

Université de Montréal

**Un rôle pour les protéines de la famille Whirly dans le
maintien de la stabilité du génome des organelles chez
Arabidopsis thaliana.**

par

Alexandre Maréchal

Département de Biochimie

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Docteur en biochimie

Juillet 2009

© Alexandre Maréchal, 2009

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

**Un rôle pour les protéines de la famille Whirly dans le
maintien de la stabilité du génome des organelles chez
Arabidopsis thaliana.**

présentée par :
Alexandre Maréchal

a été évaluée par un jury composé des personnes suivantes :

Gertraud Burger, Ph.D., président-rapporteur
Normand Brisson, Ph.D., directeur de recherche
David Morse, Ph.D., membre du jury
Gregory Brown, Ph.D., examinateur externe
Mounib Elchebly, représentant du doyen de la FES

Résumé

Le maintien de la stabilité du génome est essentiel pour la propagation de l'information génétique et pour la croissance et la survie des cellules. Tous les organismes possèdent des systèmes de prévention des dommages et des réarrangements de l'ADN et nos connaissances sur ces processus découlent principalement de l'étude des génomes bactériens et nucléaires. Comparativement peu de choses sont connues sur les systèmes de protection des génomes d'organelles. Cette étude révèle l'importance des protéines liant l'ADN simple-brin de la famille Whirly dans le maintien de la stabilité du génome des organelles de plantes.

Nous rapportons que les Whirlies sont requis pour la stabilité du génome plastidique chez *Arabidopsis thaliana* et *Zea mays*. L'absence des Whirlies plastidiques favorise une accumulation de molécules réarrangées produites par recombinaison non-homologue médiée par des régions de microhomologie. Ce mécanisme est similaire au "microhomology-mediated break-induced replication" (MMBIR) retrouvé chez les bactéries, la levure et l'humain.

Nous montrons également que les organelles de plantes peuvent réparer les bris double-brin en utilisant une voie semblable au MMBIR. La délétion de différents membres de la famille Whirly entraîne une accumulation importante de réarrangements dans le génome des organelles suite à l'induction de bris double-brin. Ces résultats indiquent que les Whirlies sont aussi importants pour la réparation fidèle des génomes d'organelles. En se basant sur des données biologiques et structurales, nous proposons un modèle où les Whirlies modulent la disponibilité de l'ADN simple-brin, régulant ainsi le choix des voies de réparation et permettant le maintien de la stabilité du génome des organelles. Les divers aspects de ce modèle seront testés au cours d'expériences futures ce qui mènera à une meilleure compréhension du maintien de la stabilité du génome des organelles.

Mots-clés : Protéines Whirly, plastides, liaison aux acides nucléiques simple-brin, mitochondries, biologie des plantes, stabilité du génome

Abstract

Maintenance of genome stability is essential for the accurate propagation of genetic information and for cell growth and survival. Organisms have therefore developed efficient strategies to prevent DNA lesions and rearrangements. Much of the information concerning these strategies has been obtained through the study of bacterial and nuclear genomes. Comparatively little is known about how organelle genomes maintain a stable structure. This study implicates the single-stranded nucleic acid-binding proteins of the Whirly family in the maintenance of plant organelle genome stability.

Here we report that the plastid-localized single-stranded DNA binding proteins of the Whirly family are required for plastid genome stability in *Arabidopsis thaliana* and *Zea mays*. Absence of plastidial Whirlies favors the accumulation of rearranged molecules that arise through a non-homologous recombination mechanism mediated by regions of microhomology. This mechanism is similar to the microhomology-mediated break-induced replication (MMBIR) described in bacteria, yeast and humans.

Additionally we show that plant organelles can repair double-strand breaks using a MMBIR-like pathway. Plants lacking Whirly proteins accumulate elevated levels of microhomology-mediated DNA rearrangements upon double-strand break induction, indicating that Whirlies also contribute to the accurate repair of plant organelle genomes. Using biological and structural data, we propose a working model in which Whirlies modulate the access of repair proteins and complementary DNA to single-stranded regions, thereby regulating the choice of repair pathways and maintaining plant organelle genome stability. The various aspects of this model will be tested in future experiments which should allow a better understanding of the mechanisms underlying genome stability in plant organelles.

Keywords : Whirly proteins, single-stranded nucleic acids, maintenance of genome stability, plant biology, plastids, mitochondria

Table des matières

Résumé.....	iii
Abstract.....	iv
Table des matières.....	v
Liste des tableaux.....	xi
Liste des figures.....	xii
Liste des abbréviations.....	xiv
Remerciements.....	xvii
CHAPITRE 1. Introduction.....	1
1.1 Régulation de l'expression des gènes dans les organelles des plantes.....	2
1.1.1 L'origine endosymbiotique des organelles.....	2
1.1.2 Systèmes d'importation de protéines dans les organelles de plantes.....	4
1.1.3 Contenu des génomes mitochondriaux et chloroplastiques.....	7
1.1.4 Organisation des génomes mitochondriaux et plastidiques.....	11
1.1.5 Réplication, recombinaison et réparation de l'ADN des organelles.....	13
1.1.5.1 Maintien de la stabilité des génomes des organelles.....	13
1.1.5.2 Réplication.....	14
1.1.5.3 Réparation.....	18
1.1.5.4 Photoréactivation.....	19
1.1.5.5 Réparation par excision de nucléotides.....	20
1.1.5.6 Réparation par excision de bases.....	20
1.1.5.7 Réparation de mésappariements.....	21
1.1.5.8 Les différents mécanismes de recombinaison.....	22
1.1.5.9 Recombinaison homologue.....	23
1.1.5.10 Recombinaison non-homologue.....	26
1.1.5.11 La recombinaison dans les organelles de plantes.....	28
1.1.5.12 Protéines impliquées dans la recombinaison dans les organelles de plantes.....	31
1.1.6 Transcription des gènes des organelles.....	37
1.1.7 Régulation post-transcriptionnelle.....	40

1.1.7.1 Épissage des introns	40
1.1.7.2 Édition de l'ARN	43
1.1.7.3 Maturation des extrémités 5' et 3' et stabilité des ARNs	45
1.1.7.4 Régulation traductionnelle et au niveau de l'assemblage des complexes protéiques	46
1.2 Les Whirlies	48
1.2.1 Découverte d'une nouvelle famille de protéines de liaison au ssDNA.....	48
1.2.2 Localisation intracellulaire.....	49
1.2.3 Structure et conservation des Whirlies de plantes	50
1.2.4 Modifications post-traductionnelles des Whirlies.....	52
1.2.5 Les rôles nucléaires des Whirlies.....	53
1.2.6 Les rôles des Whirlies dans les organelles.....	55
1.2.7 TIF1, un homologue des Whirlies.....	56
1.2.8 Les MRPs et les SALPs, des homologues structuraux des Whirlies	59
1.3 Hypothèses et objectifs de travail	63
Chapitre 2. Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function	64
Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function.....	65
2.1 Mise en contexte	66
2.2 Abstract.....	67
2.2.1 Background	67
2.2.2 Results.....	67
2.2.3 Conclusions.....	67
2.3 Background	68
2.4 Results.....	69
2.4.1 Overexpression of AtWhy2 perturbs mitochondrial function.....	69
2.4.2 General downregulation of mitochondrial gene expression and mtDNA levels in plants overproducing AtWhy2.....	71
2.4.3 AtWhy2 is associated with mitochondrial DNA.....	75

2.5 Discussion	76
2.5.1 Putative functions of mitochondrial Whirlies	77
2.6 Conclusions	79
2.7 Methods	80
2.7.1 T-DNA insertion line characterization	80
2.7.2 Production of plants constitutively overexpressing AtWhy2 and AtWhy2-myc	80
2.7.3 Antibody production	81
2.7.4 DNA, RNA gel blots and oligonucleotides	81
2.7.5 Blue native-polyacrylamide gel electrophoresis (BN-PAGE)	83
2.7.6 Semi-quantitative RT-PCR	84
2.7.7 Real-Time Quantitative PCR	84
2.7.8 Microscopy	85
2.7.9 DNA immunoprecipitation in mitochondria	85
2.8 Authors' contributions	87
2.9 Acknowledgements	87
2.10 References	88
2.11 Figure Legends	92
Chapitre 3. Whirly Proteins Maintain Plastid Genome Stability in <i>Arabidopsis</i>	101
3.1 Mise en contexte	103
3.2 Abstract	103
3.3 Introduction	104
3.4 Results	106
3.4.1 <i>Arabidopsis</i> ptWhirlies are Involved in Chloroplast Biogenesis	106
3.4.2 The Variegated Phenotype of KO1/3 Plants is Maternally-Inherited	107
3.4.3 Amplification of Reorganized ptDNA Regions in Variegated Plants	108
3.4.4 Reorganization of ptDNA is Caused by Recombination Between Short Direct Repeats	108
3.4.5 Illegitimate Recombination is Increased in the Absence of <i>Arabidopsis</i> ptWhirlies	109

3.4.6 Illegitimate Recombination is Increased in Maize Whirly Mutants	110
3.5 Discussion	111
3.5.1 Whirlies Prevent Short Homology-Dependent Illegitimate Recombination	111
3.5.3 A Model for the Variegation in <i>Arabidopsis</i> Plants lacking ptWhirlies	111
3.5.4 Conservation of the Role of ptWhirlies in the Maintenance of Plastid Genome Stability	113
3.5.5 Characteristics of the Short Direct Repeats	114
3.5.6 Whirlies Could Prevent Illegitimate Recombination by Protecting ssDNA.....	114
3.6 Materials and Methods.....	115
3.6.1 Mutant characterization.....	115
3.6.2 Antibody production	115
3.6.3 DNA gel blots	115
3.6.4 Detection of DNA rearrangements by PCR in <i>Arabidopsis</i>	116
3.7 Acknowledgments.....	116
3.8 References.....	116
3.9 Figure Legends.....	120
.....	126
Figure 18. Illegitimate recombination is increased in the absence of <i>Arabidopsis</i> ptWhirlies.....	126
3.10 Supporting Information.....	132
3.10.1 Plastid Whirlies are Associated with Nucleoids	132
3.10.2 Plastidial Whirlies are Associated with ptDNA in a Non-Sequence Specific Manner	132
3.10.3 Supporting Information Materials and Methods.....	133
3.10.3.1 <i>Arabidopsis</i> mutant characterization.....	133
3.10.3.2 Microscopy.....	134
3.10.3.3 Plastid DNA immunoprecipitation	135
3.10.3.4 Electrophoretic mobility shift assays	136
3.10.3.5 Isolation of RNA.....	137

3.10.3.6 Detection of rearranged DNA by PCR in maize	137
3.10.3.7 Oligonucleotides used in this study	137
3.10.4 References	141
3.10.5 Supporting Information Figure Legends	142
Chapitre 4. Structure of DNA-Whirly Complexes Provides Insights into the Maintenance of Plant Organelle Genome Stability	164
4.1 Mise en contexte	166
4.2 Summary	167
4.3 Introduction	167
4.4 RESULTS	169
4.4.1 DSB induce MHMR accumulation in the plastids of <i>Arabidopsis</i>	169
4.4.2 DSB induce MHMR accumulation in the mitochondria of <i>Arabidopsis</i>	171
4.4.4 Most of the detected DNA rearrangements are mediated by microhomologies	171
4.4.5 The crystal structure of StWhy2 reveals how ssDNA is stabilized	172
4.4.6 StWhy2 can bind nucleobases that differ in size and in functional groups	174
4.4.7 StWhy2 can destabilize dsDNA and protect ssDNA against nuclease degradation	175
4.5 DISCUSSION	175
4.5.1 Plant organelles can repair DSB through a MHMR mechanism	175
4.5.2 Mechanism of ssDNA-binding of Whirly proteins	177
4.5.3 Possible mechanisms for Whirly-mediated repair of DSB	179
4.6 Experimental Procedures	180
4.6.1 Plant material	180
4.6.2 Detection of DNA rearrangements	181
4.6.3 StWhy2 cloning, expression and purification	181
4.6.4 Data collection and structure determination	182
4.6.5 Electrophoretic mobility shift assays	183
4.6.6 Accession Numbers	183
4.7 Acknowledgments	183

4.8 References.....	184
4.9 Figures Legends.....	188
4.10 Supplemental Data.....	200
4.10.1 Supplemental Experimental Procedures.....	200
4.10.1.1 DNA destabilization experiment.....	200
4.10.1.2 Mung Bean protection experiment.....	200
4.10.2 Supplemental Figure Legends.....	200
Chapitre 5. Discussion.....	223
5.1 Un rôle pour les Whirlies dans la stabilité du génome des organelles de plantes.....	224
5.1.1 Évidences pour un mécanisme de réparation semblable au MMBIR dans les organelles de plantes.....	224
5.1.2 Quels sont les fonctions moléculaires possibles des Whirlies dans le maintien de la stabilité du génome des organelles ?.....	226
5.1.3 Un modèle pour élucider la fonction des Whirlies dans le maintien de la stabilité du génome des organelles.....	230
5.2 Quelles sont les autres fonctions des Whirlies dans les organelles ?.....	233
5.3 Spécificité de liaison des Whirlies aux acides nucléiques.....	236
5.4 Localisation intracellulaire des Whirlies.....	241
5.5 Maintien de la stabilité génomique, maladies et évolution.....	243
Conclusion.....	246
Bibliographie.....	I

Liste des tableaux

Tableau I. Percentage of plants showing a variegated phenotype.

Tableau II. Characteristics of amplified regions in variegated plants.

Tableau III. Representative recombination events between short direct repeats in *Arabidopsis* plants lacking ptWhirlies.

Tableau IV. Short direct repeat-mediated recombination frequency in various *Arabidopsis* genotypes.

Tableau VIII. DNA Rearrangement Statistics.

Tableau IX. Representative DNA Rearrangements Observed in *Arabidopsis* Organelle Genomes

Tableau X. Data Collection and Refinement Statistics.

Tableau XI. DNA Rearrangement Products Identified in the Mitochondrial and in the Plastid Genomes.

Liste des figures

Figure 1. Sort des gènes d'origine cyanobactérienne chez *Arabidopsis thaliana* et ciblage de leurs protéines.

Figure 2. Génome mitochondrial de la plante modèle *Arabidopsis thaliana*.

Figure 3. Génome plastidique de la plante modèle *Arabidopsis thaliana*.

Figure 4. Modèles pour la réplication de l'ADN plastidique.

Figure 5. Mécanismes de réparation d'un bris double-brin par recombinaison homologue.

Figure 6. Mécanismes de réparation d'un bris double-brin par recombinaison non-homologue.

Figure 7. Structure cristallographique de StWhy1.

Figure 8. Conservation de séquence primaire et de structure entre les Whirlies de plantes à fleurs et la protéine TIF1 de *T. thermophyla*.

Figure 9. Homologues structuraux des Whirlies.

Figure 10. Production of plants with altered expression of *atwhy2*.

Figure 11. Phenotypic characterization of plants with altered expression of *atwhy2*.

Figure 12. Mitochondrial perturbation due to overproduction of AtWhy2.

Figure 13. Gene expression of mtDNA- and nuclear-encoded genes coding for mitochondrial proteins.

Figure 14. Mitochondrial, plastidial and nuclear DNA content in plants with altered expression of *atwhy2*.

Figure 15. AtWhy2 is associated with mitochondrial DNA.

Figure 16. AtWhy1 and AtWhy3 are involved in the biogenesis of chloroplasts.

Figure 17. Variegated plants contain rearranged amplified plastid DNA regions.

Figure 18. Illegitimate recombination is increased in the absence of *Arabidopsis* ptWhirlies.

Figure 19. Model for the appearance of variegation in plants lacking Whirlies.

Figure 20. ssDNA-binding activity of AtWhy1/3 in *Arabidopsis* ptWhirlies mutants.

Figure 21. Co-localization of StWhy1 with chloroplast DNA.

Figure 22. AtWhy1 and AtWhy3 interact with chloroplast DNA.

Figure 23. Long (16 hrs) and short (4 hrs) exposure of panel 2 from Figure 2A.

Figure 24. A gene map of the ptDNA region containing the amplified rearranged regions found in Var A and Var B plants was drawn using the web-based tool OGDRAW (9).

Figure 25. Absence of subgenomic amplicons in *ZmWhy1* maize mutants.

Figure 26. Levels of plastid 23S and 16S rRNA in *Arabidopsis* ptWhirlies mutants.

Figure 27. DNA Rearrangements Accumulate in Organelles Following Treatment with the Gyrase-Inhibitor Ciprofloxacin.

Figure 28. Crystal Structures of StWhy2 in the Free Form and Bound to ERE32 at 2.20-Å Resolution.

Figure 29. StWhy2 Binds ssDNA with Little Sequence Specificity.

Figure 30. Comparison between Mitochondria- and Plastid-Targeted Whirlies.

Figure 31. Phenotypes of WT and Plastid Whirly Deficient *Arabidopsis* Plants Treated With Gyrase Inhibitors Ciprofloxacin and Novobiocin.

Figure 32. Phenotypes of WT and Mitochondrial Whirly Deficient *Arabidopsis* Plants Treated With Gyrase Inhibitors Ciprofloxacin and Novobiocin.

Figure 33. Compensating Interactions Enable StWhy2 to Bind DNA Nucleobases that Differ in Size and in Functional Groups.

Figure 34. StWhy2 can Destabilize dsDNA and Protect ssDNA against Mung Bean Degradation.

Figure 35. StWhy2 and MRP1/2 bind nucleic acids through different mechanisms.

Figure 36. Modèle du glissement réplicatif à l'origine de certains réarrangements génomiques.

Figure 37. Modèle du « microhomology-mediated break-induced replication » (MMBIR).

Figure 38. Modèle du mécanisme d'action des Whirlies dans le maintien de la stabilité du genome des organelles.

Figure 39. Affinité des Whirlies pour les acides nucléiques simple-brin.

Liste des abréviations

ADN	: Acide désoxyribonucléique
ADNmt	: ADN mitochondrial
ADNpt	: ADN plastidique
ADNr	: ADN ribosomal
ADNsb	: ADN simple-brin
AP	: site apurinique ou apyrimidique
ARN	: Acide ribonucléique
ATR	: Ataxia Telangiectasia and Rad3-related
BER	: Base Excision Repair
BIR	: Break-Induced Replication
BN-PAGE	: Blue-Native Polyacrylamide Gel Electrophoresis
CMS	: Cytoplasmic Male Sterility
DSBR	: Double-Strand Break Repair
ERE	: Elicitor Response Element
FISH	: Fluorescence In Situ Hybridization
gARN	: ARN guide
GFP	: Green Fluorescent Protein
HR	: Recombinaison Homologue
Kb	: kilo paires de bases
KO	: Knockout
MHMR	: Microhomology-Mediated Rearrangement

MMBIR	: Microhomology-Mediated Break-Induced Replication
MMEJ	: Microhomology-Mediated End-Joining
MMR	: Mismatch Repair
NEP	: Nuclear-Encoded Polymerase
NER	: Nucleotide Excision Repair
NHEJ	: Non-Homologous End-Joining
NHR	: Recombinaison Non-Homologue
ORF	: Open Reading Frame
Pb	: paires de bases
PCR	: Polymerase Chain Reaction
PEP	: Plastid-Encoded Polymerase
PPR	: Pentatricopeptide Repeat
PT	: Peptide Transit
RDR	: Réplication dépendante de la recombinaison
RPA	: Replication Protein A
ROS	: Espèces réactives d'oxygène
RRR	: Réplication, recombinaison et réparation
RT-PCR	: Reverse Transcriptase Polymerase Chain Reaction
SD	: Shine-Dalgarno
SDS-PAGE	: SodiumDodecylSulfate Polyacrylamide Gel Electrophoresis
SDSA	: Synthesis-Dependent Strand Annealing
SSA	: Single-Strand Annealing
VIGS	: Virus-Induced Gene Silencing
WT	: Wild-type

*L'imprévisible est dans la nature
même de l'entreprise
scientifique. Si ce qu'on va
trouver est vraiment nouveau,
alors c'est par définition
quelque chose d'inconnu à
l'avance. (François Jacob)*

Remerciements

Les travaux présentés dans cette thèse sont le résultat de la collaboration et de nombreuses discussions avec les autres membres de la petite équipe de recherche mise en place par le Dr. Normand Brisson. L'atmosphère particulière qui régnait au laboratoire a grandement favorisé la mise en commun de résultats et la naissance de nouvelles idées. Merci à Normand pour avoir entretenu ce climat propice à la bonne science et pour m'avoir laissé toute la latitude nécessaire à mon développement en tant que scientifique.

Je voudrais remercier particulièrement Jean-Sébastien Parent, sans qui mon doctorat aurait été beaucoup moins agréable et sûrement plus long... J'espère que tu as appris autant de moi que moi de toi. Un gros merci aussi à Laurent Cappadocia, qui a sû me montrer l'importance de la biologie structurale dans la compréhension des mécanismes de la vie et qui à mes yeux incarnera toujours la persévérance et la rigueur scientifique. Les gars, j'espère fortement que nos chemins se croiseront de nouveau...

Je voudrais remercier également tous les stagiaires que j'ai eu le plaisir de diriger et qui m'ont aidé à mener à bien ce projet : Jonathan Poulin, Charbel Abou-Rached, Félix Véronneau-Lafortune, Alexandre Moreau, Paul Tan, Étienne Lepage et Sandra Grondin. Chacun à votre manière, vous m'avez enseigné l'art de la gestion...

Merci au Dr. Franz Lang pour son aide dans le développement de mon projet de recherche et surtout pour m'avoir enseigné un style d'écriture scientifique disons plus incisif. Merci aux Drs Marie Trudel et Charles Després pour avoir fait parti de mon comité de thèse et de mon comité prédoctoral avec le Dr. Gerardo Ferbeyre, les conseils qu'ils m'ont prodigué m'ont été bien utiles. Merci au personnel administratif du département, tout particulièrement à Sylvie Beauchemin, pour leur efficacité et leur gentillesse.

Merci à Mathieu Tremblay pour ses encouragements constants, son aide et son amitié. Merci également à ma famille qui m'a toujours soutenu.

Merci finalement à Pascale, pour avoir rendu tout cela beaucoup plus facile par sa simple présence dans ma vie. Je t'aime mon amour ! Merci à Laurence pour m'avoir rempli de joie depuis qu'elle est là...

CHAPITRE 1. Introduction

1.1 Régulation de l'expression des gènes dans les organelles des plantes

1.1.1 L'origine endosymbiotique des organelles

Les plastides forment un groupe d'organelles aux fonctions variées qui sont dérivées des proplastides immatures situés dans les régions méristématiques de la plante (tissus prolifératifs non-différenciés). À partir de ces proplastides, la plante peut fabriquer des chloroplastes dont la fonction principale est la photosynthèse, des amyloplastides qui peuvent emmagasiner de grandes quantités d'énergie de réserve sous forme d'amidon, des chromoplastes qui contiennent des pigments colorés donnant leurs couleurs caractéristiques aux fruits et aux pétales et fleurs et plusieurs autres types de plastides aux fonctions diverses (revue dans (4)). Certaines formes de plastides matures peuvent également être interconverties entre elles. Comme chez les autres eucaryotes, la fonction principale des mitochondries de plantes est de produire de l'énergie par phosphorylation oxydative. Les mitochondries de plantes sont également très importantes pour la fixation de l'azote, la biosynthèse de plusieurs acides aminés et la production de vitamines co-facteurs (revue dans (5)).

L'hypothèse selon laquelle les plastides seraient le résultat d'une association symbiotique entre deux organismes a été clairement formulée pour la première fois par le botaniste russe Constantin Mereschkowski au début du siècle dernier (6). Une origine endosymbiotique pour la mitochondrie a également été proposée dans les années 1920 entre autres par le biologiste américain Ivan E. Wallin (7). Ces hypothèses se basaient surtout sur l'observation par microscopie de ressemblances structurales entre ces organelles et certains types de bactéries. La découverte que les organelles contiennent également un génome est venue renforcer dramatiquement cette théorie (8, 9). Aujourd'hui, il est généralement accepté par la communauté scientifique que des événements d'endosymbiose distincts sont à l'origine de l'intégration des mitochondries et des chloroplastes dans les cellules eucaryotes (revue dans (10)).

Des analyses phylogénétiques ont permis d'établir que l'ancêtre des mitochondries serait un organisme proche des α -protéobactéries, tandis que les plastides seraient issus d'un organisme semblable aux cyanobactéries (revue dans (11)). Des données très récentes indiquent que le nombre de gènes codant pour des protéines dans le génome des organelles varie selon les espèces et se situe entre 3 à 67 pour les mitochondries et 15 à 209 pour les plastides (12). La majorité des gènes initialement présents dans ces ancêtres eubactériens ont été soit perdus ou encore intégrés au génome nucléaire eucaryote lors d'épisodes massifs de transfert horizontal. Par exemple, le génome plastidique des plantes à fleurs code pour moins d'une centaine de polypeptides, mais on estime que les chloroplastes contiennent environ 3000 protéines différentes (13). De façon similaire, le génome de la mitochondrie des angiospermes encode moins de 60 protéines mais le protéome mitochondrial pourrait contenir jusqu'à 3000 protéines différentes (14). La vaste majorité des protéines permettant la biogénèse et le bon fonctionnement des organelles sont donc encodées par le génome nucléaire.

Le transfert d'ADN des organelles au noyau s'est possiblement effectué sous la forme d'ADN codant réverse-transcrit mais surtout via de longs segments d'ADN génomique (revue dans (15) et (16)). Ce genre d'événement a encore lieu aujourd'hui et est bien illustré par le chromosome II d'*Arabidopsis* sur lequel on retrouve une copie presque entière du génome mitochondrial (17). Ces transferts pourraient être favorisés par la lyse accidentelle d'organelles à l'intérieur de la cellule, par des processus d'autophagie et par des fusions transitoires entre les organelles et le noyau. Un mécanisme privilégié pour l'insertion de l'ADN organellique dans le génome nucléaire semble être la réparation de bris double brins par jonction des extrémités non-homologues (Non-Homologous End-Joining (NHEJ)) (18, 19) (Voir Section 1.1.5.10).

La contribution des organelles au pool génétique des eucaryotes n'est pas négligeable. Ainsi, des analyses bioinformatiques ont permis d'établir qu'environ 18 % des gènes nucléaires d'*Arabidopsis* étaient d'origine cyanobactérienne (20). Il est estimé que seulement la moitié des protéines encodées par ces gènes sont redirigés vers les organelles,

les autres ayant acquis de nouveaux rôles. De plus, il a été démontré que des séquences d'ADN organellique peuvent s'insérer dans des exons de gènes nucléaires et ainsi ajouter de nouvelles régions fonctionnelles à des protéines préexistantes (21). Tout ceci souligne le fait qu'une partie importante du matériel génétique présent dans les ancêtres procaryotes des organelles a été recyclé et intégré aux processus biologiques déjà en place chez la cellule hôte (**Figure 1**). Depuis plus d'un milliard d'années, la cellule hôte et les organelles ont ainsi co-évolué et leur interdépendance est aujourd'hui quasi-totale. Le génome nucléaire est maintenant nécessaire à la fonction des organelles et en retour, les plastides et mitochondries sont devenues essentielles pour la production de l'énergie ainsi que pour la fabrication et l'entreposage de plusieurs métabolites nécessaires à la survie de la cellule.

1.1.2 Systèmes d'importation de protéines dans les organelles de plantes

Afin de demeurer fonctionnel, un gène d'organelle transféré au noyau doit rapidement acquérir une région promotrice reconnue par la machinerie transcriptionnelle eucaryote. De plus, pour que ce gène puisse continuer à jouer un rôle dans les organelles, il doit également posséder une séquence de localisation à son extrémité 5' qui permettra à la protéine qu'il encode d'être dirigée vers l'une et/ou l'autre des deux organelles. Ces séquences transit sont habituellement situées à l'extrémité N-terminale des protéines ciblées aux organelles et peuvent être reconnues par la machinerie d'importation des plastides et/ou des mitochondries. Des séquences de reconnaissance supplémentaires permettent à certaines protéines d'être ciblées plus précisément dans des emplacements particuliers à l'intérieur des organelles telles les protéines membranaires des thylakoïdes chloroplastiques ou de la membrane externe mitochondriale (revue dans (13) et (22)).

Chez les plantes à fleurs, les peptides transit (PT) de base permettant de cibler les protéines dans la matrice des mitochondries ou le stroma des chloroplastes ont plusieurs caractéristiques en commun. Ils s'étendent généralement sur 20 à plus de 70 acides aminés, n'adoptent pas de structure précise en solution, sont riches en acides aminés hydrophobes et

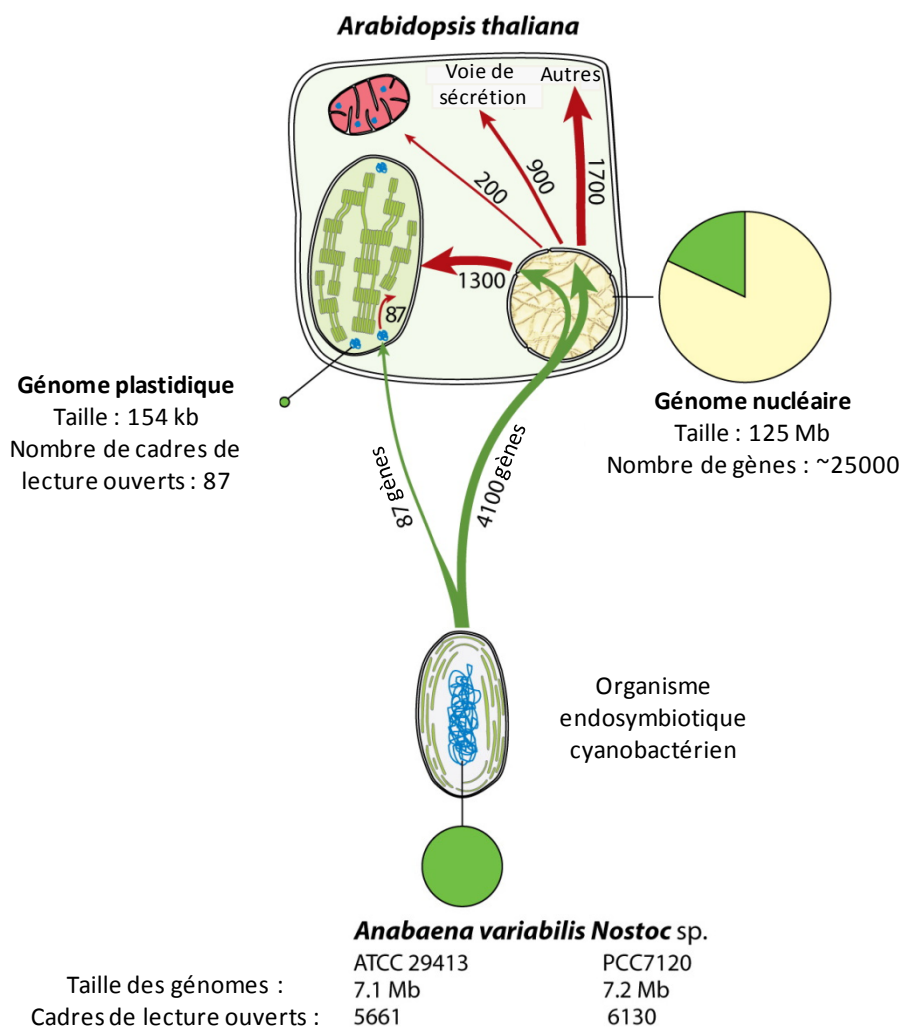


Figure 1. Sort des gènes d'origine cyanobactérienne chez *Arabidopsis thaliana* et ciblage de leurs protéines. Les gènes cyanobactériens sont en vert et les flèches rouges représentent la localisation prédite par analyse bioinformatique des protéines encodées par ces gènes (20, 23, 24). Les flèches vertes indiquent le transfert de gènes d'origine cyanobactérienne dans les différents compartiments de la cellule de plante (Figure adaptée de (16)).

hydroxylés (sérine et thréonine). Ils sont aussi généralement pauvres en résidus acides ce qui leur confère une légère charge positive. De nombreux cas de protéines ciblées aux deux organelles ont été recensés. Plus de 50 cas ont été démontrés expérimentalement et des études bio-informatiques récentes prédisent que plus de 400 protéines auraient une double localisation dans les plantes terrestres (25). La double localisation de ces protéines peut être causée par l'utilisation de différents sites d'initiation de la transcription et/ou de la traduction, menant à la production de protéines ayant des PT distincts. Il a également été démontré que plusieurs PT de protéines doublement localisées pouvaient être directement reconnus par les deux machineries d'importation à cause de leur composition en acides aminés particulière (revue dans (26)). De façon intéressante, un grand nombre de ces protéines doublement localisées sont impliquées dans les mécanismes de régulation de l'expression des gènes dans les organelles soulignant une similarité au niveau de ces processus chez les deux types d'organelles (27).

Il existe plusieurs similarités conceptuelles entre les mécanismes d'importation des protéines mitochondriales et chloroplastiques (revue dans (28) et (29)). La plupart des protéines sont importées de façon post-traductionnelle dans les organelles, bien qu'il ait été démontré que certains polypeptides mitochondriaux pouvaient être importés co-traductionnellement. De façon générale, deux complexes multi-protéiques de translocation contrôlent le passage des protéines à travers les deux membranes entourant les organelles. Les complexes permettant la traversée de la membrane externe sont appelés TOM (Translocase of the Outer Membrane) pour la mitochondrie et TOC (Translocase of the Outer membrane of the Chloroplast) pour les plastides. TIM (Translocase of the Inner Membrane) et TIC sont les complexes de translocation présents dans la membrane interne de la mitochondrie et des plastides respectivement. Ces translocons comportent un grand nombre de sous-unités et ont une origine mixte. Dans le cas de la mitochondrie, la majorité des sous-unités ne semblent pas avoir une origine eubactérienne mais auraient plutôt été créés *de novo* dans un ancêtre commun aux cellules eucaryotes (30). Par opposition, les translocons plastidiques seraient majoritairement d'origine cyanobactérienne (31). Des chaperones moléculaires cytoplasmiques des familles Hsp70 et Hsp90 permettent aux

préprotéines d'adopter une conformation dépliée compatible avec leur importation à travers les pores des complexes de translocation. Récemment, il a également été proposé que certaines chaperones pourraient faciliter l'acheminement des préprotéines au pore de translocation du chloroplaste via des interactions avec certaines sous-unités du complexe TOC (32). Des chaperones et des chaperonines eubactériennes des familles HSP70, 60 et 10 sont également présentes dans les organelles de façon à faciliter le passage dans les pores de translocation. Elles favorisent aussi le repliement fonctionnel des protéines suite à leur importation. Une fois que les protéines ont atteint leur destination finale, les PT devenus inutiles sont clivés par des peptidases localisées dans les différents compartiments des organelles et les protéines matures peuvent alors entrer en fonction. Ainsi, les machineries d'importation des protéines dans les plastides et les mitochondries représentent des cas-école de la coévolution de l'hôte et des organelles : elles sont toutes les deux d'origine chimérique et comportent des parties récupérées des eubactéries et des parties ajoutées suite à l'endosymbiose par la cellule eucaryote. Bien d'autres fonctions biologiques des organelles constituent des mosaïques évolutives puisqu'elles nécessitent d'anciens gènes bactériens issus de transferts horizontaux ainsi que de nouveaux gènes d'origine eucaryote.

1.1.3 Contenu des génomes mitochondriaux et chloroplastiques

D'une taille très variable oscillant entre 367 kb (kilopaires de bases) chez *Arabidopsis thaliana* à 2400 kb chez *Cucumis melo* (melon), le génome mitochondrial des plantes à fleurs (angiospermes) est beaucoup plus grand que celui des animaux qui se situe généralement autour de 16 kb ((33) et revue dans (15)). Cette différence s'explique surtout par un accroissement de la taille des régions non-codantes. En effet, contrairement à l'ADN mitochondrial (ADNmt) des animaux qui est majoritairement composé de gènes (68 % du génome chez l'homme (34)), la plus grande partie de l'ADNmt des angiospermes est non-codante. Chez *Arabidopsis*, par exemple, le génome mitochondrial contient 57 gènes qui, combinés avec leurs introns, représentent seulement 18 % de l'ADNmt. Le reste de ce génome consiste en des duplications (7 %), des cadres de lecture non-conservés (10 %), des

restes de rétrotransposons d'origine nucléaire (4 %) et des séquences d'origine plastidique (1 %). Plus de 60 % de l'ADNmt est donc non-codant et ne présente aucune similarité avec d'autres séquences d'ADN connues incluant les génomes mitochondriaux des autres plantes (2). Ceci implique que l'acquisition et la perte de nouvelles séquences d'ADN par le génome mitochondrial des plantes est un phénomène fréquent. Il a été proposé que ces séquences pourraient être le résultat d'un brassage très rapide de l'ADN nouvellement intégré pour lequel il ne devrait pas exister de pression sélective importante (revue dans (35)). Le génome mitochondrial des plantes contient habituellement trois ARNs ribosomiaux permettant l'assemblage des ribosomes de type eubactérien de ces organelles. Il renferme également plusieurs tRNAs et de nombreuses protéines des ribosomes. Outre ces gènes importants pour la synthèse de l'appareillage traductionnel de la mitochondrie, le reste de la portion codante du génome mitochondrial est composé d'un certain nombre de sous-unités de la chaîne respiratoire (**Figure 2**).

Le génome chloroplastique quant à lui est plus constant en taille, se situant généralement autour de 150 kb chez les plantes à fleurs (154 kb chez *A. thaliana*) (33). Le génome plastidique de la plupart des angiospermes contient quatre régions distinctes : deux portions retrouvées en une seule copie (Large Single Copy (LSC) et Small Single Copy (SSC)) et 2 copies d'une région répétée inversée (Inverted Repeats) qui sépare le LSC du SSC (**Figure 3**). Des trois génomes des plantes à fleurs, le plastome, avec une centaine de gènes dans sa séquence est celui qui a la plus forte densité en gènes (~50 % de séquence codante dans le génome). C'est un génome riche en AT (autour de 60 %) et dans les régions non-codantes du génome cette proportion peut atteindre les 80 % (revue dans (36)). Dans le génome chloroplastique, on retrouve plusieurs gènes de la machinerie de traduction (ARNs ribosomiaux et de transfert, plusieurs protéines des deux sous-unités du ribosome et le gène *infA* qui produit le facteur d'initiation de la traduction IF-1). On retrouve également un opéron encodant différentes sous-unités d'une ARN polymérase de type eubactérien qui est impliquée dans la transcription des gènes du génome plastidique ainsi que le gène *ClpP* encodant une protéase. Le reste des gènes codent pour des protéines impliquées dans l'activité photosynthétique du chloroplaste.

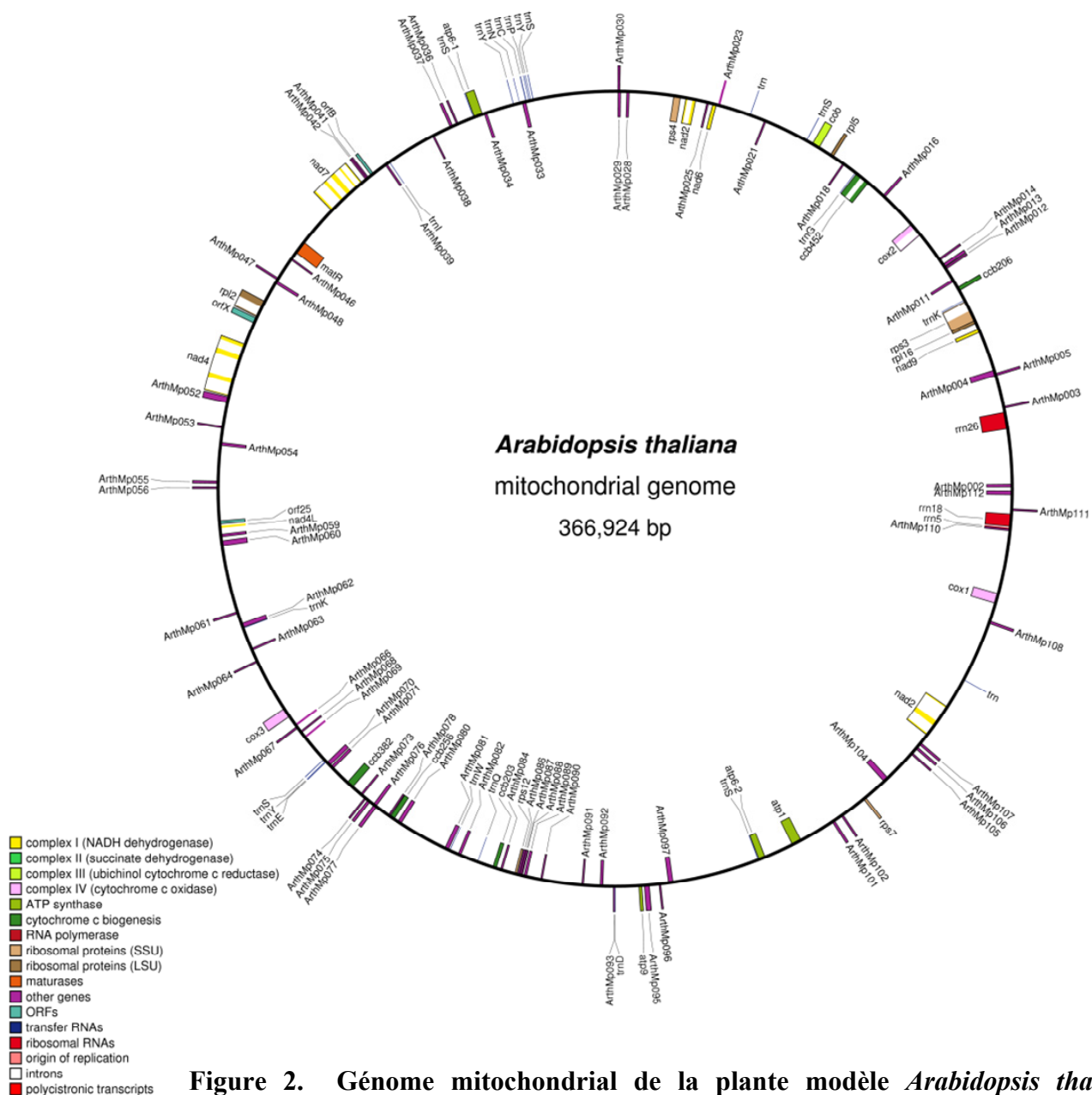


Figure 2. Génome mitochondrial de la plante modèle *Arabidopsis thaliana*. Représentation schématique du génome de la mitochondrie (2) d'*Arabidopsis* réalisée à l'aide du logiciel OGDRAW (3). Les couleurs utilisées pour représenter les différents gènes indiquent leur fonction.

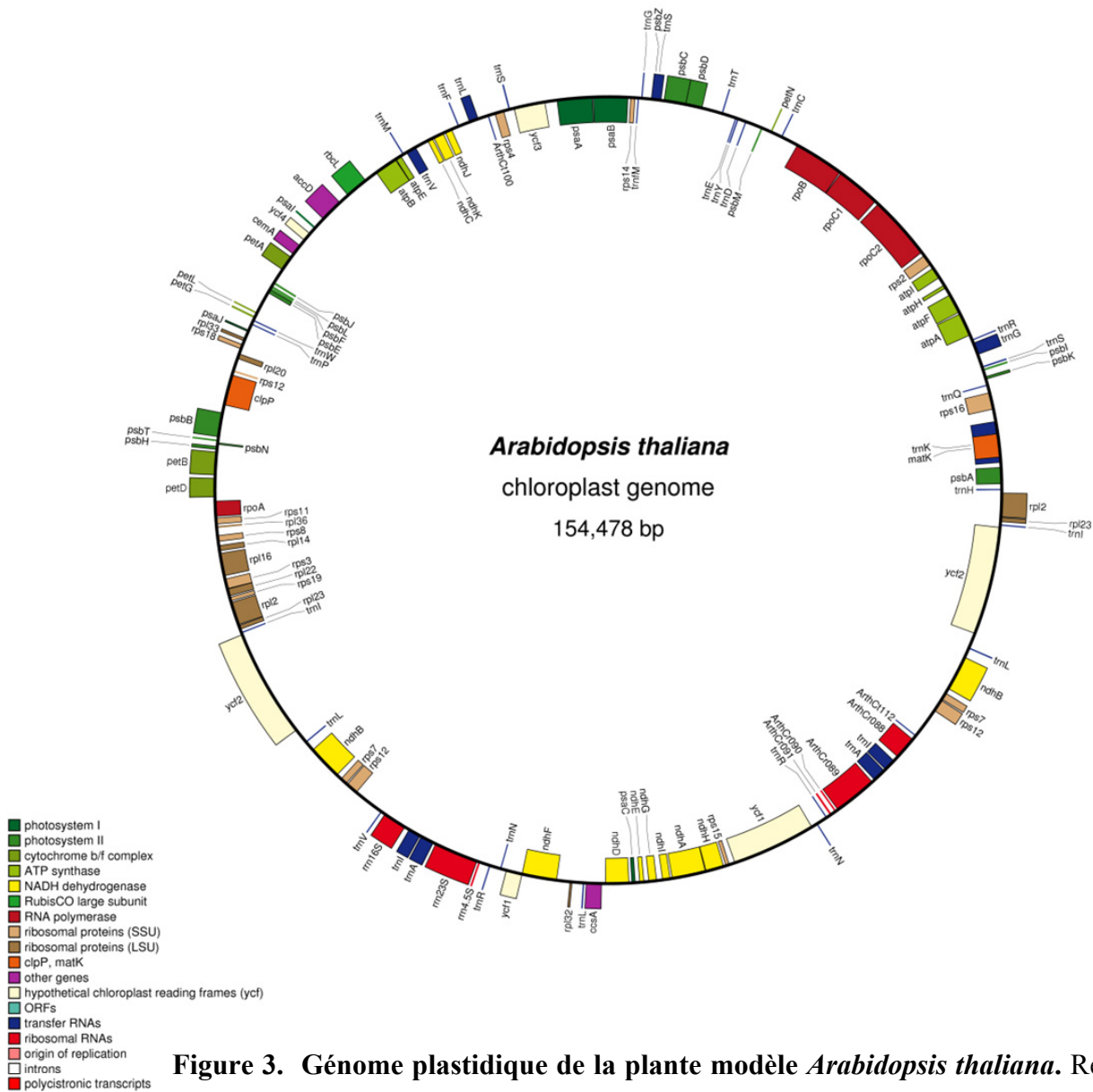


Figure 3. Génome plastidique de la plante modèle *Arabidopsis thaliana*. Représentation schématique du génome du plastide (1) d'*Arabidopsis* réalisée à l'aide du logiciel OGDRAW (3). Les couleurs utilisées pour représenter les différents gènes indiquent leur fonction.

Les gènes restants dans les génomes des organelles sont donc surtout importants pour le maintien de la production d'énergie vitale pour le bon fonctionnement de la cellule.

1.1.4 Organisation des génomes mitochondriaux et plastidiques

On a longtemps pensé que l'ADN des organelles de plantes adoptait uniquement une forme circulaire *in vivo*. En effet, le résultat attendu pour expliquer facilement la propagation du génome des organelles et pour se conformer à ce qu'on savait des génomes bactériens et des mitochondries animales à l'époque était que l'ADN des organelles de plantes devait être retrouvé principalement sous forme d'un cercle monomérique contenant une seule copie du génome. Pour expliquer l'abondance de molécules linéaires dans les préparations d'ADN de chloroplastes (ADN_{pt}) et de mitochondries de plantes observées par microscopie électronique, on invoquait la possibilité que ces molécules linéaires étaient en fait des morceaux de molécules circulaires brisées au cours de la procédure d'isolation de l'ADN (revue dans (37)). Cette situation a perduré longtemps et constitue un bon exemple de la résistance et de la facilité de propagation des dogmes. L'hypothèse des cercles brisés a finalement pu être infirmée lorsque des techniques plus douces d'isolation de l'ADN (électrophorèse en champ pulsé) ont permis d'observer l'ADN d'organelle avec la certitude que celui-ci était intact. Il est maintenant bien établi qu'à l'intérieur des organelles de plantes, l'ADN est présent sous forme d'un mélange de molécules mono- et multi-mériques circulaires et linéaires ainsi que sous forme de structures branchées hautement complexes ((38-42) et revue dans (43, 44)).

Dans les mitochondries et les plastides, l'ADN est généralement assemblé en nucléoïdes, des complexes nucléoprotéiques contenant de multiples copies du génome des organelles (revue dans (45) et (36) et **Figure 21**). Ceci rappelle l'organisation des génomes bactériens. Plusieurs nucléoïdes sont présents dans chaque organelle, leur comportement est dynamique et leur organisation peut changer au cours du développement de l'organelle (46). Lorsque purifiés, les nucléoïdes conservent leur capacité de réplication et de transcription de l'ADN, indiquant qu'ils contiennent tous les facteurs protéiques nécessaires à ces activités. Bien que la purification de nucléoïdes mitochondriaux de

plantes ait été rapportée, la composition protéique de ces complexes demeure largement inconnue (47). Une association des nucléoïdes mitochondriaux avec l'actine a été démontrée, ce qui pourrait suggérer un rôle du cytosquelette dans la répartition de l'ADN lors de la division des mitochondries (48).

Les nucléoïdes des chloroplastes sont beaucoup mieux caractérisés. Il est intéressant de noter que les angiospermes ne contiennent pas les protéines structurales majeures des nucléoïdes retrouvées dans les eubactéries soient les protéines HU (Histone-like U93) et IHF (integration host factor). Cependant, le génome plastidique de certaines algues rouges encode une protéine semblable à HU (49). Chez les angiospermes, de nouvelles protéines associées à l'ADN_{pt} avec une faible spécificité de séquence semblent avoir pris la relève des protéines structurales eubactériennes.

La ferredoxine : sulfite réductase constitue une de ces nouvelles protéines structurales et a ainsi été retrouvée associée avec les nucléoïdes de soja, de pois, et d'*Arabidopsis* (50-53). Le rôle principal de cette enzyme est de transformer le sulfite en sulfide pour favoriser son incorporation dans la cystéine suite à l'action de la cystéine synthase. En plus de son rôle métabolique, cette sulfite réductase peut compacter l'ADN_{pt}, ce qui inhibe la réplication et la transcription de celui-ci (50, 53, 54). La présence d'enzymes métaboliques ayant un second rôle dans le maintien de la structure de l'ADN est un thème récurrent dans la mitochondrie des levures et des mammifères (revue dans (45)). Une telle association pourrait permettre un couplage entre le niveau de métabolisme des organelles et l'activité transcriptionnelle et répllicative des nucléoïdes.

La protéase CND41 (41 kDa Chloroplast Nucleoid DNA binding protein) est également une protéine abondante dans les nucléoïdes. Elle se lie à l'ADN_{pt} de manière non-spécifique et joue aussi un rôle dans la dégradation de certaines protéines au cours de la sénescence (55-57). Une corrélation négative entre la quantité de CND41 et l'abondance de certains transcrits chloroplastiques suggère que cette protéine agirait comme un régulateur négatif de l'expression des gènes dans les plastides.

La protéine PEND (Plastid ENvelope DNA-binding protein) a originellement été décrite comme étant localisée à la membrane interne des chloroplastes durant les premières étapes de leur développement via un domaine transmembranaire et capable de se lier aux nucléoïdes (58). Il s'agit d'une protéine à zipper de leucine (bZIP) capable de dimériser et de lier l'ADN via un domaine basique (59). La surexpression de PEND entraîne une forte perturbation du développement des plastides chez *Brassica napus* ce qui a été attribué au fait que PEND semble avoir une fonction importante dans l'ancrage des nucléoïdes à la membrane interne au début du développement des chloroplastes. Il a été proposé que la surexpression ectopique de PEND empêcherait le relâchement de l'ADNpt nécessaire au bon déroulement des étapes subséquentes de la maturation des plastides (60).

Finalement, la protéine MFP1 est une protéine à structure coiled-coil associée aux membranes thylakoïdes des chloroplastes et se liant à plusieurs régions différentes de l'ADNpt avec des affinités comparables (61). L'activité de liaison à l'ADN de cette protéine semble être régulée négativement lorsqu'elle est phosphorylée par la caséine kinase II plastidique (62). Il a été proposé que cette protéine participerait à l'ancrage des nucléoïdes aux thylakoïdes dans les chloroplastes matures. De façon intéressante, il a récemment été démontré que cette protéine en plus d'être localisée aux plastides, est également ciblée au noyau (63). La fonction nucléaire de MFP1 demeure cependant inconnue mais elle pourrait possiblement ancrer la chromatine à la matrice nucléaire (64). Bien que toutes ces protéines semblent avoir un rôle dans le maintien de la structure des nucléoïdes, les évidences biochimiques accumulées jusqu'à présent auraient besoin d'être complétées par des études de génétique inverse afin de clarifier leur fonction *in vivo*.

1.1.5 Réplication, recombinaison et réparation de l'ADN des organelles

1.1.5.1 Maintien de la stabilité des génomes des organelles

Le maintien de la stabilité du génome est essentiel à la croissance cellulaire et à la survie des organismes. Une déstabilisation du génome résulte en l'accumulation de

mutations et de réarrangements dans l'ADN, ce qui peut rapidement devenir dommageable. L'évolution a donc permis l'élaboration de mécanismes efficaces de prévention et de réparation pour éviter ces réarrangements. Plus particulièrement, les processus de réplication, recombinaison et réparation de l'ADN permettent le maintien de la stabilité génomique face à un éventail impressionnant de stress ayant la capacité de détruire le génome (revue dans (65)). Dans le cas des génomes mitochondriaux et plastidiques, les protéines régulant la réplication, la réparation et la recombinaison (protéines RRR) de l'ADN des organelles sont toutes encodées par le génome nucléaire (revue dans (15) et (36)).

1.1.5.2 Réplication

Les mécanismes de réplication de l'ADN des organelles de plantes demeurent mal compris. La réplication de l'ADN_{mt} en particulier a été très peu étudiée. Des évidences récentes suggèrent l'existence d'un mécanisme de réplication par cercle roulant (rolling circle) pour le génome de cette organelle (66). La réplication de l'ADN_{pt} est un peu mieux comprise. Ainsi, l'observation par microscopie électronique de structures semblables à des boucles D et des cercles roulants, des intermédiaires réplicatifs bien connus, a permis d'établir le modèle classique pour la réplication de l'ADN_{pt} (67). Celui-ci propose l'existence de deux origines de réplication sur le génome monomérique circulaire qui donnent naissance à des boucles D initiées sur des brins opposés. En polymérisant l'ADN, les boucles convergent l'une vers l'autre et sont converties en une bulle répllicative bidirectionnelle. Les deux fourches de réplication de la bulle finissent par se rencontrer à 180° des origines de réplication et à cet endroit, la réplication cesse. Ceci donne naissance à deux duplexes circulaires contenant chacun un bris simple-brin. Ce bris peut être réparé par ligation, produisant un nouveau génome circulaire monomérique ou encore l'extrémité 3'-OH peut être allongée par une ADN polymérase, ce qui résultera en un cercle-roulant capable de produire des génomes concatémériques (**Figure 4**). Ce modèle a lancé une course à la découverte des origines de réplication de l'ADN_{pt}. Le clonage de fragments de génomes couplé à des essais de réplication *in vitro* ont permis d'identifier un certain nombre

d'origines de réplication potentielles dans l'ADNpt de plusieurs espèces de plantes (revue dans (44)). Les études les plus précises à ce jour ont établi que dans le chloroplaste de tabac, deux origines de réplication (OriA et OriB) se trouvaient dans les régions répétées inversées (68). Cependant, l'abolition ou la mutation de ces origines de réplication *in vivo* n'affecte pas substantiellement les niveaux d'ADNpt dans la plante, suggérant que d'autres origines de réplication existent et/ou sont utilisées en absence de OriA et OriB. Il est aussi possible que le mécanisme de réplication de l'ADNpt ne nécessite pas d'origine de réplication particulière (69, 70). Récemment, la découverte de structures complexes branchées dans l'ADN des plastides et des mitochondries a permis de proposer un mécanisme de réplication de l'ADN des organelles dépendant de la recombinaison (RDR). Ce mode réplcatif est semblable à celui utilisé normalement par le bactériophage T4 ou celui de la bactérie lorsque la réponse SOS aux dommages à l'ADN est induite (38, 41, 42, 71, 72) (**Figure 4**). Il est en effet maintenant bien établi que la réplication et la recombinaison ne constituent pas des mécanismes indépendants mais sont plutôt fortement interconnectés (73). Un tel mécanisme pourrait facilement produire des concatémères, des cercles, des molécules linéaires et des structures branchées et complexes qui autrement sont difficiles à expliquer en utilisant le modèle de réplication classique ori-dépendent.

La compréhension définitive du mécanisme de réplication de l'ADN dans les organelles de plante ne pourra être possible que lorsque les acteurs protéiques de ce processus auront été identifiés. Encore une fois, la plupart de ces protéines demeurent inconnues. Chez *Arabidopsis*, deux ADN polymérasés (At3g20540;AtPoly1 et At1g50840l;AtPoly2) de type γ sont ciblées aux plastides et aux mitochondries. Ce type de polymérase est responsable de la réplication et de la réparation de l'ADN dans les mitochondries des animaux et des levures (74). AtPoly1 encode un PT qui la cible aux deux organelles tandis qu'AtPoly2 produit deux PT différents en utilisant des codons d'initiation alternatifs (75). Récemment, deux homologues de ces protéines dans le tabac NtPoll-like1 et NtPoll-like 2 ont également été montrés comme étant ciblés aux deux organelles (76).

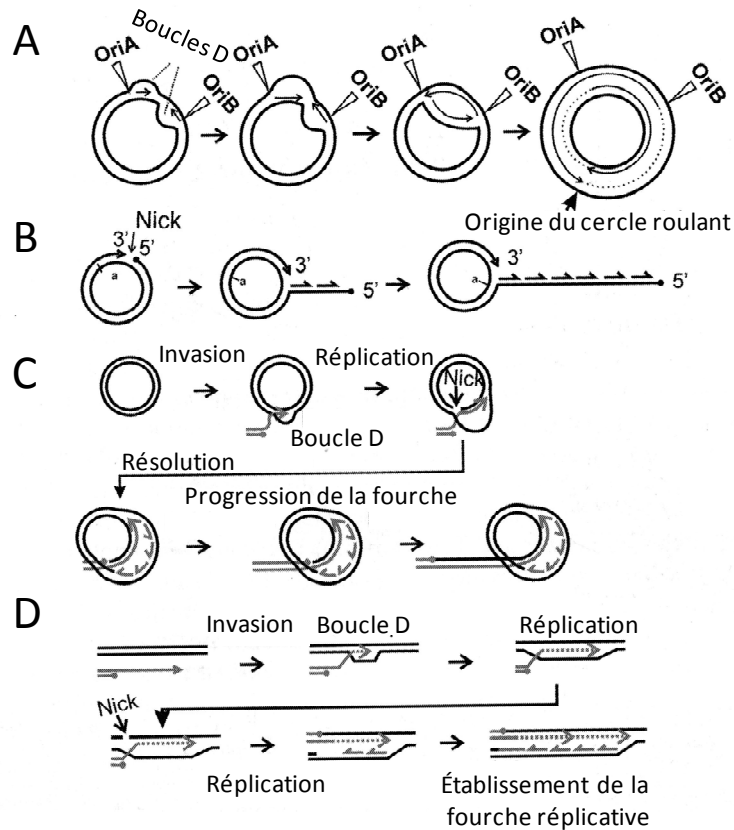


Figure 4. Modèles pour la réplication de l'ADN plastidique. **A.** Modèle des deux boucles D. **B.** Réplication en cercle-roulant initiée par une brisure simpli-brin (nick). **C.** Réplication dépendante de la recombinaison sur une matrice circulaire produisant un cercle contenant une bulle réplivative unidirectionnelle avec une queue. **D.** Réplication dépendante de la recombinaison sur une matrice linéaire produisant une molécule branchée. Pour plus de détail voir la description de chacun des mécanismes dans le texte (Figure adaptée de (44)).

Dans cet article, les auteurs classifient les polymérases d'organelles de plantes comme faisant partie des polymérases de type I plutôt que du type γ . Les polymérases eubactériennes de type I sont considérées comme étant le prototype à partir duquel les homologues eucaryotes nucléaires et mitochondriaux ont évolué. La preuve génétique de l'implication de ces polymérases dans la réplication, la recombinaison ou la réparation de l'ADN des organelles n'a cependant pas encore été produite.

Les gyrases sont des topoisomérases de type II d'origine eubactérienne qui régulent le niveau de surenroulement de l'ADN et qui sont donc importantes pour la réplication, la recombinaison, la réparation ainsi que la transcription. Elles peuvent créer de l'ADN surenroulé négativement ou relaxer l'ADN surenroulé positivement en utilisant l'ATP comme source d'énergie (revue dans (77)). Les gyrases s'assemblent généralement sous forme de tétramères contenant deux sous-unités de chacune des gyrases A et B. Des homologues des 2 types de gyrases ont été trouvés chez les plantes et sont localisés dans les organelles (78, 79). Chez *Arabidopsis*, l'homologue de la gyrase A (At3g10690) encode une protéine ciblée aux deux organelles. Deux homologues des gyrases B sont aussi ciblés aux organelles: At3g10270, dont la protéine est doublement ciblée grâce à l'utilisation de codons d'initiation alternatifs et At5g04130, dont la protéine est localisée uniquement dans les plastides (75, 79). La délétion de ces gènes peut être létale au niveau embryonnaire (gyrase A) ou encore produire des plantes difformes, naines et incapables de se reproduire (gyrases B). Ceci souligne l'importance de ces gènes pour la survie de la plante (79). L'utilisation de la technique du « virus-induced gene silencing » (VIGS) chez *Nicotiana benthamiana*, un proche parent du tabac, a permis de mieux comprendre les rôles des gyrases dans les organelles de plantes (78). Les gyrases A et B de cette espèce possèdent aussi des PT qui ciblent ces protéines aux deux organelles. Une diminution des niveaux d'expression de ces gènes par VIGS perturbe le développement des chloroplastes, ce qui se traduit par la formation de secteurs jaunes et blancs à la surface des feuilles dans les régions où le virus a été injecté. La visualisation des mitochondries de ces secteurs a montré qu'elles avaient aussi une morphologie altérée. Le contenu en ADN des organelles dans les secteurs affectés était plus élevé que dans les régions vertes des feuilles et des analyses par

électrophorèse en champ pulsé ont établi que la structure de l'ADN des chloroplastes était fortement modifiée. Les auteurs de cette étude suggèrent que les gyrases d'organelles seraient importantes pour la partition des nucléoïdes entre organelles-filles ainsi que pour la résolution des structures branchées produites par la réplication et la recombinaison. L'implication des gyrases dans la réplication de l'ADN_{pt} est également supportée par l'inhibition de ce processus par des inhibiteurs spécifiques de l'activité des gyrases (80). À ce jour, les gyrases sont les protéines impliquées dans la réplication de l'ADN des organelles de plante les mieux caractérisées. De façon intéressante, elles constituent également des cibles pharmacologiques de choix dans la lutte contre le paludisme. En effet, l'apicoplaste (plastide non-photosynthétique) de *Plasmodium falciparum*, l'agent pathogène causant la malaria, utilise des homologues de gyrases dans le métabolisme de son ADN et des inhibiteurs de gyrases inhibent la croissance de cet organisme (revue dans (81)).

Le mécanisme exact de réplication de l'ADN des organelles de plantes demeure nébuleux. Il est probable que plusieurs mécanismes différents permettent la passation de l'information génétique aux organelles-filles. D'autres études seront nécessaires avant de pouvoir identifier précisément quels mécanismes sont utilisés, dans quelles circonstances et quels sont les acteurs protéiques qui régulent la réplication.

1.1.5.3 Réparation

L'intérieur des organelles de plantes contient une grande concentration d'espèces réactives d'oxygène (ROS) qui ont la capacité d'endommager l'ADN. L'ADN_{pt} en particulier devrait avoir un taux de mutations élevé étant donnée sa proximité avec la machinerie photosynthétique qui produit de grandes quantités d'oxygène. Selon Muller, la reproduction asexuée des organelles devrait également favoriser l'accumulation de mutations délétères (82). Pourtant, les génomes des organelles de plantes à fleur ont un taux de substitutions synonymes significativement plus faible que celui du génome nucléaire (3 fois et 16 fois plus faible pour les plastides et les mitochondries respectivement) (83, 84). Ceci suggère que des mécanismes très efficaces de réparation de l'ADN ont été mis en

place par les angiospermes, que la réplication de l'ADN s'effectue de manière extrêmement précise dans les organelles ou encore que l'ADN des organelles particulièrement bien protégé contre les ROS.

Les différents types de dommages à l'ADN sont pris en charge par différentes machineries de réparation de l'ADN chez la plante. Les dommages peuvent être provoqués par des sources environnementales externes par exemple suite à l'exposition de la plante à des substances chimiques, des radiations ou même à des toxines fongiques ou bactériennes. Les dommages peuvent également être issus de source endogènes lorsqu'ils sont le résultat d'une grande concentration de ROS à proximité du matériel génétique ou encore d'erreurs commises par les machineries de réplication, recombinaison et réparation de l'ADN (revue dans (85)).

1.1.5.4 Photoréactivation

Plusieurs mécanismes classiques de réparation de l'ADN retrouvés chez la plupart des organismes maintiennent la stabilité du génome nucléaire des plantes mais leur implication dans le maintien de la stabilité du génome des organelles demeure encore incertaine (revue dans (86)). La principale source de dommage environnemental pour les plantes est l'exposition aux rayons UV-B qui est inévitable étant donnée la nécessité de la lumière pour la photosynthèse. Cela produit principalement des dimères de pyrimidines et des photo-produits 6,4 qui perturbent la transcription et la réplication de l'ADN et induisent également des mutations. Durant le processus de photo-réactivation, la lumière est utilisée afin de réparer ces mutations dans les trois génomes de la plante (87). Dans le noyau, les photolyases sont responsables de la photo-réactivation; elles se lient directement au site endommagé et le réparent en utilisant l'énergie de la lumière bleue (300-600 nm) pour réduire les dimères de pyrimidines en monomères. Le chloroplaste de soja (*Glycine max*) semble également utiliser un mécanisme de réparation dépendant de la lumière mais la cinétique de réparation des dommages provoqués par une irradiation aux rayons UV est plus lente que dans le noyau ce qui suggère que des protéines différentes des photolyases

seraient impliquées dans cette voie (88). La photoréactivation de l'ADN mitochondrial et chloroplastique a également été documentée chez *Arabidopsis* mais aucune photolyase encodée par le génome nucléaire de cette plante ne semble posséder un PT qui la ciblerait aux organelles (89). Il est à noter que la réparation de photo-produits dans les génomes des organelles de plantes juvéniles (5 jours) d'*Arabidopsis* s'effectue de façon beaucoup plus lente ($t_{1/2} = 24$ hrs) que dans des plantes matures ($t_{1/2} = 2-3$ hrs). Ceci suggère la présence de mécanismes de tolérance aux dommages comme l'utilisation d'ADN polymérase capables de poursuivre la synthèse en présence de lésions, ou l'utilisation d'autres voies de réparation dans les plantes jeunes comme le changement de matrice de réplication par recombinaison homologue (template-switching) (90).

1.1.5.5 Réparation par excision de nucléotides

La réparation par excision de nucléotides (NER) est un autre mécanisme qui permet de réparer les photo-produits dans le génome nucléaire. Ce type de réparation reconnaît d'abord le site de dommage, déroule la double-hélice de l'ADN, enlève les nucléotides endommagés (20 à 30 paires de bases autour du site) puis repolymérise l'ADN dans la région simple-brin ainsi formée. Une étape de ligation vient ensuite terminer le processus. Le NER peut avoir lieu de façon globale ou encore être couplé à la transcription et réparer les dommages spécifiquement sur le brin en cours de transcription. Bien que la plupart des composantes du NER soient présentes dans les plantes et que son importance dans la réparation du génome nucléaire ait été établie, aucune évidence pour le NER dans les organelles de plantes n'a jusqu'à présent été produite (revue dans (91)).

1.1.5.6 Réparation par excision de bases

La réparation par excision de bases (BER) est un mécanisme qui permet de réparer des nucléotides dont les bases azotées ont été modifiées (e.g. par oxydation, déamination, alkylation...). Certaines de ces bases modifiées peuvent entraîner des mutations car elles sont soit mal reconnues par les polymérase soit non-codantes et entraînent un arrêt de la

machinerie répliquative ou transcriptionnelle. Au cours du BER, les bases endommagées sont d'abord retirées par des glycosylases spécifiques, créant des sites apuriniques ou apyrimidiques (AP). Le sucre restant est alors retiré par une AP endonucléase. Le bris simple-brin sera ensuite comblé par une polymérase et une ligase permettra la recréation d'un double-brin intact. Ce type de BER est appelé « short patch », puisqu'il ne remplace qu'un seul nucléotide. Alternativement, une courte série de nucléotides peut être remplacée dans la version « long patch » du BER. Dans ce type de BER, le bris simple-brin créé par les AP endonucléases est transloqué par polymérisation sur une longueur de quelques nucléotides. Ceci déplace le brin originel qui est alors éliminé par une endonucléase « flap ». La ligation vient ensuite terminer la réparation (revue dans (86)).

Chez les mammifères, de nombreuses protéines impliquées dans la réplication et la recombinaison sont localisées à la fois dans le noyau et dans la mitochondrie. Celles-ci comprennent des glycosylases et des AP endonucléases nucléaires, ce qui indique la présence du BER dans ces organelles (revue dans (92) et (93)). Une activité uracile-ADN glycosylase a été détectée dans les chloroplastes de maïs et très récemment, une activité glycosylase/endonucléase présente dans les chloroplastes d'*Arabidopsis* a été mise en évidence ce qui confirme la présence du BER dans ces organelles (94, 95). Trois gènes (At2g31450, At1g05900 et At2g41460) encodant respectivement deux homologues de la glycosylase endonucléase III (Nth1 et 2) et une AP endonucléase (ARP) ont été démontrés comme co-localisant avec les nucléoïdes de chloroplastes. Cependant, même une plante dans lesquels ces trois gènes avaient été mutés ne montrait pas de sensibilité particulière à différents stress oxydatifs suggérant la présence d'autres mécanismes de réparation dans les plastides (95).

1.1.5.7 Réparation de mésappariements

Des erreurs de réplication ou encore des événements de recombinaison entre des séquences imparfaitement homologues (« homéologues ») peuvent créer des hétéroduplexes d'ADN contenant des bases mésappariées. Le mécanisme de « mismatch

repair » (MMR) est alors utilisé pour corriger ces erreurs. Ce mécanisme est bien caractérisé chez la bactérie *E. coli* où les mésappariements sur le brin naissant non-méthylé peuvent être facilement reconnus par les protéines à activité ATPase MutS et MutL. L'endonucléase MutH vient alors cliver l'ADN sur le brin naissant à proximité de l'erreur, ce qui permet le déroulement de l'ADN par des hélicases. L'ADN simple brin ainsi formé est digéré par des exonucléases et une nouvelle ronde de polymérisation produit un brin intact (revue dans (96)). De nombreux homologues de MutS et MutL ont été retrouvés chez la plupart des eucaryotes et l'existence d'un processus de MMR maintenant la stabilité du génome nucléaire des plantes est bien établie (86, 97-99). Plusieurs protéines du MMR peuvent aussi inhiber la recombinaison homéologue au cours des premières étapes de reconnaissance entre les molécules d'ADN impliquées dans la recombinaison. Ce rôle est important puisqu'il renforce la stabilité du génome en diminuant la probabilité que des régions répétées se recombinent entre elles (100). Finalement, certaines protéines du MMR sont impliquées dans les étapes de recombinaison homologue de la méiose chez les eucaryotes (101). Récemment, la protéine MSH5 a été impliquée dans la formation des chiasmata (lieux d'échange d'information entre chromosomes homologues) lors de la méiose chez *Arabidopsis* (102).

De façon intéressante, MSH1, un homologue de MutS, est localisée dans les plastides et les mitochondries chez *Arabidopsis* (75). L'inactivation de cette protéine provoque des réarrangements importants de l'ADNmt ainsi que l'accumulation de ces régions réarrangées (103-106) (Voir Section 1.1.5.13). Une fonction de maintien de la stabilité du génome mitochondrial a également été retrouvée chez un homologue de MSH1 de la levure *S. cerevisiae* (107). Jusqu'à présent, le rôle classique du MMR dans la réparation de mésappariements n'a toutefois pas été établi dans les organelles de plantes.

1.1.5.8 Les différents mécanismes de recombinaison

Les bris double-brins peuvent survenir suite à une exposition à des agents chimiques ou à des radiations, être provoqués par un stress mécanique ou une digestion par

endonucléase. Ils peuvent aussi survenir suite à la réplication d'une région d'ADN contenant un bris simple-brin ou même être causés par réversion de la fourche suite à un arrêt de la réplication (revue dans (108) et (109)). La réparation de ces bris est extrêmement importante car ils peuvent potentiellement entraîner des réarrangements dans le génome. Dans les organelles de plantes, la réparation des bris double-brins par recombinaison est bien documentée.

Généralement, on distingue deux types principaux de recombinaison. La recombinaison homologue (HR) repose sur l'échange d'information entre deux régions d'ADN homologues quasi-identiques sur une longueur de l'ordre des centaines de nucléotides. Par opposition, la recombinaison illégitime ou non-homologue (NHR) implique la jonction de fragments d'ADN ne comportant que de courtes régions de microhomologie (moins de 30 pb) ou sans homologie (revue dans (110)).

1.1.5.9 Recombinaison homologue

La recombinaison homologue classique comprend plusieurs voies différentes permettant l'échange d'information entre des brins d'ADN contenant des régions d'homologie (**Figure 5**). La conversion génique (gene conversion) est définie comme étant le transfert non-réciproque d'information génétique d'une molécule à son homologue. Ce phénomène est fréquemment associé à un transfert réciproque (crossing-over) d'autres régions sur les molécules d'ADN ayant participé à la conversion génique. La conversion génique peut s'expliquer par deux mécanismes pour lesquels des évidences expérimentales substantielles ont été obtenues. Le premier modèle proposé par Szostak et al est la réparation de bris double-brins par recombinaison homologue (DSBR) qui produit parfois des événements de crossing-over (111). Le second modèle, le « synthesis-dependent strand annealing » (SDSA) ne produit pas de crossing-over.

La conversion génique débute par la création d'un bris double-brin dans une molécule d'ADN. Ce bris est ensuite digéré par des exonucléases 5'→3' pour créer des extrémités 3'-OH libres qui pourront alors envahir une matrice homologue. Dans le modèle

DSBR, la réplication de la nouvelle matrice mène à la production d'une double jonction d'Holliday, une structure branchée composée de 4 hélices doubles brins reliées entre elles par de courtes régions simple-brin. La résolution de ces structures par des endonucléases spécifiques mène à la production de produits de conversion génique avec ou sans événements de crossing-over associés. Dans le modèle SDSA, le brin envahissant la nouvelle matrice est déplacé à mesure que la synthèse d'ADN a lieu, ce qui empêche la formation de jonctions de Holliday. Ceci pourrait être dû à l'action d'hélicases ou de topoisomérases qui détruiraient activement la bulle de réplication ou encore à la formation d'une très petite bulle de réplication facilitant le déroulement du brin nouvellement synthétisé. Suite au SDSA, il y a donc uniquement création de produits de conversion génique sans crossing-over. Un autre mécanisme de réparation par recombinaison homologue consiste en l'hybridation de séquences répétées (> 30 nts) localisées de part et d'autre du bris double-brin, un phénomène appelé le single-strand annealing (SSA) (revue dans (112)). Suite à cette réhybridation, les régions superflues sont enlevées par des nucléases, les brins sont reliés et les bouts manquants sont repolymérisés entre eux. Contrairement au DSBR et au SDSA, ce mécanisme n'est pas parfaitement conservateur car il entraîne la délétion de la région située entre les 2 répétitions. La réplication induite par les bris (break-induced replication ou BIR) est une autre forme de recombinaison homologue. Elle implique l'établissement d'une fourche répliquative complète contrairement aux modèles de DSBR, SDSA et SSA classiques qui ne requièrent que la synthèse du brin avancé (leading strand) (revue dans (113)). Chez la levure et les mammifères, le BIR est utilisé pour la réplication des télomères indépendante de la télomérase ((114) et revue dans (115)). Ce type de mécanisme est présent également chez *E. coli* et semble être très utilisé pour réparer les bris double-brins à une seule extrémité (revue dans (116)). Ces bris sont fréquents et arrivent régulièrement dans le métabolisme normal des cellules notamment suite à l'arrêt de fourches de réplication (117). Il a même été suggéré que la fonction principale de la recombinaison était d'assurer une relance des fourches de réplication arrêtées (118). Une caractéristique de ce type de bris est que suite à l'invasion de l'extrémité 3'-OH, il n'y a pas de possibilité de stabilisation du brin déplacé par hybridation

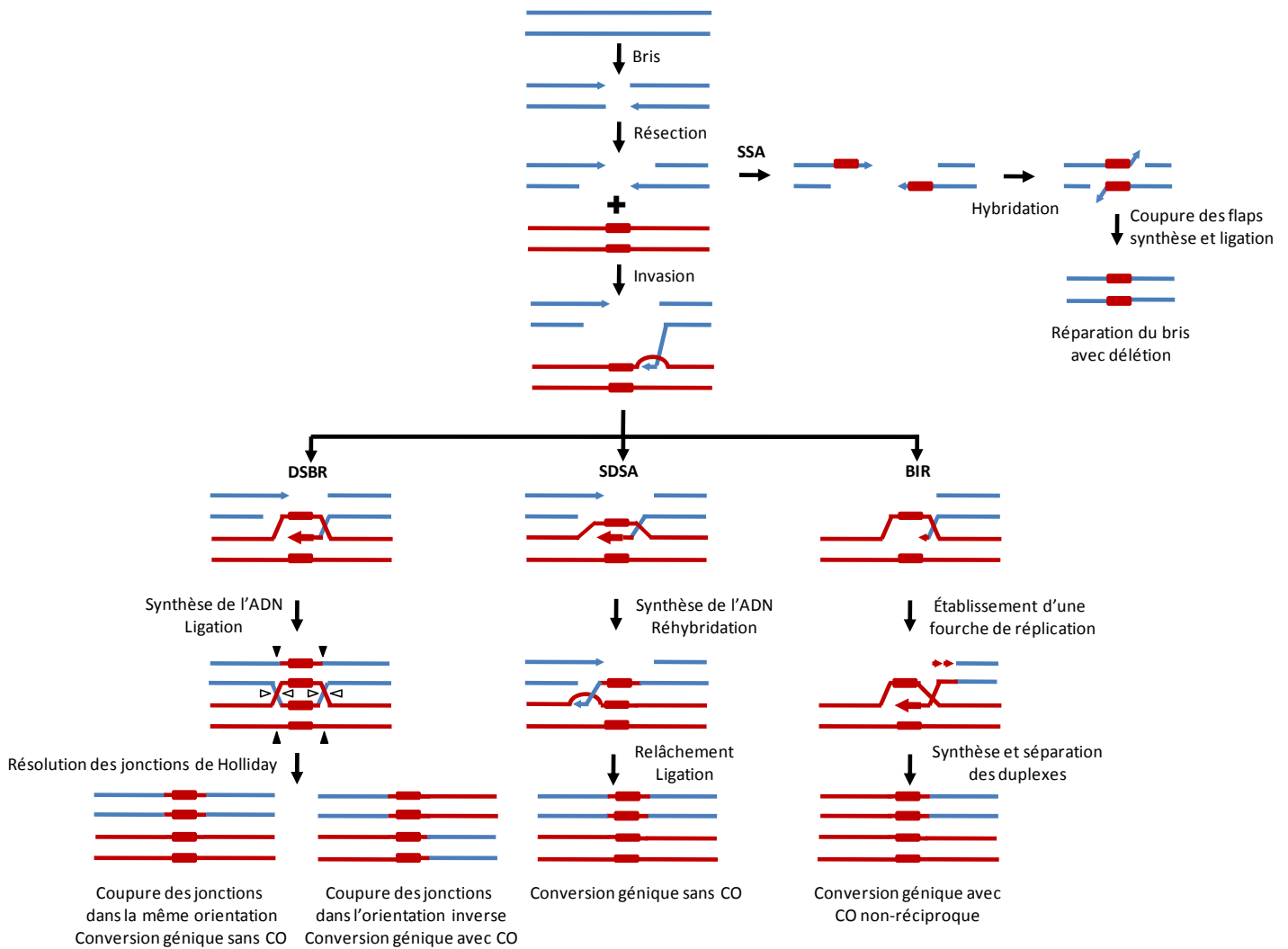


Figure 5. Mécanismes de réparation d'un bris double-brin par recombinaison homologue. SSA, Single-Strand Annealing; DSBR, Double-Strand Break Repair; SDSA, Synthesis-Dependent Strand Annealing; BIR, Break-Induced Replication; CO, Crossing-Over. Pour plus de détails voir la description de chacun des mécanismes dans le texte. Les lignes bleues représentent le duplex receveur endommagé et les lignes rouges représentent le donneur homologue. Les rectangles rouges correspondent à des régions d'homologie.

avec la deuxième extrémité du bris, la synthèse du nouveau brin se poursuit donc jusqu'à l'extrémité de la matrice ce qui produit un événement de crossing-over non-réciproque. Dans un organisme diploïde, ceci peut avoir de graves conséquences sur la viabilité car il y a perte d'hétérozygotie sur toute la partie d'ADN nouvellement répliquée, ce qui rend le génome homozygote pour toutes les mutations présentes dans cette région. Chez l'humain par exemple, la perte d'hétérozygotie est à l'origine de nombreuses maladies génétiques (119).

