Plant Single-Stranded DNA Viruses Selectively Re-Purpose the Host DNA Replication Machinery for Their Multiplication

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Chaonan Shi
aus Henan, China

Tübingen

2025

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	02.10.2025
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Rosa Lozano-Durán
2. Berichterstatter/-in:	Prof. Dr. Thorsten Nürnberger

Erklärung

Hiermit erkläre ich,

- dass ich diese Arbeit selbst verfasst habe.

- dass ich keine anderen als die angegebenen Quellen benutzt und, dass ich alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe.

- dass die eingereichte Arbeit weder vollständig, noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist.

Tübingen, den 09.07.2025

Chaonan Shi

Acknowledgements

First of all, I would like to sincerely thank my supervisor Prof. Dr. Rosa Lozano-Durán, for offering me the opportunity to come to Germany and start my PhD journey. I have truly enjoyed the past four years. Her support extended far beyond my PhD studies; she also gave me many supports in daily life. Her dedication and passion for science will continue to inspire me throughout my future career. I am deeply grateful for her endless patience in guiding me through critical thinking, writing, and preparing various presentations, abstracts, and also my thesis.

I would also thank all the people who have contributed to this project. Special thanks to Laura Medina-Puche and Delphine Pott for your efforts in completing additional replicates of some qPCR experiments. I am also grateful for all the fun conversations I had with Laura, which brightened up many long lab days. Your support, especially during my early days in Germany, meant a lot to me. Thanks to Hua Wei for helping to generate the proximity labelling data in Shanghai, and also for the delicious food you treated me to. I also appreciate the help from Viviane Heise, a Master's student, for her work on the geminiviral origin licensing experiments.

I’m thankful to all the members of the GT lab, especially Huang Tan, who patiently answered my countless scientific questions and supported me not only in research but also throughout the writing process. His kindness and patience have been invaluable for me. I also want to sincerely thank Man Gao and Yu Zhou for their tremendous help, both scientifically and in daily life. Special thanks to Gao Man for being such a wonderful and supportive listener throughout my PhD. My gratitude also goes to Zhihao for sharing your ideas and helpful suggestions on my project. Many thanks to Shaojun Pan and Shuyi Luo for their kindness, support, and all the fun conversations we had. I also would like to say thank you to Clémence Marchal for sharing the first glimpse of our fascinating Rep hexamer and for patiently explaining protein structure predictions to me. I’m also grateful to Kyoka Kuroiwa, the holiday we spent together in Miami remains one of the most cherished memories of my PhD. Thanks to Zhenggang Li for his help with processing the phylogenetic tree. I would also like to highly acknowledge Ying Li and Bettina Stadelhofer, our fantastic technicians, for their kind assistance and for always providing the materials we needed to keep our experiments running smoothly.

I appreciate the excellent support from the ZMBP platform, especially the greenhouse team and the microscopy facility, your efforts greatly accelerated

the progress of my project.

Finally, I would like to express my heartfelt thanks to my parents, sisters, and also my friends. Your warm words have given me tremendous strength and kept me going. A last special thank-you to myself.

Table of Contents

Abstract

Geminiviruses are a family of plant-infecting viruses characterized by twin icosahedral capsids and circular single-stranded (ss) DNA genomes. Members of this family cause devastating diseases in crops worldwide. Geminiviruses replicate in the cell nucleus of the infected plant mainly through a rolling-circle replication (RCR) mechanism. As none of the viral proteins possesses DNA polymerase activity, these viruses heavily rely on the host replication machinery for successful replication. As viral replication is a fundamental step in the infection cycle, the host factors essential for this process can serve as potential targets to develop strategies to control infection. We have previously shown that DNA polymerase α and δ are required for viral replication. However, the composition of the viral replisome remains mostly elusive. The viral replication associated (Rep) protein is highly conserved, and the only virus-encoded protein essential for this process. Here, we used Rep from the geminivirus tomato yellow leaf curl virus (TYLCV) as a bait to capture host factors involved in viral replication in *Nicotiana benthamiana* via TurboID-based proximity labelling (PL) followed by mass spectrometry (MS) analysis. Our data demonstrate that geminiviruses exploit the molecular machinery mediating eukaryotic leading-strand, but not lagging-strand, DNA synthesis in the bidirectional replication fork. An exception is the DNA polymerases, which are swapped. Furthermore, our findings suggest that the composition of the viral replisome is conserved within the phylum Circular Rep-Encoding Single-Stranded (CRESS) DNA viruses. In addition, our work shows that eukaryotic DNA helicase loaders, but not the helicase itself, contribute to the viral genome replication, underscoring a mechanistic divergence in replication initiation between geminiviruses and their plant hosts. Structural prediction indicates that geminiviral Rep assembles into a homo-hexameric complex, analogous to the hetero-hexameric eukaryotic helicase. Taken together, our results demonstrate that geminiviruses selectively repurpose the plant DNA replication machinery to carry out RCR. Collectively, our study provides new insight into viral replication strategies and lays the foundation for the identification of new antiviral targets.

Zusammenfassung

Geminiviren sind eine Familie von Pflanzenviren, die sich durch doppelte ikosaedrische Kapside und zirkuläre einzelsträngige (ss) DNA-Genome auszeichnen. Mitglieder dieser Familie verursachen weltweit verheerende Krankheiten in Nutzpflanzen. Geminiviren replizieren sich im Zellkern der infizierten Pflanze hauptsächlich durch einen Rolling-Circle-Replikationsmechanismus (RCR). Da keines der viralen Proteine DNA-Polymerase-Aktivität besitzt, sind diese Viren für eine erfolgreiche Replikation stark auf den Replikationsapparat des Wirts angewiesen. Da die virale Replikation ein grundlegender Schritt im Infektionszyklus ist, können die für diesen Prozess wesentlichen Wirtsfaktoren als potenzielle Ziele für die Entwicklung von Strategien zur Infektionsbekämpfung dienen. Wir haben bereits gezeigt, dass die DNA-Polymerasen α und δ für die virale Replikation erforderlich sind. Die Zusammensetzung des viralen Replikationskomplexes ist jedoch noch weitgehend unbekannt. Das virale Replikationsprotein (Rep) ist hochkonserviert und das einzige viruskodierte Protein, das für diesen Prozess unerlässlich ist. Hier verwendeten wir Rep aus dem Geminivirus tomato yellow leaf ccurl virus (TYLCV) als Köder, um Wirtsfaktoren zu erfassen, die an der viralen Replikation in Nicotiana benthamiana beteiligt sind, und zwar mittels TurboID-basierter Proximity Labeling (PL) gefolgt von einer Massenspektrometrie (MS)-Analyse. Unsere Daten zeigen, dass Geminiviren die molekulare Maschinerie nutzen, die die DNA-Synthese des führenden Strangs, nicht jedoch des verzögerten Strangs in der bidirektionalen Replikationsgabel vermittelt. Eine Ausnahme bilden die DNA-Polymerasen, die ausgetauscht werden. Darüber hinaus deuten unsere Ergebnisse darauf hin, dass die Zusammensetzung des viralen Replikationskomplexes innerhalb der Phylum Circular Rep-Encoding Single-Stranded (CRESS) DNA-Viren konserviert ist. Darüber hinaus zeigen unsere Arbeiten, dass eukaryotische DNA-Helikase-Loader, nicht jedoch die Helikase selbst, zur viralen Genomreplikation beitragen, was eine mechanistische Divergenz bei der Replikationsinitiierung zwischen Geminiviren und ihren Pflanzenwirten unterstreicht. Strukturelle Vorhersagen deuten darauf hin, dass sich das Geminivirus-Rep zu einem homohexameren Komplex zusammenlagert, analog zur heterohexameren eukaryotischen Helikase. Zusammengenommen zeigen unsere Ergebnisse, dass Geminiviren den DNA-Replikationsapparat der Pflanzen selektiv für die RCR umfunktionieren. Insgesamt liefert unsere Studie neue Erkenntnisse über virale Replikationsstrategien und legt den Grundstein für die Identifizierung neuer antiviraler Zielmoleküle.

List of Figures

List of Tables

List of Abbreviations

AAA+	ATPase associated with various cellular activities
AAV	Adeno-associated virus
AbMV	Abutilon mosaic virus
ACS	ARS consensus sequence
AD	ATPase domain
APC/C	Anaphase-promoting complex/cyclosome
ARS	Autonomously replicating sequence
BcssDV1	*Botrytis cinerea* ssDNA virus 1
BCTV	Beet curly top virus
BeYD	Bean yellow dwarf virus
CDC6	Cell division cycle 6
CDC20	Cell division cycle 20
CDC45	Cell division cycle 45
CDC48	Cell division cycle 48
CDK	Cyclin-dependent kinase
CDS	Coding sequence
CDT1	CDC10-dependent transcript 1
CFDV	Coconut foliar decay virus
ChiLCV	Chilli leaf curl virus
CMG	CDC45-MCM-GINS
CMV	Cauliflower mosaic virus
Co-IP	Co-immunoprecipitation
CP	Capsid protein
CRESS DNA virus	Circular Rep-encoding single-stranded DNA virus
cryo-EM	cryo-electron microscopy
CsalDNAV01	chaetoceros salsugineum DNA virus 01
DDK	Dbf4-dependent kinase
DPB11	DNA polymerase B 11
dsDNA	Double-stranded DNA
EBNA1	Epstein-Barr nuclear antigen 1
EBV	Epstein-Barr virus
ED	Endonuclease domain
EV	Empty vector
FBNYN	Faba bean necrotic yellows virus
FEN1	Flap endonuclease 1
FgGMTV1	Fusarium graminearum gemytripvirus 1
GINS	Sld5-Psf1-Psf2-Psf3
GOIs	Genes of interest

GRS	Geminivirus Rep Sequence
HPeCV	Human periodontal circular-like virus
HUH	His-hydrophobe-His
ICTV	International Committee on Taxonomy of Viruses
iPOND	Isolation of proteins on nascent DNA
IR	Intergenic region
OD	Oligomerization domain
ORC	Origin recognition complex
ORFs	Open reading frames
PCV2	Porcine circovirus 2
PICh	Proteomics of Isolated Chromatin segments
PCNA	Proliferating cell nuclear antigen
PDB	Protein Data Bank
PDS	Phytoene desaturase
PL	Proximity labelling
PNYDV	Pea necrotic yellow dwarf virus
Pol α-primase	DNA polymerase α-primase
Pol δ	DNA polymerase δ
Pol ε	DNA polymerase ε
pre-RC	pre-replicative complex
LIG1	DNA ligase I
LIR	Large intergenic region
MCM	Minichromosome maintenance
MP	Movement protein
MS	Mass spectrometry
MSV	Maize streak virus
MVM	Minute virus of mice
NSP	Nuclear shuttle protein
RBR	Retinoblasmota-related
RCR	Rolling-circle replication
REn	Replication Enhancer protein
Rep	Rep-associated protein
RFC	Replication factor C
RPA	Replication protein A
RPM	Reads per million
SF3	Superfamily 3
SIR	Short intergenic region
SNP	Single nucleotide polymorphism
SR	Silencing-resistant
ssDNA	Single-stranded DNA

SsHADV-1	Sclerotiorum hypovirulence-associated DNA virus 1
SV40	Simian virus 40
TFDaV	Temperate fruit decay-associated virus
TGMV	Tomato golden mosaic virus
ToLCGuV	Tomato leaf curl Gujarat virus
ToLCNDV	Tomato leaf curl New Delhi virus
TrAP	Transcriptional Activator Protein
TRV	Tobacco rattle virus
TYLCV	Tomato yellow curl virus
VIGS	Virus-induced gene silencing
WT	Wild-type
Y2H	Yeast two-hybrid
YMaCV	Yerba mate-associated circular DNA virus

List of Publications

1. **Shi C***, Pott D, Medina-Puche L, Heise V, Wei H, Lozano-Durán R. Plant single-stranded DNA viruses selectively re-purpose the host DNA replication machinery for their multiplication. *In preparation*.

2. **Shi C***, Tan H*, Lozano-Durán R. TurboID-based proximity labelling in plants. *Methods Mol. Biol.* 2025. *In press*.

(*) These authors contributed equally to this work.

3. Tan H*, **Shi C***, Macho AP, Lozano-Durán R. The cartography of plant immunity: Proximity labeling puts a novel SGT1-NSL1 regulatory module on the map. *Mol. Plant.* (Spotlight) 2024 Nov 4;17(11):1645–1647.

(*) These authors contributed equally to this work.

4. Wang L, Tan H, Medina-Puche L, Wu M, Garnelo Gomez B, Gao M, **Shi C**, Jimenez-Gongora T, Fan P, Ding X, Zhang D, Ding Y, Rosas-Díaz T, Liu Y, Aguilar E, Fu X, Lozano-Durán R. Combinatorial interactions between viral proteins expand the potential functional landscape of the tomato yellow leaf curl virus proteome. *PLoS Pathog.* 2022 Oct. PMID: 36256684.

5. Garnelo Gómez B, Holzwart E, **Shi C**, Lozano-Durán R, Wolf S. Phosphorylation-dependent routing of RLP44 towards brassinosteroid or phytosulfokine signalling. J Cell Sci. 2021 Oct 15;134(20):jcs259134. doi: 10.1242/jcs.259134. Epub 2021 Oct 20. PMID: 34569597; PMCID: PMC8572011.

6. Garnelo Gómez B, Rosas-Díaz T, **Shi C**, Fan P, Zhang D, Rufián JS, Lozano-Durán R. The viral silencing suppressor P19 interacts with the receptor-like kinases BAM1 and BAM2 and suppresses the cell-to-cell movement of RNA silencing independently of its ability to bind sRNA. New Phytol. 2021 Feb;229(4):1840-1843. doi: 10.1111/nph.16981. Epub 2020 Oct 29. PMID: 33007117.

1. Introduction

1.1 CRESS DNA viruses

1.1.1 Overview of CRESS DNA viruses

Circular Rep-encoding single-stranded (CRESS) DNA viruses were first described by Rosario *et al*. (2012)[1] as a group of single-stranded (ss) DNA viruses that encode a replication-associated (Rep) protein and are classified within the phylum *Cressdnaviricota*. Members of this phylum possess small, circular genomes, typically ranging in size from approximately 1.7 to 6 kilobases (kb)[2]. A defining feature of CRESS DNA viruses is their conserved Rep protein, which harbors characteristic motifs associated with endonuclease and helicase functions. These motifs are essential for initiating genome replication via a rolling-circle replication mechanisms that takes place within the nucleus of infected host cells[3]. In addition to the Rep, all CRESS DNA viruses encode at least one non-orthologous capsid protein (CP), which is required for virion assembly[4,5]. This phylum exhibits a remarkably broad host range, having been identified in nearly all eukaryotic organisms, including plants, animals, fungi, and unicellular algae such as diatoms[6–8]. Several CRESS viruses have been reported to pose a threat to animal health and contribute to significant crop losses over the past few decades[9–14]. Strikingly, advances in high-throughput sequencing technologies have revealed the widespread prevalence of these viruses in diverse environments, including various soil types[15], marine samples[16], and host-associated niches such as human vaginal secretions, the respiratory tract, gingival tissue[2,17,18].

1.1.2 Taxonomy and diversity of CRESS DNA viruses

Until 2019, the International Committee on Taxonomy of Viruses (ICTV) had identified seven families with the phylum *Cressdnaviricata*: *Geminiviridae*, *Circoviridae*, *Nanoviridae*, *Genomoviridae*, *Bacilladnaviridae*, *Smacoviridae*, and *Redondoviridae*[4,6]. Viruses belonging to these families infect a diverse range of eukaryotic hosts, including animals (*Circoviridae*), diatoms (*Bacilladnaviridae*), fungi (*Genomoviridae*), and plants (*Nanoviridae* and *Geminiviridae*). Additionally, two more viral families—*Smacoviridae* and *Redondoviridae*—which are suspected to infect animals were recently identified through metagenomics[17,19]. In 2021, ICTV proposed coconut foliar decay virus (CFDV) as the sole member of the newly established family *Metaxyviridae*, further expanding the phylum. This represents another family of plant-infecting viruses, in addition to those mentioned earlier[20]. By 2022, ICTV ratified three additional families associated with protozoan parasites—*Naryaviridae*,

Nenyaviridae, and *Vilyaviridae*—increasing the number of families in *Cressdnaviricoda* from eight to eleven[21]. Most recently, a fourth family of plant-infecting viruses, *Amesuviridae*, was also classified within this phylum by ICTV[22]. Thus far, a total of twelve families have been identified within *Cressdnaviricoda*. This phylum is divided into two classes, *Arfiviricetes*, which includes the majority of families (*Bacilladnaviridae*, *Circoviridae*, *Smacoviridae*, *Metaxyviridae*, *Nanoviridae*, *Redondoviridae*, *Naryaviridae*, *Nenyaviridae*, *Vilyaviridae*), and *Repensiviricetes*, which currently consists of all known fugal and plant viruses within the families *Genomoviridae* and *Geminiviridae*. Given the relevance of this study and the available published information, we have only summarized multicellular eukaryote-infecting viruses (nine families) from *Cressdnaviricata* in Table 1.1. A phylogenetic tree based on the amino acid sequences of Rep proteins within this phylum is presented in Figure 1.1. A general overview of these nine viral families is provided below.

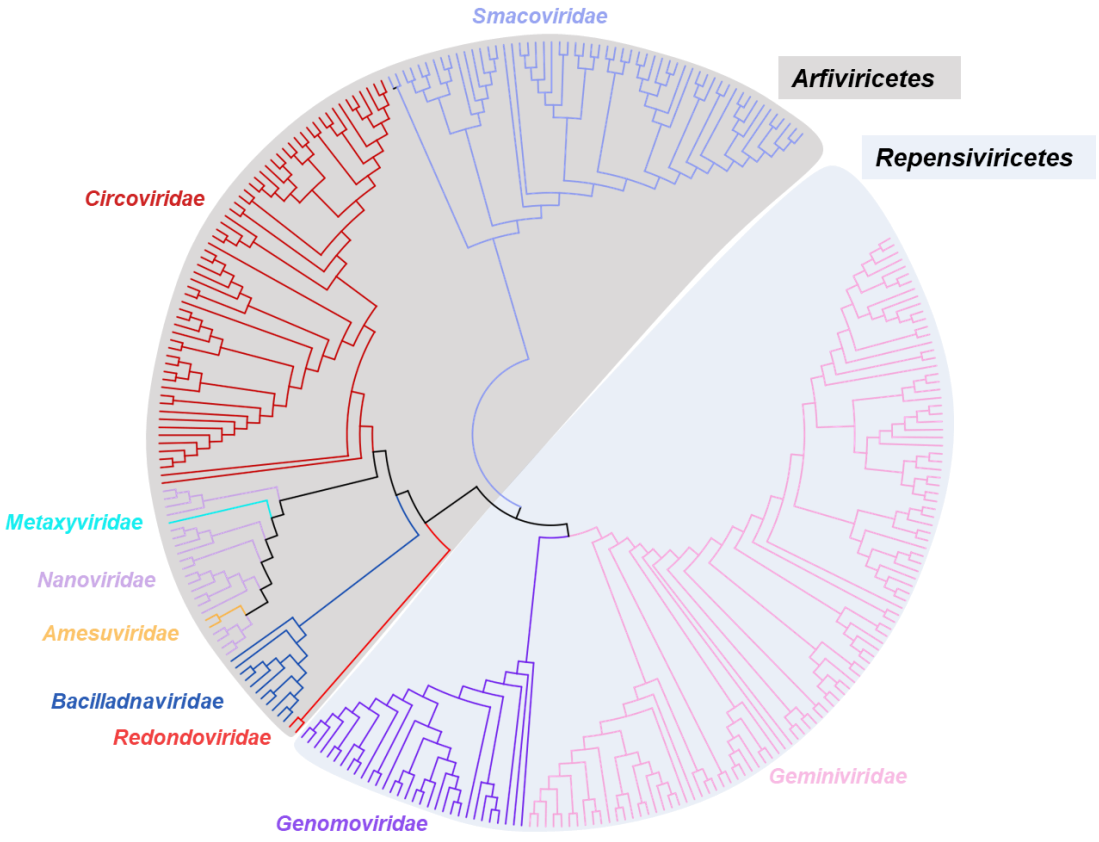


Figure 1.1 Maximum likelihood phylogenetic tree of Rep proteins from representative multicellular eukaryote-infecting CRESS DNA viruses. The phylogenetic tree was constructed using the maximum likelihood method based on

amino acid sequences of Rep proteins from CRESS DNA viruses. Sequences were obtained from Kazlauskas *et al*. (2018) and supplemented with updated entries retrieved from the International Committee on Taxonomy of Viruses (ICTV) and the NCBI database. Multiple sequence alignment was performed using MUSCLE, and the Maximum likelihood tree was generated in MEGA 12 software using the JTT substitution model. Bootstrap analysis was performed with 50 replicates. The resulting tree highlights the evolutionary relationships among CRESS DNA viruses. Viral families are color-coded and labelled in branches. Viruses belonging to the two viral classes are also indicated by two shape-filled colors.

Table 1.1 Basic information of eukaryotic CRESS DNA viral families, modified from Zhao *et al*., 2019

Family	Hosts	Genome size (kb)	Virion diameter (nm)	No. of canonical ORFs	Replication protein	ORF orientation	Origin of replication
Geminiviridae	Plants	2.5-5.2 per molecule (1-2 molecule)	22 x 38	4-8	HUH-Rep	ambisense (or segmented)	TAATATTAC
Circoviridae	Animals	1.7-2.1	15-25	2	HUH-Rep	ambisense	TAGTATTAC
Nanoviridae	Plants	0.98-1.1 per molecule (6-8 molecules)	17-19	5-8	HUH-Rep	segmented	TATTATTAC
Genomoviridae	Fungi, insects	2-2.4	20-22	2/3	HUH-Rep	ambisense	TAATATTAT
Bacilliadnaviridae	Algae (Diatoms)	5.5-6	33-38	3	HUH-Rep	ambisense	Unknown
Smacoviridae	Unknown	2.3-2.9	Unknown	2	HUH-Rep	ambisense	NAGTATTAC
Redondoviridae	Unknown	3.0-3.1	Unknown	3	HUH-Rep	ambisense	TATTATTAT
Metaxyviridae	Plants	0.6-1.3 per molecule (3 molecule plus 9 alphasatellites)	20 nm	3	HUH-Rep	segmented	TAGTATTAC
Amesuviridae	Plants	2.7-3.4	Unknown	3–5	HUH-Rep	ambisense	TAGTATTAC/CATTATTAC

The family *Geminiviridae* consists of plant-infecting viruses and is classified into 15 genera—*Becurtovirus*, *Begomovirus*, *Capulavirus*, *Citlodavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Maldovirus*, *Mastrevirus*, *Mulcrilevirus*, *Opunvirus*, *Topilevirus*, *Topocuvirus*, *Turncurtovirus*, and *Welwivirus*—based on genome organization, host range, and insect vectors. These viruses comprise more than 500 species and are considered the most speciose family in the phylum[23] (https://ictv.global/report/chapter/geminiviridae). A detailed description is provided in Section 1.1.3.

The family *Circoviridae* consists of two genera, *Circovirus* and *Cyclovirus*, which include the smallest known animal-infecting viruses, with genome sizes ranging from 1.7-2.1 kb[6]. Viruses in this viral family possess two open reading frames (ORFs) encoding Rep and CP. The two genera are distinguished by their genome organization. In *Circovirus*, the *Rep* gene is located on the virion-sense (positive-sense) strand, while *CP* is transcribed from the antisense strand. In contrast, in *Cyclovirus*, the ORFs are arranged in the opposite orientation, with the *Rep* gene located on the antisense strand and *CP* gene on the virion-sense strand. Additionally, an intron has been identified in the *Rep* gene of *Cyclovirus* but not in *Circovirus*[24]. Members of the family *Circoviridae* have been found to infect a broad of hosts, including mammals, fish, birds, and insects[25].

The family *Nanoviridae* comprises two genera, *Babuvirus* and *Nanovirus*. This family consists of plant-infecting viruses that are unique in possessing multipartite, circular, ssDNA genomes, composed of 6-8 components, each approximately 1kb in length[13]. Viruses in both genera share five homologous DNA components: DNA-R (master replication initiation protein, M-Rep), DNA-S (capsid protein, CP), DNA-C (cell-cycle link protein, Clink), DNA-M (movement protein, MP) and DNA-N (nuclear shuttle protein, NSP)[26–29]. However, three components encode proteins of unknown function (DNA-U1, DNA-U2, and DNA-U4) are exclusive to the genus *Nanovirus*, while DNA-U3 is specific for genus *Babuvirus*[30]. Babuviruses comprise three species, each with six genomic components, whereas the twelve species are classified under genus *Nanovirus* typically have eight components, except for parsley severe stunt associated virus (PSSaV), which have only seven components due to the absence of DNA-U4 (https://ictv.global/report/chapter/nanoviridae/). Nanoviruses infect dicotyledonous hosts, predominantly legumes, whereas the species from babuviruses mainly infect tropical monocotyledonous crops, including bananas, abaca, taro and cardamom[31]. The characteristic symptoms of nanovirus infection in plants include yellowing, stunting, and necrosis of plant

tissues[32–35]. The insect vector for the *Nanoviridae* family is an aphid, which transmits the viruses in a circulative, nonpropagative manner[36]. Not all of viral proteins encoded by nanoviruses are required for successful infection in plants. It has been demonstrated that DNA-R (M-Rep), DNA-S (CP), and DNA-M (MP) are crucial for symptom induction by faba bean necrotic yellows virus (FBNYN) in plants using *Agrobacterium tumefaciens*-mediated transformation[36]. Several nanoviruses genome components encode Rep-related proteins, but only master Rep (M-Rep/DNA-R) is necessary and sufficient for the replication of other ssDNA components. None of the additional Rep-related proteins are capable of independently inducing replication, except for their own genome components[37].

The family *Genomoviridae* was created by ICTV in 2016[38], and together with *Geminiviridae*, it belongs to the class *Repensiviricetes*, while the remaining of CRESS DNA viruses are classified under the class *Arfiviricetese*[39] (see Figure 1.1). According to ICTV, *Genomoviridae* currently comprises 10 genera and 237 species, categorized as follows: *Gemycircularvirus* (126 species), *Gemyduguivirus* (12 species), *Gemygorvirus* (8 species), *Gemykibivirus* (50 species), *Gemykolovirus* (16 species), *Gemykrogvirus* (13 species), *Gemykroznavirus* (7 species), *Gemytondvirus* (1 species), *Gemytripvirus* (1 species), and *Gemyvongvirus* (3 species)[40,41]. Typically, members of this family possess a monopartite, circular, ssDNA genome (1.8-2.4 kb) encoding two ORFs: Rep and CP. These genes are arranged in an ambisense orientation. One notable example is sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1), the first known ssDNA virus to infect fungi. The Rep protein of SsHADV-1 shares the highest similarity with those of geminiviruses, while its CP does not exhibit sequence similarity to capsid proteins from any other known viral taxa[41]. In contrast to the predominantly monopartite genomes of genomoviruses, fusarium graminearum gemytripvirus 1 (FgGMTV1), a member of the newly established genera *Gemytripvirus*, possesses a tripartite genome: DNA-A encodes a Rep protein; DNA-B encodes a geminivirus-like protein CP, while DNA-C encodes a protein of unknown function[42]. Intriguingly, *Botrytis cinerea* ssDNA virus 1 (BcssDV1), previously believed to possess a monosegmented genome[43], was recently characterized as a tetrasemented virus based on next-generation sequencing and Sanger sequencing[44]. The four genomic segments, DNA-A, DNA-B, DNA-C, and DNA-D, encode a Rep, CP, and two hypothetical proteins, respectively.

The *Bacilladnaviridae* family consists of marine algae-infecting viruses and is composed of seven genera: *Aberdnavirus*, *Diatodnavirus*, *Keisodnavirus*,

Kieseladnavirus, *Protobacilladnavirus*, *Puahadnavirus*, *Seawadnavirus* (https://ictv.global/taxonomy, 2025). Viruses belonging to bacilladnaviruses contain three genes responsible for Rep, CP and a protein of unknown function. Notably, the CP protein of bacilladnaviruses was likely acquired through horizontal gene transfer from an ancestral nadavirus, a group of ssRNA viruses[45]. Unlike viruses from other families, bacilladnaviruses possess a partially double stranded (~700 to 800 bp) region within their ~6 kb ssDNA genome[46]. Despite the first diatom-infecting DNA virus, chaetoceros salsugineum DNA virus 01 (CsalDNAV01), being identified over a decade ago, the documentation and characterization of ssDNA viruses within this family remain less developed compare to other CRESS DNA viral families[47]. Future research is warranted, as diatoms infected by bacilladnaviruses play a crucial role in marine and freshwater ecosystems, potentially influencing global biogeochemical cycles and primary production.

The family *Smacoviridae* has been officially classified into twelve genera, comprising 143 species, by ICTV: *Babosmacovirus*, *Bonzesmacovirus*, *Bostasmacovirus*, *Bovismacovirus*, *Cosmacovirus*, *Dragsmacovirus*, *Drosmacovirus*, *Felismacovirus*, *Huchismacoviruses*, *Inpeasmacovirus*, *Porprismacovirus*, and *Simismacovirus*[48]. Similar to most viruses in the families *Circoviridae* and *Genomoviridae*, members of *Smacoviridae* encode two viral proteins, Rep and CP, which are arranged in either a unidirectional orientation (exclusively for genus *Felismacovirus*) or an ambisense orientation (the rest of the genera). Although smacoviruses have been detected in various environmental and faecal samples derived from a wide range of animals, no definitive hosts have been identified for any member of this family to date[48]. The discovery of majority of these viruses has relied on metagenomic analyses, with subsequent validation performed using Sanger sequencing[19]. However, the study by Díez-Villaseñor and Rodriguez-Valera (2019) suggested that methanogenic archaea in the gut, belonging to the genus *Methanomassiliicoccus*, serve as potential hosts of smacoviruses using CRISPR analysis[49].

The family *Redondoviridae* consists of a single genus, *Torbevirus*, which currently includes two species[50]. Viruses in this family possess three ORFs, with the two largest encoding Rep and CP in opposite orientations. The third ORF shares the same orientation as CP and is entirely embedded within *CP*-coding region[50]. Redondoviruses, identified in the human respiratory tract through metagenomic sequencing, are considered the second most prevalent viruses in human respiratory samples, surpassed only by members of the family

Anelloviridae, which are known to be ubiquitous in humans[17,18]. More recently, five novel redondoviruses, collectively referred to as “human periodontal circular-like virus (HPeCV)”, were detected in the gingival tissue of patients with chronic periodontitis[51]. However, whether redondoviruses actively replicate in humans, infect other eukaryotic hosts in the respiratory tract or gingiva, or are merely passively inhaled and deposited on respiratory surfaces or in the oral cavity remains largely unknown.

The family *Metaxyviridae* was established by ICTV in 2021, making it the third identified group of plant-infecting viruses[20]. CFDV, which is transmitted in a persistent-circulative manner by the cixiid plant hopper *Myndus taffini*, is currently the sole member of this family[52]. The CFDV genome is suggested to be tripartite, consisting of DNA-S.1 (CP), DNA-S.2 (CP) and DNA-gamma (MP). Notably, the proteins encoded by DNA-S.1 and DNA-S.2 share 94% amino acid identity and exhibit high similarity to the capsid protein of grapevine red blotch virus (family *Geminiviridae*, genus *Grablovirus*). There is no clear candidate for Rep protein based solely on sequence homology. However, CFDV is associated with nine alphasatellites, one of which is hypothesized to encode a master Rep protein, designated CFDV DNA-R.

The family *Amesuviridae* is the fourth identified group of plant-infecting viruses and was officially established in 2023[22]. It is classified into two genera, *Temfrudevirus* (1 species) and *Yermavirus* (1 species) based on genome organization. A novel, highly divergent virus named temperate fruit decay-associated virus (TFDaV), was described to infect temperate fruit trees (apple, pear, and grapevine) in Brazil in 2015[53]. In 2018, another novel ssDNA virus, yerba mate-associated circular DNA virus (YMaCV), was discovered in Argentina, associated with yerba mate plant[54]. Both viruses were later classified under family *Amesuviridae,* with TFDaV assigned to the genus *Temfrudevirus* and YMaCV to the genus *Yermavirus*. The genomes of both viruses contain 5 ORFs: C1 and C2 are located on the complementary strand, while V1, V2, and V3 are positioned on the viral-sense strand. However, their genome functions differ: In TFDaV, V1 encodes a MP, while V2 encodes the CP. In YMaCV, their functions are reversed, with V1 encoding the CP and V2 encoding the MP. In both viruses, C1 encodes the replication-associated protein, Rep. Thus far, the natural insect vector responsible for the transmission of viruses in this family remains unknown[53,54].

Information on representative classified species, along with their accession numbers across different viral families, is shown in Table 1.2.

Table 1.2 Overview of CRESS DNA viral families: genera, number of classified species, genome accession number, and genome organization

Family	Genus	Species no.	Type species	Accession number	Genomic organization
Geminiviridae	*Becurtovirus*	3	*Beet curly top Iran virus*	EU273818	Monopartite
	Begomovirus	445	*Bean golden yellow mosaic virus*	DNA-A: L01635; DNA-B: L01636	Mono- and bipartite
	Capulavirus	5	*Euphorbia caput-medusae latent virus*	HF921459	Monopartite
	Citlodavirus	6	*Camellia chlorotic dwarf-associated virus*	MG452759	Monopartite
	Curtovirus	3	*Beet curly top virus*	M24597	Monopartite
	Eragrovirus	1	*Eragrostis curvula sreak virus*	FJ665631	Monopartite
	Grablovirus	3	*Grapevine red blotch virus*	JQ901105	Monopartite
	Maldovirus	3	*Apple geminivirus 1*	KM386645	Monopartite
	Mastrevirus	50	*Maize streak virus*	Y00514	Monopartite
	Mulcrilevirus	2	*Mulberry crinkle leaf virus*	KR131749	Monopartite
	Opunvirus	1	*Opuntia virus 1*	MN100000	Monopartite
	Topilevirus	2	*Tomato apical leaf curl virus*	MG491195	Monopartite
	Topocuvirus	1	*Tomato pseudo-curly top virus*	X84735	Monopartite
	Turncurtovirus	3	*Turnip curly top virus*	GU456685	Monopartite
	Welwivirus	2	*Welwitschia mirabilis associated geminivirus B*	BK061150	Monopartite
Circoviridae	*Circovirus*	70	*Porcine circovirus 2*	AY651850	Monopartite
	Cyclovirus	90	*Human associated cyclovirus 1*	GQ404847	Monopartite
Nanoviridae	*Nanovirus*	12	*Pea necrotic yellow dwarf virus*	DNA-C: JN133280; DNA-M: JN133281; DNA-N: JN133282; DNA-R: GU553134; DNA-S: JN133279; DNA-U1: JN133283; DNA-U2: JN133284; DNA-U4: JN133285	Multipartite
	Babuvirus	3	*Banana bunchy top virus*	DNA-C: L41578; DNA-M: L41575; DNA-N: L41577; DNA-R: S56276; DNA-S: L41574; DNA-U3: L41576	Multipartite
Genomoviridae	*Gemycircularvirus*	43	*Sclerotinia gemycircularvirus 1*	GQ365709	Monopartite
	Gemyduguivirus	1	*Dragonfly associated gemyduguivirus 1*	JX185428	Monopartite
	Gemygorvirus	5	*Staling associated gemygorvirus 1*	KF371632	Monopartite
	Gemykibivirus	16	*Dragonfly associated gemykibivirus 1*	JX185430	Monopartite
	Gemykolovirus	2	*Pteropus associated gemykolovirus 1*	LK931484	Monopartite

	Gemykrogvirus	3	*Pteropus associated gemykolovirus 1*	LK931484	Monopartite
	Gemykroznavirus	1	*Rabbit associated gemykronznavirus 1*	KF371631	Monopartite
	Gemytondvirus	1	*Ostrich associated gemytondvirus 1*	KF371630	Monopartite
	Gemyvongvirus	1	*Human associated gemyvongvirus 1*	KP974693	Monopartite
	Gemytripvirus	1	*Fusarium graminearum gemytripvirus 1*	DNA-A: MK430076; DNA-B: MK430077; DNA-C: MK430078	Tripartite
Bacilladnaviridae	*Aberdnavirus*	1	*Avonheates virus SG_61*	OM154946	Monopartite
	Diatodnavirus	1	*Chaetoceros diatodnavirus 1*	AB781089	Monopartite
	Keisodnavirus	1	*Avon-Heathcote Estuary associated kieseladnavirus*	AQA27298	Monopartite
	Kieseladnavirus	4	*Chaetoceros tenuissimus DNA virus SS12-43V*	LC379168	Monopartite
	Protobacilladnavirus	8	*Chaetoceros protobacilladnavirus 1*	AB193315	Monopartite
	Puahadnavirus	4	*Avonheates virus SG_479*	OM154950	Monopartite
	Seawadnavirus	3	*Bacillariodnavirus LDMD-2013*	KF133809	Monopartite
Smacoviridae	*Babosmacovirus*	3	*Bovine associated bovismacovirus 1*	JN634851	Monopartite
	Bonzesmacovirus	2	*Bovine faeces associated smacovirus 1*	KT862223	Monopartite
	Bostasmacovirus	1	*Bovine faeces associated smacovirus 6*	KT862229	Monopartite
	Bovismacovirus	7	*Bovine faeces associated smacovirus 3*	KT862222	Monopartite
	Cosmacovirus	1	*Bovine associated cosmacovirus 1*	KT862228	Monopartite
	Dragsmacovirus	3	*Odonata-associated circular virus-5*	KM598410	Monopartite
	Drosmacovirus	7	*Camel associated drosmacovirus 1*	KM573769	Monopartite
	Felismacovirus	23	*Chicken virus mg6_1197*	MN379595	Monopartite
	Huchismacovirus	5	*Human associated huchismacovirus 1*	KP233180	Monopartite
	Inpeasmacovirus	3	*Inpeasmacovirus humas1*	MH500283	Monopartite
	Porprismacovirus	88	*Chimpanzee associated porprismacovirus 1*	GQ351272	Monopartite
	Simismacovirus	2	*Chlorocebus cynosuros associated smacovirus*	LC386199	Monopartite
Redondoviridae	*Torbevirus*	2	*Human lung-associated vientovirus*	MK059763	Monopartite
Metaxyviridae	*Cofodevirus*	1	*Coconut foliar decay virus*	DNA-S.1: MF926436; DNA-S.2: MF926439; DNA-gamma: MF926441; DNA-R: MF926434	Tripartite
Amesuviridae	*Temfrudevirus*	1	*Temperate fruit decay-associated virus*	KR134312	Monopartite
	Yermavirus	1	*Yerba mate-associated circular DNA virus*	MG748715	Monopartite

1.1.3 Overview of the family *Geminiviridae*

The name “geminiviruses” is derived from the characteristic twinned or geminate virions in this viral family, which encapsidate the circular, ssDNA viral genomes, ranging from 2.5 to 5.2 kb in size[55]. As the most species-rich family within the phylum *Cressdnaviricota*, *Geminiviridae* is classified into 15 genera: *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Citlodavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Maldovirus*, *Mastrevirus*, *Mulcrilevirus*, *Opunvirus*, *Topilevirus*, *Topocuvirus*, *Turncurtovirus*, and *Welwivirus*. This family encompasses over 500 viral species, classified based on genome organization, host range, and insect vectors[56,57]. Among these, the genus *Begomovirus* comprises the largest number of species, with 450 identified to date (see Table 1.2). Notably, only viruses within this genus have either one (monopartite) or two (bipartite, DNA-A and DNA-B) genome components, while viruses classified within the remaining of 14 genera contain only a single DNA genome[56].

Table 1.3 Plant hosts, insect vectors, and origin of replication of different genera across *Geminiviridae*

Genera	Plant hosts	Insect vectors	Origin of replication
Becurtovirus	Dicots	Leafhoppers (*Circulifer haematoceps*)	TAAGATTCC
Begomovirus	Dicots	Whiteflies (*Bemisia tabaci*)	TAATATTAC
Capulavirus	Dicots	Aphids	TAATATTAC
Citlodavirus	Dicots	Unknown	TAATATTAC
Curtovirus	Dicots	Leafhoppers	TAATATTAC
Eragrovirus	Monocots	Unknown	TAAGATTCC
Grablovirus	Dicots	Treehoppers (*Spissistilus festinus*)	TAATATTAC
Maldovirus	Dicots and monocots	Unknown	TAATATTAC
Mastrevirus	Most monocots and some dicots	Leafhoppers	TAATATTAC
Mulcrilevirus	Dicots	Leafhoppers (*Tautoneura mori*)	TAATATTAC
Opunvirus	Dicots	Cochineals	TAATATTAC
Topilevirus	Dicots	Unknown	TAATATTAC
Topocuvirus	Dicots	Treehoppers (*Micrutalis malleifera* Fowler)	TAATATTAC
Turncurtovirus	Dicots	Leafhoppers (*Circulifer haematoceps*)	TAATATTAC
Welwivirus	Monocots	Unknown	TAATATTAC

The genome organization of representative members across the 15 genera within the family *Geminiviridae* is depicted in Figure 1.2. Generally, viruses in this family exhibit a limited coding capacity, with partially overlapping ORFs encoding only 4 to 8 canonical proteins in an ambisense orientation[58–60]. The DNA-A component of bipartite geminiviruses encodes proteins associated with viral DNA replication, transcription, suppression of gene silencing,

encapsidation, and vector transmission, designated as AC1/C1/Rep (Replication-associated protein), AC3/C3/REn (Replication Enhancer protein), AC2/C2/TrAP (Transcriptional Activator Protein), AC4/C4, AV2/V2, and CP (Capsid/Coat Protein). The DNA-B component encodes two proteins: BC1/MP (Movement Protein) and BV1/NSP (Nuclear Shuttle Protein), which are involved in cell-to-cell movement and the intracellular trafficking of viral DNA, respectively. The genomic component of monopartite geminiviruses is homologous to the DNA-A of bipartite geminiviruses, encoding the proteins C1/Rep, C2/TrAP, C3/REn, and C4. Besides, V2 and CP, located on the viral-sense strand, are involved in viral movement[61–63]. Recently, a number of works revealed the existence of additional ORFs in geminivirus genomes, expanding their limited coding capacity[64–72]. Additionally, at least one intergenic region (IR or LIR, large intergenic region), approximately 300 bp in length and exhibiting high sequence similarity among geminiviruses, is consistently found within the viral genome and lies between genes encoded on the viral strand and the complementary strand. This IR plays a crucial role in viral replication, as it carries the origin of replication (see Table 1.3), which is essential for the cleavage of the viral DNA during initiation of rolling-circle replication[73]. Viruses from certain genera, such as mastreviruses, also possess a short intergenic region (SIR) on the opposite site of the LIR, which likely serves as a primer for the synthesis of complementary strand[74].

Despite their small genomes, geminiviruses cause substantial economic losses on a variety of vegetable and field crops, including beans, cassava, maize, wheat, cucurbits, pepper, and tomato[58,61,75,76]. The propagation and prevalence of geminiviruses in nature are heavily reliant on insect transmission, with insects vectors including whiteflies, leafhoppers, treehoppers, and aphids[77–80]. An overview of plant hosts, insect vectors, and origin of replication across different genera within *Geminiviridae* is presented in Table 1.3. It is worth noting that insect vectors for viruses within certain genera, such as *Eragrovirus* (established in 2014), *Citlodavirus*, *Maldovirus*, and *Topilevirus* (reported in 2022), as well as recently classified *Welwivirus* (proposed in 2023), remain unidentified to date[57,81,82].

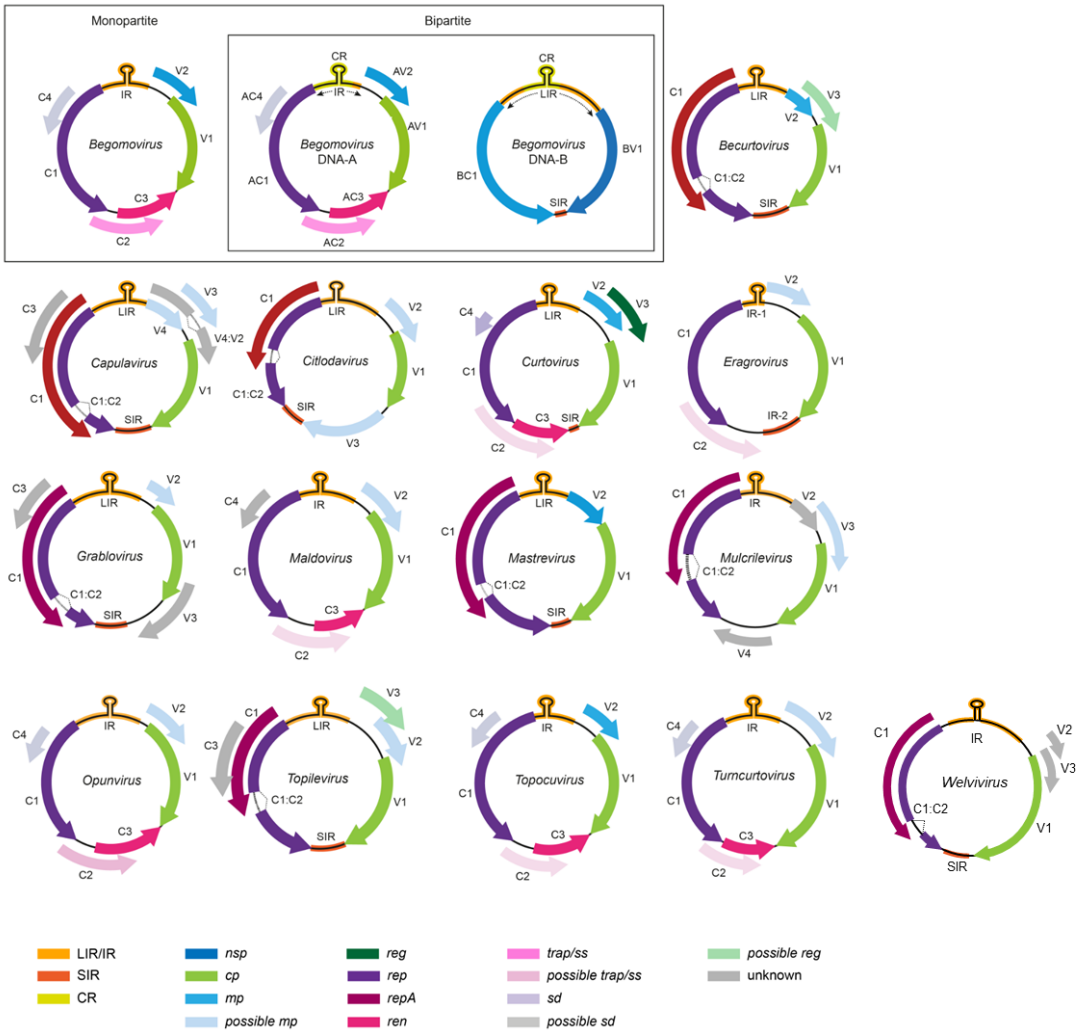


Figure 1.2 Genome organization of representative members of the fifteen genera in the family *Geminiviridae*. The stem-loop structure containing the nonanucleotide sequence located in the intergenic region (IR/IR-1) and long intergenic region (LIR) is indicated. ORFs are denoted as being encoded on the virion-sense (V) or complementary-sense (C) strand, and corresponding protein products are coded by color. C1/AC1: replication associated protein (Rep); C2/AC2: transcriptional activator protein (TrAP); C3/AC3: replication enhancer protein (REn); V1/AV1: coat protein (CP); V2/AV2: pre-coat protein; V3/AV3: cell-to-cell movement in genus *Curtovirus*. BC1: movement protein (MP) and BV1: nuclear shuttle protein (NSP). CR, SCR, SIR and LIR refer to Common Region, Satellite Conserved Region, Short-Intergenic Region, and Long-Intergenic Region, respectively. Figures are modified from the ICTV online report on the family *Geminiviridae*: https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/geminiviridae.

