



The replication machinery of LUCA: common origin of DNA replication and transcription

Eugene Koonin, Mart Krupovic, Sonoko Ishino, Yoshizumi Ishino

► To cite this version:

Eugene Koonin, Mart Krupovic, Sonoko Ishino, Yoshizumi Ishino. The replication machinery of LUCA: common origin of DNA replication and transcription. BMC Biology, 2020, 18 (1), pp.61. 10.1186/s12915-020-00800-9 . pasteur-02862286

HAL Id: pasteur-02862286

<https://pasteur.hal.science/pasteur-02862286>

Submitted on 9 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

OPINION

Open Access



The replication machinery of LUCA: common origin of DNA replication and transcription

Eugene V. Koonin^{1*} , Mart Krupovic², Sonoko Ishino³ and Yoshizumi Ishino³

Abstract

Origin of DNA replication is an enigma because the replicative DNA polymerases (DNAPs) are not homologous among the three domains of life, Bacteria, Archaea, and Eukarya. The homology between the archaeal replicative DNAP (PolD) and the large subunits of the universal RNA polymerase (RNAP) responsible for transcription suggests a parsimonious evolutionary scenario. Under this model, RNAPs and replicative DNAPs evolved from a common ancestor that functioned as an RNA-dependent RNA polymerase in the RNA-protein world that predated the advent of DNA replication. The replicative DNAP of the Last Universal Cellular Ancestor (LUCA) would be the ancestor of the archaeal PolD.

Introduction

DNA replication is a central process for all living cells [1]. Therefore, it is astonishing that the key enzymes involved in DNA replication, in particular, the replicative DNA polymerases (rDNAP), are unrelated among the 3 domains of life, Bacteria, Archaea, and Eukarya [2, 3]. This diversity of the replication machineries sharply contrasts with the conservation of the proteins involved in the other key processes of information transfer, namely, transcription and translation, as well as some key metabolic processes, such as nucleotide biosynthesis [4]. The lack of conservation of the rDNAPs and some other key components of the replication machinery, such as helicases and primases, complicates the reconstruction of the replicative apparatus of the ancestral life forms. Moreover, the existence of DNA genomes and, accordingly, DNA replication in the LUCA have been questioned, and the possibility of an RNA-based LUCA has been considered [2, 5]. Here, we focus on the nature of the DNA replication machinery of LUCA and, particularly, the rDNAP and the

major evolutionary events that led to the drastic transformation of this machinery in the three domains of life and in viruses.

There are several families of DNAPs that are involved in replication, repair, or both types of processes [6–8]. The replicative DNAPs of bacteria, archaea, and eukaryotes belong to 3 distinct protein families, and the core catalytic domains of these 3 DNAPs are unrelated to each other, i.e., adopt different protein folds as their catalytic cores (Fig. 1) and therefore are unlikely to share common ancestry (Table 1). The great majority of dsDNA viruses that infect either prokaryotes or eukaryotes and encode their own rDNAPs have the B family polymerase (PolB) that is also responsible for the replication in eukaryotes [9] (Table 1). Archaea encode multiple PolB copies, and with the exception of members of the order Crenarchaeota and some thermophilic members of the Thaumarchaeota [10, 11], also the distinct family D DNAP (PolD) [12–14]. In archaea that possess both DNAPs, it has been recently demonstrated that PolD, rather than PolB, is responsible for the synthesis of both DNA strands [15–18]. The structure of PolD has been recently solved, resulting in a surprising discovery that the catalytic core of PolD is homologous to that of

* Correspondence: koonin@ncbi.nlm.nih.gov

¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

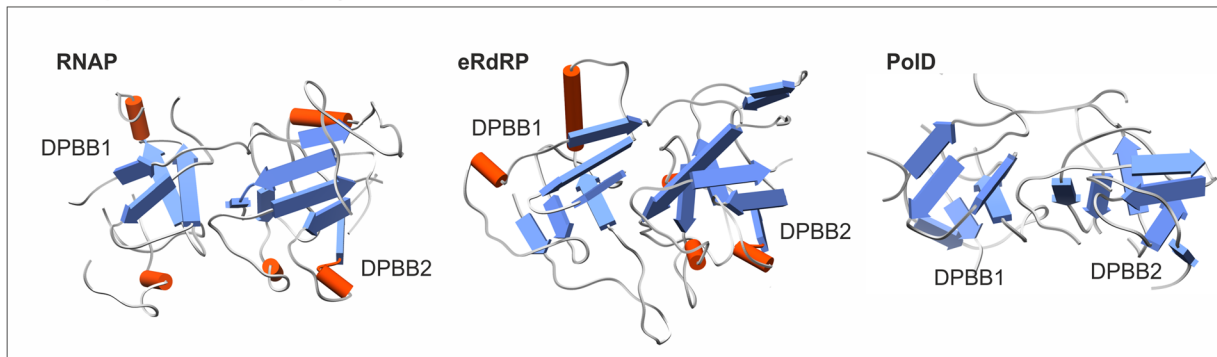
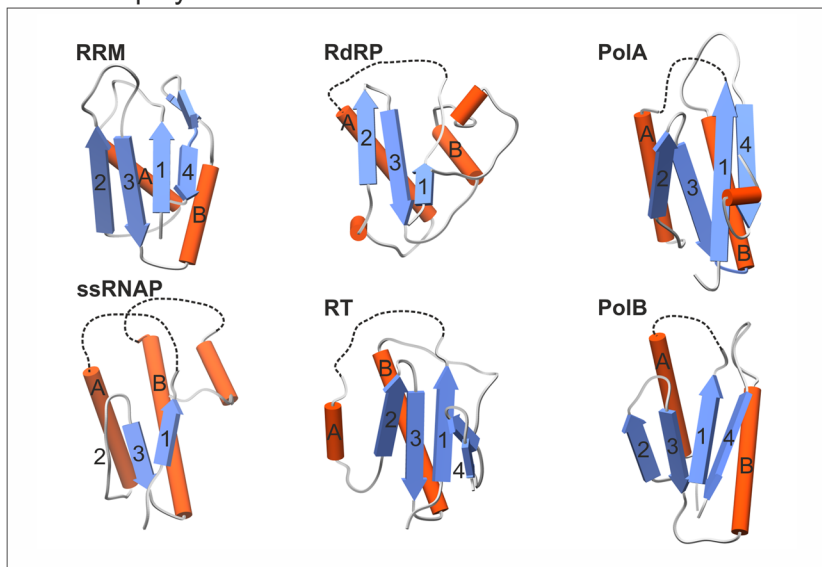
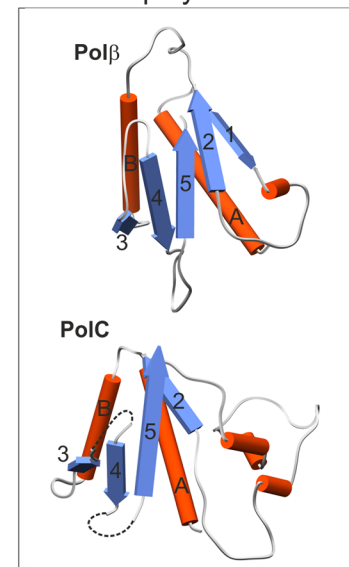
A Double psi-beta barrel polymerases**B RRM-fold polymerases****C PolC-like polymerases**

