

Genome-Wide Analysis of the Core DNA Replication Machinery in the Higher Plants *Arabidopsis* and Rice^{1[W][OA]}

Randall W. Shultz², Vinaya M. Tatineni, Linda Hanley-Bowdoin*, and William F. Thompson

Department of Plant Biology (R.W.S., W.F.T.), Department of Statistical Genetics and Bioinformatics (V.M.T.), and Department of Molecular and Structural Biochemistry (L.H.-B.), North Carolina State University, Raleigh, North Carolina 27695

Core DNA replication proteins mediate the initiation, elongation, and Okazaki fragment maturation functions of DNA replication. Although this 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. We have combined genome-wide bioinformatic analyses of *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*) with published experimental data to provide a comprehensive view of the core DNA replication machinery in plants. Many components identified in this analysis have not been studied previously in plant systems, including the GINS (go ichi ni san) complex (PSF1, PSF2, PSF3, and SLD5), MCM8, MCM9, MCM10, NÖC3, POLA2, POLA3, POLA4, POLD3, POLD4, and RNASEH2. Our results indicate 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. However, we also uncovered some important differences between plants and vertebrate machinery. For example, we did not identify *geminin* or *RNASEH1* genes in plants. Our analyses also indicate that plants may be unique among eukaryotes in that they have multiple copies of numerous core DNA replication genes. This finding raises the question of whether specialized functions have evolved in some cases. This analysis establishes that the core DNA replication machinery is highly conserved across plant species and displays many features in common with other eukaryotes and some characteristics that are unique to plants.

DNA replication depends on the coordinated action of numerous multiprotein complexes. At the simplest level, it requires an initiator to establish the site of replication initiation, a helicase to unwind DNA, a polymerase to synthesize new DNA, and machinery to process the Okazaki fragments generated during discontinuous synthesis. Much is known about the DNA replication machinery in yeast (*Saccharomyces cerevisiae*) and animal model systems, but relatively little is known about the apparatus in plants. To gain insight into plant DNA replication components, we have

combined published experimental information with our own bioinformatic analysis of genomic sequence data to examine the core DNA replication machinery in the model plants *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*).

Figure 1 depicts a model eukaryotic DNA replication fork and illustrates the protein complexes known or suspected to be part of the core DNA replication machine. These complexes mediate the initiation, elongation, and maturation stages of DNA replication and, as such, constitute the core eukaryotic DNA replication machinery. The events leading to the formation of an active DNA replication fork occur in a stepwise fashion, but our understanding of the timing and specific details of how these events unfold in diverse eukaryotes is limited, and there are a growing number of examples of variations between model systems (Bell, 2002; Bell and Dutta, 2002; Kearsley and Cotterill, 2003).

In recent years, there has been increased interest in plant DNA replication and in using plants as models for understanding DNA replication in eukaryotes. A detailed understanding of the core DNA replication machinery in plants will provide researchers with an important tool for understanding what makes plants unique with respect to replicative and developmental capacity and for investigating how plant strategies compare to the mechanisms employed by animals.

¹ This work was supported by the National Science Foundation Plant Genome Research Initiative (grant no. 0421651) and an Integrative Graduate Education and Research Traineeship from the National Science Foundation (to R.W.S.).

² Present address: Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695.

* Corresponding author; e-mail linda_hanley-bowdoin@ncsu.edu; fax 919-513-1209.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Linda Hanley-Bowdoin (linda_hanley-bowdoin@ncsu.edu).

^[W] The online version of this article contains Web-only data.

^[OA] Open Access articles can be viewed online without a subscription. www.plantphysiol.org/cgi/doi/10.1104/pp.107.101105

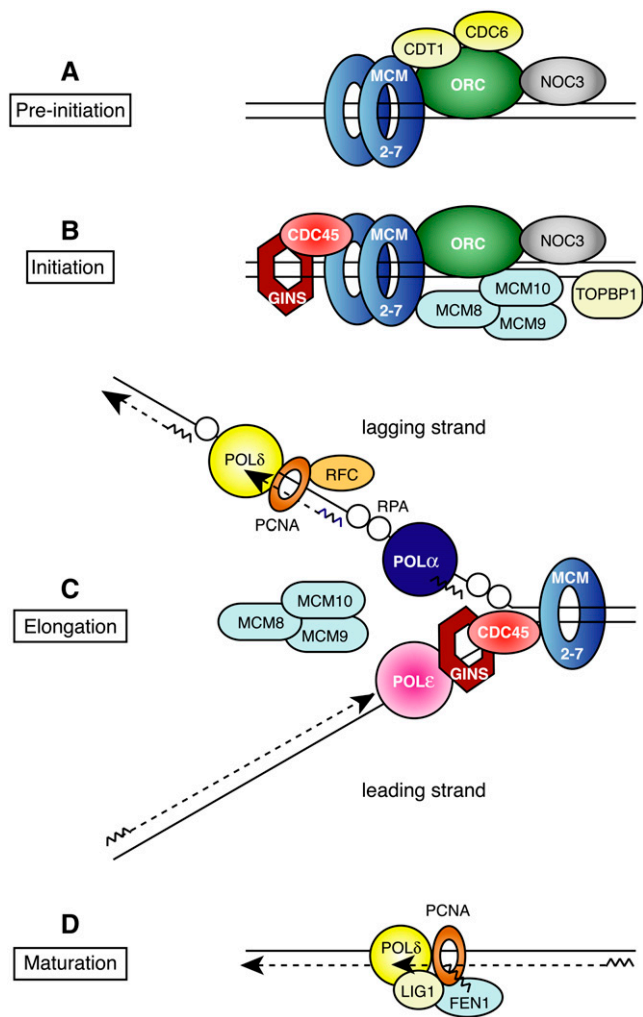


Figure 1. Model depicting the core eukaryotic DNA replication machinery from initiation through Okazaki fragment maturation. A, Components of the preinitiation complex. DNA bound ORC recruits NOC3, CDC6, and CDT1 in early G1. Reiterative loading of 10 to 40 MCM complexes forms a licensed origin. After MCM loading is complete, CDC6 and CDT1 dissociate from the origin. B, At the G1/S transition a subset of licensed origins transition to an initiation complex. The precise order of events is not clear and may vary between systems. CDC45, TOPBP1, and MCM8-10 contribute to GINS complex loading, DNA unwinding, and recruitment of the polymerases. C, Components of the active DNA replication fork. MCM2-7, CDC45, and GINS unwind the duplex DNA. Leading strand synthesis is accomplished primarily by POLE. GINS increases the processivity of POLE. On the lagging strand, RPA stabilizes ssDNA, POLA lays down a short RNA/DNA primer and then is replaced by POLD, which completes the Okazaki fragment. RFC loads PCNA, which increases the processivity of POLD. The precise role of MCM8-10 in this process is not clear. D, The dominant mechanism of Okazaki fragment maturation requires FEN1 to cleave the RNA/DNA flap, resulting in a nick that is sealed by LIG1.

RESULTS AND DISCUSSION

Strategy

To identify the core DNA replication genes in Arabidopsis and rice, we developed an approach

that incorporated experimental data from the literature with homology-based computational gene annotation. First, we assembled a database of yeast and animal proteins that have been determined experimentally to be part of the core eukaryotic DNA replication machinery. The BLAST algorithm was used to search against the translated Arabidopsis genome database at The Arabidopsis Information Resource (TAIR), and sequences with significant similarity were assigned putative annotations based on their functions in yeast and animal systems. The Arabidopsis sequences were then used to identify putative homologs in The Institute for Genomic Research (TIGR) rice genome database. Next, we searched the primary literature and, when available, incorporated experimental results that pertained to plant systems to validate the annotation (relevant plant literature is listed in Table I). In cases where no experimental data from plants could be found, we generated protein sequence alignments from diverse eukaryotes and considered the validity of putative annotations based on the quality of the alignment and the presence of highly conserved domains. Using this strategy, we report the core DNA replication machinery in the dicot Arabidopsis and the monocot rice (Table I). Together, these results established that there is a general conservation of DNA replication machinery in plants.

We encountered numerous instances where the existing gene model resulted in a protein that either lacked highly conserved sequences or contained additional residues compared to other eukaryotic proteins. When available, plant-derived transcripts from GenBank and the TIGR Plant Transcript Assembly databases (TIGR-TA; Childs et al., 2007) were used to guide the prediction of a new gene model. In cases where full-length transcripts could not be identified, we used protein sequence alignments to create gene models that maximized sequence conservation with other eukaryotes (Table I). The predicted coding and resulting protein sequences are provided in Supplemental Text S1.

Preinitiation

One of the first steps toward establishing a functional DNA replication fork is binding of the eukaryotic initiator complex termed the origin recognition complex (ORC) to DNA in late M and early G1 phases of the cell division cycle (Dutta and Bell, 1997; Bryant et al., 2001; Bell, 2002; Bell and Dutta, 2002; DePamphilis, 2003). Next, CDC6 interacts with the origin-bound ORC, which together recruit a CDT1/MCM2-7 complex of proteins. Hydrolysis of ATP by ORC/CDC6 causes the release of CDT1 and structural alteration of the ring-shaped MCM complex, leading to its loading around DNA (Randell et al., 2006; Ranjan and Gossen, 2006; Waga and Zembutsu, 2006). MCM loading is reiterated through the action of a single ORC/CDC6 complex resulting in the recruitment of 10 to 40 MCM complexes at each potential origin (Blow and Dutta,

Table 1. Core DNA replication genes in *Arabidopsis* and rice

Gene Description	Arabidopsis Locus	% Amino Acid Identity/Similarity ^a			Rice Locus	References ^b
		At versus Hs	At versus Sc	At versus Os		
Preinitiation						
Origin recognition						
ORC1	At4g14700	36/56	30/47	59/73	Os06g08790	1–6
	At4g12620	37/56	30/47	58/73		
ORC2	At2g37560	35/56	27/46	60/77	Os03g08640	2–8
ORC3	At5g16690	22/42	ns	39/57	Os10g26280	2–6
ORC4	At2g01120	31/54	25/46	61/81	Os01g49010	2–6
ORC5	At4g29910	26/45	ns	46/63	Os03g55200	2–6
ORC6	At1g26840	32/56	ns	61/76	Os07g43540	3–5
NOC3	At1g79150 ^c	27/46	24/43	49/69	Os06g30320	
Replicative helicase						
MCM2	At1g44900	49/66	51/72	72/83	Os11g29380	4
MCM3	At5g46280	48/65	49/65	67/79	Os05g39850	4, 9–13
MCM4	At2g16440	43/62	44/67	67/79	Os01g36390	4
MCM5	At2g07690	48/65	46/63	73/85	Os02g55410	4
MCM6	At5g44635	45/63	41/59	79/89	Os05g14590	4, 14
MCM7	At4g02060	49/66	46/65	80/89	Os12g37400	4, 15–18
Helicase loading factors						
CDC6	At2g29680	28/48	24/43	51/66	Os01g63710	4, 5, 12, 19–22
	At1g07270	31/54	26/46	59/74		
CDT1	At2g31270	20/37	ns	37/54	Os04g10650	4, 22, 23
	At3g54710	23/50	ns	28/47	Os10g34820	
Geminin	NA				NA	
Initiation						
CDC45	At3g25100	28/49	23/42	65/79	Os12g03130	24
				65/79	Os11g03430	
MCM8	At3g09660	44/62	NA	60/74	Os05g38850 ^c	14
MCM9	At2g14050 ^c	48/66	NA	70/82	Os06g11500	14
MCM10	At2g20980	33/54	19/39	44/62	Os09g36820	
TOPBP1	At1g77320	23/38	ns	38/52	Os11g08660	39–41
	At4g02110	26/46	ns	27/44	Os07g49010	
				29/44	Os03g19190	
GINS complex						
PSF1	At1g80190	29/50	29/50	56/76	Os05g37980 ^c	
PSF2	At3g12530	42/64	27/54	66/79	Os01g14610	
PSF3	At1g19080 ^c	29/45	22/38	62/77	Os08g17294	
	At3g55490	29/45	22/38	62/77		
SLD5	At5g49010	31/52	23/44	58/75	Os05g05150	
Elongation						
DNA polymerase α						
POLA1	At5g67100	34/52	32/50	56/71	Os01g64820 ^c	25, 26
POLA2	At1g67630	37/56	24/45	51/70	Os12g13950 ^c	
POLA3	At1g67320	41/59	35/56	67/83	Os07g22400	
POLA4	At5g41880	47/64	39/57	62/77	Os05g29010 ^c	
Elongation						
DNA polymerase δ						
POLD1	At5g63960	54/69	49/66	80/88	Os11g08330	26–28
POLD2	At2g42120	45/62	32/50	71/85	Os03g03650	27
POLD3	At1g78650	23/38	ns	37/53	Os01g10690	
POLD4	At1g09815	50/70	NA	38/50	Os08g43080	
				41/61	Os09g34850	
DNA polymerase ϵ						
POLE1	At1g08260 ^c	42/60	39/58	66/79	Os02g30800 ^c	26, 37, 38
	At2g27120	41/59	38/57	63/76		
POLE2	At5g22110	33/52	28/46	65/79	Os05g06840	37, 38
				32/41	Os08g36330	
POLE3	Gene family				Gene family	37
POLE4	Gene family				Gene family	37

(Table continues on following page.)

