Chloroplast DNAs Diversify Nuclear and Mitochondrial Genomes in Plants

Dong Wang

A thesis submitted for the degree of Doctor of Philosophy

Discipline of Genetics

School of Molecular and Biomedical Science

The University of Adelaide

March 2013

Table of Contents

Table of Contents I

Abstract II

Declaration IV

Acknowledgements VI

Chapter1 Introduction 1

Chapter 2 Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants 12

Chapter 3 Single Molecule PCR Reveals Similar Patterns of Non-Homologous DSB Repair in Tobacco and Arabidopsis 14

Chapter 4 Cytoplasmic Organelle DNA Preferentially Inserts into Open Chromatin Regions 15

Chapter 5 Plastid Sequences Contribute to Some Plant Mitochondrial Genes 23

Chapter 6 Conclusions and Future Directions 24

Reference 27

Abstract

Mitochondria and chloroplasts in eukaryotic cells originated more than a billion years ago when an ancestor of the nucleated cell engulfed two different prokaryotes in separate sequential events. Extant cytoplasmic organellar genomes contain very few genes compared with their candidate free-living ancestors, as most have functionally relocated to the nucleus. The first step in functional relocation involves the integration of cytoplasmic organllar DNA fragments into nuclear chromosomes and this process continues at high frequency with attendant genetic, genomic and evolutionary consequences. The frequency of DNA transposition from plastid (chloroplast) to nucleus has been measured experimentally in tobacco plants (*Nicotiana tabacum*) growing in ideal growth conditions.

To monitor the effects of environmental stress on the rate of DNA transfer from plastid to nucleus, two different transplastomic tobacco lines were used and it was shown that DNA migration from chloroplasts to the nucleus was markedly increased by mild heat stress. In addition, manually induced DNA double-strand breaks (DSBs) were made using the rare-cutting endonuclease I-SceI in tobacco and Arabidopsis and this system was used to investigate the role of DSBs repair during organellar DNA insertion into the nuclear genome. Integrants of none organelle DNA origin were found at the break points when plants grown at normal temperature. In contrast, insertions of mitochondrial DNA fragments occurred during the repair of induced DSBs were only observed in tobacco when plants were heat treated. This finding suggested that the frequency of mitochondrial DNA migration was also increased by mild heat stress.

To further investigate whether the DSB repair is involved in plastid DNA integration into the nuclear genome, 14 nuclear insertions of chloroplast DNA (*nupts*) that are unique to *Oryza sativa* subsp. *indica* were indentified. Comparisons with the nuclear pre-insertion loci (identified in the related subspecies, *O. sativa* subsp. *Japonica* which lacked these *nupts*) indicated that chloroplast DNA had integrated by non-homologous end joining. Combined with analyzing available DNase-seq data, this analysis also revealed that *nupts* were significantly more

frequent in open chromatin regions of the nucleus. The generality of this insertion site preference was tested in the chimpanzee genome by comparing nuclear loci containing integrants of mitochondrial DNA (*numts*) with *numt*-lacking preinsertion sites in the human genome. Mitochondrial DNAs also tended to insert more frequently into regions of open chromatin revealed by human DNase-seq and FAIRE-seq databases.

Chloroplast DNA movement is not limited to the nucleus and it is also found within the mitochondrial genome in most plants. However, the functions of these plastid-derived DNA tracts in mitochondrial genomes (also called *mtpt* for *mit*ochondrial *p*las*t*id DNA) have been considered to be limited to rare instances where plastid tRNA genes have replaced their mitochondrial counterparts, where short patches of mitochondrial genes evolved using their homologous plastidic copies by gene conversion, or where a new promoter region is created. In this thesis it is demonstrated that some *mtpts* contribute codons to unrelated mitochondrial protein-coding sequences and others may have a role in post-transcriptional RNA processing.

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 Dong Wang 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, 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

Wang, D., Lloyd, A.H. and Timmis, J.N. (2012) Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc Natl Acad Sci USA* **109**, 2444-244.
Copyright: © 2012 National Academy of Sciences

Wang, D., Lloyd, A.H. and Timmis, J.N. (2012) Nuclear genome diversity in somatic cells is accelerated by environmental stress. *Plant Signal Behav* **7**, 595-597.
Copyright: © 2012 Landes Bioscience

Lloyd, A.H., Wang, D. and Timmis, J.N. (2012) Single molecule PCR reveals similar patterns of non-homologous DSB repair in tobacco and Arabidopsis. *PLoS One*, **7**, e32255.
Copyright: © 2012 Lloyd et al.

Wang, D., Rousseau-Gueutin, M. and Timmis, J.N. (2012) Plastid sequences contribute to some plant mitochondrial genes. *Mol Biol Evol* **29**, 1707-1711.
Copyright: © The Author 2012

Acknowledgements

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

Jeremy Timmis, for supervision and guidance, and for saving my research career. I am so lucky to have such a great supervisor. Jeremy, the knowledge that I learnt from you is not only just valuable for my PhD, and will benefit for my whole academic career.

Dave Adelson and Jack Da Silva, my co-supervisors, for providing me the server and teaching me statistics during my PhD.

Joan Kelly, for providing me space to finish my last experiment and thesis writing.

Members of the Timmis lab, past and present, for making it such a supportive and enjoyable place to work. Andrew, Mat, Rory, Yuan, and Zhipeng from the Adelson lab for helpful discussion to generate interesting idea and a lot of fun.

Sim Lim, Hani and everyone else in the Genetics Discipline and MLS building who has helped me along this journey.

My parents, for their endless love and encouragement throughout my life.

I would also like to specially thank the China Scholarship Council (CSC) and The University of Adelaide, who provided me this scholarship to support my PhD.

Chapter1 Introduction

After an α-proteobacterium was captured by an ancestor of the nucleated eukaryotic host, a second separate event followed which captured a cyanobacterium that gave rise to photosynthetic eukaryotes. These engulfed prokaryotes evolved through endosymbiosis into mitochondria (Gray et al., 1999) and chloroplasts (Rodriguez-Ezpeleta et al., 2005), respectively. Following the engulfment of the two free-living prokaryotes, there were extensive rearrangements and quantitative changes within the genomes of both host and symbionts. These changes involved large-scale translocation of genes and DNA fragments from the endosymbiont's genome to the nuclear genome of the host cell (Blanchard and Schmidt, 1995;Martin et al., 1998). Consequently, compared with their free-living ancestors, contemporary organellar genomes are greatly decreased in size, containing only a small proportion of the genes that their free-living ancestors possessed. Different present-day species show some variation in the genetic complexity of their mitochondrial genomes, varying in coding capacity between 3 and 67 proteins (Gray et al., 1999;Lang et al., 1999;Timmis et al., 2004). Likewise the chloroplast genomes of different species encode between 15 and 209 proteins (Kaneko et al., 1996;Timmis et al., 2004;Keeling and Palmer, 2008). However, the function and biogenesis of the cytoplasmic organelles each require approximately the same number of proteins as the candidate free-living prokaryotic relatives of their ancestors. This can be explained by the observation that cytoplasmic products of nuclear genes are imported into the cytoplasmic organelles (Martin and Herrmann, 1998). These sets of nuclear genes that function in mitochondrial and chloroplast biogenesis derive predominantly from DNA transferred from the prokaryote endosymbionts, a process that has proceeded over long evolutionary time. However, phylogenetic (Adams et al., 2000;Millen et al., 2001;Adams et al., 2002) and experimental (Thorsness and Fox, 1990; 1993;Campbell and Thorsness, 1998;Huang et al., 2003;Stegemann et al., 2003) analyses suggest that DNA transfer is ubiquitous in eukaryotes and, in some species functional gene transfer, is still ongoing.

Relocate, or retain, that is the question?

Why do a lot of organelle genes relocate to the nuclear genome, while some organellar genes hold on in the organelle? Several explanations have been posited for functional gene transfer from cytoplasmic organelle-to-nucleus. Firstly, DNA in the highly energetic cytoplasmic organellar genetic compartments of eukaryotes is subjected to high levels of stress damage from the oxygen free radicals produced during respiration and photosynthesis. Also the asexual, usually polyhaploid, organellar genomic environment may accumulate deleterious mutations (Muller, 1964). Thus, organllar gene migration to the nucleus may be beneficial to by providing sexual recombination and DNA repair. This not only avoids free-radical-induced mutation (Allen and Raven, 1996) but also avoids genetic decay *via* Muller's ratchet (Andersson and Kurland, 1998;Martin and Herrmann, 1998). Secondly, smaller organellar genomes may be preferred (Selosse et al., 2001) and selection for rapid replication may be a significant cause of the reduction of cytoplasmic organellar genomes. The genome of nucleus is less constrained, as shown by the presence of non-coding DNA and the occurrence of massive nuclear genomes (Beaton and Cavalier-Smitht, 1999). Thirdly, gene transfer is more frequent from a multicopy genome to a single copy genome than reverse (Selosse et al., 2001). Thus the existence of multiple organelles favours the transfer from organelle-to-nucleus, whereas the unique nucleus precludes the transposition in the reverse direction.

On the other hand, there are also some hypotheses that may account for the observation that some genes still reside in the organelle. One suggests that some proteins, for example hydrophobic proteins, are difficult to import into organelles and must be encoded by organellar genome (Adams and Palmer, 2003). The other is the co-location for redox regulation (CORR) hypothesis in which co-located genes and gene products permit direct redox regulation of gene expression that may be applied to individual cytoplasmic organelles (Allen, 2003).

The fate of post-transfer organelle gene

After organellar genes move from a prokaryotic environment to the eukaryotic host nucleus, they usually cannot immediately become functional. Acquisition of regulatory elements is required to activate the organelle gene in the nucleus (Stegemann and Bock,

2006;Lloyd and Timmis, 2011), and a target peptide is generally needed if the product is required in the cytoplasmic organelle. Interestingly, a prokaryotic promoter has been observed to work weakly in the nucleus (Cornelissen and Vandewiele, 1989), and some plant mitochondrial genes already contain the required targeting information before they transfer to nucleus (Ueda et al., 2008). These findings suggest that it is not difficult for an organelle gene to become functional in the nuclear environment. However a new problem immediately arises for the organism: how to balance the two functional gene copies that are located in different genetic compartments; which should be maintained - the nuclear copy, or the organellar version? If the nuclear copy is lost, gene transfer from organelle to nucleus can be repeated but if the organellar copy is lost, then the gene is permanently established in the nucleus (Doolittle, 1998).

Genes of novel function also originate by endosymbiotic DNA transfer

It has been estimated that approximately 18% (4,500) of *Arabidopsis thaliana* nuclear protein-coding genes (Martin et al., 2002) and 10.8% (132) of *Cyanophora paradoxa* annotated nuclear gene sets (Reyes-Prieto et al., 2006) were acquired from the cyanobacterial ancestor of the plastid. Endosymbiotic gene transfer also provides a way to create new genes in addition to creating nuclear versions of cytoplasmic organellar genes. Thus, while many transferred genes maintain chloroplast biogenesis and function, others encode products that are targeted to cell compartments other than the chloroplast (Martin et al., 2002). Although the majority of organellar DNA transposition is considered to result in nonfunctional nuclear sequences, functional nuclear exons are generated by insertions of organellar DNA (Noutsos et al., 2007). For example, the Arabidopsis *At2g28820* gene, encoding a protein with similarity to alanine aminotransferases (AlaATs) contains a plastid DNA insert in the nuclear DNA terminal exon. This gene contains a 3' terminal domain derived from the plastid-encoded gene encoding ribosomal protein L16 (PRPL16). The level and spectrum of sequence diversification of the *At2g28820* nuclear integrant of organelle DNA (*norg*) from its organellar *PRPL16* template indicates that natural selection has reshaped the protein sequence to create a novel biochemical function. Compared with the corresponding plastid sequence, the nuclear derivative contains three indels (two 3-bp deletions and one 3-bp insertion) and a total of 60 nucleotide substitutions. This case, and others,

suggest that nuclear integrants of mitochondrial DNAs (*numts*) and nuclear integrants of plastid DNAs (*nupts*) represent a major source of nuclear chromosomal variation and genomic diversity (Lough et al., 2008).

Experimental analyses of DNA transfer from cytoplasmic organelle to nucleus

The first instance of organelle-to-nucleus DNA transfer utilised a yeast mitochondrial plasmid containing a nucleus-specific gene *URA3*, a mitochondrion-specific gene *COX2* and the yeast 2μ origin of replication. After insertion of this cassette into mitochondria, the *URA3* gene was demonstrated to migrate to the nucleus in *Saccharomyces cerevisiae* (Thorsness and Fox, 1990). A strain lacking *URA3* that was auxotrophic for uracil was converted to Ura^{+} prototrophy through movement of this plasmid from mitochondria to nucleus. The frequency of DNA escape from mitochondria and entry into the nucleus was surprisingly high: approximately 2×10^{-5} per cell per generation (Thorsness and Fox, 1990). DNA movement in the other direction, i.e. transfer of DNA from nucleus to mitochondria was not inspected. Later, a nuclear DNA fragment including the *TRP1* gene was shown to escape from mitochondria and insert into the nuclear genome in yeast (Thorsness and Fox, 1993). Under these selection conditions, the frequency of DNA escape from mitochondria and the subsequent migration to the nucleus was at least one in 5×10^{6} cells.

To determine the frequency of DNA transfer from chloroplast (plastid) to nucleus in a higher plant, a nucleus-specific neomycin phosphotransferase gene (*neoSTLS2*) was experimentally integrated into the tobacco chloroplast genome and the progeny of transplastomic plants were screened for kanamycin resistance (Huang et al., 2003). Approximately ~250,000 progeny produced by fertilization of wild-type females with pollen from homotransplastomic plants containing chloroplast-located *neoSTLS2* were screened for kanamycin resistance. Sixteen plants contained the *neoSTLS2* marker integrated into a nuclear chromosome. This equated to a chloroplast-to-nucleus DNA transfer frequency of one in 16,000 male gametes tested (Huang et al., 2003). In contrast, the frequency of transplastomic *neoSTLS2* transfer in the female nuclear germline of tobacco was very much lower: only one transposition was identified in a

screen of 273,000 transplastomic ovules (Sheppard et al., 2008). A similar strategy was used to detect chloroplast-to-nucleus transposition in somatic cells (Stegemann et al., 2003). One transposition event was estimated to occur in about 5 million somatic cells of leaf explants in culture (Stegemann et al., 2003). Therefore, taken at face value, the frequency of plastid-to-nucleus transposition in pollen is much higher than that in ovules or somatic cells. There is ample evidence that both loss of plastids and degradation of plastid DNA occur during pollen grain development, particularly in species which show maternal inheritance of plastid genes (including tobacco) (Mogensen, 1996). This process may release fragmented plastid DNA to the cytoplasm, enabling it to transfect the nucleus. It is probable therefore, that the programmed degeneration of plastids occurring during pollen development explains the higher rate of plastid-to-nucleus transfer through the male germline (Timmis et al., 2004). In the majority of these experiments the size of the de novo *nupts* was very large such that the *aadA* gene used as a selectable marker for chloroplast transformation was nearly always co-transferred to the nucleus along with *neoSTLS2*, but it remained inactive in the new location. However, the *aadA* gene was shown to be activated by acquiring a eukaryotic promoter at a frequency that is measurable under experimental conditions (Stegemann and Bock, 2006;Lloyd and Timmis, 2011). However, neither of these scenarios required the experimental reporter gene to acquire a transit peptide-encoding sequence. Therefore, while DNA transfer from chloroplast to nucleus is a very frequent event, functional transfer, that recapitulates endosymbiotic relocation of prokaryotic genes to the nucleus, is very much rarer.

The unicellular green alga *Chlamydomonas reinhardtii* has also been used to screen for chloroplast DNA transfer to the nucleus (Lister et al., 2003). A selectable marker gene (*ble*, conferring nucleus-specific zeomycin resistance) was integrated into chloroplast genome. Extensive antibiotic screens of several billion haploid cells revealed no evidence of stable nuclear integration of chloroplast DNA. This finding implies that the frequency of chloroplast-to-nucleus transfer is much lower in *C. reinhardtii* than tobacco and is consistent with the finding of very few *nupts C. reinhardtii* (Richly and Leister, 2004). One explanation for absence of experimental identification of chloroplast-to-nucleus transposition in *C. reinhardtii* is the 'limited transfer window' hypothesis, which posits that cells with multiple organelles per cell are more likely to show endosymbiotic gene transfer than those with a single organelle like *C. reinhardtii*

which contains only one copy of chloroplast. This is because lysis of a single organelle to release the DNA necessary for nuclear transfection into the cytoplasm would result in the catastrophic loss of the cellular functions of the plastid (Barbrook et al., 2006;Smith et al., 2011).

Mechanisms of organelle DNA transfer to the nucleus

Direct DNA or RNA intermediates?

