Plant Omics Journal POJ

POJ 6(4):302-310 (2013) ISSN:1836-3644

Cloning and expression profiling of polycomb gene, DNA Polymerase Alpha (POLα) from tomato *Solanum lycopersicum* L.

Zainab M. Almutairi[1]*, Monther T. Sadder[2,3]

[1]Biology Department, College of Science and Humanity Studies, P. O. Box 83, Salman Bin Abdulaziz University, Alkharj 11942, Saudi Arabia
[2]Center of Excellence in Biotechnology Research, King Saud University, PO Box 2455, Riyadh 11451, Saudi Arabia
[3]Plant Biotechnology Lab, Faculty of Agriculture, The University of Jordan, Amman, 11942, Jordan

***Corresponding author: z.almutairi@sau.edu.sa**

Abstract

The decision to replicate its DNA is of crucial importance for every cell and, in many organisms, is decisive for the progression through the entire cell cycle. DNA polymerases are required to maintain the integrity of genome during DNA replication. DNA Polymerase Alpha (POLα) encodes the putative catalytic subunit of the DNA polymerase α which plays an essential role in the initiation of DNA replication. Here, we present the cloning, characterization and expression of a putative *SlPOLα* gene in tomato by isolating cDNA clones corresponding to *SlPOLα* gene from tomato using primers designed from available ESTs based on conserved sequences between PcG genes in *Arabidopsis thaliana* and tomato. The *SlPOLα* cDNAs were cloned into pBS plasmid and sequenced. Both 5' and 3' RACE were generated and sequenced. The flcDNA for *SlPOLα* gene length was composed of 4,944 bp, containing 103 bp in the 5'UTR; 4,683 bp in the ORF; and 158 bp in 3'UTR. The translated ORF encodes a polypeptide of 1,561 amino acids. Alignment of deduced amino acids indicated that there are highly conserved regions between tomato *SlPOLα* predicted protein and hypothetical plant POLα gene family members. Both unrooted phylogenetic trees, constructed using maximum parsimony and maximum likelihood methods indicated that there is a close relationship between *SlPOLα* predicted protein and POLα protein of *Mimulus guttatus*. *SlPOLα* was strongly expressed in proliferating tissues (closed floral buds) and moderately expressed in flower tissues, whereas, it was weakly expressed in non-proliferating tissues (unripe fruit).

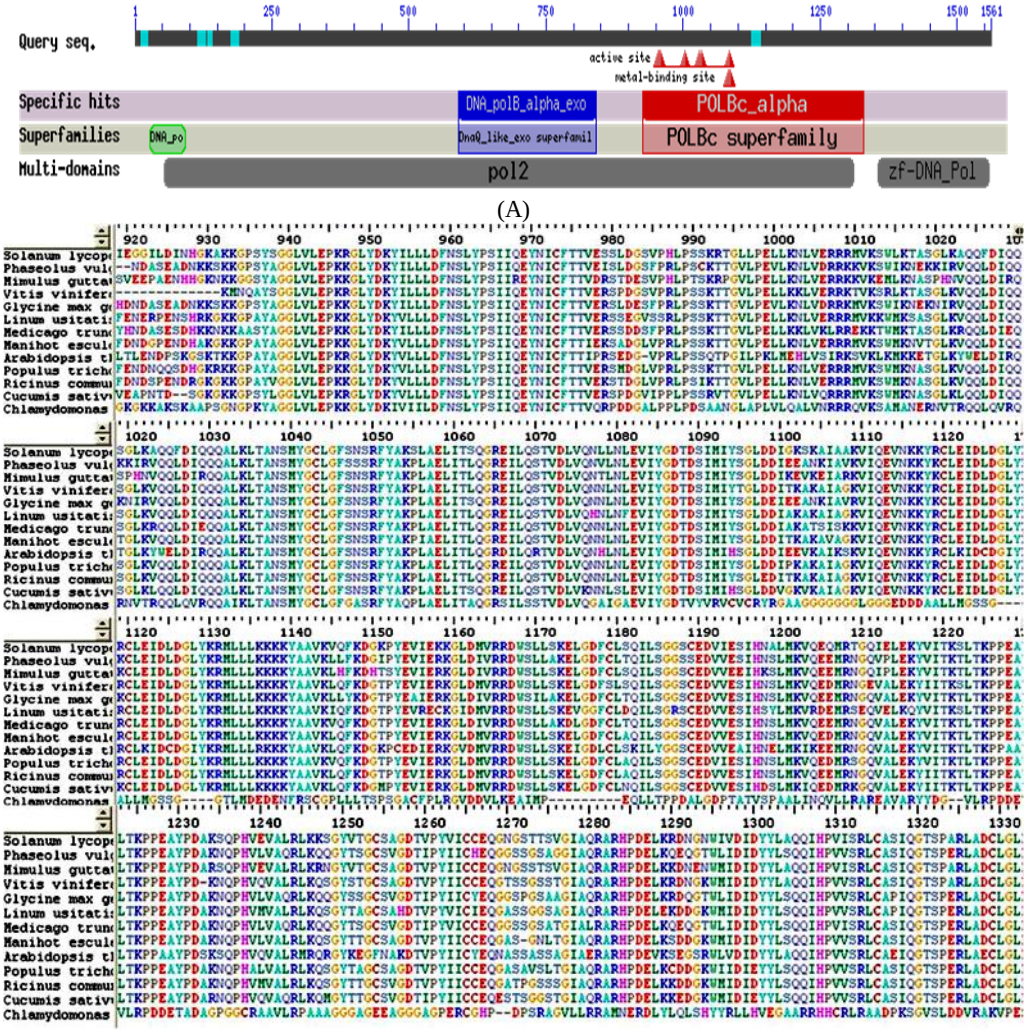
Keywords: Tomato, DNA polα gene, flcDNA, phylogenetics.
Abbreviations: *SlPOLα_Solanum lycopersicum* L. DNA Polymerase Alpha; POL_DNA polymerase; PcG_Polycomb group; SGN_ Solanaceae Genomics Network database; CDD_NCBI Conserved Domain database; UTR_Untranslated region; ICU2_ INCURVATA 2; ORF_Open reading frame; flcDNA_Full length cDNA; RACE_Rapid amplification of cDNA ends; EST_ Expressed sequence tag; qRT-PCR_Quantitative real time PCR.