Table I. (Continued from previous page.)

Gene Description	Arabidopsis Locus	% Amino Acid Identity/Similarity ^a			Rice Locus	References ^b
		At versus Hs	At versus Sc	At versus Os		
POLD clamp						
PCNA	At1g07370	65/85	40/66	88/96	Os02g56130	29–34
	At2g29570	66/86	39/66	89/95		
PCNA loading complex						
RFC1	At5g22010	32/48	30/48	59/72	Os11g36390	35
RFC2	At1g63160	67/81	62/78	85/93	Os04g48060	35
RFC3	At5g27740	53/75	41/62	78/91	Os03g57870	35
RFC4	At1g21690	56/74	51/68	82/90	Os12g07720	35
RFC5	At1g77470	57/76	49/69	73/83	Os02g53500	35, 36
ssDNA binding						
RPA1	At2g06510	35/55	30/53	59/75	Os02g53680	42–45
	At5g08020	30/51	33/53	65/78	Os03g11540	
	At5g45400	36/55	36/56	40/58	Os05g02040	
	At5g61000	32/51	30/50			
	At4g19130	37/58	33/55			
RPA2	At2g24490	28/49	26/44	37/59	Os02g58220	42–46
	At3g02920	29/45	22/40	36/57	Os02g42230	
					Os06g47830	
RPA3	At3g52630	ns	ns	52/74	Os01g14980	44, 45
	At4g18590			51/75		
Maturation						
FEN1	At5g26680	55/74	48/69	84/92	Os05g46270	29, 47, 48
				67/81	Os03g61820	
DNA2	At1g08840	33/51	30/51	54/69	Os04g49860	
RNASE H1	NA				NA	
RNASEH2A	At2g25100	51/70	40/54	69/83	Os11g05570	
LIG1	At1g08130	49/69	41/62	63/77	Os10g34750	49–52
	At1g49250	46/66	44/65	62/80		

^aAt, Hs, Os, and Sc are Arabidopsis, human, rice, and yeast, respectively. ^bReferences: 1, Kimura et al. (2000a); 2, Witmer et al. (2003); 3, Collinge et al. (2004); 4, Masuda et al. (2004); 5, Diaz-Trivino et al. (2005); 6, Mori et al. (2005); 7, Gavin et al. (1995); 8, Li et al. (2005); 9, Sabelli et al. (1996); 10, Sabelli et al. (1999); 11, Stevens et al. (2002); 12, Dambrauskas et al. (2003); 13, Taliencio et al. (2005); 14, Dresselhaus et al. (2006); 15, Springer et al. (1995); 16, Springer et al. (2000); 17, Bastida and Puigdomenech (2002); 18, Holding and Springer (2002); 19, Castellano et al. (2001); 20, de Jager et al. (2001); 21, Ramos et al. (2001); 22, Castellano et al. (2004); 23, Raynaud et al. (2005); 24, Stevens et al. (2004); 25, Yokoi et al. (1997); 26, Burgers et al. (2001); 27, Uchiyama et al. (2002); 28, Garcia et al. (2006); 29, Kimura et al. (2001); 30, Egelkrout et al. (2002); 31, Kosugi and Ohashi (2002); 32, Toueille et al. (2002); 33, Castillo et al. (2003); 34, Raynaud et al. (2006); 35, Furukawa et al. (2003); 36, Furukawa et al. (2001); 37, Jenik et al. (2005); 38, Ronceret et al. (2005); 39, He and Mascarenhas (1998); 40, Yang and Sheila (2002); 41, Grelon et al. (2003); 42, Ishibashi et al. (2001); 43, Marwedel et al. (2003); 44, Ishibashi et al. (2005); 45, Ishibashi et al. (2006); 46, Xia et al. (2006); 47, Kimura et al. (2003); 48, Kimura et al. (2000b); 49, Taylor et al. (1998); 50, Sunderland et al. (2004); 51, Bonatto et al. (2005); 52, Sunderland et al. (2006). ^cNew gene models were predicted for these loci. NA, Gene was not identified; ns, no significant similarity was found.

2005). The resulting protein/DNA assembly consisting of ORC1-6/CDC6/MCM2-7 is termed the prereplication complex (pre-RC), and sites containing this complex are considered licensed with the potential to serve as origins of replication (Fig. 1A).

All six ORC genes have been identified in Arabidopsis (Gavin et al., 1995; Collinge et al., 2004; Masuda et al., 2004), and genes encoding ORC1 to 5 have been reported for rice (Kimura et al., 2000a; Li et al., 2005; Mori et al., 2005) and maize (*Zea mays*; Witmer et al., 2003). The Arabidopsis ORC proteins show 22% to 37% amino acid identity with human ORC subunits (Table I). Consistent with a role in DNA replication, Arabidopsis ORC transcripts have been shown to be abundant in proliferating tissues such as root tips, young leaves, and flower buds, and their expression induced upon cell cycle reentry following Suc starvation of cultured suspension cells (Masuda et al., 2004;

Diaz-Trivino et al., 2005). Interestingly, *ZmORC3* (Witmer et al., 2003) and *AtORC5-6* (Diaz-Trivino et al., 2005) transcripts are also abundant in postmitotic tissues, suggesting that plant ORC subunits may have additional functions in mature tissues.

Two analyses of the ORC complex in rice (Mori et al., 2005) and maize (Witmer et al., 2003) failed to identify a candidate ORC6 gene. The authors of each of these studies suggested that ORC6 is poorly conserved in plants. However, our query of the TIGR-TA database using AtORC6 yielded strong hits from a diverse array of plant species. Alignment of five dicot, two monocot, and two conifer ORC6 proteins showed significant sequence similarity within plants and between plants and other eukaryotes (Fig. 2A). This alignment supports the conclusion that like other ORC subunits, ORC6 is conserved in plants. Interestingly, our analysis indicates that the ORC6 C terminus is conserved

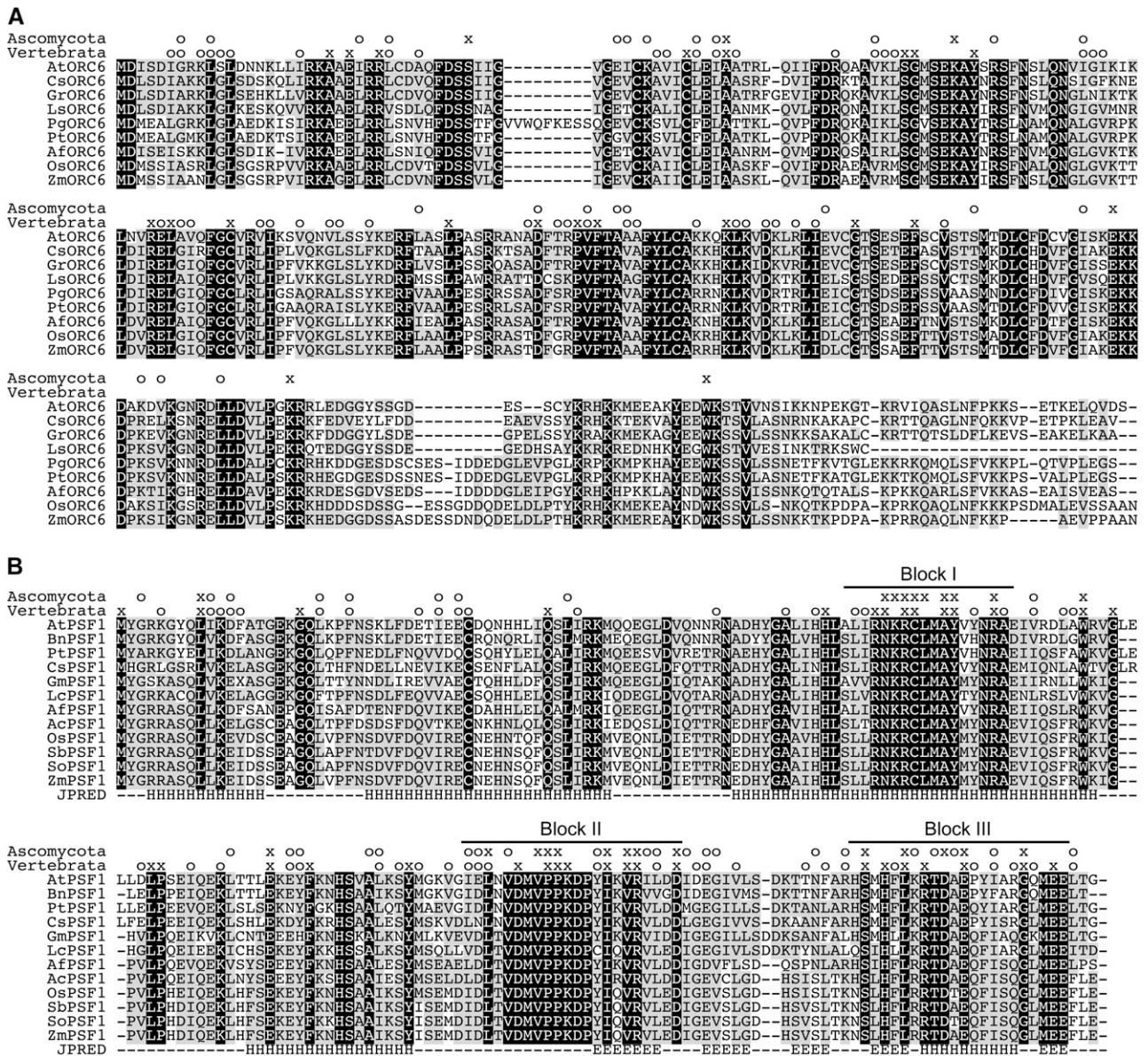


Figure 2. Multiple sequence alignments of plant ORC6 and GINS complex proteins. A, ORC6. B, PSF1. C, PSF2. D, PSF3. E, SLD5. For all sections, protein sequences from the indicated plant species were aligned using the Clustal W algorithm. Black shading indicates identical residues in all aligned sequences. Gray shading denotes residues with similar chemical properties that are conserved in >50% of sequences aligned. Similar chemical properties of amino acid residues were defined as follows: DE, Acidic; AGILV, aliphatic; NQ, amide; FWY, aromatic; RHK, basic; ST, hydroxyl; CM, sulphur. Plant sequences were also aligned with proteins from other eukaryotes. Ascomycota includes sequences from yeast and *Schizosaccharomyces pombe*. Vertebrata includes *Homo sapiens*, *Danio rerio*, *X. laevis*, and *Gallus gallus*. An x indicates residues that are identical in all sequences from plants and the specified group. An o denotes residues with similar chemical properties that are conserved in >50% of plant sequences and 100% of sequences from the specified group. The plant species are designated as Ac, *Ananas comosus*; Af, *Aquilegia formosa* × *Aquilegia pubescens*; At, *Arabidopsis*; Bn, *Brassica napus*; Br, *B. rapa*; Cs, *Citrus sinensis*; Ee, *Euphorbia esula*; Ga, *Gossypium arboreum*; Gm, *G. max*; Gr, *Gossypium raimondii*; Ha, *Helianthus annuus*; Ht, *Helianthus petiolaris*; Hv, *Hordeum vulgare*; In, *Ipomoea nil*; Lc, *Lotus corniculatus*; Ls, *Lactuca serriola*; Lv, *Lactuca virosa*; Mt, *Medicago truncatula*; Nt, *N. tabacum*; Os, *O. sativa*; Pg, *Picea glauca*; Ps, *Picea sitchensis*; Pt, *Populus trichocarpa*; Pta, *Pinus taeda*; Sa, *Senecio aethnensis*; Sb, *Sorghum bicolor*; Sch, *Solanum chacoense*; So, *Saccharum officinarum*; St, *Solanum tuberosum*; Ta, *Triticum aestivum*; Zm, *Z. mays*.

