Evolution of DNA Replication Protein Complexes in Eukaryotes and Archaea

Nicholas Chia[1,2]*, Isaac Cann[1,3,4], Gary J. Olsen[1,3]

1 Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **2** Loomis Laboratory of Physics and the Physics Frontier Center: Physics of the Living Cell, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **3** Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America, **4** Laboratory of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States of America

Abstract

Background: The replication of DNA in Archaea and eukaryotes requires several ancillary complexes, including proliferating cell nuclear antigen (PCNA), replication factor C (RFC), and the minichromosome maintenance (MCM) complex. Bacterial DNA replication utilizes comparable proteins, but these are distantly related phylogenetically to their archaeal and eukaryotic counterparts at best.

Methodology/Principal Findings: While the structures of each of the complexes do not differ significantly between the archaeal and eukaryotic versions thereof, the evolutionary dynamic in the two cases does. The number of subunits in each complex is constant across all taxa. However, they vary subtly with regard to composition. In some taxa the subunits are all identical in sequence, while in others some are homologous rather than identical. In the case of eukaryotes, there is no phylogenetic variation in the makeup of each complex—all appear to derive from a common eukaryotic ancestor. This is not the case in Archaea, where the relationship between the subunits within each complex varies taxon-to-taxon. We have performed a detailed phylogenetic analysis of these relationships in order to better understand the gene duplications and divergences that gave rise to the homologous subunits in Archaea.

Conclusion/Significance: This domain level difference in evolution suggests that different forces have driven the evolution of DNA replication proteins in each of these two domains. In addition, the phylogenies of all three gene families support the distinctiveness of the proposed archaeal phylum Thaumarchaeota.

Introduction

DNA replication is one of the defining processes of modern life. The spread of DNA replication likely represents a major evolutionary transition in early life. Duplication of DNA content allows organisms to pass genetic information onto future generations. Mutations during the duplication process enable populations to evolve and adapt. The centrality of DNA replication to such important life processes makes the evolution of the DNA replication machinery all the more significant for understanding the evolution of life.

Chromosome replication in Archaea and eukaryotes requires three ancillary complexes—the proliferating cell nuclear antigen (PCNA), replication factor C (RFC), and the minichromosome maintenance complex (MCM) [1–3]. Each of these three complexes plays an essential role in DNA replication. The MCM complex is thought to function as replicative DNA helicases that unwind the DNA at the replication fork, and PCNA and RFC, known as the clamp and clamp loader, respectively, confer the processive DNA synthesis to the DNA polymerase [1–3].

Without them, large genomes would be extremely difficult to sustain.

We refer the interested reader to Refs. [1–3] for more in-depth reviews of the proteins that act at the replication fork; here we provide only an outline sufficient to introduce the three complexes that we analyze. The process of DNA replication generally begins at specific sites known as origins of replication. The double-stranded DNA is unwound and the two single strands form the templates for replication of the chromosome. The site of DNA replication activity is known as the replication fork, and the supramolecular assembly carrying out the process of replication is known as the replisome. The replisome consists of a large number of protein complexes. Replicative DNA polymerases are incapable of *de novo* DNA synthesis. Therefore, once the single stranded DNA template is generated by the replicative helicase, an RNA primer is initially synthesized by a DNA primase to create a primer/template junction. The primer/template junction is recognized by the clamp loader, which loads the clamp onto this DNA structure. The clamp then recruits the DNA polymerase to the single stranded DNA to perform the actual template guided process of

DNA replication. The function of PCNA is to encircle the DNA and affix, or clamp, the polymerase to the template. In a role analogous to the bacterial beta clamp, PCNA enhances the speed and efficiency of DNA polymerase by enabling the polymerase to synthesize the complementary strand continuously without frequent dissociation.

Figure 1 shows the general subunit organization of PCNA, RFC, and MCM in the archaeal and eukaryotic domains [3,4]. A common theme of these complexes is the repetitive use of homologous or identical subunits. For instance, although PCNA is always a trimer, with the three subunits in a ring (Fig. 1a), the subunits can be of 1, 2, or 3 different sequence types corresponding to α_3, $\alpha_2\beta$, and $\alpha\beta\gamma$ subunit compositions. In eukaryotes, the subunits are all identical, forming a homotrimer, but among the Archaea there is a greater diversity. In the case of RFC, there is always the distinct large subunit (RFCL), while the smaller subunits (RFCS) are of 1,2, or 4 different sequence types. In the case of MCM helicase, the six subunits are drawn from 1, 2, 3, 4, 6, or 8 distinct sequence types, depending on the phylogenetic group. The diversity of sequence types is summarized by phylogeny in Table 1.

In all cases where distinct sequence types are observed within a complex, the proteins are sufficiently similar to imply a common ancestry. For over 40 years it has been observed that gene duplication followed by divergence is an important source of new or modified protein functions [5,6]. The globins are one of the earliest elucidated examples of a protein family that arose from gene duplications [7,8]. Gene family expansions are often associated with the emergence of organismal complexity [5,9]. The number of examples linking increasing organismal complexity and gene duplication continues to grow [10,11]. In fact, the *Saccharomyces cerevisiae* genome appears to be the result of the duplication of a smaller ancestral genome [12]. Such genome duplications have been postulated to be key steps in the increasing complexity of microbes [13] and vertebrates [5].