1.1.4 Replication of CRESS viruses

Rolling circle replication was first described as the mechanism of genome

replication in the *Escherichia coli* bacteriophage ΦX174[83,84]. Since then, numerous studies have documented that rolling-circle replication is utilized by a wide range of genetic elements, including bacterial plasmids, archaeal plasmids, transposable elements, and CRESS DNA viruses[85–88]. A defining characteristic of these elements is the presence of a Rep protein, which plays a crucial role in rolling-circle replication. However, Rep lacks DNA polymerase activity, making these elements entirely dependent on host replication machinery, which recruited by Rep, for genome synthesis. The Rep protein, highly conserved among CRESS DNA viruses, consists of an HUH (His-hydrophobe-His) endonuclease domain (ED) at the N-terminus, an oligomerization domain in the middle, and a helicase/ATPase domain at the C-terminus (Figure 1.3)[89–91]. The structure of Rep enables it to initiate, process and complete rolling-circle replication within the host cell nucleus. During rolling-circle replication in CRESS DNA viruses, a double-stranded (ds) DNA intermediate, derived from the conversion of ssDNA, serves as the replication template. Rep binds to a nonanucleotide motif (see Table 1.1) located in a hairpin structure and introduces a site-specific nick between positions seventh and eighth of the motif (e.g., NANTAT↓AC), which is situated within the IR of the (+)-strand (viral-sense strand)[92–96]. The cleavage generates a 3'-OH end, which serves as a primer for unidirectional synthesis by host DNA polymerases[6]. Meanwhile, the 5' end remains covalently linked to the Rep protein and is displaced by the newly synthesized (+)-strand until the completion of rolling-circle replication. The DNA unwinding of the DNA duplex during this process is facilitated by either cellular or viral helicases[90]. Several studies have demonstrated that Rep from circoviruses and geminiviruses exhibits DNA helicase activity, suggesting that Rep may function as a helicase during rolling-circle replication to unwind the DNA duplex[90,97,98]. At the termination of rolling-circle replication, Rep catalyzes the joining reaction at the viral origin of replication, leading to the release of the initial ssDNA from the dsDNA replicative form[73,93,95]. A rolling-circle replication model of CRESS DNA virus replication is shown in Figure 1.4. Experimental evidence supporting rolling-circle replication as a replication strategy employed by geminiviruses has been well-documented[73,99]. Furthermore, certain eukaryotic ssDNA viruses, particularly geminiviruses, have been reported to replicate their genome through recombination-dependent replication[99–101], a mechanism also utilized for the bacteriophage T4[102]. Notably, in contrast to the study in which both rolling-circle replication and recombination-dependent replication occur simultaneously in the naturally infected leaves of abutilon mosaic virus (AbMV, genus *Begomovirus*), the replication modes of bacteriophage T4 occur sequentially: Rolling-circle replication dominates during the early phage of

infection, but once viral genome accumulation reaches a certain threshold, replication switches to recombination-dependent replication. However, evidence is lacking as to whether both replication modes occur simultaneously within the same cell or different cells, potentially containing different stages of infection, as the leaf tissue used in experiments were a mixture[99]. Therefore, future studies focusing on single-cell resolution approaches may be valuable in addressing this question. To date, no direct experimental evidence supports the involvement of recombination-dependent replication in the replication of viruses from other families within the phylum *Cressdnaviricoda,* such as *Nanoviridae*[103].

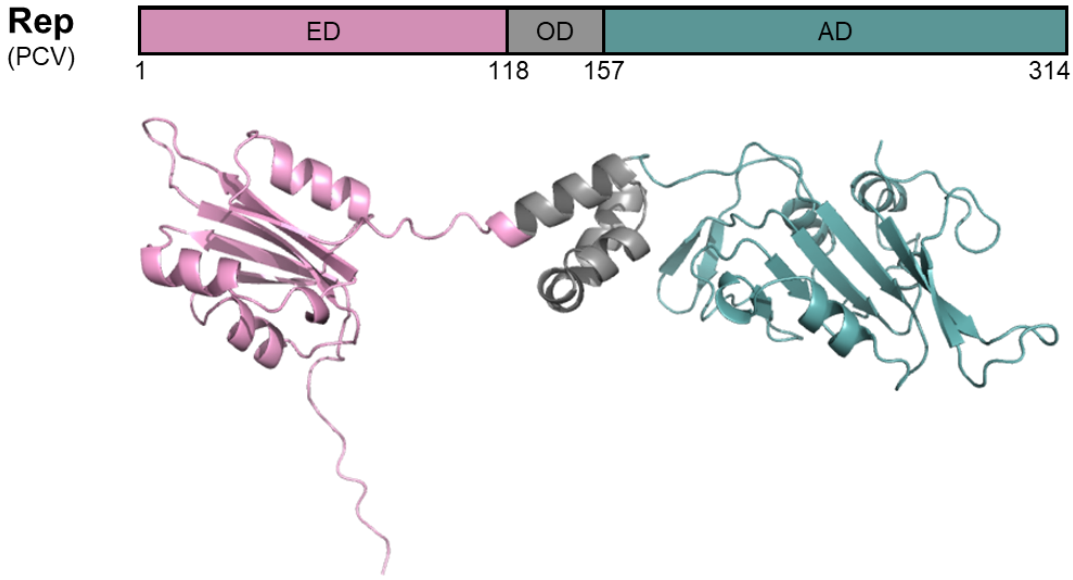


Figure 1.3 Domain architecture (upper panel) and structural organization (lower panel) of Rep from CRESS DNA viruses. Schematic representation of Rep, with porcine circovirus 2 (PCV2, genus *Circovirus*, family *Circoviridae*, phylum *Cressdnaviricoda*) as an example. The endonuclease domain (ED, pink), oligomerization domain (OD, grey), and ATPase domain (AD, light teal) are indicated and labelled. Figure is modified from Tarasova *et al*. (2021) with minor modifications[90].

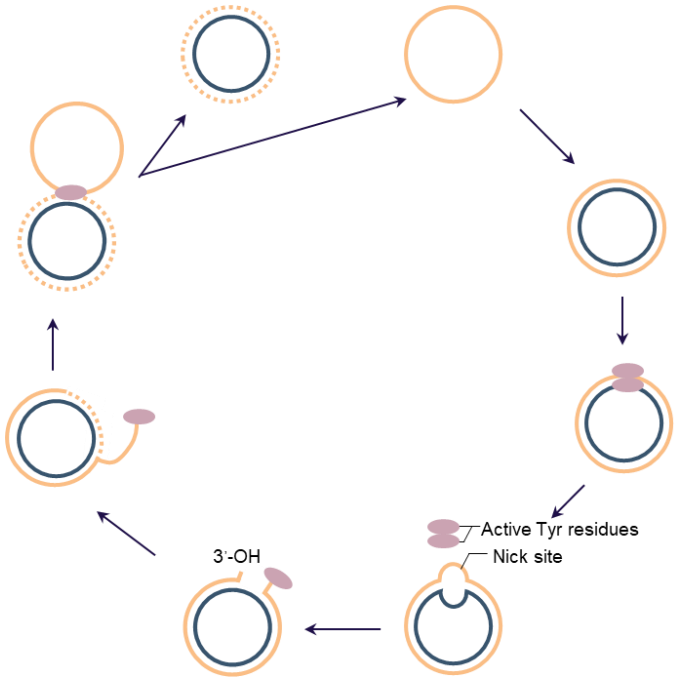


Figure 1.4 Schematic representation of rolling-circle replication (RCR) employed by CRESS DNA viruses. The ssDNA genome (orange) is converted to a dsDNA intermediate (the complementary strand is shown in navy blue) by (a) host DNA polymerase(s). The first synthesized Rep protein (pink, with one tyrosine in the motif III) recognizes and binds to the intergenic region (IR) during initiation. Rep introduces a nick in the nonanucleotide motif; the released 3'-OH is further synthesized (dashed orange line) by host proteins through replacing the original viral strand, while the 5' end is covalently bound by Rep during elongation. At the end of RCR, Rep catalyzes the joining reaction, leading to the release of the initial viral strand from the dsDNA replicative form. Figure is modified from Zhao *et al*., 2019 with minor modifications[6].

1.2 Eukaryotic DNA replication

DNA replication is a fundamental process in all living organisms, producing two identical copies of DNA from a single original DNA molecule[104–107]. It proceeds through three main stages: initiation, elongation, and termination. In eukaryotes, DNA replication is highly conserved, with each stage tightly regulated by numerous proteins. Each of these three stages is discussed in detail in the following subsections.

1.2.1 Eukaryotic DNA replication initiation

Eukaryotic DNA replication initiates at numerous sites known as replication origins. The number of replication origins varies among organisms and is primarily determined by genome size. For example, the aforementioned CRESS DNA viruses, as well as bacteria and archaea, possess a small circular chromosome, thereby having only a single replication origin[6,108,109]. In contrast,

to efficiently duplicate their large chromosomes during S phase, eukaryotic genomes contain multiple replication origins, ranging from approximately 500 in budding yeast *Saccharomyces cerevisiae* to 30,000-50,000 in human cells[110,111]. This enables replication forks to be established at multiple locations, ensuring timely genome duplication.

Although replication factors are highly conserved across viruses, prokaryotes and eukaryotes, cumulative studies indicate that each eukaryotic system exhibits unique regulatory features. While DNA replication has been extensively studied in yeasts and mammalian cells, research on replication mechanisms in land plants remains relatively limited. Here, we mainly use the best characterized system to date—budding yeast—as a model to illustrate the replication initiation stage.

Eukaryotic replication initiation occurs in two subsequential steps: (1) Origin licensing: During the G1 phase, the replicative helicase, composed of six minichromosome maintenance 2-7 (MCM2-7) subunits, is loaded onto origin DNA; (2) origin firing: During G1-to-S phase transition, the helicase is activated, and pre-RCs (pre-replication complexes) are converted into two active bidirectional replication fork[112,113]. The origin recognition complex (ORC), composed of six distinct subunits (ORC1-6), binds to yeast replication origins in a sequence-specific manner. In budding yeast, replication origins are determined by autonomously replicating sequence (ARSs), which contain a T-rich, nonpalindromic ARS consensus sequence (ACS)[114–116]. The ACS, along with downstream B elements, forms a binding site for ORC. Specifically, ORC binds and bends the ACS/B1 site, and the ATPase activity of cell division cycle 6 (CDC6), which shares sequence similarity with the ORC1 subunit, stabilizes ORC binding at the origin. Another key origin licensing factor is CDT1 (CDC10-dependent transcript 1, also known as DNA replication factor). In yeast, CDT1 stably associates with MCM and contains an inactive dioxygenase fold important for loading[117,118], a feature not observed in multicellular eukaryotes. The ORC-CDC6 complex (product 1, OC) recruits the first CDT1-MCM complex to replication origins. The MCM complex encircles the DNA duplex through the MCM2-5 gate, forming the ORC1-CDC6-CDT1-MCM complex (product 2, OCCM), in which ORC binds the C-terminal side of the MCM ring. Upon ATP hydrolysis, CDC6 and CDT1 are released, and ORC transitions to the opposite N-tier side of the first loaded MCM ring, forming the MCM-ORC complex (product 3, MO). CDC6 re-engages ORC, recruiting a second CDT1-MCM complex, which assembles into a head-to-head (N-face to N-face) MCM2-7 double hexamer (DH) encircling the DNA duplex (product 4, MM). All of the above processes constitute origin licensing, and the associated factors

collectively form pre-RC.

Origin firing activates the helicase and converts pre-RC into two active bidirectional replication forks. This process requires multiple factors, including Sld3-7, Sld2, GINS (5-1-2-3 in Japanese, Sld5-Psf1-Psf2-Psf3), DNA polymerase B 11 (Dpb11) and the largest subunit of DNA polymerase ε (Pol ε), POL2 (POLE1 in humans)[119,120]. In addition to these factors, two key protein kinases—cyclin-dependent kinase (CDK) and Dbf4-dependent kinase (DDK)—play essential roles in origin firing[121]. The activities of CDK and DDK are regulated by anaphase promoting complex/cyclosome (APC/C), which targets and degrades the regulatory subunits of both kinases (cyclin and Dbf4, respectively). At the end of G1 phase, APC/C inactivation leads to the accumulation of cyclin and Dbf4, thereby activating CDK and DDK. DDK efficiently phosphorylates the MCM4 and MCM6 subunits of the MCM complex, but only when the complex is loaded as a double hexamer onto the DNA duplex. Phosphorylated MCM creates binding sites for Sld3 protein, which associates with Sld7 and subsequently recruits cell division cycle 45 (CDC45) to the MCM double hexamer. CDK phosphorylates Sld2 and Sld3, generating binding sites for Dpb11, which in turn recruits Sld2 to the MCM double hexamer. Sld2 then facilitates the recruitment of Pol ε and GINS to MCM[122–127]. CDC45 and GINS stably associate with MCM, forming the CMG (CDC45-MCM-GINS) complex, which consists of 11 subunits. CMG assembly is accompanied by initial untwisting of DNA and separation of the MCM double hexamer into two discrete but inactive CMG helicases. Subsequently, the MCM10 protein further promotes DNA untwisting and facilitates helicase activation in a ATP hydrolysis-dependent manner[128]. The two CMG pass each other, positioning the N-terminal side of MCM near the replication fork, where they translocate along the leading-strand template in the 3'-to-5' direction[128,129]. A schematic overview of the stepwise processes of eukaryotic DNA replication initiation is depicted in Figure 1.5.

Notably, not all organisms contain a conserved binding site at replication origins that is recognized by replication initiation proteins, such as ORC/CDC6 in eukaryotes or DnaA in bacteria. In stark contrast, origins of replication in most of eukaryotes lack strict DNA sequence specificity. Even in some other yeasts, such as S*chizosaccharomyces pombe*, replication origins tend to be AT-rich but lack a consensus sequence. The AT-hook domain located at the N-terminus of the ORC4 subunit can bind to AT-rich DNA sequences and is necessary for origin organization[130,131]. In metazoans, ORC was shown to bind to DNA without sequence specificity as well[132–134]. In this case, a G quadruplexes was formed by a noncanonical four-stranded helical structure with G-rich repeated element

and may play a role in defining replication origins[135–138]. However, fusing a Gal4 protein to ORC can force it sequence-specifically bind to the chromosomal Gal4 DNA binding site in *Drosophila melanogaster* and further promote the formation of a pre-replication complex, but not occurring in the negative control[139], this raises the possibility that ORC might interact with sequence-specific proteins and establish the initiation of DNA replication at a specific chromosomal site.

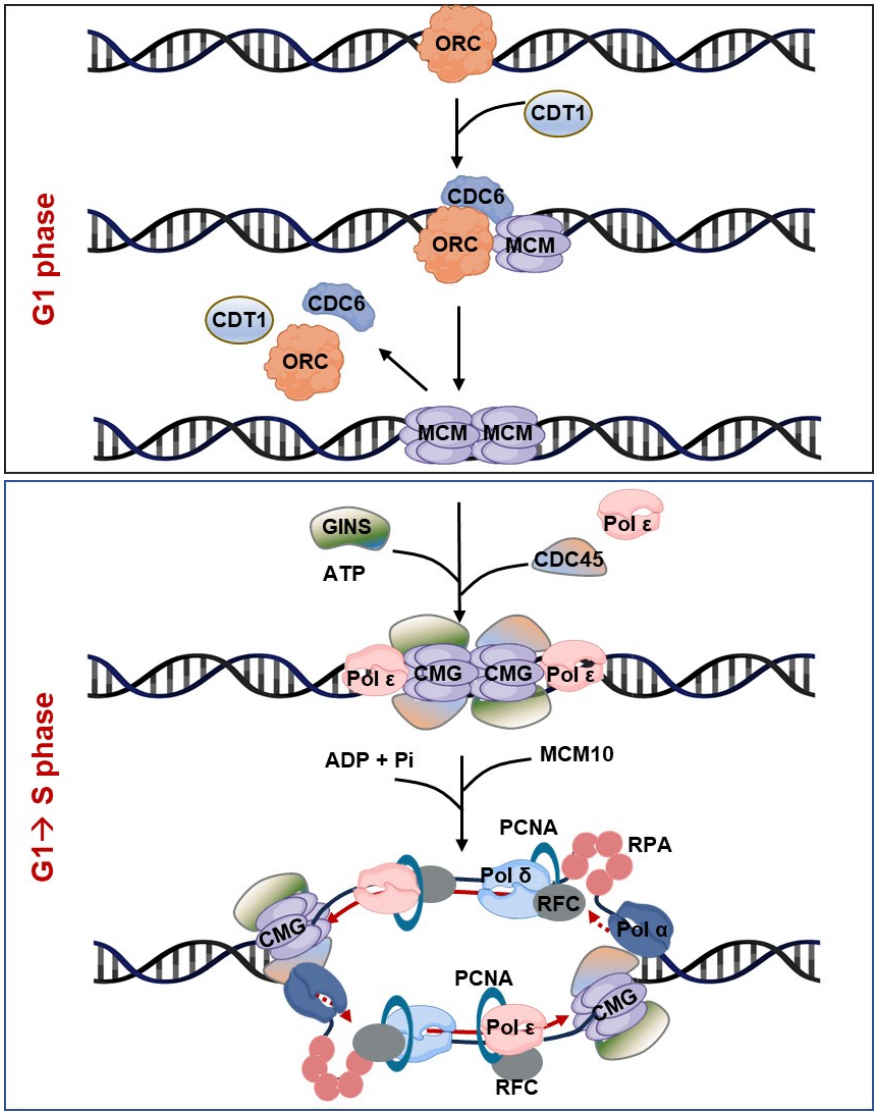


Figure 1.5 Schematic representation of eukaryotic replication initiation. This diagram illustrates the two stages of replication initiation in eukaryotic cells. (1) Origin licensing in G1 phase: During G1, the origin recognition complex (ORC) binds to replication origins and recruits CDC6 and CDT1, which together facilitate the loading of the MCM2-7 helicase complex onto DNA to form the pre-replicative complex (pre-RC). This step ensures that replication origins are licensed by loading the MCM helicase, which remains inactive until S phase; (2) helicase activation at G1-to-S phase transition: Upon entry into S phase, S-phase kinases trigger the conversion of the pre-

RC into an active helicase complex. This involves the recruitment of CDC45 and the GINS complex, which associate with MCM2-7 to form the active CMG helicase complex (CDC45-MCM-GINS). Concurrently, DNA polymerase ε (Pol ε) and other replication factors are recruited to form the pre-initiation complex, initiating bidirectional replication fork establishment and DNA synthesis.

1.2.2 Eukaryotic DNA replication elongation

During the origin firing process, multiple factors are required for helicase activation, including Pol ε, which is responsible for leading-strand synthesis during replication elongation. Pol ε is a heterotetramer composed of a catalytic subunit (POL2) and three regulatory subunits (DPB2, DPB3, and DPB4) in budding yeast, and of a catalytic subunit (POLE1) and three regulatory subunits (POLE2, POLE3, and POLE4) in multicellular eukaryotes. It has been demonstrated that the CMG complex and Pol ε form a stable assembly on the DNA duplex, referred to as CMGE, through reconstitution using purified proteins from budding yeast[120,140–142]. Upon activation by MCM10, the lagging-strand template is ejected from the central channel of the CMG complex[128]. As the CMGE complex translocate along the leading-strand template and establish the replication fork, a Y-shaped structure where the parent DNA duplex is separated into leading- and lagging-stand templates[106]. At the replication fork, the leading strand is synthesized continuously in the same direction as the movement of the replication fork, whereas lagging strand is synthesized discontinuously in the opposite direction, resulting in the formation of Okazaki fragments[104].

Three DNA polymerases are essential for duplicating the parental DNA during unperturbed replication elongation: DNA polymerase α (Pol α), DNA polymerase δ (Pol δ), and the aforementioned Pol ε. Pol α is a Pol α-primase complex, which consists of two polymerase subunits (POL1 and POL12 in budding yeast, and POLA1 and POLA2 in multicellular eukaryotes; POL1/POLA1 serves as the catalytic subunit, and POL12/POLA2 as the regulatory subunit) and two primase subunits (PRI1 and PRI2 in budding yeast, and PRIM1 and PRIM2 in multicellular eukaryotes; PRI1/PRIM1 is the catalytic subunit, and PRI2/PRIM2 is the regulatory subunit)[143,144]. It was previously accepted that Pol α synthesizes short RNA/DNA primer on both leading and lagging strand, a model first proposed in 1990[145]. However, in 2018, Aria and Yeeles demonstrated that the start sites of the leading-strand replication—for example, in the leftward direction—are located on the right side of the origin of replication, suggesting that the leading-strand synthesis is established from primers initially synthesized on the lagging strand by Pol α-primase[146]. The lagging-strand template, generated by the translocation of CMG helicase along

the leading-strand template, is targeted by Pol α-primase for priming. This targeting is mediated by an interaction between Pol α-primase and the CMG complex through the replisome-organizing factor Ctf4[147–150]. Ctf4 is a homo-trimeric complex that functions as a replisome hub connecting DNA replication and other proteins via Ctf4-interacting peptides[151]. A study by Jones *et al*. (2023) further uncovered a conserved mechanism in which the interaction between the non-catalytic subunit of primase and the CMG complex positions the catalytic subunit PRIM1/Pri1 directly above the exit channel for the lagging-strand template. This structural arrangement, revealed by cryo-electron microscopy (cryo-EM) analysis of Pol α-primase from both budding yeast and humans, explains why priming synthesis occurs efficiently and specifically on the lagging-strand template[152].

Pol δ is another major replicative DNA polymerase, composed of a catalytic subunit (POL3) and two regulatory subunits (POL31 and POL32) in budding yeast, and a catalytic subunit (POLD1) together with three regulatory subunits (POLD2, POLD3, and POLD4) in multicellular eukaryotes. It is primarily responsible for lagging-strand synthesis during unperturbed eukaryotic DNA replication. Recent studies have also demonstrated that Pol δ plays an important role in both the initiation and termination of leading-strand synthesis[146,153]. Moreover, under replication stress conditions, such as during break-induced replication, Pol δ is responsible for synthesizing both the leading and lagging strands[154–156]. Additional studies indicate that Pol δ can contribute to leading-strand synthesis when either the non-catalytic subunit Dpb2 of DNA Pol ε or the CMG helicase complex is impaired[157,158]. Interestingly, the dsDNA virus Simian virus 40 (SV40) also exclusively utilizes Pol δ to extend the primers synthesized by Pol α-primase on both the leading and lagging strands during viral replication[159,160].

At the replication fork, the RNA/DNA primer synthesized by Pol α-primase on the lagging strand is extended by Pol δ, in cooperation with the sliding clamp proliferating cell nuclear antigen (PCNA) and the clamp loader replication factor C (RFC), to produce Okazaki fragments approximately 200 nucleotides (nt) in length[161]. Upon encountering the 5'-end of the downstream Okazaki fragment, Pol δ displaces the existing primer, generating a single-stranded flap structure that is subsequently cleaved by flap endonuclease 1 (FEN1). However, a subset of these flap structures can evade FEN1 cleavage and become bound and stabilized by the ssDNA-binding protein replication protein A (RPA), which prevents their degradation by nucleases. RPA-coated flaps are resistant to FEN1 activity but can be processed by the alternative nuclease, DNA2. DNA2 displaces RPA and cleaves the long flap to generate a short flap of

approximately 5-6 nt, which can then be efficiently removed by FEN1. The resulting nicks are sealed by DNA ligase I (LIG1; known as Cdc9 in budding yeast), ultimately generating an intact lagging-strand DNA from thousands to millions of Okazaki fragments[161–164]. This process is catalyzed by three key enzymes—Pol δ, FEN1, and LIG1—that together constitute the core Okazaki fragment maturation machinery. Notably, all three enzymes contain one or more PCNA-interacting motifs that mediate their association with the PCNA homotrimer. A “toolbelt” model has been proposed, in which each of these enzymes simultaneously binds to a separate monomer of the PCNA trimer, thereby coordinating processive Okazaki fragment maturation[165].

Leading-strand synthesis is primarily catalyzed by the highly processive Pol ε. Pol ε interacts with PCNA, which further enhances its intrinsic processivity. This interaction is mediated through the nonessential regulatory subunits Dpb3 and Dpb4 of Pol ε[166]. As the replication fork progresses, the unwinding of the parental duplex generates positive supercoils ahead of the fork. To alleviate the torsional stress and prevent excessive supercoiling that could impede replication fork progression, DNA topoisomerases of type I and type II are recruited to relax these supercoils and maintain replication fork stability[167]. A schematic representation of replication fork progression during eukaryotic DNA replication is shown in Figure 1.6.

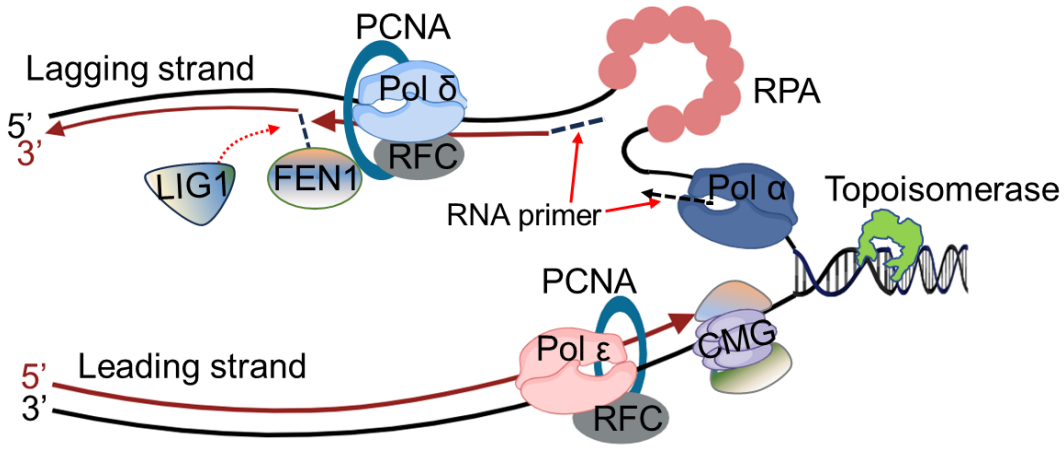


Figure 1.6 Schematic of the eukaryotic replication fork. The diagram illustrates the main proteins involved in the progression of the replication fork. The CMG complex functions as the replicative helicase, unwinding the parental DNA duplex ahead of the fork. Topoisomerase relieves supercoiling tension generated ahead of the fork by CMG helicase. On the leading strand, Pol ε carries out continuous DNA synthesis in the 5’ to 3’ direction. In contrast, the lagging strand is replicated discontinuously through the formation of Okazaki fragments synthesized by DNA polymerase δ (Pol δ). Each Okazaki fragment is initiated by the DNA polymerase α-primase (Pol α-primase)

complex, which synthesizes short RNA-DNA primers. When Pol δ encounters the downstream RNA-DNA primers, it displaces them, generating a short flap structure that is removed by flap endonuclease 1 (FEN1). DNA ligase 1 (LIG1) subsequently seals the resulting nicks, completing the maturation of the Okazaki fragment. Replication protein A (RPA) binds to the exposed ssDNA to stabilize it and prevent from degradation. The proliferating cell nuclear antigen (PCNA) acts as a sliding clamp that enhances the processivity of Pol δ and Pol ε, and plays critical role in Okazaki fragment maturation. Replication factor C (RFC) functions as the clamp loader, loading PCNA onto DNA. The coordinated action of these factors ensures accurate and efficient DNA replication.

1.2.3 Eukaryotic DNA replication termination

DNA replication initiation and elongation are coupled with the assembly of replisome components, which facilitate the duplication of parental DNA at the replication fork. Conversely, replication termination involves the disassembly of the replisome, occurring when two adjacent DNA replication fork converge. Unlike the extensively studied molecular mechanisms of eukaryotic DNA replication and elongation, the process of replication termination remains relatively limited[168,169]. The first insights into replication termination were described by Maric *et al.* and Priego Moreno *et al.* in 2014[170,171]. Their studies demonstrated that ubiquitylation of MCM7 within the CMG helicase complex occurs exclusively under conditions that permit replication completion in both budding yeast and *Xenopus laevis*[172].

Like initiation and elongation, DNA replication termination occurs throughout the S-phase[173]. When two adjacent forks converge, the CMG helicases pass one another and reach a ssDNA-dsDNA junction, which comprises the 5’ end of the downstream Okazaki fragment. The CMGs continue translocating along DNA substrates and eventually position themselves on dsDNA. Importantly, CMG disassembly occurs only when the helicase is bound to dsDNA, as this process is suppressed by replication fork DNA[174–177]. As a key event in eukaryotic replication termination, CMG disassembly was first described in 2014 in budding yeast and *X. laevis*[170,171]. These studies demonstrated that K48-linked ubiquitin chains are assembled on the MCM7 subunit of the replicative helicase, marking it for disassembly. In budding yeast, this polyubiquitylation event requires the Cullin1-RING E3 ubiquitin ligase SCF^{Dia2} (Skp, Cullin, F-box protein Dia2), which is specific to yeast. The E3 ligase responsible for MCM7 ubiquitylation and CMG unloading is $CUL\text{-}2^{LRR\text{-}1}$ (Cullin RING ligase 2 associated with LRR1), a complex specific to metazoans, including organisms such as worms, frogs, mice, and human cell lines[178–181]. Despite differences in E3 ligases across eukaryotes, a recent study

demonstrated that the leucine-rich repeat domains of both Dia2 and LRR1 directly interact with the N-terminal tier of the MCM2-7 helicase, specifically binding to the MCM3 and MCM5 zinc-finger domains. These domains engage with the excluded DNA strand (lagging-strand template) during replication elongation, thereby preventing premature MCM7 ubiquitylation and CMG disassembly[182]. Furthermore, this study showed that replisome disassembly can be triggered if the excluded strand is mispositioned, suggesting a potential mechanism for maintaining replication fork stability under replication stress conditions.

The ubiquitylated MCM7 subunit can be recognized and extracted from chromatin by the ATPase cell cycle division 48 (CDC48, also known as p97 or valosin-containing protein, VCP) in cooperation with cofactors[170,171,183]. CDC48 is a double-ring ATPase with a central pore, composed of six CDC48 monomers. Each monomer consists of an N-terminal domain and two ATPase domains, D1 and D2[184,185]. Similar to the mechanism described by Twomey *et al*. (2019), in which CDC48, in complex with UFD1-NPL4, one of its most well-characterized cofactors, initiates substrate extraction by primarily unfolding the ubiquitin molecule[186], CDC48 also recognizes the ubiquitin chains on MCM7 rather than MCM7 itself during replication termination[183,187]. Whether MCM7 is the sole CMG subunit targeted for ubiquitylation and subsequent extraction by CDC48 remains an open question. A study by Deegan *et al*. (2020) provided some insights into this by detecting a low level of ubiquitylated MCM4 in a reconstituted system. However, MCM7 was the only subunit actively unfolded by the CDC48-UFD1-NPL4 complex[176]. Notably, MCM4 is positioned adjacent to MCM7 within the hexameric MCM2-7 helicase complex[188]. Future studies could investigate whether additional subunits of the MCM2-7 complex undergo ubiquitylation *in vivo*. Furthermore, it would be interesting to explore whether viral helicases involved in viral DNA replication, such as SV40 large T antigen and PCV2 Rep, which form homo-hexameric structures, undergo a similar ubiquitylation-mediated disassembly process during replication termination. A diagram illustrating the eukaryotic replication termination is shown in Figure 1.7.

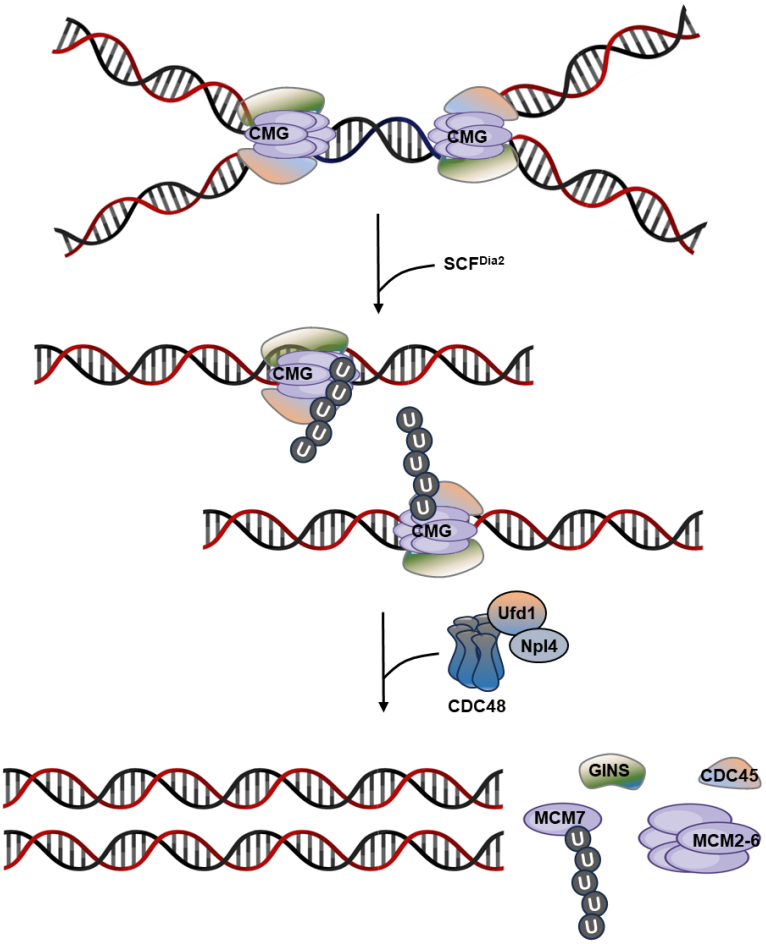


Figure 1.7 Schematic of eukaryotic DNA replication termination. Replication termination occurs when two converging replication forks meet. The MCM7 subunit of the CMG complex is polyubiquitylated by the E3 ubiquitin ligase SCF^{Dia2} (in budding yeast). The ubiquitylated CMG is then recognized by the Ufd1–Npl4 adaptor complex, which recruits the AAA+ (ATPase associated with various cellular activities) CDC48 (also known as p97). CDC48 extracts MCM7 from chromatin, leading to the disassembly of the CMG complex. This coordinated process ensures the completion of DNA synthesis, resulting in two fully replicated dsDNA molecules and the release of individual replisome components. This figure is modified from Mukherjee and Labib (2019) with minor modifications[189].

1.3 Geminiviral replication

1.3.1 Overview of geminiviral replication

All CRESS DNA viruses, including geminiviruses, do not encode their own DNA polymerases and hence heavily rely on host replication machinery. At the onset of geminiviral infection in a plant host, the virus-carrying insect vector feeds on the plant's sap, releasing virion to phloem-associated cells[58]. The priming

events that convert viral ssDNA into dsDNA vary among different geminiviral genera. In the genus *Mastrevirus*, a small primer complementary to the SIR severs as the replication primer. However, viruses from other genera, such as *Begomovirus*, possess genomes that lack an SIR. In these viruses, priming events instead occur within the IR, which contains cis-acting element that define the origin of dsDNA synthesis[74,190]. Nevertheless, the precise molecular mechanisms governing the conversion of virion-derived ssDNA into dsDNA in the host nucleus—prior to the production of any viral proteins—remains largely elusive. Interestingly, host-encoded primase subunits (PRIM1 and PRIM2) and Pol α subunits (POLA1 and POLA2), which are responsible for RNA primer synthesis, have been identified as essential for the replication of tomato yellow curl virus (TYLCV, genus *Begomovirus*) in the model *Solanaceae* host *Nicotiana benthamiana*[191,192]. Notably, *PRIM1* has been identified as a resistance gene against the geminivirus tomato leaf curl New Delhi virus (ToLCNDV)[193]. Furthermore, the study by Wu *et al*. confirmed that Pol α is required for the conversion of ssDNA to dsDNA. Despite the known physical interaction between the regulatory subunit (POLA2) of Pol α and C3, the involvement of Pol α in the synthesis of dsDNA intermediate in the absence of viral proteins remains poorly understood at the mechanistic level. Once the dsDNA intermediate is formed, it assembles into a minichromosome with 11-13 nucleosomes and serves as a template for both viral transcription by host RNA polymerase II and rolling-circle replication with involving the recruitment of the host DNA replication machinery[194]. When the first viral protein Rep is translated, it binds the origin of replication in a sequence-specific manner and nicks the nonanucleotide motif within the IR, initiating rolling-circle replication (see detailed information in Section 1.1.4). A recent breakthrough study demonstrated that replicative Pol α and Pol δ are essential for geminiviral DNA replication. This discovery resolves the long-standing question of which host DNA polymerases are exploited by geminiviruses[192]. Following one complete round of rolling-circle replication, the original ssDNA released from the dsDNA intermediate can either be encapsidated by the CP protein or undergo another round of rolling-circle replication, while the dsDNA intermediate serves as template for viral replication. A simplified model illustrating the geminiviral life cycle is presented in Figure 1.8. In addition to rolling-circle replication, geminiviruses also employ recombination-dependent replication, a mechanism widely characterized in the bacteriophage T4[99,100,195,196]. In this process, partially replicated or host-degraded ssDNA invades the loop structure of dsDNA through homologous recombination, and subsequently undergo elongation by host replication-related proteins. The newly synthesized ssDNA generated via recombination-dependent replication can then serve as a

template for the conversion of ssDNA to dsDNA[197]. In the bacteriophage T4, UvsX, a functional homolog of the eukaryotic recombinase Rad51, is the key factor required for strand-invasion during recombination-dependent replication[195,198]. Intriguingly, a previous study demonstrated that Rad51 is involved in the replication of mungbean yellow mosaic India virus (MYMIV, genus *Begomovirus*) in *Arabidopsis thaliana*, suggesting a potential role of Rad51 in geminiviral recombination-dependent replication process[199].

Notably, most geminiviruses are confined to the phloem companion cells, while some are also capable of infecting mesophyll cells[196,200]. In these highly differentiated cells, the levels of replicative enzymes required for geminiviral genome replication become undetectable. To overcome this barrier, geminiviruses must reprogram the host cell cycle, inducing S-phase re-entry to create a cellular environment permissive for DNA synthesis[201]. The retinoblastoma (Rb) (or retinoblasmota-related, RBR, in plants)/E2F pathway is a key regulatory module controlling the G1/S transition of the cell cycle and is highly conserved across all eukaryotes[202]. The E2F transcription factors, in association with their dimerization partner proteins, regulate the expression of genes involved in DNA replication, thereby controlling the onset of S phase[203]. The activity of E2Fs is, in turn, regulated by RBR, which binds to E2Fs via a docking site located within its pocket domain[202]. During an active cell cycle, RBR is phosphorylated by upstream CDKs, leading to the release of E2F, which subsequently actives the expression of cell cycle-related genes[204]. A growing body of evidence suggests that geminiviral Rep and/or RepA proteins interact with RBR through a canonical RBR-binding motif (LxCxE), resulting in the release of E2F and the induction of a cellular environment suitable for geminiviral replication. For instance, the RepA proteins from the mastreviruses wheat dwarf virus (WDV), bean yellow dwarf virus (BeYD), and maize streak virus (MSV), have been shown to interact with RBR via the canonical motif[205–208]. Additionally, the Rep protein from begomoviruses, including tomato golden mosaic virus (TGMV), also binds to RBR, but via a novel motif, suggesting that alternative interaction mechanisms may be employed[209].

The anaphase-promoting complex/cyclosome (APC/C) is a multisubunit E3-ubiquitin ligase complex that plays a crucial role in cell-cycle progression by recognizing specific targets and facilitating their degradation in corporation with its co-activator, cell division cycle 20 (CDC20)[210]. A recent study demonstrated that geminiviruses employ two novel modes to reprogram the host cell cycle: (i) the Rep protein of TYLCV interacts with CDC20, leading to APC/C activation and the subsequent degradation of RBR; (ii) the TYLCV C2 protein inhibits APC/C^{CDC20}, thereby promoting cyclin-mediate RBR depletion via the ubiquitin-

proteasome pathway. The degradation of RBR results in the release of E2F, which subsequently activates the transcription of DNA polymerase genes, creating a nuclear environment favorable for viral replication[211]. However, the study does not clarify whether RBR is phosphorylated by cyclins or not (mode ii).

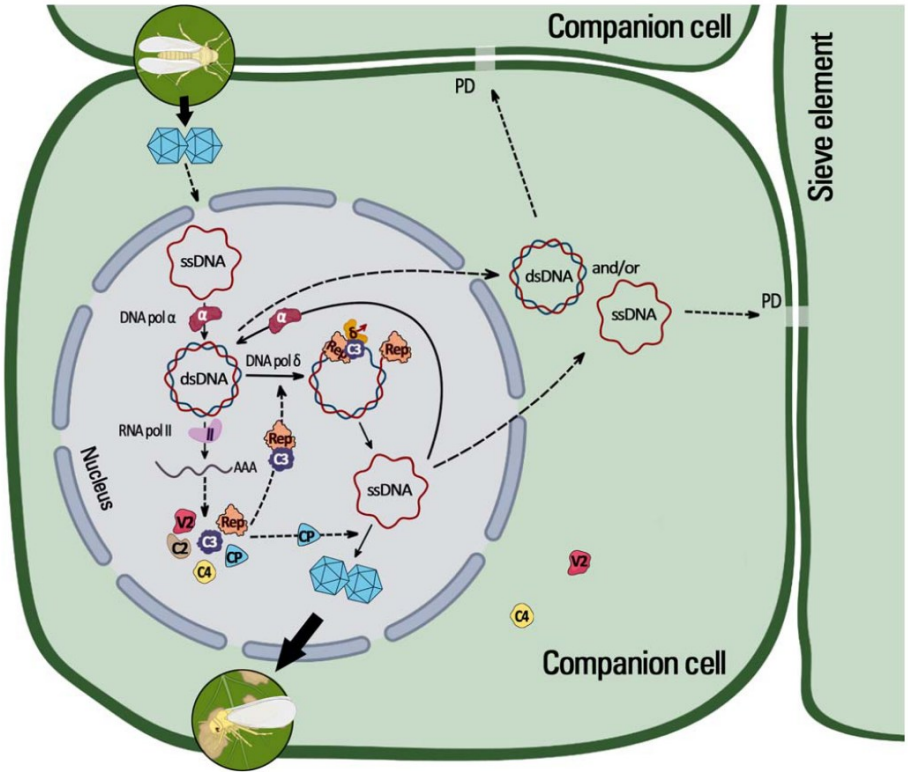


Figure 1.8 Life cycle of geminiviruses. In this diagram, a monopartite begomovirus is used as an example. When insect vector whitefly carrying geminiviruses feeds on the sap of plants, the virions were released to the phloem-associated cells. The capsid protein is disassembled and the released viral genome reaches the nucleus, where viral replication occurs. In the nucleus, viral ssDNA is converted into a dsDNA intermediate with the help of DNA polymerase α (Pol α). The dsDNA intermediate is serving as the template for the bi-directional transcription with RNA polymerase II and unidirectional RCR, respectively. Viral proteins will be ultimately produced, including Rep, C2, C3, C4, CP, and V2. Rep binds to the origin of replication and introduces a nick to initiate RCR. Together with C3, they recruit host replication machinery components to the viral genome, for which Pol δ is required. Newly synthesized viral DNA can move intra- and intercellularly, establishing a systemic infection. In addition, viral ssDNA can be encapsidated and acquired by insect vectors, hence spreading to other plants. This figure is taken from Wu *et al.*, 2022[63].

1.3.2 Rep/RepA and C3

The Rep protein—designated AC1 in the DNA-A component of bipartite

begomoviruses and C1 in monopartite begomoviruses and viruses from other geminiviral genera—is encoded on the complementary-sense strand of the geminiviral genome and is the only viral protein essential for genome replication. As a highly conserved protein among members of the phylum *Cressdnaviricota*, geminiviral Rep shares structural similarities and unique features on its domain architecture. The N-terminal of geminiviral Rep contains a DNA binding and DNA nicking domain, a characteristic feature reminiscent of the endonuclease domain found in CRESS DNA virus Rep proteins. Additionally, a highly conserved Geminivirus Rep Sequence (GRS) motif is embedded in the N-terminus, serving as a distinctive signature exclusive to all geminivirus Rep proteins[212]. Furthermore, motifs I, II, and III, located in the N-terminal region of geminiviral Rep, show high similarity to rolling-circle replication initiator proteins of eubacterial plasmids[88]. A highly conserved tyrosine residue within motif III is involved in DNA cleavage at the onset of rolling-circle replication[94]. Interestingly, based on the replication mechanism of ΦX174, where two active tyrosine residues within the gpA protein are essential for continuous replication, it has been speculated that two geminiviral Rep proteins may function in concert to nick the nonanulceotide sequence during rolling-circle replication[213]. The middle and C-terminal regions of Rep contain an oligomerization domain and a helicase/ATPase domain, respectively. Notably, members of some geminivirus genera encode an additional viral protein called RepA, also encoded on the complementary-sense strand. RepA shares high sequence similarity with Rep, particularly in domain organization, and plays an essential role in viral replication[214,215]. In genera such as *Mastrevirus*, Rep is translated from a spliced mRNA transcribed from *C1* and *C2* ORFs, whereas RepA is derived from the unspliced *C1* ORF. Similar to Rep protein, RepA contains a DNA binding domain, DNA nicking domain, and motifs I, II, and III in the N-terminal region, allowing it to recognize viral DNA and exert nicking activity during rolling-circle replication. However, unlike Rep, RepA lacks the helicase/ATPase domain in its C-terminal region, which is likely responsible for the DNA helicase activity of the former during viral replication[97,98]. A RBR binding site embedded within the oligomerization domain of RepA mediates cell cycle reprogramming through its interaction with host RBR[205–207,216]. The C-terminal region of Rep contains the highly conserved Walker A and Walker B motifs, which are essential for the ATPase activity and classify geminiviral Rep as a superfamily 3 (SF3) helicase[217]. Similar to geminiviral Rep, the well-characterized SV40 large T antigen and PCV2 Rep are also SF3 helicases[90,218]. The domain organization of geminiviral Rep and RepA proteins is illustrated in Figure 1.9.

