In Memory of Sam Wilson

Eukaryotic DNA Replication Fidelity by the B-Family DNA Polymerases

Thomas A. Kunkel, Ph.D.

NIH National Institutes of Health

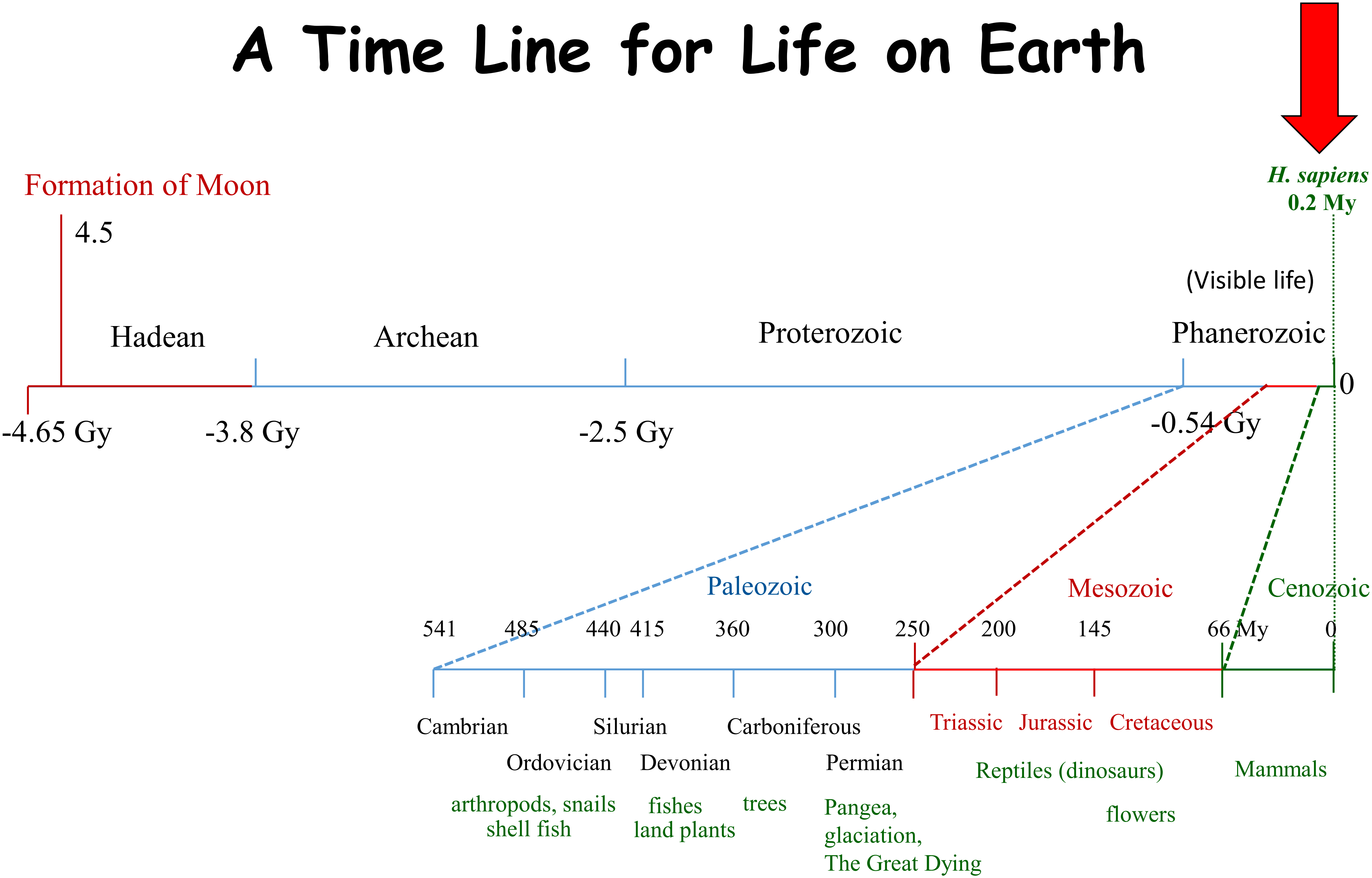
NIH National Institute of Environmental Health Sciences

Persons Currently In My Group

Kasia Bebenek
Jessica Williams
Mercedes Arana
Scott Lujan
Sarah Marks
Mahina Monsur

Postdoctoral Positions Currently Available

A Time Line for Life on Earth



H. sapiens
0.2 My

Formation of Moon
4.5

(Visible life)

Hadean
Archean
Proterozoic
Phanerozoic

-4.65 Gy
-3.8 Gy
-2.5 Gy
-0.54 Gy
0

Paleozoic
Mesozoic
Cenozoic

541
485
440
415
360
300
250
200
145
66 My
0

Cambrian
Ordovician
Silurian
Devonian
Carboniferous
Permian
Triassic
Jurassic
Cretaceous

arthropods, snails
shell fish

fishes
land plants

trees

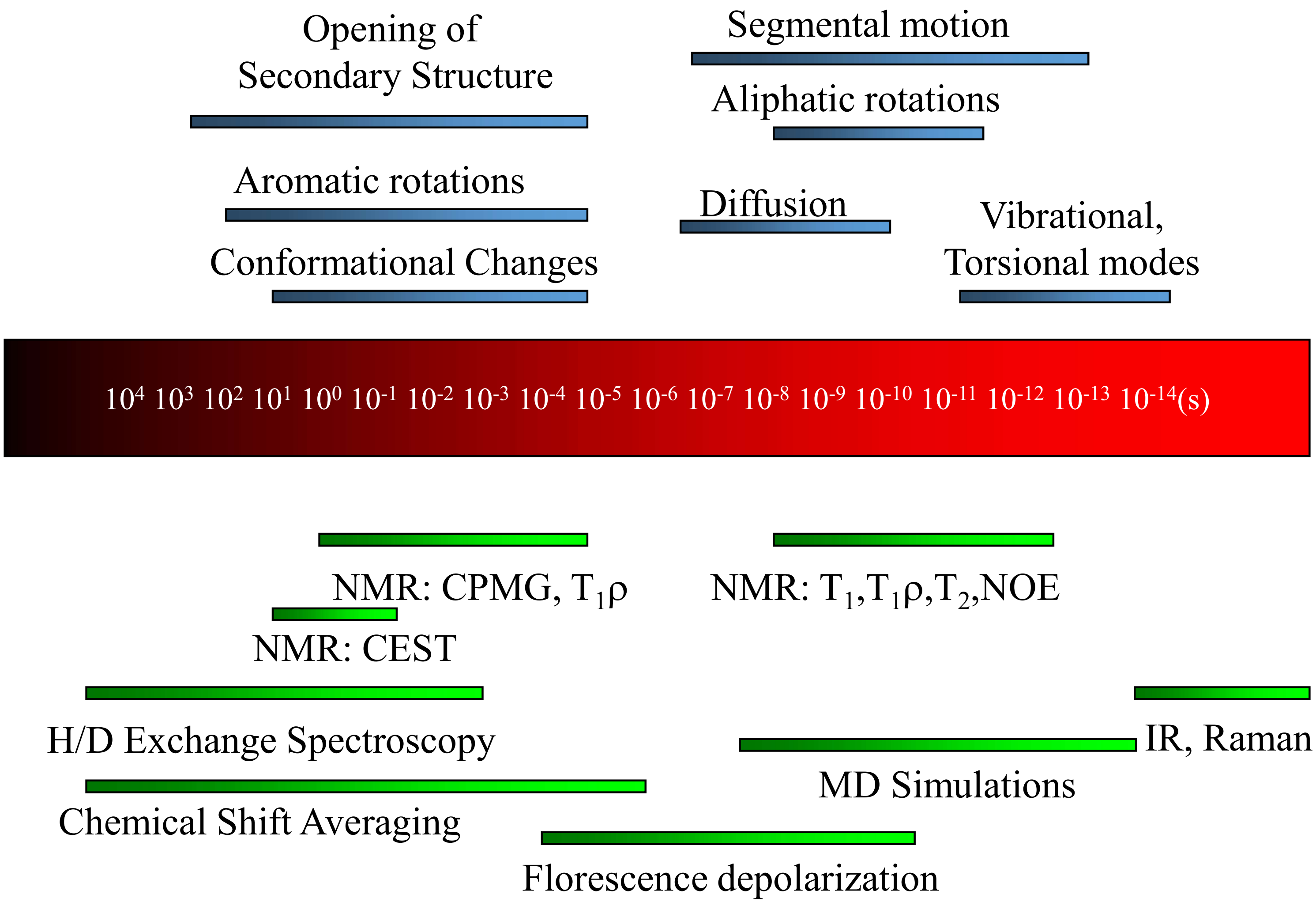
Pangea,
glaciation,
The Great Dying

Reptiles (dinosaurs)

flowers

Mammals

Protein Dynamics - Macromolecular Motions



Opening of Secondary Structure

Segmental motion

Aliphatic rotations

Aromatic rotations

Diffusion

Vibrational, Torsional modes

Conformational Changes

10^{4} 10^{3} 10^{2} 10^{1} 10^{0} 10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8} 10^{-9} 10^{-10} 10^{-11} 10^{-12} 10^{-13} 10^{-14}(s)

NMR: CPMG, $T_1\rho$

NMR: T_1,$T_1\rho$,T_2,NOE

NMR: CEST

H/D Exchange Spectroscopy

IR, Raman

MD Simulations

Chemical Shift Averaging

Florescence depolarization

Experimental Techniques

Understanding DNA Replication Fidelity

2 Ångstroms to 2 Meters

10^{-14} Seconds to > 4 Billion Years

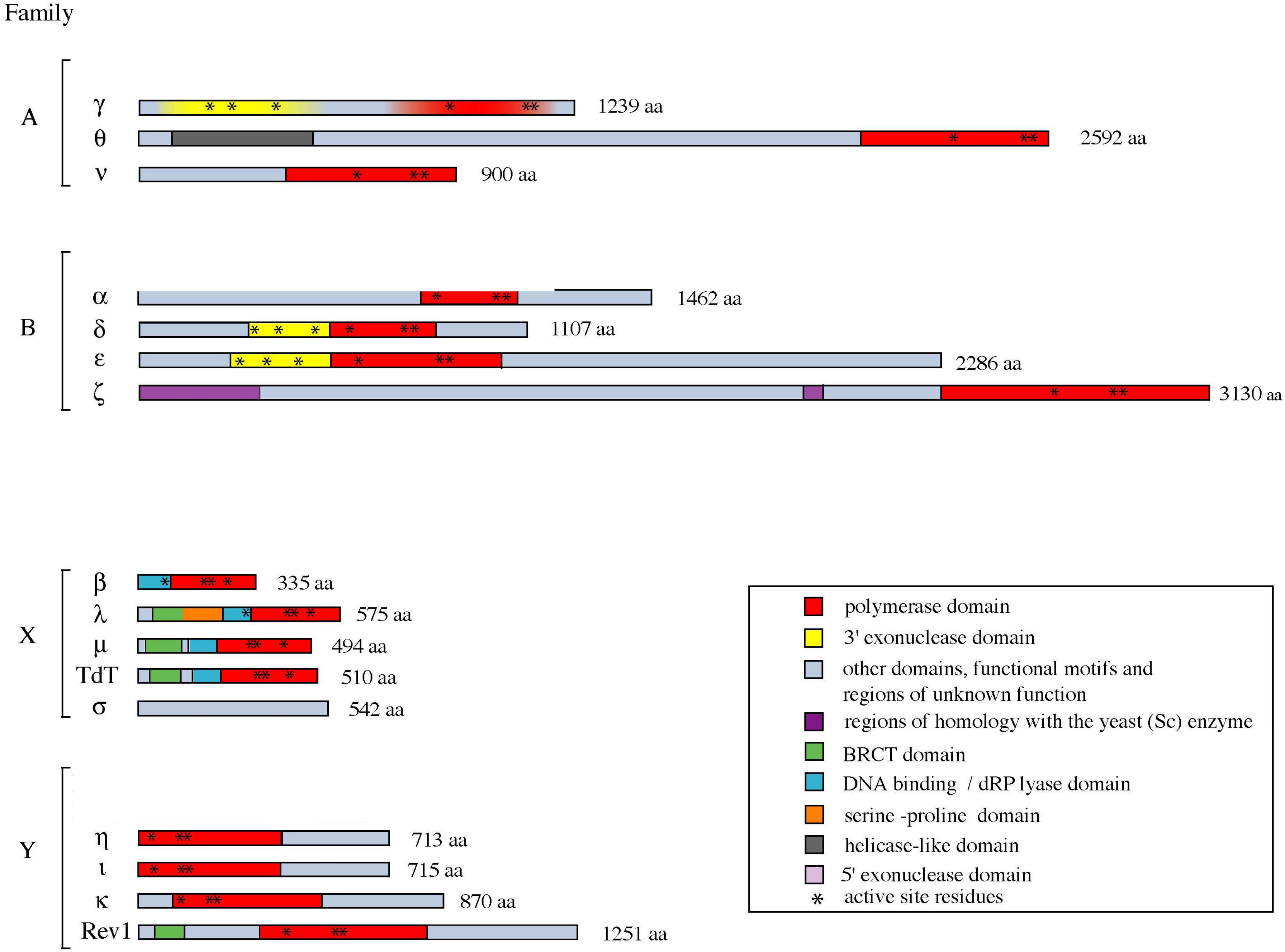
Structural Biology

Biochemistry

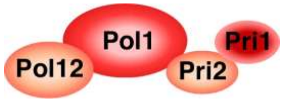
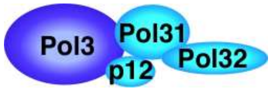
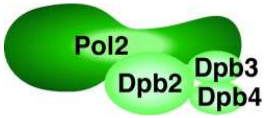
Genetics

