Spring 2024

Differential Roles of Replication Protein A Large Subunit Paralogs in DNA Damage Repair

Ian Mills
University of New Hampshire

Follow this and additional works at: https://scholars.unh.edu/dissertation

DIFFERENTIAL ROLES OF REPLICATION PROTEIN A LARGE SUBUNIT PARALOGS IN DNA DAMAGE REPAIR

By

Ian A. Mills

Bachelor of Arts, Biochemistry and Biophysics, Rensselaer Polytechnic Institute, 2017

DISSERTATION

Submitted to the University of New Hampshire

In Partial Fulfillment

of the Requirements of the Degree of

Doctor of Philosophy

In

Genetics

May 2024

This dissertation has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Genetics by Ian A. Mills

Dissertation Director, Kevin M. Culligan
Research Assistant Professor
Molecular, Cellular, & Biomedical Sciences

Anna O'Brien, Assistant Professor
Molecular, Cellular, & Biomedical Sciences

Estelle Hrabak, Associate Professor Emerita
Molecular, Cellular, & Biomedical Sciences

Subhash Minocha,
Professor Emeritus
College of Biological Sciences

Don Wojchowski, Professor
Molecular, Cellular, and Biomedical
Sciences

On May 9th, 2024

DEDICATION

To all my family, friends, and cats
Thank you all so much

ACKNOWLEDGMENTS

Thank you to my dissertation advisor Dr. Kevin Culligan for his support and guidance throughout my time as a graduate student in his lab.

I am also thankful for the encouragement and advice from my committee members.

I especially would like to thank all the undergraduate students for their help not only in completing the work but also in keeping the morale high in the lab. Noreen Syedah, Mariah Desroches, and Tallia Algiere: you are all absolute stars.

Finally, thank you so much to all my friends and family for their support and encouragement as I waffled on endlessly about plants.

TABLE OF CONTENTS

.

LIST OF TABLES

LIST OF FIGURES

Abbreviations

ABRC	*A. thaliana* Biological Resource Center
Arabidopsis	*Arabidopsis thaliana*
ATM	Ataxia-Telangiectasia Mutated
ATR	Ataxia-Telangiectasia and Rad 3 Related
ATRIP	ATR-interacting Protein
BS-I	Binding Surface I
Cas	CRISPR-associated
CPT	Camptothecin
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
C-ZFKO	RPA1C zinc finger knock-out
DBD	DNA-binding domain
DSB	Double-strand Break
dsDNA	Double-stranded DNA
E-ZFKO	RPA1E zinc finger knockout
GUS	β-glucuronidase
HMCES	5-hydroxymethylcytosine binding, ESC-specific
HRR	Homologous recombination repair
IDN2	INVOLVED IN DE NOVO2
NHEJ	Non-homologous end joining
MMEJ	Microhomology-mediated end joining
MRN	Mre11-Rad50-NBS1
PAM	Protospacer adjacent motif
PARP1	Poly (ADP-ribose) polymerase 1
PI3K	Phosphatidylinositol-3 kinase-like
Polθ	Polymerase theta
ROS	Reactive oxygen species

RPA	Replication Protein A
SDSA	Synthesis-dependent strand annealing
SSA	Single-strand annealing
SSB	Single-strand Break
SSBP	Single-stranded binding proteins
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
TALEN	Transcription activator-like effector nuclease
T-DNA	Transfer DNA
TEB	TEBICHI
X-GLUC	5-bromo-4-chloro-3-indolyl glucuronide
ZFKO	Zinc finger knockout
ZFM	Zinc finger motif
ZFN	Zinc finger nuclease

ABSTRACT

DNA damage is a constant threat for all organisms. The most severe form of DNA damage is a double-strand break (DSB) that can have disastrous consequences to the cell including genome truncation and programmed cell death. As such, organisms have multiple complex pathways dedicated to repairing DSBs. These repair pathways begin with a signaling cascade that results in the resecting of the ends of the break to create 3' OH overhangs which are immediately bound by a single-stranded binding protein called Replication Protein A (RPA). This is followed by the cell either undergoing programmed cell death or utilizing non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), or a homologous recombination repair (HRR) mechanism. HRR uses either single-strand annealing (SSA) or synthesis-dependent strand annealing (SDSA) to repair a DSB, but in either case, RPA must bind to the single-stranded DNA (ssDNA) prior to initiation of repair.

RPA is a heterotrimeric ssDNA binding protein that is highly conserved across all eukaryotes. While animal and yeast genomes typically only have a single copy of each subunit, plants have multiple paralogs of each. In plants the number of paralogs of each subunit varies considerably depending on the species. The model organism Arabidopsis (*Arabidopsis thaliana*) has five RPA1 paralogs and two each of RPA2 and RPA3. The RPA1 paralogs are divided into three groups by function. The group C paralogs (RPA1C and E in Arabidopsis) are involved in DNA damage repair and have a C-terminal extension that is only found in group C paralogs. This C-terminal extension contains a zinc finger motif (ZFM) that is highly conserved and is therefore hypothesized to be critical to the functionality of the paralogs during DNA damage repair.

This dissertation investigated the role of RPA in determining how the DSB is repaired and how RPA functions during DSB repair. The role of the C-terminal ZFM in DSB repair was examined via the CRISPR-Cas9 mediated alteration of *RPA1* genes in wild-type Arabidopsis plants. CRISPR mutant plants lacking the ZFM of the group C RPA1 paralogs were exposed to DNA damaging agents and their phenotypes were compared to wild-type Arabidopsis as well as to previously characterized T-DNA null mutants for group C paralogs (*rpa1c* and *rpa1e*). These plant lines were also used for the characterization of HRR usage with GUS gene reporter lines. This dissertation details how the ZFM of the group C paralogs is involved in the functionality of the paralogs as well as the overall usage of the group C paralogs in HRR.

CHAPTER I

INTRODUCTION

DNA damage and repair

Maintenance of the genome is a crucial intracellular process necessary for the survival and reproduction of all organisms. While alterations to the genome are essential for genetic diversity and evolution, excessive changes can have disastrous consequences (Karthika et al., 2020). Failure to maintain the integrity of the genome can have a wide array of effects on an organism, ranging from relatively minor phenotypic changes to death. As such, organisms constantly strive to maintain a stable genome. This can be challenging to achieve due to the continuous presence of DNA-damaging agents, both exogenous and endogenous (Aguilera & García-Muse, 2013). Exogenous DNA-damaging agents arise from outside the organism and include chemical mutagens, such as base analogs and intercalating agents, as well as ultraviolet and ionizing radiation (Britt, 2004; Kim et al., 2019; Manova & Gruszka, 2015; You & Chan, 2015). DNA damage can also arise following infection by pathogens, either due to direct action of the pathogen or as a consequence of the immune response of the organism (Song & Bent, 2014; Weitzman & Weitzman, 2014). In contrast, endogenous agents arise from within the organism and include reactive oxygen species (ROS) and alkylating agents (Bessho, 2003; Sharma et al., 2012). Additionally, alterations to the genome can be a direct result of the

incorrect nucleotide being added during DNA replication (Ganai & Johansson, 2016). Due to the prevalence and persistence of these potential sources of mutations, organisms have evolved multiple elaborate systems of DNA damage repair which help mitigate the effects of the mutagens.

DNA damage repair mechanisms are faced with the challenge of a wide variety of types of DNA damage. Alterations such as base substitutions during DNA replication are relatively minor compared to more deleterious damage such as single or double-strand breaks (SSBs and DSBs) and require entirely different repair mechanisms. To repair DNA damage, the cell must first sense that the damage has occurred and then initiate a response. The checkpoint kinases ataxia-telangiectasia mutated (ATM) and ataxia-telangiectasia and Rad 3 related (ATR) are the two key factors in this process (Zhou & Elledge, 2000). Both proteins are phosphatidylinositol-3 kinase-like (PI3K) protein kinases and together, these kinases serve as the primary signal transducer in DNA damage repair (Gimenez & Manzano-Agugliaro, 2017). ATM is specifically involved in repair of DSBs and is recruited by the Mre11-Rad50-Nbs1 (MRN) complex that binds to and unwinds the damaged DNA ends (Lee & Paull, 2005). ATR is a more general factor that responds to a wide variety of DNA damage including DSBs, base adducts, crosslinks, and replicative stress (Cimprich & Cortez, 2008). Recruitment of ATR does not require the MRN complex but instead involves ATR-interacting protein (ATRIP), as well as the single-stranded binding protein Replication Protein A (RPA) (Cortez et al., 2001; Zou & Elledge, 2003). RPA coats ssDNA, protecting it from unwanted interactions and RPA-coated ssDNA serves as a binding target for ATRIP, which then recruits ATR (Ball et al., 2007). Following the binding of either ATR or ATM, a signaling cascade is initiated which includes the phosphorylation of the

kinases Chk1, Chk2, and MK2 which further propagate the cascade (Q. Liu et al., 2000; Matsuoka et al., 1998; Reinhardt et al., 2007).

The initiation of this signaling cascade causes the cell cycle to halt immediately while the cell prepares to either repair the damage or initiate programmed cell death (Amiard et al., 2011). Initiating programmed cell death is a last resort but is sometimes necessary due to the possible consequences of a DSB (Roy, 2014). DSBs are the most deleterious form of DNA damage and can lead to chromosome loss or truncation (Charbonnel et al., 2011; Waterworth et al., 2011). The issue of genomic alterations is particularly problematic if the damage occurs within a germ cell (Ricaud et al., 2007). In these cases, any DNA damage that goes unrepaired, or is repaired incorrectly, will result in mutations. These mutations could then accumulate and directly or indirectly impact the reproductive fitness of the organism and any future progeny. Therefore, in cases of severe DNA damage, programmed cell death, or at the very least differentiation to a non-proliferating cell type, is sometimes the preferred cellular outcome (Fulcher & Sablowski, 2009). However, not all DSBs result in programmed cell death. Instead, multiple complex DNA repair pathways can be utilized to repair DSBs (West et al., 2004).

After the recognition of a DSB and the initiation of the signaling cascade by ATM and ATR, should the cell attempt to repair the DSB, it does so by following one of three main mechanisms: non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), or homologous recombination repair (HRR) (Figure 1.1). It is not fully understood how the cell determines which repair mechanism to utilize but it is, at least in part, based upon the type of damage and the stage of the cell cycle (Roy, 2014). HRR is typically utilized during the G2 and S phases of the cell cycle, while NHEJ can be used at any stage of the cell cycle but is most dominant during G1 and early S phases. This ability for NHEJ to be utilized at any stage of

the cell cycle is due in part to the requirements, or lack thereof, for NHEJ. Repair of a DSB via NHEJ does not require the ends of the DSB to be resectioned and can be completed with little to no sequence homology. Due to its limited requirements, NHEJ is the most common form of DSB repair and is used significantly more often than HRR or MMEJ (Kirik et al., 2000; Salomon & Puchta, 1998). MMEJ (also known as alternative end-joining or alternative non-homologous end joining) was previously considered to be a backup pathway for NHEJ but has recently been shown to have a larger, more consistent role in DSB repair (Doonan & Sablowski, 2010; Nussenzweig & Nussenzweig, 2007). MMEJ requires microhomology of 2-20 base pairs to repair a DSB and has the significant drawback that it typically results in the loss of genetic material as any sequence that is located between the microhomology and the DSB is lost (Chang et al., 2017; Puchta & Fauser, 2014). This loss of sequence is a trait shared with SSA, one of the two HRR pathways.

HRR, while less commonly used than NHEJ, is still a critical repair mechanism of DSBs. There are two variants of HRR: single-strand annealing (SSA) and synthesis-dependent strand annealing (SDSA; Figure 1.2; Puchta, 2004) . Both mechanisms have the same initial steps: the initiation of a signaling cascade by ATM and ATR. Following signaling, in both SSA and SDSA, the ends of the DSB are resectioned by the MRN complex, leaving 3' OH overhangs (Mannuss et al., 2010). RPA binds these regions of single-stranded DNA to protect the DNA from harm and any unwanted interactions. It is at this stage that the processes for SSA and SDSA diverge (Figures 1.1 and 1.2).
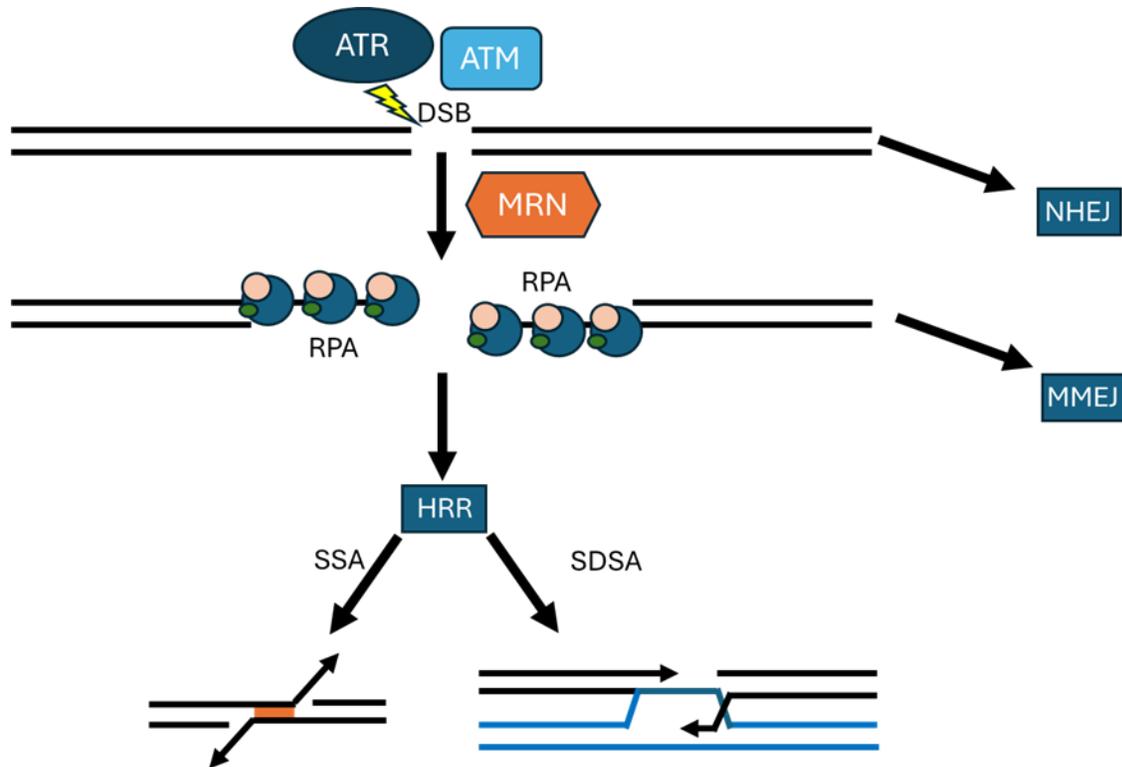
Figure 1.1. Simplified model of double-strand break repair in plants. Ataxia-Telangiectasia Mutated (ATM) and Ataxia-Telangiectasia and Rad 3 Related (ATR) initiate a signaling cascade. The signaling results in either the direct repair of the break by non-homologous end joining (NHEJ) or the resecting of the end of the break by the Mre11-Rad50-NBS1 (MRN) complex to create 3' OH overhangs which are bound by Replication Protein A (RPA). At this point, the cell either repairs the break through microhomology-mediated end joining (MMEJ) or through one of the two versions of homologous recombination repair (HRR): single-strand annealing (SSA) or synthesis dependent strand annealing (SDSA).

When a DSB is repaired via SSA, a region of homology of at least 20 base pairs is identified on either side of the DSB and the two resected single strands anneal together at that location (Figure 1.2A). Following annealing, the regions of DNA between the homologous region and the DSB are removed, and any gaps are filled in by a DNA polymerase. Finally, DNA ligase seals the final gap and completes the repair of the damage. While this method of DSB repair is effective, it also results in a significant loss of DNA sequence because any DNA between the site of the DSB and the site of homology is trimmed off (Siebert & Puchta, 2002).

This trimming process is mediated by the ERCC1/XPF complex and can be extensive enough to result in the loss of multiple kilobases of genomic sequence (Mendez-Dorantes et al., 2018). This makes SSA a non-conservative repair mechanism.

In contrast, SDSA is a conservative repair mechanism (Figure 1.2B). During SDSA, instead of looking for a region of homology on either side of the DSB, the single-stranded regions invade a homologous region, typically on a sister chromatid or a nearby region on the same chromosome (Gisler et al., 2002). Strand invasion is facilitated by Rad51 and Rad52 which are recruited to the single-stranded overhangs by RPA. After strand invasion, the homologous region is used as a template for DNA synthesis and once synthesis has passed the region of the DSB, the junction can be resolved, resulting in a completely repaired genome (Mannuss et al., 2010).While this repair method is ideal since there is no loss of genetic information, it can result in the alteration of the genome, depending on which homologous region is used (Puchta & Fauser, 2014). Additionally, it is a far less efficient repair mechanism than SSA (approximately 5x less efficient) and has the limitation of requiring a homologous region readily available (Orel et al., 2003). Homologous regions are not always available, particularly during G1 phase when there is no sister chromatid. Therefore, despite the theoretically perfect repair of the genome via SDSA, it is used far less commonly than either SSA or NHEJ (Orel et al., 2003).
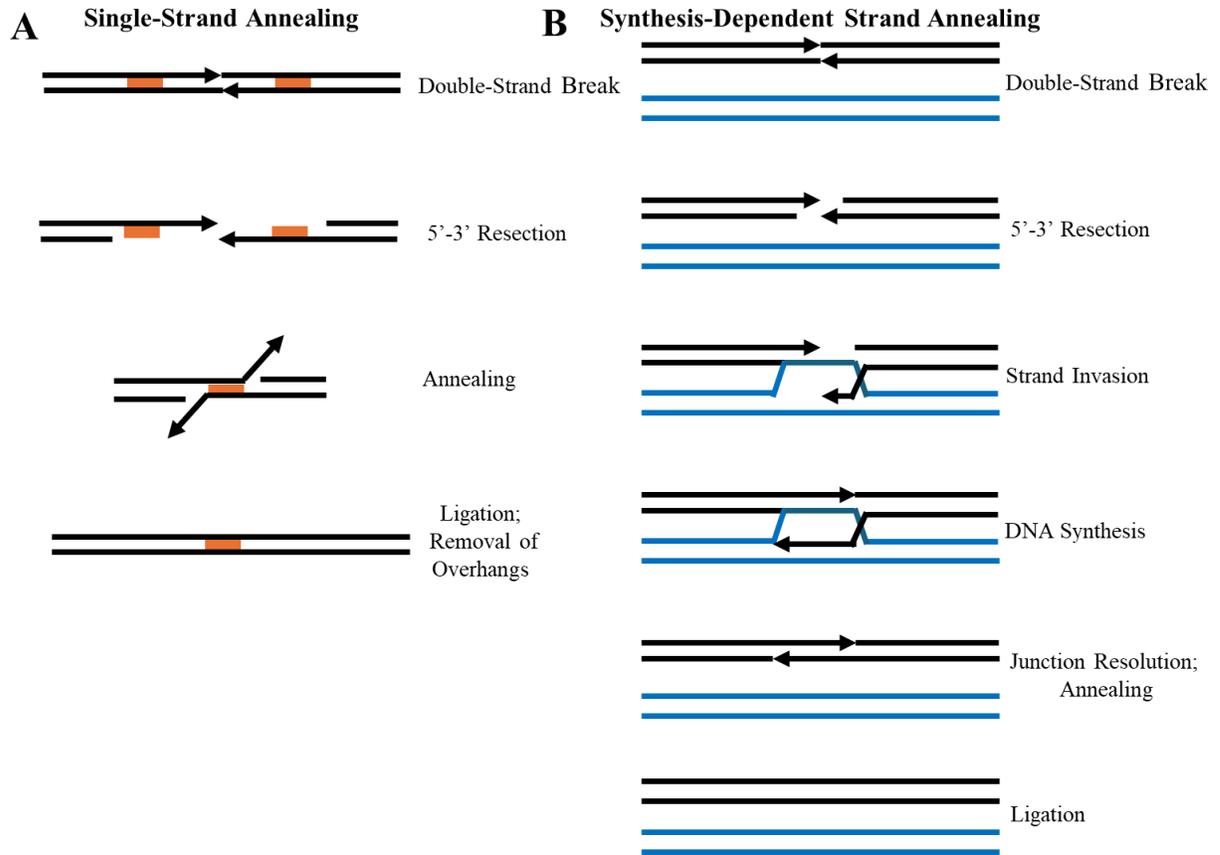
Figure 1.2. Comparison of homologous recombination repair pathways. Both pathways start with the same initial steps but then diverge. A) SSA is a non-conservative repair mechanism with few requirements except two regions of homology (orange) flanking the site of the break. B) SDSA is a conservative repair mechanism but requires a homologous region of DNA to use as a template for DNA synthesis.

Replication Protein A

While eukaryotic genomes are maintained as double-stranded DNA (dsDNA), processes such as DNA replication, recombination, and repair require segments of DNA to be unwound to single-stranded intermediates (Shereda et al., 2008). These regions of single-stranded DNA (ssDNA) lack the stability of dsDNA and as such are susceptible to chemical and nucleolytic degradation. Additionally, if left unchecked, ssDNA is capable of unwanted binding interactions such as the formation of hairpins or re-annealing to a region of complementary DNA (Marceau,

2012). To limit possible unwanted interactions of ssDNA, prokaryotic and eukaryotic cells

contain a general class of proteins called single-stranded binding proteins (SSBPs) which have

the crucial cellular role of binding to and protecting ssDNA (Pal & Levy, 2019). Beyond simply

protecting ssDNA, SSBPs also play a role in recruiting partner proteins to the DNA, aiding in the

progression of whatever process is underway.

In eukaryotes, RPA is an SSBP that is required for most cellular processes involving

ssDNA. This includes DNA replication, recombination, and repair (Wold, 1997). RPA is a

heterotrimer consisting of three subunits: RPA1 (~70 kDa), RPA2 (~32 kDa), and RPA3

(~14kDa). The RPA heterotrimer binds with extremely high (subnanomolar) affinity to ssDNA

through the use of six distinct DNA-binding domains (DBDs) which can be found across the

three subunits (Kim et al., 1994). The six DBDs are designated DBD A-F. Four of these domains

(DBD A, -B, -C, and -F) are found on the largest RPA subunit, RPA1, while the other two RPA

subunits each have a single DBD, with DBD-D on RPA2 and DBD-E on RPA3 (Figure 1.3;

Yates et al., 2018) . The term DBD is somewhat misleading, as the DBDs are also capable of

interacting with molecules other than DNA. For example, DBD-C, DBD-D, and DBD-E interact

to form the trimerization core that facilitates the formation of the RPA heterotrimer (Bochkareva

et al., 2002). These three DBDs are crucial for the formation of the heterotrimer but still retain

DNA binding capability, and as such the functionality of these DBDs is not restricted to

trimerization (Bochkarev et al., 1999). The majority of the DNA binding capability of RPA is due

to the DBDs on RPA1, particularly DBD-A, which has the greatest affinity for ssDNA (Wyka et

al., 2003). Additionally, due to the relatively short distance (11 amino acids) between DBD-A

and DBD-B, the two domains can work in conjunction, greatly enhancing the overall binding

capability of the domains and the heterotrimer as a whole (Arunkumar et al., 2003).
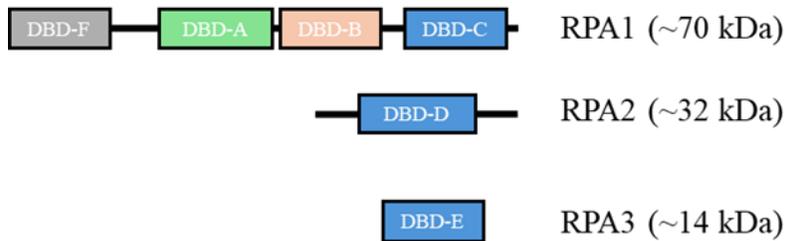
Figure 1.3. Structure of RPA subunits**.** Replication Protein A (RPA) is a heterotrimer consisting of RPA1, RPA2, and RPA3. The subunits contain six DNA binding domains (DBDs) but three of them, DBD-C, DBD-D, and DBD-E (indicated in blue) form the trimerization core. In addition to DNA binding capabilities of these DBDs, the interaction between the DBDs facilitates the formation of the heterotrimer.

Beyond DNA binding capability, the DBDs of RPA are also capable of interacting with numerous protein targets (as seen with the trimerization core). This is crucial for the ability of RPA to function within multiple pathways that involve ssDNA (Chen & Wold, 2014). RPA can interact with proteins involved in multiple processes of DNA metabolism such as replication, recombination, and repair, although it is not well understood how these interactions work or how the correct interactions are made based on what process is occurring. These protein-protein interactions are not necessarily direct and can sometimes occur due to the destabilizing effect RPA can have on dsDNA. RPA binding to ssDNA stabilizes the bound DNA which can have the effect of destabilizing nearby dsDNA in an ATP and Mg^{2+} independent manner (Nguyen et al., 2014). This destabilization can provide binding sites for other proteins, thereby enabling indirect interaction between RPA and other key proteins in DNA metabolism. This dual role of stabilizing ssDNA and enabling the binding of other proteins is why RPA is critical for DNA metabolism to proceed. Null mutations in any of the RPA subunits are typically lethal, while mutations that reduce RPA function result in genome instability and reduced DNA repair capability (Haring et al., 2008). Even a reduction in cellular levels of RPA can have a significant impact on organismal health with potential consequences including reduced lifespan, decreased DNA damage response,

and tumor growth (O'Driscoll et al., 2007; Wang et al., 2005). These varied symptoms of a deficit in RPA emphasize the importance of RPA across multiple different pathways.

In animals and yeast, the multifunctionality of RPA is due, in part, to alterations in protein structure following hyper-phosphorylation (post-translational modification). Hyper-phosphorylation by checkpoint kinases, including ATM, ATR, and DNA-PK, has been observed after both DNA damage and replication stress and is believed to alter the shape of RPA and thereby how it interacts with both DNA and proteins (Binz et al., 2004; Vassin et al., 2004). While hyper-phosphorylation has been observed in animals and yeast, there have been no reports of this activity in plants. Indeed, following induced DNA damage in rice, there was no hyper-phosphorylation detected at all (Marwedel et al., 2003). Hyper-phosphorylation alters the structure, and the function of the protein acting as a molecular 'switch' between DNA damage repair over DNA replication (Binz et al., 2004; Vassin et al., 2004). As plants do not utilize hyper-phosphorylation of RPA, there must be a different regulatory process in place to facilitate RPA functioning during different cellular processes. The presence of multiple paralogs of the RPA subunits in most plants (while animals and yeast typically have only a single copy of each subunit) presents a possible alternate mechanism for regulation (Aklilu & Culligan, 2016). Through subfunctionalization of the RPA subunit paralogs, plants may have paralogs dedicated to specific cellular processes and thereby avoid the need for hyper-phosphorylation to modulate the activity of RPA.