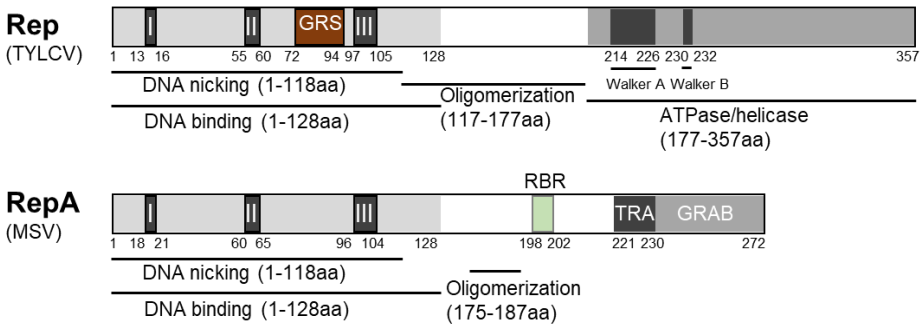


Figure 1.9 Organization of the geminiviral Rep and RepA proteins. Schematic representation of Rep from TYLCV and RepA from MSV. The functional regions responsible for DNA binding, nicking, organization, and helicase activity are indicated and labelled. The motif I, II, III, Geminivirus Rep Sequence (GRS), RBR binding site (RB), promoter-transactivation domain (TRA), and Geminivirus RepA-binding (GRAB) sites are shown. Figure is modified from Wu *et al*. (2022).

The Rep protein, which is multifunctional, interacts with both viral proteins and numerous host factors. Its characteristic features have been extensively reviewed in several studies[58,59,219–222]. As mentioned earlier, two tyrosine residues are likely required for site-specific nicking of the nonanucleotide sequence within the IR during rolling-circle replication. Therefore, Rep must interact with itself to facilitate this process. The self-interaction of Rep has been widely documented across different geminivirus species[207,223–225]. In addition to self-interaction, Rep interacts with C3, contributing to viral replication enhancement[223,226,227]. Potentially to expand the functional repertoire of the viral proteome, Rep from TYLCV has been shown to interact with several other viral proteins, including C2, C3, C4, CP and V2, in addition to itself[225]. During infection, Rep plays multiple roles, including: (1) reprogramming the host cell cycle[209,228,229] (including RepA from mastreviruses, see Section 1.3.1); (2) mediating the initiation and termination steps of rolling-circle replication[58,230]; (3) recruiting host replication machinery, such as RFC, PCNA, and RPA, to the viral genome[231–233]. Previous studies have shown that Rep from different geminiviruses interacts with several replication-associated factors, including RFC1, PCNA, RPA32, MCM2, and MCM3. However, direct evidence confirming the role of these factors in geminiviral replication is lacking[232–237]. Additionally, Rep interacts with two key components involved in homologous recombination, Rad51 and Rad54, suggesting their potential involvement in the recombination-dependent replication process, possibly through the recruitment by Rep[199,238]. Beyond its role in promoting viral DNA replication, Rep also functions as a silencing suppressor, inhibiting transcription gene silencing (TGS) by repressing the expression of host maintenance methyltransferases[239].

Moreover, Rep has been shown to interfere with the host SUMOylation system by binding to the SUMO-conjugating enzyme SCE1, thereby creating a suitable environment for viral infection[240–242]. However, Rep is also targeted by host defense mechanisms. For example, the nuclear autophagy pathway restricts viral infection by degrading Rep protein[243]. As an early viral protein, Rep represses its own transcription[244]. A study demonstrated that Rep from chilli leaf curl virus (ChiLCV, genus *Begomovirus*) binds to the viral genome and interacts with components of the monoubiquitination machinery, promoting histone post-translational modification on ChiLCV minichromosomes to enhance viral gene transcription[245]. A detailed overview of Rep/RepA interactors in plants is provided in Table 1.4.

Table 1.4 Host proteins described as interactors of geminiviral Rep proteins

Viral protein	Virus	Genus	Interactors	Function of interactor	References
Rep	TGMV, TYLCV, CaLCuV	*Begomovirus*	RBR, ZmRBR	Regulation of cell cycle progression	209,228,229
Rep	TGMV, CbLCV	*Begomovirus*	Histone H3	Core component of nucleosomes	246
Rep	TGMV, CbLCV	*Begomovirus*	GRIK	Geminivirus Rep-Interacting Kinase	246
Rep	TGMV, CbLCV	*Begomovirus*	GRIMP	Geminivirus Rep-Interacting Motor Protein	246
Rep	MYMIV	*Begomovirus*	RPA32	Single-stranded binding protein	232,247
Rep	WDV	*Mastrevirus*	RFC	Sliding clamp loader of PCNA	231
Rep	TYLCSV, IMYMV	*Begomovirus*	PCNA	DNA sliding clamp that enhances DNA replication and DNA repair	233,235
Rep	TGMV, TYLCSV	*Begomovirus*	SCE1	Conjugation of SUMO to target proteins	240–242
Rep	MYMIV	*Begomovirus*	RAD51	Recombination/repair protein	199
Rep	MYMIV	*Begomovirus*	RAD54	Recombination/repair protein	238
Rep	MYMIV	*Begomovirus*	MCM2	Subunit of minichromosome maintenance 2-7 (MCM2-7) helicase that unwinds DNA duplex	236
Rep	MYMIV	*Begomovirus*	MCM3	Subunit of MCM2-7 helicase that unwinds DNA duplex	237
Rep	MYMIV	*Begomovirus*	NAC083	Regulation of gene expression as a transcription factor	248
Rep	ChiLCV	*Begomovirus*	NbHUB1	Histone H2B monoubiquitination	245
Rep	ChiLCV	*Begomovirus*	NbUBC2	Function in monoubiquitination	245
Rep	TYLCV	*Begomovirus*	GRIEP1	EWS-like RNA-binding protein	249
Rep	TYLCV	*Begomovirus*	ALY1	mRNA export	249
RepA	WDV, MSV, BeYDV	*Mastrevirus*	p130, ZmRb	Regulation of cell cycle progression	205–208,214
RepA	WDV	*Mastrevirus*	GRAB	DNA binding and transcription Regulation	250

C3 is encoded in the complementary-sense strand in several geminivirus genera, including *Begomovirus*, *Curtovirus*, *Maldovirus*, *Opunvirus*,

Topocuvirus, and *Turncurtovirus*. It has been shown that C3 promotes viral DNA replication in both protoplast assays and leaf disc assays[226,251]. Further investigations into the interacting partners have provide insights into its potential role in enhancing viral DNA replication. These interactors including C3 itself, Rep, PCNA, and, more recently, POLA2 and POLD2[192,223,226,227,252,253]. In addition to its role in viral DNA replication, C3 may also be involved in cell cycle reprogramming through its interaction with RBR[226,227]. A list of C3 interactors is described in Table 1.5. Of note, members of certain geminivirus genera lack the C3 protein, raising the question of whether alternative molecular mechanisms might be employed by these viruses to replicate in host cells. Investigating these alternative strategies could provide valuable insights into host-virus interactions and geminivirus evolution.

Table 1.5 Host proteins described as interactors of geminiviral C3

Viral protein	Virus	Genus	Interactors	Function of interactor	References
C3	TGMV, TYLCV	*Begomovirus*	pRBR1	Regulation of cell cycle progression	[226,227]
C3	TYLCSV, TYLCV	*Begomovirus*	PCNA	DNA sliding clamp that enhances DNA replication and DNA repair	[227,235]
C3	TLCV, TGMV	*Begomovirus*	SlNAC1	Unknown	[254]
C3	TYLCV, BCTV, TGMV	*Begomovirus*, *Curtovirus*	POLA2	Subunit of DNA Pol α that synthesizes RNA primer	[192]
C3	TYLCV, BCTV, TGMV	*Begomovirus*, *Curtovirus*	POLD2	Subunit of DNA Pol δ that catalyzes the synthesis of lagging-strand	[192]

1.3.3 Host factors required for geminiviral replication

As obligate intracellular parasites, geminiviruses depend entirely on the host replication machinery for their replication and spread. As noted earlier, Rep is an essential protein for viral DNA replication, whereas C3 functions as a replication enhancer. To gain deeper insight into virus-host interactions, numerous studies have investigated potential interactors of both Rep and C3. By using these viral proteins as bait, researchers have sought to identify host factors essential for viral replication, thereby elucidating the composition of the viral replisome. Suyal *et al*. (2013) identified MCM2 as an interactor of Rep from MYMIV through a yeast two-hybrid (Y2H) screen[236]. Further validation using a yeast-based geminivirus DNA replication restoration assay, along with transient replication assays in both *mcm2*-mutant and wild-type (WT) *A. thaliana* plants, confirmed the involvement and critical role of MCM2 subunit of the MCM2-7 complex in MYMIV replication. Interestingly, Rep from the same virus also exhibits DNA helicase activity, raising the question of how the division of labor

is distributed between the host DNA helicase (MCM complex) and MYMIV Rep[97]. A study by Li *et al*. (2019) identified a broader set of host factors required for MYMIV replication using yeast temperature-sensitive (ts) mutants. This approach uncovered 131 yeast genes, categorized into functional groups related to DNA replication, DNA repair and cell cycle regulation. Notably, these included *POLD1* and *POLD2* (catalytic and regulatory subunits of Pol δ) as well as *ORC2* and *ORC5* (components of ORC complex). More recently, Wu *et al*. (2021) identified POLA2 as an interactor of TYLCV C3 through a Y2H screen using a cDNA library from infected tomato plants. This study further confirmed that both Pol α and Pol δ are required for the replication of geminiviruses across different genera. Intriguingly, the replicative Pol ε, which is primarily responsible for leading-strand replication in eukaryotes, was found to play a negative role in viral replication[192]. Shortly after, Wei and Lozano-Durán demonstrated that the primase subunits of Pol α, PRIM1 and PRIM2, also play essential roles in geminiviral replication[191]. Thus far, the essential role of Pol α-primase complex in geminiviral replication has been established.

Despite these advances, a comprehensive understanding of the composition of viral replisome remains elusive. Further investigations are needed to validate the identified host factors and uncover additional host factors involved in geminiviral replication.

1.4 Replication mode at a glance: eukaryotic and geminiviral strategies

Although geminiviral replication, like that of eukaryotic cells, occurs within the cell nucleus of infected hosts and utilized the cell's molecular machinery, there are striking differences between the two replication systems.

One key distinction is the directionality of DNA synthesis: Geminiviruses utilize a unidirectional replication mode, either rolling-circle replication or recombination-dependent replication, in contrast to the bidirectional replication characteristic of eukaryotic chromosomes. This suggests that the canonical eukaryotic mechanism involving the loading and activation of two replicative DNA helicases at replication origins is unlikely to apply to geminiviral replication initiation. Whether the host replicative helicase MCM complex is involved in geminiviral DNA replication remains an open question. If MCM does participate, it is unclear how it would operate within the framework of unidirectional replication. As mentioned earlier, Rep proteins of certain geminiviruses have been shown to possess intrinsic helicase activity, raising the possibility that Rep itself may function as the replicative helicase responsible for unwinding dsDNA during rolling-circle replication. Whether other host replication factors—such as helicase loaders (e.g., ORC and CDC6) and components of the CMG complex (CDC45 and GINS)—are involved in facilitating or regulating geminiviral DNA

replication also remain elusive.

Second, our previous results demonstrated that Pol δ, but not Pol ε, is required for the replication of geminiviruses[192]. This observation raises several important mechanistic questions. Specifically, does geminiviral DNA replication resemble a leading-strand or lagging-strand synthesis model? Is the viral genome synthesized in a continuous or discontinuous manner? Do components specifically involved in lagging-stranded synthesis, such as FEN1 and LIG1, contribute to geminiviral replication? Addressing these questions will not only advance our understanding of geminiviral replication mechanisms but also provide broader insights into the flexibility and adaptation of the eukaryotic DNA replication machinery.

Given the unidirectional nature of geminiviral DNA replication, the molecular mechanism of replication termination in geminiviruses is likely to differ from that in eukaryotic cells, where termination typically occurs upon convergence of two replication forks. In eukaryotic cells, a critical event during replication termination is the ubiquitylation of the MCM7 subunit of the MCM helicase complex. Despite the diversity of E3 ubiquitin ligases among species, the extraction of ubiquitinated MCM7 is consistently mediated by the conserved AAA+ (ATPase associated with various cellular activities) ATPase CDC48. This perspective highlights several key questions concerning the termination phase of geminiviral replication. Which helicase is responsible for unwinding DNA during geminiviral replication? Does this helicase undergo regulatory ubiquitylation similar to MCM7 in eukaryotes? Is CDC48 involved in the replication of geminiviruses? What E3 ligase catalyzes the ubiquitylation of the viral or host helicase? Answering these questions could provide valuable insights into the unique mechanisms governing geminiviral replication termination and reveal potential parallels or divergences with host cellular pathways.

1.5 Aims of this study

Geminiviruses, a family within the phylum *Cressdnaviricota* and one of the largest groups of plant-infecting viruses, pose a significant threat to global agriculture. Among them, TYLCV is one of the most widespread, capable of causing up to 100% yield loss in infected tomato plants. Severe tomato yield losses worldwide have drawn the attention of virologists seeking to identify resistance genes against TYLCV. Several resistance genes, designated *Ty-1*, *Ty-2*, *Ty-3*, *Ty-4*, *Ty-5*, and *Ty-6*, have been identified in wild tomato relatives[255]. Among these, *Ty-1* and *Ty-3* are allelic and encode an RNA-dependent RNA polymerase[256], whereas *Ty-2* encodes a nucleotide-binding leucine-rich repeat protein (NLR) protein[257,258]. *Ty-4* is another TYLCV resistance locus, mapped

to the long arm of chromosome 3 of *Solanum chilense,* which confers less effective resistance to TYLCV compared to *Ty-3*[259]. *Ty-5* gene confers resistance through a loss-of-function mutation in PELOTA, a messenger RNA surveillance factor[260]. While *Ty-1*/*Ty-3*- and *Ty-2*-based resistance has been deployed in agricultural fields, resistance breakdown has been observed in certain cases due to the high mutation rate of geminiviruses[261,262]. Consequently, there is an urgent need to identify additional resistance genes to combat the threat posed by geminiviruses.

Ty-6 was previously identified as a major resistance gene against both monopartite and bipartite begomoviruses and has been shown to complement the resistance conferred by the known *Ty-3* and *Ty-5*[263]. More recently, it was demonstrated that *Ty-6* encodes DNA polymerase delta subunit 1 (POLD1), which provides broad-spectrum resistance to geminiviruses in tomato[264]. In line with this finding, Lim *et al*. revealed that cassava plants carrying cassava mosaic disease (CMD2)-type resistance exhibit strong resistance to cassava mosaic geminiviruses (CMGs), which was attributed to nonsynonymous, single nucleotide polymorphisms (SNPs) in the *POLD1* coding sequence (CDS)[265]. Intriguingly, a recent study suggested that variant alleles of *POLD1* confer resistance to geminiviruses across diverse crop species, including cotton, soybean, and squash in addition to tomato and cassava[266].

These findings highlight a potential strategy for leveraging DNA replication-associated host proteins essential for geminiviral replication to explore natural genetic variation and engineer resistance to geminiviruses. Since viral replication is a fundamental aspect of the viral life cycle, any host factors required for this process represent a promising target for developing resistance against geminiviruses.

In this study, we use Rep, the sole viral protein essential for geminiviral replication, as bait to uncover its proxiome and identify host components required for viral DNA replication. This research will not only provide insights into geminiviral and CRESS-DNA viral replication but also serve as a valuable resource for identifying potential resistance mechanisms against geminivirus-caused diseases. Furthermore, studying viral DNA replication may also contribute to a better understanding of plant or even eukaryotic DNA replication mechanisms.

2. Results

2.1 The proxiome of the viral Rep during infection includes known DNA replication-related proteins

To gain insight into the composition of the geminiviral replisome, we used Rep as a bait to identify host proteins in its proximity during viral infection in *N. benthamiana* plants. This was achieved via the TurboID-based proximity labelling (PL) technique[267] followed by mass spectrometry (MS) to identify biotinylated proteins. Prior to TurboID-based PL, the functionality of Rep fusion proteins was evaluated through a complementation assay. TurboID was fused to the N- and C-termini of Rep proteins from two geminiviruses: TYLCV and AbMV. These fusion proteins were tested for their ability to complement Rep null TYLCV or AbMV mutants in viral replication assays in *N. benthamiana* leaves (Figure 2.1 & 2.2). Notably, we observed that only TYLCV Rep fused to a C-terminal tag, either GFP or TurboID, could complement the defective TYLCV Rep mutant, while N-terminal fusions rendered the protein incompetent to mediate viral replication (Figure 2.1A-B). Additionally, the nuclear localization of Rep[268] was not affected by the addition of the TurboID moiety, as determined upon transient expression in *N. benthamiana* plants and confocal microscopy observation (Figure 2.1C). Interestingly, Rep from AbMV fused to TurboID at either N- and C-termini could complement the defective AbMV Rep mutant to a certain extent (Figure 2.2A). This might be attributed to the longer N-terminal loop in AbMV Rep, which likely provides greater structural flexibility and allows proper folding when a protein is fused to its N-terminus (Figure 2.2B), despite the functional conservation of Rep across geminiviruses[212]. Subsequently, the protein accumulation and biotinylation activity of the functional Rep fusion proteins (Rep-TurboID) were examined in the presence and absence of exogenous biotin. As shown in Figure 2.1D & 2.2B, western blot was performed using an anti-HA antibody to confirm the accumulation of Rep fusion proteins and TurboID (harbouring a 3xHA tag). Strong accumulation of biotinylated proteins was observed upon expression of both Rep fusion proteins and TurboID alone specifically when exogenous biotin was provided. To validate the quality of the samples for MS, small aliquots from various steps were analysed to assess fusion protein level and accumulation of biotinylated proteins (Figure 2.1E & 2.2C). Accumulation of both bait proteins and biotinylated proteins after affinity purification were lower compared to the input and samples after desalting (an essential step to remove plant endogenous biotin), as longer exposure times were required to visualize the bands (Figure 2.1 E & 2.2C). This observation was consistent across all TurboID-based PL experiments aimed at identifying the proxiome of TYLCV Rep and AbMV Rep. The robust wash steps

likely contribute to this decrease by removing non-specific background noise, ultimately enhancing the specificity and reliability of the experimental results.

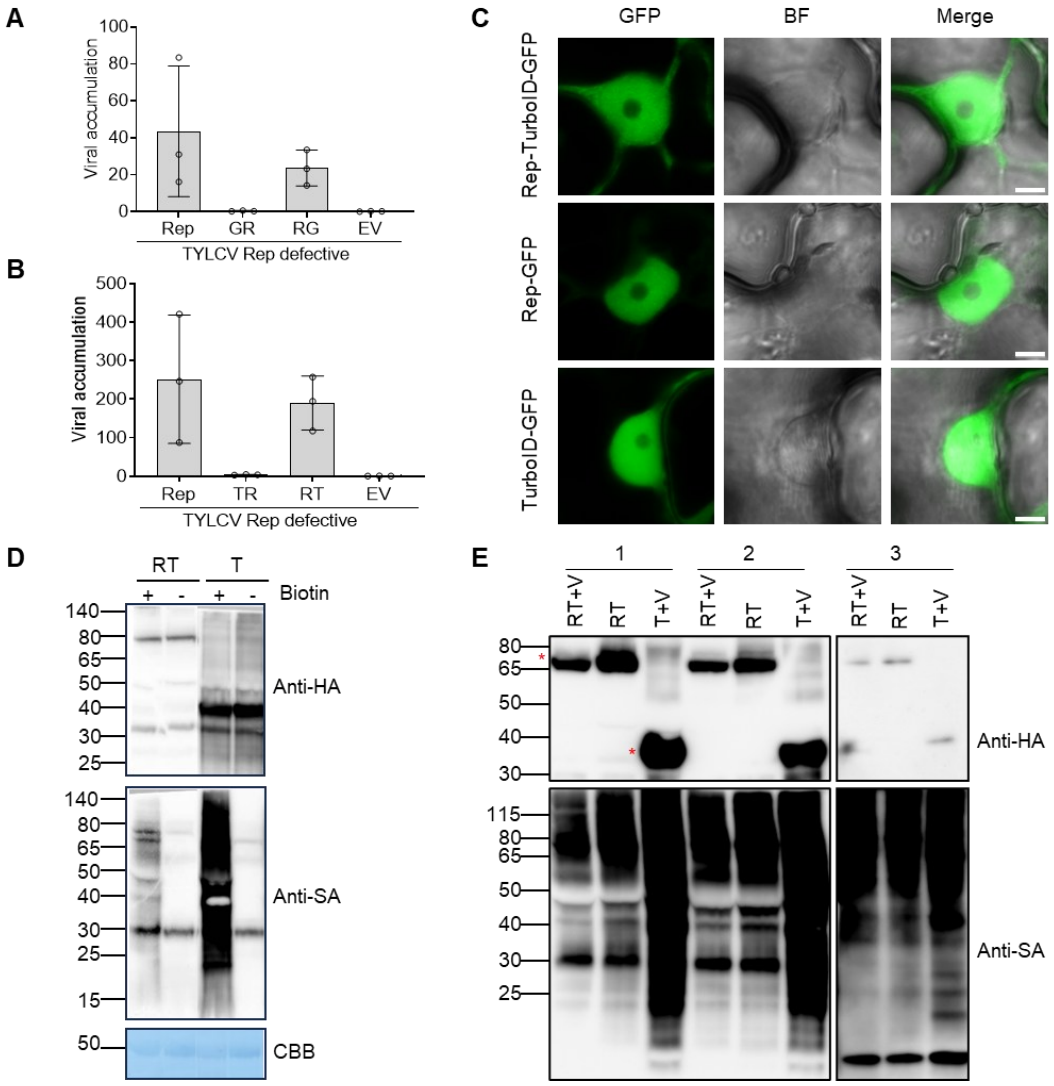


Figure 2.1 TYLCV Rep fused to a C-terminal tag mediates viral genome replication and retains biotin ligase activity. A-B. Rep from TYLCV, fused to a C-terminal tag (A: -GFP; B: -TurboID), retains its ability to mediate viral genome replication. *Agrobacterium* cells containing a TYLCV Rep null mutant were co-infiltrated into *N. benthamiana* leaves with Rep fusion proteins, untagged Rep (positive control), and empty vector (EV, negative control). Samples shown in (A) and (B) were harvested at 2 and 3 days post-infiltration, respectively. Viral DNA accumulation was measured by qPCR with 25S ribosomal DNA interspacer (ITS) an internal reference. GR, GFP-Rep; RG, Rep-GFP; TR, 3xHA TurboID-Rep; RT, Rep-3xHA TurboID. Experiments shown in (A) and (B) were repeated three times and twice, respectively, with similar results. **C.** Rep-TurboID localizes to the nucleoplasm. Transient expression of Rep-TurboID-GFP, Rep-GFP (positive control), and TurboID-GFP (negative control) in *N. benthamiana* plants. Samples were observed at 30 hours post-infiltration (hpi).

Scale bar: 5 μm. BF, brightfield. This experiment was repeated twice with similar results. **D.** Immunoblot analysis of protein accumulation (anti-HA) and biotinylation activity (Streptavidin-HRP, anti-SA) in the presence and absence of exogenous biotin. At 42 hpi, *N. benthamiana* leaves were infiltrated with either 50 μM biotin (+Biotin) or DMSO (-Biotin) and harvested 6 hours later. CBB: Coomassie Brilliant Blue. This experiment was repeated twice with similar results. **E.** Immunoblot analysis of fusion protein level (anti-HA) and biotinylated proteins (anti-SA) in the samples from input lysates (1), after desalting but without affinity purification (2), and after affinity purification using streptavidin beads (3). The position of asterisks indicates the size of RT (~78.0 KDa) and T (3xHA TurboID, ~38.7 KDa), respectively. V, TYLCV Rep null mutant. This experiment was repeated three times with similar results.

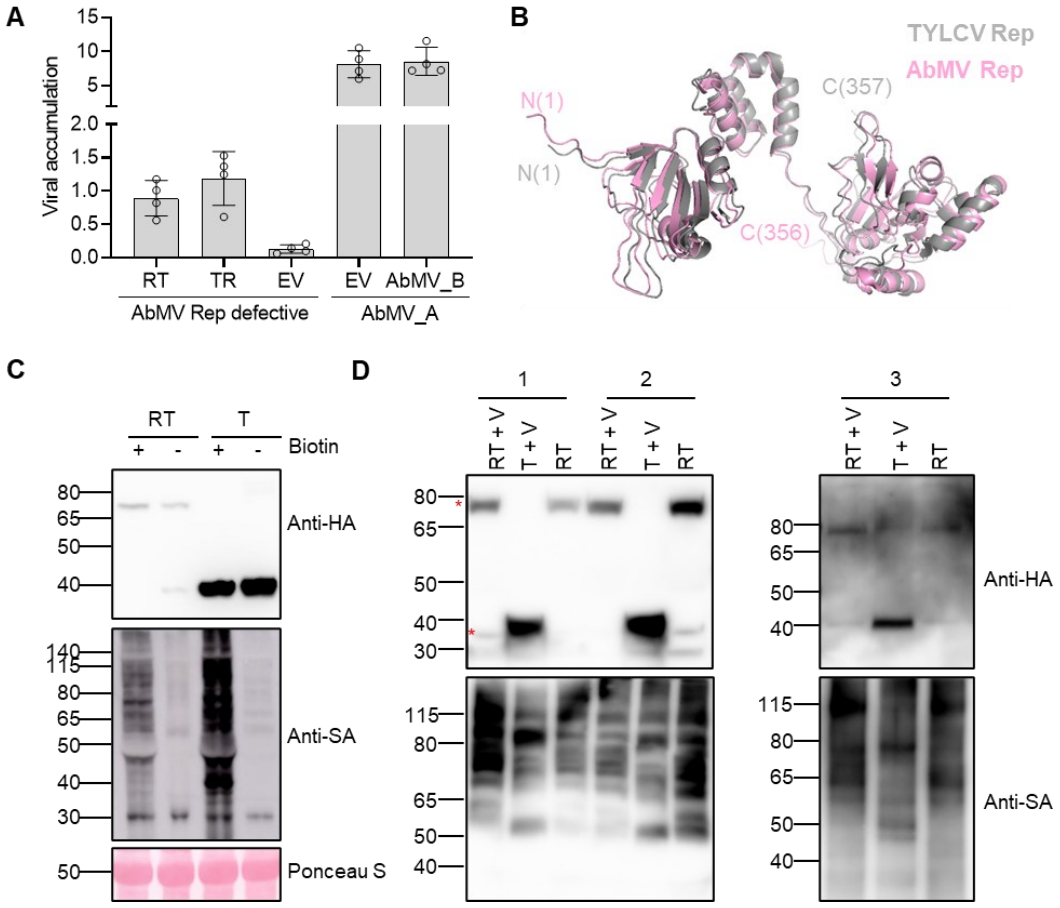


Figure 2.2 AbMV Rep fused to both N- and C-terminal tags mediates viral genome replication and retains biotin ligase activity. A. Rep from AbMV fused to both N- and C-terminal TurboID tag can mediate the replication of viral genome. *Agrobacterium* cells containing AbMV Rep null mutant were co-infiltrated into *N. benthamiana* leaves with Rep fusion proteins, and EV (negative control). Co-infiltration of AbMV A component with EV or AbMV component B were used as the positive controls. Samples were harvested at 3 days post-infiltration. Viral DNA accumulation was

measured by qPCR with 25S ribosomal DNA interspacer (ITS) an internal reference. RT, Rep-3xHA TurboID; TR, 3xHA TurboID-Rep. This experiment was repeated three times with similar results. **B.** Structural overlay of AbMV Rep onto TYLCV Rep, as predicted by AlphaFold 3. TYLCV Rep and AbMV Rep are shown in grey and pink, respectively. **C.** Immunoblot analysis of protein accumulation (anti-HA) and biotinylation activity (Streptavidin-HRP, anti-SA) in the presence and absence of exogenous biotin. At 42 hours post-infiltration, *N. benthamiana* leaves were infiltrated with either 50 μM biotin (+Biotin) or DMSO (-Biotin) and harvested 6 hours later. Ponceau S, ponceau staining. **D.** Immunoblot analysis of fusion protein level (anti-HA) and biotinylated proteins (anti-SA) in the samples from input lysates (1), after desalting but without affinity purification (2), and after affinity purification using streptavidin beads (3). The position of asterisks indicates the size of RT (~78.0 KDa) and T (3xHA TurboID, ~38.7 KDa), respectively. V, AbMV Rep null mutant. The biological replicates shown in panel D were independently performed by the author and PhD student Hua Wei. This experiment was repeated three times with similar results.

To investigate the functions of geminiviral Rep proteins that are dependent on the presence of viral DNA versus those that occur independently, we compared the proxiome of Rep in the presence (Rep-TurboID + TYLCV/AbMV Rep defective vs TurboID + TYLCV/AbMV Rep defective) and absence (Rep-TurboID alone vs TurboID + TYLCV/AbMV Rep defective) of the virus. Specifically, two biological replicates with TYLCV Rep and three replicates with AbMV Rep were performed in the absence of viral infection, while three replicates expressing both Rep proteins were conducted in the presence of their respective virus. To minimize noise, MS data filtering was performed using the following criteria: (1) protein and peptide identification confidence was set at 95%; (2) the number of peptides in the experimental sample was at least twice that observed in the TurboID negative control. The filtered data were then visualized using Venn diagrams, as shown in Figure 2.3. Thirteen proteins were identified across all three replicates in which TYLCV Rep was expressed in the presence of the virus (Figure 2.3A). By comparing the proxiome of Rep in the presence and absence of the virus, we observed several known replication-associated proteins among those labelled by TYLCV Rep in both conditions. These include the large subunit of replication factor C (RFC1), the catalytic subunit (POLD1) and regulatory subunit (POLD3) of the Pol δ holoenzyme, DNA topoisomerase I (TOPO1), and single-stranded binding protein RPA70A (Figure 2.3B). Three viral proteins, C2, V2, and CP, appeared in the presence of virus, which is in agreement with a previous study demonstrating the interaction between viral proteins[225]. Replication factor C is a heteropentameric complex composed of RFC1-5. RFC2 and RFC4 were surprisingly only present in the absence of virus (Figure 2.3B). Similar results were also observed in the

subset of AbMV Rep-labelled proteins (Figure 2.3C). Five subunits of the RFC complex (RFC1-5), POLD3, and TOPO1 were labelled by AbMV Rep in the presence and absence of virus. However, POLD1 was only labelled by AbMV Rep in the virus-free conditions. An illustration of the overlapping Rep-labelled proteins between TYLCV and AbMV Rep in the presence of their respective virus is shown in Figure 2.3D. Overall, our PL-MS data suggest that TurboID-based PL is a powerful tool to gain novel insights into viral replication. For the first time, this approach has enabled the identification of a number of known DNA replication-associated factors in proximity to the Rep protein from two distinct geminiviruses, suggesting the conservation of geminiviral replication mechanisms in both monopartite and bipartite begomoviruses and opening up new avenues of research on geminiviral replication. In addition, consistent with previous studies, our results indicate that geminiviral Rep alone is sufficient to induce the host replication machinery, as it labels certain replication-associated factors in the absence of the virus.

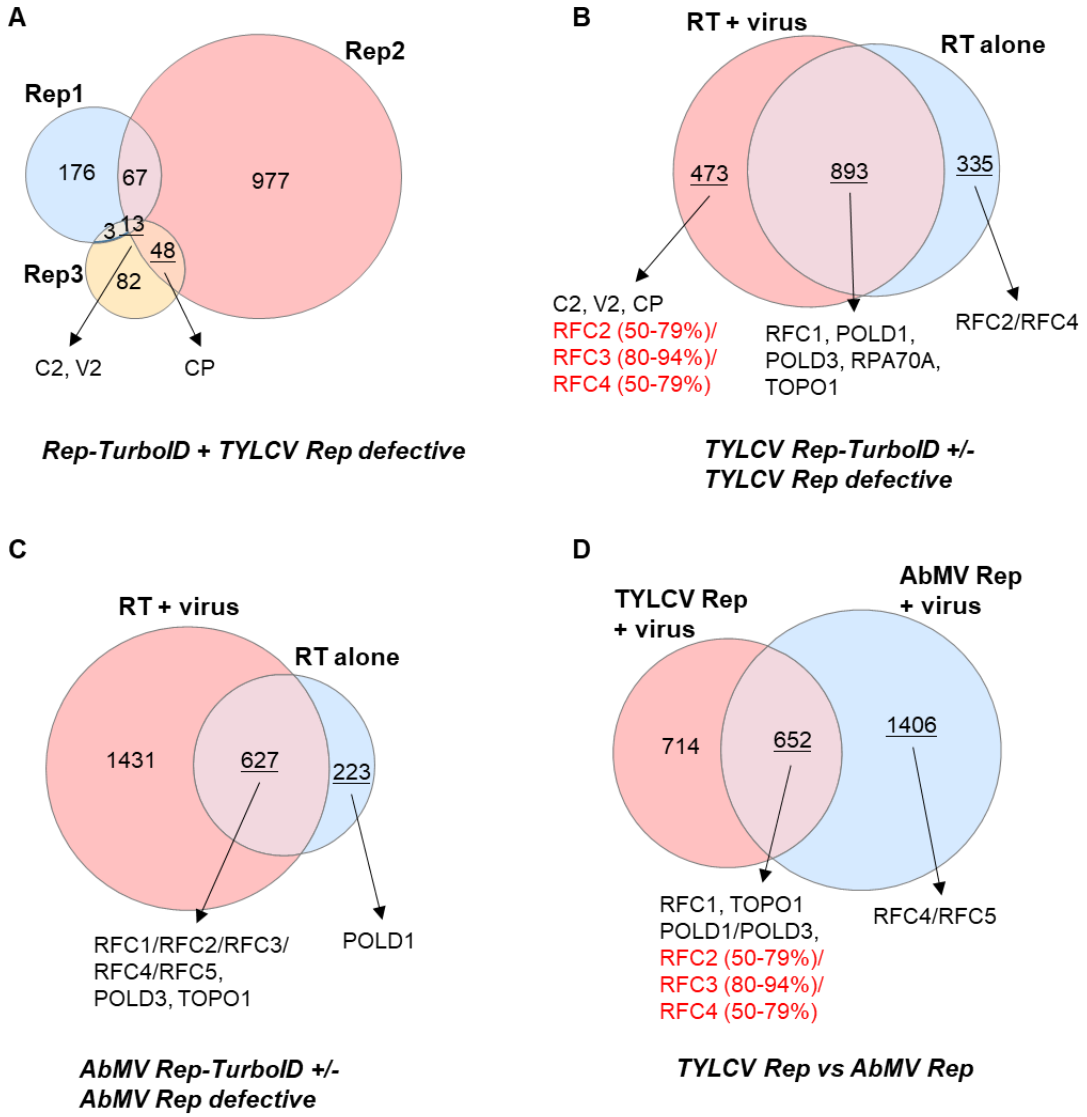


Figure 2.3 Venn diagrams illustrate the proteins labelled by TYLCV Rep and AbMV Rep. A. Venn diagram shows the overlap of Rep-labelled proteins across three independent biological replicates (Rep1-3) of TYLCV Rep in the presence of virus infection. **B-C.** Venn diagram illustrates the overlap of Rep-labelled proteins in the presence and absence of viral infection for TYLCV (B) and AbMV (C). **D.** Venn diagram shows the overlap between proteins labelled by TYLCV Rep and AbMV Rep in the presence of virus. Arrows indicate DNA replication-related proteins specifically labelled under different experimental conditions. Proteins shown in black meet the protein and peptide identification threshold of over than 95%, while those labelled in red fall below this confidence threshold, as indicated.

To further investigate the role of DNA replication-associated proteins labelled by TYLCV Rep, AbMV Rep, or both, in geminiviral replication, we selected RFC1, POLD3, POLD1, TOPO1, RPA70A, and PCNA for further analysis.

PCNA functions as a processivity factor for Pol δ and Pol ε in eukaryotes. Notably, previous studies have demonstrated interactions between geminiviral Rep proteins and PCNA from both plant and insect hosts[233,252,269], however, its functional relevance has not yet been experimentally validated. PCNA was not identified as being labelled by either TYLCV or AbMV Rep proteins in our PL assays. However, we found that TYLCV C3 labels PCNA in PL assays (unpublished). Because of this, and considering that the physical interaction between Rep and PCNA has already been demonstrated, we included this plant protein in this study. All replication-associated proteins selected for further investigation are listed in Table 2.1.

Table 2.1 DNA replication-related proteins selected from PL experiments for further study

Name	Gene identifier	Annotation	Source
RFC1	Niben101Scf01382g04008	DNA replication factor C subunit 1	TYLCV Rep & AbMV Rep PL experiments
POLD3	Niben101Scf00160g10002	DNA polymerase δ subunit 3	TYLCV Rep & AbMV Rep PL experiments
TOPO1	NbS00018030g0025	DNA topoisomerase I	TYLCV Rep & AbMV Rep PL experiments
POLD1	Niben101Scf02230g03027	DNA polymerase δ subunit 1	TYLCV Rep PL experiments
RPA70	Niben101Scf00567g04007	Replication protein A 70 kDa subunit	TYLCV Rep PL experiments
PCNA	Niben101Scf10384g02008	Proliferating cell nuclear antigen	TYLCV C3 PL experiments

Since geminiviral replication occurs in the cell nucleus, we first examined the subcellular localization of the selected replication-associated factors and their co-localization with Rep upon transient expression in *N. benthamiana* plants. Additionally, we assessed whether the presence of the virus influences protein localization patterns, as shown in Figure 2.4. As expected, all selected DNA replication-associated proteins localized to the nucleus. Notably, TOPO1 was specifically and consistently localized in the nucleolus under all tested conditions, with only a weak signal observed in the nucleoplasm. A similar localization pattern was also observed for RFC1, but its nucleolar accumulation was less pronounced, and the nucleoplasmic signal was more intense compared to that of TOPO1. RPA70A accumulated in nuclear speckles in all cases. Similar to GFP, PCNA was detected in both the nucleus and the cytoplasm under all conditions. In the presence of Rep or the virus (TYLCV), an increased number of nuclear speckles was observed for RPA70A; however, the localization patterns of the remaining proteins were not affected by either Rep or the virus.

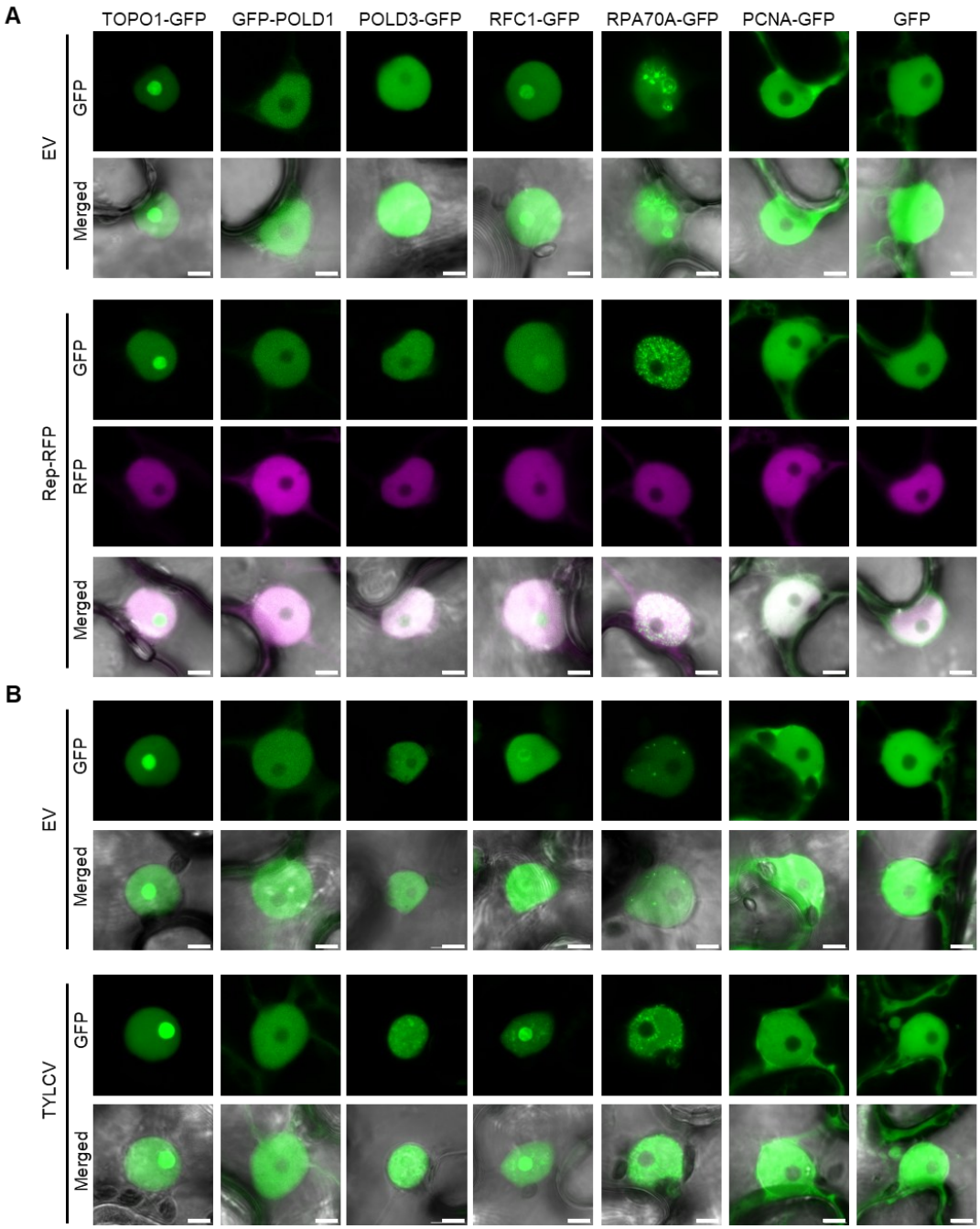


Figure 2.4 Subcellular localization of selected DNA replication-related proteins. A-B. Transient expression of TOPO1-GFP, GFP-POLD1, POLD3-GFP, RFC1-GFP, RPA70A-GFP, PCNA-GFP, and GFP in *N. benthamiana* leaves, either in the absence (EV) or presence of Rep-RFP (A) or TYLCV (B). *Agrobacterium* cells containing the respective binary vectors were co-infiltrated at a 1:1 ratio; GFP was used as a negative control; confocal images were taken at 30 hours post-infiltration. Scale bar: 5 μm. This experiment was repeated twice with similar results.

TurboID-based PL enables the identification of proteins in proximity to a protein of interest. To determine whether Rep physically associates to these DNA

replication-associated proteins, co-immunoprecipitation (co-IP) of Rep with all selected proteins was performed, as shown in Figure 2.5. Given the nuclear localization of the selected proteins, nuclei extraction was conducted prior to protein extraction to enrich protein accumulation. The results indicate that Rep associates, either directly or indirectly, with all of the selected proteins in the virus-free context.

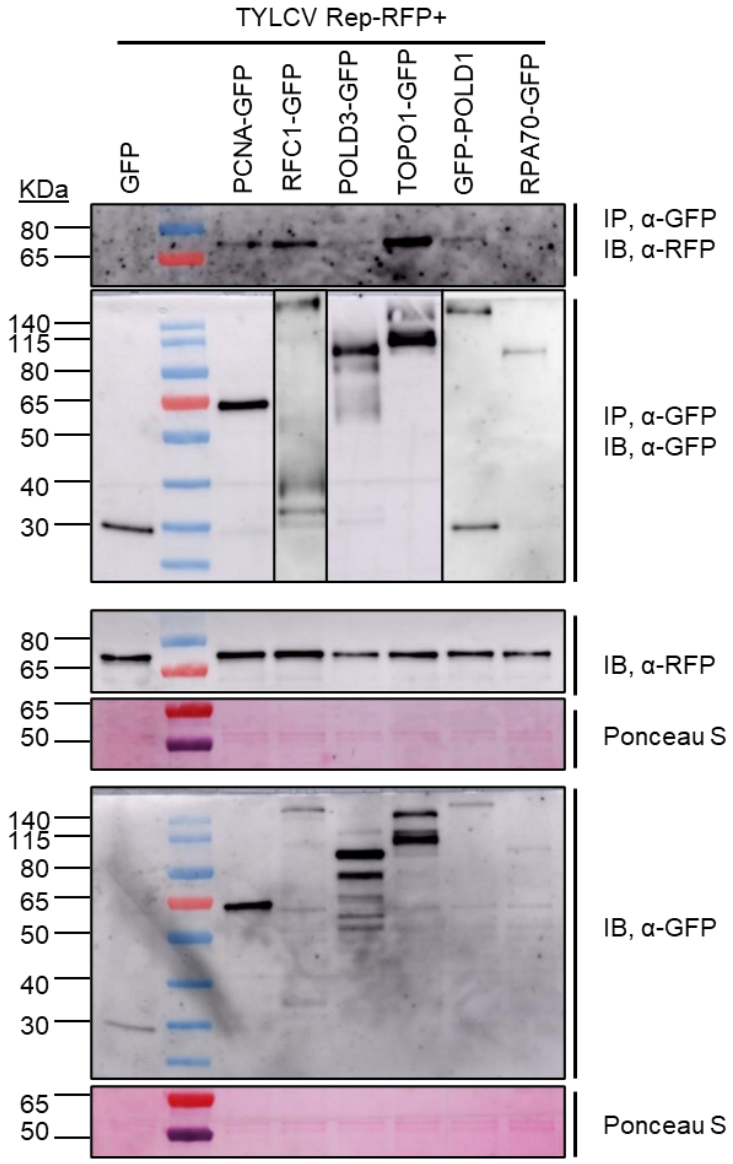


Figure 2.5 Rep associates with selected replication-associated factors by co-IP. Rep-RFP co-immunoprecipitates with RFC1-GFP, POLD3-GFP, TOPO1-GFP, GFP-POLD1, RPA70A-GFP, and PCNA-GFP (positive control), but not with GFP (negative control), upon transient expression in *N. benthamiana*. IP, immunoprecipitate; IB, immunoblotting; Ponceau S, ponceau staining. Molecular weights are indicated on the left. The predicted protein sizes are as follows: RFC1-GFP, ~134 kDa; POLD3-GFP,

~85 kDa; TOPO1-GFP, ~125 kDa; GFP-POLD1, ~148 kDa; RPA70A-GFP, ~99 kDa; PCNA-GFP, ~58 kDa; GFP, ~28 kDa. This experiment was repeated three times with similar results.

