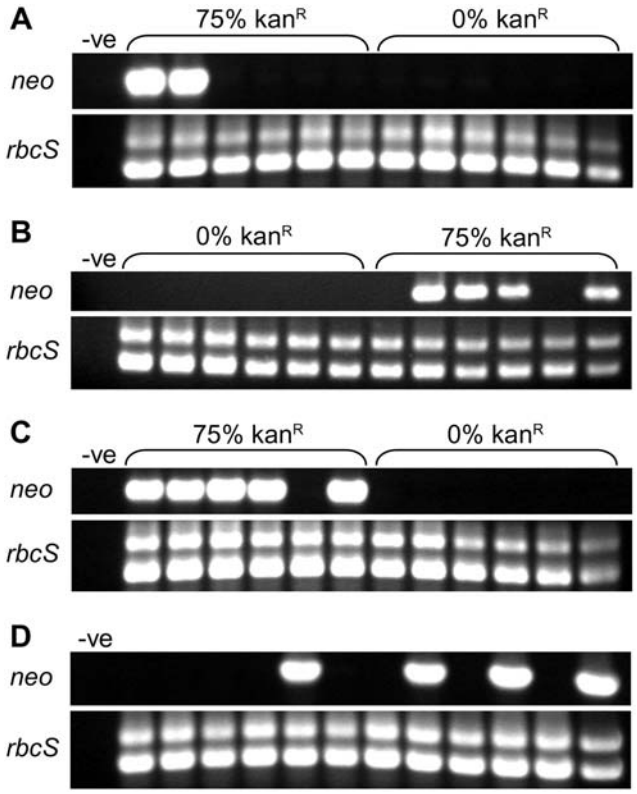

Figure 3. PCR analysis of kr lines with reduced levels of kanamycin resistance. A–C. PCR of *neo* from 6 plants grown from seeds of a normally segregating self-fertilised seed capsule (75% kan^R) and 6 plants grown from seeds of a capsule which showed no kanamycin resistant progeny (0% kan^R) for each of kr2.3 (A), kr2.5 (B) and kr2.10 (C) (see Figure 1). D. PCR of *neo* from progeny of a self-fertilised seed capsule of kr2.9 which gave 50% kanamycin resistant progeny. Results for 12 plants are shown; this is representative of a larger experiment in which 36 plants were tested (see text). In all cases control PCRs with *rbcS* primers are also shown. The *neo* primers amplify a single product of approximately 800 bp. The *rbcS* primers amplify two products of approximately 850 bp and 1 kb. -ve no template.
doi:10.1371/journal.pgen.1000323.g003

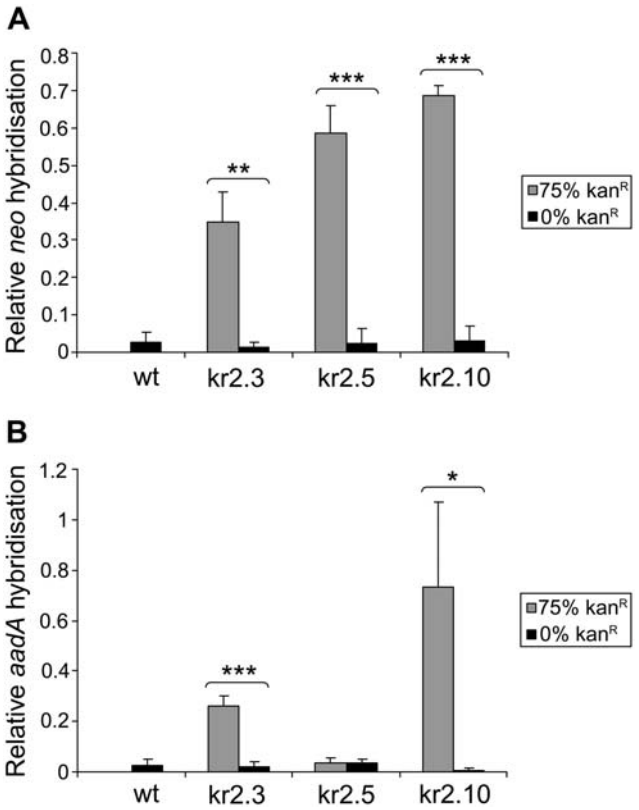

Figure 4. Quantification of DNA blot analysis. DNA slot blotting was performed using pooled DNA from at least 20 plants grown from seeds of a normally segregating self-fertilised seed capsule (75% kan^R) and pooled DNA from at least 20 plants grown from a capsule which showed no kanamycin resistant progeny (0% kan^R) for each of kr2.3, kr2.5 and kr2.10 (see Figure 1), as well as a wildtype control. Triplicates of each sample were probed with *neo* or *aadA* probes, followed by probing with ribosomal DNA as a loading control. The graphs show average hybridisation to *neo* (A) or *aadA* (B) after normalisation to ribosomal DNA hybridisation. Error bars show standard deviation. * $P<0.05$ ** $P<0.01$, *** $P<0.001$; Student's *t*-test.
doi:10.1371/journal.pgen.1000323.g004

