

Université de Montréal

**Le maintien de la stabilité génomique du plastide :
Un petit génome d'une grande importance**

Par

Étienne Lepage

Département de Biochimie et médecine moléculaire

Faculté de Médecine

Thèse présentée à la Faculté de Médecine en vue de l'obtention
du grade de doctorat en Biochimie

Avril 2015

© Étienne Lepage, 2015

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

**Le maintien de la stabilité génomique du plastide : Un
petit génome d'une grande importance**

Présenté par :

Étienne Lepage

a été évaluée par un jury composé des personnes suivantes :

Luc Desgroseillers, Ph.D., président-rapporteur

Normand Brisson, Ph.D., directeur de recherche

Marlene Oeffinger, Ph.D., membre du jury

Jean-Benoît Charron, Ph.D., examinateur externe

Jean-Claude Labbé, représentant du doyen de la FES

Résumé

Chez les plantes, le génome plastidique est continuellement exposé à divers stress mutagènes, tels l'oxydation des bases et le blocage des fourches de réplication. Étonnamment, malgré ces menaces, le génome du plastide est reconnu pour être très stable, sa stabilité dépassant même celle du génome nucléaire. Néanmoins, les mécanismes de réparation de l'ADN et du maintien de la stabilité du génome plastidique sont encore peu connus.

Afin de mieux comprendre ces processus, nous avons développé une approche, basée sur l'emploi de la ciprofloxacine, qui nous permet d'induire des bris d'ADN double-brins (DSBs) spécifiquement dans le génome des organelles. En criblant, à l'aide de ce composé, une collection de mutants d'*Arabidopsis thaliana* déficients pour des protéines du nucléoïde du plastide, nous avons identifié 16 gènes vraisemblablement impliqués dans le maintien de la stabilité génomique de cette organelle. Parmi ces gènes, ceux de la famille Whirly jouent un rôle primordial dans la protection du génome plastidique face aux réarrangements dépendants de séquences de microhomologie. Deux autres familles de gènes codant pour des protéines plastidiques, soit celle des polymérase de types-I et celle des recombinaisons, semblent davantage impliquées dans les mécanismes conservateurs de réparation des DSBs. Les relations épistatiques entre ces gènes et ceux des Whirly ont permis de définir les bases moléculaires des mécanismes de la réparation dépendante de microhomologies (MHMR) dans le plastide. Nous proposons également que ce type de mécanismes servirait en quelque sorte de roue de secours pour les mécanismes conservateurs de réparation.

Finalement, un criblage non-biaisé, utilisant une collection de plus de 50,000 lignées mutantes d'*Arabidopsis*, a été réalisé. Ce criblage a permis d'établir un lien entre la stabilité génomique et le métabolisme des espèces réactives oxygénées (ROS). En effet, la plupart des gènes identifiés lors de ce criblage sont impliqués dans la photosynthèse et la détoxification des ROS. Globalement, notre étude a permis d'élargir notre compréhension des mécanismes du maintien de la stabilité génomique dans le plastide et de mieux comprendre l'importance de ces processus.

Mots clés : Biologie des plantes, Plastide, Génome, Réarrangement génomique, Réparation de l'ADN, Stabilité génomique, Bris d'ADN double-brin, Séquençage nouvelle-génération

Abstract

The plant plastidial genome is constantly threatened by many mutagenic stresses, such as base oxidation and replication fork stalling. Despite these threats, the plastid genome has long been known to be more stable than the nuclear genome, suggesting that alterations of its structure would have dramatic consequences on plant fitness. At the moment, little is known about the genes and the pathways allowing such conservation of the organelle genome sequences.

To gain insight into these mechanisms, we developed an assay which uses ciprofloxacin, a gyrase inhibitor, to generate DNA double-strand breaks (DSBs) exclusively in plant organelles. By screening mutants deficient for proteins composing the plastid nucleoid on ciprofloxacin, we were able to identify 16 candidate genes, most likely involved in the repair of DSBs in plastid. Among these genes, those of the Whirly family of single-stranded DNA binding proteins are shown to be key factors in protecting the genome from error-prone microhomology mediated repair (MHMR). Two other family of proteins, the plastid type-I polymerases and the plastid recombinases, seem to be involved in the conservative repair pathways. The evaluation of the epistatic relationship between those two genes and the Whirly genes led us to define the molecular basis of MHMR and to propose that they might act as a backup system for conservative repair pathways.

Finally, a non-biased screen, using 50,000 different insertion lines, allowed the identification of numerous genes that were already associated with ROS homeostasis, suggesting a link between DNA repair and ROS imbalance. Globally, our study shed light on the mechanisms that allow the maintenance of plastid genome, while explaining the importance of such conservation of the plastid genome.

Keywords : Plant Biology, Plastid, Genome, Genomic Rearrangement, DNA Repair, Genomic stability, DNA Double-stranded Breaks, Next-generation Sequencing

Table des matières

RÉSUMÉ	V
ABSTRACT	VI
TABLE DES MATIÈRES	VII
LISTE DES FIGURES	XI
LISTE DES TABLEAUX	XV
LISTES D'ABRÉVIATIONS :	XVI
REMERCIEMENTS	XIX
CHAPITRE1: INTRODUCTION	1
1.1 Les plastides	1
1.1.1 L'origine endosymbiotique	2
1.1.2 Le génome des organelles	4
1.1.3 La réplication	7
1.1.4 La transcription du génome plastidique	9
1.1.5 La conservation des séquences interspécifiques.....	11
1.2 Les menaces à l'intégrité du génome plastidique et la réparation de l'ADN	13
1.2.1 La réparation par excision de base	13
1.2.2 La photoactivation	15
1.2.3 La réparation conservatrice des bris d'ADN double-brin	17
1.2.3.1 La recombinaison homologue	17
1.2.3.2 L'appariement dépendant de la synthèse d'ADN	18
1.2.3.3 La réplication induite par des bris	18
1.2.3.4 La recombinaison homologue dans le plastide	20
1.2.4 Les mécanismes sujets à l'erreur.....	20
1.2.4.1 Les mécanismes de liaison d'extrémités	21
1.2.4.2 Les mécanismes dépendants de séquences microhomologues (MHMR).....	22
1.2.3.5 Les endonucléases de homing	25
1.2.3.6 Instabilité génomique induite par la réplication	26
1.2.3.7 La transcription	29
1.3 Les facteurs génétiques de la réparation de l'ADN des organelles chez <i>Arabidopsis thaliana</i>	30
1.3.1 <i>Arabidopsis thaliana</i> , le modèle pour la génétique chez les plantes.....	30
1.3.2 La famille de la recombinase RecA	31
1.3.4 Le rôle de MSH1 dans la protection de l'ADN des organelles	32

1.3.5 La famille de protéines de liaison à l'ADN simple-brin Whirly	33
1.4 Objectifs	34
CHAPITRE 2 : RÉSULTATS PRÉLIMINAIRES.....	37
Un criblage de génétique inverse révèle l'identité des gènes impliqués dans la réparation des bris d'ADN double-brins du plastide.	37
2.1 Mise en contexte :.....	38
2.2 Matériel et méthode :	39
2.3 Résultats :	42
2.3.1 La ciprofloxacine déstabilise le génome plastidique	42
2.3.2 Identification des acteurs du maintien de la stabilité du génome du plastide par criblage génétique inverse	48
2.3.3 Les mutants ayant un taux de transcription réduit accumulent moins de réarrangements dans leur génome plastidique.....	51
2.3.3.1 La mutation de SIG6 complémente le phénotype moléculaire de why1why3	54
2.3.3.2 L'inhibition locale de la transcription protège face aux réarrangements	55
2.3.3.3 L'effet déstabilisant de la transcription est indépendant de la production de DSBs.....	57
CHAPITRE 3: PREMIER ARTICLE	59
Divergent roles for the two Poll-like organelle DNA polymerases of Arabidopsis.	59
3.1 MISE EN CONTEXTE	60
3.2 ABSTRACT	62
3.3 INTRODUCTION.....	63
3.4 RESULTS.....	65
3.5 DISCUSSION	69
3.6 MATERIAL AND METHODS	72
3.6 Supplemental Data	75
3.7 AKNOWLEDGMENTS	75
3.8 LITERATURE CITED.....	75
3.9 FIGURES LEGENDS	78

CHAPITRE 4: DEUXIÈME ARTICLE 89

Organelle DNA rearrangement mapping reveals U-turn-like inversions as a major source of genomic instability in Arabidopsis and humans. 89

4.1 MISE EN CONTEXTE 90

4.2 ABSTRACT 92

4.3 INTRODUCTION 93

4.4 RESULTS 94

4.5 DISCUSSION 99

4.6 METHODS 101

4.7 ACKNOWLEDGEMENTS 106

4.8 FIGURE LEGENDS 107

4.9 REFERENCES 110

CHAPITRE 5: TROISIÈME ARTICLE 133

Plastid genome instability leads to reactive oxygen species production and plastid-to-nucleus retrograde signaling in Arabidopsis..... 133

5.1 MISE EN CONTEXTE 134

5.2 ABSTRACT 136

5.3 INTRODUCTION 137

5.4 RESULTS 139

5.5 DISCUSSION 148

5.6 MATERIALS AND METHODS 152

5.7 SUPPLEMENTARY DATA 155

5.8 ACKNOWLEDGEMENTS 156

5.9 FIGURE TITLES 157

5.10 REFERENCES 161

CHAPITRE 6: QUATRIÈME ARTICLE 193

Time- and Cost-Efficient Identification of T-DNA Insertion Sites through Targeted Genomic Sequencing. 193

6.1 Mise en contexte:..... 194

6.2 Abstract 196

6.3 Introduction.....	197
6.4 Material and Methods	198
6.5 Results.....	200
6.6 Discussion.....	203
6.7 Acknowledgements.....	205
6.8 References.....	205
6.9 Figure Legends	207
CHAPITRE 7 : DISCUSSION	214
7.1 Survol	214
7.2 Les Whirly protègent le génome du plastide face aux réarrangements de types MHMR causés par les DSBs.	214
7.3 La réplication une source de DSBs dans les organelles ?.....	217
7.4 Le stress réplicatif et les réarrangements génomiques indépendants des DSBs.	219
7.5 La transcription et l'instabilité génomique du plastide.....	221
7.6 Un rôle pour les autres gènes identifiés par le criblage de génétique inverse	223
7.7 L'importance de la stabilité génomique dans le plastide	224
8. CONCLUSION	226
9. RÉFÉRENCES	228

Liste des figures

Figure 1: Étapes de différenciation menant aux principaux types de plastes.....	2
Figure 2 : Sort des gènes d'origine cyanobactérienne à la suite des événements d'endosymbiose chez <i>Arabidopsis thaliana</i>	4
Figure 3: Représentation schématique du génome d' <i>Arabidopsis</i>	5
Figure 4: Mécanisme de réplication de l'ADN selon le modèle de la boucle-D.....	7
Figure 5 : Mécanisme d'action des gyrases.....	9
Figure 6: L'arrangement des gènes sur le génome du plastide est fortement conservé entre les espèces.....	12
Figure 7: Structure des bases azotées lors de réaction d'oxydation.....	15
Figure 8: Modification de la structure des pyrimidines adjacentes par l'action des rayons UV.....	16
Figure 9: Mécanismes de réparation de l'ADN dépendant de la recombinaison homologue.....	19
Figure 10: Mécanismes de réparation de l'ADN par jonction d'extrémités.....	24
Figure 11: Un système conditionnel permettant d'exprimer une endonucléase dans les plastides d' <i>Arabidopsis</i>	26
Figure 12 : Mécanisme de régression de fourche permettant de redémarrer une fourche bloquée.....	28
Figure 13: La ciprofloxacine déstabilise le génome plastidique d' <i>Arabidopsis</i>	43
Figure 14: Le traitement à la ciprofloxacine affecte le taux de survie chez <i>Arabidopsis</i>	44
Figure 15: Approche PCR permettant d'amplifier spécifiquement des molécules D'ADN réarrangées dans le génome du plastide.....	45
Figure 16 La ciprofloxacine cause l'apparition de réarrangements génomiques et de structures sous-génomiques.....	46
Figure 17: La ciprofloxacine et la novobiocine inhibent l'action des gyrases à des étapes différentes du mécanisme réactionnel.....	48
Figure 18: Les mutants <i>sig2</i> et <i>sig6</i> et le traitement à la rifampicine réduisent les niveaux de réarrangements génomiques dans le plastide.....	52

Figure 19: Les taux de transcription diminuent dans les mutants <i>sig2</i> et <i>sig6</i>	53
Figure 20: L'inhibition de la transcription à haute température chez <i>hsp21</i> diminue l'apparition de réarrangements génomiques.	54
Figure 21: La mutation de <i>SIG6</i> complémente celle des Whirly au niveau de l'apparition de réarrangements génomiques.	55
Figure 22 : La mutation de <i>SIG5</i> empêche l'apparition de réarrangements génomiques autour du gène <i>PSBD</i>	56
Figure 23: L'inhibition de la transcription par la rifampicine empêche la formation de réarrangements induits par un DSB artificiel.	57
Figure 24. Isolation of PollA and PollB mutant lines.	80
FIGURE 25. Organelle DNA is less abundant in seedlings of <i>polla-1</i> and <i>pollb-1</i> mutants.	79
Figure 26. PollB mutants exhibit an increased sensitivity to ciprofloxacin (CIP) and are impaired in their ability to process DNA DSBs.	82
Figure 27. A <i>pollb-1/why1why3</i> triple mutant yields a pale-green dwarf phenotype and gives rise to abundant DNA rearrangements in plastids.	84
Figure 28. Organelle DNA is less abundant in seedlings of <i>polla-2</i>	85
Figure 29. Ciprofloxacin does not generate DSBs in the nuclear genome.	86
Figure 30. Crosses between Poll and Whirly mutants.	87
Figure 31. Global portrait of organelle genome rearrangements in Arabidopsis and human.	114
Figure 32. Analysis of organelle genome rearrangements in Arabidopsis and human.	115
Figure 33. Global portrait of plastid genome rearrangements in Arabidopsis mutant lines <i>pollb</i> , <i>reca1</i> and <i>why1why3</i>	117
Figure 34. Global portrait of plastid genome rearrangements in Arabidopsis mutant lines <i>reca1pollb</i> , <i>why1why3pollb</i> and <i>why1why3reca1</i>	118
Figure 35. Plastid DNA sequencing coverage curves for Arabidopsis lines WT, <i>why1why3pollb</i> and <i>why1why3reca1</i>	120
Figure 36. Model of microhomology-mediated U-turn-like inversions.	121

Figure 37. Level of plastid DNA rearrangements in three Arabidopsis ecotypes.....	122
Figure 38. Characterization of the <i>reca1-1</i> and <i>reca1-2</i> T-DNA insertion mutant lines.	123
Figure 39. Characterization of the <i>why1why3reca1</i> phenotype.....	124
Figure 40. Plastid DNA sequencing coverage curves for WT, <i>pollb</i> , <i>reca1</i> , <i>reca1pollb</i> and <i>why1why3</i> Arabidopsis mutant plants.	125
Figure 41. Plastid DNA quantification at three locations of the genome in WT, <i>why1why3pollb</i> and <i>why1why3reca1</i> plants.	126
Figure 42. Visualization of the distribution of the various forms of the plastid genome	127
Figure 43. Schematic representation of the analysis workflow.	128
Figure 44. Comparison of the techniques used to detect genome rearrangements.....	129
Figure 45. Local mapping of read pairs associated to specific genome rearrangements.	130
Figure 46. Plastid DNA Rearrangements in <i>why1why3pollb-1</i> Lead to Photosynthetic Impairment and Yellow Variegation.	170
Figure 47. <i>why1why3pollb-1</i> Produces Elevated Levels of Chloroplast ROS in a Light-Dependent Manner.....	171
Figure 48. Ciprofloxacin-Treated Seedlings Show Signs of Bleaching Associated With Increased Chloroplast ROS Production.	172
Figure 49. Growth under Low Light Conditions Rescues the <i>why1why3pollb-1</i> Phenotype.....	173
Figure 50. The <i>why1why3pollb-1</i> Yellow-Variegated Sectors Are not Devoid of Photosynthetically Active Chloroplasts.....	174
Figure 51. The <i>why1why3pollb-1</i> ptDNA Rearrangements and Phenotype Are not Maternally Inherited.....	175
Figure 52. The <i>why1why3pollb-1</i> Mutant Exhibits a Complex Nuclear Genetic Reprogramming. ..	176
Figure 53. The <i>why1why3pollb-1</i> Nuclear Genetic Reprogramming Is Not Observed under Low-Light Conditions.....	178
Figure 54. The Early Oxidative Stress in <i>why1why3pollb-1</i> Leads to Acquired Resistance to High Light.	179

Figure 55. Microhomology-Mediated Break-Induced Replication (MMBIR) Leads to DNA Rearrangements.	181
Figure 56. The <i>why1why3pollb-1</i> Mutant Develops Distinct Yellow and Green Sectors on Its Leaves.	182
Figure 57. Some <i>why1why3pollb-1</i> Plants Exhibit Yellow and White Variegation Simultaneously.	183
Figure 58. Ciprofloxacin Negatively Impacts Photosynthetic Efficiency in Arabidopsis.	184
Figure 59. <i>why1why3</i> and <i>why1why3pollb-1</i> Plants Are Hypersensitive to Paraquat.	185
Figure 60. The Abundance of MMBIR-Associated ptDNA Rearrangements Is Little Affected by the Light-Regimen in <i>why1why3pollb-1</i>.	186
Figure 61. Paraquat and Atrazine do not induce MMBIR-Associated ptDNA Rearrangements.	187
Figure 62. Elevated ROS Levels Do not Lead to Cell Death in <i>why1why3pollb-1</i>.	188
Figure 63. The <i>why1why3pollb-1</i> Nuclear Genetic Reprogramming Is Not Due to the <i>Pollb-1</i> and <i>why1why3</i> Genetic Backgrounds.	189
Figure 64. Forward Genetic Screen to Identify Genes Involved in the Maintenance of Organelle Genome Topology.	209
Figure 65. Overview of Targeted Genomic Sequencing. Blue rectangles represent genomic DNA, and red rectangles correspond to T-DNA insertions.	210
Figure 66. Coverage of the pSKI015 Vector Obtained by Sequencing.	211
Figure 67. Association of an Insertion Event to a Specific Line by 2D-PCR Pooling	212
Figure 68. Schematic Illustration of the Insertion Sites in the Three Novobiocin-Sensitive Mutant Lines.	213
Figure 69: Modèle pour la réparation des DSBs dans le plastide d'Arabidopsis.	217

Liste des tableaux

Tableau I : Gènes identifiés par le criblage de génétique inverse pour les mutants du nucléoïde, lors d'un traitement à la ciprofloxacine.	49
Tableau II Workflow Statistics for <i>Arabidopsis</i> plastid DNA rearrangements.	131
Tableau III Workflow Statistics for <i>Arabidopsis</i> ecotypes Ts-1 and Ws-2 plastid DNA rearrangements.....	131
Tableau IV Workflow Statistics for <i>Arabidopsis</i> mitochondria DNA rearrangements.	131
Tableau V Workflow Statistics for human brain mitochondria DNA rearrangements.	132
Tableau VI Workflow Statistics for human liver mitochondria DNA rearrangements.	132
Tableau VII Workflow Statistics for <i>E. coli</i> DNA rearrangements.	132
Tableau VIII. Primer Pairs Used for the Specific Amplification of MMBIR-Associated ptDNA Rearrangements and Described Previously (196).	190
Tableau IX. Primer Pairs Used for Quantitative RT-PCR experiments.	190

Listes d'abréviations :

pb : Paire de bases

LSC : Grande séquence unique

SSC : Petite séquence unique

IR : Séquences répétées inversées

PEP : Polymérase d'ARN encodée par le plastide

NEP : Polymérase d'ARN encodée par le noyau

ROS : Espèces réactives oxygénées

UV : Ultraviolet

BER : Réparation par excision de bases

CPD : Dimères cyclobutane de pyrimidines

DSB : Bris d'ADN double-brin

HR : Recombinaison homologue

DSBR : Réparation de bris d'ADN double-brin

SDSA : L'appariement dépendant de la synthèse d'ADN

BIR : Réplication induite par des bris

NHEJ : Liaison d'extrémités non-homologue

MHMR : Mécanismes dépendants de séquences microhomologues

MMEJ : Liaison d'extrémités dépendante de microhomologies

SSA : L'hybridation de simple-brin

MMBIR : Réplication induite par des bris dépendante de microhomologies

MMSDSA: L'appariement dépendant de la synthèse d'ADN à des sites microhomologues

PCR : Réaction de polymérisation en chaîne

WT : Type sauvage

CIP : Ciprofloxacine

Nov : Novobiocine

SSB : Single-strand binding protein

MS : Murashige-Skoog

HSP : Protéine de tolérance à la chaleur

*Pour examiner la vérité, il est besoin,
une fois dans sa vie, de mettre toutes
choses en doute autant qu'il se peut.*

René Descartes

Remerciements

Il est souvent difficile de décrire à nos proches les raisons qui nous poussent à nous lancer dans la périlleuse aventure de la recherche aux cycles supérieurs. Il est souvent difficile de décrire cette force qui nous pousse à toujours vouloir repousser les limites de nos connaissances. Cette thèse, qui est l'aboutissement de nombreuses années de travail, représente beaucoup plus à mes yeux qu'un simple amalgame d'articles. Elle me rappelle, ces six dernières années, marquées de rencontres mémorables, qui m'ont transformé profondément.

Mes premières pensées sont dirigées vers Normand, qui m'a accueilli à bras ouverts dans son laboratoire. Normand, ton approche favorisant le travail d'équipe et l'initiative m'a permis de développer des qualités de chercheurs que je n'aurais probablement pas su acquérir autrement. Même dans mes moments de doutes, tu as toujours cru en moi. Merci énormément.

L'arrivée dans un laboratoire est une expérience toujours marquante. Je n'oublierai jamais le groupe de jeunes chercheurs qui m'ont accueilli. Alexandre, tu as su allumer en moi la flamme de la curiosité scientifique. Laurent, tu as été un modèle par ta force tranquille et ta profonde gentillesse. Jean-Sébastien, ce fut un réel plaisir de travailler avec toi. Il est rare que deux personnes créent des liens aussi rapidement que nous l'avons fait. J'espère que tu auras profité autant de ma présence que j'ai apprécié la tienne.

Je ne peux non plus oublier l'équipe actuelle du laboratoire Brisson. La réputation du laboratoire n'est plus à faire et l'ambiance qui y règne pourrait rendre n'importe quelle tâche agréable. Quelque temps après mon arrivée, un jeune scientifique dynamique venait nous rejoindre. Je ne pouvais me douter à ce moment que j'y trouverais un partenaire fidèle avec qui j'allais partager la plupart de mes projets. Je tiens à te remercier Éric pour tout ce beau travail. Plus récemment, un drôle de personnage du nom de Samuel est arrivé au laboratoire. Étrange au premier abord, j'ai vite su percer sa façade et mettre à jour un être humain exceptionnel. J'espère que tu auras apprécié autant que moi toutes ces heures partagées.

Finalement, une personne formidable a croisé mon chemin au milieu de mon doctorat. Marie, tu es la personne qui a eu le plus profond impact sur ma vie. Étant bourreau de travail de nature, ta rencontre est venue remettre en perspective ce que je désire réellement. Tu as su remplir mon quotidien de bonheur et d'amour. Tu m'inspires à trouver le meilleur de moi-même. Il faut beaucoup de courage

pour suivre un jeune chercheur au milieu du labyrinthe de la recherche académique dans ce Québec morose et un peu incertain. Je t'en remercie. Je t'aime.

Ainsi s'achèvent mes études doctorales et commence un nouveau chapitre de ma vie. Prochain défi, être papa d'une merveilleuse petite crevette d'amour.

Chapitre 1 : Introduction

1.1 Les plastides

D'un point de vue métabolique, les végétaux se démarquent par leur capacité de transformer l'énergie solaire en énergie chimique par le biais de la photosynthèse (revue dans (1, 2)). Cette transformation, fondamentale au maintien de la vie sur Terre, nécessite comme substrat de base des molécules d'eau et du gaz carbonique. L'énergie solaire captée par la plante permet alors une suite de réactions anaboliques entraînant la formation de molécules organiques de complexité croissante, jusqu'à l'obtention du glucose. La photosynthèse est d'ailleurs l'un des seuls processus effectués par le vivant capable d'utiliser une autre forme d'énergie que celle emmagasinée dans les liens chimiques (revue dans (2)). Le glucose, ainsi stocké par les organismes photosynthétiques, fournit donc la vaste majorité de l'énergie aux systèmes vivants sur Terre. La photosynthèse est un processus métabolique complexe et une partie importante du métabolisme des plantes est consacrée à ce seul processus. En effet, 75 des 100 gènes les plus exprimés chez la plante *Arabidopsis thaliana* sont directement impliqués dans cette fonction métabolique (Lepage É., résultats non publiés).

Le siège de la photosynthèse chez les plantes se trouve dans le chloroplaste, une organelle spécialisée de la famille des plastides (revue dans (3)). Bien que la photosynthèse soit d'une importance capitale, il existe plusieurs autres membres à la famille des plastides qui remplissent des fonctions variées. Afin de remplir leurs rôles, les plastides doivent d'abord effectuer une étape de différenciation qui leur permet d'acquérir les caractéristiques essentielles à leurs fonctions. Les plastides débutent leur cycle de différenciation sous forme de proplastides, qui sont des plastides pluripotents retrouvés principalement dans le méristème (tissu qui sert de réserve de cellules souches chez la plante et qui est le berceau de la différenciation cellulaire). Lors de la différenciation cellulaire, les plastides adéquats seront produits selon le type de cellule généré. Les trois principaux types de plastides différenciés sont les chloroplastes, se situant dans les tissus aériens pour la photosynthèse, les amyloplastides, étant responsables du stockage de l'amidon, et les chromoplastes, servant à emmagasiner les pigments colorés dans les fleurs, les fruits et certaines racines. Le plastide est considéré comme l'un des principaux acteurs dans la perception des signaux de l'environnement chez les plantes, en grande partie grâce à sa capacité à percevoir l'intensité lumineuse. Il s'agit d'organelles

dynamiques qui s'adaptent aux conditions environnementales. L'inter-conversion entre les types de plastides est d'ailleurs observable lors des fluctuations de ces conditions (Figure1).

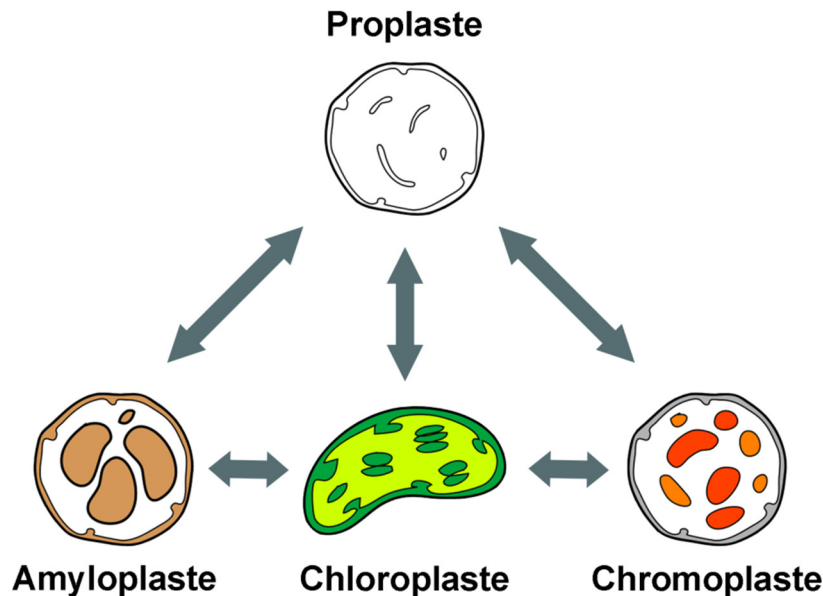


Figure 1: Étapes de différenciation menant aux principaux types de plastides. Les proplastides peuvent mener à tous les types de plastides différenciés et ceux-ci ont la capacité de réaliser certains types d'inter-conversions.

1.1.1 L'origine endosymbiotique

L'hypothèse qui stipule que les plastides proviennent d'une association symbiotique entre deux organismes remonte au début du XX^e siècle (4). Cette hypothèse, d'abord supportée par la ressemblance physiologique entre les plastides et certaines bactéries autotrophes, fut par la suite appuyée par plusieurs autres évidences biochimiques, telles la présence d'un génome et l'utilisation de ribosomes procaryotes (5, 6). Le mécanisme par lequel les végétaux ont acquis cette organelle se nomme l'endosymbiose et repose sur la phagocytose d'une bactérie. Selon un consensus général, la bactérie phagocytée dans le cas du plastide ferait partie de l'embranchement des cyanobactéries, bactéries qui ont la capacité de faire de la photosynthèse (revue dans (7)). Puisque l'incorporation de cette bactérie présentait un avantage à l'hôte, l'endosymbiote a alors été conservé par la cellule.

L'endosymbiote ainsi acquis est entouré de deux à quatre bicouches de phospholipide, deux provenant de la cyanobactérie, les autres étant des vestiges de la phagocytose (revue dans (7)).

Une seconde étape essentielle à l'établissement de l'endosymbiose est la coordination au niveau génétique de l'endosymbiote avec le génome nucléaire de l'hôte. Suivant l'étape de phagocytose, plus de 90% des gènes du plastide sont transférés vers le noyau. Ce transfert offre plusieurs avantages à l'hôte, puisqu'il permet un meilleur contrôle de l'organelle, rendant celle-ci par le fait même dépendante de l'hôte (revue dans (8)). Il est intéressant de constater que l'hôte est lui aussi dépendant de son endosymbiote et que les mutations qui empêchent la biogenèse du chloroplaste chez les plantes sont souvent embryono-létales (9). On peut alors parler d'une interdépendance.

Le transfert de gènes a grandement influencé le génome nucléaire et une analyse bio-informatique a permis de déterminer qu'environ 18% des gènes nucléaires d'*Arabidopsis* étaient d'origine cyanobactérienne (10). Les gènes désormais retrouvés dans le noyau ont soit acquis une nouvelle séquence en 5' du gène, un peptide transit, assurant le transport de la protéine vers le plastide ou obtenu une nouvelle fonction cellulaire à l'extérieur du plastide. Ce processus d'évolution génomique de l'endosymbiote est présenté dans la figure 2.

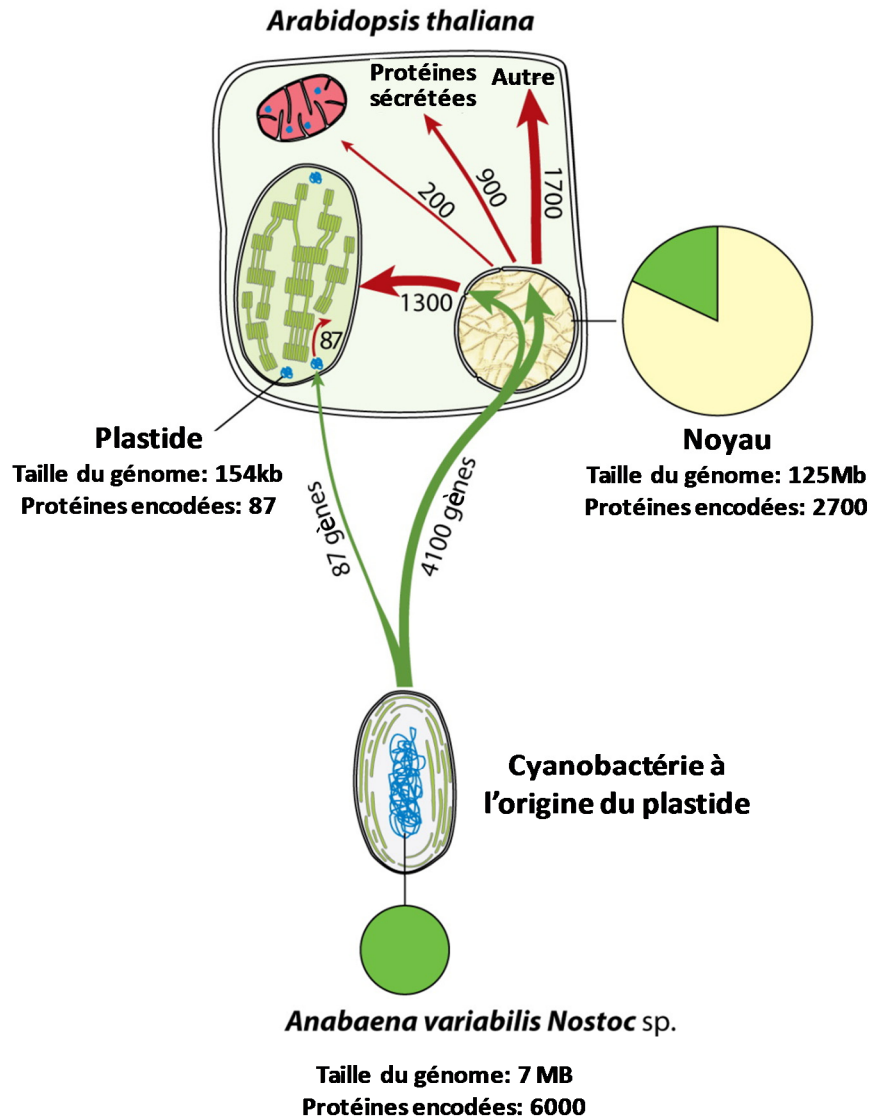


Figure 2 : Sort des gènes d'origine cyanobactérienne à la suite des événements d'endosymbiose chez Arabidopsis thaliana. La destinée des gènes de la cyanobactérie est représentée par les flèches vertes. Les flèches rouges représentent la localisation prédite par des études bio-informatiques des gènes d'origine cyanobactérienne acquis par le noyau. Les diagrammes en pointe de tarte présentent le ratio de gènes d'origine cyanobactérienne contenu dans les génomes étudiés (Figure adaptée de (11)).

1.1.2 Le génome des organelles

Le génome du chloroplaste se distingue à bien des égards. En plus d'avoir une taille très réduite, variant entre 120 et 180 kilobases chez les plantes terrestres (revue dans (12)), son niveau de conservation entre les espèces est sans pareil. Les gènes qu'on y retrouve sont hautement conservés,

que ce soit au niveau de leur séquence nucléotidique qu'au niveau de leur arrangement sur le génome. Le génome est généralement composé de deux séquences répétées inversées (IR) séparées par une grande et une petite séquences uniques, comme présenté dans l'exemple à la figure 3. Les IRs font habituellement quelques kilobases de long et encodent principalement les ARN et les protéines ribosomiques. Les 90 à 100 gènes restants sont encodés par le génome plastidique et se retrouvent dans les copies uniques. Ils sont essentiellement impliqués dans les processus de la photosynthèse et de l'expression génique. Les raisons qui supportent la conservation de cet arrangement particulier du génome demeurent encore inconnues.

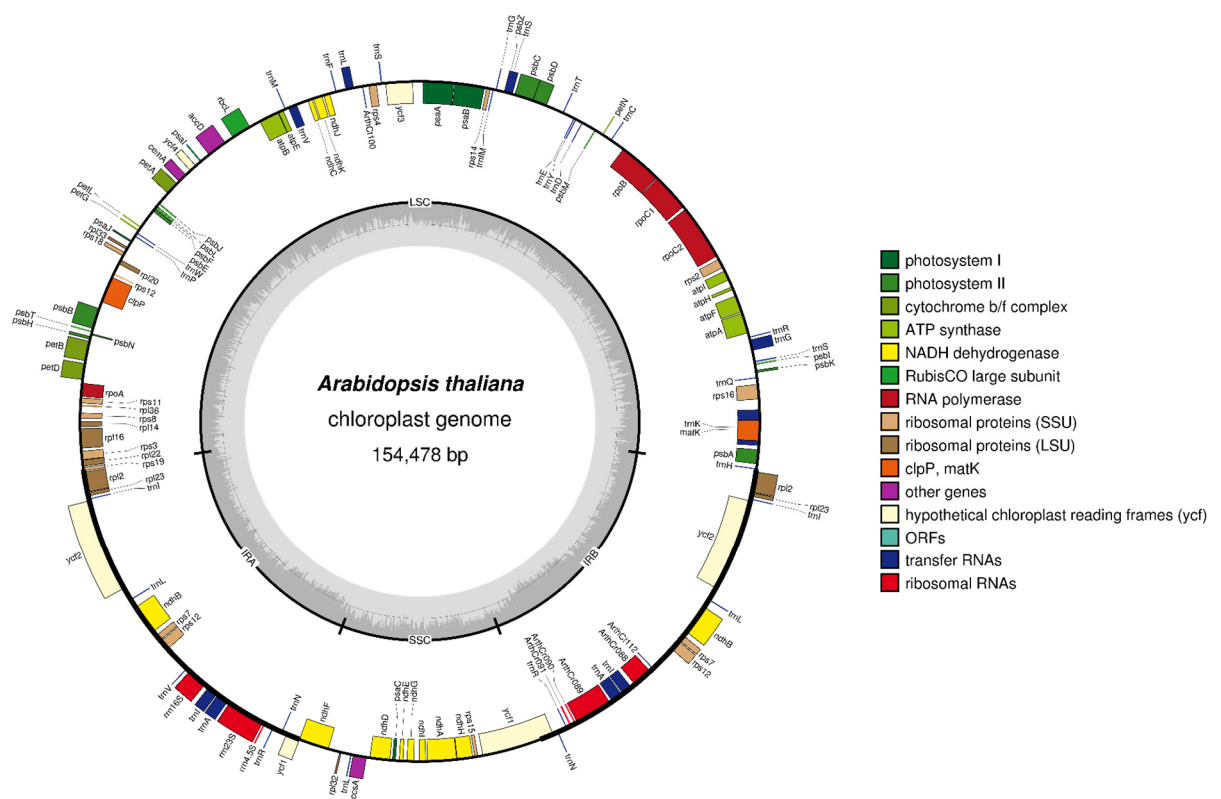


Figure 3: Représentation schématique du génome d'Arabidopsis. Carte générée à l'aide du logiciel OGDraw (13). La couleur par laquelle est représenté un gène est attribuée selon leur fonction principale. La position du gène sur le cercle (à l'intérieur ou à l'extérieur) représente le brin sur lequel le gène est encodé. L'histogramme en gris, au centre du cercle, représente le contenu en GC. Les termes LSC, SSC et IR représentent respectivement la grande séquence unique, la petite séquence unique et les copies inversées.

Comme chez la plupart des organismes, l'ADN du génome plastidique ne se retrouve pas seul dans le plastide, mais plutôt en complexe avec de nombreuses protéines. Ce complexe, nommé nucléoïde, contient des protéines impliquées dans les mécanismes de réplication de l'ADN, de transcription génique, ainsi que dans la traduction des ARN en protéines (14). Cette organisation suggère que, tout comme chez les procaryotes, les gènes du plastide sont transcrits et traduits en simultané. Des études récentes de spectrométrie de masse ont permis de dresser une liste plus exhaustive des protéines qui composent le nucléoïde (15). Plus de 300 protéines ont été identifiées dans ce complexe. On sait que la plupart sont impliquées dans la maturation de l'ARN, ce qui suggère que ce processus est particulièrement complexe dans le plastide. Sans surprise, 33 autres protéines identifiées dans cette étude sont homologues à de protéines impliquées dans le maintien de la stabilité génomique.

Le génome du plastide se retrouve dans un état de polyploïdie à l'intérieur de la cellule végétale (16). En effet, chaque cellule contient en moyenne une centaine de plastides chez *Arabidopsis*, et chacun de ces plastides possède plusieurs copies de leur génome (17). Cette polyploïdie peut avoir de nombreux effets sur l'évolution du génome plastidique. D'abord, la plupart des mutations observées dans le plastide se retrouve dans un état d'hétéroplasmie, signifiant que seulement une fraction des copies du génome a été modifiée. Cet état permet une dérive lente entre les différentes formes du génome et peut mener à la fixation des mutations qui ont un effet bénéfique (18). Au contraire, les mutations délétères sont généralement sélectionnées négativement dans un tel système.

Lors de la reproduction sexuée de la plante, les plastides sont généralement transmis par les cellules maternelles, ce qui a pour conséquence de transmettre exclusivement l'information génétique contenue dans les plastides de la mère. Les organismes qui utilisent ce mode d'héritabilité sont généralement soumis à la loi de Muller's Ratchet qui stipule que les mutations devraient s'accumuler au fil des générations à cause du faible niveau d'échange d'information génétique associé à ce mode de reproduction (19). Pourtant, au contraire, le génome plastidique est très stable et accumule peu de mutations. L'explication de cette particularité provient encore une fois de la polyploïdie qui rend possible d'échanger de l'information entre les chromosomes sœurs au sein d'un même plastide et qui permet ainsi d'échapper à l'effet Muller's Ratchet (20).

1.1.3 La réplication

Les mécanismes de réplication de l'ADN du plastide demeurent peu définis, et bien que plusieurs modèles aient été proposés à ce jour, ceux-ci sont encore contestés. À l'origine, l'observation de molécules d'ADN circulaires dans les plastides a mené à l'élaboration de différents modèles inspirés de la réplication observée chez les procaryotes. Parmi eux, celui de la boucle-D a été le plus largement accepté(21). Ce modèle repose sur des observations, faites par microscopie électronique, qui ont permis de remarquer la formation de boucles de déplacement de l'ADN qui progressent et s'élargissent sur les molécules d'ADN du plastide (21). Ce modèle propose que la réplication débute de manière unidirectionnelle et convergente à partir de deux origines de réplifications distinctes (Figure 2A). Cette réplication se poursuit ainsi jusqu'à ce que les deux fourches de réplications se rejoignent (Figure 2B). À ce point, la réplication s'établit selon un mode de synthèse bidirectionnel (Figure 2C) (brin continu et brin retardé simultanément) jusqu'à ce que la synthèse d'un brin continu rencontre le brin retardé synthétisé en aval (Figure 2D). Le brin continu est alors ligué au brin retardé et l'on obtient deux copies du génome original.

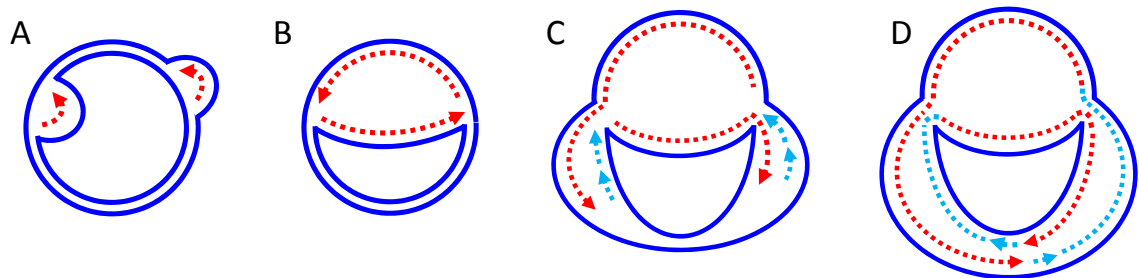


Figure 4: Mécanisme de réplication de l'ADN selon le modèle de la boucle-D. Le génome d'origine est représenté en bleu foncé, le brin direct en rouge et le brin retardé en bleu pâle. A) Début du mécanisme par la formation de deux boucles de déplacement de l'ADN (boucle-D) qui s'établissent aux origines de réplifications. B) Réplication de l'ADN à l'aide de deux fourches unidirectionnelles et convergentes. C) Initiation de la réplication bi-directionnelle une fois arrivée aux origines de réplification. D) Arrêt de la réplication lorsque le brin direct rencontre le brin retardé synthétisé par l'autre fourche. Les brins directs et retardés sont ligués pour terminer la réplication.

Les modèles basés sur l'observation de molécules d'ADN circulaire ont cependant été remis en doute par l'observation plus récente de plusieurs types de structures d'ADN dans le génome du plastide. En

effet, initialement considéré comme étant principalement circulaire, plusieurs évidences suggèrent maintenant que le génome du plastide serait plutôt composé de structures beaucoup plus complexes (22). Des études de microscopie et de migration sur gel d'agarose ont permis d'observer qu'une fraction importante des chromosomes plastidiques était davantage linéaire et formait souvent de gros complexes branchés multimériques (23). Un modèle plus récent propose que la réplication de l'ADN du plastide serait dépendante de la recombinaison, ce qui est en accord avec le type de molécules observées dans le plastide (22, 23). Néanmoins, les données expérimentales sont encore insuffisantes pour établir clairement ce modèle.

La machinerie de réplication de l'ADN du plastide demeure elle aussi peu caractérisée. Les plantes possèdent deux polymérases de type-I, similaires aux polymérases processives des procaryotes et ayant une localisation prédite au chloroplaste (24). Des études *in vitro* ont démontré que ces polymérases possédaient des taux de réplication assez élevés pour assurer la réplication complète du génome (24-27). Cependant, des études *in vivo* demeurent nécessaires pour confirmer le rôle de ces polymérases dans leur contexte biologique.

La réplication génère du surenroulement dans l'ADN, pouvant avoir des effets néfastes sur l'accessibilité des gènes (revue dans (28)). Afin de maintenir le surenroulement à un niveau optimal, le plastide possède des gyrases et des topoisomérases qui, par leur action, peuvent augmenter ou diminuer le niveau de surenroulement de l'ADN (29, 30). La vaste majorité des plantes possèdent deux gènes de gyrases, *GYRA* et *GYRB*, ayant une forte similarité avec les gyrases procaryotes. Tout comme leurs homologues procaryotes, les gyrases de plantes agissent sous forme de tétramères composés de deux unités de *GYRA* et deux unités de *GYRB* (30). Leur mécanisme d'action, présenté à la figure 5, nécessite de générer un bri d'ADN double-brin (DSB) afin de permettre le passage des deux brins l'un dans l'autre et générer ce qui est appelé du surenroulement négatif (revue dans (28)). Ce processus nécessite l'utilisation d'une molécule d'ATP et permet de diminuer le surenroulement positif causé par la réplication. La mutation des gyrases est embryo-létale chez *Arabidopsis*, supportant ainsi le rôle essentiel de ces protéines (30). Aussi, l'inhibition de ces enzymes à l'aide d'antibiotiques a permis de confirmer leur implication dans le surenroulement de l'ADN *in vivo* ainsi que leur importance dans la réplication de l'ADN du plastide (30).

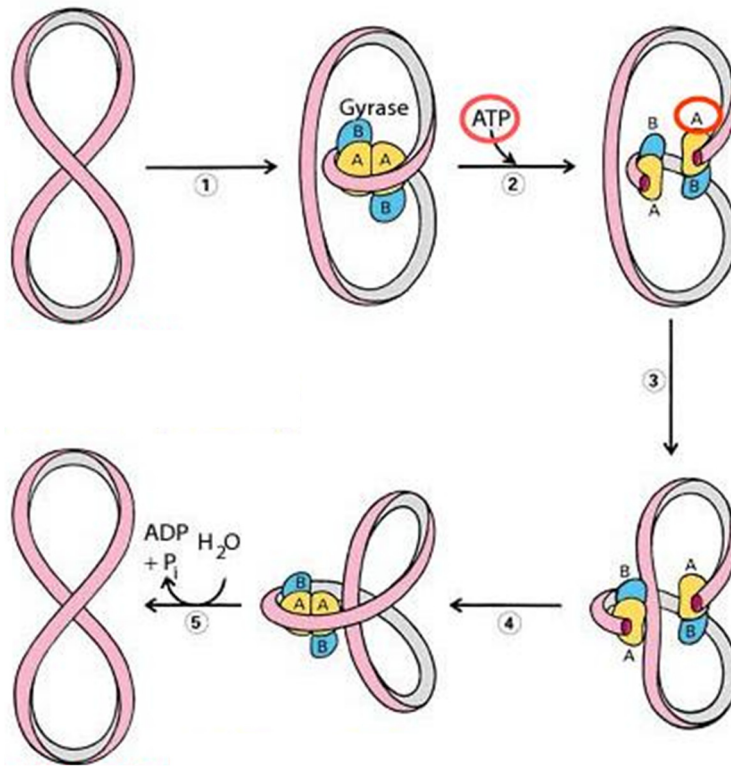


Figure 5 : Mécanisme d'action des gyrases. 1) Un tétramère de gyrases, composé de deux GYRAs et deux GYRBs se fixe à l'ADN. 2) Le complexe de gyrases utilise l'énergie d'une molécule d'ATP pour induire une cassure dans l'ADN double brin. 3) L'ADN passe à travers le bris formé pour générer un surenroulement positif. 4) L'ADN est religué par les gyrases. 5) Les gyrases quittent l'ADN.

1.1.4 La transcription du génome plastidique

Plusieurs centaines de protéines sont nécessaires au bon fonctionnement du plastide (13). Puisque ces protéines sont encodées en partie par le génome nucléaire, la coordination de l'expression des protéines entre le plastide et le noyau est essentielle. La transcription des gènes encodés par le génome du chloroplaste est assurée par trois polymérases d'ARN, soit une ARN polymérase de type procaryote qui est encodée par le plastide (PEP) et deux ARN polymérases d'origine phagique encodées dans le noyau (NEP) (31). La PEP est composée d'un complexe de quatre sous-unités protéiques, RPOA-B-C1-C2, tandis que les NEP sont composées de monomères encodés par les gènes *SCA3* et *RPOTmp* (32). L'importance relative de ces protéines varie au fil du développement. Les NEPs sont particulièrement importantes dans les premiers stades et permettent l'établissement de la

transcription dans le plastide (33). Des études ont démontré que le noyau possède tous les facteurs nécessaires à sa transcription au moment de la germination, contrairement aux plastides qui seraient inactifs au niveau transcriptionnel. Le modèle actuel propose donc que les NEPs seraient exprimées au moment de la germination et permettraient alors la transcription des gènes chloroplastiques, incluant ceux encodant la PEP. Une fois la PEP exprimée à un niveau suffisant, il y aurait une transition dans le patron d'utilisation des ARN polymérase au profit de la PEP (34). Ce modèle a cependant été récemment remis en question, suite à la caractérisation d'une activité résiduelle de la PEP dans les graines (35).

Des plantes mutantes *sca3* ont pu être isolées chez *Arabidopsis*, permettant d'en apprendre davantage sur le processus de la transcription (36). Ces mutants nécessitent un apport en glucose afin de croître, laissant supposer que RPOTmp et la PEP ne complètent que partiellement la perte de SCA3. Néanmoins, plus tard au cours de leur développement, l'expression des gènes du plastide se rétablit chez *sca3* permettant une survie autotrophe, ce qui supporte un rôle accru de la NEP dans les premiers stades de développement de la plante. À des stades plus avancés, les NEP sont principalement impliquées dans la transcription des gènes de maintenance (house-keeping), tandis que la PEP prend en charge l'expression des gènes liés à la photosynthèse (34, 36).

Une famille de facteurs de transcription d'origine procaryote, similaire aux facteurs de transcription sigma70, est essentielle à la reconnaissance de la vaste majorité des promoteurs sur le génome plastidique (revue dans (37)). Chez *Arabidopsis*, cette famille est composée de six membres, tous encodés par le noyau. Quatre membres (*SIG1-3-4-5*) sont spécialisés pour la reconnaissance de gènes spécifiques, tandis que deux autres (*SIG2* et 6) sont généraux et régulent l'expression de la majorité des gènes plastidiques. De par leur rôle central dans la transcription, la mutation de *SIG2* ou *SIG6* mène à des problèmes photosynthétiques importants associés à une diminution marquée de l'expression de plusieurs gènes chloroplastiques (38-40). Au contraire, la mutation des autres facteurs sigma n'a aucun effet sur le phénotype des plantes. Tous les facteurs sigma interagissent avec la PEP afin de moduler son activité. Le facteur sigma *SIG5* est le plus divergent au niveau de sa séquence et son affinité pour la PEP est plus faible dans des conditions optimales de croissance. En situation de stress cependant, sa liaison est favorisée et stimule alors la transcription de ces gènes cibles, en faisant un facteur critique dans la réponse à certains stimuli (41). Avec leurs activités et leurs cibles différentes, les facteurs sigmas sont des acteurs importants dans la régulation de l'expression génique du plastide et leur modulation permet un contrôle global de ce processus.

La transcription dans le plastide possède des particularités supplémentaires. À l'image des procaryotes, les gènes sont organisés en opérons et la transcription se fait en bloc. Ces blocs sont ensuite modifiés pour en isoler les gènes grâce à des étapes d'épissage qui nécessitent de très nombreuses protéines (revue dans (42)). La plupart des opérons sollicitent des facteurs spécifiques pour leur épissage et leur maturation, ce qui explique en partie la panoplie de protéines de liaison à l'ARN retrouvée dans le plastide. Une autre particularité de la transcription plastidique est que tout le génome est actif au niveau transcriptionnel, aussi bien l'espace entre les opérons que les brins anti-sens (43). Cette découverte surprenante est venue clarifier le rôle de plusieurs ribonucléases, retrouvées dans les plastides, qui digèrent les séquences non-codantes ayant été transcrites et qui pourraient avoir des effets indésirables (par exemple, ARN anti-sens). La mutation de certaines de ces ribonucléases occasionne une inhibition de la digestion d'ARN anti-sens empêchant l'expression des gènes associés (44). Ces résultats ont aussi permis de soulever l'hypothèse que les ARN anti-sens joueraient un rôle dans la régulation de l'expression des gènes du plastide .

1.1.5 La conservation des séquences interspécifiques

L'importance d'une séquence d'ADN peut être évaluée par son niveau de conservation entre les espèces (revue dans (45)). Généralement, plus la pression sélective exercée sur une séquence est élevée, plus cette séquence sera conservée. Pour sa part, tout le génome du chloroplaste est fortement conservé, ce qui supporte son importance primordiale (46). Les séquences des gènes, de même que leur orientation, sont généralement conservées dans des espèces éloignées, tel que montré à la Figure 6. Seules quelques inversions de gènes ou de locus entier sont plus fréquemment observées, comme dans l'exemple présenté ici pour le soya (*Glycine max*) (47). Certains résultats expérimentaux ont permis de proposer une hypothèse selon laquelle il y aurait une forte pression sélective sur les gènes du plastide. En effet, les niveaux de fixation des mutations non-synonymes sont particulièrement faibles dans les gènes responsables de la photosynthèse, pouvant être jusqu'à dix fois plus faibles que dans les régions non-codantes du chloroplaste (48, 49). Ceci suggère que les gènes responsables de la photosynthèse seraient optimisés pour réaliser leur fonction et que la vaste majorité des modifications dans ces séquences aboutiraient en une diminution de la compétitivité de l'organisme. Bref, ces mutations seraient contre-sélectionnées. Des études phylogéniques ont également permis de déterminer des différences significatives au niveau de la fixation des mutations

silencieuses dans le génome plastidique. Celles-ci se fixeraient environ six fois moins fréquemment que dans le génome nucléaire des plantes (50). Puisque les mutations silencieuses ne sont généralement pas soumises à la pression sélective, cette étude permet de supposer que l'ADN des plastides subirait moins de dommages que celui du noyau, ralentissant ainsi le processus de divergence de séquences. Cependant, dans certains cas, les mutations silencieuses mènent à l'utilisation d'un codon rare, ayant un impact observable sur l'expression des gènes. Ce genre de mutation est alors soumis à la pression sélective. Bref, bien qu'il est clair que le génome des plastides est très stable en comparaison au génome nucléaire des plantes, les causes de cette manifestation demeurent encore obscures.

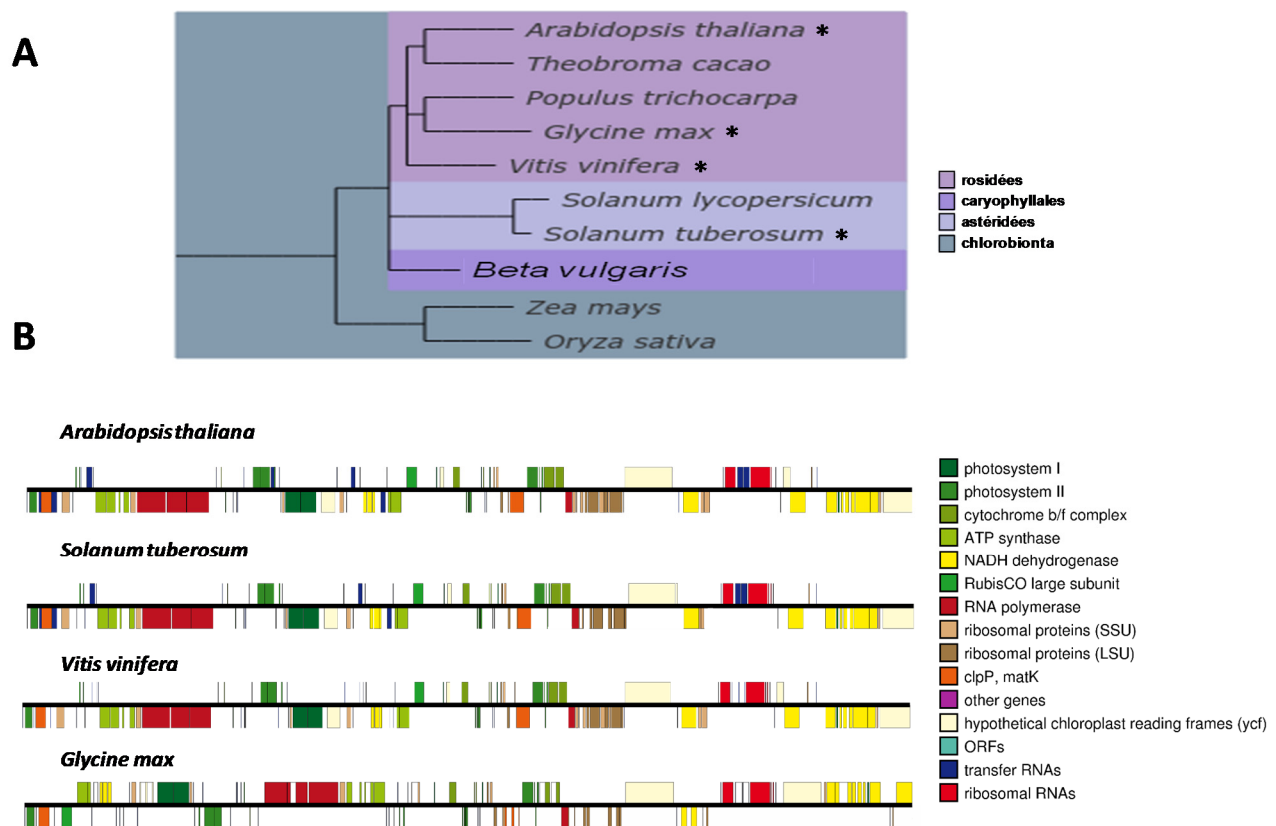


Figure 6: L'arrangement des gènes sur le génome du plastide est fortement conservé entre les espèces. A) Arbre phylogénétique représentant la distance évolutive entre les espèces appartenant au règne végétal disponibles sur la base de données Phylome DB (47). B) Représentation schématique des génomes du plastide, pour quatre espèces indiquées par un astérisque en A, réalisée à l'aide du logiciel OGDraw (13). Les gènes se voient attribuer une couleur selon leur fonction principale. La position du gène, en haut ou en bas, représente le brin sur lequel le gène est encodé.

1.2 Les menaces à l'intégrité du génome plastidique et la réparation de l'ADN

Les menaces pour l'ADN sont nombreuses et variées dans les organelles. Cependant, comme mentionné dans la section précédente, leur génome accumule très peu de mutations au cours du développement de la plante et est fortement conservé au niveau interspécifique. Pour atteindre un tel niveau de conservation, les organelles de plantes possèdent des mécanismes pour préserver la structure de leur génome, mécanismes regroupés sous l'appellation de mécanismes du maintien de la stabilité génomique (revue dans (51)).

Le maintien de la stabilité génomique est un processus complexe qui s'effectue sur plusieurs fronts à la fois. Dans un premier temps, les organismes possèdent des moyens de protéger l'intégrité de leur génome tels la détoxification des molécules mutagènes ou le recouvrement de l'ADN avec des protéines stabilisatrices. Cependant, ces mécanismes ne sont pas infaillibles et à l'occasion, des lésions apparaissent sur l'ADN. Pour atténuer l'effet néfaste de ces lésions, les organismes possèdent une panoplie d'outils permettant leur réparation et, dans une situation idéale, la restauration de la séquence génomique d'origine. La prochaine section focalisera sur l'action des différentes voies de réparation de l'ADN ainsi que leurs conséquences sur la structure génomique.

1.2.1 La réparation par excision de base

Certaines molécules ont la capacité, par leur nature oxydative, de modifier la structure des bases de l'ADN. Parmi ces molécules on retrouve les espèces réactives oxygénées (ROS), dont la production est associée à l'activité de la chaîne de transport des électrons des organelles (52-54). En effet, dans le plastide, ces molécules sont formées lorsque des électrons circulant dans la chaîne de transport ne trouvent pas leurs molécules acceptrices appropriées. Dans ce cas, la chaîne de transport des électrons n'a d'autres choix que de transférer l'électron excité à une molécule d'oxygène avoisinante, menant à l'apparition de différents types de ROS (52, 54). Autrefois considérés comme un simple dommage collatéral, on réalise aujourd'hui que les ROS sont aussi des molécules de signalisation essentielles au bon fonctionnement du chloroplaste et à la coordination de son expression génique avec le noyau (55-57). Ainsi, la vision moderne des ROS est davantage celle d'un couteau à double tranchant qui nécessite d'être particulièrement bien contrôlé.

Les plantes possèdent deux mécanismes principaux afin d'atténuer les effets néfastes des ROS à l'intérieur du chloroplaste, soit la modulation de l'activité de leur chaîne de transport des électrons et la détoxification enzymatique des ROS (58, 59). Néanmoins, ces mécanismes ne sont pas infallibles et les bases du génome plastidique sont parfois oxydées, menant à l'apparition de bases alternatives telles la 8-oxoguanine ou à la thymine glycol, présentées à la figure 7 (revue dans (52)). Ces bases oxydées retrouvées dans la séquence génomique ne pourront être reconnues efficacement par la machinerie de réplication de l'ADN, menant généralement à l'insertion d'une base erronée dans la nouvelle copie génomique (60). Cette modification porte le nom de substitution de base et il s'agit de la modification la plus fréquemment observée dans les organelles (61, 62). En plus de cette modification, les bases oxydées peuvent causer des arrêts de la fourche de réplication et mener à d'autres types de réarrangements génomiques qui seront détaillés plus loin (voir section sur les dommages liés à la réplication).

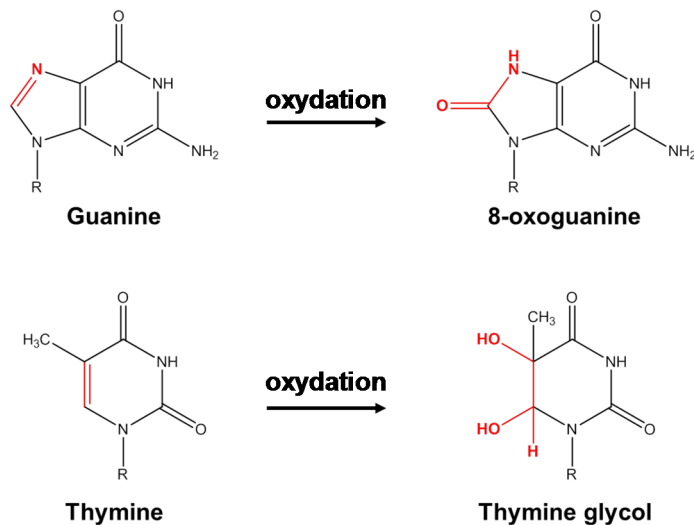


Figure 7: Structure des bases azotées lors de réaction d'oxydation. Les liaisons chimiques qui subissent des modifications lors de l'oxydation sont représentées en rouge.

Bien que loin d'être idéale, la présence de bases oxydées dans l'ADN ne mène pas nécessairement à des mutations. En effet, les plantes possèdent de nombreux mécanismes de réparation de l'ADN qui permettent d'éliminer ces bases ou de limiter leurs effets néfastes. Le mécanisme d'excision de bases (BER) permet d'enlever ces lésions avant qu'elles ne causent un dommage irréversible. Ce mécanisme est dépendant de l'activité de glycosylases d'ADN qui viennent cliver le lien N-glycosidique entre la base oxydée et le désoxyribose (63, 64). Cette étape génère un nucléotide apurinique/aprimidique qui est par la suite clivé par une endonucléase. Le trou créé par cette suppression de nucléotides dans l'ADN double-brin sera comblé à l'aide d'un ou plusieurs nucléotides complémentaires, selon le nombre de nucléotides ayant été clivés. Finalement, une étape de religation permet de retrouver la séquence originale de l'ADN double-brin.

1.2.2 La photoactivation

Les plantes, de par leur nature, sont continuellement exposées à des niveaux particulièrement élevés de radiation solaire. Parmi ces radiations, on retrouve entre autres les rayons UV, dont le potentiel

mutagène est bien établi. Étrangement, les plantes présentent des taux de mutations associés aux rayons UV similaires aux autres espèces (65).

Les rayons UV produisent deux types de lésions principales, les dimères de pyrimidine cyclobutanes (CPD) et les pyrimidine-pyrimidones, qui sont présentées à la figure 8 (revue dans (66)). Ces deux lésions sont causées par le pontage covalent de deux pyrimidines adjacentes. Afin de pallier l'effet néfaste de ces molécules, les plantes possèdent des photolyases, qui sont des enzymes capables de briser les pontages entre les bases (67-69). Ce processus, nommé photo-réactivation, nécessite l'énergie de la lumière bleue (300 à 600 nm) pour activer les enzymes et réduire les dimères de pyrimidine en monomères séparés. Étrangement, bien qu'une activité photolyase ait été confirmée dans les chloroplastes isolés de plusieurs espèces, il n'y a que chez le riz (*Oriza sativa*) que l'enzyme responsable de ce processus a été identifiée (67). En effet, chez cet organisme, une CPD-photolyase a été localisée à la fois dans le chloroplaste, la mitochondrie et le noyau. Il est aussi intéressant de constater la présence de photolyases dans le nucléoïde des plastides du maïs (*Zea mais*), bien qu'aucune caractérisation n'ait encore été réalisée pour ces protéines (15).

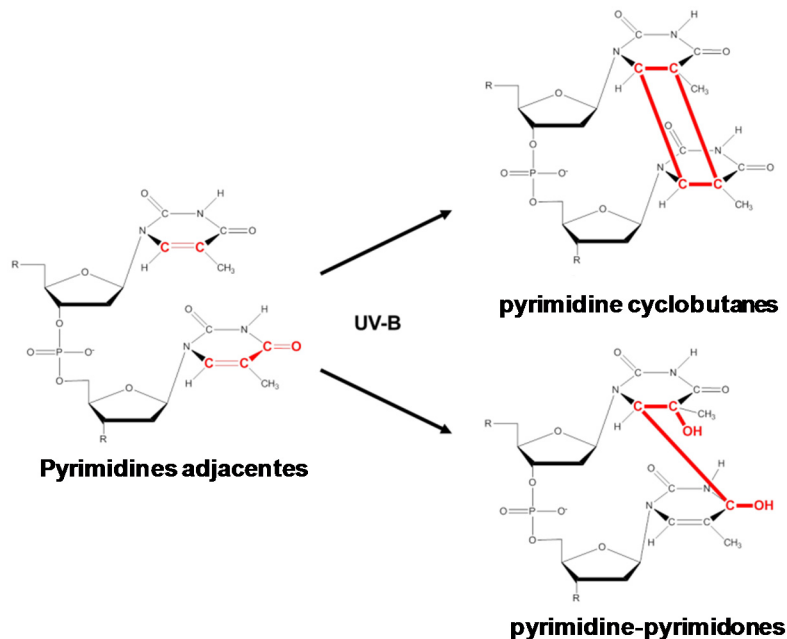


Figure 8: Modification de la structure des pyrimidines adjacentes par l'action des rayons UV. Les liaisons chimiques qui subissent les modifications par l'UV-B sont représentées en rouge ainsi que les nouveaux liens formés.

1.2.3 La réparation conservatrice des bris d'ADN double-brin

Les bris double-brin (DSBs) sont généralement considérés comme étant le type de lésion le plus déstabilisant pour les génomes. En effet, ce type de dommage est à l'origine de l'apparition de nombreux réarrangements génomiques à grande échelle telles les fusions inter-chromosomiques et les délétions de locus (revue dans (70)). Le fort potentiel mutagène des DSBs a probablement exercé une pression évolutive sur les organismes, les poussant à développer des mécanismes de réparation spécifiques à ces lésions afin de protéger leur intégrité génomique. Il est intéressant de constater que ces mécanismes sont hautement conservés entre les espèces, suggérant qu'ils seraient apparus relativement tôt chez le vivant et qu'ils auraient été conservés à cause de leur caractère essentiel (71). Il existe deux types principaux de mécanismes de réparation des DSBs : ceux qui sont dits conservateurs et ceux qui entraînent des modifications de la structure du génome.

1.2.3.1 La recombinaison homologue

Parmi les mécanismes conservateurs, on ne peut passer sous silence l'importance de la recombinaison homologue (HR). Les travaux à l'origine de la compréhension de la HR remontent au début du XX^e siècle et ont été grandement influencés par l'hypothèse de Thomas Hunt Morgan. Ce dernier suggéra que les chromosomes pouvaient former des enjambements et échanger de l'information génétique (72). Plusieurs années plus tard, Robin Holliday observa des jonctions particulières dans l'ADN permettant ce type d'échanges d'information entre les chromosomes, confirmant par le fait même l'hypothèse de son prédécesseur (73, 74). Dans les années qui suivirent, de nombreuses études sont venues clarifier les bases moléculaires de ce mécanisme et ont démontré qu'il pouvait aussi servir de mécanisme de réparation pour les DSBs en utilisant l'information contenue sur le chromosome sœur (75-77). Ce processus fut simplement dénommé réparation de bris d'ADN double-brin (DSBR).

Le mécanisme de réparation de type DSBR s'effectue en plusieurs étapes qui sont détaillées à la figure 9 (revue dans (51, 78)). Dans un premier temps, il y a résection de l'ADN au site de coupure. Cette résection est effectuée par des exonucléases qui dégradent spécifiquement l'ADN selon la polarité 5' à 3', engendrant une extrémité simple-brin (79). L'ADN simple-brin est une structure optimale pour la recherche d'homologie puisqu'elle possède une forte avidité à trouver sa séquence complémentaire. Afin de limiter les appariements indésirables de ce simple-brin nouvellement libéré, il est alors

recouvert de protéines de liaison de l'ADN simple-brin, qui viennent diminuer son avidité et stabiliser cette structure temporairement (80). Ensuite, lorsque la résection a permis de générer un brin d'ADN suffisamment long pour la recherche d'homologie, les protéines de liaison à l'ADN simple-brin sont remplacées par un long filament d'une enzyme particulière nommée recombinaise (80). Les recombinaises ainsi placées sur l'ADN permettent d'envahir un duplex d'ADN au site complémentaire à la séquence recouverte. La réplication de l'ADN permet alors de synthétiser l'information manquante pour chacune des extrémités du DSB à l'aide de l'information sur le chromosome sœur. L'étape finale de ce processus est la résolution. Cette dernière peut se produire dans deux orientations différentes, libérant alors les deux chromosomes sœurs (81). L'information échangée entre les chromosomes dépendra alors de l'orientation de la résolution.

1.2.3.2 L'appariement dépendant de la synthèse d'ADN

Bien que menant dans la plupart des cas à une réparation adéquate, la réparation de type DSBR peut générer des pertes d'informations lorsque les deux chromosomes sœurs ne sont pas en tous points identiques (revue dans (51)). Afin d'éviter cette conséquence néfaste, la cellule peut utiliser un autre mécanisme de réparation, l'appariement dépendant de la synthèse d'ADN (SDSA). Ce mécanisme présente l'avantage de simplement nécessiter la synthèse de quelques nucléotides sur le chromosome sœur, principalement ceux correspondant au site de cassure, pour ensuite permettre de relier les deux extrémités de la molécule brisée (82). Comme présentées à la figure 9, les étapes initiales sont identiques à celles de la réparation par HR. Cependant, l'envahissement ne se fait qu'à l'aide d'un seul brin et aucune résolution n'est nécessaire. Ce mécanisme ne modifie d'ailleurs aucunement le chromosome sœur.

1.2.3.3 La réplication induite par des bris

Un autre mécanisme dont la cellule dispose pour réparer de manière conservatrice les DSBs est la réplication induite par des bris (BIR) (revue dans (51)). Ce mécanisme est particulièrement utile dans le cas de bris d'ADN à une seule extrémité (Figure 9), puisqu'il permet de resynthétiser l'information qui aurait été perdue autrement. Ce type de bris est généralement produit lorsque la fourche de réplication rencontre une région d'ADN simple brin (revue dans (83)). Cette rencontre fait décrocher

un des brins de la fourche de réplication et mène à l'apparition d'un fragment d'ADN incomplet, perçu comme un DSB. Le mécanisme moléculaire du BIR, décrit à la figure 9, est lui aussi en grande partie similaire à la recombinaison homologue (revue dans (84)). Il débute en effet par les mêmes étapes de résection, de recherche de séquences homologues et d'invasion, mais ces étapes se produisent sur une seule extrémité. Une fois l'invasion sur une séquence homologue complétée, il y a établissement de la réplication à cet endroit qui permettra la synthèse complète de la partie manquante suivant le DSB.