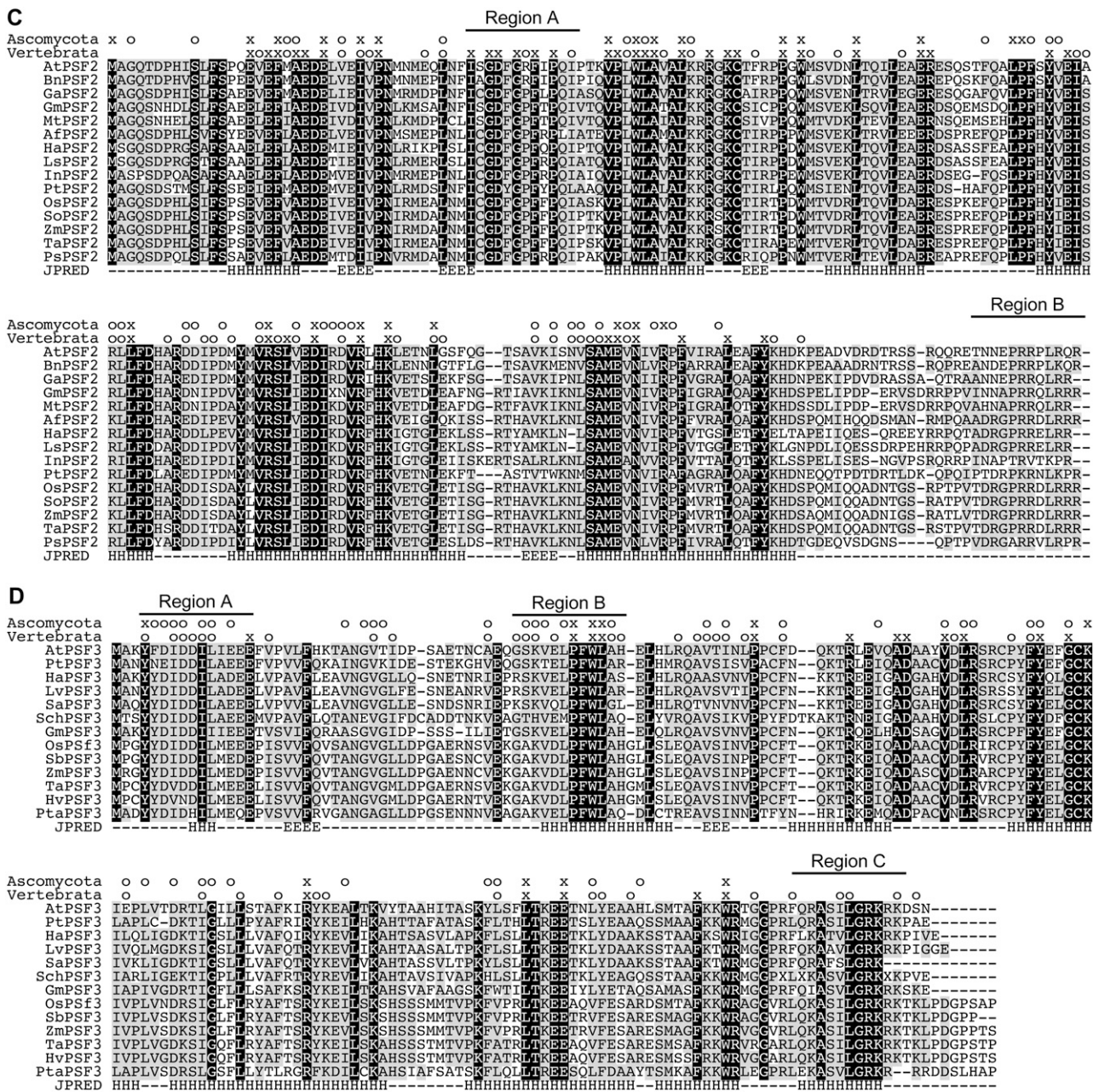


Figure 2. (Continued.)

among plants but differs from other eukaryotes, indicative of a plant-specific function.

It has been reported that Arabidopsis has two *CDC6* (Ramos et al., 2001) and two *CDT1* (Castellano et al., 2004) genes. We identified one candidate *CDC6* and two candidate *CDT1* homologs in rice (Table I). OsCDC6 shares 51% and 59% amino acid identity with AtCDC6A (At2g29680) and AtCDC6B (At1g07270), respectively. The CDT1 proteins are more divergent. AtCDT1A (At2g31270) and AtCDT1B (At3g54710) are 37% identical while the two rice CDT1 proteins are 30% identical. Between Arabidopsis and rice, the amino acid

identity ranges from 28% to 37% (Table I). Given that CDC6 and CDT1 have similar functions in all eukaryotes studied to date, it is likely that these proteins also act similarly in plants. However, the divergence between copies raises the possibility of additional activities.

The six-subunit MCM complex (MCM2-7) represents the putative eukaryotic replicative helicase (Forsburg, 2004; Masai et al., 2005; Maiorano et al., 2006), and genes encoding one copy of each subunit have been identified in Arabidopsis (Springer et al., 1995; Stevens et al., 2002; Masuda et al., 2004; Dresselhaus et al., 2006). MCM3 (Sabelli et al., 1996; Sabelli et al., 1999) and

MCM8 and MCM9 genes in rice, suggesting that these proteins are generally conserved in plants. The TIGR gene model for *OsMCM8* predicts a protein of 482 amino acids, which is considerably shorter than other MCM8 proteins and lacks several highly conserved domains. However, a sequence encoding the missing domains is present in the rice genome, supporting our prediction of a new gene model for *OsMCM8* (Supplemental Text S1).

Our examination of the Arabidopsis and poplar (*Populus* spp.) MCM9 gene sequences suggested that they also may be alternatively spliced or misrepresented by transcripts in the databases, and new gene models that maximize protein sequence conservation were predicted (Supplemental Text S1). Arabidopsis *MCM8* and *MCM9* are expressed at low levels (Schmid et al., 2005). This could explain why the functional transcripts have not been cloned, but a directed effort toward identifying primary and alternatively spliced *MCM8* and *MCM9* transcripts is needed to determine the relevance of our predicted gene models.

Like other MCM family members, the central region of plant MCM8 and MCM9 proteins contain Walker A and B NTP-binding domains, a putative zinc finger, and an Arg finger motif (Supplemental Figs. S1 and S2). In both plants and animals, the MCM8 and MCM9 proteins contain a classic GKS sequence in the Walker A motif compared to the deviant A/SKS sequence found in MCM2-7 (Maiorano et al., 2005, 2006). Arabidopsis, poplar, and rice lack the first approximately 60 amino acids of animal MCM8 proteins and are also missing a QVLTKDLEXAAXLQXDE motif found in human, chicken, frog, and sea urchin homologs (Supplemental Fig. S1, region C). Additional differences between plant and animal MCM8 proteins are indicated (Supplemental Fig. S1, regions A and B and D-I). In animals, MCM9 proteins are the largest of the MCM family members due to a long, poorly conserved C-terminal domain. Plants do not have this C-terminal extension, raising the possibility that MCM9 functions differently in plants and animals (Supplemental Fig. S2). It is also noteworthy that while all MCM2-8 proteins contain the IDEFDKM Walker B sequence, only the IDEF is conserved in MCM9 proteins (Walker et al., 1982; Neuwald et al., 1999).

MCM10, which is conserved from yeast to humans, does not contain the sequence features that define the rest of the MCM family. However, it is an essential part of the core DNA replication machinery and has been implicated in a variety of DNA replication processes, including loading and stabilizing DNA polymerase α (POLA; Ricke and Bielinsky, 2004), recruitment of CDC45 (Wohlschlegel et al., 2002; Sawyer et al., 2004), and as a component of the replisome progression complex (Gambus et al., 2006; Pacek et al., 2006). Although we found no published information regarding MCM10 proteins in plants, we identified putative MCM10 homologs in Arabidopsis, rice, maize, and the western columbine (*Aquilegia formosa*). Eukaryotic MCM10 proteins are not highly conserved (Supplemental Fig. S3).

Arabidopsis and rice MCM10 proteins show 44% amino acid identity, while Arabidopsis and human MCM10 proteins display 33% identity within the aligned region. In budding yeast, MCM10 interacts with itself through a CCCH-type zinc finger motif to form a large homocomplex that is required for DNA replication (Cook et al., 2003). The CCCH zinc finger is conserved in plant and animal MCM10 proteins, suggesting that zinc binding and homomultimerization are shared properties of MCM10 proteins (Supplemental Fig. S3).

CDC45 is essential for both the initiation and elongation stages of DNA replication (Bell and Dutta, 2002; Pollok et al., 2003; Pacek and Walter, 2004). It assembles onto the origin in late G1 after the MCM2-7 complex and concurrent with the onset of initiation (Zou and Stillman, 1998). CDC45 is required for POLA complex loading (Mimura and Takisawa, 1998) and is a component of large complexes containing MCMs and GINS. These observations have led to the suggestion that CDC45 serves as an anchor coupling POLA to the replication fork via the replisome complex (Gambus et al., 2006; Moyer et al., 2006; Pacek et al., 2006). Published results indicate that Arabidopsis *CDC45* is expressed in proliferating tissues and transcripts are most abundant at the G1/S transition (Stevens et al., 2004), consistent with a role at this cell cycle stage. Interestingly, AtCDC45 has also been implicated in meiosis—a role not yet reported for any other eukaryotes (Stevens et al., 2004). In rice, we identified two putative *CDC45* genes located on chromosomes 11 and 12. These two proteins differ only at four positions, indicative of a very recent duplication event or strong selective pressure.

The GINS complex, which consists of four proteins, PSF1, PSF2, PSF3, and SLD5, was identified recently as a critical part of the initiation process. GINS is essential for the establishment and maintenance of a functional DNA replication fork (Kanemaki et al., 2003; Kubota et al., 2003; Takayama et al., 2003; Gambus et al., 2006). The GINS proteins copurify as a tightly associated heterotetrameric complex with a ring-like structure that resembles the DNA polymerase δ (POLD) processivity factor, proliferating cell nuclear antigen (PCNA), in electron micrographs (Kubota et al., 2003). GINS has been shown to bind weakly to DNA polymerase ϵ (POLE) and specifically stimulate DNA synthesis by POLE in vitro, leading to the suggestion that GINS functions as a POLE processivity factor analogous to the function of PCNA (Seki et al., 2006). However, there is also evidence that GINS is a core component of the eukaryotic DNA replication fork helicase. GINS interacts stably with the MCM2-7 complex and CDC45, and the GINS/MCM/CDC45 supercomplex functions as a helicase in vitro (Gambus et al., 2006; Moyer et al., 2006). The precise role that GINS plays at the DNA replication fork is not yet clear, but GINS interactions with POLE and the CDC45/MCM helicase complex place it between these two complexes on the leading strand (Fig. 1C).

GINS complex proteins have been identified in a broad array of eukaryotes based on sequence similarity (Kubota et al., 2003), and experimental evidence from yeast (cited above), fly (Moyer et al., 2006), frog (Kubota et al., 2003; Pacek et al., 2006), and mouse (Ueno et al., 2005) suggests that the complex is functionally conserved in eukaryotes. Consistent with this suggestion, we identified a complete GINS complex from two dicots, *Arabidopsis* and soybean (*Glycine max*), and two monocots, rice and maize. In *Arabidopsis*, we found two loci (see Table I) that encode for nearly identical copies of the putative PSF3 protein. We also identified transcripts representing one or several of the GINS complex proteins from a diverse array of additional plant species, demonstrating that GINS is conserved broadly in plants. We performed phylogenetic analysis of the GINS complex from plants, animals, and yeasts, demonstrating that the proteins cluster primarily by subunit and secondarily by taxonomy (Fig. 3). In all cases, the vertebrates (human, chicken, zebrafish, and frog) form a tight cluster with good bootstrap support. *Drosophila* sequences cluster loosely with this group, and *C. elegans* sequences are more divergent. Plant sequences also form a highly supported cluster, with monocot and dicot sequences tending to separate into subgroups. For PSF1, PSF2, and SLD5, the plant sequences cluster with animals, while for PSF3, plants and yeast cluster, although the bootstrap support for these divisions is generally low (Fig. 3).