The extensive role and implications of gene duplication in the evolution for increasing complexity speak to a larger puzzle. The question of emergence of complexity [14,15] encompasses everything from the emergence of early life chemistry [16,17] to higher eukaryotes [5,18] and everything in between [13,19]. In this work, we examine parallel questions about the role of gene duplication and divergence in shaping complexity. The complexity we examine arises from within each of the three protein complexes, and the source of this complexity can be traced by uncovering the evolutionary relationships between the various subunits.
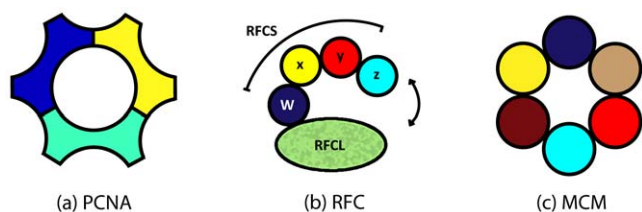

Figure 1. Structural schematic of the PCNA, RFC, and MCM complexes. (a) PCNA consists of 3 subunits forming a ring-like clamp that encloses the DNA polymerase and single stranded DNA. (b) RFC consists of a total of five subunits. Four small subunits (RFCS) form a chain, whose positions are labeled *w*, *x*, *y*, and *z*, that is anchored by *w* RFCS to one large subunit (RFCL). The complex opens between the terminal *z* RFCS and RFCL via an ATP driven conformation change. (c) The MCM complex consists of six MCM proteins in a hexameric ring.
doi:10.1371/journal.pone.0010866.g001

Table 1. Number of PCNA, RFCS, and MCM subunits found in Archaea and eukaryotes for literature [1,3,21–23,27,28,33,51,52,59,66,67] and this work.

Taxonomic Unit	Number of distinct subunits		
	PCNA	RFCS	MCM
Archaea			
Crenarchaeota	1,2,3	1,2	1
Euryarchaeota	1,2	1,2	1–4,8
Korarchaeota	1	1	1
Nanoarchaeota	1	1	1
eukaryotes	1	4	6
total number of subunits in structure	3	4	6

doi:10.1371/journal.pone.0010866.t001

Complexes consisting only of repeated identical subunits are simpler than complexes consisting entirely of homologous, but not identical, subunits. As such, the number of distinct sequence types in each complex serves as a proxy for the overall level of complexity. We trace the emergence of the distinct sequence types in order to put together a picture of how such complexity arose. For instance, where did the distinct subunits come from? Were more specialized subunits invented once and subsequently horizontally gene transferred (HGT) or did complexity increase independently in different lineages? Did simpler complexes with less specialized subunits beget the more specialized subunits in the complexes consisting of distinct subunits, or vice-versa?

Results

With these questions in mind, we examine the phylogeny of the PCNA, RFCS, and MCM subunits. The phylogenetic data is then compared in detail with the known biochemistry of each subunit, in particular, a subunits interaction partners within each complex.

Proliferating Cell Nuclear Antigen

PCNA was so named after it was found to be highly abundant in proliferating cells [20]. PCNA consists of three subunits (Figure 1a) of 1, 2, or 3 sequence types, depending on the phylogenetic group (Table 1). In the interest of clarity and consistency, we introduce our own designations of the PCNA subunits (C1, C2, C3). Table 2 translates our notation to that of previous literature [21–23].

The maximum likelihood phylogeny of the PCNA subunits is shown in Figure 2. This resultant phylogeny generally agrees with the NCBI taxonomy of the corresponding organisms. For clarity, more closely related sequences are shown as a collapsed group. The archaeal and eukaryotic sequences are grouped into separate clades. The Crenarcheota and the Euryarchaea also form distinct groups. The placement of *Nitrosopumilis* and *Cenarcheaum* in Figure 2

Table 2. Crenarchaeotal PCNA nomenclature.

Organism	PCNA C1	PCNA C2	PCNA C3	Reference
Aeropyrum pernix	ApePCNA2	ApePCNA3	ApePCNA1	[22]
Sulfolobus solfataricus	SsoPCNA2	SsoPCNA1	SsoPCNA3	[21]
Sulfolobus tokodaii	StoPCNA3	StoPCNA2	StoPCNA1	[23]

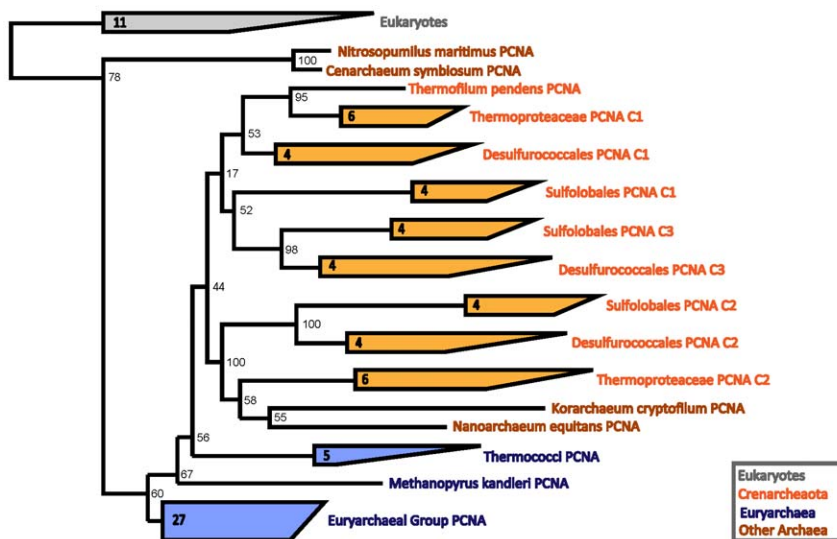doi:10.1371/journal.pone.0010866.t002

Figure 2. PCNA phylogeny, rooted between the Archaea and the eukaryotes. Tree produced using RAxML [63]. Note the proliferation of distinct subunit types in the Crenarchaeota.
doi:10.1371/journal.pone.0010866.g002

is consistent with recent proposals that these organisms belong to a phylum distict from the Crenarchaeota and Euryarchaea, which has been named Thaumarchaeota [24]. The *Korarchaeum* and *Nanoarchaeum* sequences are grouped together within those of the Crenarchaeota. Given the general agreement between the PCNA phylogeny and the organismal taxonomy, HGT does not appear to have occurred.

