

## Chapter 2

# Evolutionary Diversification of Eukaryotic DNA Replication Machinery

Stephen J. Aves, Yuan Liu, and Thomas A. Richards

**Abstract** DNA replication research to date has focused on model organisms such as the vertebrate *Xenopus laevis* and the yeast species *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. However, animals and fungi both belong to the Opisthokonta, one of about six eukaryotic phylogenetic ‘supergroups’, and therefore represent only a fraction of eukaryotic diversity. To explore evolutionary diversification of the eukaryotic DNA replication machinery a bioinformatic approach was used to investigate the presence or absence of yeast/animal replisome components in other eukaryotic taxa. A comparative genomic survey was undertaken of 59 DNA replication proteins in a diverse range of 36 eukaryotes from all six supergroups. Twenty-three proteins including Mcm2–7, Cdc45, RPA1, primase, some DNA polymerase subunits, RFC1–5, PCNA and Fen1 are present in all species examined. A further 20 proteins are present in all six eukaryotic supergroups, although not necessarily in every species: with the exception of RNase H2B and the fork protection complex component Timeless/Tof1, all of these are members of anciently derived paralogous families such as ORC, MCM, GINS or RPA. Together these form a set of 43 proteins that must have been present in the last common eukaryotic ancestor (LCEA). This minimal LCEA replisome is significantly more complex than the related replisome in Archaea, indicating evolutionary events including

---

S.J. Aves (✉)

Biosciences, College of Life and Environmental Sciences, University of Exeter,  
Geoffrey Pope Building, Stocker Road, Exeter, EX4 4QD, UK  
e-mail: s.j.aves@exeter.ac.uk

Y. Liu • T.A. Richards

Biosciences, College of Life and Environmental Sciences, University of Exeter,  
Geoffrey Pope Building, Stocker Road, Exeter, EX4 4QD, UK

Department of Zoology, Natural History Museum,  
Cromwell Road, London, SW7 5BD, UK  
e-mail: yl274@exeter.ac.uk; thomr@nhm.ac.uk

duplications of DNA replication genes in the LCEA lineage. This pattern of early diversification of the DNA replisome in the LCEA is consistent with similar patterns seen in the early evolution of other complex eukaryotic cellular features.

**Keywords** Comparative genomics • Last common eukaryotic ancestor • Opisthokonta • Phylogeny • Supergroup

## 2.1 Introduction

Most of our knowledge of eukaryotic DNA replication comes from studies on model organisms such as the fungus *S. cerevisiae* and the animal *X. laevis*. But fungi and animals belong to just one of the six major eukaryotic ‘supergroups’ (Adl et al. 2005; Simpson and Roger 2004), so variation and diversification in DNA replication systems remain largely unexplored in the diversity of eukaryotic life. This diversity covers numerous biological forms including important parasite groups, keystone species in environmental processes, and independent lineages that have evolved multicellularity, cellular differentiation and a range of reproductive systems. The recent rise in availability of genome sequence data from a range of eukaryotes allows bioinformatic investigation of the extent to which the yeast/animal replisome components are present, absent, or expanded by gene duplication in other eukaryotic groups. This comparative genomic approach is proving an important tool for understanding the evolution and diversification of numerous cellular systems (Dacks and Field 2007; Dacks et al. 2008; DeGrasse et al. 2009; Hodges et al. 2010; Ramesh et al. 2005; Richards and Cavalier-Smith 2005; Wickstead et al. 2010), providing insight into how they operate and also identifying differentially distributed gene targets for therapeutic agents. This chapter will apply similar approaches to the diversification of DNA replication machinery in extant eukaryotes and the last common eukaryotic ancestor (LCEA). As part of this work we will also compare the eukaryotic form to its homologous counterpart in Archaea, giving insight into the ancestral diversification of this core cellular system.

## 2.2 Eukaryotic Diversity

Eukaryotes have unique features such as a nucleus and other complex cell structures, but also share many cellular and molecular characteristics with one or both of the other two domains of life, the Archaea (formerly, archaebacteria) and the Bacteria (eubacteria). The evolutionary origin of eukaryotes is hotly debated with a number of contesting hypotheses (Embley and Martin 2006; Martin et al. 2001; Martin and Muller 1998), many of which posit that this ancient transition involved endosymbiotic event(s) between two or more prokaryotes, one of which was a member, close relative or ancestor of the Archaea (Martin 2005; Martin et al. 2001).

Indeed, some have claimed an archaeon was the progenitor of the nucleus and represented the first endosymbiotic event in the eukaryotic lineage (Lake and Rivera 1994). Regardless of the details of eukaryogenesis, the similarities of the eukaryote and Archaea DNA replisome and the non-homologous nature of the bacterial replisome are certainly consistent with shared ancestry between Archaea and at least a subsection of primary eukaryotic conglomerations. Whether this subsection derives from an ancestor within the Archaea, or whether Eukarya and Archaea share a common ancestor (the so-called ‘two primary domains’ or ‘three primary domains’ (2D or 3D) scenarios), is the subject of much debate (Gribaldo et al. 2010). What is certain, however, is that many complex cellular characters evolved after the initial conglomeration event(s) in the early eukaryotic lineage and before the diversification of the last common eukaryotic ancestor (LCEA) into extant and sampled taxa. These complex cellular characters include diverse elements of the cytoskeleton (Richards and Cavalier-Smith 2005; Wickstead and Gull 2011; Wickstead et al. 2010), nuclear pore complexes (DeGrasse et al. 2009), elements of the endomembrane system (Dacks and Field 2007; Dacks et al. 2008), centrioles (Hodges et al. 2010) and many genes encoding the machinery of meiosis (Ramesh et al. 2005).

