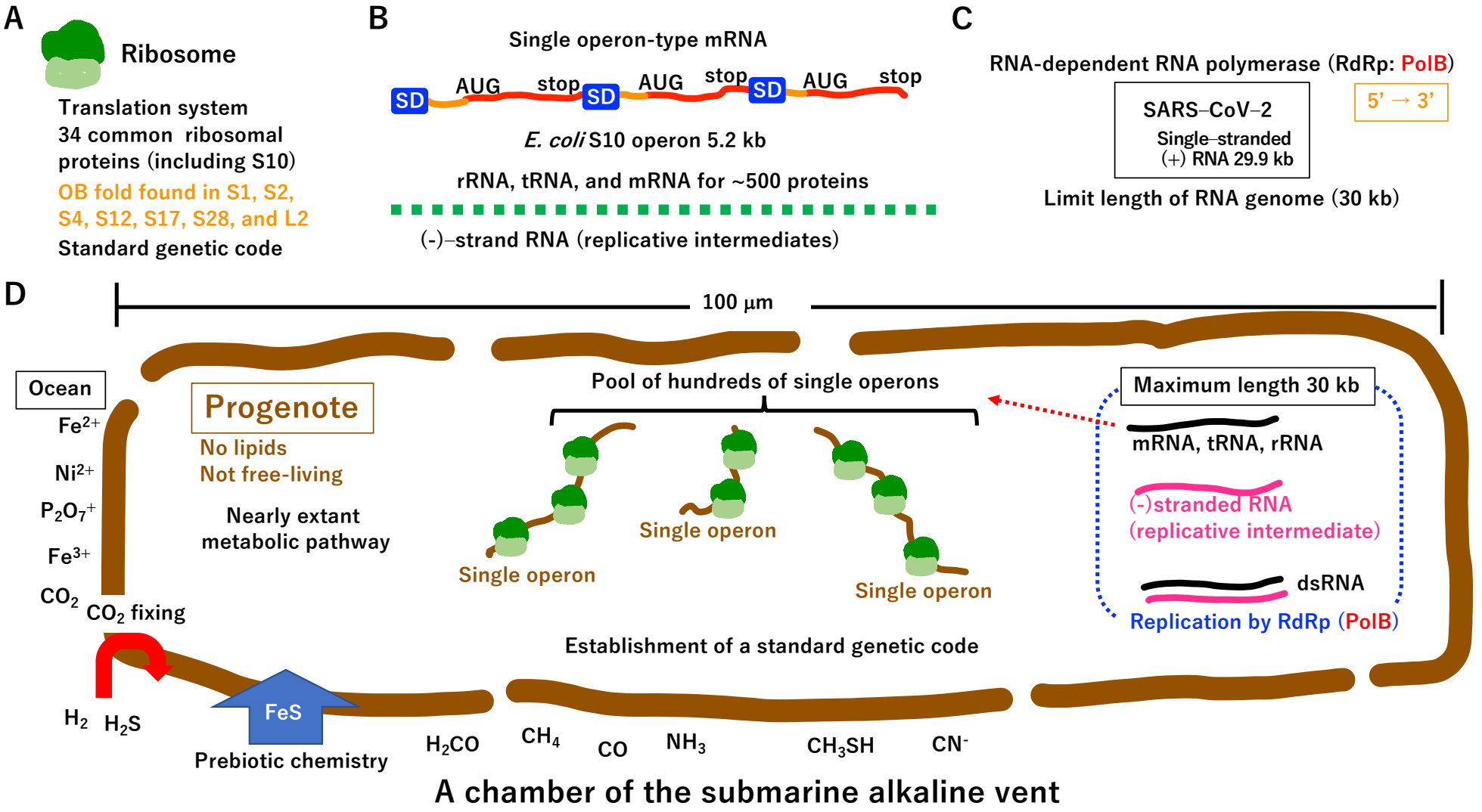
Supplementary Information

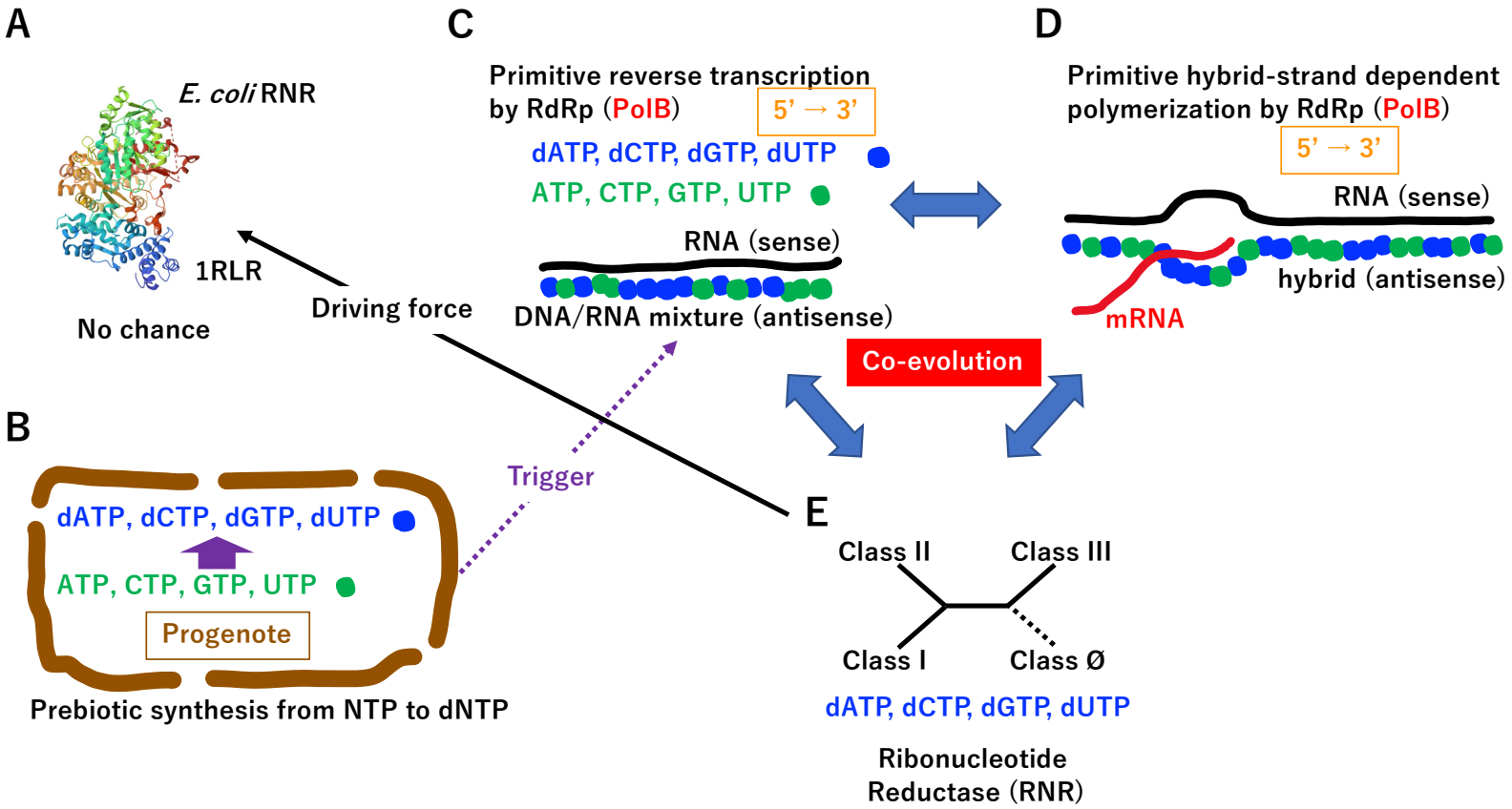
Possible crystallization process in the origin of bacteria, archaea, viruses, and mobile elements

Akari Yoshimura and Masayuki Seki

1. **Supplementary figures**
2. **Supplementary figure legends**
3. **Supplementary references**