It has been suggested that organelle-to-nucleus gene transfer involves reverse transcription of an RNA intermediate. This was based, for example, on the observation that the nuclear cox*II* sequence in some flowering plants more closely resembles mitochondrial cox*II* transcripts that had been edited inside the mitochondrion, than the gene itself (Nugent and Palmer, 1991). There is also phylogenetic evidence supporting RNA-mediated mitochondrial DNA transposition on an evolutionary timescale (Covello and Gray, 1992;Wischmann and Schuster, 1995). The mitochondrial cox*II* gene is silenced in soybean and the nuclear copy appears to have been generated through a C-to-U edited RNA intermediate (Covello and Gray, 1992). In Arabidopsis, *rps10* transfers to nucleus also via the C-to-U edited RNA intermediate and a typical organellar group II intron has been removed as well (Wischmann and Schuster, 1995). However, other interpretations of these observations are possible. For instance, direct DNA transfer and integration may have occurred prior to the evolution cytoplasmic organelle-specific RNA editing and before introns were present in mitochondrial DNA (Henze and Martin, 2001). Also, when cDNA intermediates of spliced and edited higher plant mitochondrial transcripts occur, they are likely to first arise inside the mitochondrion. Thus, assuming homologous recombination is active in mitochondria, it appears probable that cDNA intermediates will recombine with mitochondrial DNA rather than nuclear DNA, consequently erasing editing sites and removing introns in the mitochondrial genome (Henze and Martin, 2001). In addition, the frequent C to T mutations (and the complement: G to A) (Huang et al., 2005), that characterize both *nupts* and *numts*, mimic organellar mRNA editing.

Bioinformatic and experimental analysis in yeast and other eukaryotes have demonstrated evidence of recombination between escaped organelle DNA molecules and nuclear chromosomes (Thorsness and Weber, 1996;Timmis et al., 2004;Kleine et al., 2009). By comparisons between cytoplasmic organelle DNA and nuclear DNA of the same species, intergenic spacers and other non-coding regions of organelle DNA are found in nuclear-transferred copies (*numts* and *nupts*) as often as coding sequences (Timmis et al., 2004;Bock and Timmis, 2008) . Furthermore, analysis of human *numts* found that up to 30% of chromosomal *numt* content is composed of fragments that encompass two or more adjacent mitochondrial genes. Thus, there is no correlation between the abundance of mitochondrial transcripts and the multiplicity of *numt* integration, indicating that the migrations of mitochondrial DNA sequences to the nucleus are predominantly DNA mediated (Woischnik and Moraes, 2002). However, it is still formally possible that real endosymbiotic gene relocation does not start with *numt* or *nupt* formation, even though RNA-mediated transfer frequency is too low to measure experimentally.

To detect RNA/cDNA-mediated *nupt* transposition, two independent experiments were carried out in tobacco (Sheppard et al., 2011;Fuentes et al., 2012). One study introduced into the plastome a nuclear promoter-driven *neoSTLS2* gene containing several RNA editing sites including one that was located in the start codon (Sheppard et al., 2011). The other constructed a chloroplast-located reporter gene containing a chloroplast group II intron in antisense but otherwise designed to be nucleus specific (Fuentes et al., 2012). Theoretically, both these chloroplast-located reporter genes could be active and provide the kanamycin resistance to plants only after their transfer to nuclear genome in the form of RNA/cDNA. However, no antibiotic resistant seedlings containing the edited reporter gene were found using either experimental approach after screening hundreds of thousands of seeds. However, kanamycin resistant plants did emerge from both screens. Among these, the reporter gene was present in the form of DNA with unchanged RNA editing sites or the group II intron was still present in nuclear genome. Therefore the selectable marker genes, though designed to be disabled without chloroplast-specific editing and processing, were able to be expressed directly to give kanamycin resistance in the nucleus. These examples suggest that DNA-mediated chloroplast-to-nucleus gene transfer is much more common than RNA/cDNA-mediated events and the latter remain experimentally undetected.

How does organellar DNA integrate into the nuclear genome?

DNA double-strand break (DSB) repair was shown to be implicated in the integration of mitochondrial DNA into the nuclear genome in yeast (Ricchetti et al., 1999). Microhomology (2-5 bp) and longer matching sequences, together with rearrangements of plastid and nuclear DNA were found frequently near *de novo* integration sites in tobacco (Ayliffe and Timmis, 1992;Huang et al., 2004;Lloyd and Timmis, 2011), suggesting that in tobacco, as in yeast (Ricchetti et al., 1999), organelle DNA regularly integrates into the nuclear genome by nonhomologous end-joining (NHEJ) of DSBs. DSBs may be caused by environmental factors including ionizing radiation and chemical DNA-damaging agents and endogenously by reactive oxygen species, endonucleases and topoisomerases (Bleuyard et al., 2006). Repair by NHEJ – which does not require extensive homologous sequence similarity - can enable any DNA, if it is available at the time, to integrate into the nuclear genome (Ricchetti et al., 1999;van Gent et al., 2001). Chan et al (2007) showed that ionizing radiation increased the frequency of microhomology-mediated DNA integration in yeast. Irradiated yeast cells displayed 77% microhomology-mediated integration, compared with 27% in nonirradiated cells.

Given the deluge of cytoplasmic organellar DNA that constantly integrates into the nuclear genome, an ever increasing nuclear DNA content might be expected that clearly does not occur. How then does the nuclear size remain relatively stable? In general, there are deletions in different region of genome during or after the organellar DNA integration. As DSB repair is known to be a potent source of deletions (Salomon and Puchta, 1998), it may also be responsible for maintaining an equilibrium between ingress and elimination of organellar DNA fragments (Figure 1). Normally, DSB repair may balance nuclear genome size and nuclear genome rearrangement, because deletion-associated DSB repairs are frequently observed in plant and mammalian genomes (Wessler et al., 1990;Kirik et al., 2000;Lin and Waldman, 2001b;a;Miller et al., 2004;Hazkani-Covo and Covo, 2008). Additionally, bioinformatic analysis has shown that deletion of nuclear organelle DNA integrants is associated with direct repeats, indicating that DNA replication slippage and selection are able counterbalance insertion of *norg* DNAs (Noutsos et al., 2005). Recombination between homologous chromosomes may also play a role in nuclear DNA deletion events that would maintain

nuclear genome size. Similarly non-allelic homologous recombination (NAHR) can lead to DNA deletions (Stankiewicz and Lupski, 2002).

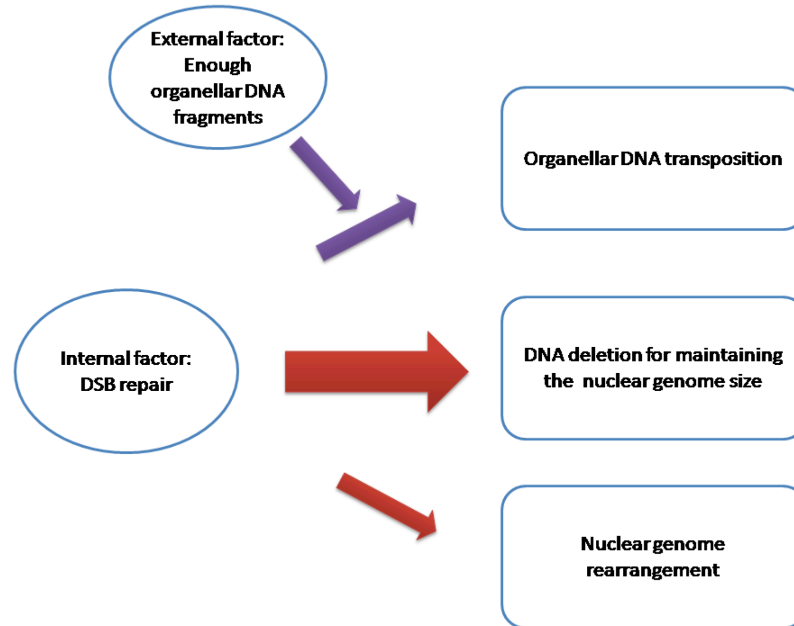

Figure 1. The role of DSB repair in organellar DNA transposition

The migration of chloroplast DNA to mitochondria

In higher plants, plastid-derived DNA constitutes between 1% and 10% of the mitochondrial genomes (Wang et al., 2007). This mitochondrial plastid DNA is called *mtpt*. Their contributions are considered to be limited to rare instances where plastid tRNA genes have substituted for their mitochondrial equivalents (Kanno et al., 1997;Miyata et al., 1998), where short patches of mitochondrial genes have evolved through their homologous plastidic counterparts by gene conversion (Hao and Palmer, 2009;Sloan et al., 2010) or where a new promoter region is generated (Nakazono et al., 1996). Thus similar sequences that appear in more than one genetic compartment were first designated "promiscuous DNA" by Ellis (1982), prompted by the finding of mitochondrial DNA resembling 12 kb of chloroplast DNA in maize (Stern and Lonsdale, 1982). Many other chloroplast-derived DNA sequences also have been found in the mitochondrial DNA of many different plant species (Smith, 2011). These

chloroplast-derived fragments vary in size from just a few recognisable nucleotides to several kilobases which may contain active genes coding for mitochondrial tRNAs (Fauron et al., 1995;Unseld et al., 1997). These chloroplast-derived mitochondrial tRNA genes are recognized on the basis that they still retain high sequence similarity with their chloroplast xenologues (Sangare et al., 1989) and some are necessary for mitochondrial translation (Marechal-Drouard et al., 1995;Miyata et al., 1998). Northern-blot hybridization revealed that plastid-derived *trnC*, *trnF*, *trnH*, *trnM*, *trnN*, *trnS* and *trnW* genes are transcribed and are precisely processed to mature tRNAs in rice (Miyata et al., 1998) suggesting that these plastid-derived tRNAs are utilized for the biosynthesis of protein in plant mitochondria. Researchers have found several distinct cases of transposition of chloroplast *rbcL* genes into mitochondria but all of the mitochondrial copies contain insertions and/or deletions that disrupt the reading frame. The mitochondrial paralogues show increased rates of nonsynonymous substitutions, providing clear evidence that they are pseudogenes (Cummings et al., 2003).

Although chloroplast transformation is possible in a small number of additional species, none of the latter has the essential abundant seed production and ease of tissue culture available in tobacco. Thus, this work is carried out mainly in tobacco by using nuclear specific reporter genes.

Aims

The general aim of this project is to investigate organelle DNA transposition. Both chloroplast-to-nucleus DNA migration and transposition from chloroplast to mitochondrion are studied.

Compared with nuclear DNA, cytoplasmic organellar DNA is arguably more susceptible to stress conditions because reactive oxygen species (ROS) are generated in mitochondria and chloroplasts as they experience frequent photooxidative stress. A possible benefit of translocation of genes from the chloroplast to the nucleus is that genes are moved away from a location with a high redox-load (Timmis et al., 2004). Thus, 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 phenomenon that is undirected. There is also some evidence and some tenuous speculation that stress has effects on plant genomes (Cullis et al., 2009;Li, 2009). For example, light apparently influences nuclear DNA content of *Helianthus annuus* (Price and Johnston, 1996). At the fixed red to far red ratio, low photon flux density increases DNA content of *Helianthus annuus*; otherwise, high photon flux density reduces the DNA content. It has not been conclusively proven that there is any effect of stress on DNA transfer from organelle to nucleus. Thorsness and Fox (Thorsness and Fox, 1990) found that higher or lower temperature increased the DNA migration from mitochondria to nucleus in yeast. Moreover, ionizing radiation increases the frequency of microhomology-mediated DNA integration in yeast (Chan et al., 2007), and NHEJ is proved to involve the mitochondrial DNA insert into nuclear genome through microhomology in yeast (Ricchetti et al., 1999). However, it remains unclear whether similar stress effects are experienced in higher organisms like angiosperms.

To explore the environmental effects on organellar DNA transfer to the nucleus, two transplastomic tobacco lines were employed for chapter 2. One line contains a plastome with a specific antibiotic resistant gene *neo* (Huang et al., 2003) that allows detection of transposition by kanamycin selection. The other (Sheppard et al., 2008) harbours a nucleus specific *gus* marker that allows detection of transposition by colorimetric assays.

Chapter 3 describes the impact of DSBs on chloroplast DNA migration to nucleus. Artificial DSBs were generated by introducing an I-SceI site and a gene encoding the endonuclease I-SceI into tobacco and Arabidopsis nuclear DNA.

Chapter 4 uses a bioinformatic approach to investigate the chromatin state of the preinsertion sites for *norg*s in rice and human nuclear DNA.

Chapter 5 investigates whether plastid DNA contributes to mitochondrial protein-coding sequence.

Chapter 2

Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants

Dong Wang, Andrew H. Lloyd and Jeremy N. Timmis

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

Proceedings of the National Academy of Sciences USA, February 2012, Vol. 109, pp. 2444-2448

STATEMENT OF AUTHORSHIP

Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants

Proceedings of the National Academy of Sciences USA, February 2012, Vol. 109, pp. 2444-2448

Dong Wang (Candidate)

Performed experiments, analysed results and wrote the manuscript.

I hereby certify that the statement of contribution is accurate

Signed. .... *Date*...6/3/2013...

Jeremy N. Timmis

Supervised development of work and assisted in analysing results and 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*.6th March 2013

Andrew H. Lloyd

Performed experiments presented in Table 2 and Figure S1, 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*...21st February 2013

Wang, D., Lloyd, A.H. & Timmis, J.N (2012) Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants.
Proceedings of the National Academy of Sciences, v. 109(7), pp. 2444-2448

NOTE:
This publication is included after page 12 in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1073/pnas.1117890109

Nuclear genome diversity in somatic cells is accelerated by environmental stress

Dong Wang, Andrew H. Lloyd and Jeremy N. Timmis

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

Plant Signaling & Behavior, May 2012, 7:5, pp. 595-597

STATEMENT OF AUTHORSHIP

Nuclear genome diversity in somatic cells is accelerated by environmental stress

Plant Signaling & Behavior, May 2012, 7:5, pp. 595-597

Dong Wang (Candidate)

Wrote the manuscript.

I hereby certify that the statement of contribution is accurate

Signed. .... *Date*......6/3/2013.........

Jeremy N. Timmis

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*...6th March 2013......

Andrew H. Lloyd

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*...21st February 2013......

Wang, D., Lloyd, A.H. & Timmis, J.N. (2012) Nuclear genome diversity in somatic cells is accelerated by environmental stress.
Plant Signaling & Behavior, v. 7(5), pp. 595-597

NOTE:
This publication is included after page 13 in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.4161/psb.19871

Chapter 3

Single Molecule PCR Reveals Similar Patterns of Non-Homologous DSB Repair in Tobacco and Arabidopsis

Andrew H. Lloyd, Dong Wang and Jeremy N. Timmis

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

PLoS ONE, February 2012, 7(2): e32255. doi:10.1371/journal.pone.0032255

STATEMENT OF AUTHORSHIP

Single Molecule PCR Reveals Similar Patterns of Non-Homologous DSB Repair in Tobacco and Arabidopsis

PLoS ONE, February 2012, 7(2): e32255. doi:10.1371/journal.pone.0032255

Dong Wang (Candidate)

Performed experiments presented in Figure 3C-D and Table 2, and assisted in writing the manuscript.

I hereby certify that the statement of contribution is accurate

Signed........ .... *Date*......6/3/2013

Jeremy N. Timmis

Supervised development of work and assisted in analysing results and 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*......6th March 2013

Andrew H. Lloyd

Performed experiments, analysed results and wrote 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*...21st February 2013

PLoS one

Single Molecule PCR Reveals Similar Patterns of Non-Homologous DSB Repair in Tobacco and *Arabidopsis*

Andrew H. Lloyd*¤, Dong Wang, Jeremy N. Timmis

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

Abstract

DNA double strand breaks (DSBs) occur constantly in eukaryotes. These potentially lethal DNA lesions are repaired efficiently by two major DSB repair pathways: homologous recombination and non-homologous end joining (NHEJ). We investigated NHEJ in *Arabidopsis thaliana* and tobacco (*Nicotiana tabacum*) by introducing DNA double-strand breaks through inducible expression of I-SceI, followed by amplification of individual repair junction sequences by single-molecule PCR. Using this process over 300 NHEJ repair junctions were analysed in each species. In contrast to previously published variation in DSB repair between *Arabidopsis* and tobacco, the two species displayed similar DSB repair profiles in our experiments. The majority of repair events resulted in no loss of sequence and small (1–20 bp) deletions occurred at a minority (25–45%) of repair junctions. Approximately ~1.5% of the observed repair events contained larger deletions (>20 bp) and a similar percentage contained insertions. Strikingly, insertion events in tobacco were associated with large genomic deletions at the site of the DSB that resulted in increased micro-homology at the sequence junctions suggesting the involvement of a non-classical NHEJ repair pathway. The generation of DSBs through inducible expression of I-SceI, in combination with single molecule PCR, provides an effective and efficient method for analysis of individual repair junctions and will prove a useful tool in the analysis of NHEJ.