Fig. 1. The core catalytic domains of DNA and RNA polymerases. **a** Double-psi beta-barrel (DPBB) polymerases. RNAP, multisubunit DNA-dependent RNA polymerase from *Thermus thermophilus* (PDB ID: 1iw7); eRdRP, eukaryotic RNA-dependent RNA polymerase from *Neurospora crassa* (PDB ID: 2j7n); PolD, DP2 subunit of family D DNA polymerase from *Pyrococcus abyssi* (PDB ID: 5ijl). **b** RRM-fold polymerases. RRM, RNA-recognition motif-containing RNA-binding domain of human nucleolin TIAR (PDB ID: 2cqi); RdRP, RNA-dependent RNA polymerase of poliovirus type 1 (PDB ID: 1ra7); PolA, family A DNA polymerase from *Thermus aquaticus* (PDB ID: 1taq); ssRNAP, single-subunit DNA-dependent RNA polymerase of bacteriophage T7 (PDB ID: 1msw); RT, reverse transcriptase of Moloney murine leukemia virus (PDB ID: 1mml); PolB, family B DNA polymerase from *Thermococcus gorgonarius* (PDB ID: 1tgo). **c** PolC-like polymerases. Polβ, DNA polymerase β from *Rattus norvegicus* (PDB ID: 1bpb); PolC, family C DNA polymerase from *Thermus aquaticus* (PDB ID: 2hpi). In **b** and **c**, the major secondary structure elements forming the core palm domain are indicated with numbers (for β-strands) and capital letters (for α-helices). Dashed lines indicate regions where insertions into the core palm domains have occurred; these have been omitted for visualization purposes

the large subunits of the DNA-directed RNA polymerases (RNAPs) that are responsible for transcription in all three domains of life and many large DNA viruses [19–21]. These findings seem to shed unexpected light on the evolution of the replication machineries in the three domains of life as well as viruses. They might even help to infer the nature of the replication machinery in the LUCA suggesting an evolutionary scenario in which PolD takes the central stage as the ancestral replicative polymerase. In the rest of this article, we discuss the reasoning behind this scenario and its implications.

Homologies among DNA and RNA polymerases

The homologous relationships among the DNA and RNA polymerases allow us to infer the evolution of the replication and transcription machineries in the 3 domains of life and in viruses. The key relationships are described below and summarized in Table 1 and Fig. 1.

DNA and RNA polymerases with a catalytic core consisting of two double-psi beta-barrel domains

PolD contains two double-psi beta-barrel (DPBB) domains that both contribute catalytic residues to the DNAP active site. These two DPBB domains have

Table 1 DNA and RNA polymerases involved in replication and transcription in cells and viruses

Polymerase family	Core catalytic domain	Organisms	Functions	DNA-RNA switches
DNA-directed DNAPs				
PolA	Derived RRM-Palm	Bacteria, some phages, mitochondria	Repair in bacteria, replication in some phages, plant and fungal mitochondria, and mitochondrial plasmids	From DNAP to single-subunit RNAP in phages
PolB	RRM-Palm	Archaea, eukaryotes, some bacteria, many large DNA viruses	Replication in Crenarchaeota, eukaryotes, viruses; repair in other archaea, bacteria	To DNA template and product at origin(?)
PolC	Polβ-like nucleotidyl-transferase	Bacteria, a few phages	Replication	PolyA polymerase in eukaryotes
PolD	2xDPBB in a single protein	Archaea except for crenarchaea and some thaumarchaea	Replication	To DNA template and product at origin(?)
AEP	RRM-Palm	Archaea, eukaryotes, some bacteria, diverse mobile elements, many large DNA viruses	RNA priming of replication in archaea, eukaryotes, and dsDNA viruses; replication of plasmids and phages	From RNA synthesis during priming to DNA synthesis during plasmid replication
DNA-directed RNAPs				
Two-DPBB RNAP, two subunits	2xDPBB, one in each subunit	Bacteria, archaea, eukaryotes, many large DNA viruses	Transcription	To DNA template at origin(?) Viroid replication: to RNA template
Two-DPBB RNAP, single subunit	2xDPBB in a single protein	Many phages and linear cytoplasmic plasmids of fungi	Transcription	Eukaryotic RNAi: to RNA template
Single-subunit RNAPs	Derived RRM-Palm	Some phages, mitochondria	Transcription	None
RdRP-RT				
RdRP	RRM-Palm	RNA viruses	Replication	To DNA synthesis in RT(?)
RT	RRM-Palm	Retroelements, reverse-transcribing viruses	Replication in viruses and retroelements; telomere synthesis in eukaryotes	Using both RNA and DNA templates
RNAi eRdRP	2xDPBB in a single protein	Eukaryotes	Small RNA amplification	None