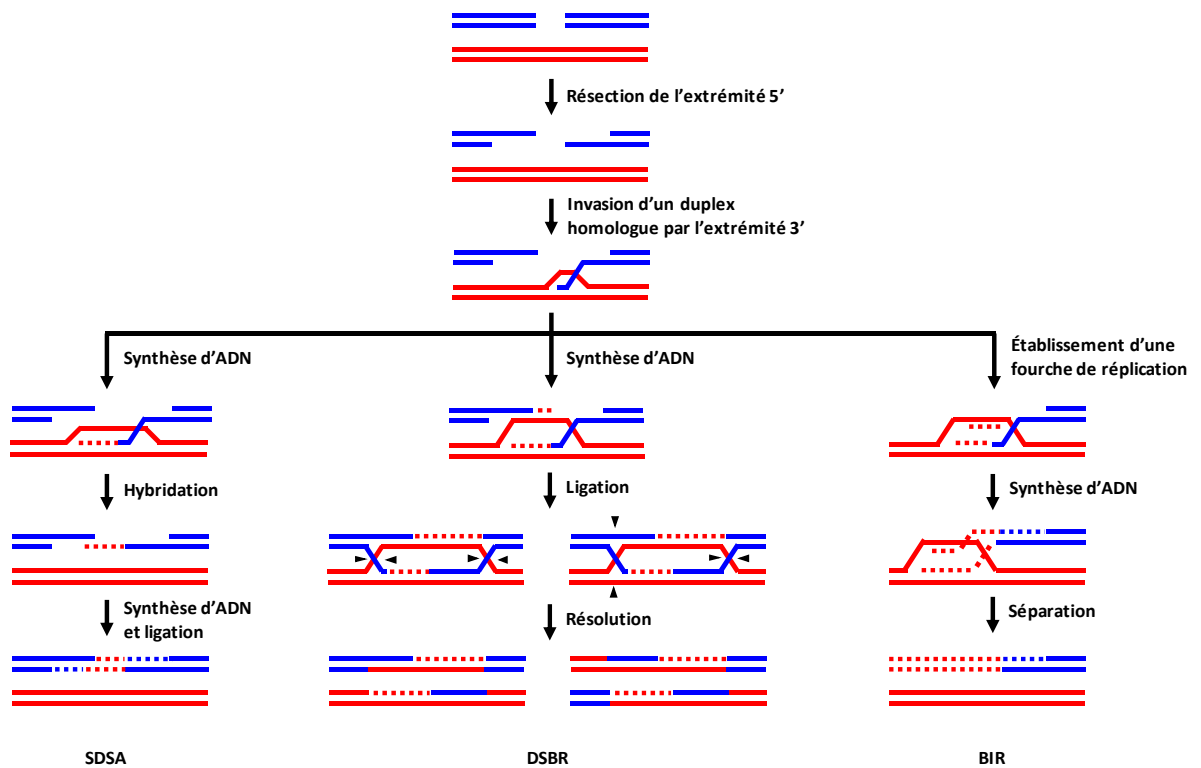


Figure 9: Mécanismes de réparation de l'ADN dépendants de la recombinaison homologue. Les brins bleus correspondent au chromosome ayant subi un bris, tandis que les brins rouges correspondent au chromosome sœur servant de modèle à la réparation. Une ligne pointillée représente un site de synthèse d'ADN. Les triangles noirs correspondent aux sites de résolutions.

1.2.3.4 La recombinaison homologue dans le plastide

La recombinaison homologue est un processus tellement fondamental qu'il n'est pas surprenant de constater qu'il se produise dans le plastide. Les séquences répétées inversées (IRs) sont particulièrement susceptibles à la HR et échangent fréquemment de l'information entre elles (85, 86). La recombinaison a aussi été démontrée pour être impliquée dans la réparation des DSBs dans le plastide. En exprimant des versions dominantes négatives de la recombinase RecA d'*Escherichia coli* dans les plastides de l'algue modèle *Chlamydomonas reinhardtii*, il a été possible d'observer une diminution de la fréquence de HR, de la capacité de réparer l'ADN et des taux de survies lors de traitement avec des agents mutagènes qui induisent des DSBs (87). Le séquençage génomique de plusieurs organismes modèles a par la suite facilité l'isolation de mutants pour les gènes de recombinases. Ces mutants ont permis de confirmer que les recombinases possédaient un rôle essentiel dans la protection du génome du plastide (88, 89). D'ailleurs, la fréquence d'utilisation de la recombinaison homologue dans le plastide corrèle avec les niveaux de DSBs, ce qui supporte encore une fois l'importance de ce mécanisme pour réparer les DSBs (90).

Avec le développement des technologies de transformation du chloroplaste, vers la fin des années 80, il devint alors possible d'évaluer les mécanismes menant à l'intégration de séquences d'ADN étrangères dans le génome plastidique. Cette intégration est dépendante des mécanismes de HR et fortement inhibée en l'absence de recombinases (92). En étudiant les mécanismes d'insertion de l'ADN étranger dans le génome plastidique, des observations ont permis de supporter que SDSA serait aussi actif dans ce compartiment (91). Le BIR, pour sa part, n'a toujours pas été clairement observé dans le plastide, malgré qu'il ait été confirmé dans les mitochondries de plantes (92). Finalement, la transformation génomique des plastides est aussi venue supporter l'hypothèse qu'il y aurait une forte pression sélective exercée sur ce génome hautement optimisé. En effet, malgré de nombreuses tentatives, toutes les modifications des gènes de la photosynthèse ont mené à une réduction de l'efficacité de ce processus et à une diminution de la compétitivité de la plante.

1.2.4 Les mécanismes sujets à l'erreur

Bien que la majorité des DSBs soit réparée de manière conservatrice, il n'en demeure pas moins que dans des cas exceptionnels, les bris d'ADN mènent à des réarrangements génomiques. Les

réarrangements génomiques ainsi formés peuvent alors avoir des effets variables selon les gènes qui ont subi une modification. La caractérisation de l'impact de ces mutations a permis d'identifier les causes moléculaires de certaines maladies chez l'humain, dont plusieurs types de cancers et de nombreuses maladies héréditaires (revue dans (78, 83)). Chez les plantes, des réarrangements génomiques ont été associés à des traits d'intérêts agricoles qui ont été sélectionnés à travers le temps (93-95). Plus spécifiquement, ce type de modification dans le génome des organelles a été associé aux phénotypes de variéation, de stérilité mâle cytoplasmique et à la tolérance à certains stress (93-164-165). Bref, une meilleure compréhension des mécanismes qui causent des modifications génomiques dans les organelles pourrait faciliter l'obtention de lignées qui possèdent ces phénotypes recherchés.

Bien que les conséquences des réarrangements génomiques aient été étudiées depuis de nombreuses années, les mécanismes qui mènent à leur formation ne sont que partiellement élucidés. Un des facteurs qui est conservé entre les mécanismes de réparation de l'ADN non-conservateur est l'absence d'étape de recherche de longue séquence d'homologie. Étant donné que cette recherche d'homologie assure la fidélité de la réparation, les mécanismes indépendants de ce processus sont reconnus pour être moins fidèles et capables d'entraîner des modifications importantes du génome. Il existe toutefois une exception à cette règle, soit la recombinaison homologue non-allélique. Ce processus nécessite deux régions fortement homologues, constituées minimalement d'une centaine de paires de bases, pour se produire (96). Ces deux régions génomiques se retrouvant à deux endroits distincts dans sur le génome peuvent ainsi mener à l'utilisation du mauvais site d'homologie sur le chromosome sœur lors de la réparation par HR, créant par le fait même un réarrangement génomique.

1.2.4.1 Les mécanismes de liaison d'extrémités

La liaison d'extrémités non-homologue (NHEJ) est un processus qui permet de lier deux extrémités d'ADN double-brin sans recherche d'homologie (Figure 10). Ce mécanisme est dépendant d'un grand nombre de facteurs, dont les protéines de liaison à l'ADN KUs qui se fixent aux extrémités franches des DSBs (97). Une fois ce complexe formé, il y a recrutement de protéines effectrices aux deux extrémités du DSB, permettant la religation et l'obtention d'une molécule continue (98, 99). Les premières études de ce mécanisme ont proposé que le choix des extrémités à liguer était aléatoire,

mais de nouveaux résultats expérimentaux suggèrent plutôt que ce choix est fortement biaisé en faveur des deux extrémités d'un même fragment original (100). Bien qu'encore au stade d'hypothèse, la raison proposée pour ce biais implique la fixation par la protéine KU80 des deux extrémités libérées lors d'un DSB, ce qui permettrait de maintenir les fragments à proximité. Cette conformation favoriserait grandement la ligation de ces deux fragments, provenant d'un même DSB, comparativement aux autres fragments d'ADN pouvant être retrouvés dans les environs (100). Dans des conditions de stress génotoxique, le NHEJ mène toutefois à l'apparition plus fréquente de réarrangements génomiques causés par l'association de deux fragments non-continus (101). Dans ces conditions, l'abondance plus élevée d'extrémités d'ADN libres augmente le nombre de molécules acceptrices potentielles, favorisant ainsi une réparation illégitime. Bref, contrairement à ce qui était proposé initialement, le NHEJ est un mécanisme majoritairement conservateur qui peut mener à l'apparition de réarrangements génomiques, particulièrement en situation de stress.

1.2.4.2 Les mécanismes dépendants de séquences microhomologues (MHMR)

Avec l'avènement des technologies d'amplification et de séquençage de l'ADN, il devint beaucoup plus facile de caractériser les différents types de réarrangements génomiques et d'obtenir leur séquence précise. Une caractéristique commune retrouvée chez plusieurs réarrangements génomiques est la présence de courtes séquences d'homologie au site de jonctions entre les deux fragments illégitimement liés (102-104). Afin de mieux comprendre les causes de l'apparition de ces molécules réarrangées particulières, plusieurs mécanismes de réparation de l'ADN dépendants de courtes séquences d'homologie ont été proposés. Le premier à avoir été clairement énoncé est la liaison d'extrémités dépendante de microhomologies (MMEJ) (105). Ce mécanisme est semblable au NHEJ et se conclut par la ligation de deux fragments d'ADN issus d'un DSB. Ce mécanisme est dépendant de MRE11, une nucléase qui se fixe rapidement au site de cassure de l'ADN et qui permet la dégradation du brin d'ADN de 5' à 3' (105). La dégradation de l'ADN par cette nucléase est inhibée lorsqu'une courte séquence d'homologie est présente entre les deux fragments d'ADN associés à cette protéine (106). Cet arrêt favorise alors la ligation des extrémités à ce site de microhomologie. Une hypothèse suggère d'ailleurs que la protéine MRE11 soit impliquée dans le maintien des deux extrémités libérées au moment de la cassure double-brin, à l'image de KU80 lors du NHEJ (106). Les protéines KUs n'étant

généralement pas impliquées dans l'établissement du MMEJ, cette hypothèse suggère aussi que MRE11 pourrait jouer un rôle similaire aux KU en se fixant aux extrémités et permettant de recruter des protéines effectrices (105).

La longueur de la séquence d'homologie nécessaire à l'établissement de la réparation est une différence importante entre le NHEJ et le MMEJ. On propose présentement que des séquences d'homologie de 5 à 25 nucléotides soient essentielles au MMEJ (revue dans (107)). Il est aussi intéressant de constater que suite à la découverte du MMEJ, des études ont permis de déterminer que le NHEJ était lui aussi favorisé par la présence d'appariements de bases entre les fragments à joindre, ces appariements pouvant être aussi courts qu'un seul nucléotide (généralement de 1 à 4) ((108). Finalement, il existe un dernier mécanisme de liaison d'extrémités, celui-ci dépendant de séquences d'homologie plus longues. En effet, l'hybridation de simple-brin (SSA) permet de joindre deux extrémités d'ADN si une séquence d'homologie de plus de 30 pb est présente à chacune des extrémités (109). Cependant, il n'est pas encore tout à fait clair s'il existe de réelles différences mécanistiques entre le MMEJ et le SSA, autre que la longueur d'homologie utilisée. La comparaison des mécanismes de liaison d'extrémités est présentée à la figure 10.

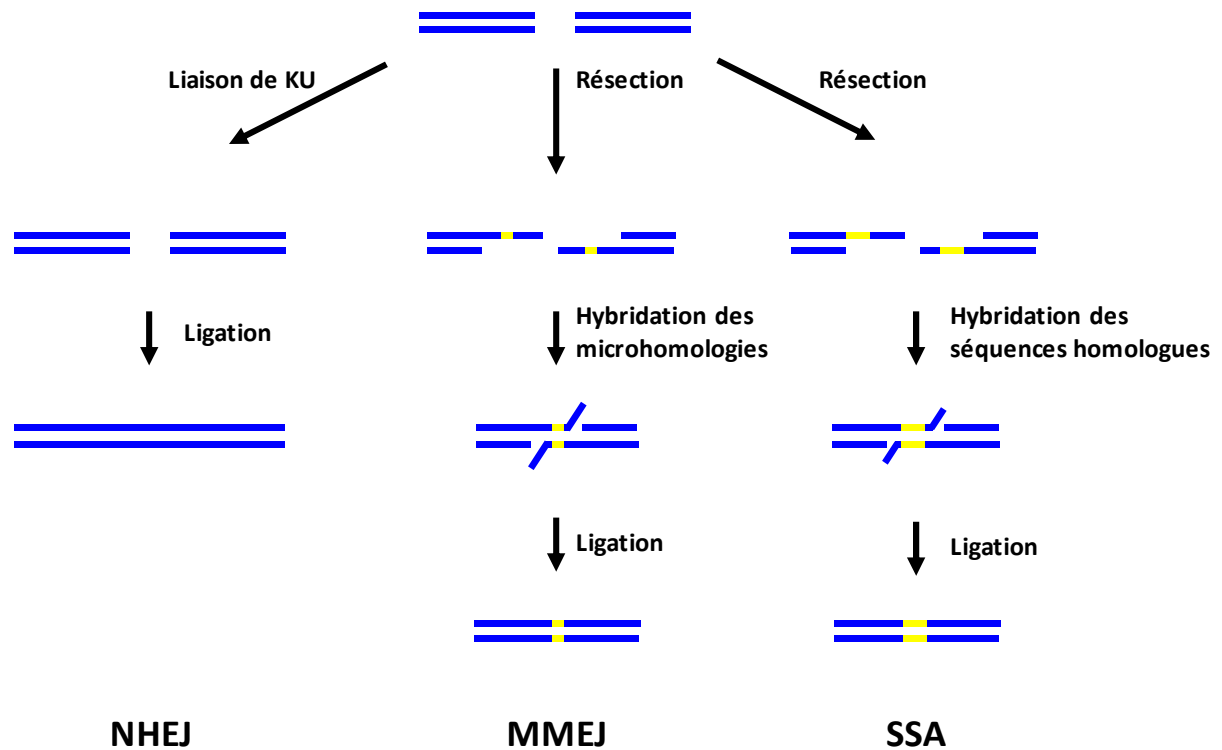


Figure 10: Mécanismes de réparation de l'ADN par jonction d'extrémités. Les brins bleus représentent un chromosome ayant subi une cassure. Les rectangles jaunes représentent des séquences d'homologie et leur longueur est relative à la longueur de microhomologies utilisée par le mécanisme (5 à 25 pb pour le MMEJ, 30 et + pour SSA).

D'autres types de réparation de DSBs dépendants de séquences microhomologues ont aussi été décrits dans la littérature. En effet, à l'instar du MMEJ qui est similaire au NHEJ, d'autres mécanismes sont similaires aux mécanismes dépendants de la HR. La réplication induite par des bris dépendante de microhomologies (MMBIR) et l'appariement dépendant de la synthèse d'ADN à des sites microhomologues (MMSDSA) sont deux de ces mécanismes (84, 110). La principale différence entre ces mécanismes et ceux dépendants de la recombinaison homologue est l'utilisation de courtes séquences d'homologie (généralement 5 à 35 bases) et l'autonomie face à l'activité recombinase. Il a même été démontré que ces mécanismes pouvaient être stimulés en l'absence d'une recombinase fonctionnelle (111). D'ailleurs, ce type de réparation semble nécessiter très peu de facteurs. Le seul ayant été identifié à ce jour étant la sous-unité de la polymérase delta, Pol32 (112). Finalement, la suppression des protéines de liaisons à l'ADN simple-brin, qui recouvrent les extrémités des bris d'ADN, stimule aussi fortement ces mécanismes (111).

Il n'est pas encore tout à fait clair si la réparation de l'ADN dépendante de courtes séquences d'homologies est véritablement un mécanisme de réparation à part entière ou plutôt une aberration causée par des erreurs lors de la réparation conservatrice. Du moins, ces processus sont plus fortement observés dans des conditions de stress génotoxiques et permettent une réparation des lésions dans l'ADN s'effectuant à l'aide de très peu de facteurs. Dans une situation où les DSBs deviennent très abondants et que la machinerie de réparation est débordée, la réparation dépendante de microhomologies peut devenir un avantage important pour empêcher que la situation devienne hors de contrôle (113). Il ne faut pas oublier que ces mécanismes peuvent aussi mener à une réparation conservatrice lorsqu'ils se produisent en utilisant l'information appropriée sur le chromosome sœur.

1.2.3.5 Les endonucléases de homing

Certaines enzymes possèdent la capacité de générer des DSBs en clivant l'ADN à des sites précis (revue dans (114)). Ces enzymes, nommées endonucléases, ont été grandement utilisées pour étudier les mécanismes de réparation de l'ADN. Les endonucléases de homing, impliquées dans la migration des introns, ont été principalement utilisées à cette fin dans les plastides. Ces endonucléases sont encodées par des introns qui, étrangement, contiennent un site de reconnaissance permettant à l'enzyme de digérer à proximité de sa séquence codante (115). Cette coupure augmente ainsi les chances que l'intron puisse recombiner et se retrouver intégré ailleurs dans le génome. Ce processus n'est pas sans rappeler celui de la transposition qui déplace un transposon dans le génome de l'hôte. Parmi les enzymes de homing, I-CREI, encodée par le génome du plastide de *Chlamydomonas reinhardtii*, est la mieux caractérisée (116-118). L'effet des coupures d'ADN double-brins par cette enzyme a été grandement étudiée dans cet organisme et a permis de mieux comprendre les phénomènes de recombinaison et de réparation de l'ADN du plastide (119).

Récemment, un système d'expression conditionnel de l'enzyme I-CREII a été développé pour le plastide d'*Arabidopsis thaliana* (120). L'enzyme I-CREII est similaire à I-CREI, mais possède un site de reconnaissance différent, qui est retrouvé à un site unique dans le génome du plastide d'*Arabidopsis*. Lors d'un traitement à l'œstradiol, le système de recombinaison CRE/LOX permet de réorganiser la séquence génétique introduite de manière artificielle chez *Arabidopsis* (Figure 11). Cette réorganisation permet l'expression de l'endonucléase, qui est alors dirigée aux plastides grâce à la séquence du peptide transit de *RBCS*. Une fois que l'enzyme est parvenue au plastide, on remarque l'apparition de DSBs au site de restriction unique retrouvé dans le gène *PSBA*. Ce système a permis d'évaluer l'impact des DSBs induits artificiellement et de déterminer que ce type de lésion est principalement réparé grâce aux mécanismes de MMEJ.

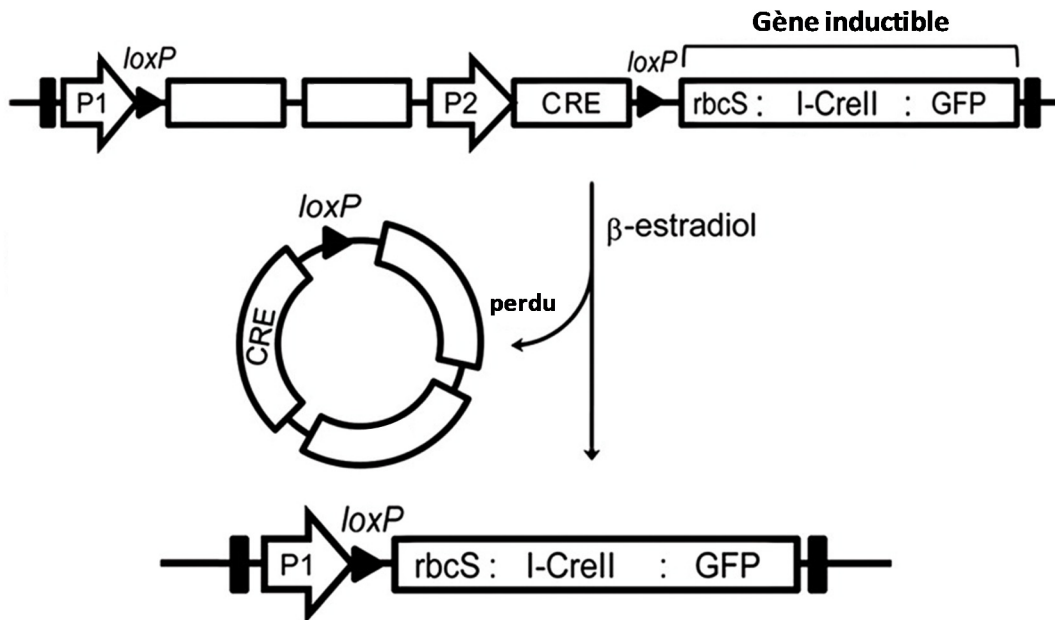


Figure 11: Un système conditionnel permettant d'exprimer une endonucléase dans les plastides d'Arabidopsis. Le traitement à la β -œstradiol active le promoteur P2 qui induit l'expression de la recombinase CRE. Une recombinaison entre les sites loxP est ensuite favorisée par cette enzyme. Cette recombinaison mène à un réarrangement génomique qui rapproche le promoteur P1 du gène chimérique composé du peptide transit de *RBCS*, de la séquence codante de I-CREII et de la séquence codante de la GFP, permettant l'expression de celui-ci.

1.2.3.6 Instabilité génomique induite par la réplication

La réplication est un processus qui nécessite de nombreuses modifications à la structure de la double-hélice d'ADN. Bien que ce processus soit essentiel, les fourches de réplication sont relativement fragiles et peuvent bloquer ou s'effondrer lorsque les conditions ne sont pas optimales (revue dans (121)). Ces deux événements peuvent avoir des conséquences différentes. Dans l'éventualité de l'effondrement d'une fourche de réplication, il y a généralement formation d'un bris d'ADN à une seule extrémité (122, 123). La réplication est d'ailleurs l'une des principales sources de DSBs à l'intérieur d'une cellule dans des conditions de croissance normales (124). Ces bris d'ADN particuliers sont généralement pris en charge par la machinerie de recombinaison homologue et réparés par le BIR. Ils peuvent aussi mener à des réarrangements génomiques, lorsqu'ils sont réparés par des mécanismes de réparation non-conservateurs, tels le NHEJ et le MMBIR.

L'instabilité génomique causée par la réplication ne se limite pas uniquement à la formation de DSBs. En effet, les fourches de réplication bloquées de manière prolongée peuvent aussi être néfastes pour le génome. Plusieurs types de lésions peuvent bloquer la réplication, mais les plus communes sont les bases modifiées (125, 126). Lorsqu'une fourche demeure bloquée à un site de lésion, la polymérase de la machinerie de réplication peut être remplacée par une polymérase d'ADN de type translésion (127). Ce type de polymérase, bien que possédant une fidélité beaucoup plus faible, est capable de faire abstraction des lésions sur l'ADN et poursuivre la réplication dans ces conditions particulières. De ce fait on remarque généralement quelques modifications de bases dans les environs du site d'arrêt de la fourche de réplication, dues à des erreurs faites par ces polymérases moins fidèles (128-130).

Les fourches de réplication bloquées peuvent aussi être redémarrées à l'aide du mécanisme de régression de fourche, qui est présenté à la figure 12 (131-133). Ce mécanisme nécessite que la lésion empêchant la réplication de l'ADN se trouve uniquement sur un seul des deux brins. De cette façon, le brin naissant non bloqué peut être répliqué en aval de la lésion et servir de matrice pour redémarrer la réplication du brin naissant bloqué (Figure 12). L'appariement entre les deux brins naissants nécessite l'activité d'une recombinaise, ce qui assure la fidélité de l'appariement préalable à l'initiation de la réplication. Une fois la réplication du brin naissant terminée, l'extrémité 3' de celui-ci se trouve en aval de la lésion et la réplication normale de l'ADN peut reprendre.

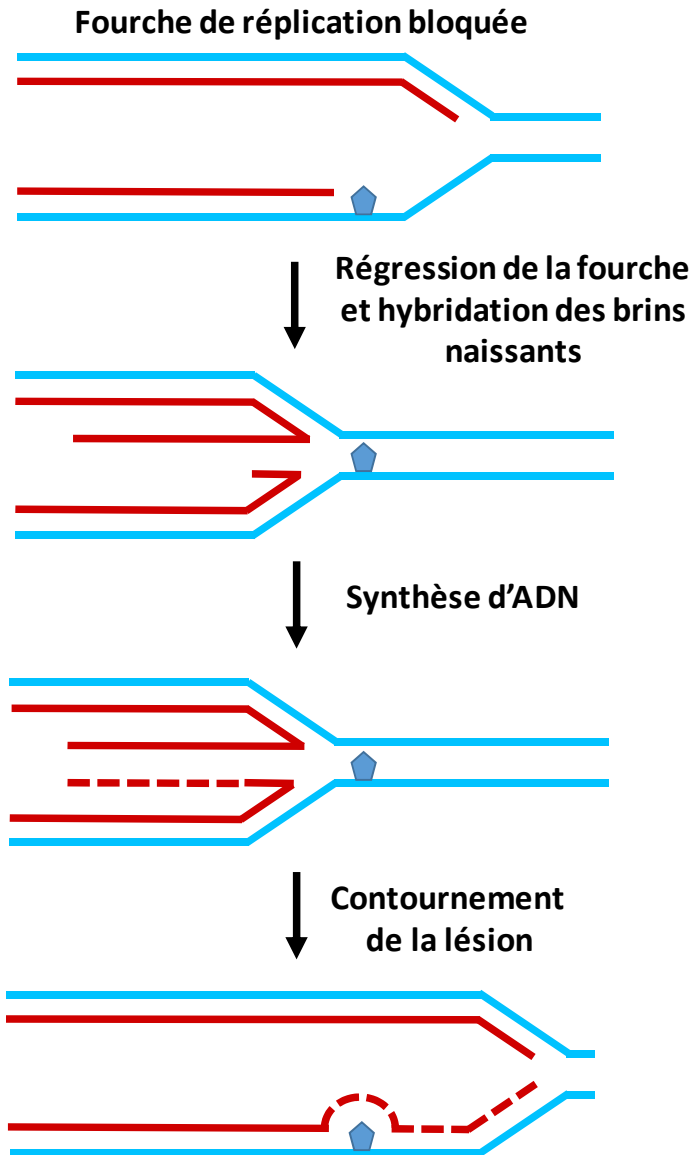


Figure 12 : Mécanisme de régression de fourche permettant de redémarrer une fourche bloquée. Les brins matrices sont représentés en bleu et les brins naissants en rouge. Les traits pointillés correspondent aux sites de synthèse de l'ADN suite au blocage de la fourche. La lésion bloquant la réplication est représentée par un pentagone bleu.

Finalement, la structure créée par une fourche de réplication bloquée est connue pour être hautement mutagène (revue dans (134)). En effet, les fourches de réplifications bloquées sont des substrats idéals pour plusieurs mécanismes menant à des réarrangements génomiques, telles la formation de chromosomes palindromiques et la formation d'inversions de séquences dans le génome

(135-139). Il a été proposé que ces mécanismes génèrent une inversion de séquence en envahissant le brin opposé de la fourche de réplication, pour redémarrer la fourche bloquée (137) . Bref, la réplication, bien qu'étant un mécanisme essentiel, est une des sources principales d'instabilité génomique.

1.2.3.7 La transcription

La transcription peut elle aussi venir déstabiliser le génome (revue dans (140)). Les collisions observées entre la fourche de transcription et la fourche de réplication sont une des causes majeures de l'instabilité génomique associée à ce processus (revue dans (141)). En effet, puisque la réplication et la transcription se produisent souvent en simultanément, il n'est pas surprenant de constater que l'ADN polymérase et l'ARN polymérase puissent parfois entrer en collision. Lors de ces contacts, il peut y avoir formation de DSBs qui seront réparés à l'aide des mécanismes décrits précédemment. Chez *S. cerevisiae*, des résultats expérimentaux ont permis d'associer les régions hautement transcrites, tels les gènes ribosomiques, à des régions particulièrement sujettes aux réarrangements génomiques (142). De plus, des études ont établi que la progression de l'ARN polymérase s'effectue dans le même sens que celui de la fourche de réplication pour la vaste majorité des gènes hautement transcrits (143). Cette co-directionnalité diminue ainsi les risques de collision et présente donc un avantage pour l'organisme en diminuant l'instabilité génomique de ces régions. Chez les procaryotes, l'inversion de l'orientation de ces gènes fortement transcrits diminue d'ailleurs grandement la capacité de répliquer l'ADN et augmente les niveaux de réarrangements génomiques (144, 145). Néanmoins, les collisions ne mènent pas automatiquement à des DSBs. Il est aussi possible d'observer un blocage prolongé de la fourche de réplication au site de collision, qui peut aussi mener à des réarrangements génomiques, comme mentionné dans la section précédente.

Une deuxième source d'instabilité liée à la transcription est la formation d'hybrides ADN:ARN nommés boucles-R (revue dans (140)). Ces structures sont produites lorsque le brin d'ARN naissant se réapparie avec son ADN complémentaire (146, 147). Cette hybridation peut se produire sur plusieurs centaines de bases, ce qui maintient ainsi une longue séquence d'ADN sous forme simple-brin. L'ADN simple-brin est connu pour être plus sensible aux agents mutagènes, tels le ROS et l'UV, et pour être plus recombinogénique que l'ADN double-brin (148-150). Bref, les boucles-R viennent déstabiliser le génome et ont été associées à deux types de modification de l'ADN, soit la mutagenèse associée à la

transcription (151), qui mène à des modifications de bases, et la recombinaison associée à la transcription (152), pouvant mener à des réarrangements génomiques par le biais de la recombinaison homologue non-allélique.

1.3 Les facteurs génétiques de la réparation de l'ADN des organelles chez *Arabidopsis thaliana*

Nous avons décrit, dans la section précédente, différents mécanismes de réparation qui se déroulent dans les organelles de plantes. Néanmoins, les acteurs qui participent à cette réparation, qui sont d'ailleurs important de connaître dans le cadre de cette thèse, n'ont pas encore été évoqués. Cette section portera donc sur les gènes impliqués dans la réparation de l'ADN du plastide ayant été identifiés chez la plante modèle *Arabidopsis thaliana*.

1.3.1 *Arabidopsis thaliana*, le modèle pour la génétique chez les plantes

De nombreux organismes modèles sont utilisés en biologie végétale, possédant tous des avantages qui leurs sont propres. Pour les études génétiques, *Arabidopsis thaliana* a su s'établir comme la référence incontestée. Cette plante possède de nombreux avantages pour les criblages à haut débit, tels une petite taille, un cycle de croissance court et une descendance abondante. De plus, son génome a été complètement séquencé en 2000, facilitant ainsi grandement les études génétiques (153). Finalement, la disponibilité de collections de mutants générées par mutagenèse aléatoire ou par insertion d'un T-DNA permet d'obtenir des mutants pour un gène d'intérêt très rapidement.

Les collections générées par mutagenèse aléatoire utilisent généralement l'éthyle-méthane-sulfonate pour induire de mutations ponctuelles à une fréquence désirée dans le génome (revue dans (154)). Par la suite, une série de croisement permet d'isoler et de cartographier la mutation qui cause un phénotype d'intérêt. Cette approche n'est pas unique à *Arabidopsis* et est utilisée chez plusieurs organismes, mais elle possède aussi l'inconvénient d'être laborieuse. Les collections de mutants d'insertions, générées en insérant d'un T-DNA aléatoirement dans le génome, sont quant à elles plus spécifiques aux plantes. En effet, cette approche tire avantage de la capacité de la bactérie

Agrobacterium tumefaciens d'introduire une courte séquence d'ADN, nommée T-DNA, dans le génome nucléaire des plantes (155). Au niveau des mutants d'insertion de T-DNA, *Arabidopsis* se distingue par l'abondance de lignées disponibles. Il existe plus de 100,000 lignées individuelles disponibles dont le site précis d'insertion du T-DNA a été séquencé. Deux banques principales ont été générées, celle du Salk institute (156), qui contient plus de 88,000 lignées et celle produite par la compagnie Syngenta, qui en contient plus de 54,000 (157). Avec toutes ces lignées disponibles à la communauté scientifique, il est possible de commander un mutant d'insertion pour pratiquement n'importe quel gène d'*Arabidopsis*, ce qui est un avantage considérable dans la caractérisation d'un gène d'intérêt.

1.3.2 La famille de la recombinase RecA

Comme mentionnée dans les chapitres précédents, la recombinaison est un mécanisme essentiel au maintien de la stabilité génomique du plastide. Des recombinases homologues à celles de la famille RecA des procaryotes ont été identifiées à la fois dans le chloroplaste et la mitochondrie des plantes terrestres. Chez *Arabidopsis*, trois membres de cette famille ont été identifiés; RECA1 dirigée au chloroplaste, RECA2 dirigée à la fois vers le chloroplaste et la mitochondrie et RECA3 dirigée vers la mitochondrie (88, 92). La protéine RECA1 est impliquée dans le maintien de la stabilité génomique du plastide (89, 158). Une analyse de l'ultra-structure de l'ADN plastidique des mutants *reca1* révèle qu'une importante proportion des molécules génomiques retrouvées dans ce mutant sont de faibles poids moléculaires et donc vraisemblablement incomplètes. De plus, *reca1* est hypersensible au traitement à l'aide d'un agent qui induit des DSBs dans les organelles. Ce même traitement chez des plantes de type sauvage occasionne une augmentation de l'abondance des molécules incomplètes, similaires à celles observées chez *reca1*.

La délétion de *RECA2* a de graves conséquences chez *Arabidopsis* et les plantes mutantes sont incapables d'atteindre leur maturité, confirmant par le fait même l'importance de la recombinaison pour le génome des organelles (88). Les plantes mutantes *reca2* présentent de nombreux réarrangements génomiques dans leur ADN mitochondrial. Ces réarrangements seraient le résultat de recombinaison non-allélique se produisant entre des séquences répétées de 150 à 700 nucléotides. Les auteurs de cette étude attribuent les problèmes de développement observés chez *reca2* à ces molécules réarrangées dans la mitochondrie. Il est toutefois important de mentionner qu'aucune

analyse du génome plastidique n'a été réalisée et que des réarrangements similaires pourraient s'y retrouver. Il a aussi été démontré que la mutation de *RECA3* déstabilisait le génome mitochondrial, mais aucune étude de localisation ne suggère que cette protéine puisse se retrouver dans le plastide (88, 92).

1.3.4 Le rôle de MSH1 dans la protection de l'ADN des organelles

L'histoire de la caractérisation du gène *MSH1* remonte aux années soixante et débute par l'identification d'une plante à l'apparence bien particulière (159). Les feuilles de cette plante se caractérisent par l'apparition aléatoire de secteurs complètement blancs, phénotype qui a été dénommé variégation. Puisque la couleur verte des feuilles est associée à la chlorophylle des chloroplastes, ce phénotype a d'abord suggéré que la mutation chez ce mutant affectait le chloroplaste. De plus, quelques années plus tard, l'héritabilité du phénotype a été confirmée pour être uniquement maternelle, suggérant une hérédité liée au génome des organelles (160). Toutes ces évidences ont alors mené les chercheurs à nommer ce mutant chloroplast mutator 1 (*chm1*).

Bien des années plus tard, la mutation chez *chm1* fut identifiée dans le gène *MSH1*, *mutS* protein homologue 1 (161). La caractérisation de la protéine MSH1 a révélé qu'elle possède un peptide transit particulier qui lui permet d'être localisée à la fois dans le chloroplaste et la mitochondrie. Puisque le génome mitochondrial est lui aussi hérité de manière uniquement maternelle, cette découverte venait poser un doute quant à l'origine du phénotype de variégation et révélait la nécessité de vérifier l'importance relative des mutations plastidiques et mitochondriales chez ce mutant. Étonnamment, différentes études ont permis de découvrir des mutations dans le génome de la mitochondrie sans pour autant en voir dans celui du chloroplaste (94, 162-164). Bien que MSH1 soit homologue à une protéine impliquée dans la réparation de mésappariement, son rôle dans la mitochondrie semble plutôt consister à limiter l'apparition de la recombinaison homologue non-allélique. En effet, le mutant *msh1* possède des réarrangements génomiques causés par une recombinaison non-allélique entre des séquences ayant entre 200 et 1000 pb de similarité (92).

Quelques années plus tard, la découverte que des mécanismes de réparation de l'ADN dépendants de microhomologies (MHMR) étaient actifs dans le plastide des plantes a ouvert une autre piste

d'explication au phénotype de variégation observé chez *msh1* (165). En effet, des réarrangements d'ADN utilisant ce type de courtes homologies furent observés dans le plastide de *msh1* (166). De plus, des lignées transgéniques exprimant MSH1 avec un transit peptide dirigeant les protéines uniquement au plastide ou à la mitochondrie ont permis de déterminer l'origine du phénotype. En effet, seule la construction exprimant une version MSH1 dirigée vers les plastides restaure le phénotype de type sauvage chez *msh1*, tout en empêchant l'apparition de réarrangements génomiques dans le plastide. De son côté, la construction exprimant une version de MSH1 dirigée vers la mitochondrie empêche l'apparition de réarrangements génomiques mitochondriaux sans restaurer le phénotype (166).

1.3.5 La famille de protéines de liaison à l'ADN simple-brin Whirly

Les Whirly forment une famille de protéines de liaison à l'ADN simple-brin spécifique aux plantes. L'identification des Whirly dans les organelles remonte aux premières études de purification du nucléoïde du chloroplaste chez *Arabidopsis* (167). Dans ce complexe, il a été possible de retrouver deux protéines de liaisons à l'ADN simple-brin, qui ont par la suite été identifiées pour être le produit du gène *WHY1* et du gène *WHY3*. L'étude phylogénique de la famille Whirly a permis de constater que ces protéines sont retrouvées chez toutes les plantes terrestres et qu'il y a minimalement une protéine Whirly dirigée vers le plastide et une vers la mitochondrie (168). *Arabidopsis* est un cas particulier avec deux Whirly localisées au plastide (*WHY1* et *WHY3*) et une à la mitochondrie (*WHY2*).

Le rôle des Whirly dans les organelles est longuement demeuré mystérieux. L'obtention de doubles mutants complètement déficients en protéines Whirly dans le chloroplaste a apporté une réponse à cette énigme (165). En effet, 5% de la population du mutant *why1why3* présente un phénotype de variégation (partie verte et blanche sur une même feuille) associé à l'apparition de chloroplastes atrophiés dans les secteurs blancs. Il est intéressant de mentionner que ce phénotype est héréditaire maternellement, comme celui de *msh1*. Grâce à des expériences de cartographie du génome des organelles des plantes variégées, il a été possible d'identifier l'apparition de concatémères tête-à-queue de certain locus du génome plastidique (165). Ces régions étaient amplifiées plusieurs dizaines de fois, rendant la stœchiométrie entre les gènes complètement déséquilibrée et menant probablement au dysfonctionnement complet du chloroplaste. Le locus d'ADN amplifié est variable d'une plante variégée à l'autre, cependant une caractéristique commune est la présence de courtes

séquences d'homologie entre les unités du concatémère. Cette observation suggère que les réarrangements génomiques observés dans ces plantes proviennent de mécanismes dépendants de courtes séquences d'homologies. Ces résultats suggèrent que le rôle des protéines Whirly dans les organelles de plantes serait de protéger les séquences d'ADN simple-brin pour limiter les événements de réparation dépendants de séquences microhomologues et ainsi favoriser la réparation conservatrice. Quelques années plus tard, une étude similaire a démontré que la réplication protéine A (RPA), une protéine liant l'ADN simple-brin nucléaire, était essentielle pour limiter l'apparition de réarrangements dépendants de courtes séquences d'homologie dans le noyau (111). Bref, il semble que les protéines de liaison à l'ADN simple-brin soient particulièrement importantes dans ce processus.

Finalement, une étude structurale des protéines Whirly en interaction avec l'ADN simple-brin a permis de déterminer que les Whirly fixaient l'ADN de manière non-spécifique par des interactions de type empilement de base (169). De plus, dans ce complexe, les bases de l'ADN se retrouvent orientées vers la protéine et non vers le solvant, ce qui limite grandement la possibilité de recherche de séquence d'homologie et irait parfaitement avec un modèle selon lequel les Whirly protégeraient l'ADN simple-brin pour éviter l'apparition de réarrangements génomiques.

1.4 Objectifs et hypothèse

Malgré l'importance du génome plastidique pour la vie sur la terre, les mécanismes permettant le maintien de la stabilité de ce génome demeurent encore peu caractérisés. La forte conservation du génome plastidique permet de formuler l'hypothèse que plusieurs mécanismes de réparation et de maintien de la stabilité génomique se déroulent dans le plastide, mécanismes qui demeurent toutefois peu caractérisés. De ce fait, l'objectif de ma thèse consiste à tracer le portrait global de ces mécanismes de réparation et de maintien de l'ADN dans le plastide, ainsi que d'identifier les acteurs principaux impliqués dans ce processus. Afin de comprendre les bases moléculaires du maintien de la stabilité génomique du plastide, nous nous sommes tournés vers la génétique, qui permet, au moyen de criblages, d'identifier les acteurs qui régissent les processus cellulaires et moléculaires. Cependant, une telle approche génétique est souvent laborieuse et nécessite la mise au point de conditions expérimentales qui permettent d'évaluer de larges collections de mutants en un temps relativement

court. Donc, en plus de mon objectif principal, décrit plus haut, je me suis appliqué à développer trois objectifs secondaires, de natures plus techniques, afin de mener à bien mon projet. Ces trois objectifs sont les suivants: Premièrement, définir des conditions de croissance qui favorisent spécifiquement l'instabilité génomique dans le plastide; deuxièmement, établir une collection de mutants pouvant faire l'objet d'un crible génétique; et troisièmement, développer des approches permettant d'évaluer le rôle spécifique de certaines protéines dans le maintien de la stabilité génomique.

Chapitre 2 : Résultats préliminaires.

Un criblage de génétique inverse révèle l'identité des gènes impliqués dans la réparation des bris d'ADN double-brins du plastide.

2.1 Mise en contexte :

L'étude des mécanismes de réparation de l'ADN a grandement été facilitée par la découverte d'agents mutagènes causant des types de dommages spécifiques dans l'ADN (revue dans (170)). En plus de permettre une meilleure compréhension de l'impact de ce type précis de lésions à l'échelle du génome, ces agents mutagènes permettent aussi l'identification des facteurs qui sont essentiels à la réparation de ces lésions particulières. Au début de mes études doctorales, aucun agent mutagène spécifique aux organelles n'avait encore été identifié, rendant l'analyse des mécanismes impliqués dans le maintien de la stabilité génomique beaucoup plus difficile dans ces compartiments. Bref, l'identification d'un tel agent constituerait un outil précieux à notre étude.

Dans cette première partie, je présente les résultats de l'effort visant à identifier un mutagène spécifique aux organelles, qui ont abouti à l'identification de la ciprofloxacine. La ciprofloxacine est un inhibiteur de gyrases qui était déjà connu pour induire des DSBs dans les génomes procaryotes (171, 172). Notre étude est venue confirmer que ce composé a un effet similaire chez les organelles de plantes, menant à l'apparition de DSBs dans le génome des plastides et des mitochondries. Ensuite, un criblage génétique utilisant ce composé a été réalisé à l'aide d'une collection de mutants d'intérêt et a permis d'identifier certains des acteurs principaux des voies de réparation des DSBs dans les organelles. Finalement, des résultats préliminaires permettent de relier la transcription à l'instabilité génomique, malgré que l'explication précise de cette relation demeure encore nébuleuse.

Les résultats présentés dans ce chapitre proviennent de différentes sources. D'abord, ceux qui ont permis d'identifier la ciprofloxacine sont issus de l'article : « Crystal Structures of DNA-Whirly Complexes and Their Role in Arabidopsis Organelle Genome Repair » disponible à l'Annexe 1. Dans un objectif de concision, la sélection des résultats présentés pour cet article se limite à ceux qui proviennent de mon travail. Les seules exceptions sont les résultats de PCR qui ont été réalisés par Jean-Sébastien Parent. Ensuite, les résultats portant sur le criblage génétique sont des résultats non-publiés qui sont néanmoins essentiels à la compréhension de ma thèse. Tous les résultats de cette section sont le fruit de mon travail, à l'exception des expériences de transcription du type run-on, réalisées par Sébastien Truche.

2.2 Matériel et méthode :

Lignées et conditions de croissance :

Les lignées mutantes d'*Arabidopsis thaliana hsp21* (CS85472) (173), *sig2* (SALK_045706) (38), *sig6* (SAIL_645_F03) (39) et *why1why3* (165) ont été rapportées précédemment. Les conditions de croissance sont les mêmes que celles décrites dans l'Annexe 1. La température normale de croissance est de 22°C, mais certaines expériences ont été réalisées à 30°C. Le pourcentage de variégation est évalué 21 jours après la germination. La sensibilité des mutants est déterminée lors d'un traitement à 0,25 µM de ciprofloxacine. On considère une lignée hypersensible lorsque plus de 50% de la population présente un phénotype de variégation lors de ce traitement. La survie des plantes est évaluée en vérifiant l'apparition des premières vraies feuilles 28 jours après la germination (ces feuilles apparaissent normalement 10 jours après la germination). Une lignée induisant l'endonucléase I-CREII lors d'un traitement à l'œstradiol a été utilisée pour générer des DSBs dans le plastide selon le protocole décrit par les auteurs (120).

Détection des réarrangements :

L'ADN total des plantes a été extrait selon le protocole du bromure de cétyl triméthylammonium (174). Pour tous les échantillons, la quantité d'ADN a été équilibrée selon l'amplification du gène *YCF2* à l'aide des amorces YCF2R ; TGT TTC GCC TCT TAC TCG GAG et YCF2F; GAT CTC TGA GAG CTG TTT CCG. L'amplification des réarrangements génomiques a été effectuée selon un protocole préalablement décrit (165) et la séquence des amorces utilisées est disponible dans l'Annexe 1. Les fragments amplifiés par PCR sont résolus sur un gel d'agarose 1% contenant le marqueur GelRed (Biotum). L'extraction des bandes d'ADN du gel, le clonage et le séquençage des réarrangements génomiques ont été réalisés selon une méthodologie décrite précédemment (165).

Quantification des réarrangements :

L'approche de PCR en temps réel a été effectuée conformément au protocole disponible à l'Annexe 1. Les amorces utilisées sont celles connues pour générer un produit unique lors de l'amplification de réarrangement : 111330R; TCA ACA GCG GGA TTA ACC GCA, 121141F; AAC GTC ATC CGG TTA CCG GTT, 50513R; GAG ACG TCT ACC ACA ACA ACC, 59311F, AGG CCT AGT CTT TCC GGC AAT. La confirmation de l'obtention d'un produit unique en évaluant la température de fusion des produits, l'évaluation de l'efficacité des amorces et l'analyse des résultats ont été réalisés à l'aide du logiciel Light-Cycler480 version 1.5. Toutes les réactions ont été faites pour des triplicatas biologiques et techniques et sont normalisées selon l'amplification du gène *YCF2*.

Analyse par buvardage de type Southern :

Le protocole utilisé pour le buvardage de type Southern a été décrit précédemment (165). La méthodologie détaillée est disponible dans l'Annexe 1.

Analyse de transcription de type run-on :

L'isolement des chloroplastes a été réalisé selon un protocole déjà établi (175) à partir de 2 g de plantes cultivées sur terre pendant 10 jours. Un gradient de Percoll 20% / 80% permet ensuite de séparer les chloroplastes intacts de ceux casés en centrifugeant 1 heure à 7000 g dans un rotor à godets mobiles refroidie à une température de 4°C. Les chloroplastes intacts sont lavés plusieurs fois avec une solution 0.33 M sorbitol, 50 mm HEPES/KOH, pH 7.9. La quantification des chloroplastes se fait en utilisant un hémocytomètre et 1 000 000 de chloroplastes sont conservés pour chaque lignée. Le protocole de transcription de type run-on est adapté d'un protocole rapporté précédemment dans la littérature (176). Un µL de Rnase A (Fermentas) est ajouté à un échantillon contrôle au début de la réaction de transcription *in vitro* afin de confirmer que le signal détecté provient de l'incorporation de nucléotides radioactifs dans l'ARN. Les sondes utilisées pour l'hybridation ont été amplifiées par PCR à l'aide des amorces, psbaF, GCGAAAGCCTATGGGGTCGC; psbaR, GGGAGCCGCCGAATACACC; rbclF, GAAGCAGGGGCTGCGGTA; rbclR, TCGCATGTACCCGCGAGTAGC; accdF, GGTGACAACGATCTGCACTATGATCC; accdR, CATAACCCCAAGGGCTACCGG; rrn16F,

ATGAACGCTGGCGGCATG; rrn16R, GCATTCACCGCTCCACCG; psaaF, CGTTCGCCGGAACCAGAAGT; psaaR, TGCCCAAGAAAGGGACCCAAG. Cinq cents ng de chaque sonde ont été fixés à la membrane Hybond-N+ (Amersham) pour l'hybridation.

2.3 Résultats :

2.3.1 La ciprofloxacine déstabilise le génome plastidique

Afin de déterminer l'effet de la ciprofloxacine chez *Arabidopsis thaliana*, une expérience de croissance en présence de différentes doses de ciprofloxacine a été réalisée. Cette expérience révèle que la ciprofloxacine réduit la croissance des plantes de type sauvage (WT), même à des doses aussi faibles que 0,25 μM (Figure 13A). La courbe de létalité, présentée à la figure 14, permet de déterminer une valeur de demi-dose létale (LD50) d'environ 0,5 μM chez *Arabidopsis*, très similaire à celle observée chez les procaryotes (variant généralement entre 0,3 et 0,7 μM (177, 178)). Ce résultat suggère que la ciprofloxacine inhibe aussi efficacement les gyrases de plantes que celles des procaryotes. De plus, lorsque l'on traite des plantes de type sauvage avec la ciprofloxacine, il est possible d'observer l'apparition sporadique d'un phénotype de variégation, phénotype ayant déjà été associé à une augmentation de l'instabilité génomique dans le plastide (165) (Figure 13B).

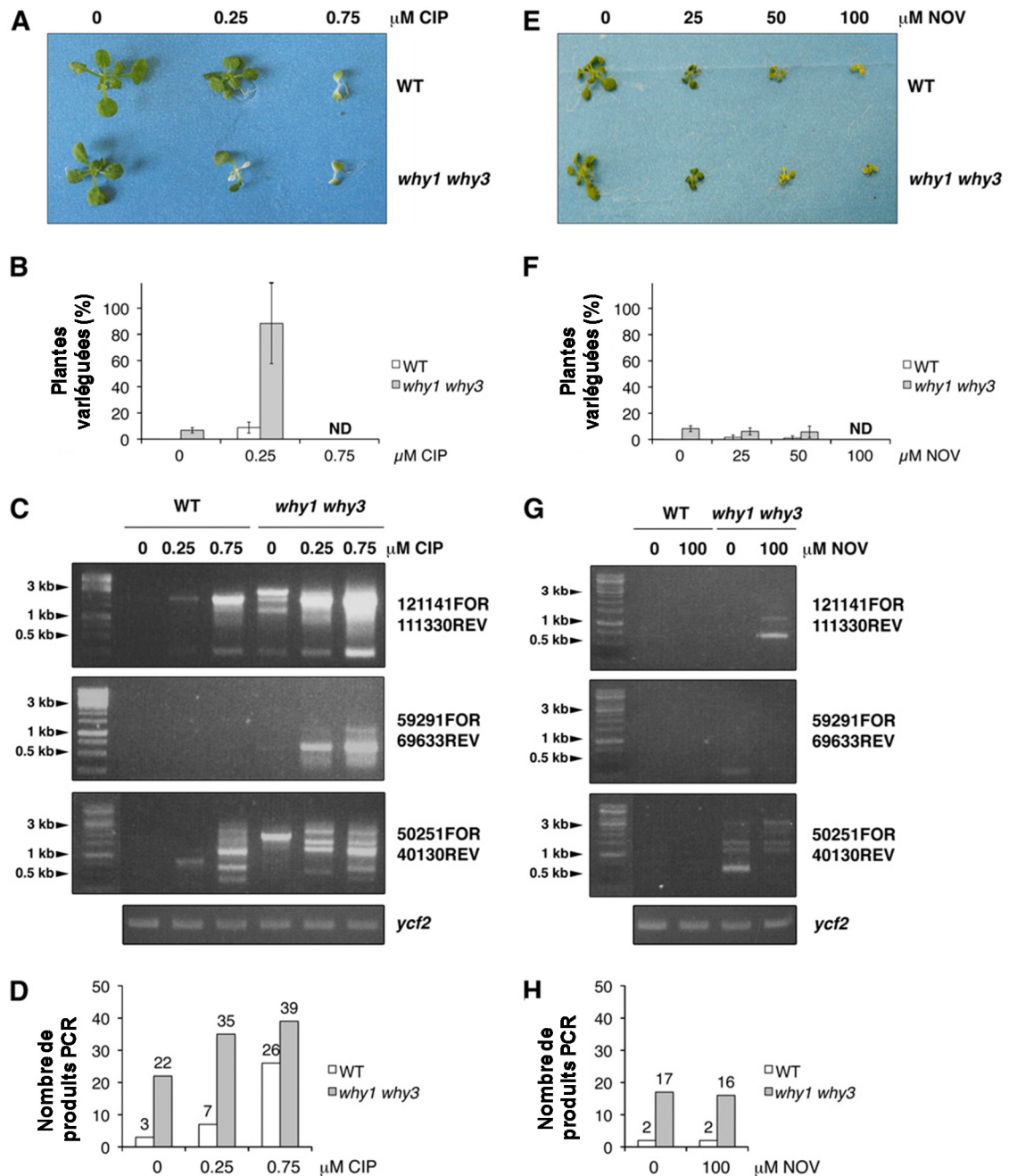


Figure 13: La ciprofloxacine déstabilise le génome plastidique d'Arabidopsis. A et E) Phénotype des plantes de type sauvage et du mutant *why1why3* après 21 jours de croissance sur milieu synthétique Murashige-Skoog (MS) supplémenté des concentrations indiquées de ciprofloxacine (A) ou de novobiocine (E). B et F) Pourcentage moyen de plantes variégues, évalué à 21 jours suite au traitement à la ciprofloxacine (B) ou à la novobiocine (F). Les barres d'erreur correspondent à l'écart type basé sur trois expériences indépendantes. La caractérisation des niveaux de variégation n'a pu être réalisée à 0,75 μM CIP et 100 μM NOVO car les premières vraies feuilles sont trop petites ou absentes. C et G) Évaluation par PCR des niveaux de réarrangements lors du traitement à la

ciprofloxacine (C) ou à la novobiocine (G). Trois réactions représentatives, issues de 13 paires d'amorces différentes, sont présentées. L'extraction de l'ADN des plantes a été réalisée à 21 jours. Les niveaux d'ADN ont été équilibrés relativement à l'amplification du gène contrôle *YCF2*. D et H) Nombre de produits PCR issus des 13 réactions PCR pour le traitement à la ciprofloxacine (D) ou à la novobiocine (H).

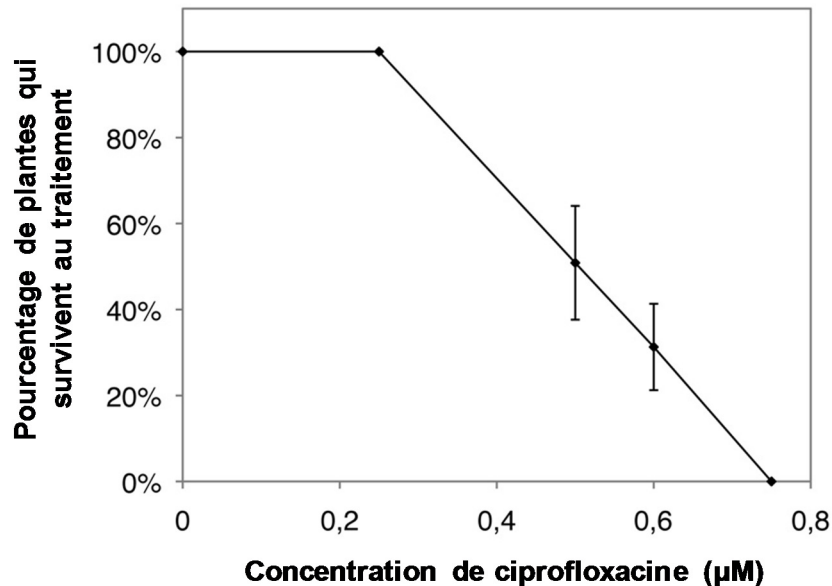


Figure 14: Le traitement à la ciprofloxacine affecte le taux de survie chez *Arabidopsis*. Graphique présentant la survie moyenne de plantes traitées à la ciprofloxacine, établie à partir de trois expériences indépendantes. Les barres d'erreur correspondent à l'écart type. La survie est déterminée par la capacité de former des premières vraies feuilles après 28 jours de croissance.

Une approche par PCR, décrite à la figure 15, a été utilisée pour vérifier l'effet de la ciprofloxacine sur le génome du plastide. Dans cette approche, des amorces réparties à travers le génome sont utilisées afin d'amplifier spécifiquement des molécules d'ADN réarrangées (165). Les résultats révèlent que, chez la plante sauvage (WT), il y a apparition de réarrangements génomiques lors du traitement à la ciprofloxacine et que leur abondance augmente de manière dose-dépendante (Figure 13, C-D). Le séquençage de ces molécules réarrangées a permis de constater la présence de séquences microhomologues aux extrémités de la vaste majorité de ces réarrangements, suggérant qu'ils seraient issus de mécanismes de réparation des DSBs dépendants de microhomologies (MHMR).

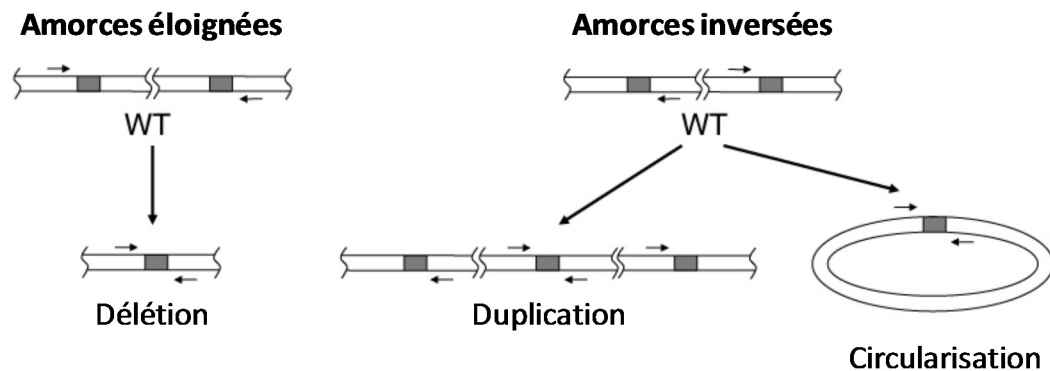


Figure 15: Approche PCR permettant d'amplifier spécifiquement des molécules d'ADN réarrangées dans le génome du plastide. Les carrés gris correspondent à des régions microhomologues (5 à 35 pb). Les coupures ondulées représentent une distance supérieure à 5 kb dans la séquence d'ADN. Les amorces séparées par ces coupures ne peuvent générer un produit dans les conditions de PCR utilisées.

L'effet de la ciprofloxacine sur le génome du plastide a ensuite été étudié à l'aide de mutants déficients pour les protéines Whirly du plastide (mutants *why1why3*). Ces mutants sont connus pour accumuler plus facilement des réarrangements génomiques (165). Il est intéressant de constater, à la figure 13A, que ces plantes mutantes sont hypersensibles au traitement à la ciprofloxacine. Le phénotype de variégation, habituellement observé dans environ 5% de la population chez *why1why3*, est maintenant observable chez environ 90% des individus traités avec 0,25 μM de ciprofloxacine (Figure 13B). Ces résultats suggèrent que les protéines Whirly jouent un rôle important dans la protection du génome face aux DSBs. De plus, les réarrangements génomiques sont aussi beaucoup plus abondants chez ce double mutant suite au traitement à la ciprofloxacine (Figure 13C-D). Ces résultats ont aussi été confirmés par des études d'immunobuvardage de type Southern à l'aide de sondes correspondant à des réarrangements génomiques (Figure 16). Étonnamment, il est souvent possible d'observer plusieurs fragments de tailles différentes pour une même sonde. Ces bandes sont uniquement observables lors du traitement à la ciprofloxacine et leur abondance corrèle avec la dose de ciprofloxacine utilisée. De plus, pour les échantillons non-digérés, on remarque des bandes de plus faible poids moléculaire. Ces résultats suggèrent que plusieurs réarrangements génomiques différents peuvent se produire à un même locus lors de conditions de stress génotoxique et que ces réarrangements peuvent mener à l'apparition de molécules sous-génomiques.

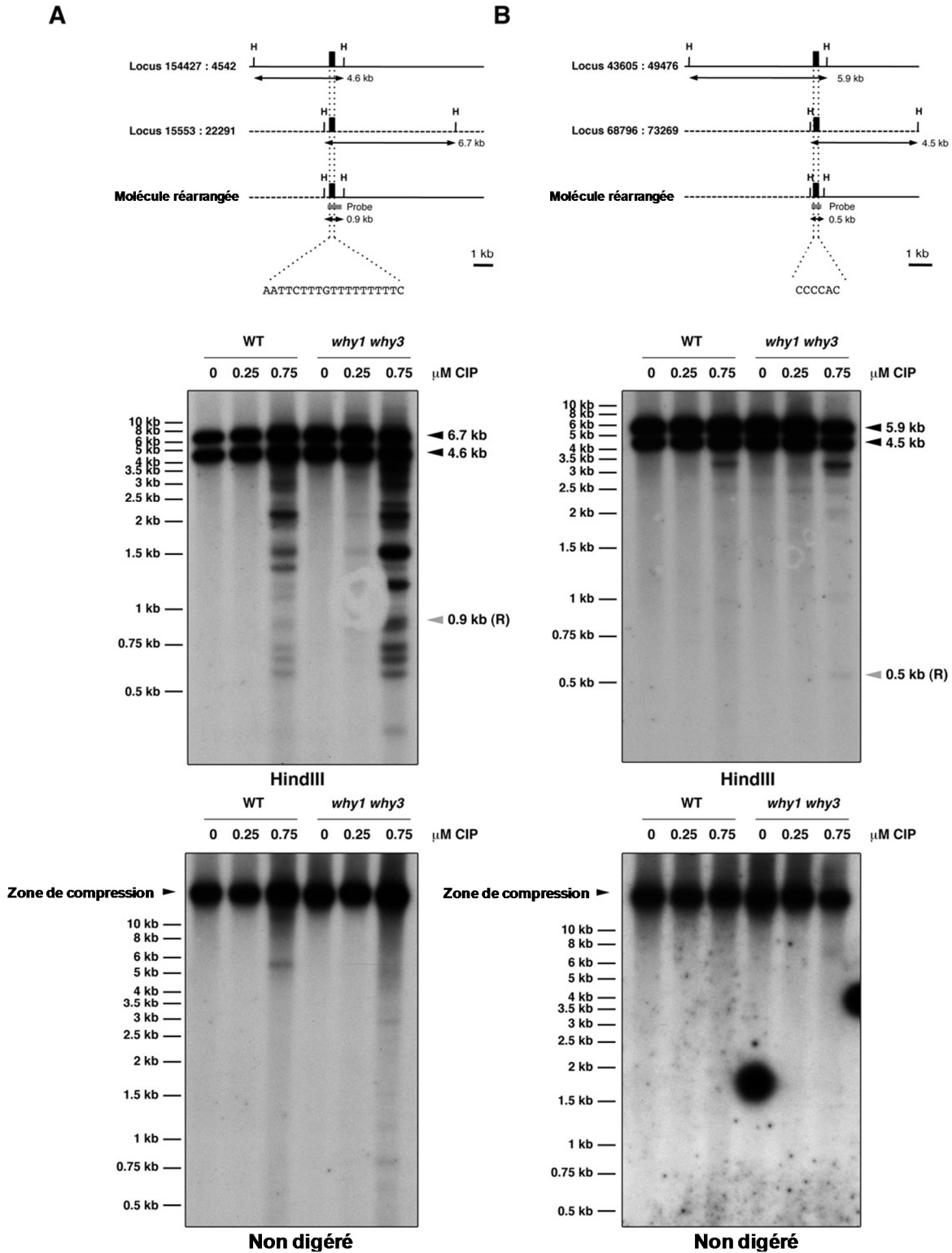


Figure 16 La ciprofloxacine cause l'apparition de réarrangements génomiques et de structures sous-génomiques. Hybridation de deux sondes radioactives différentes (A et B) correspondant à l'amplification d'un réarrangement génomique.

Panneau du haut: Représentation schématique des locus ayant été réarrangés. La sonde amplifiée y est présentée avec la séquence microhomologue ayant mené à son apparition.

Panneau du milieu: Buvardage de type Southern sur de l'ADN extrait de plantes de type sauvage (WT) et de plantes mutantes *why1why3* traitées à différentes doses de ciprofloxacine. L'extraction de l'ADN des plantes a été réalisée à 21 jours et celui-ci a été digéré avec HindIII. Les niveaux d'ADN ont été équilibrés selon l'amplification du gène contrôle *YCF2*. (R) correspond à la taille attendue du réarrangement génomique associé à la sonde.

Panneau du bas: Buvardage de type Southern sur de l'ADN non-digéré extrait de plantes de type sauvage et de plantes *why1why3* traitées à différentes doses de ciprofloxacine. L'extraction de l'ADN des plantes a été réalisée à 21 jours. Les niveaux d'ADN ont été équilibrés selon l'amplification du gène contrôle *YCF2*.

Comme mentionné précédemment, le mécanisme d'action des gyrases repose sur l'introduction contrôlée d'un DSB qui permet de générer du surenroulement négatif dans l'ADN (Figure 17). Chez les procaryotes, des études ont démontré que la ciprofloxacine inhibe l'étape de religation de l'ADN par les gyrases, laissant par le fait même un DSB dans le génome (171, 172). Afin de vérifier que les réarrangements génomiques observés lors du traitement à la ciprofloxacine sont réellement dus à la présence de DSBs, plutôt qu'à l'inhibition de l'activité de la gyrase, les expériences que nous venons de décrire ont été répétées avec la novobiocine, un autre inhibiteur de gyrases. La novobiosine agit selon un mécanisme d'action différent de celui de la ciprofloxacine. En effet, cette petite molécule inhibe l'enzyme avant l'étape du clivage de l'ADN, et n'introduit donc pas de DSBs dans le génome (voir le mécanisme à la figure 17) (179). La figure 13 (E-G) montre que la novobiocine, bien qu'inhibant efficacement la croissance des plantes, n'induit pas d'hypersensibilité chez les plantes mutantes *why1why3* et ne génère pas non plus de réarrangements génomiques dans le plastide. Ces résultats suggèrent donc fortement que l'action de la ciprofloxacine sur la survie des plantes et l'induction de réarrangements génomiques est liée à l'introduction de DSBs dans le génome des organelles. De plus, l'évaluation de la sensibilité des mutants au traitement à la ciprofloxacine semble être un outil prometteur pour l'identification de gènes impliqués dans la protection du génome plastidique.

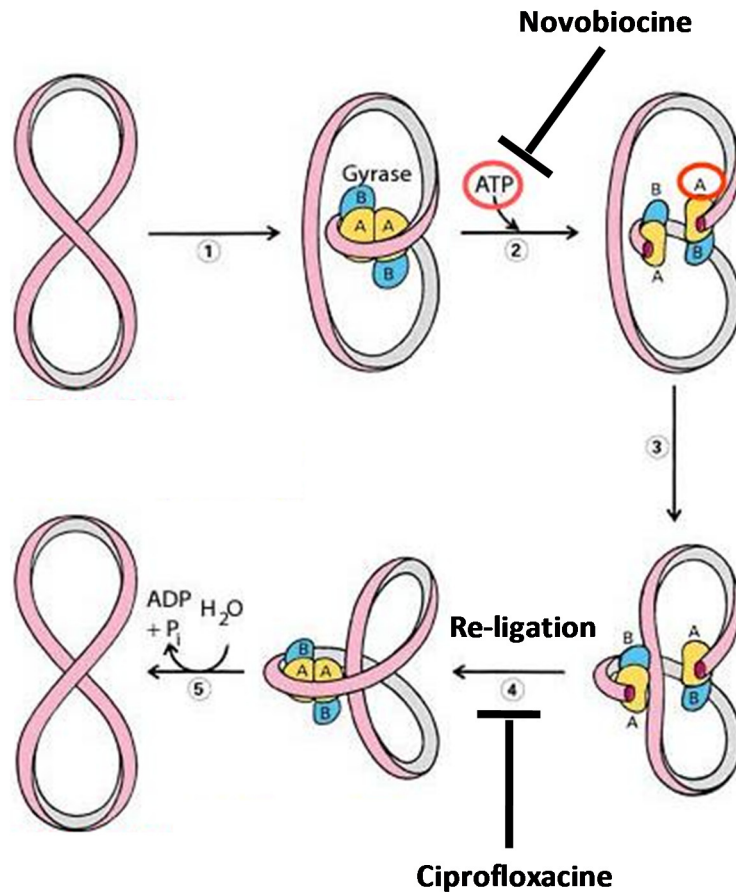


Figure 17: La ciprofloxacine et la novobiocine inhibent l'action des gyrases à des étapes différentes du mécanisme réactionnel. Représentation schématique dans laquelle l'ADN est représenté en rose, la protéine GYRA en jaune et GYRB en bleu. Figure adaptée de (180).

2.3.2 Identification des acteurs du maintien de la stabilité du génome du plastide par criblage génétique inverse

Ayant identifié un agent qui induit des DSBs dans le génome du plastide, il était maintenant possible d'élaborer un criblage génétique visant à identifier les acteurs impliqués dans le maintien de la stabilité génomique de cette organelle. Ce criblage a été conçu en tirant avantage de l'identification récente des protéines qui composent le nucléoïde du plastide (15). Dans un premier temps, nous avons constitué une collection de mutants déficients pour l'expression de protéines du nucléoïde. Cette collection est composée de 96 lignées d'insertion, obtenues des catalogues publics SALK et SAIL. Toutes les lignées ont ensuite été traitées à la ciprofloxacine (voir section Matériel et méthodes) afin

d'identifier les gènes impliqués dans la réparation des DSBs dans le plastide. Ce criblage a permis d'identifier 16 gènes dont la mutation mène à une hypersensibilité à la ciprofloxacine (Tableau I). Les gènes identifiés peuvent être regroupés en quatre catégories principales; la réparation de l'ADN, la réplication de l'ADN, la division des plastides et la transcription. On remarque aussi que la plupart de ces gènes possèdent des homologues chez les procaryotes, ce qui est peu surprenant compte tenu de l'origine endosymbiotique du plastide.

Mutant	Homologue Prokaryote	Sensibilité	Niveau de réarrangements
Réparation de l'ADN			
<i>drt111</i>	RuvC et RecG (181)	+	=
<i>reca1</i>	RecA (89)	++	=
<i>muts1</i>	MutS (182)	+	=
<i>msh1</i>	MutS (92)	+	=
<i>uvr2</i>	UvrA (182)	++	=
<i>UVR protein</i>	UvrB (182)	+	=
<i>why1why3</i>	Ssb (169)	+++	+
Réplication de l'ADN			
<i>mus308</i>	Pol ϕ (182)	+	=
<i>pollb</i>	PolA (26)	+++	=
Division du chloroplaste			
<i>ex2</i>	-	+	=
<i>vipp1</i>	-	+	=
<i>mcd1</i>	MinD (182)	+	=
Transcription			
<i>csp41-a</i>	-	+	=
<i>csp41-b</i>	-	+	=
<i>sig6</i>	SigA (39)	++	-
<i>sig2</i>	SigA (38)	+	-

Tableau I : Gènes identifiés par le criblage de génétique inverse pour les mutants du nucléoïde, lors d'un traitement à la ciprofloxacine. Les chiffres entre parenthèses correspondent aux références rapportant l'identification d'un gène homologue.

Dans la catégorie réparation de l'ADN, on retrouve entre autres les gènes *MSH1*, *RECA1*, *WHY1* et *WHY3*, tous décrits précédemment. On retrouve aussi *MUTS1*, un autre homologue du MutS

procaryote mais qui n'a pas encore été caractérisé chez les plantes, ainsi que *DRT111*, un gène qui complète une lignée d'*Escherichia coli* mutante pour les gènes RuvC et RecG (181). Les derniers membres de cette catégorie sont les gènes codant pour deux protéines ayant un domaine UVR : UVR2 et UVR protein. UVR2 a d'abord été identifié lors d'un criblage visant les gènes impliqués dans la tolérance aux UV chez *Arabidopsis* (69). Il est intéressant de constater que ce mutant est incapable d'éliminer les dimères cyclobutane de pyrimidines (CPD) dans son génome *in vivo* (69). De plus, des extraits obtenus à partir de cette lignée n'ont aucune activité photolyase (69). Il est important de mentionner que ces études ont été réalisées avec des extraits totaux et que la localisation de cette protéine n'a toujours pas été évaluée, ce qui signifie qu'elle pourrait jouer un rôle aussi bien dans le plastide que dans le noyau ou la mitochondrie. UVR Protein, pour sa part, n'a pas encore été caractérisée.

Une autre catégorie de gènes identifiée est celle de la réplication de l'ADN, qui contient deux gènes de polymérase, *POLIB* et *MUS308*. *POLIB* est l'une des polymérase du plastide étant reconnue pour avoir une forte processivité, du moins selon des études *in vitro* (25, 26). Des études plus détaillées de ce mutant seront présentées dans les chapitres suivants. Le gène *MUS308* est homologue à une polymérase de type I ayant un rôle de protection face aux dommages à l'ADN et retrouvée chez de nombreux organismes (182-184). Les mutants de *Mus308* présentent une hypersensibilité aux agents mutagènes chez la drosophile, la souris et l'humain.

Une autre catégorie du tableau I est celle des gènes impliqués dans la division des plastides. Il s'agit des gènes *MCD1* (185), *VIPP1* (186) et *EX2* (187). Ces gènes ne sont probablement pas directement impliqués dans le maintien de la stabilité génomique. En effet, il semble plus probable que l'hypersensibilité observée chez ces mutants témoigne d'une difficulté à ségréger les chromosomes réarrangés lors de la division du plastide.

