

Article

Not peer-reviewed version

Possible Crystallization Process in the Origin of Bacteria, Archaea, Viruses, and Mobile Elements

Akari Yoshimura and [Masayuki Seki](#)*

Posted Date: 11 January 2024

doi: 10.20944/preprints202401.0891.v1

Keywords: Last universal common ancestor; Evolution; Bacteria, Archaea; Virus; DNA replication



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Possible Crystallization Process in the Origin of Bacteria, Archaea, Viruses, and Mobile Elements

Akari Yoshimura and Masayuki Seki *

Division of Biochemistry, Faculty of Pharmaceutical Sciences, Tohoku Medical and Pharmaceutical University, Sendai, Miyagi 981-8558, Japan

* Correspondence: seki@tohoku-mpu.ac.jp; ORCID: 0000-0001-9824-0843

Abstract: Our understanding of the divergence of bacteria, archaea, viruses, and mobile elements from the last universal common ancestor (LUCA) has not improved substantially since the crystallization hypothesis (explaining transition from cellular machinery to cellular life) was proposed by Woese. Here, we propose a hypothesis for the simultaneous emergence of bacteria, archaea, viruses, and mobile elements by sequential and concrete biochemical and cellular pathways. According to the hypothesis, the LUCA was a non-free-living pool of single operon type genomes composed of single-stranded (ss) RNA, double-stranded (ds) RNA, RNA/DNA hybrid, ssDNA, or dsDNA at an ancient submarine alkaline vent. Each dsDNA operon was transcribed by different systems in σ , TFIIB, or TBP genomes. Upon the fusion of multiple dsDNA operons by recombinase, the transcription system (in the σ , TFIIB, or TBP genome) was preferentially selected, leading to the first genetic linkage (described by Morgan). Vertical inheritance eventually led to Bacteria (σ genome) and Archaea (TBP genome). Eigen's paradox (error catastrophe) can be overcome by the parallel gain of DNA replication and DNA repair mechanisms in both genomes. Enlarged DNA enabled efficient local biochemical reactions. Both genomes independently recruited lipids to facilitate reactions by forming coacervates (liquid droplets) at the chamber of the vent, leading to the lipid divide. Bilayer lipid membrane formation, proto-cell formation with a permeable membrane, proto-cell division, and the evolution of membrane-associated biochemistry are presented in detail. Simultaneous crystallization of systems in non-free-living bacteria and non-free-living archaea triggered the co-crystallization of primitive viruses and mobile elements. An arms race between non-free-living cells and primitive viruses finally led to free-living cells with a cell wall and mature viruses at the original vent. Cells and viruses spread to all vents on the planet, explaining the universality of the genetic code on Earth. The proposed scenario provides a plausible explanation for the origin of diverse taxa from the LUCA vertical and horizontal gene transfer.

Keywords: last universal common ancestor; evolution; bacteria; archaea; virus; DNA replication

1. Introduction

A phylogenetic analysis of ribosomal DNA (rDNA) revealed three domains of life, Bacteria, Archaea, and Eucarya (Figure 1A) [1,2]. The theory that symbioses of Bacteria and Archaea yielded Eucarya (Figure 1A(a)) and Plante (Figure 1A(b)) was proposed by Margulis [3]. Subsequently, Woese proposed an order of cellular evolution from an RNA world [4] to the last universal common ancestor (LUCA) of Bacteria and Archaea mediated by a progenote (Figure 1A) [5,6].

Koonin and Martin hypothesized that early cells evolved via 18 steps (Figure 1B(a)) at ancient submarine alkaline vents (Figure 1B(b)) [7,8]. Of note, the size of honeycomb-like chambers at both extant and fossil vents (Figure 1B(b)) [7] is nearly equivalent to the size of eukaryotic cells (Figure 1A). The LUCA was predicted to be non-free-living [5,7] with the extant genetic code, translation system, and hundreds of proteins [9,10].

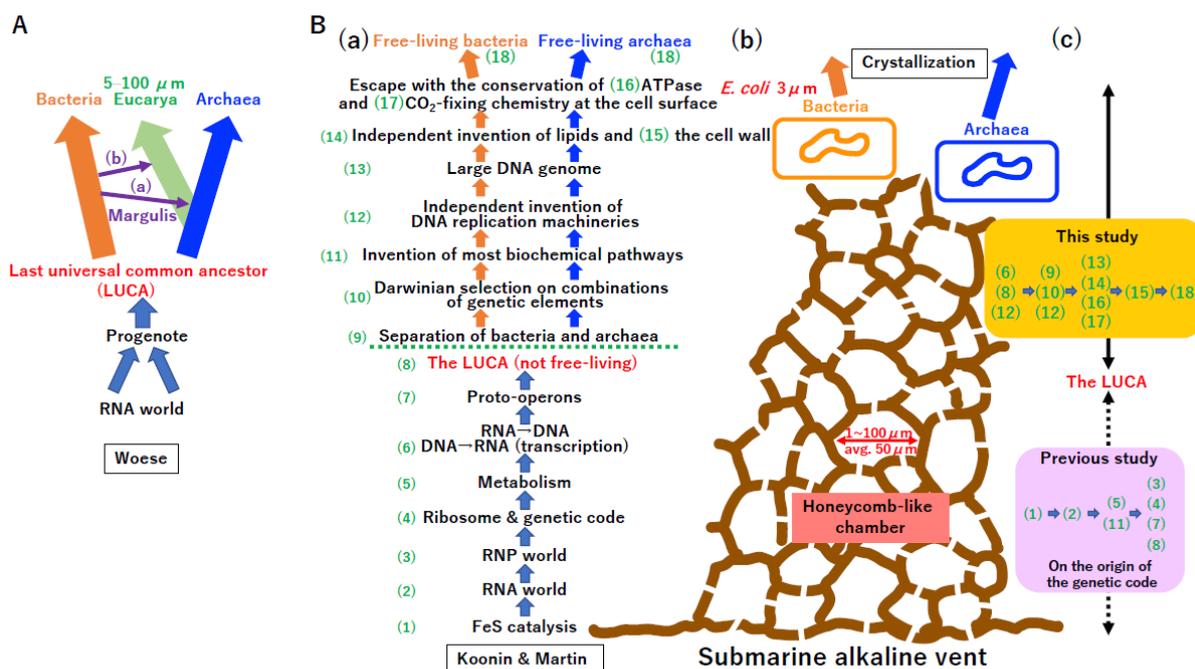


Fig.1

Figure 1. Crystallization of bacteria and archaea. (A) Woese's model [5]. Phylogenetic analysis of rDNA revealed three domains of life, Bacteria, Archaea, and Eucarya [1]. (a) Symbiosis between Archaea and α -Proteobacteria could yield Eucarya. (b) Symbiosis between Eucarya and Cyanobacteria could yield Archaeplastida. The concepts in (a) and (b) were originally proposed by Margulis [3]. Eukaryotic cells range in size from 5 to 100 μm . Woese proposed that Bacteria and Archaea arose from the crystallization of systems in the LUCA. Moreover, the LUCA might be derived from a progenote descended from an RNA world. (B) Submarine alkaline vent model. (a) The emergence of free-living bacteria and archaea at an ancient submarine alkaline vent was described as a stepwise process, (1)–(18) [7]. Koonin and Martin proposed that the LUCA was not a free-living cell. (b) Cartoon of a submarine alkaline vent. Extant vents as well as early vents have numerous honeycomb-like chambers with sizes of 1 to 100 μm (average 50 μm) [7]. *E. coli*, a descendant of a product of crystallization, bacteria, is 3 μm long. (c) Steps proposed in the current study. We have previously proposed a hypothesis for the origin of the genetic code and evolution of the translation system [11]. The order of events described in B(a) differs from that described in this previous study (purple background) [11]. In this study, we proposed evolutionary steps from the LUCA to bacteria and archaea at an ancient submarine alkaline vent (yellow background).

Previously, we proposed a hypothesis to explain the origin of the genetic code and translation system [11]. The postulated scenario from an RNA world to the LUCA (Figure 1B(c)) [11] is quite different from that proposed by Koonin and Martin (Figure 1B(a)). The final step in the emergence of Bacteria and Archaea from the LUCA was called "crystallization", as the stabilization of genetic systems, (Figure 1B(b)) by Woese [5].

In this study, we theoretically examined crystallization processes from the LUCA to extant cells. We hypothesize that crystallization was mainly driven by two physical laws (Eigen's error catastrophe and Bejan's constructal law) [12,13]. Remarkably, the order of steps (Figure 1B(c)) toward crystallization differs from that proposed by Koonin and Martin [7] (Figure 1B(a)). Complete scenarios for crystallization processes are presented.

2. The LUCA

2.1. Double-stranded U-DNA on a single operon

We have previously predicted that a pool of hundreds of single operons associated with both the translation system and extant standard genetic code existed before cells (Supplementary Figure S1) [11]. A small quantity of prebiotic dNTPs was likely essential for the RNA-to-DNA transition (Supplementary Figure S2), facilitating the cooperative emergence of ribonucleotide reductase (RNR) (Figure 2A) [14], primitive reverse transcriptase (RT) (Figure 2B), and primitive DNA-dependent RNA polymerase (DdRp) (Figure 2C). Since DNA is a more stable substance than RNA, antisense DNA is excellent template for multiple rounds of mRNA production (Figure 2D). This powerful driving force (Figure 2D) produced RT (Figure 2F) together with selective dNTP (Figure 2E), leading to T7 type DNA-dependent RNA polymerase (Figure 2G). Of note, RNA-dependent RNA polymerase (RdRp) (Supplementary Figure S1) has an RRM-palm domain (Supplementary Figure S3). Both RT (Figure 2F) and T7 phage DdRp (Figure 2G) also have the RRM-palm domain (Supplementary Figure S3). In addition to the RRM-palm fold, another DdRp derived from Bacteria or Archaea with a double psi-beta barrel domain (Supplementary Figure S3) emerged (Figure 2H) [15].

RNA digestion of RNA/DNA duplexes by RNaseH, an essential component of the retrovirus life cycle [16], could lead to single-stranded DNA (Figure 2I). To avoid the terminal replication problem, protein-primed type DNA synthesis could have emerged by utilizing a single-stranded DNA binding protein (SSB or RPA type) [17] and DNA-dependent DNA polymerases (DdDps) (Figure 2J(a)). The DdDp PolB (Figure 2J(a)) could be directly derived from T7 type DdRp (Figure 2G), primitive RdRp (Figure 2C), and RdRp (Figure S1). To simplify enzyme lineages in evolution, we marked these polymerases as PolB (Figure 2, Supplementary Figure S3). By contrast, DdDp (PolD) (Figure 2J(a)) could be derived from cellular DdRp (Figure 2H), marked PolD. Remarkably, DdDp with another fold (Supplementary Figure S3) could have emerged simultaneously, leading to PolA, PolC, and PolY (Figure 2J(a)). Then, dsDNA might emerge in the LUCA (Figure 2J(b)).

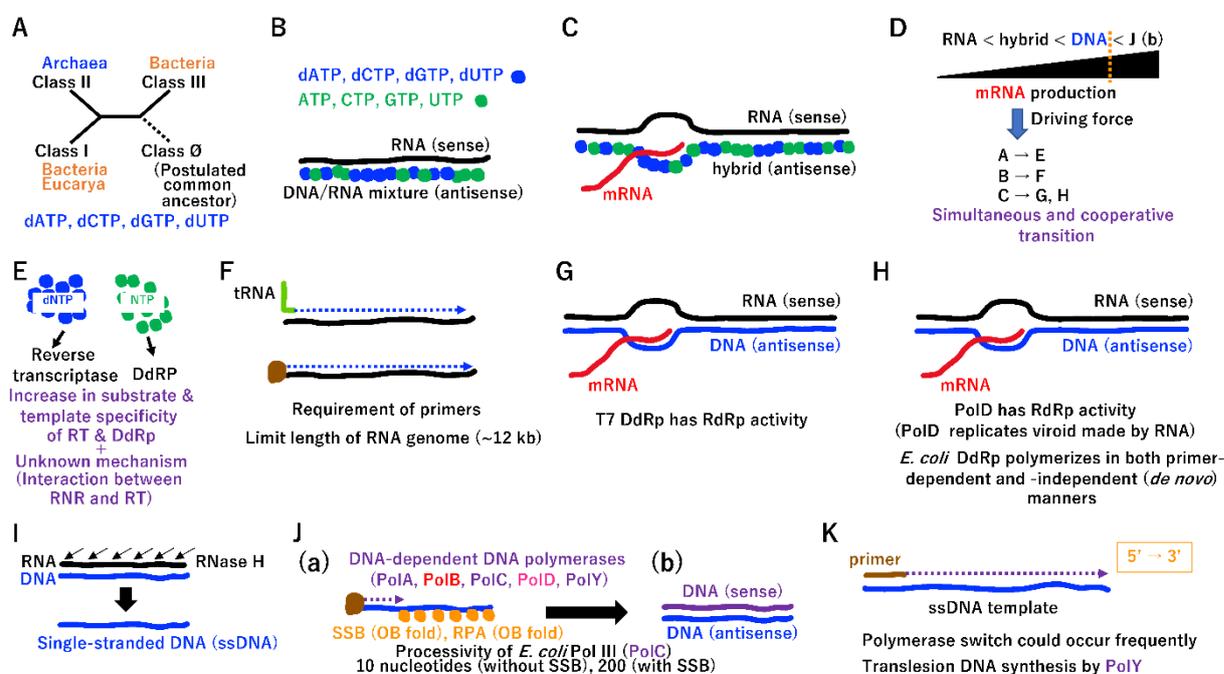


Fig.2

Figure 2. RNA to U-DNA transition. (A) Ribonucleotide reductase (RNR). Schematic phylogenetic tree of RNR classes I–III. The postulated common ancestor of RNR was reconstructed as class Ø. dATP,

dCTP, dGTP, and dUTP (not dTTP) (blue characters) could be produced at the chamber of the vent, mediated by RNR. (B) Primitive reverse transcription by RdRp (PolB). Using template sense RNA and mixed substrates of dNTP (blue) and NTP (green), primitive reverse transcriptase, a derivative of RNA-dependent RNA polymerase, could polymerize the DNA/RNA mixed antisense molecule. Blue and green dots represent incorporated dNMP and NMP, respectively. Primitive reverse transcriptase, which should have the RRM-palm domain, as marked by PolB, has 5' to 3' directional polymerization activity. (C) Primitive hybrid strand-dependent polymerization by RdRp (PolB). Primitive DNA-dependent RNA polymerase, a derivative of RNA-dependent RNA polymerase, could discriminate DNA/RNA mixed antisense molecules from RNA (sense molecule), preferentially bind the former, and then polymerize mRNA (sense RNA). Primitive DNA-dependent RNA polymerase, which should have the RRM-palm domain, marked PolB, has 5' to 3' directional polymerization activity. (D) Driving force for the transition from RNA to U-DNA. If the physical stability of the antisense strand in nucleic acid duplexes was beneficial for mRNA productivity, the transition from RNA to U-DNA would inevitably occur. < indicates the high stability of the antisense strand in a variety of duplexes. (E) Selective supply. Among mixtures of dNTP and NTP, both template and substrate specificities could increase gradually in primitive RT (B) and primitive DdRP (C) by the driving force (D). Furthermore, an unknown mechanism could selectively supply corresponding substrates for primitive RT and primitive DdRP. Finally, antisense nucleotides could transform from RNA to U-DNA. (F) Extant reverse transcriptase (RT) (PolB). All extant RT have the RRM-palm domain (marked PolB) and possess 5' to 3' activity. Although most RTs require tRNA as a primer [136], protein-primed RTs exist [137]. RT is mainly encoded by retroviruses with a genome size of less than 12 kbp. (G) T7 type DdRp (PolB). T7 DdRp has the RRM-palm domain (marked as PolB) and moves in the 5' to 3' direction. In addition to antisense DNA, T7 DdRp can synthesize RNA using an RNA template [138]. (H) Bacterial and archaeal DdRps (PolD). Bacterial and archaeal DdRps have a double phi-beta barrel domain (marked as PolD) and polymerize in the 5' to 3' direction. Since PolD-type DdRp can replicate viroid genomes [124], DdRp uses an RNA template as well as antisense DNA. Utilization of an RNA template by T7 DdRp and bacterial/archaeal DdRP could be reminiscent of primitive DdRp (C). *E. coli* DdRp polymerizes in both primer-dependent and -independent manners [139]. The former activity might be related to the gain of PolD-type DNA-dependent DNA polymerase (DdDp) (J). (I) Generation of single-stranded DNA (ssDNA). Extant retrovirus encodes RNaseH as well as RT. RNase H specifically degrades the RNA of DNA/RNA duplexes [16]. According to the driving force (D), ssDNA could be converted into double-stranded DNA (dsDNA) by DNA-dependent DNA polymerases (J). (J) Multiple DNA-dependent DNA polymerases. (a) ssDNA could be covered with single-stranded DNA binding protein (SSB) or RPA. The OB fold, which is derived from the common ribosomal subunit (Supplemental Figure S1A), is common in both SSB and RPA. Parallel emergence of multiple DdDps, including PolA, PolB, PolC, PolD, and PolY (Supplementary Figure S3). PolB and PolD could be direct descendants of PolB-type DdRp (G) and PolD-type DdRp (H), respectively. It is not clear how the terminal replication problem was solved. A simple solution is the utilization of a protein-primed mechanism, as in extant RT (F). The processivity of *E. coli* replicative DNA polymerase III (Type C polymerase) is only 10 nucleotides. In the presence of SSB, the processivity of PolIII increases to 200 nucleotides. (b) Appearance of the dsDNA single operon. (K) Kornberg's reaction. Low processivity of each DdDP could lead to cooperation rather than competition among multiple DdDp(s) to complete DNA synthesis. Thus, a polymerase switch could occur frequently. Since PolY has translesion DNA synthesis activity, the damaged ssDNA template could be converted to dsDNA. Multiple DdDp(s) would share remarkably similar biochemical properties, ssDNA-template dependency, primer-requirement, and 5' to 3' directional polymerization. These polymerase characteristics were first demonstrated by Kornberg using *E. coli* DNA polymerase I (PolA) [18].

