
Proposal for a Minimal DNA Auto-Replicative System

Agustino Martinez-Antonio,
Laura Espindola-Serna and Cesar Quiñones-Valles

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/51986>

1. Introduction

DNA replication allows cell division and population growth of living organisms. Here we will focus on DNA replication in prokaryotic single celled microorganisms. Several excellent reviews of the molecular processes that carry out DNA replication in bacteria already exist, *E. coli* being the model described in most detail (Langston LD et al., 2009; Quiñones-Valles et al., 2011). Briefly, the process begins when DnaA (DNA initiator replication protein) in its activated form (DnaA-ATP) recognizes and binds the *oriC* (origin of replication on the bacterial chromosome). In the following step, the replisome is assembled and binds to the complex of DnaA-ATP at the *oriC*. Next, the DNA strands are separated and synthesis of the complementary strands initiates followed by elongation steps. The molecular mechanisms of elongation differ depending on the strand used as a template; the leading strand is replicated continuously starting from a unique RNA primer, whereas on the lagging strand DNA polymerase III must recognize several RNA primers, previously synthesized by DnaG, and then replicate each DNA fragment (Okazaki fragments). This is followed by the replacement of RNA primers by DNA polymerase I, and removal of nicks by a DNA ligase. The whole process concludes when replisomes reach the *ter* site, almost opposite to *oriC* on the circular DNA molecule. Tus proteins are attached to the *ter* sites and when replisomes reach these complexes, they collide and finally are disassembled (see Figure 1 for an overview of the whole process).

From another aspect, one of the more challenging areas of Synthetic Biology is the design and construction of minimal cells. The accomplishment of this aim might contribute to answering basic questions about the minimal components necessary to sustain life systems, in addition to cell auto-organization, function and evolution. In a practical application, mini-

mal cells can be used as a background chassis for the generation of dedicated biological systems designed for the synthesis or degradation of diverse compounds of interest.

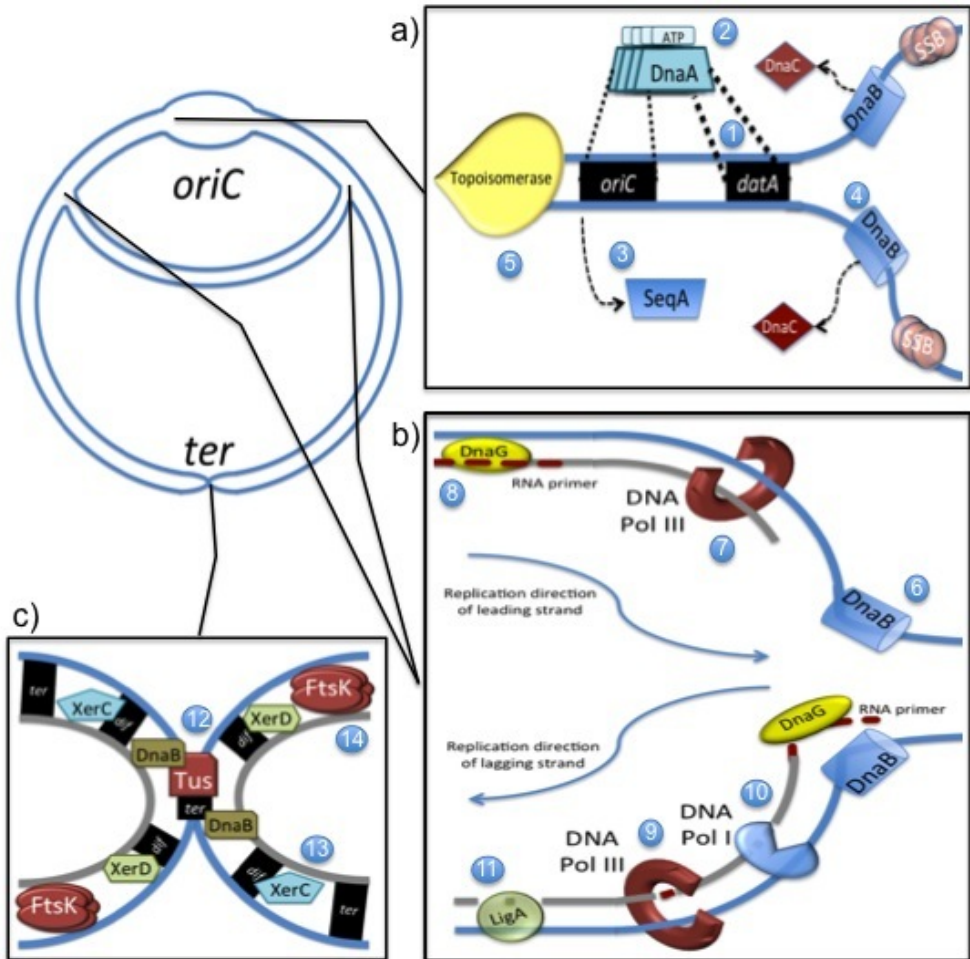


Figure 1. Main steps of DNA replication in bacteria. a) Initiation of DNA replication; the *datA* locus has a high affinity for binding DnaA (1). DnaA binds to ATP, homo-multimers of DnaA-ATP are formed (2). These homo-multimers bind to *oriC* and once replication is initiated SeqA binds this region and prevents initiation of a new replication event (3). The SSB (single strand binding) protein and DnaB assist the complex to open the DNA strands and release DnaC (4). A DNA topoisomerase helps to further unfold the DNA strands (5). b) The elongation phase; the replication fork is formed and the replisome is assembled (6). DNA polymerase III replicates the leading strand (7). DnaG incorporates RNA primers as primers for replication of the lagging strand (8). Polymerase III can now replicate Okazaki fragments on the lagging strand (9). DNA polymerase I replaces RNA nucleotides for DNA nucleotides (10). A DNA ligase (LigA) seals the nicks on contiguous DNA fragments (11). c) Termination of DNA replication; The protein Tus binds to the *ter* sites, when replisomes reach Tus, replication ceases (12). The recombinases XerC and XerD resolve the replicated DNA strands (13). Finally, FtsK translocates the DNA strands and each double-stranded DNA molecule can be liberated (14).