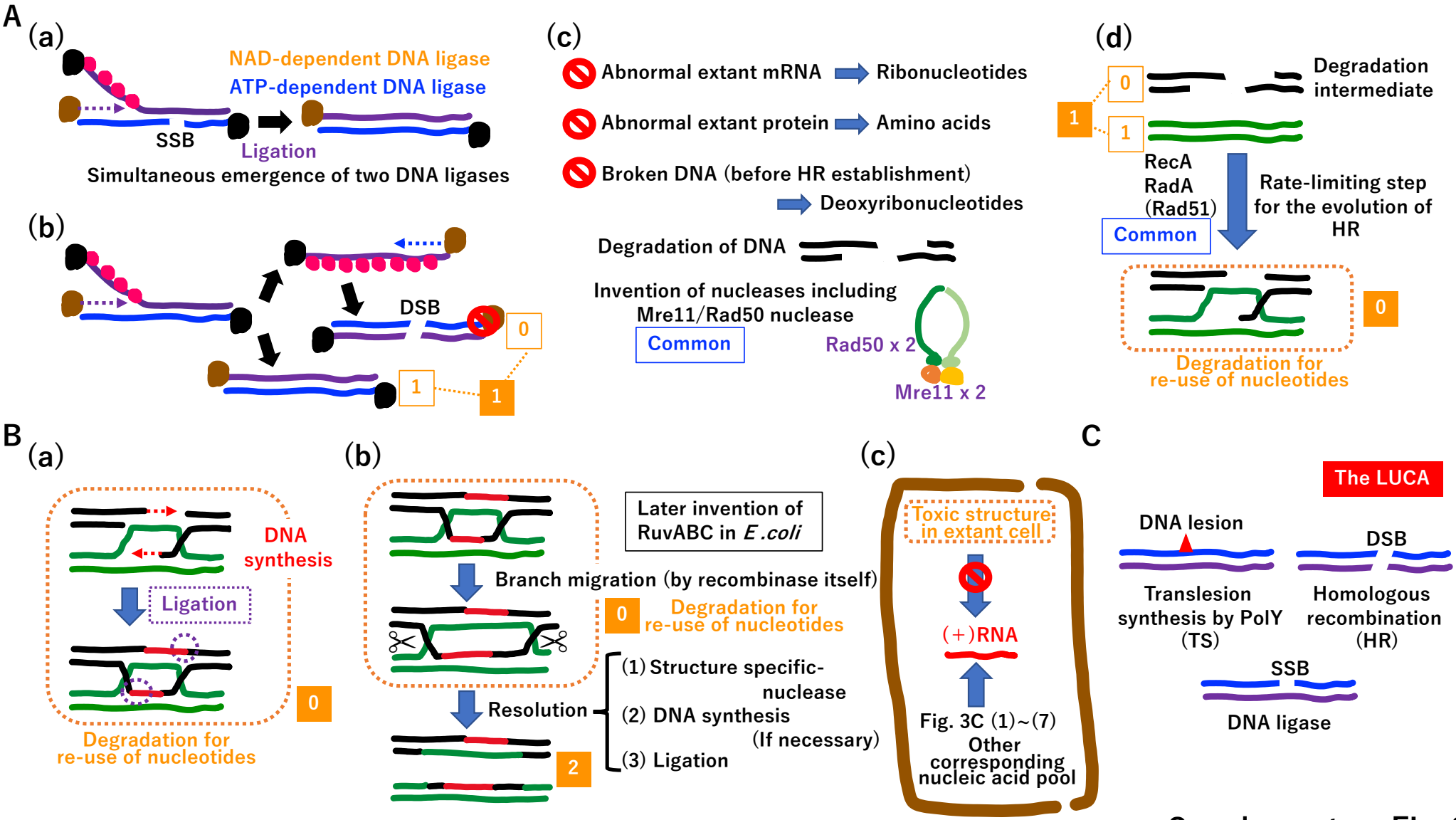
Supplementary Fig. S1



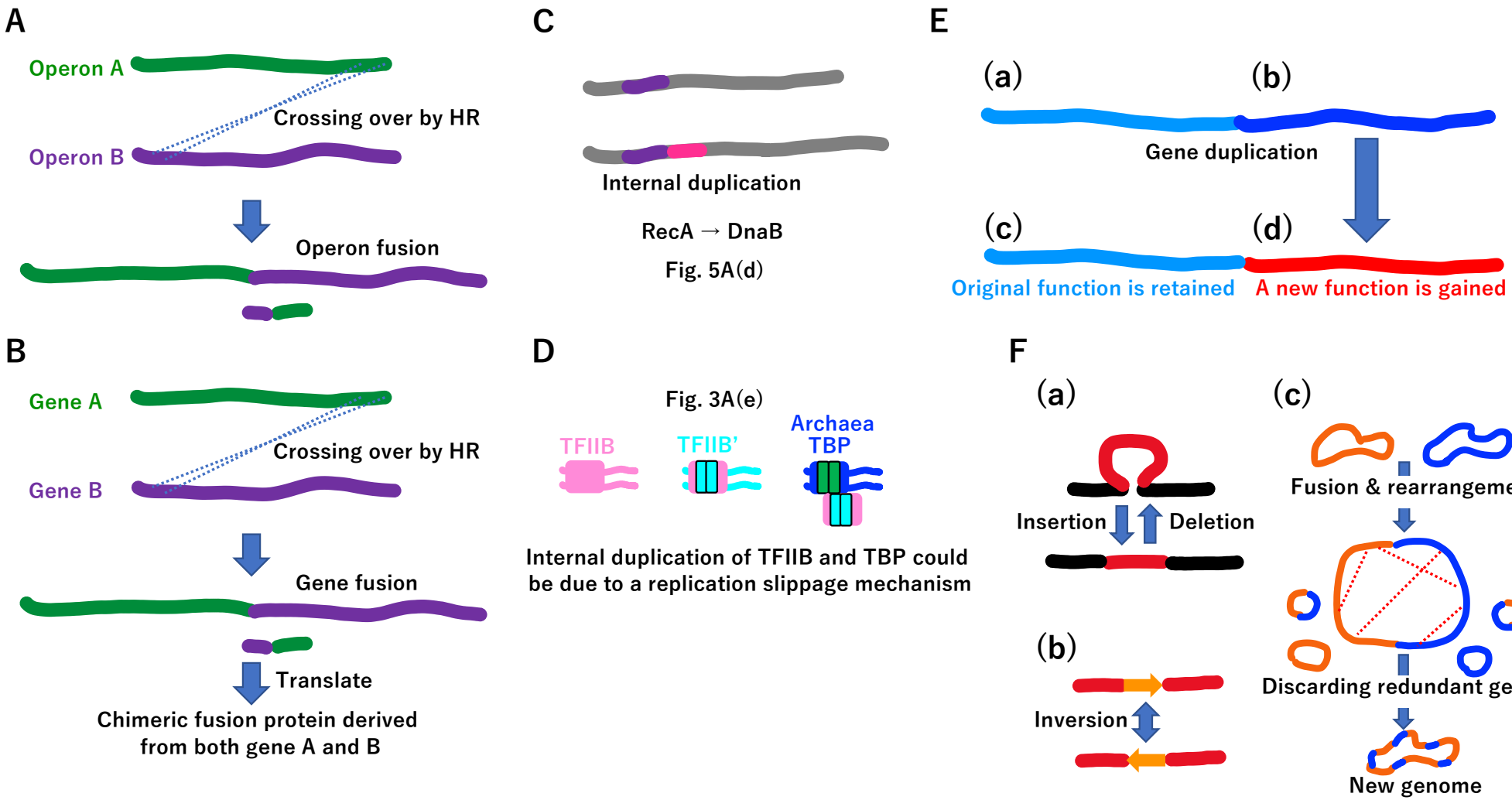
Supplementary Fig. S2

	Characteristic domain	Bacteria	Archaea		*H. sapiens*	Mobile element	Virus
		E. coli	*Eury-archaeota*	*Cren-archaeota*			
RNA-dependent RNA polymerase (RdRp) (PolB)	RRM palm	–	–	–	–	–	YES
Reverse transcriptase (RT) (PolB)	RRM palm	–	–	–	–	Group II intron	YES
T7 DNA-dependent RNA polymerase (DdRp) (PolB)	RRM palm	–	–	–	–	–	YES
DNA-dependent RNA polymerase (DdRp) (PolD)	Double psi-beta barrel	DdRp	DdRp	DdRp	Pol I, II, III	nd	YES
Type A DNA-dependent DNA polymerase (DdDp) (PolA)	RRM palm	Pol I	–	–	γ : mitochondria θ : nuclear	–	YES
Type B DdDp (PolB)	RRM palm	Pol II	DdDp	DdDp	α δ ε ζ	YES	YES
Type C DdDp (PolC)	PolC-like	Pol III	–	–	–	nd	nd
Type D DdDp (PolD)	Double psi-beta barrel	–	DdDp	–	–	nd	nd
Type X DdDp	Loop1	–	–	–	β λ μ TdT	–	nd
Type Y DdDp (PolY)	Little finger	DinB UmuC	YES	YES	η ι κ Rev1	nd	nd
Archaeo-eukaryotic primase (AEP)	RRM palm	–	–	–	PrimPol	Plasmid	YES
Rolling-circle replication endonuclease (RCRE)	RRM palm	–	–	–	–	Plasmid	YES