1.1.5.10 Recombinaison non-homologue

Puisqu'ils ne font pas appel à une molécule d'ADN homologue, les différents types de recombinaison non-homologue (NHR) commettent plus souvent des erreurs lors de la réparation de bris double-brins (**Figure 6**). Dans le noyau des cellules eucaryotes, la NHR est active tout au long du cycle cellulaire, tandis que la HR est favorisée par la présence de chromatides sœurs disponibles suite à la réplication de l'ADN (phases S et G2).

La réparation de bris double brins par jonction des extrémités non-homologues (Non-Homologous End-Joining (NHEJ)) consiste en la jonction directe des deux extrémités d'un bris double-brin. Il y a souvent utilisation de très courtes régions de microhomologies par la machinerie du NHEJ (1-4 nts) Ce type de réparation peut être conservateur ou résulter en la délétion ou l'insertion de très courts fragments d'ADN. Chez les mammifères, le NHEJ est le principal mécanisme de réparation des bris double-brins (revue dans (120)). Le microhomology-mediated end-joining (MMEJ) quant à lui utilise de courtes régions répétées comprises entre 5 à 25 nts pour réhybrider les deux extrémités d'un bris double-brins. Les régions non-complémentaires sont par la suite excisées par des nucléases, l'ADN manquant est synthétisé et des ligases viennent compléter la réparation. Compte tenu que les extrémités des bris double-brins doivent être digérées jusqu'à l'exposition des régions de microhomologie, le MMEJ entraîne une perte d'information plus grande que le NHEJ classique.

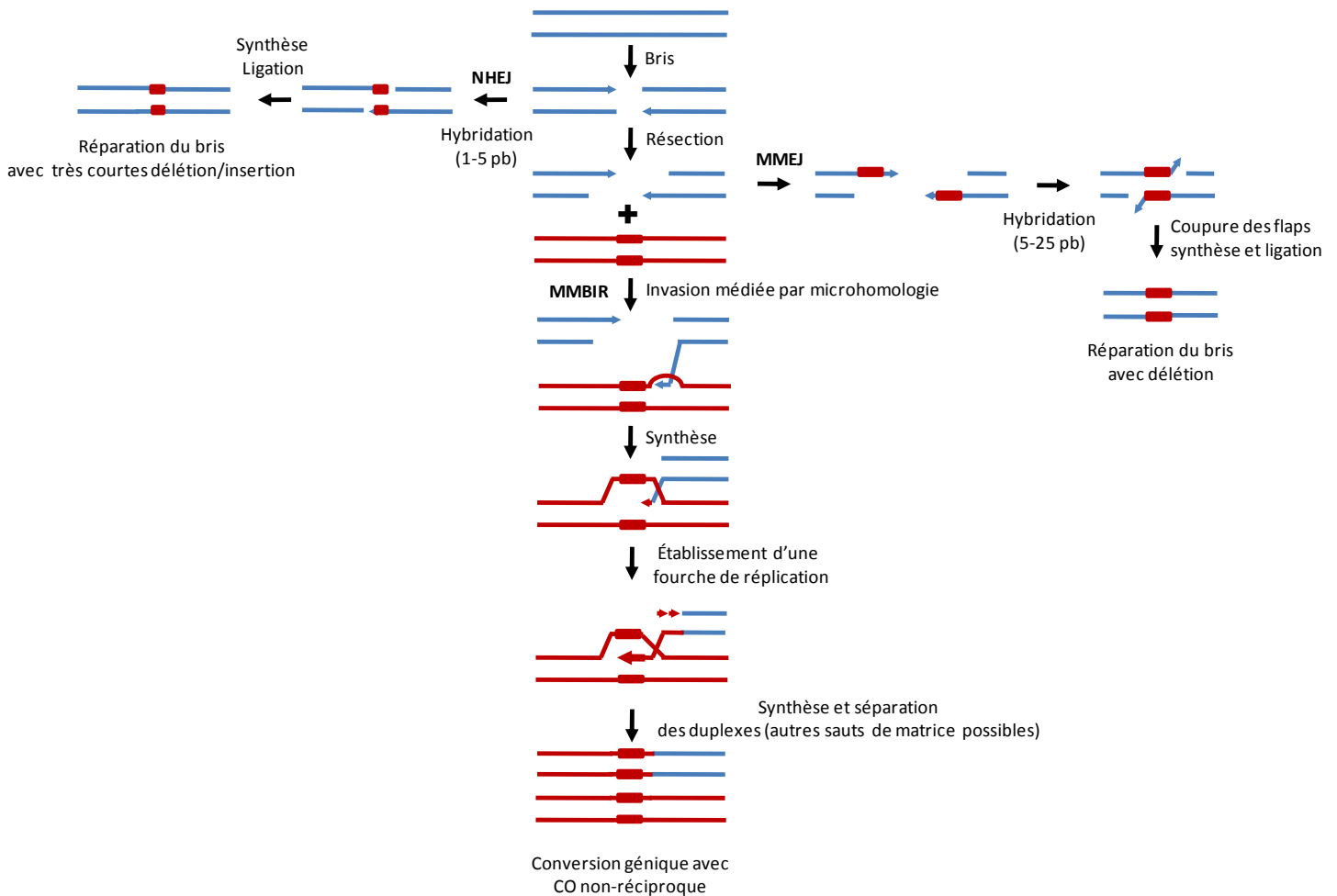


Figure 6. Mécanismes de réparation d'un bris double-brin par recombinaison non-homologue. NHEJ; Non-Homologous End Joining, MMEJ; Microhomology-Mediated End Joining, MMBIR; Microhomology-Mediated Break-Induced Replication. CO; Crossing-Over. Pour plus de détails, voir la description de chacun des mécanismes dans le texte.

Il est à noter que le rôle du MMEJ dans une cellule normale est encore mal compris puisque la plupart des données concernant cette voie alternative proviennent d'organismes mutants pour des protéines clés du NHEJ (revue dans (112)).

Une version modifiée du BIR, le microhomology-mediated break-induced replication (MMBIR), a récemment été proposée pour expliquer certains réarrangements importants dans les génomes de la levure, de la bactérie et des humains (121). Dans ce modèle, un stress réplicatif entraîne une diminution de l'efficacité de la HR et l'arrêt de plusieurs fourches de réplication. Ceci favoriserait l'invasion d'extrémités 3'-OH libres issues de fourches non-fonctionnelles dans des régions microhomologues éloignées. Suite à cette hybridation, des fourches réplicatives seraient établies, ce qui pourrait entraîner des réarrangements importants dans le génome. Souvent, ce genre de saut réplicatif (template switching) semble donner naissance à des fourches dont la processivité est réduite. On assiste alors à une série de sauts à travers le génome ce qui finit par donner naissance à des réarrangements très complexes qui peuvent être à l'origine de certaines maladies chez l'homme (122).

Les différents types de recombinaison semblent co-exister dans une sorte de hiérarchie compétitive qui créerait un équilibre entre la recombinaison conservatrice (qui ne modifie pas la séquence d'origine) lorsque possible et non-conservatrice (qui modifie la séquence d'origine) lorsque des régions substantielles d'homologie ne sont pas disponibles ou que les protéines importantes pour la recombinaison sont limitantes. Avec ces multiples couches de mécanismes de réparation et de recombinaison, les organismes sont équipés pour résister de façon remarquable aux multiples lésions génomiques auxquelles ils doivent faire face. La recombinaison permet également de générer de la diversité génétique lors de la méiose, favorisant ainsi l'adaptabilité des espèces lors de la reproduction sexuée.

1.1.5.11 La recombinaison dans les organelles de plantes

Les plantes utilisent des mécanismes de recombinaison spécifiques pour préserver les génomes de leurs organelles. Comme la plupart des génomes des organelles de plantes à

fleurs sont homoplasmiques (tous les génomes d'une organelle sont identiques) et transmis de façon uniparentale, la fonction principale de la machinerie de recombinaison des organelles consiste à réparer les dommages à l'ADN et non à générer de la diversité génétique (revue dans (44)).

L'ADN_{pt} apparaît remarquablement uniforme lorsqu'il est examiné avec des techniques classiques telles que le buvardage de type Southern. Normalement, dans les plastides de la plupart des angiospermes, le seul événement de recombinaison fréquent et facilement observable a lieu entre les deux longues séquences répétées inversées produisant deux isomères du génome (123). Le développement de techniques de transformation des chloroplastes chez une algue verte d'abord (124) puis chez les angiospermes ensuite (125) a confirmé la présence d'un système actif et efficace de HR dans ces organelles. La longueur minimale d'homologie nécessaire pour que l'ADN transformé soit intégré efficacement par HR dans le génome plastidique semble être autour de 50-100 pb (126, 127). Au cours de l'évolution des plantes, une faible fréquence de recombinaison entre de très petites régions répétées a été également observée (128, 129).

L'utilisation récente de techniques plus sensibles a révélé une complexité insoupçonnée dans le génome plastidique qui suggère que même en conditions de croissance normales, la fidélité de la machinerie de recombinaison n'est pas absolue. Des études d'hybridation fluorescente *in situ* (FISH) ont montré que des molécules réarrangées d'ADN_{pt} représentent de 0,8 à 2 % de toutes les molécules d'ADN_{pt} présents dans les chloroplastes d'*Arabidopsis* et de tabac (40). Ces molécules pourraient être issues de la recombinaison entre de courtes séquences répétées ou encore être produites par clivage et fusion de fragments du génome plastidique.

La réparation de bris par la machinerie de recombinaison dans les plastides a également été documentée. Ainsi, des évidences pour le mécanisme SDSA dans le chloroplaste de l'algue verte *Chlamydomonas reinhardtii* ont été rapportées suite à l'observation de produits de conversion génique sans crossing-over engendrés par le clivage d'une région de l'ARN23S par l'endonucléase I-CreI (130). Une étude plus récente, également chez *C. reinhardtii*, démontre que le chloroplaste de cet organisme peut utiliser 3

voies de réparation de l'ADN par HR (131). Ainsi, cette étude présente de nouvelles évidences pour le SDSA, démontre l'existence de la voie de DSB_R et également la réparation de bris double-brins par single-strand annealing (SSA). La production de délétions dans le génome du chloroplaste via des répétitions aussi petites que 15 pb suggère aussi la présence du MMEJ dans cette organelle.

Le fait que chaque organelle contienne plusieurs copies d'un même génome et que ces génomes soient assemblés sous forme de nucléoïdes pourrait favoriser des mécanismes de réparation conservateurs. En effet, la présence de plusieurs copies sauvages du génome augmente la probabilité de réparation d'un génome muté et diminue d'autant la probabilité de fixation de cette nouvelle allèle dans le génome. Un tel lien entre la polyploïdie et un faible taux de mutation a été établi dans les chloroplastes de tabac dans lesquels la conversion génique s'est avérée très efficace pour la réparation de mutations (132).

L'existence d'un mécanisme de recombinaison homologue du type NHEJ dans les plastides n'a pas encore été démontrée. De fait, des études de mobilisation d'un transposon bactérien dans le génome chloroplastique du tabac semblent indiquer que cette voie ne serait pas utilisée par les plastides. En effet, les auteurs de cette étude ont été incapables de trouver des génomes plastidiques ayant réparé le bris double-brin provoqué par l'excision du transposon via un mécanisme de NHEJ (133). Cependant, il demeure possible que ces résultats soient dus à une particularité du système de transposition utilisé.

Les évidences pour la recombinaison dans le génome mitochondrial des plantes sont également nombreuses. Le génome mitochondrial des angiospermes contient plusieurs séquences répétées relativement longues (> 1 kb) et il est bien établi que la recombinaison homologue intra- ou intermoléculaire entre ces répétitions résulte en une organisation multipartite de l'ADN_{mt} (revue dans (134)). De plus, des sous-génomes réarrangés (sublimons), produits par la recombinaison entre de courtes séquences répétées (> 100 pb) sont présents en quantités sub-stoechiométriques (moins abondants que le génome sauvage) dans les mitochondries normales de plantes. Ceci complexifie d'autant la structure du génome mitochondrial (135, 136). Étant donné qu'il est plutôt riche en séquences répétées

(22 paires de séquences répétées 100 % homologues et > 100 pb dans le génome mitochondrial de l'écotype C24 d'*Arabidopsis*), un strict contrôle sur la recombinaison homologue est important pour la stabilité du génome mitochondrial (2). Effectivement, la recombinaison non-contrôlée dans l'ADNmt des plantes peut entraîner des modifications importantes dans le patron d'expression des gènes mitochondriaux voire même la création de nouveaux cadres de lecture ouverts (ORFs). Ce phénomène s'appelle le shift stoichiométrique. Cela peut induire différents phénotypes incluant la stérilité mâle cytoplasmique (Cytoplasmic male sterility (CMS)), l'apparition de bandes non-chromosomales jaunes/blanches (Non-chromosomal stripes (NCS)) chez les monocotylédons ou encore la variégation (secteurs jaunes/blancs à la surface des feuilles) chez les dicotylédons (137). Le mauvais fonctionnement des chloroplastes dans ces plantes serait possiblement dû à une perturbation des interactions entre les plastides et les mitochondries mutantes. Le CMS quant à lui est souvent associé à une déficience dans la production de pollen fertile lorsque les mitochondries ne fonctionnent pas adéquatement (revue dans (138)).

1.1.5.12 Protéines impliquées dans la recombinaison dans les organelles de plantes

Les mtSSBs

Plusieurs familles de protéines liant l'ADN simple-brin (ADNsb) sont présentes dans les organelles de plantes et sont impliquées dans la régulation de la recombinaison. Les protéines de la famille SSB (single-stranded DNA-binding) ont des rôles démontrés dans la réplication, la recombinaison et la réparation chez la plupart des espèces (revue dans (139, 140)). Elles contiennent généralement un domaine canonique « OB-fold » (oligonucleotide/oligosaccharide-binding fold) qui permet une liaison spécifique aux acides nucléiques simple-brin. Les eucaryotes possèdent un homologue du SSB eubactérien qui est ciblé à la mitochondrie. Chez *Arabidopsis*, deux gènes nucléaires encodant des membres de la famille mtSSB ont été recensés (At4g11060 (mtSSB1) et At3g18580 (mtSSB2)) et

tous deux sont prédits comme étant mitochondriaux. La localisation de mtSSB1 à la mitochondrie a été confirmée en fusionnant son PT à la protéine fluorescente GFP (141). Une double localisation des mtSSBs aux plastides et aux mitochondries ne peut cependant pas être exclue. MtSSB1 est également capable de faciliter l'activité d'échange de brins de la protéine RecA d'*E. coli in vitro*, ce qui suggère qu'elle pourrait jouer un rôle dans la recombinaison. Comme plusieurs SSBs d'autres organismes, le gène *mtSSB1* semble essentiel à la viabilité d'*Arabidopsis* (données non-publiées dans (142)).

Les homologues de RPA

Les complexes RPA (Replication Protein A) sont composés d'une famille de protéines liant l'ADNsb et sont requis pour plusieurs aspects du métabolisme de l'ADN chez les eucaryotes, incluant la réplication, la recombinaison, la réparation de l'ADN ainsi que l'activation de points de contrôle des dommages (revue dans (143, 144)). Normalement, les complexes RPA sont assemblés en trimères de 3 sous-unités de 70, 32 et 14 kDa. La sous-unité de 70 kDa contient 4 domaines OB-fold et constitue la composante majeure de liaison de l'ADNsb du complexe trimérique. RPA32 et RPA14 ne contiennent qu'un domaine OB-fold chacune. Contrairement aux autres eucaryotes qui ne possèdent qu'un complexe RPA, les plantes en contiennent plusieurs. Par exemple, le riz contient 3 différents complexes (RPA-A, B et C) composés de 3 différentes sous-unités de 70 et 32 kDa mais de la même sous-unité de 14 kDa. Un fractionnement cellulaire couplé à une détection des différents types de RPA par Western blot semble indiquer que le complexe RPA-A est localisé au niveau du chloroplaste tandis que RPA-B et C vont au noyau (145). Il est à noter que les protéines des complexes RPA de riz ne contiennent pas de PT évidents et ne sont pas prédites comme étant localisées aux plastides. *Arabidopsis* contient également plusieurs complexes RPA différents, et au moins une des grosses sous-unités est essentielle à la viabilité (RPA70a). Bien que la localisation des différents complexes RPA d'*Arabidopsis* n'ait pas été examinée, il est possible que certains d'entre eux aient un rôle à jouer dans le métabolisme de l'ADN des plastides (146). La grande variété des différents

complexes RPA dans les plantes pourrait constituer une forme d'adaptation à leur mode de vie qui les expose à de fortes doses de rayonnement UV potentiellement mutagène.

Les homologues de RecA

Les recombinaisons eubactériennes RecA et leurs homologues eucaryotes de la famille Rad51 sont des protéines essentielles pour le bon déroulement des premières étapes de la recombinaison (147-149). Elles sont impliquées dans l'alignement des séquences homologues, dans la promotion de l'envahissement du duplex homologue par l'ADNsb et dans la migration des branches (revue dans (150)).

Très peu de gènes de la famille *RecA* ont été trouvés en dehors des procaryotes. La découverte d'un homologue de *RecA* localisé dans le stroma des plastides du pois a donc été quelque peu surprenante (151). La présence d'une activité de transfert de brin semblable à celle de RecA dans les chloroplastes de pois et les mitochondries de soya est venue confirmer la présence d'homologues actifs de RecA dans les organelles de plantes (152, 153).

Le génome d'*Arabidopsis* encode au moins 3 homologues de RecA, soit RecA1 (At1g79050), RecA2 (At2g19490) et RecA3 (At3g10140). Il existe un quatrième homologue (At3g32920), mais comme il est tronqué dans la partie C-terminale responsable de l'interaction de recA avec l'ADN, il s'agit probablement d'un pseudogène (106, 154). La localisation de RecA1, 2 et 3 a été examinée en fusionnant leurs 80 premiers acides aminés à la GFP. Ceci a permis d'établir que RecA1 est localisé aux plastides, RecA3 aux mitochondries et RecA2 est ciblé aux deux organelles. L'implication des homologues de RecA des plantes dans le maintien de la stabilité du génome de la mitochondrie a été démontrée et semble être conservée à travers les espèces. Notamment, les gènes *RecA1* et *RecA2* d'*Arabidopsis* sont essentiels à la viabilité de la plante. *RecA3* ne semble pas essentiel mais une mutation dans ce gène entraîne des réarrangements importants dans l'ADNmt. Ces réarrangements sont dus à la recombinaison homologue de courtes séquences répétées (> 150 pb) dispersées dans le génome mitochondrial. Ces

réarrangements sont réversibles puisque la réintroduction par croisement d'un gène *RecA3* fonctionnel entraîne la perte des molécules recombinées dans la majeure partie de la descendance. De façon étonnante, les plantes *recA3* n'ont pas de phénotype apparent permettant de les distinguer des plantes sauvages, ce qui suggère un certain niveau de flexibilité du génome mitochondrial face aux réarrangements (106).

Un phénomène similaire a été observé chez la mousse *Physcomitrella patens* dans laquelle un homologue de RecA mitochondrial, *PtRecA1*, était muté. Dans ce cas, des événements de recombinaison homologues entre des répétitions plus petites (< 100 pb) a provoqué des réarrangements importants dans le génome mitochondrial. Les différences apparentes de longueur des substrats de recombinaison peuvent s'expliquer par le fait que le génome mitochondrial de *P. patens* ne contient pas de répétitions de plus de 90 pb, et que l'ADNmt d'*Arabidopsis* n'a pas été systématiquement analysé pour observer les possibilités de recombinaison entre des répétitions plus courtes dans les mutants de *RecA3*. Le gène *PtRecA1* semble également important pour la réparation de l'ADNmt puisqu'il est induit transcriptionnellement lorsque la mousse est mise en présence d'agents chimiques causant des dommages à l'ADN (mitomycine C et bleomycine). De plus, la réparation de l'ADNmt est ralentie dans une souche mutante de *PtRecA1* suite à une exposition à l'agent alkylant méthyl-méthane-sulfonate (155). Il est donc clairement établi que les homologues de RecA ciblés aux organelles dans les plantes ont des rôles importants à jouer au moins dans la recombinaison et la réparation de l'ADN.

MSH1

La mutation du locus nucléaire d'*Arabidopsis* dénommé *chloroplast mutator (CHM)* produit un phénotype de variéation qui se traduit par l'apparition aléatoire sur les feuilles de secteurs jaunes/blancs symptomatiques de chloroplastes non-fonctionnels. Ce phénotype hérité de façon non-mendélienne a amené George P. Rédei, le découvreur original de cette mutation en 1973, à proposer que le phénotype serait dû à des mutations dans l'ADNpt de ces plantes (156). Presque 20 ans plus tard, on a découvert une surprenante corrélation entre la variéation des plantes *chm* et l'accumulation de réarrangements d'ADN dans le génome

mitochondrial (103). L'interdépendance entre les chloroplastes et les mitochondries a été invoquée pour expliquer l'effet de mutations mitochondriales sur la fonctionnalité des chloroplastes. Une étude subséquente a permis de mieux caractériser deux réarrangements apparaissant dans l'ADNmt de la descendance d'un croisement entre *chm1-3* (écotype *Landsberg erecta*) et une plante sauvage d'écotype *Columbia-0*. Les réarrangements observés dans les plantes *chm1-3* étaient dus à un événement de recombinaison illégitime réciproque entre de courtes régions répétées du génome mitochondrial (> 200 pb), un type d'événement typique des phénomènes de shifts stoichiométriques (142). Cette accumulation de réarrangements de l'ADNmt n'est pas réversible lorsqu'on réintroduit un allèle fonctionnelle du gène *CHM1* par croisement.

Le clonage et séquençage du gène *CHM* ayant révélé qu'il s'agissait d'un homologue particulier de la protéine de réparation des mésappariements MutS, on a donc renommé ce gène *MSH1* (MutS-homolog 1) (105). (157). L'absence de dégénérescence des plantes mutantes au fil des générations suite à l'accumulation graduelle de mutations ponctuelles suggère que *MSH1* jouerait plutôt un rôle important dans la suppression de la recombinaison illégitime dans la mitochondrie mais qu'il ne serait pas absolument requis pour le MMR. De façon intéressante, les protéines MSH1 d'*Arabidopsis* et de la tomate sont localisées dans les deux organelles, tandis que les homologues du maïs et du soya semblent être ciblés uniquement aux mitochondries (75, 158). Des rôles supplémentaires dans les plastides sont donc possibles pour les protéines de la famille MSH1.

MSH1 contient six domaines conservés dont plusieurs domaines de liaison à l'ADN et de reconnaissance des mésappariements ainsi qu'un domaine de liaison de l'ATP/activité ATPasique et un domaine endonucléase GIY-YIG semblable à celui retrouvé dans les endonucléases ciblées (« homing endonucleases »). Ce dernier domaine possède une fonction enzymatique permettant le clivage des deux brins d'ADN (158). Des mutations ponctuelles dans les domaines ATPase et endonucléase de MSH1 induisent la variéation et des réarrangements de l'ADNmt. En accord avec le rôle établi de *MSH1* chez *Arabidopsis*, une diminution de l'expression du gène *MSH1* par interférence de l'ARN (ARNi) dans la tomate et le tabac produit des réarrangements de l'ADNmt qui corrèlent avec une induction

de la stérilité mâle cytoplasmique. Il est donc possible d'induire un phénotype CMS par manipulation de l'expression d'un seul gène dans des plantes intéressantes d'un point de vue économique (157). De plus, étant donné que les modifications du génome mitochondrial induites de cette manière sont transmises par la mère et permanentes une fois établies, il serait facile de produire des plantes CMS non-transgéniques à partir de cette technique.

La famille des OSBs

Les OSBs (Organellar Single-stranded DNA-Binding proteins) constituent une famille de protéines de plantes récemment découverte et possédant une affinité particulière pour l'ADNsb (159). Chez *Arabidopsis*, cette famille compte quatre membres qui sont localisés dans la mitochondrie (OSB1; At1g47720), dans le chloroplaste (OSB2; At4g20010) ou dans les deux organelles (OSB3; At5g44785) (142). OSB4 (At1g31010) n'a pu être cloné jusqu'à présent et bien que des séquences exprimées (ESTs) existent, sa fonction et sa localisation demeurent inconnues. L'extrémité N-terminale des OSBs contient un PT qui leur permet d'être ciblées aux organelles et un domaine OB-fold est retrouvé au centre de ces protéines. La partie C-terminale des OSBs arbore un nombre variable de domaines PDF spécifiques aux plantes qui de façon surprenante semblent être responsables de la liaison de l'ADNsb à la place du domaine OB-fold. Des plantes mutantes pour les gènes *OSB1*, *OSB2* et *OSB3* ont été obtenues mais un phénotype visible a été observé uniquement dans la descendance des plantes mutantes pour *OSB1*. Celles-ci développent un phénotype de variéation assorti de défauts dans le développement de la plante qui ne deviennent apparents qu'après quelques générations de propagation des plantes mutantes et seulement chez quelques individus. Ces phénotypes corrént avec l'apparition de réarrangements impliquant plusieurs régions du génome mitochondrial des plantes affectées. Ces réarrangements sont provoqués par la recombinaison homologue entre de courtes séquences répétées (> 200 pb). Il est intéressant de noter qu'un seul des produits de HR réciproque s'accumule suffisamment pour être observable par Southern blot. L'autre produit ne peut quant à lui être observé uniquement par PCR car il est présent

en très faible quantité dans la mitochondrie. Ces résultats suggèrent qu'un mécanisme de recombinaison asymétrique est responsable des réarrangements observés dans le génome mitochondrial des plants *osb1*. L'accumulation de réarrangements dans les plants *osb1* est réversible dans un premier temps suite à la réintroduction d'une copie sauvage de *OSB1*. Cependant, lorsque la quantité du produit de recombinaison devient très grande, le phénotype devient irréversible possiblement parce que le génome mitochondrial réarrangé est devenu la composante majeure de l'ADNmt. OSB1 semble donc faire partie du système de surveillance de la recombinaison dans la mitochondrie et agit en réprimant la recombinaison illégitime entre de courtes séquences répétées de l'ADNmt. Il sera intéressant de découvrir les rôles des autres membres de cette famille localisés dans les plastides.

Finalement, bien que plusieurs protéines aient été impliquées génétiquement ou biochimiquement dans la surveillance de la recombinaison, dans la réplication et dans la réparation du génome des organelles, les mécanismes moléculaires à la base de ces processus et les interactions entre les différentes protéines RRR demeurent encore peu explorés.

1.1.6 Transcription des gènes des organelles

Le génome du plastide de la plupart des plantes contient des gènes qui codent pour les cinq sous-unités centrales d'une ARN polymérase de type eubactérien nommée PEP (Plastid-Encoded Polymerase) semblable à celle retrouvée chez les cyanobactéries. Comme pour ses homologues bactériens, la reconnaissance des promoteurs par cette polymérase est facilitée par plusieurs facteurs σ encodés par le noyau (6 différents chez *Arabidopsis* (160, 161)). La PEP partage la responsabilité du contrôle de la transcription du génome plastidique avec plusieurs autres polymérases de type phagique, les NEPs (Nucleus-Encoded Polymerases), qui sont composées d'un seul polypeptide et encodées par le noyau. Une telle polymérase de type phagique s'occupe également de la transcription du génome mitochondrial chez la levure *S. cerevisiae* et chez la plupart des eucaryotes incluant les

plantes ((162, 163) et revue dans (164)). Les mitochondries de plantes ne possèdent cependant pas d'équivalent de la PEP dans leur génome. Chez *Arabidopsis*, on dénombre trois NEPs qui sont ciblées aux plastides (RpoTp;At2g24120), à la mitochondrie (RpoTm;At1g68990) ou aux deux organelles (RpoTmp;At5g15700). Ces différentes localisations soulignent encore une fois une certaine similarité dans les mécanismes de régulation de l'expression des génomes des organelles endosymbiotiques (165, 166). Des essais de transcription *in vitro* ont montré qu'au moins deux des trois NEPs (RpoTp et RpoTm) d'*Arabidopsis* peuvent reconnaître et initier la transcription sur des promoteurs d'organelles sans l'aide d'autres protéines (167). Il pourrait cependant en être autrement *in vivo* où les contraintes topologiques de l'ADN sont différentes, ce qui rendrait nécessaire la présence de facteurs accessoires.

La division des tâches entre ces différentes ARN polymérases est assez complexe. Dans le plastide des plantes à fleurs, la plupart des gènes non-photosynthétiques importants pour les fonctions de base (« housekeeping genes » ; transcription et traduction) possèdent des promoteurs pour les deux types de polymérases (revue dans (168)). Peu de gènes plastidiques sont exclusivement transcrits par la NEP mais ceux-ci incluent l'opéron *rpoB* qui regroupe plusieurs sous-unités de la PEP (169). L'activité et l'abondance de la PEP sont donc dépendantes de la NEP. Les gènes plastidiques importants pour l'élaboration de la machinerie photosynthétique sont quant à eux principalement transcrits par la PEP.

Les promoteurs de gènes transcrits par la PEP dans les plastides des angiospermes contiennent souvent des variantes des éléments -35 (TTGaca) et -10 (TATAaT) retrouvés typiquement dans les promoteurs σ^{70} de la bactérie *E. coli*. Les promoteurs des gènes plastidiques reconnus par les NEPs peuvent être regroupés en trois types. Le type IA contient la séquence consensus YRTA, le type IB contient une séquence GAA conservée en amont du YRTA et le type II ne contient pas les séquences consensus du type I mais possède des déterminants importants encore mal définis situés en aval du site d'initiation de la transcription ((170) et revue dans (171)). Cependant, une seule séquence de type II a été répertoriée jusqu'à présent dans le promoteur du gène *ClpP* de tabac et d'*Arabidopsis*. De

façon intéressante, plusieurs exemples de gènes situés dans des opérons mais contenant quand même leur propre promoteur individuel ont été rapportés. Ceci pourrait augmenter la flexibilité de la régulation de l'expression de ces gènes, en particulier lorsque les conditions environnementales sont changeantes (172). Il ressort de ces études que les promoteurs plastidiques reconnus par les NEPs semblent être plus actifs dans les tissus jeunes et/ou non-photosynthétiques. Suite à ces observations, un modèle dans lequel la NEP serait d'abord la principale polymérase dans les proplastides a été proposé. Ainsi, elle transcrirait les gènes permettant l'élaboration des machineries transcriptionnelles (PEP) et traductionnelles nécessaires à la différenciation en chloroplastes. Suite à cette différenciation, le contrôle transcriptionnel se retrouverait majoritairement mais non-exclusivement sous la gouverne de la PEP (169, 173).

Des études de détermination des sites d'initiation de la transcription de plusieurs gènes mitochondriaux ont établi que le recours à des promoteurs multiples semble être la norme (174). Par exemple, neuf des douze gènes mitochondriaux étudiés contenaient plus d'un site d'initiation de la transcription, et six de ceux-là en utilisaient trois ou quatre différents (174). Les déterminants des promoteurs reconnus par les polymérases de type phagique dans le génome mitochondrial d'*Arabidopsis* sont habituellement assez simples. Plusieurs d'entre eux contiennent le consensus nonanucléotidique CRTAAGAGA ou CGTATATAA (les sites d'initiation de la transcription sont soulignés). Cependant, environ la moitié des promoteurs examinés ne contenaient pas ces séquences consensus. Ces derniers contiennent souvent la séquence centrale CRTA, ATTA ou RGTA, et des purines sont toujours retrouvées en positions +1 et +2 (174). Ces séquences tétranucléotidiques ressemblent aux séquences YRTA retrouvées dans certains promoteurs plastidiques normalement reconnus par une NEP (175).

Le rôle des multiples promoteurs retrouvés dans le génome mitochondrial des plantes demeure inconnu, bien que plusieurs possibilités aient été envisagées. Les transcrits issus de fleurs et de feuilles utilisent les mêmes sites d'initiation de la transcription ce qui suggère qu'une utilisation de promoteurs spécifiques à certains tissus est peu probable

(174). Des travaux supplémentaires visant à identifier les cibles transcriptionnelles particulières des différentes polymérases et les circonstances/conditions environnementales dans lesquelles les multiples promoteurs sont préférentiellement utilisés devraient nous éclairer davantage.

Un manque de corrélation entre l'activité transcriptionnelle des gènes et le niveau stationnaire d'accumulation de leurs ARNs et protéines a été observé dans les plastides et les mitochondries de plantes (176-178). Ainsi, bien que la transcription joue un rôle important dans la régulation de l'expression des gènes dans les organelles de plantes, il semblerait qu'une bonne partie de cette régulation s'effectue au niveau post-transcriptionnel lors de la maturation et de la traduction des transcrits.

1.1.7 Régulation post-transcriptionnelle

Suite à la transcription, l'ARN des gènes d'organelles de plantes doit passer par plusieurs étapes de maturation avant de pouvoir être finalement convenablement traduit en protéine fonctionnelle. Ces étapes comprennent l'épissage des introns, l'édition ainsi que la maturation des extrémités. La stabilité des différents ARNs ainsi que la régulation de leur traduction sont également des points de contrôles importants de l'expression des gènes d'organelles.

1.1.7.1 Épissage des introns

Contrairement aux génomes eubactériens desquels ils sont issus, les génomes des organelles de plantes sont riches en introns qui doivent être épissés au cours de la maturation des ARNs (revue dans (179, 180)). La plupart des introns de la mitochondrie sont arrangés en *cis*, i.e. ils joignent les deux exons consécutifs. Quelques gènes ont cependant des exons dispersés dans le génome mitochondrial et leurs introns sont épissés en *trans* pour former l'ARN mature. Les introns des organelles de plantes sont majoritairement du groupe II qui utilise un mécanisme d'épissage différent de celui des introns du groupe I. L'épissage des introns de type I requiert une guanosine exogène qui

attaque d'abord la jonction intron-exon en 5'. Le 3'-OH libéré va ensuite attaquer la jonction intron-exon en 3', ce qui lie les exons ensemble et relâche l'intron. Les introns de groupe II sont capables de s'auto-épisser *in vitro* et utilisent généralement un groupement 2'-OH d'un résidu adénosine de l'intron pour attaquer la jonction en exon-intron 5'. Cette attaque produit un lasso et l'épissage peut alors être complété par l'attaque de la jonction 3' par l'extrémité 3'-OH de l'exon, ce qui joint les deux exons ensemble et relâche l'intron contenant un lasso. Une molécule d'eau peut également réaliser la première attaque nucléophile lors de l'épissage des introns de groupe II, mais dans ces cas le lasso n'est pas observé.

Plusieurs facteurs protéiques sont impliqués dans l'épissage. Certains sont encodés par les introns eux-mêmes, les maturases, et d'autres sont présents dans le génome de l'hôte. Les maturases sont plutôt rares dans le génome des organelles de plantes; une maturase matR est retrouvée dans le génome mitochondrial des plantes et une maturase matK est présente dans le génome plastidique. D'autres maturases semblables à celles des introns de groupe II sont encodées par le noyau et sont conservées chez les plantes. Leur implication dans la maturation des ARNs des organelles n'a cependant pas été démontrée (181). Plusieurs gènes nucléaires encodant des protéines importantes pour l'épissage des introns plastidiques ont été identifiés principalement chez le maïs. De façon intéressante, plusieurs de ces gènes ont des homologues chez *Arabidopsis* qui encodent des protéines possiblement localisées dans la mitochondrie (182). Dans le maïs, plusieurs protéines telles CAF1, CAF2, CRS1, CRS2 et PPR4 ont été impliquées de façon convaincante dans l'épissage des introns plastidiques. En effet, des mutations dans les gènes encodant ces protéines diminuent l'efficacité de l'épissage de certains ARNs, et ces protéines ont été retrouvées en association avec les ARNs qu'elles épissent (182-185). D'autres protéines identifiées par leur interaction ou leur similarité de séquences avec ces facteurs ont également été impliquées dans la maturation des ARNs de plastides, levant ainsi le voile sur une machinerie complexe d'épissage et de maturation des ARNs (186-190). Ainsi, au moins une de ces protéines a pu être impliquée dans l'épissage de 16 des 17 introns de groupe II retrouvés dans le génome plastidique du maïs.

Un thème récurrent de ces divers facteurs de maturation est leur capacité de lier l'ARN, souvent de façon spécifique avec certains introns précis. Plusieurs de ces protéines (CRS1, CAF1, CAF2, CFM2, CFM3) contiennent en effet une ou plusieurs copies du domaine CRM (Chloroplast RNA-binding and ribosome Maturation) de liaison à l'ARN. On retrouve ce domaine également chez les bactéries où il est impliqué dans la maturation des ribosomes (182), ainsi que chez plusieurs membres de la famille des protéines PPRs (Pentatricopeptide Repeats). Cette famille semble être particulièrement importante lors de plusieurs étapes de la maturation, de la régulation de la stabilité et de la traduction des ARNs d'organelles chez les plantes ((191) et revue dans (192)). Les PPRs contiennent des répétitions en tandem d'un motif dégénéré de 35 acides aminés semblable à celui retrouvé dans les protéines TPR (Tetratricopeptide Repeat). Chez les plantes à fleurs, les PPRs forment une très grande famille de protéines ciblées aux plastides et aux mitochondries, soit plus de 450 membres chez *Arabidopsis*. La plupart de ces protéines semblent posséder une activité de liaison à l'ARN séquence-spécifique bien que les bases structurales de cette reconnaissance ne soient pas encore connues. L'implication des PPRs dans la maturation des ARNs d'organelles semble être conservée puisqu'un tel rôle leur a été attribué pour la première fois chez la levure *S. cerevisiae* (193). Des membres de cette famille ont été impliqués dans à peu près toutes les étapes de la vie des ARNs d'organelles de plantes incluant la transcription (194), la traduction (195), la régulation de la stabilité des ARNs (196), le clivage des ARNs (197), l'épissage (185) et l'édition (198). De façon intéressante, plusieurs protéines PPRs ont aussi été identifiées comme étant des supprimeurs de la stérilité mâle cytoplasmique et il a été démontré que certaines d'entre elles peuvent réguler l'expression des gènes recombinés associés à ce phénotype au niveau de la stabilité et de la traduction de leurs ARNs (199-203).

Bien que plusieurs protéines impliquées dans l'épissage aient été identifiées, leur fonction dans la promotion de la maturation des ARNs demeure nébuleuse. Il est aussi généralement pris pour acquis que les introns de groupe II sont autocatalytiques et que les protéines sont surtout présentes pour favoriser le repliement des introns en une conformation active. Dans le cas particulier de l'épissage en *trans*, les facteurs protéiques

pourraient promouvoir l'assemblage des fragments d'introns entre eux. Davantage d'expériences sont nécessaires pour décortiquer les composantes et les modes de fonctionnement de la machinerie d'épissage des ARNs d'organelles.

1.1.7.2 Édition de l'ARN

L'édition est un processus découvert initialement dans les mitochondries de trypanosomes. Il consiste en l'insertion/délétion ou encore en la modification de certains nucléotides de molécules d'ARN (204). Bien que ce processus n'ait pas encore été recensé chez les procaryotes, il apparaît que la maturation des ARNs d'organelles de plantes implique une étape d'édition (205-208). Le processus d'édition est beaucoup plus répandu dans les mitochondries que dans les plastides (environ 400-500 sites dans les mitochondries et autour de 30 sites dans les plastides (209, 210). L'édition des ARNs plastidiques et mitochondriaux de plantes consiste principalement en la modification de Cs en Us, et moins fréquemment de Us en Cs (revue dans (164, 180)). De façon intéressante, ces modifications entraînent des changements qui permettent une meilleure conservation évolutive au niveau de la séquence protéique. L'édition peut également créer des codons d'initiation/stop ou éliminer des codons stop ce qui permet d'assurer un démarrage et un arrêt adéquats de la traduction.

L'édition dans les organelles de plantes est régie par deux types d'éléments, soit des éléments *cis* qui entourent les sites d'édition et sont reconnus par la machinerie d'édition ainsi que des facteurs protéiques qui agissent en *trans* pour effectuer l'édition elle-même. Les éléments en *cis* demeurent assez mal définis en partie à cause de la difficulté à réaliser des essais d'édition *in vitro*. On s'accorde cependant pour dire que les nucléotides en 5' des sites d'éditions semblent être plus importants que ceux en 3'. L'absence de séquences consensus claires entre les sites d'édition suggère que plusieurs protéines différentes pourraient être requises pour la reconnaissance des différents sites ou même que chaque site pourrait avoir sa protéine attitrée. Étant données leur grande variété et leur capacité de liaison aux ARNs, les PPRs sont des candidats idéaux pour la reconnaissance des sites

d'édition. La première démonstration de l'implication d'une protéine PPR impliquée dans l'édition décrivait l'importance de CRR4 pour l'édition du codon d'initiation du gène *ndhD* encodant une sous-unité la NADH-déshydrogénase (211). Dans ce cas, il a été démontré que CRR4 interagissait directement avec le site d'édition de *ndhD*, indiquant la capacité des PPRs d'agir comme facteur de reconnaissance de motifs spécifiques d'ARN (212). Depuis, plusieurs autres membres de la famille PPR ont été impliqués dans l'édition de différents sites dans les génomes des deux organelles (198, 213-215). D'autres protéines, les cpRNPs (Chloroplast RiboNucleoProteins), ont été repêchées par association avec de petits ARNs contenant des sites d'édition. Bien que leur liaison aux ARNs soit plutôt non-spécifique, l'immunodéplétion de la protéine cp31 de tabac dans des essais d'édition *in vitro* provoque néanmoins une inhibition de l'édition suggérant que les cpRNPs sont impliquées dans ce processus (216). L'implication des cpRNPs dans l'édition et la stabilité des transcrits de chloroplastes a récemment été confirmée génétiquement chez *Arabidopsis* (217).

Bien qu'on ait identifié plusieurs protéines impliquées dans l'édition, le mécanisme de modification des bases demeure mal compris. Deux modèles sont actuellement proposés. Dans le premier, les PPRs agissent comme agents de reconnaissance des sites d'édition et recrutent une activité de déamination encore non-identifiée permettant de changer les Cs en Us. Dans le second modèle, il a été proposé qu'une sous-catégorie des PPRs possédant un motif DYW renferment une région catalytique semblable à celle retrouvée dans les cytidines déaminases, ce qui leur conférerait une fonction enzymatique en plus de leur capacité de reconnaissance des sites à éditer (218). D'autres domaines attachés aux différents PPRs pourraient également jouer un rôle catalytique dans le processus d'édition. Récemment, des essais *in vitro* ont révélé une activité endoribonucléase chez des protéines PPRs à domaine DYW. Toutefois aucune activité cytidine déaminase n'a pu être révélée (219). De plus, les domaines DYW de certaines protéines PPR se sont avérés non-nécessaire pour leur activité d'édition *in vivo* (213). De futures études permettront sûrement de déterminer quel modèle prévaut dans la machinerie d'édition des organelles chez les plantes.

1.1.7.3 Maturation des extrémités 5' et 3' et stabilité des ARNs

Chez les plastides et les mitochondries de plantes, les gènes peuvent être transcrits seuls ou co-transcrits avec d'autres gènes à l'intérieur d'opérons (revue dans (164, 168)). Les transcrits des ARNs de transfert et ribosomiaux doivent également passer par plusieurs étapes de maturation avant de pouvoir former la machinerie de traduction. La génération des extrémités 5' et 3' des ARNs matures à partir de ces différents précurseurs requiert plusieurs activités endo- et exo-nucléasiques capables de reconnaître certaines structures particulières dans leurs ARNs cibles (revue dans (164, 220)). Ces ARNases sont parfois ciblées aux deux organelles, telle la RNR1 (RNase II) chez *Arabidopsis*, tandis que d'autres sont spécifiques aux plastides ou aux mitochondries. Cette double localisation suggère encore une fois une certaine ressemblance dans les processus de maturation des ARNs d'organelles (221, 222). La maturation des extrémités 3' et 5' des ARNs peut entraîner des changements dans le niveau de traduction des différentes formes d'ARN, donc affectant la quantité des protéines qu'ils encodent (223). Mis-à-part les différentes ARNases, d'autres protéines permettent de préciser les endroits ciblés par la digestion, limiter la dégradation des ARNs par les nucléases ou même effectuer certains clivages. Notamment, des membres de la grande famille des PPRs ont été impliqués dans la maturation de plusieurs transcrits plastidiques et mitochondriaux (197, 201, 224, 225). Un modèle élaboré récemment propose que les PPRs viendraient se lier à des endroits spécifiques des différents ARNs et constitueraient ainsi des barrières qui limiteraient la progression des nucléases, ce qui spécifierait les extrémités matures des différents ARNs dans les organelles (226).

Le niveau stationnaire d'un ARN donné est défini par un équilibre entre son niveau de transcription et de dégradation. Ainsi, des mécanismes de dégradation des ARNs jouent un rôle important dans la régulation de l'expression des gènes des organelles de plantes. Certains éléments structuraux dans les transcrits sont impliqués dans le maintien de la stabilité de ces ARNs. Par exemple, la présence de tiges-boucles à l'extrémité 3' de différents transcrits peut influencer la demi-vie de ces ARNs en protégeant contre la

dégradation exonucléasique (227-229). Des facteurs agissant en *trans* de la famille des PPRs ont aussi un rôle à jouer dans la régulation de la durée de vie des ARNs mitochondriaux et plastidiques probablement également en limitant la progression des nucléases sur les ARNs (196, 201, 226).

Contrairement aux ARNs messagers nucléaires, les ARNs matures des organelles de plantes n'ont pas de queue poly-A constitutivement ajoutée à leurs extrémités 3'. Parfois, des queues de poly-A sont ajoutées en 3' des transcrits mais dirigent plutôt ces ARNs pour la dégradation et constituent un autre moyen de réguler l'expression des gènes des organelles (230-233). Suite à une diminution de l'expression de l'ARNase PNPase (Poly-Nucleotide Phosphorylase), 80 % des transcrits des gènes mitochondriaux *atp9* et *orfB* non-matures contenaient une queue de poly-A variant entre 5 et 17 nucléotides (221). Ceci suggère que la PNPase et/ou d'autres exonucléases reconnaîtraient ces queues et digéreraient les ARNs ainsi marqués, un modèle d'action qui semble également être valable dans les plastides (221, 234, 235).

1.1.7.4 Régulation traductionnelle et au niveau de l'assemblage des complexes protéiques

Très peu est connu sur la régulation de la synthèse des protéines dans les organelles de plantes, mais celle-ci semble s'effectuer surtout au niveau de l'initiation de la traduction (revue dans (236) et (237)). Plusieurs ARNs de plastides possèdent une séquence similaire à celle de SD à proximité de leur codon d'initiation mais le rôle de ces séquences n'est pas clair. En effet, malgré l'origine eubactérienne de la machinerie de traduction des organelles, les processus de traduction actuels ne semblent pas toujours utiliser une séquence Shine-Dalgarno (SD) pour initier la lecture de l'ARN par le ribosome. La mutation de plusieurs séquences semblables à des SD chez *Chlamydomonas* a eu des effets variables sur la traduction (238, 239). De plus, dans la mitochondrie, il ne semble pas y avoir de séquences SD en amont de la plupart des AUG initiateurs mais certaines séquences conservées ont été observées et pourraient avoir un rôle dans l'initiation de la traduction (240). En plus de ces

régions similaires aux séquences SD procaryotiques, d'autres éléments *cis*, souvent retrouvés sous la forme de tiges-boucles dans la région 5' non-traduite, influencent positivement ou négativement la traduction des ARNs dans les plastides (revue dans (241)).

Parmi les facteurs *trans* influençant la traduction chez les plantes à fleurs, on retrouve encore une fois des protéines PPRs (195, 202, 224). Il a été récemment suggéré que la proximité d'un PPR lié à un site de reconnaissance sur l'ARN pourrait permettre à la partie 5' non-traduite d'être disponible sous forme simple-brin, ce qui faciliterait la liaison du transcrit par le ribosome. Le PPR pourrait également interagir directement avec le ribosome et ainsi faciliter son interaction avec la séquence initiatrice de la traduction (226). Il semblerait également que les PPRs peuvent inhiber la traduction de certains ARNs en empêchant leur recrutement dans les polysomes (202).

Plusieurs des complexes protéiques présents dans les organelles contiennent à la fois des sous-unités encodées par le noyau et d'autres par le génome des organelles. L'assemblage de ces complexes constitue la dernière étape avant l'entrée en fonction des protéines encodées dans le génome des organelles. Cet assemblage pourrait également constituer un moyen de réguler l'expression des gènes des organelles, du moins dans la mitochondrie des plantes. Dans les mitochondries de cellules d'*Arabidopsis* en culture, on a constaté qu'il y avait très peu de coordination entre les variations des niveaux d'expression des gènes mitochondriaux et ceux encodés par le noyau en réponse à un retrait du sucrose dans le milieu de culture (242). De façon générale, les niveaux d'ARNs des gènes mitochondriaux ne varient que marginalement tandis que les niveaux d'ARNs des gènes nucléaires sont réduits. Suite à une expérience d'immunobuvardage, on a pu constater que les niveaux des protéines mitochondriales et nucléaires sont tout deux réduits suite au retrait du sucrose mais les protéines nucléaires semblent être plus fortement diminuées. Ce résultat suggère que les sous-unités nucléaires des complexes pourraient devenir limitantes suite à ce stress. Comme l'assemblage des complexes de la chaîne respiratoire s'effectue correctement et que la stoechiométrie entre les différentes sous-unités est respectée, les auteurs de cette étude ont suggéré que la coordination des génomes nucléaires et

mitochondriaux s'effectuait aussi de façon posttranscriptionnelle, plus particulièrement lors de l'assemblage des complexes multiprotéiques (242).

1.2 Les Whirlies

1.2.1 Découverte d'une nouvelle famille de protéines de liaison au ssDNA

Le laboratoire du Dr. Normand Brisson a établi que le niveau d'expression du gène nucléaire *PR-10a* (*Pathogenesis-related 10a*) est fortement augmenté suite à l'activation de la réponse de défense contre les pathogènes dans la pomme de terre (*Solanum tuberosum*). Ce gène est induit dans les pétioles, les feuilles, les tiges et les tubercules en réponse à des blessures, après traitement avec des éliciteurs tels que l'acide arachidonique ou encore suite à une infection par l'oomycète *Phytophthora infestans* (243-247). Des analyses de la séquence promotrice du gène *PR-10a* ont démontré que l'élément de réponse à l'éliciteur (ERE) est une séquence activatrice de la transcription de *PR-10a* qui est à la fois suffisante et nécessaire à l'induction de ce gène suite à un stress (247, 248). Le ERE est reconnu spécifiquement par le facteur PBF-2 (*PR-10a* Binding Factor 2) et l'activité de liaison à l'ADN de ce facteur corrèle avec l'accumulation du transcrit de *PR-10a* (247-249). Des expériences d'immunoprécipitation de la chromatine ont également permis d'établir que PBF-2 se lie au promoteur de *PR-10a in vivo* dans les tubercules de pomme de terre suite à une blessure ou à un traitement à l'acide arachidonique (250, 251). La surexpression de PBF-2 dans des protoplastes augmente aussi le niveau d'expression d'une construction rapporteur de façon ERE-dépendante (251). Toutes ces données suggèrent fortement que PBF-2 constitue un activateur transcriptionnel du gène *PR-10a* en conditions de stress. De façon intéressante et contrairement à la plupart des facteurs de transcription, PBF-2 a une affinité particulière pour l'ADNsb, ce qui a permis de proposer que ce facteur pourrait stabiliser l'ADN en conformation ouverte ce qui pourrait faciliter la transcription (249). La purification de PBF-2 par affinité à l'ADNsb a permis d'isoler une protéine de 24 kDa. Le séquençage de cette protéine a finalement permis le clonage du gène codant pour une sous-

unité de PBF-2 (depuis renommée *StWhy1*) qui s'est avéré être le premier membre de la famille Whirly (249).

1.2.2 Localisation intracellulaire

Bien que *StWhy1* ait été initialement décrit comme étant un activateur transcriptionnel nucléaire, une analyse de sa structure primaire révèle que cette protéine contient un PT. De plus, des logiciels de prédiction de la localisation subcellulaire indiquent qu'elle devrait être ciblée aux plastides (252). Par ailleurs, cette protéine est toujours observée *in vivo* à une taille de 24 kDa indiquant que son PT a été clivé (la séquence d'ADN codant complète prédit une protéine d'environ 30 kDa). Ceci suggère qu'elle serait bien ciblée aux plastides, du moins dans un premier temps. Chez *Arabidopsis*, on retrouve trois membres de la famille Whirly, soit *AtWhy1*, *AtWhy2* et *AtWhy3*. Des essais d'importation utilisant des organelles purifiées et des études de microscopie confocale à fluorescence utilisant des protéines fusionnées à la GFP ont établi que *AtWhy1* et *AtWhy3* sont ciblées aux plastides tandis que *AtWhy2* est dirigé vers la mitochondrie (253). De façon générale, toutes les plantes à fleurs dont les génomes ont été complétés ou pour lesquelles des programmes de séquençage d'ESTs importants ont été entrepris semblent posséder au moins un Whirly plastidique et un mitochondrial. Récemment, la double localisation nucléaire et plastidique de *HvWhy1*, un Whirly de l'orge (*Hordeum vulgare*), a été démontrée par immunolocalisation et par sous-fractionnement cellulaire suivi d'une détection par immunobuvardage (254). De façon intéressante, des constructions dans lesquelles on a retiré le PT d'*AtWhy1* ou de *HvWhy1* pour ensuite les fusionner à des fragments de la protéine fluorescente YFP ont permis de constater qu'en absence de PT, ces protéines étaient localisées principalement au noyau. Les Whirlies endogènes semblent cependant beaucoup plus abondants dans les plastides que dans le noyau et le mécanisme de localisation des Whirlies au noyau demeure encore inconnu (Voir la section Discussion).

1.2.3 Structure et conservation des Whirlies de plantes

On retrouve des homologues non-équivoques bien conservés des Whirlies dans les plantes à fleurs ainsi que chez les algues vertes (252, 255). Les Whirlies des angiospermes contiennent tous un PT à leur extrémité N-terminale ainsi qu'un domaine central hautement conservé, le domaine Whirly. Ce dernier est important pour la liaison à l'ADNsb et également pour la tétramérisation de la protéine (**Figure 7**). Des analyses de filtration sur gel des complexes ADN-Whirlies et des études de cristallographie aux rayons X ont effectivement permis d'établir que la protéine s'assemblait sous forme de tétramères (255). Finalement, l'extrémité C-terminale des Whirlies comporte un domaine acide-aromatique conservé qui a été initialement proposé comme modulant la capacité de liaison à l'ADN des Whirlies. La résolution de la structure tridimensionnelle du domaine Whirly a révélé un nouveau type de repliement permettant de lier les acides nucléiques simple-brin. Chaque protomère contient deux feuillets β perpendiculaires l'un à l'autre et comptant chacun quatre brins β agencés en conformation antiparallèle. Ces feuillets β ont été proposés comme constituant la surface de liaison de l'ADN. Une et deux hélices α se trouvent respectivement à la suite du premier et du second feuillet β . Dans le tétramère, les hélices alphas sont retrouvées au centre de la structure et sont importantes pour les interactions inter-protomères et la stabilité du tétramère en général (255). Lorsqu'on observe la structure adoptée par un Whirly sous forme tétramérique, on remarque que les feuillets bêtas rayonnent comme les pales d'une hélice à partir du centre alpha-hélical de la protéine. C'est cette structure quaternaire particulière des Whirlies, semblable à un vire-vent (« whirligig »), qui est à l'origine du nom de cette famille de protéines de liaison à l'ADNsb.

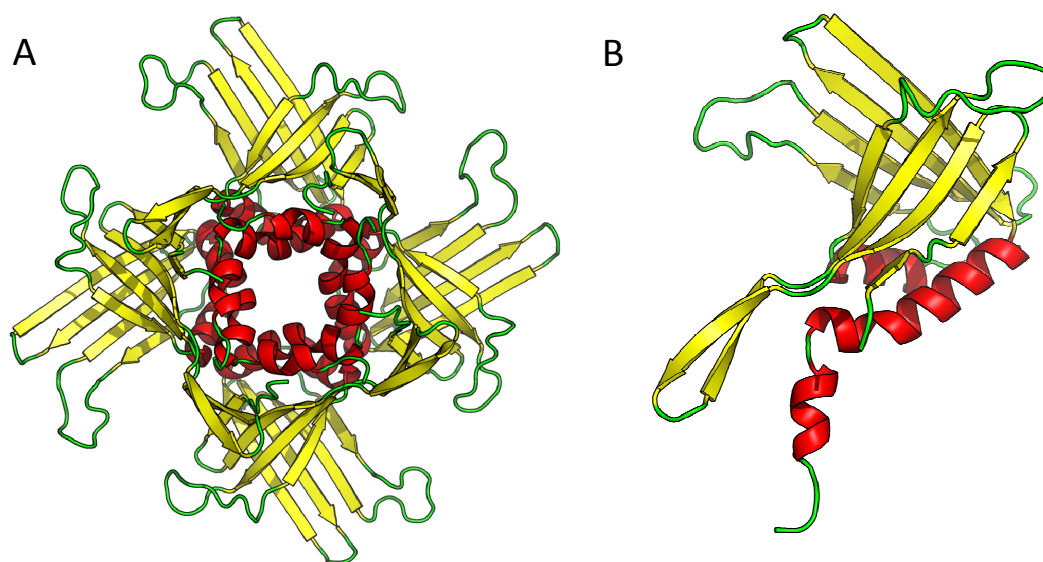


Figure 7. Structure cristallographique de StWhy1. **A.** Représentation tridimensionnelle d'un tétramère de StWhy1 semblable à un vire-vent. **B.** Représentation tridimensionnelle d'un protomère de StWhy1. Les brins β sont représentés en jaune, les hélices α en rouge et les boucles en vert (Figure adaptée de (255)).

1.2.4 Modifications post-traductionnelles des Whirlies

Plusieurs évidences suggèrent que l'activité des Whirlies est modulée par la phosphorylation de certains résidus de ces protéines (248, 256). Une première indication de la régulation des Whirlies par phosphorylation a été obtenue en démontrant que l'induction de la liaison à l'ADN de PBF-2 est perdue lorsque des extraits nucléaires de tubercules élicités ou blessés sont prétraités avec la phosphatase alcaline (248). L'activité de liaison à l'ADN de PBF-2 est également réduite après traitement par la staurosporine, un inhibiteur compétitif des protéines kinases et augmentée après traitement avec l'acide okadaïque, un inhibiteur de phosphatases. L'accumulation de la protéine PR-10a en réponse à des stress est régulée de la même façon par des inhibiteurs de kinases/phosphatases. Cette corrélation suggère un modèle dans lequel suite à un stress, l'activité de liaison à l'ADN de PBF-2 est modulée par des événements de phosphorylation. Ceci amène une augmentation de la quantité de Whirly liée au promoteur de *PR-10a* et mène éventuellement à la production de la protéine correspondante.

Dans une étude subséquente, une kinase ayant des propriétés similaires à celles de la famille des protéines kinases C (PKC) a été impliquée dans la régulation post-traductionnelle de PBF-2 (256). En effet, l'accumulation de la protéine PR-10a suite à l'élicitation des tubercules est bloquée par la sphingosine, le bisindolylmaleimide et la calphostine C, des inhibiteurs spécifiques des kinases de la famille PKC. De plus, le traitement de tubercules blessés avec du 12-O-tetradecanoylphorbol 13-acetate (TPA), un activateur des PKCs, augmente l'accumulation de PR-10a. Une corrélation a pu être établie entre l'accumulation de PR-10a suite à ces traitements et l'activité de liaison à l'ADN de PBF-2. En accord avec l'implication de kinases semblables aux PKCs dans la réponse de défense, le prétraitement des tubercules blessés avec le TPA imite l'effet de l'acide arachidonique et protège contre une infection subséquente par *P. infestans*. Les premières évidences pour une phosphorylation directe des Whirlies ont été obtenues en purifiant la protéine AtWhy3 d'*Arabidopsis* surexprimée dans la plante puis en l'analysant par spectrométrie de masse (LC-MS-MS). Cela a permis d'identifier au moins un résidu sérine

phosphorylé situé dans la partie N-terminale de la protéine (Maréchal, A., résultats non-publiés). L'identification précise des kinases responsables de la phosphorylation des Whirlies n'a toutefois pas encore été réalisée.

1.2.5 Les rôles nucléaires des Whirlies

À ce jour, le gène de pomme de terre relié à la pathogénèse *PR-10a* est la cible transcriptionnelle la mieux caractérisée pour un Whirly. Bien que la surexpression de *PR-10a* chez *S. tuberosum* ne confère pas de protection particulière contre les infections par *P. infestans* ou le Virus X de la pomme de terre, d'autres évidences plaident en faveur d'un rôle des Whirlies dans la régulation de la réponse de défense des plantes contre certains pathogènes (251, 257). En effet, des plantes mutantes arborant des mutations dans le gène *AtWhy1* se sont avérées plus sensibles à des infections médiées par des souches virulentes (la plante ne résiste normalement pas à ces souches) et avirulentes (la plante résiste normalement à ces souches) du pathogène oomycète *Hyaloperonospora parasitica* (251). Ces mêmes plantes ne se sont cependant pas avérées plus sensibles à des souches virulentes ou avirulentes de la bactérie *Pseudomonas syringae* (Maréchal A. et Parent JS., résultats non-publiés), indiquant que certaines réponses de défenses sont encore intactes malgré la mutation du gène *AtWhy1*. De façon intéressante, l'activité de liaison à l'ADN de ces mutants est diminuée et l'intensité de cette diminution corrèle positivement avec l'augmentation des symptômes suivant une infection, ce qui suggère une corrélation directe entre la liaison à l'ADN de *AtWhy1* et la résistance d'*Arabidopsis* à *H. parasitica*. *AtWhy1* semble également être requis pour l'induction de la réponse systémique acquise qui se développe dans les tissus distants des tissus infectés par une voie dépendante de l'acide salicylique, une hormone cruciale pour la réponse de défense des plantes (revue dans (258)). De nouvelles évidences suggèrent qu'une autre cible transcriptionnelle nucléaire d'*AtWhy1* et *AtWhy3* pourrait être le gène *AtKP1* qui encode pour une kinésine. *AtWhy1* et *AtWhy3* ont été purifiés par affinité pour un élément de réponse A/T-riche dans le promoteur d'*AtKP1*, le KP_{RE} (Kinesin Protein Regulatory Element). La surexpression de versions tronquées de *AtWhy1/3* auxquelles ont été retirés le PT ont pu être

immunoprécipitées spécifiquement avec le promoteur d'AtKP1. Dans les plantes surexprimant ces versions tronquées des Whirlies, l'expression du gène AtKP1 est diminuée. AtWhy1 et AtWhy3 pourraient donc avoir un rôle de répresseurs transcriptionnels de ce gène (259). La confirmation de ces rôles devrait cependant être obtenue en utilisant des plantes KO pour AtWhy1 et/ou AtWhy3. Les expériences d'immunoprécipitations devraient aussi être refaites en utilisant des anticorps contre les Whirlies endogènes.

Une autre fonction nucléaire a récemment été proposée pour AtWhy1 dans le maintien de l'homéostasie des télomères, des séquences d'ADN répétées qui protègent les extrémités des chromosomes nucléaires ((260) et revue dans (115)). Il avait précédemment été déterminé que AtWhy1 et plusieurs autres protéines de plantes possèdent une affinité pour la séquence répétée télomérique (TTTAGGG)₆ sous forme simple-brin (261). Un article plus récent visait à étudier l'effet de l'absence de AtWhy1 sur la longueur des télomères et l'activité de la télomérase (260).

Des essais *in vitro* d'activité de la télomérase ont démontré que l'ajout de la protéine AtWhy1 recombinante augmentait l'activité de la télomérase. Cependant des essais d'activité *ex vivo* utilisant des extraits de plantes KO pour *AtWhy1* indiquent plutôt que l'activité de la télomérase est augmentée en absence d'AtWhy1 ce qui corrèle avec une augmentation de la longueur des télomères dans ces plantes. Cette apparente contradiction est attribuée à la liaison d'AtWhy1 *in vitro* à certaines structures des répétitions télomériques, ce qui maintiendrait l'ADN dans une conformation appropriée pour l'action de la télomérase. *In vivo*, AtWhy1 favoriserait la formation d'une structure télomérique réfractaire à son élongation. En l'absence d'AtWhy1, les télomères semblaient aussi plus susceptibles à la dégradation par l'exonucléase Bal31 *ex vivo*. La surexpression d'*AtWhy1* rend les télomères plus courts et diminue l'activité de la télomérase dans les essais *ex vivo*. Ces évidences suggèrent qu'AtWhy1 serait impliqué dans un mécanisme de maintien des télomères dépendant de la télomérase. Les auteurs proposent qu'en se liant aux extrémités des télomères, AtWhy1 pourrait limiter l'accès de la télomérase à l'ADNsb requis pour son

activité. Il est également possible que cette liaison pourrait empêcher la machinerie de recombinaison d'avoir accès à l'ADNsb pour initier l'élongation alternative par recombinaison, un mécanisme semblable au BIR (revue dans (115)). Des croisements entre des mutants d'*AtWhy1* et différents mutants de gènes impliqués dans le maintien des télomères devraient permettre d'en apprendre plus sur l'implication des Whirlies dans l'homéostasie télomérique.

1.2.6 Les rôles des Whirlies dans les organelles

Les Whirlies plastidiques *AtWhy1* et *AtWhy3* ont été purifiés en association avec la fraction du chromosome transcriptionnellement actif du chloroplaste ce qui suggère qu'ils auraient un rôle à jouer dans le métabolisme des acides nucléiques de cette organelle (194). Récemment il a été montré qu'un Whirly du maïs, *ZmWhy1*, était également localisé dans les chloroplastes, plus précisément dans le stroma et dans la fraction des membranes thylakoides auxquelles il est relié via son interaction avec l'ADNpt (262). Les premières évidences pour un rôle des Whirlies plastidiques dans la biogénèse des plastides sont venues de l'examen de mutants du gène *ZmWhy1*. Les auteurs de cette étude ont repêché la protéine *ZmWhy1* par coimmunoprécipitation avec *CRS1*, une protéine de maturation des ARNs plastidiques requise pour l'épissage de l'intron de groupe II présent dans l'ARN du gène *atpF* (263). Des plantes mutantes hypomorphiques (ayant une perte de fonction partielle du gène muté) de *ZmWhy1* se sont révélées être pâles ou même complètement blanches selon la sévérité de la mutation du gène. Une analyse moléculaire des plantes mutantes a montré que leur contenu en ARN ribosomal plastidique était fortement diminué, les ribosomes plastidiques étant même carrément absents dans le mutant le plus sévère. L'épissage de l'intron d'*atpF* semblait quelque peu affecté dans les mutants de *ZmWhy1* mais de façon beaucoup moins marquée que dans le mutant de *CRS1*, indiquant que le rôle des Whirlies dans la maturation de cet ARN est relativement accessoire. Les auteurs ont par la suite renforcé leur étude en montrant que *ZmWhy1* était lié à quelques ARNs spécifiques dont celui d'*atpF*, mais également à des séquences d'ADN couvrant toute la longueur du génome plastidique du maïs. *In vitro*, l'affinité de *ZmWhy1* semble

cependant être plus forte pour l'ADNsb que pour l'ARN. L'analyse des autres ARNs auxquels ZmWhy1 est lié indique que cette protéine n'est pas requise pour leur maturation correcte puisque celle-ci n'est pas perturbée en absence de ZmWhy1. ZmWhy1 n'était pas requis non plus pour l'édition, la transcription et la réplication dans le chloroplaste puisqu'aucun de ces processus ne semble affecté de façon importante dans les mutants de *ZmWhy1*.

À la suite de ces travaux, l'implication exacte des Whirlies dans la perte des ribosomes chloroplastiques demeure matière à débat. ZmWhy1 semble plus important pour la biogénèse de la grande sous-unité du ribosome puisque les niveaux des ARN 23S et 4.5S sont beaucoup plus affectés par des mutations dans *ZmWhy1* que l'ARN 16S. Les auteurs proposent que ZmWhy1 pourrait agir avec d'autres protéines impliquées dans la maturation du polycistron des ARN ribosomiaux. Bien que l'interaction entre ZmWhy1 et l'ADNpt semble plus forte que celle observée pour l'ARN, le rôle des Whirlies dans le métabolisme de l'ADN reste également inconnu.

1.2.7 TIF1, un homologue des Whirlies

Chez le cilié *Tetrahymena thermophyla*, chaque cellule contient 2 noyaux : le micronoyau germinale qui transmet l'information génétique à la génération suivante et le macronoyau somatique qui sert à la transcription des gènes et est produit à la suite d'une série de réarrangements programmés du génome micronucléaire. Au cours du développement du macronoyau transcriptionnellement actif, le minichromosome de l'ADN ribosomal (ADNr) est formé. Ce minichromosome est composé par une amplification d'une région palindromique contenant 2 copies de la région de l'ADNr (environ 9000 copies par macronoyau). Cette amplification programmée constitue un modèle de choix pour l'étude de la régulation de la réplication chez les eucaryotes (264).

Les séquences particulières d'ADN appelées les éléments de type I sont situées dans la région non-transcrite de l'ADNr et régulent l'initiation de la réplication, la progression de la fourche de réplication ainsi que la transcription du minichromosome d'ADN

ribosomial de *T. thermophila*. Différentes protéines capables de lier les éléments de type I ont été identifiées et TIF1 (Type I element binding Factor 1) est la mieux caractérisée. TIF1 est une protéine de *T. thermophyla* qui s'assemble en homotétramères de 24 kDa et qui possède des régions de similarité avec les Whirlies de plantes à fleurs (**Figure 8**). Il a été démontré que TIF1 pouvait se lier aux deux brins d'ADN des éléments de type I *in vitro* et moduler l'occupation de ces sites *in vivo*. Ces résultats suggèrent un rôle pour TIF1 dans le maintien de l'origine de réplication en conformation ouverte lors des étapes d'initiation et/ou d'élongation de la réplication (265, 266). Le génome nucléaire de *T. thermophyla* est très riche en A+T (> 70 %) et les éléments de type I le sont encore plus (84 %), ce qui pourrait favoriser l'ouverture de la double-hélice de l'ADN dans les régions sans nucléosomes telles que les éléments de type I. Il a été proposé que des protéines avec une forte affinité pour l'ADN simple brin comme TIF1 pourraient alors achever de déplacer l'équilibre vers une conformation ouverte et ainsi favoriser la réplication.

Des études de génétique inverse ont permis de démontrer que *TIF1* réprime l'initiation de la réplication de l'ADNr mais qu'il est requis pour la progression normale de la phase S et la division des cellules de *T. thermophyla* (267). En effet, une diminution de l'expression de *TIF1* entraîne une activation précoce de la réplication de l'ADNr durant le cycle cellulaire couplée à un ralentissement de la progression de la phase S. Dans des souches où la production de TIF1 est diminuée, la cytokinèse est également retardée et la division du macronoyau se déroule fréquemment de façon aberrante. De plus, la passation de l'information génétique est compromise lorsque TIF1 est absent ou en quantité diminuée dans la cellule ce qui se traduit par une diminution de la quantité d'ADN dans les micronoyaux. Ces observations suggèrent un rôle pour TIF1 dans la propagation et le maintien de la stabilité des génomes micro- et macronucléaires de *T. thermophyla*.

De façon intéressante, les mutants de *TIF1* ont aussi une sensibilité exacerbée aux agents causant des dommages à l'ADN tels que l'hydroxyurée (inhibiteur réplcatif) et le méthyl méthane sulfonate (agent alkylant).

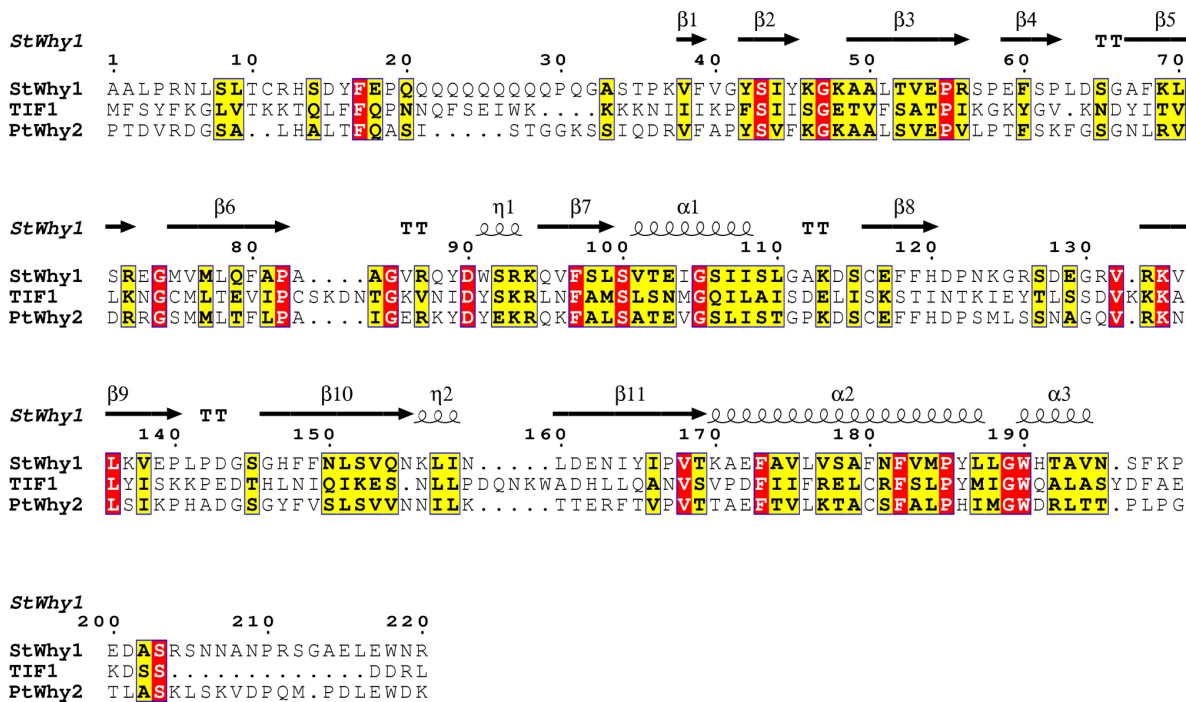


Figure 8. Conservation de séquence primaire et de structure secondaire entre les Whirlies de plantes à fleurs et la protéine TIF1 de *T. thermophyla*. Alignement de séquence de protéines Whirly matures (sans PT) et de la protéine TIF1. Les séquences du Whirly plastidique StWhy1 de la pomme de terre (*Solanum tuberosum*; AAF91282) et du Whirly mitochondrial PtWhy2 de peuplier (*Populus trichocarpa*; ABK94132) et de la protéine TIF1 du cilié *Tetrahymena thermophyla* (AAL09694) ont été utilisées. La structure secondaire de StWhy1 est représentée schématiquement au dessus de l'alignement. Une similarité de séquence de plus de 70 % est représentée en jaune et le rouge correspond à une identité. La conservation de structures secondaires similaires à celles des Whirlies pour TIF1 a été vérifiée à l'aide du serveur-web SAM-T08 (268). L'alignement a été réalisé en utilisant T-Coffee et la figure a été produite avec ESPript (269, 270).

Aussi, les voies de réparation de l'ADN dans ces cellules sont déjà activées en l'absence de stress génotoxique ce qui se traduit par une augmentation de l'expression de Rad51 (271). Par ailleurs, une absence de TIF1 provoque une perte de marqueurs génétiques sur tous les chromosomes indiquant que la stabilité du génome micronucléaire est fortement compromise. Aussi, le point de contrôle de la phase S dans les mutants de TIF1 n'est pas activé en présence d'hydroxyurée. Les auteurs suggèrent que TIF1 jouerait un rôle dans l'activation du point de contrôle de la phase S dans les deux noyaux en conjonction avec le senseur de dommage/kinase ATR, et proposent un modèle dans lequel TIF1 coopérerait avec le complexe RPA de liaison de l'ADNsb aux sites de dommages pour recruter ATR et ainsi activer la réponse aux dommages à l'ADN.