GINS complex proteins are highly conserved with respect to amino acid sequence (Supplemental Table S1). PSF2, which shares 66% identity between *Arabidopsis* and rice and 42% identity between *Arabidopsis* and human, is the most highly conserved GINS complex subunit. Amino acid sequence length and pI are also conserved features of eukaryotic GINS proteins (Supplemental Table S2). PSF1 shows the least size variability with an average of 199 amino acids and a SD of 4.4 between plants, vertebrates, and yeasts (data not shown). In budding yeast, PSF3 and SLD5 proteins are longer than plant and animal sequences due to an approximately 25 amino acid N-terminal extension and several small internal insertions. The predicted pIs of GINS complex proteins typically range from 5 to 7 (Supplemental Table S2). PSF1 in chicken, which has a predicted pI of 8.8, and PSF3 in *Arabidopsis* and rice, which have predicted pIs of 8.3 and 9.2, respectively, are notable exceptions.

To identify conserved and unique features of plant GINS complex proteins, we generated sequence alignments from diverse plant species and compared them to yeast and vertebrate GINS proteins (Fig. 2, B–E). These alignments indicated that eukaryotic PSF1 proteins are similar along their entire length, but show the highest degree of sequence conservation in the central and C-terminal regions (Fig. 2B). Two blocks of identical residues, RNKRCLMAY (block I) and VDMVPPKDP (block II), and a highly conserved motif in the C terminus (block III), are apparent in plant sequences. These domains are also highly conserved in yeast and

animals, suggesting that they are critical for PSF1 function. Supporting this conclusion, it has been shown that mutation of a conserved Arg residue in block I of budding yeast [NK(R-to-G) CL] results in cell growth arrest and morphology consistent with a DNA replication defect (Takayama et al., 2003). PSF1 is predicted to adopt primarily a helical conformation with a short elongated region (β sheet) near its C terminus. Given the structural constraints conferred by Pro residues, the conserved double Pro in block II may serve an important role in defining the structural properties of PSF1.

PSF2 proteins from yeast, animals, and plants contain tracts of identical and conserved residues spread across the length of the protein (Fig. 2C). In contrast to the rest of the protein, the C terminus stands out as being poorly conserved. Plant proteins have an additional 15 to 20 amino acids at the C terminus including a short, conserved motif, PRRxLRR (region B). Plant and vertebrate PSF2 proteins also contain a conserved sequence (region A) that is lacking in budding and fission yeasts. Alignment of PSF3 proteins reveals two conserved features in the N-terminal region (Fig. 2D, region A and region B), a high degree of similarity through the central portion of the protein, and an LGRKR motif at the C-terminal end (region C). This motif does not align with yeast and vertebrate proteins. However, vertebrate sequences have a conserved NYXKRK motif in this region, suggesting that positive charge may be important at the C terminus (data not shown).

Our analysis indicated that SLD5 proteins contain two prominent blocks of highly conserved amino acids (Fig. 2E, blocks I and II). Except for a short conserved region at the extreme C terminus (region A), the N and C termini of SLD5 are divergent. We used the COILS algorithm (Lupas et al., 1991) to predict that the plant SLD5 proteins adopt a coiled-coil structure between blocks I and II (Fig. 2E, COILS track). Coiled-coil domains, which are common in transcription factors (Leu zipper motif), SNARE complexes, and spindle-pole-body components, are thought to interact primarily with other coiled-coil domains (Lupas, 1996, 1997; Martin et al., 2004; Rose et al., 2004). We did not detect coiled-coil domains in the other GINS complex proteins, and it would be interesting to ask if SLD5 acts as a homodimer or facilitates interactions between the GINS complex and other coiled-coil proteins.

Our analysis suggested that the initiation components have largely been conserved in plants, and supports the hypothesis that similar mechanisms govern the transition from pre-RC to active replication fork in plants and animals.

Elongation Complex

Initiation of active replication at the G1/S transition requires the assembly of additional proteins including DNA polymerases and Okazaki fragment maturation factors to form a complete replication factory (Fig. 1C;

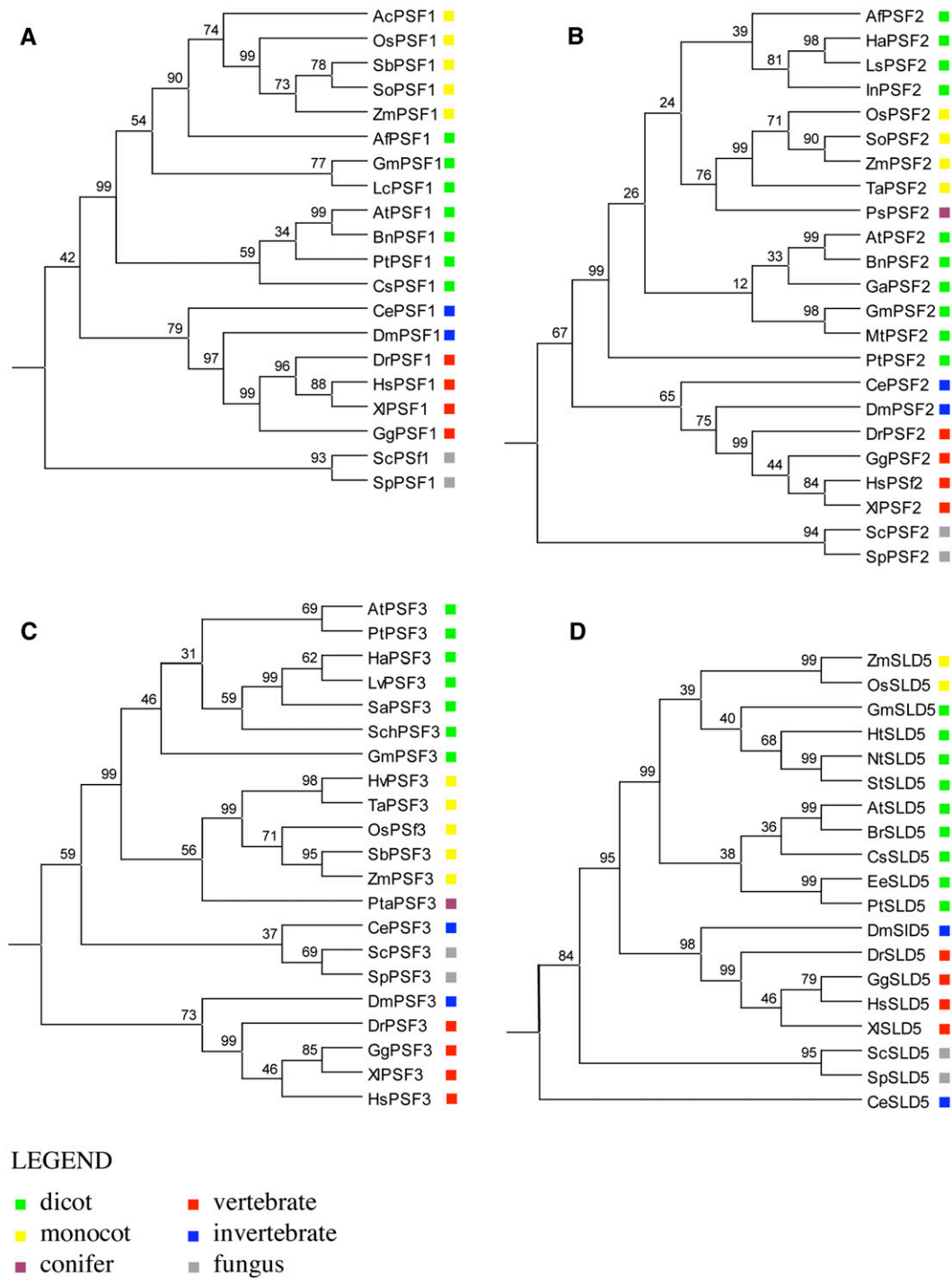


Figure 3. Phylogenetic analysis of eukaryotic GINS complex proteins. Protein sequences were aligned with ClustalW using the Gonnet scoring matrix in MEGA. A single tree containing all of the sequences was constructed by the neighbor-joining method, and was split manually into four subtrees (A–D) for visualization. Bootstrap values from 5,000 iterations are shown at each node. Species abbreviations are the same as in Figure 2.

Waga and Stillman, 1998; Bell and Dutta, 2002). Three eukaryotic DNA polymerase complexes have been implicated in DNA replication—POLA, POLD, and POLE (Burgers et al., 2001; Garg and Burgers, 2005; Johnson and O’Donnell, 2005).

The POLA complex includes a catalytic subunit (POLA1), two primase subunits (POLA3 and POLA4), and POLA2, which is thought to tether the complex to the replication fork (Frick and Richardson, 2001). Protein complexes containing polymerase and primase

activity have been purified from a variety of plant systems (Coello and Vazquez-Ramos, 1995; Garcia et al., 2002), demonstrating that a POLA-like function exists in plants. However, sequence homology has been investigated only for the POLA1 subunit in rice (Yokoi et al., 1997). Rice POLA1 was originally reported to be shorter than other eukaryotic POLA1 homologs, due to a truncated N terminus (Yokoi et al., 1997). However, publication of the rice genomic sequence (Sasaki et al., 2002) allowed us to predict a full-length OsPOLA1 (GenBank accession O48653). We identified all four putative *POLA* subunits in Arabidopsis and the remaining three subunits in rice. We predicted new gene models for rice POLA2 and POLA4 subunits, resulting in better conservation to other eukaryotes (Supplemental Text S1).

Seven protein sequence features have been established as conserved in all eukaryotic DNA polymerase catalytic subunits (Spicer et al., 1988; Wong et al., 1988), and an additional five regions are conserved in POLA1 proteins (Miyazawa et al., 1993). We found that these defined regions are conserved in the Arabidopsis and rice POLA1 proteins. Although the sequence features of POLA2 to 4 are less well characterized, Arabidopsis POLA2, POLA3, and POLA4 align with 37%, 41%, and 47% identity to their corresponding human proteins, respectively. According to our analyses, the majority of sequence features that are conserved between yeast and human are present in the corresponding Arabidopsis and rice proteins, supporting the hypothesis of conserved function. One notable exception is an YRRRLF motif of unknown function located at the N terminus of yeast and animal POLA4 proteins but absent in Arabidopsis and rice POLA4. We conclude that POLA is a four subunit complex in Arabidopsis and rice.

POLD is known to function as a heterotetramer in fission yeast and animals (POLD1–4), but only three subunits have been identified in budding yeast (POLD1–3; Johnson and O'Donnell, 2005). The largest subunit (POLD1) contains the polymerase and exonuclease activity, while the other subunits are involved in complex stabilization and interactions with PCNA. In rice, the *POLD1* and *POLD2* genes have been shown to be expressed primarily in proliferating tissues and induced upon regrowth following Suc starvation in cell culture (Uchiyama et al., 2002). Interestingly, only *POLD1* transcripts were detected in mature leaves and induced upon UV irradiation treatment (Uchiyama et al., 2002), leading to the suggestion that POLD1 has specific DNA repair functions independent of POLD2. An alternate explanation is that POLD2 protein activity does not correlate with its transcription. POLD1 and PCNA protein levels correlate in maize, further supporting the conclusion that plant POLD functions in DNA replication (Garcia et al., 2006).