2.2 Rep proximal proteins are components of the viral replisome and required for viral DNA replication

DNA replication-associated proteins are typically essential for host cell function. Since a viable host is the prerequisite for studying the role of host factors in the infection cycle of obligate parasitic viruses, gene knockdown serves as an ideal approach to assess the function of host factors labelled by Rep in geminiviral replication. In this study, a reverse genetics technique leading to knockdown of the desired gene, virus-induced gene silencing (VIGS), mediated by tobacco rattle virus (TRV), was employed to achieve the targeted gene silencing. Knockdown of *Phytoene desaturase* (*PDS*) was utilized as a positive control in all VIGS experiments due to its distinct photobleaching phenotype, while either the TRV2 empty vector (TRV-EV) or a TRV2 vector containing the exogenous gene encoding ß-glucuronidase (TRV-*GUS*) were used as negative controls. The developmental phenotype of gene-silenced *N. benthamiana* plants was observed at 14 days post-TRV inoculation, as shown in the left panel of Figure 2.6. However, an exception was made for *TOPO1*-silenced plants (right panel), which were harvested at an earlier time point (7 days post-TRV inoculation) since silencing of this gene led to plant death shortly after (see below). The height of gene-silenced plants (*RFC1*-, *POLD3*-, *POLD1*-, *PCNA*-) was lower than that of control plants, with *PCNA*-silenced plants showing the greatest reduction. Moreover, a thick-leaf phenotype was commonly observed in *N. benthamiana* plants silenced for *RFC1*, *POLD1*, and *PCNA*, but not in those silenced for *POLD3*, which displayed the mildest phenotype among all tested genes.

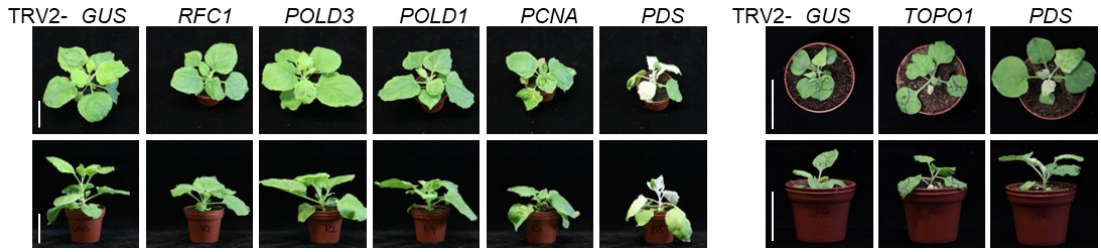


Figure 2.6 Developmental phenotypes of gene-silenced *N. benthamiana* plants. *Agrobacterium* cells containing TRV1 and TRV2-RFC1, TRV2-POLD3, TRV2-POLD1, TRV2-PCNA, and TRV2-TOPO1 were mixed at a 1:1 ratio and inoculated into 2-week-old *N. benthamiana* seedlings. Plants infiltrated with TRV2-GUS and TRV2-NbPDS were used as negative and positive controls, respectively. Scale bar: 5 cm. This

experiment was repeated multiple times with similar results.

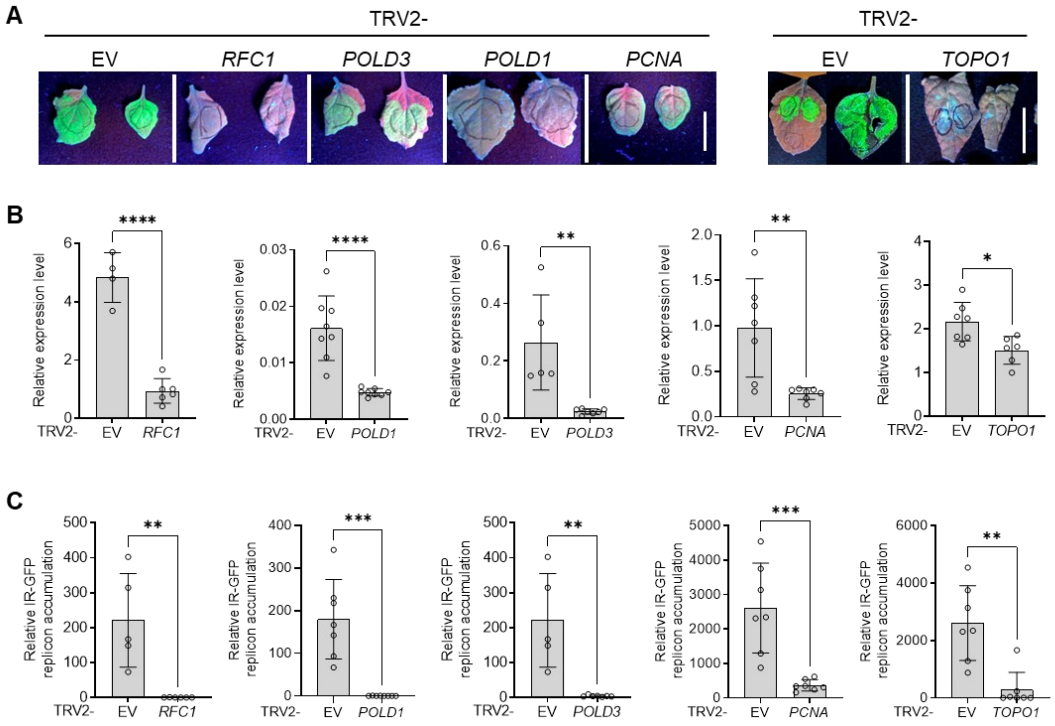


Figure 2.7 Silencing of selected replication-association factors affect Rep-mediated DNA replication in the 2IR-GFP reporter plants. A. GFP signal was observed in *RFC1*-, *POLD3*-, *POLD1*-, *PCNA*-, and *TOPO1*-silenced 2IR-GFP transgenic *N. benthamiana* plants following agroinfiltration with Rep-RFP. Scale bar: 5 cm. **B-C.** Silencing efficiency (B) of *RFC1*, *POLD3*, *POLD1*, *PCNA*, and *TOPO1* in 2IR-GFP transgenic *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and IR-GFP replicon accumulation (C), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TRV1 and TRV2-RFC1, TRV2-POLD3, TRV2-POLD1, TRV2-PCNA, and TRV2-TOPO1 were mixed at a 1:1 ratio and inoculated into 2-week-old *N. benthamiana* seedlings. Plants were then agroinfiltrated with Rep-RFP at 14 days post-TRV inoculation (dpi), except for *TOPO1*-silenced plants, which were infiltrated at 7 dpi. All samples were harvested at 2 days post-infiltration. Data are presented relative to TRV2-EV plants, with values shown as the mean of eight individual plants, and error bars indicating standard deviation (SD). Asterisks indicate statistically significant differences based on Student's t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). This experiment was repeated twice with similar results.

The 2IR-GFP reporter system described previously[249,270,271] was used to monitor Rep-mediated DNA replication in a non-destructive manner. In this system, 2IR-GFP transgenic *N. benthamiana* plants contain a GFP expression cassette flanked by direct repeats of the IR of TYLCV. During viral infection or

transient expression of TYLCV Rep, Rep site-specifically recognizes and nicks the conserved nonanucleotide sequence (TAATATTAC) within the IR, initiating rolling-circle replication and inducing the production of circular extrachromosomal IR-GFP replicon, which can be quantified by qPCR (see Figure 2.7C). The *GFP* gene in the IR-GFP replicon is driven by the cauliflower mosaic virus (CaMV) 35S promoter, which leads to production of GFP that can be visualized under a UV lamp (see Figure 2.7A). By combining the reporter system with VIGS, we can identify essential proteins required for geminiviral Rep-mediated DNA replication by detecting IR-GFP replicon and monitoring GFP accumulation in gene-silenced plants compared to control plants. The silencing efficiency of each target gene was examined by qRT-PCR (Figure 2.7B). GFP signal was significantly reduced in *RFC1*-, *POLD3*-, *POLD1*-, *PCNA*-, and *TOPO1*-silenced plants compared to the negative control (TRV2-EV). Consistent with the GFP signal observed in Figure 2.7A, the relative accumulation of IR-GFP replicon was largely abolished in the gene-silenced samples compared to that in TRV2-EV control. Altogether, these results demonstrated that these replication-associated factors labelled by Rep play a crucial role in Rep-mediated DNA replication, as evidenced by GFP signal and extrachromosomal replicon monitoring in the reporter system.

Local and systemic viral infection assays were described by Wu *et al*. (2019)[272]; in local infections, viral accumulation mainly depends on replication, whereas in systemic infections, it reflects not only replication but also viral movement, suppression of defence, and other factors. To investigate the role of selected replication-associated proteins in geminiviral replication and infection, we performed these experiments in *N. benthamiana* wild-type (WT) plants (Figure 2.8 and Figure 2.9). At 14 days post-TRV inoculation, *Agrobacterium*-mediated inoculation of the TYLCV infectious clone was conducted in gene-silenced *N. benthamiana* plants to assess the effect on local infection (Figure 2.8). Quantification of silencing efficiency and viral accumulation in *RFC*-, *POLD3*-, *POLD1*-, *PCNA*-, and *TOPO1*-silenced plants is shown in Figure 2.8A and Figure 2.8B, respectively. The expression level of all target genes was significantly reduced in gene-silenced samples compared to those in control plants. A distinct reduction in viral accumulation was consistently observed in all gene-silenced *N. benthamiana* plants, as shown in Figure 2.8B. This reduction positively correlates with the effect of knocking down these selected replication-associated genes on GFP signal in the 2IR-GFP reporter system, demonstrating their essential role in geminiviral replication. The requirement of POLD1 for geminiviral replication is consistent with our previous results[273] demonstrating the essential role of Pol δ in geminiviral replication.

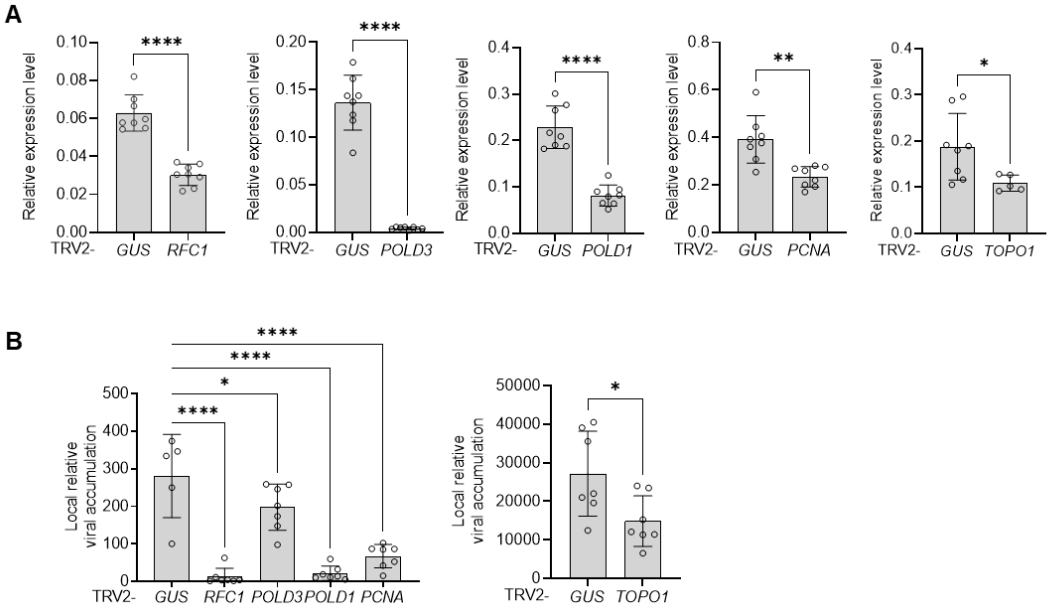


Figure 2.8 Selected replication-association factors are required for TYLCV local infection. A-B. Silencing efficiency (A) of *RFC1*, *POLD3*, *POLD1*, *PCNA*, and *TOPO1* in *N. benthamiana*, measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (B), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TRV1 and TRV2-RFC1, TRV2-POLD3, TRV2-POLD1, TRV2-PCNA, and TRV2-TOPO1 were mixed at a 1:1 ratio and inoculated in 2-week-old *N. benthamiana* seedlings. Plants were then agroinfiltrated with the TYLCV infectious clone at 14 days post-TRV inoculation (dpi), except for *TOPO1*-silenced samples, which were infiltrated at 7 dpi. All samples were harvested at 3 days post-infiltration. The biological replicates shown in panels A-B were independently performed by the author and Dr. Delphine Pott. Data are presented relative to TRV2-GUS plants, with values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student's t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). This experiment was repeated three times with similar results.

Agrobacterium cells containing VIGS constructs were co-infiltrated with either the TYLCV infectious clone or an EV control into WT *N. benthamiana* seedlings to establish a systemic infection. The symptoms observed in *RFC1*-, *POLD3*-, *POLD1*-, *PCNA*-, and *TOPO1*-silenced *N. benthamiana* inoculated with TYLCV are presented in Figure 2.9A. As previously described, typical geminiviral symptoms, such as stunted growth and leaf curling, were clearly observed in control plants (Figure 2.9A). The relative gene expression levels of each target gene in control and gene-silenced plants was evaluated and is shown in Figure 2.9B. Consistent with the results from the 2IR-GFP reporter system and viral

local infection assays, the systemic viral accumulation was dramatically reduced in gene-silenced plants (Figure 2.9C). Collectively, our data demonstrate that all selected DNA replication-associated proteins are essential not only for geminiviral genome replication, but also, as expected, for the infection process.

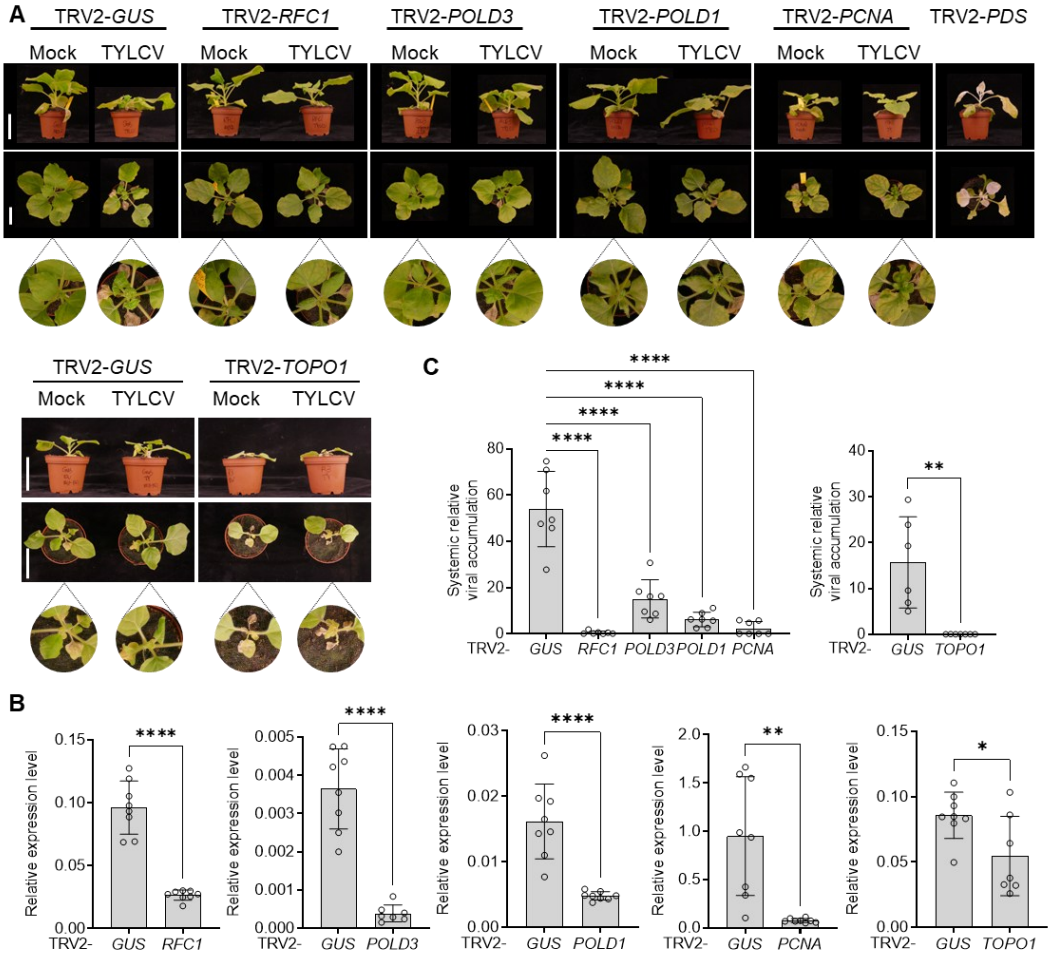


Figure 2.9 Selected replication-association factors are required for TYLCV systemic infection. A-C. Symptom observation (A, scale bar: 5 cm) in *RFC1*-, *POLD3*-, *POLD1*-, *PCNA*-, and *TOPO1*-silenced *N. benthamiana* plants inoculated with the TYLCV infectious clone. Silencing efficiency (B), measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (C), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TYLCV infectious clone, TRV1 and TRV2-RFC1, TRV2-POLD3, TRV2-POLD1, TRV2-PCNA, and TRV2-TOPO1 were mixed at a 1:1:1 ratio and inoculated in 2-week-old *N. benthamiana* seedlings. Samples were harvested at 14 days post-TRV inoculation (dpi), except for *TOPO1*-silenced samples, which were harvested at 9 dpi. Plants infiltrated with TRV2-GUS and TRV2-NbPDS were used as negative and positive controls, respectively. Data are presented relative to TRV2-GUS plants, with

values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student's t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). This experiment was repeated three times with similar results.

To identify whether these replication-associated factors are components of the viral replisome, we performed chromatin immunoprecipitation (ChIP)-qPCR assays to assess their binding to the viral DNA genome. Endogenous proteins may compete with exogenously expressed ones for interaction with the viral genome or other viral components. To minimize these effects, silencing-resistant (SR) versions of the target genes, with modifications of the third nucleotide position that do not alter the encoded protein sequence, and harboring a FLAG tag, were introduced into the corresponding gene-silenced *N. benthamiana* plants. Prior to the complementation assay, the protein accumulation of each replication-associated candidate encoded by their SR version fused to a FLAG tag was detected by western blot using an anti-FLAG antibody (Figure 2.10A). To further assess the functionality of the SR-encoded fusion proteins in viral DNA replication, we co-infiltrated Rep and either the SR version or the EV into gene-silenced 2IR-GFP transgenic *N. benthamiana* leaves at 14 days post-TRV inoculation. GFP signal was observed under UV light after an additional two days. As shown in Figure 2.10B, the leaf patches co-infiltrated with Rep and the SR version exhibited a significantly stronger GFP signal compared to the leaf patches containing Rep and EV in *RFC1*-, *POLD3*-, *POLD1*-, *PCNA*-, and *TOPO1*-silenced 2IR-GFP transgenic plants. These results indicate successful complementation by the fusion proteins encoded by the SR gene versions in the gene-silenced *N. benthamiana* background, demonstrating that the negative effect observed on viral DNA replication (Figure 2.7, 2.8, and 2.9) is indeed specifically due to the knockdown of the respective target genes. Using the same methodology described in Figure 2.10B, we performed ChIP-qPCR to assess the binding of these replication-associated proteins to the IR-GFP replicon in 2IR-GFP transgenic plants. The amplified sequences used for qPCR quantification are shown in the lower panel of Figure 2.10C. Our results indicated that all six replication-associated proteins, including RFC1, POLD3, POLD1, PCNA, TOPO1, and RPA70A, can bind to the IR-GFP replicon. Notably, the ChIP assay for RPA70 was not performed by the aforementioned strategy, as this gene could not be successfully silenced via VIGS, likely due to the presence of multiple paralogs in *N. benthamiana*. Therefore, transient expression of RPA70A-FLAG was directly performed instead. Our ChIP-qPCR results assessing the binding of DNA replication-associated proteins to the viral genome upon viral infection (Figure 2.10D) are consistent with results from the 2IR-GFP reporter system, confirming that all

selected replication-associated proteins bind to the viral genome. Next, using PCNA as proof of concept, we aimed to access whether its binding to the IR-GFP replicon in the 2IR-GFP reporter system or the viral genome during infection is dependent on the presence of Rep. To this end, we designed an experiment in which *N. benthamiana* leaves were agroinfiltrated with IR-GFP (a construct where the *GFP* gene was placed downstream of the TYLCV IR, which contains the viral replication origin), PCNA-FLAG, and either Rep-RFP or RFP as a control. We then analysed PCNA binding to the IR region by ChIP-qPCR assay (Figure 2.10E). Interestingly, PCNA was found to strongly bind to the IR in a Rep-dependent manner, indicated by the high enrichment of IR sequence in PCNA-FLAG samples, but not in the IgG control, in the presence of Rep; on the contrary, no detectable binding was observed in the absence of Rep, neither in the PCNA-FLAG samples nor in the IgG control. Notably, the IR-GFP construct used here contains a single IR, in contrast to the two IRs present in the 2IR-GFP reporter system. The single IR prevents the substantial production of IR-GFP replicon, as incomplete IR structure fails to support Rep-mediated rolling-circle replication. Therefore, the amount of IR available is not affected by the presence or absence of Rep. In summary, our results demonstrate that Rep-labelled proteins required for viral replication and infection are likely components of the viral replisome, and their association with the viral genome is dependent on the presence of the Rep protein.

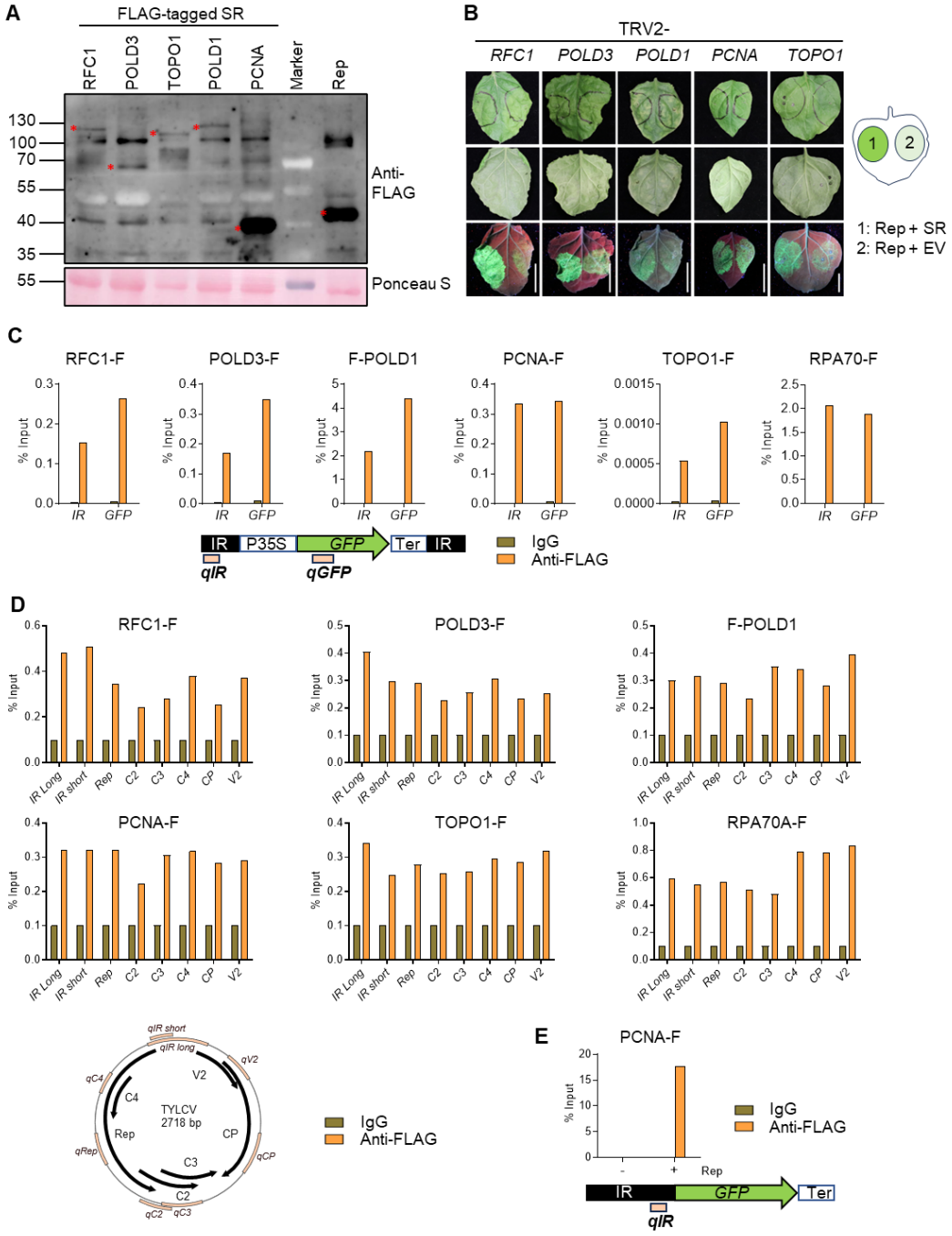


Figure 2.10 Selected replication-associated factors are likely components of the geminiviral replisome. A. Western blot analysis of the protein accumulation of silencing-resistant (SR)-encoded replication-associated proteins. Rep-FLAG was used as the positive control. Ponceau S, ponceau staining. The predicted protein sizes are as follows: RFC1-FLAG-SR, ~106 kDa; POLD3-FLAG-SR, ~57 kDa; TOPO1-FLAG-SR, ~97 kDa; FLAG-POLD1-SR, ~120 kDa; PCNA-FLAG-SR, ~30 kDa; Rep-FLAG, ~40 kDa. Asterisks with red color indicate the protein size. **B.** Functional analysis of fusion proteins encoded by the SR gene versions in *RFC1*-, *POLD3*-, *POLD1*-, *PCNA*-,

and *TOPO1*-silencing *N. benthamiana* plants in the 2IR-GFP reporter system via complementation assay. Agroinfiltration of FLAG-tagged SR-encoded proteins or EV containing FLAG tag and Rep was performed in gene-silenced 2IR-GFP transgenic *N. benthamiana* plants. GFP signal was observed at 2 days post-infiltration under UV-lamp. Scale bar: 2 cm. **C-D.** RFC1, POLD3, POLD1, PCNA, TOPO1, and RPA70A bind to the IR-GFP replicon in the 2IR-GFP reporter system (C) and viral genome upon viral infection (D). Co-infiltration of *Agrobacterium* cells containing FLAG-tagged SR-encoded proteins and Rep or the TYLCV infectious clone was performed in gene-silenced 2IR-GFP transgenic (C) or wild-type *N. benthamiana* plants (D), respectively, with the exception of RPA70A (co-expressed RPA70A-FLAG and Rep or the TYLCV infectious clone in 4-week-old 2IR-GFP transgenic or wild-type *N. benthamiana* plants, respectively). All experiments were repeated at least twice with similar results. **E.** PCNA binds the geminiviral IR region, which contains the origin of replication, in a Rep-dependent manner. Co-infiltration of PCNA-FLAG (wild-type version), Rep-RFP or RFP, and IR-GFP was performed in 4-week-old WT *N. benthamiana* plants. All experiments were repeated at least five times with similar results. Genomic locations of the IR-GFP replicon (C), the viral genome (D), and the IR region (E) amplified in the ChIP-qPCR assays are indicated in orange at the bottom of panels C, D, and E, respectively. The brown and orange bars in the graphs represent IgG (negative control) and anti-FLAG, respectively.

2.3 The viral genome follows leading-strand, not lagging-strand, DNA replication with swapped DNA polymerases

Notably, all components identified in the previous sections as required for viral replication are involved in both leading- and lagging-strand replication at the eukaryotic replication fork—namely PCNA, RFC, and RPA—while TOPO1 functions to relieve torsional stress ahead of the fork. An exception is Pol δ, which is primarily responsible for lagging-strand synthesis. In line with this work, Wu *et al*. (2021) have proven that Pol δ, but not Pol ε, is required for geminiviral genome replication[273]. Although an increasing number of studies have highlighted the role of Pol δ in initiating leading-strand synthesis[153,274], as well as its involvement in leading-strand replication during double-strand break repair[154,155], it is generally considered to be primarily responsible for lagging-strand synthesis[153,275]. Given that lagging-strand synthesis occurs in a discontinuous manner, we sought to determine whether geminiviral DNA replication, also dependent on Pol δ, proceeds in a continuous or discontinuous fashion. To address this, we examined the roles of FEN1 and LIG1—two key factors involved in lagging-strand synthesis—in geminiviral genome replication. Each paralog of *FEN1* and *LIG1* from *N. benthamiana*, as well as their expression levels in leaf tissue in TYLCV local infection or control samples (Wu

et al., 2019), is shown in Table 2.2. The VIGS fragments targeting *FEN1* and *LIG1* are shown in Figure 2.11A. The VIGS fragment for *FEN1* can theoretically target both *FEN1.1* and *FEN1.2*, as the two paralogs differ by only three single-nucleotide polymorphisms (SNPs). qPCR primers were designed to assess the expression levels of both *FEN1* paralogs. Similarly, the VIGS fragment for *LIG1* fully targets the *LIG1.1* paralog, but differs from the *LIG1.2* sequence by ten SNPs. qPCR primers were used to quantify the expression levels of both *LIG1.1* and *LIG1.2* individually. The effects of knocking down either *FEN1* or *LIG1* on geminiviral replication were examined using both the 2IR-GFP reporter system and viral local infection assays. *FEN1*- and *LIG1*-silenced *N. benthamiana* plants exhibited slight phenotypic differences compared to control plants, with *FEN1*-silenced plants appearing slightly larger and *LIG1*-silenced plants slightly smaller (Figure 2.11B). However, in the reporter system, GFP fluorescence observed under a UV lamp in gene-silenced 2IR-GFP transgenic plants showed no discernible differences compared to the control, indicating that neither FEN1 nor LIG1 plays a role in Rep-mediated DNA replication. The expression levels of *FEN1* and *LIG1* in both the reporter system and viral local infection assays are presented in Figure 2.11C and Figure 2.11E, respectively. Notably, silencing of *FEN1* and *LIG1* did not impact either IR-GFP replicon accumulation (Figure 2.11D) or viral accumulation (Figure 2.11F). Taken together, these results suggest that key host factors involved in lagging-strand synthesis are not required for geminiviral replication, providing the first experimental evidence supporting a continuous replication mechanism in rolling-circle replication of geminiviral genomes by recruiting the components mediating the leading-strand synthesis, but with a swapped DNA polymerase.

Table 2.2 Selected host factors involved in eukaryotic lagging-strand replication and the expression levels of each orthologue coding gene in *N. benthamiana*. Expression data are taken from Wu *et al*. (2019) and represent transcription abundance measured in RPM (reads per million). Data were obtained from TYLCV locally-infected samples, with agroinfiltration of an empty vector (EV) serving as the negative control.

Name	Gene identifier	Annotation	Source	EV	TYLCV
FEN1.1	Niben101Scf05932g05010	Flap endonuclease 1	TYLCV Rep & AbMV Rep PL experiments	202	191
FEN1.2	Niben101Scf02486g03012		AbMV Rep PL experiments	147	107
LIG1.1	Niben101Scf19975g01014	DNA ligase 1	Functional information	360	503
LIG1.2	Niben101Scf00436g03006			49	84

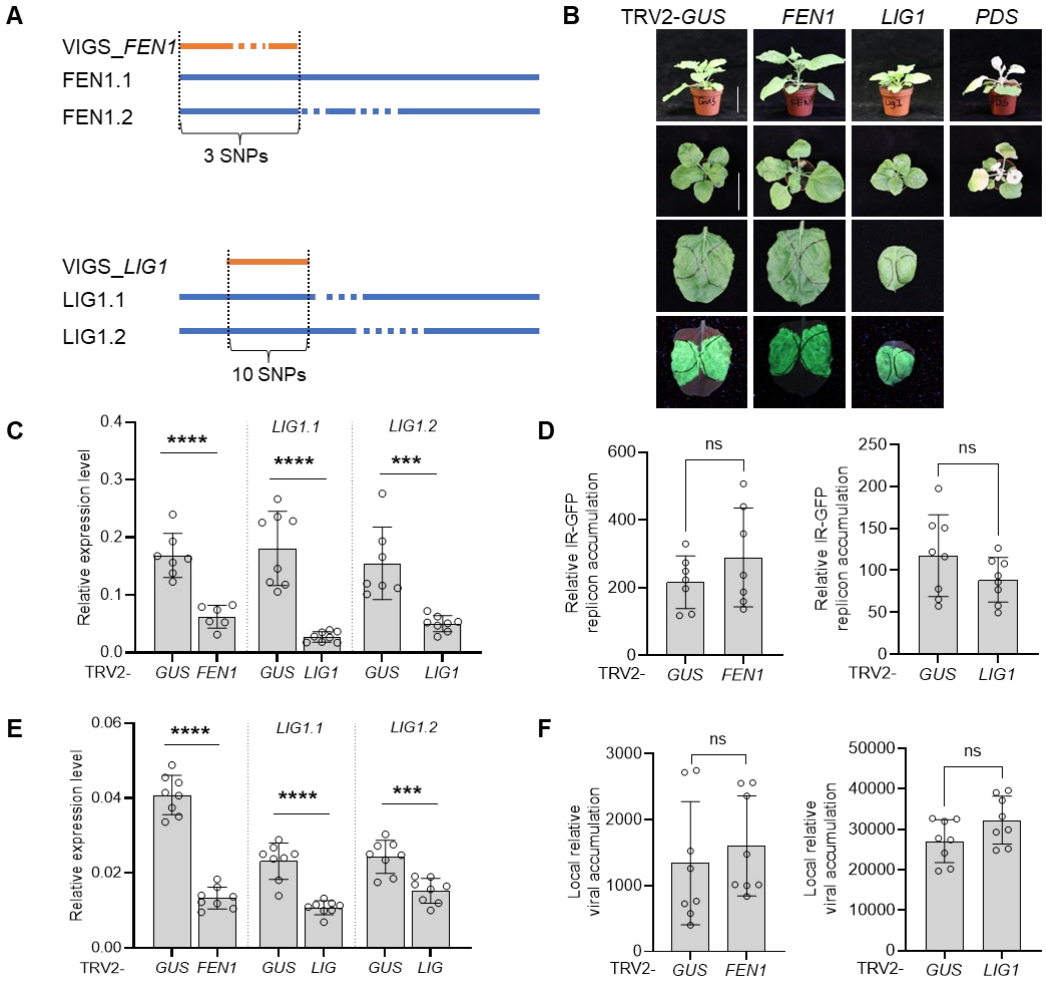


Figure 2.11 Geminiviral genome follows a leading-strand replication mode. A. Comparison of each *FEN1/LIG1* paralog and the position of the VIGS target sequence. The orange lines represent the VIGS sequence, while the blue lines represent the gene coding sequence. Dotted lines indicate deletions compared to its paralog. SNPs, single-nucleotide polymorphisms. **B.** Developmental phenotypes (upper panel) and GFP signal observation (lower panel) of *FEN1*- and *LIG1*-silenced 2IR-GFP transgenic *N. benthamiana* plants following agroinfiltration with Rep-RFP. Scale bar: 5 cm. **C-D.** Silencing efficiency (C) of *FEN1* and *LIG1* in 2IR-GFP transgenic *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and IR-GFP replicon accumulation (D), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. Experiments testing the role of *FEN1* and *LIG1* in the reporter system were repeated three times and twice, respectively, with similar results. **E-F.** Silencing efficiency (E) of *FEN1* and *LIG1* in wild-type *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (F), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. Experiments testing the role of *FEN1* and *LIG1* in TYLCV local infection were repeated twice and three times, respectively, with similar results. *Agrobacterium*

cells containing TRV1 and either TRV2-FEN1 or TRV2-LIG1 were mixed at a 1:1 ratio and inoculated in 2-week-old *N. benthamiana* seedlings. At 14 days post-TRV inoculation, plants were then agroinfiltrated with either Rep-RFP (panels B-D) or the TYLCV infectious clone (panels E-F) in 2IR-GFP transgenic or wild-type *N. benthamiana* plants, respectively. Samples were harvested at 2 days post-infiltration for the reporter system and at 3 days post-inoculation for TYLCV infection. Plants inoculated with TRV2-GUS and TRV2-PDS were used as the negative and positive controls, respectively. Data are presented relative to TRV2-GUS plants, with values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student's t-test (****, $P<0.0001$; ***, $P<0.001$; **, $P<0.01$; *, $P<0.05$).

2.4 Leading-strand replication components are required for replication of CRESS viruses

All CRESS DNA viruses ubiquitously employ a rolling-circle replication mechanism for their genome replication. To gain deeper insight into the conservation of host factors involved in CRESS DNA viral replication, we conducted a comparative analysis across different viral species. First, we assessed the effect of silencing selected DNA replication-associated genes, found to be required for TYLCV replication (*RFC1*, *POLD3*, *POLD1*, and *PCNA*), on the replication of two additional geminiviruses: AbMV and chickpea chlorotic dwarf virus (CpCDV), a member of the *Mastrevirus* genus. *TOPO1* was excluded from this analysis due to its severe phenotypic effects observed upon its silencing in *N. benthamiana* plants. As shown in Figure 2.12A, the phylogenetic tree built based on the Rep protein sequences of representative geminiviruses highlights the evolutionary relationships among them, with the viruses used to assess the roles of selected replication-associated proteins in viral DNA replication indicated on the tree. Viral infection assays with AbMV (Figure 2.12) and CpCDV (Figure 2.13) were performed as described for TYLCV. The silencing efficiency for each gene in AbMV local infection assays is presented in Figure 2.12B. Notably, local viral accumulation—reflecting mostly viral replication—was almost completely suppressed in all gene-silenced *N. benthamiana* plants infected with AbMV (Figure 2.12C). Overall, our results indicate that these selected replication-associated proteins are essential for the replication of bipartite begomovirus.

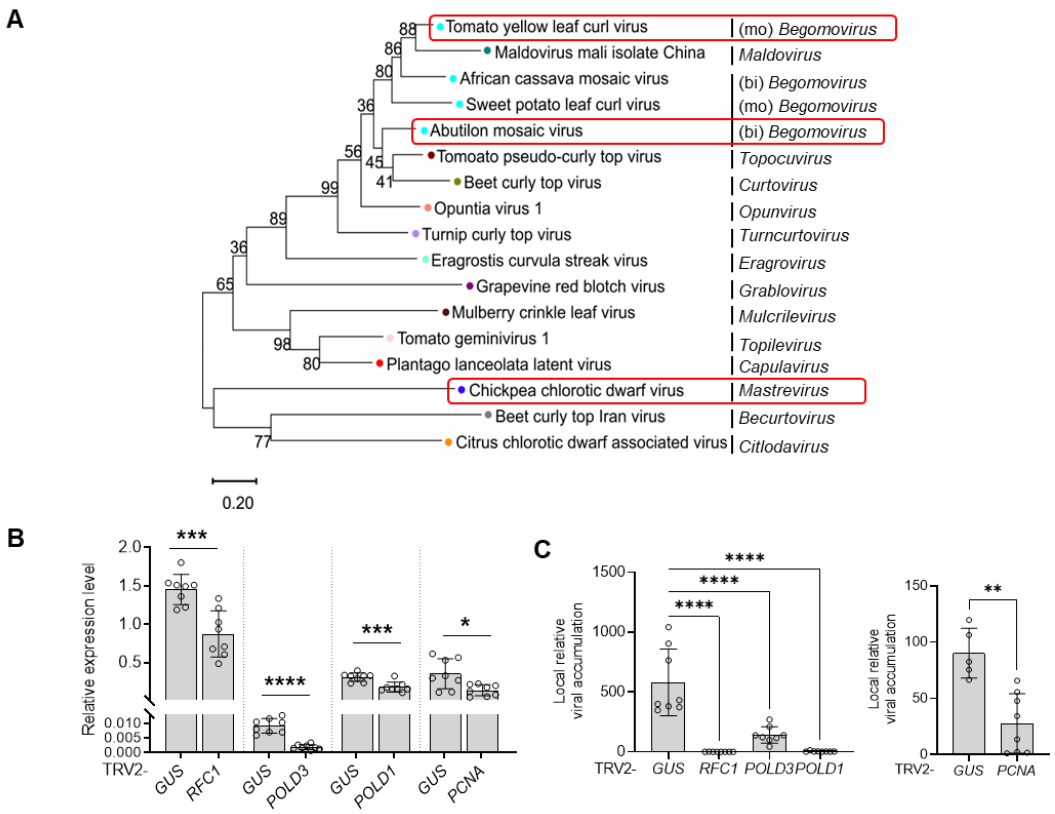


Figure 2.12 Selected DNA replication-associated proteins are required for bipartite begomovirus replication. A. Phylogenetic tree of Rep protein sequences of representative geminiviruses. The tree was built using the neighbor-joining method in MEGA11 with 1,000 bootstrap replicates. Numbers at branches and the bottom indicate bootstrap values (%) and substitutions per site, respectively. Mono- and bipartite begomovirus were labelled as “mo” and “bi”, respectively. **B-C.** Silencing efficiency (B) of *RFC1*, *POLD3*, *POLD1*, and *PCNA* in *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (C), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TRV1 and TRV2-RFC1, TRV2-POLD3, TRV2-POLD1, and TRV2-PCNA were mixed at a 1:1 ratio and inoculated in 2-week-old *N. benthamiana* seedlings. Plants were then agroinfiltrated with the AbMV infectious clone at 14 days post-TRV inoculation. All samples were harvested at 3 days post inoculation. The biological replicates shown in panels B-C were independently performed by the author and Dr. Laura Medina-Puche. Data are presented relative to TRV2-GUS plants, with values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student’s t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). All experiments were repeated at least two times with similar results.

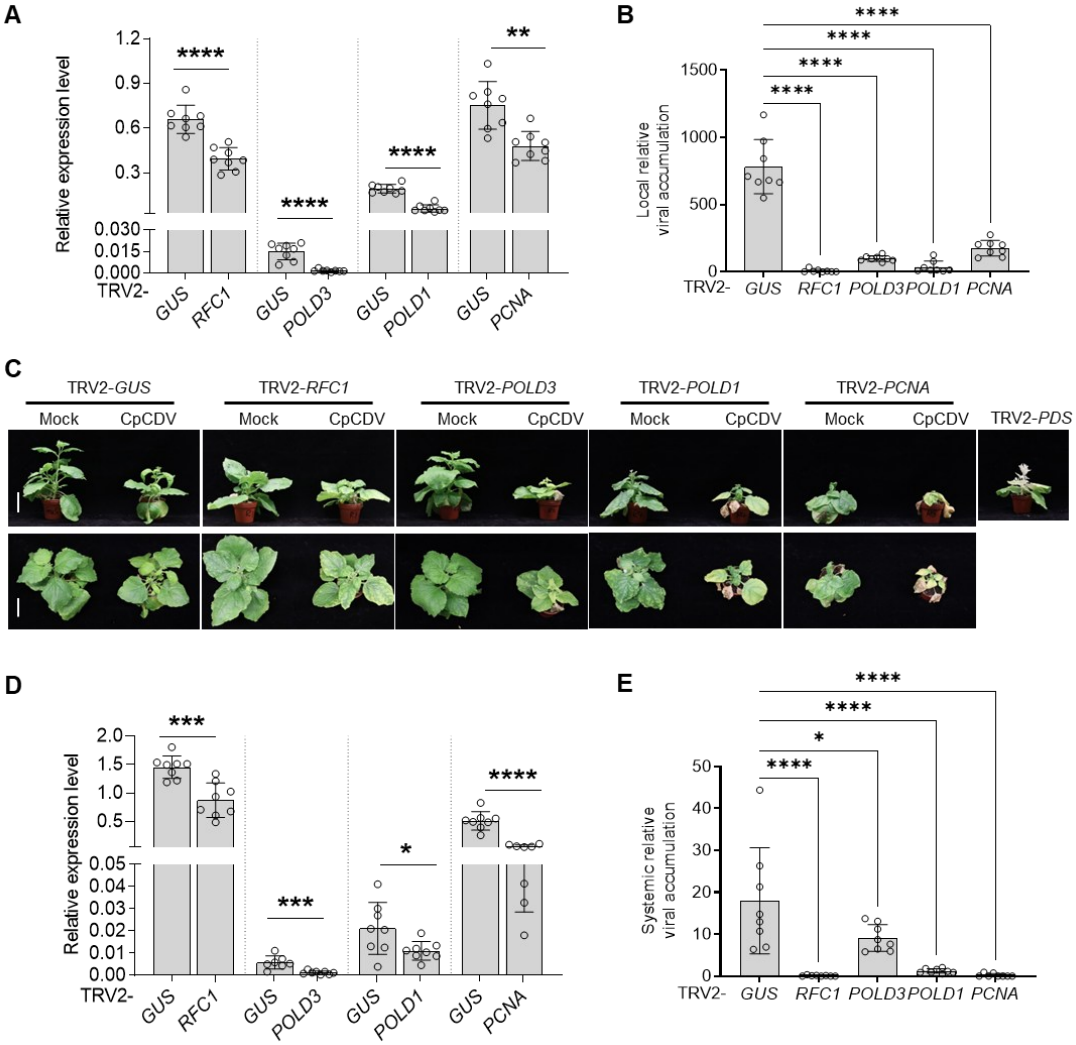


Figure 2.13 Selected replication-associated proteins are required for mastrevirus replication. A-B. Silencing efficiency (A) of *RFC1*, *POLD3*, *POLD1*, and *PCNA* in *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (B), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TRV1 and TRV2-RFC1, TRV2-POLD3, TRV2-POLD1, and TRV2-PCNA were mixed at a 1:1 ratio and inoculated into 2-week-old *N. benthamiana* seedlings. Plants were then agroinfiltrated with the CpCDV infectious clone at 14 days post-TRV inoculation (dpi). All samples were harvested at 3 days post-inoculation. **C-E.** Symptom observation (C, scale bar: 5cm) in *RFC1*-, *POLD3*-, *POLD1*-, and *PCNA*-silenced *N. benthamiana* plants inoculated with the CpCDV infectious clone. Silencing efficiency (D), measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (E), measured by qPCR with ITS as an internal reference. *Agrobacterium* cells containing CpCDV infectious clone, TRV1 and TRV2-RFC1, TRV2-POLD3, TRV2-POLD1, and TRV2-PCNA were mixed at a 1:1:1 ratio and inoculated in 2-week-old *N. benthamiana* seedlings. Samples were harvested at 14 dpi. The biological replicates shown in

panels C-E were independently performed by the author and Dr. Delphine Pott. Plants inoculated with TRV2-GUS and TRV2-PDS were used as negative and positive controls, respectively. Data are presented relative to TRV2-GUS plants, with values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student's t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). All experiments were repeated at least two times with similar results.