often quite complex [13], with various rearrangements that have occurred during, or subsequent to, integration. Nevertheless, inverse PCR has been used with some success to determine border sequences for previously generated kr lines, although this has mainly been limited to smaller integrants and internal border sequences [10,13]. Despite this partial success, it has not been possible to confirm a single pre-insertion site, let alone the complete integrant sequence or its remnants after deletion in the case of an unstable kr line. Also, the sequence information that has been determined previously is essentially limited to kr lines which appear to be stable. Furthermore, the kr lines for which *neo* deletion has been confirmed (kr2.3, kr2.5 and kr2.10) appear to have particularly large integrants on the basis of Southern analysis [11]. This makes them even more difficult to work with than the kr lines for which partial success has been reported previously [10,13]. For these reasons we have not been able to further describe the process of instability at the sequence level.

Discussion

We have shown that instability of kanamycin resistance in kr2.3, kr2.5 and kr2.10 is due to the absence of *neo*. Chimerism of the initial kr plants as an explanation for this is ruled out by the instability also being present in subsequent generations. Somatic recombination is ruled out by the observation that homozygous plants are also unstable. Therefore, the instability must be caused by the loss of *neo* due to deletion of some or all of the chromosomal sequence containing the integrant or its large-scale degeneration. It is disappointing that we have not been able to obtain sequence information to shed light on the mechanism of loss. Analyses of organelle DNA insertions in the nuclear genomes of Arabidopsis and rice have suggested that deletions occur by replication slippage, as deleted fragments are often flanked by short direct repeats [16,17]. However, the largest deletions observed in these studies were only a few hundred base pairs long. Therefore, it is not clear whether the same mechanisms are involved here, as the complete loss of *neo* and *aadA* would require a much larger deletion. We have invested much time in investigating the process of insertion and deletion at the sequence level, with minimal success. It appears that the most fruitful approach would be to construct individual BAC libraries for each unstable and deleted genotype, but even this approach could be problematic because the DNA used for construction of the libraries would be subject to variation as a result of the instability.

A

ATGAGGGAAGCGGTGATCGCCGAAGTATCGACTCAACTAT
CAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCGAC
GTTGCTGGCCGTACATTAATTCTGGGAATATTAATGATAG
ATAAAAATAGATCGACTCAACTACCAAGATTTTATCTATC
ATTAATTCTGGGAATATTAATGATAGATAAAAATAGATCG
ATGTATAATCTTTGAATGAATAATTTTAGAAAATAATGGA
ATTTCCATATTAATCGAGTATTTCTTCTTTTTAATATTTG

B

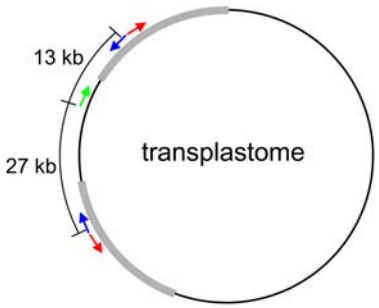

Figure 5. Origin and partial sequence of the kr2.5 nuclear integrant. A. Partial sequence of the kr2.5 nuclear integrant. Blue text indicates *aadA* sequence, beginning from the start codon, and green text indicates plastid sequence, with the two bold regions being identical. The box highlights a 3 bp region of overlap between *aadA* and plastid sequences. Black text indicates non-plastid sequence of unknown origin. B. Structure of the transplastome showing the origin of kr2.5 nuclear integrant sequences (not to scale). *Neo* and *aadA* are represented by red and blue arrows respectively. They are present in two copies as they are located within the inverted repeats (shown in grey). The green arrow indicates plastid sequence adjacent to *aadA* in the kr2.5 nuclear integrant. In each case, the direction of the arrow corresponds to the direction of the sequence as written in part A.
doi:10.1371/journal.pgen.1000323.g005

The under representation of kanamycin resistant progeny from kr2.9 also is due to the absence of *neo*, but it may be caused by a different underlying mechanism, since the kr:ks ratio is consistently altered to approximately 1:1 in self-fertilised progeny. One possibility for this could be non-transmission of *neo* through either the male or the female germline. However, backcrossing kr2.9 to male wildtype gave 24% resistant progeny (n = 232), which is not consistent with this explanation. Since the kr:ks ratio was found to be approximately the same in 39 self-fertilisations, it seems likely that *neo* is mitotically stable but meiotically unstable, with approximately 50% loss during both male and female meiosis. This loss could be occurring by a gene conversion-like process where a template is used to 'correct' the kr2.9 integrant sequence, resulting in the removal of *neo*. The template could be the native homologous or homeologous allele (*N. tabacum* is an allotetraploid), or an adjacent pre-existing *nupt* with homology to the native ptDNA that flanks *neo* in the experimental construct.