Citation: Lloyd AH, Wang D, Timmis JN (2012) Single Molecule PCR Reveals Similar Patterns of Non-Homologous DSB Repair in Tobacco and *Arabidopsis*. PLoS ONE 7(2): e32255. doi:10.1371/journal.pone.0032255

Editor: Luis Herrera-Estrella, Centro de Investigación y de Estudios Avanzados del IPN, Mexico

Received December 12, 2011; **Accepted** January 25, 2012; **Published** February 28, 2012

Copyright: © 2012 Lloyd et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported under the Australian Research Council's (www.arc.gov.au) Discovery Projects funding scheme (project number DP0986973). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: andrew.lloyd@versailles.inra.fr

¤ Current address: Institut Jean-Pierre Bourgin, UMR1318 INRA-AgroParisTech, INRA Centre de Versailles-Grignon, Versailles, France

Introduction

DNA double strand breaks (DSBs) that occur frequently in eukaryotes are potentially lethal to the cell as they lead to mitotically unstable acentric chromosome fragments and the consequent loss of essential genes [1]. In order to deal with these dangerous cellular lesions several DNA repair pathways exist. When a homologous template is available, DNA repair may occur *via* homologous recombination (HR) [2]. During HR any sequence information lost as a result of DNA damage or degradation at the break site, is recovered by using the homologous chromosome or a sister chromatid as template for repair [3]. DNA DSBs may also be repaired without the use of a homologous template by using non-homologous end joining (NHEJ) [1]. In plants this latter pathway appears responsible for the majority of DSB repair [4]. Classical-NHEJ involves the ku70/ku80 heterodimer which binds to DNA ends [5] and recruits a number of other proteins including the DNA ligase IV/XRCC4 complex which repairs the break [6]. The term alternative-NHEJ (alt-NHEJ) is generally used to describe any NHEJ event which lacks one or more of the core classical NHEJ proteins e.g. ku70, ku80, Lig4, XRCC4 [1]. Alt-NHEJ, sometimes referred to as backup-NHEJ (B-NHEJ) [7] or micro-homology-mediated end joining (MMEJ) [8], is not as well characterised and may well include several distinct repair pathways [1]. It has been suggested that alt-NHEJ is inhibited by classical-NHEJ [9,10]. Recently, there has been increased research into NHEJ in mammalian systems, as its importance with regard to cancer treatment has become clear. NHEJ promotes cancer cell survival [11] and inhibitors of NHEJ can be used to increase the sensitivity of tumours to DNA damaging drugs [12] or radiation treatment [13].

While NHEJ research is less advanced in plants there is considerable interest in the process, as it is considered the major pathway for transgene insertion by particle bombardment, *Agrobacterium* and zinc-finger nuclease mediated transformation [14,15,16] and also for the insertion of cytoplasmic organellar DNA [17,18]. A better understanding of this pathway may lead to development of more advanced transformation techniques and manipulation of the pathway may enable efficient gene targeting by HR in higher plants [19].

Despite the large body of work investigating NHEJ there are still several shortcomings in the analysis of this form of DSB repair. A number of these arise from the necessity for tissue culture selection to generate clonal cell lines arising from cells that have undergone individual repair events. The requirement for cell culture restricts analysis to tissues or cell lines able to be cultured efficiently, preventing investigation of NHEJ in some tissues of interest. In addition, the selection and maintenance of multiple cell lines is not only labour intensive and time-consuming, but it also hinders analyses needed to uncover subtle variations in NHEJ repair, and to observe rare classes of repair. These problems require the development of a high-throughput pathway for the analysis of

NHEJ repair events without the need for rounds of tissue culture, selection and plant regeneration.

Over the past decade single molecule (sm) PCR has become a powerful method for examining DNA sequences at the single cell level. It has been used previously in wide ranging applications [20,21,22] and is ideally suited to the analysis of somatic mutations as it allows amplification of a target locus from unique template DNAs [23]. Therefore this technique provides a tool with which to investigate DSB repair, enabling rapid amplification and sequencing of individual repair junctions. We have used smPCR to investigate NHEJ using a genetic system that allows induction of DNA double-strand breaks at a specific nuclear location *in planta*, followed by amplification of individual NHEJ repair junctions. We validated the use of this system in both *Arabidopsis* and tobacco, revealing similar patterns of NHEJ repair in both species and finding insertions at repair sites in *Arabidopsis* in contrast to previous studies [24]. In addition, we demonstrated that sequence insertions at sites of DSB repair in tobacco involve a non-classical NHEJ repair pathway.

In the course of this work we also evaluated the use of the *dao1* dual selectable marker gene [25] in tobacco. While both positive and negative selection worked well in seedling selection experiments, we caution that only positive selection was found to be reliable during tissue culture regeneration.

Results

The experimental system

The experimental system consists of a gene encoding the rare cutting endonuclease I-SceI under the control of an ethanol inducible promoter (Figure 1A) [26]. At a second locus two I-SceI restriction sites flank a ~3 kb 'spacer region' (Figure 1B) containing the *dao1* dual selectable marker gene [25] that includes three HincII sites.

To generate DNA DSBs, tissues of interest in a transgenic plant hemizygous for both experimental transgene cassettes (Figure 1) were sprayed with ethanol to induce I-SceI expression which then introduced DSBs at the two restriction sites flanking the spacer region. Following DSB induction the plants were left for several days to allow DSB repair to take place, after which the tissue was harvested and DNA prepared. Individual junctions that had been repaired by NHEJ were then specifically amplified by single molecule PCR using primers flanking the two I-SceI sites. As each product was generated from a single template molecule, every

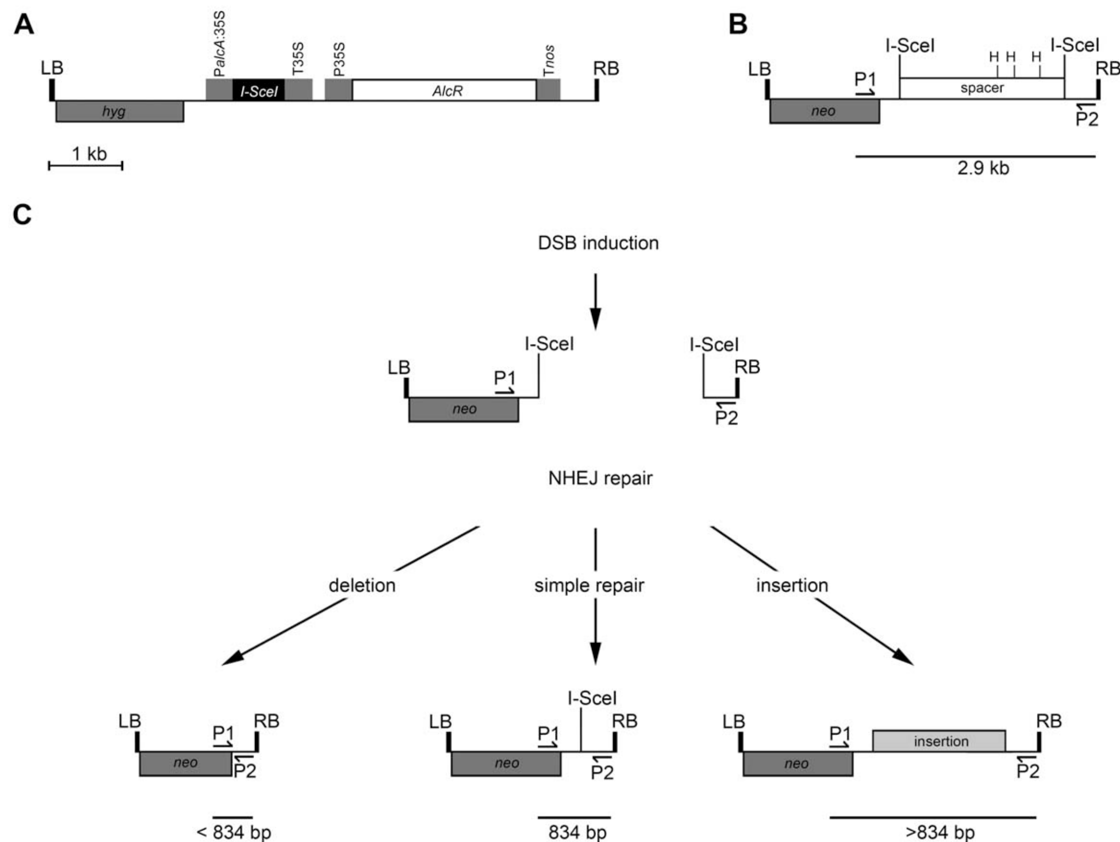

Figure 1. Overview of EtOH induced induction of DSBs. The T-DNA of vector pAlcR:ISceI (**A**) contains the left (LB) and right border (RB) sequences; a hygromycin selectable marker gene (*hyg*); the *AlcR* gene constitutively expressed from the 35S promoter; and the *I-SceI* gene driven by the *alcA*:35S promoter. In the presence of EtOH, AlcR binds to and transcriptionally activates the *alcA*:35S promoter, driving expression of *I-SceI*. The T-DNA of vector pdao1 (**B**) contains left and right border sequences; a kanamycin selectable marker gene (*neo*); and a spacer region flanked by I-SceI target sites. The spacer region also contains three HincII sites (H). (**C**) Upon I-SceI expression the I-SceI sites are cleaved leading to the excision of the spacer region. DSB repair will then result in the joining of the cleaved sequences. This may result in direct joining of the I-SceI restriction sites, deletion of sequence on either side of the DSB or insertion of sequence at the site of DSB repair. These three types of repair can be distinguished by PCR using primers P1 (DSBF1) and P2 (DSBR1) that flank the site of DSB repair. Direct joining will result in an 834 bp product whilst deletion will result in a smaller product and insertion in a larger product. In the absence of DSB induction or when DSBs are repaired by HR the spacer region will not be excised resulting in a PCR product of 2.9 kb (**B**). Amplification of this product will be prevented by digesting template DNA with HincII.
doi:10.1371/journal.pone.0032255.g001

amplicon represents a unique repair junction. Template molecules which contain the spacer region may have been repaired through homologous recombination using the sister chromatid as a template or, alternatively, these molecules may come from cells in which DSBs were not induced. These non-NHEJ template molecules were digested by HincII restriction of the spacer region such that they were not represented in the PCR products (Figure 1B). Repair events that lack the spacer region i.e. those events joining the two I-Sce-I sites to exclude the spacer (Figure 1C), will have arisen through NHEJ, as no chromosomal template molecule exists that is able to mediate such a repair *via* homologous recombination. These NHEJ-derived template molecules are not digested and remain intact, unless a *de novo* insert happens to contain HincII sites.

NHEJ repair junctions that result in no loss of sequence other than the excision of the spacer region will reform one I-SceI site from the two I-SceI half sites generated by the initial DSB induction (Figure 1C). These junctions generate PCR products of 834 bp (Figure 1C). PCR products larger or smaller than this result from NHEJ repair junctions involving insertions or deletions respectively (Figure 1C).

Generation of experimental lines

To establish the experimental system, two binary *Agrobacterium* transformation constructs, pdao1 and pAlcR:I-SceI, were generated. The pdao1 T-DNA contains *neo* for kanamycin selection and the 'spacer region' containing a 35S promoter-driven *dao1* gene flanked by two I-SceI target sites (Figure 1B). The pAlcR:I-SceI T-DNA contains *hyg* for hygromycin selection, *AlcR* driven by a 35S promoter and *I-SceI* driven by the ethanol inducible *aclA*:35S promoter [27] (Figure 1A). The pdao1 and pAlcR:ISceI constructs were individually transformed into *Arabidopsis* and tobacco by *Agrobacterium* transformation to generate D (pdao1) and A (pAlcR:ISceI) lines for both species. For *Arabidopsis* the nuclear location of the pdao1 T-DNA was determined by TAIL-PCR and comparison with the current *Arabidopsis* whole genome assembly (TAIR9).

Antibiotic resistant D and A line shoots (tobacco) or seedlings (*Arabidopsis*) were assayed by PCR to confirm the presence of the pdao1 and pAlcR:ISceI T-DNAs respectively. For PCR positive lines, T_1 progeny from self fertilised T_0 plants were grown on the appropriate antibiotic to determine segregation ratios in order to identify lines with single locus T-DNA insertions.

Homozygous, single locus, D line plants were crossed to homozygous, single locus, A line plants to generate doubly hemizygous progeny, providing the starting genotype for DSB induction. The doubly hemizygous lines were designated D4A2 (tobacco) and D19A26 (*Arabidopsis*).

dao1 enables dual selection in tobacco seedlings but not in tissue culture

The dual selectable marker gene encoding a D-amino acid oxidase (*dao1*) was included between flanking I-SceI sites to enable selection of seedlings or single cells in tissue culture for both the presence or absence of the spacer region. It was intended that this would be used in a complementary approach to identify NHEJ repair events. Effective use of *dao1* has been demonstrated for both positive and negative selection of *Arabidopsis* seedlings [25] but its use in the selection of tobacco seedlings or in explant shoot regeneration was not previously demonstrated. Experiments showed that *dao1* was effective for use both as a positive and a negative selectable marker gene for selection of germinating seedlings using concentrations of 10 mM D-alanine and 15–30 mM D-valine respectively as the selective agents (Text S1; Figures S1 and S2). In tissue culture, positive selection (D-alanine) but not negative selection (D-valine) was able to clearly distinguish between *dao1* transgenic and wild-type explants. D-valine was therefore unsuitable for negative *dao1* selection in tobacco explant regeneration (Text S1; Figures S3 and S4). As a result *dao1* acts only as the essential spacer DNA in the current experiments.

Ethanol application leads to I-SceI expression and induction of DSBs

Prior to screening, the efficiency of ethanol induction of *I-SceI* was assessed. Leaf tissue was taken from the tobacco T_0 A2 plant immediately prior to, and three days after, induction with 0.7 M ethanol. From these tissues RNA was prepared and cDNA synthesised for use in RT-PCR. A very faint gel band was observed for *I-SceI* mRNA prior to induction in leaf tissue (Figure 2A) indicating minimal leaky transcription in the absence of ethanol. After induction a strong band was observed (Figure 2A), indicating a marked increase in transcript accumulation in the presence of ethanol.

Standard PCR was used initially to confirm that I-SceI expression resulted in the induction of DSBs and excision of the spacer region by NHEJ repair. DNA was prepared from tobacco leaf samples taken from experimental lines prior to and four days after ethanol induction. The DSB repair locus was then amplified using primers flanking the two I-SceI sites. A 2.9 kb band was expected from template molecules that had not undergone spacer excision or were repaired by HR (Figure 1B). An 834 bp band was expected from template molecules arising from NHEJ repair of DSBs without any associated insertion or deletion (Figure 1C).

A 2.9 kb band, resulting from amplification of the locus without *dao1* excision, was observed for all DNA templates (Figure 2B). In addition a ~834 bp product, reflecting excision of the spacer region, was amplified when using template DNA from I-SceI induced tissue (Figure 2B). As both 2.9 kb and ~834 bp bands were present, it is evident that some template molecules originated from cells where DSBs were repaired by NHEJ and others were from cells repaired by HR (or where DSBs were not induced). Equivalent results indicated that DSB induction was efficient in *Arabidopsis* (data not shown).

HincII digest allows preferential amplification of junctions repaired by NHEJ

To favour amplification of repaired junctions arising specifically from NHEJ of DSBs, template samples were predigested with HincII which cuts three times within the spacer region but not in the flanking sequences between the I-SceI sites and primer binding sites (Figure 1B). Unavoidably digestion would also prevent amplification of template molecules arising from DSB repair events involving insertion of DNA containing HincII site(s). After HincII digestion only ~834 bp products were amplified (Figure 2C).

Single molecule PCR

To amplify PCR products representing individual DSB repair junctions smPCR was used. For smPCR the template DNA was diluted until about two thirds of samples received no template molecules at all. This ensured that ~80% of the reactions that did generate a PCR product did so from a single template molecule [28]. It was not possible to calculate the average template molecules in a given weight of DNA, as HincII digestion reduced the number of template molecules per unit weight DNA in a manner dependent upon the unknown efficiency of excision of the spacer region. Instead, a series of DNA concentrations were tested to arrive at the desired empirical concentration.