Genomics

Catalytic Subunits of Human Pols in Four Families



Major DNA polymerases at the eukaryotic nuclear DNA replication fork

	Pol α-Primase	**Pol δ**	**Pol ε**
Subunit organization			
Genes and subunit sizes			
S. cerevisiae	Pol1-p167 Pol12-p79 Pri1-p48 Pri2-p62	Pol3-p125 Pol31-p55 Pol32-p40 -	Pol2-p256 Dpb2-p78 Dpb3-p23 Dpb4-p22
S. pombe	Pol1-p159 Pol12-p64 Pri1-p52 Spp2-p53	Pol3-p124 Cdc1-p51 Cdc27-p42 Cdm1-p19	Pol2-p253 Dpb2-p67 Dpb3-p22 Dpb4-p24
Human	PolA1-p166 PolA2-p68 Prim1-p48 Prim2A-p58	PolD1-p124 PolD2-p51 PolD3-p66 PolD4-p12	PolE-p261 PolE2-p59 PolE3-p17 PolE4-p12
Activity	Polymerase Primase	Polymerase 3’-Exonuclease	Polymerase 3’-Exonuclease dsDNA binding
Fidelity	10^{-4}-10^{-5}	10^{-6}-10^{-7}	10^{-6}-10^{-7}
Function	Initiation of replication Initiation of Okazaki fragments	Elongation and maturation of Okazaki fragments DNA repair Mutagenesis	Replisome assembly Leading strand synthesis Replication checkpoint

Eukaryotic nuclear DNA replication fidelity depends on:

1. The concentrations of dNTPs and rNTPs
    a. Absolute concentrations
    b. Relative concentrations
2. The selectivity of DNA polymerases for
    a. A correct, properly aligned dNTP
    b. The correct sugar moiety
3. Proofreading *during* replication
    a. Intrinsic proofreading
    b. Extrinsic proofreading
4. Repair of errors *after* replication by
    a. DNA Mismatch Repair
    b. Ribonucleotide Excision Repair

Number of DNA processing events that control replication fidelity in humans

Process	Events/cell cycle
Chain elongation on an open template	6,000,000,000
Okazaki fragment maturation	25,000,000
Ribonucleotide excision repair	3,000,000
Mismatch repair	≤ 100

Reasons why I am interested:

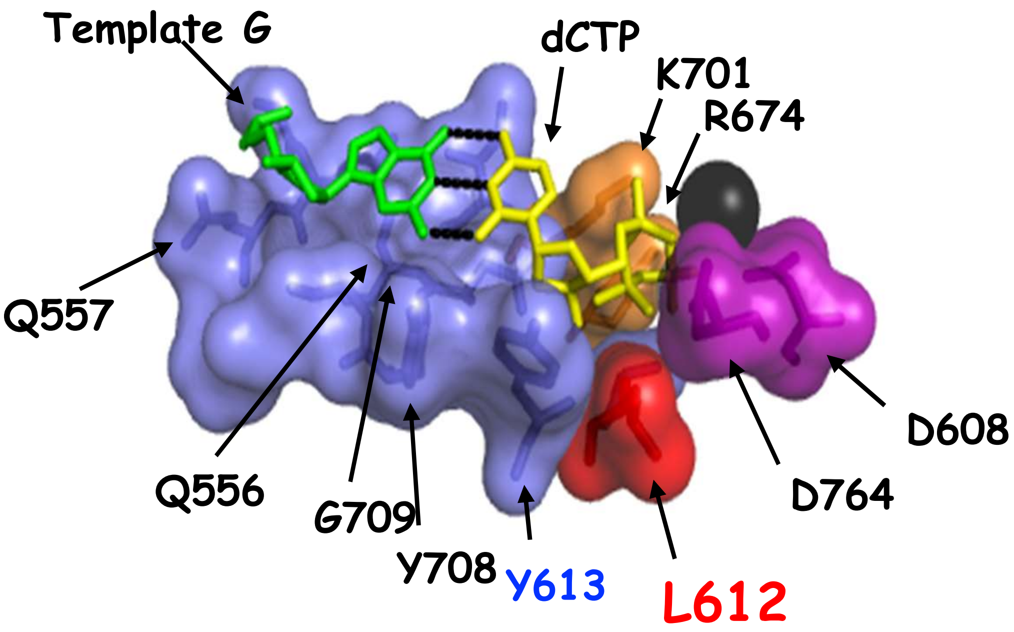
Responsible Mechanisms

Connections to Human Health

Evolutionary Conservation

Pol δ Binding Pocket

Swan *et al.*, *NSMB* 16: 979 (2009)



S.cer α	863	M	D	F	N	S	L	Y	P	S	I	I	Q	E	F	N	877
S.cer δ	607	L	D	F	N	S	L	Y	P	S	I	M	M	A	H	N	621
S.cer ε	639	V	D	V	A	S	M	Y	P	N	I	M	T	T	N	R	653
S.cer ζ	974	L	D	F	Q	S	L	Y	P	S	I	M	I	G	Y	N	988

Whole Genome Mutation Rate Analysis of Eight Diploid Yeast Strains

STRAINS

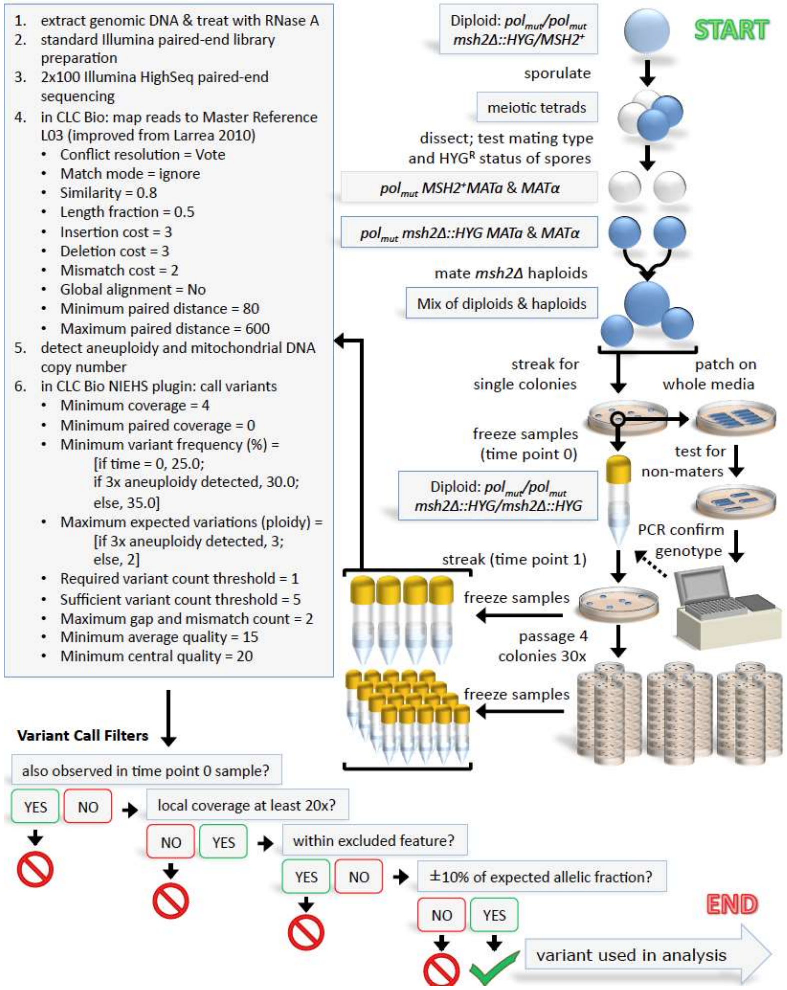
Wild type ± MMR

Pol ε - M644G ± MMR

Pol α - L868M ± MMR

Pol δ - L612M ± MMR

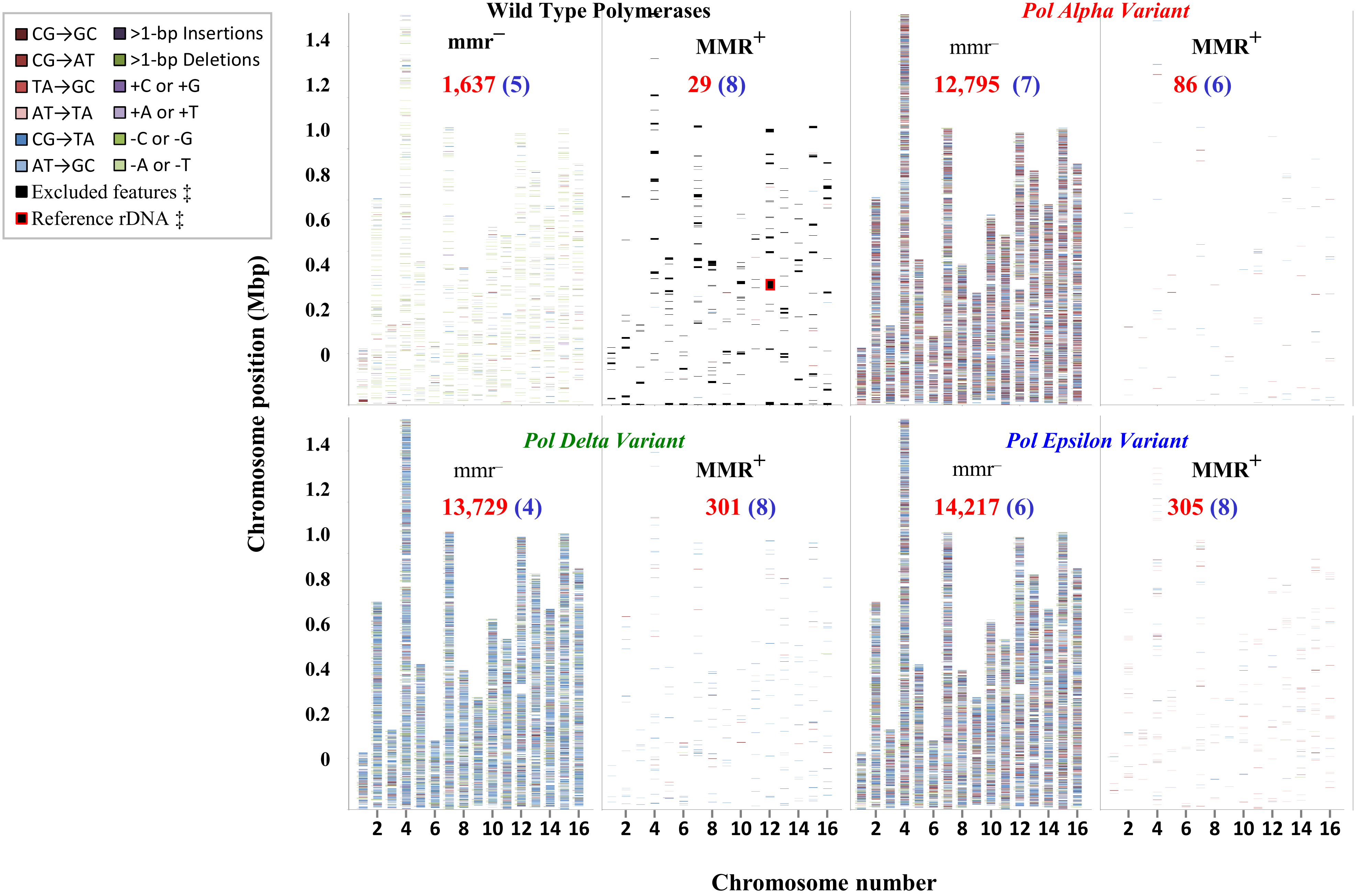
WGS PROTOCOL



1. extract genomic DNA & treat with RNase A
2. standard Illumina paired-end library preparation
3. 2x100 Illumina HighSeq paired-end sequencing
4. in CLC Bio: map reads to Master Reference L03 (improved from Larrea 2010)
   - Conflict resolution = Vote
   - Match mode = ignore
   - Similarity = 0.8
   - Length fraction = 0.5
   - Insertion cost = 3
   - Deletion cost = 3
   - Mismatch cost = 2
   - Global alignment = No
   - Minimum paired distance = 80
   - Maximum paired distance = 600
5. detect aneuploidy and mitochondrial DNA copy number
6. in CLC Bio NIEHS plugin: call variants
   - Minimum coverage = 4
   - Minimum paired coverage = 0
   - Minimum variant frequency (%) = [if time = 0, 25.0; if 3x aneuploidy detected, 30.0; else, 35.0]
   - Maximum expected variations (ploidy) = [if 3x aneuploidy detected, 3; else, 2]
   - Required variant count threshold = 1
   - Sufficient variant count threshold = 5
   - Maximum gap and mismatch count = 2
   - Minimum average quality = 15
   - Minimum central quality = 20