In recent years, the essential properties and capabilities necessary to develop minimum cells have been broadly speculated (MacDonald et al., 2011). Among these characteristics it is evident that DNA replication should be a fundamental property of these biosystems. Many genes for DNA replication are found to be conserved when comparative analysis of bacterial genomes is carried out. These types of genes are considered as informational genes, in charge of maintaining the genetic code, and are among the genes less frequently found to be horizontally transferable (Jain et al, 1999). Therefore by genomic comparisons and functional analyses it is possible to propose a minimum core of genes capable of supporting the process of DNA replication.

From a genetic point of view, and for the purpose of this study it is important to state our definition of a minimal DNA auto-replicative system (MiDARS) as: *a genetic system comprising the minimum number of DNA components, including regulatory elements and gene products necessary for the auto-replication of the DNA molecule on which they are encoded, functioning in an in vitro condition.*

In this chapter we will develop a proposal for the construction of such an auto-replicative DNA system. This system is designed to serve as a scaffold for the incorporation of additional biological functions such as transcription and translation, etc. For the scaffold design we exploit information of genes necessary for replication in *E. coli* that are highly conserved in bacteria with extremely reduced genomes and analyze their functional role in DNA replication in order to finally propose a minimal genetic system with a DNA auto-replicative function.

2. Minimal cells and minimal genetic systems

A minimal cell can be defined as a biological system that has the minimal number of genetic parts and molecular components for supporting life functions under defined growth conditions. In other words, it includes only the necessary number of genes and derived biomolecular machinery that are considered basic to support life functions (Jewett and Forster, 2010).

The concept of life is intrinsically complex; in biochemical terms it could be defined by three basic characteristics (Luisi et al., 2006):

1. auto-regulation of metabolism,
2. auto-replication of the genetic material and,
3. controlled evolution of their components and functions.

The design and synthesis of minimal cells depends on the environmental conditions the systems will be exposed to. Initially, we might consider that a minimal cell should be exposed to the most favorable conditions in order to facilitate its conception and function. These favorable conditions will require an environment where the cell is not suffering any kind of environmental stress. Nonetheless, even this ideal scenario is a challenging condition to direct the rational design of components of a minimal cellular system since the genes for many

cellular functions are not yet totally defined. What we could do is to start to reconstruct minimal biological functions that are more or less well defined. These might be the processes relating to the central dogma of molecular biology: DNA replication, DNA transcription and mRNA translation (Figure 2). Some of these functions have been the object of different studies; e. g. transcription and translation were successfully recreated in the experiment of Asahara (2010) by separately expressing the components of the *E. coli* RNA polymerase, including the sigma70 factor and reconstituting the function of the complete enzyme *in vitro*.

Since one of the fundamental characteristics of life systems is the replication of their own genetic material, we can consider the design of minimal genetic systems that sustain DNA auto-replication as an important starting point.

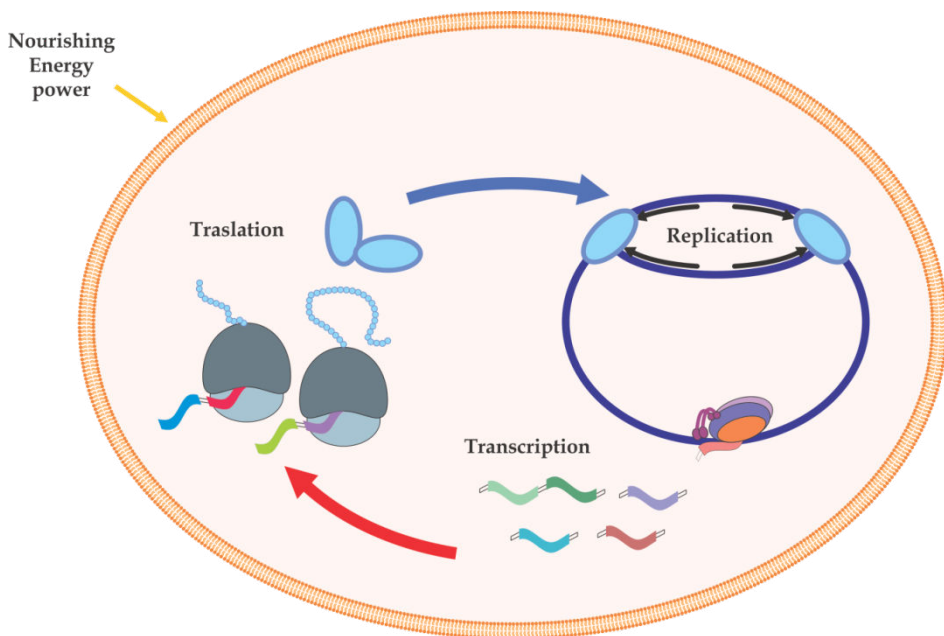


Figure 2. Representation of a hypothetical minimal auto-replicative system. One of the key features of the minimal cell is that it should perform basic functions such as transcription, translation and replication of the genetic information contained in its genome.