The eukaryotes and the Euryarchaeota contain only one PCNA gene, with the exception of a few near identical copies of unknown functionality in *Drosphila*, *Arabidopsis*, and *Thermococcus* (see Figure S1) that are generally not present in closely related taxa (data not shown). By contrast, the Crenarchaeota show deep branchings between PCNA subunits. *Cenarchaeum symbiosum* contains one PCNA gene, while the Thermoproteales have either one, as in *Thermofilum pendens*, or two distinct PCNA encoding genes, as in the Thermoprotaeceae. The Desulfurococcales and the Sulfolobales both encode three distinct PCNA subunits.

The phylogenetic relationships between the distinct sequence types yield an interesting picture—one that is consistent with their known biochemical properties. Note that the three distinct types of PCNA roughly group into three clades labeled C1, C2, and C3. Sulfolobales PCNA C1 appears slightly more related to PCNA C3, but not significantly so. We tested this further by constructing a phylogeny of sequences from organisms with more than one distinct sequence type. As shown in Figure 3, in this more focused phylogeny, the PCNA subunits C1, C2, and C3 all group separately.

Furthermore, within each of these three groups, the subunits share similar interaction properties. PCNA C1 appears to have preserved the most ancestral function, sharing the most properties in common with the homotrimeric PCNA subunit. C1 has the most stable dimeric interactions with the other subunits [21–23] and in *Aeropyrum pernix*, C1 is capable of forming a homotrimer [22]. In addition, C1 is present in all heterotrimeric configurations of PCNA (C1-C2-C3, C1-C1-C2, and C1-C2-C2) [21–23]. Phylogenetically, C1 is also the most closely related to the homotrimeric PCNA of *Thermofilum pendens* (Figure 2).

In contrast, C3 takes part only in C1-C2-C3 heterotrimer arrangements [21–23]. Data suggest that in *Sulfolobus solfataricus*, C3 is the last to be recruited into the PCNA trimer [21]. Overall,

C3 has the least interactions with the other subunits [21–23] and appears to be the most functionally divergent of the three subunits from homotrimeric PCNA.

The results for PCNA are consistent with a simpler ancestral homotrimeric PCNA subunit and subsequent duplication and divergence of the distinct subunit types. The archaeal and eukaryotic PCNA both appear to have diverged from a homotrimeric form. Then, in the crenarcheotes, more specialized PCNA sequence types appear to have originated from gene duplications, while the eukaryotes and Euryarchaea retained the ancestral configuration.

The Clamp Loader: Replication Factor C

The RFC complex consists of five subunits, one large (RFCL) and four small (RFCS). The RFC complex opens between the *z*-position RFCS and the RFCL (Figure 1b) in order to open and close PCNA about the DNA polymerase at the replication fork [25,26]. The RFC complex is made up of either 1, 2, or 4 distinct RFCS sequence types, depending on phylogenetic group (Table 1).

The maximum likelihood phylogeny of the RFCS subunits is shown in Figure 4. Again, the phylogeny shows general agreement with the NCBI taxonomy of the corresponding organisms. As such, HGT does not appear in the phylogeny of the RFCS subunits. The eukaryotes, crenarchaeotes, and Euryarchaea form separate groups. As with PCNA, the RFCS tree places the *Cenarcheaum* deep in the branching of archaeal sequences, again consistent with proposals that it be a member of a distinct phylum. The Korarchaea and Nanoarchaea sequences cluster with those of the Euryarchaea. The rooting between the eukaryotes and Archaea follows the canonical pattern, dividing the crenarchaeotes and the Euryarchaea at the base of the archaeal clade.