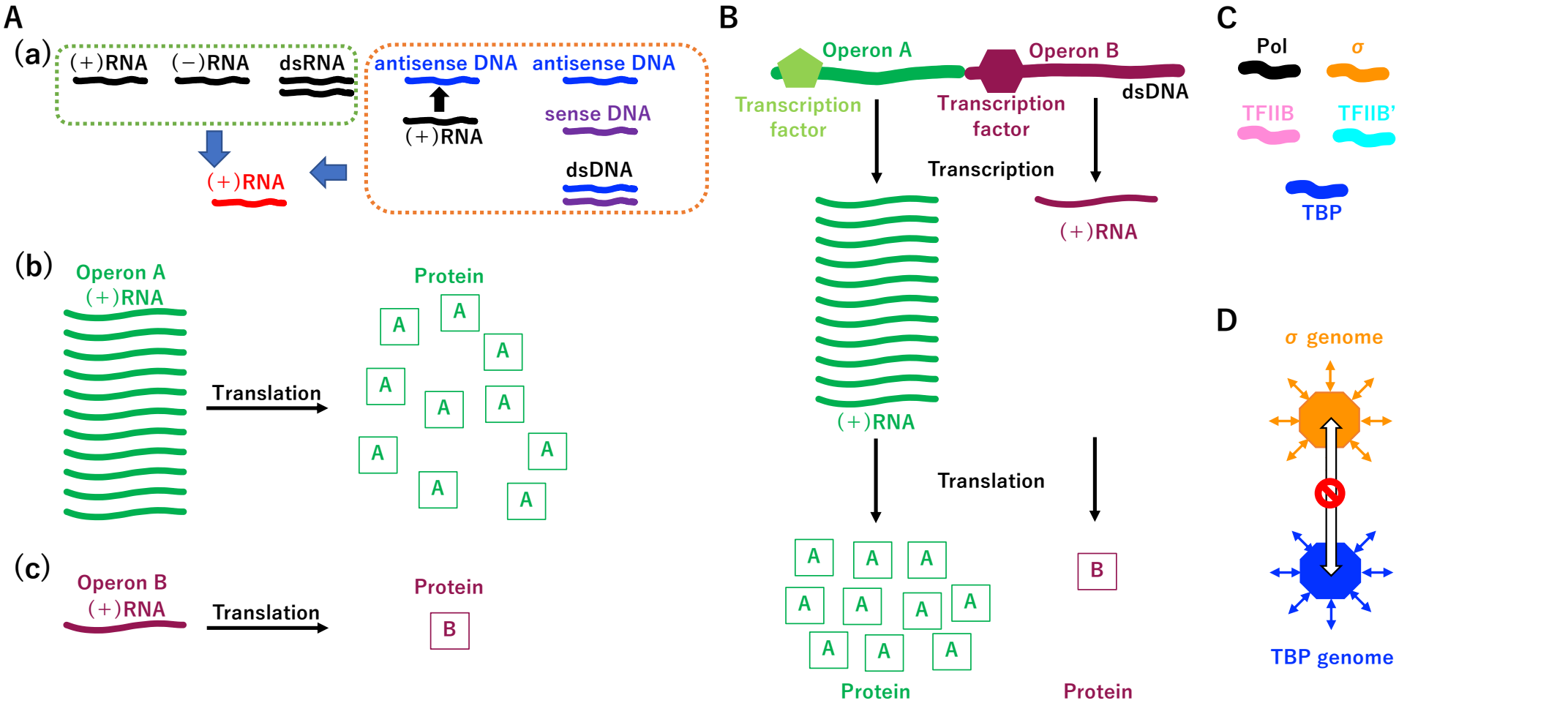
Supplementary Fig. S3



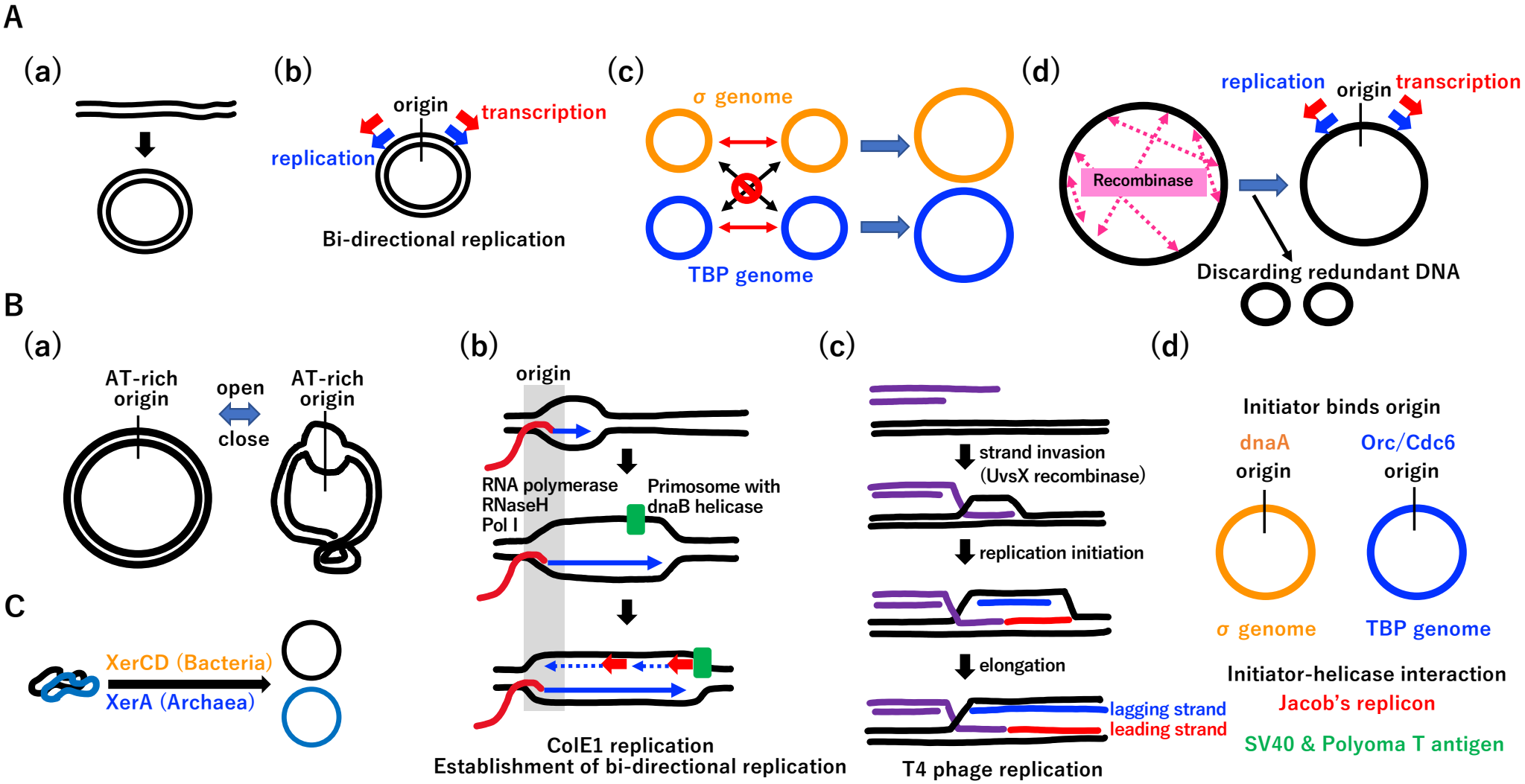
Supplementary Fig. S4



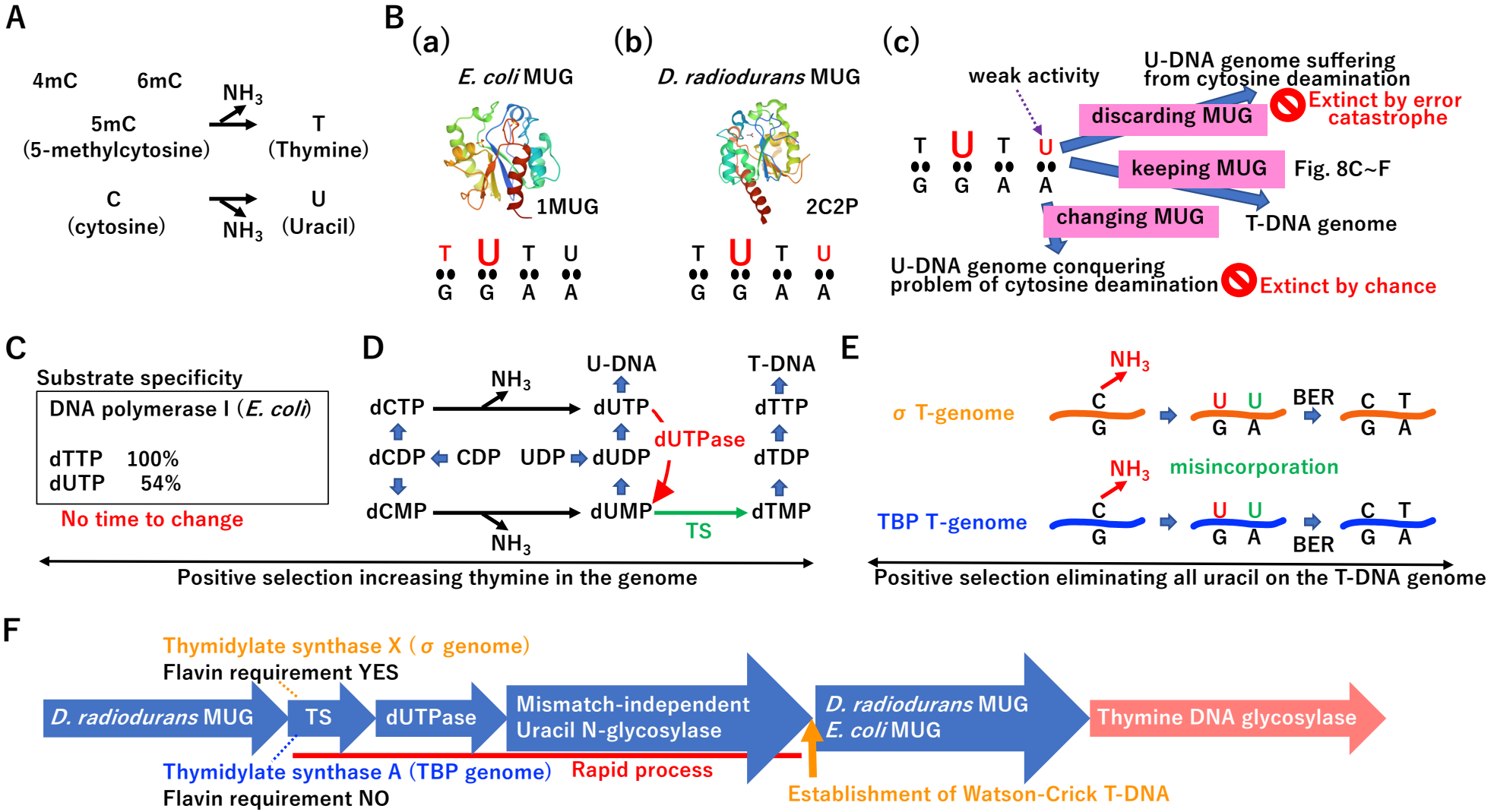
Supplementary Fig. S5



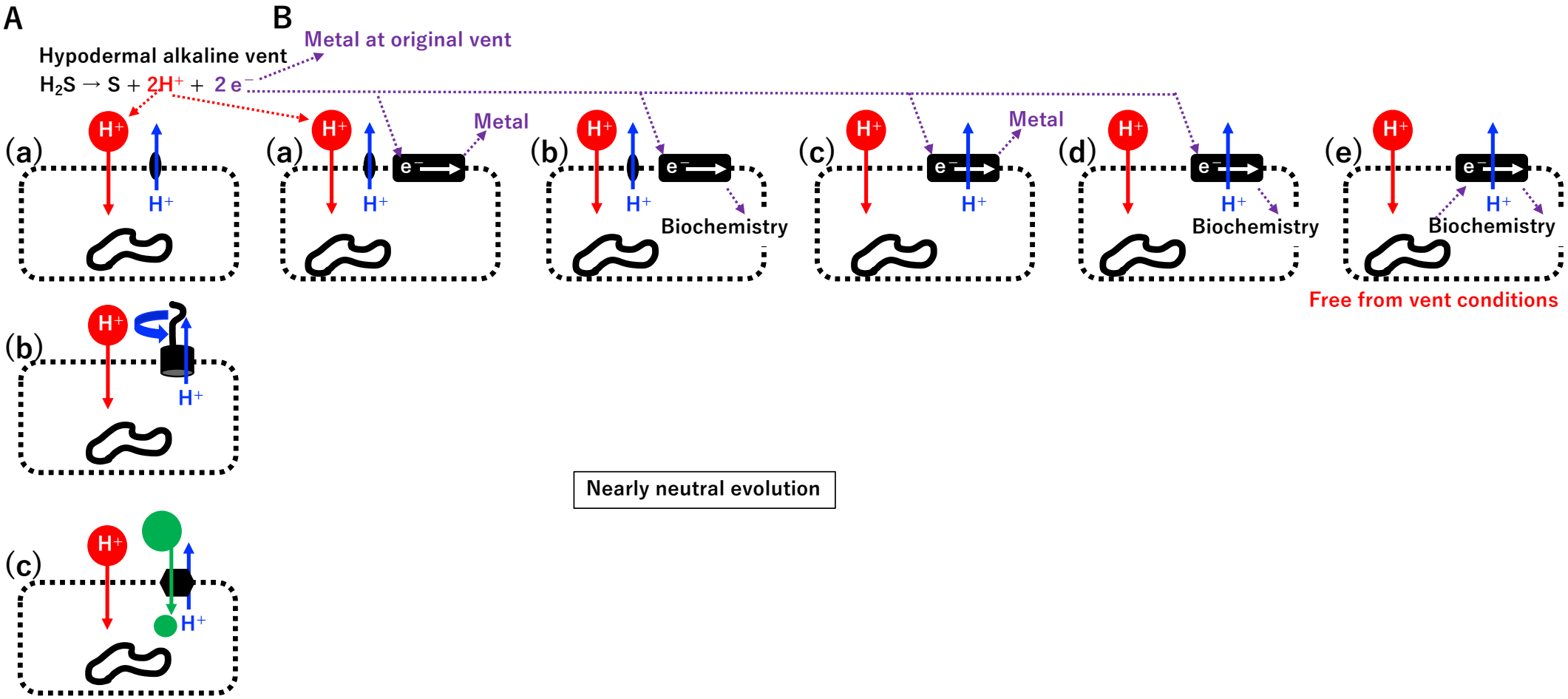
Supplementary Fig. S6



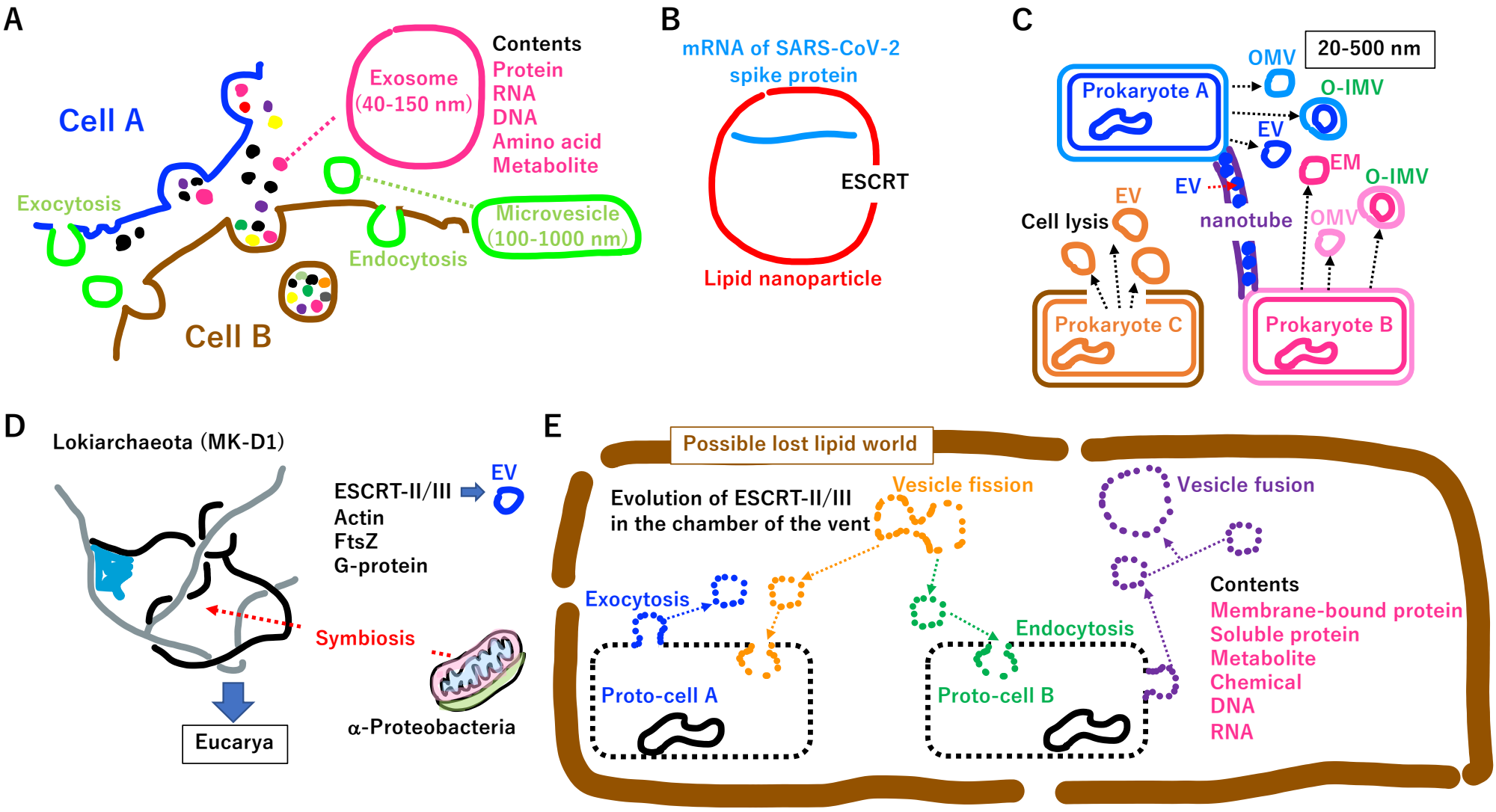
Supplementary Fig. S7



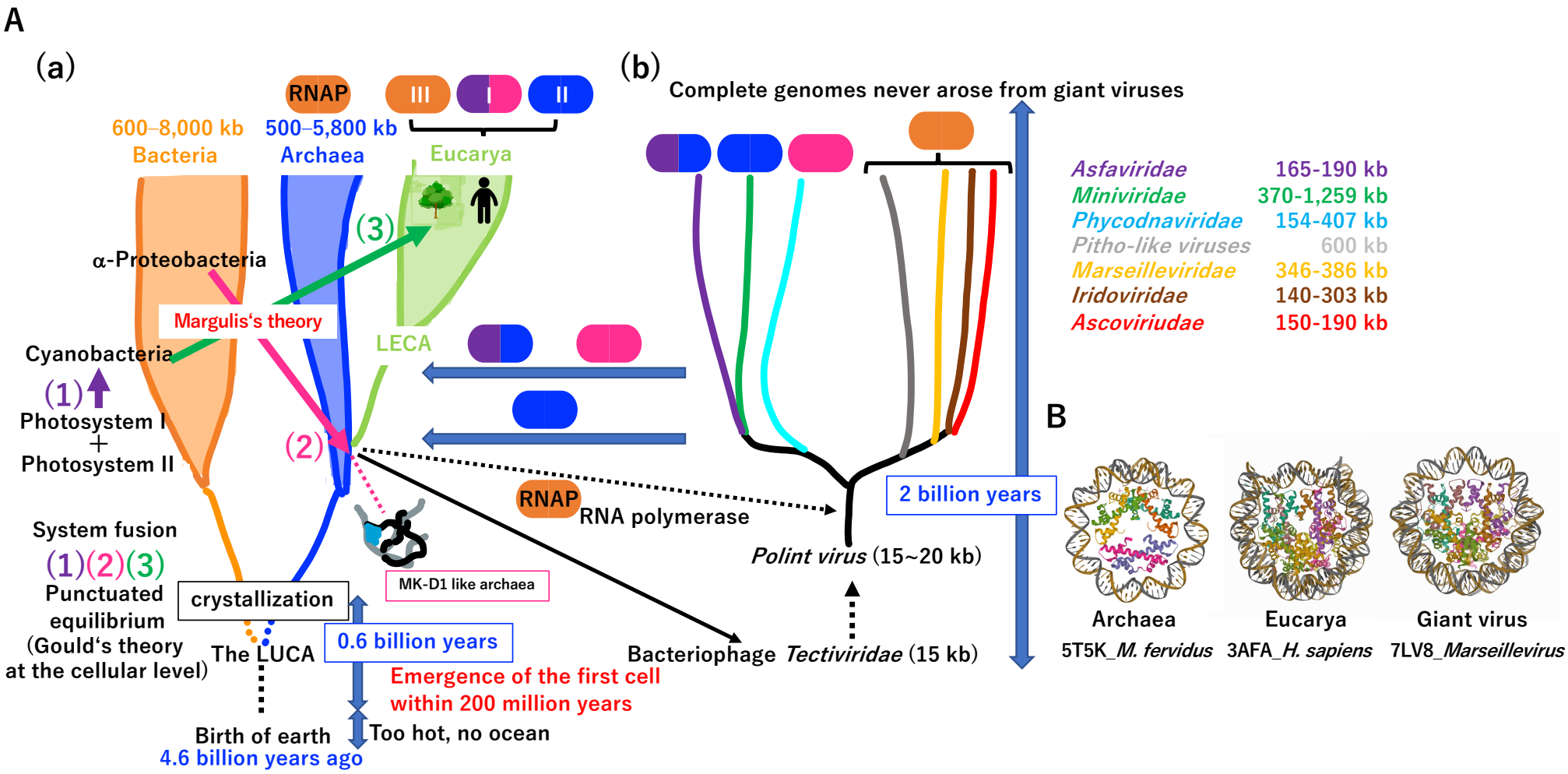
Supplementary Fig. S8



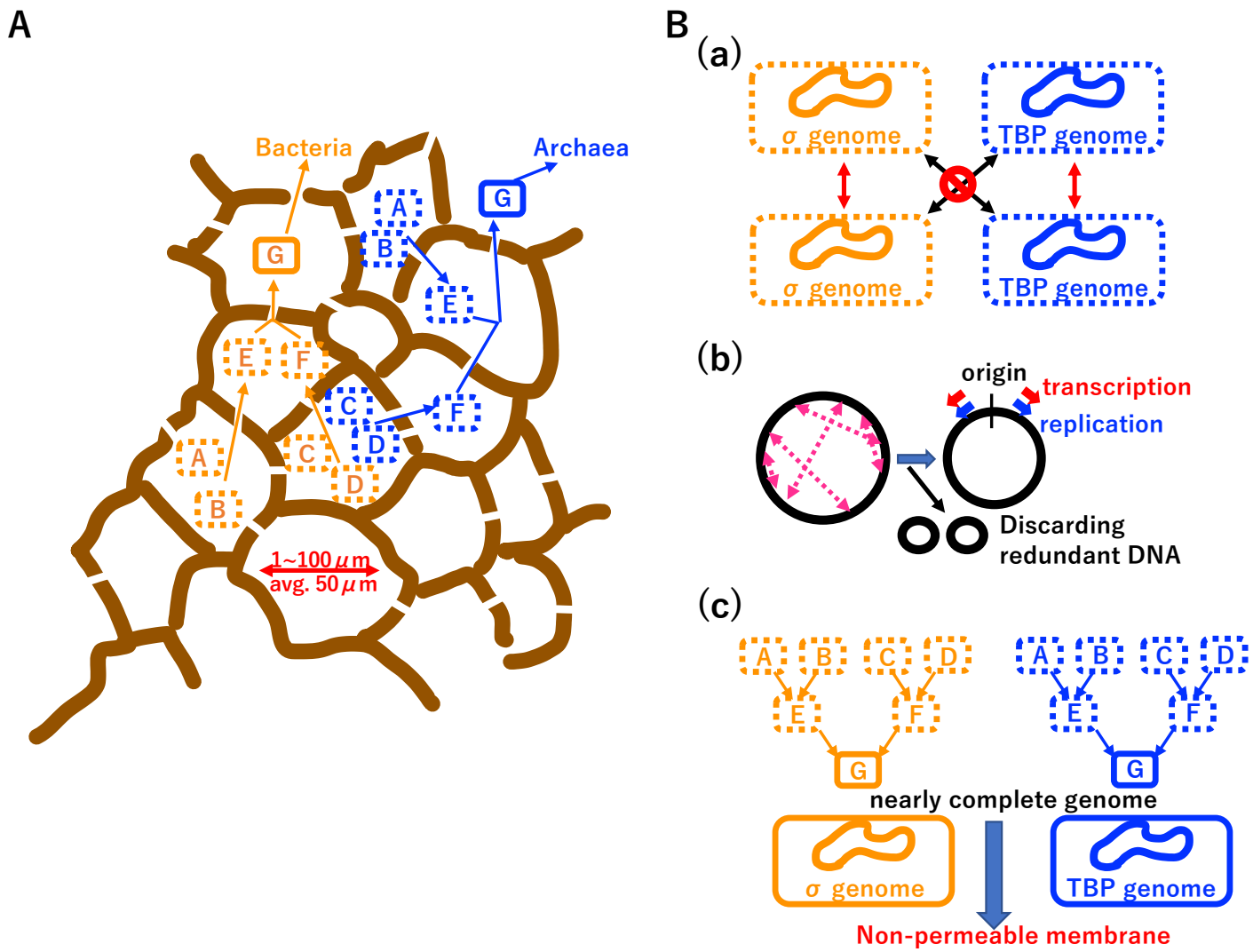
Supplementary Fig. S9



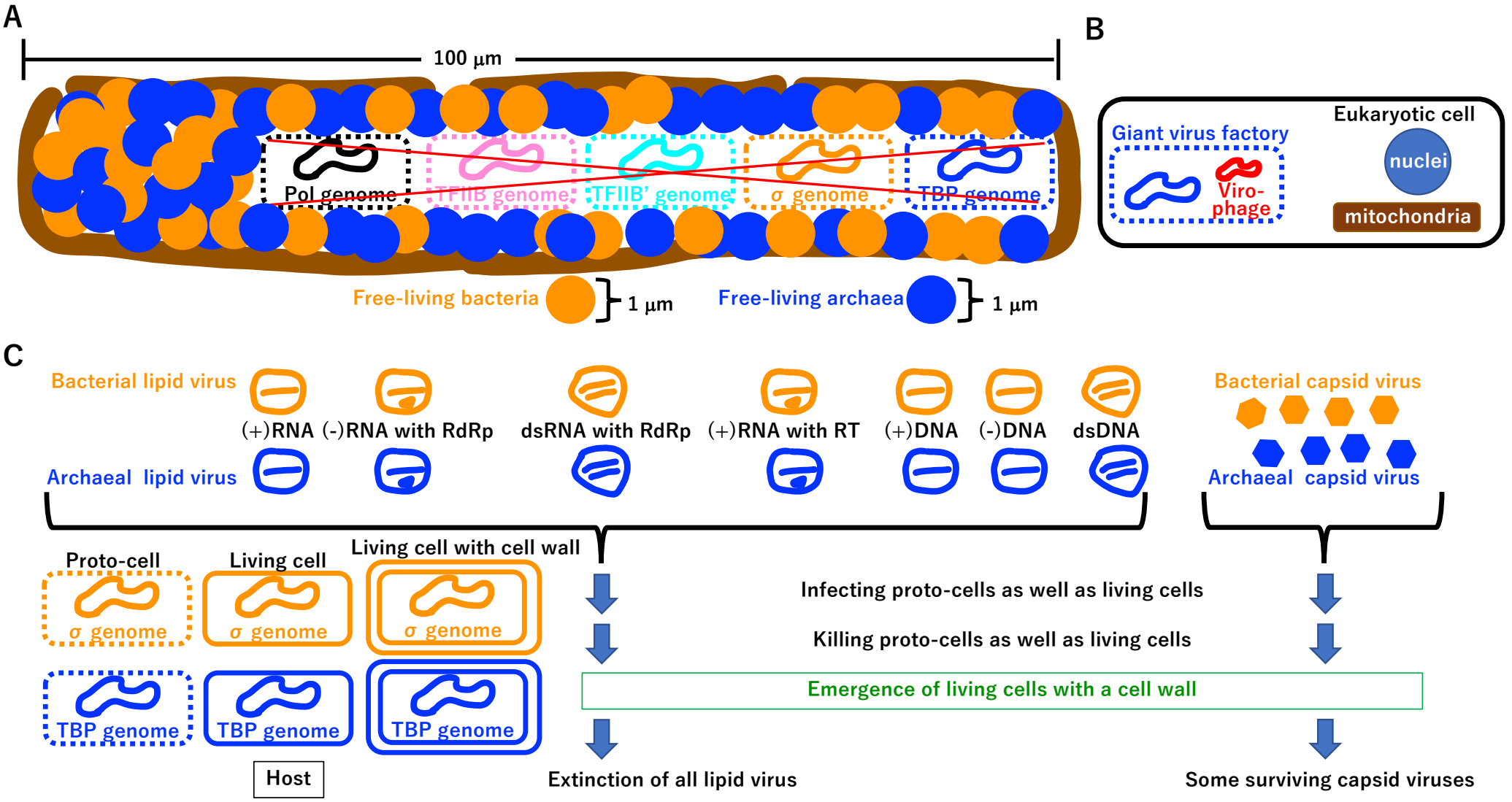
Supplementary Fig. S10



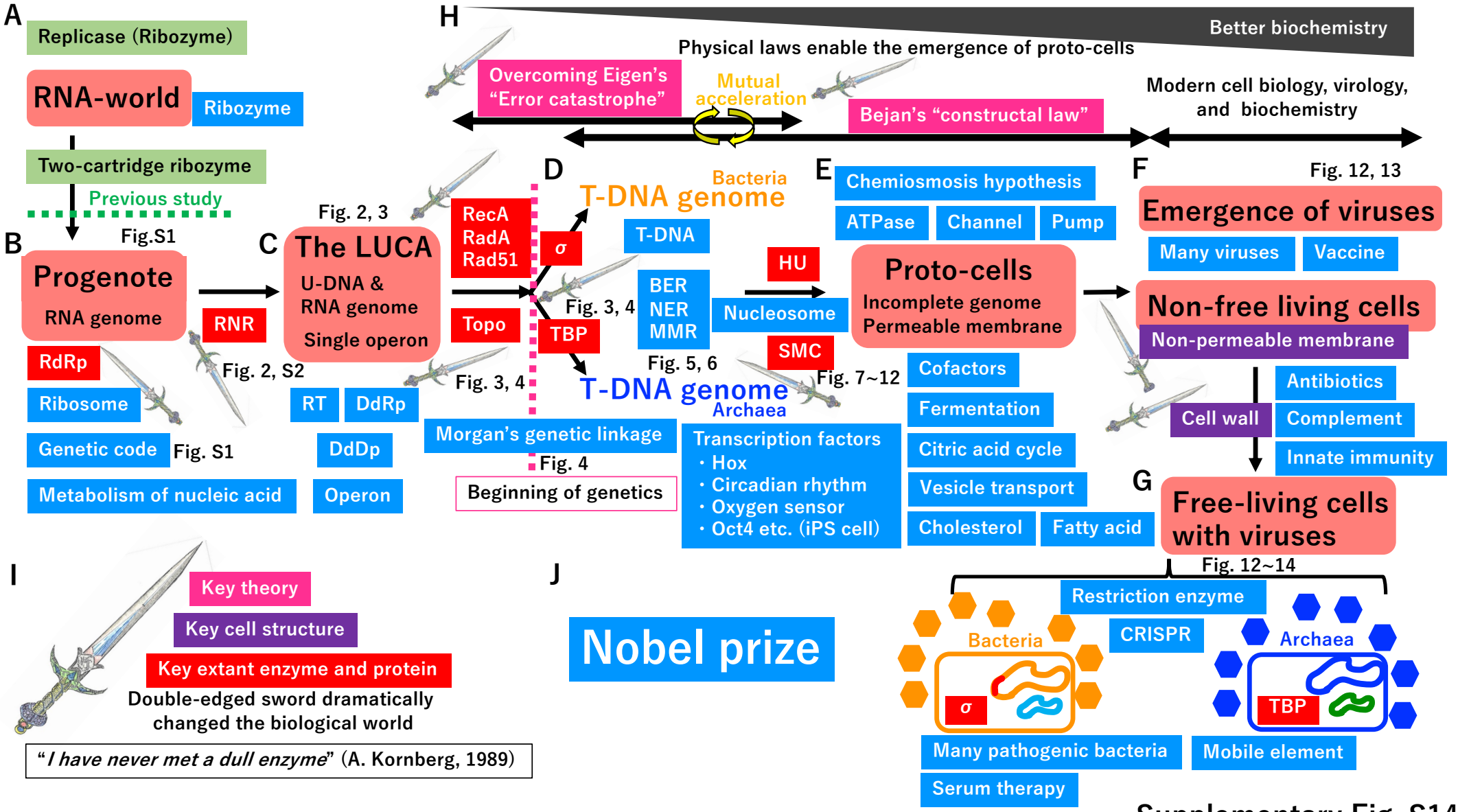
Supplementary Fig. S11



Supplementary Fig. S12



Supplementary Fig. S13



Supplementary Fig. S14

A Synthetic U-DNA cells

(a) Synthetic biology

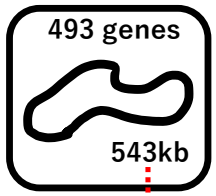


Fig. 7B

0.185 mm

1μm

JCVI-syn3A

(U-DNA)

(b) *E. coli* MUG

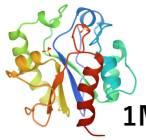


Fig. S8

1MUG

T U T U

G G A A

(c) Lack of dUTPase and uracil glycosylase

B *In vitro* evaluation of biochemical efficiency in a variety of liquid droplets by both wet and dry experiments

(a) Oparin's coacervate (Fig. 8A (c))

(b) Liquid droplet containing lipid (Fig. 8B (a)(b)(c))

(c) Phase transition from liquid droplet to permeable membrane (Fig. 8B(d))

C *In vitro* evaluation of the biochemical nature of extant membrane-bound proteins on a permeable membrane by wet and dry experiments

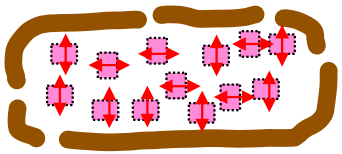
(a) Channel, pump (Fig. 9A)

(b) ATPase (Fig. 9C)

(c) Electron transport system (Fig. 9C)

D Evaluation of entropy in the chamber by "omnis cellula a cellula" of the proto-cell

Fig. 10

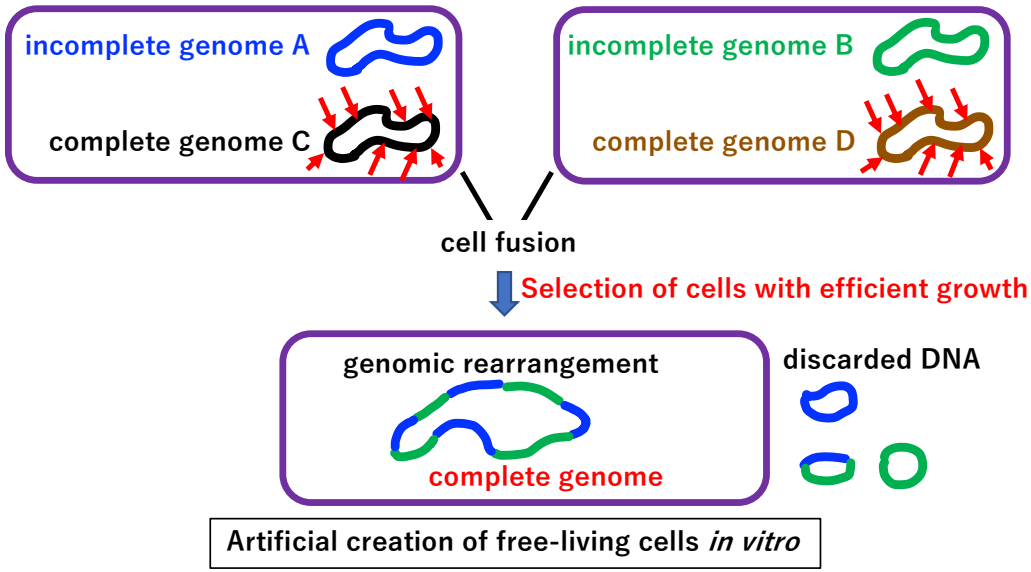


Toward maximum entropy

E Simulation of the macrovesicle-like interchange system at the chamber