Replication Protein A paralogs

With a few exceptions, plants have multiple paralogs of each RPA subunit, while animals and yeast typically have a single version of each subunit. For example, some mammals,

including humans, have two RPA2 paralogs (Keshav et al., 1995). While all plants appear to have multiple RPA subunit paralogs, the number of paralogs of each subunit varies between plants (Shultz et al., 2007). Rice has three paralogs of both RPA1 and RPA2 and only a single RPA3 variant, while *Arabidopsis thaliana* has five RPA1 paralogs (RPA1A, RPA1B, RPA1C, RPA1D, and RPA1E) and two each of RPA2 (RPA2A and RPA2B) and RPA3 (RPA3A and RPA3B) (Figure 1.4; Aklilu et al., 2014; Ganpudi & Schroeder, 2011; Ishibashi et al., 2006). This matches the overall trend found through the analysis of > 20 complete genomes which found that most plants have three RPA1 paralogs, but those belonging to the family *Brassicaceae* (including Arabidopsis) instead have five RPA1 paralogs (Aklilu & Culligan, 2016). All of the Arabidopsis paralogs for each of the three subunits are functional and capable of assembling into a heterotrimer, but different combinations of paralogs form heterotrimers with divergent binding capabilities and cellular function (Eschbach & Kobbe, 2014). These possible heterotrimers have not yet all been determined and classified, but the existence and differing functionality of the heterotrimers emphasizes the difference in the regulation of RPA that must exist between plants and other eukaryotes due to the presence of multiple subunit paralogs. While other eukaryotes must utilize post-translational modifications such as hyper-phosphorylation to adjust the function of RPA for different pathways, plants may use an alternate system utilizing different combinations of paralogs to build functionally diverse RPA heterotrimers. In this system, the paralogs for each subunit may function only in a specific pathway and only assemble into a heterotrimer with paralogs of the other subunits which also work in that specific pathway.
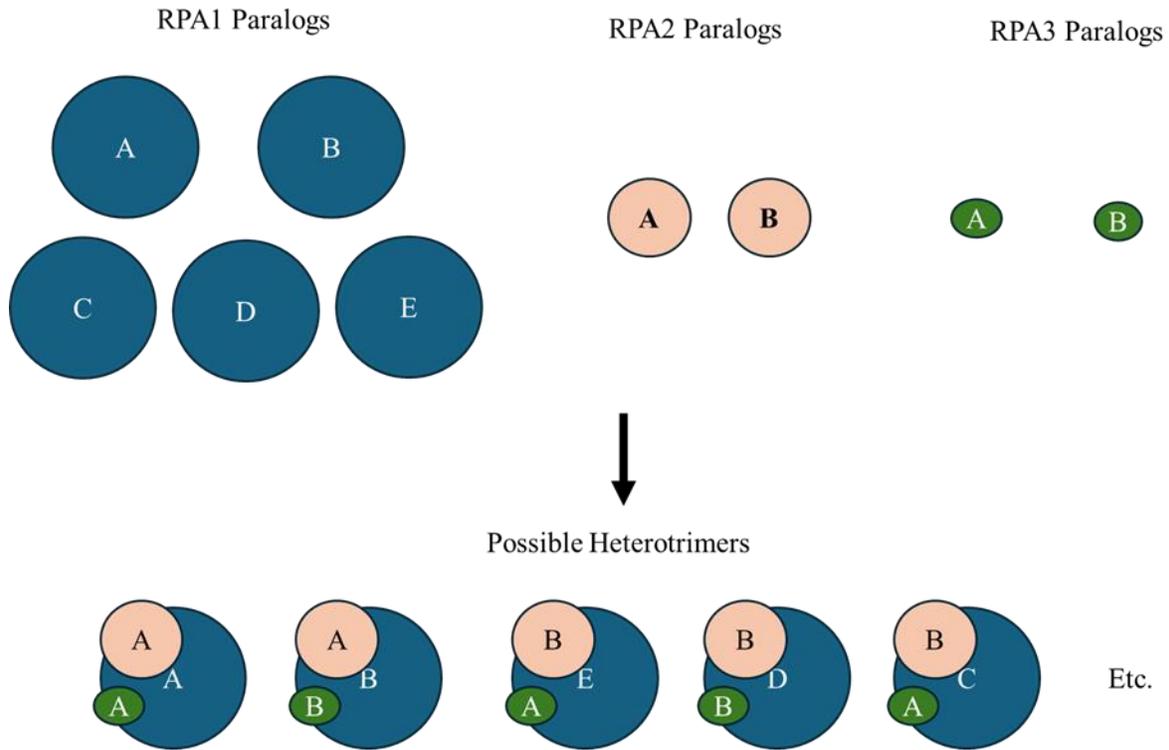
Figure 1.4. Model of Arabidopsis thaliana RPA subunits and heterotrimer assembly. *Arabidopsis thaliana* has multiple paralogs of each of the three RPA subunits. These different paralogs are all capable of assembling into a functional RPA heterotrimer, but different paralogs combine to form heterotrimers with differing functionality (Eschbach & Kobbe, 2014).

While striking, this difference in paralogs between animals/yeast and plants is somewhat to be expected. Paralogs arise through genetic duplication such as whole-genome duplication, segmental duplication, or tandem gene duplication, and genetic duplication is relatively common in plants (Lockton & Gaut, 2005). Following a gene duplication event, there are three possible outcomes. First, one copy might undergo pseudogenization and slowly lose its function(s) over time. Second, the two duplicate genes might undergo subfunctionalization and both lose different functional aspects while maintaining approximately equal functionalities overall as the original gene prior to duplication. Third, one gene may undergo neofunctionalization and gain a new,

beneficial function while losing its original functionality altogether, while the other copy retains its original function (Louis, 2007; Lynch & Conery, 2000; Moore & Purugganan, 2005).

In Arabidopsis, the RPA1 paralogs have undergone subfunctionalization, while still maintaining some degree of overlapping functionality (Aklilu et al., 2014). The five Arabidopsis RPA1 paralogs fall into one of three groups by function. Group A paralogs (RPA1A) are involved in the progression of meiosis, group B paralogs (RPA1B and D) are involved in replication, and group C paralogs (RPA1C and E) are involved in DNA damage repair. This partitioning of the paralogs into three distinct groups is seen across most plants but the presence of multiple paralogs within a group is less universal. There is considerable variance among different species and families, with some, such as soybean (*Glycine max*) and maize (*Zea mays*) possessing two group B RPA1 paralogs while sorghum (*Sorghum bicolor*) and millet (*Setaria italica*) have four and two group C RPA1 paralogs, respectively. Interestingly, the presence of two group B RPA1 and group C RPA1 paralogs that is seen in Arabidopsis seems to be unique to the *Brassicaceae* family (*A. thaliana, Arabidopsis lyrata, and Capsella rubella*), suggesting that the gene duplication occurred relatively recently.
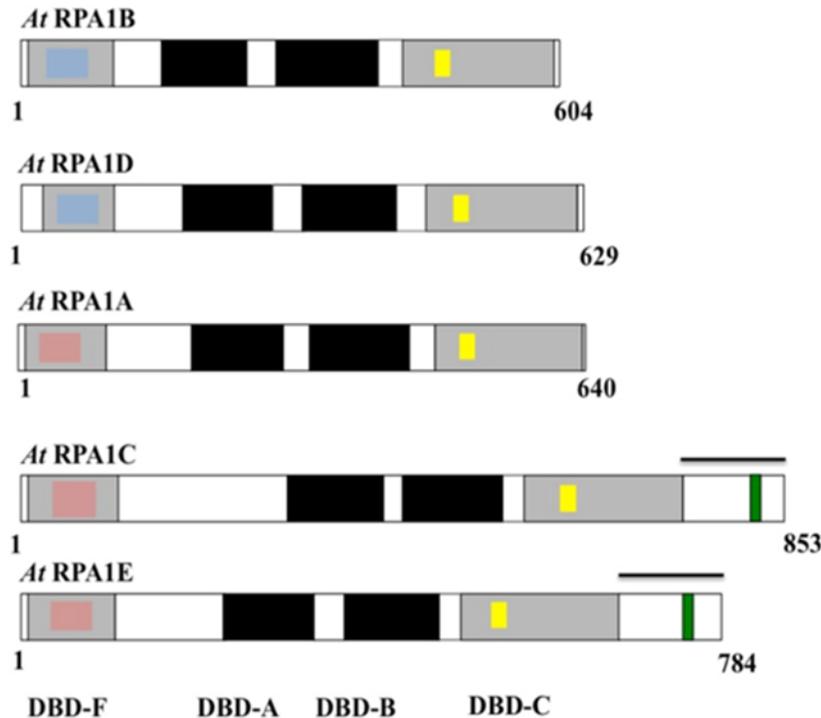
Figure 1.5. Model of Arabidopsis thaliana RPA1 paralogs. *Arabidopsis thaliana* has five RPA1 paralogs. The paralogs all share the same four DNA binding domains (DBD-A, DBD-B, DBD-C, and DBD-F) but Binding Surface I (BS-I, pink boxes) is only found in RPA1A, RPA1C, and RPA1E. DBD-F in RPA1B and RPA1D does not contain the BS-I subdomain (blue boxes). All RPA1 paralogs contain a zinc finger motif (ZFM) within DBD-C (yellow boxes). RPA1C and RPA1E also have a unique C-terminal extension which contains a CCHC-type type ZFM (green; Aklilu & Culligan, 2016)

Arabidopsis RPA1 paralogs are structurally very similar (Figure 1.5; Aklilu & Culligan, 2016). All five paralogs maintain the same four DNA binding domains (DBD-A, DBD-B, DBD-C, and DBD-F) but the subdomain Binding Surface I (BS-I) can only be found in RPA1A, RPA1C, and RPA1E. BS-I is required for functionality during DNA damage repair but unnecessary for activity during DNA replication (Haring et al., 2008; Longhese et al., 1994; Umezu et al., 1998). The other distinguishing feature between the paralogs is the C-terminal extension, which is found only in group C paralogs. This region of ~176 amino acids in RPA1C paralogs and ~119 amino acids in RPA1E paralogs always contains at least one CCHC-type

(CX2CX4HX4C) zinc finger motif (ZFM) (Aklilu & Culligan, 2016). In the *Brassicaceae* plants, each group C paralog has a single ZFM, at approximately the same location. However, in non-*Brassicaceae* plants, which have only a single group C paralog, two or more ZFMs are found within the C-terminal extension. These non-*Brassicaceae* ZFMs fall into two distinct clusters, each with its own unique sequence identity. Typically (as seen in tomato, cucumber, and maize, among others), the clusters are separated by ~30 amino acids within the C-terminal extension (Aklilu & Culligan, 2016). Maintenance of both ZFM groups suggests that both groups may have distinct and important roles in the functionality of RPA1 group C paralogs.

This dissertation details my investigation into the roles of RPA1C and RPA1E in DSB repair in *Arabidopsis thaliana*. Chapter II covers the investigation into the ZFM motifs of the group C paralogs and how the motifs contribute to DSB repair. Chapter III describes the study of the differential contributions of the group C paralogs to the DSB repair mechanisms of SSA and SDSA and what role the C-terminal ZFMs may play in this differentiation.

CHAPTER II

ARABIDOPISIS THALIANA RPA1C AND RPA1E C-TERMINAL EXTENSION ZINC-FINGER MOTIFS ARE CRUCIAL TO THE FUNCTIONALITY OF THE PARALOGS IN DOUBLE-STRAND BREAK REPAIR

ABSTRACT

Maintenance of the genome is crucial for the health of an organism as well as its reproductive fitness, making the repair of DNA damage a key process. The most deleterious form of DNA damage is a double-strand break (DSB), which, if left unrepaired, can lead to genome loss or truncation and potentially programmed cellular death. After a DSB has been detected, a signaling cascade halts the progression of the cell cycle and prepares the DSB for repair by resection of the DNA ends, creating 3' OH overhangs. These single-stranded overhangs are immediately bound by single-stranded binding protein Replication Protein A (RPA) both to protect it and to help guide the next steps of the process.

RPA is a heterotrimeric single-stranded DNA (ssDNA) binding protein which is well conserved across all eukaryotes. Plants typically have multiple RPA paralogs while animals and yeast typically have only a single copy of each subunit. Multiple RPA subunit paralogs can be

found in most plants, but in the model organism Arabidopsis (*Arabidopsis thaliana*) there are five RPA1 paralogs and two each of RPA2 and RPA3 (Aklilu et al., 2014).

In plants, the RPA1 paralogs are divided into three groups by function: group A paralogs (RPA1A in Arabidopsis) are involved in the progression of meiosis, group B paralogs (RPA1B and RPA1D in Arabidopsis) are involved in replication, and group C paralogs (RPA1C and RPA1E in Arabidopsis) are involved in DNA damage repair (Aklilu et al., 2014). In Arabidopsis, the RPA1 paralogs are structurally similar, apart from a C-terminal extension which is present only in the group C paralogs. For this reason, I hypothesize that this C-terminal region, and the CCHC-type zinc-finger motif (ZFM) within it, are necessary for the DNA damage repair functionality of group C paralogs.

To test this hypothesis, a CRISPR/Cas9 system was utilized to generate Arabidopsis that lack the C-terminal zinc finger domain of either RPA1C or RPA1E but are otherwise fully functional. These mutant lines were then compared to wild-type Arabidopsis and to previously characterized T-DNA insertion null mutants *rpa1c* and *rpa1e (Aklilu et al., 2014)*. The CRISPR mutant lines and the respective null mutants had normal growth and germination under standard conditions. The CRIPSR mutants were then exposed to DSB inducing agents, specifically either gamma radiation (100 Gy) or camptothecin (CPT) (15 nM). Both the RPA1C and RPA1E null mutants and the respective RPA1 CRISPR mutants displayed identical phenotypes and displayed hypersensitivity to both CPT and gamma radiation. Furthermore, the *rpa1c* and RPA1C CRISPR mutants displayed greater hypersensitivity to both CPT and gamma radiation than the *rpa1e* and RPA1E CRSIPR mutants. Overall, these data suggest that the CCHC-type zinc-finger in the C-terminal extension of group C RPA1 paralogs is crucial for the paralogs' functionality in DSB repair.

INTRODUCTION

CRISPR-Cas9 genome editing system

All organisms are under constant threat of potential attack by pathogenic viruses. Eukaryotes have complex immune systems which help to mitigate the danger of infection and to preserve the health of the organism. Upon recognition of an infection, both plants and animals immediately begin up- or down-regulation of key genes resulting in the production of antimicrobial compounds, activation of programmed cell death in infected cells, and production of signaling molecules to create a systemic immune response (Roudaire et al., 2021). While bacteria do not share these same pathways, even microbes must face the challenge of potential viral infections and have evolved unique mechanisms to reduce the risk and danger associated with a viral infection. One such mechanism is the CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR-associated) prokaryotic immune system.

The CRISPR-Cas system can be divided into two main classes which are further subdivided into six types and several additional sub-types (Makarova et al., 2015). While these different classes and types differ somewhat in terms of exact functionality, the same base of utilizing an adaptive genomic locus (CRISPR) to specifically target invading foreign nucleic acids is shared across all known classes. The genomic locus maintains unique sequences (known as spacers) obtained over time from bacteriophages, transposons, or plasmids, each separated by short repetitive elements (known as repeats). These spacers serve a similar role to memory B cells in the human immune system and allow the bacteria to recognize foreign nucleic acids (Hille & Charpentier, 2016). Each spacer can be transcribed and utilized as CRISPR RNA

(crRNA), which binds to a Cas endonuclease and conveys it to an invading nucleic acid which is complementary to the crRNA. This process results in the cleaving of the foreign nucleic acid and the effective protection of the bacteria from attempted infection by the bacteriophage. The targeting of the endonuclease is limited by the need for a protospacer adjacent motif (PAM), a short sequence that must be directly downstream of the crRNA binding site for the endonuclease to function correctly (Ran et al., 2013).

This use of targeted endonucleases by bacteria did not go unnoticed by researchers and awakened the possibility of co-opting the system and using it as a gene editing tool. While gene-editing tools such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been in use for years, these methodologies have limitations in terms of flexibility and ease of use (Bibikova et al., 2002; Christian et al., 2010). Meanwhile, the CRISPR-Cas system is limited by the need for a PAM but otherwise can be targeted to any genomic sequence using a manufactured crRNA and Watson-Crick base pairing. The best characterized of the CRISPR-Cas systems is the type II CRISPR-Cas9 commandeered from *Streptococcus pyogenes* which requires only a single crRNA and has a relatively short PAM (NGG), making it extremely versatile and accessible (Jiang & Doudna, 2017).

Agrobacterium tumefaciens-mediated plant transformation

Agrobacteria are a genus of bacteria that are largely pathogenic. Depending on the species and the strain, the host range includes dicot and monocot angiosperm species, gymnosperms, and even some fungi including yeasts (Anderson & Moore, 1979). Infections have ramifications including crown gall disease and hairy root disease and typically involve the insertion of bacterial DNA into the host genome. This transferred DNA (T-DNA) seems to be

integrated randomly into the genome and includes genes that aid in the infection process and are involved in the production of tumors in which the bacteria can thrive (Gelvin, 2003; Kim et al., 2007; Tinland et al., 1994). The T-DNA originates from a plasmid within the bacterium and only a small portion of the plasmid, bound on either side by T-DNA border sequences, is transferred into the host (Jouanin et al., 1989).

Following the initial understanding of how the T-DNA insertion process functions, the possibility of utilizing the process for the genetic manipulation of plants swiftly followed. This usage generally falls into one of two categories: insertion of T-DNA to directly disrupt gene function or insertion of T-DNA to facilitate the addition of desired gene(s) or other DNA elements (Kohli et al., 2003; Krysan et al., 1999).

The use of Agrobacteria to modify the genome of a plant relies on the knowledge of the T-DNA border sequences and the necessary components for T-DNA insertion to occur. Using this knowledge, it is possible to engineer a plasmid that contains only the necessary components of the T-DNA insertion machinery and no longer has the genes involved in tumor formation and pathogenicity. This engineered plasmid can then be transformed into Agrobacteria, which will then be allowed to attempt to infect your plant of interest. Instead of a typical infection, the plant will now contain the non-pathogenic T-DNA inserted somewhere within its genome. While targeted disruption of a gene is not possible due to the unpredictability of the area of T-DNA insertion, generating a multitude of lines with T-DNA insertions eventually results in an insertion landing at essentially every desired target (O'Malley et al., 2015). The exception to this is any genes which are essential for survival of the plant cannot contain a T-DNA insert as that would not be compatible with life. In addition to interrupting genes T-DNA inserts can be used to insert desired DNA sequences into the genome of the plant. When engineering plasmids to remove the

pathogenic genes, any relatively small region of DNA can be inserted into the plasmid instead. This DNA could be one or more genes and promoters or any of several selectable markers (Hwang et al., 2017). Provided the inserted genes and promoter are compatible with the host plant, any genes contained within the T-DNA will be active and functional unless silenced by host mechanisms.

MATERIALS AND METHODS

Plant materials and growth

Salk T-DNA insertion mutants were obtained from the *A. thaliana* Biological Resource Center (ABRC). The Salk IDs for each mutant line are respectively: *rpa1c*, Salk_085556; rpa*1e*, Salk_120368. Homozygous mutants for these lines were identified via PCR using gene and T-DNA specific primers (Table 2.1). Prior to germination, seeds were surface sterilized by agitation in a solution of 20% bleach and 0.02% Tween-20 for five minutes. Seeds were then rinsed with autoclaved double distilled water three times, with further agitation used during each rinse. Following sterilization, seeds were sown on nutrient phytoagar plates containing 1x MS salts (PlantMedia, Dublin, Ohio, USA) pH 5.7, 0.05 g/L MES and 1.0% (w/v) phytoagar (PlantMedia, Dublin, Ohio, USA). Seeds were stratified at 4°C for two days in the dark before plates were placed vertically in a growth chamber under cool-white lights at an intensity of 100-150 mmol/m^2/sec at 22°C and a photoperiod of 16hr light/ 8hr dark. After approximately one week of growth, the plants were transferred to soil growth medium (SUNGRO Horticulture, Seba Beach, Canada) with each plant given its own pot. Plants were watered every 3 days with

tap water supplemented by Miracle-Gro® 15-30-15 plant fertilizer at a concentration of 0.45

g/liter (Scotts Miracle-Gro products Inc., Marysville Ohio, USA).

Table 2.1. Primers for sequencing and screening of CRISPR lines and plasmids.

Primer Name	Sequence (5'-3')	Purpose
M13F	GTAAAACGACGGCCAG	PCR Screening (Plasmid) Sequencing
M13R	CAGGAAACAGCTATGAC	Sequencing
hPR93	CTGACTCGGATACTTACGTCACGTC	PCR Screening (T-DNA/plasmid) Sequencing
hPR95	CGGGTCTTAATTAACTCTCTAGACTCACCTAG	Sequencing
hPR96	GTTGAGTAATATATCTACTCACGATTATGG	PCR Screening (T-DNA/plasmid) Sequencing
RPA1C Fwd2	ATGCGGAAAAACTTAACTGAGGTA	PCR Screening (RPA1C ZFKO)
RPA1C Rev2	CTCATAAGGGTTGGGCAGTTTG	PCR Screening (RPA1C ZFKO)
RPA1E Fwd2	TCTGGTGTGTTTCCGGTTCT	PCR Screening (RPA1E ZFKO)
RPA1E Rev1	CTCCATAACTCCCCGTCGTG	PCR Screening (RPA1E ZFKO)
rpa1c, Salk_085556 (F)	GAGAACAACAGCACCACTGATGTA	PCR Screening (RPA1C)
rpa1c, Salk_085556 (R)	GTCTCTAGTTCCTGAGGTTCCA	PCR Screening (RPA1C)
rpa1e, Salk_120368 (F)	TGGTATTGTGTCATCTATCA	PCR Screening (RPA1E)
rpa1e, Salk_120368 (R)	AACCTTACGGATGATATCTT	PCR Screening (RPA1E)
Lba1	TGGTTCACGTAGTGGGCCATCG	PCR Screening (RPA1C and E)

Generation of plasmids for CRISPR-Cas9 engineering

crRNAs for use in the CRISPR-Cas9 mediated modification of Arabidopsis were

designed and chosen by screening the Arabidopsis genome within 300 base pairs of the C-

terminal extension for regions containing the PAM (NGG) necessary for Cas endonuclease

binding. This was done using the online tool CRISPR-P (Lei et al., 2014). These regions were

then checked for problematic features such as hairpin formation and examined using BLASTn

for specificity. Two crRNAs were chosen upstream and two downstream of the ZFM in both

RPA1C and RPA1E (Figure 2.1). crRNAs were purchased from IDT (Integrated DNA

Technologies, Inc. Coralville, Iowa USA). The two strands of each crRNA were ordered separately and upon receipt were rehydrated and mixed at a concentration of 25 μM each. This mixture was incubated at 95ºC for 5 minutes and then allowed to cool to room temperature over the course of 20 minutes to encourage effective annealing.

Table 2.2. Oligos for CRISPR RNA creation. Each pair of forward and reverse oligos were annealed and then inserted into either pEN_Comaira.1 or pEN_Comaira.2. All "1" crRNAs were designed to bind upstream of the zinc finger while each "2" crRNA binds downstream. Bolded regions are overhangs used to anneal crRNAs into Comaira plasmids during cloning.

crRNA Name	Forward Oligo sequence (5'-3')	Reverse Oligo Sequence (5'-3')
RPA1C 1A	**ATTG**TGGCAGACAAGAAGTTCGCA	**AAAC**TGCGAACTTCTTGTCTGCCA
RPA1C 1B	**ATTG**GGTTGAAAATTTCTGCTATA	**AAAC**TATAGCAGAAATTTTCAACC
RPA1C 2A	**ATTG**TGAGAAGTGTGACAAATGCG	**AAAC**CGCATTTGTCACACTTCTCA
RPA1C 2B	**ATTG**GGACCATACAGACCTAACCT	**AAAC**AGGTTAGGTCTGTATGGTCC
RPA1E 1A	**ATTG**GCCGGAAAACTCCCGAGATA	**AAAC**TATCTCGGGAGTTTTCCGGC
RPA1E 1B	**ATTG**GGACGATCACCGTTCATGAT	**AAAC**ATCATGAACGGTGATCGTCC
RPA1E 2A	**ATTG**TTGCGTCACAGCATTCCAGG	**AAAC**CCTGGAATGCTGTGACGCAA
RPA1E 2B	**ATTG**TTATGTGAAGAATGAACACA	**AAAC**TGTGTTCATTCACATAA

```
ctacacagcttgtcctatcatgaacggtgatcgtccatgcagcaaaaaggtaacagacaatggagatggg
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
gatgtgtcgaacaggatagtacttgccactagcaggtacgtcgtttttccattgtctgttacctctaccc
```

Zinc Finger

RPA1E 1B

```
acatggcggtgcgaaaagtgtgataaaagcgttgatgaatgtgactacaggtatatactgcagcttcaga
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
tgtaccgccacgcttttcacactattttcgcaactacttacactgatgtccatatatgacgtcgaagtct
```

Zinc Finger

```
tacaggaccatacagatcttacttgcgtcacagcattccaggaggctggtgaagagataatgggcatatc
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
atgtcctggtatgtctagaatgaacgcagtgtcgtaggtcctccgaccacttctctattacccgtatag
```

RPA1E 2A

Figure 2.1. Example of CRISPR RNA binding and expected cut sites. This pair was used in the creation of E-ZFKO. The red line indicates the approximate expected cut site, three bases upstream of the PAM region (green).

Plasmids pEG_302v2, pEN_Comaira.1, pEN_Comaira.2 and pEn_RC9 (Table 2.3; Figure 2.2) were acquired from the laboratory of Dr. Han Tan at the University of Maine, Orono. pEG_302v2 was transformed into the ccdB tolerant cell line *E. coli* One Shot™ ccdB Survival™ 2 T1R Competent Cell (Thermofisher). Transformation was achieved via heat-shock and was performed following the manufacturer's recommended procedure. Transformants were selected on LB medium containing Kanamycin at 50 mg/L, Carbenicillin at 50 mg/L, and Chloramphenicol at 10 mg/L. pEN_Comaira.1, pEN_Comaira.2 and pEn_RC9 were transformed into chemically competent Mach1 E. coli cells (Thermofisher) following the provided procedure and selected for on LB medium containing Kanamycin at 50 mg/L. Plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol and plasmids were then adjusted to 50 ng/µL for pEG_302v2 and 20 ng/µL for pEN_Comaira.1, pEN_Comaira.2,

and pEn_RC9 using double distilled water. Prior to dilution of the plasmid, DNA quantification was done using a DeNovix DS-11+ spectrophotometer, and the final concentration was confirmed in the same manner.
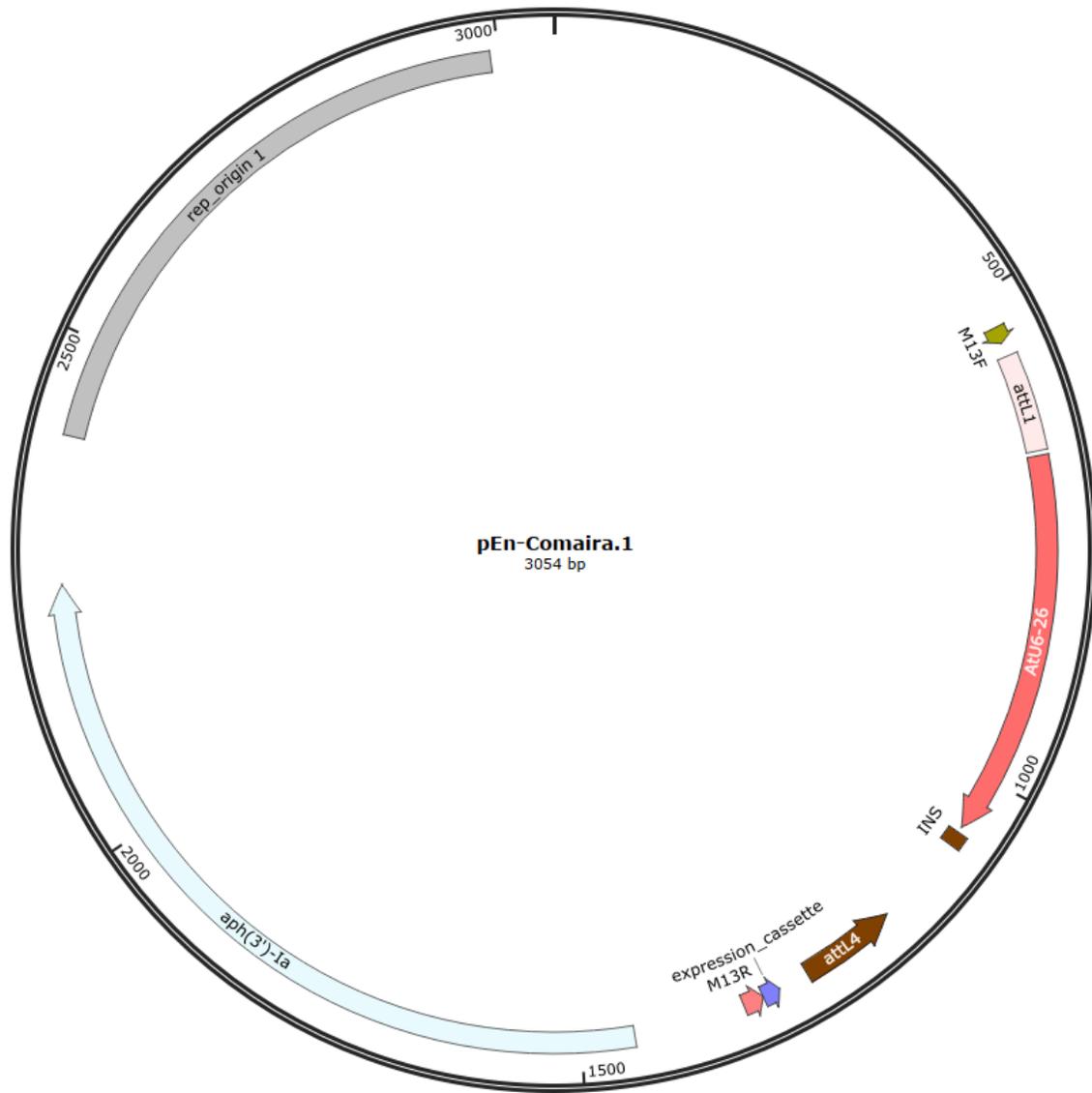
Figure 2.2. Map of pEN_Comaira.1. The Comaira plasmids contain an origin of replication (rep_origin 1), and a kanamycin resistance gene (aph(3')-Ia)- to allow for cloning. The plasmids also contain att sites (attL4 and attL1) for Gateway Cloning. The insertion site for the designed crRNA (INS) is between the att sites and upon insertion the crRNA is fused with the AtU6-26 promoter. Verification of insertion was accomplished using the M13F and M13R primers (Table 2.1).

Assembly of Comaira plasmids was begun by separately digesting 500 ng of

pEn_Comaira.1 or pEN_Comaira.2 with BsaI-HF (NEB R3535S) with the CutSmart buffer.