The kr2.5 integrant has been partially characterised and *aadA* is adjacent to a *nupt* sequence which is physically distant from *aadA* in the transplastome. There are several possible explanations for this observation. Firstly, this plastid sequence may be part of a pre-existing *nupt*. However, sequencing of 2.6 kb revealed perfect identity to the plastid genome (data not shown), indicating that this sequence is of very recent origin in the nucleus. Secondly, it may have integrated into the nuclear genome as part of the same event as *neo* and *aadA* with rearrangement occurring at the time of integration. Finally, it may have integrated into the nuclear genome with *neo* and *aadA* as a continuous sequence from the transplastome and subsequent rearrangement and / or deletion may have then brought it into the vicinity of *aadA*. This final possibility could be explained by mechanisms similar to those envisaged for the deletion of *neo*.

Why do some kr lines show a high level of instability, while others appear to be more stable? One possibility is that the chromosomal location and sequence context of the integrant determines the level of stability. For example, nuclear integration of organellar sequences may be dependent on the formation of double strand breaks (DSBs) [13,18,19] and if some regions of the genome are particularly prone to DSBs, as is the case for meiotic recombination hotspots in yeast [20], this could facilitate both integration and removal of *nupts* in these regions. Differing levels of stability could represent differing tendencies to sustain DSBs. Another possibility is that the level of stability depends on the sequence of the integrant itself, rather than the surrounding sequence. In this case *nupts* may be recognised as foreign DNA and subsequently removed. For example, the recognition could occur via differences in methylation status, as plant nuclear DNA is highly methylated and ptDNA is not [21,22]. Certain plastid sequences may be more prone to elimination than others, or alternatively the level of stability may depend on the size of the integrant. Differing levels of stability also may be related to differences in transgene copy number. Kr2.3, kr2.5, kr2.7 and kr2.10 all display instability, but kr2.7 appears to be the least unstable of this group (Figures 1 and 2). Southern blotting indicates that kr2.3, kr2.5 and kr2.10 have single or low copy insertions while kr2.7 appears to have several copies of *neo* [11]. Therefore it may be that in the case of kr2.7, the loss of kanamycin resistance requires several deletion events or an infrequent large deletion.

Some nuclear genomes contain large numbers of *nupts* [9]. The high level of deletion that has been observed in this study raises the question of why supposedly non-functional *nupts* are retained in nuclear genomes for long periods. Firstly, it appears that some integrants are more stable than others in terms of deletion frequency, so it is possible that a small proportion of *nupts* are retained by chance because deletion occurs only very rarely. In support of this idea, bioinformatic analysis has indicated that recent *nupt* insertions are far more prevalent than older ones as assessed by their close similarity to extant *bona fide* ptDNA [9]. It is also possible that selection plays a role in the retention of some *nupts*. Clearly if a *nupt* is functional and it confers a selective advantage then it is likely to be retained even in the presence of some level of genomic instability. Therefore, some *nupts* that have been retained in nuclear genomes may have as yet unidentified functional significance. In addition, a *nupt* that integrates near a gene or other functionally significant region of the genome may be retained even though the *nupt* itself does not confer a selective advantage because deletions would tend to disrupt the nearby functional sequences and therefore be selected against.

Physical loss of transgenes has been reported in a wide range of plants [23–28], but little is known about the causes and mechanisms involved in transgene elimination. Furthermore, we are not aware of any studies where variation in transgene instability has been examined within a single plant. Therefore it is possible that similar mechanisms are involved in the removal of both transgenes and *nupts*. In this case, it may be that transgenic lines which appear to be stable on the basis of limited progeny testing are actually relatively unstable (as in Figure 1), an undesirable trait in biotechnological applications. Therefore the work described here is relevant not only in the context of endosymbiotic evolution, but also in the broader context of transgenic plant research and crop production.