Fig. S10

F Reconstruction of the final crystallization step *in vitro* using Venter's "Synthetic Biology" strategy Fig. 11, Fig. S11, Fig. S12



Supplementary Fig. S15

(2) Supplementary figure legends

Supplementary Fig. S1. Progenote

(A) Genetic code and translation system. Bacteria and Archaea share a genetic code and translation system. There are 34 ribosomal proteins shared by both lineages.

(B) Single operon-type mRNA. The common ribosomal protein S10 is encoded by the S10 operon in both lineages [1], suggesting that a single operon-type mRNA existed in the common ancestor of both lineages. There were around 500 genes in the LUCA [2, 3].

(C) Maximum length of single operon-type mRNA. Extant SARS-CoV-2 has a 30 kb RNA genome [4], one of largest genomes of RNA viruses. Thus, it is likely that 30 kb was the maximum size of single operon-type RNA in the ancient world.

(D) A chamber of the submarine alkaline vent. The progenote [5] could be surrounded by an inorganic chamber [6], with a similar size to that of extant eucaryotic cells (A). The progenote might contain the extant metabolic map (except lipid metabolism), supported by a continuous supply of prebiotic resources and energy from the alkaline vent as well as enzymes encoded by the pool of RNA.

Supplementary Fig. S2. Darwinian selection on the transition from RNA to U-DNA.

(A) Structure of RNR. Structure of *E. coli* ribonucleotide reductase (RNR) is presented with the PDB number. It is possible that there were no opportunities for the spontaneous emergence of sophisticated and complex enzymes, like RNR.