Lujan *et al.* (2014) *Genome Research*

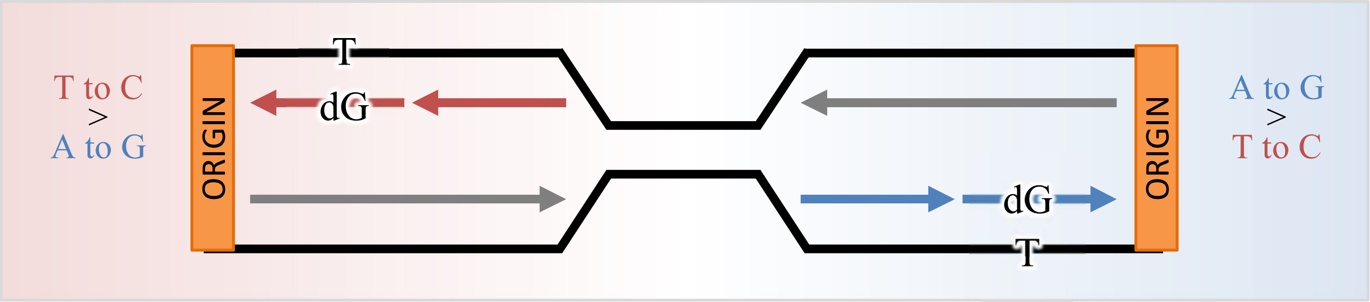
Number & Distribution of Single Base Changes

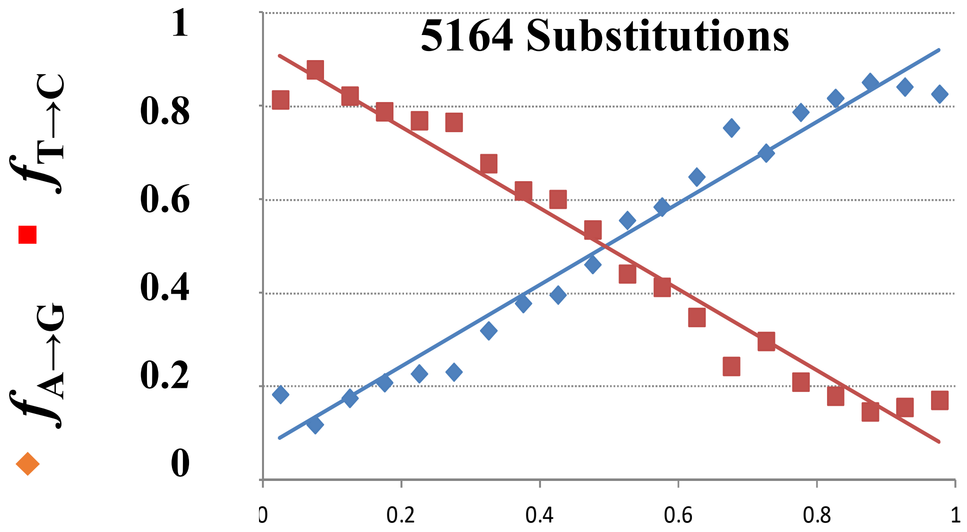


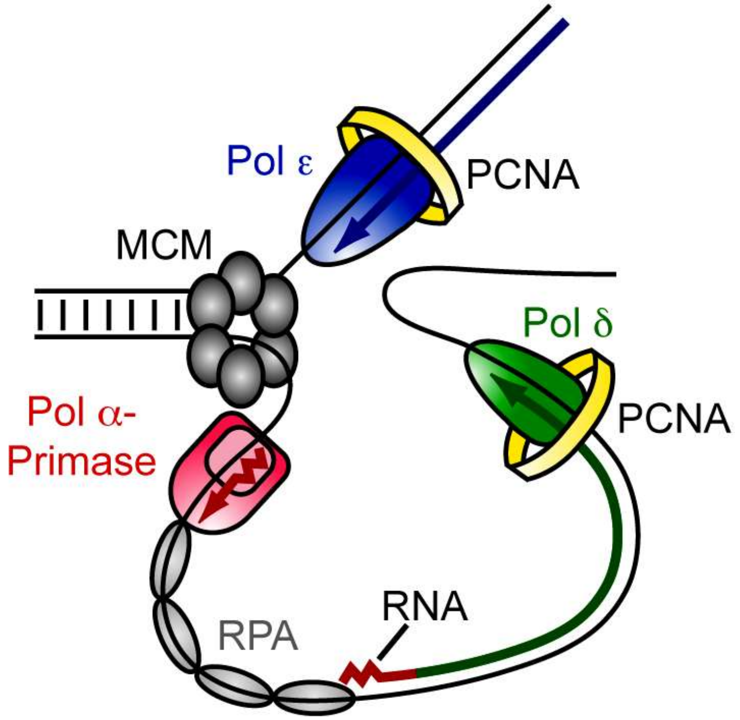
Lujan *et al.* (2014) *Genome Research*

Distribution of Substitutions Relative to Origins

L612M Pol δ
in vitro
T•dG : A•dC
28 : 1

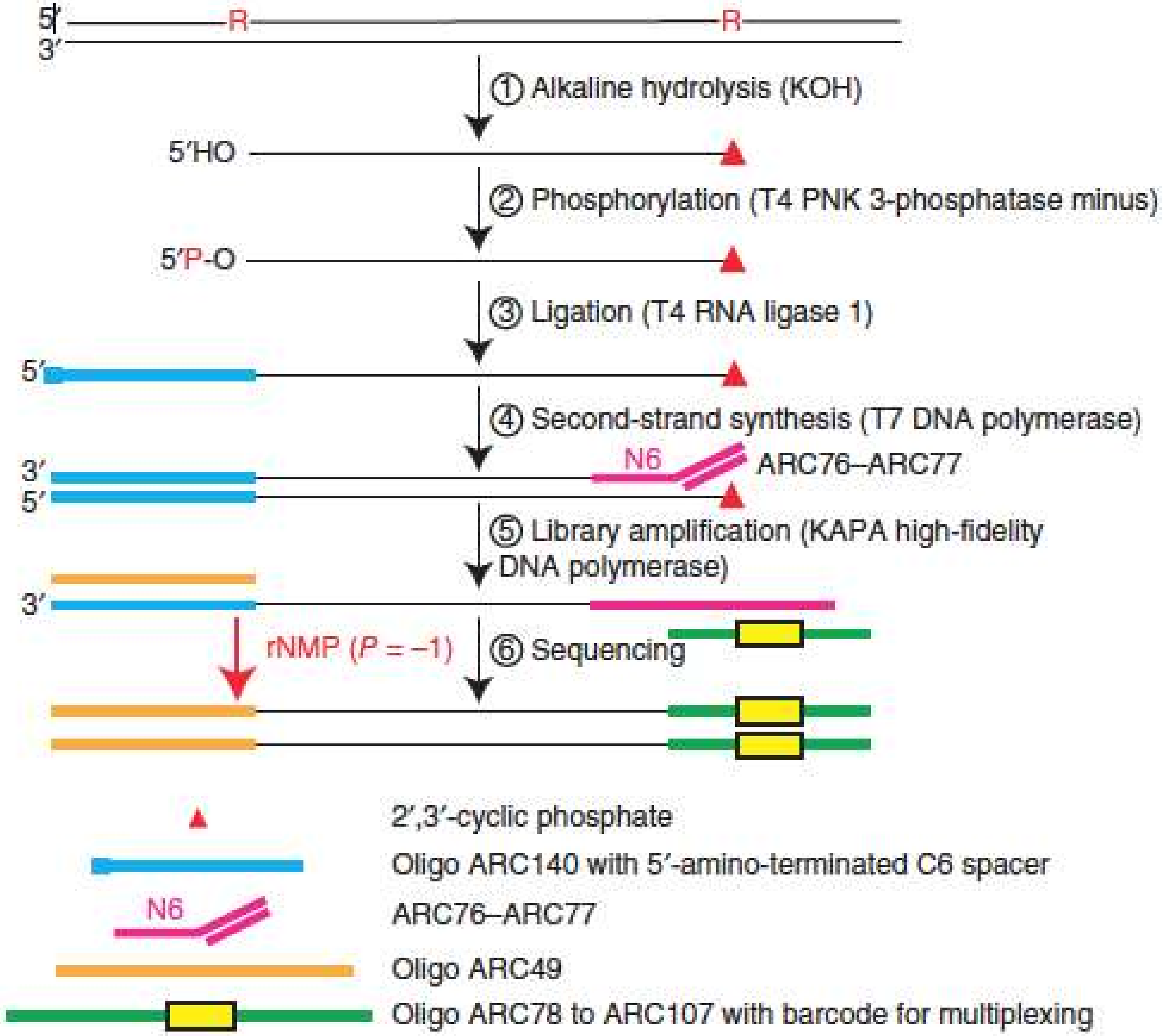




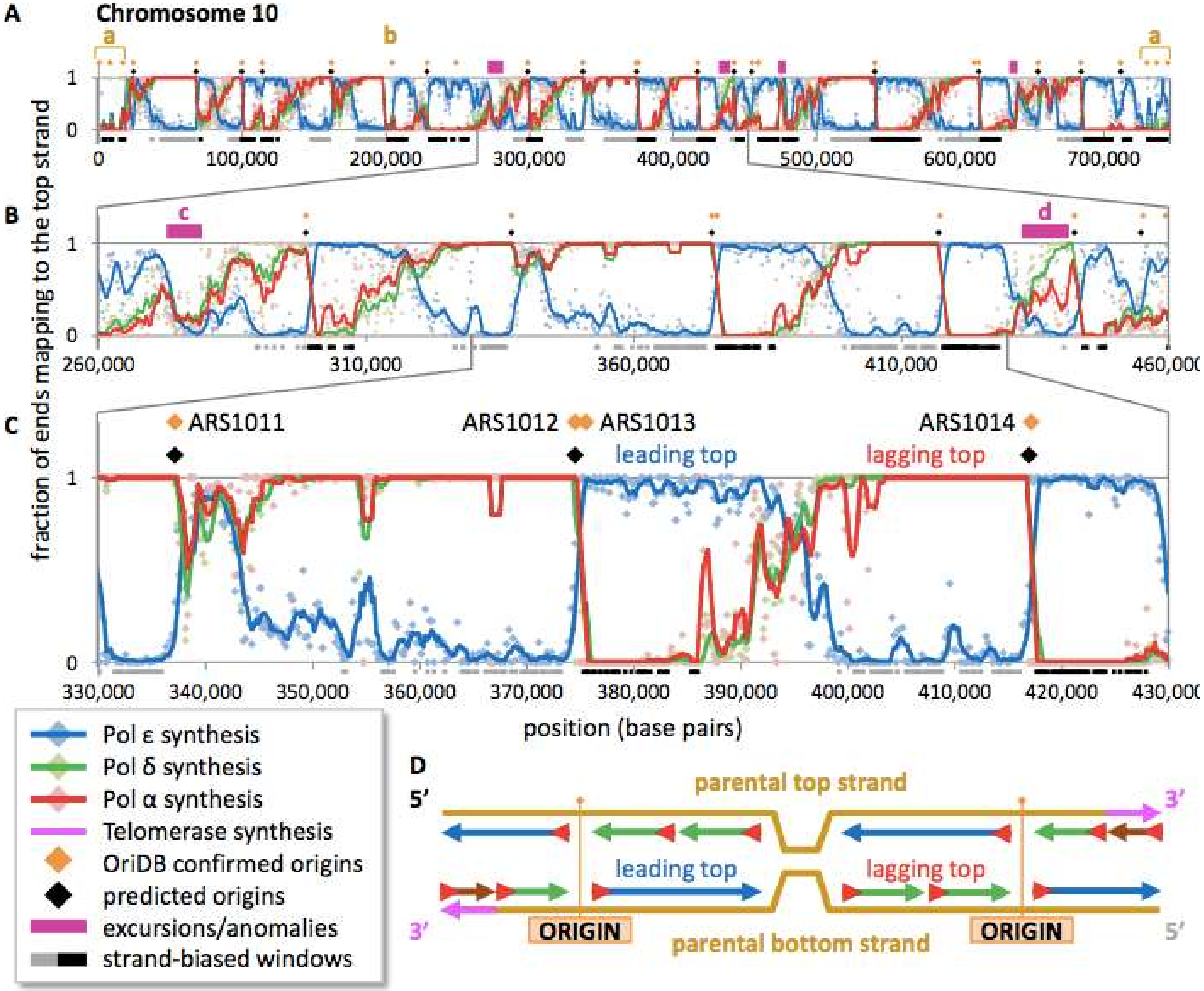


Hydrolytic DNA End-Sequencing

(HydEn-Seq)