Purified *Escherichia coli* replicative Pol III (PolC) has a processivity of only 10 or 200 nucleotides in the absence or presence of SSB, respectively [18]. To complete the DNA replication of a single S10 operon (5.2 kbp) (Supplementary Figure S1B), DdDp should be loaded on ssDNA template, more than 26 (5,200/200) times. Since multiple DdDps, including lesion bypass-type PolY, could function cooperatively in the completion of DNA replication (Figure 2K), they share similar characteristics

depending on the ssDNA template and primer for 5' to 3' directional DNA synthesis, as firstly described by Kornberg [18].

2.2. Discrimination of sense and antisense DNA

Extant T7 RNA polymerase (PolB) and bacterial/archaeal RNA polymerase (PolD) cannot discriminate sense or antisense strands of duplex DNA without promoters (Figure 3A(a)(b)). T7 RNA polymerase directly recognizes the promoter. *E. coli* RNA polymerase recognizes the promoter *via* σ (Figure 3A(c)). TFIIB in both Achaea and Eucarya interacts with the surface of RNA polymerase, which is structurally equivalent to the σ -interacting surface of *E. coli* RNA polymerase [19,20]. TFIIB may have been the primary promoter recognition protein in the ancient world (Figure 3A(c)) [19–21]. Although the internal duplication of TFIIB yielded TFIIB', there was no internal duplication in TBP at that time (Figure 3A(d)) [20]. Extant TBP, the primary site of promoter recognition, has an internal duplication (Figure 3A(e)) [20]. Importantly, various transcription systems produce mRNA equally (Figure 3A(c)(d)(e)), implicating that such weak selection allowed divergence of transcription initiations. Ohta proposed the nearly neutral theory explaining how weak selection shapes rates of evolution and patterns of genetic variation (allele frequency distributions) under different population sizes [22]. Similar to Ohta's concept, the nearly neutral selection on mRNA may explain patterns of molecular variation in these systems.

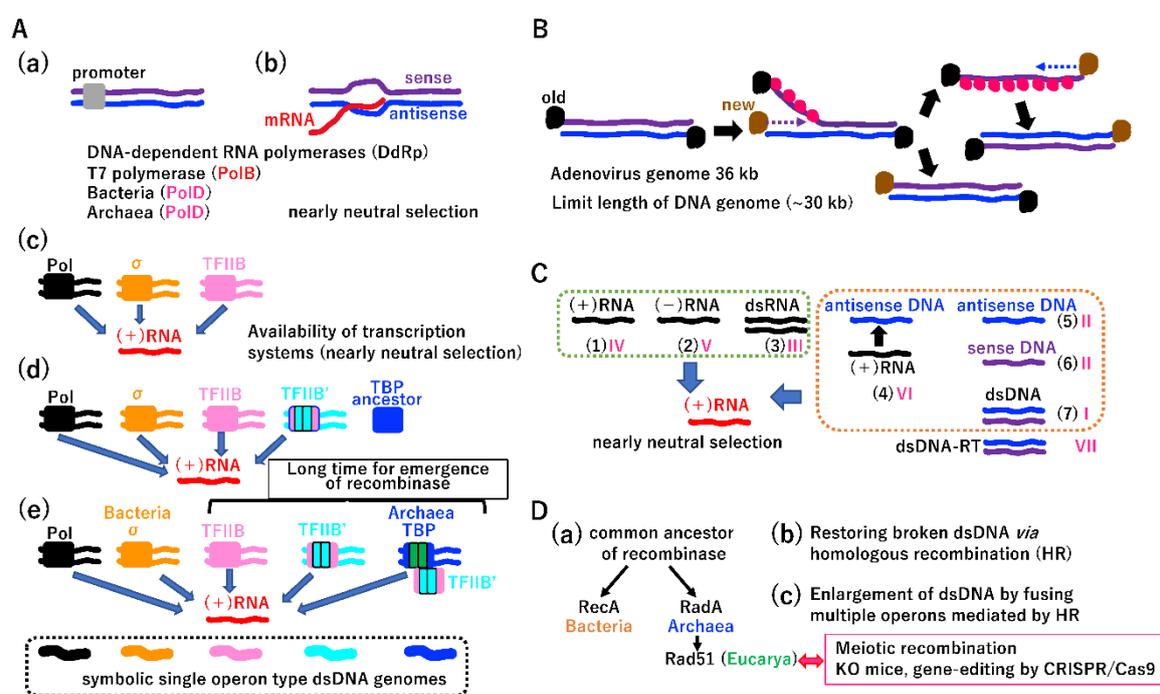


Fig.3

Figure 3. The LUCA—Garden of Baltimore (A) Different promoter recognition systems in the era of single operons. (a) (b) Necessity of promoters. Since sense and antisense ssDNAs in duplex DNA share the same physical properties, specific RNA synthesis on antisense DNA, as a template, required promoter sequences. Extant T7 and bacterial/Archaeal DdRp can specifically transcribe antisense DNA. (c) T7 DdRp directly binds promoters. Bacterial DdRp recognizes the promoter *via* the σ subunit. Although Eucarya (derived from Archaea) DdRp recognizes promoters *via* TBP, a basic transcription factor, it is predicted that TFIIB, another basic transcription factor, initially recognized the promoter in early evolution (see text for details). (d) Internal duplication of TFIIB prior to that of TBP. Both extant TFIIB and TBP underwent internal duplication. A phylogenetic analysis of direct repeats revealed that duplication in TFIIB occurred earlier than that in TBP. Thus, TFIIB' with an internal duplication could recognize the promoter, similar to the original TFIIB. At that time, the role

of TBP in transcription initiation cannot be predicted. (e) The promoter was mainly recognized by TBP. Since extant TBP had an internal duplication and initial promoter recognition factor, at least five different transcription initiation systems may have been present in the LUCA. Importantly, any systems described in (a)–(e) produce the same (+) RNA, consistent with the nearly neutral theory. (B) Strand displacement and unidirectional replication in the single operon era. The same tool kit (protein-priming, ssDNA binding proteins, and multiple DdDp(s)) shown in Figure 2J could also be used to replicate dsDNA. Just like adenovirus replication, protein-primed DNA synthesis can occur on one strand of duplex DNA, coupled with displacement of the other strand. After the replication of one strand is complete, subsequent DNA synthesis occurs on the remaining ssDNA coated with SSB, as a template. Black and brown dots represent old and new terminal proteins, required for protein priming. Pink dots represent SSB. (C) Garden of Baltimore in the single operon era. There were three types of RNA, (1)–(3), three types of DNA, (5)–(7), and ssDNA (4) reverse transcribed from RNA in the LUCA. All nucleic acids directly and/or indirectly produce (+) RNA, consistent with the nearly neutral theory. Nucleic acids (1)–(7) correspond to virus classifications I–VI by the Baltimore definition. Double-stranded DNA mediated by RT, corresponding to Baltimore class VII, is described in Figure 12D. There was a tremendous capacity for multiple DNA polymerases (Figure 2J), multiple transcription initiation systems (A), and seven types of nucleic acids in the single operon era (C), essentially the golden era, at the time of the LUCA. Since most nucleic acids corresponding to the Baltimore definition coexisted, the LUCA could be described as a garden of Baltimore. (D) Evolution of recombinase (triggering the end of the LUCA). (a) Common recombinase emerged in the LUCA. RecA of Bacteria and RadA of Archaea (the latter derived from Eucarya Rad51) are key recombinases for homologous recombination (HR) [140–142]. (b) HR could restore broken dsDNA using homologous sequences of other DNA (Supplemental Figure S4). (c) Fusion of multiple operons in the LUCA. In addition to its repair ability, HR facilitates gene duplication [143] and DNA rearrangements (Supplemental Figure S5), thereby accelerating evolution tremendously [26]. Eucarya Rad51 is the basis for meiotic recombination-mediated biological diversity, and biotechnologies, such as knockouts and CRISPR/Cas9-mediated gene editing.

2.3. Strand displacement and unidirectional replication

Duplex DNA (up to 30 kbp) could be replicated, as follows (Figure 3B). As in extant adenovirus DNA replication, protein-primed 5' to 3' directional replication of one template strand occurs and the other strand is displaced and stabilized by SSB [18,23]. After the replication of one strand is completed, the replication of the other strand occurs (Figure 3B). DNA polymerases and SSB (or RPA) are required for duplex DNA replication, suggesting that DdDp and SSB (RPA) described in Figure 2J were available. In addition to three types of RNA (Figure 3C (1)(2)(3), Supplementary Figure S1) in the progenote, the LUCA had four types of DNA (Figure 2F, 2J, 3B, 3C(4)(5)(6)(7)). Various nucleic acids could directly or indirectly produce mRNA, as predicted by the nearly neutral theory [22]. Nucleic acids (1)–(7) in Figure 3C correspond to types I–VI among VII types in a classification of viruses [24]. Considering the diversification of RNR (Figure 2A), multiple DdRps (Figure 2G, 2H), multiple DdDps and SSBs (RPA) (Figure 2J), multiple transcription initiation systems (Figure 3A), two-types of DNA ligases (Supplementary Figure S4A), and the variety of nucleic acids (Figure 3C), substantial diversification from the LUCA could occur. Thus, the LUCA can be considered the garden of Baltimore [24].

3. End of the LUCA

3.1. Beginning of genetic linkage and vertical inheritance

The balance among single operons encoding various products (Figure 3C) could be rapidly destabilized by an increase in dsDNA due to the gain of the homologous recombination (HR) system, which relies on common recombinases (Figure 3D(a)). Since the maximum genome length of extant RNA viruses is 30 kbp (Supplementary Figure S1C), the superior physical nature of dsDNA enabled the generation of genomes exceeding 30 kbp. The evolution of HR to repair errors during DNA replication (Figure 3D(b)) could have been a long process in the robust LUCA (Figure S4A, S4B). In

addition, two independent DNA ligases [25] could have emerged to repair template ssDNA with single-stranded breaks (Figure S4A(a)). Thus, the basic tool kit for the completion of DNA replication on damaged templates could have been established in the LUCA (Supplementary Figure S4C).

In addition to its repair function, extant HR accelerates a variety of genomic rearrangements, including duplications (Supplementary Figure S5A) [26]. Furthermore, HR greatly accelerates protein evolution (Figure S5B, S5C, S5E). Eukaryotic HR-mediated meiotic recombination prompted species diversification [27]. Moreover, biotechnologies, such as knockouts and gene-editing by CRISPR/Cas9, rely on cellular HR [28,29]. Thus, HR has been a key event in Darwinian evolution since it arose in the LUCA.

Although the *yellow body* phenotype is not biochemically related to the *white eye* phenotype in *Drosophila melanogaster*, the alleles are inherited together due to their linkage on the same chromosome (the same dsDNA molecule) (Figure 4A) [30]. Morgan's genetic linkage could be the first step in the fusion of two single operons by HR. Transcription initiation could be divided into at least five systems (Figure 3A(e)), and homologous transcriptional system fusion could be beneficial over heterologous one (Figure 3B). In the LUCA, genes associated with different phenotypes could be genetically linked, as observed in the extant fruit fly. For instance, the σ genome (bacterial lineage) harbored genes encoding DNA polymerases (Figure 2J), including PolA, PolB, PolC, and PolY, and NAD-dependent DNA ligase (Supplementary Figure S4A(a)). By contrast, the TBP genome (archaeal lineage) encoded the DNA polymerases (Figure 2J) PolB, PolD, and PolY and ATP-dependent DNA ligase (Supplementary Figure S4A(a)). Thus, the difference in transcription initiation easily created the so-called DNA replication divide between Bacteria and Archaea [31]. A network of transcription factors evolved rapidly to regulate biological processes in both σ - and TBP genomes lineages (Supplementary Figure S6). Extinct genomes not found in free-living cells (Figure 4C(a)(c)(d)) could also acquire DNA polymerases and DNA ligases in different combinations. We speculated that initial replication mechanism evolved more than twice in the LUCA.