Evolutionary and taxonomic explanations for the diversity of present-day eukaryotic forms are in a state of flux, with different datasets and rival hypotheses identifying a number of different phylogenetic trees and taxonomic hierarchies. These phylogenetic trees reveal between three and eight major eukaryotic clades, the exact number depending on the analysis performed and the dataset used (Baptiste et al. 2002; Burki et al. 2007, 2008; Hampl et al. 2009; Rodriguez-Ezpeleta et al. 2005, 2007). Animals and fungi, together with some unicellular organisms such as free-living choanoflagellates, parasitic Ichthyosporidia, and amoeboid organisms known as nucleariids, belong to the **Opisthokonta**, which is currently recognised as one of the six major eukaryotic phylogenetic ‘supergroups’ (Adl et al. 2005; Simpson and Roger 2004). ‘Opisthokont’ means ‘posterior flagellum’ and refers to the characteristic single rear organ of motility possessed by some animal and fungal cells (think sperm, or the motile zoospores of chytrid fungi) and represents one of the most consistently recovered phylogenetic groupings (Burki et al. 2007, 2008). Flattened mitochondrial cristae are the other ancestral defining feature of this supergroup (Patterson 1999). These cytological characteristics and molecular phylogenies have been used to demonstrate that this group represents a holophyletic clade (Cavalier-Smith 2003; Lang et al. 2002), which helps to explain why yeasts are useful model organisms for biomedical studies. However, we note that both yeast species commonly used for experimental study have undergone relatively recent gene loss events, in some cases limiting their use as comparative models; we discuss examples of this below. For comparative genomics, the opisthokonts represent one of the best sampled groups, with over 100 fungal genomes reported and numerous animal genomes representing the wide diversity of metazoan forms. Increasing effort has been applied to genome sequencing of single cellular relatives of the fungi and animals, including the choanoflagellate *Monosiga brevicollis* (King et al. 2008), while a sequencing initiative to sample further opisthokont taxa

that branch in and around the fungi and the animal radiations is also underway (Ruiz-Trillo et al. 2007).

A range of molecular evidence suggests that the opisthokonts form a sister branch to the **Amoebozoa** supergroup (Baptiste et al. 2002; Burki et al. 2008; Richards and Cavalier-Smith 2005), which includes diverse forms of amoebic protozoa. In terms of genome projects this supergroup is less well represented, with genomes of the cellular slime mould *Dictyostelium discoideum* and the anaerobic dysentery pathogen *Entamoeba histolytica* completed, and that of *Acanthamoeba castellanii* underway.

The positions of the remaining groups, and indeed the number of major clades and how they branch relative to the root of the eukaryotes, remain unclear. However, recognised major groups include the **Plantae** supergroup (also known as Archaeplastida – referring to the ancient primary endosymbiosis of a cyanobacterium – (Adl et al. 2005; Gould et al. 2008)). This contains the familiar land plants (e.g. *Arabidopsis thaliana* and the moss *Physcomitrella patens* genomes) and green algae (e.g. *Chlamydomonas reinhardtii* and *Ostreococcus tauri* genomes), as well as the red algae (rhodophytes – e.g. *Cyanidioschyzon merolae* genome), and a small group of unicellular algae, the glaucophytes. Other algal groups can be found in the **Chromalveolata**, **Rhizaria** and **Excavata**, and are all the product of multiple secondary and/or tertiary endosymbiotic transfers of plastids (Archibald 2009).

The supergroup **Chromalveolata** has changed in terms of constituent groups on a number of occasions. It was originally proposed as a major grouping united by an ancient secondary endosymbiosis of a red alga (Cavalier-Smith 2000). This larger grouping (sometimes called Chromista (Cavalier-Smith 1987, 1998)) has undergone a number of revisions (Burki et al. 2007, 2008) and recent phylogenetic data suggest that there were two separate red algal endosymbioses (Baurain et al. 2010). As such, current versions of the Chromalveolata encompass the alveolates and the stramenopiles which include for example the photosynthetic diatoms (e.g. *Thalassiosira pseudonana* and *Phaeodactylum tricornerutum* genomes), brown algae (e.g. *Ectocarpus siliculosus* and the microalga *Aureococcus anophagefferens*), dinoflagellates, *Chromera* and their non-photosynthetic relatives such as the oomycete potato blight pathogen *Phytophthora*, ciliates (e.g. *Tetrahymena* and *Paramecium*), and parasitic apicomplexa. Many of the apicomplexa possess a remnant plastid organelle, the apicoplast, for example the causative agents of toxoplasmosis and malaria (e.g. *Toxoplasma gondii* and *Plasmodium falciparum* genomes).

Also traditionally included within the Chromalveolata are a group now sometimes referred to as ‘Hacrobia’ – the haptophytes and cryptomonads (cryptophytes). Haptophytes include the coccolithophores, such as *Emiliania huxleyi*, which are ecologically and geologically important phytoplankton, capable of forming huge blooms and whose calcareous platelets form a major constituent of chalk and limestone sedimentary rocks. The Hacrobia acquired their plastids from a red algal endosymbiosis, and current data suggest they constitute a monophyletic group (Okamoto et al. 2009; Patron et al. 2007) along with several heterotrophic protists e.g. the Katablepharids and Telonemids (Burki et al. 2008). At present Hacrobia are

poorly represented by genome sequences and are in a state of phylogenetic limbo as recent analyses suggest the possibility that they may belong to the Plantae supergroup rather than the Chromalveolata (Burki et al. 2008; Hampl et al. 2009; Patron et al. 2007); they are not included in this analysis.

The **Rhizaria** supergroup was defined from molecular data (Archibald et al. 2003; Bass et al. 2005) and unites a diversity of planktonic and benthic heterotrophs with phototrophs derived from another secondary endosymbiosis, in this case a green algal endosymbiosis (e.g. *Bigeloviella natans* for which the genome is currently being sequenced). Some phylogenetic studies indicate affinity between the Rhizaria and certain chromalveolate groups (Burki et al. 2007), but deep evolutionary relationships between the supergroups remain controversial and the Rhizaria will be treated as a separate supergroup in this discussion consistent with the current taxonomic framework (Adl et al. 2005).

The final supergroup, the **Excavata**, comprises mainly flagellates with a wide diversity of morphological forms, most notably the agents that cause sleeping sickness (e.g. *Trypanosoma brucei* genome), giardiasis (e.g. *Giardia intestinalis* genome), and trichomoniasis (e.g. *Trichomonas vaginalis* genome). The Excavata has been a contentious grouping because they share no single defining morphological character – rather they possess a suite of overlapping cellular characters (Simpson et al. 2006). Attempts to test the phylogenetic relationships of these groups have been greatly affected by artefacts such as long-branch attraction (Philippe 2000; Rodriguez-Ezpeleta et al. 2007). However, a recent phylogenomic analysis focused on correcting such artefacts supports the monophyly of the Excavata and confirms a subsection of the excavates including the Discoba (e.g. *Trypanosoma*, *Naegleria* and *Euglena*), metamonads and *Malawimonas* is monophyletic when only slowly-evolving sites are sampled for phylogenetic analysis (Hampl et al. 2009; Rodriguez-Ezpeleta et al. 2007). The status of this group remains controversial however: it includes several long-branch forming taxa, which group together in the Metamonada (e.g. *Giardia* and *Trichomonas*) (Cavalier-Smith 2003). This very group has been suggested to include the primary branch in the eukaryotic radiation (Morrison et al. 2007), implying the root of the eukaryotes may lie within a subsection of the excavates and this may therefore not be a holophyletic group when rooted.