Previous studies have shown that ptDNA is integrated into the nucleus at high frequencies that can be measured in the laboratory [10–12]. Here we have shown that, in some cases, newly integrated ptDNA is also removed from the nuclear genome at high frequency within a single generation. Presumably a more thorough analysis over many generations would reveal losses in other lines as well. This provides an explanation for the avoidance of increasing genome size in the presence of such a high transfer frequency. What is the functional significance of having such high frequencies of ptDNA integration and subsequent deletion? Many nuclear genes are plastid derived [7] so it is feasible that a high transfer frequency provides more opportunity for the evolution of these plastid-derived nuclear genes. However, it is clear that the vast majority of ptDNA integrations result in non-functional sequences. Therefore, if deletions and rearrangements involving part of the integrant and / or flanking sequence are frequent, this not only counterbalances the problem of increasing genome size associated with a large transfer frequency, but also provides many more opportunities for ptDNA to attain functional sequence contexts in the nucleus. In support of this idea, it has been shown that partial *nupt* deletions resulting in nuclear activation of a plastid gene can be detected in the laboratory [6]. Furthermore, it has recently been shown that novel nuclear exons can be generated from non-coding organellar DNA sequences [29]. The mechanism of *nupt* deletion described here may therefore be fundamentally important in eukaryotic evolution by providing a significant source of new functional sequences in the nuclear genome.

Materials and Methods

Plant Growth Conditions

Nicotiana tabacum plants were grown in soil in a controlled environment chamber with a 14 hr light/10 hr dark and 25°C day/18°C night growth regime.

Analysis of Kanamycin Resistance

Kanamycin selection was performed by plating surface-sterilised seeds on ½ MS salt medium [30] containing 150 μg ml^{-1} kanamycin. Plates were incubated at 25°C with 16 hr light/8 hr dark.

DNA Extraction

DNA extraction was performed using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

PCR

Standard PCR was performed in 25 μl reactions with 1.5 U *Taq* DNA Polymerase (New England Biolabs, Ipswich, MA), 1× ThermoPol Reaction Buffer, 10 pmol each primer, 0.5 mM dNTPs, and ~100 ng genomic DNA or ~1 ng plasmid DNA template. Cycling was performed with an initial denaturation step at 94°C for 2 min followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 30 sec and extension at 72°C for 1 min. Primers used were neoF and neoR for *neo* PCRs and rbcSF and rbcSR for *rbcS* PCRs. TAIL-PCR was performed as described [31] using the degenerate primer AD1 [32] and specific primers TaadA1, TaadA2 and TaadA3. Standard PCR to extend the ptDNA sequence obtained from TAIL-PCR of kr2.5 was performed using primers TaadA2 and cp130676R as above except that a 3 min extension time was used. See Table 3 for primer sequences.

Table 3. Primers used in this study.

Primer name	Sequence (5′–3′)
neoF	TTGAACAAGATGGATTGCACGCAGG
neoR	GAACTCGTCAAGAAGGCGATAGAAGG
rbcSF	GGTGGGCAACTATGCAATGACC
rbcSR	CTTGACGCACGTTGTCGAATCC
aadAF	AGTATCGACTCAACTATCAGAGG
aadAR	GACTACCTTGGTGATCTCGCCTTTC
TaadA1	CCAAGATTTTACCATGAGGGAAGCGGTG
TaadA2	GCCGAAGTATCGACTCAACTATCAGAGG
TaadA3	GTCATCGAGCGCCATCTCGAACCGAC
cp130676R	GGAAGAAATCCGAGTGAATG

doi:10.1371/journal.pgen.1000323.t003

DNA Blot Analysis

For DNA slot blotting, 2 μg of DNA per slot was transferred to Amersham Hybond-N+ membrane (GE Healthcare, Buckinghamshire, UK) using a SRC 072/0 Minifold II slot blotting apparatus (Schleicher & Schuell). Membranes were probed with [^{32}P]-dATP labelled probe. Detection and quantification was performed using a Typhoon Trio imaging system and ImageQuant TL software (GE Healthcare, Buckinghamshire, UK). *Neo* and *aadA* probes were generated by PCR using primers neoF and neoR or aadAF and aadAR with pPRV111A::neoSTLS2 [10] as template. pCU5 [33] was used as a ribosomal DNA probe.

Statistical Analysis

Significance of deviation from an expected Mendelian ratio was determined using a Chi-squared test. Due to the large number of these tests performed, only P values <0.01 were considered to be significant in order to minimise the number of false positives.

Acknowledgments

We thank Yuan Li for technical assistance and Ralph Bock for discussion.

Author Contributions

Conceived and designed the experiments: AES JNT. Performed the experiments: AES. Analyzed the data: AES. Wrote the paper: AES JNT.

References

1. Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. Science 283: 1476–1481.
2. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, et al. (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol 15: 1325–1330.
3. Adams KL, Daley DO, Qiu YL, Whelan J, Palmer JD (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature 408: 354–357.
4. Adams KL, Qiu YL, Stoutemyer M, Palmer JD (2002) Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U S A 99: 9905–9912.
5. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, et al. (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13: 645–658.