(B) Prebiotic synthesis of deoxyribonucleotides. Since the conversion from ribonucleotides to deoxyribonucleotides has been demonstrated *in vitro* in a prebiotic synthesis manner [7-10], the small number of dNTPs could trigger coevolution among (C), (D), (E), and (A), as described in the text.

(C) (D) (E) are the same as in Fig. 2.

Supplementary Fig. S3. Variety of replicases. As described previously [11-14], various replicases in the three domains of life are summarized. YES, indicates existence. n.d. indicates not described in the above references. Orange background indicates bacterial lineages. By contrast, a blue background indicates archaeal lineages.

Supplementary Fig. S4. Establishment of homologous recombination.

(A) Single-stranded and double-stranded breaks. (a) Origin of DNA ligases. DNA polymerase at SSB (single-stranded break) on the template leads to lethal damage, double stranded-break (DSB). To prevent such DSB and complete DNA replication, two types of DNA ligases could simultaneously arise [15]. (b) DSB. If DSBs occur in dsDNA, the broken DNA would be inert. Live dsDNA is marked 1 and inert dsDNA is marked 0. (c) The

degradation of abnormal macromolecules is beneficial for metabolism due to the reuse of monomer molecules. Thus, nuclease-mediated broken DNA degradation could emerge in the ancient world. Since Mre11/Rad50-type nucleases commonly exist in Bacteria and Archaea [16, 17], they might be among the nucleases that arose in these taxa. (d) Birth of recombinase. Degradation intermediates of dsDNA might provide a basis for the evolution of recombinase, an ancestor of recA, radA, or Rad51 in Bacteria, Archaea, or Eucarya, respectively [18, 19]. Of note, the synapsis of two dsDNA inactivates the template (marked by 0) for DNA replication and transcription.