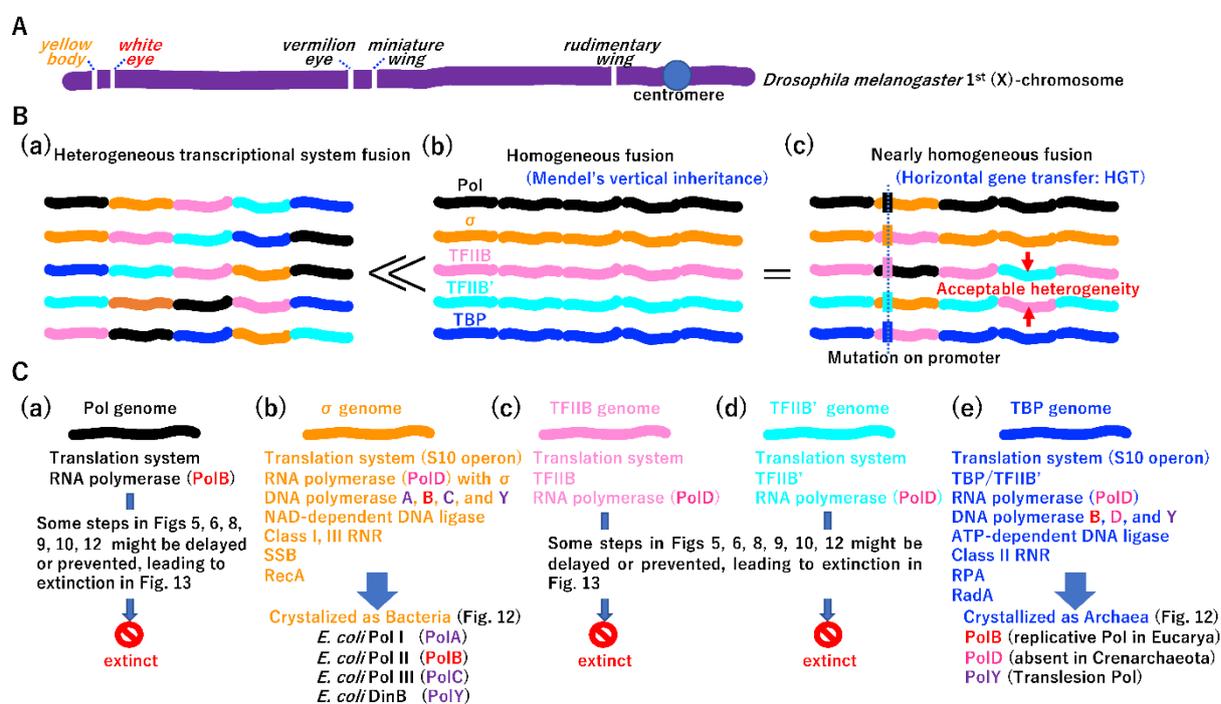


Fig.4

Figure 4. Origin of genetics. (A) Well-known example of genetic linkage. Schematic diagram of the *Drosophila melanogaster* 1st (X)-chromosome. Genes affecting a variety of phenotypes are arranged linearly on a single chromosome, as demonstrated by Morgan et al. [30]. (B) Early genetics in the fused operon era. HR-mediated fusion of five single operons could lead to (a) heterologous transcriptional system fusion, (b) homologous transcriptional system fusion, and (c) nearly homologous

transcriptional system fusion. Each operon is marked by a different color, representing the transcription initiation system (Figure 3A(e)). << Integrity of (b) type fused operons were superior to that of (a) type. = If the promoter of the heterogenous operon in the (c) type genome could be adapted to surrounding operons by mutation, the integrity of the type (c) genome is equivalent to that of the type (b) genome. Red arrows indicate acceptable heterogeneity, because TFIIB' might have similar biochemical activity to that of TFIIB (Figure 3A). Since each operon represents a different biochemistry (different phenotype), different phenotypes are in linkage due to the linear arrangement of operons, as in the fly chromosome (A). Thus, Morgan's genetic linkage and Mendel's vertical inheritance existed since the first operon fusion at the LUCA. The adaptation of the heterogeneous operon in the surrounding genome (B(c)) is the same process as horizontal gene transfer (HGT). Thus, HGT began at the first operon fusion. (C) Persistence of two of five genomes. Although extant cells have either the σ genome (b) or TBP genome (d), each operated by a σ - or TBP-mediated transcription initiation system (Figure 3A(e)), other genomes (a)(c)(e) could have emerged at the LUCA. Each genome contained ribosomal proteins, DdRP, DdDp, DNA ligase, diverse RNR, and ssDNA binding protein, by a variety of combinations. After different biochemistries, such as transcription and replication, were fused, established linkage was vertically inherited to extant cells, including human cells. Thus, *E. coli* (bacteria) has σ factor and replicative PolIII (PolC). Human cells (Eucarya), descendants of Archaea, have TBP and DNA polymerase α , β , γ (all are PolB). The so-called replication divide between bacteria (σ genome) and archaea (TBP genome) began since the first operon fusion at the LUCA. Thus, HR-mediated operon fusion ended the single operon era of the LUCA, an active time in the garden of Baltimore.

3.2. Beginning of horizontal gene transfer (HGT)

The integration of DNA fragments into the genome by HGT under different transcription initiation systems and subsequent mutations in promoter regions was an important process (Figure 4B(c)). Genomes incorporated genes encoding distinct proteins by vertical transmission (Figure 4B(b)), which was initially experimentally demonstrated by Mendel [32]. Over time, beneficial mutations could be transferred among genomes by HGT (Figure 4B(c)).

3.3. Evolution of two different DNA replication machineries

The replication of large DNA molecules took more time than that required for the replication of small DNA. Accordingly, Mullis-type DNA replication (i.e., the simultaneous replication of both strands) emerged (Figure 5A(a)). Furthermore, replication and transcription occurred simultaneously and collision between these processes should occur occasionally (Figure 5A(b)) [33–36]. Topological torsional stress has to be increased, as shown in Figure 5A(a) (b). After the publication of the DNA duplex model [37], Delbrück claimed that it is impossible to separate dsDNA into two ssDNAs due to a topological constraint [38]; however, the σ genome and TBP genome easily resolved this topological issue by the independent evolution of topoisomerases, TopoIIA and TopoIIB, respectively (Figure 5A(c)) [39,40]. TopoIA is common to both lineages (Figure 5A(c)), which may be explained by HGT. The *TopoIB* gene is distributed in various extant Bacteria and Archaea (Figure 5A(c)), suggesting that TopoIB arose in either Bacteria or Archaea, was transferred to other lineages by HGT, and was inherited by Eucarya.

Resolving topological issues enabled independent origins of dnaB helicase and MCM helicase in each genome (Figure 5A(d)). The dnaB helicase moves in the 5' to 3' direction [41]. MCM is translocated in the opposite direction [42]. Of note, dnaB could be derived from the duplication of the recA domain (Figure 5A(d), Figure S5C) [43]. Other helicases, including superfamily I and II helicases, may have arisen [44] and spread to both genomes by HGT (Figure 5A(e)). The rapid progression of the replication fork by topoisomerase and helicase enabled high-speed leading strand DNA synthesis, triggering the evolution of the clamp loader and clamp (Figure 5B(a)(b)) [45,46]. Since both the clamp loader and clamp are sequentially and structurally similar in both genomes [47,48], after they arose in either genome, they could be transferred to the other genome and thereby exert important functions.

High-speed replication fork opening and leading strand synthesis provided space for lagging strand synthesis, triggering a two-fold increase in the total replication speed. Under the same driving force, dnaG primase in the σ genome and PriS/PriL primase in the TBP genome could have arose independently (Figure 5C) [49]. Remarkably, dnaG could be derived from the previously established Topo domain [50] (Figure 5A(c)). The Okazaki fragment [51] is 1,000 or 100 nucleotides in Bacteria or Archaea, respectively (Figure 5C) [52, 53].

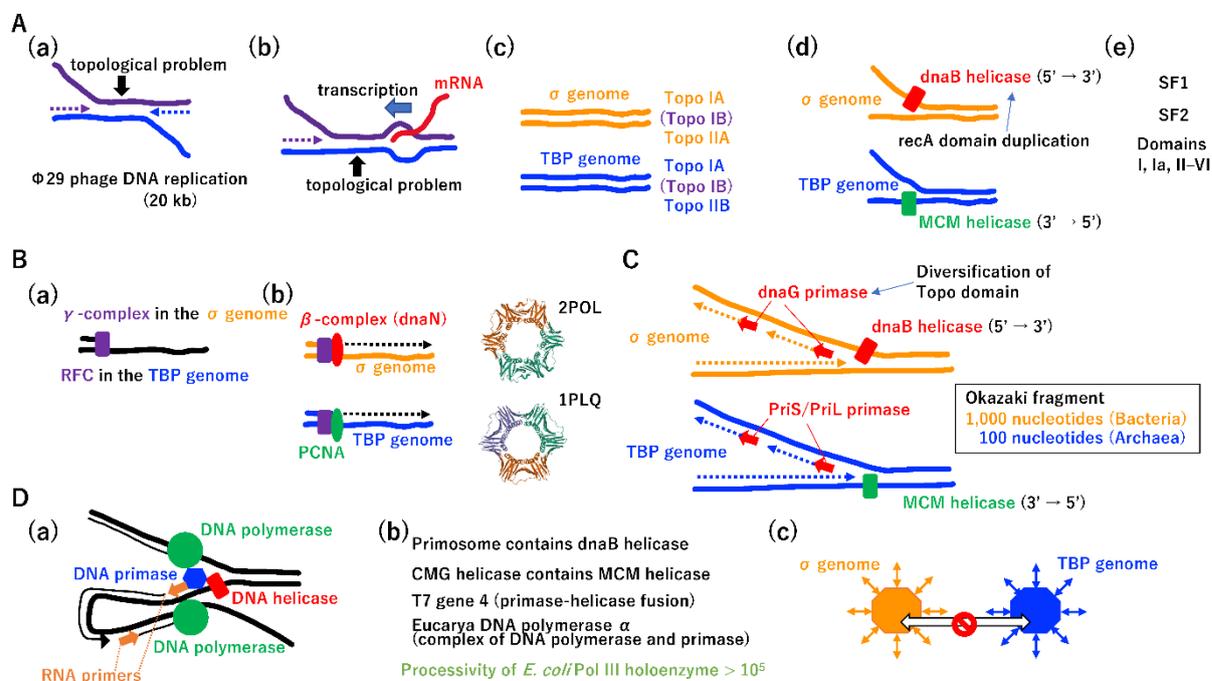


Fig.5

Figure 5. DNA replication divide. (A) Overcoming the topological problem. (a) Mullis' reaction. Strand displacement-type DNA replication of $\Phi 29$ phage (20 kbp) occurs simultaneously from both ends [18], as in PCR invented by Mullis. The speed of the DNA replication of $\Phi 29$ phage is up to twice that of adenovirus DNA replication (Figure 3B). Unwinding from both ends creates torsional stress on the dsDNA region. (b) Collision between transcription and replication. During DNA enlargement, transcription and DNA replication should occur simultaneously. The collision of transcription and replication leads to torsional stress on dsDNA. (c) Solution to Delbrück's claim. Type IIA and IIB topoisomerases were invented in the σ genome and TBP genome, respectively. Topoisomerase IA is common to both. Eucarya DNA topoisomerase IB is found in some Bacteria and some Archaea, suggesting that the development of topoisomerase IB and subsequent HGT across the three kingdoms (Bacteria, Archaea, and Eucarya) occurred. (d) Birth of replicative DNA helicase. Topoisomerase offered the opportunity for the high-speed unwinding of dsDNA by DNA helicase. DnaB helicase and MCM helicase were invented in the σ genome and TBP genome, respectively. DnaB helicase arose from the duplication of the recA domain. (e) Other helicases. In addition to replicative helicases, a helicase superfamily including SF1 and SF2 arose and these loci were transferred between the σ and TBP genomes. (B) High-speed leading strand synthesis. (a) Clamp loader. The clamp loaders γ -complex (σ genome) and RFC (TBP genome) are common. (b) Clamp. The clamps dnaN (σ genome) and PCNA (TBP genome) are structurally common. 2POL (dnaN) and 1PLQ (PCNA) are PDB IDs. (C) Emergence of the lagging strand. Invention of (A) and (B) provides a basis for the development of lagging strand synthesis. DNA primases, dnaG and PriS/PriL, evolved independently in the σ and TBP genomes, respectively. DnaG shares a common domain with topoisomerase, Toprim. Although replication mechanisms are similar in Bacteria and Archaea, the Okazaki fragment in Archaea (including its descendant Eucarya) is shorter than that of Bacteria. (D) Coordination of both strand DNA synthesis. (a) Schematic view of coordinated DNA replication. (b) Mutual interactions among replicative proteins. Leading and lagging strand synthesis coupled with duplex DNA unwinding are

achieved by highly mutual interactions among replicative proteins. DnaB helicase and MCM helicase became components of large protein complexes in the σ and TBP genomes, respectively. In addition, DNA primase and DNA helicase are fused in T7 gene 4. Eucarya DNA polymerase α (PolB) is a complex of DNA polymerase and DNA primase. Although the processivity of *E. coli* PolIII (PolC) is only 10 nucleotides, the processivity of the holoenzyme including PolIII is more than 10^5 nucleotides. (c) Completion of the DNA replication divide. The independent recruitment of replicative proteins (polymerase, helicase, and primase) and highly mutual interactions among replicative proteins led to the complete DNA replication divide between the σ genome (Bacteria) and TBP genome (Archaea) due to the incompatibility of exchanges between equivalent replicative proteins in different lineages.

The coordination of leading and lagging strand synthesis (Figure 5D(a)) [54] could be achieved by highly mutual interactions among replicative proteins (Figure 5D(b)). For instance, dnaB helicase and MCM helicase formed large protein complexes, the primosome and CMG (Cdc45-Mcm-GINS) helicase in Bacteria and Eucarya, respectively [55,56]. In addition to MCM, Cdc45 and GINS [57] are conserved in Eucarya and Archaea. T7 phage gene 4 encodes a primase and helicase fusion protein [58]. Mouse DNA polymerase α is a polymerase and primase complex [59]. Although *E. coli* PolIII has a processivity of only 10 nucleotides (Figure 2J), that of PolIII holoenzyme is more than 10^5 nucleotides [18]. Thus, the equivalent replication components of the σ genome and TBP genome were no longer exchangeable, leading to a complete DNA replication divide between Bacteria and Archaea (Figure 5D(c)) [31]. Further evolution of DNA replication machineries in the σ genome and TBP genome, based on Jacob's replicon theory [60], is summarized in Supplementary Figure S7.

3.4. Necessity of DNA repair for the increase in DNA genome size

According to Eigen's error-catastrophe theory (Figure 6A) [12], it is impossible for the DNA genome to increase in size without the evolution of a DNA repair system. Translesion polymerase (PolY) (Figure 2K, 6B(a)) [61], SSB repair by DNA ligases (Supplementary Figure S4A), and recombinases (Figure 3D, 6B(a)) may have been established in the LUCA (before genome fusion). During genome enlargement, proof-reading of DNA polymerases [18] and a mismatch repair system could have arisen independently in each genome (Figure 6B(a)) [62–64]. Enzymes involved in base excision repair (BER) [65–67] are largely shared between Bacteria and Archaea, suggesting that genes encoding the enzymes involved in BER in either the σ genome or TBP genome are easily transferred into the other genome by HGT (Figure 4B(c)) [68–70].

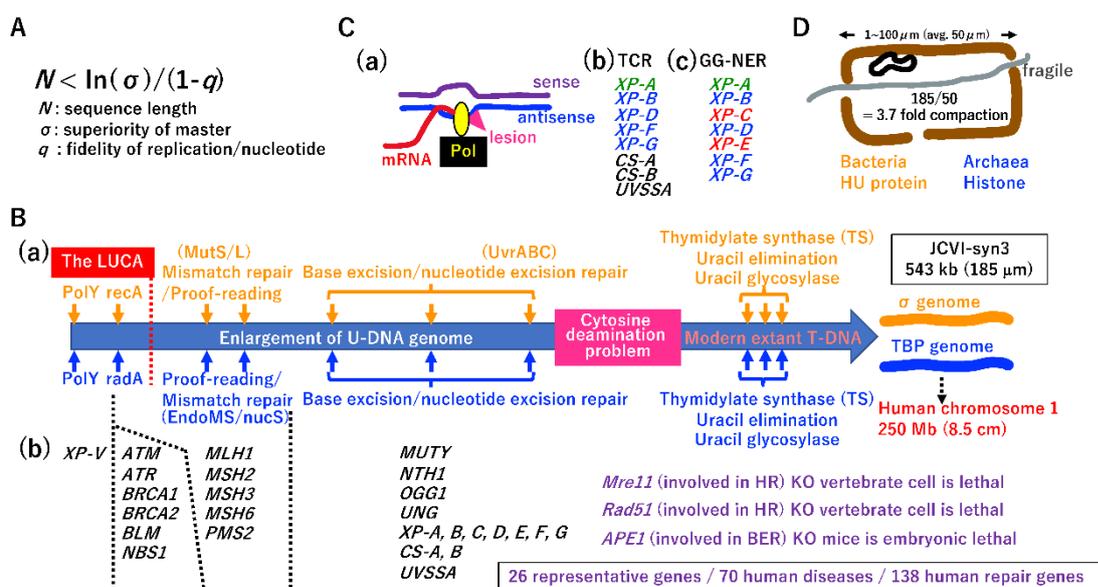


Fig.6

Figure 6. Coevolution of DNA enlargement and the DNA repair system. (A) Eigen's error catastrophe. (B) (a) Schematic diagram of a proposed mechanism for the evolution of DNA repair. As described in Figures 2J and 3D, translesion synthesis (PolY) and homologous recombination (*recA*, *radA*) could be established in the LUCA. Except for base excision repair, DNA repair mechanisms might have evolved independently in the σ or TBP genomes. Toward the thymine-based DNA genome (T-DNA), the cytosine deamination problem should be resolved (see Supplemental Figure S8 for details). Although DNA can increase to up to 8.5 cm (corresponding to human chromosome 1), 185- μ m-long T-DNA was expected based on the genome size of the synthetic minimum cell JCVI-syn3. (b) Human repair system. A lack of either *Mre11* or *Rad51*, both of which are involved in homologous recombination repair, is lethal in vertebrate cells [144,145]. A lack of *Ape1* in mice, a component of BER, leads to embryonic lethality [146]. In addition, 70 of 138 DNA repair-related genes are associated with hereditary diseases in humans [72]. Among these 70 genes, 26 representative genes in the DNA repair pathway of B(a) are listed. (C) Hanawalt's transcription-coupled repair. (a) Hanawalt first recognized the existence of transcription-coupled DNA repair (TCR). (b) In human cells, XP-B, -D, -F, and -G, all of which are conserved in Archaea, are required for TCR. Since there is no counterpart to the *E. coli* UvrABC system in Archaea, TCR-NER mainly acts in Archaea. (c) Global genome-NER. In human cells, GG-NER could arise from XP-A, -C, and -E, all of which do not exist in Archaea. (D) DNA compaction and protection. Elongated and naked dsDNA is physically fragile. Since the predicted average size of chambers in the alkaline vent is 50 μ m, the minimum genome size of 185 μ m should be compacted and protected by dsDNA binding proteins. HU protein and histone evolved independently in each genome.