Introduction

The tomato genome was sequenced as the cornerstone of an International Solanaceae Genome Initiative, a project that aims to develop the family Solanaceae as a model for biological systems for understanding plant adaptation and diversification. Tomato genome comprises approximately 950 Mb of DNA, more than 75% of which is heterochromatin and largely devoid of genes (SGN, 2013). Tomato genome has recently been achieved and published on the Solanaceae Genomics Network database (SGN) (http://solgenomics.net/ tomato/), as well as the sequence of a wild relative. Of estimated 950 Mb genome size, 760 Mb were assembled in 91 scaffolds aligned to the 12 tomato chromosomes (The Tomato Genome Consortium, 2012). The tomato POLα gene is homologous to the *Arabidopsis thaliana* POLα gene (also called INCURVATA 2 or ICU2) which encodes the catalytic subunit of the DNA polymerase α (Barrero et al., 2007) and one of the various proteins that participate in DNA replication, repair and recombination; these include chromatin assembly factors, POLδ, POLε, DNA primase, replication factor C (RFC), proliferating cell nuclear antigen (PCNA) and other proteins (Garg and Burgers, 2005; Moldovan et al., 2007). *Incurvata* mutants were isolated by G Röbbelen in 1957 and were found to belong to *A. thaliana* as per the Arabidopsis Information Service Form Mutants collection. Thereafter, INCURVATA genes were identified as genes involved in the control of leaf morphogenesis by screening 13 *A. thaliana* mutants with curled and involute leaves (Serrano-Cartagena et al., 2000). The strong *icu2-2* and *icu2-3* insertional alleles caused fully penetrant zygotic lethality when homozygous and incompletely penetrant gametophytic lethality, probably because of loss of DNA polymerase activity. The weak *icu2-1* allele carried a point mutation and caused early flowering, leaf incurvature, and homeotic transformations of sepals into carpels and of petals into stamens (Barrero et al., 2007). In eukaryotic cells, DNA polymerases are required to maintain integrity of the genome during biological processes such as replication, various repair events, translesion synthesis and recombination and also in regulatory events such as cell control and DNA damage checkpoint function (Hübscher et al., 2000). Perhaps the most important function of DNA polymerases is to synthesize an exact replica of the genome during the replication process. Eukaryotic nuclear replication is a bidirectional process initiated at one of the several origins of replication, where a carefully regulated complex of proteins unwinds the DNA and facilitates the assembly of a replication fork (Garg and

Table 1. The eleven primer pairs used to amplify cDNA clones in this study. 'F' and 'R' in the primer names indicate direction (forward or reverse) of the primer.

Primer name	Primer sequence
*SlPOLα*F01	5′-TGA GAA CGG CGG ATT ACA GG-3′
*SlPOLα*R01	5′-TAC ATC TCT GGA GAG CTT GG-3′
*SlPOLα*F02	5′-GGC TGA TAG GGA GAG AAG GCG AAG AGG-3′
*SlPOLα*R02	5′-GGA CTA TCT ACA ATG ACC TCA A-3′
*SlPOLα*F03	5′-CAG TGG GAA TAA GGG ATG GTG-3′
*SlPOLα*R03	5′-CCC ATA GGA AAT ATG CCT CCC TC-3′
*SlPOLα*F04	5′-CCA AGC TCT CCA GAG ATG TA-3′
*SlPOLα*R04	5′-TCC CCA AAG GTT ACC ACT GAT-3′
*SlPOLα*F05	5′-ATG CGC CCA TGT TAA CCT CC-3′
*SlPOLα*R05	5′-GGG CCT TGA GAC CAG ATG CT-3′
*SlPOLα*F06	5′-CAT CTA AGT GTT CTT CCT TTG ACT CG-3′
*SlPOLα*R06	5′-GGC CAG TCC TCA TTT GCT C-3′
*SlPOLα*F07	5′-ATC AGT GGT AAC CTT TGG GG-3′
*SlPOLα*R07	5′-TGG AGA ATC CTA GAC ATC CG-3′
*SlPOLα*F08	5′-GAA GGT GGG ATT CTG GAC ATC AAC C-3′
*SlPOLα*R08	5′-GGC AGC TAT TGC TTT GGA TTT TCC-3′
*SlPOLα*F09	5′-GCA AAG TCG CTA GCA GAG C-3′
*SlPOLα*R09	5′-ATT CCC CTC ACT CTC CTC TGG GC-3′
*SlPOLα*F10	5′-CGT GAC TGG AGC TTG CTG TC-3′
*SlPOLα*R10	5′-TTC AAC CAC AGG CCG TAT CC-3′
*SlPOLα*F11	5′-GCC CAG AGG AGA GTG AGG GGA AT-3′
*SlPOLα*R11	5′-CGG TAT TAG CCT TGC TAG AAT G-3′



(A)

(B)

Fig 1. Conserved domains identified in the sequence of the *SlPOLα* protein using NCBI CDD. (A): Conserved domains in *SlPOLα* protein; DNA polα N, DNA polα B exo, POLBcα and zf-DNA Pol. (B): Multiple sequence alignments of conserved POLBcα domain sequences from amino acids 926 to 1,329. Alignment with 12 POLα protein sequences of other plant species was performed using ClustalW 1.83 by BioEdit software.