Even from the brief outline presented here it can be seen that there is a huge diversity of eukaryotic life, and that each of the supergroups contains organisms of great ecological and medical importance. To what extent is the process of DNA replication conserved or diverged across these taxa? Notwithstanding some experimental data for plants (Bryant 2010) and trypanosomes (e.g. Dang and Li 2011), little replication research has been carried out on non-animal/fungi organisms, so this question is being addressed by bioinformatic studies using completed genome sequences. These comparisons also enable us to identify which features of the DNA replication system are conserved and ancestral to all sampled eukaryotic forms, and which features are derived. Such analysis is important for comparisons with prokaryotic replication systems, understanding how the replisome has diversified as cell complexity has evolved, and identifying therapeutic targets.

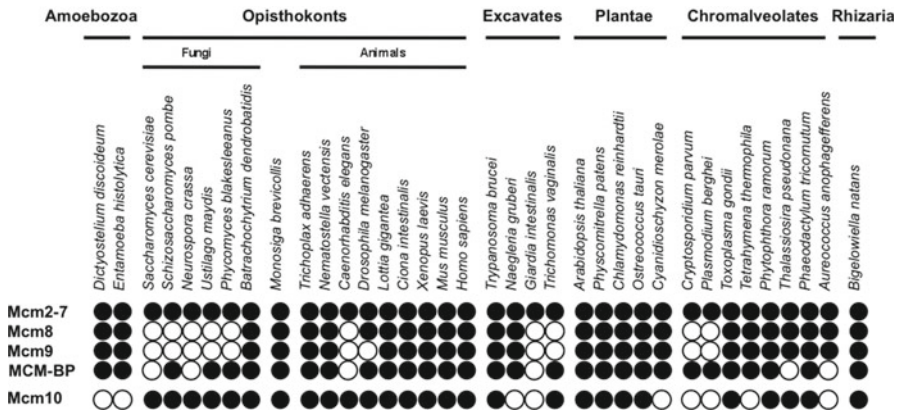
### 2.3 Conservation of Replisome Proteins

A comparative genomic survey of MCM proteins (see Chaps. 6 and 7 for detailed description) in a diverse range of 36 eukaryotes from all six supergroups is shown in Fig. 2.1. BLAST, PSI-BLAST (Altschul et al. 1997), and local Pfam (Bateman et al. 2004) searches using Hidden Markov Models were performed to identify MCM orthologues, with phylogenetic analysis to confirm the identities of the individual MCM paralogues (Liu et al. 2009). In cases of apparent absence, Expressed Sequence Tag (EST) and Genome Survey Sequences (GSS) data of closely-related species were also searched. This analysis enabled us to identify the distribution of DNA replication proteins across the extant eukaryotes. We do not use these data to identify duplication events within each DNA replication subfamily; as such all references to gene duplications refer to anciently derived paralogues present in the LCEA.

All six of the Mcm2–7 helicase subunits were found to be present in all 36 eukaryotes sampled, consistent with the essential roles of all six subunits in the replicative helicase. However, the same pattern was not observed for the Mcm10 replisome protein (see Chap. 11) which in animals/fungi is required for replication initiation and elongation (Gambus et al. 2006; Moore and Aves 2008; Pacek et al. 2006). Mcm10, which is not related to Mcm2–7 as has no identifiable sequence similarity, appears absent from at least some species in three supergroups, and from both Amoebozoa species sampled. While it cannot of course be ruled out that homologues were not detected due to low homology, or that individual genome sequences may not have 100% coverage, this implies that, although Mcm10 has widespread distribution across the eukaryotes, in some species its replication roles are either not required or are provided by other factors.

The Mcm2–7 paralogues Mcm8, Mcm9 and MCM-BP also show widespread but patchy distributions across the eukaryotes, implying gene loss events in more than one lineage (Fig. 2.1). These proteins have received relatively little experimental attention, possibly because they are absent in *S. cerevisiae*, but in vertebrates they have been reported to function in aspects of DNA replication (Gozuacik et al. 2003; Kinoshita et al. 2008; Lutzmann and Mechali 2008; Maiorano et al. 2005; Volkening and Hoffmann 2005). Particularly notable in Fig. 2.1 is the concordant pattern of presence/absence of Mcm8 and Mcm9: in all but one case the absence of one gene corresponds with the absence of the other. This suggests that Mcm8 and Mcm9 may have associated functions in the cell. Phylogenetic analysis groups Mcm8 and Mcm9 as sister paralogues indicating that they also share co-ancestry. The one exception to the co-ordinate loss pattern is *Drosophila melanogaster* in which Mcm8 is present but Mcm9 is absent. This may be the exception that proves the rule however, because closer inspection reveals that all *Drosophila* species have a highly divergent Mcm8 which has a meiotic role; *Drosophila* therefore may not be a good model for Mcm8 in other organisms (Blanton et al. 2005; Liu et al. 2009; Matsubayashi and Yamamoto 2003).

MCM binding protein (MCM-BP) shares only limited homology with Mcm2–9 (Sakwe et al. 2007). However, MCM-BP interacts with MCM proteins and, at least

