Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants?

C. A. Cullis[1,*], B. J. Vorster[2], C. Van Der Vyver[3] and K. J. Kunert[2]

[1] Case Western Reserve University, Department of Biology, Cleveland, OH 4404, USA
[2] Forestry and Agricultural Biotechnology Institute, Plant Science Department, **University of Pretoria**, Pretoria 0002, South Africa
[3] Institute for Plant Biotechnology, Department of Genetics, University of Stellenbosch, Stellenbosch, 7602, South Africa

[*] For correspondence. E-mail cac5@case.edu

Abstract

Background: The presence of chloroplast-related DNA sequences in the nuclear genome is generally regarded as a relic of the process by which genes have been transferred from the chloroplast to the nucleus. The remaining chloroplast encoded genes are not identical across the plant kingdom indicating an ongoing transfer of genes from the organelle to the nucleus.

Scope: This review focuses on the active processes by which the nuclear genome might be acquiring or removing DNA sequences from the chloroplast genome. Present knowledge of the contribution to the nuclear genome of DNA originating from the chloroplast will be reviewed. In particular, the possible effects of stressful environments on the transfer of genetic material between the chloroplast and nucleus will be considered. The significance of this research and suggestions for the future research directions to identify drivers, such as stress, of the nuclear incorporation of plastid sequences are discussed.

Conclusions: The transfer to the nuclear genome of most of the protein-encoding functions for chloroplast-located proteins facilitates the control of gene expression. The continual transfer of fragments, including complete functional genes, from the chloroplast to the nucleus has been observed. However, the mechanisms by which the loss of functions and physical DNA elimination from the chloroplast genome following the transfer of those functions to the nucleus remains obscure. The frequency of polymorphism across chloroplast-related DNA fragments within a species will indicate

the rate at which these DNA fragments are incorporated and removed from the chromosomes.

Introduction

The generally accepted endosymbiotic theory for the presence of chloroplasts in the eukaryotic cell requires that many of the genes for functions that were originally encoded in the pro-chloroplast genome have been transferred to the nucleus. Typical chloroplast genomes only contain about 50–200 genes, with a core constituency of about 40 protein-encoding genes (Woodson and Chory, 2008), mostly for chloroplast proteins and the protein-synthesizing machinery (Martin, 2003; Table 1), whereas cyanobacterial genomes encode several thousand genes. On the other hand, the estimate from proteomics is that organelles contain up to several thousand different proteins (Andersson *et al.*, 2003; Richly and Leister, 2004) which is within the range of the number of predicted protein-encoding genes in current-day cyanobacteria and proteobacteria. Therefore, since most proteins that are found in chloroplasts are encoded in the nucleus, synthesized in the cytoplasm and then transported into the organelles, at some point, following the transfer and active expression of these genes in the nucleus, they must have been lost from the chloroplast genome.

Table 1. Chloroplast genome sizes and number of genes from various plant species

Chloroplast genome	Length	Unique genes	tRNA genes	Reference
Bigelowiella natans	69 166	98*	27	Rogers *et al.*, 2007
Physcomitrella patens	122 890	118	31	Sugiura *et al.*, 2003
Oryza sativa	13 4525	159*	30	Hiratsuka *et al.*, 1989
Hordeaum vulgare	136 462	113	30	Saki *et al.*, 2006
Agrostis stolonifera	136 584	113	30	Saki *et al.*, 2006
Zea mays	140 386	158*	30	Maier *et al.*, 1995
Sorghum bicolor	140 754	113	30	Saki *et al.*, 2007
Lotus japonicus	150 519	128*	30	Kato *et al.*, 2000
Glycine max	152 218	111	30	Saki *et al.*, 2005
Arabidopsis thaliana	154 478	129*	30	Sato *et al.*, 1999
Coffea arabica	155 189	112	29	Samson *et al.*, 2007

Solanum tuberosum	155 312	113	30	Chung *et al.*, 2006
Solanum lycopersicum	155 371	113	30	Daniell *et al.*, 2006
Daucus carota	155 911	115	30	Ruhlman *et al.*, 2006
Citrus sinensis	160 129	113	30	Bausher *et al.*, 2006
Eucalyptus globulus	160 286	112	30	Steane, 2005
Gossypium hirsutum	160 301	112	30	Lee *et al.*, 2006
Anthoceros formosae	161 162	121	32	Kugita *et al.*, 2003
Pelargonium x *hortorum*	217 942	110	30	Chumley *et al.*, 2006

*Data obtained from the NCBI database.

Furthermore, data from genome sequencing projects have incontrovertibly demonstrated that fragments of the extant chloroplast DNA are present in both the nuclear and chloroplast genomes (Farrelly and Butow, 1983; Scott and Timmis, 1984; Ayliffe and Timmis, 1992; Thorsness and Fox, 1990; Sun and Callis, 1993). This DNA is still being actively transferred from the chloroplast to the nucleus in an ongoing and frequent process (Ayliffe and Timmis, 1992; Huang *et al.*, 2003; Shahmuradov *et al.*, 2003; Stegemann *et al.*, 2003; Matsuo *et al.*, 2005).

This review will consider the current contribution of chloroplast DNA to the nuclear genome, the interplay between the nucleus and chloroplast with respect to the transfer of DNA sequences between the two cellular components, the persistence of these sequences, the possibility of roles played by the nuclear plastidic DNAs (NUPTs) in mutation buffering and the possible influence of stress on the frequency or reasons for these DNA transfer events. The effects of stress on the nuclear genome are the subject of much discussion, both with respect to the effect on methylation (Lukens and Zhan, 2007) and the actual genome itself (Cullis, 2005; Lexer and Fay, 2005; Madlung and Comai, 2004). The availability of chloroplast genome fragments for transfer to the nucleus might also be optimal during stress episodes since, at this time, there are regulatory signals passing between the nucleus and chloroplast in response to stress.