3. Approaches for the development of minimal genetic systems

Currently there are two approaches for the study of minimal biological systems. These are the *top down* and *bottom up* strategies (Delays L & Moya A, 2009; Murtas, 2009). The *top down* approach considers the analysis of existing biological systems and, by following a reductionist approach, looks to minimize the number of components either by searching for con-

served genetic elements or by experimentally reducing the genome without losing functionality. This strategy was used to reduce the *E. coli* genome by 15% by deleting non-essential genes, recombinogenic and mobile DNA elements, and cryptic genes. The resulting cells had good growth profiles and showed improved performance for protein production (Pósfai et al., 2006). Another focus of this approach is to carry out comparative genomics and define a set of conserved genes such as those in charge of specific functions (Gil et al., 2004; Forster & Church, 2006).

On the other hand, the *bottom up* approach involves the construction of complex systems starting from relatively simple molecular precursors. A classical example is the experiment of Miller, who obtained amino acids from a mixture of simple organic and inorganic molecules (Miller, 1953).

Considering the design and construction of minimal genetic systems, benefits should be obtained by employing both complementary *top down* and *bottom up* approaches.

4. *Escherichia coli* as a model organism for the design of a minimal DNA auto-replicative system

Escherichia coli is a bacillary Gram-negative, aerobic, facultative and non-sporulating organism. It was discovered in 1885 by the physician Theodore von Escherich and is now classified as part of the Enterobacteriaceae family of the Gamma-proteobacterias (Blattner et al., 1997).

This bacterium lives in the intestine of mammals, and assists its hosts with assimilation of nutrients, providing some vitamins and preventing the establishment of bacterial pathogens. Since its discovery, *E. coli* has been widely used as a working model in the laboratory to study biochemistry and diverse molecular processes. In addition, it has been widely used in biotechnology as a vehicle for the expression of multiple recombinant proteins and whole metabolic pathways.

Arthur Kornberg was one of the most prominent investigators in molecular biology and a pioneer in the description of the replicative process using *E. coli* as a model. For his accomplishments in the field he was awarded the Nobel Prize in Physiology and Medicine in 1959. He discovered DNA polymerase I (Bessman et al., 1958; Lehman et al., 1958a), and describes the synthesis of DNA as a process based on the use of a single strand of DNA as a template (Lehman et al., 1958b). Later, Kornberg and his collaborators discovered additional enzymes involved in DNA replication: DNA primase, DNA helicase, DnaA, PriA among others. Nowadays, the replication process and the replicative enzymes of *E. coli* are the best understood and characterized of any organism.

From a biotechnological standpoint, *E. coli* shows three important characteristics that make it an ideal organism to serve as the platform for the design of a synthetic cellular program (Foley & Shuler, 2010):

1. functionally it is the organism best characterized at the molecular and biochemical levels in terms of components of metabolism,
2. it has proven to be a robust vehicle for the expression of multiple biotechnological processes,
3. it has a short growth cycle and is easy to manipulate genetically.

Additionally, the genome of *E. coli* serves as the principal source of standardized genetic parts for the construction of genetic circuits, the “BioBricks”, in a project whose aim is to standardize genetic parts to facilitate biological engineering (<http://partsregistry.org>), (Smolke, 2009). Most biobricks are designed to function in *E. coli*, therefore, we think *E. coli* is the best organism of choice for the design of a DNA auto-replicative system.

5. Comparison of the DNA replicative machinery of *E. coli* with that of bacteria with reduced genomes

Comparative genomics is a powerful approach that allows the identification of genetic sequences sharing identity/similarity among different organisms. Through these comparisons it is possible to identify conserved genes and predict the components of the replicative machinery in several different organisms.

For our purpose, among the organisms of interest to consider in our design are those with extremely reduced genomes. A characteristic of these organisms is that they are incapable of growth in a free-living manner. The genomes of organisms with these characteristics correspond to those having the minimum number of genes possible in nature. From these we chose the 25 organisms with the most reduced genomes known to date (Table 1). All of these genomes contain less than 1,200 kbp of DNA and all are endosymbiotic bacteria, most of which are thought to survive at the expense of the host.

In these organisms, we searched for genes encoding enzymes involved in DNA replicative functions with orthology to the replicative machinery from *E. coli* (Table 2). To find orthologous genes we followed two complementary strategies: we looked for Clusters of Orthologous Groups (COGs, Tatusov et al., 2003) and also used bidirectional best blast hits (Moreno-Hagelsieb & Latimer, 2008). In Table 2 the blue cells indicate where genes orthologous to *E. coli* are present in the target organism. In the table we show orthologous genes to be present in at least fifteen of these bacteria. Remarkably, bacteria with the most reduced genomes; *Carsonella ruddii* PV (Nakabachi et al., 2006; Tamames et al., 2007), *Hodgkinia cicadicola* Dsem (McCutcheon et al., 2009) and *Tremblaya princeps* PCIT (López-Madrigal et al., 2011; McCutcheon & Moran, 2011) had only 5, 3 and 5 genes related to replication respectively which were orthologous to *E. coli*. These three organisms are strict endosymbionts of insects, with the smallest genomes known to date (Table 2). The fact that these bacteria showed fewer genes related to DNA replication in comparison to bacteria with larger genomes (Figure 3), indicates that the minimal replicative machinery in these organisms might be composed by a small number of constituents. This observation raises many open questions, for instance:

| Organism | Abrev | Total length | Total genes | Replic. genes | Classification | Symbiotic Relation/interaction | Host |
|---|-------|--------------|-------------|---------------|----------------------|--------------------------------|---|
| <i>Escherichia coli</i> K-12 MG1655 | eco | 4,639,675 | 4,146 | 228 | Proteobacteria-gamma | Free living | <i>Homo sapiens</i> |
| <i>Anaplasma marginale</i> St. Maries | ama | 1,197,687 | 948 | 55 | Proteobacteria-alpha | Parasitic, Endosymbiotic | <i>Homo sapiens</i> |
| <i>Mitochondria</i> IricVA | mmn | 1,183,732 | 1,211 | 65 | Proteobacteria-alpha | Pathogen | <i>Ixodes ricinus</i> |
| <i>Chlamydomonas reinhardtii</i> G115 | cca | 1,181,356 | 1,005 | 85 | Chlamydiae | Parasitic, Endosymbiotic | <i>Homo sapiens</i> , swine |
| <i>Ruthnia magnifica</i> Cm | rma | 1,160,782 | 976 | 61 | Proteobacteria-gamma | Endosymbiotic | <i>Calypogena magnifica</i> |
| <i>Treponema pallidum</i> Nichols | tpa | 1,138,011 | 1,036 | 74 | Spirochaetes | Host-associated Free living | <i>Homo sapiens</i> |
| <i>Rickettsia prowazekii</i> Madrid E | rpr | 1,111,523 | 835 | 85 | Proteobacteria-alpha | Parasitic-endosymbiotic | <i>Homo sapiens</i> |
| <i>Chlamydia trachomatis</i> D/UW- | ctr | 1,042,519 | 895 | 83 | Chlamydiae | Parasitic, Endosymbiotic | <i>Homo sapiens</i> |
| <i>Vesicomycosocius okutanii</i> HA | vok | 1,022,154 | 937 | 49 | Proteobacteria-gamma | Endosymbiotic | <i>Calypogena okutanii</i> |
| <i>Tropheryma whippelii</i> TW08/27 | tws | 925,938 | 783 | 68 | Actinobacteria | Pathogen | <i>Homo sapiens</i> |
| <i>Neorickettsia sennetsu</i> Miyayama | nse | 859,006 | 932 | 46 | Proteobacteria-alpha | Pathogen | <i>Homo sapiens</i> |
| <i>Mesoplasma florum</i> L1; ATCC 33453 | mfl | 793,224 | 682 | 85 | Firmicutes | Parasite, non-pathogen | <i>Citrus limon</i> |
| <i>Ureaplasma parvum</i> ATCC 700970 | uur | 751,719 | 614 | 52 | Tenericutes | Pathogen | <i>Homo sapiens</i> |
| <i>Blochmannia floridanus</i> | bfl | 705,557 | 583 | 37 | Proteobacteria-gamma | Endosymbiont | <i>Camponotus floridanus</i> |
| <i>Wigglesworthia glossinidia</i> | wbr | 703,004 | 617 | 41 | Proteobacteria-gamma | Endosymbiotic | <i>Glossina brevipalpis</i> |
| <i>Baumannia cicadellincola</i> Hc | bci | 686,194 | 595 | 46 | Proteobacteria-gamma | Endosymbiotic | Insect endosymbiont |
| <i>Buchnera aphidicola</i> APS | buc | 655,725 | 574 | 48 | Proteobacteria-gamma | Mutualistic, Endosymbiotic | <i>Acyrtosiphon pisum</i> |
| <i>Blattabacterium</i> sp. BPLAN | bpi | 640,442 | 581 | 39 | Bacteroidetes | Endosymbiotic | <i>Periplaneta americana</i> |
| <i>Riesia pediculicola</i> USDA | rip | 582,127 | 556 | 31 | Proteobacteria-gamma | Endosymbiotic | <i>Pediculus humanus</i> |
| <i>Mycoplasma genitalium</i> G37 | mge | 580,076 | 475 | 40 | Tenericutes | Host-associated | <i>Homo sapiens</i> |
| <i>Moranella endobia</i> PCIT | men | 538,294 | 406 | 41 | Proteobacteria-gamma | Host-associated | <i>Planococcus citri</i> |
| <i>Sulcia muelleri</i> GWSS | smg | 245,530 | 227 | 11 | Bacteroidetes | Endosymbiotic | Glassy-winged sharpshooter |
| <i>Zindera insecticola</i> CARI | zin | 208,564 | 202 | 15 | Proteobacteria beta | Endosymbiotic | Spittlebugs |
| <i>Carsonella ruddii</i> PV | cpr | 159,662 | 182 | 5 | Proteobacteria-gamma | Endosymbiotic intracellular | <i>Pachypsylla venusta</i> |
| <i>Hodgkinia cicadicola</i> Dsem | hci | 143,795 | 169 | 3 | Proteobacteria-alpha | Endosymbiotic | <i>Diceroprocta semicincta</i> |
| <i>Tremblaya princeps</i> PCIT | tpn | 138,927 | 121 | 5 | Proteobacteria-beta | Endosymbiotic | Mealybug <i>Planococcus citri</i> (Risso) |