A previously published alignment of Arabidopsis, soybean, rice, and maize sequences indicated that plant POLD1 proteins contain most of the conserved domains present in other eukaryotic POLD1 proteins, but the dicot sequences lacked two C-terminal zinc finger mo-

tifs (Garcia et al., 2006). This result was surprising because these motifs are highly conserved in other eukaryotes and are critical for interaction with POLD2. The AtPOLD1 and GmPOLD1 sequences were derived from transcripts in GenBank (accessions NP_201201 and AAC18443, respectively). We identified a second Arabidopsis transcript (accession ABA41487) that utilizes an alternative splice donor site (GT) at the end of exon 27 (data not shown) that results in a frameshift, which restores conservation in the C terminus, including the zinc finger motifs. Similarly, we searched the TIGR-TA database and identified an assembly (TA66266_3847) for *GmPOLD1* that encodes a protein with these zinc finger motifs. It is not known if these transcripts represent bona fide alternative splicing events or artifacts, but it is clear that dicots produce transcripts specifying proteins that contain these important zinc finger domains.

Our analysis indicated that Arabidopsis and rice POLD2 proteins also contain all of the sequence features conserved between animals and yeasts (data not shown). In humans, a region of hydrophobic residues (MRPFL) near the N terminus of POLD2 has been shown to mediate interaction with PCNA (Lu et al., 2002). We found a similar sequence (MRT/NLL) in Arabidopsis and rice POL2D proteins at this position. We also identified a conserved PCNA-binding motif in Arabidopsis and rice POLD3, suggesting that multiple POLD subunits mediate PCNA interactions in plants as has been reported for human (Ducoux et al., 2001). The N termini of plant and animal POLD3 also show significant sequence similarity, while their central regions are more divergent (data not shown).

The POLD4 subunit is not essential for growth in fission yeast (Reynolds et al., 1998), but increases the processivity of both fission yeast and human POLD1 to 3 complexes in vitro (Zou and Stillman, 2000; Li et al., 2006). Human POLD4 also stabilizes the POLD complex and participates in interactions with PCNA (Li et al., 2006). A POLD4 homolog has not been identified in budding yeast, and it has been uncertain whether the plant POLD complex consists of three or four subunits. We identified a single Arabidopsis *POLD4* and two putative *POLD4* genes in rice (Table I), indicating that POLD consists of at least four subunits in plants.

Four POLE subunits have been identified in vertebrates (*POLE1–4*), budding yeast (*POL2*, *DPB2*, *DPB4*, and *DPB3*), and other eukaryotes (Johnson and O'Donnell, 2005; the vertebrate nomenclature is used here). In Arabidopsis, two genes encoding the catalytic subunit (*POLE1A*, At1g08260 and *POLE1B*, At2g27120) and a single gene encoding the second largest subunit (*POLE2*) have been reported (Ronceret et al., 2005). Both AtPOLE1A and AtPOLE1B were shown to contain the conserved domains of other eukaryotic homologs (Ronceret et al., 2005). Mutations in either *AtPOLE1A* or *AtPOLE2* result in DNA replication defects (Jenik et al., 2005; Ronceret et al., 2005), while *AtPOLE1B* mutants do not exhibit visible phenotypic effects (Ronceret et al., 2005). These results indicate that AtPOLE1B is not required when AtPOLE1A is present.

However, *AtPOLE1A/B* double mutants arrest earlier than single *AtPOLE1A* mutants, suggesting some functional overlap in vivo (Jenik et al., 2005). We identified a single *POLE1* gene in rice that specifies a protein that is 66% and 63% identical to *AtPOLE1A* and *AtPOLE1B*, respectively (Table I). Like the *AtPOLE1* proteins, *OsPOLE1* contains all of the functional domains conserved in other eukaryotes.

We identified two candidate *POLE2* genes in the rice genome (Table I). The gene models for these loci (*OsPOLE2A*, LOC_Os05g06840.1 and *OsPOLE2B*, LOC_Os08g36330.1) predict proteins that are considerably shorter than other eukaryotic *POLE2* proteins and are missing several highly conserved domains. Because these gene models were derived solely by computational methods, we searched the TIGR-TA database for biological transcripts. We identified a single transcript assembly (TA60386_4530) representing *OsPOLE2*. This transcript aligns to the *OsPOLE2A* locus but has a different intron/exon structure than the computational gene model. Translation of this sequence results in a protein containing the domains missing from the computational model and likely specifies a functional *OsPOLE2* protein. We were unable to detect any biological transcripts for the *OsPOLE2B* gene, and stop codons in the genomic sequence prevent the prediction of a full-length coding sequence that would contain all of the conserved domains. As a consequence, we suggest that *OsPOLE2B* is a pseudogene.

Our search for *POLE3* and *POLE4* homologs in *Arabidopsis* and rice returned a family of histone-fold proteins, which includes the core histones as well as a large number of CCAAT box-binding transcription factors. Histone-fold proteins share a conserved three-dimensional conformation but are only distantly related in primary sequence (Arents and Moudrianakis, 1995; Marino-Ramirez et al., 2006). We were unable to specify *POLE3* and *POLE4* homologs based on sequence similarity. However, the *POLE* small subunits have been identified and functionally verified in a multitude of other eukaryotes (Garg and Burgers, 2005), and it is likely that a directed experimental approach will identify plant homologs.

PCNA, the processivity clamp for POLD, is highly conserved among eukaryotes and is structurally related to the bacterial β -sliding clamp (Maga and Hubscher, 2003; Naryzhny et al., 2005). PCNA homologs have been described in numerous plants (Toueilie et al., 2002) and will not be described in detail here.

Replication factor C (RFC) is a five-subunit clamp loader complex that uses ATP to load PCNA onto DNA (Ellison and Stillman, 1998; Venclovas et al., 2002; Majka and Burgers, 2004). Variable nomenclature has been used to describe the subunits in yeasts and animals with HsRFC1-5, corresponding to budding yeast ScRFC1, ScRFC4, ScRFC5, ScRFC2, and ScRFC3, respectively (we have adopted the human nomenclature here). All five RFC subunits have been identified in both *Arabidopsis* and rice, and contain conserved sequence motifs characteristic of other eukaryotic

RFCs (Furukawa et al., 2003). In rice, the RFC subunits are expressed in proliferating tissues and transcript levels respond to chemical treatments that arrest cell cycle progression (Furukawa et al., 2003). The sequence conservation and experimental data indicate that like other eukaryotes, plants utilize a five-subunit RFC complex to load PCNA.

Budding yeast DPB11 is essential for the recruitment of *POLE* and *POLA* complexes to origins (Masumoto et al., 2000; Bell and Dutta, 2002). Fission yeast RAD4, human TOPBP1, *Drosophila* MUS101, and *Xenopus* CUT5 proteins are all thought to be functional homologs of ScDPB11 (Hashimoto and Takisawa, 2003; Kim et al., 2005). Amino acid sequence conservation between the yeast and animal proteins is limited, but all contain copies of the breast cancer 1 gene (*BRCA1*) C-terminal domain (BRCT). Four BRCT domains are present in ScDPB11, while HsTOPBP1 and DmMUS101 contain eight and seven copies of the BRCT domain, respectively (Makiniemi et al., 2001; Kim et al., 2005). We searched for an *Arabidopsis* DPB11/TOPBP1 homolog and found two BRCT domain-containing proteins, meiosis 1 (*MEI1*, At1g77320), and At4g02110 (Table I). *AtMEI1* contains five BRCT domains and plays an essential role in DNA repair during meiosis (Grelon et al., 2003). At4g02110 contains only two BRCT domains (Pfam data not shown) but does show significant similarity to other TOPBP1 proteins (Table I). Based on sequence similarity, it is not possible to determine if one, both, or neither of these proteins are true homologs of TOPBP1. *AtMEI1* mutants do not display any visible mitotic phenotypes, suggesting that plants do not require a TOPBP1 homolog or that another protein performs this function (Grelon et al., 2003). In rice, we identified one protein (*Os11g08660*) that is most similar to *AtMEI1* (38% identity) and two proteins with similarity to At4g02110 (Table I). Considering that TOPBP1/DPB11 function is conserved from yeast to human, a directed effort to identify a functional homolog in plants would be worthwhile.

Okazaki Fragment Maturation

Semidiscontinuous replication requires machinery to process the Okazaki fragments generated during lagging strand synthesis (Fig. 1D). As POLD/PCNA extends the Okazaki fragment, it encounters the 5' end of the downstream replication product and displaces it from the template strand, generating a flap (Maga et al., 2001). The flap is then cleaved to generate a nick, which is ligated to form the intact nascent strand (Kao and Bambara, 2003). The dominant mechanism of flap cleavage requires Flap Endonuclease1 (FEN1) to cleave the 5' flap structure and DNA Ligase1 (LIG1) to seal the nick (Kao et al., 2004; Rossi and Bambara, 2006). Both FEN1 and LIG1 homologs have been described in plants (Table I). Other models of Okazaki fragment maturation require DNA2 and/or RNASE H in addition to FEN1 for efficient processing of the flap structure (Qiu et al., 1999; Masuda-Sasa et al., 2006;

Stewart et al., 2006). We identified a single putative *DNA2* gene in both the Arabidopsis and rice genomes (Table I). AtDNA2 is 33% identical to HsDNA2 and 54% identical to the putative OsDNA2 protein. We were unable to identify a RNASEH1 homolog in any plant species, but both Arabidopsis and rice encode a RNASE H2 homolog. Perhaps RNASE H2 is the dominant RNASE H enzyme in plants.

Multiple Copy Core DNA Replication Genes

Plants may be unique among eukaryotes in that they have multiple copies of numerous core DNA replication genes (Table II). This raises the question of whether some copies have evolved specialized functions. Indeed, this has been demonstrated for the single-stranded DNA (ssDNA)-binding RPA complex in rice (see references cited below).

RPA functions as a heterotrimeric complex to stabilize ssDNA during replication, repair, and transcription (Iftode et al., 1999; Fanning et al., 2006; Zou et al., 2006). The largest subunit (RPA1) contains the primary ssDNA-binding activity, while the two smaller subunits (RPA2 and RPA3) stabilize the complex and mediate interactions with replication and repair machinery (Zou et al., 2006). Aside from humans, which have two *RPA2* homologs (Keshav et al., 1995), plants are the only eukaryotes that possess multiple copies of an *RPA* gene. Rice has three copies each of *RPA1* and *RPA2*, and a single *RPA3* gene (Table II). Arabidopsis has five putative *RPA1* genes and two copies each of *RPA2* and *RPA3* (Table II). Three distinct RPA complexes, termed A type, B type, and C type have been characterized in rice (Ishibashi et al., 2005, 2006). The A-type complex localizes to the chloroplast and, thus, is expected to function primarily in organelle processes. The B- and C-type

complexes both localize to the nuclear compartment, suggesting that they act in nuclear processes, but the precise function of each complex remains to be determined. It is not known whether analogous complexes occur in Arabidopsis, but mutation of Arabidopsis *RPA1A* (At2g06510) is lethal while mutation of a different *RPA1* copy (At5g08020) results in viable but mutagen-sensitive plants (Ishibashi et al., 2005).

Three members of the pre-RC (ORC1, CDC6, and CDT1) are duplicated in Arabidopsis. Both the *AtORC1A* (At4g14700) and *AtORC1B* (At4g12620) promoters have been shown to contain consensus E2F-binding sites (Masuda et al., 2004), but only *AtORC1A* transcripts were found to be elevated in tissues that undergo extra endocycles (Diaz-Trivino et al., 2005), suggesting distinct functions with respect to mitotic and endocycling cells. Similarly, *AtCDC6A* (At2g29680) and *AtCDC6B* (At1g07270) have distinct expression profiles (Masuda et al., 2004). The only other eukaryote known to have multiple *CDC6* genes is *Xenopus laevis*, where *XICDC6A* and *XICDC6B* have distinct N-terminal regulatory motifs and different expression patterns in the developing frog embryo (Tikhmyanova and Coleman, 2003). *XICDC6A* acts prior to the mid-blastula transition, after which *XICDC6B* becomes the dominant protein (Tikhmyanova and Coleman, 2003). The midblastula transition coincides with extensive chromatin remodeling, activation of zygotic transcription, and a clear shift in the regulation of origin usage. It has been suggested that *XICDC6A* and *XICDC6B* play key roles in determining origin usage during development. An understanding of whether the two *AtCDC6* genes are functionally distinct awaits further analysis. We identified only single copies of ORC1 and CDC6 in rice, indicating that any specialized functions that may have evolved in Arabidopsis are not required for plant development. Rice contains two nearly identical *CDC45* proteins while Arabidopsis only has one, but both Arabidopsis and rice contain two *CDT1* genes. A directed effort to understand whether these multicopy DNA replication genes in plants have functional significance would be worthwhile.

