2019-03-01

Interactions Between the Organellar Pol1A, Pol1B, and Twinkle DNA Replication Proteins and Their Role in Plant Organelle DNA Replication

Stewart Anthony Morley
Brigham Young University

Interactions Between the Organellar Pol1A, Pol1B, and Twinkle DNA Replication Proteins and Their Role in Plant Organelle DNA Replication

Stewart Anthony Morley

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Brent L. Nielsen, Chair
Laura Bridgewater
Joel S. Griffitts
William R. McCleary
Perry Ridge

Department of Microbiology and Molecular Biology

Brigham Young University

ABSTRACT

Interactions Between the Organellar Pol1A, Pol1B, and Twinkle DNA Replication Proteins and Their Role in Plant Organelle DNA Replication

Stewart Anthony Morley
Department of Microbiology and Molecular Biology, BYU
Doctor of Philosophy

Plants maintain organelle genomes that are descended from ancient microbes. Ages ago, these ancient microbes were engulfed by larger cells, beginning a process of co-evolution we now call the endo-symbiotic theory. Over time, DNA from the engulfed microbe was transferred to the genome of the larger engulfing cell, eventually losing the ability to be free-living, and establishing a permanent residency in the larger cell. Similarly, the larger cell came to rely so much on the microbe it had engulfed, that it too lost its ability to survive without it. Thus, mitochondria and plastids were born. Nearly all multicellular eukaryotes possess mitochondria; however, different evolutionary pressures have created drastically different genomes in plants versus animals. For one, animals have very compact, efficient mitochondrial genomes, with about 97% of the DNA coding for genes. These genomes are very consistent in size across different animal species. Plants, on the other hand, have mitochondrial genomes 10 to more than 100 times as large as animal mitochondrial genomes.

Plants also use a variety of mechanisms to replicate and maintain their DNA. Central to these mechanisms are nuclear-encoded, organelle targeted replication proteins. To date, there are two DNA polymerases that have been identified in plant mitochondria and chloroplasts, Pol1A and Pol1B. There is also a DNA helicase-primase that localizes to mitochondria and chloroplasts called Twinkle, which has similarities to the gp4 protein from T7 phage. In this dissertation, we discuss the roles of the polymerases and the effects of mutating the Pol1A and Pol1B genes respectively. We show that organelle genome copy number decreases slightly and over time but with little effect on plant development. We also detail the interactions between Twinkle and Pol1A or Pol1B. Plants possess the same organellar proteins found in animal mitochondria, which are homologs to T7 phage DNA replication proteins. We show that similar to animals and some phage, plants utilize the same proteins in similar interactions to form the basis of a DNA replisome. However, we also show that plants mutated for Twinkle protein show no discernable growth defects, suggesting there are alternative replication mechanisms available to plant mitochondria that are not accessible in animals.
Keywords: Arabidopsis, DNA replication, plant organelles, DNA polymerase, DNA helicase-primase, qPCR, yeast-two-hybrid

ACKNOWLEDGEMENTS

Graduate work is hard, and it would be much harder without my family. Firstly, my wife Danyelle, you have been patient and supportive while I vent frustrations and celebrate successes. Without you, I have nothing. My parents, brother, and sisters have all loved me unfailingly despite surpassing me in many ways. My friends for telling me I'm smart even though I think I have them fooled.

Years ago when I began my Ph.D. at BYU I wasn't sure I'd be able to work in the lab I liked. Dr. Nielsen did not need any new Ph.D. candidates and was working on finishing with his current graduate student. However, he did accept me, and I am forever grateful for the years he spent with me.

I thank my graduate committee for guiding me towards fruitful research avenues and discouraging me from stupid ones. Sweating in front of them every six months taught me that I had little to fear if I worked hard. Thanks to the wonderful faculty of the Microbiology and Molecular Biology Department who are always happy to answer my questions whenever I pop in.

And lastly, thank you BYU and Provo. You have been part of my life for nearly a decade. I will be happy to move on, but also sad to see you go.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

Discovery of the organelles

In 1665 Robert Hooke became the first person to observe cells with a simple microscope.[1] Almost one hundred and fifty years later in 1804 Franz Bauer described the discovery of the first observed organelle, the nucleus.[2] Finally, it would take nearly eighty more years before Richard Altmann would publish and describe what he called "bio-blasts" in 1890 or what we now call mitochondria.[3] Around the same time in 1883 A. F. W. Schimper described "chloroplastids", what we now know as chloroplasts.[4]

Today it's common knowledge that nearly all eukaryotes possess mitochondria. Using the citric acid cycle[5] and oxidative electron transport chain[6] these organelles produce ATP, which the cells use as a main source of energy. In addition to mitochondria, plant possess chloroplasts which serve a similar but slightly different role. Chloroplasts turn light energy into chemical energy via photosynthesis and fix carbon into sugars using the Calvin cycle.[7] These seemingly simple discoveries were extremely important to our understanding of cell physiology; however, there was still a major discovery to be made. This was the discovery of organellar DNA in mitochondria and chloroplasts (mtDNA and ctDNA respectively).[8,9]

Evolutionary origins of each organelle

Once we knew mitochondria and chloroplasts possess their own unique DNA two crucial questions arose: Why do organelles have their own DNA and where does it

come from? The answer to these questions takes shape in the endosymbiotic theory first proposed by Konstantin Mereschkowski in 1905[10]. This theory states that through endosymbiosis, a larger cell engulfed a smaller one and rather than destroying it, developed a symbiotic relationship with it. Over time, more and more DNA from the smaller engulfed cell was shuttled to the nucleus of the larger cell until it eventually lost the ability to be free living, transforming into an organelle. Likewise, the larger cell became so dependent on the smaller engulfed cell that it too could not survive without it. This theory was developed before knowing these organelles contained DNA, a remarkable achievement.

In general terms, mitochondria are descendants of ancient proteobacteria and cyanobacteria, respectively.[11,12] Access to DNA sequence data of chloroplasts and mitochondria provide more insight into the endosymbiotic theory. Based on phylogenetic data, mitochondria are thought to be descendants of *Rickettsiales*.[13] *Rickettsiales* is an order of small alphaproteobacteria that form endosymbionts with eukaryotic cells. Culturing *Rickettsiales* is difficult, as they require a eukaryotic host to survive. Furthermore, microbes in the *Rickettsiales* order actively undergo genome reduction of their already relatively small genome (<1.5 Mbp).[14] These facts provide strong support that *Rickettsiales* represents an ancient ancestor of modern mitochondria.

Chloroplast origins are more complicated. Currently, scientists believe there are three lineages of chloroplast that split from one endosymbiotic event; *Glaucophyta*,

Rhodophyceae, and *Chloroplastida*.[15,16] *Glaucophytes* form the smallest of the three lineages, consisting of freshwater algae and is also thought to be the first group to branch off.[17] *Rhodophytes* consist of a diverse group of red algae whose chloroplasts are also called rhodoplasts, translating literally to "red chloroplasts." Although a major lineage of chloroplasts, *Rhodophytes* do not represent any of the chloroplasts found in plants.[18] Plant chloroplasts are represented by *Chloroplastida*, the largest and most diverse lineage. Host organisms for *Chloroplastida* include green algae and land plants. In general, chloroplast discussions revolve around the *Chloroplastida* lineage.[19]

DNA replication in plant and animal organelles

Genome size

Endosymbionts are subject to genome modification via the following processes: mutation, selection, genetic drift, and recombination.[20] Almost all endosymbionts have their genomes reduced in size as genetic material gets transferred to the nucleus of the host organism. This has been observed in many bacterial species[21,22] and other studies show that reduction of the endosymbionts' genome can promote gene transfer to the nucleus.[23]

The same process has occurred in eukaryotic mitochondria and chloroplasts. However, mitochondrial genomes in plants have evolved almost in complete contrast to animal mitochondrial genomes (Figure 1-1). Most animal mitochondrial genomes are roughly 16 kb in size and seem to have been selected for the economy of their small

size.[24] The number of DNA copies per organelle varies from study to study. Older estimates place as many as 10 copies per organelle[25] whereas more recent data suggests it may be as low as 1.[26] Regardless of the actual number, mitochondrial genome copy number is thought to be tightly regulated in animal cells.[27]

In contrast, plant mitochondrial genomes are not only much bigger, but can vary in size from species to species. Sizes of plant mitochondrial genomes range from 187 to 2400 kb in size[28] and the number of copies per organelle varies widely based on tissue type, development, and in some cases, researcher opinion. One study observed certain tissues possessed less than 1 genome copy per mitochondrion whereas other tissues contained over 100 per organelle.[29] The same study also noted that as the plants developed, mtDNA levels declined. However, contrary to this study, a separate publication found that genome copies per mitochondrion varied between 40 in young leaves to 280 in mature leaves in *Nicotiana tabacum*.[30] Consistent in all of these studies is a higher concentration of genome copies per mitochondrion in root tip tissues than in other parts of the plant. The high variance in size and copy number of plant mtDNA suggests that these genomes evolved over a much longer period of time than their animal counterparts. Although the exact copy number appears to vary widely, we can safely assert that plants maintain much higher genome copy number per mitochondrion than animals. Why plants maintain such large mitochondrial genomes at such high levels is still a subject of debate, with no clear answer.

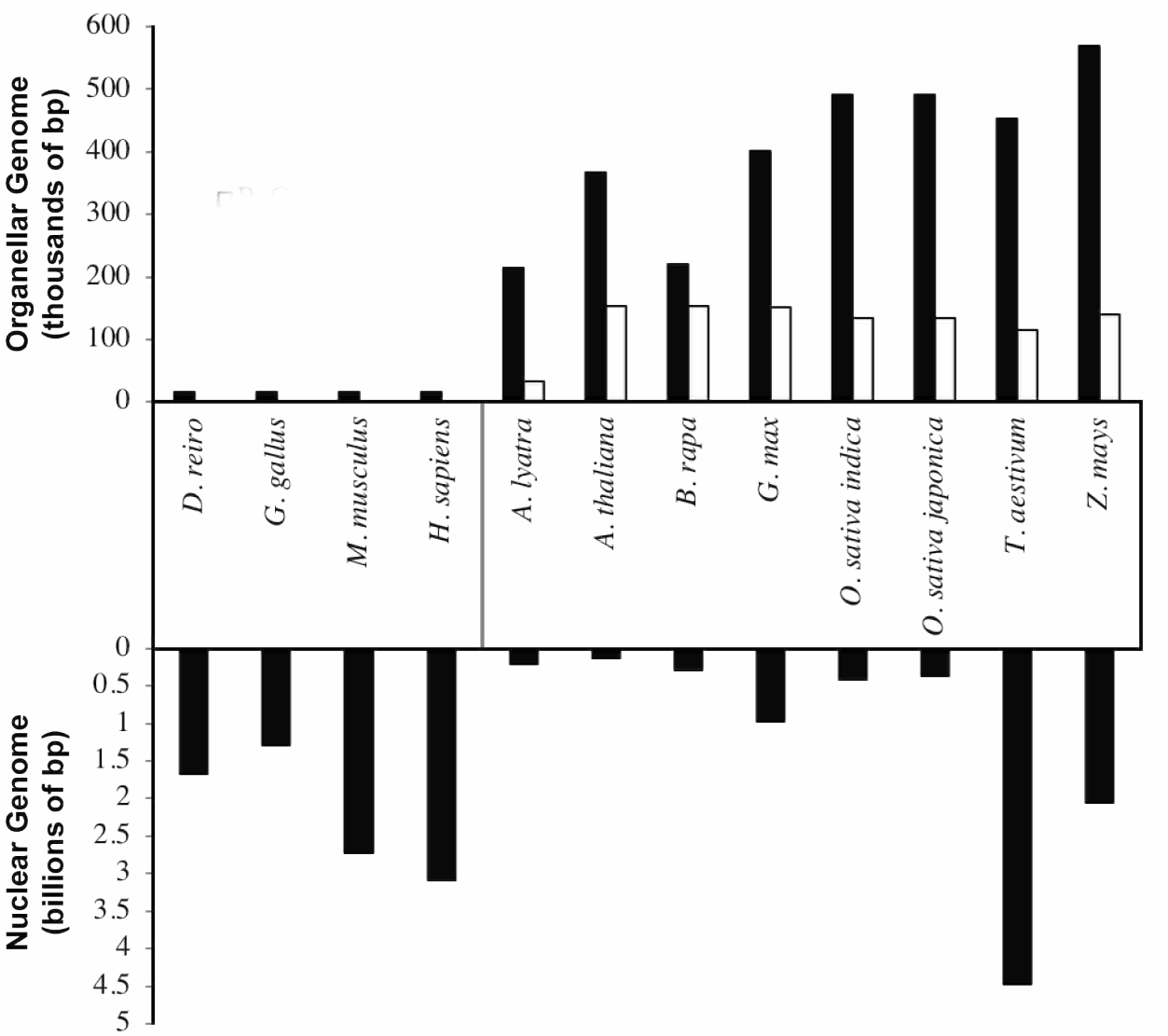

Figure 1-1. Nuclear and organelle genome sizes among different organisms. Mitochondrial genomes among animals are compact and remarkably similar in size; ~16.5 kb. Plants however have mitochondrial genomes that dwarf those found in animals and vary in size from species to species. Chloroplast genomes vary less in size from organism to organism but still are relatively large compared to animal mitochondrial genomes. Organelle genome sizes are independent of nuclear genome size.

Compared to mitochondria, chloroplast genomes are fairly uniform ranging between 120 and 160 kb in size with some exceptions as large as 2000 kb.[31,32] This uniformity indicates that genome reduction in chloroplasts took place in a relatively

short period of time soon after endosymbiosis.[33] Genome copy numbers for chloroplasts seem to be more related to plastid size rather than tissue type. Genome copy number estimates per chloroplast range anywhere from several hundred to nearly two thousand per organelle.[34,35] Substantial evidence has also shown that as plants age, ctDNA levels decline.[36-39]

Genome structure and content

The coding content of plant versus animal organelles varies significantly. For the most part, animal mtDNA possesses the same 37 genes; two for rRNAs, 13 for proteins and 22 for tRNAs (Table 1-1).[24] All 37 of these genes possess homologs in plants, fungi, and protists. To date, mtDNA gene content among animals only varies in nematodes,[40] a bivalve,[41] and cnidarians.[42] In these exceptions, there have been losses and gains of different mitochondrial genes, mostly tRNA genes.

In contrast to animals, plant mtDNA also contains many more genes and large portions of non-coding or undefined DNA.[43] A typical plant mitochondrial genome encodes anywhere between 50 and 100 genes (Table 1-1)[44] but this does not account for the large size of the genomes we observe. The amount of non-coding DNA in plant mitochondria can vary widely but in general nearly half cannot be assigned a function.[45] Much of this non-coding DNA is made up of introns, repeats, and duplications of regions of the genome.[43] The known genes consist of rRNA and tRNA genes as well as subunits for oxidative phosphorylation chain complexes.[46] Considering the harsh

selection animals made toward compact mitochondrial genomes, it's interesting to observe how plants seemingly maintained or even duplicated introns in theirs. This includes small non-coding DNA repeats found throughout plant mitochondrial genomes. These small repeats can often undergo expansion, duplicating themselves multiple times. The leading hypothesis explaining this phenomenon suggests a bias in conserving genes to keep mutation rates low, whereas break-induced replication in noncoding regions explains expansion and rearrangements.[47]

In general, genes are conserved in most chloroplast genomes. For the most part these consist of rRNA, tRNA, and genes involved in photosynthesis (Table 1-1).[31,32] Loss of genes seems to be the only difference in gene content when comparing genomes. In these cases, essential genes have been lost from the chloroplast genome and transferred to the nucleus. Considering the three lineages of chloroplast genomes, it's interesting to observe the relatively conserved number and type of genes found in these genomes. Such patterns suggest key genes are crucial to maintaining multi-subunit complexes or that significant barriers exist to transporting them outside the organelle environment.

Table 1-1. Genome size and content of model plants, fungi, algae, and animals

Species	mtDNA genome size	Genes encoded			Ref.
		Protein	rRNA	tRNA	
Land Plants					
Arabidopsis thaliana	366,924	33/117*	3	21	[48]
Beta Vulgaris	368,801	27/140*	5	26	[49]
Brassica rapa	219,747	34/78*	3	18	[50]
Glycine Max	402,558	36/88*	3	19	[51]
Gossypium raimondii	676,078	39	6	30	[52]
Nicotiana tabacum	430,597	37/156*	4	23	[53]
Oryza sativa	491,515	33/54*	6	33	[54]
Triticum aestivum	452,528	35/39*	9	25	[55]
Zea Mays	569,630	39/163*	4	29	[56]
Fungi					
Ashbya gossypii	23,564	8	2	23	[57]
Neurospora crassa	64,840	22/28*	-	28	[58]
Saccharomyces cerevisiae	78,917	8	2	24	[59]
Schizosaccharomyces pombe	19,431	6/10*	2	25	[60]
Algae					
Chlamydomonas reinhardtii	15,758	8	14	3	[61]
Dictyostelium discoideum	55,564	33/42*	2	18	[62]
Animal					
Homo sapiens	16,569	13	2	22	[63]
Mus Musculus	16,299	13	2	22	[64]

*Includes predicted/hypothetical genes that have not been annotated or reviewed.

Animal mtDNA consists of a singular circular molecule[65] and is very gene dense, with about 97% of the DNA coding for functional genes.[63,66] In most animals, the small non-coding region is a control region that has important elements to regulate DNA replication and gene transcription.[67] One large exception to these norms can be seen in non-bilaterian animals, which possess large segments of non-coding DNA as well as varying levels of linear and circular DNA molecules.[66]

Although plant mitochondrial genomes are mapped as circular molecules (sometimes called master circles), circular molecules equal to a genome equivalent have only been observed in cultured liverwort cells.[68] Typically, plant mtDNA is observed primarily as large sub-genomic linear molecules. Other structures found in lower abundance include lariats, rosette like structures, catenane molecules, and branched linear molecules.[69-71] When observed on a pulsed field gel, a large collection of plant mitochondrial DNA remains locked in highly complex arrangements.[72] Viewed by electron microscopy, these complex arrangements form DNA 'rosettes' and branched molecules indicating high levels of recombination. Other high molecular weight plant mtDNA simply does not enter the gel at all and has been theorized to be relaxed circle DNA, other replication intermediates, or DNA somehow bound to a matrix of other materials.

Plant mtDNA is also subject to much more rearrangement due to the many repeat regions found throughout the genome.[73] An example of this can be observed in *Arabidopsis thaliana* ecotypes Col-0 and C24. These two ecotypes have genetically identical mitochondrial genomes, but arrange their genes in different orders. Interestingly, this rearrangement occurred recently, only 200,000 years ago.[47] Despite the high rate of rearrangement in plant mtDNA, the mutation rates of these genomes are extremely low. This is directly contrary to patterns observed in humans and animals, in which the mutation rate in the mitochondrial genome is high enough that

different ethnicities and population can be tracked completely through sequencing of the mitochondrial genome. This pattern of strict gene arrangement and no recombination is maintained throughout the animal kingdom.[74]

Chloroplast genomes are much simpler and exist primarily as homogeneous closed circle DNA molecules.[75,76] A small portion of these molecules exist as circular dimers.[77] One exception to these observations can be seen among two species of brown algae.[78] These organisms display a collection of circular chloroplast molecules that differ in size. The hypothesis in this instance is that the collection of these heterogeneous molecules make up the entirety of the chloroplast genome of these brown algae.

Most chloroplast genomes possess a large inverted repeat. In spinach, lettuce, and corn, this repeat is between 22–24 kb in length, nearly 16% of the chloroplast genome.[77] Studies have shown that when this repeat is removed there are more recombination events and fewer nucleotide substitution events.[79,80] Therefore, these large inverted repeats are thought to be used by the chloroplasts as a way to maintain fidelity, and correct mutations or errors in replicated DNA.

Genome replication mechanisms

In animal mitochondria

Several modes of DNA replication in animals have been proposed (Figure 1-2). These include rolling circle, theta replication, strand-displacement, and RITOLS (Ribonucleotide Incorporation ThroughOut the Lagging Strand)/bootlace.[81] Rolling

circle replication assures efficient reproduction of genomes exploiting a bacteriophage-like mechanism. However, amongst animals only nematode worms have been observed employing rolling circle DNA replication in mitochondria.[82] Theta replication can be commonly observed among other invertebrates and was first characterized in *Drosophila melanogaster*.

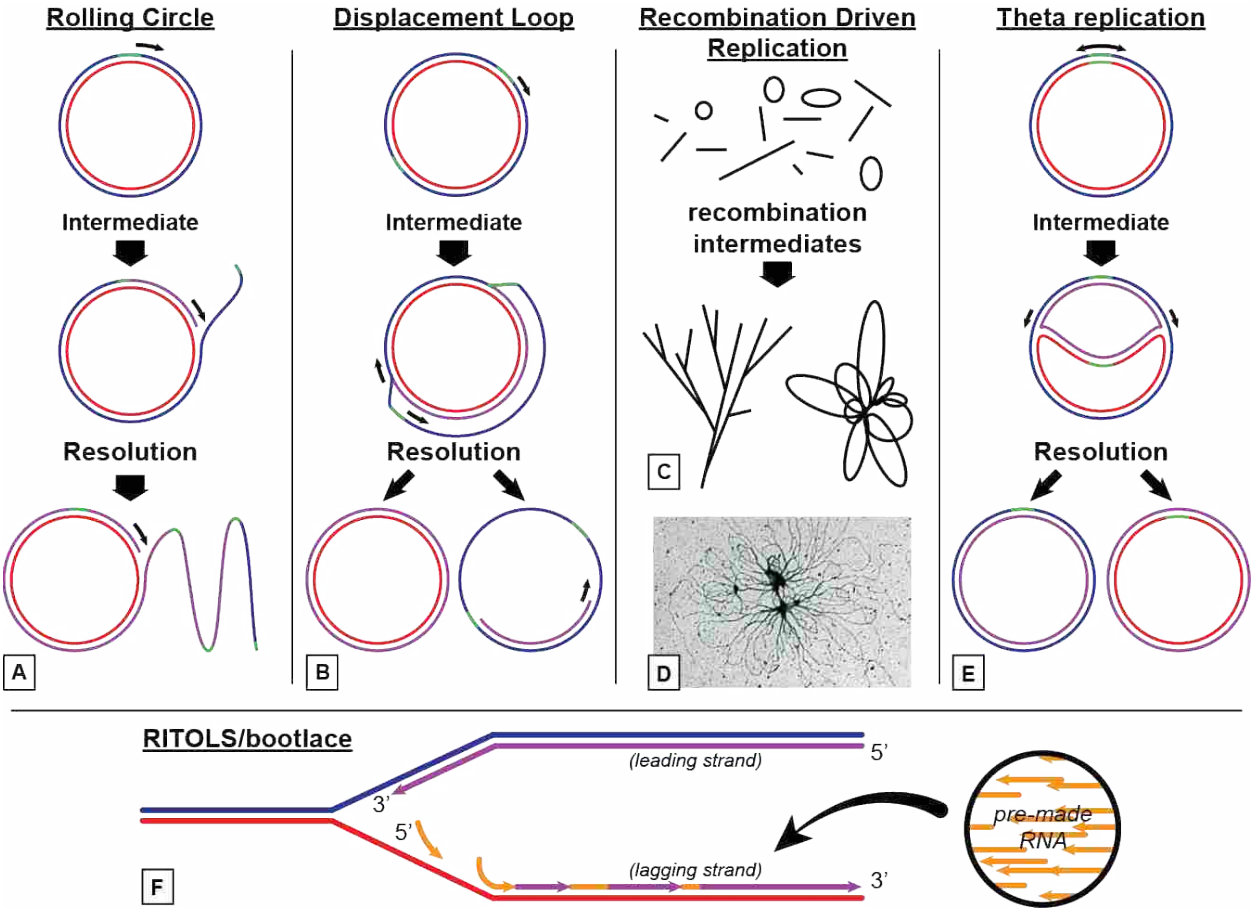

Figure 1-2. Replication methods of mitochondrial and chloroplast DNA. (A) Rolling circle replication involves unidirectional replication after DNA nicking of one strand of DNA. DNA replication continues along the circular molecule displacing the nicked strand. Upon reaching the initial start site, the displaced strand may be nicked and ligated to form a new single stranded circular molecule or synthesis may continue, creating a linear concatemeric molecule which is later converted into multiple single stranded circular copies of the parent molecule. (B) Similar to rolling circle replication, Displacement loop replication proceeds unidirectionally by first displacing one of the strands of DNA. Unlike rolling circle, the displaced strand is not nicked

and maintains its circular form. Upon synthesizing a certain portion of the genome (commonly 2/3) a second origin site is revealed that allows synthesis in the opposite direction. By the time the first double stranded DNA molecule is finished, synthesis on the parent strand is still completing. Once replication reaches the initial start site, the parent strand is displaced as a single stranded circular DNA molecule. The single stranded circular molecules formed by rolling circle and displacement loop replication are later turned into double stranded copies by DNA replication machinery. (C) Recombination Driven Replication involves the use of many linear and circular pieces of DNA that share homology. These pieces recombine to form branched linear and "rosette" like intermediates that are copied and replicated by DNA machinery. (D) Electron Micrograph image of DNA forming a "rosette" as a result of recombination. (E) Theta replication is so named because of the intermediate it forms as a result of bi-directional DNA replication. Replication initiates bi-directionally at an origin of replication, forming two replication forks. When these replication forks meet, the two double stranded circular molecules are separated. (F) The RITOLS (Ribonucleotide Incorporation ThroughOut the Lagging Strand)/Bootlace strategy of replication involves the lagging strand of a replication fork. While the leading strand replicates normally, pre-synthesized RNA molecules hybridize to the lagging strand of DNA. Gaps are filled in and the primers removed by DNA replication machinery.

In nearly all vertebrate animals, mitochondrial DNA is replicated via a mechanism called strand-displacement replication, commonly referred to as D-loop displacement. In this method, a single strand of DNA is displaced as replication proceeds. When nearly two thirds of the molecule is replicated, the site for lagging strand synthesis is revealed and initiates.[83] Simultaneous replication forks have also been observed. Replication via strand displacement or D-loop displacement always initiates from the non-coding control region of animal mtDNA.

RITOLS/bootlace replication is another form of strand displacement replication that involves a unique and rather odd mechanism. RITOLS was coined after scientists observed replication intermediates that were resistant to DNA endonucleases but sensitive to RNaseH.[84] Later, they found that a substantial amount of RNA was present

in these replication intermediates.[85] Thus they proposed that pre-synthesized RNAs hybridized to the lagging strand of DNA synthesis rather than being synthesized concomitantly with the leading strand. These RNAs hybridize to ensure there are no gaps of single stranded DNA.

In all of the above mentioned cases animals utilize a simple minimal DNA replisome. This replisome has been reconstituted *in vitro* and is made up of DNA helicase TWINKLE, and DNA polymerase POL γ.[86] Together, these two enzymes are fairly processive and can create molecules about 2kb in length. The addition of single strand binding protein to TWINKLE and POL γ increases the processivity of this replisome to create genome sized molecules of 16 kb.

In plant mitochondria

Plants most likely use multiple mechanisms within the same mitochondrion due to the complex structure of their mtDNA. The structure of plant mitochondrial DNA makes strand displacement or D-loop replication implausible although there is one instance of this mechanism observed in petunia flowers.[87] Rolling circle replication has also been observed in *Chenopodium album* suggesting there may be more widespread use in other plants as well.[88,89] The main difference that distinguishes plants from animals is the use of recombination to initiate DNA replication. In contrast to animals which always begin replicating mtDNA in the non-coding control region of the genome, we cannot predict where mitochondrial DNA replication initiates in plants. This is due to

the large amount of non-coding DNA, repeat regions, and complex replication intermediates frequently observed in plants. Because of this, many scientists hypothesize that the main mechanisms plants use for mtDNA comes from a combination of recombination driven replication (RDR) and recombination independent rolling circle replication.

The polymerases used to replicate plant mtDNA are low-fidelity DNA polymerases. Interestingly, despite their low-fidelity and the propensity of plant mitochondrial genome to undergo frequent rearrangements, plant mtDNA remains remarkably resistant to mutations in coding regions. This could be due to multiple DNA replication strategies and mechanisms that are not present in animal mitochondria, which use only one set of enzymes. The complexity of plant mtDNA replication may actually serve as a boon to the plant, allowing multiple synthesis and repair pathways to correct mutations.

The size, complexity, and variation of plant mtDNA from species to species make it difficult to clearly define essential replication mechanisms. A lack of mutants and easily accessible genetics make certain details such as recombination machinery, initiations sites, and DNA replisome proteins vague and unclear.

In plant chloroplasts

ctDNA replication is much less complicated and better understood than plant mitochondrial DNA replication. Chloroplasts utilize a double displacement loop

strategy to initiate DNA replication.[90] The two displacement loops begin on opposite strands and begin replicating unidirectionally towards each other until they meet. At this point the displacement loops fuse forming a Cairns or theta structure and DNA replication continues bidirectionally until two daughter molecules are created.

Some exceptions to the double D-loop replication model exist but they do not involve complicated recombination, rolling circle, or branched linear structures we see in plant mitochondria. For example, *Chlamydomonas* and *Oenothera* possess two displacement loops, but discontinuous DNA replication begins shortly after initiation rather than after the fusion of the two D-loops.[91,92] *Euglena* possesses only one origin of replication site and appears to replicate bidirectionally from this site rather than forming two displacement loops.[93]

Similarity to T7 bacteriophage

T7 phage replicates DNA with a simple DNA replisome (Figure 1-3) consisting of proteins gp5 (T7 DNA polymerase), gp4 (DNA helicase/primase) and gp2.5 (DNA single stranded binding protein). *E. coli* thioredoxin also binds to gp5 to increase the processivity of the enzyme.[94] Animal mitochondria use a similar system consisting of DNA POL γ, TWINKLE, and SSB1 protein.[95] Since plant organelles possess the same proteins, one could logically assume that the same replisome is tasked with maintaining and replicating DNA in chloroplasts and plant mitochondria. However, while Twinkle knockouts in animals cause embryo lethal effects, Twinkle knockouts in plants lead to

no distinguishable phenotype. Genome copy numbers in organelles also remain unchanged. This presents a fatal flaw to the idea that a Twinkle-Pol1A/B replisome is the main driver of DNA synthesis in plant organelles. This also highlights the likelihood of plants utilizing multiple methods to replicate organellar DNA rather than a dependence on one mechanism.

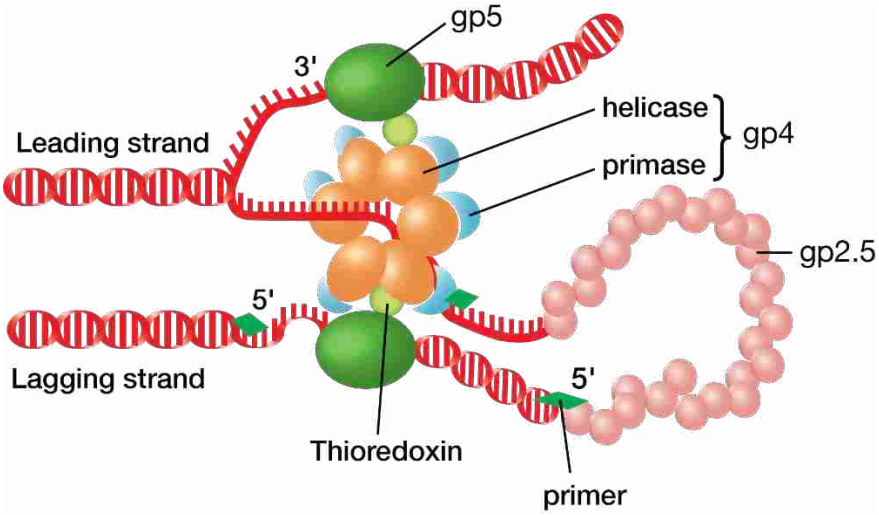

Figure 1-3. Minimal DNA replisome of T7 phage. There are four proteins involved in the minimal DNA replisome of T7 phage. These are; gp5 DNA polymerase, gp4 DNA helicase/primase, gp2.5 single stranded binding protein, and E. coli thioredoxin. Thioredoxin is the only host protein involved in this replisome and binds to gp5. Without thioredoxin, gp5 loses processivity.

Organelle DNA replication proteins

Table 1-2 summarizes the necessary proteins and functions required for DNA replication and maintenance in plant organelles.

Table 1-2. Proteins involved in organellar DNA replication in *Arabiodpsis thaliana*

Function	Protein Name	TAIR	Homology	Localization*	Ref.
DNA polymerase	Pol1A or Pol gamma 2	At1g50840	Bacterial	M, P	96-98
	Pol1B or Pol gamma 2	At3g20540	Bacterial	M, P	96-98
Helicase	Twinkle	At1g30680	Phage	M, P	99,100
	DNA2	At1g08840	Mammalian	?	101,102
Priming	Twinkle	At1g30680	Phage	M, P	99,100
	RNA polymerase:				103-106
	RpoT1	At1g68990	Phage	M	
	RpoT2	At5g15700	Phage	M, P	
	RpoT3	At2g24120	Phage	P	
	rpoA	AtCg00740	Bacterial	P	
	rpoB	AtCg00190	Bacterial	P	
	rpoC1	AtCg00180	Bacterial	P	
	rpoC2	AtCg00170	Bacterial	P	
Primer Removal	RNaseH				
	AtRNH1B	At5g51080	Bacterial	M, P	107
	AtRNH1C	At1g24090	Bacterial	P	107
	EXO1	?			108
	EXO2	?			108
ssDNA binding, recombination monitoring	SSB1	At4g11060	Bacterial	M, P	46,109
	SSB2	At3g18580	Bacterial	M	46
	OSB1	At3g18580	Bacterial-like, but unique to plants	M	110
	OSB2	At4g20010	Unique to plants	P	46
	OSB3	At5g44785	Unique to plants	M, P	104,110
	OSB4	At1g31010	Unique to plants	M	110
	WHY1	At1g14410	Unique to plants	P	111-113
	WHY2	At1g71260	Unique to plants	M	111-113
	WHY3	At2g02740	Unique to plants	P	111-113
	ODB1	At1g71310		M, N?	

	ODB2	At5g47870		P, N?	
Recombination	RecA1	At1g79050	Bacterial	P	[114]
	RecA2	At2g19490	Bacterial	M, P	[114]
	RecA3	At3g10140	Bacterial	M	[114,115]
	MSH1	At3g24320	Bacterial	M, P	[116]
Topoisomerase	Topoisomerase I	At4g31210	Bacterial	M, P	[104]
	DNA Gyrase A	At3g10690	Bacterial	M, P	[117]
	DNA Gyrase B1	At3g10270	Bacterial	P	[117]
	DNA Gyrase B2	At5g04130	Bacterial	M	[117]
	DNA Gyrase B3	At5g04110	Eukaryotic	N	[117]
Ligation	LIG1	At1g08130	Bacterial	M, N	[46]

* Localization to M (mitochondria) or P (plastids)

Key functions required for DNA replication in plant organelles include; polymerization, DNA unwinding, priming, strand separation, recombination, and ligation. These functions are carried out by nuclear encoded proteins that target to either the mitochondria, chloroplasts, or both. For the sake of simplicity we will only discuss those replication proteins described in *Arabidopsis* as homologs exist in all vascular plants.

An interesting point to mention is that DNA replication proteins in plant organelles have interesting ancestral sources. For example, the DNA polymerases are bacterial in origin, however Twinkle DNA helicase-primase and RNA polymerases are phage-like. Other proteins like the Whirly class of single strand DNA binding proteins are unique to plants. Thus we see an interesting mosaic of proteins from different clades of life involved in replicating organellar DNA.

Polymerization

To date, two organellar DNA polymerases have been discovered in both mitochondria and chloroplasts. The nomenclature of these plant organellar DNA polymerases has a convoluted history. Initially, they were called Polymerase gamma 2 and 1, in reference to human mitochondrial DNA POL γ.[118] However, further studies disputed this characterization and found they had more in common with bacterial DNA polymerase I; so they are also commonly referred to as Pol1A and Pol1B.[119] Furthermore, other studies frequently refer to these proteins as plant organellar DNA polymerases, or POPs.[120] Thus, we have three different names for the same protein. We will refer to the proteins as DNA polymerase 1A and 1B. These are the only DNA polymerases known to function within the chloroplasts and mitochondria in plants.