Finalement la dernière catégorie identifiée est celle des gènes impliqués dans le métabolisme de l'ARN, soit *CSP41-a*, *CSP41-b* (188), *SIG2* (38) et *SIG6* (39). Le rôle des protéines CSP41-a et CSP41-b n'est pas bien caractérisé, mis à part le fait que leur mutation mène à une diminution de la transcription dans le plastide. Ces protéines semblent aussi posséder une activité ribonucléase et joueraient possiblement un rôle dans la maturation de l'ARN (188). Les gènes *SIG2* et *SIG6*, quant à eux, sont des facteurs de transcription ayant de très nombreuses cibles dans le plastide (revue dans (37)).

Des études récentes ont révélé une implication intrigante de la transcription dans la déstabilisation génomique chez les procaryotes et le génome nucléaire eucaryote (revue dans (140)). Ayant identifié différents facteurs impliqués dans la transcription plastidique, nous voulions clarifier le lien entre ce processus et l'instabilité génomique dans le plastide. La prochaine section présente des résultats qui suggèrent que la transcription déstabiliserait le génome plastidique en favorisant l'apparition de réarrangements dépendants de microhomologies. Ces résultats sont toutefois préliminaires et davantage d'expériences seront nécessaires pour venir appuyer les conclusions que nous en tirons.

2.3.3 Les mutants ayant un taux de transcription réduit accumulent moins de réarrangements dans leur génome plastidique

Les niveaux de réarrangements génomiques ont été évalués pour tous les mutants identifiés lors du criblage génétique. Curieusement, tel que détecté par l'approche PCR, seul le mutant *why1why3* montre une augmentation importante de réarrangements génomiques dans le plastide. À l'opposé, les mutants *sig2* et *sig6* présentent des taux de réarrangements génomiques plus faibles que les plantes de type sauvage (Figure 18). Afin d'avoir une vision plus quantitative de l'abondance des réarrangements génomiques chez ces mutants, des études de PCR en temps réel ont été réalisées. Les résultats révèlent une diminution de plus de 80% des molécules réarrangées pour les mutants *sig2* et *sig6*. Ce résultat suggère que l'inhibition de la transcription protégerait le génome du plastide. Si tel est le cas, l'inhibition de la transcription par traitement avec une petite molécule devrait avoir un effet similaire aux mutations *sig2* et *sig6* sur la stabilité du génome. Afin de tester cette hypothèse, des plantes mutantes *why1why3* furent traitées avec 1 mM de rifampicine, un inhibiteur de la transcription plastidique (189). Le mutant *why1why3* fût choisi pour ces expériences à cause de ses niveaux élevés de réarrangements génomiques. Les résultats de cette expérience révèlent que le traitement à la rifampicine mène à une forte diminution des niveaux de réarrangements génomiques de ce mutant (Figure 18C). Cette diminution suggère donc que la transcription aurait un important effet déstabilisant dans le plastide.

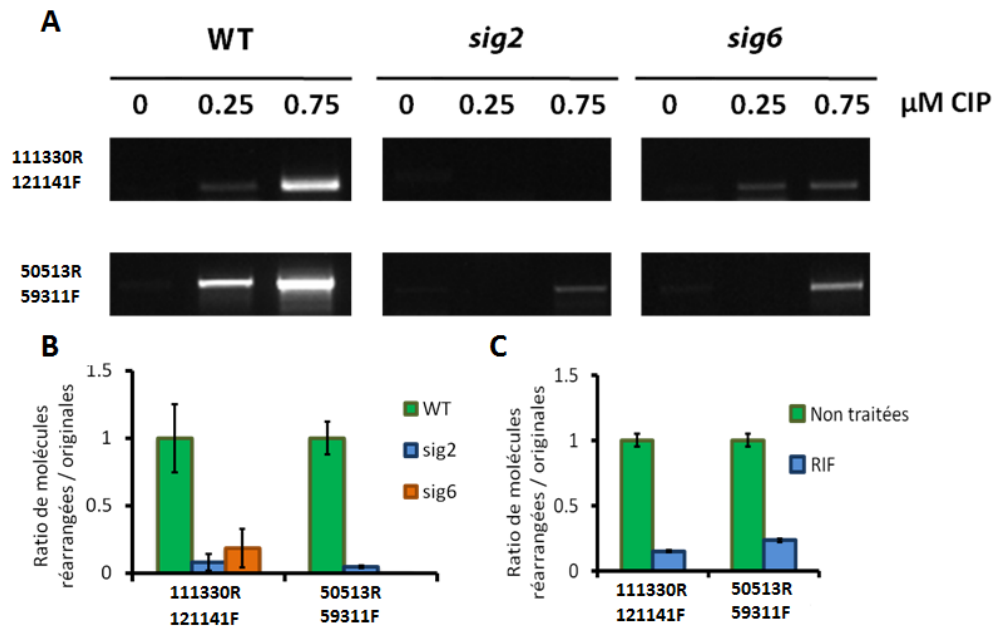


Figure 18: Les mutants *sig2* et *sig6* et le traitement à la rifampicine réduisent les niveaux de réarrangements génomiques dans le plastide. A) Évaluation par PCR des niveaux de réarrangements génomiques lors du traitement à la ciprofloxacine pour les plantes de type sauvage (WT), *sig2* et *sig6*. Photographies de gels pour deux réactions représentatives, issues de 8 paires d'amorces différentes. L'extraction de l'ADN des plantes a été réalisée à 14 jours. Les niveaux d'ADN ont été équilibrés selon l'amplification du gène contrôle *YCF2*. B) Quantification par PCR en temps réel des réarrangements suite à un traitement avec 0,75 μM de ciprofloxacine (triplicata biologique de plantes de type sauvage, *sig2* et *sig6*). Les barres d'erreur correspondent à l'écart type. L'extraction de l'ADN des plantes a été réalisée à 14 jours. Le gène *YCF2* est utilisé pour équilibrer les niveaux d'ADN plastidique. C) Quantification des réarrangements par PCR en temps réel chez le mutant *why1why3* traité avec 1mM de rifampicine (triplicata biologique). Les barres d'erreur correspondent à l'écart type. L'extraction de l'ADN des plantes a été réalisée à 14 jours. Le gène *YCF2* est utilisé pour équilibrer les niveaux d'ADN plastidique.

Afin de vérifier si les taux de transcription sont réellement réduits chez les mutants étudiés, des tests de transcription de type run-on ont été réalisés. Sans surprise, la mutation de *SIG2* ou *SIG6* diminue fortement l'expression des cinq gènes plastidiques évalués (Figure 19). Afin de confirmer que le signal détecté lors de l'hybridation provient réellement de la transcription, un traitement à la ribonucléase a été réalisé sur l'ARN extrait d'un échantillon de chloroplastes isolés de la lignée de type sauvage. Aucun signal ne peut être observé suite à ce traitement, confirmant que les nucléotides radioactifs étaient incorporés dans l'ARN.

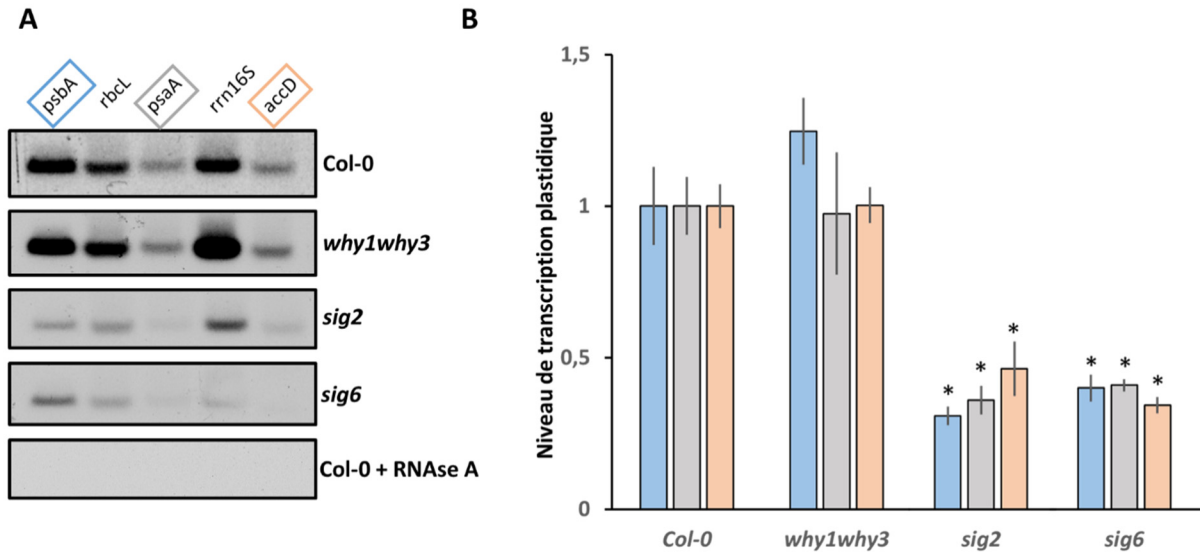


Figure 19: Les taux de transcription diminuent dans les mutants *sig2* et *sig6*. A) Autoradiographie d'ARN pour des gènes transcrits dans des chloroplastes isolés des lignées de type sauvage (WT), *why1why3*, *sig2* et *sig6*. Les autoradiogrammes présentés sont représentatifs d'un triplicata biologique. Chaque sonde fixée à la membrane est spécifique à un seul gène et est composée de 500 nucléotides. Le nombre de chloroplastes a été déterminé à l'aide d'un hémocytomètre et normalisé entre les lignées. B) Histogramme représentant l'intensité du signal radioactif provenant de l'hybridation à la sonde d'ADN. Les gènes *PSBA*, *PSAA* et *ACCD* sont respectivement présentés en bleu, mauve et rose. Les barres d'erreur correspondent à l'écart type. L'astérisque représente une différence significative selon un test de student avec un $p < 0.01$.

Un système permettant d'induire la transcription de manière contrôlée serait un outil précieux afin de confirmer l'effet de la transcription sur la stabilité génomique. Pour obtenir un tel système, nous avons utilisé le mutant *hsp21*, qui est déficient pour une protéine de tolérance à la chaleur (HSP). Cette HSP se fixe à la PEP et permet d'assurer la transcription lorsque la plante est exposée à de hautes températures. Des études de run-on ont révélé qu'un traitement à haute température inhibe modérément la transcription chez ce mutant (173). Les résultats présentés à la figure 20 indiquent que les niveaux de réarrangements sont similaires dans les plantes de type sauvage et le mutant *hsp21* à 22°C. Cependant, dans des conditions de croissances à température élevée (30°C), les réarrangements génomiques deviennent beaucoup plus faibles chez *hsp21*. Ces résultats, bien que préliminaires, supportent nos conclusions précédentes concernant l'effet de la transcription sur la stabilité du génome plastidique. Il sera aussi possible dans le futur d'utiliser ce mutant afin de déterminer la chronologie de l'apparition des réarrangements génomiques.

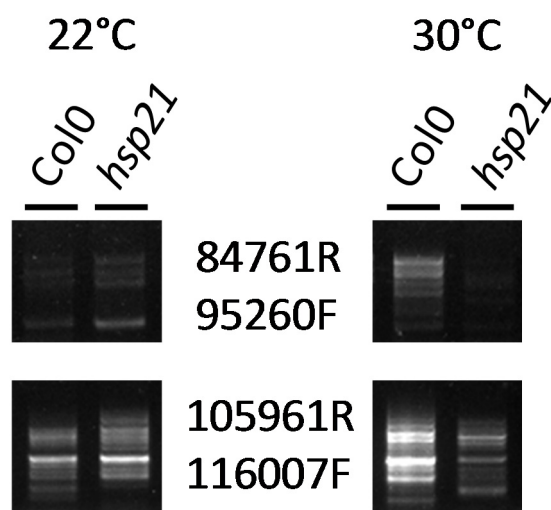


Figure 20: L'inhibition de la transcription à haute température chez *hsp21* diminue l'apparition de réarrangements génomiques. Évaluation par PCR des niveaux de réarrangements pour les plantes de type sauvage et *hsp21* à 22°C (A) ou 30°C (B). Photographies de 2 réactions représentatives, issues de 13 paires d'amorces différentes distribuées à travers le génome. L'extraction de l'ADN des plantes a été réalisée à 14 jours. Les niveaux d'ADN ont été équilibrés relativement à l'amplification du gène contrôle *YCF2*.

2.3.3.1 La mutation de *SIG6* complémente le phénotype moléculaire de *why1why3*

Puisque l'inhibition de la transcription diminue l'abondance de réarrangements génomiques dans le plastide, on peut supposer que la mutation des facteurs sigmas complémenterait l'absence des protéines Whirly, qui elles ont un effet inverse sur la stabilité génomique. Afin de tester cette hypothèse, des plantes issues d'un croisement génétique entre les plantes *why1why3* et *sig6* ont été soumises à une analyse PCR. La figure 21 montre que la mutation de *SIG6* a un effet marqué dans ce contexte et rétablit la stabilité du génome plastidique. Les niveaux de réarrangements du triple mutant *why1why3sig6* sont même légèrement plus faibles que ceux observés chez la plante de type sauvage. Suite à un traitement à la ciprofloxacine, les niveaux de réarrangements génomiques demeurent très bas dans ce triple mutant. Étrangement, le phénotype de variéation est toujours observable chez environ 5% de la population. Cependant, la taille et l'abondance des secteurs blancs sont plus faibles chez les plantes affectées (Lepage É., résultats non publiés). Ces résultats, combinés à

ceux obtenus grâce au traitement à la rifampicine, permettent de supposer que l'instabilité génomique causée par la perte des Whirly est principalement liée à la transcription.

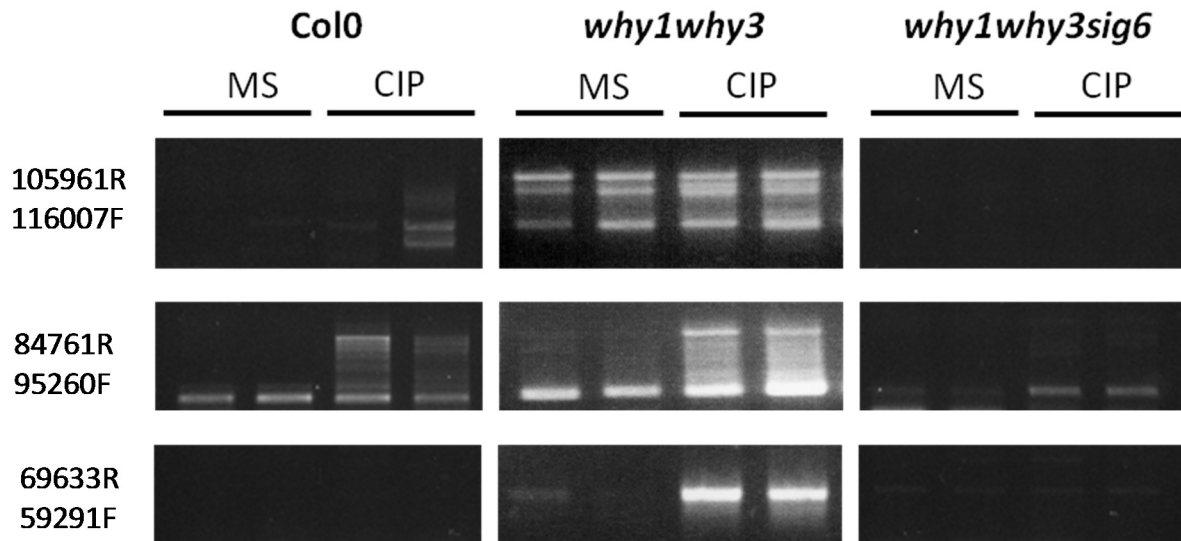


Figure 21: La mutation de *SIG6* complète celle des Whirly au niveau de l'apparition de réarrangements génomiques. Évaluation par PCR des niveaux de réarrangements pour des plantes cultivées sur milieu Murashige-Skoog (MS) avec ou sans ajout de 0,5 μ M ciprofloxacine (CIP). Photographies de 3 réactions représentatives, issues de 8 paires d'amorces différentes. L'extraction de l'ADN des plantes a été réalisée à 14 jours. Les niveaux d'ADN ont été équilibrés selon l'amplification du gène contrôle *YCF2*.

2.3.3.2 L'inhibition locale de la transcription protège face aux réarrangements

Certains facteurs sigmas du plastide reconnaissent des promoteurs qui leur sont propres. L'un de ceux-ci est le facteur sigma 5, qui est spécifique au promoteur de *PSBD*. En l'absence de SIG5, le niveau d'expression de *PSBD* est fortement diminué (41). Afin de vérifier si l'inhibition locale de la transcription, au niveau du promoteur de SIG5, pouvait avoir des effets similaires sur la stabilité du génome, des plantes mutantes *sig5* ont été traitées à la ciprofloxacine. De façon intéressante, l'apparition de réarrangements génomiques est fortement inhibée dans ce mutant dans la région de *PSBD* (Figure 22A). Toutefois, les réarrangements se produisant ailleurs dans le génome semblent peu

affectés (Figure 22B), ce qui suggère encore une fois que la transcription aurait un effet déstabilisant sur le génome. Ces résultats suggèrent aussi que la transcription doit être active à proximité des sites où se produisent les réarrangements observables par PCR.

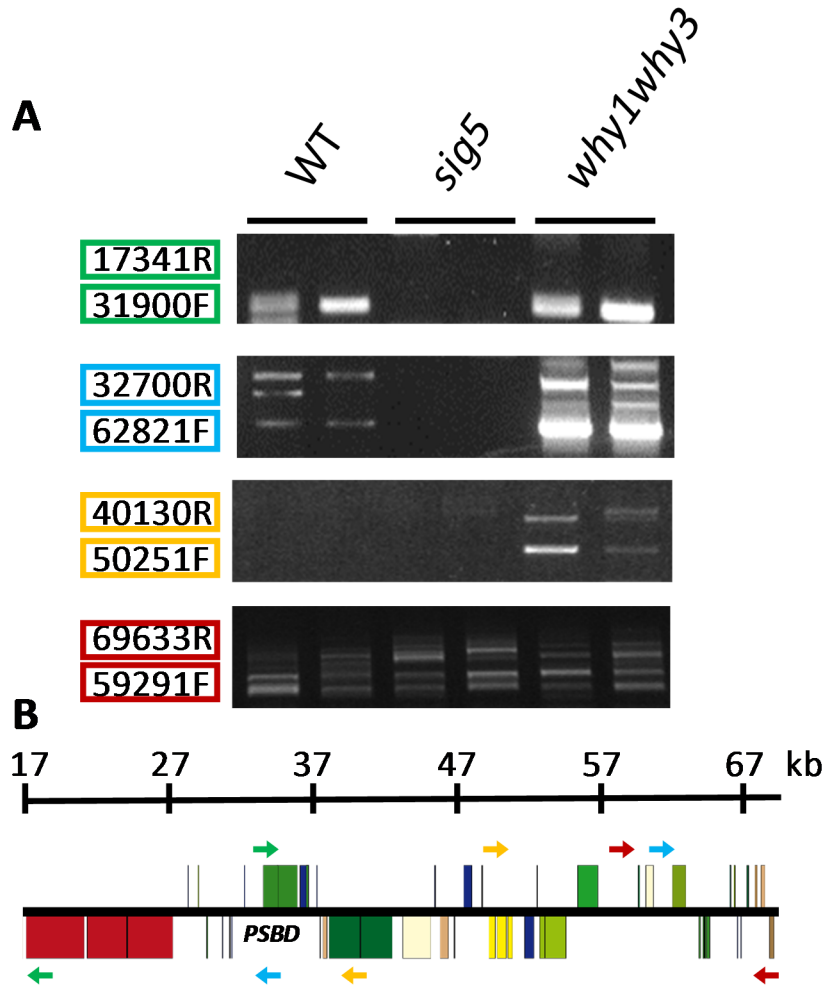


Figure 22 : La mutation de *SIG5* empêche l'apparition de réarrangements génomiques autour du gène *PSBD*. Évaluation par PCR des niveaux de réarrangements pour les plantes de type sauvage, *sig5* et *why1why3*. A) Photographies de 4 réactions représentatives, issues de 13 paires d'amorces différentes distribuées tout autour du génome. L'extraction de l'ADN des plantes a été réalisée à 21 jours. Les niveaux d'ADN ont été équilibrés relativement à l'amplification du gène contrôle *YCF2*. B) Représentation schématique de l'environnement génétique de *PSBD*. Les flèches de couleur représentent les amorces utilisées en A. La ligne au-dessus du schéma indique les positions en kilobase sur le génome complet. Le code de couleur pour les gènes est le même qu'à la figure 3.

2.3.3.3 L'effet déstabilisant de la transcription est indépendant de la production de DSBs

Une façon par laquelle la transcription pourrait induire de l'instabilité génomique dans le plastide serait via la formation de DSBs. Chez *Arabidopsis*, des lignées induisant l'endonucléase I-CREII lors d'un traitement à l'œstradiol permettent de générer un DSB uniquement dans le gène *PSBA* (120). Un tel système représente un outil précieux pour étudier la réparation de l'ADN à un site spécifique de DSB. La figure 23 montre que l'induction de l'endonucléase à l'aide de l'œstradiol mène à l'apparition de réarrangements génomiques dans les environs du gène *PSBA*. Le séquençage de ces réarrangements, déjà publié, a permis de constater qu'ils sont issus du MHMR (120). Cependant, un traitement combiné à la rifampicine (1 mM) et à l'œstradiol inhibe complètement l'apparition de ces mêmes réarrangements génomiques. Il semble donc que l'inhibition de la transcription protège aussi le génome dans une situation où les DSBs sont induits de manière exogène, soit ici par une endonucléase. Puisque les DSBs dans cette situation ne sont pas causés directement par la transcription, ce résultat suggère que l'impact de la transcription sur la stabilité du génome ne soit pas limité à l'induction de DSBs. Ces résultats sont cependant préliminaires et il serait nécessaire d'effectuer certains contrôles, tels que vérifier l'expression et l'activité de l'endonucléase lors du traitement à la rifampicine, avant de pouvoir émettre des conclusions définitives.

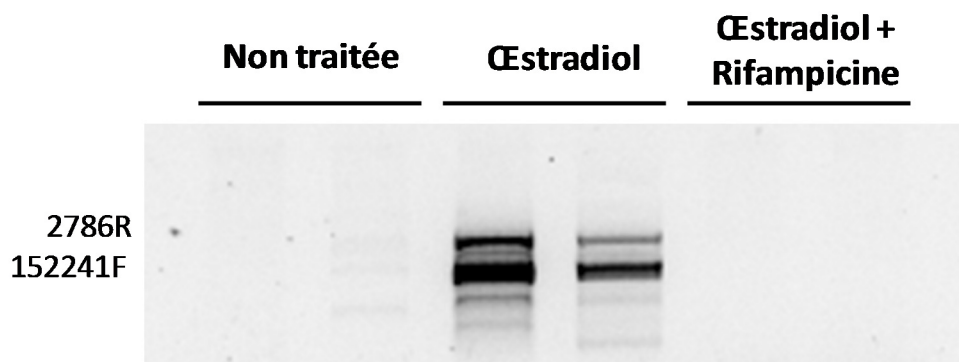


Figure 23: L'inhibition de la transcription par la rifampicine empêche la formation de réarrangements induits par un DSB artificiel. Évaluation par PCR des niveaux de réarrangements causée par les coupures de I-CREII dans le gène *PSBA*. L'extraction de l'ADN des plantes a été réalisée à 21 jours. Les niveaux d'ADN ont été équilibrés relativement à l'amplification du gène contrôle *YCF2*.

Chapitre 3: Premier article

Divergent roles for the two Poll-like organelle DNA polymerases of Arabidopsis.

Article publié dans le journal *Plant Physiol* **156**: 254-262 (2011)

Copyright American Society of Plant Biologists

3.1 MISE EN CONTEXTE

Le maintien de la stabilité génomique est un processus complexe qui nécessite la coordination de nombreux mécanismes aux fonctions variées. Le criblage génétique présenté au chapitre 2 a permis d'identifier plusieurs gènes vraisemblablement impliqués dans ce processus essentiel. Pour faire suite à ces résultats, il est nécessaire de déterminer le rôle précis de chacun des gènes identifiés, dans l'espoir d'en apprendre davantage sur les mécanismes nécessaires au maintien de la stabilité génomique du plastide. Un des gènes identifiés qui a rapidement suscité notre intérêt est celui de la polymérase de type I, *POLIB*. Notre intérêt envers ce gène découlait de deux motifs principaux. D'abord, la machinerie de réplication de l'ADN du plastide n'est que très peu caractérisée. Ensuite, la famille de polymérases de type-I ayant un peptide transit pour les organelles est composée de deux membres chez *Arabidopsis*, *POLIA* et *POLIB*. Or, étonnamment, notre criblage indique que seuls les mutants de *POLIB* sont hypersensibles au traitement à la ciprofloxacine. Cette observation suggère que, bien que très similaires, ces deux gènes ont probablement acquis des rôles distincts. Toutes ces raisons nous ont poussés à clarifier le rôle de ces deux polymérases de type-I dans les organelles d'*Arabidopsis*.

Mon rôle dans cette étude consista à déterminer la sensibilité des différentes lignées face au traitement à la ciprofloxacine, à développer une approche pour évaluer la fréquence de DSBs et à caractériser les lignées issues de croisements génétiques. JSP a pour sa part isolé les différentes lignées mutantes, évalué les quantités d'ADN dans les organelles par PCR quantitatif et participé à la caractérisation des mutants. L'analyse des résultats et l'écriture de l'article furent un travail conjoint entre JSP, Normand Brisson et moi-même. J'ai réalisé les figures de l'article.

Divergent roles for the two Poll-like organelle DNA polymerases of *Arabidopsis thaliana*

Jean-Sébastien Parent^{2,3}, Etienne Lepage² and Normand Brisson*

Department of Biochemistry, Université de Montréal, C.P. 6128, PO Box Centre-Ville,
Montréal, Qc, Canada H3C 3J7

¹This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). E.L. was supported by a scholarship from NSERC.

²These authors contributed equally to this work.

³Current address : Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique, 78026 Versailles Cedex, France.

3.2 ABSTRACT

DNA polymerases play a central role in the process of DNA replication. Yet, the proteins in charge of replication of plant organelle DNA have not been unambiguously identified. There are, however, many indications that a family of proteins homologous to bacterial DNA polymerase I is implicated in organelle DNA replication. Here, we have isolated mutant lines of the *PollA* and *PollB* genes of *Arabidopsis thaliana* to test this hypothesis. We find that mutation of both genes is lethal, thus confirming an essential and redundant role for these two proteins. However, the mutation of a single gene is sufficient to cause a reduction in the levels of DNA in both mitochondria and plastids. We also demonstrate that *pollb*, but not *polla* mutant lines, are hypersensitive to ciprofloxacin, a small molecule that specifically induces DNA double-strand breaks in plant organelles, suggesting a function for PolIB in DNA repair. In agreement with this result, a cross between *pollb* and a plastid *Whirly* mutant line yielded plants with high levels of DNA rearrangements and severe growth defects, indicating impairments in plastid DNA repair pathways. Taken together, this work provides further evidences for the involvement of the plant *Poll-like* genes in organelle DNA replication and suggests an additional role for PolIB in DNA repair.

3.3 INTRODUCTION

Faithful replication of DNA is one of the most crucial processes in living organisms. Plant cells are particular in this regard because of their need to maintain genomes in three separate compartments: the nucleus, the plastid and the mitochondria. While nuclear DNA replication and repair have been extensively studied, little is known about these processes in plant organelles. For example, the DNA polymerases responsible for the replication of the genome of mitochondria and plastids have not yet been clearly identified.

In higher plants, DNA polymerase activity has been purified from chloroplast extracts of *Spinacia oleracea* (spinach) (Spencer and Whitfield, 1969; Sala et al., 1980) and *Pisum sativum* (pea) (McKown and Tewari, 1984), while a mitochondrial DNA polymerase activity has been purified from *Triticum aestivum* (wheat) (Castroviejo et al., 1979) and the flowering plant *Chenopodium album* (Meissner et al., 1993). The DNA-polymerase activities from these various sources have similar characteristics, including insensitivity to aphidicolin, a known inhibitor of nuclear eukaryotic α DNA polymerases, a broad alkaline pH optimal activity and the requirement for a divalent metal cation (Mg^{2+} or Mn^{2+}). This suggests that plant organelles, like animal mitochondria, possess their own specialized enzymes, with distinct characteristics, for the replication of their genome (Sala et al., 1980).

Isolation of DNA polymerases from both the chloroplast and mitochondria of *Glycine max* (soybean) and *Nicotiana tabacum* (tobacco) cell cultures allowed further comparison of the two enzymes contained in these compartments (Heinhorst et al., 1990; Sakai et al., 1999). Interestingly, the biochemical properties and molecular weights of these enzymes were shown to be very similar. They reach maximal activity under the same experimental conditions and are sensitive to the same chemical agents, suggesting that they are at least homologous proteins. These enzymes were also shown to share characteristics with the γ DNA polymerases (Poly) found in the mitochondria of animals (Heinhorst et al., 1990; Sakai et al., 1999), such as insensitivity to aphidicolin, requirement for KCl and template preference. Another important characteristic shared by these purified enzymes and Poly is that they possess a 3'-5' exonuclease activity, as demonstrated for the DNA polymerases from spinach chloroplasts (Keim and Mosbaugh, 1991) and *C. album* mitochondria (Meissner et al., 1993).

More recently, genes encoding proteins targeted to organelles and with similarity to the gene coding for *Escherichia coli* DNA polymerase I (Poll) were identified in *Oryza sativa* (rice) (Kimura et al., 2002), *Arabidopsis thaliana* (Arabidopsis) (Mori et al., 2005) and tobacco (Ono et al., 2007). Full-length cDNAs were cloned and regions corresponding to the DNA polymerase domains of these enzymes were produced in bacteria and tested for template preference, biochemical properties and sensitivity to inhibitors. Like the enzymes purified from plants, the recombinant proteins are insensitive to aphidicolin and dideoxynucleotides (ddNTP) (Kimura et al., 2002; Mori et al., 2005; Ono et al., 2007), suggesting that these DNA Poll-like enzymes are involved in DNA replication in plants organelles.

It therefore appears that a conserved family of eubacteria-like DNA polymerases is responsible for the replication of DNA inside plant organelles. Indeed, every plant genome that has been sequenced to date possesses two *Poll-like* genes which, in Arabidopsis and tobacco, encode proteins that localize to both organelles, thus suggesting an important conserved function (Elo et al., 2003; Christensen et al., 2005; Ono et al., 2007; Wamboldt et al., 2009). In addition, the recombinant proteins have a high processivity, which would allow them to replicate complete organelle genomes (Kimura et al., 2002; Mori et al., 2005; Ono et al., 2007). All these characteristics make the Poll proteins ideal candidates for a role in organelle DNA replication. However, this function has not yet been confirmed *in vivo*. Here we report the isolation of mutant lines for the two Arabidopsis *Poll-like* genes At1g50840 and At3g20540 (named *PollIA* and *PollIB* henceforth). While no visible phenotype was observed with two mutant alleles of *PollIA*, a strong mutant allele of *PollIB* yielded plants that grew more slowly than wild-type (WT) plants. More importantly, we were unable to recover plants bearing homozygous mutations in both genes, indicating that they play an essential, at least partially redundant role in Arabidopsis. All mutants tested showed reduced levels of organelle DNA relative to nuclear DNA, suggesting that they are involved in genome replication. Interestingly, only the mutants for *PollIB* showed increase sensitivity to the organelle-specific DNA double-strand break-inducing agent ciprofloxacin, suggesting that one of the two Poll proteins has specialized in DNA repair.

3.4 RESULTS

***POLIA* and *POLIB* Share a Common Essential Function in Organelles**

To provide genetic evidence for the implication of the Arabidopsis *PollA* and *PollB* genes in organelle DNA replication, we have obtained two T-DNA insertion mutant lines for each gene. Figure 24A shows the position of the insertion in each line. Figure 24B shows *PollA* mRNA levels, measured by qRT-PCR, in the *polla* lines. The two mutations in *PollA* caused a marked reduction in the expression level of the gene. The *polla-1* line (SALK_022638) showed a 536 fold reduction and the *polla-2* line (SALK_065221) a 12 fold reduction in mRNA level as compared to WT. Mutations in the *PollB* gene had a different effect depending on the region of the gene that was amplified. Figure 24C shows the results of two qRT-PCR reactions using primers located either in the middle of the gene or at the 3' end (see Fig. 24A). The *pollb-1* line (SALK_134274) showed a 33 fold reduction in expression in the mid-gene section but only a 3 fold reduction near the end of the gene. The *pollb-2* line (WiscDsLoxHs021_09D) resulted in a 1.5 fold reduction of the mid-gene section and a 365 fold reduction at the 3' end. This difference in signal level between two parts of the same messenger could be due to differential processing of the mRNA caused by the T-DNA insertion, or to leaky transcription from the 35S promoter inside the T-DNA insert. This could lead to different results when using probes placed on either side of the insertion site, as it is the case for *pollb-2*.

The phenotype of the two *polla* lines did not differ from WT (Fig. 24D). Interestingly, while the *pollb-1* mutation also yielded no differential phenotype, the *pollb-2* line showed growth retardation (Fig. 24D). This difference in phenotype compared to *pollb-1* might be attributed to the lower expression of *PollB* (maximal reduction of 365 fold vs 33 fold for the *pollb-1* line) or to the site of the insertion (5th exon vs 1st intron). We then attempted to cross *polla-1* with *pollb-1*, and *polla-2* with *pollb-2* to obtain mutant plants devoid of Poll proteins. We were unable to isolate double homozygous mutant plants in the progeny of double heterozygous plants. By testing the germination of seeds from the cross on MS medium, we did not observe any aborted seed and we concluded that the double homozygous individuals do not produce mature seeds. We went a step further for the *polla-1/pollb-1* cross and isolated one

plant homozygous for *pollb-1*, but heterozygous for *polla-1*. We then tested 48 plants from the progeny of this plant and again could not recover any double homozygous plant. The significance of this result was confirmed by a χ^2 test ($p=0$). These results indicate that at least one intact *Poll* gene is essential for plant viability. Therefore, in the rest of our work we focused on the analysis of single mutant plants.

Mutants for the *POL* Genes Have Less Organelle DNA

Mutants for DNA polymerase genes are expected to have a reduced capacity to synthesize and accumulate DNA. We therefore measured the relative level of organelle to nuclear DNA in the *Poll* mutants in order to evaluate plastid genome copy number in *Arabidopsis* (Zoschke et al., 2007; Rowan et al., 2009). WT, *polla-1* and *pollb-1* mutant lines were grown on MS medium, whole seedlings (including the roots) were harvested at 3, 4 and 5 days after germination and their total DNA isolated. The level of DNA was measured by qPCR using primers specific for a given DNA sequence inside the nucleus, the mitochondria or the plastid. The levels for the mitochondria and plastid DNA (mtDNA and ptDNA) are reported as a ratio to the nuclear DNA level. Figure 25A shows that, in 3 and 4 day old seedlings, the relative level of mtDNA is lower in both *polla-1* and *pollb-1* as compared to the WT plants, but that this difference is less pronounced by day 5. A similar trend is observed when testing ptDNA, although in this case the difference between WT and *pollb-1* is still pronounced at day 5 (Fig. 25B). Similar results were obtained with the *polla-2* line (Fig. 28). However, it was not possible to test the *pollb-2* line at these early time points due to its growth retardation defect. These results show that *PollA* and *PollB* are both required for maximal DNA accumulation in *Arabidopsis* plastids and mitochondria.

Mutants for the *POLIB* but not *POLIA* Gene Are Impaired in their Capacity to Repair DNA Double-Strand Breaks

DNA-synthesis activity is required for the repair of DNA damages by mechanisms such as homologous recombination (HR). To test for a role in DNA repair, the *Poll* mutants were treated with ciprofloxacin, a molecule that induces DNA double-strand breaks (DSBs) specifically in plant organelle DNA (Wall et al., 2004) and was recently shown to affect ptDNA conformation (Rowan et al., 2010). Interestingly, only the *pollb* mutants showed

increased sensitivity to ciprofloxacin when compared to WT plants (Fig. 26A and B). At 0.25 μM of ciprofloxacin, *pollb-2* plants were more sensitive to the drug than *pollb-1* plants. In this condition, the *pollb-1* plants were able to produce small first true leaves with white regions while *pollb-2* plants produced abnormal first true leaves that were completely white (Fig. 26B). At 0.5 μM ciprofloxacin very few true leaves were observed for either *pollb* alleles. Figures 26C and D present, for each line and condition tested, the percentage of plants showing first true leaves with white regions or no first true leaves, respectively.

These results suggest that in addition to a role in DNA replication, PolIB is also required for the repair of DNA. In that case one can predict that the DSBs induced by ciprofloxacin will accumulate to higher levels in *pollb-1* plants than in WT or *polla-1* plants. This was tested by measuring the level of DNA lesions in plastid and mitochondrial DNA using a PCR approach (Yakes and Van Houten, 1997). Figure 26E and F show a clear, dose dependent decrease in the relative amplification of a long DNA fragment when using *pollb-1* DNA isolated from plants treated with ciprofloxacin as template. This indicates the presence of a higher level of DNA lesions in the organelle genomes of these plants as compared to WT and *polla-1* plants. Although a similar trend was observed when comparing the results for ptDNA (Fig. 26E) and mtDNA (Fig. 26F), the difference with WT DNA was less marked when testing mtDNA. This difference was nevertheless significant ($p < 0.05$ in Student t-test) for the highest concentration of the antibiotic. To confirm that the DSBs induced by ciprofloxacin are specific to organellar DNA, we designed a similar experiment with nuclear probes and found, as expected, no increase in DNA lesions in WT and *pollb-1* plants treated with 0.5 μM ciprofloxacin (Fig. 29). Taken together, these results indicate a role for PolIB in DNA repair in both the mitochondria and the plastid.

Genetic Interaction between *POLIB* and Whirly Genes

The above results suggest that PolIB is specifically recruited at DSB sites and contribute to their repair. In prokaryotes, the recruitment of repair proteins to the DNA is orchestrated by the single-stranded DNA-binding proteins (SSBs) (reviewed in (Shereda et al., 2008)). We have recently reported that the single-stranded DNA-binding proteins Whirly are required for accurate repair of DSBs in plant organelles and proposed that their function in this

process is comparable to that of SSBs in prokaryotes (Maréchal et al., 2009; Cappadocia et al., 2010). To test whether a genetic interaction exists between the *Whirly* and the *Poll* genes, we crossed the *polla-1* and the *pollb-1* mutants with the mitochondria (*why2-1*) or the plastid (*why1why3*) *Whirly* mutants reported previously (Maréchal et al., 2009; Cappadocia et al., 2010). All but one of the crosses yielded plants with no difference in phenotype when compared to WT plants, except for the occasional appearance of variegation in lines bearing the *why1why3* mutations (Maréchal et al., 2009) (Fig. 27A and Fig. 30). Interestingly, the *pollb-1/why1why3* homozygous triple mutant exhibited a severe growth defect and pale leaves, indicating that a synergistic interaction exists between plastid Whirlies and PolIB.

The *why1why3* mutant line of *Arabidopsis* has been shown to promote the repair of DSBs through an error-prone microhomology-mediated break-induced replication (MMBIR) pathway (Cappadocia et al., 2010) that leads to major genome rearrangements (see (Hastings et al., 2009; Maréchal and Brisson, 2010) for reviews). About 5% of the *why1why3* plants also exhibit a variegated leaf phenotype that correlates with selectively amplified genomic regions (Maréchal et al., 2009). As the results in Figure 26E show that the amount of DSBs is increased in ptDNA from ciprofloxacin-treated *pollb-1* plants, it suggests that the severe phenotype of the *pollb-1/why1why3* plants is also linked to an accumulation of rearranged DNA products in the plastid genome. This was tested by monitoring the extent of genome rearrangements in the mutant lines using eight primer pairs employed in our previous study to survey ptDNA rearrangements in *Whirly* mutant plants (Cappadocia et al., 2010). These primers are either in an opposite orientation or placed far apart (more than 5 kb) from each other (Fig. 27C) and amplify respectively junctions of duplicated/circular or deleted DNA regions produced by the MMBIR pathway in plastids. Figure 27B shows the results of representative PCR reactions with three primer pairs. A graphic representation of the main rearranged molecule is illustrated for each reaction in Figure 27D. Untreated WT and *pollb-1* plants show no detectable rearrangements, while the same genotypes treated with 0.25 μ M ciprofloxacin show only few rearrangements (Fig. 27B). The *why1why3* plants contain a constitutive level of rearranged DNA molecules and, as reported before (Cappadocia et al., 2010), this level increases noticeably after treatment with 0.25 μ M ciprofloxacin. The level of DNA rearrangements is however markedly higher in the untreated triple mutant as compared to

untreated *why1why3* plants. This level is actually similar to that of the treated *why1why3* plants and appears to further increase upon treatment with the drug. Therefore, the severe phenotype of the *polb1-1/why1why3* plants correlates with increased rearrangements of the plastid genome.

3.5 DISCUSSION

An Essential Role for the *POLI* Genes in Higher Plants

Our understanding of plant organelle DNA replication is still very incomplete (Bendich, 2007; Nielsen et al., 2010). This is largely due to our ignorance concerning some of the most basic factors, such as DNA polymerases, that contribute to this fundamental process in plants. Many studies have suggested that the PolII-like DNA polymerases found in higher plants are responsible for most if not all the DNA-synthesis activity found in mitochondria and plastids (Kimura et al., 2002; Mori et al., 2005; Ono et al., 2007). This is based on the biochemical properties of the recombinant proteins, which indicate that these enzymes have sufficient processivity to be able to replicate both genomes. In addition, the measured *in vitro* characteristics of these enzymes correspond almost perfectly to those measured *ex vivo* from the purified DNA-synthesis activity of mitochondria and plastids (Spencer and Whitfeld, 1969; Sala et al., 1980; McKown and Tewari, 1984; Heinhorst et al., 1990; Meissner et al., 1993). All these elements, plus the fact that they are present in both organelles (Elo et al., 2003; Christensen et al., 2005; Mori et al., 2005; Ono et al., 2007; Wamboldt et al., 2009), suggest that the PolII DNA polymerases are responsible for DNA replication in plant organelles.

In the present study, we bring further arguments to support a role for PolII proteins in the replication of plastid and mitochondria DNA. Using Arabidopsis mutant plants for the *PolIA* and *PolIB* genes, we show that concurrent mutation of the two genes is lethal, indicating that they share an essential function in plants. Furthermore, each gene is able to partially compensate for the function of the other gene as all single *PolII* mutants tested are viable. However, the fact that the *polib-2* line displays a growth retardation phenotype suggests that *PolIA* cannot entirely compensate for the function of *PolIB*, or that *PolIB* carries another function not shared by *PolIA*.

A role for the PolI proteins in DNA replication is also supported by our results showing that, as compared to WT plants, the relative level of organelle DNA is reduced in plastids and mitochondria of both *polla-1* and *pollb-1* plants. The reduction is most pronounced at 3 days post-germination, which is consistent with expression data showing that the *PolI* genes are most highly expressed in young, meristematic tissues (Kimura et al., 2002; Mori et al., 2005; Ono et al., 2007). These results are also in line with previous studies showing that DNA replication in organelles is highly active in 3 day old Arabidopsis seedlings (Fujie et al., 1993; Fujie et al., 1994). Our results therefore provide a functional link between the *PolI* genes and DNA accumulation inside plant mitochondria and plastids.

Interestingly, the ratios of organelle to nuclear DNA appear to decrease in WT plants between day 3 and 5 (Fig. 25 and Fig. 28). This could either mean that nuclear DNA ploidy is increasing or that organelle DNA is less abundant in cells of older seedlings. However, since WT and mutant plants of the same age are being compared, any variations in nuclear or organellar DNA levels should not affect our results. It is interesting that previous studies reported a similar trend for ptDNA (Zoschke et al., 2007) and mtDNA (Preuten et al., 2010) ratios in aging cotyledons of Arabidopsis, although different samples were used in these studies (cotyledon alone vs. whole seedling with roots).

A Specific Role for POLIB in DNA DSB Repair

DNA repair activity is essential for correcting DNA damage caused by exogenous stress or that occur naturally during replication. Many mechanisms used to repair DNA, such as HR and MMBIR, involve a DNA-synthesis step. Since, as mentioned above, PolIA does not fully compensate for the loss of PolIB in the *pollb-2* line, there is a possibility that PolIB could be involved in an additional process, such as synthesis of DNA during repair. This was tested by inducing DSBs in the organelle genome of all the *PolI* mutants and by measuring the effect of this stress on plant phenotype and on the level of DSBs. The fact that the growth of the *pollb* lines, but not of the *polla* lines, was more affected than WT plants by the genotoxic stress provides a first evidence for a role of PolIB in DNA repair. Similar increased sensitivity to ciprofloxacin was also reported for Whirly (Cappadocia et al., 2010) and cpRecA (Rowan et al., 2010) mutants, which are both implicated in the maintenance of plastid genome stability. A

role in DNA repair for PolIB is further supported by the observation that the *pollb-1* line accumulates higher levels of DNA lesions than the *polla-1* line or WT plants when treated with ciprofloxacin.

The higher levels of DNA lesions in plastids compared to mitochondria suggest a specialized role for PolIB in this organelle. However, it could also simply reflect the presence of alternate repair mechanisms more active in the mitochondria. For example homologous recombination between short repeats (>100 bp and <1000 bp, reviewed in (Maréchal and Brisson, 2010)), a process very active in plant mitochondria, would provide efficient means for the repair of DSBs. Also, we cannot exclude the possibility that ciprofloxacin could be less active in mitochondria. The different response of the *polla* and *pollb* mutant lines to the genotoxic stress is surprising, considering that the two PolI proteins have over 72% identity at the amino acid sequence level (Mori et al., 2005) and that they are nearly identical regarding their 3'-5' exonuclease and DNA polymerase domains. Indeed, there does not seem to be any distinctive motif between PolIA and PolIB when comparing available amino acid sequences of PolI proteins from various plant species. Therefore our results indicate that Arabidopsis PolIB has either evolved a specific repair activity or that it is specifically recruited to damage sites. This additional function in DNA repair might also explain why one of the *PolIB* mutants shows a growth retardation phenotype while both *PolIA* mutants are undistinguishable from WT plants. However, we cannot exclude that PolIA is involved in some other DNA repair mechanisms that do not involve DSBs.

A Genetic Interaction between *PolIB* and Plastid *Whirly* Genes

Previous studies have proposed that Whirly proteins play a role in organelle DNA replication (Maréchal et al., 2008; Prikryl et al., 2008). More recently, these proteins were shown to be involved in the maintenance of plastid genome stability (Maréchal et al., 2009) and the repair of DSBs in plastids and mitochondria (Cappadocia et al., 2010). It is therefore interesting that a strong synergistic interaction exists between *PolIB* and the plastid *Whirly* genes. This suggests that PolIB and the plastid Whirlies function in separate, but complementary pathways. The large increase of MMBIR rearrangements in ptDNA observed in the *pollb-1/why1why3* mutant in the absence of genotoxic stress supports this conclusion

and indicates that PolIB is not required or essential for DNA repair by MMBIR. This is reminiscent of the situation in yeast where break-induced replication (BIR) as well as MMBIR are dependent on a nonessential subunit of Pol δ called POL32 (Lydeard et al., 2007; Payen et al., 2008) and in *E. coli* where it is dependent on DinB (Galhardo et al., 2009). Our results suggest that, in the absence of PolIB, conservative repair of DSBs is reduced, leading to the accumulation of substrates (DSBs) that are repaired by the error prone MMBIR in absence of Whirlies. It also indicates that another DNA polymerase is responsible for the synthesis of rearranged DNA molecules, at least regarding the MMBIR process.

The fact that no phenotype is observed when *polIb-1* is crossed with the mitochondrial Whirly mutant allele *why2-1* supports once again the presence of an alternative DNA repair mechanism in this organelle. This is in agreement with our previous results where no visible or molecular phenotype were observed in the *why2-1* mutant unless an exogenous stress was applied (Cappadocia et al., 2010). This contrast with the situation with the *why1why3* plastid Whirly mutant where all the plants showed DNA rearrangements and 4.6% were variegated (Maréchal et al., 2009).

Taken together, our results indicate that the action of PolIB in DNA repair most likely involves a classical repair mechanism, such as HR. It is thus expected that reduced expression or inactivation of *PolIB* in Arabidopsis would lead to an increase in unattended DSBs and/or to slower repairs, even in the absence of exogenous stress. These breaks or arrested replication forks would be more likely to be corrected by other mechanisms, such as break-induced replication. In the absence of Whirly proteins, however, the breaks would also be repaired through error-prone mechanisms such as MMBIR, leading to the accumulation of rearranged DNA molecules such as those observed in *polIb-1/why1why3* and that could account for the severe phenotype of this line. It will be interesting to confirm this model in the future by further genetic crosses between the *polIb* alleles and other mutant alleles for genes involved in organelle DNA repair.

3.6 MATERIAL AND METHODS

Plant Material and Growth Conditions

Arabidopsis thaliana (ecotype Columbia-0) mutant lines *polla-1* (SALK_022638), *polla-2* (SALK_065221), *pollb-1* (SALK_134274) and *pollb-2* (WiscDsLoxHs021_09D) were obtained from the Arabidopsis Biological Resource Center (ABRC). Seeds were either sown on soil or sterilized and placed on petri dishes containing Murashige and Skoog (MS) basal media (Sigma-Aldrich, www.sigmaaldrich.com) supplemented with 1 % sucrose and 0.8 % agar with or without ciprofloxacin (Sigma-Aldrich). After three days of vernalization at 4°C, the seeds were placed under light (100 $\mu\text{mol m}^{-2} \text{s}^{-1}$, 16 h day/8 h dark) at an average temperature of 22°C. Whole plants grown on MS medium were harvested at 3, 4, 5 (for seedlings), 14 (5 to 6 true leaves) or 21 (mature plants) days after germination by pulling the plants (including the roots) carefully from the agar and instantly frozen in liquid nitrogen. Plants were pooled to make 75 mg samples.

Nucleic Acids Isolation

DNA was isolated using a cetyl trimethylammonium bromide (CTAB) protocol (Weigel and Glazebrook, 2002). RNA was isolated using TRIzol® Reagent (Invitrogen, www.invitrogen.com) according to the manufacturer instructions. RNA pellets were resuspended in 30 μL of nuclease-free water and kept at -80°C. cDNA was synthesized using the First Strand cDNA Synthesis kit from Fermentas (www.fermentas.com) according to the manufacturer instructions. cDNA reactions were diluted 10 fold and kept at -20°C.

Quantitative PCR

Real-time qPCR reactions were done using the SYBR® Green Master Mix (SABiosciences, www.sabiosciences.com) according to the manufacturer instructions. Primers used for qRT-PCR experiments are as follow : 5'Tubu-qRT, CCT TCT CGG TGT TCC CTT CAC C, 3'Tubu-qRT, GGG AAC CTC AGA CAG CAA GTC AC, 5'PolIA-qRTa, ACC GCG GTC GGG CTT TCT AGA, 3'PolIA-qRTa, GCA CGT GAC TTT GAC GCC GGA, 5'PolIB-qRTb, GCG GAG ATC GCA AAG GAC ATA GTT GT, 3'PolIB-qRTb, TCC TTC AAC TAC GGC CGG TCT CT, 5'PolIB-qRTc, GAC GCT TAT CGG CTA GAA GGC CA and 3'PolIB-qRTc, CTT GAG TGG AAG TCT CCA CCA GCT. Primers used for qPCR experiments are as follow : 5'ncDNA-qPCR, GTT GAA GCC TCC GTT CCC TGC TA, 3'ncDNA-qPCR, CTC TTC CAC CGT GCA TGG CTT GT, 5'mtDNA-qPCR, CCT GAT

TCT GCG CGT AGA AGA CCT , 3'mtDNA-qPCR, AGG CGT AAG CGC AGC AGT TAG A, 5'ptDNA-qPCR, CCC TCT CTC TCG TAG TGT GGG GAA and 3'ptDNA-qPCR, TCG AAA GGG GTT ACC CCA TGA ATG G. The amplification efficiency of each reaction was determined using dilution of template cDNA or DNA. Calibration runs determined the use of 1/10 dilution for cDNA and 1/100 for DNA. Melting curves confirmed the presence of a unique PCR product in each case. Amplification was carried out by using the LightCycler®480 (Roche, www.roche-applied-science.com) and all the data were analyzed using the LightCycler®480 Software version 1.5.

Long-Amplification PCR

DNA lesions were quantified using the technique described in Yakes and Van Houten (1997). Briefly, DNA was amplified using GeneAmp® XL PCR kit (Applied Biosystems, www.appliedbiosystems.com) following the manufacturer recommendations. For mitochondria, plastids and the nucleus we amplified a long (about 20 kb) and a short (about 200 bp) PCR fragment. It is expected that the long amplification will be more susceptible to breaks than the short amplification. The primers used for ptDNA were as follow: 5'ptLongAmp, CCG GGA CTC GAA CCC GGA ACT A, 3'ptLongAmp, CTT CCT CCC GAG TTG AGA CCC A, 5'ptShortAmp, CTC CAG TGC ATT TCG CCC TCT GA and 3'ptShortAmp, ACG TAC ACG TCA CGG GCA TCC T. The primers used for mtDNA were as follow: 5'mtLongAmp, CCT ATC AGT AGC TAG GCC CTG GTC C, 3'mtLongAmp, CCT AAC TAC ACC GAC CTC CGA GTC C, 5'mtShortAmp, CCT GAT TCT GCG CGT AGA AGA CCT and 3'mtShortAmp, AGG CGT AAG CGC AGC AGT TAG A. The primers used for ncDNA were as follow: 5'ncLongAmp, TCG CGG AGA TCT TTC TCT CGC AC, 3'ncLongAmp, CTG CCT CCA CCA ACT CGT TCA C, 5'ncShortAmp, GTT GAA GCC TCC GTT CCC TGC TA and 3'ncShortAmp, CTC TTC CAC CGT GCA TGG CTT GT. DNA quantification of the reactions was done using PicoGreen® Reagent (Invitrogen) following the manufacturer instructions. Fluorescence was measured using a Perkin Elmer/Packard Fusion Alpha-FP Microplate Analyzer (PerkinElmer, www.perkinelmer.com). The level of PCR products for the long amplification was reported on the level of the short product for each sample and normalized to 100 with WT untreated signal.

Total DNA from MS medium-grown 21 day old plants was used to survey ptDNA rearrangements. DNA was equilibrated with a low-cycle PCR amplification of a Ycf2 DNA fragment. The PCR strategy and primers used were described in Cappadocia et al. (2010). PCR reactions products were migrated on ethidium bromide stained agarose gels.

3.6 Supplemental Data

The following materials are available on the online version of this article.

Figure 28. Organelle DNA is less abundant in seedlings of polla-2.

Figure 29. Ciprofloxacin does not generate DSB in the nuclear genome.

Figure 30. Crosses between Poll and Whirly mutants.

3.7 ACKNOWLEDGMENTS

The assistance of Alexandre Maréchal in isolating and crossing mutant lines is gratefully acknowledged. We thank B. Franz Lang and all the members of the Brisson's laboratory for valuable discussions.

3.8 LITERATURE CITED

- Bendich AJ** (2007) The size and form of chromosomes are constant in the nucleus, but highly variable in bacteria, mitochondria and chloroplasts. *Bioessays* **29**: 474-483
- Cappadocia L, Maréchal A, Parent JS, Lepage E, Sygusch J, Brisson N** (2010) Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. *Plant Cell* **22**: 1849-1867
- Castroviejo M, Tharaud D, Tarrago-Litvak L, Litvak S** (1979) Multiple deoxyribonucleic acid polymerases from quiescent wheat embryos. Purification and characterization of three

- enzymes from the soluble cytoplasm and one from purified mitochondria. *Biochem J* **181**: 183-191
- Christensen AC, Lyznik A, Mohammed S, Elowsky CG, Elo A, Yule R, Mackenzie SA** (2005) Dual-domain, dual-targeting organellar protein presequences in *Arabidopsis* can use non-AUG start codons. *Plant Cell* **17**: 2805-2816
- Elo A, Lyznik A, Gonzalez DO, Kachman SD, Mackenzie SA** (2003) Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the *Arabidopsis* genome. *Plant Cell* **15**: 1619-1631
- Fujie M, Kuroiwa H, Kawano S, Mutoh S, Kuroiwa T** (1994) Behavior of organelles and their nucleoids in the shoot apical meristem during leaf development in *Arabidopsis thaliana* L. *Planta* **194**: 395-405
- Fujie M, Kuroiwa H, Suzuki T, Kawano S, Kuroiwa T** (1993) Organelle DNA synthesis in the quiescent center of *Arabidopsis thaliana* (col.). *Journal of Experimental Botany* **44**: 689-693
- Galhardo RS, Do R, Yamada M, Friedberg EC, Hastings PJ, Nohmi T, Rosenberg SM** (2009) DinB upregulation is the sole role of the SOS response in stress-induced mutagenesis in *Escherichia coli*. *Genetics* **182**: 55-68
- 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**: e1000327
- Heinhorst S, Cannon GC, Weissbach A** (1990) Chloroplast and mitochondrial DNA polymerases from cultured soybean cells. *Plant Physiol* **92**: 939-945
- Keim CA, Mosbaugh DW** (1991) Identification and characterization of a 3' to 5' exonuclease associated with spinach chloroplast DNA polymerase. *Biochemistry* **30**: 11109-11118
- Kimura S, Uchiyama Y, Kasai N, Namekawa S, Saotome A, Ueda T, Ando T, Ishibashi T, Oshige M, Furukawa T, Yamamoto T, Hashimoto J, Sakaguchi K** (2002) A novel DNA polymerase homologous to *Escherichia coli* DNA polymerase I from a higher plant, rice (*Oryza sativa* L.). *Nucleic Acids Res* **30**: 1585-1592
- Lydeard JR, Jain S, Yamaguchi M, Haber JE** (2007) Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **448**: 820-823
- Maréchal A, Brisson N** (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytol* **186**: 299-317
- Maréchal A, Parent JS, Sabar M, Véronneau-Lafortune F, Abou-Rached C, Brisson N** (2008) Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol* **8**: 42
- Maréchal A, Parent JS, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N** (2009) Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc Natl Acad Sci U S A* **106**: 14693-14698
- McKown RL, Tewari KK** (1984) Purification and properties of a pea chloroplast DNA polymerase. *Proc Natl Acad Sci U S A* **81**: 2354-2358
- Meissner K, Heinhorst S, Cannon GC, Börner T** (1993) Purification and characterization of a gamma-like DNA polymerase from *Chenopodium album* L. *Nucleic Acids Res* **21**: 4893-4899
- Mori Y, Kimura S, Saotome A, Kasai N, Sakaguchi N, Uchiyama Y, Ishibashi T, Yamamoto T, Chiku H, Sakaguchi K** (2005) Plastid DNA polymerases from higher plants, *Arabidopsis thaliana*. *Biochem Biophys Res Commun* **334**: 43-50

- Nielsen BL, Cupp JD, Brammer J** (2010) Mechanisms for maintenance, replication, and repair of the chloroplast genome in plants. *J Exp Bot* **61**: 2535-2537
- Ono Y, Sakai A, Takechi K, Takio S, Takusagawa M, Takano H** (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**: 1679-1692
- 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**: e1000175
- Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Borner T** (2010) Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *Plant J* **64**: 948-959
- 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**: 5152-5165
- Rowan BA, Oldenburg DJ, Bendich AJ** (2009) A multiple-method approach reveals a declining amount of chloroplast DNA during development in Arabidopsis. *BMC Plant Biol* **9**: 3
- Rowan BA, Oldenburg DJ, Bendich AJ** (2010) RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. *J Exp Bot* **61**: 2575-2588
- Sakai A, Suzuki T, Nagata N, Sasaki N, Miyazawa Y, Saito C, Inada N, Nishimura Y, Kuroiwa T** (1999) Comparative analysis of DNA synthesis activity in plastid-nuclei and mitochondrial-nuclei simultaneously isolated from cultured tobacco cells. *Plant Science* **140**: 9-19
- Sala F, Amileni AR, Parisi B, Spadari S** (1980) A gamma-like DNA polymerase in spinach chloroplasts. *Eur J Biochem* **112**: 211-217
- 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**: 289-318
- Spencer D, Whitfeld PR** (1969) The characteristics of spinach chloroplast DNA polymerase. *Arch Biochem Biophys* **132**: 477-488
- 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**: 7821-7826
- Wamboldt Y, Mohammed S, Elowsky C, Wittgren C, de Paula WB, Mackenzie SA** (2009) Participation of leaky ribosome scanning in protein dual targeting by alternative translation initiation in higher plants. *Plant Cell* **21**: 157-167
- Weigel D, Glazebrook J** (2002) *Arabidopsis*: a laboratory manual. Cold Spring Harbor Laboratory Press
- Yakes FM, Van Houten B** (1997) Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. *Proc Natl Acad Sci U S A* **94**: 514-519
- Zoschke R, Liere K, Borner 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**: 710-722

3.9 FIGURES LEGENDS

Figure 24. Isolation of *PollA* and *PollB* mutant lines. A, Schematic representation of *PollA* and *PollB* genes. Exons are denoted by blue boxes, the 3'UTR by a red box and the 5'UTR by a white box. The pale blue lines represent introns. The position of the T-DNA insertions is indicated by black triangles. Half-arrows stand for the primers used for the qRT-PCR experiments. B, Histogram presenting the results of the qRT-PCR experiments measuring the level of expression of the *PollA* gene inside the *polla* mutant plants grown on MS medium for 14 days. Expression levels are represented as percentages of WT level. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicates. C, Histogram presenting the results of the qRT-PCR experiments measuring the level of expression of the *PollB* gene inside the *pollb* mutant plants grown on MS medium for 14 days. Expression levels are represented as percentages of WT level. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicates. D, Representative photographs of four week old Arabidopsis plants for WT and *Poll* mutant lines.

Figure 25. Organelle DNA is less abundant in seedlings of *polla-1* and *pollb-1* mutants. A, Histogram presenting the results of the qPCR analysis of mitochondrial genome copy number relative to nuclear genome copy number for 3, 4 and 5 day old seedlings. B, Histogram presenting the results of the qPCR analysis of plastid genome copy number relative to nuclear genome copy number for 3, 4 and 5 day old seedlings. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicate.

Figure 26. *PollB* mutants exhibit an increased sensitivity to ciprofloxacin (CIP) and are impaired in their ability to process DNA DSBs. A, Representative photographs of the different *poll* mutants grown for 21 days on MS medium containing 0, 0.25 or 0.5 μ M of ciprofloxacin. B, Representative photographs of the first true leaves of *pollb-1* and *pollb-2* plants grown for 21 days on MS medium containing 0.25 μ M of ciprofloxacin. C, Histogram representing the percentage of plants with white/variegated first true leaves after 21 days of growth on medium with different concentrations of ciprofloxacin. Error bars represent the standard deviation of the counts made on different plates. D, Histogram representing the percentage of plants without any first true leaves after 21 days of growth on medium with different concentrations of ciprofloxacin. Error bars represent the standard deviation of the counts made on different plates. E, Histogram presenting the results of the PCR experiments to evaluate the abundance of DNA lesions upon ciprofloxacin treatment in plastid DNA. DNA lesions are significantly more abundant for the *pollb-1* mutant when treated with both 0.25 and 0.5 μ M of ciprofloxacin. Error bars represent the standard error of the mean of three separate experiments where WT signal at 0 μ M of ciprofloxacin is expressed as 100 for each repetition.

One asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test and two asterisks, a difference with a $p < 0.01$. F, Histogram presenting the results of the PCR experiments to evaluate the abundance of DNA lesions in mitochondrial DNA upon treatment with ciprofloxacin. DNA lesions are significantly more abundant for the *pollb-1* mutant when treated with $0.5 \mu\text{M}$ of ciprofloxacin. Error bars represent the standard error of the mean of three separate experiments where WT signal at $0 \mu\text{M}$ of ciprofloxacin is expressed as 100 for each repetition. Asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test.

Figure 27. A *pollb-1/why1why3* triple mutant yields a pale-green dwarf phenotype and gives rise to abundant DNA rearrangements in plastids. A, Representative photographs of four week old plants for the genotypes used in the cross and the triple mutant progeny compared to WT. B, Representative PCR reactions to evaluate the abundance of DNA rearrangements dependant on microhomology in plastids of the plants presented in A. The plants were grown for 21 days on MS medium containing 0 or $0.25 \mu\text{M}$ of ciprofloxacin. Two DNA samples were tested for each condition. PCR reactions were done using the primers indicated on the right of the gels. C, Schematic representation of DNA rearrangements evaluated by the PCR approach. Inward-facing primers placed far apart will yield a PCR product only if a deletion has occurred in the region that separates the primers. Outward-facing primers will yield a PCR product only if the template DNA is duplicated or circularized. Primers are represented by black half-arrows. D, Schematic representation of the DNA rearrangements amplified by PCR. The upper panel shows the position of the primers used to test the plastid genome of Arabidopsis. The lower panel gives a graphical representation of the rearranged DNA molecules. The sequences in the white boxes are the microhomologous sequences that were used to prime the MMBIR process. The original position of the microhomologous sequences in WT ptDNA is indicated.

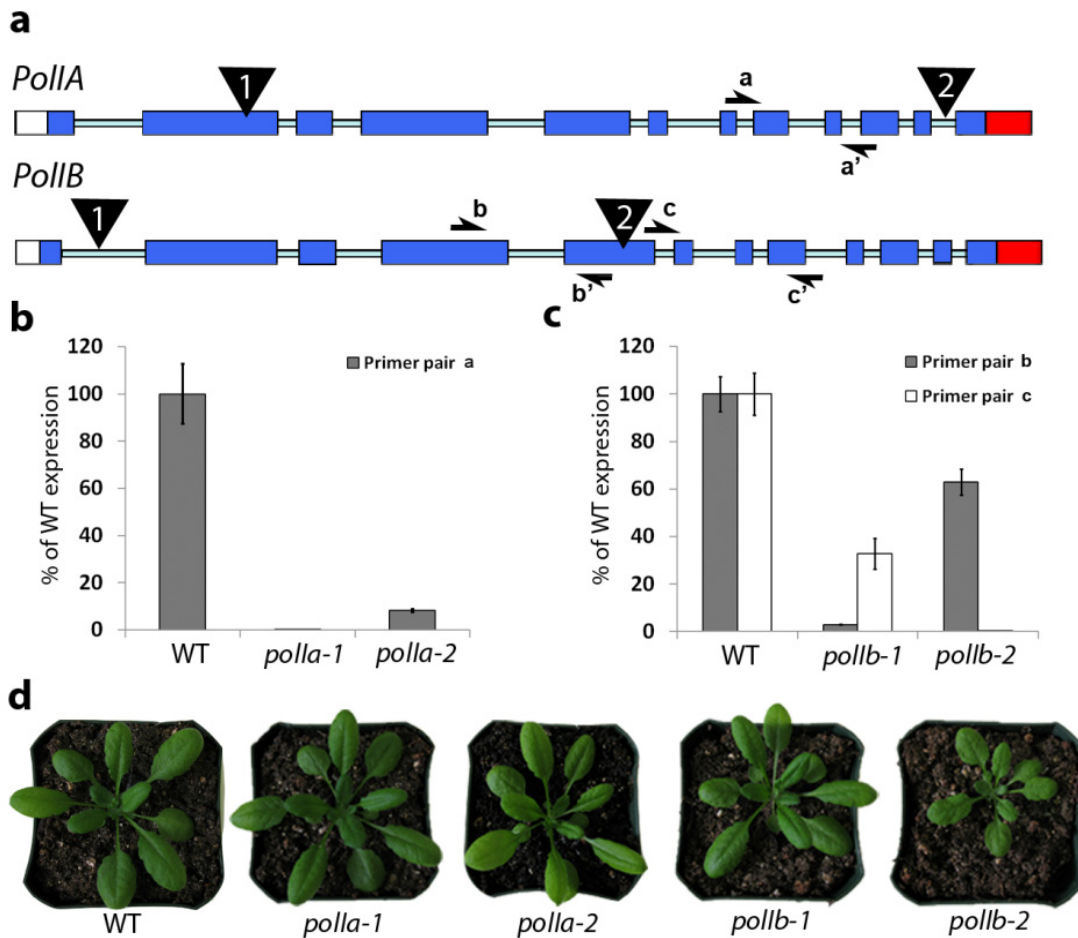


Figure 24. Isolation of *PollIA* and *PollIB* mutant lines. A, Schematic representation of *PollIA* and *PollIB* genes. Exons are denoted by blue boxes, the 3'UTR by a red box and the 5'UTR by a white box. The pale blue lines represent introns. The position of the T-DNA insertions is indicated by black triangles. Half-arrows stand for the primers used for the qRT-PCR experiments. B, Histogram presenting the results of the qRT-PCR experiments measuring the level of expression of the *PollIA* gene inside the *polla* mutant plants grown on MS medium for 14 days. Expression levels are represented as percentages of WT level. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicates. C, Histogram presenting the results of the qRT-PCR experiments measuring the level of expression of the *PollIB* gene inside the *pollb* mutant plants grown on MS medium for 14 days. Expression levels are represented as percentages of WT level. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicates. D, Representative photographs of four week old *Arabidopsis* plants for WT and *Poll* mutant lines.

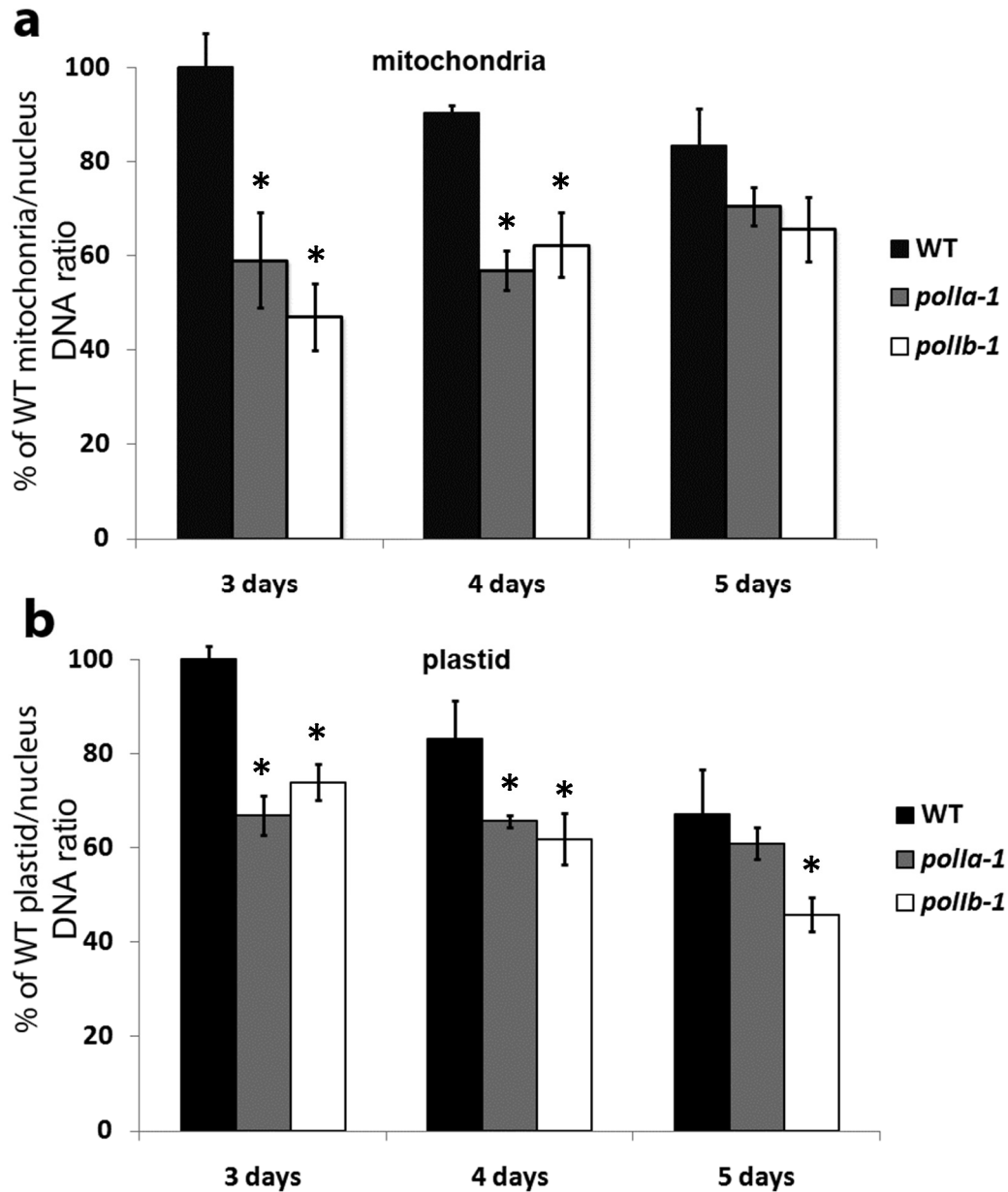


Figure 25. Organelle DNA is less abundant in seedlings of *polla-1* and *pollb-1* mutants. A, Histogram presenting the results of the qPCR analysis of mitochondrial genome copy number relative to nuclear genome copy number for 3, 4 and 5 day old seedlings. One asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test when compared to the WT. B, Histogram presenting the results of the qPCR analysis of plastid genome copy number relative to nuclear genome copy number for 3, 4 and 5 day old seedlings. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicate. One asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test when compared to the WT.