A summary of many of the interactions between the chloroplast and the nucleus is shown in Fig. 1. Signals pass in both directions between all the organelles, although only those between the chloroplast and nucleus are shown in detail since the mitochondrion is not being considered here. The nucleus controls the synthesis of most of the proteins in the chloroplast. Those proteins encoded in the chloroplast genome are part of multi-protein complexes which facilitate the control of protein levels in the chloroplast. The chloroplast is responsible for signalling to the nucleus especially with respect to the presence of both reactive oxygen species (ROS) and redox signals which then alter nuclear gene

expression. The majority of reactive oxygen intermediates are produced in the peroxisomes and chloroplasts (Pitzscheke *et al.*, 2006) and they activate signalling pathways, such as through the MAPK (mitogen-activated protein kinase) pathway in arabidopsis. It is through these pathways that the nucleus is affected with little evidence that ROS are directly responsible for the responses in the nucleus. As part of this stress signalling, it is possible that chloroplast DNA fragments from damaged chloroplasts enter the nucleus and, at an undetermined rate, become incorporated into the nuclear genome (Martin, 2003).

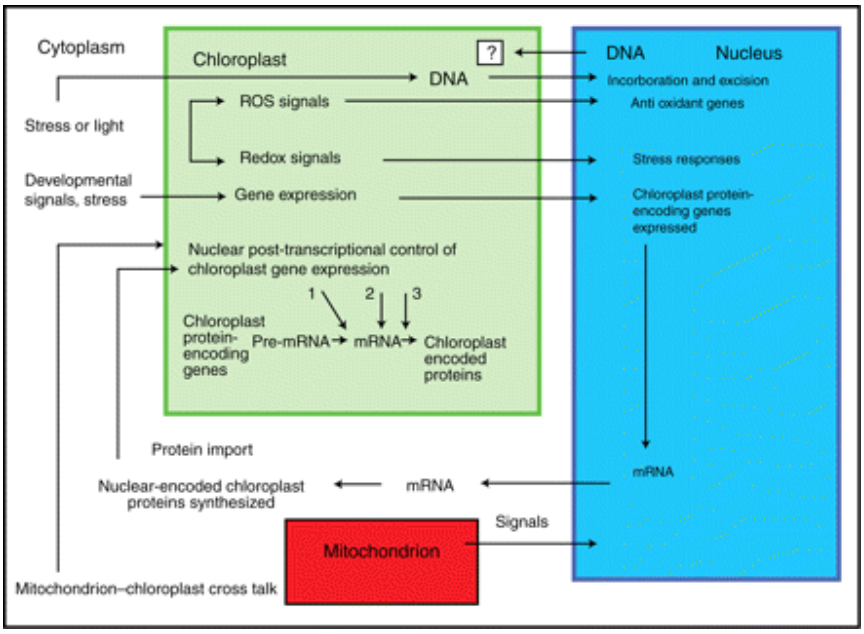

Fig. 1. The interactions between nucleus and chloroplast. Signals are transferred in both directions with the chloroplast signalling in response to developmental or stress effects while the nucleus responds with changes in gene expression for proteins to be synthesized within the cytoplasm and then imported into the chloroplast. Whether or not it is part of any signalling, chloroplast DNA fragments can be transferred to the nucleus but the reverse transfer has not been documented. Many of the same interactions also occur between the mitochondrion and the nucleus which are not documented here. Further interactions can occur between the chloroplast and mitochondrion including possible DNA transfers (adapted from Woodson and Chory, 2008).

Gene-regulatory processes under the control of the nucleus are more complex and interrelated than those under the control of organelles (Herrmann, 1997). The continuous transfer of DNA from the plastid to the nucleus must either have a neutral effect or confer some sort of positive selective advantage, otherwise natural selection would have selected against a phenomena that is undirected. A possible benefit of translocation of genes from

the plastids to the nucleus is that genes are moved from the plastid with no recombination and a high redox-load to an environment with recombination and without the associated redox-load of the plastids. This would be beneficial to the genes of the plastids but the continuous addition of more DNA into the nucleus could eventually lead to 'genome obesity' (Bennetzen and Kellogg, 1997)

Chloroplast DNA insertions into the Nuclear DNA

The completed arabidopsis and rice genome sequences have facilitated the search for evolutionary evidence of the transfer of genes from the original endosymbiont to the nucleus. Proteins encoded by the arabidopsis nuclear genome that are most similar to proteins encoded by other species' chloroplast genomes (44 plastid genes) have been identified (Arabidopsis Genome Initiative, 2000). These genes missing from the arabidopsis chloroplast genome are presumed to represent organelle-to-nuclear gene transfers that have occurred sometime after the divergence of the organelle-containing lineages. The completion of the arabidopsis genome sequence also facilitated the identification of possible nuclear sequences originating from the proto-plastids. For example, in arabidopsis there are 806 predicted nuclear-encoded proteins for which the best significant database match are 404 different proteins from the cyanobacterium *Synechocystis* (Arabidopsis Genome Initiative, 2000). These 806 predicted proteins and many others, up to 1700 (Martin *et al.*, 2002), of greatly diverse function, which are present in chloroplasts but not encoded by the chloroplast genome, are likely to have been derived from the chloroplast progenitor. Therefore, the evolutionary transfer of genes from the original captured endosymbiont to the plant nucleus and their subsequent expression is well established. There is a similar movement of mitochondrial sequences into the nuclear genome. However, the pattern can be somewhat different between sequences arising from the two organelles. The plant mitochondrial genome is generally much larger than the chloroplast genome, yet there in not the expected relatively larger contribution of mitochondria-derived sequences. Added to this is that the relative contribution across the mitochondrial genome is less representative with some regions being present in high multiplicity and others at very low multiplicity (Vorster, 2008). Therefore, it would appear that either the transfer to, or elimination from, the nuclear genome of the two organelle genomes follows different paths.