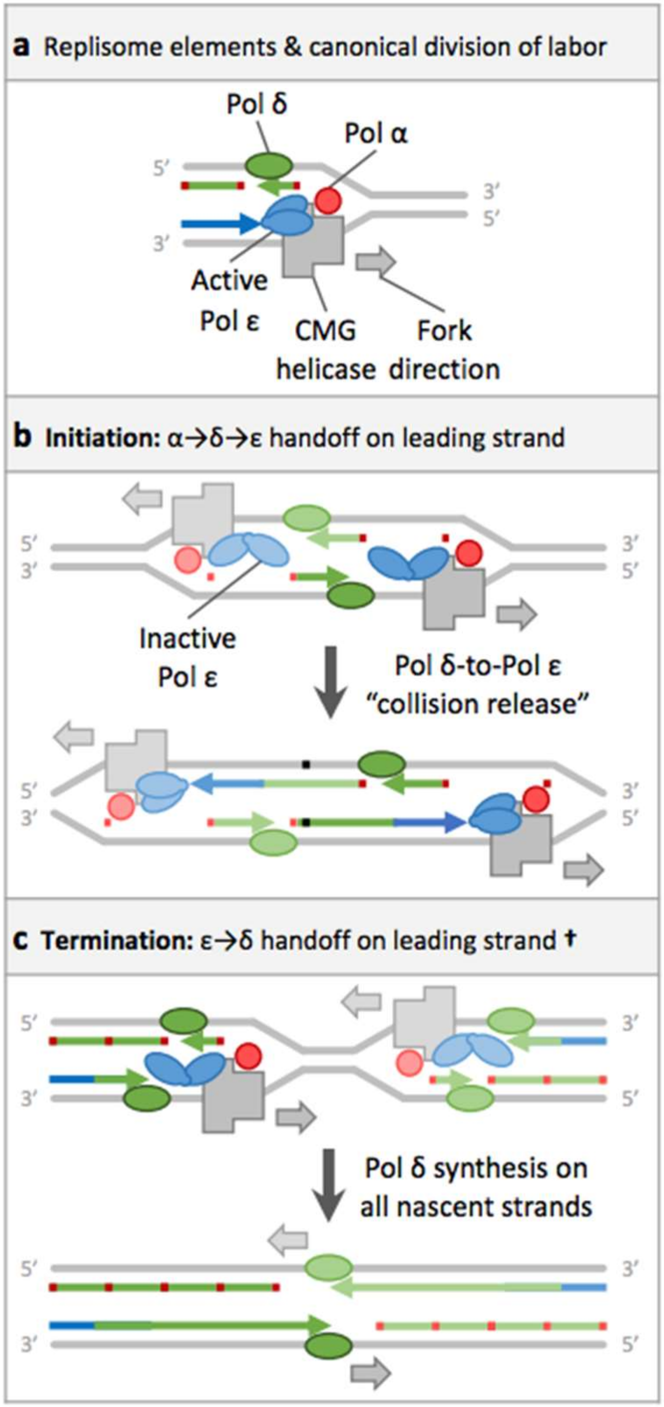


HydEn-Seq Mapping of Ribonucleotides in RER-Deficient *S. cerevisiae* - 2015



Indicates Roles of Replicases, Identifies Replication Origins

Current Model



1989 - 1990

Catherine Joyce, *J. Biological Chemistry*
Fred Perrino and Larry Loeb, *PNAS*

Biochemical evidence suggests that after a polymerase generates a mismatch, the mismatch can be removed by an exonuclease in a separate protein.

Intrinsic Proofreading

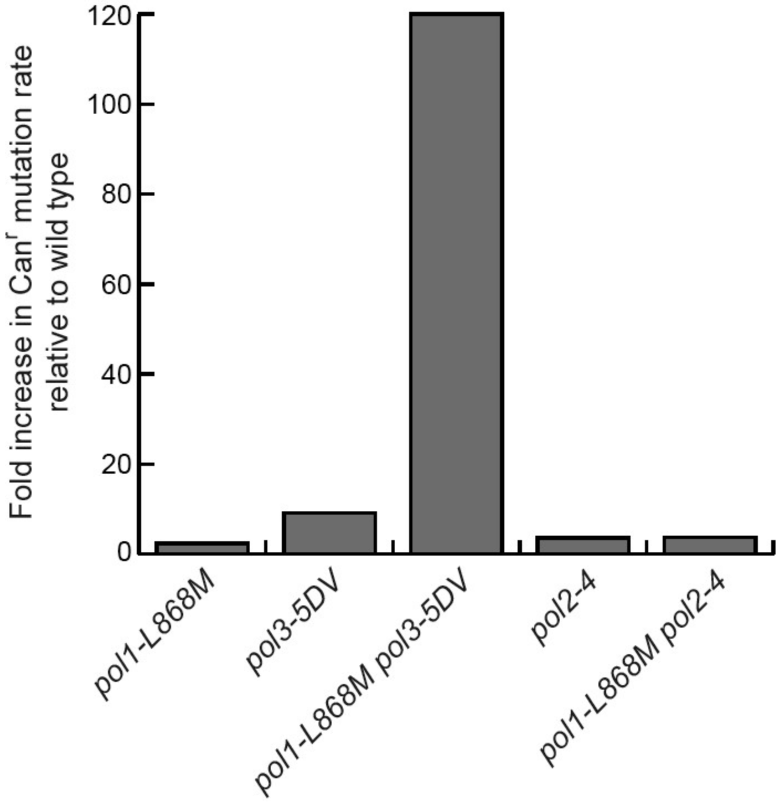
A mismatch is made and is then removed without intervening dissociation of the polymerase

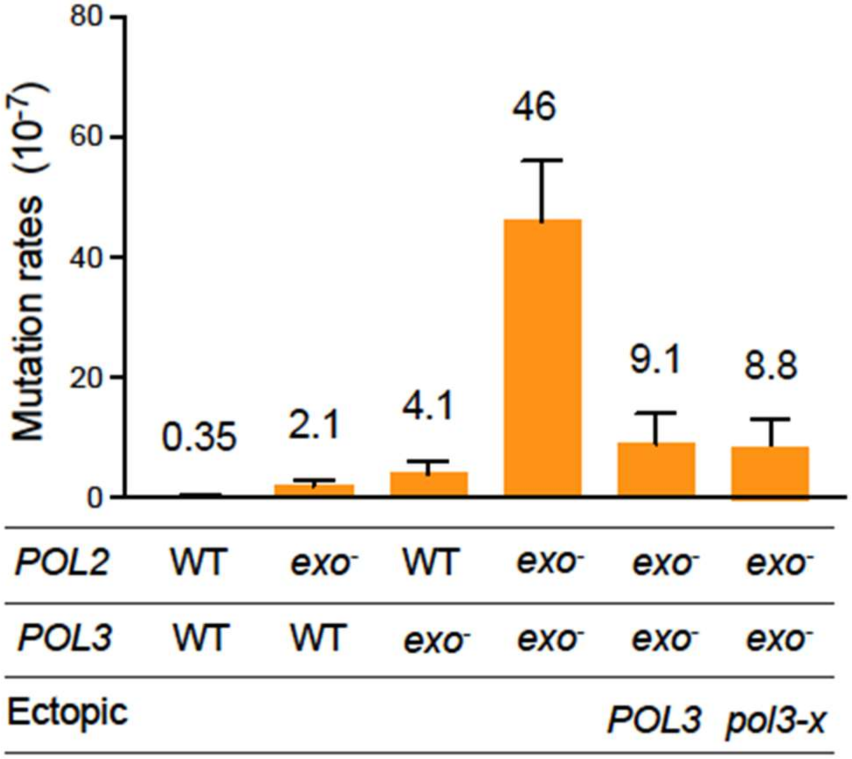
Extrinsic Proofreading

A mismatch is made, the polymerase dissociates, and an enzyme then binds the mismatch in its exonuclease active site and removes it.

Evidence for extrinsic proofreading of replication errors by Pol δ

Pavlov *et al.*, *Current Biology*, 2006; Zhou *et al.*, *Nature Structural Molecular Biology*, 2021





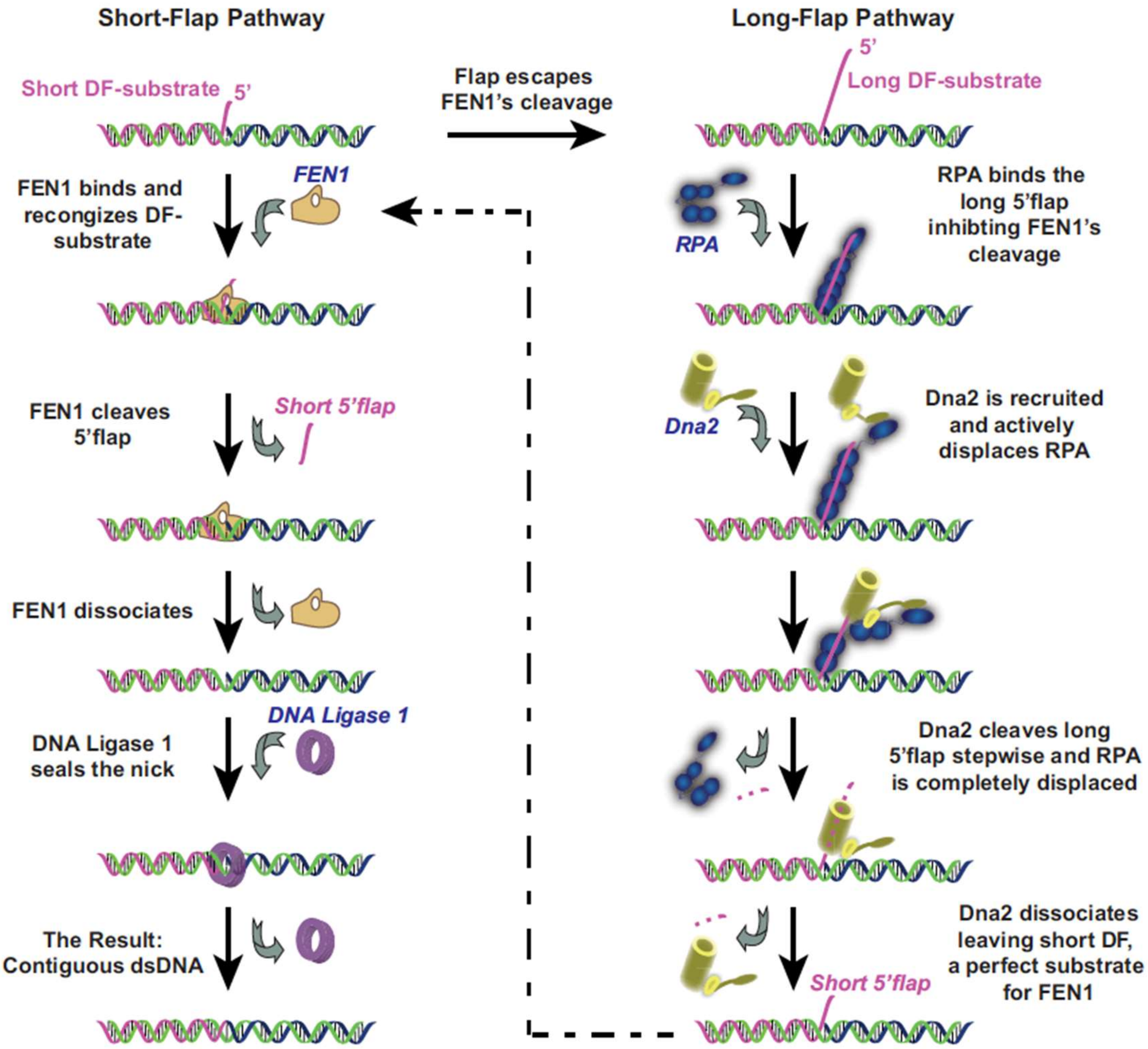
POL2	WT	*exo*$^-$	WT	*exo*$^-$	*exo*$^-$	*exo*$^-$
POL3	WT	WT	*exo*$^-$	*exo*$^-$	*exo*$^-$	*exo*$^-$
Ectopic					*POL3*	*pol3-x*

These data are consistent with the hypothesis that Pol δ can extrinsically proofread replication errors *in vivo*.