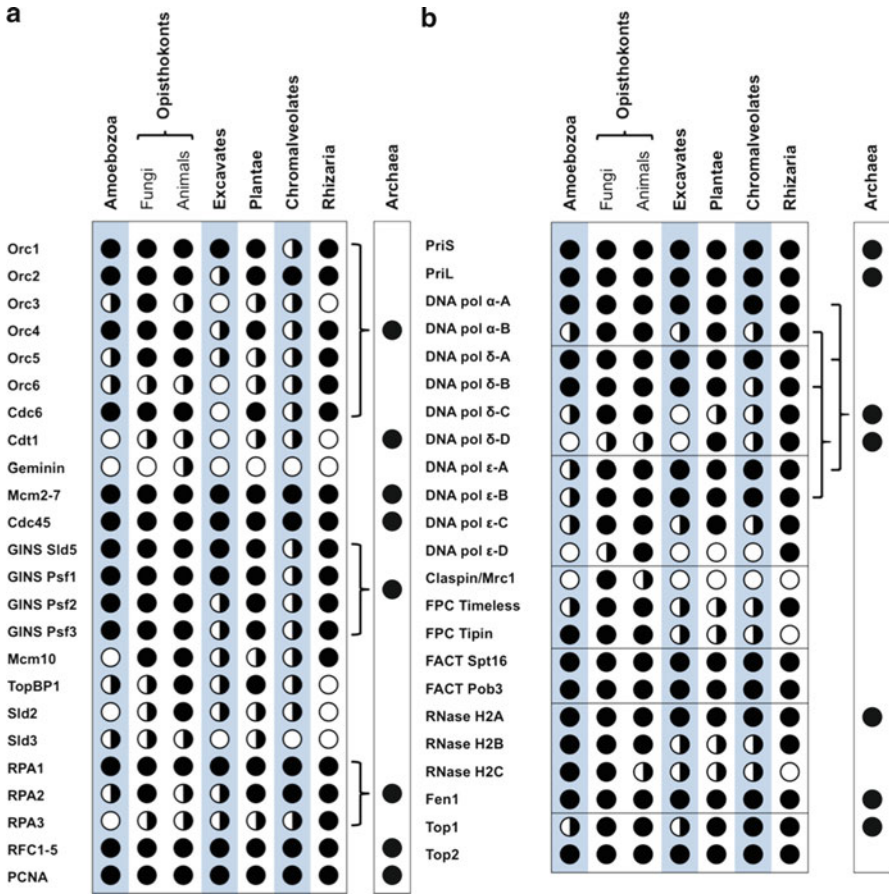


**Fig. 2.1** Distribution of MCM proteins in eukaryotes. *Black circles* indicate detections and *white circles* indicate no homologue detected in a comparative genomic survey of 36 species (Figure adapted from Liu et al. 2009)

in animals and fission yeast, can form an alternative complex in which Mcm2 is replaced by MCM-BP (MCM<sup>MCM-BP</sup>) (Ding and Forsburg 2011; Li et al. 2011; Nishiyama et al. 2011; Sakwe et al. 2007; Takahashi et al. 2008). *Xenopus* MCM-BP has been reported to participate in unloading of the Mcm2–7 complex from chromatin in late S-phase (Nishiyama et al. 2011). MCM-BP is widely distributed across eukaryote taxa but its patchy distribution is different from that of Mcm8/9 and also from that of Mcm10 (Fig. 2.1); this suggests that it does not function in association with these proteins and that its roles are dispensable, or are provided by other components in species such as *S. cerevisiae* and *Caenorhabditis elegans* that lack MCM-BP.

Comparative genomic surveys of 50 other replisome proteins, carried out across a diversity of eukaryotes as for the MCM proteins, are summarised in Fig. 2.2. It can be seen that some replication proteins, like Mcm2–7, are completely conserved in all species sampled – these include Cdc45, RPA1, primase, some DNA polymerase subunits, RFC1–5, PCNA and Fen1 – and are likely to be conserved because they perform a core function in the DNA replisome such as DNA unwinding, single-strand DNA binding, priming, DNA synthesis, clamp loading (where PCNA is the sliding clamp, see Chap. 15) or Okazaki fragment processing.

Other gene families, like Mcm8 and Mcm9, have a widespread distribution across all six supergroups but are absent from individual species, suggesting they have a shared and ancient ancestry but have been lost on multiple occasions. The third category of proteins is those, like Mcm10, which appear to be absent from one or more supergroups, although in almost all cases they have taxonomic distributions well beyond the opisthokonts. This demonstrates a high degree of conservation of the replisome system in eukaryotes. Figure 2.2 also indicates those eukaryotic replisome proteins which have homologues in Archaea.



**Fig. 2.2** Distribution of DNA replication proteins across eukaryotic supergroups. *Black dot* indicates proteins present in all species; *black/white dot* indicates proteins present in some species; *white dot* indicates undetected proteins. See Fig. 2.1 for genomes analysed. Replication proteins with established archaeal homologues are indicated (final column: *black dots*). Lines connecting the eukaryote rows to the Archaea rows indicate paralogue relationships. (a) Initiation, sliding clamp and clamp loader proteins. Distributions of Mcm8, Mcm9 and MCM-BP are in Fig. 2.1. (b) DNA synthesis and associated proteins. DNA polymerase subunits labelled ‘A’ and primase subunit PriS are catalytic; ‘DNA pol ε-C’ and ‘DNA pol ε-D’ designate Dpb3 and Dpb4 subunits respectively. *FACT* Facilitates Chromatin Transcription; *FPC* fork protection complex

## 2.4 Indispensable Replisome Proteins

Many replisome proteins are found in all eukaryotic species (indicated by a row of black dots in Fig. 2.2). These proteins therefore appear to be indispensable components of the replisome: certainly they have remained steadfast components of



the eukaryotic genome during the at least one billion years of evolution that has generated the huge diversity of eukaryotic forms (Berney and Pawlowski 2006; Parfrey et al. 2011). We predict these indispensable proteins provide key functions in the DNA replication process. Interestingly, almost all of them have homologues in archaeal genomes (Fig. 2.2).

The set of indispensable replication proteins includes the Mcm2–7 hexamer plus its accessory factor Cdc45, and the largest subunit of the RPA single-stranded DNA binding protein (see Chap. 10). These represent the key initiation function of DNA unwinding.

For the DNA synthesis functions, the sliding clamp PCNA plus all five subunits of the clamp loader RFC are completely conserved in eukaryotes (Fig. 2.2) (Chia et al. 2010), as are both primase subunits, the catalytic subunit of the initiating DNA polymerase  $\alpha$  and the catalytic subunit of the processive DNA polymerase  $\delta$  (see Chaps. 9 and 12). With the exception of the dysentery pathogen *Entamoeba histolytica*, the catalytic and B-subunit of the leading-strand processive DNA polymerase  $\epsilon$  (Chap. 13) are also completely conserved in eukaryotes. Together these represent all the key activities for DNA synthesis on leading and lagging strands. For processing Okazaki fragments on the lagging strand, indispensable replication proteins ribonuclease H2A and flap endonuclease Fen1 (Chap. 16) are conserved (Fig. 2.2) and although not part of this study, it is likely that DNA ligase I (Chap. 17) can also be added to this list (Ellenberger and Tomkinson 2008). And for chromatin configuration, topoisomerase IIA (Top2) is conserved, as is the FACT (facilitates chromatin transcription) complex of Spt16 and Pob3/SSRP1 for histone interactions and nucleosome disassembly/reassembly (Formosa 2012).