It is impossible to derive the origin and the specific sequence of events that resulted in the capture and control of all these chloroplast-localized functional proteins. However, the continual transfer of chloroplast sequences is one window into the process. What is the nature of the DNA sequences present in the nuclear genome that originated from sequences which are still present in the chloroplast genome? From the arabidopsis and rice genome sequences, different patterns of chloroplast DNA fragments in the nuclear genome have emerged. In arabidopsis, only relatively few NUPTs were found with estimates varying from 11 kbp (Arabidopsis Genome Initiative, 2000) to 20 kbp (Shahmuradov *et al.*, 2003). These fragments were generally <1 kb in length. This small

amount of chloroplast DNA is in contrast to that found in the rice genome where the DNA from the rice chloroplast aligned with a total of between 780 000 and 933 600 bp in the DNA of the nuclear genome (International Rice Genome Sequencing Project, 2005). In our own bioinformatic analysis of the chloroplast contribution to the rice genome, we found that 778 678 bp of chloroplast DNA were present in fragments greater than 100 bp in the rice genome (Vorster, 2008). This chloroplast DNA was present in >2000 different fragments, of which only 212 were of length greater than 1000 bp. It is also evident in rice that the chloroplast insertions are not evenly distributed over the chromosomes. Approximately 19·4% of the fragments were located on chromosome 1 and 22·5% on chromosome 10. In contrast, chromosome 11, which is larger than chromosome 10, only contains 0·9% of these chloroplast sequences. This analysis also showed that while most of the chloroplast genome is present to some extent, some components, notably the rRNA subunits, are fully or partially present at a much higher ratio – up to 36 (16S) and 49 (23S) copies in the nuclear genome of rice. The fragments of this linear region in the chloroplast, 23S, 16S, 5S rRNAs and the spacer between the 4·5S and 5S chloroplast rRNA genes are not uniformly represented. In Fig. 2, the positions of all the insertions of these complete or partial fragments across the rice chromosomes are shown. The height of the peak is proportional to the number of copies of the sequence in that region. It is clear that this region has very different representation with spacer being highly underrepresented and the 5S being significantly underrepresented.

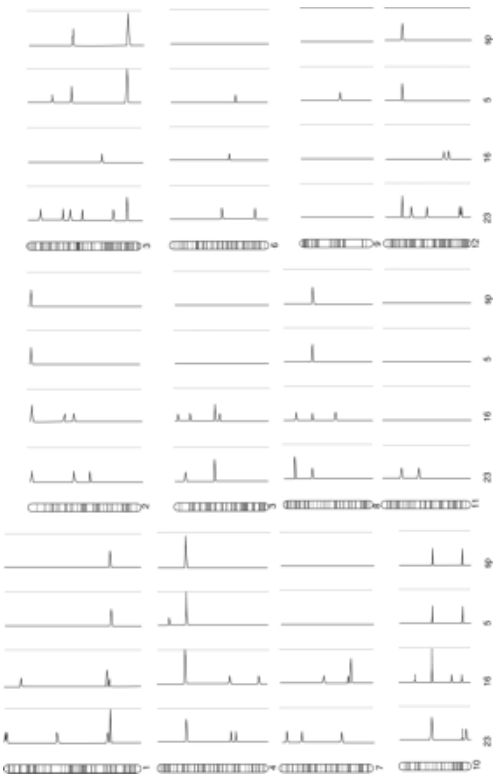

Fig. 2. Comparison of the nuclear insertions of the 23S 16S and 5S cp rRNA sequences

and that of the cp spacer between the 4·5S and 5S cp rRNA genes. The height of each peak is proportional to the number of copies of a fragment in the rice nuclear genome with similarity to one of the regions of the chloroplast rRNA genes.

The question, however, remains: is the reason for this that these sequences are more often inserted, less often removed or a combination of both? Matsou *et al.* (2005) using the rice nuclear genome data reported that the plant nuclear genome is in equilibrium between frequent integration and rapid elimination of the chloroplast genome and that the pericentromeric regions play a significant role in facilitating the chloroplast–nuclear DNA flux. This equilibrium between integration and deletion is necessary to prevent the nuclear genome continuously expanding with the contribution of chloroplast-related sequences manifestly increasing. However, the differential distribution of the chloroplast-related DNA fragments along a chromosome does not explain the differential content between chromosomes. Without any control on the site of integration into the nucleus, these chloroplast fragments have the potential to impact the function of nuclear genes and genome organization through insertional mutagenesis or the generation of potential sites for ectopic recombination resulting in chromosomal rearrangements.