Table 1. Bacteria with genome sizes less than 1200 kbp

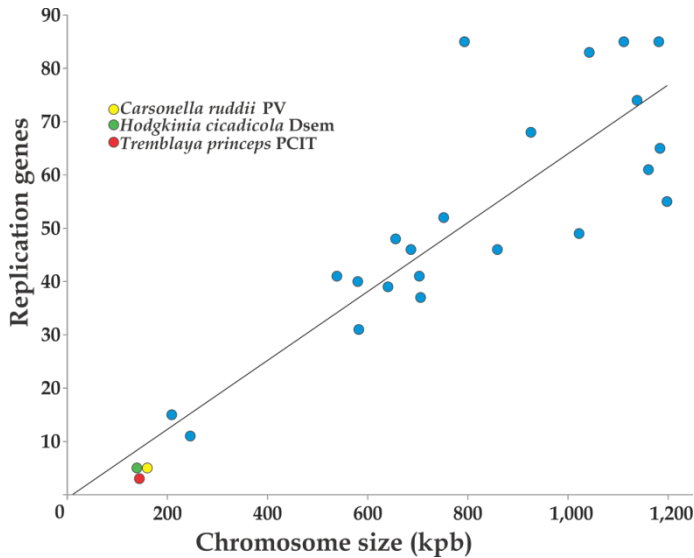


Figure 3. Conservation of DNA replicative machinery in bacteria with reduced genomes. Graph showing the relationship between number of genes annotated with DNA replicative functions versus genome sizes.

- i. Are these genes sufficient to sustain the process of replication of an entire chromosome?;
- ii. Does the host supply the missing elements for replication of the endosymbionts DNA? and,
- iii. Do these organisms use additional proteins in comparison to those currently described for the process of DNA replication?

The apparent requirement of only a handful of genes for DNA replication in extremely reduced genomes, compared with the 228 annotated in *E. coli*, might suggest a parsimonious mechanism of DNA replication in endosymbiont bacteria since they are always living in stable environments. The genes which are more highly conserved in both reduced genomes and *E. coli* are those whose products form the replisome (*dnaE*, *dnaB*, *dnaN*, *dnaG*, *dnaX*, *dnaQ*, *ssb*, *holA* and *holB*), the genes encoding for DNA topoisomerase type II (*gyrA* and *gyrB*), and the gene for the NAD(+)-dependent DNA-ligase, *ligA* (Figure 4).

6. Components of a Minimal DNA Auto-Replicative System (MiDARS)

Of the three organisms with the most reduced genomes in nature, *Carsonella ruddii* is the more closely related phylogenetically to *E. coli* (Nakabachi et al., 2006). For this reason in our design we used the information of the replicative machinery in *Carsonella ruddii* and the functions known in *E. coli*. For the physical construction of the systems, however, we will use genes from *E. coli* for two main reasons:

| Organisms/ Gene | e c o | r m a | t p a | c a a | a m a | r p r | t w s | t r | v o k | m m n | m f l | u r | w b r | b f l | b c i | n s e | b p i | r i p | b u c | m g e | m e n | s m g | z i n | c r p | h c i | t p n | |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------|-------------|-------------|-------------|--------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--|
| <i>dnaA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>priA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>dnaB</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>dnaG</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>ssb</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>dnaE</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>dnaN</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>dnaQ</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>dnaX</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>holA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>holB</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>polA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>ligA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>gyrA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>gyrB</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>topA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>recA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>rnhA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>ruvA</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>ruvB</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>tatD</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>ung</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>uvrD</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>yagF</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Table 2. Conservation of replicative genetic machinery in bacteria with less than 1200 kbp

1. it is difficult to obtain genomic DNA from *Carsonella ruddii* since it cannot be cultured *in vitro*.
2. *E. coli* is the best chassis for applications in synthetic biology as mentioned above and therefore adequate for the incorporation of additional functions.

For the design of the minimal DNA auto-replicative system we will attempt to include the minimal elements present in *Carsonella* and -in a conservative manner- those we presume as necessary to perform the process of DNA replication in *E. coli*. In addition to the coding sequences, it is also necessary to define the regulatory regions of the genes and we propose to conserve the operative regions as defined for *E. coli* with the future aim of expanding the minimal functions of *E. coli* including the regulatory functions. Other important regions to include in the design are: the DNA replication origin (*oriC*) and the signals for termination of replication. Below we propose the genetic components that would constitute a MiDARS.

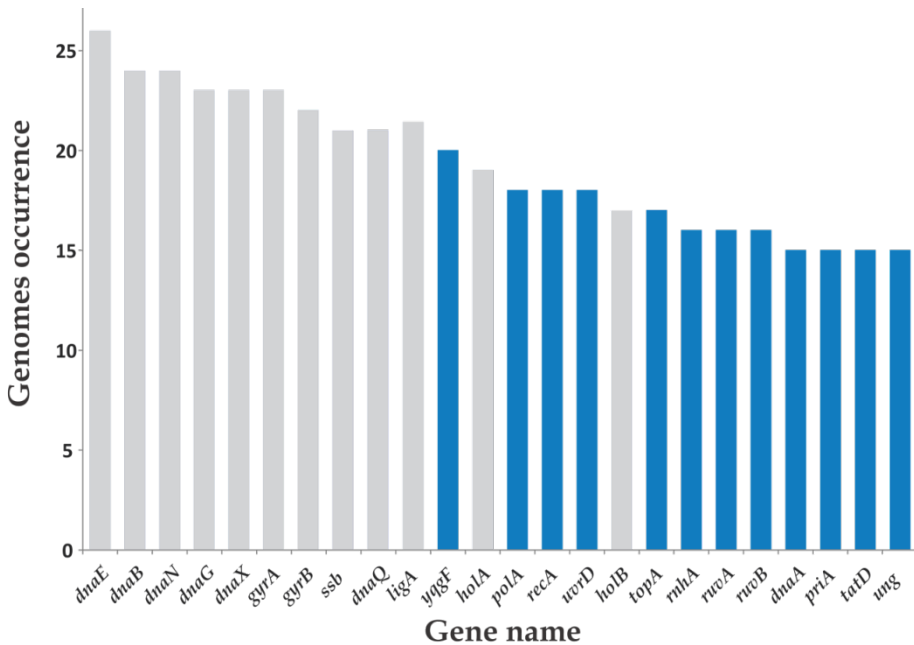


Figure 4. Conservation of genes for DNA replication in 25 reduced genomes and in *E. coli*. Grey bars represent the genes proposed to be essential for DNA auto-replication in a minimal genetic system.