Although *E. coli* has an efficient nucleotide excision repair (NER) system mediated by UvrABC [71], an NER enzyme is lacking in most species of Archaea. This point will be discussed later (see Figure 6C). Among 138 human DNA repair genes, 70 genes are associated with hereditary diseases [72]. Among these 70 genes, 26 representative genes are listed in Figure 6B(b), such as *XP-V* [73], *NBS* [74], and *XP-A* [75]. Thus, Eigen's error-catastrophe, a physical law, has dominantly governed biological processes in all organisms, including humans, since the LUCA era.

3.5. Transition from U-DNA to T-DNA

Cytosine often changes into uracil *via* deamination, leading to mutations during subsequent DNA replication [18]. Extant living cells resolve this problem using T-DNA, rather than U-DNA. This transition simultaneously requires the invention of thymidylate synthetase (TS), uracil elimination by dUMPase, and removal of mis-incorporated uracil by glycosylase (Figure 6B(a)) [18,76]. One scenario for the transition from U-DNA to T-DNA is described in Supplementary Figure S8.

3.6. Transcription-coupled NER

Hanawalt's group first proposed transcription-coupled repair [77]. RNA polymerase in the elongation step is an excellent sensor of damaged antisense DNA (Figure 6C(a)), and the preferential repair of antisense DNA is highly beneficial for restoring normal transcription. In Eucarya, transcription-coupled NER [78] is operated by various proteins, including XP-B, XP-D, XP-F, and XP-G, all of which are conserved in Archaea (Figure 6C(b)). Thus, TCR-NER might mainly function in Archaea to repair DNA damage. It is possible that global genome (GG)-NER in Eucarya (Figure 6C(c)) [79] evolved from TCR-NER via the Eucarya-specific evolution of XP-A, XP-C, and XP-E.

3.7. Necessity of DNA compaction

After the evolution of DNA replication (Figure 5, Supplemental Figure S7) and DNA repair systems (Figure 6B,C), the DNA genome can increase 8.5 cm in length (Figure 6B(a)). Venter's minimal synthetic cell JCVI-syn3 has a 543 kbp DNA genome [80], with an estimated length of 185 μ m. Since the average size of each chamber in a hydrothermal vent (Figure 1B(b), 6D) is 50 μ m, a 185- μ m-long DNA genome should be compacted 3.7-fold in the chamber. HU and histone evolution in

the σ genome and TBP genome, respectively [81,82], easily solved the compaction problem (Figure 6D).

4. Toward proto-cells

4.1. A lack of essential genes is lethal in extant cells

Giant viruses carrying a 1,259 kb genome (larger than the genome of free-living JCVI-syn3) (Figure 7B(a)) [83,84] are not free-living but are very active in host cells (Figure 7A). The virus factory in Eucarya freely imports and/or exports necessary biomolecules. Thus, an open system is required for the survival of taxa with incomplete genomes, like giant viruses. If extant cells lose a single essential gene, they cannot survive (Figure 7B(b)). For example, the lack of arginine biosynthesis in *Neurospora crassa* leads to cell death [85]. The lethality of *cdc2* mutations in *Schizosaccharomyces pombe* was rescued by the *Homo sapiens* homologue of the *cdc2* gene [86]. As previously reported [87,88], we hypothesized that intermediate proto-cells with an incomplete genome before reaching the free-living state had a permeable membrane to freely exchange necessary biomolecules within the chamber (Figure 7C). A variety of membrane-bound proteins could have evolved. Since the cell size of JCVI-syn3A is 1 μm , the 185 μm DNA genome of JCVI-syn3A should be compacted more than 185-fold (Figure 7B(a), 7D). Such compaction could be achieved by the SMC protein [89], which could have arisen in either the σ genome or TBP genome, followed by HGT into the other genome (Figure 7D). In Eucarya, condensation compacts metaphase chromosomes [90] and cohesion attaches two sister chromatids [91]. The consideration on Figure 7C,D, prompts the question of what type of Darwinian driving force creates proto-cells with permeable membrane and highly compacted DNA genomes.

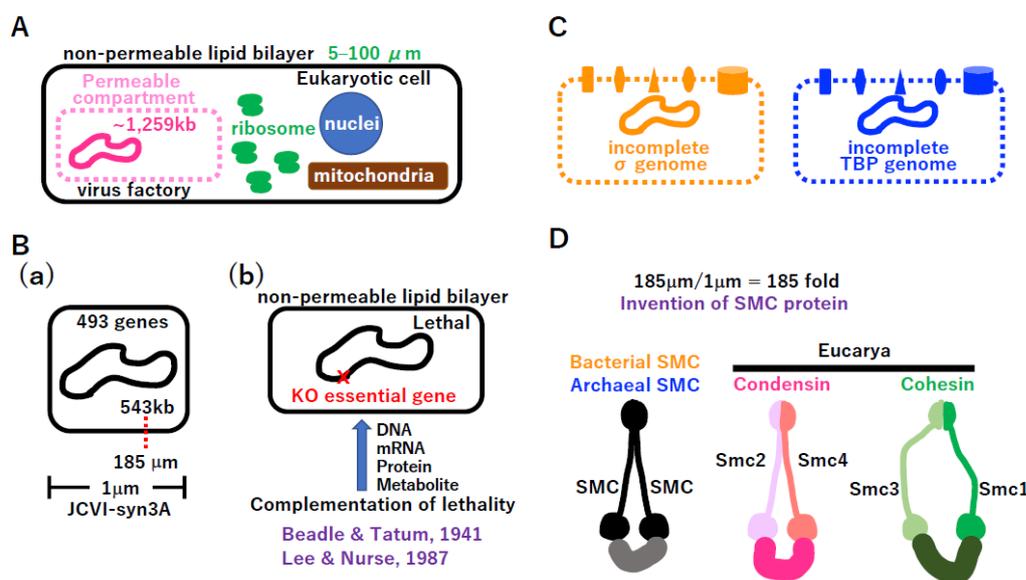


Fig.7

Figure 7. Lethality of a non-permeable membrane surrounding the incomplete genome. (A) Giant virus factory. Although mimivirus has a 1,259 kbp genome, larger than that of JCVI-syn3, it is not free-living. The permeable compartment surrounding the mimivirus genome is called a virus factory. Thus, the permeability of the compartment might be essential for incomplete genomes, such as mimivirus. (B) Incomplete *vs.* complete genome. (a) Synthetic biology (JCVI-syn3A). A compartment surrounded by a non-permeable membrane and cell wall and including a complete genome (543 kbp) leads to free-living JCVI-syn3A cells. (b) Cell lethality. If an essential gene is lacking, cell death will

occur without complementation of the missing component, similar to experimental results of Beadle and Tatum or Lee and Nurse. (C) Membrane-bound proteins. An incomplete genome should be surrounded by a permeable membrane. Essential membrane-bound proteins (transporter, channel, and ATPase) should evolve on the permeable membrane. (D) Evolution of SMC proteins. Expected size of the first Bacteria and Archaea might be equivalent to that of JCVI0syn3, 1 μm . Since JCVI-syn3 has 185- μm -long DNA, compaction of more than 185-fold is necessary. SMC proteins could arise in either the σ or TBP genome, followed by transfer to the other genome by HGT. The descendants of prokaryotic SMC in Eucarya are condensin and cohesin.

4.2. Coacervate formation triggering lipid divide

The reconstructed metabolism of the LUCA is quite similar to that of extant cells [9,10]. A symbolic metabolic map of the LUCA is illustrated in Figure 8A(a). DNA rearrangement and DNA compaction via SMC and transcription factors enabled operons encoding components of metabolic pathways (1), (2), and (3) in Figure 8A(a)(b) to become closer in limited space (Figure 8A(b)), as observed in Eukarya [92]. High concentrations of ribosomes, enzymes, and metabolites could trigger the formation of coacervate [93], also called liquid droplets (Figure 8A(c)) [94–98], leading to efficient biochemistry.

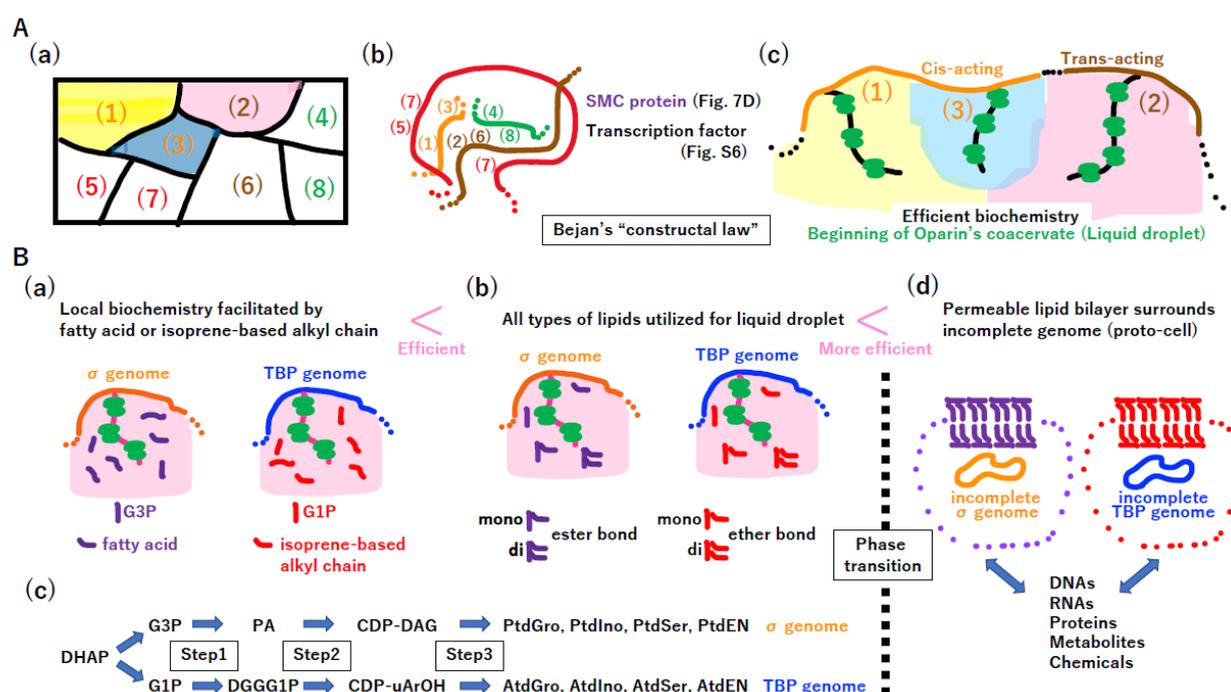


Fig.8

Figure 8. Incomplete large DNA genome surrounded by a permeable membrane. (A) Large DNA genome triggering Oparin's coacervate formation. (a) Schematic representative of metabolic pathways encoded by an ancient uncomplete large genome. Each number (1)-(8) represents a metabolic pathway, such as glycolysis. (b) DNA rearrangement and compaction. Curved line represents dsDNA. The color and number of curved lines represent the coding region of the cluster of the same metabolic pathway. High local concentrations of related enzymes in the metabolic map in limited space, enabled by genome rearrangement, SMC-mediated compaction, and concerted action of transcription factors. (c) Coacervate. The color and number of curved lines are the same as those in (b). Black curved line represents mRNA. Two tandem green dots represent ribosomal large and small subunits. Enlarged view of genes (1), (2), and (3) in A(b). Metabolically related enzymes accumulate in a narrow area, leading to efficient biochemistry due to high concentrations of related metabolites (colored background, corresponding to those in metabolic map (a)). The DNA-mediated hot spot in A(b)(c) might trigger coacervate formation, originally proposed by Oparin. (B) Independent evolution of the synthesis lipid bilayer in a step-by-step manner. (a) Fatty acid or fatty

alcohol. G3P (glycerol-3 phosphate) and G1P (glycerol-1 phosphate) are derived from the metabolic intermediate DHAP (dihydroxyacetonephosphate) in glycolysis (see (c)). Fatty acid or fatty alcohol improve the efficiency of biochemical reactions in the coacervate of A(c). (b) Mono- and di-glyceride of fatty acids and their equivalents in fatty alcohol. All types of lipids could improve the biochemistry in the coacervate. < superior efficiency. (c) Lipid divide. DHAP is common in both genomes. σ -genome; G3P, PA (phosphatidic acid), CDP-DAG (CDP-diacylglycerol), PtdGro (phosphatidyl glycerol), PtdInoP (phosphatidyl inositol), PtdSer (phosphatidyl serine), PtdEN (phosphatidyl ethanolamine), TBP genome; G1P (glycerol-1-phosphate), DGGG-1-P (digeranylgeranyl G1P), CDP-uArOH (CDP-unsaturated archaeol), AtdGro (archaetidyl glycerol), AtdIno (archaetidyl inositol), AtdSer (archaetidyl serine), AtdEN (archaetidyl ethanolamine) [99]. Genes encoding enzymes in step 1 might have arisen independently in the σ and TBP genomes, followed by vertical inheritance (Figure 4). By contrast, genes encoding enzymes in steps 2 and 3 are common in both genomes [99], suggesting that HGT occurred under the same Darwinian driving force (A, B) at the chamber of the vent. Thus, the independent invention of enzymes in step 1 in each genome could easily create the so-called “lipid divide” between bacteria and archaea. (d) Phase transition. < indicates a superior efficiency. The physical nature of di-glyceride of fatty acid or its counterpart in fatty alcohol might surrounding the incomplete DNA genome might lead to the formation of proto-cells with a permeable membrane. The incompleteness of proto-cells should require an exchange system for all types of biomolecules.

The lipid divide between Bacteria (σ genome) and Archaea (TBP genome) [99] can be explained by equal contributions of fatty acids and fatty alcohols to the biochemistry of coacervate (Figure 8B(a)). Moreover, if di-acyl-glycerol is a better biochemical facilitator than is mono-acyl-glycerol, which is a better biochemical facilitator than either glycerol or fatty acids (Figure 8B(b)), then the σ genome could obtain all biochemical steps one-by-one (Figure 8B(c) upper column). A similar scenario (Figure 8B(b)) could occur independently in the TBP genome (Figure 8B(c) lower column).

Simultaneously and independently, a phase transition could happen in both σ and TBP incomplete genomes, leading to the development of a permeable membrane (Figure 8B(d)). Importantly, if the hypotheses described in Figure 8 are correct, the enlargement, rearrangement, and proper compaction of the DNA genome (Figure 5, 6), enabling better biochemistry, will inevitably produce proto-cells in the chamber. Then, both the replication divide and lipid divide between Bacteria and Archaea could be fully completed at this stage (Figure 8B(d)).

4.3. Bejan's “constructal law” in transporter and channel creation

The coevolution of the membrane and membrane-bound proteins was proposed [48]. We hypothesize that such coevolution occurs under Bejan's “constructal law,” a law of thermodynamics [13]. Bejan stated that “For a finite-size system to persist in time (to live), it must evolve in such a way that it provides easier access to the imposed (global) currents that flow through it.” [100]. Under the constructal law, a high concentration of substance A inside a proto-cell moves outward through a permeable membrane; then, to improve this flow, transporter A from the inner to outer space evolved (Figure 9A(a)). Similarly, transporter B and dual transporters C/D emerged. Since the proto-cell produces substance A and consumes substance B, such transporters could facilitate metabolism in proto-cells substantially.