(B) Completion of HR. (a) Synapsis, DNA synthesis, and ligation. Intermediate A(d), after synapsis, could be a template for DNA polymerase (Fig. 2, Supplementary Fig. S3) and subsequent DNA ligation (A)(a), leading to a double Holliday junction-type intermediate. Importantly, such an intermediate might exhibit a lack of activity (marked 0). (b) Recombinases itself, such as recA and Rad51, have primitive branch migration activity [20, 21]. Since evolution is not linear, dHJ (inert intermediate marked 0) might be degraded and reused for mono deoxynucleotides. Particular digestion of dHJ could yield two active dsDNAs (marked 2). (c) Robustness in the LUCA. Importantly, dsDNA produced the same RNA in the LUCA (Fig. 3C). Thus, null intermediates (A)(d), (B)(a), and (B)(b) might survive in the LUCA by nearly neutral selection [22]. Such a robust system could support the evolution of a complicated series of homologous recombination reactions. Branch migration and dHJ resolution by RuvABC [23] could have arisen specifically in bacterial lineages.

(C) Basic tool kit for the completion of dsDNA replication on damaged DNA. Transletion polymerase PolY (Fig. 2J, 2K, Supplementary Fig. S3) could mediate replication on damaged templates [24]. SSB on the template could be sealed by DNA ligase before encountering DNA polymerase. DSBs on dsDNA could be repaired by using homologous dsDNA in the chamber of the vent. Thus, the primitive tool kit for the completion of dsDNA replication could be established under the robust circumstances of the LUCA.

Supplementary Fig. S5. Recombination-mediated genome rearrangement.

(A) Operon fusion. Crossing over-mediated homologous recombination facilitates the fusion of two operons.

(B) Gene fusion. Upon HR-mediated fusion between gene A and gene B, a novel fusion protein could easily emerge.

(C) Internal duplication. HR-mediated internal duplication could create dnaB from recA (Fig. 5A) [25].

(D) Replication slippage-mediated internal duplication. TFIIB and TBP have internal duplications (Fig. 3A) [26]. Since we predict that such duplications occurred before the establishment of HR, they might have occurred by replication slippage [27].

(E) Gene duplication. Ohno originally proposed that gene duplication facilitates evolution [28]. (a) Original gene. (b) Duplicated gene. (c) Gene maintaining the original function. (d) Gene with a gain-of-function mutation.

(F) Genome rearrangement. (a) Insertion and deletion mediated by HR. (b) Inversion by HR. (c) New genome. Two genomes (orange and blue) could be fused and undergo rearrangements and gene loss, yielding a new genome. Whole genome rearrangement requires HR.

Supplementary Fig. S6. Transcriptional regulation.

(A) Copy number of nucleic acids regulates the amount of protein. (a) Seven types of nucleic acids might directly or indirectly produce mRNA, which is translated into protein in the LUCA (Fig. 3C). (b) (c) Amounts of protein A and B. The amount of protein A could be ten-fold larger than that of protein B due to the amount of mRNA derived from operons A and B.

(B) Two operon fusion. After the fusion of dsDNA on operons A and B, one copy of DNA should produce a ten-fold difference between levels of protein A and B. Transcriptional regulation mediated by transcription factors [29] could emerge immediately after this fusion event. The basic unit of transcriptional regulation is promoter recognition (Pol, σ, TFIIB, TFIIB', and TBP/TFIIB') (Fig. 3A). Since the same promoter recognition system might be beneficial for the development of transcriptional regulatory networks (B), the transcriptional regulation divide should be accelerated (C) (Fig. 3A).

(C) Transcriptional divide based on the promoter recognition system.

(D) σ genome and TBP genome. The σ [30] and TBP [31] genomes could give rise to Bacteria and Archaea, respectively. Bacterial proteins cannot be exchanged with archaeal proteins. Such incompatibility between lineages might have begun at the time of the two-operon fusion (B) (Fig. 4B, 4C, Fig. S5A).

Supplementary Fig. S7. Establishment of Jacob's replicon.

(A) Ideal origin defined by genome rearrangements on circular dsDNA. (a) To avoid the terminal replication problem, a circular dsDNA genome could emerge on both the σ and TBP genome. (b) Bi-directional replication. Compared to unidirectional replication, bi-directional replication could speed up the completion of DNA replication of the whole genome. To minimize collision between replication and transcription (Fig. 5A(b)), the direction of replication and transcription tends to be the same [32] via HR. (c) Two system fusion. Genomes carrying the same transcriptional operation system could be preferentially fused. (d) Novel fused genome. The fused genome could undergo rearrangement by HR (left column) as well as the loss of redundant genes. A new genome (right column) could arise under the same process (b).

(B) Evolution of an origin firing mechanism. (a) Mutation at the origin. A pre-defined origin on a new genome (A(d)) could acquire mutations toward an AT-rich sequence for easy opening. (b) Firing by RNA polymerase. The origin could be opened by RNA polymerase, similar to extant ColE1 plasmid replication [33]. (c) Firing by recombination. The origin could be opened by recombination, similar to T4 phage replication [34]. (d) Evolution of initiators. dnaA [35] and Orc/Cdc6 [36] could arise independently in the σ and TBP genomes, respectively. The initial initiator might only bind the replication origin. Later, interactions between the initiator and replicative DNA helicase could evolve [37-42]. Thus, the replicon model for DNA replication [43-45] could be established in a step-by-step manner, as described in Figs. 2, 5, and Supplementary Fig. S7. An extreme case of an initiator-helicase interaction is achieved in SV40 and polyoma virus T-antigen [46, 47].

(C) Unlinking of duplicated DNA. After the completion of DNA replication on the circular DNA genome, two daughter DNAs are topologically linked. The unlinking of daughter DNA by Xer C/D and Xer A might emerge in Bacteria and Archaea, respectively [48].

Supplementary Fig. S8. Possible scenario for the transition from U-DNA to T-DNA.

(A) Spontaneous deamination. 5-Methylcytosine and cytosine can be spontaneously deaminated, yielding thymine and uracil, respectively [49].

(B) Evolution of mismatch-specific uracil glycosylase (MUG). (a) *E. coli* MUG (structure shown with PDB ID) recognizes U/G and T/G mismatches. (b) *D. radiodurans* MUG (structure shown with PDB ID) recognizes the U/G mismatch and U/A non-mismatch [50]. (c) Possible substrate specificity of ancient MUG. We hypothesize that ancient MUG commonly emerged in both the σ and TBP genomes and has similar substrate specificity to that of *D. radiodurans* MUG. The U/T non-mismatch pair can be cut by the postulated ancient MUG. If the hypothesis is correct, there are three scenarios for evolution in the ancient alkaline vent. First, discarding the toxic *D. radiodurans* MUG-type enzyme led to cytosine deamination in the U-DNA genome, resulting in extinction. Second, keeping the *D. radiodurans* MUG-type enzyme drove the U-DNA to T-DNA transition (C–E). Third, changing the substrate specificity of *D. radiodurans* MUG toward that of *E. coli* MUG resolved the cytosine damination problem in the U-DNA genome. Although the third scenario predicts the existence of the U-DNA genome, cells with a U-DNA genome do not exist at present, suggesting that it went extinct. Since both extant bacteria (σ-genome) and archaea (TBP-genome) have T-DNA genomes, the second evolutionary route was most likely at the original alkaline vent.

(C) DNA polymerases efficiently utilize dUTP. DNA polymerases, including *E. coli* DNA polymerase I (Pol I), cannot discriminate between dTTP and dUTP as substrates for DNA synthesis [49], suggesting that the transition from U-DNA to T-DNA occurred without changing the DNA polymerase substrate specificity. The lack of time to change

strongly suggests that rapid evolution increasing dTTP coupled with decreasing dUTP occurred in the ancient alkaline vent.

(D) Simultaneous evolution of TS and dUTPase. Metabolic map for dTTP synthesis is shown. Thymidylate synthetase (TS) might have evolved independently in the σ and TBP genome (F), leading to increased dTTP concentrations. By contrast, dUTPase decreases the dUTP concentration and supplies dUMP as a substrate for TS. The evolution of both TS and dUTPase enabled the transition from U-DNA to T-DNA.

(E) Elimination of mis-incorporated dUMP from the T-DNA genome. Due to the substrate specificity of DNA polymerase (C), dUMP can be frequently mis-incorporated into T-DNA. To eliminate uracil from T-DNA, mismatch-independent uracil glycosylase could arise. Coordinated activity of TS, dUTPase, and mismatch-independent uracil glycosylase-mediated BER could establish the T-DNA genome, whose structure was solved in 1953 [51].

(F) Possible steps in the transition from U-DNA to T-DNA. Summary of postulated events (B–E). After establishment of the T-DNA genome, the substrate specificities of *E. coli* and *D. radiodurans* MUGs were compatible with the T-DNA genome. Moreover, thymine DNA glycosylase might arise later.

Supplementary Fig. S9. Electron transport system.

(A) Possible collection of outward proton transporters. External protons supplied from the inorganic chamber could trigger the evolution of inward and outward proton transport mechanisms [52, 53] based on the maximization of entropy inside and outside of proto-cells. (a)(b)(c) represent possible outward proton transporters. Importantly, any of these might be under nearly neutral selection [22].

(B) A variety of electron transport systems. In extant alkaline vents, electrons derived from H_2S convert O_2 to H_2O. Under a lack of O_2 in ancient oceans, electron derived from H_2S could be transferred to surrounding metals at the vent (A). (a)(b) Electron transport on the permeable membrane of proto-cells without coupling to outward proton transport. Outward proton transport might be independently derived from electron transport. (c)–(e) A variety of electron transport systems coupled with outward proton transport. Importantly, any of these systems (a)–(e) might be under nearly neutral selection [22]. System (e) could be selected in ancient cells (Fig. 12) because electron transport could be driven independently from the surrounding inorganic vent. Thus, independence from the original alkaline vent could be achieved by free-living cells.

Supplementary Fig. S10. Postulated ancient vesicle transport system.

(A) Extant exosome and microvesicle in Eucarya. Membrane-based transport systems are present in Eucarya [54, 55].

(B) mRNA vaccines. An mRNA vaccine for the SARS-CoV-2 spike protein mimics the membrane-based transport system (A) [56].

(C) Extant bacteria/archaea extracellular exchange system. A variety of extra-membrane vesicles are exchanged among cells [57, 58]. In a biofilm, a nanotube containing multiple extra-membrane vesicles is connected between two different cells. EV: Extracellular vesicle, OMV: Outer membrane vesicle, O-IMV: Outer-inner membrane vesicle.

(D) The ESCRT system for vesicle-based transport. Some Archaea, such as Lokiarchaeota MK-D1, have the ESCRT system, suggesting that MK-D1 cells have an internal vesicle transport system, as in extant Eucarya. Some Archaea produce extramembrane vesicles (EVs) via the ESCRT system [58]. Moreover, enveloped archaeal viruses might bud from host cells via the ESCRT system [59]. It was proposed that ancient Archaea, like MK-D1 cells, undergo symbiosis with α-proteobacteria, leading to eucaryotic lineage cells [60].

(E) Postulated ancient vesicle transport system. Accounting for (A), (C), (D), and the concepts outlined in Fig. 7B, a primitive macrovesicle-like transport system at the chamber of the vent was essential for survival of proto-cells. We hypothesize that ESCRT-dependent vesicle transport was invented at the ancient chamber. The postulated vesicle might carry any macromolecules.

Supplementary Fig. S11. Puncted equilibrium at the cellular level.