Extrinsic Proofreading By Pol δ

- Corrects mismatches made by all three replicases
- Is efficient
- Is independent of the polymerase activity of Pol δ
- Is largely independent of DNA mismatch repair
- Its specificity differs from intrinsic proofreading
- Balances leading and lagging strand replication fidelity
- Is relevant to origins of cancer and to evolution

Okazaki Fragment Maturation



Short-Flap Pathway

FEN1 binds and recongizes DF-substrate

FEN1 cleaves 5'flap

FEN1 dissociates

DNA Ligase 1 seals the nick

The Result: Contiguous dsDNA

Flap escapes FEN1's cleavage

Long-Flap Pathway

RPA binds the long 5'flap inhibting FEN1's cleavage

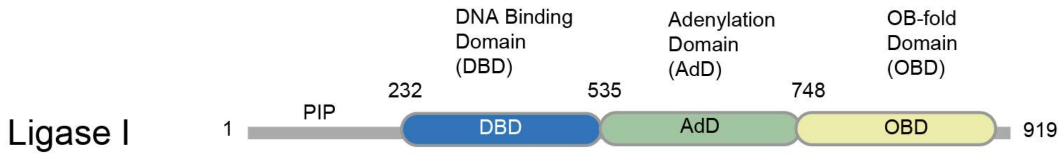
Dna2 is recruited and actively displaces RPA

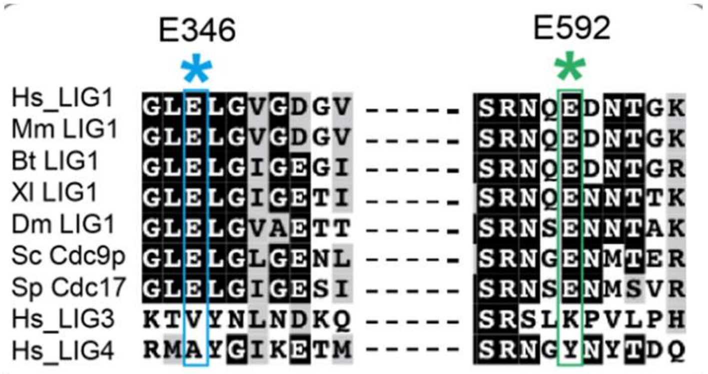
Dna2 cleaves long 5'flap stepwise and RPA is completely displaced

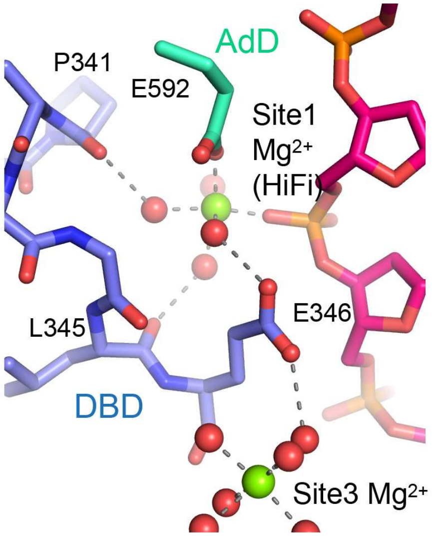
Dna2 dissociates leaving short DF, a perfect substrate for FEN1

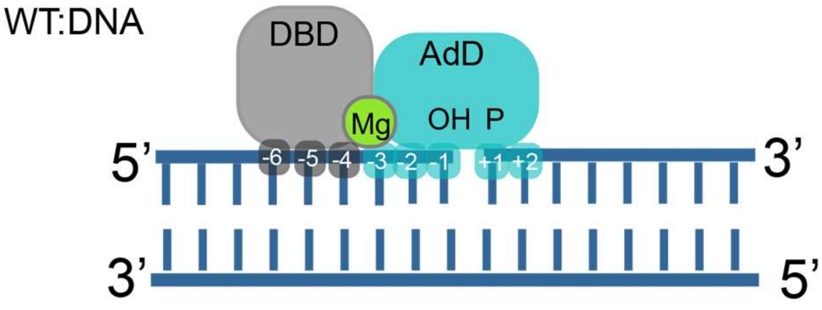
High-resolution LIG1-DNA structure

R.S. Williams *et al.*, *Nature Communications* (2019)









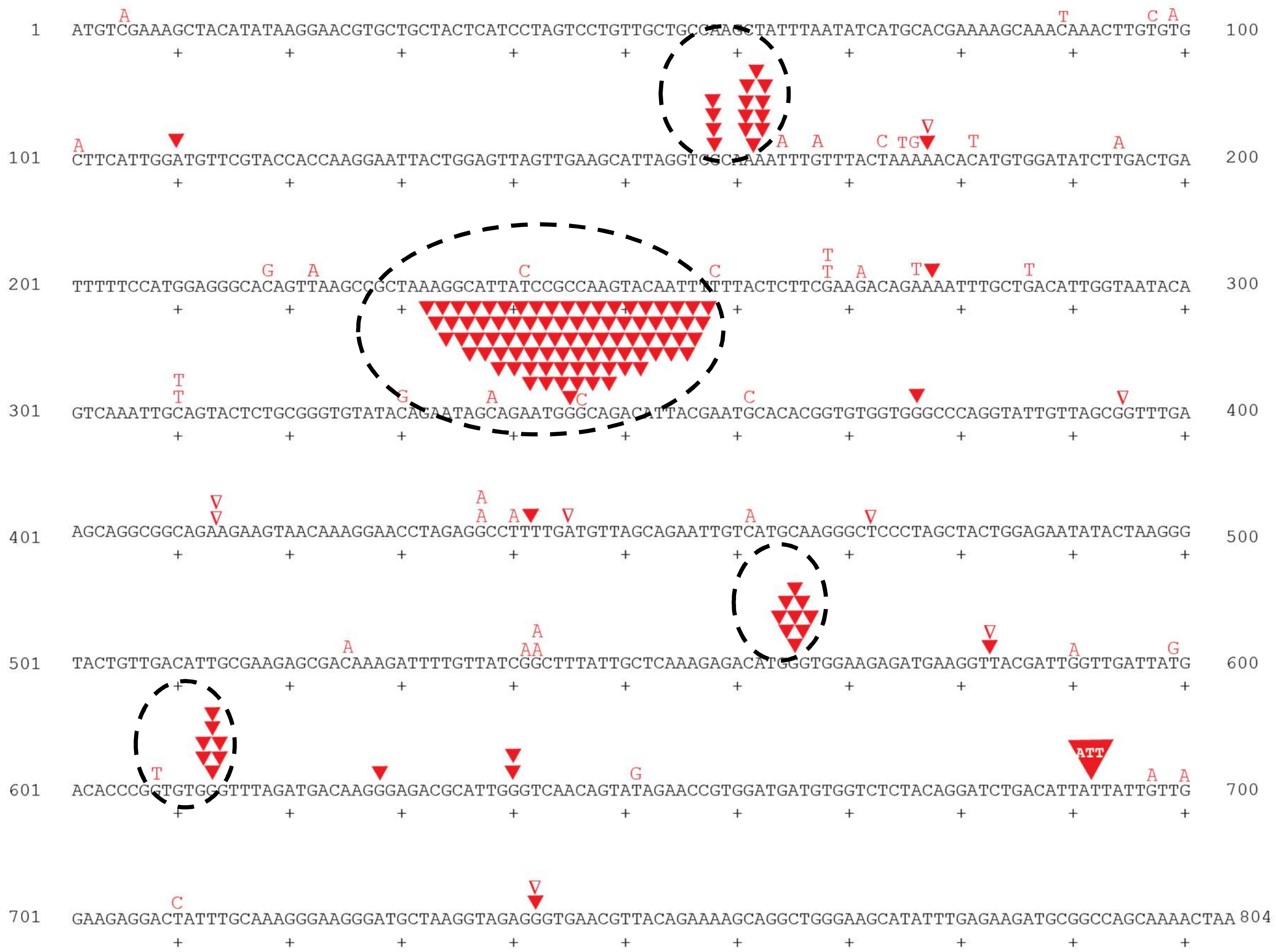
Hypothesis:

DNA ligase 1 fidelity is important for genome stability

Approach:

Use budding yeast to probe the biological consequences of expressing a *cdc9-EE/AA* mutant of DNA ligase 1

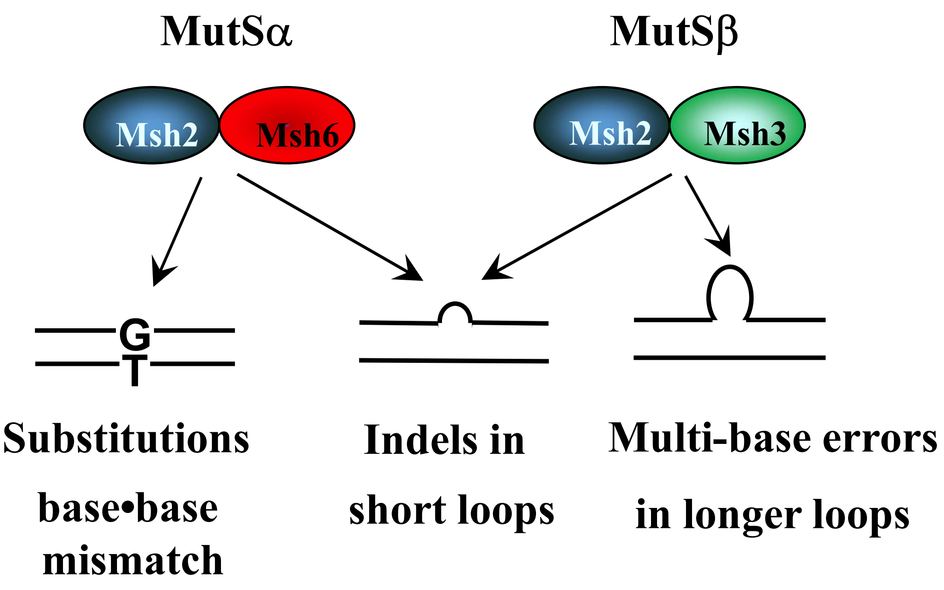
Spectrum of *ura3* mutants in Lig1 EE-AA mutant



Dominated by one base additions in short mononucleotide runs

How might the +1 insertion mutations in the *cdc9-EE/AA* strain be corrected?