When comparing *Arabidopsis* Pol1A and Pol1B with *E. coli* DNA polymerase I, several notable structural differences can be observed. The most obvious is that the plant polymerases lack a 5′-3′ exonuclease domain present in *E. coli* DNA polymerase I. In its place is a long sequence of amino acids with no functional assignment. It is possible that as plants evolved this 5′-3′ exonuclease function became irrelevant and was not maintained as mutations accumulated in this domain of the polymerases. The second difference between the plant and bacterial DNA polymerases is a stretch of amino acids that has been inserted between the 3′-5′ exonuclease domain and the polymerase domain.

Although Pol1A and Pol1B are very similar to each other, notable differences between the two have been observed. Pol1B knockouts grow slightly slower and have fewer genome copy numbers per organelle than Pol1A knockouts.[121] Additionally, Pol1B mutants show increased sensitivity to double stranded DNA breaks.[119] However, recent studies show that Pol1A replicates DNA with more fidelity and has an increased ability to displace DNA when replicating over short single stranded gaps of DNA.[122,123] When determining the importance of each polymerase, these two findings seem to contradict each other. The growth delay and lower genome copy numbers in Pol1B mutants suggests Pol1B is more essential to DNA replication, however the greater fidelity and strand displacement displayed by Pol1A goes contrary to this statement. It's important to note that a double mutant for both DNA polymerases has not successfully been created and is seed lethal. However, heterozygous plants containing a single copy of either Pol1A or Pol1B are able to grow to maturity. In summary, it appears that Pol1A is a more processive DNA polymerase than Pol1B, which is more involved in DNA damage repair. This conclusion is somewhat circumstantial, as this hypothesis has not been directly tested in a singular study.

Pol1A and Pol1B are processive enough to replicate an entire genome equivalent in both mitochondria and chloroplasts.[119,124] They are much more processive than *E. coli* DNA polymerase I which is involved mostly in Okazaki fragment processing and DNA repair and cannot replicate much more than several dozen bases.[125] Interestingly,

recombinant versions of *E. coli* DNA polymerase I that are able to bind thioredoxin display a dramatic increase in processivity.[126] This increase in processivity mimics the behavior of T7 DNA polymerase (gp5) which binds thioredoxin to improve processivity. Plant organelle DNA polymerases may also bind thioredoxin to achieve high processivity, although this has yet to be shown. If the plant enzymes bind thioredoxin it would also help explain why they are so much more processive than their cousin found in *E coli.*

Both Pol1A and Pol1B are able to bypass DNA lesions and continue replicating DNA. Typically, translesion bypass DNA polymerases do not replicate DNA with great fidelity, for example POLQ DNA polymerase in humans makes an error approximately every 200 bases and many yeast translesion DNA polymerases have error rates even higher than this.[127,128] For comparison, human POLG makes an error approximately every 100,000 – 1,000,000 bases. *Arabidopsis* Pol1B will misincorporate one nucleotide approximately every 2,000 bases, much more error-prone than POLG, but highly faithful compared to other translesion synthesis DNA polymerases. One reason for the greater fidelity is likely because Pol1A and Pol1B possess 3'-5' proofreading exonuclease domains which are typically not present in translesion synthesis DNA polymerases.

Three unique amino acid insertions have been identified in both Pol1A and Pol1B that confer translesion activity. Two of these insertions exist in the 'thumb'

domain of the polymerase (insertions 1 and 2) and the third resides in the 'finger' domain (insertion 3). These appear to be flexible elements as mutants lacking all three of these insertions are still able to synthesize DNA.[128] The translesion activity of Pol1A and Pol1B is negatively affected by the removal of insertions 1 and 3, indicating that these enzymes acquired translesion activity through evolution and the acquisition of these insertions.

Ligation

The amount of information surrounding plant organellar DNA ligases is extremely lacking and understudied. Although we have some information concerning DNA ligases in plant mitochondria, no DNA ligase has been confirmed or observed functioning in plastids. This presents a potential avenue of research, as the activity of this enzyme in both organelles must be present, but little to no work physically identifies the enzyme responsible.

Four DNA ligase genes have been identified in the genome of *Arabidopsis*; however, only DNA ligase 1 (LIG1) has been identified in mitochondria. If LIG1 is transcribed from a second alternate start codon it localizes to the nucleus.[129] No confirmation of LIG1 activity or any other DNA ligase has been confirmed in chloroplasts. Plant LIG1 knockouts are seed-lethal and knockdown mutants display severe growth defects due to effects on the nuclear genome rather than the mitochondrial genome.[130]

DNA ligase III (LIG3) in animals has been shown to have important roles in mitochondria. LIG3 knockouts in mice are embryo lethal, and have mitochondria with abnormal morphology and reduced DNA content.[131] The same study shows that LIG3 is essential for mitochondrial DNA integrity but not for Xrcc-1 mediated nuclear DNA repair. A homolog to LIG3 with a potential mitochondrial localizing signal has been identified in the genome of *Hordeum vulgare* but a cDNA has not been identified.[132] LIG3 has not been observed in *Arabidopsis* and most other plants.

DNA unwinding

In *Arabidopsis,* two well studied helicases unwind DNA in mitochondria and chloroplasts: Twinkle, and DNA2. *Arabidopsis* also employs several Gyrases to relieve tension in DNA. There may be other DNA unwinding enzymes active in the organelles but these have not been identified. Twinkle and DNA2 are assumed to be responsible for most of this activity.

Twinkle gets its name from the bacteriophage T7 gp4 DNA primase/helicase protein (T7 gp4-like protein With Intramitochondrial Nucleoid + K + Localization + E).[133] Like its name implies, Twinkle is a dual function protein that unwinds and primes DNA for synthesis. When unwinding DNA, Twinkle forms a hexamer or 'lock washer' around a single strand of DNA. This hexamer is not a perfect ring, with a gap between the first and last unit of Twinkle, taking on the appearance of a lock washer.[134] In T7 phage, Twinkle unwinds DNA using a 'hand-over-hand' mechanism in which the

lagging Twinkle protein within the hexamer translocates 2 nt along the DNA by ATP hydrolysis (Figure 1-4). This mechanism is similar to mechanisms observed in *E. coli* DnaB helicases.

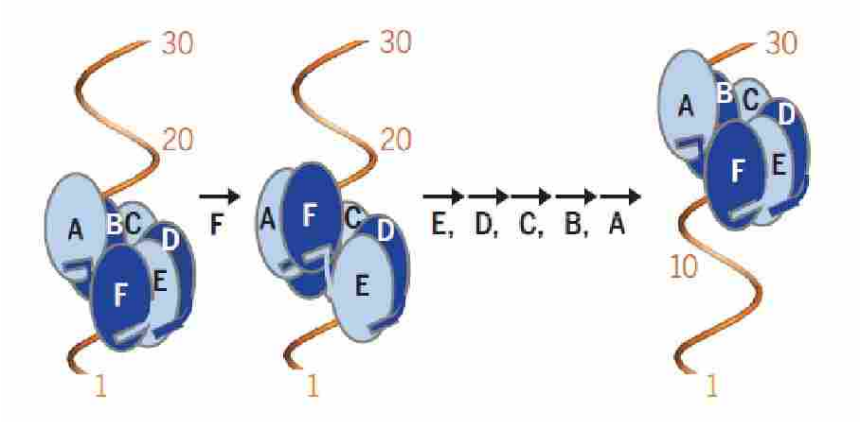

Figure 1-4. 'Hand-over-hand' mechanism of Twinkle helicase in T7 phage. In this diagram the different units of Twinkle are labeled A-F. When forming a hexamer, the ring of Twinkle proteins form a 'lock washer' in which unit A is slightly separated from unit F. The lagging unit F translocates along single stranded DNA by 2 nt upon the hydrolysis of ATP. This repeats for every lagging unit of Twinkle. Taken from Gao et al.[134]

In *Arabidopsis* and T7 phage, Twinkle possesses a zinc finger domain in the priming domain of the protein that binds DNA and synthesizes RNA primers for replication. In humans, amino acid changes to this area of the protein have removed the zinc finger domain and its priming ability.[135] In plants, both abilities remain active.[94,99,136] When compared to T7 gp4, *Arabidopsis* Twinkle has a slight extension between the primase and helicase domains when compared to phage gp4 and a longer N-terminal region. Twinkle localizes to both chloroplasts and mitochondria in plants.

Arabidopsis also possesses a truncated form of Twinkle commonly referred to as Twinky. This truncation lacks the C-terminal helicase domain of Twinkle but maintains

the primase domain. Very little work has been performed on Twinky and whether it is active in priming DNA or is simply a pseudogene is unknown.

JHS1 is the gene in *Arabidopsis* homologous to human and yeast nuclease/helicase DNA2. In humans and yeast, DNA2 cleaves single stranded DNA where there is a junction between ssDNA and dsDNA. The helicase activity of the protein unwinds these junctions to allow for cleavage to occur. Experiments with human DNA2 and DNA POL γ have shown a positive interaction and the ability to unwind DNA without cleaving the D-loop structure observed in human mitochondria.[102] Unlike the yeast version of the same protein, human DNA2 does not localize to the nucleus but to the mitochondria.

Arabidopsis DNA2 has not been heavily investigated in plants and may localize to either organelle. Because of its unique nuclease/helicase activity and unknown localization, most scientists assume the bulk of DNA unwinding activity is carried out by Twinkle protein. Homozygous DNA2 *Arabidopsis* mutants are seed lethal but this is most likely due to the important functions performed by the protein within the nucleus of the cell.[137]

Organelle gyrases

One key study by Wall et al has identified four bacterial-like DNA Gyrase genes in *Arabidopsis,* one GYRA and three GYRB genes.[117] DNA Gyrase consists of an A_2B_2 tetramer in bacteria and the same holds true in *Arabidopsis.* Interestingly, *Arabidopsis*

GYRA has two alternative translation start sites that allow it to dual-localize to both mitochondria and chloroplasts. On the other hand GYRB1 is specific to chloroplasts and GYRB2 to mitochondria. The third GYRB3 localized to the nucleus. Mutation of GYRA leads to an embryo-lethal phenotype while mutation of any of the GYRB genes leads to a seedling-lethal phenotype that varies in intensity. Interestingly, GYRA, GYRB1, and GYRB2 all show common ancestry with cyanobacteria whereas GYRB3 aligns more closely with eukaryotic type II topoisomerases. This helps explain its unique nuclear localization but also brings into question whether this gene actually encodes a functional Gyrase as some of the conserved topo regions are truncated.

Homologues of the *Arabidopsis* GYRA and GYRB genes have been identified in *Nicotiana benthamiana.* Depletion of these genes in *Nicotiana* leads to abnormally high levels of DNA in both mitochondria and chloroplasts.[138] Furthermore, DAPI staining of affected chloroplasts reveals one or a few large nucleoids rather than many small nucleoids present in wild-type plants. These results, in conjunction with those found in *Arabidopsis* indicates that organellar DNA Gyrases are critical for nucleoid partitioning by regulating DNA topology.

Priming

Arabidopsis utilizes Twinkle and RNA polymerase (RNAP) to prime organellar DNA replication. As mentioned previously, Twinkle is similar to T7 gp4 protein and possesses both helicase and primase activity. Using Twinkle to prime organellar DNA

synthesis is unique to plants, as animals utilize RNA polymerase to prime their mtDNA. Nonetheless, plants do contain organellar RNA polymerases that could complement the activity of Twinkle.

Twinkle uses a unique recognition sequence to begin ribonucleotide synthesis and appears to prefer cytosine and guanine incorporation over uracil and adenine.[136] The recognition sequence is 5′-(G/C)GGA-3′ where the underlined nucleotides are cryptic, meaning they are essential for template recognition but do not become incorporated into the extended RNA molecule. If either of the cryptic nucleotides or the guanine directly upstream from them are substituted, RNA synthesis is abolished. This is unique from other DNA primases, in that two cryptic nucleotides are required for synthesis whereas other primases often only require one. The exact mechanism of Twinkle association with template DNA is currently unknown.

Twinkle preferentially incorporates CTP and GTP, which is curious as nearly all plant mitochondrial and chloroplast genomes are highly A/T rich.[139] Why then would a plant organellar primase preferentially incorporate CTP and GTP? One theory points to *Aquifex aeolicus*, a primitive thermophilic bacteria. *Aquifex aeolicus* initiates primer synthesis from a trinucleotide sequence composed of cytosines and guanines much like *Arabidopsis* Twinkle.[140] This G-C rich sequence is hypothesized to provide stability during primer extension. Similarly, plants may rely on the stability of the template sequence 5′-(G/C)GGA-3′ paired with preferential CTP and GTP incorporation to

provide thermodynamic stability. Another leading hypothesis is that preferential incorporation of CTP and GTP aid in determining Okazaki fragment length.[136]

The co-evolution of nuclear, plastid, and mitochondrial genomes in plants has led to an interesting arrangement of RNA polymerases (RNAP) in the organelles. Unlike animal mitochondria which utilize a single RNA polymerase[141] plant organelles require multiple; at least two for plastids and one for mitochondria. These genes are designated as *"RpoT"* genes, the *"T"* indicating their similarity to the single subunit RNA polymerases of T3 and T7 phage.[142] RNAP that targets to mitochondria are designated RpoTm and those that target to plastids are called RpoTp. RpoTmp represent RNAP that target to both organelles. Different species may possess multiple copies of these nuclear encoded organellar proteins, but the earliest phylogenetic versions of these enzymes exist in the waterlily *Nuphar advena*, a basal angiosperm.[143] Mitochondria do not possess RNAP genes within their genome, therefore all RNA polymerases for the mitochondria are nuclear encoded. Plastids on the other hand use both nuclear and plastid-encoded RNAPs. Extensive research has been performed on how plant RNA polymerases recognize promoters and transcribe genes. This work focuses on the potential RNAP has to prime DNA for synthesis rather than its role in transcribing genes.

Three single-subunit mitochondrial RpoT genes have been identified in *Arabidopsis*; however, only two have been proven to localize to mitochondria.[106,142] A

duplication of one of these genes has led to the creation of a dual targeted mitochondrial-plastid RNAP (RpoTmp). How these enzymes coordinate synthesis of RNA is largely unstudied although some research suggests RpoTmp is responsible for gene synthesis in early seedling development and RpoTm and RpoTp take over once the plant has developed more fully.[144] Plant organellar RNA polymerases show high degrees of conservation with both T3/T7 phage and human orthologs.

In *Arabidopsis*, plastids require at least two RNA polymerases; one nuclear encoded RNAP, and one eubacterial plastid encoded RNAP. The nuclear and plastid encoded version of RNAP are distinct and do not share subunits.[145] The nuclear-encoded plastid RNA polymerase is homologous to the phage-like single subunit RNA polymerases found in mitochondria. This RNAP is represented as RpoTp and is thought to be a duplication of the mitochondrial RpoTm.

A nuclear-encoded RNAP has been isolated from *P. sativum* that seems to act more as a primase than an RNAP.[146] This is because the isolated enzyme was able to synthesize primers shorter than expected for transcription but larger than primers generated by other primases. The isolated enzyme also prefers binding to single-stranded rather than double stranded DNA, a feature common to primases. Finally, the isolated enzyme is resistant to inhibition by tagetitoxin, a specific chloroplast RNA polymerase inhibitor, as well as polyclonal antibodies specific to purified pea chloroplast RNAP. These findings suggest that plastids and probably mitochondria

possess an RNAP gene that functions as a DNA primase, although further research on this topic is needed.

Unlike the mitochondrial RNA polymerases, the plastid encoded RNA polymerase is made up of multiple subunits that share homology with the core subunits of *E. coli* RNAP; α, β, β', and β''. These subunits are encoded from the genes rpoA, rpoB, rpoC1, and rpoC2 respectively. Sigma factors for the plastid-encoded RNAP are required for promoter recognition but these are nuclear encoded.[103] In agreement with the theory of endosymbiosis, the core enzyme of the nuclear-encoded plastid RNAP is also homologous to multi subunit RNA polymerases of cyanobacteria.[142]

In humans and yeast, RNAPs are required to initiate and prime DNA for replication in mitochondria.[147,148] This makes sense if you recall that Twinkle, a helicase-primase, is present in animals but lacks the primase activity observed in both phage and plants. Therefore, plants may also use their organellar RNA polymerases to prime DNA for synthesis. Unfortunately, the ability and scale on which this actually happens is understudied, most likely due to the assumption that organellar DNA is primed by mimicking the simple replisome found in T7 phage. However, unlike animals in which mutation of Twinkle helicase/primase is embryo-lethal, plants with Twinkle knock-out mutations grow well, and display no phenotypic defects. Therefore, the ability of RNA polymerase to prime DNA for synthesis may be extremely important to plants and could be a fruitful area of research.

Primer removal

In *E. coli*, RNA primers are removed from DNA-RNA hybrids by the 5′-3′ exonuclease activity of DNA polymerase I. Since Pol1A and Pol1B lack 5′-3′ exonuclease activity, primer removal must be carried out by another enzyme. In humans, RNase H in mitochondria removes these primers. *Arabidopsis* possesses at least two RNaseH enzymes that have been shown to localize to mitochondria and chloroplasts.[107] One study references two exonucleases with homology to the 5′-3′ exonuclease domain of *E. coli* DNA polymerase I (5′-3′ EXO1 and 2) are predicted to localize to chloroplasts or mitochondria although research on their organellar activity has not been performed nor are gene designations assigned.[108]

Strand separation

Plants utilize at least two types of single stranded DNA binding proteins in their organelles. The first is similar to bacterial single stranded binding proteins (SSB). *Arabidopsis* encodes at least two of these genes, called SSB1 and SSB2, although less is known about SSB2.[46,109] SSB1 functions at replication forks by coating single stranded DNA to prevent fork collapse. This protein localizes to both mitochondria and chloroplasts and stimulates the activity of *E. coli* recombination protein RecA.

The second class of single stranded binding proteins are called organellar single-stranded DNA binding proteins (OSB). OSB proteins are distinct from the bacterial SSB versions and are unique to plant organelles. At least four OSB genes are transcribed in

Arabidopsis with OSB proteins localizing to both mitochondria and organelles. Although the function of these molecules has not been completely detailed, mutants for OSB1 and OSB4 accumulate aberrant mitochondrial DNA sequences resulting from recombination products.[110] Therefore, OSB proteins are most likely involved in recombination surveillance and preventing transmission of incorrect recombination products to new mitochondria.

In addition to OSB proteins, plants utilize Whirly (WHY) and organellar DNA binding (ODB) proteins. WHY proteins form tetramers that take on the appearance of a whirligig, hence the name Whirly. These proteins are unique to plants and research into their function is beginning to attract interest but at the moment is still somewhat scarce. There are three Whirly proteins, WHY1, WHY2, and WHY3.[111] WHY1 and WHY3 localize to chloroplasts while WHY2 localizes to mitochondria. Like OSB, Whirly proteins appear to be involved in recombination surveillance but also have been shown associating with RNA as well. Some researchers have hypothesized that this activity is indicative of Whirly facilitating RNA maturation. Whirly proteins also have been associated with double stranded DNA repair[112,113] and as transcription factors for defensive gene expression.[149] Plants mutated for WHY1 and WHY3 have variegated green/white/yellow leaves indicative of nonfunctional chloroplasts. Further investigation of these chloroplasts reveals higher levels of illegitimate recombination in these mutants.[113] Interestingly, overexpression of WHY2 leads to dysfunctional

mitochondria with lower respiratory activity. Inactivation of WHY2 does not affect plant development nor mitochondrion morphology.[150] Additional research shows these proteins act as transcriptional regulators within both the nucleus and organelles of plants and are involved in organelle-to-nucleus communication. In summary, the Whirlies seem to possess many critical functions in plant organelles despite their relatively small size. The Whirlies present many fruitful research avenues and may point out key points in evolution where plant organelles diverged from animals.

At least two ODB genes have been identified but the vast majority of research has examined ODB1. ODB1 can bind to both single and double stranded DNA, although it has a much higher affinity for single stranded DNA. ODB1 co-purifies with WHY2 protein and also monitors DNA recombination.[151] ODB1 is identical to RAD52, a gene involved in homologous recombination. ODB1 is also distantly related to the yeast gene MGM101, which is involved in homologous DNA repair. Predictably, ODB1 knockout mutants in *Arabidopsis* have an impaired ability to repair DNA breaks via homologous recombination.

Recombination

As mentioned in the previous section OSB, ODB, and Whirly proteins all participate in monitoring and facilitating correct homologous recombination in mitochondria. These proteins are all primarily DNA binding proteins but plant organelles also contain proteins whose primary responsibility is DNA recombination.

There are two classes of proteins dedicated to recombination in plant organelles. One is a MutS homolog called MSH1, and the others are RecA homologs.

MSH1 is a nuclear encoded gene and mutants display patchy green/white/yellow leaves symptomatic of dysfunctional chloroplasts. This phenotype is passed down in a non-Mendelian fashion and was originally thought to be the result of mutations in the plastid genome.[152] Because of this assumption, this gene was originally called *chm* for *chloroplast mutator*. Later, it was discovered that *chm* mutants cause rearrangements to the mitochondrial genome that lead to the observed phenotypes and defective chloroplasts. Despite extensive searching, no mutation or rearrangement of the plastid genome has been observed in *chm* mutants.[153] Being homologous to MutS from *E. coli* and MSH1 from yeast, the gene was consequently renamed MSH1 and mutants designated as *msh1* plants.[154]

MutS from *E. coli* is part of the mismatch repair (MMR) pathway and corrects point mutations and small insertions and deletions by preventing recombination between partially homologous DNA sequences. Mutations of yeast *msh1* lead to a petite phenotype indicative of poor respiration. Mutation of yeast *msh1* is also accompanied by large-scale point mutations and rearrangements in the mitochondrial genome.[155] Interestingly, plant *MSH1* mutants do not accumulate point mutations over time, suggesting that the plant MSH1 specializes primarily in recombination and is not essential for correcting mismatches.

E. coli utilizes the adaptor protein MutL and endonuclease MutH to assist in the mismatch repair pathway, but no identified homologs to MutL and MutH have been identified in plant organelles. Instead, plant MSH1 possesses three recognizable domains and three unknown domains to facilitate mismatch repair. These include a conserved DNA binding and mismatch recognition domain and an ATPase domain homologous to those in *E. coli* MutS. Plant MSH1 also has a unique GIY-YIG endonuclease. Point mutations to the ATPase and endonuclease domains of plant MSH1 lead to the defective chloroplast phenotype suggesting that plant MSH1 represents a compact system of mismatch recognition and base excision without the need for MutL or MutH homologs.[156]

The results of these studies suggest the main purpose of plant MSH1 is to maintain mitochondrial genome stability by rejecting heteroduplex formation. Interestingly, MSH1 has been shown to localize to both mitochondria and plastids in some but not all plants.[156,157] Despite the ability to dual-localize, no effect on plastid genome integrity has been observed, indicating an additional yet to be discovered function of MSH1 in plastids.

RecA facilitates homologous recombination by correctly pairing homologous sequences, promoting strand invasion, and migrating DNA branches during recombination. Eukaryote versions of this protein are called RAD51 homologues. All homologous recombination begins with strand invasion mediated by RecA family

proteins, making this protein crucial for this type of repair. RecA functions by coating single stranded DNA at lesions to form presynaptic filaments. This complex will then search for homology within intact double stranded DNA. Once homology has been established, the presynaptic complex will destabilize the double stranded DNA promoting strand exchange and D-loop formation.

Arabidopsis encodes three RecA proteins, RecA1, RecA2, and RecA3. RecA1 localizes to plastids, RecA3 to mitochondria, and RecA2 to both organelles.[114] Mutations to RecA3 cause mitochondrial rearrangements distinct from those observed in MSH1 mutants. The rearrangements observed in RecA3 mutants are due to homologous recombination of repeated sequences in the mitochondrial genome. Reintroducing RecA3 into these mutants results in a reversal of this effect in most of the progeny by abolishing aberrant mitochondrial DNA molecules. RecA1 and RecA2 appear to be even more essential to homologous recombination as mutations in these genes cause a seed-lethal phenotype. This may be explained by the lack of the C-terminal domain found in both RecA1 and RecA2 as well as bacterial homologs. In bacteria, deletion of this C-terminal domain enhances the activity of RecA, suggesting that it is involved in autoregulation.

CHAPTER 2: Chloroplast DNA copy number changes during plant development in organelle DNA polymerase mutants

The following chapter is taken from an article published in Frontiers in Plant Science. Content and figures have been formatted for this dissertation but otherwise it is unchanged.

Abstract

Chloroplast genome copy number has been shown to be very high in leaf tissue, with upwards of 10,000 or more copies of the chloroplast DNA (ctDNA) per leaf cell. This is often promoted as a major advantage for engineering the plastid genome, as it provides high gene copy number and thus is expected to result in high expression of foreign proteins from integrated genes. However, it is also known that ctDNA copy number and ctDNA integrity decrease as cells age. Quantitative PCR (qPCR) allows measurement of organelle DNA levels relative to a nuclear gene target. We have used this approach to determine changes in copy number of ctDNA relative to the nuclear genome at different ages of Arabidopsis plant growth and in organellar DNA polymerase mutants. The mutant plant lines have T-DNA insertions in genes encoding the two organelle localized DNA polymerases (Pol1A and Pol1B). Each of these mutant lines exhibits some delay in plant growth and development as compared to wild-type

plants, with the Pol1B plants having a more pronounced delay. Both mutant lines develop to maturity and produce viable seeds. Mutants for both proteins were observed to have a reduction in ctDNA and mtDNA copy number relative to wild type plants, measured by quantitative PCR. Both DNA polymerase mutants had a decrease in ctDNA copy number, while the Pol1B mutant had a greater effect of reduction in mtDNA levels. However, despite similar decreases in genome copy number, photosynthesis assays of 14 dpi plants show an increased rate of photosynthesis in Pol1A -/- mutants. These results indicate that the two DNA polymerases are both important in ctDNA replication, and they are not fully redundant to each other, suggesting each has a specific function in plant organelles.

Introduction

Through the process of endosymbiosis, ancient bacteria were engulfed by precursors of eukaryotic cells, and over time most of the genes required for organelle function from these ancestral bacteria have been moved into the nucleus. This raises the question, if most genes have migrated to the nucleus, why not all of them? How do chloroplasts benefit from maintaining their genomes? Most evidence suggests that the unique physiological environment of chloroplasts is required for proper regulation of chloroplast-specific genes. John Allen proposes, supported by significant evidence from the literature, that redox regulation of gene expression is required within the membrane-bound compartment.[158] A chloroplast sensor kinase may detect disruptions

in the photosynthetic electron transport chain, which responds to changes in redox conditions to activate or repress chloroplast gene expression, allowing response and regulation of photosynthesis to changing environmental conditions.[158] Light has been shown to affect the amount of chloroplast DNA (ctDNA) during plant development[159]. Evidence for regulation of chloroplast DNA (ctDNA) by the redox state of cells has been reported in *Chlamydomonas reinhardtii*,[160] and similarly for yeast mitochondrial DNA (mtDNA).[161]

Despite the importance of these organelles, chloroplast and mitochondrial genomes possess relatively few of the genes required for their functions in photosynthesis and respiration. In Arabidopsis thaliana chloroplasts there are 87 protein-coding genes and 41 rRNA and tRNA genes.[162] These numbers are very similar in chloroplast genomes from other higher plant species.[32] The organelle genomes require fully functional transcriptional and translational machinery for expression of the genes. However, plant organelles do not use nuclear DNA replication proteins. Instead, they utilize their own unique set of nuclear-encoded organellar localized DNA replication proteins to maintain their genomes. Many of these are dual-localized to chloroplasts and mitochondria.[46,71,124,157]

In this chapter we focus on chloroplast genome replication and maintenance. CtDNA in higher plants has been shown to replicate by a double-displacement loop mechanism from two specific replication origins[163,164] but may also replicate by a

recombination-dependent (RDR) mechanism.[165-167] The use of two distinct replication mechanisms has been observed for many bacterial virus genomes,[168] where one mechanism is used during the initial stage of infection and another (RDR or rolling circle (RC) replication) for rapid replication of the phage genome for incorporation into new phage particles. The use of two or more mechanisms has been discussed as a possibility for ctDNA replication in plants.[167] Replication via a double-displacement mechanism from specific origins may be involved in maintaining low levels of the chloroplast genome in mature or quiescent cells, while recombination-dependent replication may drive rapid replication to generate high copy numbers of the genome during early stages of plant development.

Tobacco and Arabidopsis have been found to encode two closely related bacterial-like DNA polymerases, which have been designated Pol1A and Pol1B.[97,119,157] Both are dual-localized to chloroplasts and mitochondria in these species.[157] Pol1B has been shown to play a role in ctDNA repair and mtDNA maintenance, photosynthesis, and respiration.[119,169,170] However, in rice and maize a single chloroplast-localized DNA polymerase has been identified.[36,171] By analysis of mutants the maize enzyme, encoded by the w2 gene, appears to be the only DNA polymerase that functions in chloroplasts and may also function in mitochondria. There is a paralog of this gene in maize, but the protein has not been detected in chloroplasts. Both maize proteins appear to be involved in mtDNA replication.[36]

Although the identification and biochemical analysis of plant organelle-localized DNA polymerases has been progressing, limited research has been reported on the role and degree of redundancy of the two DNA polymerases that are found in Arabidopsis and some other species. We have examined the effects of mutations in the A. thaliana organellar DNA polymerases on ctDNA replication by quantitative PCR (qPCR) analysis of organelle DNA levels. We provide an analysis of the effects of T-DNA insertion mutations in either of the DNA polymerase genes on plant growth and development and chloroplast genome copy numbers.

Material and Methods

Planting and growing conditions

We obtained the following T-DNA insertion lines from the Arabidopsis Biological Resource Center (Figure 2-1; ABRC; www.arabidopsis.org): Salk_022624 for Pol1A (At1g50840); Salk_134274 for Pol1B (At3g20540). Pots with the approximate dimensions 3 × 3 × 4 (width × length × height) inches were firmly packed with potting soil and placed in a tray. The soil was then saturated with nutrient water prepared with water-soluble fertilizer (Peter's Houseplant Food). Arabidopsis seeds were planted directly onto the surface of the soil and placed in a 4°C cold room in the dark for up to 3 days. Plants were then moved to a growth room maintained at 22°C with an average surface-light exposure of 80–100 μmol m−2 s−1. During the first 5 days of germination

trays were covered with transparent plastic covers to maintain humidity and prevent drying, after which the covers were removed.

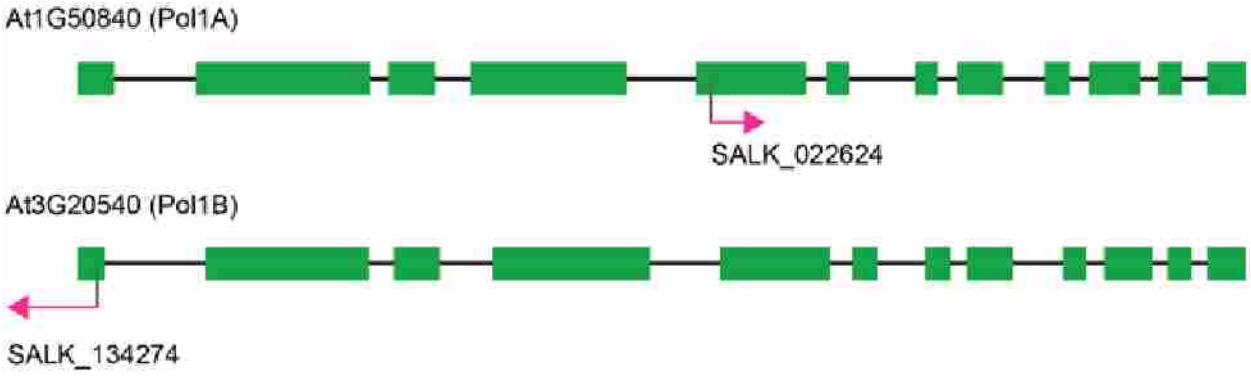

Figure 2-1. Map of the DNA polymerase genes and T-DNA insertions. Note the overall similarity between both genes for Pol1A and Pol1B. Both genes possess 12 exons although SALK_022624 inserts in the fifth exon of Pol1A whereas SALK_134274 inserts in the first exon of Pol1B.

Tissue harvesting and DNA extraction

Leaf tissue was harvested from plants at 7, 10, 14, and 21 dpi (day's post-imbibition). Genomic DNA from these plants was then isolated following a cetyltrimethylammonium bromide (CTAB) method for isolating high quality DNA.[172]

Screening of T-DNA insertion lines

To determine if the T-DNA insertion was present, T-DNA specific primers were used in conjunction with native gene primers. Primers were designed so that native gene primers produced a PCR product about 1 kb in length, and that the T-DNA insertion primer paired with the native gene primer produced a PCR product ~500 b in length. Details of the primers used in zygosity screening are shown in Supplementary Table 2-3.

In order to obtain plants that were heterozygous for Pol1A and Pol1B genes, homozygous Pol1A and Pol1B plants were emasculated and then pollinated from either homozygous Pol1A or Pol1B flowers. This cross generated offspring that were heterozygous for both Pol1A and Pol1B, confirmed via PCR. Seeds from the first generation of heterozygous plants were collected to screen for all possible combinations of Pol1A and Pol1B using PCR as described above.

Genome copy number analysis

Mitochondrial and chloroplast genome copy number was analyzed using an Applied Biosystems StepOne Plus qPCR machine and PowerUp SYBR green reagents. To analyze genome copy number, sequences unique to either ctDNA or mtDNA were identified. For ctDNA analysis, the targets psbK, petD, and ndhH were used. For mtDNA analysis, these targets included nad9, orf25, and cox1. The housekeeping gene AtRpoTp was used as a positive nuclear control and a reference for ΔΔCt calculations. A summary of these targets and their specific genes are listed in Supplementary Table 2-1. Technical and biological replicates were compiled and analyzed using the ΔΔCt method.[170,173]

Analysis of gene expression analysis in Pol1A insertion line

mRNA was isolated from 7 dpi plants using PureLink Plant RNA Reagent (Life Technologies). RNase free DNaseI was added to remove residual DNA. Purity of mRNA was confirmed by running a small amount on a gel and checking for the absence

of large DNA bands. cDNA for RT-PCR was generated from the purified mRNA using SuperScript III reverse transcriptase (Thermo Fisher). Primers for RT-PCR were designed to amplify a portion of the gene near the 3′ end of the mRNA. Primers for RT-PCR are described in Supplementary Table 2-2.

Photosynthesis assays

Seeds from each mutant were germinated in plastic scintillation vials and grown under the same conditions as described above. At 14 dpi the vials were placed in a Licor 6400-22 Lighted Conifer Chamber Package connected to a Licor Li-6400XT analyzer. This system has the ability to measure photosynthetic rates and can automatically generate CO2 and light response curves. For this study, net photosynthetic rates of Pol1A and Pol1B mutants were calculated by measuring total leaf surface area. Total leaf area was calculated by scanning each plant and using ImageJ to trace and calculate surface area.

Results

Phenotype and expression analysis of organelle DNA polymerase mutants

The T-DNA insertion in Pol1A is in the fifth exon of the gene, while the insertion in Pol1B is in the first exon (Figure 2-1). The homozygous single mutant plants exhibited slight growth delays but both grow to maturity and produce seeds. Mutants in Pol1B mutant plants exhibit a slower growth rate than the Pol1A mutants. This

pattern is consistent over time and reproducible (Figure 2-2). This indicates that neither DNA polymerase is completely essential for development.

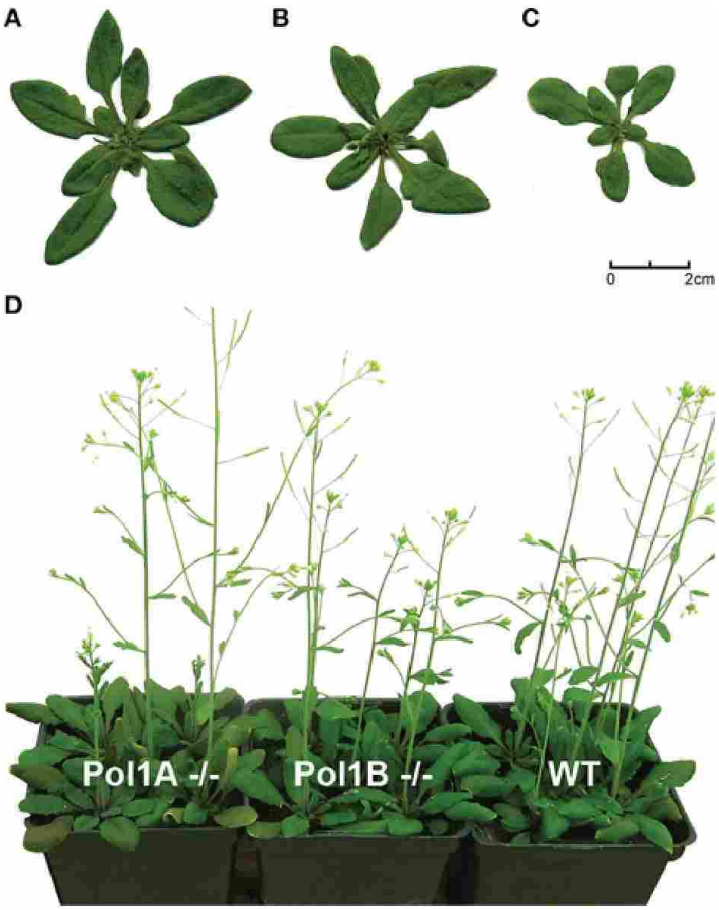

Figure 2-2. Side by side comparison of 23 dpi Arabidopsis. WT (A), Pol1A−/− (B), and Pol1B−/− (C) plants. Note the slightly delayed growth of Pol1B−/− plants and the lack of a distinguishable phenotype between WT and Pol1A−/− plants (D).