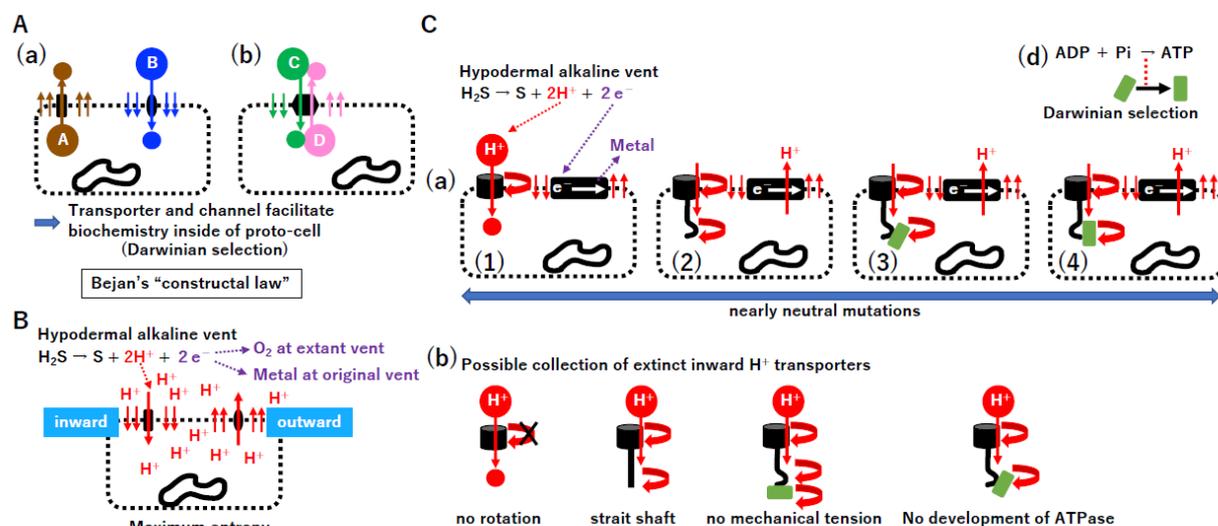


Fig.9

Figure 9. Constructal law. (A) Transporter and channel. Large and small closed circles represent concentrations of corresponding substances. Small thin arrows and large thick arrows represent passive and protein-mediated transport. Dashed rectangles represent a permeable membrane encircling an incomplete genome. (a) Inward and outward flow of substances. Difference in concentrations between the inside and outside of the proto-cell could create a natural flow of substances via the permeable membrane, resulting in equal concentrations of substances inside and outside of the proto-cell. According to Bejan's constructal law, a law of thermodynamics, an outward transporter/channel of substance A and inward transporter/channel of substance B should evolve. (b) Double transporter. Similar logic to that in (a) should result in a double transporter of substances C and D. (B) Inward and outward transporter for protons. Small thin arrows and large thick arrows represent passive and protein-mediated proton transport. The extant submarine alkaline vent provides continuous supplies of protons and electrons. A high concentration of protons outside of the proto-cell could result in inward proton transport by the constructal law. By contrast, an outward flow of protons could raise by increasing maximum entropy, due to the second law of thermodynamics. (C) Birth of ATPase. (a) Schematic overview of steps in the evolution of proton-driven membrane-bound ATPase (1)–(4). (b) Various types of inward proton transporter/channel/rotator could arise simultaneously in each proto-cell. (d) Only extant ATPase persisted because it is extremely useful in producing ATP in the proto-cell and was subjected to powerful Darwinian selection. A possible scenario for the evolution of an electron transport-coupled outward proton transport system is shown in Supplementary Figure S9.

4.4. ATPase and the electron transport system

Protons and electrons are produced from H_2S in the chambers of extant alkaline vents [7,8,101–106]. External protons move into proto-cells through the permeable membrane, accumulate in proto-cells, and move outside, reaching maximum entropy by the second law of thermodynamics (Figure 9B). Under the constructal law, both inward and outward transporters (channel) of protons inevitably emerged (Figure 9B). A variety of inward and outward proton transporters likely arose (Figure 9C, Supplementary Figure S9). Importantly, all proton transporters contributed equally to the improved current under constructal law, as predicted by the nearly neutral theory [22]. However, among these, proton current-driven ATPase (Figure 9C(d)) and outward proton pumps coupled with electron transport (Figure S9B(e)) were positively selected due to the benefits with respect to energy and biochemistry for proto-cells. After the ATPase or electron transport system arose in either the σ or TBP genome, the corresponding gene could be transferred into the other genome and retained due

to its important function. In addition to the transport of small molecules, incomplete genomes also require macromolecule transport mechanisms. As hypothesized previously [107,108], a proposed ancient membrane-based macromolecule exchange system is summarized in Supplemental Figure S10.

4.5. *Omnis cellula a cellula*

The continuous supply of lipids to membrane vesicles *in vitro* inevitably leads to vesicle division due to the physical nature of the lipid bilayer [96]. Thus, under constructal law, FtsZ-mediated cytokinesis [109,110] arose (Figure 10A(a)(b)). Since the concentration of all molecules on one side after cytokinesis is counter to the second law of thermodynamics (Figure 10A(c)), after proto-cell division, equal segregation of chromosomes yields maximum entropy in the production of two daughter proto-cells (Figure 10A(d)(e)). Since pathways for the unlinking of replicated DNA were already established (Supplementary Figure S7C) and membrane-vesicle division is an intrinsic property of a lipid bilayer [96], the development of anchoring DNA on the membrane enabled the equal segregation of duplicated DNA genomes into daughter cells (Figure 10B); subsequently, ParA, ParB, and ParS in the σ genome and SegA and SegB in the TBP genome could emerge (Figure 10B) [111].

Coordination among DNA replication/segregation, membrane growth, anchoring DNA to the membrane, and cytokinesis could establish primitive proto-cell division (Figure 10C). Importantly, the equal division of proto-cells into daughter cells maximizes entropy within the proto-cell (Figure 10C). Moreover, the proliferation of proto-cells in a manner consistent with “*omnis cellula a cellula*” [112] maximizes total entropy in the chamber as well as entropy in each proto-cell (Figure 10D). Thus, the second law of thermodynamics could promote the establishment of extant living cells, following “*omnis cellula a cellula*.”

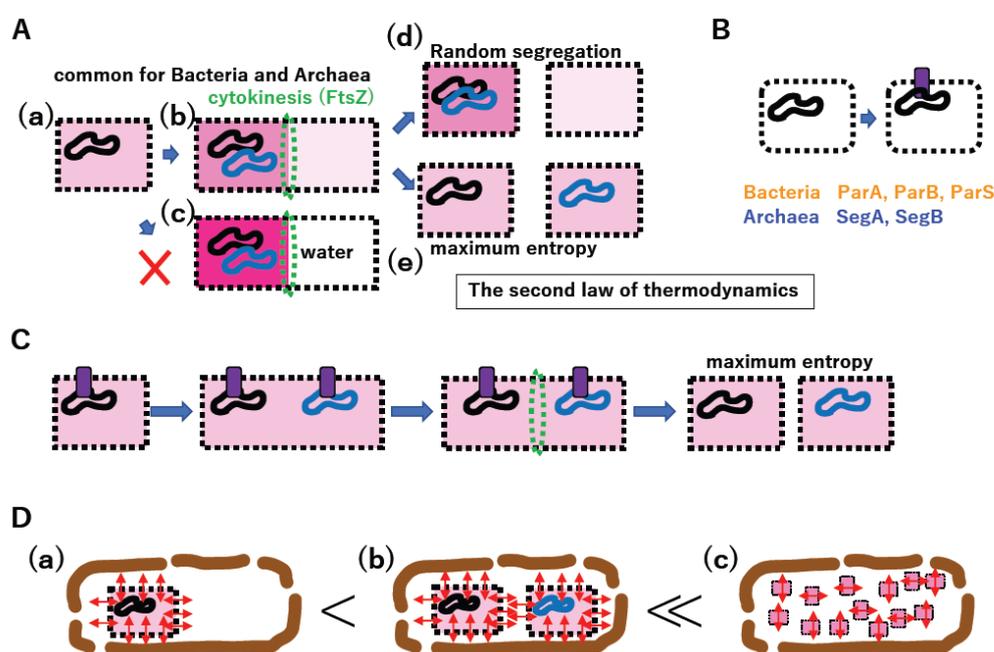


Fig.10

Figure 10. *Omnis cellula a cellula*. (A) Evolution of the lipid bilayer. (a) Proto-cell. (b) The proto-cell contains a duplicated genome and large lipid bilayer. Although cytokinesis of the enlarged proto-cell might occur naturally, FtsZ (green dashed circle), common to both the σ and TBP genome, could arise in accordance with the constructal law. (c) Extreme unequal cytokinesis. Complete unequal cytokinesis including all macromolecules and small molecules (difference in the background pink

color in left and right compartments) is counter to the second law of thermodynamics. (d) Random segregation. If one daughter cell and the other inherit two or no copies of the genome, some unequal components between daughter cells remain. (e) Equal genome segregation. Different from the random process in (d), equal genome segregation could maximize entropy within two daughter cells. Thus, equal proto-cell division could evolve under the second law of thermodynamics as well as the constructal law (b)(c). (B) Evolution of genome anchoring to the lipid bilayer. Purple rectangle represents a membrane-bound protein anchoring genomic DNA. (C) Origin of Virchow's "*Omnis cellula a cellula*" at the proto-cell. After DNA replication, duplicated DNA segregation could arise (Supplementary Figure S7C). An increasing lipid bilayer intrinsically divides A(b)–(d). Thus, concerted effects of duplicated DNA separation, the anchoring of each genome to lipids, and increasing membrane size could yield equal proto-cell division. Importantly, this proto-cell division might maximize entropy in the two daughter cells. (D) "*Omnis cellula a cellula*" in the chamber driven by the second law of thermodynamics. (a) As shown in Figure 9 and Supplementary Figure S9, each proto-cell could maximize inner and outer entropy. (b) Furthermore, equal proto-cell division could increase entropy in two daughter cells. (c) Thus, increasing numbers of proto-cells in a manner consistent with "*omnis cellula a cellula*" could maximize the total entropy in the chamber of the vent. Then, "*omnis cellula a cellula*" could become deeply and genetically engrained in the σ and TBP genomes.

4.6. Punctuated equilibrium by the fusion of different systems

GoULD proposed the concept of punctuated equilibrium based on fossil data [113]. In the history of life on our planet, the symbiosis of Archaea and Bacteria (Figure 1A(a), Supplementary Figure S11A(a)(2)) and the symbiosis of Eucarya and Cyanobacteria (Figure 1A(b), Supplementary Figure S11A(a)(3)) can be described as examples of punctuated equilibrium at the cellular level. Moreover, another example of punctuated equilibrium, the fusion of photosystem I and II, could yield Cyanobacteria (Supplementary Figure S11A(a)(1)), which produced oxygen and changed the environment of the whole planet.

Considering that free-living cells derived from giant viruses did not arise for 2 billion years (Supplementary Figure S11(b)) [83], punctuated equilibrium by system fusion likely occurred in the chamber of the original alkaline vent (Figure 11) during a relatively narrow time window (Supplementary Figure S11A(a)). Since we hypothesize that membrane-vesicle fusion and the split system were present at the chamber (Supplementary Figure S10), a cascade of proto-cell fusion could rapidly create new and complicated proto-cells in a pattern of punctuated equilibrium (Figure 11A). During proto-cell fusion, DNA rearrangement by HR could create new genomes (Figure 11B). Once useful protein machinery arose in either the σ genome or TBP genome, DNA encoding such machinery could be transferred into the other genome by HGT (Figure 11C). For instance, ATPase and SRP are common in Bacteria and Archaea [9,48,114]. By a cascade of proto-cell fusion (Supplementary Figure S12A) associated with genomic rearrangement (Supplementary Figure S12B(a)(b)), proto-cells carrying a nearly complete σ genome or TBP genome may have arisen, after which the permeable membrane could be transformed into a non-permeable membrane (Supplementary Figure S12B(c)).

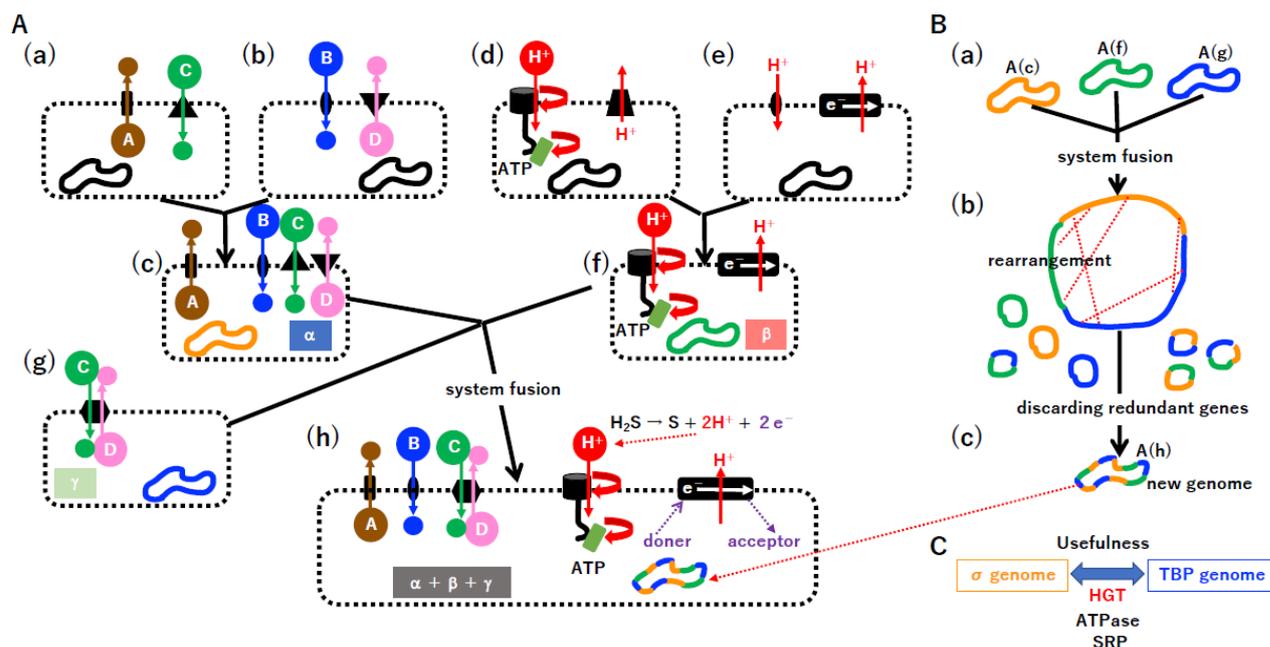


Fig.11

Figure 11. Gould's punctuated equilibrium at the proto-cell level. (A) Punctuated equilibrium. The putative lipid vesicle fusion (including proto-cell fusion) system in the ancient world is described in Supplementary Figure S10. Among proto-cells, proto-cell (c) was derived from the fusion of (a) and (b) and proto-cell (f) was from (d) and (e). The fusion of proto-cells (c), (f), and (g), could yield proto-cell (h), which is selected by Darwinian selection. The ideal combination of membrane-bound proteins and combined biochemistry (α , β , and γ) could be achieved. (B) Genome rearrangement. (a) (b) (c) During the fusion steps of A(c), A(f), and A(g), each genome fusion, rearrangement, and loss created a new genome for proto-cell A(h). (C) HGT. Beneficial genomic elements arising in either the σ or TBP genome could be transferred to the other genome by HGT. Membrane-bound proteins, ATPase, and SRP are common to Bacteria and Archaea, suggesting that HGT occurred in an ancient chamber of the vent.

4.7. Selfish cells trigger the emergence of selfish genes

Primitive living cells could develop active transporters using ATP created by ATPase (Figure 12A) and start to use all energy and resources at the chamber in a selfish manner, as do extant Bacteria and Archaea. Thus, all intermediates described in Figures 2–11, including proto-cells with incomplete genomes, would go extinct (Supplementary Figure S13A, see Figure 13A). Small genes encoding replicases, such as RdRp, RT, and DdDp, could become selfish DNA as in primitive viruses, and an arms race between selfish cells and selfish replicators is possible (Figure 12A,B). Most deadly viruses would be enveloped because the membrane-vesicle-mediated macromolecule transport system supported metabolism at the chamber (Supplementary Figure S10E). Furthermore, primitive viruses could infect proto-cells as well as primitive living cells, just like mimivirus (giant virus) infection by virophage (Supplementary Figure S13B) [115]. Thus, proto-cells should be affected by a shortage of energy and resources as well as viral infection, leading to rapid extinction at the original vent.