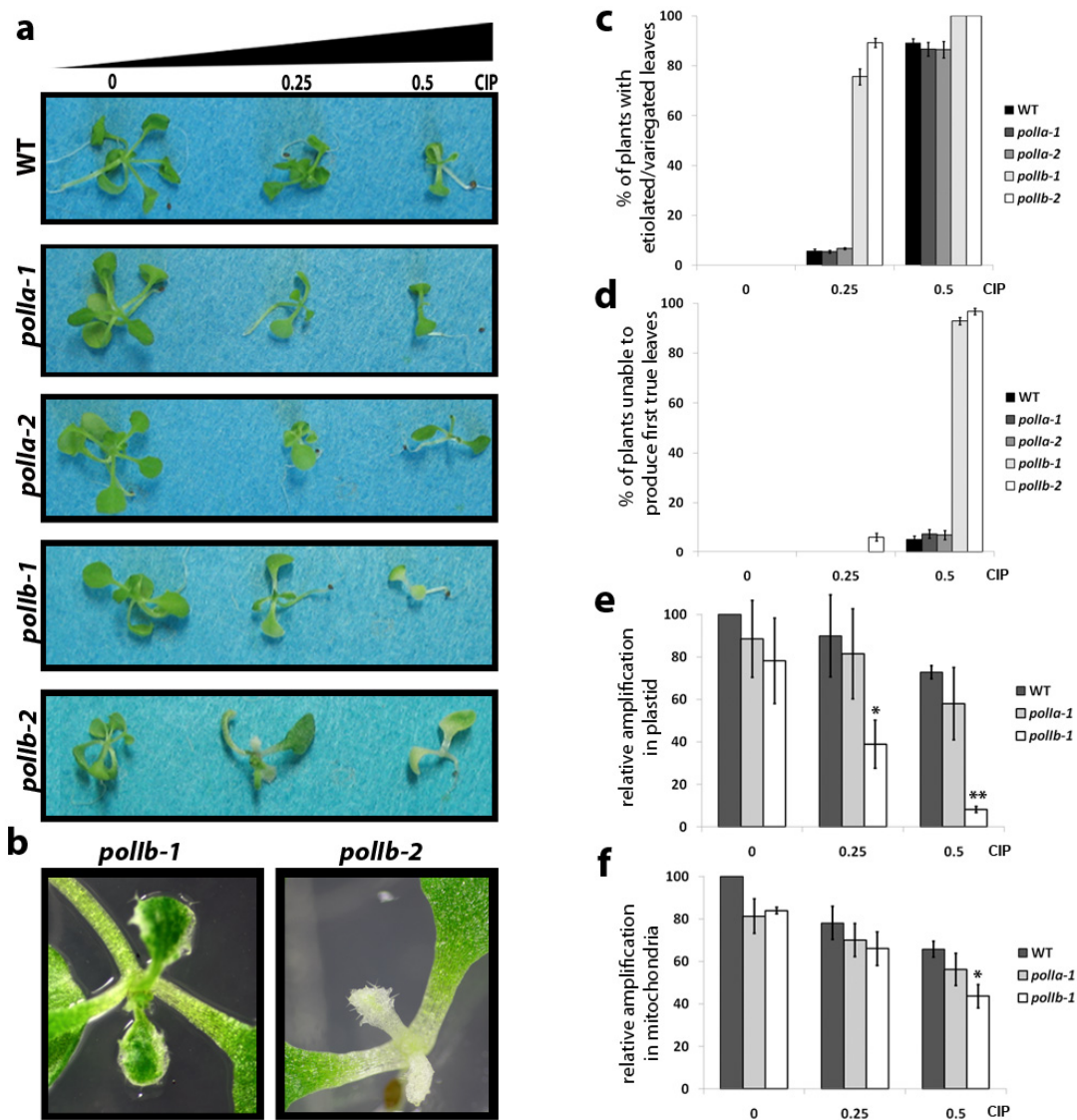


Figure 26. PolIB mutants exhibit an increased sensitivity to ciprofloxacin (CIP) and are impaired in their ability to process DNA DSBs. A, Representative photographs of the different *poll* mutants grown for 21 days on MS medium containing 0, 0.25 or 0.5 μ M of ciprofloxacin. B, Representative photographs of the first true leaves of *pollb-1* and *pollb-2* plants grown for 21 days on MS medium containing 0.25 μ M of ciprofloxacin. C, Histogram representing the percentage of plants with white/variegated first true leaves after 21 days of growth on medium with different concentrations of ciprofloxacin. Error bars represent the standard deviation of the counts made on different plates. D, Histogram representing the percentage of plants without any first true leaves after 21 days of growth on medium with different concentrations of

ciprofloxacin. Error bars represent the standard deviation of the counts made on different plates. E, Histogram presenting the results of the PCR experiments to evaluate the abundance of DNA lesions upon ciprofloxacin treatment in plastid DNA. DNA lesions are significantly more abundant for the *pollb-1* mutant when treated with both 0.25 and 0.5 μM of ciprofloxacin. Error bars represent the standard error of the mean of three separate experiments where WT signal at 0 μM of ciprofloxacin is expressed as 100 for each repetition. One asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test and two asterisks, a difference with a $p < 0.01$. F, Histogram presenting the results of the PCR experiments to evaluate the abundance of DNA lesions in mitochondrial DNA upon treatment with ciprofloxacin. DNA lesions are significantly more abundant for the *pollb-1* mutant when treated with 0.5 μM of ciprofloxacin. Error bars represent the standard error of the mean of three separate experiments where WT signal at 0 μM of ciprofloxacin is expressed as 100 for each repetition. Asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test.

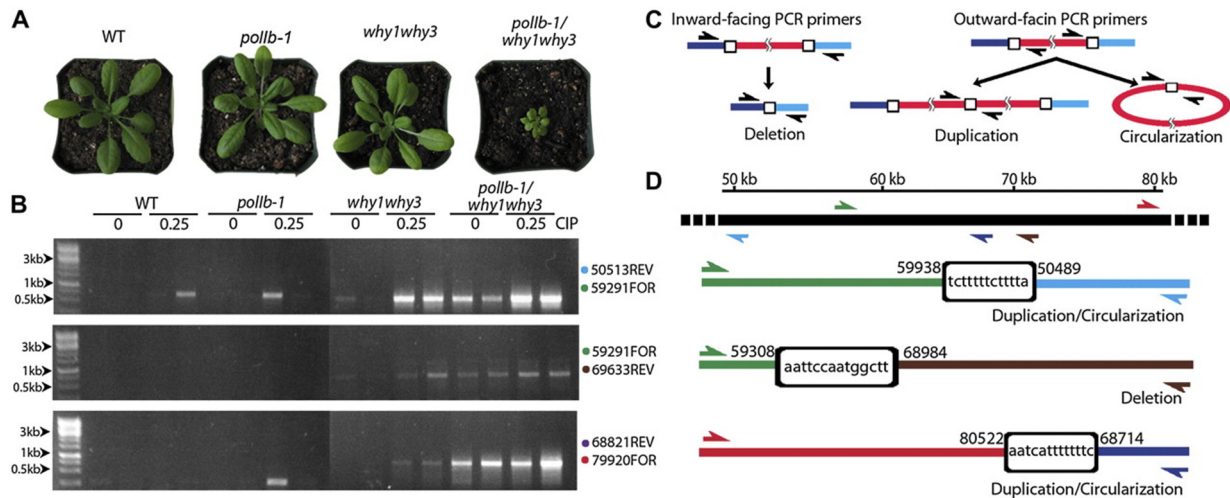


Figure 27. A *pollb-1/why1why3* triple mutant yields a pale-green dwarf phenotype and gives rise to abundant DNA rearrangements in plastids. A, Representative photographs of four week old plants for the genotypes used in the cross and the triple mutant progeny compared to WT. B, Representative PCR reactions to evaluate the abundance of DNA rearrangements dependant on microhomology in plastids of the plants presented in A. The plants were grown for 21 days on MS medium containing 0 or 0.25 μ M of ciprofloxacin. Two DNA samples were tested for each condition. PCR reactions were done using the primers indicated on the right of the gels. C, Schematic representation of DNA rearrangements evaluated by the PCR approach. Inward-facing primers placed far apart will yield a PCR product only if a deletion has occurred in the region that separates the primers. Outward-facing primers will yield a PCR product only if the template DNA is duplicated or circularized. Primers are represented by black half-arrows. D, Schematic representation of the DNA rearrangements amplified by PCR. The upper panel shows the position of the primers used to test the plastid genome of Arabidopsis. The lower panel gives a graphical representation of the rearranged DNA molecules. The sequences in the white boxes are the microhomologous sequences that were used to prime the MMBIR process. The original position of the microhomologous sequences in WT ptDNA is indicated.

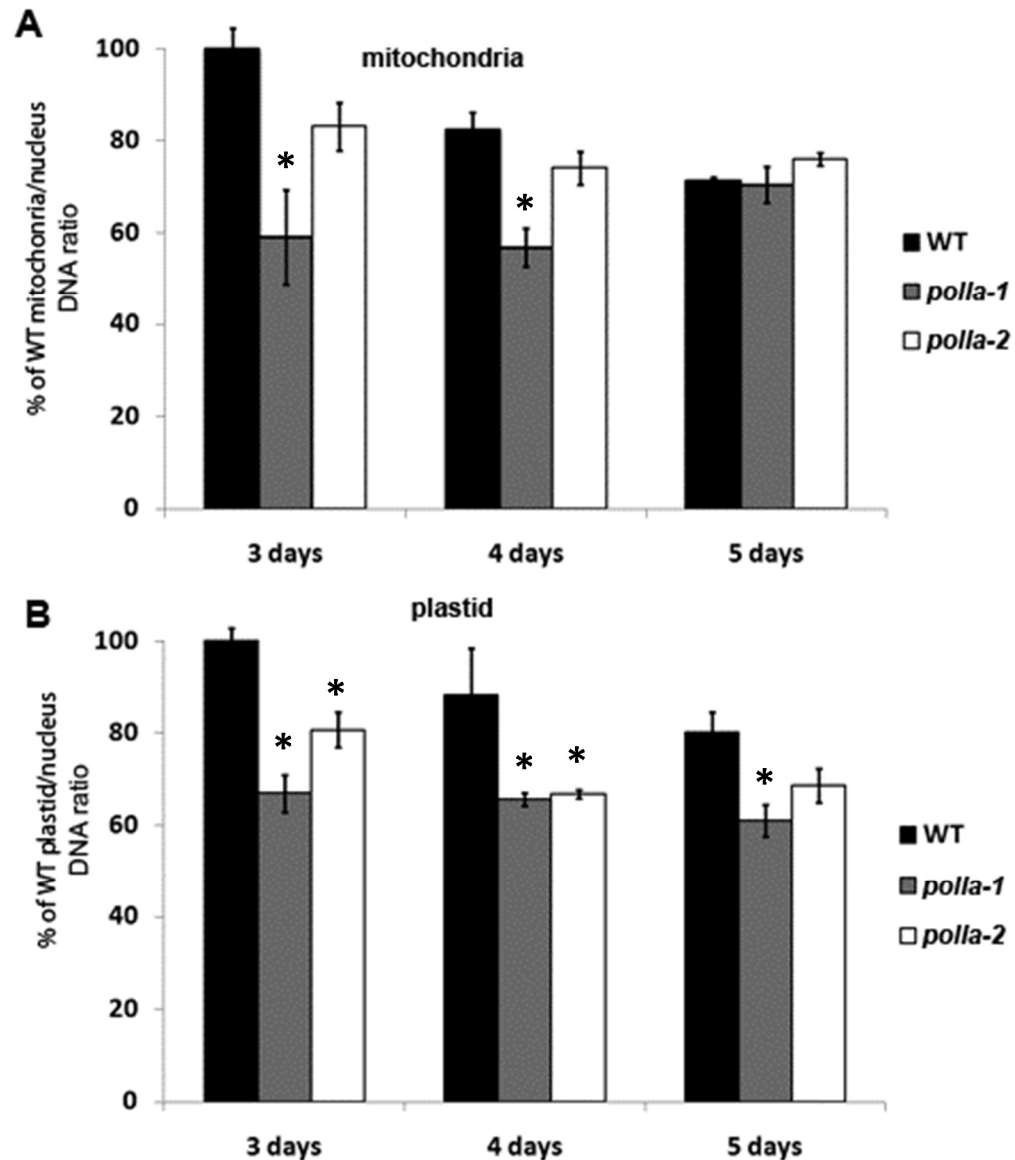


Figure 28. Organelle DNA is less abundant in seedlings of *polla-2*. A, Histogram presenting the results of the qPCR analysis of mitochondrial genome copy number relative to nuclear genome copy number for 3, 4 and 5 day old seedlings. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicate. The results for the *polla-1* line (Fig. 2) are also presented. One asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test when compared to the WT. B, Histogram presenting the results of the qPCR analysis of plastid genome copy number relative to nuclear genome copy number for 3, 4 and 5 day old seedlings. The experiments were done in technical and biological triplicates. The error bars represent the standard error of the mean on the biological triplicate. The results for the *polla-1* line (Fig. 2) are also presented. One asterisk indicates a significant difference with a $p < 0.05$ in the Student t-test when compared to the WT.

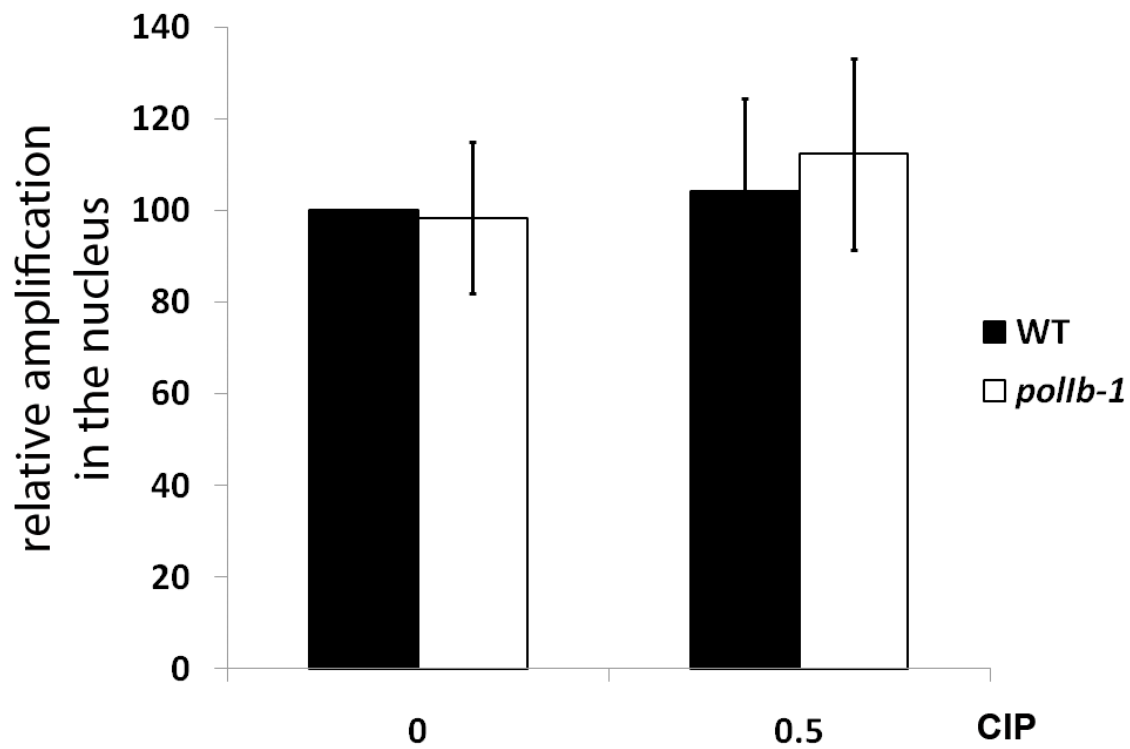


Figure 29. Ciprofloxacin does not generate DSBs in the nuclear genome. Histogram presenting the relative amplification of a long nuclear DNA fragment and a short nuclear DNA fragment for WT and *pollb-1* plants. Error bars represent the standard error of the mean of three separate experiments where WT signal at 0 μM of ciprofloxacin is expressed as 100 for each repetition.

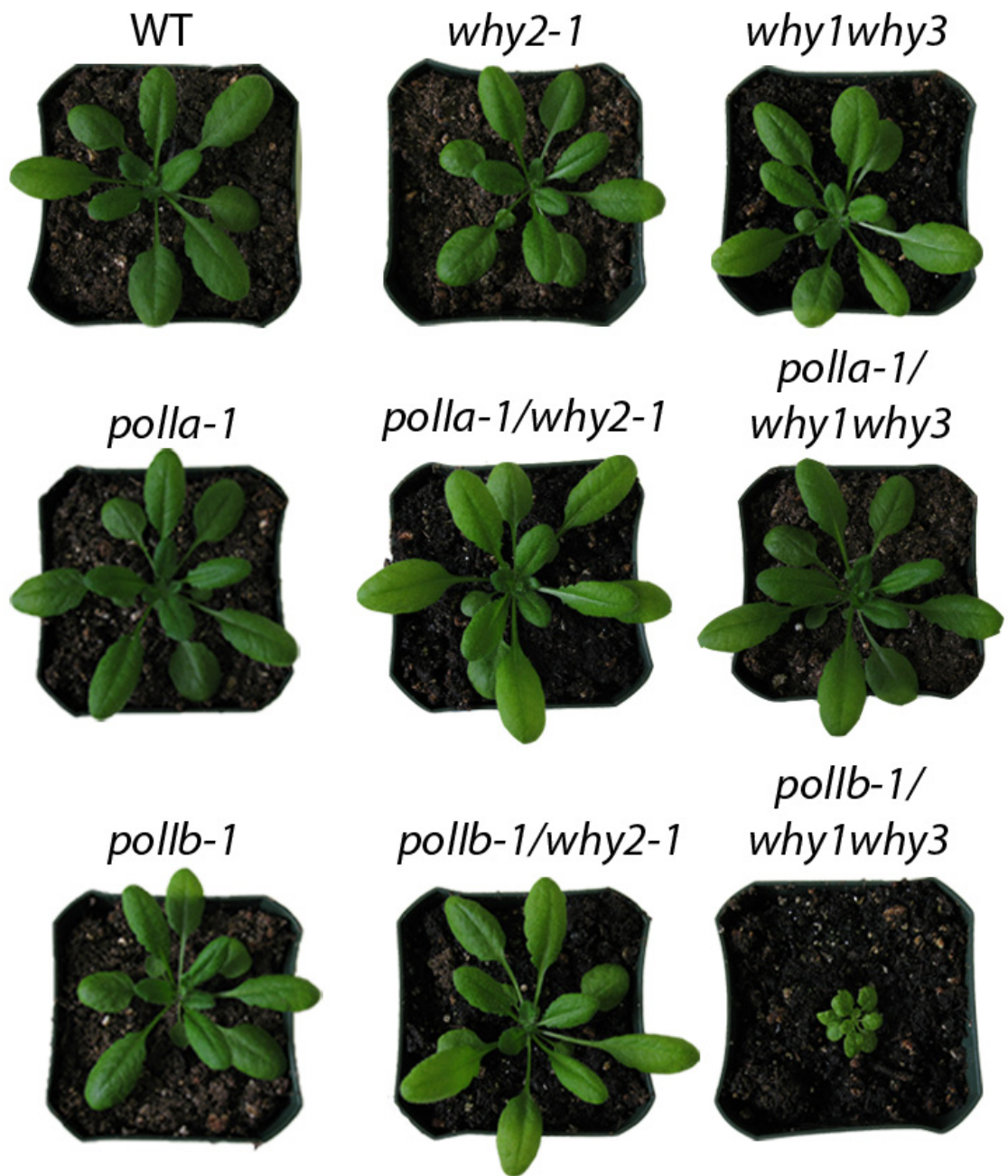


Figure 30. Crosses between Poll and Whirly mutants. Representative photographs of WT and mutant plants from the indicated genotypes.

Chapitre 4: Deuxième article

Organelle DNA rearrangement mapping reveals U-turn-like inversions as a major source of genomic instability in Arabidopsis and humans.

Article publié dans le journal Genome Research (publication en ligne en avance, gr.188573.114.)

4.1 MISE EN CONTEXTE:

Il est bien connu que des erreurs survenant lors de la réparation de l'ADN peuvent mener à l'apparition de réarrangements génomiques. Dans cet ouvrage, nous avons démontré que le plastide n'échappe pas à cette règle et que des réarrangements génomiques s'y produisent suite à l'induction de DSBs par la ciprofloxacine ou suite à la mutation de protéines impliquées dans le maintien de la stabilité génomique. Pour caractériser ces réarrangements, nous avons déjà mis au point des techniques d'amplification PCR permettant d'amplifier et de séquencer les réarrangements génomiques individuels, menant à l'identification générale du mécanisme responsable de leur formation. Cependant, ces techniques demeurent laborieuses et ne fournissent pas un aperçu global de l'instabilité génomique. Pour remédier à cette limitation et faciliter la caractérisation des mutants identifiés par notre criblage génétique, nous avons développé une méthode d'analyse des résultats de séquençage nouvelle-génération, nous permettant de caractériser l'instabilité génomique à l'échelle du génome. Cette méthode combine les avantages de différentes approches qui ont déjà été utilisées pour caractériser l'instabilité génomique dans la mitochondrie animale (190, 191), tout en ajoutant de nouvelles caractéristiques pour améliorer sa sensibilité et sa spécificité. Dans ce chapitre, je présente les résultats de séquençage nouvelle-génération obtenus pour plusieurs mutants identifiés lors du criblage génétique. Ces résultats nous permettent de proposer un nouveau modèle pour la formation fréquente d'inversions génomiques dans le génome des organelles.

Dans cette étude, j'ai participé à l'élaboration de l'approche bio-informatique menant à l'identification des réarrangements dans les génomes des organelles en collaboration avec Éric Zampini et Samuel Tremblay-Belzile. EZ a réalisé les expériences de laboratoire, soit l'extraction de l'ADN et la caractérisation des lignées mutantes. L'analyse des résultats et l'écriture de l'article fut un travail conjoint entre EZ, STB, NB et moi-même.

Title:

Organelle DNA rearrangement mapping reveals U-turn-like inversions as a major source of genomic instability in Arabidopsis and humans

Authors:

Éric Zampini ^{1,2}, Étienne Lepage ^{1,2}, Samuel Tremblay-Belzile ^{1,2}, Sébastien Truche ² and Normand Brisson ²

¹ These authors contributed equally to this work

² Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Québec, Canada, H3C 3J7.

Corresponding author:

Normand Brisson, Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, P.O. Box 6128, Station Centre-ville, Montréal, Québec, Canada, H3C 3J7

Tel: 514-343-6984

Fax: 514-343-2210

Running title:

U-turn inversions destabilize organelle genomes

Keywords:

Genome rearrangements, plastid, mitochondrion, Illumina, RecA.

Manuscript type: Research

4.2 ABSTRACT

Failure to maintain organelle genome stability has been linked to numerous phenotypes, including variegation and cytosolic male sterility (CMS) in plants, as well as cancer and neurodegenerative diseases in mammals. Here we describe a next-generation sequencing approach that precisely maps and characterizes organelle DNA rearrangements in a single genome-wide experiment. In addition to displaying global portraits of genomic instability, it surprisingly unveiled an abundance of short-range rearrangements in *Arabidopsis thaliana* and human organelles. Among these, short-range U-turn-like inversions reach 25% of total rearrangements in wild-type *Arabidopsis* plastids and 60% in human mitochondria. Furthermore, we show that replication stress correlates with the accumulation of this type of rearrangements, suggesting that U-turn-like rearrangements could be the outcome of a replication-dependent mechanism. We also show that U-turn-like rearrangements are mostly generated using microhomologies and are repressed in plastids by Whirly proteins WHY1 and WHY3. A synergistic interaction is also observed between the genes for the plastid DNA recombinase *RECA1* and those encoding plastid Whirly proteins, and the triple mutant *why1why3reca1* accumulates almost 60 times the WT levels of U-turn-like rearrangements. We thus propose that the process leading to U-turn-like rearrangements may constitute a RecA-independent mechanism to restart stalled forks. Our results reveal that short-range rearrangements, and especially U-turn-like rearrangements, are a major factor of genomic instability in organelles and raise the question of whether they could have been underestimated in diseases associated with mitochondrial dysfunction.

4.3 INTRODUCTION

The endosymbiosis events at the origin of mitochondria and plastids have been a key step in the appearance of eukaryotic cells. During evolution, most of the genes present in the genome of the endosymbionts have been lost or transferred to the nucleus, resulting in organelles with small, dense genomes that almost exclusively encode essential genes. There is therefore a strong selective pressure for the maintenance of genome stability in organelles (Wolfe et al. 1987). In plants, *Arabidopsis thaliana* has emerged as a useful model to study organelle genome rearrangements. Several actors involved in the maintenance of *Arabidopsis* plastid and mitochondrion genomes have been identified, among which are the Whirly proteins, a family of single-stranded DNA-binding proteins that guard organelles against genomic rearrangements (Maréchal et al. 2009; Cappadocia et al. 2010). It has been proposed that these proteins stabilize single-stranded DNA and guide it through conservative repair mechanisms such as homologous recombination (Maréchal et al. 2009; Cappadocia et al. 2010). Whirly proteins are found in all plant organelles and, in *Arabidopsis*, WHY1 and WHY3 are targeted to the plastids while WHY2 is targeted to the mitochondria (Krause et al. 2005). The RecA family of DNA-binding proteins also maintains organelle genome stability through their central role in homologous recombination. In *Arabidopsis*, three RecA proteins are found in organelles, with RECA1 targeted to the chloroplast, RECA3 to the mitochondrion, and RECA2 targeted to both organelles (Shedge et al. 2007). These proteins were shown to be involved in DNA double-strand break (DSB) repair and, in plastids, to maintain the structure of the genome (Rowan et al. 2010; Miller-Messmer et al. 2012). Prokaryote RecA proteins were also shown to be essential for fork reversal, a mechanism that allows the accurate restart of paused replication forks, thereby promoting fork progression in conditions of replication stress (Seigneur et al. 2000; Robu et al. 2001; Costes and Lambert 2012).

Replication represents a major challenge to plastid and mitochondrion genome stability. For example, it was shown that the mutation of the type-I polymerase *POLIB* in *Arabidopsis* causes replication stress at early developmental stages and increases the amount of DSBs upon genotoxic stress treatment (Parent et al. 2011). Interestingly, mutation of the mammalian mitochondrial DNA polymerase gamma has also been linked to replication stress and DSBs (Vermulst et al. 2008; Ameer et al. 2011). Replication-dependent DSBs are known to arise when a fork collapses or encounters a nick in the matrix DNA (Zeman and Cimprich 2014). These DSBs can subsequently be repaired by homologous recombination or by error-prone mechanisms such as microhomology-mediated recombination (MHMR) and non-homologous end-joining (NHEJ) (Lundin et al. 2002; Lee et al. 2007; Hastings et al.

2009). Replication-dependent genomic instability is however not solely induced by DSB repair mechanisms (Zeman and Cimprich 2014). Indeed, stalled forks have recently been shown to produce DSB-independent fusions of nearby inverted-repeats, which lead to the formation of palindromic chromosomes (Mizuno et al. 2009; Paek et al. 2009). The mechanisms by which these fusions take place remain a subject of debate, and three distinct possibilities have been proposed: faulty template switching, tandem inversion duplications, and replication U-turns (Mizuno et al. 2009; Paek et al. 2009; Kugelberg et al. 2010; Seier et al. 2012; Mizuno et al. 2013).

To date, our understanding of how organisms deal with replication-associated genomic instability has mostly been obtained using reporter systems. A drawback of these systems is that they are impractical for the study of organelle genomes. To overcome this limitation, we developed a next-generation sequencing approach which allows the characterization of organelle DNA instability at a genome-wide level while providing information about the mechanism underlying each rearrangement formation.

4.4 RESULTS

Short-range rearrangements are abundant in organelle genomes

We first evaluated overall plastid DNA (ptDNA) rearrangements in the *Arabidopsis thaliana* ecotype Col-0. To obtain a global and quantitative portrait of ptDNA instability, a heat map representation was employed in which all rearrangements are reported at the intersection of the two genomic coordinates that correspond to each side of the rearrangement junction. This analysis shows that many rearrangements occur apparently randomly in this plastid genome (Fig. 31A). It also reveals an over-representation of short-range rearrangements (<1,000 bp), as indicated by the higher intensity of the heat map diagonal (Fig. 31A,D). To validate our approach, the same analysis was repeated with the *Arabidopsis thaliana* ecotypes Ts-1 (Tossa de Mar) and Ws-2 (Wassilewskija) (Weigel and Mott 2009). Results indicate that the overall pattern and relative level of plastid DNA rearrangements are strikingly similar between ecotypes (Fig. 37).

To verify if mitochondria exhibit a similar pattern of genomic instability, we generated a heat map for *Arabidopsis thaliana* (Col-0) mitochondrial genome rearrangements. Interestingly, the vast majority of rearrangements also occur at short range in this organelle (Fig. 31B,D). To our knowledge, this is the first time that such a high level of short-range rearrangements is reported in organelle genomes, which raises the question of whether these could also be abundant in animal mitochondria. To test

this hypothesis, we subjected publicly available datasets from human brain and liver to the same analysis. This revealed a similar pattern of short-range genomic instability, reaching 86% and 84% of total rearrangements in brain and liver cells, respectively (Fig. 31C-D). In contrast, analysis of genomic instability in *E. coli* reveals a much less striking prominence of short-range rearrangements (<http://genome.cshlp.org/content/suppl/2015/03/20/gr.188573.114.DC1/SuppTableS4.xlsx>), indicating that the detection of high levels of these rearrangements in organelles does not occur systematically in our approach. Globally, these results indicate that short-range rearrangements are a major factor of genomic instability in both plant and animal organelles.

Microhomology and non-microhomology repair happen at similar rates in wild-type organelles

To get insights into the mechanisms involved in the formation of the DNA rearrangements in organelles, we further analyzed the reads corresponding to rearranged genome molecules. Because these reads are always composed of two alignments mapping distinct regions of the genome, two types of DNA rearrangements can easily be discriminated: those that possess a microhomology at their junction, most likely formed by MHMR, and those without microhomology, reminiscent of NHEJ repair (Fig. 32A). This analysis revealed that in wild-type Col-0 (WT) plastids and mitochondria, 55% and 64% respectively of genome rearrangements detected arose from microhomology-dependent pathways with microhomologies of 5 bp or more (Fig. 32B). Similarly, rearrangements dependent on microhomologies represent 64% and 67% of total genomic instability in human brain and liver mitochondria, respectively (Fig. 32B).

U-turn-like rearrangements are ubiquitous among organelle short-range rearrangements

The fact that short-range rearrangements represent a large part of the total junctions identified in organelle genomes suggests that these rearrangements could arise through a specific mechanism taking place at nearby microhomologies. Analysis of rearrangement junctions at the base pair level offers the key advantage of allowing the characterization of these mechanisms. We therefore measured the proportion of rearrangements involving inversions among the short-range rearrangements (<1,000 bp) according to their gap length in comparison to other types of rearrangements (duplications and deletions) in *Arabidopsis* and human organelles. This revealed that inversions occurring over distances smaller than 50 bp are largely overrepresented (Fig. 32C). This type of genomic instability seems to be of particular importance in organelles since it represents approximately 26% and 38% of total rearrangements in *Arabidopsis* plastids and mitochondria,

respectively (Fig. 32E). This proportion is even higher in human mitochondria, reaching 64% and 56% of the total brain and liver genomic instability, respectively (Fig. 32E). A closer look at the junctions of these inversions indicates that they mainly occur upstream of the junction breakpoint, on the opposite strand (Fig. 32D-F). Interestingly, although often imperfect, approximately 85% of these short-range inversions present a microhomology at their junction in plastids. The overall characteristics of the DNA junctions observed in these short range inversions can best be explained by a model in which replication executes a U-turn upstream on the opposite strand (Fig. 32D). It also suggests that the inversions take place on the opposite strand between the 5' end of the lagging strand and the site where the DNA is unwound in the replication fork.

Whirly proteins protect the plastid genome from MHMR but not from NHEJ

The organelle DNA polymerase POLIB and Whirly proteins have previously been shown to protect plastid genomes against long-range deletions and duplications mediated by microhomologies. To verify whether these proteins do also offer a protection against short-range DNA rearrangements, we further characterized the *Arabidopsis* mutant lines *why1why3* and *pollb* by sequencing. We also included in our analysis the mutant line *reca1-1* (hereafter called *reca1*), which is mutated for plastid *RECA1* (Fig. 38). While only a small increase in ptDNA rearrangements is observed in *pollb* and *reca1* as compared to the Col-0 parent line, a much higher level of genomic instability is observed in the *why1why3* line (Fig. 33A-B). Interestingly, the heat map diagonal region of all three mutant lines also shows a much higher intensity than the rest of the genome (Fig. 33A). It thus seems that short-range rearrangements also constitute an important part of the total ptDNA instability in these lines and that their occurrence is increased when genes involved in DNA metabolism are mutated. To get insights into the mechanisms involved in the formation of these rearrangements, we determined the prevalence of microhomology usage for each line. While approximately 0.12 DNA rearrangements per genome arose from microhomologies in WT plants, this increased to 0.8 microhomology-dependent rearrangements per genome in *why1why3* (Fig. 33B). This shift toward microhomology usage in *why1why3* is also confirmed by the analysis of the lengths of the microhomologies leading to rearrangements. Indeed, while microhomology-mediated rearrangements in WT, *reca1* and *pollb* plants are in large part produced by microhomologies of 5 to 9 bp, those generated from microhomologies of 10 to 14 bp are the most prevalent in the Whirly mutant (Fig. 33C). In contrast, the level of DNA rearrangements generated independently of microhomologies is similar in all three lines (Fig. 33B), suggesting that Whirly proteins mainly suppress the appearance of microhomology-

dependent rearrangements, and do not affect microhomology-independent pathways. The *pollb*, *reca1* and *why1why3* mutations however have little effect on the accumulation of U-turn-like rearrangements (inversions occurring over distances smaller than 50 bp), which only slightly increase in all three mutant lines (Fig. 33D).

Whirly proteins, POLIB and RECA1 all act to maintain stability in the plastid genome

Genome maintenance is a tightly controlled process in which many proteins act in concert to repress the accumulation of DNA rearrangements. Consequently, the mutation of a single gene may not be sufficient to significantly destabilize the genome. We verified whether *POLIB*, *RECA1* and plastid Whirly genes have additive or synergistic effects by combining mutations for these genes in *Arabidopsis*. While the phenotype of a *reca1pollb* line is indistinguishable from that of WT plants (Fig. 34A), *why1why3pollb* is characterized by a severe growth retardation phenotype, in addition to yellow-variegation (Fig 34A) (Parent et al. 2011; Lepage et al. 2013). *why1why3reca1* triple mutants show severe growth retardation as well and display white variegation and leaf distortion (Fig 34A and Fig. 39A). Also, embryo lethality is observed in *why1why3reca1*, with 59% of seeds being unable to germinate (Fig. 39B). The quadruple mutant *why1why3pollbreca1* could not be isolated from the progeny of *why1why3^{-/-}pollb^{+/-}reca1^{+/-}* plants, nor from *why1why3^{-/-}pollb^{-/-}reca1^{+/-}* plants (Stouffer's Z test p-value <0.05), suggesting that high levels of plastid genome instability leads to embryo lethality. While the overall pattern of ptDNA rearrangements in *reca1pollb* was similar to that of WT plants, an important increase was observed in both *why1why3pollb* and *why1why3reca1* (Fig. 34B). Furthermore, a hotspot for rearrangements is present in the heat map of these two triple mutants, at the intersection of regions 86-101 kilobases (kb) and 68-100 kb. By contrast, the IR region from 101-108 kb, which encodes both plastid ribosomal RNAs, displays lower levels of instability.

Comparison of ptDNA rearrangements between *why1why3pollb* and *why1why3reca1* indicates that their amount of long-range rearrangements (≥ 1000 bp), arising with or without the use of microhomologies, are strikingly similar (Fig. 34D-E). This suggests that these types of rearrangements are unlikely to be responsible for the difference in phenotypes observed between the two lines. In contrast, rearrangements having occurred between regions separated by less than 1 kb were much more abundant in *why1why3reca1* than in *why1why3pollb* (Fig. 34C-D). *why1why3reca1* plastids also contain a 60-fold increase in U-turn-like rearrangements compared to WT plastids (Fig. 34F), which corresponds to approximately 60% of all short-range rearrangements present in this line. These results

suggest that Whirly proteins and RECA1 both act to suppress the appearance of U-turn-like rearrangements in plastids. In addition, this high level of U-turn-like rearrangements could account for the severe phenotype observed in *why1why3reca1* plants (Fig. 34A).

U-turn-like rearrangements are associated to replication stress

The characteristics of the U-turn-like rearrangements identified above suggest that they might arise during replication, and that their formation could therefore be linked to replication stress. To test this hypothesis, we verified if a replication stress could be observed in *why1why3pollb* and *why1why3reca1*. Replication stress is defined here as the slowing of replication fork progression during DNA synthesis (Zeman and Cimprich 2014) which, in next-generation sequencing data, has been associated to a progressive, directional decrease in DNA copy number along the genome (Slager et al. 2014). We therefore compared ptDNA coverage curves for each mutant line to the WT. Results indicate that the coverage in WT plants has a similar pattern to those previously reported in the literature for plastid genomes (Wu et al. 2012; Ferrarini et al. 2013), and to that of all single and double mutant lines used in this study (Fig. 40). However, the pattern for ptDNA coverage appears quite different in *why1why3pollb* and *why1why3reca1* (Fig. 35A). Regression analysis within the large single-copy region (LSC, 0-84 kb) of *why1why3pollb* reveals a steeper slope than in the WT (Fig. 35B), suggesting that replication is affected in this mutant line and is unidirectional along the LSC, going from 84 kb toward the beginning of the genome. However, yet another pattern is observed in *why1why3reca1*, with two slopes converging in the middle of the LSC in a manner consistent with bidirectional replication (Fig. 35B). Both of these patterns were also confirmed using quantitative PCR (Fig. 41). These results therefore suggest that the *pollb* and *reca1* mutations, when combined to the *why1why3* mutations, affect replication differently. Nevertheless, the steeper slopes in the LSC observed for *why1why3pollb* and *why1why3reca1* suggest a replication stress, which could be at the origin of the increase in short-range inversions observed in the plastid genome of both of these mutant lines (Fig. 34F). Interestingly, the essential role of RecA in replication fork reversal and restart was previously shown in bacteria and could account for the replication stress observed in *why1why3reca1*. This result thus supports a role for RECA1 in plastid DNA replication.

Since replication stress is generally associated to incomplete replication of chromosomes, pulse-field gel electrophoresis (PFGE) was used to visualize the distribution of plastid genomic molecules in the previous mutant lines. This confirmed an earlier study which demonstrated that, in contrast to the

WT, no monomeric form of the chloroplast genome is observed in *reca1* mutant plants (Rowan et al. 2010). Similar to *reca1*, disappearance of the monomeric form is observed in *why1why3pollb* and *why1why3reca1* (Fig. 42). However, a smear of subgenomic molecules, likely associated with abortive rounds of replication, is instead observed in these mutants, supporting the hypothesis that the elevated levels of U-turn-like rearrangements in these lines could be the consequence of an ongoing replication stress. Note that subgenomic molecules are also observed in the *why1why3* mutant, suggesting that Whirly proteins are also important for replication.

4.5 DISCUSSION

By sequencing DNA rearrangement junctions, we have been able to obtain a genome-wide portrait of DNA instability in the genomes of organelles and of a prokaryote. Unlike paired-end sequencing approaches used to detect rearrangements, our method is not affected by the distance over which the rearrangements occur. It therefore provides a view of rearrangements occurring both at short-range and long-range, while also allowing some characterization of the DNA repair mechanisms used (Fig. 43-44). Indeed, it revealed that in *Arabidopsis* and human organelles, both microhomology-dependent and independent pathways constitute an important driving force of genome variation, with pathways using microhomologies being used slightly more often (Fig. 32B). Although illegitimate recombination has been known to occur in organelles (Small et al. 1987; Ogihara et al. 1988), there was no evidence that microhomology-independent repair, such as NHEJ, accounted for such a substantial proportion of organelle DNA rearrangements. Our approach also unveiled an unexpected pattern of genomic instability in organelles of both *Arabidopsis* and humans, with a strong propensity to generate rearrangements between closely spaced regions of the genome. Since DNA rearrangements are often associated to errors occurring during DNA repair (Lundin et al. 2002; Lee et al. 2007; Hastings et al. 2009), the high occurrence of short range rearrangements suggests that, in organelle genomes, inaccurate DNA repair takes place predominantly in the vicinity of the damaged DNA.

Our findings also reveal that, in organelles, a previously-unreported but frequent error-prone mechanism exists that most likely contributes to the restart of stalled replication forks. Indeed, we observed that U-turn-like rearrangements are particularly abundant among short-range DNA rearrangements in both *Arabidopsis* and human organelles. The finding that an apparent DNA replication stress correlates with the appearance of U-turn-like rearrangements in plastids suggests that they are initiated in a replication-dependent manner. Paused forks are unstable structures often

formed in conditions of replication stress, which leads us to hypothesize that they act as a template for U-turn-like rearrangements. The tendency of these U-turn-like rearrangements to occur upstream and the presence of short inverted repeats at most of their junctions suggest that a small inverted repeat in the 3' end of the nascent strand would misanneal to its complement upstream on the opposite strand and lead to inaccurate fork restart (Fig. 36).

The fact that microhomologies favor U-turn-like rearrangements suggests that *WHY1* and *WHY3*, which limit the accumulation of microhomology-dependent rearrangements, prevent this form of instability by repressing the misannealing of closely-spaced microhomologies. Also, the observation that single mutations of *RECA1* and *POLIB* have no visible impact on U-turn-like rearrangements but cause major changes only when combined to the mutations of *WHY1* and *WHY3* indicate the synergistic interaction between these genes and the Whirly genes. In the absence of Whirly proteins, replication stress could explain an increase in U-turn-like rearrangements, as observed in *why1why3polib*. In contrast, we postulate that replication stress alone does not likely account for the much larger increase observed in *why1why3reca1*. We propose instead that *RECA1* is involved in a conservative repair pathway that directly competes with replication U-turns. Interestingly, the recombinase *RecA* has been shown to be essential in bacteria to efficiently bypass lesions and restart replication by promoting replication fork reversal (Seigneur et al. 2000; Robu et al. 2001; Costes and Lambert 2012). It can thus be hypothesized that *RECA1* also promotes lesion bypass and accurate fork restart in plastids, thereby limiting the accumulation of U-turn-like rearrangements. The severe phenotype observed in the *why1why3reca1* mutant also indicates that U-turn-like rearrangements have deleterious effects, and that their occurrence must be limited by multiple checkpoints.

Taken together, our results suggest a model for the generation of U-turn-like rearrangements in which progression of the leading strand polymerase is arrested during replication as a consequence of replication stress (Fig. 36). If the fork remains paused and eventually collapses, mechanisms such as *RecA*-dependent fork-reversal can accurately restart replication. In the absence of *RECA1* and Whirly proteins, forks can restart inaccurately by the misannealing of a small inverted repeat in the 3' end of the nascent strand to its complement upstream on the opposite strand. Replication would then be re-initiated on the opposite strand until it either displaces the lagging strand or is ligated to its 5' end, resulting in the formation of a palindromic chromosome. Although this mechanism generates genomic rearrangements, it might serve as an alternative mechanism to restart replication forks. Whirly

proteins would avert these rearrangements by binding to the single-stranded DNA and thus preventing misannealing of small inverted repeats.

Overall, our approach reveals that short-range rearrangements, and especially U-turn-like inversions, have been largely underestimated in organelle genomes. In this study, we report that high levels of U-turn-like rearrangements correlate with the appearance of the severe phenotype observed in *why1why3reca1* and to its 59% seed lethality. Since this type of genomic instability also constitute the majority of the DNA rearrangements occurring in human mitochondria, it warrants further investigation into its link to the development of some of poorly understood mitochondrial disorders. Mitochondrial genome instability has indeed been observed in many clinical disorders including Parkinson's disease (Bender et al. 2006; Kraytsberg et al. 2006), inclusion body myositis (Moslemi et al. 1997) and cancer (Ju et al. 2014). In this regard, it will be interesting to evaluate if particular patterns of mitochondrial genomic instability are observed in the context of these disorders. Genomic rearrangements associated to inversions in the human nuclear genome have also been linked to leukemia, autism, and intellectual disability (Pui et al. 1992; Hermetz et al. 2014). It would therefore also be interesting to determine whether such rearrangements also occur through a U-turn-like mechanism and lead to the onset of such phenotypes.

4.6 METHODS

Plant material and growth conditions. *Arabidopsis thaliana* (ecotype Columbia-0) mutant lines *reca1-2* (SALK_072979) (Rowan et al. 2010), *pollb-1* (SALK_134274), *why1why3*, and *why1why3pollb-1* were reported previously (Maréchal et al. 2009; Parent et al. 2011). The *reca1-1* (SALK_057982) mutant line was isolated from heterozygous seeds obtained from ABRC (Alonso et al. 2003). Seeds were sown on soil, vernalized for 3 days at 4°C, and grown under laboratory normal light (100 mmol m⁻² s⁻¹), at 22°C on a 16-h-day/8-h-dark cycle. Representative photographs were taken at 21 days. To determine statistical significance of *why1why3pollbreca1* lethality, chi-square tests were performed on progeny of *why1why3^{-/-}pollb^{+/-}reca1-1^{+/-}* and *why1why3^{-/-}pollb^{-/-}reca1-1^{+/-}* plants. Chi-square *p*-values were then combined using Stouffer's Z test (Whitlock 2005).

DNA isolation and DNA-seq. For Col-0, *pollb*, *reca1*, *why1why3*, *reca1pollb*, *why1why3pollb* and *why1why3reca1*, total DNA was isolated from ≈400 mg pools of 14-day-old *Arabidopsis* plants using the cetyl trimethyl-ammonium DNA extraction protocol (Weigel and Glazebrook 2002). DNA was

fragmented to \approx 200-500 bp using S-Series Covaris according to Illumina's specifications. Libraries were prepared using the TruSeq DNA library preparation kit (Illumina) according to manufacturer's instructions. Efficient library generation was then assessed using a Bioanalyzer platform (Agilent) and Illumina MiSeq-QC run was performed. Sequencing was performed using an Illumina HiSeq 2000 using TruSeq SBS v3 chemistry at the Institute for Research in Immunology and Cancer's Genomics Platform (Université de Montréal). Cluster density was targeted at around 600 to 800 kilo clusters mm^{-2} . Sequencing data is available on the NCBI Sequence Read Archive (SRA) under the BioProject Number SRP051208 (Col-0: SRX883065, *pollb*: SRX813508, *reca1*: SRX883066, *why1why3*: SRX883067, *reca1pollb*: SRX883068, *why1why3pollb*: SRX883069, *why1why3reca1*: SRX883070).

Publicly available Illumina whole-genome sequencing datasets. NCBI SRA accession numbers for paired-end Illumina whole-genome sequencing datasets for *Arabidopsis* ecotypes Ts-1 and Ws-2 are, respectively, SRX145018 and SRX145037 (submitted by the SALK Institute for Biological Studies). SRA accession numbers for paired-end Illumina whole-genome sequencing datasets for four human brain samples and four liver samples are, respectively, ERX385572, ERX385573, ERX385574, ERX385575 and ERX385576, ERX385577, ERX385578, ERX385579 (submitted by the Institute for Molecular Bioscience - The University of Queensland). SRA accession numbers for paired-end Illumina whole-genome sequencing datasets for *E. coli* are SRX154301, SRX154337, SRX154338 and SRX154342 (submitted by Indiana University).

Enrichment for reads with potential junctions. The Galaxy online software suite was used to develop a workflow that enriches reads spanning potential junctions from paired-end Illumina datasets (Fig. 43) (Goecks et al. 2010). As part of the workflow, quality filtering was performed to keep pairs for which both reads measure at least 40 bases, have an average quality of at least 20 and display no more than 50 bases outside of the quality range (Blankenberg et al. 2010). Using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009), both reads from each pair were mapped individually against the whole reference genome (GRCh38 with alternate loci for *Homo sapiens* and Galaxy built-in reference genomes *Arabidopsis_thaliana_TAIR10* for *Arabidopsis thaliana* and *eschColi_K12* for *E. coli*) to remove pairs of read that do not contain a junction. This also allows the removal of nuclear plastid DNA sequences (NUPTs) and nuclear mitochondrial DNA sequences (NUMTs) that would otherwise be flagged as rearrangements. The alignment was performed without using a subsequence as a seed and parameters were set to a maximum edit distance of 7, a maximum number of gap opens of 1, a maximum number of gap extensions of 3, a mismatch penalty of 3, a gap open penalty of 11 and a gap

extension penalty of 4. Pairs with at least one unmapped read were then selected and the first 25 bases of both reads were mapped against the desired organelle genome (GenBank: AP000423.1 for *Arabidopsis* plastid, GenBank: Y08501.2 for *Arabidopsis* mitochondria, NCBI: NC_012920.1 for human mitochondria and Galaxy built-in reference genome eschColi_K12 for *E. coli*) using Bowtie for Illumina with default parameters (Langmead et al. 2009). Aligning the 25 first bases of both reads is critical, as rearrangement junctions would prevent the alignment of the whole rearranged read. The distance between each alignment position of a pair was then computed and pairs containing an unmapped read were discarded. Only pairs that fulfil one of the following conditions were then kept: 1- Each read of a pair is mapped in opposite orientation, 2- Each read of a pair is mapped in the same orientation with a distance of at least 3 bases. Finally, pairs were filtered to discard those that correspond to DNA fragments spanning junctions created by the different isoforms of the plastid genome or the circular nature of organelle DNA. Workflow statistics are listed in Tables 2 to 7. Enriched reads spanning potential junctions are provided for all lines and organisms used in this study (http://genome.cshlp.org/content/suppl/2015/03/20/gr.188573.114.DC1/Supplemental_Data_File_1.xlsx).

Analysis of rearranged reads. Sequences of reads containing a potential junction were aligned against the organelle genome (Positions 1 to 128214 of GenBank: AP000423.1 for *Arabidopsis* plastid, GenBank: Y08501.2 and JF729201.1 for *Arabidopsis* mitochondria, NCBI: NC_012920.1 for human mitochondria and NCBI: NC_000913.2 for *E. coli*) using BLAST+ (Camacho et al. 2009) and the following parameters: “blastn -query potential_junctions.fasta -db organelle_genome.fasta -out output.txt -word_size 10 -max_target_seqs 2 -evalue 0.0001 -outfmt 6”. Only reads with two alignments were kept for further analysis. Among the reads that output two alignments, the longest alignment was subtracted to the length of the read, and reads with values of 5 and less were discarded. BLAST+ outputs for rearrangements reads are provided for all lines and organisms used in this study (http://genome.cshlp.org/content/suppl/2015/03/20/gr.188573.114.DC1/Supplemental_Data_File_1.xlsx). To ensure that errors in wild-type plastid genome annotations do not generate false-positive rearrangements, the genomic sequence was confirmed using BLAST+ for each locus at which rearrangements are more abundant than 5% of the average coverage. The reference locus sequence was used as the subject sequence to align the total reads of the samples. When rearrangements were found to be in homoplasmy, these rearrangements were considered as false-positives and were discarded. For *Arabidopsis* Col-0 mitochondria, two genome annotations exist (GenBank: Y08501.2

and JF729201.1). To remove false-positives, only reads considered as rearrangements on both annotations were kept. The following algorithms were then used to analyze the rearrangement junctions according to the alignment position output by BLAST+.

Distance measurement algorithm evaluates the difference between the reference genome positions of the 3' end of the first alignment and the 5' end of the second alignment. Overlap of alignments algorithm corresponds to the subtraction of the total read length to the sum of the lengths of both alignments. Rearrangements with an overlap of at least 5 bases are considered to have occurred through the use of microhomology, while the rest are assigned to the "No Microhomology" group. This high cut-off was used to ensure the stringency of microhomology calling. Inversion algorithm corresponds to rearrangements for which the reference genome positions from the 5' end to the 3' end are in ascending order for one alignment and in descending order for the other. All remaining rearrangements are treated as deletions or duplications. U-turn-like rearrangements are defined as those which present an inversion with a distance parameter of less than 50 bases. Downstream, same base and upstream U-turn-like rearrangements are determined by the position on the reference genome of the 5' end of the second alignment in relation to the 3' end and direction of the first alignment. The polarity of U-turn-like rearrangements identified in R2 reads has been inverted. Local mapping for read pairs associated to a single rearrangement were selected and aligned on the *Arabidopsis thaliana* (TAIR10) genome using the Golden Helix GenomeBrowse® visualization tool (Version 2.0.7). Representative results for a duplication, a deletion and an inversion are presented in Figure 45.

To evaluate the proportion of U-turn-like rearrangements that occurred through the use of imperfect microhomologies, reads corresponding to this type of genomic rearrangement were aligned against the plastid genome using BLAST+ and the following parameters: "blastn -query U-Turns.fasta -db plastid_genome.fasta -out output.txt -word_size 10 -max_target_seqs 2 -evalue 0.002 -outfmt 6 -penalty -1 -gap open 0 -gap extend 2". The previous algorithms were then used to evaluate the proportion of rearrangements harboring a microhomology at the junction.

The combination of the previous workflow to enrich reads with potential junctions and these algorithms achieved 64.0% sensitivity for detection of reads containing junctions and 97.5% specificity for accurate rearrangement type calling. For sensitivity, 200 randomly chosen reads from the potential junctions obtained following the Galaxy workflow for the plastid DNA were blind-tested individually to assess if they correspond to a rearrangement or not. For specificity, 480 randomly chosen

rearrangements analyzed by the algorithm were analyzed for accurate rearrangement type calling. For both sensitivity and specificity, the online BLAST interface (<http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi>) was used with the lowest stringency settings.

Plastid sequencing coverage analysis. Pairs with both reads fully aligned against the reference genome using BWA during enrichment for potential junctions were filtered to keep only those mapping the plastid genome. Positions of each read were rounded down to the nearest kb and all reads mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The number of reads mapping each 1 kb range was measured and normalized relative to 1,000,000 plastid reads.

Pulse-field gel electrophoresis (PFGE) analysis of the plastid genome. Chloroplasts were isolated from 21-day-old *Arabidopsis* plants as described previously (Little 1997). Isolated chloroplasts were resuspended in homogenization buffer and then mixed 1:1 to 45°C 1% low melting point agarose in TE buffer and allowed to fix at 4°C. Isolated chloroplasts concentration were adjusted according to a low-cycle amplification of a DNA fragment of the hypothetical protein RF2 (YCF2) with the following primers: YCF2FOR, GAT CTC TGA GAG CTG TTT CCG; YCF2REV, TGT TTC GCC TCT TAC TCG GAG. Agarose plugs were then soaked overnight at 50°C in lysis buffer (0.45 M EDTA pH 8.0, 1% (w/v) sarkosyl, 1 mg/mL proteinase K) and washed in storage solution (0.45 M EDTA pH 8.0, 1% (w/v) sarkosyl). Migration was performed in 1.5% agarose gel in 0.5X TBE buffer for 46 hours at 12°C using a Bio-Rad CHEF-DR® III system. Pulses switch times were set to 120 seconds at 5 V/cm using a 120 degrees angle. The gel was then soaked successively in 0.5X TBE buffer supplemented with 1 µg/mL RNase A for 2 hours at room temperature, in 0.5X TBE buffer supplemented with 5 µg/mL ethidium bromide for thirty minutes and finally in 0.5X TBE buffer to wash the gel. Southern hybridization was then performed as described previously (Maréchal et al. 2009) using a chloroplast probe amplified using the following primers: 49741FOR, CCT TAC GTA AAG GCC ACC CTA; 54551REV, TGG GAC GCA TAA CCG GAT ATG.

Quantitative PCR analysis of ptDNA levels. Total DNA was isolated from 14-day-old *Arabidopsis* Col-0, *why1why3pollb* and *why1why3reca1* plants using the cetyl trimethyl-ammonium DNA extraction protocol (Weigel and Glazebrook 2002). Primers used for qPCR were calibrated to ensure the amplification of a unique PCR product and efficiency between 1.90 and 2.05. Every reaction was carried out on biological and technical triplicates relative to the amplification of nuclear DNA. Primers sequences are as follows: 7660FOR, TGA TCC AGG ACG TAA TCC GGG AC; 7802REV, CGA ATC CCT CTC

TTT CCC CTT CTC C; 45345FOR, TTG GCA ATT CCT CAG GGG CAG; 45525REV, TTG ACT ATT CCT CAA GCG CGC C; 81312FOR, AGC TAC CCA ATA CTC AGG GGA TCC; 81460REV, AAA TAG AAG CAG GGC GAC GCG; nucDNA-FOR, GTT GAA GCC TCC GTT CCC TGC TA; nucDNA-REV, CTC TTC CAC CGT GCA TGG CTT GT. The Power SYBR® Green PCR Master Mix (Applied Biosystems) was used according to manufacturer's instructions. qPCR experiments and analysis were carried out using LightCycler 480 (Roche) and the LightCycler 480 software version 1.5, respectively.

Quantitative PCR analysis of *RECA1* expression in *reca1* mutants. Quantitative PCR analysis of *RECA1* expression was performed as described previously for 21-day-old plants (Lepage et al. 2013). Every reaction was carried out on biological and technical triplicates relative to the amplification of beta tubulin. Primers used for qRT-PCR were calibrated to ensure the amplification of a unique PCR product and efficiency between 1.90 and 2.05. Primers sequences are as follows: RECA1FOR, GGT GGA GGC CTA CCA AAG GG; RECA1REV, GGT GGA GGC CTA CCA AAG GG; BetaTubFOR, TCG TTG GGA GGA GGC ACA GGT; BetaTubREV, GCT GAG TTT GAG GGT ACG GAA GCA G. The Power SYBR® Green PCR Master Mix (Applied Biosystems) was used according to manufacturer's instructions. qPCR experiments and analysis were carried out using LightCycler 480 (Roche) and the LightCycler 480 software version 1.5, respectively.

DATA ACCESS

All newly generated sequencing data is freely available on the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under the BioProject Number SRP051208 (Col-0: SRX883065, *pollb*: SRX813508, *reca1*: SRX883066, *why1why3*: SRX883067, *reca1pollb*: SRX883068, *why1why3pollb*: SRX883069, *why1why3reca1*: SRX883070). The Galaxy workflow is freely available on Galaxy's Published Workflows section under the title "Rearrangement Junction Detection".

4.7 ACKNOWLEDGEMENTS

We thank M. Arteau, M. Valach and the Galaxy Team for assistance with the DNA-seq, PFGE and data processing, respectively. We also thank T. Vincent for his help in isolating mutant lines. This work was supported by scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC) to É.L., É.Z. and S.T.B., from the Fonds de Recherche du Québec – Nature et Technologies to S.T.B. and grants from NSERC to N.B.

AUTHOR CONTRIBUTIONS

É.Z. isolated plant DNA for sequencing and performed the PFGE and qPCR experiments. É.Z., É.L. and S.T.B. developed the bioinformatics approach. É.Z., É.L., S.T.B., S.T. and N.B. analyzed and interpreted the data. É.Z., É.L., S.T.B. and N.B. wrote the manuscript.

DISCLOSURE DECLARATION

The authors declare they have no conflict of interest.

4.8 FIGURE LEGENDS

Figure 31. Global portrait of organelle genome rearrangements in *Arabidopsis* and human. Heat maps depict each rearrangement as the intersection of the two genomic positions corresponding to the nucleotide on each side of the junction. (A) Rearrangement breakpoint positions of *Arabidopsis thaliana* (Col-0) plastids. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long grey bar, a red bar and a short grey bar, respectively. (B) Rearrangement breakpoint positions of wild-type *Arabidopsis thaliana* (Col-0) mitochondria. Each tile represents a region spanning 10 kb along each axis. Tile intensity represents the number of rearrangements per 100 mitochondrion genomes. (C) Rearrangement breakpoint positions of a representative sample for human brain (ERR419275) and liver (ERR419281) mitochondria. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 mitochondrion genomes. (D) Proportion of short-range (breakpoint positions less than 1,000 bp apart) and long-range (breakpoint positions at least 1,000 bp apart) rearrangements of *Arabidopsis thaliana* (Col-0) plastid, mitochondrion and of the mean of the four samples for brain and liver mitochondria. Ara.: *Arabidopsis*, mito.: mitochondrion.

Figure 32. Analysis of organelle genome rearrangements in *Arabidopsis* and human. (A) Depiction of the types of rearrangement junctions observed by next-generation sequencing. Large arrows

represent microhomologies and their orientation. Small triangles represent relative strand directions. (B) Proportion of rearrangements identified displaying a microhomology (≥ 5 bp) at their junction in each organelle. The mean of the four samples for brain and liver mitochondria is presented. (C) Proportion of short-range deletions/duplications and inversions ($<1,000$ bp) displaying a gap of the indicated length. Y axis represents the percentage of each type of rearrangement relative to its class size (gap length). A representative sample is shown for human brain (ERR419275) and liver (ERR419281) mitochondria. (D) Schematic representation of short-range inversions displaying a junction upstream, at the same base, or downstream on the opposite strand. The matrix strands are shown in blue, and the nascent strands in red with the junction gap shown as a dotted line. (E) Proportion of total rearrangements corresponding to short-range inversions (<50 bp, U-turn-like rearrangements) in each organelle genome. The mean of the four samples for brain and liver mitochondria is presented. (F) Proportion of short-range inversions (<50 bp) displaying a junction upstream, at the same base, or downstream on the opposite strand in each organelle genome. A representative sample is shown for human brain (ERR419275) and liver (ERR419281) mitochondria. Ara.: *Arabidopsis*, mito.: mitochondrion, del.: deletion, dup.: duplication.

Figure 33. Global portrait of plastid genome rearrangements in *Arabidopsis* mutant lines *pollb*, *reca1* and *why1why3*. (A) Plastid rearrangement breakpoint positions of the indicated *Arabidopsis* mutant lines. Heat maps depict each rearrangement as the intersection of the two genomic positions corresponding to the nucleotide on each side of the junction. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long grey bar, a red bar and a short grey bar, respectively. (B) Normalized amount of rearrangements per plastid genome displaying a microhomology (≥ 5 bp) (red) or not displaying a microhomology (blue) at their junction for each of the indicated mutant lines. (C) Proportion of total plastid rearrangements displaying a microhomology of given lengths, in bases, in each mutant line. Y axis represents the percentage of usage of each microhomology length. (D) Normalized amount of rearrangements corresponding to short-range inversions (<50 bp, U-turn-like rearrangements, red) and short-range deletions/duplications (<50 bp, blue) in each mutant line. Data were normalized on the total number of plastid genome reads. Del.: deletion, dup.: duplication. Two asterisks indicate a significant difference with p -value < 0.00001 using a chi-square test.

Figure 34. Global portrait of plastid genome rearrangements in *Arabidopsis* mutant lines *reca1pollb*, *why1why3pollb* and *why1why3reca1*. (A) Representative photographs of 21-d-old WT, *reca1pollb*, *why1why3pollb* and *why1why3reca1* *Arabidopsis* mutant plants. (B) Plastid rearrangement breakpoint positions of the indicated *Arabidopsis* mutant lines. Heat maps depict each rearrangement as the intersection of the two genomic positions corresponding to the nucleotide on each side of the junction. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long grey bar, a red bar and a short grey bar, respectively. (C) Short-range (<1000 bp) plastid rearrangement breakpoint positions of the indicated *Arabidopsis* mutant lines. Each heat map tile represents a region spanning 1 kb along the genome and the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. (D) Normalized amount of short-range (<1000 bp) and long-range (≥ 1000 bp) rearrangements per plastid genome for each of the indicated mutant lines. (E) Normalized amount of long-range rearrangements (≥ 1000 bp) per plastid genome displaying a microhomology (≥ 5 bp) (red) or not displaying a microhomology (blue) at their junction for each of the indicated mutant lines. (F) Normalized amount of rearrangements corresponding to short-range inversions (<50 bp, U-turn-like rearrangements) in each mutant line. Data were normalized on the total number of plastid genome reads. Two asterisks indicate a significant difference with p -value <0.00001 using a chi-square test.

Figure 35. Plastid DNA sequencing coverage curves for *Arabidopsis* lines WT, *why1why3pollb* and *why1why3reca1*. (A) Plastid sequencing coverage of pools of 14-day-old *Arabidopsis* seedlings of the indicated genotypes. Positions were rounded down to 1 kb. All reads mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long blue bar, a red bar and a short blue bar, respectively. Y axis represents the number of reads per 1,000,000 total plastid reads. (B) Regression analysis of the plastid large single-copy (LSC) sequencing coverage of WT, *why1why3pollb*, *why1why3reca1* seedlings. Positions were rounded down to 1 kb. Y axis represents the number of reads per 1,000,000 total plastid reads.

Figure 36. Model of microhomology-mediated U-turn-like inversions. The red shape represents the DNA polymerase and the blue shape represents the DNA helicase. Yellow and green strands represent inverted microhomologies. Impediments to replication fork progression might force the leading strand polymerase to pause and eventually unload. In the presence of RecA, the impediment may be bypassed to restart replication accurately. In the absence of RecA and Whirly proteins, a microhomology located at the 3' end of the nascent strand can anneal to its complementary inverted repeat on the opposite strand and restart the replication fork on the wrong strand. The dotted arrow represents a multistep process.

4.9 REFERENCES

- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R et al. 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**(5633): 653-657.
- Ameur A, Stewart JB, Freyer C, Hagstrom E, Ingman M, Larsson NG, Gyllensten U. 2011. Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. *PLoS Genet* **7**(3): e1002028.
- Bender A, Krishnan KJ, Morris CM, Taylor GA, Reeve AK, Perry RH, Jaros E, Hershenson JS, Betts J, Klopstock T et al. 2006. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat Genet* **38**(5): 515-517.
- Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy T. 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**(14): 1783-1785.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Cappadocia L, Maréchal A, Parent JS, Lepage E, Sygusch J, Brisson N. 2010. Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair. *Plant Cell* **22**(6): 1849-1867.
- Costes A, Lambert SA. 2012. Homologous recombination as a replication fork escort: fork-protection and recovery. *Biomolecules* **3**(1): 39-71.
- Ferrarini M, Moretto M, Ward JA, Surbanovski N, Stevanovic V, Giongo L, Viola R, Cavalieri D, Velasco R, Cestaro A et al. 2013. An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics* **14**: 670.

- Goecks J, Nekrutenko A, Taylor J, Galaxy T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology* **11**(8): R86.
- Hastings PJ, Ira G, Lupski JR. 2009. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS genetics* **5**(1): e1000327.
- Hermetz KE, Newman S, Conneely KN, Martin CL, Ballif BC, Shaffer LG, Cody JD, Rudd MK. 2014. Large inverted duplications in the human genome form via a fold-back mechanism. *PLoS Genet* **10**(1): e1004139.
- Ju YS, Alexandrov LB, Gerstung M, Martincorena I, Nik-Zainal S, Ramakrishna M, Davies HR, Papaemmanuil E, Gundem G, Shlien A et al. 2014. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *eLife* **3**.
- Krause K, Kilbiński I, Mulisch M, Rodiger A, Schafer A, Krupinska K. 2005. DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS letters* **579**(17): 3707-3712.
- Kraytsberg Y, Kudryavtseva E, McKee AC, Geula C, Kowall NW, Khrapko K. 2006. Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nat Genet* **38**(5): 518-520.
- Kugelberg E, Kofoid E, Andersson DI, Lu Y, Mellor J, Roth FP, Roth JR. 2010. The tandem inversion duplication in *Salmonella enterica*: selection drives unstable precursors to final mutation types. *Genetics* **185**(1): 65-80.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**(3): R25.
- Lee JA, Carvalho CM, Lupski JR. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**(7): 1235-1247.
- Lepage E, Zampini E, Brisson N. 2013. Plastid genome instability leads to reactive oxygen species production and plastid-to-nucleus retrograde signaling in *Arabidopsis*. *Plant Physiol* **163**(2): 867-881.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.
- Little SE. 1997. Large- or Small-Scale Isolation of Chloroplasts Using the Avanti® J Series of High Performance Centrifuges. *Beckman Instruments, Inc.*

- Lundin C, Erixon K, Arnaudeau C, Schultz N, Jenssen D, Meuth M, Helleday T. 2002. Different roles for nonhomologous end joining and homologous recombination following replication arrest in mammalian cells. *Mol Cell Biol* **22**(16): 5869-5878.
- Maréchal A, Parent JS, Veronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. 2009. Whirly proteins maintain plastid genome stability in Arabidopsis. *Proc Natl Acad Sci U S A* **106**(34): 14693-14698.
- Miller-Messmer M, Kuhn K, Bichara M, Le Ret M, Imbault P, Gualberto JM. 2012. RecA-dependent DNA repair results in increased heteroplasmy of the Arabidopsis mitochondrial genome. *Plant Physiol* **159**(1): 211-226.
- Mizuno K, Lambert S, Baldacci G, Murray JM, Carr AM. 2009. Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. *Genes Dev* **23**(24): 2876-2886.
- Mizuno K, Miyabe I, Schalbetter SA, Carr AM, Murray JM. 2013. Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature* **493**(7431): 246-249.
- Moslemi AR, Lindberg C, Oldfors A. 1997. Analysis of multiple mitochondrial DNA deletions in inclusion body myositis. *Hum Mutat* **10**(5): 381-386.
- 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.
- Paek AL, Kaochar S, Jones H, Elezaby A, Shanks L, Weinert T. 2009. Fusion of nearby inverted repeats by a replication-based mechanism leads to formation of dicentric and acentric chromosomes that cause genome instability in budding yeast. *Genes Dev* **23**(24): 2861-2875.
- Parent JS, Lepage E, Brisson N. 2011. Divergent roles for the two Poll-like organelle DNA polymerases of Arabidopsis. *Plant Physiol* **156**(1): 254-262.
- Pui CH, Carroll AJ, Raimondi SC, Schell MJ, Head DR, Shuster JJ, Crist WM, Borowitz MJ, Link MP, Behm FG et al. 1992. Isochromosomes in childhood acute lymphoblastic leukemia: a collaborative study of 83 cases. *Blood* **79**(9): 2384-2391.
- Robu ME, Inman RB, Cox MM. 2001. RecA protein promotes the regression of stalled replication forks in vitro. *Proc Natl Acad Sci U S A* **98**(15): 8211-8218.
- Rowan BA, Oldenburg DJ, Bendich AJ. 2010. RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. *Journal of experimental botany* **61**(10): 2575-2588.

- Seier T, Zilberberg G, Zeiger DM, Lovett ST. 2012. Azidothymidine and other chain terminators are mutagenic for template-switch-generated genetic mutations. *Proc Natl Acad Sci U S A* **109**(16): 6171-6174.
- Seigneur M, Ehrlich SD, Michel B. 2000. RuvABC-dependent double-strand breaks in dnaBts mutants require recA. *Mol Microbiol* **38**(3): 565-574.
- 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.
- Slager J, Kjos M, Attaiech L, Veening JW. 2014. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* **157**(2): 395-406.
- 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.
- Vermulst M, Wanagat J, Kujoth GC, Bielas JH, Rabinovitch PS, Prolla TA, Loeb LA. 2008. DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. *Nat Genet* **40**(4): 392-394.
- Weigel D, Glazebrook J. 2002. *Arabidopsis: A Laboratory Manual*. Cold Spring Harbor Lab Press, Cold Spring Harbor, NY.
- Weigel D, Mott R. 2009. The 1001 genomes project for Arabidopsis thaliana. *Genome biology* **10**(5): 107.
- Whitlock MC. 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* **18**(5): 1368-1373.
- 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.
- Wu J, Liu B, Cheng F, Ramchiary N, Choi SR, Lim YP, Wang XW. 2012. Sequencing of chloroplast genome using whole cellular DNA and solexa sequencing technology. *Frontiers in plant science* **3**: 243.
- Zeman MK, Cimprich KA. 2014. Causes and consequences of replication stress. *Nat Cell Biol* **16**(1): 2-9.

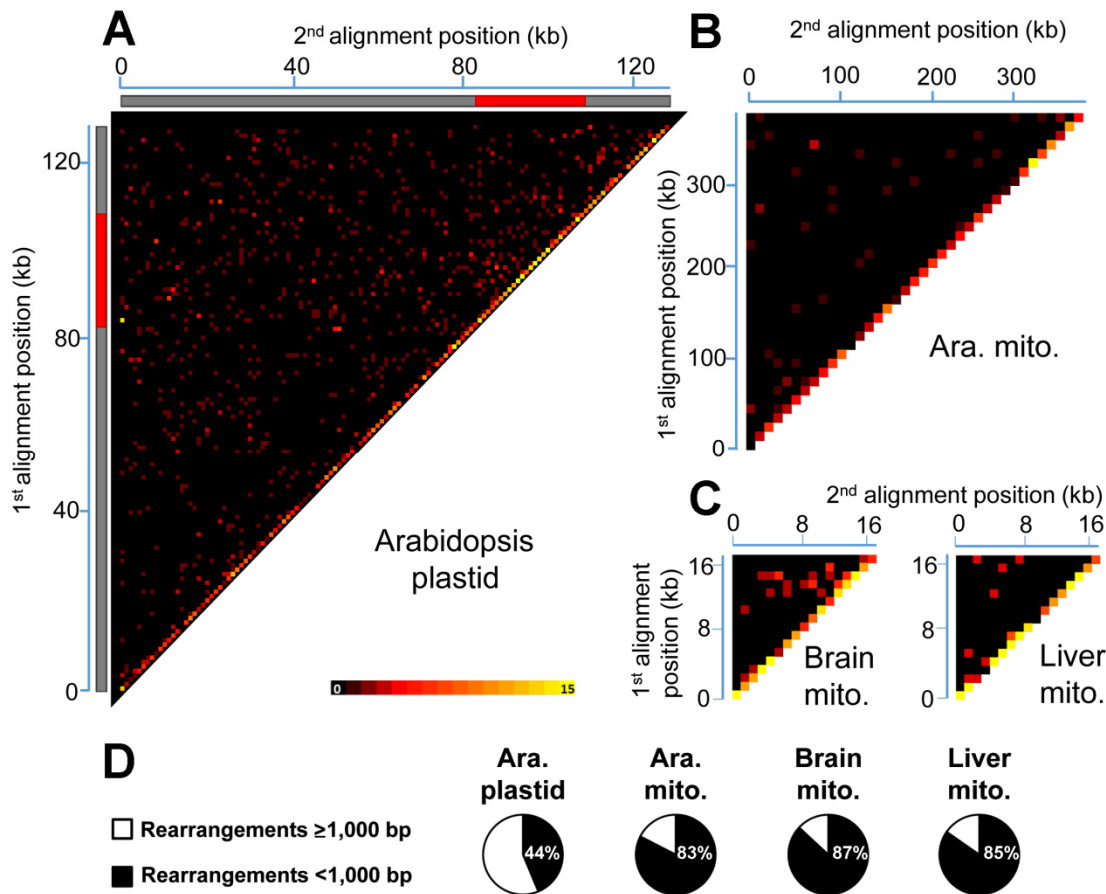


Figure 31. Global portrait of organelle genome rearrangements in Arabidopsis and human. Heat maps depict each rearrangement as the intersection of the two genomic positions corresponding to the nucleotide on each side of the junction. (A) Rearrangement breakpoint positions of *Arabidopsis thaliana* (Col-0) plastids. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long grey bar, a red bar and a short grey bar, respectively. (B) Rearrangement breakpoint positions of wild-type *Arabidopsis thaliana* (Col-0) mitochondria. Each tile represents a region spanning 10 kb along each axis. Tile intensity represents the number of rearrangements per 100 mitochondrion genomes. (C) Rearrangement breakpoint positions of a representative sample for human brain (ERR419275) and liver (ERR419281) mitochondria. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 mitochondrion genomes. (D) Proportion of short-range (breakpoint positions less than 1,000 bp apart) and long-range (breakpoint positions at least 1,000 bp apart) rearrangements of *Arabidopsis thaliana* (Col-0) plastid, mitochondrion and of the mean of the four samples for brain and liver mitochondria. Ara.: *Arabidopsis*, mito.: mitochondrion.