The phylogeny of the RFCS subunits shows that a RFC with four distinct RFCS sequence types seems to have been present in a common eukaryotic ancestor. This can be seen from the four eukaryotic RFCS clades—one for each RFCS position. On the other hand, the archaeal RFC consists of one or two distinct RFCS subunits [27,28]. Archaea containing only one distinct RFCS form the RFC complex with the same RFCS in all four positions [25]. Euryarchaeal RFC complexes with two distinct RFCS subunits are
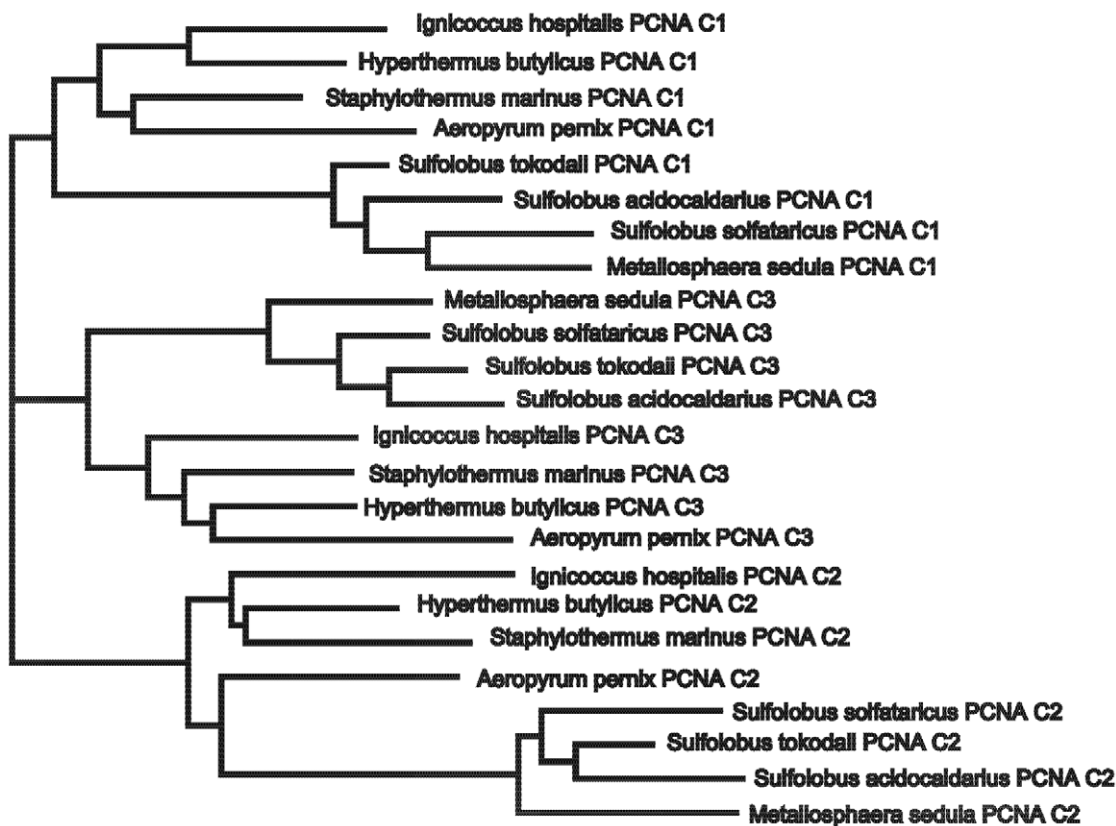
Figure 3. Desulfurococcales and Sulfolobales PCNA phylogeny rooted between PCNA C1, C2, and C3. The branching indicated here lends further support to the three PCNA C1, C2, and C3 groupings.
doi:10.1371/journal.pone.0010866.g003

composed of three RFCS1 at positions w, x, and y, and a single RFCS2 at position z [29]. The configuration of RFC in crenarchaeotes with two distinct subunits has not yet been elucidated.

In Euryarchaeota, the specialization of RFCS into RFCS1 and RFCS2 appears to have occurred before the split between Methanomicrobia and Halobacteria. Following the RFCS1-RFCS2 divergence, there appear to be two independent losses of RFCS2 in the Methanomicrobia, indicated by stars in Figure 4. On the other hand, RFCS1 and RFCS2 could have evolved independently in the Halobacteria and Methanomicrobia—a hypothesis that we do not have enough phylogenetic resolution to affirm or reject. However, data from gene context of RFCS1, shown in Figure S4, is consistent with the phylogeny. (For a more general study of gene context of archaeal DNA replication proteins, we refer the interested reader to Ref. [30]). Also, RFCS1-RFCL complexes have been shown to have some functional activity, further lending plausibility to the notion of independent gene losses [29].

Note that the long branch of RFCS2 corresponds to a change of function. Unlike RFCS and RFCS1, RFCS2 is unable to further extend the small subunit chain since it contains only one RFCS-RFCS binding site [29]. Thus, very conserved amino acid positions in RFCS and RFCS1 corresponding to the second RFCS-RFCS binding site have been allowed to drift in RFCS2 [29], resulting in the long RFCS2 branch seen in Figure 4. Also note that the RFCL rooting of the RFCS tree places the root within the eukaryotes, but is not in significant disagreement with the more sensible rooting between Archaea and eukaryotes (Figure S2).

The results for RFCS are consistent with a simpler ancestral RFC complex containing RFCL and four identical RFCS subunits. In the Archaea, we see subsequent multiple independent duplications and divergences of the distinct subunit types in both crenarchaeotes and Euryarchaea. In eukaryotes, we do not see any intermediate forms with fewer than four distinct RFCS types.

Minichromosome Maintenance Complex

MCM complex plays a role in replication licensing [31] and DNA duplex unwinding [32]. The MCM complex consists of six homologous subunits arranged in a hexameric ring (Figure 1c). The six MCM subunits are drawn from 1, 2, 3, 4, 6, or 8 distinct sequence types, depending on phylogenetic lineage (Table 1).

The phylogeny of the MCM subunits is shown in Figure 5 (shown uncondensed in Figure S3). As in the case of PCNA and RFCS, this phylogeny also shows general agreement with the NCBI taxonomy of the corresponding organisms. The eukaryotes, crenarchaeotes, and Euryarchaea form separate groups. Once again the basal position of *Nitrosopumilus* and *Cenarcheaum* is consistent with a distinct phylum level group, the proposed Thaumarchaeota [24]. Also as in Figures 2 and 4, the Korarchaea and Nanoarchaea sequences group with those of the Euryarchaea. Once again, given the general agreement between gene and organismal relationships, HGT between distantly related organisms does not appear in the phylogeny of the MCM subunits.