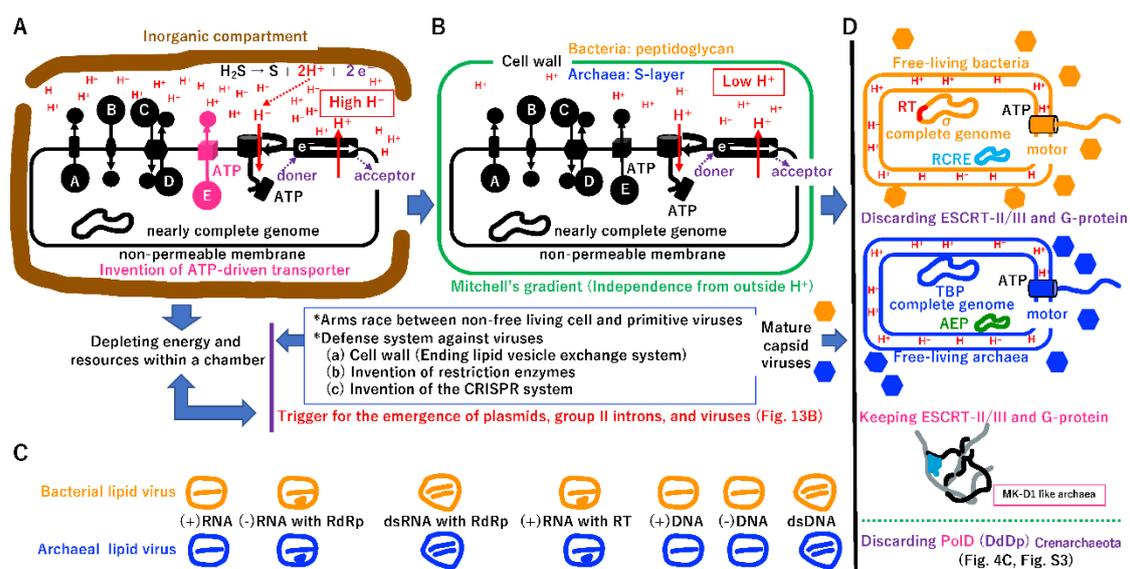


Fig.12

Figure 12. Crystallization of systems in Bacteria, Archaea, viruses, and mobile elements (A) Living cell with a nearly complete genome and non-permeable membrane. Living cells dependent on the proton supply from the inorganic chamber of the vent could obtain enough ATP to develop an ATP-driven transporter for substance E. Living cells might start to occupy all areas of the original vent, depleting energy and resources. These events could trigger a mass extinction of intermediates from the primitive cell to proto-cell (Figures 2–11). During the mass extinction, surviving intermediates might transform to proto-viruses (see Figure 13B). One of the most deadly proto-viruses could be made by lipid vesicles (Supplementary Figure S13C). (B) Cell wall. An arms race between proto-cells/non-free-living cells and proto-viruses could strongly drive the formation of a cell wall of non-free-living cells (A). Although a cell wall could prevent the entrance of enveloped viruses (C), an arms race could continue between cells and capsid-based viruses with DNA or RNA [129,147]. The arms race could drive the evolution of restriction enzyme and CRISPR defense systems in cells, and a DNA injection system of capsid viruses. Although cell wall formation could limit the utility of the external proton supply, some cells could produce their own proton gradient and acquire independence from the chamber made by inorganic materials. (C) Possible extinct proto-viruses based on lipids. Seven types of nucleic acids encoding replicases enveloped by a lipid bilayer could exist in an ancient world. (D) Crystallization. Systems in Bacteria and Archaea with their viruses could undergo crystallization simultaneously at the original vent. Motor protein complexes embedded in the cell wall could arise in both lineages [119], leading to free-swimming Bacteria and Archaea. Mobile elements, such as plasmids encoding RCRE (pale blue) and AEP (green) (Supplementary Figure S3), could be crystallized mainly in Bacteria and Archaea, respectively (see Figure 13B). Mobile elements integrated in the bacterial genome could encode RT (red). Since ESCRT-dependent vesicle transport/cell-fusion (Supplementary Figure S10) might not be necessary for both Bacteria and Archaea, the ESCRT system could be abandoned in both lineages. Some Archaea, including MK-D1, could retain the ESCRT system for intracellular vesicle transport. Ancestral Archaea (D: upper Archaea marked in blue) might have had both PolB and PolD DNA polymerases. In *Crenarchaeota*, the duplication of PolB and loss of PolD might have occurred during evolution.

5. Independent time

5.1. Independent cell wall formation

To prevent the entry of enveloped viruses into primitive cells (Figure 12A), a cell wall may have evolved independently in primitive cells containing the σ or TBP genome (Figure 12B) [116]. The cell

wall would be an intrinsically double-edged sword for Bacteria and Archaea. The endless supply of protons from the vent (Figure 12A) would not be accessible. However, the coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic mechanism [117] could be retained in cells with a cell wall (Figure 12B). Additionally, cell fusion (Figure 11) and membrane-mediated transport (Supplementary Figure S10E) would no longer be possible.

By contrast, primitive cells (Figure 12A) would no longer depend on the inorganic chamber of the original vent (Figure 12B). The cell wall could eliminate enveloped viruses (Figure 12C); however, capsid-based viruses would remain [118]. A flagellin motor could be anchored to the solid cell wall [119,120], leading to free-swimming Bacteria and Archaea (Figure 12D).

Genes encoding the postulated ESCRT-mediated membrane-vesicle transport system (Supplementary Figure S10E) would be lost in all Bacteria and most Archaea (Figure 12D) due to lack of necessity. Similarly, an archaeal lineage, *Crenarchaeote*, has two copies of PolB and could discard PolD [121] due to the redundancy of DNA polymerases (Figure 12D, Supplementary Figure S3). Extant slow-growing Archaea (with a doubling time of one month) with an elongated, branched form [122] maintains the ESCRT system for internal vesicle trafficking. Thus, ESCRT in some Archaea might be a molecular fossil remain of an ancient extinct membrane-vesicle transporting system at the chamber of the original vent (Supplementary Figure S10E). Restriction enzymes [123] and the CRISPR/Cas9 system [29] could have evolved during an arms race between capsid-based viruses and their hosts (Figure 12A, 12B, Supplementary Figure S14).

5.2. Why is the genetic code universal?

Free-living Bacteria and Archaea at the original vent would lead to the mass extinction of intermediates other than free-living cells (Figure 13A(a)(b)(c), Supplementary Figure S13). During the depletion of energy and resources at the original vent, viroid-type RNA [124] (Wang 2021) would survive (Figure 13B(a)). Small nucleic acids of any of the seven types summarized in Figure 3C could be synthesized and protected by the single jellyroll (SHR) protein derived from the proto-cell [118] or the membrane (Supplementary Figure S10E). The evolutionary process described in Figure 13B could rapidly create a variety of viruses. Since cell wall formation in primitive cells (Figure 12) could eliminate enveloped viruses, only capsid-based viruses would remain at the original vent.

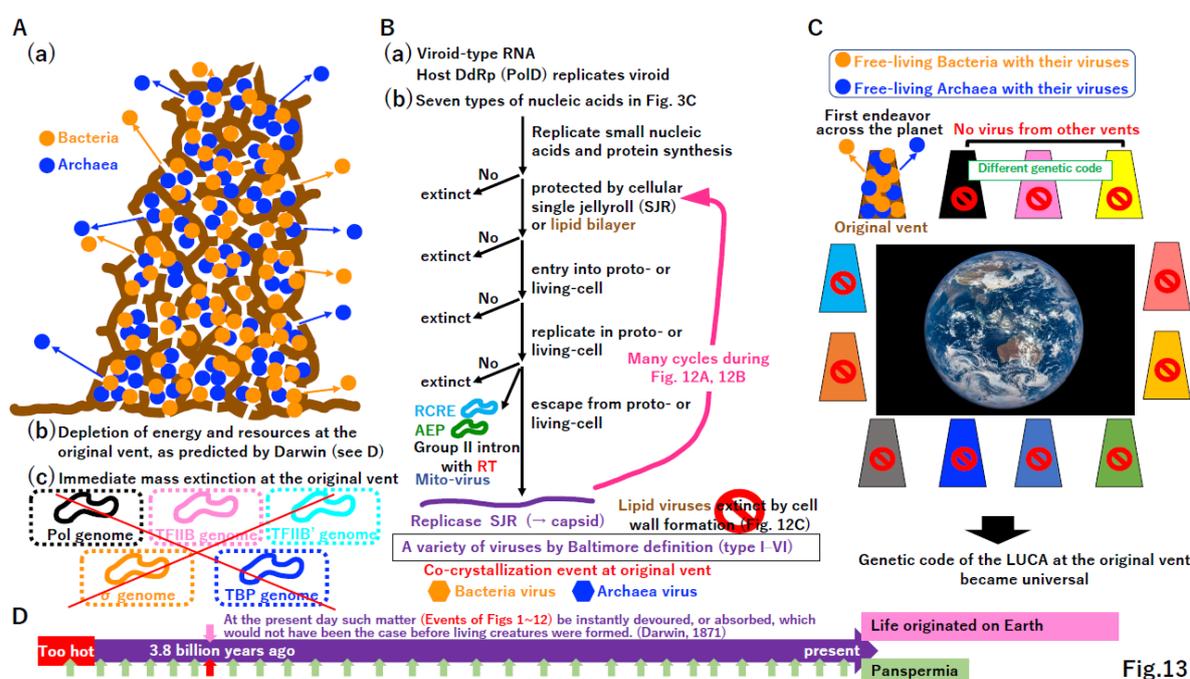


Fig.13

Figure 13. Explanation for the universal standard genetic code. (A) Mass extinction at the original vent. (a) The original alkaline vent might have been occupied by free-living Bacteria and Archaea. (b)

Event (a) might exhaust all energy and resources at the vent. (c) Events (a) and (b) might lead to the mass extinction of most of intermediates described in Figures 2–11. This would reduce the activity of the proto-cell, which is completely dependent on such energy and resources at the vent. (B) Purification of survivors. (a) Extant viroid RNAs can be copied by host DNA-dependent RNA polymerase. Thus, ancient viroid-type RNAs could survive a mass extinction (A). (b) During the exhaustion of energy and resources at the vent, small nucleic acids might be replicated to avoid extinction. The remaining nucleic acids could be protected by single jellyroll protein (SJR), which is derived from the proto-cell and later became a capsid protein, or by a lipid bilayer to prevent extinction. Nucleic acids covered with SJR or a membrane might enter living cells. Nucleic acids in cells could replicate. Thus, nucleic acids should encode replicase. Replicated molecules might remain in cells, leading to mobile elements. The extant mobile elements encode RCRE or RT in Bacteria and AEP in Archaea (Supplementary Figure S3), as observed in mitovirus, found in some extant mitochondria. Nucleic acids covered with SJR or lipids could escape from cells, leading to proto-viruses. During many infection cycles described in (A)(B), proto-viruses could rapidly evolve to become deadly viruses. Since cell wall formation (Figure 12B), enveloped viruses might be lost due to an inability of enter cells. Systems in ancestors of extant viruses could be crystallized at the original vent. (C) Mass extinction on the planet. Bacteria and Archaea with their corresponding viruses might leave the original vent and spread widely, eventually inhabiting all alkaline vents similar to the original one. Since other vents might use different genetic codes, process (B) might not produce viruses at these other vents. Thus, the genetic code at the original vent could become universal. (D) Against panspermia. Living cells with viruses were brought to the Earth 3.8 billion years ago according to the panspermia hypothesis (red arrow) (lower column), raising the question: why has other life (many green arrows) not been brought to the planet? Thus, it is likely that life originated on the Earth through the process described in Figures 2–13, 3.8 billion years ago (upper column).

Plasmids encoding RCRE (PolB) or AEP (PolB) and mobile elements encoding RT (PolB) (Supplementary Figure S3) cannot escape from host cells, just like Mitovirus [125]. We speculated that these plasmids and mobile elements lost capsids or were introduced from a primitive enveloped virus. In the latter case, cell wall formation in host cells could permanently trap RCRE, AEP, and RT (Figure 12D, 13B). The remaining nucleic acids not existing at the time of the LUCA (Baltimore definition class VII) (Figure 3C) [24] appeared in the σ genome as mobile elements (i.e., group II introns) (Figure 12D, 13B).

When Bacteria and Archaea with their viruses migrated from the original vent to the ocean, cells with viruses occupied all alkaline vents on the planet. Although intermediates described in Figures 2–11 could be present at some vents, they would eventually go extinct (Figure 13C), as in the original vent (Figure 13A). Gene flow between taxa and populations at different vents was not possible (Figure 13B), because the genetic code likely varied among vents [11], leading to the standard genetic code of the progenote (Supplementary Figure S1) and the LUCA (Figure 3) from the original vent and its global spread (Figure 13C).

According to Darwin's statement "At the present day such matter (Events in Figures 1–12) be instantly devoured, or absorbed, which would not have been the case before creatures were formed" [126] (Figure 13D), we predict that the crystallization event only occurred once at the original vent 3.8 billion years ago (Figure 13D, upper column). Although an alternative evolutionary theory, panspermia [127] (Crick and Orgel 1973), has been proposed repeatedly [128], it does not effectively explain why Bacteria and Archaea reaching the Earth 3.8 billion years ago grew, while other taxa did not survive (Figure 13D, lower column).

6. Conclusion

Although viruses are believed to be parasites of host cells (Figure 14A(a)), the first free-living cells would likely be selfish parasites of the original vent (Figure 14(b)). Massive numbers of viruses may have been produced from the LUCA [129] at the original vent (Figure 13A(a), 14A(c)). Moreover, there are far more viruses than taxa belonging to Bacteria, Archaea, and Eucarya in the virosphere

(Figure 14A(c)) [130] or cells on the human body (Figure 14A(d)) [131]. Thus, viruses should be included in the definition of life [132,133].

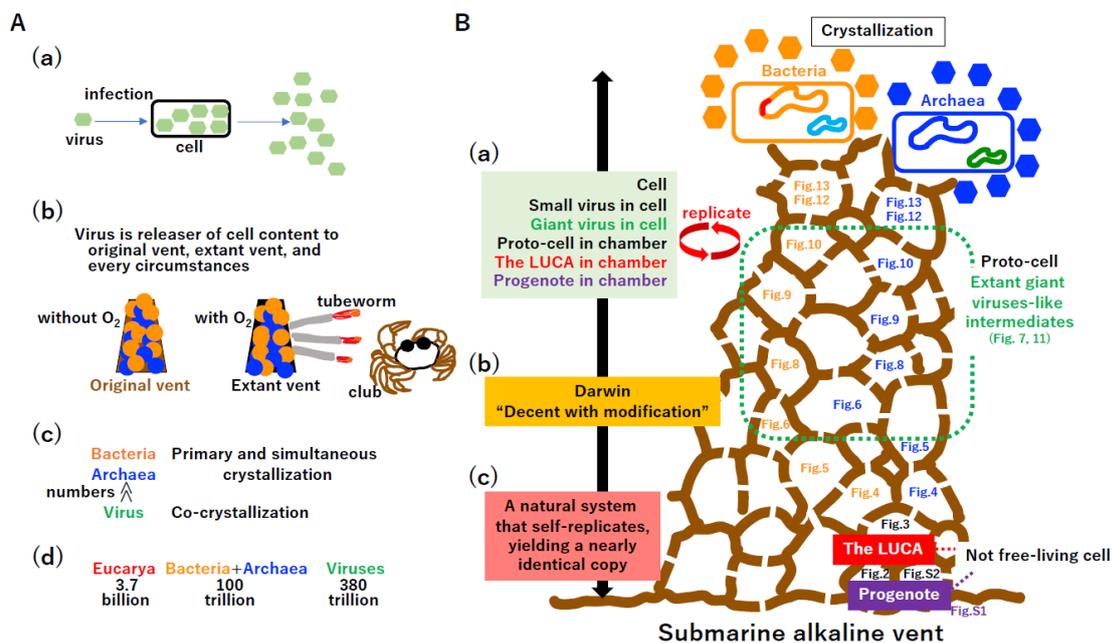


Fig.14

Figure 14. Definition of life. (A) Two opposite perspectives. (a) Viruses are parasites of living cells. (b) Cells are parasites at the original vent and extant vents. Extant vents are occupied by Bacteria and Archaea. Tubeworms and other taxa in the extant vent are completely dependent on such prokaryotes. Since extant viruses release cell contents to the extant vent, ancient viruses could release cell contents to the original vent. (c) Virosphere. Number of viruses exceeds that of cells. (d) Number of viruses in the human body. (B) Definition of life. A schematic diagram of the crystallization of processes described in this study is presented on the right. (a) Replicator. In addition to free-living Bacteria and Archaea, other intermediates, including small and large viruses, proto-cells, the LUCA, and progenote, were not free-living. (b) Darwin's word. (c) Definition of life, including (a) all.