DNA Mismatch Repair

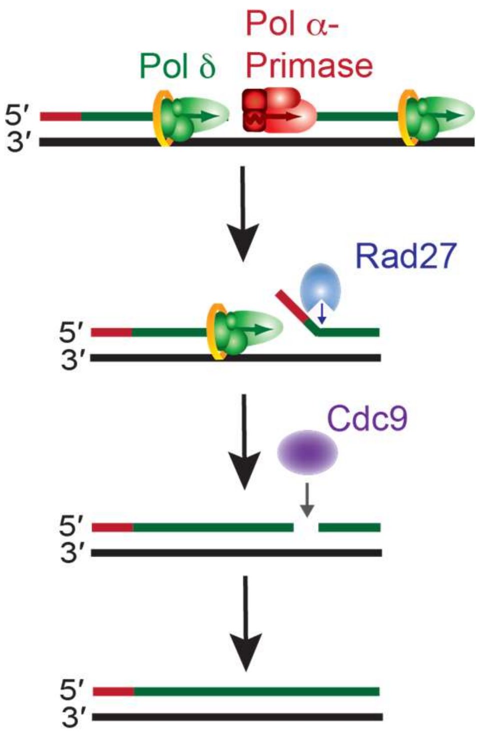


Substitutions base•base mismatch

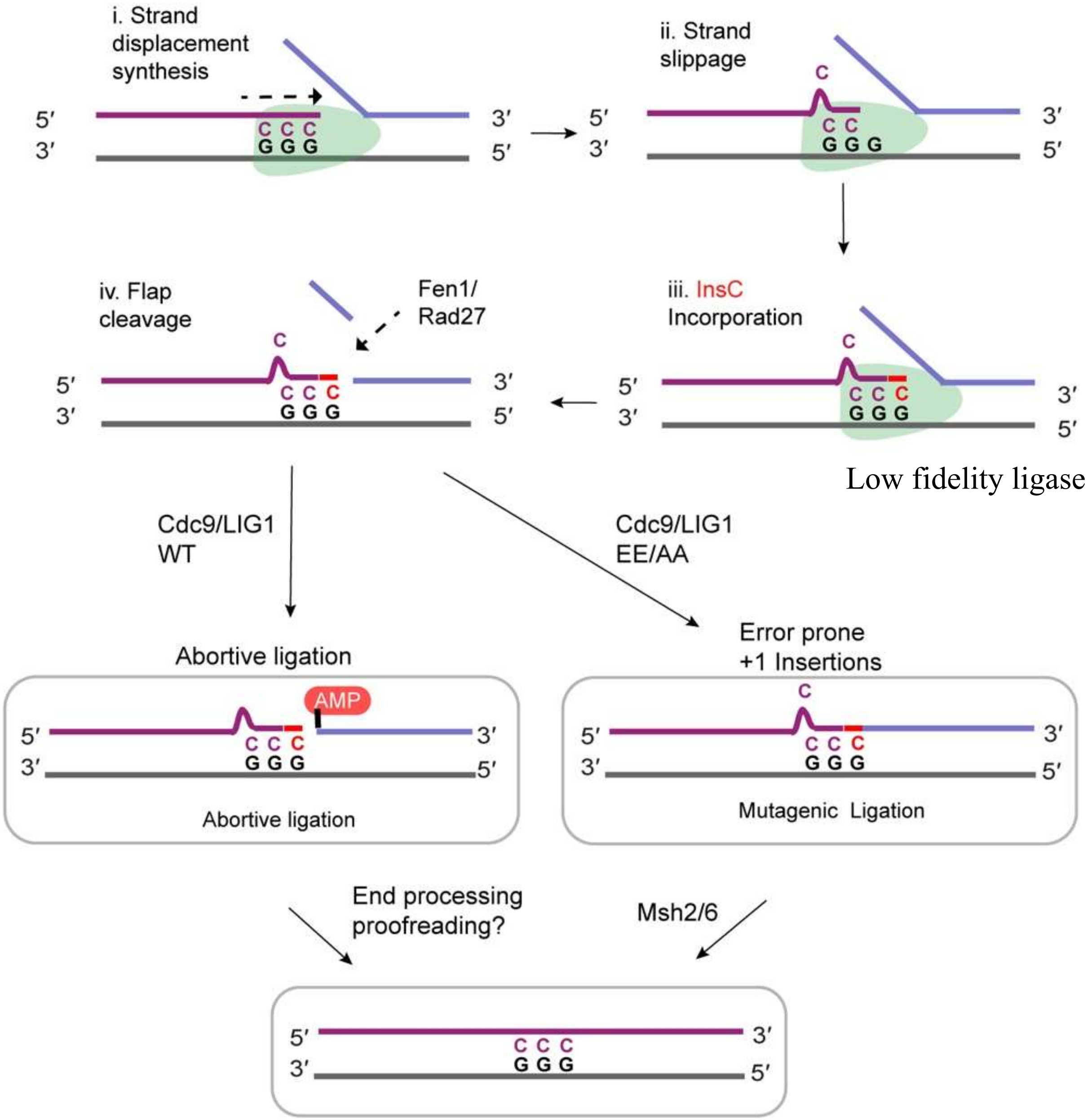
Indels in short loops

Multi-base errors in longer loops

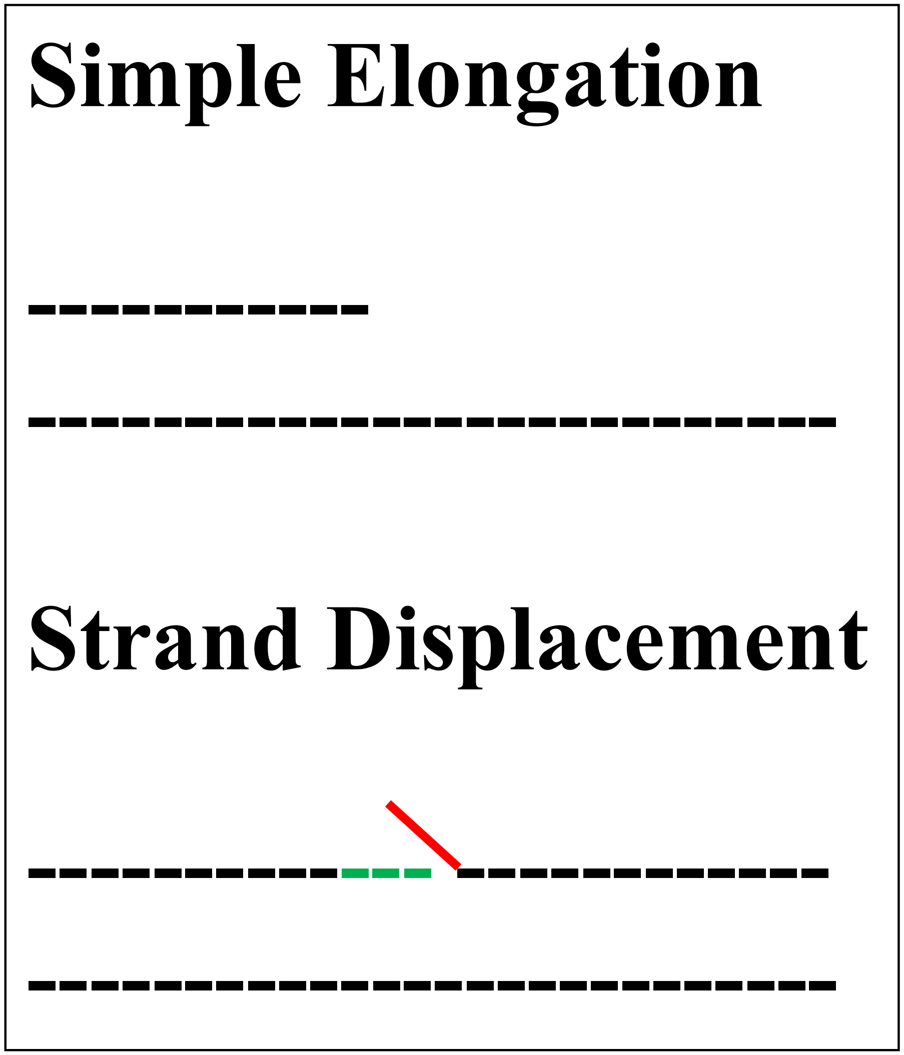
Flap Processing During OFM



DNA ligase 1 influences the fidelity of Okazaki fragment maturation



Low fidelity ligase



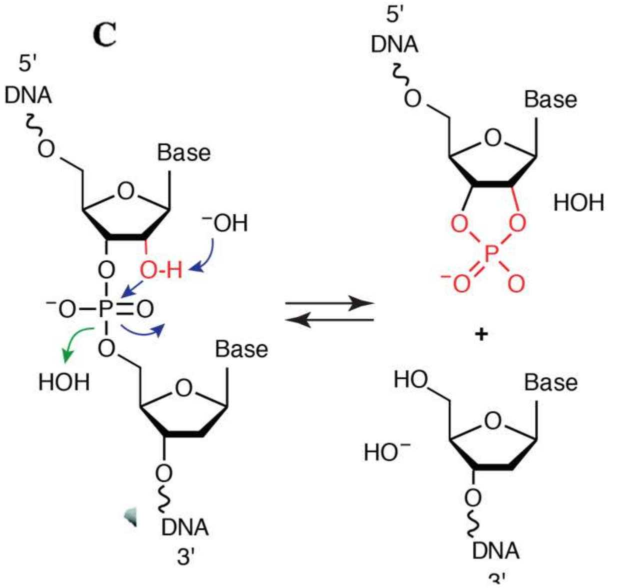
Simple Elongation

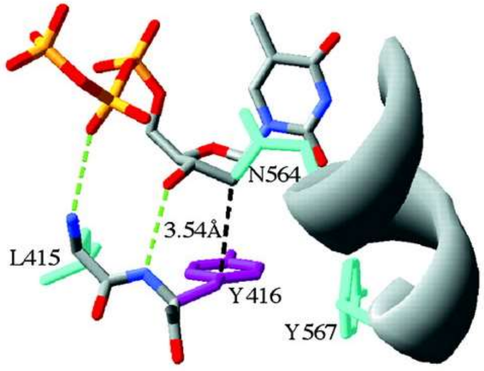
Strand Displacement

Are +1 errors common to strand displacement synthesis reactions?

Are rNTPs incorporated during DNA replication?

(Fidelity issue, but for sugar, not base)



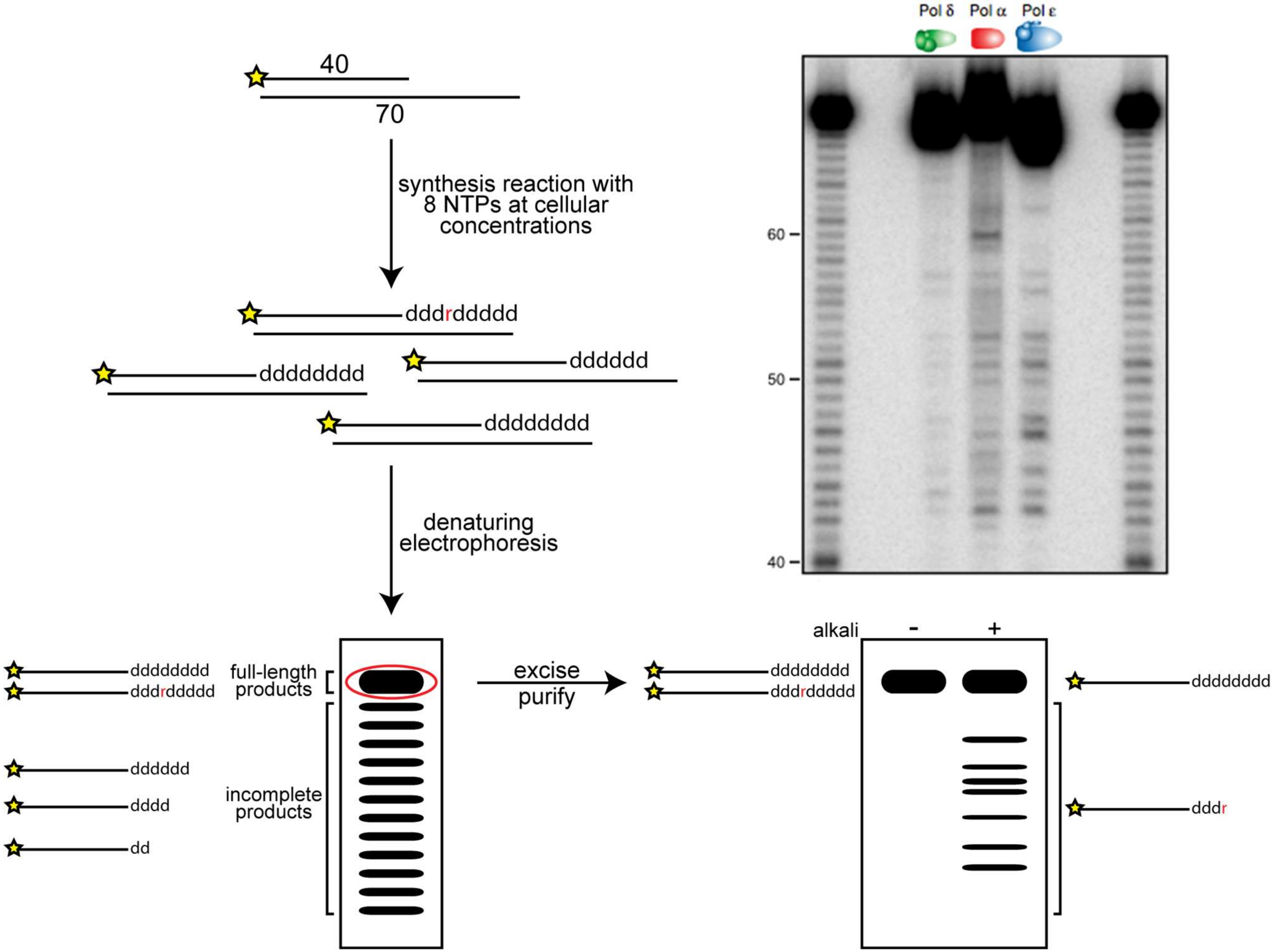


dNTP and rNTP Pools in Yeast

(Courtesy of Andrei Chabes)

dNTP	Concentration *in vivo* (μM)	rNTP	Concentration *in vivo* (μM)	Fold Difference
dA	16	rA	3000	190
dC	14	rC	500	36
dG	12	rG	700	58
dT	30	rU	1700	57