Digestion was completed overnight at 37ºC, followed by heat inactivation at 65ºC for 20

minutes. Digests were gel purified using a 1% agarose gel with DNA recovered using the

Monarch® DNA Gel Extraction Kit (New England Biolabs). DNA extraction was done

following the manufacturer's recommended procedure and the extracted DNA was quantified

and diluted to 25 ng/µL. Ligation of crRNAs into Comaira plasmids was carried out as follows:

25 ng BsaI-digested plasmid DNA was combined with 1 µL (25µM) annealed crRNA, 400 units

T4 DNA ligase (New England Biolabs), and 1x T4 DNA ligase reaction buffer (New England

Biolabs) in a final volume of 10 µL. All crRNAs targeting sites upstream of the ZFM were

ligated into pEN_Comaira.1 backbones while the downstream-targeting crRNAs were ligated

into pEN_Comaira.2 backbones (Table 2.3). The mixture was incubated overnight at room

temperature, and then transformed into chemically competent Mach1 *E. coli* cells (Thermofisher)

following the provided procedure and selected for on LB medium containing Kanamycin at 50

mg/L.

Plasmids were isolated from Comaira transformants as described previously. The success

of annealing and transformation was confirmed first using PCR, and subsequently via

sequencing. PCR was done using GoTaq Green master mix kit (Promega, Madison, WI)

according to the manufacturer's instructions and with the M13F primer and corresponding

reverse oligo from the inserted crRNA (Table 2.2). PCR was performed using an Applied

Biosystems Proflex PCR system, thermocycler parameters were: 95° C for 2 minutes; 30 cycles

of 95° C for 30 seconds, 56° C for 45 seconds, 72° C for 30 seconds; 72° C for 5 minutes.

Following PCR, results were visualized on a 1% agarose gel. Sequencing was done by

GENEWIZ (Azenta Life Sciences) using the M13R primer (Table 2.2).

After confirmation of crRNA insertion into Comaira plasmids, Gateway Cloning was

used to produce eight unique plasmids, four of which contained crRNAs targeting RPA1C and

four targeting RPA1E. Gateway Cloning using the LR Clonase™ II Plus enzyme (Thermofisher)

was conducted as follows: 20 ng pEN-Comaira.1, 20 ng uL pEN-Comaira.2, 20 ng pEN-RC9.3,

50 ng pEG-302v2 , and 1x LR Clonase II Plus enzyme mix were combined in a final volume of 5

uL, vortexed briefly, and incubated at room temperature overnight. Enzyme was inactivated by

the addition of 1 uL proteinase K (Sigma-Aldrich) and a 10-minute incubation at 37ºC. 5 µL of

reaction mixture was used for transformation into MAX Efficiency™ DH5α Competent Cells

(Thermofisher) following the manufacturer's recommended procedure and selecting for mutants

on LB medium containing 50 mg/L Carbenicillin. Plasmids were isolated following the previous

procedure and were sequenced to confirm successful assembly. Sequencing was done by

GENEWIZ (Azenta Life Sciences) using the hpr93, hpr95, and hpr96 primers (Table 2.2).

Table 2.3. Plasmids used during creation of CRISPR lines. Every plasmid was verified by sequencing.

Plasmid Name	Plasmid Size (bp)	Source
pEN_Comaira.1	3054	Dr. Han Tan (University of Maine, Orono)
pEN_Comaira.2	3161	Dr. Han Tan (University of Maine, Orono)
pEn_RC9	8840	Dr. Han Tan (University of Maine, Orono)
pEG_302v2	11571	Dr. Han Tan (University of Maine, Orono)
pCom1CA	3058	Comaira plasmid generated in this study containing crRNA RPA1C 1A
pCom1CB	3058	Comaira plasmid generated in this study containing crRNA RPA1C 1B
pCom2CA	3165	Comaira plasmid generated in this study containing crRNA RPA1C 2A
pCom2CB	3165	Comaira plasmid generated in this study containing crRNA RPA1C 2B
pCom1EA	3058	Comaira plasmid generated in this study containing crRNA RPA1E 1A
pCom1EB	3058	Comaira plasmid generated in this study containing crRNA RPA1E 1B
pCom2EA	3165	Comaira plasmid generated in this study containing crRNA RPA1E 2A
pCom2EB	3165	Comaira plasmid generated in this study containing crRNA RPA1E 2B
pCAA	16050	pEG302 plasmid generated in this study targeting RPA1C
pCBB	16050	pEG302 plasmid generated in this study targeting RPA1C
pEBA	16050	pEG302 plasmid generated in this study targeting RPA1E

Generation of Agrobacterium for CRISPR-Cas9 genetic engineering

Agrobacterium tumefaciens GV3101 was acquired from the lab of Dr. Estelle Hrabak

(University of New Hampshire) and made electrocompetent via the following procedure: 3 mL

seed culture in LB broth was grown overnight at 28ºC with vigorous shaking. 2.5 mL of seed

culture was added to 250 mL of LB broth and grown at 28ºC with vigorous shaking until an OD$_{600}$ of 0.550 was reached. The culture was chilled on ice then transferred to centrifuge bottles and spun down at 4000 x g for 10 minutes at 4ºC. The supernatant was decanted, and cells were resuspended in 125 mL sterile ice-cold water. Cells were spun down at 4000 x g for 10 minutes at 4ºC, the supernatant was again discarded, and the cells were resuspended in 75 mL sterile ice-cold water. Cells were again spun down at 4000 x g for 10 minutes at 4ºC and following the removal of the supernatant, were resuspended in 7.5 mL ice-cold 10% glycerol. Cells were spun down at 4000 x g for 10 minutes at 4ºC for a final time, followed by removal of supernatant and resuspension in 1 mL sterile ice-cold glycerol. 50 µL aliquots were created and frozen in dry ice before storage at -80ºC.

Transformation of electrocompetent Agrobacteria was conducted as follows: 3 µL of 15 ng/µL plasmid was added to 50 µL of electrocompetent Agrobacteria. Bacterial-DNA mixture was added to a pre-chilled cuvette. Cells were shocked with 1670 volts followed by immediate rescue with 1 mL of SOC growth medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl$_2$, 10 mM MgSO$_4$, 20 mM Glucose). The mixture was transferred to a clean 1.5 mL microfuge tube and incubated at 28ºC with vigorous shaking. Following a 90-minute outgrowth period, selection was done using LB medium plates containing Kanamycin at 50 mg/L and Carbenicillin at 50 mg/L. Transformation into Agrobacteria was confirmed via colony PCR performed using an Applied Biosystems Proflex PCR system, and the primers hpr93 and hpr96 (Table 2.2). To generate the template for PCR, an Agrobacterium colony was touched with a pipette tip which was then swirled in 200 µL of double distilled water. 5 µL of this solution was then used as a template for PCR. Thermocycler parameters were: 95° C for 2

minutes; 30 cycles of 95° C for 30 seconds, 56° C for 45 seconds, 72° C for 80 seconds; 72° C for 5 minutes.

Generation of CRISPR-Cas9 engineered Arabidopsis

Following confirmation of successful transformation, the Agrobacterium were used to infect wild-type (Col-0) Arabidopsis plants. Agrobacteria were grown in LB broth with Kanamycin at 50 mg/L to approximately mid-log phase then centrifuged at 4000 x g for 10 minutes at 4ºC and resuspended in 5% sucrose to an OD_{600} of 0.8. Silwet L-77® (Phytotechnology Laboratories, Shawnee Mission, Kansas) was added to a concentration of 0.05% (500 µL/L). Any existing siliques were removed from flowering Arabidopsis plants that were approximately one month old. The bulk of the plant, including all of the flowers, was then immersed into the bacterial cell slurry for 2-3 seconds with gentle swirling before being placed horizontally, returned to the growth chamber, and covered with a dome for 24 hours. Following this period, the plants were returned to a vertical position and allowed to grow as previously described. The floral dip was repeated 7 days later following the same procedure. After maturation, seeds were collected. Seeds containing the T-DNA insert were selected for using phytoagar plates containing 1x MS salts (PlantMedia, Dublin, Ohio, USA) pH 5.7, 0.05 g/L MES, 1.0% (w/v) phytoagar (PlantMedia, Dublin, Ohio, USA), and 10 mg/L Glufosinate ammonium (Bio-World).

After one week, any plants that grew on the herbicide media were transferred to pots containing soil growth medium and grown as previously described. DNA was isolated from a segment of leaf approximately 1 cm^2 taken from plants that were approximately three weeks old. The leaf piece was placed in 500 µl extraction buffer (0.2 M Tris-HCl, (pH9.0), 0.4 M LiCl, 25

mM EDTA, 1% SDS). The tissue was ground using a micro pestle and, once homogenous, was centrifuged in an Eppendorf microcentrifuge at 14,000 x g for five minutes at room temperature. 350 µl of supernatant was transferred to Eppendorf tubes containing 350 µl isopropanol and mixed by inversion. Tubes were centrifuged at 14,000 x g for ten minutes at room temperature. The pellet was allowed to air dry for twenty minutes and was resuspended in 80 µl double distilled water.

Plants were screened via PCR for both the presence of the T-DNA insertion and for editing of the C-terminal extension. Screening for the T-DNA insertion was completed using primers hpr93 and hpr96 as previously described. Screening for editing of the RPA1C C-terminal extension was done using primers RPA1C Fwd2 and RPA1C Rev2 (Table 2.2). Screening for editing of the RPA1E C-terminal extension was done using primers RPA1E Fwd2 and RPA1E Rev1 (Table 2.2). Both sets of primers span the region of the expected deletion and editing was identified by observing the difference in band size compared to Col-0 on a 1.2% agarose gel. Any plants displaying a reduced band size were sequenced by GENEWIZ using the same primers as used for PCR screening (Table 2.4).

Table 2.4. Zinc finger deletions used in this study. In each case a portion of the genome including the C-terminal zinc finger motif was deleted. C-ZFKO1 was generated using crRNAs RPA1C 1A and RPA1C 2A. C-ZFKO2 was generated using RPA1C 1B and RPA1C 2B. E-ZFKO was generated using RPA1E 1B and RPA1E 2A.

crRNA	Expected Deletion Size (bp)	Actual Deletion Size (bp)
C-ZFKO1	206	202
C-ZFKO2	160	225
E-ZFKO	160	159

Plants with a deleted ZFM were backcrossed with Col-0 Arabidopsis in order to generate plants which lacked the T-DNA insert. This was done to remove the possibility of off-target CRISPR activity in future generations and of the T-DNA insertion altering the phenotype of the plant due to its insertion site. The offspring were screened for both the presence of the ZFM and the inserted T-DNA region as previously described. Any plants that were heterozygous for the removal of the ZFM were allowed to self-fertilize. Seeds were collected from these self-fertilizations and seedlings were screened for the presence of both the ZFM and the T-DNA region. Plants that were homozygous for the removal of the ZFM and also lacked the T-DNA insert were identified and utilized for all future experiments.

DNA damage hypersensitivity assays

For DNA damage hypersensitivity assays, approximately 50 surface-sterilized wild-type or mutant seeds were sown on 3 replicate plates containing 1X MS phytoagar media with or without camptothecin (CPT) (SIGMA, St. Louis, MO, USA) at a concentration of 15nM. Seeds were stratified and then grown as previously described. Plates were photographed, and root length measured following 11 days of growth. Root measurements were of primary root growth from the root junction to the root tip.

For gamma-radiation assays, Arabidopsis seeds or plants were irradiated using a Cs 137 source (Massachusetts Institute of Technology, Cambridge, MA, USA), dose rate 81 rad/minute. Sterilized seeds were imbibed in water at 4ºC for 2 days, irradiated to a dosage of 200 Gy (20,000 rad) and then immediately placed on 1X MS phytoagar plates for germination in the growth chamber. For irradiation of seedlings, plate-grown 5-day-old seedlings were irradiated to a dosage of 100 Gy and then immediately returned to the growth chamber. For both

the experiments using seeds and seedlings, growth continued until an age of 11 days at which point the plants were photographed and primary root growth was measured.

Comet assay

Following imaging and root measurement, 100-150 mg of Arabidopsis tissue from 11-day-old seedlings was frozen in liquid nitrogen and stored at -80º C. The tissue was thawed on ice in 400 µL PBS (160 mM NaCl, 8 mM Na_2HPO_4, 4 mM NaH_2PO_4, pH 7.0) containing 50 mM EDTA (Menke et al., 2001). The tissue was chopped vigorously for 15 seconds with a fresh razor blade (Pourrut et al., 2015). 60 µL of the suspension was mixed with an equal volume of 1% low melting point agarose (Thomas Scientific) at 42º C and pipetted onto a slide precoated with 1% normal melting point agarose. The suspension was then covered with a coverslip and allowed to solidify for five minutes at room temperature. The coverslip was removed, and slides were subjected to lysis via immersion in a high salt buffer (2.5 M $NaCl_2$, 10 mM Tris-HCl, pH 7.5, 100 mM EDTA) for 20 minutes at room temperature, followed by equilibration in 1x TAE for 3x5 minutes on ice. Electrophoresis was conducted in 1x TAE at 23 V (1V/cm) (17mA) for 6 minutes. Following electrophoresis, slides were soaked in 1% Triton for 10 minutes, and then dehydrated for 5 minutes in 70% ethanol followed by 5 minutes in 95% ethanol. Slides were then allowed to fully air dry. 20 µL of 5 ug/mL ethidium bromide was added to the fully dried gels and covered with a coverslip. Slides were imaged using a NIKON A1R - HD Confocal microscope at 20x magnification. Analysis of resulting images was done using the OpenComet program in ImageJ (Gyori et al., 2014).

RESULTS

To better understand the functionality of RPA1 group C paralogs in DNA damage repair, several mutant lines were generated using the CRISPR-Cas9 gene-editing system. These mutant lines were based on wild-type Columbia (Col-0) plants but were edited to remove the ZFM within C-terminal extension of either RPA1C or RPA1E. I hypothesize here that the ZFM motifs are crucial to the functionality of these paralogs during DNA damage repair. This hypothesis was tested by exposing these new mutant lines to the DNA damaging agent camptothecin (CPT) and to gamma radiation. The susceptibility of these lines to DNA damage was compared to that of both Col-0 plants and the null mutant lines SALK_ 085556 (*rpa1c*) and SALK_120368 (*rpa1e*). The response of the null mutant lines to these stressors was documented previously by our lab; so the null mutant lines served as an effective benchmark for the performance of the CRISPR lines (Aklilu et al., 2014). Exposure to CPT inhibits the functionality of DNA Topoisomerase I, resulting in the induction of DSBs in S and G2 phase, while gamma radiation results in oxidative stress during all stages of the cell cycle (Kovács & Keresztes, 2002; L. F. Liu et al., 2000).

CRISPR-generated mutant lines

The CRISPR-Cas9 gene editing system was used to remove the C-terminal ZFM of RPA1C and RPA1E in Arabidopsis. Sequencing of the resulting plants showed that the ZFM was removed along with some sequence on either side of the ZFM. (Figure 2.3). In two of the three mutants the deleted region closely matched the expected deletion but in C-ZFKO2 an additional 65 bases were deleted (Table 2.4). Analysis of the predicted amino acid sequence of the generated ZFKO lines showed that the deletions in both C-ZFKO2 and E-ZFKO were in-frame,

meaning the original stop codon was preserved (Figure 2.4). The deletion in C-ZFKO1 was not

in frame and resulted in a stop codon 20 amino acids after the ZFM (Figure 2.4A)
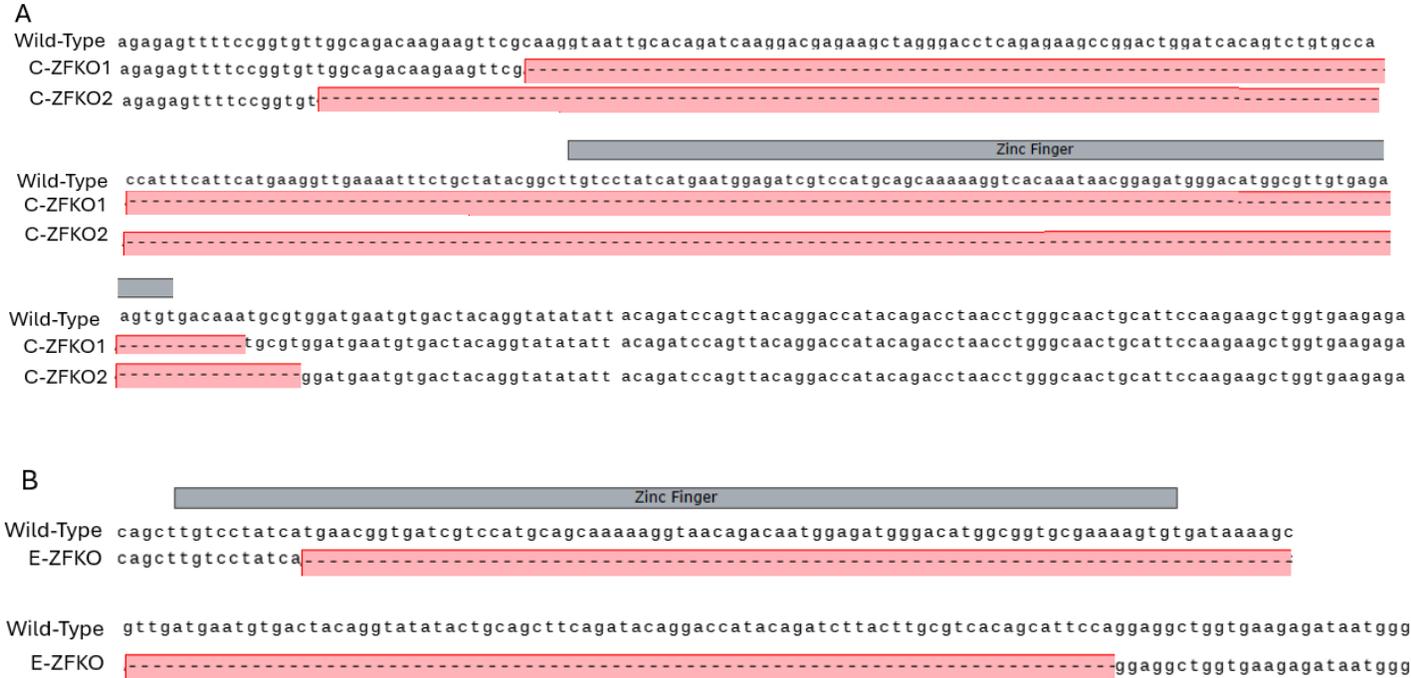

Figure 2.3. Nucleotide sequences of zinc finger knock-out (ZFKO) lines. A) The RPA1C zinc finger knock-out (C-ZFKO) lines 1 and 2 both entirely removed the zinc finger motif (ZFM) as well as a portion of the surrounding region. B) The RPA1E zinc finger knock-out (E-ZFKO) removed the majority of the ZFM as well as 84 downstream bases.
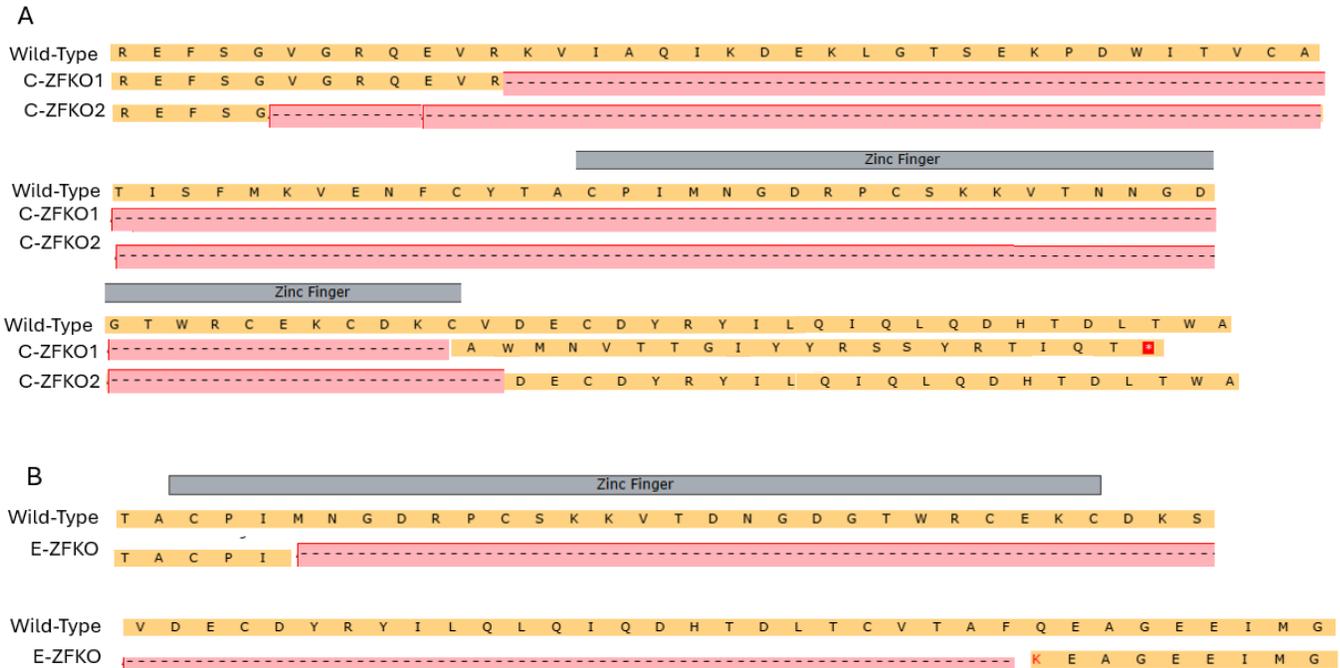
Figure 2.4. Amino acid sequences of zinc finger knock-out (ZFKO) lines. A) The RPA1C zinc finger knock-out (C-ZFKO) line 1 removed the zinc finger motif (ZFM) but resulted in a shifted reading frame and generated a stop codon 20 amino acids after the ZFM. C-ZFKO2 removed the ZFM and maintained the original reading frame. B) The RPA1E zinc finger knock-out (E-ZFKO) removed ZFM and maintained the original reading frame.

Susceptibility of RPA1 mutants to camptothecin-induced DSBs

To test the susceptibility of RPA1 mutant lines to DSBs, seedlings were sown on medium with or without 15nM CPT and grown for 11 days (Figure 2.5). Both *rpa1c* and the CRISPR RPA1C mutants, C-ZFKO1 and C-ZFKO2 displayed hypersensitivity to CPT exposure compared to Col-0 plants but had identical phenotypes to Col-0 in the absence of CPT (Figures 2.6 and 2.7). The *RPA1C* mutant lines also showed failure of cellular division at the meristem, resulting in new meristems forming higher up on the root and a "fuzzy" root tip (Figure 2.5B). This supports previous work which suggests that RPA1C is involved in DSB repair and indicates

that the ZFM in RPA1C is crucial to its ability to repair DSBs during S and G2 phases. Both *rpa1e* and its corresponding ZFM knockout, E-ZFKO, displayed hypersensitivity to CPT but the extent of the hypersensitivity was less than that seen in RPA1C mutant lines (Figures 2.6 and 2.7). These results suggest that, while RPA1E does not have as large of a role in DSB repair as RPA1C does, it nevertheless contributes to the process. Furthermore, the ZFM in RPA1E also appears to be vital for its ability to function during DSB repair in S and G2 phases.

Exposure to CPT significantly reduced the root length of all lines, including Col-0, but the reduction was greatest in *RPA1C* and *RPA1E* mutant lines (Figure 2.8). When comparing the root length of both plants exposed to CPT and those unexposed, a significant difference in root length is seen in the unexposed plants between the Col-0 plants and all *RPA1* mutant lines. This may be an indication of a minor reduced root length phenotype in *RPA1* mutants or simply an effect of the lower LSD value following the inclusion of the data from the CPT exposed plants.

A. -CPT

Col-0 　　　*rpa1c* 　　　C-ZFKO1 　　C-ZFKO2 　　*rpa1e* 　　E-ZFKO

B. +CPT

Col-0 　　　*rpa1c* 　　　C-ZFKO1 　　C-ZFKO2 　　*rpa1e* 　　E-ZFKO
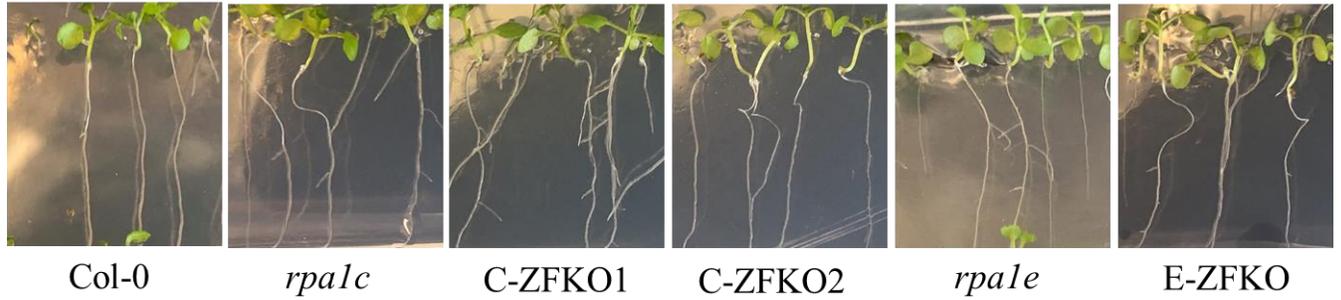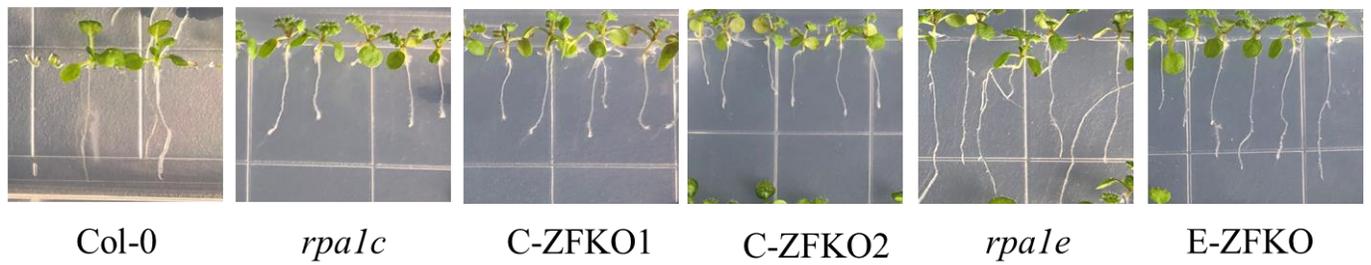
Figure 2.5. Hypersensitivity assay of RPA1 mutants to damage caused by CPT. 11-day-old wild-type (Col-0) and *rpa1* null and ZFKO mutant seedlings grown on MS medium supplemented with either 0 nM (A) or 15 nM CPT (B).
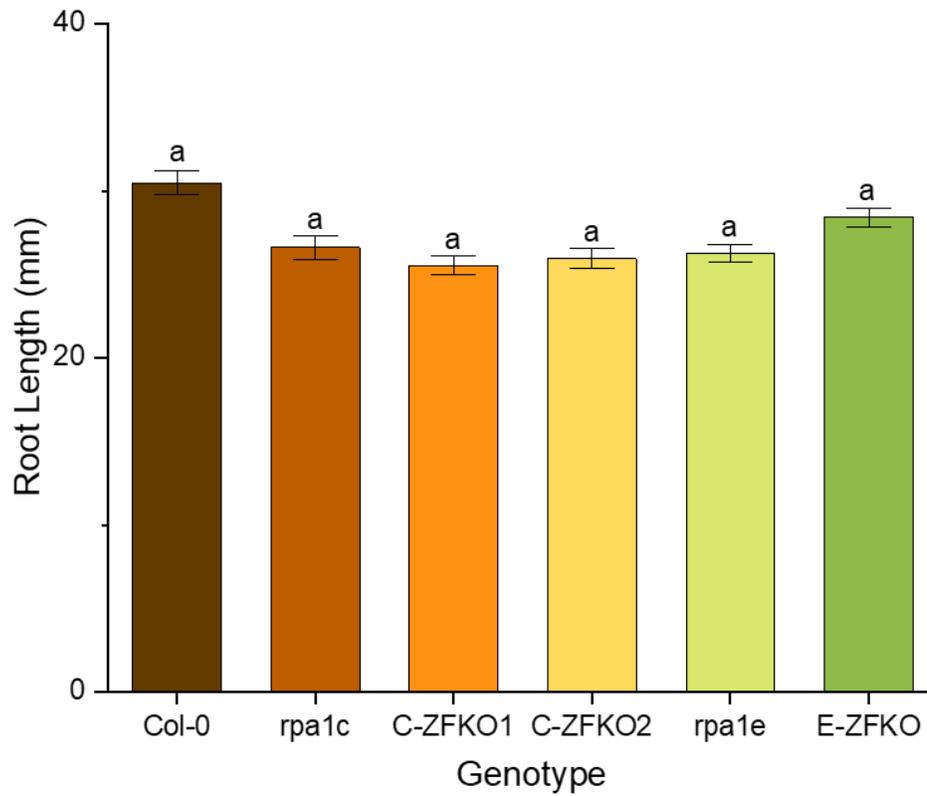
Figure 2.6. Observed root length of CPT hypersensitivity assay control plants. No significant differences were observed between any of the tested lines. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error (n>60).