Key innovations for evolution from the progenote to free-living cells with viruses (Figure 14B, right cartoon of the vent) are summarized in Supplementary Figure S14. Some evolutionary steps described in Figures 7–11 and S10–S12 can be tested *in vitro* and/or *in silico* (Supplementary Figure S15). Kornberg stated “*I have never met a dull enzyme*” [134], and key enzymes and proteins forming the extant biological world have been never dull since they emerged at the LUCA (Figure S14).

Although the progenote (Figure S1), the LUCA (Figure 3), and intermediates were not free-living (Figures 4–11) (Figure 14B), all extant life including cells, viruses, and mobile elements are clearly descendants of these non-free-living materials. Cells, viruses, the LUCA, and progenote commonly replicate (Figure 14(a)). Although Darwin articulately described life as “*descent with modification*” (Figure 14(b)) [135], this statement could apply to man-made artifacts, such as smartphones, personal computers, televisions, and cars. Thus, we propose that life, including cells, viruses, mobile elements, the LUCA, and progenote, can be defined as “a natural system that self-replicates, yielding a nearly identical copy” (Figure 14B(c)).

Supplementary Information: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: All conceptual ideas were discussed by AY and MS. The manuscript was drafted by AY and MS and all authors contributed to subsequent editing. The authors read and approved the final manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data are included in the published article.

Acknowledgments: We thank Ms. M. Seki for illustrating some of the figures. This study was financially supported by the Tohoku Medical and Pharmaceutical University, whose founding spirit is “We will open the gate of truth.” We wish to extend our gratitude to Dr. M. Takayanagi, chairman of the University.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could appear to influence the work reported in this paper.

References

1. Balch, W. E.; Magrum, L. J.; Fox, G. E.; Wolfe, R. S.; Woese, C. R. An ancient divergence among the bacteria. *J. Mol. Evol.* **1977**, *9*, 305–311.
2. Woese, C. R. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA.* **2000**, *97*, 8392–8396.
3. Margulis, L. Symbiosis and evolution. *Sci. Am.* **1971**, *225*, 48–57.
4. Gilbert, W. Origin of life: RNA world. *Nature* **1986**, *319*, 618.
5. Woese, C. R. The universal ancestor. *Proc. Natl. Acad. Sci. USA.* **1998**, *95*, 6854–6859.
6. Woese, C. R. On the evolution of cells. *Proc. Natl. Acad. Sci. USA.* **2002**, *99*, 8742–8747.
7. Koonin, E. V.; Martin, W. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **2005**, *21*, 647–654.
8. Branscomb, E.; Russell, M. J. Why the submarine alkaline vent is the most reasonable explanation for the emergence of life. *BioEssays* **2019**, *41*, e1800208.
9. Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **2003**, *1*, 127–136.
10. Tuller, T.; Birin, H.; Gophna, U.; Kupiec, M.; Ruppim, E. Reconstructing ancestral gene content by coevolution. *Genome Res.* **2010**, *20*, 122–132.
11. Seki, M. On the origin of the genetic code. *Genes Genet. Syst.* **2023**, *98*, 9–24.
12. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523.
13. Bejan, A.; Lorente, S. The constructal law and the evolution of design in nature. *Phys. Life Rev.* **2011**, *8*, 209–240.
14. Burnim, A. A.; Spence, M. A.; Xu, D.; Jackson, C. J.; Ando, N. Comprehensive phylogenetic analysis of the ribonucleotide reductase family reveals an ancestral clade. *Elife* **2022**, *11*, e79790.
15. Koonin, E. V.; Krupovic, M.; Ishino, S.; Ishino, Y. The replication machinery of LUCA: common origin of DNA replication and transcription. *BMC Biol.* **2020**, *18*, 61.
16. Iliina, T. V.; Brosenitsch, T.; Sluis-Cremer, N.; Ishima, R. Retroviral RNase H: Structure, mechanism, and inhibition. *Enzymes* **2021**, *50*, 227–247.
17. Taib, N.; Gribaldo, S.; MacNeill, S. A. Single-stranded DNA-binding proteins in the archaea. *Methods Mol. Biol.* **2021**, *2281*, 23–47.
18. Kornberg, A.; Baker, T. *DNA Replication, Second Edition*. Freeman, New York, 1992.
19. Burton, S. P.; Burton, Z. F. The σ enigma: bacterial σ factors, archaeal TFB and eukaryotic TFIIB are homologs. *Transcription* **2014**, *5*, e967599.
20. Adachi, N.; Senda, T.; Horikoshi, M. Uncovering ancient transcription systems with a novel evolutionary indicator. *Sci. Rep.* **2016**, *6*, 27922.
21. Lei, L.; Burton, Z. F. Early evolution of transcription systems and divergence of Archaea and Bacteria. *Front. Mol. Biosci.* **2021**, *8*, 651134.
22. Ohta, T. Simulation studies on the evolution of amino acid sequences. *J. Mol. Evol.* **1976**, *8*, 1–12.
23. Ikeda, J. E.; Enomoto, T.; Hurwitz, J. Adenoviral protein-primed initiation of DNA chains *in vitro*. *Proc. Natl. Acad. Sci. USA.* **1982**, *79*, 2442–2446.
24. Baltimore, D. Expression of animal virus genomes. *Bacteriol. Rev.* **1971**, *35*, 235–241.
25. Pergolizzi, G.; Wagner, G. K.; Bowater, R. P. Biochemical and structural characterisation of DNA ligases from bacteria and archaea. *Biosci. Rep.* **2016**, *36*, art: e00391.
26. Lobkovsky, A. E.; Wolf, Y. I.; Koonin, E. V. Evolvability of an optimal recombination rate. *Genome Biol. Evol.* **2015**, *8*, 70–77.
27. Henderson, I. R.; Bomblies, K. Evolution and plasticity of genome-wide meiotic recombination rates. *Annu. Rev. Genet.* **2021**, *55*, 23–43.

28. Thomas, K. R.; Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **1987**, *51*, 503–512.
29. Charpentier, E.; Doudna, J. A. Biotechnology: Rewriting a genome. *Nature* **2013**, *495*, 50–51.
30. Morgan, T. H.; Sturtevant, A. H.; Bridge, C. B. The evidence for the linear order of the genes. *Proc. Natl. Acad. Sci. USA*. **1920**, *6*, 162–164.
31. Leipe, D. D.; Aravind, L.; Koonin, E. V. Did DNA replication evolve twice independently? *Nucleic Acids Res.* **1999**, *27*, 3389–3401.
32. Mendel, G. *Experiments in plant hybridization*. Royal Horticultural Society of London, translation (1938), Harvard University Press, Cambridge, MA. 1866.
33. Liu, L. F.; Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA*. **1987**, *84*, 7024–7027.
34. Mirkin, E. V.; Mirkin, S. M. Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell. Biol.* **2005**, *25*, 888–895.
35. Pomerantz, R. T.; O'Donnell, M. What happens when replication and transcription complexes collide? *Cell Cycle* **2010**, *9*, 2537–2543.
36. St Germain, C.; Zhao, H.; Barlow, J. H. Transcription-replication collisions—A series of unfortunate events. *Biomolecules* **2021**, *11*, 1249.
37. Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738.
38. Delbrück, M. On the replication of deoxyribonucleic acid (DNA). *Proc. Natl. Acad. Sci. USA*. **1954**, *40*, 783–788.
39. Wang, J. C. DNA topoisomerases. *Annu. Rev. Biochem.* **1985**, *54*, 665–697.
40. Forterre, P.; Gribaldo, S.; Gabelle, D.; Serre, M. C. Origin and evolution of DNA topoisomerases. *Biochimie* **2007**, *89*, 427–446.
41. LeBowitz, J. H.; McMacken, R. The *Escherichia coli* dnaB replication protein is a DNA helicase. *J. Biol. Chem.* **1986**, *261*, 4738–4748.
42. Ishimi, Y. A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *J. Biol. Chem.* **1997**, *272*, 24508–24513.
43. Leipe, D. D.; Aravind, L.; Grishin, N. V.; Koonin, E. V. The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res.* **2000**, *10*, 5–16.
44. Goralenya, A. E.; Koonin, E. V.; Donchenko, A. P.; Blinov, V. M. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **1989**, *17*, 4713–4730.
45. Maki, S.; Kornberg, A. DNA polymerase III holoenzyme of *Escherichia coli*. II. A novel complex including the gamma subunit essential for processive synthesis. *J. Biol. Chem.* **1988**, *263*, 6555–6560.
46. Tsurimoto, T.; Stillman, B. Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication *in vitro*. *Mol. Cell. Biol.* **1989**, *9*, 609–619.
47. Krishna, T. S.; Kong, X. P.; Gary, S.; Burgers, P. M.; Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **1994**, *79*, 1233–1243.
48. Koonin, E. V. The origins of cellular life. *Antonie Van Leeuwenhoek*. **2014**, *106*, 27–41.
49. Bocquier, A. A.; Liu, L.; Cann, I. K.; Komori, K.; Kohda, D.; Ishino, Y. Archaeal primase: bridging the gap between RNA and DNA polymerases. *Curr. Biol.* **2011**, *11*, 452–456.
50. Aravind, L.; Leipe, D. D.; Koonin, E. V. Toprim-a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **1998**, *26*, 4205–4213.
51. Sugino, A.; Hirose, S.; Okazaki, R. RNA-linked nascent DNA fragments in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. **1972**, *69*, 1863–1867.
52. Matsunaga, F.; Norais, C.; Forterre, P.; Myllykallio, H. Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep.* **2003**, *4*, 154–158.
53. Samson, R. Y.; Bell, S. D. Archaeal chromosome biology. *J. Mol. Microbiol. Biotechnol.* **2014**, *24*, 420–427.
54. McNerney, P.; Johnson, A.; Katz, F.; O'Donnell, M. Characterization of a triple DNA polymerase replisome. *Mol. Cell* **2007**, *27*, 527–538.

55. Masai, H.; Nomura, N.; Kubota, Y.; Arai, K. Roles of phi X174 type primosome- and G4 type primase-dependent primings in initiation of lagging and leading strand syntheses of DNA replication. *J. Biol. Chem.* **1990**, *265*, 15124–15133.
56. Tanaka, S.; Araki, H. Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a010371.
57. Takayama, Y.; Kamimura, Y.; Okawa, M.; Muramatsu, S.; Sugino, A.; Araki, H. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev.* **2003**, *17*, 1153–1165.
58. Hamdan, S. M, Richardson CC. Motors, switches, and contacts in the replisome. *Annu. Rev. Biochem.* **2009**, *78*, 205–243.
59. Suzuki, M.; Enomoto, T.; Masutani, C.; Hanaoka, F.; Yamada, M.; Ui, M. DNA primase-DNA polymerase α assembly from mouse FM3A cells. Purification of constituting enzymes, reconstitution, and analysis of RNA priming as coupled to DNA synthesis. *J. Biol. Chem.* **1989**, *264*, 10065–10071.
60. Jacob, F. On regulation of DNA replication in Bacteria. *Cold Spring Harbor Symposia on Quantttative Biology* **1963**, *28*, 329.
61. Ohmori, H.; Friedberg, E. C.; Fuchs, R. P.; Goodman, M. F.; Hanaoka, F.; Hinkle, D.; Kunkel, T. A.; Lawrence, C. W.; Livneh, Z.; Nohmi, T.; Prakash, L.; Prakash, S.; Todo, T.; Walker, G. C.; Wang, Z.; Woodgate, R. M. The Y-family of DNA polymerases. *Mol. Cell* **2001**, *8*, 7–8.
62. Lu, A. L.; Clark, S.; Modrich, P. Methyl-directed repair of DNA base-pair mismatches *in vitro*. *Proc. Natl. Acad. Sci. USA.* **1983**, *80*, 4639–4643.
63. Nakae, S.; Hijikata, A.; Tsuji, T.; Yonezawa, K.; Kouyama, K. I.; Mayanagi, K.; Ishino, S.; Ishino, Y.; Shirai, T. Structure of the EndoMS-DNA complex as mismatch restriction endonuclease. *Structure* **2016**, *24*, 1960–1971.
64. Cebrián-Sastre, E.; Martín-Blecua, I.; Gullón, S.; Blázquez, J.; Castañeda-García, A. Control of genome stability by EndoMS/NucS-mediated non-canonical mismatch repair. *Cells* **2021**, *10*, 1314.
65. Maki, H.; Sekiguchi, M. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* **1992**, *355*, 273–275.
66. Tajiri, T.; Maki, H.; Sekiguchi, M. Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli*. *Mutat. Res.* **1995**, *336*, 257–267.
67. Takao, M.; Kanno, S.; Shiromoto, T.; Hasegawa, R.; Ide, H.; Ikeda, S.; Sarker, A. H.; Seki, S.; Xing, J. Z.; Le, X. C.; Weinfeld, M.; Kobayashi, K.; Miyazaki, J.; Muijtjens, M.; Hoeijmakers, J. H.; van der Horst, G.; Yasui, A. Novel nuclear and mitochondrial glycosylases revealed by disruption of the mouse Nth1 gene encoding an endonuclease III homolog for repair of thymine glycols. *EMBO J.* **2002**, *21*, 3486–3493.
68. White, M. F.; Allers, T. DNA repair in the archaea - an emerging picture. *FEMS Microbiol. Rev.* **2018**, *42*, 514–526.
69. Marshall, C. J.; Santangelo, T. J. Archaeal DNA repair mechanisms. *Biomolecules* **2020**, *10*, 1472.
70. Pérez-Arnaiz, P.; Dattani, A.; Smith, V.; Allers, T. *Haloferax volcanii* - a model archaeon for studying DNA replication and repair. *Open Biol.* **2020**, *10*, 200293.
71. Sancar, A.; Rupp, W. D. A novel repair enzyme: UVRABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell* **1983**, *33*, 249–260.
72. Scheijen, E. E. M.; Wilson, D. M. 3rd. Genome integrity and neurological disease. *Int. J. Mol. Sci.* **2022**, *23*, 4142.
73. Masutani, C.; Kusumoto, R.; Yamada, A.; Dohmae, N.; Yokoi, M.; Yuasa, M.; Araki, M.; Iwai, S.; Takio, K.; Hanaoka, F. The XPV (*xeroderma pigmentosum* variant) gene encodes human DNA polymerase θ . *Nature* **1999**, *399*, 700–704.
74. Matsuura, S.; Tauchi, H.; Nakamura, A.; Kondo, N.; Sakamoto, S.; Endo, S.; Smeets, D.; Solder, B.; Belohradsky, B. H.; Der Kaloustian, V. M.; Oshimura, M.; Isomura, M.; Nakamura, Y.; Komatsu, K. Positional cloning of the gene for *Nijmegen* breakage syndrome. *Nat. Genet.* **1998**, *19*, 179–181.
75. Tanaka, K.; Miura, N.; Satokata, I.; Miyamoto, I.; Yoshida, M. C.; Satoh, Y.; Kondo, S.; Yasui, A.; Okayama, H.; Okada, Y. Analysis of a human DNA excision repair gene involved in group A *xeroderma pigmentosum* and containing a zinc-finger domain. *Nature* **1990**, *348*, 73–76.
76. Lindahl, T. An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl. Acad. Sci. USA.* **1974**, *71*, 3649–3653.