Even when there is a concentration of chloroplast fragments within a region of a chromosome, the chloroplast sequences within that region are not in a linear order (Fig. 3B). Here a region contains five chloroplast fragments varying in length from 288 to 2916 bp but the fragments are from a region of the chloroplast spanning nearly 82 000 bp. Therefore, the question has to be asked is this the result of the insertion of a large fragment that has been eliminated and re-arranged substantially or it is the result of multiple insertions in a single region of the chromosome? One indication for an answer is that this region resembles, in some ways, the pattern of the complex arrangements of multiple transposable elements in maize where nested insertions of families occur. The multiple insertions of transposable elements in maize are either due to low selection for deleterious effects in these regions or to a structure that is available so that preferred sites for insertion are developed. The non-random distribution of chloroplast fragments may have arisen by similar mechanisms. The presence of multiple regions of the chloroplast ribosomal RNA genes may also facilitate ectopic recombination but this has yet to be determined. Further, it has to be asked, if this pattern is the result of many independent events, then what is special about this region? A search through the rice germplasm of the structure of this region, including chromatin structure and modification, will possibly indicate the mechanism and timing of the insertion of the chloroplast fragments into this

complex region.

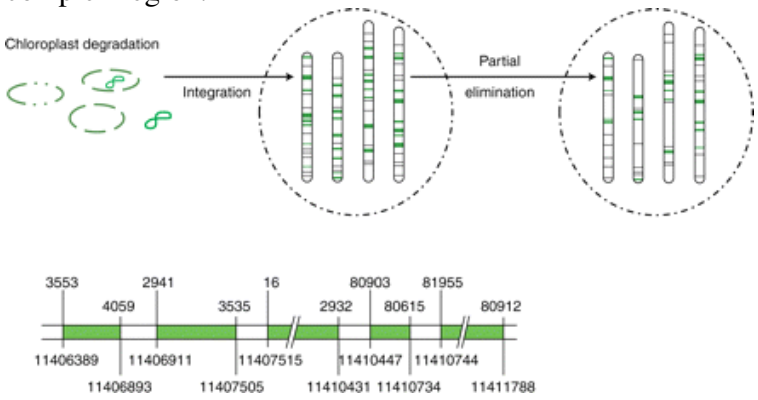

Fig. 3. (A) Model of the incorporation of chloroplast DNA into the nuclear chromosomes after chloroplast degradation, either as a result of stress or pollen formation. Chloroplast fragments are imported into the nucleus where they undergo recombination into the chromosomes. Many of the fragments are removed either completely or partially and they can also be rearranged. (B) An example of a series of non-contiguous fragments of various sizes from the chloroplast DNA inserted into a single region of a rice chromosome. The upper numbers refer to the positions on the chloroplast genome (green sections) of the regions of chloroplast homology, while the lower numbers refer to the position along the rice chromosome.

Of further interest will be data from additional complete genomes to determine if the chloroplast DNA representation in the nuclear DNA is a function of genome size as is indicated from the arabidopsis and rice data. One possibility is that plants with small genomes efficiently reject additional DNA sequences either by not incorporating, or by rapidly removing, any plastidic sequences that migrate to the nucleus as part of an overall strategy for maintaining such a small genome. The characterization of chloroplast DNA components localized in the nucleus in the many current plant genome sequencing projects will throw light on this question.

How frequently does the transfer of chloroplast DNA fragments (rather than complete genes) currently occur in plants? A high frequency of chloroplast DNA transfer has been inferred from data on the age of chloroplast insertions within the rice nuclear genome (Matsou *et al.*, 2005) This high frequency of transfer has been experimentally confirmed. Two different techniques, either by using selection in regenerated tobacco plants or by measuring transfer following meiosis (Stegemann *et al.*, 2003; Huang *et al.*, 2003) have demonstrated the transfer of functional genes from the chloroplast to the nucleus. The estimate of the frequency of functional gene transfer made by measuring the rate at which selectable plants arose in progeny of plants containing a nuclear-functioning selectable gene that had been integrated into the tobacco chloroplast (Huang *et al.*, 2003) demonstrated that functionally intact DNA was transferred from the chloroplast and integrated into, and functional in, the nucleus at a frequency of one in approx. 16 000 tobacco pollen grains. Since the selection was for a functional gene of >1·5 kb, only the

transfer of a large fragment would effectively allow resistance to be expressed. Therefore, considering the sizes of chloroplast fragments found in plant genomes this estimate of 1 in 16 000 is likely to be a very great underestimate of the actual rate of transfer of parts of this fragment of chloroplast DNA to the nucleus. In addition, there is no *a priori* reason to expect that the selected fragment would be preferentially transferred as supported by the approximately even distribution of fragments across the chloroplast genome that have been integrated into the rice genome. The apparent lack of preference for specific chloroplast regions results in an expected overall rate of transfer of chloroplast DNA to the nucleus of about 100-fold greater than that observed in the experiments. A final correction to the transfer rate calculations comes from the suggestion that <10% of the rice NUPTs were >1 kbp. Therefore, applying these corrections to the tobacco data would result in 1 out of every 16 pollen grains having a new chloroplast DNA insertion. The long-term corollary to these experiments, namely how long does it take for the transgene to be eliminated from the chloroplast has yet to be completed. The fate of these chloroplast-located transgenes and their degradation when no selective pressure is applied may indicate the pathway that loss of information from the chloroplast genome proceeds. Alternatively, inserting a functional chloroplast copy of a nuclear gene, thought to be originally derived from the proto-endosymbiont, and charting its decay might be a better model for describing the endogenous process.

The equilibrium within the nuclear genome between integration and deletion of chloroplast sequences is necessary to prevent the nuclear genome from continuously expanding with the contribution of chloroplast-related sequences manifestly increasing. Additionally, without some control of the position of integration into the nucleus, these chloroplast fragments have the potential to impact on the function of nuclear genes and genome organization through insertional mutagenesis or generation of potential sites for recombination resulting in chromosomal rearrangements.