Virtually all of the key indispensable replication proteins outlined above have homologues in Archaea but not in Bacteria (Barry and Bell 2006; Chia et al. 2010; Edgell and Doolittle 1997; Forterre and Gabelle 2009; Johansson and MacNeill 2010; MacNeill 2011; Marinsek et al. 2006; Robbins et al. 2005; Robinson and Bell 2007); only the FACT complex and possibly topoisomerase type IIA (Forterre and Gabelle 2009) appear to be eukaryotic innovations. Again this confirms that the DNA replisome was derived from a lineage within Archaea or a close relative, consistent with models of eukaryotic genesis that suggest an Archaea or Archaea-like entity contributed to the primary eukaryotic conglomeration. In many cases the eukaryotic core replication apparatus contains paralogues which in many Archaea are represented by a single ancestrally derived orthologue, for example the Mcm2–7 heterohexamer is present in all eukaryotes whereas many Archaea have a homohexameric replicative helicase; in those cases where Archaea possess multiple MCM proteins these are best explained by Archaea-specific gene duplications (Chia et al. 2010; Liu et al. 2009). Conserved eukaryotic paralogues such as Mcm2–7 arose by early gene duplication events of an archaeal-like MCM after this gene family was acquired by the eukaryotic progenitor cell prior to the LCEA (Liu et al. 2009). These wider observations suggest that a pattern of ancient gene duplication was important in the early evolution of the eukaryotic DNA replisome prior to the LCEA.

## 2.5 Replisome Proteins Present in All Eukaryotic Supergroups

In addition to the ‘indispensable’ eukaryotic replisome proteins, many other proteins are present in members of all six eukaryotic supergroups, although missing from particular species. These ‘anciently acquired but dispensable’ proteins must therefore represent gene products which were present in the LCEA but have been lost from different lineages; for example, Mcm8 and 9 have been lost on at least five occasions in evolutionary history (Liu et al. 2009). Each ‘anciently acquired but dispensable’ protein must either not be absolutely required for DNA replication, or its function can be substituted by other protein(s). In this context, it is notable that all but two of these 20 proteins are members of anciently derived paralogous gene families (ORC/Cdc6; Mcm2–9; GINS; RPA; DNA pol B; topoisomerase IB) (Figs. 2.1 and 2.2) with only RNase H2B and the fork protection complex (FPC) subunit Timeless (Tim1) having no evidence of ancient gene duplication and paralogues but being differentially lost. Note that these proteins are ‘dispensable’ only in an evolutionary sense: in any one species they may be performing an essential function (e.g. Orc6 is essential in *S. cerevisiae* (Li and Herskowitz 1993) but absent from the related ascomycete fungus *Neurospora crassa*). Examples of replication proteins in this ‘anciently acquired but dispensable’ category are ORC subunits Orc1, Orc2, Orc4 and Orc5; RPA subunit Rpa2; ribonuclease H2B; topoisomerase IB (Top1); the regulatory B-subunits of all three replicative DNA polymerases; and the Dpb3 subunit of DNA polymerase  $\epsilon$ .

Interestingly, the individual ORC/Cdc6 and GINS subunits (see Chaps. 3 and 8) appear to be dispensable. While all species sampled in the Amoebozoa, Opisthokonta, Plantae and Rhizaria possess all four GINS subunits, individual subunits are absent in particular species of the Excavata and the Chromalveolata (Fig. 2.2). It is noteworthy that many Archaea possess only one GINS protein in their replisomes which has homology to two eukaryotic GINS subunits (Yoshimochi et al. 2008). In eukaryotes, the GINS and ORC/Cdc6 complexes are the only anciently derived paralogous gene families amongst the DNA replication proteins which do not contain at least one ‘indispensable’ member (Fig. 2.2).

Aside from the RNase H2B subunit, the Timeless (Tim1) protein of the FPC is the only protein with no evidence of anciently derived paralogues which is ‘anciently acquired but dispensable’. The FPC appears to be a eukaryotic innovation which is conserved across all supergroups but may be dispensable, in whole or in part, in particular species. The two components of the FPC, Timeless (Tof1 in *S. cerevisiae*; Swi1 in *S. pombe*) and Tipin (*ScCsm3*; *SpSwi3*) together function in yeasts and Metazoa to stabilise the paused replisome, activate the replication checkpoint and facilitate chromatin cohesion, thereby contributing to genome stability (Leman et al. 2010; McFarlane et al. 2010). It may be that in certain species both subunits are not required, or this function may be provided in a different manner, or may be less important due to the biology of the organism e.g. faster generation time or tolerance of higher mutation rates.

## 2.6 Replisome Proteins Not Present in All Supergroups

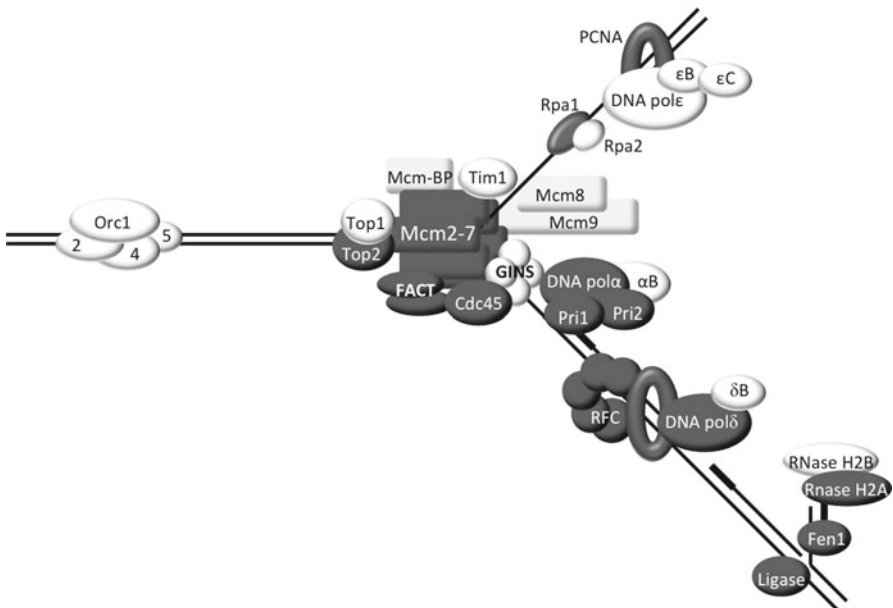
A minority of replisome proteins are only present in some supergroups. Some, like Mcm10, TopBP1/Dpb11, ORC subunits Orc3 and Orc6, RPA subunit 3, RNase H2C and subunits of DNA polymerases  $\delta$  and  $\epsilon$ , have widespread distribution and may possibly have been present in the LCEA but have not been detected in one or two supergroups to date (Figs. 2.1 and 2.2). A few proteins have a more limited distribution and may represent regulatory variations between taxa despite conserved DNA replication mechanisms (Errico and Costanzo 2010; Kearsey and Cotterill 2003). For example the FPC-interacting checkpoint mediator protein Claspin/Mrc1 is limited to opisthokonts, and geminin is an animal-specific inhibitor of the MCM loading factor Cdt1 (Fig. 2.2). It is possible that alternative factors act as regulators of Cdt1 in different eukaryotic taxa, such as the GEM protein in plants (Caro et al. 2007; Caro and Gutierrez 2007).