6. Stegemann S, Bock R (2006) Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. Plant Cell 18: 2869–2878.
7. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci U S A 99: 12246–12251.
8. Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D (2006) Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr Biol 16: 2320–2325.
9. Richly E, Leister D (2004) NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Mol Biol Evol 21: 1972–1980.
10. Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature 422: 72–76.
11. Sheppard AE, Ayliffe MA, Blatch L, Day A, Delaney SK, et al. (2008) Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. Plant Physiol 148: 328–336.
12. Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. Proc Natl Acad Sci U S A 100: 8828–8833.
13. Huang CY, Ayliffe MA, Timmis JN (2004) Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. Proc Natl Acad Sci U S A 101: 9710–9715.
14. Matzke AJ, Matzke MA (1998) Position effects and epigenetic silencing of plant transgenes. Curr Opin Plant Biol 1: 142–148.
15. Ayliffe MA, Timmis JN (1992) Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. Theor Appl Genet 85: 229–238.
16. Huang CY, Grunheit N, Ahmadinejad N, Timmis JN, Martin W (2005) Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiol 138: 1723–1733.
17. Noutsos C, Richly E, Leister D (2005) Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res 15: 616–628.
18. Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 402: 96–100.
19. Yu X, Gabriel A (1999) Patching broken chromosomes with extranuclear cellular DNA. Mol Cell 4: 873–881.
20. Baudat F, Nicolas A (1997) Clustering of meiotic double-strand breaks on yeast chromosome III. Proc Natl Acad Sci U S A 94: 5213–5218.
21. Ayliffe MA, Scott NS, Timmis JN (1998) Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. Mol Biol Evol 15: 738–745.
22. Montero LM, Filipski J, Gil P, Capel J, Martinez-Zapater JM, et al. (1992) The distribution of 5-methylcytosine in the nuclear genome of plants. Nucleic Acids Res 20: 3207–3210.
23. Romano E, Soares A, Proite K, Neiva S, Grossi M, et al. (2005) Transgene elimination in genetically modified dry bean and soybean lines. Genet Mol Res 4: 177–184.
24. Limanton-Grevet A, Jullien M (2001) *Agrobacterium*-mediated transformation of *Asparagus officinalis* L.: molecular and genetic analysis of transgenic plants. Mol Breed 7: 141–150.
25. Joersbo M, Brunstedt J, Marcussen J, Okkels FT (1999) Transformation of the endospermous legume guar (*Cyamopsis tetragonoloba* L.) and analysis of transgene transmission. Mol Breed 5: 521–529.
26. Fladung M (1999) Gene stability in transgenic aspen (*Populus*). I. Flanking DNA sequences and T-DNA structure. Mol Gen Genet 260: 574–581.
27. Srivastava V, Vasil V, Vasil IK (1996) Molecular characterization of the fate of transgenes in transformed wheat (*Triticum aestivum* L.). Theor Appl Genet 92: 1031–1037.
28. Cherdshewasart W, Ghartichhetri GB, Saul MW, Jacobs M, Negrutiu I (1993) Expression instability and genetic disorders in transgenic *Nicotiana plumbaginifolia* L. plants. Transgenic Res 2: 307–320.
29. Noutsos C, Kleine T, Armbruster U, DalCorso G, Leister D (2007) Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. Trends Genet 23: 597–601.
30. Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15: 473–497.
31. Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J 8: 457–463.
32. Sessions A, Burke E, Presting G, Aux G, McElver J, et al. (2002) A high-throughput Arabidopsis reverse genetics system. Plant Cell 14: 2985–2994.
33. Kavanagh TA, Timmis JN (1986) Heterogeneity in cucumber ribosomal DNA. Theor Appl Genet 72: 337–345.

Chapter 5: Conclusions and Future Directions

Plastid-to-nucleus DNA transfer has led to the establishment of functional nuclear sequences in several different ways. Firstly, many plastid genes have relocated to the nuclear genome with the protein product now being imported into the plastid rather than encoded in the plastid's own genome. As well as DNA transfer, this also requires the gene to acquire nuclear regulatory sequences and an organelle targeting sequence. Secondly, many plastid genes have moved to the nucleus where they now perform non-plastid functions. Thirdly, a number of short regions of plastid DNA have been found which contribute sequence to nuclear exons, so it appears that plastid sequences are able to modulate the functions of existing genes. It should be noted that there are severe limitations in detecting this last type of integrant because they may be very small and diverged from the query plastid sequence, making them impossible to detect using conventional search strategies. It is also possible that plastid sequences could modulate the functions of nuclear genes in other ways, for example by contributing to regulatory sequences or by simple insertional inactivation.