In CpCDV locally and systemically infected plants, the reduced gene expression levels in each gene-silenced *N. benthamiana* plants compared to the negative control were confirmed (Figure 2.13A and Figure 2.13D). Strikingly, the viral accumulation in gene-silenced samples were significantly decreased upon both local and systemic infection compared to the control samples (Figure 2.13B and Figure 2.13E). Furthermore, the symptoms observed in *N. benthamiana* plants infected with CpCDV—including stunting, yellowing, and downward curling of leaves—were consistent with previous reports[276] (Figure 2.13C). Taken together, these results demonstrate the essential roles of replication associated proteins labelled by TYLCV and AbMV in mastrevirus replication, despite the phylogenetic divergence between CpCDV Rep and TYLCV Rep, as illustrated in Figure 2.12A. This further suggests that the function of these replication-related proteins in geminiviral replication is likely conserved across different geminiviral species.

Nanoviridae is a family of plant viruses belonging to the CRESS DNA phylum. Unlike most geminiviruses, nanoviruses are multipartite, consisting of 6-8 circular ssDNA molecules of approximately 1 kb in size[13]. Similar to geminiviral Rep, the Rep protein encoded by nanoviruses is the only viral protein essential for viral DNA replication. To further investigate whether host factors labelled by geminiviral Rep are also required for replication of other CRESS DNA viruses, we examined their roles in DNA replication mediated by the Rep protein of pea necrotic yellow dwarf virus (PNYDV; genus *Nanovirus*, family *Nanoviridae*). As mentioned earlier, viruses in this family possess a multipartite genome, with each component encoding a single viral protein. Among them, only the DNA-R, which encodes a Rep protein, is capable of replicating autonomously. Although PNYDV lacks infectivity in *N. benthamiana*, its DNA-R has the ability to replicate in this species. Therefore, *Agrobacterium*-mediated delivery of the circular PNYDV DNA-R replicon into silenced *N. benthamiana* plants allowed us to assess the role of geminiviral Rep-labelled proteins in PNYDV Rep-mediated DNA-R replication. As shown in Figure 2.14, PNYDV Rep DNA levels were dramatically decreased in gene-silenced *N. benthamiana* plants compared to control samples. This strongly indicates that RFC1, POLD3, POLD1 and PCNA

are essential for the genome replication of nanoviruses.

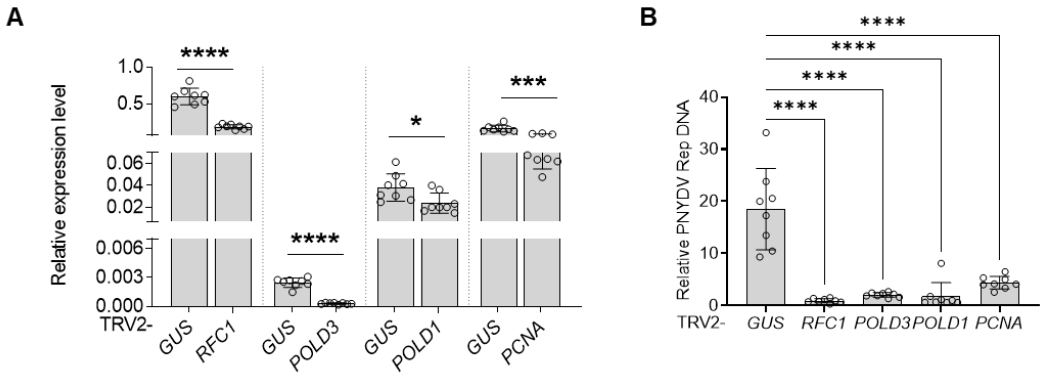


Figure 2.14 Selected replication-associated proteins are required for nanovirus replication. A-B. Silencing efficiency (A) of *RFC1*, *POLD3*, *POLD1*, and *PCNA* in *N. benthamiana* infiltrated with *Agrobacterium* cells containing PNYDV DNA-R, measured by qRT-PCR using *NbActin* as an internal reference, and PNYDV DNA-R quantification (B), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. Samples were harvested at 2 days post-infiltration. Data are presented relative to TRV2-GUS plants, with values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student's t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). All experiments were repeated at least two times with similar results.

Overall, our results suggest that CRESS DNA viruses exploit host factors involved in leading-strand synthesis—rather than lagging-strand synthesis—in eukaryotic bi-directional DNA replication with a swapped DNA polymerase. This implies a conserved replication strategy among CRESS DNA viruses, wherein host factors involved in eukaryotic DNA replication are selectively recruited by CRESS DNA virus Rep protein to perform rolling-circle replication.

2.5 Rep recruits helicase loaders to initiate geminiviral replication and potentially acts as the replicative helicase

All the aforementioned host factors recruited by Rep protein as components of the viral replisome are involved in the elongation stage of DNA replication. However, the molecular mechanisms acting upstream, underlying geminiviral replication initiation, as well as the host factors crucial for this step remain elusive. CRESS DNA viral Rep proteins are known to recognize the viral origin of replication and introduce a site-specific nick to initiate rolling-circle replication[6]. However, whether host DNA replication machinery is recruited to and assembled on the geminiviral genome at the onset of replication remains an open question. It should be noted that no replication initiation-associated proteins were identified in our PL-MS dataset. This is likely due to the limited

duration of the replication initiation step compared to the elongation process.

To investigate whether host DNA replication initiation components are recruited and functionally coordinated during the initiation of geminiviral replication, we examined the effects of silencing *ORC1* (the large subunit of ORC complex) and *CDC6*, two eukaryotic DNA helicase loaders, on geminiviral genome replication. All paralogs of *ORC1* and *CDC6* in *N. benthamiana*, along with their transcript abundance (measured in RPM^{273}) following TYLCV local infection, are summarized in Table 2.3. Due to the very low read counts of *CDC6.2* in both the presence and absence of TYLCV, only *CDC6.1* was selected for further experiments and is hereafter referred as *CDC6*. As previously performed for Rep-labelled replication-associated proteins, we observed the subcellular localization of ORC1.1-GFP and GFP-CDC6, in the presence and absence of the Rep protein or the TYLCV (Figure 2.15). Data for ORC1.2 is not shown, as it exhibited an almost undetectable signal under the confocal microscope. Both ORC1.1 and CDC6 localized to the nucleus and co-localized with Rep. Notably, TYLCV inoculation induced a striking redistribution of ORC1.1 and CDC6, with increased accumulation at the periphery of the nucleus and the nucleolus, accompanied by a substantial reduction in signal intensity within the nuclear interior (Figure 2.15B).

Table 2.3 Selected host factors involved in eukaryotic DNA replication initiation and the expression levels of each orthologue coding gene in *N. benthamiana.* Expression data are taken from Wu *et al*. (2019) and represent transcription abundance measured in RPM (reads per million). Data were obtained from TYLCV locally-infected samples, with agroinfiltration of an empty vector (EV) serving as the negative control.

Name	Gene identifier	Annotation	Source	EV	TYLCV
ORC1.1	Niben101Scf02055g00001	Origin recognition complex 1	Functional information	164	47
ORC1.2	Niben101Scf02706g02004			200	117
CDC6.1	Niben101Scf06541g00002	Cell division control 6	Functional information	217	8
CDC6.2	Niben101Scf01694g12008			1	2

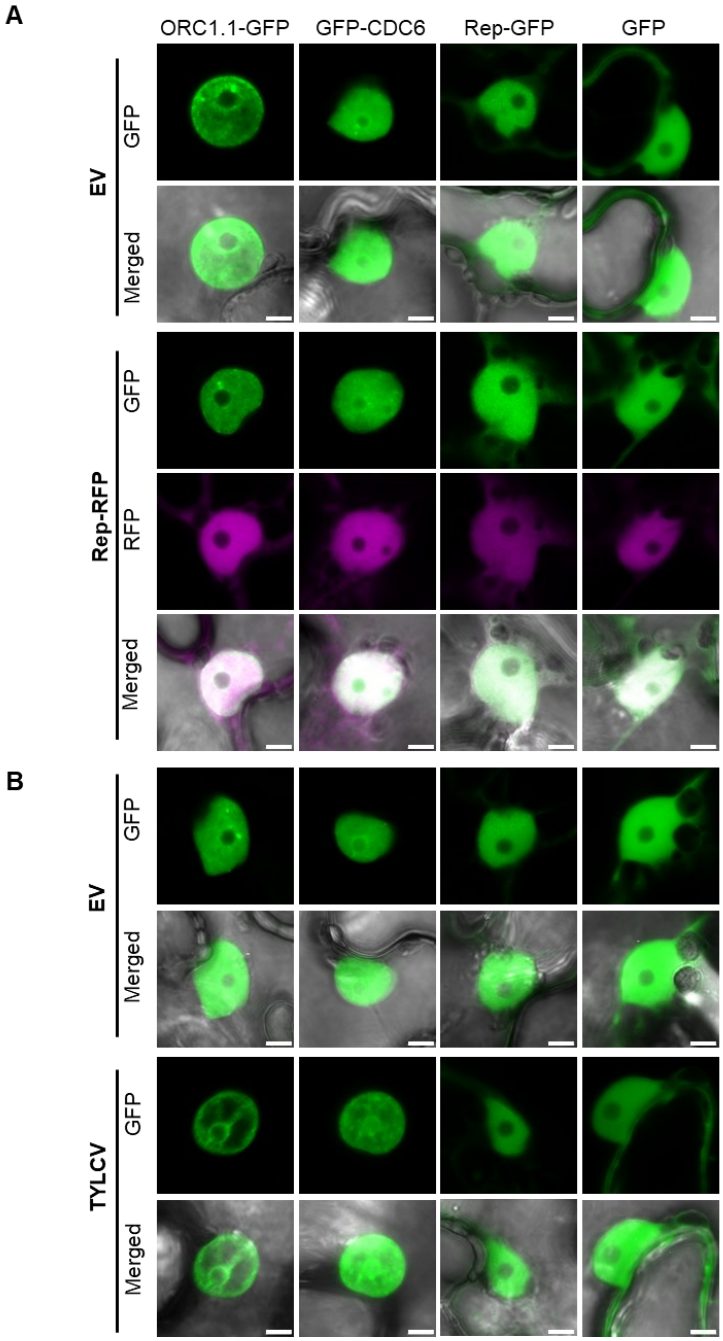


Figure 2.15 Subcellular localization of eukaryotic DNA helicase loaders. A-B. Transient expression of ORC1.1-GFP, GFP-CDC6, Rep-GFP, and GFP in *N. benthamiana* leaves, either in the absence (EV) or presence of Rep-RFP (A) or TYLCV (B). *Agrobacterium* cells containing the respective binary vectors were co-infiltrated at a 1:1 ratio; Rep-GFP and GFP were used as positive and negative controls, respectively; confocal images were taken at 30 hours post-infiltration. Scale bar: 5 µm. This experiment was repeated twice with similar results.

Next, we aimed to evaluate the role of ORC1 and CDC6 in geminiviral replication using both the 2IR-GFP reporter system and viral infection assays.

However, we encountered difficulties in effectively detecting silencing of *CDC6* in *N. benthamiana* plants via VIGS, likely due to the low expression of this gene in infected samples[273]; as a result, only the role of ORC1 in viral replication was examined through this approach. Given the crucial role of CDC6 in loading eukaryotic DNA helicases onto the DNA duplex in cooperation with ORC during replication initiation process, we have nevertheless included CDC6 in the following experiments. *ORC1*-silenced *N. benthamiana* plants displayed a reduction of height and thick leaves compared with control plants (Figure 2.16A). The significant contribution of ORC1 to viral replication was confirmed using both the 2IR-GFP reporter system (Figure 2.16B & C) and virus-infected samples (Figure 2.16D & E). These results suggest that ORC, which binds to the replication origin in eukaryotes, also plays a crucial role in the initiation of geminiviral replication.

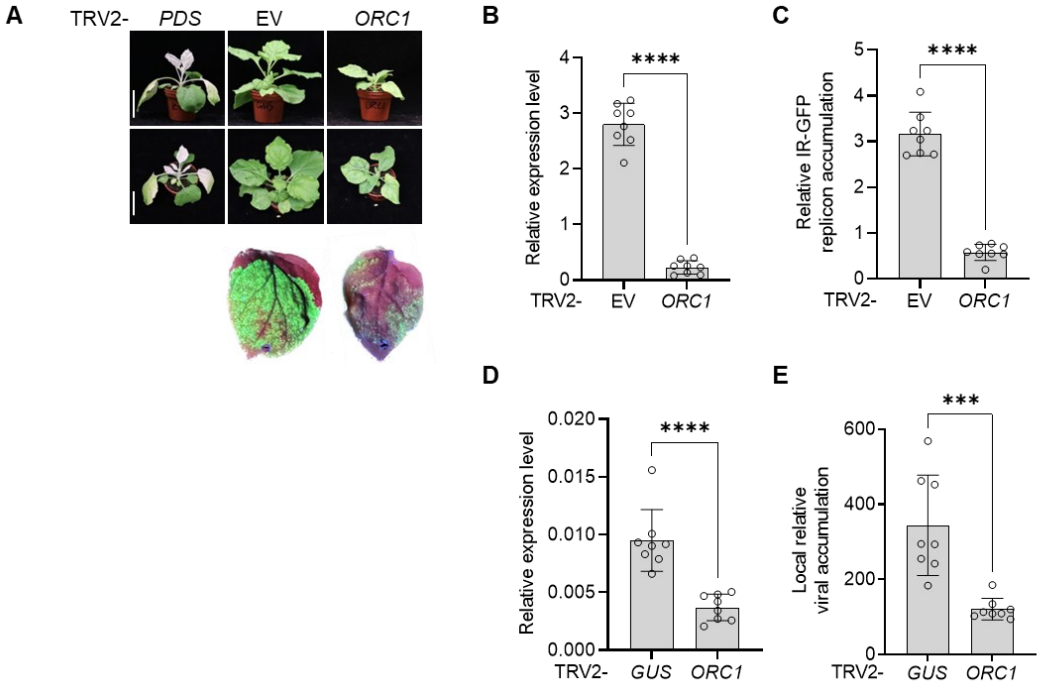


Figure 2.16 Eukaryotic replicative DNA helicase loaders contribute to geminiviral replication. A. Developmental phenotype (upper panel) and GFP signal observation (lower panel) of *ORC1*-silenced 2IR-GFP transgenic *N. benthamiana* plants following agroinfiltration with Rep-RFP. Scale bar: 5 cm. **B-C.** Silencing efficiency (B) of *ORC1* in 2IR-GFP transgenic *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and IR-GFP replicon accumulation (C), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. The results shown in panels A-C were generated by Master's student Vivian Heise. **D-E.** Silencing efficiency (D) of *ORC1* in wild-type *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (E), measured by qPCR

with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TRV1 and TRV2-ORC1 were mixed at a 1:1 ratio and inoculated in 2-week-old *N. benthamiana* seedlings. At 14 days post-TRV inoculation, plants were then agroinfiltrated with either Rep-RFP (panels A-C) or the TYLCV infectious clone (panels D-E) in 2IR-GFP transgenic or wild-type *N. benthamiana* plants, respectively. Samples were harvested at 2 days post-infiltration for the reporter system and 3 days post-inoculation for TYLCV infection. All experiments were repeated three times with similar results. Plants inoculated with TRV2-GUS/EV and TRV2-PDS were used as negative and positive controls, respectively. Data are presented relative to TRV2-GUS/EV plants, with values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student's t-test (****, $P<0.0001$; ***, $P<0.001$; **, $P<0.01$; *, $P<0.05$). All experiments were repeated at least two times with similar results.

Given that Rep functions as the initiator protein in geminiviral replication, while ORC and CDC6 act as initiator proteins in eukaryotic replication, and that ORC1 has been shown to be required for geminiviral replication, we hypothesized that a function interplay may exist between Rep and these two host initiators. To test whether Rep physically interacts with ORC1 and CDC6, a co-IP experiment was performed. Our results indicate that Rep associates with both ORC1 and CDC6, but not the negative control, GFP (Figure 2.17).

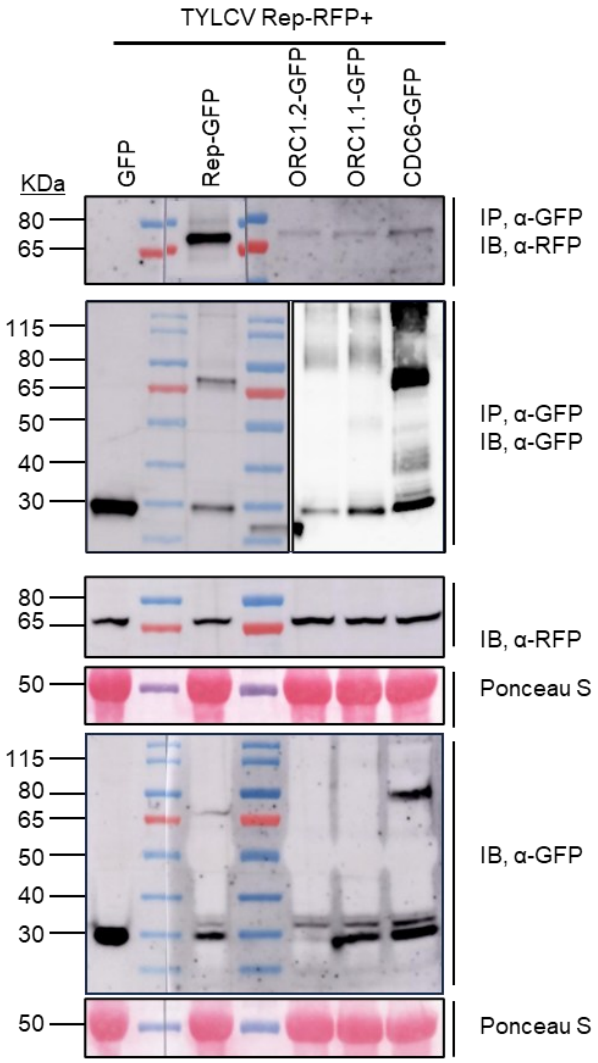


Figure 2.17 Rep associates with ORC1 and CDC6 by co-IP. Rep co-immunoprecipitates with ORC1.1-GFP, ORC1.2-GFP, POLD3-CDC6, and Rep-GFP (positive control), but not with GFP (negative control), upon transient expression in *N. benhtamiana*. IP: immunoprecipitate; IB: immunoblotting; Ponceau S, ponceau staining. Molecular weights are indicated on the left. The predicted protein sizes are as follows: ORC1.1-GFP, ~103 kDa; ORC1.2-GFP, ~99 kDa; CDC6-POLD1, ~92 kDa; Rep-GFP, ~68 kDa; GFP, ~28 kDa. This experiment was repeated twice with similar results.

Next, we sought to examine whether the helicase loaders ORC and CDC6 bind to the viral origin of replication, and whether this binding depends on Rep, which inherently binds the geminiviral origin. For this purpose, we evaluated the binding ability of ORC1 and CDC6 to the geminiviral IR in the presence and absence of TYLCV Rep using ChIP-qPCR, following the strategy described previously in Section 2.1.2. *N. benthamiana* plants were agroinfiltrated with constructs encoding IR-GFP, FLAG-tagged proteins of interest, and either Rep-

RFP or a negative control (RFP), then analyzed at 2 days post-infiltration, as shown in Figure 2.18. PCNA-FLAG was used as a positive control, as demonstrated in Figure 2.10. Both CDC6 and ORC1.1, as well as the positive control, were observed to weakly bind to IR in the absence of Rep. However, when Rep is present, the binding of CDC6, ORC1.1, ORC1.2, and PCNA to the IR cassette showed a dramatic increase. Although the role of CDC6 in geminiviral replication was not definitively established, our results strongly suggest that Rep recruits both CDC6 and ORC1, significantly enhancing their binding to the IR. This implies a potential involvement of both ORC1 and CDC6 in the viral replication initiation step.

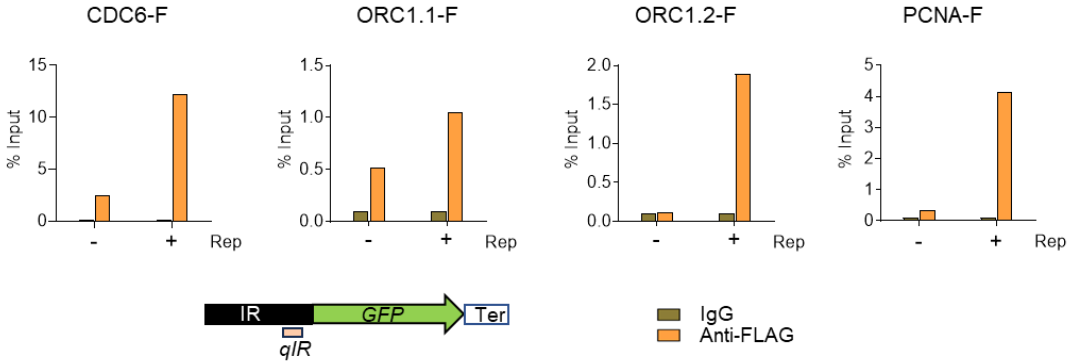


Figure 2.18 Rep significantly stimulates DNA helicase loaders binding to the geminiviral intergenic region. CDC6, ORC1.1, ORC1.2 and PCNA (positive control) bind the geminiviral IR region, including the origin of replication, in a Rep-dependent manner. Co-expression of FLAG-tagged proteins, Rep-RFP or RFP, and TYLCV IR was performed in 4-week-old WT *N. benthamiana* plants. Genomic location of the IR region amplified in the ChIP-qPCR assays is indicated in orange at the bottom of panel. The brown and orange bars in the graphs represent IgG (negative control) and protein binding, respectively. F, FLAG. All experiments were repeated at least twice with similar results.

The ORC is a multi-subunit protein complex comprising six subunits that play a crucial role in recognizing and binding replication origins in eukaryotes. In contrast, geminiviral Rep protein has been shown to site-specifically recognize their cognate origin of replication[277,278]. To assess whether ORC influences the binding of Rep to the viral origin of replication, we performed a ChIP-qPCR assay to evaluate Rep binding to the IR region in *ORC1*-silenced and control plants, as well as in *N. benthamiana* plants transiently expressing either GFP or ORC1-GFP. For this, we co-expressed Rep-FLAG and IR-GFP in *ORC1*-silenced and control plants at 14 days post-TRV inoculation and harvested samples after two days. Additionally, we performed agroinfiltration of constructs encoding Rep-FLAG, IR-GFP and either GFP or ORC1-GFP in *N. benthamiana*

plants two days before sample collection. As shown in Figure 2.19, the binding of Rep to the IR region in *ORC1*-silenced plants was reduced by approximately 50% compared to control plants. However, western blot analysis revealed a lower level of Rep-FLAG protein in *ORC1*-silenced plants, which likely accounts for the observed reduction in Rep binding to the IR. Furthermore, overexpression of ORC did not significantly enhance Rep binding to the geminiviral IR. Taken together, these results indicate that Rep itself is sufficient for binding to the geminiviral origin of replication, and ORC binding occurs downstream and, in a Rep-dependent manner.

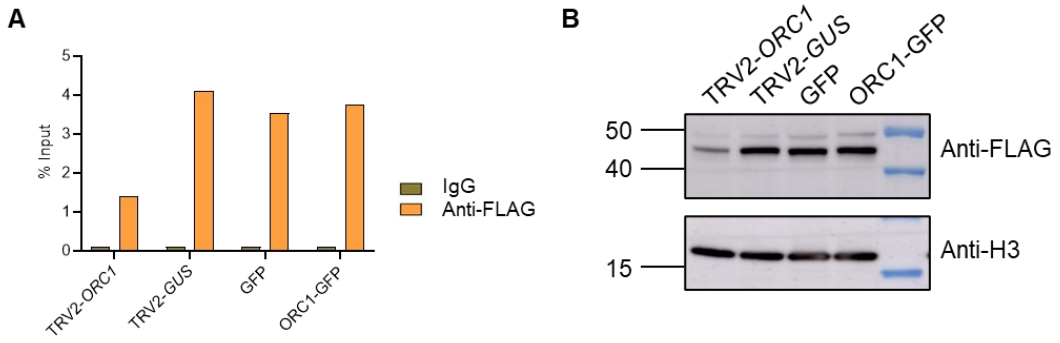


Figure 2.19 ORC1 does not contribute to Rep binding to geminiviral IR. A. Agroinfiltration of Rep-FLAG and TYLCV IR into *ORC1*-silenced *N. benthamiana* plants at 14 days post-TRV inoculation, with TRV2-GUS-inoculated plants used as controls (columns 1-4). Co-expression of Rep-FLAG, TYLCV IR, and either GFP or ORC1-GFP was performed in 3- or 4-week-old WT *N. benthamiana* plants (column 5-8). Samples were harvested at 2 days post-infiltration. The brown and orange bars in the graphs represent IgG (negative control) and anti-FLAG, respectively. **B.** Western blot analysis of Rep-FLAG protein level in different backgrounds. H3 served as an internal control. Molecular weights are indicated on the left. Rep-FLAG, ~41 kDa; H3, ~15 kDa. All experiments were repeated at least twice with similar results.

The DNA helicase loaders ORC and CDC6, along with CDT1, mediate the loading of the replicative MCM2-7 helicase complex onto replication origins—a highly conserved process across eukaryotes, including yeast, animals, and plants[106]. Given that ORC, and likely CDC6 as well, play important roles in the initiation of geminiviral replication, it is worthwhile to investigate the potential involvement the MCM helicase complex in this process. Interestingly, by examining the role of MCM3, a subunit of the MCM2-7 complex, in Rep-mediated replication using the 2IR-GFP reporter system, we observed that GFP signal in *MCM3*-silenced plants was comparable to that in control samples (Figure 2.20A). The expression level of *MCM3* was reduced in the silenced 2IR-GFP transgenic plants compared to the control (Figure 2.20B). Consist with the

observed GFP signal, qPCR analysis showed that IR-GFP replicon accumulation in *MCM3*-silenced plants was similar to that in control plants (Figure 2.20C). Considering that the MCM2-7 complex consists of six subunits arranged in the architecture shown in Figure 2.20D, and based on our results, we speculate that the MCM2-7 complex may not be involved geminiviral replication. However, how the parental DNA duplex is unwound during rolling-circle replication remains an open question.

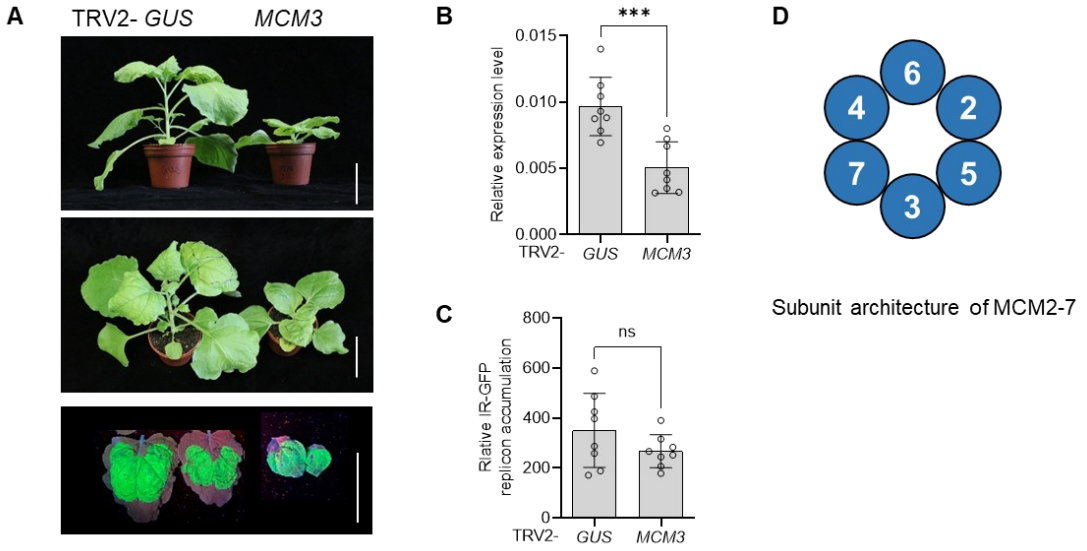
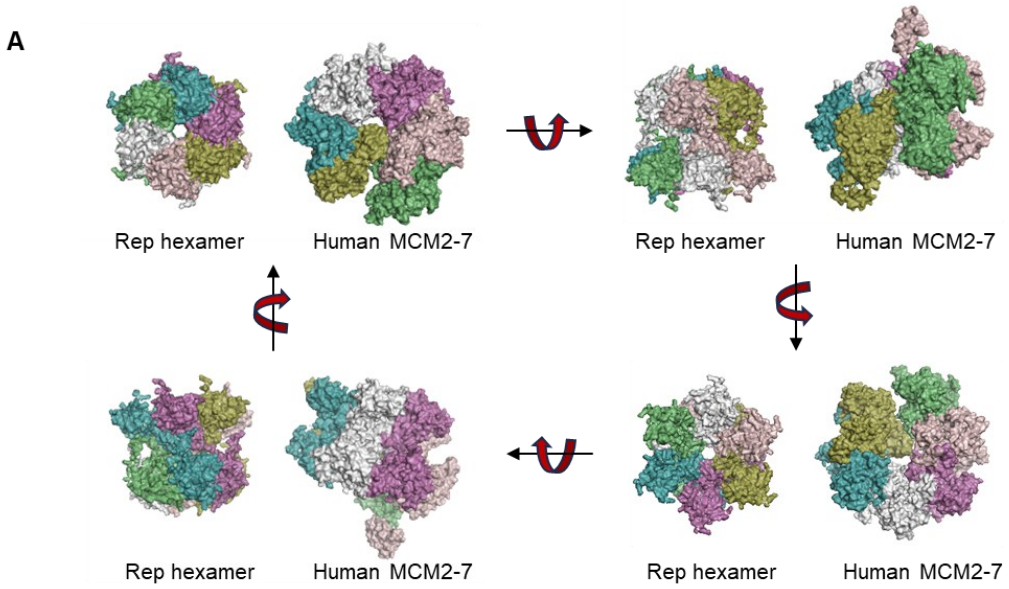


Figure 2.20 MCM3 is not required for Rep-dependent replication. A. Developmental phenotypes (upper panel) and GFP signal observation (lower panel) of *MCM3*-silenced 2IR-GFP transgenic *N. benthamiana* plants following agroinfiltration with Rep-RFP. Scale bar: 5 cm. **B-C.** Silencing efficiency (B) of *MCM3* in 2IR-GFP transgenic *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and IR-GFP replicon accumulation (C), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TRV1 and TRV2-MCM3 were mixed at a 1:1 ratio and inoculated into 2-week-old *N. benthamiana* seedlings. Plants were then agroinfiltrated with Rep-RFP at 14 days post-TRV inoculation. All samples were harvested at 2 days post-infiltration. Data are presented relative to TRV2-GUS plants, with values shown as the mean of eight individual plants, and error bars indicating standard deviation (SD). Asterisks indicate statistically significant differences based on Student's t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). This experiment was repeated twice with similar results. **D.** Subunit organization of MCM2-7, modified from Vijayraghavan and Schwacha[279].

Interestingly, previous studies have shown that Rep proteins from certain geminiviruses possess DNA helicase activity[97,98,217,280]. Notably, the Rep protein from tomato yellow leaf curl Sardinia virus (TYLCSV) forms an oligomer

consisting of approximately 5.8 monomers, likely assembling into a hexamer[98], similar to most replicative DNA helicases[281]. These studies raise the possibility that geminiviral Rep may harbour intrinsic helicase activity essential for DNA duplex unwinding during rolling-circle replication, analogous to the role of the MCM2-7 in eukaryotic bi-directional DNA replication. To explore this hypothesis, we predicted the structures of various Rep oligomers using AlphaFold 3 (Figure 2.21A). Among the predicted oligomeric structures, the structure of Rep hexamer and Rep heptamer showed the highest confidence scores (as indicated by ipTM and pTM values); only the Rep hexamer structure is presented here. A side-by-side comparison of the Rep hexameric structure with the human MCM2-7 complex[188] reveals a strikingly similarity in overall assembly, but with the Rep complex being notably smaller in size (left) than the MCM2-7 complex (right). Next, we performed blue-native PAGE to analyse the oligomeric state of the Rep protein following transient expression of Rep-FLAG in *N. benthamiana*. Our preliminary data (Figure 2.21B) showed a band, marked by a red asterisk at the top, corresponding in size to a Rep hexamer or heptamer. However, due to the limitations of this in vivo experiment, it remains possible that the observed complexes correspond to the interaction between Rep and host proteins. Further experiments are required to determine whether Rep indeed self-oligomerises into a hexamer. For example, assessing whether a Rep mutant deficient in self-interaction can still form a complex at the same position would help determine the specificity of the observed band. Additionally, hexamer formation and the potential DNA helicase activity of Rep could be assessed through in vitro assays. Nonetheless, self-interaction of Rep has been demonstrated in previous studies[223–225,282].



Predicted Rep hexamer (ipTM: 0.65, pTM: 0.66; left); the human MCM2-7 complex (PDB ID: 7W68; right)

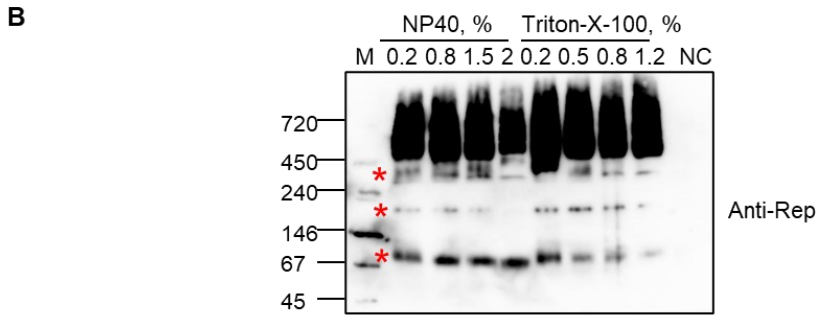


Figure 2.21 Predicted hexameric structure of Rep. A. Front, side, back, and bottom views of the geminiviral Rep hexamer (left; predicted by AlphaFold 3) and the human MCM2–7 complex (right; as described in Xu *et al*., 2022) are shown for visual comparison. Each color represents an individual Rep/MCM monomer. The confidence value of Rep hexamer and Protein Data Bank (PDB) ID of the MCM2-7 complex are indicated below. **B.** Blue-native PAGE analysis of oligomeric Rep protein upon transient expression of Rep-FLAG in *N. benthamiana* plants. Samples were harvested at 2 days post-infiltration and extracted using different concentrations of NP40 and Triton-X-100 detergents. The red asterisks at the lower, middle and top positions indicate the putative Rep dimer, tetramer or pentamer, and hexamer, respectively. Predicted oligomer sizes: monomer (~41 kDa), dimer (~82 kDa), trimer (~123 kDa), tetramer (~164 kDa), pentamer (~205 kDa), hexamer (~246 kDa), and heptamer (~287 kDa).

2.6 Potential roles of CDC48 and RuvB in geminiviral replication

In addition to the Rep-labelled DNA replication-associated proteins discussed

earlier, we identified two proteins in our PL experiments, cell division cycle 48 (CDC48) and holliday junction branch migration complex subunit (RuvB), that may contribute to geminiviral DNA replication. CDC48 is a ubiquitin-dependent unfoldase that plays a critical role in the termination of DNA replication by extracting the polyubiquitylated helicase subunit (MCM7) from the replisome complex and further facilitating its degradation via the 26S proteasome. This function has been documented across organisms including budding yeast, worm, frogs, mice, and human cell lines[171,178–181]. Furthermore, the role of CDC48 in cell cycle progression has been validated by several studies[283–286]. RuvB, a member of the AAA^+ family, works alongside with RuvA to process the Holiday junction in bacteria, a key intermediate in homologous recombination[287,288]. Both CDC48 and RuvB were found to be labelled by both TYLCV and AbMV Rep (Table 2.4). However, their potential roles in geminiviral replication remain undetermined. Here, we investigated their involvement in geminiviral replication, as shown in Figure 2.22-2.25. Upon *Agrobacterium*-mediated transformation of GFP tagged fusion proteins along with either Rep-RFP or TYLCV in *N. benthamiana*, both CDC48-GFP and RuvB-GFP were observed to co-localize with TYLCV Rep protein in the nucleus; the presence of the virus does not affect the localization pattern of these proteins (Figure 2.22). Additionally, cytoplasmic signal was also observed for CDC48, similar to that of GFP. To assess the physical interaction between Rep and CDC48 or RuvB, co-IP was performed. The results demonstrated that Rep associates with both CDC48 and RuvB, as well as the positive control PCNA-GFP, but not with the negative control, GFP (Figure 2.23). Next, the effects of silencing either *CDC48* or *RuvB* on IR-GFP replicon accumulation and viral accumulation were assessed in the experiments shown in Figure 2.24, where *CDC48*- and *RuvB*-silenced *N. benthamiana* plants exhibited a significant reduction in IR-GFP replicon level and viral accumulation. Notably, *N. benthamiana* plants in which *CDC48* was silenced displayed severe developmental phenotypes, with reduction of plant height and curling and swelling leaves; these observations correlate with the known roles of CDC48 in multiple cellular processes[184]. Furthermore, we verified the binding of these two proteins to the IR-GFP replicon in 2IR-GFP transgenic *N. benthamiana* plants using ChIP-qPCR. Similar to the experimental design described in Figure 2.10 for replication-associated proteins, the fusion proteins used in Figure 2.25 were encoded by their SR version genes. Taken together, our results demonstrate that CDC48 and RuvB are involved in geminiviral genome replication.

Table 2.4 Additional selected candidates based on functional information.

Name	Gene identifier	Annotation	Source

CDC48	Niben101Scf01911g15015	Cell division cycle protein 48	TYLCV Rep & AbMV Rep PL experiments
RuvB	Niben101Scf01281g02006	Holliday junction ATP dependent DNA helicase	TYLCV Rep & AbMV Rep PL experiments

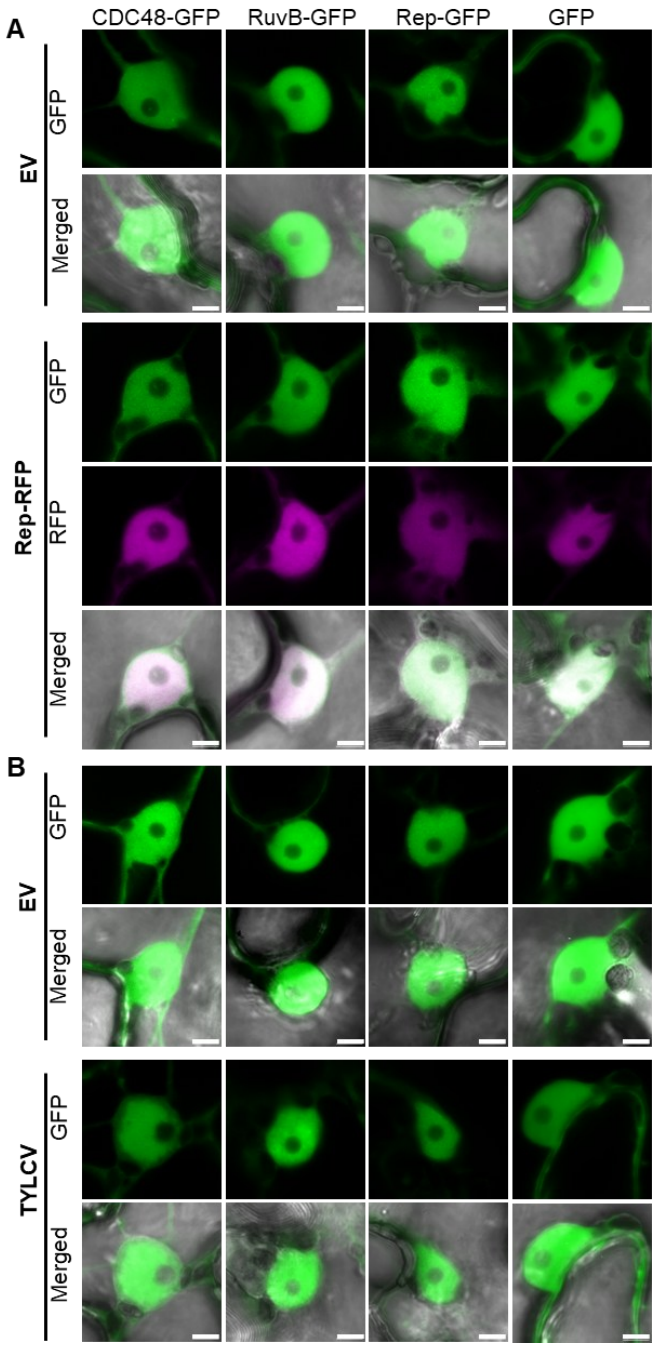


Figure 2.22 Subcellular localization of CDC48 and RuvB. A-B. Transient expression of CDC48-GFP, RuvB-GFP, and GFP in *N. benthamiana* leaves, either in the absence (EV) or presence of Rep-RFP (A) or TYLCV (B). *Agrobacterium* cells containing the respective binary vectors were co-infiltrated at 1:1 ratio; confocal images were taken at 30 hours post-infiltration. This experiment was performed in

parallel with ORC1 and CDC6. The same images of the positive control (Rep-GFP) and negative control (GFP) as shown in Figure 2.14 were used. Scale bar: 5 μm. This experiment was repeated twice with similar results.

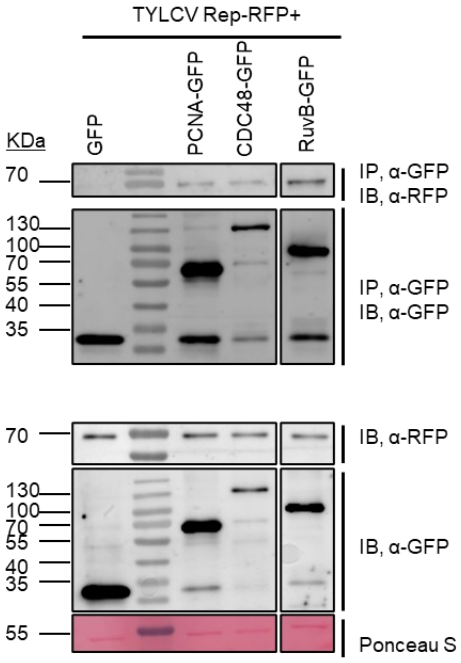


Figure 2.23 Rep associates with CDC48 and RuvB by co-IP. Rep co-immunoprecipitates with CDC48-GFP, RuvB-GFP, and Rep-GFP (positive control), but not with GFP (negative control), upon transient expression in *N. benthamiana*. IP: immunoprecipitate; IB: immunoblot; Ponceau S, ponceau staining. Molecular weights are indicated on the left. The predicted protein sizes are as follows: CDC48-GFP, ~117 kDa; RuvB-GFP, ~80 kDa; Rep-GFP, ~68 kDa; GFP, ~28 kDa. This experiment was repeated twice with similar results.

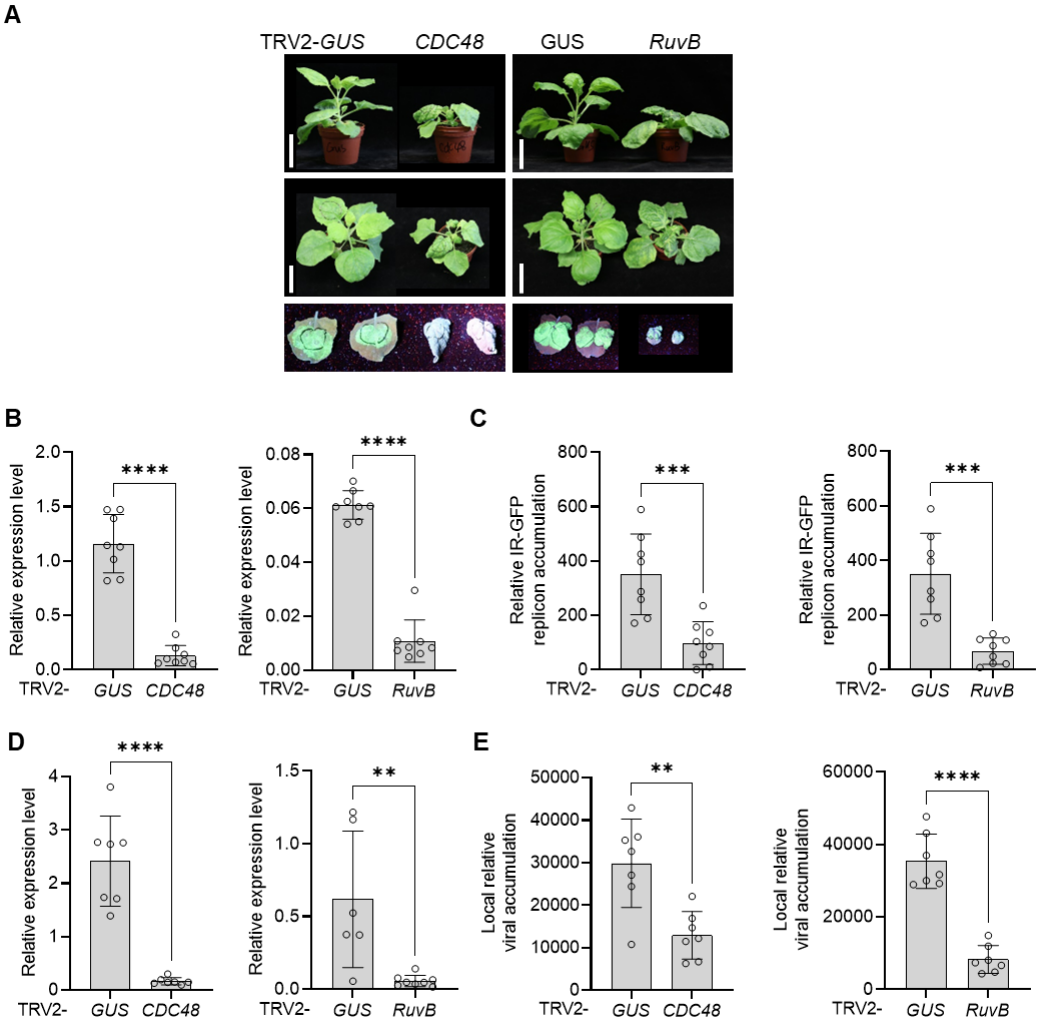


Figure 2.24 CDC48 and RuvB are required for geminiviral replication. A. Developmental phenotype (upper panel) and GFP signal observation (lower panel) of *CDC48*- and *RuvB*-silenced 2IR-GFP transgenic *N. benthamiana* plants following agroinfiltration with Rep-RFP. Scale bar: 5 cm. **B-C.** Silencing efficiency (B) of *CDC48* and *RuvB* in 2IR-GFP transgenic *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and IR-GFP replicon accumulation (D), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. **D-E.** Silencing efficiency (D) of *CDC48* and *RuvB* in wild-type *N. benthamiana* plants, measured by qRT-PCR using *NbActin* as an internal reference, and viral accumulation (E), measured by qPCR with 25S ribosomal DNA interspacer (ITS) as an internal reference. *Agrobacterium* cells containing TRV1 and either TRV2-CDC48 or TRV2-RuvB were mixed at a 1:1 ratio and inoculated in 2-week-old *N. benthamiana* seedlings. At 14 days post-TRV inoculation, plants were then agroinfiltrated with either Rep-RFP (panels A-C) or the TYLCV infectious clone (panels D-E) in 2IR-GFP transgenic or wild-type *N. benthamiana* plants, respectively. Samples are harvested at 2 days post-infiltration for the reporter system and 3 days post-inoculation for TYLCV infection.