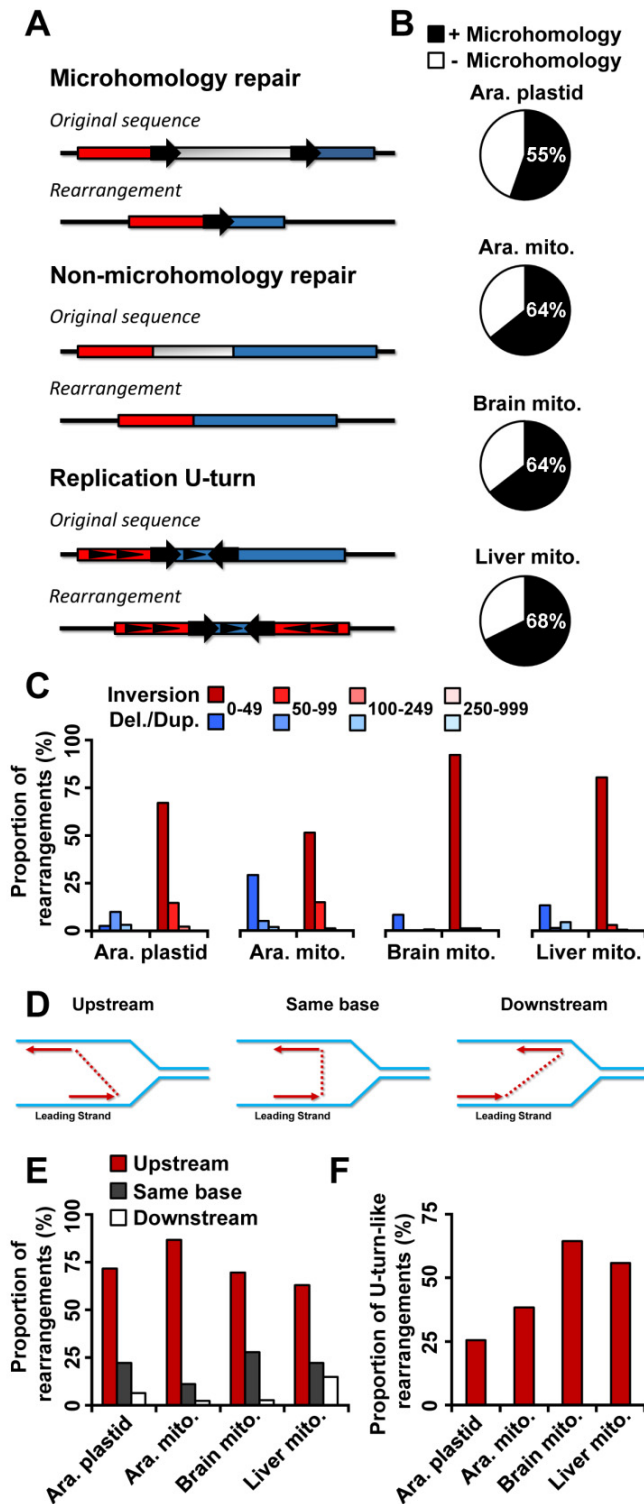


Figure 32. Analysis of organelle genome rearrangements in Arabidopsis and human. (A) Depiction of the types of rearrangement junctions observed by next-generation sequencing. Large arrows

represent microhomologies and their orientation. Small triangles represent relative strand directions. (B) Proportion of rearrangements identified displaying a microhomology (≥ 5 bp) at their junction in each organelle. The mean of the four samples for brain and liver mitochondria is presented. (C) Proportion of short-range deletions/duplications and inversions ($<1,000$ bp) displaying a gap of the indicated length. Y axis represents the percentage of each type of rearrangement relative to its class size (gap length). A representative sample is shown for human brain (ERR419275) and liver (ERR419281) mitochondria. (D) Schematic representation of short-range inversions displaying a junction upstream, at the same base, or downstream on the opposite strand. The matrix strands are shown in blue, and the nascent strands in red with the junction gap shown as a dotted line. (E) Proportion of total rearrangements corresponding to short-range inversions (<50 bp, U-turn-like rearrangements) in each organelle genome. The mean of the four samples for brain and liver mitochondria is presented. (F) Proportion of short-range inversions (<50 bp) displaying a junction upstream, at the same base, or downstream on the opposite strand in each organelle genome. A representative sample is shown for human brain (ERR419275) and liver (ERR419281) mitochondria. Ara.: *Arabidopsis*, mito.: mitochondrion, del.: deletion, dup.: duplication.

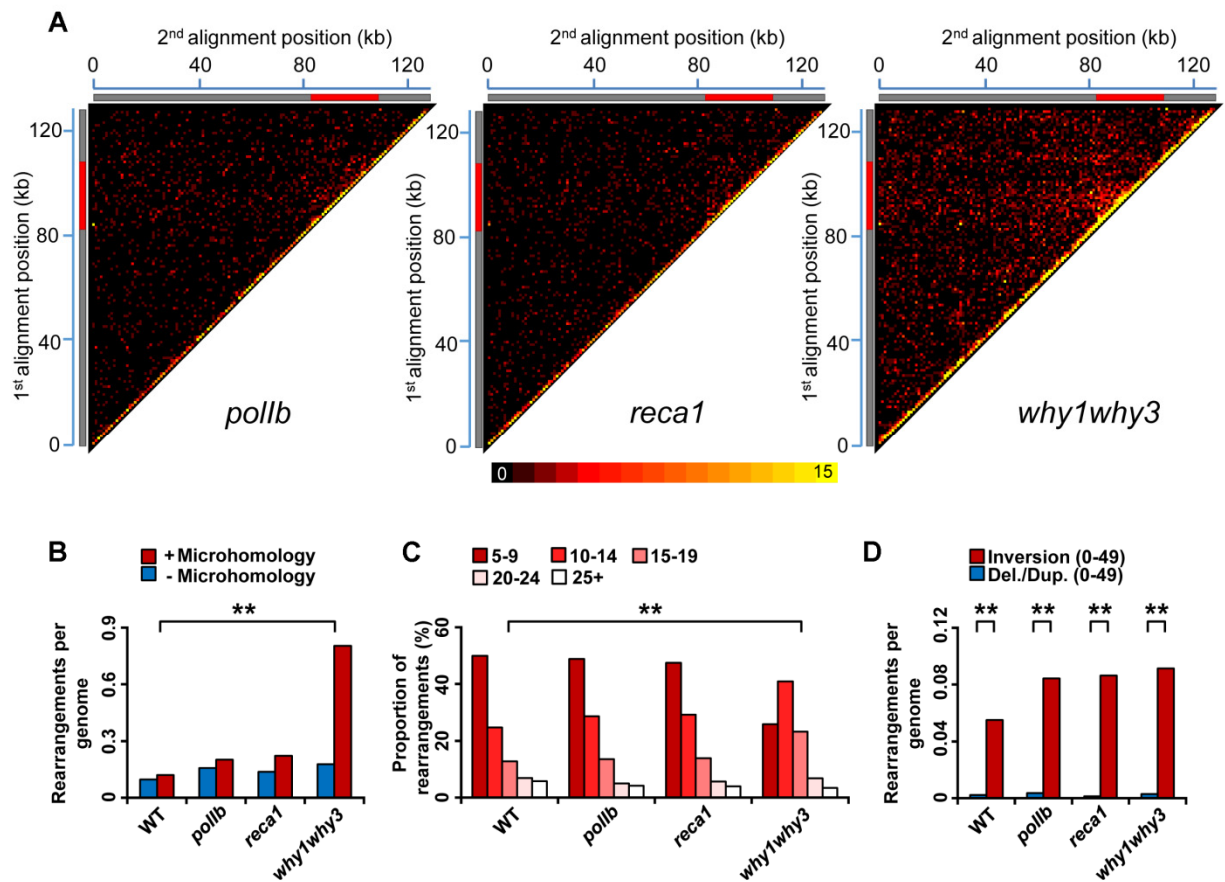


Figure 33. Global portrait of plastid genome rearrangements in *Arabidopsis* mutant lines *pollb*, *reca1* and *why1why3*. (A) Plastid rearrangement breakpoint positions of the indicated *Arabidopsis* mutant lines. Heat maps depict each rearrangement as the intersection of the two genomic positions corresponding to the nucleotide on each side of the junction. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long grey bar, a red bar and a short grey bar, respectively. (B) Normalized amount of rearrangements per plastid genome displaying a microhomology (≥ 5 bp) (red) or not displaying a microhomology (blue) at their junction for each of the indicated mutant lines. (C) Proportion of total plastid rearrangements displaying a microhomology of given lengths, in bases, in each mutant line. Y axis represents the percentage of usage of each microhomology length. (D) Normalized amount of rearrangements corresponding to short-range inversions (<50 bp, U-turn-like rearrangements, red) and short-range deletions/duplications (<50 bp, blue) in each mutant line. Data were normalized on the total number of plastid genome reads. Del.: deletion, dup.: duplication. Two asterisks indicate a significant difference with p -value < 0.00001 using a chi-square test.

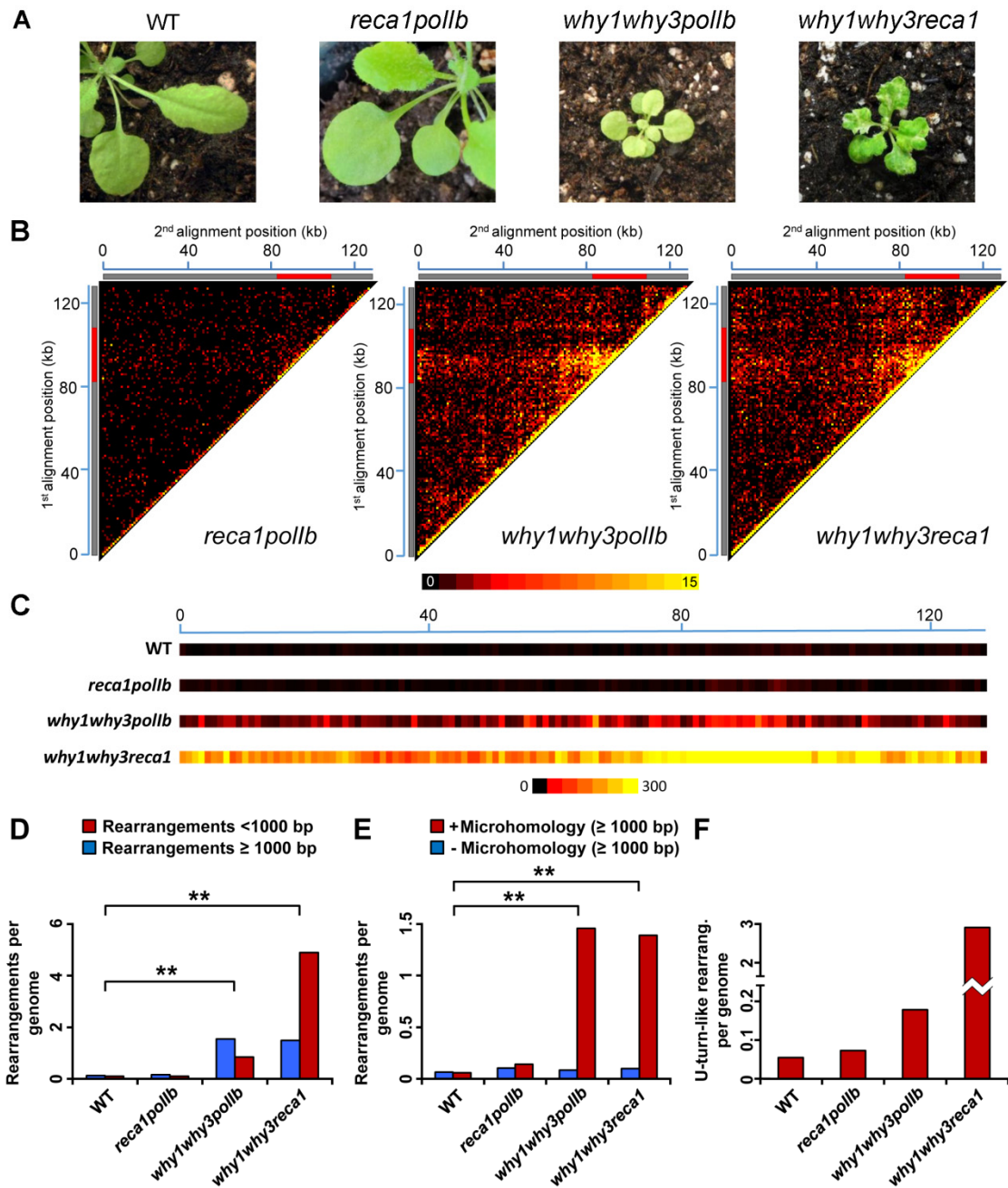


Figure 34. Global portrait of plastid genome rearrangements in *Arabidopsis* mutant lines *reca1pollb*, *why1why3pollb* and *why1why3reca1*. (A) Representative photographs of 21-d-old WT, *reca1pollb*, *why1why3pollb* and *why1why3reca1* *Arabidopsis* mutant plants. (B) Plastid rearrangement breakpoint positions of the indicated *Arabidopsis* mutant lines. Heat maps depict each rearrangement as the intersection of the two genomic positions corresponding to the nucleotide on each side of the junction. Each tile represents a region spanning 1 kb along each axis. Tile intensity represents the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid

large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long grey bar, a red bar and a short grey bar, respectively. (C) Short-range (<1000 bp) plastid rearrangement breakpoint positions of the indicated *Arabidopsis* mutant lines. Each heat map tile represents a region spanning 1 kb along the genome and the number of rearrangements per 10,000 plastid genomes. All rearrangements mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. (D) Normalized amount of short-range (<1000 bp) and long-range (≥ 1000 bp) rearrangements per plastid genome for each of the indicated mutant lines. (E) Normalized amount of long-range rearrangements (≥ 1000 bp) per plastid genome displaying a microhomology (≥ 5 bp) (red) or not displaying a microhomology (blue) at their junction for each of the indicated mutant lines. (F) Normalized amount of rearrangements corresponding to short-range inversions (<50 bp, U-turn-like rearrangements) in each mutant line. Data were normalized on the total number of plastid genome reads. Two asterisks indicate a significant difference with p -value <0.00001 using a chi-square test.

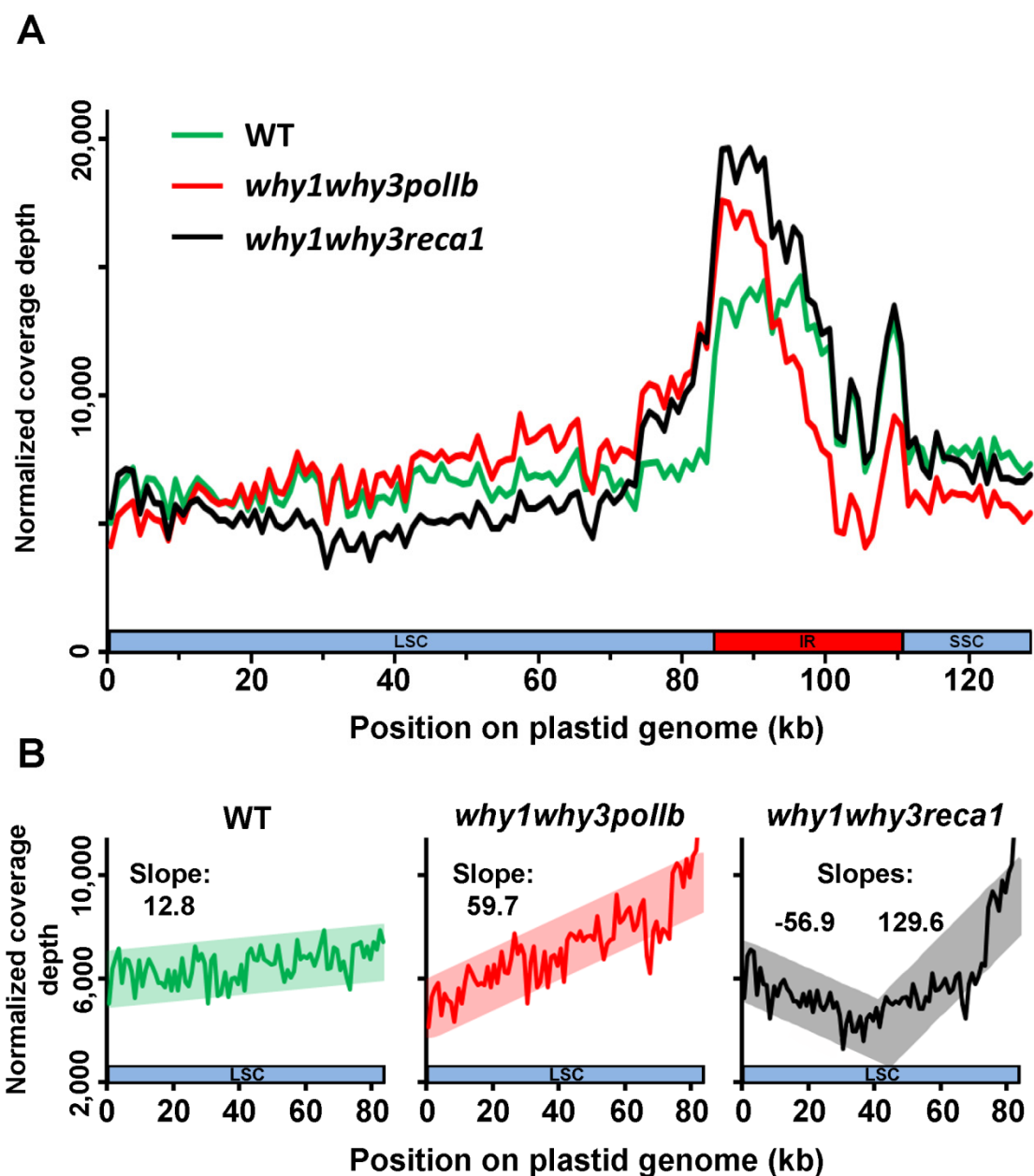


Figure 35. Plastid DNA sequencing coverage curves for *Arabidopsis* lines WT, *why1why3pollb* and *why1why3reca1*. (A) Plastid sequencing coverage of pools of 14-day-old *Arabidopsis* seedlings of the indicated genotypes. Positions were rounded down to 1 kb. All reads mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long blue bar, a red bar and a short blue bar, respectively. Y axis represents the number of reads per 1,000,000 total plastid reads. (B) Regression analysis of the plastid large single-copy (LSC) sequencing coverage of WT, *why1why3pollb*, *why1why3reca1* seedlings. Positions were rounded down to 1 kb. Y axis represents the number of reads per 1,000,000 total plastid reads.

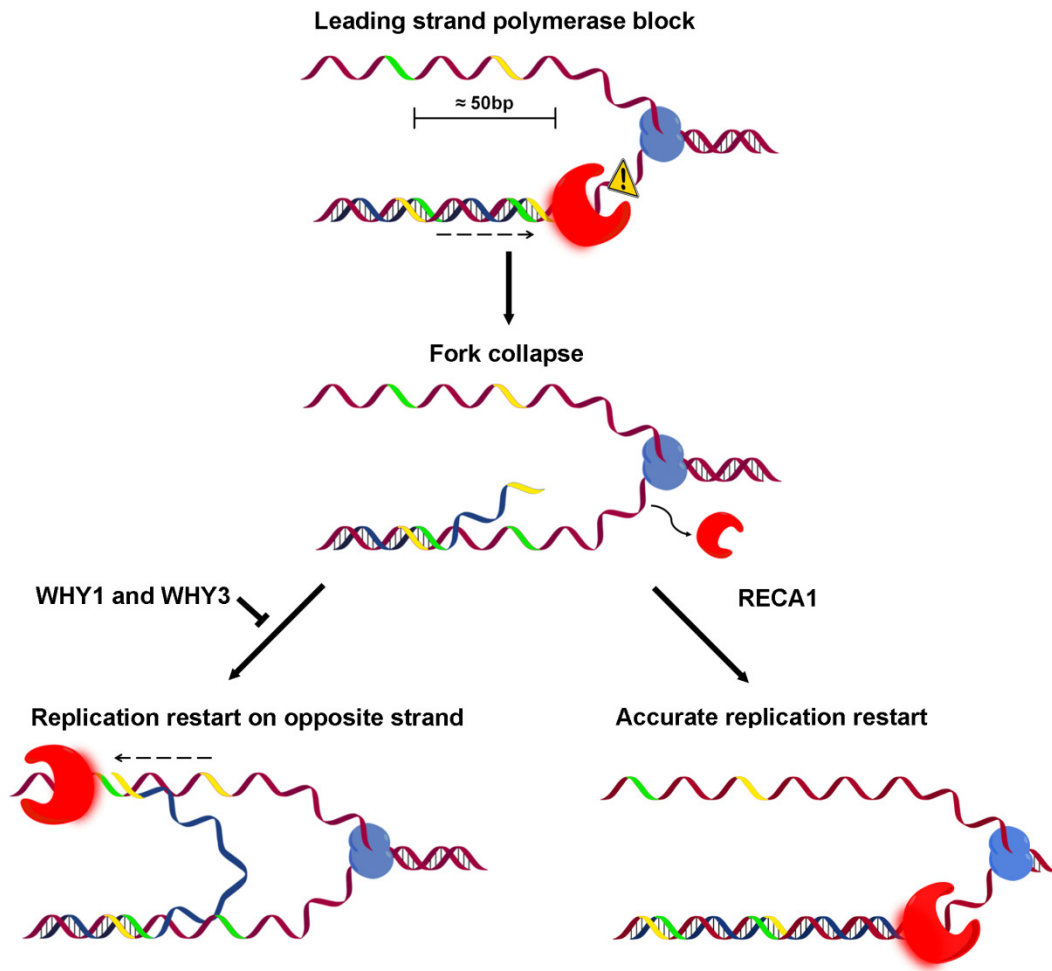


Figure 36. Model of microhomology-mediated U-turn-like inversions. The red shape represents the DNA polymerase and the blue shape represents the DNA helicase. Yellow and green strands represent inverted microhomologies. Impediments to replication fork progression might force the leading strand polymerase to pause and eventually unload. In the presence of RecA, the impediment may be bypassed to restart replication accurately. In the absence of RecA and Whirly proteins, a microhomology located at the 3' end of the nascent strand can anneal to its complementary inverted repeat on the opposite strand and restart the replication fork on the wrong strand. The dotted arrow represents a multistep process.

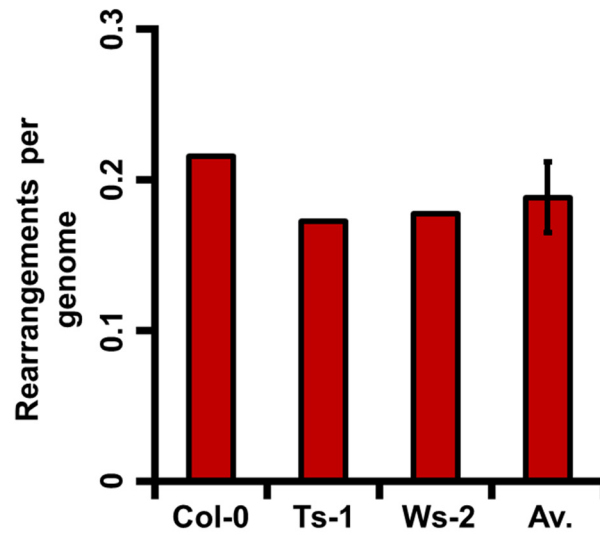


Figure 37. Level of plastid DNA rearrangements in three Arabidopsis ecotypes. Total number of rearrangements identified for each of the Col-0, Ts-1 and Ws-2 ecotypes. Y axis represents the number of rearrangements per 1,000,000 plastid reads. Av.: average of the three ecotypes. The error bar represents the standard deviation.

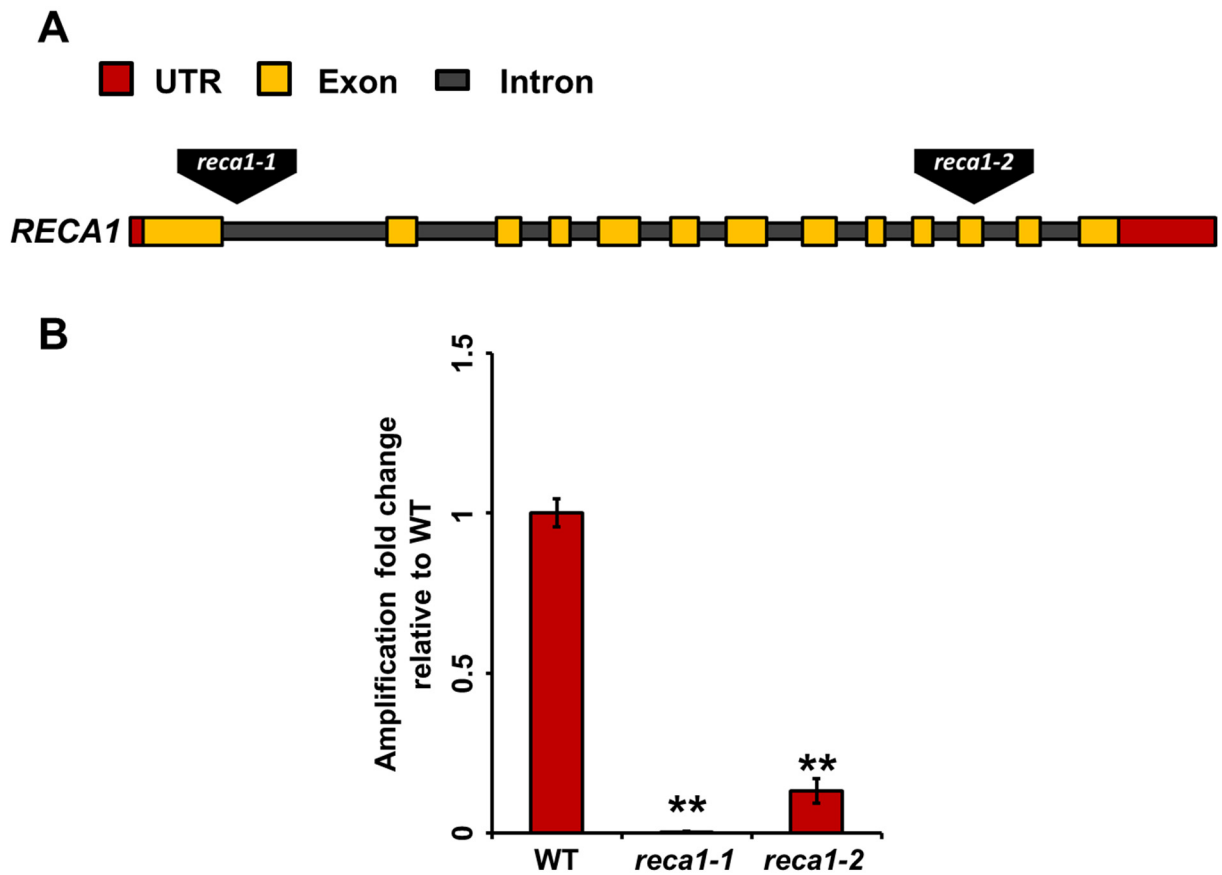


Figure 38. Characterization of the *reca1-1* and *reca1-2* T-DNA insertion mutant lines. (A) Schematic representation of the *RECA1* gene and insertion positions of the *reca1-1* and *reca1-2* insertions. (B) Quantitative PCR measurement of the *RECA1* expression levels for the *reca1-1* and *reca1-2* mutant lines relative to WT plants. Error bars represent the standard error of the mean of three biological replicates. Asterisks indicate a significant difference of a Student's *t* test *p*-value ≤ 0.01 with the WT.

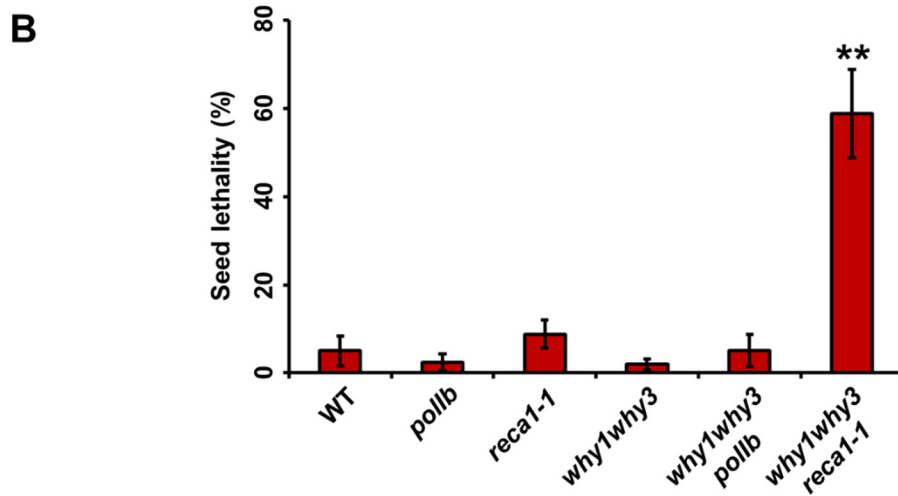
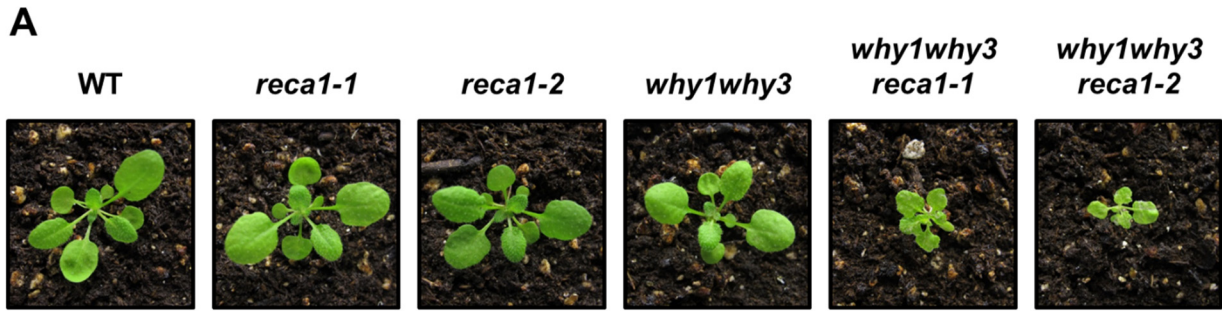


Figure 39. Characterization of the *why1why3reca1* phenotype (A) Representative photographs of 21-day-old WT, *reca1-1*, *reca1-2*, *why1why3*, *why1why3reca1-1* and *why1why3reca1-2* *Arabidopsis* mutant plants. (B) Proportion of non-germinated seeds of WT, *pollb*, *reca1-1*, *why1why3*, *why1why3pollb* and *why1why3reca1-1*, four days after vernalization. Two asterisks indicate a significant difference of a Student's *t* test *p*-value ≤ 0.01 with the WT.

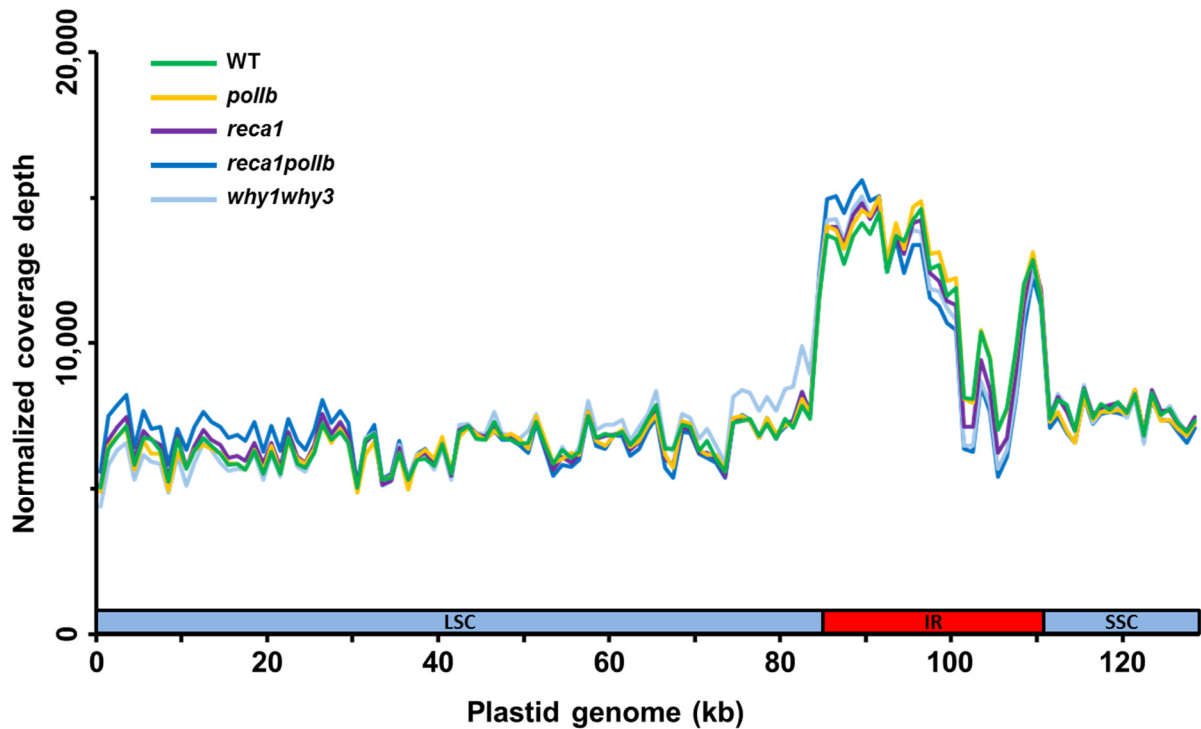


Figure 40. Plastid DNA sequencing coverage curves for WT, *pollb*, *reca1*, *reca1pollb* and *why1why3* Arabidopsis mutant plants. Plastid sequencing coverage of pools of 14-day-old *Arabidopsis* seedlings of the indicated genotypes. Positions were rounded down to 1 kb. All reads mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. Y axis represents the number of reads per 1,000,000 total plastid reads. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long blue bar, a red bar and a short blue bar, respectively.

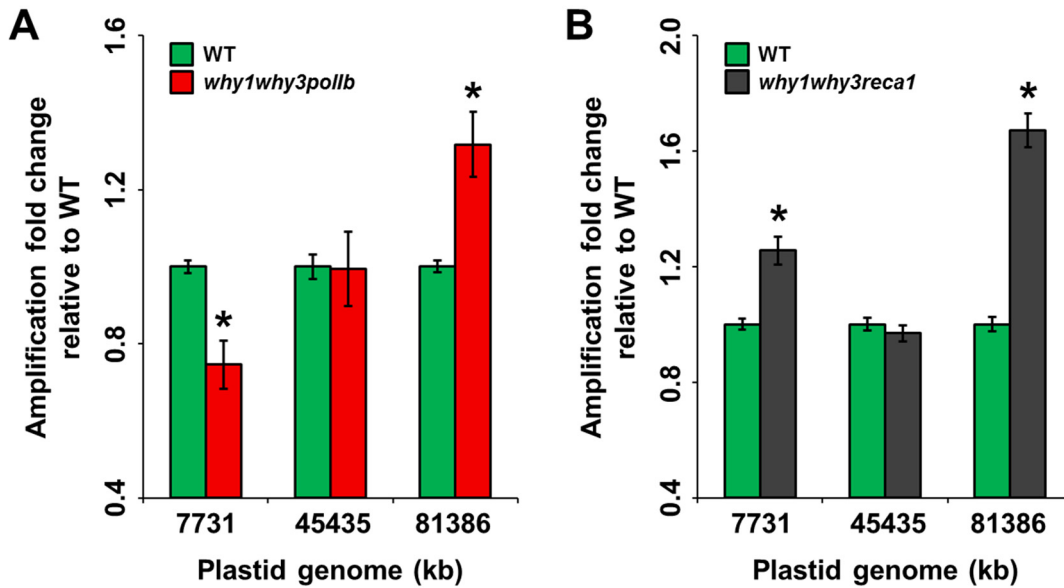


Figure 41. Plastid DNA quantification at three locations of the genome in WT, *why1why3pollb* and *why1why3reca1* plants. (A) Quantitative PCR measurement of ptDNA levels at three sites of the LSC in *why1why3pollb* relative to WT plants. (B) Quantitative PCR measurement of ptDNA levels at three sites of the LSC in *why1why3reca1* relative to WT plants. Error bars represent the standard error of the mean of three biological replicates. Asterisks indicate a significant difference of a Student's *t* test *p*-value ≤ 0.05 with the WT.

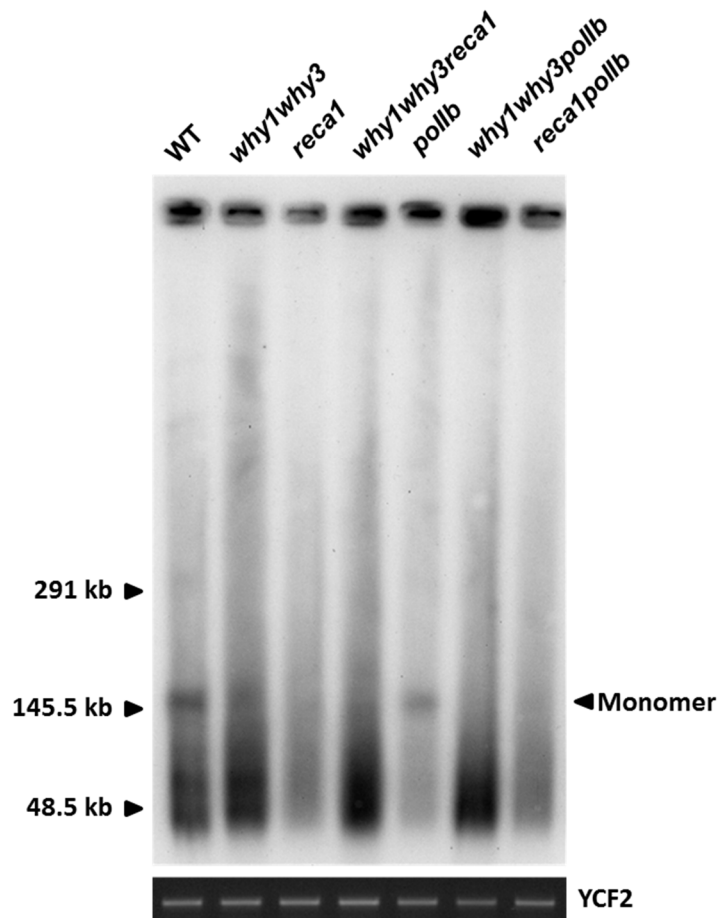


Figure 42. Visualization of the distribution of the various forms of the plastid genome. Pulsed-Field Gel Electrophoresis (PFGE) analysis of ptDNA in the indicated genotypes. The amount of DNA loaded in each well was equilibrated relative to the amplification of a *YCF2* fragment.

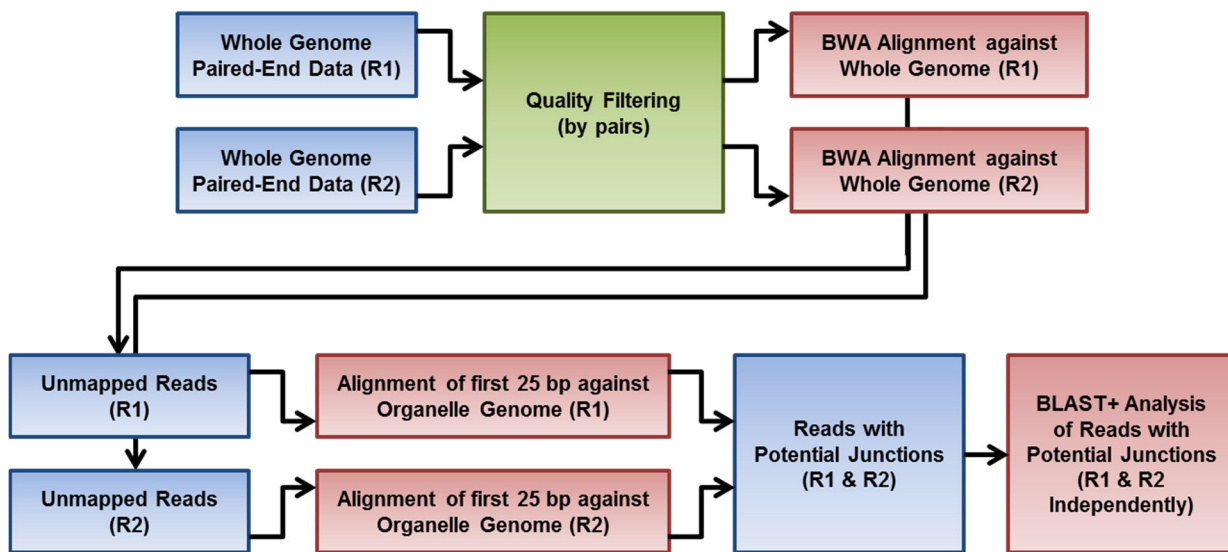


Figure 43. Schematic representation of the analysis workflow. Blue background boxes represent text manipulation steps, while the green and red backgrounds stand for quality filtering and mapping, respectively. Sequencing was performed on both ends of DNA fragments (R1 and R2). BWA: Burrows-Wheeler Aligner. R1 & R2: Paired-end sequencing read1 and read2.

	PCR and Southern Blot	Reporter Systems	Paired-End Analysis	Junction Analysis
Genome-wide	No <i>Primers and probe are specific to a single region</i>	No <i>Only the inserted region is observed</i>	Yes	Yes
Can be performed without genetic modification	Yes	No <i>Requires insertion of exogenous DNA into an organism's genome</i>	Yes	Yes
Allows comparison of samples	Yes	Yes	No <i>Detection is heavily dependent on DNA fragment length, which varies between samples</i>	Yes
Provides information about mechanism	Yes <i>Individual rearrangements need to be cloned and sequenced to obtain sequence</i>	Yes <i>Individual parameters can be varied to assess their importance</i>	No	Yes <i>Exact junction sequence provides some information about mechanism</i>
Provides a view of both short- and long-range rearrangements	No <i>Primers and probe are specific to a single region</i>	No <i>A single mechanism is observed at any time</i>	No <i>Bias toward long-range rearrangements</i>	Yes
Allows detection of short-ranged rearrangements	Yes	Yes	No <i>Detection of short-range rearrangements is limited by DNA fragment size</i>	Yes

Figure 44. Comparison of the techniques used to detect genome rearrangements. Details are provided when relevant.

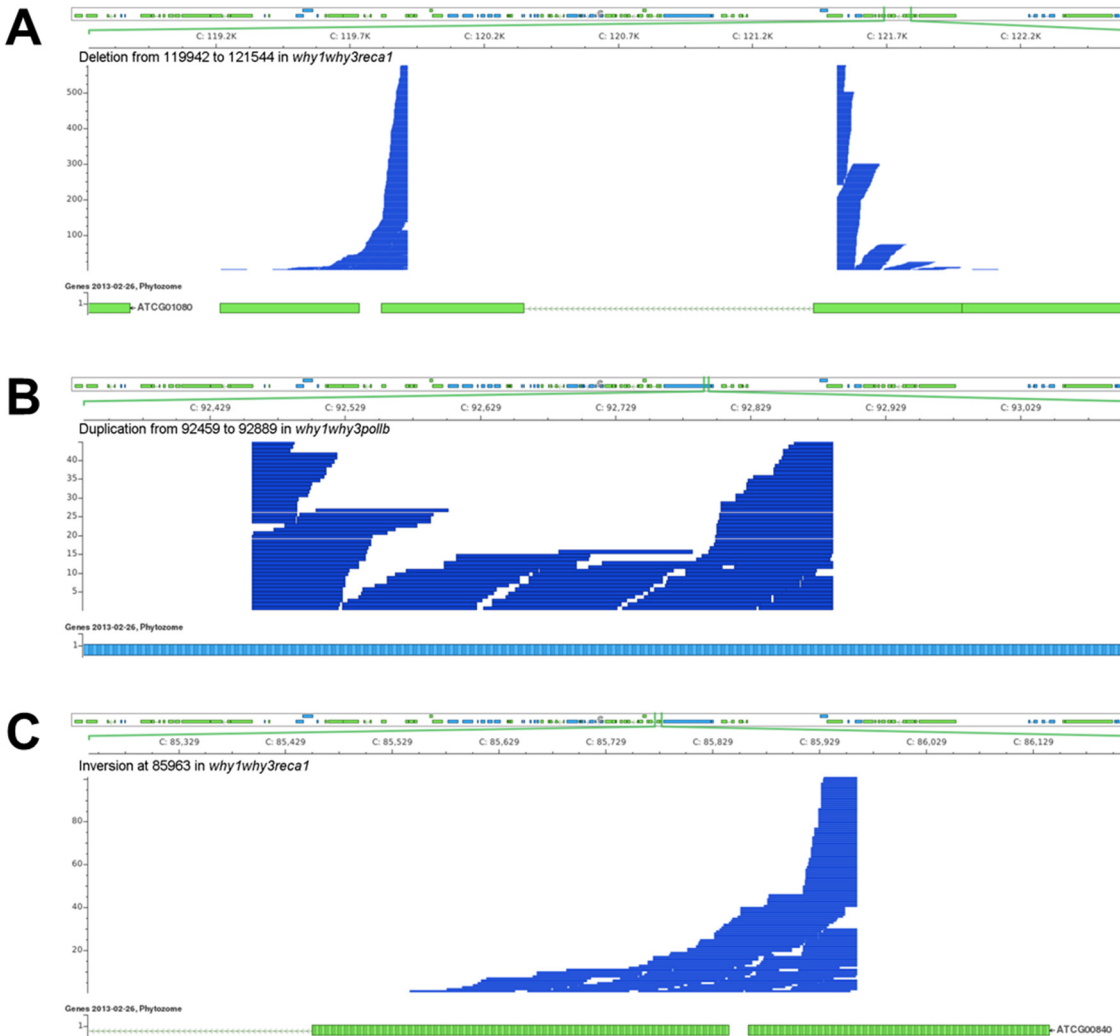


Figure 45. Local mapping of read pairs associated to specific genome rearrangements. The top box in each panel shows a map of the complete plastid genome on which genes on the forward strand are presented in blue and those on the reverse strand in green. The position encompassed by the zoom on the genome are presented below. The mapping of each read is presented as a blue rectangle. The bottom box shows a map of genes in the zoom. A) Representative local mapping for a deletion. This deletion was observed 537 times in *why1why3reca1*. B) Representative local mapping for a duplication. This duplication was observed 77 times in *why1why3pollb*. C) Representative local mapping for an inversion. This inversion was observed 81 times in *why1why3reca1*.

	Col-0	<i>pollb</i>	<i>reca1</i>	<i>reca1pollb</i>	<i>why1why3</i>	<i>why1why3</i> <i>pollb</i>	<i>why1w</i> <i>reca</i>
Read Pairs	31241719	32810928	30409865	31132957	31438219	32941816	31699
Pairs with average quality >20	28771745	31343328	28967978	29738528	29944699	31308538	30156
Plastid Pairs (%)	17,69	16,77	19,51	13,35	20,35	19,16	20,5

Tableau II Workflow Statistics for *Arabidopsis* plastid DNA rearrangements.

	Ts-1 (SRX145018)	Ws-2 (SRX145037)
Read Pairs	30012578	45521342
Pairs with average quality >20	26774195	40165795
Plastid Pairs (%)	9,89	9,54

Tableau III Workflow Statistics for *Arabidopsis* ecotypes Ts-1 and Ws-2 plastid DNA rearrangements.

	Col-0
Read Pairs	31241719
Pairs with average quality >20	28771745
Mitochondria Pairs (%)	1,27

Tableau IV Workflow Statistics for *Arabidopsis* mitochondria DNA rearrangements.

	erx385572	erx385573	erx385574	erx385575
Read Pairs	78908272	79738231	54905866	53255234
Pairs with average quality >20	54109701	54783744	52966448	51137017
Mitochondria Pairs (%)	0,55	0,55	0,61	0,64

Tableau V Workflow Statistics for human brain mitochondria DNA rearrangements.

	erx385576	erx385577	erx385578	erx385579
Read Pairs	74027979	75446145	47333976	46047382
Pairs with average quality >20	51040536	52250953	45679517	44520994
Mitochondria Pairs (%)	0,45	0,49	0,52	0,53

Tableau VI Workflow Statistics for human liver mitochondria DNA rearrangements.

	SRX154301	SRX154337	SRX154338	SRX154342
Read Pairs	2611112	2611112	2611112	2611112
Pairs with average quality >20	2611099	2611096	2611098	2611089
<i>E. coli</i> Pairs (%)	99,85	99,83	99,85	99,84

Tableau VII Workflow Statistics for *E. coli* DNA rearrangements.

Chapitre 5: Troisième article

Plastid genome instability leads to reactive oxygen species production and plastid-to-nucleus retrograde signaling in Arabidopsis.

Article publié dans le journal *Plant Physiol* **163**:867-81 (2013)

Copyright American Society of Plant Biologists

5.1 Mise en contexte:

Les réarrangements génomiques peuvent avoir des effets variés selon leur abondance et l'identité des gènes affectés. L'instabilité génomique a été associée à l'apparition de nombreuses pathologies chez l'humain, menant entre autres à différents types de cancer et à diverses maladies héréditaires (revue dans (78, 192, 193)). Au niveau des organelles, les modifications du génome mitochondrial ont aussi été associées à diverses problématiques cellulaires, tels le débalancement du métabolisme des ROS et la diminution de l'efficacité de la phosphorylation oxydative (194). Récemment, des études sont même venues relier les niveaux de réarrangements génomiques dans les mitochondries à la tumorigénicité de certains cancers (195). Néanmoins, chez les plantes, l'impact de l'instabilité génomique demeure très peu caractérisé dans les organelles. L'identification de nombreux mutants chez *Arabidopsis* ayant des génomes plastidiques moins stables peut servir de base à la compréhension de l'impact physiologique de l'instabilité génomique dans le plastide. Les résultats présentés dans ce chapitre focaliseront sur les mutants *why1why3* et *why1why3pollb* possédant des niveaux croissants de réarrangements génomiques. De façon intéressante, les niveaux élevés de réarrangements génomiques observés chez *why1why3pollb* corréleront avec d'importants problèmes photosynthétiques et nous désirons comprendre davantage les causes et les conséquences de ceux-ci.

Dans cette étude, j'ai participé aux expériences qui ont permis de déterminer la corrélation entre l'accumulation de réarrangements génomiques dans le plastide et le débalancement de la production d'espèces réactives oxygénées dans cette organelle. J'ai également participé aux expériences de transcriptomique pour évaluer la réponse nucléaire des plantes ayant des niveaux plus élevés de réarrangements génomiques dans le chloroplaste. EZ a contribué à la réalisation des expériences de quantification des réarrangements génomiques, d'évaluation du stress photosynthétique et aux expériences de séquençage de l'ARN. L'analyse des résultats, la rédaction de l'article et l'élaboration des figures a été un travail commun entre NB, EZ et moi.

Plastid Genome Instability Leads to ROS Production and Plastid-to-Nucleus Retrograde Signaling in Arabidopsis

Étienne Lepage¹, Éric Zampini¹, Normand Brisson²

Department of Biochemistry, Université de Montréal, Montréal, Québec, Canada H3C 3J7.

¹ These authors contributed equally to this work

One-sentence summary: Instability of the plastid genome leads to the production of reactive oxygen species and to nuclear genetic reprogramming.

Funding:

E.Z. and E.L. were funded by scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC). This work was supported by a grant from NSERC to N.B.

5.2 ABSTRACT

The plastid genome is highly conserved among plant species, suggesting that alterations of its structure would have dramatic impacts on plant fitness. Nevertheless, little is known about the direct consequences of plastid genome instability. Recently, it was reported that the plastid Whirly proteins, WHY1 and WHY3, and a specialized type-I polymerase, POLIB, act as safeguards against plastid genome instability in *Arabidopsis thaliana*. In the present study, we use ciprofloxacin, an organelle double-strand break inducing agent, and the *why1why3pollb-1* variegated mutant to evaluate the impact of generalized plastid DNA instability. First, we show that in *why1why3pollb-1* and ciprofloxacin treated plants, plastid genome instability is associated to increased ROS production. Then, using different light regimens, we show that the elevated ROS production correlates with the appearance of a yellow-variegated phenotype in the *why1why3pollb-1* population. This redox imbalance also correlates to modifications of nuclear gene expression patterns, which in turn leads to acclimation to high light. Taken together, these results indicate that plastid genome instability induces an oxidative burst that favors, through nuclear genetic reprogramming, adaptation to subsequent oxidative stresses.

5.3 INTRODUCTION

As opposed to most non-photosynthetic eukaryotes, plants possess an additional endosymbiotic organelle, the plastid, which is the seat of photosynthesis. In addition to this essential process, plastids have a variety of other important biological roles. Of these, chloroplasts serve as sensors of environmental conditions and guide the plant toward proper adaptation (Pfannschmidt et al., 1999). Indeed, as a result of its light-harvesting capacity, the chloroplast is directly associated with light-absorption and, accordingly, is ideally placed to detect environmental fluctuations (Ruckle et al., 2012). The fluctuating environmental conditions are perceived through several important molecular checkpoints, such as the redox state of the photosynthetic electron transport chain (PET), often measured as the excitation pressure of the photosystem II (PSII), or the concentration of reactive oxygen species (ROS) produced by the PET (Huner et al., 2012).

Even though photosynthesis serves as the main energy source for autotrophic organisms, this process also constantly generates ROS by-products that threaten all cellular components by their strong propensity for oxidation (Asada, 2006). ROS production is notably exacerbated by abiotic stresses, such as a sudden exposure to high light intensities, which leads to excessive photon absorption, over-reduction of the plastoquinone (PQ) pool, and electron transfer to molecular oxygen. If improperly scavenged, ROS can have a variety of deleterious effects such as base oxidation, lipid peroxidation, and in some cases cell death (Przybyla et al., 2008; Triantaphylides et al., 2008).

Beyond their deleterious effects, ROS also act as important signaling molecules, especially under stress conditions. Indeed, ROS-dependent signaling has been observed in the vast majority of biotic and abiotic stress responses (Mittler, 2002; Miller et al., 2008; Torres, 2010). Recently, different studies have highlighted specific reprogramming of the nuclear transcriptome depending on the site of production within the cell and the type of ROS involved (Laloi et al., 2007; Maruta et al., 2012; Ramel et al., 2013). When produced in the chloroplast, ROS initiate plastid-to-nucleus retrograde signaling and modify nuclear and chloroplast gene expression patterns to adjust the photosynthetic activity. For instance, transcriptome alterations

during chloroplast oxidative stress often include repression of the light-harvesting complexes and accumulation of high levels of antioxidant proteins (Escoubas et al., 1995; Grace and Logan, 1996).

The negative correlation between PET efficiency and chloroplast ROS production suggests that any disequilibrium in PET homeostasis can lead to redox imbalance. Accordingly, mutations of nuclear genes involved in the quality control of the PET proteins, such as the FtsHs and Deg proteases, induce developmental defects and redox imbalance (Zaltsman et al., 2005; Stroher and Dietz, 2008; Kato et al., 2009; Sun et al., 2010). The chloroplast gene expression status acts as an additional checkpoint to ensure PET homeostasis. Indeed, although plastid DNA (ptDNA) only encodes a small fraction of chloroplast proteins, the retained plastid genes are mostly involved in photosynthesis (Green, 2011). Thus, inhibition of chloroplast gene expression induced by mutations or chemical treatments leads to impairment of photosynthesis and increased chloroplast ROS production (Myouga et al., 2008; Arsova et al., 2010; Kindgren et al., 2012). This suggests that the maintenance of plastid genome stability is crucial to ensure PET efficiency. However, although the function of many nuclear genes involved in chloroplast DNA repair is emerging, little is known about the direct outcome of generalized plastid genome instability on chloroplast function.

Among the different types of DNA damages leading to genome instability, double-strand breaks (DSBs) are considered the most threatening since they can give rise to severe chromosomal rearrangements if improperly repaired (Odom et al., 2008). This type of damage can spontaneously arise in the chloroplast genome as a consequence of replication fork collapse or can be specifically induced by ciprofloxacin (CIP), a gyrase inhibitor that produces DSBs in a replication-dependent manner (Drlica and Zhao, 1997; Parent et al., 2011). In *Arabidopsis*, PolIA and PolIB, two type-I DNA polymerases, are proposed to perform chloroplast DNA replication (Mori et al., 2005; Parent et al., 2011). The mutation of either of the genes encoding these proteins is enough to cause replication stress and PolIB has been demonstrated to have additional roles in DNA repair (Parent et al., 2011) and mitochondrial homeostasis (Cupp and Nielsen, 2012). In addition to PolIB, plant organelles possess Whirly proteins that act as safeguards of plastid genome stability by binding and stabilising single-stranded DNA, thereby limiting the occurrence of illegitimate recombination (Maréchal et al.,

2009; Cappadocia et al., 2010). Arabidopsis mutants for the two plastid Whirly genes, *WHY1* and *WHY3*, are sensitive to CIP and exhibit ptDNA rearrangements that are thought to occur through a microhomology-mediated break-induced replication (MMBIR) mechanism (Hastings et al., 2009; Maréchal et al., 2009; Cappadocia et al., 2010). Interestingly, in the *why1why3* double mutant, the absence of chloroplast Whirly proteins causes variegation in about 5 % of the plant population (Maréchal et al., 2009). Moreover, the mutation of *PollB* has, together with the mutations of *WHY1* and *WHY3*, a synergistic effect on the appearance of variegation and accumulation of ptDNA rearrangements. The *why1why3pollb-1* triple mutant displays a 100 % penetrance yellow-variegated phenotype and accumulates even larger amounts of MMBIR-associated rearrangements than *why1why3* plants (Parent et al., 2011).

Here we report that ptDNA instability, induced by two distinct approaches, leads to an altered cellular redox status. We show that CIP treatment, which induces DSBs in the chloroplast, and the *why1why3pollb-1* mutations lead to decreased PET efficiency and elevated chloroplast ROS production in Arabidopsis. Our results also reveal that ptDNA rearrangements and increased ROS levels in *why1why3pollb-1* trigger chloroplast-to-nucleus retrograde signaling and overexpression of many abiotic stress-related genes. Finally, we show that this ROS-dependent genetic reprogramming promotes an acquired adaptation to high light.

5.4 RESULTS

Plastid Genome Instability Negatively Impacts Photosynthetic Efficiency

Leaf variegation has been observed in many different mutants of Arabidopsis. The severity of this phenotype varies among these mutants and is sometimes associated to dwarfism (Rédei, 1963; Martínez-Zapater, 1993; Maréchal et al., 2009; Parent et al., 2011; Xu et al., 2011). In Arabidopsis, the variegation observed in *why1why3* and *why1why3pollb-1* (Fig. 46A-B) has been linked to DNA rearrangements produced by the MMBIR pathway in the chloroplast genome (Fig. 55) (Maréchal et al., 2009; Parent et al., 2011). While the *why1why3* plants present a moderate amount of MMBIR-associated rearrangements, these rearrangements are present at much higher levels in the triple mutant *why1why3pollb-1* (Parent et al., 2011) (Fig. 46C). Interestingly, the high levels of rearrangements in *why1why3pollb-1* do not

exacerbate the *why1why3* white variegation phenotype. Instead, a distinct yellow-variegated phenotype is observed in *why1why3pollb-1* (Fig. 46B and Fig. 56). The fact that *why1why3pollb-1* also presents white variegation in an equivalent fraction of its population further supports that white and yellow variegation are distinct phenotypes (Fig. 57).

Although the roles of Why1, Why3 and PolIB in plastid genome stability are becoming well established, the consequences of the ptDNA rearrangements associated with their mutations are still poorly understood. Since the plastid genome encodes many components of the PET, it is conceivable that plastid genome instability could impair photosynthesis. Therefore, to evaluate the potential impacts of MMBIR-associated ptDNA rearrangements on the PET photosynthetic capacity, the efficiency of the PSII was measured in wild type, *pollb-1*, *why1why3* and *why1why3pollb-1* plants grown for four weeks under normal light conditions. Measurement of the Fv/Fm ratio, a widely used parameter for the measurement of the PSII efficiency, indicates that the *why1why3pollb-1* mutant exhibits a significant decrease in its photosynthetic capacity (Fig. 46D). To confirm the link between ptDNA rearrangements and photosynthetic defects, DSBs were introduced in the organelle genomes of WT plants by treatment with ciprofloxacin (CIP). Similarly to *why1why3pollb-1*, wild type plants treated with CIP presented a significant decrease in their PSII efficiency as revealed by a lower Fv/Fm ratio (Fig. 58). No such decrease was observed when the plants were treated with novobiocin (NOVO) (Fig. 58), a gyrase inhibitor that induces neither DSBs nor ptDNA rearrangements (Hardy and Cozzarelli, 2003; Cappadocia et al., 2010). Taken together, the lowered Fv/Fm ratios observed in *why1why3pollb-1* and upon CIP treatment clearly suggest that ptDNA rearrangements negatively impact PET performance.

Plastid Genome Instability Leads to Generation of ROS in the Chloroplast

Since PET defects and imbalance are often associated with increased ROS production (Zaltsman et al., 2005; Kato et al., 2009), we verified whether the ptDNA rearrangements observed in Figure 46C are also associated with oxidative stress in the chloroplasts. Leaves from four-week-old Arabidopsis plants were incubated in the presence of the general oxidative stress fluorescent probe H₂DCFDA. Microscopic detection of the oxidized fluorescent form of the molecule revealed ROS-generating foci in *why1why3pollb-1* plants, but not in wild type, *why1why3* and *pollb-1* plants (Fig. 47A). Moreover, the fluorescent H₂DCFDA colocalizes

with chlorophyll and structures resembling plastids, suggesting that chloroplasts are the origin of this redox imbalance (Fig. 47B). To further characterize ROS production in *why1why3pollb-1*, detached leaves from four-week-old plants were incubated in the dark in the presence of the colorimetric agent nitroblue tetrazolium (NBT), which forms an insoluble quantifiable blue precipitate, the formazan, upon reduction from superoxide (O_2^-) (Driever et al., 2009). While the color intensity of *pollb-1* and *why1why3* leaves was indistinguishable from that of wild type, *why1why3pollb-1* leaves showed a dense shadowed blue coloration all over their surface (Fig. 47C, Dark). Measurement of the formazan produced in those leaves revealed an approximate two fold increase in superoxide concentration as compared to WT leaves (Fig. 47D). ROS concentration was further analysed by switching the dark-incubated leaves to high light intensity. As expected, this treatment induced ROS production in the leaves of every line tested (Fig. 47C, Dark \rightarrow Light and 2E). However, the increase was more pronounced in the *why1why3pollb-1* mutant, further suggesting that the PET is involved in the redox imbalance of this mutant.

To verify that the elevated ROS levels in *why1why3pollb-1* are truly the consequence of the ptDNA rearrangements, ROS production was assessed in wild type plants treated either with CIP or NOVO. As depicted in Figure 48A, both treatments effectively affected plant development. Indeed, whereas the NOVO treatment arrested plant growth, the CIP treatment bleached the sides of the new emerging leaves (Fig. 48A). Treated plants were then incubated in presence of H_2DCFDA to assess the redox imbalance. Similarly to the *why1why3pollb-1* mutant, the CIP treatment also led to the formation of ROS foci that colocalized with chloroplasts at the subcellular level (Fig. 48B and 48C). No fluorescence could be observed in the NOVO-treated plants, suggesting that the redox imbalance observed is caused by ptDNA rearrangements, and not by gyrase inhibition.

The *why1why3pollb-1* Phenotype Depends on Elevated ROS Levels

Considering that the redox imbalance observed in *why1why3pollb-1* could be responsible for the appearance of the yellow-variegated phenotype, it is likely that decreasing the electron flow across the PET would rescue the phenotype. Therefore, to minimize the pressure on the electron transport chain, plants were grown for three weeks at low light and their phenotypes compared to those grown at normal light. When grown for three weeks under

normal light conditions, *why1why3pollb-1* plants developed a severe dwarf and yellow-variegated phenotype (Fig. 46B and Fig. 49A, NL) that is associated with low chlorophyll content (Fig. 49B). In comparison, although all seedlings grown at low light for three weeks exhibited severe growth retardation, no additional growth delay and yellow variegation could be observed for *why1why3pollb-1* (Fig. 49A, LL 21 days). The rescue of the yellow-variegated phenotype persisted even after sixty days of growth under the same conditions, although at this stage the plants had smaller petioles than wild type and *why1why3* plants (Fig. 49A, LL 60 days). Consistent with the rescue of the phenotype, low light grown *why1why3pollb-1* plants exhibited chlorophyll content comparable to that of control plants (Fig. 49B). This low light dependent rescue of the phenotype thus suggests that elevated ROS levels play a role in the appearance of the *why1why3pollb-1* yellow-variegated phenotype. In agreement with this, *why1why3* and *why1why3pollb-1* plants are hypersensitive to paraquat and, therefore, affected by an aggravation of the redox imbalance (Fig. 59).

In contrast to the observed rescue of the phenotype, low light conditions did not restore PSII efficiency in *why1why3pollb-1* plants (Fig. 49C). This is in agreement with our hypothesis that the poor PET efficiency is due to ptDNA rearrangements, the abundance of which is also little affected by the light regimen (Fig. 60). Thus, it seems likely that ptDNA rearrangements do not originate from ROS-associated DNA modification. Consistent with this idea, the ROS-inducing agents paraquat and atrazine do not lead to an increased accumulation of MMBIR-associated ptDNA rearrangements (Fig. 61).

Yellow-Variegated Sectors in *why1why3pollb-1* Appear Photosynthetically Active and Are not Related to Cell Death

ROS readily damage all components of the cell and induce cell death if improperly scavenged (Mittler, 2002; Przybyla et al., 2008; Triantaphylides et al., 2008). To verify if the high ROS concentration in the *why1why3pollb-1* mutant is associated with cell death, four-week-old leaves were stained with trypan blue to selectively color dead cells (Weigel and Glazebrook, 2002). Interestingly, although the *why1why3pollb-1* mutant produces high levels of ROS, cell death did not appear higher in leaves of this mutant than in leaves of wild type plants (Fig. 62). This result suggests that the *why1why3pollb-1* yellow-variegated phenotype is not the direct consequence of ROS-associated cell death. To verify if chloroplasts in

why1why3pollb-1 appear functional, chloroplast ultrastructure was examined in green and yellow sectors of this mutant using transmission electron microscopy. Remarkably, despite the redox imbalance, thylakoid stacks appear normal in both green and yellow sectors of *why1why3pollb-1* (Fig. 50). In comparison, chloroplast ultrastructure is completely altered in white variegated sectors of *why1why3* (Fig. 50A) (Maréchal et al., 2009). Furthermore, whereas the white-variegated sectors of *why1why3* present no trace of chlorophyll, the yellow-variegated sectors of *why1why3pollb-1* are not devoid of photosynthetically active chloroplasts, as monitored by chlorophyll autofluorescence (Fig. 50B). Altogether, these results suggest that chloroplasts in *why1why3pollb-1* are still functional despite the redox imbalance.

The *why1why3pollb-1* Yellow-Variegated Phenotype and Rearrangements Are not Maternally Inherited

The appearance of variegation in many mutants has been linked to maternal transmission of “permanently-defective” chloroplasts (Yu et al., 2007). However, the fact that chloroplasts in *why1why3pollb-1* appear functional and that the phenotype does not worsen in a generation-dependent manner suggests that its yellow-variegated phenotype is not maternally inherited. This was tested by backcrossing the *why1why3pollb-1* mutant with wild type plants and examining the transmitted phenotype. For each of the reciprocal crosses, 72 plants from five distinct crosses were examined. As expected, when *why1why3pollb-1* plants were backcrossed as male parents, no transmission of the phenotype was observed (Fig. 51A). Consistent with the idea that chloroplasts are functional in this mutant, the phenotype also disappeared in all plants issued from the reciprocal maternal crosses. Figure 51B indicates that ptDNA rearrangements were also eliminated in two-week-old plants of the F1 progeny, in agreement with the lack of inheritance of the phenotype in each reciprocal cross. Restoration of the alleles thus probably prevents generation of *de novo* ptDNA rearrangements in addition to allowing dilution of the inherited rearrangements throughout the entire plant. These results further support the idea that *why1why3pollb-1* chloroplasts function normally despite the redox imbalance.

***why1why3pollb-1* Exhibits a Photosynthetic Stress-Related Nuclear Genetic Reprogramming**

We then tested whether the elevated ROS levels could lead to nuclear genetic reprogramming in *why1why3pollb-1*, enabling the plant to adapt to its particular photosynthetic conditions. To test this hypothesis, a RNA-Sequencing (RNA-Seq) approach was used to sequence and quantify the entire *why1why3pollb-1* nuclear transcriptome. Comparison of this transcriptome to that of wild type plants revealed a complex genetic reprogramming with 340 and 157 transcripts up and down-regulated at least two-fold, respectively (http://genome.cshlp.org/content/suppl/2015/03/20/gr.188573.114.DC1/Supplemental_Data_File_2.xlsx).

To characterize gene expression reprogramming in *why1why3pollb-1*, its entire nuclear transcriptome was subjected to MapMan analysis and the biological processes with the most significant alterations were identified (Thimm et al., 2004). Regulation of transcription and RNA processing were the most significant terms, closely followed by photosynthesis and, particularly, the light reactions (Fig. 52A). This was followed by metabolism, especially that of carbohydrate metabolism, which is not surprising considering that photosynthesis is affected in *why1why3pollb-1* (Fig. 46D). Among the other terms significantly overrepresented were the ascorbate and glutathione-associated redox processes as well as the abiotic stress responses. Interestingly, ethylene signal transduction was significantly overrepresented, similarly to a number of abiotic stresses that are also associated with redox imbalance (Ecker, 1995; Miller et al., 2008; Cheng et al., 2013).

To independently characterize genes up and down-regulated at least two fold in *why1why3pollb-1*, these two subsets of genes were analysed using Pageman (Usadel et al., 2006). Similarly to the MapMan analysis of the entire transcriptome, assignment of each of the up-regulated genes to a given biological process mainly resulted in the terms RNA, protein and signaling (Fig. 52B). Regarding the down-regulated genes, the terms hormone and stress were strongly represented in addition to genes of various functions, the miscellaneous class, which accounts alone for more than half of the genes. However, the main terms found using Pageman belongs to large classes containing large numbers of genes. To avoid this bias toward the largest classes, enrichment of each biological process was calculated relative to its class size

(Fig. 52C). Interestingly, this normalization revealed a net enrichment for the terms hormone, stress, redox, nucleotide metabolism, RNA and signaling in the up-regulated genes. The down-regulated genes showed a clear enrichment for the terms photosynthesis, hormone and stress. Globally, a net enrichment for biological processes linked to photosynthetic stress is found within the most differentially expressed genes in *why1why3pollb-1*.

The *why1why3pollb-1* Nuclear Genetic Reprogramming Closely Resembles ROS, Abiotic Stress and *msh1* Responses

Numerous stresses such as chloroplast redox imbalance, abiotic stress and chloroplast biogenesis defects have been shown to initiate plastid-to-nucleus retrograde signaling pathways and modify the nuclear genetic expression patterns. To assess if the genetic reprogramming observed in *why1why3pollb-1* is similar to that observed in other conditions known to alter the nuclear transcriptome, the expression level of the fifty most overexpressed genes in this mutant was compared to that observed in other transcriptomic studies (Fig. 52D). The *why1why3pollb-1* transcriptome was first compared to that of other conditions known to produce specific types of ROS, including those of the *flu* mutant, which exhibits elevated levels of singlet oxygen upon a dark to light shift (op den Camp et al., 2003), of the treatment of Col-0 with paraquat, which produces superoxide in the PET (Chia et al., 1982; Ramel et al., 2009), and of Col-0 treated with hydrogen peroxide. Interestingly, of the 43 overexpressed *why1why3pollb-1* genes found on the ATH1 chip (probes for seven of the *why1why3pollb-1* overexpressed genes are not present on this chip), 26, 18 and 15 were also overexpressed in *flu*, paraquat treatment and hydrogen peroxide treatment, respectively.