PCNA and *POLE1* genes have also been duplicated in Arabidopsis (Table II). Interestingly, we observed that the *AtCDC6B* (At1g07270), *AtPCNA1* (At1g07370), and *AtPOLE1A* (At1g08260) genes are located in close physical proximity on chromosome 1, and the other copies of these genes, *AtCDC6A* (At2g29680), *AtPCNA2* (At2g29570), and *AtPOLE1B* (At2g27120), are clustered on chromosome 2 (data not shown). A published analysis of segmental duplications in the Arabidopsis genome indicated that this region was duplicated in a polyploidy event approximately 24 to 40 million years ago, prior to the *Arabidopsis/Brassica rapa* split (Blanc et al., 2003). We compared the sequences and found that the levels of nucleotide conservation between each copy of *AtCDC6*, *AtPCNA*, and *AtPOLE1* are similar at 82%, 85%, and 85%, respectively. However, at the amino acid level, the two copies of *AtPCNA* and *AtPOLE1* are identical at 96%

Table II. Copy numbers of core DNA replication genes

Gene	Hs	Sc	At	Os
ORC1	1	1	2	1
CDC6	1	1	2	1
CDT1	1	1	2	2
CDC45	1	1	1	2
TOPBP1 ^a	1	1	2	3
PSF3	1	1	2	1
POLD4	1	0	1	2
POLE1	1	1	2	1
POLE2	1	1	1	2
POLE3	1	1	GF	GF
POLE4	1	1	GF	GF
PCNA	1	1	2	1
RPA1	1	1	5	3
RPA2 ^b	2	1	2	3
RPA3	1	1	2	1
FEN1	1	1	1	2
LIG1	1	1	3	1

^aTOPBP1 aliases include DPB11, CUT5, MUS101, and RAD4. ^bThe second copy of RPA2 in human is named RPA4. GF, Putative gene family. Species abbreviations: Hs, Human; Sc, yeast; At, Arabidopsis; Os, rice.

and 90% of residues, respectively, while the two copies of AtCDC6 only show 72% identity. This situation provides an excellent opportunity for a more detailed analysis of the different evolutionary pressures on these genes.

CONCLUSION

Through genome-wide bioinformatic analysis of Arabidopsis and rice and a comprehensive review of the extant literature, we report that the core DNA replication machinery of animals and yeasts is conserved in plants. Generalization to other plant species is supported by the inclusion of both a monocot and a dicot in this analysis. Identification of components that have not previously been reported from any plant, including the GINS complex, MCM10, NOC3, POLA2 to 4, POLD3 and 4, and RNASEH2, will open up new avenues of research. Additionally, extension of many previously reported components to include both monocot and dicot proteins should facilitate comparison within plants and between plants and other eukaryotes.

We did not detect candidate homologs for RNASE H1 or geminin, leading us to suggest that these proteins are not conserved in plants. Geminin is a critical regulator of CDT1 activity in some metazoans (Ballabeni et al., 2004; Lutzmann et al., 2006) but has not been identified in yeast. It would be interesting to determine how CDT1 activity is regulated in plants. An intriguing possibility is that one or more members of the large CDK and cyclin gene families encoded by plant genomes (Vandepoele et al., 2002) function as CDT1 regulators. Very little is known about RNASE H enzymes in plants, and the lack of an obvious RNASE H1 homolog in plants suggests that plants may be different from other eukaryotic organisms in this regard. However, we did identify an RNASE H2 gene in both Arabidopsis and rice, and an effort to define the functional capacity of this RNASE H2 enzyme in plants is needed.

Our analysis also indicated that core DNA replication proteins from plants are more similar to human than to budding yeast proteins (Table I). This observation holds true for the majority of proteins listed in Table I and is most striking for ORC3, ORC5, ORC6, CDT1, TOPBP1, and POLD3, for which no significant alignment between Arabidopsis and budding yeast proteins could be generated. The parallels in the core DNA replication machinery between plants and animals are not limited to amino acid similarity. For example, budding yeast have only three POLD subunits, while animals have four, and there are four strong POLD candidates in both Arabidopsis and rice. The Arabidopsis and rice genomes also encode putative MCM8 and MCM9 proteins, which are part of the replication initiation complex in animals but not in yeasts. In summary, the available data suggest that animal systems may be more relevant models than budding yeast for plant DNA replication.

We also found numerous components of the core DNA replication machine that are encoded by small gene families in both Arabidopsis and rice. With few exceptions, this situation seems to be unique to plants. There is some evidence of functional divergence between copies, and it would be interesting to investigate the evolutionary relationships and functional roles of these genes in greater detail. There are many examples of overlapping functions in DNA replication and repair machinery (Kimura and Sakaguchi, 2006), and it is attractive to hypothesize that some members of DNA replication gene families have specialized roles related to repair. Similarly, plant cells often undergo endoreduplication as part of normal development, and there is some evidence suggesting that members of DNA replication gene families have specialized roles in this process. A systematic approach to determining the function of each gene copy in such families would provide an important contribution to the fields of DNA replication and plant developmental biology.

MATERIALS AND METHODS

Assembly of Yeast and Animal Reference Sequences

Core DNA replication genes were defined primarily by review of the literature. The STRING database (Snel et al., 2000; von Mering et al., 2007) was also used to supplement known protein interaction networks. Yeast (*Saccharomyces cerevisiae*) sequences were downloaded from the National Center for Biotechnology Information (NCBI) and the Saccharomyces Genome Database. Animal sequences were downloaded from the NCBI RefSeq database when possible, and the NCBI nonredundant database otherwise. The NCBI HomoloGene database was useful for identifying homologs in various organisms. The BLAST programs (BLASTP and TBLASTN) were used to query sequence databases. Vector NTI (Invitrogen) was used to manage and analyze sequences in house. GenBank accession numbers are provided in Supplemental File S1.

Plant Sequences

To identify core DNA replication proteins in Arabidopsis (*Arabidopsis thaliana*), yeast and animal proteins were used to query (BLASTP) the Arabidopsis genome databases at TAIR, TIGR, and NCBI. Sequences with significant similarity were downloaded into our Vector NTI database and putative annotations were assigned based on the function of yeast and animal proteins. Next, the NCBI PubMed and ISI Web of Science literature databases were queried for publications relevant to each protein in a plant system. Pertinent information was used to manually curate the putative annotations we assigned based on sequence similarity. This curated list of Arabidopsis proteins was then used to query (BLASTP and TBLASTN) the rice (*Oryza sativa* L. sp. *japonica*) genome database managed by TIGR. Sequences with significant similarity to Arabidopsis proteins were downloaded into our Vector NTI database and annotated accordingly. Transcripts representing core DNA replication proteins from all other plants in this analysis were downloaded from either the TIGR plant transcript assembly (Childs et al., 2007) or NCBI databases. Protein translations were performed using the Vector NTI software package.

Protein Sequence and Phylogenetic Analyses

Percent amino acid identity and similarity values reported in Table I, Supplemental Table S1, and in the text were generated by pairwise BLAST on the NCBI Web site using the default parameters. The percentages reported correspond to regions of the proteins that were aligned by the algorithms. In

cases where we revised gene models (Supplemental Text S1), we used those revised models in making the alignments. Multiple sequence alignments were performed using the Clustal W algorithm within the Vector NTI suite and the BLOSUM62 scoring matrix. Similar amino acids were defined based on the chemical properties of residue side chains as follows: acidic, Asp (D), and Glu (E); aliphatic, Gly (G), Ala (A), Val (V), Leu (L), and Ile (I); amide, Asn (N) and Gln (Q); aromatic, Phe (F), Tyr (Y), and Trp (W); basic, His (H), Lys (K), and Arg (R); hydroxyl, Ser (S) and Thr (T); and sulfur containing, Met (M) and Cys (C). Conserved sequence features were annotated by searching a variety of databases including Pfam, SMART, and the NCBI conserved domain database. Phylogenetic trees were constructed using the neighbor-joining method (Saitou and Nei, 1987) within the Molecular Evolutionary Genetics Analysis software package (MEGA3; Kumar et al., 2004). Alignment for tree construction was done using ClustalW with the Gonnet scoring matrix. For bootstrap tests, the *P*-distance method and 5,000 iterations were selected.

Accession numbers and locus identifiers for sequences used in these analyses are provided in Supplemental File S1.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Multiple sequence alignment of MCM8 proteins.

Supplemental Figure S2. Multiple sequence alignment of MCM9 proteins.

Supplemental Figure S3. Multiple sequence alignment of MCM10 proteins.

Supplemental Table S1. Pairwise BLAST of GINS complex proteins.

Supplemental Table S2. Properties of GINS complex proteins.

Supplemental Text S1. Text file containing nucleotide coding sequences and amino acid sequences in FASTA format for new gene models predicted in these analyses.

Supplemental File S1. Microsoft Excel file containing accession numbers for sequences used in these analyses.

ACKNOWLEDGMENT

We would like to thank Dr. George Allen (Department of Horticultural Sciences, North Carolina State University) for critical review of the manuscript.

Received April 16, 2007; accepted May 29, 2007; published June 7, 2007.

LITERATURE CITED

- Arents G, Moudrianakis EN** (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci USA* **92**: 11170–11174
- Ballabeni A, Melixetian M, Zamponi R, Masiero L, Marinoni F, Helin K** (2004) Human geminin promotes pre-RC formation and DNA replication by stabilizing CDT1 in mitosis. *EMBO J* **23**: 3122–3132
- Bastida M, Puigdomenech P** (2002) Specific expression of ZmPRL, the maize homolog of MCM7, during early embryogenesis. *Plant Sci* **162**: 97–106
- Bell SP** (2002) The origin recognition complex: from simple origins to complex functions. *Genes Dev* **16**: 659–672
- Bell SP, Dutta A** (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* **71**: 333–374
- Blanc G, Hokamp K, Wolfe KH** (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Res* **13**: 137–144
- Blanton HL, Radford SJ, McMahan S, Kearney HM, Ibrahim JG, Sekelsky J** (2005) REC, Drosophila MCM8, drives formation of meiotic crossovers. *PLoS Genet* **1**: e40
- Blow JJ, Dutta A** (2005) Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol* **6**: 476–486
- Bonatto D, Brendel M, Henriques JAP** (2005) A new group of plant-specific ATP-dependent DNA ligases identified by protein phylogeny, hydrophobic cluster analysis and 3-dimensional modelling. *Funct Plant Biol* **32**: 161–174
- Bryant JA, Moore K, Aves SJ** (2001) Origins and complexes: the initiation of DNA replication. *J Exp Bot* **52**: 193–202
- Burgers PM, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, Copeland WC, Friedberg EC, Hanaoka F, Hinkle DC, et al** (2001) Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J Biol Chem* **276**: 43487–43490
- Castellano MM, Boniotti MB, Caro E, Schnittger A, Gutierrez C** (2004) DNA replication licensing affects cell proliferation or endoreplication in a cell type-specific manner. *Plant Cell* **16**: 2380–2393
- Castellano MM, del Pozo JC, Ramirez-Parra E, Brown S, Gutierrez C** (2001) Expression and stability of Arabidopsis CDC6 are associated with endoreplication. *Plant Cell* **13**: 2671–2686
- Castillo AG, Collinet D, Deret S, Kashoggi A, Bejarano ER** (2003) Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology* **312**: 381–394
- Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP** (2007) The TIGR plant transcript assemblies database. *Nucleic Acids Res* **35**: D846–851
- Coello P, Vazquez-Ramos JM** (1995) Maize DNA polymerase 2 is a phosphoprotein with increasing activity during germination. *Eur J Biochem* **231**: 99–103
- Collinge MA, Spillane C, Kohler C, Gheyselinck J, Grossniklaus U** (2004) Genetic interaction of an origin recognition complex subunit and the Polycomb group gene MEDEA during seed development. *Plant Cell* **16**: 1035–1046
- Cook CR, Kung G, Peterson FC, Volkman BF, Lei M** (2003) A novel zinc finger is required for Mcm10 homocomplex assembly. *J Biol Chem* **278**: 36051–36058
- Dambrauskas G, Aves SJ, Bryant JA, Francis D, Rogers HJ** (2003) Genes encoding two essential DNA replication activation proteins, Cdc6 and Mcm3, exhibit very different patterns of expression in the tobacco BY-2 cell cycle. *J Exp Bot* **54**: 699–706
- de Jager SM, Menges M, Bauer UM, Murray JAH** (2001) Arabidopsis E2F1 binds a sequence present in the promoter of S-phase-regulated gene AtCDC6 and is a member of a multigene family with differential activities. *Plant Mol Biol* **47**: 555–568
- DePamphilis ML** (2003) The “ORC cycle”: a novel pathway for regulating eukaryotic DNA replication. *Gene* **310**: 1–15
- DePamphilis ML, Blow JJ, Ghosh S, Saha T, Noguchi K, Vassilev A** (2006) Regulating the licensing of DNA replication origins in metazoa. *Curr Opin Cell Biol* **18**: 231–239
- Diaz-Trivino S, del Mar Castellano M, de la Paz Sanchez M, Ramirez-Parra E, Desvoves B, Gutierrez C** (2005) The genes encoding Arabidopsis ORC subunits are E2F targets and the two ORC1 genes are differentially expressed in proliferating and endoreplicating cells. *Nucleic Acids Res* **33**: 5404–5414
- Dresselhaus T, Srilunchang KO, Leljak-Levanic D, Schreiber DN, Garg P** (2006) The fertilization induced DNA replication factor MCM6 of maize shuttles between cytoplasm and nucleus, and is essential for plant growth and development. *Plant Physiol* **140**: 512–527
- Ducoux M, Urbach S, Baldacci G, Hubscher U, Koundrioukoff S, Christensen J, Hughes P** (2001) Mediation of proliferating cell nuclear antigen (PCNA)-dependent DNA replication through a conserved p21(Cip1)-like PCNA-binding motif present in the third subunit of human DNA polymerase delta. *J Biol Chem* **276**: 49258–49266
- Dutta A, Bell SP** (1997) Initiation of DNA replication in eukaryotic cells. *Annu Rev Cell Dev Biol* **13**: 293–332
- Egelkrout EM, Mariconti L, Settlage SB, Cella R, Robertson D, Hanley-Bowdoin L** (2002) Two E2F elements regulate the proliferating cell nuclear antigen promoter differently during leaf development. *Plant Cell* **14**: 3225–3236
- Ellison V, Stillman B** (1998) Reconstitution of recombinant human replication factor C (RFC) and identification of an RFC subcomplex possessing DNA-dependent ATPase activity. *J Biol Chem* **273**: 5979–5987
- Fanning E, Klimovich V, Nager AR** (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* **34**: 4126–4137