AEP archaeal-eukaryotic primase, DNAP DNA polymerase, DPBB double-psi beta-barrel, RNAP RNA polymerase, RdRP RNA-dependent RNA polymerase, RRM RNA recognition motif, RT reverse transcriptase

homologous counterparts in the RNAP that is universally conserved in all 3 domains of life (Fig. 1a) and is also encoded by many large viruses with dsDNA genomes (Table 1) [19, 21–25]. In cellular RNAPs, the two DPBB domains reside in separate large subunits [26, 27]. Numerous large viruses with dsDNA genomes, including some tailed bacteriophages as well as nucleocytoplasmic large DNA viruses (NCLDV), such as the thoroughly characterized vaccinia virus and the giant mimiviruses, and baculo-like viruses of eukaryotes, also encode RNAPs with two DPBB domain-containing subunits [28–32]. Other viruses, for example, bacillus subtilis SPβ prophage and thermus thermophilus phage P23-45, encode single-subunit RNAPs that contain both DPBB domains within a single polypeptide and are highly divergent, with a limited sequence similarity to the cellular RNAPs [33, 34]. The eukaryotic RNA-dependent RNA polymerase (eRdRP, also known as QDE1 family, after the well-characterized representative from *Neurospora*

crassa) that is responsible for small RNA amplification in eukaryotic RNA interference is another homolog of the RNAP catalytic subunits, in which the two DPBB domains are combined within the same protein (Fig. 1a) [27, 35]. Unexpectedly, the eRdRPs appear to be most closely related to a particular group of bacteriophage single subunit 2xDPBB RNAPs [27, 36]. The two large RNAP subunits are also fused in linear cytoplasmic plasmids from plants and fungi [37], suggesting that such fusions occur repeatedly in the course of RNAP evolution.

Polymerases with a catalytic core based on the RRM domain

PolBs are also represented in all archaea [12, 38] but appear to function as the main replicative polymerases only in Crenarchaeota [39]. Multiple PolB paralogs are fully responsible for DNA replication in eukaryotes [40]. Additionally, PolBs are encoded by numerous viruses and some non-viral mobile genetic elements from all 3

domains of life [9, 29, 41–43]. Some bacteria encode PolBs of apparent virus origin that are involved in repair functions [44]. The catalytic core of PolB is the RNA recognition motif (RRM) domain (also known as the Palm domain in the case of polymerases) that is unrelated to DPBB but is homologous to the core catalytic domains of reverse transcriptases (RT) and viral RNA-dependent RNA polymerases (RdRP) (Fig. 1b) [8, 45].

Conserved C-terminal domains in archaeal and eukaryotic replicative DNA polymerases

The archaeal and eukaryotic replicative DNAPs, respectively, PolD and PolB, share a homologous C-terminal domain (CTD), suggesting that this domain could be specifically important for replication [40, 46]. The CTD contains a distinct Zn-finger as well as the PCNA-binding “PIP motif” [47–52] and interacts with the small DNAP subunit in both archaea and eukaryotes via a conserved interface [20].

Polymerases and nucleotidyltransferases with a Pol β family core domain

The catalytic core of PolC, the bacterial replicative DNAP, is a distant homolog of the catalytic domain of the Pol β family nucleotidyltransferases. This family includes eukaryotic repair DNAPs (Fig. 1c), polyA polymerases and a variety of small-molecule nucleotidyltransferases which are likely to represent the ancestral state of the family [53, 54]. Only a few, poorly characterized phages encode PolC homologs, presumably a late acquisition in virus evolution [9].

DNA and RNA polymerases related to bacterial PolA

PolA, a distant homolog of PolB, is a bacterial repair polymerase. Some phages, Sputnik-like virophages (a group of eukaryotic viruses within the family *Lavidaviridae*), as well as mitochondria and chloroplasts employ PolA homologs as rDNAPs [55–58]. In addition, some phages (e.g., the thoroughly studied phage T7) encode distant homologs of PolA which function as single-subunit RNAPs [59]. These single-subunit phage RNAPs are the apparent ancestors of the mitochondrial RNAP in most eukaryotes [60], with the exception of jacobids that retain the multisubunit, DPBB-based RNAP [61].