(A) Evidence of system fusion and exchange during the history of life on Earth. (a) Three revolutionary system fusions. (1) Photosynthesis expending CO_2 and producing O_2 might be derived from the fusion of photosystems I and II. Event (1) created Cyanobacteria and irreversibly changed the conditions on Earth toward an O_2-available world. (2) Symbiosis between α-Proteobacteria and Archaea [61], like MK-D1 cells [60], led to the ancestor of extant Eucarya. (3) Symbiosis between Cyanobacteria and Eucarya led to an archaeplasrtid ancestor of extant plants [61]. (b) Giant virus-mediated system exchange. Extant giant virus carry a large DNA genome (up to 1,259 kbp) [62]. During symbiosis (2), the bacteriophage Tectiviridae (15 kb), whose host is α-Proteobacteria, might be introduced into the ancestor of Eucarya [63]. The coevolution of the ancestor of Eucarya and giant viruses could yield the LECA (last eukaryotic common ancestor). Eucarya has three DNA-dependent RNA polymerases (DdRp), I, II, and III. A phylogenetic analysis of DdRp revealed that multiple gene transfer events and rearrangements [64] created three DdRp(s) leading to the establishment of the LECA.

(B) Coevolution of the nucleosome structure. Just like DdRp, the nucleosome structure might also co-evolve in the ancestor of Eucarya and giant viruses [64]. Representative nucleosome structures of Archaea [65], Eucarya, and giant viruses [66], presented with their PDB IDs.

Although Mimiviridae has a 1,259 kbp DNA genome, which is larger than the genome of JCVI-syn3, the minimum cell, giant viruses including Mimivoridae never reached free-living cells over 2 billion years. By contrast, free-living Bacteria and free-living Archaea emerged simultaneously in a narrow time window (an estimated 200 million years). Thus, we hypothesize that system fusion consistent with punctuated equilibrium occurred in the ancient alkaline vent (Fig. 11, Supplementary Fig. S12).

Supplementary Fig. S12. Emergence of a nearly complete genome with a non-permeable membrane.

(A) Cascade of proto-cell fusion. Proto-cell E, the fusion product of proto-cells A and B, was fused to proto-cell F, the fusion product of proto-cell C and D, yielding cell G, with a nearly complete genome surrounded by a non-permeable membrane. The proto-cell and cell are surrounded by dashed (permeable membrane) and open (non-permeable membrane) boxes, respectively. Orange and blue cells represent bacterial and archaeal lineages, respectively.

(B) Crystallization process by a cascade of proto-cell fusion coupled with genome rearrangement. (a) Homologous system fusion. The same mechanisms underlying transcriptional regulation are preferentially fused (Fig. 4). (b) Genome rearrangement. Upon the fusion of two systems, genome rearrangement occurs, as described in Supplementary Fig. S7A. (c) Emergence of cells. After gaining a nearly complete genome, a permeable membrane (dashed box) changed into a non-permeable membrane (open box).

Supplementary Fig. S13. Cell wall formation and the extinction of enveloped viruses.

Orange and blue represent bacterial and archaeal lineages, respectively.

(A) Free-living cells occupied all chambers of the original vent. All replicative intermediates described in Figs. 2–11, including proto-cells, suffered from shortages of energy and resources. Most intermediates went extinct, except for some primitive viruses (Fig. 13B).

(B) Virophage. Since extant virophages can infect giant viruses [67], primitive viruses in ancient vents also infected proto-cells as well as free-living cells.

(C) Cell wall formation. Among primitive viruses, enveloped viruses might have went extinct as a result of the inability to penetrate the cell wall. Some capsid viruses remained.

Supplementary Fig. S14. New proposal for the crystallization process in the origin of bacteria, archaea, viruses, and mobile elements.

Main and supplementary figures related to each process involved in crystallization are marked in (B)–(H).

(A) RNA world. We proposed a mechanism for the change from an RNA world to a progenote (Supplementary Fig. S1) [68]. A replicase made by RNA might arise [69-71]. A two-cartridge ribozyme might be a key for the evolution of the genetic code and translation system [68].

(B) Progenote. A key innovation for the maintenance of the progenote is RNA-dependent RNA polymerase (RdRp; PolB) (Supplementary Fig. S1).

(C) The LUCA. The transition from the progenote to the LUCA required multiple enzymes, including ribonucleotide reductase (RNR) [72] (Fig. 2A, Supplementary Fig. S2). Enlargement of U-DNA might require a recombinase (Fig. 3D) and topoisomerase (Fig. 5A).

(D) Origin of genetics. Transcriptional initiation relied on σ or TBP in extant bacteria and archaea, respectively (Fig. 3A). This divide begins at the initial stage of operon fusion, leading to genetic linkage (described by Morgan) and vertical inheritance (described by Mendel). Although the DNA replication machinery and DNA repair arose independently and simultaneously in both lineages (Figs. 2, 5, 6), genes encoding common enzymes might have arisen in either lineage, followed by horizontal gene transfer. Since the size of the chamber in the original vent might be similar to the size of extant eukaryotic cells, enlarged DNA could be compacted by HU and histones in bacteria and archaea, respectively (Fig. 6D). Further compaction could require SMC in both lineages (Fig. 7D).

(E) Proto-cell. Compacted DNA might trigger coacervate [73] formation, lipid innovation, and proto-cells with a permeable membrane (Fig. 8).

(F) Non-free-living cells with a nearly complete genome surrounded by a non-permeable membrane. Cells started to occupy all chambers in the vent in a selfish manner, leading to the emergence of primitive enveloped and capsid viruses. An arms race between non-free living cells and primitive viruses similar to that between extant cells and viruses might have occurred (Figs. 12, 13).

(G) Free-living cells with viruses. Cell wall formation and independence from the inorganic vent yielded free-living cells with capsid viruses and mobile elements. During the course of cell wall formation, primitive enveloped viruses might have gone extinct (Figs. 12, 13).

(H) Better biochemistry. The crystallization process could be driven by the laws of physics. The enlargement of DNA could overcome the “error-catastrophe” [74]. Proto-cell formation could be driven by Bejan’s “constructal law” [75] and the second law of thermodynamics. Overcoming “error-catastrophe” and Bejan’s “constructal law” could improve biochemical reactions and the crystallization process described in (D)–(F). Thus, σ and TBP genomes produce distinct lineages of free-living cells from the LUCA. Free-living cells occupied all chambers of the original vent. All replicative intermediates described in Figs. 2–11, including proto-cells, suffered from a shortage of energy and resources. Most intermediates went extinct, except for some primitive viruses (Fig. 13B).

(I) Key enzymes and proteins. Key theories, cell structures, and enzymes/proteins are marked by closed rectangles with a variety of colors. Kornberg stated "*I have never met a dull enzyme*" [76]. Enzymes that arose during the crystallization process in the original vent contributed to various intermediates, including free-living cells. Each innovation dramatically changed the destinies of developing life, symbolically marked as the double-edged sword.

(J) Not awarded a "Nobel prize". Advances in blue boxes have been awarded Nobel prizes, whereas those shown in red and pink boxes, which are keys for the emergence of free-living cells, have not. Forterre stated that "*the discovery of these enzymes (topoisomerases) was a great leap forward in our understanding of cell biology which may have deserved a Nobel Prize*" [77]. We believe that the key findings in the red and pink boxes, including the evolution of topoisomerases, are critical achievements.

Supplementary Fig. S15. Possible *in vitro* and *in silico* experiments to test the hypotheses presented in this study. Corresponding figure numbers are marked in blue.

(A) Establish synthetic cells with U-DNA. Since *E. coli* MUG (Fig. S8) can repair both T/G and U/G mismatches caused by the deamination of 5-methylcytosine and cytosine, deleting genes encoding TS, dUTPase, or uracil glycosylase could generate cells carrying U-DNA by a synthetic biology approach (Fig. 7B).

(B) Liquid droplets nucleated around compacted DNA. *In vitro* tests are neeed to evaluate whether compacted DNA can nucleate and form a coacervate and whether such a coacervate could recruit lipid components and yield a proto-cell-like structure surrounded by a permeable membrane. Many experiments using droplets [78-80] provide a basis for these analyses.

(C) Physical nature of channels, pumps, and ATPases on a permeable membrane. *In vitro* reconstruction of Bejan's "constructal law" on proto-cells is possible using vesicles with a permeable membrane, extant membrane bound proteins, and endless supplies of electrons, protons, and resources. We predict that extant membrane-bound proteins function efficiently under such conditions.

(D) Toward maximum entropy. Whether the principal of "*omnis cellula a cellula*" applies to proto-cells in the chamber of the alkaline vent under the second law of thermodynamics is an important question and can be evaluated by *in silico* experiments.

(E) Postulated vesicle transport system in the ancient vent. Although a variety of membrane-bound proteins evolved in the ancient vent, the transport of essential macromolecules inside and outside of proto-cells might be difficult. Thus, we hypothesized that a vesicle transport system was established in the ancient vent (Fig. S10). This hypothesis can be evaluated by *in silico* experiments.

(F) Reconstruction of the postulated final crystallization process. Using advanced synthetic biology approaches [81, 82], artificial fusion among proto-cells containing an incomplete genome could yield free-living cells. Artificially constructed incomplete genomes A and B can be introduced into each host cells. C and D, which are genomes for host cells, should be degraded during synthetic biology approaches.

(3) Supplementary references

1. Scholzen, T.; Arndt, E. Organization and nucleotide sequence of ten ribosomal protein genes from the region equivalent to the spectinomycin operon in the archaebacterium *Halobacterium marismortui*. *Mol. Gen. Genet.* **1991**, *228*, 70–80.
2. Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **2003**, *1*, 127–136.
3. Tuller, T.; Birin, H.; Gophna, U.; Kupiec, M.; Ruppin, E. Reconstructing ancestral gene content by coevolution. *Genome Res.* **2010**, *20*, 122–132.
4. Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Höbartner, C.; Cramer, P. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **2021**, *12*, 279.
5. Woese, C. The universal ancestor. *Proc. Natl. Acad. Sci. USA*. 1998, *95*, 6854–6859.
6. Koonin, E. V.; Martin, W. T. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **2005**, *21*, 647–654.
7. Powner, M. W.; Zheng, S. L.; Szostak, J. W. Multicomponent assembly of proposed DNA precursors in water. *J. Am. Chem. Soc.* **2012**, *134*, 13889–13895.
8. Dragičević, I.; Barić, D.; Kovačević, B.; Golding, B. T.; Smith, D. M. Non-enzymatic ribonucleotide reduction in the prebiotic context. *Chemistry* **2015,** *21*, 6132–6143.
9. Kim, S. C.; Zhou, L.; Zhang, W.; O'Flaherty, D. K.; Rondo-Brovetto, V.; Szostak, J. W. A model for the emergence of RNA from a prebiotically plausible mixture of ribonucleotides, arabinonucleotides, and 2'-deoxynucleotides. *J. Am. Chem. Soc.* **2020**, *142*, 2317–2326.
10. Xu, J.; Chmela, V.; Green, N. J.; Russell, D. A.; Janicki, M. J.; Góra, R. W.; Szabla, R.; Bond, A. D.; Sutherland, J. D. Selective prebiotic formation of RNA pyrimidine and DNA purine nucleosides. *Nature* **2020**, *582*, 60–66.
11. Loc'h, J.; Gerodimos, C. A.; Rosario, S.; Tekpinar, M.; Lieber, M. R.; Delarue, M. Structural evidence for an in trans base selection mechanism involving Loop1 in polymerase μ at an NHEJ double-strand break junction. *J. Biol. Chem.* **2019**, *294*, 10579–10595.
12. Acharya, N.; Khandagale, P.; Thakur, S.; Sahu, J. K.; Utkalaja, B. G. Quaternary structural diversity in eukaryotic DNA polymerases: monomeric to multimeric form. *Curr. Genet.* **2020**, *66*, 635–655.
13. 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.