The phylogeny of the MCM subunits shows that MCM with six distinct sequence types seems to have been present in a common eukaryotic ancestor, a result previously noted by Liu et al. [33]. By contrast, the archaeal genomes vary in the number of distinct MCM sequence types they contain. The crenarchaeotes appear to
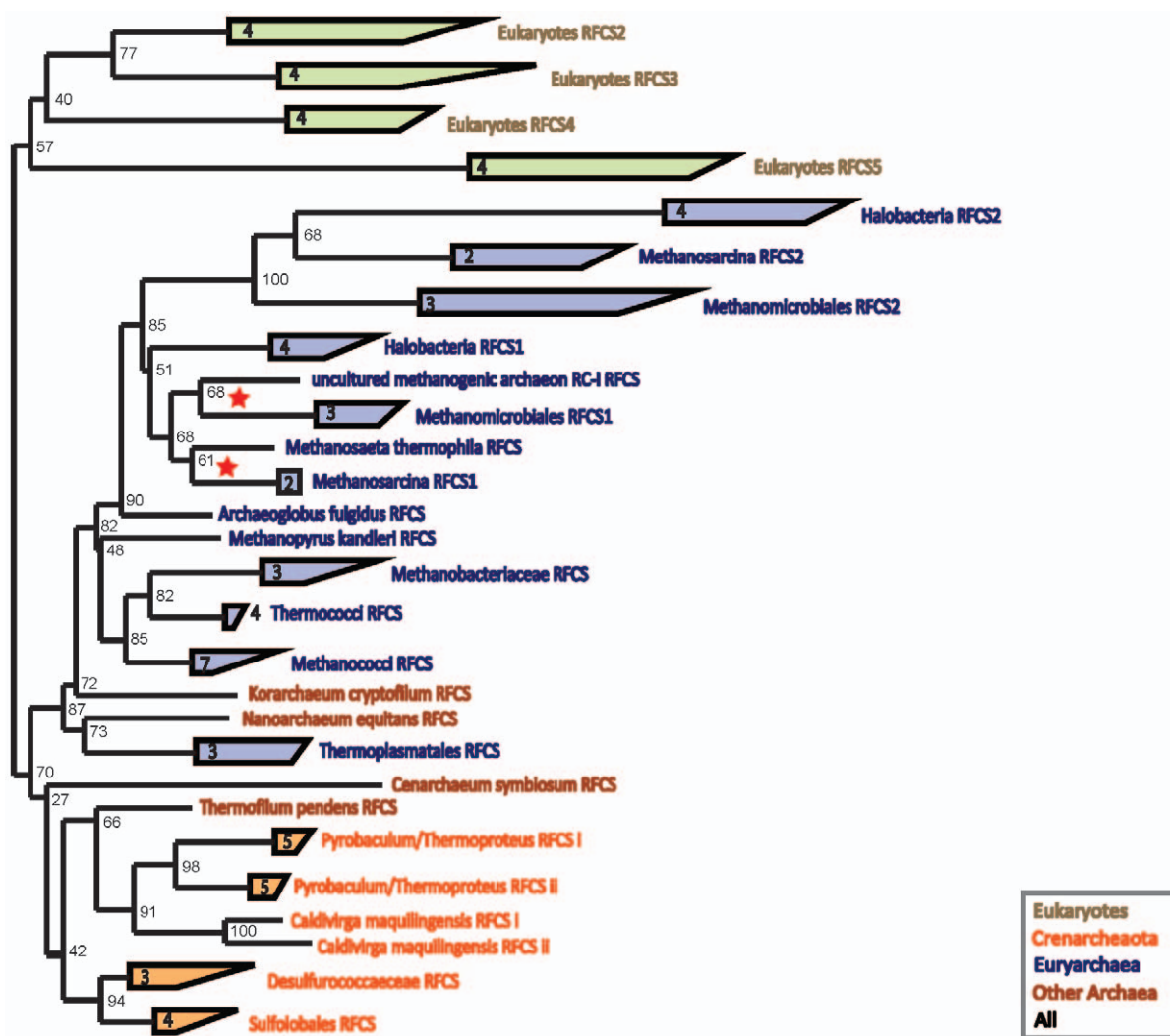
Figure 4. RFCS subunit phylogeny rooted between the Archaea and the eukaryotes. The red stars indicate splits between RFCS and RFCS1 subunit types in the Methanomicrobia, possibly from loss of RFCS2.
doi:10.1371/journal.pone.0010866.g004

contain only a single distinct MCM subunit. On the other hand, the euryarchaeotal genomes contain up to eight distinct MCM subunit genes.

The largest number of MCM genes can be found in the Methanococci. The Methanococci subunits in Figure 5 are labeled based on their phylogeny. The branch lengths between the labeled groups appear indicative of distinct roles among the subunits. The organismal members of each group vary—an indication of gene gains and losses in the Methanococci. For instance, *Methanococcus aeolicus* appears to have lost MCM III while *Methanococcus maripaludis* C6 has five MCM V sequences.

There are multiple eukaryotic MCM complexes. At least two different complexes are known to play a role in unwinding dsDNA [34], MCM2-7 [35] and MCM467 [32,36]. MCM2467 and MCM35 complexes have also been observed [37]. In Archaea, MCM has mostly been characterized in single MCM containing organisms, and several of these MCM proteins have been shown to function as homohexamers [38–44]. It is worth noting, however, that MCM in *Pyrococcus furiosus* requires the presence of accessory protein GINS for unwinding DNA activity [43]. Recently it has been demonstrated that coexpression of the four MCM homologs in *Methanococcus maripaludis* S2 result in the

formation of a heterohexameric complex [45]. Since *M. maripaludis* has a very robust genetic system, we anticipate that subsequent studies will reveal the need for multiple MCM homologs in this archaeon, instead of the usual single homolog in most archaea.