1.2.8 Les MRPs et les SALPs, des homologues structuraux des Whirlies

Les kinétoplastidés forment un ordre de protistes qui comprend plusieurs membres responsables de maladies humaines comme la maladie du sommeil causée par plusieurs espèces de trypanosomes (revue dans (272)). Une caractéristique remarquable et conservée chez les kinétoplastidés est le processus d'édition très complexe impliqué dans la maturation des ARNs mitochondriaux de ces organismes. Ce processus permet l'insertion et la délétion de nombreux nucléotides uridines dans au moins douze des dix-huit pré-ARNs mitochondriaux du kinétoplastidé *Trypanosoma brucei* permettant ainsi à ces transcrits d'être traduits adéquatement. Un cas très extrême d'édition est celui du transcrit de la cytochrome C oxydase III de *T. brucei* dans lequel 547 Us sont ajoutés et 41 sont délétés pour finalement produire l'ARN mature (273). L'édition chez les trypanosomes est dirigée par de petits ARNs de 50-70 nucléotides qu'on appelle les ARNs-guides (gARNs) et qui spécifient les endroits où la machinerie d'édition doit insérer ou déléter des Us dans les pré-ARNs. Les gARNs contiennent à leur extrémité 5' une tige-boucle ancre (I) qui leur permet de s'apparier par complémentarité au pré-ARN en 3' du site à éditer. Dans les gARNs, on retrouve également une autre tige-boucle (II) qui contient l'information nécessaire au processus d'édition et qui est complémentaire à la séquence d'ADN édité. Les gARNs se terminent par une queue poly-U qui pourrait aider à stabiliser le complexe en

interagissant avec les régions riches en purines souvent retrouvées en amont du site d'édition sur les pré-mARNs (revue dans (274)).

En plus du complexe éditosome qui contient 15-20 protéines et qui est en charge des processus enzymatiques requis pour l'édition, plusieurs protéines accessoires sont nécessaires pour le bon déroulement de la maturation des ARNs mitochondriaux. Le complexe MRP fait partie de celles-ci. MRP1 est une protéine basique qui a d'abord été identifiée par pontage aux rayons UV avec des gARNs (275). Elle est capable de faciliter l'hybridation des gARNs avec leur pré-ARN complémentaire (276). Cette étude a proposé la présence de deux domaines différents de liaison de l'ARN sur MRP1, un premier capable d'accommoder la tige-boucle II et un second capable d'exposer la tige-boucle I ancre pour favoriser son hybridation au pré-ARN. Il a par la suite été montré que MRP1 s'associe *in vivo* avec MRP2 et que chacune de ces protéines est nécessaire pour la stabilité de l'autre (277). Le complexe MRP1/2 ne semble pas absolument requis pour l'édition mais son abolition entraîne des défauts dans l'édition de certains ARNs et une diminution des niveaux d'autres ARNs qui ne sont jamais édités, suggérant des rôles supplémentaires pour ce complexe *in vivo*.

Bien que MRP1 et MRP2 n'aient aucune similarité de séquence avec les Whirlies, des études cristallographiques ont montré que ces protéines pouvaient s'assembler en hétérotétramères et que chaque protomère adopte un repliement très semblable à celui des Whirlies (β - β - β - β - α - β - β - β - β - α) (278) (**Figure 9**). De plus, bien que MRP1 et MRP2 possèdent seulement 18 % d'identité entre elles, elles adoptent tout de même un repliement quasi-identique. Par ailleurs, l'arrangement tétramérique du complexe MRP1/2 a été confirmé par observation en microscopie électronique de complexes purifiés directement de *T. brucei* (279). Le complexe MRP1/2 lie l'ARN principalement par des interactions électrostatiques et avec une faible spécificité de séquence. En accord avec les prédictions issues d'études *in vitro* sur la protéine MRP1, la résolution de la structure de MRP1/2 en complexe avec un gARN montre que la tige-boucle II est associée à la surface basique des feuillets β de MRP2 tandis que la tige-boucle I contenant la région d'ancrage du gARN est

liée à la surface basique des feuilletts β de MRP1 et est déstabilisée, exposant ses bases au solvant (276, 278). Ceci suggère un mécanisme pour expliquer le rôle de MRP1/2 dans la promotion de l'hybridation du gARN avec son pré-ARN dans lequel le complexe représente une plateforme pour exposer les régions complémentaires des gARNs. Cependant, comme l'affinité de MRP1/2 pour le gARN est très semblable à celle pour le pré-ARN, le modèle actuel propose que la liaison de ces deux types d'ARN à la région basique du complexe favoriserait leur exposition à l'ARN complémentaire (279). Une comparaison des résidus se liant à l'ARN dans le complexe MRP1/2 aux résidus situés à des positions similaires dans StWhy1 suggère que le mécanisme de liaison à l'ADNsb des Whirlies de plantes est probablement différent. En effet, chez les Whirlies de plantes ces résidus sont plutôt aromatiques et hydrophobes tandis qu'ils sont basiques chez MRP1/2. La conservation de la structure Whirly suggère cependant que ce type de repliement est particulièrement approprié pour la liaison d'acides nucléiques de manière séquence-indépendante.

Le repliement caractéristique des Whirlies a également été retrouvé très récemment chez les protéines SSGA-like (SALPs), impliquées dans la sporulation chez les actinomycètes du genre *Streptomyces*. Ces protéines ne possèdent pas de similarité de séquence avec les Whirlies de plantes mais sont assemblées sous forme homotrimériques et chaque protomère adopte une conformation semblable aux Whirlies (β - β - β - β - α - β - β - β - β - α) (**Figure 9**). Cette similarité structurale est cependant plus faible que celle observée entre les Whirlies et le complexe MRP1/2. Dans ce cas, comme SsgB ne semble pas interagir avec l'ADN génomique et que le caractère électronégatif des feuilletts β est quelque peu incompatible avec un rôle dans la liaison des acides nucléiques, cette surface a plutôt été proposée comme étant une plateforme pour des interactions protéines-protéines encore inconnues (280). Cette nouvelle découverte vient souligner la flexibilité fonctionnelle du repliement Whirly.

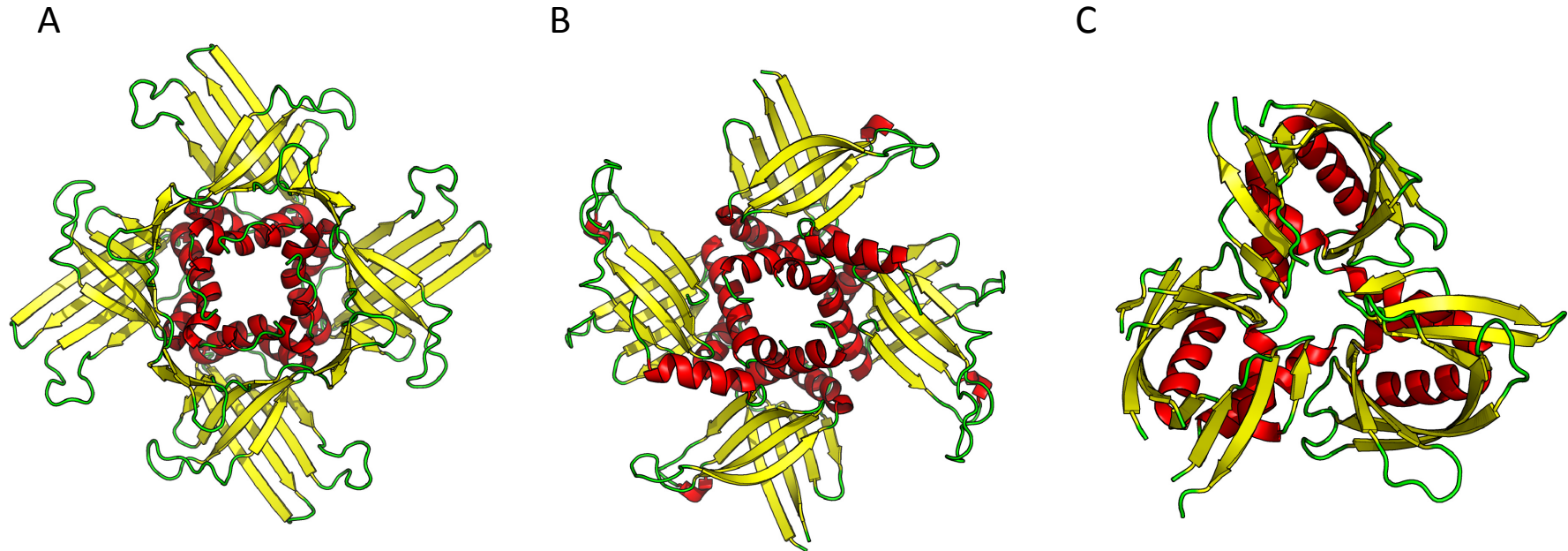


Figure 9. Homologues structuraux des Whirlies. A. Représentation tridimensionnelle d'un homotétramère de StWhy1 de *Solanum tuberosum*. B. Représentation tridimensionnelle d'un hétérotétramère du complexe MRP1/2 de *Trypanosoma brucei*. C. Représentation tridimensionnelle d'un homotrimère de SsgB de *Thermobifida fusca*. Les brins β sont représentés en jaune, les hélices α en rouge et les boucles en vert (Figure adaptée de (255, 278, 280)).

1.3 Hypothèses et objectifs de travail

Lorsque j'ai débuté mes études doctorales, les rôles des Whirlies dans les organelles étaient encore complètement inconnus et beaucoup de questions demeuraient ouvertes quant à leur implication dans la réponse de défense des plantes contre les pathogènes. L'objectif principal de mon travail était donc de préciser les fonctions des Whirlies en utilisant des approches de génétique inverse dans la plante modèle *A. thaliana*. Jean-Sébastien Parent, un autre étudiant au laboratoire, avait préalablement isolé des mutants KO des trois Whirlies d'*Arabidopsis* mais aucune de ces plantes ne semblait démontrer de phénotype particulier que ce soit au niveau des organelles ou encore de la susceptibilité aux infections par la bactérie *Pseudomonas syringae*. Finalement, l'apparition d'un phénotype de variégation indiquant que certains chloroplastes deviennent non-fonctionnels dans des plantes KO1/3 n'exprimant aucun Whirly plastidique a confirmé que ces protéines ont un rôle important dans la biogénèse des organelles. Une autre approche pour identifier la fonction d'une protéine consiste à la surexprimer de façon constitutive. Cette approche nous a permis de constater rapidement que le métabolisme des acides nucléiques dans les plastides et ou les mitochondries de plantes surexprimant des protéines AtWhy1, 2 ou 3 fusionnées à des épitopes-myc était fortement perturbé. Ces constatations ont recentré mon projet de recherche plus spécifiquement sur le rôle des Whirlies dans le métabolisme des acides nucléiques des organelles.

Chapitre 2. Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function

Article publié dans BMC Plant Biology en avril 2008

**Overexpression of mtDNA-associated AtWhy2
compromises mitochondrial function**

Alexandre Maréchal, Jean-Sébastien Parent, Mohammed Sabar, Félix
Véronneau-Lafortune, Charbel Abou-Rached and Normand Brisson

Department of Biochemistry, Université de Montréal, 2900 Édouard-Montpetit, Montréal,
Québec, Canada, H3C 3J7

2.1 Mise en contexte

Cet article avait pour objectif de caractériser la fonction des Whirlies mitochondriaux en examinant les phénotypes produits par des altérations du niveau d'expression du gène *AtWhy2* d'*Arabidopsis*. Il a permis d'établir que la surexpression d'*AtWhy2* entraînait un mauvais fonctionnement des mitochondries en diminuant fortement le niveau d'expression des gènes de cette organelle ainsi que la quantité d'ADNmt de la plante. Ces expériences n'ont cependant pas permis d'identifier précisément un rôle pour les Whirlies mitochondriaux. Pour cet article, j'ai produit et caractérisé les lignées de plantes transgéniques surexprimant *AtWhy2* et *AtWhy2-myc*. J'ai aussi caractérisé l'effet de modifications du niveau d'*AtWhy2* sur la régulation de l'expression des gènes de la mitochondrie (Figures 10C, 11A-D, 12B, 13A, 14 A et B). J'ai également mis au point la technique d'immunoprécipitation de l'ADNmt (inspiré par un protocole d'Alexandre Joyeux) et supervisé Félix Véronneau-Lafortune pour l'obtention de la Figure 15). Pour la microscopie électronique, j'ai fixé les tissus et la préparation a été réalisée par Richard Janvier à l'Université Laval. Jean-Sébastien Parent a obtenu les plantes KO*Why2* et effectué les analyses d'expression de gènes par RT-PCR (Figures 10A et B, 11E, 13B). Mohammed Sabar a effectué les analyses d'activité des complexes protéiques de la chaîne respiratoire par BN-PAGE (Figure 12A). Jean-Sébastien et moi avons produit les anticorps polyclonaux anti-*AtWhy2*. Finalement, j'ai écrit cet article au complet et Normand Brisson l'a révisé.

2.2 Abstract

2.2.1 Background

StWhy1, a member of the plant-specific Whirly single-stranded DNA-binding protein family, was first characterized as a transcription factor implicated in the activation of the nuclear *PR-10a* gene following defense-related stress in potato. In *Arabidopsis thaliana*, Whirlies have recently been shown to be primarily localized in organelles. Two representatives of the family, AtWhy1 and AtWhy3 are imported into plastids while AtWhy2 localizes to mitochondria. Their function in organelles is currently unknown.

2.2.2 Results

To understand the role of mitochondrial Whirlies in higher plants, we produced *A. thaliana* lines with altered expression of the *atwhy2* gene. Organellar DNA immunoprecipitation experiments demonstrated that AtWhy2 binds to the mitochondrial genome. Overexpression of *atwhy2* in plants perturbs mitochondrial function by causing a diminution in transcript levels and mtDNA content which translates into a low activity level of respiratory chain complexes containing mtDNA-encoded subunits. This lowered activity of mitochondria yielded plants that were reduced in size and had distorted leaves that exhibited accelerated senescence. Inactivation of the *atwhy2* gene did not affect plant development and had no detectable effect on mitochondrial morphology, activity of respiratory chain complexes, transcription or the amount of mtDNA present. This lack of phenotype upon abrogation of *atwhy2* expression suggests the presence of functional homologues of the Whirlies or the activation of compensating mechanisms in mitochondria.

2.2.3 Conclusions

AtWhy2 is associated with mtDNA and its overexpression results in the production of dysfunctional mitochondria. This report constitutes the first evidence of a function for the

Whirlies in organelles. We propose that they could play a role in the regulation of the gene expression machinery of organelles.

2.3 Background

Plant cells comprise three organelles that possess and maintain genetic information. Coordination of the gene expression in the nucleus, plastids and mitochondria is critical for plant development and survival [1] [2]. Since the endosymbiosis events that have resulted in the integration of plastids and mitochondria into eukaryotic cells, most of the genetic information found in the cyanobacterial and α -proteobacterial ancestors has been transferred to the nucleus. Nevertheless, remnants of the original genomes are still found in organelles. In *Arabidopsis*, the mitochondrial genome contains coding sequences for approximately 87 genes encoding mainly components of the translational apparatus and of the electron transport chain [3]. Since no protein involved in general DNA metabolism is present in the mitochondrial genome of *Arabidopsis*, gene expression in this organelle is under nuclear control. A consequence of this is that extensive anterograde (nucleus to organelle) and retrograde (organelle to nucleus) signalling is required for co-regulation of nuclear and organellar genes that encode proteins working cooperatively in organelles as well as for the general homeostasis of mitochondria.

Whirlies form a small family of single-stranded DNA (ssDNA) binding proteins found mainly in the plant kingdom. StWhy1, the prototypical Whirly from *Solanum tuberosum*, has been characterized as a transcriptional activator of the pathogenesis-related gene *PR-10a* following elicitation or wounding of potato tubers [4], [5], [6]. Following stress, it was shown to bind with high affinity to a single-stranded form of an inverted-repeat-containing region located in the promoter of *PR-10a* called the elicitor response element (ERE) both *in vitro* and *in vivo* [6]. Analysis of the crystal structure of StWhy1 revealed that *in vivo* Whirlies adopt a tetrameric form. Each protomer consists of 2 antiparallel β sheets packed perpendicularly against each other forming blade-like extensions which protrude out of an α -helical core that allows formation of a stable

tetramer. The surface formed by these “blades” was proposed to form the Whirly ssDNA-binding domain [7]. In accordance with the role of StWhy1 in *S. tuberosum*, the *Arabidopsis* homolog AtWhy1 was shown to be required for both full basal and specific disease resistance responses to the obligate biotroph *Peronospora parasitica* [6].

Based on analysis of the primary sequence of Whirly proteins from a variety of flowering plants, we predicted that they could localize to organelles [5]. Recently, this was confirmed for the *Arabidopsis thaliana* Whirly representatives. Two of those Whirlies, AtWhy1 [TAIR:At1g14410] and AtWhy3 [TAIR:At2g02740] are imported in plastids whereas AtWhy2 [TAIR:At1g71260] is targeted to the mitochondria ([8] and our unpublished data). Remarkably, all flowering plants, when sufficient sequence information is available, contain at least two Whirly representatives, one predicted to be plastid-localized while the other is expected to be in mitochondria. In a recent turn of events, another nuclear function has been proposed for the Whirlies as AtWhy1 was shown to be implicated in telomere length homeostasis [9]. Although dual-localization of Whirlies to nucleus and organelles in the same cell remains to be shown, it is possible that certain circumstances, such as specific stresses or developmental cues could allow Whirlies to shuttle between cell compartments, thus constituting good candidates as mediators of antero/retrograde signalling. In a first step towards a better understanding of the relationship between the nuclear functions of the Whirlies and their primary localization to organelles, we decided to elucidate the functions of the Whirlies in mitochondria.

2.4 Results

2.4.1 Overexpression of AtWhy2 perturbs mitochondrial function

To investigate the role of mitochondrial Whirlies, we produced plants with altered expression of the *atwhy2* (At1g71260) gene. Homozygous plants carrying a T-DNA insertion in the 3' untranslated region and completely devoid of *Atwhy2* expression were obtained (Figure 10A and B). In addition, plants constitutively overexpressing a myc-

tagged version of AtWhy2 under the control of the CaMV35S promoter were produced (Figure 10C). While the knock-out plants (KO) showed no visible phenotype, plants overexpressing *atwhy2* (OEX) were smaller and produced shorter siliques which contained about half the amount of seeds found in wild-type plants (Figure 11B and C). Interestingly, OEX plants also developed dark green distorted leaves and their mature leaves exhibited signs of early senescence when compared to wild-type (Col-0) or KO plants (Figure 11A and 11D). To document this accelerated cell death, we monitored the expression levels of a number of previously described senescence-associated genes (SAGs) in the third and fourth leaves of 5 week old plants from each genotype using RT-PCR (Figure 2E) [10] [11]. These genes have all been described as molecular markers of leaf senescence because their abundance is significantly increased during this genetically programmed phenomenon, *YLS6* in particular is specifically activated in conditions that result in cell-death [12]. The mRNAs of all tested SAGs were clearly more abundant in OEX compared to Col-0 and KO plants, thereby confirming that the early yellowing of leaves is an indication that a senescent state is reached more rapidly in leaves of plants overexpressing *atwhy2*. To ascertain that the observed phenotypes were not due to a non-specific effect of the overexpression of a mitochondria-targeted myc-tag, we produced transgenic plants constitutively expressing an untagged version of AtWhy2 and observed the apparition of the same phenotypes (data not shown).

Recent reports have highlighted the implication of mitochondria in regulation of senescence in numerous organisms including yeast, flowering plants and mammals [13], [14], [15], [16]. As it has been demonstrated that AtWhy2 is imported into mitochondria *in vivo*, these observations prompted us to verify the functionality of mitochondria in plants with altered *atwhy2* expression. In order to monitor the activity of the mitochondrial respiratory chain complexes, we used blue-native polyacrylamide gel electrophoresis (BN-PAGE) coupled to in-gel histochemical staining of enzymatic activities [17]. Using this procedure, we were able to evaluate the individual activities of NADH dehydrogenases, succinate dehydrogenase (Complex II) and cytochrome C oxidase (Complex IV). No differences in activity for all the observed complexes could be found between Col-0 and

KO plants (Figure 3A). In contrast, OEX plants exhibited strong deficiencies in Complexes I and IV while their alternative NADH dehydrogenase and Complex II remained as competent as those found in wild-type and KO. Interestingly, the affected complexes in OEX plants were those that contained subunits encoded by the mitochondrial genome while those that are only composed of products encoded in the nuclear genome were intact. These observations prompted us to use electron microscopy to monitor the quantity and ultrastructure of organelles present in the various lines. The number of mitochondria found in all the plants was approximately the same and they were of similar size (Figure 12B upper panel). Upon examination of individual organelles at higher magnification, mitochondria from OEX plants exhibited a simpler structure than the same organelles in Col-0 and KO plants. In general, cristae were slightly less abundant in OEX plants (Figure 12B lower panel (black arrows)). The inner membrane invaginations were counted on 10 mitochondrial sections of similar size for each of the genotypes. Cristae density averages for all observed sections (in cristae/ μm^2) were 24.2 ± 7.9 , 23.3 ± 10.3 and 17.1 ± 7.1 for Col-0, KO and OEX plants respectively. Altogether, these results indicate that mitochondrial function is compromised upon overexpression of *Atwhy2*.

2.4.2 General downregulation of mitochondrial gene expression and mtDNA levels in plants overproducing AtWhy2

The ssDNA-binding capacity of the Whirlies could be an important regulator of gene expression in organelles. Since StWhy1 has been shown to act as a transcriptional activator for the *PR-10a* nuclear gene in tubers following elicitation, it is plausible that AtWhy2 could take part in the regulation of transcription in mitochondria [4] [6]. This eventual implication was investigated by monitoring mitochondrial gene expression in plants with altered *Atwhy2* content using RNA gel blots.

Since respiratory chain complexes I and IV are clearly incompetent in OEX plants, we evaluated the expression levels of 3 subunits from each of these complexes that are encoded by the mitochondrial genome. As shown in figure 13A, steady-state RNA levels, particularly the larger transcripts of the *nad3*, *nad4*, *nad7* and for *cox1*, *cox2* and *cox3*

genes were all significantly diminished in OEX plants compared to Col-0 and KO plants. No change could be observed between Col-0 and KO plants for the steady state RNA levels detected with all of these probes. This is in agreement with the lack of difference in the activity of the respective supercomplexes (Figure 12A). Similar results were obtained for mitochondrial genes *atp8*, *atp9*, *orf240a*, *rps3*, *rpl16*. Upon closer examination of the RNA gel blots, we observed that usually the smallest RNA forms, presumably representing the mature translated forms are less affected than the larger forms, which may represent the primary transcripts. Surprisingly, for the *rpl16* gene, the smallest RNA products that hybridized to the probe were more abundant in OEX compared to KO and wild-type plants. We propose that these differences between the abundance of RNAs of different sizes could be due to post-transcriptional stabilization compensating for the diminution in production of the large primary transcripts.

Although AtWhy2 has never been detected in the nucleus and is consistently described as a mitochondrial protein, it still remained possible that at least part of the deficiency in mitochondrial complexes observed in OEX plants could be due to a defect in the expression of nuclear genome-encoded respiratory chain subunits or to a general defect in RNA metabolism ([8], [18] and our unpublished data). As a first control, nuclear 25S rRNA levels visualized by ethidium bromide staining of the gels did not appear to vary. Additionally, we measured the levels of *act3* mRNA which codes for an actin family protein and showed that they were constant between the different plants. We also evaluated the expression of 2 different nuclear genes encoding subunits of both the NADH dehydrogenase (Complex I) and of the cytochrome oxidase (Complex IV) using semi-quantitative RT-PCR. No differences in expression levels could be observed among Col-0, KO and OEX plants for the *nadb18*, *nad51*, *coxVb* and *cox6b* genes. Oligonucleotides designed to amplify the cDNA encoding a β -tubulin family protein were used as a loading control. Amplifications of *nad4* and *cox1* cDNA were used as a mitochondrial RNA level control and are clearly diminished in OEX plants as shown previously by RNA gel blot. Altogether, the data presented here underlines the specific perturbation of mitochondrial RNA accumulation in OEX plants. These observations suggest that AtWhy2 might affect

maturation of RNA in mitochondria. Alternatively, the observed general diminution in the amount of mitochondrial RNA suggests that overexpression of AtWhy2 could affect the transcriptional process or lead to a diminution in the quantity of mitochondrial DNA (mtDNA) which in turn might result in a drop of transcription rate for all the genes encoded.

To assess this latter possibility, we monitored the level of mtDNA in the plants. Probes directed against the *atp9*, *rpl16* and *orf240a* coding sequences of the mitochondrial genome were used to evaluate the amount of mtDNA present in Col-0, KO and OEX plants by Southern blot. All mitochondrial DNA regions detected by these probes were found to be highly diminished in plants overexpressing AtWhy2 (Figure 14A). No difference in mtDNA quantity could be highlighted between wild-type and KO individuals. The *rpl16* probe detects a mitochondrial band of 6.1 kb following digestion with the *BamHI* digestion of total DNA the amount of which is significantly lower in OEX plants. This probe also detects a nuclear sequence found on chromosome II in the form of a band of 1.1 kb that remains unaffected in plants overexpressing *atwhy2*, indicating that the amount of nuclear DNA in OEX plants is similar to that of KO and Col-0 plants. To confirm that mtDNA is specifically affected by the overexpression of AtWhy2, we have used probes directed against the plastidial gene *rbcL* and the nuclear *25S rRNA*. Both of these sequences are found in equal amounts in Col-0, KO and OEX plants. Thus the observed diminution in the expression level of mitochondrial genes correlates with a diminished level of mtDNA. To further substantiate the mtDNA variation observed upon overexpression of *atwhy2*, quantitative PCR (qPCR) was used to evaluate the nuclear to mitochondrial DNA ratios for all mutant plants. Primers were designed to amplify unique regions of the mitochondrial and nuclear genomes. As with the regions quantified by Southern blotting, no difference between mt/nuclear DNA ratios could be highlighted by comparing Col-0 and KO DNA samples. In agreement with the DNA gel blots results, both OEX mitochondrial regions gave significantly diminished mt/nuclear DNA ratios (7,1 % of wild-type ratio for the intron between exons d and e of *nad5* and 28,2 % of wild-type ratio for a portion of the *rRNA18S* gene).

Perturbations in higher plant mitochondrial recombination regulation system result in the accumulation of normally substoichiometric DNA molecules (sublimons) [19], [20], [21]. As Whirlies bind to ssDNA, a role in DNA recombination cannot be excluded. A particularly active site of recombination in the mitochondrial genome of *A. thaliana* is located around the *atp9* gene. This region contains numerous small repeats that are able to recombine to produce new DNA molecules. To investigate a potential role of AtWhy2 as a regulator of mtDNA recombination, we digested total DNA from the three lines with the BamHI restriction enzyme and observed the various bands detected with a probe encompassing the whole *atp9* gene and the 5' portion of *orf262*. The expected 4 bands of 1.1, 1.7, 3.2 and 4.0 kb were present in all of the genotypes though in diminished amount in OEX (see [19] for identification of the individual bands) (Figure 14A). AtOSB1, a mitochondria-localized member of the Organellar Single-stranded DNA Binding protein family, RecA3 and MSH1, a MutS family homolog (formerly CHM1) have been implicated in the regulation of mitochondria recombinational activity [19] [22] [21] [20]. When these genes are knocked-out, they produce phenotypes similar to that observed upon overexpression of AtWhy2 (distorted leaves). Absence of these genes also results in the accumulation of aberrant recombination products that are visible by Southern blotting when using probes directed against the *atp9* region. In contrast to these mutants though, no new band appeared in the AtWhy2 KO or OEX plants. Additionally, with two other probes directed against *rpl16* and *orf240a*, we could not observe an accumulation of putative aberrant recombination products suggesting that AtWhy2 is not implicated in the regulation of recombination in mitochondria. Although we could not detect a perturbation around the *atp9*, *rpl16* and *orf240a* regions, we examined only 28.2 out of the approximately 350 Kb (8.2 %) *Arabidopsis* mitochondrial genome, thus, we cannot eliminate the possibility that recombination might have occurred among other repeats present in mtDNA of OEX plants.

2.4.3 AtWhy2 is associated with mitochondrial DNA

In *Chenopodium album* mitochondrial DNA, long and abundant stretches of single-stranded DNA are present and distributed all over the genome [23]. Presence of ssDNA has also been reported in the mitochondria of cultured tobacco cells [24]. Therefore, some of these regions could be bound by mitochondria-targeted Whirlies to control a yet-to-be-defined process in plants. To determine if and which DNA regions are bound by AtWhy2 in *Arabidopsis*, we immunoprecipitated AtWhy2 and detected DNA bound to it by PCR amplification (Figure 15). Oligonucleotides were designed to amplify 4 types of DNA regions based on their biological significance : in the vicinity of transcription start sites [25], inside genes, inside isolated regions that are devoid of any coding sequence and inside the two large repeated regions of the mitochondrial genome (Figure 15A).

In these experiments, we predict that a region that is specifically pulled-down using a polyclonal anti-AtWhy2 antibody should amplify up to a certain level for Col-0 plants and should amplify to equal or higher levels in the OEX. In KO plants, PCR products should be absent or lower than in both OEX and Col-0 plants (mtDNA being especially abundant in plant cells, some non-specific immunoprecipitation may occur). Figure 15B (Input), indicates that for all amplified regions, the amount of mtDNA was significantly lower in the OEX than in the Col-0 and KO plants, confirming our previous observations about the reduced amount of mtDNA in this mutant line. After immunoprecipitation, all tested DNA regions were pulled-down in the Col-0 and OEX plants but not, or to a much lesser extent in the KO plants, indicating that the Whirlies constitutively interact with these different regions of the mitochondrial genome (Figure 15B, IP). Interestingly, even though the input amount of mtDNA in OEX plants was lower than in Col-0 or KO, we were usually able to immunoprecipitate equivalent or higher amounts of DNA from these extracts compared to wild-type lysates, indicating that a much higher proportion of the total mtDNA was bound by AtWhy2 in overexpressing plants. Therefore, the results suggest that AtWhy2 binding to mtDNA does not discriminate between coding and non-coding sequences.

2.5 Discussion

To narrow down the array of potential roles of mitochondrial Whirlies in higher plants, we produced *A. thaliana* lines with altered expression of the *Atwhy2* gene. Overexpression of *Atwhy2* in *Arabidopsis* compromised mitochondrial function by causing a diminution in transcript levels and mtDNA content. This resulted in a lowered activity of the electron transport chain complexes containing subunits encoded in the chondriome. As no significant diminution in the amount of mitochondria per leaf area was showed, the reduced amount of mtDNA in the OEX indicates that these organelles possess lower genome copy number than their counterparts in Col-0 and KO plants. Our results suggest that the modified activity of mitochondria in OEX plants may be responsible for the general smaller size and the more rapid leaf senescence phenotypes observed. It has recently been shown that in yeast, a drop in mitochondrial gene expression correlates with a diminution in chronological lifespan. Moreover, an imbalance in the various components of the electron transport chain seems to be even more detrimental to yeast lifespan (reviewed in [13]). In the case of OEX plants, both of these conditions are present, as we observed a general downregulation of mtDNA-encoded gene expression coupled to an imbalance in the oxidative phosphorylation machinery. Indeed, the activity of the succinate dehydrogenase complex, which is entirely encoded in the nucleus in *Arabidopsis*, is unaffected in the OEX plants, while activity of the mt-encoded complexes is greatly perturbed (Figure 12A). Thus, our results suggest that similar mitochondrial dysfunction in yeast and plants are able to affect the effective lifespan of cells.

The plants with altered mitochondrial Whirly expression allowed us to determine that AtWhy2 is constitutively associated with mtDNA, most probably through a direct interaction of the Whirly domain with the abundant ssDNA found in mitochondrial nucleoids. Binding sites seem to occur along the length of the mitochondrial genome since we were unable to detect a region where Whirlies are not bound to mtDNA. Interestingly, plastidial Whirlies have been isolated within *Arabidopsis* and mustard purified fractions containing transcriptionally active chromosomes suggesting that they also interact with nucleoid DNA in chloroplasts [26]. Because a higher proportion of mtDNA is bound by

AtWhy2 in OEX plants (Figure 15B), it is possible that an overabundance of mitochondrial Whirlies could perturb the regulation of gene expression in mitochondria by rendering nucleoid DNA unavailable for binding by components of the transcription and/or replication machinery, thereby diminishing the steady-state RNA levels and the amount of mtDNA produced.

Since their discovery, members of the Whirly family have been purified independently using ssDNA affinity chromatography by ours and other groups ([4], [18], [27]). While our lab used a single-stranded form of the inverted-repeat-containing elicitor response element as a bait [4], Vermel and colleagues used an affinity column exhibiting random bovine ssDNA sequences [18] and yet another group used an hexameric repeat of *Arabidopsis* telomere ends [27]. The fact that Whirlies are able to bind with high affinity to all these different DNAs indicates weak sequence specificity at least *in vitro*. Our mtDNA immunoprecipitation results suggest that this low sequence specificity is also true for mitochondrial Whirlies *in vivo*. As the size of DNA fragments following sonication was determined to be approximately between 350 and 1000 bp (data not shown), an alternative explanation for the obtained results could be that binding of AtWhy2 to mtDNA is specific to certain sequences that are common to the various immunoprecipitated regions.

2.5.1 Putative functions of mitochondrial Whirlies

Our results suggest that AtWhy2 is implicated in the regulation of mitochondrial gene expression. Whirlies are found in mitochondria and plastids from higher plants. In plastids, transcription is under the control of two types of RNA polymerases, a unique eubacterial-type plastid-encoded polymerase (PEP) and multiple phage-type nucleus-encoded polymerases (NEPs) (reviewed in [28] and [29]). These RNA-polymerases specifically regulate the transcription of different subsets of genes but can also co-regulate a portion of the plastidial genes [30]. In mitochondria of higher plants, transcription is regulated by NEPs only. These facts, coupled to the demonstrated role of StWhy1 as a transcriptional activator suggest that the Whirlies could also act as transcriptional regulators in organelles. For example, they might constitute processivity cofactors for phage-type

RNA polymerases. It can be argued that since binding of AtWhy2 is also observed in non-coding regions of the mitochondrial genome, transcription might not constitute the sole function of this Whirly. However, binding to non-coding regions and to non-consensus sites has been reported for numerous classical nuclear transcription factors during genome-wide protein-DNA interaction studies (reviewed in [31]), hence a role in transcription of mtDNA-encoded genes for AtWhy2 remains possible.

In mammals, transcription is important for the replication of mitochondrial DNA, providing the primers necessary for initiation (reviewed in [32]). In higher plants however, such an interrelationship between these processes remains to be shown. The diminution in mtDNA content observed in OEX plants hints at a role for AtWhy2 in replication/maintenance of the mitochondrial genome. A distant homolog of the Whirlies, the TIF1 protein from the protist *Tetrahymena thermophyla* has been implicated in replication of DNA as well as in the faithful intergenerational transmission of genetic information and is required for activation of an intra-S-phase checkpoint triggered by DNA-damage [33], [34], [35]. Thus, mitochondrial Whirlies could potentially play multiple roles in the replication, maintenance and transcription of the chondriome.

Since overexpression of AtWhy2 differentially affects various forms of mitochondrial RNA, it is plausible that Whirlies might be implicated in the maturation of mtRNA. For example they might stabilize specific forms of RNA or facilitate processing from polycistronic immature RNA to the final translatable RNA form. Recently, the crystallographic structure of a heterotetramer of the MRP1 and MRP2 mitochondrial RNA-binding proteins from *Trypanosoma brucei* was solved. Although these two proteins share no primary sequence similarity with the Whirlies, they adopt a Whirly-like fold. These proteins are involved in RNA editing and bind to guide RNAs (gRNAs) that specify where uracil residues are added or removed in the sequence of the mitochondrial pre-mRNA [36]. We tested whether the Whirlies could have a role in editing by cloning and sequencing *nad7* and *ccb206* cDNAs. These RNA are normally edited at 28 (*nad7*) and 39 sites (*ccb206*) in *Arabidopsis* [37]. No significant differences could be found in the edition frequency for these 2 genes between WT, KO and OEX plants (data not shown). Thus,

Whirlies do not appear to be required for the correct editing process of mtRNA in flowering plants. This is not unexpected since the mechanism of editing in flowering plants is very different than in *Trypanosoma brucei* (reviewed in [38], [39]). In addition, representatives of the Whirly family are found in the green algae *Chlamydomonas reinhardtii* and *Ostreococcus taurii* where the editing process of organellar RNA appears to be absent [40]. The absence of a role for AtWhy2 in the editing process does not preclude the implication of the Whirlies in other steps of the maturation of RNA.

Finally, although overexpression of *Atwhy2* led to impaired mitochondria, inactivation of this gene in the KO line did not affect plant development or fertility and had no detectable effect on mitochondrial morphology and the activity of its respiratory chain complexes. Additionally, no alteration in steady state mtRNA levels or the amount of mtDNA present could be observed. AtWhy2 is thus not essential for the viability of *Arabidopsis thaliana* and the correct function of mitochondria in normal controlled growth conditions. This could be explained by functional redundancy with other proteins present in the organelle. The role of mitochondrial Whirlies could also be highlighted when a particular stress is applied to the plant akin to the transcriptional activator role of StWhy1 which correlates to the induction of its ssDNA-binding activity following biotic stresses. Further experiments will be required on both plastidial and mitochondrial Whirlies before we can actually pinpoint the role(s) of these intriguing proteins in organelles.

2.6 Conclusions

We have shown that AtWhy2 is constitutively bound to mtDNA demonstrating for the first time that Whirlies can associate with DNA in organelles. The interaction between AtWhy2 and mtDNA seems to occur along the entire mitochondrial genome, suggesting that the Whirlies might regulate a variety of processes important for organellar nucleic acid metabolism. Ectopically overproducing AtWhy2 clearly affected the gene expression machinery of mitochondria, diminishing both the amount of mtDNA and the level of expression of the genes it encodes. These deficiencies led to smaller plants that exhibited

accelerated senescence in leaves. The details in the mechanism of regulation of gene expression by *AtWhy2* remain to be unravelled.

2.7 Methods

2.7.1 T-DNA insertion line characterization

The SALK Institute Genomic Analysis Laboratory provided the sequence-indexed T-DNA insertion line SALK_118907 [41]. These plants contain an insertion at nucleotide +29 relative to the STOP codon. Homozygosity of the insertion was verified by PCR using the following primers: P3, 5'-GTAGCGGCTACTTCATCTCA-3', P4, 5'-TTCGTGTGCACCAAATGCCA-3' and Lb1, 5'-GCGTGGACCGCTTGCTGCAACT-3'. RNA was isolated from homozygous mutants using TRIZOL® reagent (Invitrogen) according to manufacturer's instructions. Absence of expression of *Atwhy2* was confirmed by RT-PCR using the following primers : P1, 5'-ATGAAGCAAGCCCGCTCTTT-3' and P2, 5'-CAGCTTTTGTGACAGGAACC-3' and P3 and P4.

2.7.2 Production of plants constitutively overexpressing *AtWhy2* and *AtWhy2-myc*

Six copies of the c-myc epitope were amplified by PCR from the pCR-blunt-II-TOPO-myc vector (kind gift of Dr. Jeff L. Dangl, University of North Carolina, NC, USA) using the following primers : 5'-CCCAAGCTTGCCCTTCCGGTCGACAAAGCTATG-3' and 5'-CCGCTCGAGTCATCGATTTTCGAACCCGGGGTAC-3'. The amplicon was then digested with HindIII and XhoI and cloned into a pBS-SK(+) vector. The full length *Atwhy2* was amplified from cDNA using the following primers : 5'-ATAAGAATGCGGCCGCATGATGAAGCAAGCCCGCTCTTTG-3' and 5'-GACTAGTTTTATCCCACTCCAGCTCTAACTG-3'. The amplicon was subsequently digested with NotI and SpeI restriction enzymes and cloned in frame with the c-myc epitopes in the pBS-SK(+)-6-c-myc vector. The tagged *Atwhy2* was then reamplified using the following primers: 5'-TCTAGAGGCGCGCCATGATGAAGCAAGCCCGCTCTT

TG-3' and 5'-TCTAGAAGGCCTTCATCGATTTCTGAACCCGGGGTAC-3' cloned into the a pGREENII022935S vector [42]. This construct was cotransformed with companion vector pSOUP into a GV3101 pMP90 *Agrobacterium tumefaciens* strain. Plant transformation was carried out using the floral dip approach as described [43]. Transformed plants were selected on soil using the BASTA resistance conferred by the pGREENII0229 vector. All plants that expressed the construct exhibited the phenotype presented here. Because homozygous lines were almost completely infertile, heterozygous lines were derived from the T1 plants and used for the experiments shown here. WT plants were identified by their normal appearance and removed for the experiments involving mature plants. For experiments with seedlings, such a selection could not be performed. Plants overexpressing an untagged version of *Atwhy2* under the control of a CaMV35S promoter phenocopied the plants overproducing the *AtWhy2*-myc construct, thereby eliminating the possibility that the phenotype could be a result of some unspecific effect of the epitope on the biogenesis of mitochondria.

2.7.3 Antibody production

Recombinant *AtWhy2* was expressed and purified as described previously [6]. Rabbits were immunized and antiserum was collected. For protein gel blot analysis, the antiserum was used typically at a concentration of 1:1000 to 1:2500.

2.7.4 DNA, RNA gel blots and oligonucleotides

RNA was isolated from 4 week old plants using TRIZOL® reagent (Invitrogen) according to manufacturer's instructions. DNA was isolated from plants using a CTAB DNA extraction protocol [44]. Running of the samples and blotting of the gels was performed as described [45]. All PCR products in this study were amplified by PCR from DNA using the following primers : RRNA26SMFOR, 5'-GCGTACCTTTTGCATGATGGG-3', RRNA26SMREV, 5'-CCTAGCCCATTGAGTGCTCTA-3', NAD7FOR, 5'-TTGCGAGGTACCATTACGAGC-3', NAD7REV, 5'-CACCACTGAATCCCCAATCCT-3', ATP9FOR, 5'-ACCCGAGATGTTAGAAGGTGC-3', ATP9REV, 5'-GGCCATCATTGG

GGCAAACAA-3', COX1FOR, 5'-GTCTATCCGCCCTTAAGTGGT-3' , COX1REV, 5'-TACACCTCTGGATGACCGAAG-3', ATP8FOR, 5'-GGAGATGG AGTACTTG GGATC-3' , ATP8REV, 5'-TCCATTCCCTCGTGAGCCACTT-3', RPS3FOR, 5'-CCATGACCGATTACCCTCGATA-3', RPS3REV, 5'-CTGTAAGCTTCTTCCCTGTGC-3', RPL16FOR, 5'-CCTGCGGAAGTATCTACTCGT-3', RPL16REV, 5'-ATGTAGCGG CTTGTCGAGCAT-3', ORF240aFOR, 5'-GCGTTCTAAGATCACTGAGGG-3', ORF240 aREV, 5'-AGCTCGTTGAGATCGAGAGCA-3', ATP9ORF262FOR, 5'-GTATAATTC TCAACCCGAGATGTTAGAA-3', ATP9ORF262REV, 5'-GGCTAGATAGCACCA TTGTGTCA-3', PSBAFOR, 5'-GGAAGCTGCATCCGTTGATGA-3', PSBAREV, 5'-CCGAATACACCAGCTACACCT-3', UPCOX1FOR, 5'-GCTAGCTCATGGCAGGA AATC-3', UPCOX1REV, 5'-GTAACGTCCGTTCCGTGA TCT-3', UPRPS3FOR, 5'-GGGCCCTTTTACCAAGTCTAC-3', UPRPS3REV, 5'-CGCGCTAAGAAAGGTTG CTTC-3', UPRRNA26SFOR, 5'-GCACTCTACCAGAGCTACTAC-3', UPRRNA26S REV, 5'-GGCGCCTTTGTATGTCTCAAC-3', UPRRNA18SFOR, 5'-CCAAGCCT GGA TAAGCAACTG-3', UPRRNA18SREV, 5'-CAGTTGCTTATCCAGGCTTGG-3', 5000FOR, 5'-CGATTCATCTCGGCCTTAGAC-3' 5000REV, 5'-GTCCTGGAAAGAG CAGATGTG-3', UPATP9652FOR, 5'-GGA TCGCATGAGTACGAGAGA-3', UPATP9652REV, 5'-GCACGAGTCGATCCCT ATTCT-3', REPEATIIBFOR, 5'-CTAAACGAGACAGCAGCTACC-3', REPEATII REV, 5'-AACAGCCGGAAGGTATGAAGC-3', TRNAGLYFOR, 5'-GTTGGACA TCTGCTCTATCC-3', TRNAGLYREV, 5'-AAGACAGATGCCGCTACCTA-3', TRNAFMETFOR, 5'-GGTTGTTGGTCCAACGACTCT-3', TRNAFMETREV, 5'-AGTACACACTGTGCACCACGA-3', ER265000FOR, 5'- GTGAATGCCCAGTG CAATCCA-3', ER265000REV, 5'-TTCTCCCTGTGCTTACATGGG-3', ER84000F OR, 5'-TTTCTCGTGCTCTCCGTCCAA-3', ER84000REV, 5'-AGACAGAAGA GCTGGTAGGAC-3', REPEAT1AFOR, 5'-GCCATTCTTCTCTACCCATGC-3', REPEAT1AREV, 5'-CTCACAGAGTCATCGGTATCC-3', REPEAT1BFOR, 5'-TGGACTACCAAAGACCCAGAC-3', REPEAT1BREV, 5'- TCGGATTCTTGCCA TCACTGG-3', REPEATIIAFOR, 5'-CGGATCGACTCGACTGATATG-3', REPEA

TIIAREV, 5'-GTCCTGATCGAGCAACTAGTC-3', NAD3FOR, 5'-TTT ACT CCC GAT CCG AAG CAC-3', NAD3REV, 5'-TTT GAT CCT ACT CGG TGT TCC-3', NAD4FOR, 5'-TTC CAA ACA GGA ACC ACC GAT-3', NAD4REV, 5'-CAC CAG ATT CAT ATG GGC TAC-3', COX3FOR, 5'-CAA GTC CAT GGC CTA TTT CGG-3', COX3REV, 5'-ACC ATG CAG CTG CTT CAA AGC-3', COX2FOR, 5'-TTG TGA TGC AGC GGA ACC ATG-3', COX2REV, 5'-TGA CAC CTG AGG AAG GTA CAG-3'. All probes were subsequently cloned into a pDRIVE cloning vector (QIAGEN) and sequenced.

2.7.5 Blue native-polyacrylamide gel electrophoresis (BN-PAGE)

Around 200 mg of leaves were homogenized in 0.33M Sucrose, 50 mM Bis-tris, pH 7.0 and 750 mM aminocaproic acid. The homogenate was filtered and spun at 1000g for 10 min. The pellet was discarded and the supernatant spun for 15 min at 10000 g. The pellet enriched with crude mitochondria was resuspended in 50 mM Bis-tris, pH 7.0, 750 mM ACA and 0.5 mM EDTA and solubilised with 1% n-dodecyl-maltoside on ice for 5 min. After centrifugation at 25000 g for 5 min, the supernatant enriched with mitochondrial complexes was supplemented with a 5% (w/v) stock solution of Coomassie blue (Serva) in solubilisation medium to a final ratio of 1:4 (w:w) of Coomassie to detergent, and subjected to BN-PAGE according to [46]. The separating gel consisted of a linear gradient of 5-13% (w/v) acrylamide and a stacking gel of 4.5% (w/v) acrylamide. The cathode buffer used was 50 mM Tricine, 15 mM Bis-Tris pH 7.0 at 4 °C and 0.02% Coomassie blue (w/v), freshly prepared and supplemented with 0.02% detergent. Electrophoresis was carried out at 65 V for 1 hour at 4°C or until the proteins entered the stacking gel and then at 140 V, constant voltage, overnight. Following electrophoresis, the gels were rinsed briefly with MilliQ water and then equilibrated in a reaction buffer without reagents for 10 min. The gels were then incubated in fresh buffer plus specific reagents. All steps were carried out at room temperature. The reaction media were as described [17]. The reactions were stopped at various lengths of time by fixing the gels in 45% methanol (v/v) and 10 % acetic acid (v/v). The gels were destained overnight in the same solution to remove residual Coomassie blue.

2.7.6 Semi-quantitative RT-PCR

cDNA was produced from total RNA isolated from 4 week old plants of the corresponding genotypes using RevertAid™ First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer's instructions. PCR was performed using the following oligonucleotides: NADB18FOR, 5'-ATG GAG GTT CCA GGT TCA TCG-3', NADB18REV, 5'-AGG GAT AAG AGG AAC AGC AGC-3', NAD51FOR, 5'-CAC GTC CAC GAA GTC CAG ATT-3', NAD51REV, 5'-TTC TTG GCT TGC AAA GGG CAG-3', COXVBFOR, 5'-GTG ATG GTC ATC ACC ATG TCC-3', COXVBREV, 5'-GGA GGA GAA TCG TTT CAT CGC-3', COX6BFOR, 5'-TCA AAC TCC ATC GCT CTC CGA-3', COX6BREV, 5'-AGA GGA CCA GGG AAT GTT CCA-3', 5At2g45570, 5'-AGC TCT CGC AAG GTC GTT CCA TCT-3', 3At2g45570, 5'-CTT CAC CTT GTT CTC CCT CAA CGT-3', 5At1g47128, 5'-GCA GTT GCT CAT CAA CCC ATT AGC-3', 3At1g47128, 5'-GTC ACA TTG GGT TGG AGG CTT GAT-3', 5At2g38860, 5'-CGG CGG TTT ATG ACC TTG AGG ATG-3', 3At2g38860, 5'-AAA AAG GCA GGC AGA TCC GTG GCT-3', 5-TUBURT, 5'-GGT GGA GCC TTA CAA CGC TAC TTT-3', 3-TUBURT, 5'-TCG CCT GAA CAT CTC TTG GAT CGA-3'.

2.7.7 Real-Time Quantitative PCR

DNA extracted from equal fresh weight of each genotype was used as a template. PCR reactions for 384 well plate formats were performed using 2 µl of each DNA sample, 5 µl of the TaqMan PCR Master Mix (Applied Biosystems, CA), 2 µM of each primer and 1 µM of the Universal TaqMan probe in a total volume of 10 µl. Primers for genomic DNA were selected to amplify a unique nuclear region in gene AT5G08530 : 5'-GCGTTTGAATCTAGAGAAGGCTA-3' and 5'-CAGAACCAC ATGCATTCTTCC-3'. The Taqman probe #54 from a Universal Library was selected. Primers were selected to amplify a unique mitochondrial region found in the *18S rRNA* gene : 5'-TTTCGAAACCAATTCACCTTGAG-3' and 5'-TGTAGATTTC ACCCCTCCACA-3'. The Taqman probe #82 from a Universal Library was selected. Another unique mitochondrial region found in the intron between exons d and e of the *nad5* gene was also examined using

the following primers: 5'-CTCTCCGCAGGGGAATCT-3' and 5'- AACCCCCATGAT GTGGTAA-3'. The Taqman probe #38 from a Universal Library was selected. The ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) was used to detect amplification products and was programmed to an initial step of 10 minutes at 95°C, followed by 45 cycles of 15 seconds at 95°C and 1 minute at 60°C. All reactions were run in triplicate and the average values were used for quantification. The relative quantification of target genes was determined by using the $\Delta\Delta\text{CT}$ method. Briefly, the Ct (threshold cycle) values of the mitochondrial region were normalized to the nuclear region ($\Delta\text{CT} = \text{Ct mitochondrial} - \text{Ct nuclear}$) and compared with a calibrator (Col-0 samples): $\Delta\Delta\text{CT} = \Delta\text{Ct Sample} - \Delta\text{Ct Col-0 samples}$. Relative expression (RQ) was calculated using the Sequence Detection System 2.2.2 software (Applied Biosystems) and the formula $\text{RQ} = 2^{-\Delta\Delta\text{CT}}$.

2.7.8 Microscopy

For transmission electron microscopy, leaf mesophyll tissue from 4 week old plants was cut into 1-2 mm² pieces which were washed twice with 0.2 M sodium cacodylate buffer pH 7.4 for 10 minutes. Tissue was fixed by incubation for 6 hours in cacodylate buffer containing 1 % EM-grade glutaraldehyde. Samples were coated with epoxy resin, cut into 80 nm sections and mounted onto nickel/formvar grids. Staining of the samples was done using uranyl acetate and lead citrate. Observations were done using a JEOL (JEM1230) transmission electron microscope at 80 kV. Photographs were taken using a Gatan DualVision camera.

2.7.9 DNA immunoprecipitation in mitochondria

Leaf tissue was fixed in 1 % formaldehyde for 15 minutes under vacuum. Glycine was added to 0.125 M to titrate the remaining formaldehyde and fixed tissue was washed 3 times with distilled water. Crude mitochondria were prepared by grinding in mitochondria extraction buffer (350 mM mannitol, 30 mM MOPS pH 7.3, 0.2 % bovine serum albumin, 0.6% polyvinylpyrrolidone, 1 mM EDTA) using a mortar and pestle. The ground tissue was

filtered through 2 layers of Miracloth (Calbiochem) and spun at 1000 g for 5 minutes in a microcentrifuge to pellet most plastids. The supernatant was recuperated and spun at 11000 g in a microcentrifuge to pellet mitochondria and remaining plastids. The pellet was resuspended in IPP buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 % NP-40, 0.5 % sodium deoxycholate 1 tablet Complete protease inhibitor (Roche)), lysis was carried out on ice for 2 minutes and the lysate was spun at 11000 g for 15 minutes. Supernatant was recuperated and sonicated using a Branson microtip sonicator. Sonicated material was precleared by incubating with protein-A-agarose beads (Roche) for 1 hour at 4°C on a rotating platform. The precleared material was spun at 11000 g for 15 minutes and the supernatant was transferred to a fresh tube. Proteins were quantified and 100 µg was kept as input material while 500 µg was used for immunoprecipitation using 1/200 rabbit polyclonal anti-AtWhy2 sera in 1 mL final volume. Immunoprecipitations were carried out for 1 hour at 4°C before adding 50 µl of protein-A-agarose beads. Complexes were allowed to form overnight. Beads were pelleted at 100 g for 1 minute and were washed 2 times 5 minutes with IPP buffer, 2 times 5 minutes with low salt buffer (50 mM Tris-HCl pH 7.5, 0.1 % NP-40, 0.05% sodium deoxycholate), 2 times 5 minutes with high salt buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.1 % NP-40, 0.05% sodium deoxycholate), 2 times 5 minutes with LiCl buffer (50 mM Tris-HCl pH 7.5, 0.1 % NP-40, 0.05 % sodium deoxycholate, 250 mM LiCl) and 2 times 5 minutes with TE pH 7.5 (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Beads were pelleted at 11000 g for 20 seconds and resuspended in TE buffer containing 1 % SDS. Complexes were eluted from the beads by heating at 65 °C for 15 minutes. Beads were spun down at 11000 g for 5 minutes and the supernatant transferred to a new tube. One tenth volume of 2M NaCl was added and the eluted complexes were incubated at 65 °C overnight for decrosslinking. Glycogen was added to a final concentration of 0.5 µg/ul and twice the volume of 100 % ethanol was added. DNA was precipitated at -80 °C for 1 hour. Tubes were spun at 11000 g for 20 minutes. DNA pellets were resuspended in TE buffer. DNA was subjected to a phenol/chloroform extraction and to a chloroform extraction before being reprecipitated as above. After a final wash with 70 % ethanol, DNA was resuspended in water. This DNA was used for PCR reactions. To

further control the specificity of the experiment, we determined the size of DNA fragments following sonication and found that they were between 350 and 1000 bp in length, indicating that for any given region immunoprecipitated specifically, it can be assumed that there is at least one Whirly binding-site within a maximum of 1000 bp starting from the center of the amplified region. Additionally, we used primers to amplify a plastidial DNA region in the *psbA* gene as a negative control since AtWhy2 exclusively localizes to mitochondria and thus should not interact with the chloroplast genome. We found that we could only amplify similar background levels of plastidial genes for Col-0, KO and OEX plants confirming the lack of specific interaction between AtWhy2 and plastidial DNA *in vivo* (data not shown).

2.8 Authors' contributions

AM produced the AtWhy2 and AtWhy2-myc overexpressing lines, characterized the phenotypes of the mitochondrial Whirly mutants, designed the study and drafted the manuscript. JSP obtained the *Atwhy2* KO plants, helped producing antibodies and performed RT-PCR experiments. MS did the BN-PAGE and in-gel enzymatic assays and participated in the design of the study. FVL participated in the mtDNA immunoprecipitation experiments. CAR participated in the RNA and DNA gel blot analysis of the plants. NB participated in the design and coordination of the study and helped to draft the manuscript. All authors read and approved the final manuscript.

2.9 Acknowledgements

We thank Richard Janvier for excellent technical assistance with the electron microscopy. This research was supported by a grant from the Natural Science and Engineering Research Council of Canada (NSERC). AM was supported by scholarships from NSERC and from the Fonds Québécois de la Recherche sur la Nature et les Technologies.

2.10 References

1. Nott A, Jung HS, Koussevitzky S, Chory J: Plastid-to-nucleus retrograde signaling. *Annual review of plant biology* 2006, 57:739-759.
2. Liu Z, Butow RA: Mitochondrial retrograde signaling. *Annual review of genetics* 2006, 40:159-185.
3. Unseld M, Marienfeld JR, Brandt P, Brennicke A: The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature genetics* 1997, 15(1):57-61.
4. Desveaux D, Despres C, Joyeux A, Subramaniam R, Brisson N: PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. *The Plant cell* 2000, 12(8):1477-1489.
5. Desveaux D, Marechal A, Brisson N: Whirly transcription factors: defense gene regulation and beyond. *Trends in plant science* 2005, 10(2):95-102.
6. Desveaux D, Subramaniam R, Despres C, Mess JN, Levesque C, Fobert PR, Dangl JL, Brisson N: A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*. *Developmental cell* 2004, 6(2):229-240.
7. Desveaux D, Allard J, Brisson N, Sygusch J: A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nature structural biology* 2002, 9(7):512-517.
8. Krause K, Kilbienski I, Mulisch M, Rodiger A, Schafer A, Krupinska K: DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS letters* 2005, 579(17):3707-3712.
9. Yoo HH, Kwon C, Lee MM, Chung IK: Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in *Arabidopsis*. *Plant J* 2007, 49(3):442-451.
10. Gepstein S, Sabehi G, Carp MJ, Hajouj T, Neshher MF, Yariv I, Dor C, Bassani M: Large-scale identification of leaf senescence-associated genes. *Plant J* 2003, 36(5):629-642.
11. Yoshida S, Ito M, Nishida I, Watanabe A: Isolation and RNA gel blot analysis of genes that could serve as potential molecular markers for leaf senescence in *Arabidopsis thaliana*. *Plant Cell Physiol* 2001, 42(2):170-178.

12. Godiard L, Sauviac L, Dalbin N, Liaubet L, Callard D, Czernic P, Marco Y: CYP76C2, an *Arabidopsis thaliana* cytochrome P450 gene expressed during hypersensitive and developmental cell death. *FEBS letters* 1998, 438(3):245-249.
13. Bonawitz ND, Shadel GS: Rethinking the mitochondrial theory of aging: the role of mitochondrial gene expression in lifespan determination. *Cell cycle (Georgetown, Tex)* 2007, 6(13):1574-1578.
14. Maxwell DP, Nickels R, McIntosh L: Evidence of mitochondrial involvement in the transduction of signals required for the induction of genes associated with pathogen attack and senescence. *Plant J* 2002, 29(3):269-279.
15. Ahn CS, Lee JH, Reum Hwang A, Kim WT, Pai HS: Prohibitin is involved in mitochondrial biogenesis in plants. *Plant J* 2006, 46(4):658-667.
16. Kujoth GC, Leeuwenburgh C, Prolla TA: Mitochondrial DNA mutations and apoptosis in mammalian aging. *Cancer Res* 2006, 66(15):7386-7389.
17. Sabar M, Balk J, Leaver CJ: Histochemical staining and quantification of plant mitochondrial respiratory chain complexes using blue-native polyacrylamide gel electrophoresis. *Plant J* 2005, 44(5):893-901.
18. Vermel M, Guermann B, Delage L, Grienenberger JM, Marechal-Drouard L, Gualberto JM: A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* 2002, 99(9):5866-5871.
19. Zaegel V, Guermann B, Le Ret M, Andres C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P: The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *The Plant cell* 2006, 18(12):3548-3563.
20. Shedje V, Arrieta-Montiel M, Christensen AC, Mackenzie SA: Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *The Plant cell* 2007, 19(4):1251-1264.
21. Sakamoto W, Kondo H, Murata M, Motoyoshi F: Altered mitochondrial gene expression in a maternal distorted leaf mutant of *Arabidopsis* induced by chloroplast mutator. *The Plant cell* 1996, 8(8):1377-1390.
22. Martinez-Zapater JM, Gil P, Capel J, Somerville CR: Mutations at the *Arabidopsis* CHM locus promote rearrangements of the mitochondrial genome. *The Plant cell* 1992, 4(8):889-899.

23. Backert S, Lurz R, Oyarzabal OA, Borner T: High content, size and distribution of single-stranded DNA in the mitochondria of *Chenopodium album* (L.). *Plant molecular biology* 1997, 33(6):1037-1050.
24. Oldenburg DJ, Bendich AJ: Size and Structure of Replicating Mitochondrial DNA in Cultured Tobacco Cells. *The Plant cell* 1996, 8(3):447-461.
25. Kuhn K, Weihe A, Borner T: Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucleic acids research* 2005, 33(1):337-346.
26. Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmuller R: pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *The Plant cell* 2006, 18(1):176-197.
27. Kwon C, Chung IK: Interaction of an *Arabidopsis* RNA-binding protein with plant single-stranded telomeric DNA modulates telomerase activity. *The Journal of biological chemistry* 2004, 279(13):12812-12818.
28. Sakai A, Takano H, Kuroiwa T: Organelle nuclei in higher plants: structure, composition, function, and evolution. *International review of cytology* 2004, 238:59-118.
29. Shiina T, Tsunoyama Y, Nakahira Y, Khan MS: Plastid RNA polymerases, promoters, and transcription regulators in higher plants. *International review of cytology* 2005, 244:1-68.
30. Hajdukiewicz PT, Allison LA, Maliga P: The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *The EMBO journal* 1997, 16(13):4041-4048.
31. Sikder D, Kodadek T: Genomic studies of transcription factor-DNA interactions. *Current opinion in chemical biology* 2005, 9(1):38-45.
32. Bonawitz ND, Clayton DA, Shadel GS: Initiation and beyond: multiple functions of the human mitochondrial transcription machinery. *Molecular cell* 2006, 24(6):813-825.
33. Morrison TL, Yakisich JS, Cassidy-Hanley D, Kapler GM: TIF1 Represses rDNA replication initiation, but promotes normal S phase progression and chromosome transmission in Tetrahymena. *Molecular biology of the cell* 2005, 16(6):2624-2635.
34. Saha S, Nicholson A, Kapler GM: Cloning and biochemical analysis of the tetrahymena origin binding protein TIF1: competitive DNA binding in vitro and in

- vivo to critical rDNA replication determinants. *The Journal of biological chemistry* 2001, 276(48):45417-45426.
35. Yakisich JS, Sandoval PY, Morrison TL, Kapler GM: TIF1 activates the intra-S-phase checkpoint response in the diploid micronucleus and amitotic polyploid macronucleus of Tetrahymena. *Molecular biology of the cell* 2006, 17(12):5185-5197.
 36. Schumacher MA, Karamooz E, Zikova A, Trantirek L, Lukes J: Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell* 2006, 126(4):701-711.
 37. Giege P, Brennicke A: RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96(26):15324-15329.
 38. Shikanai T: RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* 2006, 63(6):698-708.
 39. Lukes J, Hashimi H, Zikova A: Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Current genetics* 2005, 48(5):277-299.
 40. Hiesel R, Combettes B, Brennicke A: Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proceedings of the National Academy of Sciences of the United States of America* 1994, 91(2):629-633.
 41. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R *et al*: Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science (New York, NY)* 2003, 301(5633):653-657.
 42. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM: pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant molecular biology* 2000, 42(6):819-832.
 43. Clough SJ, Bent AF: Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 1998, 16(6):735-743.
 44. Weigel D, Glazebrook J: *Arabidopsis* : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2002.
 45. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning* : a laboratory manual, 2nd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1989.

46. Schagger H, von Jagow G: Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Analytical biochemistry* 1991, 199(2):223-231.

2.11 Figure Legends

Figure 10. Production of plants with altered expression of *atwhy2*.

A. Physical map of the *atwhy2* (AT1G71260) gene. The position of the T-DNA insertion in the KO line is indicated. The small arrows symbolize the primers used to amplify *atwhy2* mRNA by RT-PCR.

B. Molecular analysis of plants homozygous for the disrupted *atwhy2* allele. RT-PCR was performed on Col-0 and KO total RNA samples using two sets of primers (P1/P2 and P3/P4). Semi-quantitative conditions were used and primers for tubulin amplification were used as a control.

C. Levels of the AtWhy2-myc fusion protein in OEX and wild-type (Col-0) plants were monitored by Western blot using a monoclonal antibody against the c-myc epitope.

Figure 11. Phenotypic characterization of plants with altered expression of *atwhy2*.

A. Four week old plants of the indicated genotypes grown in soil were photographed.

B. Representative inflorescences and individual siliques taken from six week old plants of the indicated genotypes were photographed.

C. Twenty individual mature siliques from the indicated genotypes were dissected and their average seed content was calculated.

D. Equivalent leaves were taken from 6 week old plants and photographed. Leaves are ordered by age from left to right.

E. Early accumulation of senescence markers in mature leaves of OEX plants. RT-PCR was performed on Col-0, KO and OEX RNA samples taken from the 3rd (L3) and 4th (L4) leaves of 5 week old plants using oligonucleotides designed to amplify specifically the

following genes: At1g47128 : Cystein protease RD21A, At2g45570 : YLS6, Cell-death-associated cytochrome and At2g38860 : YLS5, Protease I. Semi-quantitative conditions and primers for β -tubulin amplification were used to ensure adequate loading for all samples.

Figure 12. Mitochondrial perturbation due to overproduction of AtWhy2.

A. Activity of mitochondrial respiratory chain complexes in plants with altered expression of *atwhy2*. BN-PAGE was used to separate membrane protein complexes from crude mitochondrial fraction taken from plants of the indicated genotypes. Activity of the different complexes was evaluated by in-gel enzymatic assays using equivalent protein amounts for all plants.

B. Mitochondria ultrastructure was evaluated using transmission electron microscopy. In the upper panel, representative mitochondria from the indicated genotypes were photographed at 25000 X magnification. The bar represents 0.5 μ m. In the lower panel, organelles were observed at 100000 X magnification. The bar represents 100 nm. Black arrows point to invaginations of the inner membrane (cristae).

Figure 13. Gene expression of mtDNA- and nuclear-encoded genes coding for mitochondrial proteins. A.

RNA gel blot analysis of mitochondrial genes. 10 μ g of total RNA extracted from 100 mg of mature leaf tissue were separated on a 1.2 % agarose gel and blotted onto a nylon membrane. PCR-amplified probes were used for the detection of the indicated genes. Ethidium bromide-stained 25S rRNA provides the RNA loading control. **B.** RT-PCR was performed on Col-0, KO and OEX total RNA extracted from leaves of 4 week old plants. Amplified genes are indicated. Semi-quantitative conditions and primers for β -tubulin amplification were used to ensure adequate loading for all samples.

Figure 14. Mitochondrial, plastidial and nuclear DNA content in plants with altered expression of *atwhy2*.

A. DNA gel blot analysis of mitochondrial, plastidial and nuclear genes. 10 μ g of total DNA digested with the indicated restriction enzymes were separated on a 1 % agarose gel, blotted onto a nylon membrane and hybridized with probes representing the indicated genes. Numbers on the left indicate the molecular weight of the observed bands in kilobasepairs (Kb) as predicted by restriction mapping of mtDNA.

B. Ratio of mitochondrial to nuclear genomes was evaluated by Quantitative Real-Time PCR using 2 different mitochondrial gene regions and a single nuclear genome region. The mt/nuclear DNA ratio of Col-0 plants was arbitrarily fixed as 100 %.

Figure 15. AtWhy2 is associated with mitochondrial DNA.

A. Position of the amplified regions on the mitochondrial genome.

B. PCR amplification of mtDNA regions following immunoprecipitation on crude mitochondrial extracts of the indicated genotypes. Inputs represent 20 % of the total DNA used in the immunoprecipitation. Representative regions are shown here, other amplified regions gave similar results showing specific interaction with AtWhy2.

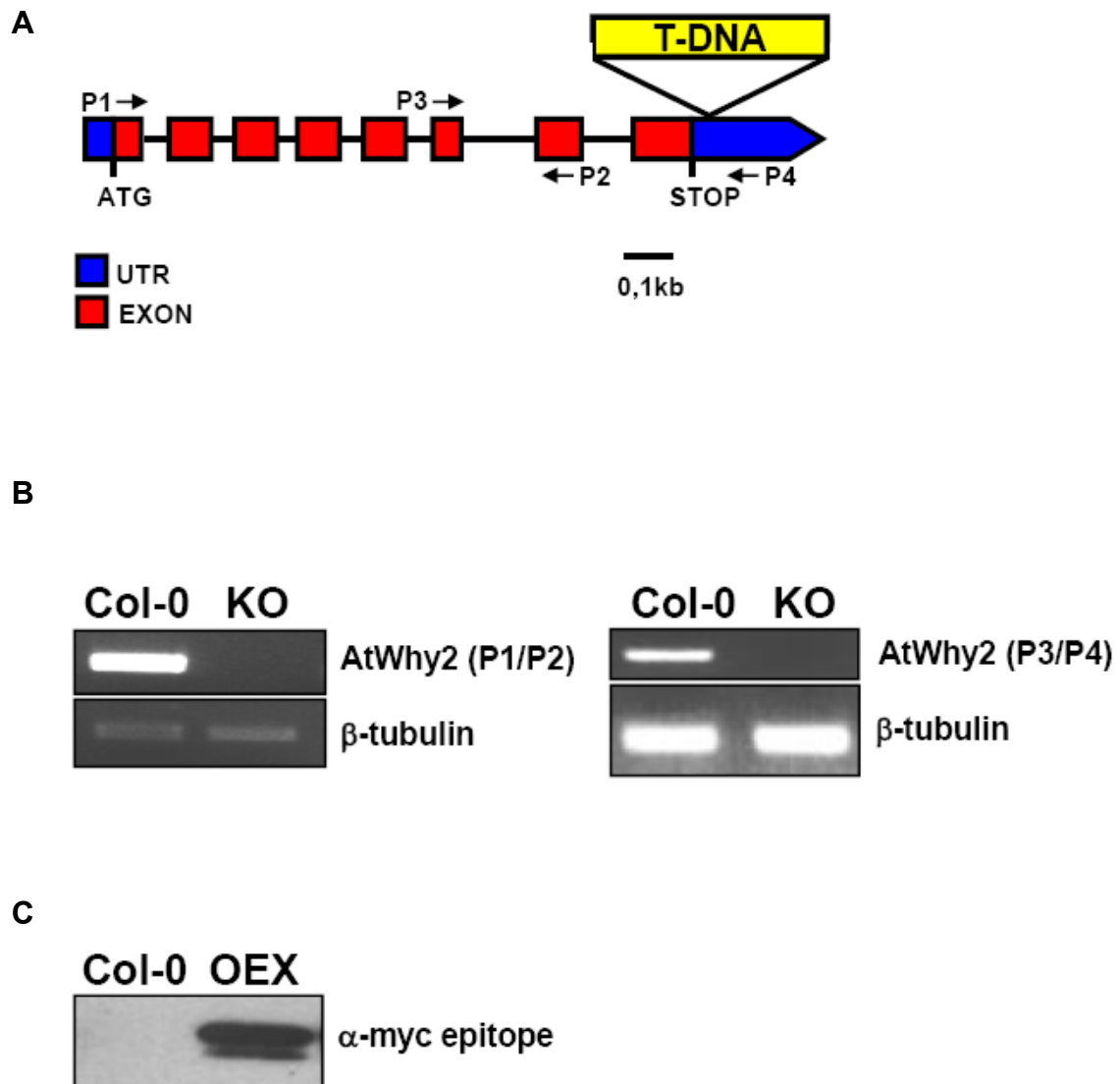


Figure 10. Production of plants with altered expression of *atwhy2*.

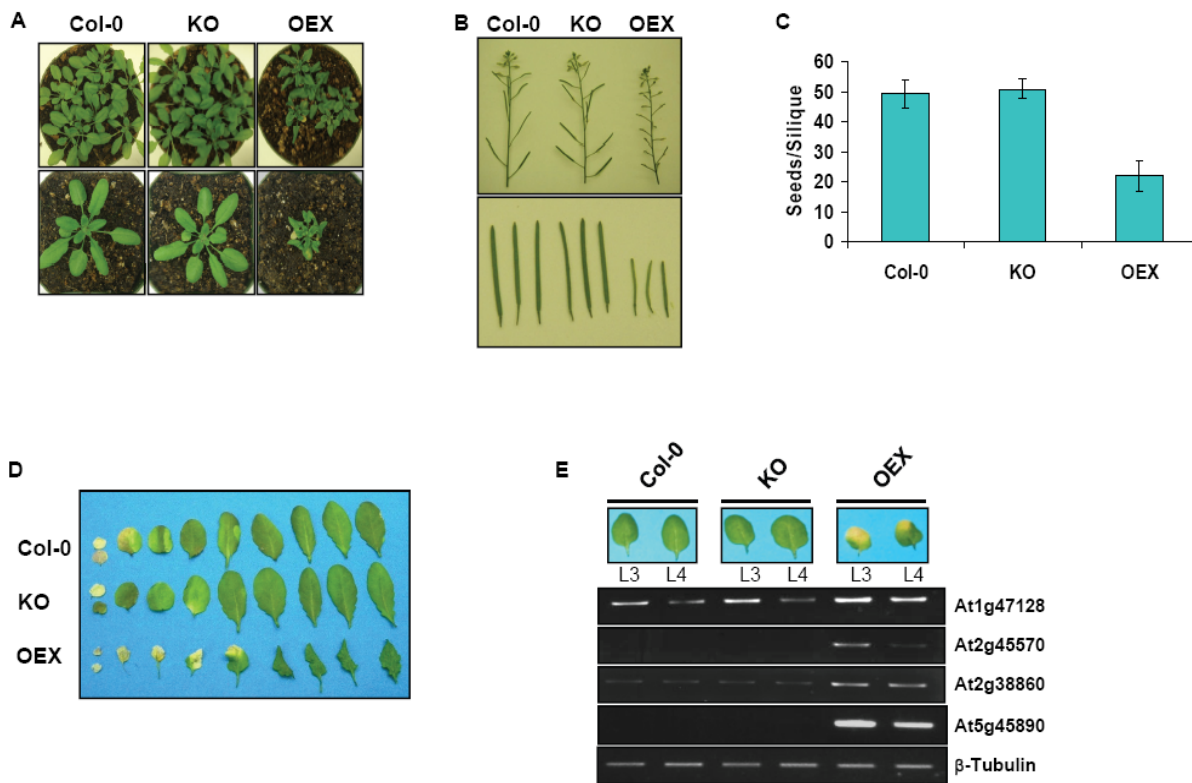


Figure 11. Phenotypic characterization of plants with altered expression of *atwhy2*.

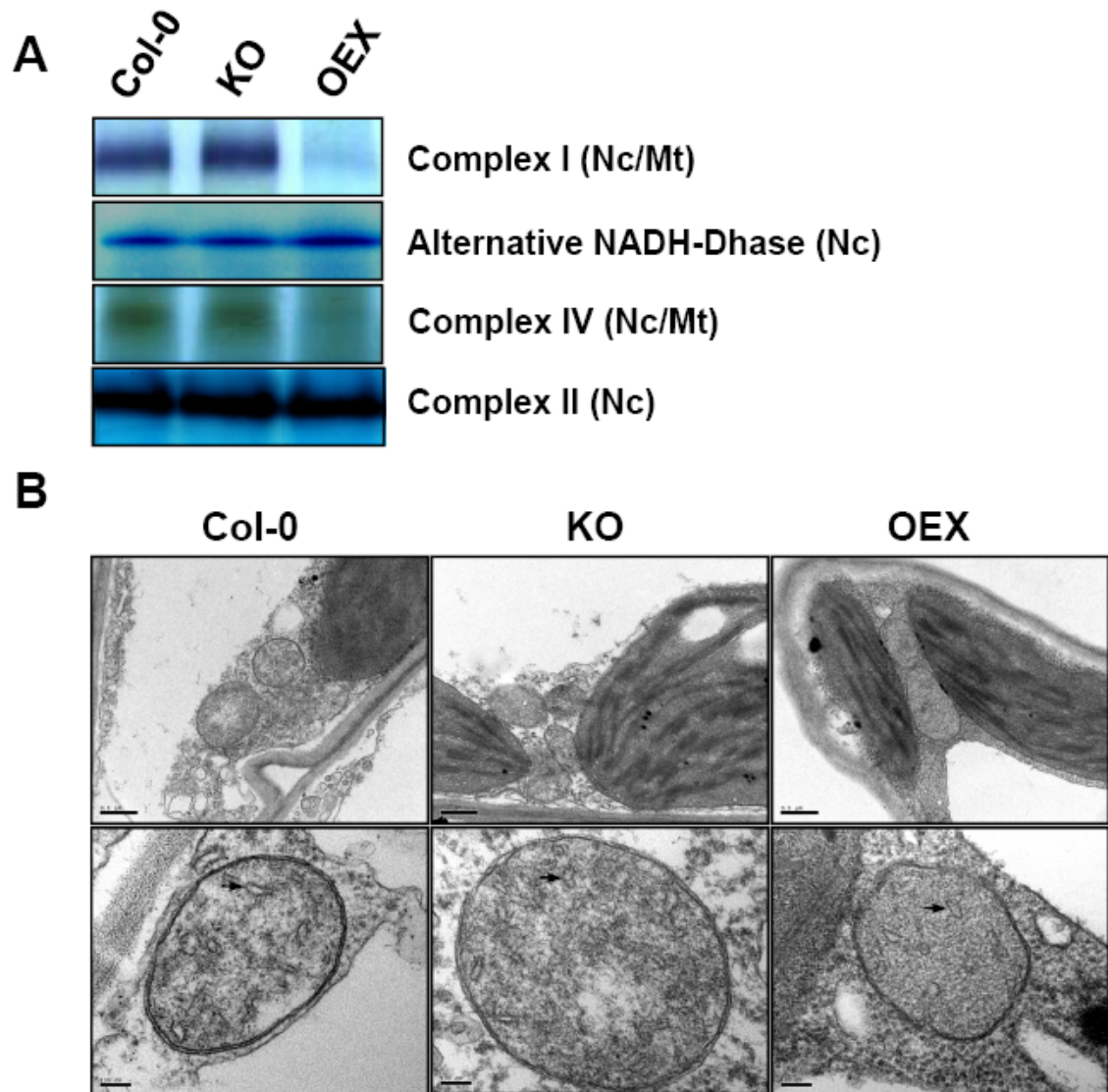


Figure 12. Mitochondrial perturbation due to overproduction of AtWhy2.