We previously showed that both DNA polymerases are expressed in most plant tissues during development, but there is a difference when comparing expression levels of the two genes. Pol1A is most highly expressed (relative to Pol1B) in rosette leaves, while Pol1B is expressed more abundantly in non-photosynthetic tissue. We previously reported that in Pol1B mutant plants, when expression of Pol1B is knocked down a substantial increase (60–70%) in Pol1A expression was observed by qRT-PCR

analysis.[170] We were interested to determine if a similar compensatory effect occurs for the Pol1A mutant. However, relative expression of Pol1B in Pol1A mutant plants was not significantly different from wild-type levels (Figure 2-3). This suggests an important role for DNA Pol1A in chloroplasts and ctDNA maintenance, while Pol1B may play a more significant role in mtDNA replication and maintenance.

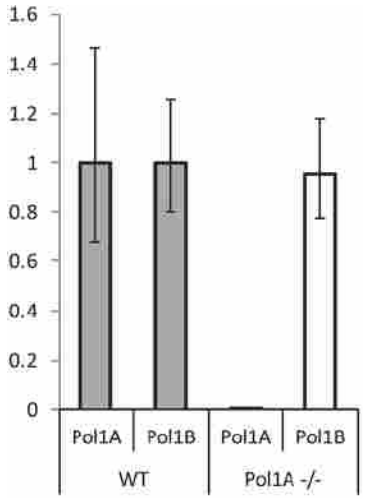

Figure 2-3. RT-PCR of Pol1A and Pol1B expression in Pol1A mutant plants. Although previous work has suggested that mutation in Pol1B causes an increase in Pol1A expression, mutation of Pol1A does not affect expression of Pol1B. This experiment shows relative levels of each polymerase transcript normalized against Actin mRNA. Although mutation in Pol1A knocks down its expression, no significant change in Pol1B expression can be observed.

Our findings are consistent with expression of the Arabidopsis DNA Pol1A gene compiled from microarray analysis in the Arabidopsis eFP browser (http://bar.utoronto.ca/~dev/eplant/). Pol1A expression is highest in rosette leaves of wild-type plants, especially the youngest leaves, but is also high in imbibed seeds and developing flowers, and remains relatively high in cauline and older leaves. Expression of Pol1A is low in embryos and siliques and in pollen (Figure 2-4), and is stimulated by

drought and greatly repressed by osmotic stress.[174,175] Coexpression data (ATTED-II) indicates that the Pol1A gene is co-expressed along with chloroplast-localized RecA, OSB2 (a single-stranded DNA binding protein[46]) and some helicase genes. These proteins may all be involved in ctDNA replication, which would be compatible with the involvement of DNA recombination in chloroplast genome replication (RDR) and/or repair. There is very little information available for DNA Pol1B in these databases.

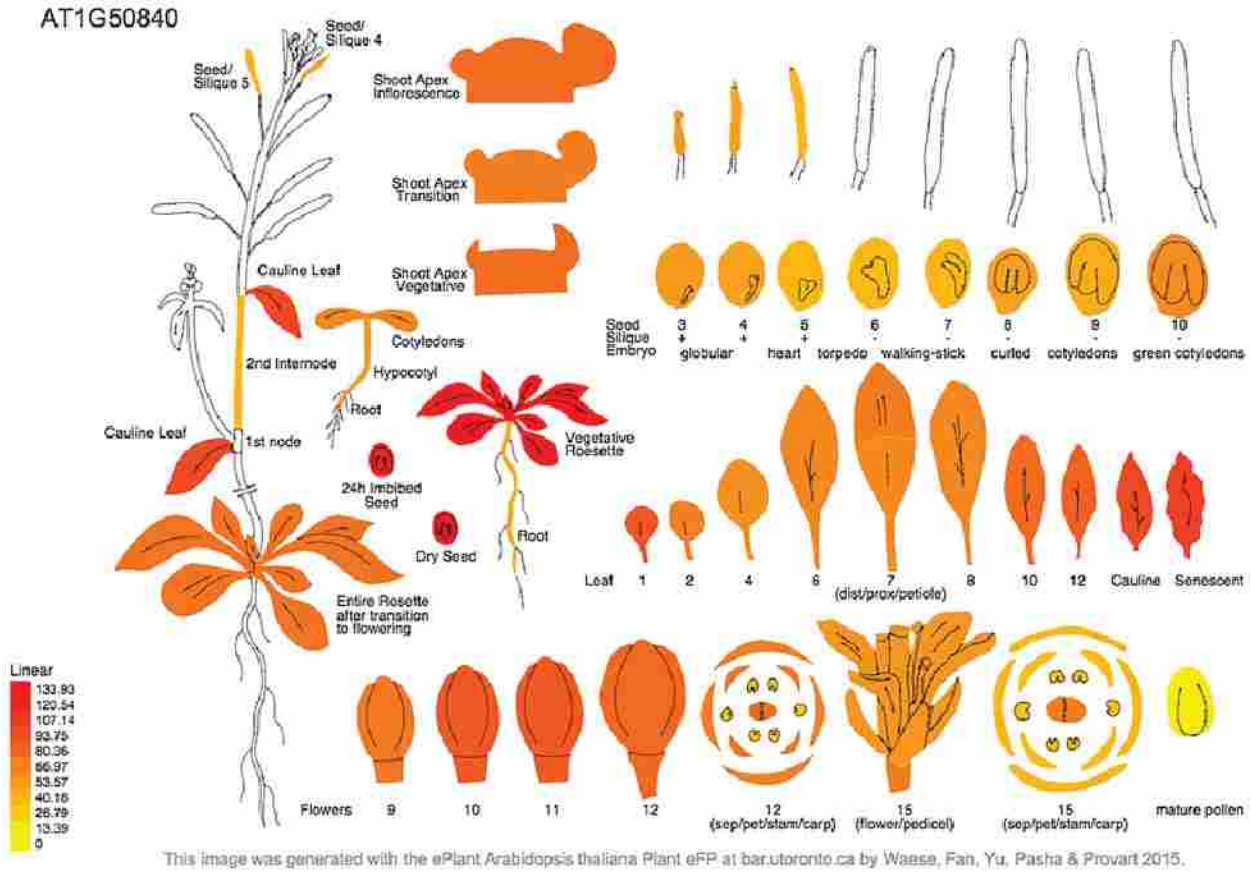

Figure 2-4. Arabidopsis eFP browser report showing predicted Pol1A gene expression in different plant tissues. Expression of Pol1A is highest in rosette leaves, particularly at a young age, however, expression remains relatively high even in senescing leaves. Expression of Pol1A is lowest in seed embryos and pollen.

Field-inversion gel electrophoresis (FIGE) and restriction pattern analysis of ctDNA from the mutants showed no discernable differences in the mutants compared to wild-type plants (data not shown). We used a PCR assay to detect any differences in rearrangement frequency in the mitochondrial genome, as has been observed for mutants affected in mtDNA recombination.[116] However, the Pol1A and Pol1B mutants showed no differences in rearrangement frequency, indicating that there is no major disruption or change in the mechanism for DNA replication/recombination in the individual gene mutants for ctDNA or mtDNA (not shown).

CtDNA and mtDNA copy number determination

qPCR analysis of ctDNA and mtDNA levels in each of the DNA polymerase mutant lines compared to wild-type showed that relative ctDNA levels and mtDNA levels, compared to the nuclear genome, are reduced in both Pol1A and Pol1B mutants, similar to what has been reported before for single time points.[119,170] To determine DNA levels at additional stages of growth, we analyzed samples at different time points. We examined DNA levels at 7, 10, 14, and 21 days of growth. At all time points there is a decrease in organelle DNA copy number in both mutants compared to wild-type plants of the same age for all 3 separate targets for each organelle genome at each age (Figure 2-5). Both Pol1A and Pol1B mutants showed a ~30% reduction in ctDNA at 7 days, a ~40% reduction at 10 and 14 days, and a 50% reduction at 21 days. At 21 days, there is a slightly greater reduction in the Pol1B mutant (~60% decrease) compared to the Pol1A

mutant (~50% decrease). These results indicate that both DNA polymerases affect ctDNA copy number, in contrast with the finding in maize that a single DNA polymerase is responsible for ctDNA replication.[36]

Similar but slightly different results were observed with the two mutant lines when mtDNA targets were analyzed. At 7 days the Pol1A mutant showed only a slight drop in mtDNA copy number, while Pol1B showed nearly a 40% drop (Figure 2-5), similar to what we previously reported.[170] At 10 and 14 days the Pol1A mutant had a 20–40% drop in mtDNA copy number, while in Pol1B the decrease was about 50%. At 21 days, the Pol1A mutant had a 40% decrease in mtDNA, while the PoI1B mutant showed a decrease of more than 60%. These results suggest that while both DNA polymerases contribute to mtDNA copy numbers, Pol1B appears to play a greater role in maintenance of the mitochondrial genome. While qPCR analysis does not directly address quality of the DNA, it does show trends over time for the mutants compared to wild-type plants, indicating changes in organelle DNA levels during development in the mutants compared to wild-type plants.

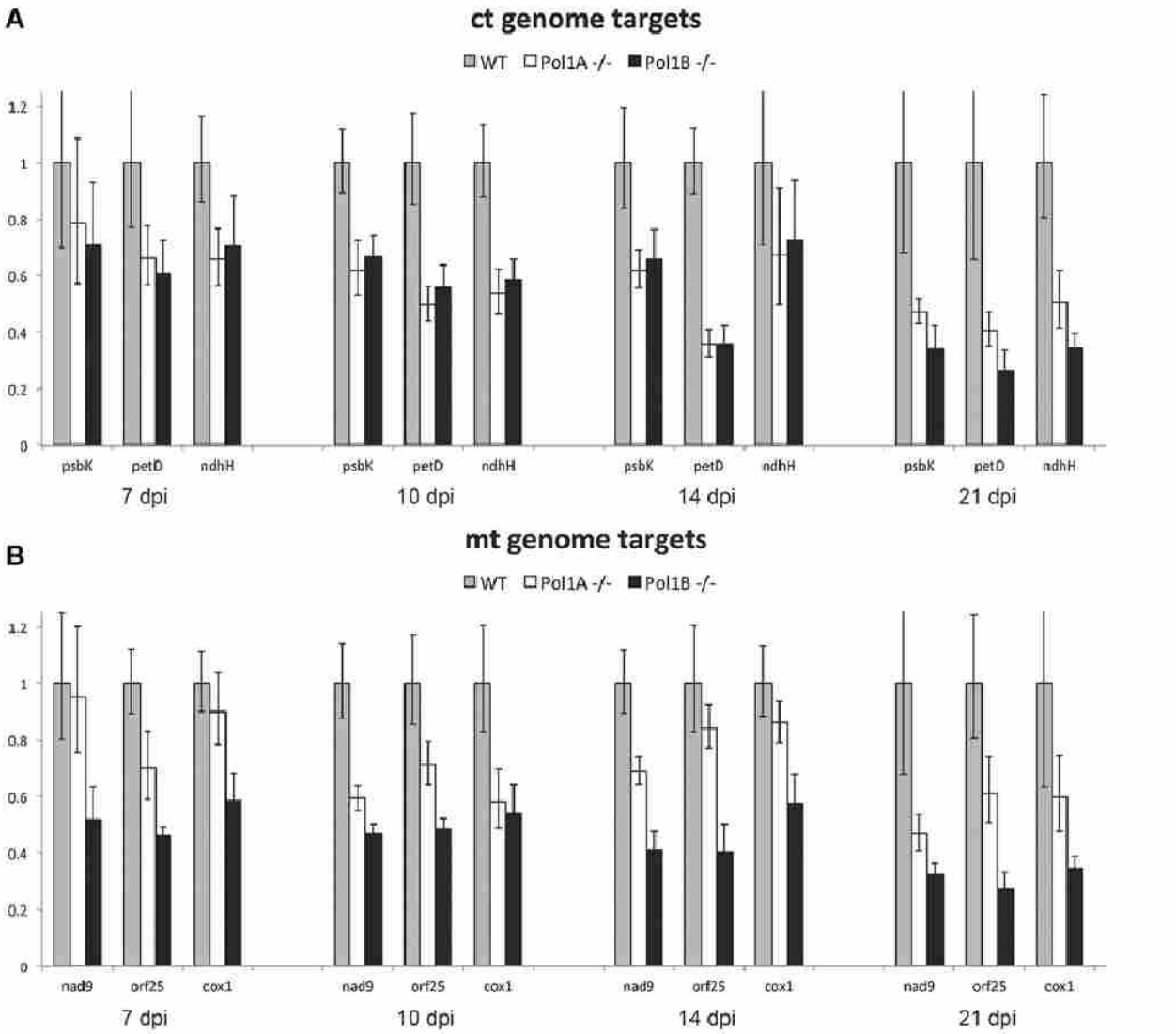

Figure 2-5. Change in relative chloroplast and mitochondrial genome copy number. Note that mutations in Pol1A and Pol1B affect chloroplast genome copy number equally (A) however mutation of Pol1B causes a more severe drop in mitochondrial genome copy number (B). In both mutants, genome copy number gradually decreases but remains lower than wild type as the plants age.

Analysis of photosynthesis in DNA polymerase mutants

The decreases in organelle DNA copy number in the mutants raises a question as to whether these changes affect photosynthesis. In previous work with Pol1B mutants increases in photosynthesis and related parameters were observed.[170] Current

measurements showed an increase in net photosynthesis was observed in 14 dpi Pol1A−/− plants (Figure 2-6). However, we acknowledge that despite careful controls during experimentation, the observed data for Pol1A−/− plants may not be completely accurate. Despite this difficulty in making highly precise measurements, the data suggests that there is an increase in photosynthesis in Pol1A−/− plants, although it cannot be accurately quantified at this time.

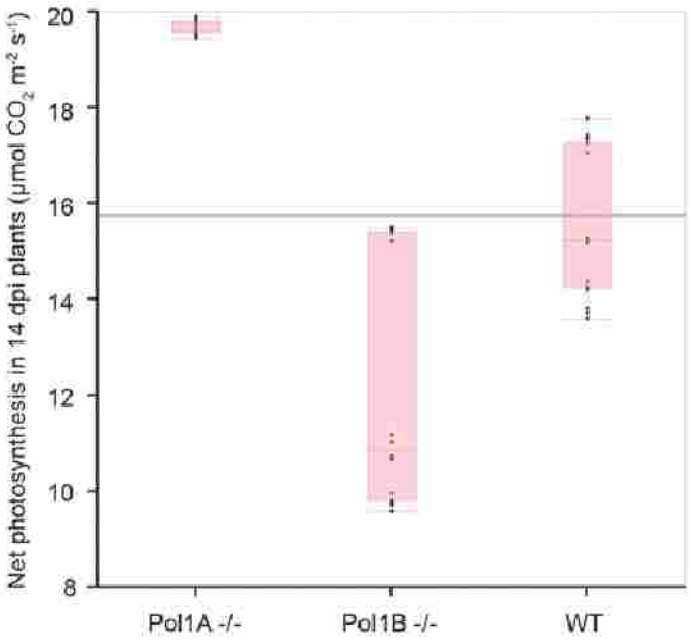

Figure 2-6. Net photosynthetic rates in mutant plants. Observed photosynthesis rates appear to increase in Pol1A−/− mutants.

Analysis of Pol1A × Pol1B partial double mutants

The results of qPCR analysis and previous genotyping experiments led us to believe that certain genotypes would be more beneficial to plant survival than others. To test this theory, we planted seeds on soil in the same manner described above and genotyped all plants that were able to successfully germinate and grow. As expected,

none of the surviving plants were homozygous for T-DNA insertions in both DNA polymerase genes as this most likely is lethal to the plant (Figure 2-7). We also noticed that survival for plants possessing only one functioning DNA polymerase gene was poor. Interestingly we observed strong pressure to maintain both copies of Pol1B with at least one functioning copy of Pol1A. The pressure to maintain both copies of Pol1B suggests higher levels of this polymerase are required to maintain healthy plants.

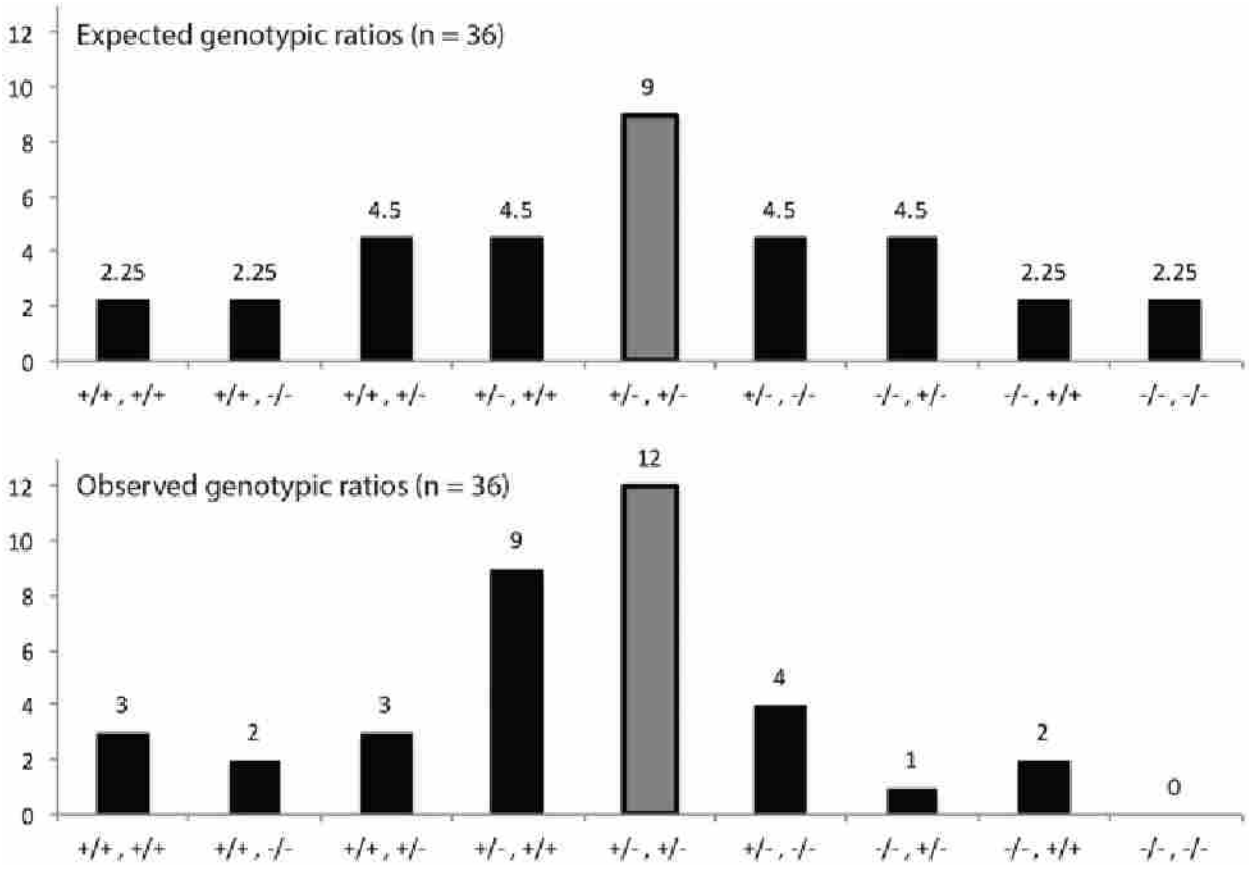

Figure 2-7. Proportion of genotypes from DNA Pol1A × Pol1B crosses. The results come from 36 plants that were able to successfully grow on soil. The horizontal axes represent the possible genotype combinations starting with Pol1A and followed by Pol1B (e.g., +/−, +/− represents Pol1A+/−, Pol1B+/−, respectively). The middle bar represents the heterozygous combination of genes and is highlighted gray for convenience. Because the results are only from surviving plants, certain genotypes were not observed, such as Pol1A−/−, Pol1B−/− as this combination most likely is lethal to the plant. A particularly interesting genotype was Pol1A+/−, Pol1B+/+ which was present in an uncharacteristically high number of plants.

Discussion

Analysis of mutations in the genes encoding the organellar DNA polymerases can provide helpful information for understanding their role in chloroplast DNA replication and genome maintenance. However, at the current time analysis of organelle DNA polymerase mutants has apparently only been done for Arabidopsis[119,170] and maize.[36] In maize it was shown that a single nuclear-encoded chloroplast-localized DNA polymerase (encoded by the w2 gene) is responsible for nearly all ctDNA replication.[36] In contrast, our results show that both Pol1A and Pol1B are required to maintain normal growth of A. thaliana (Figure 2-2, Supplementary Movie 2-1).[170]

Both of the previous reports on Arabidopsis focused on Pol1B, which indicated effects on mtDNA copy number and mitochondrial structure and on plastid DNA repair. In this chapter, we have focused on Pol1A, and show that it plays a role along with Pol1B in ctDNA replication as measured by copy number analysis. This analysis also indicates that Pol1A contributes to a lesser extent in mtDNA maintenance. Mutants in each DNA polymerase gene have a limited effect on phenotype, with Pol1B plants growing the slowest, while Pol1A plants grow only slightly slower than wild type plants.

Analysis of partial double mutants indicates a strong preference for at least one copy of the Pol1B gene. As expected, no viable homozygous double mutants were observed, indicating that at least one copy of one of the DNA polymerases is required

for growth, although growth is progressively affected by the loss of either the second Pol1A or Pol1B allele. As mentioned previously, there is a strong pressure to maintain at least two functioning copies of either DNA polymerase gene, and an even stronger pressure to maintain both Pol1B genes with at least one functioning Pol1A gene. This suggests that Pol1B is much more essential to plant survival and may also be needed at higher expression levels to support a healthy plant. This is in line with our previous report that Pol1B mutants are haploinsufficient while Pol1A is not, which suggests an additive effect of functional Pol1B gene copy number.[170]

Expression of the DNA polymerase genes appears to be very high in young developing tissues, especially in meristems.[171] Pol1A is expressed most abundantly in developing and rosette leaves (Figure 2-4), which agrees with the data available from online expression databases. In contrast, Pol1B is expressed highly (relative to Pol1A) in non-photosynthetic tissues. However, both are expressed in all tissues. The higher expression of Pol1A in leaves suggests that it may play an important role in ctDNA replication. However, the small effect of a homozygous insertion mutant for this gene on plant growth indicates that the Pol1B gene can at least partially complement the Pol1A mutation.

A significant increase in Pol1A expression was observed in homozygous mutant Pol1B plants.[170] In contrast, in homozygous Pol1A mutants there is no significant change in Pol1B gene expression (Figure 2-3). Pol1A homozygous mutants show an increase in

net photosynthesis (Figure 2-6). Photosynthesis was also affected in Pol1B mutants. There may be an inverse relationship between mtDNA levels and net photosynthesis. It may be a decrease in mtDNA, which would affect mitochondrial function, causes a compensatory increase in chloroplast function, including photosynthesis. Thus, while mutants in both genes share some similarities (reduction in growth rate and organelle genome copy numbers and effect on photosynthesis), there are differences in the levels of these effects that strongly suggest different functions for the two DNA polymerases.

Although, both DNA polymerases have been shown to be dual targeted to chloroplasts and mitochondria, we hypothesize that chloroplasts rely more on Pol1A whereas mitochondria rely more on Pol1B for DNA replication. We hypothesize that a mutation in Pol1B causes increased expression of Pol1A to make up for the loss of function of Pol1B proteins. In the reverse scenario, mutation of Pol1A has a less severe effect, and Pol1B may compensate for loss of function of Pol1A without the need for higher Pol1B expression. Further supporting this hypothesis are localization predictions based on protein sequence analysis. When the protein sequences for Pol1A and Pol1B are analyzed by localization prediction programs Target P[176], PCLR,[177] and Predotar (https://urgi.versailles.inra.fr/predotar/predotar.html), Pol1A is consistently predicted to localize to chloroplasts more strongly than mitochondria while Pol1B is most strongly predicted to localize to mitochondria. A summary of these results can be found in Table 2-1. A more detailed analysis using ChloroP predicts that the first 91 residues

of Pol1A whereas only the first 36 of Pol1B serve as a signal peptide for Pol1B, which may help explain the differences in preferred localization.[178] However, Pol1B maintains high homology with Pol1A beyond its predicted signal for ~60 residues (Figure 2-8). Thus, while the genes and protein products are highly homologous, they have some significant differences at the N-terminal and other internal regions, contributing to the observation that the two DNA polymerases are not fully redundant to each other.

Table 2-1. Prediction of Pol1A and Pol1B organelle localization

Prediction program	Pol1A*		Pol1B*	
	Ct	Mt	Ct	Mt
TargetP	0.928	0.314	0.588	0.741
PCLR	0.995	-	0.915	-
Predotar	0.950	0.100	0.600	0.450

*Each prediction program returns the likelihood of each resulting protein localizing to either chloroplasts (Ct) or mitochondria (Mt). This prediction is made based on the amino acid sequence of each polymerase.

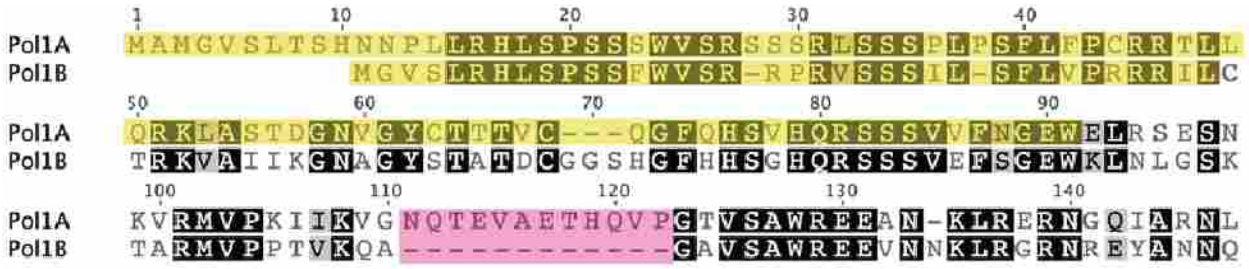

Figure 2-8. Predicted signal peptides of Pol1A and Pol1B and sequence homology in the early region of each polymerase protein. Predictions of each protein's signal peptide was made using ChloroP. Residues highlighted in yellow represent the predicted signal peptide to be cleaved after localization. Note that despite a much shorter predicted signal peptide, Pol1B continues to maintain high homology with Pol1A for ~60 more residues. The first region of dissimilarity between the two polymerases is highlighted in pink.

In contrast to the computer predictions, both Pol1A and Pol1B have been shown to be dual-targeted to chloroplasts and mitochondria.[157] However, the two DNA polymerases may not be equally localized to both organelles at all stages of plant development. It was reported that plastid localization of Pol1A was only obtained when the entire 5′UTR was included in the GFP fusion construct. When the UTR was deleted, initiation of protein synthesis occurred only at the annotated start codon and localization became dual-targeted. The 5′UTR lacks an in-frame upstream start codon, suggesting that an alternate non-AUG start codon was used.[157] Localization may vary depending on growth conditions, which could dictate which form of the protein is translated and thus which organelle it is targeted to. This may also play a role in the localization of the proteins when one of the DNA polymerase genes is knocked out in the T-DNA insertion lines. The absence of one DNA polymerase may trigger signal(s)

for expression of a form of the other DNA polymerase that can compensate for the mutated enzyme. This could explain some of the slight differences in growth rate and other characteristics between the two mutants. The proposed presence of an alternate mechanism for ctDNA replication could also explain why disruption of one or both of the mapped origins (ori) is not lethal, while some of the linear fragments generated still map near the mapped ori regions.[179,180] The confirmation and characterization of different replication mechanisms and differential localization of the organellar DNA polymerases during plant development or in response to mutation or stresses deserves further study.

It is interesting that of the four species for which organellar DNA polymerase genes have been characterized, Arabidopsis and tobacco, which are dicotyledonous plants, have two organelle localized DNA polymerases that both appear to be essential for normal growth and replication of chloroplast and mitochondrial genomes. In contrast, maize and rice, which are monocots, appear to have a single DNA polymerase that is responsible for substantially all ctDNA replication. Analysis of organelle DNA polymerases in additional species will be required to determine whether this is a consistent pattern, which would suggest significant differences in the replication machinery for plants from these two lineages.

Chloroplast genome copy numbers per cell are highest in young photosynthetically active leaves. Chloroplast genome copy number varies widely

between tissues, ranging from 3 to 275 copies per plastid in leaf cells of different developmental stages.[34,181] For other species there are 10–400 copies of the chloroplast genome per plastid, translating to 1000–50,000 genome copies per plant cell.[182,183] This number has been given as a compelling basis for chloroplast genetic engineering. Such high copy numbers could theoretically lead to high expression of introduced genes. Indeed, high yields of gene products in engineered chloroplasts have been reported.[184,185]

Conclusion

In summary, there are two closely related organelle-localized DNA polymerases in *A. thaliana*. While mutants in either gene have only a slight effect on plant growth and net photosynthesis, the two enzymes do not appear to be fully redundant. Mutation of Pol1B causes a more drastic effect on growth compared to the effect of mutation in Pol1A. This is supported by genome copy number analysis. Mutation of either DNA polymerase causes a similar decrease in ctDNA copy number, while mutation of Pol1B causes a more substantial reduction in mtDNA genome copy number than Pol1A mutation. While knockdown of Pol1B resulted in increased expression of Pol1A, suggesting compensation for the loss of Pol1B,[170] knockdown of Pol1A did not lead to any significant change in Pol1B expression (this work). However, Pol1A mutants exhibit a small increase in net photosynthesis, suggesting some adjustment in plants to the reduction in organelle DNA levels. Analysis of double mutants suggests that while

homozygous mutants of either DNA polymerase are still viable, there is a strong pressure to maintain two functioning copies of Pol1B or at the least two functioning copies of either DNA polymerase. These findings indicate that both are important for plant organelle genome replication and plant development, and suggest distinct roles for Pol1A and Pol1B in Arabidopsis. A better understanding of the dynamics and controls of ctDNA copy numbers are important to improve chloroplast genetic engineering to overexpress introduced genes, which is relevant to this special topic issue.

CHAPTER 3: *Arabidopsis* thaliana organelles mimic T7 DNA replisome with specific interactions between Twinkle protein and DNA polymerases Pol1A and Pol1B

The following chapter is taken from an article submitted for publication. At the time of submission the article had not officially been accepted. Therefore there may be slight differences between the completed article and the version that appears here.

Abstract

Background

Plant chloroplasts and mitochondria utilize nuclear encoded proteins to replicate their DNA. These proteins are purpose built for replication in the organelle environment and are distinct from those involved in replication of the nuclear genome. These organelle localized proteins also have ancestral roots in bacterial and bacteriophage genes, supporting the endosymbiotic theory of their origin. We examined the interactions between three of these proteins from Arabidopsis thaliana: a DNA helicase-primase similar to bacteriophage T7 gp4 protein and animal mitochondrial Twinkle, and two DNA polymerases, Pol1A and Pol1B. We used a three-pronged approach to analyze the interactions, including Yeast-two-hybrid analysis, Direct Coupling Analysis (DCA), and thermophoresis.

Results

Yeast-two-Hybrid analysis reveals residues 120-295 of Twinkle as the minimal region that can still interact with Pol1A or Pol1B. This region is a part of the primase domain of the protein and slightly overlaps the zinc-finger and RNA polymerase subdomains located within. Additionally, we observed that Arabidopsis Twinkle interacts much more strongly with Pol1A versus Pol1B. Thermophoresis also confirms that the primase domain of Twinkle has higher binding affinity than any other region of the protein. Direct-Coupling-Analysis identified specific residues in Twinkle and the DNA polymerases critical to positive interaction between the two proteins.

Conclusions

The interaction of Twinkle with Pol1A or Pol1B mimics the minimal DNA replisomes of T7 phage and those present in mammalian mitochondria. However, while T7 and mammals absolutely require their homolog of Twinkle DNA helicase-primase, Arabidopsis Twinkle mutants are seemingly unaffected by this loss. This implies that while Arabidopsis mitochondria mimic minimal replisomes from T7 and mammalian mitochondria, there is an extra level of redundancy specific to loss of Twinkle function.

Introduction

Mitochondria and chloroplasts possess their own unique genomes. These genomes originated from small free living organisms that were engulfed by larger ones during the process of endosymbiosis.[186] Over time, the engulfed organisms lost more

and more of their DNA coding capacity to the nucleus of the larger organism until they lost the ability to be free living. However, time has not removed all DNA from these organelles, leaving behind the modern mitochondrial and chloroplast genomes. This DNA is not an artifact, it is fully functional with genes that are replicated, transcribed and translated to produce essential proteins for organelle function.[44,187] The mechanisms in place for maintenance of this DNA are similar to bacteriophage systems and are much simpler than the mechanisms involved in eukaryotic nuclear DNA replication and bacterial chromosomal replication.[71,188]

Many plant organellar DNA replication proteins have been identified and characterized in various species.[46,71] In this report we focus on the interactions between the Arabidopsis organelle DNA replication proteins Twinkle and DNA Polymerases 1A and 1B. We propose that a minimal functioning DNA replisome in plant organelles consists of the DNA primase-helicase protein Twinkle along with DNA polymerase 1A or 1B. We also propose that SSB1 participates in this minimal replisome. This system has been described as the minimal mitochondrial DNA replisome in mammals,[95] and mimics the replication machinery of T7 bacteriophage.[94]

Arabidopsis Twinkle protein gets its name from the human homolog,[133] which is similar to the bacteriophage T7 gp4 DNA primase/helicase protein. In humans this protein lacks primase activity due to amino acid sequence differences in the primase domain.[135] In plants, this protein has both DNA helicase and DNA primase

activities.[99,136] T7 gp4 protein functions as both a DNA helicase and primase and is central to the replication machinery of its genome.[94] Structurally, *Arabidopsis* Twinkle and T7 gp4 are very similar; however, *Arabidopsis* Twinkle has a slight extension between the primase and helicase domains when compared to phage gp4 as well as a longer N-terminal region.

Zinc fingers are typically associated with DNA binding domains, with many examples found in transcription factors. Zinc fingers may also be involved in protein folding and assembly, RNA packaging, and lipid binding.[189] *Arabidopsis* Twinkle protein mimics T7 gp4 protein by having a functional zinc finger domain. The zinc finger of T7 gp4 is required for a functional primase domain.[190] In humans and mammals, mutations in the ancestral zinc finger motifs lead to a non-functional primase domain because of the absence of the cysteine residues necessary to coordinate a Zn^{+2} atom.

The ancestral origin of plant organellar DNA polymerases is convoluted. Initially, *Arabidopsis* proteins Polymerase 1A and 1B were designated Polymerase gamma 2 and 1, respectively, in reference to human mitochondrial DNA POL γ.[118] However, further studies disputed this characterization and found that the plant proteins have more in common with bacterial DNA polymerase I; hence the names Pol1A and Pol1B.[119] Furthermore, other studies frequently refer to these proteins as plant organellar DNA polymerases, or POPs.[120] Thus, three different names exist for the same protein. In this chapter, we will refer to the proteins as DNA polymerase 1A and

1B. These are the only DNA polymerases known to function within the chloroplasts and mitochondria in plants. The phage T7 homolog of Pol1A and Pol1B is the gp5 protein. Although Pol1A and Pol1B share approximately 70% amino acid identity with each other, studies have identified several different characteristics for each one. Parent et al reported that Pol1B serves more of a role in DNA repair than Pol1A,[119] and Pol1A replicates with more fidelity and is able to displace DNA more effectively than Pol1B.[128,191] These previous works show that Pol1A and Pol1B are not merely redundant DNA polymerases but each has specific functions.