rNTP Incorporation *In Vitro*



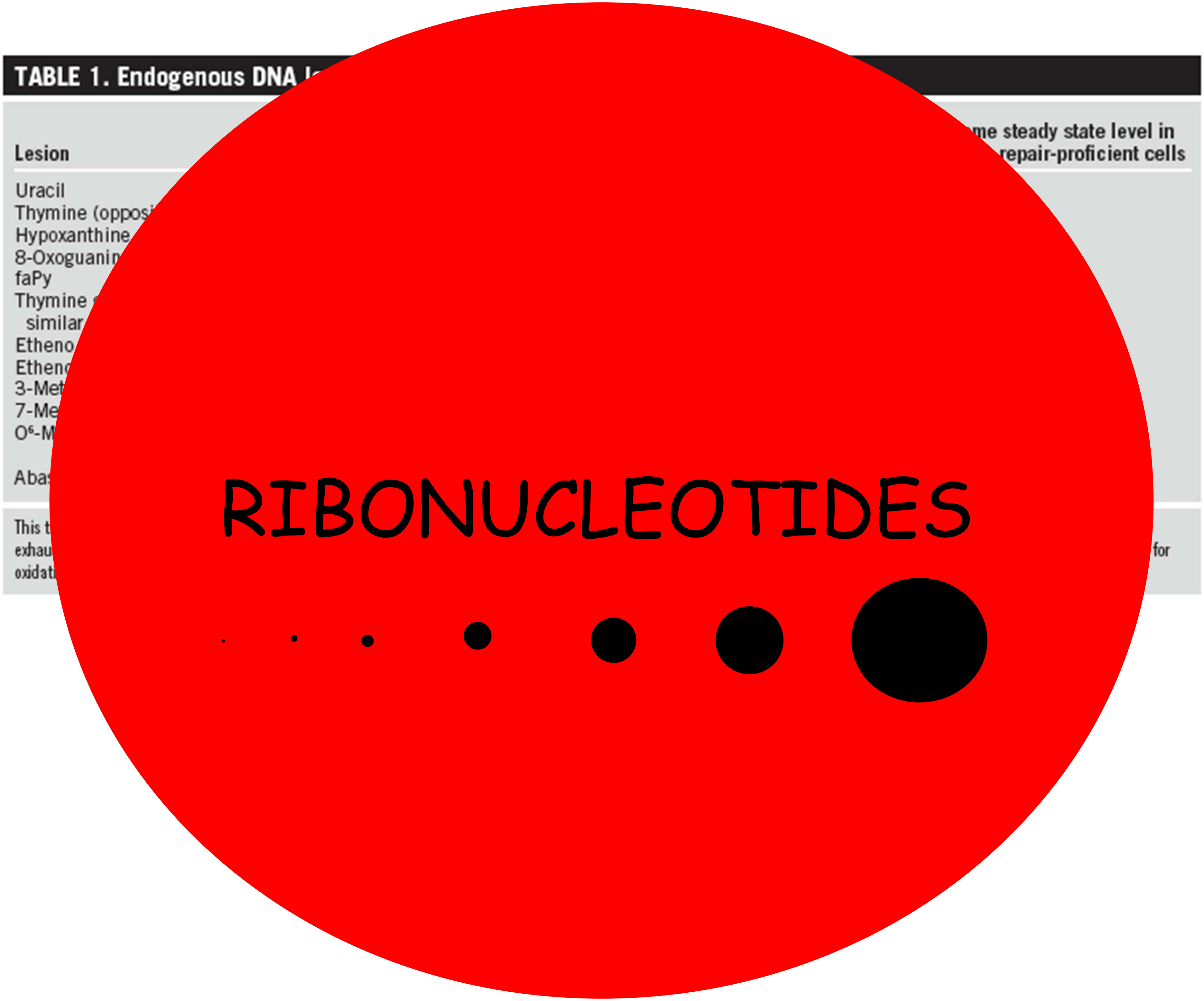


TABLE 1. Endogenous DNA le

Lesion

me steady state level in
repair-proficient cells

Uracil
Thymine (opposi
Hypoxanthine
8-Oxoguanin
faPy
Thymine
similar
Etheno
Etheno
3-Met
7-Me
O⁶-M
Abas

This t
exhau
oxidati
for

RIBONUCLEOTIDES

Ribonucleotides in DNA
Origins, Repair and Consequences

Williams & Kunkel, *Annual Reviews Biochemistry*, online

DNA Replicases

~ 13,000 in Yeast

>1,100,000 in Mouse

Mitochondrial Pol γ

RNA Primase

~150,000,000

Prim-Pol

X Family Pols

Y Family Pols

DNA Replicases

Proofreading (weak)

RNase H2-Dependent RER

Top1 Removal

RNA Primase

Okazaki Fragment Maturation

Replication Stress

Genome Instability

(short deletions, GCRs, LOH)

Cell Death

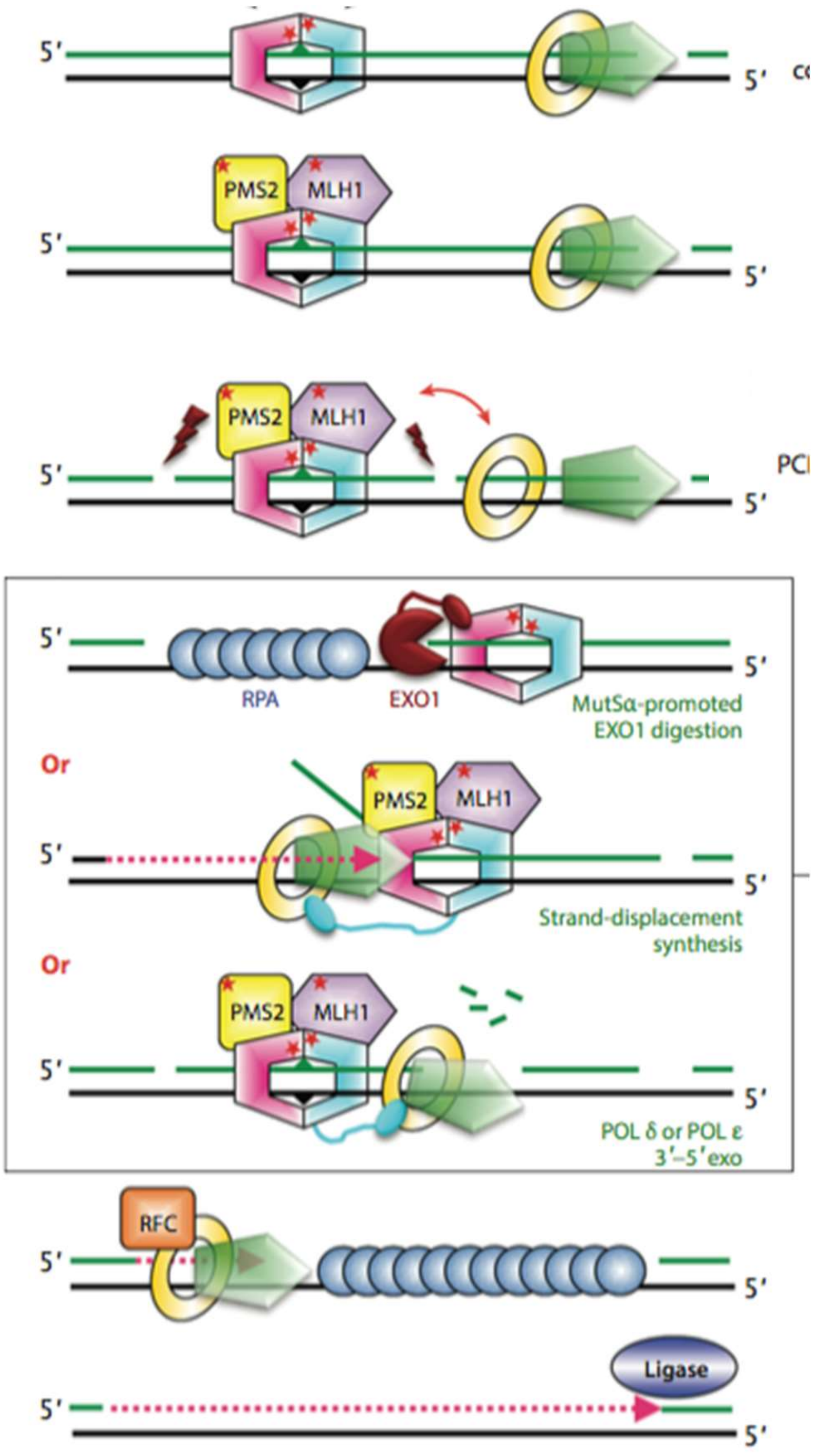
Diseases

MMR

Mating Type Switching

Some effects may be related to other roles of RNase H2 (e.g., resolving R-loops)

Eukaryotic DNA Mismatch Repair



MutSα recognition

MutLα binding

MutLα incision

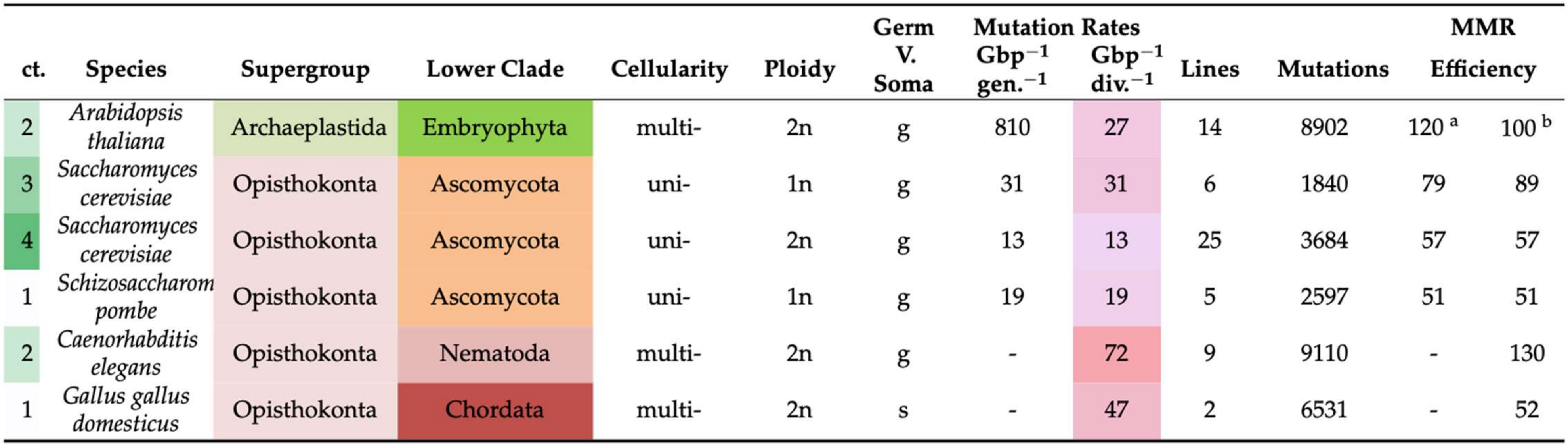
Mismatch removal

DNA synthesis

Ligation

Evolutionary conservation of MMR efficiency

Lujan & Kunkel, *Cells* 10, 1224 (2021)

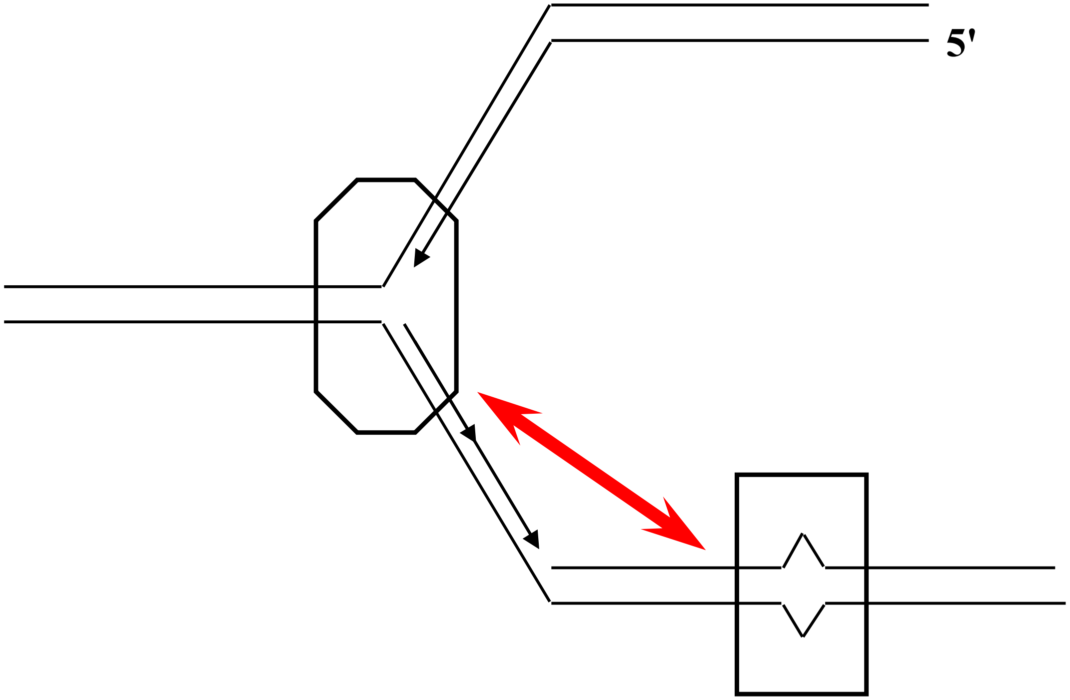


ct.	Species	Supergroup	Lower Clade	Cellularity	Ploidy	Germ V. Soma	Mutation Rates Gbp^{-1} gen.$^{-1}$	Gbp^{-1} div.$^{-1}$	Lines	Mutations	MMR Efficiency	
2	*Arabidopsis thaliana*	Archaeplastida	Embryophyta	multi-	2n	g	810	27	14	8902	120[a]	100[b]
3	*Saccharomyces cerevisiae*	Opisthokonta	Ascomycota	uni-	1n	g	31	31	6	1840	79	89
4	*Saccharomyces cerevisiae*	Opisthokonta	Ascomycota	uni-	2n	g	13	13	25	3684	57	57
1	*Schizosaccharom pombe*	Opisthokonta	Ascomycota	uni-	1n	g	19	19	5	2597	51	51
2	*Caenorhabditis elegans*	Opisthokonta	Nematoda	multi-	2n	g	-	72	9	9110	-	130
1	*Gallus gallus domesticus*	Opisthokonta	Chordata	multi-	2n	s	-	47	2	6531	-	52

Studying Relationships Between Replication and MMR *In Vivo*

"--- it may be that mismatch repair acts in a directed manner,
-- possibly because of a *special relation* to the replication complex."