The effect of stress on the appearance of resistant progeny has so far not been determined., although in the case of the Stegemann *et al.* (2003) experiment, where approx. 1 in 5 x 10^6 cells assayed had the transfer of a functional gene, the cells were under stress (they were dying unless the gene was transferred and functional). The overall frequency of the transfer of chloroplast sequences to the nucleus when plants are grown under optimal or stress conditions could be experimentally determined. DNA could be isolated from nuclei of individual progeny plants from inbred lines grown under different conditions. The total chloroplast contribution to the nuclear DNA could be determined by quantitative polymerase chain reaction (PCR) using a set of primers spanning the complete chloroplast genome. Changes in the relative amounts of the chloroplast, and more particularly changes in the relative numbers of copies of specific fragments, could indicate either increased or decreased integration of chloroplast genome regions into the nucleus.

The transfer of organellar DNA to the nucleus is likely to have been a driving force in the evolution of eukaryotic cells. However, it is not as clear how the sequences, even once they had become functional in the nucleus, were then lost from the plastid genome. One possible clue to this process is in the occurrence of albino plants arising from anther

culture (Day and Ellis, 1985). In these plants, the defects are usually in a rearranged chloroplast genome, but there is evidence for the involvement of both nuclear and chloroplast sequences in the process (Ankele *et al.*, 2005). Therefore, selection would eliminate these plants with a defective chloroplast genome, which would be non-viable. However, if the inactivated chloroplast genes had already been transferred and become functional in the nucleus, then such a re-arranged chloroplast genome could survive. The primary rate of transposition of chloroplast sequences into the nucleus is unlikely to be the rate-determining step for the reduction of chloroplast genome size. The newly acquired NUPT would need to obtain appropriate nuclear signals for transcription, translation and finally for protein transport back into the chloroplast to become functional. Therefore, these would need to be developed without any specific selection pressure and await active participation in the cellular machinery until the chloroplast-located copy of the gene has been inactivated in some fashion. However, as described earlier, there seems to be a rapid turnover of NUPTs. The question still is: how do all these genes manage to survive long enough to become active and replace the equivalent chloroplast gene?

Much of the information about DNA insertions is derived from bioinformatics analyses of genome sequence data. However, there is evidence for this rapid turnover of NUPTs in data derived from comparisons of closely related genomes (C. A. Cullis *et al.*, unpubl. res.). Representational difference analysis (RDA) uses DNA subtraction, PCR and a sampling of restriction fragments to increase the enrichment of target sequences (Lisitsyn, 1995). Essentially, the procedure removes all the sequences that two DNA samples have in common by hybridization with one of the DNAs in vast excess. Only those fragments present in the minor DNA representation are then amplified and characterized. The procedure requires the two DNAs to be digested to completion by a restriction enzyme, adaptor sequences added to the ends of the fragments and then the whole genome amplified by PCR. Those fragments between 2 kbp and 200 bp are isolated and reamplified to produce the original amplicons. The adaptor sequences are removed from all the amplicons and a new set of adaptors are attached to only one of the DNA samples (now called the tester). Then a large excess amplicons without the new adaptor (the driver) is mixed with a small amount of the tester, denatured and then allowed to reanneal. Following hybridization the mixture is amplified using the second adaptor as primer and only fragments that contain two tester molecules annealed together will amplify exponentially. The process is repeated up to four times with a new adaptor attached to the tester after each round of hybridization/amplification and increasing ratios of driver to tester DNAs. The result is the isolation of all of those restriction fragments that differed between the starting DNAs. Thus it is a powerful tool for identifying recent insertion events in DNAs from closely related individuals. When RDA has been applied to identify genomic variation among closely related plants within a species, different products containing short chloroplast fragments have invariably been identified. These fragments are frequently, but not exclusively, related the ribosomal RNA genes. Such fragments have been identified in subtractions involving flax, tomato, date and oil palm, tobacco, cowpea and banana. Therefore, it would appear that the identification of NUPTs in the genome could be a useful source of polymorphisms. In fact, the idea that the NUPT component of the genome is a very variable one could be tested by amplifying across

such regions in different accessions of the same species and determining the level of polymorphisms compared with regions not containing such NUPTs. This would be a direct measurement of the relative genomic stability of such regions.

Stress responses

The utility of having the chloroplast-located genes in the nucleus appears to be one of co-ordination of control of expression and perhaps less exposure to ROS. Since only a single source of new proteins is needed in response to changes in the cell, either due to development or stress, then with nuclear copies, the changes in all the chloroplasts can be centrally co-ordinated. Organelles experience shifting environmental conditions, including rapid redox changes, oxidative damage and changes in nutrient availability. Signalling from the chloroplast to the nucleus in response to such changes permits the co-ordination of nuclear and organelle genomes with every chloroplast in the cell co-ordinately regulated, rather than each organelle being regulated independently. Exposure to both biotic and abiotic stresses can lead to the increased accumulation of ROS, which can cause irreversible oxidative damage to cells. ROS that are generated in chloroplasts, or the damage that they cause, act as signals to modify nuclear gene expression to counteract this damage (Woodson and Chory, 2008). Since there is movement of protein signals from the chloroplast to the nucleus under stress conditions, it is possible that these exchanges also facilitate the transfer of chloroplast genome fragments to the nucleus where they could be incorporated in to the nuclear genome.