Several of the 50 most overexpressed genes in *why1why3pollb-1* encode key components of the abiotic stress responses, including the Early Light-Inducible Protein ELIP1, the transcriptional regulators ZAT10 and ZAT12, the ERF/AP2 transcription factor CBF2, and the mRNA deadenylases CAF1A and CAF1B (Gilmour et al., 1998; Hutin et al., 2003; Sakamoto et al., 2004; Davletova et al., 2005; Mittler et al., 2006; Walley et al., 2010). Not surprisingly, a striking similarity was observed when comparison was made with overexpressed genes in abiotic stress responses. Indeed, 20, 26 and 22 genes were also up-regulated upon high light, cold and drought stresses, respectively (Fig. 52D) (Vogel et al.,

2005; Gonzalez-Perez et al., 2011). A total of 15 genes that are overexpressed during all three abiotic stresses were also up-regulated in *why1why3pollb-1*.

Finally, since most variegated mutants are also linked to an alteration of the PET, we were interested in verifying the similitudes between the *why1why3pollb-1* transcriptome and those of the variegated mutants *msh1*, *var2* and *immutans* (Aluru et al., 2009; Miura et al., 2010; Xu et al., 2011). Although the *var2* and *immutans* mutants also destabilize the PET, only limited similarity was observed between their nuclear transcriptome and that of *why1why3pollb-1* (Fig. 52D). Only 12 of the 43 genes were up-regulated in *immutans*. Surprisingly, although 28 genes were differentially expressed in *var2*, 13 of them were repressed and thus represent a distinct genetic response than the one observed in *why1why3pollb-1*. Nevertheless, a strong correlation was observed with *msh1* since 31 genes were also up-regulated in this mutant. Our own analysis of the published *msh1* transcriptome revealed a high similitude with the *flu* mutant and the high light treatment. Indeed, 40% and 32% of the *msh1* two fold up-regulated genes are also overexpressed in *flu* and high light treatment, respectively. The fact that the MSH1 protein is involved in the maintenance of plastid and mitochondria genome stability (Xu et al., 2011) further supports the relationship between the observed transcriptome modifications and plastid genome rearrangements.

The *why1why3pollb-1* Nuclear Genetic Reprogramming Is not Observed Under Low-Light Conditions

The important signaling functions carried out by ROS in triggering stress responses have been extensively documented (op den Camp et al., 2003; Maruta et al., 2012; Ramel et al., 2013). To assess if the nuclear genetic reprogramming observed in *why1why3pollb-1* depends on elevated chloroplast ROS production, we compared, in plants grown either under normal or low-light conditions, the expression of six genes whose expression is up-regulated in *why1why3pollb-1*. These genes are the Nudix hydrolase homolog 4 (*NUDT4*) (Ogawa et al., 2005), the oxidative stress marker *UPOX* (AT2G21640) (Sweetlove et al., 2002; Gadjev et al., 2006), the ACC synthase 6 (*ACS6*) (Arteca and Arteca, 1999), the ARM repeat AT3G06530 and the two Early Light-Inducible Proteins (*ELIP1* and *ELIP2*) (Kimura et al., 2001; Hutin et al., 2003; Tzvetkova-Chevolleau et al., 2007). The overexpression of these *why1why3pollb-1* up-regulated genes was first confirmed by quantitative RT-PCR to validate the RNA-

Sequencing results. qRT-PCR confirmed that all these genes are significantly overexpressed in *why1why3pollb-1* (Fig. 53, NL). This overexpression was, however, not observed in *pollb-1* or *why1why3* (Fig. 63). Expression of these markers was then measured in plants grown for 21 days under low light conditions, where no difference in phenotype is observed between *why1why3pollb-1*, WT or *why1why3* (Fig. 49) and where ROS levels are low. Figure 53 (LL) shows that, under these conditions, there is no significant difference in the expression of these genes. These results again suggest that it is the ROS produced under normal light conditions that act as signals to initiate the genetic reprogramming in response to the *why1wh3pollb-1* ptDNA rearrangements.

Genetic Reprogramming in *why1why3pollb-1* Allows Adaptation to High Light Stress

Reprogramming of the *why1why3pollb-1* nuclear transcriptome suggests that this mutant line could adapt to elevated ROS concentrations. This was verified by growing plants for three weeks at normal light, followed by transfer to high light conditions for an additional week. Remarkably, the *why1why3pollb-1* plants were considerably less affected by this treatment than the wild type, *why1why3* or *pollb-1* plants (Fig. 54A). Under these conditions, the *why1why3pollb-1* plants accumulated about 20 times less anthocyanin in their leaves than control plants (Fig. 54C). The greater tolerance to high light in addition to the reduced accumulation of the antioxidant anthocyanin in response to this treatment thus supports an acquired adaptation to elevated ROS levels in *why1why3pollb-1*.

To confirm that this adaptation to high light conditions is acquired rather than innate, plants were germinated and allowed to grow for four weeks under constant high light conditions. The lack of pre-adaptation under normal light conditions abolished the tolerance of *why1why3pollb-1* plants to high light, as these mutants were considerably more affected than control plants (Fig. 54B). Expression pattern of the high light and oxidative stress markers *ELIP1* and *ELIP2* supported that an acquired adaptation is taking place in this mutant (Kimura et al., 2001; Hutin et al., 2003; Tzvetkova-Chevolleau et al., 2007). Indeed, when measured in 5-day-old seedlings grown under normal light, expression levels of *ELIP1* and *ELIP2* were low and did not differ in wild type and *why1why3pollb-1* plants (Fig. 54D). Their expression was however significantly higher than in wild type plants after 14 days in normal light. As these genes are known to be induced by ROS, these results further support the idea that modification

of the nuclear expression and adaptation to subsequent oxidative stresses are processes that first require an early accumulation of ROS.

5.5 DISCUSSION

Plastid Genome Instability Leads to Impairment of Photosynthesis and Increased Chloroplast ROS Production

Until now, few mutations in the plastid genome have been associated to photosynthetic impairment and lethality (Petersen et al., 2011), and the impact of more generalized plastid genome instability is still largely unknown. In this article, we used the *why1why3pollb-1* mutant line and the organelle specific DSB-inducing agent, CIP, to analyse the consequences of generalized plastid genome instability.

First, the *why1why3pollb-1* mutations and CIP treatment were both shown to reduce plant photosynthetic capacity, as indicated by the reduction of the Fv/Fm ratio (Fig. 46D and Fig. 58). The effect of CIP is most likely due to its DSB-inducing activity as NOVO, another gyrase inhibitor that does not induce DSBs, does not influence the Fv/Fm parameter. One interpretation of these results is that, since many ptDNA genes encode components of the PET, ptDNA rearrangements will invariably affect these genes, leading to a modification of the stoichiometry of the components of the PET. The fact that an increased chloroplast ROS production could only be observed in the *why1why3pollb-1* mutant and the CIP-treated plants supports this link between ptDNA rearrangements and severe photosynthetic defects (Figs. 47 and 48). Moreover, as shown by the NBT staining experiment, the abundance of superoxide in *why1why3pollb-1* is strongly and quickly enhanced by light, which further supports an imbalance at the level of the PET (Asada, 1999). Interestingly, in mammals, a similar redox imbalance was also observed in mitochondria, where genome rearrangements are linked to oxidative phosphorylation defects and ROS production (Ishikawa et al., 2008; Fan et al., 2012).

The plastid genome has long been known to be more stable than its nuclear counterpart (Wolfe et al., 1987), but the reason for such a high stability remains unknown. The developmental defects and redox imbalance observed in *why1why3pollb-1* and upon CIP

treatment demonstrate the deleterious consequences associated with plastid genome instability. Indeed, CIP-treated seedlings exhibit bleaching of their new emerging leaves and *why1why3pollb-1* plants present a yellow-variegated dwarf phenotype. Consistent with this finding, the severity of the *why1why3pollb-1* phenotype correlates with the intensity of chloroplast ROS production (Figs. 49, 54B and 59). All these deleterious effects associated with plastid genome instability might explain the importance of maintaining its highly optimized genome structure.

ptDNA Rearrangements in *why1why3pollb-1* Do not Lead to “Permanently-Defective” Chloroplasts

Variation has been studied for half a century, and the mode of inheritance of almost every variegated mutant identified to date has been established. While some variegated mutants show maternal inheritance, others are inherited in a Mendelian fashion (Yu et al., 2007). It has been proposed that in the case of maternally inherited variegation, a perturbation would lead to “permanently-defective” chloroplasts. In *Arabidopsis* and most angiosperms, these “permanently-defective” chloroplasts are then maternally transmitted to the progeny, with the extent of variegation usually correlating with the proportion of transmission. In *why1why3pollb-1*, chloroplasts appear to be photosynthetically active and their ultrastructure is maintained. Accordingly, the yellow variegation observed in this mutant is inherited in the classical Mendelian way, supporting the idea that the chloroplasts are not “permanently-defective”. Conversely, since white variegation appears in *why1why3pollb-1* at the same frequency as in the *why1why3* population, it suggests that both phenotypes are actually the same. Thus, the white variegated sectors of *why1why3pollb-1* most likely possess “permanently-defective” chloroplasts with an altered ultrastructure.

The early studies on ptDNA inheritance were done with lines carrying point mutations at or near homoplasmy in their plastid genome (Avni et al., 1989; Triboush et al., 1999). In such cases, the transmission of the mutation is directly linked to the transmission of the chloroplast. In *why1why3pollb-1*, the progeny from the reciprocal crosses with wild type plants shows wild type levels of DNA rearrangements. This suggests that the rearranged molecules are present in heteroplasmy and relatively low copy numbers, which results mainly in the

transmission of normal copies of the plastid genome. Once again the fact that both the phenotype and the rearrangements are not maternally inherited supports the importance of these DNA rearrangements for the appearance of the yellow-variegated phenotype. In addition, restoration of the alleles in the crossing experiments prevents the formation of *de novo* ptDNA rearrangements and favors the dilution of potentially inherited rearrangements. Furthermore, the *why1why3pollb-1* phenotype remains constant through many generations, which suggests that most DNA rearrangements observed in its plastid genome are generated during plant development rather than inherited from the female gametes.

In *why1why3* plants, rearrangements throughout the plastid genome are also observed, although at a much lower frequency than in *why1why3pollb-1*. In approximately 5% of the *why1why3* plants, the amplification of specific parts of the genome, probably near homoplasmy, leads to the appearance of white variegated sectors (Maréchal et al., 2009). The progeny of these variegated plants is composed of a large fraction of variegated plants (between 20% and 50%) (Maréchal et al., 2009). Interestingly, the white variegated sectors in these plants show abnormal cell morphology and absence of chlorophyll autofluorescence. One could thus assume that the specific amplification causing the white-variegation leads to “permanently-defective” chloroplasts. White variegation is also observed in an equivalent fraction of the *why1why3pollb-1* population, which suggests that these two phenotypes are distinct and independent.

The ROS-Associated *why1why3pollb-1* Nuclear Genetic Reprogramming Promotes Acquired Resistance to High Light

ROS can have various deleterious effects on plant cell development. Therefore, plants have evolved mechanisms to counteract this threat. However, ROS are also important signaling molecules that ensure the response to biotic and abiotic stresses (Apel and Hirt, 2004; Mittler et al., 2004). We propose that in *why1why3pollb-1*, the elevated chloroplast ROS levels are perceived by the cells and lead to genetic reprogramming, as indicated by the transcriptome modifications observed in this mutant. The high similitude of this genetic expression remodelling with that of many ROS-inducing conditions (Fig. 52D) suggests that ROS, and not ptDNA rearrangements, mediate this response. This is further supported by the apparent absence of genetic reprogramming in low light grown *why1why3pollb-1* plants (Fig. 53) where

ptDNA rearrangements are still present (Fig. 60). At this time it is however not possible to completely rule out that the excitation pressure, which is also influenced by light, contributes to the observed genetic reprogramming (Rosso et al., 2009).

Surprisingly, no genes involved in the maintenance of plastid genome integrity were differentially expressed in *why1why3pollb-1*. This is rather unexpected as preventing ptDNA rearrangements could prevent ROS formation and its damages to cellular components. Chloroplast ROS therefore do not appear to act as direct signals of genome perturbation but rather provide a warning that the PET has been affected and that the plant must adapt to stress. This could explain the striking resemblance observed between the genetic reprogramming in *why1why3pollb-1* and the responses to abiotic stresses (Fig. 52D). Examples of this are the early light induced proteins *ELIP1* and *ELIP2*, often considered as markers of high light stress (Kimura et al., 2001; Hutin et al., 2003; Tzvetkova-Chevolleau et al., 2007), which are significantly overexpressed in *why1why3pollb-1*. Overexpression of the abiotic-stress-responsive transcriptional activators *ZAT10*, *ZAT12*, *CBF1* and *CBF2* also supports this hypothesis (Gilmour et al., 1998; Sakamoto et al., 2004; Davletova et al., 2005; Mittler et al., 2006). Finally, a large proportion of the overexpressed *why1why3pollb-1* genes were shown to be involved in calcium and ethylene signaling pathways. The important roles mediated by these pathways in the regulation and signalization of abiotic stresses suggest that these molecules might also be important for proper adaption to plastid genome instability (Rentel and Knight, 2004; Galon et al., 2010; Shi et al., 2012). This adaptation appears to take place between days 5 and 14 (Fig. 54D) and promotes acquired resistance to high light. Although the redox imbalance observed in *why1why3pollb-1* is associated with deleterious phenotypes, the early oxidative stress experienced in this mutant might trigger this acquired resistance. This type of adaptation has also been reported for other chloroplast stresses and is generally referred to as the acclimation response (Prasad et al., 1994; Coll et al., 2009; Ramel et al., 2012; Šimková et al., 2012; Gordon et al., 2013). Indeed, many genes differentially expressed in *why1why3pollb-1* were also shown to be part of the high light systemic acquired acclimation (HL SAA) response (Gordon et al., 2013) and support that the *whywhy3pollb-1* nuclear gene expression reprogramming would be at the origin of the acquired tolerance to high light observed in this mutant. Interestingly, a similar high light tolerance was reported in *msh1*, but the exact

cause of this adaptation remains elusive, as both chloroplast and mitochondria exhibit genomic instability in this mutant (Xu et al., 2011). Similarly to Msh1, PolIB is known to act in both organelles (Parent et al., 2011; Cupp and Nielsen, 2012) and thus it is possible that alterations in DNA metabolism from both chloroplast and mitochondria contribute to the nuclear genetic reprogramming observed in *why1why3pollb-1*. However, the synergistic increase in ptDNA rearrangements observed when plastid-localized Whirly proteins and PolIB are affected suggests that chloroplast defects are the main factor at the origin of changes in the nuclear gene expression.

Taken together, our results indicate that DNA rearrangements in the chloroplast genome lead to a severe PET imbalance and an important increase in photosynthesis-dependent ROS production. This elevated ROS production would then trigger plastid-to-nucleus retrograde signaling pathways that, in turn, reprogram the nuclear transcriptome. Although the *why1why3pollb-1* mutant exhibits a dwarf and yellow-variegated phenotype, this reprogramming would allow the plants to survive to its inner oxidative stress and even acquire a high light tolerance.

5.6 MATERIALS AND METHODS

Plant Material and Growth Conditions

Arabidopsis (*Arabidopsis thaliana*; ecotype Columbia-0) mutant lines *pollb-1* (SALK_134274), *why1why3* and *why1why3pollb-1* were reported previously (Maréchal et al., 2009; Parent et al., 2011). Seeds were sown on soil, vernalized for 3 days at 4°C and grown either under laboratory normal light (100 $\mu\text{mol m}^{-2} \text{s}^{-1}$), very low light (20 $\mu\text{mol m}^{-2} \text{s}^{-1}$) or moderate high light conditions (225 $\mu\text{mol m}^{-2} \text{s}^{-1}$) at 22°C on a 16 h day/8 h dark cycle. When grown at very low light, plants were germinated at 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for one week prior to their transfer at 20 $\mu\text{mol m}^{-2} \text{s}^{-1}$. For paraquat and atrazine experiments, seeds were sterilized and sown on Murashige and Skoog basal media (MS medium, Sigma-Aldrich) supplemented with 1 % sucrose, 0.8 % agar and the indicated amount of drug. For ciprofloxacin and novobiocin treatments, seeds were grown on solid MS medium for eight days and transferred in liquid MS

medium for two days to allow adaptation to the new growth conditions. The indicated amount of drug was then added at day 10.

Detection of ptDNA Rearrangements

Total DNA was isolated for each plant sample using a cetyl trimethylammonium bromide DNA extraction protocol (Weigel and Glazebrook, 2002). For every DNA samples, ptDNA amount was equilibrated with a low-cycle amplification of a *YCF2* DNA fragment. Primers and approach for rearrangements detection were described previously (Cappadocia et al., 2010) and PCR products were visualized after migration on GelRed (Biotium) stained agarose gels. Every reaction shown is representative of the eight primer pairs presented in Table 7.

Measurement of PSII Efficiency

For each line analysed, Fv/Fm measurements were performed on whole, four-week-old plants grown under normal light conditions or whole, 60-day-old plants grown under low light conditions with an Open FluorCam (Photon Systems Instruments) according to the manufacturer's instructions.

Chlorophyll Extraction and Quantification

For each line analyzed, 10 mg of three-week-old plants grown under normal light conditions or 60-day-old plants grown under low light conditions were grinded in 400 μ L dimethylformamide and centrifuged for one minute at 13000 RPM. The supernatant was recovered and absorbance was measured at A_{645} and A_{663} . Chlorophyll content was then assessed as described previously (Porra et al., 1989).

Chloroplast Ultrastructure Imaging

Imaging of *why1why3pollb-1* chloroplast ultrastructure was performed in dissected yellow and green sectors as described previously (Maréchal et al., 2009).

ROS Detection, Quantification and Microscopic Analyses

The NBT superoxide coloration experiment was conducted as described previously (Driever et al., 2009). Briefly, petioles of 28-day-old Arabidopsis leaves were cut twice under water to

avoid air obstruction of the vasculature. For probe feeding of the leaves, petioles were soaked in the dark for 2 hours in 6 mM nitrotetrazolium blue chloride (Sigma-Aldrich) dissolved in 10 mM potassium phosphate pH 7.8. To assess the effect of photosynthesis on superoxide production, fed leaves were transferred at high light for fifteen minutes. To remove chlorophyll, colored leaves were boiled for five minutes in an ethanol, lactic acid and glycerol solution (4:1:1, v/v/v). To quantify formazan, leaves were boiled in dimethyl sulfoxide until leaves were clear and formazan concentration was measured spectrophotometrically at A560. Absorbance was measured in five leaves for every line and conditions tested and then normalized relative to the weight of each leaf. For the H2DCFDA experiment, protocol was adapted from a previous report (Leshem et al., 2006). Briefly, 10 day-old seedlings, acclimated for 48 hours in liquid MS medium, were incubated in 10 μ M H2DCFDA (Sigma-Aldrich) in PBS buffer for 30 minutes under high light exposure and washed twice for 5 minutes in PBS buffer. Chlorophyll autofluorescence and oxidized H2DCFDA were visualized using an Olympus FV300 confocal microscope. Chlorophyll and H2DCFDA were excited with a 488 nm argon laser. H2DCFDA fluorescence was detected at 535 nm and chlorophyll at 650 nm and over.

Anthocyanin Extraction and Quantification

For each line analyzed, three-week-old plants grown under normal light conditions were transferred for one week at high light. Anthocyanin extraction was then conducted as described previously (Cottage et al., 2010) for five 30 mg replicates of each line.

RNA Isolation, RNA-Sequencing and Quantitative RT-PCR Analysis

For each plant sample total RNA was isolated using TRIzol reagent (Invitrogen) and the integrity of every sample assessed using a Bioanalyzer platform (Agilent). Libraries were prepared using the TruSeq RNA Sample Preparation v2 kit (Illumina) according to the manufacturer's instructions. Samples were then pooled together and ran on an Illumina MiSeq to further verify the libraries quality. Sequencing was performed using the Illumina HiSeq2000 machine running TruSeq SBS v3 chemistry at the genomic platform of the Institute for Research in Immunology and Cancer, Université de Montréal. Cluster density was targeted at around 600-800k clusters/mm². Analysis was then carried out using the Galaxy online software

suite (Goecks et al., 2010). Transcripts with FPKM values lower than one were discarded from the analysis to ensure optimal quantification. MapMan was used to retrieve GO annotations associated with the up- and down-regulated genes of the entire *why1why3pollb-1* transcriptome (Thimm et al., 2004). Pageman web browsing tool was used to retrieve the GO annotations associated with the significantly up- and down-regulated *why1why3pollb-1* genes (Usadel et al., 2006). Genevestigator was used to compare the 50 most overexpressed *why1why3pollb-1* genes to that of other conditions (Hruz et al., 2008). Genevestigator transcriptomes for paraquat, hydrogen peroxide and drought treatments were, respectively, AT-120:Kudla/Puchta/Bartels/Harter/Nover, AT-185:Mittler and AT-58:Warren.

For quantitative RT-PCR analysis, cDNA synthesis was carried out using a first-strand cDNA synthesis kit (Fermentas) on RNAs from 14 and 21 day-old seedlings grown under normal and low-light conditions, respectively. cDNAs were diluted 20-fold prior to qRT-PCR experiments. Primers used for qRT-PCR experiments were designed for randomly-selected intron-containing genes and the 6 primer pairs with acceptable efficiency were selected (efficiency between 1.90 and 2.05). These primers are listed in Table 8 Every reaction was carried out on biological and technical triplicates and normalized relative to β -Tubulin. Real-time qPCR reactions were done using the SYBR green master mix (SABiosciences, www.sabiosciences.com) according to the manufacturer's instructions. Melting curves confirmed the amplification of a unique PCR product for each of the qRT-PCR primer pairs used. LightCycler480 (Roche) was used for qRT-PCR experiments and data analyzed using the Light-Cycler480 software version 1.5.

Accession Numbers

Sequence data from this article can be found under the following accession numbers: WHY1 (AT1G14410), WHY3 (AT2G02740), POLIB (AT3G20540).

5.7 SUPPLEMENTARY DATA

The following data are available in the online version of this article.

Figure 55. Microhomology-Mediated Break-Induced Replication (MMBIR) Leads to DNA Rearrangements.

Figure 56. The *why1why3pollb-1* Mutant Develops Distincts Yellow and Green Sectors on Its Leaves.

Figure 57. Some *why1why3pollb-1* Plants Exhibit Yellow and White Variegation Simultaneously.

Figure 58. Ciprofloxacin Negatively Impacts Photosynthetic Efficiency in Arabidopsis.

Figure 59. *why1why3* and *why1why3pollb-1* Are Hypersensitive to Paraquat.

Figure 60. The Abundance of MMBIR-Associated ptDNA Rearrangements Is Little Affected by the Light-Regimen in *why1why3pollb-1*.

Figure 61. Paraquat and Atrazine do not induce MMBIR-Associated ptDNA Rearrangements.

Figure 62. Elevated ROS Levels Do not Lead to Cell Death in *why1why3pollb-1*.

Figure 63. The *why1why3pollb-1* Nuclear Genetic Reprogramming Is Not Due to the *pollb-1* and *why1why3* Genetic Backgrounds.

Tableau VIII. Primer Pairs Used for the Specific Amplification of MMBIR-Associated ptDNA Rearrangements.

Tableau IX. Primer Pairs Used for Quantitative RT-PCR experiments.

5.8 ACKNOWLEDGEMENTS

The assistance of Jean-Benoit Charron, Marianne Arteau and Sébastien Truche with the Fv/Fm, RNA-Sequencing and microscopy experiments, respectively, is gratefully acknowledged. We gratefully acknowledge Samuel Tremblay-Belzile for reading the manuscript.

AUTHOR CONTRIBUTIONS

E.L., E.Z. and N.B. designed research. E.L. and E.Z. performed research. E.L., E.Z. and N.B. analyzed the data. E.L., E.Z. and N.B. wrote the article.

5.9 FIGURE TITLES

Figure 46. Plastid DNA Rearrangements in *why1why3pollb-1* Lead to Photosynthetic Impairment and Yellow Variegation.

(A) Representative photographs of 21-day-old wild type and *pollb-1*, *why1why3* and *why1why3pollb-1* Arabidopsis mutants grown under normal light conditions. White bars represent 1 cm.

(B) Enlargements showing the variegated sectors of *why1why3* and *why1why3pollb-1*. White bars represent 1 cm.

(C) Representative outward-facing PCR reactions showing the abundance of MMBIR-associated plastid DNA rearrangements in 14-day-old seedlings of the mutant lines presented in (A). PCR reactions were carried out using the indicated primer pairs. The amount of DNA loaded in each reaction was equilibrated relative to the amplification of a *YCF2* fragment. Every reaction was performed on three different biological samples and a representative biological duplicate is presented.

(D) Fv/Fm ratios for 21-day-old plants of the indicated genotype grown under normal light conditions. The error bars represent the standard error of the mean (SEM). Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type. Every measurement was performed independently on at least five biological samples

Figure 47. *why1why3pollb-1* Produces Elevated Levels of Chloroplast ROS in a Light-Dependent Manner.

(A) Microscopic analysis of 28-day-old Arabidopsis leaves incubated with H₂DCFDA. Oxidized H₂DCFDA is represented in green, chlorophyll in magenta and the merge yields a white color.

(B) Inset on H₂DCFDA foci, which colocalize with chloroplasts.

(C) Representative photographs of NBT stained leaves from 28-day-old Arabidopsis plants. Blue color on the leaves is caused by formazan, the reduced form of NBT.

(D) Histogram presenting the relative formazan concentration of the dark-incubated leaves. Five different leaves were tested for every condition. The error bars represent the standard deviation of the five extractions. One asterisk indicates a significant difference of a Student's t-test P-value ≤ 0.05 with the wild type.

(E) Histogram presenting the relative formazan concentration of dark-incubated leaves switched to high light for 15 minutes. Five different leaves were tested for every condition. The error bars represent the standard deviation of the five extractions. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type.

Figure 48. Ciprofloxacin-Treated Seedlings Show Signs of Bleaching Associated With Increased Chloroplast ROS Production.

(A) Representative photographs of 15-day-old plants exposed from day 10 to 15 with the indicated concentration of drug.

(B) Microscopic analysis of plants exposed to the prior treatment with H₂DCFDA. Photographs were taken on day 11, 24 hours after the beginning of the treatment. Oxidized H₂DCFDA is represented in green, chlorophyll in magenta and the merge yields a white color.

(C) Inset on H₂DCFDA foci, which colocalize with chloroplasts.

Figure 49. Growth under Low Light Conditions Rescues the *why1why3pollb-1* Phenotype.

(A) Representative photographs of wild type, *why1why3* and *why1why3pollb-1* Arabidopsis plants grown under normal (NL) or low light (LL) regimens.

(B) Chlorophyll content of each line grown under normal or low-light conditions for 21 and 60 days, respectively. White and Gray bars represent respectively NL and LL experiments. The error bars represent the standard deviation for five measurements on independent biological samples. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type.

(C) Fv/Fm ratios for 60-day-old plants of the indicated genotype grown under low-light conditions. The error bars represent the standard error of the mean (SEM). Two asterisks

indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type. Every measurement was performed independently on at least five biological samples.

Figure 50. The *why1why3pollb-1* Yellow-Variigated Sectors Are not Devoid of Photosynthetically Active Chloroplasts.

(A) Chloroplast ultrastructure observed by transmission electronic microscopy of green, white and yellow sectors of *why1why3* and *why1why3pollb-1* leaves.

(B) Microscopic analysis of chlorophyll autofluorescence in green, white and yellow sectors of 28-day-old *why1why3* and *why1why3pollb-1* leaves. Chlorophyll fluorescence is represented in magenta.

Figure 51. The *why1why3pollb-1* ptDNA Rearrangements and Phenotype Are not Maternally Inherited.

(A) Representative photographs of 21-day-old backcrossed *why1why3pollb-1* plants.

(B) Representative outward-facing PCR reactions showing the abundance of MMBIR-associated plastid DNA rearrangements of 14-day-old backcrossed *why1why3pollb-1* plants. PCR reactions were carried out using the indicated primer pairs. The amount of DNA loaded in each reaction was equilibrated relative to the amplification of a *YCF2* fragment. Every reaction was performed on three different biological samples and a representative biological duplicate is presented.

Figure 52. The *why1why3pollb-1* Mutant Exhibits a Complex Nuclear Genetic Reprogramming.

(A) MapMan analysis for GO terms enrichment of the entire *why1why3pollb-1* nuclear transcriptome.

(B) Pie charts presenting the biological processes associated to the significantly *why1why3pollb-1* up- and down-regulated genes (2-fold). To lighten the figure, the GO terms “unknown” and “not assigned” were removed from the analysis.

(C) Histogram presenting the enrichment of each biological process of the *why1why3pollb-1* transcriptome.

(D) Heatmap of the 50 most up-regulated *why1why3pollb-1* genes compared to other ROS-inducing conditions, abiotic stress responses and variegated mutants transcriptomes.

Figure 53. The *why1why3pollb-1* Nuclear Genetic Reprogramming Is Not Observed under Low-Light Conditions.

Histograms presenting the real-time PCR measurements of *NUDT4*, *UPOX* (AT2G21640), *ACS6*, AT3G06530, *ELIP1* and *ELIP2* expression in *why1why3pollb-1* relative to wild type plants for normal light (NL) and low light (LL) conditions. The error bars represent the standard error of the mean (SEM) of the biological triplicates. One and two asterisks indicate, respectively, a significant difference of a Student’s t-test P-value ≤ 0.05 and 0.01 with the wild type.

Figure 54. The Early Oxidative Stress in *why1why3pollb-1* Leads to Acquired Resistance to High Light.

(A) Photographs of representative Arabidopsis plants grown for 21 days at normal light (NL) and switched to high light (HL) for an additional seven days.

(B) Photographs of representative Arabidopsis plants grown for 28 days at high light (HL).

(C) Histogram presenting the relative anthocyanin content for the mutant lines and wild type plants switched for seven days at high light. The error bars represent the standard deviation of the biological triplicates. Two asterisks indicate a significant difference of a Student’s t-test P-value ≤ 0.01 with the wild type.

(D) Histogram presenting the real-time PCR measurements of the *ELIP1* and *ELIP2* expression in *why1why3pollb-1* relative to wild type plants grown under normal light conditions for 5 and

14 days. The error bars represent the standard error of the mean (SEM) of the biological triplicates. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type.

5.10 REFERENCES

- Aluru, MR, Zola, J, Foudree, A, and Rodermel, SR** (2009) Chloroplast photooxidation-induced transcriptome reprogramming in *Arabidopsis immutans* white leaf sectors. *Plant Physiol* **150**: 904-923
- Apel, K, and Hirt, H** (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu Rev Plant Biol* **55**: 373-399
- Arsova, B, Hoja, U, Wimmelbacher, M, Greiner, E, Ustun, S, Melzer, M, Petersen, K, Lein, W, and Bornke, F** (2010) Plastidial thioredoxin z interacts with two fructokinase-like proteins in a thiol-dependent manner: evidence for an essential role in chloroplast development in *Arabidopsis* and *Nicotiana benthamiana*. *Plant Cell* **22**: 1498-1515
- Arteca, JM, and Arteca, RN** (1999) A multi-responsive gene encoding 1-aminocyclopropane-1-carboxylate synthase (ACS6) in mature *Arabidopsis* leaves. *Plant Mol Biol* **39**: 209-219
- Asada, K** (1999) THE WATER-WATER CYCLE IN CHLOROPLASTS: Scavenging of Active Oxygens and Dissipation of Excess Photons. *Annu Rev Plant Physiol Plant Mol Biol* **50**: 601-639
- Asada, K** (2006) Production and scavenging of reactive oxygen species in chloroplasts and their functions. *Plant Physiol* **141**: 391-396
- Avni, A, Edelman, M, Rachailovich, I, Aviv, D, and Fluhr, R** (1989) A Point Mutation in the Gene for the Large Subunit of Ribulose 1,5-Bisphosphate Carboxylase Oxygenase Affects Holoenzyme Assembly in *Nicotiana-Tabacum*. *EMBO J* **8**: 1915-1918
- Cappadocia, L, Maréchal, A, Parent, JS, Lepage, E, Sygusch, J, and Brisson, N** (2010) Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair. *Plant Cell* **22**: 1849-1867
- Cheng, MC, Liao, PM, Kuo, WW, and Lin, TP** (2013) The *Arabidopsis* ETHYLENE-RESPONSE-FACTOR1 Regulates Abiotic-Stress-Responsive Gene Expression by Binding to Different cis-Acting Elements in Response to Different Stress Signals. *Plant Physiol*

- Chia, LS, McRae, DG, and Thompson, JE** (1982) Light-Dependence of Paraquat-Initiated Membrane Deterioration in Bean-Plants - Evidence for the Involvement of Superoxide. *Physiol Plant* **56**: 492-499
- Coll, NS, Danon, A, Meurer, J, Cho, WK, and Apel, K** (2009) Characterization of *soldat8*, a suppressor of singlet oxygen-induced cell death in *Arabidopsis* seedlings. *Plant Cell Physiol* **50**: 707-718
- Cottage, A, Mott, EK, Kempster, JA, and Gray, JC** (2010) The *Arabidopsis* plastid-signalling mutant *gun1* (*genomes uncoupled1*) shows altered sensitivity to sucrose and abscisic acid and alterations in early seedling development. *J Exp Bot* **61**: 3773-3786
- Cupp, JD, and Nielsen, BL** (2012) *Arabidopsis thaliana* organellar DNA polymerase IB mutants exhibit reduced mtDNA levels with a decrease in mitochondrial area density. *Physiol Plant*
- Davletova, S, Schlauch, K, Coutu, J, and Mittler, R** (2005) The zinc-finger protein *Zat12* plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. *Plant Physiol* **139**: 847-856
- Driever, SM, Fryer, MJ, Mullineaux, PM, and Baker, NR** (2009) Imaging of reactive oxygen species in vivo. *Methods Mol Biol* **479**: 109-116
- Drlica, K, and Zhao, X** (1997) DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol Mol Biol Rev* **61**: 377-392
- Ecker, JR** (1995) The ethylene signal transduction pathway in plants. *Science* **268**: 667-675
- Escoubas, JM, Lomas, M, LaRoche, J, and Falkowski, PG** (1995) Light intensity regulation of *cab* gene transcription is signaled by the redox state of the plastoquinone pool. *Proc Natl Acad Sci U S A* **92**: 10237-10241
- Fan, W, Lin, CS, Potluri, P, Procaccio, V, and Wallace, DC** (2012) mtDNA lineage analysis of mouse L-cell lines reveals the accumulation of multiple mtDNA mutants and intermolecular recombination. *Genes Dev* **26**: 384-394
- Gadjev, I, Vanderauwera, S, Gechev, TS, Laloi, C, Minkov, IN, Shulaev, V, Apel, K, Inze, D, Mittler, R, and Van Breusegem, F** (2006) Transcriptomic footprints disclose specificity of reactive oxygen species signaling in *Arabidopsis*. *Plant Physiol* **141**: 436-445
- Galon, Y, Finkler, A, and Fromm, H** (2010) Calcium-regulated transcription in plants. *Molecular plant* **3**: 653-669
- Gilmour, SJ, Zarka, DG, Stockinger, EJ, Salazar, MP, Houghton, JM, and Thomashow, MF** (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. *Plant J* **16**: 433-442

- Gonzalez-Perez, S, Gutierrez, J, Garcia-Garcia, F, Osuna, D, Dopazo, J, Lorenzo, O, Revuelta, JL, and Arellano, JB** (2011) Early transcriptional defense responses in Arabidopsis cell suspension culture under high-light conditions. *Plant Physiol* **156**: 1439-1456
- Gordon, MJ, Carmody, M, Albrecht, V, and Pogson, B** (2013) Systemic and Local Responses to Repeated HL Stress-Induced Retrograde Signaling in Arabidopsis. *Frontiers in plant science* **3**: 303
- Grace, SC, and Logan, BA** (1996) Acclimation of Foliar Antioxidant Systems to Growth Irradiance in Three Broad-Leaved Evergreen Species. *Plant Physiol* **112**: 1631-1640
- Green, BR** (2011) Chloroplast genomes of photosynthetic eukaryotes. *Plant J* **66**: 34-44
- Hardy, CD, and Cozzarelli, NR** (2003) Alteration of Escherichia coli topoisomerase IV to novobiocin resistance. *Antimicrob Agents Chemother* **47**: 941-947
- Hastings, PJ, Ira, G, and Lupski, JR** (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* **5**: e1000327
- Hruz, T, Laule, O, Szabo, G, Wessendorp, F, Bleuler, S, Oertle, L, Widmayer, P, Gruissem, W, and Zimmermann, P** (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advances in bioinformatics* **2008**: 420747
- Huner, NP, Bode, R, Dahal, K, Hollis, L, Rosso, D, Krol, M, and Ivanov, AG** (2012) Chloroplast redox imbalance governs phenotypic plasticity: the "grand design of photosynthesis" revisited. *Frontiers in plant science* **3**: 255
- Hutin, C, Nussaume, L, Moise, N, Moya, I, Kloppstech, K, and Havaux, M** (2003) Early light-induced proteins protect Arabidopsis from photooxidative stress. *Proc Natl Acad Sci U S A* **100**: 4921-4926
- Ishikawa, K, Takenaga, K, Akimoto, M, Koshikawa, N, Yamaguchi, A, Imanishi, H, Nakada, K, Honma, Y, and Hayashi, J** (2008) ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis. *Science* **320**: 661-664
- Kato, Y, Miura, E, Ido, K, Ifuku, K, and Sakamoto, W** (2009) The variegated mutants lacking chloroplastic FtsHs are defective in D1 degradation and accumulate reactive oxygen species. *Plant Physiol* **151**: 1790-1801
- Kimura, M, Yoshizumi, T, Manabe, K, Yamamoto, YY, and Matsui, M** (2001) Arabidopsis transcriptional regulation by light stress via hydrogen peroxide-dependent and -independent pathways. *Genes Cells* **6**: 607-617

- Kindgren, P, Kremnev, D, Blanco, NE, de Dios Barajas Lopez, J, Fernandez, AP, Tellgren-Roth, C, Kleine, T, Small, I, and Strand, A** (2012) The plastid redox insensitive 2 mutant of Arabidopsis is impaired in PEP activity and high light-dependent plastid redox signalling to the nucleus. *Plant J* **70**: 279-291
- Laloi, C, Stachowiak, M, Pers-Kamczyc, E, Warzych, E, Murgia, I, and Apel, K** (2007) Cross-talk between singlet oxygen- and hydrogen peroxide-dependent signaling of stress responses in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **104**: 672-677
- Leshem, Y, Melamed-Book, N, Cagnac, O, Ronen, G, Nishri, Y, Solomon, M, Cohen, G, and Levine, A** (2006) Suppression of Arabidopsis vesicle-SNARE expression inhibited fusion of H₂O₂-containing vesicles with tonoplast and increased salt tolerance. *Proc Natl Acad Sci U S A* **103**: 18008-18013
- Maréchal, A, Parent, JS, Veronneau-Lafortune, F, Joyeux, A, Lang, BF, and Brisson, N** (2009) Whirly proteins maintain plastid genome stability in Arabidopsis. *Proc Natl Acad Sci U S A* **106**: 14693-14698
- Martínez-Zapater, JM** (1993) Genetic-Analysis of Variegated Mutants in Arabidopsis. *J Hered* **84**: 138-140
- Maruta, T, Noshi, M, Tanouchi, A, Tamoi, M, Yabuta, Y, Yoshimura, K, Ishikawa, T, and Shigeoka, S** (2012) H₂O₂-triggered retrograde signaling from chloroplasts to nucleus plays specific role in response to stress. *J Biol Chem* **287**: 11717-11729
- Miller, G, Shulaev, V, and Mittler, R** (2008) Reactive oxygen signaling and abiotic stress. *Physiol Plant* **133**: 481-489
- Mittler, R** (2002) Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci* **7**: 405-410
- Mittler, R, Vanderauwera, S, Gollery, M, and Van Breusegem, F** (2004) Reactive oxygen gene network of plants. *Trends Plant Sci* **9**: 490-498
- Mittler, R, Kim, Y, Song, L, Coutu, J, Coutu, A, Ciftci-Yilmaz, S, Lee, H, Stevenson, B, and Zhu, JK** (2006) Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. *FEBS Lett* **580**: 6537-6542
- Miura, E, Kato, Y, and Sakamoto, W** (2010) Comparative transcriptome analysis of green/white variegated sectors in Arabidopsis yellow variegated2: responses to oxidative and other stresses in white sectors. *J Exp Bot* **61**: 2433-2445

- Mori, Y, Kimura, S, Saotome, A, Kasai, N, Sakaguchi, N, Uchiyama, Y, Ishibashi, T, Yamamoto, T, Chiku, H, and Sakaguchi, K** (2005) Plastid DNA polymerases from higher plants, *Arabidopsis thaliana*. *Biochem Biophys Res Commun* **334**: 43-50
- Myouga, F, Hosoda, C, Umezawa, T, Iizumi, H, Kuromori, T, Motohashi, R, Shono, Y, Nagata, N, Ikeuchi, M, and Shinozaki, K** (2008) A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell* **20**: 3148-3162
- Odom, OW, Baek, KH, Dani, RN, and 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**: 842-853
- Ogawa, T, Ueda, Y, Yoshimura, K, and Shigeoka, S** (2005) Comprehensive analysis of cytosolic Nudix hydrolases in *Arabidopsis thaliana*. *J Biol Chem* **280**: 25277-25283
- op den Camp, RG, Przybyla, D, Ochsenein, C, Laloi, C, Kim, C, Danon, A, Wagner, D, Hideg, E, Gobel, C, Feussner, I, Nater, M, and Apel, K** (2003) Rapid induction of distinct stress responses after the release of singlet oxygen in *Arabidopsis*. *Plant Cell* **15**: 2320-2332
- Parent, JS, Lepage, E, and Brisson, N** (2011) Divergent roles for the two PolII-like organelle DNA polymerases of *Arabidopsis*. *Plant Physiol* **156**: 254-262
- Petersen, K, Schöttler, MA, Karcher, D, Thiele, W, and Bock, R** (2011) Elimination of a group II intron from a plastid gene causes a mutant phenotype. *Nucleic Acids Res* **39**: 5181-5192
- Pfannschmidt, T, Nilsson, A, and Allen, JF** (1999) Photosynthetic control of chloroplast gene expression. *Nature* **397**: 625-628
- Porra, RJ, Thompson, WA, and Kriedemann, PE** (1989) Determination of Accurate Extinction Coefficients and Simultaneous-Equations for Assaying Chlorophyll-a and Chlorophyll-B Extracted with 4 Different Solvents - Verification of the Concentration of Chlorophyll Standards by Atomic-Absorption Spectroscopy. *Biochim Biophys Acta* **975**: 384-394
- Prasad, TK, Anderson, MD, Martin, BA, and Stewart, CR** (1994) Evidence for Chilling-Induced Oxidative Stress in Maize Seedlings and a Regulatory Role for Hydrogen Peroxide. *Plant Cell* **6**: 65-74
- Przybyla, D, Gobel, C, Imboden, A, Hamberg, M, Feussner, I, and Apel, K** (2008) Enzymatic, but not non-enzymatic, $1O_2$ -mediated peroxidation of polyunsaturated fatty acids forms part of the EXECUTER1-dependent stress response program in the flu mutant of *Arabidopsis thaliana*. *Plant J* **54**: 236-248

- Ramel, F, Sulmon, C, Bogard, M, Couée, I, and Gouesbet, G** (2009) Differential patterns of reactive oxygen species and antioxidative mechanisms during atrazine injury and sucrose-induced tolerance in *Arabidopsis thaliana* plantlets. *BMC Plant Biol* **9**: 28
- Ramel, F, Birtic, S, Ginies, C, Soubigou-Taconnat, L, Triantaphylides, C, and Havaux, M** (2012) Carotenoid oxidation products are stress signals that mediate gene responses to singlet oxygen in plants. *Proc Natl Acad Sci U S A* **109**: 5535-5540
- Ramel, F, Ksas, B, Akkari, E, Mialoundama, AS, Monnet, F, Krieger-Liszkay, A, Ravanat, JL, Mueller, MJ, Bouvier, F, and Havaux, M** (2013) Light-Induced Acclimation of the *Arabidopsis chlorina1* Mutant to Singlet Oxygen. *Plant Cell* **25**: 1445-1462
- Rédei, GP** (1963) Somatic Instability Caused by a Cysteine-Sensitive Gene in *Arabidopsis*. *Science* **139**: 767-769
- Rentel, MC, and Knight, MR** (2004) Oxidative stress-induced calcium signaling in *Arabidopsis*. *Plant Physiol* **135**: 1471-1479
- Rosso, D, Bode, R, Li, W, Krol, M, Saccon, D, Wang, S, Schillaci, LA, Rodermel, SR, Maxwell, DP, and Hüner, NP** (2009) Photosynthetic redox imbalance governs leaf sectoring in the *Arabidopsis thaliana* variegation mutants *immutans*, *spotty*, *var1*, and *var2*. *Plant Cell* **21**: 3473-3492
- Ruckle, ME, Burgoon, LD, Lawrence, LA, Sinkler, CA, and Larkin, RM** (2012) Plastids are major regulators of light signaling in *Arabidopsis*. *Plant Physiol* **159**: 366-390
- Sakamoto, H, Maruyama, K, Sakuma, Y, Meshi, T, Iwabuchi, M, Shinozaki, K, and Yamaguchi-Shinozaki, K** (2004) *Arabidopsis* Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. *Plant Physiol* **136**: 2734-2746
- Shi, Y, Tian, S, Hou, L, Huang, X, Zhang, X, Guo, H, and Yang, S** (2012) Ethylene signaling negatively regulates freezing tolerance by repressing expression of CBF and type-A ARR genes in *Arabidopsis*. *Plant Cell* **24**: 2578-2595
- Šimková, K, Kim, C, Gacek, K, Baruah, A, Laloi, C, and Apel, K** (2012) The chloroplast division mutant *caa33* of *Arabidopsis thaliana* reveals the crucial impact of chloroplast homeostasis on stress acclimation and retrograde plastid-to-nucleus signaling. *Plant J* **69**: 701-712
- Stroher, E, and Dietz, KJ** (2008) The dynamic thiol-disulphide redox proteome of the *Arabidopsis thaliana* chloroplast as revealed by differential electrophoretic mobility. *Physiol Plant* **133**: 566-583

- Sun, X, Fu, T, Chen, N, Guo, J, Ma, J, Zou, M, Lu, C, and Zhang, L** (2010) The stromal chloroplast Deg7 protease participates in the repair of photosystem II after photoinhibition in Arabidopsis. *Plant Physiol* **152**: 1263-1273
- Sweetlove, LJ, Heazlewood, JL, Herald, V, Holtzapffel, R, Day, DA, Leaver, CJ, and Millar, AH** (2002) The impact of oxidative stress on Arabidopsis mitochondria. *Plant J* **32**: 891-904
- Thimm, O, Blasing, O, Gibon, Y, Nagel, A, Meyer, S, Kruger, P, Selbig, J, Muller, LA, Rhee, SY, and Stitt, M** (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* **37**: 914-939
- Torres, MA** (2010) ROS in biotic interactions. *Physiol Plant* **138**: 414-429
- Triantaphylides, C, Krischke, M, Hoerberichts, FA, Ksas, B, Gresser, G, Havaux, M, Van Breusegem, F, and Mueller, MJ** (2008) Singlet oxygen is the major reactive oxygen species involved in photooxidative damage to plants. *Plant Physiol* **148**: 960-968
- Triboush, SO, Danilenko, NG, Ulitcheva, II, and Davydenko, OG** (1999) Location of induced mutations and reversions in the chloroplast genome of *Helianthus annuus*. *Plant Growth Regul* **27**: 75-81
- Tzvetkova-Chevolleau, T, Franck, F, Alawady, AE, Dall'Osto, L, Carriere, F, Bassi, R, Grimm, B, Nussaume, L, and Havaux, M** (2007) The light stress-induced protein ELIP2 is a regulator of chlorophyll synthesis in *Arabidopsis thaliana*. *Plant J* **50**: 795-809
- Usadel, B, Nagel, A, Steinhauser, D, Gibon, Y, Blasing, OE, Redestig, H, Sreenivasulu, N, Krall, L, Hannah, MA, Poree, F, Fernie, AR, and Stitt, M** (2006) PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC Bioinformatics* **7**: 535
- Vogel, JT, Zarka, DG, Van Buskirk, HA, Fowler, SG, and Thomashow, MF** (2005) Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. *Plant J* **41**: 195-211
- Walley, JW, Kelley, DR, Nestorova, G, Hirschberg, DL, and Dehesh, K** (2010) *Arabidopsis* deadenylases AtCAF1a and AtCAF1b play overlapping and distinct roles in mediating environmental stress responses. *Plant Physiol* **152**: 866-875
- Weigel, D, and Glazebrook, J** (2002) *Arabidopsis: A Laboratory Manual* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY)
- Wolfe, KH, Li, WH, and Sharp, PM** (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A* **84**: 9054-9058

- Xu, YZ, Arrieta-Montiel, MP, Viridi, KS, de Paula, WB, Widhalm, JR, Basset, GJ, Davila, JI, Elthon, TE, Elowsky, CG, Sato, SJ, Clemente, TE, and Mackenzie, SA** (2011) MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. *Plant Cell* **23**: 3428-3441
- Yu, F, Fu, A, Aluru, M, Park, S, Xu, Y, Liu, H, Liu, X, Foudree, A, Nambogga, M, and Rodermel, S** (2007) Variegation mutants and mechanisms of chloroplast biogenesis. *Plant Cell Environ* **30**: 350-365
- Zaltsman, A, Ori, N, and Adam, Z** (2005) Two types of FtsH protease subunits are required for chloroplast biogenesis and Photosystem II repair in Arabidopsis. *Plant Cell* **17**: 2782-2790

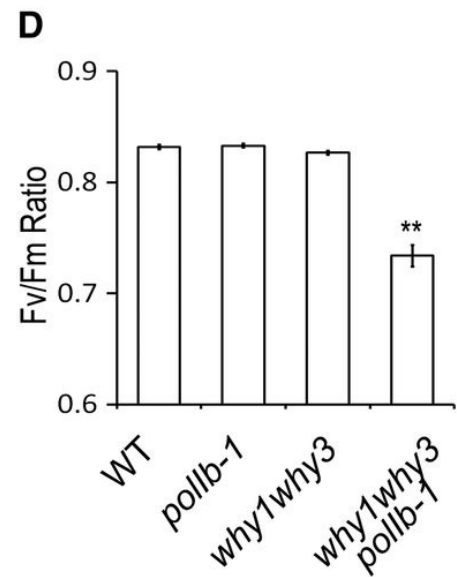
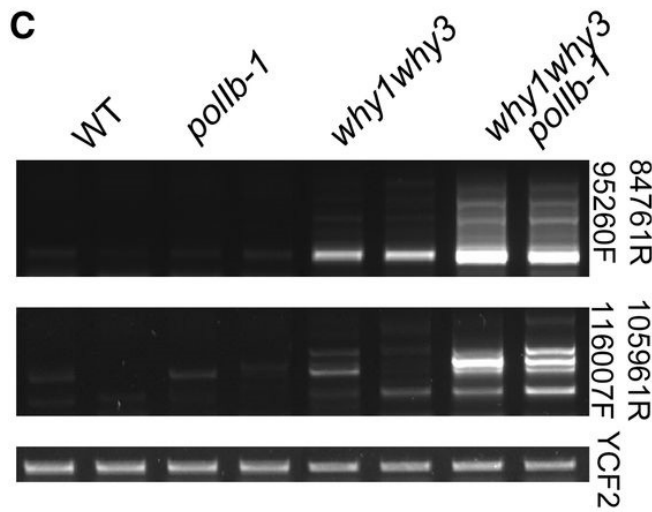
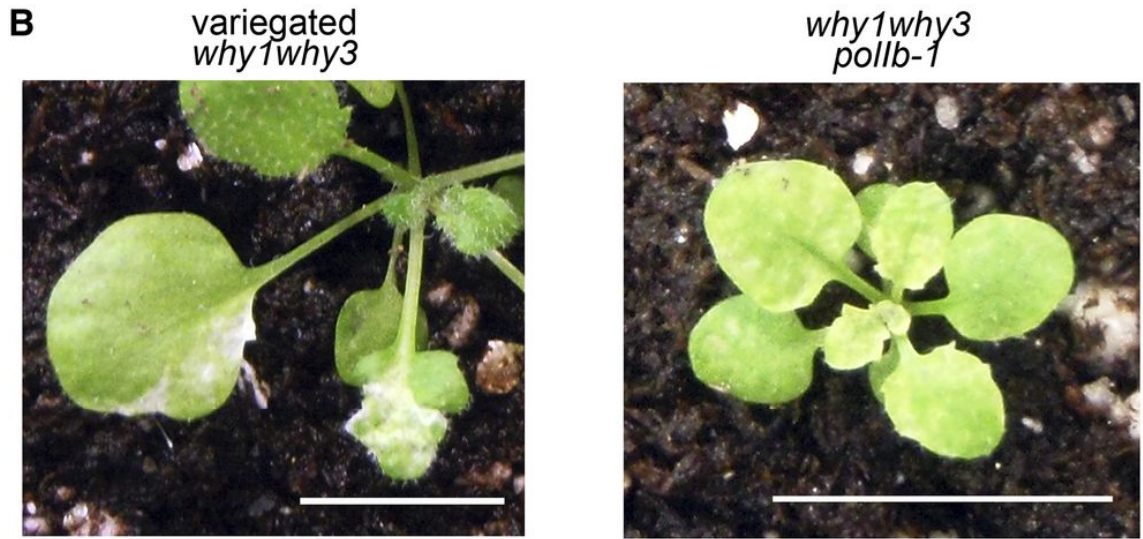


Figure 46. Plastid DNA Rearrangements in *why1why3pollb-1* Lead to Photosynthetic Impairment and Yellow Variegation.

(A) Representative photographs of 21-day-old wild type and *pollb-1*, *why1why3* and *why1why3pollb-1* Arabidopsis mutants grown under normal light conditions. White bars represent 1 cm.

(B) Enlargements showing the variegated sectors of *why1why3* and *why1why3pollb-1*. White bars represent 1 cm.

(C) Representative outward-facing PCR reactions showing the abundance of MMBIR-associated plastid DNA rearrangements in 14-day-old seedlings of the mutant lines presented in (A). PCR reactions were carried out using the indicated primer pairs. The amount of DNA loaded in each reaction was equilibrated relative to the amplification of a *YCF2* fragment. Every reaction was performed on three different biological samples and a representative biological duplicate is presented.

(D) Fv/Fm ratios for 21-day-old plants of the indicated genotype grown under normal light conditions. The error bars represent the standard error of the mean (SEM). Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type. Every measurement was performed independently on at least five biological samples

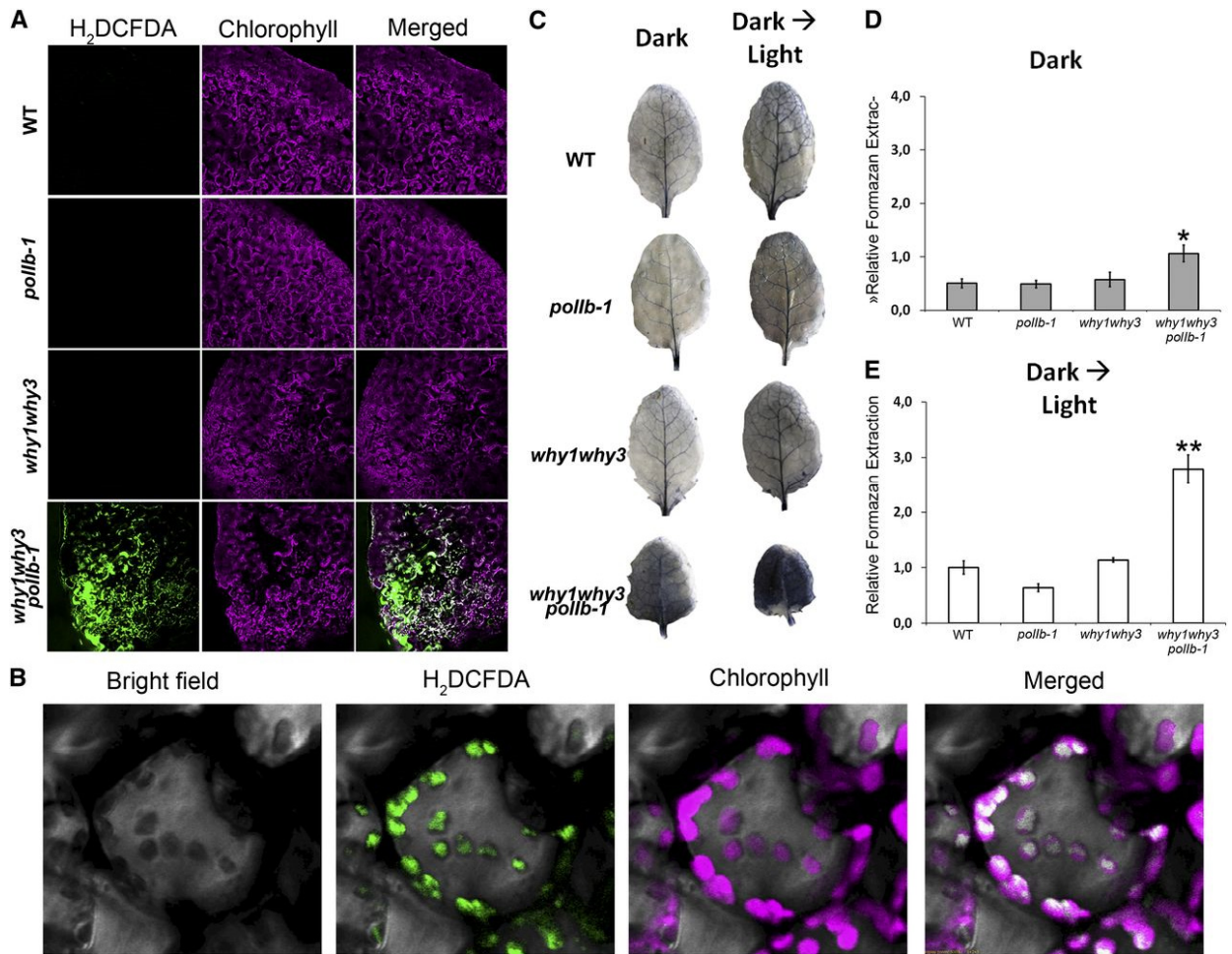


Figure 47. *why1why3pollb-1* Produces Elevated Levels of Chloroplast ROS in a Light-Dependent Manner.

(A) Microscopic analysis of 28-day-old Arabidopsis leaves incubated with H₂DCFDA. Oxidized H₂DCFDA is represented in green, chlorophyll in magenta and the merge yields a white color.

(B) Inset on H₂DCFDA foci, which colocalize with chloroplasts.

(C) Representative photographs of NBT stained leaves from 28-day-old Arabidopsis plants. Blue color on the leaves is caused by formazan, the reduced form of NBT.

(D) Histogram presenting the relative formazan concentration of the dark-incubated leaves. Five different leaves were tested for every condition. The error bars represent the standard deviation of the five extractions. One asterisk indicates a significant difference of a Student's t-test P-value ≤ 0.05 with the wild type.

(E) Histogram presenting the relative formazan concentration of dark-incubated leaves switched to high light for 15 minutes. Five different leaves were tested for every condition. The error bars represent the standard deviation of the five extractions. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type.

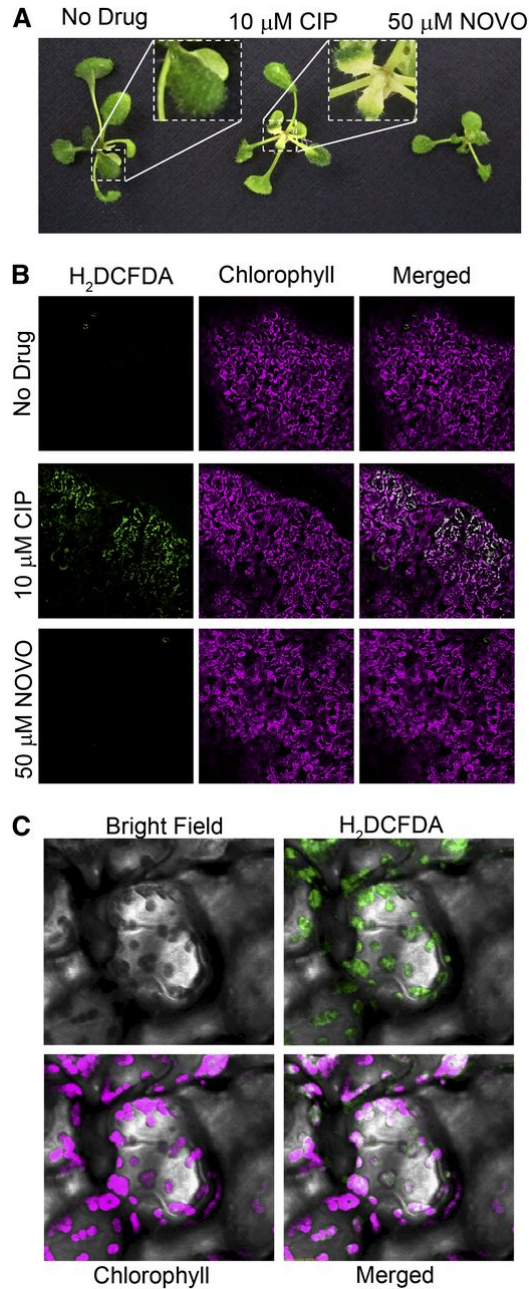


Figure 48. Ciprofloxacin-Treated Seedlings Show Signs of Bleaching Associated With Increased Chloroplast ROS Production.

(A) Representative photographs of 15-day-old plants exposed from day 10 to 15 with the indicated concentration of drug.

(B) Microscopic analysis of plants exposed to the prior treatment with H₂DCFDA. Photographs were taken on day 11, 24 hours after the beginning of the treatment. Oxidized H₂DCFDA is represented in green, chlorophyll in magenta and the merge yields a white color.

(C) Inset on H₂DCFDA foci, which colocalize with chloroplasts.

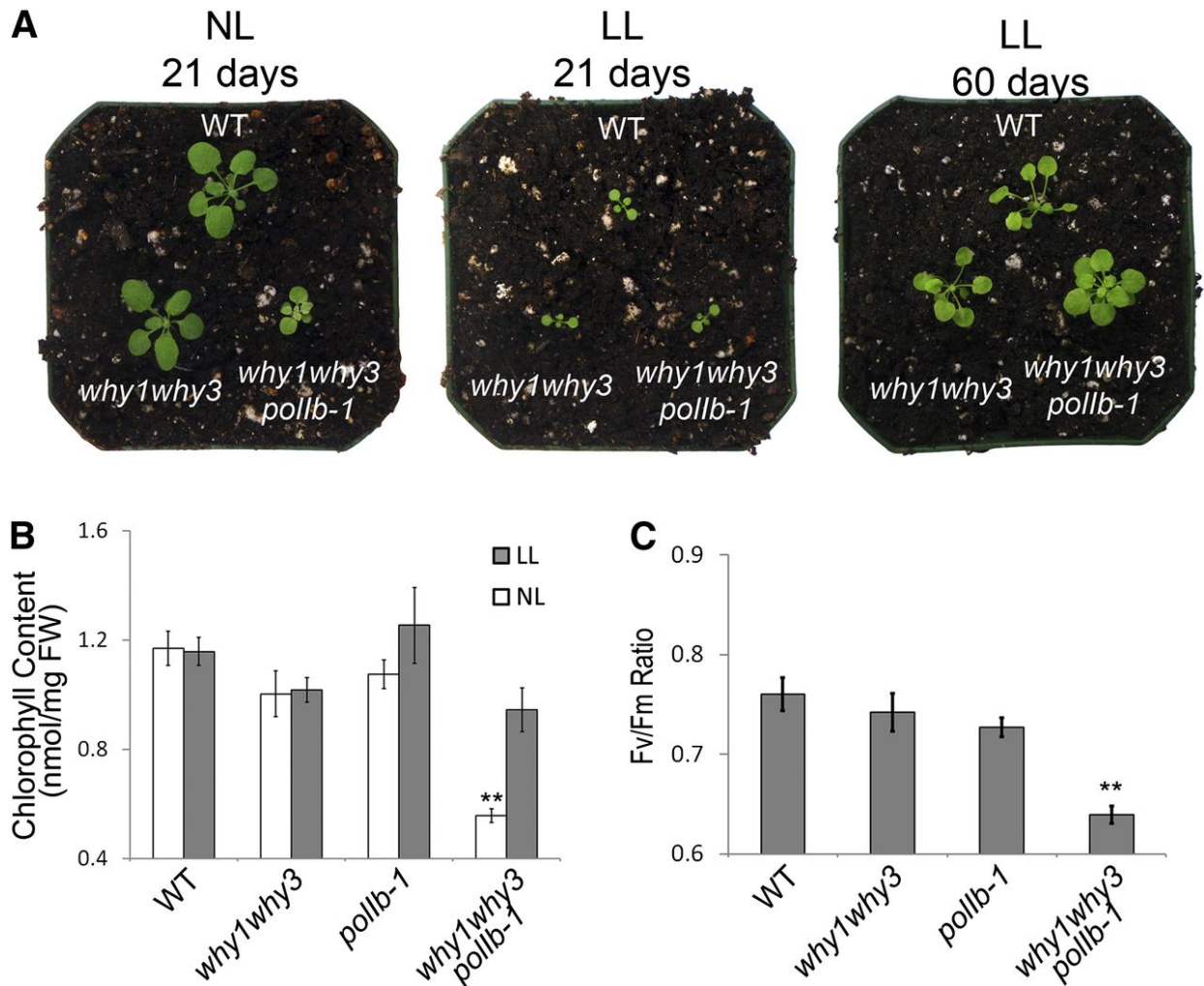


Figure 49. Growth under Low Light Conditions Rescues the *why1why3pollb-1* Phenotype.

(A) Representative photographs of wild type, *why1why3* and *why1why3pollb-1* Arabidopsis plants grown under normal (NL) or low light (LL) regimens.

(B) Chlorophyll content of each line grown under normal or low-light conditions for 21 and 60 days, respectively. White and Gray bars represent respectively NL and LL experiments. The error bars represent the standard deviation for five measurements on independent biological samples. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type.

(C) Fv/Fm ratios for 60-day-old plants of the indicated genotype grown under low-light conditions. The error bars represent the standard error of the mean (SEM). Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type. Every measurement was performed independently on at least five biological samples.

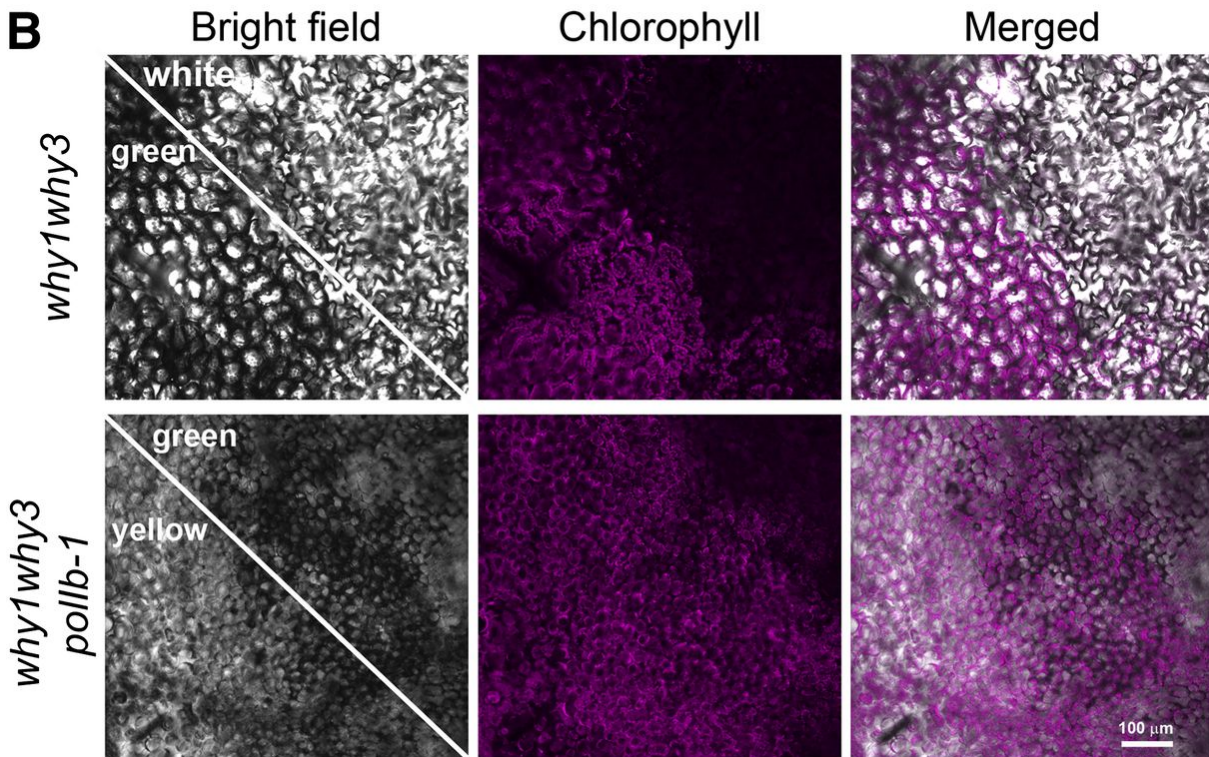
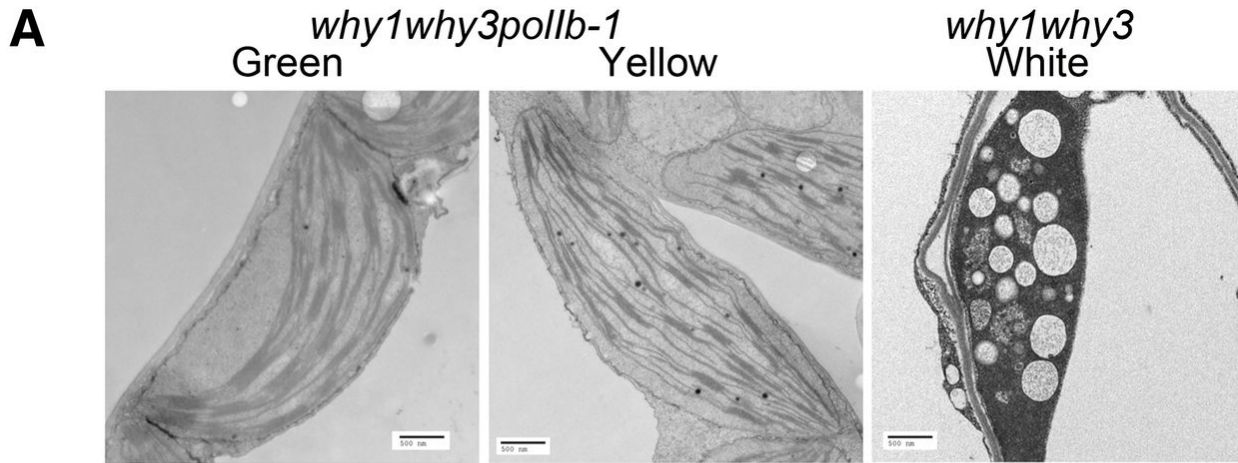


Figure 50. The *why1why3pollb-1* Yellow-Variiegated Sectors Are not Devoid of Photosynthetically Active Chloroplasts.

(A) Chloroplast ultrastructure observed by transmission electronic microscopy of green, white and yellow sectors of *why1why3* and *why1why3pollb-1* leaves.

(B) Microscopic analysis of chlorophyll autofluorescence in green, white and yellow sectors of 28-day-old *why1why3* and *why1why3pollb-1* leaves. Chlorophyll fluorescence is represented in magenta.

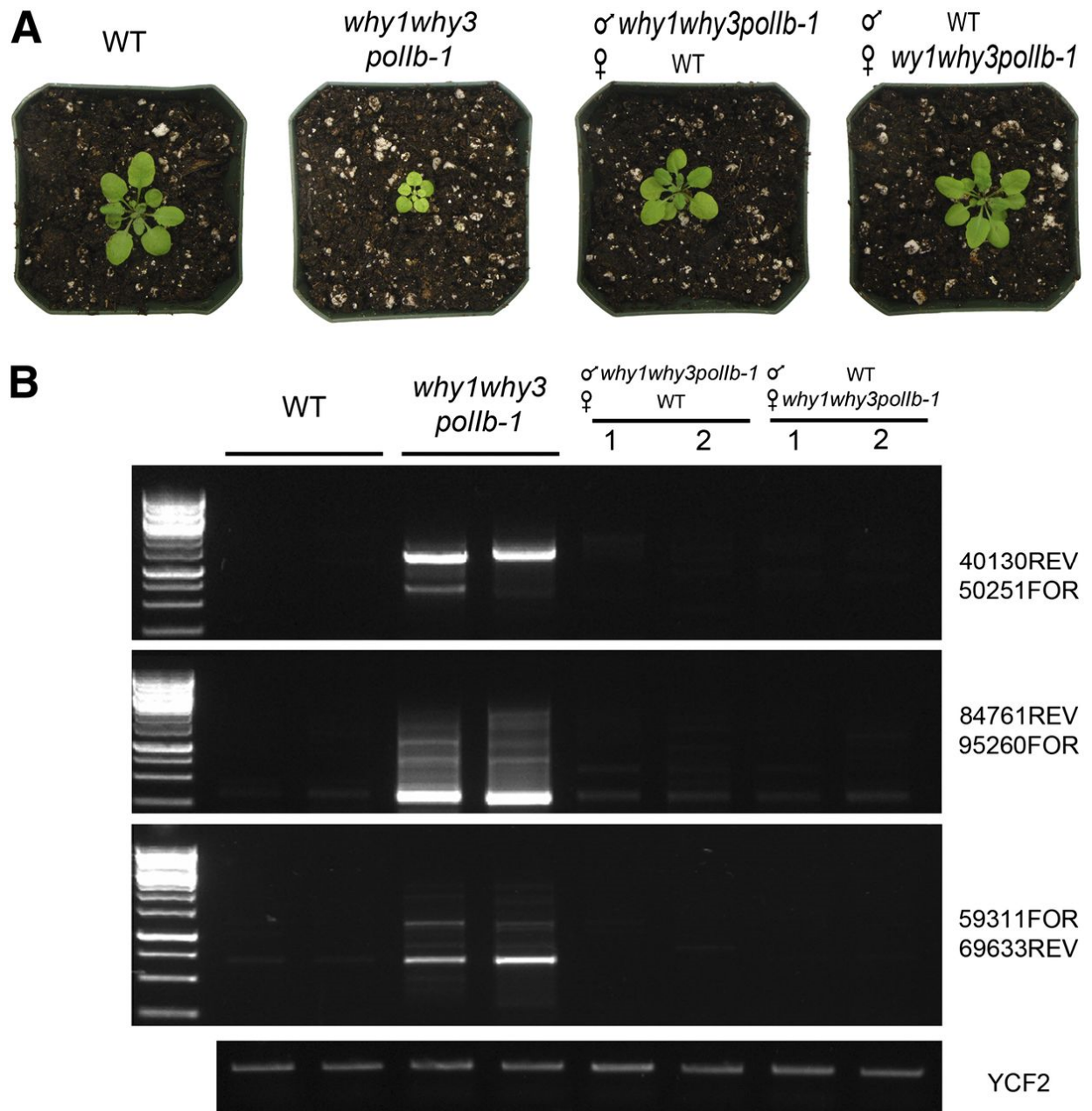


Figure 51. The *why1why3pollb-1* ptDNA Rearrangements and Phenotype Are not Maternally Inherited.