Burgers, 2005; Kunkel and Burgers, 2008; Yeeles et al., 2013). Eukaryotic cells have at least four major nuclear DNA polymerases: POLα, POLß, POLδ and POLε. POLγ is found in mitochondria, although it is encoded by a nuclear gene. Plant chloroplasts also contain their own DNA polymerase that appears to be similar to POLγ (γ-like DNA polymerase) (Wang et al., 1991; Nielsen et al., 2010). POLα plays an essential role in the initiation of DNA replication. The POLα complex is composed of four subunits: the catalytic subunit POLα 1, the regulatory subunit POLα 2 and the small and the large primase subunits PRIM1 and PRIM2. The primase subunit of the POLα complex initiates DNA synthesis by oligomerising short RNA primers on both leading and lagging strands (Muzi-Falconi et al., 2003; Lévy et al., 2009). These primers are initially extended by the POLα catalytic subunit and subsequently transferred to POLδ and POLε for processive synthesis on the lagging and leading strand, respectively (Garg and Burgers, 2005; Vaara et al., 2012). In *A. thaliana*, 12 different DNA polymerases have been identified by sequence similarities. They fall into four main classes of eukaryotic DNA polymerases: classes A, B, X and Y (Burgers et al., 2001). The class B comprises four DNA polymerases, POLα, POLδ, POLε and POLζ. POLα, POLδ and POLε participate in nuclear DNA replication; POLδ and POLε are also required for proliferating cell nuclear antigen-dependent long-patch base-excision repair. POLζ involved for mutagenic DNA replication. These functions in plants are only speculative since they have been deduced from the functions attributed to their orthologues in other organisms (Garcia-Diaz and Bebenek, 2008; Sanchez et al., 2012). During DNA replication, not only nucleic acids but also DNA methylation, histone modifications and chromatin structure are accurately replicated (Ekanayake and Sabatini, 2011). Understanding the stable transmission of epigenetic information through multiple cell cycles has been the main focus of recent studies concerning epigenetic regulation (Sanchez et al., 2012; Yeeles et al., 2013). The epigenetic role of ICU2 was revealed in *A. thaliana*, where the mutation of *polα*/*icu2* reduced dimethylation of histone H3 lysine 9 (H3K9 me2) in the 35S promoter. The *polα* mutation also influences development of the shoot apical meristem, and delays the G2/M phase with high expression of a G2/M marker gene CycB1;1:GUS (Liu et al., 2010). Recent studies revealed that ICU2 ensures stable maintenance of repressive histone modifications and facilitates histone assembly in dividing cells, suggesting a possible mechanism for ICU2-mediated epigenetic maintenance (Hyun et al., 2013). FlcDNA clones are fundamental resources for molecular biology experimental investigations of gene functions, as well as for detection of intron–exon structures in genomes. The general aim of this work is to identify a putative PcG gene, *S. lycopersicum L. POLα (SlPOLα)*, from tomato by cloning and sequencing its flcDNA of mRNA transcripts in tomato tissues. *SlPOLα* gene has an epigenetic role in flowering and floral organs development and as far as we know, this is the first DNA polymerase gene to be sequenced and characterised from Solanaceae.

Results

Construction of full length cDNA for tomato SlPOLα gene

To isolate cDNA encoding for *SlPOLα* gene, PCR was performed using a primer set that targets the *SlPOLα* gene in cDNA synthesised from total RNA extracted from tomato closed floral buds. Amplification products were cloned in cloning vector, pBS, linearised with *Eco*RV restriction enzyme and transformed into *Escherichia coli* DH5α. The *SlPOLα* cDNA contig was obtained by overlapping 11 ESTs. The lengths of ESTs varied between 460 and 1,250 bp with an average length of 981 bp. Based on the sequence of the *SlPOLα* cDNA contig, 5′ and 3′ RACE gene specific primers were designed and used in 5′ and 3′ RACE amplification. The generated 5′ RACE fragment was 738 bp in length, and 3′ RACE was 505 bp. Both 5′ and 3′ RACE fragments were sequenced and assembled with obtained ESTs to construct *SlPOLα* flcDNA contig. Computer analysis using the BLASTN algorithm in SGN database confirmed that the obtained sequence corresponded to *SlPOLα* gene. The obtained *SlPOLα* flcDNA was located on tomato chromosome 2 in genomic region about 12,782 bp. Thirty-one exons were determined using flcDNA and BLASTN search in SGN database (Table 2; Supplementary Fig 1). The constructed *SlPOLα* flcDNA comprised 4,944 bp, containing 103 bp in 5′ UTR, 4,683 bp in ORF and 158 bp in 3′ UTR without poly (A) tail (Supplementary Fig 2). ORF encodes a polypeptide of 1,561 amino acids, spanning from np 104, where the first ATG codon is located, to np 4786, adjacent to a termination codon (TAG) (Supplementary Fig 3). The flcDNA sequence has been submitted to the NCBI GenBank as accession number JQ669019, where the deduced amino acid sequence is found under accession number AFD98849.

ORF analysis and protein homology

The 4,683 bp ORF from *SlPOLα* flcDNA encodes 1,561 amino acids. The calculated molecular mass of the predicted protein is 172.7 kDa, with an estimated isoelectric point of 7.02. Domain prediction using NCBI Conserved Domain Database (CDD) indicates that tomato *SlPOLα* protein includes, like in most others plants, four conserved domains. DNA polymerase α subunit p180 N terminal domain (DNA polα N) from amino acid 27 to 93, DNA polymerase family α B exonuclease domain (DNA polα B exo) from amino acid 591 to 842, DNA polymerase type-B α subfamily catalytic domain (POLBcα) from amino acids 926 to 1,329 and DNA Polymerase α zinc finger domain (zf-DNA Pol) from amino acid 1,355 to 1,557. The four domains appeared as conserved sequences in multiple sequence alignment for *SlPOLα* predicted protein and other homologous proteins as shown in POLBcα domain alignment (Fig 1). Protein homology search which was performed with BLASTP program shows that there is a high homology between the deduced amino acids sequence of tomato *SlPOLα*-predicted protein and POLα of *Vitis vinifera* which revealed the highest score value (2,184 bits). The second highest score value of 2,149 bits was observed with POLα of *Ricinus communis*. Sequence identity observed with angiosperms species, both monocot and dicot, ranged between 55% and 75% for best 13 hits. Homologous POLα proteins in non-seed vascular plant, *Selaginella moellendorffii* and non-vascular plant, *Physcomitrella patens*, revealed 64% and 58% sequence identity with tomato *SlPOLα* protein, respectively.

Phylogenetic relationships between SlPOLα and homologous proteins

To investigate the relationships between known POLα proteins from different plant species, phylogenetic analyses were conducted on homologous protein sequences of other plant species of hypothetical gene family, reported in Phytozome database (code #31830449). Protein sequences for 12 members, which are related to *SlPOLα* gene, were chosen for phylogenetic analysis. Phytozome accession numbers for analysed protein sequences are as follows: *A.*

Table 2. Sizes of the exon/intron for SlPOLα gene. Thirty-one exons were determined by presence within flcDNA using BLASTN search in SGN database. Complete nucleotide sequence was derived from 11 ESTs with 5′ and 3′ RACE fragments.