An additional important assumption is the RNA world hypothesis [62, 63]. Specifically, we assume a stage in the evolution of life, subsequent to the RNA-only era, when the replicating genomes of the protocellular life forms consisted of RNA including mRNA translated into proteins including RdRPs [64, 65]. We assume that translation evolved within the RNA world, giving rise to an RNA-protein world that presaged the advent of DNA as the dedicated information carrier and the DNA

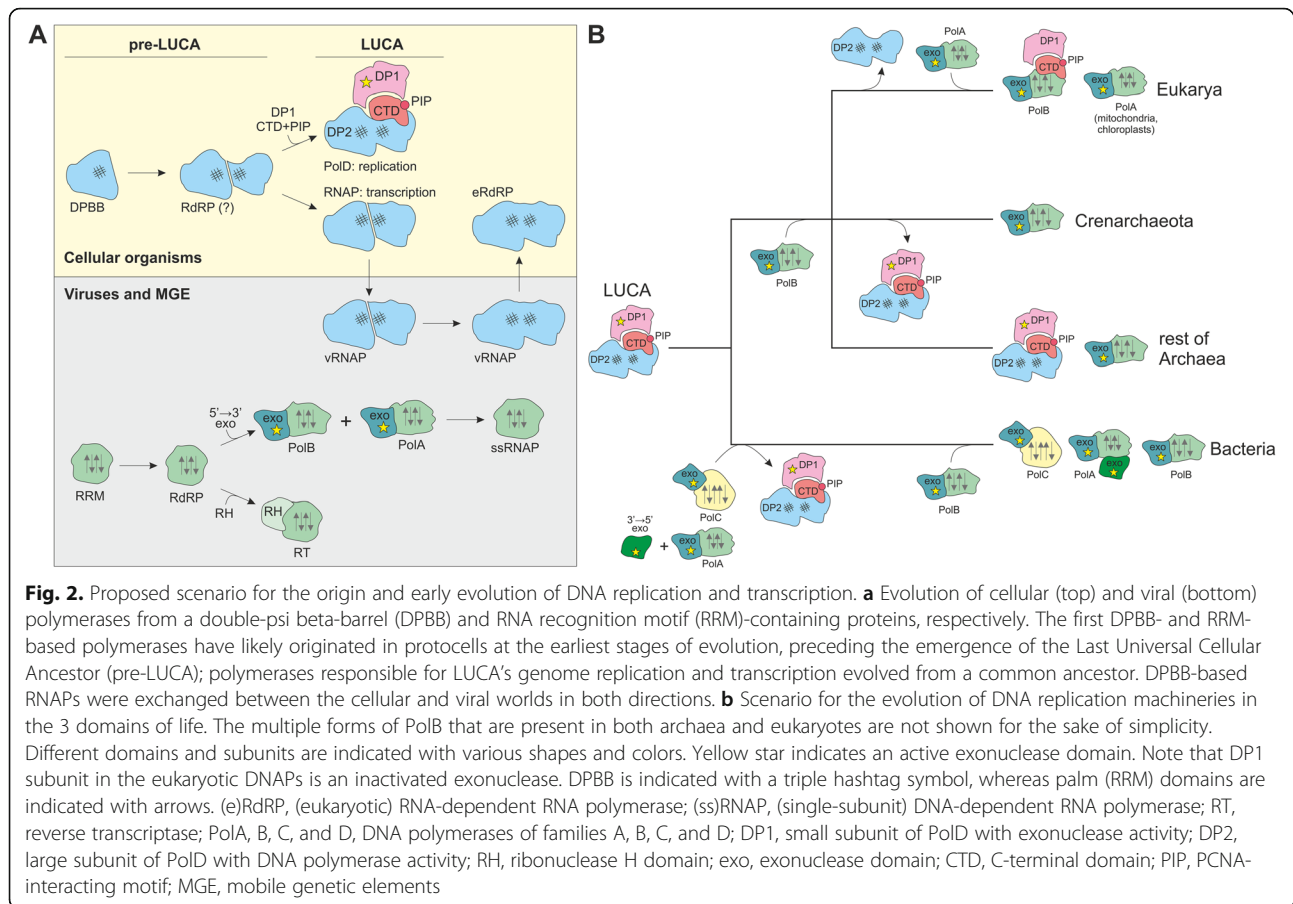
replication machinery, probably, via a reverse transcription stage.

Origins of DNA replication and transcription

The evolutionary connections among the polymerases of cellular organisms and viruses can be superimposed over the evolutionary tree of life that is based on the phylogenies of the universal proteins, namely, translation system components and the large RNAP subunits. This superposition suggests a plausible evolutionary scenario for the evolution of the replicative DNAPs that is intertwined with the evolution of RNAPs and RdRPs (Fig. 2). Given the ubiquity of the RNAPs with two DPBB-containing subunits in all 3 domains of life, this enzyme, obviously, predates the LUCA (Fig. 2a). The extant RNAPs readily assume RdRP activity, as demonstrated by the apparent evolutionary derivation of eRdRP from the catalytic subunit of phage RNAP [27], the involvement of plant RNAP II in viroid replication [66, 67] and animal RNAP II in hepatitis delta virus replication [68], and experimental data on the ability of RNAPs to use RNA as a template in vitro under certain conditions, such as molecular crowding [69].

Thus, it appears likely that the ancestral DPBB polymerase was an RdRP that antedated the origin of DNA replication (i.e., the advent of DNA as the genetic material) [8, 21] and could have been at least one of the enzymes responsible for RNA replication (see below). Given that all extant polymerases in this lineage contain two DPBB domains, it appears most likely that the primordial replicative polymerase already possessed this characteristic pair of DPBB domains that both contribute essential amino acid residues to the catalytic site. These domains conceivably evolved via duplication of a single ancestral DPBB domain (single DPBB domains are present in a variety of metabolic enzymes [21]) and could have resided in either a single or in two subunits (Fig. 2a). The ancestral DPBB form that gave rise to the first protein RdRP might have started as a non-catalytic RNA-binding domain that functioned as a cofactor to ribozyme RdRPs, but following the duplication, evolved the polymerase activity and displaced the ribozyme. Notably, given the apparent origin of the DPBB fold from the so-called RIFT barrel found in such proteins as EF-Tu-like translation factors and ribosomal protein L3 [70], the DPBB-based replication and transcription machineries might be rooted in the translation apparatus which predated DNA genomes.