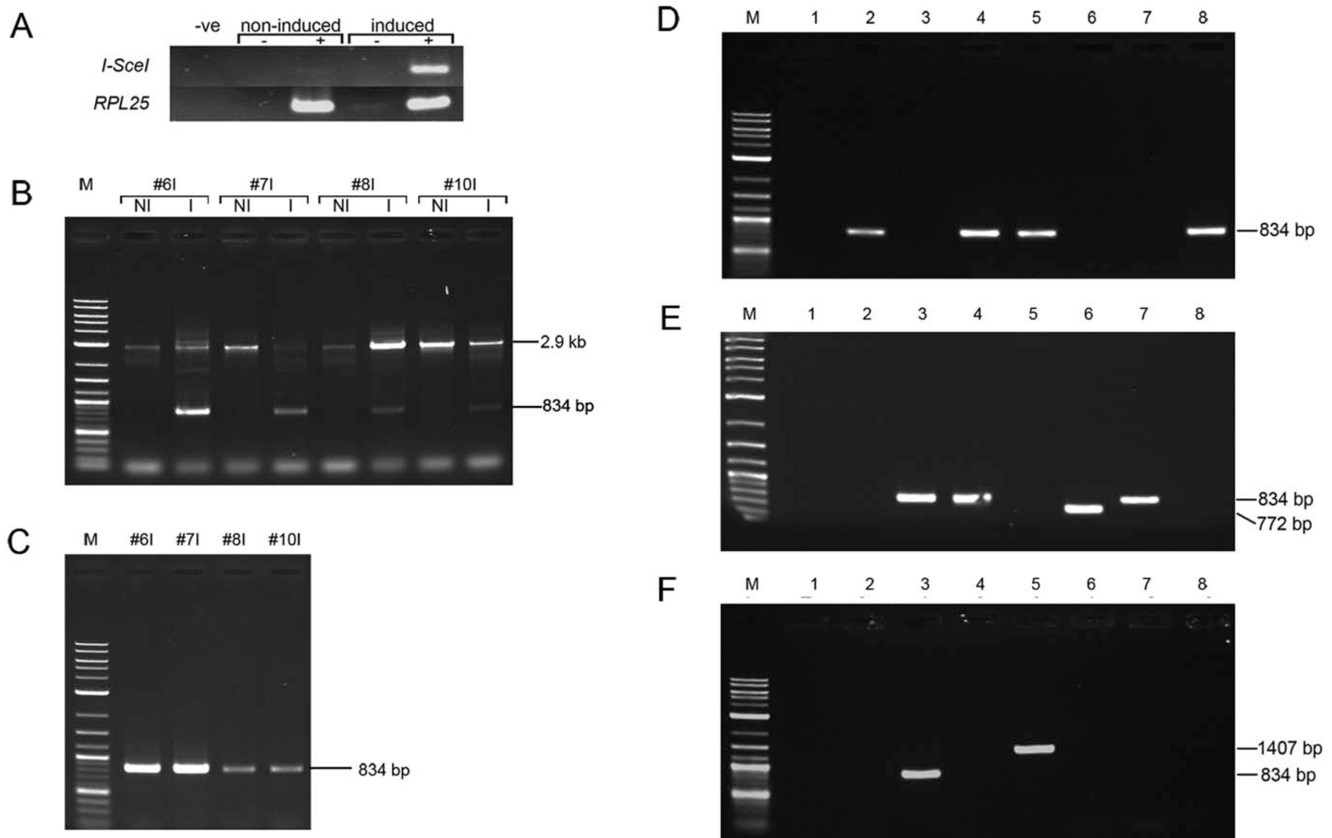

Figure 2. PCR analysis of DSB induction and repair. (**A**) RT-PCR (+) demonstrates increased *I-SceI* mRNA accumulation after induction with 0.7 M ethanol in tobacco leaf tissue. Low levels of I-SceI mRNA accumulate in non-induced leaf tissue. No reverse transcriptase (−) and no template (-ve) controls are shown. Template control RT-PCRs used *RPL25* mRNA primers. (**B**) The DSB region was PCR amplified from 4 tobacco D4A2 plants using primers DSBF1 and DSBR1 which flank the two I-SceI sites. Only the full length 2.9 kb band is amplified from template DNA extracted prior to DSB induction (NI). An additional ~834 bp band is amplified from template DNA extracted after DSB induction (I). 834 bp is the expected size of the DSB region after excision of the spacer region. (**C**) After HincII digestion of induced template DNA only the 834 bp band is amplified. No amplification is observed from those molecules which have not undergone *dao1* excision. Individual repair junctions were amplified by smPCR (**D**–**F**). The majority of products amplified were ~834 bp in size. Some repair events resulted in deletions leading to products <834 bp (**E**) while others resulted in insertion leading to products >834 bp (**F**). Examples shown are amplified from D4A2#6I template DNA.
doi:10.1371/journal.pone.0032255.g002

In tobacco, two independent doubly hemizygous plants (D4A2#2I and D4A2#6I) were tested using an optimised DNA concentration of 110–130 pg per reaction. This resulted in a product being amplified in 33–38% of reactions. For both plants this corresponds to one template molecule in ~275 pg of genomic DNA, or one DSB repair by NHEJ in every 24 genomes. In *Arabidopsis*, one plant (D19A26#1I) was tested and a DNA concentration of 1 pg/reaction was chosen. This resulted in a product being amplified in 27% of reactions, corresponding to one template molecule in ~3.7 pg, or, one DSB repair by NHEJ in every 9 genomes [29]. These results clearly indicate efficient induction of DSBs and subsequent NHEJ repair.

Arabidopsis and tobacco have similar patterns of non-homologous DNA repair

389 and 311 unique repair junctions were amplified in tobacco and *Arabidopsis* respectively. The majority of PCR products were ~834 bp in length (Figure 2D): the size expected with simple joining of the two I-SceI half sites (Figure 1C). For both species ~1.5% of PCR products were significantly smaller, corresponding to large (>50 bp) deletions (one example is shown in Figure 2E) and ~1.5% were significantly larger indicating net insertions (one example is shown in Figure 2F). Deletions that resulted in the loss of one or both primer binding sites would not have been observed in this analysis such that a maximum symmetrical deletion size of ~750 bp could be amplified by PCR using these primers. As a result, 1.3–1.5% is a minimum estimate of the proportion of repair events that involve large deletions. In addition amplification of junction sequences involving insertions would not be possible if the insert was too large for PCR or if the insert contained a HincII restriction site. 1.5% is therefore a conservative estimate of proportion of repair events involving insertion. Occasionally, a 2.9 kb product was amplified from a template molecule containing the spacer region which had not been digested by HincII (data not shown) demonstrating that inserts of at least 2 kb could be amplified effectively using this method.

For both plant species, all the insertion events recovered together with 20 randomly chosen smPCR products were sequenced. The smPCR products sequenced that were clearly the result of reactions containing more than one template molecule (~20%) were discarded. In tobacco and *Arabidopsis* respectively, 45% (9/20) and 75% (15/20) of repair junctions without insertion resulted in no loss of sequence due to simple ligation of the two I-SceI half sites (Figure 3A and 3C) excluding the spacer region.

A

```
                                          I-SceI
    GGCCCCCCCTCGAGATTACCCTGTTAT                    CCCTAGAGCTCCAGCTTTTGTTC
    CCGGGGGGGAGCTCTAATGGGAC                    AATAGGGATCTCGAGGTCGAAAACAAG
                                                                          -(bp) +(bp) MH(bp)
 1  GGCCCCCCCTCGAGA............                 ....32 bp deleted....GGGT  45     -     1
 2  GGCCCCCCCTCGAGATTACC.......                 ...........AGCTTTTGTTC    19     -     2
 3  GGCCCCCCCTCGAGATTACCCTGTTA.                 .......GCTCCAGCTTTTGTTC    8     -     1
 4  GGCCCCCCCTCGAGATTACCCTGTTA.                 .......GCTCCAGCTTTTGTTC    8     -     1
 5  GGCCCCCCCTCGAGATTACCCTGTTAT      T          .......GCTCCAGCTTTTGTTC    7     1     -
 6  GGCCCCCCCTCGAGATTACCCTGTTA.                 .....GAGCTCCAGCTTTTGTTC    6     -     2
 7  GGCCCCCCCTCGAGATTACCCTG....                 CCCTAGAGCTCCAGCTTTTGTTC    4     -     -
 8  GGCCCCCCCTCGAGATTACCCTGT                  ATCCCTAGAGCTCCAGCTTTTGTTC    1     -     1
 9  GGCCCCCCCTCGAGATTACCCTGT                  ATCCCTAGAGCTCCAGCTTTTGTTC    1     -     1
10  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
11  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
12  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
13  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
14  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
15  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
16  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
17  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
18  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
19  GGCCCCCCCTCGAGATTACCCTGTT                TATCCCTAGAGCTCCAGCTTTTGTTC    -     -     -
20  GGCCCCCCCTCGAGATTACCCTGTT                TATCCCTAGAGCTCCAGCTTTTGTTC    -     -     -
```

B

```
                                          I-SceI
    GGCCCCCCCTCGAGATTACCCTGTTAT                    CCCTAGAGCTCCAGCTTTTGTTC
    CCGGGGGGGAGCTCTAATGGGAC                    AATAGGGATCTCGAGGTCGAAAACAAG
                                                                          -(bp) +(bp) MH(bp)
                                                                                      (L/R)
 1    ............nt............ +>138 bp ....108 bp deleted..CACAA  ≥108  >138  (nt/nt)
 2  TGGGT.....31 bp deleted......  +380 bp ............AGCTTTTGTTC    43   380  (nt/F)
 3    GG.........................  +433 bp ....110 bp deleted..CAACA  133   433  (2/nt)
 4    GGCCCCCCCTCGAGATTAC........  +613 bp .....32 bp deleted..AGGGT   40   613  (nt/nt)
 5    GGCCCCCCCTCGAGATTACCCTGTT..  +127 bp .....26 bp deleted..TTAGT   28   127  (2/4)
 6  GCGTTCTA..84 bp deleted......  +677 bp .....99 bp deleted..ACAAT  183   677  (7/4)
```

C

```
                                          I-SceI
    GGCCCCCCCTCGAGATTACCCTGTTAT                    CCCTAGAGCTCCAGCTTTTGTTC
    CCGGGGGGGAGCTCTAATGGGAC                    AATAGGGATCTCGAGGTCGAAAACAAG
                                                                          -(bp) +(bp) MH(bp)
 1  GGCCCCCCCTCGAGATTA.........                 CCCTAGAGCTCCAGCTTTTGTTC    9     -     3
 2  GGCCCCCCCTCGAGATTA.........                 CCCTAGAGCTCCAGCTTTTGTTC    9     -     3
 3  GGCCCCCCCTCGAGATTA........                 TCCCTAGAGCTCCAGCTTTTGTTC    8     -     -
 4  GGCCCCCCCTCGAGATTACCCTG.                 TATCCCTAGAGCTCCAGCTTTTGTTC    1     -     1
 5  GGCCCCCCCTCGAGATTACCCTG.                 TATCCCTAGAGCTCCAGCTTTTGTTC    1     -     1
 6  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
 7  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
 8  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
 9  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
10  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
11  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
12  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
13  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
14  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
15  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
16  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
17  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
18  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
19  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
20  GGCCCCCCCTCGAGATTACCCTGTTAT                 CCCTAGAGCTCCAGCTTTTGTTC    -     -     -
```

D

```
                                          I-SceI
    GGCCCCCCCTCGAGATTACCCTGTTAT                    CCCTAGAGCTCCAGCTTTTGTTC
    CCGGGGGGGAGCTCTAATGGGAC                    AATAGGGATCTCGAGGTCGAAAACAAG
                                                                          -(bp) +(bp) MH(bp)
                                                                                      (L/R)
 1    GGCCCCCCCTCGAGATTACCCTGTTA  +154 bp TTATCCCTAGAGCTCCAGCTTTTGTTC     -    154  (4/2)
 2  TCAGTG....351 bp deleted....   +80 bp TTATCCCTAGAGCTCCAGCTTTTGTTC   351     80  (4/-)
 3    GGCCCCCCCTCGAGATTACCCTG...  +328 bp ..................CTTTTGTTC    21    328  (-/F)
 4    GGCCCCCCCTCGAGATTACCCTGTTAT +534 bp TTATCCCTAGAGCTCCAGCTTTTGTTC     -    534  (-/F)
```

E

Figure 3. Sequence of double strand break repair events. The sequence surrounding the junction sites is shown for 20 randomly chosen repair products (tobacco, **A**; *Arabidopsis*, **C**), and the repair products that harboured insertions (tobacco, **B**; *Arabidopsis*, **D**). The original sequence generated by I-SceI cleavage is shown at the top of **A–D** (the sequence of both strands is shown). Bases from the I-SceI site upstream of *dao1* are shown in red, bases from the I-SceI site downstream of *dao1* are shown in green. Inserted bases are shown in blue. In some instances microhomology was observed between the terminal bases of the fragments being joined (bold italics). Columns to the right of **A–D** indicate the total length of deletion (−), insertion (+) and microhomology (MH, not including I-SceI site overlap), nt signifies not testable, F indicates the insertion of filler DNA. Numbers in brackets indicate length of deletion or microhomology observed at the junction either upstream (L) or downstream (R) of the insertion. In tobacco the median deletion size was considerably larger in DSB repair events involving insertion than in repair events not involving insertion (**E**). The box-and-whisker plot shows the median (red line), the first and third quartiles, and the upper and lower limits of the length of deletions (two-tailed Mann-Whitney U test).
doi:10.1371/journal.pone.0032255.g003

Comparable experiments in mammals showed a similar percentage (40%) of I-SceI site reformation [30]. Two junctions in tobacco were joined by inexact ligation of the two 4 bp I-SceI half site overhangs, resulting in the addition of an extra nucleotide. The remaining junctions, 45% (9/20) for tobacco and 25% (5/20) for *Arabidopsis* had small (1–45 bp) deletions (Figure 3A and 3C). In some instances micro-homology was observed between the terminal bases of the fragments being joined (Figure 3A and 3C) although there was no indication that the amount of micro-homology observed was greater than that expected by chance.

In both tobacco and *Arabidopsis* the average deletion size (~14 bp and ~9 bp respectively) was much smaller than the average insertion size (~95 bp and ~274 bp respectively). However, as deletions occurred far more frequently than insertions, there was no net loss or gain of sequence at sites of DSB repair in either species.

Sequences inserted at sites of DSB repair are nuclear in origin

The six insertions in tobacco ranged from 127–677 bp in length (Figure 3B; Table 1) and in all cases insertion was accompanied by deletion of the starting sequence (Figure 3B). Part of insertion NTI1 shared 97% identity with the *Arabidopsis* isoleucine tRNA gene suggesting that it may be SINE-derived sequence [31]. All other inserts showed partial identity to uncharacterised EST clones from tobacco or other Solanaceous species (Table 1), indicating that all insertions were probably of nuclear origin. The four insertions in *Arabidopsis* ranged from 80 to 534 bp in length (Figure 3D; Table 2). Insertion ATI1 originated from an intergenic region on chromosome 1. The DSB locus in line D19 is located on chromosome 5 indicating that insertion ATI1 did not originate from an adjacent or remote syntenic region. Insertion ATI2 was accompanied by a large deletion upstream of the left hand I-SceI site (Figure 3D). The insert was derived from part of this deleted region but was inserted in the opposite polarity. ATI3 also originated from DNA found upstream of the left hand I-SceI recognition sequence (−33 to −352). This sequence was inserted in the same orientation as the original sequence, effectively generating a tandem duplication. Insertion ATI4 was derived from the region excised between the two I-SceI sites. This 498 bp section of the spacer region did not contain any HincII sites, enabling this junction to be amplified by smPCR. This observation implies that similar insertions of segments of the spacer region may have occurred but were missed in this screen through HincII digestion.

In both *Arabidopsis* and tobacco, short stretches of filler DNA (1–36 bp) were inserted at some junctions (Figure S5). Filler DNA was usually derived from a short stretch, or multiple stretches, of nearby sequence that probably primed ligation (Figure S5). Filler DNA has been associated previously with the insertion of T-DNA [32] and organelle DNA [18] suggesting that both integrate during repair of nuclear DSBs.

Insertion at sites of DSB repair in tobacco is associated with genomic deletion and increased micro-homology at the sequence junctions

In tobacco, the median size of deletion was found to be significantly larger in DSB repair events involving insertion than in those that did not (Figure 3E, $p = 0.003$, two-tailed Mann-Whitney U test). DSB repair events that resulted in direct ligation of the two I-SceI half sites where excluded from this analysis as the two I-SceI half sites have complementary 4 bp single stranded overlaps that may promote joining without deletion. In addition, only those DSB repair events harbouring insertions >1 bp were included.

Investigation of the presence of micro-homology at repair junctions involving insertion was possible for six junctions. At these, the bases flanking the insert sequence in its original context could be inferred from the EST sequence to which the insert matched. For the other junctions, BLAST searches only identified accessions with limited identity to the insert sequence, preventing unequivocal assessment of micro-homology. This is a limitation of analyses such as this where the sequence from which the insert originates is unknown. Five of the six junctions that could be assessed showed micro-homology (2–7 bp, Figure 3B). The remaining junction showed a 1 bp insertion of filler DNA (Figure 3B, Figure S5). Overall the level of micro-homology observed in insertion repair events was greater than that expected by chance ($p = 0.048$, $n = 6$, two-tailed Mann-Whitney U test). The presence of large deletions and micro-homology at repair junctions is indicative of insertion via MMEJ or SDSA.

Discussion

DSBs have a number of different causes including reactions with oxygen free radicals generated during aerobic respiration, ionizing radiation and faulty action of nuclear enzymes [1]. To deal with these cellular lesions a highly flexible pathway of NHEJ repair has evolved, which enables efficient joining of the many types of damaged DNA ends generated by DSBs. In addition to the mechanistic flexibility of some of the proteins involved [33], NHEJ exhibits multiple levels of redundancy enabling it to function even when some components are lacking [7,34]. Whether these alternative forms of repair constitute distinct pathways, or are essentially the same flexible pathway with one or more enzymes being substituted, is still unclear [1,10].

We developed a genetic system that allows induction of DSBs at a known nuclear locus through ethanol inducible expression of the

Table 1. Origin of tobacco insertions.

Insertion	Length	Origin	Accession
NTI1	>138	isoleucine tRNA gene	partial match to AC009755 (97%, $e = 2\times10^{-18}$)
NTI2	379	unknown nuclear	partial match to AM843263 (78%, $e = 6\times10^{-57}$)
NTI3	430	unknown nuclear	partial match to EB695504 (97%, $e = 2\times10^{-38}$)
NTI4	127	unknown nuclear	complete match to FS392274 (99%, $e = 2\times10^{-56}$)
NTI5	613	unknown nuclear	partial match to BP133287 (92%, $e = 3\times10^{-18}$)
NTI6	677	unknown nuclear	partial match to FN014067 (90%, $e = 2\times10^{-96}$) and partial match to AM847760 (80%, $e = 6\times10^{-53}$)

doi:10.1371/journal.pone.0032255.t001

Table 2. Origin of *Arabidopsis* insertions.