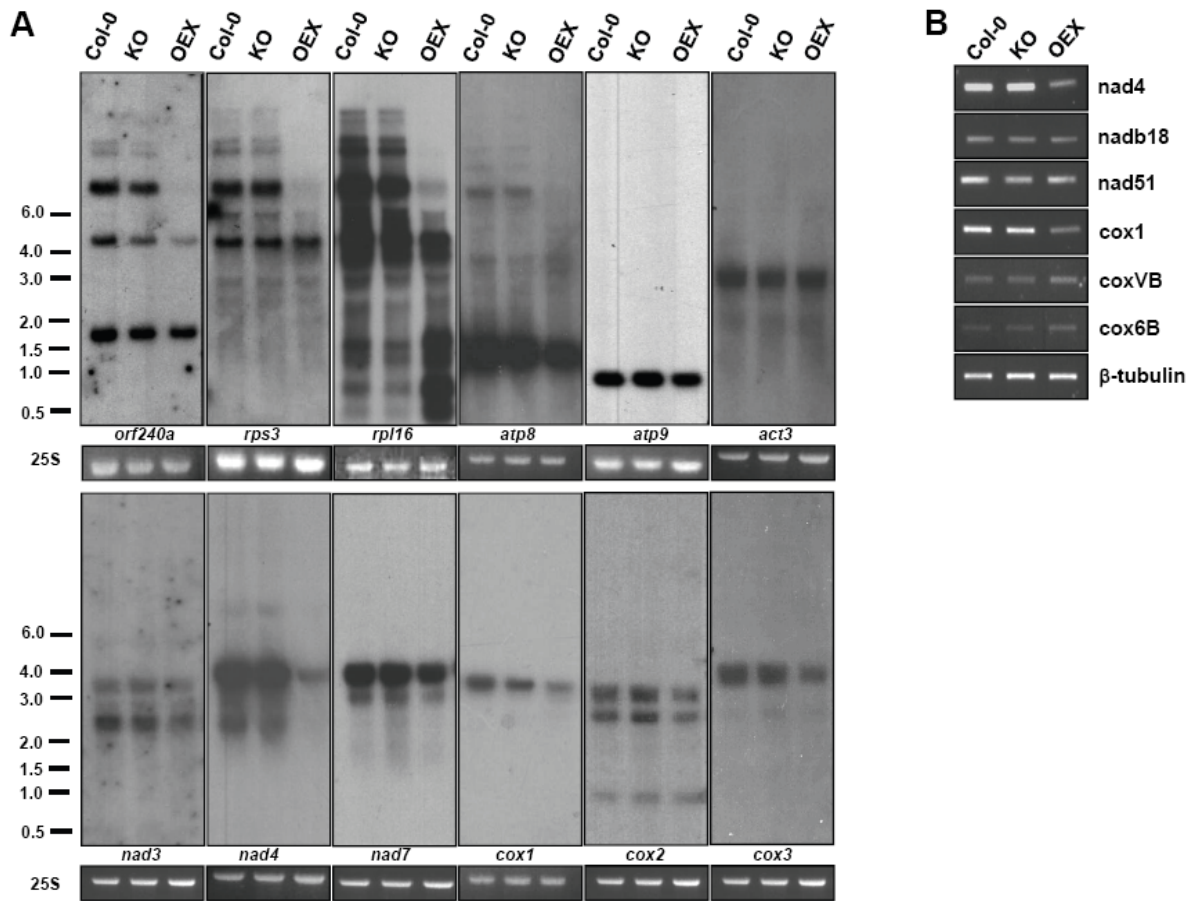


Figure 13. Gene expression of mtDNA- and nuclear-encoded genes coding for mitochondrial proteins.

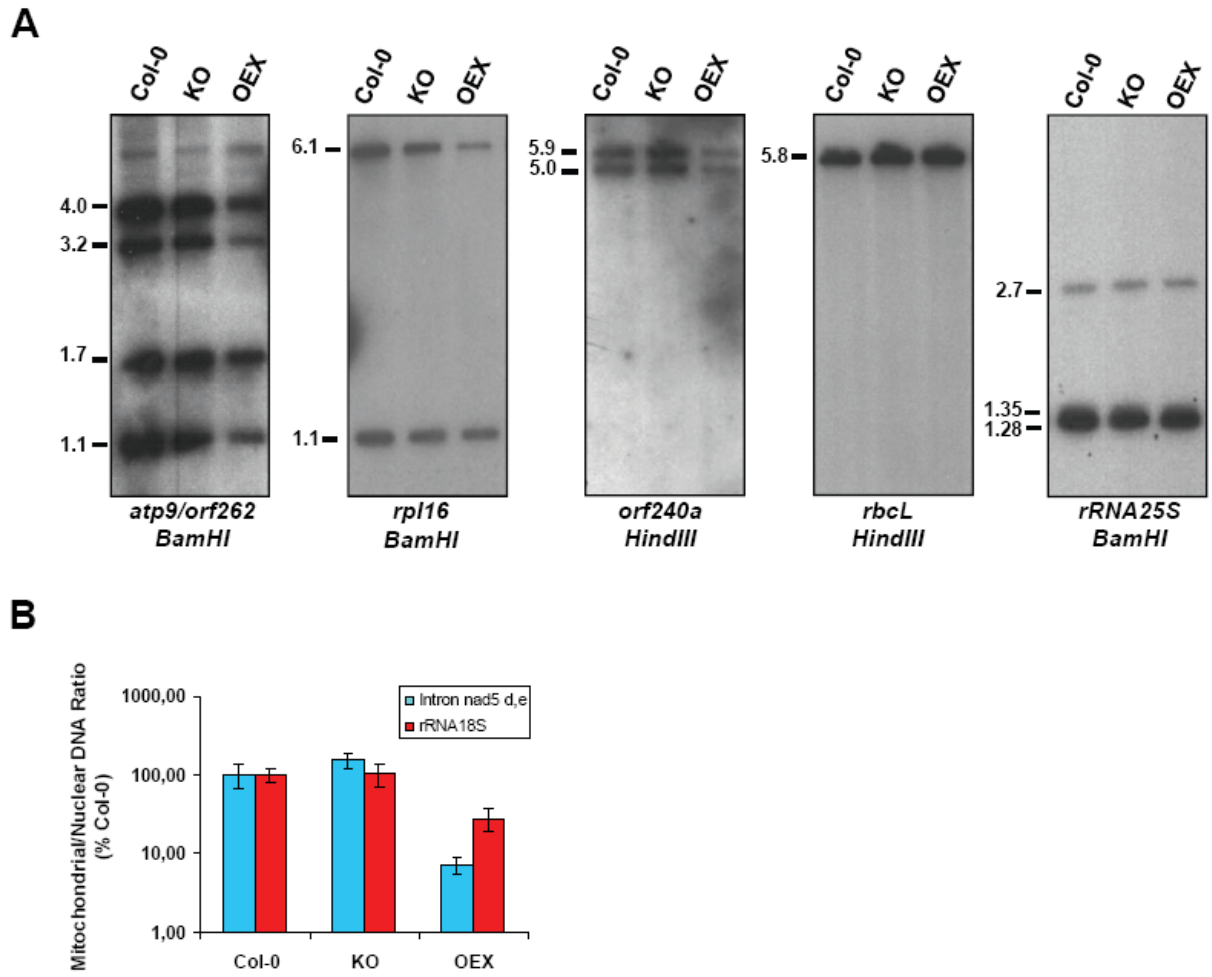


Figure 14. Mitochondrial, plastidial and nuclear DNA content in plants with altered expression of *atwhy2*.

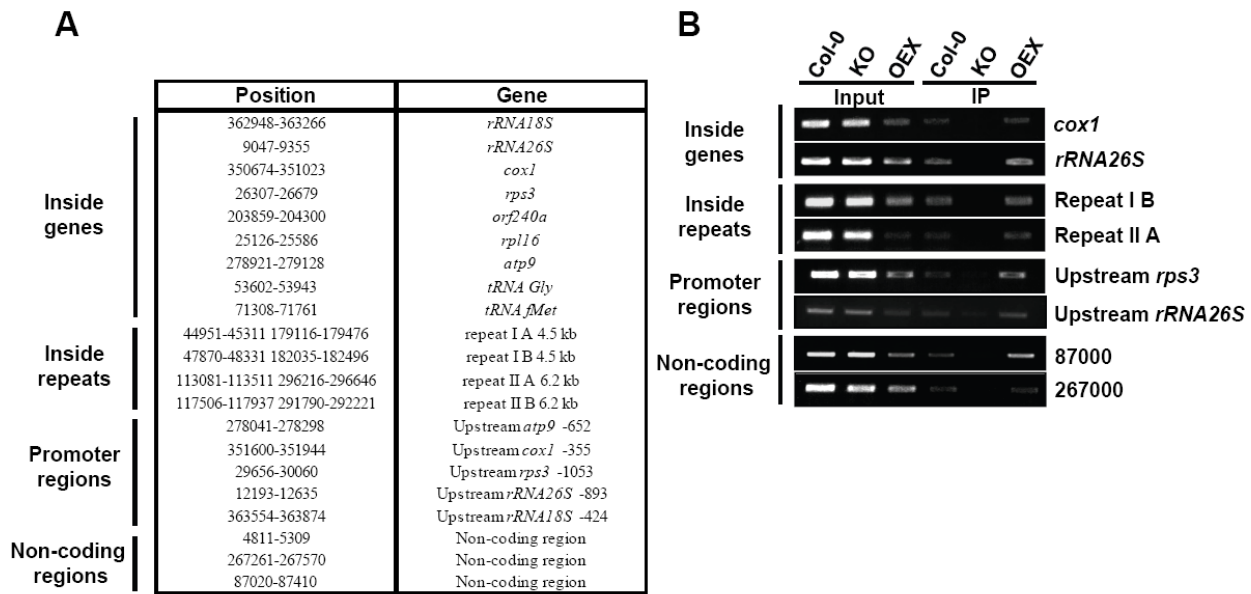


Figure 15. AtWhy2 is associated with mitochondrial DNA.

Chapitre 3. Whirly Proteins Maintain Plastid Genome Stability in *Arabidopsis*

Article sous presse dans Proceedings of the National Academy of Sciences of the United States of America

Whirly Proteins Maintain Plastid Genome Stability in *Arabidopsis*

Alexandre Maréchal*, Jean-Sébastien Parent*, Félix Véronneau-Lafortune,
Alexandre Joyeux, B. Franz Lang and Normand Brisson

*These authors contributed equally to this work.

Department of Biochemistry, Université de Montréal, PO Box 6128, Station
Centre-ville, Montréal, Québec, Canada, H3C 3J7.

Corresponding author: Normand Brisson,
Department of Biochemistry, Université de Montréal, PO Box 6128, Station Centre-ville,
Montréal, Québec, Canada, H3C 3J7 Phone : 514-343-6984 Fax : 514-343-2210
BIOLOGICAL SCIENCES: Plant Biology

19 text pages, 4 figures, and 4 tables.

Author contributions: A.M., J-S. P., A.J., B.F.L. and N.B. designed research; A.M., J-S. P.,
F. V.-L. and A.J. performed research; A.M., J.-S. P., A.J., B.F.L. and N.B. analyzed data;
and A.M., J-S. P., A.J., B.F.L. and N.B. wrote the paper.

The authors declare no conflict of interest.

3.1 Mise en contexte

L'objectif de cet article était de caractériser la fonction des Whirlies plastidiques. Pour ce faire, nous avons analysé les phénotypes de plantes incapables de produire les protéines AtWhy1 et/ou AtWhy3. Ceci nous a permis de découvrir un rôle important de ces protéines dans le maintien de la stabilité du génome plastidique. Pour cet article, j'ai caractérisé l'effet d'une délétion des gènes AtWhy1 et/ou AtWhy3 sur la stabilité du génome chloroplastique d'*Arabidopsis* (Figure 16C, D, E et le tableau 1). J'ai aussi réalisé les buvardages Southern (Figure 17 A, B, C, E, F) et le clonage des extrémités des régions amplifiées (Tableau 2). J'ai supervisé Sandra Grondin et participé à la réalisation des buvardages Southern de la Figure 25. J'ai également mis au point la technique d'immunoprécipitation de l'ADNpt (inspiré d'un protocole d'Alexandre Joyeux) et supervisé Félix Véronneau-Lafortune pour l'obtention de la Figure 22). J'ai produit les anticorps polyclonaux anti-AtWhy1/3. Pour la microscopie électronique, j'ai fixé les tissus et la préparation a été réalisée par Richard Janvier à l'Université Laval. J'ai aussi effectué les essais de retard sur gel de la figure 20 et produit la figure 24. Jean-Sébastien Parent a obtenu les plantes KOWhy1, KOWhy3 et KOWhy1/3 et produit le Western Blot final (Figure 1A et B). Il a également effectué les expériences de croisements pour démontrer l'héritabilité maternelle de la variégation. JSP et moi-même avons produit la figure 17D, la figure 18 ainsi que les tableaux 3, 4, 5 et 6. Alexandre Joyeux a réalisé la figure de microscopie confocale de localisation de StWhy1 (Figure 21). Finalement, j'ai écrit les sections résultats, discussion (incluant le modèle de la Figure 19) et informations supplémentaires qui ont ensuite été revues et corrigées par JSP, Normand Brisson et B. Franz Lang. JSP et NB ont écrit l'introduction qui a été revue par BFL et moi-même.

3.2 Abstract

Maintenance of genome stability is essential for the accurate propagation of genetic information and for cell growth and survival. Organisms have therefore developed efficient strategies to prevent DNA lesions and rearrangements. Much of the information concerning

these strategies has been obtained through the study of bacterial and nuclear genomes. Comparatively, little is known about how organelle genomes maintain a stable structure. Here we report that the plastid-localized Whirly single-stranded DNA binding proteins are required for plastid genome stability in *Arabidopsis*. We show that a double knockout of the genes *AtWhy1* and *AtWhy3* leads to the appearance of plants with variegated green/white/yellow leaves, symptomatic of non-functional chloroplasts. This variegation is maternally-inherited, indicating defects in the plastid genome. Indeed, in all variegated lines examined, reorganized regions of plastid DNA are amplified as circular and/or head-tail concatemers. All amplified regions are delimited by short direct repeats of 10 to 18 bp, strongly suggesting that these regions result from illegitimate recombination between repeated sequences. This type of recombination occurs frequently in plants lacking both Whirlies, to a lesser extent in single knockout plants and rarely in wild-type individuals. Maize mutants for the *ZmWhy1* Whirly protein also show an increase in the frequency of illegitimate recombination. We propose a model where Whirlies contribute to plastid genome stability by protecting against illegitimate repeat-mediated recombination.

3.3 Introduction

Plastids play diverse and essential roles in plants. Despite this central importance, surprisingly little is known about even the most basic aspects of the plastid genome structure, maintenance and propagation. For example, while the textbook depiction of plastid DNA (ptDNA) is that of a genome-sized circular DNA molecule, recent evidence suggests instead that most of the ptDNA is organized in concatenated, branched linear forms, with T4 phage-like features (1). This change of perception of plastid genome architecture requires a re-evaluation of the current rolling-circle model for plastid genome replication. It is now considered that a recombination-dependent replication process is responsible for the branched, multi-genomic structures present in plastids (1).

Recombination is also expected to play a crucial role in plastid genome maintenance. Indeed, due to its exposure to radiation and reactive oxygen species, the plastid genome is expected to accumulate mutations at a high rate. This stresses the importance of efficient DNA-replication, recombination and repair (DNA-RRR) pathways in these organelles (2). However, to date the mechanisms and enzymes involved in these pathways remain poorly characterized.

Evidence for recombination in plastid genomes abound in the literature (2). An example is the recombination between the large inverted repeat sequences present in many plastid genomes (3). This flip-flop recombination is responsible for the two isomers of ptDNA, which differ only with respect to the orientation of the single-copy regions. More direct evidence of recombination comes from plastid transformation experiments, which demonstrate that foreign DNA is integrated into ptDNA by homologous recombination (4).

Homologues of bacterial genes involved in DNA-RRR are present in the nuclear genome of plants, and some of their encoded proteins are targeted to plastids. These include the *recA* homologs *RECA1* (5) and *RECA2*, whose disruption is lethal in *Arabidopsis* (6), a Rec Q-like DNA helicase from rice (7), and genes for a gyrase A-like and two gyrase B-like subunits in *Arabidopsis* (8). Recently, two homologs of OSB1, a single-stranded DNA (ssDNA) binding protein that regulates recombination in mitochondria (9), were shown to localize to plastids. However, no role has yet been ascribed to these proteins. Finally, homologs of RPA, another ssDNA-binding protein family which plays an essential role in mammalian DNA-RRR, have recently been identified in plants. One member of this family is targeted to plastids (10).

Similar to many DNA-RRR proteins, Whirlies form a small family of ssDNA-binding proteins (11). They are involved in a variety of phenomena, ranging from pathogen defence (12) to telomeric homeostasis (13). In *Arabidopsis*, three Whirly genes are present and their proteins localize to organelles; AtWhy1 and AtWhy3 are targeted to plastids and AtWhy2 to mitochondria (14, 15).

Recent evidence indicates that Whirlies bind organelle DNA without apparent sequence-specificity *in vivo*. In *Arabidopsis*, AtWhy2 binds to many regions of the

mitochondrial genome with no obvious sequence consensus (15). Similarly, in maize, the plastid-localized *ZmWhy1* interacts with DNA from throughout the plastid genome (16). Knockdown mutations of *ZmWhy1* lead to ivory or pale-green plants, indicating that this Whirly is involved in chloroplast biogenesis. This phenotype was attributed to a defect in the maturation of the *atpF* and *23S* ribosomal RNAs, but the authors did not rule out participation of *ZmWhy1* in DNA recombination or repair.

To better understand the role of plastid-targeted Whirlies (ptWhirlies), we characterized an *Arabidopsis* double knockout (KO) line of the *AtWhy1* and *AtWhy3* genes (KO1/3). Variegation patterns, which appear on leaves in about 4.6% of the progeny, correlated with the selective rearrangement and amplification of large regions of the plastid genome. We show that the rearrangements are produced by illegitimate recombination at short direct repeats that border the amplified regions in intact ptDNA. We suggest that *AtWhy1* and *AtWhy3* function as anti-recombination proteins, contributing to safeguard plastid genome integrity.

3.4 Results

3.4.1 *Arabidopsis* ptWhirlies are Involved in Chloroplast Biogenesis

To investigate the role of Whirlies in plastids, we obtained mutant plants that no longer produce the *AtWhy1* (KO1) and/or *AtWhy3* (KO3) proteins (Fig. 16 A and B). Only one Whirly is detected in the single KOs, and no Whirlies are detected in KO1/3 extracts. We then monitored the ssDNA-binding activity of Whirlies by electrophoretic mobility shift assay. As shown in Figure 20 a strong DNA-binding signal is observed with Col-0 (wild-type) proteins, whereas less intense signals are detected with KO1 and KO3, respectively. No signal was observed in KO1/3, confirming the absence of ptWhirlies in these plants.

The KO1 and KO3 plants have no apparent phenotype. Interestingly, while most KO1/3 plants also have a wild-type (WT) appearance, some individuals have a smaller size and a variegated phenotype with white/yellow sectors on some leaves (Fig. 16C, upper

panel). In these sectors, a strong diminution in chlorophyll autofluorescence is observed (Fig. 16C, lower panel). Examination of plastids by electron microscopy reveals that in the white sectors, thylakoid stacks are replaced by large round vesicles, indicating that plastid development is compromised (Fig. 16D, right panels). By contrast, plastids from green sectors of the same leaf appear normal (Fig. 16D, left panels). We then evaluated the frequency of sectorized individuals in large populations of Col-0 and Whirly mutant plants (Tableau I). While no variegated individuals were recovered from Col-0, KO1 and KO3 populations, 4.6 % of KO1/3 plants had at least one variegated leaf sector (Tableau I). These data indicate that elimination of both *ptWhirlies* triggers changes that ultimately lead to strong interference with chloroplast development and function.

3.4.2 The Variegated Phenotype of KO1/3 Plants is Maternally-Inherited

The severity of the variegation phenotype in the KO1/3 population is variable. While some plants exhibit chlorosis on a single leaf, others have most of their leaves covered by yellow/white sectors (data not shown and Fig. 16E). Phenotypic differences were also observed between different variegated individuals, suggesting that the defects responsible for variegation differ from one plant to another (Fig. 16E). From this first generation of variegated KO1/3 plants, we selected two lines with a strong variegation phenotype (Var A and Var B) and set out to define the molecular basis of impaired plastid development.

Maternal inheritance of variegation is often linked to modifications of organelle genomes (9, 17, 18). Crosses were performed between variegated line Var B and Col-0 plants to determine the inheritance mode of the variegation. When variegated plants were used as male parents, no variegation was observed in the heterozygous F1 progeny. However, when Var B plants were used as female in the same cross, variegation was found in 46% of the progeny (47 out of 103 F1 plants). This indicates that variegation is maternally inherited and that reintroduction of *Whirlies* is unable to completely rescue the plastid defect, suggesting that the plastids in variegated sectors are irreversibly damaged, most likely at the genetic level.

3.4.3 Amplification of Reorganized ptDNA Regions in Variegated Plants

Since Whirlies bind DNA in organelles ((15, 16); Fig. 21 and 22), it is likely that variegation in KO plants results from mutations in their ptDNA. We thus searched for ptDNA rearrangements by DNA hybridization. Although no change was detected when comparing HindIII-digested Col-0 to non-variegated (green) KO1/3 DNA, unique amplified regions were identified in digested ptDNA of the Var A and Var B lines (Fig. 17A and B). Amplified DNA was estimated to be 10 to 25 times more abundant than WT ptDNA. In addition, for both variegated lines, new HindIII fragments (~5.8 kb in Var A and ~6.3 kb in Var B) were detected using probes located at the extremities of the amplified regions. These new bands indicate that the amplified regions are reorganized either as circular molecules and/or head-tail concatemers (Fig. 17C).

3.4.4 Reorganization of ptDNA is Caused by Recombination Between Short Direct Repeats

To map the extremities of the amplified DNA, we designed PCR oligonucleotides close to the ends of the amplified regions of Var A and Var B, facing opposite directions (outward-facing PCR; Fig. 17C). In a WT plastid genome, these primers would yield no product. However in a rearranged genome where circular or head-tail concatemers are present, we predict a new fragment containing both extremities of the amplified regions. As expected, when PCR is performed with DNA from Col-0 and green KO1/3 plants, no product is observed with both Var A and Var B primers (Fig. 17D). However, when the same primers are used on Var A and Var B DNA, amplification products appear in both cases. Interestingly, more amplified products are detected in variegated *versus* non-variegated leaves of the same plants, suggesting that the appearance of defective chloroplasts is related to the abundance of reorganized ptDNA molecules (Fig. 17D; compare leaves 4 and 8).

The position of the ends of amplified regions was determined by cloning and sequencing the PCR products. These positions matched perfectly with the HindIII digestion patterns of the reorganized regions (Fig. 17A, B, C and Table 2). Remarkably, the regions

of WT plastid genome that correspond to the amplified regions in Var A and Var B are bordered by short direct repeats of 10 and 14 bp, respectively (Tableau II). In the PCR fragments, both ends of the amplified regions are joined by a single repeat, indicating that recombination occurred between these short sequences. Analysis of three additional variegated lines also revealed unique amplified regions with ends joined by recombination at short repeats (Tableau II). Interestingly, in the Var C and Var E lines, the bordering repeats are not identical and carry a few mismatches. Altogether, our data strongly suggest that variegation is induced by an overabundance of amplified recombined ptDNA regions.

Additional DNA hybridization experiments were performed to determine the arrangement of the amplified DNA. DNA from Col-0, green KO1/3 and Var A plants was digested either with XhoI and PstI restriction enzymes that cut on both sides of the Var A region in WT ptDNA, or with KpnI which cleaves once in this region (Fig. 17E and F). A probe specific for the Var A amplified region was used for detection. Intact Col-0 and green KO1/3 DNA migrated within the unresolved compression zone (above ~ 10 kb). Six bands were detected in undigested Var A DNA, including a fast-migrating band of ~ 4.5 kb. When DNA was digested with XhoI and PstI, the Var A specific bands were unaffected (Fig. 17E), indicating that these DNA molecules are extrachromosomal. Finally, after KpnI digestion, all Var A specific bands collapsed into a single 6.25 kb band, showing that these forms are concatemers of the same repeat unit. This also suggests that the fast migrating band in undigested Var A DNA corresponds to a circular molecule, possibly a monomeric one. Similar results were obtained with the Var B line (data not shown). We thus conclude that amplified ptDNA is present as both circular and/or head-tail concatemers in variegated plants.

3.4.5 Illegitimate Recombination is Increased in the Absence of *Arabidopsis* ptWhirlies

Our results suggest that the absence of ptWhirlies induces plastid genome instability through an increase in recombination between short direct repeats. One could therefore expect to detect low-level recombination events in non-variegated Whirly mutant plants.

We tested this hypothesis using the outward-facing PCR approach described above, on each of four independent DNA pools from Col-0, KO1, KO3 and green KO1/3 plants. Representative results of the PCR amplification are shown in Fig. 18 (panels A, B and C). Rearranged products were observed in all genotypes.

Cloning and sequencing of rearranged DNA confirmed that illegitimate recombination is strongly increased in green KO1/3 plants, where 40 different recombination products are identified out of 30 PCR reactions (Tableaux III, IV and V). Recombination products were also detected in single KOs, although less frequently, indicating that the depletion of AtWhy1 or AtWhy3 is sufficient to increase spurious recombination. Surprisingly, two recombination events were detected in Col-0 samples. These were present in all genotypes tested (Tableaux III and V), suggesting the presence of a small sub-population of rearranged ptDNA even in WT plants.

3.4.6 Illegitimate Recombination is Increased in Maize Whirly Mutants

We verified in three maize *ZmWhy1* mutant lines whether mutation of monocot ptWhirlies also affects plastid genome stability. The lines *ZmWhy1-1*, *ZmWhy1-2* and the heteroallelic progeny of complementation crosses *ZmWhy1-2/-1* all have a reduced level of the plastid-localized protein ZmWhy1. These lines exhibit ivory (*ZmWhy1-1*), pale green (*ZmWhy1-2*) and intermediate (*ZmWhy1-2/-1*) leaf phenotypes (16). Our Southern hybridizations did not reveal amplified plastid genome regions in any of the three lines (Fig. 25), indicating that the defect in chloroplast biogenesis in *ZmWhy1* mutants is not linked to the presence of amplified ptDNA regions. Outward-facing PCR analysis of the B73 control maize line revealed a single short-repeat mediated illegitimate recombination event for the 19 primer pairs tested (Tableaux VI and VII). The same primers revealed 3 events in lines *ZmWhy1-2* and *ZmWhy1-2/-1*, respectively. Illegitimate recombination was highest in the most severe mutant line (*ZmWhy1-1*), with a total of 14 different events (Tableaux VI and VII). These results suggest that Whirlies are also involved in stabilizing the plastid genome in maize.

3.5 Discussion

3.5.1 Whirlies Prevent Short Homology-Dependent Illegitimate Recombination

Accumulation of reorganized DNA through illegitimate recombination has also been observed in bacteria. In *E. coli*, stress can induce selective amplification of genome regions conferring tolerance to the applied stress, a phenomenon called adaptive amplification. These regions (~10 to 30 kb) are first duplicated through illegitimate recombination between microhomologous repeats (5-15 bp), and are then further replicated through a mechanism requiring the homologous recombination machinery (reviewed in (19)). Another recombination process, called 'short homology-dependent illegitimate recombination' exists in bacteria (reviewed in (20)). It also requires microhomology (3-20 bp) at recombination sites, and its occurrence is increased by various DNA-damaging agents. Similarly, microhomology-mediated recombination has recently been proposed to trigger segmental DNA duplications that are linked to some genetic diseases in humans and to copy number variations (CNV) among individual genomes (21-23). It was also observed in yeast cells undergoing segmental duplications (24). Interestingly, in bacteria, DNA-RRR proteins such as the subunits of the UvrAB complex, the RecQ helicase, the histone-like H-NS and the single-stranded DNA binding protein SSB can suppress microhomology-mediated recombination (25-29). However, the genetics requirements for this suppression may vary with the system used to monitor recombination. We propose a similar role for Whirlies in plastids, where they contribute to maintain genome stability by preventing accumulation of illegitimate recombination products.

3.5.3 A Model for the Variegation in *Arabidopsis* Plants lacking ptWhirlies

The direct repeat sequences at *Arabidopsis* ptDNA junctions are too short to serve as substrates for homologous recombination (30, 31). Recently, an alternative model was proposed to explain how microhomologies may generate large DNA rearrangements in *E.*

coli, yeast and human (21). In this model, the collapse of a replication fork leads to the generation of a 3' protruding ssDNA end. This 3' tail may anneal to a microhomologous ssDNA sequence at a different location, allowing re-initiation of replication at this new site. When such template switching occurs between two different DNA molecules, either duplications (switch to a position behind the fork) or deletions (switch to a position ahead of the fork) will occur. Alternatively, switching on the same DNA molecule behind a stalled replication fork leads to rolling-circle replication which produces concatemers of subgenomic regions, and eventually circular DNA products (Fig. 17E). In addition to duplication/circularization events that were detected by DNA hybridization and outward-facing PCR, deletions were also identified using inward-facing PCR in *Arabidopsis* Whirly mutants (Tableau V). In summary, the above replication-based mechanism might explain the variety of microhomology-mediated rearrangements observed in the plastid genome of Whirly mutants.

Based on our results, we propose the following model for the emergence of variegation in KO1/3 plants (Fig. 19). In Col-0, the genome surveillance machinery repairs DNA lesions and, with the help of Whirlies, prevents accumulation of illegitimate DNA recombination products, which are rarely detected in these plants. In the single KO plants, the absence of one ptWhirly is enough to induce a low-level accumulation of rearranged molecules. However, when both Whirlies are eliminated, aberrant molecules accumulate to high levels. Some of these molecules, most likely circularized products, could replicate independently of the main plastid genome and eventually become overabundant. In each of the five variegated lines analysed in detail (Tableau II) only one rearranged molecule is amplified. Although a certain level of these molecules can be tolerated by the plastid, as demonstrated by their presence in green tissue of variegated plants (Fig. 17D), their accumulation eventually gives rise to plastids that are non-functional, thus leading to the appearance of variegated leaf sectors. No deletions resulting in variegation were detected, suggesting that their accumulation is either lethal and/or that they are more easily eliminated by gene conversion (32).

3.5.4 Conservation of the Role of ptWhirlies in the Maintenance of Plastid Genome Stability

The finding that illegitimate recombination mediated by short repeats is also enhanced in a severely affected maize Whirly mutant line suggests a conserved function for Whirlies in maintaining plastid genome stability. *ZmWhy1* hypomorphic mutants show a less clear trend than the *ZmWhy1-1* mutant, which might be due to the presence of a higher residual level of ZmWhy1 protein in these lines (16). This situation resembles that of the single KO1 and KO3 *Arabidopsis* Whirly mutants, which show less illegitimate recombination than the double KO. The low level of ZmWhy1 protein present in all *ZmWhy1* mutants may also explain why no amplified ptDNA regions are found in these plants. In fact, the level of residual ptWhirly protein in the maize mutants appears similar to that present in the *Arabidopsis* KO1 mutant, which also does not contain amplicons. Amplified regions were only observed in KO1/3, the only line analysed which is completely devoid of ptWhirlies, suggesting that a low level of Whirlies is sufficient to prevent amplification of recombined ptDNA molecules.

The absence of detectable amplicons and the severe first generation phenotypes in the maize Whirly mutants indicate that a different mechanism contributes to the formation of defective chloroplasts in this species. Actually, the *ZmWhy1* mutants exhibit a reduced content of plastid ribosomes, with the albino *ZmWhy1-1* seedlings almost completely lacking plastid rRNA (16). This ribosome deficiency was proposed to result from a defect in the biogenesis of the large ribosomal subunit due to an aberrant 23S rRNA metabolism, suggesting a role for Whirlies in plastid RNA metabolism (16). Although we cannot rule out such a role in *Arabidopsis*, it is unlikely that the more subtle variegation phenotype observed in this species is due to aberrant rRNA metabolism as no major changes are detected in 23S and 16S rRNA levels in the *Arabidopsis* ptWhirly mutants (Fig. 26). Thus, Whirlies appear to form a flexible family of single-stranded nucleic acid binding proteins that can fulfill a variety of roles, depending on the cellular context and/or plant species.

3.5.5 Characteristics of the Short Direct Repeats

Some of the direct repeats reported here appear to be particularly prone to illegitimate recombination as they are used repeatedly and at multiple different positions of the plastid genome (see related events in Tableaux V and VI). This suggests that a specific sequence motif or structure promotes their use as substrates. Homopolymeric A/T stretches appear to be over-represented among the direct repeats (Tableaux II, III and Tableaux V and VI). This type of sequence produces unusual structures called slipped-strand DNA, which can cause replication stalling (reviewed in (33)). Stalled replication forks can be reinitiated by template-switching, and can also generate double-strand ends (reviewed in (34)), potentially lead to an increase in illegitimate recombination. Alternatively, these poly A/T sequences could have a greater tendency to melt, providing easier access to invading ssDNA 3' tails and thus favouring ptDNA rearrangements.

3.5.6 Whirlies Could Prevent Illegitimate Recombination by Protecting ssDNA

Many ssDNA binding proteins bind non-specifically to DNA and play important roles in recombination processes. Bacterial SSBs and their eukaryotic nuclear counterparts such as RPA play essential protective roles in genome biology by protecting ssDNA from damage and preventing spurious DNA annealing. They also serve to recruit a large number of genome maintenance proteins (reviewed in (35, 36)). In *E. coli*, SSB also suppresses short homology-dependent illegitimate recombination (27, 28). It is thus tempting to speculate that in plastids, Whirlies play a role similar to that of SSB in *E. coli*. This is supported by our data (Fig. 22 and (15)) and those of Prikryl *et al.* (16), indicating that Whirlies bind DNA in a non-sequence specific manner. In addition, similar to bacterial SSBs, AtWhy1 protects ssDNA against degradation by nucleases *in vitro* (Cappadocia, L., Maréchal, A., Lepage, E., Sygush, J. and Brisson, N., unpublished results). We therefore propose that in plastids, the absence of Whirlies enhances the availability of free ssDNA. This might lead to increased DNA damage due to the vulnerability of the exposed DNA,

with a subsequent increase in recombination-repair mechanisms. More importantly, the damaged ssDNA might increase the rate of illegitimate recombination by increasing the frequency of collapsed replication forks (reviewed in (37)). In any case, stresses able to induce DNA damage would also increase ssDNA availability, as happens after processing of double-strand breaks, leading to increased recombination (28).

Our experiments define Whirlies as new and important components of the plastid genome maintenance machinery. As a follow-up, it will be interesting to determine if stresses that alter the integrity of ptDNA also lead to increased illegitimate recombination both in Whirly mutants and in plants deficient for putative plastid DNA-RRR proteins. This will help precise the function of Whirlies in DNA protection/repair processes and will permit identification of new regulators of plastid genome stability.

3.6 Materials and Methods

3.6.1 Mutant characterization

The SALK Institute Genomic Analysis Laboratory provided the sequence-indexed T-DNA insertion line (38). The Seattle TILLING Project (39) provided plants with mutations in the *AtWhy3* gene. The maize mutant lines *ZmWhy1-1*, *ZmWhy1-2* and *ZmWhy1-2/-1* were obtained from Dr. Alice Barkan (16).

3.6.2 Antibody production

Recombinant AtWhy1 was purified as described previously (40). Rabbits were immunized and antiserum was collected. For protein gel blot analysis, the antiserum was used at a concentration of 1:4000.

3.6.3 DNA gel blots

DNA was isolated from plants using a Cetyl trimethylammonium bromide DNA extraction protocol (41). Running of the samples and blotting of the gels was performed as described (15).

3.6.4 Detection of DNA rearrangements by PCR in *Arabidopsis*

PCR reactions were conducted using the Taq polymerase from Genscript and a series of outward-facing oligonucleotides spaced by approximately 10-20 kb. Inward-facing primers spaced by 10-20 kb were used to detect deletion events. Thirty reactions were performed on each of four independent DNA samples from Col-0, KO1, KO3 and green KO1/3 plants and analysed by gel electrophoresis. All visible DNA bands were isolated, cloned and sequenced. See SI Text.

3.7 Acknowledgments

We thank Dr. C.S. Hardtke for his help with the *Arabidopsis* crosses, R.M. Gaafar for help with genotyping, S. Grondin, P. Tan and A. Moreau for technical assistance and L. Forget for the sequencing of ptDNA. We thank Dr. A. Barkan for the generous gift of the *ZmWhy1* mutant lines. A.M. was supported by scholarships from the Natural Science and Engineering Research Council of Canada (NSERC) and from the Fonds Québécois de la Recherche sur la Nature et les Technologies. This research was supported by a grant from NSERC.

3.8 References

1. Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16(7):1661-1666.
2. Day A, Madesis, P. (2007) DNA replication, recombination, and repair in plastids. *Cell and Molecular Biology of Plastids, Topics in Current Genetics*, ed Bock R (Springer-Verlag Berlin Heidelberg, Leipzig), Vol 19, pp 65-119.
3. Palmer JD (1983) Chloroplast DNA exists in 2 orientations. *Nature* 301:92-93.
4. Staub JM & Maliga P (1992) Long regions of homologous DNA are incorporated into the tobacco plastid genome by transformation. *Plant Cell* 4(1):39-45.
5. Cao J, Combs C, & Jagendorf AT (1997) The chloroplast-located homolog of bacterial DNA recombinase. *Plant Cell Physiol* 38(12):1319-1325.

6. Shedge V, Arrieta-Montiel M, Christensen AC, & Mackenzie SA (2007) Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell* 19(4):1251-1264.
7. Saotome A, et al. (2006) Characterization of four RecQ homologues from rice (*Oryza sativa* L. cv. Nipponbare). *Biochem Biophys Res Commun* 345(4):1283-1291.
8. Wall MK, Mitchenall LA, & Maxwell A (2004) *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci U S A* 101(20):7821-7826.
9. Zaegel V, et al. (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell* 18(12):3548-3563.
10. Ishibashi T, Kimura S, & Sakaguchi K (2006) A higher plant has three different types of RPA heterotrimeric complex. *J Biochem* 139(1):99-104.
11. Desveaux D, Maréchal A, & Brisson N (2005) Whirly transcription factors: defense gene regulation and beyond. *Trends Plant Sci* 10(2):95-102.
12. Desveaux D, Després C, Joyeux A, Subramaniam R, & Brisson N (2000) PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. *Plant Cell* 12(8):1477-1489.
13. Yoo HH, Kwon C, Lee MM, & Chung IK (2007) Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in *Arabidopsis*. *Plant J* 49(3):442-451.
14. Krause K, et al. (2005) DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS Lett* 579(17):3707-3712.
15. Maréchal A, et al. (2008) Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol* 8:42 .
16. Prikryl J, Watkins KP, Friso G, van Wijk KJ, & Barkan A (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res* 36(16):5152-5165.
17. Chang TL, et al. (1996) Characterization of primary lesions caused by the plastome mutator of *Oenothera*. *Curr Genet* 30(6):522-530.

18. Martinez-Zapater JM, Gil P, Capel J, & Somerville CR (1992) Mutations at the *Arabidopsis* CHM locus promote rearrangements of the mitochondrial genome. *Plant Cell* 4(8):889-899.
19. Hastings PJ (2007) Adaptive amplification. *Crit Rev Biochem Mol Biol* 42(4):271-283 .
20. Ehrlich SD (1989) Illegitimate recombination in bacteria. *Mobile DNA*, ed Berg DE, M-Howe, S.M.), pp 799-831.
21. Hastings PJ, Ira G, & Lupski JR (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 5(1):e1000327.
22. Redon R, et al. (2006) Global variation in copy number in the human genome. *Nature* 444(7118):444-454.
23. Lee JA, Carvalho CM, & Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131(7):1235-1247.
24. Payen C, Koszul R, Dujon B, & Fischer G (2008) Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet* 4(9):e1000175.
25. Hanada K, Iwasaki M, Ihashi S, & Ikeda H (2000) UvrA and UvrB suppress illegitimate recombination: synergistic action with RecQ helicase. *Proc Natl Acad Sci U S A* 97(11):5989-5994.
26. Hanada K, et al. (1997) RecQ DNA helicase is a suppressor of illegitimate recombination in *Escherichia coli*. *Proc Natl Acad Sci U S A* 94(8):3860-3865.
27. Mukaihara T & Enomoto M (1997) Deletion formation between the two *Salmonella typhimurium* flagellin genes encoded on the mini F plasmid: *Escherichia coli* *ssb* alleles enhance deletion rates and change hot-spot preference for deletion endpoints. *Genetics* 145(3):563-572.
28. Reddy M & Gowrishankar J (1997) Identification and characterization of *ssb* and *uup* mutants with increased frequency of precise excision of transposon Tn10 derivatives: nucleotide sequence of *uup* in *Escherichia coli*. *J Bacteriol* 179(9):2892-2899.
29. Shanado Y, Hanada K, & Ikeda H (2001) Suppression of gamma ray-induced illegitimate recombination in *Escherichia coli* by the DNA-binding protein H-NS. *Mol Genet Genomics* 265(2):242-248 .

30. Lovett ST, Hurley RL, Sutera VA, Jr., Aubuchon RH, & Lebedeva MA (2002) Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics* 160(3):851-859.
31. Liskay RM, Letsou A, & Stachelek JL (1987) Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* 115(1):161-167.
32. Khakhlova O & Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J* 46(1):85-94.
33. Mirkin EV & Mirkin SM (2007) Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 71(1):13-35.
34. Michel B, Grompone G, Flores MJ, & Bidnenko V (2004) Multiple pathways process stalled replication forks. *Proc Natl Acad Sci U S A* 101(35):12783-12788.
35. Zou Y, Liu Y, Wu X, & Shell SM (2006) Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. *J Cell Physiol* 208(2):267-273.
36. Shereda RD, Kozlov AG, Lohman TM, Cox MM, & Keck JL (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol* 43(5):289-318.
37. Persky NS & Lovett ST (2008) Mechanisms of recombination: lessons from *E. coli*. *Crit Rev Biochem Mol Biol* 43(6):347-370.
38. Alonso JM, et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301(5633):653-657.
39. Till BJ, et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* 13(3):524-530.
40. Desveaux D, et al. (2004) A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*. *Dev Cell* 6(2):229-240.
41. Weigel D & Glazebrook J (2002) *Arabidopsis* : a laboratory manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) pp xii, 354 p.
42. Sato S, Nakamura Y, Kaneko T, Asamizu E, & Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res* 6(5):283-290.

43. Kunnimalaiyaan M & Nielsen BL (1997) Fine mapping of replication origins (ori A and ori B) in *Nicotiana tabacum* chloroplast DNA. *Nucleic Acids Res* 25(18):3681-3686.

3.9 Figure Legends

Fig. 16. *AtWhy1* and *AtWhy3* are involved in the biogenesis of chloroplasts. (A) Physical maps of the *AtWhy1* (AT1G14410) and *AtWhy3* (AT2G02740) genes. The position of the T-DNA insertion in the KO1 line is indicated. KO3 is a TILLING line with a mutation that changes a TGG codon to a TGA stop codon in the *AtWhy3* gene. An asterisk indicates the position of the mutation. (B) Protein gel blot analysis of simple and double *ptWhirlies* knockout plants. Crude plastid proteins were separated by SDS-PAGE on a 15 % polyacrylamide gel. Whirlies were detected using an anti-*AtWhy1/3* antibody. A section of the blot stained with Ponceau red is presented in the lower panel as a loading control. (C) In the upper panel, 4 week-old individuals of the indicated genotypes are shown. In the lower panel, fluorescence of chlorophyll was visualized by confocal microscopy. (D) Transmission electron microscopy of sections from green (left panels) and white (right panels) sectors of variegated leaves of KO1/3 plants. The bar represents 10 μm in the upper panels and 2 μm in the lower panels. (E) Variegation phenotype varies between independent lines. Four week-old individuals from the two variegated lines Var A and Var B are shown.

Fig. 17. Variegated plants contain rearranged amplified plastid DNA regions. (A) DNA gel blot (10 $\mu\text{g}/\text{lane}$) of total leaf DNA digested with *HindIII* and hybridized with the probes indicated below the gels. The probe numbers refer to the nucleotides of the published *Arabidopsis* chloroplast genome (42). Expected fragments from restriction analysis of Col-0 *ptDNA* and the size of new fragments observed in variegated lines are presented below the probes. A lower exposition of the second panel allowing better visualization of the amplified bands in VarA is presented in Fig. 23. (B) Restriction map of the reorganized regions in variegated lines. The red and green arrows represent the amplified regions in Var A and Var B respectively. Probes used in DNA gel blot experiments are represented as blue

lines. The HindIII restriction sites are represented as vertical black lines and the horizontal black line corresponds to ptDNA. The expected restriction fragments in base pairs (bp) are shown underneath the ptDNA. A gene map of this region of *Arabidopsis* ptDNA is presented in Fig. 24. (C) Schematic of the possible arrangements of the reorganized ptDNA in variegated lines. A head-tail dimer and a monomeric circular molecule are represented for Var A (red) and Var B (green). Oligonucleotides used for the PCR-amplification of the junctions of reorganized ptDNA are represented by small black arrows. (D) PCR amplification of fragments containing the junctions of reorganized ptDNA in Var A and B plants (bottom panel). DNA from leaves 4 (variegated) and 8 (non-variegated) was isolated (upper panel). The plastidial *ycf2* gene was used as a loading control. (E) DNA gel blot analysis showing the arrangement of amplified ptDNA in the Var A line. DNA from the indicated genotypes was digested with the indicated restriction enzymes and separated on an agarose gel. The DNA was hybridized with the probe depicted in F. A 9.2 kb band corresponding to the WT DNA fragment appears in all samples digested with XhoI and PstI restriction enzymes. A band of 9.4 kb expected from digestion of the WT plastid genome with KpnI was found in all genotypes. The asterisks indicate putative circular molecules. (F) Restriction map of the reorganized regions in the Var A line. The red arrow represents the amplified region in Var A. The probe used is represented as a blue line. A portion of ptDNA is represented as a black horizontal line. The restriction sites are indicated by vertical black lines. K = KpnI, P = PstI, X = XhoI. A circular monomer is represented on the right with the expected linear digestion product of this molecule.

Fig. 18. Illegitimate recombination is increased in the absence of *Arabidopsis* ptWhirlies. For each genotype, PCR reactions using outward or inward-facing PCR primers were performed on 4 pools of DNA from 4 different plants. Reactions were run on agarose gels containing ethidium bromide. (A-C) Representative PCR reactions are shown. The oligonucleotides used are indicated above each panel. Individual bands (white numbers) were cut, cloned and sequenced. Each band represents a unique recombination product (Table 3). The asterisks indicate non-recombinant products arising from non-specific

hybridization of the 69633F primer at positions 58720-58733 of the plastid genome. (D)
The plastidial *ycf2* gene was used as a loading control.

Fig. 19. Model for the appearance of variegation in plants lacking Whirlies. Very low levels of rearranged products are found in Col-0 plants. These products are more abundant in single Whirly KO plants. Rearranged product levels are highest in KO1/3 plants. In these plants, some rearranged molecules can reach 10-25 times the levels of the normal plastome and thus compromise the development and function of chloroplasts.

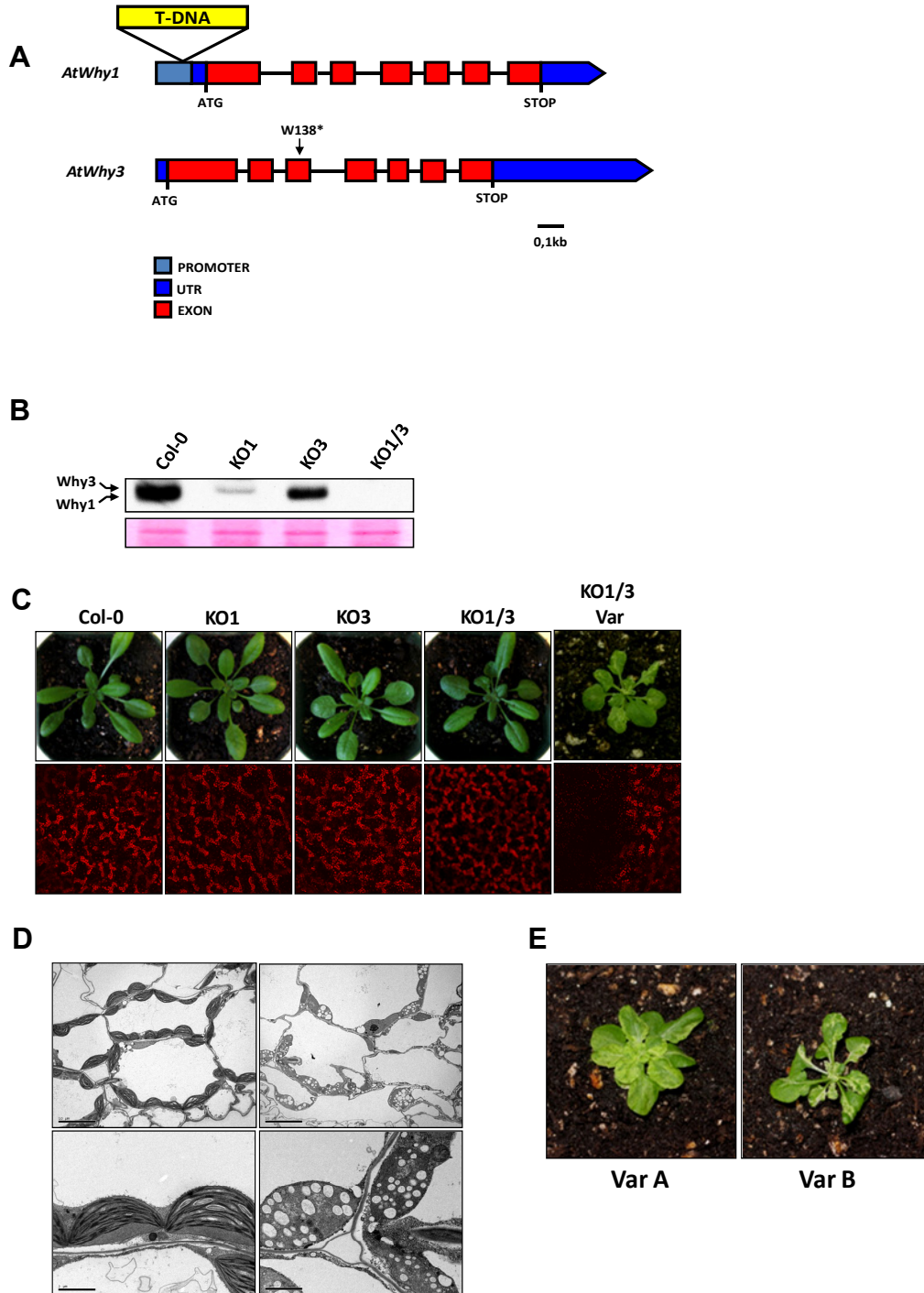


Figure 16. *AtWhy1* and *AtWhy3* are involved in the biogenesis of chloroplasts.

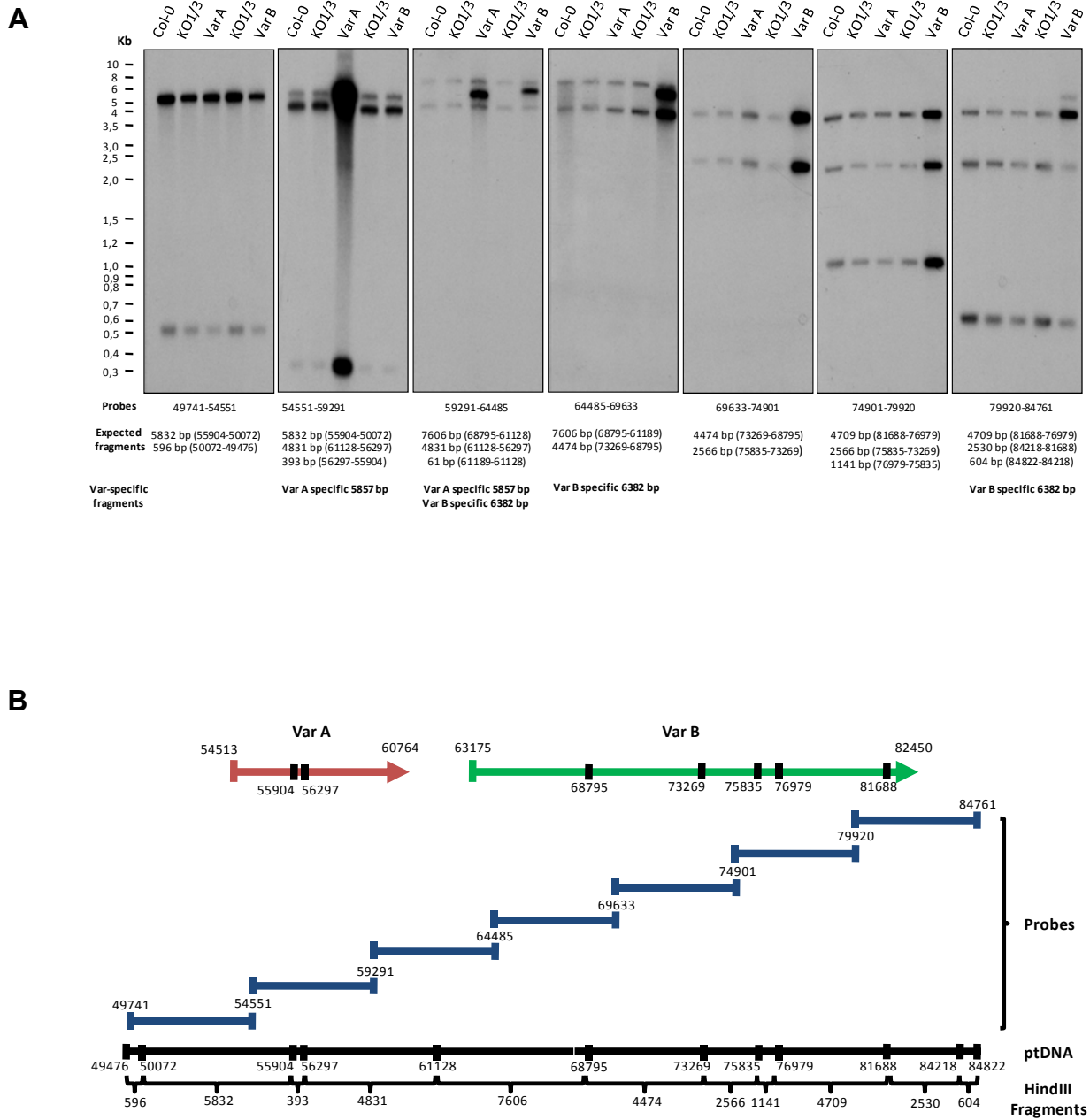


Figure 17. Variegated plants contain rearranged amplified plastid DNA regions.

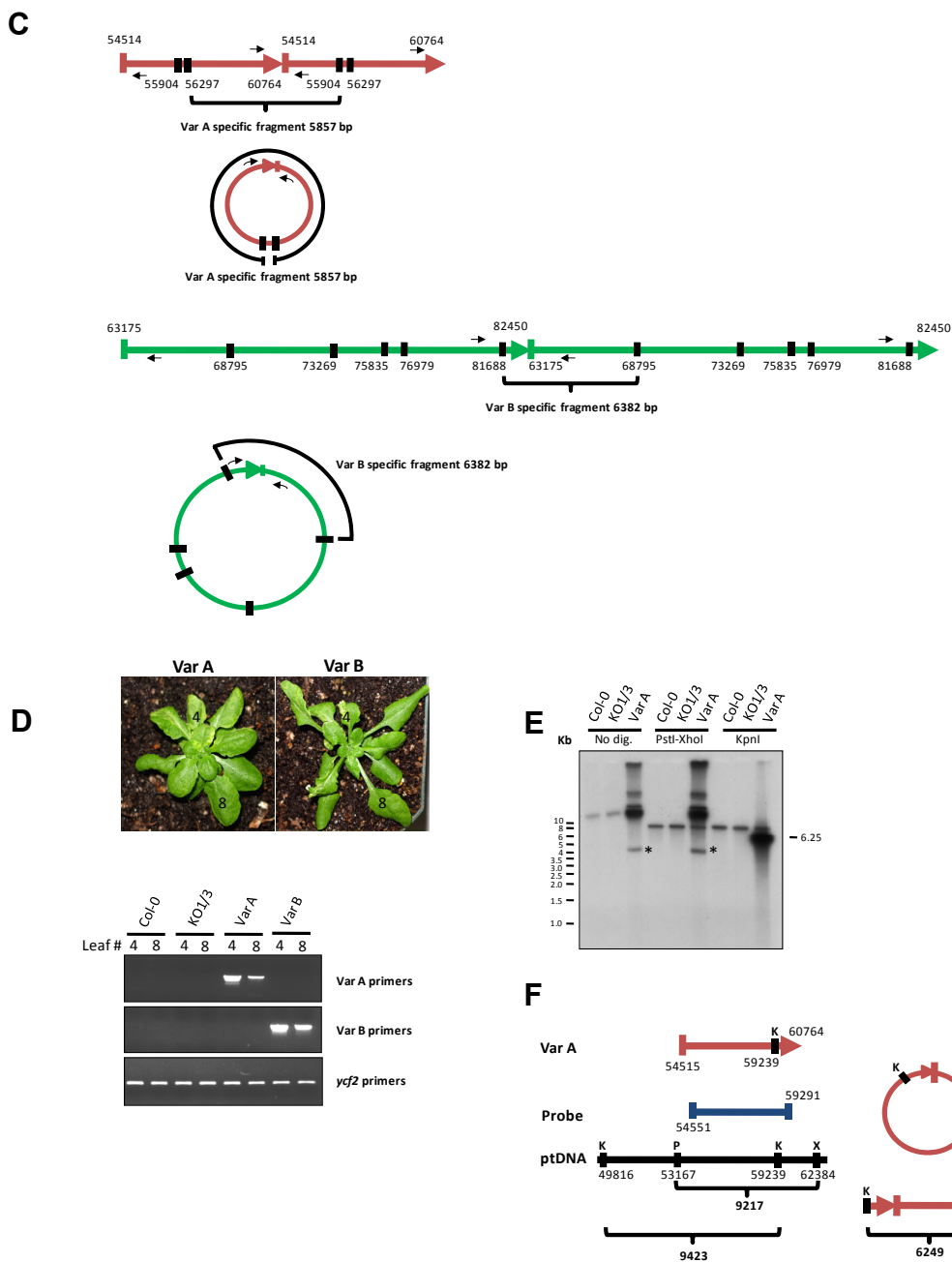


Figure 17 (suite). Variegated plants contain rearranged amplified plastid DNA regions.

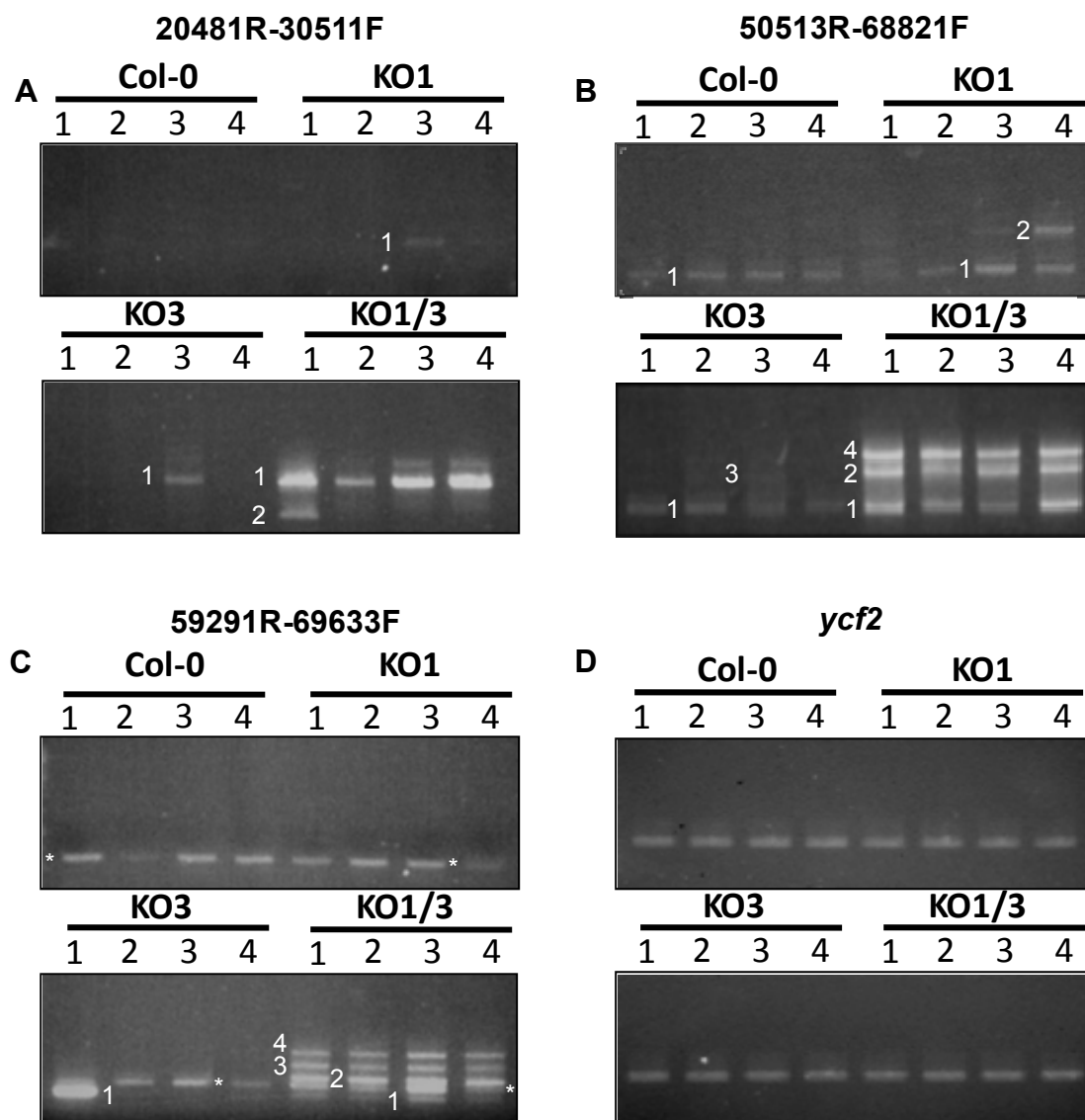


Figure 18. Illegitimate recombination is increased in the absence of *Arabidopsis* ptWhirlies.

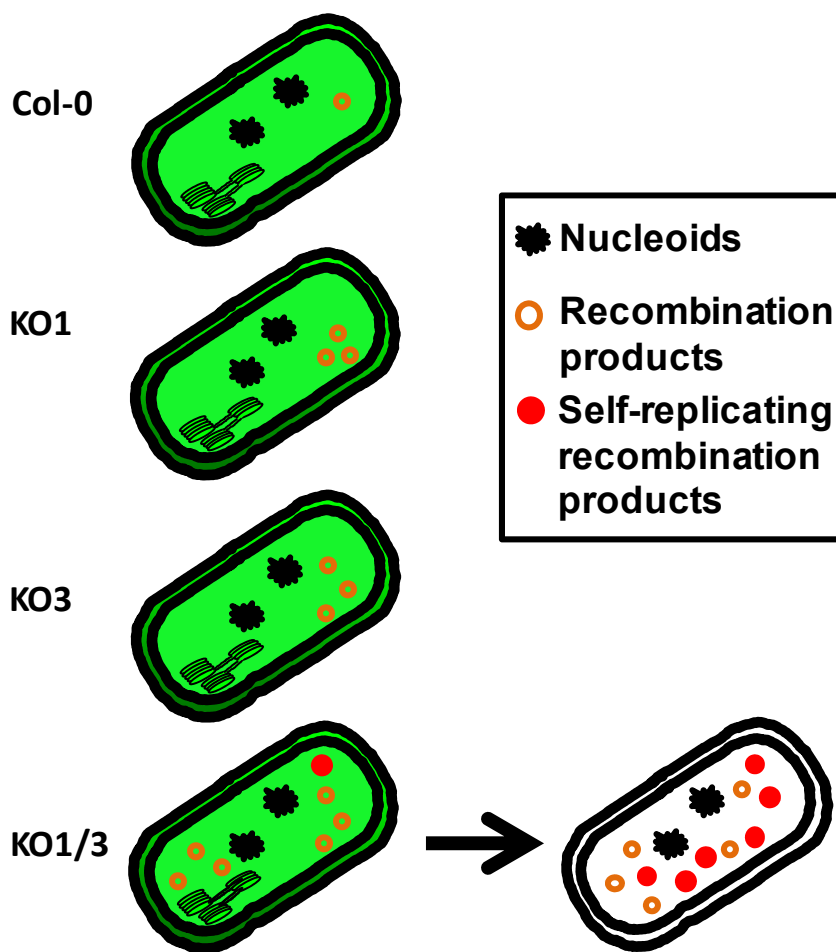


Figure 19. Model for the appearance of variegation in plants lacking Whirlies.

Tableau I. Percentage of plants showing a variegated phenotype.

Genotype	No. of Plants	No. of Variegated Plants	% Variegation
Col-0	2287	0	0
KO1	1367	0	0
KO3	2736	0	0
KO1/3	2879	133	4.6

Tableau II. Characteristics of amplified regions in variegated plants.

Plant line	Length of amplified region	Direct repeats (DR) positions ¹	Short direct repeat sequences and junction	DR length and (mismatches)
Var A	6250 bp	54514 recombinant 60764	atctcattagCCTTTTTTTTTcgtatTTtca agcattcattCCTTTTTTTTTcgtatTTtca agcattcattCCTTTTTTTTTatTTtctatc	(0) 10 bp (0)
Var B	19275 bp	63165 recombinant 82450	TTTTgtagtTTCTTTTTTTTTTTTTtcaatTTtg tagttcTTtaTTCTTTTTTTTTTTTTtcaatTTtg tagttcTTtaTTCTTTTTTTTTTTTTcagttgctac	(0) 14 bp (0)
Var C	25940 bp	9992 recombinant 35932	gTtctagttgCTCTTGAATACCTTCTTTcaaaaagctt attctcttatCTCTTGAATACCTTCTTTcaaaaagctt attctcttatCTCTTGAATTCATT <u>CGTT</u> gcagatccaa	(0) 18 bp (3)
Var D	19216 bp	61921 recombinant 80507	tgaagcagttGTTAAAATTCCTTATgatatgcaac cccatgataaGTTAAAATTCCTTATgatatgcaac cccatgataaGTTAAAATTCCTTATtgTtctataa	(0) 15 bp (0)
Var E ²	46310 bp	124537 recombinant 6019	tcaataatgtAAAAGTATACTGTCTCTTGgtagactaa tatattacaaAAAAGTGTAG-GACTCTTGgtagactaa tatattacaaAAAAGTGTAG-GACTCTTGtaaaatagaa	(3) 18 bp (0)

¹ The positions of the direct repeats refer to the numbers of the nucleotides in the published *Arabidopsis* chloroplast genome sequence (42)

² Presence of sequence motifs similar to OriA and OriB replication origins of *Nicotiana tabacum* (68) in the amplified region.

Tableau III. Representative recombination events between short direct repeats in *Arabidopsis* plants lacking ptWhirlies. 130

Reaction	Band #	Plant lines	DR positions	Short direct repeat sequences	DR length and (mismatches)
A	1	KO1, KO3 KO1/3	20366 recombinant 31005	aatatatataCAAAAATTTTCTTTTTTTacacttctta gttaagttatCAAAAATTTGCTTTTTTTacacttctta gttaagttatCAAAAATTTGCTTTTTTTtttttttagaa	(1) 18 bp (0)
A	2*	KO1/3	30300 recombinant 30827	gtcatttatgCAAAAAAAAAAAAAAatggttatgta attggctctaaCATAAAAAAAAAAAAAAatggttatgta attggctctaaCATAAAAAAAAAAAAAAagaaaaaaag	(1) 14 bp (0)
B	1	Col-0 KO1,KO3, KO1/3	50427 recombinant 69424	aattccaatcGGAGTGGATTGGgcaaggata tctaaaaaatGGAGTTGATTGGgcaaggata tctaaaaaatGGAGTTGATTGGatttgcacca	(1) 12 bp (0)
B	2	KO1 KO1/3	49181 recombinant 68999	tttttatttcCCCCACACCTTTTTTatataaaatt ctctaaccttCCCCACACCTTTTTTatataaaatt ctctaaccttCCCCACCACTATTTTTtgctaggta	(0) 15 bp (3)
B	3	KO3	49972 recombinant 70402	gttttttttaCTTTTTTTTTTTTTAttattgtatc cccaaagtgtCTTTTTTTTTTTTTAttattgtatc cccaaagtgtCTTTTTTTTTTTTTAacgggtgtgaaa	(0) 15 bp (0)
B	4	KO1/3	48553 recombinant 68998	tagaattgtaTCCCCCCTTCATTTATTGCTttccgatcct actctaacctTCCCCCCTTCATTTATTGCTttccgatcct actctaacctTCCCCACCACTATTTTTGCTaggatatttc	(0) 21 bp (5)
C	1	KO3 KO1/3	59043 recombinant 69807	tttttgtagcTATTTTTTTATTctatttctat ttctcttcaaTATTTTTTTATTctatttctat ttctcttcaaTATTTTTTTATTtttatattga	(0) 12 bp (0)
C	2	KO1/3	59287 recombinant 70228	acttaccctcTATTTTTGTGCCTTtagtaggcct ccgtacaggcTTTTTTTTGTGCCTTtagtaggcct ccgtacaggcTTTTTTTTGTGCATTgcatacggct	(1) 14 bp (1)
C	3	KO1/3	59028 recombinant 70225	tatttagtttGGCTTTTTTTGTacctattttt taaccgtacaGGCTTTTTTTGTacctattttt taaccgtacaGGCTTTTTTTGTgcattgcata	(0) 12 bp (0)
C	4	KO1/3	58721 recombinant 70186	attagactagACAAACAAAAAAAAAgttcattttc agacgggataACAAACAAAAAAAAAgttcattttc agacgggataACAAACAAAAAAAAatagataaat	(0) 14 bp (1)

2*: Non-specific annealing of primer 20481REV at positions 30367-30346 yielded this product.

Tableau IV. Short direct repeat-mediated recombination frequency in various *Arabidopsis* genotypes.

Plant lines	PCR reactions	Total events
Col-0	30	2
KO1	30	7
KO3	30	6
KO1/3	30	40

3.10 Supporting Information

3.10.1 Plastid Whirlies are Associated with Nucleoids

It has been previously shown that Whirlies of *Arabidopsis* localize to either plastids or mitochondria (1). An AtWhy1-green fluorescent protein (GFP) fusion produces punctuate foci of fluorescence in the plastids of transformed potato protoplasts, a pattern that is reminiscent of nucleoids (condensed DNA-protein complexes containing the complete plastid genome). Yet, co-localization of AtWhy1-GFP with plastid DNA has so far not been demonstrated. Therefore, we produced transgenic tobacco plants expressing a GFP fusion protein of StWhy1 (*Solanum tuberosum* Whirly 1), a close homolog of AtWhy1 (68% amino acid identity). Localization of StWhy1-GFP was monitored by laser-scanning confocal microscopy (LSCM). Examination of StWhy1-GFP fluorescence in transgenic leaf guard cells indicated that StWhy1 was also present in speckles within chloroplasts (Fig. 21A). To determine if ptWhirlies are associated with nucleoids, we transiently transfected protoplasts produced from tobacco mesophyll leaf cells with StWhy1-GFP. As shown in the upper right panel of Fig. 21B, StWhy1-GFP appeared as punctuate foci of fluorescence distributed in the chloroplasts. The transformed protoplasts were subsequently stained with the DNA-dye Syto85 which allowed the visualisation of nucleoids (lower left panel of Fig. 21B). Co-localization of Syto85 fluorescence with that of StWhy1-GFP in chloroplasts of tobacco mesophyll protoplasts (lower right panel of Fig. 21B) confirmed that StWhy1 co-localizes with DNA in plastids.

3.10.2 Plastidial Whirlies are Associated with ptDNA in a Non-Sequence Specific Manner

To characterize the DNA regions that are bound by AtWhy1 and AtWhy3, we performed DNA immunoprecipitation experiments followed by PCR detection of precise regions distributed all over the plastid genome : inside genes, around the determined transcription start sites of genes, in isolated regions that are devoid of any coding sequence and inside

putative replication origins (Fig. 22A). No differences could be observed for the immunoprecipitation of all types of regions between Col-0, KO1 and KO3 extracts, whereas no ptDNA could be immunoprecipitated using the KO1/3 extracts (Fig. 22B). To control for the specificity of interaction, we tested that the sonication regimen to which ptDNA was subjected yielded fragments of an average length of 500-1000 base pairs (Fig. 22C). This indicated that for any given region immunoprecipitated specifically, it can be assumed that there is at least one Whirly binding-site within a maximum of 1200 bp relative to the center of the amplified region. Additionally, we used primers to amplify a mitochondrial DNA region in the *atp9* gene as a negative control since AtWhy1 and AtWhy3 exclusively localize to chloroplasts and thus should not interact with the mitochondrial genome. We found that we could only amplify similar background levels of mitochondrial DNA for all genotypes, confirming the lack of specific interaction between AtWhy1/3 and mitochondrial DNA *in vivo* (Fig. 22D). These results indicate that ptWhirlies are able to bind many if not all ptDNA regions *in vivo*.

3.10.3 Supporting Information Materials and Methods

3.10.3.1 *Arabidopsis* mutant characterization

The SALK Institute Genomic Analysis Laboratory provided the sequence-indexed T-DNA insertion line SALK_099937 (2). The insertion inside the *AtWhy1* gene was mapped at nucleotide -102 relative to the initial ATG. The Seattle TILLING Project (3) provided plants with mutations in the *AtWhy3* gene. One line isolated and identified as il68D8 changed the TGG codon 138 to a TGA stop codon. The heterozygous M₂ plants were backcrossed with Col-0 eight consecutive times to eliminate background mutations. The homozygous progeny obtained was used for experiments. These *AtWhy3* mutant plants (KO3) were crossed with *AtWhy1* mutant plants (KO1) and the progeny was genotyped to isolate double mutants. The insertion mutant was genotyped using PCR primers specific for the wild-type or mutant allele. The KO3 point mutation was detected using PCR

amplification with dCAPS primer W138for (CAGCTGGTGTTCGTCAATATGCCTG) coupled with primer W138rev (GCCCAAGGCTAACTAGATTACCGAT) and digestion of the product with the MvaI restriction enzyme (Fermentas).

3.10.3.2 Microscopy

Chlorophyll autofluorescence in *Arabidopsis*

Mature leaf disks were taken from 4 week old plants and mounted on glass slides. A confocal microscope OLYMPUS IX71 was used with a 488 nm laser. Chlorophyll autofluorescence was visualised through a 660 nm long pass filter.

Transmission electron microscopy

Leaf mesophyll tissue from 4 week old plants was cut into 1-2 mm² pieces which were washed twice with 0.2 M sodium cacodylate buffer pH 7.4 for 10 minutes. Tissue was fixed by incubation for 6 hours in cacodylate buffer containing 1 % EM-grade glutaraldehyde. Samples were coated with epoxy resin, cut into 80 nm sections and mounted onto nickel/formvar grids. Staining of the samples was done using uranyl acetate and lead citrate. Observations were done using a JEOL (JEM1230) transmission electron microscope at 80 kV. Photographs were taken using a Gatan DualVision camera.

Intracellular localization of StWhy1

The StWhy1-GFP construct was produced by inserting the Emerald GFP (Clontech) sequence downstream of *StWhy1* into pBluescript (Stratagene). The StWhy1-GFP fusion comprising all 274 amino acids of StWhy1 was then inserted into the pBin19 vector which contains 2 copies of the CaMV35S promoter. All constructs were electroporated into *Agrobacterium tumefaciens* strain LBA4404 and transgenic plants were obtained as described (4). For transient expression, the StWhy1-GFP fusion in pBin19 was transferred to the pBI223 vector which also contains the CaMV35S promoter. Leaf mesophyll protoplasts were isolated and transformed as described. GFP fluorescence (500-530 nm) was visualised with a Leica DM IRB/E laser-scanning confocal microscope using a 488 nm

laser excitation source. Chlorophyll was excited using a 543 nm laser and its autofluorescence was visualised at 575-630 nm. For DNA staining, protoplasts were incubated with 5 μ M Syto85 (Molecular Probes) for 10-30 min and then washed with fresh culture medium. Syto85 fluorescence (570-600 nm) was visualised using a 568 nm laser excitation source. GFP and Syto85 images were collected sequentially, and no fluorescence cross-talk was observed under our conditions (data not shown). Pseudocoloring of the images, maximal projections, and image overlays were done using the Leica confocal software. Hand-made thin sections of fresh leaves from 4 weeks-old plants were examined in sterile water. Confocal microscopy was carried out using an Olympus FV300 microscope.

3.10.3.3 Plastid DNA immunoprecipitation

Leaf tissue was fixed in 1 % formaldehyde for 15 minutes under vacuum. Glycine was added to 0.125 M to titrate the remaining formaldehyde and fixed tissue was washed 3 times with distilled water. Crude plastids were prepared by grinding in chloroplast extraction buffer (0.33M sucrose, 10 mM MES-NaOH pH 4.6, 10 mM NaCl, 1 mM EDTA, 10 mM DTT) using a mortar and pestle. The ground tissue was filtered through 2 layers of Miracloth (Calbiochem) and spun at 1000 g for 5 minutes in a microcentrifuge to pellet most plastids. The pellet was resuspended in IPP buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 % NP-40, 0.5 % sodium deoxycholate and 1 tablet of Complete Protease Inhibitor (Roche)), lysis was carried out on ice for 2 minutes and the lysate was spun at 11000 g for 15 minutes. Supernatant was sonicated using a Branson microtip sonicator. Sonicated material was precleared by incubating with protein-A-agarose beads (Roche) for 1 hour at 4°C on a rotating platform. The precleared material was spun at 11000 g for 15 minutes and the supernatant was transferred to a fresh tube. Proteins were quantified and an 100 μ g aliquot was kept as input material while 500 μ g was used for immunoprecipitation using 1/200 rabbit polyclonal anti-AtWhy1 sera in 1 mL final volume. Immunoprecipitations were carried out for 1 hour at 4°C before adding 50 μ l of protein-A-agarose beads. Complexes were allowed to form overnight. Beads were pelleted at 100 g for 1 minute and

were washed 2 times 5 minutes with IPP buffer, 2 times 5 minutes with low salt buffer (50 mM Tris-HCl pH 7.5, 0.1 % NP-40, 0.05% sodium deoxycholate), 2 times 5 minutes with high salt buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.1 % NP-40, 0.05% sodium deoxycholate), 2 times 5 minutes with LiCl buffer (50 mM Tris-HCl pH 7.5, 0.1 % NP-40, 0.05 % sodium deoxycholate, 250 mM LiCl) and 2 times 5 minutes with TE pH 7.5 (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Beads were pelleted at 11000 g for 20 seconds and resuspended in TE buffer containing 1 % SDS. Complexes were eluted from the beads by heating at 65°C for 15 minutes. Beads were spun down at 11000 g for 5 minutes and the supernatant transferred to a new tube. One tenth volume of 2 M NaCl was added and the eluted complexes were incubated at 65°C overnight for de-crosslinking. Glycogen was added to a final concentration of 0.5 µg/µl and two volumes of 100 % ethanol was added. DNA was precipitated at -80°C for 1 hour. Tubes were spun at 11000 g for 20 minutes and DNA pellets were resuspended in TE buffer. DNA was subjected to a phenol/chloroform extraction and to a chloroform extraction before being reprecipitated as above. After a final wash with 70% ethanol, DNA was resuspended in water. This DNA was used for PCR reactions. To further control the specificity of the experiment, we determined the size of DNA fragments following sonication and found that they were between 500 and 1200 bp in length. For the determination of the specificity of the interaction between AtWhy1 and AtWhy3 and plastid DNA, mitochondrial DNA immunoprecipitation was performed as described previously (5).