Exon no.	Coordinate	Size in bp	Intron no.	Size in bp
5′ UTR	–103 . .–1	103		
Exon01	1 . . 1,200	1,200	Intron01	417
Exon02	1,201 . . 1,371	171	Intron02	109
Exon03	1,372 . . 1,454	83	Intron03	126
Exon04	1,455 . . 1,529	75	Intron04	177
Exon05	1,530 . . 1,593	64	Intron05	384
Exon06	1,594 . . 1,687	94	Intron06	86
Exon07	1,688 . . 1,874	187	Intron07	1,672
Exon08	1,875 . . 2,037	163	Intron08	86
Exon09	2,038 . . 2,153	116	Intron09	332
Exon10	2,154 . . 2,334	181	Intron10	94
Exon11	2,335 . . 2,448	114	Intron11	544
Exon12	2,449 . . 2,566	118	Intron12	91
Exon13	2,567 . . 2,891	325	Intron13	278
Exon14	2,892 . . 2,984	93	Intron14	411
Exon15	2,985 . . 3,102	118	Intron15	121
Exon16	3,103 . . 3,177	75	Intron16	87
Exon17	3,178 . . 3,234	57	Intron17	88
Exon18	3,235 . . 3,327	93	Intron18	540
Exon19	3,328 . . 3,450	123	Intron19	79
Exon20	3,451 . . 3,548	98	Intron20	92
Exon21	3,549 . .3,594	46	Intron21	506
Exon22	3,595 . . 3,709	115	Intron22	82
Exon23	3,710 . . 3,791	82	Intron23	97
Exon24	3,792 . . 3,904	113	Intron24	110
Exon25	3,905 . . 3,995	91	Intron25	84
Exon26	3,996 . . 4,072	77	Intron26	307
Exon27	4,073 . . 4,289	217	Intron27	547
Exon28	4,290 . . 4,343	54	Intron28	84
Exon29	4,368 . . 4,462	119	Intron29	99
Exon30	4,463 . . 4,540	78	Intron30	108
Exon31	4,541 . . 4683	143		
3′ UTR	4684 . . 4814	158		

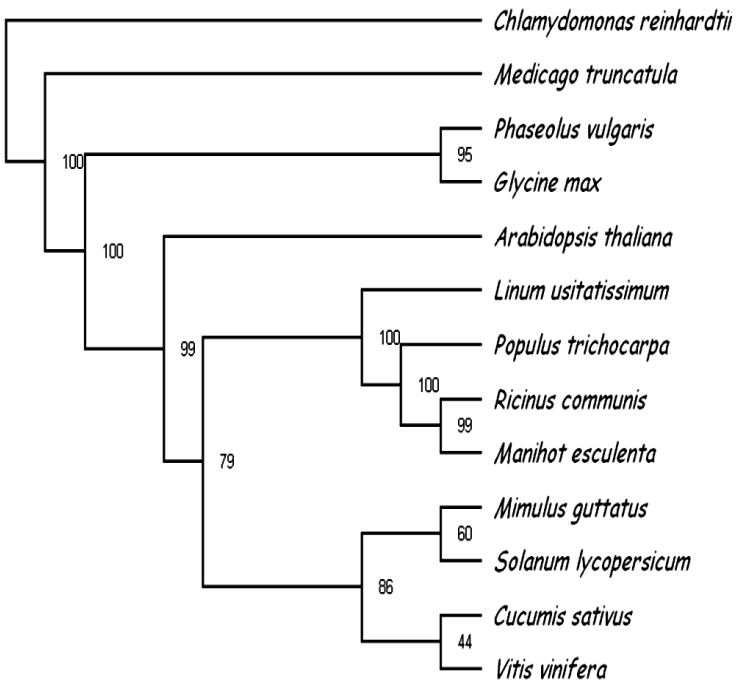


Fig 2. Unrooted phylogenetic tree of 12 hypothetical POLα protein sequences with tomato SlPOLα-predicted protein sequence constructed using maximum parsimony method. Chlamydomonas reinhardtii was used as outgroup. Numbers on nodes are bootstrap percentages supporting a given partitioning.

thaliana (AT5G67100), barrel medick; *Medicago truncatula* (Medtr5g017000), *Chlamydomonas reinhardtii* (Cre04. g214350), cucumbers; *Cucumis sativus* (Cucsa.367180), soya beans; *Glycine max* (Glyma01g39430), cassava; *Manihot esculenta* (cassava4.1_000188 m.g), monkey flower; *Mimulus guttatus* (mgv1a000225m.g), poplar; *Populus trichocarpa* (POPTR_0007s10880), flax; *Linum usitatissimum* (Lus10019338.g), common bean; *Phaseolus vulgaris* (Phvulv091019163 m.g), castor oil; *Ricinus communis* (30115.t000057), grape; *Vitis vinifera* (GSVIVG01022113001). Both phylogenetic trees constructed by maximum parsimony and maximum likelihood methods (Fig 2 and 3) confirmed close relationship between *SlPOLα*-predicted protein and POLα of *M. guttatus*, included by the same clade. *V. vinifera* and *C. sativus* appeared to be closely related to tomato *SlPOLα*-predicted protein and POLα of *M. guttatus* due to the inclusion in the same cluster, although they formed other clade in the maximum parsimony tree and ingroups within subcluster in the maximum likelihood tree.

Expression profiling of SlPOLα gene

ΔΔCT method was used to compare expression of *SlPOLα* gene in tomato tissues using leaf tissue as calibrator. *SlPOLα* was found to be expressed in closed floral buds, flowers, unripe fruits and mature leaves, but the relative expression levels varied. The calculated $2^{-\Delta\Delta Ct}$ values showed the highest transcript abundance (4.47 fold) in closed floral bud tissues, moderate levels in flower tissues (1.74 fold) and the lowest level in unripe fruit tissues (0.09 fold; Fig 4).