Insertion	Length	Origin
ATI1	154	chromosome 1 (12,474,035…12,474,189)
ATI2	80	DSB left flanking region (−39…−133)[1]
ATI3	328	DSB left flanking region (−352…−33)[1]
ATI4	534	spacer region (+5…+502)[1]

[1]co-ordinates given are relative to the left hand I-SceI cleavage site.
doi:10.1371/journal.pone.0032255.t002

rare cutting endonuclease I-SceI. This approach allowed individual NHEJ repair junctions to be efficiently amplified by smPCR. Using this system we observed a total of ~700 unique NHEJ junctions in *Arabidopsis* and tobacco, facilitating comparison of NHEJ repair in these two plant species.

In general, *Arabidopsis* and tobacco were found to have very similar patterns of NHEJ repair. In both species, the vast majority of repair events resulted in relatively conservative repair with either no loss of sequence (*At*, 75%; *Nt*, 55%) or small deletions (*At*, 25%; *Nt*, 45%) at the repair junction. In a small percentage of junctions, repair was less conservative and involved large deletions or insertions. Although the average insertion size was greater than the average deletion size, the greater frequency of deletions meant that there was no net loss or gain of sequence at sites of DSB repair. Overall this picture is similar to that observed in mammalian NHEJ [30], highlighting the high degree of conservation in this important pathway.

Our findings provide a similar general picture of NHEJ in the two species which contrasts markedly with previous comparisons that uncovered large differences between specific types of NHEJ repair in *Arabidopsis* and tobacco. Kirik *et al.* [24] investigated NHEJ repair events associated with deletions and observed insertions at a high proportion of repair junctions in tobacco, whereas deletions were larger and insertions were entirely absent in *Arabidopsis*. It is clear from our more extensive results, however, that insertion during NHEJ repair occurs at a similarly low frequency in both tobacco and *Arabidopsis*.

These apparently contradictory findings are best explained by our observation that insertion events in tobacco are associated with high levels of micro-homology and large deletions. The earlier study, by only observing NHEJ events associated with deletion (which was necessary to eliminate the activity of a negative selectable marker gene) may have been strongly biased toward observation of insertion events in tobacco.

The presence of micro-homology and increased deletion size during insertion in tobacco indicates that insertion is not mediated by the classical NHEJ repair pathway. There are (at least) two alternative mechanisms that could explain these observations. One possibility is that these sequences are not inserted *per se* but rather copied into the break site by synthesis dependent strand annealing (SDSA) [35,36]. In this model, 3′ ends generated by the DSB invade a nearby double stranded DNA molecule and short regions of micro-homology prime synthesis along this template. Template jumping to other nearby sequences may then occur resulting in the synthesis of chimeric insertion sequences. Finally, newly generated complementarity is used to bridge the gap to other side of the DSB. The second possibility is that free-floating DNA fragments close to (or recruited to) the DSB site are inserted, with the insert and DSB ends being joined by small regions of micro homology (MMEJ) [37] or by synthesis dependent MMEJ [38]. Almost all sequence outcomes may be explained equally well by both possible pathways making it essentially impossible to distinguish between the two based on the junction sequence alone. The observation that many insertions are derived from sequences close to the site of the DSB, suggests the SDSA model [39]. It is also clear, however, that many insertions (such as T-DNA insertions and the insertion of organelle DNA) are derived from free floating fragments in the nucleus. In these cases it is likely that the DNA ends are treated similarly to those of a chromosomal DSB and enter the DSB repair pathway leading to insertion.

Increased micro-homology and deletion size are also associated with insertion during NHEJ in mammals [10]. Given the wide conservation of this phenomenon, it is surprising that previous analysis has not found deletions during DSB repair to be associated with insertions in *Arabidopsis* [24]. One possibility is that decreased stability of free DNA ends in *Arabidopsis* [40] may result in larger and more frequent deletions during NHEJ repair. If deletions occur frequently at all NHEJ junctions then they would not be differentially associated with insertion events. We observed no difference, however, between the number of deletions in *Arabidopsis* and tobacco. As we were only able to detect deletion events of up to 750 bp, it is possible that larger deletions, which are known to occur in *Arabidopsis* during NHEJ [24], were missed, concealing a higher frequency of deletion. Indeed, 750 bp is the maximum and deletions between the primers must be symmetrical. As soon as one primer site is deleted, it is no longer possible to amplify the target.

Interestingly, there is recent evidence that chromatin state can affect the pathway of DSB repair, with ku-dependent NHEJ occurring preferentially in euchromatin and ATM mediated DSB repair occurring in heterochromatic regions [41]. ATM is both recruited by [42] and essential for [43] normal DSB repair by the MRN complex which, as well as having a central role in HR [2], is involved in MMEJ repair [44,45]. Given our finding that DNA insertion during repair of DSBs may be mediated by MMEJ, it is possible that insertion events may occur preferentially in heterochromatic regions. This is a possible explanation for the observation that insertions of mobile elements and organelle DNA tend to occur at heterochromatic pericentromeres [46,47]. Such a bias would minimise the chance of insertion events disrupting genes while maintaining genome stability and avoiding the loss of potentially useful genetic information. Contradictory to this hypothesis, ATM has been found to suppress MMEJ in mammalian cells [48] but this suppression occurred in plasmid re-circularisation assays and is unlikely to be representative of DSB repair in heterochromatin as efficient nuclear repair is dependent upon distinct histone epigenetic marks [49,50].

Conclusion

This study has shown smPCR in this transgenic system to be an efficient method for screening large numbers of DSB repair events and has the potential to be used in wide ranging investigations of DSB repair. Analysis of ~700 DSB repair events were analysed and, in contrast to previously published evidence suggesting differences in DSB repair between *Arabidopsis* and tobacco, the two species displayed similar DSB repair profiles in our experiments. The majority of repair events were essentially conservative resulting in no, or little, loss of sequence at the junction. A small percentage of repair events resulted in larger deletions or insertion. In tobacco, insertions were associated with larger deletions and micro-homology indicative of insertion via MMEJ or SDSA.

Materials and Methods

Plant growth and nucleic acid isolation

Nicotiana tabacum and *Arabidopsis thaliana* (Col-0) plants were grown either in soil (in pots) or in tissue culture jars containing

0.5×MS salt medium [51] and 0.8% agar (0.5×MS agar). Soil grown plants were grown in a controlled environment chamber with a 14 hr light/10 hr dark and 25°C day/18°C night growth regime. *In vitro* grown plants were grown in a controlled temperature room with a 16 hr light/8 hr dark cycle at 25°C. For *Arabidopsis*, presumed double hemizygous progeny, resulting from crosses between homozygous A and D line plants were initially grown on 0.5×MS agar medium with 50 mg L^{-1} kanamycin and 15 mg L^{-1} hygromycin to confirm the presence of both T-DNAs, before transferring plants to soil. DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions, or by phenol/chloroform extraction [52]. RNA was extracted using an RNeasy Plant Mini Kit (Qiagen) according to manufacturer's instructions

Plasmid construction and plant transformation

pAlcR:ISceI. The *AlcR* expression cassette containing the 35S promoter, *AlcR* ORF, and *nos* terminator was isolated as a NcoI/HindIII fragment from pbinSRN [26]. This cassette was blunt ended using the Klenow fragment of DNA pol I and cloned into SmaI cut pGreen0179 to generate pG.AlcR. The *I-SceI* coding region was excised from pCISceI [53] and inserted between the alcA:35S promoter and *nos* terminator in Alc-pUC (kindly provided by Dr V. Radchuk), using BamHI. Primers AlcF_NcoI (TTCCATGGGA-TAGTTCCGACCTAGGATGG) and AlcR_NcoI (TTCCATGG-GGCGATTAAGTTGGGTAACG) were then used to amplify the *I-SceI* expression cassette and the product was ligated into pG.AlcR using NcoI to generate pAlcR:ISceI.

pdao1. The 35S terminator from pPRVIIIA::neoSTLS2 [54] and 35S promoter from p35S (kindly provided by Dr S. Delaney) were cloned into pGreen0029 [55] using HindIII/BamHI and NotI/XbaI respectively. The *dao1* coding sequence was amplified from pVC_RLM_1qcz (kindly provided by Dr. A. Renz, BASF Plant Science) using primers dao1F (GAGAAAGGAAGGGAA-GAAAGC) and dao1R_XbaI (ACTCTAGACCTACAACTT-CGACTCCCG), the PCR product was then digested with XbaI and cloned into the pGreen0029 vector containing the 35S promoter and terminator, thus generating pG.dao1. A multiple cloning site containing two I-SceI restriction sites flanking HindIII and NotI sites was generated by annealing two complementary oligonucleotides I-SceIMCS1 (CTAGGGATAACAGGGTAA-TAAGCTTGCGGCCGCTAGGGATAACAGGGTAATC) and I-SceIMCS2 (TCGAGATTACCCTGTTATCCCTAGCGGC-CGCAAGCTTATTACCCTGTTATCCCTAGAGCT). This double stranded MCS had 4 bp overhangs at each end allowing ligation into SacI and XhoI cut pGreen0029, generating pG.MCS. The *dao1* expression cassette was excised from pG.dao1 with HindIII and NotI and cloned into HindIII/NotI digested pG.MCS to generate pdao1.

Transformation was performed using the pGreen system of binary transformation vectors [55]. Transgenic tobacco lines were generated using a standard leaf disc method [56]. Transgenic Arabidopsis lines were generated using the simplified floral-dip method [57] with a rapid selection protocol [58]. Putative D line and A line transformants were confirmed by PCR using primer pairs dao1F2/dao1R2 (GGCAAACCGTCCTCGTCAAG/TG-ACCTCCTTCTCCTTCGCC) and AlcRF1/AlcRR1 (CGTC-GTTCTTATTCACTCGTTTGC/TTGGAGGATGGGAAAT-GCGTTAG) respectively.

Evaluation of *dao1* selection

To evaluate the use of *dao1* as a selectable marker gene, both wild type and *dao1* transgenic seedlings were grown on 0.5×MS agar medium containing various concentrations of D-alanine (positive selection) or D-valine (negative selection). In addition, leaf explants from both wild type and *dao1* transgenic plants were grown on regeneration medium containing various concentrations of D-alanine and D-valine. For full methods see Text S1.

Experimental induction of DSBs

For RT–PCR DNA was removed from RNA samples using a TURBO DNA-free kit (Ambion, Austin, TX). Reverse transcription was performed using an Advantage RT-for-PCR kit (Clontech, Mountain View, CA) with an oligo(dT) primer in accordance with the manufacturer's instructions. Samples were also prepared without RT. For amplification of *I-SceI*, primers I-SceIF1 (ACAAACTGGCTAACCTGTTCATCGT), and I-SceIR1 (TTCGGAGGAGATAGTGTTCGGCA) were used. RPL25 mRNA was amplified using primers L25F (AAAATCT-GACCCCAAGGCAC) and L25R (GCTTTCTTCGTCCCAT-CAGG). For tobacco *I-SceI* expression was induced in the leaves of one month old D4A2 plants grown in tissue culture jars. Leaves were sprayed with 1–2 mL of 0.7 M ethanol and the jar lids replaced. For *Arabidopsis* one month old plants selected on petri dishes and then grown in soil were sprayed with 1–2 mL of 0.7 M ethanol and covered with a plastic bag to maintain the presence of ethanol vapour. The plants were then left for 4 days to allow time for I-SceI expression, the generation of DSBs and their subsequent repair. After 4 days leaf tissue was sampled.

PCR

TAIL-PCR was undertaken as described [59] using degenerate primer AD2 [59] and pdao1 T-DNA specific primers dao1T1 (T-CTTCCGCTTCCTCGCTCACTGACTCG), dao1T2 (CTCA-CTCAAAGGCGGTAATACGGTTATCCA) and dao1T3 (CC-ACAGAATCAGGGGATAACGCAGGAAAG). Standard PCR was performed using taq polymerase (ROCHE, Basel, Switzerland), using suggested PCR conditions. DSB PCR products were amplified with LongAmp taq DNA polymerase (New England Biolabs, Ipswich, MA) using suggested PCR conditions, primers DSBF1 (GATAGTGACCTTAGGCGACTTTTGAACG) and DSBR1 (TCCCCTGATTCTGTGGATAACCGT), an annealing temperature of 59°C and 40 ng template DNA. Non-induced, induced/undigested and induced/digested DNA was used as template. For digested template, 2 µg genomic DNA was digested overnight at 37°C using 20 u HincII (New England Biolabs) in a 20 µL reaction and purified using a PCR purification kit (QIAGEN) according to manufacturer's instructions.

Single molecule PCR was performed using LongAmp taq DNA polymerase and HincII digested DNA as template. Reactions were 2 µL in volume and contained 0.3 mM dNTPs, 0.4 µM primers (DSBF1 DSBR1), 0.2 u LongAmp taq DNA polymerase, 1× LongAmp buffer and 110–130 pg template DNA for tobacco or 1 pg template DNA for *Arabidopsis*. Reactions were overlayed with mineral oil to prevent evaporation. Cycle conditions were as follows: Initial denaturation 95°C 30′ then 45 cycles of 95°C 20′; 59°C 20′; and 65°C 4″ followed by a final extension at 65°C for 10″. After PCR, 18 µL of H_2O was added to each reaction to give a total volume of 20 µL. 5 µL was analysed by standard agarose gel electrophoresis and the remainder used in subsequent sequencing.

Statistical and sequence analysis

Statistical analysis of deletion size and use of micro-homology (two-tailed Mann-Whitney U test) was performed using Prism 5 (GraphPad Software). Junction sequences were analysed using Geneious (verion 5.3, Drummond AJ et al. 2010 [http://www.geneious.com/]). BLAST analysis was performed on several

databases, including NCBI's non-redundant nucleotide collection (nr/nt) and non-human non- mouse ESTs (est_others).

Supporting Information

Figure S1 D-alanine and D-valine are suitable for positive and negative selection of dao1 respectively in tobacco. Seedlings of transgenic lines containing *dao1* and wild-type (wt) seedlings were grown on various concentrations of D-alanine (**A**) and D-valine (**B**) or media containing neither amino acid (**A–B**). D-alanine was most effective at a concentration of 10 mM leading to a strong reduction in the growth of wt seedlings while not affecting the growth of transgenic seedlings (**A**). D-valine was most effective at a concentration of 30 mM leading to a marked reduction in the growth of transgenic seedlings while not affecting the growth of wt seedlings. 50 mM D-valine was toxic to both transgenic and wt seedlings and wt seedlings grown at this concentration were unable to be distinguished from transgenic seedlings. Error bars for both A and B show SD.
(TIF)

Figure S2 *dao1* transgenic and wild-type seedlings were easily distinguishable by sight when grown on both 10 mM D-alanine and 30 mM D-valine. *dao1* transgenic seedlings grown on 10 mM D-alanine showed strong growth (**A,C**), wild-type (wt) seedlings grown on the same medium bleached soon after germination (**B,D**). *dao1* transgenic seedlings grown on 30 mM D-valine had reduced growth (**E,G**) although seedlings did not bleach, cotyledons failed to fully expand and there was no growth of the first true leaf, wt seedlings grown on the same medium showed strong growth (**F,H**). Scale bars for A, B, E and F = 5 mm, scale bars for C, D, G and H = 2 mm.
(TIF)

Figure S3 10 mM D-alanine is suitable for positive selection of tobacco leaf tissue explants but 30 mM D-valine is not suitable for negative selection. Leaf explants taken from wild-type plants (wt) were killed when grown on regeneration medium containing 10 mM D-alanine (**A**). Resistant shoots were generated from *dao1* positive leaf explants grown on same media (**B**). Leaf explants from both wt and *dao1* positive plants were killed when grown on regeneration medium containing 10 mM D-valine (**E–F**). Scale bar = 10 mm.
(TIF)

Figure S4 D-valine is not suitable for negative selection of tobacco leaf tissue explants. At concentrations of both 15 mM and 5 mM D-valine both *dao1* positive and wild-type (wt) explants failed to generate resistant shoots (**A–D**). At a concentration of 2 mM D-valine both *dao1* positive and wt explants (white boxed area) generated shoots (**E**). Scale bar = 10 mm.
(TIF)

Figure S5 Filler DNA at repair junctions was derived from short stretches of flanking sequence. Filler DNA (pink) was observed at three sites of DSB repair, each involving an insertion (NTI1, ATI3 and ATI4). The filler DNA found between the insert sequence (blue) and the original DSB locus sequence (black) was derived from short stretches of DNA flanking the junction (underlined, bold). The homology at the filler DNA donor sites (bold) often extends into sequence flanking the filler DNA suggesting that several base pairs of micro-homology were used to prime the synthesis of the filler DNA promoting joining of the loose DNA ends. Numbers in brackets indicate bases missing from the diagram.
(TIF)

Text S1 Full methods and supporting material.
(DOC)

Author Contributions

Conceived and designed the experiments: AHL JNT. Performed the experiments: AHL DW. Analyzed the data: AHL DW. Contributed reagents/materials/analysis tools: JNT. Wrote the paper: AHL DW JNT.