Data are presented relative to TRV2-GUS plants, with values shown as the mean of eight individual plants, and error bars indicating SD. Asterisks indicate statistically significant differences based on Student's t-test (****, P<0.0001; ***, P<0.001; **, P<0.01; *, P<0.05). All experiments were repeated at least two times with similar results.

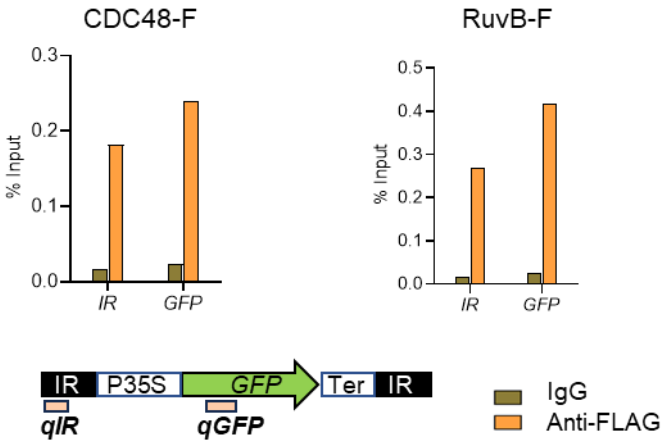


Figure 2.25 CDC48 and RuvB bind to IR-GFP replicon. CDC48 and RuvB bind to the IR-GFP replicon in the 2IR-GFP reporter system. Both proteins were encoded by their SR version genes. Co-infiltration of FLAG-tagged SR and Rep was performed in gene-silenced 2IR-GFP transgenic *N. benthamiana* plants. Genomic location of the IR-GFP replicon amplified in the ChIP-qPCR assays is indicated in orange at the bottom of panel. The brown and orange bars in the graphs represent IgG (negative control) and anti-FLAG, respectively. All experiments were repeated at least twice with similar results.

In this chapter, we examined the role of TOPO1, replication factor C largest subunit (RFC1), two subunits of Pol δ (POLD1 and POLD3), PCNA, FEN1, LIG1, large subunit of origin recognition complex (ORC1), CDC6, CDC48, and RuvB in geminiviral genome replication via VIGS upon transient expression in *N. benthamiana*. To determine whether *Agrobacterium*-mediated transformation efficiency was affected in these gene-silenced plants, we agroinfiltrated GFP into the silenced plants and assessed the protein accumulation using a western blot assay, as shown in Figure 2.23. Our results indicate that *Agrobacterium*-mediated transient expression of GFP was not severely affected in most gene-silenced plants, with the exception of those in which *PCNA*, *ORC1*, and *CDC48*, have been silenced, which showed noticeably lower protein accumulation compared to the control samples. However, the drastic reduction in viral accumulation observed in *PCNA*-silenced samples is unlikely to be solely due to impaired *Agrobacterium*-mediated transformation. Instead, it is more likely attributed to the essential role

of PCNA in geminiviral replication, as viral accumulation in *PCNA*-silenced plants was nearly undetectable (Figure 2.8B, Figure 2.9C, Figure 2.12C, Figure 2.13E and Figure 2.14B). Additionally, the binding of PCNA to the IR-GFP replicon in the 2IR-GFP reporter system, as well as to the viral genome during infection, further supports its role as a component of the viral replisome (Figure 2.10C & D). Although the efficiency of *Agrobacterium*-mediated transformation appeared to be diminished in *CDC48*-silenced samples, as indicated by reduced GFP protein accumulation compared to the control (Figure 2.24), CDC48 was shown to interact with Rep through co-IP (Figure 2.23) and to bind IR-GFP replicon in the reporter system (Figure 2.25). Furthermore, a previous study using yeast temperature-sensitive mutants library identified CDC48 as a factor involved in the replication of MYMIV[289]. Taken together, these observations support a putative role of CDC48 in the geminiviral DNA replication process. *Agrobacterium*-mediated transformation in *ORC1*-silenced plants was also slightly impeded. Nevertheless, strong recruitment of ORC1 by Rep to the viral IR was clearly observed in Figure 2.18. In the same study[289], ORC2 and ORC5, two subunits of origin recognition complex were identified to be involved in the replication of MYMIV, suggesting a potential role of the ORC complex in geminiviral DNA replication.

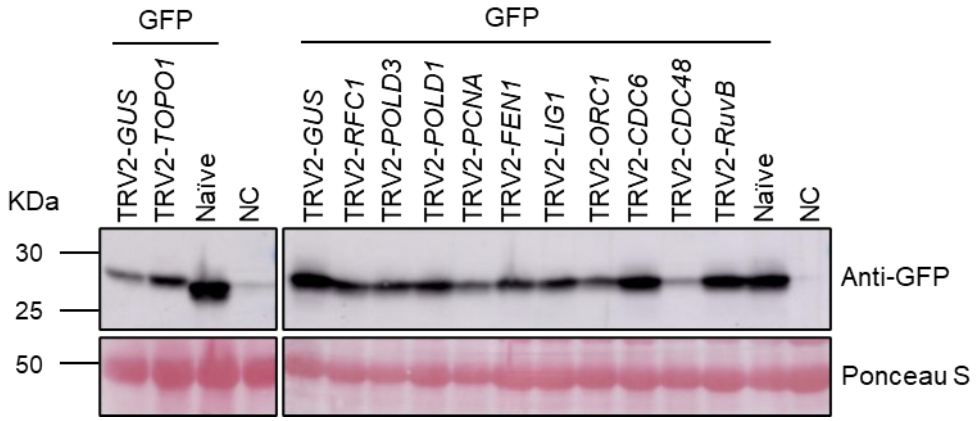


Figure 2.26 Assessment of *A. tumefaciens*-mediated transient expression of GFP in gene-silenced *N. benthamiana* plants. Western blot analysis of GFP accumulation in *N. benthamiana* plants silenced for *TOPO1*, *RFC1*, *POLD3*, *POLD1*, *PCNA*, *FEN1*, *LIG1*, *ORC1*, *CDC6*, *CDC48*, and *RuvB*, as well as naïve plant. TRV2-GUS-inoculated and naïve samples served as positive controls, while NC represents the negative control (infiltrated with an empty vector lacking GFP). Samples were taken at 2 days post-infiltration. Molecular weights are indicated on the left. Ponceaus S, ponceaus staining. GFP, ~28 kDa. This experiment was repeated twice with similar results.

Collectively, based on the results presented in this chapter, we propose a model illustrating the molecular mechanisms underlying geminiviral rolling-circle

replication, which utilizes a dsDNA intermediate converted from the single-stranded viral genome (Figure 2.27). At the onset of viral replication, the Rep protein, likely in its dimeric form[213], recognizes the viral origin of replication and introduces a site-specific nick at the conserved nonanucleotide motif within the stem-loop structure. Oligomeric Rep, likely in a non-hexameric form, subsequently recruits the host factors ORC and CDC6 to the viral IR through physical interaction, contributing to the replication initiation. Following origin recognition, Rep facilitates the recruitment of host DNA replication machinery, primarily components involved in leading-strand synthesis and Pol δ, to the viral genome to initiate rolling-circle replication. The introduced nick generates a 5’ end and a 3’ hydroxyl group. The 5′ end is covalently bound by Rep, while the 3′ end serves as a primer for extension by Pol δ, assisted by the sliding clamp PCNA, which is loaded onto DNA duplex by clamp loader RFC. Simultaneously, hexameric Rep likely functions as a replicative helicase, unwinding the dsDNA. As the replication fork progresses, TOPO1, labelled by Rep, functions on alleviating supercoiling ahead of the fork. To protect the exposed ssDNA generated during replication, RPA binds to it, preventing degradation. At the termination stage, Rep catalyzes the re-ligation of the original nick introduced at the initiation step, resulting in the production of the initial ssDNA released from the circular dsDNA genome. During this process, and considering the role of CDC48 in geminiviral replication, it is plausible that CDC48 contributes to the disassembly of the viral replisome, a role analogous to its function during eukaryotic replication termination. However, further investigation is required to determine whether the viral helicase is ubiquitylated and subsequently extracted by CDC48. Additionally, RuvB may be involved in the replication-recombination replication of geminiviruses. Taken together, our study demonstrates that geminiviruses exploit the molecular machinery mediating eukaryotic leading-strand, but not lagging-strand, DNA synthesis in the bidirectional replication fork. An exception is the DNA polymerases, which are swapped, perhaps contributing to the high mutation rates observed in these viruses[290–293], given the error-prone feature of Pol δ when functioning on the leading strand during break-induced replication[154,155]. Furthermore, our data also demonstrate that, in addition to geminiviruses, other CRESS viruses—which also utilize rolling-circle replication—likely co-opt this conserved replication strategy for their replication of their genomes.

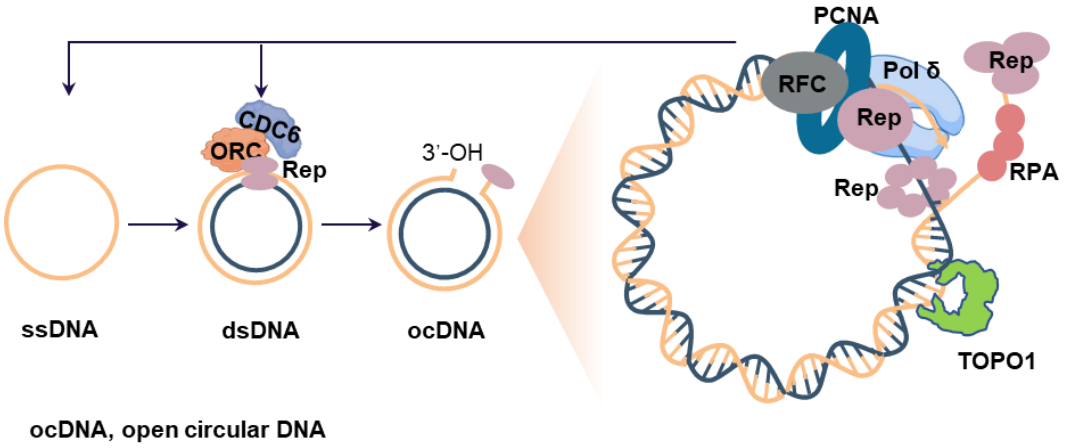


Figure 2.27 Proposed hypothetical model of CRESS DNA viral replication. The dsDNA intermediate, derived from the ssDNA genome, serves as the template for rolling-circle replication. The Rep protein specifically binds to and nicks the conserved nonanucleotide sequence within the viral origin of replication. ORC and CDC6, which are recruited to the IR by Rep, contribute to the onset of rolling-circle replication through a mechanism that remains to be elucidated. Following replication initiation, the hexameric form of Rep is likely involved in unwinding the parental DNA duplex. On the other hand, the host replication machinery factors, including TOPO1, RFC, PCNA, Pol δ, and RPA constitute the composition of viral replisome, promoting the elongation of the viral genome replication.

3. Discussion

3.1 TurboID-based PL as a useful tool to study geminiviral replication

3.1.1 TurboID-based PL overcomes the limitations of traditional approaches

TurboID-based PL has emerged as a powerful approach and has been applied across multiple organisms to study complex cellular proximal interaction networks. In this technique, TurboID—an engineered biotin ligase—can be fused to a protein of interest and directed to a specific cellular compartment by attaching a localization signal peptide. Similar to other engineered biotin ligases such as BioID (BirA*)[294], BioID2[295], miniTurboID[295], and AirID[296], TurboID converts supplemented biotin into a short-lived, diffusible intermediate—biotin-5′-AMP—which promiscuously labels the lysine residues of proximal proteins within a ~10-nm radius. The biotinylated proteins can be subsequently enriched using streptavidin beads and identified via MS[297]. Due to the labelling of transient or weak protein interactions, PL overcomes limitations of traditional affinity purification-MS (AP-MS). This was evident by comparing to our previous works, 1) we found there was no any interactors related to DNA replication in our unpublished Y2H screen data using Rep as bait against a library from TYLCV-infected tomato plants; 2) immunoprecipitation-MS (IP-MS) using TYLCV Rep as a bait also failed to detect replication-associated proteins[298]. Another independent study using the same method did not find out replication-related interactors of TYLCV Rep either[249]. By contrast, TurboID-based PL enables us, for the first time, to label replication-associated proteins in proximity to Rep. We further confirmed that some of these host factors, which serve as components of the viral replisome, are required not only for geminiviral replication but also for the replication of another plant-infecting CRESS DNA virus—pea necrotic yellow dwarf virus, a member of the *Nanoviridae* family. Notably, we have, for the first time, complemented a defective TYLCV Rep mutant with a functional Rep fusion protein. This strategy not only provides a native viral replication context but also ensures that the Rep proxiome is derived exclusively from the exogenously expressed Rep fusion protein, thereby avoiding competition with endogenous Rep.

3.1.2 It captures a broader range of host proxiome

Our PL-MS dataset revealed that certain host factors labelled by Rep proteins may function beyond viral DNA replication.

For instance, several nucleosome assembly related proteins, such as histone H1 and histone H2B, were labelled by Rep (the raw data of PL-MS are available

at the following link: https://www.dropbox.com/scl/fi/0cp3kt7bwk0hjt41mnm19/PL-raw-data_after-filtering.xlsx?rlkey=5rxtl71a4435v4s02iyt2lsmg&st=sn1z95i0&dl=0), supporting the mechanism by which geminiviruses exploit host histones to form minichromosomes[299], which serve as templates for viral transcription and rolling-circle replication[300].

Geminiviruses rely on host RNA polymerase II for viral gene transcription[58]. In agreement with this, our PL-MS data revealed that Rep from TYLCV, AbMV or both labelled three out of twelve RNA Pol II subunits: RPB2, RPB4, and RPB9[301].

Moreover, we identified a set of splicing-related host factors labelled by Rep proteins that may contribute to additional layers of regulation during viral infection. These include, but not limited to, U4/U6 small nuclear ribonucleoprotein Prp3, nuclear speckle regulatory protein 1, pre-mRNA splicing factor SYF2, and pre-mRNA splicing factor 38A/40A. However, the involvement of host alternative splicing machinery in geminivirus infection remains unclear. Interestingly, in viruses belonging to the genus mastrevirus, the Rep is produced from a spliced C1 and C2 ORFs, whereas RepA is derived from the unspliced C1 ORF, with both proteins being essential for viral replication and sharing functional similarities[63]. However, the molecular mechanisms regulating these splicing events are not well understood. It also remains unknown whether additional alternative splicing events occur in *Rep* mRNA or in transcripts encoding other canonical viral proteins such as C2, C3, CP, C4, V2, and CP in other geminivirus genera. Furthermore, whether host factors involved in viral infection undergo splicing events that are co-opted by geminiviruses to promote infection remains an open question. Future studies are needed to determine whether alternative splicing plays a role in viral infection.

We also detected translational machinery components among labelled proteins, including eukaryotic translation initiation factor (eIF2B and eIF3) and eukaryotic small and large ribosomal subunits such as 40s and 60s ribosomal proteins, this may indicate the possible engagement of host translation systems by geminiviruses. Notably, translation process typically occurs in the cytoplasm, but Rep protein localizes in the nucleoplasm. This raises the question of how Rep labels these translation-related proteins. One possibility is that Rep-TurboID may come into proximity with the translation machinery during its own translation. Subsequently, the mature Rep protein may further label components of the host translation machinery to facilitate viral infection. Although the mechanism by which viral proteins are initially translated in the

absence of Rep remains unknown, it is possible that Rep-mediated labelling of translation machinery could further promote viral infection.

Among these annotated host proteins identified in proximity to Rep, we confirmed that several replication-associated factors are essential for geminiviral replication. In addition, proteins involved in other biological processes may contribute to distinct stages of geminivirus infection and warrant further investigation. Besides, a number of proteins labelled by Rep in our dataset are annotated as having unknown function. Could this represent previously unrecognized components involved in geminiviral DNA replication? Might this dataset be leveraged to uncover novel host factors contributing to viral DNA replication? Addressing these questions may provide a more comprehensive understanding of geminiviral infection and facilitate the identification of the complete repertoire of factors constituting the viral replisome. Ultimately, these insights could advance the development of in vitro systems for studying geminiviral replication.

In conclusion, TurboID based PL is a versatile and sensitive tool for investigating geminiviral replication. However, its high sensitivity also increases the risk of nonspecific labelling, underscoring the importance of stringent MS data filtering to accurately distinguish true proximal proteins from background signals.

3.1.3 Integration of TurboID-based PL with other techniques for a comprehensive view

Despite the broad range of host factors involved in DNA replication, transcription, alternative splicing, translation, nucleosome assembly, and proteins of unknown function that were labelled by geminiviral Rep proteins through TurboID-based PL, notably, none of the host factors directly associated with DNA replication initiation were detected in close proximity to Rep. This absence is likely due to the transient and spatially restricted nature of the replication initiation step compared to the more prolonged viral replication elongation process. Moreover, since it is unclear whether Rep or C3 participates in the conversion of ssDNA to dsDNA, or is solely involved in the rolling-circle replication process, it remains challenging to determine which specific stages of geminiviral DNA replication the host proteins labelled by these viral proteins are associated with. Additionally, TurboID-based PL captures a pool of spatially adjacent proteins without temporal resolution, further complicating the ability to discern which host factors are functionally involved in each specific stage of viral replication. Besides, focusing on the proxiome or interactome of a single protein of interest, using it as a bait, may overlook critical components of the broader protein network, even if that protein plays an

essential role in a specific biological process. To overcome these limitations, complementary techniques are needed to further explore the proteome associated with viral replication. One promising approach is Proteomics of Isolated Chromatin segments (PICh)[302], which employs sequence-specific nucleic acid probes to isolate genomic DNA along with its associated protein complexes. By designing probes targeting the viral replication origin, the viral-sense strand, and the complementary-sense strand, it may be possible to selectively enrich and identify host proteins involved in distinct stages of geminiviral replication—such as replication initiation, ssDNA-to-dsDNA conversion, and rolling-circle replication, respectively. Another technique, known as isolation of proteins on nascent DNA (iPOND), enables the purification of proteins that are directly or indirectly associated with nascent DNA at the replication fork through the incorporation of the thymidine analog EdU, followed by click chemistry and MS[303,304]. This approach has the potential to identify a comprehensive set of host proteins involved in DNA replication, without being limited to those interacting with a single viral protein. However, since iPOND also captures proteins involved in host DNA replication, optimizing the method to specifically enrich for proteins contributing to viral DNA replication is a critical prerequisite.

3.2 Geminiviruses selectively recruit host factors to facilitate the viral replication

In Chapter II, we demonstrated the essential roles of several host factors labelled by Rep, including TOPO1, DNA Pol δ, PCNA, and RFC, in geminiviral DNA replication. Notably, two key components of the eukaryotic lagging-strand synthesis machinery, FEN1 and LIG1, were found to be dispensable for geminiviral genome replication, as determined using VIGS in combination with either the 2IR-GFP reporter system or viral infection assays. Importantly, conventional loss-of-function approaches like gene editing or mutagenesis are not feasible for studying genes that are essential for host viability, such as those involved in DNA replication, as is the case in our study. In contrast, VIGS enables partial and transient knockdown of gene expression rather than complete knockout, making it a powerful tool for investigating the roles of genes essential for host survival.

Furthermore, we confirmed that these proteins required for viral DNA replication, along with RPA, constitute components of the geminiviral replisome. This finding indicates that geminiviruses utilize the eukaryotic leading-strand DNA synthesis machinery, but not the lagging-strand machinery, at the bi-directional replication fork. An exception is the DNA polymerases, which are swapped, perhaps contributing to the high mutation rates observed in these viruses[290–293].

Notably, our findings suggest that the composition of the viral replisome is likely conserved among CRESS DNA viruses.

As obligate intracellular parasites with limited coding capacity, geminiviruses fully rely on the host replication machinery. Despite extensive research, the specific host proteins co-opted by geminiviruses for viral replication have remained largely unidentified for decades. While several studies have examined the host interactors of the viral proteins Rep and C3, known to be involved in viral DNA replication, many have not validated the roles of these host proteins in viral genome replication, nor have they confirmed their binding to the viral DNA. As a result, the composition of the geminiviral replisome remains largely elusive. In 2019, Li *et al*. identified a number of factors required for geminiviral replication through a genetic screen, including the DNA Pol δ subunits POLD1 and POLD2, as well as the catalytic subunit of DNA Pol α POLA1. However, their direct involvement as components of the viral replisome was not confirmed[289]. Subsequently, in 2021, Wu *et al*. demonstrated that DNA polymerases α and δ are essential components of the geminiviral replisome, a finding validated in both TYLCV and beet curly top virus (BCTV), which belong to different genera[192]. Building upon this, our study further confirmed the critical role of Pol δ not only in geminiviral genome replication but also in nanovirus DNA replication. Additionally, we validated the essential roles of processivity factor of DNA Pol δ sliding clamp PCNA and clamp loader RFC, as well as TOPO1, which alleviates torsional stress ahead of the replication fork, in geminiviral DNA replication. Moreover, we demonstrated that Rep interacts with RPA70—the largest subunit of the single-stranded DNA-binding protein RPA—via co-IP, and that RPA strongly binds to the viral genome. These results highly suggest that RPA is a component of the viral replisome. Considering Wu *et al*. (2021) and our own results, we proposed a model in which the geminiviral replisome comprises DNA Pol α, DNA Pol δ, PCNA, RFC, RPA, and TOPO1. Importantly, the host factors exploited by these viruses appear to be conserved among CRESS DNA viruses, as evidenced by both our findings and previous studies. For example, we have confirmed that the host components required for geminivirus replication also play essential roles in the genome replication of nanoviruses, another group of plant-infecting CRESS DNA viruses. Moreover, in porcine circovirus (PCV)—an animal-infecting CRESS DNA virus—replication also requires Pol α and Pol δ[305].

Notably, DNA Pol δ, rather than Pol ε, is utilized not only by viruses in the phylum CRESS DNA viruses but also by other DNA viruses for their genome replication. For example, parvoviruses, with linear ssDNA genomes that replicate via a rolling-hairpin replication mechanism, rely on Pol δ, along with

RFC, PCNA, and RPA, in reconstituted replication systems[306–309]. Likewise, the well-studied dsDNA virus SV40, which replicates its genome via a bidirectional mechanism—unlike the unidirectional strategies used in rolling-circle replication and rolling-hairpin replication—also employs Pol δ, not Pol ε, for both leading- and lagging-strand synthesis[310–313]. This recurrent usage of Pol δ across a wide range of small DNA viruses, including both ssDNA and dsDNA viruses that employ either bidirectional or unidirectional replication mechanisms, raises a fundamental question: Why do these DNA viruses preferentially recruit Pol δ rather than Pol ε for their genome replication? It is noted that all of these viruses that utilize Pol δ possess relatively small genome, each less than 8 kb in size. In addition, they generally exhibit high mutation rates. For instance, it has been shown that CRESS DNA viruses, including geminiviruses and nanoviruses, as well as parvoviruses, could evolve at rates close to those of RNA viruses[314]. Nevertheless, these viruses differ in their host range, replication mechanisms, and viral genome organization. As mentioned earlier, Pol δ has been shown to introduce more mutations during leading-strand synthesis in the context of break-induced replication. Whether these viruses exploit Pol δ's tendency to introduce mutations during certain DNA replication processes to promote their evolutionary adaptability remains an intriguing question for future investigation.

Although our study confirmed that FEN1 is dispensable for geminiviral replication, it is noteworthy that FEN1 was identified as a Rep-labelled protein in our PL-MS dataset (Table 2.2). This is likely due to its close spatial proximity to Pol δ, which plays a crucial role in geminiviral replication. This observation aligns with the proposed "toolbelt" model, in which FEN1, Pol δ, and LIG1 interact with individual monomers of the PCNA trimer during the maturation of Okazaki fragments in host plants[315]. LIG1 is the key replicative ligase that joins Okazaki fragments during DNA replication and seals nicks generated during long patch base excision repair[316,317]. Interestingly, there are some common components utilized by both pathways, including PCNA, FEN1, and LIG1. Given that our results indicate that LIG1, as well as FEN1, are not involved in geminiviral DNA replication, it suggests that (1) geminiviral replication follows a leading-strand replication mode, in a continuously manner; (2) geminiviral infection may not induce long patch base excision repair pathway.

Understanding the mechanisms of geminiviral genome replication, one of the most fundamental stages of the viral life cycle, provides a valuable framework for developing antiviral strategies that target replication-specific host factors. This approach is feasible, as studies have shown that DNA replication-associated host proteins required for geminivirus replication can serve as

targets for engineering resistance. For example, nonsynonymous single nucleotide polymorphisms (SNPs) in *POLD1* and *PRIM* have been associated with resistance to geminiviruses[193,264–266]. Notably, POLD1 variant alleles have been shown to confer resistance across a broad range of crop species[266]. Given that our study suggests CRESS DNA viruses may exploit same host replication factors, it is worth exploring whether SNP variants in replication-associated genes in different host species also confer resistance to these viruses. This can be achieved by mapping key residues in host proteins that mediate interactions with viral Rep proteins, and screening for natural allelic variants at these sites. In conclusion, our findings not only deepen the understanding of geminiviral replication mechanisms but also offer valuable insights and resources for the identification and engineering of resistance genes against geminiviruses and potentially, other CRESS DNA viruses.

3.3 Gaining insights into geminiviral replication initiation and termination

In Chapter II, our results suggest that ORC, and likely CDC6 as well, are involved in geminiviral replication. The potential roles of these two eukaryotic DNA helicase loaders may occur during the initiation phase of geminiviral replication, given their established functions in eukaryotic systems and their strong Rep-dependent binding to the geminiviral IR. In addition, the dispensability of the MCM3 subunit in geminiviral replication suggest that the MCM2-7 complex is unlikely to be involved in this process. Furthermore, hexameric predictions of Rep and the observed molecular size of Rep upon overexpression in *N. benthamiana*, support the notion that Rep—classified as an SF3 helicase[217]—likely substitutes for the host DNA helicase by assembling into a hexameric complex with helicase activity. This is reminiscent of other SF3 helicases such as the large T antigen of SV40 and NS1/Rep of parvoviruses, which function to unwind DNA during viral genome replication[218,307,318,319]. Although our study provides the first insight into the initiation of geminiviral replication, many key questions remain unanswered and warrant future exploration.

Our co-IP results suggest that TYLCV Rep interacts with both ORC1 and CDC6. However, the oligomeric state of Rep in this context, whether monomeric, dimeric, or higher-order, remains unclear. It is possible that the same form of Rep is responsible for both recruiting ORC and CDC6 to the DNA duplex and mediating sequence-specific cleavage at the onset of rolling-circle replication. This hypothesis is supported by the role of ORC in eukaryotes, where it interacts with sequence-specific proteins to initiate replication at defined chromosomal loci[139]. Rep may adopt a dimeric form during this stage, as suggested by the replication mechanism of ΦX174, in which two active tyrosine

residues, separated by three amino acids within the gpA protein, are essential for origin-specific cleavage. In contrast, only one conserved tyrosine residue within motif III of Rep has been shown to be required for site-specific nicking of rolling-circle replication[94,320,321]. Further experiments are necessary to determine the oligomeric form of Rep in different functional contexts.

Another important question concerns the precise roles of ORC and CDC6 in geminiviral replication initiation. In eukaryotes, ORC recognizes replication origins and, together with CDC6, recruits the CDT1-MCM complex to initiate DNA replication[106]. In contrast, geminiviral Rep proteins are known to recognize and bind replication origins in a sequence-specific manner[73]. Furthermore, our data indicate that overexpression of ORC does not affect Rep binding to the IR, and the binding of ORC and CDC6 to the viral genome is Rep-dependent. Together, these results suggest that ORC and CDC6 acts downstream of Rep during geminiviral replication initiation. Notably, a previous study demonstrated that ORC can promote pre-replication complex formation at a Gal4-binding site when fused to a protein containing a Gal4-binding domain[139]. ORC may use a same mechanism in geminiviruses, where it interacts with Rep to recruit CDC6 and other host factors to the geminiviral IR, thereby facilitating replication initiation. Interestingly, Epstein-Barr nuclear antigen 1 (EBNA1) of Epstein-Barr virus (EBV), a dsDNA herpesvirus that establishes persistent infection in human cells, has been shown to selectively recruit ORC to the viral origin of replication, triggering the assembly of pre-RC[322–324]. Moreover, the interaction between EBNA1 and CDC6 further enhances ORC association with the origin of replication[325].

Beyond its canonical role in loading the replicative helicase onto replication origins in eukaryotes, a recent study demonstrated that ORC also acts as a master regulator of nucleosome organization at replication origins. This function is critical for efficient chromosome replication and is mediated through the recruitment of chromatin remodelers such as INO80, ISW1a, ISW2, and Chd1[326]. Intriguingly, our TurboID-based PL-MS analysis identified INO80 and SWI_SNF complex-related proteins (SNF12 and SWI4) as host proteins labelled by Rep from both TYLCV and AbMV, raising the possibility that ORC may contribute to origin-adjacent nucleosome remodeling in geminiviral minichromosome via interaction with INO80. This hypothesis is supported by the fact that geminiviral minichromosome contains approximately 12 nucleosomes, with an open region localized at the viral origin of replication[194,300,327]. Notably, although Rep has been experimentally confirmed to bind and nick the nonanucleotide sequence within the geminiviral replication origin[73], several studies have demonstrated that Rep exhibits stronger,

sequence-specific binding to the regions upstream and/or downstream of the nicking site[277,328,329]. This observation has been linked to the potential role of Rep in transcription regulation. Intriguingly, a recent study reported that Rep does not bind to the dsDNA containing the nonanucleotide sequence, but instead binds strongly to the upstream adjacent region of the IR; nevertheless, Rep was shown to recognize and cleave the ssDNA form harbouring the replication origin site[330]. Taken together, it is plausible that Rep binding to the adjacent region of the IR contributes to nucleosome organization, potentially in coordination with ORC1, to establish a nucleosome-free region at the viral replication origin. The labelling of various nucleosome remodelling factors and chromatin assembly-related proteins by Rep, as discussed previously, further support the potential role of Rep in organizing the viral nucleosomes. To functionally validate whether ORC plays a role in nucleosome positioning during geminiviral replication, future studies should focus on generating ORC1 mutants that selectively disrupt nucleosome phasing activity—such as those targeting the Walker B motif—while preserving its ability to load the MCM complex. Such mutants would allow a mechanistic dissection of ORC's chromatin regulatory role independent of its replication origin licensing function.

Our results from the 2IR-GFP reporter system indicate that MCM3 is dispensable for geminiviral genome replication, suggesting that eukaryotic replicative helicase MCM2-7 may not participate in this process. This raises a critical question: Which factors mediates duplex DNA unwinding during geminiviral replication? Interestingly, Rep protein from certain geminivurses, including TYLCSV, MYMIV, tomato leaf curl Gujarat virus (ToLCGuV; genus *Begomovirus*, family *Geminiviridae*), and ToLCNDV, have been experimentally confirmed to possess DNA helicase activity[97,98,217,280]. To evaluate whether TYLCV Rep could function as a DNA helicase analogous to the MCM2-7 complex in eukaryotes, we predicted its structure and examined its oligomeric state following transient expression of Rep-FLAG in *N. benthamiana*. Our preliminary results suggest that TYLCV Rep likely forms a homo-hexamer. This is consistent with previous findings showing that Rep from the closely related virus TYLCSV—which shares 77% amino acid identity with TYLCV Rep—exists as approximately 5.8 monomers and is proposed to form a hexameric complex[98]. Similarly, the Rep of PCV, another member of the CRESS DNA virus family, has been shown to form a hexameric structure. Cryo-EM studies revealed that the N-terminus of PCV Rep is oriented near the dsDNA fork, while the ssDNA is translocated through the central pore of the oligomerization domain[90]. Whether TYLCV Rep exhibits DNA helicase activity and adopts a hexameric state during rolling-circle replication remains to be experimentally determined in future studies.

Several studies have shown that host replicative helicase MCM complex plays a crucial role in the genome replication of parvoviruses, a group of ssDNA viruses that replicate via a rolling-hairpin replication strategy[331–333]. Notably, Nash *et al*. (2008) reconstituted a minimal host-virus replication complex for adeno-associated virus (AAV; family *Parvoviridae*), consisting of host proteins MCM, Pol δ, RFC, and PCNA, in addition to the viral SF3 helicase NS1[308,332]. Conversely, previous study by Christensen and Tattersall (2002) showed that NS1 from minute virus of mice (MVM; family *Parvoviridae*) can serve as the replicative helicase in reconstituted in vitro system[307], though NS1 from MVM and other parvoviruses belong to SF3 helicase[334]. SV40 large T antigen was demonstrated to be a functional homologue of the eukaryotic MCM complex, functions on unwinding viral DNA at the replication fork[335]. Whether MCM2-7 subunits other than MCM3 contribute to geminiviral DNA replication remains unclear. Nonetheless, it is worth noting that the replication strategies between geminiviruses and their plant hosts are distinct, eukaryotic DNA replication is bi-directional and relies on the MCM2-7 complex, which assembles as a double hexamer and that translocates bi-directionally along the leading-strand template. This helicase forms a stable 15-subunit CMGE complex with CDC45, GINS, and Pol ε during fork progression[141]. In contrast, geminiviruses replicate their genomes via a unidirectional rolling-circle replication mechanism and rely on Pol δ rather than Pol ε. These mechanistic and component differences imply that MCM2-7 is unlikely to serve as the helicase responsible for unwinding the geminiviral DNA duplex during replication. Future investigations into the involvement of individual MCM subunits except for MCM3 in geminiviral replication will be essential to test and validate this hypothesis.

While our PL-MS data did not reveal clear evidence of host factors involved in the termination step of rolling-circle replication, the potential involvement of CDC48 remains an open question. Given its role as an unfoldase in extracting ubiquitylated MCM7 from the CMG complex to disassemble the eukaryotic replisome, CDC48 may play a similar role in terminating geminiviral replication, but this hypothesis requires further investigation.

In eukaryotic replication, termination is coupled with the disassembly of the replisome, a process marked by the ubiquitylation of MCM7—one of the core components of the CMG complex. The AAA+ ATPase CDC48 functions as an unfoldase, extracting ubiquitylated MCM7 from the CMG complex to facilitate replisome disassembly during replication termination[170–172,336,337]. However, the mechanisms underlying replisome disassembly in CRESS DNA viruses remains largely unexplored, even in the well-characterized SV40 system. In our study, we demonstrated that CDC48 plays an important role in geminiviral

replication. Notably, the identification of its co-factor Npl4 in our PL-MS data further underscores the involvement of CDC48 in geminiviral replication. Nevertheless, the precise molecular mechanisms by which CDC48 contributes to geminiviral DNA replication remain to be elucidated. Given CDC48's established function in eukaryotic replication termination and our hypothesis regarding the potential oligomeric behavior of the geminiviral Rep protein, we propose that Rep may undergo ubiquitylation prior to its disassembly. CDC48 could then act to extract ubiquitylated Rep subunits from the hexameric complex. Beyond its role in disassembling ubiquitylated MCM7, CDC48 is also critical for mediating the degradation of CDT1, a licensing factor required for loading the MCM complex onto replication origins in eukaryotic cells[94,338]. Tight regulation of CDT1 is essential, as its accumulation can lead to re-replication and genomic instability. Whether CDT1 is involved in geminiviral replication—and if CDC48 contributes to viral genome replication by modulating CDT1 stability—remains an open and intriguing question. Further studies are necessary to elucidate these potential regulatory mechanisms and define the broader role of CDC48 in geminiviral DNA replication.

In summary, our work demonstrates that geminiviruses exploit the host's leading-strand DNA synthesis machinery, but utilizing Pol δ, to complete their genome replication. Furthermore, we show that the selection of co-opted host replication factors is not only conserved within the *Geminiviridae* family but also extends more broadly across the phylum of CRESS DNA viruses. Previous studies have shown that SNPs in host factors required for geminivirus replication, including *POLD1* and *PRIM1*, could confer resistance to these viruses. This highlights the potential of our findings as a valuable resource for the development of resistance strategies through targeting such host genes. Collectively, this study advances our understanding of geminivirus replication and opens new avenues for future research.

4. Materials and Methods

4.1 Materials

4.1.1 Bacterial and viral strains

Escherichia coli strain TOP10 was used for construct transformation; *E. coli* strain DB3.1 was used for propagating plasmids containing the ccdB gene. *A. tumefaciens* strain GV3101 was used for plant transformation. Viral strain TYLCV-Almeria (GenBank accession no. AJ489258) was used for infection assays and as a template for further constructs; viral strain CpCDV (GenBank accession no. DQ458791) was used for infection assays; AbMV-A and AbMV-B (GenBank accession no. NC_001928 and NC_001929) were used for infection assays; PNYDV-R (GenBank accession no. GU553134) was used to access PNYDV DNA replication by agroinfiltration assay.

4.1.2 Antibiotics

Table 4.1 Antibiotics used in this study

Antibiotic	Abbreviation	Stock concentration	Working concentration
Zeocin	Zeo	100 mg/mL	50 µg/mL
Kanamycin	Kan	50 mg/mL	50 µg/mL
Spectinomycin	Spec	50 mg/mL	50 µg/mL
Gentamycin	Gent	50 mg/mL	50 µg/mL
Rifampicin	Rif	50 mg/mL	50 µg/mL

4.1.3 Enzymes

Phusion™ High-Fidelity DNA-Polymerase (Thermo Scientific) or Q5® High-Fidelity DNA Polymerase (NEB) was used for gene amplification; Gateway™ BP Clonase™ II Enzyme mix (Thermo Scientific) and Gateway™ LR Clonase™ II Enzyme mix (Thermo Scientific) were used for Gateway cloning; restriction enzymes, BamHI-HF®, EcoRI-HF®, and KpnI-HF® (NEB), and T4 DNA ligase (NEB) were used for generating VIGS constructs to silence desire genes.