- Forsburg SL (2004) Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev* **68**: 109–131
- Frick DN, Richardson CC (2001) DNA primases. *Annu Rev Biochem* **70**: 39–80
- Furukawa T, Ishibashi T, Kimura S, Tanaka H, Hashimoto J, Sakaguchi K (2003) Characterization of all the subunits of replication factor C from a higher plant, rice (*Oryza sativa* L.), and their relation to development. *Plant Mol Biol* **53**: 15–25
- Furukawa T, Kimura S, Ishibashi T, Hashimoto J, Sakaguchi K (2001) A plant homologue of 36 kDa subunit of replication factor C: molecular cloning and characterization. *Plant Sci* **161**: 99–106
- Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* **8**: 358–366
- Garcia E, Laquel P, Castroviejo M, Plasencia J, Vazquez-Ramos JM (2002) Maize replicative alpha-type DNA polymerase: separation of polymerase and primase activities and recognition of primase subunits. *Physiol Plant* **114**: 533–539
- Garcia E, Quiroz F, Uchiyama Y, Sakaguchi K, Vazquez-Ramos JM (2006) Expression of a maize delta-type DNA polymerase during seed germination. *Physiol Plant* **127**: 268–276
- Garg P, Burgers PMJ (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit Rev Biochem Mol Biol* **40**: 115–128
- Gavin KA, Hidaka M, Stillman B (1995) Conserved initiator proteins in eukaryotes. *Science* **270**: 1667–1671
- Gozuacik D, Chami M, Lagorce D, Favre J, Murakami Y, Poch O, Biermann E, Knippers R, Brechot C, Paterlini-Brechot P (2003) Identification and functional characterization of a new member of the human Mcm protein family: hMcm8. *Nucleic Acids Res* **31**: 570–579
- Grelon M, Gendrot G, Vezon D, Pelletier G (2003) The Arabidopsis MEI1 gene encodes a protein with five BRCT domains that is involved in meiosis-specific DNA repair events independent of SPO11-induced DSBs. *Plant J* **35**: 465–475
- Hashimoto Y, Takisawa H (2003) Xenopus Cut5 is essential for a CDK-dependent process in the initiation of DNA replication. *EMBO J* **22**: 2526–2535
- He C, Mascarenhas JP (1998) MEI1, an Arabidopsis gene required for male meiosis: isolation and characterization. *Sex Plant Reprod* **11**: 199–207
- Holding DR, Springer PS (2002) The Arabidopsis gene PROLIFERA is required for proper cytokinesis during seed development. *Planta* **214**: 373–382
- Iftode C, Daniely Y, Borowiec JA (1999) Replication protein A (RPA): the eukaryotic SSB. *Crit Rev Biochem Mol Biol* **34**: 141–180
- Ishibashi T, Kimura S, Furukawa T, Hatanaka M, Hashimoto J, Sakaguchi K (2001) Two types of replication protein A 70 kDa subunit in rice, *Oryza sativa*: molecular cloning, characterization, and cellular & tissue distribution. *Gene* **272**: 335–343
- Ishibashi T, Kimura S, Sakaguchi K (2006) A higher plant has three different types of RPA heterotrimeric complex. *J Biochem (Tokyo)* **139**: 99–104
- Ishibashi T, Koga A, Yamamoto T, Uchiyama Y, Mori Y, Hashimoto J, Kimura S, Sakaguchi K (2005) Two types of replication protein A in seed plants. *FEBS J* **272**: 3270–3281
- Jenik PD, Jurkuta RE, Barton MK (2005) Interactions between the cell cycle and embryonic patterning in Arabidopsis uncovered by a mutation in DNA polymerase epsilon. *Plant Cell* **17**: 3362–3377
- Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. *Annu Rev Biochem* **74**: 283–315
- Kanemaki M, Sanchez-Diaz A, Gambus A, Labib K (2003) Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. *Nature* **423**: 720–724
- Kao HI, Bambara RA (2003) The protein components and mechanism of eukaryotic Okazaki fragment maturation. *Crit Rev Biochem Mol Biol* **38**: 433–452
- Kao HI, Veeraraghavan J, Polaczek P, Campbell JL, Bambara RA (2004) On the roles of *Saccharomyces cerevisiae* Dna2p and Flap endonuclease 1 in Okazaki fragment processing. *J Biol Chem* **279**: 15014–15024
- Kearsey SE, Cotterill S (2003) Enigmatic variations: divergent modes of regulating eukaryotic DNA replication. *Mol Cell* **12**: 1067–1075
- Keshav KF, Chen C, Dutta A (1995) Rpa4, a homolog of the 34-kilodalton subunit of the replication protein A complex. *Mol Cell Biol* **15**: 3119–3128
- Kim JE, McAvoy SA, Smith DI, Chen J (2005) Human TopBP1 ensures genome integrity during normal S phase. *Mol Cell Biol* **25**: 10907–10915
- Kimura S, Furukawa T, Kasai N, Mori Y, Kitamoto HK, Sugawara F, Hashimoto J, Sakaguchi K (2003) Functional characterization of two flap endonuclease-1 homologues in rice. *Gene* **314**: 63–71
- Kimura S, Ishibashi T, Hatanaka M, Sakakibara Y, Hashimoto J, Sakaguchi K (2000a) Molecular cloning and characterization of a plant homologue of the origin recognition complex 1 (ORC1). *Plant Sci* **158**: 33–39
- Kimura S, Sakaguchi K (2006) DNA repair in plants. *Chem Rev* **106**: 753–766
- Kimura S, Suzuki T, Yanagawa Y, Yamamoto T, Nakagawa H, Tanaka I, Hashimoto J, Sakaguchi K (2001) Characterization of plant proliferating cell nuclear antigen (PCNA) and flap endonuclease-1 (FEN-1), and their distribution in mitotic and meiotic cell cycles. *Plant J* **28**: 643–653
- Kimura S, Ueda T, Hatanaka M, Takenouchi M, Hashimoto J, Sakaguchi K (2000b) Plant homologue of flap endonuclease-1: molecular cloning, characterization, and evidence of expression in meristematic tissues. *Plant Mol Biol* **42**: 415–427
- Kosugi S, Ohashi Y (2002) E2F sites that can interact with E2F proteins cloned from rice are required for meristematic tissue-specific expression of rice and tobacco proliferating cell nuclear antigen promoters. *Plant J* **29**: 45–59
- Kubota Y, Takase Y, Komori Y, Hashimoto Y, Arata T, Kamimura Y, Araki H, Takisawa H (2003) A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication. *Genes Dev* **17**: 1141–1152
- Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150–163
- Li H, Xie B, Zhou Y, Rahmeh A, Trusa S, Zhang S, Gao Y, Lee EY, Lee MY (2006) Functional roles of p12, the fourth subunit of human DNA polymerase delta. *J Biol Chem* **281**: 14748–14755
- Li KG, Yang JS, Attia K, Su W, He GM, Qian XY (2005) Cloning and characterization of OsORC2, a new member of rice origin recognition complex. *Biotechnol Lett* **27**: 1355–1359
- Lu X, Tan CK, Zhou JQ, You M, Carastro LM, Downey KM, So AG (2002) Direct interaction of proliferating cell nuclear antigen with the small subunit of DNA polymerase delta. *J Biol Chem* **277**: 24340–24345
- Lupas A (1996) Coiled coils: new structures and new functions. *Trends Biochem Sci* **21**: 375–382
- Lupas A (1997) Predicting coiled-coil regions in proteins. *Curr Opin Struct Biol* **7**: 388–393
- Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* **252**: 1162–1164
- Lutzmann M, Maiorano D, Mechali M (2006) A Cdt1-geminin complex licenses chromatin for DNA replication and prevents rereplication during S phase in Xenopus. *EMBO J* **25**: 5764–5774
- Maga G, Hubscher U (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci* **116**: 3051–3060
- Maga G, Villani G, Tillement V, Stucki M, Locatelli GA, Frouin I, Spadari S, Hubscher U (2001) Okazaki fragment processing: modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1. *Proc Natl Acad Sci USA* **98**: 14298–14303
- Maiorano D, Cuvier O, Danis E, Mechali M (2005) MCM8 is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation. *Cell* **120**: 315–328
- Maiorano D, Lutzmann M, Mechali M (2006) MCM proteins and DNA replication. *Curr Opin Cell Biol* **18**: 130–136
- Majka J, Burgers PM (2004) The PCNA-RFC families of DNA clamps and clamp loaders. *Prog Nucleic Acid Res Mol Biol* **78**: 227–260
- Makiniemi M, Hillukkala T, Tuusa J, Reini K, Vaara M, Huang D, Pospiech H, Majuri I, Westerling T, Makela TP, et al (2001) BRCT domain-containing protein TopBP1 functions in DNA replication and damage response. *J Biol Chem* **276**: 30399–30406
- Marino-Ramirez L, Hsu B, Baxevanis AD, Landsman D (2006) The histone database: a comprehensive resource for histones and histone fold-containing proteins. *Proteins* **62**: 838–842
- Martin J, Gruber M, Lupas AN (2004) Coiled coils meet the chaperone world. *Trends Biochem Sci* **29**: 455–458
- Marwedel T, Ishibashi T, Lorbiecke R, Jacob S, Sakaguchi K, Sauter M (2003) Plant-specific regulation of replication protein A2 (OsRPA2) from