Experimental evidence that plant proteins could traffic through plasmodesmata was gained from studies conducted on *KNOTTED1* (*KN1*), a homeodomain transcription factor known to be involved in controlling cell fate in the plant meristem (Jackson and Hake, 1997). Microinjection experiments performed with recombinant KN1 revealed that this plant protein has functional properties almost identical to those of viral movement proteins (Lucas, 1995). Introduction of KN1 into cells resulted in an increase in plasmodesmata molecular-size exclusion limit and the rapid cell-to-cell movement of the fluorescently labelled protein (Jackson, 2002, Kim *et al.*, 2003, 2005).

Plants viruses spread throughout their hosts through the plasmodesmata (Lucas, 1995). Pioneering studies on plant viruses revealed that plasmodesmata allow the cell-to-cell trafficking of virally encoded proteins, termed the movement proteins (Atabekov and Dorokhov, 1984). DNA viruses replicate in the nucleus, and the export of progeny DNA (both ss- and ds-DNA) to and from the nucleus is mediated by BV1, a nuclear shuttle protein (Noueiry *et al.*, 1994; Sanderfoot and Lazarowitz, 1995). An interaction between BV1 and BC1 is thought to be required for the transfer of the DNA into a BC1-associated complex that can then be trafficked through the plasmodesmata (Sanderfoot and Lazarowitz, 1996). This combination of movement protein (BC1) and the ancillary nuclear shuttle protein (BV1) appears to be essential for the begomoviruses to exploit the endogenous RNA trafficking system of their hosts (Gilbertson *et al.*, 2003). This non-cell-autonomous protein pathway is similarly employed by the host to traffic macromolecules. Viral movement proteins bind RNA–DNA in a sequence-nonspecific manner to form nucleoprotein complexes. Host proteins are then involved in the delivery of movement proteins and nucleoprotein complexes to the plasmodesmata orifice, and a role for the cytoskeleton has also been implicated (Lucas, 2006). The delivery of chloroplast fragments to the nucleus could be facilitated by the active two-way macromolecular transport process between the nucleoplasm and the cytoplasm.

Alternatively, and more simply, DNA fragments released from damaged chloroplasts might randomly arrive in the nucleus where they are incorporated into the chromosomes (Fig. 3A).

Stress, in all its forms exerts a strong evolutionary pressure on organisms. The physiological responses of plants to stresses are well documented, but the effects on the genome are less well known. However, it is clear that genome can undergo both epigenetic and genetic remodelling caused by DNA rearrangements and transposition in response to stress (Madlung and Comai, 2004). The genome does not appear to be randomly rearranged in response. Thus specific DNA polymorphisms arise in flax in certain varieties in response to nutritional stresses (Cullis, 2005). Hot spots of DNA instability have been revealed through the study of somaclonal variation in rye (Linacero *et al.*, 2000). This increasing direct molecular evidence for stress-induced whole genome responses at the molecular level is consistent with the data from a series of experiments to identify the transfer of chloroplast-genome located sequences to the nucleus. The differences between varieties in NUPTs are striking when comparisons are made between plants that have undergone some stress conditions. In particular, genomic subtraction experiments involving RDA among plants regenerated from tissue culture have been informative. In one study, tobacco plants were regenerated *in vitro* from leaf discs into functional plants and compared with control tobacco plants not produced via a tissue culture route. In all the subtractions between plants that have been regenerated *in vitro*, chloroplast DNA fragments were found amongst the different products (C. van der Vyver, 2003). This observation has two important aspects. First, it confirms the lability of the NUPTs, in particular under stress, since there are differences within the same plant lines and is consistent with the data of Stegemann *et al.* (2003). Secondly, it indicates that the NUPTs may be particularly responsive to being mobilized when the plant encounters stress conditions. However, more data are needed to determine if the major events are either the insertion of new chloroplast DNA fragments into the nuclear genome or the deletion of NUPTs from the nuclear genome, or whether both processes occur at approximately the same rate.

A second interesting observation has been the identification of NUPTs that have varied following gamma irradiation of cowpeas. Again, the altered NUPTs were originally identified through subtraction experiments involving RDA of genomic DNA from irradiated and unirradiated plants (C. van der Vyver, unpubl. res.). As much as 40% of these isolated subtraction products were in chloroplast-related regions. However, it is unlikely that these regions represent actual mutations in the cowpea chloroplast genome for two reasons: (1) most of the observed mutations in the chloroplast fragments would be lethal to the chloroplast; and (2) non-chloroplast sequences were found at the ends of some of these fragments (C. van der Vyver, unpubl. res.). Therefore, primers were designed across some of the variable regions and used to screen newly irradiated M1 cowpea seeds. It was observed that a significant proportion, up to 24% in one specific chloroplast region, of these newly irradiated seeds had similar mutations to those identified in the previous subtraction products. Therefore, these NUPTs appear to be very susceptible to irradiation damage. From both the irradiation and tissue culture experiments we might speculate that the regions containing the NUPTs act as a mutation

buffer under various forms of stress. The reasons for this are unknown but two possibilities need to be distinguished. First, are these sequences particularly labile because of their position in the genome, i.e. are they integrated in these positions because these are very receptive regions of the genome? Alternatively, are they recognized as dispensable fragments in some way (such as a specific chromatin structure) so that they provide a sink for mutations and protect more essential regions of the genome? The NUPTs can possibly be seen as something functional, where they provide a vital function by absorbing mutational energy of stressful environments, especially radiation and free radicals, and thereby protect the sequences functioning as genes, regulators and promoters.

A final question which has received little attention relates to the possibility of the reverse movement of DNA sequences, namely from the nucleus to the chloroplast. Clearly these events are more difficult to detect and as yet there is no evidences of such transfer.