Given the various ways in which plastid sequences have contributed to nuclear genomes, it appears that the plastid genome has been an important source of novel sequences which the nucleus can tinker with and occasionally generate something functionally useful. Therefore, in an evolutionary sense, plastid-to-nucleus DNA transfer can be viewed as a process somewhat analogous to mutation. The reasons for this analogy are as follows: both processes generate genetic variation, most of the time their effects will be neutral or deleterious, but occasionally they result in a genetic change which selection favours. Clearly if evolution is to occur, there must be process(es) that generate genetic variation. Therefore the process of plastid-to-nucleus DNA transfer may be beneficial, as long as it does not occur at such a high frequency that there are too many deleterious insertions.

If plastid-to-nucleus DNA transfer is beneficial, then cellular processes that favour transfer are likely to have been selected for. As described in chapter 2, the frequency of plastid-to-nucleus transfer is much higher in the male germline than in the female germline. The elevated transfer frequency in the male germline could be a consequence of programmed plastid degradation which occurs during pollen development as a mechanism for maintaining maternal inheritance of plastids. However, the situation can also be viewed the other way around. In this way programmed plastid degradation may have evolved as a way of promoting plastid-to-nucleus transfer, but at the same time confining it to a short and specific developmental stage. Having the degradation occur in the germline would ensure the transmission of nupts to the next generation and having it occur only in one sex may help to ensure faithful transmission of the plastid genome

through the germline of the other sex. Therefore, plastid-to-nucleus DNA transfer may have been a contributing selective force that favoured uniparental inheritance.

Chapter 2 also described experiments using a *gus* reporter gene to detect plastid-to-nucleus DNA transfer. This was used to investigate transfer in somatic tissues and revealed a startlingly high frequency of transfer, although the experimental design was such that some events involving plastid-to-nucleus DNA transfer without integration into a nuclear chromosome would also have been detected. Nevertheless, this work, in conjunction with previous studies (Stegemann *et al.* 2003), has led to the surprising realisation that even within a single leaf, it is likely that not all cells are genetically identical with respect to their nupt content. Since the germline is derived from somatic cells in plants, the high level of somatic nupt variation should lead to increased genetic variation in the next generation, which may be advantageous.

Chapter 3 described attempts to determine whether plastid-to-nucleus transfer occurs by a direct DNA mechanism or whether an RNA intermediate is involved. For technical reasons the experimental setup was not able to distinguish between these two modes of transfer. Numerous studies have suggested that direct DNA transfer is the predominant mechanism (although this has not been verified experimentally), but it is not clear whether RNA-mediated transfer occurs as well. If RNA-mediated transfer could be detected in the laboratory, this would indicate that RNA-mediated transfer may have an important role to play in evolution. This could be significant in two ways. Firstly, functional gene relocation from the plastid to the nucleus could occur more easily for genes containing introns or RNA editing sites. Secondly, having both direct DNA transfer and RNA-mediated transfer occurring would maximise the genetic variation generated by the transfer process (since a consequence of RNA processing is that RNA molecules do not always have the same sequence as the DNA template from which they were transcribed). As described above, this may be beneficial.

Plastid-to-nucleus DNA transfer creates genetic novelty that can lead to the generation of new nuclear genes (or parts of genes). The sequence context of plastid sequences in the nucleus is an important determinant of this, as in the case of a whole gene transfer, nuclear regulatory sequences must be acquired and in the case of modification of an existing nuclear gene, the plastid sequence must clearly be associated with that gene. Therefore post-insertional rearrangements involving nupts may be vital for the generation of anything functionally useful. Chapter 4 described a series of experiments analysing the stability of the *neo* gene following plastid-to-nucleus transfer. It was found that *neo* was highly unstable in around half of the lines tested, with deletion of *neo* occurring often within a single generation. While the deletion events have not been characterised at the sequence level, it is possible that, at least in some cases, the deletions do not involve the entire nupt. In this case the deletion process may be important in

generating the post-insertional rearrangements described above. The remarkably high occurrence of deletion that has been observed indicates that this process may be much more frequent than might be expected. Future experiments could include further attempts at sequence characterisation of the deletion events to see whether this hypothesis is supported (i.e. how big the deletions are in relation to integrant size and whether any other rearrangements are associated with the deletions). However, as discussed in detail in chapter 4, the tobacco system imposes severe limitations on the characterisation of nupts.