4.1.4 Plasmids and constructs

Table 4.2 Infectious clones used in this study

Infectious clones	**Source**	**Vector**	**Description**
TYLCV WT	Rosas-Diaz *et al*., 2018	pGWB501	1.2 TYLCV monomer
TYLCV Rep defective	This study	pGWB501	1.2 TYLCV monomer with Rep null mutant (M1-stop; L14-stop; E53-stop; E63-stop)
AbMV-A	Sicking and Krenz, 2022	pGreen	1.2 AbMV A component
AbMV-B	Sicking and Krenz, 2022	pGreen	Full tandem repeat of AbMV B component
CpCDV WT	Generated by Dr. Yu Zhou in our laboratory (unpublished).	pGWB501	1.2 CpCDV monomer

Table 4.3 Expression constructs used in this study

Expression construct	Source	Vector	Species	Identifier
35S:RFC1-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf01382g04008
35S:POLD3-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf00160g10002
35S:TOPO1-GFP	This study	pGWB505	*N. benthamiana*	NbS00018030g0025
35S:GFP-POLD1	This study	pGWB506	*N. benthamiana*	Niben101Scf02230g03027
35S:RPA70A-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf00567g04007
35S:PCNA-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf10384g02008
35S:GFP-CDC6	This study	pGWB506	*N. benthamiana*	Niben101Scf06541g00002
35S:ORC1.1-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf02055g00001
35S:ORC1.2-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf02706g02004
35S:CDC48-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf01911g15015
35S:RuvB-GFP	This study	pGWB505	*N. benthamiana*	Niben101Scf01281g02006
35S:RFC1-SR-FLAG	This study	pGWB511	*N. benthamiana*; SR: silencing resistant version	Niben101Scf01382g04008
35S:POLD3-SR-FLAG	This study	pGWB511	*N. benthamiana*; SR: silencing resistant version	Niben101Scf00160g10002
35S:TOPO1-SR-FLAG	This study	pGWB511	*N. benthamiana*; SR: silencing resistant version	NbS00018030g0025
35S:FLAG-POLD1-SR	This study	pGWB512	*N. benthamiana*; SR: silencing resistant version	Niben101Scf02230g03027
35S:PCNA-SR-FLAG	This study	pGWB511	*N. benthamiana*; SR: silencing resistant version	Niben101Scf10384g02008
35S:FLAG-CDC6	This study	pGWB512	*N. benthamiana*	Niben101Scf06541g00002
35S:ORC1.1-FLAG	This study	pGWB511	*N. benthamiana*	Niben101Scf02055g00001
35S:ORC1.2-FLAG	This study	pGWB511	*N. benthamiana*	Niben101Scf02706g02004
35S:CDC48-SR-FLAG	This study	pGWB511	*N. benthamiana*; SR: silencing resistant version	Niben101Scf01911g15015
35S:RuvB-SR-FLAG	This study	pGWB511	*N. benthamiana*; SR: silencing resistant version	Niben101Scf01281g02006
TRV2-GUS	This study	pTRV2	ß-glucoronidase from *E. coli*	946713
TRV2:RFC1	This study	pTRV2	*N. benthamiana*	Niben101Scf01382g04008
TRV2:POLD3	This study	pTRV2	*N. benthamiana*	Niben101Scf00160g10002
TRV2:TOPO1	This study	pTRV2	*N. benthamiana*	NbS00018030g0025
TRV2:POLD1	This study	pTRV2	*N. benthamiana*	Niben101Scf02230g03027
TRV2:PCNA	This study	pTRV2	*N. benthamiana*	Niben101Scf10384g02008
TRV2:FEN1	This study	pTRV2	*N. benthamiana*	Niben101Scf05932g05010/ Niben101Scf02486g03012
TRV2:LIG1	This study	pTRV2	*N. benthamiana*	Niben101Scf19975g01014/ Niben101Scf00436g03006
TRV2:ORC1	This study	pTRV2	*N. benthamiana*	Niben101Scf02055g00001
TRV2:CDC48	This study	pTRV2	*N. benthamiana*	Niben101Scf01911g15015
TRV2:RuvB	This study	pTRV2	*N. benthamiana*	Niben101Scf01281g02006
35S:GUSi	This study	pK7GWIWG	ß-glucoronidase from *E.*	946713

			coli	
35S:LIG1.2i	This study	pK7GWIWG	*N. benthamiana*	Niben101Scf00436g03006
PNYDV Rep	Grigoras *et al*., 2018	pBin	PNYDV DNA-R replicon	GU553134
35S:Rep-GFP	Ding *et al*., 2019	pGWB505	TYLCV	CAD33251
35S:GFP-Rep	Ding *et al*., 2019	pGWB506	TYLCV	CAD33251
35S:Rep	Generated by Dr. Xue Ding in our laboratory (unpublished)	pGWB502	TYLCV	CAD33251
35S:Rep-FLAG	This study	pGWB502	TYLCV	CAD33251
35S:Rep-TurboID	This study	pGWB502	TYLCV	CAD33251
35S:TurboID-Rep	This study	pGWB502	TYLCV	CAD33251
35S:AC1-TurboID	This study	pGWB502	AbMV	NP_047218.2
35S:TurboID-AC1	This study	pGWB502	AbMV	NP_047218.2
35S:RepF13S-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepP36Q-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepP36A-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepL59R-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepL59V-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepL59F-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepQ90P-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepY101C-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepS138L-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepS138P-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepS138T-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepL145A-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepR46I-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepG52E-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepV113F-GFP	This study	pGWB505	TYLCV	CAD33251
35S:RepL145A-GFP	This study	pGWB505	TYLCV	CAD33251
TYLCV RepP36A	This study	pGWB501	TYLCV	AJ489258
TYLCV RepL59F	This study	pGWB501	TYLCV	AJ489258
TYLCV RepQ90P	This study	pGWB501	TYLCV	AJ489258
TYLCV RepS138T	This study	pGWB501	TYLCV	AJ489258

4.1.5 Oligonucleotides

Table 4.4 Oligonucleotides used in this study

Amplification target	**Source**	**Sequence 5'-3'**	**Purpose**	**Identifier**
Oligonucleotides to generate TurboID-containing fusion protein				
TOPO-*TurboID+Rep*	This study	F:CACCATGAAAGACAATACTGTGCCTC	For cloning *TurboID-Rep* to pDONRTM/D-TOPO, 1068 bp	N/A
		R:TTACGCCTTATTGGTTTCTTCTTGG		
pDONR/Zeo-	This	F:GGGGACAAGTTTGTACAAAAAAG	For cloning *Rep-TurboID*	N/A

Rep+TurboID	study	CAGGCTTCATGCCTCGTTTATTTAAAAT R:GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTTTTCGGCAGACCGCAGAC	to pDONR™/Zeo, 1068 bp	
pDONR/Zeo-*TurboID*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTACCCGTATGATGTTCC R:GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTTTTCGGCAGACCGCAGAC	For cloning *TurboID* harbouring HA tag to pDONR™/Zeo, 1056 bp	N/A
Oligonucleotides to clone in pDONR™/Zeo				
pDONR/Zeo-*RFC1*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTCAGATATCAGAAAATGGT R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACCTCTTTCTCTTGGTAG R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCCCTCTTTCTCTTGGTAG	Cloning of *N. benthamiana RFC1* into pDONR™/Zeo: ST (with stop codon, 2997 bp) and NS (no stop codon, 2994 bp)	Niben101Scf01382g04008
pDONR/Zeo-*POLD3*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGCGGTTGACATGGAA R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAAGCCTTCTTGAAAAATGATAG R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCAGCCTTCTTGAAAAATGATAG	Cloning of *N. benthamiana POLD3* into pDONR™/Zeo: ST (with stop codon, 1545 bp) and NS (no stop codon, 1542 bp)	Niben101Scf00160g10002
pDONR/Zeo-*TOPO1*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGCTGTTGAGGCTTTTTC R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGAATGTGAAGCTTGGTT R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCGAATGTGAAGCTTGGTT	Cloning of *N. benthamiana TOPO1* into pDONR™/Zeo: ST (with stop codon, 2634 bp) and NS (no stop codon, 2631 bp)	NbS00018030g0025
pDONR/Zeo-*POLD1*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGAACAACAGCAAATCGAG R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAATATCTCCTGTCATCTAATCT R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCATATCTCCTGTCATCTAATCT	Cloning of *N. benthamiana POLD1* into pDONR™/Zeo: ST (with stop codon, 3270 bp) and NS (no stop codon, 3267 bp)	Niben101Scf02230g03027
pDONR/Zeo-*PCNA*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTTGGAATTACGGCTTGTTC R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAGGTTTAGTTTCCTCTTCATCC R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCAGGTTTAGTTTCCTCTTCATCC	Cloning of *N. benthamiana PCNA* into pDONR™/Zeo: ST (with stop codon, 795 bp) and NS (no stop codon, 792 bp)	Niben101Scf10384g02008
pDONR/Zeo-*RPA70A*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGCCTGTGAACTTGACTG R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAGAAGGACGGAATTTTGAA R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCAGAAGGACGGAATTTTGAA	Cloning of *N. benthamiana RPA70* into pDONR™/Zeo: ST (with stop codon, 1917 bp) and NS (no stop codon, 1914 bp)	Niben101Scf00567g04007
pDONR/Zeo-	This	F:GGGGACAAGTTTGTACAAAAAAG	Cloning of *N.*	Niben101Scf

CDC6	study	CAGGCTCCATGCCGACAATCCCCGTC R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGATGTTCTTGTTCCTTCTATCC R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCGATGTTCTTGTTCCTTCTATCC	*benthamiana CDC6* into pDONR™/Zeo: ST (with stop codon, 1602 bp) and NS (no stop codon, 1599 bp)	06541g00002
pDONR/Zeo-*ORC1.1*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGCAGAAACCCCAAAGAAG R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACGCAGCTCGCCTACATATTT R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCCGCAGCTCGCCTACATATTT	Cloning of *N. benthamiana ORC1.1* into pDONR™/Zeo: ST (with stop codon, 2037 bp) and NS (no stop codon, 2034 bp)	Niben101Scf02055g00001
pDONR/Zeo-*ORC1.2*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGTTGAATGTGATGAATGCT R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATAGATACTTGGCCAACCAT R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGATACTTGGCCAACCAT	Cloning of *N. benthamiana ORC1.2* into pDONR™/Zeo: ST (with stop codon, 1929 bp) and NS (no stop codon, 1926 bp)	Niben101Scf02706g02004
pDONR/Zeo-*CDC48*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGAGTCACCAGGCCGAGT R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGCTATACAAGTCATCATCATCAGCT R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCGCTATACAAGTCATCATCATCAGCT	Cloning of *N. benthamiana CDC48* into pDONR™/Zeo: ST (with stop codon, 2418 bp) and NS (no stop codon, 2415 bp)	Niben101Scf01911g15015
pDONR/Zeo-*RuvB*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGCGGAGGTAAAAATCTCAG R_ST:GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGAGACCATGGCAGTAG R_NS:GGGGACCACTTTGTACAAGAAAGCTGGGTCGGAGACCATGGCAGTAG	Cloning of *N. benthamiana RuvB* into pDONR™/Zeo: ST (with stop codon, 1404 bp) and NS (no stop codon, 1401 bp)	Niben101Scf01281g02006
Oligonucleotides to clone into VIGS (pTRV2)/RNAi (pK7GWIWG(II)) constructs				
TRV2:*GUS*	This study	F:GGAATTCTGTGGGTCAATAATCAGGAAGTGAT R:GGGGTACCCGCGTGGTTACAGTCTTGC	For cloning 311 bp *GUS* fragment to pYL156 (TRV2)	946713
pK7GWIWG(II): *GUS*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTTACGTCCTGTAGAAACC R:GGGGACCACTTTGTACAAGAAAGCTGGGTCTTGTTCACACAAACGGTGATAC	For cloning 407 bp *GUS* fragment to pK7GWIWG(II)	946713
TRV2:*RFC1*	This study	F:GGAATTCCCCTCAGCTGATAAGGACACC R:GGGGTACCCCTCGTTTTCTTGCTAACAGATCCT	For cloning 300 bp *RFC1* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf01382g04008
TRV2:*POLD3*	This study	F:GGAATTCCATGGCGGTTGACATGGAAAC R:GGGGTACCCCGCTATATACCTGAACTGAGCAGT	For cloning 300 bp *POLD3* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf00160g10002

TRV2:*TOPO1*	This study	F:GGAATTCCGTAAAGTCCGCTGCTCAG R:GGGGTACCCCCAATGTAGTCCATTTTTGCCCTT	For cloning 300 bp *TOPO1* fragment from *N. benthamiana* to pYL156 (TRV2)	NbS00018030g0025
TRV2:*POLD1*	This study	F:GGAATTCGGATCTGAGCAAGGAACATATG R:GGGGTACCAAATCTGCTTTTGCTCTTTTGC	For cloning 299 bp *POLD1* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf02230g03027
TRV2:*PCNA*	This study	F:GGAATTCCATGTTGGAATTACGGCTTGTTCAG R:GGGGTACCCCAGTGACGGTGTCACTGCC	For cloning 300 bp *PCNA* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf10384g02008
TRV2:*FEN1*	This study	F:GGAATTCATCCCCAAAATGGGAATTAAGGGT R:GGGGTACCCTCTGTAGCCTTAGCCAAATCAT	For cloning 300 bp *FEN1* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf05932g05010 Niben101Scf02486g03012
TRV2:*LIG1*	This study	F: GGAATTCGACGCGATTTCGAAGGAGTC R:CGGGATCCTGGTTTTCGCATCAGAGGCT	For cloning 300 bp *LIG1* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf19975g01014 Niben101Scf00436g03006
pK7GWIWG(II): *LIG1.2*	This study	F:GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTTGCCGTTGCGCTCAT R:GGGGACCACTTTGTACAAGAAAGCTGGGTCCTCTCACCATCCCTCCAATATGC	For cloning 476 bp *LIG1.2* fragment from *N. benthamiana* to pK7GWIWG(II)	Niben101Scf00436g03006
TRV2:*ORC1*	This study	F: GGAATTCACCGCCACTACCTTCG R: CGGGATCCGCATCTTCTCTTCTCTT	For cloning 300 bp *ORC1* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf02055g00001
TRV2:*CDC48*	This study	F:GGAATTCAGGAAAGGGGACAATTTCC R:GGGGTACCAGGGGGTCCATAAAGTAAAATC	For cloning 300 bp *CDC48* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf01911g15015
TRV2:*RuvB*	This study	F:GGAATTCGCACTTATGCAGGCATTTCG R:CGGGATCCCTTAGTCTGGGGTCCCATAG	For cloning 321 bp *RuvB* fragment from *N. benthamiana* to pYL156 (TRV2)	Niben101Scf01281g02006
Oligonucleotides for qPCR and qRT-PCR				
NbRFC1	This study	F: ATATTTGGGAAGGGCAAACC R: CAGCTTTCACAACAGGCTGA	For qPCR (224 bp); primer efficiency (E = 89.3 %)	Niben101Scf01382g04008
NbPOLD3	This study	F: CATGTGGTTGATGCCAAAAG R: AGCCGAAACACTGTCAGCTT	For qPCR (207 bp); primer efficiency (E = 91 %)	Niben101Scf00160g10002
NbTOPO1	This study	F: GGGAATGAGAAGGATGACGA R: GCTCATACACAGCCACCTCA	For qPCR (151 bp); primer efficiency (E = 107.5 %)	NbS00018030g0025
NbPOLD1	This study	F: CAGATTCGGTGATGGTGCAG R: TTGTCCATAGAAGGCCAGCA	For qPCR (180 bp); primer efficiency (E = 96.2 %)	Niben101Scf02230g03027
NbPCNA	This study	F: CAAGCCTGAAGAAGCCACAG R: AGCCCATCTCAGCAATCTTG	For qPCR (170 bp); primer efficiency (E = 102 %)	Niben101Scf10384g02008
NbFEN1	This study	F: GCTAGACGGCTCTTCAAGGA R: CGTCCCTGAGAAGACTTGGT	For qPCR (182 bp); primer efficiency (E = 109.6 %)	Niben101Scf05932g05010 Niben101Scf

				02486g03012
NbLIG1.1	This study	F: TGTTGGAGGTGAAATGGTGC R: GTAAGTTCCCGCTACGCATG	For qPCR (214 bp); primer efficiency (E = 94.3 %)	Niben101Scf 19975g01014
NbLIG1.2	This study	F: ATTCGCCTCCTCCAGACAAA R: GCTTCCAAACACCATCAGCT	For qPCR (196 bp); primer efficiency (E = 106.1 %)	Niben101Scf 00436g03006
NbORC1	This study	F: TCCTGAAAAGTTGCTTCCGC R: CCAACCTTCCTTGACGCAAA	For qPCR (162 bp); primer efficiency (E = 118.7 %)	Niben101Scf 02055g00001
NbCDC48	This study	F: GGAGAGAGTGAGGCCAATGT R: TGTCAGTTGGTTCAGGA	For qPCR (159 bp); primer efficiency (E = 109.7 %)	Niben101Scf 01911g15015
NbRuvB	This study	F: GCGGAAAGGGAAGATTGTGG R: CAGGAGACCATGGCAGTAGT	For qPCR (168 bp); primer efficiency (E = 104 %)	Niben101Scf 01281g02006
25SrRNA (ITS)	Rosas-Diaz *et al*., 2018	F: ATAACCGCATCAGGTCTCCA R: CCGAAGTTACGGATCCATTT	For qPCR (90 bp); primer efficiency (E = 101.1 %)	X13557.1
NbActin	Maimbo *et al*., 2010	F: CGGAATCCACGAGACTACATAC R: GGGAAGCCAAGATAGAGC	For qPCR (230 bp); primer efficiency (E = 108 %)	X69885
Rep (TYLCV)	Wang *et al*., 2017	F: TGAGAACGTCGTGTCTTCCG R: TGACGTTGTACCACGCATCA	For viral accumulation and ChIP-qPCR (176 bp); primer efficiency (E = 94.8 %)	CAD3325
C2 (TYLCV)	Wang *et al*., 2017	F: ACCTTCGTCACCCTCTACGA R: AAACGCCATTCTCTGCCTGA	For ChIP-qPCR (198 bp); primer efficiency (E = 93.5 %)	CAD33250
C3 (TYLCV)	Wu *et al*., 2021	F: TGGACGACATTACAGCCTCA R: ACAATACATGATCAACTGCTCTGA	For ChIP-qPCR (219 bp); primer efficiency (E = 94.3 %)	CAD33249
C4 (TYLCV)	Rosas-Diaz *et al*., 2018	F: ATCCGAACATTCAGGCAGCT R: TGCTGACCTCCTCTAGCTGA	For ChIP-qPCR (121 bp); primer efficiency (E = 90.9 %)	CAD33252
CP (TYLCV)	Wang et al., 2017	F: TGGAAGCAGCCCAATGGATT R: GTTCTCGTACTTGGCTGCCT	For ChIP-qPCR (229 bp); primer efficiency (E = 91.6 %)	CAD33248
V2 (TYLCV)	Wang *et al*., 2017	F: ATCTGTTGTAAGGGCCCGTG R: CTTTCGGTACATGGGCCTGT	For ChIP-qPCR (182 bp); primer efficiency (E = 92.7 %)	CAD33247
IR-GFP	This study	F: TGTTCCATGGCCAACACTTG R: ACGTGTCTTGTAGTTCCCGT	For qPCR (166 bp) in reporter system; primer efficiency (E = 102.8 %)	N/A
IR_long	This study	F: GTGGTTCCCCATTCTCGTG R: CCAAATAGCCATTAGGTGTCCAG	For ChIP-qPCR (250 bp), used in the reporter system and viral infection; primer efficiency (E = 98.1 %)	N/A
IR_short	This study	GAATCGGTGTCCCTCAAAGC R: CCAAATAGCCATTAGGTGTCCAG	For ChIP-qPCR (79 bp), used in the reporter system and viral infection; primer efficiency (E = 99.4 %)	N/A
Oligonucleotides for Rep mutagenesis				
RepF13S	This study	F:TATATGCCAAAAATTATTCCCTAACATATCCCAATTG R:CAATTGGGATATGTTAGGGAATAATTTTTTGGCATATA	Introduction of a point mutation in Rep: TTC→TCC (F13→S13)	GU553134
RepP36Q	This study	F:AAAAACCTAGAAACCCAAACAAATAAAAAATACATC R:GATGTATTTTTTATTTGTTTGGGTTTCTAGGTTTTT	Introduction of a point mutation in Rep: CCA→CAA (P36→Q36)	GU553134
RepP36A	This study	F:AAAAACCTAGAAACCGCAACAAATAAAAAATACATC	Introduction of a point mutation in Rep:	GU553134

		R:GATGTATTTTTTATTTGTTGCGGTTTCTAGGTTTTT	CCA→GCA (P36→A36)	
RepL59R	This study	F:GAACCACATCTCCATGTGCGTATCCAATTCGAAGGCAAAT R:ATTTGCCTTCGAATTGGATACGCACATGGAGATGTGGTTC	Introduction of a point mutation in Rep: CTT→CGT (L59→R59)	GU553134
RepL59V	This study	F:GAACCACATCTCCATGTGGTTATCCAATTCGAAGGCAAAT R:ATTTGCCTTCGAATTGGATAACCACATGGAGATGTGGTTC	Introduction of a point mutation in Rep: CTT→GTT (L59→V59)	GU553134
RepL59F	This study	F:GAACCACATCTCCATGTGTTTATCCAATTCGAAGGCAAAT R:ATTTGCCTTCGAATTGGATAAACACATGGAGATGTGGTTC	Introduction of a point mutation in Rep: CTT→TTT (L59→F59)	GU553134
RepQ90P	This study	F:GCACATTTCCATCCGAACATTCCGGCAGCTAAGAGCTCAACAGATG R:CATCTGTTGAGCTCTTAGCTGCCGGAATGTTCGGATGGAAATGTGC	Introduction of a point mutation in Rep: CAG→CCG (Q90→P90)	GU553134
RepY101C	This study	F:ACAGATGTCAAGACCTGCGTGGAGAAAGACGGAGAC R:GTCTCCGTCTTTCTCCACGCAGGTCTTGACATCTGT	Introduction of a point mutation in Rep: TAC→TGC (Y101→C101)	GU553134
RepS138L	This study	F:ATGCCGAAGCACTCAATTTAGGCAATAAATCCGAGGCC R:GGCCTCGGATTTATTGCCTAAATTGAGTGCTTCGGCAT	Introduction of a point mutation in Rep: TCA→TTA (S138→L138)	GU553134
RepS138P	This study	F:ATGCCGAAGCACTCAATCCAGGCAATAAATCCGAGGCC R:GGCCTCGGATTTATTGCCTGGATTGAGTGCTTCGGCAT	Introduction of a point mutation in Rep: TCA→CCA (S138→P138)	GU553134
RepS138T	This study	F:ATGCCGAAGCACTCAATACAGGCAATAAATCCGAGGCC R:GGCCTCGGATTTATTGCCTGTATTGAGTGCTTCGGCAT	Introduction of a point mutation in Rep: TCA→ACA (S138→T138)	GU553134
RepL145A	This study	F:AGGCAATAAATCCGAGGCCGCCAATATATTAAAAGAGAAGGCC R:GGCCTTCTCTTTTAATATATTGGCGGCCTCGGATTTATTGCCT	Introduction of a point mutation in Rep: CTC→GCC (L145→A145)	GU553134
RepR46I	This study	F:CAAAGTTTGCAtAGAACTCCACGAGAAT R:ATGTATTTTTTATTTGTTGGGGTTTCTAGGT	Introduction of a point mutation in Rep: AGA→ATA (R46→I46)	GU553134
RepG52E	This study	F:TCCACGAGAATGaGGAACCACATCT R: GTTCTCTGCAAACTTTGATGT	Introduction of a point mutation in Rep: GGG→GAG (G52→E52)	GU553134
RepV113F	This study	F:CATTGATTTTGGAtTTTTCCAAATCGATG R: AAGTCTCCGTCTTTCTCCAC	Introduction of a point mutation in Rep: GTT→TTT (V113→F113)	GU553134
RepY285H	This study	F: ACACAAAGcACGGGAAGCCCAT R: TGCTTTGCCAGTCCCTCT	Introduction of a point mutation in Rep: TAC→CAC (Y285→H285)	GU553134

4.1.6 Plant materials

Both wild-type (WT) and 2IR-GFP transgenic *N. benthamiana* plants[249,270,271] were used in this study.

Computer programs

The following computer programs were used in this study:

- Microsoft Excel 2021
- Microsoft Word 2021
- Microsoft PowerPoint 2021
- Snapgene 6.0.2
- Graphpad prism 9.5.1
- Zotero 7.0.11
- Bio-Rad CFX Maestro 1.0
- Mega 11/12
- Inkscape 1.3.2
- PyMOL 2.0

4.2 Methods

4.2.1 Plant growth

N. benthamiana plants were grown in a controlled growth room under long-day conditions (16 hours light/8 hours dark) at 22 °C and 35% humidity.

4.2.2 Cloning

All constructs used in this study were generated using one of two cloning methods: Gateway cloning (when cloning into plasmids of the pGWB series or pK7GWIWG) and traditional restriction/ligation cloning (when cloning into the pTRV2 vector).

1. Gateway cloning

a. The genes of interest (GOIs) were amplified with gene-specific primers containing the attB sites for the Gateway system using High-Fidelity DNA Polymerase (Phusion™ or Q5®), and the PCR products were purified with GeneJET PCR purification Kit (Thermo Scientific), following the manufacturer's instructions.

b. The GOIs were cloned into a Gateway entry vector (pDONR™/Zeo or pDONR™/D-TOPO, Thermo Scientific) through a Gateway BP reaction, following the manufacturer's instructions.

c. The BP reaction was then transformed into *E. coli* competent cells (TOP10) following the heat-shock transformation protocol. The cells were grown on selective LB plates containing the appropriate antibiotics, and incubated at 37°C overnight.

d. Transformants were confirmed by colony PCR, and a positive colony was isolated and grown overnight in liquid culture. Miniprep was subsequently performed, and the cloned sequence was verified via Sanger sequencing.

e. The GOI from the pDONR vector was subcloned into the desired expression binary vector (e.g., plasmids of the pGWB series or pK7GWIWG, as shown in Table 5.3) via a Gateway LR reaction, following the manufacturer's instructions. Subsequent experiments were conducted to verify the sequence accuracy of the constructs in the expression vector, as described in step d above.

2. Restriction/ligation cloning

a. The 300 bp virus-induced gene silencing (VIGS) fragment was amplified using High-Fidelity DNA polymerase (Phusion ™ or Q5®), with gene-specific primers containing the corresponding restriction enzyme sites. PCR products were then purified with GeneJET PCR purification Kit (Thermo Scientific), following the manufacturer's instructions. The VIGS fragments were designed using the online tool available at https://vigs.solgenomics.net/.

b. Both the VIGS fragment and pTRV2 vector were digested with double enzymes using the online tool available at https://nebcloner.neb.com/#!/redigest, and the digested products were purified with GeneJET PCR purification Kit (Thermo Scientific), following the manufacturer's instructions.

c. The double enzyme-digested VIGS fragment was cloned into pTRV2 vector via a T4 ligation (NEB) reaction, following the manufacturer's instructions.

d. The T4 ligation products were transformed into *E. coli* competent cells (TOP10) using the heat-shock transformation protocol. The cells were grown on selective LB plates containing the appropriate antibiotics and incubated at 37°C overnight. Subsequent experiments were conducted and the cloned sequence was verified via Sanger sequencing.

4.2.3 Site-directed mutagenesis

All mutations in either Rep or TYLCV infectious clone were generated using the QuickChange Site-Directed Mutagenesis Kit (Agilent) with gene-specific primers using QuickChange Primer design. The mutations were confirmed by Sanger sequencing. Subsequently, the Rep or TYLCV mutants in the entry vector were subcloned into expression vectors from the pGWB series via a Gateway LR reaction, following the manufacturer's protocol.

4.2.4 Overlapping PCR

Overlapping PCR was used to generate all TurboID-containing constructs, as described in Heckman and Pease, 2007. The overlapping fragments were then cloned into pDONR vector, and subcloned into the expression vectors through a Gateway LR reaction.

4.2.5 *Agrobacterium*-mediated transformation

The expression constructs generated in Section 5.2.2-5.2.4 were transformed into *A. tumefaciens* strain GV3101 competent cells using the freeze/thaw method. The transformed cells were grown on solid LB plates containing the appropriate antibiotics, and incubated at 28°C for 2 days. The transformants were confirmed by colony PCR. The positive transformants were stored at -80°C for future experiments.

4.2.6 Virus-induced gene silencing (VIGS)

a. Tobacco rattle virus (TRV)-mediated virus-induced gene silencing (VIGS) assays were performed as described by Wu *et al*. (2021)[192]. In brief, *Agrobacterium* cells containing pTRV1 (helper plasmid), pTRV2-GUS (negative control), or constructs cloned in pTRV2 (Table 5.3) were grown in selective LB liquid culture containing the appropriate antibiotics at 28 °C overnight.

b. Bacterial cells were collected by centrifugation at 4000 rpm for 10 min at room temperature (RT) and resuspended in the infiltration buffer (10 mM $MgCl_2$, 10 mM MES, pH 5.6, and 150 µM acetosyringone).

c. The optical density (OD_{600}) of the *Agrobacterium* cells containing pTRV1, pTRV2-GUS, or constructs cloned in pTRV2 was adjusted to 0.2. *Agrobacterium* cells containing either pTRV2-GUS or constructs cloned in pTRV2 were mixed with pTRV1 at a 1:1 ratio. The mixture was then incubated in the dark for 2-4 hours prior to infiltration.

d. The above *Agrobacterium* culture was infiltrated into 2-week-old 2IR-GFP transgenic or WT *N. benthamiana* seedlings (see Section 5.2.8 for details) using a 1 mL needleless syringe.

e. The plants were kept in the controlled growth room for 2 weeks before downstream analyses.

4.2.7 Visualization of protein subcellular localization

Confocal microscopy was used to visualize the subcellular localization in transiently transformed *N. benthamiana* leaves. Briefly, *Agrobacterium* cells containing the constructs with GFP-tagged proteins were mixed with RFP-tagged proteins at a 1:1 ratio. The mixture was infiltrated into fully expanded leaves from 4-week-old *N. benthamiana* plants with a 1 mL needleless syringe. The samples were collected at 30-48 hours post-infiltration and imaged with a Zeiss LSM 880 upright confocal microscope. The preset settings for GFP and RFP are Ex: 488 nm, Em: 500–550 nm and Ex: 554 nm, Em: 570-620 nm, respectively.

4.2.8 Local and systemic viral infection

Local and systemic viral infection were performed as described by Wu *et al*. (2021)[192]. Briefly, for local infection, the *Agrobacterium* cells containing the infectious clones were infiltrated into WT *N. benthamiana* plants and samples were harvested at 3 days post-inoculation. For systemic infection, *Agrobacterium* cells containing geminiviruses and pTRV1 and pTRV2-based constructs were co-infiltrated into 2-week-old *N. benthamiana* seedlings. Samples were harvested at 14 days post-TRV inoculation.

4.2.9 qPCR and qRT-PCR

DNA and RNA were extracted as described by Oñate-Sánchez and Vicente-Carbajosa (2008)[339]. To detect gene expression in *N. benthamiana* plant tissue, 5 µg of crude RNA was digested with DNase I (Thermo Scientific) to remove genomic DNA. cDNA synthesis was then performed using PrimeScript RT Master Mix (Takara) following the manufacturer's instructions. *NbActin* was used to as the reference gene[340]. For DNA samples from local viral infections, *Dpn*I (Thermo Scientific) was used to digest plasmid DNA derived from *Agrobacterium* prior to qPCR, in order to eliminate the influence of plasmid-derived viral genomes on the results. The 25S ribosomal DNA interspacer (*ITS*) was used as the reference gene[341]. PowerTrack™ SYBR Green Mastermix (Thermo Scientific) was used for the qPCR assay in a BioRad CFX384 real-time system. The procedure was performed as follows: initial denaturation at 95 °C for 3 minutes, followed by 40 cycles of 95 °C for 15 seconds and 60 °C for 30 seconds. At least two technical replicates were performed for each experiment. qPCR primers were designed using the online tool Primer 3, and are shown in Table 5.4.

4.2.10 ChIP assay

ChIP assays were performed as described by Nie *et al*. (2019)[342] with minor modifications. *Agrobacterium* cells carrying the desired constructs to express Flag-tagged fusion protein (as shown in Table 5.3) were co-infiltrated with either Rep-RFP or TYLCV infectious clone into 2IR-GFP transgenic or WT *N. benthamiana* plants, respectively. Samples were harvested at 2 days post-infiltration. Crosslinking was performed using 1% formaldehyde in 1 x PBS by vacuum infiltration. The crosslinking reaction was then quenched with 0.125 M glycine in 1 x PBS. The following steps were carried out as described by Nie *et al*. (2019)[342].

For the IP step, 2 µg of Anti-FLAG (SICGEN) or IgG (Sigma) were used. DNA was purified using the MinElute PCR Purification Kit (QIAGEN, No. 28004). ChIP input and IP products were diluted to a 1:1000 and 1:10 ratio, respectively.

1 µL was taken for qPCR analysis. The primers used in this experiment are listed in Table 5.4.

4.2.11 Protein extraction and co-IP

Agrobacterium cells carrying the desired constructs were infiltrated into 4-week-old expanded *N. benthamiana* leaves, and samples were taken at 2 days post-infiltration. The samples were ground into a fine powder with liquid nitrogen. For proteins with exclusively nuclear localization, Honda buffer was used to extract the nuclei. Protein extraction and co-IP with GFP-Trap agarose beads (Chromotek, GTA-20) were performed as described by Wu *et al*. (2021)[273].

The antibodies used were as follows: anti-GFP (SICGEN, AB0020-500; dilution ratio: 1:5000), anti-RFP (Chromotek, 6G6; dilution ratio: 1:5000), anti-goat IgG (Sigma, A8919-2ML; dilution ratio: 1:15,000), and anti-mouse IgG (Sigma, A2554; dilution ratio: 1:15,000).

4.2.12 TurboID-based PL

TurboID-based PL was performed as described by Zhang *et al*. (2019)[267]. Prior to the preparation of samples for MS, the biotin ligase activity of TurboID-containing constructs was tested. In brief, *Agrobacterium* cells harboring TurboID-fused protein of interest and control constructs were infiltrated into fully expanded leaves of 3- to 4-week-old *N. benthamiana* plants using a 1 mL needleless syringe. Biotin solution (50 µM biotin and 2.5 mM $MgCl_2$) or control solution (same amount of DMSO and 2.5 mM $MgCl_2$) was infiltrated into the same leaf tissues at 42 hours post-infiltration. Samples were harvested six hours after biotin treatment. Immunoblot assay was then performed to detect the protein accumulation and biotinylation activity with anti-HA and streptavidin-HRP antibodies, respectively.

Sample preparation for MS was performed as described by Zhang *et al*. (2019) with minor modifications. Briefly, biotin-treated samples were harvested as described above and ground to a fine power in liquid nitrogen, and protein extraction (as described in Section 5.2.11) and removal of free biotin were carried out. A 100 µL-aliquot of lysate was taken from each sample before and after removal of free biotin. The lysates were incubated with streptavidin-C1-conjugated magnetic beads (Invitrogen, 65001) at 4°C for 3-4 hours on a rotator to enrich for biotinylated proteins. In the final wash step, 10% of the suspension was taken to confirm the successful enrichment of biotinylated proteins. Immunoblot analysis of protein enrichment and biotinylation was performed on all lysates taken from the different processing steps.

The antibodies used in this experiment are as follows: Streptavidin-HRP

(Abcam, Ab7403; dilution ratio: 1:15,000), anti-HA (Roche, 12013819001; dilution ratio: 1: 5000), anti-Rat IgG (Abcam, Ab7097; dilution ratio: 1:15,000).

4.2.13 Statistical analysis

For qPCR and RT-qPCR analyses, the statistical differences between two samples were evaluated using Student's t-test, with a significance level (α) set at 0.05. To compare the statistical differences between more than two samples, one-way ANOVA was used. Asterisks are used to indicate significant statistical difference (P>0.05: no significance (ns); P≤0.05: *; P≤0.01: **; P≤0.001: ***; P≤0.0001: ****.

5. References

1. Rosario, K., Duffy, S. & Breitbart, M. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch. Virol.* **157**, 1851–1871 (2012).
2. Ramos, E. D. S. F. *et al.* Characterization of CRESS-DNA viruses in human vaginal secretions: An exploratory metagenomic investigation. *J. Med. Virol.* **96**, e29750 (2024).
3. Chandler, M. *et al.* Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat. Rev. Microbiol.* **11**, 525–538 (2013).
4. Krupovic, M. *et al. Cressdnaviricota* : a Virus Phylum Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes. *J. Virol.* **94**, e00582-20 (2020).
5. Desingu, P. A. & Nagarajan, K. Genetic Diversity and Characterization of Circular Replication (Rep)-Encoding Single-Stranded (CRESS) DNA Viruses. *Microbiol. Spectr.* **10**, e01057-22 (2022).
6. Zhao, L., Rosario, K., Breitbart, M. & Duffy, S. Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) Viruses: Ubiquitous Viruses With Small Genomes and a Diverse Host Range. in *Advances in Virus Research* vol. 103 71–133 (Elsevier, 2019).
7. Kazlauskas, D., Varsani, A., Koonin, E. V. & Krupovic, M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat. Commun.* **10**, 3425 (2019).
8. Krupovic, M. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Curr. Opin. Virol.* **3**, 578–586 (2013).
9. Turlewicz-Podbielska, H., Augustyniak, A. & Pomorska-Mól, M. Novel Porcine Circoviruses in View of Lessons Learned from Porcine Circovirus Type 2-Epidemiology and Threat to Pigs and Other Species. *Viruses* **14**, 261 (2022).
10. Alarcon, P., Rushton, J. & Wieland, B. Cost of post-weaning multi-systemic wasting syndrome and porcine circovirus type-2 subclinical infection in England – An economic disease model. *Prev. Vet. Med.* **110**, 88–102 (2013).
11. Sánchez-Sánchez, M., Aispuro-Hernandez, E., Quintana-Obregon, E. A., Vargas-Arispuro, I. D. C. & Martinez-Tellez, M. A. Estimating tomato production losses due to plant viruses, a look at the past and new challenges. *Comun. Sci.* **15**, e4247 (2024).
12. Varma, A. & Malathi, V. G. Emerging geminivirus problems: A serious threat to crop production. *Ann. Appl. Biol.* **142**, 145–164 (2003).
13. Lal, A. *et al.* Nanovirus Disease Complexes: An Emerging Threat in the Modern Era. *Front. Plant Sci.* **11**, 558403 (2020).

14. http://www.apsnet.org/publications/apsnetfeatures/Pages/cassava.
15. Reavy, B. *et al.* Distinct Circular Single-Stranded DNA Viruses Exist in Different Soil Types. *Appl. Environ. Microbiol.* **81**, 3934–3945 (2015).
16. Rosario, K., Schenck, R. O., Harbeitner, R. C., Lawler, S. N. & Breitbart, M. Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. *Front. Microbiol.* **6**, (2015).
17. Abbas, A. A. *et al.* Redondoviridae, a Family of Small, Circular DNA Viruses of the Human Oro-Respiratory Tract Associated with Periodontitis and Critical Illness. *Cell Host Microbe* **25**, 719-729.e4 (2019).
18. Spandole, S., Cimponeriu, D., Berca, L. M. & Mihăescu, G. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch. Virol.* **160**, 893–908 (2015).
19. Varsani, A. & Krupovic, M. Smacoviridae: a new family of animal-associated single-stranded DNA viruses. *Arch. Virol.* **163**, 2005–2015 (2018).
20. Gronenborn, B., Randles, J., Vetten, H. & Thomas, J. Create one new family (Metaxyviridae) with one new genus (Cofodevirus) and one species (Coconut foliar decay virus) moved from the family Nanoviridae (Mulpavirales). *Int. Comm. Taxon. Viruses Propos. Taxoprop Number 2020022P* (2021).
21. Krupovic, M. & Varsani, A. Naryaviridae, Nenyaviridae, and Vilyaviridae: three new families of single-stranded DNA viruses in the phylum Cressdnaviricota. *Arch. Virol.* **167**, 2907–2921 (2022).
22. Da Silva, J. P. H. *et al.* Amesuviridae: a new family of plant-infecting viruses in the phylum Cressdnaviricota, realm Monodnaviria. *Arch. Virol.* **168**, 223 (2023).
23. Zhang, J., Ma, M., Liu, Y. & Ismayil, A. Plant Defense and Viral Counter-Defense during Plant–Geminivirus Interactions. *Viruses* **15**, 510 (2023).
24. Rosario, K. *et al.* Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus. *Arch. Virol.* **162**, 1447–1463 (2017).
25. Breitbart, M. *et al.* ICTV Virus Taxonomy Profile: Circoviridae. *J. Gen. Virol.* **98**, 1997–1998 (2017).
26. Andrew, M. K. *Virus Taxonomy Ninth Report of the International Committee on Taxonomy of Viruses*.
27. Aronson, M. N. *et al.* Clink, a Nanovirus-Encoded Protein, Binds both pRB and SKP1. *J. Virol.* **74**, 2967–2972 (2000).
28. Wanitchakorn, R., Harding, R. M. & Dale, J. L. Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Arch. Virol.* **145**, 593–602 (2000).
29. Wanitchakorn, R., Hafner, G. J., Harding, R. M. & Dale, J. L. Functional

analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *Microbiology* **81**, 299–306 (2000).
30.Thomas, J. E. *et al.* ICTV Virus Taxonomy Profile: Nanoviridae. *J. Gen. Virol.* **102**, (2021).
31. Stainton, D. *et al.* The global distribution of *Banana bunchy top virus* reveals little evidence for frequent recent, human-mediated long distance dispersal events. *Virus Evol.* **1**, vev009 (2015).
32. Gaafar, Y., Nielsen, G. C. & Ziebell, H. Molecular characterisation of the first occurrence of *Pea necrotic yellow dwarf virus* in Denmark. *New Dis. Rep.* **37**, 16–16 (2018).
33. Gaafar, Y., Grausgruber-Gröger, S. & Ziebell, H. *Vicia faba, V. sativa* and *Lens culinaris* as new hosts for *Pea necrotic yellow dwarf virus* in Germany and Austria. *New Dis. Rep.* **34**, 28–28 (2016).
34. Gaafar, Y., Timchenko, T. & Ziebell, H. First report of *Pea necrotic yellow dwarf virus* in The Netherlands. *New Dis. Rep.* **35**, 23–23 (2017).
35. Saucke, H., Uteau, D., Brinkmann, K. & Ziebell, H. Symptomology and yield impact of pea necrotic yellow dwarf virus (PNYDV) in faba bean (Vicia faba L. minor). *Eur. J. Plant Pathol.* **153**, 1299–1315 (2019).
36. Gaafar, Y. Z. A. & Ziebell, H. Aphid transmission of nanoviruses. *Arch. Insect Biochem. Physiol.* **104**, e21668 (2020).
37. Timchenko, T. *et al.* The Master Rep Concept in Nanovirus Replication: Identification of Missing Genome Components and Potential for Natural Genetic Reassortment. *Virology* **274**, 189–195 (2000).
38. Krupovic, M., Ghabrial, S. A., Jiang, D. & Varsani, A. Genomoviridae: a new family of widespread single-stranded DNA viruses. *Arch. Virol.* **161**, 2633–2643 (2016).
39. Couto, R. D. S. *et al.* Genomoviruses in Liver Samples of Molossus molossus Bats. *Microorganisms* **12**, 688 (2024).
40. Varsani, A. & Krupovic, M. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. *Virus Evol.* **3**, (2017).
41. Varsani, A. & Krupovic, M. Family Genomoviridae: 2021 taxonomy update. *Arch. Virol.* **166**, 2911–2926 (2021).
42. Li, P. *et al.* A tripartite ssDNA mycovirus from a plant pathogenic fungus is infectious as cloned DNA and purified virions. *Sci. Adv.* **6**, eaay9634 (2020).
43.Ruiz-Padilla, A., Rodríguez-Romero, J., Gómez-Cid, I., Pacifico, D. & Ayllón, M. A. Novel Mycoviruses Discovered in the Mycovirome of a Necrotrophic Fungus. *mBio* **12**, e03705-20 (2021).
44. Ruiz-Padilla, A., Turina, M. & Ayllón, M. A. Molecular characterization of a tetra segmented ssDNA virus infecting Botrytis cinerea worldwide. *Virol. J.* **20**,

306 (2023).
45. Kazlauskas, D. *et al.* Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. *Virology* **504**, 114–121 (2017).
46. Kimura, K. & Tomaru, Y. Discovery of Two Novel Viruses Expands the Diversity of Single-Stranded DNA and Single-Stranded RNA Viruses Infecting a Cosmopolitan Marine Diatom. *Appl. Environ. Microbiol.* **81**, 1120–1131 (2015).
47. Nagasaki, K. *et al.* Previously Unknown Virus Infects Marine Diatom. *Appl. Environ. Microbiol.* **71**, 3528–3535 (2005).
48. Varsani, A. & Krupovic, M. 2024 Smacoviridae family update: 59 new species in seven genera. *Arch. Virol.* **169**, 184 (2024).
49. Díez-Villaseñor, C. & Rodriguez-Valera, F. CRISPR analysis suggests that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. *Nat. Commun.* **10**, 294 (2019).
50. Abbas, A., Taylor, L. J., Collman, R. G., Bushman, F. D., & ICTV Report Consortium. ICTV Virus Taxonomy Profile: Redondoviridae: This article is part of the ICTV Virus Taxonomy Profiles collection. *J. Gen. Virol.* **102**, (2021).
51. Zhang, Y., Wang, C., Feng, X., Chen, X. & Zhang, W. *Redondoviridae* and periodontitis: a case–control study and identification of five novel redondoviruses from periodontal tissues. *Virus Evol.* **7**, veab033 (2021).
52. Gronenborn, B. *et al.* Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. *Sci. Rep.* **8**, 5698 (2018).
53. Basso, M. F., Da Silva, J. C. F., Fajardo, T. V. M., Fontes, E. P. B. & Zerbini, F. M. A novel, highly divergent ssDNA virus identified in Brazil infecting apple, pear and grapevine. *Virus Res.* **210**, 27–33 (2015).
54. Bejerman, N., De Breuil, S. & Nome, C. Identification and molecular characterization of a novel circular single-stranded DNA virus associated with yerba mate in Argentina. *Arch. Virol.* **163**, 2811–2815 (2018).
55. Kulshrestha, S., Bhardwaj, A., & Vanshika. Geminiviruses: Taxonomic Structure and Diversity in Genomic Organization. *Recent Pat. Biotechnol.* **14**, 86–98 (2020).
56. Fiallo-Olivé, E. *et al.* ICTV Virus Taxonomy Profile: Geminiviridae 2021: This article is part of the ICTV Virus Taxonomy Profiles collection. *J. Gen. Virol.* **102**, (2021).
57. Bejerman, N. & Debat, H. Create a new genus Welwivirus in the family Geminiviridae, in the order Geplafuvirales, including two species, Welwivirus welwitschiae and Welwivirus mirabilis. (2023).
58. Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D. & Mansoor, S. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat.*

Rev. Microbiol. **11**, 777–788 (2013).
59.Fondong, V. N. Geminivirus protein structure and function. *Mol. Plant Pathol.* **14**, 635–649 (2013).
60. Luna, A. P. & Lozano-Durán, R. Geminivirus-Encoded Proteins: Not All Positional Homologs Are Made Equal. *Front. Microbiol.* **11**, 878 (2020).
61. Jeske, H. Geminiviruses. in *TT Viruses* (eds. De Villiers, E.-M. & Hausen, H. Z.) vol. 331 185–226 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009).
62. Gutierrez, C. Geminivirus DNA replication. *Cell. Mol. Life Sci. CMLS* **56**, 313–329 (1999).
63. Wu, M., Bejarano, E. R., Castillo, A. G. & Lozano-Durán, R. Geminivirus DNA replication in plants. in *Geminivirus : Detection, Diagnosis and Management* 323–346 (Elsevier, 2022). doi:10.1016/B978-0-323-90587-9.00038-9.
64. Tan, H., Li, F., Zhou, X. & Lozano-Durán, R. New genes on the block: additional open reading frames in the genomes of geminiviruses. *Trends Microbiol.* S0966842X25001507 (2025) doi:10.1016/j.tim.2025.05.003.
65. Gong, P. *et al.* Geminiviruses encode additional small proteins with specific subcellular localizations and virulence function. *Nat. Commun.* **12**, 4278 (2021).
66. Chiu, C.-W., Li, Y.-R., Lin, C.-Y., Yeh, H.-H. & Liu, M.-J. Translation initiation landscape profiling reveals hidden open-reading frames required for the pathogenesis of tomato yellow leaf curl Thailand virus. *Plant Cell* **34**, 1804–1821 (2022).
67. Zhao, S. *et al.* The novel C5 protein from tomato yellow leaf curl virus is a virulence factor and suppressor of gene silencing. *Stress Biol.* **2**, 19 (2022).
68. Zhao, S. *et al.* Geminivirus C5 proteins mediate formation of virus complexes at plasmodesmata for viral intercellular movement. *Plant Physiol.* **193**, 322–338 (2023).
69. Wang, Z., Wang, Y., Lozano-Duran, R., Hu, T. & Zhou, X. Identification of a novel C6 protein encoded by tomato leaf curl China virus. *Phytopathol. Res.* **4**, 46 (2022).
70. Liu, H. *et al.* Functional identification of a novel C7 protein of tomato yellow leaf curl virus. *Virology* **585**, 117–126 (2023).
71. Gong, P. *et al.* Tomato yellow leaf curl virus V3 protein traffics along microfilaments to plasmodesmata to promote virus cell-to-cell movement. *Sci. China Life Sci.* **65**, 1046–1049 (2022).
72. Li, P. *et al.* The C5 protein encoded by Ageratum leaf curl Sichuan virus is a virulence factor and contributes to the virus infection. *Mol. Plant Pathol.* **22**, 1149–1158 (2021).
73. Laufs, J. *et al.* In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc. Natl.*

Acad. Sci. **92**, 3879–3883 (1995).
74. Saunders, K., Lucy, A. & Stanley, J. RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Res.* **20**, 6311–6315 (1992).
75. Rojas, M. R. *et al.* World Management of Geminiviruses. *Annu. Rev. Phytopathol.* **56**, 637–677 (2018).
76. Rojas, M. R., Hagen, C., Lucas, W. J. & Gilbertson, R. L. Exploiting Chinks in the Plant's Armor: Evolution and Emergence of Geminiviruses. *Annu. Rev. Phytopathol.* **43**, 361–394 (2005).
77. Navas-Castillo, J., Fiallo-Olivé, E. & Sánchez-Campos, S. Emerging Virus Diseases Transmitted by Whiteflies. *Annu. Rev. Phytopathol.* **49**, 219–248 (2011).
78. Soto, M. J. & Gilbertson, R. L. Distribution and Rate of Movement of the Curtovirus *Beet mild curly top virus* (Family *Geminiviridae*) in the Beet Leafhopper. *Phytopathology*® **93**, 478–484 (2003).
79. Briddon, R. W., Bedford, I. D., Tsai, J. H. & Markham, P. G. Analysis of the Nucleotide Sequence of the Treehopper-Transmitted Geminivirus, Tomato Pseudo-Curly Top Virus, Suggests a Recombinant Origin. *Virology* **219**, 387–394 (1996).
80. Roumagnac, P. *et al.* Alfalfa Leaf Curl Virus: an Aphid-Transmitted Geminivirus. *J. Virol.* **89**, 9683–9688 (2015).
81. Varsani, A. *et al.* Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch. Virol.* **159**, 2193–2203 (2014).
82. Roumagnac, P. *et al.* Establishment of five new genera in the family Geminiviridae: Citlodavirus, Maldovirus, Mulcrilevirus, Opunvirus, and Topilevirus. *Arch. Virol.* **167**, 695–710 (2022).
83. Gilbert, W. & Dressler, D. DNA Replication: The Rolling Circle Model. *Cold Spring Harb. Symp. Quant. Biol.* **33**, 473–484 (1968).
84. Dressler, D. The Rolling Circle for ϕX DNA Replication, II. Synthesis of Single-Stranded Circles. *Proc. Natl. Acad. Sci.* **67**, 1934–1942 (1970).
85.Baas, P. D. DNA replication of single-stranded Escherichia coli DNA phages. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* **825**, 111–139 (1985).
86. Ruiz-Masó, J. A. *et al.* Plasmid Rolling-Circle Replication. *Microbiol. Spectr.* **3**, 3.1.16 (2015).
87. Thomas, J. & Pritham, E. J. *Helitrons* , the Eukaryotic Rolling-circle Transposable Elements. *Microbiol. Spectr.* **3**, 3.4.03 (2015).
88. Koonin, E. V. & Ilyina, T. V. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *J. Gen. Virol.* **73**, 2763–2766 (1992).