The origin of DNA-based cells involved differentiation of the primordial two-DPBB RdRP into two distinct lineages: (1) the first replicative DNAP homologous to the extant archaeal PolD and (2) RNAP responsible for transcription (Fig. 2a). The separation of the replication and transcription machineries could have been precipitated



by the accretion of additional domains in both classes of enzymes. The emergence of a DNAP capable of processive DNA synthesis that is required for replication was enabled by the fusion of the DPBB polymerase with a Zn-finger-containing DNA-binding and PCNA-binding protein that gave rise to the CTD and a separate fusion to an RNA-binding KH domain that became the N-terminal domain of the DNAP [20]. The original function of the sliding clamp remains obscure, but given its essentiality in all three domains of life, it is highly likely that a PCNA-like sliding clamp was a component of the LUCA's replisome. The conservation of the PIP motif in both PolD and eukaryotic PolB [52], indeed, strongly suggests that PCNA binding and utilization as the sliding clamp during replication are ancestral features. Accordingly, under this scenario, the replicative DNAP of LUCA was a DPBB-CTD enzyme that subsequently survived as PolD and retained its role in replication, in all archaea except for Crenarchaeota (Fig. 2b). Additionally, either already in LUCA or at an early stage of archaeal evolution, PolD acquired a distinct small subunit, a phosphoesterase that became the proofreading exonuclease [20, 71]. Assuming that the ancestral RdRP contained the two DPBB domains within a single

polypeptide, in the transcription lineage, the ancestral two-DPBB enzyme split into the two subunits each of which captured multiple additional domains including a clamp unrelated to PCNA [21, 26]. An alternative possibility is the fusion of the two ancestral DPBB-containing subunits in PolD. The subsequent evolution of RNAPs involved multiple, independent secondary fusions in archaea and bacteria as well as one or more fusion events that gave rise to the DPBB-based single-subunit bacteriophage RNAPs, one of which was recruited for the eRdRP function in eukaryotic RNAi (Fig. 2a) [72].

Post-LUCA, at the point of divergence between archaea and bacteria, the ancestral DPBB-containing replicative DNAP was displaced by PolC in the bacterial lineage (Fig. 2b). The bacterial DNAP apparently originated from an ancestral Pol β family nucleotidyltransferase although high divergence obscures its specific ancestry. The evolution of archaea involved the acquisition of multiple B family DNAPs (Fig. 2b). Given the widespread of this DNAP family in viruses, the virus origin of archaeal PolBs appears most likely. Ultimately, given the conservation of the core RRM domain, which most likely originated in the RNA-protein world (Fig. 2a), PolB, conceivably, evolved within the pool of mobile

genetic elements including primordial viruses that would parasitize on protocells even in the pre-LUCA era. Specifically, PolB could originate from the RT of primordial retroelements. PolBs were similarly acquired by several groups of bacteria, apparently, at later stages of evolution (Fig. 2b), and in these cases, clearly, from bacteriophages. Thus, the PolB line of descent seems to represent the second, after the DPBB line, evolutionary path from a primordial RNA-binding domain (i.e., RRM) to both RNA and DNA polymerases.

In most archaea, PolBs are not involved in replication but rather in repair-related functions. However, in Crenarchaeota, two paralogous PolB forms replaced the ancestral PolD as the replicative DNAPs. A similar displacement occurred at the onset of the evolution of eukaryotes. In this case, PolB apparently recombined with PolD, replacing the polymerase domain but retaining the CTD [40] (Fig. 2b). Subsequent duplications of PolB at the onset of the evolution of eukaryotes yielded DNAPs ϵ , δ , α , and ξ , the first two of which are responsible for replication. The evolution of PolB in eukaryotes also involved inactivation of the small exonuclease subunit (archaeal DP1) that retained a structural role. Conceivably, the exonuclease activity of the small subunit became dispensable in eukaryotes due to its functional redundancy with the exonuclease domain of the PolB which replaced the PolD large subunit (archaeal DP2) [71].

The transition from DNA to RNA synthesis occurred also in the evolution of the family A of DNAPs. The origin of PolA that is conserved in nearly all bacteria and clearly is ancestral in the bacterial domain remains uncertain. One possibility is that PolA was derived from an ancestral RRM polymerase, perhaps, in a virus, and then was captured by the bacterial ancestor. In bacteria, PolA is a repair enzyme that is not directly involved in replication, but it functions as the replicative polymerase in some viruses and in eukaryotic mitochondria. Notably, PolA was captured by a group of phages as a single-subunit RNAP and was subsequently recruited in the same capacity by eukaryotic mitochondria, in all likelihood, from a phage [73, 74]. Thus, recruitment of viral polymerases, which are often more catalytically efficient than cellular counterparts [75, 76], by cellular organisms appears to be a recurrent theme in evolution, with postulated replacement of PolD by PolB at the onset of eukaryotes being but one example (albeit one of major importance).

Finally, a notable case of switching from RNA to DNA synthesis is the family of archaeal-eukaryotic primases (AEP), another group of RRM (Palm) domain polymerases [77, 78]. The primary function of AEP appears to be the synthesis of RNA primers in archaea, eukaryotes, and many large viruses, such as the NCLDV and herpesviruses. However, many plasmids and other mobile

genetic elements in prokaryotes apparently employ AEP (also known as PrimPol) as the replicative DNAP [79].