Discussion

Since the completion of sequencing of the *A. thaliana* genome in 2000, it has become clear that information about the genome of a particular plant species can have dramatic benefits in promoting plant molecular genetics in general. *A. thaliana* genome sequence provided functional clues of genes from different species (AGI, 2000). The present study focused on the identification and characterization of a tomato gene, *SlPOLα*, which belongs to the PcG gene family, using sequence homology to the orthologous gene in *A. thaliana*. Schubert et al. (2005) discussed the epigenetic role of this gene in flowering and floral organ development in *A. thaliana*. To our knowledge, this is the first gene of DNA polymerases to be sequenced and characterised from Solanaceae. We obtained the complete sequence of flcDNA for the tomato *SlPOLα* gene. The available ESTs sequence at SGN database covered just 18.91% of the total length that was cloned and presented in this study. The exon/intron regions were determined, ORF was predicted and the 5′ and 3′ UTR for flcDNA were determined. POLα encodes the putative catalytic subunit of the DNA polymerase α. This protein is one of the various proteins that participate in DNA replication, repair and recombination (Barrero et al., 2007). The *SlPOLα* predicted protein contains four conserved domains, DNA polα N, POLBcα, DNA polα B exo and zf-DNA Pol. POLα is responsible for the initiation of DNA replication at origins and during lagging strand synthesis of Okazaki fragments (Kunkel and Burgers, 2008). Two of the conserved domains, zf-DNA Pol and DNA polα N, participate directly in the active site required for substrate deoxynucleotide interaction. The zf-DNA Pol domain towards the carboxyl-terminus has the potential to be the DNA interacting domain (Evanics et al., 2003), whereas a DNA polα N, which has potential DNA primase interaction, is predicted to be towards the amino-terminus (Wang et al., 1996). In zf-DNA Pol domain, zinc binding occurs due to the presence of four cysteine residues positioned to bind the metal centre in a tetrahedral coordination geometry. The function of this domain is uncertain. It has been proposed that the zinc finger motif may be an essential part of the DNA binding domain (Evanics et al., 2003). DNA polα B exo domain has 3′ to 5′ exonuclease activity for the catalytic subunit of polymerase α. The 3′-5′ exonuclease domain contains three sequence motifs termed ExoI, ExoII and ExoIII, without the four conserved acidic residues that are crucial for metal binding and catalysis. This explains why in most organisms, that no specific repair role, other than check point control, has been assigned to this enzyme. The exonuclease domain may have a structural role (Wang et al., 2004). POLBcα domain, has been identified as essential for nuclear DNA replication in eukaryotes. It is responsible for the template-directed polymerization of dNTPs onto the growing primer strand of duplex DNA that is usually magnesium dependent (Wang et al., 1997). The four domains, DNA polα N, DNA polα B exo, POLBcα and zf-DNA Pol, were present in POLα genes of all analysed plant species except zf-DNA Pol domain, which was absent in *Medicago truncatula*, according to NCBI domain prediction search. Homology analysis for *SlPOLα* predicted protein against sequenced plant genomes revealed strong homology with POLα protein of *V. vinifera*. Strong homology also was observed with angiosperms species, both monocot and dicot, non-seed vascular plant, *Selaginella moellendorffii* and non-vascular plant, *Physcomitrella patens*. Although DNA replication process is generally conserved in eukaryotes, important differences in the molecular architecture of the DNA replication machine and the function of individual subunits have been reported in various model systems. A recent study indicated that the core DNA replication machinery from plants is more similar to vertebrates than single-celled yeasts (*Saccharomyces cerevisiae*), suggesting that animal models may be more relevant to plant systems. This analysis established that the core DNA replication machinery is highly conserved across plant species and displayed many features in common with other eukaryotes and some characteristics that are unique to plants (Shultz et al., 2007; Perera et al., 2013). Both constructed phylogenetic trees for *SlPOLα* protein sequences phylogeny confirmed close relationships between *SlPOLα* predicted protein and POLα of *M. guttatus* which grouped into the same clade in both constructed trees. Monkey flower (*M. guttatus*) belongs to the order of Lamiales which descends with Solanales order of the clade of Asterids, whereas other accessions in phylogenetic analysis belong to clade of Rosids. Thus, of the analysed accessions, tomato and *M. guttatus* were the most related accessions based on plant evolutionary tree. POLα proteins of grapes and cucumbers also appeared to be closely related to *SlPOLα* protein. Our research focused on the expression levels of *SlPOLα*, in various tomato tissues which reflected its function in cell differentiation and development (Seto et al., 1998; Sanchez et al., 2012). We found that *SlPOLα* transcripts were strongly expressed in proliferating tissues such as closed floral buds, weakly expressed in the unripe fruit tissues, where no cell proliferation was observed, and the expression in flower tissues was between strong and weak. This result indicates that the expression of *SlPOLα* is correlated with cell proliferation and may have special roles in DNA replication and repair in the closed floral buds; thus, it has an important role in maintaining the integrity of the genome during replication and various repair events and also in regulatory events such as cell control and DNA damage checkpoint function (Hübscher et al., 2000). In plants, DNA replication takes place in distinct somatic cell types:

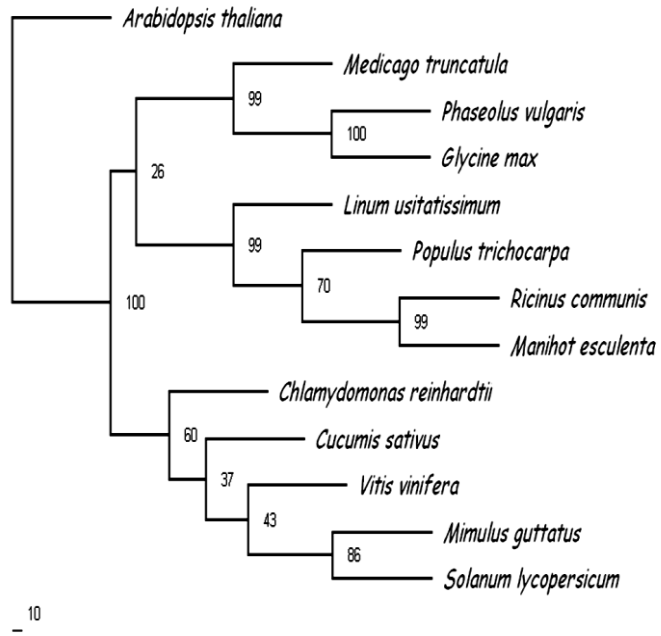


Fig 3. Unrooted phylogenetic tree of 12 hypothetical POLα protein sequences with tomato *SlPOLα*-predicted protein sequence obtained using maximum likelihood method. *Chlamydomonas reinhardtii* was used as outgroup. Numbers on branches are bootstrap percentages supporting a given partitioning. The scale below the tree measures the distance between protein sequences where the lengths of branches are in proportion to the number of amino acid changes that have occurred along the branch.