References

1. Lieber MR (2010) The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. Annual Review of Biochemistry Vol 79 79: 181–211.
2. Mazon G, Mimitou EP, Symington LS (2010) SnapShot: Homologous Recombination in DNA Double-Strand Break Repair. Cell 142: 646.
3. van den Bosch M, Lohman PHM, Pastink A (2002) DNA double-strand break repair by homologous recombination. Biological Chemistry 383: 873–892.
4. Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Biol 56: 1–14.
5. Downs JA, Jackson SP (2004) A means to a DNA end: The many roles of Ku. Nature Reviews Molecular Cell Biology 5: 367–378.
6. Grawunder U, Wilm M, Wu XT, Kulesza P, Wilson TE, et al. (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature 388: 492–495.
7. Wang HC, Perrault AR, Takeda Y, Qin W, Wang HY, et al. (2003) Biochemical evidence for Ku-independent backup pathways of NHEJ. Nucleic Acids Research 31: 5377–5388.
8. McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends in Genetics 24: 529–538.
9. Fattah F, Lee EH, Weisensel N, Wang YB, Lichter N, et al. (2010) Ku Regulates the Non-Homologous End Joining Pathway Choice of DNA Double-Strand Break Repair in Human Somatic Cells. Plos Genetics 6: e1000855.
10. Simsek D, Jasin M (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. Nature Structural & Molecular Biology 17: 410–443.
11. Yang Y-N, Chou K-M, Pan W-Y, Chen Y-W, Tsou T-C, et al. (2011) Enhancement of non-homologous end joining DNA repair capacity confers cancer cells resistance to the novel selenophene compound, D-501036. Cancer Letters 309: 110–118.
12. Davidson D, Coulombe Y, Martinez-Marignac VL, Amrein L, Grenier J, et al. (2011) Irinotecan and DNA-PKcs inhibitors synergize in killing of colon cancer cells. Invest New Drugs;doi: 10.1007/s10637-10010-19626-10639.
13. Chen H, Ma Z, Vanderwaal RP, Feng Z, Gonzalez-Suarez I, et al. (2011) The mTOR Inhibitor Rapamycin Suppresses DNA Double-Strand Break Repair. Radiation Research 175: 214–224.
14. Friesner J, Britt AB (2003) Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. Plant Journal 34: 427–440.
15. Kohli A, Twyman RM, Abranches R, Wegel E, Stoger E, et al. (2003) Transgene integration, organization and interaction in plants. Plant Mol Biol 52: 247–258.
16. Wright DA, Townsend JA, Winfrey RJ, Jr., Irwin PA, Rajagopal J, et al. (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J 44: 693–705.
17. Huang CY, Ayliffe MA, Timmis JN (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: 9710–9715.
18. Lloyd AH, Timmis JN (2011) The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. Mol Biol Evol 28: 2019–2028.
19. Tanaka S, Ishii C, Hatakeyama S, Inoue H (2010) High efficient gene targeting on the AGAMOUS gene in an Arabidopsis AtLIG4 mutant. Biochemical and Biophysical Research Communications 396: 289–293.
20. Ben Yehezkel T, Linshiz G, Buaron H, Kaplan S, Shabi U, et al. (2008) De novo DNA synthesis using single molecule PCR. Nucleic Acids Research 36: e107.
21. Chhibber A, Schroeder BG (2008) Single-molecule polymerase chain reaction reduces bias: Application to DNA methylation analysis by bisulfite sequencing. Analytical Biochemistry 377: 46–54.
22. Kraytsberg Y, Bodyak N, Myerow S, Nicholas A, Ebralidze K, et al. (2009) Quantitative Analysis of Somatic Mitochondrial DNA Mutations by Single-Cell Single-Molecule PCR. Methods in Molecular Biology 554: 329–369.
23. Kraytsberg Y, Khrapko K (2005) Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations. Expert Review of Molecular Diagnostics 5: 809–815.
24. Kirik A, Salomon S, Puchta H (2000) Species-specific double-strand break repair and genome evolution in plants. Embo Journal 19: 5562–5566.
25. Erikson O, Hertzberg M, Nasholm T (2004) A conditional marker gene allowing both positive and negative selection in plants. Nature Biotechnology 22: 455–458.

26. Caddick MX, Greenland AJ, Jepson I, Krause KP, Qu N, et al. (1998) An ethanol inducible gene switch for plants used to manipulate carbon metabolism. Nature Biotechnology 16: 177–180.
27. Salter MG, Paine JA, Riddell KV, Jepson I, Greenland AJ, et al. (1998) Characterisation of the ethanol-inducible alc gene expression system for transgenic plants. Plant Journal 16: 127–132.
28. Nakano M, Komatsu J, Matsuura S, Takashima K, Katsura S, et al. (2003) Single-molecule PCR using water-in-oil emulsion. Journal of Biotechnology 102: 117–124.
29. Schmuths H, Meister A, Horres R, Bachmann K (2004) Genome size variation among accessions of Arabidopsis thaliana. Ann Bot 93: 317–321.
30. Bennardo N, Cheng A, Huang N, Stark JM (2008) Alternative-NHEJ Is a Mechanistically Distinct Pathway of Mammalian Chromosome Break Repair. Plos Genetics 4: e1000110.
31. Ohshima K, Okada N (1994) Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of 3 different tRNA-derived retrotransposons in the octopus. Journal of Molecular Biology 243: 25–37.
32. Windels P, De Buck S, Van Bockstaele E, De Loose M, Depicker A (2003) T-DNA integration in Arabidopsis chromosomes. Presence and origin of filler DNA sequences. Plant Physiol 133: 2061–2068.
33. Lieber MR, Lu H, Gu J, Schwarz K (2008) Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate nonhomologous DNA end joining: relevance to cancer, aging, and the immune system. Cell Research 18: 125–133.
34. Liang L, Deng L, Nguyen SC, Zhao X, Maulion CD, et al. (2008) Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. Nucleic Acids Research 36: 3297–3310.
35. Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Research 25: 4650–4657.
36. Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO J 17: 6086–6095.
37. Ma JL, Kim EM, Haber JE, Lee SE (2003) Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Molecular and Cellular Biology 23: 8820–8828.
38. Yu AM, McVey M (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res 38: 5706–5717.
39. Pace JK, II, Sen SK, Batzer MA, Feschotte C (2009) Repair-mediated duplication by capture of proximal chromosomal DNA has shaped vertebrate genome evolution. PLoS Genet 5: e1000469.
40. Orel N, Puchta H (2003) Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. Plant Mol Biol 51: 523–531.
41. Kongruttanachok N, Phuangphairoj C, Thongnak A, Ponyeam W, Rattanatanyong P, et al. (2010) Research Replication independent DNA double-strand break retention may prevent genomic instability. Molecular Cancer 9: 70.
42. Lee JH, Paull TT (2005) ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science 308: 551–554.
43. Wu XH, Ranganathan V, Weisman DS, Heine WF, Ciccone DN, et al. (2000) ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. Nature 405: 477–482.
44. Della-Maria J, Zhou Y, Tsai MS, Kuhnlein J, Carney J, et al. (2011) hMre11/hRad50/Nbs1 and DNA ligase IIIα/XRCC1 act together in an alternative non-homologous end joining pathway. J Biol Chem;doi: 10.1074/jbc.M1111.274159.
45. Taylor EM, Cecillon SM, Bonis A, Chapman JR, Povirk LF, et al. (2010) The Mre11/Rad50/Nbs1 complex functions in resection-based DNA end joining in *Xenopus laevis*. Nucleic Acids Research 38: 441–454.
46. Hall AE, Kettler GC, Preuss D (2006) Dynamic evolution at pericentromeres. Genome Research 16: 355–364.
47. 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: 665–675.
48. Rahal EA, Henricksen LA, Li Y, Williams RS, Tainer JA, et al. (2010) ATM regulates Mre11-dependent DNA end-degradation and microhomology-mediated end joining. Cell Cycle 9: 2866–2877.
49. Fnu S, Williamson EA, De Haro LP, Brenneman M, Wray J, et al. (2011) Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proceedings of the National Academy of Sciences of the United States of America 108: 540–545.
50. Peng JC, Karpen GH (2009) Heterochromatic Genome Stability Requires Regulators of Histone H3 K9 Methylation. Plos Genetics 5: e1000435.
51. Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plantarum 15: 473–497.
52. Pallotta MA, Graham RD, Langridge P, Sparrow DHB, Barker SJ (2000) RFLP mapping of manganese efficiency in barley. Theoretical and Applied Genetics 101: 1100–1108.
53. Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proceedings of the National Academy of Sciences of the United States of America 93: 5055–5060.
54. Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature 422: 72–76.
55. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM (2000) pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. Plant Mol Biol 42: 819–832.
56. Mathis NL, Hinchee AW (1994) *Agrobacterium* inoculation techniques for plant tissues. In: Gelvin SB, Schilperoort RA, eds. Plant Molecular Biology Manual. 2nd ed. Dordrecht, The Netherlands: Kluwer Academic Publishers. pp 1–9.
57. Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant Journal 16: 735–743.
58. Harrison SJ, Mott EK, Parsley K, Aspinall S, Gray JC, et al. (2006) A rapid and robust method of identifying transformed *Arabidopsis thaliana* seedlings following floral dip transformation. Plant methods 2: 19.
59. 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.

Chapter 4 Cytoplasmic Organelle DNA Preferentially Inserts into Open Chromatin Regions

Introduction

Endosymbiont DNAs have constantly bombarded the nucleus since the appearance of eukaryotes and it is usual for nuclear genomes to contain multiple chromosomal integrants derived from cytoplasmic organellar genomes (Timmis et al., 2004). This process of DNA escape and integration has resulted in massive functional relocation to the nucleus of genes that once belonged to the prokaryotic ancestors of mitochondria and chloroplasts. Simple DNA transposition and functional gene relocation from the extant organellar genomes have both been demonstrated experimentally (Thorsness and Fox, 1990;Huang et al., 2003;Stegemann et al., 2003) and found to occur at previously unexpectedly high frequencies. However, the mechanisms responsible for organellar DNA escape and incorporation into the nuclear genome have not been extensively investigated and relatively little is known about how relocated genes become functional.

DNA double-strand break (DSB) repair sometimes results in the integration of mitochondrial DNA into the nuclear genome in yeast (Ricchetti et al., 2004) and tobacco (Wang et al., 2012a), and chloroplast DNA incorporation is also implicated during repair of DSBs by non-homologous end joining (NHEJ) (Lloyd and Timmis, 2011). Bioinformatic analysis also suggests that NHEJ is involved in the formation of nuclear integrants of mitochondrial DNAs (*numt*s) in primate genomes (Hazkani-Covo and Covo, 2008). Moreover, recent human *numt*s were shown to insert preferentially into genes, especially into introns (Ricchetti et al., 2004) that are often flanked by regions of open chromatin (Tsuji et al., 2012). However, whether this is a general rule, and whether it applies to the plastid counterparts of *numt*s (nuclear integrants of plastid DNAs (*nupt*s)), has not been investigated. A recent study in tobacco showed that mild heat stress increases DNA transfer from chloroplast-to-nucleus (Wang et al., 2012a) suggesting the possibility that heterochromatin relaxation that is associated with heat stress (Pecinka et al., 2010), may be responsible. We have investigated the hypothesis that cytoplasmic organelle DNA tends to integrate preferentially into DNA in open chromatin regions by comparing nuclear organelle DNAs (*norg*) maps with chromatin

status as revealed by DNase I hypersensitivity (DH) sites or Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE) (Boyle et al., 2008;Song et al., 2011).

We identified recent *nupt*s that are unique to *Oryza sativa* subsp. *indica* which had inserted into the nuclear genome after its very recent divergence from *O. sativa* subsp. *japonica*. The *nupt* insertion sites were studied to determine whether they include, or are flanked by, DH sites in chromatin derived from seedling or callus tissues. A related study (Tsuji et al., 2012) found that *numt* loci in the human nucleus were often found in open regions of open chromatin. However, a problem with this approach is that it examines the chromatin status of *numt* junction sites after, rather than before insertion. Thus, there is no certainty that mitochondrial DNA inserted into pre-existing open chromatin and it is possible that the *numt* insertion event may cause chromatin relaxation. To avoid this ambiguity, and to investigate the generality of *norg* insertion mechanisms, we identified chimpanzee-specific *numt*s and characterised the equivalent loci in the human genome, allowing us to analyze the probable chromatin status before mitochondrial DNA insertion.

Results

Comparative Analysis of *nupt* Integration Sites in Oryza species Supports NHEJ-mediated Chloroplast DNA insertion

Using *O. rufipogon*, the wild progenitor of *O. sativa* (Khush, 1997) as a control, we indentified *nupt*s that had inserted into the nuclear genome of *O. sativa* subsp. *indica* after its divergence from subsp. *japonica* (Figure 1). By comparing the same loci and their flanking genomic regions between the two subspecies we were able to deduce the mechanism of DSB repair (Hazkani-Covo and Covo, 2008). We reasoned that, if *O. sativa* subsp. *indica* contains a *nupt* that is absent at the equivalent loci in both *O. sativa* subsp. *japonica* and *O. rufipogon,* the latter two taxa will reveal the *nupt* preinsertion site. Therefore, the differences among the chromosomal *nupt* loci of these three Oryza taxa may be considered as record of the insertion process.

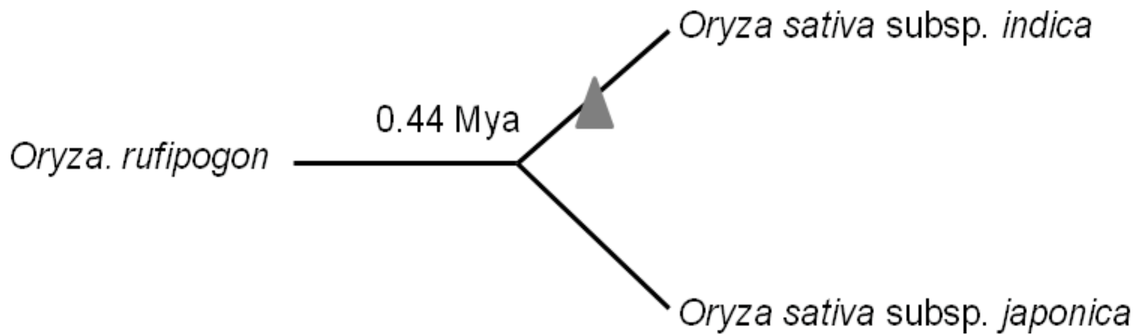

Figure 1. A phylogenetic tree of *O. rufipogon*, *O. Sativa* subsp. *indica* and subsp. *japonica* showing the recent *nupt* insertions (grey triangle) used to investigate *nupt* insertion mechanisms. Subspecies *indica* and *japonica* diverged 0.44 Mya (Ma and Bennetzen, 2004).

Among the 14 insertional events with their 14×2 molecular ligation points, 8 involved perfect or slightly imperfect microhomology of more than 1 bp (Figure 2 and Table S1), with a single matching base seen at 6 other junctions (Table S1), implicating DSB repair by NHEJ. The remaining 14 junctions involved blunt-end ligation (Table S1). Consistent with the observations in primate *numt*s (Hazkani-Covo and Covo, 2008), only 2 of the 14 *nupt* insertions resulted in deletion of nucleotides, suggesting that DSB repair with cytoplasmic organelle DNA insertion reduces sequence loss when the break is healed. It is known that DSB repair of incompatible ends always involves deletion of a few nucleotides (Guirouilh-Barbat et al., 2004;Nick McElhinny et al., 2005;Lloyd et al., 2012).

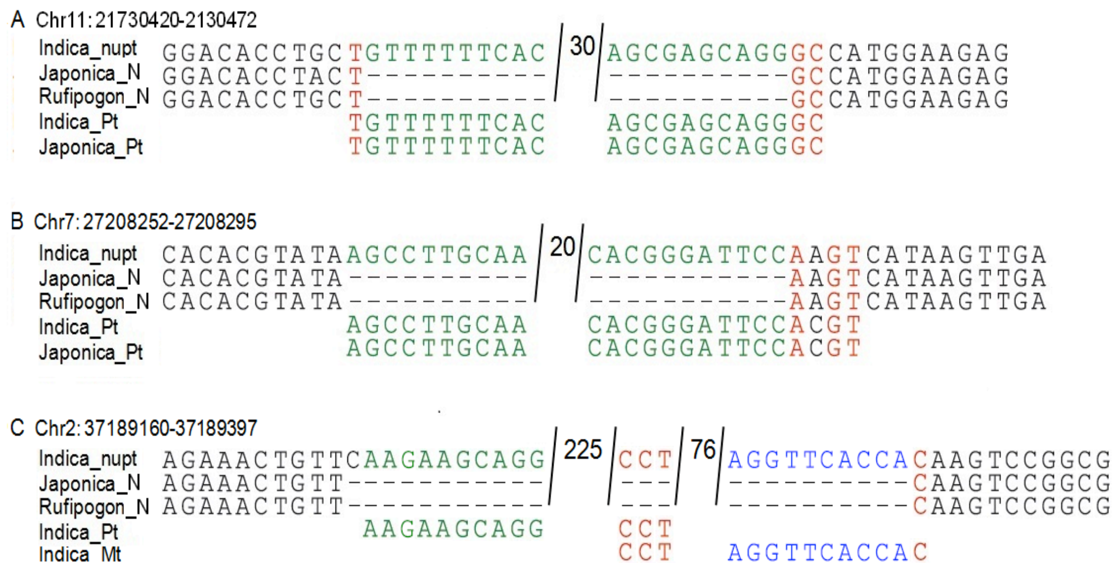

Figure 2. Examples of *nupt*-mediated DSB repair reflect NHEJ. Alignments show *O. sativa* subsp. *indica* loci containing a *nupt*. The corresponding nuclear sequence in subsp. *japonica* and *O. rufipogon* are shown (TIGR database release 5). The chloroplast DNA sequences (Pt) from subsp. *indica* and *japonica* are also presented. (A) A *nupt* insertion (shown in green) that involved short microhomology (1 or 2 bp shown in red) at both fusion points. (B) A *nupt* locus that involved imperfect microhomology at the right fusion point and blunt-end repair at the left fusion point. (C) A chimeric insertion containing a *nupt* of 225 bp (in green) and a *numt* of 76 bp (in blue). Complementary microhomology of AGG in the *norg* with CCT in chloroplast DNA and mitochondrial DNA is marked red. The subsp. *indica* mitochondrial DNA sequence (Mt) is shown in (C).