Figure 2.7. Observed root length of CPT hypersensitivity assay experimental group. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>60).
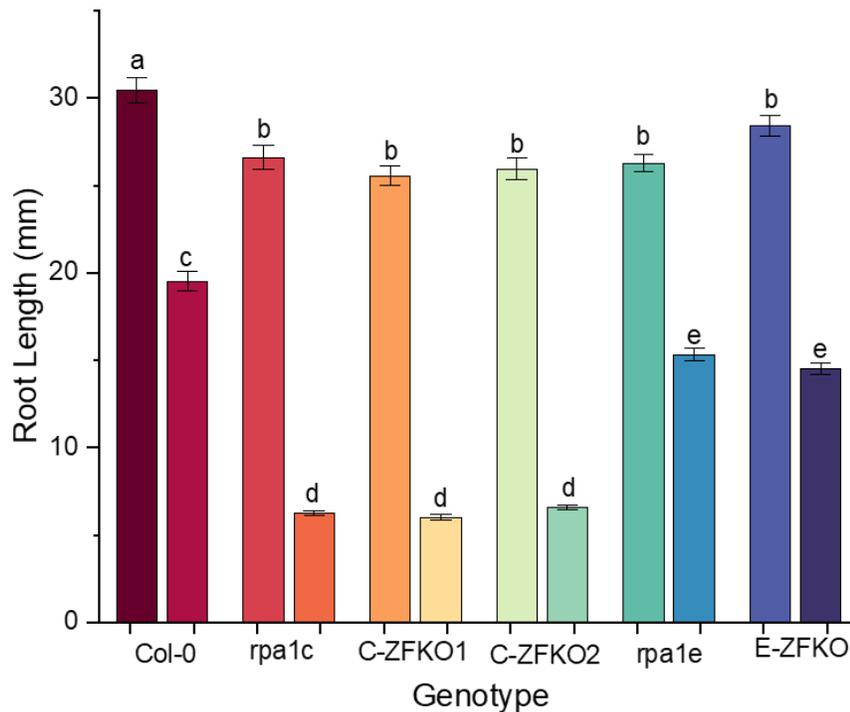
Figure 2.8. Root length measurements of all control (left) and experimental (right) plants used in CPT exposure experiment. Statistical analysis was done using F-test (ANOVA) and LSD at P ≤ 0.05. Error bars denote standard error and bars with different letters indicate significant differences (n>60).

Susceptibility of RPA1 mutant seedlings to gamma radiation

To determine the level of susceptibility of RPA1 mutant lines to DNA damage caused by gamma radiation, seedlings were sown on MS phytoagar and grown under normal conditions until the plants were five days old. The seedlings were exposed to a gamma radiation dosage of 100 Gy, and immediately returned to the growth chamber and allowed to grow for an additional 6 days. Both *rpa1c* and the RPA1C CRISPR mutants, C-ZFKO1 and C-ZFKO2, exhibited hypersensitivity to gamma radiation with near-identical phenotypes (Figures 2.9, 2.10 and 2.11). All lines had significantly reduced root length following exposure to gamma radiation, with average root length dropping by half or more in all cases (Figure 2.12). These data echo what

was found in the CPT experiment and indicate that RPA1C is a crucial component of the DSB

repair machinery and the C-terminal ZFM is critical to the functionality of the protein during the

DNA damage repair process. Additionally, both *rpa1e* and E-ZFKO displayed hypersensitivity

to gamma radiation. While the degree of hypersensitivity was not as extreme as seen with *rpa1c*,

it was still a significant increase in hypersensitivity compared to Col-0. This is consistent with

the results from the CPT assay and suggests that while RPA1E has a role in DSB repair, it is a

lesser role than RPA1C and that the ZFM is critical to the function of the protein during this

process.
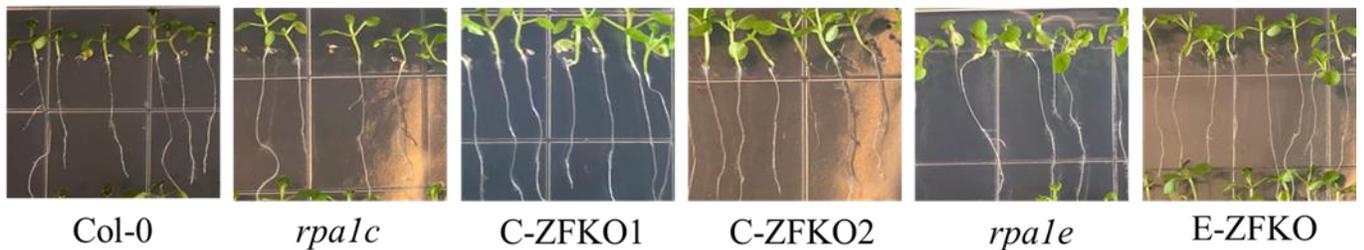
Figure 2.9. Hypersensitivity assay of RPA1 mutant seedlings to damage caused by gamma radiation. 11-day-old wild-type (Col-0) and *rpa1* null and ZFKO mutant seedlings grown on MS medium and treated with 0 (A) or 100 (B) Gy of gamma radiation when 5 days old.

Figure 2.10. Observed root length of seedling gamma radiation hypersensitivity assay control plants. No significant differences were observed between any of the tested lines. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error (n>60).
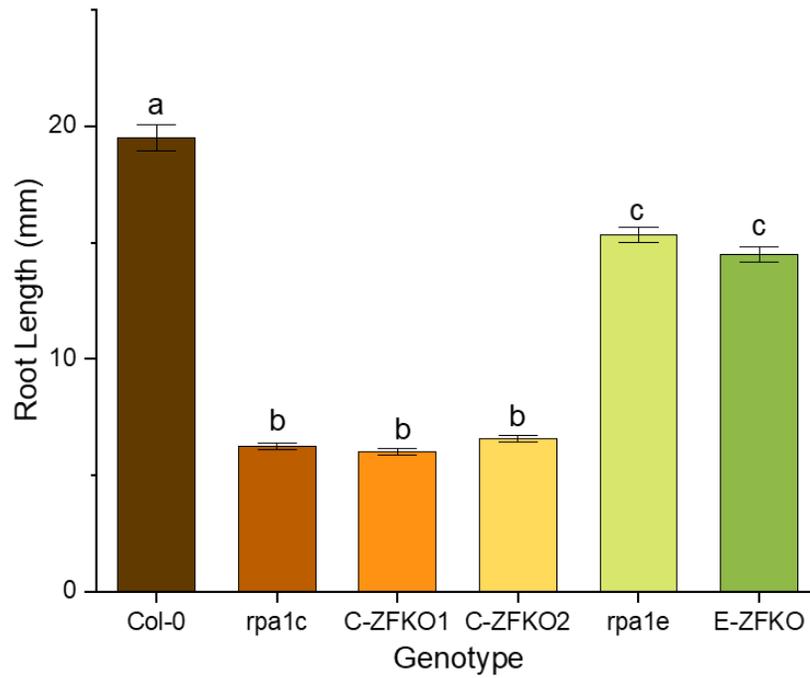
Figure 2.11. Observed root length of seedling gamma radiation hypersensitivity assay experimental group. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>60).
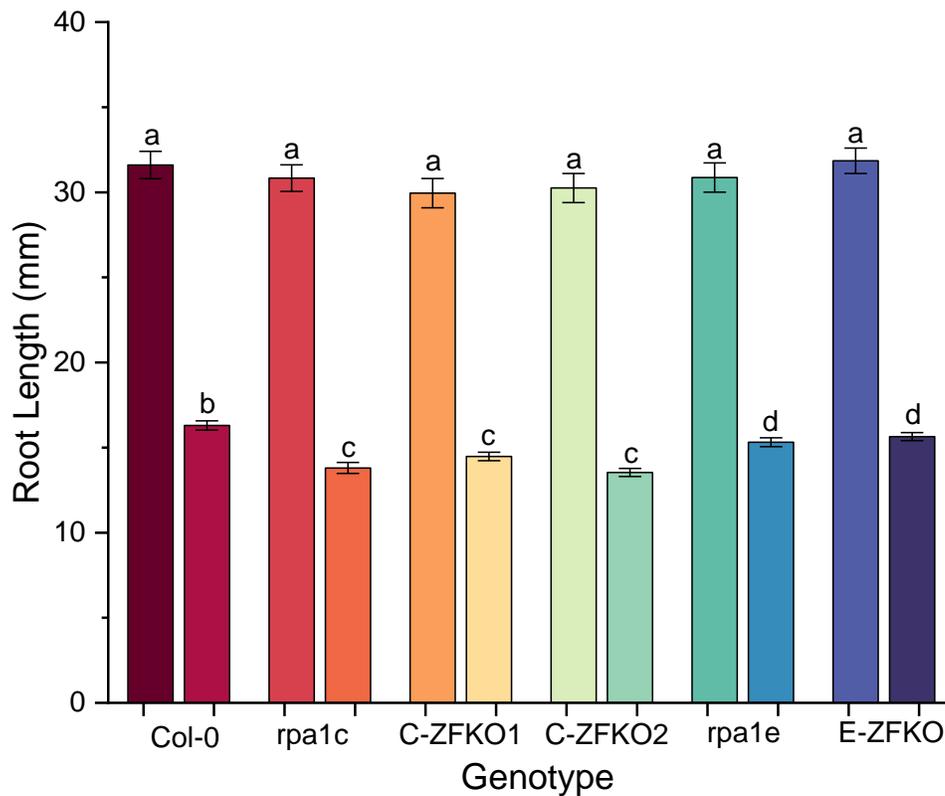
Figure 2.12. Root length measurements of all control (left) and experimental (right) plants exposed to gamma radiation as seedlings. Statistical analysis was done using F-test (ANOVA) and LSD at $P \le 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>60).

Susceptibility of RPA1 mutant seeds to gamma radiation

To thoroughly examine the level of susceptibility of RPA1 mutant lines to DNA damage caused by gamma radiation, seeds were surface sterilized, then imbibed in water and stratified at 4ºC for two days. The seeds were exposed to a gamma radiation dosage of 200 Gy, sown on 1x MS phytoagar, and grown for 11 days. As with the gamma exposure of seedlings, *rpa1c*, C-ZFKO1, and C-ZFKO2 all exhibited hypersensitivity to gamma radiation (Figures 2.13, 2.14 and 2.15). This again reinforces the idea that RPA1C is essential for DSB repair and that the ZFM

with the C-terminal extension of RPA1C is pivotal to the function of RPA1C during DSB repair.

There was no hypersensitivity to gamma radiation observed in either *rpa1e* or E-ZFKO with this

assay, but all lines had significantly reduced root length following exposure to gamma radiation,

with average root length dropping by half or more in all cases (Figure 2.16).

A. Seeds
0 Gy

| Col-0 | *rpa1c* | C-ZFKO1 | C-ZFKO2 | *rpa1e* | E-ZFKO |

B. Seeds
200 Gy

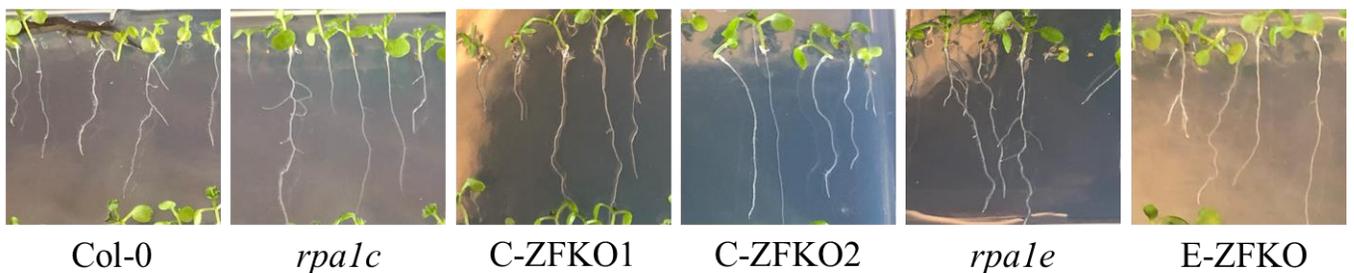| Col-0 | *rpa1c* | C-ZFKO1 | C-ZFKO2 | *rpa1e* | E-ZFKO |
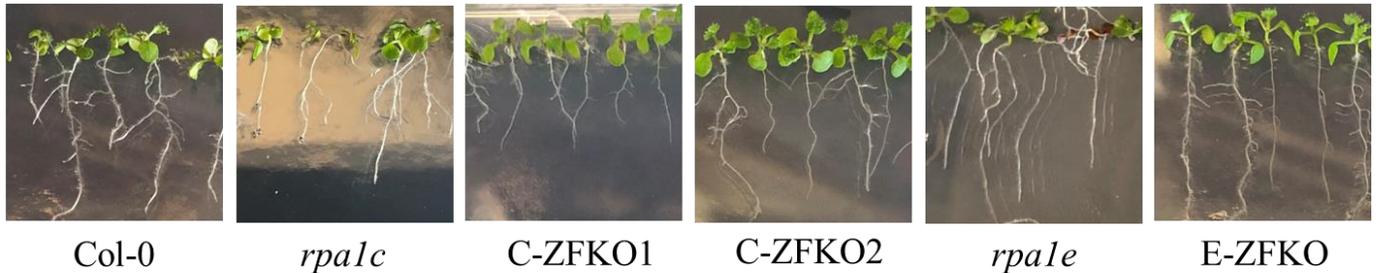
Figure 2.13. Hypersensitivity assay of RPA1 mutant seed to damage caused by gamma radiation. 11-day-old wild-type (Col-0) and *rpa1* null and ZFKO mutant seedlings grown on MS medium and treated with 0 (A) or 200 (B) Gy of gamma radiation as seeds immediately prior to sowing and transfer to the growth chamber.
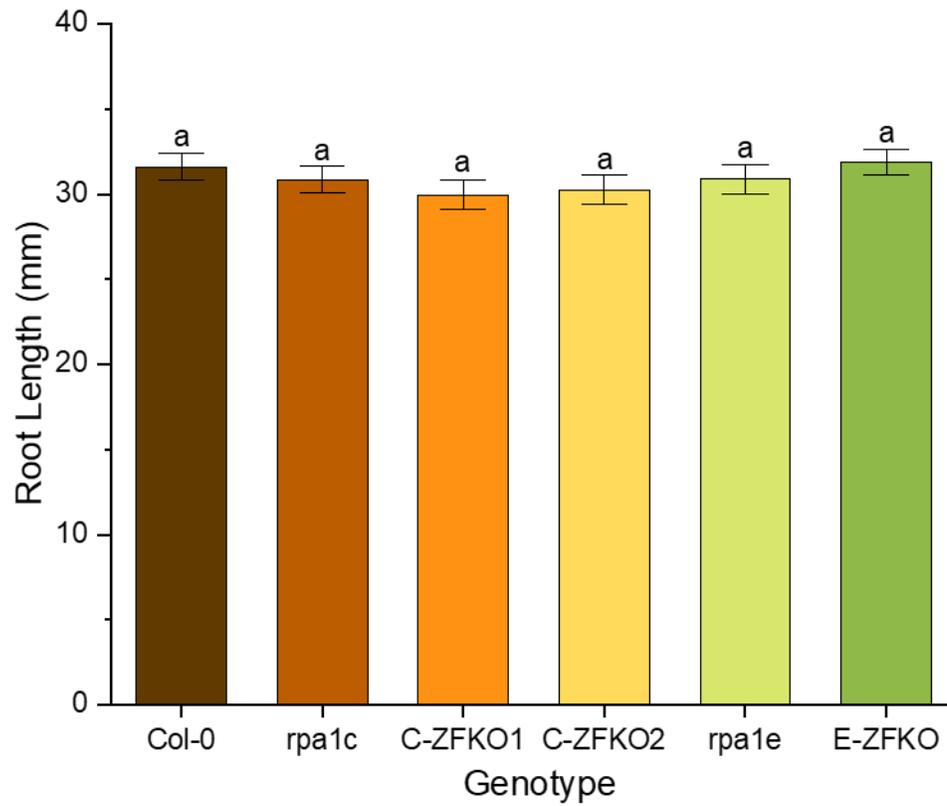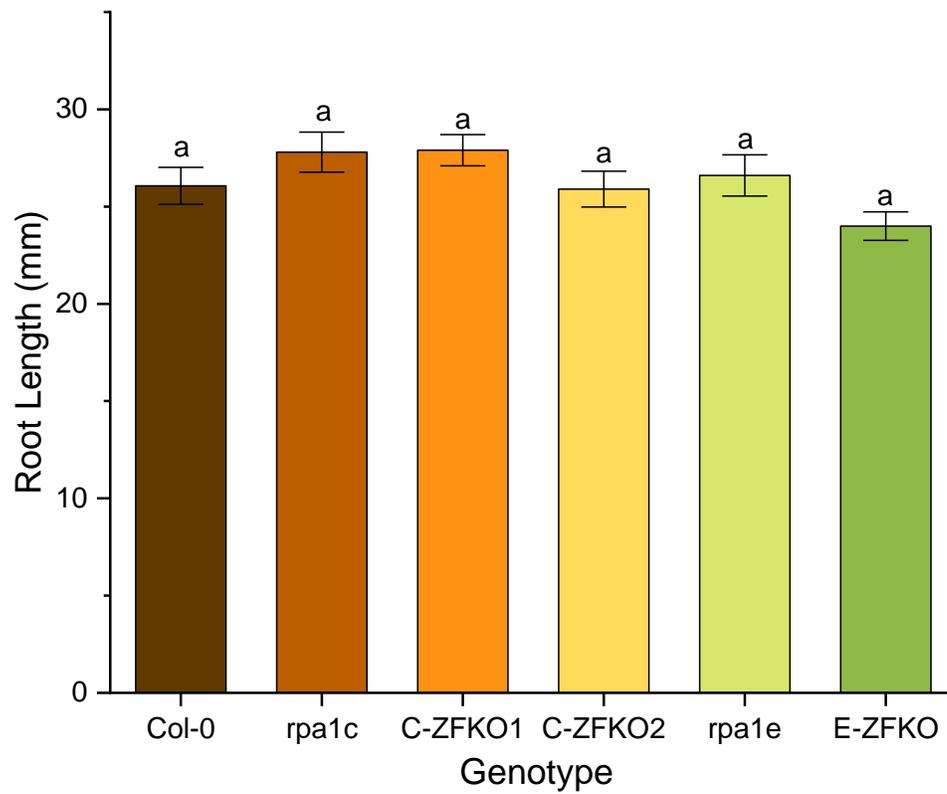
Figure 2.14. Observed root length of seed gamma radiation hypersensitivity assay control plants. There was no significant difference found between any of the tested lines. Statistical analysis was done using F-test (ANOVA) and LSD at $P \le 0.05$. Error bars denote standard error (n>60).
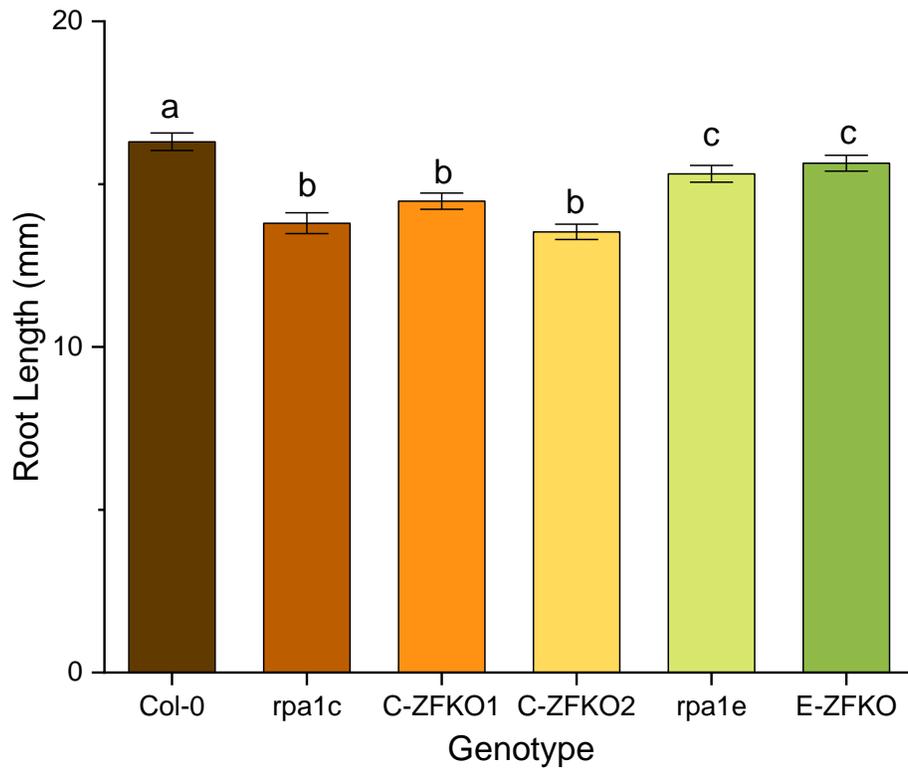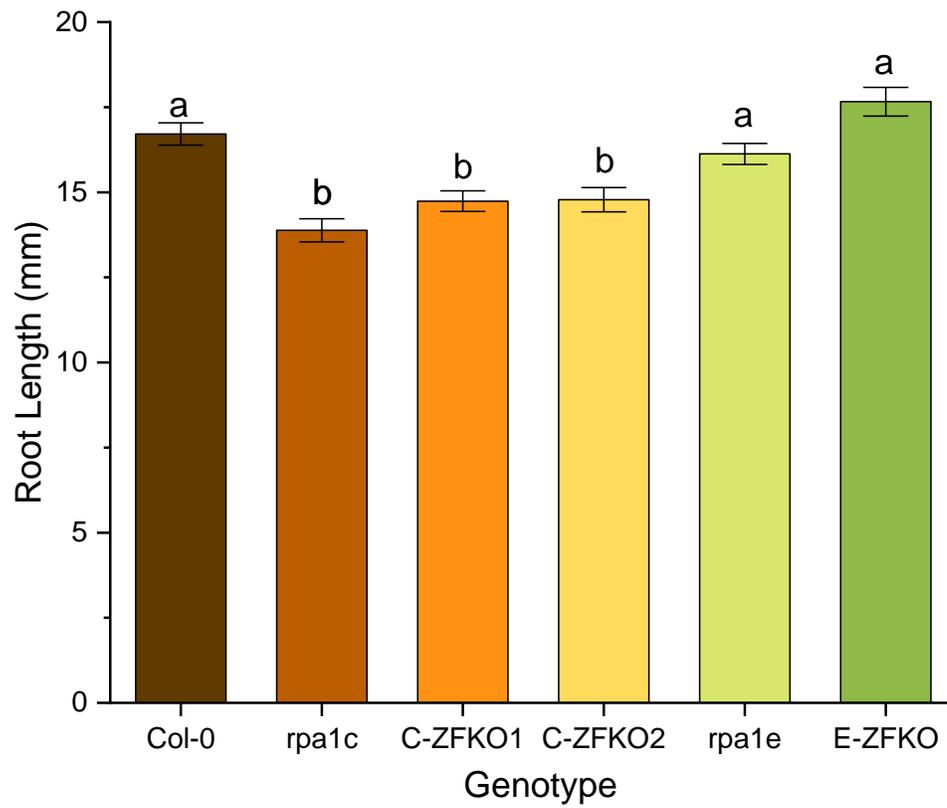
Figure 2.15. Observed root length of seed gamma radiation hypersensitivity assay experimental group plants. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>60).
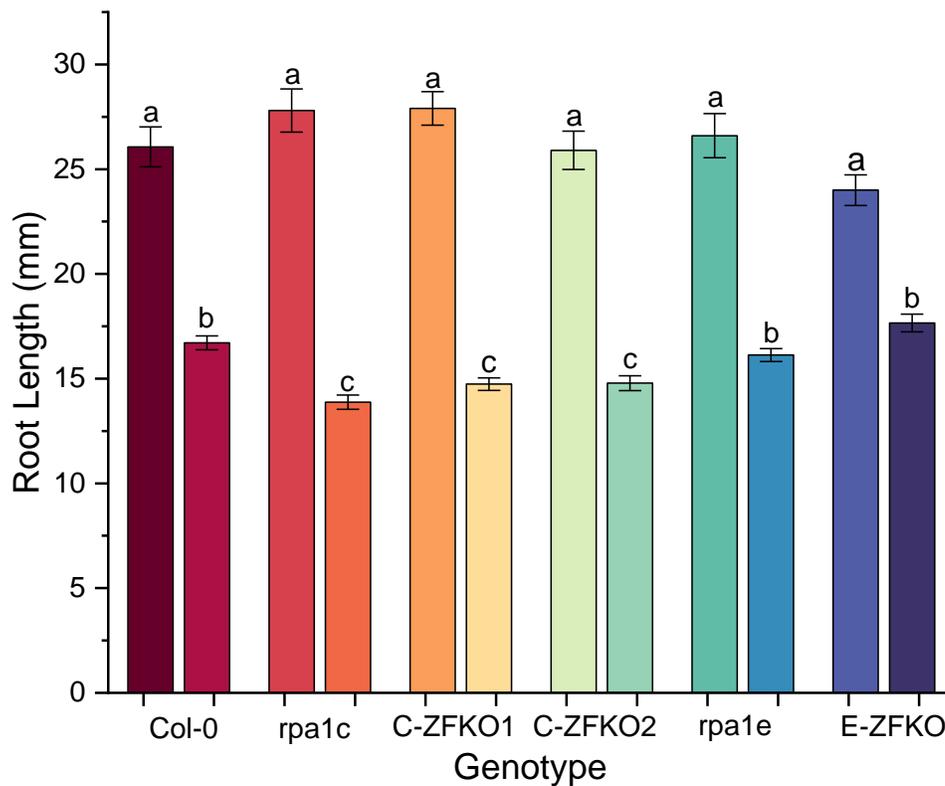
Figure 2.16. Root length measurements of all control (left) and experimental (right) plants exposed to gamma radiation as seeds. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>60)

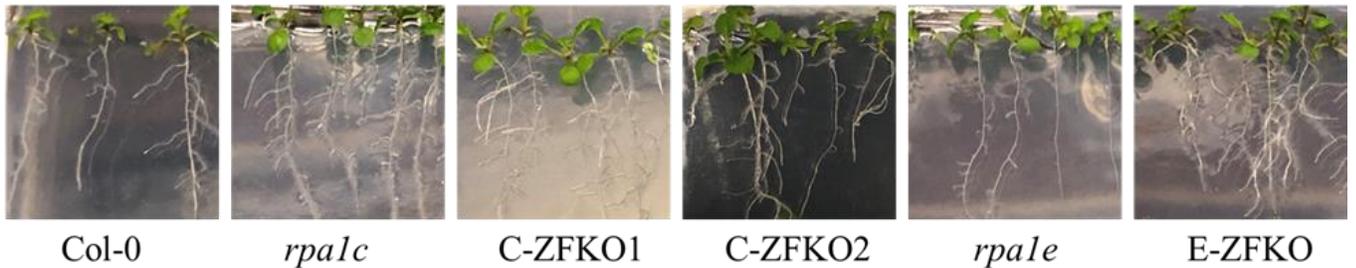## Measurement of DNA damage via comet assay

To directly examine the degree of accumulated DNA damage in group C RPA1 paralog mutants, following root measurement, seedlings exposed to gamma radiation utilized for a comet assay. No significant differences were detected between any of the sampled tissues (Figure 2.17). This demonstrates that while *rpa1c*, *rpa1e*, C-ZFKO1, C-ZFKO2, and E-ZFKO are more susceptible to DNA damage, not all DNA damage machinery is inactivated and therefore the plants are still capable of complete DNA damage repair given enough time to recuperate

following exposure to a damaging agent. For the gamma irradiated seedlings, the six days

following gamma exposure was sufficient to repair the DNA but the phenotype of reduced root

length was still visible.
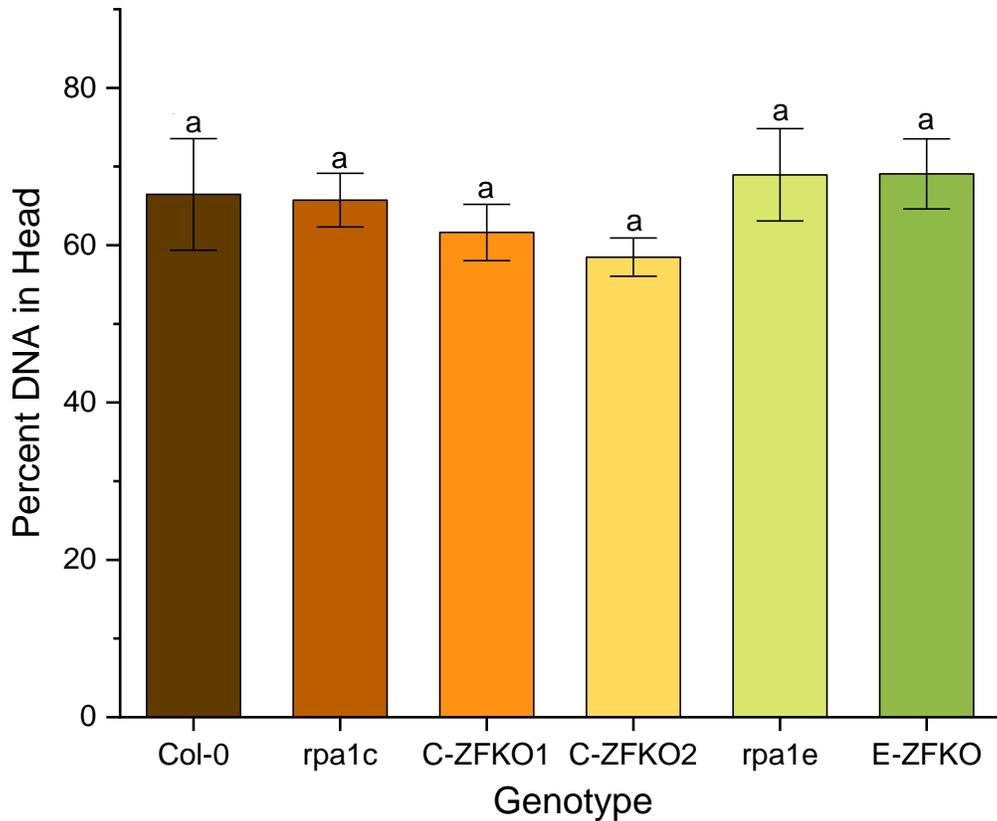
Figure 2.17. Percent DNA in head of comet assay conducted on gamma irradiated seedlings. No significant difference was observed between any of the tested tissues. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error (n>40).