There are a few structural differences between *E. coli* DNA polymerase I and *Arabidopsis* Pol1A and Pol1B. The most obvious is that Pol1A and Pol1B lack the 5′-3′ exonuclease domain found in *E. coli* DNA polymerase I. In its place is a long sequence of amino acids with no functional assignment. In *E. coli,* this 5′-3′ exonuclease activity is involved in removal of Okazaki fragments. Lack of this domain in the plant DNA polymerases suggests that primer removal has been replaced by another mechanism or by other proteins. Therefore, it is possible that as the plant DNA polymerases evolved, the 5′-3′ exonuclease domain was not maintained and became non-functional as mutations accumulated. The second difference between the plant and bacterial DNA polymerases is a stretch of amino acids that has been inserted between the 3′-5′ exonuclease domain and the polymerase domain (Figure 3-1). In structural predictions this takes on a large looping structure.

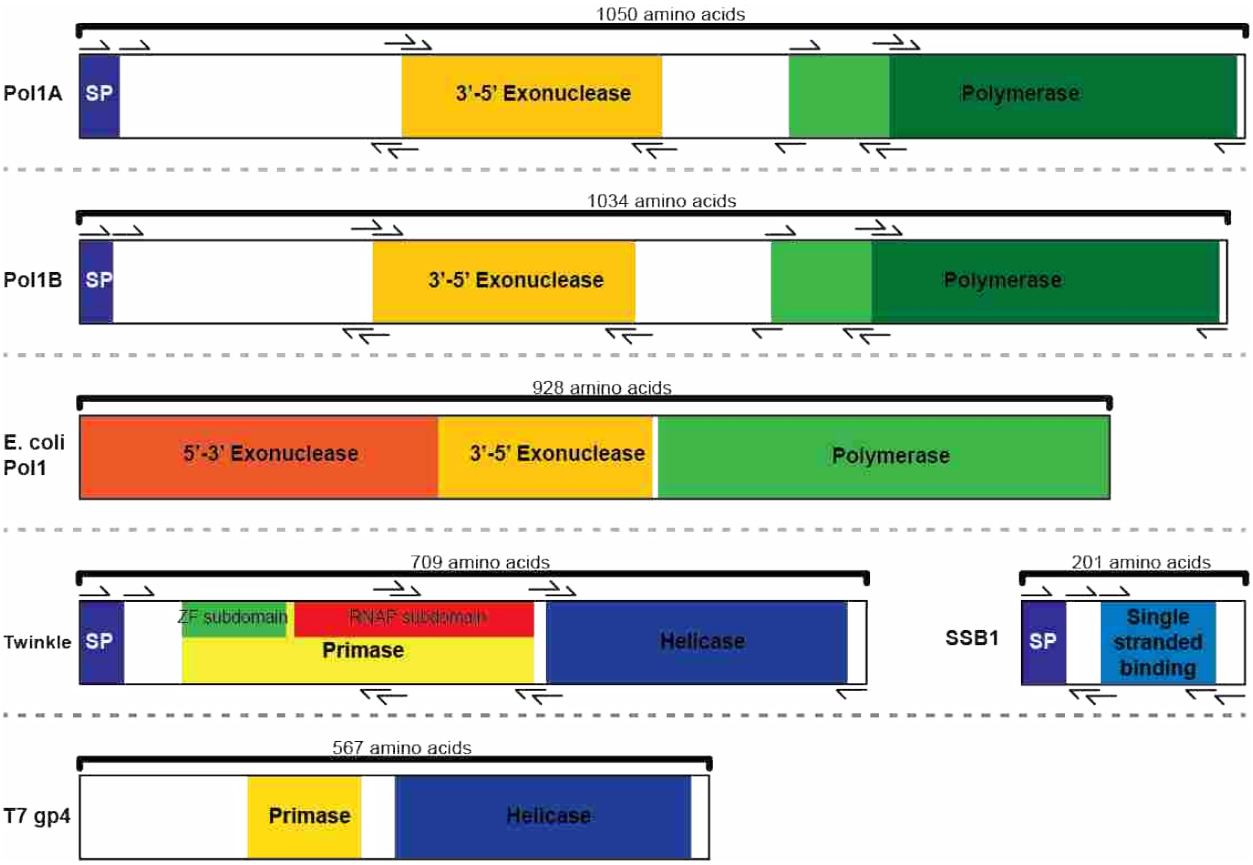

Figure 3-1. Arabidopsis Pol1A, Pol1B, Twinkle, SSB1, E. coli DNA polymerase I, and T7 gp4. "SP" stands for "signal peptide" as predicted by the SignalP server. The light green region in Pol1A and Pol1B represents the border of the plastid-like DNA pol a domain which overlaps with the dark green region representing the more general DNA pol a domain. Genes are shown to scale in terms of cDNA and amino acid length. Black arrows represent primers used to create truncations of each protein.

SSB1 is a single stranded DNA binding protein with homologs in many other organisms. When making comparisons, this chapter will refer to the T7 phage SSB1 homolog, gp2.5. An earlier study has shown that in humans, minimal mitochondrial DNA replisomes can be formed in vitro with just TWINKLE and POL γ.[95] Addition of mtSSB1 protein enhances this replisome. This mimics the simple system found in T7

phage which consists of gp4 DNA primase-helicase, gp5 (DNA polymerase), and gp2.5 (SSB) proteins.[94]

It stands to reason that plants utilize similar replisomes to maintain their organellar DNA. We propose that the minimal DNA replisome in plant organelles consists of similar components found in humans and utilized by T7 phage, namely Twinkle, Pol1A or Pol1B, and SSB1 proteins. In this chapter, we will illustrate the interactions of these proteins, focusing on the specifics of Twinkle and the DNA polymerases. We have identified one specific region of Twinkle with a predicted disordered structure, and we show that this region is key for Twinkle's interactions with Pol1A and Pol1B.

Results

Visual summaries of the *Arabidopsis* DNA replication proteins included in this study with their functional domains are shown in Figure 3-1. SSB1, *E. coli* DNA polymerase 1, and bacteriophage T7 gp4 primase/helicase are included for reference. Functional domains are illustrated based on results from the NCBI conserved domain database, Uniprot[192] databases and published works by Bernstein,[193] Gray,[100] and Richardson.[194] Signal peptide regions were predicted using the SignalP[195] server. With the exception of SSB1, these proteins are unusually large, especially when compared to the amino acid length from orthologous plant and animal proteins.[196]

We hypothesized that the disordered N-terminal regions of Pol1A, Pol1B, and Twinkle interact with each other. This was supported based on multiple sequence alignments of orthologs from other plant species. These alignments revealed high conservation in the functional domains and low conservation in unassigned regions of each protein (Figure 3-2). Structure prediction also differed in depicting the disordered regions of Twinkle and the DNA polymerases. Using the iTASSER structure prediction server,[197-199] we can see clear divisions between the different functional regions of Twinkle (Figure 3-3). These include the primase domain; which includes zinc finer and RNA polymerase subdomains, and the helicase domain.

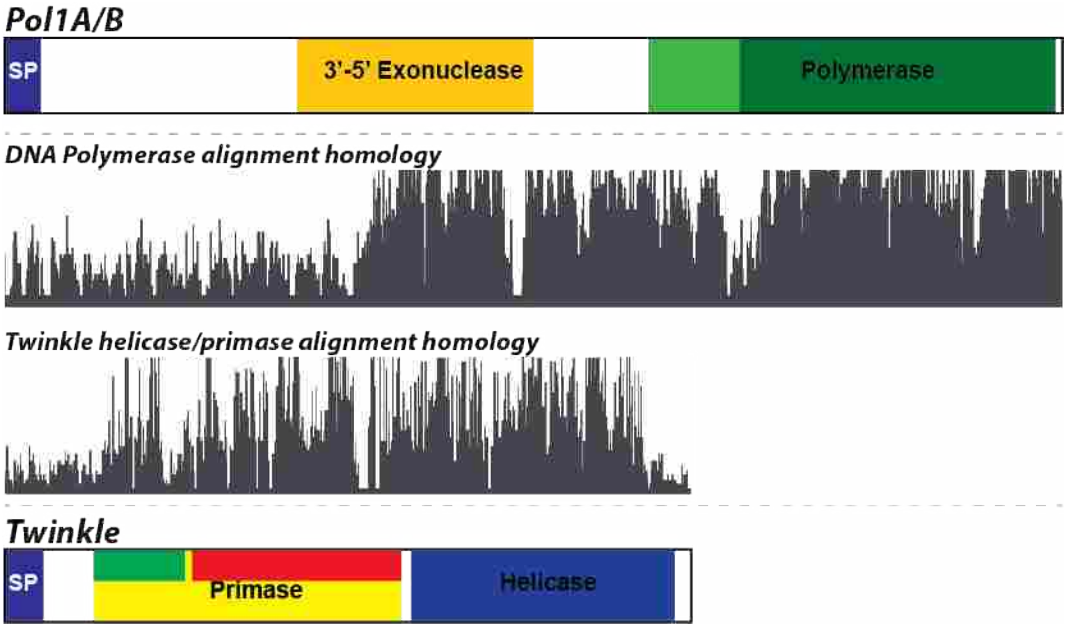

Figure 3-2. Multiple Sequence Alignment histograms of Pol1A/B and Twinkle showing residue conservation. Multiple sequence alignment was performed using other vascular plant sequences and the MAFFT server. High conservation can be observed in the functional domains of each protein, particularly in the polymerases. Low conservation occurs most often in the regions where no functional domain is predicted.

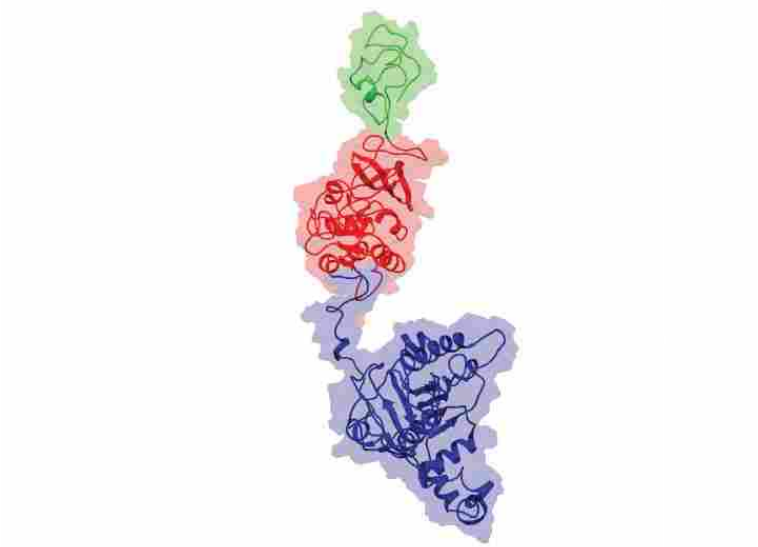

Figure 3-3. Functional regions of Twinkle modeled by iTASSER. This model shows residues 127-708 of Arabidopsis Twinkle. Clear divisions between the zinc finger (green) RNA polymerase (red) and helicase (blue) functional regions can be observed. The primase domain of Twinkle is the combination of the zinc finger and RNA polymerase regions.

To test our hypothesis we adopted a yeast-two-hybrid approach. To ensure we weren't missing any interactions, we created a series of protein truncations. The design of these truncations is shown in Figure 3-1. These truncations were compiled into yeast libraries which were mated against one another to select for interactions. The resulting interactions between the truncated protein constructs are summarized in Figure 3-4.

Our library screen showed consistent interactions between specific regions of Twinkle and three other proteins; Pol1A, Pol1B, and SSB1. In particular, the region of Twinkle downstream from the predicted signal peptide and before the helicase domain appeared repeatedly in many interactions. A similar pattern occurs with the DNA polymerases; the region downstream from the predicted signal peptide and slightly inside the polymerase domain appear in many interactions. The interacting regions are

highlighted in Figure 3-5. These results support our hypothesis that the N-terminal region of Twinkle is responsible for interacting with the DNA polymerases. However, this does not necessarily indicate that only the N-terminal regions of the polymerases interact with Twinkle. The observed interacting regions of the polymerases are very large and eclipse several domains of the protein, not just the N-terminal region. Because the regions we identified in both proteins are quite large, we decided to further investigate the interaction between Twinkle and the polymerases by truncating these regions further.

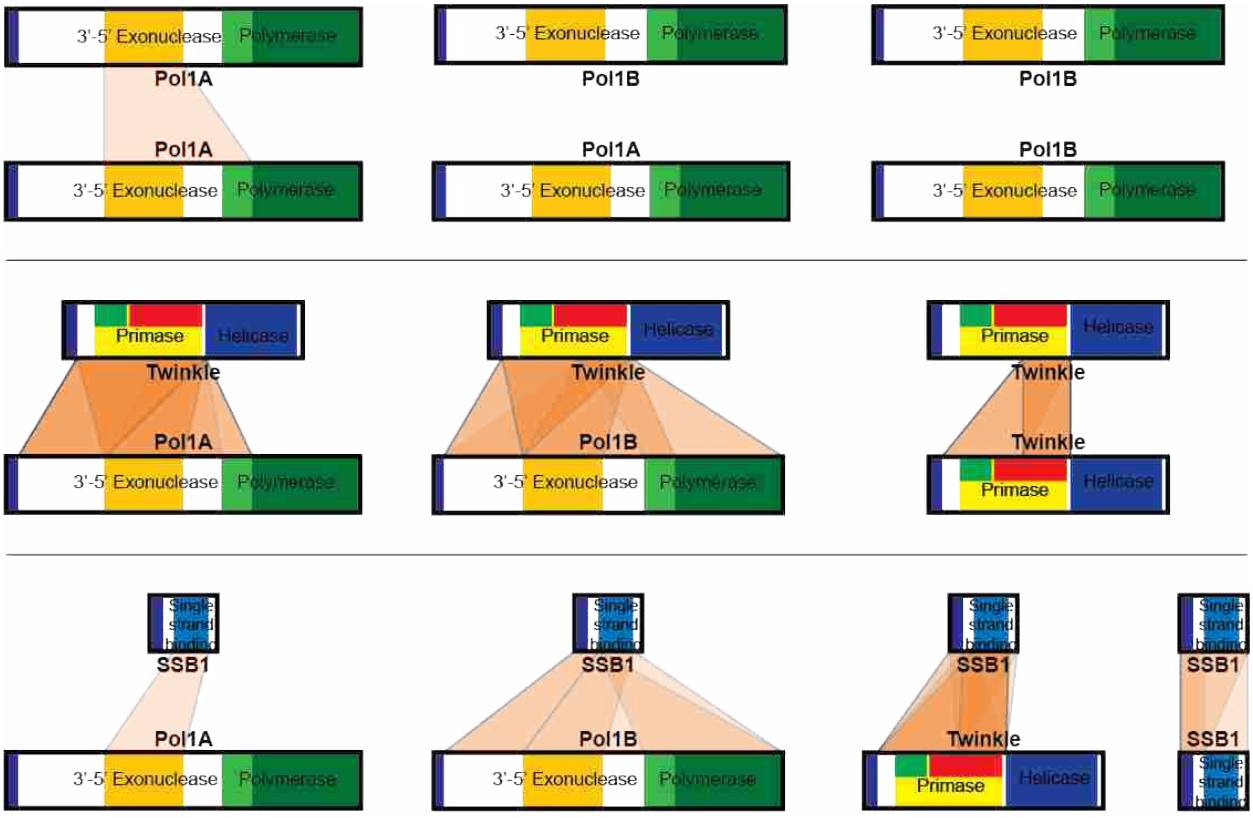

Figure 3-4. Summary of interaction after yeast library mating. Orange bridges between each protein represent unique interactions observed between specific gene truncations. Thus, higher numbers of unique interactions leads to darker bridges between proteins.

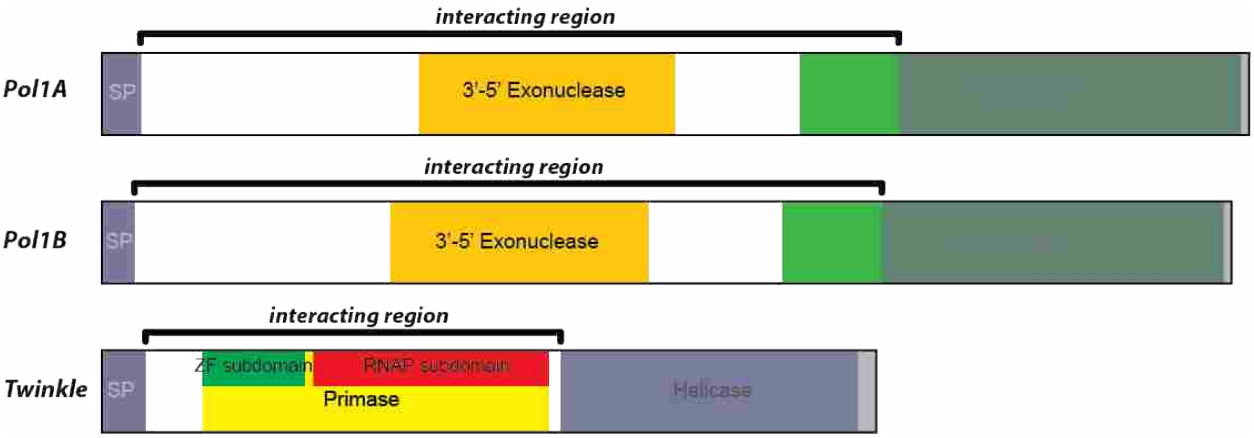

Figure 3-5. Regions of Pol1A, Pol1B, and Twinkle involved in the highest number of yeast-two-hybrid interactions. The regions highlighted above are still rather large and were selected for further truncation to specify exactly which regions of the protein led to positive interaction.

Truncating Twinkle

To make Twinkle truncations we designed primers that shortened the protein in 10 amino acid increments from either the N-terminal or C-terminal side of the interacting region. We amplified and tested different truncations until the interaction was lost. After finding the maximum truncation from both the N-terminal and C-terminal regions that retained a positive interaction, we created a final truncation that combined these borders. Residues 120-295 of Twinkle make up the smallest truncation that maintains interaction with Pol1A or Pol1B, as illustrated in Figure 3-6. This region does not reside in any predicted functional domain and further strengthens our hypothesis that the N-terminal region of Twinkle coordinates interaction with DNA polymerase.

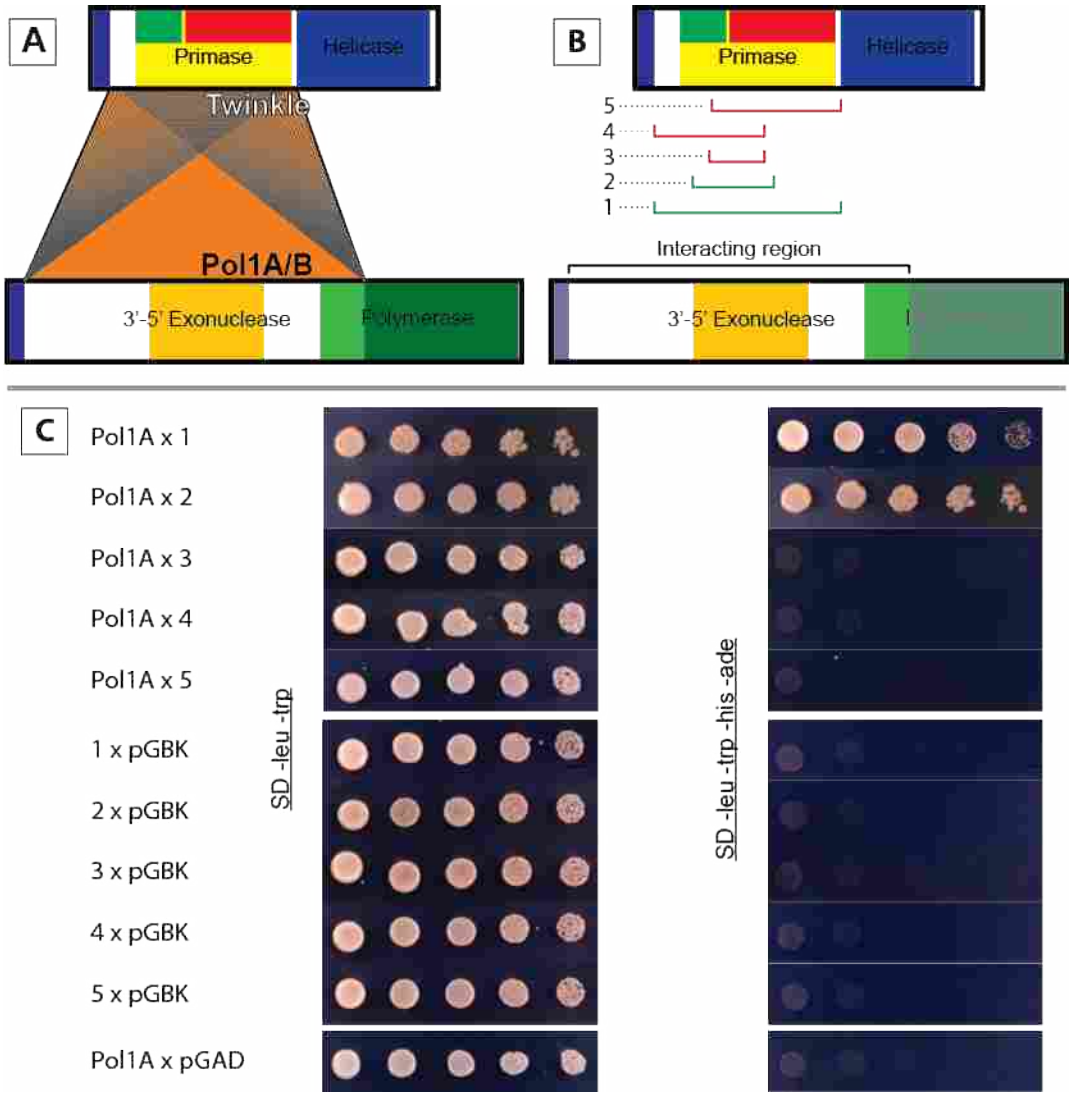

Figure 3-6. Design and results of Twinkle protein truncations. (A) Schematic showing the identified regions of Twinkle and Pol1A/B with a high number of positive interactions. Each grey line represents a potential 10 amino acid truncation from either the N-terminal or C-terminal side of Twinkle. (B) Truncations that succeeded or failed to interact with the interacting region of Pol1A or Pol1B. (C) Spot dilutions for interactions between various truncations of Twinkle and Pol1A. To keep figures concise, we opted to only display the spot dilutions made with Twinkle and Pol1A. Spot dilutions against empty bait (pGBK) or prey (pGAD) vectors are included as negative controls.

Truncating the DNA polymerases

We next truncated the DNA polymerase genes as we did with Twinkle in order to hone in on a specific interacting region. To do this we used the minimal region identified from the Twinkle truncations as a binding partner. Before designing primers for the 10 amino acid deletions, we produced truncations using the primers originally designed in our library approach (Figure 3-1). The purpose was to identify a significantly smaller region of the DNA polymerases that would still interact with Twinkle in order to reduce the amount of effort required to produce dozens of 10 amino acid truncations. However, we soon discovered we could not reduce the interacting region of the polymerases much further than our initial screen had revealed (Figure 3-7). We also noticed that different regions of Pol1A and Pol1B were able to interact with the small region of Twinkle we had identified. Spot dilutions of each interaction allude to a much stronger association between Pol1A and Twinkle versus Pol1B with Twinkle.

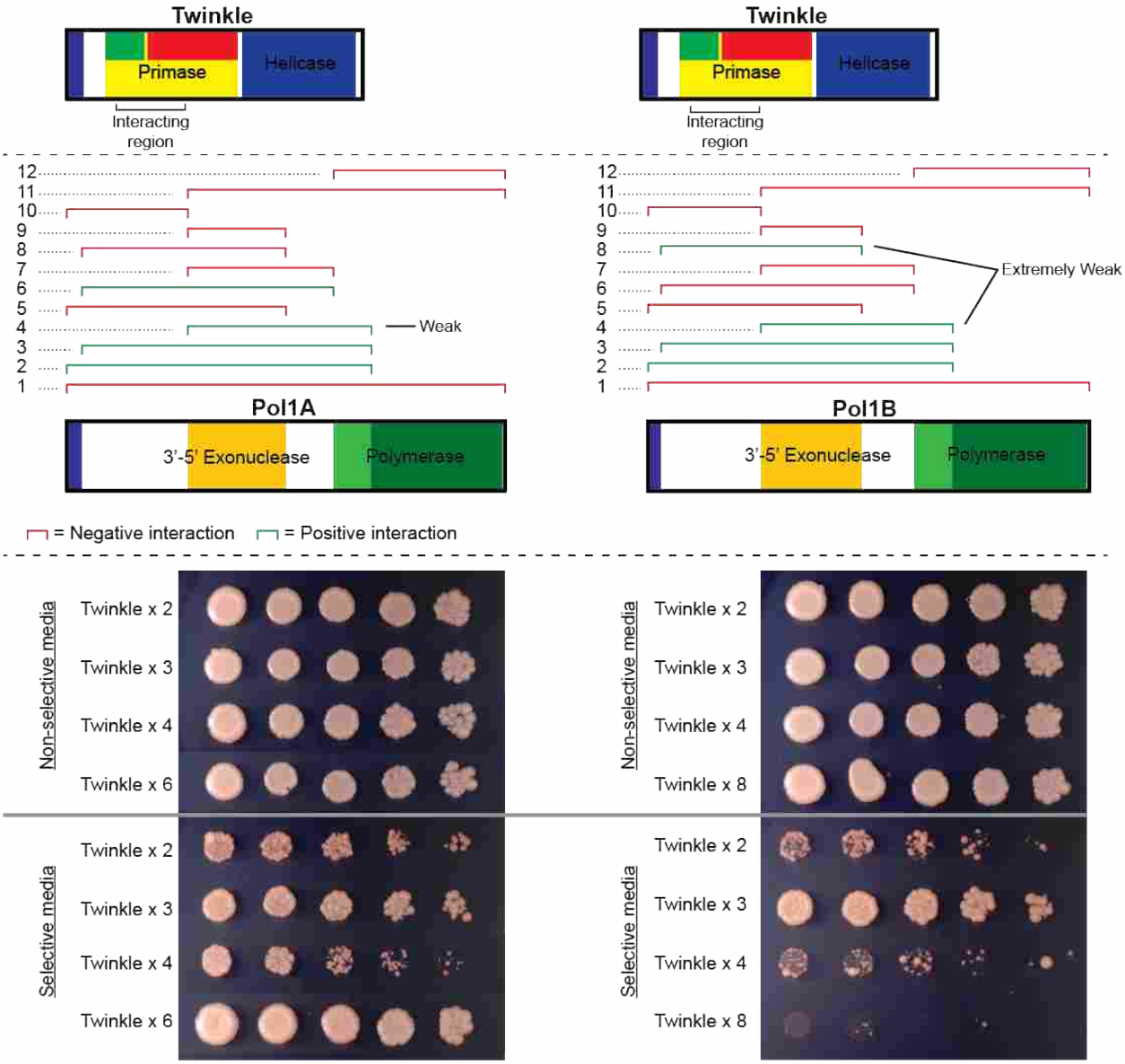

Figure 3-7. Truncations of Pol1A/Pol1B interacting with residues 120-295 of Twinkle. Truncations are numbered 1-12 and span Pol1A and Pol1B as designated above. Each truncation was tested for association with the interacting region of Twinkle and plated in 1:5 spot dilutions on selective and non-selective media as a control. Non-selective media is SD -leu -trp and the selective media is SD -leu -trp -his -ade. Note the stronger interaction between Pol1A and Twinkle versus Pol1B and Twinkle.

Direct Coupling Analysis of residues potentially involved in interactions

Direct coupling analysis (DCA) is a bioinformatic technique that quantifies interactions between two positions in a sequence and can be used to evaluate evolutionary constraints by examining orthologous gene conservation.[200] In our case, we calculated the direct coupling of Twinkle and the DNA polymerase orthologs taken from 90 plant species. Although DCA is typically conducted on hundreds of protein sequences, since our model was intended to assess direct coupling in *Arabidopsis thaliana*, we limited the multiple sequence alignment to include only protein sequences originating in plant genomes. Furthermore, DCA is shown to systematically improve when sequence similarity between the protein sequences is less than 80%.[200] To ensure a balance between the number of sequences and potential biases incurred by sampling only a few species with very similar protein sequences, we included only a single protein sequence from each species, sampling from 90 available genomes from different species. From this data we generated a heat map and selected 10 residues from each protein with the highest likelihood of interaction. Of the 10 residues identified in Twinkle, 8 reside in the region we had previously categorized by yeast-two-hybrid analysis as important in protein-protein interactions (Figure 3-8). For the DNA polymerases, the top 10 residues are spread out much further throughout the protein. The exact position of each candidate residue can be seen in Table 3-1.

Table 3-1. Top 10 residues likely to display interaction between Twinkle protein and Pol1A or Pol1B as predicted by Direct Coupling Analysis (DCA). Amino acid designations and tri-alanine mutants for each of these residues is also displayed.

Twinkle

Residue number	Amino acid	Tri-Alanine Mutant created
124	Glu	Δ 123,124,125
126	Cys	Δ 126,127,128
127	Pro	
159	Leu	Δ 158,159,160
177	Ala	Δ 176,177,178
185	Asp	Δ 184,185,186
189	Lys	Δ 188,189,190
233	Gly	Δ 232,233,234
430	Thr	Δ 429,430,431
431	His	

Pol1A/B

Residue number	Amino acid (Pol1A/Pol1B)	Tri-Alanine Mutant created*
118	Val/Glu	Δ 117,118,119
248	Arg/Asp	Δ 247,248,249
255	Val/Glu	Δ 254,255,256
412	Pro/Lys	Δ 411,412,413
422	Glu/Asp	Δ 421,422,423
427	Leu/Lys	Δ 426,427,428
495	His/Phe	Δ 494,495,496
591	Val/Lys	Δ 590,591,592
634	Asn/Val	Δ 633,634,635
695	Thr/Gly	Δ 694,695,696

*Mutants were created for Pol1A and not Pol1B.

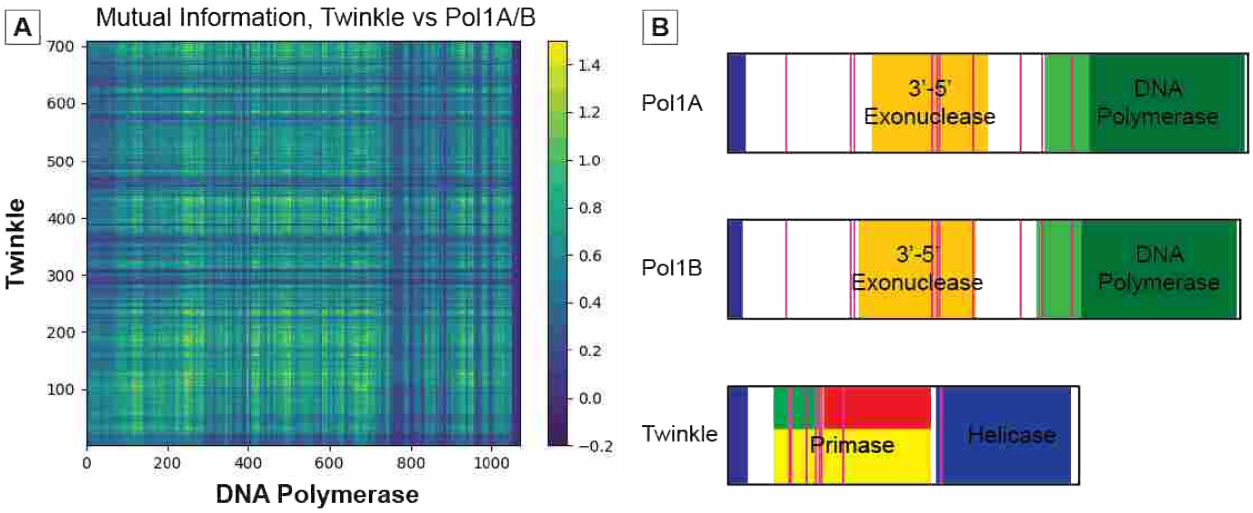

Figure 3-8. Heat map of residues most likely to interact based on Direct Coupling Analysis. (A) Heat map showing the likelihood of residues from Twinkle (y-axis) to interact with residues from Pol1A or Pol1B (x-axis) based on mutual information generated from DCA. The raw data used to make this heat map was analyzed to identify 10 residues in both the polymerases and Twinkle with the highest likelihood for interaction. (B) Location of the 10 residues identified from DCA analysis. These residues were picked for point mutation analysis to see how they would affect the Polymerase-Twinkle interaction.

We mutated each of these 10 residues and their immediate neighbors in Twinkle and Pol1A to determine whether they would disrupt interactions. We chose to mutate Pol1A as our previous work showed more positive associations with Twinkle and therefore a loss of interaction would be more distinct. All mutations were made by substituting the original residues and their neighbors with alanine creating tri-alanine mutants. Because DCA analysis is based on gene alignments from many species, we mutated the candidate residues and their immediate neighbors to correct for imperfect positioning of critical interacting residues in Twinkle or Pol1A. Some of the candidate residues in Twinkle were so close together we were able to mutate both of them with one clone. Similar to the previous analysis, we observed much stronger interactions

between Twinkle and Pol1A vs Pol1B (Figure 3-9). We also detected more interaction disruptions between Pol1B and Twinkle due to point mutations.

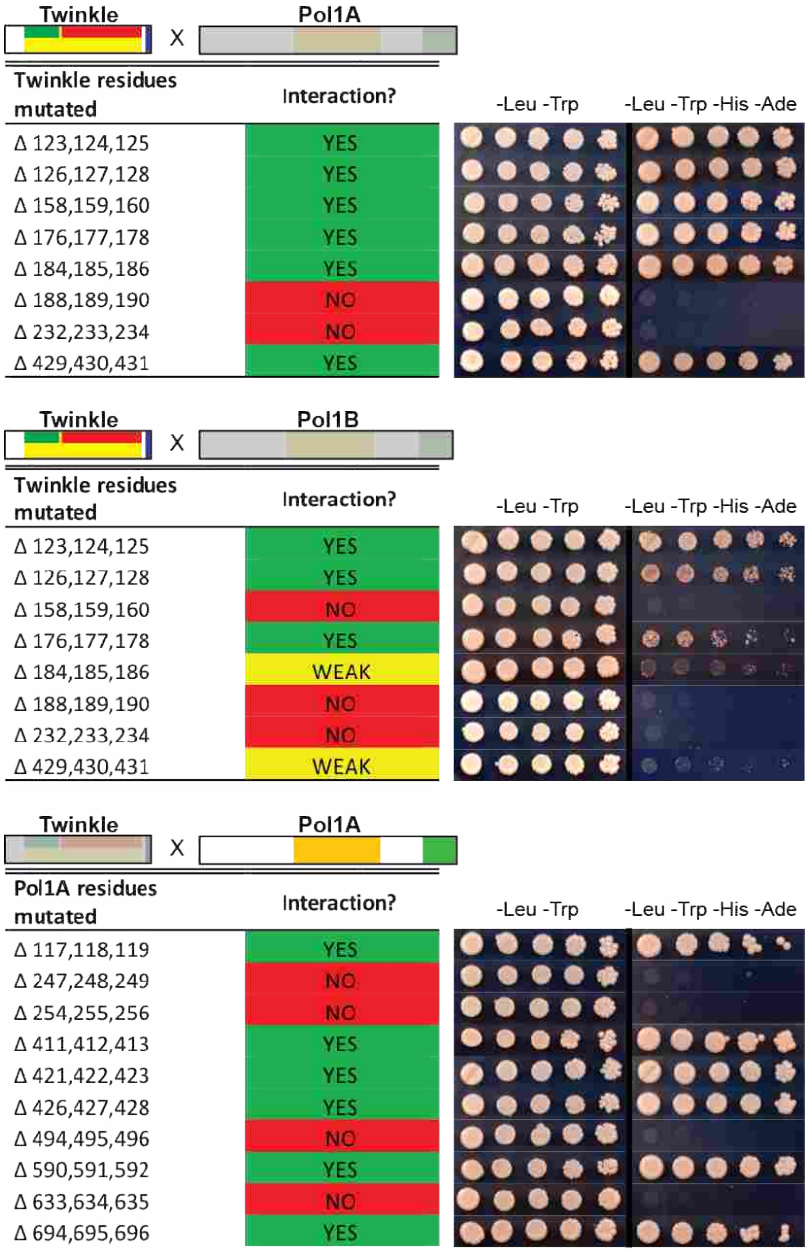

Figure 3-9. Results of tri-alanine substitution mutation analysis. Tri-alanine mutants were created in Twinkle and Pol1A to test for interaction disruption. Mutated residues were selected based off of DCA analysis results. While only two mutations caused a disruption of Twinkle with Pol1A, most mutations weakened or completely disrupted interaction with Pol1B. Mutations in Pol1A were much more distinct, either completely disrupting interaction with Twinkle, or failing to affect this interaction at all.