Several chimeric insertions involving both mitochondrial- and chloroplast-derived sequences were observed, confirming their relative abundance as described previously (Lloyd and Timmis, 2011;Wang et al., 2012a). The *nupt*s examined show extensive variation, consisting of DNA fragments originating from different parts of the chloroplast circular genome, mixed chloroplast and mitochondrial DNA integrants (Figure 2), and mosaic DNA inserts containing nuclear, mitochondrial DNA and chloroplast DNA (Table S1). Some of the loci reveal that mitochondrial and chloroplast DNA fragments also join together through microhomology (Figure 2C).

These results suggest that *norg* insertion precludes deletion at DSBs and involves both blunt-end repair and variable lengths of microhomology whereas, if other filler DNAs are included in repairs, there is often some deletion at the preinsertion site (Lin and Waldman, 2001a;b). For example capture of Adeno-Associated Virus (AAV) in I-SceI induced breaks is associated with a high frequency of deletion (Miller et al., 2004;Hazkani-Covo and Covo, 2008). Thus, our results suggest that organelle DNAs play a role in preserving genome integrity during potentially deleterious DSB repair.

Norg Insertion Sites Prefer Open Chromatin Regions

Open chromatin regions often show nucleosome depletion that allows genomic DNA segments to be exposed to interacting molecules (Hogan et al., 2006;Kim et al., 2007;Song et al., 2011). The recent *nupt*s we indentified in subsp. *indica* provided an opportunity to examine whether open chromatin is more accessible to organellar DNA

insertion. For each subsp. *indica* specific *nupt* we reasoned that the preinsertion sites located in the *O. sativa* subsp. *japonica* genome, would represent the chromatin status prior to integration. Therefore, DNase-seq data for *O. sativa* subsp. *japonica* shed light on subsp. *indica* chromatin status prior to *nupt* insertion. Crosschecking subsp. *japonica* preinsertion sites together with 1 kb of flanking DNA revealed that 6 cases of the 14 examined were located in open chromatin regions in seedling chromatin, and 8 were seen in callus (Table 1). The rice genome contains 420 Mb of DNA, of which only 5% and 7% is DNase I hypersensitive in seedling and callus chromatin respectively (Zhang et al., 2012). Therefore, on average one of these 14 recent insertions is expected in 30 Mb of the genome if the events are random. However, one of these recent *nupts* occurs in 16 kb (seedling) or 19 kb (callus) of DH DNA, indicating that *nup*t insertion strongly favours open chromatin (p <0.0001, χ^2 test).

Table 1. Open chromatin data and *Norg* descriptions used in this analysis

Nupt			*Numt*				
Tissue	DH sites	No. of *nupt*s	Cell line	DH sites	No. of *numt*s	FAIRE sites	No. of *numt*s
Seedling	97,975 (5%)	6	GM12878	103,075 (1.528%)	6	146,147 (0.728%)	10
Callus	155,025 (7%)	8	HeLa-S3	142,403 (2.174%)	7	131,935 (0.694%)	2
			H1-ES	138,025 (3.224%)	7	126,439 (0.695%)	6
			HUVEC	133,091 (2.259%)	3	225,564 (1.723%)	5

Table 1. Co-location of *norg*s with open chromatin regions. Percentages (in parentheses) indicate the proportion of the rice or human genome identified as open chromatin regions. Example 1: 6 of the 14 ssp. *indica nupts* (43%) examined were located in 5% of the genome identified as DH sites in ssp. *japonica* seedling tissues. Example 2: 10 of the 52 (19%) chimpanzee *numts* examined were located in 0.728% of the human genome identified as FAIRE sites in human cell line GM12878. Human cell line descriptions: GM12878, lymphoblast; Hela-S3, cervical carcinoma; H1-ES, human embryonic stem cells; HUVEC, human umbilical vein endothelial cells.

In order to investigate whether *norg* insertion prefers into open chromatin is generalised in eukaryotes, a similar analysis was carried out for 52 previously-indentified species-specific *numt*s in the chimpanzee genome (version, panTro 2) (Hazkani-Covo and

Covo, 2008). The corresponding loci were then unequivocally located in the human genome (version, hg 18) (Karolchik et al., 2004) and 1 kb of flanking DNA was compared with open chromatin coordinates defined by DNase-seq and FAIRE-seq in 4 different human cell lines (Table 1). In common with their *nupt* counterparts in *Oryza*, chimpanzee *numt* insertion sites strongly favoured open chromatin in all the cell lines tested (p<0.0001, χ^2 test). Furthermore, different cell lines showed different degrees of preference for *norg* insertion, reflecting their known differences in chromatin status (Song et al., 2011). Consistent with these observations, more *norg*s correlate with open chromatin in parallel with the number of DH or FAIRE sites reported (Table 1) in different human cell lines or in different plant tissues. We conclude that cytoplasmic organelle DNA preferentially inserts into open chromatin regions in diverse eukaryotes. Open chromatin regions are likely to be more accessible to the proteins involved DNA breakage and repair, necessarily leading both to more cleavage and their more successful healing, sometimes with the incorporation of available mitochondrial or chloroplast DNA fragments.

Discussion

NHEJ has been suggested to associate with chloroplast DNA insertion into nuclear genome (Lloyd and Timmis, 2011) and this is supported by the current comparative analysis using Oryza genome sequence data. Moreover, NHEJ-mediated DSB repair that includes chloroplast DNA insertion may protect genome integrity by precluding deletions, though insertional mutagenesis may be a by-product. However, mutation may be alleviated as approximately half of *de novo nupt*s are unstable and may be very quickly deleted from the genome (Sheppard and Timmis, 2009), though the precision or otherwise of excision remain to be established. The frequency of DNA transfer from chloroplast to nucleus is different in different tissues (Huang et al., 2003;Stegemann et al., 2003;Sheppard et al., 2008) and it is known that the frequency of organellar DNA transposition is positively correlated with the amount of available organelle DNA (Wang et al., 2012a;b). Here we describe that DNA transfer from organelle to nucleus also tends to occur in regions of open chromatin. Therefore there is evidence that *norg*s insert into open chromatin (this paper) and the loci probably remain transcriptionally active after insertion (Tsuji et al., 2012). Thus chromatin state appears to be a

significant contributor to the successful relocation of cytoplasmic organellar genes to the nucleus. The accessibility of chromatin is also modified by stress (Pecinka et al., 2010) and this may account for the significant increase in stable integration of plastid DNA after heat stress (Wang et al., 2012a).

Materials AND Methods

Identification of *O. sativa* subsp. *indica* specific *nupt*s

The nuclear, chloroplast, and mitochondrial genome sequences of *O. sativa* subsp. *indica* and *O. sativa* subsp. *japonica* were downloaded from NCBI. *O. rufipogon* nuclear genome is acquired from Rice Haplotype Map Project Database (Huang et al., 2012). *Nupt*s present in the subsp. *indica* genome were identified using BlastN (version 2.2.23) (Altschul et al., 1990). Local BlastN with the parameters previously described (Wang et al., 2012c) was carried out. *Nupt*s that appeared in *O. sativa* subsp. *indica* that were not present at the same unequivocally-identified loci in *O. sativa* subsp. *japonica* and *O. rufipogon* were indentified. *Nupt*s that appeared in *O. sativa* subsp. *indica* but could not be located in *O. sativa* subsp. *japonica* and *O. rufipogon* genomes were eliminated from this study. A total of 14 *nupt*s were identified that fitted these criteria and each was analysed in detail.

NHEJ analysis

The NHEJ analysis was as previously described (Hazkani-Covo and Covo, 2008). In short, *nupt*s were classified by known NHEJ patterns (microhomology and blunt end repair). Microhomology was indentified only if the nucleotide adjacent to the fusion point was shared among the *nupt*, the corresponding subsp. *japonica* and *rufipogon* nuclear sequences, and the plastomes of subsp. *japonica* or *indica*. If no microhomology was found, then the NHEJ was classified as a blunt-end repair. Any sequences of less than 10 nucleotides found at the junction sites, other than known nuclear, mitochondrial or chloroplast DNA, was classified as a non-template insertion.

Open chromatin regions

To test whether insertion sites of *norg*s correlate with open chromatin regions, we downloaded the open chromatin data generated by DNase-seq for *O. sativa* subsp. *japonica* (Zhang et al., 2012) and human (Song et al., 2011), and the human database of FAIRE-seq (Song et al., 2011).The position of individual *norg*s and their flanking regions was superimposed on these coordinates to identify the corresponding chromatin status.

Legends for supplemental table

Table S1. ***Nupt* information**

Nupt number	Organism	Chromosome	chromosome start	Chromosome end	Insertion Origin	Blunt-end /Microhomology	Microhomology bases	non template insertion	Blunt-end /Microhomology	Microhomology bases	non template insertion	Deletion size	Deleted sequence
1	indica	chr1	8577632	8577955	ct_33489_33812	Blunt-end		a	Blunt-end		gg	0	
	japonica	chr1	7925581	7925582									
2	indica	chr1	33910711	33910824	ct_32397_32510	Blunt-end			Microhomology	tggg		0	
	japonica	chr1	32511931	32511933									
3	indica	chr1	26634425	26634456	chimeric:chr4_4090343_4090309; Oryza sativa indica Group mitochondrial plasmid-like B4 like DNA_145_180;mt_188766_188742;ct_19036_19067	Microhomology	ttttccataggattg		Blunt-end		atta	0	
	japonica	chr1	25469478	25469479									
4	indica	chr2	37189160	37189397	chimeric:ct_54202_54439; mt_53959_54035	Blunt-end		c	Microhomology	c		0	
	japonica	chr2	35932857	35932858									
5	indica	chr3	31777840	31777885	ct_69255_69300	Blunt-end		g	Microhomology	a		0	
	japonica	chr3	28519261	28519262									
6	indica	chr3	40007968	40008171	ct_93530_93787/ct_121263_121520	Blunt-end		cgaaccga	Microhomology	t		0	
	japonica	chr3	36476270	36476271									
7	indica	chr5	29644072	29644480	chimeric:ct_111367_111336;ct_106348_106727	Microhomolgy	aat		Blunt-end			0	
	japonica	chr5	28457616	28457617									
8	indica	chr7	27208252	27208295	ct_95561_95604/ct_119446_119489	Blunt-end			Microhomology	aagt(imperfect)		0	
	japonica	chr7	29817075	29817076									
9	indica	chr7	9362429	9362525	ct_119052_119148/ct_95902_95998	Blunt-end		tg	Microhomology	t		0	
	japonica	chr7	9419151	9419152									
10	indica	chr7	21006376	21006515	ct_32024_32163	Microhomology	ccatta(imperfect)		Microhomology	ctt		79	tgataatttcgtaaatcttagtaatgcatgaaaaatggagacacctataataactttgtgaatcttaaaatatactatt
	japonica	chr7	23656765	23656845									
11	indica	chr9	9414984	9415088	ct_47119_47223	Blunt-end			Microhomolgy	g		0	
	japonica	chr9	11511536	11511537									
12	indica	chr11	21730420	21730472	ct_28784_28836	Microhomology	t		Microhomoly	gc		0	
	japonica	chr11	29648997	29648998									
13	indica	chr11	14418470	14418755	chimeric:ct_122498_122677/ct_92373_92552;ct_69866_69972	Blunt-end			Blunt-end			22	tttcaaactgttaaatgatatg
	japonica	chr11	19574533	19574556									
14	indica	chr11	13759923	13760035	chimeric:ct_123596_123629/ct_91421_91454;mt_385614_385652/mt_466727_466765;ct_107199_107239	Blunt-end		g	Microhomology	ggg		0	
	japonica	chr11	19080457	19080458									

Genome Version used in this study

indica		japonica	
chr1	CM000126.1	chr1	AP008207.2
chr2	CM000127.1	chr2	AP008208.2
chr3	CM000128.1	chr3	AP008209.2
chr4	CM000129.1	chr4	AP008210.2
chr5	CM000130.1	chr5	AP008211.2
chr6	CM000131.1	chr6	AP008212.2
chr7	CM000132.1	chr7	AP008213.2
chr8	CM000133.1	chr8	AP008214.2
chr9	CM000134.1	chr9	AP008215.2
chr10	CM000135.1	chr10	AP008216.2
chr11	CM000136.1	chr11	AP008217.2
chr12	CM000137.1	chr12	AP008218.2
chloroplast	NC_008155.1	chloroplast	NC_001320.1
mitochondrion	NC_007886.1		

Chapter 5

Plastid Sequences Contribute to Some Plant Mitochondrial Genes

Dong Wang, Mathieu Rousseau-Gueutin and Jeremy N. Timmis

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

Molecular Biology and Evolution, July 2012, 29(7): 1707-1711

STATEMENT OF AUTHORSHIP

Plastid Sequences Contribute to Some Plant Mitochondrial Genes

Molecular Biology and Evolution, July 2012, 29(7): 1707-1711

Dong Wang (Candidate)

Performed experiments, analysed results and wrote the manuscript.

I hereby certify that the statement of contribution is accurate

Signed...... .... *Date*....6/3/2013....

Jeremy N. Timmis

Supervised development of work and assisted in analysing results and 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*....6th Feb 2013....

Mathieu Rousseau-Gueutin

Assisted in analysing results and 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*....20/02/2013........

Wang, D., Rousseau-Gueutin, M. & Timmis, J.N. (2012) Plastid sequences contribute to some plant mitochondrial genes.
Molecular Biology and Evolution, v. 29(7), pp. 1707-1711

NOTE:
This publication is included after page 23 in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1093/molbev/mss016

Chapter 6 Conclusions and Future Directions

The escape of ancestral and extant plastidic and mitochondrial DNA between genetic compartments in eukaryotes is now considered to be a crucial step in the transfer of cytoplasmic organellar genes to the nucleus. The process has had profound consequences for eukaryote evolution (Timmis et al., 2004). Endosymbiont DNAs have constantly bombarded the nucleus since the appearance of eukaryotes and DNA migration from chloroplast to nucleus, the focus of this thesis, is an ongoing process in seed plants. DNA transfer from chloroplast to nucleus has been demonstrated in tobacco and it has a very high frequency that far exceeds previous expectations (Huang et al., 2003;Stegemann et al., 2003;Sheppard et al., 2008).

Chapter 2 describes how mild heat stress increases very significantly the frequency of DNA transfer from chloroplast to nucleus. This result indicates that the frequency of cytoplasmic organelle DNA migration can be changed by stress. Under stress conditions, more organellar DNA fragments or whole genomes may appear transiently in the nucleus such that more may insert into the nuclear genome, suggesting that the nuclear genome is diversified by the environment. This finding may partly explain the genome diversity found in different ecotypes of *Arabidopsis thaliana* originating from different habitats and climatic conditions (Cao et al., 2011).

Chapter 3 reports results that investigate the mechanisms of organellar DNA insertion into the nuclear genome. After environmental stress, short fragments of mitochondrial DNA are introduced at sites of DSB repair and this is not seen when plants are grown under optimal growth conditions when only DNA of non-organellar is inserted. A large amount of free chloroplast DNA was shown to be liberated after heat treatment that caused disruption of organellar membranes.

Chloroplast DNA integration during DSB repair is not observed in Chapter 3 and Chapter 2 that demonstrated only the incorporation of short pieces of mitochondrial DNA. The apparent lack of chloroplast DNA at healed DSBs may be due to the PCR techniques used. There is a strong possibility that chloroplast DNAs are inserted in very

large tracts as previously demonstrated (Huang et al., 2004;Lloyd and Timmis, 2011) that would not be amplified using the methods employed.