DISCUSSION

In this study, we find that the C-terminal extension of the group C paralogs and specifically the CCHC-type ZFM within the C-terminal extension is crucial to the functionality of the paralogs during DNA damage repair. The CRISPR-Cas9 gene editing system was used to remove the C-terminal ZFM domains from RPA1C or RPA1E. The removal of the ZFM resulted in hypersensitivity not significantly different from the phenotypes of RPA1 null mutants (*rpa1c* and *rpa1e)* in response to damage from both CPT and gamma radiation. Furthermore, the RPA1C null and ZFKO lines both displayed significantly greater susceptibility to DNA damage upon exposure to both CPT and gamma radiation treatment than the RPA1E null and ZFKO lines. These results support previous work suggesting that while both RPA1C and RPA1E are involved in DNA damage repair, RPA1C plays a more significant role than RPA1E. This suggests that RPA1E has a more specialized role in DNA damage repair compared to the more general role of RPA1C.

RPA1C has a larger role in DSB repair than RPA1E

In this study, we hypothesize that group C RPA1 (RPA1C) paralogs function during DNA damage repair requires the specific ZFM contained within their C-terminal extension. To test this hypothesis, a CRISPR-Cas9 approach was utilized to remove the ZFM from the coding region of either RPA1C or RPA1E of *Arabidopsis thaliana.* These mutant lines were evaluated for susceptibility to DNA damage via a root length assay following exposure to either the double-strand break inducing agent CPT or to gamma radiation. *RPA1C* and *RPA1E* null mutants (*rpa1c* and *rpa1e*) as well as Col-0 were used as positive and negative controls, respectively. Previous research found that both *rpa1c* and *rpa1e* are hypersensitive to gamma radiation while

only *rpa1c* was hypersensitive to CPT (Aklilu et al., 2014). In the current study, we find slightly different results upon exposure to CPT, with hypersensitivity found in both *rpa1c* and *rpa1e* lines (Figure 2.7). However, the degree of hypersensitivity found in the *rpa1c* line is significantly greater than that found in *rpa1e* which had a relatively minor, but still significant reduction in root length. In the previous study, there was a decrease in the root length of *rpa1e,* but it was not significant. Given the identical conditions in this study, the difference in results may be due to an increased sample size, resulting in overall more consistent data and greater statistical power.

In addition, we find *rpa1c* and *rpa1e* to be hypersensitive to gamma radiation when exposed as seedlings (Figure 2.11), matching previous work, but only *rpa1c* was hypersensitive when exposed to gamma radiation prior to germination (Figure 2.15). There are several possible explanations for this observed difference in susceptibility. In seeds exposed to gamma radiation. the germination rate is significantly reduced (Babina et al., 2020; Marcu et al., 2013). The seeds used in this study that did not germinate were likely those most negatively impacted by the gamma radiation and had impaired progression. This is supported by previous research showing that DSBs in seeds result in delayed germination (Waterworth et al., 2016). In this study, seeds with delayed germination were not included in the root length assay counts as only germinated seedlings with measurable roots were considered. This method effectively skews the data towards the less affected individuals, possibly explaining the lack of any observable significant differences between the *rpa1e* and Col-0 seeds.

Additionally, seeds, particularly those in dormancy, have two factors that result in greater tolerance to gamma radiation. The primary way in which gamma radiation induces DNA damage is through the ionizing of water to create free radicals, including the reactive oxygen species

(ROS) OH•, which damages the DNA and other molecules (Caplin & Willey, 2018; Esnault et al., 2010). There is much less water present within cells of seeds than those of seedlings, meaning that there is less water available to be ionized and cause damage (Goldberg et al., 1994). Additionally, prior to germination. there is relatively little gene expression, particularly if the seeds are in a dormant state (Le et al., 2010). This lack of expression is controlled in part by epigenetic regulation which makes the genes themselves more difficult to access (Ding et al., 2022). The epigenetic regulation prevents transcription factors from reaching genes but also makes the genes less accessible for ROS, thereby making the genes less likely to be damaged by any free radicals created from gamma radiation. In any case, these data reinforce previous findings that RPA1C has a major role in DSB repair while RPA1E has a lesser, more auxiliary role. This role may be focused on repair only under certain conditions, such as particular phases of the cell cycle, in specific tissues, or in response to a specific damage source.

In the experiment utilizing CPT as a source of DNA damage, the degree of hypersensitivity seen in the *RPA1C* and *RPA1E* mutant lines was greater than that observed following exposure to gamma radiation (Figures 2.8, 2.12, and 2.16). Regardless of whether the gamma radiation was applied to seeds or seedlings, the difference in root length between the Col-0 plants and the *RPA1* mutants was much less drastic in the irradiation experiments but the reduction in root length of the Col-0 plants was greater. This may indicate that RPA is of particular importance in the repair of DSBs. It may also be due, in part, to the method used to determine susceptibility. During the CPT experiment, the roots were in continual contact with the CPT-containing growth medium and CPT specifically affects dividing cells. This resulted in the stalling and failure of the meristem seen in the *RPA1C* mutants (Figure 2.5B). This specific dosage of the roots with a DSB-inducing agent that targets dividing cells would logically result

in reduced root growth in plants with deficient DSB repair machinery. While gamma radiation also induces DSBs and stalls root growth it is a less targeted and more generalized source of damage and was not continually applied to the roots as the CPT was, explaining the lower observed level of hypersensitivity to gamma radiation as a DNA damage source.

When analyzing the CPT experiment, when the control and experimental conditions are combined into a single data set, a slight but significant difference in root length is found between the Col-0 and RPA1 mutant line, even in the absence of CPT (Figure 2.8). This difference could be indicative of a minor root length phenotype under normal growth conditions, but it is more likely to be due to the lowered LSD value upon the inclusion of the experimental data. The lack of root length phenotype is supported by the lack of any difference in root length seen when comparing only those plants grown under normal conditions (Figure 2.6). Additional support to this is lent by the fact that the control plants in the seedlings exposed to gamma radiation, which were grown under the same conditions as the CPT control plants, showed no difference in root length between the Col-0 and RPA1 mutant lines (Figures 2.10 and 2.12).

RPA1 group C paralogs require the C-terminal extension to function during DSB repair.

When observing the susceptibility to DNA damaging agents, in every case, the ZFKO and the respective null mutant lines displayed identical phenotypes. These identical phenotypes were observed for both the assay using CPT as the source of DNA damage as well as the assay utilizing gamma radiation. CPT exposure specifically creates DSBs in actively replicating cells in S and G2 phases while gamma radiation results in oxidative stress during all stages of the cell cycle (Kovács & Keresztes, 2002; L. F. Liu et al., 2000). Null lines and ZFKO lines displaying

identical phenotypes upon exposure to either CPT or gamma radiation strongly indicates that the ZFM contained within the C-terminal extension of the group C RPA1 paralogs is pivotal to the overall functionality of the protein during DNA damage repair throughout the cell cycle.

Furthermore, while CPT causes only DSBs, gamma radiation can result in myriad types of DNA damage including abasic sites, SSBs, and DSBs (Gill et al., 2015; Rastogi et al., 2010). Despite this more generalized DNA damage caused by gamma radiation, there was no statistical difference observed between the ZFKO lines and the respective null mutants. This indicates that the ZFM is likely needed for all activity of RPA1 during DNA damage repair and that it is not a specialized component for repairing DSBs.

The ZFM is crucial during DNA damage repair though both its exact role and mechanism of action are currently unclear. Zinc fingers are capable of binding to a wide variety of amino acid and nucleic acid targets but the zinc finger CCHC motif is most known for its binding to single-stranded nucleic acids and particularly to single-stranded RNA (ssRNA) (Summers, 1991; Wang et al., 2021). Additionally, the region surrounding the ZFM motif in the RPA1 C-terminal extension is both glycine and serine rich, similar to regions found in RNA-binding plant proteins involved in the post-transcriptional regulation of gene expression during cold stress response (Karlson & Imai, 2003; Karlson et al., 2002). This could indicate a regulatory role for the ZFM in DNA damage repair, possibly by stabilizing transcripts of proteins involved in DNA damage repair for translation or interfering with undesirable transcripts.

A potential regulatory role of the ZFM may not be sufficient to explain the susceptibility caused by its deletion. It is possible that, instead of ssRNA binding, the ZFM is instead involved in binding to ssDNA during the DNA damage response. It has previously been found that "Binding Surface I" (BS-I), a subdomain within DBD-F of RPA1, is required for RPA1 to

function during DNA damage repair in humans and yeast, and BS-I is also found in the vast majority group C RPA1 paralogs in plants (Aklilu & Culligan, 2016; Bochkareva et al., 2005; Haring et al., 2008; Xu et al., 2008). While BS-I is known to be required for binding during DSB repair it is possible that other factors are also necessary. The C-terminal extension ZFM may also be involved and may serve to ensure that binding is successful or that RPA stays bound long enough for the damage to be successfully repaired. Given that CCHC ZFMs are known for strong binding capacity to single-stranded nucleic acids, the possibility of the C-terminal extension ZFM being involved in DNA binding does seem likely.

There has been some work indicating that CCHC type ZFMs are also capable of protein binding (Liew et al., 2000; Matthews et al., 2000; Tapia-Ramírez et al., 1997). If that is the case for the C-terminal extension ZFM, then its role is likely to aid in the recruitment of proteins such as Rad51 and Rad54 to the site of DNA damage. This possibility does make sense in light of structural comparisons with RPA1 in humans and yeast which do not contain a C-terminal extension region (Aklilu & Culligan, 2016). If the C-terminal extension was necessary for RPA1 to be able to bind to damaged DNA at all, it would be reasonable to expect it to be conserved across all eukaryotes instead of just plants. Instead, its presence exclusively in plants, combined with its necessity for functionality of RPA, indicates that its role may be specific to plants. While DNA damage response machinery is largely conserved between all eukaryotes, there are notable differences such as the lack of hyper-phosphorylation of RPA2 in response to damage in plants and the plant-specific DNA damage response transcription factor *SOG1* (Marwedel et al., 2003; Yoshiyama et al., 2009; Yoshiyama et al., 2013). It is possible that the C-terminal ZFM interacts with a plant-specific protein or RNA, making it crucial for the plant-specific repair process but unneeded in other eukaryotes. Alternatively, if the C-terminal ZFM does bind to a factor that is

found across all eukaryotes, then it is likely an indication of differentiation in the regulation of DNA damage repair between plants and other eukaryotes. In this case, the ZFM would be fulfilling a role that is typically fulfilled by another protein in non-plant eukaryotes, or possibly by RPA which has undergone some form of post-translational modification.

In addition to its possible role as a binding partner to a nucleic acid or protein partner, there is an additional possible explanation for the complete loss of function of RPA upon the removal of the C-terminal extension ZFM from RPA1. It is possible that the removal of the ZFM and the surrounding region disables RPA directly by preventing it from properly folding and/or assembling into a heterotrimer. We did not provide verification of the expression of the gene or the creation of the final protein product during this experiment and as such, it is possible that the removal of the ZFM resulted in compromised heterotrimer formation. If this is the case, then the observed phenotype would be due to the failure to assemble a heterotrimer rather than simply the loss of the ZFM binding capability. However, this concern can be partially mitigated by considering the other RPA1 paralogs. Upon the removal of the ZFM and surrounding region, RPA1C and RPA1E become very similar in terms of primary protein structure to the other RPA1 paralogs which can fold and form a heterotrimer without any difficulties. This is particularly true for C-ZFKO2 and E-ZFKO, which both maintained the original stop codon (Figure 2.4). Therefore, upon the removal of the ZFM, it is likely that the CRISPR-mutated group C paralogs were still capable of folding correctly and forming a heterotrimer, and that the observed change in phenotype was due to the loss of binding capability of the ZFM rather than the loss of the heterotrimer.

CHAPTER III

DIFFERENTIAL USAGE OF HOMOLOGOUS RECOMBINATION REPAIR STRATEGIES IN *ARABIDOPISIS THALIANA* RPA1C AND RPA1E MUTANTS

ABSTRACT

Repair of genomic double-strand breaks (DSBs) is crucial to prevent disastrous cellular consequences such as loss of genes or chromosomes and cell death. DSBs can be repaired non-conservatively via non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ) or by one of the two homologous recombination repair (HRR) methodologies: synthesis-dependent strand annealing (SDSA) or single-strand annealing (SSA). It is not currently well understood how the cell determines which repair process is utilized but it is at least partially dependent on the stage of the cell cycle as SDSA is only used in G2 or S phase. For the majority of these repair mechanisms, the repair begins following the resecting of the ends of the DSB and the binding of RPA to the resulting single-stranded overhangs. Given its presence on the DNA when the repair process is determined, RPA is a strong candidate for involvement in the decision-making process.

In *Arabidopsis thaliana* (Arabidopsis)*,* there are two RPA1 paralogs that are specifically involved in DNA damage repair: *RPA1C* and *RPA1E*. To further understand the paralogs' unique roles in damage repair, we employed GUS reporter lines that are selectively activated by either SSA or SDSA. These reporter lines were crossed with *RPA1C* and *RPA1E* null lines or zinc finger motif knockout (ZFKO) lines, creating plants with selectively active GUS reporter

systems and mutant RPA1C or RPA1E genes. These plants were used to determine the relative levels of SSA and SDSA in wild-type (Col-0) Arabidopsis compared to plants that had either a completely non-functional RPA1C or RPA1E or to plants that lacked only the C-terminal extension Zinc finger motif (ZFM) in RPA1C or RPA1E.

Our results demonstrate that *rpa1c* and *rpa1e* null mutants showed significantly decreased activity of both SDSA and SSA compared to wild-type Arabidopsis, with *rpa1e* mutants having an approximately twofold reduction while *rpa1c* null mutants displayed a tenfold reduction. The *rpa1c* mutants displayed a much larger decrease in SSA activity than SDSA activity, suggesting a possible preference for repair via SSA following RPA1C binding. Additionally, each ZFKO line displayed identical phenotypes to its respective null mutant. This suggests that both RPA1C and RPA1E are important for both methods of HRR and that the C-terminal extension ZFM is necessary to fulfill this role.

INTRODUCTION

Double-strand break repair

Following the introduction of a double strand break (DSB) in the genome of an organism, there are several possible outcomes. In cases of severe damage, the cell will initiate programmed cell death but in other cases, the damage will instead be repaired. There are three main mechanisms for DSB repair: Non-Homologous End Joining (NHEJ), microhomology-mediated end joining (MMEJ), and homologous recombination repair (HRR; Figure 3.1; Vítor et al., 2020). HRR is further divided into two pathways: synthesis-dependent strand annealing (SDSA)

or single-strand annealing (SSA). Of all the mechanisms of DSB repair, NHEJ is the most used, partially due to availability throughout the cell cycle and few requirements for activation (Kirik et al., 2000). NHEJ requires the factors Ku70 and Ku80 and is most active in G1 phase but can be used during any stage of the cell cycle. NHEJ differs from the other repair mechanisms as it does not require any homology or the resection of the ends of the DNA to repair a DSB while the other repair processes require varying degrees of homology (Han & Yu, 2008; Lieber, 2010).

As the name indicates, MMEJ requires very little homology to repair a DSB, with as few as two bases allowing for DSB repair. For MMEJ to proceed, the ends of the DSB first need to be resectioned by the MRN complex. In mammalian cells, the next step of MMEJ is the binding of 5-hydroxymethylcytosine binding, ESC-specific protein (HMCES). HMCES binding precedes the binding of other proteins including Polymerase theta (Polθ) and poly (ADP-ribose) polymerase 1 (PARP1) which initiate MMEJ (Patterson-Fortin & D'Andrea, 2020; Shukla et al., 2020). MMEJ results in some sequence loss as any sequence between the site of the DSB and the microhomology used for repair is lost.

Sequence is also lost during repair via the HRR repair methodology, SSA. SSA proceeds similarly to MMEJ, with the use of the MRN complex to resect the ends of the DSB, but during SSA, RPA binds in the place of HMCES (Figure 3.2A). Additionally, SSA does not require Polθ or PARP1 and instead requires RAD52 (Bhargava et al., 2016). Another key difference between MMEJ and SSA is that SSA requires a larger region of homology of at least 20 bp (Chang et al., 2017). This greater homology requirement means that SSA typically requires a larger region of

DNA to be resectioned and thereby leads to a greater loss of sequence than MMEJ, sometimes

leading to the loss of kilobases of genomic sequence (Mendez-Dorantes et al., 2018).

In contrast, SDSA is a conservative repair mechanism that does not lead to the loss of any

sequence. Like SSA, it requires MRN to resection the ends of the DNA and for RPA to then bind

to these ends. Additionally, the binding of Rad52 to the RPA-coated single-stranded DNA

(ssDNA) occurs in both SSA and SDSA. However, unlike in SSA, in SDSA Rad51 is also

recruited to the DNA (Serra et al., 2013). Recruitment of Rad51, along with other factors, allows

the regions of ssDNA to invade nearby homologous regions of DNA and use them as a template

(Figure 3.2B). The use of a homologous region allows for perfect repair of the DNA but can

result in alterations to the genome if an imperfect template is used (Puchta & Fauser, 2014).

SDSA is also a far less efficient repair mechanism and as such is used much less commonly than

SSA (Orel et al., 2003).
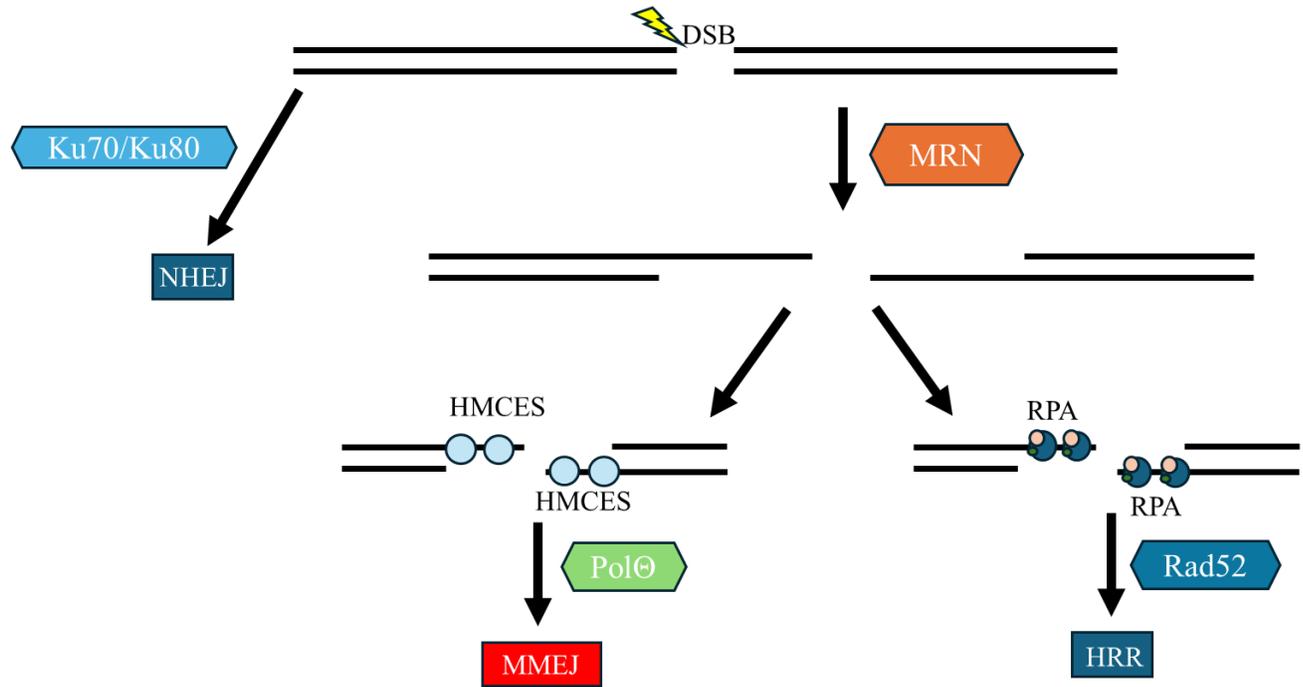
Figure 3.1. Simplified model of the mammalian double-strand break (DSB) repair. DSBs are primarily directly repaired via Ku70/Ku80 mediated non-homologous end joining (NHEJ). Alternatively, following end resection by the Mre11-Rad50-NBS1 (MRN) complex, the 3' OH overhangs can be bound either by Replication Protein A (RPA) or 5-hydroxymethylcytosine binding, ESC-specific protein (HMCES). Following HMCES binding the break is repaired by microhomology-mediated end joining (MMEJ), mediated by polymerase theta (Polθ). Following RPA binding the break is instead repaired by homologous recombination repair (HRR), mediated by RAD52.
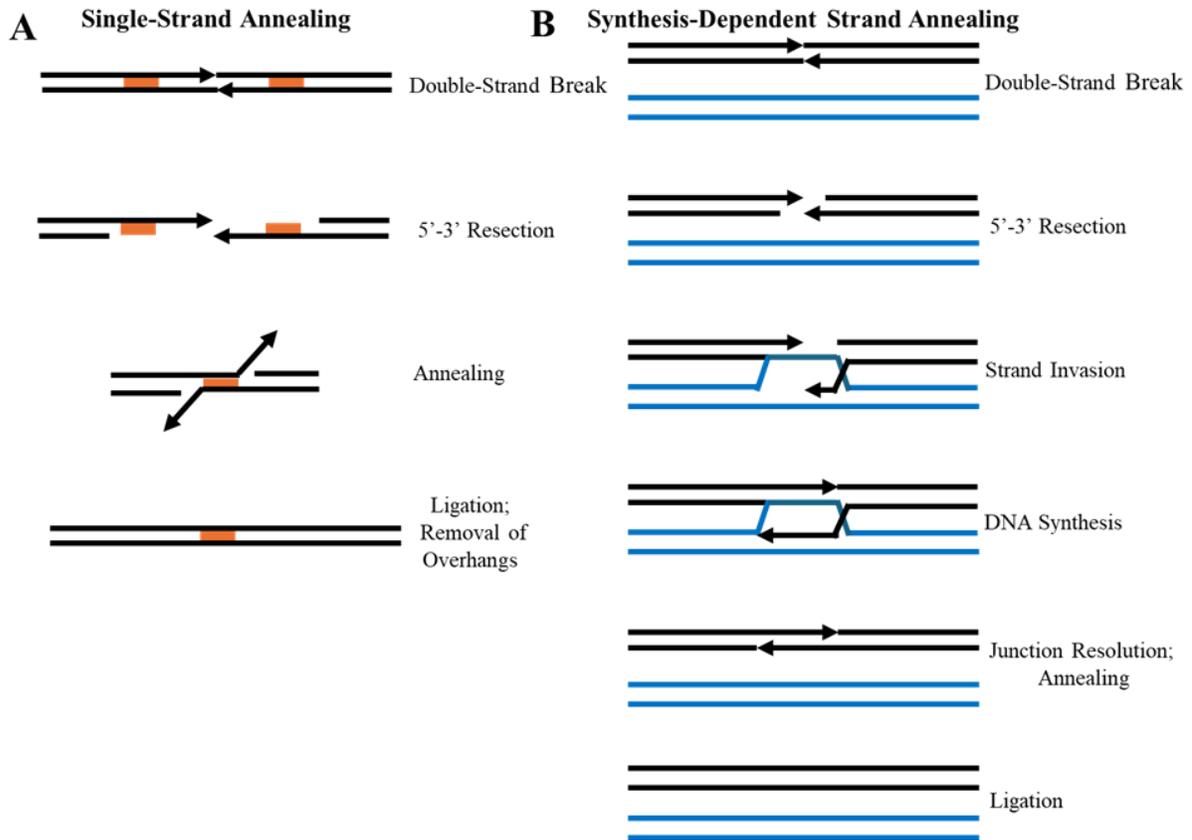
Figure 3.2. Comparison of homologous recombination repair pathways. Both pathways start with the same initial steps but then diverge. A) SSA is a non-conservative repair mechanism with few requirements except two regions of homology (orange) flanking the site of the break. B) SDSA is a conservative repair mechanism but requires a homologous region of DNA to use as a template.

It is not currently well understood how a cell determines which DSB repair mechanism is utilized but the decision is at least partially based on the stage of the cell cycle, with NHEJ dominating in G1 phase while the processes requiring the MRN complex are typically only used during G2 and S phases (Roy, 2014). Additionally, the pathways are at least partially competitive as decreased NHEJ capability leads to increased MMEJ and the binding of RPA inhibits MMEJ (Mateos-Gomez et al., 2017; Osakabe et al., 2010)

Group C RPA1 paralog zinc finger motifs

As previously discussed in Chapters I and II, the RPA1 group C paralogs of Arabidopsis contain an extra region, a C-terminal extension, not found in any of the other paralogs (groups A, B, and D). This C terminal extension is also found in most other plants, but only in group C paralogs. However, there is a discrepancy between plants within the family *Brassicaceae* versus non-Brassica species (Aklilu & Culligan, 2016). In non-Brassica plants, there is only a single group C paralog, while Brassicas, such as Arabidopsis typically have two or more. The major domain found within this C-terminal extension is a CCHC-type zinc finger motif (ZFM), commonly known as a zinc knuckle. In non-Brassica plants, two distinct zinc knuckles are found within the C-terminal extension, but in Brassicas these zinc knuckles are instead distributed among the paralogs C/E (Figure 3.3). The two ZFMs are distinct and are evolutionary maintained, suggesting that the two domains have differential roles.
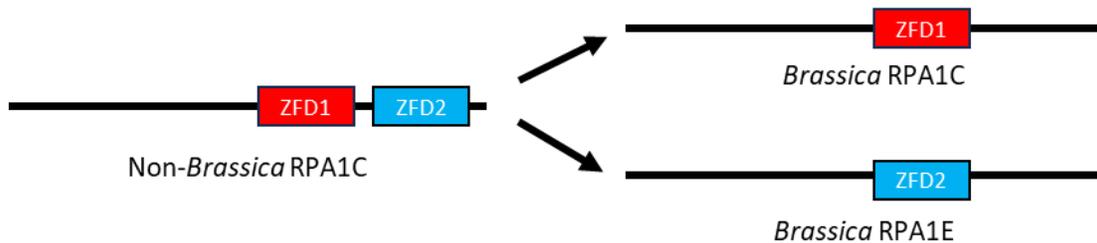
Figure 3.3. Model of distribution of group C RPA1 C-terminal extension zinc finger motifs in Brassicaceae and non-Brassicaceae plants. Plants outside of the family *Brassicaceae* contain only a single group C RPA1 paralog which contains two distinct zinc knuckle domains in their C-terminal extension. In plants within the family *Brassicaceae*, which typically have multiple group C RPA1 paralogs, the zinc knuckle domains are conserved but are distributed among the group C paralogs.

GUS gene reporter system

The *Escherichia coli* β-glucuronidase gene (GUS) is a common reporter gene system that has been used in plants for decades (Jefferson et al., 1987). In this system, 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) is hydrolyzed by GUS to produce 5,5'-dibromo-4,4'-dichloro-indigo, a blue pigment which can be easily observed. This pigment is localized to the cells that have an active GUS gene, and as such the GUS system is often used as a marker to check for the activity of a promoter or transgene in plants.

A slightly alternative use for the GUS system is the use of an inducible GUS gene which is only active under certain conditions. One such system, relies upon GUS genes which are interrupted by a cut site for the restriction endonuclease I-SceI (Orel et al., 2003). These GUS genes are inactive and do not function unless the genes have first been cut by the I-SceI endonuclease and then repaired by a specific repair mechanism (Figure 3.4). There are two separate GUS lines, DGU.US (which is only active after having been repaired via SSA, Figure 3.4A) and IU.GUS (which is only active after having been repaired via SDSA, Figure 3.4B). The GUS gene in the DGU.US line contains a duplicated region which must be removed as part of the resectioning during SSA to produce a functional GUS gene. The GUS gene in the IU.GUS line does not have a duplicate region but instead contains a nonfunctional fragment of the GUS gene in addition to the GUS gene which is interrupted by an I-SceI cut site. The fragment serves as a homologous region to be used as a template for DNA synthesis during SDSA.