89. Wawrzyniak, P., Płucienniczak, G. & Bartosik, D. The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins. *Front. Microbiol.* **8**, 2353 (2017).
90. Tarasova, E., Dhindwal, S., Popp, M., Hussain, S. & Khayat, R. Mechanism of DNA Interaction and Translocation by the Replicase of a Circular Rep-Encoding Single-Stranded DNA Virus. *mBio* **12**, e00763-21 (2021).
91. Tarasova, E. & Khayat, R. A Structural Perspective of Reps from CRESS-DNA Viruses and Their Bacterial Plasmid Homologues. *Viruses* **14**, 37 (2021).
92. Cheung, A. K. Identification of an octanucleotide motif sequence essential for viral protein, DNA, and progeny virus biosynthesis at the origin of DNA replication of porcine circovirus type 2. *Virology* **324**, 28–36 (2004).
93. Hafner, G. J., Dale, J. L., Harding, R. M., Wolter, L. C. & Stafford, M. R. Nicking and joining activity of banana bunchy top virus replication protein in vitro. *J. Gen. Virol.* **78**, 1795–1799 (1997).
94. Laufs, J. Geminivirus replication: Genetic and biochemical characterization of rep protein function, a review. *Biochimie* **77**, 765–773 (1995).
95. Steinfeldt, T., Finsterbusch, T. & Mankertz, A. Demonstration of Nicking/Joining Activity at the Origin of DNA Replication Associated with the Rep and Rep′ Proteins of Porcine Circovirus Type 1. *J. Virol.* **80**, 6225–6234 (2006).
96. Timchenko, T. *et al.* A Single Rep Protein Initiates Replication of Multiple Genome Components of Faba Bean Necrotic Yellows Virus, a Single-Stranded DNA Virus of Plants. *J. Virol.* **73**, 10173–10182 (1999).
97. Choudhury, N. R. *et al.* The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. *Nucleic Acids Res.* **34**, 6362–6377 (2006).
98. Clérot, D. & Bernardi, F. DNA Helicase Activity Is Associated with the Replication Initiator Protein Rep of Tomato Yellow Leaf Curl Geminivirus. *J. Virol.* **80**, 11322–11330 (2006).
99. Jeske, H. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO J.* **20**, 6158–6167 (2001).
100. Preiss, W. & Jeske, H. Multitasking in Replication Is Common among Geminiviruses. *J. Virol.* **77**, 2972–2980 (2003).
101. Erdmann, J. B. *et al.* Replicative intermediates of maize streak virus found during leaf development. *J. Gen. Virol.* **91**, 1077–1081 (2010).
102. Kreuzer, K. N. Recombination-dependent DNA replication in phage T4. *Trends Biochem. Sci.* **25**, 165–173 (2000).
103. Bonnamy, M., Blanc, S. & Michalakis, Y. Replication mechanisms of circular ssDNA plant viruses and their potential implication in viral gene expression regulation. *mBio* **14**, e01692-23 (2023).

104. Burgers, P. M. J. & Kunkel, T. A. Eukaryotic DNA Replication Fork. *Annu. Rev. Biochem.* **86**, 417–438 (2017).
105. Bleichert, F., Botchan, M. R. & Berger, J. M. Mechanisms for initiating cellular DNA replication. *Science* **355**, eaah6317 (2017).
106. Costa, A. & Diffley, J. F. X. The Initiation of Eukaryotic DNA Replication. *Annu. Rev. Biochem.* **91**, 107–131 (2022).
107. Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N. & Oda, M. Eukaryotic Chromosome DNA Replication: Where, When, and How? *Annu. Rev. Biochem.* **79**, 89–130 (2010).
108. Gao, F. & Zhang, C.-T. DoriC: a database of *oriC* regions in bacterial genomes. *Bioinformatics* **23**, 1866–1867 (2007).
109. Barry, E. R. & Bell, S. D. DNA Replication in the Archaea. *Microbiol. Mol. Biol. Rev.* **70**, 876–887 (2006).
110. Foss, E. J. *et al.* Identification of 1600 replication origins in S. cerevisiae. *eLife* **12**, RP88087 (2024).
111. Tian, M. *et al.* Integrative analysis of DNA replication origins and ORC/MCM binding sites in human cells reveals a lack of overlap. Preprint at https://doi.org/10.7554/eLife.89548.3 (2024).
112. Fragkos, M., Ganier, O., Coulombe, P. & Méchali, M. DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.* **16**, 360–374 (2015).
113. Boos, D. & Ferreira, P. Origin Firing Regulations to Control Genome Replication Timing. *Genes* **10**, 199 (2019).
114. Palzkill, T. G. & Newlon, C. S. A yeast replication origin consists of multiple copies of a small conserved sequence. *Cell* **53**, 441–450 (1988).
115. Marahrens, Y. & Stillman, B. A Yeast Chromosomal Origin of DNA Replication Defined by Multiple Functional Elements. *Science* **255**, 817–823 (1992).
116. Theis, J. F. & Newlon, C. S. The *ARS309* chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. *Proc. Natl. Acad. Sci.* **94**, 10786–10791 (1997).
117. Frigola, J. *et al.* Cdt1 stabilizes an open MCM ring for helicase loading. *Nat. Commun.* **8**, 15720 (2017).
118. Zhai, Y. *et al.* Open-ringed structure of the Cdt1–Mcm2–7 complex as a precursor of the MCM double hexamer. *Nat. Struct. Mol. Biol.* **24**, 300–308 (2017).
119. Araki, H. Initiation of chromosomal DNA replication in eukaryotic cells; contribution of yeast genetics to the elucidation. *Genes Genet. Syst.* **86**, 141–149 (2011).
120. Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. X. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*

519, 431–435 (2015).
121. Labib, K. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev.* **24**, 1208–1219 (2010).
122. Deegan, T. D., Yeeles, J. T. & Diffley, J. F. Phosphopeptide binding by Sld3 links Dbf4-dependent kinase to MCM replicative helicase activation. *EMBO J.* **35**, 961–973 (2016).
123. Tanaka, S., Nakato, R., Katou, Y., Shirahige, K. & Araki, H. Origin Association of Sld3, Sld7, and Cdc45 Proteins Is a Key Step for Determination of Origin-Firing Timing. *Curr. Biol.* **21**, 2055–2063 (2011).
124. Tanaka, S. *et al.* CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* **445**, 328–332 (2007).
125. Masumoto, H., Muramatsu, S., Kamimura, Y. & Araki, H. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature* **415**, 651–655 (2002).
126. Zegerman, P. & Diffley, J. F. X. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* **445**, 281–285 (2007).
127. Muramatsu, S., Hirai, K., Tak, Y.-S., Kamimura, Y. & Araki, H. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ε, and GINS in budding yeast. *Genes Dev.* **24**, 602–612 (2010).
128. Douglas, M. E., Ali, F. A., Costa, A. & Diffley, J. F. X. The mechanism of eukaryotic CMG helicase activation. *Nature* **555**, 265–268 (2018).
129. Georgescu, R. *et al.* Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc. Natl. Acad. Sci.* **114**, (2017).
130. Chuang, R.-Y. & Kelly, T. J. The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc. Natl. Acad. Sci.* **96**, 2656–2661 (1999).
131. Lee, J.-K., Moon, K.-Y., Jiang, Y. & Hurwitz, J. The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc. Natl. Acad. Sci.* **98**, 13589–13594 (2001).
132. Vashee, S. *et al.* Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.* **17**, 1894–1908 (2003).
133. Remus, D., Beall, E. L. & Botchan, M. R. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC–DNA binding. *EMBO J.* **23**, 897–907 (2004).
134. Schaarschmidt, D., Baltin, J., Stehle, I. M., Lipps, H. J. & Knippers, R. An

episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J.* **23**, 191–201 (2004).
135. Cayrou, C. *et al.* Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* **21**, 1438–1449 (2011).
136. Cayrou, C. *et al.* New insights into replication origin characteristics in metazoans. *Cell Cycle* **11**, 658–667 (2012).
137. Delgado, S. Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J.* **17**, 2426–2435 (1998).
138. Besnard, E. *et al.* Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
139. Crevel, G. & Cotterill, S. Forced binding of the origin of replication complex to chromosomal sites in *Drosophila* S2 cells creates an origin of replication. *J. Cell Sci.* **125**, 965–972 (2012).
140. Georgescu, R. E. *et al.* Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 (2014).
141. Langston, L. D. *et al.* CMG helicase and DNA polymerase ε form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci.* **111**, 15390–15395 (2014).
142. Zhou, J. C. *et al.* CMG–Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proc. Natl. Acad. Sci.* **114**, 4141–4146 (2017).
143. Jain, R., Aggarwal, A. K. & Rechkoblit, O. Eukaryotic DNA polymerases. *Curr. Opin. Struct. Biol.* **53**, 77–87 (2018).
144. Pavlov, Y. I., Zhuk, A. S. & Stepchenkova, E. I. DNA Polymerases at the Eukaryotic Replication Fork Thirty Years after: Connection to Cancer. *Cancers* **12**, 3489 (2020).
145. Morrison, A., Araki, H., Clark, A. B., Hamatake, R. K. & Sugino, A. A third essential DNA polymerase in S. cerevisiae. *Cell* **62**, 1143–1151 (1990).
146. Aria, V. & Yeeles, J. T. P. Mechanism of Bidirectional Leading-Strand Synthesis Establishment at Eukaryotic DNA Replication Origins. *Mol. Cell* **73**, 199-211.e10 (2019).
147. Simon, A. C. *et al.* A Ctf4 trimer couples the CMG helicase to DNA polymerase α in the eukaryotic replisome. *Nature* **510**, 293–297 (2014).
148. Villa, F. *et al.* Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. *Mol. Cell* **63**, 385–396 (2016).
149. Samora, C. P. *et al.* Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome. *Mol. Cell* **63**, 371–384 (2016).

150. Yuan, Z. *et al.* Ctf4 organizes sister replisomes and Pol α into a replication factory. *eLife* **8**, e47405 (2019).
151. Branzei, D., Bene, S., Gangwani, L. & Szakal, B. The multifaceted roles of the Ctf4 replisome hub in the maintenance of genome integrity. *DNA Repair* **142**, 103742 (2024).
152. Jones, M. L., Aria, V., Baris, Y. & Yeeles, J. T. P. How Pol α-primase is targeted to replisomes to prime eukaryotic DNA replication. *Mol. Cell* **83**, 2911-2924.e16 (2023).
153. Zhou, Z.-X., Lujan, S. A., Burkholder, A. B., Garbacz, M. A. & Kunkel, T. A. Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat. Commun.* **10**, 3992 (2019).
154. Deem, A. *et al.* Break-Induced Replication Is Highly Inaccurate. *PLoS Biol.* **9**, e1000594 (2011).
155. Hicks, W. M., Kim, M. & Haber, J. E. Increased Mutagenesis and Unique Mutation Signature Associated with Mitotic Gene Conversion. *Science* **329**, 82–85 (2010).
156. Donnianni, R. A. *et al.* DNA Polymerase Delta Synthesizes Both Strands during Break-Induced Replication. *Mol. Cell* **76**, 371-381.e4 (2019).
157. Dmowski, M., Makiela-Dzbenska, K., Sharma, S., Chabes, A. & Fijalkowska, I. J. Impairment of the non-catalytic subunit Dpb2 of DNA Pol ε results in increased involvement of Pol δ on the leading strand. *DNA Repair* **129**, 103541 (2023).
158. Dmowski, M. *et al.* Increased contribution of DNA polymerase delta to the leading strand replication in yeast with an impaired CMG helicase complex. *DNA Repair* **110**, 103272 (2022).
159. Waga, S. & Stillman, B. Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. *Nature* **369**, 207–212 (1994).
160. Zlotkin, T. *et al.* DNA polymerase epsilon may be dispensable for SV40-but not cellular-DNA replication. *EMBO J.* **15**, 2298–2305 (1996).
161. Balakrishnan, L. & Bambara, R. A. Okazaki Fragment Metabolism. *Cold Spring Harb. Perspect. Biol.* **5**, a010173–a010173 (2013).
162. Liu, B., Hu, J., Wang, J. & Kong, D. Direct Visualization of RNA-DNA Primer Removal from Okazaki Fragments Provides Support for Flap Cleavage and Exonucleolytic Pathways in Eukaryotic Cells. *J. Biol. Chem.* **292**, 4777–4788 (2017).
163. Rossi, M. L. & Bambara, R. A. Reconstituted Okazaki Fragment Processing Indicates Two Pathways of Primer Removal. *J. Biol. Chem.* **281**, 26051–26061 (2006).
164. Sun, H. *et al.* Okazaki fragment maturation: DNA flap dynamics for cell proliferation and survival. *Trends Cell Biol.* **33**, 221–234 (2023).

165. Stodola, J. L. & Burgers, P. M. Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nat. Struct. Mol. Biol.* **23**, 402–408 (2016).
166. Chilkova, O. *et al.* The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res.* **35**, 6588–6597 (2007).
167. Garinther, W. I. & Schultz, M. C. Topoisomerase Function during Replication-Independent Chromatin Assembly in Yeast. *Mol. Cell. Biol.* **17**, 3520–3526 (1997).
168. Bailey, R., Priego Moreno, S. & Gambus, A. Termination of DNA replication forks: “Breaking up is hard to do”. *Nucleus* **6**, 187–196 (2015).
169. Dewar, J. M. & Walter, J. C. Mechanisms of DNA replication termination. *Nat. Rev. Mol. Cell Biol.* **18**, 507–516 (2017).
170. Priego Moreno, S., Bailey, R., Campion, N., Herron, S. & Gambus, A. Polyubiquitylation drives replisome disassembly at the termination of DNA replication. *Science* **346**, 477–481 (2014).
171. Maric, M., Maculins, T., De Piccoli, G. & Labib, K. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* **346**, 1253596 (2014).
172. Bell, S. P. Terminating the replisome. *Science* **346**, 418–419 (2014).
173. McGuffee, S. R., Smith, D. J. & Whitehouse, I. Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. *Mol. Cell* **50**, 123–135 (2013).
174. Dewar, J. M., Budzowska, M. & Walter, J. C. The mechanism of DNA replication termination in vertebrates. *Nature* **525**, 345–350 (2015).
175. Low, E., Chistol, G., Zaher, M. S., Kochenova, O. V. & Walter, J. C. The DNA replication fork suppresses CMG unloading from chromatin before termination. *Genes Dev.* **34**, 1534–1545 (2020).
176. Deegan, T. D., Mukherjee, P. P., Fujisawa, R., Polo Rivera, C. & Labib, K. CMG helicase disassembly is controlled by replication fork DNA, replisome components and a ubiquitin threshold. *eLife* **9**, e60371 (2020).
177. Vrtis, K. B. *et al.* Single-strand DNA breaks cause replisome disassembly. *Mol. Cell* **81**, 1309-1318.e6 (2021).
178. Sonneville, R. *et al.* CUL-2LRR-1 and UBXN-3 drive replisome disassembly during DNA replication termination and mitosis. *Nat. Cell Biol.* **19**, 468–479 (2017).
179. Dewar, J. M., Low, E., Mann, M., Räschle, M. & Walter, J. C. $CRL2^{Lrr1}$ promotes unloading of the vertebrate replisome from chromatin during replication termination. *Genes Dev.* **31**, 275–290 (2017).

180. Villa, F. *et al.* $CUL2^{LRR1}$, TRAIP and p97 control CMG helicase disassembly in the mammalian cell cycle. *EMBO Rep.* **22**, e52164 (2021).
181. Fan, Y. *et al.* LRR1-mediated replisome disassembly promotes DNA replication by recycling replisome components. *J. Cell Biol.* **220**, e202009147 (2021).
182. Jenkyn-Bedford, M. *et al.* A conserved mechanism for regulating replisome disassembly in eukaryotes. *Nature* **600**, 743–747 (2021).
183. Kochenova, O. V., Mukkavalli, S., Raman, M. & Walter, J. C. Cooperative assembly of p97 complexes involved in replication termination. *Nat. Commun.* **13**, 6591 (2022).
184. Bodnar, N. & Rapoport, T. Toward an understanding of the Cdc48/p97 ATPase. *F1000Research* **6**, 1318 (2017).
185. Van Den Boom, J. & Meyer, H. VCP/p97-Mediated Unfolding as a Principle in Protein Homeostasis and Signaling. *Mol. Cell* **69**, 182–194 (2018).
186. Twomey, E. C. *et al.* Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. *Science* **365**, eaax1033 (2019).
187. Ji, Z. *et al.* Translocation of polyubiquitinated protein substrates by the hexameric Cdc48 ATPase. *Mol. Cell* **82**, 570-584.e8 (2022).
188. Xu, N. *et al.* Cryo-EM structure of human hexameric MCM2-7 complex. *iScience* **25**, 104976 (2022).
189. Mukherjee, P. P. & Labib, K. P. M. In Vitro Reconstitution Defines the Minimal Requirements for Cdc48-Dependent Disassembly of the CMG Helicase in Budding Yeast. *Cell Rep.* **28**, 2777-2783.e4 (2019).
190. Donson, J., Morris-Krsinich, B. A., Mullineaux, P. M., Boulton, M. I. & Davies, J. W. A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. *EMBO J.* **3**, 3069–3073 (1984).
191. Wei, H. & Lozano-Durán, R. The primase subunits of DNA polymerase α, PRIM1 and PRIM2, are required for the replication of the geminivirus tomato yellow leaf curl virus in the host plant. *Open Access*.
192. Wu, M. *et al.* Plant DNA polymerases α and δ mediate replication of geminiviruses. *Nat. Commun.* **12**, 2780 (2021).
193. Siskos, L. *et al.* DNA primase large subunit is an essential plant gene for geminiviruses, putatively priming viral ss-DNA replication. *Front. Plant Sci.* **14**, 1130723 (2023).
194. Paprotka, T., Deuschle, K., Pilartz, M. & Jeske, H. Form follows function in geminiviral minichromosome architecture. *Virus Res.* **196**, 44–55 (2015).
195. 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**, 358 (2010).
196. Richter, K. S., Kleinow, T. & Jeske, H. Somatic homologous recombination

in plants is promoted by a geminivirus in a tissue-selective manner. *Virology* **452–453**, 287–296 (2014).
197. Wyant_11.
198. Liu, J. & Morrical, S. W. Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. *Virol. J.* **7**, 357 (2010).
199. Suyal, G., Mukherjee, S. K. & Choudhury, N. R. The host factor RAD51 is involved in mungbean yellow mosaic India virus (MYMIV) DNA replication. *Arch. Virol.* **158**, 1931–1941 (2013).
200. Morra, M. R. & Petty, I. T. D. Tissue Specificity of Geminivirus Infection Is Genetically Determined. *Plant Cell* **12**, 2259–2270 (2000).
201. Hanley-Bowdoin, L., Settlage, S. B. & Robertson, D. Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Mol. Plant Pathol.* **5**, 149–156 (2004).
202. Desvoyes, B. & Gutierrez, C. Roles of plant retinoblastoma protein: cell cycle and beyond. *EMBO J.* **39**, e105802 (2020).
203. Wang, L., Chen, H., Wang, C., Hu, Z. & Yan, S. Negative regulator of E2F transcription factors links cell cycle checkpoint and DNA damage repair. *Proc. Natl. Acad. Sci.* **115**, (2018).
204. Inzé, D. Green light for the cell cycle. *EMBO J.* **24**, 657–662 (2005).
205. Grafi, G. *et al.* A maize cDNA encoding a member of the retinoblastoma protein family: involvement in endoreduplication. *Proc. Natl. Acad. Sci.* **93**, 8962–8967 (1996).
206. Liu, L., Saunders, K., Thomas, C. L., Davies, J. W. & Stanley, J. Bean Yellow Dwarf Virus RepA, but Not Rep, Binds to Maize Retinoblastoma Protein, and the Virus Tolerates Mutations in the Consensus Binding Motif. *Virology* **256**, 270–279 (1999).
207. Horvath, G. V. *et al.* Prediction of functional regions of the maize streak virus replication-associated proteins by protein-protein interaction analysis.
208. Xie, Q., Sanz-Burgos, A. P., Hannon, G. J. & Gutiérrez, C. Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. *EMBO J.* **15**, 4900–4908 (1996).
209. Arguello-Astorga, G. *et al.* A Novel Motif in Geminivirus Replication Proteins Interacts with the Plant Retinoblastoma-Related Protein. *J. Virol.* **78**, 4817–4826 (2004).
210. The Plant Anaphase-Promoting Complex Cyclosome.
211. Zhang, X. *et al.* Plant cell-cycle regulators control the nuclear environment for viral pathogenesis. *Cell Host Microbe* S193131282500054X (2025) doi:10.1016/j.chom.2025.02.006.
212. Nash, T. E. *et al.* Functional Analysis of a Novel Motif Conserved across Geminivirus Rep Proteins. *J. Virol.* **85**, 1182–1192 (2011).

213.Gros, M. F., Te Riele, H. & Ehrlich, S. D. Replication origin of a single-stranded DNA plasmid pC194. *EMBO J.* **8**, 2711–2716 (1989).
214.Xie, Q., Suárez-López, P. & Gutiérrez, C. Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO J.* **14**, 4073–4082 (1995).
215.Schalk, H. J., Matzeit, V., Schiller, B., Schell, J. & Gronenborn, B. Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication. *EMBO J.* **8**, 359–364 (1989).
216.Xie, Q., Sanz-Burgos, A. P., Hannon, G. J. & Gutiérrez, C. Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. *EMBO J.* **15**, 4900–4908 (1996).
217.Ruhel, R. *et al.* Functional implications of residues of the B′ motif of geminivirus replication initiator protein in its helicase activity. *FEBS J.* **288**, 6492–6509 (2021).
218.Gorbalenya, A. E. & Koonin, E. V. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419–429 (1993).
219.Ruhel, R. & Chakraborty, S. Multifunctional roles of geminivirus encoded replication initiator protein. *VirusDisease* **30**, 66–73 (2019).
220.Rizvi, I., Choudhury, N. R. & Tuteja, N. Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. *Arch. Virol.* **160**, 375–387 (2015).
221.Shakir, S. *et al.* REPercussions: how geminiviruses recruit host factors for replication. *Front. Microbiol.* **14**, 1224221 (2023).
222.Kamal, H. *et al.* Functional role of geminivirus encoded proteins in the host: Past and present. *Biotechnol. J.* **19**, 2300736 (2024).
223.Settlage, S. B., Miller, A. B. & Hanley-Bowdoin, L. Interactions between geminivirus replication proteins. *J. Virol.* **70**, 6790–6795 (1996).
224.Orozco, B. M., Miller, A. B., Settlage, S. B. & Hanley-Bowdoin, L. Functional Domains of a Geminivirus Replication Protein. *J. Biol. Chem.* **272**, 9840–9846 (1997).
225.Wang, L. *et al.* Combinatorial interactions between viral proteins expand the potential functional landscape of the tomato yellow leaf curl virus proteome. *PLOS Pathog.* **18**, e1010909 (2022).
226.Settlage, S. B., See, R. G. & Hanley-Bowdoin, L. Geminivirus C3 Protein: Replication Enhancement and Protein Interactions. *J. Virol.* **79**, 9885–9895 (2005).
227.Settlage, S. B., Miller, A. B., Gruissem, W. & Hanley-Bowdoin, L. Dual Interaction of a Geminivirus Replication Accessory Factor with a Viral

Replication Protein and a Plant Cell Cycle Regulator. *Virology* **279**, 570–576 (2001).
228. Ach, R. A. *et al. RRB1* and *RRB2* Encode Maize Retinoblastoma-Related Proteins That Interact with a Plant D-Type Cyclin and Geminivirus Replication Protein. *Mol. Cell. Biol.* **17**, 5077–5086 (1997).
229. Kong, L.-J. A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO J.* **19**, 3485–3495 (2000).
230. Aguilar, E., Garnelo Gomez, B. & Lozano-Duran, R. Recent advances on the plant manipulation by geminiviruses. *Curr. Opin. Plant Biol.* **56**, 56–64 (2020).
231. Luque, A., Sanz-Burgos, A. P., Ramirez-Parra, E., Castellano, M. M. & Gutierrez, C. Interaction of Geminivirus Rep Protein with Replication Factor C and Its Potential Role during Geminivirus DNA Replication. *Virology* **302**, 83–94 (2002).
232. Singh, D. K., Islam, M. N., Choudhury, N. R., Karjee, S. & Mukherjee, S. K. The 32 kDa subunit of replication protein A (RPA) participates in the DNA replication of Mung bean yellow mosaic India virus (MYMIV) by interacting with the viral Rep protein. *Nucleic Acids Res.* **35**, 755–770 (2007).
233. Bagewadi, B., Chen, S., Lal, S. K., Choudhury, N. R. & Mukherjee, S. K. PCNA Interacts with Indian Mung Bean Yellow Mosaic Virus Rep and Downregulates Rep Activity. *J. Virol.* **78**, 11890–11903 (2004).
234. Luque, A., Sanz-Burgos, A. P., Ramirez-Parra, E., Castellano, M. M. & Gutierrez, C. Interaction of Geminivirus Rep Protein with Replication Factor C and Its Potential Role during Geminivirus DNA Replication. *Virology* **302**, 83–94 (2002).
235. Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A. & Bejarano, E. R. Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology* **312**, 381–394 (2003).
236. Suyal, G., Mukherjee, S. K., Srivastava, P. S. & Choudhury, N. R. Arabidopsis thaliana MCM2 plays role(s) in mungbean yellow mosaic India virus (MYMIV) DNA replication. *Arch. Virol.* **158**, 981–992 (2013).
237. Rizvi, I., Hisamuddin, M., Malik, A. & Khan, R. H. Identification of mungbean yellow mosaic India virus (MYMIV) Rep interacting partners using phage display and influence of *Arabidopsis thaliana* MCM3 on geminivirus DNA replication. *J. Biomol. Struct. Dyn.* **40**, 10507–10517 (2022).
238. Kaliappan, K., Choudhury, N. R., Suyal, G. & Mukherjee, S. K. A novel role for RAD54: this host protein modulates geminiviral DNA replication. *FASEB J.* **26**, 1142–1160 (2012).
239. Rodríguez-Negrete, E. *et al.* Geminivirus R ep protein interferes with the

plant DNA methylation machinery and suppresses transcriptional gene silencing. *New Phytol.* **199**, 464–475 (2013).
240. Arroyo-Mateos, M. *et al.* Geminivirus Replication Protein Impairs SUMO Conjugation of Proliferating Cellular Nuclear Antigen at Two Acceptor Sites. *J. Virol.* **92**, e00611-18 (2018).
241. Sánchez-Durán, M. A. *et al.* Interaction between Geminivirus Replication Protein and the SUMO-Conjugating Enzyme Is Required for Viral Infection. *J. Virol.* **85**, 9789–9800 (2011).
242. Castillo, A. G., Kong, L. J., Hanley-Bowdoin, L. & Bejarano, E. R. Interaction between a Geminivirus Replication Protein and the Plant Sumoylation System. *J. Virol.* **78**, 2758–2769 (2004).
243. Li, F., Zhang, M., Zhang, C. & Zhou, X. Nuclear autophagy degrades a geminivirus nuclear protein to restrict viral infection in solanaceous plants. *New Phytol.* **225**, 1746–1761 (2020).
244. Eagle, P. A., Orozco, B. M. & Hanley-Bowdoin, L. A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *Plant Cell* **6**, 1157–1170 (1994).
245. Kushwaha, N. K., Bhardwaj, M. & Chakraborty, S. The replication initiator protein of a geminivirus interacts with host monoubiquitination machinery and stimulates transcription of the viral genome. *PLOS Pathog.* **13**, e1006587 (2017).
246. Kong, L.-J. & Hanley-Bowdoin, L. A Geminivirus Replication Protein Interacts with a Protein Kinase and a Motor Protein That Display Different Expression Patterns during Plant Development and Infection. *Plant Cell* **14**, 1817–1832 (2002).
247. Singh, D. K., Malik, P. S., Choudhury, N. R. & Mukherjee, S. K. MYMIV replication initiator protein (Rep): Roles at the initiation and elongation steps of MYMIV DNA replication. *Virology* **380**, 75–83 (2008).
248. Suyal, G., Rana, V. S., Mukherjee, S. K., Wajid, S. & Choudhury, N. R. Arabidopsis thaliana NAC083 protein interacts with Mungbean yellow mosaic India virus (MYMIV) Rep protein. *Virus Genes* **48**, 486–493 (2014).
249. Maio, F. *et al.* Identification of Tomato Proteins That Interact With Replication Initiator Protein (Rep) of the Geminivirus TYLCV. *Front. Plant Sci.* **11**, 1069 (2020).
250. Xie, Q., Sanz-Burgos, A. P., Guo, H. & Garcıa, J. A. GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein.
251. Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L. & Bisaro, D. M. Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication.

Virology **179**, 69–77 (1990).
252. Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A. & Bejarano, E. R. Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology* **312**, 381–394 (2003).
253. Pasumarthy, K. K., Choudhury, N. R. & Mukherjee, S. K. Tomato leaf curl Kerala virus (ToLCKeV) AC3 protein forms a higher order oligomer and enhances ATPase activity of replication initiator protein (Rep/AC1). *Virol. J.* **7**, 128 (2010).
254. Selth, L. A. *et al.* A NAC Domain Protein Interacts with *Tomato leaf curl virus* Replication Accessory Protein and Enhances Viral Replication. *Plant Cell* **17**, 311–325 (2005).
255. Yan, Z., Wolters, A.-M. A., Navas-Castillo, J. & Bai, Y. The Global Dimension of Tomato Yellow Leaf Curl Disease: Current Status and Breeding Perspectives. *Microorganisms* **9**, 740 (2021).
256. Verlaan, M. G. *et al.* The Tomato Yellow Leaf Curl Virus Resistance Genes Ty-1 and Ty-3 Are Allelic and Code for DFDGD-Class RNA–Dependent RNA Polymerases. *PLoS Genet.* **9**, e1003399 (2013).
257. Shen, X. *et al.* The NLR Protein Encoded by the Resistance Gene Ty-2 Is Triggered by the Replication-Associated Protein Rep/C1 of Tomato Yellow Leaf Curl Virus. *Front. Plant Sci.* **11**, 545306 (2020).
258. Yamaguchi, H. *et al.* An NB-LRR gene, TYNBS1, is responsible for resistance mediated by the Ty-2 Begomovirus resistance locus of tomato. *Theor. Appl. Genet.* **131**, 1345–1362 (2018).
259. Ji, Y., Scott, J. W., Schuster, D. J. & Maxwell, D. P. Molecular Mapping of Ty-4, a New Tomato Yellow Leaf Curl Virus Resistance Locus on Chromosome 3 of Tomato. *J. Am. Soc. Hortic. Sci.* **134**, 281–288 (2009).
260. Lapidot, M. *et al.* A Novel Route Controlling Begomovirus Resistance by the Messenger RNA Surveillance Factor Pelota. *PLOS Genet.* **11**, e1005538 (2015).
261. Barbieri, M. *et al.* INTROGRESSION OF RESISTANCE TO TWO MEDITERRANEAN VIRUS SPECIES CAUSING TOMATO YELLOW LEAF CURL INTO A VALUABLE TRADITIONAL TOMATO VARIETY.
262. Ohnishi, J., Yamaguchi, H. & Saito, A. Analysis of the Mild strain of tomato yellow leaf curl virus, which overcomes Ty-2 gene–mediated resistance in tomato line H24. *Arch. Virol.* **161**, 2207–2217 (2016).
263. Gill, U. *et al.* Ty-6, a major begomovirus resistance gene on chromosome 10, is effective against Tomato yellow leaf curl virus and Tomato mottle virus. *Theor. Appl. Genet.* **132**, 1543–1554 (2019).
264. Shen, X. *et al.* The tomato gene Ty-6, encoding DNA polymerase delta subunit 1, confers broad resistance to Geminiviruses. *Theor. Appl. Genet.* **138**,

22 (2025).
265. Lim, Y.-W. *et al.* Mutations in DNA polymerase δ subunit 1 co-segregate with CMD2-type resistance to Cassava Mosaic Geminiviruses. *Nat. Commun.* **13**, 3933 (2022).
266. Gilbert, K. B. *et al.* Evidence that variation in DNA polymerase delta subunit 1 underlies geminivirus resistance in diverse plants. Preprint at https://doi.org/10.21203/rs.3.rs-6330309/v1 (2025).
267. Zhang, Y. *et al.* TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity. *Nat. Commun.* **10**, 3252 (2019).
268. Ding, X., Jimenez-Gongora, T., Krenz, B. & Lozano-Duran, R. Chloroplast clustering around the nucleus is a general response to pathogen perception in *Nicotiana benthamiana*. *Mol. Plant Pathol.* **20**, 1298–1306 (2019).
269. He, Y.-Z. *et al.* A plant DNA virus replicates in the salivary glands of its insect vector via recruitment of host DNA synthesis machinery. *Proc. Natl. Acad. Sci.* **117**, 16928–16937 (2020).
270. Maio, F. *et al.* A Lysine Residue Essential for Geminivirus Replication Also Controls Nuclear Localization of the Tomato Yellow Leaf Curl Virus Rep Protein. *J. Virol.* **93**, e01910-18 (2019).
271. Czosnek, H. *et al.* Discovering Host Genes Involved in the Infection by the Tomato Yellow Leaf Curl Virus Complex and in the Establishment of Resistance to the Virus Using Tobacco Rattle Virus-based Post Transcriptional Gene Silencing. *Viruses* **5**, 998–1022 (2013).
272. Wu, M., Ding, X., Fu, X. & Lozano-Duran, R. Transcriptional reprogramming caused by the geminivirus Tomato yellow leaf curl virus in local or systemic infections in Nicotiana benthamiana. *BMC Genomics* **20**, 542 (2019).
273. Wu, M. *et al.* Plant DNA polymerases α and δ mediate replication of geminiviruses. *Nat. Commun.* **12**, 2780 (2021).
274. Garbacz, M. A. *et al.* Evidence that DNA polymerase δ contributes to initiating leading strand DNA replication in Saccharomyces cerevisiae. *Nat. Commun.* **9**, 858 (2018).
275. Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. J. & Kunkel, T. A. Division of Labor at the Eukaryotic Replication Fork. *Mol. Cell* **30**, 137–144 (2008).
276. Kanakala, S. & Kuria, P. Chickpea chlorotic dwarf virus: An Emerging Monopartite Dicot Infecting Mastrevirus. *Viruses* **11**, 5 (2018).
277. Castellano, M. M., Sanz-Burgos, A. P. & Gutiérrez, C. Initiation of DNA replication in a eukaryotic rolling-circle replicon: identification of multiple DNA-protein complexes at the geminivirus origin 1 1Edited by I. B. Holland. *J. Mol.*

Biol. **290**, 639–652 (1999).
278. Akbar Behjatnia, S. Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. *Nucleic Acids Res.* **26**, 925–931 (1998).
279. Vijayraghavan, S. & Schwacha, A. The Eukaryotic Mcm2-7 Replicative Helicase. in *The Eukaryotic Replisome: a Guide to Protein Structure and Function* (ed. MacNeill, S.) vol. 62 113–134 (Springer Netherlands, Dordrecht, 2012).
280. George, B. *et al.* Mutational analysis of the helicase domain of a replication initiator protein reveals critical roles of Lys 272 of the B′ motif and Lys 289 of the β-hairpin loop in geminivirus replication. *J. Gen. Virol.* **95**, 1591–1602 (2014).
281. O'Donnell, M. E. & Li, H. The ring-shaped hexameric helicases that function at DNA replication forks. *Nat. Struct. Mol. Biol.* **25**, 122–130 (2018).
282. Horvath, G. V. *et al.* Prediction of functional regions of the maize streak virus replication-associated proteins by protein-protein interaction analysis.
283. Feiler, H. S. *et al.* The higher plant Arabidopsis thaliana encodes a functional CDC48 homologue which is highly expressed in dividing and expanding cells. *EMBO J.* **14**, 5626–5637 (1995).
284. Parisi, E., Yahya, G., Flores, A. & Aldea, M. Cdc48/p97 segregase is modulated by cyclin-dependent kinase to determine cyclin fate during G1 progression. *EMBO J.* **37**, e98724 (2018).
285. Fu, X., Ng, C., Feng, D. & Liang, C. Cdc48p is required for the cell cycle commitment point at Start via degradation of the G1-CDK inhibitor Far1p. *J. Cell Biol.* **163**, 21–26 (2003).
286. Mouysset, J. *et al.* Cell cycle progression requires the CDC-48$^{UFD-1/NPL-4}$ complex for efficient DNA replication. *Proc. Natl. Acad. Sci.* **105**, 12879–12884 (2008).
287. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA^{+} : A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes. *Genome Res.* **9**, 27–43 (1999).
288. Parsons, C. A., Tsaneva, I., Lloyd, R. G. & West, S. C. Interaction of Escherichia coli RuvA and RuvB proteins with synthetic Holliday junctions. *Proc. Natl. Acad. Sci.* **89**, 5452–5456 (1992).
289. Li, F., Xu, X., Li, Z., Wang, Y. & Zhou, X. Identification of Yeast Factors Involved in the Replication of Mungbean Yellow Mosaic India Virus Using Yeast Temperature-Sensitive Mutants. *Virol. Sin.* **35**, 120–123 (2020).
290. Ge, L., Zhang, J., Zhou, X. & Li, H. Genetic Structure and Population Variability of Tomato Yellow Leaf Curl China Virus. *J. Virol.* **81**, 5902–5907 (2007).

291. Duffy, S. & Holmes, E. C. Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus *Tomato Yellow Leaf Curl Virus*. *J. Virol.* **82**, 957–965 (2008).
292. Urbino, C., Thébaud, G., Granier, M., Blanc, S. & Peterschmitt, M. A novel cloning strategy for isolating, genotyping and phenotyping genetic variants of geminiviruses. *Virol. J.* **5**, 135 (2008).
293. Sánchez-Campos, S. *et al.* Differential Shape of Geminivirus Mutant Spectra Across Cultivated and Wild Hosts With Invariant Viral Consensus Sequences. *Front. Plant Sci.* **9**, 932 (2018).
294. Kim, D. I. *et al.* Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci.* **111**, (2014).
295. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).
296. Kido, K. *et al.* AirID, a novel proximity biotinylation enzyme, for analysis of protein–protein interactions. *eLife* **9**, e54983 (2020).
297. Xu, S.-L., Shrestha, R., Karunadasa, S. S. & Xie, P.-Q. Proximity Labeling in Plants. *Annu. Rev. Plant Biol.* **74**, 285–312 (2023).
298. Wang, L. *et al.* Inference of a Geminivirus−Host Protein−Protein Interaction Network through Affinity Purification and Mass Spectrometry Analysis. *Viruses* **9**, 275 (2017).
299. Pilartz, M. & Jeske, H. Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. *Virology* **189**, 800–802 (1992).
300. Pilartz, M. & Jeske, H. Mapping of Abutilon Mosaic Geminivirus Minichromosomes. *J. Virol.* **77**, 10808–10818 (2003).
301. Schier, A. C. & Taatjes, D. J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* **34**, 465–488 (2020).
302. Déjardin, J. & Kingston, R. E. Purification of Proteins Associated with Specific Genomic Loci. *Cell* **136**, 175–186 (2009).
303. Dungrawala, H. & Cortez, D. Purification of Proteins on Newly Synthesized DNA Using iPOND. in *The Nucleus* (ed. Hancock, R.) vol. 1228 123–131 (Springer New York, New York, NY, 2015).
304. Sirbu, B. M., Couch, F. B. & Cortez, D. Monitoring the spatiotemporal dynamics of proteins at replication forks and in assembled chromatin using isolation of proteins on nascent DNA. *Nat. Protoc.* **7**, 594–605 (2012).
305. Gassmann, M. *et al.* Replication of single-stranded porcine circovirus DNA by DNA polymerases α and δ. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* **951**, 280–289 (1988).
306. Tattersall, P. & Ward, D. C. Rolling hairpin model for replication of parvovirus and linear chromosomal DNA. *Nature* **263**, 106–109 (1976).
307. Christensen, J. & Tattersall, P. Parvovirus Initiator Protein NS1 and RPA

Coordinate Replication Fork Progression in a Reconstituted DNA Replication System. *J. Virol.* **76**, 6518–6531 (2002).
308. Nash, K., Chen, W., McDonald, W. F., Zhou, X. & Muzyczka, N. Purification of Host Cell Enzymes Involved in Adeno-Associated Virus DNA Replication. *J. Virol.* **81**, 5777–5787 (2007).
309. Ni, T.-H. *et al.* Cellular Proteins Required for Adeno-Associated Virus DNA Replication in the Absence of Adenovirus Coinfection. *J. Virol.* **72**, 2777–2787 (1998).
310. Waga, S. & Stillman, B. THE DNA REPLICATION FORK IN EUKARYOTIC CELLS. *Annu. Rev. Biochem.* **67**, 721–751 (1998).
311. Lee, S. H., Pan, Z. Q., Kwong, A. D., Burgers, P. M. & Hurwitz, J. Synthesis of DNA by DNA polymerase epsilon in vitro. *J. Biol. Chem.* **266**, 22707–22717 (1991).
312. Stillman, B. Reconsidering DNA Polymerases at the Replication Fork in Eukaryotes. *Mol. Cell* **59**, 139–141 (2015).
313. Weinberg, D. H. *et al.* Reconstitution of simian virus 40 DNA replication with purified proteins. *Proc. Natl. Acad. Sci.* **87**, 8692–8696 (1990).
314. Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **73**, 4433–4448 (2016).
315. Dovrat, D., Stodola, J. L., Burgers, P. M. J. & Aharoni, A. Sequential switching of binding partners on PCNA during in vitro Okazaki fragment maturation. *Proc. Natl. Acad. Sci.* **111**, 14118–14123 (2014).
316. Levin, D. S., McKenna, A. E., Motycka, T. A., Matsumoto, Y. & Tomkinson, A. E. Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Curr. Biol.* **10**, 919-S2 (2000).
317. Johnson, A. & O'Donnell, M. DNA Ligase: Getting a Grip to Seal the Deal. *Curr. Biol.* **15**, R90–R92 (2005).
318. Stahl, H., Droge, P. & Knippers, R. DNA helicase activity of SV40 large tumor antigen.
319. Wiekowski, M. Simian virus 40 large T antigen DNA helicase. Characterization of the ATPase-dependent DNA unwinding activity and its substrate requirements.
320. Hanai, R. & Wang, J. C. The mechanism of sequence-specific DNA cleavage and strand transfer by phi X174 gene A* protein. *J. Biol. Chem.* **268**, 23830–23836 (1993).
321. Van Mansfeld, A. D. M., Van Teeffelen, H. A. A. M., Bass, P. D. & Jansz, H. S. Two juxtaposed tyrosyl-OH groups participate in 0X174 gene A protein catalysed cleavage and ligation of DNA. *Nucleic Acids Res.* **14**, 4229–4238 (1986).

322. Chaudhuri, B., Xu, H., Todorov, I., Dutta, A. & Yates, J. L. Human DNA replication initiation factors, ORC and MCM, associate with *oriP* of Epstein–Barr virus. *Proc. Natl. Acad. Sci.* **98**, 10085–10089 (2001).
323. Dhar, S. K. *et al.* Replication from oriP of Epstein-Barr Virus Requires Human ORC and Is Inhibited by Geminin. *Cell* **106**, 287–296 (2001).
324. Schepers, A. Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus. *EMBO J.* **20**, 4588–4602 (2001).
325. Moriyama, K., Yoshizawa-Sugata, N., Obuse, C., Tsurimoto, T. & Masai, H. Epstein-Barr Nuclear Antigen 1 (EBNA1)-dependent Recruitment of Origin Recognition Complex (Orc) on oriP of Epstein-Barr Virus with Purified Proteins. *J. Biol. Chem.* **287**, 23977–23994 (2012).
326. Chacin, E. *et al.* Establishment and function of chromatin organization at replication origins. *Nature* **616**, 836–842 (2023).
327. Deuschle, K., Kepp, G. & Jeske, H. Differential methylation of the circular DNA in geminiviral minichromosomes. *Virology* **499**, 243–258 (2016).
328. Argüello-Astorga, G. R., Guevara-González, R. G., Herrera-Estrella, L. R. & Rivera-Bustamante, R. F. Geminivirus Replication Origins Have a Group-Specific Organization of Iterative Elements: A Model for Replication. *Virology* **203**, 90–100 (1994).
329. Fontes, E. P., Luckow, V. A. & Hanley-Bowdoin, L. A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell* **4**, 597–608 (1992).
330. Wang, C. *et al.* Structural basis of DNA recognition of tomato yellow leaf curl virus replication-associated protein. *Int. J. Biol. Macromol.* **205**, 316–328 (2022).
331. Ganaie, S. S. *et al.* Phosphorylated STAT5 directly facilitates parvovirus B19 DNA replication in human erythroid progenitors through interaction with the MCM complex. *PLOS Pathog.* **13**, e1006370 (2017).
332. Zou, W. *et al.* Human Parvovirus B19 Utilizes Cellular DNA Replication Machinery for Viral DNA Replication. *J. Virol.* **92**, e01881-17 (2018).
333. Nash, K., Chen, W. & Muzyczka, N. Complete In Vitro Reconstitution of Adeno-Associated Virus DNA Replication Requires the Minichromosome Maintenance Complex Proteins. *J. Virol.* **82**, 1458–1464 (2008).
334. Cotmore, S. F. & Tattersall, P. Parvoviruses: Small Does Not Mean Simple. *Annu. Rev. Virol.* **1**, 517–537 (2014).
335. Li, D. *et al.* Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* **423**, 512–518 (2003).
336. Gambus, A. Termination of Eukaryotic Replication Forks. in *DNA Replication* (eds. Masai, H. & Foiani, M.) vol. 1042 163–187 (Springer

Singapore, Singapore, 2017).
337. Jones, R. M., Reynolds-Winczura, A. & Gambus, A. A Decade of Discovery—Eukaryotic Replisome Disassembly at Replication Termination. *Biology* **13**, 233 (2024).
338. Franz, A. *et al.* CDC-48/p97 Coordinates CDT-1 Degradation with GINS Chromatin Dissociation to Ensure Faithful DNA Replication. *Mol. Cell* **44**, 85–96 (2011).
339. Oñate-Sánchez, L. & Vicente-Carbajosa, J. DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. *BMC Res. Notes* **1**, 93 (2008).
340. Maimbo, M., Ohnishi, K., Hikichi, Y., Yoshioka, H. & Kiba, A. S-Glycoprotein-Like Protein Regulates Defense Responses in *Nicotiana* Plants against *Ralstonia solanacearum*. *Plant Physiol.* **152**, 2023–2035 (2010).
341. Mason, G., Caciagli, P., Accotto, G. P. & Noris, E. Real-time PCR for the quantitation of Tomato yellow leaf curl Sardinia virus in tomato plants and in Bemisia tabaci. *J. Virol. Methods* **147**, 282–289 (2008).
342. Nie, W.-F. *et al.* Histone acetylation recruits the SWR1 complex to regulate active DNA demethylation in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **116**, 16641–16650 (2019).