(A) Representative photographs of 21-day-old backcrossed *why1why3pollb-1* plants.

(B) Representative outward-facing PCR reactions showing the abundance of MMBIR-associated plastid DNA rearrangements of 14-day-old backcrossed *why1why3pollb-1* plants. PCR reactions were carried out using the indicated primer pairs. The amount of DNA loaded in each reaction was equilibrated relative to the amplification of a *YCF2* fragment. Every reaction was performed on three different biological samples and a representative biological duplicate is presented.

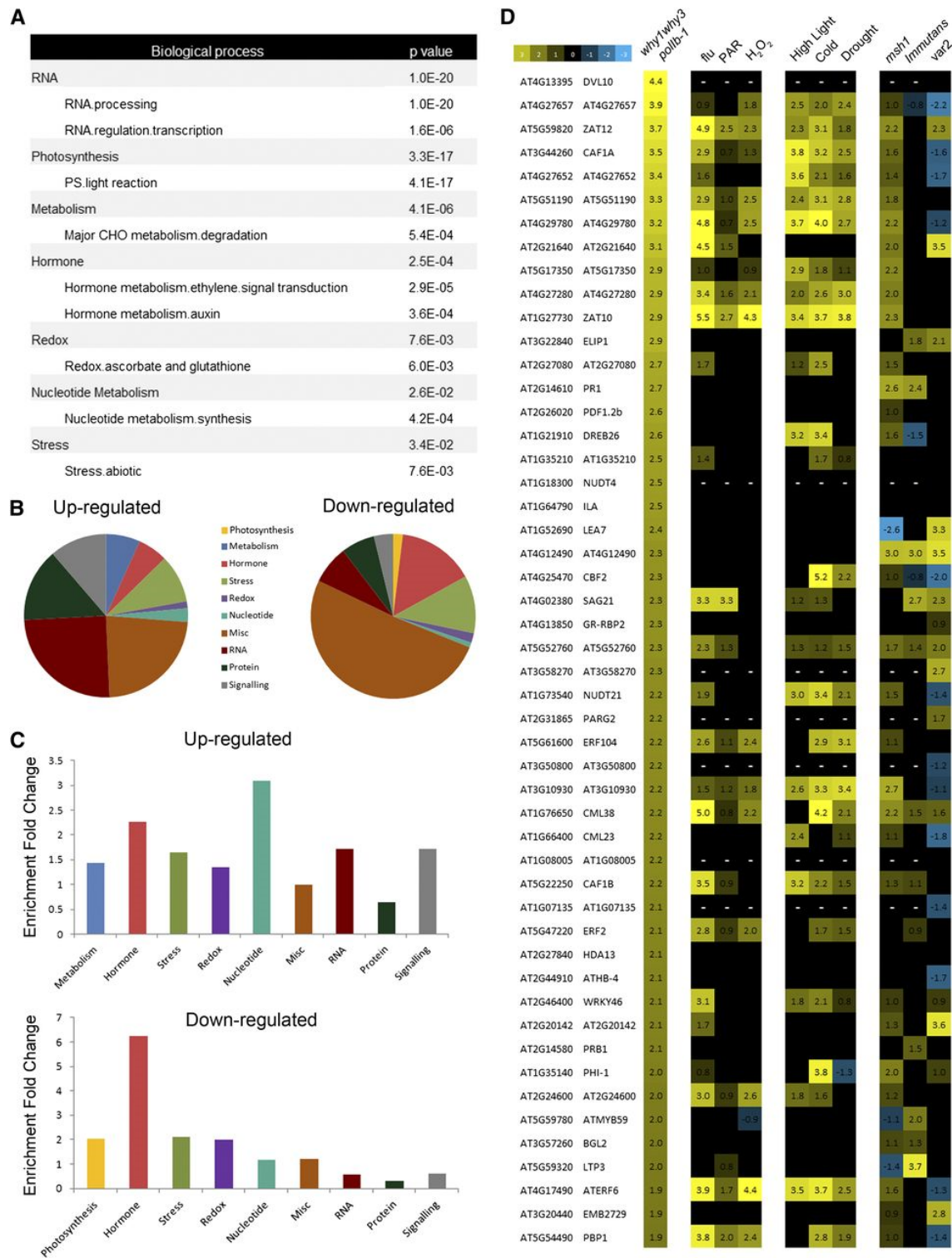


Figure 52. The *why1why3pollb-1* Mutant Exhibits a Complex Nuclear Genetic Reprogramming.

- (A)** MapMan analysis for GO terms enrichment of the entire *why1why3pollb-1* nuclear transcriptome.
- (B)** Pie charts presenting the biological processes associated to the significantly *why1why3pollb-1* up- and down-regulated genes (2-fold). To lighten the figure, the GO terms “unknown” and “not assigned” were removed from the analysis.
- (C)** Histogram presenting the enrichment of each biological process of the *why1why3pollb-1* transcriptome.
- (D)** Heatmap of the 50 most up-regulated *why1why3pollb-1* genes compared to other ROS-inducing conditions, abiotic stress responses and variegated mutants transcriptomes.

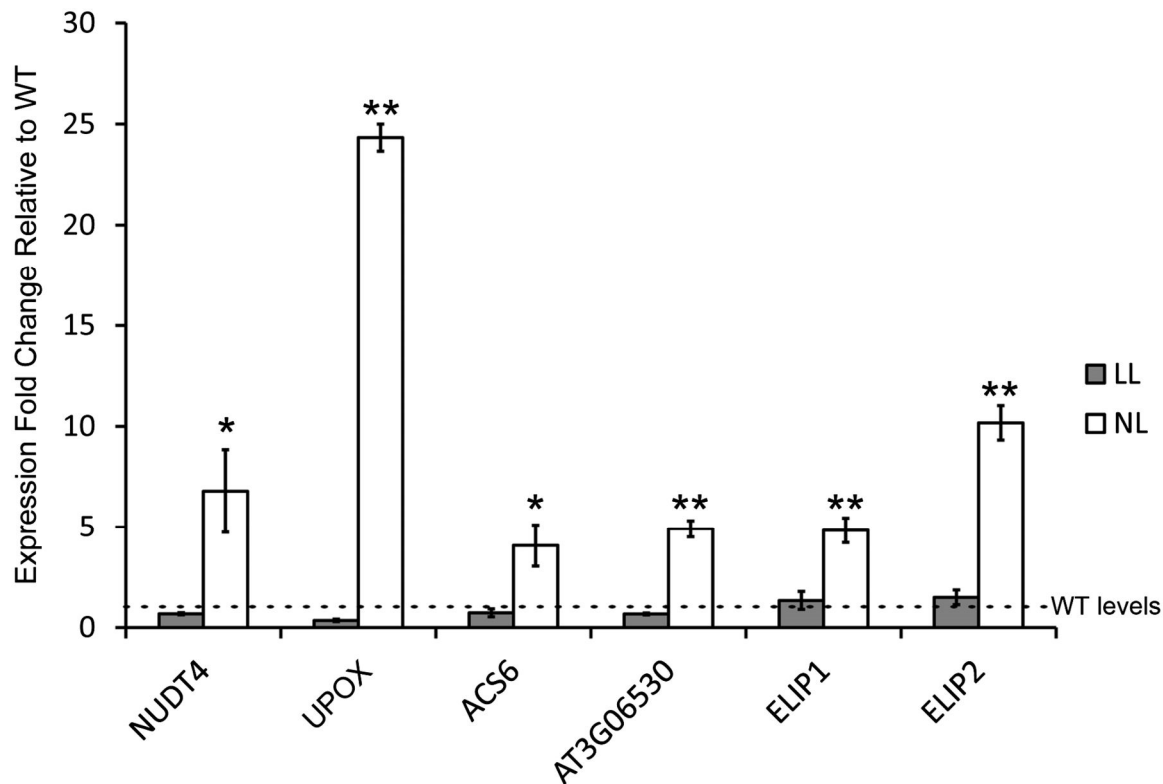


Figure 53. The *why1why3pollb-1* Nuclear Genetic Reprogramming Is Not Observed under Low-Light Conditions.

Histograms presenting the real-time PCR measurements of *NUDT4*, *UPOX* (AT2G21640), *ACS6*, *AT3G06530*, *ELIP1* and *ELIP2* expression in *why1why3pollb-1* relative to wild type plants for normal light (NL) and low light (LL) conditions. The error bars represent the standard error of the mean (SEM) of the biological triplicates. One and two asterisks indicate, respectively, a significant difference of a Student's t-test P-value ≤ 0.05 and 0.01 with the wild type.

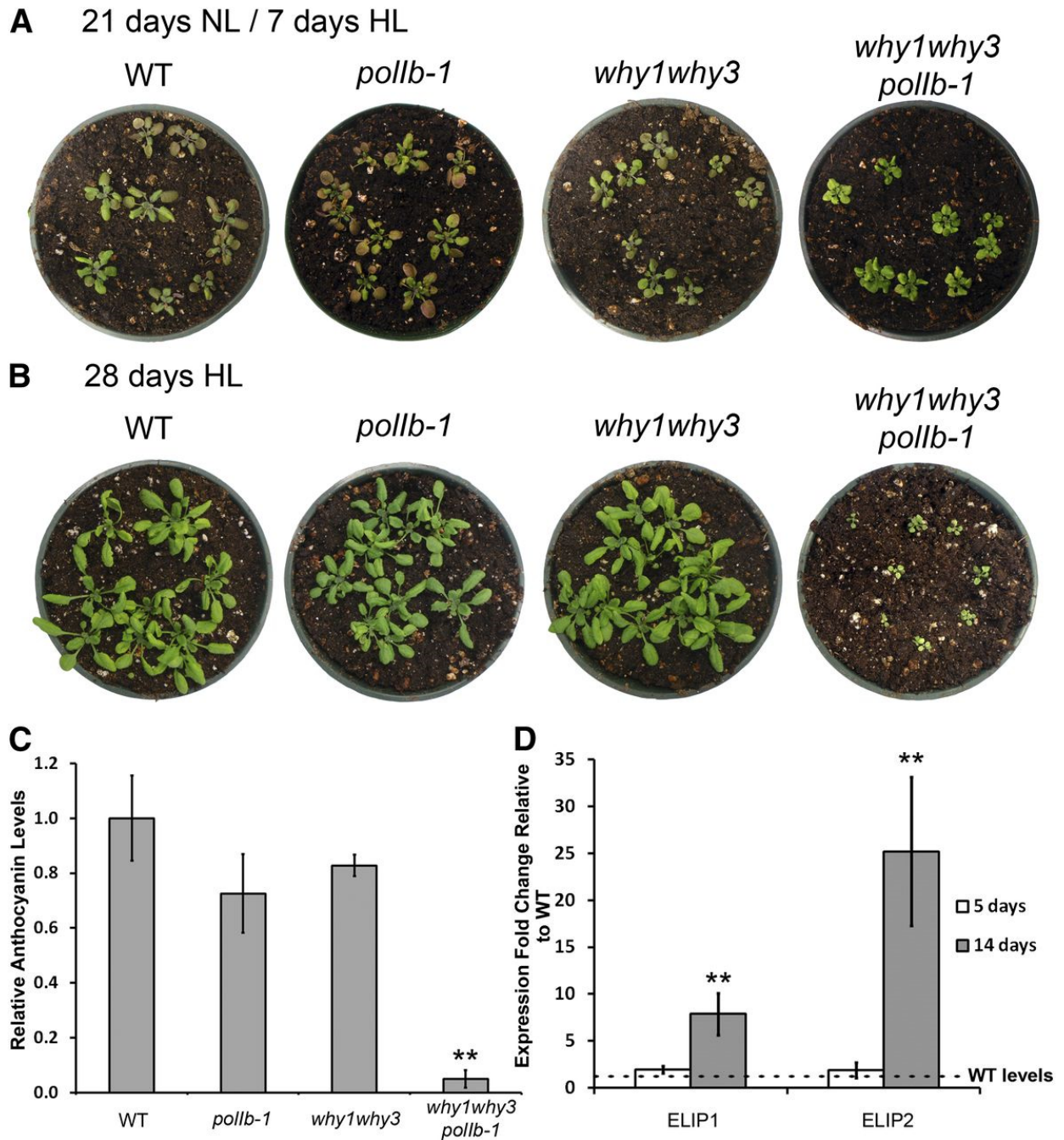


Figure 54. The Early Oxidative Stress in *why1why3pollb-1* Leads to Acquired Resistance to High Light.

(A) Photographs of representative *Arabidopsis* plants grown for 21 days at normal light (NL) and switched to high light (HL) for an additional seven days.

(B) Photographs of representative *Arabidopsis* plants grown for 28 days at high light (HL).

(C) Histogram presenting the relative anthocyanin content for the mutant lines and wild type plants switched for seven days at high light. The error bars represent the standard deviation of the biological

triplicates. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type.

(D) Histogram presenting the real-time PCR measurements of the *ELIP1* and *ELIP2* expression in *why1why3pollb-1* relative to wild type plants grown under normal light conditions for 5 and 14 days. The error bars represent the standard error of the mean (SEM) of the biological triplicates. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with the wild type.

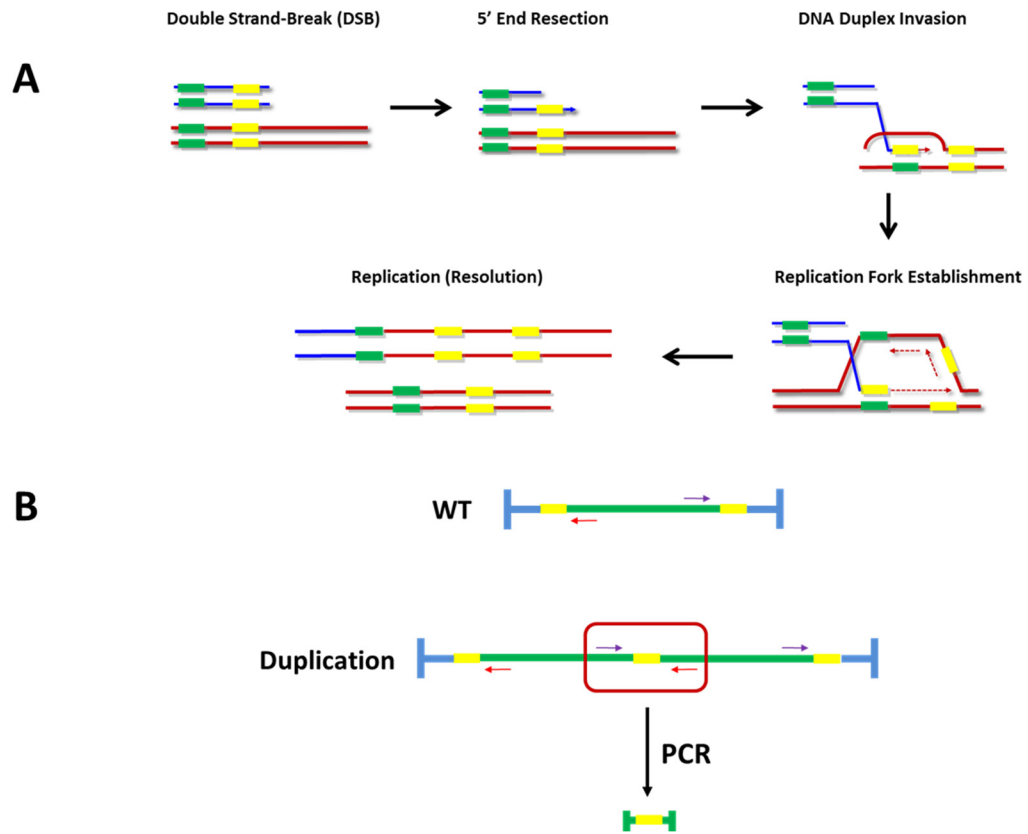


Figure 55. Microhomology-Mediated Break-Induced Replication (MMBIR) Leads to DNA Rearrangements.

A) Scheme presenting the repair of a DNA double-strand break by MMBIR. Dotted lines represent newly synthesized DNA. Yellow and green boxes represent microhomologous repeat sequences.

B) Scheme presenting the detection of MMBIR-associated DNA rearrangements by outward-facing PCR. Arrows represent the position of the PCR primers.

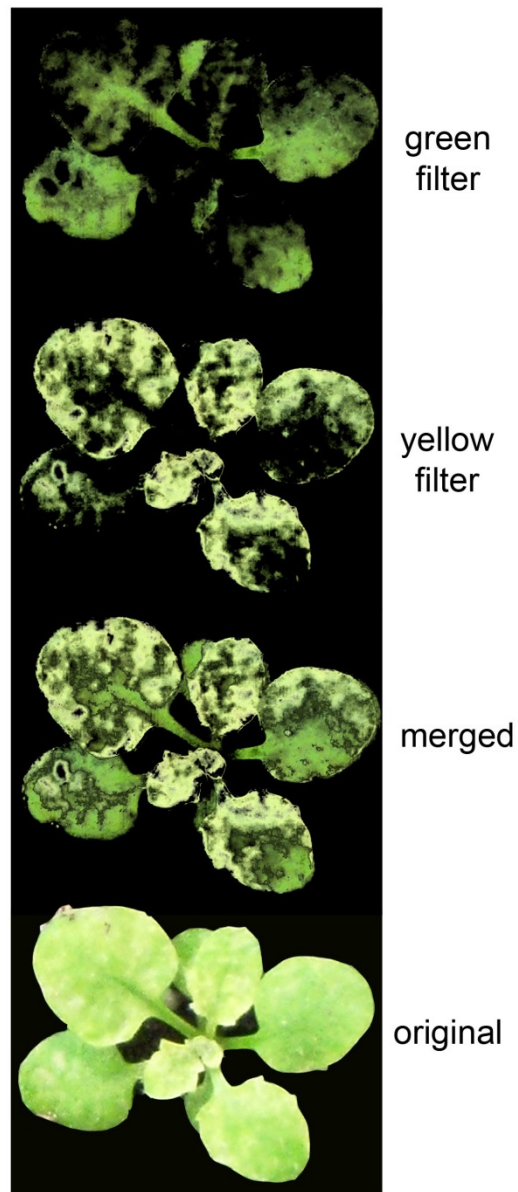


Figure 56. The *why1why3pollb-1* Mutant Develops Distinct Yellow and Green Sectors on Its Leaves.

Photographs of a representative 28-day-old *why1why3pollb-1* yellow variegated plant. From top to bottom, the images represent the use of a green filter, a yellow filter, a merge of the two filters and the original picture. Color filters were applied using Photoshop CS4 color range function.



Figure 57. Some *why1why3pollb-1* Plants Exhibit Yellow and White Variegation Simultaneously.

Photograph of a representative 28-day-old *why1why3pollb-1* yellow and white variegated plant.

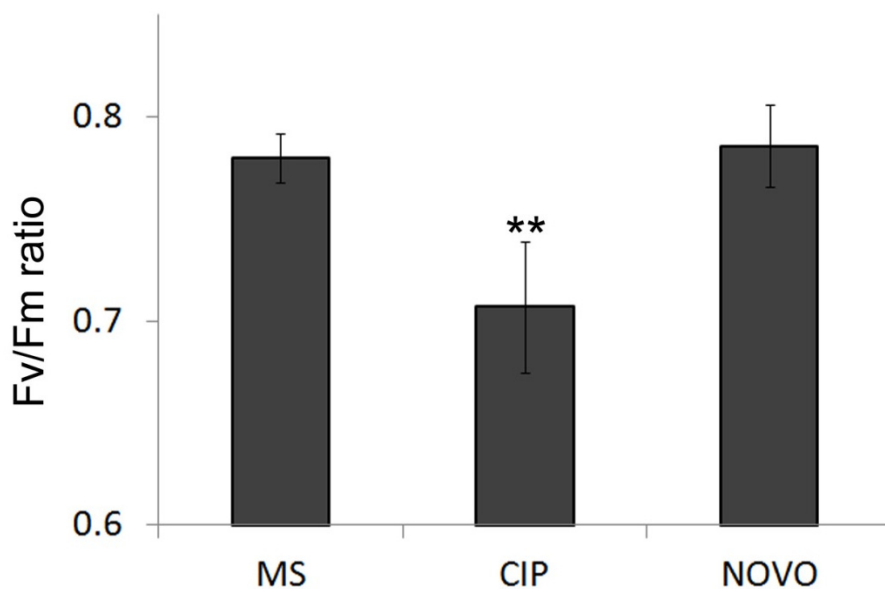


Figure 58. Ciprofloxacin Negatively Impacts Photosynthetic Efficiency in Arabidopsis.

Fv/Fm ratios of 15-day-old plants treated from days 10 to 15 to 50 μ M NOVO, 10 μ M CIP or no drug (MS). The error bars represent the standard deviation. Two asterisks indicate a significant difference of a Student's t-test P-value ≤ 0.01 with MS. Measurements were performed independently on at least five biological samples.

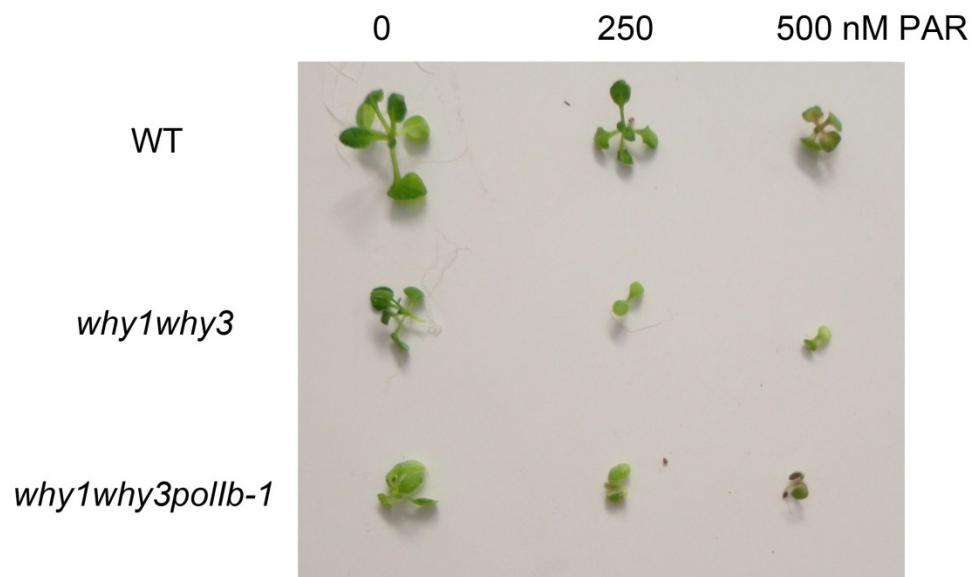


Figure 59. *why1why3* and *why1why3pollb-1* Plants Are Hypersensitive to Paraquat.

Representative photograph of 3-week-old plants grown under normal light conditions on MS basal media containing 0, 250 or 500 nm paraquat (PAR).

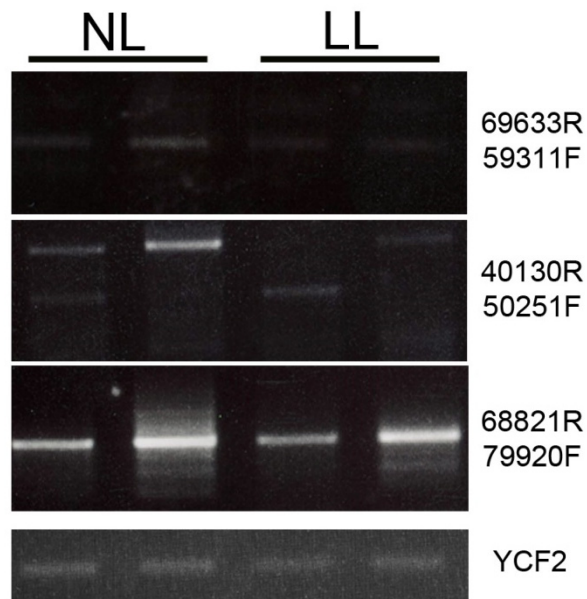


Figure 60. The Abundance of MMBIR-Associated ptDNA Rearrangements Is Little Affected by the Light-Regimen in *why1why3pollb-1*.

Representative outward-facing PCR reactions showing the abundance of MMBIR-associated plastid DNA rearrangements of 3-week-old *why1why3pollb-1* plants grown either under normal (NL) or low-light conditions (LL). PCR reactions were carried out using the indicated primer pairs. The amount of DNA loaded in each reaction was equilibrated relative to the amplification of a *YCF2* fragment. Every reaction was performed on three different biological samples and a representative duplicate is presented.

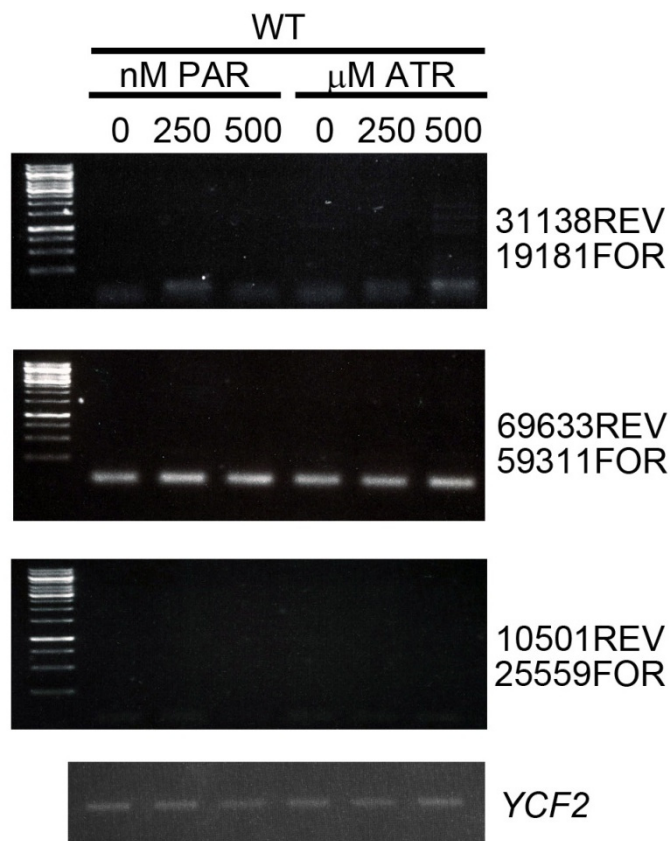
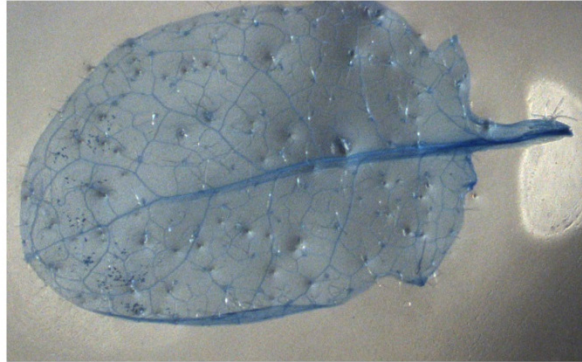


Figure 61. Paraquat and Atrazine do not induce MMBIR-Associated ptDNA Rearrangements.

Representative outward-facing PCR reactions showing the abundance of MMBIR-associated plastid DNA rearrangements of 3-week-old *Arabidopsis* plants grown in the presence of the ROS-inducing agents paraquat (PAR) and atrazine (ATR). PCR reactions were carried out using the indicated primer pairs. The amount of DNA loaded in each reaction was equilibrated relative to the amplification of a *YCF2* fragment. Every reaction was performed on three different biological samples and a representative reaction is presented. The results indicate that MMBIR-associated ptDNA rearrangements are not the consequence of superoxide, singlet oxygen or hydrogen peroxide oxidation.

WT



why1why3pollb-1

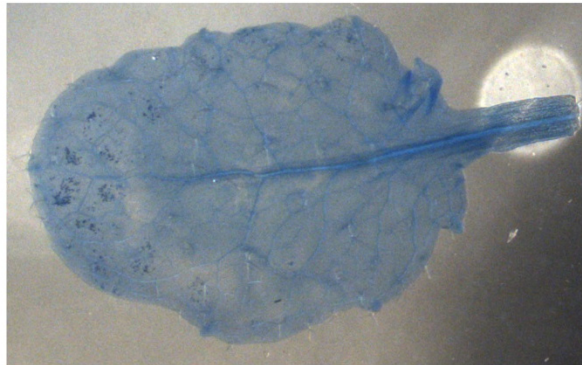


Figure 62. Elevated ROS Levels Do not Lead to Cell Death in *why1why3pollb-1*.

Representative photographs of 4-week-old plants grown under normal light conditions and stained with trypan blue.

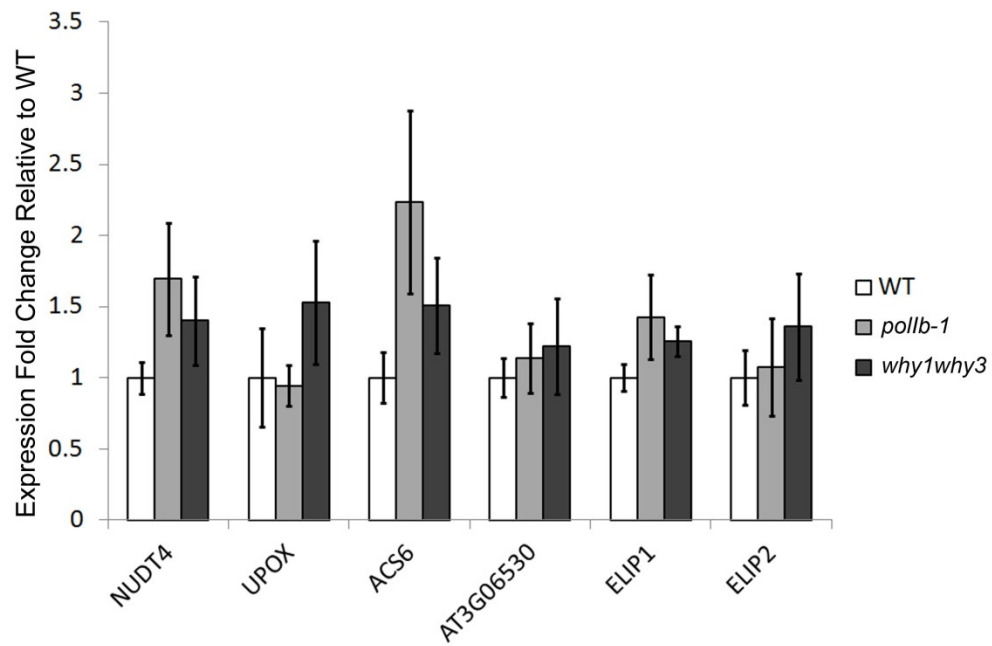


Figure 63. The *why1why3pollb-1* Nuclear Genetic Reprogramming Is Not Due to the *Pollb-1* and *why1why3* Genetic Backgrounds.

Histograms presenting the real-time PCR measurements of *NUDT4*, *UPOX* (AT2G21640), *ACS6*, *AT3G06530*, *ELIP1* and *ELIP2* expression in wild type, *pollb-1* and *why1why3* plants. The error bars represent the standard error of the mean (SEM) of the biological triplicates.

Tableau VIII. Primer Pairs Used for the Specific Amplification of MMBIR-Associated ptDNA Rearrangements and Described Previously (196).

Primer number	Primer name	Sequence
1	31138REV	AAT AGC CTA GCG CAT CGA ACG
	19181FOR	CAG CAT CTG ATG TGC GTA CAG
2	40130REV	GGG AAG AGG AGG AAC ATG TCA
	50251FOR	GGC CCC TTC CTA ATT GGA GAT
3	84761REV	CCT TAG GCA CGG CCA TAC ATA
	95260FOR	AAC CTG CTA GTG GAG GAA GAC
4	105961REV	GCG TTA CTC AAG CCG ACA TTC
	116007FOR	AAA CCA CTC ATG CCC GGC AAT
5	69633REV	GAG GAA CAT GTA CTC GGG TGT
	59311FOR	AGG CCT AGT CTT TCC GGG AAT
7	50461REV	GAG ACG TCT ACC ACA ACA ACC
	59311FOR	AGG CCT AGT CTT TCC GGC AAT
8	68821REV	ACG CGG ATA TAT AGC TCG GAG
	79920FOR	CCA TGC CGT TCG GAA GTG ATT
9	10501REV	TGA CTG CCT TAC CAA TCG TCG
	25559FOR	CGG CGA CCA ATC CTT CCT AAT
ycf2	ycf2REV	TGT TTC GCC TCT TAC TCG GAG
	ycf2FOR	GAT CTC TGA GAG CTG TTT CCG

Tableau IX. Primer Pairs Used for Quantitative RT-PCR experiments.

Gene	Forward Primer	Reverse Primer
B-Tub	TCGTTGGGAGGAGGCACAGGT	GCTGAGTTTGAGGGTACGGAAGCAG
ACS6	CCTGACCGGATTGTTATGAGCGG	TCTATCAAACCCTGGATAATAAGGGGTTGG
AT2G21640	CCATGGACAAGATATTCGGGAGCTTCTC	GCTGTCTGAAGAGGAGAGTGAATCTCG
AT3G06530	CGCTGAACCACGAGGTGCTAATG	GCCAAGAATGGAATGGTTTCAGCAAGC
ELIP1	CTCACCTTAGGTTTCATAGGAGGAGG	GCAACAGCATCGTTCAACATGC
ELIP2	GCAAGCCTAAGGTGAGTACGAAG	ACCGTCAGAGATCTGAGCAAACACG
NUDT4	CCCGCACAGATCGAGAAAGTTGTC	ACGGTACACATCCGACGACTTGAC
UBQ5	GTAACGTAGGTGAGTCCA	GACGCTTCATCTCGTCC

SUPPLEMENTAL MATERIALS AND METHODS

Plant Material and Growth Conditions

Arabidopsis (*Arabidopsis thaliana*; ecotype Columbia-0) mutant lines *why1why3* and *why1why3pollb-1* were reported previously (197, 198). Seeds were sown on soil, vernalised for 3 days at 4°C and grown either under normal light (100 $\mu\text{mol m}^{-2} \text{s}^{-1}$) or low light (20 $\mu\text{mol m}^{-2} \text{s}^{-1}$) at 22°C on a 16 h day/8 h dark cycle. When grown at low light, plants were germinated at 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for one week prior to their transfer at 20 $\mu\text{mol m}^{-2} \text{s}^{-1}$. For drug hypersensitivity experiments, seeds were sterilized and sown on Murashige and Skoog basal media (MS medium, Sigma-Aldrich) supplemented with 1 % sucrose, 0.8 % agar and the indicated amount of drug.

Measurement of PSII Efficiency

For each condition, Fv/Fm measurements were performed on 15-day-old plants treated from days 10 to 15 with 50 μM NOVO, 10 μM CIP or no drug with an Open FluorCam (Photon Systems Instruments) according to the manufacturer's instructions.

Detection of ptDNA Rearrangements

Total DNA was isolated for each plant sample using a cetyl trimethylammonium bromide DNA extraction protocol (174). For every DNA samples, ptDNA amount was equilibrated with a low-cycle amplification of a *YCF2* DNA fragment. Primers and approaches for rearrangements detection were described previously (196) and PCR products were visualized after migration on GelRed (Biotium) stained agarose gels. Every reaction shown is representative of the eight primer pairs presented in Supplemental Table S2.

Trypan blue staining

The trypan blue staining of dead cells was conducted as described previously (174) with 4-week-old *Arabidopsis* leaves.

Quantitative RT-PCR Analysis

cDNA synthesis was carried out using a first-strand cDNA synthesis kit (Fermentas) according to manufacturer's instructions. cDNAs were diluted 20-fold prior to qRT-PCR experiments. Primers used for qRT-PCR are listed in Supplemental Table S3. Every reaction was carried out on biological and technical triplicates and normalised relative to UBI5. Real-time qPCR reactions were done using the SYBR green master mix (SABiosciences, www.sabiosciences.com) according to the manufacturer's instructions. Melting curves confirmed the amplification of a unique PCR product for each of the qRT-PCR primer pairs used. LightCycler480 (Roche) was used for qRT-PCR experiments and data analyzed using the Light-Cycler480 software version 1.5.

SUPPLEMENTAL REFERENCES

- Cappadocia, L., Maréchal, A., Parent, J.S., Lepage, E., Sygusch, J., and Brisson, N.** (2010). Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. *Plant Cell* **22**: 1849-1867.
- Maréchal, A., Parent, J.S., Veronneau-Lafortune, F., Joyeux, A., Lang, B.F., and Brisson, N.** (2009). Whirly proteins maintain plastid genome stability in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 14693-14698.
- Parent, J.S., Lepage, E., and Brisson, N.** (2011). Divergent roles for the two Poll-like organelle DNA polymerases of Arabidopsis. *Plant Physiol.* **156**: 254-262.
- Weigel, D., and Glazebrook, J.** (2002). *Arabidopsis: A Laboratory Manual* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).

Chapitre 6: Quatrième article

Time- and Cost-Efficient Identification of T-DNA Insertion Sites through Targeted Genomic Sequencing.

Article publié dans le journal PLoS One, 8: e70912 (2013)

6.1 Mise en contexte:

Tous les mutants présentés jusqu'ici ont été identifiés à l'aide d'un criblage génétique se limitant uniquement aux protéines retrouvées dans le nucléoïde. Bien que ce criblage ait grandement élargi notre compréhension des processus permettant le maintien de la stabilité génomique du plastide, un regard plus global sur ce processus permettrait d'identifier d'autres acteurs qui agissent sans être recrutés à l'ADN. Pour nous débarrasser de ce biais, nous avons procédé à un criblage génétique à l'aide d'une imposante collection de plus de 50, 000 lignées d'insertions différentes (199). Un des désavantages principaux de ce genre de criblage est la difficulté à identifier le site de mutation dans les plantes d'intérêt puisque les mutants n'ont pas été préalablement isolés. Au moment où nous entreprenions ce projet, tous les protocoles pour identifier les sites d'insertions se basaient sur des techniques laborieuses et inefficaces d'amplification PCR (156, 200). Afin d'améliorer l'efficacité de cette étape, nous avons développé une approche de séquençage de nouvelle-génération pour localiser à haut débit le site d'insertion dans les plantes d'intérêt. Les résultats obtenus par cette technique ont permis d'identifier de nouveaux facteurs possiblement impliqués dans les processus de maintien de la stabilité du génome plastidique.

Les résultats présentés ici sont principalement le fruit de mon travail. EZ a aidé lors de l'analyse des résultats et l'écriture de l'article. Brian Boyle a réalisé les expériences de séquençage et a participé à l'élaboration du projet, ainsi qu'à l'écriture de l'article. NB a participé à l'élaboration du projet et à l'écriture.

Time- and Cost-Efficient Identification of T-DNA Insertion Sites through Targeted Genomic Sequencing

Étienne Lepage¹, Éric Zampini¹, Brian Boyle² and Normand Brisson^{1,3}

¹Department of Biochemistry, Université de Montréal, Montréal, Qc, Canada.

²Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, Qc, Canada.

6.2 Abstract

Forward genetic screens enable the unbiased identification of genes involved in biological processes. In *Arabidopsis*, several mutant collections are publicly available, which greatly facilitates such practice. Most of these collections were generated by agrotransformation of a T-DNA at random sites in the plant genome. However, precise mapping of T-DNA insertion sites in mutants isolated from such screens is a laborious and time-consuming task. Here we report a simple, low-cost and time efficient approach to precisely map T-DNA insertions simultaneously in many different mutants. By combining sequence capture, next-generation sequencing and 2D-PCR pooling, we developed a new method that allowed the rapid localization of T-DNA insertion sites in 55 out of 64 mutant plants isolated in a screen for gyrase inhibition hypersensitivity.

6.3 Introduction

Genetic studies paved the way in understanding most biochemical processes in plants. Forward genetic screening requires mutant collections produced either by chemical mutagenesis using mutagens such as ethyl methanesulfonate, or by insertional mutagenesis, obtained by the agrotransformation of a T-DNA cassette [1]. The public availability of T-DNA insertion lines collections is particularly interesting as it considerably shortens the time required to perform a forward genetic screen. The function of many *Arabidopsis* genes have thus been characterized using the SALK collection, which is composed of more than 88,000 lines containing an inactivating T-DNA [2]. Additionally, collections of activation lines, transformed with a T-DNA containing repeated cauliflower mosaic virus (CaMV) 35S enhancer regions, can also be used to characterize gene function [3,4]. In these collections, in addition to disrupting the coding sequences, insertion of the T-DNA near the promoter region of a gene gives rise to overexpressors. These overexpressors allow the characterization of large families of genes, which might be redundant and generally missed by loss-of-function genetic screens [5].

The laborious work of mapping the T-DNA insertion sites in each mutant is one of the main problems that limit the use of large mutant collections for forward genetic screens. Many techniques have been proposed to identify T-DNA insertion sites, such as Tail-PCR [6], inverse PCR (IPCR) [7] and restriction site extension PCR (RSE-PCR) [8], but these methods present important limitations such as inefficient ligation step, the need of restriction enzymes that cut both the T-DNA and the genomic part at a reasonable distance and the generation of non-specific products through PCR [8]. Taken together, these limitations greatly affect the scalability and processivity of these techniques.

Recently, some studies have started to demonstrate the huge potential of next-generation sequencing to identify insertion sites. Indeed Illumina sequencing has been used to i) map the position of transposons in highly transposable maize lines [9], ii) identify insertion sites of LORE1 retrotransposon in *Lotus japonicus* [10], and iii) identify mutants for leaf shape abnormalities by the low-coverage sequencing of a pool containing genomic DNA from four distinct mutants [11].

Here, we describe targeted genomic sequencing, a new technique that allows the simultaneous identification of multiple insertion sites in a complex DNA sample. Using biotinylated primers specific for the extremities of the T-DNA, the regions flanking the insertion sites of 64 different mutants pooled together were specifically enriched from total genomic DNA. Roche GS-FLX + sequencing allowed the identification of 31 genes in which the T-DNA cassette was inserted. As an example, our approach is described with a screen carried out using gyrase inhibitors that specifically affect plant organelle genomes topology [12,13,14].

6.4 Material and Methods

Plant Material, Growth Conditions

The *Arabidopsis* (*Arabidopsis thaliana*; ecotype Columbia-4) SK mutant collection was kindly provided by Dr. Isobel Parkin [3]. Seeds were sterilized, sown on Murashige and Skoog basal media (Sigma-Aldrich) supplemented with 1 % sucrose and 0.8 % agar and vernalised for 3 days in the dark at 4 °C. Plants were grown under normal light (100 $\mu\text{mol m}^{-2} \text{s}^{-1}$), at 22 °C on a 16 h day/8 h dark cycle and the phenotype was assessed at seven days.

Mutant Collection Screening

Approximately 150,000 plants of the SK collection were screened on 0.125 μM ciprofloxacin (CIP). Mutants presenting white sectors on the first true leaves were transplanted on soil to ensure their survival and allow them to set seeds. The progeny of each mutant was grown under the same conditions on medium containing either 0.125 μM CIP, 50 μM novobiocin (NOVO) or no drugs.

Library Preparation

Total DNA was isolated for each plant sample using a cetyl trimethylammonium bromide (CTAB) DNA extraction protocol [15]. DNA from all samples was pooled in an equimolar ratio and a single GS-FLX+ rapid library was produced according to the manufacturer instructions (Roche, 454 Sequencing). The library was amplified by ligation mediated PCR

using the 454 A and B primers as described in the general guidelines provided in the NimbleGen SeqCap EZ Library LR User's guide.

Target Enrichment

Previous work demonstrated high specificity of 70 mer oligonucleotides in microarray analysis [16]. Therefore, three biotinylated 70 nucleotides long probes were designed to each extremity of the T-DNA sequence, ensuring that there was no significant sequence similarity to the Arabidopsis genome by blast analysis. The sequences are as follow: For the right border, RB1; AAC ATG GTG GAG CAC GAC ACT CTC GTC TAC TCC AAG AAT ATC AAA GAT ACA GTC TCA GAA GAC CAG AGG G, RB2; CTA TTG AGA CTT TTC AAC AAA GGG TAA TAT CGG GAA ACC TCC TCG GAT TCC ATT GCC CAG CTA TCT GTC A, RB3; TCA AAG ATA CAG TCT CAG AAG ACC AGA GGG CTA TTG AGA CTT TTC AAC AAA GGG TAA TAT CGG GAA ACC T. For the left border, LB1; ATG GAA ATT ATC TGC CTA ACC GGC TCA GTT CTG CGT AGA AAC CAA CAT GCA AGC TCC ACC GGG TGC AAA G, LB2; AAC GCC ATC CGA CGG ATG ATG TTT AAA AGT CCC ATG TGG ATC ACT CCG TTG CCC CGT CGC TCA CCG TGT T, LB3; AAG GTG CAC ATG GCT CAG TTC TCA ATG GAA ATT ATC TGC CTA ACC GGC TCA GTT CTG CGT AGA AAC CAA C. Target enrichment was performed using the SeqCap EZ hybridization and wash kit (Roche Nimblegen) using the general guidelines provided in the NimbleGen SeqCap EZ Library LR User's guide. Briefly, 10 µl of plant capture enhancer (Roche Nimblegen) and 5 µl of 100 µM hyb enhancing 454 A and B primers were added to 1 µg of amplified library and then dried. The A and B primers are added to inhibit unspecific interactions between the flanking primer regions in the DNA molecules of the library. The dried mixture was resuspended in 7.5 µl of 2X SC hybridization buffer and 3 µl of SC component A and heated to 70 °C for 10 minutes. After a quick spin, 4.5 µl of the capture oligonucleotides solution in water (3.75×10^6 molecules of each biotinylated oligonucleotide) were added. The amount of oligonucleotides added represents about a 5 fold excess of capture oligonucleotides to the number of genome molecules present. The hybridization mixture was incubated at 95 °C for 10 minutes and then at 47.5 °C for 40 hours. The hybridization mixture was put in contact with Streptavidin beads (Invitrogen) and non-captured material was washed away according to the

NimbleGen SeqCap EZ Library LR User's guide. Captured material was amplified with the 454 A and B primers.

GS-FLX + Sequencing and Analysis

Emulsion PCR and GS-FLX+ sequencing was performed according to manufacturer's instructions at the Plateforme d'Analyses Génomiques of the Institut de Biologie Intégrative et des Systèmes (Laval University, Québec, Canada). Raw sequencing reads were mapped to the T-DNA sequence vector (pSKI015 – Genbank [AF187951](#)) and the Arabidopsis genome using the gsMapper module of Newbler v.2.5.3. We used the Linux/Unix grep command to rapidly pullout the lines containing the word “partial” from the 454ReadStatus.txt accessory file that is created by Newbler following the mapping of the reads to either the Arabidopsis genomes or the T-DNA vector. Alternatively, the 454ReadStatus file could have been filtered in a spreadsheet.

2D-PCR Pooling

Two distinct sets of 8 pools were prepared with an equivalent concentration of genomic DNA extracted from the different mutants in such a way that each mutant line is represented in a unique pool combination. PCR was conducted on 100 ng of the pooled genomic DNA with a T-DNA and a gene specific primer.

6.5 Results

Chemical Screening using Gyrase Inhibitors Identified 64 Sensitive Plants

To identify the genes involved in the maintenance of organelle genome topology, we set up a screen for gyrase inhibition hypersensitivity using the SK collection, which is composed of approximately 55,000 different activation tagged lines [3]. A global view of the screen is summarized in Figure 64. The first round of screening consisted of growing 150,000 tagged lines on 0.125 μ M ciprofloxacin (CIP), a gyrase inhibitor that introduces DNA double strand breaks (DSBs) in plant organelle genomes, and isolating the plants that showed signs of

hypersensitivity [17]. Progeny of these plants were then submitted to a second round of screening, in the presence or absence of CIP (Figure 1). In addition, plants were also germinated on a medium containing novobiocin, a gyrase inhibitor that does not introduce DSBs [18]. The second round of screening identified 64 plants hypersensitive to CIP, of which eight were sensitive to both novobiocin and ciprofloxacin. We then proceeded to identify the T-DNA insertion sites for all the 64 plants.

Targeted Genomic Sequencing for High-Throughput Insertion Sites Identification

The identification of T-DNA insertion sites in 64 plants by classical methods such as TAIL-PCR or IPCR represents a highly laborious and time-consuming task [8]. However, given the recent advances in next-generation sequencing coupled with targeted gene enrichment, it was foreseeable to achieve multiple insertion sites identification in a cost and time efficient manner. Figure 65 describes the different steps used in this method. Briefly, a next-generation shotgun library was prepared with a pool containing equivalent amounts of CTAB-extracted genomic DNA from every line. Then, biotinylated primers complementary to the T-DNA ends were hybridized to the genomic DNA library and hybridized target T-DNA recovered using the SeqCap EZ hybridization kit. It was expected that the regions flanking the T-DNA insertion sites would be enriched as well during this procedure. Following amplification of the captured material using 454 specific primers, the efficiency of the T-DNA capture was assessed by qPCR. Finally, ROCHE 454 GS-FLX + sequencing was carried out to identify the region flanking the T-DNA in each line. A major advantage of GS-FLX + is that the read length can reach a thousand bases, increasing the probability of getting a hybrid fragment composed of a T-DNA and a genomic part. The sequencing reads were aligned against the T-DNA sequence using the gsMapper module of Newbler v2.5.3. Of the 115,000 reads obtained, 28,023 reads mapped to a unique position of the T-DNA cassette and 19,090 reads mapped to the enhancer repeat region, indicating that the sequence capture worked efficiently, with more than 40% of the reads mapping to the T-DNA (Figure 66). Most of these reads fully mapped to the T-DNA. Reads mapping closely to the T-DNA right border were more abundant compared to the ones mapping closely to the left border, most likely due to the design of the probes within the 4x repeated CaMV 35S enhancer adjacent to the right border (Figure 66). Nevertheless, approximately 4,000 reads that partially mapped to the T-DNA had a remaining portion also

mapping to the Arabidopsis genome. The sequencing identified 31 genes, suggesting that more than one plant could be mutated for the same gene.

Tracking the Mutations by Pool-PCR.

Many next-generation sequencing approaches take advantage of barcoding, which consists of adding a unique short DNA sequence to each genomic sample to easily distinguish individuals. However, library production costs can be prohibitive when dealing with many different samples. In order to minimize these costs, 2D-PCR pooling was used in the present study to match the identified mutations to each CIP-sensitive plants (Figure 67). Genomic DNA from the 64 CIP-sensitive plants was split into two sets of eight pools in a manner ensuring that each genomic sample would be present in a unique set combination [19]. Figure 67B presents an example for the preparation of the pools for 16 plants. Then a specific PCR reaction was carried out for every candidate gene (31 in our case) and the amplified products were separated on gel. Depending on the number of lines that carries the tested insertion, a specific pattern of band is obtained. When the mutation is present only once, a single band is detected in the 1st and 2nd pool which can be linked to a single plant (Figure 67C, upper part). When more than one plant has the insertion, different outcomes are possible and require confirmation of the insertion for every candidate gene (Figure 67C, lower part). However, even when dealing with the first scenario, it is always best to confirm the insertion with a second PCR.

Although more time-consuming than barcoding, this straight-forward approach was shown to be very efficient as 86% of the plants (55/64) could be successfully genotyped. The remaining nine plants either contain an insertion site that could not be identified by our technique or other DNA modifications that would cause the phenotype. From the 64 plants isolated by the screen, eight were hypersensitive to both ciprofloxacin and novobiocin, which suggests a role for the mutated genes in the control of organelle genome topology, and 56 plants were solely sensitive to ciprofloxacin. By 2D-PCR pooling, we determined that the eight mutants sensitive to both gyrase inhibitors were only associated to three different insertions and thus represent three lines. Two plants were mutated in the 5' UTR of the *ARL2* gene (AT1G59980), three in the intergenic space between *Wrky49* (AT5G43290) and *GDPD3* (AT5G43300), and three in the 6th intron of AT2G24350. It is not surprising that we obtained more than one plant for these

lines as we screened more plants than the total expected number of lines present in the collection. The mapping of the insertion sites of these three different lines is presented in Figure 68. The other 28 mutations, which we could associate to 47 of the 64 isolated plants, were not found among the eight CIP/NOVO-sensitive plants and will be described elsewhere.

6.6 Discussion

In this report we describe a time-efficient and low cost targeted genomic sequencing technique for high-throughput identification of insertion lines in forward genetic screens. Indeed, with the combination of next-generation sequencing and 2D-PCR pooling, we show that it is possible to identify most insertion sites in a large collection of mutants. The approach described here to capture the T-DNA ends is very efficient since more than 40% of the reads obtained corresponded to the T-DNA cassette. This is particularly impressive considering the complexity of the genomic sample used, which was composed of a genomic DNA pool from 64 different genomic extractions. The fact that a majority of the captured sequences corresponded to the T-DNA right border, which contains the repeated CaMV 35S enhancer, indicates that repetitive sequences should be avoided from primer design since it leads to over representation of this sequence, therefore increasing background noise.

The efficiency of the technique was demonstrated in a forward genetic screen for gyrase inhibition hypersensitivity. In this screen, we isolated 56 plants sensitive specifically to ciprofloxacin and 8 sensitive to both ciprofloxacin and novobiocin. The mutants sensitive specifically to ciprofloxacin might be involved in the maintenance of organelle genome stability, since ciprofloxacin generates DNA double-strand breaks (DSBs) in the organelles. Conversely, novobiocin inhibits organelle gyrases without inducing DSBs, thus the phenotype observed for the 3 lines (8 isolated plants) also sensitive to novobiocin would most likely be related to gyrase inhibition itself. Given the depth of the sequencing, with more than 4000 reads representing the 31 genes identified by the screen, it seems likely that no other insertional mutation would be present in those three lines. Nevertheless, we cannot exclude that the

phenotype observed in those lines could be linked to an insertion that could not be identified by our technique or a DNA rearrangement induced by the agrotransformation process [20].

The approach reported here to identify T-DNA insertion sites on a large scale presents several advantages over other NGS techniques. Because of the sequence capture step that eliminates most unwanted genomic DNA, it is more scalable than the low coverage sequencing approach described by Polko et al. (2012), which is one of the least expensive NGS techniques available. In addition, the 2D-PCR pooling step significantly reduces the cost of our approach as compared to the Illumina barcoded library preparation [9]. Finally, the use of Roche 454 GS-FLX + facilitates the identification of flanking sequences when dealing with variable borders, such as those present in T-DNA cassettes, by giving longer reads and clear hybrid T-DNA/flanking region fragments. On the other hand, one must consider that 2D-PCR pooling is more time-consuming than barcoding, as separate PCR reactions must be carried out to confirm each insertion. Altogether, our technique is well suited to identify mutants isolated from a forward genetic screen that has a limited number of insertions (usually 1 to 5), at a medium throughput. However 2D-PCR becomes too laborious when dealing with numerous insertions, which is the case in most retrotransposon experiments, or in very large mutant collections. Although 3D-PCR pooling [19] and more in depth sequencing could provide a higher throughput, this would still be more laborious than barcoding.

A great advantage of our procedure is that a single 454 library is made from multiple samples, followed by a single capture experiment requiring a single sequencing run that leads to the identification of more than 80% of the insertions sites. The whole process from DNA to sequencing reads takes only 2 weeks. In comparison, classical PCR based techniques needs to be performed on individual samples and requires 2 to 3 nested PCR steps that often lead to multiple PCR bands. This is in addition of the important limitations such as the inefficient ligation step or the need for restriction enzymes that cut both the T-DNA and the genomic part [8]. A hybrid version of classical and NGS approaches has recently been described in *Lotus japonicus* [10]. PCR amplification of flanking regions of a retrotransposon was used to enrich for the junctions between retrotransposons and genomic DNA prior to Illumina sequencing on a single barcoded pool. The PCR amplification efficiency was superior to our sequence capture approach yielding 73% of the reads linked to the retrotransposon insertion sites. This technique

has the advantage of being more scalable while still being affordable. However, the variable length and structure of T-DNA borders inserted in Arabidopsis [3] would greatly complicate the design of primers and be a hindrance for the PCR amplification required in this technique. This problem is solved by the sequence capture step described here. The simplicity and rapidity of our procedure compared to PCR-based methods are its most attractive features.

In conclusion, we show that the efficiency of sequence capture and 2D-PCR pooling can greatly diminish the cost of next-generation sequencing when dealing with many different biological samples. Furthermore, this technique is not restricted to the T-DNA sequence or to Arabidopsis, but is also suitable for the identification of unknown regions flanking any known DNA sequence, in any organism.

6.7 Acknowledgements

We thank Dr. Isobel Parkin for the generous gift of the Arabidopsis SK mutant collection. The sequencing work has been done at the Institut de Biologie Intégrative et des Systèmes, Université Laval (<http://www.ibis.ulaval.ca/sequencage.shtml>).

6.8 References

1. Alonso JM, Ecker JR (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat Rev Genet* 7: 524-536.
2. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, et al. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* 301: 653-657.
3. Robinson SJ, Tang LH, Mooney BA, McKay SJ, Clarke WE, et al. (2009) An archived activation tagged population of Arabidopsis thaliana to facilitate forward genetics approaches. *BMC Plant Biol* 9: 101.
4. Sessions A, Burke E, Presting G, Aux G, McElver J, et al. (2002) A high-throughput Arabidopsis reverse genetics system. *Plant Cell* 14: 2985-2994.

5. Ostergaard L, Yanofsky MF (2004) Establishing gene function by mutagenesis in *Arabidopsis thaliana*. *Plant J* 39: 682-696.
6. Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J* 8: 457-463.
7. Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120: 621-623.
8. Ji J, Braam J (2010) Restriction site extension PCR: a novel method for high-throughput characterization of tagged DNA fragments and genome walking. *PLoS One* 5: e10577.
9. Williams-Carrier R, Stiffler N, Belcher S, Kroeger T, Stern DB, et al. (2010) Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. *Plant J* 63: 167-177.
10. Urbanski DF, Malolepszy A, Stougaard J, Andersen SU (2012) Genome-wide LORE1 retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *Plant J* 69: 731-741.
11. Polko JK, Temanni MR, van Zanten M, van Workum W, Iburg S, et al. (2012) Illumina sequencing technology as a method of identifying T-DNA insertion loci in activation-tagged *Arabidopsis thaliana* plants. *Mol Plant* 5: 948-950.
12. 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: 7821-7826.
13. Rowan BA, Oldenburg DJ, Bendich AJ (2010) RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. *J Exp Bot* 61: 2575-2588.
14. Cappadocia L, Marechal A, Parent JS, Lepage E, Sygusch J, et al. (2010) Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair. *Plant Cell* 22: 1849-1867.
15. Weigel D, Glazebrook J (2002) *Arabidopsis: A Laboratory Manual*. Cold Spring Harbor Lab Press, Cold Spring Harbor, NY.
16. Raheison E, Rigault P, Caron S, Poulin PL, Boyle B, et al. (2012) Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC Genomics* 13: 434.
17. Parent JS, Lepage E, Brisson N (2011) Divergent roles for the two Poll-like organelle DNA polymerases of *Arabidopsis*. *Plant Physiol* 156: 254-262.
18. Collin F, Karkare S, Maxwell A (2011) Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Appl Microbiol Biotechnol* 92: 479-497.

19. Barillot E, Lacroix B, Cohen D (1991) Theoretical analysis of library screening using a N-dimensional pooling strategy. *Nucleic Acids Res* 19: 6241-6247.
20. Nacry P, Camilleri C, Courtial B, Caboche M, Bouchez D (1998) Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics* 149: 641-650.

6.9 Figure Legends

Figure 64. Forward Genetic Screen to Identify Genes Involved in the Maintenance of Organelle Genome Topology. Schematic representation of the different steps of the forward genetic screen. Plants with white first true leaves represent the mutants sensitive to ciprofloxacin (CIP) or novobiocin (NOVO).

Figure 65. Overview of Targeted Genomic Sequencing. Blue rectangles represent genomic DNA, and red rectangles correspond to T-DNA insertions. The grey squares represent the 454 specific primers added in order to bind the sequencing beads (purple circles). The green circles correspond to biotin bound to a red T-DNA specific primer and hybridized to T-DNA. Hybridized sequences are then enriched by capture on streptavidin beads (orange circles).

Figure 66. Coverage of the pSKI015 Vector Obtained by Sequencing. Features of the pSKI015 are summarized below the coverage graph. The blue rectangles represent the T-DNA cassette with the right (RB) and left (LB) borders in green. The position of the 35S enhancers are indicated by blue open end arrows. The red lines represent the annealing regions of the three biotinylated primers for each border. The position where the repeated reads align is indicated by the double red arrowhead line on the coverage graph.

Figure 67. Association of an Insertion Event to a Specific Line by 2D-PCR Pooling. A. Workflow of the 2D-PCR pooling B. An example of the pooling design for 16 plants. Each plant genomic DNA is pooled in a unique set combination. The plants encompass by the colored rectangle associate to the pool of the same color. C. Data analysis to identify the positive line. All bands on a given gel correspond to the same amplification product in different pools.

Figure 68. Schematic Illustration of the Insertion Sites in the Three Novobiocin-Sensitive Mutant Lines. The small black arrows represent the orientation of the CaMV 35S enhancers within the T-DNA (rectangle). For Insertion 3, a different part of the plasmid still containing the enhancer region has been inserted.

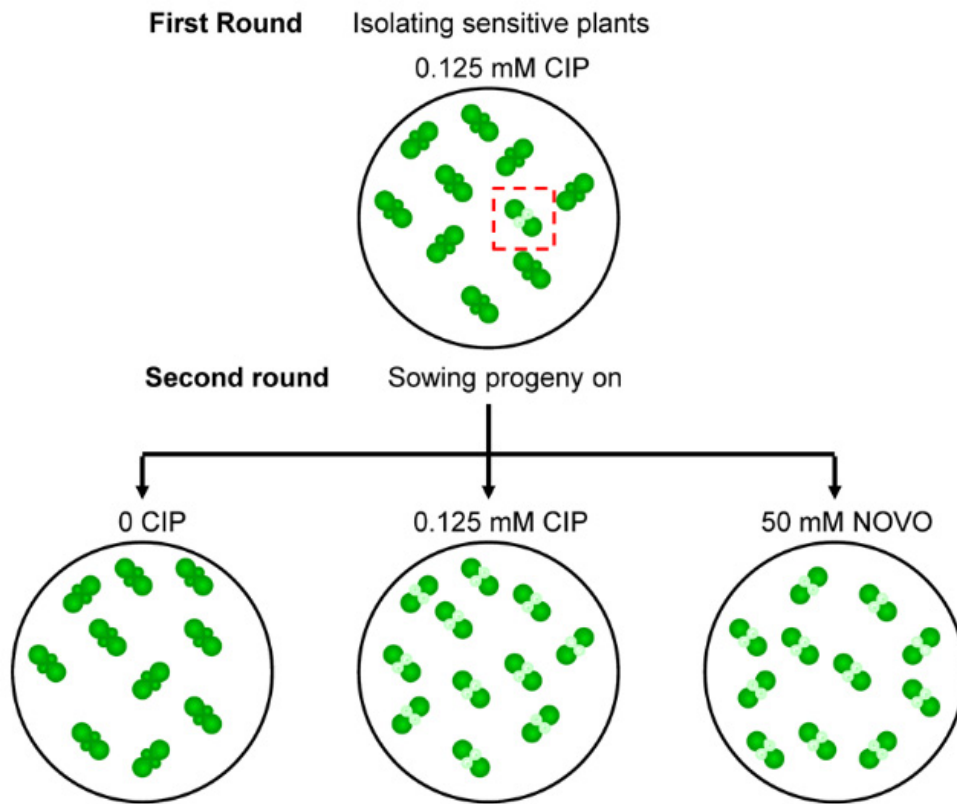


Figure 64. Forward Genetic Screen to Identify Genes Involved in the Maintenance of Organelle Genome Topology. Schematic representation of the different steps of the forward genetic screen. Plants with white first true leaves represent the mutants sensitive to ciprofloxacin (CIP) or novobiocin (NOVO).

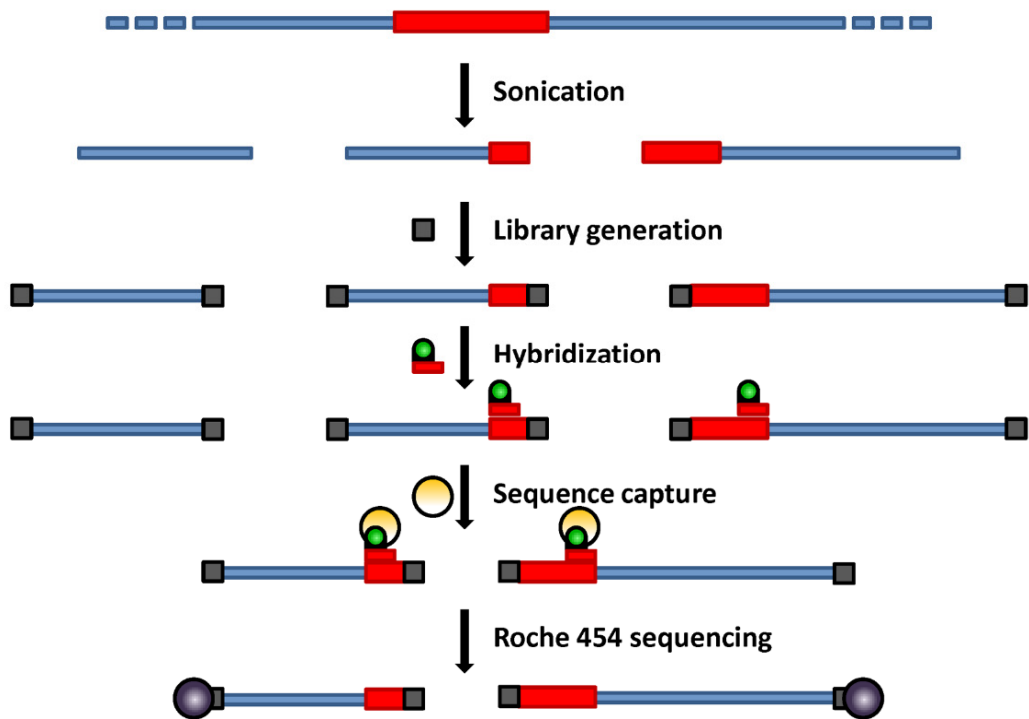


Figure 65. Overview of Targeted Genomic Sequencing. Blue rectangles represent genomic DNA, and red rectangles correspond to T-DNA insertions. The grey squares represent the 454 specific primers added in order to bind the sequencing beads (purple circles). The green circles correspond to biotin bound to a red T-DNA specific primer and hybridized to T-DNA. Hybridized sequences are then enriched by capture on streptavidin beads (orange circles).

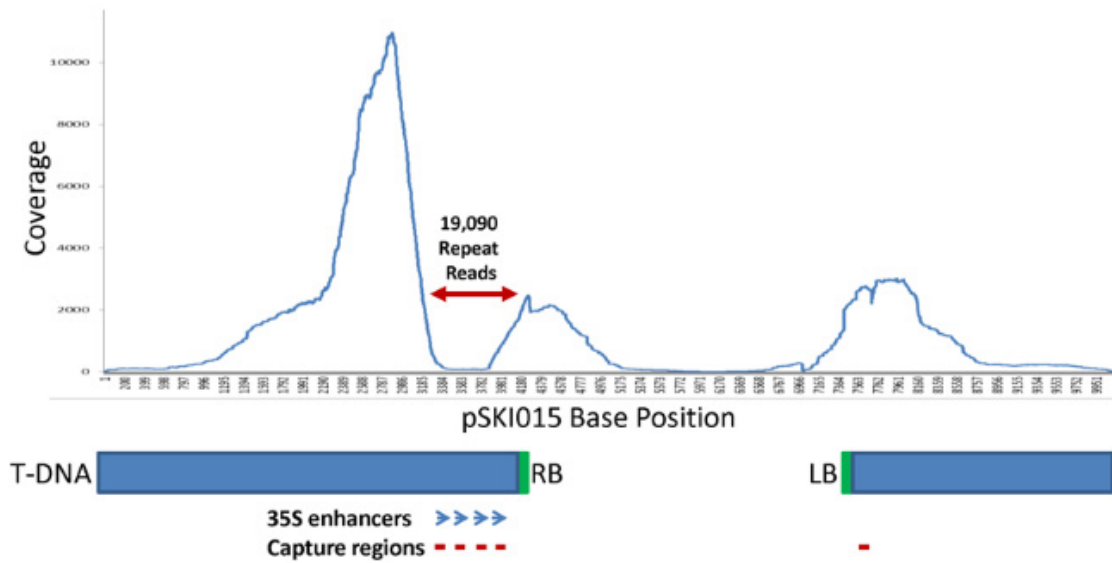


Figure 66. Coverage of the pSKI015 Vector Obtained by Sequencing. Features of the pSKI015 are summarized below the coverage graph. The blue rectangles represent the T-DNA cassette with the right (RB) and left (LB) borders in green. The position of the 35S enhancers are indicated by blue open end arrows. The red lines represent the annealing regions of the three biotinylated primers for each border. The position where the repeated reads align is indicated by the double red arrowhead line on the coverage graph.

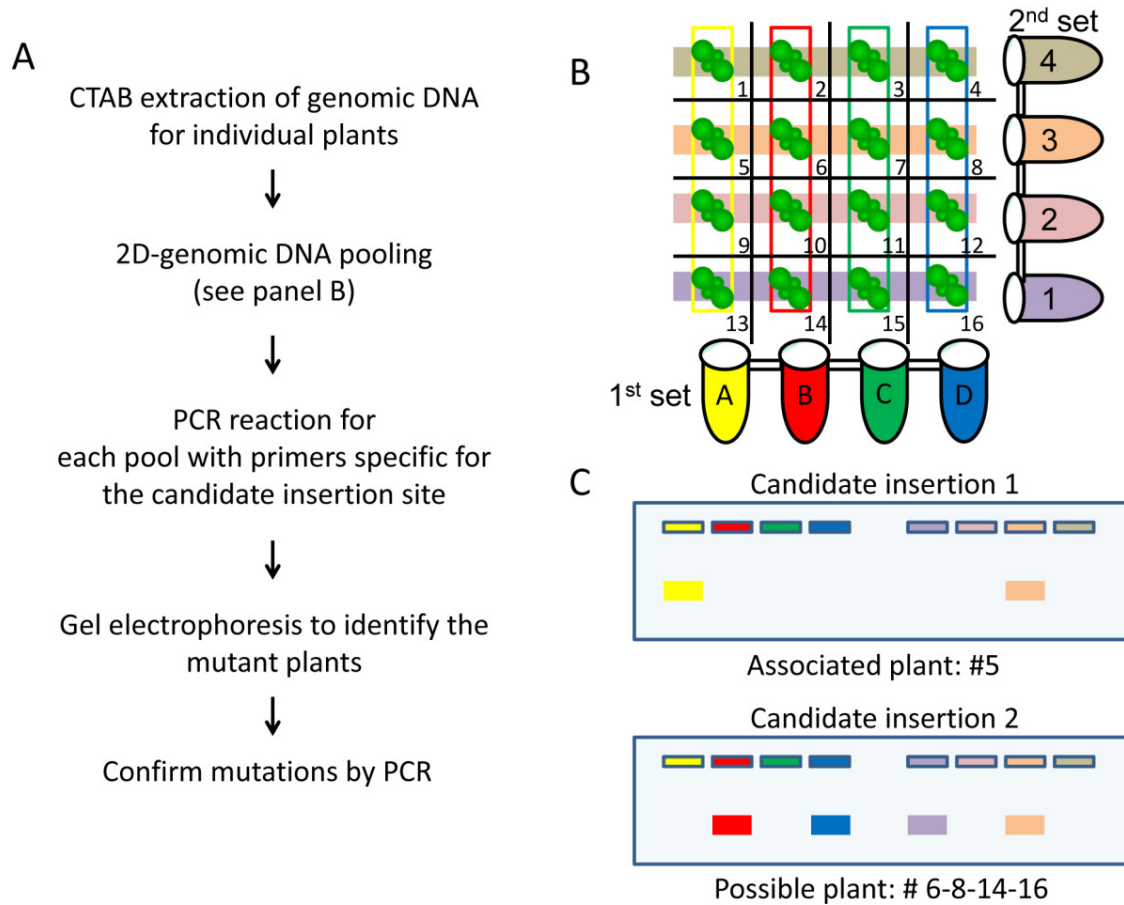


Figure 67. Association of an Insertion Event to a Specific Line by 2D-PCR Pooling. A. Workflow of the 2D-PCR pooling B. An example of the pooling design for 16 plants. Each plant genomic DNA is pooled in a unique set combination. The plants encompass by the colored rectangle associate to the pool of the same color. C. Data analysis to identify the positive line. All bands on a given gel correspond to the same amplification product in different pools.