In plants containing either one of the GUS genes, as well as a gene to produce the I-SceI endonuclease, appearance of blue coloration is a direct result of the specific repair process occurring. Therefore, this feature allows the indirect measurement of the activity of these repair

processes by observing the degree of coloration in plants containing these elements. This reporter system has been previously used and verified for multiple proteins involved in DNA damage repair including RAD5A, RECQ4A, MUS81, RPA2A, and RPA2B (Liu et al., 2017; Mannuss et al., 2010). The usefulness of the system comes from being able to observe the activity of both SSA and SDSA independently, thereby allowing for comparison of the differential activity of the two systems for a single plant line and how different plant lines compare within the same HRR system. These comparisons can be achieved by breeding the GUS lines into any genetic background desired, thereby enabling the quantification and comparison of HRR in plants with differing mutant backgrounds.
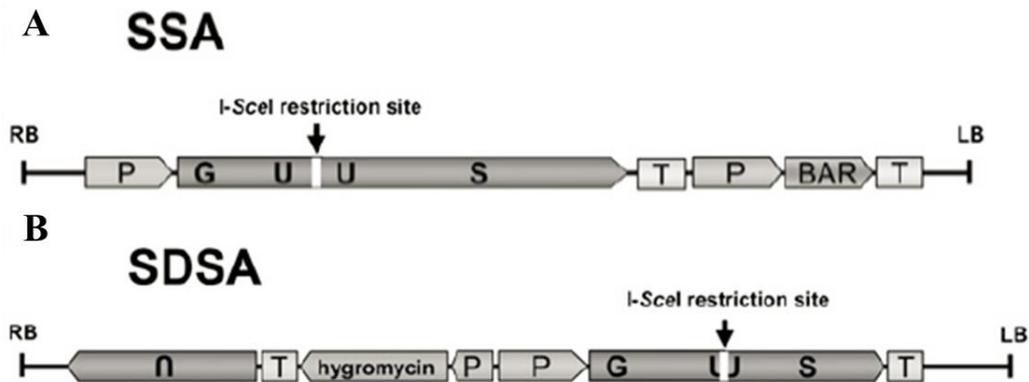
Figure 3.4. Model of the GUS reporter genes used in this study. The DGU.US line (A) is only active after being cut by I-SceI the repaired via single-strand annealing (SSA) while the IU.GUS line (B) is instead activated following repair via synthesis-dependent strand annealing (SDSA). Both GUS genes are fused to a 35S promoter (P) and NOS terminator (T). The DGU.US line contains a phosphinothricin resistance gene (BAR) fused to a 35S promoter and a 35S terminator while the IU.GUS line contains a hygromycin resistance gene fused to a NOS promoter and a NOS terminator. Both gene systems are contained within a T-DNA region bounded by the right border (RB) and left border (LB) regions (Mannuss et al., 2010).

MATERIALS AND METHODS

Plant materials and growth

Plant lines carrying IU.GUS, DGU.US and I-SceI were acquired courtesy of H. Puchta (KIT Germany). Lines were confirmed as homozygous via gene-specific primers (Table 3.1). All three acquired lines were crossed with the following Arabidopsis lines: *rpa1c*, *rpa1e*, C-ZFKO1, E-ZFKO and the F1 offspring were allowed to self-fertilize to produce F2 plants homozygous for a mutant *RPA1C* or *RPA1E* background and either I-SceI or a GUS reporter. Finally, crosses between F2 lines yielded plants heterozygous for both I-SceI and one GUS reporter line in a homozygous *RPA1C* or *RPA1E* mutant background (Figure 3.5). For a control, plants with a GUS reporter gene and with I-SceI were crossed to produce lines that had a wild-type background and were heterozygous for both I-SceI and one GUS reporter. These lines served as controls to determine how the level of HRR activity was affected in plants with mutant *RPA1C* or *RPA1E* backgrounds. For all backgrounds, the final cross was conducted with the I-SceI containing plant as the female. After each cross, except for the final one, the resulting plants were screened with gene-specific primers (Table 3.1) to screen for the desired genotype.

Table 3.1. Primers for sequencing and screening of GUS lines.

Primer Name	Sequence (5'-3')	Purpose
rpa1c, Salk_085556 (F)	GAGAACAACAGCACCACTGATGTA	PCR Screening (RPA1C)
rpa1c, Salk_085556 (R)	GTCTCTAGTTCCTGAGGTTCCA	PCR Screening (RPA1C)
rpa1e, Salk_120368 (F)	TGGTATTGTGTCATCTATCA	PCR Screening (RPA1E)
rpa1e, Salk_120368 (R)	AACCTTACGGATGATATCTT	PCR Screening (RPA1E)
Lba1	TGGTTCACGTAGTGGGCCATCG	PCR Screening (RPA1C and E)
SSA gFW	TCTCCAATTTGCATACCTCATACC	PCR Screening (DGU.US)
TGIUS-RB5	GTTGCGCAGCCTGAATGGC	PCR Screening (DGU.US)
SSA gRV	TCAGCTTTGTCATTGTGGGATAA	PCR Screening (DGU.US)
IU8 gFW	ACTTATTTGCGTGTCTGTTACTT	PCR Screening (IU.GUS)
GABI-LB	TTGGACGTGAATGTAGACAC	PCR Screening (IU.GUS)
IU8 gRV	GAAATCGATTACTAGCCACCACTC	PCR Screening (IU.GUS)
2S-I-SceI#19 gFW	GGCTTGACAAAACCGAAATAC	PCR Screening (I-SceI)
2S-I-SceI#19 gRV	AAGTGGGGAGTGAAGGACATA	PCR Screening (I-SceI)
UFW TGIUS LB4	CTCACGTGTTGAGCATATAAG	PCR Screening (I-SceI)
RPA1C Fwd2	ATGCGGAAAAACTTAACTGAGGTA	PCR Screening (RPA1C ZFKO)
RPA1C Rev2	CTCATAAGGGTTGGGCAGTTTG	PCR Screening (RPA1C ZFKO)
RPA1E Fwd2	TCTGGTGTGTTTCCGGTTCT	PCR Screening (RPA1E ZFKO)
RPA1E Rev1	CTCCATAACTCCCCGTCGTG	PCR Screening (RPA1E ZFKO)
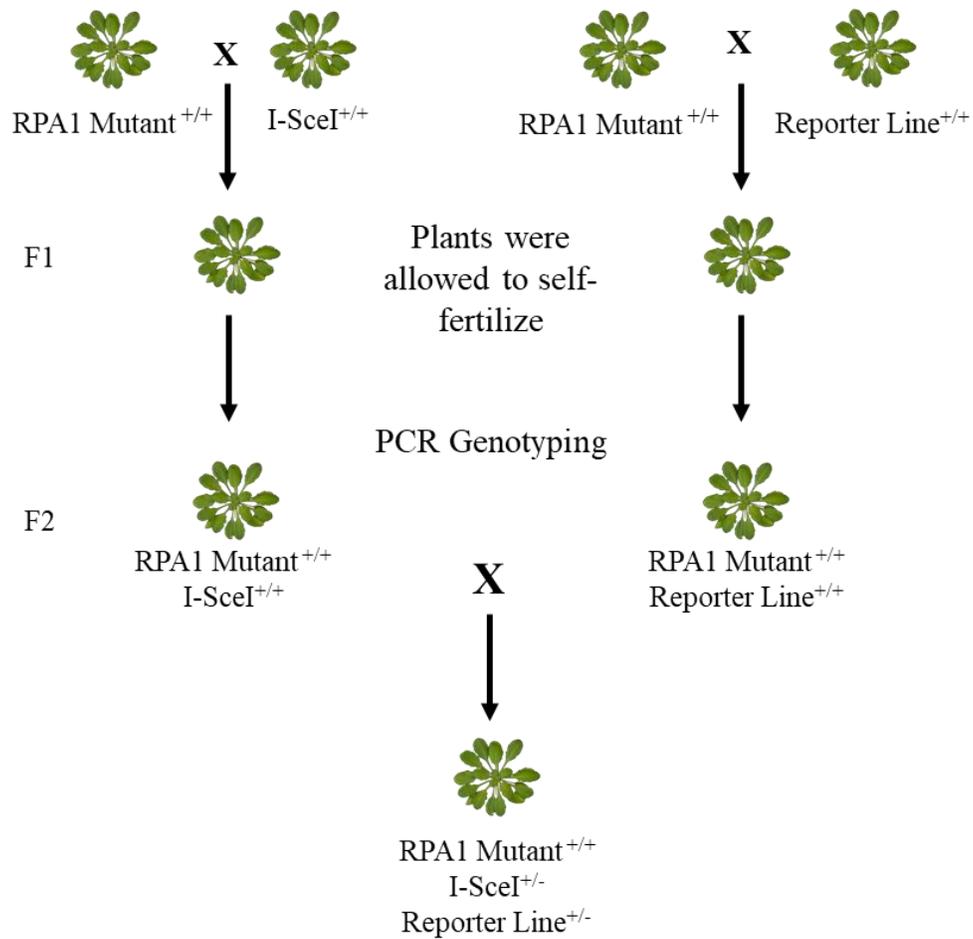
Figure 3.5. Model of the breeding strategy used to generate GUS lines with mutant RPA1 backgrounds. After each generation all plants were screened via PCR for the desired genotype.

Seeds harvested from the final round of crossing were presumed to be homozygous for the *RPA1C* or *RPA1E* background (or homozygous Col-0 for the control lines) and heterozygous for both I-SceI and one of the GUS reporters. These lines were not further propagated and were utilized for GUS assays prior to reaching sufficient growth for harvesting tissue for DNA isolation so PCR genotyping was not conducted. All generated lines, as well as Col-0 seeds, were surface sterilized by agitation in a solution of 20% bleach and 0.02% Tween-20 for five minutes. Seeds were then rinsed with autoclaved double distilled water three times, with further agitation

used during each rinse. Following sterilization, seeds were sown on nutrient phytoagar plates containing 1x MS salts (PlantMedia, Dublin, Ohio, USA) pH 5.7, 0.05 g/L MES and 1.0% (w/v) phytoagar (PlantMedia, Dublin, Ohio, USA) which were supplemented with either 40 mg/L Hygromycin-B (Thermofisher) or 10 mg/L glufosinate ammonium (bioWORLD, Dublin, Ohio, USA).

All DGU.US lines were sown on medium containing glufosinate ammonium, and all IU.GUS lines were sown on medium containing hygromycin. These lines contain resistance genes to these specific chemicals, so this methodology prevented the growth and analysis of any failed crosses that produced seeds following self-fertilization of the I-SceI plant. I-SceI lines do not have any relevant resistance genes and therefore would not be capable of growth on the medium used. Additionally, Col-0 plants were grown on both selections to serve as a control for the efficacy of the chemicals. Seeds were stratified at 4°C for two days in the dark before the plates were placed vertically in a growth chamber under cool-white lights at an intensity of 100-150 mmol/m2/sec at 22°C, and a photoperiod of 16hr light/ 8hr dark for two weeks.

GUS assay

Prior to application of the X-Gluc reagent, plants were fixed in cold 90% acetone for 20 minutes. Following fixation, plants were incubated in 50 mM $NaPO_4$, pH 7.2, 0.5 mM $K_3Fe(CN)_6$, 0.5 mM $K_4Fe(CN)_6$, 10 mM EDTA, 0.01% Trition X-100, and 2 mM X-Gluc (Gold Biotechnology, St. Louis, MO, USA) overnight at 37°C. Plants were decolorized in 95% ethanol for 3x30 minutes. Plants were directly imaged under an Olympus SZ51 light microscope. Quantitative evaluation of GUS staining was done by examining leaves of at least four plants per genotype per trial. Each individual instance of coloration was counted as a single "speck".

Results were averaged to find the number of specks per leaf for each line. Values were compared via one-way ANOVA followed by LSD to determine significant differences between groups.

Results

Previous work has indicated that Arabidopsis deficient in RPA1C group C paralogs are susceptible to DNA damage, with *rpa1c* lines being particularly susceptible to DSBs from CPT exposure (Aklilu et al., 2014). However, the mechanism behind this susceptibility is currently unknown. Given that RPA is present on the DNA just prior to the initiation of DSB repair, it has been hypothesized that RPA is partially responsible for determining the repair mechanism used (Binz et al., 2004). Additionally, the two distinct ZFMs found in the group C paralogs of *Brassica* plants suggest possible differential function. Therefore, we hypothesize that Arabidopsis RPA1C and RPA1E have differential usage during the HRR of DSBs. To test this hypothesis, we utilized two distinct GUS reporter gene systems. Each reporter can only be activated by first being cut by the endonuclease I-SceI and then being repaired specifically via SSA or SDSA (Mannuss et al., 2010; Orel et al., 2003). A single GUS reporter gene and I-SceI gene were introduced into the mutant *RPA1C* or *RPA1E* backgrounds: either T-DNA null mutants which have been studied previously in our lab, or mutants with the C-terminal extension ZFM removed using CRISPR technology as described in Chapter II. The hybridized plants were observed to determine the amount of GUS activity and therefore the usage of each method of HRR in each line.

Relative level of single-strand annealing using DGU.US construct

Two-week-old seedlings assayed for GUS activity from the DGU.US construct (Figure 3.6). All lines containing an *RPA1C* or *RPA1E* mutant background showed less SSA activity than the wild-type background. Compared to Col-0, both the *rpa1c* and the C-ZFKO1 background showed a ten-fold reduction in SSA activity, while both *rpa1e* and E-ZFKO showed an approximate two-fold reduction (Figures 3.7 and 3.8). This indicates that both RPA1C and RPA1E are involved in DSB repair using SSA and reinforces previous work which showed that RPA1C has a larger role in DNA damage repair than RPA1E. It may also indicate a larger role in SSA for RPA1C than RPA1E.

| Col-0 | *rpa1c* | C-ZFKO | *rpa1e* | E-ZFKO |

Figure 3.6. Representative images of GUS activity in single-strand annealing GUS assay. The level of activity was quantified by counting and averaging the number of individual specks per leaf.

Figure 3.7. Single-strand annealing (SSA) activity of two-week-old seedlings. The level of activity was quantified by counting and averaging the number of individual specks per leaf. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>40).
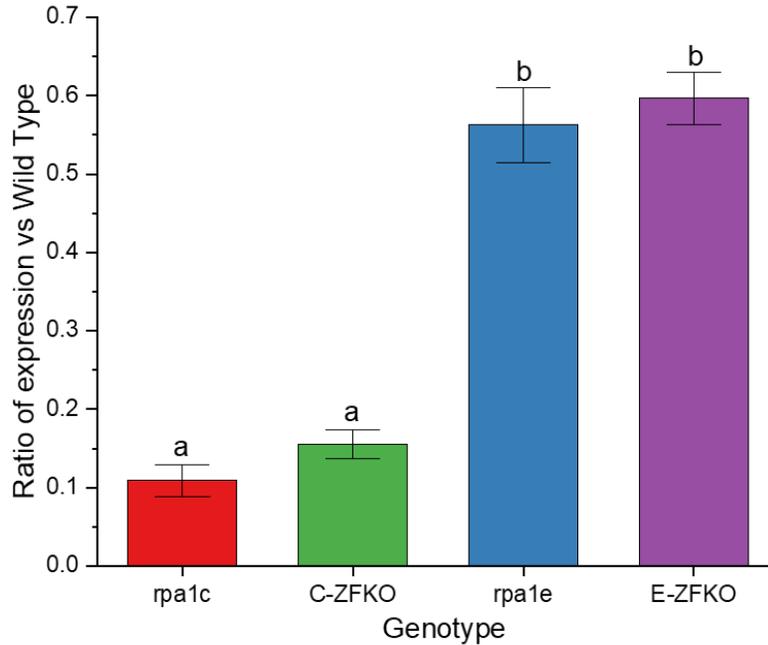
Figure 3.8. Single-strand annealing (SSA) activity of *Arabidopsis thaliana* mutant lines as a fraction of wild-type expression. *RPA1C* mutant lines displayed a much lower fraction of activity than *RPA1E* mutant lines. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>40).

Relative level of synthesis-dependent strand annealing using IU.GUS construct

Two-week-old seedlings were assayed for GUS activity from the IU.GUS reporter (Figure 3.9). All lines containing an RPA1 group C mutant background showed a significant reduction in SDSA activity compared to that in the Col-0 background (Figure 3.10). However, no distinction could be made between plants with a mutant *RPA1C* gene and a mutant *RPA1E* gene. All tested lines showed an approximate two-fold reduction in SDSA compared to Col-0 (Figure 3.11). This suggests that RPA1C and RPA1E may be of equal importance for DNA damage repair via SDSA despite the larger overall role of RPA1C in DSB repair.
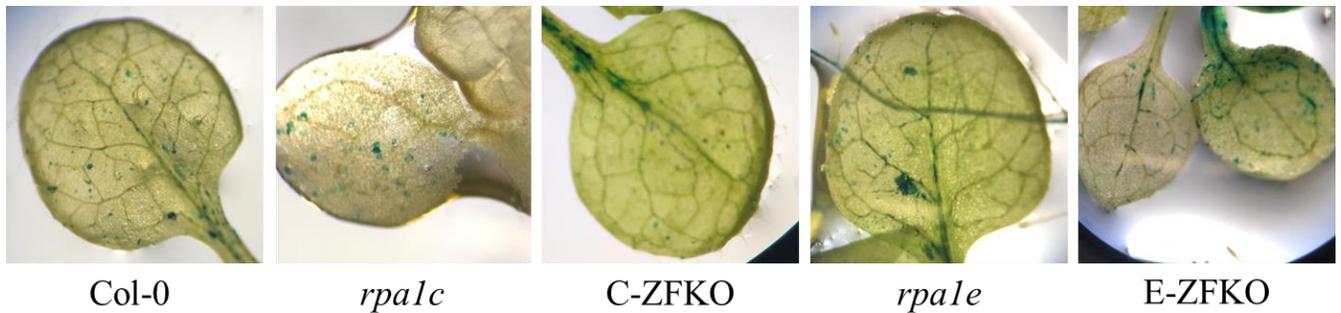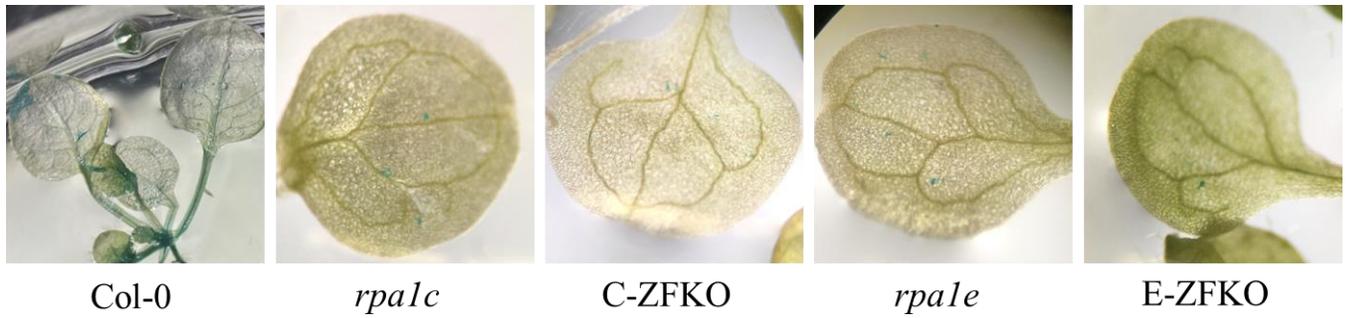
Figure 3.9. Representative images of GUS activity in synthesis-dependent strand annealing GUS assay. The level of activity was quantified by counting and averaging the number of individual specks per leaf.
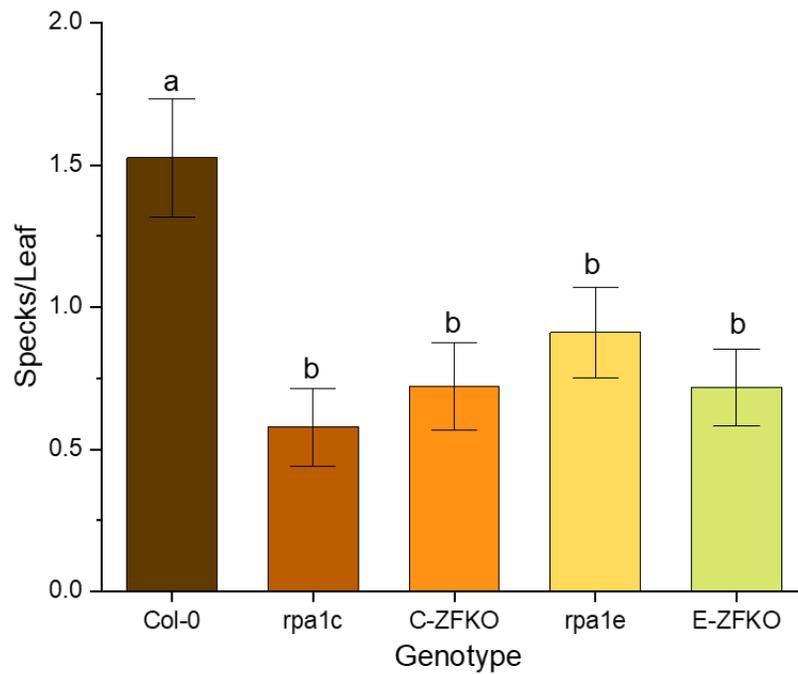
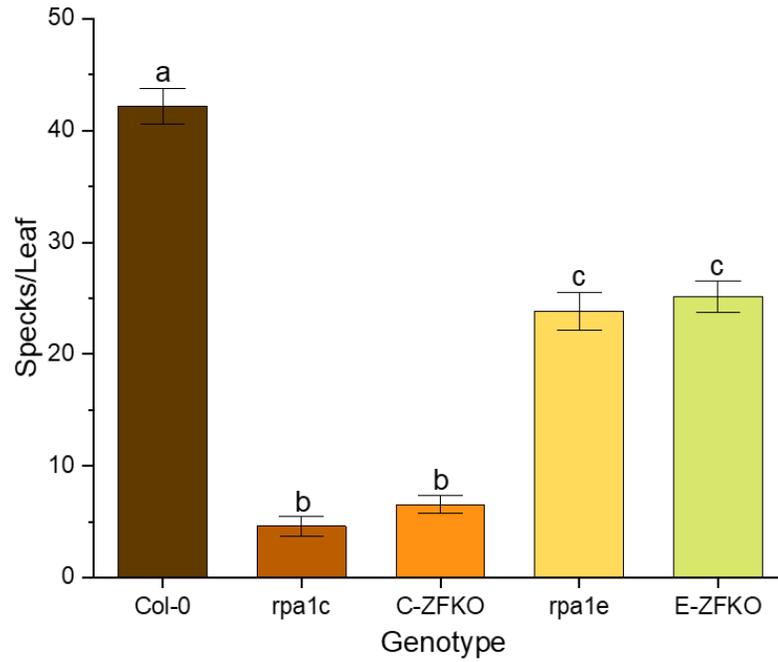**Figure 3.10.** Synthesis-dependent strand (SDSA) activity of two-week-old seedlings. The level of activity was quantified by counting and averaging the number of individual specks per leaf. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error and bars with different letters indicate significant differences (n>40).
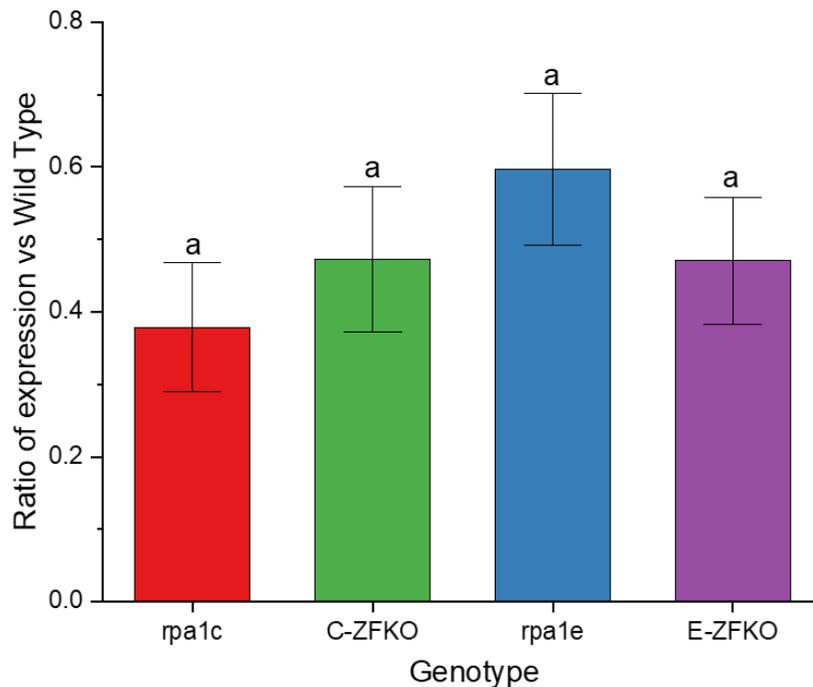
Figure 3.11. Synthesis-dependent strand (SDSA) activity of *Arabidopsis thaliana* mutant lines as a fraction of wild-type expression. All lines displayed approximately the same fraction of the wild-type expression. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.05$. Error bars denote standard error (n>40).

Relative levels of SSA versus SDSA

For all observed lines, the amount of SDSA was significantly lower than the SSA activity in the corresponding line with a ~7-36-fold difference between the SSA and SDSA lines, depending on the mutant background (Figure 3.12). This is consistent with previous work indicating that SSA is the dominant HRR mechanism and occurs at a much higher rate than SDSA. The ratios for *RPA1E* mutants and wild-type lines are very similar (~27-36), while *RPA1C* mutants had a much lower ratio (~7).
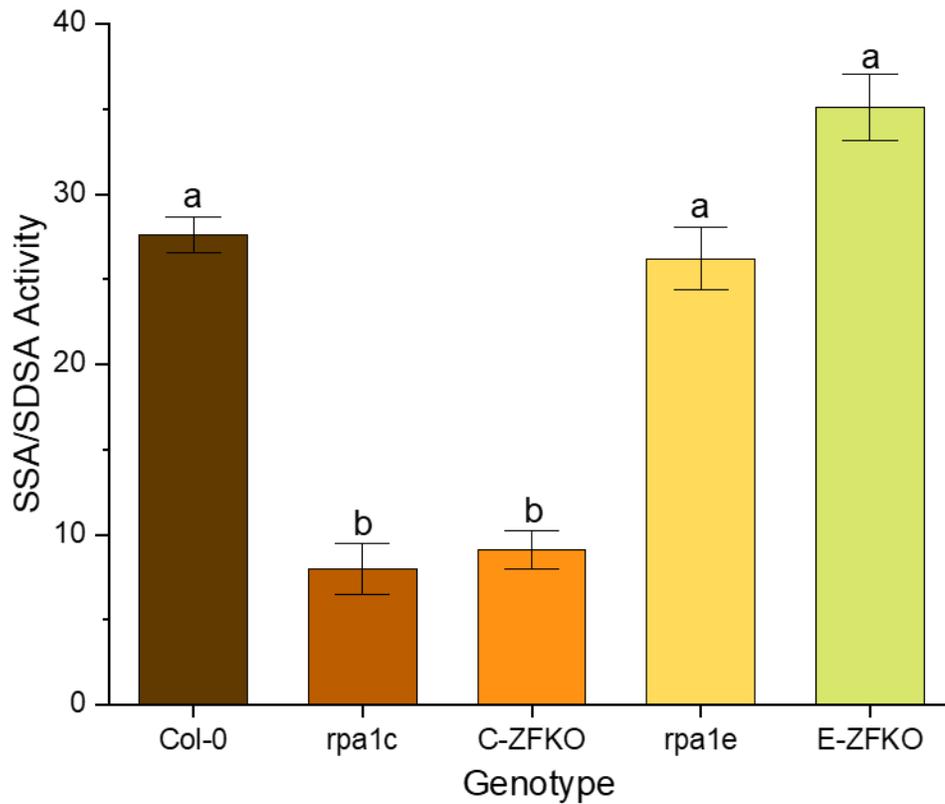
Figure 3.12. Relative levels of single-strand annealing (SSA) and synthesis-dependent strand annealing (SDSA) activity. Statistical analysis was done using F-test (ANOVA) and LSD at $P \leq 0.005$. Error bars denote standard error and bars with different letters indicate significant differences (n>40).

DISCUSSION

In this study, we find that Arabidopsis with non-functional RPA1C or RPA1E had significantly decreased capability for both SSA and SDSA. The *RPA1C* null line (*rpa1c*) displayed a larger decrease in SSA activity than the *RPA1E* null line (*rpa1e*). However, no difference was observed between mutant lines when observing SDSA activity. Furthermore, CRISPR-generated Arabidopsis lines lacking the C-terminal extension ZFM from RPA1C or

RPA1E displayed identical phenotypes to their respective null mutant. These data indicate that both RPA1C and RPA1E are involved in both HRR repair processes. The greater decrease in SSA activity in *RPA1C* mutants compared to *RPA1E* mutants could indicate a preference towards SSA following RPA1C binding but could also reflect the overall greater importance of RPA1C in DSB repair. Additionally, the identical phenotype of the null mutant lines and ZFKO lines indicates that the functionality of both RPA1C and RPA1 in both HRR pathways is reliant upon the presence of the C-terminal extension ZFM.