If plastid-to-nucleus DNA transfer is beneficial, it would be expected that evolution has selected for a relatively high transfer frequency. Furthermore, if post-insertional rearrangements are important for the generation of functionally useful sequence contexts then it would be expected that there would be selection for a high frequency of rearrangement, including large scale duplication and deletion, following transfer. Here deletion could be important not only in terms of partial nupt deletions potentially establishing useful sequence contexts, but also in terms of preventing genome size increases as a result of having a high transfer frequency. These predictions are consistent with the observations described in chapters 2 and 4. However, it is not known whether tobacco is typical with respect to nupt integration and deletion. Therefore, an important extension of the work described here will be to examine these processes in other systems. Plastid transformation technology is becoming available in other plant species, so future experiments will be able to make use of this to determine more general aspects of nupt integration and deletion. It may also be possible to provide more detailed sequence characterisation using species that have smaller, less repetitive genomes than tobacco. If other plant species show high levels of nupt integration and post-insertional rearrangement and deletion similar to the results obtained from tobacco, plastid DNA transfer to the nucleus may indeed have been a fundamental process in shaping eukaryotic nuclear genomes.

References for Chapters 1, 3 and 5

Adams K.L., Daley D.O., Qiu Y.L., Whelan J. & Palmer J.D. (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408**, 354-7.

Adams K.L. & Palmer J.D. (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* **29**, 380-95.

Ayliffe M.A., Scott N.S. & Timmis J.N. (1998) Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. *Molecular Biology and Evolution* **15**, 738-45.

Ayliffe M.A. & Timmis J.N. (1992a) Plastid DNA sequence homologies in the tobacco nuclear genome. *Molecular & General Genetics* **236**, 105-12.

Ayliffe M.A. & Timmis J.N. (1992b) Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. *Theoretical and Applied Genetics* **85**, 229-38.

Behura S.K. (2007) Analysis of nuclear copies of mitochondrial sequences in honeybee (*Apis mellifera*) genome. *Molecular Biology and Evolution* **24**, 1492-505.

Bensasson D., Feldman M.W. & Petrov D.A. (2003) Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *J Mol Evol* **57**, 343-54.

Bock R. (2000) Sense from nonsense: How the genetic information of chloroplasts is altered by RNA editing. *Biochimie* **82**, 549-57.

Bock R., Hermann M. & Kossel H. (1996) *In vivo* dissection of *cis*-acting determinants for plastid RNA editing. *Embo Journal* **15**, 5052-9.

Boore J.L. (1999) Animal mitochondrial genomes. *Nucleic Acids Res* **27**, 1767-80.

Chaudhuri S., Carrer H. & Maliga P. (1995) Site-specific factor involved in the editing of the *psbL* mRNA in tobacco plastids. *Embo Journal* **14**, 2951-7.

Depeiges A., Degroote F., Espagnol M.C. & Picard G. (2006) Translation initiation by non-AUG codons in *Arabidopsis thaliana* transgenic plants. *Plant Cell Reports* **25**, 55-61.

Deusch O., Landan G., Roettger M., Gruenheit N., Kowallik K.V., Allen J.F., Martin W. & Dagan T. (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Molecular Biology and Evolution* **25**, 748-61.

Farrelly F. & Butow R.A. (1983) Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* **301**, 296-301.

Grohmann L., Brennicke A. & Schuster W. (1992) The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in Oenothera. *Nucleic Acids Res* **20**, 5641-6.

Huang C.Y., Ayliffe M.A. & Timmis J.N. (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422**, 72-6.

Huang C.Y., Ayliffe M.A. & Timmis J.N. (2004) Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9710-5.

Jacobs H.T., Posakony J.W., Grula J.W., Roberts J.W., Xin J.H., Britten R.J. & Davidson E.H. (1983) Mitochondrial DNA sequences in the nuclear genome of Strongylocentrotus purpuratus. *J Mol Biol* **165**, 609-32.

Keeling P.J. & Palmer J.D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* **9**, 605-18.

Kobayashi Y., Dokiya Y., Kumazawa Y. & Sugita M. (2002) Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in *Nicotiana sylvestris*. *Biochemical and Biophysical Research Communications* **299**, 57-61.

Kubo N., Harada K., Hirai A. & Kadowaki K. (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: Common use of the same mitochondrial targeting signal for different proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9207-11.

Lister D.L., Bateman J.M., Purton S. & Howe C.J. (2003) DNA transfer from chloroplast to nucleus is much rarer in *Chlamydomonas* than in tobacco. *Gene* **316**, 33-8.

Liu S.L., Zhuang Y., Zhang P. & Adams K.L. (2009) Comparative Analysis of Structural Diversity and Sequence Evolution in Plant Mitochondrial Genes Transferred to the Nucleus. *Molecular Biology and Evolution* **26**, 875-91.

Maliga P. (2004) Plastid transformation in higher plants. *Annual Review of Plant Biology* **55**, 289-313.