Thermophoresis analysis of Twinkle/DNA polymerase interactions

Although the yeast-two hybrid experiments showed a stronger interaction between Pol1A and Twinkle, Pol1B was used to determine the binding constant with Twinkle and its different structural modules because this recombinant protein has a better yield during protein purification and because of the 70% amino acid identity present between Pol1A and Pol1B. In this experiment a polyhistidine-tagged Pol1B and the ligands (full-length Twinkle, Zn^{++} Finger domain, RNAP domain, primase module and helicase module) without a histidine tag were used. The his-tagged Pol1B was labeled with a fluorescent dye and the amount was kept constant at 10 nM during the binding experiments. After incubation with increasing concentrations of ligand, samples were loaded into MST standard capillaries. Pol1B showed specific binding to full length Twinkle, the primase and helicase domains, and the zinc finger subdomain. The RNAP region did not show appreciable interaction with Pol1B (data not shown; region indicated in Figure 3-10). The fitted values from the thermophoretic analysis yielded dissociation constants for the ligand partners Pol1B-Twinkle of 1.26 μM, Pol1B-Primase of 0.81 μM, pol1B-Helicase of 1.97 μM, pol1B-zf of 2.17 μM (Figure 3-10). These values are comparable with the binding affinities between T7 DNA polymerase and the bifunctional T7 primase-helicase.[201-203] The dissociation constants for this interaction are in the order of 0.44 μM in the presence of DNA [204]. The DNA primase and DNA helicase activities are encoded as two independent polypeptides in bacteria. In *Bacillus subtilis*

the dnaG primase protein has been documented to interact with the DNA polymerase dnaE helicase with a dissociation constant of 0.75 μM.[205] Finally, zinc finger containing proteins have been described as interacting partners with DNA polymerases, in which cases the zinc finger is especially responsible for physical interaction between both proteins.[206,207] The confirmation of the interaction between the Twinkle and Pol1B provide further support for the role of these enzymes in replicating the *Arabidopsis thaliana* organellar genome.

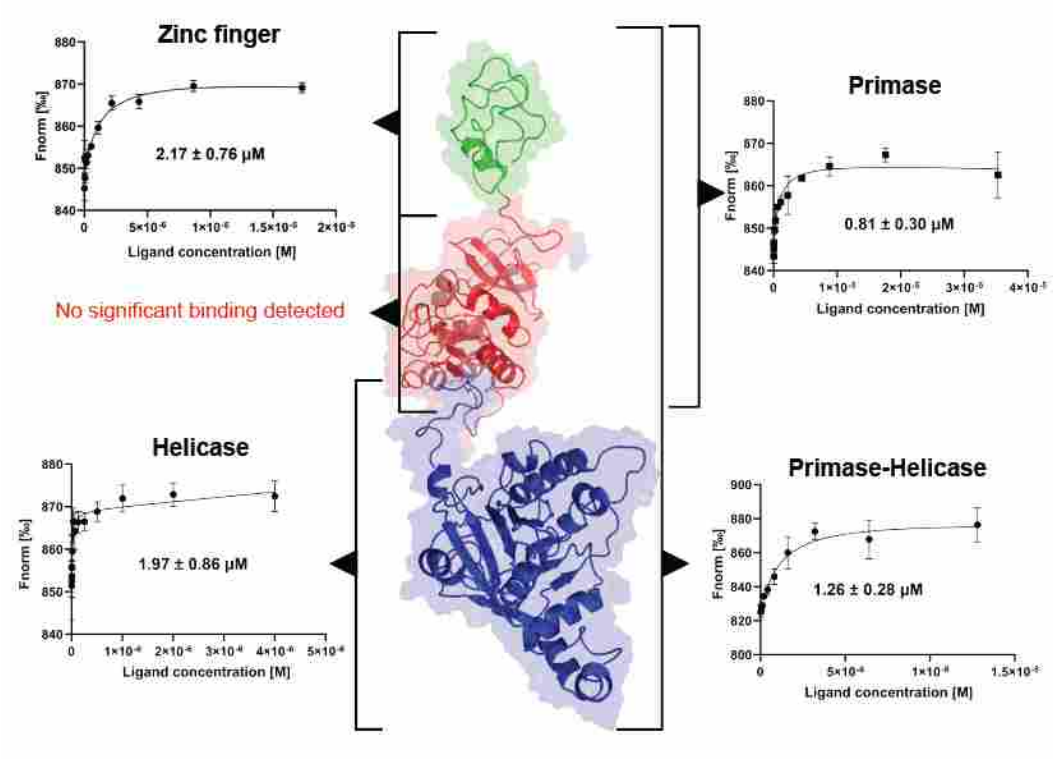

Figure 3-10. Binding of DNA polymerase B with Twinkle domains. The DNA polymerase B was labeled as the target protein at 20 nM concentration and assayed in microscale thermophoresis experiments with Twinkle, the primase and helicase domains, and the zinc finger subdomain. The thermophoretic data was fitted to the Kd equation described in materials and methods. The zinc finger subdomain was able to bind to Pol1B and a Kd was measured at 2.17 μM. The RNAP subdomain (red) did not show binding (data not shown). The primase and helicase domain were able interact with the polymerase with a Kd of 0.81 and 1.97 μM respectively. The

full-length protein interacts with a Kd of 1.26 μM. all proteins were titrated in 16 serial dilutions from different concentrations. Graphs were plotted at x axis with enzyme concentration and at the y axis with normalized fluorescence. Error bars represent the standard error for three measurements.

Discussion

We have identified a region of *Arabidopsis* Twinkle DNA primase/helicase that is crucial for interaction with the two organellar DNA polymerases. Within this region, mutation of several key residues results in complete dissociation from Pol1A or Pol1B as determined by yeast-two-hybrid analysis. We have also shown that mutation of key residues in Pol1A disrupts interaction with Twinkle. Unlike Twinkle, these key residues in the DNA polymerases are spaced much further apart. We suspect this was the main reason we could not produce a smaller truncation of Pol1A or Pol1B that maintained interaction with Twinkle, whereas the interacting region of Twinkle was localized to a very narrow region of the protein. This supports our hypothesis that the N-terminal region of Twinkle coordinates an interaction with Pol1A or Pol1B. However, it does not confirm that only the N-terminal region of the DNA polymerases is crucial for positive interaction with Twinkle. Results from our previous screens also show this same region of Twinkle associates with SSB1, supporting the idea that Twinkle coordinates the assembly of a minimal DNA replisome. As most vascular plants possess the same orthologs of these organellar replication proteins, this pattern is likely to be repeated in higher plant chloroplasts and mitochondria of other species.

Although we have provided evidence showing that Twinkle and the DNA polymerases likely form a minimal DNA replisome, we do not know if this is the sole DNA replisome utilized in plant organelles. Additionally, other proteins may provide accessory functions such as in *E. coli*.[208,209] For example, almost no research has been performed on primer removal. Two 5 '-3' exonucleases have been identified in *Arabidopsis* organelles,[210,211] one localizing to the mitochondria and one localizing to the chloroplasts, but there not have been any studies on their mechanisms of action. Ribonucleases have been reported in plant mitochondria and chloroplasts, but their characterized functions are in mRNA and tRNA processing.[212] We have also shown strong interactions of *Arabidopsis* Twinkle with SSB1, which may form a crucial part of a DNA replisome.

Pol1A and Pol1B have been shown to be processive enough to replicate an entire organelle genome equivalent.[119,124] This is much more processive than *E. coli* DNA polymerase I, which is involved mostly in Okazaki fragment processing and DNA repair and cannot replicate much more than several dozen bases.[125] By itself, T7 DNA polymerase gp5 also lacks processivity, but when bound to *E. coli* thioredoxin becomes highly processive. Interestingly, recombinant versions of *E. coli* DNA polymerase I that are able to bind thioredoxin display a dramatic increase in processivity.[126] Plant organelle DNA polymerases may also bind thioredoxin or other processivity factor to achieve their greater processivity, but this has never been shown. If this occurs, it would

help explain why these enzymes possess much greater processivity than E. coli DNA polymerase I, although they are otherwise quite similar in sequence and function.

To date, no other organellar DNA polymerases have been identified in plant chloroplasts and mitochondria other than Pol1A and Pol1B. In addition to DNA polymerase(s), determining what helps unwind and prime DNA for replication is crucial to our understanding of DNA replication in plant organelles. We have demonstrated an association between the DNA polymerases and Twinkle suggesting that these two proteins are part of the DNA replisome in these organelles. However, Twinkle T-DNA insertion mutants in *Arabidopsis* show no noticeable defect in plant growth (Nielsen et al., unpublished data) and there is no noticeable decrease in organelle genome copy number compared to wild-type plants (Supplementary Figure 3-1). The plants grow similarly to WT. This is puzzling, as similar interactions between T7 gp4 helicase-primase are essential for processive replication of phage DNA.[203] In addition, conditional Twinkle knockout mice fail to survive and display a rapid depletion of mitochondrial DNA in both heart and skeletal muscle tissue.[213] If Twinkle knockout plants grow well, this strongly suggests that another protein may provide the primase activity required for DNA polymerases to function. However, previous work has shown that Twinkle efficiently primes DNA synthesis by Pol1A and Pol1B.[136] This same study demonstrates that *Arabidopsis* Twinkle cannot prime T7 or *E. coli* DNA polymerase I. We and others have shown that plants can survive with almost no visible

growth defects if they have at least one functional organellar DNA polymerase. Despite this lack of phenotype, mutating Pol1A or Pol1B leads to an approximate 30% decrease in organelle genome copy number.[121] Growth at these decreased organelle genome copy numbers does not appear to be affected by low light conditions or drought (data not shown).

A simple explanation for the reason Twinkle knockout plants grow normally is that there is another DNA replication protein that is either 1) the main helicase/primase active at replication forks or 2) another protein compensates for the loss of Twinkle activity. In either case, there is an apparent redundancy in Arabidopsis for Twinkle function. Other candidates include Twinky (At1g30660), a truncated version of Twinkle that retains only the DNA primase domain, and PrimPol (At5g52800), a unique primase/polymerase protein. However, T-DNA insertion mutations in either of these proteins also lead to no discernable growth defect (Supplementary Figure 3-1). In any case, Pol1A and Pol1B must be involved, since they are the only DNA polymerases found in the organelles and we and others have been unable to make double homozygous mutants in both genes.

In previous studies, we found that plants with Pol1B mutations had a greater decrease in organelle genome copy number relative to plants with Pol1A mutations, particularly in mitochondria.[121,214] We also observed a slight delay in growth of these plants. This suggests that Pol1B is more important for DNA replication in the organelles

than Pol1A. If the minimal replisome consists of Twinkle and a DNA polymerase, we expected based on our previous results that Twinkle would show a strong association with Pol1B. However, our research shows the opposite is true. This could be explained by different scenarios: a strong interaction with Twinkle may be detrimental to DNA replication, or the roles of Pol1A and Pol1B may be less redundant and more distinct from each other. Supportive of this, one study demonstrated that Pol1B is primarily involved in DNA repair.[119] Other studies show that both DNA polymerases can perform translesion repair but Pol1B is more effective at strand displacement than Pol1A.[122,128] This suggests that Pol1A may be more involved in DNA replication whereas Pol1B is more involved in DNA repair. This contrasts with findings in maize, where only one of the two organellar DNA polymerases was shown to be responsible for replication of the maize plastid genome, as mutation of this single gene essentially abolished chloroplast DNA replication.[36] This maize mutant had about a two-fold reduction in mitochondrial DNA, inferring that the second DNA polymerase may function in mitochondria.

In addition, recombination dependent replication (RDR) could explain the lack of a phenotype in Twinkle mutants. Extensive use of RDR in plant organelles has been demonstrated as a means of maintenance and repair of mitochondrial and plastid DNA.[215] This is especially true in mitochondria where direct and inverted sequence repeats of ≥50 bp are present throughout the genome. Mutations in recombination proteins lead to genome instability and often plant death. One prominent example is

RecA, a homolog of the bacterial replication protein RecA. Arabidopsis possess three RecA homologs that localize to the organelles (RecA1 [At1g79050], RecA2 [At2g19490], RecA3 [At3g10140]). Mutations to these homologs lead to delayed phenotypes, increased recombination, and in the case of RecA2 plant death beyond the seedling stage.[114,166,216] Other recombination proteins that localize to the organelles include single stranded DNA binding proteins Whirly[111,113] and OSB,[110] and a MutS homolog called Msh1.[154] Mutations to any of these proteins lead to an increase in illegitimate recombination and adversely affect plant development.

Conclusion

We have used three independent approaches to confirm a positive interaction between Twinkle and the organellar DNA polymerases Pol1A and Pol1B. We have used a classic molecular biology technique (yeast-two-hybrid), bioinformatics (DCA), and biochemistry (thermophoresis) to define important regions and residues that are key to this interaction. This three pronged approach provides confirmation of the interactions, and can be applied to other protein-protein interaction studies.

Further work examining the complete DNA replisome of plant organelles will identify other proteins involved in mitochondrial DNA replication in *Arabidopsis*, including when Twinkle is mutated. Candidates for study include Twinky, PrimPol, RecA, Whirly, and MutS. The same approach we have used to prove the interaction

between Twinkle and the DNA polymerases may reveal specific regions in these candidate proteins that are crucial for DNA replisome assembly and function.

Materials/Methods

Yeast-two-hybrid analysis

Yeast-two-hybrid analysis was performed using materials and protocols from Clontech. This included using the Matchmaker gold strain of yeast for small scale yeast transformations following the lithium acetate protocol outlined by Clontech. Different truncations of Pol1A and Pol1B were used in yeast-two-hybrid to test for regions of interactions *in vivo.* All constructs were tested for autoactivation by transforming the target protein against an empty bait or prey plasmid to eliminate the possibility of false positive interactions. For mating experiments, Matchmaker gold yeast was used in conjunction with an Arabidopsis library of normalized cDNAs purchased from Clontech.

Cloning of constructs in yeast was performed via traditional methods to create gene truncations (primers shown in Supplementary Table 3-1 and 3-2). Constructs were inserted into either the pGADT7 (prey) or pGBKT7 (bait) plasmids and transformed into *E. coli.* The resulting *E. coli* clones were grown up and plasmid DNA harvested via miniprep columns from Midsci or Zymoresearch for transformation into Matchmaker gold yeast.

Tri-alanine substitution mutants were created to test the effects of mutating the residues highlighted by DCA analysis. These mutants were created by utilizing 'around-the-horn' PCR. In this approach we used pGAD or pGBK plasmid with Twinkle or Pol1A correctly inserted as template DNA for PCR. Using a forward primer and a reverse primer with 5' ends that line up back-to-back, we extended the length of the plasmid using high fidelity DNA polymerase in a traditional PCR reaction. The forward primer possessed a tri-alanine mismatch region flanked by homologous sequences of ~15 bp for Twinkle or Pol1A. The resulting PCR product was a plasmid sized blunt ended DNA molecule possessing the tri-alanine substitution we had designed. The blunt ends of this molecule were ligated together and transformed into *E. coli.* Colonies were picked, grown and harvested for plasmid DNA which was checked for correct insertion of the tri-alanine substitution via Sanger sequencing. Once verified, the plasmids were transformed into yeast and measured for interaction using selective media.

Direct Coupling Analysis

We calculated both direct and indirect interactions between amino acid residues in the DNA polymerase and Twinkle genes through a direct coupling analysis of sequences from 90 plant species. First, we manually downloaded each DNA polymerase and Twinkle gene from the National Center for Biotechnology Information (NCBI),[217] ensuring that each plant species included complete gene annotations for both genes.

After separating the downloaded genes into two FASTA files, one for each gene, we performed a multiple sequence alignment on each FASTA file separately. We used the following CLUSTAL OMEGA [218] command, where ${INPUT} is the FASTA file containing the unaligned genes and ${OUTPUT} is the FASTA file containing the aligned genes:

```
clustalo -i ${INPUT} > ${OUTPUT}
```

A comparison of different multiple sequence aligners by Pais, et al. [219] shows that CLUSTAL OMEGA performs relatively well for full-length gene sequences, similar to sequences used in our analysis.

Following the individual alignments of each set of sequences, we joined the aligned sequences for each species into a single FASTA file by adding 20 asparagine residues to the end of the Polymerase II gene followed by concatenating the Twinkle sequence for each species. This artificial buffer ensured that interactions identified between the two genes were not affected by combining the genes (e.g., residue proximity), and facilitated our ability to quickly differentiate between the two genes. We used a MATLAB implementation of DCA[220] to identify direct information and mutual information in each pairwise residue comparison between the two genes. The output file contains four columns: the position of the first amino acid residue, the position of the compared amino acid residue, the amount of direct information between the two residues, and the amount of mutual information between the two residues. We

used the following command to perform the direct coupling analysis, where ${INPUT} is the combined FASTA file and ${OUTPUT} contains the direct information and mutual information for each pairwise comparison:

```
matlab -nodisplay -nojvm -nosplash -r "dca ${INPUT} ${OUTPUT}"
```

Since the amino acid residues reported in the DCA analysis came from the combined multiple sequence alignments (i.e., the position of the first residue of Twinkle starts after the 20 asparagine residues that follow the last residue of Polymerase II), we determined which residues corresponded to the residues in the original unaligned *Arabidopsis thaliana* sequences and extracted only those pairwise comparisons from the DCA output file. We then renumbered the amino acid residues to be congruent with the original unaligned sequences (i.e., the first residue of Polymerase II was labeled position one in the first column of the output file and the first residue of Twinkle was labeled position one in the second column of the output file). Finally, using these labels, we created a heat map of the calculated mutual information using the matplotlib and scipy.interpolate.griddata libraries in Python version 2.7. We used the heat map to visually identify areas of higher mutual information between the two genes.

Thermophoresis

Cloning and expression

Twinkle lacking the first 91 codons was used as template for the construction of the DNA primase and helicase domains, and the RNAP and zinc finger subdomains.

The primase domain (residues 92-410), RNAP and zinc finger subdomains (residues 194-410 and 92-186, respectively) were cloned into pET-19b and purified as described.[128,136] The helicase domain (401-709) was cloned into the pCri-1b vector[221] and purified as described before, changing Tris-HCl to Potassium phosphate 7.0 in the buffer composition. Tags were removed using PreScission protease for the pET constructs and TEV protease for the construct in pCri-1b.

Microscale thermophoresis

Pol1B was labeled using NanoTemper's Monolith His-Tag Labeling Kit RED-tris-NTA at 100 nM concentration. The fluorescent labeled protein was used at a constant concentration of 20 nM. The ligands were titrated against labeled Pol1B in 16 serial dilutions from 12.8 μM for Twinkle, 30 μM for the helicase domain, 70 μM for the primase domain, 30 μM for the RNAP subdomain and 270 μM for the zinc finger subdomain. The reactions were incubated in PBS buffer + 0.05% Tween. The measurements were performed using a NanoTemper Monolith NT.115pico instrument and the analysis was conducted at 10% LED power and 50% MST power with standard capillaries. The data from the thermophoretic change was fitted according to the equation:

$$(concentration) = U + \frac{(B-U)(C + C_T + K_d - \sqrt{(C + C_T + K_d)^2 - 4 * C * C_T})}{2C_T}$$

where U is the response value of the unbound state, B is the response value of the bound state and CT is the final concentration of the fluorescent molecule.

qPCR

Leaf tissue was harvested from wild-type and Twinkle and PrimPol mutant plants at 7 and 14 dpi. High quality DNA suitable for qPCR was isolated from tissue samples via a cetyltrimethylammonium bromide (CTAB) protocol adapted from Minas et al.[172] qPCR was used to measure organelle genome copy number by measuring 3 mitochondrial and 3 chloroplast DNA targets (Supplementary Table 3-3). Mitochondrial targets included nad9 (NADH dehydrogenase iron-sulfur protein 3), orf25 (open reading frame, encodes plant b subunit of mitochondrial ATP synthase based on structural similarity), and cox1 (Cytochrome c oxidase subunit 1). Chloroplast targets included psbK (photosystem II reaction center protein K precursor), petD (Cytochrome b6-f complex subunit 4) and ndhH (NAD(P)H-quinone oxidoreductase subunit H). These targets are unique to mitochondria and chloroplast DNA to avoid overlap with the nuclear genome. AtRpoTp (phage-like RNA polymerase, nuclear encoded, plastid localized) was used as a standard to measure copy levels. Relative genome copy numbers were analyzed using an Applied Biosystems StepOne Plus qPCR machine and PowerUp SYBR green reagents. Technical and biological replicates were compiled and analyzed using the ΔΔCt method.[173]

APPENDIX 1: Co-IP of *Arabidopsis* Pol1A, Pol1B, and Twinkle

The following section contains unpublished data from experiments designed to support results found from the yeast-two-hybrid experiments performed in Chapter 3.

Abstract

Arabidopsis organellar proteins Pol1A, Pol1B, and Twinkle have shown positive interactions when tested using a yeast-two-hybrid system. To verify our findings from the yeast-two-hybrid, we designed a co-immunoprecipitation experiment to provide additional support. By modifying the bait and prey plasmids used for the yeast-two-hybrid analysis, we were able to perform this experiment while still taking advantage of cloning our products into yeast. Tagged proteins were successfully identified from total protein extracts. Pulldown experiments failed to display Pol1A, Pol1B, and Twinkle in the predicted patterns, but instead displayed a high abundance of Twinkle with no detection of Pol1A or Pol1B. This result is particularly odd, as Pol1A and Pol1B were both targets of the pulldown. If any result were expected, it would be to see high levels of Pol1A and Pol1B and none of Twinkle; however, the opposite result was observed. Development of this assay ceased after we were able to utilize thermophoresis to verify the results of our yeast-two-hybrid.

Introduction

Previous studies in our lab showed positive protein interactions of Twinkle with Pol1A and Pol1B. However, yeast-two-hybrids are notorious for high rates of false positives. To validate our yeast-two-hybrid experiments we designed a co-immunoprecipitation experiment that targeted Pol1A and Pol1B as pulldown targets. Based on our yeast-two-hybrid results, we predicted that Twinkle would pulldown with either one of the DNA polymerases.

We designed the pulldown experiments by modifying the same pGAD and pGBK vectors provided with the Clontech yeast-two-hybrid system. When used for yeast-two-hybrid analysis, these vectors attach GAL binding and activating domains on the N-terminal end of each target protein. We created new vectors from pGAD and pGBK that deleted these GAL domains (Figure A1-1). These plasmids retained the ADH1 promoter, allowing us to express the *Arabidopsis* proteins in yeast.

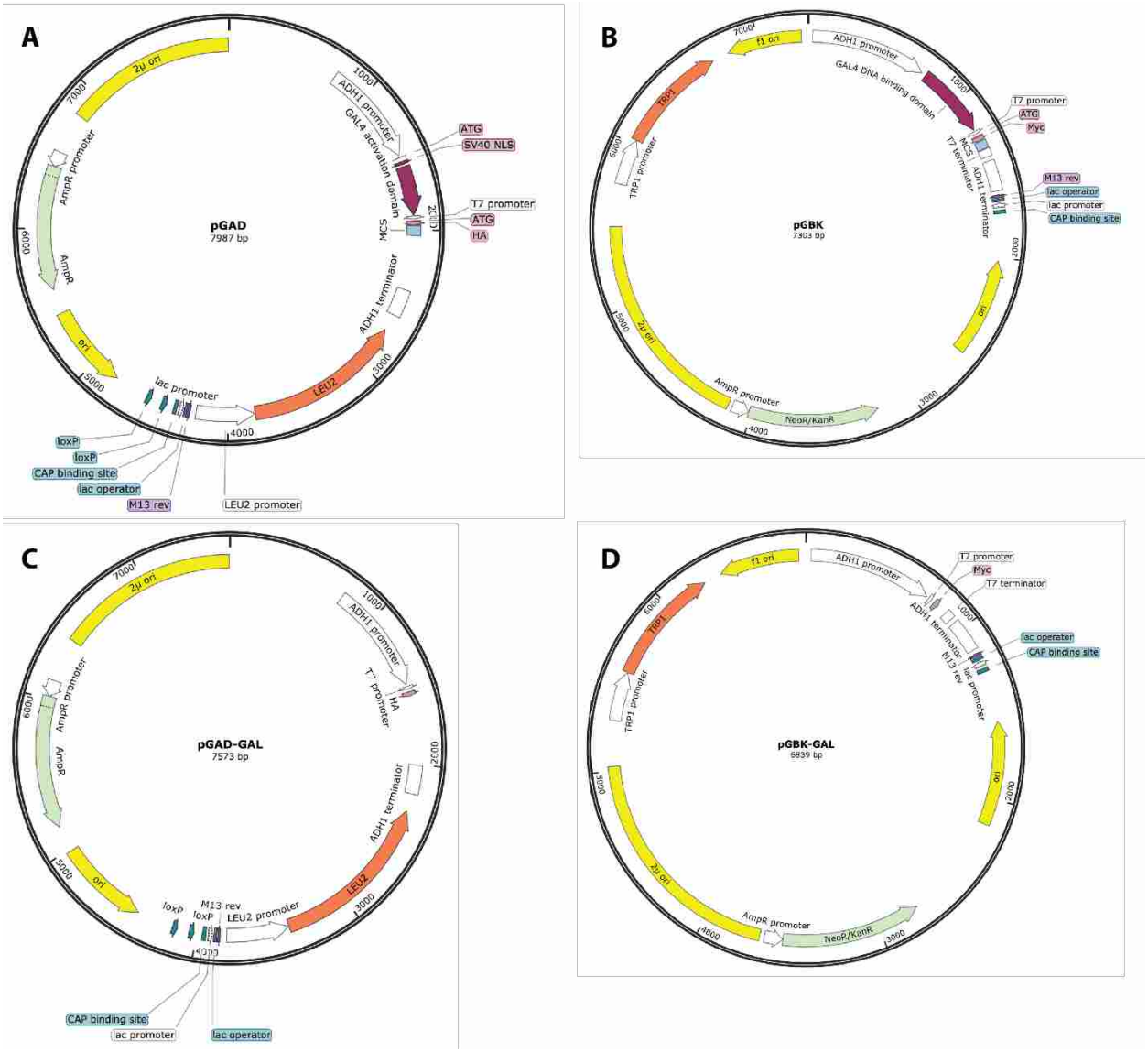

*Figure A1-1. Vectors used for protein expression in yeast cells. Pictured you can see features present in the original pGAD (**A**) and pGBK (**B**) vectors used in the Clontech yeast-two-hybrid system. Modified pGAD and pGBK vectors retain all features except for the GAL activating and binding domains (**C** and **D** respectively).*

Results

Modified pGAD and pGBK vectors were able to successfully transform into yeast cells. Expression of Twinkle, Pol1A, and Pol1B in yeast cells was confirmed by western blot analysis of total protein.

Twinkle was observed in pulldown experiments, however we did not see Pol1A or Pol1B (Figure A1-2).

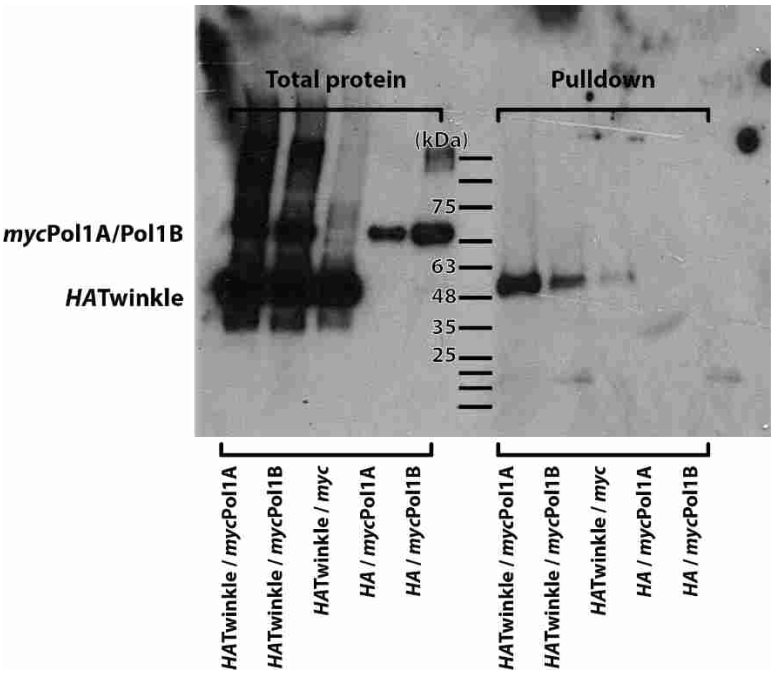

Figure A1-2. Co-immunoprecipitation of Twinkle with Pol1A and Pol1B. HATwinkle represent residues 40-435 of Twinkle protein with an HA tag present on the N-terminal end of the protein. mycPol1A/Pol1B represents residues 36-730 or 30-714 of Pol1A or Pol1B respectively. Pol1A and Pol1B both possess myc tags on the Ni-terminal end of the protein. In the images above we can see the attempts to pulldown Twinkle with Pol1A and Pol1B. As can be seen in the total protein images, each protein appears in the total protein blots but only Twinkle appears in the pulldown experiments. This is an unexpected result, as we performed the immunoprecipitation with a myc tag and Twinkle protein possesses an HA tag. It could indicate that the experiment was in fact successful and that levels of Pol1A and Pol1B were simply very low.

Conclusions

The presence of Twinkle in our pulldown experiments is troubling. Pulldown was performed targeting Pol1A and Pol1B as they both possessed the myc epitope whereas Twinkle possessed an HA epitope. If anything, we would expect to see high levels of Pol1A and Pol1B in the pulldown rather than Twinkle. Several explanations for

these results include; (1) Twinkle was able to pulldown effectively, but Pol1A and Pol1B were present in such low amounts they were undetectable or were degraded at some point in the experiment. (2) Twinkle was produced in such high abundance it was difficult to remove from the pulldown protein fractions. (3) Twinkle exhibited some type of affinity for the magnetic beads used in the pulldown.

These results led us to adopt an alternate approach. We used microscale thermophoresis to support our yeast-two-hybrid results.

Experimental design/methods

The vectors used to perform the yeast-two-hybrid analysis were modified from the pGAD and pGBK vectors used in the Clontech Matchmaker gold system. We modified each of these vectors to delete the GAL domains present in both the bait and prey plasmids. This created protein products that did not possess GAL domains, leaving only the protein along with an epitope that was already built into each vector. For this experiment, Twinkle was inserted into a modified pGAD vector with an HA epitope on the N-terminal end of the protein. Pol1A and Pol1B were inserted into modified pGBK vectors and possessed myc epitopes. Tags in both proteins were located on the N-terminal end of each protein. Expression of each protein was under the control of the constitutive ADH1 yeast promoter. Production of each protein was performed by growing the yeast in selective media for 24 hours at 30°, followed by mechanical lysis.

Protein pulldown was performed using magnetic beads purchased from Cell Signaling and following protocols outlined by the same company. These beads were covalently linked to monoclonal antibodies specific for the myc epitope that was attached to Pol1A and Pol1B (cell signaling #9B11). After pulldown, the purified products were treated with SDS and separated in mini SDS-PAGE gels. Western blots were performed by transferring gels to PVDF membranes using the iBlot system. Primary antibodies (rabbit monoclonal) specific to myc or HA tags were used to probe for Pol1A/B or Twinkle (Cell Signaling #71D10 and #C29F4 respectively). Secondary antibodies (anti-rabbit IgG Cell Signaling #7074) conjugated to horseradish peroxidase were used to prepare the blot for treatment with Luminol. Detection of proteins was performed using Luminol and horseradish peroxidase secondary antibodies provided by Clontech (Cell Signaling product #7074). Blots were exposed to X-ray film and developed to visualize interactions.

APPENDIX 2: Bacterial growth curves

The following appendix contains unpublished data from a study to see if *Arabidopsis* Pol1A or Pol1B can complement the activity of *E. coli* DNA polymerase 1.

Abstract

Arabidopsis organellar DNA polymerases Pol1A and Pol1B are bacteria-like proteins based on structural and phylogenetic characteristics. However, millions of years have given Pol1A and Pol1B time to evolve beyond their bacterial roots into more complex proteins. We tested *in vivo* whether Pol1A and Pol1B still have the ability to function in place of *E. coli* DNA polymerase I, or if evolutionary processes have moved it beyond that capacity. To do this, we transformed full length and Klenow fragments of Pol1A and Pol1B into *E. coli* DNA polymerase I mutants. We also cloned the *E. coli* gene polA, which encodes for DNA polymerase I, back into these mutants as a positive control. Unfortunately, we were unable to completely knock down expression of polA in our positive control *E. coli* strain. Despite this, we observed that the plant proteins were unable to complement the loss of DNA polymerase I in *E. coli* mutants.

Introduction

Pol1A and Pol1B are organellar DNA polymerases found in *Arabidopsis*. These proteins have gone through a number of names including polymerase gamma,[118] and

plant organellar polymerases (POPs).[120] Today, they are called Pol1A and Pol1B because of their similarity to bacterial DNA polymerase I.[119] In *E. coli* the gene responsible for coding DNA polymerase I is called polA.

The question we wanted to answer was simple: can *Arabidopsis* Pol1A and Pol1B function in place of *E. coli* DNA polymerase I? One key difference in the plant proteins is the lack of an identifiable 5′-3′ exonuclease domain as well as an overall increase in the length of the transcribed gene. The lack of this exonuclease region highly suggests that the plant proteins have evolved to function as organelle specialists, and work in conjunction with other enzymes that have taken over the 5′-3′ exonuclease role. Determining if Pol1A and Pol1B are able to function in *E. coli* will help us determine how far removed Pol1A and Pol1B are from their bacterial cousin.

To answer this question we obtained *E. coli* strains that were mutated for DNA polymerase I and complemented these mutations by transforming Pol1A, Pol1B, and the native *E. coli* polA gene back into the bacteria. In addition to full length Pol1A and Pol1B, we created Klenow only versions of Pol1A and Pol1B. This was to test if extensions in the other areas of the plant organelle DNA polymerases were responsible for a loss of function in *E. coli*.

Results

WT *E. coli* grew well as predicted at 42°C while mutant strain RS5065 failed to grow at all. RS5065 transformed with polA expression vectors also successfully grew

whether induced with arabinose or not. RS5065 transformed with full length and Klenow only Pol1A and Pol1B failed to grow successfully under any conditions.

Conclusions

Our intent was to create an expression system that would tightly control expression of *E. coli* polA as well as *Arabidopsis* Pol1A and Pol1B. In our positive control strains, induction was performed using arabinose and the cloned polA gene was able to successfully complement the exonuclease mutation of RS5065. However, we were unable to completely quench basal expression of *E. coli* DNA polymerase I in our expression plasmids. Therefore our negative control performed without induction still grew, suggesting that even the smallest amount of polA expression was able to rescue RS5065 mutants. (Figure A2-1).

In either case, *E. coli* transformed with full length or Klenow fragment Pol1A and Pol1B failed to grow when induced (Figure A2-1). Therefore, despite the failure of our polA controls, we can confidently state that the plant organellar DNA polymerases do not retain enough bacterial characteristics to function in place of *E. coli* DNA polymerase I.