RPA1C and RPA1E are both involved in both HRR pathways

In this study, we hypothesized that, in *Arabidopsis thaliana*, RPA1C and RPA1E have differential usage between the two methods of HRR. To test this hypothesis, *RPA1* mutants were studied using GUS constructs that were specifically activated by either SSA or SDSA (Mannuss et al., 2010). We found that both *RPA1C* and *RPA1E* mutants have decreased activity of both HRR pathways compared to wild-type Arabidopsis. In the SSA assay, *RPA1C* mutants displayed a more drastic reduction in activity than *RPA1E* mutants (Figure 3.8). It should be noted that this result differs somewhat from previously published work, which found approximately equal decreases in SSA in both *RPA1C* and *RPA1E* null mutants (Liu et al., 2017). However, the previous work verified GUS expression via PCR, and in that assay, there does appear to be a greater decrease in the *RPA1C* null line, matching what was found in this study. It is possible that this difference in phenotype reflects a greater role for RPA1C during SSA than that of RPA1E and specialization of the paralog to this repair mechanism. Previous work has indicated that RPA1C has a greater overall role in DSB repair than that of RPA1E and this has been supported by the work detailed in Chapter II of this thesis (Aklilu et al., 2014). Therefore, the difference in SSA activity between *RPA1C* and *RPA1E* mutants may instead be a reflection of the larger role of

RPA1C in DSB repair and not an indication of specialization. Close examination of the SDSA assay and how it compares to the SSA assay may provide a hint as to which possibility is more likely.

Unlike in the SSA assay, there was no difference in phenotype between the *RPA1C* and *RPA1E* mutants observed in the SDSA assay, as mutants for both paralogs had roughly equivalent decreases in SDSA activity versus wild-type (Figure 3.11). This lack of differentiation between RPA1C and RPA1E may support the proposal that RPA1C has a larger role in SSA. The equality in phenotype for SDSA suggests that the difference seen for SSA may be due to specialization rather than the greater overall role of RPA1C in DSB repair. This is also supported by an examination of the fold-difference from wild-type seen in both assays. While *RPA1E* mutants show an approximately two-fold decrease in expression compared to wild-type in both assays, the amount of decrease seen in *RPA1C* mutants is considerably larger for SSA (9.2 vs 2.6-fold decreases) (Figures 3.7 and 3.10). Furthermore, the ratio of SSA/SDSA is considerably lower in *RPA1C* mutants than in wild-type or *RPA1E* mutant lines (Figure 3.11). This variation could indicate that a DSB that is repaired following RPA1C binding is more likely to result in SSA, while RPA1E binding has no preferred HRR methodology.

If RPA1C does favor SSA, the question of how and why SSA is favored has several possible answers. It is known that RPA-coated ssDNA is the binding target of multiple proteins involved in the DNA damage response, such as Rad9, ATRIP, and Rad52 (Caldwell & Spies, 2020; Xu et al., 2008; Zou & Elledge, 2003). Many of these factors are required for both SSA and SDSA, but some are unique for each pathway (Mannuss et al., 2010; Roth et al., 2012). It is therefore possible that RPA1C and RPA1E have differing binding capabilities for these factors, leading to RPA1C favoring SSA. For example, Rad51 is necessary only for SDSA (SSA is

sometimes referred to as "Rad51-independent repair) and recruited to the site of a DSB following the binding of Rad52 to RPA coated ssDNA (Golub et al., 1998; Serra et al., 2013). It is possible that RPA1C does not interact as favorably with Rad51 in this scenario as RPA1E, thereby favoring DSB repair via SSA. A previous study found that deletion of Rad51 leads to an increase in SSA; therefore lesser binding of Rad51 would similarly lead to an increase in SSA (Agmon et al., 2009).

While differential binding to factors involved specifically in SDSA or SSA is a plausible hypothesis, it is also possible that another DSB repair pathway is involved. Microhomology-mediated end joining (MMEJ) is another repair pathway that repairs DSBs in a very similar manner to SSA (Boulton & Jackson, 1996). A key difference between the processes is that MMEJ requires a smaller region of homology than SSA (2-20 bp vs >20 bp), but it is unclear whether there is an upper limit to the region of homology used for MMEJ (McVey & Lee, 2008; Sugawara et al., 2000). Therefore, it is entirely possible that the GUS assay investigating SSA activity actually recorded both SSA and MMEJ activity. If this is the case, then the differentiation between RPA1C and RPA1E may be due to altered binding to proteins specific to MMEJ. DNA polymerase theta (Polθ) is a prime candidate for this differential binding. Polθ is necessary for MMEJ but not involved in SSA or SDSA, and furthermore has been implicated in preventing the recruitment of Rad51 to ssDNA (Ceccaldi et al., 2015; Koole et al., 2014). RPA1C possessing an increased affinity for Polθ, either through direct binding or an indirect methodology, would thereby result in increased MMEJ activity, mimicking SSA activity and accounting for the differential ratios observed between *RPA1C* mutants and *RPA1E* mutants.

Previous work investigating the interactions of RPA and Polθ indicates that RPA inhibits MMEJ and that its ATP-dependent displacement by Polθ constitutes one of the first steps in

MMEJ (Deng et al., 2014; Mateos-Gomez et al., 2017). However, these studies were conducted using either yeast or human cells, and the RPA1 homolog in these species does not have a C-terminal extension (Aklilu & Culligan, 2016). Therefore, the C-terminal extension may alter the relationship between RPA and Polθ, instead causing the RPA heterotrimer to recruit Polθ and push the repair towards MMEJ. While there have not been any studies investigating the interaction between RPA and Polθ in Arabidopsis, Polθ itself has been studied in both Arabidopsis and other plant models. In Arabidopsis, the gene coding for Polθ is known as *TEBICHI* (*TEB*). In Arabidopsis, Polθ has been found to be involved in multiple cellular processes and is thought to be crucial for resolving replicative stress and activating MMEJ, with *TEB* null mutants exhibiting severe developmental defects (Inagaki et al., 2009; Inagaki et al., 2006). Additionally, Polθ is increasingly shown to have a role in both integrating T-DNA and mediating the repair of DSBs induced by CRISPR-Cas9, making it a promising target of future studies to optimize gene editing in plants (Kamoen et al., 2024; van Kregten et al., 2016; van Tol et al., 2022). These potential future studies would likely delve into how the MMEJ process occurs in plants and would necessitate further investigation into how RPA is involved.

Another possible rationale for the use of RPA in the promotion of MMEJ in plants can be found by examining the role of 5-hydroxymethylcytosine binding, ESC-specific protein (HMCES) in mammalian MMEJ. In the current mammalian model, HMCES binds in the place of RPA and promotes MMEJ while RPA binding leads to repair via HRR (Patterson-Fortin & D'Andrea, 2020). However, there have not yet been any investigations into this process in plants, and there is some evidence to suggest that HMCES may not perform the same role in plants that it does in mammals. 5-hydroxymethylcytosine, the namesake and presumed primary binding target of HMCES, is a modified nucleotide that is common in mammals but is notably

uncommon in Arabidopsis (Erdmann et al., 2014; Spruijt et al., 2013). This indicates that the role of HMCES in mammals may not be shared in plants. In this case, it is possible that in plants RPA fulfills the proposed mammalian role of HMCES and promotes MMEJ while also maintaining its role in promoting HRR.

Should RPA be involved in the promotion of MMEJ and HRR in plants, there remains an issue of how regulation of such a system would work. Perhaps the most straightforward possibility is that of competitive regulation. The concept of competition between DSB repair mechanisms has some precedent with previous research finding that there is competition between NHEJ and MMEJ (Osakabe et al., 2010). In a competitive model, the repair mechanism is determined by whether Polθ and other MMEJ related factors or Rad52 and other HRR related factors bind to the RPA-coated ssDNA (Figure 3.13). This binding could also be regulated by varying the level of expression of these factors based on cellular conditions and the phase of the cell cycle to favor one mechanism over another. Having the factors available for all viable repair mechanisms does make sense from a physiological perspective, because when faced with a DSB, successful repair is more important than which repair mechanism is used. An extensive decision-making process when faced with a DSB could slow the initiation of repair while determining which repair mechanism to use based on which factor binds first would lead to a more rapid response with the cost of a potential accumulation of mutations.

This trade-off of more rapid repair of DSBs in exchange for the greater likelihood of errors occurring during the repair process would make sense to exist in plants. Unlike most eukaryotes, plants are sessile organisms and therefore are unable to move to avoid stressors that may cause DSBs. This means that a plant's ability to repair damage such as DSBs is paramount, even if it is using an error-prone methodology such as MMEJ. Error-prone DSB repair in

mammals can lead to the accumulation of mutations and eventually serious consequences such as cancer, but the risk of error-prone repair mechanisms is much lesser in plants due to the inability to get cancer (Doonan & Sablowski, 2010). This greater tolerance of errors during DSB repair combined with the greater risk of DNA damage from environmental mutagens would make the evolution of a regulatory system that favors MMEJ more beneficial and could explain why the C-terminal extension is ubiquitous in plants but not found elsewhere.
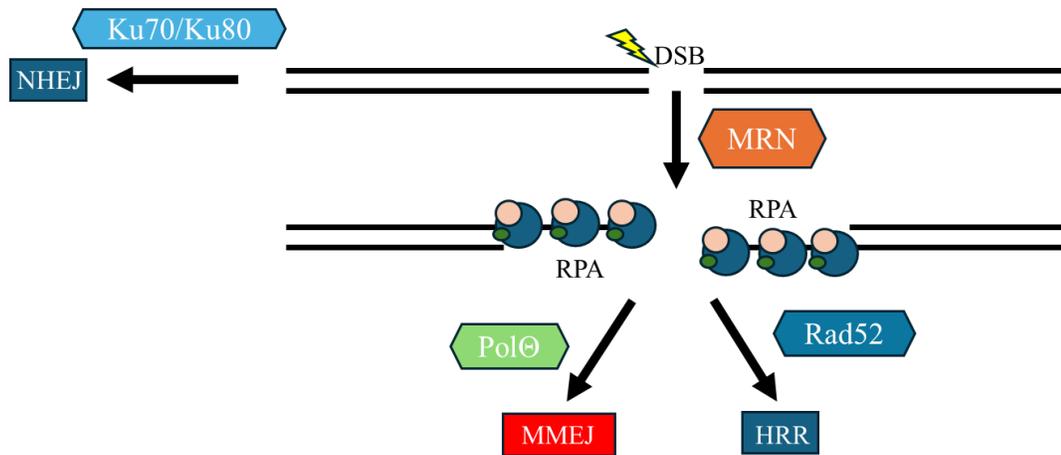
Figure 3.13. Potential model for double-strand break (DSB) repair in plants. Break can be repaired directly by via Ku70/Ku80 mediated non-homologous end joining (NHEJ). Alternatively, the ends can be resected by the Mre11-Rad50-NBS1 (MRN) complex and bound by Replication Protein A (RPA). Following the binding of RPA the DSB can be repaired either by microhomology-mediated end joining (MMEJ) or homologous recombination repair (HRR). This decision may be determined through the competitive binding of either polymerase theta (Polθ) and other MMEJ related factors or Rad52 and other HRR related factors.

The difference between the wild-type vs *RPA1C* mutant ratios when comparing the SSA and SDSA assays is compelling and has several potential implications, but a possible shortfall of the assay should also be considered. The rate of SDSA found in all the plant lines examined was extremely low, with even the wild-type plants having an average of 1.5 repair events per leaf. With such a low maximum level of expression, any observed trends are somewhat unreliable. It

is possible that with an altered protocol, an increased level of expression could be observed, resulting in more definitive results.

RPA1 group C paralogs require the C-terminal extension ZFM for functionality in HRR pathways

We hypothesized that the C-terminal extension of RPA1C and RPA1E, and specifically the ZFM contained therein, is crucial to the functionality of these paralogs in DNA damage repair. As detailed in Chapter II, this hypothesis was supported for DNA damage caused by both CPT and gamma radiation. We further hypothesized that the C-terminal ZFM may influence how the RPA1 paralogs carry out HRR, possibly leading to the favoring of one pathway over the other. To test this hypothesis, in addition to studying the rate of SSA vs SDSA in *RPA1* null mutants, mutant *RPA1* lines that lacked the C-terminal domain ZFM were investigated in the same manner. For both RPA1C and RPA1E, the levels of SSA and SDSA were nearly identical between the ZFM knockout (ZFKO) line and its respective null mutant (Figures 3.6 and 3.9). These data indicate that the C-terminal ZFMs are necessary for functionality in both HRR pathways, mirroring the results found in Chapter II and emphasizing the importance of the ZFMs to the functionality of the paralogs.

The work in Chapter II suggests that the C-terminal ZFM is crucial for DNA damage repair. However, it does leave open the possibility that RPA1 paralogs are capable of functioning in some lesser-used DNA repair pathways even with the removal of the C-terminal extension. The majority of DSB repair in higher plants is conducted via NHEJ; so the HRR pathways would be prime candidates for possible pathways to remain intact (Chen et al., 2022). The results of the GUS assay indicate that this is not the case, as no difference could be observed between the

RPA1 ZFKO lines and the corresponding null mutant. This result, combined with what was observed in Chapter II, indicates that the functionality of RPA1C and RPA1E in DSB repair is reliant on the presence of the C-terminal ZFM and that it is unlikely that the paralogs can function in any DSB repair pathway when missing this domain.

RPA1C and RPA1E are very similar proteins, which both contain the same domains in roughly the same locations (Aklilu & Culligan, 2016). A key difference between the two is the C-terminal extension ZFM, which while a CCHC-type in both, varies in sequence between RPA1C and RPA1E. These distinct ZFMs are maintained in non-Brassica plants that have only a single group C RPA1 paralog, indicating the ZFMs may have distinct functionality (Figure 3.3). As previously discussed, RPA1C and RPA1E have differing roles in DNA damage repair, with RPA1C playing a leading role and RPA1E seemingly filling a more auxiliary role. A discussion of some of the details of this variation can be found in the previous section. Given the variation in the role of RPA1C and RPA1E, as well as the variation in the C-terminal ZFM and the importance of that domain to the overall function of the protein, we hypothesize that the ZFMs may be responsible for the differential roles of the paralogs. While it has not been experimentally determined, the two ZFMs likely have distinct binding capabilities which may alter the ability of the RPA heterotrimer to bind under different circumstances and alter which proteins are recruited or with what affinity. This would account not only for the split duties of RPA1C and RPA1E but would also provide a potential explanation for the retention of both ZFMs in non-Brassica. Two adjacent ZFMs with differing binding capabilities could allow for more flexible binding patterns and interactions and potentially could act synergistically to bind a single target. Both possibilities could enhance the functionality of RPA in DNA damage repair, an extremely beneficial trait for any organism.

SUMMARY AND CONCLUSION

The C-terminal extension ZFM is crucial to the function of RPA1 group C paralogs during DSB repair.

A CRISPR-Cas9 system was utilized to remove the C-terminal ZFM from RPA1C and RPA1E in wild-type *Arabidopsis thaliana*. After verification of the gene editing by both PCR and sequencing, the resulting plants were tested for susceptibility to DNA damage via exposure to either CPT or gamma radiation along with *RPA1C* and *RPA1E* null mutant lines (*rpa1c* and *rpa1e*). In both cases the ZFKO and null mutant lines presented identical phenotypes, suggesting that the ZFM is necessary for the functionality of the paralogs during DNA damage repair. Additionally, both of the *RPA1C* ZFKO lines and *rpa1c* displayed greater hypersensitivity to both CPT and gamma radiation than the *RPA1E* mutant lines, supporting previous work done in our laboratory suggesting that RPA1C has a leading role in repairing DSBs, while RPA1E has a more auxiliary role (Aklilu et al., 2014). The exact role of RPA1E is currently unclear. Due to the importance of DNA damage repair, it may simply exist as a back-up to ensure that RPA of some type is always available during DNA damage repair. As plants are sessile organisms and cannot move to avoid stressors, redundancy on such a key protein would be beneficial. Alternatively, RPA1E may be specialized to function during a specific stage of the cell cycle or in response to a specific type of damage in addition to functioning as a back-up. A possible future study would be to measure the level of expression of both RPA1C and RPA1E in different phases of the cell cycle to determine if differential expression may play a role in the differential function of the paralogs.

Further investigation into the roles of RPA1C and RPA1E was conducted using GUS reporter lines which are activated specifically through either SSA or SDSA. This experiment found that both RPA1C and RPA1E are needed for both HRR pathways, but while SDSA is likely reduced equally in *RPA1C* and *RPA1E* mutants, SSA is significantly reduced in *RPA1C* mutants versus *RPA1E* mutants. As with the earlier DNA damage assays, the ZFKO lines and respective null mutants were indistinguishable. These results indicate that the ZFMs are needed for functionality in both HRR pathways and that RPA1C may have an enhanced role in SSA.

It is possible that the SSA assay represents not only SSA activity but MMEJ activity as well, and that the greater decrease in activity seen for RPA1C is due to a preference for MMEJ. In this model, Polθ would likely have a greater affinity for ssDNA coated with an RPA heterotrimer containing RPA1C than RPA1E, resulting in the RPA1C-containing complex recruiting Polθ more often and thereby undergoing MMEJ more readily. To test this experimentally, the first step would be to generate and isolate specific heterotrimers (Aklilu et al., 2020; Eschbach & Kobbe, 2014). These heterotrimers could be used to test variation in affinity for ssDNA via gel mobility shift assays, as it is possible that the varied phenotypes of *RPA1C* and *RPA1E* null mutants is simply due to differential ssDNA binding affinity. Following this, ssDNA bound by RPA could be combined in solution with Polθ and a similar gel mobility shift assay could be used to determine whether the RPA1 subunit alters Polθ recruitment. There are also a multitude of other proteins that bind to RPA-coated ssDNA, each of which may have differential binding depending on the RPA1 subunit and could be similarly tested.

Along with experiments to further examine the differentiation between RPA1C and RPA1E, more work could be done to evaluate the impact of the removal of the C-terminal ZFM

of these paralogs. Our research indicates that the ZFMs are necessary for the functionality of the paralogs, as in every assay completed, plants lacking the ZFM presented identical phenotypes to null mutants. However, it is possible that this is due to the inability to form the RPA heterotrimer rather than the loss of the ZFMs binding capability. To examine this, RPA1C and RPA1E subunits with the same sequence as the ZFKO lines could be generated and evaluated for viability in heterotrimer formation and for the ssDNA binding capability of each heterotrimer. If ZFKO lines are capable of heterotrimer formation, then the same experiments proposed for examining affinity to Polθ and other potential binding partners discussed above could be conducted with these heterotrimers to directly assess how the absence of the ZFM alters the behavior of RPA1C and RPA1E.

One additional set of binding partners to investigate are those related to INVOLVED IN DE NOVO2 (IDN2), a dsRNA binding protein recently found to bind RPA during HRR based DSB repair (Liu et al., 2017). IDN2 was found to specifically bind to RPA2B and not to any of the RPA1 paralogs but given its presence alongside RPA during the HRR process it, along with its binding partners, would be worthwhile to investigate. Even without direct binding to RPA1, the presence or absence of the ZFM may be a key factor in allowing for IDN2 binding. Liu et al. (2017) also found that RPA2B only formed a complex with RPA1C or RPA1E, so it is possible that the ZFM is needed for interactions with other RPA subunits to form a heterotrimer that is active in DNA damage repair or that it otherwise helps to facilitate the binding of IDN2 and RPA2B (Figure 3.14).
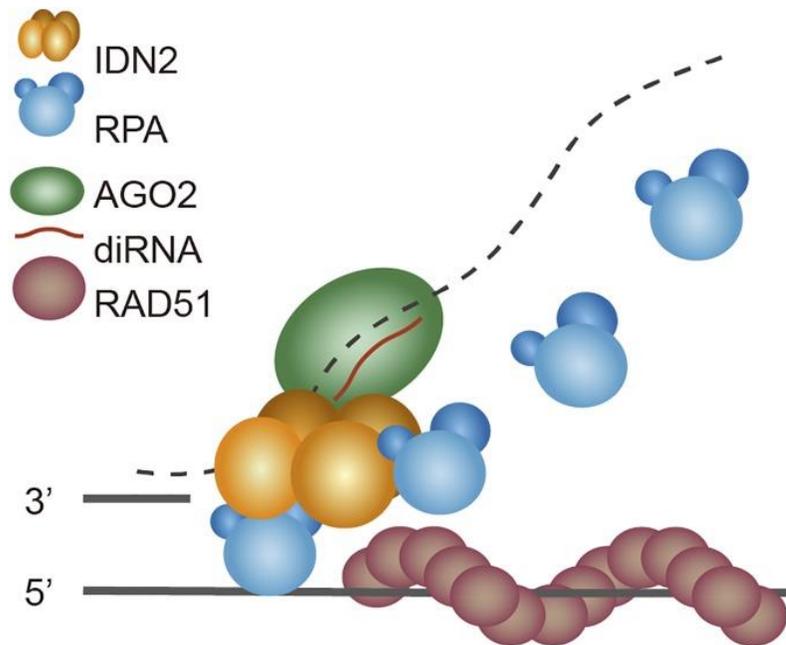
Figure 3.14. Hypothesized model for the interaction of INVOLVED IN DE NOVO2 (IDN2) and Replication Protein A (RPA). In this model AGO2/diRNA complexes are recruited to a DSB, providing a template for IDN2 binding. IDN2 binding then facilitates the replacement of RPA with Rad51 on the ssDNA (Liu et al., 2017).

Another methodology to examine the roles of RPA1C and RPA1E C-terminal ZFMs is to test what occurs if neither is present. Double mutant lines that lack both RPA1C and RPA1E have previously been found to have supra-additive hypersensitivity to DNA damaging agents (Aklilu et al., 2014). These same double mutants could be used for the HRR assay to observe whether those processes are similarly reduced further in double mutants. Furthermore, the ZFKO lines could be combined both with each other and with the opposing null mutant to examine how DNA damage hypersensitivity and overall growth may vary. One portion of these experiments could be a comet assay with a slightly altered procedure. The comet assays conducted as a part of this study were completed a week or more after exposure to a DNA damaging agent and primarily showed that even plants with mutant RPA1C or RPA1E subunits were capable of

eventually completing DNA damage repair. Conducting these same assays much sooner after exposure to the DNA damaging agent, possibly with multiple time points, would better examine how mutant RPA1 subunits affect DNA damage repair. This assay could be conducted both with the currently existing null mutants and all possible double mutant combinations.

While there is always more work to be done and more to be discovered, the work presented here provides a strong platform for future endeavors. We have shown that RPA1C and RPA1E are both involved in DNA damage repair throughout the cell cycle and in both HRR repair pathways. Furthermore, we have reinforced previous work indicating that RPA1C has a larger role in DNA damage repair than RPA1E and have gone further to implicate SSA or MMEJ as a possible area where the greater role of RPA1C is emphasized. Alongside the work investigating the overall roles of RPA1C and RPA1E, the CRISPR-Cas9 gene editing system was used to generate mutant lines that lack the C-terminal extension ZFM of either RPA1C or RPA1E, which are now available for future experiments and investigations. The C-terminal ZFM of both paralogs was found to be necessary for the function of the paralogs in DNA damage repair, and plants lacking an RPA1 C-terminal ZFM presented identical phenotypes as the respective null mutants in every experiment. While the exact role of the ZFM is yet to be determined, it is clear that the domain is important for the role RPA1 plays in DNA damage repair in plants. Further investigation building upon the results found in this study could lead to a greater understanding of how RPA1 functions in DNA damage repair in plants and how its role and regulation varies between plants and other eukaryotes.

REFERENCES

Agmon, N., Pur, S., Liefshitz, B., & Kupiec, M. (2009). Analysis of repair mechanism choice during homologous recombination. *Nucleic Acids Res*, *37*(15), 5081-5092. https://doi.org/10.1093/nar/gkp495

Aguilera, A., & García-Muse, T. (2013). Causes of genome instability. *Annu Rev Genet*, *47*, 1-32. https://doi.org/10.1146/annurev-genet-111212-133232

Aklilu, B. B., & Culligan, K. M. (2016). Molecular Evolution and Functional Diversification of Replication Protein A1 in Plants [Original Research]. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00033

Aklilu, B. B., Peurois, F., Saintomé, C., Culligan, K. M., Kobbe, D., Leasure, C., Chung, M., Cattoor, M., Lynch, R., Sampson, L., Fatora, J., & Shippen, D. E. (2020). Functional Diversification of Replication Protein A Paralogs and Telomere Length Maintenance in Arabidopsis. *Genetics*, *215*(4), 989-1002. https://doi.org/10.1534/genetics.120.303222

Aklilu, B. B., Soderquist, R. S., & Culligan, K. M. (2014). Genetic analysis of the Replication Protein A large subunit family in Arabidopsis reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. *Nucleic Acids Res*, *42*(5), 3104-3118. https://doi.org/10.1093/nar/gkt1292

Amiard, S., Depeiges, A., Allain, E., White, C. I., & Gallego, M. E. (2011). Arabidopsis ATM and ATR kinases prevent propagation of genome damage caused by telomere dysfunction. *Plant Cell*, *23*(12), 4254-4265. https://doi.org/10.1105/tpc.111.092387

Anderson, A., & Moore, L. (1979). Host specificity in the genus Agrobacterium. *Phytopathology*, *69*(4), 320-323.