These results are consistent with an ancestral homohexameric MCM complex. In the Archaea, we see subsequent multiple independent duplications and divergences of the distinct subunit types in the Euryarchaea. The crenarchaeotes, on the other hand, retain the simpler ancestral configuration. In eukaryotes, we do not see any intermediate forms with fewer then six distinct sequence types implying a common eukaryotic ancestor containing six distinct MCM subunits.

Discussion

The different numbers of distinct but homologous subunits utilized in the formation of these three complexes in different taxa represent different levels of refinement in the structure and interactions of the complexes. Complexes made up of identical subunits retain the least possibilities for refinement and specialization, while complexes made up entirely of distinct subunits hold the most possibilities for refinement and specialized interactions of

Figure 5. MCM phylogeny, rooted between the Archaea and the Eukaryota. The Methanococci MCM sequences show abundant gene duplication and divergence. They have been labeled I, II, III, IV, and V according to the phylogeny.
doi:10.1371/journal.pone.0010866.g005

each subunit. For example, the eukaryotic RFCS subunits have been shown to play a role in cell cycle regulation, serving as sensors for important processes such as cell cycle arrest and DNA damage repair [46–48]. Likewise, the eukaryotic MCM helicase has been shown to serve as a regulatory target in cell cycle

regulation [48]. From the robust genetic system in *M. maripaludis*, we anticipate that subsequent studies will reveal the need for multiple MCM homologs in this archaeon, instead of the usual single homolog in most archaea. Similarly specialized roles have yet to be identified in the archaeal analogs of these proteins, but

hints of additional function exist. Crenarchaeota exhibit differences in the PCNA interacting protein (PIP) box of proteins such as FEN1 and DNA polymerase B1-differences that are not found in the exclusively homotrimeric PCNA-containing eukaryotes, Euryarchaeota, *Cenarchaeum*, and *Nitrosopumilus* [49]. Thus, while PIP-box containing proteins in the euryarchaeota and the eukaryotes may be able to bind any of the three binding sites in the homotrimeric PCNA, PCNA interacting proteins in the crenarchaeota are known to have preferred interaction partners [21]. This suggests that functional differences may exist between homo- and heterotrimeric PCNA. We can surmise that the level of refinement of the crenarchaeotal PCNA as well as eukaryotic RFC and MCM may play a role in providing additional functionality. If true, we would expect the archaeal subunits from less refined complexes to have lesser roles than those from more refined complexes.

The archaeal branch always begins with complexes formed from exactly one PCNA, RFCS, or MCM distinct subunit type. Thereafter, the archaeal subunits duplicate and diverge, resulting in complexes with a greater level of refinement. In other words, the number of distinct subunits is always *increasing*. These refinements sometimes occur independently in multiple archaeal lineages with no evidence for HGT of distinct subunit types between different species. The agreement among our phylogenies and the concurance with other results supports the conclusions of Brochier et al. [50] that organismal phylogenies can be reconstructed from protein coding genes. It is particularly noteworthy that in all three phylogenies we discuss, the *Nitrosopumilus* and *Cenarcheaum* data are consistent with the proposal for an additional archaeal phylum, the Thaumarchaeota [24].

On the other hand, eukaryotes exhibit no changes in the number of distinct subunits. Instead, the level of refinement remains that of an ancestral Eukaryote from which the modern eukaryotes derive. In two of the cases, RFC and MCM, the ancestral eukaryotic complexes contained the maximum number of possible distinct subunits. In the other case, PCNA, the ancestral eukaryotic complex was made from three identical copies of a single distinct subunit. The same level of refinement has been retained in all modern eukaryotes surveyed in the literature [33,51,52] and during the course of this work.

When the number of distinct subunits increases, the duplication is followed by an initially faster evolution. This can be seen from the longer branch lengths that lead into some subunit clades, for example, the long branches of RFCS2 in Figure 4 or the long branches leading up to PCNA C1, C2, and C3 in Figure 2. This is consistent with a change in the selection on these subunits, i.e., positive selection for a different functional role [53].

Similar patterns of early complexity increase (subunit differentiation) in the common ancestral line of eukaryotes, followed by relatively stable conservation of the composition throughout subsequent speciation has been previously observed in other complexes including the α and β subunits of the proteasome [54] and the core histone subunits [55]. In other words, when the eukaryotic subunits are specialized, intermediate forms are often lacking. We therefore cannot be certain how the eukaryotic complexity arose in these cases. However, we can state with certainty that the many distinct archaeal subunits in the three present cases do not derive from reductive evolution of the eukaryotic complexes, as their subunit proliferation is phylogenetically independent.

Finally, it is interesting to consider the role of DNA processivity within the larger scheme of evolution in early life. Processivity was likely a requirement for the replication of large chromosomes on competitive timescales. One consequence of increased processivity in DNA replication would be the ability to retain additional copies of genes that could then potentially specialize and form more refined complexes. Ironically, the initial evolution of these three complexes may have provided themselves with the means necessary for their own subsequent refinements.

Materials and Methods

Sequences were collected from the NCBI database and identified using BLAST [56] by their similarity to proteins identified experimentally [21–23,26–28,34,35,57–60]. Sequences used in this study are listed in Table S1. Multiple alignments were based on MUSCLE [61] and edited by hand using Jalview [62], and are available upon request. Columns that were judged to be poorly resolved or lacking in information content were removed prior to the maximum likelihood phylogeny. The maximum likelihood phylogeny was performed by RAxML [63] using command line arguments of the form:

./raxmlHPC-PTHREADS -T 8 -f a -x 57843 -p 83755 -N 10000 -m PROTMIXDAYHOFF
    -s alignment_file.phy

The trees presented in the main article were condensed in ARB [64]. Bootstrap values were calculated using PhyML 3.0 (http://www.atgx-montpellier.fr/phyml/) the RAxML-generated trees with their corresponding multiple alignments as the initial input [65].