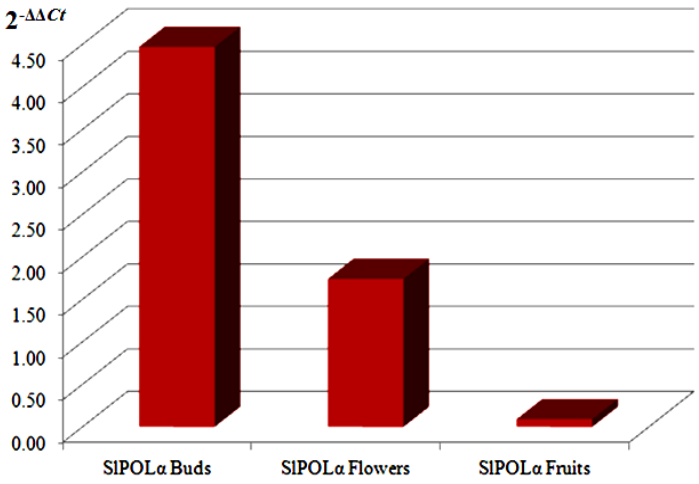


Fig 4. Expression analysis of the tomato PcG gene, *SlPOLα*, in various tomato tissues. The shown $2^{-\Delta\Delta Ct}$ values (*y* axis) are an average of three replicates. Expression level of *SlPOLα* gene was highest in closed floral bud tissue (4.47 fold) and lowest was in unripe fruits (0.09 fold), whereas the expression in flowers tissue was 1.74 fold.

meristematic and differentiated cells. Meristematic DNA replication is followed by cell division that is a prerequisite for plant growth and morphogenesis (Kimura et al., 2004). The promoter of *A. thaliana* POLδ1, one of the DNA polymerases, was also found to be more active in young emerging leaves than in mature leaves (Schuermann et al., 2009). The amounts of the POLδ1 and POLδ2 mRNAs in the rice cells changed rapidly during cell proliferation (Uchiyama et al., 2002). Further studies are needed to identify other PcG genes in tomato and clarify the epigenetic mechanisms and dynamic control of cell fate and differentiation in tomato in distinct developmental stages, which will hopefully aid in understanding the role of cell cycle regulators and chromatin complexes involved in developmental process in the family of Solanaceae.

Experimental procedures

Samples collection

Tomato plants, *Solanum lycopersicum* L., Castlerock variety from Pacific Seed Company, were grown in a greenhouse under standard culture conditions. Various tissues were collected from tomato plants at different stages, quickly frozen in liquid nitrogen and then stored at −80°C until RNA extraction.

Primers designing

A. thaliana POLα protein, available in The *Arabidopsis* Information Resource database (http://www.arabidopsis.org/) under accession No. AT5G67100, was compared with the

tomato proteins available in SGN database. A considerable similarity was found between *A. thaliana* POLα protein and tomato BAC (C02HBa0064B17.3) and EST (SGN-U585459) in SGN database. Predicted *SlPOLα* gene in tomato was constructed on the basis of high similarity found between POLα protein in *A. thaliana* and the mentioned BAC and ESTs using VectorNTI software version 11 (Invitrogen, USA). Forward and reverse primers were designed for *SlPOLα* gene from predicted exon regions and used to amplify the cDNA clones from tomato flower tissues. The primers that amplify cDNAs are listed in table (1).

Reverse transcriptase PCR

cDNA was synthesised using SuperScript® III Reverse Transcriptase (Invitrogen, USA) of total RNA extracted by Chomczynski and Mackey (1995) method, from various tissues with TRI reagent (Sigma, USA), with oligo(dt) primers. Synthesised cDNA was used as a template for PCR to amplify tomato *SlPOLα* cDNAs with designed primers using 2X PCR ready mix (KAPA, USA), 0.4 µM of each primer and 3 µl of cDNA as template with the following program: an initial denaturation at 94°C for 5 min and then 35 cycles of denaturation at 94°C for 1 min, annealing at 52°C for 1 min and extension at 72°C for 2 min. An extension at 72°C for 7 min was performed after all cycles were complete. The cDNA clone was re-amplified using Platinum® *Pfx* DNA polymerase (Invitrogen, USA) which produces PCR products with blunt ends for cloning. The obtained blunt ended fragment was purified by extracting of gel using gel extraction kit (Omega Bio-tek, USA) according to the manufacturer's instruction.

Cloning of SlPOLα cDNA

The cloning vector pBluescript II SK (+) (pBS) from Stratagene (USA) was prepared from bacterial cells using QIAprep spin Miniprep kit (QIAgen, USA) according to the manufacturer's protocol, digested using *Eco*RV restriction enzyme (Promega, USA) to produce blunt ends and then collected from gel. The purified cDNA clones were ligated into the linearised plasmids with T4 DNA ligase (Promega, USA). The recombinant plasmids were transformed into *E. coli* DH5α host cells prepared to be competent by the calcium chloride method (Sambrook and Russel, 2001). Transformation was performed by heat shock protocol (Froger and Hall, 2007). The cells were spread in LB plates (Display Systems Biotech, USA) containing ampicillin 100 µg/ml H2O, X-gal 20% and IPTG 2.4% and incubated at 37°C overnight. Recombinant plasmids were prepared from positive clones to be used as template in sequencing reaction.

5′ and 3′ RACE cDNA amplification

Rapid amplification of cDNA ends (RACE) was performed to determine the 5′ and 3′ non-coding region of the tomato *SlPOLα* gene, using the SMART RACE cDNA amplification kit (Clontech, USA). Total RNA from tomato flower tissues was used as a template to synthesize the first-strand cDNAs for 5′ and 3′ ends following the manufacturer's instruction. The *SlPOLα*-specific primers were designed from sequence information of *SlPOLα* contig, constructed by overlapping sequenced ESTs; *SlPOLα*-specific primer 1 for 5′ ends amplification (5′-CAA TCC TCT TCC TGG CCT TCA TCG CC-3′) and *SlPOLα*-specific primer 2 for 3′ end amplification (5′-GCC CAG AGG AGA GTG AGG GGA AT-3′).