An alternative explanation for a limited distribution of a regulatory replication protein is that it may be poorly conserved at the sequence level and therefore difficult to detect across supergroups using bioinformatic methods. Sld3 is a case in point: this replication initiation protein was initially thought to be restricted to fungi, but experimental clues and advanced bioinformatic analysis revealed homology with the vertebrate Treslin/Ticrr protein (Kumagai et al. 2010; Sansam et al. 2010) and identified Sld3 homologues in the Plantae and Amoebozoa supergroups (Sanchez-Pulido et al. 2010). Sld3 function as well as structure is conserved between yeast and vertebrates: in yeast, phosphorylation of Sld3 and Sld2/Drc1 by cyclin-dependent kinase (CDK) leads to the formation of a ternary complex with the BRCT-domain protein Dpb11, which is required for CMG complex formation and initiation of DNA replication (Tanaka et al. 2007; Zegerman and Diffley 2007). Similarly, CDK-dependent phosphorylation of Treslin/Ticrr is required for binding to BRCT-domains of TopBP1, the vertebrate Dpb11, and initiation of DNA replication in both *Xenopus* and humans (Boos et al. 2011; Kumagai et al. 2010, 2011). Sld3 phosphorylation sites and the binding region of Dpb11 are conserved in metazoans: phosphorylated Treslin/Ticrr binds to BRCT repeats 1 and 2 of TopBP1, which are homologous to the Sld3-binding BRCT repeats 1 and 2 in Dpb11 (Boos et al. 2011).

The Dpb11 protein has homologues in at least five eukaryotic supergroups (also known as Mei1 in *Arabidopsis*; Mus101 in *Drosophila*; Rad4/Cut5 in *S. pombe*; TopBP1 in humans)(Garcia et al. 2005) which suggests that Sld3 and Sld2 may also be widely conserved. However, the situation for Sld2 is not straightforward in that its apparent animal homologue, the RecQL4 helicase, only shares homology in the N-terminal domain and, although it is required for initiation of DNA replication (Im et al. 2009; Matsuno et al. 2006; Sangrithi et al. 2005; Xu et al. 2009), it is not clear if CDK phosphorylation is conserved (Boos et al. 2011). Other TopBP1-binding proteins may also play a role in initiation of vertebrate DNA replication (Balestrini et al. 2010; Chowdhury et al. 2010). The extent of RecQL4 functional similarity with yeast Sld2 therefore remains to be determined (Masai 2011).

## 2.7 A Complex Ancestral Replisome

An important evolutionary point about replisome proteins represented in all six supergroups, regardless of their dispensability or otherwise, is that these must all have been present in the LCEA. This assumes that horizontal gene transfer is not a factor (Keeling and Palmer 2008; Richards et al. 2011) which is consistent with the complexity hypothesis which suggests gene transfer is rare in DNA replication-encoding gene families (Cotton and McInerney 2010; Jain et al. 1999). It is thus possible to deduce a core replisome present in the LCEA from the sum of the ‘indispensable’ and ‘anciently acquired but dispensable’ replication proteins (Fig. 2.3). It is immediately clear that this is much more complex than the ‘core’ archaeal replisome, i.e. involving additional novel gene families and duplicated members of the archaeal form. This indicates that many events occurred early in the evolution of the eukaryotic cell to produce the replisome of the LCEA, most notably a series of gene duplications to give rise to anciently derived paralogues of single proteins (MCM, GINS, RPA, B-family DNA polymerase, etc.) present in replisomes of extant Archaea. This observation is consistent with many other cellular systems e.g. nuclear pore complexes, membrane trafficking systems, molecular motors, protein complexes that control meiosis, where a large proportion



**Fig. 2.3** Schematic diagram of the possible replisome in the LCEA with ‘indispensable’ proteins in *black* and others (‘anciently acquired but dispensable’) in *white*. DNA ligase I was not part of this study but is included in this diagram as it is likely to be conserved in eukaryotes (Ellenberger and Tomkinson 2008)

of the features are derived in the LCEA (Dacks and Field 2007; Dacks et al. 2008; DeGrass et al. 2009; Hodges et al. 2010; Ramesh et al. 2005; Richards and Cavalier-Smith 2005; Wickstead et al. 2010).

## 2.8 Conclusions

A high level of conservation across all six eukaryotic phylogenetic supergroups indicates that the last common eukaryotic ancestor (LCEA) possessed a complex DNA replication machinery comprising at least 43 proteins. Twenty-three of these ancestral replication proteins appear to be indispensable, in that they are present in the genome of all species sampled; the remaining 20 have been lost in some taxa implying that their function is not essential or can be provided by other factors. The replisome of the LCEA was significantly more complex than replisomes of related Archaea, possessing novel eukaryotic components and multiple paralogues. This indicates evolutionary events including gene duplications in the lineage leading to the LCEA, paralleling the acquisition of other complex cellular features in early eukaryotic evolution.

DNA replication research to date has been heavily concentrated on model opisthokonts. Studies should now be carried out on representatives of other phylogenetic supergroups to both test bioinformatic predictions and to seek other DNA replication components within the diversity of eukaryotic life.