The DNA initiator protein (*dnaA*)

At the beginning of the replication process, check-point proteins have to recognize and unfold the initiation site for replication at *oriC*. In *E. coli* DnaA is the principal protein employed for this purpose and is highly conserved among bacteria with reduced genomes. Therefore, *dnaA* should be present in the MiDARS.

The DNA helicase (*dnaB*)

The next candidate gene is *dnaB*, which encodes a DNA helicase. The role of the product of this gene is to unwind the DNA strands, a very important process during the elongation stage of replication.

The DNA primase (*dnaG*)

The gene that encodes the primase (*dnaG*) should also be considered. It is important for the synthesis of the RNA primers that permit the elongation of new DNA strands.

The single strand stabilization protein (*ssb*)

Another important function is the stabilization of single strands, carried out by the SSB protein, encoded by the *ssb* gene.

The core components of DNA polymerase III (*dnaE* and *dnaQ*)

The gene for the α subunit (*dnaE*) of DNA polymerase III is present in all 25 organisms with reduced genomes and the gene for the ϵ subunit (*dnaQ*) in twenty-one. These proteins form part of the core of DNA polymerase III, which carries out the essential polymerization and proofreading activities during DNA synthesis.

The clamp components (*dnaX*, *holA*, *holB*, *dnaN*)

During the elongation stage, two very important structures are formed; the leader and slider clamps. The first has the function of anchorage between DNA polymerase III and the DNA helicase; (Reyes-Lamothe R. et al., 2010) allowing the synthesis of the DNA in a synchronized manner between the leading and lagging strand. It is composed of the following subunits (genes): τ (*dnaX*), γ (*dnaX*), δ (*holA*) and δ' (*holB*). The circular slider clamp is constituted by two β -subunits (both products of *dnaN* gene), that recognize and bind to DNA-RNA hybrids (Georgescu R. et al., 2010). The slider clamp assists the core of DNA pol III to bind the lagging strand and allows the extension of the Okazaki fragments.

The DNA ligase (*ligA*)

The function of a ligase is needed for sealing nicks formed when the RNA primers are removed and replaced by DNA in the Okazaki fragments on the lagging strand.

Type II DNA topoisomerase (*gyrA* and *gyrB*)

We consider that a relaxing system produced by a DNA helicase may be necessary. This could be provided by the DNA gyrase complex (Type II Topoisomerase) composed of the A (*gyrA*) and B (*gyrB*) subunits.

Protein for termination of replication (*tus*)

Although there are several proteins that could contribute to termination of DNA replication we think that in a minimal system, the action of Tus could be enough to ensure this.

Origin of DNA replication (*oriC*)

This DNA sequence of around 245 bp in *E. coli* (Tabata et al., 1983) is needed to enable the DnaA protein to initiate the process of DNA replication

Termination of DNA replication (*terB* and *terC*)

These sequences are used by the Tus proteins to form the trap which terminates DNA replication.

The proposed elements that constitute the auto-replicative system are also listed in Table 3. This proposal is somewhat similar to previous reports, where genes that could constitute a minimal cell based on a comparative genomics study among various endosymbionts are described (Gil et al., 2004). In the present study however we also considered the inclusion of the DNA regions for initiation and termination of replication, as well as the *dnaA*, *ssb* and *tus* genes.

| Gene/DNA element | Product | Size (bp. <i>E. coli</i>) |
|------------------|--|----------------------------|
| <i>dnaA</i> | Chromosomal replication initiator protein DnaA | 1404 |
| <i>dnaB</i> | Replicative DNA helicase | 1416 |
| <i>dnaG</i> | DNA primase | 1746 |
| <i>ssb</i> | Single-stranded DNA-binding protein | 537 |
| <i>dnaE</i> | DNA polymerase III α subunit | 3483 |
| <i>dnaN</i> | DNA polymerase III, β subunit | 1101 |
| <i>dnaQ</i> | DNA polymerase III ϵ subunit | 732 |
| <i>dnaX</i> | DNA polymerase III, τ and γ subunits | 1932 |
| <i>holA</i> | DNA polymerase III, δ subunit | 1032 |
| <i>holB</i> | DNA polymerase III, δ' subunit | 1005 |
| <i>polA</i> | DNA pol I 5'-3' and 3'-5' exonuclease ; 3'-5' polymerase | 2787 |
| <i>ligA</i> | DNA ligase, NAD(+)-dependent | 2016 |
| <i>gyrA</i> | DNA gyrase (type II topoisomerase), subunit A | 2628 |
| <i>gyrB</i> | DNA gyrase, subunit B | 2415 |
| <i>tus</i> | Termination DNA replication protein | 930 |
| <i>oriC</i> | DNA region for initiation, origin of replication | 245 |
| <i>ter</i> | DNA region for termination of replication | 23 |

Table 3. Components of a minimal DNA auto-replicative system.