Supporting Information

Figure S1 Uncondensed PCNA phylogeny.
Found at: doi:10.1371/journal.pone.0010866.s001 (0.03 MB TIF)

Figure S2 Uncondensed RFCS phylogeny, rooted by RFCL.
Found at: doi:10.1371/journal.pone.0010866.s002 (0.03 MB TIF)

Figure S3 Uncondensed MCM phylogeny.
Found at: doi:10.1371/journal.pone.0010866.s003 (0.04 MB TIF)

Figure S4 Genome context for the Methanomicrobiales, Methanosarcinales, Methanosaeta thermophila, and uncultured archaeon RC-I. The key shows the genes that are conserved across contexts. Uncolored genes denote that there was no homolog among these seven contexts.
Found at: doi:10.1371/journal.pone.0010866.s004 (0.27 MB TIF)

Table S1 List of sequences used in this study.
Found at: doi:10.1371/journal.pone.0010866.s005 (0.10 MB PDF)

Author Contributions

Conceived and designed the experiments: NC. Performed the experiments: NC. Analyzed the data: NC IC GO. Wrote the paper: NC IC GO.

References

1. Kornberg A, Baker TA (1992) DNA replication. New York: University Science Books.

2. Grabowski B, Kelman Z (2003) Archaeal DNA replication: Eukaryal proteins in a bacterial context. Annu Rev Microbiology 57: 487–516.

3. Barry ER, Bell SD (2006) DNA replication in the Archaea. Microbiol Mol Biol Rev 70: 876–887.

4. Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. Annu Rev Biochem 71: 333–374.

5. Ohno S (1970) Evolution by gene duplication. New York: Springer-Verlag.

6. Taylor JS, Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 38: 615–643.

7. Hunt LT, Dayhoff MO (1972) The origin of the genetic material in the abnormally long human hemoglobin and chains. Biochem Biophys Res Comm 47: 699–704.

8. Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, et al. (1980) The structure and evolution of the human β-globin gene family. Cell 21: 653–668.

9. Holland PW, Garcia-Fernandez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Development 120: 125–133.

10. Skaer N, Pistillo D, Gibert JM, Lio P, Wülbeck C, et al. (2002) Gene duplication at the achaete–scute complex and morphological complexity of the peripheral nervous system in Diptera. Trends Genet 18: 399–405.

11. Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, et al. (2004) Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol 2: 937–954.

12. Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.

13. Zipkas D, Riley M (1975) Proposal concerning mechanism of evolution of the genome of Escherichia coli. Proc Natl Acad Sci USA 72: 1354–1358.

14. Ohta T (1991) Multigene families and the evolution of complexity. J Mol Evol 33: 34–41.

15. Kauffman SA (1993) The origins of order: Self organization and selection in evolution. New York: Oxford University Press.

16. Oparin AI (1964) The chemical origin of life. Springfield: Charles C Thomas.

17. Morowitz HJ (1993) Beginnings of cellular life: metabolism recapitulates biogenesis. New Haven: Yale University Press.

18. Carroll SB (2005) Endless forms most beautiful: The new science of evo devo and the making of the animal kingdom. New York: W.W. Norton & Company.

19. Olendzenski L, Gogarten JP (1998) Deciphering the molecular record for the early evolution of life: Gene duplication and horizontal gene transfer. In: Wiegel J, Adams M, eds. Thermophiles: The Keys to Molecular Evolution and the Origin of Life. Boca Raton: CRC Press. pp 165–176.

20. Miyachi K, Fritzler MJ, Tan EM (1978) Autoantibody to a nuclear antigen in proliferating cells. J Immunol 121: 2228–2234.

21. Dionne I, Nookala RK, Jackson SP, Doherty AJ, Bell SD (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. Mol Cell 11: 275–282.

22. Imamura K, Fukunaga K, Kawarabayasi Y, Ishino Y (2007) Specific interactions of three proliferating cell nuclear antigens with replication-related proteins in Aeropyrum pernix. Mol Microbiol 64: 308–318.

23. Lu S, Li Z, Wang Z, Ma X, Sheng D, et al. (2008) Spatial subunit distribution and in vitro functions of the novel trimeric PCNA complex from Sulfolobus tokodaii. Biochem Biophys Res Comm 376: 369–374.

24. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nature Rev Microbiol 6: 245–252.

25. Oyama T, Ishino Y, Cann IKO, Ishino S, Morikawa K (2001) Atomic structure of the clamp loader small subunit from Pyrococcus furiosus. Mol Cell 8: 455–463.

26. Bowman GD, O'Donnell M, Kuriyan J (2004) Structural analysis of a eukaryotic sliding DNA clamp–clamp loader complex. Nature 429: 724–730.

27. Pisani FM, De Felice M, Carpentieri F, Rossi M (2000) Biochemical characterization of a clamp-loader complex homologous to eukaryotic replication factor C from the hyperthermophilic archaeon Sulfolobus solfataricus. J Mol Biol 301: 61–73.

28. Chen YH, Kocherginskaya SA, Lin Y, Sriratana B, Lagunas AM, et al. (2005) Biochemical and mutational analyses of a unique clamp loader complex in the archaeon Methanosarcina acetivorans. J Biol Chem 280: 41852–41863.

29. Chen YH, Lin Y, Yoshinaga A, Chhotani B, Lorenzini JL, et al. (2009) Molecular analyses of a three-subunit euryarchaeal clamp loader complex from Methanosarcina acetivorans. J Bact 191: 6539–6549.