3.10.3.4 Electrophoretic mobility shift assays

Five hundred mg of leaves taken from 4 week old plants were ground in 5 mL of ice-cold chloroplast isolation buffer (0.33 M sucrose, 10 mM MES-NaOH pH 4.6, 10 mM NaCl, 1 mM EDTA, 10 mM DTT). The resulting slurry was filtered through 2 layers of Miracloth (Calbiochem). The suspension was then centrifuged at 3000 rpm in a microcentrifuge for 5 minutes. The crude plastids were resuspended in IPP buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 % NP-40, 0.5 % sodium deoxycholate, 1X Complete protease inhibitor cocktail (Roche)) and incubated 2 minutes on ice. The lysed plastids were centrifuged at

maximal speed in a microcentrifuge for 15 min and the supernatant was taken. Twenty μ g of the crude plastid extracts were used for ssDNA binding assays. The same extracts were used in the protein gel blot presented Figure 1B thereby providing a loading control. For electrophoretic mobility shift assays, crude plastid proteins were incubated in IPP buffer with 30000 counts per minute of end-labelled modified 4XPB (6) oligonucleotide (5'-TGTCATTTTTGTCATTTTTGTCATTTTTGTCA-3') for 30 minutes at room temperature. The anti-AtWhy1/3 antibody or preimmune serum were added to the Col-0 extracts before adding the oligonucleotide. The reactions were then separated on a 5.4% polyacrylamide Tris-Borate-EDTA gel as described (6).

3.10.3.5 Isolation of RNA

RNA was isolated from 4 week old plants using TRIZOL® reagent (Invitrogen) according to manufacturer's instructions. Running of the samples was performed as described (7). For the Var A and Var B samples, only highly variegated tissue was used.

3.10.3.6 Detection of rearranged DNA by PCR in maize

PCR reactions were conducted using the Ex Taq polymerase from Takara Bio Inc. according to the manufacturer's instructions. The search for rearranged products was performed using a series of outward-facing oligonucleotides spaced by approximately 5-50 kb. A total of 19 PCR reactions were performed on each of two independent DNA samples from B73 inbred line (WT) and each of the *ZmWhy1* mutant lines and analysed by gel electrophoresis. The DNA samples for all the plants were pre-adjusted to the same amount of ptDNA using semi-quantitative amplification of the maize *rpoB* gene sequence before performing the outward-facing PCR. All visible DNA bands were isolated, cloned and sequenced.

3.10.3.7 Oligonucleotides used in this study

The following oligonucleotides were used to identify rearranged products in *Arabidopsis*:
 5291REV, TCC CAA CGA GCC GTT TAT CGA ; 9121REV, AAG ATA GAT CGG GCC AAC TCG ; 9640FOR, GGT TCA AAT CCT ATT GGA CGC; 19181FOR, CAG CAT

CTG ATG TGC GTA CAG ; 19309FOR, TCG TCT GCT AAG ACA CGA CCA ; 20481,
 FOR GCG TTG ATC TAA TTG CCA CCG ; 20481REV, CGG TGG CAA TTA GAT
 CAA CGC ; 30511FOR, GCT GCC TCC TTG AAA GAG AGA ; 30511REV, TCT CTC
 TTT CAA GGA GGC AGC ; 31138REV, AAT AGC CTA GCG CAT CGA ACG ;
 40130FOR, TGA CAT GTT CCT CCT CTT CCC ; 40130REV, GGG AAG AGG AGG
 AAC ATG TCA ; 48481FOR, AGG GCA AAA ATA CCC CAA GGT ; 50251FOR, GGC
 CCC TTC CTA ATT GGA GAT ; 50513FOR, GGT TGT TGT GGT AGA CGT CTC ;
 50513REV, GAG ACG TCT ACC ACA ACA ACC ; 54551REV, TGG GAC GCA TAA
 CCG GAT ATG ; 58730FOR, CCT CCC TTG CTT GCA TAT GTA; 59291FOR, AGG
 CCT AGT CTT TCC GGC AAT ; 59291REV, ATT GCC GGA AAG ACT AGG CCT ;
 60182REV, ATA TAA AGG ACA CGC CGG GCA ; 65482REV, GCA TCT GCA CTA
 TCT CAG GAG ; 68821FOR, CTC CGA GCT ATA TAT CCG CGT; 68821REV, ACG
 CGG ATA TAT AGC TCG GAG; 69633FOR, ACA CCC GAG TAC ATG TTC CTC;
 69633REV, GAG GAA CAT GTA CTC GGG TGT; 69941REV, AGC AAC AGA AGC
 CCA AGC TCA; 71221REV, TTC TTG TTC CTG AAC GGG CTT ; 79920FOR, CCA
 TGC CGT TCG GAA GTG ATT ; 79920REV, AAT CAC TTC CGA ACG GCA TGG ;
 84761FOR, TAT GTA TGG CCG TGC CTA AGG ; 84761REV, CCT TAG GCA CGG
 CCA TAC ATA ; 95260FOR, AAC CTG CTA GTG GAG GAA GAC; 95260REV, GTC
 TTC CTC CAC TAG CAG GTT ; 104831FOR, CTG AAT CCA TGG GCA GGC AAG ;
 105983REV, GCG TTA CTC AAG CCG ACA TTC ; 111330FOR, TGC GGT TAA TCC
 CGC TGT TGA ; 111330REV, TCA ACA GCG GGA TTA ACC GCA ; 115987FOR,
 AAA CCA CTC ATG CCC GGC AAT ; 116007FOR, AAA CCA CTC ATG CCC GGC
 AAT ; 116007REV, ATT GCC GGG CAT GAG TGG TTT ; 121141FOR, AAC GTC ATC
 CGG TTA CCG GTT.

The following oligonucleotides were used to identify rearranged products in maize:

Zm33FOR, GGG TAT TTC CGG CTT TCC TTC ; Zm33REV, GAA GGA AAG CCG
 GAA ATA CCC ; Zm4761FOR, TCT CCT CAT ACG GCT CGA GAA ; Zm4761REV,
 TTC TCG AGC CGT ATG AGG AGA ; Zm10113REV, TTC CAC GTG GTA GAA CCT
 CCT ; Zm15001REV, TAC CCC TTG GAA ATG GGG TAC ; Zm20058FOR, CCA ACT

TAG GGG GGT ACG AAT ; Zm20058REV, ATT CGT ACC CCC CTA AGT TGG ;
 Zm25060FOR, GCA TGT CCA GTG ACT CAT GTG ; Zm25060REV, CAC ATG AGT
 CAC TGG ACA TGC ; Zm29971REV, AAC AAG AGA AGC ACG AGC TCC,
 Zm35230FOR, GAA AGG GAG TGT GTG CGA GTT ; Zm40020FOR, CCC AAG GTA
 TGG AAT CCT AGG ; Zm40020REV, CCT AGG ATT CCA TAC CTT GGG ;
 Zm45152FOR, TTT CGT TCT AGC GCC CGG AAA; Zm45152REV, TTT CCG GGC
 GCT AGA ACG AAA ; Zm50141FOR, AGA GGT GAG ATT AGG TGC CCT ;
 Zm50141REV, AGG GCA CCT AAT CTC ACC TCT ; Zm54971FOR, TGC CAA CAA
 TCC GAG GCT GTA ; Zm54971REV, TAC AGC CTC GGA TTG TTG GCA;
 Zm59880FOR, CGG GCC TAT TCA TTA GCT CCT ; Zm59880REV, AGG AGC TAA
 TGA ATA GGC CCG ; Zm64883FOR, GGG GTC GTG GAA TGC TTT TCT ;
 Zm64883REV, AGA AAA GCA TTC CAC GAC CCC ; Zm69981FOR, ATT AGC CAA
 CCG CCA AGG GAA ; Zm69981REV, TTC CCT TGG CGG TTG GCT AAT ;
 Zm74851FOR, GCG TAA GTC CCC TAT CTA GGA ; Zm85072FOR, ACC TCT CCG
 GAT CCT CGA ATT ; Zm89967FOR, GAA AAG CAA CGA CTG GAG TGG ;
 Zm89967REV, CCA CTC CAG TCG TTG CTT TTC ; Zm100051REV, TCG ACG ATG
 AAG CTT ATC CCC ; Zm105133FOR, ATT CGC CCT CCC CCT ACA TAT ;
 Zm105133REV, ATA TGT AGG GGG AGG GCG AAT ; Zm110101FOR, CTG GAA
 TAT AGA GCG GAC TGG ; Zm129981FOR, GAG CAG GCT ACC ATG AGA CAA ;
 Zm129981REV, TTG TCT CAT GGT AGC CTG CTC ; Zm135011FOR, TCA GTG ACT
 TTG GCA CTG GAC.

The following oligonucleotides were used to generate probes for DNA gel blots in *Arabidopsis*:

421FOR, AGC AGC TAG GTC TAG AGG GAA ; 5291REV, TCC CAA CGA GCC GTT
 TAT CGA ; 10230FOR, GCA ATC GTT GAC CTC TTG CCA ; 15461REV, CAA TAC
 GGG CTC GGT GTC ATT ; 30511FOR, GCT GCC TCC TTG AAA GAG AGA ;
 35850REV, ACC AAC CAT CAG GAG ACG CAA ; 40130FOR, TGA CAT GTT CCT
 CCT CTT CCC ; 44916REV, GTA CGG AAA GAG AGG GAT TCG ; 49741FOR, CCT
 TAC GTA AAG GCC ACC CTA ; 54551REV, TGG GAC GCA TAA CCG GAT ATG ;

54551FOR, CAT ATC CGG TTA TGC GTC CCA ; 59291REV, ATT GCC GGA AAG ACT AGG CCT ; 59291FOR, AGG CCT AGT CTT TCC GGC AAT ; 64485REV, CGT TGC TGT GTC AGA AGA AGG ; 69633FOR, ACA CCC GAG TAC ATG TTC CTC ; 74901FOR, CAG AGA ATG GGG TCT GTC ATC ; 79920REV, AAT CAC TTC CGA ACG GCA TGG ; 79920FOR, CCA TGC CGT TCG GAA GTG ATT ; 84761REV, CCT TAG GCA CGG CCA TAC ATA ; 84761FOR, TAT GTA TGG CCG TGC CTA AG ; 90150REV, TCC ATC TGC CCT TTG TCA ACG ; 100770FOR, GAA GTC ATC AGT TCG AGC CTG ; 105983REV, GCG TTA CTC AAG CCG ACA TTC ; 105983FOR, GAA TGT CGG CTT GAG TAA CGC ; 111330REV, TCA ACA GCG GGA TTA ACC GCA ; 111330FOR, TGC GGT TAA TCC CGC TGT TGA ; 116007REV, ATT GCC GGG CAT GAG TGG TTT.

The following oligonucleotides were used to determine the ends of amplified regions in *Arabidopsis* variegated lines:

56061REV, AGA ACA CCA GGT AGT GAG ACC ; 60211FOR, GAG GTC AAG GGG CTA TTC CTT ; 63990REV, CAT GGA CTA GCT GTA CCT ACC ; 81360FOR, CGC AGG TCT TAC TGT AAC TGG ; 10230REV, TGG CAA GAG GTC AAC GAT TGC ; 30511FOR, GCT GCC TCC TTG AAA GAG AGA ; 62161REV, TAG CAG GGT CTG GAG CAA GAA ; 75961FOR, TTA CTG GTG TGG TTC TGG GTG ; 120601FOR, GCG CTT CGG AAT TCA TCT CAC ; 4341FOR, CCG AGT ACT CTA CCG TTG AGT.

The following oligonucleotides were used for adjusting DNA levels in the PCR experiments:

YCF2FOR, GAT CTC TGA GAG CTG TTT CCG ; YCF2REV, TGT TTC GCC TCT TAC TCG GAG ; ZmRpoBFOR, GCT TGG CGG AAG AAC TTG AGA ; ZmRpoBREV, TGT TTC GCC TCT TAC TCG GAG.

The following oligonucleotides were used to generate probes for DNA gel blots in maize:

Zm4761FOR, TCT CCT CAT ACG GCT CGA GAA ; Zm10113REV, TTC CAC GTG GTA GAA CCT CCT ; Zm10113FOR, AGG AGG TTC TAC CAC GTG GAA ; Zm15001REV, TAC CCC TTG GAA ATG GGG TAC ; Zm15001FOR, GTA CCC CAT TTC CAA GGG GTA ; Zm20058REV, ATT CGT ACC CCC CTA AGT TGG ;

Zm25060FOR, GCA TGT CCA GTG ACT CAT GTG ; Zm29971REV, AAC AAG AGA AGC ACG AGC TCC ; Zm29971FOR, GGA GCT CGT GCT TCT CTT GTT ; Zm35230REV, AAC TCG CAC ACA CTC CCT TTC ; Zm35230FOR, GAA AGG GAG TGT GTG CGA GTT ; Zm40020REV, CCT AGG ATT CCA TAC CTT GGG ; Zm40020FOR, CCC AAG GTA TGG AAT CCT AGG ; Zm45152REV, TTT CCG GGC GCT AGA ACG AAA ; Zm45152FOR, TTT CGT TCT AGC GCC CGG AAA ; Zm50141REV, AGG GCA CCT AAT CTC ACC TCT ; Zm50141FOR, AGA GGT GAG ATT AGG TGC CCT ; Zm54971REV, TAC AGC CTC GGA TTG TTG GCA ; Zm59880FOR, CGG GCC TAT TCA TTA GCT CCT ; Zm64883REV, AGA AAA GCA TTC CAC GAC CCC ; Zm69981FOR, ATT AGC CAA CCG CCA AGG GAA ; Zm74851REV, TCC TAG ATA GGG GAC TTA CGC ; Zm74851FOR, GCG TAA GTC CCC TAT CTA GGA ; Zm79860REV, TGA AGG GAA AAT CCT GCC GAG ; Zm85072FOR, ACC TCT CCG GAT CCT CGA ATT ; Zm89967REV, CCA CTC CAG TCG TTG CTT TTC ; Zm95042FOR, GTG GCG AAC TCC AGG CTA ATA ; Zm100051REV, TCG ACG ATG AAG CTT ATC CCC ; Zm100051FOR, GGG GAT AAG CTT CAT CGT CGA ; Zm105133REV, ATA TGT AGG GGG AGG GCG AAT ; Zm105133FOR, ATT CGC CCT CCC CCT ACA TAT ; Zm110101REV, CCA GTC CGC TCT ATA TTC CAG ; ZmMt255542FOR, GCA GTA TTG GAA CCA TCC TCG ; ZmMt259720REV, CCC ATG ATG TGG TAA AGG GAG.

3.10.4 References

1. Krause K, *et al.* (2005) DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS Lett* 579(17):3707-3712 .
2. Alonso JM, *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301(5633):653-657 .
3. Till BJ, *et al.* (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* 13(3):524-530 .
4. Matton DP, Prescott G, Bertrand C, Camirand A, & Brisson N (1993) Identification of cis-acting elements involved in the regulation of the pathogenesis-related gene STH-2 in potato. *Plant Mol Biol* 22(2):279-291 .
5. Maréchal A, *et al.* (2008) Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol* 8:42 .

6. Desveaux D, *et al.* (2004) A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*. *Dev Cell* 6(2):229-240 .
7. Sambrook J, Fritsch EF, & Maniatis T (1989) *Molecular cloning : a laboratory manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) 2nd Ed.
8. Sato S, Nakamura Y, Kaneko T, Asamizu E, & Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res* 6(5):283-290 .
9. Lohse M, Drechsel O, & Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 52(5-6):267-274 .
10. Maier RM, Neckermann K, Igloi GL, & Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* 251(5):614-628 .
11. Prikryl J, Watkins KP, Friso G, van Wijk KJ, & Barkan A (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res* 36(16):5152-5165 .
12. Clifton SW, *et al.* (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol* 136(3):3486-3503 .

3.10.5 Supporting Information Figure Legends

Fig. 20. ssDNA-binding activity of AtWhy1/3 in *Arabidopsis* ptWhirlies mutants. Electrophoretic mobility shift assay was performed using 20 µg of crude plastid proteins isolated from plants of the indicated genotypes and a radiolabeled probe of 32 nucleotides (5'-TGTCATTTTTGTCATTTTTGTCATTTTTGTC-3'). As a control, an anti-AtWhy1/3 antibody (α -Why1/3) or preimmune serum (PI) was preincubated with Col-0 extracts before adding the probe. Addition of the antibody eliminated the signal in the Col-0 extract, confirming that this signal corresponds to a complex between Whirlies and ssDNA.

Fig. 21. Co-localization of StWhy1 with chloroplast DNA. (A) LSCM of StWhy1-GFP tobacco leaf guard cells. Upper left panel is chlorophyll autofluorescence pseudo-coloured in red. Upper right panel is GFP fluorescence pseudo-coloured in green. Bottom left panel is overlay of chlorophyll and GFP fluorescence. Bottom right panel is corresponding phase-contrast image. (B) LSCM of a tobacco leaf mesophyll protoplast transiently expressing StWhy1-GFP and stained with the DNA dye Syto85. Upper left panel is chlorophyll autofluorescence pseudo-coloured in red. Upper right panel is GFP fluorescence pseudo-coloured in green. Bottom left panel is Syto85 fluorescence pseudo-coloured in blue.

Bottom right panel is overlay of all three images. Maximum projections are shown. Scale bars = 8 μm .

Fig. 22. AtWhy1 and AtWhy3 interact with chloroplast DNA. (A) Position of the amplified regions on the plastid genome. (B) PCR amplification of ptDNA regions following immunoprecipitation on crude plastid extracts of the indicated genotypes. Inputs represent 20 % of the total DNA used in the immunoprecipitation. Representative regions are shown here indicating specific interaction with AtWhy1/3. (C) Total plastid DNA following sonication was purified and used as a template for PCR. Oligonucleotides designed to amplify regions of increasing length were used to verify the efficiency of the sonication regimen. (D) Immunoprecipitation on total crude organelles was performed and plastid or mitochondrial DNA was assessed by PCR using oligonucleotides designed to amplify part of the plastid *psbA* or mitochondrial *atp9* genes.

Fig. 23. Long (16 hrs) and short (4 hrs) exposure of panel 2 from Figure 2A. DNA gel blot (10 $\mu\text{g}/\text{lane}$) of total leaf DNA digested with HindIII and hybridized with the probes indicated below the gel. The probe numbers refer to the nucleotides of the published *Arabidopsis* chloroplast genome (8). Expected fragments from restriction analysis of Col-0 ptDNA and the size of new fragments observed in variegated lines are presented below the probes.

Fig. 24. A gene map of the ptDNA region containing the amplified rearranged regions found in Var A and Var B plants was drawn using the web-based tool OGDRAW (9).

Fig. 25. Absence of subgenomic amplicons in *ZmWhy1* maize mutants. (A-E) DNA gel blots (10 $\mu\text{g}/\text{lane}$) of total leaf DNA digested with HindIII and hybridized with the probes indicated below the gel. The probe numbers refer to the nucleotides of the published maize chloroplast genome (10). Expected fragments from restriction analysis of WT ptDNA are presented below the probes. B73 inbred line was used as the source for WT DNA. For each of the three *ZmWhy1* mutants, a segregating green plant issued from heterozygous seeds was used as a control (g). The pale green *ZmWhy1-2* (pg), the yellow (intermediate) *ZmWhy1-2/-1* (y) and the albino *ZmWhy1-1* (al) lines were previously described (11). The corresponding ethidium bromide staining of total digested DNA is presented below each

panel. (E) The ptDNA level in the albino mutant was slightly lower than in the other lines where no significant difference was observed compared to the WT. To verify this, a membrane was hybridized with plastid probe 95042-100051, partially stripped and re-blotted with mitochondrial probe 255542-259720 (12). The residual signal from the plastid genome is marked by an asterisk. This confirmed that the level of ptDNA was specifically diminished in the albino *ZmWhy1-1* mutant.

Fig. 26. Levels of plastid 23S and 16S rRNA in *Arabidopsis* ptWhirlies mutants. Total RNA was extracted from 100 mg of mature leaf tissue from the indicated plants. Ten μ g of purified RNA was migrated on a 1.2 % agarose gel and stained with ethidium bromide.

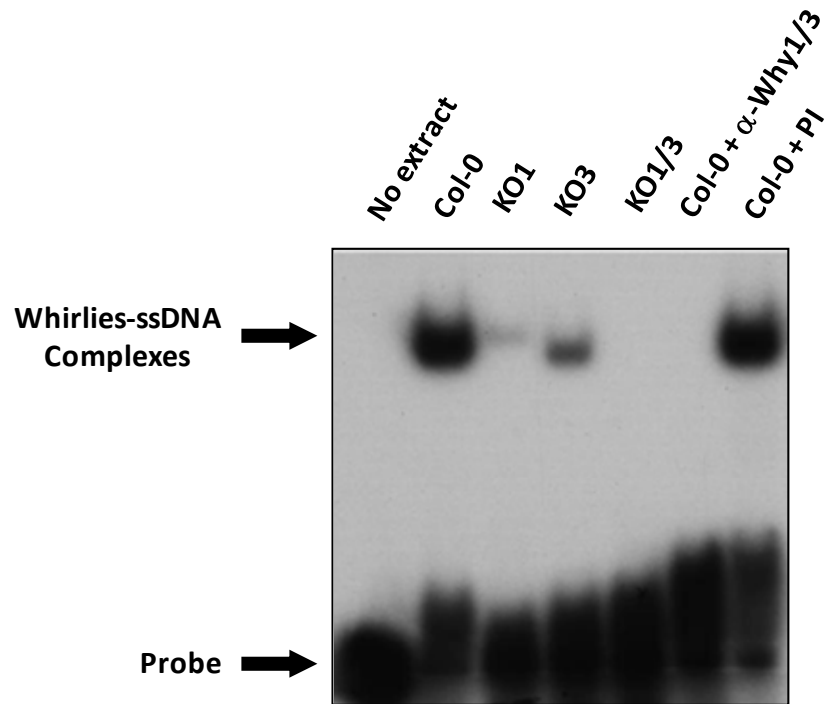
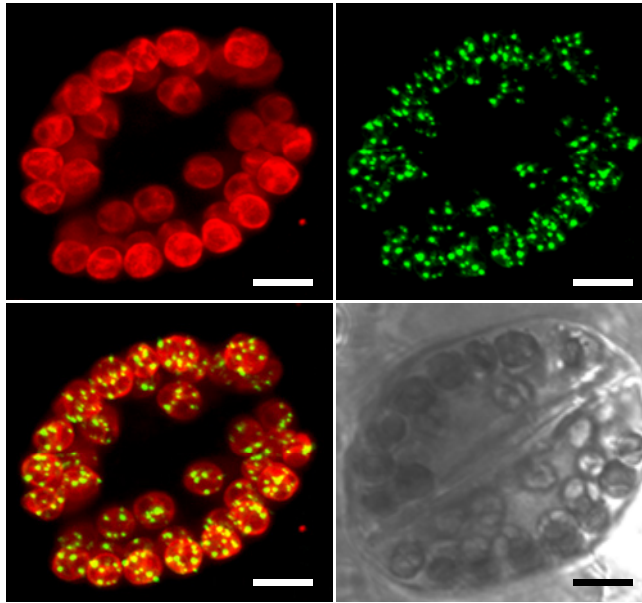


Figure 20. ssDNA-binding activity of AtWhy1/3 in *Arabidopsis* ptWhirlies mutants.

A

146



B

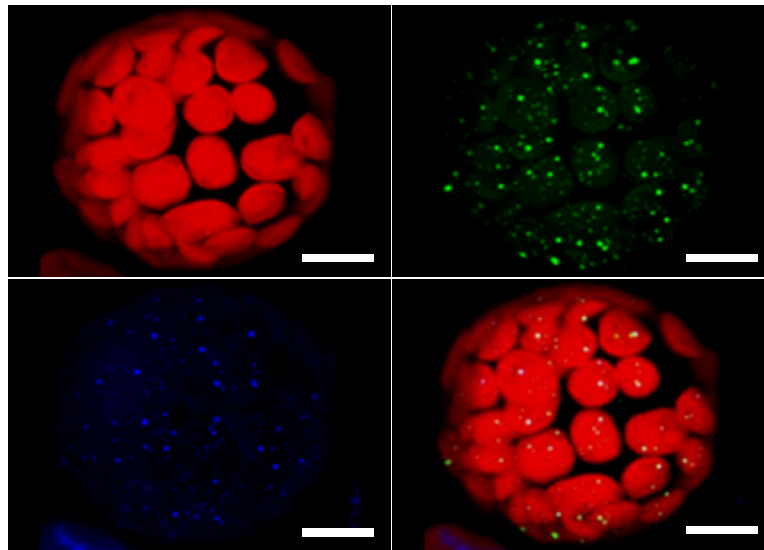


Figure 21. Co-localization of StWhy1 with chloroplast DNA.

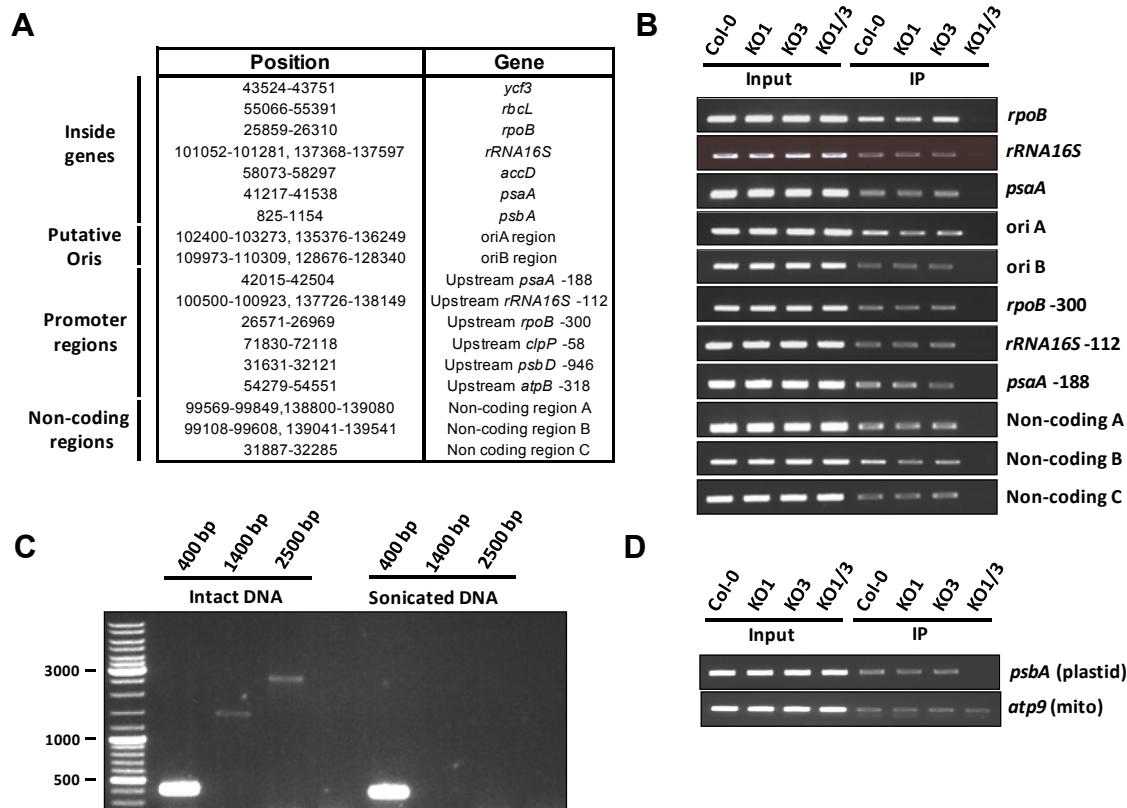


Figure 22. AtWhy1 and AtWhy3 interact with chloroplast DNA.

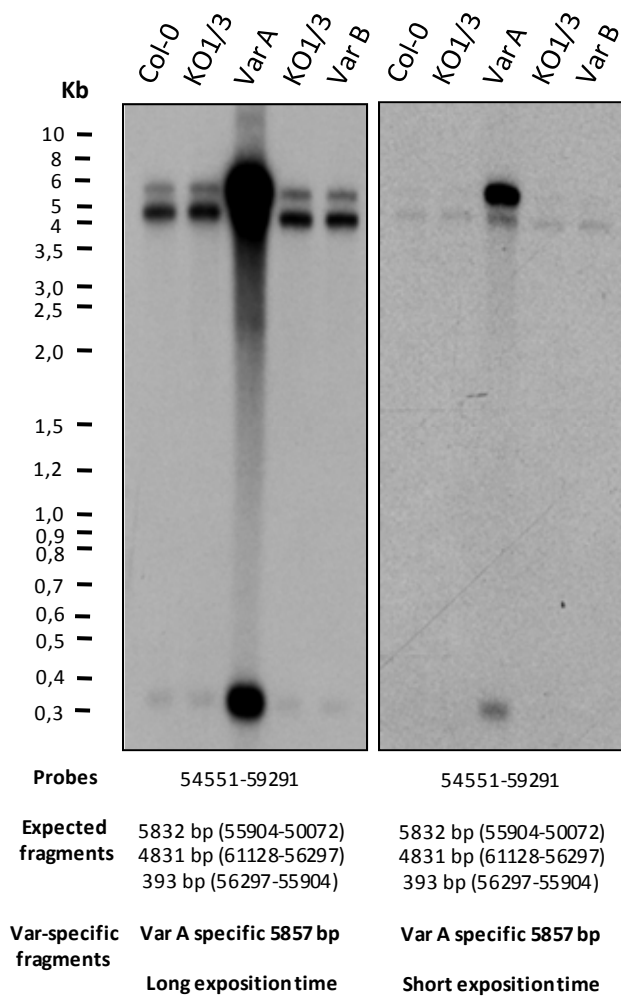


Figure 23. Long (16 hrs) and short (4 hrs) exposure of panel 2 from Figure 2A.

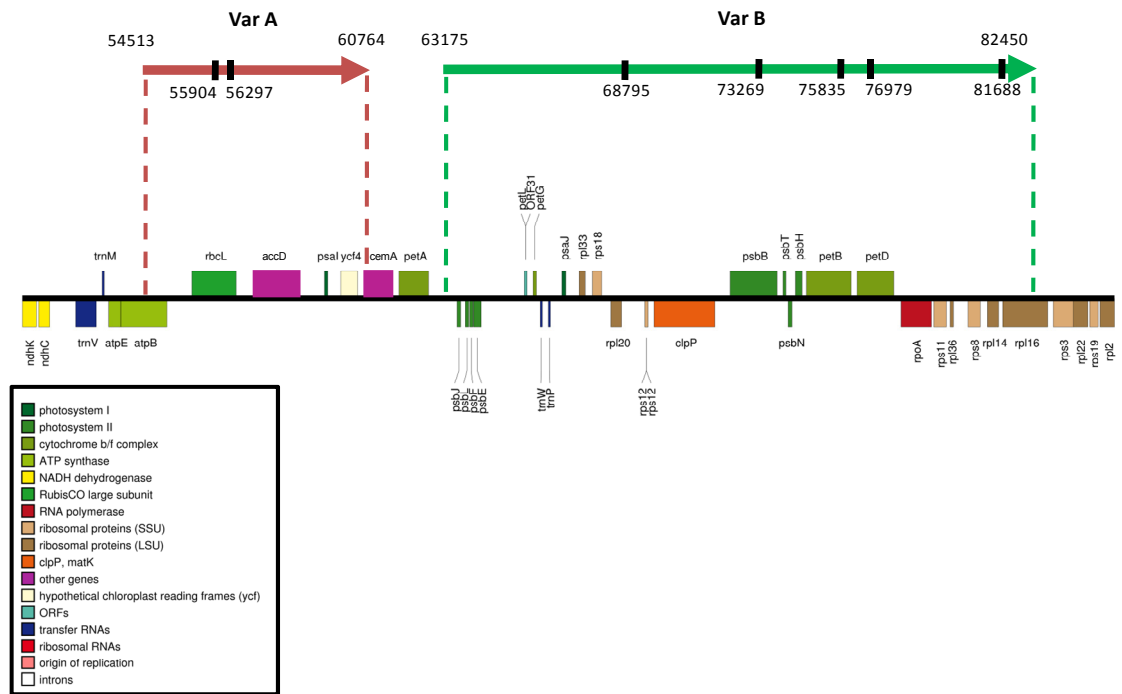


Figure 24. A gene map of the ptDNA region containing the amplified rearranged regions found in Var A and Var B plants was drawn using the web-based tool OGDRAW (9).

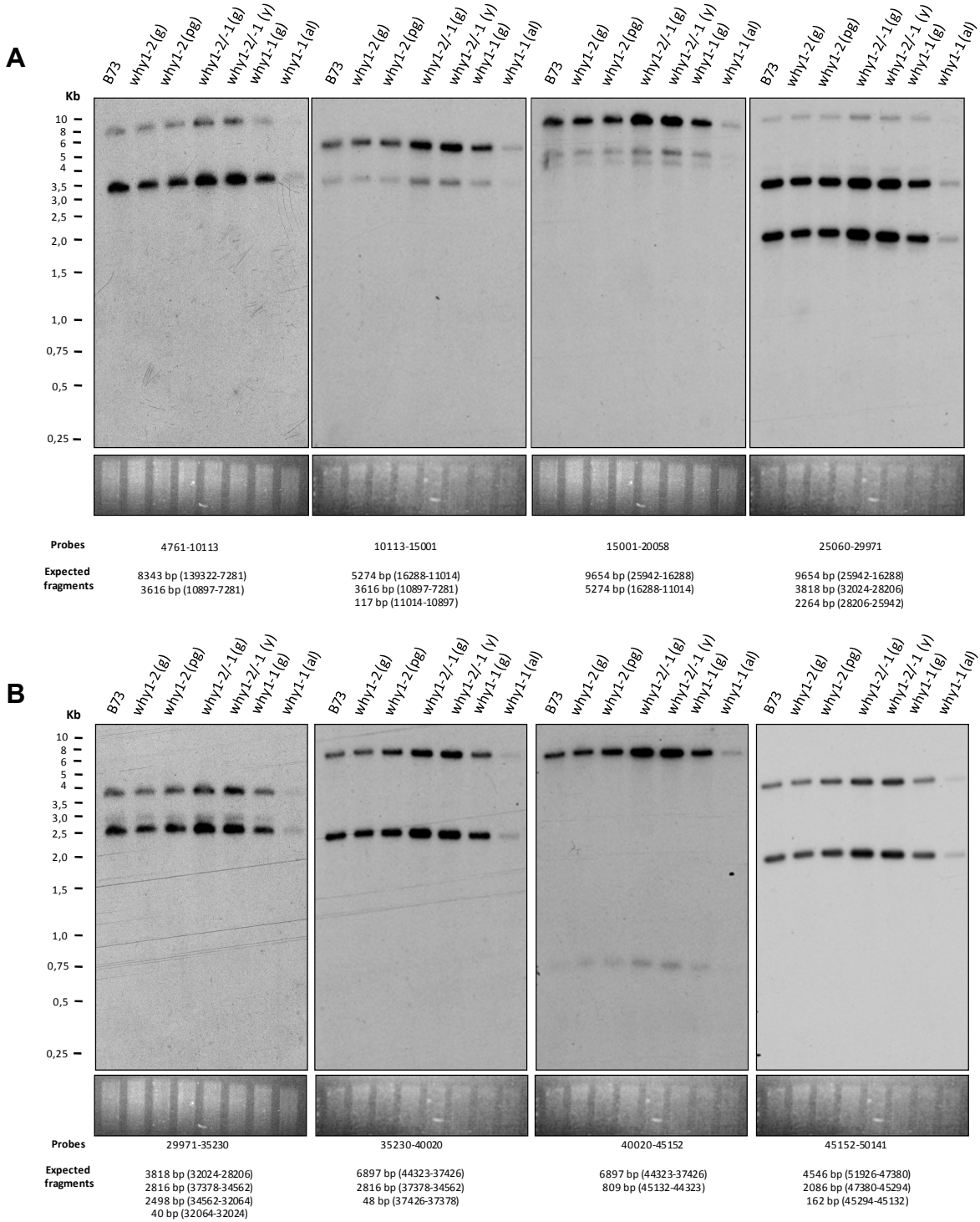


Figure 25. Absence of subgenomic amplicons in *ZmWhy1* maize mutants.

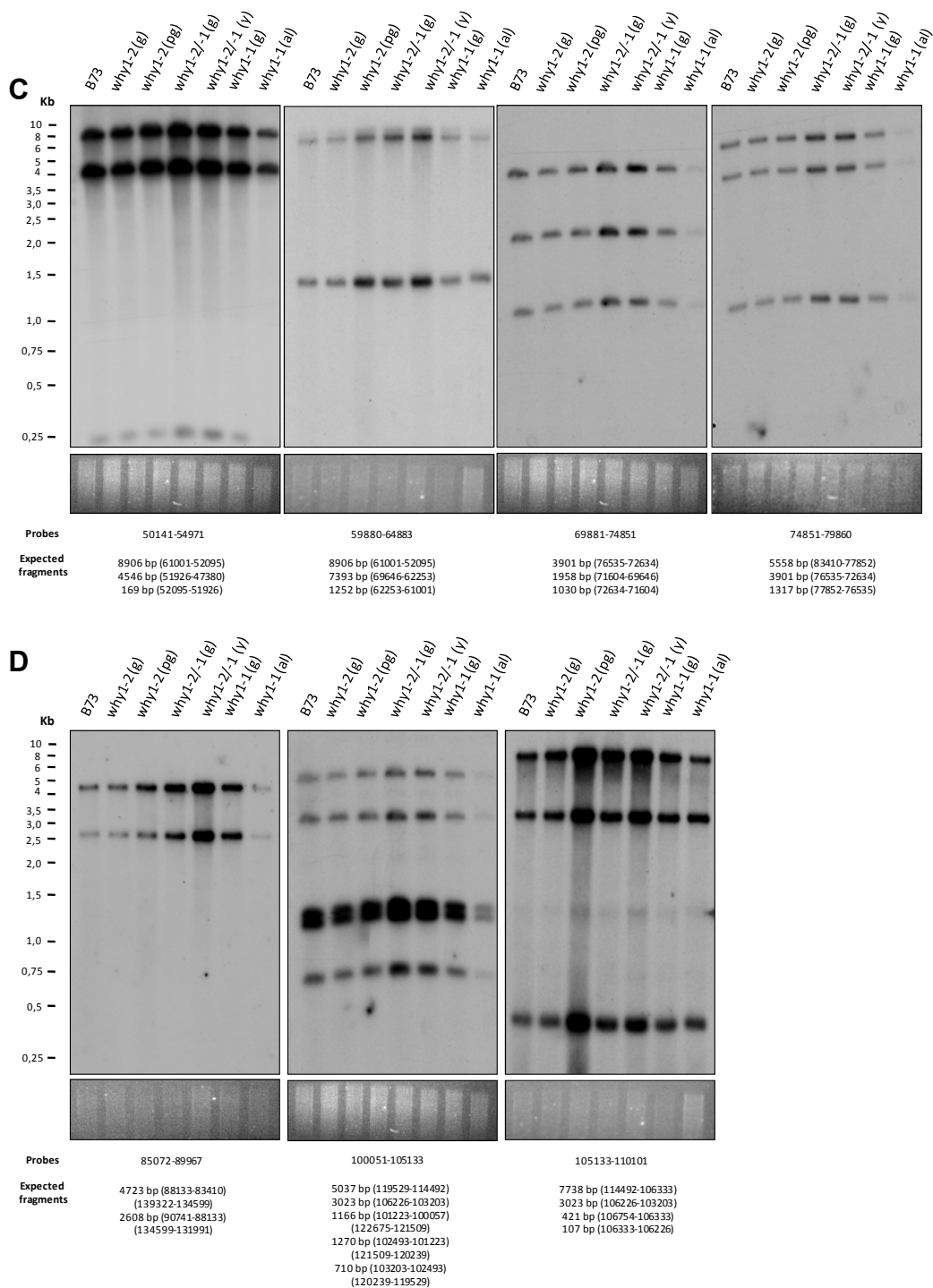


Figure 25 (suite). Absence of subgenomic amplicons in *ZmWhy1* maize mutants.

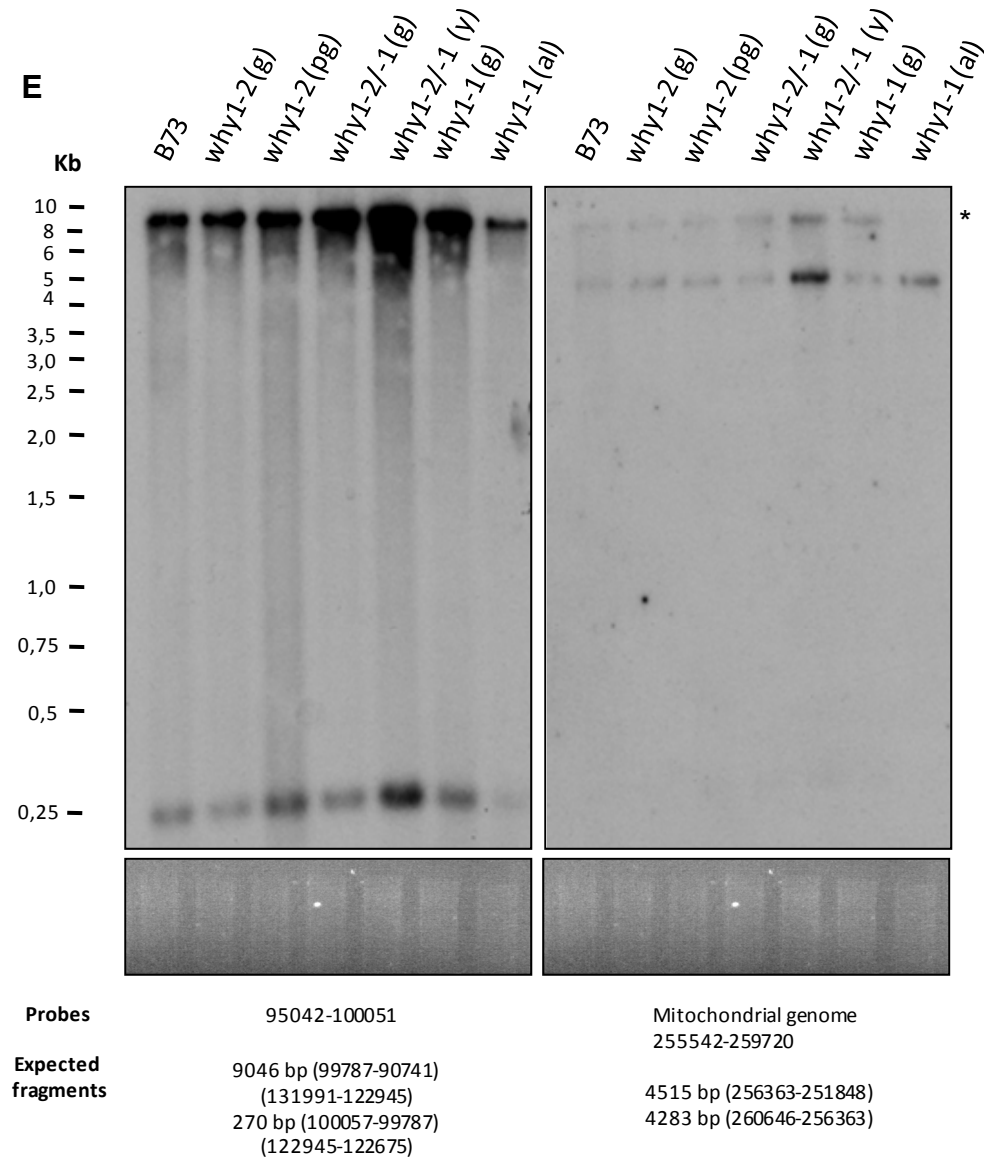


Figure 25 (suite). Absence of subgenomic amplicons in *ZmWhy1* maize mutants.

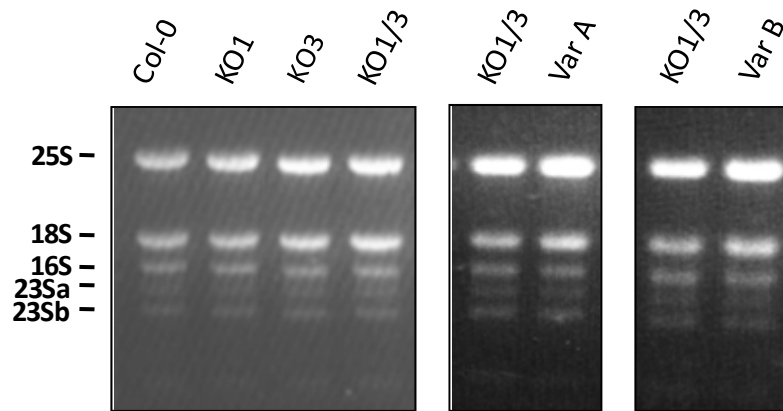


Figure 26. Levels of plastid 23S and 16S rRNA in *Arabidopsis* ptWhirlies mutants.

Table V. Recombination events mediated by short direct repeats in the *Arabidopsis* plastid genome

Event	Related ¹ Events	Reaction	Plant lines	DR positions	Short direct repeat sequences and junction	DR length and (mismatches)
1		5291REV DUP/CIR ² 20481FOR	KO1/3	5204 recombinant 21835	aaaagtttgaTTAGTTATGGGatcataaaaa agtaggttttTTAGTTATGGGatcataaaaa agtaggttttTTAGTTATGGGcctagcaaaa	(0) 11 bp (0)
2		19181FOR DEL ³ 31138REV	KO1/3	19766 recombinant 30894	ccatatctcaATAGATTGgcgtaatttt ccatatctcaATAGATTGcctcatctca actttccttgATAGATTGcctcatctca	(0) 8 bp (0)
3		20481REV DUP/CIR 30511FOR	KO1,KO3 KO1/3	20366 recombinant 31005	aatatatataCAAAAATTTTCTTTTTTTacacttctta gttaagttatCAAAAATTTGCTTTTTTTacacttctta gttaagttatCAAAAATTTGCTTTTTTTtttttagaa	(1) 18 bp (0)
4 ⁴		20481REV DUP/CIR 30511FOR	KO1/3	30300 recombinant 30827	gtcatttatgCAAAAAAAAAAAAAAatggttatgta attggtctaaCATAAAAAAAAAAAAAAatggttatgta attggtctaaCATAAAAAAAAAAAAAAagaaaaaag	(1) 14 bp (0)
5		30511REV DUP/CIR 40130FOR	KO1/3	29711 recombinant 40160	gaccctttctTTTGCTTTATgaaaaaagaa aaaaccaagaTTTGCTTTATgaaaaaagaa aaaaccaagaTTTGCTTTATctggtattaa	(0) 10 bp (0)
6	23	40130REV DUP/CIR 50251FOR	KO1/3	38989 recombinant 50299	cccctggatgCAGGAATAGCgacatggact aatgccaaaaCAGGAATAGCgacatggact aatgccaaaaCAGGAATAGCacttgatatt	(0) 10 bp (0)
7		40130REV DUP/CIR 50251FOR	KO1/3	39109 recombinant 50758	cttggtttccATTTTGGTTGtaggtgtaac gatagattgaATTTTGGTTGtaggtgtaac gatagattgaATTTTGGTTGaatttaatta	(0) 10 bp (0)
8		40130REV DUP/CIR 50251FOR	KO1/3	39266 recombinant 51611	cacctcctcgAGTAaacgcttcca cattctacctAGCAaacgcttcca cattctacctAGCAatgagtagac	(1) 4 bp (0)

9		40130REV DUP/CIR 50251FOR	KO1	39411 recombinant 50322	gtccacagaaAAAttattgctaa tgatattagtAAAttattgctaa tgatattagtAAAaatgccaga	(0) 3bp (0)
10		40130REV DUP/CIR 50251FOR	KO3	39586 recombinant 50325	ttaatgccatAATGCCtttcaaatcc tattagtaaaAATGCCtttcaaatcc tattagtaaaAATGCCcagaaaatat	(0) 6 bp (0)
11	25,35	50251FOR DEL 60182REV	KO1/3	50780 recombinant 59406	ttaattagaTTTCTTTTTTTATTTTTAAAcaggccgtg ttaattagaTTTCTTTTTTTATTTTTAAAaacttcgatt tcatatcaccTTTTTTTTTTATTTTTAAAaacttcgatt	(0) 20 bp (2)
12		50251FOR DEL 60182REV	KO1/3	50847 recombinant 59726	gggatacaaaAAAAAAAAAGAAAGTattatgaact gggatacaaaAAAAAAAAAGAAAGTcatattttct ctttgaaaacAAAAAAAAAGAAAGTcatattttct	(1) 15 bp (0)
13	14,18,21,22, 24	50251FOR DEL 60182REV	KO1/3	50474 recombinant 59953	aggttaatcgAATCATTTTTTTTCgtttggttg aggttaatcgAATCATTTTTTTTCcacaagggat cgtcccagcaAATCATTTTTTTTCcacaagggat	(0) 14 bp (0)
14	13,18,21,22, 24	50513REV DUP/CIR 68821FOR	KO1/3	50488 recombinant 70515	aaattcaggtTAATCGAATCATTTTTTTTCgtttggttg tttcaatataTAATCGAATCATTTTTTTTCgtttggttg tttcaatataTAATCGAA--TTTTTTTTtaatctaaaa	(0) 19 bp (3)
15		50513REV DUP/CIR 68821FOR	Col-0 KO1,KO3 KO1/3	50427 recombinant 69424	aattccaatcGGAGTGGATTGGgcaaggata tctaaaaaatGGAGTTGATTGGgcaaggata tctaaaaaatGGAGTTGATTGGatttgcacca	(1) 12 bp (0)
16	17	50513REV DUP/CIR 68821FOR	KO1/3	48553 recombinant 68998	tagaattgtaTCCCCCCTTCATTTATTGCTttccgatctt actctaacctTCCCCCCTTCATTTATTGCTttccgatctt actctaacctTCCCC <u>CCACT</u> ATTTTTTGCTaggtattttc	(0) 21 bp (5)
17	16	50513REV DUP/CIR 68821FOR	KO1 KO1/3	49181 recombinant 68999	tttttatttcCCCCACACCTTTTTTatataaaatt ctctaaccttCCCCACACCTTTTTTatataaaatt ctctaaccttCCCC <u>CCACT</u> ATTTTTtgctaggta	(0) 15 bp (3)

18	13,14,21,22	50513REV DUP/CIR 68821FOR	KO1	50482 recombinant 70117	caaaaattcaGGTTAATCGAATCATTtttttttcggt cgaactagcggGTTGAT-GGATCATTtttttttcggt cgaactagcggGTTGAT-GGATCATTaccctgatga	(3) 15 bp (0)
19		50513REV DUP/CIR 68821FOR	KO1/3	50243 recombinant 70143	tacctgattcATAACTAGAAAagtttctccg ccctgatgatATAA-TAGAAAagtttctccg ccctgatgatATAA-TAGAAAaggtttttct	(1) 10 bp (0)
20		50513REV DUP/CIR 68821FOR	KO3	49972 recombinant 70402	gttttttttaCTTTTTTTTTTTTTAttattgtatc cccaaagtgtCTTTTTTTTTTTTTAttattgtatc cccaaagtgtCTTTTTTTTTTTTTAcggtgtgaaa	(0) 15 bp (0)
21	13,14,18,22, 24	50513REV DUP/CIR 59291REV	Col-0 KO1,KO3 KO1/3	50489 recombinant 59938	aggttaatcgAATCATTTTTTTTCgtttggtgt cgtcccagcaAATCATTTTTTTTCgtttggtgt cgtcccagcaAATCATTTTTTTTCcacaagggat	(0) 14 bp (0)
22	13,14,18,21, 24	50513REV DUP/CIR 59291REV	KO1/3	50489 recombinant 59936	tcaggttaatCGAATCATTTTTTTTTCgtttggttg tccgtcccagCGAATCATTTTTTTTTCgtttggttg tccgtcccagCAAATCATTTTTTTTTCcacaagggat	(0) 16 bp (1)
23	6	50513REV DUP/CIR 59291REV	KO1/3	50309 recombinant 60067	atgccaaaacAGGAATAGcacttgatat accgaaaagaAAGAATAGcacttgatat accgaaaagaAGGAATAGtgcgtat	(0) 8 bp (0)
24	13,14,21,22	50513REV DUP/CIR 59291REV	KO1/3	50495 recombinant 60081	gaatcattttTTTTCGTTGGTtggttggt tagtgcgatTTTTCGTTGGGtggttggt tagtgcgatTTTTCGTTGGGgatttctgg	(1) 11 bp (0)
25	11,32,35	54551REV DUP/CIR 58730FOR	KO1/3	54523 recombinant 59387	ctcattagccTTTTTTTCGTATTTTcatttagca catatcacctTTTTTTTCGTATTTTcatttagca catatcacctTTTTTTT-TTATTTTaaaaacttcg	(0) 16 bp (2)
26		59291REV DUP/CIR 69633FOR	KO1/3	58721 recombinant 70186	attagactagACAAACAAAAAAAgttcattttc agacgggataACAAACAAAAAAAgttcattttc agacgggataACAAACAAAAAAAtagataaat	(0) 14 bp (1)

27	28	59291REV DUP/CIR 69633FOR	KO1/3	59287 recombinant 70228	acttacctcT <u>ATTTTTGTGCCTT</u> tagtaggcct ccgtacaggcT <u>TTTTTTGTGCCTT</u> tagtaggcct ccgtacaggcT <u>TTTTTTGTGCATT</u> gcatacggct	(1) 14 bp (1)
28	27	59291REV DUP/CIR 69633FOR	KO1/3	59028 recombinant 70225	tatttagtttGGCT <u>TTTTTTGT</u> acctattttt taaccgtacaGGCT <u>TTTTTTGT</u> acctattttt taaccgtacaGGCT <u>TTTTTTGT</u> gcattgcata	(0) 12 bp (0)
29	30, 36	59291REV DUP/CIR 69633FOR	KO1/3	59039 recombinant 69901	ttttttgtaCCT <u>TTTTTT</u> tattctattt tgaaatcgctCCT <u>TTTTTT</u> tattctattt tgaaatcgctCCT <u>TTTTTT</u> tattgaaccgc	(0) 10 bp (0)
30	29,35,36	59291REV DUP/CIR 69633FOR	KO3 KO1/3	59043 recombinant 69807	tttttgtaCCT <u>TTTTTT</u> TATTctatttctat ttctcttcaaT <u>TTTTTT</u> TATTctatttctat ttctcttcaaT <u>TTTTTT</u> TATTtttatattga	(0) 12 bp (0)
31		59311FOR DEL 69633REV	KO1/3	59307 recombinant 68984	tctttccggcAATTGCAATGGCTTctttatttct tctttccggcAATTGCAATGGCTTtctgctactct gtataaaaaAATT <u>CCAATGGCTT</u> tctgctactct	(0) 14 bp (1)
32	25,36	59311FOR DEL 69633REV	KO1/3	59380 recombinant 68522	accgaatcatATCACCT <u>TTTTTTTT</u> tattttaaaa accgaatcatATCCCCT <u>TTTTTTTT</u> atgtcattcg attcttccaATCCCCT <u>TTTTTTTT</u> atgtcattcg	(2) 16 bp (0)
33		58730FOR DEL 69941REV	KO1/3	58973 recombinant 69303	acatctagttGTAGTTGATTTTGAAGATGAataagtcatt acatctagttGTAGTTGATTTTGAAGATGAataggtcatt tacagtaacgGT <u>ATTTT</u> CATTTTGA <u>AAAT</u> TGAataggtcatt	(0) 21 bp (4)
34		58730FOR DEL 69941REV	KO1/3	59590 recombinant 69740	gtatataactATAGGGGATCATATAAAggaaggagat gtatataactATAGGGGATCATATAAAaatgggctg aatggttggcATAC <u>GGAAT</u> CATATAAAaatgggctg	(0) 17 bp (2)
35	11,25,30,32	58730FOR DEL 69941REV	KO1/3	59390 recombinant 69822	atcaccttttTTTTTTTATTTTaaaaacttcg atcaccttttTTTTTTTATTTTtatattgaag ctcttcaataTTTTTTTATTTTtatattgaag	(0) 12 bp (0)

36	30,32	58730FOR DEL 69941REV	KO1/3	59030 recombinant 68524	tttttgtaccTATTTTTTTATtctatttcta tttttgtaccTATTTTTTTATgtcattcgaa ccaatcccctTATTTTTTTATgtcattcgaa	(0) 11 bp (0)
37		58730FOR DEL 69941REV	KO1/3	59074 recombinant 69714	ctattctataTATTTCTATTAGGTTGTATAttaatattag ctattctataTATTTCTATTAGGTTGTTTAatggttggca tggctgtctgTATTTCTAATAAGTTGTTTAatggttggca	(1) 20 bp (2)
38		69633REV DUP/CIR 79920FOR	KO1/3	69468 recombinant 80032	atgtaaaccaTAAGTTTCAGAcacaatagag tcacaaattgTAAGTTACAGAcacaatagag tcacaaattgTAAGTTACAGAtatctaattt	(1) 11 bp (0)
39		79920REV DUP/CIR 111330FOR	KO1/3	78747 recombinant 111896	aagctcttcgCATcgcaatgcct ggtaaccataCATcgcaatgcct ggtaaccataCATgaagaggaaa	(0) 3 bp (0)
40		79920REV DUP/CIR 111330FOR	KO1/3	76183 recombinant 111428	tgggcagcaaAGGGTGTTCCTCTTcttgtacccc ccacaaagagAGAGTGTTCCTCTTcttgtacccc ccacaaagagAGAGTGTTCCTATTaaaaaggccg	(1) 14 bp (1)
41		79920REV DUP/CIR 111330FOR	KO1/3	78830 recombinant 111688	ataataaagaCGCTTACTGTCTCTTTttgattcaac ggctaaaccCGCTTACTGTCTCTTTttgattcaac ggctaaaccCGCTTAATGTCTTTTTgagcaagagc	(0) 16 bp (2)
42		84761REV DUP/CIR 95260FOR	KO1/3	84370 recombinant 95699	acggcgacgaAGAATCAAAGTctcactatatt ctggagtgggAGAATCAAAGTctcactatatt ctggagtgggAGAGTCAGAGTCgaaaagagga	(0) 13 bp (2)
43		105983REV DUP/CIR 116007FOR	KO1/3	104982 recombinant 116127	gtgggagagcAAAAAAGcgtcgtgctg tcctgccaagAAAAAAGcgtcgtgctc tcctgccaagAAAAAAGtgcagcgcca	(0) 8 bp (0)
44		105983REV DUP/CIR 116007FOR	KO1 KO1/3	105322 recombinant 116009	tggcttgggtAAGGGAACCCACCGgagccgtagc ccggcaatgcAAGGGAACCCACCGgagccgtagc ccggcaatgcAAGGGAAGCCATCGataaaatagt	(0) 14 bp (2)

¹ Related events implicate at least part of a direct repeat that is found independently in another recombination event.

² **DUP/CIR = Illegitimate recombination event yielding a duplicated/circularized product and detected using outward-facing PCR.**

³ **DEL = Illegitimate recombination event yielding a deletion-containing product and detected using inward-facing PCR.**

⁴ **Non-specific annealing of primer 20481REV at positions 30367-30346 yielded this product.**

Table VI. Recombination events mediated by short direct repeats in the maize plastid genome

Events	Related Events	Reaction	Plant lines	DR positions ²	Short direct repeat sequences and junctions	DR length and (mismatches)
1		15001REV DUP/CIR ¹ 50141FOR	Why1-1	14923 recombinant 50999	aattcgactcCCAACCTTTTTTTtaga aattagataaCCAACCTTTTTTTtaga aattagataaCCAACCTgctgcattgt	(0) 8 bp (0)
2		15001REV DUP/CIR 50141FOR	Why1-1	14492 recombinant 50341	aggagtgagaTGACTTTCCAAAatTTTTcata aacaatatgTGATTTTCCAAAatTTTTcata aacaatatgTGATTTTCCAAAatTggaaaata	(1) 12 bp (0)
3		25060REV DUP/CIR 35230FOR	Why1-2/-1	24850 recombinant 36373	tggcgaagtcACAAGGccctccactt aacaagctgtACAAGGccctccactt aacaagctgtACAAGGagctctagga	(0) 6 bp (0)
4		25060REV DUP/CIR 35230FOR	Why1-2	25015 recombinant 35708	ggggtagaatTTgttgattctc ataaaaggaaTTgttgattctc ataaaaggaaTTctatcaattt	(0) 2 bp (0)
5		25060REV DUP/CIR 35230FOR	Why1-1	24796 recombinant 35785	tttccctcaACAAATAAAGgcttgggcta cagagcaaacACAAATAAAGgcttgggcta cagagcaaacACAAATAAAGaaacaacttt	(0) 10 bp (0)
6		25060REV DUP/CIR 35230FOR	Why1-1	24258 recombinant 35542	cccgagtaccCGGGAAAAGCagaatatttg atgcttttaaCGGGAAATAGCagaatatttg atgcttttaaCGGGAAATAGCtagtgtagaa	(1) 11 bp (0)
7		29971REV DUP/CIR 54971FOR	Why1-1	29627 recombinant 55252	tttctgaagAGGCGGATAagatattagg ctgcctaataAGGCGGATAagatattagg ctgcctaataAGGCGGATActtccgatcc	(0) 9 bp (0)
8		29971REV DUP/CIR 54971FOR	Why1-1	29507 recombinant 55692	attattggagTCGATAcacaactcac gtatctaactCGATAcacaactcac gtatctaactCGATAaaggcaggcg	(0) 6 bp (0)

9		29971REV DUP/CIR 54971FOR	Why1-1	29434 recombinant 55437	gtagatcggTTTTTTTCATTcttcaagaac ttcctcaataTTTTTTCATTcttcaagaac ttcctcaataTTTTTTCATTaattactccg	(0) 11 bp (0)
10	11	29971REV DUP/CIR 54971FOR	Why1-1 Why1-2	26297 recombinant 56132	ctaaggaatcAAAAAAAAAGGAAaaattggggtt ataatTTTcaAAAAAAAAAGGAAaaattggggtt ataatTTTcaAAAAAAAAAGGAAattgtcgaaa	(0) 12 bp (0)
11	10	29971REV DUP/CIR 54971FOR	WT	26691 recombinant 56131	tctaaggaatCAAAAAAAGgaaaaattg acataatTTTCAAAAAAAGgaaaaattg acataatTTTCAAAAAAAGgaaattgt	(0) 9 bp (0)
12		40020REV DUP/CIR 45152FOR	Why1-2/-1	39662 recombinant 45567	ggcgcgcaAGGAAACTATAcccgaatcc ttaatgctagAGGAAACTATAcccgaatcc ttaatgctagAGGAAAACGATAgctccaagaa	(0) 12 bp (1)
13		45152REV DUP/CIR 59880FOR	Why1-2	44350 recombinant 60179	tccgcaatTTCCGAATCGCCCTgtagaatggc caaaaagctgCCGAATCGCCCTgtagaatggc caaaaagctgCCGAATGGCCTatTTcttgcg	(0) 12 bp (2)
14		45152REV DUP/CIR 59880FOR	Why1-1	45065 recombinant 60562	agtattcactAGAAA--ATTTCCTCCTTTCTTTTTTTcttc atgaaaaaaaaAGAAAGCATTGCCTCCTTTCTTTTTTTcttc atgaaaaaaaaAGAAAGCATTGCCTTTCTTTCTTatatcttgta	(3) 22 bp (1)
15		45152REV DUP/CIR 59880FOR	Why1-1	44815 recombinant 60324	gattcctaaaTTTGTCCcatatcgtgg ttcgaatttgTTTTGTCCcatatcgtgg ttcgaatttgTTTTGTCCaagtgagata	(0) 8 bp (0)
16		45152REV DUP/CIR 59880FOR	Why1-1	45075 recombinant 61177	ttcctcctTTCTTTTTTTcttctttct aaatTTTgggTCTTTTTTTcttctttct aaatTTTgggTCTTTTTTTgtttaaccg	(0) 9 bp (0)
17		45152REV DUP/CIR 59880FOR	Why1-2/-1	44253 recombinant 60796	cgtaatcTTTAAACAGTTCtgtgcttcaa acacatgtacAAACAGTTCtgtgcttcaa acacatgtacAAAACTTCctataggaat	(1) 9 bp (2)

18		54971REV DUP/CIR 64883FOR	Why1-1	54631 recombinant 65071	ataaagattt A ATTT C ATTT C TT C AATttgctctcct ccgaaaatcc A ATTT C ATTT C TT C AATttgctctcct ccgaaaatcc A ATTT C ATTT- T TC A ATgggggtagat	(0) 17 bp (1)
19		59880REV DUP/CIR 74851FOR	Why1-1	58828 recombinant 74980	caactgggta T T C T A TT C CA T TT C T A CTTTTT A aattaaaggg ccaattcaatTTTT A TT C CA T TT C T A CTTTTT A aattaaaggg ccaattcaatTTTT A TT A CA T T- A CTTTTT A ttatagtatg	(1) 22 bp (3)
20		59880REV DUP/CIR 74851FOR	Why1-1	59266 recombinant 76157	tatttcgggc A ATT G CAAT G ACTTccttatttct aagaggaaaa A ATT G CAAT G ACTTccttatttct aagaggaaaa A ATT G CAAT G A T TTaaagcgagaa	(0) 14 bp (1)

¹ DUP/CIR = Illegitimate recombination event yielding a duplicated/circularized product and detected using outward-facing PCR.

² The positions of the direct repeats refer to the numbers of the nucleotides in the published Maize Chloroplast Genome Sequence (10)

Table VII. Frequency of recombination events mediated by short direct repeats in the maize plastid genome

Plant line	PCR reactions	Recombination events
WT	19	1
Why1-2	19	3
Why1-2/-1	19	3
Why1-1	19	14

**Chapitre 4. Structure of DNA-Whirly Complexes
Provides Insights into the Maintenance of Plant
Organelle Genome Stability**

Article soumis à *Molecular Cell*

Structure of DNA-Whirly Complexes Provides Insights into the Maintenance of Plant Organelle Genome Stability

Laurent Cappadocia¹, Alexandre Maréchal¹, Jean-Sébastien Parent¹, Étienne Lepage¹, Jurgén Sygusch^{1,*} and Normand Brisson^{1,*}

¹Department of Biochemistry, Université de Montréal, PO Box 6128, Station Centre-Ville, Montréal, Québec, H3C 3J7, Canada

*Correspondence should be addressed to J.S. or N.B.

Running title: **Whirly-DNA complexes in organelle genome stability**

4.1 Mise en contexte

Cet article fait suite à celui présenté dans le Chapitre 3 et avait pour objectif de caractériser la fonction des Whirlies dans la réparation de l'ADN des génomes d'organelles suite à l'induction de bris double-brins. Il visait également la description moléculaire du mécanisme de liaison de l'ADNsb par les protéines de la famille Whirly afin de préciser le rôle de ces protéines dans le maintien de la stabilité du génome des organelles. Il a permis d'établir que les Whirlies ont un rôle à jouer dans la réparation des bris double-brins de façon conservatrice dans les organelles. Il a aussi permis de comprendre en détail les interactions entre les Whirlies et l'ADNsb et de proposer que les Whirlies pourraient moduler l'accès des protéines de réparation à l'ADNsb et empêcher ce dernier de former des duplexes avec des séquences d'ADN complémentaires de façon à favoriser la prise en charge des bris par des mécanismes fidèles de réparation de l'ADN. Pour cet article, j'ai proposé l'utilisation de la ciprofloxacine et de la novobiocine afin d'étudier la réponse des organelles aux bris double-brins en présence et en absence de Whirlies plastidiques et/ou mitochondriaux. J'ai supervisé les stagiaires Étienne Lepage et Sandra Grondin dans la réalisation des tests de résistance à ces composés et à d'autres stress reliés à l'ADN (Figures 31-32 et résultats non-publiés) et également collaboré avec Jean-Sébastien Parent et Laurent Cappadocia pour la réalisation des figures 27A-E et des tableaux VIII, IX et XI. LC a réalisé toutes les expériences de détermination de structure des complexes Whirly-ADN ainsi que les essais d'activité *in vitro* (Figures 28-30, 33-35). LC a écrit une première version de l'article qui a été corrigée par JSP et moi-même puis par Normand Brisson et Jurgen Sygusch.

4.2 Summary

Double-strand breaks are highly detrimental to genomes and need to be quickly repaired. Here, we show that plant mitochondria and plastids can repair such lesions through a pathway similar to the microhomology-mediated break-induced replication observed in humans, yeast and bacteria. Plants lacking ssDNA-binding proteins from the Whirly family accumulate elevated levels of DNA rearrangements upon double-strand break induction indicating that Whirlies contribute to the accurate repair of plant organelle genomes. Significantly, crystal structures of Whirly-DNA complexes reveal a conserved ssDNA-binding mechanism where DNA is stabilized between domains of adjacent subunits and rendered unavailable for duplex formation and/or protein interactions. Like eubacterial single-stranded DNA-binding protein SSB, Whirlies possess an acidic/aromatic C-terminus that is displaced upon ssDNA-binding making it available for potential recruitment of protein partners. We propose that Whirly proteins modulate the access of repair proteins to ssDNA, thereby regulating the choice of repair pathways and maintaining plant organelle genome stability.

4.3 Introduction

Maintaining a proper balance between genome stability and instability is crucial for all organisms. Cells have therefore developed numerous mechanisms to ensure that their genomes remain functional yet dynamic. These mechanisms mainly rely on DNA replication, recombination and repair (DNA-RRR) pathways. Single-stranded DNA-binding (SSB) proteins are critical players in these pathways, as single-stranded DNA (ssDNA) is an obligate intermediate during DNA-RRR processes. Often envisioned as simple ssDNA-coating proteins, the SSB proteins actually work as organizational scaffolds that can recruit genome maintenance complexes when and where they are most required (Cox, 2007; Shereda et al., 2008). Consistent with their important role, SSB proteins are present in all realms of life (Shereda et al., 2008) and their alteration can lead to strong genomic instability and even cell death (Porter and Black, 1991; Symington, 2002).

Whirlies are proteins that bind preferentially to ssDNA (Desveaux et al., 2002; Desveaux et al., 2000; Desveaux et al., 2005) and that are mainly found in the plant kingdom, where they localize to the nucleus, plastids or mitochondria (Desveaux et al., 2005; Krause et al., 2005). In *Arabidopsis*, the Whirly proteins AtWhy1 and AtWhy3 are targeted to the plastids, whereas AtWhy2 is targeted to the mitochondria (Krause et al., 2005; Maréchal et al., 2008; Maréchal et al., In press). The Whirlies are characterized by a transit peptide at their N-terminus, a Whirly domain that accounts for ssDNA binding and a conserved acidic-aromatic C-terminal tail (Desveaux et al., 2002). The crystal structure of a Whirly protein in the free form revealed that these proteins form tetramers resembling whirligigs (Desveaux et al., 2002). Binding to ssDNA is thought to occur on a curved β -sheet surface, but the precise mode of ssDNA binding remains elusive (Desveaux et al., 2002). Like other SSB, Whirly proteins perform numerous activities related to the metabolism of ssDNA, including regulation of transcription (Desveaux et al., 2000) (Desveaux et al., 2004) and modulation of telomere length (Yoo et al., 2007). Whirlies have also been shown to bind certain chloroplast RNA and to be important for the proper biogenesis of chloroplasts (Prikryl et al., 2008).

Recently, we demonstrated that the plastid genome of *Arabidopsis* plants becomes unstable and accumulates rearranged DNA molecules in the absence of both AtWhy1 and AtWhy3 (Maréchal et al., In press). This phenomenon is not unique to *Arabidopsis*, as rearranged DNA molecules also accumulate in maize lines that have reduced levels of ZmWhy1, a plastid-localized Whirly protein. The rearranged molecules correspond to deletion, duplication and/or circularization events that are mediated by microhomologies (Maréchal et al., In press). Microhomology-mediated DNA rearrangements (MHMR) are widespread throughout evolution, and have been described in humans, yeasts and bacteria (Hastings et al., 2009). In humans, numerous copy number variants, as well as several diseases, are associated with MHMR (Lupski, 2009). These rearrangements appear following DNA-related stresses and notably in response to double-stranded DNA breaks (DSB) (Hastings et al., 2009), suggesting that they could be a consequence of an error-prone DNA repair pathway.

The accumulation of MHMR in *Arabidopsis* plastids devoid of AtWhy1 and AtWhy3 (Maréchal et al., In press) suggests that Whirly proteins could protect organelle genomes against rearrangements occurring under DNA stresses. Here, we report that treatment of *Arabidopsis* plants with ciprofloxacin, a compound that induces DSB in genomes by inhibiting the DNA cleavage/resealing reaction of DNA gyrases (Drlica and Zhao, 1997; Gmuender et al., 2001; Nöllmann et al., 2007), leads to an accumulation of DNA rearrangements including MHMR in both plastid and mitochondrial genomes. Importantly, plants lacking plastid or mitochondrial Whirlies accumulate higher levels of MHMR than the wild-type controls, suggesting that Whirlies are components of the organelle repair machinery. To further understand the role of Whirly proteins in the repair response, we obtained the crystal structures of the mitochondrial Whirly protein StWhy2 in complex with various ssDNA oligonucleotides. These reveal how Whirly proteins protect and stabilize the single-stranded form of DNA and accommodate ssDNA molecules that differ in sequence.

4.4 RESULTS

4.4.1 DSB induce MHMR accumulation in the plastids of *Arabidopsis*

We previously demonstrated that, in the absence of AtWhy1 and AtWhy3, rearranged DNA products accumulate in the plastid genome of *Arabidopsis*. We proposed that these rearrangements could occur following replication fork collapse or DSB (Maréchal *et al.*, In press). To verify whether DSB are indeed involved in the formation of rearranged DNA molecules, we treated Col-0 (WT) and AtWhy1/3 double knock-out (KO1/3) lines of *A. thaliana* with ciprofloxacin, an inhibitor of organelle-targeted DNA gyrases (Wall et al., 2004) that induces DSB in both plastids and mitochondria (Figure 31A). Ciprofloxacin induced leaf etiolation and inhibited seedling growth in both WT and KO1/3 lines. However, KO1/3 seedlings were more affected than WT plants (Figures 31A and 31B). We then monitored DNA rearrangements using a PCR approach (Figure 27A) and primer pairs scattered throughout the plastid genome. We used outward-facing PCR

primers to monitor DNA duplication/circularization events and inward-facing PCR primers to detect deletions.

The size and the abundance of the PCR products varied according to the primer pairs and the individual plants used. A common trend was nevertheless observed for all reactions yielding PCR products (Figure 27B). Indeed, for WT untreated plants, no more than one PCR product could be detected per PCR reaction. At intermediate and high ciprofloxacin concentrations, PCR products became more diverse and/or more abundant. For KO1/3 plants, PCR products were diverse and/or abundant even for untreated plants, in agreement with our previous finding that untreated KO1/3 plants accumulate elevated levels of rearranged DNA molecules (Maréchal et al., In press). Ciprofloxacin-treated KO1/3 plants had more diverse and/or more abundant PCR products than both untreated KO1/3 plants and ciprofloxacin-treated WT plants. Our results thus indicate that ciprofloxacin induces DNA rearrangement accumulation in both genotypes and that KO1/3 plants are more sensitive to ciprofloxacin than WT plants.

To ensure that DNA rearrangements were due to DSB induction rather than DNA gyrase inhibition, we treated WT and KO1/3 plants with novobiocin, a DNA gyrase inhibitor that does not produce DSB (Gmuender et al., 2001). Novobiocin inhibited seedlings growth in both WT and KO1/3 lines (Figure 31A) without significant etiolation of the leaves (Figure 31B). In contrast to ciprofloxacin, novobiocin-treated plants did not show correlation between inhibitor concentration and PCR product accumulation (Figure 27C). Indeed, while some reactions showed higher PCR product accumulation at 100 μ M novobiocin, other reactions showed higher PCR product accumulation at 0 μ M novobiocin. This likely reflects a variation between individual plants rather than a novobiocin effect on MHMR accumulation, thus confirming that DSB induction, rather than DNA gyrase inhibition, leads to DNA rearrangements in plastids.

4.4.2 DSB induce MHMR accumulation in the mitochondria of *Arabidopsis*

In contrast to AtWhy1 and AtWhy3, no function has yet been ascribed for the mitochondria-localized protein AtWhy2 in DNA metabolism. To verify if AtWhy2 could be important to maintain mitochondrial genome stability under DNA stress conditions, we repeated the experiment with gyrase inhibitors on AtWhy2-deficient (KO2) lines. No phenotypic differences could be observed between KO2 and WT plants in the presence or absence of ciprofloxacin or novobiocin (Figures 32A and 32B). We then monitored rearrangements in the mitochondrial DNA following ciprofloxacin treatment using the PCR approach described above and primer pairs scattered throughout the mitochondrial genome. At 0 μM ciprofloxacin, no DNA rearrangements were detected in either WT or KO2 plants (Figure 27D). At 0.25 μM ciprofloxacin, PCR products accumulated, but only in KO2 plants (Figure 27D). At 0.75 μM ciprofloxacin, PCR products accumulated to high levels in both WT and KO2 plants (Figure 27D). No PCR products could be detected in plants treated with 100 μM novobiocin (Figure 27E), thus confirming that mitochondrial DNA rearrangements depend on DSB induction. These results indicate that KO2 plants accumulate DNA rearrangements in the mitochondrial genome more readily than WT plants.

4.4.4 Most of the detected DNA rearrangements are mediated by microhomologies

A total of 189 DNA rearrangement events could be detected through cloning and sequencing of PCR products (Table XI). We classified those rearrangements in two groups: MHMR products that had at least 5 bp of microhomology at the junction and nonhomologous end joining (NHEJ)-like products that had less than 5 bp of microhomology at the junction (McVey and Lee, 2008). MHMR and NHEJ-like products

constituted respectively 84 % and 16 % of all DNA rearrangements (Table VIII) and were present in both deletion and duplication/circularization events (Table XI). Although MHMR junctions varied in sequence, they were frequently A/T rich and often contained homopolymeric tracts (Table IX). Some MHMR junctions appeared recurrently in different genotypes and/or in different plant treatments and were counted as independent events. The 48 mitochondrial MHMR had a mean size of 12.6 ± 5.2 bp whereas the 104 plastid MHMR had a mean size of 14.5 ± 5.5 bp. Globally, the size of the MHMR events varied from 5 to 33 bp. No significant difference was observed between the mean size of plastid junctions from ciprofloxacin-treated and untreated plants and that of mitochondrial junctions from ciprofloxacin-treated plants. This, plus the fact that MHMR were more abundant in Whirly-deficient than in WT organelles (Table VIII), indicates that a microhomology-dependent repair pathway is common to both organelles and is more active in Whirly-deficient plants.