Martin W., Rujan T., Richly E., Hansen A., Cornelsen S., Lins T., Leister D., Stoebe B., Hasegawa M. & Penny D. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12246-51.

Matsuo M., Ito Y., Yamauchi R. & Obokata J. (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. *Plant Cell* **17**, 665-75.

Matzke A.J. & Matzke M.A. (1998) Position effects and epigenetic silencing of plant transgenes. *Curr Opin Plant Biol* **1**, 142-8.

Millen R.S., Olmstead R.G., Adams K.L., Palmer J.D., Lao N.T., Heggie L., Kavanagh T.A., Hibberd J.M., Gray J.C., Morden C.W., Calie P.J., Jermiin L.S. & Wolfe K.H. (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**, 645-58.

Mishmar D., Ruiz-Pesini E., Brandon M. & Wallace D.C. (2004) Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. *Hum Mutat* **23**, 125-33.

Mourier T., Hansen A.J., Willerslev E. & Arctander P. (2001) The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol Biol Evol* **18**, 1833-7.

Murashige T. & Skoog F. (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**, 473-&.

Noutsos C., Kleine T., Armbruster U., DalCorso G. & Leister D. (2007) Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. *Trends in Genetics* **23**, 597-601.

Noutsos C., Richly E. & Leister D. (2005) Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Research* **15**, 616-28.

Nugent J.M. & Palmer J.D. (1991) RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. *Cell* **66**, 473-81.

Pamilo P., Viljakainen L. & Vihavainen A. (2007) Exceptionally high density of NUMTs in the honeybee genome. *Molecular Biology and Evolution* **24**, 1340-6.

Pereira S.L. & Baker A.J. (2004) Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. *BMC Evolutionary Biology* **4**, 8.

Ricchetti M., Tekaia F. & Dujon B. (2004) Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol* **2**, E273.

Richly E. & Leister D. (2004a) NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol* **21**, 1081-4.

Richly E. & Leister D. (2004b) NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol Biol Evol* **21**, 1972-80.

Shafer K.S., Hanekamp T., White K.H. & Thorsness P.E. (1999) Mechanisms of mitochondrial DNA escape to the nucleus in the yeast Saccharomyces cerevisiae. *Curr Genet* **36**, 183-94.

Sheppard A.E., Ayliffe M.A., Blatch L., Day A., Delaney S.K., Khairul-Fahmy N., Li Y., Madesis P., Pryor A.J. & Timmis J.N. (2008) Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. *Plant Physiology* **148**, 328-36.

Stegemann S. & Bock R. (2006) Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell* **18**, 2869-78.

Stegemann S., Hartmann S., Ruf S. & Bock R. (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci U S A* **100**, 8828-33.

Stupar R.M., Lilly J.W., Town C.D., Cheng Z., Kaul S., Buell C.R. & Jiang J. (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2:

implication of potential sequencing errors caused by large-unit repeats. *Proc Natl Acad Sci U S A* **98**, 5099-103.

Thorsness P.E. & Fox T.D. (1993) Nuclear mutations in Saccharomyces cerevisiae that affect the escape of DNA from mitochondria to the nucleus. *Genetics* **134**, 21-8.

Thorsness P.E. & Weber E.R. (1996) Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. *Int Rev Cytol* **165**, 207-34.

Timmis J.N., Ayliffe M.A., Huang C.Y. & Martin W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* **5**, 123-35.

Timmis J.N. & Scott N.S. (1983) Sequence Homology between Spinach Nuclear and Chloroplast Genomes. *Nature* **305**, 65-7.

Tourmen Y., Baris O., Dessen P., Jacques C., Malthiery Y. & Reynier P. (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. *Genomics* **80**, 71-7.

Ueda M., Fujimoto M., Arimura S.I., Tsutsumi N. & Kadowaki K.I. (2008) Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. *Molecular Biology and Evolution* **25**, 1791-3.

Wamboldt Y., Mohammed S., Elowsky C., Wittgren C., de Paula W.B.M. & Mackenzie S.A. (2009) Participation of Leaky Ribosome Scanning in Protein Dual Targeting by Alternative Translation Initiation in Higher Plants. *Plant Cell* **21**, 157-67.

Wang H.H., Yin W.B. & Hu Z.M. (2009) Advances in chloroplast engineering. *Journal of Genetics and Genomics* **36**, 387-98.

Woischnik M. & Moraes C.T. (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res* **12**, 885-93.

Yu Y.S., Rambo T., Currie J., Saski C., Kim H.R., Collura K., Thompson S., Simmons J., Yang T.J., Nah G., Patel A.J., Thurmond S., Henry D., Oates R., Palmer M. *et al.* (2003) In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* **300**, 1566-9.