30. Berthon J, Cortez D, Forterre P (2008) Genomic context analysis in Archaea suggests previously unrecognized links between DNA replication and translation. Genome Biol 9: R71.

31. Thömmes P, Kubota Y, Takisawa H, Blow JJ (1997) The RLF-M component of the replication licensing system forms complexes containing all six MCM/P1 polypeptides. EMBO J 16: 3312–3319.

32. Ishimi Y (1997) A DNA helicase activity is associated with an MCM4,-6, and-7 protein complex. J Biol Chem 272: 24508–24513.

33. Liu Y, Richards TA, Aves SJ (2009) Ancient diversification of eukaryotic MCM DNA replication proteins. BMC Evol Biol 9: 60.

34. Kanter DM, Bruck I, Kaplan DL (2008) MCM subunits can assemble into two different active unwinding complexes. J Biol Chem 283: 31172–31182.

35. Bochman ML, Bell SP, Schwacha A (2008) Subunit organization of MCM2-7 and the unequal role of active sites in ATP hydrolysis and viability. Mol Cell Biol 28: 5865–5873.

36. You Z, Komamura Y, Ishimi Y (1999) Biochemical analysis of the intrinsic MCM4-MCM6-MCM7 DNA helicase activity. Mol Cell Biol 19: 8003–8015.

37. Lee JK, Hurwitz J (2000) Isolation and characterization of various complexes of the minichromosome maintenance proteins of Schizosaccharomyces pombe. J Biol Chem 275: 18871–18878.

38. Pape T, Meka H, Chen S, Vicentini G, Van Heel M, et al. (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. EMBO Reports 4: 1079–1083.

39. Kasiviswanathan R, Shin JH, Melamud E, Kelman Z (2004) Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains. J Biol Chem 279: 28358–28366.

40. Costa A, Pape T, van Heel M, Brick P, Patwardhan A, et al. (2006) Structural basis of the Methanothermobacter thermautotrophicus MCM helicase activity. Nucleic Acids Research 34: 5829–5838.

41. Haugland GT, Shin JH, Birkeland NK, Kelman Z (2006) Stimulation of MCM helicase activity by a Cdc 6 protein in the archaeon Thermoplasma acidophilum. Nucleic acids research 34: 6337–6344.

42. Atanassova N, Grainge I (2008) Biochemical characterization of the mini-chromosome maintenance (MCM) protein of the Crenarchaeote Aeropyrum pernix and its interactions with the origin recognition complex (ORC) proteins. Biochemistry 47: 13362–13370.

43. 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.

44. Shin JH, Heo GY, Kelman Z (2009) The Methanothermobacter thermautotrophicus MCM helicase is active as a hexameric ring. J Biol Chem 284: 540–546.

45. Walters AD, Chong JPJ (2010) An archaeal order with multiple minichromosome maintenance genes. Microbiology, in press.

46. Zhou BBS, Elledge SJ (2000) The DNA damage response: putting checkpoints in perspective. Nature 408: 433–439.

47. Rouse J, Jackson SP (2002) Interfaces between the detection, signaling, and repair of DNA damage. Science 297: 547–551.

48. Sclafani RA, Holzen TM (2007) Cell cycle regulation of DNA replication. Annu Rev Genet 41: 237–280.

49. Lin IJ, Yoshinaga A, Lin Y, Guzman C, Chen YH, et al. (2010) Molecular analyses of an unusual translesion DNA polymerase from Methanosarcina acetivorans C2A. J Mol Biol 397: 13–30.

50. Brochier C, Forterre P, Gribaldo S (2005) An emerging phylogenetic core of Archaea: phylogenies of transcription and translation machineries converge following addition of new genome sequences. BMC Evol Biol 5: 36.

51. Waga S, Stillman B (1998) The DNA replication fork in eukaryotic cells. Annu Rev Biochem 67: 721–751.

52. Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem 74: 283–315.

53. Pál C, Papp B, Lercher MJ (2006) An integrated view of protein evolution. Nature Rev Genet 7: 337–348.

54. Bouzat JL, McNeil LK, Robertson HM, Solter LF, Nixon JE, et al. (2000) Phylogenomic analysis of the proteasome gene family from early-diverging eukaryotes. J Mol Evol 51: 532–543.

55. Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. Nature Struct Biol 10: 882–891.

56. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3390–3402.

57. Kuriyan J, O'Donnell M (1993) Sliding clamps of DNA polymerases. J Mol Biol 234: 915–925.

58. Krishna TS, Kong XP, Gary S, Burgers PM, Kuriyan J (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell 79: 1233–1243.

59. Kelman Z, Lee JK, Hurwitz J (1999) The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum ΔH contains DNA helicase activity. Proc Natl Acad Sci USA 96: 14783–14788.

60. Cann IKO, Ishino S, Yuasa M, Daiyasu H, Toh H, et al. (2001) Biochemical analysis of replication factor C from the hyperthermophilic archaeon Pyrococcus furiosus. J Bact 183: 2614–2623.

61. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

62. Clamp M, Cuff J, Searle SM, Barton GJ (2004) The jalview java alignment editor. Bioinformatics 20: 426–427.

63. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.

64. Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32: 1363–1371.

65. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704.

66. Carpentieri F, De Felice M, De Falco M, Rossi M, Pisani FM (2002) Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon Sulfolobus solfataricus. J Biol Chem 277: 12118–12127.

67. Grainge I, Scaife S, Wigley DB (2003) Biochemical analysis of components of the pre-replication complex of Archaeoglobus fulgidus. Nucleic Acids Res 31: 4888–4898.