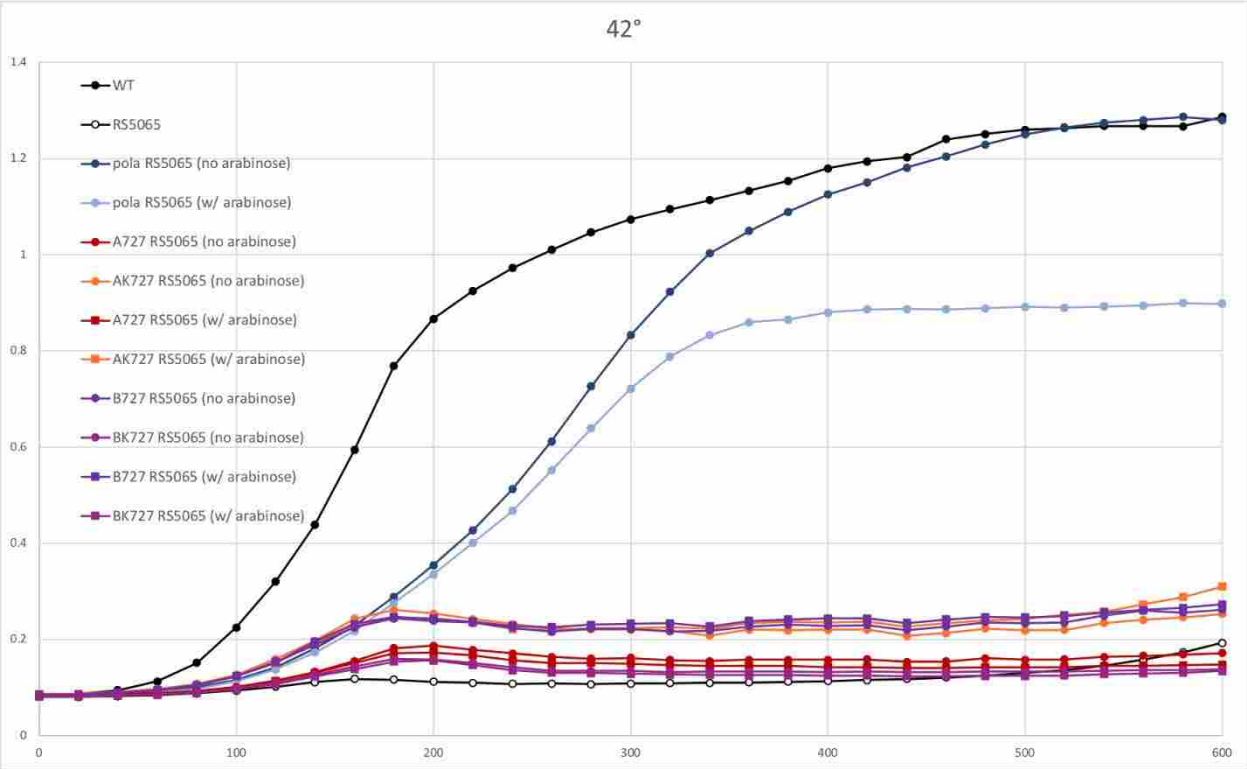

Figure A2-1. Bacterial growth curves of E. coli DNA polymerase I mutants. As pictured above, WT E. coli grows well at 42°C while the mutant RS5065 fails to grow at all. Positive and negative control polA clones both succeed to grow, indicating that even the smallest amount of polA expression in un-induced cultures is enough to complement the mutation in RS5065. Full length and Klenow only fragments of Pol1A and Pol1B were unable to grow when induced, indicating that these proteins no longer provide the necessary functions in bacterial cells.

Experimental design/methods

E. coli polA mutants were obtained from the *E. coli* Genetic Stock Center located at Yale University. The strain used for this study is designated RS5065, and was first described by Konrad.[222] The genotype of this mutant is: λ^-, *trpA33, IN(rrnD-rrnE)1, polA546*(tx,EX). This mutant is not a DNA polymerase I knockout but has mutated exonuclease activity when elevated to 42° C. This temperature sensitive mutation

effectively removes all DNA polymerase I activity, as exonuclease function is essential to the protein function.

Full length and Klenow only fragments of *Arabidopsis* Pol1A and Pol1B were cloned into expression vector pJG727 obtained from Dr. Joel Griffitts. This vector possesses the arabinose operon allowing for tight expression of cloned products (Figure A2-2). Klenow only fragments of Pol1A and Pol1B were made by creating clones that possessed only the conserved polymerase domain of each protein. For our experiments, Klenow only versions of these proteins consisted of residues 715-1050 and 699-1034 of Pol1A and Pol1B respectively. *E. coli* polA was also cloned into the same expression vector as a positive control. These were in turn transformed into RS5065.

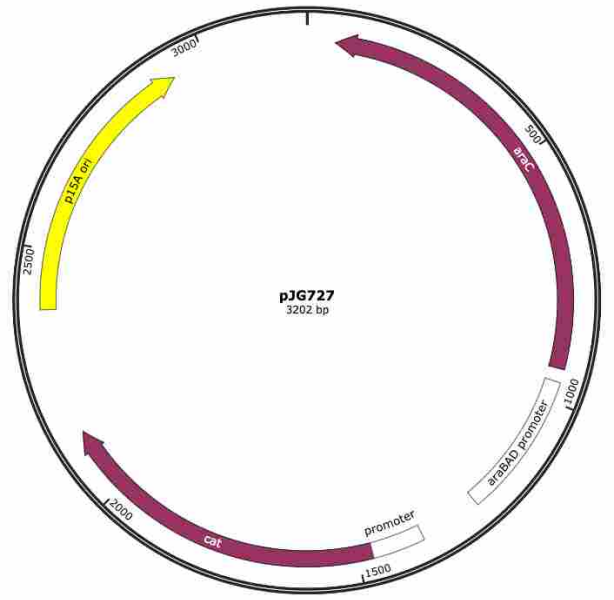

Figure A2-2. Expression vector pJG727. Important features to note include the chloramphenicol resistance gene (cat), araBAD promoter, and araC gene.

Growth curve experiments of these mutants and clones were conducted at 42° C to induce the temperature sensitive mutation in RS5065 and to measure for rescue by the native *E. coli* polA gene and *Arabidopsis* Pol1A and Pol1B genes. Cultures were grown to an approximate OD600 of 0.4 and induced so that the final concentration of arabinose was 0.1%. Growth was measured using a BioScreen C instrument which recorded OD600 absorbance every 15 minutes for a maximum of 24 hours.

APPENDIX 3: Yeast cDNA screen against Twinkle

The following appendix contains unpublished data from a cDNA library screen of all *Arabidopsis* genes against *Arabidopsis* organellar DNA helicase-primase Twinkle.

Abstract

The minimal DNA replisome in mammals has been characterized *in situ* consisting of DNA polymerase gamma (pol γ), and DNA helicase-primase TWINKLE. SSB1 protein also aids in replication by contributing to processivity of the complex. A similar system in T7 phage has been described consisting of gp5 DNA polymerase, gp4 DNA helicase-primase, and gp2.5 single stranded DNA binding protein. Similar proteins have been described in *Arabidopsis* and most likely form the basis of a minimal DNA replisome in plant organelles. We attempted to identify other proteins that may participate in the minimal DNA replisome of plant organelles. To identify these proteins we performed a yeast-two-hybrid screen against *Arabidopsis* Twinkle using a normalized yeast library that contained all cDNAs from *Arabidopsis.* We decided to use Twinkle as the bait protein as it most likely forms the central or main "hub" of protein interactions within the minimal DNA replisome of plant organelles. We were able to successfully identify and verify the interaction of 23 unique binding proteins, including Twinkle itself.

Introduction

Animal mitochondria and T7 phage form minimal DNA replisomes that are able to efficiently replicate DNA with a minimum of proteins. In animals, these proteins include DNA polymerase gamma (POL γ) and DNA helicase-primase TWINKLE. Single stranded DNA binding protein SSB1 is not required but when present confers proccessivity to the minimal replisome formed by POL γ and TWINKLE. T7 phage possesses homologs of POL γ, TWINKLE, and SSB1 called gp5, gp4, and gp2.5 respectively.

Arabidopsis also maintains homologs of these animal and phage proteins within its organelles. However, this does not exclude the possibility that other factors interact, enhance, or regulate the plant proteins. For example, T7 DNA polymerase protein gp5 requires *E. coli* thioredoxin as a host factor to become a processive enzyme.[223,224] Because of the greater complexity of *Arabidopsis*, we hypothesized that there could be many different binding partners that would be able to regulate the activity of Twinkle and control DNA replication.

Using a normalized yeast library that contains all the cDNAs of *Arabidopsis*, we probed for other proteins that may bind to Twinkle.

Results

Of the 24 positive results, we identified 23 potential novel binding partners of *Arabidopsis* Twinkle. The 24th result includes Twinkle itself. This should be expected as

Twinkle forms a hexamer or heptamer *in vivo.* Identified biding partners are listed in Table A3-1.

Conclusions

We identified 23 potential binding partners of Twinkle. However; analysis of these proteins has not led to any distinguishable pattern of functional groups with a possible role in organelle DNA replication.

A normalized yeast *Arabidopsis* cDNA library was purchased from Clontech. This library arrived in the prey vector (pGAD) provided by Clontech. Mating of yeast libraries was performed by co-culturing the yeast library with yeast possessing Twinkle in the bait vector (pGBK). Overnight mating experiments were performed in accordance to Clontech's recommended protocols.

Table A3-1. Binding partners of Twinkle discovered from *Arabidopsis* cDNA yeast library

Protein Name	NCBI reference
6-phosphogluconate dehydrogenase family protein	NM_001341465.1
SPFH/Band 7/PHB domain-containing membrane-associated protein	NM_124536.4
Genomic Chromosome 5 DNA, 2 iron, 2 sulfur cluster binding protein	NM_124551.5
lipoic acid synthase 1 (LIP1)	NM_001335715.1
Aldolase-type TIM barrel family protein (RSR4)	NM_001342589.1
GDP-L-galactose phosphorylase 1 (VTC2)	NM_118819.3
myb domain protein 25 (MYB25)	NM_129546.2
toprim domain-containing protein	NM_179404.3
cell wall invertase 2 (CWINV2)	NM_001339580.1
Rubisco methyltransferase family protein	NM_121430.4
FASCICLIN-like arabinogalactan protein 15 precursor	NM_115097.3
folic acid binding / transferase	NM_001335710.1
SWIB/MDM2 domain superfamily protein (CHC1)	NM_121421.4
temperature-induced lipocalin (TIL)	NM_125192.4
Disease resistance-responsive (dirigent-like protein) family protein	NM_117432.3
2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase protein	NM_124567.5
HSP40/DnaJ peptide-binding protein	NM_127615.3
Aldolase superfamily protein (FBA6)	NM_001336598.1
similar to GTP-binding protein (T7I23.11)	NM_100108.5
P-type ATP-ase 1 (PAA1)	NM_001342218.1
FK506-binding protein 16-2 (PnsL4)	NM_001342549.1
Translation initiation factor 2, small GTP-binding protein	NM_001340703.1
glutamine synthetase 1.3 (GLN1.3)	NM_112663.3
high chlorophyll fluorescence phenotype 173 (HCF173)	NM_001332254.1

Experimental design/methods

Mated yeast was plated on SD –leu –trp to ensure yeast had mated correctly and each cell possessed both pGAD and pGBK plasmids. Mating was performed to allow a minimum of 10X coverage of each *Arabidopsis* gene with Twinkle. This coverage was verified by counting the number of colonies that formed on SD –leu –trp media. Mated yeast was also plated on SD –leu –trp –his –ade media that selected for positive

interaction between bait (Twinkle) and prey (*Arabidopsis* cDNA) proteins. Colonies that successfully grew on SD –leu –trp –his –ade media were patched onto fresh plates, followed by DNA extraction, PCR amplification, and sequencing to identify the interacting cDNA gene.

Identified cDNA products were then tested for autoactivation. This was done by growing the target yeast to saturation in liquid SD –leu –trp for 2-3 days followed by zymolyase treatment to lyse the cells. Zymolyase was purchased from Zymo research and used in accordance to their recommended protocols. After zymolyase treatment, plasmid DNA was harvested from the yeast by subjecting the sample to a standard plasmid miniprep kit. The total quantity of plasmid DNA harvested by this method is very low, but high enough to successfully transform into *E. coli*. Transformed *E. coli* were subsequently grown in liquid culture and harvested using a standard plasmid mini kit to yield high concentrations of cDNA plasmid. This plasmid, possessing the cDNA from the original yeast cell, was then re-transformed into yeast along with an empty bait (pGBK) vector to test for autoactivation. Growth on SD –leu –trp –his –ade indicates that the cDNA product is able to activate the GAL promoter in the yeast without the need for a bait protein. This step revealed that over 60% of identified binding partners were able to autoactivate. Specifically, of the 62 potential binding partners we identified, only 24 passed the autoactivation test.

APPENDIX 4: Leaf-area and mass measurements of *Arabidopsis* Pol1A/Pol1B heterozygotes

The following appendix contains unpublished data that was taken to determine if there was a distinct phenotype associated with specific Pol1A/Pol1B genotypes. Some of the following was presented at the ICAR conference held in St. Louis MO in 2017.

Abstract

Arabidopsis possesses two organellar DNA polymerases called Pol1A and Pol1B. Previous research has shown that these two are somewhat redundant and able to complement the loss of the other. We wanted to see if there were phenotypic differences in Pol1A/Pol1B heterozygous plants that could easily be overlooked by casual observing. To create heterozygous plants, we crossed homozygous Pol1A and Pol1B T-DNA knockouts to create a completely heterozygous F1 generation. These plants were grown to create an F2 generation with every possible combination of heterozygosity for both Pol1A and Pol1B. Seeds from the F2 generation were planted to create an F3 generation. To accurately phenotype Pol1A/Pol1B heterozygotes, hundreds of F3 plants were digitally measured for shoot leaf area at 26 days post imbibition. These plants were also measured for mass at the same time. The results of our phenotyping

experiments did not determine any statistically significant difference in size or weight based on genotype.

Introduction

Phenotyping plants can provide valuable information on the physical effects certain genotypes have on plant development. *Arabidopsis* grows relatively quickly, and subsequent plants are genetically identical. *Arabidopsis* is also diploid, simplifying genetic analysis which can quickly become complicated in other plant species. Under identical growing conditions, any variance in phenotype can be reasonably assumed to be a result of altered genetics.

Previous experience in our lab has shown little difference in plant growth or development due to homozygous loss of either Pol1A or Pol1B. Heterozygotes also display little negative effect as long as at least one functional copy of Pol1A or Pol1B is present. We were interested in seeing if there were in fact phenotypic differences that could be easily missed by casual observing.

We possess *Arabidopsis* mutants that are T-DNA knockouts for Pol1A and Pol1B. By breeding these plants, we created completely heterozygous F1 plants. Seeds produced by the F2 generation are every possible combination of heterozygosity of Pol1A and Pol1B. These plants were then planted and grown under identical circumstances, grown and harvested at 26 days post imbibition (dpi). They were then

measured for shoot mass, and digitally scanned for analysis of shoot leaf area. After mass and area measurements, the plants were genotyped.

We were also interested to see if there were any significant differences in plant development between wild-type and homozygous Pol1A and Pol1B knockouts under low light conditions. We therefore analyzed WT and Pol1A and Po1lB homozygous knockouts under 100, 50, and 25 μmol $m^{-2}s^{-1}$ light conditions.

Results

F1 plants were confirmed to be Pol1A/Pol1B heterozygotes. F2 plants could be any one of nine different genotypes based on our breeding of homozygous Pol1A and Pol1B plants. These genotypes ranged from WT genotypes to complete knockouts for both Pol1A and Pol1B, although complete knockouts were not observed as this is most likely embryo-lethal.

Mass and leaf area measurements did not reveal any distinct phenotypes based on any particular genotype (Figure A4-1). Similarly, we did not detect any distinguishable difference in growth of WT and homozygous Pol1A and Pol1B knockouts under varying light conditions (Figure A4-2). It is important to note that these experiments were performed in simple growth chambers that do not have tight controls for humidity or light quality and intensity. Furthermore, planting soil conditions and watering can have great effects on plant development. If these

experiments were to be repeated they would require growth chambers with tight controls as well as consistent planting/soil conditions.

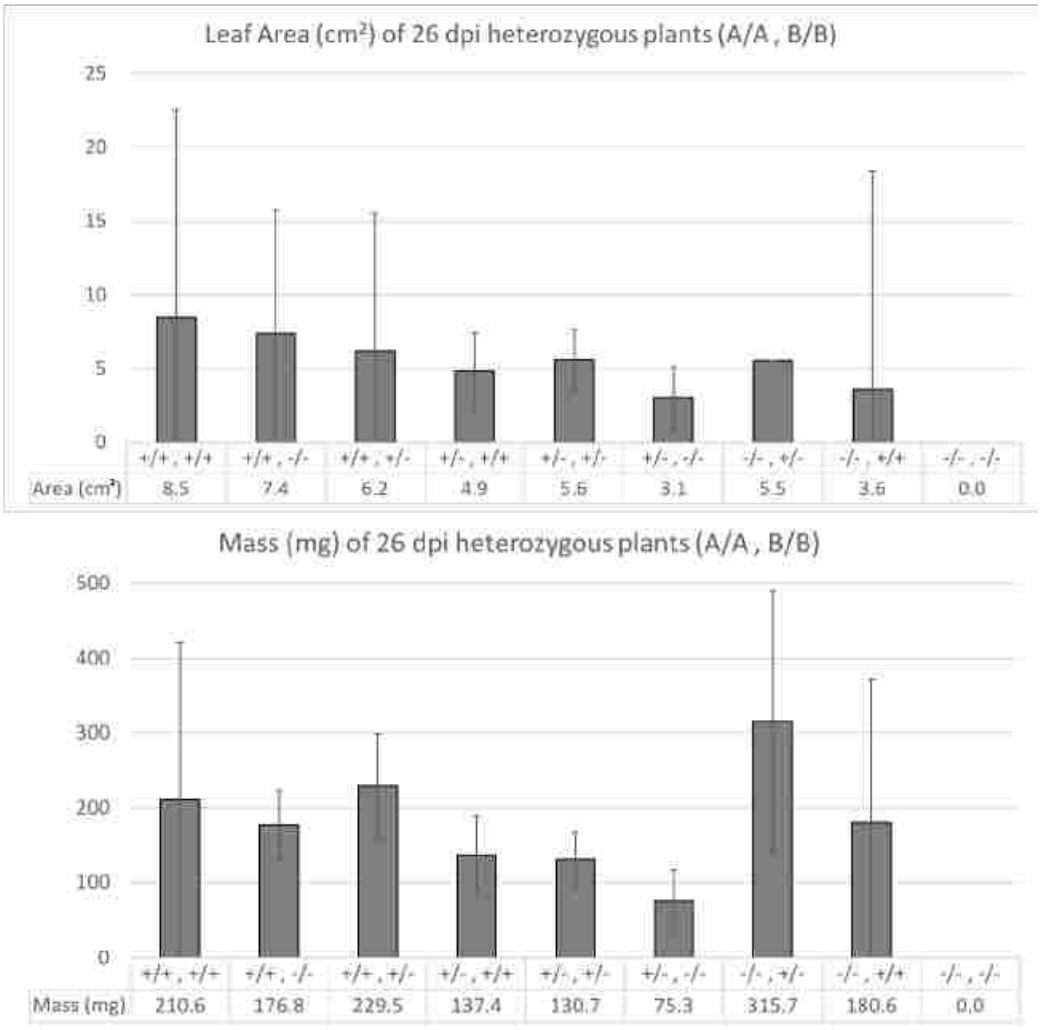

Figure A4-1. Area and mass measurements of heterozygous Pol1A/Pol1B plants. Genotype designations are as follows: (Pol1A/Pol1A, Pol1B/Pol1). For example, the designation +/-, -/- indicates the plant is heterozygous for Pol1A (Pol1A^{+}/Pol1A^{-}) and homozygous negative for Pol1B (Pol1B^{-}/Pol1B^{-}). In this figure, we see that as long as the plant maintains one functioning copy of Pol1A or Pol1B there is little to no effect on plant size or mass.

Conclusions

Lack of observable phenotypes suggest that one copy of either Pol1A or Pol1B is sufficient for plant survival. Pol1A and Pol1B share significant sequence homology, and

although evidence has supported somewhat specialized roles for each polymerase, there is likely overlap and redundancy between the two. A similar conclusion can be drawn from the growth analysis of WT and Pol1A and Pol1B homozygous mutants grown under varying light conditions.

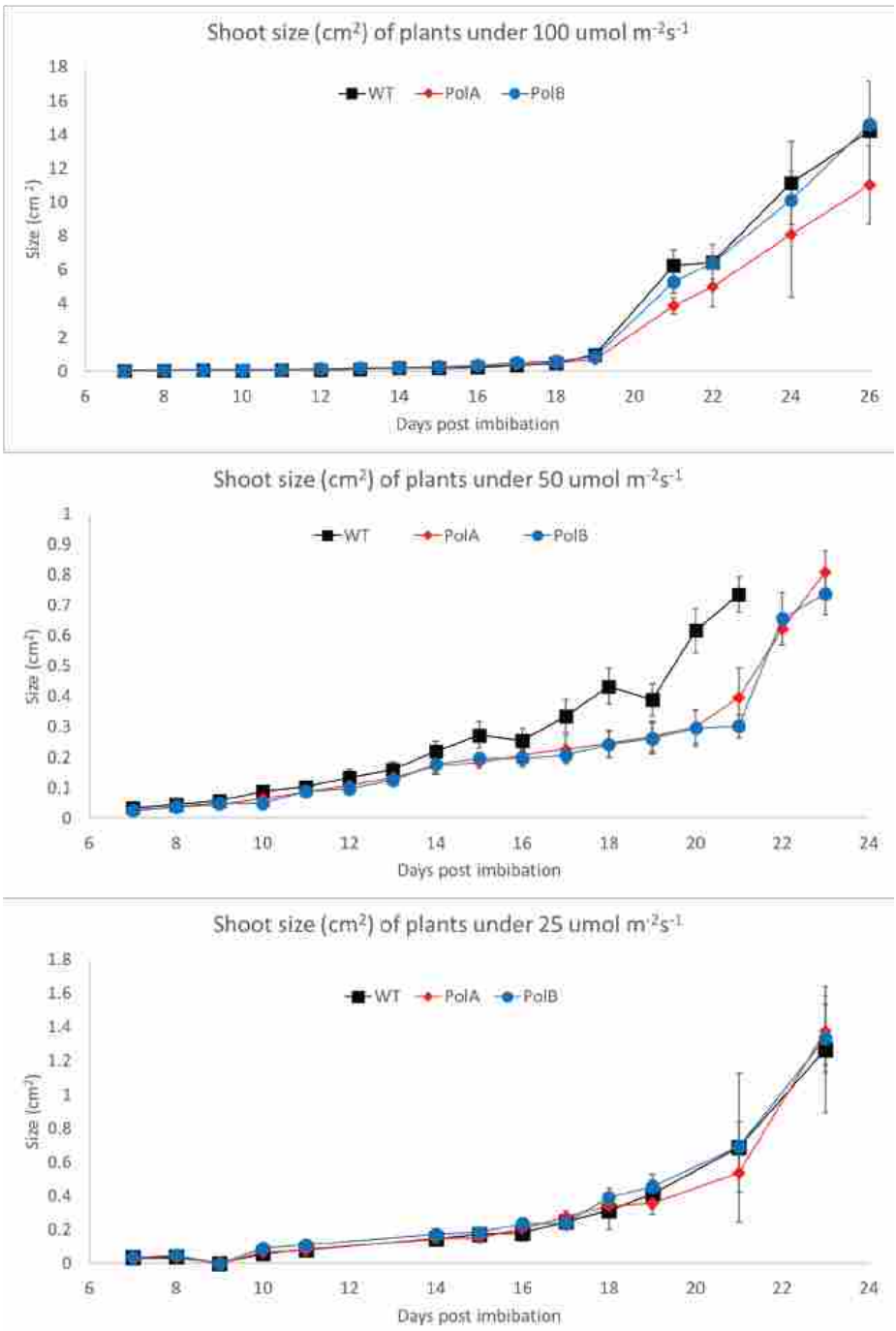

Figure A4-2. Growth of WT, Pol1A, Pol1B under 100, 50, and 25 µmol $m^{-2}s^{-1}$ of light. The results of this experiment were inconclusive and did not show a significant difference in plant size when grown under differing light conditions.
Experimental design/methods

Homozygous Pol1A and Pol1B knockout mutants were bred to create a fully heterozygous F1 generation. Seeds from the F1 generation were then harvested and planted as the F2 generation. At various time points during F2 development, plants were collected and measured for shoot mass and leaf area.

Shoot mass was measured by clipping the plants at the soil barrier and placing the resulting plant on a table top analytical scale. Leaf area measurements were made by taking the cut shoot of the plant and scanning the image with a simple desktop scanner. The resulting images were scaled and pixel counted using ImageJ to determine the area of the leaf.

Plant genotype was determined by taking the weighed and measured plant and subjecting it to a simple DNA isolation followed by PCR. Genotype was determined using primers specific to the gene or the T-DNA insert. Primers used for this zygosity testing are listed in Supplementary Table 2-3.

APPENDIX 5: mtDNA sequencing of *Arabidopsis* mitochondrial genome

The following is unpublished data that was submitted as a report to the Office of Research and Creative Activities. This report was in partial fulfillment of receiving a BYU graduate studies fellowship.

Abstract

The Arabidopsis mitochondrial genome is represented as a large circular DNA molecule. In actuality, single molecules of this size have never been observed. Rather, smaller circular and linear sub-genomic molecules of Arabidopsis mitochondrial DNA (mtDNA) are seen. We believe these smaller molecules constantly recombine with each other to initiate synthesis of new Arabidopsis mtDNA. To identify specific regions of DNA responsible for this initiation, we performed next generation PacBio and Illumina sequencing on samples of enriched Arabidopsis mtDNA. Doing so produced coverage maps that show DNA regions that are sequenced more often. We believe higher coverage regions correspond to DNA that is replicated more frequently and therefore identify specific regions actively involved in initiation of DNA replication.

Introduction

The inspiration for this experiment came from a study that showed origins of replication could be identified in the archaea *Haloferax volcanii* by deep sequencing the

genome and identifying 'peaks' in the sequence coverage.[225] This study identified four distinct peaks in sequence coverage which corresponded to four unique origins of replication. Using this logic, we determined that a similar approach could be applied to the mitochondrial genome of *Arabidopsis.*

The mitochondrial genomes of plants possess many repeats that make genome assembly tricky. For our experiment, we were worried that traditional short read Illumina sequencing would lead to high sequence coverage in certain regions that was not accurate. To overcome this, we performed long-read PacBio sequencing. The advantage of the long reads was to overcome repeats and correctly map the genome.

Results

Results of PacBio and Illumina sequencing were plotted on a graph of mitochondrial genome position versus coverage depth (Figure A5-1). The resulting graph exhibits large spikes at regions on the mitochondrial genome that correspond of chloroplast DNA contamination. Because of these large spikes we determined the best approach would be to perform whole genome mapping and then examine how the reads mapped differently to the mitochondrial genome. The mapping was performed using BYU's MaryLou supercomputer and graphed as described previously. To remove noise, a 5,000 and 10,000 point boxcar average was plotted alongside the raw data (Figure A5-2). This allowed for easier visualization of high copy regions.

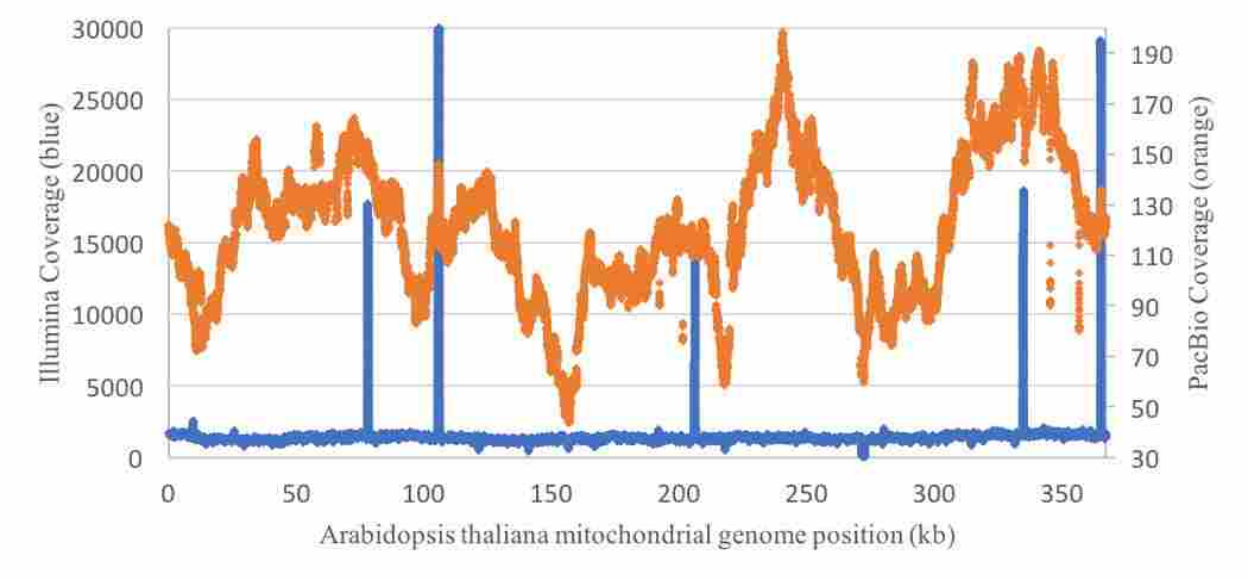

Figure A5-1. Raw mapping of PacBio and Illumina reads to the Arabidopsis mitochondrial genome. Note the 5 sharp peaks present in the Ilumina (blue) data. These peaks correspond to chloroplast DNA and are not representative of the mitochondria. These results led us to conduct a second round of mapping using the whole genome.

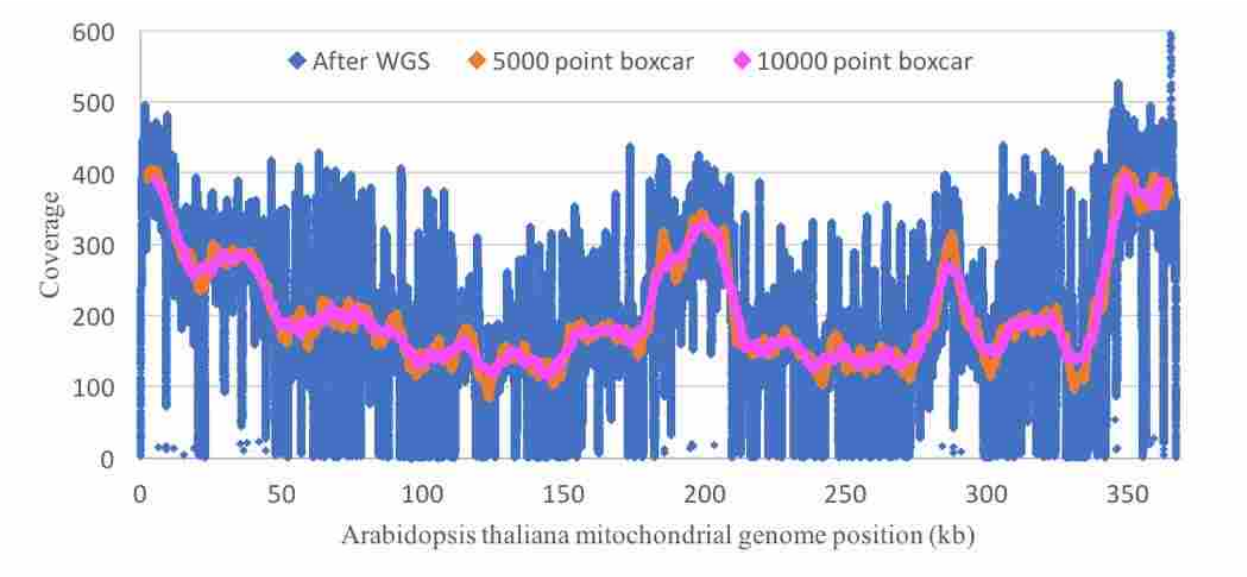

Figure A5-2. Mapping of Illumina reads to the mitochondrial genome after whole genome sequencing. To better visualize and smooth out the data a plot of a 5000 point boxcar average (orange) and a 10000 point boxcar average (pink) were superimposed on the original data (blue).

Coverage analysis after mapping to the entire genome reveals at least four locations of interest where distinct peaks can be seen. Peaks of these locations occur at approximately positions 0-8000, 193400-206400, 280500-294500, and 346500-357500. Analysis of these regions reveals that these peaks lie in relatively gene poor regions of the mitochondrial genome.

Conclusions

We have developed a method for enriching and analyzing non-pure mitochondrial DNA for deep sequencing. We found that a combination of Illumina, PacBio, and bioinformatics workflows is crucial to further defining high copy regions of the mitochondrial genome. We have identified four regions of the mitochondrial genome that may play a pivotal role in initiating DNA replication in the mitochondria. Many smaller features can be observed residing between the four larger peaks but as of this report an exhaustive analysis has not been performed. These smaller features likely correspond to areas where gene transcription is necessary. The data generated from these experiments provides a good base for bioinformatics analysis.

Experimental design/methods

Initially we attempted to isolate pure mitochondria so that we would be able to sequence only mitochondrial DNA. We were able to successfully isolate pure mitochondria using the following gradient centrifugation protocol modified from Lamppa et al[226] (steps can be scaled as needed):

Mito homogenization buffer

- Mannitol 31.9 g
- 1 M MOPS (pH 7.3) 15 mL
- 0.5 M EDTA (pH8) 1 mL
- Distilled H_2O to 500 mL

After autoclaving, cool solution and add 1g of BSA and 3g of polyvinvylpyrrolidone.

Percoll (for two gradients)

	Volume	Percoll	1 M sucrose	MOPS	dH_2O	10% BSA
60%	15 mL	9 mL	3.75 mL	150 µL	1.8 mL	300 µL
45%	30 mL	13.5 mL	7.5 mL	300 µL	8.1 mL	600 µL
27%	20 mL	5.4 mL	5 mL	200 µL	9 mL	400 µL
21%	20 mL	4.2 mL	5 mL	200 µL	10.2 mL	400 µL

Note: Except for steps 2-4, all procedures should be performed at 4°C.

1. Surface sterilize 2g of Arabidopsis seeds and imbibe in water for 3-4 days at 4°C.
2. Sow seeds on sterile soil covered with sterile cheesecloth or grow in a large flask with liquid media.
3. Grow seedlings for 7 days under continuous light or 16/8 light/dark cycles. (Seedlings will grow through cheesecloth).
4. Using a razor blade and a sawing motion, cut away seedlings from cheesecloth and place in a beaker on ice. If grown in liquid media seedlings are filtered through the cheesecloth and placed in the beaker.

5. Seedlings are homogenized at 4°C using a household blender and Mito homogenization buffer. Use two to four 5 second high speed bursts. Do not over homogenize.
6. Filter ground tissue through two layers of Miracloth, pre wet with Mito homogenization buffer.
7. Pellet chloroplasts in Sorvall GSA rotor (or equivalent) at 5500 rpm for 5 minutes.
8. Recover supernatant.
9. Pellet mitochondria by centrifuging in Sorvall SS-34 rotor at 13,000 rpm for 10 minutes.
10. While centrifuging, prepare step gradient by carefully layering the following Percoll concentration solutions, bottom to top: 6 mL 60%, 14 mL 45%, 8 mL 27%, and 8 mL 21%. Mark the border between the 60% and 45% layer.
    a. *During centrifugation, the step gradient will smooth out. Isolated mitochondria may not be visible after centrifuging but will band at the border of the 60% and 45% Percoll gradients.*
11. After centrifugation carefully decant supernatant. Using a paintbrush pre-wetted with Mito homogenization buffer, carefully resuspend the pellet.

12. Gently layer resuspended pellet on the top of the gradient created in step 10. Centrifuge in Sorvall SS-34 rotor at 10,500 rpm for 30 minutes with the brake off.
    a. *Remaining or broken chloroplasts are likely still visible after centrifuging. These will travel to the bottom and tops of the tube. A faint brownish band of mitochondria may or may not be visible at the border of the 60%-45% Percoll gradients.*
13. Use a pipette to transfer fractions from the top of the gradient.
14. Pool the fractions containing mitochondria.
15. Pellet mitochondria by centrifuging in a Sorvall SS-34 rotor at 13,000 rpm for 10 minutes.
16. (Optional) Decant supernatant and gently resuspend pellet with Mito homogenization buffer. Pellet mitochondria in Sorvall SS-34 rotor at 13,000 rpm for 10 minutes.

Although pure, the amount of isolated mitochondria was small, and the resulting DNA isolated was minimal. We needed much more DNA to successfully perform both PacBio and Illumina sequencing.

We then decided to enrich rather than purify mitochondria. We decided to do this because we would still get sufficient coverage of the mitochondrial genome. To

enrich for mitochondria we utilized differential centrifugation. We did an initial centrifugation at 1,000g for 5 minutes to remove large cell debris, followed by two 6,000 g spins at 10 minutes each to remove a large fraction of chloroplasts. The final centrifugation was performed at 21,000g for 30 minutes to enrich for mitochondria.