Concluding remarks

The origin of DNA replication is one of the most enigmatic subjects in the reconstruction of the early stages in the evolution of life because the replicative DNAPs (as well as primases and the main helicases involved in replication) are not homologous among bacteria, archaea, and eukaryotes. Until recently, this lack of conservation of the key elements of the DNA replication machinery precluded reconstruction of the ancestral state, suggesting multiple origins for DNA replication and even the possibility that LUCA was an RNA-based cell [2, 5]. However, given the universal conservation of other components of the replication apparatus, such as PCNA (sliding clamp), clamp loader ATPase, and ssDNA-binding protein, along with the inferred relatively high complexity of LUCA, comparable to that of modern prokaryotes, such scenarios appear unlikely. The line of reasoning developed here, based primarily on the recently discovered evolutionary connection between PolD and the universally conserved RNAP, allows inference of the ancestral DNAP. Under this scenario, the first transcriptase (RNAP) and the first replicative DNAP evolved from a common ancestor that probably functioned as an RdRP. Thus, the replicative DNAP of the LUCA was the direct ancestor of the extant archaeal replicative DNAP, PolD. The proposed evolutionary scenario appears parsimonious in that the two key processes associated with the advent of DNA genomes, replication and transcription, derive from a common ancestor. An alternative candidate for the role of the replicative DNAP of LUCA potentially could be PolB. However, a PolB-centered scenario for the evolution of replication lacks the symmetry in the evolution of replication and transcription. Besides, PolB is the replicative DNAP only in Crenarchaeota, eukaryotes, and in diverse viruses infecting hosts in all three cellular domains which seem to be best compatible with an origin in viruses or mobile genetic elements.

The proposed scenario traces two lines of descent from primordial RNA-binding domains, DPBB and RRM, to RdRPs (RTs) to RNAPs and DNAPs (Fig. 2a). Among these evolutionary lineages, the DPBB one is associated with the evolution of cells and the RRM one, with the evolution of viruses and mobile genetic elements. The causes of such asymmetry between hosts and parasites remain enigmatic. A notable aspect of the emerging picture of the evolution of replication and transcription is the switch between RNA and DNA template and products that, clearly, occurred on multiple occasions in evolution. Although highly challenging, validation of the current evolutionary scenario by experimental reconstruction of ancestral forms of RNA and DNA polymerases does not seem to be out of the question.

Acknowledgements

EVK is grateful to Dr. Purificación López-García for the valuable discussions.

Authors' contributions

EVK conceived the project. EVK, MK, SI, and YI contributed to the conceptualization and discussion. EVK and MK wrote the manuscript that was read and approved by all authors.

Funding

EVK is supported through the Intramural Research Program of the National Institutes of Health of the USA. MK was supported by l'Agence Nationale de la Recherche (project ENVIRA, no. ANR-17-CE15-0005-01).

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA. ²Archaeal Virology Unit, Institut Pasteur, 75015 Paris, France. ³Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Fukuoka 819-0395, Japan.