Wagner & Meselson (1976) *PNAS*



Does MMR Efficiency Depend On:

Polymerase that makes error

MMR sub-pathway

Mismatch composition

Leading v Lagging Strand

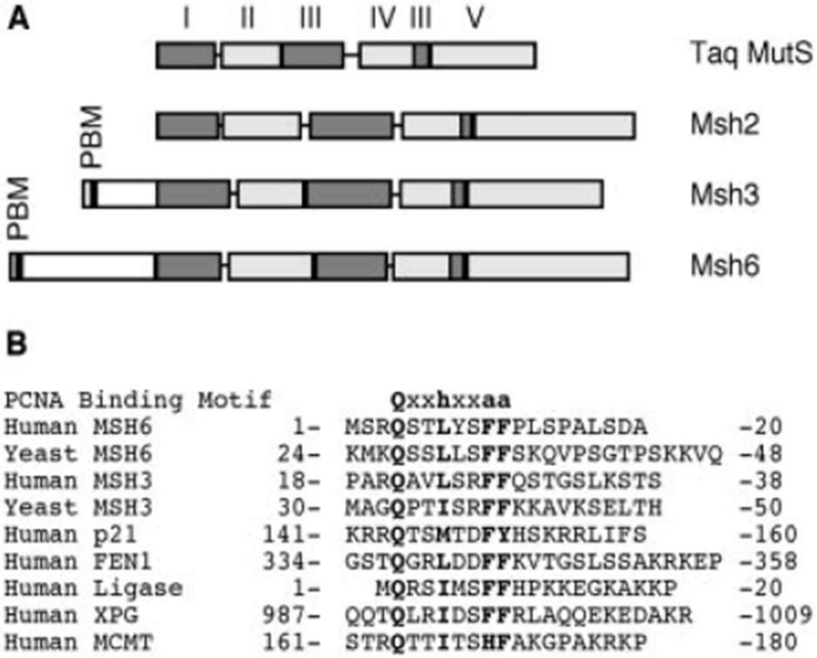
Sequence context (e.g., non-B DNA)

Replication timing

Chromosomal location

Chromatin status

Requirement for PCNA in DNA Mismatch Repair at a Step Preceding DNA Resynthesis



PCNA Binding Motif		**Q**xx**h**xx**aa**	
Human MSH6	1-	MSR**Q**ST**L**YS**FF**PLSPALSDA	-20
Yeast MSH6	24-	KMK**Q**SS**L**LS**FF**SKQVPSGTPSKKVQ	-48
Human MSH3	18-	PAR**Q**AV**L**SR**FF**QSTGSLKSTS	-38
Yeast MSH3	30-	MAG**Q**PT**I**SR**FF**KKAVKSELTH	-50
Human p21	141-	KRR**Q**TS**M**TD**FY**HSKRRLIFS	-160
Human FEN1	334-	GST**Q**GR**L**DD**FF**KVTGSLSSAKRKEP	-358
Human Ligase	1-	M**Q**RS**I**MS**FF**HPKKEGKAKKP	-20
Human XPG	987-	QQT**Q**LR**I**DS**FF**RLAQQEKEDAKR	-1009
Human MCMT	161-	STR**Q**TT**I**TS**HF**AKGPAKRKP	-180

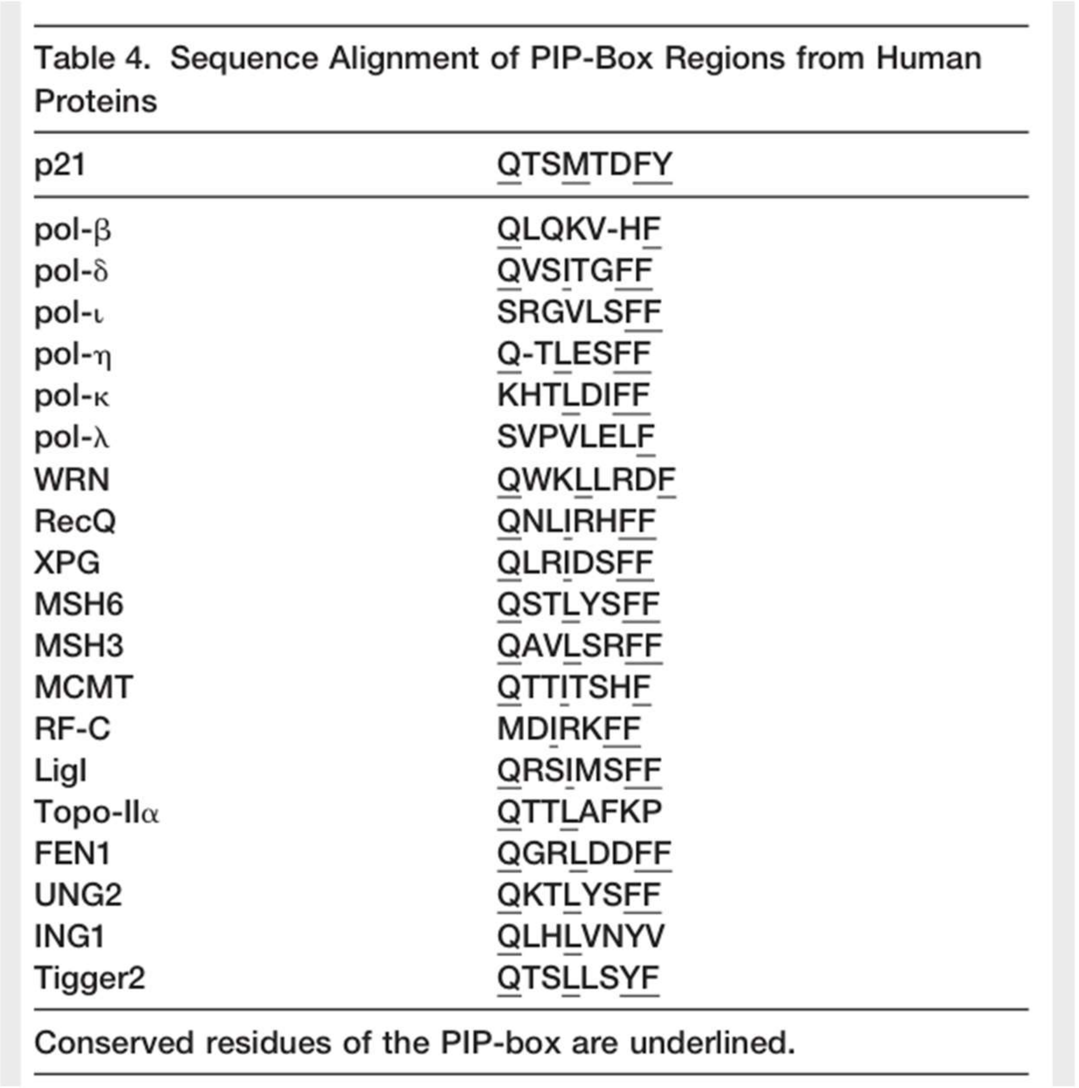


Table 4. Sequence Alignment of PIP-Box Regions from Human Proteins

p21	QTSMTDFY
pol-β	QLQKV-HF
pol-δ	QVSITGFF
pol-ι	SRGVLSFF
pol-η	Q-TLESFF
pol-κ	KHTLDIFF
pol-λ	SVPVLELF
WRN	QWKLLRDF
RecQ	QNLIRHFF
XPG	QLRIDSFF
MSH6	QSTLYSFF
MSH3	QAVLSRFF
MCMT	QTTITSHF
RF-C	MDIRKFF
LigI	QRSIMSFF
Topo-IIα	QTTLAFKP
FEN1	QGRLDDFF
UNG2	QKTLYSFF
ING1	QLHLVNYV
Tigger2	QTSLLSYF

Conserved residues of the PIP-box are underlined.

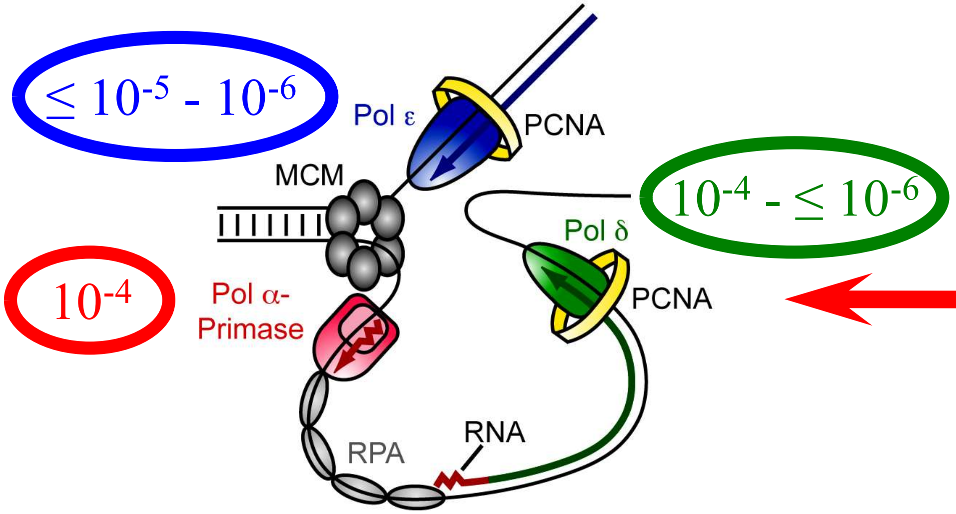
Umar *et al.*, *Cell* (1996)

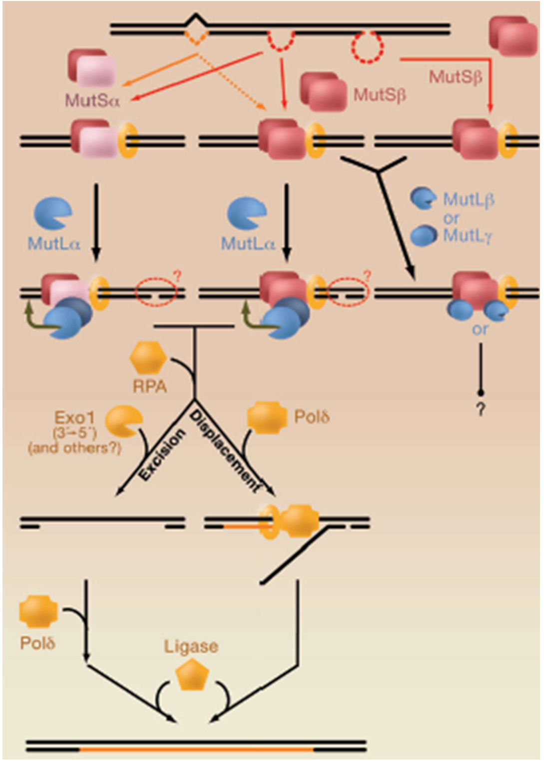
Clark *et al.*, *JBC* (2000)

Replication of Undamaged Eukaryotic Nuclear DNA

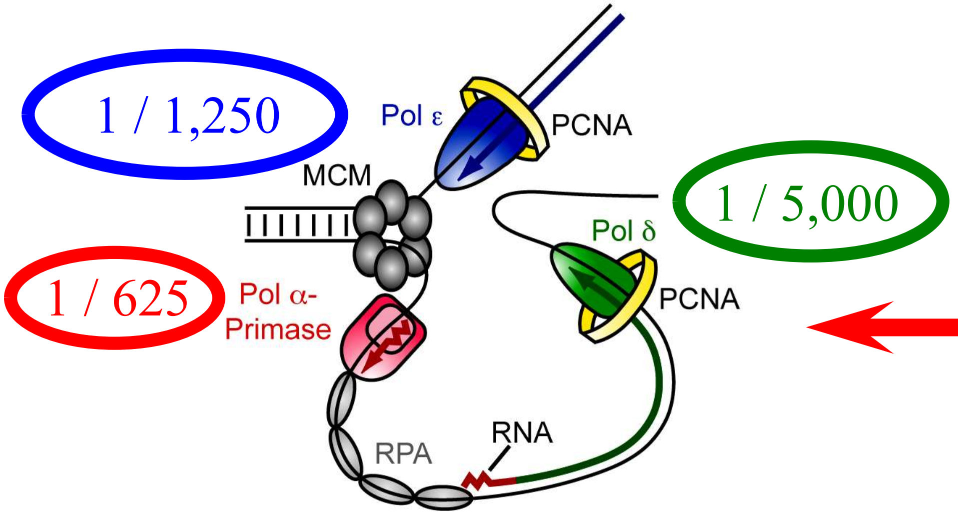
Mismatch Repair

Substitutions





Ribonucleotides



Ribonucleotide Excision Repair