Sequence analysis and phylogenetics

The obtained ESTs and generated 5′ and 3′ RACE fragments were used as templates in sequencing reaction by dideoxynucleotide method (Sanger et al., 1977). Sequencing results were assembled using contigExpress VectorNTI. ORF was predicted and translated using VectorNTI ORF Finder tool. Deduced amino acids were subjected to BLASTP program in http://www.phytozome.net/ database to obtain homologous protein sequences within plant species. Hypothetical POLα protein sequences of plant species were aligned by ClustalW 1.83 multiple sequence alignments (Thompson et al., 1997) using BioEdit software. The phylogenetic analysis was performed using two methods, maximum parsimony (Eck and Dayhoff, 1966; Fitch, 1971) and maximum likelihood (Felsenstein and Churchill, 1996), with the help of the Phylip-3.68 package. *Chlamydomonas reinhardtii* was used as outgroup. Domains were predicted in the deduced *SlPOLα* protein using NCBI CDD (Marchler-Bauer and Bryant, 2004; Marchler-Bauer et al., 2009; 2011).

Quantitative real time PCR

Expression level of the tomato *SlPOLα* gene was detected using qRT-PCR technique. First-strand cDNA was synthesised from total RNA extracted from four tomato tissues; closed floral buds, flowers, fruits and leaves, and it was assayed for *SlPOLα* gene expression level. Gene specific primers (forward: 5′-TGG TAA CCT TTG GGG AA AAA-3′ and reverse: 5′-GTC CAG AAT CCC ACC TTC AA-3′) were used. Actin 2/7 gene was used as a constitutively expressed gene control with forward primer: 5′-GGA CTC TGG TGA TGG TGT TAG-3′ and reverse: 5′-CCG TTC AGC AGT AGT GGT G-3′. qRT-PCR was performed in triplicate. Each reaction (20 µl) consisted of 2X Hot Start SYBR ready mix (QIAgen, USA), forward and reverse primers (0.125 µM each), 3.5 µl of first-strand cDNA and H_2O to 20 µl. qRT-PCR conditions were as follows: 50°C for 2 min, 95°C for 15 min, 40 cycles (95°C for 15 sec, 60°C for 30 sec). Comparative or ΔΔCT method (Livak and Schmittgen, 2001) was used to measure *SlPOLα* gene expression. First, the threshold cycle (Ct) values for each sample were determined at the respective crossing point indicating the cycle at which PCR amplification begins its exponential phase. ΔCt was calculated in each tissue by normalizing all Ct values by subtracting the Ct value for Actin 2/7 gene in each tissue from the Ct value for *SlPOLα* gene in the same tissue. ΔCt values were calibrated using the expression of *SlPOLα* gene in leaf tissues as calibrator, and $2^{-\Delta\Delta Ct}$ was calculated using the equation below. $2^{-\Delta\Delta Ct}$ values were calculated and charts were created using Microsoft Excel application.

$$\Delta Ct = Ct_{target\ gene} - Ct_{Actin2/7}$$

$$\Delta\Delta Ct = \Delta Ct\ tested\ tissue - \Delta Ct\ calibrator\ tissue$$

$$Ratio = 2^{-\Delta\Delta Ct}$$

Conclusion

The molecular machinery that regulates plant flowering and development is conserved throughout the plant kingdom. The POLα gene, which is one of various proteins participate in DNA replication, repair and recombination, plays an important role in epigenetic regulation of flowering and floral organ development. Here we isolated and characterised the tomato *SlPOLα* gene by cloning and sequencing. Phylogenetic relationships with homologous proteins available in GenBank were investigated. Expression profiling

for *SlPOLα* gene reflected its function in cell differentiation and development at different stages of tomato life cycle.

Acknowledgments

Authors are greatly indebted to Dr. Khalid Abu Khabar, director of Biomolecular Research Department in King Faisal Specialist Hospital & Research Centre in Riyadh, and other staff members for giving an opportunity to work in the laboratory and for technical assistance with cDNA sequencing. This work was supported by a grant from Center of Excellence in Biotechnology Research in King Saud University (grant No. CEBR-GSRP 01/2010).

References

AGI (Arabidopsis Genome Initiative) (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814):796-815

Barrero JM, González-Bayón R, del Pozo JC, Ponce MR, Micol JL (2007) *INCURVATA2* encodes the catalytic subunit of DNA Polymerase alpha and interacts with genes involved in chromatin-mediated cellular memory in *Arabidopsis thaliana*. Plant Cell 19:2822-2838

Burgers PM, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, Copeland WC, Friedberg EC, Hanaoka F, Hinkle DC, Lawrence CW, Nakanishi M, Weill JC, Woodgate R (2001) Eukaryotic DNA polymerases: proposal for a revised nomenclature. J Biol Chem 276:43487-34390

Chomczynski P, Mackey K (1995) Short technical report. Modification of the TRIZOL reagent procedure for isolation of RNA from Polysaccharide-and proteoglycan-rich sources. Biotechniques 19(6):942-945

Eck RV, Dayhoff MO (1966) Atlas of Protein Sequence and Structure 1966. National Biomedical Research Foundation, Silver Spring, Maryland

Ekanayake D, Sabatini R (2011) Epigenetic regulation of polymerase II transcription initiation in Trypanosoma cruzi: modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. Eukaryotic Cell 10(11):1465-72

Evanics F, Maurmann L, Yang WW, Bose RN (2003) Nuclear magnetic resonance structures of the zinc finger domain of human DNA polymerase-alpha. Biochim Biophys Acta 1651(1-2):163-171

Felsenstein J, Churchill GA (1996) A hidden Markov model approach to variation among sites in rate of evolution. Mol Biol Evol 13:93-104

Fitch WM (1971) Toward defining the course of evolution: minimum change for a specified tree topology. Syst Zool 20:406-416

Froger A, Hall JE (2007) Transformation of plasmid DNA into *E. coli* using the heat shock method. J Vis Exp 6:253

Garcia-Diaz M, Bebenek K (2008) Multiple functions of DNA polymerases. CRC Crit Rev Plant Sci 26(2):105-122

Garg P, Burgers PM (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol 40:115-128.