7. Expression of the replicative proteins of the MiDARS

A primary condition for the operation of an auto replicative system is that the protein-machinery encoded in it should be expressed. For transcription of the assembled group of genes, we propose use the *E. coli* RNA polymerase and its transcription factor sigma70 since all the genes of the system have a sigma70 factor promoter. The essential components of the RNA polymerase and their sigma70 factors have previously been successfully expressed separately and their activity reconstituted as mentioned previously (Asahara & Chong, 2010). We propose these components can be assembled as an additional functional module whose activity can be assayed separately and subsequently integrated into the system. The resulting mRNA (16) could be translated in an *in vitro* system such as the Pure System™ (Ueda et al., 1992; Shimizu & Ueda, 2010); containing ribosomes, aminoacyl-tRNAs, chaperones and initiation, elongation and termination factors among other elements essential for translation. Once protein synthesis is completed, the products could initiate replication of the DNA molecule for which the addition of deoxynucleotide triphosphates (dNTPs) and the appropriate buffers will be necessary. The source of energy for the

system will be creatine phosphate with the creatine kinase enzyme as the regenerator (Shimizu et al., 2006). An outline for the operation of the DNA auto-replicative system is shown in Figure 5.

8. Perspectives

Previous efforts have been made to propose the design of minimal cells however this objective is still far from being accomplished. From the standpoint of Synthetic Biology, biological systems that are robust, predictable in performance and highly efficient are desired (Jewett & Forster, 2010). In this work, we present a proposal to build an auto-replicative DNA system as the first step toward the development of synthetic biosystems. Additional cellular processes will need to be designed and constructed in a modular way including: transcriptional and translational functions and a minimal metabolism in order to maintain cell growth and produce energy.

Once this first prototype has been constructed and tested for performance, some further reduced combinations of the proposed number of genes could be tested to determine the absolute minimum set of genes sufficient to sustain DNA auto-replication; e.g. the few genes present in *Carsonella ruddii* PV.

The system proposed in this work can be assembled using methodologies such as that used when working with Biobricks (Smolke, 2009). Once the mini-chromosome is assembled it could function in cell-free systems, in nucleated mini-cells (Adler et al., 1967), in spores that lack DNA (Siccardi et al., 1975), in micelles or lipidic vesicles, and in some commercial systems. An important achievement in this sense has previously been reported by another research group, namely DNA replication achieved by using the Phi29 DNA polymerase, inside a lipidic vesicle. In this report only one strand was linearly replicated and circularized (Kurihara, 2011).

The successful development of a DNA auto-replicative system as proposed here could be a very important platform for the development of synthetic biology and the potential for such a system is great:

1. in the refinement of biotechnological processes since cellular energy could be directed to the desired biosynthetic pathways;
2. in the study of synthetic or natural circuits at a higher resolution and sharpness due to the minimalization of cellular noise; and
3. to test some evolutionary hypotheses, such as the proposed components of last common ancestors and components of rudimentary first cells, among others.