- rice during the cell cycle and in response to ultraviolet light exposure. *Planta* **217**: 457–465
- Masai H, You Z, Arai K (2005) Control of DNA replication: regulation and activation of eukaryotic replicative helicase, MCM. *IUBMB Life* **57**: 323–335
- Masuda HP, Ramos GB, de Almeida-Engler J, Cabral LM, Coqueiro VM, Macrini CM, Ferreira PC, Hemerly AS (2004) Genome based identification and analysis of the pre-replicative complex of *Arabidopsis thaliana*. *FEBS Lett* **574**: 192–202
- Masuda-Sasa T, Imamura O, Campbell JL (2006) Biochemical analysis of human Dna2. *Nucleic Acids Res* **34**: 1865–1875
- Masumoto H, Sugino A, Araki H (2000) Dpb11 controls the association between DNA polymerases alpha and epsilon and the autonomously replicating sequence region of budding yeast. *Mol Cell Biol* **20**: 2809–2817
- Milkereit P, Gadal O, Podtelejnikov A, Trumtel S, Gas N, Petfalski E, Tollervey D, Mann M, Hurt E, Tschochner H (2001) Maturation and intranuclear transport of pre-ribosomes requires Noc proteins. *Cell* **105**: 499–509
- Mimura S, Takisawa H (1998) *Xenopus* Cdc45-dependent loading of DNA polymerase alpha onto chromatin under the control of S-phase Cdk. *EMBO J* **17**: 5699–5707
- Miyazawa H, Izumi M, Tada S, Takada R, Masutani M, Ui M, Hanaoka F (1993) Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase alpha-primase complex and their gene expression during cell proliferation and the cell cycle. *J Biol Chem* **268**: 8111–8122
- Mori Y, Yamamoto T, Sakaguchi N, Ishibashi T, Furukawa T, Kadota Y, Kuchitsu K, Hashimoto J, Kimura S, Sakaguchi K (2005) Characterization of the origin recognition complex (ORC) from a higher plant, rice (*Oryza sativa* L.). *Gene* **353**: 23–30
- Moyer SE, Lewis PW, Botchan MR (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci USA* **103**: 10236–10241
- Naryzhny SN, Zhao H, Lee H (2005) Proliferating cell nuclear antigen (PCNA) may function as a double homotrimer complex in the mammalian cell. *J Biol Chem* **280**: 13888–13894
- Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**: 27–43
- Pacek M, Tutter AV, Kubota Y, Takisawa H, Walter JC (2006) Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* **21**: 581–587
- Pacek M, Walter JC (2004) A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. *EMBO J* **23**: 3667–3676
- Pollok S, Stoepel J, Bauerschmidt C, Kremmer E, Nasheuer HP (2003) Regulation of eukaryotic DNA replication at the initiation step. *Biochem Soc Trans* **31**: 266–269
- Qiu J, Qian Y, Frank P, Wintersberger U, Shen B (1999) *Saccharomyces cerevisiae* RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Mol Cell Biol* **19**: 8361–8371
- Ramos GBA, Engler JD, Ferreira PCG, Hemerly AS (2001) DNA replication in plants: characterization of a cdc6 homologue from *Arabidopsis thaliana*. *J Exp Bot* **52**: 2239–2240
- Randell JC, Bowers JL, Rodriguez HK, Bell SP (2006) Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* **21**: 29–39
- Ranjan A, Gossen M (2006) A structural role for ATP in the formation and stability of the human origin recognition complex. *Proc Natl Acad Sci USA* **103**: 4864–4869
- Raynaud C, Perennes C, Reuzeau C, Catrice O, Brown S, Bergounioux C (2005) Cell and plastid division are coordinated through the prereplication factor AtCDT1. *Proc Natl Acad Sci USA* **102**: 8216–8221
- Raynaud C, Sozzani R, Glab N, Domenichini S, Perennes C, Cella R, Kondorosi E, Bergounioux C (2006) Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in *Arabidopsis*. *Plant J* **47**: 395–407
- Reynolds N, Watt A, Fantes PA, MacNeill SA (1998) Cdm1, the smallest subunit of DNA polymerase delta in the fission yeast *Schizosaccharomyces pombe*, is non-essential for growth and division. *Curr Genet* **34**: 250–258
- Ricke RM, Bielinsky AK (2004) Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. *Mol Cell* **16**: 173–185
- Ronceret A, Guillemot J, Lincker F, Gadea-Vacas J, Delorme V, Bechtold N, Pelletier G, Delseny M, Chaboute ME, Devic M (2005) Genetic analysis of two *Arabidopsis* DNA polymerase epsilon subunits during early embryogenesis. *Plant J* **44**: 223–236
- Rose A, Manikantan S, Schraegle SJ, Maloy MA, Stahlberg EA, Meier I (2004) Genome-wide identification of *Arabidopsis* coiled-coil proteins and establishment of the ARABI-COIL database. *Plant Physiol* **134**: 927–939
- Rossi ML, Bambara RA (2006) Reconstituted Okazaki fragment processing indicates two pathways of primer removal. *J Biol Chem* **281**: 26051–26061
- Sabelli PA, Burgess SR, Kush AK, Young MR, Shewry PR (1996) cDNA cloning and characterisation of a maize homologue of the MCM proteins required for the initiation of DNA replication. *Mol Genet* **252**: 125–136
- Sabelli PA, Parker JS, Barlow PW (1999) cDNA and promoter sequences for MCM3 homologues from maize, and protein localization in cycling cells. *J Exp Bot* **50**: 1315–1322
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425
- Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, Nagamura Y, et al (2002) The genome sequence and structure of rice chromosome 1. *Nature* **420**: 312–316
- Sawyer SL, Cheng IH, Chai W, Tye BK (2004) Mcm10 and Cdc45 cooperate in origin activation in *Saccharomyces cerevisiae*. *J Mol Biol* **340**: 195–202
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**: 501–506
- Seki T, Akita M, Kamimura Y, Muramatsu S, Araki H, Sugino A (2006) GINS is a DNA polymerase epsilon accessory factor during chromosomal DNA replication in budding yeast. *J Biol Chem* **281**: 21422–21432
- Snel B, Lehmann G, Bork P, Huynen MA (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* **28**: 3442–3444
- Spicer EK, Rush J, Fung C, Reha-Krantz LJ, Karam JD, Konigsberg WH (1988) Primary structure of T4 DNA polymerase: evolutionary relatedness to eucaryotic and other procaryotic DNA polymerases. *J Biol Chem* **263**: 7478–7486
- Springer PS, Holding DR, Groover A, Yordan C, Martienssen RA (2000) The essential Mcm7 protein PROLIFERA is localized to the nucleus of dividing cells during the G(1) phase and is required maternally for early *Arabidopsis* development. *Development* **127**: 1815–1822
- Springer PS, McCombie WR, Sundaresan V, Martienssen RA (1995) Gene trap tagging of PROLIFERA, an essential MCM2-3-5-like gene in *Arabidopsis*. *Science* **268**: 877–880
- Stevens R, Grelon M, Vezon D, Oh J, Meyer P, Perennes C, Domenichini S, Bergounioux C (2004) A CDC45 homologue in *Arabidopsis* is essential for meiosis, as shown by RNA interference-induced gene silencing. *Plant Cell* **16**: 99–113
- Stevens R, Mariconti L, Rossignol P, Perennes C, Cella R, Bergounioux C (2002) Two E2F sites in the *Arabidopsis* MCM3 promoter have different roles in cell cycle activation and meristematic expression. *J Biol Chem* **277**: 32978–32984
- Stewart JA, Campbell JL, Bambara RA (2006) Flap endonuclease disengages Dna2 helicase/nuclease from Okazaki fragment flaps. *J Biol Chem* **281**: 38565–38572
- Sunderland PA, West CE, Waterworth WM, Bray CM (2004) Choice of a start codon in a single transcript determines DNA ligase 1 isoform production and intracellular targeting in *Arabidopsis thaliana*. *Biochem Soc Trans* **32**: 614–616
- Sunderland PA, West CE, Waterworth WM, Bray CM (2006) An evolutionarily conserved translation initiation mechanism regulates nuclear or mitochondrial targeting of DNA ligase 1 in *Arabidopsis thaliana*. *Plant J* **47**: 356–367
- Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev* **17**: 1153–1165
- Taliercio E, Hendrix B, Stewart JM (2005) DNA content and expression of genes related to cell cycling in developing *Gossypium hirsutum* (Malvaceae) fibers. *Am J Bot* **92**: 1942–1947

- Taylor RM, Hamer MJ, Rosamond J, Bray CM** (1998) Molecular cloning and functional analysis of the *Arabidopsis thaliana* DNA ligase I homologue. *Plant J* **14**: 75–81
- Tikhmyanova N, Coleman TR** (2003) Isoform switching of Cdc6 contributes to developmental cell cycle remodeling. *Dev Biol* **260**: 362–375
- Tominaga K, Johmura Y, Nishizuka M, Imagawa M** (2004) Fad24, a mammalian homolog of Noc3p, is a positive regulator in adipocyte differentiation. *J Cell Sci* **117**: 6217–6226
- Touelle M, Saint-Jean B, Rome C, Couillaud F, Castroviejo M, Benedetto JP** (2002) Two distinct proliferating cell nuclear antigens are present in the wheat cell. *Plant Physiol Biochem* **40**: 743–748
- 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
- Ueno M, Itoh M, Kong L, Sugihara K, Asano M, Takakura N** (2005) PSF1 is essential for early embryogenesis in mice. *Mol Cell Biol* **25**: 10528–10532
- Vandepoele K, Raes J, De Veylder L, Rouze P, Rombauts S, Inze D** (2002) Genome-wide analysis of core cell cycle genes in *Arabidopsis*. *Plant Cell* **14**: 903–916
- Venclovas C, Colvin ME, Thelen MP** (2002) Molecular modeling-based analysis of interactions in the RFC-dependent clamp-loading process. *Protein Sci* **11**: 2403–2416
- von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P** (2007) STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**: D358–362
- Waga S, Stillman B** (1998) The DNA replication fork in eukaryotic cells. *Annu Rev Biochem* **67**: 721–751
- Waga S, Zembutsu A** (2006) Dynamics of DNA binding of replication initiation proteins during de novo formation of pre-replicative complexes in *Xenopus* egg extracts. *J Biol Chem* **281**: 10926–10934
- Walker JE, Saraste M, Runswick MJ, Gay NJ** (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a noncommon nucleotide binding fold. *EMBO J* **1**: 945–951
- Witmer X, Alvarez-Venegas R, San-Miguel P, Danilevskaya O, Avramova Z** (2003) Putative subunits of the maize origin of replication recognition complex ZmORC1-ZmORC5. *Nucleic Acids Res* **31**: 619–628
- Wohlschlegel JA, Dhar SK, Prokhorova TA, Dutta A, Walter JC** (2002) *Xenopus* Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45. *Mol Cell* **9**: 233–240
- Wong SW, Wahl AF, Yuan PM, Arai N, Pearson BE, Arai K, Korn D, Hunkapiller MW, Wang TS** (1988) Human DNA polymerase alpha gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. *EMBO J* **7**: 37–47
- Xia R, Wang J, Liu C, Wang Y, Wang Y, Zhai J, Liu J, Hong X, Cao X, Zhu JK, et al** (2006) ROR1/RPA2A, a putative replication protein A2, functions in epigenetic gene silencing and in regulation of meristem development in *Arabidopsis*. *Plant Cell* **18**: 85–103
- Yang M, Sheila M** (2002) The *Arabidopsis* MEI1 gene likely encodes a protein with BRCT domains. *Sex Plant Reprod* **14**: 355–357
- Yokoi M, Ito M, Izumi M, Miyazawa H, Nakai H, Hanaoka F** (1997) Molecular cloning of the cDNA for the catalytic subunit of plant DNA polymerase alpha and its cell-cycle dependent expression. *Genes Cells* **2**: 695–709
- Yoshida K** (2005) Identification of a novel cell-cycle-induced MCM family protein MCM9. *Biochem Biophys Res Commun* **331**: 669–674
- Zhang YX, Yu ZL, Fu XR, Liang C** (2002) Noc3p, a bHLH protein, plays an integral role in the initiation of DNA replication in budding yeast. *Cell* **109**: 849–860
- Zou L, Stillman B** (1998) Formation of a preinitiation complex by S-phase cyclin CDK-dependent loading of Cdc45p onto chromatin. *Science* **280**: 593–596
- Zou L, Stillman B** (2000) Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. *Mol Cell Biol* **20**: 3086–3096
- Zou Y, Liu Y, Wu X, Shell SM** (2006) Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. *J Cell Physiol* **208**: 267–273