Hübscher U, Nasheuer HP, Syvaoja JE (2000) Eukaryotic DNA polymerases, a growing family. Trends Biochem Sci 25:143-147

Hyun Y, Yun H, Park K, Ohr H, Lee O, Kim DH, Sung S, Choi Y (2013) Stable inheritance of epigenetic histone modifications requires INCURVATA2, the catalytic subunit of Arabidopsis DNA polymerase [alpha]. Development 140(1):156-66

Kimura S, Tahira Y, Ishibashi T, Mori Y, Mori T, Hashimoto J, Sakaguchi K (2004) DNA repair in higher plants; photoreactivation is the major DNA repair pathway in non-proliferating cells while excision repair (nucleotide excision repair and base excision repair) is active in proliferating cells. Nucleic Acids Res 32:2760-2767

Kunkel TA, Burgers PM (2008) Dividing the workload at a eukaryotic replication fork. Trends Cell Biol 18:521-527

Lévy N, Oehlmann M, Delalande F, Nasheuer HP, Dorsselaer AV, Schreiber V, de Murcia G, Murcia JM, Maiorano D, Bresson A (2009) XRCC1 interacts with the p58 subunit of DNA Pol alpha-primase and may coordinate DNA repair and replication during S phase. Nucleic Acids Res 37(10):3177-88

Liu J, Ren X, Yin H, Wang Y, Xia R, Wang Y, Gong Z (2010) Mutation in the catalytic subunit of DNA polymerase α influences transcriptional gene silencing and homologous recombination in *Arabidopsis*. Plant J 61:36-45

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression using realtime quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25:402-408

Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32(W):327-331

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39(D):225-229

Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Tasneem A, Thanki N, Yamashita RA, Zhang D, Zhang N, Bryant SH (2009) CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res 37(D):205-210

Moldovan GL, Pfander B, Jentsch S (2007) PCNA, the maestro of thereplication fork. Cell 129:665-679

Muzi-Falconi M, Liberi G, Lucca C, Foiani M (2003) Mechanisms controlling the integrity of replicating chromosomes in budding yeast. Cell Cycle 2:564-567

Nielsen BL, Cupp JD, Brammer J (2010) Mechanisms for maintenance, replication, and repair of the chloroplast genome in plants. J Exp Bot 61(10)2535-2537

Perera RL, Torella R, Klinge S, Kilkenny ML, Maman JD, Pellegrini L (2013) Mechanism for priming DNA synthesis by yeast DNA Polymerase α. eLife 2:e00482

Röbbelen G (1957) UÈ ber Heterophyllie bei *Arabidopsis thaliana* (L.) Heynh. Ber Dtsch Bot Ges 70:39-44

Sambrook J, Russell DW (2001) Molecular cloning: A laboratory manual. 3rd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; pp:132-150

Sanchez Mde L, Costas C, Sequeira-Mendes J, Gutierrez C (2012) Regulating DNA Replication in Plants. Cold Spring Harb Perspect Biol 1;4(12)

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74(12):5463-5467

Schubert D, Clarenz O, Goodrich J (2005) Epigenetic control of plant development by Polycomb-group proteins. Curr Opin Plant Biol 8:553-561

Schuermann D, Fritsch O, Lucht JM, Hohn B (2009) Replication stress leads to genome instabilities in *Arabidopsis* DNA polymerase delta mutants. Plant Cell 21(9):2700-14

Serrano-Cartagena J, Candela H, Robles P, Ponce MR, Perezperez JM, Piqueras P, Micol JL (2000) Genetic analysis of incurvata mutants reveals three independent genetic operations at work in *Arabidopsis* leaf morphogenesis. Genetics 156:1363-1377

Seto H, Hatanaka M, Kimura S, Oshige M, Tsuya Y, Mizushina Y, Sawado T, Aoyagi N, Matsumoto T, Hashimoto J, Sakaguchi K (1998) Purification and characterization of a 100 kDa DNA polymerase from cauliflower inflorescence. Biochem J 332:557-563

SGN (Sol Genomics Network) [http://www.sgn.cornell.edu/index.pl]. Retrieved on 2 May 2013

Shultz RW, Tatineni VM, Hanley-Bowdoin L, Thompson WF (2007) Genome-wide analysis of the core DNA replication machinery in the higher plants Arabidopsis and rice. Plant Physiol 144:1697-1714

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635-641

Thompson JD, Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-4882

Uchiyama Y, Hatanaka M, Kimura S, Ishibashi T, Ueda T, Sakakibara Y, Matsumoto T, Furukawa T, Hashimoto J, Sakaguchi K (2002) Characterization of DNA polymerase delta from a higher plant, rice (*Oryza sativa* L.). Gene 295:19-26

Vaara M, Itkonen H, Hillukkala T, Liu Z, Nasheuer HP, Schaarschmidt D, Pospiech H, Syväoja JE (2012) Segregation of replicative DNA polymerases during S phase: DNA polymerase ε, but not DNA polymerases α/δ, are associated with lamins throughout S phase in human cells. J Biol Chem 28; 287(40):33327-38

Wang CX, Zakharova E, Li J, Joyce CM, Wang J, Konigsberg W (2004) Pre-steady-state kinetics of RB69 DNA polymerase and its exo domain mutants: effect of pH and thiphosphoryl linkages on the 3'-5' exonuclease activity. Biochemistry 43:3853-3861

Wang J, Sattar AKMA, Wang CC, Karam JD, Konigsberg WH, Steitz TA (1997) Crystal structure of a pol_ replication DNA polymerase from bacteriophage RB69. Cell 89:1087-1089

Wang J, Yu P, Lin TC, Konigsberg WH, Steitz TA (1996) Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. Biochemistry 35:8110-8119

Wang ZF, Yang J, Nie ZQ, Wu M (1991) Purification and characterization of a y-like polymerase from *Chlamydomonris reinhardii.* Biochemistry 30:1127-1131

Yeeles JTP, Poli J, Marians KJ, Pasero P (2013) Rescuing Stalled or Damaged Replication Forks. Cold Spring Harb Perspect Biol 5:a012815