Conclusions and future research direction

All these observations reinforce the problem of understanding the stabilization of chloroplast fragments in the nuclear genome, allowing them to acquire the requisite signals to become functional in the nucleus and then being available to substitute for the endogenous chloroplast gene when it is inactivated, without any direct selection pressure driving the process.

Apart from the question of the transfer of functionality, questions concern the purpose and mechanisms of the transfer of exogenous DNA fragments into and out of the plant nuclear genome. Questions to be asked include: are stress-induced DNA transfers from the chloroplasts (or from other external sources such as pathogens and symbionts) beneficial for plants, allowing them to survive the stress? Are the proposed stress-related transfers of DNA directed or random, i.e. do the stress conditions simply facilitate the acceptance of exogenous DNA fragments or do they direct them to specific regions of the genome? To answer these questions more data need to be collected. In particular, comparison of the regions surrounding the identified chloroplast fragments for polymorphisms will indicate the frequency of insertion into (and deletion from) specific sites.

Further, do all stresses facilitate or mandate the integration of exogenous DNA fragments, or is either the intensity or nature of the stress a determining factor? How is the integration of these fragments of chloroplast related to other stress responses? Is the suite of stress responses at the level of gene expression exhausted before DNA transfer occurs, i.e. is the destabilization of the genome a last resort or part of the normal stress response? Are fragments of chloroplast DNA available for integration at all stages of vegetative growth, or only during meiosis, unless otherwise produced such as in the experiments involving irradiation? Many of these questions can be addressed as more genomes are completely sequenced especially with respect to multiple ecotypes and varieties of the same species. An understanding of the origin and persistence of these fragments of

organellar DNA may enhance the understanding of the complexity of adaptations to stress, namely, are these types of genomic modifications an integral part of the mechanisms by which plants do adapt to stress environments?

Footnotes

†These two authors contributed equally to the manuscript.

Literature cited

Andersson SG, Karlberg O, Canback B, Kurland CG. On the origin of mitochondria: a genomics perspective. Philosophical Transactions of the Royal Society London. B Biological Science (2003) 358:165–177.

Ankele E, Heberle-Bors E, Pfosser MF, Hofinger BJ. Searching for mechanisms leading to albino plant formation in cereals. Acta Physiologiae Plantarum (2005) 27:651–665.

Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature (2000) 408:796–815.

Atabekov JG, Dorokhov YL. Plant virus-specific transport function and resistance of plants to viruses. Advances in Virus Research (1984) 29:313–364.

Ayliffe MF, Timmis JN. Plastid DNA sequence homologies in the tobacco nuclear genome. Molecular General Genetics (1992) 236:105–112.

Bausher MG, Singh ND, Lee S-B, Jansen RK, Daniell H. The complete genome sequence of *Citrus sinensis* (L.) Osbeck var. 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. BMC Plant Biology (2006) 6:21.

Bennetzen J, Kellogg EA. Do plants have a one-way ticket to genomic obesity? The Plant Cell (1997) 9:1509–1514.

Chumley TW, Palmer JD, Mower JP, Fourade HM, Calie PJ, Boore JL, et al. The complete genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Molecular Biology Evolution (2006) 23:2175–2190.

Chung H-J, Jung JD, Park H-W, Kim J-H, Cha HW, Min SR, et al. The complete chloroplast genome of *Solanum tuberosum* and the comparative analysis with Solanaceae species identified the presence of a 241 bp deletion in cultivated potato chloroplast DNA sequence. Plant Cell Reports (2006) 25:1369–1379.

Cullis CA. Mechanisms and control of rapid genomic changes in flax. Annals of Botany (2005) 95:201–206.

Daniell H, Lee S-B, Grevich J, Saki C, Quesada-Vargas T, Guda C, et al. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analysis with other Solanaceae genomes. Theoretical Applied Genetics (2006) 112:1503–1518.

Day A, Ellis THN. Deleted forms of plastid DNA in albino plants from cereal anther culture. Current Genetics (1985) 9:671–678.

Farrelly F, Butow RA. Rearranged mitochondrial genes in the yeast nuclear genome. Nature (1983) 301:296–301.

Gilbertson RL, Sudarshana M, Jiang H, Rojas MR, Lucas WJ. Limitations on geminivirus genome size imposed by plasmodesmata and virus-encoded movement protein: insights into DNA trafficking. The Plant Cell (2003) 15:2578–2591.

Herrmann RG. Eukaryotism, towards a new interpretation. Eukaryotism and symbiosis— Schenk HEA, Herrmann RG, Jeon KW, Muller NE, Schwemmler W, eds. (1997) Springer: Vienna. 73–118.

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, et al. The complete sequence of the rice (Oryza sativa) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Molecular and General Genetics (1989) 217:185–194.

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

International Rice Genome Sequencing Project. The map-based sequence of the rice genome. Nature (2005) 436:793–800.

Jackson D. Double labeling of KNOTTED1 mRNA and protein reveals multiple potential sites of protein trafficking in the shoot apex. Plant Physiology (2002) 129:1423–1429.

Jackson D, Hake S. Morphogenesis on the move: cell-to-cell trafficking of plant regulatory proteins. Current Opinion in Genetics and Development (1997) 7:495–500.

Kato T, Kaneko T, Sato S, Nakumura Y, Tabata S. Complete structure of the chloroplast genome of a legume, *Lotus japonicus.* DNA Research (2000) 7:323–330.