Chapter 4 shows that NHEJ-mediated chloroplast DNA insertion is supported by bioinformatic analysis of naturally-occurring *nupts*. However, the *nupt*s studied are rather short and would have been revealed in the PCR screen in Chapter 3. Therefore, the lack of chloroplast DNA at healed DSBs in the experimental data suggests that *nupt*s are initially very long and that the shorter, naturally occurring examples found by bioinformatic analysis, may be highly derived, despite their recent origin as assessed by their sequence similarity to the extant plastome. This result is consistent with the high level of instability of *nupt*s, ~50% of which are subject to deletions within one generation after insertion (Sheppard and Timmis, 2009). Thus the small chloroplast DNA integrants studied in Chapter 4 may be remnants of deletions inside initially large chloroplast DNA integrants. The genomic sequence analyses also show that chloroplast DNA may play a role in preserving genome integrity during potentially deleterious DSB repair, which is consistent with previous reports that mitochondrial DNA insertion reduces the sequence loss when a break is healed in primates (Hazkani-Covo and Covo, 2008). Recently, heterochromatin relaxation associated with heat stress was reported in *Arabidopsis thaliana* (Pecinka et al., 2010). Therefore, given the results observed in Chapter 2, the correlation between chromatin status and cytoplasmic organelle DNA integrants is studied in Chapter 4. Organellar DNAs were found to insert preferentially into regions of open chromatin. Thus, it is possible that open chromatin regions are more accessible to the proteins involved DSB repair, necessarily leading to more successful healing of DSBs, sometimes with the incorporation of available mitochondrial or chloroplast DNA fragments.

Transfer of genetic material is not limited to cytoplasmic organellar DNA to the nucleus; migration from chloroplast to mitochondrion also occurs to give *mtpt*s (Kleine et al., 2009). While chloroplast and mitochondrial DNA integrants can create novel functional exon sequences in the nuclear genome (Noutsos et al., 2007), *mtpt*s have been assumed to be "virtually entirely" (Hao and Palmer, 2009) non-functional with only rare instances of replacement of mitochondrial tRNA genes by their plastid equivalents (Kanno et al., 1997;Miyata et al., 1998) or of the creation of new promoter regions (Nakazono et al., 1996). In contrast with this assumption of non-functionality, Chapter

5 demonstrates that chloroplast DNA has created novel mitochondrial genes whose functions in the mitochondrion are unrelated to their plastidic origins.

Together the findings of this thesis indicate that the chloroplast genome provides novel genes and potential mutational material for both nuclear and mitochondrial genomes and that the genetic consequences are influenced by the environment.

It will be interesting to test whether factors causing DSBs in the nuclear genome also increase organllar DNA migration and insertion. Ultraviolet light and chemical agents such as zeocin and bleomycin are good candidates for future work. On the other hand, direct experimental evidence of chloroplast DNA insertion through DSB repair is lacking and it is necessary to design experiments to test this hypothesis. If this phenomenon is confirmed, more light will be shed on the mechanisms of instability of chloroplast DNA integrants in the nuclear genome (Sheppard and Timmis, 2009). The known flanking sequences of chloroplast DNA integration sites may facilitate monitoring the mechanisms of deletion occurring inside novel chloroplast integrants. Chapter 4 demonstrates that cytoplasmic organelle DNA tends to insert into regions of open chromatin. Previous research has shown that recent human *numt*s insert into, or at least survive, more frequently in genes, especially in introns (Ricchetti et al., 2004) and are often located in, or adjacent to, regions of open chromatin (Tsuji et al., 2012). It has been known that open chromatin regions cause genomic DNA segments to be hyper-exposed to interacting molecules (Song et al., 2011), and a chloroplast promoter has been reported to work weakly in the nucleus (Cornelissen and Vandewiele, 1989). Taking these results together suggests that organellar gene may be transcribed directly after migration to the nucleus without acquiring a nuclear promoter.

Reference

Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J., and Palmer, J.D. (2000). Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408**,** 354-357.

Adams, K.L., and Palmer, J.D. (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylog Evol* 29**,** 380-395.

Adams, K.L., Qiu, Y.L., Stoutemyer, M., and Palmer, J.D. (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 USA* 99**,** 9905-9912.

Allen, J.F. (2003). The function of genomes in bioenergetic organelles. *Phil Trans Roy Soc B* 358**,** 19-37.

Allen, J.F., and Raven, J.A. (1996). Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J Mol Evol* 42**,** 482-492.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215**,** 403-410.

Andersson, S.G.E., and Kurland, C.G. (1998). Reductive evolution of resident genomes. *Trends in Microbiology* 6**,** 263-268.

Ayliffe, M.A., and Timmis, J.N. (1992). Plastid DNA sequence homologies in the tobacco nuclear genome. *Mol Gen Genet* 236**,** 105-112.

Barbrook, A.C., Howe, C.J., and Purton, S. (2006). Why are plastid genomes retained in non-photosynthetic organisms? *Trends in plant science* 11**,** 101-108.

Beaton, M.J., and Cavalier-Smitht, T. (1999). Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes. *Proc Biol Sci* 266**,** 2053-2059.

Blanchard, J.L., and Schmidt, G.W. (1995). Pervasive migration of organellar DNA to the nucleus in plants. *J Mol Evol* 41**,** 397-406.

Bleuyard, J.Y., Gallego, M.E., and White, C.I. (2006). Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair* 5**,** 1-12.

Bock, R., and Timmis, J.N. (2008). Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays* 30**,** 556-566.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132**,** 311-322.

Campbell, C.L., and Thorsness, P.E. (1998). Escape of mitochondrial DNA to the nucleus in *yme1* yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. *J Cell Sci* 111 (Pt 16)**,** 2455-2464.

Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Muller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K.J., and Weigel, D. (2011). Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nat Genet* 43**,** 956-963.

Chan, C.Y., Kiechle, M., Manivasakam, P., and Schiestl, R.H. (2007). Ionizing radiation and restriction enzymes induce microhomology-mediated illegitimate recombination in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 35**,** 5051-5059.

Cornelissen, M., and Vandewiele, M. (1989). Nuclear transcriptional activity of the tobacco plastid psbA promoter. *Nucleic Acids Res* 17**,** 19-29.

Covello, P.S., and Gray, M.W. (1992). Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: evidence for RNA-mediated gene transfer. *EMBO J* 11**,** 3815-3820.

Cullis, C.A., Vorster, B.J., Van Der Vyver, C., and Kunert, K.J. (2009). Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? *Ann Bot* 103**,** 625-633.

Cummings, M.P., Nugent, J.M., Olmstead, R.G., and Palmer, J.D. (2003). Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms. *Curr Genet* 43**,** 131-138.

Doolittle, W.F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* 14**,** 307-311.

Ellis, J. (1982). Promiscuous DNA - chloroplast genes inside plant mitochondria. *Nature* 299**,** 678-679.

Fauron, C., Casper, M., Gao, Y., and Moore, B. (1995). The maize mitochondrial genome: dynamic, yet functional. *Trends Genet* 11**,** 228-235.

Fuentes, I., Karcher, D., and Bock, R. (2012). Experimental reconstruction of the functional transfer of intron-containing plastid genes to the nucleus. *Curr Biol* 22**,** 763-771.

Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. *Science* 283**,** 1476-1481.

Guirouilh-Barbat, J., Huck, S., Bertrand, P., Pirzio, L., Desmaze, C., Sabatier, L., and Lopez, B.S. (2004). Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. *Mol Cell* 14**,** 611-623.

Hao, W., and Palmer, J.D. (2009). Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. *Proc Natl Acad Sci USA* 106**,** 16728-16733.

Hazkani-Covo, E., and Covo, S. (2008). Numt-mediated double-strand break repair mitigates deletions during primate genome evolution. *PLoS Genet* 4**,** e1000237.

Henze, K., and Martin, W. (2001). How do mitochondrial genes get into the nucleus? *Trends Genet* 17**,** 383-387.

Hogan, G.J., Lee, C.K., and Lieb, J.D. (2006). Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet* 2**,** e158.

Huang, C.Y., Ayliffe, M.A., and Timmis, J.N. (2003). Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422**,** 72-76.

Huang, C.Y., Ayliffe, M.A., and Timmis, J.N. (2004). Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proc Natl Acad Sci USA* 101**,** 9710-9715.

Huang, C.Y., Grunheit, N., Ahmadinejad, N., Timmis, J.N., and Martin, W. (2005). Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiol* 138**,** 1723-1733.

Huang, X., Kurata, N., Wei, X., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T., Zhang, L., Wang, Y., Feng, L., Furuumi, H., Kubo, T., Miyabayashi, T., Yuan, X., Xu, Q., Dong, G., Zhan, Q., Li, C., Fujiyama, A., Toyoda, A., Lu, T., Feng, Q., Qian, Q., Li, J., and Han, B. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490**,** 497-501.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* 3**,** 109-136.

Kanno, A., Nakazono, M., Hirai, A., and Kameya, T. (1997). Maintenance of chloroplast-derived sequences in the mitochondrial DNA of Gramineae. *Curr Genet* 32**,** 413-419.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32**,** D493-496.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Rev Genet* 9**,** 605-618.

Khush, G.S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 35**,** 25-34.

Kim, A., Song, S.H., Brand, M., and Dean, A. (2007). Nucleosome and transcription activator antagonism at human beta-globin locus control region DNase I hypersensitive sites. *Nucleic Acids Res* 35**,** 5831-5838.

Kirik, A., Salomon, S., and Puchta, H. (2000). Species-specific double-strand break repair and genome evolution in plants. *EMBO J* 19**,** 5562-5566.

Kleine, T., Maier, U.G., and Leister, D. (2009). DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu Rev plant Biol* 60**,** 115-138.

Lang, B.F., Gray, M.W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 33**,** 351-397.

Li, X.Q. (2009). Developmental and environmental variation in genomes. *Heredity* 102**,** 323-329.

Lin, Y., and Waldman, A.S. (2001a). Capture of DNA sequences at double-strand breaks in mammalian chromosomes. *Genetics* 158**,** 1665-1674.

Lin, Y., and Waldman, A.S. (2001b). Promiscuous patching of broken chromosomes in mammalian cells with extrachromosomal DNA. *Nucleic Acids Res* 29**,** 3975-3981.

Lister, D.L., Bateman, J.M., Purton, S., and Howe, C.J. (2003). DNA transfer from chloroplast to nucleus is much rarer in Chlamydomonas than in tobacco. *Gene* 316**,** 33-38.

Lloyd, A.H., and Timmis, J.N. (2011). The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. *Mol Biol Evol* 28**,** 2019-2028.

Lloyd, A.H., Wang, D., and Timmis, J.N. (2012). Single molecule PCR reveals similar patterns of non-homologous DSB repair in tobacco and Arabidopsis. *PLoS One* 7**,** e32255.

Lough, A.N., Roark, L.M., Kato, A., Ream, T.S., Lamb, J.C., Birchler, J.A., and Newton, K.J. (2008). Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. *Genetics* 178**,** 47-55.

Ma, J., and Bennetzen, J.L. (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 101**,** 12404-12410.

Marechal-Drouard, L., Small, I., Weil, J.H., and Dietrich, A. (1995). Transfer RNA import into plant mitochondria. *Methods Enzymol* 260**,** 310-327.

Martin, W., and Herrmann, R.G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant Physiol* 118**,** 9-17.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and 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.

Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., and Kowallik, K.V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393**,** 162-165.

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., and 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, 645-658.

Miller, D.G., Petek, L.M., and Russell, D.W. (2004). Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat Genet* 36, 767-773.

Miyata, S., Nakazono, M., and Hirai, A. (1998). Transcription of plastid-derived tRNA genes in rice mitochondria. *Curr Genet* 34, 216-220.

Mogensen, H.L. (1996). The hows and whys of cytoplasmic inheritance in seed plants. *Am J Bot* 83, 383-404.

Muller, H.J. (1964). The Relation of Recombination to Mutational Advance. *Mut Res* 106, 2-9.

Nakazono, M., Nishiwaki, S., Tsutsumi, N., and Hirai, A. (1996). A chloroplast-derived sequence is utilized as a source of promoter sequences for the gene for subunit 9 of NADH dehydrogenase (nad9) in rice mitochondria. *Mol Gen Genet* 252, 371-378.

Nick Mcelhinny, S.A., Havener, J.M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B.L., Blanco, L., Kunkel, T.A., and Ramsden, D.A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* 19, 357-366.

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

Noutsos, C., Richly, E., and Leister, D. (2005). Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res* 15, 616-628.

Nugent, J.M., and Palmer, J.D. (1991). RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66, 473-481.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., and Mittelsten Scheid, O. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell* 22, 3118-3129.

Price, H.J., and Johnston, J.S. (1996). Influence of light on DNA content of *Helianthus annuus* Linnaeus. *Proc Natl Acad Sci USA* 93, 11264-11267.

Reyes-Prieto, A., Hackett, J.D., Soares, M.B., Bonaldo, M.F., and Bhattacharya, D. (2006). Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr Biol* 16, 2320-2325.

Ricchetti, M., Fairhead, C., and Dujon, B. (1999). Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* 402, 96-100.

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

Richly, E., and Leister, D. (2004). NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol Biol Evol* 21, 1972-1980.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15, 1325-1330.

Salomon, S., and Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J* 17, 6086-6095.

Sangare, A., Lonsdale, D., Weil, J.H., and Grienenberger, J.M. (1989). Sequence analysis of the $tRNA^{Tyr}$ and $tRNA^{Lys}$ genes and evidence for the transcription of a chloroplast-like $tRNA^{Met}$ in maize mitochondria. *Curr Genet* 16, 195-201.

Selosse, M., Albert, B., and Godelle, B. (2001). Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends in ecology & evolution* 16, 135-141.

Sheppard, A.E., Ayliffe, M.A., Blatch, L., Day, A., Delaney, S.K., Khairul-Fahmy, N., Li, Y., Madesis, P., Pryor, A.J., and Timmis, J.N. (2008). Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. *Plant Physiol* 148, 328-336.

Sheppard, A.E., Madesis, P., Lloyd, A.H., Day, A., Ayliffe, M.A., and Timmis, J.N. (2011). Introducing an RNA editing requirement into a plastid-localised transgene reduces but does not eliminate functional gene transfer to the nucleus. *Plant Mol Biol* 76, 299-309.

Sheppard, A.E., and Timmis, J.N. (2009). Instability of plastid DNA in the nuclear genome. *PLoS Genet* 5, e1000323.

Sloan, D.B., Alverson, A.J., Storchova, H., Palmer, J.D., and Taylor, D.R. (2010). Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. *BMC Evol Biol* 10, 274.

Smith, D.R. (2011). Extending the limited transfer window hypothesis to inter-organelle DNA migration. *Genome Biol Evol* 3, 743-748.

Smith, D.R., Crosby, K., and Lee, R.W. (2011). Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. *Genome Biol Evol* 3, 365-371.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., Graf, S., Huss, M., Keefe, D., Liu, Z., London, D., Mcdaniell, R.M., Shibata, Y., Showers, K.A., Simon, J.M., Vales, T., Wang, T., Winter, D., Clarke, N.D., Birney, E., Iyer, V.R., Crawford, G.E., Lieb, J.D., and Furey, T.S. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* 21, 1757-1767.

Stankiewicz, P., and Lupski, J.R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18, 74-82.

Stegemann, S., and Bock, R. (2006). Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant cell* 18**,** 2869-2878.

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

Stern, D.B., and Lonsdale, D.M. (1982). Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature* 299**,** 698-702.

Thorsness, P.E., and Fox, T.D. (1990). Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. *Nature* 346**,** 376-379.

Thorsness, P.E., and Fox, T.D. (1993). Nuclear mutations in Saccharomyces cerevisiae that affect the escape of DNA from mitochondria to the nucleus. *Genetics* 134**,** 21-28.

Thorsness, P.E., and Weber, E.R. (1996). Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. *Int Rev Cytol* 165**,** 207-234.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Rev Genet* 5**,** 123-135.

Tsuji, J., Frith, M.C., Tomii, K., and Horton, P. (2012). Mammalian NUMT insertion is non-random. *Nucleic Acids Res* 40**,** 9073-9088.

Ueda, M., Fujimoto, M., Arimura, S.I., Tsutsumi, N., and Kadowaki, K.I. (2008). Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. *Mol Biol Evol* 25**,** 1791-1793.

Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nat Genet* 15**,** 57-61.

Van Gent, D.C., Hoeijmakers, J.H., and Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2**,** 196-206.

Wang, D., Lloyd, A.H., and Timmis, J.N. (2012a). Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc Natl Acad Sci USA* 109**,** 2444-2448.

Wang, D., Lloyd, A.H., and Timmis, J.N. (2012b). Nuclear genome diversity in somatic cells is accelerated by environmental stress. *Plant signaling & behavior* 7**,** 595-597.

Wang, D., Rousseau-Gueutin, M., and Timmis, J.N. (2012c). Plastid sequences contribute to some plant mitochondrial genes. *Mol Biol Evol* 29**,** 1707-1711.

Wang, D., Wu, Y.W., Shih, A.C., Wu, C.S., Wang, Y.N., and Chaw, S.M. (2007). Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Mol Biol Evol* 24**,** 2040-2048.

Wessler, S., Tarpley, A., Purugganan, M., Spell, M., and Okagaki, R. (1990). Filler DNA is associated with spontaneous deletions in maize. *Proc Natl Acad Sci USA* 87**,** 8731-8735.

Wischmann, C., and Schuster, W. (1995). Transfer of rps10 from the mitochondrion to the nucleus in Arabidopsis thaliana: evidence for RNA-mediated transfer and exon shuffling at the integration site. *FEBS Lett* 374**,** 152-156.

Woischnik, M., and Moraes, C.T. (2002). Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res* 12**,** 885-893.

Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J. (2012). High-resolution mapping of open chromatin in the rice genome. *Genome Res* 22**,** 151-162.