Arunkumar, A. I., Stauffer, M. E., Bochkareva, E., Bochkarev, A., & Chazin, W. J. (2003). Independent and coordinated functions of replication protein A tandem high affinity single-stranded DNA binding domains. *J Biol Chem*, *278*(42), 41077-41082. https://doi.org/10.1074/jbc.M305871200

Babina, D., Podobed, M., Bondarenko, E., Kazakova, E., Bitarishvili, S., Podlutskii, M., Mitsenyk, A., Prazyan, A., Gorbatova, I., Shesterikova, E., & Volkova, P. (2020). Seed Gamma Irradiation of Arabidopsis thaliana ABA-Mutant Lines Alters Germination and Does Not Inhibit the Photosynthetic Efficiency of Juvenile Plants. *Dose Response*, *18*(4), 1559325820979249. https://doi.org/10.1177/1559325820979249

Ball, H. L., Ehrhardt, M. R., Mordes, D. A., Glick, G. G., Chazin, W. J., & Cortez, D. (2007). Function of a conserved checkpoint recruitment domain in ATRIP proteins. *Mol Cell Biol*, *27*(9), 3367-3377. https://doi.org/10.1128/mcb.02238-06

Bessho, T. (2003). Induction of DNA Replication-mediated Double Strand Breaks by Psoralen DNA Interstrand Cross-links *. *Journal of Biological Chemistry*, *278*(7), 5250-5254. https://doi.org/10.1074/jbc.M212323200

Bhargava, R., Onyango, D. O., & Stark, J. M. (2016). Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet*, *32*(9), 566-575. https://doi.org/10.1016/j.tig.2016.06.007

Bibikova, M., Golic, M., Golic, K. G., & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*, *161*(3), 1169-1175. https://doi.org/10.1093/genetics/161.3.1169

Binz, Sheehan, A. M., & Wold, M. S. (2004). Replication protein A phosphorylation and the cellular response to DNA damage. *DNA Repair (Amst)*, *3*(8-9), 1015-1024. https://doi.org/10.1016/j.dnarep.2004.03.028

Bochkarev, A., Bochkareva, E., Frappier, L., & Edwards, A. M. (1999). The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *Embo j*, *18*(16), 4498-4504. https://doi.org/10.1093/emboj/18.16.4498

Bochkareva, E., Kaustov, L., Ayed, A., Yi, G. S., Lu, Y., Pineda-Lucena, A., Liao, J. C., Okorokov, A. L., Milner, J., Arrowsmith, C. H., & Bochkarev, A. (2005). Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc Natl Acad Sci U S A*, *102*(43), 15412-15417. https://doi.org/10.1073/pnas.0504614102

Bochkareva, E., Korolev, S., Lees-Miller, S. P., & Bochkarev, A. (2002). Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *The EMBO Journal*, *21*(7), 1855-1863. https://doi.org/https://doi.org/10.1093/emboj/21.7.1855

Boulton, S. J., & Jackson, S. P. (1996). Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res*, *24*(23), 4639-4648. https://doi.org/10.1093/nar/24.23.4639

Britt, A. B. (2004). Repair of DNA Damage Induced by Solar UV. *Photosynthesis Research*, *81*(2), 105-112. https://doi.org/10.1023/B:PRES.0000035035.12340.58

Caldwell, C. C., & Spies, M. (2020). Dynamic elements of replication protein A at the crossroads of DNA replication, recombination, and repair. *Crit Rev Biochem Mol Biol*, *55*(5), 482-507. https://doi.org/10.1080/10409238.2020.1813070

Caplin, N., & Willey, N. (2018). Ionizing Radiation, Higher Plants, and Radioprotection: From Acute High Doses to Chronic Low Doses. *Front Plant Sci*, *9*, 847. https://doi.org/10.3389/fpls.2018.00847

Ceccaldi, R., Liu, J. C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M. I., O'Connor, K. W., Konstantinopoulos, P. A., Elledge, S. J., Boulton, S. J., Yusufzai, T., & D'Andrea, A. D. (2015). Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature*, *518*(7538), 258-262. https://doi.org/10.1038/nature14184

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*, *18*(8), 495-506. https://doi.org/10.1038/nrm.2017.48

Charbonnel, C., Allain, E., Gallego, M. E., & White, C. I. (2011). Kinetic analysis of DNA double-strand break repair pathways in Arabidopsis. *DNA Repair*, *10*(6), 611-619. https://doi.org/https://doi.org/10.1016/j.dnarep.2011.04.002

Chen, H., Neubauer, M., & Wang, J. P. (2022). Enhancing HR Frequency for Precise Genome Editing in Plants [Perspective]. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.883421

Chen, R., & Wold, M. S. (2014). Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays*, *36*(12), 1156-1161. https://doi.org/10.1002/bies.201400107

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., & Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, *186*(2), 757-761. https://doi.org/10.1534/genetics.110.120717

Cimprich, K. A., & Cortez, D. (2008). ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol*, *9*(8), 616-627. https://doi.org/10.1038/nrm2450

Cortez, D., Guntuku, S., Qin, J., & Elledge, S. J. (2001). ATR and ATRIP: partners in checkpoint signaling. *Science*, *294*(5547), 1713-1716. https://doi.org/10.1126/science.1065521

Deng, S. K., Gibb, B., de Almeida, M. J., Greene, E. C., & Symington, L. S. (2014). RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat Struct Mol Biol*, *21*(4), 405-412. https://doi.org/10.1038/nsmb.2786

Ding, X., Jia, X., Xiang, Y., & Jiang, W. (2022). Histone Modification and Chromatin Remodeling During the Seed Life Cycle. *Front Plant Sci*, *13*, 865361. https://doi.org/10.3389/fpls.2022.865361

Doonan, J. H., & Sablowski, R. (2010). Walls around tumours — why plants do not develop cancer. *Nature Reviews Cancer*, *10*(11), 794-802. https://doi.org/10.1038/nrc2942

Erdmann, R. M., Souza, A. L., Clish, C. B., & Gehring, M. (2014). 5-hydroxymethylcytosine is not present in appreciable quantities in Arabidopsis DNA. *G3 (Bethesda)*, *5*(1), 1-8. https://doi.org/10.1534/g3.114.014670

Eschbach, V., & Kobbe, D. (2014). Different Replication Protein A Complexes of Arabidopsis thaliana Have Different DNA-Binding Properties as a Function of Heterotrimer Composition. *Plant and Cell Physiology*, *55*(8), 1460-1472. https://doi.org/10.1093/pcp/pcu076

Esnault, M.-A., Legue, F., & Chenal, C. (2010). Ionizing radiation: Advances in plant response. *Environmental and Experimental Botany*, *68*(3), 231-237. https://doi.org/https://doi.org/10.1016/j.envexpbot.2010.01.007

Fulcher, N., & Sablowski, R. (2009). Hypersensitivity to DNA damage in plant stem cell niches. *Proc Natl Acad Sci U S A*, *106*(49), 20984-20988. https://doi.org/10.1073/pnas.0909218106

Ganai, R. A., & Johansson, E. (2016). DNA Replication-A Matter of Fidelity. *Mol Cell*, *62*(5), 745-755. https://doi.org/10.1016/j.molcel.2016.05.003

Ganpudi, A., & Schroeder, D. (2011). UV Damaged DNA Repair & Tolerance in Plants, Selected Topics in DNA Repair. *InTech*.

Gelvin, S. B. (2003). Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev*, *67*(1), 16-37, table of contents. https://doi.org/10.1128/mmbr.67.1.16-37.2003

Gill, S. S., Anjum, N. A., Gill, R., Jha, M., & Tuteja, N. (2015). DNA damage and repair in plants under ultraviolet and ionizing radiations. *ScientificWorldJournal*, *2015*, 250158. https://doi.org/10.1155/2015/250158

Gimenez, E., & Manzano-Agugliaro, F. (2017). DNA Damage Repair System in Plants: A Worldwide Research Update. *Genes (Basel)*, *8*(11). https://doi.org/10.3390/genes8110299

Gisler, B., Salomon, S., & Puchta, H. (2002). The role of double-strand break-induced allelic homologous recombination in somatic plant cells. *The Plant Journal*, *32*(3), 277-284. https://doi.org/https://doi.org/10.1046/j.1365-313X.2002.01421.x

Goldberg, R. B., de Paiva, G., & Yadegari, R. (1994). Plant Embryogenesis: Zygote to Seed. *Science*, *266*(5185), 605-614. https://doi.org/doi:10.1126/science.266.5185.605

Golub, E. I., Gupta, R. C., Haaf, T., Wold, M. S., & Radding, C. M. (1998). Interaction of human rad51 recombination protein with single-stranded DNA binding protein, RPA. *Nucleic Acids Res*, *26*(23), 5388-5393. https://doi.org/10.1093/nar/26.23.5388

Gyori, B. M., Venkatachalam, G., Thiagarajan, P. S., Hsu, D., & Clement, M. V. (2014). OpenComet: an automated tool for comet assay image analysis. *Redox Biol*, *2*, 457-465. https://doi.org/10.1016/j.redox.2013.12.020

Han, L., & Yu, K. (2008). Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV-deficient B cells. *J Exp Med*, *205*(12), 2745-2753. https://doi.org/10.1084/jem.20081623

Haring, S. J., Mason, A. C., Binz, S. K., & Wold, M. S. (2008). Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints. *J Biol Chem*, *283*(27), 19095-19111. https://doi.org/10.1074/jbc.M800881200

Hille, F., & Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philos Trans R Soc Lond B Biol Sci*, *371*(1707). https://doi.org/10.1098/rstb.2015.0496

Hwang, H.-H., Yu, M., & Lai, E.-M. (2017). *Agrobacterium*-Mediated Plant Transformation: Biology and Applications. *The Arabidopsis Book*, *2017*(15). https://doi.org/10.1199/tab.0186

Inagaki, S., Nakamura, K., & Morikami, A. (2009). A link among DNA replication, recombination, and gene expression revealed by genetic and genomic analysis of TEBICHI gene of Arabidopsis thaliana. *PLoS Genet*, *5*(8), e1000613. https://doi.org/10.1371/journal.pgen.1000613

Inagaki, S., Suzuki, T., Ohto, M. A., Urawa, H., Horiuchi, T., Nakamura, K., & Morikami, A. (2006). Arabidopsis TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. *Plant Cell*, *18*(4), 879-892. https://doi.org/10.1105/tpc.105.036798

Ishibashi, T., Kimura, S., & Sakaguchi, K. (2006). A higher plant has three different types of RPA heterotrimeric complex. *J Biochem*, *139*(1), 99-104. https://doi.org/10.1093/jb/mvj014

Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *Embo j*, *6*(13), 3901-3907. https://doi.org/10.1002/j.1460-2075.1987.tb02730.x

Jiang, F., & Doudna, J. A. (2017). CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys*, *46*, 505-529. https://doi.org/10.1146/annurev-biophys-062215-010822

Jouanin, L., Bouchez, D., Drong, R. F., Tepfer, D., & Slightom, J. L. (1989). Analysis of TR-DNA/plant junctions in the genome of a Convolvulus arvensis clone transformed by Agrobacterium rhizogenes strain A4. *Plant Mol Biol*, *12*(1), 75-85. https://doi.org/10.1007/bf00017449

Kamoen, L., Kralemann, L. E. M., van Schendel, R., van Tol, N., Hooykaas, P. J. J., de Pater, S., & Tijsterman, M. (2024). Genetic dissection of mutagenic repair and T-DNA capture at CRISPR-induced DNA breaks in Arabidopsis thaliana. *PNAS Nexus*, *3*(3). https://doi.org/10.1093/pnasnexus/pgae094

Karlson, D., & Imai, R. (2003). Conservation of the cold shock domain protein family in plants. *Plant Physiol*, *131*(1), 12-15. https://doi.org/10.1104/pp.014472

Karlson, D., Nakaminami, K., Toyomasu, T., & Imai, R. (2002). A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. *J Biol Chem*, *277*(38), 35248-35256. https://doi.org/10.1074/jbc.M205774200

Karthika, V., Babitha, K. C., Kiranmai, K., Shankar, A. G., Vemanna, R. S., & Udayakumar, M. (2020). Involvement of DNA mismatch repair systems to create genetic diversity in plants for speed breeding programs. *Plant Physiology Reports*, *25*(2), 185-199. https://doi.org/10.1007/s40502-020-00521-9

Keshav, K. F., Chen, C., & Dutta, A. (1995). Rpa4, a homolog of the 34-kilodalton subunit of the replication protein A complex. *Mol Cell Biol*, *15*(6), 3119-3128. https://doi.org/10.1128/mcb.15.6.3119

Kim, C., Paulus, B. F., & Wold, M. S. (1994). Interactions of human replication protein A with oligonucleotides. *Biochemistry*, *33*(47), 14197-14206. https://doi.org/10.1021/bi00251a031

Kim, J. H., Ryu, T. H., Lee, S. S., Lee, S., & Chung, B. Y. (2019). Ionizing radiation manifesting DNA damage response in plants: An overview of DNA damage signaling and repair mechanisms in plants. *Plant Sci*, *278*, 44-53. https://doi.org/10.1016/j.plantsci.2018.10.013

Kim, S. I., Veena, & Gelvin, S. B. (2007). Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *Plant J*, *51*(5), 779-791. https://doi.org/10.1111/j.1365-313X.2007.03183.x

Kirik, A., Salomon, S., & Puchta, H. (2000). Species-specific double-strand break repair and genome evolution in plants. *Embo j*, *19*(20), 5562-5566. https://doi.org/10.1093/emboj/19.20.5562

Kohli, A., Twyman, R. M., Abranches, R., Wegel, E., Stoger, E., & Christou, P. (2003). Transgene integration, organization and interaction in plants. *Plant Mol Biol*, *52*(2), 247-258. https://doi.org/10.1023/a:1023941407376

Koole, W., van Schendel, R., Karambelas, A. E., van Heteren, J. T., Okihara, K. L., & Tijsterman, M. (2014). A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun*, *5*, 3216. https://doi.org/10.1038/ncomms4216

Kovács, E., & Keresztes, Á. (2002). Effect of gamma and UV-B/C radiation on plant cells. *Micron*, *33*(2), 199-210. https://doi.org/https://doi.org/10.1016/S0968-4328(01)00012-9

Krysan, P. J., Young, J. C., & Sussman, M. R. (1999). T-DNA as an Insertional Mutagen in Arabidopsis. *The Plant Cell*, *11*(12), 2283-2290. https://doi.org/10.1105/tpc.11.12.2283

Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G. N., Fischer, R. L., Okamuro, J. K., Harada, J. J., & Goldberg, R. B. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci U S A*, *107*(18), 8063-8070. https://doi.org/10.1073/pnas.1003530107

Lee, J. H., & Paull, T. T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science*, *308*(5721), 551-554. https://doi.org/10.1126/science.1108297

Lei, Y., Lu, L., Liu, H.-Y., Li, S., Xing, F., & Chen, L.-L. (2014). CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants. *Molecular Plant*, *7*(9), 1494-1496. https://doi.org/10.1093/mp/ssu044

Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*, *79*, 181-211. https://doi.org/10.1146/annurev.biochem.052308.093131

Liew, C. K., Kowalski, K., Fox, A. H., Newton, A., Sharpe, B. K., Crossley, M., & Mackay, J. P. (2000). Solution Structures of Two CCHC Zinc Fingers from the FOG Family Protein U-Shaped that Mediate Protein–Protein Interactions. *Structure*, *8*(11), 1157-1166. https://doi.org/10.1016/S0969-2126(00)00527-X

Liu, L. F., Desai, S. D., Li, T. K., Mao, Y., Sun, M., & Sim, S. P. (2000). Mechanism of action of camptothecin. *Ann N Y Acad Sci*, *922*, 1-10. https://doi.org/10.1111/j.1749-6632.2000.tb07020.x

Liu, M., Ba, Z., Costa-Nunes, P., Wei, W., Li, L., Kong, F., Li, Y., Chai, J., Pontes, O., & Qi, Y. (2017). IDN2 Interacts with RPA and Facilitates DNA Double-Strand Break Repair by Homologous Recombination in Arabidopsis. *Plant Cell*, *29*(3), 589-599. https://doi.org/10.1105/tpc.16.00769

Liu, Q., Guntuku, S., Cui, X. S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A., Donehower, L. A., & Elledge, S. J. (2000). Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev*, *14*(12), 1448-1459.

Lockton, S., & Gaut, B. S. (2005). Plant conserved non-coding sequences and paralogue evolution. *Trends Genet*, *21*(1), 60-65. https://doi.org/10.1016/j.tig.2004.11.013

Longhese, M. P., Plevani, P., & Lucchini, G. (1994). Replication factor A is required in vivo for DNA replication, repair, and recombination. *Mol Cell Biol*, *14*(12), 7884-7890. https://doi.org/10.1128/mcb.14.12.7884-7890.1994

Louis, E. J. (2007). Evolutionary genetics: making the most of redundancy. *Nature*, *449*(7163), 673-674. https://doi.org/10.1038/449673a

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, *290*(5494), 1151-1155. https://doi.org/10.1126/science.290.5494.1151

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., . . . Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*, *13*(11), 722-736. https://doi.org/10.1038/nrmicro3569

Mannuss, A., Dukowic-Schulze, S., Suer, S., Hartung, F., Pacher, M., & Puchta, H. (2010). RAD5A, RECQ4A, and MUS81 have specific functions in homologous recombination and define different pathways of DNA repair in Arabidopsis thaliana. *Plant Cell*, *22*(10), 3318-3330. https://doi.org/10.1105/tpc.110.078568

Manova, V., & Gruszka, D. (2015). DNA damage and repair in plants – from models to crops [Review]. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00885

Marceau, A. H. (2012). Functions of Single-Strand DNA-Binding Proteins in DNA Replication, Recombination, and Repair. In J. L. Keck (Ed.), *Single-Stranded DNA Binding Proteins: Methods and Protocols* (pp. 1-21). Humana Press. https://doi.org/10.1007/978-1-62703-032-8_1

Marcu, D., Damian, G., Cosma, C., & Cristea, V. (2013). Gamma radiation effects on seed germination, growth and pigment content, and ESR study of induced free radicals in maize (Zea mays). *J Biol Phys*, *39*(4), 625-634. https://doi.org/10.1007/s10867-013-9322-z

Marwedel, T., Ishibashi, T., Lorbiecke, R., Jacob, S., Sakaguchi, K., & Sauter, M. (2003). Plant-specific regulation of replication protein A2 (OsRPA2) from rice during the cell cycle and in response to ultraviolet light exposure. *Planta*, *217*(3), 457-465. https://doi.org/10.1007/s00425-003-1001-z

Mateos-Gomez, P. A., Kent, T., Deng, S. K., McDevitt, S., Kashkina, E., Hoang, T. M., Pomerantz, R. T., & Sfeir, A. (2017). The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. *Nat Struct Mol Biol*, *24*(12), 1116-1123. https://doi.org/10.1038/nsmb.3494

Matsuoka, S., Huang, M., & Elledge, S. J. (1998). Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science*, *282*(5395), 1893-1897. https://doi.org/10.1126/science.282.5395.1893

Matthews, J. M., Kowalski, K., Liew, C. K., Sharpe, B. K., Fox, A. H., Crossley, M., & Mackay, J. P. (2000). A class of zinc fingers involved in protein–protein interactions. *European Journal of Biochemistry*, *267*(4), 1030-1038. https://doi.org/https://doi.org/10.1046/j.1432-1327.2000.01095.x

McVey, M., & Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet*, *24*(11), 529-538. https://doi.org/10.1016/j.tig.2008.08.007

Mendez-Dorantes, C., Bhargava, R., & Stark, J. M. (2018). Repeat-mediated deletions can be induced by a chromosomal break far from a repeat, but multiple pathways suppress such rearrangements. *Genes Dev*, *32*(7-8), 524-536. https://doi.org/10.1101/gad.311084.117

Menke, M., Chen, I. P., Angelis, K. J., & Schubert, I. (2001). DNA damage and repair in Arabidopsis thaliana as measured by the comet assay after treatment with different classes of genotoxins. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *493*(1), 87-93. https://doi.org/https://doi.org/10.1016/S1383-5718(01)00165-6

Moore, R. C., & Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol*, *8*(2), 122-128. https://doi.org/10.1016/j.pbi.2004.12.001

Nguyen, B., Sokoloski, J., Galletto, R., Elson, E. L., Wold, M. S., & Lohman, T. M. (2014). Diffusion of human replication protein A along single-stranded DNA. *J Mol Biol*, *426*(19), 3246-3261. https://doi.org/10.1016/j.jmb.2014.07.014

Nussenzweig, A., & Nussenzweig, M. C. (2007). A backup DNA repair pathway moves to the forefront. *Cell*, *131*(2), 223-225. https://doi.org/10.1016/j.cell.2007.10.005

O'Driscoll, M., Dobyns, W. B., van Hagen, J. M., & Jeggo, P. A. (2007). Cellular and clinical impact of haploinsufficiency for genes involved in ATR signaling. *Am J Hum Genet*, *81*(1), 77-86. https://doi.org/10.1086/518696

O'Malley, R. C., Barragan, C. C., & Ecker, J. R. (2015). A user's guide to the Arabidopsis T-DNA insertion mutant collections. *Methods Mol Biol*, *1284*, 323-342. https://doi.org/10.1007/978-1-4939-2444-8_16

Orel, N., Kyryk, A., & Puchta, H. (2003). Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *Plant J*, *35*(5), 604-612. https://doi.org/10.1046/j.1365-313x.2003.01832.x

Osakabe, K., Osakabe, Y., & Toki, S. (2010). Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. *Proc Natl Acad Sci U S A*, *107*(26), 12034-12039. https://doi.org/10.1073/pnas.1000234107

Pal, A., & Levy, Y. (2019). Structure, stability and specificity of the binding of ssDNA and ssRNA with proteins. *PLoS Comput Biol*, *15*(4), e1006768. https://doi.org/10.1371/journal.pcbi.1006768

Patterson-Fortin, J., & D'Andrea, A. D. (2020). Exploiting the Microhomology-Mediated End-Joining Pathway in Cancer Therapy. *Cancer Research*, *80*(21), 4593-4600. https://doi.org/10.1158/0008-5472.Can-20-1672

Pourrut, B., Pinelli, E., Celiz Mendiola, V., Silvestre, J., & Douay, F. (2015). Recommendations for increasing alkaline comet assay reliability in plants. *Mutagenesis*, *30*(1), 37-43. https://doi.org/10.1093/mutage/geu075

Puchta, H. (2004). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *Journal of Experimental Botany*, *56*(409), 1-14. https://doi.org/10.1093/jxb/eri025

Puchta, H., & Fauser, F. (2014). Synthetic nucleases for genome engineering in plants: prospects for a bright future. *The Plant Journal*, *78*(5), 727-741. https://doi.org/https://doi.org/10.1111/tpj.12338

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, *8*(11), 2281-2308. https://doi.org/10.1038/nprot.2013.143

Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B., & Sinha, R. P. (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J Nucleic Acids*, *2010*, 592980. https://doi.org/10.4061/2010/592980

Reinhardt, H. C., Aslanian, A. S., Lees, J. A., & Yaffe, M. B. (2007). p53-deficient cells rely on ATM- and ATR-mediated checkpoint signaling through the p38MAPK/MK2 pathway for survival after DNA damage. *Cancer Cell*, *11*(2), 175-189. https://doi.org/10.1016/j.ccr.2006.11.024

Ricaud, L., Proux, C., Renou, J. P., Pichon, O., Fochesato, S., Ortet, P., & Montané, M. H. (2007). ATM-mediated transcriptional and developmental responses to gamma-rays in Arabidopsis. *PLoS One*, *2*(5), e430. https://doi.org/10.1371/journal.pone.0000430

Roth, N., Klimesch, J., Dukowic-Schulze, S., Pacher, M., Mannuss, A., & Puchta, H. (2012). The requirement for recombination factors differs considerably between different pathways of homologous double-strand break repair in somatic plant cells. *The Plant Journal*, *72*(5), 781-790. https://doi.org/https://doi.org/10.1111/j.1365-313X.2012.05119.x

Roudaire, T., Héloir, M.-C., Wendehenne, D., Zadoroznyj, A., Dubrez, L., & Poinssot, B. (2021). Cross Kingdom Immunity: The Role of Immune Receptors and Downstream Signaling in Animal and Plant Cell Death [Review]. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.612452

Roy, S. (2014). Maintenance of genome stability in plants: repairing DNA double strand breaks and chromatin structure stability [Mini Review]. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00487

Salomon, S., & Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *Embo j*, *17*(20), 6086-6095. https://doi.org/10.1093/emboj/17.20.6086

Serra, H., Da Ines, O., Degroote, F., Gallego, M. E., & White, C. I. (2013). Roles of XRCC2, RAD51B and RAD51D in RAD51-independent SSA recombination. *PLoS Genet*, *9*(11), e1003971. https://doi.org/10.1371/journal.pgen.1003971

Sharma, P., Jha, A. B., Dubey, R. S., & Pessarakli, M. (2012). Reactive Oxygen Species, Oxidative Damage, and Antioxidative Defense Mechanism in Plants under Stressful Conditions. *Journal of Botany*, *2012*, 217037. https://doi.org/10.1155/2012/217037

Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., & Keck, J. L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol*, *43*(5), 289-318. https://doi.org/10.1080/10409230802341296

Shukla, V., Halabelian, L., Balagere, S., Samaniego-Castruita, D., Feldman, D. E., Arrowsmith, C. H., Rao, A., & Aravind, L. (2020). HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells. *Mol Cell*, *77*(2), 384-394.e384. https://doi.org/10.1016/j.molcel.2019.10.031

Shultz, R. W., Tatineni, V. M., Hanley-Bowdoin, L., & Thompson, W. F. (2007). Genome-wide analysis of the core DNA replication machinery in the higher plants Arabidopsis and rice. *Plant Physiol*, *144*(4), 1697-1714. https://doi.org/10.1104/pp.107.101105

Siebert, R., & Puchta, H. (2002). Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell*, *14*(5), 1121-1131. https://doi.org/10.1105/tpc.001727

Song, J., & Bent, A. F. (2014). Microbial Pathogens Trigger Host DNA Double-Strand Breaks Whose Abundance Is Reduced by Plant Defense Responses. *PLOS Pathogens*, *10*(4), e1004030. https://doi.org/10.1371/journal.ppat.1004030

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., . . . Vermeulen, M. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, *152*(5), 1146-1159. https://doi.org/10.1016/j.cell.2013.02.004

Sugawara, N., Ira, G., & Haber, J. E. (2000). DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair. *Mol Cell Biol*, *20*(14), 5300-5309. https://doi.org/10.1128/mcb.20.14.5300-5309.2000

Summers, M. F. (1991). Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance. *J Cell Biochem*, *45*(1), 41-48. https://doi.org/10.1002/jcb.240450110

Tapia-Ramírez, J., Eggen, B. J., Peral-Rubio, M. J., Toledo-Aral, J. J., & Mandel, G. (1997). A single zinc finger motif in the silencing factor REST represses the neural-specific type II sodium channel promoter. *Proc Natl Acad Sci U S A*, *94*(4), 1177-1182. https://doi.org/10.1073/pnas.94.4.1177

Tinland, B., Hohn, B., & Puchta, H. (1994). Agrobacterium tumefaciens transfers single-stranded transferred DNA (T-DNA) into the plant cell nucleus. *Proceedings of the National Academy of Sciences*, *91*(17), 8000-8004. https://doi.org/doi:10.1073/pnas.91.17.8000

Umezu, K., Sugawara, N., Chen, C., Haber, J. E., & Kolodner, R. D. (1998). Genetic analysis of yeast RPA1 reveals its multiple functions in DNA metabolism. *Genetics*, *148*(3), 989-1005. https://doi.org/10.1093/genetics/148.3.989

van Kregten, M., de Pater, S., Romeijn, R., van Schendel, R., Hooykaas, P. J., & Tijsterman, M. (2016). T-DNA integration in plants results from polymerase-θ-mediated DNA repair. *Nat Plants*, *2*(11), 16164. https://doi.org/10.1038/nplants.2016.164

van Tol, N., van Schendel, R., Bos, A., van Kregten, M., de Pater, S., Hooykaas, P. J. J., & Tijsterman, M. (2022). Gene targeting in polymerase theta-deficient Arabidopsis thaliana. *Plant J*, *109*(1), 112-125. https://doi.org/10.1111/tpj.15557

Vassin, V. M., Wold, M. S., & Borowiec, J. A. (2004). Replication Protein A (RPA) Phosphorylation Prevents RPA Association with Replication Centers. *Molecular and Cellular Biology*, *24*(5), 1930-1943. https://doi.org/10.1128/MCB.24.5.1930-1943.2004

Vítor, A. C., Huertas, P., Legube, G., & de Almeida, S. F. (2020). Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox [Review]. *Frontiers in Molecular Biosciences*, *7*. https://doi.org/10.3389/fmolb.2020.00024

Wang, Y., Putnam, C. D., Kane, M. F., Zhang, W., Edelmann, L., Russell, R., Carrión, D. V., Chin, L., Kucherlapati, R., Kolodner, R. D., & Edelmann, W. (2005). Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice. *Nat Genet*, *37*(7), 750-755. https://doi.org/10.1038/ng1587

Wang, Y., Yu, Y., Pang, Y., Yu, H., Zhang, W., Zhao, X., & Yu, J. (2021). The distinct roles of zinc finger CCHC-type (ZCCHC) superfamily proteins in the regulation of RNA metabolism. *RNA Biology*, *18*(12), 2107-2126. https://doi.org/10.1080/15476286.2021.1909320

Waterworth, W. M., Drury, G. E., Bray, C. M., & West, C. E. (2011). Repairing breaks in the plant genome: the importance of keeping it together. *New Phytol*, *192*(4), 805-822. https://doi.org/10.1111/j.1469-8137.2011.03926.x

Waterworth, W. M., Footitt, S., Bray, C. M., Finch-Savage, W. E., & West, C. E. (2016). DNA damage checkpoint kinase ATM regulates germination and maintains genome stability in seeds. *Proc Natl Acad Sci U S A*, *113*(34), 9647-9652. https://doi.org/10.1073/pnas.1608829113

Weitzman, M. D., & Weitzman, J. B. (2014). What's the damage? The impact of pathogens on pathways that maintain host genome integrity. *Cell Host Microbe*, *15*(3), 283-294. https://doi.org/10.1016/j.chom.2014.02.010

West, C. E., Waterworth, W. M., Sunderland, P. A., & Bray, C. M. (2004). Arabidopsis DNA double-strand break repair pathways. *Biochemical Society Transactions*, *32*(6), 964-966. https://doi.org/10.1042/bst0320964

Wold, M. S. (1997). REPLICATION PROTEIN A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. *Annual Review of Biochemistry*, *66*(1), 61-92. https://doi.org/10.1146/annurev.biochem.66.1.61

Wyka, I. M., Dhar, K., Binz, S. K., & Wold, M. S. (2003). Replication Protein A interactions with DNA: differential binding of the core domains and analysis of the DNA interaction surface. *Biochemistry*, *42*(44), 12909-12918. https://doi.org/10.1021/bi034930h

Xu, X., Vaithiyalingam, S., Glick, G. G., Mordes, D. A., Chazin, W. J., & Cortez, D. (2008). The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling. *Mol Cell Biol*, *28*(24), 7345-7353. https://doi.org/10.1128/mcb.01079-08

Yates, L. A., Aramayo, R. J., Pokhrel, N., Caldwell, C. C., Kaplan, J. A., Perera, R. L., Spies, M., Antony, E., & Zhang, X. (2018). A structural and dynamic model for the assembly of Replication Protein A on single-stranded DNA. *Nature Communications*, *9*(1), 5447. https://doi.org/10.1038/s41467-018-07883-7

Yoshiyama, K., Conklin, P. A., Huefner, N. D., & Britt, A. B. (2009). Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc Natl Acad Sci U S A*, *106*(31), 12843-12848. https://doi.org/10.1073/pnas.0810304106

Yoshiyama, K. O., Sakaguchi, K., & Kimura, S. (2013). DNA damage response in plants: conserved and variable response compared to animals. *Biology (Basel)*, *2*(4), 1338-1356. https://doi.org/10.3390/biology2041338

You, J., & Chan, Z. (2015). ROS Regulation During Abiotic Stress Responses in Crop Plants. *Front Plant Sci*, *6*, 1092. https://doi.org/10.3389/fpls.2015.01092

Zhou, B. B., & Elledge, S. J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature*, *408*(6811), 433-439. https://doi.org/10.1038/35044005

Zou, L., & Elledge, S. J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science*, *300*(5625), 1542-1548. https://doi.org/10.1126/science.1083430