Kim JY, Yuan Z, Jackson D. Developmental regulation and significance of KNOX protein trafficking in *Arabidopsis.* Development (2003) 130:4351–4362.

Kim JY, Rim Y, Wang L, Jackson D. A novel cell-to-cell trafficking assay indicates that the KNOX homeodomain is necessary and sufficient for intercellular protein and mRNA trafficking. Genes and Development (2005) 19:788–793.

Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K. The complete nucleotide sequence of the hornwort (*Anthoceros formosae)* chloroplast genome: insight into the earliest land plants. Nucleic Acids Research (2003) 31:716–721.

Lee S-B, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. The complete chloroplast genome sequence of *Gossypium hirsutum*: organization and phylogenetic relationships to other angiosperms. BMC Genomics (2006) 7:61.

Lexer C, Fay MF. Adaptation to environmental stress: a rare or frequent driver of speciation? Journal of Evolutionary Biology (2005) 18:893–900.

Linacero R, Alves EF, Vázquez AM. Hot spots of DNA instability revealed through the study of somaclonal variation in rye. Theoretical and Applied Genetics (2000) 100:506–511.

Lisitsyn N. Representational difference analysis: finding the differences between genomes. Trends in Genetics (1995) 11:303–307.

Lucas WJ. Plasmodesmata-intercellular channels for macromolecular transport in plants. Current Opinion in Cell Biology (1995) 7:673–680.

Lucas WJ. Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes. Virology (2006) 344:169–184.

Lukens LN, Zhan S. The plant genome's methylation status and response to stress: implications for plant improvement. Current Opinion in Plant Biology (2007) 10:317–322.

Madlung A, Comai L. The effect of stress on genome regulation and structure. Annals of Botany (2004) 94:481–495.

Maier RM, Neckermann K, Igloi GL, Kossel H. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. Journal Molecular Biology (1995) 251:614–628.

Martin W. Gene transfer from the organelles to the nucleus: frequent and in big chunks. Proceedings of the National Academy of Sciences of the USA (2003) 100:8612–8614.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. 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 USA (2002) 99:12246–12251.

Matsuo M, Ito Y, Yamauchi R, Obokata J. The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear genome DNA flux. The Plant Cell (2005) 17:665–675.

Noueiry AO, Lucas WJ, Gilbertson RL. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. Cell (1994) 76:925–932.

Pitzscheke A, Forzani C, Heribert H. Reactive oxygen species signaling in plants. Antioxidants & Redox Signalling (2006) 8:1757–1764.

Richly E, Leister D. An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. Gene (2004) 329:1–16.

Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. The complete chloroplast genome of the Chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. Molecular Biology Evolution (2007) 24:54–62.

Ruhlman T, Lee S-B, Jansen RK, Hostetler JB, Tallon LJ, Town CD, et al. Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. BMC Genomics (2006) 7:222.

Saki C, Lee S-B, Daniell H, Wood TC, Tomkins J, Kim H-G, Jansen RK. Complete chloroplast genome sequence of Glycine max and comparative analysis with other legume genomes. Plant Molecular Biology (2005) 59:309–322.

Saki C, Lee S-B, Fjellheim S, Guda C, Jansen RK, Lou H, et al. Compete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analysis with other grass genomes. Theoretical and Applied Genetics (2007) 115:571–590.

Samson N, Bausher MG, Lee S-B, Jansen RK, Daniell H. The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications of biotechnology and the phylogenetic relationship amongst angiosperms. Plant Biotechnology Journal (2007) 5:339–353.

Sanderfoot AA, Lazarowitz SG. Cooperation in viral movement—the geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. The Plant Cell (1995) 7:1185–1194.

Sanderfoot AA, Lazarowitz SG. Getting it together in plant virus movement: cooperative interactions between bipartite geminivirus movement proteins. Trends in Cell Biology (1996) 6:353–358.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana.* DNA Research (1999) 6:283–290.

Scott NS, Timmis JN. Homologies between nuclear and plastid DNA in spinach. Theoretical and Applied Genetics (1984) 67:279–288.

Shahmuradov IA, Akbarova YY, Solovyev VV, Aliye JA. Abundance of plastid DNA insertions in nuclear genomes of rice and Arabidopsis. Plant Molecular Biology (2003) 52:923–934.

Steane DA. Complete nucleotide sequence of the chloroplast genome from the tasmainian blue gum, *Eucalyptus globules* (Myrtaceae). DNA Research (2005) 12:215–220.

Stegemann S, Hartmann S, Ruf S, Bock R. High-frequency gene transfer from the chloroplast to the nucleus. Proceedings of the National Academy of Sciences of the USA (2003) 100:8828–8833.

Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and the relocation of *rpoA* from the chloroplast to the nucleus. Nucleic Acids Research (2003) 31:5324–5331.

Sun CW, Callis J. Recent stable insertion of mitochondrial DNA into and Arabidopsis polyubiquitin gene by nonhomologous recombination. The Plant Cell (1993) 5:97–107.

Thorsness M, Fox TD. Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondria to the nucleus. Genetics (1990) 134:21–28.

Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organellar genomes forge eukaryotic chromosomes. Nature Reviews Genetics (2004) 5:123–135.

Vorster J. Comparative whole genome analysis detects similarities between plants and microbes. (2008) Pretoria, South Africa: University of Pretoria. PhD Thesis.
van der Vyver C. Stress-induced genome changes in plants (2003) Pretoria, South Africa: University of Pretoria. PhD Thesis.

Woodson JD, Chory J. Coordination of gene expression between organellar and nuclear genomes. Nature Reviews Genetics (2008) 9:383–395.