DNA from the enriched samples was extracted following a CTAB, phenol-chloroform protocol.[172] After extraction, we verified that our enriched sample was fifteen times more concentrated for mitochondrial DNA using qPCR. Large amounts of high quality high molecular weight DNA (~5 μg) were required for PacBio sequencing. Illumina sequencing only required a maximum of 1 μg of DNA although half or even a quarter of this amount was sufficient to gather the needed data.

Following high quality DNA extraction from enriched organelles, DNA was prepared for sequencing by the DNA sequencing center of Brigham Young University. DNA was prepared for PacBio sequencing by following two protocols recommended by PacBio. These include 'Guidelines for Using a Salt:Chloroform Wash to Clean Up gDNA' and 'Guidelines for Preparing 20 kb SMRTbell™ Templates' both available from PacBio. DNA for 250 bp paired end Illumina sequencing was prepared by shearing DNA on a Covaris ultrasonicator and cleaning up the DNA with an AMPure cleanup treatment. DNA was then processed using the NEBNext® Ultra™ II End Repair/dA-Tailing Module followed by NEBNext® Ultra™ II Ligation Module. Standard Illumina index primers purchased from IDTDNA were used as ligated adaptors.

Conclusions and future directions

In this dissertation we have compared the organellar DNA polymerases Pol1A and Pol1B of *Arabidopsis* to bacterial DNA polymerase I and DNA helicase-primase Twinkle to T7 phage gp4 protein. We have demonstrated that mutation of Pol1A and Pol1B have only small effects on plant organelle genome copy number. We have also shown how, despite lower copy numbers, plant development continues mostly unaffected as long as a single copy of either DNA polymerase is present. In addition, we show that Pol1A or Pol1B interact with Twinkle, similar to minimal DNA replisomes in animal mitochondria and T7 phage.

It is tempting at this point to claim that we have proven that plants utilize a minimal DNA replisome equivalent to those in animal mitochondria and T7 phage. However, as discussed in chapter 3, plants that are homozygous for a T-DNA insertion mutation in Twinkle display no growth defects and appear as healthy as WT plants. These mutant plants also do not display any differences in organelle genome copy number. Mutants of Twinkle in animals result in embryo lethal phenotypes, and T7 cannot function without its appropriate homolog. The major question left to answer is this: If plants possess functioning Twinkle and it interacts with Pol1A or Pol1B to replicate DNA similar to animals and T7 phage, why are plants totally unaffected by its loss?

While we cannot answer that question here, it's important to remember that plant organelle genomes are much larger and complex than T7 phage and animal mitochondria. T7 phage possesses a genome approximately 40 kb in size and nearly all animal mitochondrial genomes are about 16.5 kb. *Arabidopsis* chloroplast and mitochondrial genomes are orders of magnitude larger, 155 kb and 366 kb respectively. This stark difference in size clearly displays different evolutionary pressure between plants and animals. And while mitochondrial genomes in animals are always small circular molecules, there are many different DNA structures in plant organelles, particularly mitochondria. These structures include linear, linear branched, rosette-like, and catenane DNA molecules.

These stark differences make DNA replication in plant organelles much more complicated. Replication in animal mitochondria is fairly homogenous, always involving the same proteins, always initiating from the same site, and always resolving the same way. I believe plants use many other proteins to drive replication in plant organelles under different circumstances and based on different DNA structures. The next major advancement in studying plant organelle DNA replication will detail how these proteins affect DNA replication under different circumstances. I believe part of that research will involve reconstituting a functioning DNA replisome *in vitro*. Other studies have successfully accomplished this feat using human mitochondrial proteins,[95] but this was a very simple system for a much smaller genome. The minimal replisome

of plants likely utilizes more proteins and will be much more difficult to assemble. And while animals maintain only one minimal mitochondrial DNA replisome, I believe plants maintain two or more, based on the environment and based on the DNA structure being replicated. For example, DNA recombination occurs frequently in plant mitochondria. Plant mitochondria possess many recombination proteins that are bacteria-like and others that are unique to plants. These proteins are most likely involved in DNA replisomes present in the organelles.

Candidates for future study include the Whirly proteins (Figure C-1). As mentioned in chapter 1, these unique-to-plant proteins are heavily involved in recombination, organelle biogenesis and health, transcription regulation, and potential nuclear-organelle communication. These functions have been described from only a handful of publications, showing there is much more to be learned by studying the Whirlies. Studying how the Whirlies came to be unique plant proteins may also highlight key moments in evolution where animal mitochondria diverged from plants. The Whirly proteins also contribute to the unique mosaic of proteins involved in organelle DNA replication. This mosaic includes bacterial, phage, and plant proteins.

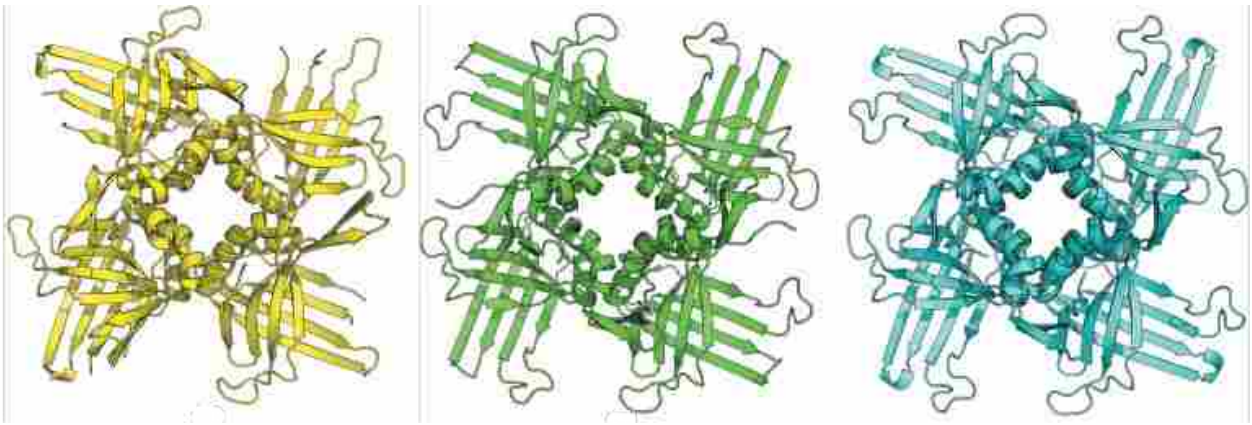

Figure C-1. The Whirly proteins. Whirly1 (yellow), Whirly2 (green), and Whirly3 (blue), are recombination proteins that are unique to plants. These proteins provide interesting research opportunities and will most likely be involved in the next major research avenue of plant organelle DNA replication. Taken from Cappadocia et al.[227]

Supplementary materials

Supplementary Table 2-1. Primers used for qPCR analysis of mitochondrial and chloroplast genome copy number

Target	Genome	Gene function
AtRpoTp	Nuclear	DNA-directed RNA polymerase 3
nad9	Mitochondrial	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3
orf25	Mitochondrial	Predicted ATP synthase b subunit
cox1	Mitochondrial	Cytochrome c oxidase subunit 1
psbK	Chloroplast	Photosystem II reaction center protein K
petD	Chloroplast	Cytochrome b6-f complex subunit 4
ndhH	Chloroplast	NAD(P)H-quinone oxidoreductase subunit H

Primer	Sequence	Tm °C
5′ AtRpoTp	CTGAATGCAGGTCGAAACTCGGG	60.1
3′ AtRpoTp	GCTTGGAAGCCGTCTGCTAGAAC	60.1
5′ nad9	GTGGGAGCGAGAAGTTTGGGATATG	59.4
3′ nad9	GGGTCATCTCAATGGGTTCAGAAACC	59.5
5′ orf25	TCAAAGTGACTCTCGACGGGAGC	60.6
3′ orf25	TGCCACAAATTCGCAAGCTGATCC	60.5
5′ cox1	GAAGTAGGTAGCGGCACTGGG	59.7
3′ cox1	ATTCCAGGTCCACGCATGTTGAAG	59.7
5′ psbK	GTCGCCAAATTGCCAGAGGC	59.7
3′ psbK	CGGCTTGCCAAACAAAGGCTAAGAG	60.7
5′ petD	TATTACGGGGAACCCGCATGG	63.6
3′ petD	GCAAAAGGATCCGCAGGTTCACC	60.9
5′ ndhH	GACTTCCAGGGGGTCCCTATGAG	60.5
3′ ndhH	CCCAACTCCCCTTTTGGAGCTTC	60.2

Supplementary Table 2-2. Primers used for RT-PCR analysis of Pol1A and Pol1B expression

Primer	Sequence	Tm °C	Target/Purpose
Actin_F	TCCCTCAGCACATTCCTGCAGAT	60.5	Nuclear control/reference
Actin_R	AACGATTCCTGGACCTGCCTC	60.8	for expression comparisons
RTPolA_F	TTCCGGCGTCAAAGTCACGTGC	62.6	Pol1A gene
RTPolA_R	TGCACTTCCCTGGACTGGAGTGT	62.4	
RTPolB_F	CCTGAATACCGTTCACGTGCCCA	61.5	Pol1B gene
RTPolB_R	AGCCGCACTTCCCTGAACAGGA	63.1	

Supplementary Table 2-3. Primers used for zygosity testing

Primer	Sequence	Tm °C	Target/Purpose
PolA_F	TTGAAGAGCTTCAGCGAGAAG	54.9	Pol1A gene
PolA_R	TAGCATGACATGCCTCCTTTC	54.9	
PolB_F	TTACCAAAAGCATCATCCTGG	53.0	Pol1B gene
PolB_R	AGAGTTTTCGTGTTCCCCATC	55.0	
Lbb1.3-1	ATTTTGCCGATTTCGGAAC	51.5	T-DNA specific primer

Supplementary Table 3-1. Primers used to create gene truncations

Primer name	Sequence	Pos.	Features/Restriction cut site
[A1]5' (1–)	GATC CCCGGG T* ATGGCCATGGGGGTTTC	1	SmaI
[A2]5' (36–)	GATC CCCGGG T* CCACTCCCTTCCTTCCTC	36	Excludes predicted signal peptide. SmaI
[A3]5' (275–)	GATC CCCGGG T* GCGAAGGATACCGTGGC	275	15 AA before 3'-5' exonuclease domain. SmaI
[A4]5' (290–)	GATC CCCGGG T* GTCCATTCCTGTGATACAGAGGT	290	3'-5' exonuclease domain border. SmaI
[A5]5' (640–)	GATC CCCGGG T* TGGCCCTCTGTAGGTGG	640	Starts at DNA polA superfamily domain. SmaI
[A6]5' (715–)	GATC CCCGGG T* TGTCATGCTATTGCCTCATTATGTG	715	15 AA before polymerase domain. SmaI
[A7]5' (730–)	GATC CCCGGG T* TTGATCTCAAATTTTATTCTTCCGTTACAGG	730	Polymerase domain border. SmaI
[A8]3' (–290)	GATC GGATCC TTA GACATGATTCCTAAACTGATTCACGAG	290	3'-5' exonuclease border. Stop codon. BamHI
[A9]3' (–305)	GATC GGATCC TTA TTCTTCCTTAACCTCAATCCCG	305	15 AA into 3'-5' exonuclease. Stop codon. BamHI
[A10]3' (–523)	GATC GGATCC TTA TTTTACAAGAAGTTCACCAAAGGGTC	523	3'-5' exonuclease border. Stop codon. BamHI
[A11]3' (–538)	GATC GGATCC TTA CTCAGCAAGATACTCTCTATCTACAAGT	538	15 AA after 3'-5' exonuclease. Stop codon. BamHI
[A12]3' (–654)	GATC GGATCC TTA TTTCCCAGCTAACTCTTTCAAAACATC	654	15 AA into DNA polA superfamily. Stop codon. BamHI
[A13]3' (–730)	GATC GGATCC TTA CAAAGAGTCTATAGAGCAAACTTCACA	730	Polymerase domain border. Stop codon. BamHI
[A14]3' (–745)	GATC GGATCC TTA GCCTGATACATTACTTCCCTGTAAC	745	15 AA into polymerase domain. Stop codon. BamHI
[A15]3' (–1050)	GATC GGATCC TTA CTATTTGGCAGCATACCAGTTTTGA	1050	Last AA of Pol1A. Stop codon. BamHI
[B1]5' (1–)	GATC GAATTC ATGGGGGTTTCTCTTCGTCA	1	First AA of Pol1B. EcoRI
[B2]5' (30–)	GATC GAATTC GTCCCTCGCCGTCGAAT	30	Excludes predicted signal peptide. EcoRI
[B3]5' (249–)	GATC GAATTC AATGTGTCTAGTGCAAAGGAAACC	249	15 AA before 3'-5' exonuclease. EcoRI
[B4]5' (264–)	GATC GAATTC TATAGGAATCTTGTCCATGCTTGC	264	3'-5' exonuclease domain border. EcoRI

[B5]5' (618–)	GATC GAATTC GGCTGGCCCTCTGTTAG	618	DNA polA superfamily border. EcoRI
[B6]5' (699–)	GATC GAATTC TGCCATGCTATTGCTGCA	699	15 AA before polymerase domain. EcoRI
[B7]5' (714–)	GATC GAATTC TTAATATCAAATTTTATCCTTCCTTTACAGGGAA GATC GGATCC TTA	714 264	Polymerase domain border. EcoRI
[B8]3' (–264)	ATATTGATTCATGAGCAGAGCCAC GATC GGATCC TTA	279	3'-5' exonuclease border. Stop codon. BamHI
[B9]3' (–279)	ATCAATCCTGGATACCTCTGTATCG		15 AA into 3'-5' exonuclease. Stop codon. BamHI
[B10]3' (–501)	GATC GGATCC TTA TTTGGCAAGAAGTTCACCAAAAG	501	3'-5' exonuclease domain border. Stop codon. BamHI
[B11]3' (–516)	GATC GGATCC TTA CTGCGCCAAATAATCCCTATCT	516	15 AA after 3'-5' exonuclease. Stop codon. BamHI
[B12]3' (–633)	GATC GGATCC TTA TTTCCCAGCTAAGGCTTTCAAG	633	15 AA into DNA polA superfamily. Stop codon. BamHI
[B13]3' (–714)	GATC GGATCC TTA TAAGGAATCAATGGAGCAAACTTCAC	714	Polymerase domain border. Stop codon. BamHI
[B14]3' (–729)	GATC GGATCC TTA TCCTGACACGTTACTTCCCT	729	15 AA into polymerase domain. Stop codon. BamHI
[B15]3' (–1034)	GATC GGATCC TTA TTATTTGCCAGCATACCAGTTCTG	1034	Last AA of Pol1B. Stop codon. BamHI

[S1]5' (1–)	GATC GAATTC ATGAACTCACTCGCCATTAGAGT	1	First AA of SSB1. EcoRI
[S2]5' (55–)	GATC GAATTC CTTCAACCTCATGGAGTTGATCC	55	15 AA before SSB1 protein family domain. EcoRI
[S3]5' (70–)	GATC GAATTC GGTGTTCATAGGGCGATTATTTGT	70	SSB1 family domain border. EcoRI
[S4]3' (–70)	GATC GGATCC TTA ACCGCGAAATCCCCAAC	70	SSB1 family domain border. Stop codon. BamHI
[S5]3' (–85)	GATC GGATCC TTA TAACGGTGCTTGCCCTAC	85	15 AA into SSB1 family domain. Stop codon. BamHI
[S6]3' (–175)	GATC GGATCC TTA ACGACGAACGCAAATCTCAG	175	SSB1 family domain border. Stop codon. BamHI
[S7]3' (–201)	GATC GGATCC TTA CTAAATCAATCCTTCTTTTAGCTCATCAAAAG	201	Last AA of SSB1. Stop codon. BamHI

[T1]5' (1–)	GATC CATATG ATGCGATTTTTGCTTCGTTTACCA	1	First AA of Twinkle. NdeI
[T2]5' (40–)	GATC CATATG TACCCTTCTTCTCCTTCTTATTCTTCA	40	Excludes signal peptide. NdeI
[T3]5' (265–)	GATC CATATG AAGACACGGAGGATCTTATATGGT	265	15 AA before Uniprot primase domain. NdeI
[T4] 5' (280–)	GATC CATATG TCTGAAGTCATTATAGTTGAAGGGGA	280	Uniprot primase domain border. NdeI
[T5] 5' (405–)	GATC CATATG GCTGAGCCATATCCTATACTAGGA	405	15 AA before helicase domain. NdeI

	GATC CATATG	420	
[T6] 5' (420–)	TTTGATGAAATTGATGCCTACTATGATAGAAC		Helicase domain border. NdeI
	GATC GAATTC TTA	280	
[T7] 3' (–280)	AGATGTTTTTTCTATGTCATCAAGACCA		Uniprot Primase domain border. Stop codon. EcoRI
	GATC GAATTC TTA	295	
[T8] 3' (–295)	TTCCATTGCAAGTTTATCTATCTCCC		15 AA into primase domain. Stop codon. EcoRI
	GATC GAATTC TTA	420	
[T9] 3' (–420)	AAAGAAATCTTTGAAGGAGAATAATCCTAGTATAGG		Helicase domain border. Stop codon. EcoRI
[T10] 3' (–435)	GATC GAATTC TTA ATACTCGTGCCCATGTGTTC	435	15 AA into helicase domain. Stop codon. EcoRI
[T11] 3' (–709)	GATC GAATTC TTA TCAGTACCGCTTGGGTGA	709	Last AA of Twinkle. Stop codon. EcoRI

*Extra thymine nucleotide added to keep truncation in frame with cloning plasmids

Supplementary Table 3-2. Primers used to create 10 residue truncations of Twinkle

Primer Name	Sequence
Forward primers	
[T2.1] 5' (AA:50–)	GATC CATATG AGACAAGTATCTTCCGTTTCTAGAAGA
[T2.2] 5' (AA:60–)	GATC CATATG CGACCAGTTTTGGCCTC
[T2.3] 5' (AA:70–)	GATC CATATG AAAAACAGTCCTTATTACCAAAGGAC
[T2.4] 5' (AA:80–)	GATC CATATG GGTTTATCATCTTACAATTCAATCCCC
[T2.5] 5' (AA:90–)	GATC CATATG GTCCCAACTCCTGTTGATACTG
[T2.6] 5' (AA:100–)	GATC CATATG GCAGATAAGAGGGTTGTTCTATCT
[T2.7] 5' (AA:110–)	GATC CATATG GTGACTTTGAGGCGTAAATTGG
[T2.8] 5' (AA:120–)	GATC CATATG GGAGTTGATGCTGAAAACTGC
[T2.9] 5' (AA:130–)	GATC CATATG CAACATAGTGGCTTGATATGTCC
[T2.10] 5' (AA:140–)	GATC CATATG GAAGGTGGAAACTCTGGAGA
[T2.11] 5' (AA:150–)	GATC CATATG TCTCTTTTTATAGCCCCTGATGG
[T2.12] 5' (AA:160–)	GATC CATATG GCTACATGGAATTGCTTTAGGG
[T2.13] 5' (AA:170–)	GATC CATATG GGGTTAAAAGGTGGAGTTCG
[T2.14] 5' (AA:180–)	GATC CATATG GGGTTGGCATCTGCTGAT
[T2.15] 5' (AA:190–)	GATC CATATG GTTGAAAGAAAAATTACGGTGGAGG
[T2.16] 5' (AA:200–)	GATC CATATG GAGCTAGAACCTCTCTGTGAT
[T2.17] 5' (AA:210–)	GATC CATATG GATTATTTCGCTGCAAGAGCG
[T2.18] 5' (AA:220–)	GATC CATATG AAAACACTCGAGAGAAATCGGG
[T2.19] 5' (AA:230–)	GATC CATATG AAAAGAATAGGTGACGAGATTGTAATTG
[T2.20] 5' (AA:240–)	GATC CATATG TTTACTTATTGGCAAAGAGGGGAG
[T2.21] 5' (AA:250–)	GATC CATATG AGTTGCAAGTACCGGTCTC
[T2.22] 5' (AA:260–)	GATC CATATG TTCTTTCAGGAAAGGAAGACACG
[T2.23] 5' (AA:270–)	GATC CATATG TTATATGGTCTTGATGACATAGAAAAAACA
[T2.24] 5' (AA:280–)	GATC CATATG TCTGAAGTCATTATAGTTGAAGGGG
[T2.25] 5' (AA:290–)	GATC CATATG GATAAACTTGCAATGGAAGAAGCT
Reverse primers	
[T10.25] 3' (AA:–185)	GATC GAATTC TTA ATCAGCAGATGCCAACCC
[T10.24] 3' (AA:–195)	GATC GAATTC TTA CGTAATTTTTCTTTCAACCTTCTCTATAGG
[T10.23] 3' (AA:–205)	GATC GAATTC TTA ACAGAGAGGTTCTAGCTCTATACC
[T10.22] 3' (AA:–215)	GATC GAATTC TTA TCTTGCAGCGAAATAATCTTGAATC
[T10.21] 3' (AA:–225)	GATC GAATTC TTA ATTTCTCTCGAGTGTTTTCCGT
[T10.20] 3' (AA:–235)	GATC GAATTC TTA CTCGTCACCTATTCTTTTCTGC
[T10.19] 3' (AA:–245)	GATC GAATTC TTA TCTTTGCCAATAAGTAAACGCAAT
[T10.18] 3' (AA:–255)	GATC GAATTC TTA AGACCGGTACTTGCAACTC
[T10.17] 3' (AA:–265)	GATC GAATTC TTA CTTCCTTTCCTGAAAGAACATCTTAG
[T10.16] 3' (AA:–275)	GATC GAATTC TTA GTCATCAAGACCATATAAGATCCTCC

[T10.15] 3' (AA:–285)	GATC GAATTC TTA AACTATAATGACTTCAGATGTTTTTTCTATGT
[T10.14] 3' (AA:–295)	GATC GAATTC TTA TTCCATTGCAAGTTTATCTATCTCCC
[T10.13] 3' (AA:–305)	GATC GAATTC TTA AACGGATACACAATTGAGAAAACCA
[T10.12] 3' (AA:–315)	GATC GAATTC TTA CGAAGAAACCTTCGCTGGA
[T10.11] 3' (AA:–325)	GATC GAATTC TTA CGTGTCCTTGTCTTCCGAT
[T10.10] 3' (AA:–335)	GATC GAATTC TTA GTCATTGCAATTCCATAGAAATTTATACT
[T10.9] 3' (AA:–345)	GATC GAATTC TTA AATAACAATTCGAGACGCCTTTTTTAG
[T10.8] 3' (AA:–355)	GATC GAATTC TTA AGCTTGACCAGGTCCATC
[T10.7] 3' (AA:–365)	GATC GAATTC TTA ACCCAAACGCCGTGC
[T10.6] 3' (AA:–375)	GATC GAATTC TTA CGGCCACTTGACACGC
[T10.5] 3' (AA:–385)	GATC GAATTC TTA ATCTTTAAAATGTTCATCCTCACTTTTCT
[T10.4] 3' (AA:–395)	GATC GAATTC TTA AGGTCCCTTAGACATAAGAACC
[T10.3] 3' (AA:–40)	GATC GAATTC TTA AGCATCTAAAATAGCTTCCTTGAGT
[T10.2] 3' (AA:–415)	GATC GAATTC TTA GGAGAATAATCCTAGTATAGGATATGGC
[T10.1] 3' (AA:–425)	GATC GAATTC TTA GGCATCAATTTCATCAAAGAAATCTTTG

Supplementary Table 3-3. Primers used for qPCR analysis

Primer name	Sequence	Genome targeted	Gene target
5′ AtRpoTp	CTGAATGCAGGTCGAAACTCGGG	Nuclear	RNA polymerase
3′ AtRpoTp	GCTTGGAAGCCGTCTGCTAGAAC	Nuclear	
5′ nad9	GTGGGAGCGAGAAGTTTGGGATATG	Mitochondrial	NADH dehydrogenase
3′ nad9	GGGTCATCTCAATGGGTTCAGAAACC	Mitochondrial	
5′ orf25	TCAAAGTGACTCTCGACGGGAGC	Mitochondrial	B subunit of ATP synthase
3′ orf25	TGCCACAAATTCGCAAGCTGATCC	Mitochondrial	
5′ cox1	GAAGTAGGTAGCGGCACTGGG	Mitochondrial	Cytochrome c oxidase subunit 1
3′ cox1	ATTCCAGGTCCACGCATGTTGAAG	Mitochondrial	
5′ psbK	GTCGCCAAATTGCCAGAGGC	Plastid	PSII K protein
3′ psbK	CGGCTTGCCAAACAAAGGCTAAGAG	Plastid	
5′ petD	TATTACGGGGAACCCGCATGG	Plastid	Cytochrome b6-f complex subunit 4
3′ petD	GCAAAAGGATCCGCAGGTTCACC	Plastid	
5′ ndhH	GACTTCCAGGGGGTCCCTATGAG	Plastid	NAD(P)H-quinone oxidoreductase
3′ ndhH	CCCAACTCCCCTTTTGGAGCTTC	Plastid	

REFERENCES

1 Hooke, R. *Micrographia : or some physilogical descriptions of minute bodies made by magnifying glasses with observations and inquiries thereupon*. Facsim. edn, (Science Heritage, Ltd., 1987).

2 Harris, H. *The birth of the cell*. (Yale University Press, 1999).

3 Ernster, L. & Schatz, G. Mitochondria - a Historical Review. *J Cell Biol* **91**, S227-S255, doi:DOI 10.1083/jcb.91.3.227s (1981).

4 Schimper, A. F. W. Ueber die Entwickelung der Chlorophyllkörner und Farbkörper. (1883).

5 Kornberg, H. L. Krebs Citric-Acid Cycle - Half a Century and Still Turning - Introductory. *Biochem Soc Symp*, 1-2 (1987).

6 Ling, G. N. Oxidative-Phosphorylation and Mitochondrial Physiology - a Critical-Review of Chemiosmotic Theory, and Reinterpretation by the Association-Induction Hypothesis. *Physiol Chem Phys M* **13**, 29-96 (1981).

7 Calvin, M. The Path of Carbon in Photosynthesis. *Harvey Lect* **46**, 218-251 (1951).

8 Nass, M. M. K. & Nass, S. Intramitochondrial Fibers with DNA Characteristics .1. Fixation and Electron Staining Reactions. *J Cell Biol* **19**, 593-&, doi:DOI 10.1083/jcb.19.3.593 (1963).

9 Green, B. R. & Gordon, M. P. The satellite DNA's of some higher plants. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **145**, 378-390, doi:https://doi.org/10.1016/0005-2787(67)90056-1 (1967).

10 Martin, W. & Kowallik, K. Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren imPflanzenreiche'.

European Journal of Phycology **34**, 287-295, doi:10.1080/09670269910001736342 (1999).

11 Thrash, J. C. *et al.* Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci Rep-Uk* **1**, doi:ARTN 13 10.1038/srep00013 (2011).

12 Chu, K. H., Qi, J., Yu, Z. G. & Anh, V. Origin and phylogeny of chloroplasts revealed by a simple correlation analysis of complete genomes. *Mol Biol Evol* **21**, 200-206, doi:10.1093/molbev/msh002 (2004).

13 Emelyanov, V. V. Evolutionary relationship of Rickettsiae and mitochondria. *Febs Lett* **501**, 11-18, doi:Doi 10.1016/S0014-5793(01)02618-7 (2001).

14 Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J. & Andersson, S. G. E. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet* **23**, 511-520, doi:10.1016/j.tig.2007.08.002 (2007).

15 McFadden, G. I. & van Dooren, G. G. Evolution: Red algal genome affirms a common origin of all plastids. *Curr Biol* **14**, R514-R516, doi:10.1016/j.cub.2004.06.041 (2004).

16 Keeling, P. J. The endosymbiotic origin, diversification and fate of plastids. *Philos T R Soc B* **365**, 729-748, doi:10.1098/rstb.2009.0103 (2010).

17 Jackson, C., Clayden, S. & Reyes-Prieto, A. The Glaucophyta: the blue-green plants in a nutshell. *Acta Soc Bot Pol* **84**, 149-165, doi:10.5586/asbp.2015.020 (2015).

18 Gabrielson, P. W. & Garbary, D. Systematics of Red Algae (Rhodophyta). *Crit Rev Plant Sci* **3**, 325-366, doi:Doi 10.1080/07352688609382215 (1986).

19 Keeling, P. J. Diversity and evolutionary history of plastids and their hosts. *Am J Bot* **91**, 1481-1493, doi:DOI 10.3732/ajb.91.10.1481 (2004).

20 Wernegreen, J. J. Endosymbiont evolution: predictions from theory and surprises from genomes. *Year in Evolutionary Biology* **1360**, 16-35, doi:10.1111/nyas.12740 (2015).

21 Nilsson, A. I. *et al.* Bacterial genome size reduction by experimental evolution. *P Natl Acad Sci USA* **102**, 12112-12116, doi:10.1073/pnas.0503654102 (2005).

22 McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**, 13-26, doi:10.1038/nrmicro2670 (2012).

23 Selosse, M. A., Albert, B. R. & Godelle, B. Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends Ecol Evol* **16**, 135-141, doi:Doi 10.1016/S0169-5347(00)02084-X (2001).

24 Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res* **27**, 1767-1780, doi:DOI 10.1093/nar/27.8.1767 (1999).

25 Iborra, F. J., Kimura, H. & Cook, P. R. The functional organization of mitochondrial genomes in human cells. *Bmc Biol* **2**, doi:Artn 9 Doi 10.1186/1741-7007-2-9 (2004).

26 Kukat, C. *et al.* Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *P Natl Acad Sci USA* **108**, 13534-13539, doi:10.1073/pnas.1109263108 (2011).

27 Montier, L. L. C., Deng, J. J. & Bai, Y. D. Number matters: control of mammalian mitochondrial DNA copy number. *J Genet Genomics* **36**, 125-131, doi:10.1016/S1673-8527(08)60099-5 (2009).

28 Fauron, C., Allen, J., Clifton, S. & Newton, K. in *Molecular Biology and Biotechnology of Plant Organelles: Chloroplasts and Mitochondria* (eds Henry Daniell & Christine Chase) 151-177 (Springer Netherlands, 2004).

29 Oldenburg, D. J., Kumar, R. A. & Bendich, A. J. The amount and integrity of mtDNA in maize decline with development. *Planta* **237**, 603-617, doi:10.1007/s00425-012-1802-z (2013).

30 Preuten, T. *et al.* Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *Plant J* **64**, 948-959, doi:10.1111/j.1365-313X.2010.04389.x (2010).

31 Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* **17**, doi:ARTN 134 10.1186/s13059-016-1004-2 (2016).

32 Palmer, J. D. Comparative Organization of Chloroplast Genomes. *Annu Rev Genet* **19**, 325-354, doi:DOI 10.1146/annurev.ge.19.120185.001545 (1985).

33 Gray, M. W. The Bacterial Ancestry of Plastids and Mitochondria. *BioScience* **33**, 693-699, doi:10.2307/1309349 (1983).

34 Zoschke, R., Liere, K. & Borner, T. From seedling to mature plant: Arabidopsis plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. *Plant J* **50**, 710-722, doi:10.1111/j.1365-313X.2007.03084.x (2007).

35 Zheng, Q., Oldenburg, D. J. & Bendich, A. J. Independent effects of leaf growth and light on the development of the plastid and its DNA content in Zea species. *J Exp Bot* **62**, 2715-2730, doi:10.1093/jxb/erq441 (2011).

36 Udy, D. B., Belcher, S., Williams-Carrier, R., Gualberto, J. M. & Barkan, A. Effects of Reduced Chloroplast Gene Copy Number on Chloroplast Gene Expression in Maize. *Plant Physiol* **160**, 1420-1431, doi:10.1104/pp.112.204198 (2012).

37 Shaver, J. M., Oldenburg, D. J. & Bendich, A. J. Changes in chloroplast DNA during development in tobacco, Medicago truncatula, pea, and maize. *Planta* **224**, 72-82, doi:10.1007/s00425-005-0195-7 (2006).

38 Rowan, B. A., Oldenburg, D. J. & Bendich, A. J. The demise of chloroplast DNA in Arabidopsis. *Curr Genet* **46**, 176-181, doi:10.1007/s00294-004-0515-7 (2004).

39 Rowan, B. A., Oldenburg, D. J. & Bendich, A. J. A multiple-method approach reveals a declining amount of chloroplast DNA during development in Arabidopsis. *Bmc Plant Biol* **9**, doi:Artn 3 10.1186/1471-2229-9-3 (2009).

40 Okimoto, R., Macfarlane, J. L., Clary, D. O. & Wolstenholme, D. R. The mitochondrial genomes of two nematodes, Caenorhabditis elegans and Ascaris suum. *Genetics* **130**, 471-498 (1992).

41 Hoffmann, R. J., Boore, J. L. & Brown, W. M. A Novel Mitochondrial Genome Organization for the Blue Mussel, Mytilus-Edulis. *Genetics* **131**, 397-412 (1992).

42 Beagley, C. T. *et al.* Mitochondrial genomes of anthozoa (Cnidaria). *Prog Cell R* **5**, 149-153 (1995).

43 Palmer, J. D. *et al.* Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *P Natl Acad Sci USA* **97**, 6960-6966, doi:DOI 10.1073/pnas.97.13.6960 (2000).

44 Morley, S. A. & Nielsen, B. L. Plant mitochondrial DNA. *Front Biosci-Landmrk* **22**, 1023-1032, doi:10.2741/4531 (2017).

45 Burger, G., Gray, M. W. & Lang, B. F. Mitochondrial genomes: anything goes. *Trends Genet* **19**, 709-716, doi:10.1016/j.tig.2003.10.012 (2003).

46 Gualberto, J. M. *et al.* The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **100**, 107-120, doi:10.1016/j.biochi.2013.09.016 (2014).

47 Christensen, A. C. Plant Mitochondrial Genome Evolution Can Be Explained by DNA Repair Mechanisms. *Genome Biology and Evolution* **5**, 1079-1086, doi:10.1093/gbe/evt069 (2013).

48 Unseld, M., Marienfeld, J. R., Brandt, P. & Brennicke, A. The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nature genetics* **15**, 57-61, doi:DOI 10.1038/ng0197-57 (1997).

49 Kubo, T. *et al.* The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNA(Cys)(GCA). *Nucleic acids research* **28**, 2571-2576, doi:DOI 10.1093/nar/28.13.2571 (2000).

50 Chang, S. X. *et al.* Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in Brassica. *BMC genomics* **12**, doi:Artn 497 10.1186/1471-2164-12-497 (2011).

51 Chang, S. X. *et al.* The Mitochondrial Genome of Soybean Reveals Complex Genome Structures and Gene Evolution at Intercellular and Phylogenetic Levels. *PloS one* **8**, doi:ARTN e56502 10.1371/journal.pone.0056502 (2013).

52 Liu, G. Z. *et al.* The Complete Mitochondrial Genome of Gossypium hirsutum and Evolutionary Analysis of Higher Plant Mitochondrial Genomes. *PloS one* **8**, doi:ARTN e69476 10.1371/journal.pone.0069476 (2013).

53 Sugiyama, Y. *et al.* The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Molecular Genetics and Genomics* **272**, 603-615, doi:10.1007/s00438-004-1075-8 (2005).

54 Tian, X. J., Zheng, J., Hu, S. N. & Yu, J. The rice mitochondrial genomes and their variations. *Plant physiology* **140**, 401-410, doi:10.1104/pp.105.070060 (2006).

55 Ogihara, Y. *et al.* Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic acids research* **33**, 6235-6250, doi:10.1093/nar/gki925 (2005).

56 Clifton, S. W. *et al.* Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant physiology* **136**, 3486-3503, doi:10.1104/pp.104.044602 (2004).

57 Dietrich, F. S. *et al.* The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. *Science* **304**, 304-307, doi:10.1126/science.1095781 (2004).

58 Nowrousian, M. *et al.* De novo Assembly of a 40 Mb Eukaryotic Genome from Short Sequence Reads: Sordaria macrospora, a Model Organism for Fungal Morphogenesis. *PLoS genetics* **6**, doi:ARTN e1000891 10.1371/journal.pgen.1000891 (2010).

59 Nakao, Y. *et al.* Genome Sequence of the Lager Brewing Yeast, an Interspecies Hybrid. *DNA Res* **16**, 115-129, doi:10.1093/dnares/dsp003 (2009).