4.4.5 The crystal structure of StWhy2 reveals how ssDNA is stabilized

To understand how Whirly proteins prevent MHMR in the organelle genomes of *Arabidopsis*, we sought to determine how these proteins bind ssDNA. We thus solved the crystal structure of StWhy248-216, a close homolog of AtWhy2 that also localizes to mitochondria (Vermel et al., 2002), in the free form and bound to ssDNA. These structures were obtained at 2.20-Å resolution (Table X) (Cappadocia et al., 2008). The StWhy248-216 construct, hereafter called StWhy2, encompasses the entire ssDNA-binding domain but lacks the mitochondria targeting sequence and the acidic/aromatic C-terminus (Figure 28A). For the ssDNA sequence, we chose ERE32, a 32-nucleotide ssDNA derived from the Elicitor Response Element (ERE), as the interaction between this sequence and plant Whirlies is well documented (Desveaux et al., 2000; Desveaux et al., 2004). StWhy2 and StWhy2-ERE32 crystallized in different chemical environments, yet they both gave rise to crystals of the F432 space group. In the two crystals, StWhy2 tetramers are disposed around four-fold crystallographic axes (Figures 28B and 28C).

In the StWhy2-ERE32 structure, electron density corresponding to ssDNA was clearly visible on the edges and in between the β -sheets of adjacent protomers (Figure 28C), enabling us to build a nine-nucleotide model. Although we could observe electron density for most nucleobases, we were unable to assign the ERE32 sequence into the electron density. We thus modeled the thymine-rich ERE32 oligonucleotide as an all-thymine oligonucleotide. No clear density was observed for ssDNA in the β -sheets (Figure 28C). This was unexpected, since most single-stranded nucleic acid proteins use the core of their β -sheets as a primary binding platform (Horvath, 2008). The absence of clear electron density in the β -sheets suggests that the ssDNA is disordered and/or adopts multiple conformations in this location, probably because the ssDNA can either go from one ssDNA-binding site to the other or enter/leave the complex at this location.

In each ssDNA-binding site, StWhy2 maintains the ssDNA in an extended conformation (Figure 28D). The single-strand preference of StWhy2 arises from steric impediment for the binding of a second strand at several positions on the binding surface. The presence of abrupt twists in the DNA backbone, such as between nucleotides 2 and 3 (Figure 28D), also prevents the binding of a double-stranded DNA helix. The mode of ssDNA binding is dominated by stacking and hydrophobic interactions between adjacent nucleobases and between nucleobases and aromatic/hydrophobic protein residues, consistent with the salt resistant binding of StWhy2 to ssDNA (Vermel et al., 2002). The polar contribution to DNA binding involves seven hydrogen bonds plus one water-mediated interaction (Figure 28D). We did not observe strong sequence-specific interactions between StWhy2 and ERE32. Indeed, most of the nucleobases have their edges, containing the sequence-specific binding moieties, exposed to the solvent whereas the faces of the nucleobases make intimate contact with residues of the protein surface. StWhy2 only forms two sequence-specific hydrogen bonds with ERE32, which are directed toward nucleotides 3 and 7 (Figures 33A and 33B).

4.4.6 StWhy2 can bind nucleobases that differ in size and in functional groups

The local disorder of the electron density around the nucleobases, as well as the scarcity of sequence-specific contacts between StWhy2 and ERE32, suggests that StWhy2 could bind other ssDNA sequences with high affinity. We thus assayed the binding of StWhy2 to different A/T-rich ssDNA, as plant organelle genomes are generally A/T-rich (Fauron et al., 2004; Ravi et al., 2008). We compared the binding of the thymine-rich ERE32 oligonucleotide to that of a dT32 oligonucleotide, and to that of two adenine-rich oligonucleotides, cERE32, which is complementary to ERE32 and rcERE32, which is the reverse-complementary to ERE32. Quantitative electrophoretic mobility shift assay (EMSA) revealed that, despite their different nucleotide sequences, ERE32, cERE32, rcERE32 and dT32 all interacted with StWhy2 with similar nanomolar affinities (Figures 29A and 29B). To gain structural insights on how StWhy2 could accommodate these DNA sequences, we obtained the crystal structure of these oligonucleotides bound to StWhy2 (Tableau X) and we compared them to that of StWhy2-ERE32. The four oligonucleotides adopt similar conformations (Figure 29C), suggesting that they interact with StWhy2 through a common mechanism. The structures also reveal how StWhy2 can accommodate bases that differ in size and in functional groups. Indeed, the two sequence-specific interactions formed with thymine residues in the StWhy2-dT32 and StWhy2-ERE32 structures make equivalent interactions in the structures of StWhy2-cERE32 and StWhy2-rcERE32, which are Ser62-A7 and Asp140-A3 (Figures 33C and 33D). These compensating interactions demonstrate the capacity of StWhy2 to bind different DNA sequences with high affinity and provide an explanation for the binding of Whirly proteins to every tested region of plastid and mitochondrial DNA *in vivo* (Maréchal et al., 2008; Maréchal et al., In press; Prikryl et al., 2008).

4.4.7 StWhy2 can destabilize dsDNA and protect ssDNA against nuclease degradation

Whirly proteins were shown to bind ssDNA more strongly than dsDNA (Desveaux et al., 2000; Prikryl et al., 2008). To gain further insight into the biochemical properties of Whirly proteins, we incubated a radiolabeled DNA duplex with StWhy2 protein (Figure 34A) and monitored the StWhy2-induced denaturation of dsDNA. As the concentration of StWhy2 is augmented, formation of ssDNA is detected. This reveals that StWhy2 can destabilize a DNA duplex, probably by binding more avidly to the single-strand form of DNA and driving the equilibrium toward the formation of ssDNA. In addition to promoting the formation of ssDNA, StWhy2 effectively protects DNA against nuclease degradation (Figure 34B). In this experiment, M13mp18 ssDNA was incubated with a saturating amount of StWhy2 before addition of Mung Bean nuclease. The protection of the ssDNA against degradation suggests that StWhy2 can efficiently prevent Mung Bean nucleases from accessing the ssDNA.

4.5 DISCUSSION

4.5.1 Plant organelles can repair DSB through a MHMR mechanism

Plants treated with ciprofloxacin accumulate high levels of rearranged DNA molecules in both plastids and mitochondria. Most of the DNA rearrangements contain microhomologies at their junction, thus suggesting that ciprofloxacin-induced DSB are repaired through an error-prone microhomology-dependent DNA rearrangement pathway. Some of the rearranged molecules contain less than 5bp of homology at their junctions and thus, a NHEJ-like repair pathway may be involved in their formation. However, there is little support for the existence of such a repair pathway in plant organelles and notably in plastids (Kohl and Bock, 2009; Odom et al., 2008). Furthermore, whereas MHMR are

detected in absence or in presence of ciprofloxacin, all but one NHEJ-like events are detected in ciprofloxacin-treated plants, thus raising the possibility that NHEJ-like events could be byproducts of the ciprofloxacin-mediated DNA gyrase inhibition. As DNA gyrases are tetramers, DNA rearrangements could result from an exchange of DNA gyrase subunits bound to different sites (Marvo et al., 1983).

In contrast to NHEJ-like events, MHMR with similar junction sizes could be detected in both treated and untreated plants, at least for plastid events, thus suggesting that MHMR formation occurs through a common pathway in both treated and untreated plants. We previously proposed that a microhomology-mediated break-induced replication (MMBIR) pathway could explain the accumulation of MHMR in both untreated KO1/3 *Arabidopsis* plants and in untreated ZmWhy1 knockdown maize plants (Maréchal et al., In press). The MMBIR pathway relies on the microhomology-dependent restart of DNA synthesis on a different template following collapse of the replication fork (Hastings et al., 2009). Microhomologies are sufficient to initiate DNA synthesis as certain DNA polymerases can efficiently use mismatched primers or primers as short as 2-3 bp as substrates (Cannistraro and Taylor, 2007; Hastings et al., 2009). As our DSB-induced MHMR have similar junction sizes as uninduced MHMR, at least in the plastids, it is possible that these DSB are repaired through a MMBIR pathway. MMBIR can explain our observation of both duplications/circularization and deletion events, whereas pathways such as microhomology-mediated end-joining or single-strand annealing can only account for deletion events (McVey and Lee, 2008). Additionally, we observed that 91 out of 156 MHMR junctions contain mismatches relative to one or both parental strands (Table IX). This further underlines the low-specificity requirements of the pairing process. Finally, MMBIR can account for complex cases of DNA rearrangements (Hastings et al., 2009), including the formation of molecules containing multiple MHMR which have also been observed in our study (Table IX). The characteristics of the MHMR events obtained in this study are thus in agreement with the idea that MMBIR acts as a repair pathway in plant organelles and that Whirly proteins modulate this pathway.

Intriguingly, some PCR reactions yielded numerous DNA rearrangement products, whereas others yielded no products (Table XI). Along with the recurrence of certain DNA rearrangements, this may suggest (i) that DNA rearrangement hotspots exist, (ii) that only certain regions can tolerate DNA rearrangements, and (iii) that DSB have a non-random distribution due to the binding of DNA gyrases at preferred locations (Fisher et al., 1981).

Our finding that DSB can lead to MHMR product accumulation has important evolutionary implications. In many organisms, MHMR could be detected when canonical repair pathways such as NHEJ or homologous recombination were overloaded or challenged, leading to the idea that microhomology-mediated DNA repair acts as a backup DNA-repair pathway (Wang et al., 2003). In *Arabidopsis* organelles, multiple DSB may cause depletion of repair proteins due to their recruitment to multiple damage sites, allowing the accumulation of MHMR. The use of an error-prone pathway under DNA stress may however be more than a backup pathway. Indeed, as MHMR enables DNA rearrangements including copy number variation (Hastings et al., 2009), it promotes rapid changes in the genome. Some of these changes may confer plants selective advantages under DNA stress. A similar mechanism, termed stress-induced mutation, exists in bacteria and is also dependent on error-prone repair of DSB (Ponder et al., 2005). The fact that certain plants maintain DNA rearrangements mediated by microhomologies in their organellar genome provides indirect support for their selective advantage (Kanno et al., 1993; Moeykens et al., 1995; Ogihara et al., 1988).

4.5.2 Mechanism of ssDNA-binding of Whirly proteins

The structure of StWhy2 is well suited for binding and protecting ssDNA with only modest sequence preference. Indeed, as the ssDNA is sandwiched between adjacent protomers, it becomes unavailable for interacting with other proteins. This also restricts the possibility of sequence specific interactions between the sandwiched DNA and the Whirly protein. This nucleic acid binding mechanism contrasts with those of MRP1/2 (Figures 35A and 35B), two trypanosomatid proteins that also contain a Whirly domain but that share

essentially no sequence similarity with plant Whirlies (Schumacher et al., 2006). Indeed, MRP1/2 recognize RNA mainly through electrostatic contacts whereas StWhy2 recognize ssDNA mainly through hydrophobic/aromatic interactions. These differences outline the versatile nature of the Whirly fold and its adaptation in the context of plant organelles.

Interestingly, StWhy2 shares a number of common characteristics with bacterial SSB proteins (Meyer and Laine, 1990). They are both tetrameric proteins that bind ssDNA with high affinity and little sequence specificity; they can destabilize a DNA duplex, prevent degradation of ssDNA by nucleases and they both possess a C-terminal acidic/aromatic tail. However, the structure of the DNA-binding domain, along with the mechanism of ssDNA recognition, differs between Whirly proteins and bacterial SSB proteins. Indeed, Whirly proteins rely less on the binding of the DNA by a single domain than on the binding of the DNA between properly positioned domains. As a consequence, binding to ssDNA exploits but also depends on the C4 symmetry of the Whirly proteins. The rmsd of 0.5 Å for all C α atoms between the free and the DNA-bound forms of StWhy2 suggests that plant Whirlies act as a preformed scaffold for ssDNA binding. The requirement for properly positioned amino acids could also explain the strong structure conservation between StWhy1 (PDB 1L3A) and StWhy2 with a root-mean-square deviation (rmsd) of 1.4 Å for 159 matched C α atoms.

In addition to the strict conservation of the Whirly fold between plastid- and mitochondria-targeted Whirlies, the sequence alignment of StWhy1 with StWhy2 reveals that most of the ssDNA-binding residues are conserved between these two proteins (Figure 30A). Structural superposition of StWhy1 with StWhy2 also indicates that key ssDNA-binding residues are properly positioned in StWhy1 to contact ssDNA, thus supporting the hypothesis that plastid- and mitochondria-targeted Whirlies have closely related ssDNA-binding interfaces. Importantly, ssDNA-binding residues were also conserved between StWhy2 and AtWhy1, AtWhy2 and AtWhy3 (Figure 30A), in agreement with our observation that Whirly proteins fulfill similar functions in mitochondria and in plastids.

Structural comparison between StWhy1 and StWhy2 further revealed that the C-terminus in StWhy1 and the ssDNA in StWhy2 both interact with the same residues in the β -sheets (Figures 30A, 30B and 30C). Specifically, two protein-phosphate interactions present in the StWhy2-ssDNA structure (Figure 30B) are mimicked by the interactions between E271 and H174 and between the backbone of E271 and K153 in the StWhy1 structure (Figure 30C). This demonstrates that the C-terminus and the ssDNA can compete for the same binding surface on the Whirly proteins and that their interaction with the Whirly domain is mutually exclusive. The competition between ssDNA and an acidic/aromatic C-terminal tail is consistent with experiments showing that a StWhy1 protein mutated at its C-terminus has an increased affinity for ssDNA (Desveaux et al., 2005). A similar competition has been observed in numerous SSB proteins (Williams et al., 1983) and supports a model where binding of SSB proteins to ssDNA would render the acidic/aromatic C-terminus available for interacting with target proteins (Shereda et al., 2008). Recent structural findings even established how the C-terminus of SSB proteins is accommodated on conserved basic/aromatic patches on proteins involved in DNA metabolism (Lu and Keck, 2008; Shereda et al., 2009). Our results are consistent with such an interpretation where, upon binding of Whirly proteins to ssDNA, the acidic-aromatic C-terminal tails would become available for interacting with target proteins.

4.5.3 Possible mechanisms for Whirly-mediated repair of DSB

Whirlies could maintain genome stability by favoring the repair of DSB by error-free homologous recombination and/or by restricting the DSB repair by error-prone MMBIR. The ssDNA-binding characteristics of Whirly proteins enable them to bind ssDNA throughout the organelle genomes, in agreement with *in vivo* experiments (Maréchal et al., 2008; Maréchal et al., In press; Prikryl et al., 2008). By doing so, Whirlies could compete with repair proteins that bind ssDNA and regulate their access to ssDNA, thereby influencing the choice of a repair pathway. The protection of ssDNA against Mung Bean nuclease provides conceptual grounds for a competition between proteins that bind ssDNA. By binding to resected DNA at the breakage sites, Whirlies could also promote the

accurate repair of DSB by recruiting proteins involved in homologous recombination through their acidic/aromatic C-terminus in an SSB-like fashion. Finally, we cannot exclude that Whirly proteins could indirectly affect the balance between homologous recombination and microhomology-dependent repair by modifying the topology of DNA. Indeed, the binding of Whirlies to multiple DNA sequences in the genome could induce torsional stress in the DNA. In turn, this could influence DNA-related processes such as transcription, replication, repair and recombination. However, as processes such as transcription and replication appear to be unaffected in the absence of Whirly proteins (Maréchal et al., 2008; Prikryl et al., 2008), it seems unlikely that changes in DNA torsion could explain all the phenotypes that we observed.

Our work thus defined Whirly proteins as components of the repair machinery in plant organelles. Interestingly, TIF1, a distant homolog of the Whirlies (Saha et al., 2001), safeguards chromosomes from DNA-damage in both nuclei of the ciliate *Tetrahymena thermophyla*. TIF1-deficient cells are hypersensitive to genotoxic stress and are defective in the activation of the intra-S phase checkpoint mediated by the sensor/transducer kinase ATR (Yakisich et al., 2006). These studies suggest that the role of Whirlies in the maintenance of genome stability could be conserved between plants and ciliates. Further work should allow the identification of the players involved in plant organelle repair pathways as well as the reconstruction of the sequence of events leading to DNA repair.

4.6 Experimental Procedures

4.6.1 Plant material

Sterile seeds of *A. thaliana* ecotype Columbia (WT), *A. thaliana* lacking both AtWhy1 and AtWhy3 (KO1/3) and *A. thaliana* lacking AtWhy2 (KO2) were germinated on Murashige and Skoog basal media (Sigma) containing ciprofloxacin or novobiocin at the indicated concentration and collected after 3 weeks of growth under long day conditions (16 hours light / 8 hours dark). KO1/3 and KO2 lines were described previously (Maréchal et al.,

2008; Maréchal *et al.*, in press). Similar results were obtained with an additional *A. thaliana* line lacking *AtWhy2* (KO2b; SALK_016156).

4.6.2 Detection of DNA rearrangements

DNA was isolated from plants using a cetyl trimethylammonium bromide DNA extraction protocol (Weigel and Glazebrook, 2002). PCR reactions were conducted using the Taq polymerase (Genscript) according to the manufacturer's instructions. DNA rearrangement events were detected using both outward- and inward- facing PCR primers spaced by approximately 5-30 kb. Thirteen and 10 PCR primer pairs scattered in the plastid genome were used for ciprofloxacin- and novobiocin-treated plants, respectively. Twenty and 10 PCR primer pairs scattered in the mitochondrial genome were used for ciprofloxacin- and novobiocin-treated plants, respectively. The sequences of the PCR primers are listed in Table S2. PCR reactions were performed on DNA samples from WT, KO2 and KO1/3 plants and analyzed by gel electrophoresis. All visible DNA bands were isolated and cloned into the pDrive vector (Qiagen). DH5 α strain was used for transformation and the plasmids from two randomly picked colonies were sequenced. Sequencing of visible bands confirmed that more than 96 % of all clonable PCR products were due to DNA rearrangements, the remaining events being caused by non-specific annealing of the oligonucleotides at undesired locations. DNA rearrangements are given in Table XI.

4.6.3 StWhy2 cloning, expression and purification

Details concerning cloning, expression and purification of StWhy2 are reported elsewhere (Cappadocia *et al.*, 2008). Briefly, StWhy248-216 was amplified from *Solanum tuberosum* total RNA and cloned into the plasmid vector pET-21a (Novagen) which encodes a C-terminal hexahistidine tag. StWhy2 was expressed in *E. coli* BL21 (DE3) strain with 1 mM isopropyl β -D-thiogalactopyranoside for 3 hours at 37°C. The cells were lysed by alumina grinding and the lysate resuspended in 20 mM sodium phosphate buffer pH 7.5 containing 500 mM NaCl and 25 mM imidazole. The recombinant protein was purified by applying

the supernatant onto a HiTrap Chelating column (GE Healthcare) followed by size-exclusion chromatography on a Superdex 200 16/60 (GE Healthcare) in 20 mM Tris-HCl pH 8.0, 100 mM NaCl.

4.6.4 Data collection and structure determination

StWhy2 in the free form and StWhy2-ssDNA complexes were crystallized as previously described (Cappadocia et al., 2008). Briefly, protein-DNA complexes were crystallized by the hanging drop vapor diffusion method using a precipitant solution containing 100 mM Tris/HCl pH 8.0, 13.5–25 % (w/v) PEG6000 and 1–2 M LiCl. The free protein was crystallized by the hanging drop vapor diffusion method using a precipitant solution containing 100 mM MOPS pH 7.0, 21–27 % (w/v) PEG1000 and 100–400 mM NH₄H₂PO₄. Diffraction data were collected using ADSC Quantum 315 CCD detectors at beamlines X25 and X29 at the National Synchrotron Light Source (NSLS) at the Brookhaven National Laboratory (BNL, USA) or using a MAR-225 CCD detector at beamline 22-ID at the Advanced Photon Source (APS) at the Argonne National Laboratory (ANL, USA). Initial phases for the StWhy2-ERE32 complex were obtained by molecular replacement using StWhy1 (PDB 1L3A) as a search template. Phases were improved by iterative cycles of model building with Coot (Emsley and Cowtan, 2004) and refinement with CNS (Brunger et al., 1998) and Phenix (Adams et al., 2002). DNA nucleotides were modeled one at a time according to the FoFc electron density map. The refined structure of StWhy2-ERE32 was used as a molecular replacement search template for the other datasets. Clear density was visible for protein residues 55-215. Since we could not attribute the specific sequence of the ERE32, rcERE32 and cERE32 oligonucleotides to the electron density, we assigned these oligonucleotides as all-thymine, all-adenine and all-adenine oligonucleotides, respectively. The occupancy of the ssDNA was set at 0.8 since each tetramer effectively binds a 32-nucleotide ssDNA whereas the binding site on a single tetramer can accommodate a 40-nucleotide ssDNA. Test data sets were randomly selected from the observed reflections prior to refinement. The figures were prepared with PyMOL (DeLano, 2002).

4.6.5 Electrophoretic mobility shift assays

The apparent dissociation constants for StWhy2 bound to DNA were determined by EMSA. The ssDNA oligonucleotides were obtained from Integrated DNA Technologies (Coralville, Iowa, USA) and radiolabeled at the 5' end using [γ - 32 P]ATP (6,000 Ci/mmol, PerkinElmer Life Science) and T4 polynucleotide kinase (Fermentas) following manufacturer instructions. Binding reactions were done for 30 min at 20°C by incubating 15,000 cpm radiolabeled ssDNA (less than 100 pM) with the serially diluted StWhy2 protein in a buffer containing 10 mM Tris-HCl pH 8.0, 150 mM KCl, 0.5 mM EDTA pH 8.0 and 15 % (v/v) glycerol. Reactions were loaded on 10% (w/w) 29:1 acrylamide:bisacrylamide gels with 4.5 mM Tris, 4.5 mM boric acid and 1 mM EDTA pH 8.0 at 20°C under 150 V constant voltage for 20 min and then subjected to autoradiography. The autoradiograms were scanned and the intensity of the bands corresponding to bound and free radiolabeled ssDNA were quantified using ImageJ (NIH, <http://rsb.info.nih.gov/ij/>). The data was plotted and analyzed using Prism 5 Demo (GraphPad Software). The apparent dissociation constants were calculated by fitting the plot of the fraction of ssDNA bound versus protein concentration to the Hill equation. Each gel was done in duplicate. n represents the number of independent experiments.

4.6.6 Accession Numbers

The atomic coordinates of StWhy2 in the free form, StWhy2-ERE32, StWhy2-dT32, StWhy2-cERE32 and StWhy2-rcERE32, have been deposited in the Protein Data Bank (www.rcsb.org) under ID codes VVVV, WWWW, XXXX, YYYY and ZZZZ, respectively.

4.7 Acknowledgments

The assistance of Mathieu Coinçon, Sandra Grondin and Bruno Piché is gratefully acknowledged. Thanks to Dr. James Omichinski for valuable discussions. We also thank

the High Throughput Crystallization Service at the Hauptman-Woodward Institute, Buffalo for determining the initial crystallization conditions. Research carried out at the NSLS and at the APS was supported by the U.S. Department of Energy, Division of Materials Sciences and Division of Chemical Sciences and by the United States Department of Energy, Office of Science, respectively. Assistance by X25, X29 and 22ID beamline personnel is gratefully appreciated. L.C. and A.M. were supported by fellowships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Fonds québécois de la recherche sur la nature et les technologies (FQRNT). This research was supported by a grant from NSERC and FQRNT to both N.B. and J.S. and from the Canadian Institutes of Health Research to J.S.

4.8 References

- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58, 1948-1954.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54, 905-921.
- Cannistraro, V.J., and Taylor, J.S. (2007). Ability of polymerase eta and T7 DNA polymerase to bypass bulge structures. *J Biol Chem* 282, 11188-11196.
- Cappadocia, L., Sygusch, J., and Brisson, N. (2008). Purification, crystallization and preliminary X-ray diffraction analysis of the Whirly domain of StWhy2 in complex with single-stranded DNA. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64, 1056-1059.
- Cox, M.M. (2007). Regulation of bacterial RecA protein function. *Crit Rev Biochem Mol Biol* 42, 41-63.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35, W375-383.

- DeLano, W.L. (2002). The Pymol Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.
- Desveaux, D., Allard, J., Brisson, N., and Sygusch, J. (2002). A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nat Struct Biol* 9, 512-517.
- Desveaux, D., Després, C., Joyeux, A., Subramaniam, R., and Brisson, N. (2000). PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. *Plant Cell* 12, 1477-1489.
- Desveaux, D., Maréchal, A., and Brisson, N. (2005). Whirly transcription factors: defense gene regulation and beyond. *Trends Plant Sci* 10, 95-102.
- Desveaux, D., Subramaniam, R., Despres, C., Mess, J.N., Lévesque, C., Fobert, P.R., Dangl, J.L., and Brisson, N. (2004). A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis*. *Dev Cell* 6, 229-240.
- Drlica, K., and Zhao, X. (1997). DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol Mol Biol Rev* 61, 377-392.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Fauron, C., Allen, J., Clifton, S., and Newton, K. (2004). Plant mitochondrial genomes. In *Molecular Biology and Biotechnology of Plant Organelles*, H. Daniell, and C. Chase, eds. (Dordrecht: Springer), pp. 151-177.
- Fisher, L.M., Mizuuchi, K., O'Dea, M.H., Ohmori, H., and Gellert, M. (1981). Site-specific interaction of DNA gyrase with DNA. *Proc Natl Acad Sci U S A* 78, 4165-4169.
- Gmuender, H., Kuratli, K., K, D.P., Gray, C.P., Keck, W., and Evers, S. (2001). Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. *Genome Res* 11, 28-42.
- Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31, 3320-3323.
- Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 5, e1000327.
- Horvath, M.P. (2008). Single-stranded Nucleic Acid-binding Proteins. In *Protein-Nucleic Acid Interactions: Structural Biology*, P.A. Rice, and C.C. Correll, eds. (London: Royal Society of Chemistry), pp. 91-128.

Kanno, A., Watanabe, N., Nakamura, I., and Hirai, A. (1993). Variations in chloroplast dna from rice (*Oryza-sativa*) - differences between deletions mediated by short direct-repeat sequences within a single-species. In *Theor Appl Genet*, pp. 579-584.

Kohl, S., and Bock, R. (2009). Transposition of a bacterial insertion sequence in chloroplasts. *Plant J*.

Krause, K., Kilbiński, I., Mulisch, M., Rödiger, A., Schäfer, A., and Krupinska, K. (2005). DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS Lett* 579, 3707-3712.

Lu, D., and Keck, J.L. (2008). Structural basis of *Escherichia coli* single-stranded DNA-binding protein stimulation of exonuclease I. *Proc Natl Acad Sci USA* 105, 9169-9174.

Lupski, J.R. (2009). Genomic disorders ten years on. *Genome Med* 1, 42.

Maréchal, A., Parent, J.S., Sabar, M., Véronneau-Lafortune, F., Abou-Rached, C., and Brisson, N. (2008). Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol* 8, 42.

Maréchal, A., Parent, J.S., Véronneau-Lafortune, F., Joyeux, A., Lang, B.F., and Brisson, N. (In press). Whirly Proteins Maintain Plastid Genome Stability in *Arabidopsis*. *Proc Natl Acad Sci USA*.

Marvo, S.L., King, S.R., and Jaskunas, S.R. (1983). Role of short regions of homology in intermolecular illegitimate recombination events. *Proc Natl Acad Sci USA* 80, 2452-2456.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24, 529-538.

Meyer, R.R., and Laine, P.S. (1990). The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev* 54, 342-380.

Moeykens, C.A., Mackenzie, S.A., and Shoemaker, R.C. (1995). Mitochondrial genome diversity in soybean: repeats and rearrangements. *Plant Mol Biol* 29, 245-254.

Nöllmann, M., Crisona, N.J., and Arimondo, P.B. (2007). Thirty years of *Escherichia coli* DNA gyrase: from in vivo function to single-molecule mechanism. *Biochimie* 89, 490-499.

Odom, O.W., Baek, K.H., Dani, R.N., and Herrin, D.L. (2008). *Chlamydomonas* chloroplasts can use short dispersed repeats and multiple pathways to repair a double-strand break in the genome. *Plant J* 53, 842-853.

- Ogihara, Y., Terachi, T., and Sasakuma, T. (1988). Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc Natl Acad Sci USA* 85, 8573-8577.
- Ponder, R.G., Fonville, N.C., and Rosenberg, S.M. (2005). A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol Cell* 19, 791-804.
- Porter, R.D., and Black, S. (1991). The single-stranded-DNA-binding protein encoded by the *Escherichia coli* F factor can complement a deletion of the chromosomal *ssb* gene. *J Bacteriol* 173, 2720-2723.
- Prikryl, J., Watkins, K.P., Friso, G., Wijk, K.J., and Barkan, A. (2008). A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res.*
- Ravi, V., Khurana, J.P., Takagi, A.K., and Khurana, P. (2008). An update on chloroplast genomes. *Plant Systematics and Evolution* 271, 101-122.
- Saha, S., Nicholson, A., and Kapler, G.M. (2001). Cloning and biochemical analysis of the tetrahymena origin binding protein TIF1: competitive DNA binding in vitro and in vivo to critical rDNA replication determinants. *J Biol Chem* 276, 45417-45426.
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res* 6, 283-290.
- Schumacher, M.A., Karamooz, E., Ziková, A., Trantírek, L., and Lukes, J. (2006). Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell* 126, 701-711.
- Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M., and Keck, J.L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol* 43, 289-318.
- Shereda, R.D., Reiter, N.J., Butcher, S.E., and Keck, J.L. (2009). Identification of the SSB Binding Site on *E. coli* RecQ Reveals a Conserved Surface for Binding SSB's C Terminus. *J Mol Biol.*
- Symington, L.S. (2002). Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol Mol Biol Rev* 66, 630-670, table of contents.
- Unsold, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15, 57-61.

Vermel, M., Guermann, B., Delage, L., Grienenberger, J.M., Maréchal-Drouard, L., and Gualberto, J.M. (2002). A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proc Natl Acad Sci USA* 99, 5866-5871.

Wall, M.K., Mitchenall, L.A., and Maxwell, A. (2004). *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci USA* 101, 7821-7826.

Wang, H., Perrault, A.R., Takeda, Y., Qin, W., and Iliakis, G. (2003). Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res* 31, 5377-5388.

Weigel, D., and Glazebrook, J. (2002). *Arabidopsis: A Laboratory Manual* (Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory Press).

Williams, K.R., Spicer, E.K., LoPresti, M.B., Guggenheimer, R.A., and Chase, J.W. (1983). Limited proteolysis studies on the *Escherichia coli* single-stranded DNA binding protein. Evidence for a functionally homologous domain in both the *Escherichia coli* and T4 DNA binding proteins. *J Biol Chem* 258, 3346-3355.

Yakisich, J.S., Sandoval, P.Y., Morrison, T.L., and Kapler, G.M. (2006). TIF1 activates the intra-S-phase checkpoint response in the diploid micronucleus and amitotic polyploid macronucleus of *Tetrahymena*. *Mol Biol Cell* 17, 5185-5197.

Yoo, H.H., Kwon, C., Lee, M.M., and Chung, I.K. (2007). Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in *Arabidopsis*. *Plant J* 49, 442-451.

4.9 Figures Legends

Figure 27. DNA Rearrangements Accumulate in Organelles Following Treatment with the Gyrase-Inhibitor Ciprofloxacin

(A) PCR strategy to detect DNA rearrangements. Outward-facing PCR primers are used to monitor DNA duplication/circularization events whereas inward-facing PCR primers detect deletions. In both cases, PCR amplification only occurs if a DNA rearrangement brings together the annealing sites of the primers.

(B) Electrophoretic analysis of representative PCR reactions performed with outward- or inward-facing plastid genome-specific PCR primers on DNA from WT and KO1/3 plants grown on solid media containing the indicated concentrations of ciprofloxacin (CIP). The

oligonucleotides used for each PCR reaction are indicated. Individual bands were cut from the gel, cloned and sequenced. DNA rearrangements are listed in Table XI. Amplification of the plastid *ycf2* gene was used as a loading control.

(C) Electrophoretic analysis of representative PCR reactions performed with outward- or inward-facing plastid genome-specific PCR primers on DNA from WT and KO1/3 plants grown on solid media containing the indicated concentrations of novobiocin (NOV). Amplification of the plastid *ycf2* gene was used as a loading control.

(D) Electrophoretic analysis of representative PCR reactions performed with outward- or inward-facing mitochondria genome-specific PCR primers on DNA from WT and KO2 plants grown on solid media containing the indicated concentrations of ciprofloxacin (CIP). Amplification of the mitochondrial *cox1* gene was used as a loading control.

(E) Electrophoretic analysis of representative PCR reactions performed with outward- or inward-facing mitochondria genome-specific PCR primers on DNA from WT and KO2 plants grown on solid media containing the indicated concentrations of novobiocin (NOV). Amplification of the mitochondrial *cox1* gene was used as a loading control.

Figure 28. Crystal Structures of StWhy2 in the Free Form and Bound to ERE32 at 2.20-Å Resolution

(A) Schematic representation of StWhy2. Filled boxes indicate the position of the mitochondria-targeting sequence (MITO), the Whirly domain and the acidic/aromatic C-terminal tail (CT). A green dotted line represents the construct used for structure determination.

(B) Overall view of StWhy2 in the free form in cartoon representation. The tetramer was generated by applying the crystallographic symmetry along the four-fold axis. Protomers are colored in yellow, orange, pink and green.

(C) Surface representation of the protein moiety in the StWhy2-ERE32 complex. Difference electron density was calculated from a FoFc simulated annealing omit map encompassing DNA, contoured at 1.6 Å and colored in blue. The tetramer was generated by applying the crystallographic symmetry along the four-fold axis.

(D) Close-up view of the protein-DNA interactions in the StWhy2-ERE32 complex. The DNA and the DNA-interacting residues are in stick representation with carbon atoms colored in yellow and in gray, respectively. Nucleotides and protein residues are labeled. *, residues that contact DNA through their main-chain. +, residue that contact the DNA through both its main-chain and its side-chain. Hydrogen bonds are represented as yellow dashed lines. A water molecule is represented as a red sphere.

Figure 29. StWhy2 Binds ssDNA with Little Sequence Specificity

(A) Representative EMSA results showing the binding of StWhy2 to four different ssDNA sequences. Increasing amounts of StWhy2 were incubated with target oligonucleotides ERE32, dT32, rcERE32 or cERE32 and the complexes resolved on a 10 % (w/v) polyacrylamide gel.

(B) Sequences of the oligonucleotides ERE32, dT32, rcERE32 and cERE32 and apparent dissociation constant for StWhy2 ± standard error of the mean (SEM), as measured by EMSA. n represents the number of independent experiments.

(C) Crystal structures of four different ssDNA sequences bound by StWhy2. DNA molecules are presented as stick models with carbon atoms colored in yellow. The 2FoFc electron density maps corresponding to DNA is contoured at 1 Å and colored in gray. DNA molecules are presented in the same order as for Figure 2A.

Figure 30. Comparison between Mitochondria- and Plastid-Targeted Whirlies

(A) Sequence alignment of the Whirly domain of Whirly proteins from *Arabidopsis thaliana* (AtWhy1, AtWhy2 and AtWhy3) and *Solanum tuberosum* (StWhy1 and StWhy2)

with schematic secondary structural elements from StWhy2. Similarity above 70% is depicted in yellow whereas perfect conservation is depicted in red. A green star indicates residues of StWhy1 that interact with the C-terminal acidic-aromatic tail in the StWhy1 structure (PDB 1L3A) whereas a blue star indicates residues of StWhy2 that interact with ssDNA. A bi-color star indicates residues that can interact with both the C-terminus and the ssDNA. The figure was prepared using ESPript (Gouet et al., 2003).

(B) Cartoon representing part of the ERE32-binding site in the StWhy2 structure. The representation is similar as in Figure 28D.

(C) Cartoon representation of the binding site of the acidic/aromatic C-terminal tail in the StWhy1 structure (PDB 1L3A). The orientation is similar as that presented in Figure 30B.

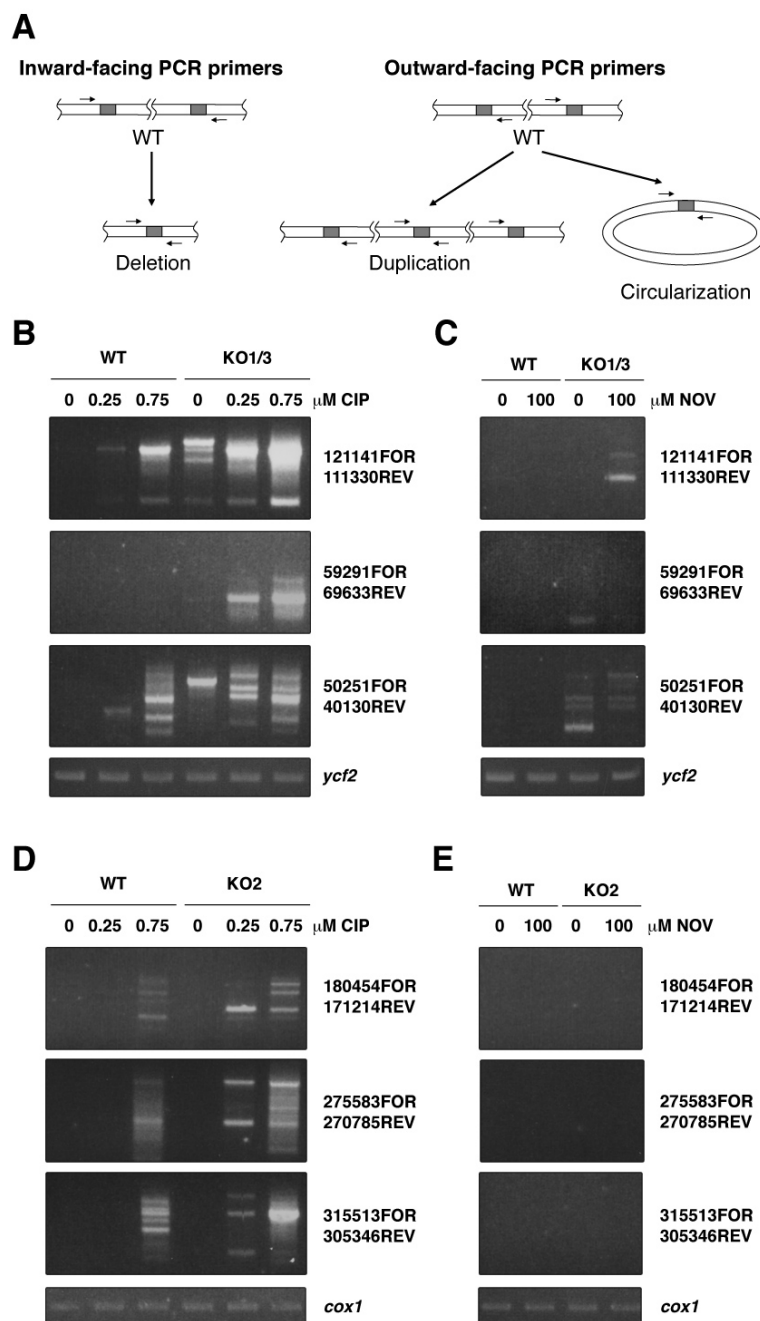


Figure 27. DNA Rearrangements Accumulate in Organelles Following Treatment with the Gyrase-Inhibitor Ciprofloxacin.

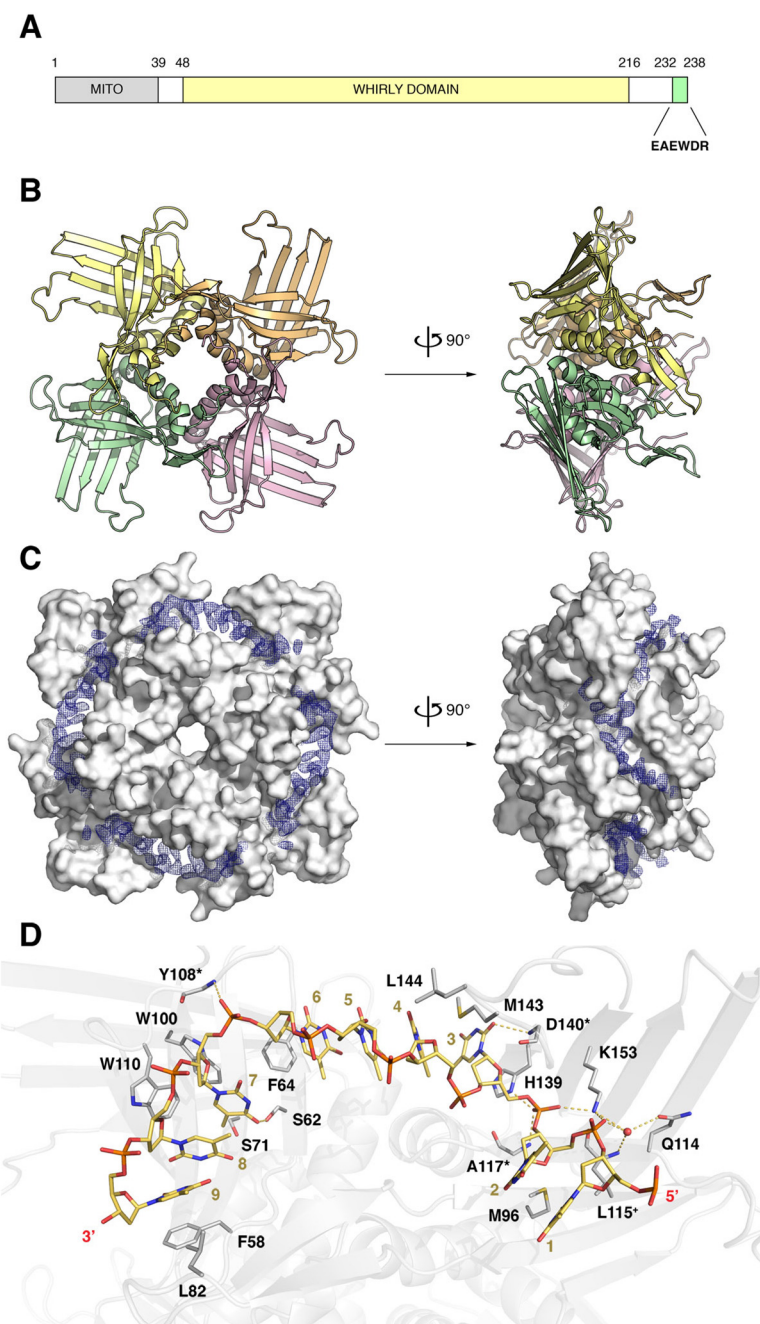


Figure 28. Crystal Structures of StWhy2 in the Free Form and Bound to ERE32 at 2.20-Å Resolution.

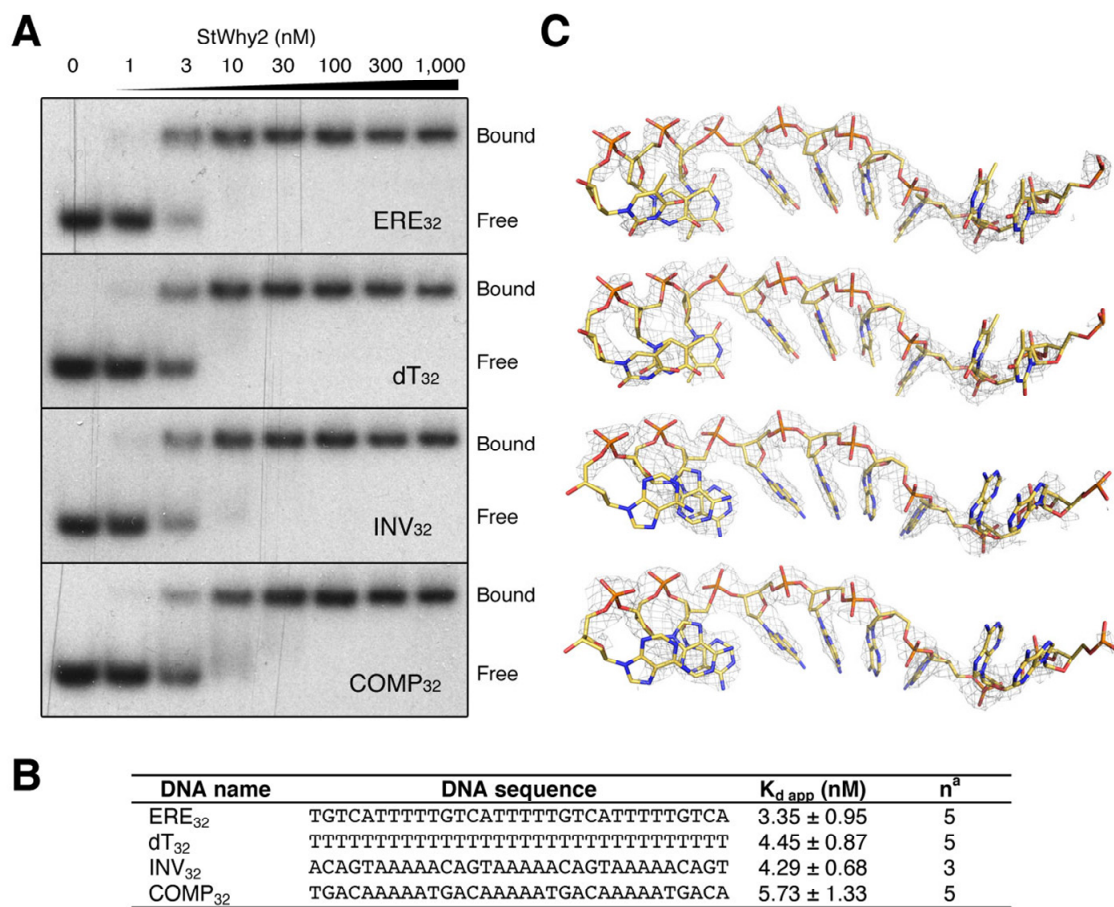


Figure 29. StWhy2 Binds ssDNA with Little Sequence Specificity.

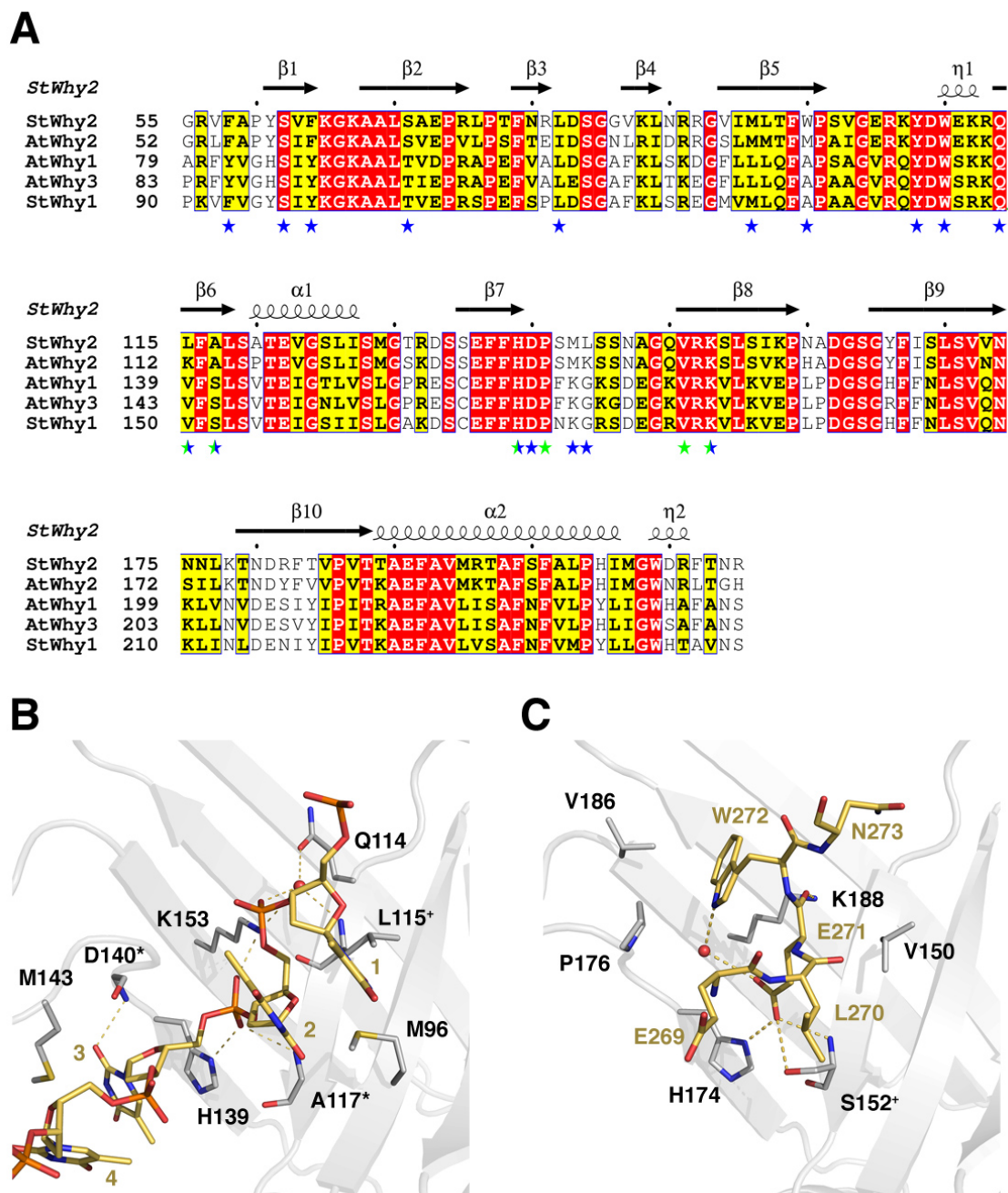


Figure 30. Comparison between Mitochondria- and Plastid-Targeted Whirlies.

Tableau VIII. DNA Rearrangement Statistics.

Treatment	Type of DNA Rearrangement	Plastid Events ^a		Mitochondrial Events ^b	
		Col-0	KO1/3	Col-0	KO2
CIP 0 μ M	NHEJ-like	0	0	0	0
	MHMR (mismatches)	2 (1)	17 (13)	0 (0)	0 (0)
	Total Events	2	17	0	0
CIP 0.25 μ M	NHEJ-like	1	0	0	1
	MHMR (mismatches)	7 (5)	25 (16)	1 (0)	13 (8)
	Total Events	8	25	1	14
CIP 0.75 μ M	NHEJ-like	7	4	11	8
	MHMR (mismatches)	10 (4)	26 (15)	10 (7)	25 (15)
	Total Events	17	30	21	33
NOV 0 μ M	NHEJ-like	0	1	0	0
	MHMR (mismatches)	0 (0)	11 (4)	0 (0)	0 (0)
	Total Events	0	12	0	0
NOV 100 μ M	NHEJ-like	0	0	0	0
	MHMR (mismatches)	1 (0)	8 (3)	0 (0)	0 (0)
	Total Events	1	8	0	0

NHEJ-like are DNA junctions with four or less homologous base pairs. MHMR are DNA junctions with five or more homologous base pair. Values in parentheses designate the number of MHMR events that contain at least one mismatch. CIP relates to plant treated with ciprofloxacin whereas NOV relates to plants treated with novobiocin. DNA rearrangements containing multiple DNA junctions were counted as single events.

^a Plastid events are based on 13 primer pairs for ciprofloxacin treatment and 10 primer pairs for novobiocin treatment.

^b Mitochondrial events are based on 20 primer pairs for ciprofloxacin treatment and 10 primer pairs for novobiocin treatment.

Tableau IX. Representative DNA Rearrangements Observed in *Arabidopsis* Organelle Genomes

Event	Plant	Plant	Junction	Junction Sequences ^c	Length of	Junction
-------	-------	-------	----------	---------------------------------	-----------	----------

Treatment ^a	Genotype	Endpoints ^b	Junctions (Mismatches)	Types ^d		
NHEJ-like DNA junction						
M3-2	CIP 0.75 μM	Col-0	55001 Recombinant 65518	cgaaaggagc cataccggaa tgagccgctg cataccggaa tgagccgctg tactgtggca	0 bp	NHEJ- like
MHMR DNA junction						
M3-3	CIP 0.75 μM	KO2	53421 Recombinant 65476	gctccgtcca TGGAGCGTATG atthtccttg ctctaaagct TGGAGCGTATG atthtccttg ctctaaagct TGGAGCGTATG gtaacggtta	(0) 11 bp (0)	MHMR
MHMR DNA junction containing a mismatch in one parental strand						
M6-1	CIP 0.25 μM	KO2	234779 Recombinant 246725	gatgaggatg AAGTAAAGGAATT actctggttg acgaaagtga AAGTAAAGGAATT actctggttg acgaaagtga AAGTAAAGGGATT gtctggccct	(0) 13 bp (1)	MHMR*
MHMR DNA junction containing a mismatch in both parental strands						
M7-7	CIP 0.75 μM	Col-0	304900 Recombinant 316104	tggaatgctc TCGTTGTTAAAAAAA ctacgagcgg caggggggat TCTTTGTCGATAAAA ctacgagcgg caggggggat TCTTTGTCGATAGAA ggcattcgcg	(4) 15 bp (1)	MHMR*
Same MHMR DNA junction containing different mismatches						
M3-6	CIP 0.75 μM	KO2	54234 Recombinant 65096	taggaatcct AGGTTACCTCCAG ccttctgaag ccactaggta AGGTTACCTCCAG ccttctgaag ccactaggta AGGTTACTTCCAG agcggagtat	(0) 13 bp (1)	MHMR*
M3-7	CIP 0.75 μM	KO2	54234 Recombinant 65096	taggaatcct AGGTTACCTCCAG ccttctgaag ccactaggta AGGTTACTTCCAG ccttctgaag ccactaggta AGGTTACTTCCAG agcggagtat	(1) 13 bp (0)	MHMR*
MHMR DNA junctions with homopolymeric tracts						

M12-1	CIP 0.25 μ M	Col-0	29609 Recombinant 80994	tttttcccct attagacgctc attagacgctc	CCCCCCCCCA CCCCCCCCCA CCCCCCCCCA	aaaaaaaaaa aaaaaaaaag tgctacggtt	(0) 10 bp (0)	MHMR
P10-14	CIP 0.75 μ M	KO1/3	59719 Recombinant 50843	tactttgaaa gagtggggat gagtggggat	ACAAAAAAAAA ACAAAAAAAAA ACAAAAAAAAA	ggaaagtcacat ggaaagtcacat aaagaaagta	(0) 10 bp (0)	MHMR
Molecules containing multiple MHMR								
P6-3	CIP 0 μ M	KO1/3	19786 Recombinant 110850	catatctcaa caaaattttg caaaattttg	TAGATTGGCGT-AATTTTTT TAGATTGGCGT-AATTTTTT TAGATTGG <u>TATGA</u> AATTTTTT	ctaccgcatg ctaccgcatg gagaaatgca	(0) 18 bp (3)	MHMR*
P6-3	CIP 0 μ M	KO1/3	19867 Recombinant 122003	taattatggg gccagccctt gccagccctt	AGAAGTTGATCGTTAAAATGAT AGAAGTTGATCGTTAAAATGAT AGAAG <u>GT</u> -ATCGTTAAA <u>A</u> -GAT	aatttatgca aatttatgca catcaattcc	(0) 22 bp (3)	MHMR*

A complete list of DNA rearrangements can be found in Table XI.

^a Treatment of plants with ciprofloxacin (CIP) or novobiocin (NOV) at the indicated concentrations.

^b The junction endpoint is the position of the first nucleotide preceding the DNA junction. The accession numbers for the mitochondrial and plastid genomes are NC_001284 (281) and NC_000932 (1), respectively.

^c Alignment of the sequenced DNA junction (center) with homologous sequences (top and bottom). Mismatches are underlined.

^d Product types were classified according to junction length with NHEJ-like for junction of less than five nucleotides, MHMR for junctions of five or more nucleotides and MHMR* for junctions of five or more nucleotides containing at least one mismatch.

Tableau X. Data Collection and Refinement Statistics.

Protein	StWhy2 Free Form	StWhy2 -ERE ₃₂	StWhy2 - (dT) ₃₂	StWhy2 -cERE ₃₂	StWhy2 -rcERE ₃₂
Beamline/Wavelength	NSLS- X25/1.08 Å	NSLS- X29/1.08 Å	NSLS- X25/1.08 Å	APS- 22ID/1.10 Å	APS- 22ID/1.10 Å
Space group	F432	F432	F432	F432	F432
Cell dimensions : <i>a, b, c</i> (Å)	164.58, 164.58, 164.58	167.16, 167.16, 167.16	166.51, 166.51, 166.51	166.93, 166.93, 166.93	166.67, 166.67, 166.67
Resolution (Å)	50–2.20 (2.28–2.20)	50–2.20 (2.28–2.20)	50–2.65 (2.74–2.65)	50–2.70 (2.80–2.70)	50–2.35 (2.43–2.35)
Total Reflections	380467	382748	118110	178903	253462
Unique Reflections	10208	10665	6188	5865	8716
R _{sym} (%)	7.6 (92)	5.1 (56)	11.8 (87)	9.0 (68)	8.8 (75)
I / σ I	16.9 (3.8)	23.1 (3.8)	9.1 (3.0)	12.7 (2.2)	16.5 (2.3)
Completeness (%)	100.0 (99.9)	99.8 (98.7)	100.0 (100.0)	99.6 (96.0)	99.5 (95.4)
Multiplicity	37.3 (19.5)	35.9 (9.9)	19.1 (12.7)	30.5 (5.9)	29.1 (6.5)
Refinement Statistics					
Resolution (Å)	50–2.20	50–2.20	50–2.65	50–2.70	50–2.35
Reflections (total/test)	9873/532	9843/752	6012/603	5625/628	8437/900
R _{work} / R _{free} (%)	22.45/25.00	23.03/26.57	22.82/25.63	22.34/25.97	23.86/26.63
Number of atoms					
Protein	1267	1267	1267	1267	1267
DNA/Phosphate	10	180	180	189	189
Water	69	67	60	56	60
B-factors					
Protein	49.7	54.9	53.8	60.8	62.8
DNA/Phosphate	50.3	65.3	66.0	76.6	80.5
Water	51.5	54.4	54.3	58.4	61.2
RMSDs					
Bond lengths (Å)	0.004	0.003	0.003	0.003	0.004
Bond angles (°)	0.799	0.808	0.683	0.694	0.843
Ramachandran^a					
Favored (%)	98.1	97.5	95.6	97.5	95.6
Outliers (%)	0	0	0	0	0

Values in parentheses are for highest-resolution shell.

$R_{\text{sym}} = \sum_{hkl} [\sum_i |I_{hkl,i} - \langle I_{hkl} \rangle|] / \sum_{hkl,i} \langle I_{hkl} \rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices hkl and $\langle I_{hkl} \rangle$ is the mean intensity of that reflection.

^a MolProbity analysis (282)

4.10 Supplemental Data

4.10.1 Supplemental Experimental Procedures

4.10.1.1 DNA destabilization experiment

Oligonucleotide R2 (CCA ATG ACT TGC GAG GAG CCC AAA CGT CAA TCG TAG ATG TCG GCA GCG TTG ACT AGA CGG ATT CGA TA) was radiolabeled at the 5' end using [γ - 32 P]ATP (6,000 Ci/mmol, PerkinElmer Life Science) and T4 polynucleotide kinase (Fermentas) following manufacturer instructions and annealed with the oligonucleotide R1 (TAT CGA ATC CGT CTA GTC AAC GCT GCC GAC ATC TAC GAT TGA CGT TTG GGC TCC TCG CAA GTC ATT GG). The annealed duplex was incubated with StWhy2 or BSA for 30 min at 20°C. The reactions were stopped by the addition of a 40 fold excess of unlabeled R2 and SDS to a final concentration of 0.2 %. Reactions were loaded on 7.5% (w/w) 29:1 acrylamide:bisacrylamide gels with 4.5 mM Tris, 4.5 mM boric acid and 1 mM EDTA pH 8.0 at 20°C under 130 V constant voltage for 30 min.

4.10.1.2 Mung Bean protection experiment

Two hundred nanograms of M13mp18 ssDNA (USB) either naked or prebound for 30 min at 20°C with StWhy2 at a 1:10 protein/nucleotide ratio were incubated with 1U of Mung Bean nuclease (NEB) for 0, 15, 30 or 60 min at 30°C. Reactions were stopped by incubation on ice and addition of SDS and EDTA to final concentrations of 0.1 % (w/v) and 1.5 mM, respectively. Reactions were loaded on 0.7 % (w/w) agarose gels under 100 V constant voltage.

4.10.2 Supplemental Figure Legends

Figure 31. Phenotypes of WT and Plastid Whirly Deficient Arabidopsis Plants Treated With Gyrase Inhibitors Ciprofloxacin and Novobiocin

(A) Phenotypic effects of various concentrations of gyrase inhibitors ciprofloxacin and novobiocin on WT and KO1/3 Arabidopsis plants. Plants were grown on solid media containing the indicated concentrations of ciprofloxacin or novobiocin.

(B) Etiolated leaves were quantified at each gyrase inhibitor concentration for each plant genotype. Each data point represents an average \pm standard error of the mean (SEM). ND : Etiolated leaves could not be counted at 0.75 μ M ciprofloxacin and at 100 μ M novobiocin.

Figure 32. Phenotypes of WT and Mitochondrial Whirly Deficient Arabidopsis Plants Treated With Gyrase Inhibitors Ciprofloxacin and Novobiocin

(A) Phenotypic effects of various concentrations of gyrase inhibitors ciprofloxacin and novobiocin on WT and KO2 Arabidopsis plants. Plants were grown on solid media containing the indicated concentrations of ciprofloxacin or novobiocin.

(B) Etiolated leaves were quantified at each gyrase inhibitor concentration for each plant genotype. Each data point represents an average \pm SEM. ND : Etiolated leaves could not be counted at 0.75 μ M ciprofloxacin and at 100 μ M novobiocin.

Figure 33. Compensating Interactions Enable StWhy2 to Bind DNA Nucleobases that Differ in Size and in Functional Groups

(A) Interactions between StWhy2 and the edge of T3 in the StWhy2-ERE32 structure. The tetramers were obtained by applying the crystallographic symmetry and are in cartoon representation. Selected residues are in stick representation. Hydrogen bonds are symbolized by yellow dashes. A red sphere corresponds to a water molecule.

(B) Interactions between StWhy2 and the edge of T7 in the StWhy2-ERE32 structure. Same representation as in Figure S3A.

(C) Interactions between StWhy2 and the edge of A3 in the StWhy2-rcERE32 structure. Same representation and orientation as in Figure S3A.

(D) Interactions between StWhy2 and the edge of A7 in the StWhy2-rcERE32 structure. Same representation as in Figure S3A and same orientation as in Figure S3B.

Figure 34. StWhy2 can Destabilize dsDNA and Protect ssDNA against Mung Bean Degradation

(A) StWhy2 destabilizes a DNA duplex. A R1-R2 DNA duplex of which the R2 strand was ³²P-labeled was incubated with StWhy2 or BSA for 30 min at 20°C. The reactions were stopped by the addition of a 40 fold excess of unlabeled R2 and SDS to a final concentration of 0.2 %. DNA were resolved on a 7.5 % acrylamide gel.

(B) StWhy2 protects DNA against Mung Bean nuclease degradation. M13mp18 ssDNA either naked or prebound for 30 min at 20°C with StWhy2 at a 1:10 protein/nucleotide ratio was incubated with 1U of Mung Bean nuclease at 30°C for the indicated amount of time. Reactions were stopped by incubation on ice and addition of SDS and EDTA to final concentrations of 0.1 % (w/v) and 1.5 mM, respectively. Black/white inverted image are shown. M represents the molecular weight markers.

Figure 35. StWhy2 and MRP1/2 bind nucleic acids through different mechanisms

(A) Crystal structure of StWhy2 in complex with ERE32. The StWhy2 protein is in gray cartoon whereas the ssDNA is in gold stick representation.

(B) Crystal structure of MRP1/2 in complex with a guide RNA (PDB 2GJE). The MRP1/2 proteins are in gray cartoon whereas RNA is in gold stick representation.

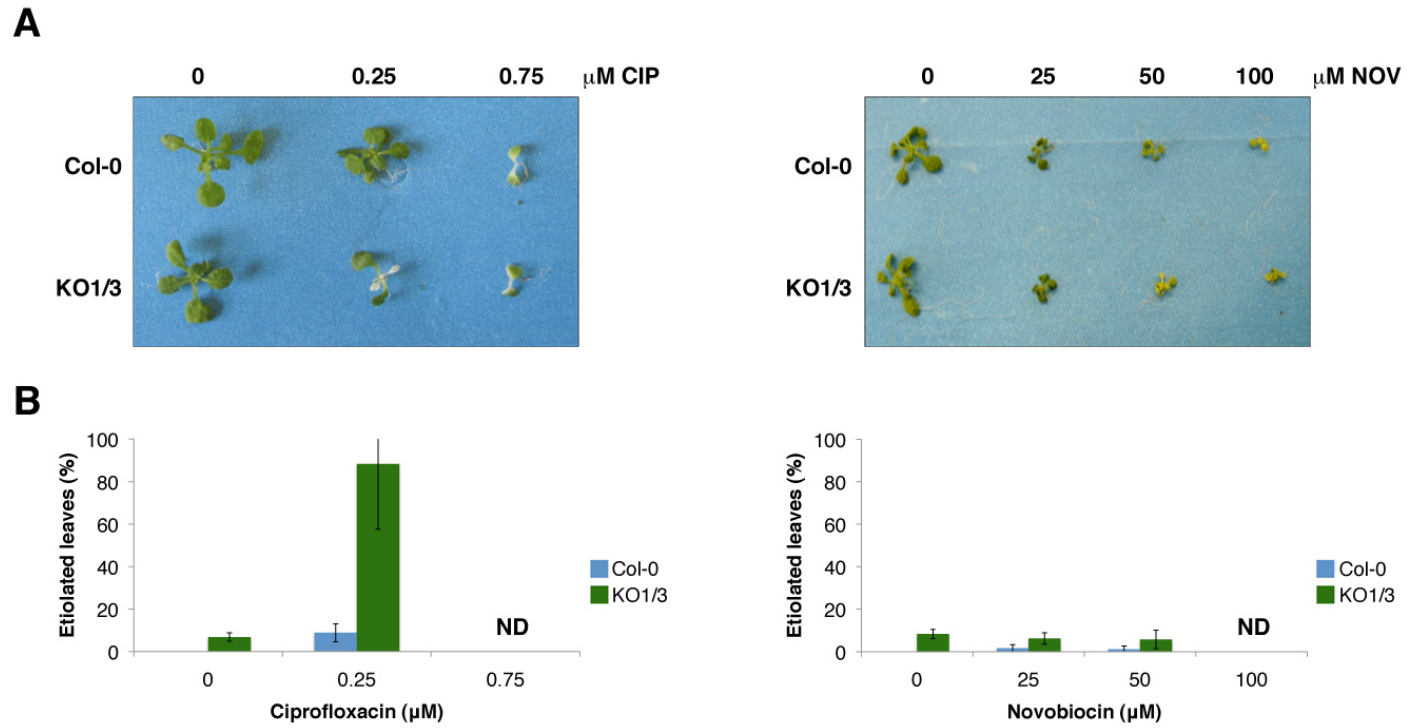


Figure 31. Phenotypes of WT and Plastid Whirly Deficient Arabidopsis Plants Treated With Gyrase Inhibitors Ciprofloxacin and Novobiocin.

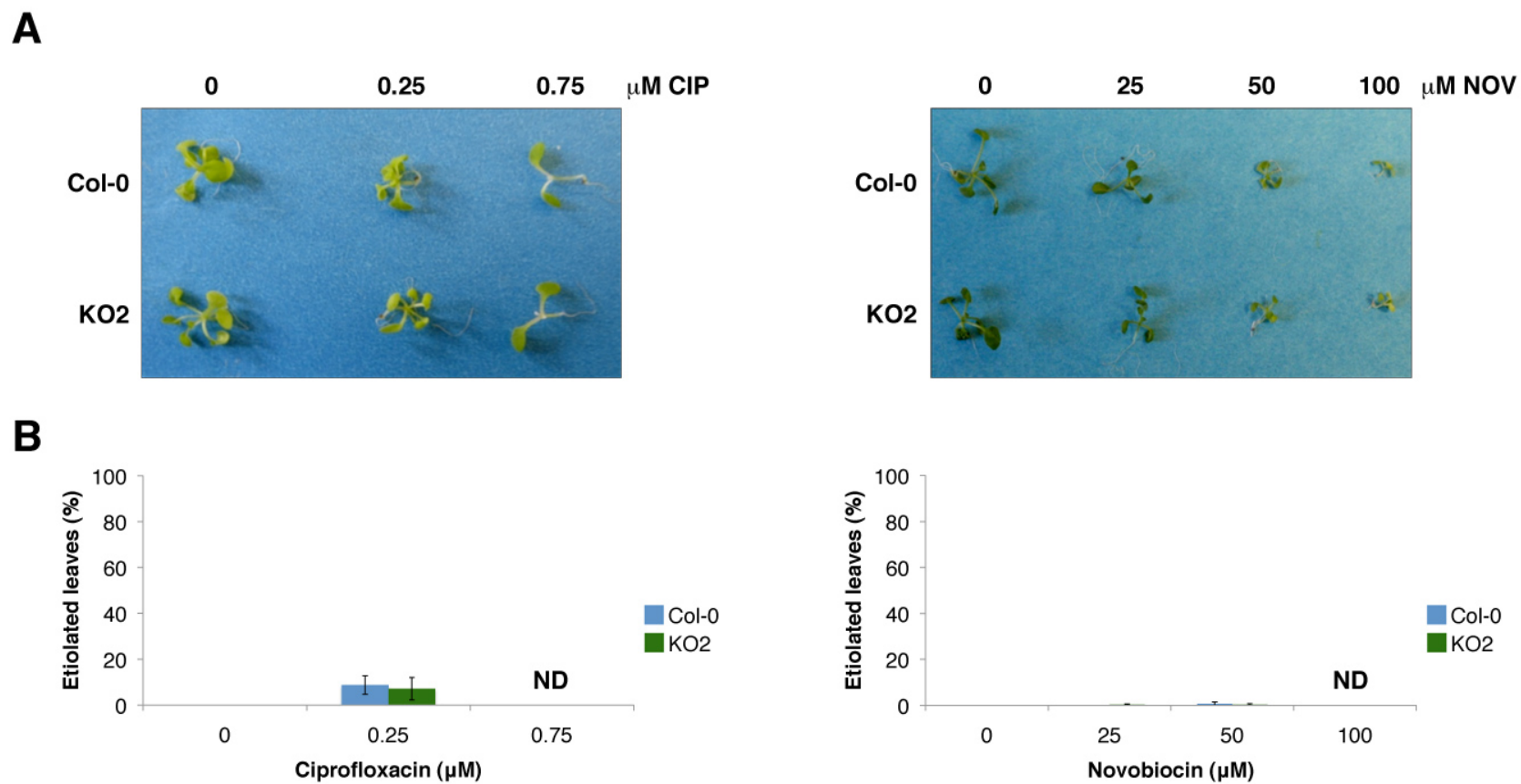


Figure 32. Phenotypes of WT and Mitochondrial Whirly Deficient Arabidopsis Plants Treated With Gyrase Inhibitors Ciprofloxacin and Novobiocin.

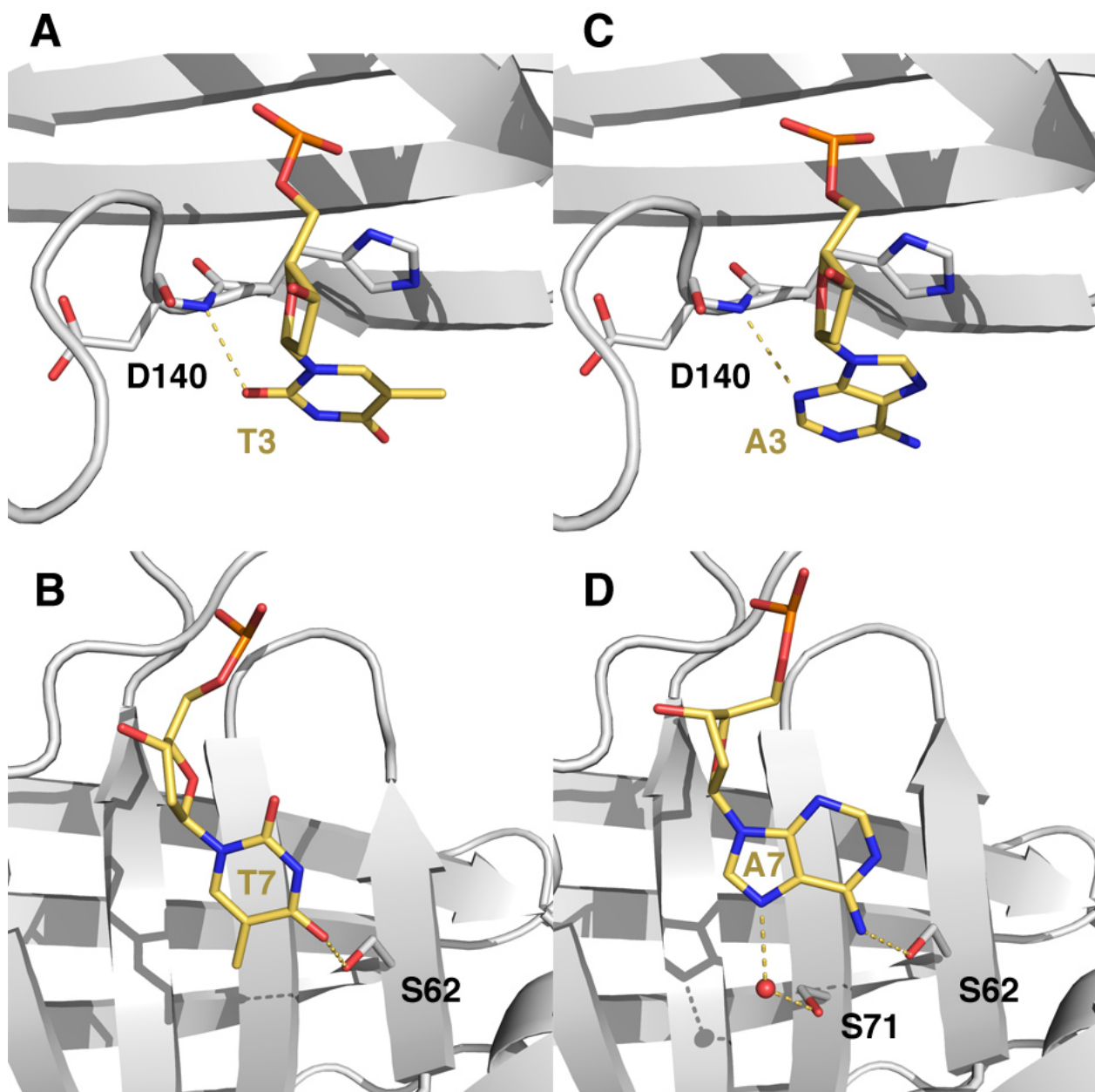


Figure 33. Compensating Interactions Enable StWhy2 to Bind DNA Nucleobases that Differ in Size and in Functional Groups.

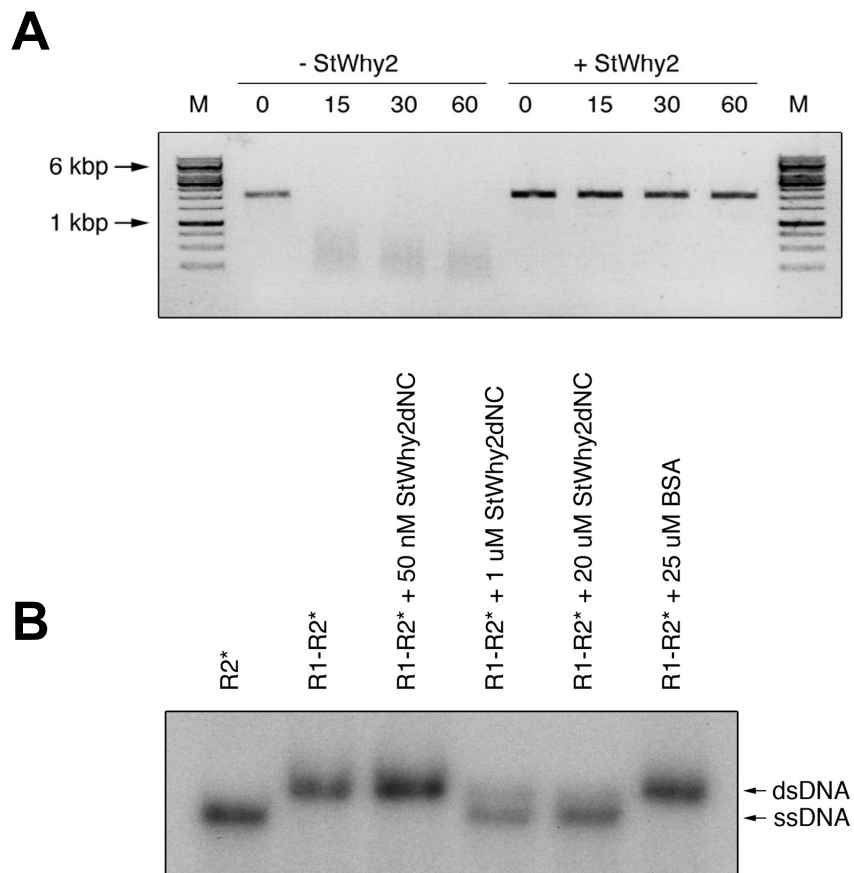


Figure 34. StWhy2 can Destabilize dsDNA and Protect ssDNA against Mung Bean Degradation.