14. de Farias, S. T.; Furtado, A. N. M.; Dos Santos Junior, A. P.: José, M. V. Natural history of DNA-dependent DNA polymerases: Multiple pathways to the origins of DNA. *Viruses* **2023**, *15*, 749.
15. 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.
16. Symington, L. S. End resection at double-strand breaks: mechanism and regulation. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016436.
17. Paull, T. T. 20 years of Mre11 biology: No end in sight. *Mol. Cell.* **2018**, *71*, 419–427.
18. Stasiak, A.; Di Capua, E. The helicity of DNA in complexes with recA protein. *Nature* **1982**, *299*, 185–186.
19. 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.
20. Cunningham, R. P.; DasGupta, C.; Shibata, T.; Radding, C. M. Homologous pairing in genetic recombination: recA protein makes joint molecules of gapped circular DNA and closed circular DNA. *Cell* **1980**, *20*, 223–235.
21. Murayama, Y.; Kurokawa, Y.; Mayanagi, K.; Iwasaki, H. Formation and branch migration of Holliday junctions mediated by eukaryotic recombinases. *Nature* **2008**, *451*, 1018–1021.
22. Ohta, T. Simulation studies on the evolution of amino acid sequences. *J. Mol. Evol.***1976**, *8*, 1–12
23. Ariyoshi, M.; Vassylyev, D. G.; Iwasaki, H.; Nakamura, H.; Shinagawa, H.; Morikawa, K. Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from *E. coli*. *Cell* **1994**, *78*, 1063–1072
24. Sale, J. E.; Lehmann, A. R.; Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell. Biol.* **2012**, *13*, 141–152
25. 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
26. Adachi, N.; Senda, T.; Horikoshi, M. Uncovering ancient transcription systems with a novel evolutionary indicator. *Sci. Rep.* **2016**, *6*, 27922.
27. Yang, G.; Zheng, R. Y.; Tan, Q.; Dong, C. J.; Jin, Z. S. Clinical characteristics and responses to chemotherapy and immune checkpoint inhibitor treatment for microsatellite instability gastric cancer. *Am. J. Cancer Res.* **2020**, *10*, 4123–4133
28. Ohno, S.; Wolf, U.; Atkin, N. B. Evolution from fish to mammals by gene duplication. *Hereditas* **1968**, *59*, 169–187.

29. Jacob, F.; Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **1961**, *3*, 318–356
30. Sugiura, M.; Okamoto, T.; Takanami, M. RNA polymerase sigma-factor and the selection of initiation site. *Nature* **1970**, *225*, 598–600.
31. Horikoshi, M.; Wang, C. K.; Fujii, H.; Cromlish, J. A.; Weil, P. A.; Roeder, R. G. Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. *Nature* **1989**, *341*, 299–303.
32. Blattner, F. R.; Plunkett III, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. The complete genome sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1474.
33. Masukata, H.; Tomizawa, J. Control of primer formation for ColE1 plasmid replication: conformational change of the primer transcript. *Cell* **1986**, *44*, 125–136.
34. Formosa, T.; Alberts, B. M. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell* **1986**, *47*, 793–806.
35. Fuller, R. S.; Kornberg, A. Purified dnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. *Proc. Natl. Acad. Sci. USA.* **1983**, *80*, 5817–5821.
36. Bell, S. P.; Mitchell, J.; Leber, J.; Kobayashi, R.; Stillman, B. The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **1995**, *83*, 563–568.
37. Bramhill, D.; Kornberg, A. A model for initiation at origins of DNA replication. *Cell* **1987**, *54*, 915–918.
38. Blow, J. J.; Laskey, R. A. A role for the nuclear envelope in controlling DNA replication within the cell cycle. *Nature* **1988**, *332*, 546–548.
39. Kubota, Y.; Mimura, S.; Nishimoto, S.; Takisawa, H.; Nojima, H. Identification of the yeast MCM3-related protein as a component of *Xenopus* DNA replication licensing factor. *Cell* **1995**, *81*, 601–609.
40. Nishitani, H.; Lygerou, Z.; Nishimoto, T.; Nurse, P. The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* **2000**, *404*, 625–628
41. Tada, S.; Li, A.; Maiorano, D.; Méchali, M.; Blow, J. J. Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. *Nat. Cell Biol.* **2001**, *3*, 107–113.

42. Hayashi, C.; Miyazaki, E.; Ozaki, S.; Abe, Y.; Katayama, T. DnaB helicase is recruited to the replication initiation complex via binding of DnaA domain I to the lateral surface of the DnaB N-terminal domain. *J. Biol. Chem.* **2020**, *295*, 11131–11143.
43. Jacob, F. On regulation of DNA replication in Bacteria. *Cold Spring Harbor Symposia on Quantttative Biology.* **1963**, *28*, 329.
44. Michel, B.; Bernander, R. Chromosome replication origins: do we really need them? *Bioessays* **2014**, *36*, 585–590.
45. Masai, H. Replicon hypothesis revisited. *Biochem. Biophys. Res. Commun.* **2022**, *633*, 77–80.
46. Dean, F. B.; Bullock, P.; Murakami, Y.; Wobbe, C. R.; Weissbach, L.; Hurwitz, J. Simian virus 40 (SV40) DNA replication: SV40 large T antigen unwinds DNA containing the SV40 origin of replication. *Proc. Natl. Acad. Sci. USA.* **1987**, *84*, 16–20.
47. Seki, M.; Enomoto, T.; Eki. T.; Miyajima, A.; Murakami, Y.; Hanaoka, F.; Ui, M. DNA helicase and nucleoside-5'-triphosphatase activities of polyoma virus large tumor antigen. *Biochemistry* **1990**, *29*, 1003–1009.
48. Cortez, D.; Quevillon-Cheruel, S.; Gribaldo, S.; Desnoues, N.; Sezonov, G.; Forterre, P.; Serre, M. C. Evidence for a Xer/dif system for chromosome resolution in archaea. *PLoS Genet.* **2010**, *6*, e1001166.
49. Kornberg, A.; Baker, T. *DNA Replication, Second Edition.* Freeman, New York. 1992.
50. Moe, E.; Leiros, I.; Smalås, A. O.; McSweeney, S. The crystal structure of mismatch-specific uracil-DNA glycosylase (MUG) from *Deinococcus radiodurans* reveals a novel catalytic residue and broad substrate specificity. *J. Biol. Chem.* **2006**, *281*, 569–577.
51. Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738.
52. Lane, N.; Allen, J. F.; Martin, W. How did LUCA make a living? Chemiosmosis in the origin of life. *Bioessays* **2010**, *32*, 271–280
53. Martin, W. F.; Sousa, F. L.; Lane, N. Evolution. Energy at life's origin. *Science* **2014**, *344*, 1092–1093.
54. Tkach, M.; Théry, C. Communication by extracellular vesicles: Where we are and where we need to go. *Cell* **2016**, *164*, 1226–1232.
55. Mathieu, M.; Martin-Jaular, L.; Lavieu, G.; Théry, C. Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. *Nat. Cell. Biol.* **2019**, *21*, 9–17.
56. Hoffmann, M. A. G.; Yang, Z.; Huey-Tubman, K. E.; Cohen, A. A.; Gnanapragasam, P. N. P.; Nakatomi, L. M.; Storm, K. N.; Moon, W. J.; Lin, P. J. C.; West Jr, A. P.;

Bjorkman, P. J. ESCRT recruitment to SARS-CoV-2 spike induces virus-like particles that improve mRNA vaccines. *Cell* **2023**, *186*, 2283–2492.

57. Gill, S.; Catchpole, R.; Forterre, P. Extracellular membrane vesicles in the three domains of life and beyond. *FEMS Microbiol. Rev.* **2019**, *43*, 273–303.
58. 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.
59. Quemin, E. R.; Chlanda, P.; Sachse, M.; Forterre, P.; Prangishvili, D.; Krupovic, M. Eukaryotic-like virus budding in archaea. *mBio* **2016**, *7*, e01439-16.
60. 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.
61. Margulis, L. Symbiosis and evolution. *Sci. Am.* **1971**, *225*, 48–57.
62. 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.
63. 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.
64. Guglielmini, J.; Woo, A. C.; Krupovic, M.; Forterre, P.; Gaia, M. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc. Natl. Acad. Sci. USA*. **2019**, *116*, 19585–19592
65. Mattiroli, 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.
66. Liu, Y.; Bisio, H.; Toner, C. M.; Jeudy, S.; Philippe, N.; Zhou, K.; Bowerman, S.; White, A.; Edwards, G.; Abergel, C.; Luger, K. Virus-encoded histone doublets are essential and form nucleosome-like structures. *Cell* **2021**, *184*, 4237–4250
67. 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
68. Seki, M. On the origin of the genetic code. *Genes Genet. Syst.* **2023**, *98*, 9–24

69. Bartel, D. P.; Doudna, J. A.; Usman, N.; Szostak, J. W. Template-directed primer extension catalyzed by the *Tetrahymena* ribozyme. *Mol. Cell. Biol.* **1991**, *11*, 3390–3394.
70. Been, M. D.; Cech, T. R. RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science* **1988**, *239*, 1412–1416.
71. McGinness, K. E.; Joyce, G. F. In search of an RNA replicase ribozyme. *Chem. Biol.* **2003**, *10*, 5–14.
72. Greene, B. L.; Kang, G.; Cui, C.; Bennati, M.; Nocera, D. G.; Drennan, C. L.; Stubbe, J. Ribonucleotide reductases: Structure, chemistry, and metabolism suggest new therapeutic targets. *Annu. Rev. Biochem.* **2020**, *89*, 45–75
73. Oparin, A. I. *The origin of life.* McMillan, New York, USA. 1938.
74. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523.
75. Bejan, A.; Lorente, S. The constructal law and the evolution of design in nature. *Phys. Life Rev.* **2011**, *8*, 209–240.
76. Kornberg, A. Never a dull enzyme. *Annu. Rev. Biochem.* **1989**, *58*, 1–30
77. Forterre, P.; Gribaldo, S.; Gadelle, D.; Serre, M. C. Origin and evolution of DNA topoisomerases. *Biochimie* **2007**, *89*, 427–446.
78. Courchaine, E. M.; Lu, A.; Neugebauer, K. M. Droplet organelles? *EMBO J.* **2016**, *35*, 1603–1612.
79. Feng, Z. Y.; Liu, T. T.; Sang. Z. T.; Lin, Z. S.; Su, X.; Sun, X. T.; Yang, H. Z.; Wang, T.; Guo, S. Microfluidic preparation of janus microparticles with temperature and pH triggered degradationproperties. *Front. Bioeng. Biotechnol.* **2021**, *9*, 756758.
80. Xu, C.; Martin, N.; Li, M.; Mann, S. Living material assembly of bacteriogenic protocells. *Nature* **2022**, *609*, 1029–1037.
81. Olivi, L.; Berger, M.; Creyghton, R. N. P.; De Franceschi, N.; Dekker, C.; Mulder, B. M.; Claassens, N. J.; Ten Wolde, P. R.; van der Oost, J. Towards a synthetic cell cycle. *Nat. Commun.* **2021**, *12*, 4531.
82. Venter, J. C.; Glass, J. I.; Hutchison, CA 3rd.; Vashee, S. Synthetic chromosomes, genomes, viruses, and cells. *Cell* **2022**, *185*, 2708–2724.