60 Lang, B. F. The Mitochondrial Genome of the Fission Yeast Schizosaccharomyces-Pombe - Highly Homologous Introns Are Inserted at the Same Position of the Otherwise Less Conserved Coxl Genes in Schizosaccharomyces-Pombe and Aspergillus-Nidulans. *Embo Journal* **3**, 2129-2136 (1984).

61 Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B. & Michaelis, G. Mitochondrial-DNA of Chlamydomonas-Reinhardtii - the Structure of the Ends of the Linear 15.8-Kb Genome Suggests Mechanisms for DNA-Replication. *Curr Genet* **24**, 241-247, doi:Doi 10.1007/Bf00351798 (1993).

62 Ogawa, S. *et al.* The mitochondrial DNA of Dictyostelium discoideum: complete sequence, gene content and genome organization. *Molecular and General Genetics* **263**, 514-519, doi:Doi 10.1007/Pl00008685 (2000).

63 Anderson, S. *et al.* Sequence and Organization of the Human Mitochondrial Genome. *Nature* **290**, 457-465, doi:DOI 10.1038/290457a0 (1981).

64 Bayona-Bafaluy, M. P. *et al.* Revisiting the mouse mitochondrial DNA sequence. *Nucleic acids research* **31**, 5349-5355, doi:10.1093/nar/gkg739 (2003).

65 Clayton, D. A. Replication of Animal Mitochondrial-DNA. *Cell* **28**, 693-705, doi:Doi 10.1016/0092-8674(82)90049-6 (1982).

66 Lavrov, D. V. & Pett, W. Animal Mitochondrial DNA as We Do Not Know It: mt-Genome Organization and Evolution in Nonbilaterian Lineages. *Genome Biol Evol* **8**, 2896-2913, doi:10.1093/gbe/evw195 (2016).

67 Shadel, G. S. & Clayton, D. A. Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem* **66**, 409-435, doi:DOI 10.1146/annurev.biochem.66.1.409 (1997).

68 Oda, K., Kohchi, T. & Ohyama, K. Mitochondrial-DNA of Marchantia-Polymorpha as a Single Circular Form with No Incorporation of Foreign DNA. *Biosci Biotech Bioch* **56**, 132-135, doi:DOI 10.1271/bbb.56.132 (1992).

69 Backert, S., Nielsen, B. L. & Borner, T. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci* **2**, 477-483, doi:Doi 10.1016/S1360-1385(97)01148-5 (1997).

70 Sloan, D. B. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytol* **200**, 978-985, doi:10.1111/nph.12395 (2013).

71 Cupp, J. D. & Nielsen, B. L. Minireview: DNA replication in plant mitochondria. *Mitochondrion* **19**, 231-237, doi:10.1016/j.mito.2014.03.008 (2014).

72 Schuster, W. & Brennicke, A. The Plant Mitochondrial Genome - Physical Structure, Information-Content, Rna Editing, and Gene Migration to the Nucleus. *Annu Rev Plant Phys* **45**, 61-78 (1994).

73 Palmer, J. D. & Herbon, L. A. Plant Mitochondrial-DNA Evolves Rapidly in Structure, but Slowly in Sequence. *J Mol Evol* **28**, 87-97, doi:Doi 10.1007/Bf02143500 (1988).

74 Parsons, T. J. *et al.* A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* **15**, 363-368, doi:DOI 10.1038/ng0497-363 (1997).

75 Sugiura, M. The chloroplast genome. *Essays Biochem* **30**, 49-57 (1995).

76 Palmer, J. D. & Stein, D. B. Conservation of Chloroplast Genome Structure among Vascular Plants. *Curr Genet* **10**, 823-833, doi:Doi 10.1007/Bf00418529 (1986).

77 Kolodner, R. & Tewari, K. K. Inverted Repeats in Chloroplast DNA from Higher-Plants. *P Natl Acad Sci USA* **76**, 41-45, doi:DOI 10.1073/pnas.76.1.41 (1979).

78 Dalmon, J., Loiseaux, S. & Bazetoux, S. Heterogeneity of plastid DNA of two species of brown algae. *Plant Science Letters* **29**, 243-253, doi:https://doi.org/10.1016/0304-4211(83)90149-9 (1983).

79 Palmer, J. D. & Thompson, W. F. Chloroplast DNA Rearrangements Are More Frequent When a Large Inverted Repeat Sequence Is Lost. *Cell* **29**, 537-550, doi:Doi 10.1016/0092-8674(82)90170-2 (1982).

80 Zhu, A. D., Guo, W. H., Gupta, S., Fan, W. S. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and

loss on substitution rates. *New Phytol* **209**, 1747-1756, doi:10.1111/nph.13743 (2016).

81 Ciesielski, G. L., Oliveira, M. T. & Kaguni, L. S. Animal Mitochondrial DNA Replication. *Enzymes* **39**, 255-292, doi:10.1016/bs.enz.2016.03.006 (2016).

82 Lewis, S. C. *et al.* A Rolling Circle Replication Mechanism Produces Multimeric Lariats of Mitochondrial DNA in Caenorhabditis elegans. *Plos Genet* **11**, doi:ARTN e1004985 10.1371/journal.pgen.1004985 (2015).

83 Fish, J., Raule, N. & Attardi, G. Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis. *Science* **306**, 2098-2101, doi:10.1126/science.1102077 (2004).

84 Holt, I. J., Lorimer, H. E. & Jacobs, H. T. Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* **100**, 515-524, doi:Doi 10.1016/S0092-8674(00)80688-1 (2000).

85 Yang, M. Y. *et al.* Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* **111**, 495-505, doi:Doi 10.1016/S0092-8674(02)01075-9 (2002).

86 Korhonen, J. A., Pham, X. H., Pellegrini, M. & Falkenberg, M. Reconstitution of a minimal mtDNA replisome in vitro. *Embo J* **23**, 2423-2429, doi:10.1038/sj.emboj.7600257 (2004).

87 Dehaas, J. M. *et al.* 2 Potential Petunia-Hybrida Mitochondrial-DNA Replication Origins Show Structural and Invitro Functional Homology with the Animal Mitochondrial-DNA Heavy and Light Strand Replication Origins. *Curr Genet* **20**, 503-513, doi:Doi 10.1007/Bf00334779 (1991).

88 Backert, S., Dorfel, P., Lurz, R. & Borner, T. Rolling-circle replication of mitochondrial DNA in the higher plant Chenopodium album (L). *Mol Cell Biol* **16**, 6285-6294 (1996).

89 Backert, S. & Borner, T. Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant Chenopodium album (L.). *Curr Genet* **37**, 304-314, doi:DOI 10.1007/s002940050532 (2000).

90 Heinhorst, S. & Cannon, G. C. DNA-Replication in Chloroplasts. *J Cell Sci* **104**, 1-9 (1993).

91 Chiu, W. L. & Sears, B. B. Electron-Microscopic Localization of Replication Origins in Oenothera Chloroplast DNA. *Mol Gen Genet* **232**, 33-39, doi:Doi 10.1007/Bf00299134 (1992).

92 Waddell, J., Wang, X. M. & Wu, M. Electron-Microscopic Localization of the Chloroplast DNA Replicative Origins in Chlamydomonas-Reinhardii. *Nucleic Acids Res* **12**, 3843-3856, doi:DOI 10.1093/nar/12.9.3843 (1984).

93 Ravelchapuis, P., Heizmann, P. & Nigon, V. Electron-Microscopic Localization of the Replication Origin of Euglena-Gracilis Chloroplast DNA. *Nature* **300**, 78-81, doi:DOI 10.1038/300078a0 (1982).

94 Lee, S. J. & Richardson, C. C. Choreography of bacteriophage T7 DNA replication. *Curr Opin Chem Biol* **15**, 580-586, doi:10.1016/j.cbpa.2011.07.024 (2011).

95 Korhonen, J. A., Pham, X. H., Pellegrini, M. & Falkenberg, M. Reconstitution of a minimal mtDNA replisome in vitro. *Embo J* **23**, 2423-2429, doi:10.1038/sj.emboj.7600257 (2004).

96 Elo, A., Lyznik, A., Gonzalez, D. O., Kachman, S. D. & Mackenzie, S. A. Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are

clustered in the Arabidopsis genome. *Plant Cell* **15**, 1619-1631, doi:10.1105/tpc.010009 (2003).

97 Ono, Y. *et al.* NtPolI-like1 and NtPolI-like2, bacterial DNA polymerase I homologs isolated from BY-2 cultured tobacco cells, encode DNA polymerases engaged in DNA replication in both plastids and mitochondria. *Plant Cell Physiol* **48**, 1679-1692, doi:10.1093/pcp/pcm140 (2007).

98 Carrie, C. *et al.* Approaches to defining dual-targeted proteins in Arabidopsis. *Plant J* **57**, 1128-1139, doi:10.1111/j.1365-313X.2008.03745.x (2009).

99 Diray-Arce, J., Liu, B., Cupp, J. D., Hunt, T. & Nielsen, B. L. The Arabidopsis At1g30680 gene encodes a homologue to the phage T7 gp4 protein that has both DNA primase and DNA helicase activities. *Bmc Plant Biol* **13**, doi:Artn 36 10.1186/1471-2229-13-36 (2013).

100 Shutt, T. E. & Gray, M. W. Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes. *J Mol Evol* **62**, 588-599, doi:10.1007/s00239-005-0162-8 (2006).

101 Duxin, J. P. *et al.* Human Dna2 Is a Nuclear and Mitochondrial DNA Maintenance Protein. *Mol Cell Biol* **29**, 4274-4282, doi:10.1128/Mcb.01834-08 (2009).

102 Zheng, L. *et al.* Human DNA2 Is a Mitochondrial Nuclease/Helicase for Efficient Processing of DNA Replication and Repair Intermediates. *Mol Cell* **32**, 325-336, doi:10.1016/j.molcel.2008.09.024 (2008).

103 Liere, K., Weihe, A. & Borner, T. The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J Plant Physiol* **168**, 1345-1360, doi:10.1016/j.jplph.2011.01.005 (2011).

104 Carrie, C. & Small, I. A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Bba-Mol Cell Res* **1833**, 253-259, doi:10.1016/j.bbamcr.2012.05.029 (2013).

105 Hedtke, B., Borner, T. & Weihe, A. One RNA polymerase serving two genomes. *Embo Reports* **1**, 435-440, doi:10.1093/embo-reports/kvd086 (2000).

106 Hedtke, B., Borner, T. & Weihe, A. Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science* **277**, 809-811, doi:DOI 10.1126/science.277.5327.809 (1997).

107 Yang, Z. *et al.* RNase H1 Cooperates with DNA Gyrases to Restrict R-Loops and Maintain Genome Integrity in Arabidopsis Chloroplasts. *Plant Cell* **29**, 2478-2497, doi:10.1105/tpc.17.00305 (2017).

108 Sato, N., Terasawa, K., Miyajima, K. & Kabeya, Y. Organization, developmental dynamics, and evolution of plastid nucleoids. *Int Rev Cytol* **232**, 217-262, doi:Doi 10.1016/S0074-7696(03)32006-6 (2003).

109 Edmondson, A. C. *et al.* Characterization of a mitochondrially targeted single-stranded DNA-binding protein in Arabidopsis thaliana. *Mol Genet Genomics* **273**, 115-122, doi:10.1007/s00438-004-1106-5 (2005).

110 Zaegel, V. *et al.* The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in Arabidopsis. *Plant Cell* **18**, 3548-3563, doi:10.1105/tpc.106.042028 (2006).

111 Krause, K. *et al.* DNA-binding proteins of the Whirly family in Arabidopsis thaliana are targeted to the organelles. *Febs Letters* **579**, 3707-3712, doi:10.1016/j.febslet.2005.05.059 (2005).

112 Cappadocia, L. *et al.* Crystal Structures of DNA-Whirly Complexes and Their Role in Arabidopsis Organelle Genome Repair. *Plant Cell* **22**, 1849-1867, doi:10.1105/tpc.109.071399 (2010).

113 Marechal, A. *et al.* Whirly proteins maintain plastid genome stability in Arabidopsis. *P Natl Acad Sci USA* **106**, 14693-14698, doi:10.1073/pnas.0901710106 (2009).

114 Shedge, V., Arrieta-Montiel, M., Christensen, A. C. & Mackenzie, S. A. Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell* **19**, 1251-1264, doi:10.1105/tpc.106.048355 (2007).

115 Khazi, F. R., Edmondson, A. C. & Nielsen, B. L. An Arabidopsis homologue of bacterial RecA that complements an E-coli recA deletion is targeted to plant mitochondria. *Mol Genet Genomics* **269**, 454-463, doi:10.1007/s00438-003-0859-6 (2003).

116 Xu, Y. Z. *et al.* MutS HOMOLOG1 Is a Nucleoid Protein That Alters Mitochondrial and Plastid Properties and Plant Response to High Light. *Plant Cell* **23**, 3428-3441, doi:10.1105/tpc.111.089136 (2011).

117 Wall, M. K., Mitchenall, L. A. & Maxwell, A. Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria. *P Natl Acad Sci USA* **101**, 7821-7826, doi:10.1073/pnas.0400836101 (2004).

118 Roy, S., Choudhury, S. R., Singh, S. K. & Das, K. P. AtPol lambda, A Homolog of Mammalian DNA Polymerase lambda in Arabidopsis thaliana, is Involved in the Repair of UV-B Induced DNA Damage Through the Dark Repair Pathway. *Plant Cell Physiol* **52**, 448-467, doi:10.1093/pcp/pcr002 (2011).

119 Parent, J. S., Lepage, E. & Brisson, N. Divergent Roles for the Two PolI-Like Organelle DNA Polymerases of Arabidopsis. *Plant Physiol* **156**, 254-262, doi:10.1104/pp.111.173849 (2011).

120 Moriyama, T., Terasawa, K. & Sato, N. Conservation of POPs, the Plant Organellar DNA Polymerases, in Eukaryotes. *Protist* **162**, 177-187, doi:10.1016/j.protis.2010.06.001 (2011).

121 Morley, S. A. & Nielsen, B. L. Chloroplast DNA Copy Number Changes during Plant Development in Organelle DNA Polymerase Mutants. *Front Plant Sci* **7**, doi:ARTN 57 10.3389/fpls.2016.00057 (2016).

122 Trasvina-Arenas, C. H. *et al.* Identification of a unique insertion in plant organellar DNA polymerases responsible for 5 '-dRP lyase and strand-displacement activities: Implications for Base Excision Repair. *DNA Repair* **65**, 1-10, doi:10.1016/j.dnarep.2018.02.010 (2018).

123 Ayala-Garcia, V. M., Baruch-Torres, N., Garcia-Medel, P. & Brieba, L. G. Plant organellar DNA polymerases paralogs exhibit dissimilar nucleotide incorporation fidelity. *Febs J*, doi:10.1111/febs.14645 (2018).

124 Moriyama, T. & Sato, N. Enzymes involved in organellar DNA replication in photosynthetic eukaryotes. *Front Plant Sci* **5**, doi:ARTN 480 10.3389/fpls.2014.00480 (2014).

125 Eun, H.-M. in *Enzymology Primer for Recombinant DNA Technology* (ed Hyone-Myong Eun) 345-489 (Academic Press, 1996).

126 Bedford, E., Tabor, S. & Richardson, C. C. The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on Escherichia coli DNA polymerase I. *P Natl Acad Sci USA* **94**, 479-484, doi:DOI 10.1073/pnas.94.2.479 (1997).

127 McCulloch, S. D. & Kunkel, T. A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**, 148-161, doi:10.1038/cr.2008.4 (2008).

128 Baruch-Torres, N. & Brieba, L. G. Plant organellar DNA polymerases are replicative and translesion DNA synthesis polymerases. *Nucleic Acids Res* **45**, 10751-10763, doi:10.1093/nar/gkx744 (2017).

129 Sunderland, P. A., West, C. E., Waterworth, W. M. & Bray, C. M. An evolutionarily conserved translation initiation mechanism regulates nuclear or mitochondrial targeting of DNA ligase 1 in Arabidopsis thaliana. *Plant J* **47**, 356-367, doi:10.1111/j.1365-313X.2006.02791.x (2006).

130 Waterworth, W. M. *et al.* DNA ligase 1 deficient plants display severe growth defects and delayed repair of both DNA single and double strand breaks. *Bmc Plant Biol* **9**, doi:Artn 79 10.1186/1471-2229-9-79 (2009).

131 Gao, Y. K. *et al.* DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature* **471**, 240-U134, doi:10.1038/nature09773 (2011).

132 Matsumoto, T. *et al.* Comprehensive Sequence Analysis of 24,783 Barley Full-Length cDNAs Derived from 12 Clone Libraries. *Plant Physiol* **156**, 20-28, doi:10.1104/pp.110.171579 (2011).

133 Spelbrink, J. N. *et al.* Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria (vol 28, pg 223, 2001). *Nat Genet* **29**, 100-100 (2001).

134 Gao, Y. *et al.* Structures and operating principles of the replisome. *Science* **363**, doi:10.1126/science.aav7003 (2019).

135 Korhonen, J. A., Gaspari, M. & Falkenberg, M. TWINKLE has 5 '-> 3 ' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *J Biol Chem* **278**, 48627-48632, doi:10.1074/jbc.M306981200 (2003).

136 Peralta-Castro, A., Baruch-Torres, N. & Brieba, L. G. Plant organellar DNA primase-helicase synthesizes RNA primers for organellar DNA polymerases using a unique recognition sequence. *Nucleic Acids Res* **45**, 10764-10774, doi:10.1093/nar/gkx745 (2017).

137 Jia, N., Liu, X. M. & Gao, H. B. A DNA2 Homolog Is Required for DNA Damage Repair, Cell Cycle Regulation, and Meristem Maintenance in Plants. *Plant Physiol* **171**, 318-333, doi:10.1104/pp.16.00312 (2016).

138 Cho, H. S. *et al.* DNA gyrase is involved in chloroplast nucleoid partitioning. *Plant Cell* **16**, 2665-2682, doi:DOI 10.1105/tpc.104.024281 (2004).

139 Smith, D. R. Updating Our View of Organelle Genome Nucleotide Landscape. *Frontiers in Genetics* **3**, 175, doi:10.3389/fgene.2012.00175 (2012).

140 Larson, M. A. *et al.* Hyperthermophilic Aquifex aeolicus initiates primer synthesis on a limited set of trinucleotides comprised of cytosines and guanines. *Nucleic Acids Res* **36**, 5260-5269, doi:10.1093/nar/gkn461 (2008).

141 Arnold, J. J., Smidansky, E. D., Moustafa, I. M. & Cameron, C. E. Human mitochondrial RNA polymerase: Structure-function, mechanism and inhibition. *Bba-Gene Regul Mech* **1819**, 948-960, doi:10.1016/j.bbagrm.2012.04.002 (2012).

142 Hess, W. R. & Borner, T. Organellar RNA polymerases of higher plants. *International Review of Cytology - a Survey of Cell Biology, Vol 190* **190**, 1-59, doi:Doi 10.1016/S0074-7696(08)62145-2 (1999).

143 Yin, C., Richter, U., Borner, T. & Weihe, A. Evolution of plant phage-type RNA polymerases: the genome of the basal angiosperm Nuphar advena encodes two mitochondrial and one plastid phage-type RNA polymerases. *Bmc Evol Biol* **10**, doi:Artn 379 10.1186/1471-2148-10-379 (2010).

144 Baba, K. *et al.* Organellar gene transcription and early seedling development are affected in the rpoT;2 mutant of Arabidopsis. *Plant J* **38**, 38-48, doi:10.1111/j.1365-313X.2004.02022.x (2004).

145 Serino, G. & Maliga, P. RNA polymerase subunits encoded by the plastid rpo genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiol* **117**, 1165-1170, doi:DOI 10.1104/pp.117.4.1165 (1998).

146 Nielsen, B. L., Rajasekhar, V. K. & Tewari, K. K. Pea Chloroplast DNA Primase - Characterization and Role in Initiation of Replication. *Plant Mol Biol* **16**, 1019-1034, doi:Doi 10.1007/Bf00016074 (1991).

147 Fuste, J. M. *et al.* Mitochondrial RNA Polymerase Is Needed for Activation of the Origin of Light-Strand DNA Replication. *Mol Cell* **37**, 67-78, doi:10.1016/j.molcel.2009.12.021 (2010).

148 Ramachandran, A. *et al.* The Yeast Mitochondrial RNA Polymerase and Transcription Factor Complex Catalyzes Efficient Priming of DNA Synthesis on Single-stranded DNA. *J Biol Chem* **291**, 16828-16839, doi:10.1074/jbc.M116.740282 (2016).

149 Desveaux, D., Marechal, A. & Brisson, N. Whirly transcription factors: defense gene regulation and beyond. *Trends Plant Sci* **10**, 95-102, doi:10.1016/jtplants.2004.12.008 (2005).

150 Marechal, A. *et al.* Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *Bmc Plant Biol* **8**, doi:Artn 42 10.1186/1471-2229-8-42 (2008).

151 Janicka, S. *et al.* A RAD52-like single-stranded DNA binding protein affects mitochondrial DNA repair by recombination. *Plant J* **72**, 423-435, doi:10.1111/j.1365-313X.2012.05097.x (2012).

152 Redei, G. P. Extrachromosomal Mutability Determined by a Nuclear Gene Locus in Arabidopsis. *Mutat Res* **18**, 149-162, doi:Doi 10.1016/0027-5107(73)90031-6 (1973).

153 Martinezzapater, J. M., Gil, P., Capel, J. & Somerville, C. R. Mutations at the Arabidopsis Chm Locus Promote Rearrangements of the Mitochondrial Genome. *Plant Cell* **4**, 889-899 (1992).

154 Abdelnoor, R. V. *et al.* Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *P Natl Acad Sci USA* **100**, 5968-5973, doi:10.1073/pnas.1037651100 (2003).

155 Reenan, R. A. G. & Kolodner, R. D. Characterization of Insertion Mutations in the Saccharomyces-Cerevisiae Msh1 and Msh2 Genes - Evidence for Separate Mitochondrial and Nuclear Functions. *Genetics* **132**, 975-985 (1992).

156 Abdelnoor, R. V. *et al.* Mitochondrial genome dynamics in plants and animals: Convergent gene fusions of a MutS homologue. *J Mol Evol* **63**, 165-173, doi:10.1007/s00239-005-0226-9 (2006).

157 Christensen, A. C. *et al.* Dual-domain, dual-targeting organellar protein presequences in Arabidopsis can use non-AUG start codons. *Plant Cell* **17**, 2805-2816, doi:10.1105/tpc.105.035287 (2005).

158 Allen, J. F. Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *P Natl Acad Sci USA* **112**, 10231-10238, doi:10.1073/pnas.1500012112 (2015).

159 Shaver, J. M., Oldenburg, D. J. & Bendich, A. J. The structure of chloroplast DNA molecules and the effects of light on the amount of chloroplast DNA during development in Medicago truncatula. *Plant Physiol* **146**, 1064-1074, doi:10.1104/pp.107.112946 (2008).

160 Kabeya, Y. & Miyagishima, S. Chloroplast DNA Replication Is Regulated by the Redox State Independently of Chloroplast Division in Chlamydomonas reinhardtii. *Plant Physiol* **161**, 2102-2112, doi:10.1104/pp.113.216291 (2013).

161 Hori, A., Yoshida, M., Shibata, T. & Ling, F. Reactive oxygen species regulate DNA copy number in isolated yeast mitochondria by triggering recombination-mediated replication. *Nucleic Acids Res* **37**, 749-761, doi:10.1093/nar/gkn993 (2009).

162 Kato, T., Kaneko, T., Sato, S., Nakamura, Y. & Tabata, S. Complete structure of the chloroplast genome of a legume, Lotus japonicus. *DNA Res* **7**, 323-330, doi:DOI 10.1093/dnares/7.6.323 (2000).

163 Kolodner Rd Fau - Tewari, K. K. & Tewari, K. K. Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism.

164 Kunnimalaiyaan, M. & Nielsen, B. L. Chloroplast DNA replication: Mechanism, enzymes and replication origins. *J Plant Biochem Biot* **6**, 1-7 (1997).

165 Oldenburg, D. J. & Bendich, A. J. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *J Mol Biol* **335**, 953-970, doi:10.1016/j.jmb.2003.11.020 (2004).

166 Rowan, B. A., Oldenburg, D. J. & Bendich, A. J. RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. *J Exp Bot* **61**, 2575-2588, doi:10.1093/jxb/erq088 (2010).

167 Nielsen, B. L., Cupp, J. D. & Brammer, J. Mechanisms for maintenance, replication, and repair of the chloroplast genome in plants. *J Exp Bot* **61**, 2535-2537, doi:10.1093/jxb/erq163 (2010).

168 Kreuzer, K. N. & Brister, J. R. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virol J* **7**, doi:Artn 358 10.1186/1743-422x-7-358 (2010).

169 Mori, Y. *et al.* Plastid DNA polymerases from higher plants, Arabidopsis thaliana. *Biochem Bioph Res Co* **334**, 43-50, doi:10.1016/j.bbrc.2005.06.052 (2005).

170 Cupp, J. D. & Nielsen, B. L. Arabidopsis thaliana organellar DNA polymerase IB mutants exhibit reduced mtDNA levels with a decrease in mitochondrial area density.

171 Kimura, S. *et al.* A novel DNA polymerase homologous to Escherichia coli DNA polymerase I from a higher plant, rice (Oryza sativa L.). *Nucleic Acids Res* **30**, 1585-1592, doi:DOI 10.1093/nar/30.7.1585 (2002).

172 Minas, K., McEwan, N. R., Newbold, C. J. & Scott, K. P. Optimization of a high-throughput CTAB-based protocol for the extraction of qPCR-grade DNA from rumen fluid, plant and bacterial pure cultures. *Fems Microbiol Lett* **325**, 162-169, doi:10.1111/j.1574-6968.2011.02424.x (2011).

173 Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C-T method. *Nat Protoc* **3**, 1101-1108, doi:10.1038/nprot.2008.73 (2008).

174 Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y. & Nambara, E. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J* **41**, 697-709, doi:10.1111/j.1365-313X.2005.02337.x (2005).

175 Schmid, M. *et al.* A gene expression map of Arabidopsis thaliana development. *Nat Genet* **37**, 501-506, doi:10.1038/ng1543 (2005).

176 Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953-971, doi:10.1038/nprot.2007.131 (2007).

177 Schein, A. I., Kissinger, J. C. & Ungar, L. H. Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic Acids Res* **29**, doi:ARTN e82 DOI 10.1093/nar/29.16.e82 (2001).

178 Emanuelsson, O., Nielsen, H. & Von Heijne, G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* **8**, 978-984, doi:DOI 10.1110/ps.8.5.978 (1999).

179 Muhlbauer, S. K., Lossl, A., Tzekova, L., Zou, Z. R. & Koop, H. U. Functional analysis of plastid DNA replication origins in tobacco by targeted inactivation. *Plant J* **32**, 175-184 (2002).

180 Scharff, L. B. & Koop, H. U. Linear molecules of tobacco ptDNA end at known replication origins and additional loci. *Plant Mol Biol* **62**, 611-621, doi:10.1007/s11103-006-9042-x (2006).

181 Liere, K. & Börner, T. in *Plastid Development in Leaves during Growth and Senescence* (eds Basanti Biswal, Karin Krupinska, & Udaya C. Biswal) 215-237 (Springer Netherlands, 2013).

182 Boffey, S. A. & Leech, R. M. Chloroplast DNA Levels and the Control of Chloroplast Division in Light-Grown Wheat Leaves. *Plant Physiol* **69**, 1387-1391, doi:DOI 10.1104/pp.69.6.1387 (1982).

183 Tymms, M. J., Scott, N. S. & Possingham, J. V. DNA Content of Beta-Vulgaris Chloroplasts during Leaf Cell Expansion. *Plant Physiol* **71**, 785-788, doi:DOI 10.1104/pp.71.4.785 (1983).

184 Grevich, J. J. & Daniell, H. Chloroplast genetic engineering: Recent advances and future perspectives. *Crit Rev Plant Sci* **24**, 83-107, doi:10.1080/07352680590935387 (2005).

185 Maliga, P. & Bock, R. Plastid Biotechnology: Food, Fuel, and Medicine for the 21st Century. *Plant Physiol* **155**, 1501-1510, doi:10.1104/pp.110.170969 (2011).

186 Archibald, J. M. Endosymbiosis and Eukaryotic Cell Evolution. *Curr Biol* **25**, R911-R921, doi:10.1016/j.cub.2015.07.055 (2015).

187 Hadariova, L., Vesteg, M., Hampl, V. & Krajcovic, J. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr Genet* **64**, 365-387, doi:10.1007/s00294-017-0761-0 (2018).

188 Oldenburg, D. J. & Bendich, A. J. DNA maintenance in plastids and mitochondria of plants. *Front Plant Sci* **6**, doi:ARTN 883 10.3389/fpls.2015.00883 (2015).

189 Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struc Biol* **11**, 39-46, doi:Doi 10.1016/S0959-440x(00)00167-6 (2001).

190 Kulczyk, A. W. & Richardson, C. C. Molecular interactions in the priming complex of bacteriophage T7. *P Natl Acad Sci USA* **109**, 9408-9413, doi:10.1073/pnas.1207033109 (2012).

191 Ayala-Garcia, V. M., Baruch-Torres, N., Garcia-Medel, P. L. & Brieba, L. G. Plant organellar DNA polymerases paralogs exhibit dissimilar nucleotide incorporation fidelity. *FEBS J* **285**, 4005-4018, doi:10.1111/febs.14645 (2018).

192 Renaux, A. & Consortium, U. UniProt: the universal protein knowledgebase (vol 45, pg D158, 2017). *Nucleic Acids Res* **46**, 2699-2699, doi:10.1093/nar/gky092 (2018).

193 Bernstein, J. A. & Richardson, C. C. A 7-kDa region of the bacteriophage T7 gene 4 protein is required for primase but not for helicase activity. *Proc Natl Acad Sci U S A* **85**, 396-400 (1988).

194 Kusakabe, T. & Richardson, C. C. The role of the zinc motif in sequence recognition by DNA primases. *J Biol Chem* **271**, 19563-19570, doi:DOI 10.1074/jbc.271.32.19563 (1996).

195 Nielsen, H. in *Protein Function Prediction: Methods and Protocols* (ed Daisuke Kihara) 59-73 (Springer New York, 2017).

196 Ramirez-Sanchez, O., Perez-Rodriguez, P., Delaye, L. & Tiessen, A. Plant Proteins Are Smaller Because They Are Encoded by Fewer Exons than Animal Proteins. *Genom Proteom Bioinf* **14**, 357-370, doi:10.1016/j.gpb.2016.06.003 (2016).

197 Yang, J. Y. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2015).

198 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738, doi:10.1038/nprot.2010.5 (2010).

199 Zhang, Y. I-TASSER server for protein 3D structure prediction. *Bmc Bioinformatics* **9**, doi:Artn 40 10.1186/1471-2105-9-40 (2008).

200 Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *P Natl Acad Sci USA* **108**, E1293-E1301, doi:10.1073/pnas.1111471108 (2011).

201 Hamdan, S. M. *et al.* Dynamic DNA helicase-DNA polymerase interactions assure processive replication fork movement. *Mol Cell* **27**, 539-549, doi:10.1016/j.molcel.2007.06.020 (2007).

202 Zhang, H. D. *et al.* Helicase-DNA polymerase interaction is critical to initiate leading-strand DNA synthesis. *P Natl Acad Sci USA* **108**, 9372-9377, doi:10.1073/pnas.1106678108 (2011).

203 Kulczyk, A. W. *et al.* An Interaction between DNA Polymerase and Helicase Is Essential for the High Processivity of the Bacteriophage T7 Replisome. *J Biol Chem* **287**, 39050-39060, doi:10.1074/jbc.M112.410647 (2012).

204 Zhang, H. D. *et al.* Binding Affinities among DNA Helicase-Primase, DNA Polymerase, and Replication Intermediates in the Replisome of Bacteriophage T7. *J Biol Chem* **291**, 1472-1480, doi:10.1074/jbc.M115.698233 (2016).

205 Rannou, O. *et al.* Functional interplay of DnaE polymerase, DnaG primase and DnaC helicase within a ternary complex, and primase to polymerase hand-off during lagging strand DNA replication in Bacillus subtilis. *Nucleic Acids Res* **41**, 5303-5320, doi:10.1093/nar/gkt207 (2013).

206 Braun, K. A., Lao, Y., He, Z. G., Ingles, C. J. & Wold, M. S. Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase a by multiple mechanisms. *Biochemistry-Us* **36**, 8443-8454, doi:DOI 10.1021/bi970473r (1997).

207 De, A. & Campbell, C. A novel interaction between DNA ligase III and DNA polymerase gamma plays an essential role in mitochondrial DNA stability. *Biochem J* **402**, 175-186, doi:10.1042/Bj20061004 (2007).

208 Benkovic, S. J., Valentine, A. M. & Salinas, F. Replisome-mediated DNA replication. *Annu Rev Biochem* **70**, 181-208, doi:DOI 10.1146/annurev.biochem.70.1.181 (2001).

209 Yao, N. & O'Donnell, M. Bacterial and Eukaryotic Replisome Machines. *JSM biochemistry and molecular biology* **3**, 1013 (2016).

210 Theologis, A. *et al.* Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. *Nature* **408**, 816-820, doi:Doi 10.1038/35048500 (2000).

211 Salanoubat, M. *et al.* Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana. *Nature* **408**, 820-822 (2000).

212 Green, P. J. The Ribonucleases of Higher-Plants. *Annu Rev Plant Phys* **45**, 421-445, doi:DOI 10.1146/annurev.pp.45.060194.002225 (1994).

213 Kaguni, L. S. & Oliveira, M. T. Structure, function and evolution of the animal mitochondrial replicative DNA helicase. *Crit Rev Biochem Mol* **51**, 53-64, doi:10.3109/10409238.2015.1117056 (2016).

214 Cupp, J. D. & Nielsen, B. L. Arabidopsis thaliana organellar DNA polymerase IB mutants exhibit reduced mtDNA levels with a decrease in mitochondrial area density. *Physiol Plantarum* **149**, 91-103, doi:10.1111/ppl.12009 (2013).

215 Marechal, A. & Brisson, N. Recombination and the maintenance of plant organelle genome stability. *New Phytol* **186**, 299-317, doi:10.1111/j.1469-8137.2010.03195.x (2010).

216 Miller-Messmer, M. *et al.* RecA-Dependent DNA Repair Results in Increased Heteroplasmy of the Arabidopsis Mitochondrial Genome. *Plant Physiol* **159**, 211-226, doi:10.1104/pp.112.194720 (2012).

217 Coordinators, N. R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **46**, D8-D13, doi:10.1093/nar/gkx1095 (2018).

218 Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135-145, doi:10.1002/pro.3290 (2018).

219 Pais, F. S., Ruy, P. C., Oliveira, G. & Coimbra, R. S. Assessing the efficiency of multiple sequence alignment programs. *Algorithms Mol Biol* **9**, 4, doi:10.1186/1748-7188-9-4 (2014).

220 Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A* **108**, E1293-1301, doi:10.1073/pnas.1111471108 (2011).

221 Goulas, T. *et al.* The pCri System: a vector collection for recombinant protein expression and purification. *Plos One* **9**, e112643, doi:10.1371/journal.pone.0112643 (2014).

222 Konrad, E. B., Modrich, P. & Lehman, I. R. Genetic and Enzymatic Characterization of a Conditional Lethal Mutant of Escherichia-Coli-K12 with a Temperature-Sensitive DNA Ligase. *J Mol Biol* **77**, 519-529, doi:Doi 10.1016/0022-2836(73)90220-9 (1973).

223 Ghosh, S., Hamdan, S. M., Cook, T. E. & Richardson, C. C. Interactions of Escherichia coli Thioredoxin, the Processivity Factor, with Bacteriophage T7 DNA Polymerase and Helicase. *J Biol Chem* **283**, 32077-32084, doi:10.1074/jbc.M805062200 (2008).

224 Huber, H. E., Tabor, S. & Richardson, C. C. Escherichia-Coli Thioredoxin Stabilizes Complexes of Bacteriophage-T7 DNA-Polymerase and Primed Templates. *J Biol Chem* **262**, 16224-16232 (1987).

225 Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A. & Allers, T. Accelerated growth in the absence of DNA replication origins. *Nature* **503**, 544-+, doi:10.1038/nature12650 (2013).

226 Lamppa, G. Methods in Plant Molecular Biology. (1995).

227 Cappadocia, L., Parent, J. S., Sygusch, J. & Brisson, N. A family portrait: structural comparison of the Whirly proteins from Arabidopsis thaliana and Solanum tuberosum. *Acta Crystallogr F* **69**, 1207-1211, doi:10.1107/S1744309113028698 (2013).