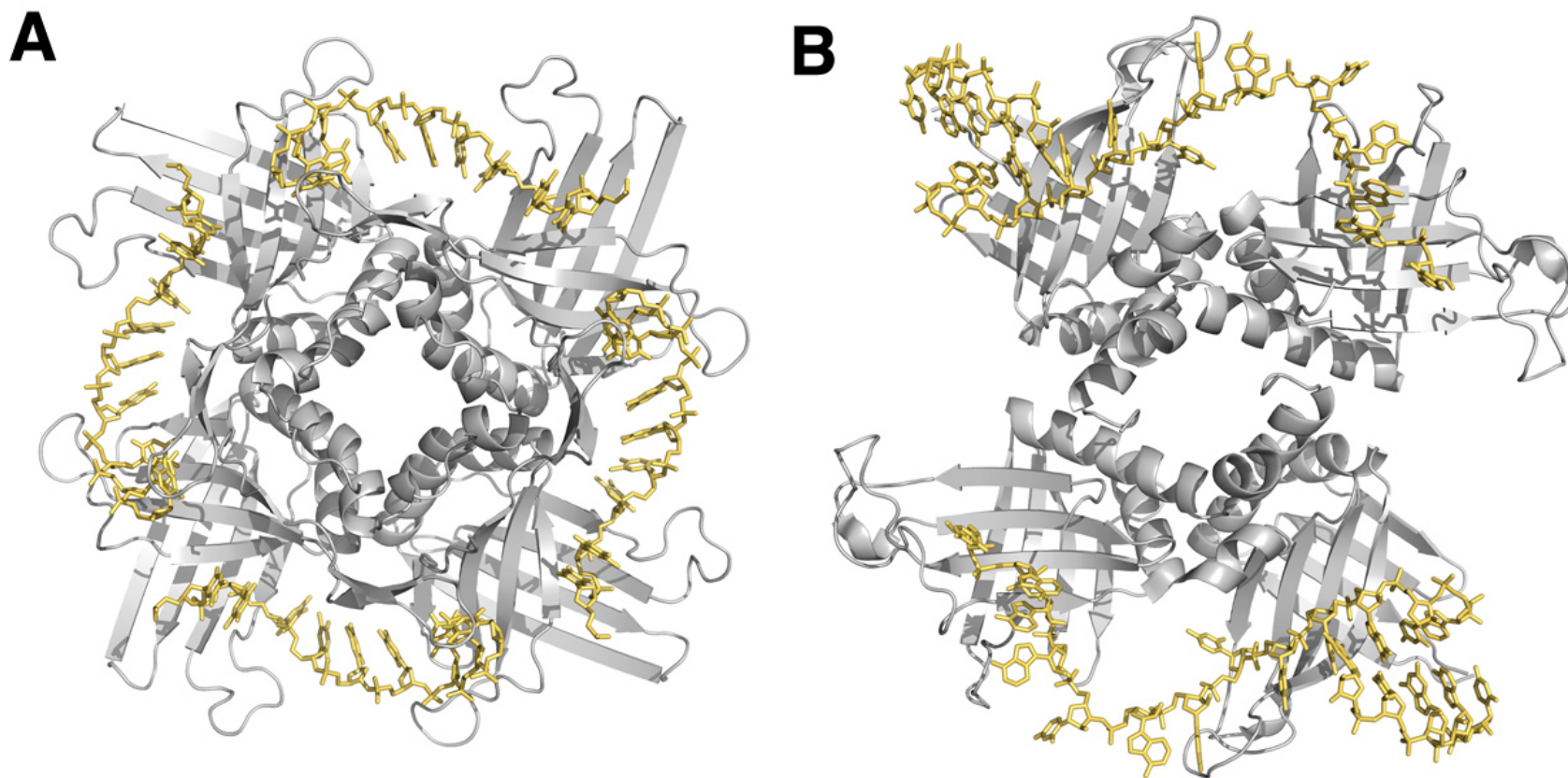


Figure 35. StWhy2 and MRP1/2 bind nucleic acids through different mechanisms.

Tableau XI. DNA Rearrangement Products Identified in the Mitochondrial and in the Plastid Genomes.

Event	Plant Treatment ^a	Plant Genotype	Junction Endpoints ^b	Junction Sequences ^c	Length of Junctions (Mismatches)	Junction Types ^d
Mitochondrial reaction 1 with primers 20041REV and 30001FOR^{e,f}						
Mitochondrial reaction 2 with primers 30001REV and 50341FOR						
Mitochondrial reaction 3 with primers 55021REV and 65041FOR						
M3-1	CIP 0.75 µM	Col-0	54739 Recombinant 65547	aaaaaaaggg G aaagaactac gtcgtgtac G aaagaactac gtcgtgtac G ttcattggctt	1 bp	NHEJ-like
M3-2	CIP 0.75 µM	Col-0	55001 Recombinant 65518	cgaaggagc cataccggaa tgagccgctg cataccggaa tgagccgctg tactgtggca	0 bp	NHEJ-like
M3-3	CIP 0.75 µM	KO2	53421 Recombinant 65476	gctccgtcca TGGAGCGTATG attttccttg ctctaaagct TGGAGCGTATG attttccttg ctctaaagct TGGAGCGTATG gtaacgggta	(0) 11 bp (0)	MHMR
M3-4	CIP 0.75 µM	KO2	54638 Recombinant 65742	ccttattagc ctaatatctt atctccttct ctaatatctt atctccttct tcctttgact	0 bp	NHEJ-like
M3-5	CIP 0.75 µM	KO2	54463 Recombinant 65476	tcgtgccttc gccttaggct ctctaaagct gccttaggct ctctaaagct tggagcgtat	0 bp	NHEJ-like
M3-6	CIP 0.75 µM	KO2	54234 Recombinant 65096	taggaatctt AGGTACCTCCAG ccttctgaag ccactaggta AGGTACCTCCAG ccttctgaag ccactaggta AGGTACTTCCAG agcggagtat	(0) 13 bp (1)	MHMR*
M3-7	CIP 0.75 µM	KO2	54234 Recombinant 65096	taggaatctt AGGTACCTCCAG ccttctgaag ccactaggta AGGTACTTCCAG ccttctgaag ccactaggta AGGTACTTCCAG agcggagtat	(1) 13 bp (0)	MHMR*
M3-8	CIP 0.75 µM	KO2	54393 Recombinant 65074	cgcgacttgg GAAACCCTT tcttcattcg ctgagctgta GAAACCCTT tcttcattcg ctgagctgta GAAACCCTT gatccactag	(0) 9 bp (0)	MHMR
M3-9	CIP 0.75 µM	KO2	54483 Recombinant 65089	tgccttcgct CTAGGCTCGGCTACTTC tttctcaaata ccttgatcca CTAGGCTCGGCTACTTC tttctcgaata ccttgatcca CTAGGTAAGGTTACTTC cagagcggagt	(0) 17 bp (4)	MHMR*
Mitochondrial reaction 4 with primers 130413REV and 151514FOR						
Mitochondrial reaction 5 with primers 180454REV and 200500FOR						
M5-1	CIP 0.75 µM	KO2	179517 Recombinant 200771	atgacggggt ACCCGG ttcaaggctt aaggtaagag ACCCGG ttcaaggctt aaggtaagag ACCCGG gaaaggcggg	(0) 6 bp (0)	MHMR
Mitochondrial reaction 6 with primers 235252REV and 245966FOR						
M6-1	CIP 0.25 µM	KO2	234779 Recombinant 246725	gatgaggatg AAGTAAAGGAATT actctggttg acgaaagtga AAGTAAAGGAATT actctggttg acgaaagtga AAGTAAAGGATT gtctggccct	(0) 13 bp (1)	MHMR*

Mitochondrial reaction 7 with primers 305346REV and 315513FOR^{ef}

M7-1	CIP 0.25 μ M	KO2	305245 Recombinant 316714	cgaagataca AGCGCCCG gtcctcttct ttaaagtcctat AGCGCCCG gtcctcttct ttaaagtcctat AGCGCCCG attcgcgatcac	(0) 8 bp (0)	MHMR
M7-2	CIP 0.25 μ M	KO2	304953 Recombinant 316326	gataaaaggta AGGTCACACT ctcactagta taacagggcgt AGATCACACT ctcactagta taacagggcgt AGATCACACT ggaactggag	(1) 10 bp (0)	MHMR*
M7-3	CIP 0.25 μ M	KO2	305304 Recombinant 315835	aggttttgcc TCACAGAAGGAGTGG gaaattgggg aaagactacg TCACTGAAG-AGTGG gaaattgggg aaagactacg TCACTGAAG-AGTGG tttgaagaac	(2) 14 bp (0)	MHMR*
M7-4	CIP 0.75 μ M	Col-0	304556 Recombinant 316456	agaaggagaa AGGCTATTAAACATAAAAAAAAAATATAA agtccaaatg atgcctctc AGGCTAATGAAAGCAAAAAAAAAATAGAA agtccaaatg atgcctctc AGGCTAATGAAAGCAAAAAAAAAAGCAGAA ggccttact	(6) 28 bp (2)	MHMR*
M7-5	CIP 0.75 μ M	Col-0	305222 Recombinant 317005	cgataactcc tatcccgaag agctaggcag tatcccgaag agctaggcag atagatagac	(0) 0 bp (0)	NHEJ- like
M7-6	CIP 0.75 μ M	Col-0	305245 Recombinant 316714	cgaagataca AGCGCCCG gtcctcttct ttaaagtcctat AGCGCCCG gtcctcttct ttaaagtcctat AGCGCCCG attcgcgatcac	(0) 8 bp (0)	MHMR
M7-7	CIP 0.75 μ M	Col-0	304900 Recombinant 316104	tggaatgctc TCGTTGTTAAAAA ctacgagcgg caggggggat TCTTTGTCGATAAAA ctacgagcgg caggggggat TCTTTGTCGATAGAA ggcattcgcg	(4) 15 bp (1)	MHMR*
M7-8	CIP 0.75 μ M	Col-0	305105 Recombinant 316064	ccagacctaa CGGCTA caagaataaa tagcaaacac CGGCTA caagaataaa tagcaaacac CGGCTA gataagtggg	(0) 6 bp (0)	MHMR
M7-9	CIP 0.75 μ M	KO2	305257 Recombinant 316731	cggtcctctt CTCT tgtgtcgaag gattcgcgatca CTCT tgtgtcgaag gattcgcgatca CTCT atcgaataga	(0) 4 bp (0)	NHEJ- like
M7-10	CIP 0.75 μ M	KO2	305245 Recombinant 316714	cgaagataca AGCGCCCG gtcctcttct ttaaagtcctat AGCGCCCG gtcctcttct ttaaagtcctat AGCGCCCG attcgcgatcaC	(0) 8 bp (0)	MHMR
M7-11	CIP 0.75 μ M	KO2	305252 Recombinant 316371	agatacaagc GCCCGTCTCT tctcttctgtgt gacgaacaat GCTCGGCACTCT tctcttctgtgt gacgaacaat GCTCGGCACTCT ggtcaagtga	(3) 12 bp (0)	MHMR*
M7-12	CIP 0.75 μ M	KO2	305132 Recombinant 315565	aaaagtacag GCAGAGCCAA acccatccta attcagatct GCAGAGCCAA acccatccta attcagatct GCAGAGACAA ctggaacaga	(0) 10 bp (1)	MHMR*

Mitochondrial reaction 8 with primers 165904REV and 200500FOR

Mitochondrial reaction 9 with primers 155034REV and 200500FOR

Mitochondrial reaction 10 with primers 10081REV and 15001FOR

M10-1	CIP 0.75 μ M	Col-0	9505 Recombinant 16023	atcttagett G agcctacctt aagtgaaggg G agcctacctt aagtgaaggg G cgctatccgg	1 bp	NHEJ- like
M10-2	CIP 0.75 μ M	KO2	9725 Recombinant 16003	ggttgctact GCGAAGCTA-TAGTGAAGGTGC acgggggtctt tcctagagta GCGAAGCTACAAGTGAAGGTGC acgggggtctt tcctagagta GCGAAACTACAAGTGAAGGGGC gctatccgga	(2) 22 bp (2)	MHMR*
M10-3	CIP 0.75 μ M	KO2	9511 Recombinant 15480	gcttgagcct A ccttcgcaca acgaagggtc A ccttcgcaca acgaagggtc A tccagaagcg	1 bp	NHEJ- like
Mitochondrial reaction 11 with primers 20461REV and 24961FOR						
M11-1	CIP 0.75 μ M	Col-0	19795 Recombinant 25300	gaatccgagg ATTTGCTTGGATTGTCCGCT tcatctcccc atatcttacc ATTTCTTCGGAATTGTCCGCT tcatctcccc atatcttacc ATTTCTTCGGAATTGTCCGCT catagcacga	(4) 21 bp (0)	MHMR*
M11-2	CIP 0.75 μ M	KO2	19317 Recombinant 25309	caagacgata GATAGGCCGG ttatgaaaaa catttcttcg GAATTGGCCG ttatgaaaaa catttcttcg GAATTGTCCG ctcatagcac	(5) 10 bp (1)	MHMR*
M11-3	CIP 0.75 μ M	KO2	19795 Recombinant 25300	gaatccgagg ATTTGCTTGGATTGTCCGCT tcatctcccc atatcttacc ATTTCTTCGGAATTGTCCGCT tcatctcccc atatcttacc ATTTCTTCGGAATTGTCCGCT catagcacga	(4) 21 bp (0)	MHMR*
Mitochondrial reaction 12 with primers 30001REV and 80161FOR						
M12-1	CIP 0.25 μ M	Col-0	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaaaaag attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-2	CIP 0.25 μ M	KO2	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaaaaag attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-3	CIP 0.25 μ M	KO2	29610 Recombinant 80713	acttttttc CCT-CCCCCCCCCA aaaaaaaaaaag cccgggcaaa CCT-CCCCCCCCCA aaaaaaaaaaag cccgggcaaa CCAACCCCCCTCCAA gaggaattgt	(0) 14 bp (3)	MHMR*
M12-4	CIP 0.75 μ M	Col-0	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaaaaag attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-5	CIP 0.75 μ M	KO2	29609 Recombinant 80994	tttttccct CCCCCCCCCA aaaaaaaaaa attagacgtc CCCCCCCCCA aaaaaaaaaaag attagacgtc CCCCCCCCCA tgctacggtt	(0) 10 bp (0)	MHMR
M12-6	CIP 0.75 μ M	KO2	29610 Recombinant 80713	acttttttc CCT-CCCCCCCCCA aaaaaaaaaaag cccgggcaaa CCT-CCCCCCCCCA aaaaaaaaaaag cccgggcaaa CCAACCCCCCTCCAA gaggaattgt	(0) 14 bp (3)	MHMR*
Mitochondrial reaction 13 with primers 84901REV and 151514FOR^f						
Mitochondrial reaction 14 with primers 210239REV and 245966FOR^f						
Mitochondrial reaction 15 with primers 1REV and 4981FOR^f						

M15-1	CIP 0.25 μ M	KO2	366037 Recombinant 5462	ctaactgata ggccttggct ctttacctat ggccttggct ctttacctat ctgggtgag	0 bp	NHEJ- like
M15-2	CIP 0.25 μ M	KO2	366764 Recombinant 5145	gcgaatagaa GGCTT-TTGCCAC ctcctcaac gattcttctg GGCTTGTGCCAC ctcctcaac gattcttctg GGCTTGTGCCAC taaagaaaga	(1) 13 bp (0)	MHMR*
M15-3	CIP 0.75 μ M	Col-0	366038 Recombinant 5171	taactgatag gccctggctg agaaagagtt gccctggctg agaaagagtt ctgccttcta	0 bp	NHEJ- like
M15-4	CIP 0.75 μ M	Col-0	366042 Recombinant 5080	tgataggccc tggctggacc ccgtaggaaa tggctggacc ccgtaggaaa agccgttgat	0 bp	NHEJ- like
M15-5	CIP 0.75 μ M	Col-0	366617 Recombinant 5215	gccgacttta AGTCAAAGGAGAAAGAGTG cctactgcca gcttcccagc AGTCAAAGGAGAAAGAGTG cctactgcca gcttcccagc AGTCAA <u>CA</u> AGAAAGAGTG tttctccct	(0) 19 bp (3)	MHMR*
M15-6	CIP 0.75 μ M	KO2	366041 Recombinant 5298	actgataggc C ctggctggac cacatctgct C ctggctggac cacatctgct C tttccaggac	1 bp	NHEJ- like
M15-7	CIP 0.75 μ M	KO2	366856 Recombinant 5607	agaagaggtc TCAACGAGCC tctgaattt aaactcccat TCAAGGAGCC tctgaattt aaactcccat TCAAGG <u>G</u> CC aatcattcat	(1) 10 bp (1)	MHMR*
M15-8	CIP 0.75 μ M	KO2	366856 Recombinant 5577	aggtctcaac GAGCC tctgaattt ttttttataa GAGCC tctgaattt ttttttataa GAGCC attggcagga	(0) 5 bp (0)	MHMR
Mitochondrial reaction 16 with primers 125781REV and 115816FOR^f – Deletion Events						
M16-1	CIP 0.25 μ M	KO2	125708 Recombinant 116865	taatgctttc TCTGCTTCCCTCT tggtttcggt ggttgctgag TCTGCTTCCCTCT tggtttcggt ggttgctgag TCTGCTTCCCTCT ctatcaaagc	(0) 14 bp (0)	MHMR
M16-2	CIP 0.75 μ M	Col-0	125359 Recombinant 116280	gggtacgtcc CCTATC gcctttcatt gagcaggtaa CCTACC gcctttcatt gagcaggtaa CCTACC ttctaaacca	(1) 6 bp (0)	MHMR*
M16-3	CIP 0.75 μ M	KO2	125529 Recombinant 116981	ttagctgagc ATCAATCCCAT tctttgcgga caggtatgaa ATCAATCCCAT tctttgcgga caggtatgaa ATCAATCCCAT gttggtcgta	(0) 11 bp (0)	MHMR
M16-4	CIP 0.75 μ M	KO2	125708 Recombinant 116865	taatgctttc TCTGCTTCCCTCT tggtttcggt ggttgctgag TCTGCTTCCCTCT tggtttcggt ggttgctgag TCTGCTTCCCTCT ctatcaaagc	(0) 14 bp (0)	MHMR
Mitochondrial reaction 17 with primers 135180REV and 140053FOR^f						
Mitochondrial reaction 18 with primers 171214REV and 180454FOR^f						

M18-1	CIP 0.25 μ M	KO2	170939 Recombinant 181121	attctataga TAC CTCTGGTAG gtaaagcact ggacaggtag TAG CTCTGGTAG gtaaagcact ggacaggtag TAG CTCTGGTAG agcgagggac	(1) 12 bp (0)	MHMR*
M18-2	CIP 0.25 μ M	KO2	171149 Recombinant 181066	aaccttcgtc TCAATCTTCATAA tcaattttatg tcacaaaacta TCAATCTTCATAA tcaattttatg tcacaaaacta TCAAT-<u>TTC</u>ATAA gagaagacga	(0) 13 bp (1)	MHMR*
M18-3	CIP 0.75 μ M	Col-0	170678 Recombinant 181451	gttttgatg gagctactgg attctccgtt gagctactgg attctccgtt tcgatattaa	0 bp	NHEJ- like
M18-4	CIP 0.75 μ M	Col-0	170609 Recombinant 180538	ttgtgatgac TGAACATGA-AACGATCGATATA gatcggcgta agtcaaactg TGAACATGA-AACGATCGATATA gatcggcgta agtcaaactg TGAACAT<u>CAGA</u>AGGATCGATATA ttggtaatag	(0) 22 bp (3)	MHMR*
M18-5	CIP 0.75 μ M	KO2	170939 Recombinant 181121	attctataga TAC CTCTGGTAG gtaaagcact ggacaggtag TAG CTCTGGTAG gtaaagcact ggacaggtag TAG CTCTGGTAG agcgagggac	(1) 12 bp (0)	MHMR*
M18-6	CIP 0.75 μ M	KO2	171149 Recombinant 181066	aaccttcgtc TCAATCTTCATAA tcaattttatg tcacaaaacta TCAATCTTCATAA tcaattttatg tcacaaaacta TCAAT-<u>TTC</u>ATAA gagaagacga	(0) 13 bp (1)	MHMR*
Mitochondrial reaction 19 with primers 245966REV and 239584FOR^f – Deletion Event						
M19-1	CIP 0.75 μ M	KO2	244727 Recombinant 240924	gtaacccct AAAAGGAAACTCACCACCAAAGA- aaaaggtagg aaaaaaagaa AAAAGGAAACTCACCACCAAAGAT aaaaggtagg aaaaaaagaa AAAAGGAAACTCACCACCAAAGAT cagtccaata	(1) 25 bp (0)	MHMR*
Mitochondrial reaction 20 with primers 270785REV and 275583FOR^f						
M20-1	CIP 0.25 μ M	KO2	269741 Recombinant 276747	attgaagaaa GG AAGAAAGCCTGTAAA ctaccttact tccggtcatt GG TAGAAAGCCTGTAAA ctaccttact tccggtcatt GG TAGAAAGCCTGTAAA tggatttat	(1) 17 bp (0)	MHMR*
M20-2	CIP 0.25 μ M	KO2	269069 Recombinant 275816	gtagtcccg CCGGATC aatagattgg tctcccaaac CCGGATC aatagattgg tctcccaaac CCGGATC cgcaagtcgg	(0) 7 bp (0)	MHMR
M20-3	CIP 0.25 μ M	KO2	270780 Recombinant 276407	tccaagccta ATAGGA acatcgctag tattccttcc ATAGGA acatcgctag tattccttcc ATAGGA ggacaccttg	(0) 6 bp (0)	MHMR
M20-4	CIP 0.75 μ M	Col-0	269800 Recombinant 276776	cagtaaac CCTCTCCTT<u>C</u>CAGT gttttttcct gaatttatct CCTCTCCTT<u>C</u>CAGT gttttttcct gaatttatct CCTCTCCTT<u>C</u>CAGT cgagttccct	(1) 14 bp (0)	MHMR*
M20-5	CIP 0.75 μ M	Col-0	270679 Recombinant 276212	aggacggaac caccgaggaa tttgggtggtt caccgaggaa tttgggtggtt ttccacgacc	0 bp	NHEJ- like
M20-6	CIP 0.75 μ M	Col-0	270497 Recombinant 276011	cattcatggc C ttttttggtt gagatgcgtg C ttttttggtt gagatgcgtg C gcacgggtga	1 bp	NHEJ- like

M20-7	CIP 0.75 µM	Col-0	270716 Recombinant 275872	agtaggatgt atactgtgct tgctacctcc atactgtgct tgctacctcc caacagacgt	0 bp	NHEJ- like
M20-8	CIP 0.75 µM	Col-0	270715 Recombinant 275738	tagtaggatg tatactgtgc ggaatatggc tatactgtgc ggaatatggc caccgtagag	0 bp	NHEJ- like
M20-9	CIP 0.75 µM	KO2	269741 Recombinant 276747	attgaagaaa GGAAGAAAGCCTGTAAA ctaccttact tccggtcatt GGTAGAAAGCCTGTAAA ctaccttact tccggtcatt GGTAGAAAGCCTGTAAA tggaaattat	(1) 17 bp (0)	MHMR*
M20-10	CIP 0.75 µM	KO2	270561 Recombinant 276533	ctgaggcgtt CCTCTCAGTCA tacaaaaaag atctcctacg CCTCTCAGTCA tacaaaaaag atctcctacg CCTCTCAGTCA aaagagtgcg	(0) 11 bp (0)	MHMR
M20-11	CIP 0.75 µM	KO2	270682 Recombinant 276415	ggacggaacc AC cgaggaatac ccataggagg AC cgaggaatac ccataggagg AC accttggttc	2 bp	NHEJ- like
M20-12	CIP 0.75 µM	KO2	270521 Recombinant 276184	ttctttgatt GATCTC cctttcgtat tcttccttaa GATCTC cctttcgtat tcttccttaa GATCTC aaccaaccac	(0) 6 bp (0)	MHMR
M20-13	CIP 0.75 µM	KO2	270135 Recombinant 275738	gatgtggaag CAC atcttagcaa ggaatatggc CAC atcttagcaa ggaatatggc CAC cggatagagc	3 bp	NHEJ- like
M20-14	CIP 0.75 µM	KO2	270445 Recombinant 275819	gcatttatca GATCAGC tcccagagta cccaaaccg GATCAGC tcccagagta cccaaaccg GATCAGC aagtcgggtc	(0) 7 bp (1)	MHMR*
M20-15	CIP 0.75 µM	KO2	270679 Recombinant 275979	aggacggaac caccgaggaa agaaaggctc caccgaggaa agaaaggctc tatcggctcag	0 bp	NHEJ- like
Plastid reaction 1 with primers 31138REV and 19181FOR^f – Deletion Events						
P1-1	CIP 0.25 µM	Col-0	30922 Recombinant 20410	aaatagtagt atccctgtct atgatagtta atccctgtct atgatagtta ccaaagatt	0 bp	NHEJ- like
P1-2	CIP 0.25 µM	Col-0	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-3	CIP 0.25 µM	KO1/3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-4	CIP 0.75 µM	KO1/3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTT acacttctta	(1) 18 bp (0)	MHMR*

P1-5	NOV 0 μ M	KO1/3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-6	NOV 100 μ M	KO1/3	31024 Recombinant 20347	gttaagttat CAAAAATTTGCTTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTTT ttttttagaa aatatatata CAAAAATTTCTTTTTTT acacttctta	(1) 18 bp (0)	MHMR*
P1-7	NOV 100 μ M	KO1/3	30618 Recombinant 19900	tataatgaaa TGAAATGG tatgactagc caattcctaa TGAAATGG tatgactagc caattcctaa TGAAATGG atgtagcagt	(0) 8 bp (0)	MHMR
Plastid reaction 2 with primers 40130REV and 50251FOR^{e f}						
P2-1	CIP 0 μ M	Col-0	39586 Recombinant 50325	ttaatgccat AATGCC tttcaaatcc tattagtaaa AATGCC tttcaaatcc tattagtaaa AATGCC cagaaaatat	(0) 6 bp (0)	MHMR
P2-2	CIP 0 μ M	KO1/3	38759 Recombinant 50687	ttgcgatagc TAAATGATGATGT gccatatcgg cttcgttttt TAAATGATGATGT gccatatcgg cttcgttttt TAAATGATGACGT atcaaaaaat	(0) 13 bp (1)	MHMR*
P2-3	CIP 0 μ M	KO1/3	38989 Recombinant 50299	cccctggatg CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC acttgatatt	(0) 10 bp (0)	MHMR
P2-4	CIP 0.25 μ M	Col-0	39586 Recombinant 50325	ttaatgccat AATGCC tttcaaatcc tattagtaaa AATGCC tttcaaatcc tattagtaaa AATGCC cagaaaatat	(0) 6 bp (0)	MHMR
P2-5	CIP 0.25 μ M	KO1/3	39109 Recombinant 50758	cttggtttcc ATTTTGGTTG taggtgtaac gatagattga ATTTTGGTTG taggtgtaac gatagattga ATTTTGGTTG aatttaatta	(0) 10 bp (0)	MHMR
P2-6	CIP 0.25 μ M	KO1/3	38989 Recombinant 50299	cccctggatg CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC acttgatatt	(0) 10 bp (0)	MHMR
P2-7	CIP 0.25 μ M	KO1/3	39517 Recombinant 50285	tcatgtgctg TAGCAATACCAACCA aatacgaacga ctccggaaat TAGCAATACCAACCA aatacgaacga ctccggaaat TAGAATGCCAAAACA ggaatagcac	(0) 16 bp (3)	MHMR*
P2-8	CIP 0.25 μ M	KO1/3	39582 Recombinant 50289	tggaaatcct AATGCCATAA tgcctttcaa ggaaattaga AATGCCATAA tgcctttcaa ggaaattaga AATGCCAAA caggaatagc	(0) 10 bp (1)	MHMR*
P2-9	CIP 0.75 μ M	Col-0	37846 Recombinant 50396	ccagaaaact GC caataaaat aaaaatcccc GC caataaaat aaaaatcccc GC ttagtcaatt	(0) 2 bp	NHEJ- like
P2-10	CIP 0.75 μ M	Col-0	39655 Recombinant 50363	attccttgcta AGAAGAACGCCCATG-TTGTGGCAA ttccaccag gcagaaacat AGACGAACGCCCATG-TTGTGGCAA ttccaccag gcagaaacat AGACGAACCTCTATGATGTGGAAA aaataccgc	(1) 24 bp (5)	MHMR*

P2-11	CIP 0.75 μ M	Col-0	39192 Recombinant 50308	ttcattagta C gtaaaccgat acaggaatag C gtaaaccgat acaggaatag C acttgatatt	1 bp	NHEJ- like
P2-12	CIP 0.75 μ M	Col-0	39586 Recombinant 50325	ttaatgccat AATGCC tttcaaatcc tattagtaaa AATGCC tttcaaatcc tattagtaaa AATGCC cagaaaatat	(0) 6 bp (0)	MHMR
P2-13	CIP 0.75 μ M	Col-0	39845 Recombinant 50392	gagcacctag gaaaaaaga tgaaaaaaat gaaaaaaga tgaaaaaaat acccgcttag	0 bp	NHEJ- like
P2-14	CIP 0.75 μ M	KO1/3	37647 Recombinant 50326	aagaacatcc ATGCCCAGA ctgataaact attagtaaaa ATGCCCAGA ctgataaact attagtaaaa ATGCCCAGA aaatatcata	(0) 9 bp (0)	MHMR
P2-15	CIP 0.75 μ M	KO1/3	39109 Recombinant 50758	cttggtttcc ATTTTGGTTG tagtgtaac gatagattga ATTTTGGTTG tagtgtaac gatagattga ATTTTGGTTG aatttaatta	(0) 10 bp (0)	MHMR
P2-16	CIP 0.75 μ M	KO1/3	38989 Recombinant 50299	cccttgatg CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC acttgatatt	(0) 10 bp (0)	MHMR
P2-17	CIP 0.75 μ M	KO1/3	39536 Recombinant 50834	tacgacgagt AGTGGGG tctgagcta accaagattg AGTGGGG tctgagcta accaagattg AGTGGGG atacaaaaaa	(0) 7 bp (0)	MHMR
P2-18	CIP 0.75 μ M	KO1/3	39195 Recombinant 50397	ttcattagta CGTA aaccgattgt aaaatacccg CGTA aaccgattgt aaaatacccg CTA gtcaattcca	4 bp	NHEJ- like
P2-19	CIP 0.75 μ M	KO1/3	39613 Recombinant 50397	tagccattat CCTA ctgcaataat aaaatacccg CCTA ctgcaataat aaaatacccg CTA gtcaattcca	4 bp	NHEJ- like
P2-20	CIP 0.75 μ M	KO1/3	39586 Recombinant 50325	ttaatgccat AATGCC tttcaaatcc tattagtaaa AATGCC tttcaaatcc tattagtaaa AATGCC cagaaaatat	(0) 6 bp (0)	MHMR
P2-21	CIP 0.75 μ M	KO1/3	39655 Recombinant 50363	attcttgcta AGAAGAACGCCCATG-TTGTGGCAA ttccaccag gcagaaacat AGACGAACGCCCATG-TTGTGGCAA ttccaccag gcagaaacat AGACGAACCTCTATGATGTGGAA aaatacccg	(1) 24 bp (5)	MHMR*
P2-22	NOV 0 μ M	KO1/3	38889 Recombinant 50299	cccttgatg CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC acttgatatt	(0) 10 bp (0)	MHMR
P2-23	NOV 0 μ M	KO1/3	39109 Recombinant 50758	cttggtttcc ATTTTGGTTG tagtgtaac gatagattga ATTTTGGTTG tagtgtaac gatagattga ATTTTGGTTG aatttaatta	(0) 10 bp (0)	MHMR
P2-24	NOV 0 μ M	KO1/3	39602 Recombinant 50362	caaatcctcc TAG ccattatcct agcagaaaca TAG ccattatcct agcagaaaca TAG acgaactcct	3 bp	NHEJ- like

P2-25	NOV 100 µM	KO1/3	38889 Recombinant 50299	cccctggatg CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC gacatggact aatgccaaaa CAGGAATAGC acttgatatt	(0) 10 bp (0)	MHMR
Plastid reaction 3 with primers 84761REV and 95260FOR^{g f}						
P3-1	CIP 0.25 µM	KO1/3	83151 Recombinant 95532	ggaaattctg GTAATAGCAATG ttgagttttc gcttgtttga GTAATAGCAATG ttgagttttc gcttgtttga GTAATAGCAATG agattcccga	(0) 12 bp (0)	MHMR
P3-2	CIP 0.25 µM	KO1/3	142740 Recombinant 36	aaaatcagtc ATTCACAA-CCACTG atgaaggatt gcgatggtga ATTCACAATCCACTG atgaaggatt gcgatggtga ATTCACAATCCACTG ccttaatcca	(1) 15 bp (0)	MHMR*
P3-3	CIP 0.25 µM	KO1/3	84368 Recombinant 95699	acggcgcacga AGAATCAAAGTC tcaactatatt ctggagtggg AGAATCAAAGTC tcaactatatt ctggagtggg AGAGTCAGAGTC gaaaagagga	(0) 12 bp (2)	MHMR*
P3-4	CIP 0.25 µM	KO1/3	84661 Recombinant 95481	tttaatgtgg CCGATTT ccctctttt aattacatat CCGATTT ccctctttt aattacatat CCGATTT gacctatgga	(0) 7 bp (0)	MHMR
P3-5	CIP 0.75 µM	Col-0	84464 Recombinant 96061	gtcgacggat gctcctatta gtcgacggat ccacctccat cctcccctca ccacctccat	(0) 0 bp	NHEJ- like
P3-6	CIP 0.75 µM	KO1/3	83151 Recombinant 95532	ggaaattctg GTAATAGCAATG ttgagttttc gcttgtttga GTAATAGCAATG ttgagttttc gcttgtttga GTAATAGCAATG agattcccga	(0) 12 bp (0)	MHMR
P3-7	CIP 0.75 µM	KO1/3	83532 Recombinant 95421	ttagatttct GTTTTCAT tgctgcatcc cagcatataa GTTTTCAT tgctgcatcc cagcatataa GTTTTCAT gctcgcata	(0) 8 bp (1)	MHMR*
P3-8	CIP 0.75 µM	KO1/3	142740 Recombinant 36	aaaatcagtc ATTCACAA-CCACTG atgaaggatt gcgatggtga ATTCACAATCCACTG atgaaggatt gcgatggtga ATTCACAATCCACTG ccttaatcca	(1) 15 bp (0)	MHMR*
P3-9	CIP 0.75 µM	KO1/3	143168 Recombinant 153988	tccataggtc AAATCGG atatgtaatt aaaagagggg AAATCGG atatgtaatt aaaagagggg AAATCGG ccacattaaa	(0) 7 bp (0)	MHMR
Plastid reaction 4 with primers 105961REV and 116007FOR^f						
P4-1	CIP 0.25 µM	KO1/3	105738 Recombinant 116787	agctaaggcc CCTAAATGACCGCT cagtgataaa agaaaatgat CCTAAATGACCGCT cagtgataaa agaaaatgat CCTATTTGACCGCT gtacattgct	(0) 14 bp (2)	MHMR*
P4-2	CIP 0.25 µM	KO1/3	105452 Recombinant 116276	ggaggtccga ACCGACTGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA gctgcataaa	(3) 16 bp (0)	MHMR*
P4-3	CIP 0.25 µM	KO1/3	105322 Recombinant 116009	tggtcttggtt AAGGGAA CC CA CC CG gaggccgtagc cgggcaatgc AAGGGAA CC CA CC CG gaggccgtagc cgggcaatgc AAGGGAA CC CA CC CG ataaaaatagt	(2) 14 bp (0)	MHMR*

P4-4	CIP 0.75 μ M	Col-0	105739 Recombinant 116394	ctaaatgacc GCTC agtgataaag caaccatag GCTC agtgataaag caaccatag GCTC ccatttttaa	4 bp	NHEJ- like
P4-5	CIP 0.75 μ M	KO1/3	105452 Recombinant 116276	ggaggtccga ACCGACTGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA aaatcagcgg aattacgttg ACCGGGAGATGTTGAA gctgcataaa	(3) 16 bp (0)	MHMR*
P4-6	CIP 0.75 μ M	KO1/3	105821 Recombinant 116393	gtgcgtaata GCTC actgatcgag caaccatag GCTC actgatcgag caaccatag GCTC ccatttttaa	4 bp	NHEJ- like
Plastid reaction 5 with primers 69633REV and 59291FOR^{e f} – Deletion Events						
P5-1	CIP 0.25 μ M	KO1/3	68984 Recombinant 59308	tctttccggc AATTGCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ttgctactct	(1) 14 bp (0)	MHMR*
P5-2	CIP 0.75 μ M	KO1/3	68984 Recombinant 59308	tctttccggc AATTGCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ttgctactct	(1) 14 bp (0)	MHMR*
P5-3	NOV 0 μ M	KO1/3	68522 Recombinant 59380	atttcttcca ATCCCTTATTTTTTTT atgtcattcg accgaatcat ATCACCTTTTTTTTTT atgtcattcg accgaatcat ATCACCTTTTTTTTTT tattttaaaa	(2) 16 bp (0)	MHMR*
P5-4	NOV 100 μ M	KO1/3	68984 Recombinant 59308	tctttccggc AATTGCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ctttatttct gtataaaaa AATTCCAATGGCTT ttgctactct	(1) 14 bp (0)	MHMR*
Plastid reaction 6 with primers 111330REV and 121141FOR^{e f}						
P6-1	CIP 0 μ M	KO1/3	109932 Recombinant 121745	atgcaacaac AAGATGTTATTTGTAACAAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT cccatcggct	(4) 27 bp (0)	MHMR*
P6-2	CIP 0 μ M	KO1/3	111043 Recombinant 122213	cggataatag AAAGTCTTGTATT aagtccaaaa actcttacat AAAGTCTTGTATT aagtccaaaa actcttacat AAAGTCTTGTATT gataattcaa	(0) 14 bp (1)	MHMR*
P6-3	CIP 0 μ M	KO1/3	19786 Recombinant 110850	catatctcaa TAGATTGGCGT-AATTTTT ctaccgcatg caaaattttg TAGATTGGCGT-AATTTTT ctaccgcatg caaaattttg TAGATTGGTATGAATTTTT gagaaatgca	(0) 18 bp (3)	MHMR*
P6-3	CIP 0 μ M	KO1/3	19867 Recombinant 122003	taattatggg AGAAGTTGATCGTTAAAATGAT aatttatgca gccagccctt AGAAGTTGATCGTTAAAATGAT aatttatgca gccagccctt AGAAGGT-ATCGTTAAAA-GAT catcaattcc	(0) 22 bp (3)	MHMR*
P6-4	CIP 0 μ M	KO1/3	111244 Recombinant 121304	tagaaggagc CACTTTTTTTTCACT gtaatttaga aatttcagag CAGTTTTTTTCACT gtaatttaga aatttcagag CAGTTTTTTTCACT catatataac	(2) 16 bp (0)	MHMR*
P6-5	CIP 0.25 μ M	Col-0	109932 Recombinant 121745	atgcaacaac AAGATGTTATTTGTAACAAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT cccatcggct	(4) 27 bp (0)	MHMR*

P6-6	CIP 0.25 μ M	Col-0	109826 Recombinant 121273	cctgaataat CTCATTTTTTC aattattcaa ggaaatgaga CTCATTTTTTC aattattcaa ggaaatgaga CTCATTTTTTC gttttactgc	(0) 10 bp (0)	MHMR
P6-7	CIP 0.25 μ M	Col-0	111043 Recombinant 122213	cggataatag AAAGTTCCTGTATT aagtccaaaa actcctacat AAAGTTCCTGTATT aagtccaaaa actcctacat AAAGTTCCTGTATT gataattcaa	(0) 14 bp (1)	MHMR*
P6-8	CIP 0.25 μ M	Col-0	111244 Recombinant 121304	tagaaggagc CACTTTTTTTTCCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT catatataac	(2) 16 bp (0)	MHMR*
P6-9	CIP 0.25 μ M	KO1/3	109932 Recombinant 121745	atgcaacaac AAGATGTTATTTGTAACAAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT cccatcggct	(4) 27 bp (0)	MHMR*
P6-10	CIP 0.25 μ M	KO1/3	111244 Recombinant 121304	tagaaggagc CACTTTTTTTTCCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT catatataac	(2) 16 bp (0)	MHMR*
P6-11	CIP 0.75 μ M	Col-0	111043 Recombinant 122213	cggataatag AAAGTTCCTGTATT aagtccaaaa actcctacat AAAGTTCCTGTATT aagtccaaaa actcctacat AAAGTTCCTGTATT gataattcaa	(0) 14 bp (1)	MHMR*
P6-12	CIP 0.75 μ M	Col-0	111244 Recombinant 121304	tagaaggagc CACTTTTTTTTCCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT catatataac	(2) 16 bp (0)	MHMR*
P6-13	CIP 0.75 μ M	KO1/3	109932 Recombinant 121745	atgcaacaac AAGATGTTATTTGTAACAAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT tggttggtta tagatggacg AAGATTTCTTTG-AA-AAGTAGTTTTGT cccatcggct	(4) 27 bp (0)	MHMR*
P6-14	CIP 0.75 μ M	KO1/3	111244 Recombinant 121304	tagaaggagc CACTTTTTTTTCCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT gtaatttaga aatttcagag CAGTTTTTTTTTCACT catatataac	(2) 16 bp (0)	MHMR*
P6-15	NOV 0 μ M	KO1/3	111189 Recombinant 121305	atthtttatta AGTTTTTTTTT cttctctctt atthtcagagc AGTTTTTTTTT cttctctctt atthtcagagc AGTTTTTTTTT tcaactcatat	(0) 10 bp (0)	MHMR
Plastid reaction 7 with primers 50513REV and 59311FOR^f						
P7-1	CIP 0.25 μ M	KO1/3	50489 Recombinant 59938	aggттаатсg AATCATTTTTTTTC gtttggttgt cgtcccagca AATCATTTTTTTTC gtttggttgt cgtcccagca AATCATTTTTTTTC cacaagggat	(0) 14 bp (0)	MHMR
P7-2	CIP 0.75 μ M	Col-0	50489 Recombinant 59938	aggттаатсg AATCATTTTTTTTC gtttggttgt cgtcccagca AATCATTTTTTTTC gtttggttgt cgtcccagca AATCATTTTTTTTC cacaagggat	(0) 14 bp (0)	MHMR
P7-3	CIP 0.75 μ M	KO1/3	50489 Recombinant 59938	aggттаатсg AATCATTTTTTTTC gtttggttgt cgtcccagca AATCATTTTTTTTC gtttggttgt cgtcccagca AATCATTTTTTTTC cacaagggat	(0) 14 bp (0)	MHMR
P7-4	CIP 0.75 μ M	KO1/3	50462 Recombinant 59705	ctcttgagtc AAAACAAAAA ttcagggttaa agttactttg AAAACAAAAA ttcagggttaa agttactttg AAAACAAAAA aaaggaaagt	(0) 10 bp (0)	MHMR

P7-5	NOV 0 μ M	KO1/3	50489 Recombinant 59938	aggттаатсg AATCATTTTTTTTC gtttggttg cgtcccagca AATCATTTTTTTTC gtttggttg cgtcccagca AATCATTTTTTTTC cacaagggat	(0) 14 bp (0)	MHMR
Plastid reaction 8 with primers 68821REV and 79920FOR^f						
P8-1	CIP 0 μ M	KO1/3	68809 Recombinant 81088	cctcgaaagc TTGATGCAAA taaacgaagt atctcttctt TTAATGCAAA taaacgaagt atctcttctt TTAATGCAAA ggacgaagta	(1) 10 bp (0)	MHMR*
P8-2	CIP 0 μ M	KO1/3	68714 Recombinant 80522	ggcgaaaatc TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA tttatatata	(0) 13 bp (0)	MHMR
P8-3	CIP 0.25 μ M	KO1/3	68714 Recombinant 80522	ggcgaaaatc TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA tttatatata	(0) 13 bp (0)	MHMR
P8-4	CIP 0.75 μ M	Col-0	68733 Recombinant 81329	cgatgagcc G aaactaaagc aatactcagg G aaactaaagc aatactcagg G gatcctttcc	1 bp	NHEJ- like
P8-5	CIP 0.75 μ M	Col-0	68714 Recombinant 80522	ggcgaaaatc TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA tttatatata	(0) 13 bp (0)	MHMR
P8-6	CIP 0.75 μ M	KO1/3	68733 Recombinant 82098	cgatgagcc G aaactaaagc tatattcatg G aaactaaagc tatattcatg G gtccgctcg	1 bp	NHEJ- like
P8-7	CIP 0.75 μ M	KO1/3	68714 Recombinant 80522	ggcgaaaatc TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA tttatatata	(0) 13 bp (0)	MHMR
P8-8	NOV 0 μ M	KO1/3	68522 Recombinant 80652	atccccttat TTTTT atgtcattcg tcaatacctg TTTTT atgtcattcg tcaatacctg TTTTT taatccgata	(0) 6 bp (0)	MHMR
P8-9	NOV 0 μ M	KO1/3	68714 Recombinant 80522	ggcgaaaatc TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA tttatatata	(0) 13 bp (0)	MHMR
P8-10	NOV 100 μ M	Col-0	68714 Recombinant 80522	ggcgaaaatc TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA tttatatata	(0) 13 bp (0)	MHMR
P8-11	NOV 100 μ M	KO1/3	68819 Recombinant 81100	atgcaaataa ACGAAGT tttgtttctac aatgcaaagg ACGAAGT tttgtttctac aatgcaaagg ACGAAGT aaaaaaatat	(0) 7 bp (0)	MHMR
P8-12	NOV 100 μ M	KO1/3	68714 Recombinant 80522	ggcgaaaatc TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA cccctatccc atcaacatgt TCTTTTCTTTTA tttatatata	(0) 13 bp (0)	MHMR
Plastid reaction 9 with primers 10230REV and 30511FOR^f						

P9-1	CIP 0 μ M	KO1/3	9940 Recombinant 30859	agaataatcc GTTTCGTTTTTT atactttctc agagggttaa GTTTCGTTTTTT atactttctc agagggttaa GTTTCGTTTTTT tttactttc	(0) 12 bp (0)	MHMR
P9-2	CIP 0 μ M	KO1/3	9948 Recombinant 30859	agaataatcc GTTTCGTTTTTT- <u>ATACTTTC</u> tcttgaagta agagggttaa GTTTCGTTTTTTTTACTTTC tcttgaagta agagggttaa GTTTCGTTTTTTTTACTTTC cttgatagat	(3) 22 bp (0)	MHMR*
P9-3	CIP 0 μ M	KO1/3	9895 Recombinant 30750	taaaaaatta ATTAAAATTAAAGAGTAAA attgactaaa aagtggaaat ATTAAAATTAAAGAGTAAA attgactaaa aagtggaaat ATTAAAATTAA <u>AA</u> TAAA taaaaaaaa	(0) 19 bp (2)	MHMR*
P9-4	CIP 0.25 μ M	Col-0	9948 Recombinant 30859	agaataatcc GTTTCGTTTTTT- <u>ATACTTTC</u> tcttgaagta agagggttaa GTTTCGTTTTTTTTACTTTC tcttgaagta agagggttaa GTTTCGTTTTTTTTACTTTC cttgatagat	(3) 22 bp (0)	MHMR*
P9-5	CIP 0.25 μ M	KO1/3	9476 Recombinant 31012	aaattcacia TTGCCCTTTTTTTTTATTAGAGAAAATTGTT atattagaat tatcaaaaat TTGCCCTTTTTTTTTATTAGAGAAAATTGTT atattagaat tatcaaaaat TTG- <u>CTTTTTTTTTTTAGAAAATTATTGTT</u> cagaggataa	(0) 32 bp (7)	MHMR*
P9-6	CIP 0.25 μ M	KO1/3	9948 Recombinant 30864	agaataatcc GTTTCGTTTTTTATACTTTC tcttgaagta gttaagtttc GTTTCGTTTTTTATACTTTC tcttgaagta gttaagtttc GTTT- <u>TTTTTT-TACTTTC</u> cttgatagat	(0) 20 bp (3)	MHMR*
P9-7	CIP 0.25 μ M	KO1/3	9584 Recombinant 30885	ttttaacaat AGGAAAGTAAAAAAA gcgaaaagcg caatctatca AGGAAAGTAAAAAAA gcgaaaagcg caatctatca AGGAAAGTAAAAAAA aacgaaactt	(0) 16 bp (0)	MHMR
P9-7	CIP 0.25 μ M	KO1/3	31015 Recombinant 30829	ttttctaaaa AAAAAAAA gcaaattttt ggtctaacat AAAAAAAA gcaaattttt ggtctaacat AAAAAAAA aaagaaaaaa	(0) 9 bp (0)	MHMR
P9-8	CIP 0.75 μ M	Col-0	9472 Recombinant 31015	cacaattgcc CTTTTTTTTATTAGAGAAAATT tggtatatta caaaaatttg CTTTTTTTTATTAGAGAAAATT tggtatatta caaaaatttg CTTTTTTTT- <u>TTTTAGAAAATT</u> attgttcaga	(0) 23 bp (3)	MHMR*
P9-9	CIP 0.75 μ M	Col-0	9933 Recombinant 31046	tcaaaaatta AGAATAATCCGTTTC gttttttata tattgttcag AGGATAATCCGTTTC gttttttata tattgttcag AGGATAATCCG <u>TATC</u> tcttcatcac	(1) 15 bp (1)	MHMR*
P9-10	CIP 0.75 μ M	KO1/3	9933 Recombinant 31046	tcaaaaatta AGAATAATCCGTTTC gttttttata tattgttcag AGGATAATCCGTTTC gttttttata tattgttcag AGGATAATCCG <u>TATC</u> tcttcatcac	(1) 15 bp (1)	MHMR*
P9-11	CIP 0.75 μ M	KO1/3	9948 Recombinant 30859	agaataatcc GTTTCGTTTTTT- <u>ATACTTTC</u> tcttgaagta agagggttaa GTTTCGTTTTTTTTACTTTC tcttgaagta agagggttaa GTTTCGTTTTTTTTACTTTC cttgatagat	(3) 22 bp (0)	MHMR*
P9-12	CIP 0.75 μ M	KO1/3	9883 Recombinant 30757	ctagaatttc TTAAAAATTAATTAAAA ttaaagagta aatattaaaa TTAAAAATTAATAAAA ttaaagagta aatattaaaa TTAAAAATTAATAAAA aaaatctaaa	(2) 18 bp (0)	MHMR*
Plastid reaction 10 with primers 60182REV and 50251FOR^f – Deletion Events						
P10-1	CIP 0 μ M	KO1/3	59060 Recombinant 50621	gctttttttg T <u>ACTTATTTTTTTATTCTATTTCTATT</u> cggttctatt atctgattaa T <u>ACTAGTAATTTTTTTATTCTGTTTCTATT</u> cggttctatt atctgattaa T <u>ACTAGTAATTTTTTTATTCTGTTT<u>TG</u>T</u> taaaaaaaaatt	(7) 33 bp (2)	MHMR*

P10-2	CIP 0 μ M	KO1/3	59724 Recombinant 51123	gagttacttt <u>GAAA</u> ACAAAAAAAAAGGAAA gtcatatattt tagaaataaa GAATACAAAAAAAAAGGAAA gtcatatattt tagaaataaa GAATACAAAAAAAAAGTAAA aaaatattatc	(1) 19 bp (1)	MHMR*
P10-3	CIP 0 μ M	KO1/3	59726 Recombinant 50847	ctttgaaaac AAAAAAAAAAGGAAAGT catatatttct gggatacaaaa AAAAAAAAAAGGAAAGT catatatttct gggatacaaaa AAAAAAAAAAGGAAAGT attatgaact	(0) 15 bp (1)	MHMR*
P10-4	CIP 0 μ M	KO1/3	59953 Recombinant 50474	cgtcccagca AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC gtttggttgt	(0) 14 bp (0)	MHMR
P10-5	CIP 0.25 μ M	KO1/3	59724 Recombinant 51123	gagttacttt <u>GAAA</u> ACAAAAAAAAAGGAAA gtcatatattt tagaaataaa GAATACAAAAAAAAAGGAAA gtcatatattt tagaaataaa GAATACAAAAAAAAAGTAAA aaaatattatc	(2) 19 bp (2)	MHMR*
P10-6	CIP 0.25 μ M	KO1/3	59726 Recombinant 50847	ctttgaaaac AAAAAAAAAAGGAAAGT catatatttct gggatacaaaa AAAAAAAAAAGGAAAGT catatatttct gggatacaaaa AAAAAAAAAAGGAAAGT attatgaact	(0) 15 bp (1)	MHMR*
P10-7	CIP 0.25 μ M	KO1/3	59726 Recombinant 50845	ctttgaaaac AAAAAAAAA-GGAAAGT catatatttct tggggataca AAAAAAAAA-GGAAAGT catatatttct tggggataca AAAAAAAAAAGGAAAGT attatgaact	(0) 16 bp (2)	MHMR*
P10-8	CIP 0.25 μ M	KO1/3	60097 Recombinant 50836	tatttttctg TGGGGAT ttctctggaaa caagattgag TGGGGAT ttctctggaaa caagattgag TGGGGAT acaaaaaaaaa	(0) 7 bp (0)	MHMR
P10-9	CIP 0.25 μ M	KO1/3	59953 Recombinant 50474	cgtcccagca AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC gtttggttgt	(0) 14 bp (0)	MHMR
P10-10	CIP 0.75 μ M	Col-0	59846 Recombinant 51200	aatttctgct GGGC ctttatccta ttcttagtta GGGC ctttatccta ttcttagtta GGGC tatacggact	(0) 4 bp	NHEJ- like
P10-11	CIP 0.75 μ M	Col-0	59953 Recombinant 50474	cgtcccagca AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC gtttggttgt	(0) 14 bp (0)	MHMR
P10-12	CIP 0.75 μ M	KO1/3	58619 Recombinant 50471	catttttctt <u>TC</u> AATCATTTTTTTT tatcgatatt ttcaggttaa TCGAATCATTTTTTTT tatcgatatt ttcaggttaa TCGAATCATTTTTTTT cgtttggttg	(1) 16 bp (0)	MHMR*
P10-13	CIP 0.75 μ M	KO1/3	59398 Recombinant 50631	atcacctttt TTTTTT attttaaaaa taactagtaa TTTTTT attttaaaaac taactagtaa TTTTTT tatttctggt	(0) 7 bp (0)	MHMR
P10-14	CIP 0.75 μ M	KO1/3	59719 Recombinant 50843	tactttgaaa AAAAAAAAAA ggaagtcacat gagtggggat AAAAAAAAAA ggaagtcacat gagtggggat AAAAAAAAAA aaagaaagta	(0) 10 bp (0)	MHMR
P10-15	CIP 0.75 μ M	KO1/3	59953 Recombinant 50474	cgtcccagca AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC cacaagggat aggттаатсg AATCATTTTTTTTTC gtttggttgt	(0) 14 bp (0)	MHMR

P10-16	NOV 0 μ M	KO1/3	59724 Recombinant 51123	gagttacttt <u>GAAA</u> ACAAAAAAGGAAA gtcatat ^{ttt} tagaaataaa <u>GAAT</u> ACAAAAAAGGAAA gtcatat ^{ttt} tagaaataaa <u>GAAT</u> ACAAAAAAGTAAA aaaat ^{ttatc}	(1) 19 bp (1)	MHMR*
P10-17	NOV 0 μ M	KO1/3	59726 Recombinant 50847	ctttgaaaac <u>AAAA</u> AAAAAGGAAAGT catat ^{tttct} ggggatacaa <u>AAAA</u> AAAAAGGAAAGT catat ^{tttct} ggggatacaa <u>AAAA</u> AAAAAGGAAAGT attat ^{gaact}	(0) 15 bp (1)	MHMR*
P10-18	NOV 0 μ M	KO1/3	59953 Recombinant 50474	cgtcccagca <u>AATC</u> ATTTTTTTTC cacaagg ^{gat} agg ^{ttaatcg} <u>AATC</u> ATTTTTTTTC cacaagg ^{gat} agg ^{ttaatcg} <u>AATC</u> ATTTTTTTTC g ^{tttgg} ttgt	(0) 14 bp (0)	MHMR
P10-19	NOV 100 μ M	KO1/3	59724 Recombinant 51123	gagttacttt <u>GAAA</u> ACAAAAAAGGAAA gtcatat ^{ttt} tagaaataaa <u>GAAT</u> ACAAAAAAGGAAA gtcatat ^{ttt} tagaaataaa <u>GAAT</u> ACAAAAAAGTAAA aaaat ^{ttatc}	(1) 19 bp (1)	MHMR*
P10-20	NOV 100 μ M	KO1/3	59953 Recombinant 50474	cgtcccagca <u>AATC</u> ATTTTTTTTC cacaagg ^{gat} agg ^{ttaatcg} <u>AATC</u> ATTTTTTTTC cacaagg ^{gat} agg ^{ttaatcg} <u>AATC</u> ATTTTTTTTC g ^{tttgg} ttgt	(0) 14 bp (0)	MHMR
Plastid reaction 11 with primers 31381REV and 19309FOR						
Plastid reaction 12 with primers 69633REV and 79920FOR						
Plastid reaction 13 with primers 50513REV and 68821FOR						
P13-1	CIP 0 μ M	Col-0	50427 Recombinant 69421	agtcaattcc <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AAT_GGAGTGG</u> ATTGG atttgc ^{acca}	(1) 16 bp (1)	MHMR*
P13-2	CIP 0 μ M	KO1/3	50427 Recombinant 69421	agtcaattcc <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AAT_GGAGTGG</u> ATTGG atttgc ^{acca}	(1) 16 bp (1)	MHMR*
P13-3	CIP 0 μ M	KO1/3	50394 Recombinant 69096	aatgtg ^g aaa <u>AAATA</u> cccg ^t tagt aga ^{at} tg ^g at <u>AAATA</u> cccg ^t tagt aga ^{at} tg ^g at <u>AAATA</u> g ^t gg ^g ttcca	(0) 5 bp (0)	MHMR
P13-4	CIP 0.75 μ M	Col-0	50427 Recombinant 69421	agtcaattcc <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AAT_GGAGTGG</u> ATTGG atttgc ^{acca}	(1) 16 bp (1)	MHMR*
P13-5	CIP 0.75 μ M	KO1/3	50427 Recombinant 69421	agtcaattcc <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AATCGGAGTGG</u> ATTGG gcaagg ^{gata} ttttc ^t aaaa <u>AAT_GGAGTGG</u> ATTGG atttgc ^{acca}	(0) 16 bp (2)	MHMR*

^a Treatment of plants with ciprofloxacin (CIP) or novobiocin (NOV) at the indicated concentrations.

^b The junction endpoint is the position of the first nucleotide preceding the DNA junction. The accession numbers for the mitochondrial and plastid genomes are NC_001284 and NC_000932, respectively.

^c Alignment of the sequenced DNA junctions (center) with homologous sequences (top and bottom). Mismatches are underlined.

^d Product types were classified according to junction length with NHEJ-like for junction of less than five nucleotides, MHMR for junctions of five or more nucleotides and MHMR* for junctions of five or more nucleotides containing at least one mismatch.

^e These reactions were done twice with a ciprofloxacin treatment. Only nonredundant products were included in the table.

^f These reactions were done with plants treated with ciprofloxacin and with plants treated with novobiocin.

^g These primers anneal in the Large Inverted Repeats.

Chapitre 5. Discussion

5.1 Un rôle pour les Whirlies dans la stabilité du génome des organelles de plantes

La recherche présentée ici visait à comprendre le rôle des Whirlies dans les organelles de plantes. Les résultats obtenus au cours de cette étude suggèrent que les protéines de la famille Whirly ont un rôle conservé dans le maintien de la stabilité du génome des organelles de plantes. Nous avons démontré qu'une diminution importante du niveau d'expression des Whirlies plastidiques entraîne une accumulation de molécules réarrangées chez deux espèces de plantes éloignées d'un point de vue évolutif : *Arabidopsis thaliana* et *Zea mays*. De plus, la réparation des dommages dans les organelles semble s'effectuer plus souvent d'une manière non-conservatrice en l'absence de Whirlies.

5.1.1 Évidences pour un mécanisme de réparation semblable au MMBIR dans les organelles de plantes

Le mécanisme précis de la déstabilisation des génomes des organelles observée au cours de nos études n'est pas encore complètement compris mais plusieurs observations indiquent que les réarrangements sont probablement le résultat d'un mécanisme de recombinaison non-homologue basé sur la réplication.

Ainsi, les courtes répétitions à l'origine des réarrangements observés dans les organelles de nos plantes ont une taille moyenne d'environ 13 paires de bases. Chez *E. coli*, la recombinaison homologue entre deux molécules nécessite au moins 50 pb (283). Il est donc peu probable que les réarrangements dans les organelles soient produits par recombinaison homologue.

Ensuite, les réarrangements des génomes d'organelles observés impliquent des régions microhomologues distantes de 5 à 50 kb et semblent donc indépendants de la distance entre les répétitions, contrairement à d'autres types de réarrangements. Par exemple, le glissement réplicatif (« Replication Slippage ») est un mécanisme capable

d'utiliser de courtes régions répétées rapprochées semblables à celles trouvées dans nos études pour générer des duplications ou des délétions. Au cours de la réplication ces régions répétées sont susceptibles de provoquer un glissement du brin matriciel ou du brin en élongation (« template-switching »), ce qui produira des duplications ou des délétions des courtes répétitions selon la direction du glissement (**Figure 36**). Comme ces glissements ont lieu à l'intérieur d'une seule bulle réplivative, ils sont fortement dépendants de la distance entre les répétitions et sont virtuellement éliminés lorsque celle-ci devient supérieure à 10 kb. Nos réarrangements utilisent donc probablement un mécanisme de template-switching mais celui-ci est capable de fonctionner entre des courtes répétitions très éloignées les unes des autres.

Récemment, le modèle du « microhomology-mediated break-induced replication » (MMBIR) a été proposé pour expliquer la production de nombreux réarrangements complexes dans les génomes de la bactérie, de la levure et de l'homme (121) (**Figure 37**). Comme dans le glissement réplivative, une fourche de réplication bloquée est réinitialisée par template-switching médié par des régions de microhomologies. Cependant, dans le MMBIR, le changement de matrice est indépendant de la distance, c'est-à-dire qu'il peut s'effectuer sur de longues distances sur la même molécule d'ADN ou même sur une autre molécule. Le changement de matrice peut aussi survenir plusieurs fois de façon consécutive, ce qui permet la production de réarrangements hautement complexes. Dépendamment de l'endroit où s'effectue le changement de matrice par rapport à la fourche réplivative initiale, il peut y avoir production de délétions, de duplications, d'inversions et génération de cercles-roulants qui entraînent la production de concatémères. Cette variété de réarrangements corrèle bien avec ce que nous avons observé dans les plastides de nos plantes KO1/3. Le MMBIR partage donc de nombreuses caractéristiques avec le mécanisme de réarrangements des génomes d'organelles décrit ici. Nous proposons donc qu'un mécanisme semblable au MMBIR existe dans les organelles de plantes et peut être utilisé pour réparer leurs génomes.

Une autre possibilité serait que la recombinaison non-homologue observée dans les mutants de Whirlies utilise un mécanisme de microhomology-mediated end-joining (MMEJ) (**Figure 6**). Jusqu'à présent, le MMEJ a été décrit comme produisant uniquement des délétions en reliant les extrémités d'un même bris double-brin (revue dans (112)). Des événements de délétions ont bien été détectés au cours de nos études, mais nous avons aussi observé plusieurs événements de circularisation/duplication et de concatémérisation. Pour expliquer l'apparition de telles molécules à l'aide d'un mécanisme de type MMEJ, il faudrait postuler l'existence de deux bris double-brins indépendants qui seraient reliés ensemble via leurs régions microhomologues. Un tel mécanisme est cependant moins parcimonieux que le MMBIR.

5.1.2 Quels sont les fonctions moléculaires possibles des Whirlies dans le maintien de la stabilité du génome des organelles ?

Plusieurs hypothèses peuvent être avancées pour expliquer l'augmentation du nombre de molécules réarrangées suite à la délétion des Whirlies. Il est possible que l'absence de Whirlies entraîne une augmentation des bris dans l'ADN des organelles provoquée par une plus grande accessibilité, et donc vulnérabilité de l'ADNsb. Les événements de BIR étant favorisés lorsque des bris double-brins à une seule extrémité sont produits, il est également possible que l'absence de Whirlies entraîne une plus haute fréquence d'arrêt de la fourche répliquative. Ces arrêts pourraient favoriser la réversion de la fourche ce qui produirait des bris double-brins à une seule extrémité (revue dans (284)). Cependant, des expériences préliminaires d'analyse des bris d'ADN sur gel alcalin indiquent que la fréquence d'apparition de bris dans l'ADNpt des plantes KO1/3 n'est pas significativement plus élevée que dans les plantes WT en absence ou en présence de ciprofloxacine, un agent capable d'induire des bris double brins dans l'ADN des organelles de plantes. De plus, des essais de susceptibilité à la DNase I de l'ADNpt n'ont pas révélé de différences significatives en présence ou en absence d'AtWhy1/3 (Lepage, É. et Maréchal, A., résultats non-publiés). Ces résultats couplés à la plus grande augmentation des MHMR (« MicroHomology-Mediated Rearrangements ») dans les mutants de

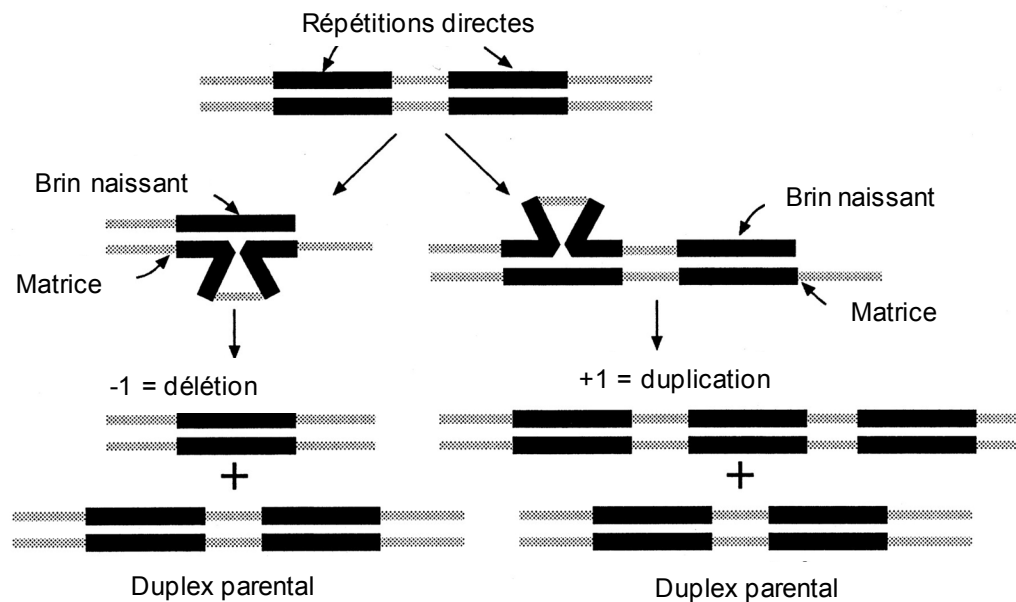


Figure 36. Modèle du glissement répliatif à l'origine de certains réarrangements génomiques. (Figure adaptée de (285)). Des erreurs peuvent survenir lorsqu'une fourche répliatrice change de matrice lors de l'élongation. Ces changements de matrice se font normalement entre des régions répétées. Lorsque ces séquences répétées sont relativement rapprochées l'une de l'autre, il peut y avoir glissement des brins à l'intérieur de la bulle répliatrice qui peuvent se réhybrider au niveau des différentes répétitions. Dépendamment du type de glissement il y a production de délétions ou de duplications de la région répétée.

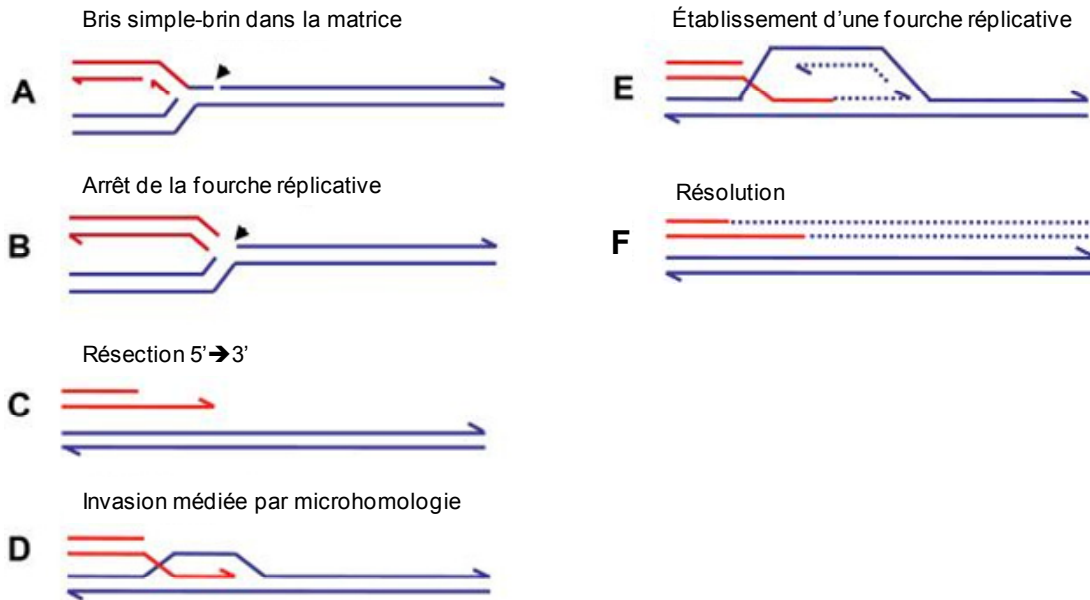


Figure 37. Modèle du « microhomology-mediated break-induced replication » (MMBIR). Lorsque la fourche répliquative rencontre une lésion, il peut y avoir production d'un bris double-brin à une seule extrémité. Ce bris double-brin est ensuite résecté pour produire une extrémité 3'-OH simple-brin libre. Cet ADNsb peut alors envahir une autre matrice via des régions de microhomologie. Une fourche répliquative peut éventuellement être créée ce qui résultera en la production d'une molécule réarrangée. De multiples sauts de la polymérase peuvent avoir lieu, ce qui causera des réarrangements complexes (Figure adaptée de (121)).

Whirlies suite à l'induction de bris double-brins par la ciprofloxacine supportent plutôt un rôle pour les Whirlies dans la réparation de ces bris.

L'accumulation de MHMR pourrait donc être due à une fonction des Whirlies dans le choix des voies de réparation des bris. Dans ce cas, les Whirlies pourraient être importants dans la promotion de la recombinaison homologe ou encore limiter activement la recombinaison non-homologue à l'origine des MHMR. Pour distinguer entre ces deux possibilités, il serait possible de mettre au point un système rapporteur dans des chloroplastes transgéniques permettant de mesurer l'efficacité de différents types de recombinaison en présence ou en absence de Whirlies (132). De tels systèmes existent mais la transformation routinière des chloroplastes demeure pour l'instant confinée au tabac. Néanmoins, on pourrait envisager d'intégrer deux morceaux d'un gène rapporteur tel celui de la luciférase dans le génome chloroplastique du tabac. Ces fragments contiendraient des régions d'homologies plus grandes ou plus petites que 50 pb afin de mesurer les fréquences de recombinaison homologe et non-homologue respectivement. L'un de ces fragments pourrait contenir un site de reconnaissance spécifique pour une enzyme de restriction. En transformant les plantes avec une construction inductible de l'enzyme de restriction fusionnée à un PT plastidique, il serait possible d'induire un bris qui favoriserait la recombinaison entre les deux fragments de la luciférase. Cette recombinaison serait évaluée en produisant des extraits de feuilles et en mesurant l'activité luciférase présente. Pour évaluer l'importance des Whirlies dans la régulation de ces deux types de recombinaison, il serait possible d'utiliser le VIGS pour diminuer l'expression des Whirlies plastidiques du tabac. Malgré l'effort requis pour sa mise en place, un tel système pourrait être utilisé pour décortiquer la machinerie de recombinaison des chloroplastes.

Il est intéressant de noter que l'activité de liaison à l'ADN la plus abondante du plastide semble correspondre à celle des Whirlies. En effet, lorsqu'un essai de retard sur gel est réalisé en utilisant une sonde d'ADNsb et des extraits plastidiques, le principal complexe observé est celui formé entre les Whirlies et l'ADNsb (**Figure 20**). Cette

prépondérance des Whirlies pourrait leur permettre d'interagir avec l'ADNsb produit suite à des bris dans l'ADN des plastides.

5.1.3 Un modèle pour élucider la fonction des Whirlies dans le maintien de la stabilité du génome des organelles.

À l'aide des données expérimentales générées lors de mes études doctorales et de la structure cristallographique du complexe StWhy2-ADNsb, nous proposons un modèle de travail afin d'élucider plus précisément le mécanisme par lequel les Whirlies préviennent l'accumulation de MHMR dans les organelles.

Lorsqu'un bris double-brin survient dans le génome plastidique, il subit une résection qui permet la production de fragments d'ADNsb 3'-OH. En présence de Whirlies, ces extrémités simple-brin seraient reconnues et protégées ce qui empêcherait l'interaction de l'ADNsb avec de l'ADN complémentaire ou d'autres protéines. Par un mécanisme encore inconnu, la machinerie de recombinaison homologue serait ensuite recrutée au site de dommage qu'elle pourrait réparer de façon conservatrice, favorisant ainsi la stabilité du génome. Par contre, en l'absence de Whirlies, l'ADNsb serait libre d'interagir avec d'autres régions simple-brin via des régions de microhomologies ou encore de se lier avec d'autres protéines, ce qui favoriserait la production de MHMR et donc une déstabilisation du génome (**Figure 38**).

En support à ce modèle, on note qu'en l'absence de plusieurs exonucléases simple-brin, une voie de recombinaison homologue indépendante de *recA* est révélée chez *E. coli* (286). Ceci suggère que la disponibilité de l'ADNsb est un élément limitant l'utilisation de cette voie en temps normal. Cette nouvelle voie est capable de réparer l'ADN par recombinaison en utilisant des fragments longs d'ADN (recombinaison homologue) mais elle est également capable d'utiliser de courtes séquences d'homologie (25 pb). Une telle voie parallèle, possiblement analogue au BIR et au MMBIR, pourrait être activée dans les mutants de Whirlies où la disponibilité de l'ADNsb est augmentée. À cause d'un

racourcissement de la région d'homologie nécessaire à la recombinaison, il y aurait alors une plus grande production de molécules réarrangées.

Le modèle que nous proposons permet un certain nombre de prédictions qui, lorsque vérifiées expérimentalement, permettront de mieux comprendre l'implication des Whirlies dans la stabilité des génomes. Une première prédiction du modèle est que les mutants de Whirlies devraient être plus sensibles aux dommages à l'ADN dans les organelles. Nous avons effectivement pu déterminer que les plantes KO1/3 étaient plus sensibles à la ciprofloxacine. Le traitement de mutants de Whirlies avec cette molécule entraîne également une plus grande accumulation de MHMR. Afin de pousser plus loin l'exploration du rôle des Whirlies dans le maintien de la stabilité du génome des organelles, plusieurs expériences pourraient être effectuées. Par exemple, il serait intéressant de croiser les mutants de Whirlies avec d'autres mutants ayant des niveaux endogènes de dommages oxydatifs élevés et avec des mutants d'autres protéines potentiellement impliquées dans le maintien de la stabilité du génome des organelles (106, 142, 194, 287). Une diminution de la vitesse de réparation des dommages à l'ADN des organelles dans des mutants de Whirlies, ainsi qu'une augmentation du contenu en Whirlies suite à une exposition à des agents génotoxiques supporteraient également l'implication des Whirlies dans la réparation des génomes (155).

Selon le modèle, les Whirlies devraient aussi s'accumuler à proximité des régions d'ADN endommagées. Un moyen de vérifier cette prédiction serait d'immunoprécipiter les Whirlies afin de déterminer s'ils sont liés plus ou moins fortement avec l'ADN en présence ou en absence de bris. Il serait également possible de mettre au point un système permettant de générer des bris à des endroits précis du génome des organelles. La production de plantes transgéniques dans lesquelles une enzyme de restriction fusionnée à un PT sous le contrôle d'un promoteur inductible nous permettrait un plus grand contrôle sur ce type d'expériences. Suite à l'induction de bris, il serait possible de suivre le recrutement des Whirlies aux sites de dommage par immunoprécipitation de l'ADN et PCR.

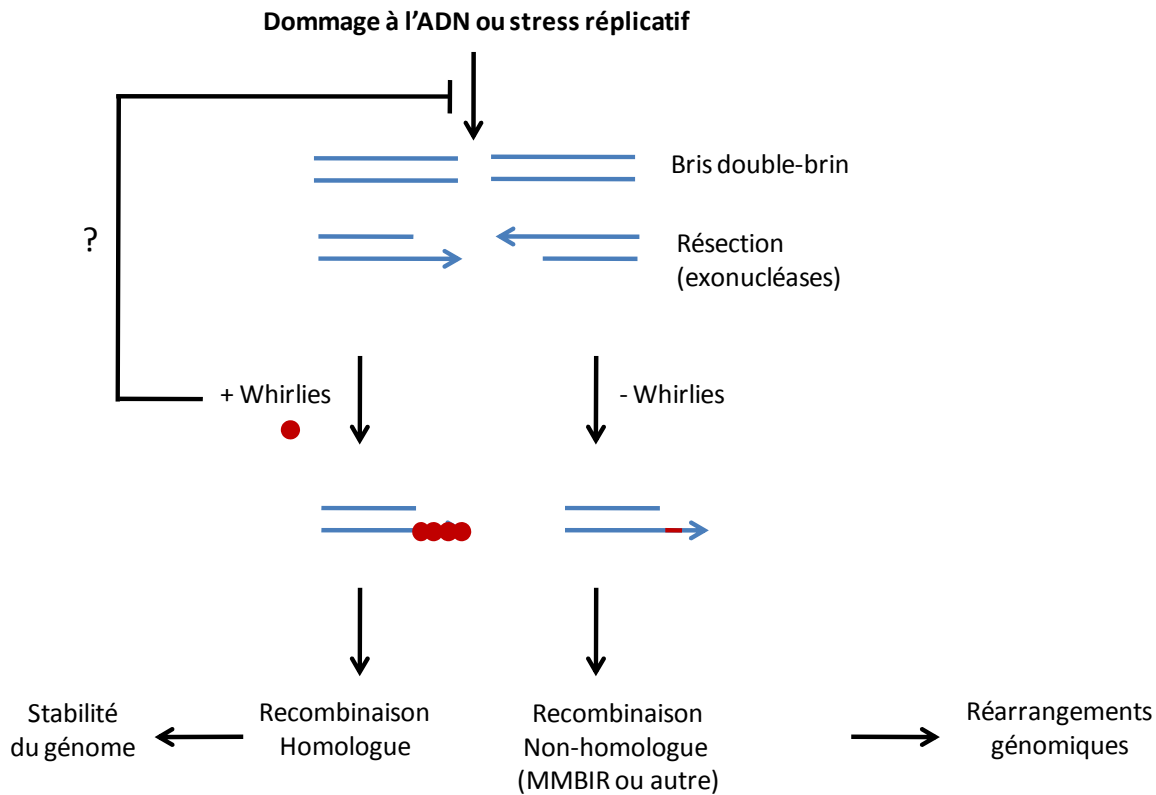


Figure 38. Modèle du mécanisme d'action des Whirlies dans le maintien de la stabilité du génome des organelles. Différents stress peuvent produire des bris double-brins dans l'ADN. Ces bris double-brins sont transformés en extrémités 3'-OH simple-brin libres par des exonucléases. En présence de Whirlies, l'ADNs_b serait protégé contre les interactions avec de l'ADN complémentaires ou d'autres protéines. Ceci permettrait à un mécanisme de recombinaison homologue de réparer le génome de façon conservatrice. En absence de Whirlies, des interactions non-régulées avec d'autre ADN via des régions de microhomologie ou d'autres protéines favoriseraient les réarrangements du génome via la recombinaison non-homologue. Bien que des résultats préliminaires suggèrent le contraire, il est possible que les Whirlies préviennent l'apparition de lésions sur l'ADN ou diminuent l'incidence de stress réplicatif.

L'élaboration d'un tel système permettrait aussi d'évaluer la localisation d'autres protéines potentiellement impliquées dans la recombinaison des génomes d'organelles et de déterminer l'ordre dans lequel elles sont recrutées aux sites de dommage (288, 289).