Published online: 09 June 2020

References

- Kornberg A, Baker T. DNA replication. 2nd ed. New York: W. H. Freeman and Co; 1992.
- Leipe DD, Aravind L, Koonin EV. Did DNA replication evolve twice independently? *Nucleic Acids Res*. 1999;27(17):3389–401.
- Brown JR, Doolittle WF. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev*. 1997;61(4):456–502.
- Koonin EV. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol*. 2003;1:127–36.
- Forterre P. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci U S A*. 2006;103(10):3669–74.
- Burgers PM, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, Copeland WC, Friedberg EC, Hanaoka F, Hinkle DC, et al. Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J Biol Chem*. 2001;276(47):43487–90.
- Pavlov YI, Shcherbakova PV, Rogozin IB. Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. *Int Rev Cytol*. 2006;255:41–132.
- Raia P, Delarue M, Sauguet L. An updated structural classification of replicative DNA polymerases. *Biochem Soc Trans*. 2019;47(1):239–49.
- Koonin EV. Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases. *Biol Direct*. 2006;1:39.
- Abby SS, Melcher M, Kerou M, Krupovic M, Stieglmeier M, Rossel C, Pfeifer K, Schleper C. *Candidatus Nitrosocaldus cavascurens*, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome. *Front Microbiol*. 2018;9:28.
- Daebeler A, Herbold CW, Vierheilig J, Sedlacek CJ, Pjevac P, Albertsen M, Kirkegaard RH, de la Torre JR, Daims H, Wagner M. Cultivation and genomic analysis of “*Candidatus Nitrosocaldus islandicus*,” an obligately thermophilic, ammonia-oxidizing Thaumarchaeon from a hot spring biofilm in Graendalur Valley, Iceland. *Front Microbiol*. 2018;9:193.
- Makarova KS, Krupovic M, Koonin EV. Evolution of replicative DNA polymerases in archaea and their contributions to the eukaryotic replication machinery. *Front Microbiol*. 2014;5:354.
- Cann IK, Ishino Y. Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics*. 1999;152(4):1249–67.
- Cann IKO, Komori K, Toh H, Kanai S, Ishino Y. A heterodimeric DNA polymerase: evidence that members of euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci U S A*. 1998;95(24):14250–5.
- Kushida T, Narumi I, Ishino S, Ishino Y, Fujiwara S, Imanaka T, Higashibata H. Pol B, a family B DNA polymerase, in *Thermococcus kodakarensis* is important for DNA repair, but not DNA replication. *Microbes Environ*. 2019;34(3):316–26.
- Cubonova L, Richardson T, Burkhart BW, Kelman Z, Connolly BA, Reeve JN, Santangelo TJ. Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. *J Bacteriol*. 2013;195(10):2322–8.
- Li Z, Santangelo TJ, Cubonova L, Reeve JN, Kelman Z. Affinity purification of an archaeal DNA replication protein network. *MBio*. 2010;1(5):e00221–10.
- Sarmiento F, Mrazek J, Whitman WB. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A*. 2013;110(12):4726–31.
- Sauguet L, Raia P, Henneke G, Delarue M. Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography. *Nat Commun*. 2016;7:12227.
- Raia P, Carroni M, Henry E, Pehau-Arnaudet G, Brule S, Beguin P, Henneke G, Lindahl E, Delarue M, Sauguet L. Structure of the DP1-DP2 PolD complex bound with DNA and its implications for the evolutionary history of DNA and RNA polymerases. *PLoS Biol*. 2019;17(1):e3000122.
- Sauguet L. The extended “two-barrel” polymerases superfamily: structure, function and evolution. *J Mol Biol*. 2019;431(20):4167–83.
- Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol*. 2011;9(2):85–98.
- Fouqueau T, Blombach F, Werner F. Evolutionary origins of two-barrel RNA polymerases and site-specific transcription initiation. *Annu Rev Microbiol*. 2017;71:331–48.
- Lane WJ, Darst SA. Molecular evolution of multisubunit RNA polymerases: structural analysis. *J Mol Biol*. 2009;395(4):686–704.
- Lane WJ, Darst SA. Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J Mol Biol*. 2010;395(4):671–85.
- Iyer LM, Koonin EV, Aravind L. Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene*. 2004;335:73–88.
- Iyer LM, Koonin EV, Aravind L. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol*. 2003;3:1.
- Iyer LM, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res*. 2006;117(1):156–84.
- Yutin N, Koonin EV. Hidden evolutionary complexity of nucleocytoplasmic large DNA viruses of eukaryotes. *Virology*. 2012;9:161.
- Passarelli AL, Guarino LA. Baculovirus late and very late gene regulation. *Curr Drug Targets*. 2007;8(10):1103–15.
- Hillen HS, Bartuli J, Grimm C, Dienemann C, Bedenk K, Szalay AA, Fischer U, Cramer P. Structural basis of poxvirus transcription: transcribing and capping Vaccinia complexes. *Cell*. 2019;179(7):1525–36 e1512.
- Grimm C, Hillen HS, Bedenk K, Bartuli J, Neyer S, Zhang Q, Huttenhofer A, Erlacher M, Dienemann C, Schlosser A, et al. Structural basis of poxvirus transcription: Vaccinia RNA polymerase complexes. *Cell*. 2019;179(7):1537–50 e1519.
- Berdysheva Z, Westblade LF, Florens L, Koonin EV, Chait BT, Ramanculov E, Washburn MP, Darst SA, Severinov K, Minakhin L. Temporal regulation of gene expression of the *Thermus thermophilus* bacteriophage P23-45. *J Mol Biol*. 2011;405(1):125–42.
- Forrest D. Unusual relatives of the multisubunit RNA polymerase. *Biochem Soc Trans*. 2019;47(1):219–28.
- Salgado PS, Koivunen MR, Makeyev EV, Bamford DH, Stuart DI, Grimes JM. The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol*. 2006;4(12):e434.
- Drobysheva AV, Panafidina SA, Kolesnik MV, Klimuk EI, Minakhin L, Yakunia MV, Borukhov S, Nilsson E, Holmfeldt K, Yutin N, et al. Structure and function of virion RNA polymerase of crAss-like phage. <https://www.biorxiv.org/content/10.1101/2020.03.07.982082v1>.
- Sykora M, Pospisek M, Novak J, Mrnova S, Krasny L, Vopalensky V. Transcription apparatus of the yeast virus-like elements: architecture, function, and evolutionary origin. *PLoS Pathog*. 2018;14(10):e1007377.
- Rogozin IB, Makarova KS, Pavlov YI, Koonin EV. A highly conserved family of inactivated archaeal B family DNA polymerases. *Biol Direct*. 2008;3:32.
- Yan J, Beattie TR, Rojas AL, Schermerhorn K, Gristwood T, Trinidad JC, Albers SV, Roversi P, Gardner AF, Abrescia NGA, et al. Identification and characterization of a heterotrimeric archaeal DNA polymerase holoenzyme. *Nat Commun*. 2017;8:15075.

40. Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin EV. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol ϵ and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct*. 2009;4:11.
41. Kazlauskas D, Krupovic M, Venclovas C. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic Acids Res*. 2016;44(10):4551–64.
42. Redrejo-Rodriguez M, Ordóñez CD, Berjon-Otero M, Moreno-Gonzalez J, Aparicio-Maldonado C, Forterre P, Salas M, Krupovic M. Primer-independent DNA synthesis by a family B DNA polymerase from self-replicating mobile genetic elements. *Cell Rep*. 2017;21(6):1574–87.
43. Krupovic M, Koonin EV. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat Rev Microbiol*. 2015;13(2):105–15.
44. Dapa T, Fleuriel S, Bredeche MF, Matic I. The SOS and RpoS regulons contribute to bacterial cell robustness to genotoxic stress by synergistically regulating DNA polymerase Pol II. *Genetics*. 2017;206(3):1349–60.
45. Krupovic M, Dolja VV, Koonin EV. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat Rev Microbiol*. 2019;17(7):449–58.
46. Shen Y, Tang XF, Matsui E, Matsui I. Subunit interaction and regulation of activity through terminal domains of the family D DNA polymerase from *Pyrococcus horikoshii*. *Biochem Soc Trans*. 2004;32(Pt 2):245–9.
47. Bauer RJ, Wolff ID, Zuo X, Lin HK, Trakselis MA. Assembly and distributive action of an archaeal DNA polymerase holoenzyme. *J Mol Biol*. 2013;425(23):4820–36.
48. Xu X, Yan C, Kossmann BR, Ivanov I. Secondary interaction interfaces with PCNA control conformational switching of DNA polymerase PolB from polymerization to editing. *J Phys Chem B*. 2016;120(33):8379–88.
49. Tori K, Kimizu M, Ishino S, Ishino Y. DNA polymerases BI and D from the hyperthermophilic archaeon *Pyrococcus furiosus* both bind to proliferating cell nuclear antigen with their C-terminal PIP-box motifs. *J Bacteriol*. 2007;189(15):5652–7.
50. Kuba Y, Ishino S, Yamagami T, Tokuhara M, Kanai T, Fujikane R, Daiyasu H, Atomi H, Ishino Y. Comparative analyses of the two proliferating cell nuclear antigens from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells*. 2012;17(11):923–37.
51. Castrec B, Rouillon C, Henneke G, Flament D, Querellou J, Raffin JP. Binding to PCNA in euryarchaeal DNA replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B. *J Mol Biol*. 2009;394(2):209–18.
52. Madru C, Henneke G, Raia P, Hugonnet-Beaufet I, Pehau-Arnauudet G, England P, Lindahl E, Delarue M, Carroni M, Sauguet L. Structural basis for the increased processivity of D-family DNA polymerases in complex with PCNA. *Nat Commun*. 2020;11(1):1591.
53. Aravind L, Koonin EV. DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res*. 1999;27(7):1609–18.
54. Bailey S, Wing RA, Steitz TA. The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell*. 2006;126(5):893–904.
55. Morley SA, Peralta-Castro A, Bribea LG, Miller J, Ong KL, Ridge PG, Oliphant A, Aldous S, Nielsen BL. Arabidopsis thaliana organelles mimic the T7 phage DNA replisome with specific interactions between Twinkle protein and DNA polymerases Pol1A and Pol1B. *BMC Plant Biol*. 2019;19(1):241.
56. Oldenburg DJ, Bendich AJ. DNA maintenance in plastids and mitochondria of plants. *Front Plant Sci*. 2015;6:883.
57. Lee SJ, Richardson CC. Choreography of bacteriophage T7 DNA replication. *Curr Opin Chem Biol*. 2011;15(5):580–6.
58. Iyer LM, Abhiman S, Aravind L. A new family of polymerases related to superfamily A DNA polymerases and T7-like DNA-dependent RNA polymerases. *Biol Direct*. 2008;3:39.
59. Cheetham GM, Steitz TA. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr Opin Struct Biol*. 2000;10(1):117–23.
60. Schwinghammer K, Cheung AC, Morozov YI, Agaronyan K, Temiakov D, Cramer P. Structure of human mitochondrial RNA polymerase elongation complex. *Nat Struct Mol Biol*. 2013;20(11):1298–303.
61. Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*. 1997;387(6632):493–7.
62. Higgs PG, Lehman N. The RNA world: molecular cooperation at the origins of life. *Nat Rev Genet*. 2015;16(1):7–17.
63. Joyce GF. The antiquity of RNA-based evolution. *Nature*. 2002;418(6894):214–21.
64. Forterre P. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie*. 2005;87(9–10):793–803.
65. Wolf YI, Koonin EV. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Direct*. 2007;2:14.
66. Wang Y, Qu J, Ji S, Wallace AJ, Wu J, Li Y, Gopalan V, Ding B. A land plant-specific transcription factor directly enhances transcription of a pathogenic noncoding RNA template by DNA-dependent RNA polymerase II. *Plant Cell*. 2016;28(5):1094–107.
67. Ding B. Viroids: self-replicating, mobile, and fast-evolving noncoding regulatory RNAs. *Wiley Interdiscip Rev RNA*. 2010;1(3):362–75.
68. Modahl LE, Macnaughton TB, Zhu N, Johnson DL, Lai MM. RNA-dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. *Mol Cell Biol*. 2000;20(16):6030–9.
69. Takahashi S, Okura H, Sugimoto N. Bisubstrate function of RNA polymerases triggered by molecular crowding conditions. *Biochemistry*. 2019;58(8):1081–93.
70. Coles M, Hulko M, Djuranovic S, Truffault V, Koretke K, Martin J, Lupas AN. Common evolutionary origin of swapped-hairpin and double-psi beta barrels. *Structure*. 2006;14(10):1489–98.
71. Aravind L, Koonin EV. Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res*. 1998;26(16):3746–52.
72. Iyer LM, Aravind L. Insights from the architecture of the bacterial transcription apparatus. *J Struct Biol*. 2012;179(3):299–319.
73. Falkenberg M, Larsson NG, Gustafsson CM. DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem*. 2007;76:679–99.
74. Filee J, Forterre P. Viral proteins functioning in organelles: a cryptic origin? *Trends Microbiol*. 2005;13(11):510–3.
75. Menendez-Arias L, Andino R. Viral polymerases. *Virus Res*. 2017;234:1–3.
76. Salas M, de Vega M. Protein-primed replication of bacteriophage Phi29 DNA. *Enzymes*. 2016;39:137–67.
77. Iyer LM, Koonin EV, Leipe DD, Aravind L. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res*. 2005;33(12):3875–96.
78. Kazlauskas D, Sezonov G, Charpin N, Venclovas C, Forterre P, Krupovic M. Novel families of archaeo-eukaryotic primases associated with mobile genetic elements of Bacteria and Archaea. *J Mol Biol*. 2018;430(5):737–50.
79. Zhu B, Wang L, Mitsunobu H, Lu X, Hernandez AJ, Yoshida-Takashima Y, Nunoura T, Tabor S, Richardson CC. Deep-sea vent phage DNA polymerase specifically initiates DNA synthesis in the absence of primers. *Proc Natl Acad Sci U S A*. 2017;114(12):E2310–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.