## References

- Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* 52:399–451
- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Archibald JM, Longet D, Pawlowski J, Keeling PJ (2003) A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Mol Biol Evol* Jan; 20(1):62–6. PubMed PMID: 12519907.
- Archibald JM (2009) The puzzle of plastid evolution. *Curr Biol* 19:R81–R88
- Balestrini A, Cosentino C, Errico A, Garner E, Costanzo V (2010) GEMC1 is a TopBP1-interacting protein required for chromosomal DNA replication. *Nat Cell Biol* 12:484–496
- Baptiste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Duruffe L, Gaasterland T, Lopez P, Muller M, Philippe H (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc Natl Acad Sci USA* 99:1414–1419
- Barry ER, Bell SD (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev* 70:876–887

- Bass D, Moreira D, López-García P, Polet S, Chao EE, von der Heyden S, Pawlowski J, Cavalier-Smith T (2005) Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. *Protist*. Aug; 156(2):149-61. PubMed PMID: 16171183.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR (2004) The Pfam protein families database. *Nucleic Acids Res* 32:D138–D141
- Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H (2010) Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* 27:1698–1709
- Berney C, Pawlowski J (2006) A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc Biol Sci* 273:1867–1872
- Blanton HL, Radford SJ, McMahan S, Kearney HM, Ibrahim JG, Sekelsky J (2005) REC, *Drosophila* MCM8, drives formation of meiotic crossovers. *PLoS Genet* 1:343–354
- Boos D, Sanchez-Pulido L, Rappas M, Pearl LH, Oliver AW, Ponting CP, Diffley JFX (2011) Regulation of DNA replication through Sld3-Dpb11 interaction is conserved from yeast to humans. *Curr Biol* 21:1152–1157
- Bryant JA (2010) Replication of nuclear DNA. In: Lüttge U, Beyschlag W, Büdel B, Francis D (eds) *Progress in botany*, vol 71. Springer, Berlin/Heidelberg, pp 25–60
- Burki F, Shalchian-Tabrizi K, Minge M, Skjaveland A, Nikolaev SI, Jakobsen KS, Pawlowski J (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2:e790
- Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new ‘megagroup’ including most photosynthetic eukaryotes. *Biol Lett* 4:366–369
- Caro E, Gutierrez C (2007) A green GEM: intriguing analogies with animal geminin. *Trends Cell Biol* 17:580–585
- Caro E, Castellano MM, Gutierrez C (2007) A chromatin link that couples cell division to root epidermis patterning in *Arabidopsis*. *Nature* 447:213–216
- Cavalier-Smith T (1987) The kingdom Chromista: origin and systematics. In: Round FE, Chapman DJ (eds) *Progress in phycological research*, vol 4. BioPress, Bristol, pp 309–348
- Cavalier-Smith T (1998) A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* 73:203–266
- Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. *Trends Plant Sci* 5: 174–182
- Cavalier-Smith T (2003) The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): their evolutionary affinities and new higher taxa. *Int J Syst Evol Microbiol* 53:1741–1758
- Chia N, Cann I, Olsen GJ (2010) Evolution of DNA replication protein complexes in eukaryotes and archaea. *PLoS One* 5:e10866
- Chowdhury A, Liu G, Kemp M, Chen X, Katrangi N, Myers S, Ghosh M, Yao J, Gao Y, Bubulya P, Leffak M (2010) The DNA unwinding element binding protein DUE-B interacts with Cdc45 in preinitiation complex formation. *Mol Cell Biol* 30:1495–1507
- Cotton JA, McInerney JO (2010) Eukaryotic genes of archaeobacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc Natl Acad Sci USA* 107:17252–17255
- Dacks JB, Field MC (2007) Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J Cell Sci* 120:2977–2985
- Dacks JB, Poon PP, Field MC (2008) Phylogeny of endocytic components yields insight into the process of nonendosymbiotic organelle evolution. *Proc Natl Acad Sci USA* 105:588–593
- Dang HQ, Li Z (2011) The Cdc45-Mcm2-7-GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. *J Biol Chem* 286:32424–32435
- DeGrasse JA, DuBois KN, Devos D, Siegel TN, Sali A, Field MC, Rout MP, Chait BT (2009) Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Mol Cell Proteomics* 8:2119–2130
- Ding L, Forsburg SL (2011) *Schizosaccharomyces pombe* minichromosome maintenance-binding protein (MCM-BP) antagonizes MCM helicase. *J Biol Chem* 286:32918–32930

- Edgell DR, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* 89:995–998
- Ellenberger T, Tomkinson AE (2008) Eukaryotic DNA ligases: structural and functional insights. *Annu Rev Biochem* 77:313–338
- Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440:623–630
- Errico A, Costanzo V (2010) Differences in the DNA replication of unicellular eukaryotes and metazoans: known unknowns. *EMBO Rep* 11:270–278
- Formosa T (2012) The role of FACT in making and breaking nucleosomes. *Biochim Biophys Acta* 1819:247–255
- Forterre P, Gadhelle D (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res* 37:679–692
- 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 V, Furuya K, Carr AM (2005) Identification and functional analysis of TopBP1 and its homologs. *DNA Repair* 4:1227–1239
- Gould SB, Waller RR, McFadden GI (2008) Plastid evolution. *Annu Rev Plant Biol* 59:491–517
- Gozuacik D, Chami M, Lagorce D, Faivre 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
- Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C (2010) The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol* 8:743–752
- Hampel V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc Natl Acad Sci USA* 106:3859–3864
- Hodges ME, Scheumann N, Wickstead B, Langdale JA, Gull K (2010) Reconstructing the evolutionary history of the centriole from protein components. *J Cell Sci* 123:1407–1413
- Im J-S, Ki S-H, Farina A, Jung S, Hurwitz J, Lee J-K (2009) Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. *Proc Natl Acad Sci USA* 106:15628–15632
- Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* 96:3801–3806
- Johansson E, MacNeill SA (2010) The eukaryotic replicative DNA polymerases take shape. *Trends Biochem Sci* 35:339–347
- Kearsey SE, Cotterill S (2003) Enigmatic variations: divergent modes of regulating eukaryotic DNA replication. *Mol Cell* 12:1067–1075
- Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9:605–618
- King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li WQ, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D, Sequencing JGI (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788
- Kinoshita Y, Johnson EM, Gordon RE, Negri-Bell H, Evans MT, Coolbaugh J, Rosario-Peralta Y, Samet J, Slusser E, Birkenbach MP, Daniel DC (2008) Colocalization of MCM8 and MCM7 with proteins involved in distinct aspects of DNA replication. *Microsc Res Tech* 71:288–297
- Kumagai A, Shevchenko A, Shevchenko A, Dunphy WG (2010) Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. *Cell* 140:349–359
- Kumagai A, Shevchenko A, Shevchenko A, Dunphy WG (2011) Direct regulation of Treslin by cyclin-dependent kinase is essential for the onset of DNA replication. *J Cell Biol* 193:995–1007
- Lake JA, Rivera MC (1994) Was the nucleus the first endosymbiont? *Proc Natl Acad Sci USA* 91:2880–2881

- Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G (2002) The closest unicellular relatives of animals. *Curr Biol* 12:1773–1778
- Leman AR, Noguchi C, Lee CY, Noguchi E (2010) Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion. *J Cell Sci* 123:660–670
- Li JJ, Herskowitz I (1993) Isolation of Orc6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* 262:1870–1874
- Li J-J, Schnick J, Hayles J, MacNeill SA (2011) Purification and functional inactivation of the fission yeast MCM<sup>MCM-BP</sup> complex. *FEBS Lett* 585:3850–3855
- Liu Y, Richards TA, Aves SJ (2009) Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evol Biol* 9:60
- Lutzmann M, Mechali M (2008) MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. *Mol Cell* 31:190–200
- MacNeill SA (2011) Protein-protein interactions in the archaeal core replisome. *Biochem Soc Trans* 39:163–168
- 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
- Marinsek N, Barry ER, Makarova KS, Dionne I, Koonin EV, Bell SD (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* 7:539–545
- Martin W (2005) Archaeobacteria (Archaea) and the origin of the eukaryotic nucleus. *Curr Opin Microbiol* 8:630–637
- Martin W, Muller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392:37–41
- Martin W, Hoffmeister M, Rotte C, Henze K (2001) An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol Chem* 382:1521–1539
- Masai H (2011) RecQL4: a helicase linking formation and maintenance of a replication fork. *J Biochem* 149:629–631
- Matsubayashi H, Yamamoto MT (2003) REC, a new member of the MCM-related protein family, is required for meiotic recombination in *Drosophila*. *Genes Genet Syst* 78:363–371
- Matsuno K, Kumano M, Kubota Y, Hashimoto Y, Takisawa H (2006) The N-terminal noncatalytic region of *Xenopus* RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Mol Cell Biol* 26:4843–4852
- McFarlane RJ, Mian S, Dalgaard JZ (2010) The many facets of the Tim-Tipin protein families' roles in chromosome biology. *Cell Cycle* 9:700–705
- Moore K, Aves SJ (2008) Mcm10 and DNA replication in fission yeast. *SEB Exp Biol Ser* 59:45–69
- Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317:1921–1926
- Nishiyama A, Frappier L, Mechali M (2011) MCM-BP regulates unloading of the MCM2-7 helicase in late S phase. *Genes Dev* 25:165–175
- Okamoto N, Chantangsi C, Horak A, Leander BS, Keeling PJ (2009) Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PLoS One* 4:e7080
- 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
- Parfrey LW, Lahr DJG, Knoll AH, Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci USA* 108:13624–13629
- Patron NJ, Inagaki Y, Keeling PJ (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol* 17:887–891
- Patterson DJ (1999) The diversity of eukaryotes. *Am Nat* 154:S96–S124
- Philippe H (2000) Opinion: long branch attraction and protist phylogeny. *Protist* 151:307–316



- Ramesh MA, Malik SB, Logsdon JM (2005) A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol* 15:185–191
- Richards TA, Cavalier-Smith T (2005) Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436:1113–1118
- Richards TA, Leonard G, Soanes DM, Talbot NJ (2011) Gene transfer into the fungi. *Fungal Biol Rev* 25:98–110
- Robbins JB, McKinney MC, Guzman CE, Sriratana B, Fitz-Gibbon S, Ha T, Cann IKO (2005) The euryarchaeota, nature's medium for engineering of single-stranded DNA-binding proteins. *J Biol Chem* 280:15325–15339
- Robinson NP, Bell SD (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci USA* 104:5806–5811
- Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, Bohnert HJ, Philippe H, Lang BF (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330
- Rodriguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF (2007) Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr Biol* 17:1420–1425
- Ruiz-Trillo I, Burger G, Holland PWH, King N, Lang BF, Roger AJ, Gray MW (2007) The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet* 23:113–118
- Sakwe AM, Nguyen T, Athanasopoulos V, Shire K, Frappier L (2007) Identification and characterization of a novel component of the human minichromosome maintenance complex. *Mol Cell Biol* 27:3044–3055
- Sanchez-Pulido L, Diffley JFX, Ponting CP (2010) Homology explains the functional similarities of Treslin/Ticrr and Sld3. *Curr Biol* 20:R509–R510
- Sangrithi MN, Bernal JA, Madine M, Philpott A, Lee J, Dunphy WG, Venkitaraman AR (2005) Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund-Thomson syndrome. *Cell* 121:887–898
- Sansam CL, Cruz NM, Danielian PS, Amsterdam A, Lau ML, Hopkins N, Lees JA (2010) A vertebrate gene, *ticrr*, is an essential checkpoint and replication regulator. *Genes Dev* 24:183–194
- Simpson AGB, Roger AJ (2004) The real 'kingdoms' of eukaryotes. *Curr Biol* 14:R693–R696
- Simpson AG, Inagaki Y, Roger AJ (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol Biol Evol* 23:615–625
- Takahashi N, Lammens T, Boudolf V, Maes S, Yoshizumi T, De Jaeger G, Witters E, Inze D, De Veylder L (2008) The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1. *EMBO J* 27:1840–1851
- Tanaka S, Umemori T, Hirai K, Muramatsu S, Kamimura Y, Araki H (2007) CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* 445:328–332
- Volkening M, Hoffmann I (2005) Involvement of human MCM8 in prereplication complex assembly by recruiting hcd6 to chromatin. *Mol Cell Biol* 25:1560–1568
- Wickstead B, Gull K (2011) The evolution of the cytoskeleton. *J Cell Biol* 194:513–525
- Wickstead B, Gull K, Richards TA (2010) Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. *BMC Evol Biol* 10:110
- Xu Y, Lei Z, Huang H, Dui W, Liang X, Ma J, Jiao R (2009) dRecQ4 is required for DNA synthesis and essential for cell proliferation in *Drosophila*. *PLoS One* 4:e6107
- Yoshimochi T, Fujikane R, Kawanami M, Matsunaga F, Ishino Y (2008) The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J Biol Chem* 283:1601–1609
- Zegerman P, Diffley JFX (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445:281–285