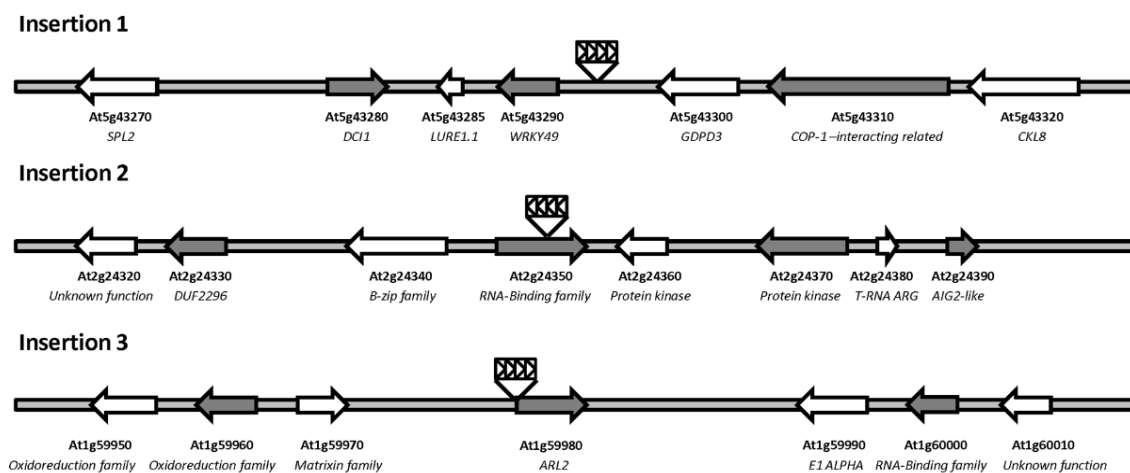


Figure 68. Schematic Illustration of the Insertion Sites in the Three Novobiocin-Sensitive Mutant Lines. The small black arrows represent the orientation of the CaMV 35S enhancers within the T-DNA (rectangle). For Insertion 3, a different part of the plasmid still containing the enhancer region has been inserted.

Chapitre 7 : Discussion

7.1 Survol

Les chapitres précédents illustrent l'importance du maintien de la stabilité génomique pour le fonctionnement du plastide. En effet, les rares défaillances lors de ce processus mènent à la formation de réarrangements génomiques qui peuvent avoir des effets néfastes sur le développement de la plante. Les résultats présentés dans cet ouvrage dressent un portrait beaucoup plus exhaustif des acteurs impliqués dans ces mécanismes à l'intérieur du plastide. Dans un premier temps, le criblage de génétique inverse a permis de discerner le rôle de plusieurs acteurs retrouvés en interaction avec l'ADN dans les processus de réparation des DSBs. La caractérisation moléculaire de quatre familles de gènes, identifiées par ce criblage, est venue poser les bases de certains mécanismes de réparation de l'ADN dans les organelles, particulièrement ceux dépendants de microhomologies. Dans un second temps, les résultats obtenus lors du criblage génétique non-biaisé ont permis de constater l'importance de la stabilité génomique du plastide pour le développement de la plante et la tolérance face aux ROS (Annexe 2). En effet, plus de la moitié des gènes identifiés sont impliqués dans les processus de la photosynthèse ou de la détoxification des ROS. La caractérisation des mutants issus de ce criblage génétique n'a pas encore été effectuée, mais elle devrait permettre de mieux comprendre les effets de l'instabilité génomique dans le plastide ainsi qu'au niveau global chez la plante.

7.2 Les Whirly protègent le génome du plastide face aux réarrangements de types MHMR causés par les DSBs.

Les résultats présentés ici ne constituent pas un premier témoignage suggérant l'importance de la famille de protéines de liaison à l'ADN simple-brin, de type Whirly, dans le maintien de la stabilité génomique du plastide (165). En effet, une étude antérieure a permis de démontrer que le mutant des Whirly, *why1why3*, présente un phénotype de variégation chez 5% de sa population, phénotype causé par l'apparition de réarrangements génomiques dépendants de séquences microhomologues dans le génome plastidique (165). La protéine Whirly 1 du maïs remplit aussi une fonction similaire dans le plastide de cette espèce. Cette conservation du rôle des protéines Whirly entre des représentants

aussi éloignés du règne végétal suggère qu'elle remplit une fonction d'une grande importance, conservée chez les plantes.

Les résultats décrits dans cet ouvrage viennent approfondir notre compréhension du rôle des Whirly dans les mécanismes de maintien de la stabilité génomique. Dans un premier temps, la sensibilité des mutants *why1why3* au traitement à la ciprofloxacine suggère que ces protéines seraient impliquées dans la réparation des DSBs. De plus, les études de PCR et de séquençage nouvelle-génération indiquent que l'apparition de réarrangements dépendants de microhomologies est grandement facilitée dans les lignées mutantes n'exprimant pas les Whirly plastidiques. Les autres types de réarrangements, plus similaires à ceux issus du NHEJ, demeurent à des niveaux pratiquement inchangés dans ces lignées. Ces mêmes études d'amplification PCR révèlent aussi que, suite à un traitement à la ciprofloxacine, les plantes de type sauvage accumulent des réarrangements dépendants de microhomologies similaires à ceux observés constitutivement chez *why1why3*. De plus, ce type de réarrangements a aussi été observé dans les plantes de type sauvage non traitées, par séquençage nouvelle-génération. Ces résultats suggèrent donc que les Whirly protégeraient l'ADN plastidique contre les réarrangements dépendants de microhomologies, particulièrement lorsque des DSBs sont générés dans le génome. Ils indiquent aussi que ce type de réarrangements se produit dans un contexte endogène chez les plantes de type sauvage.

Les Whirly ne sont pas les seules protéines de liaison à l'ADN simple-brin jouant un rôle primordial dans le maintien de la stabilité génomique. Les protéines Single-stranded DNA Binding Proteins (SSB), retrouvées chez les procaryotes, recouvrent les extrémités d'ADN simple-brin générées lors de l'étape de résection nécessaire à la réparation de la plupart des DSBs (80). Par la suite, elles seront enlevées uniquement lorsqu'un filament d'ADN simple-brin, assez long pour permettre la recombinaison homologue, aura été produit. L'inactivation des SSB entraîne d'ailleurs une plus grande sensibilité face aux agents mutagènes qui induisent des DSBs, suggérant que leur absence nuit à la réparation de ce type de lésion (201, 202). Dans le noyau des eucaryotes, la protéine Replication Protein A (RPA) est connue pour jouer un rôle similaire. Chez *S. cerevisiae*, cette protéine jouerait d'ailleurs un rôle essentiel dans la protection du génome face aux réarrangements dépendants de microhomologies. Les auteurs de cette étude ont effectué de nombreux croisements génétiques entre différents mutants pour des gènes impliqués dans la réparation des DSBs, mais la mutation de RPA demeurait toujours une condition essentielle à l'augmentation importante de la fréquence d'établissement du MHMR (111). Le fait que la mutation des Whirly mène aussi à une plus forte sensibilité aux DSBs et à

l'apparition fréquente de molécules réarrangées issues du MHMR suggère que cette famille de protéines posséderait un rôle similaire aux SSBs procaryotes et à RPA.

La caractérisation de la structure des Whirly a aussi permis d'en apprendre davantage sur le mécanisme d'action de la protéine. À l'image des SSBs, les Whirly forment un tétramère et se lient de façon non-spécifique à l'ADN. Lors de cette liaison, les bases de l'ADN sont tournées face à la protéine, limitant la possibilité d'appariement à une autre molécule d'ADN simple-brin. Il semble donc que les Whirly fixent l'ADN simple-brin afin de protéger le génome des effets délétères que peut avoir une telle structure, tels que la recombinaison illégitime avec une autre molécule d'ADN simple-brin. Globalement, nos résultats permettent de formuler un modèle, présenté à la figure 69, décrivant le rôle des Whirly dans la réparation des DSBs. Ce modèle propose que les Whirly se lient aux extrémités d'ADN simple-brin, formées lors de l'étape de résection nécessaire à la réparation de certains DSBs, afin d'empêcher les appariements indésirables du simple-brin et favoriser ainsi la réparation conservatrice. Cependant, en l'absence des Whirly, ou dans des cas de dommages intenses à l'ADN, les extrémités d'ADN peuvent se retrouver libres et mener à des réarrangements génomiques dépendant de courtes séquences d'homologies.

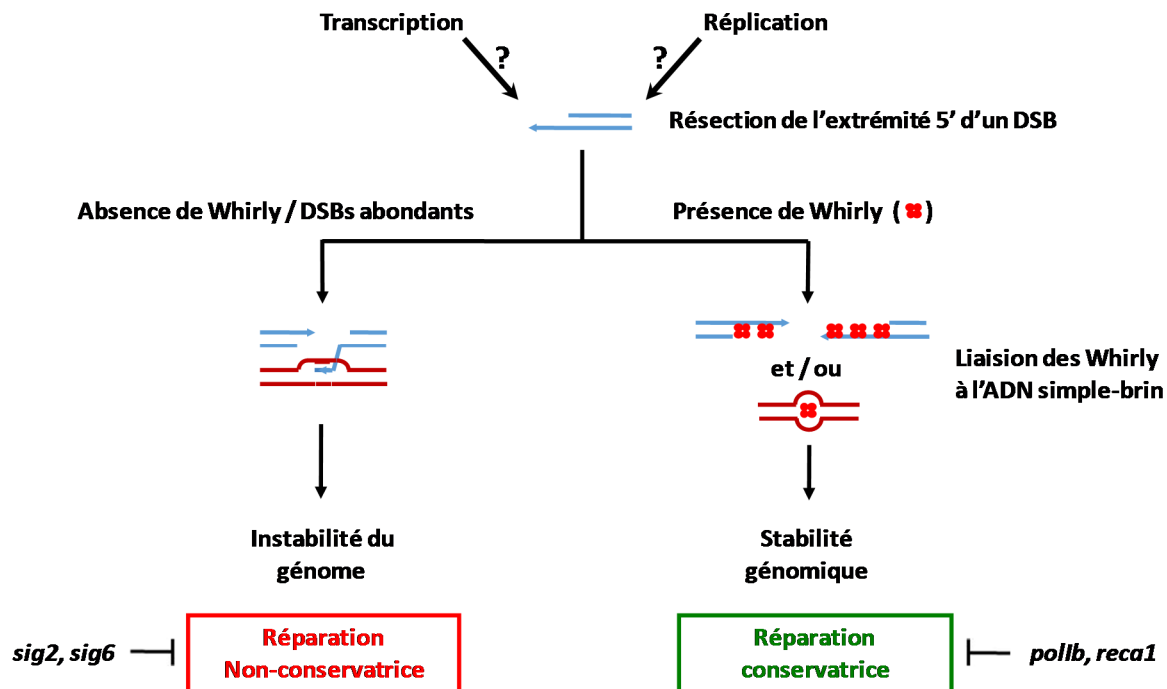


Figure 69: Modèle pour la réparation des DSBs dans le plastide d'Arabidopsis. Différents stress peuvent mener à la formation de DSBs dans le plastide. Les Whirly, représentées par un tétramère de boules rouges, sont des acteurs déterminants dans le choix du type de réparation pour un DSB. Suite à l'étape de résection, les Whirly se lient à l'ADN ayant subi un bris, présenté en bleu, afin de limiter sa capacité recombinogénique. Dans cette conformation, les Whirly guident la réparation vers les mécanismes conservateurs, régulés par différents facteurs, dont font vraisemblablement partie les gènes *POLIB* et *RECA1*. En l'absence de Whirly ou dans une situation de DSBs abondants, les mécanismes de réparation de l'ADN dépendant de microhomologies sont favorisés. Bien que les résultats soient préliminaires, il est possible que les facteurs de transcription *SIG2* et *SIG6* soient essentiels à ces mécanismes. Il est aussi possible que les Whirly se lient à l'ADN simple-brin sur une molécule génomique intacte, illustrée en rouge, afin de limiter les réactions d'invasion à ces sites.

Notre approche expérimentale s'est principalement concentrée sur la réparation des DSBs. Cependant, plusieurs autres processus, tels que la réplication et la transcription, peuvent mener à l'apparition d'ADN simple-brin, connu pour être plus facilement altérable et plus sensible aux modifications, telles les substitutions de bases (revue dans (83)). Il serait donc intéressant d'évaluer plus précisément le rôle des Whirly dans la protection de l'ADN simple-brin en dehors du contexte des DSBs.

7.3 La réplication une source de DSBs dans les organelles ?

Comme mentionnée précédemment, la réplication de l'ADN nécessite des modifications importantes à la structure de l'ADN, qui peuvent en retour faciliter l'apparition de plusieurs types de réarrangements génomiques. Au chapitre 3, nous avons rapporté une augmentation de l'abondance de DSBs chez les mutants *pollb*. Habituellement, lorsqu'il y a augmentation de la fréquence d'apparition de DSBs chez un organisme, le niveau de réarrangement de type NHEJ augmente de manière proportionnelle (101). Étonnamment, les résultats de séquençage de nouvelle-génération

révèlent que ce type de réarrangements n'est pas significativement plus abondant chez *pollb* et chez le triple mutant *why1why3pollb* que chez les plantes de type sauvage. De plus, un stress réplicatif important devrait causer l'apparition de DSBs de manière constitutive chez ces mutants (revue dans (121)). Toutefois, l'analyse par PCR amplifiant de longs fragments, présentée au chapitre 3, révèle des différences dans les niveaux de DSBs uniquement lors du traitement à la ciprofloxacine. Bien que le stress réplicatif soit connu pour favoriser l'apparition de DSBs (revue dans (121)), nos résultats suggèrent plutôt que leur réparation est moins efficace chez *pollb*. En effet, une réparation plus lente des DSBs mènerait à une détection accrue de cassures dans le génome, suite à un traitement induisant des DSBs, sans pour autant être discernable en l'absence de stress, vu la faible utilisation des mécanismes de réparation dans cette situation.

Cette hypothèse, que la réparation des DSBs soit moins efficace chez *pollb*, suggère par le fait même que cette polymérase serait impliquée dans les mécanismes de réparation de l'ADN conservateurs, telle la HR. La littérature regorge d'ailleurs d'exemples de polymérases impliquées dans ces processus qui nécessitent généralement la synthèse d'ADN (revue dans (203)). Un ralentissement de la réparation de l'ADN a déjà été rapporté pour les mutants de certaines polymérases chez *E. coli* (204). Chez les mammifères, la réplication de l'ADN dans la mitochondrie est accomplie par un autre type de polymérase, soit la polymérase Gamma (PolG). Des études de séquençage de nouvelle-génération ont révélé que les mutants exprimant une version moins efficace de cette polymérase accumulaient des réarrangements génomiques dans leur mitochondrie, ainsi que de forts niveaux de modifications de bases (190). Ces modifications ont d'ailleurs été associées à plusieurs maladies héréditaires de façon maternelle chez l'humain, ainsi qu'à un phénotype de vieillissement prématuré (revue dans (205)). Le rôle précis de PolG dans les mécanismes de maintien de la stabilité génomique de la mitochondrie demeure toutefois nébuleux. Néanmoins, la mutation d'une telle polymérase processive dans les organelles est maintenant connue pour avoir un effet déstabilisant sur le génome, et ce à travers différents règnes chez les eucaryotes.

Le maintien de la stabilité du génome plastidique nécessite la coordination de nombreux mécanismes de réparation de l'ADN, aux finalités diverses. À la lumière des résultats présentés dans cette thèse, nous proposons que les mécanismes de réparation de l'ADN dépendants de séquences microhomologues sont en compétition avec les mécanismes conservateurs. Donc, ralentir la réparation conservatrice favoriserait celle non-conservatrice. Une telle interaction a déjà été proposée entre le MMEJ et le NHEJ, suggérant que le MMEJ serait en quelque sorte une roue de secours lorsque

le NHEJ est surchargé ou dysfonctionnel. L'interaction génétique entre *pollB* et les Whirly supporte un tel modèle, suggérant qu'une réparation plus lente des DSBs causée par la mutation de *POLIB* favorise l'apparition de réarrangements dépendants de microhomologies. Il est intéressant de constater que la mutation de *RECA1*, lorsque combinée à celle des Whirly, mène aussi à une augmentation des niveaux de réarrangements génomiques dans le plastide. Bref, on peut supposer que la mutation de *POLIB*, comme celle de *RECA1*, diminue l'efficacité de la réparation conservatrice. Étonnamment, les simples mutants *pollb* et *reca1* n'accumulent pas significativement plus de réarrangements dépendants de microhomologies, suggérant que la mutation des Whirly est le facteur limitant dans l'apparition de ce type de réarrangements.

Les mécanismes de type MHMR nécessitent généralement une étape de synthèse d'ADN afin d'assurer la réparation des DSBs. Chez la levure et l'humain, une polymérase de type-I, pol32, est essentielle à l'établissement de ces mécanismes (112). Chez les organelles de plantes, l'identité de la polymérase qui remplit une telle fonction demeure inconnue. Bien que la sensibilité des mutants *pollb* à la ciprofloxacine aurait pu supporter la candidature de cette polymérase, des duplications dépendantes de microhomologies sont observées chez les mutants *pollb* et *why1why3pollb*, indiquant qu'elle n'est pas essentielle au MHMR. Il serait donc intéressant de tester le rôle des autres polymérases d'ADN présentes dans le plastide lors de ces processus, d'autant plus qu'une autre polymérase, MUS308, a été identifiée par le criblage de génétique inverse utilisant la ciprofloxacine.

7.4 Le stress répliatif et les réarrangements génomiques indépendants des DSBs.

Le stress répliatif, généralement décrit comme le ralentissement ou l'arrêt fréquent de la fourche de répliation, peut lui aussi déstabiliser le génome des organelles (121). En effet, les fourches de répliation bloquées adoptent une structure relativement fragile, connue pour être un substrat idéal pour la formation de plusieurs types de réarrangements génomiques. Afin d'atténuer l'effet négatif de ces fourches de répliation bloquées, certains mécanismes permettent de les redémarrer de façon conservatrice (121). Parmi eux, la régression de fourche est particulièrement efficace et fortement conservée entre les espèces (131). Ce processus permet de redémarrer les fourches ayant un seul des deux brins naissants entravé. Dans une telle situation, un des brins naissants peut être synthétisé en

aval de la lésion. Par la suite, une recombinaison permet aux deux brins naissants de s'appareiller, et le plus long peut alors servir de matrice pour allonger le brin naissant initialement bloqué (voir figure 12, page 28). Bien que ce mécanisme n'ait pas encore été directement observé dans le plastide, les résultats obtenus par séquençage nouvelle-génération suggèrent qu'un mécanisme similaire à la régression de fourche mène à l'apparition d'inversions dans le génome du plastide. En effet, une forte proportion des réarrangements observés dans le plastide est constituée d'inversions, générées à courte distance, qui semblent s'être produites lors de l'appariement erroné d'un brin naissant sur le brin matrice opposé (voir figure 12, page 28). Ces réarrangements de type U-turn-like sont majoritairement produits à l'aide de séquences microhomologues et sont plus abondants en l'absence du gène *RECA1*. Ces caractéristiques ne sont pas sans rappeler celles du MMBIR ou du MMEJ, qui sont tous deux stimulés lorsque leur mécanisme associé (BIR et NHEJ) est inhibé. De plus, à l'image du MMBIR et du MMEJ, la présence de Whirly inhibe fortement l'apparition de ce type de réarrangements. Il est alors tentant de spéculer que les réarrangements de type « U-turn-like » sont issus d'un mécanisme de régression de fourche alternatif, dépendant de microhomologies.

D'autres études ont aussi rapporté l'apparition de molécules réarrangées similaires à celles décrites au chapitre 4 (136-138, 206). L'une d'entre elles propose un modèle (137), nommé « réplication U-turn », qui suggère qu'en cas de stress réplicatif, la polymérase peut bloquer et redémarrer sur le brin matrice opposé. Ce mécanisme produit des isochromosomes, qui sont des chromosomes palindromiques dupliqués et inversés à partir d'une position génomique précise. La fréquence à laquelle se produit le mécanisme de réplication U-turn est aussi influencée par le stress réplicatif, qui semble favoriser le décrochage de la polymérase réplicative (137). De plus, l'apparition de ces réarrangements ne nécessite pas l'intervention des acteurs couramment impliqués dans la réparation des DSBs (135). Bref, les similarités entre le modèle de « réplication U-turn » et le nôtre sont nombreuses, suggérant qu'il pourrait s'agir d'un mécanisme conservé entre le noyau de la levure et les organelles.

Comme mentionnées en introduction, les inversions de locus sont les modifications génomiques les plus fréquemment observées dans le plastide (207, 208). Jusqu'à maintenant aucun mécanisme ne permettait d'expliquer l'apparition de ces fréquentes inversions. Les abondantes molécules inversées de type « U-turn-like » observées par NGS constituent peut-être une avenue intéressante pour expliquer ces fréquentes modifications génomiques. En effet, il a été proposé que les inversions de type « U-turn-like » serviraient de point de départ à une suite d'événements de réarrangements

génomiques pouvant mener à l'inversion d'une partie de génome (135). Bien qu'il ne s'agisse actuellement que d'une hypothèse, il serait intéressant de vérifier si de telles suites d'événements de modifications génomiques se produisent dans le plastide d'*Arabidopsis*.

Une autre caractéristique des réarrangements observés dans les organelles de plantes est qu'ils se produisent majoritairement à courte distance. En effet, près de la moitié de tous les réarrangements génomiques identifiés se produisent avec une distance inférieure à un kilobase entre les locus réarrangés. Le changement de matrice (template switching) est un mécanisme de la famille du MHMR qui se produit lors de stress réplicatifs et qui mène à un déplacement de la fourche de réplication vers une répétition proximale, engendrant une duplication ou une délétion (139, 209, 210). Ces réarrangements génomiques se produisent aussi généralement à courte distance, puisque ce mécanisme favorise le déplacement de la fourche au sein de la bulle de réplication. La prédominance de réarrangements à courte distance dans les organelles suggère donc que le changement de matrice serait actif dans ces compartiments. D'autres mécanismes, comme le MMBIR, sont aussi favorisés à courte distance, mais l'impact de la distance entre les répétitions est beaucoup plus faible. Il est donc probable que les réarrangements à longue distance soient principalement issus du MMBIR, tandis que les délétions et les duplications à plus courte distance soient issues du changement de matrice.

7.5 La transcription et l'instabilité génomique du plastide.

Il est maintenant bien établi que la transcription est une source importante d'instabilité génomique (revue dans (140)). Ce processus peut mener à la formation de DSBs lorsqu'une fourche de transcription entre en collision avec une fourche de réplication. Dans cette optique, il n'est pas surprenant d'observer une diminution des taux de réarrangements génomiques lorsque certains facteurs responsables de la transcription, tels *SIG2* et *SIG6*, sont mutés. Des études réalisées à l'aide de systèmes rapporteurs chez les procaryotes ont d'ailleurs établi que l'intensité de la transcription corrélait directement avec la fréquence des collisions entre ces deux types de fourches (144). Une autre étude évaluant l'impact de la rifampicine a aussi révélé que l'inhibition de la transcription diminuait la fréquence de ces collisions chez *Bacillus subtilis* (211). Bref, on peut supposer que la

mutation de *SIG2* ou de *SIG6*, de même que le traitement à la rifampicine, diminuent le nombre de collisions entre les polymérase dans le plastide, réduisant ainsi les niveaux de DSBs.

Les résultats obtenus grâce aux lignées induisant l'expression de I-CREII sont quant à eux plus difficiles à expliquer. L'inhibition de la transcription par la rifampicine, lors de l'induction de I-CREII, prévient complètement l'apparition de réarrangements génomiques au site de coupure de l'endonucléase. Ce résultat indique que l'impact de la transcription dans l'instabilité génomique ne serait pas uniquement limité à la formation de DSBs. Puisque les réarrangements causés par la coupure de I-CREII proviennent des mécanismes de MHMR, une hypothèse attrayante pour expliquer ce résultat serait que la transcription soit essentielle à l'établissement de la réparation dépendante de microhomologies. En effet, la diminution importante des niveaux de réarrangements génomiques de type MHMR dans les mutants *sig2* et *sig6*, ainsi que lors du traitement à la rifampicine, est très similaire à celle observée chez les mutants de la polymérase responsable de ces mécanismes chez la levure, soit Pol32 (112). Il reste cependant à évaluer comment la transcription serait impliquée dans ce processus. Plusieurs articles publiés au cours de dernières années suggèrent différents rôles pour la transcription dans les mécanismes de réparation, mais aucun ne relie l'activité de la transcription à la réparation dépendante de microhomologies (212).

L'implication directe de la transcription dans les mécanismes de MHMR n'est toutefois pas la seule hypothèse qui pourrait réconcilier tous nos résultats. Une autre hypothèse serait que l'instabilité génomique reliée à la transcription se produit via la formation de boucles-R. Les boucles-R sont produites lorsque l'ARN messager se réapparie à sa matrice d'ADN. Il en résulte le déplacement d'une longue séquence d'ADN simple-brin, connu pour être plus instable que l'ADN double-brin. Le filament d'ADN simple-brin ainsi produit peut aussi faciliter les réactions d'invasions nécessaires lors de la HR et lors des mécanismes de MHMR (revue dans ((140)). L'inhibition de la transcription devrait statistiquement diminuer l'abondance des boucles-R, et donc protéger le génome. Bref, l'évaluation plus détaillée de l'implication des boucles-R dans l'instabilité génomique du plastide devrait permettre de mieux comprendre ces processus. Il est important de mentionner que les résultats obtenus dans cette section sont préliminaires et que des expériences supplémentaires seront nécessaires afin de clarifier l'implication de la transcription dans l'instabilité génomique du plastide. Il serait aussi prématuré de favoriser une hypothèse au détriment d'une autre, considérant que ces hypothèses ne sont pas mutuellement exclusives.

7.6 Un rôle pour les autres gènes identifiés par le criblage de génétique inverse

Les relations épistatiques observées entre les gènes identifiés par le criblage de génétique inverse suggèrent qu'il y a un fort niveau d'interaction entre les différents mécanismes de réparation de l'ADN du plastide. Lorsqu'un mécanisme est inhibé, un autre prend la relève. Les gènes identifiés lors du criblage présentent un niveau impressionnant d'interactions génétiques les uns avec les autres, particulièrement les Whirly qui interagissent avec *POLIB*, *SIG6* et *RECA1*. Certains gènes identifiés par le criblage de génétique inverse n'ont pas encore été caractérisés et il est possible de supposer, à la lumière de résultats préliminaires, que plusieurs interagiront ensemble. Bien que leur caractérisation ne soit que très superficielle pour l'instant, des études réalisées chez des homologues de ces protéines permettent de spéculer sur leur rôle potentiel dans le maintien de la stabilité génomique.

Les protéines UVR sont des candidats particulièrement intéressants puisque les protéines UVRA-B-C limitent l'apparition de molécules d'ADN réarrangés dépendantes de microhomologies chez *E. coli* (213). Cette inhibition des mécanismes de MHMR est observable à la fois en présence ou en l'absence de traitement aux UV, ce qui suggère que le rôle de ces protéines est plus vaste que la seule réparation des dommages causés par les UV. Les auteurs proposent que UVRA protège les extrémités d'ADN, dans un modèle similaire à celui proposé pour les Whirly (213). UVRB possède quant à lui un domaine hélicase qui pourrait servir à défaire les intermédiaires de recombinaison formés par le mésappariement de microhomologies (213). Il serait intéressant de vérifier si les deux protéines UVR identifiées lors du criblage de génétique inverse peuvent jouer des rôles similaires dans le plastide d'*Arabidopsis*.

Deux gènes plus traditionnellement associés à la réparation de mésappariements, faisant partie de la famille MutS, ont aussi été identifiés au cours du criblage. Néanmoins, depuis quelques années, de nombreux articles sont venus confirmer le rôle de cette famille de protéines dans la réparation des DSBs. Des homologues de ces protéines chez l'humain, nommées hMSH4 et hMSH5, agissent dans la recombinaison lors de la méiose et lors de la réparation des DSBs (214). hMSH2 est quant à elle impliquée dans le MMEJ et sa mutation diminue grandement l'efficacité de ce processus (215). Le rôle précis de MSH2 n'a pas encore été élucidé dans ces processus. Par contre, les mutants *hmsH2* tolèrent mieux les appariements imparfaits lors de la recombinaison et sont incapables de corriger les mésappariements qui sont souvent retrouvés aux jonctions des réarrangements génomiques (215,

216). Tous ces résultats permettent de supposer que les protéines de la famille MSH, identifiées par le criblage génétique, sont impliquées directement dans la réparation des DSBs. De plus, comme mentionnée en introduction, la mutation de *MSH1* a déjà été associée à l'apparition de réarrangements génomiques à la fois dans le plastide et la mitochondrie (162, 166).

7.7 L'importance de la stabilité génomique dans le plastide

Les deux derniers chapitres de cet ouvrage ont permis de mieux comprendre l'importance du maintien de la stabilité génomique du plastide. En effet, le chapitre 5 révèle que des niveaux élevés de réarrangements génomiques peuvent mener à une production accrue de ROS par le chloroplaste. Ce déséquilibre du métabolisme des ROS est d'ailleurs associé à de graves conséquences développementales chez la plante. Les résultats obtenus par le criblage de génétique non biaisé, présentés au chapitre 6, confirment aussi le lien entre la stabilité génomique et le métabolisme des ROS. En effet, plus de la moitié des gènes identifiés pour leur sensibilité à la ciprofloxacine sont impliqués de façon directe ou indirecte dans le métabolisme des ROS (Annexe 2). Il est intéressant de mentionner que la mutation de ces gènes mène généralement à une sensibilité supérieure à celle observée pour les mutants des gènes de réparation des DSBs. Bref, déstabiliser le génome du plastide présente une menace réelle pour la plante et cette menace semble être reliée au déséquilibre des niveaux de ROS.

Ces résultats ne sont pas nécessairement surprenants puisque le génome plastidique encode plusieurs composants de la chaîne de transport des électrons, qui est la source principale de ROS dans le chloroplaste (revue dans (52)). Néanmoins, le mécanisme exact par lequel les réarrangements mènent à une augmentation des ROS demeure nébuleux. Est-ce que les réarrangements produisent des gènes chimériques exprimant des protéines non fonctionnelles qui s'intercalent dans la chaîne de transport des électrons? Est-ce que les génomes réarrangés possèdent une stœchiométrie différente pour leurs gènes photosynthétiques, menant à un déséquilibre des composants de la chaîne de transport des électrons? Plusieurs hypothèses peuvent être avancées, mais la question demeure ouverte.

L'incapacité à maintenir la stabilité génomique des organelles a grandement été étudiée chez la mitochondrie humaine et a permis de poser les bases moléculaires de nombreuses maladies (revue

dans (217)). Les phénotypes observés chez certains mutants identifiés lors de nos criblages, tels les troubles de production énergétique et la production accrue de ROS, sont d'ailleurs souvent observés chez les patients atteints de maladies associées à des dysfonctions mitochondriales. Parmi ces maladies, on retrouve la maladie de Leigh (218), le syndrome Pearson et le syndrome de Kearns Sayre (revue dans (219)) qui sont toutes causées par la délétion de sections spécifiques du génome mitochondrial. De plus, l'impact de l'instabilité génomique mitochondriale commence aussi à se clarifier pour certains types de cancer, particulièrement en ce qui concerne la tumorigénèse (195). En effet, celle-ci a été associée à l'apparition de réarrangements génomiques dans l'ADN mitochondrial et à la production de ROS dans la mitochondrie (220). Toutes ces observations illustrent bien l'importance du maintien de la stabilité génomique des organelles et la nécessité d'élargir notre compréhension de ce processus.

Bien que nous ayons pu caractériser plusieurs effets néfastes associés à l'apparition de réarrangements génomiques dans les organelles, il est toutefois surprenant de constater la résilience des organelles face aux altérations de structures de leur génome. En effet, le niveau de réarrangements génomiques observé dans les plantes de type sauvage est tel qu'il suggère que plus de 10% des molécules génomiques possèdent une altération. Pourtant, les dérèglements photosynthétiques et les déséquilibres du métabolisme des ROS sont extrêmement rares chez les plantes de type sauvage. Cette observation signifie que les conséquences de ces réarrangements dans ces plantes sont négligeables. Dans le cas du mutant *why1why3*, les niveaux de réarrangements se situent à environ un réarrangement par génome. Cette augmentation de l'instabilité génomique ne cause toutefois l'apparition d'un phénotype de variabilité que chez seulement 5% de la population.

Par ailleurs, il semble y avoir un effet de seuil dans la tolérance aux réarrangements. En effet, lorsque l'on observe les mutants *why1why3pollb* et *why1why3reca1*, qui ont respectivement des niveaux de réarrangements 3 et 10 fois supérieurs à *why1why3*, on remarque l'apparition de troubles sévères du développement chez l'entièreté de la population. Ces observations suggèrent que, contrairement à ce qui est rapporté dans la littérature, le génome du plastide subit de fréquentes modifications. Alors, pourquoi est-il si conservé à travers les espèces? Nous ne pouvons que supposer qu'il existe une forte pression sélective qui empêche la fixation de ces modifications dans le génome du plastide.

8. Conclusion

Le maintien de la stabilité génomique est un processus essentiel à la transmission d'informations génétiques. Bien qu'un certain niveau d'instabilité soit nécessaire pour permettre l'apparition de nouveaux traits essentiels au processus d'évolution, une trop grande instabilité rendrait le message héréditaire inutilisable. L'objectif primaire des travaux présentés dans cet ouvrage était de caractériser les mécanismes permettant le maintien de la stabilité génomique dans le plastide d'Arabidopsis. D'entrée de jeu, ces travaux ont permis de distinguer deux classes importantes d'instabilité génomique dans le plastide, une dépendante de séquences microhomologues, l'autre formée par la ligation franche de deux fragments non-continus. Ces deux types de réarrangements sont d'ailleurs présents en quantité similaire chez les plantes de type sauvage.

La caractérisation de quatre familles de gènes, identifiés à l'aide d'un criblage de génétique inverse, a contribué à révéler les bases moléculaires de plusieurs mécanismes de réparation des DSBs dans ce compartiment. Parmi ces familles, celle des protéines de liaison à l'ADN simple-brin de type Whirly joue un rôle primordial dans la protection du génome face aux mécanismes de réparation sujets à l'erreur de type MHMR. Nous proposons ici un modèle qui stipule que les protéines Whirly se fixent à l'ADN simple-brin généré lors de la réparation des DSBs et guident le DSB vers une réparation conservatrice. D'autres gènes identifiés par les criblages, tels que *POLIB* et *RECA1*, semblent aussi jouer un rôle dans les mécanismes de réparation conservateurs. La mutation de *POLIB* ou de *RECA1*, lorsque combinée à celle des Whirly, déstabilise encore davantage le génome du plastide, suggérant que le MHMR est en quelque sorte un mécanisme de secours lorsque la réparation conservatrice est entravée. En plus de cette augmentation globale du MHMR, le triple mutant *why1why3reca1* a permis d'identifier un nouveau mécanisme, très actif dans les organelles, et similaire à une régression de fourches, dépendante de microhomologies. Nous proposons que ce mécanisme permet de redémarrer les fourches bloquées, particulièrement lors de stress réplcatifs. Finalement, des résultats préliminaires suggèrent que la transcription aurait un rôle central dans la déstabilisation du génome plastidique, malgré que ce rôle demeure nébuleux. Davantage de travaux seront nécessaires pour venir confirmer le rôle de ce processus, ainsi que le rôle de plusieurs autres gènes identifiés lors de nos criblages, dans le maintien de la stabilité génomique du plastide.

Pour conclure, la découverte de la ciprofloxacine, comme agent induisant des DSBs spécifiquement dans les organelles, fournit un outil précieux pour évaluer les mécanismes de maintien de la stabilité génomique du plastide. De plus, la ciprofloxacine offre un moyen de déstabiliser uniquement le génome des organelles, ce qui est particulièrement intéressant pour des approches plus appliquées de sélection de traits d'intérêt reliés à des mutations dans les organelles. Chez les plantes, la variégation, qui est un phénotype très recherché en horticulture(165), la stérilité mâle cytoplasmique, qui facilite les croisements génétiques en agriculture (93), et la tolérance à certains stress (164), ont toutes été associées à des modifications du génome des organelles. Il est donc envisageable que le traitement à la ciprofloxacine pourrait faciliter l'obtention de cultivars présentant des traits d'intérêts en agriculture.

9. Références

1. Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C. 2011. *The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J Exp Bot* 62: 1775-801
2. Croce R, van Amerongen H. 2014. *Natural strategies for photosynthetic light harvesting. Nat Chem Biol* 10: 492-501
3. Lopez-Juez E, Pyke KA. 2005. *Plastids unleashed: their development and their integration in plant development. Int J Dev Biol* 49: 557-77
4. Martin W, Kowallik KV. 1999. *Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. European Journal of Phycology* 34: 287-95
5. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, et al. 1998. *The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature* 396: 133-40
6. Loening UE, Ingle J. 1967. *Diversity of RNA components in green plant tissues. Nature* 215: 363-7
7. Dyall SD, Brown MT, Johnson PJ. 2004. *Ancient invasions: from endosymbionts to organelles. Science* 304: 253-7
8. McFadden GI. 2014. *Origin and evolution of plastids and photosynthesis in eukaryotes. Cold Spring Harb Perspect Biol* 6: a016105
9. Hsu SC, Belmonte MF, Harada JJ, Inoue K. 2010. *Indispensable Roles of Plastids in Arabidopsis thaliana Embryogenesis. Curr Genomics* 11: 338-49
10. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. 2002. *Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci U S A* 99: 12246-51
11. Kleine T, Maier UG, Leister D. 2009. *DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol* 60: 115-38
12. Green BR. 2011. *Chloroplast genomes of photosynthetic eukaryotes. Plant J* 66: 34-44
13. 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. Current Genetics* 52: 267-74

14. Hallick RB, Lipper C, Richards OC, Rutter WJ. 1976. *Isolation of a transcriptionally active chromosome from chloroplasts of Euglena gracilis*. *Biochemistry* 15: 3039-45
15. Majeran W, Friso G, Asakura Y, Qu X, Huang M, et al. 2012. *Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions*. *Plant Physiol* 158: 156-89
16. Lamppa GK, Bendich AJ. 1979. *Changes in Chloroplast DNA Levels during Development of Pea (Pisum sativum)*. *Plant Physiol* 64: 126-30
17. Leutwiler LS, Houghévans BR, Meyerowitz EM. 1984. *The DNA of Arabidopsis-Thaliana*. *Molecular & General Genetics* 194: 15-23
18. Woloszynska M. 2010. *Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes-though this be madness, yet there's method in't*. *J Exp Bot* 61: 657-71
19. Muller HJ. 1964. *The Relation of Recombination to Mutational Advance*. *Mutat Res* 106: 2-9
20. Khakhlova O, Bock R. 2006. *Elimination of deleterious mutations in plastid genomes by gene conversion*. *Plant J* 46: 85-94
21. Kolodner R, Tewari KK. 1975. *Presence of displacement loops in the covalently closed circular chloroplast deoxyribonucleic acid from higher plants*. *J Biol Chem* 250: 8840-7
22. Bendich AJ. 2004. *Circular chloroplast chromosomes: the grand illusion*. *Plant Cell* 16: 1661-6
23. Oldenburg DJ, Bendich AJ. 2004. *Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms*. *Journal of Molecular Biology* 335: 953-70
24. Christensen AC, Lyznik A, Mohammed S, Elowsky CG, Elo A, et al. 2005. *Dual-domain, dual-targeting organellar protein presequences in Arabidopsis can use non-AUG start codons*. *Plant Cell* 17: 2805-16
25. Kimura S, Uchiyama Y, Kasai N, Namekawa S, Saotome A, et al. 2002. *A novel DNA polymerase homologous to Escherichia coli DNA polymerase I from a higher plant, rice (Oryza sativa L.)*. *Nucleic Acids Res* 30: 1585-92
26. Mori Y, Kimura S, Saotome A, Kasai N, Sakaguchi N, et al. 2005. *Plastid DNA polymerases from higher plants, Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* 334: 43-50
27. Ono Y, Sakai A, Takechi K, Takio S, Takusagawa M, Takano H. 2007. *NtPoll-like1 and NtPoll-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 and Cell Physiology* 48: 1679-92

28. Gubaev A, Klostermeier D. 2014. *The mechanism of negative DNA supercoiling: A cascade of DNA-induced conformational changes prepares gyrase for strand passage (Reprinted from DNA Repair, VOL 16C, PG 23-34, 2014). DNA Repair (Amst) 20: 130-41*
29. Cho HS, Lee SS, Kim KD, Hwang I, Lim JS, et al. 2004. *DNA gyrase is involved in chloroplast nucleoid partitioning. Plant Cell 16: 2665-82*
30. 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: 7821-6*
31. Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K, Borner T. 2008. *Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. Nucleic Acids Res 36: 785-92*
32. Serino G, Maliga P. 1998. *RNA polymerase subunits encoded by the plastid rpo genes are not shared with the nucleus-encoded plastid enzyme. Plant Physiol 117: 1165-70*
33. Mullet JE. 1993. *Dynamic Regulation of Chloroplast Transcription. Plant Physiol 103: 309-13*
34. Hanaoka M, Kanamaru K, Fujiwara M, Takahashi H, Tanaka K. 2005. *Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis. Embo Reports 6: 545-50*
35. Kremnev D, Strand A. 2014. *Plastid encoded RNA polymerase activity and expression of photosynthesis genes required for embryo and seed development in Arabidopsis. Front Plant Sci 5: 385*
36. Hricova A, Quesada V, Micol JL. 2006. *The SCABRA3 nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in Arabidopsis. Plant Physiol 141: 942-56*
37. Chi W, He B, Mao J, Jiang J, Zhang L. 2015. *Plastid sigma factors: Their individual functions and regulation in transcription. Biochim Biophys Acta*
38. Kanamaru K, Nagashima A, Fujiwara M, Shimada H, Shirano Y, et al. 2001. *An Arabidopsis sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. Plant and Cell Physiology 42: 1034-43*
39. Ishizaki Y, Tsunoyama Y, Hatano K, Ando K, Kato K, et al. 2005. *A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. Plant J 42: 133-44*
40. Loschelder H, Schweer J, Link B, Link G. 2006. *Dual temporal role of plastid sigma factor 6 in Arabidopsis development. Plant Physiol 142: 642-50*
41. Nagashima A, Hanaoka M, Shikanai T, Fujiwara M, Kanamaru K, et al. 2004. *The multiple-stress responsive plastid sigma factor, SIG5, directs activation of the psbD blue light-responsive promoter (BLRP) in Arabidopsis thaliana. Plant and Cell Physiology 45: 357-68*

42. Germain A, Hotto AM, Barkan A, Stern DB. 2013. *RNA processing and decay in plastids*. *Wiley Interdiscip Rev RNA* 4: 295-316
43. Hotto AM, Schmitz RJ, Fei Z, Ecker JR, Stern DB. 2011. *Unexpected Diversity of Chloroplast Noncoding RNAs as Revealed by Deep Sequencing of the Arabidopsis Transcriptome*. *G3 (Bethesda)* 1: 559-70
44. Hotto AM, Castandet B, Gilet L, Higdon A, Condon C, Stern DB. 2015. *Arabidopsis Chloroplast Mini-Ribonuclease III Participates in rRNA Maturation and Intron Recycling*. *Plant Cell* 27: 724-40
45. Harmston N, Baresic A, Lenhard B. 2013. *The mystery of extreme non-coding conservation*. *Philos Trans R Soc Lond B Biol Sci* 368: 20130021
46. Huang H, Shi C, Liu Y, Mao SY, Gao LZ. 2014. *Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships*. *BMC Evol Biol* 14: 151
47. Huerta-Cepas J, Bueno A, Dopazo J, Gabaldon T. 2008. *PhylomeDB: a database for genome-wide collections of gene phylogenies*. *Nucleic Acids Res* 36: D491-6
48. 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: 9054-8
49. Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, et al. 2010. *Localized hypermutation and associated gene losses in legume chloroplast genomes*. *Genome Research* 20: 1700-10
50. Drouin G, Daoud H, Xia J. 2008. *Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants*. *Molecular Phylogenetics and Evolution* 49: 827-31
51. Marechal A, Brisson N. 2010. *Recombination and the maintenance of plant organelle genome stability*. *New Phytologist* 186: 299-317
52. Asada K. 2006. *Production and scavenging of reactive oxygen species in chloroplasts and their functions*. *Plant Physiol* 141: 391-6
53. Kuroda H, Maliga P. 2003. *The plastid clpP1 protease gene is essential for plant development*. *Nature* 425: 86-9
54. Sun X, Fu T, Chen N, Guo J, Ma J, et al. 2010. *The stromal chloroplast Deg7 protease participates in the repair of photosystem II after photoinhibition in Arabidopsis*. *Plant Physiol* 152: 1263-73
55. Mittler R. 2002. *Oxidative stress, antioxidants and stress tolerance*. *Trends Plant Sci* 7: 405-10

56. Laloi C, Stachowiak M, Pers-Kamczyc E, Warzych E, Murgia I, Apel K. 2007. *Cross-talk between singlet oxygen- and hydrogen peroxide-dependent signaling of stress responses in Arabidopsis thaliana. Proc Natl Acad Sci U S A* 104: 672-7
57. Maruta T, Noshi M, Tanouchi A, Tamoi M, Yabuta Y, et al. 2012. *H₂O₂-triggered Retrograde Signaling from Chloroplasts to Nucleus Plays Specific Role in Response to Stress. Journal of Biological Chemistry* 287: 11717-29
58. Pfannschmidt T, Nilsson A, Allen JF. 1999. *Photosynthetic control of chloroplast gene expression. Nature* 397: 625-8
59. Asada K. 1999. *The water-water cycle in chloroplasts: Scavenging of active oxygens and dissipation of excess photons. Annual Review of Plant Physiology and Plant Molecular Biology* 50: 601-39
60. Hatahet Z, Zhou MX, Reha-Krantz LJ, Morriscal SW, Wallace SS. 1998. *In search of a mutational hotspot. Proc Natl Acad Sci U S A* 95: 8556-61
61. Decker-Walters DS, Chung SM, Staub JE. 2004. *Plastid sequence evolution: A new pattern of nucleotide substitutions in the Cucurbitaceae. Journal of Molecular Evolution* 58: 606-14
62. Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. *Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci U S A* 101: 17741-6
63. Bensen RJ, Warner HR. 1987. *Partial Purification and Characterization of Uracil-DNA Glycosylase Activity from Chloroplasts of Zea mays Seedlings. Plant Physiol* 84: 1102-6
64. Gutman BL, Niyogi KK. 2009. *Evidence for base excision repair of oxidative DNA damage in chloroplasts of Arabidopsis thaliana. J Biol Chem* 284: 17006-12
65. Tuteja N, Ahmad P, Panda BB, Tuteja R. 2009. *Genotoxic stress in plants: Shedding light on DNA damage, repair and DNA repair helicases. Mutation Research-Reviews in Mutation Research* 681: 134-49
66. Palomera-Sanchez Z, Zurita M. 2011. *Open, repair and close again: Chromatin dynamics and the response to UV-induced DNA damage. DNA Repair (Amst)* 10: 119-25
67. Takahashi M, Teranishi M, Ishida H, Kawasaki J, Takeuchi A, et al. 2011. *Cyclobutane pyrimidine dimer (CPD) photolyase repairs ultraviolet-B-induced CPDs in rice chloroplast and mitochondrial DNA. Plant Journal* 66: 433-42
68. Kaiser G, Kleiner O, Beisswenger C, Batschauer A. 2009. *Increased DNA repair in Arabidopsis plants overexpressing CPD photolyase. Planta* 230: 505-15
69. Landry LG, Stapleton AE, Lim J, Hoffman P, Hays JB, et al. 1997. *An Arabidopsis photolyase mutant is hypersensitive to ultraviolet-B radiation. Proc Natl Acad Sci U S A* 94: 328-32

70. Malkova A, Haber JE. 2012. *Mutations Arising During Repair of Chromosome Breaks. Annual Review of Genetics, Vol 46* 46: 455-73
71. Sonoda E, Hohegger H, Saberi A, Taniguchi Y, Takeda S. 2006. *Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. DNA Repair (Amst)* 5: 1021-9
72. Keros T, Borovecki F, Jemersic L, Konjevic D, Roic B, Balatinec J. 2010. *The Centenary Progress of Molecular Genetics. A 100(th) Anniversary of T. H. Morgan's Discoveries. Collegium Antropologicum* 34: 1167-74
73. Holliday R. 1961. *Induced Mitotic Crossing-over in Ustilago Maydis. Genetical Research* 2: 231-&
74. Holliday R. 1965. *Induced Mitotic Crossing-over in Relation to Genetic Replication in Synchronously Dividing Cells of Ustilago Maydis. Genetics Research* 6: 104-+
75. Bridges BA. 1971. *RecA⁺ -Dependent Repair of Gamma-Ray Damage to Escherichia-Coli Does Not Require Recombination between Existing Homologous Chromosomes. Journal of Bacteriology* 108: 944-&
76. Rupp WD, Howard-Flanders P. 2005. *Discontinuities in the DNA synthesized in an excision-defective strain of Escherichia coli following ultraviolet irradiation. DNA Repair (Amst)* 4: 620-34
77. Green MH, Bridges BA, Riazuddin S. 1971. *Effect of gamma-radiation on the donor ability of recA and recA⁺ strains of Escherichia coli. Journal of General Microbiology* 67: 63-8
78. Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. 2007. *Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet* 8: 762-75
79. Bressan DA, Baxter BK, Petrini JH. 1999. *The Mre11-Rad50-Xrs2 protein complex facilitates homologous recombination-based double-strand break repair in Saccharomyces cerevisiae. Mol Cell Biol* 19: 7681-7
80. Kowalczykowski SC, Krupp RA. 1987. *Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. Journal of Molecular Biology* 193: 97-113
81. Symington LS, Kolodner R. 1985. *Partial-Purification of an Enzyme from Saccharomyces-Cerevisiae That Cleaves Holliday Junctions. Proc Natl Acad Sci U S A* 82: 7247-51
82. Paques F, Leung WY, Haber JE. 1998. *Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol Cell Biol* 18: 2045-54

83. Caldecott KW. 2008. *Single-strand break repair and genetic disease. Nature Reviews Genetics* 9: 619-31
84. 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: e1000327
85. Palmer JD. 1983. *Chloroplast DNA Exists in 2 Orientations. Nature* 301: 92-3
86. Guo WH, Grewe F, Cobo-Clark A, Fan WS, Duan ZL, et al. 2014. *Predominant and Substoichiometric Isomers of the Plastid Genome Coexist within Juniperus Plants and Have Shifted Multiple Times during Cupressophyte Evolution. Genome Biology and Evolution* 6: 580-90
87. Cerutti H, Johnson AM, Boynton JE, Gillham NW. 1995. *Inhibition of Chloroplast DNA Recombination and Repair by Dominant-Negative Mutants of Escherichia-Coli RecA. Mol Cell Biol* 15: 3003-11
88. Miller-Messmer M, Kuhn K, Bichara M, Le Ret M, Imbault P, Gualberto JM. 2012. *RecA-Dependent DNA Repair Results in Increased Heteroplasmy of the Arabidopsis Mitochondrial Genome. Plant Physiol* 159: 211-26
89. Rowan BA, Oldenburg DJ, Bendich AJ. 2010. *RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. J Exp Bot* 61: 2575-88
90. Durrenberger F, Thompson AJ, Herrin DL, Rochaix JD. 1996. *Double strand break-induced recombination in Chlamydomonas reinhardtii chloroplasts. Nucleic Acids Res* 24: 3323-31
91. 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: 842-53
92. Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. 2007. *Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. Plant Cell* 19: 1251-64
93. Luo DP, Xu H, Liu ZL, Guo JX, Li HY, et al. 2013. *A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nature Genetics* 45: 573-U157
94. 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: 1377-90
95. Qi JJ, Liu X, Shen D, Miao H, Xie BY, et al. 2013. *A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nature Genetics* 45: 1510-U149
96. Liu PF, Lacia M, Zhang F, Withers M, Hastings PJ, Lupski JR. 2011. *Frequency of Nonallelic Homologous Recombination Is Correlated with Length of Homology: Evidence that Ectopic Synapsis Precedes Ectopic Crossing-Over. American Journal of Human Genetics* 89: 580-8

97. Chen FQ, Peterson SR, Story MD, Chen DJ. 1996. *Disruption of DNA-PK in Ku80 mutant xrs-6 and the implications in DNA double-strand break repair. Mutation Research-DNA Repair* 362: 9-19
98. McElhinny SAN, Snowden CM, McCarville J, Ramsden DA. 2000. *Ku recruits the XRCC4-ligase IV complex to DNA ends. Mol Cell Biol* 20: 2996-3003
99. Calsou P, Delteil C, Frit P, Droulet J, Salles B. 2003. *Coordinated assembly of Ku and p460 subunits of the DNA-dependent protein kinase on DNA ends is necessary for XRCC4-ligase IV recruitment. Journal of Molecular Biology* 326: 93-103
100. Betermier M, Bertrand P, Lopez BS. 2014. *Is non-homologous end-joining really an inherently error-prone process? PLoS Genet* 10: e1004086
101. Rothkamm K, Kuhne M, Jeggo PA, Lohrich M. 2001. *Radiation-induced genomic rearrangements formed by nonhomologous end-joining of DNA double-strand breaks. Cancer Research* 61: 3886-93
102. Boulton SJ, Jackson SP. 1998. *Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. Embo Journal* 17: 1819-28
103. Roth DB, Wilson JH. 1986. *Nonhomologous Recombination in Mammalian-Cells - Role for Short Sequence Homologies in the Joining Reaction. Mol Cell Biol* 6: 4295-304
104. Ottaviani D, LeCain M, Sheer D. 2014. *The role of microhomology in genomic structural variation. Trends in Genetics* 30: 85-94
105. Ma JL, Kim EM, Haber JE, Lee SE. 2003. *Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Mol Cell Biol* 23: 8820-8
106. Paull TT, Gellert M. 2000. *A mechanistic basis for Mre11-directed DNA joining at microhomologies. Proc Natl Acad Sci U S A* 97: 6409-14
107. McVey M, Lee SE. 2008. *MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends in Genetics* 24: 529-38
108. Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF, Lieber MR. 2007. *XRCC4 : DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps (vol 26, pg 1010, 2007). Embo Journal* 26: 3506-7
109. Lin FL, Sperle K, Sternberg N. 1984. *Model for Homologous Recombination during Transfer of DNA into Mouse L-Cells - Role for DNA Ends in the Recombination Process. Mol Cell Biol* 4: 1020-34

110. Hicks WM. 2010. *Increased mutagenesis and unique mutation signature associated with mitotic gene conversion (vol 329, pg 82, 2010). Science* 330: 448-
111. Deng SK, Gibb B, de Almeida MJ, Greene EC, Symington LS. 2014. *RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. Nature Structural & Molecular Biology* 21: 405-U152
112. Villarreal DD, Lee K, Deem A, Shim EY, Malkova A, Lee SE. 2012. *Microhomology Directs Diverse DNA Break Repair Pathways and Chromosomal Translocations. PLoS Genet* 8
113. Katsura Y, Sasaki S, Sato M, Yamaoka K, Suzukawa K, et al. 2007. *Involvement of Ku80 in microhomology-mediated end joining for DNA double-strand breaks in vivo. DNA Repair (Amst)* 6: 639-48
114. Pingoud A, Wilson GG, Wende W. 2014. *Type II restriction endonucleases-a historical perspective and more. Nucleic Acids Res* 42: 7489-527
115. Muscarella DE, Ellison EL, Ruoff BM, Vogt VM. 1990. *Characterization of I-Ppo, an Intron-Encoded Endonuclease That Mediates Homing of a Group-I Intron in the Ribosomal DNA of Physarum-Polycephalum. Mol Cell Biol* 10: 3386-96
116. Seligman LM, Stephens KM, Savage JH, Monnat RJ, Jr. 1997. *Genetic analysis of the Chlamydomonas reinhardtii I-CreI mobile intron homing system in Escherichia coli. Genetics* 147: 1653-64
117. Thompson AJ, Yuan X, Kudlicki W, Herrin DL. 1992. *Cleavage and recognition pattern of a double-strand-specific endonuclease (I-creI) encoded by the chloroplast 23S rRNA intron of Chlamydomonas reinhardtii. Gene* 119: 247-51
118. Durrenberger F, Rochaix JD. 1993. *Characterization of the cleavage site and the recognition sequence of the I-CreI DNA endonuclease encoded by the chloroplast ribosomal intron of Chlamydomonas reinhardtii. Mol Gen Genet* 236: 409-14
119. Durrenberger F, Thompson AJ, Herrin DL, Rochaix JD. 1996. *Double strand break-induced recombination in Chlamydomonas reinhardtii chloroplasts. Nucleic Acids Res* 24: 3323-31
120. Kwon T, Huq E, Herrin DL. 2010. *Microhomology-mediated and nonhomologous repair of a double-strand break in the chloroplast genome of Arabidopsis. Proc Natl Acad Sci U S A* 107: 13954-9
121. Zeman MK, Cimprich KA. 2014. *Causes and consequences of replication stress. Nat Cell Biol* 16: 2-9
122. Hinz JM, Urbin SS, Thompson LH. 2009. *RAD51D- and FANCG-dependent base substitution mutagenesis at the ATP1A1 locus in mammalian cells. Mutat Res* 665: 61-6

123. Langerak P, Mejia-Ramirez E, Limbo O, Russell P. 2011. *Release of Ku and MRN from DNA ends by Mre11 nuclease activity and Ctp1 is required for homologous recombination repair of double-strand breaks.* *PLoS Genet* 7: e1002271
124. Kondratova A, Watanabe T, Marotta M, Cannon M, Segall AM, et al. 2015. *Replication fork integrity and intra-S phase checkpoint suppress gene amplification.* *Nucleic Acids Res* 43: 2678-90
125. Tolentino JH, Burke TJ, Mukhopadhyay S, McGregor WG, Basu AK. 2008. *Inhibition of DNA replication fork progression and mutagenic potential of 1, N6-ethenoadenine and 8-oxoguanine in human cell extracts.* *Nucleic Acids Res* 36: 1300-8
126. Rudolph CJ, Upton AL, Lloyd RG. 2007. *Replication fork stalling and cell cycle arrest in UV-irradiated Escherichia coli.* *Genes Dev* 21: 668-81
127. Indiani C, Patel M, Goodman MF, O'Donnell ME. 2013. *RecA acts as a switch to regulate polymerase occupancy in a moving replication fork.* *Proc Natl Acad Sci U S A* 110: 5410-5
128. Ziv O, Zeisel A, Mirlas-Neisberg N, Swain U, Nevo R, et al. 2014. *Identification of novel DNA-damage tolerance genes reveals regulation of translesion DNA synthesis by nucleophosmin.* *Nat Commun* 5: 5437
129. Lin YC, Li L, Makarova AV, Burgers PM, Stone MP, Lloyd RS. 2014. *Error-prone replication bypass of the primary aflatoxin B1 DNA adduct, AFB1-N7-Gua.* *J Biol Chem* 289: 18497-506
130. Yuan F, Xu Z, Yang M, Wei Q, Zhang Y, et al. 2013. *Overexpressed DNA polymerase iota regulated by JNK/c-Jun contributes to hypermutagenesis in bladder cancer.* *PLoS One* 8: e69317
131. Robu ME, Inman RB, Cox MM. 2001. *RecA protein promotes the regression of stalled replication forks in vitro.* *Proc Natl Acad Sci U S A* 98: 8211-8
132. Seigneur M, Ehrlich SD, Michel B. 2000. *RuvABC-dependent double-strand breaks in dnaBts mutants require recA.* *Mol Microbiol* 38: 565-74
133. Costes A, Lambert SA. 2012. *Homologous recombination as a replication fork escort: fork-protection and recovery.* *Biomolecules* 3: 39-71
134. Branzei D, Foiani M. 2010. *Maintaining genome stability at the replication fork.* *Nat Rev Mol Cell Biol* 11: 208-19
135. Paek AL, Kaochar S, Jones H, Elezaby A, Shanks L, Weinert T. 2009. *Fusion of nearby inverted repeats by a replication-based mechanism leads to formation of dicentric and acentric chromosomes that cause genome instability in budding yeast.* *Genes Dev* 23: 2861-75

136. Mizuno K, Lambert S, Baldacci G, Murray JM, Carr AM. 2009. *Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. Genes Dev* 23: 2876-86
137. Mizuno K, Miyabe I, Schalbetter SA, Carr AM, Murray JM. 2013. *Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. Nature* 493: 246-9
138. Kugelberg E, Kofoed E, Andersson DI, Lu Y, Mellor J, et al. 2010. *The tandem inversion duplication in Salmonella enterica: selection drives unstable precursors to final mutation types. Genetics* 185: 65-80
139. Seier T, Padgett DR, Zilberberg G, Sutera VA, Jr., Toha N, Lovett ST. 2011. *Insights into mutagenesis using Escherichia coli chromosomal lacZ strains that enable detection of a wide spectrum of mutational events. Genetics* 188: 247-62
140. Aguilera A. 2002. *The connection between transcription and genomic instability. EMBO J* 21: 195-201
141. Helmrich A, Ballarino M, Nudler E, Tora L. 2013. *Transcription-replication encounters, consequences and genomic instability. Nature Structural & Molecular Biology* 20: 412-8
142. Takeuchi Y, Horiuchi T, Kobayashi T. 2003. *Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev* 17: 1497-506
143. Brewer BJ. 1988. *When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. Cell* 53: 679-86
144. Mirkin EV, Mirkin SM. 2005. *Mechanisms of transcription-replication collisions in bacteria. Mol Cell Biol* 25: 888-95
145. Rudolph CJ, Upton AL, Stockum A, Nieduszynski CA, Lloyd RG. 2013. *Avoiding chromosome pathology when replication forks collide. Nature* 500: 608-11
146. Thomas M, White RL, Davis RW. 1976. *Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc Natl Acad Sci U S A* 73: 2294-8
147. Drolet M, Phoenix P, Menzel R, Masse E, Liu LF, Crouch RJ. 1995. *Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proc Natl Acad Sci U S A* 92: 3526-30
148. Chan K, Sterling JF, Roberts SA, Bhagwat AS, Resnick MA, Gordenin DA. 2012. *Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. PLoS Genet* 8: e1003149
149. Ashton NW, Bolderson E, Cubeddu L, O'Byrne KJ, Richard DJ. 2013. *Human single-stranded DNA binding proteins are essential for maintaining genomic stability. BMC Mol Biol* 14: 9

150. Hoffman EA, McCulley A, Haarer B, Arnak R, Feng W. 2015. *Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. Genome Res* 25: 402-12
151. Beletskii A, Bhagwat AS. 1996. *Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. Proc Natl Acad Sci U S A* 93: 13919-24
152. Huertas P, Aguilera A. 2003. *Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell* 12: 711-21
153. 2000. *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature* 408: 796-815
154. Hoffmann GR. 1980. *Genetic effects of dimethyl sulfate, diethyl sulfate, and related compounds. Mutat Res* 75: 63-129
155. 1985. *A simple and general method for transferring genes into plants. Science* 227: 1229-31
156. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, et al. 2003. *Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science* 301: 653-7
157. Sessions A, Burke E, Presting G, Aux G, McElver J, et al. 2002. *A high-throughput Arabidopsis reverse genetics system. Plant Cell* 14: 2985-94
158. Jeon H, Jin YM, Choi MH, Lee H, Kim M. 2013. *Chloroplast-targeted bacterial RecA proteins confer tolerance to chloroplast DNA damage by methyl viologen or UV-C radiation in tobacco (Nicotiana tabacum) plants. Physiol Plant* 147: 218-33
159. Redei GP. 1963. *Somatic Instability Caused by a Cysteine-Sensitive Gene in Arabidopsis. Science* 139: 767-9
160. Redei GP, Plurad SB. 1973. *Hereditary Structural Alterations of Plastids Induced by a Nuclear Mutator Gene in Arabidopsis. Protoplasma* 77: 361-80
161. Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA. 2003. *Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. Proc Natl Acad Sci U S A* 100: 5968-73
162. 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: 889-99
163. Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, et al. 2011. *Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC Biol* 9: 64

164. Shedge V, Davila J, Arrieta-Montiel MP, Mohammed S, Mackenzie SA. 2010. *Extensive rearrangement of the Arabidopsis mitochondrial genome elicits cellular conditions for thermotolerance. Plant Physiol* 152: 1960-70
165. Marechal A, Parent JS, Veronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. 2009. *Whirly proteins maintain plastid genome stability in Arabidopsis. Proc Natl Acad Sci U S A* 106: 14693-8
166. Xu YZ, Arrieta-Montiel MP, Viridi KS, de Paula WB, Widhalm JR, et al. 2011. *MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. Plant Cell* 23: 3428-41
167. Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmüller R. 2006. *pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. Plant Cell* 18: 176-97
168. Desveaux D, Marechal A, Brisson N. 2005. *Whirly transcription factors: defense gene regulation and beyond. Trends Plant Sci* 10: 95-102
169. Cappadocia L, Marechal A, Parent JS, Lepage E, Sygusch J, Brisson N. 2010. *Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. Plant Cell* 22: 1849-67
170. Povirk LF. 1996. *DNA damage and mutagenesis by radiomimetic DNA-cleaving agents: bleomycin, neocarzinostatin and other enediynes. Mutat Res* 355: 71-89
171. Zhao X, Xu C, Domagala J, Drlica K. 1997. *DNA topoisomerase targets of the fluoroquinolones: a strategy for avoiding bacterial resistance. Proc Natl Acad Sci U S A* 94: 13991-6
172. Tamayo M, Santiso R, Gosalvez J, Bou G, Fernandez JL. 2009. *Rapid assessment of the effect of ciprofloxacin on chromosomal DNA from Escherichia coli using an in situ DNA fragmentation assay. BMC Microbiol* 9: 69
173. Zhong L, Zhou W, Wang H, Ding S, Lu Q, et al. 2013. *Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in Arabidopsis under heat stress. Plant Cell* 25: 2925-43
174. Weigel D, Glazebrook J. 2002. *Arabidopsis: A Laboratory Manual*. Cold Spring Harbor Lab Press, Cold Spring Harbor, NY
175. Cappadocia L, Parent JS, Zampini E, Lepage E, Sygusch J, Brisson N. 2012. *A conserved lysine residue of plant Whirly proteins is necessary for higher order protein assembly and protection against DNA damage. Nucleic Acids Res* 40: 258-69
176. Chi W, Mao J, Li Q, Ji D, Zou M, et al. 2010. *Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor SIG6 in the regulation of chloroplast gene expression in Arabidopsis cotyledons. Plant J* 64: 14-25

177. Mouneimne H, Robert J, Jarlier V, Cambau E. 1999. *Type II topoisomerase mutations in ciprofloxacin-resistant strains of Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 43: 62-6
178. Rafii F, Park M. 2005. *Effects of gyrase mutation on the growth kinetics of ciprofloxacin-resistant strains of Clostridium perfringens*. *Anaerobe* 11: 201-5
179. Lewis RJ, Tsai FT, Wigley DB. 1996. *Molecular mechanisms of drug inhibition of DNA gyrase*. *Bioessays* 18: 661-71
180. Mathews CK, Van Holde KE. 1996. *Biochemistry*. Menlo Park, Calif.: Benjamin/Cummings Pub. Co., Inc. xxvii, 1159 p. pp.
181. Pang Q, Hays JB, Rajagopal I. 1993. *Two cDNAs from the plant Arabidopsis thaliana that partially restore recombination proficiency and DNA-damage resistance to E. coli mutants lacking recombination-intermediate-resolution activities*. *Nucleic Acids Res* 21: 1647-53
182. Chan SH, Yu AM, McVey M. 2010. *Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila*. *PLoS Genet* 6: e1001005
183. Yu AM, McVey M. 2010. *Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions*. *Nucleic Acids Res* 38: 5706-17
184. Moldovan GL, Madhavan MV, Mirchandani KD, McCaffrey RM, Vinciguerra P, D'Andrea AD. 2010. *DNA polymerase POLN participates in cross-link repair and homologous recombination*. *Mol Cell Biol* 30: 1088-96
185. Nakanishi H, Suzuki K, Kabeya Y, Miyagishima SY. 2009. *Plant-specific protein MCD1 determines the site of chloroplast division in concert with bacteria-derived MinD*. *Curr Biol* 19: 151-6
186. Kroll D, Meierhoff K, Bechtold N, Kinoshita M, Westphal S, et al. 2001. *VIPP1, a nuclear gene of Arabidopsis thaliana essential for thylakoid membrane formation*. *Proc Natl Acad Sci U S A* 98: 4238-42
187. Lee KP, Kim C, Landgraf F, Apel K. 2007. *EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 104: 10270-5
188. Qi Y, Armbruster U, Schmitz-Linneweber C, Delannoy E, de Longevialle AF, et al. 2012. *Arabidopsis CSP41 proteins form multimeric complexes that bind and stabilize distinct plastid transcripts*. *J Exp Bot* 63: 1251-70
189. Pfannschmidt T, Link G. 1997. *The A and B forms of plastid DNA-dependent RNA polymerase from mustard (Sinapis alba L.) transcribe the same genes in a different developmental context*. *Mol Gen Genet* 257: 35-44

190. Ameer A, Stewart JB, Freyer C, Hagstrom E, Ingman M, et al. 2011. *Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. PLoS Genet* 7: e1002028
191. Williams SL, Mash DC, Zuchner S, Moraes CT. 2013. *Somatic mtDNA mutation spectra in the aging human putamen. PLoS Genet* 9: e1003990
192. Liu P, Carvalho CM, Hastings PJ, Lupski JR. 2012. *Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Genet Dev* 22: 211-20
193. Lupski JR. 1998. *Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends in Genetics* 14: 417-22
194. Vermulst M, Wanagat J, Kujoth GC, Bielas JH, Rabinovitch PS, et al. 2008. *DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. Nat Genet* 40: 392-4
195. Singh KK, Ayyasamy V, Owens KM, Koul MS, Vujcic M. 2009. *Mutations in mitochondrial DNA polymerase-gamma promote breast tumorigenesis. J Hum Genet* 54: 516-24
196. Cappadocia L, Maréchal A, Parent JS, Lepage E, Sygusch J, Brisson N. 2010. *Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. Plant Cell* 22: 1849-67
197. Maréchal A, Parent JS, Veronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. 2009. *Whirly proteins maintain plastid genome stability in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America* 106: 14693-8
198. Parent JS, Lepage E, Brisson N. 2011. *Divergent roles for the two Poll-like organelle DNA polymerases of Arabidopsis. Plant Physiology* 156: 254-62
199. Robinson SJ, Tang LH, Mooney BA, McKay SJ, Clarke WE, et al. 2009. *An archived activation tagged population of Arabidopsis thaliana to facilitate forward genetics approaches. BMC Plant Biol* 9: 101
200. Ji J, Braam J. 2010. *Restriction site extension PCR: a novel method for high-throughput characterization of tagged DNA fragments and genome walking. PLoS One* 5: e10577
201. Chase JW, Murphy JB, Whittier RF, Lorensen E, Sninsky JJ. 1983. *Amplification of ssb-1 mutant single-stranded DNA-binding protein in Escherichia coli. Journal of Molecular Biology* 164: 193-211
202. Meyer RR, Glassberg J, Kornberg A. 1979. *An Escherichia coli mutant defective in single-strand binding protein is defective in DNA replication. Proc Natl Acad Sci U S A* 76: 1702-5
203. Burgers PM. 1998. *Eukaryotic DNA polymerases in DNA replication and DNA repair. Chromosoma* 107: 218-27

204. Dorson JW, Deutsch WA, Moses RE. 1978. *Role of DNA polymerases in excision repair in Escherichia coli.* *J Biol Chem* 253: 660-4
205. Copeland WC, Longley MJ. 2003. *DNA polymerase gamma in mitochondrial DNA replication and repair.* *ScientificWorldJournal* 3: 34-44
206. Branzei D, Foiani M. 2010. *Leaping forks at inverted repeats.* *Genes Dev* 24: 5-9
207. Kim KJ, Choi KS, Jansen RK. 2005. *Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae).* *Mol Biol Evol* 22: 1783-92
208. Cosner ME, Jansen RK, Palmer JD, Downie SR. 1997. *The highly rearranged chloroplast genome of Trachelium caeruleum (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families.* *Current Genetics* 31: 419-29
209. 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: 8319-25
210. Dutra BE, Lovett ST. 2006. *Cis and trans-acting effects on a mutational hotspot involving a replication template switch.* *Journal of Molecular Biology* 356: 300-11
211. Wang JD, Berkmen MB, Grossman AD. 2007. *Genome-wide coorientation of replication and transcription reduces adverse effects on replication in Bacillus subtilis.* *Proc Natl Acad Sci U S A* 104: 5608-13
212. Kamarthapu V, Nudler E. 2015. *Rethinking transcription coupled DNA repair.* *Curr Opin Microbiol* 24: 15-20
213. 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: 5989-94
214. Snowden T, Acharya S, Butz C, Berardini M, Fishel R. 2004. *hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes.* *Mol Cell* 15: 437-51
215. Larijani M, Zaheen A, Frieder D, Wang Y, Wu GE, et al. 2005. *Lack of MSH2 involvement differentiates V(D)J recombination from other non-homologous end joining events.* *Nucleic Acids Res* 33: 6733-42
216. Honda M, Okuno Y, Hengel SR, Martin-Lopez JV, Cook CP, et al. 2014. *Mismatch repair protein hMSH2-hMSH6 recognizes mismatches and forms sliding clamps within a D-loop recombination intermediate.* *Proc Natl Acad Sci U S A* 111: E316-25
217. Copeland WC, Longley MJ. 2014. *Mitochondrial genome maintenance in health and disease.* *DNA Repair (Amst)* 19: 190-8

218. Vellarikkal SK, Dhiman H, Joshi K, Hasija Y, Sivasubbu S, Scaria V. 2015. *mit-o-matic: A Comprehensive Computational Pipeline for Clinical Evaluation of Mitochondrial Variations from Next-Generation Sequencing Datasets*. *Hum Mutat* 36: 419-24
219. Tonin Y, Entelis N. 2014. *[Mitochondrial DNA diseases and therapeutic strategies]*. *Med Sci (Paris)* 30: 1101-9
220. Fan W, Lin CS, Potluri P, Procaccio V, Wallace DC. 2012. *mtDNA lineage analysis of mouse L-cell lines reveals the accumulation of multiple mtDNA mutants and intermolecular recombination*. *Genes Dev* 26: 384-94