77. Madhani, H. D.; Bohr, V. A.; Hanawalt, P. C. Differential DNA repair in transcriptionally active and inactive proto-oncogenes: *c-abl* and *c-mos*. *Cell* **1986**, *45*, 417–423.
78. Zhang, X.; Horibata, K.; Saijo, M.; Ishigami, C.; Ukai, A.; Kanno, S.; Tahara, H.; Neilan, E. G.; Honma, M.; Nohmi, T.; Yasui, A.; Tanaka, K. Mutations in *UVSSA* cause UV-sensitive syndrome and destabilize ERCC6 in transcription-coupled DNA repair. *Nat. Genet.* **2012**, *44*, 593–597.
79. Kim, J.; Li, C. L.; Chen, X.; Cui, Y.; Golebiowski, F. M.; Wang, H.; Hanaoka, F.; Sugawara, K.; Yang, W. Lesion recognition by XPC, TFIIH and XPA in DNA excision repair. *Nature* **2023**, *617*, 170–175.
80. Hutchison, C. A. 3rd; Chuang, R. Y.; Noskov, V. N.; Assad-Garcia, N.; Deerinck, T. J.; Ellisman, M. H.; Gill, J.; Kannan, K.; Karas, B. J.; Ma, L.; Pelletier, J. F.; Qi, Z. Q.; Richter, R. A.; Strychalski, E. A.; Sun, L.; Suzuki, Y.; Tsvetanova, B.; Wise, K. S.; Smith, H. O.; Glass, J. I.; Merryman, C.; Gibson, D. G.; Venter, J. C. Design and synthesis of a minimal bacterial genome. *Science* **2016**, *351*, aad6253.
81. Macvanin, M.; Adhya, S. Architectural organization in *E. coli* nucleoid. *Biochim. Biophys. Acta.* **2012**, *1819*, 830–835.
82. Mattioli, F.; Bhattacharyya, S.; Dyer, P. N.; White, A. E.; Sandman, K.; Burkhart, B. W. Byrne, K. R.; Lee, T.; Ahn, N. G.; Santangelo, T. J.; Reeve, J. N.; Luger, K. Structure of histone-based chromatin in Archaea. *Science* **2017**, *357*, 609–612.
83. Koonin, E. V.; Yutin, N. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv. Virus Res.* **2019**, *103*, 167–202.
84. Queiroz, V. F.; Rodrigues, R. A. L.; de Miranda Boratto, P. V.; La Scola, B.; Andreani, J.; Abrahão, J. S. Amoebae: Hiding in plain sight: unappreciated hosts for the very large viruses. *Annu. Rev. Virol.* **2022**, *9*, 79–98.
85. Beadle, G. W.; Tatum, E. L. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA.* **1941**, *27*, 499–506.
86. Lee, M. G.; Nurse, P. Complementation used to clone a human homologue of the fission yeast cell cycle control gene *cdc2*. *Nature* **1987**, *327*, 31–35.
87. Paula, S.; Volkov, A. G.; Van Hoek, A. N.; Haines, T. H.; Deamer, D. W. Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys. J.* **1996**, *70*, 339–348.
88. Deamer, D. W. Origins of life: How leaky were primitive cells? *Nature* **2008**, *454*, 37–38.
89. Niki, H.; Jaffé, A.; Imamura, R.; Ogura, T.; Hiraga, S. The new gene *mukB* codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of *E. coli*. *EMBO J.* **1991**, *10*, 183–193.
90. Hirano, T.; Mitchison, T. J. A heterodimeric coiled-coil protein required for mitotic chromosome condensation *in vitro*. *Cell* **1994**, *79*, 449–458.
91. Michaelis, C.; Ciosk, R.; Nasmyth, K. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **1997**, *91*, 35–45.
92. Lengronne, A.; Katou, Y.; Mori, S.; Yokobayashi, S.; Kelly, G. P.; Itoh, T.; Watanabe, Y.; Shirahige, K.; Uhlmann, F. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **2004**, *430*, 573–578.
93. Oparin, A. I. *The origin of life*. McMillan, New York, USA. 1938.
94. Gomes, E.; Shorter, J. The molecular language of membraneless organelles. *J. Biol. Chem.* **2019**, *294*, 7115–7127.
95. Deng, N. N. Complex coacervates as artificial membraneless organelles and protocells. *Biomicrofluidics* **2020**, *14*, 051301.
96. Matsuo, M.; Kurihara, K. Proliferating coacervate droplets as the missing link between chemistry and biology in the origins of life. *Nat. Commun.* **2021**, *12*, 5487.
97. Matsuo, M.; Toyota, T.; Suzuki, K.; Sugawara, T. Evolution of proliferative model protocells highly responsive to the environment. *Life (Basel)* **2022**, *12*, 1635.
98. Gao, N.; Mann, S. Membranized coacervate microdroplets: from versatile protocell models to cytomimetic materials. *Acc. Chem. Res.* **2023**, *56*, 297–307.
99. Koga, Y. From promiscuity to the lipid divide: on the evolution of distinct membranes in Archaea and Bacteria. *J. Mol. Evol.* **2014**, *78*, 234–242.
100. Bejan, A. Constructal-theory network of conducting paths for cooling a heat generating volume. *Int. J. Heat Mass Transfer.* **1997**, *40*, 799–816.

101. Nakamura, R.; Takashima, T.; Kato, S.; Takai, K.; Yamamoto, M.; Hashimoto, K. Electrical current generation across a black smoker chimney. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 7692–7694.
102. Yamamoto, M.; Nakamura, R.; Oguri, K.; Kawagucci, S.; Suzuki, K.; Hashimoto, K.; Takai, K. Generation of electricity and illumination by an environmental fuel cell in deep-sea hydrothermal vents. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 10758–10761.
103. Martin, W. F.; Sousa, F. L.; Lane, N. Evolution. Energy at life's origin. *Science* **2014**, *344*, 1092–1093.
104. Ishii, T.; Kawaichi, S.; Nakagawa, H.; Hashimoto, K.; Nakamura, R. From chemolithoautotrophs to electrolithoautotrophs: CO₂ fixation by Fe(II)-oxidizing bacteria coupled with direct uptake of electrons from solid electron sources. *Front. Microbiol.* **2015**, *6*, 994.
105. McGlynn, S. E.; Chadwick, G. L.; Kempes, C. P.; Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **2015**, *526*, 531–535.
106. Lane, N.; Allen, J. F.; Martin, W. How did LUCA make a living? Chemiosmosis in the origin of life. *Bioessays* **2010**, *32*, 271–280.
107. Gill, S.; Catchpole, R.; Forterre, P. Extracellular membrane vesicles in the three domains of life and beyond. *FEMS Microbiol. Rev.* **2019**, *43*, 273–303.
108. Liu, J.; Cvirkaite-Krupovic, V.; Commere, P. H.; Yang, Y.; Zhou, F.; Forterre, P.; Shen, Y.; Krupovic, M. Archaeal extracellular vesicles are produced in an ESCRT-dependent manner and promote gene transfer and nutrient cycling in extreme environments. *ISME J.* **2021**, *15*, 2892–2905.
109. Pende, N.; Sogues, A.; Megrian, D.; Sartori-Rupp, A.; England, P.; Palabikyan, H.; Rittmann, S. K. R.; Graña, M.; Wehenkel, A. M.; Alzari, P. M.; Gribaldo, S. SepF is the FtsZ anchor in archaea, with features of an ancestral cell division system. *Nat. Commun.* **2021**, *12*, 3214.
110. Santana-Molina, C.; Del Saz-Navarro, D.; Devos, D. P. Early origin and evolution of the FtsZ/tubulin protein family. *Front. Microbiol.* **2023**, *13*, 1100249.
111. Badrinarayanan, A.; Le, T. B.; Laub, M. T. Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.* **2015**, *31*, 171–199.
112. Ribatti, D. Rudolf Virchow, the founder of cellular pathology. *Rom. J. Morphol. Embryol.* **2019**, *60*, 1381–1382.
113. Gould, S. J. Punctuated equilibrium in fact and theory. *J. Social Biol. Struct.* **1989**, *12*, 117–136.
114. Mulkidjanian, A. Y.; Makarova, K. S.; Galperin, M. Y.; Koonin, E. V. Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. *Nat. Rev. Microbiol.* **2007**, *5*, 892–899.
115. La Scola, B.; Desnues, C.; Pagnier, I.; Robert, C.; Barrassi, L.; Fournous, G.; Merchat, M.; Suzan-Monti, M.; Forterre, P.; Koonin, E. V.; Raoult, D. The virophage as a unique parasite of the giant mimivirus. *Nature* **2008**, *455*, 100–104.
116. van Wolferen, M.; Pulschen, A. A.; Baum, B.; Gribaldo, S.; Albers, S. V. The cell biology of archaea. *Nat. Microbiol.* **2022**, *7*, 1744–1755.
117. Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature* **1961**, *191*, 144–148.
118. Krupovic, M.; Dolja, V. V.; Koonin, E. V. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* **2019**, *17*, 449–458.
119. Miyata, M.; Robinson, R. C.; Uyeda, T. Q. P.; Fukumori, Y.; Fukushima, S. I.; Haruta, S.; Homma, M.; Inaba, K.; Ito, M.; Kaito, C.; Kato, K.; Kenri, T.; Kinosita, Y.; Kojima, S.; Minamino, T.; Mori, H.; Nakamura, S.; Nakane, D.; Nakayama, K.; Nishiyama, M.; Shibata, S.; Shimabukuro, K.; Tamakoshi, M.; Taoka, A.; Tashiro, Y.; Tulum, I.; Wada, H.; Wakabayashi, K. I. Tree of motility - A proposed history of motility systems in the tree of life. *Genes Cells* **2020**, *25*, 6–21.
120. Matzke, N. J.; Lin, A.; Stone, M.; Baker, M. A. B. Flagellar export apparatus and ATP synthetase: homology evidenced by synteny predating the Last Universal Common Ancestor. *Bioessays* **2021**, *43*, e2100004.
121. Daimon, K.; Ishino, S.; Imai, N.; Nagumo, S.; Yamagami, T.; Matsukawa, H.; Ishino, Y. Two family B DNA polymerases from *Aeropyrum pernix*, based on revised translational frames. *Front. Mol. Biosci.* **2018**, *5*, 37.
122. Imachi, H.; Nobu, M. K.; Nakahara, N.; Morono, Y.; Ogawara, M.; Takaki, Y.; Takano, Y.; Uematsu, K.; Ikuta, T.; Ito, M.; Matsui, Y.; Miyazaki, M.; Murata, K.; Saito, Y.; Sakai, S.; Song, C.; Tasumi, E.; Yamanaka, Y.; Yamaguchi, T.; Kamagata, Y.; Tamaki, H.; Takai, K. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **2020**, *577*, 519–525.
123. Arber, W. Restriction endonucleases. *Angew. Chem. Int. Ed. Engl.* **1978**, *17*, 73–79.
124. Wang, Y. Current view and perspectives in viroid replication. *Curr. Opin. Virol.* **2021**, *47*, 32–37.
125. Hillman, B. I.; Cai, G. The family narnaviridae: simplest of RNA viruses. *Adv. Virus Res.* **2013**, *86*, 149–176.

126. Follmann, H.; Brownson, C. Darwin's warm little pond revisited: from molecules to the origin of life. *Naturwissenschaften* **2009**, *96*, 1265–1292.
127. Crick, F. H. C.; Orgel, L. E. (1973) Directed panspermia. *Icarus* **1973**, *19*, 341–346.
128. Aydinoglu, A. U.; Taskin, Z. Origins of life research: a bibliometric approach. *Orig. Life Evol. Biosph.* **2018**, *48*, 55–71.
129. Krupovic, M.; Dolja, V. V.; Koonin, E. V. The LUCA and its complex virome. *Nat. Rev. Microbiol.* **2020**, *18*, 661–670.
130. Call, L.; Nayfach, S.; Kyrpides, N. C. Illuminating the virosphere through global metagenomics. *Annu. Rev. Biomed. Data. Sci.* **2021**, *4*, 369–391.
131. Liang, G.; Bushman, F. D. The human virome: assembly, composition and host interactions. *Nat. Rev. Microbiol.* **2021**, *19*, 514–527.
132. Forterre, P. Viruses in the 21st century: from the curiosity-driven discovery of giant viruses to new concepts and definition of life. *Clin. Infect Dis.* **2017**, *65*(suppl 1), S74–S79.
133. Harris, H. M. B.; Hillm C. A place for viruses on the tree of life. *Front. Microbiol.* **2021**, *11*, 604048.
134. Kornberg, A. Never a dull enzyme. *Annu. Rev. Biochem.* **1989**, *58*, 1–30.
135. Penny, D. Darwin's theory of descent with modification, versus the biblical tree of life. *PLoS Biol.* **2011**, *9*, e1001096.
136. Mak, J.; Kleiman, L. Primer tRNAs for reverse transcription. *J. Virol.* **1997**, *71*, 8087–8095.
137. Beck, J.; Nassal, M. A Tyr residue in the reverse transcriptase domain can mimic the protein-priming Tyr residue in the terminal protein domain of a hepadnavirus P protein. *J. Virol.* **2011**, *85*, 7742–7753.
138. Jain, N.; Blanch, L. R.; Szymanski, M. R.; Das, R.; Tang, S. K. Y.; Yin, Y. W.; Fire, A. Z. Transcription polymerase-catalyzed emergence of novel RNA replicons. *Science* **2020**, *368*, eaay0688.
139. Skalenko, K. S.; Li, L.; Zhang, Y.; Vvedenskaya, I. O.; Winkelman, J. T.; Cope, A. L. Taylor, D. M.; Shah, P.; Ebright, R. H.; Kinney, J. B.; Zhang, Y.; Nickels, B. E. Promoter-sequence determinants and structural basis of primer-dependent transcription initiation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* **2021**, *118*, e2106388118.
140. Stasiak, A.; Di Capua, E. The helicity of DNA in complexes with recA protein. *Nature* **1982**, *299*, 185–186.
141. Shinohara, A.; Ogawa, H.; Ogawa, T. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **1992**, *69*, 457–470.
142. Haruta, N.; Kurokawa, Y.; Murayama, Y.; Akamatsu, Y.; Unzai, S.; Tsutsui, Y.; Iwasaki, H. The Swi5-Sfr1 complex stimulates Rhp51/Rad51- and Dmc1-mediated DNA strand exchange in vitro. *Nat. Struct. Mol. Biol.* **2006**, *13*, 823–830.
143. Ohno, S.; Wolf, U.; Atkin, N. B. Evolution from fish to mammals by gene duplication. *Hereditas* **1968**, *59*, 169–187.
144. Sonoda, E.; Sasaki, M. S.; Buerstedde, J. M.; Bezzubova, O.; Shinohara, A.; Ogawa, H.; Takata, M.; Yamaguchi-Iwai, Y.; Takeda, S. Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO J.* **1998**, *17*, 598–608.
145. Yamaguchi-Iwai, Y.; Sonoda, E.; Sasaki, M. S.; Morrison, C.; Haraguchi, T.; Hiraoka, Y.; Yamashita, Y. M.; Yagi, T.; Takata, M.; Price, C.; Kakazu, N.; Takeda, S. Mre11 is essential for the maintenance of chromosomal DNA in vertebrate cells. *EMBO J.* **1999**, *18*, 6619–6629.
146. Izumi, T.; Brown, D. B.; Naidu, C. V.; Bhakat, K. K.; Macinnes, M. A.; Saito, H.; Chen, D. J.; Mitra, S. Two essential but distinct functions of the mammalian abasic endonuclease. *Proc. Natl. Acad. Sci. USA.* **2005**, *102*, 5739–5743.
147. Neri, U.; Wolf, Y. I.; Roux, S.; Camargo, A. P.; Lee, B.; Kazlauskas, D.; Chen, I. M.; Ivanova, N.; Zeigler Allen, L.; Paez-Espino, D.; Bryant, D. A.; Bhaya, D.; RNA Virus Discovery Consortium; Krupovic, M.; Dolja, V. V.; Kyrpides, N. C.; Koonin, E. V.; Gophna, U. Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell* **2022**, *185*, 4023–4037.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.