Une autre possibilité soulevée par le modèle est que les Whirlies pourraient aussi être impliqués dans le recrutement de la machinerie de recombinaison. En effet, les protéines de liaison à l'ADNsb SSB chez les bactéries et RPA chez les eucaryotes sont capables d'attirer une variété impressionnante de protéines à proximité des régions endommagées de l'ADN (revue dans (139) et (144)). Dans cette optique, il est important de noter que le domaine C-terminal acide/hydrophobe conservé des Whirlies d'angiospermes compétitionne avec l'ADNsb pour la surface de liaison de l'ADN. Ceci suggère que lorsque les Whirlies sont liés à l'ADNsb, le domaine C-terminal pourrait être libre d'interagir avec des partenaires protéiques des Whirlies. Un tel mécanisme élégant est retrouvé chez le SSB eubactérien qui utilise son domaine C-terminal acide/hydrophobe pour recruter plusieurs protéines (290, 291). La découverte d'interacteurs des Whirlies par double-hybride et purification par affinité permettra de mieux comprendre cet aspect du modèle et éventuellement d'explorer d'autres fonctions des Whirlies dans la cellule.

5.2 Quelles sont les autres fonctions des Whirlies dans les organelles ?

Malgré les évidences que les Whirlies préviennent l'accumulation de MHMR, il est possible que ces protéines aient d'autres fonctions dans les organelles. Comme les Whirlies ont une affinité particulière pour l'ADNsb, ils pourraient jouer des rôles non-essentiels lors de la réplication et de la transcription de l'ADN dans les organelles. Effectivement, plusieurs liens ont été établis entre des perturbations de la réplication et de la transcription et une augmentation de l'instabilité génomique (revue dans (292)). Notamment, la mutation de plusieurs protéines impliquées dans la réplication entraîne une augmentation de la recombinaison et/ou des réarrangements génomiques dans une variété d'organismes. Un traitement avec des inhibiteurs de la réplication augmente aussi l'instabilité génomique

(293). Ces traitements ou ces mutations favorisent des arrêts de la fourche de réplication qui peuvent être rescapés par des événements de changements de matrice médiés par recombinaison homologue. Une augmentation des arrêts de la réplication des génomes d'organelles en absence des Whirlies pourrait entraîner une augmentation des réarrangements en surtaxant la machinerie de HR, ce qui favoriserait l'utilisation de voies de réparation alternatives telles que le MMBIR.

Il est bien établi que la transcription stimule la recombinaison (revue dans (294)). Plusieurs intermédiaires produits lors de la transcription peuvent favoriser la recombinaison. En effet, la transcription entraîne l'accumulation de surenroulement négatif en amont de l'ARN polymérase, ce qui favorise une ouverture de la double-hélice d'ADN. Ceci crée de l'ADNsb qui est plus susceptible aux dommages et prompt à la formation de structures secondaires qui viendraient par la suite perturber la réplication. Par ailleurs, le brin non-transcrit est aussi présent sous forme simple-brin au cours de la réplication. La protection de ces régions sensibles par les Whirlies pourrait diminuer la fréquence de réarrangements. Il est aussi possible que les Whirlies favorisent la progression de la fourche transcriptionnelle de manière à diminuer la durée d'exposition des régions simple-brin durant la transcription.

Malgré les liens évidents entre la réplication, la transcription et la recombinaison, les évidences en faveur d'une implication des Whirlies dans la réplication et/ou la transcription des gènes d'organelles sont ténues et les résultats obtenus dans différents systèmes sont parfois contradictoires. La surexpression d'*AtWhy2* entraîne une diminution importante de la quantité d'ADN et d'ARN mitochondriaux. Par contre, l'absence d'*AtWhy2* n'entraîne pas de modifications importantes dans les niveaux d'acides nucléiques de la mitochondrie (295). La quantité d'ADNpt dans un mutant très sévère de *ZmWhy1* est fortement diminuée mais ce n'est pas le cas dans les plantes KO1/3 d'*Arabidopsis*. De plus, une diminution de la quantité de *ZmWhy1* chez les mutants de maïs les moins affectés semble augmenter légèrement le niveau d'expression de plusieurs gènes plastidiques, phénomène qui n'est cependant pas observé chez les plantes KO1/3

d'*Arabidopsis* (262). L'activité transcriptionnelle de plusieurs gènes chloroplastiques a été évaluée chez le maïs mais elle ne semblait pas être affectée significativement dans des mutants de *ZmWhy1*. Finalement, bien que l'expression de tous les gènes d'organelles examinés chez *Arabidopsis* ne soit pas affectée par l'absence des Whirlies plastidiques ou mitochondriaux, nous n'avons examiné qu'environ 10 % des gènes des organelles dans ces mutants. Des études plus poussées pourront permettre d'évaluer plus précisément la contribution des Whirlies au bon déroulement de la réplication et de la transcription des génomes d'organelles mais pour le moment cette contribution semble au mieux n'être qu'accessoire.

La protéine de liaison à l'ADNsb RPA étant impliquée dans l'activation du point de contrôle des dommages à l'ADN dans le génome nucléaire eucaryote, il serait intéressant de déterminer si un tel point de contrôle existe chez les organelles et quel serait le rôle des Whirlies dans ce processus (revue dans (144)). Il est bien établi que plusieurs gènes potentiellement impliqués dans la réparation de l'ADN dans les organelles de plantes sont induits suite à un traitement de la plante avec des agents génotoxiques qui endommagent les trois génomes de la cellule (296-298). Cependant, il serait important de vérifier si l'expression de ces gènes est également induite lorsque des dommages sont provoqués spécifiquement au niveau du génome des organelles, par exemple suite à une exposition à la ciprofloxacine. Ceci révélerait l'existence d'un système de communication organelles-noyau permettant d'augmenter la résilience du génome des organelles en présence de stress. Il sera important de déterminer le niveau d'activation des voies de réparation de l'ADN des organelles en présence ou en absence de Whirlies et/ou d'agents génotoxiques.

Les études de l'homologue de Whirly TIF1 apportent des arguments en faveur d'un rôle pour les Whirlies dans un éventuel point de contrôle des dommages à l'ADN dans les organelles. En effet, chez *T. thermophila*, TIF1 est important pour l'activation de la réponse de dommage médiée par le senseur de dommages/kinase ATR (265-266). Suite à un traitement à l'hydroxyurée, les mutants de TIF1 sont incapables d'activer le point de contrôle intra-phase S qui permet normalement de réparer les dommages à l'ADN

encourus. En effet, malgré des dommages importants aux chromosomes, les mutants sortent quand même de la phase S. Ces mutants de TIF1 sont aussi hypersensibles à l'hydroxyurée et au méthylméthanesulfonate, des agents génotoxiques et les voies de réparation de l'ADN sont déjà préactivées dans des mutants de TIF1 en absence de stress. Bien qu'assez éloignés d'un point de vue évolutif, de nombreux parallèles peuvent être tracés entre les fonctions de TIF1 dans les deux noyaux de *Tetrahymena* et celles des Whirlies dans les organelles des plantes à fleurs et il sera intéressant de voir jusqu'à quel point ces protéines se ressemblent au niveau de leurs rôles dans la cellule.

5.3 Spécificité de liaison des Whirlies aux acides nucléiques

Les Whirlies ont d'abord été décrits comme possédant une certaine spécificité dans leurs interactions avec l'ADNsb dans le noyau. Cette spécificité est fortement soutenue par des expériences d'immunoprécipitation de la chromatine qui ont démontré que la liaison de StWhy1 au promoteur du gène *PR-10a* dans les tubercules de pomme de terre avait lieu uniquement suite à des stress reliés à la réponse de défense (247-251).

Dans le laboratoire du Dr. Brisson, StWhy1 a été purifié par affinité à une forme simple-brin de l'élément de réponse à l'éliciteur (ERE) (5'-CTAGACCATTTTTGACAT TTGTGTCATTTTATCTAG-3') (249). Un autre laboratoire a été en mesure de purifier StWhy2, le Whirly mitochondrial de pomme de terre en utilisant une colonne d'ADNsb de séquences aléatoires. Finalement, AtWhy1 a pu être isolé par affinité pour la répétition hexamérique des télomères d'*Arabidopsis* (5'-TTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGG-3') (159, 261). Ces différentes stratégies d'isolation indiquent que les Whirlies peuvent se lier à des séquences d'ADN passablement différentes. Ceci est soutenu par des études d'immunoprécipitation qui ont révélé une faible spécificité de séquence dans les interactions entre l'ADN des organelles et les Whirlies (262, 295). À la limite, ces expériences suggèrent que les Whirlies pourraient être associés à tout le génome des organelles. En accord avec le comportement des Whirlies dans les organelles, l'étude cristallographique de la protéine StWhy2 en complexe avec différentes molécules

d'ADNsb a révélé un mécanisme de liaison aux acides nucléiques simple-brin généralement séquence-indépendant basé surtout sur des interactions d'empilement des bases (« base stacking »). De plus, la constante de dissociation de StWhy2 pour plusieurs oligonucléotides différents varie peu et se situe aux environs de 3-5 nM.

Ces observations semblent indiquer qu'intrinsèquement, les Whirlies n'ont pas une grande spécificité de reconnaissance au niveau de la séquence des nucléotides des acides nucléiques. La liaison plus spécifique des Whirlies au promoteur de *PR-10a* et aux répétitions télomériques pourrait s'expliquer par le caractère simple-brin naturellement plus élevé de ces régions *in vivo*. En effet, les télomères possèdent de longues portions d'ADNsb à leurs extrémités et le ERE est relativement riche en A/T et contient une région palindromique ce qui pourrait augmenter son caractère simple-brin (249, 299). Une modification du contexte particulier de la chromatine dans le promoteur suite à l'activation de la transcription de *PR-10a* pourrait aussi faciliter la liaison des Whirlies. De façon similaire, la protéine TIF1, un homologue des Whirlies, se lie aux éléments de type I du minichromosome de l'ADN ribosomal de *T. thermophila* à proximité d'un promoteur (266). Ces séquences sont riches en A/T et sont situées dans des régions sans nucléosomes. Le caractère A/T-riche des éléments de type I couplé à cette absence d'histones est susceptible de favoriser l'ouverture de la double-hélice qui pourrait alors être stabilisée en conformation ouverte par la liaison de TIF1. Une augmentation de l'activité transcriptionnelle en aval augmente le stress torsionnel dans les séquences situées en amont, favorisant la séparation des deux brins (revue dans (300)). Ainsi, dans les deux cas décrits ci-haut, une augmentation du surenroulement négatif en amont du promoteur suite à l'activation de la transcription entraînerait une ouverture de la double-hélice favorisée par la forte proportion en résidus A/T du ERE et des éléments de type I. Ceci rendrait ces régions plus disponibles pour être liées par StWhy1 ou TIF1. Une autre possibilité envisageable serait que les Whirlies soient recrutés à certains endroits spécifiques du génome nucléaire par d'autres protéines. Ces deux hypothèses permettent d'expliquer la localisation spécifique des Whirlies à certains endroits du génome sans invoquer un mécanisme de reconnaissance spécifique de certaines séquences nucléotidiques.

L'identification de nouvelles cibles transcriptionnelles nucléaires, la démonstration *in vivo* de la liaison des Whirlies avec les télomères ainsi que l'identification de partenaires protéiques pourra nous permettre de mieux comprendre les déterminants des interactions entre les Whirlies et l'ADN *in vivo*.

En plus de se lier à des séquences représentant l'ensemble du génome plastidique, ZmWhy1 est capable de lier certains ARNs plastidiques *in vivo* (262). ZmWhy1 s'associe principalement à un intron dans l'ARN du gène *atpF* mais interagit également avec les ARNs des gènes *rps12*, *ndhA*, *rpl16*, *ycf3* et *rps14* dans les chloroplastes. *In vitro*, ZmWhy1 est capable de lier l'intron de l'ARN du gène *atpF* mais se lie aussi à d'autres ARNs de séquences différentes avec une affinité similaire. AtWhy1 possède aussi une certaine affinité pour l'ARN (10 à 100 X moins élevée que pour l'ADNsb). Des essais de compétitions ont cependant révélé que lorsqu'AtWhy1 est lié à de l'ADNsb, une séquence équivalente d'ARN est incapable de déplacer l'ADNsb tandis que l'ADNsb remplace facilement l'ARN sur la surface de liaison des acides nucléiques des Whirlies (Maréchal A. et Cappadocia L, résultats non-publiés (**Figure 39**)). Ces résultats démontrent une préférence des Whirlies pour l'ADNsb mais suggèrent également que la capacité de lier l'ARN est une propriété commune à tous les Whirlies de plantes. Des expériences d'immunoprécipitation d'ARN avec les Whirlies d'*Arabidopsis* suivies de Northern blots ou d'hybridation sur des puces à ADN permettraient de déterminer si la liaison d'ARN dans les organelles est une fonction conservée des Whirlies *in vivo* (301).

Récemment, des protéines importantes dans la maturation des ARNs ont été impliquées dans le maintien de la stabilité du génome chez les mammifères et les levures ((302-304) et revue dans (305)). Une déstabilisation du génome serait provoquée par la formation plus fréquente de boucles R stables lors de l'élongation de l'ARN messager lorsque la maturation de celui-ci est perturbée. Les boucles R sont habituellement formées lorsque l'ARN transcrit forme un duplex avec son brin matrice d'ADN, un processus favorisé par le surenroulement négatif présent en amont de la polymérase.

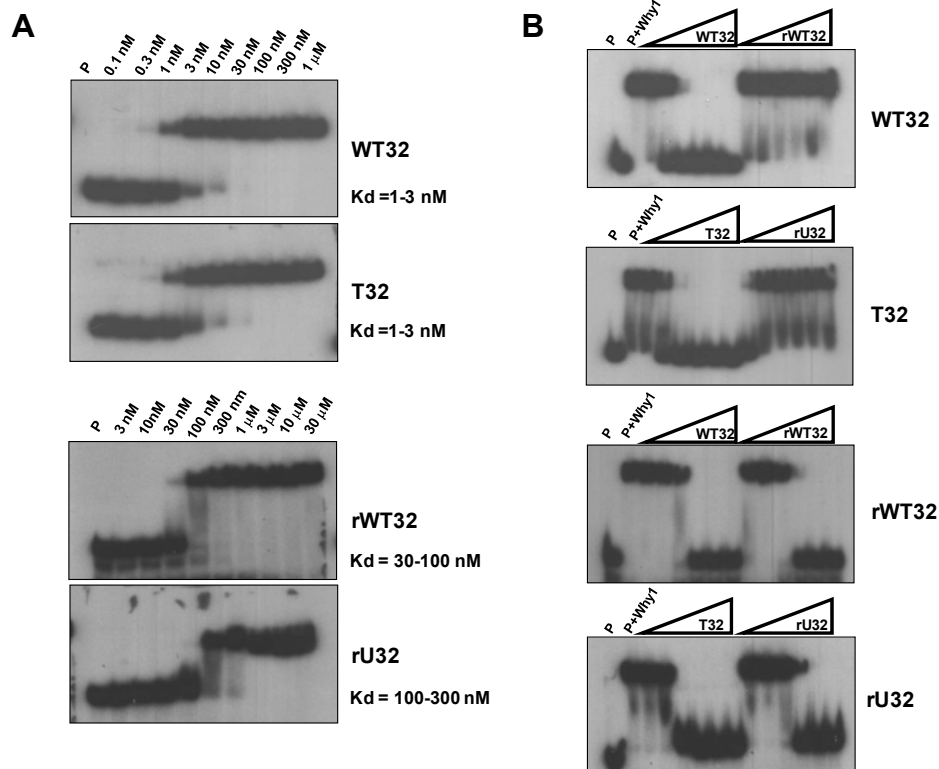


Figure 39. Affinité des Whirlies pour les acides nucléiques simple-brin. **A.** Affinité d'AtWhy1 pour différents acides nucléiques simple-brin. Des concentrations croissantes de la protéine recombinante ont été incubées avec une quantité donnée de différents oligonucléotides radiomarqués. Les complexes ont ensuite été séparés sur un gel non-dénaturant d'acrylamide 10 %. Les constantes de dissociation ont été estimées visuellement. **B.** AtWhy1 se lie préférentiellement à l'ADNs. 20 nM (pour l'ADNs) ou 300 nM (pour l'ARN) de protéine recombinante a été préincubée avec différents oligonucléotides radiomarqués de façon à ce que les sondes soient complètement liées par AtWhy1. Des concentrations croissantes d'oligonucléotides froids ont été ajoutées passant de 0,1 pM à 0,1 μ M. Les complexes ont ensuite été séparés sur un gel non-dénaturant. P; probe. ADN: WT32; 5'-TGTCATTTTGTGTCATTTTGTGTCATTTTGTGCA-3', T32; 5'- TTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTT-3', ARN: rWT32; 5'-UGUCAUUUUUGUCAUUUUUGUCAUUUUUGUCA-3', rU32; 5'-UUUUUUUUUUUUUUUUUUUUUUUUUUUU-3'.

Une augmentation de la stabilité de telles structures pourrait entraîner des arrêts de la fourche répliquative et favoriser les bris dans l'ADNsb du brin non-transcrit. Il serait donc possible que les Whirlies, en stimulant la maturation des ARNs des organelles, limitent les réarrangements de leurs génomes.

Les nucléoïdes des organelles de plantes à fleurs ne possèdent pas d'homologues des protéines structurales principales retrouvées chez les bactéries (HU, H-NS, IHF...). Comme les Whirlies semblent être associés non-spécifiquement à l'ADN des organelles, elles pourraient prendre en charge certaines des fonctions des protéines structurales bactériennes. Plusieurs ressemblances existent entre les Whirlies et la protéine HU (Histone-like U93), une composante majeure des nucléoïdes bactériens. HU est associée non-spécifiquement à l'ADN double-brin mais possède une affinité particulière pour l'ADNsb et plusieurs intermédiaires de la réparation et de la recombinaison (306, 307). HU est aussi capable de lier certains ARNs *in vitro* et *in vivo* (308, 309).

Cette protéine est impliquée dans la recombinaison, la réparation, la protection contre les dommages et la régulation de la topologie de l'ADN bactérien (310-312). Un rôle architectural et/ou topologique éventuel des Whirlies pourrait donc avoir un impact sur le maintien de la stabilité du génome. Pour établir un tel rôle, d'autres expériences sont nécessaires. Par exemple, il faudrait observer la structure de l'ADN des organelles en présence ou en absence de Whirlies en utilisant l'électrophorèse en champ pulsé ou la microscopie électronique. Des essais de sédimentations des nucléoïdes permettraient également de détecter une éventuelle modification de leur état de condensation lorsque les Whirlies sont absents (313). La perte de fonction de HU chez *E. coli* peut être compensée par des mutations dans les gyrases permettant la réintroduction d'un surenroulement négatif et le retour à une topologie plus ou moins normale de l'ADN. Il serait intéressant de voir si des mécanismes compensatoires similaires sont activés dans les organelles de plantes en absence de Whirlies. La complémentation d'une perte de HU par l'expression des Whirlies dans la bactérie serait aussi une expérience pertinente à réaliser.

5.4 Localisation intracellulaire des Whirlies

De nombreux complexes protéiques localisés dans les organelles sont composés de sous-unités encodées par le génome des organelles et par le génome nucléaire. Afin d'optimiser l'assemblage de ces complexes, un certain nombre de mécanismes appelés globalement signalisation antérograde permettent au noyau d'adapter les niveaux des protéines ciblées vers les organelles en fonction de l'état métabolique de la cellule. Des signaux moléculaires issus des organelles sont aussi détectés par le noyau afin que celui-ci réagisse de façon appropriée aux besoins des plastides et des mitochondries : il s'agit de la signalisation rétrograde (revue dans (314-316)).

Les médiateurs protéiques permettant l'intercommunication organelles-noyau demeurent mal connus, mais des candidats potentiels intéressants seraient des protéines localisées à la fois aux organelles et au noyau (revue dans (317)). Les Whirlies font partie de ces candidats avec des rôles et une localisation démontrés dans les plastides, les mitochondries et le noyau ((251, 254, 260, 262, 295, 318)).

Les Whirlies ciblés aux plastides étant uniquement détectables sous leur forme mature, il est possible qu'ils soient d'abord ciblés aux plastides où leur PT serait clivé. Ils pourraient ensuite être relocalisés au noyau dans certains tissus particuliers (i.e. des tissus non-photosynthétiques comme les tubercules de pomme de terre) ou suite à certains stress. Une telle relocalisation pourrait s'effectuer via les stromules, des extensions du stroma qui sont souvent observées entre les différents plastides d'une même cellule et qui peuvent également relier un plastide à la membrane nucléaire (revue dans (319)). Pour le moment, aucune évidence en faveur d'une localisation nucléaire des Whirlies mitochondriaux n'a été rapportée. Toutes les expériences effectuées jusqu'à présent, dans lesquelles un Whirly pleine longueur a été fusionné à la GFP ont toujours montré les Whirlies exclusivement localisés dans les organelles (253, 254). Cependant, des expériences d'immunolocalisation indiquent que HvWhy1, un Whirly plastidique de l'orge est localisé dans le noyau et les chloroplastes d'une même cellule de feuille. Des expériences d'immunoprécipitation de la

chromatine ont aussi démontré une localisation nucléaire des Whirlies dans les tubercules de pomme de terre (251). Il est donc possible que l'ajout d'une GFP en C-terminal empêche la relocalisation des Whirlies dans le noyau. Des expériences d'immunolocalisation pourraient être effectuées en utilisant les anticorps α -AtWhy1/3 et α -AtWhy2 afin de déterminer si la double localisation des Whirlies est également présente chez les Whirlies plastidiques et mitochondriaux d'*Arabidopsis*.

Plusieurs fonctions dans les différents compartiments de la cellule ont été proposées jusqu'à présent pour les Whirlies. Parmi celles-ci, le maintien de la stabilité du génome des organelles semble être la mieux conservée. Cette fonction est effectivement partagée par les Whirlies plastidiques et mitochondriaux d'*Arabidopsis* ainsi que par le Whirly plastidique du maïs, suggérant sa conservation chez toutes les plantes à fleurs. Des fonctions dans la régulation transcriptionnelle des gènes de défense chez la pomme de terre, dans l'homéostasie des télomères chez *Arabidopsis* et dans la maturation des ARNs chez le maïs ont également été démontrées. Il semblerait donc que la capacité de liaison aux acides nucléiques simple-brin soit à la base d'une grande flexibilité fonctionnelle de cette famille de protéine. L'étendue de la conservation de ces autres rôles demeure cependant à explorer.

Un lien intéressant peut être fait entre les rôles d'AtWhy1 dans le maintien de la stabilité du génome des organelles et dans l'homéostasie des télomères nucléaires. En effet, les télomères peuvent être allongés via différents mécanismes dépendants ou indépendants de la télomérase. Le mécanisme d'élongation des télomères indépendant de la télomérase retrouvé chez la levure et les humains ressemble beaucoup au BIR et dépend de la machinerie de recombinaison (revue dans (115)). La ré-invasion d'une séquence télomérique par son extrémité simple-brin favorise l'établissement d'une fourche de réplication qui permet l'élongation des télomères sans avoir recours à la télomérase. Comme un KO d'AtWhy1 entraîne une augmentation de la longueur des télomères, AtWhy1 pourrait se lier à la partie simple-brin des télomères et empêcher l'élongation des télomères indépendante de la télomérase (260). Il serait donc possible qu'AtWhy1 agisse

comme suppresseur de la recombinaison en se liant aux extrémités simple-brin libres dans deux compartiments différents de la cellule.

5.5 Maintien de la stabilité génomique, maladies et évolution

Des défauts dans le maintien de la stabilité des génomes sont à la base de plusieurs pathologies humaines, incluant le cancer (320). Au cours de notre étude, nous avons démontré qu'une diminution de la stabilité du génome plastidique entraîne des réarrangements importants dans l'ADN qui, à terme, compromettent sérieusement le fonctionnement des chloroplastes. À l'échelle d'un organisme, il semble donc préférable de maintenir un génome parfaitement intact. Cependant, un certain niveau d'instabilité génomique est crucial pour l'évolution des espèces puisque cela génère une diversité génétique entre les individus. Les phénotypes conférés par cette variabilité peuvent ensuite faire l'objet d'une sélection permettant la dissémination de nouveaux allèles dans une population.

L'instabilité génomique engendrée par une absence de Whirlies permet la production d'une grande diversité de nouveaux génomes. Bien que ceux-ci soient présents en faible quantité, nous avons été en mesure de détecter des génomes réarrangés dans les plastides de plantes sauvages. Des génomes réarrangés contenant habituellement des délétions existent aussi dans les mitochondries humaines (321). Bien que cela demeure controversé, leur accumulation serait un des facteurs favorisant le vieillissement et le développement de certaines maladies neurodégénératives. Cette situation est similaire à l'apparition de la variégation consécutive à une accumulation de MHMR.

Plusieurs explications sont possibles pour relier l'accumulation de génomes réarrangés et le mauvais fonctionnement des organelles. Dans les mitochondries humaines, des mutations ponctuelles ou la délétion de certains gènes impliqués dans la chaîne de transport des électrons peuvent perturber la phosphorylation oxydative, le cycle de Krebs et les autres processus métaboliques reliés à cette organelle. Dans le cas des plantes

variéguées, il apparaît que les gènes contenus dans les régions du génome plastidique amplifiées sont très fortement surexprimés. Des formes non-matures des ARNs de plusieurs de ces gènes sont également observées, ce qui suggère que la machinerie de maturation des ARNs pourrait être titrée par le surplus de transcrits avec des conséquences néfastes pour la biogénèse des plastides (Maréchal, A., résultats non-publiés). La possibilité que la recombinaison médiée par microhomologie produise de nouveaux cadres de lectures ouverts potentiellement toxiques pour les organelles existe aussi. De telles protéines chimériques sont souvent responsables de la stérilité cytoplasmique mâle chez les plantes à fleurs (revue dans (322)).

Malgré les conséquences néfastes qu'entraîne l'accumulation de réarrangements dans les organelles, ceux-ci semblent importants pour la modification du génome à l'échelle évolutive. En effet, la recombinaison entre des régions répétées très courtes a été observée dans le génome plastidique de différentes sous-espèces de blé (128, 129). Des molécules réarrangées très peu abondantes appelées sublimons ont aussi été proposées comme étant des intermédiaires importants au cours de l'évolution du génome mitochondrial des plantes (135). Ainsi, des génomes réarrangés peuvent devenir prédominants dans les générations subséquentes. Les réarrangements pourraient donc être un agent important au cours de l'évolution des génomes des organelles.

Un mécanisme dépendant de la recombinaison semblable au BIR a été proposé pour expliquer la réplication de l'ADN dans les organelles de plantes (revue dans (44)). Nous avons également montré, dans la présente étude, que les organelles sont capables de réparer des bris double-brin en utilisant une voie semblable au MMBIR. L'évolution des génomes mitochondriaux des plantes à fleurs s'effectue plus lentement que celle des génomes plastidiques qui à leur tour évoluent plus lentement que les génomes nucléaires (83, 84). Une explication plausible de ces observations pourrait être que les mécanismes de réparation de l'ADN des organelles sont plus performants que ceux du noyau. Une utilisation généralisée de mécanismes du type BIR dépendants de la machinerie de recombinaison pour réparer les bris doubles brins ou les arrêts de fourche répliquatives dans

les organelles de plantes pourrait favoriser le maintien d'une homogénéité génétique. En effet, cela permettrait aux génomes mutants de se conformer à la majorité des autres génomes d'une même organelle sur une plus grande partie de leur séquence que lorsqu'ils sont réparés par simple recombinaison homologue.

Le MMBIR a récemment été proposé comme un mécanisme important permettant de générer une variation du nombre de copies de certaines régions des génomes dans une variété d'organismes (revue dans (121)). Le « copy-number variation » (CNV) consiste en la délétion ou la multiplication de différentes régions du génome dans les populations et durant l'évolution. Chez l'homme, jusqu'à 12 % du génome est sujet au CNV et le nombre de copie de ces régions est donc variable entre les individus (323). Au niveau du nombre de nucléotides affectés, le CNV serait même plus important que les polymorphismes nucléotidiques simples (SNPs) pourtant très étudiés. Comme dans le cas des SNPs, il a été démontré que certains CNVs sont à l'origine de maladies alors que d'autres sont plutôt sous sélection positive (122, 324, 325). Le CNV est donc une composante cruciale de la diversité génétique et de l'évolution en général. Certaines des jonctions des réarrangements à l'origine des CNVs sont clairement générées par recombinaison homologue entre des régions répétées assez longues. Par contre, une autre classe de réarrangements comporte des jonctions microhomologues. Ces microhomologies et la complexité de certains réarrangements chromosomiques ont suggéré l'implication d'un mécanisme de type MMBIR dans la formation des CNVs. Dans notre étude, nous démontrons que le MMBIR est aussi présent dans les organelles de plantes où il semble particulièrement important pour la réparation des bris double-brins, suggérant ainsi l'universalité d'un tel mécanisme. Comme le MMBIR semble être favorisé en conditions de stress et qu'il ne dépend pas des recombinases classiques telles Rad51 chez la levure et RecA chez *E. coli*, il pourrait constituer un mécanisme d'urgence ancestral qui interviendrait dans des conditions où la machinerie de réparation conservatrice est surchargée ou moins fonctionnelle, par exemple en absence de Whirlies (326, 327).

Conclusion

Les travaux effectués au cours de mon doctorat m'ont permis de découvrir un nouveau rôle pour les protéines de la famille Whirly dans le maintien de la stabilité du génome des organelles. La très forte affinité de cette famille de protéines pour les acides nucléiques simple-brin leur confère une flexibilité fonctionnelle hors norme. Le rôle de gardiens des génomes des Whirlies est cependant celui qui semble être le mieux conservé à travers les espèces. Il est en effet retrouvé dans les deux branches principales des plantes à fleurs, les eudicotylédons et les monocotylédons. De plus, l'homologue distant des Whirlies, TIF1, maintient la stabilité des deux génomes nucléaires du cilié *T. thermophila*. Il sera intéressant de vérifier le niveau de conservation des autres rôles des Whirlies. La découverte du MMBIR dans les organelles de plantes suggère également qu'il s'agit d'un mécanisme universel important de réparation des dommages à l'ADN basé sur la réplication. Finalement, nos travaux ont permis d'identifier les premiers régulateurs de la recombinaison dans les plastides. L'élucidation des interactions entre les Whirlies et les autres protéines de réplication/réparation/recombinaison des organelles devraient nous permettre de mieux comprendre le maintien de la stabilité du génome des organelles. Ces études devraient également nous renseigner sur les mécanismes moléculaires de production de réarrangements à l'origine des variations du nombre de copies et du développement de certaines maladies génétiques chez l'homme.

Bibliographie

1. Sato S, Nakamura Y, Kaneko T, Asamizu E, & Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res* 6(5):283-290.
2. Unseld M, Marienfeld JR, Brandt P, & Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet* 15(1):57-61.
3. Lohse M, Drechsel O, & Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 52(5-6):267-274.
4. Pyke K (2007) Plastid Biogenesis and Differentiation. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag, Heidelberg), Vol 19, pp 1-28.
5. Nunes-Nesi A, Fernie, AR. (2007) Mitochondrial metabolism. *Plant Mitochondria*, Annual Plant Reviews, ed Logan DC (Blackwell Publishing, Oxford), Vol 31, pp 212-277.
6. Martin W, Kowallik, KV (1999) Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche' *Eur. J. Phycol.*, **34**: 287-295.
7. Wallin IE (1927) *Symbiogenesis and the origin of species* (Williams and Wilkins Company, Baltimore).
8. Hagemann R (2000) Erwin Baur or Carl Correns: who really created the theory of plastid inheritance? *J Hered* 91(6):435-440.
9. Ephrussi B & Hottinguer H (1951) On an unstable cell state in yeast. *Cold Spring Harb Symp Quant Biol* 16:75-85.
10. Dyall SD, Brown MT, & Johnson PJ (2004) Ancient invasions: from endosymbionts to organelles. *Science* 304(5668):253-257.
11. Gould SB, Waller RF, & McFadden GI (2008) Plastid evolution. *Annu Rev Plant Biol* 59:491-517.
12. Keeling PJ & Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9(8):605-618.
13. Jarvis P (2008) Targeting of nucleus-encoded proteins to chloroplasts in plants. *New Phytol* 179(2):257-285.
14. Millar AH, Heazlewood JL, Kristensen BK, Braun HP, & Moller IM (2005) The plant mitochondrial proteome. *Trends Plant Sci* 10(1):36-43.
15. Adams K, Daley, DO. (2004) Plant Mitochondrial Genome Evolution and Gene Transfer to the Nucleus. *Plant Mitochondria: From Genome to Function*, ed Day D, Millar, H., Whelan, J. (Kluwer Academic Publishers, Dordrecht), pp 107-120.

16. Kleine T, Maier UG, & Leister D (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol* 60:115-138.
17. Stupar RM, *et al.* (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc Natl Acad Sci U S A* 98(9):5099-5103.
18. Stegemann S & Bock R (2006) Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell* 18(11):2869-2878.
19. Ricchetti M, Fairhead C, & Dujon B (1999) Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* 402(6757):96-100.
20. Martin W, *et al.* (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A* 99(19):12246-12251.
21. Noutsos C, Kleine T, Armbruster U, DalCorso G, & Leister D (2007) Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. *Trends Genet* 23(12):597-601.
22. Rapaport D (2003) Finding the right organelle. Targeting signals in mitochondrial outer-membrane proteins. *EMBO Rep* 4(10):948-952.
23. Richly E & Leister D (2004) An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. *Gene* 329:11-16.
24. Deusch O, *et al.* (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol* 25(4):748-761.
25. Mitschke J, *et al.* (2009) Prediction of dual protein targeting to plant organelles. *New Phytol* .
26. Mackenzie SA (2005) Plant organellar protein targeting: a traffic plan still under construction. *Trends Cell Biol* 15(10):548-554.
27. Carrie C, Giraud E, & Whelan J (2009) Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. *FEBS J* 276(5):1187-1195.
28. Glaser EaWJ (2007) Import of nuclear-encoded mitochondrial proteins. *Plant Mitochondria*, ed Logan DC (Blackwell Publishing, Oxford), Vol 31, pp 97-140.
29. Agne BaK, F. (2007) Protein import into plastids. *Cell and Molecular Biology of Plastids*, ed Bock R (Springer-Verlag, Heidelberg), Vol 19, pp 339-370.
30. Dolezal P, Likic V, Tachezy J, & Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. *Science* 313(5785):314-318.
31. Jarvis P & Soll J (2002) Toc, tic, and chloroplast protein import. *Biochim Biophys Acta* 1590(1-3):177-189.
32. Qbadou S, *et al.* (2006) The molecular chaperone Hsp90 delivers precursor proteins to the chloroplast import receptor Toc64. *EMBO J* 25(9):1836-1847.

33. O'Brien EA, *et al.* (2009) GOBASE: an organelle genome database. *Nucleic Acids Res* 37(Database issue):D946-950.
34. Anderson S, *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290(5806):457-465.
35. Tomohiko Kubo TM (2007) Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum* 129(1):6-13.
36. Bock R (2007) Structure, function and inheritance of plastid genomes. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag Berlin Heidelberg, Heidelberg), Vol 19, pp 29-63.
37. Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16(7):1661-1666.
38. Backert S & Borner T (2000) Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L.). *Curr Genet* 37(5):304-314.
39. Backert S, Lurz R, & Borner T (1996) Electron microscopic investigation of mitochondrial DNA from *Chenopodium album* (L.). *Curr Genet* 29(5):427-436.
40. Lilly JW, Havey MJ, Jackson SA, & Jiang J (2001) Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. *Plant Cell* 13(2):245-254.
41. Oldenburg DJ & Bendich AJ (1996) Size and Structure of Replicating Mitochondrial DNA in Cultured Tobacco Cells. *Plant Cell* 8(3):447-461.
42. Oldenburg DJ & Bendich AJ (2004) Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *J Mol Biol* 335(4):953-970.
43. Bendich AJ (1993) Reaching for the ring: the study of mitochondrial genome structure. *Curr Genet* 24(4):279-290.
44. Day A, Madesis, P. (2007) DNA replication, recombination, and repair in plastids. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag Berlin Heidelberg, Leipzig), Vol 19, pp 65-119.
45. Kucej M & Butow RA (2007) Evolutionary tinkering with mitochondrial nucleoids. *Trends Cell Biol* 17(12):586-592.
46. Sakai A, Takano H, & Kuroiwa T (2004) Organelle nuclei in higher plants: structure, composition, function, and evolution. *Int Rev Cytol* 238:59-118.
47. Fey J, Vermel M, Grienberger J, Marechal-Drouard L, & Gualberto JM (1999) Characterization of a plant mitochondrial active chromosome. *FEBS Lett* 458(2):124-128.
48. Dai H, *et al.* (2005) Structural and functional characterizations of mung bean mitochondrial nucleoids. *Nucleic Acids Res* 33(15):4725-4739.
49. Kobayashi T, *et al.* (2002) Detection and localization of a chloroplast-encoded HU-like protein that organizes chloroplast nucleoids. *Plant Cell* 14(7):1579-1589.

50. Cannon GC, Ward LN, Case CI, & Heinhorst S (1999) The 68 kDa DNA compacting nucleoid protein from soybean chloroplasts inhibits DNA synthesis in vitro. *Plant Mol Biol* 39(4):835-845.
51. Sato N, Nakayama M, & Hase T (2001) The 70-kDa major DNA-compacting protein of the chloroplast nucleoid is sulfite reductase. *FEBS Lett* 487(3):347-350.
52. Chi-Ham CL, Keaton MA, Cannon GC, & Heinhorst S (2002) The DNA-compacting protein DCP68 from soybean chloroplasts is ferredoxin:sulfite reductase and co-localizes with the organellar nucleoid. *Plant Mol Biol* 49(6):621-631.
53. Sekine K, *et al.* (2007) DNA binding and partial nucleoid localization of the chloroplast stromal enzyme ferredoxin:sulfite reductase. *FEBS J* 274(8):2054-2069.
54. Sekine K, Hase T, & Sato N (2002) Reversible DNA compaction by sulfite reductase regulates transcriptional activity of chloroplast nucleoids. *J Biol Chem* 277(27):24399-24404.
55. Kato Y, *et al.* (2004) The DNA-binding protease, CND41, and the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. *Planta* 220(1):97-104.
56. Kato Y, Yamamoto Y, Murakami S, & Sato F (2005) Post-translational regulation of CND41 protease activity in senescent tobacco leaves. *Planta* 222(4):643-651.
57. Nakano T, *et al.* (1997) A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *Plant Cell* 9(9):1673-1682.
58. Sato N, *et al.* (1998) Molecular characterization of the PEND protein, a novel bZIP protein present in the envelope membrane that is the site of nucleoid replication in developing plastids. *Plant Cell* 10(5):859-872.
59. Sato N & Ohta N (2001) DNA-binding specificity and dimerization of the DNA-binding domain of the PEND protein in the chloroplast envelope membrane. *Nucleic Acids Res* 29(11):2244-2250.
60. Wycliffe P, *et al.* (2005) Continuous expression in tobacco leaves of a Brassica napus PEND homologue blocks differentiation of plastids and development of palisade cells. *Plant J* 44(1):1-15.
61. Jeong SY, Rose A, & Meier I (2003) MFP1 is a thylakoid-associated, nucleoid-binding protein with a coiled-coil structure. *Nucleic Acids Res* 31(17):5175-5185.
62. Jeong SY, Peffer N, & Meier I (2004) Phosphorylation by protein kinase CKII modulates the DNA-binding activity of a chloroplast nucleoid-associated protein. *Planta* 219(2):298-302.
63. Samaniego R, Jeong SY, Meier I, & de la Espina SM (2006) Dual location of MAR-binding, filament-like protein 1 in Arabidopsis, tobacco, and tomato. *Planta* 223(6):1201-1206.
64. Meier I, Phelan T, Gruissem W, Spiker S, & Schneider D (1996) MFP1, a novel plant filament-like protein with affinity for matrix attachment region DNA. *Plant Cell* 8(11):2105-2115.
65. Hanawalt PC (2007) Paradigms for the three rs: DNA replication, recombination, and repair. *Mol Cell* 28(5):702-707.
66. Backert S, Dorfel P, Lurz R, & Borner T (1996) Rolling-circle replication of mitochondrial DNA in the higher plant *Chenopodium album* (L.). *Mol Cell Biol* 16(11):6285-6294.

67. Kolodner RD & Tewari KK (1975) Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism. *Nature* 256(5520):708-711.
68. Kunnimalaiyaan M & Nielsen BL (1997) Fine mapping of replication origins (ori A and ori B) in *Nicotiana tabacum* chloroplast DNA. *Nucleic Acids Res* 25(18):3681-3686.
69. Muhlbauer SK, Lossl A, Tzekova L, Zou Z, & Koop HU (2002) Functional analysis of plastid DNA replication origins in tobacco by targeted inactivation. *Plant J* 32(2):175-184.
70. Scharff LB & Koop HU (2007) Targeted inactivation of the tobacco plastome origins of replication A and B. *Plant J* 50(5):782-794.
71. Luder A & Mosig G (1982) Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. *Proc Natl Acad Sci U S A* 79(4):1101-1105.
72. Kogoma T (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* 61(2):212-238.
73. Kreuzer KN (2005) Interplay between DNA replication and recombination in prokaryotes. *Annu Rev Microbiol* 59:43-67.
74. Wang TS (1991) Eukaryotic DNA polymerases. *Annu Rev Biochem* 60:513-552.
75. Christensen AC, *et al.* (2005) Dual-domain, dual-targeting organellar protein presequences in *Arabidopsis* can use non-AUG start codons. *Plant Cell* 17(10):2805-2816.
76. Ono Y, *et al.* (2007) NtPolI-like1 and NtPolI-like2, bacterial DNA polymerase I homologs isolated from BY-2 cultured tobacco cells, encode DNA polymerases engaged in DNA replication in both plastids and mitochondria. *Plant Cell Physiol* 48(12):1679-1692.
77. Singh BN, Sopory SK, & Reddy MK (2004) Plant DNA Topoisomerases: Structure, Function, and Cellular Roles in Plant Development. *Critical Reviews in Plant Sciences* 23(3):251 - 269.
78. Cho HS, *et al.* (2004) DNA gyrase is involved in chloroplast nucleoid partitioning. *Plant Cell* 16(10):2665-2682.
79. Wall MK, Mitchenall LA, & Maxwell A (2004) *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci U S A* 101(20):7821-7826.
80. Woelfle MA, Thompson RJ, & Mosig G (1993) Roles of novobiocin-sensitive topoisomerases in chloroplast DNA replication in *Chlamydomonas reinhardtii*. *Nucleic Acids Res* 21(18):4231-4238.
81. Wiesner J, Reichenberg A, Heinrich S, Schlitzer M, & Jomaa H (2008) The plastid-like organelle of apicomplexan parasites as drug target. *Curr Pharm Des* 14(9):855-871.
82. Muller HJ (1964) The Relation of Recombination to Mutational Advance. *Mutat Res* 106:2-9.
83. Wolfe KH, Li WH, & Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A* 84(24):9054-9058.

84. Drouin G, Daoud H, & Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol* 49(3):827-831.
85. Tuteja N, Ahmad P, Panda BB, & Tuteja R (2009) Genotoxic stress in plants: shedding light on DNA damage, repair and DNA repair helicases. *Mutat Res* 681(2-3):134-149.
86. Kimura S & Sakaguchi K (2006) DNA repair in plants. *Chem Rev* 106(2):753-766.
87. A. E. STAPLETON, C. S. THORNER, & V. WALBOT (1997) UV-B component of sunlight causes measurable damage in field-grown maize (*Zea mays* L.): developmental and cellular heterogeneity of damage and repair. *Plant, Cell and Environment* 20(3):279-290.
88. Cannon GC, Hedrick LA, & Heinhorst S (1995) Repair mechanisms of UV-induced DNA damage in soybean chloroplasts. *Plant Mol Biol* 29(6):1267-1277.
89. Draper CK & Hays JB (2000) Replication of chloroplast, mitochondrial and nuclear DNA during growth of unirradiated and UVB-irradiated Arabidopsis leaves. *Plant J* 23(2):255-265.
90. Chen JJ, Jiang CZ, & Britt AB (1996) Little or No Repair of Cyclobutyl Pyrimidine Dimers Is Observed in the Organellar Genomes of the Young Arabidopsis Seedling. *Plant Physiol* 111(1):19-25.
91. Kunz BA, Anderson HJ, Osmond MJ, & Vonarx EJ (2005) Components of nucleotide excision repair and DNA damage tolerance in Arabidopsis thaliana. *Environ Mol Mutagen* 45(2-3):115-127.
92. Holt IJ (2009) Mitochondrial DNA replication and repair: all a flap. *Trends Biochem Sci* .
93. de Souza-Pinto NC, Wilson DM, 3rd, Stevensner TV, & Bohr VA (2008) Mitochondrial DNA, base excision repair and neurodegeneration. *DNA Repair (Amst)* 7(7):1098-1109.
94. Bensen RJ & Warner HR (1987) Partial Purification and Characterization of Uracil-DNA Glycosylase Activity from Chloroplasts of Zea mays Seedlings. *Plant Physiol* 84(4):1102-1106.
95. Gutman BL & Niyogi KK (2009) Evidence for Base Excision Repair of Oxidative DNA Damage in Chloroplasts of Arabidopsis thaliana. *J Biol Chem* 284(25):17006-17012.
96. Hsieh P & Yamane K (2008) DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mech Ageing Dev* 129(7-8):391-407.
97. Leonard JM, Bollmann SR, & Hays JB (2003) Reduction of stability of arabidopsis genomic and transgenic DNA-repeat sequences (microsatellites) by inactivation of AtMSH2 mismatch-repair function. *Plant Physiol* 133(1):328-338.
98. Hoffman PD, Leonard JM, Lindberg GE, Bollmann SR, & Hays JB (2004) Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective Arabidopsis. *Genes Dev* 18(21):2676-2685.
99. Dion E, Li L, Jean M, & Belzile F (2007) An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. *Plant J* 51(3):431-440.
100. Chen S, Bigner SH, & Modrich P (2001) High rate of CAD gene amplification in human cells deficient in MLH1 or MSH6. *Proc Natl Acad Sci U S A* 98(24):13802-13807.

101. Ross-Macdonald P & Roeder GS (1994) Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell* 79(6):1069-1080.
102. Lu X, *et al.* (2008) The Arabidopsis MutS homolog AtMSH5 is required for normal meiosis. *Cell Res* 18(5):589-599.
103. Martinez-Zapater JM, Gil P, Capel J, & Somerville CR (1992) Mutations at the Arabidopsis CHM locus promote rearrangements of the mitochondrial genome. *Plant Cell* 4(8):889-899.
104. Sakamoto W, Kondo H, Murata M, & Motoyoshi F (1996) Altered mitochondrial gene expression in a maternal distorted leaf mutant of Arabidopsis induced by chloroplast mutator. *Plant Cell* 8(8):1377-1390.
105. Abdelnoor RV, *et al.* (2003) Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc Natl Acad Sci U S A* 100(10):5968-5973.
106. Shedge V, Arrieta-Montiel M, Christensen AC, & Mackenzie SA (2007) Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell* 19(4):1251-1264.
107. Reenan RA & Kolodner RD (1992) Characterization of insertion mutations in the *Saccharomyces cerevisiae* MSH1 and MSH2 genes: evidence for separate mitochondrial and nuclear functions. *Genetics* 132(4):975-985.
108. Wyman C & Kanaar R (2006) DNA double-strand break repair: all's well that ends well. *Annu Rev Genet* 40:363-383.
109. Michel B, Grompone G, Flores MJ, & Bidnenko V (2004) Multiple pathways process stalled replication forks. *Proc Natl Acad Sci U S A* 101(35):12783-12788.
110. Pâques F & Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63(2):349-404.
111. Szostak JW, Orr-Weaver TL, Rothstein RJ, & Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* 33(1):25-35.
112. McVey M & Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24(11):529-538.
113. Llorente B, Smith CE, & Symington LS (2008) Break-induced replication: what is it and what is it for? *Cell Cycle* 7(7):859-864.
114. Lydeard JR, Jain S, Yamaguchi M, & Haber JE (2007) Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* 448(7155):820-823.
115. McEachern MJ & Haber JE (2006) Break-induced replication and recombinational telomere elongation in yeast. *Annu Rev Biochem* 75:111-135.
116. Persky NS & Lovett ST (2008) Mechanisms of recombination: lessons from *E. coli*. *Crit Rev Biochem Mol Biol* 43(6):347-370.
117. Pennington JM & Rosenberg SM (2007) Spontaneous DNA breakage in single living *Escherichia coli* cells. *Nat Genet* 39(6):797-802.

118. Cox MM, *et al.* (2000) The importance of repairing stalled replication forks. *Nature* 404(6773):37-41.
119. Spence JE, *et al.* (1988) Uniparental disomy as a mechanism for human genetic disease. *Am J Hum Genet* 42(2):217-226.
120. Mahaney BL, Meek K, & Lees-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 417(3):639-650.
121. Hastings PJ, Ira G, & Lupski JR (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 5(1):e1000327.
122. Lee JA, Carvalho CM, & Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131(7):1235-1247.
123. Palmer JD (1983) Chloroplast DNA exists in 2 orientations. *Nature* 301:92-93.
124. Boynton JE, *et al.* (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* 240(4858):1534-1538.
125. Svab Z, Hajdukiewicz P, & Maliga P (1990) Stable transformation of plastids in higher plants. *Proc Natl Acad Sci U S A* 87(21):8526-8530.
126. Iamtham S & Day A (2000) Removal of antibiotic resistance genes from transgenic tobacco plastids. *Nat Biotechnol* 18(11):1172-1176.
127. Dauvillee D, Hilbig L, Preiss S, & Johanningmeier U (2004) Minimal Extent of Sequence Homology Required for Homologous Recombination at the *psbA* Locus in *Chlamydomonas reinhardtii* Chloroplasts using PCR-generated DNA Fragments. *Photosynth Res* 79(2):219-224.
128. Ogihara Y, Terachi T, & Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc Natl Acad Sci U S A* 85(22):8573-8577.
129. Ogihara Y, Terachi T, & Sasakuma T (1992) Structural analysis of length mutations in a hot-spot region of wheat chloroplast DNAs. *Curr Genet* 22(3):251-258.
130. Durrenberger F, Thompson AJ, Herrin DL, & Rochaix JD (1996) Double strand break-induced recombination in *Chlamydomonas reinhardtii* chloroplasts. *Nucleic Acids Res* 24(17):3323-3331.
131. Odom OW, Baek KH, Dani RN, & Herrin DL (2008) *Chlamydomonas* chloroplasts can use short dispersed repeats and multiple pathways to repair a double-strand break in the genome. *Plant J* 53(5):842-853.
132. Khakhlova O & Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J* 46(1):85-94.
133. Kohl S & Bock R (2009) Transposition of a bacterial insertion sequence in chloroplasts. *Plant J* 58(3):423-436.
134. Fauron C, Casper M, Gao Y, & Moore B (1995) The maize mitochondrial genome: dynamic, yet functional. *Trends Genet* 11(6):228-235.

135. Small I, Suffolk R, & Leaver CJ (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell* 58(1):69-76.
136. Small ID, Isaac PG, & Leaver CJ (1987) Stoichiometric differences in DNA molecules containing the *atpA* gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. *EMBO J* 6(4):865-869.
137. Newton KJ & Coe EH (1986) Mitochondrial DNA changes in abnormal growth (nonchromosomal stripe) mutants of maize. *Proc Natl Acad Sci U S A* 83(19):7363-7366.
138. Mackenzie SA (2005) The mitochondrial genome of higher plants: a target for natural adaptation. *Plant diversity and evolution genotypic and phenotypic variation in higher plants*, ed Henry RJ (CABI Publishing, Wallingford), pp 69-79.
139. Shereda RD, Kozlov AG, Lohman TM, Cox MM, & Keck JL (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol* 43(5):289-318.
140. Richard DJ, Bolderson E, & Khanna KK (2009) Multiple human single-stranded DNA binding proteins function in genome maintenance: structural, biochemical and functional analysis. *Critical Reviews in Biochemistry and Molecular Biology* 99999(1):1 - 19.
141. Edmondson AC, *et al.* (2005) Characterization of a mitochondrially targeted single-stranded DNA-binding protein in *Arabidopsis thaliana*. *Mol Genet Genomics* 273(2):115-122.
142. Zaegel V, *et al.* (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell* 18(12):3548-3563.
143. Sakaguchi K, Ishibashi T, Uchiyama Y, & Iwabata K (2009) The multi-replication protein A (RPA) system--a new perspective. *FEBS J* 276(4):943-963.
144. Zou Y, Liu Y, Wu X, & Shell SM (2006) Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. *J Cell Physiol* 208(2):267-273.
145. Ishibashi T, Kimura S, & Sakaguchi K (2006) A higher plant has three different types of RPA heterotrimeric complex. *J Biochem* 139(1):99-104.
146. Ishibashi T, *et al.* (2005) Two types of replication protein A in seed plants. *FEBS J* 272(13):3270-3281.
147. Shinohara A, Ogawa H, & Ogawa T (1992) Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* 69(3):457-470.
148. Bishop DK, Park D, Xu L, & Kleckner N (1992) DMC1: a meiosis-specific yeast homolog of *E. coli* *recA* required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* 69(3):439-456.
149. Sancar A & Rupp WD (1979) Physical map of the *recA* gene. *Proc Natl Acad Sci U S A* 76(7):3144-3148.
150. Cox MM (2007) Regulation of bacterial RecA protein function. *Crit Rev Biochem Mol Biol* 42(1):41-63.

151. Cerutti H, Osman M, Grandoni P, & Jagendorf AT (1992) A homolog of Escherichia coli RecA protein in plastids of higher plants. *Proc Natl Acad Sci U S A* 89(17):8068-8072.
152. Manchekar M, *et al.* (2006) DNA recombination activity in soybean mitochondria. *J Mol Biol* 356(2):288-299.
153. Cerutti H & Jagendorf AT (1993) DNA Strand-Transfer Activity in Pea (*Pisum sativum* L.) Chloroplasts. *Plant Physiol* 102(1):145-153.
154. Khazi FR, Edmondson AC, & Nielsen BL (2003) An Arabidopsis homologue of bacterial RecA that complements an E. coli recA deletion is targeted to plant mitochondria. *Mol Genet Genomics* 269(4):454-463.
155. Odahara M, Inouye T, Fujita T, Hasebe M, & Sekine Y (2007) Involvement of mitochondrial-targeted RecA in the repair of mitochondrial DNA in the moss, *Physcomitrella patens*. *Genes Genet Syst* 82(1):43-51.
156. Rédei GP (1973) Extra-chromosomal mutability determined by a nuclear gene locus in arabidopsis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 18(2):149-162.
157. Sandhu AP, Abdelnoor RV, & Mackenzie SA (2007) Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. *Proc Natl Acad Sci U S A* 104(6):1766-1770.
158. Abdelnoor RV, *et al.* (2006) Mitochondrial genome dynamics in plants and animals: convergent gene fusions of a MutS homologue. *J Mol Evol* 63(2):165-173.
159. Vermel M, *et al.* (2002) A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proc Natl Acad Sci U S A* 99(9):5866-5871.
160. Isono K, *et al.* (1997) Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of sigma70 factors of bacterial RNA polymerases in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 94(26):14948-14953.
161. Fujiwara M, Nagashima A, Kanamaru K, Tanaka K, & Takahashi H (2000) Three new nuclear genes, sigD, sigE and sigF, encoding putative plastid RNA polymerase sigma factors in Arabidopsis thaliana. *FEBS Lett* 481(1):47-52.
162. Masters BS, Stohl LL, & Clayton DA (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* 51(1):89-99.
163. Hedtke B, Borner T, & Weihe A (1997) Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science* 277(5327):809-811.
164. Gagliardi D, Binder, S (2007) Expression of the plant mitochondrial genome. *Plant Mitochondria*, Annual Plant Reviews, ed Logan DC (Blackwell Publishing, Oxford), Vol 31, pp 50-96.
165. Hedtke B, *et al.* (1999) Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo. *Plant J* 17(5):557-561.
166. Hedtke B, Borner T, & Weihe A (2000) One RNA polymerase serving two genomes. *EMBO Rep* 1(5):435-440.

167. Kuhn K, Bohne AV, Liere K, Weihe A, & Börner T (2007) Arabidopsis phage-type RNA polymerases: accurate in vitro transcription of organellar genes. *Plant Cell* 19(3):959-971.
168. Liere K, Börner, T (2007) Transcription and transcriptional regulation in plastids. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag, Berlin), Vol 19, pp 121-174.
169. Hajdukiewicz PT, Allison LA, & Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J* 16(13):4041-4048.
170. Swiatecka-Hagenbruch M, Liere K, & Börner T (2007) High diversity of plastidial promoters in Arabidopsis thaliana. *Mol Genet Genomics* 277(6):725-734.
171. Weihe A & Börner T (1999) Transcription and the architecture of promoters in chloroplasts. *Trends Plant Sci* 4(5):169-170.
172. Haley J & Bogorad L (1990) Alternative promoters are used for genes within maize chloroplast polycistronic transcription units. *Plant Cell* 2(4):323-333.
173. Mullet JE (1993) Dynamic regulation of chloroplast transcription. *Plant Physiol* 103(2):309-313.
174. Kuhn K, Weihe A, & Börner T (2005) Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. *Nucleic Acids Res* 33(1):337-346.
175. Liere K, Kaden D, Maliga P, & Börner T (2004) Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type. *Nucleic Acids Res* 32(3):1159-1165.
176. Zoschke R, Liere K, & Börner T (2007) From seedling to mature plant: arabidopsis plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. *Plant J* 50(4):710-722.
177. Giege P, Hoffmann M, Binder S, & Brennicke A (2000) RNA degradation buffers asymmetries of transcription in Arabidopsis mitochondria. *EMBO Rep* 1(2):164-170.
178. Leino M, Landgren M, & Glimelius K (2005) Alloplasmic effects on mitochondrial transcriptional activity and RNA turnover result in accumulated transcripts of Arabidopsis orfs in cytoplasmic male-sterile Brassica napus. *Plant J* 42(4):469-480.
179. Bonen L (2008) Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion* 8(1):26-34.
180. Schmitz-Linneweber C, Barkan, A (2007) RNA splicing and RNA editing in chloroplasts. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag, Heidelberg), Vol 19, pp 213-248.
181. Mohr G & Lambowitz AM (2003) Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res* 31(2):647-652.
182. Ostheimer GJ, *et al.* (2003) Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. *EMBO J* 22(15):3919-3929.

183. Jenkins BD & Barkan A (2001) Recruitment of a peptidyl-tRNA hydrolase as a facilitator of group II intron splicing in chloroplasts. *EMBO J* 20(4):872-879.
184. Till B, Schmitz-Linneweber C, Williams-Carrier R, & Barkan A (2001) CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA* 7(9):1227-1238.
185. Schmitz-Linneweber C, *et al.* (2006) A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell* 18(10):2650-2663.
186. Watkins KP, *et al.* (2007) A ribonuclease III domain protein functions in group II intron splicing in maize chloroplasts. *Plant Cell* 19(8):2606-2623.
187. Asakura Y & Barkan A (2007) A CRM domain protein functions dually in group I and group II intron splicing in land plant chloroplasts. *Plant Cell* 19(12):3864-3875.
188. Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, & Barkan A (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol Cell Biol* 28(17):5337-5347.
189. Asakura Y, Bayraktar OA, & Barkan A (2008) Two CRM protein subfamilies cooperate in the splicing of group IIB introns in chloroplasts. *RNA* 14(11):2319-2332.
190. Kroeger TS, Watkins KP, Friso G, van Wijk KJ, & Barkan A (2009) A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. *Proc Natl Acad Sci U S A* 106(11):4537-4542.
191. Lurin C, *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16(8):2089-2103.
192. Schmitz-Linneweber C & Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* 13(12):663-670.
193. Manthey GM & McEwen JE (1995) The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of *Saccharomyces cerevisiae*. *EMBO J* 14(16):4031-4043.
194. Pfalz J, Liere K, Kandlbinder A, Dietz KJ, & Oelmuller R (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18(1):176-197.
195. Barkan A, Walker M, Nolasco M, & Johnson D (1994) A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J* 13(13):3170-3181.
196. Yamazaki H, Tasaka M, & Shikanai T (2004) PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in Arabidopsis. *Plant J* 38(1):152-163.
197. Meierhoff K, Felder S, Nakamura T, Bechtold N, & Schuster G (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* 15(6):1480-1495.

198. Okuda K, Myouga F, Motohashi R, Shinozaki K, & Shikanai T (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* 104(19):8178-8183.
199. Brown GG, *et al.* (2003) The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J* 35(2):262-272.
200. Gillman JD, Bentolila S, & Hanson MR (2007) The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. *Plant J* 49(2):217-227.
201. Wang Z, *et al.* (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* 18(3):676-687.
202. Kazama T, Nakamura T, Watanabe M, Sugita M, & Toriyama K (2008) Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. *Plant J* 55(4):619-628.
203. Uyttewaal M, *et al.* (2008) Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell* 20(12):3331-3345.
204. Benne R, *et al.* (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46(6):819-826.
205. Covello PS & Gray MW (1989) RNA editing in plant mitochondria. *Nature* 341(6243):662-666.
206. Gualberto JM, Lamattina L, Bonnard G, Weil JH, & Grienenberger JM (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature* 341(6243):660-662.
207. Hiesel R, Wissinger B, Schuster W, & Brennicke A (1989) RNA editing in plant mitochondria. *Science* 246(4937):1632-1634.
208. Hoch B, Maier RM, Appel K, Igloi GL, & Kossel H (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353(6340):178-180.
209. Giege P & Brennicke A (1999) RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A* 96(26):15324-15329.
210. Maier RM, Neckermann K, Igloi GL, & Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* 251(5):614-628.
211. Kotera E, Tasaka M, & Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433(7023):326-330.
212. Okuda K, Nakamura T, Sugita M, Shimizu T, & Shikanai T (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281(49):37661-37667.
213. Okuda K, *et al.* (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell* 21(1):146-156.

214. Yu QB, Jiang Y, Chong K, & Yang ZN (2009) AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript accD RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. *Plant J*.
215. Kim SR, *et al.* (2009) Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. *Plant J*.
216. Hirose T & Sugiura M (2001) Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system. *EMBO J* 20(5):1144-1152.
217. Tillich M, *et al.* (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A* 106(14):6002-6007.
218. Salone V, *et al.* (2007) A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* 581(22):4132-4138.
219. Nakamura T & Sugita M (2008) A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett* 582(30):4163-4168.
220. Bollenbach T, Schuster, G, Portnoy, V, Stern, DB (2007) Processing, degradation, and polyadenylation of chloroplast transcripts. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag, Heidelberg), pp 175-211.
221. Perrin R, *et al.* (2004) Two exoribonucleases act sequentially to process mature 3'-ends of atp9 mRNAs in *Arabidopsis* mitochondria. *J Biol Chem* 279(24):25440-25446.
222. Bollenbach TJ, *et al.* (2005) RNR1, a 3'-5' exoribonuclease belonging to the RNR superfamily, catalyzes 3' maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*. *Nucleic Acids Res* 33(8):2751-2763.
223. Yukawa M, Kuroda H, & Sugiura M (2007) A new in vitro translation system for non-radioactive assay from tobacco chloroplasts: effect of pre-mRNA processing on translation in vitro. *Plant J* 49(2):367-376.
224. Fisk DG, Walker MB, & Barkan A (1999) Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J* 18(9):2621-2630.
225. Hattori M, Miyake H, & Sugita M (2007) A Pentatricopeptide repeat protein is required for RNA processing of clpP Pre-mRNA in moss chloroplasts. *J Biol Chem* 282(14):10773-10782.
226. Pfalz J, Bayraktar OA, Prikryl J, & Barkan A (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J*.
227. Kaleikau EK, Andre CP, & Walbot V (1992) Structure and expression of the rice mitochondrial apocytochrome b gene (cob-1) and pseudogene (cob-2). *Curr Genet* 22(6):463-470.
228. Bellaoui M, Pelletier G, & Budar F (1997) The steady-state level of mRNA from the Ogura cytoplasmic male sterility locus in *Brassica* cybrids is determined post-transcriptionally by its 3' region. *EMBO J* 16(16):5057-5068.

229. Kuhn J, Tengler U, & Binder S (2001) Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria. *Mol Cell Biol* 21(3):731-742.
230. Lisitsky I, Klaff P, & Schuster G (1996) Addition of destabilizing poly (A)-rich sequences to endonuclease cleavage sites during the degradation of chloroplast mRNA. *Proc Natl Acad Sci U S A* 93(23):13398-13403.
231. Kudla J, Hayes R, & Gruissem W (1996) Polyadenylation accelerates degradation of chloroplast mRNA. *EMBO J* 15(24):7137-7146.
232. Gagliardi D & Leaver CJ (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. *EMBO J* 18(13):3757-3766.
233. Zimmer SL, Schein A, Zipor G, Stern DB, & Schuster G (2009) Polyadenylation in Arabidopsis and Chlamydomonas organelles: the input of nucleotidyltransferases, poly(A) polymerases and polynucleotide phosphorylase. *Plant J* 59(1):88-99.
234. Walter M, Kilian J, & Kudla J (2002) PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. *EMBO J* 21(24):6905-6914.
235. Holec S, *et al.* (2006) Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol Cell Biol* 26(7):2869-2876.
236. Peled-Zehavi H, Danon, A (2007) Translation and translational regulation in chloroplasts. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag, Heidelberg), pp 249-281.
237. Binder S & Brennicke A (2003) Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos Trans R Soc Lond B Biol Sci* 358(1429):181-188; discussion 188-189.
238. Mayfield SP, Cohen A, Danon A, & Yohn CB (1994) Translation of the psbA mRNA of Chlamydomonas reinhardtii requires a structured RNA element contained within the 5' untranslated region. *J Cell Biol* 127(6 Pt 1):1537-1545.
239. Fargo DC, Zhang M, Gillham NW, & Boynton JE (1998) Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in Chlamydomonas reinhardtii chloroplasts or in Escherichia coli. *Mol Gen Genet* 257(3):271-282.
240. Pring DR, Mullen JA, & Kempken F (1992) Conserved sequence blocks 5' to start codons of plant mitochondrial genes. *Plant Mol Biol* 19(2):313-317.
241. Zerges W (2000) Translation in chloroplasts. *Biochimie* 82(6-7):583-601.
242. Giege P, Sweetlove LJ, Cognat V, & Leaver CJ (2005) Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in Arabidopsis. *Plant Cell* 17(5):1497-1512.
243. Marineau C, Matton DP, & Brisson N (1987) Differential accumulation of potato tuber mRNAs during the hypersensitive response induced by arachidonic acid elicitor. *Plant Molecular Biology* 9(4):335-342.

244. Matton DP & Brisson N (1989) Cloning, expression, and sequence conservation of pathogenesis-related gene transcripts of potato. *Mol Plant Microbe Interact* 2(6):325-331.
245. Matton DP, Bell B, & Brisson N (1990) Nucleotide sequence of a pathogenesis-related gene of potato. *Plant Mol Biol* 14(5):863-865.
246. Constabel CP & Brisson N (1992) The defense-related STH-2 gene product of potato shows race-specific accumulation after inoculation with low concentrations of *Phytophthora infestans* zoospores. *Planta* 188(3):289-295.
247. Matton DP, Prescott G, Bertrand C, Camirand A, & Brisson N (1993) Identification of cis-acting elements involved in the regulation of the pathogenesis-related gene STH-2 in potato. *Plant Mol Biol* 22(2):279-291.
248. Després C, Subramaniam R, Matton DP, & Brisson N (1995) The Activation of the Potato PR-10a Gene Requires the Phosphorylation of the Nuclear Factor PBF-1. *Plant Cell* 7(5):589-598.
249. Desveaux D, Després C, Joyeux A, Subramaniam R, & Brisson N (2000) PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. *Plant Cell* 12(8):1477-1489.
250. Gonzalez-Lamothe R, *et al.* (2008) The transcriptional activator Pti4 is required for the recruitment of a repressosome nucleated by repressor SEBF at the potato PR-10a gene. *Plant Cell* 20(11):3136-3147.
251. Desveaux D, *et al.* (2004) A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in Arabidopsis. *Dev Cell* 6(2):229-240.
252. Desveaux D, Maréchal A, & Brisson N (2005) Whirly transcription factors: defense gene regulation and beyond. *Trends Plant Sci* 10(2):95-102.
253. Krause K, *et al.* (2005) DNA-binding proteins of the Whirly family in Arabidopsis thaliana are targeted to the organelles. *FEBS Lett* 579(17):3707-3712.
254. Grabowski E, Miao Y, Mulisch M, & Krupinska K (2008) Single-stranded DNA-binding protein Whirly1 in barley leaves is located in plastids and the nucleus of the same cell. *Plant Physiol* 147(4):1800-1804.
255. Desveaux D, Allard J, Brisson N, & Sygusch J (2002) A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nat Struct Biol* 9(7):512-517.
256. Subramaniam R, Despres C, & Brisson N (1997) A functional homolog of mammalian protein kinase C participates in the elicitor-induced defense response in potato. *Plant Cell* 9(4):653-664.
257. Constabel CP, Bertrand C, & Brisson N (1993) Transgenic potato plants overexpressing the pathogenesis-related STH-2 gene show unaltered susceptibility to *Phytophthora infestans* and potato virus X. *Plant Mol Biol* 22(5):775-782.
258. Durrant WE & Dong X (2004) Systemic acquired resistance. *Annu Rev Phytopathol* 42:185-209.
259. Xiong JY, *et al.* (2009) Recruitment of AtWHY1 and AtWHY3 by a distal element upstream of the kinesin gene AtKP1 to mediate transcriptional repression. *Plant Mol Biol*.

260. Yoo HH, Kwon C, Lee MM, & Chung IK (2007) Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in Arabidopsis. *Plant J* 49(3):442-451.
261. Kwon C & Chung IK (2004) Interaction of an Arabidopsis RNA-binding protein with plant single-stranded telomeric DNA modulates telomerase activity. *J Biol Chem* 279(13):12812-12818.
262. Prikryl J, Watkins KP, Friso G, van Wijk KJ, & Barkan A (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res* 36(16):5152-5165.
263. Jenkins BD, Kulhanek DJ, & Barkan A (1997) Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* 9(3):283-296.
264. Eisen JA, *et al.* (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 4(9):e286.
265. Saha S & Kapler GM (2000) Allele-specific protein-DNA interactions between the single-stranded DNA-binding protein, ssA-TIBF, and DNA replication determinants in *Tetrahymena*. *J Mol Biol* 295(3):423-439.
266. Saha S, Nicholson A, & Kapler GM (2001) Cloning and biochemical analysis of the tetrahymena origin binding protein TIF1: competitive DNA binding in vitro and in vivo to critical rDNA replication determinants. *J Biol Chem* 276(48):45417-45426.
267. Morrison TL, Yakisich JS, Cassidy-Hanley D, & Kapler GM (2005) TIF1 Represses rDNA replication initiation, but promotes normal S phase progression and chromosome transmission in *Tetrahymena*. *Mol Biol Cell* 16(6):2624-2635.
268. Karplus K (2009) SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Res* 37(Web Server issue):W492-497.
269. Notredame C, Higgins DG, & Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302(1):205-217.
270. Gouet P, Robert X, & Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31(13):3320-3323.
271. Yakisich JS, Sandoval PY, Morrison TL, & Kapler GM (2006) TIF1 activates the intra-S-phase checkpoint response in the diploid micronucleus and amitotic polyploid macronucleus of *Tetrahymena*. *Mol Biol Cell* 17(12):5185-5197.
272. Simpson AGB, Stevens JR, & Lukes J (2006) The evolution and diversity of kinetoplastid flagellates. *Trends in Parasitology* 22(4):168-174.
273. Feagin JE, Abraham JM, & Stuart K (1988) Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* 53(3):413-422.
274. Lukes J, Hashimi H, & Zikova A (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr Genet* 48(5):277-299.
275. Koller J, *et al.* (1997) *Trypanosoma brucei* gBP21. An arginine-rich mitochondrial protein that binds to guide RNA with high affinity. *J Biol Chem* 272(6):3749-3757.

276. Muller UF, Lambert L, & Goring HU (2001) Annealing of RNA editing substrates facilitated by guide RNA-binding protein gBP21. *EMBO J* 20(6):1394-1404.
277. Vondruskova E, *et al.* (2005) RNA interference analyses suggest a transcript-specific regulatory role for mitochondrial RNA-binding proteins MRP1 and MRP2 in RNA editing and other RNA processing in *Trypanosoma brucei*. *J Biol Chem* 280(4):2429-2438.
278. Schumacher MA, Karamooz E, Zikova A, Trantirek L, & Lukes J (2006) Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell* 126(4):701-711.
279. Zikova A, *et al.* (2008) Structure and function of the native and recombinant mitochondrial MRP1/MRP2 complex from *Trypanosoma brucei*. *Int J Parasitol* 38(8-9):901-912.
280. Wilson IA (2009) Structural and functional characterizations of SSGB, a conserved activator of developmental cell division in morphologically complex actinomycetes. *J Biol Chem* .
281. Unseld M, Marienfeld JR, Brandt P, & Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15(1):57-61.
282. Davis IW, *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375-383.
283. Lovett ST, Hurley RL, Sutera VA, Jr., Aubuchon RH, & Lebedeva MA (2002) Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics* 160(3):851-859.
284. Atkinson J & McGlynn P (2009) Replication fork reversal and the maintenance of genome stability. *Nucleic Acids Res* .
285. Bzymek M & Lovett ST (2001) Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc Natl Acad Sci U S A* 98(15):8319-8325.
286. Dutra BE, Sutera VA, Jr., & Lovett ST (2007) RecA-independent recombination is efficient but limited by exonucleases. *Proc Natl Acad Sci U S A* 104(1):216-221.
287. Myouga F, *et al.* (2008) A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell* 20(11):3148-3162.
288. Lisby M, Barlow JH, Burgess RC, & Rothstein R (2004) Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* 118(6):699-713.
289. Srivastava S & Moraes CT (2005) Double-strand breaks of mouse muscle mtDNA promote large deletions similar to multiple mtDNA deletions in humans. *Hum Mol Genet* 14(7):893-902.
290. Shereda RD, Reiter NJ, Butcher SE, & Keck JL (2009) Identification of the SSB binding site on *E. coli* RecQ reveals a conserved surface for binding SSB's C terminus. *J Mol Biol* 386(3):612-625.
291. Lu D & Keck JL (2008) Structural basis of *Escherichia coli* single-stranded DNA-binding protein stimulation of exonuclease I. *Proc Natl Acad Sci U S A* 105(27):9169-9174.

292. Aguilera A & Gomez-Gonzalez B (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet* 9(3):204-217.
293. Arlt MF, *et al.* (2009) Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet* 84(3):339-350.
294. Aguilera A (2002) The connection between transcription and genomic instability. *EMBO J* 21(3):195-201.
295. Maréchal A, *et al.* (2008) Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol* 8:42.
296. Inouye T, Odahara M, Fujita T, Hasebe M, & Sekine Y (2008) Expression and complementation analyses of a chloroplast-localized homolog of bacterial RecA in the moss *Physcomitrella patens*. *Biosci Biotechnol Biochem* 72(5):1340-1347.
297. Nakazato E, Fukuzawa H, Tabata S, Takahashi H, & Tanaka K (2003) Identification and expression analysis of cDNA encoding a chloroplast recombination protein REC1, the chloroplast RecA homologue in *Chlamydomonas reinhardtii*. *Biosci Biotechnol Biochem* 67(12):2608-2613.
298. Cao J, Combs C, & Jagendorf AT (1997) The chloroplast-located homolog of bacterial DNA recombinase. *Plant Cell Physiol* 38(12):1319-1325.
299. Griffith JD, *et al.* (1999) Mammalian telomeres end in a large duplex loop. *Cell* 97(4):503-514.
300. Schoeffler AJ & Berger JM (2008) DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Q Rev Biophys* 41(1):41-101.
301. Schmitz-Linneweber C, Williams-Carrier R, & Barkan A (2005) RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. *Plant Cell* 17(10):2791-2804.
302. Paulsen RD, *et al.* (2009) A Genome-wide siRNA Screen Reveals Diverse Cellular Processes and Pathways that Mediate Genome Stability. *Mol Cell* 35(2):228-239.
303. Li X & Manley JL (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122(3):365-378.
304. Huertas P & Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 12(3):711-721.
305. Li X & Manley JL (2006) Cotranscriptional processes and their influence on genome stability. *Genes Dev* 20(14):1838-1847.
306. Kamashev D, Balandina A, Mazur AK, Arimondo PB, & Rouviere-Yaniv J (2008) HU binds and folds single-stranded DNA. *Nucleic Acids Res* 36(3):1026-1036.
307. Kamashev D & Rouviere-Yaniv J (2000) The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J* 19(23):6527-6535.
308. Balandina A, Kamashev D, & Rouviere-Yaniv J (2002) The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J Biol Chem* 277(31):27622-27628.

309. Balandina A, Claret L, Hengge-Aronis R, & Rouviere-Yaniv J (2001) The Escherichia coli histone-like protein HU regulates rpoS translation. *Mol Microbiol* 39(4):1069-1079.
310. Malik M, Bensaïd A, Rouviere-Yaniv J, & Drlica K (1996) Histone-like protein HU and bacterial DNA topology: suppression of an HU deficiency by gyrase mutations. *J Mol Biol* 256(1):66-76.
311. Boubrik F & Rouviere-Yaniv J (1995) Increased sensitivity to gamma irradiation in bacteria lacking protein HU. *Proc Natl Acad Sci U S A* 92(9):3958-3962.
312. Li S & Waters R (1998) Escherichia coli strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination. *J Bacteriol* 180(15):3750-3756.
313. Stoilov LM, *et al.* (1988) Supercoils in plant DNA: nucleoid sedimentation studies. *J Cell Sci* 89 (Pt 2):243-252.
314. Liu Z & Butow RA (2006) Mitochondrial retrograde signaling. *Annu Rev Genet* 40:159-185.
315. Pesaresi P, Schneider A, Kleine T, & Leister D (2007) Interorganellar communication. *Curr Opin Plant Biol* 10(6):600-606.
316. Woodson JD & Chory J (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* 9(5):383-395.
317. Krause K & Krupinska K (2009) Nuclear regulators with a second home in organelles. *Trends Plant Sci* 14(4):194-199.
318. Marechal A, *et al.* (2009) Whirly proteins maintain plastid genome stability in Arabidopsis. *Proc Natl Acad Sci U S A* .
319. Hanson MR & Sattarzadeh A (2008) Dynamic morphology of plastids and stromules in angiosperm plants. *Plant Cell Environ* 31(5):646-657.
320. Halazonetis TD, Gorgoulis VG, & Bartek J (2008) An oncogene-induced DNA damage model for cancer development. *Science* 319(5868):1352-1355.
321. Khrapko K & Vijg J (2009) Mitochondrial DNA mutations and aging: devils in the details? *Trends Genet* 25(2):91-98.
322. Fujii S & Toriyama K (2008) Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. *Plant Cell Physiol* 49(10):1484-1494.
323. Redon R, *et al.* (2006) Global variation in copy number in the human genome. *Nature* 444(7118):444-454.
324. Zhang F, *et al.* (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* 41(7):849-853.
325. Perry GH, *et al.* (2007) Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39(10):1256-1260.

326. Payen C, Koszul R, Dujon B, & Fischer G (2008) Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet* 4(9):e1000175.
327. Slack A, Thornton PC, Magner DB, Rosenberg SM, & Hastings PJ (2006) On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet* 2(4):e48.