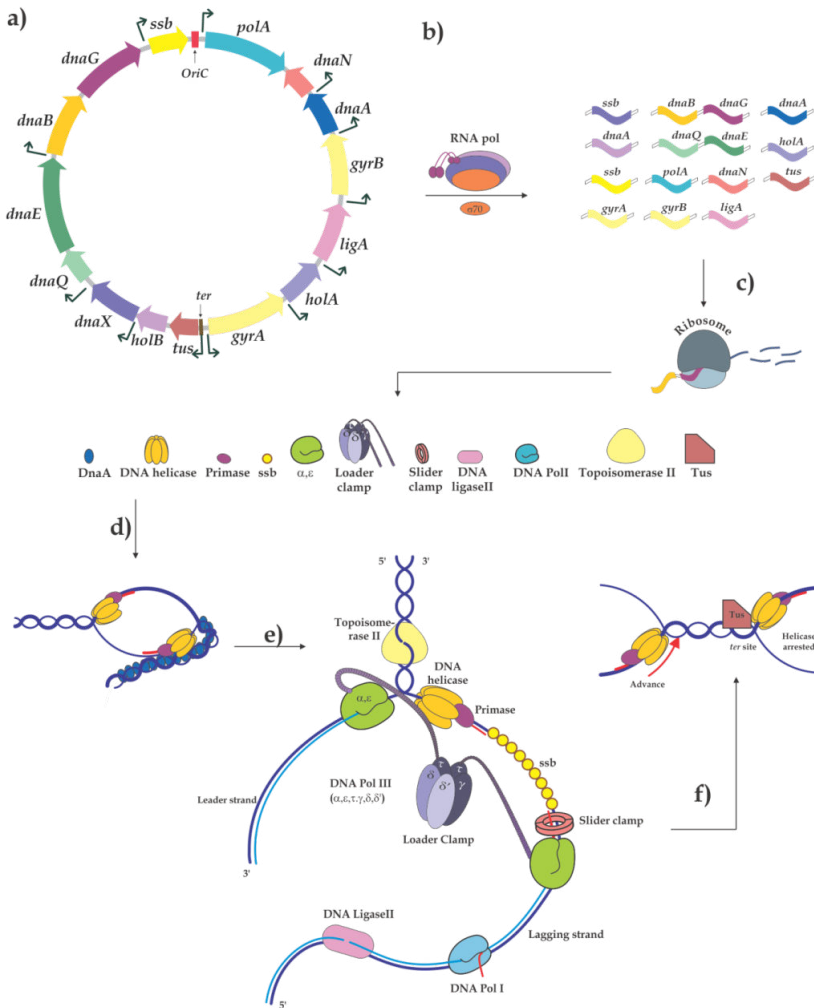


Figure 5. Proposal for the minimal components of a MiDARS and their function. a) The genetic system is a simplified version of a prokaryotic DNA mini-chromosome (25432 bp). The system contains the initiation region (*oriC*) and termination (*ter*) sites for DNA replication as well as a set of genes from *E. coli* (Table 4). The genes can be organized in the same order as in the native chromosome and contain their native operator regions to control expression. b) Transcription and, c) Translation can be carried out in solution using commercial kits (e.g. Pure System), RNA Polymerase and the *E. coli* sigma70 factor. d) The initiation of replication is regulated by DnaA-ATP and the helicase will join to the lagging strand in order to form the replication forks. The primase will bind to the helicase to carry out the synthesis of RNA primers that permit the activity of DNA pol III. The SSB protein stabilizes single strands of DNA. e) Two core subunits (α and ϵ) of the DNA Pol III, perform the elongation and proofreading of DNA. The DNA ligase and DNA Pol I replace the RNA primers, sealing the nicks between contiguous DNA fragments on the lagging strand. Topoisomerase II will relax the DNA template as the replication fork progresses. f) The Tus protein bound to the *ter* sites serves as a trap for the replicative machinery headed by the DNA helicase, stopping its movement and promoting the separation of the new MiDARS.

9. Conclusions

Here we propose a design for the construction of a minimal genetic system for DNA auto-replication. This proposal is based on the consideration of the latest knowledge of the details of the mechanisms and controls of DNA replication in *E. coli* and by taking into account the conservation of the replicative machinery in bacteria with extremely reduced genomes particularly those present in *Carsonella ruddii* PV.

The proposed auto-replicative device consists of 17 DNA elements (27822 bp including their operator regions) taken from the *E. coli* genome and incorporating the most conserved elements of the replicative machinery found in bacteria with extremely reduced genomes. These genetic elements will maintain their native operator and termination regions. Their products encode proteins encompassing the minimal number of predicted activities involved in DNA replication. Finally we propose some conditions in which the system might function.

Acknowledgments

Authors thank June Simpson for critical comments to the ms. This work was supported by CONACYT grants (102854 and 103686) given to AM-A. LE-S and CQ-V thank to CONACYT for PhD scholarships (208153 and 206011)

Author details

Agustino Martinez-Antonio, Laura Espindola-Serna and Cesar Quiñones-Valles

Departamento de Ingeniería Genética, Cinvestav, Irapuato-León, Irapuato Gto, México

References

- [1] Adler, H., Fisher, W. D., Cohen, A., & Hardigree, A. (1967). Miniature *Escherichia coli* cells deficient in DNA. *Proc. Natl. Acad. Sci. USA*, 57, 321-326.
- [2] Asahara, H., & Chong, S. (2010). In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. *Nucleic Acids Res* 3813:e141.
- [3] Bessman, M. J., Lehman, I. R., Simms, E. S., & Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction. *J Biol Chem* ., 233, 171-177.

- [4] Blattner, F. R., Plunkett, 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. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, 277, 1453-1462.
- [5] Delaye, L., & Moya, A. (2009). Evolution of reduced prokaryotic genomes and the minimal cell concept: Variations on a theme. *BioEssays*, 32, 281-287.
- [6] Foley, P. L., & Shuler, M. L. (2010). Considerations for the design and construction of a synthetic platform cell for biotechnological applications. *Biotechnol Bioeng*, 105, 26-36.
- [7] Forster, A. C., & Church, G. M. (2006). Towards synthesis of a minimal cell. *Mol Syst Biol* 2:45.
- [8] Georgescu, R., Yao, N. Y., & O'Donnell, O. (2010). Single-molecule analysis of the *Escherichia coli* replisome and use of clamps to bypass replication barriers. *FEBS Letters*, 584, 2596-2605.
- [9] Gil, R., Silva, F. J., Peretó, J., & Moya, A. (2004). Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev*, 68, 518-37.
- [10] Jain, R., Rivera, M. C., & Lake, J. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *PNAS*, 96, 3801-3806.
- [11] Jewett, M. C., & Forster, A. C. (2010). Update on designing and building minimal cells. *Curr Opin Biotechnol*, 21, 697-703.
- [12] Keyamura, K., & Katayama, T. (2011). DnaA protein DNA-binding domain binds to Hda protein to promote inter-AAA+ domain interaction involved in regulatory inactivation of DnaA. *J Biol Chem*, 286, 29336-29346.
- [13] Kurihara, K., Tamura, M., Shohda, K., Toyota, T., Suzuki, K., & Sugawara, T. (2011). Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. *Nature Chemistry*, 3, 775-781.
- [14] Langston, L. D., Indiani, C., & O'Donnell, M. (2009). Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle*, 8, 2686-91.
- [15] Lehman, I. R., Bessman, M. J., Simms, E. S., & Kornberg, A. (1958a). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J. Biol Chem*, 233, 163-170.
- [16] Lehman, I. R., Zimmerman, S. B., Adler, J., Bessman, M. J., Simms, E. S., & Kornberg, A. (1958b). Enzymatic synthesis of deoxyribonucleic acid. V. Chemical composition of enzymatically synthesized deoxyribonucleic acid. *Proc Natl Acad. Sci U S A*, 44, 1191-1196.

- [17] López-Madrugal, S., Latorre, A., Porcar, M., Moya, A., & Gil, R. (2011). Complete genome sequence of "Candidatus *Tremblaya princeps*" strain PCVAL, an intriguing translational machine below the living-cell status. *J Bacteriol.* , 193, 5587-8.
- [18] Luisi, P. L., Ferri, F., & Stanó, P. (2006). Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften*, 93, 1-13.
- [19] Mac, Donald. J. T., Barnes, C., Kitney, R. I., Freemont, P. S., & Stan, G. B. (2011). Computational design approaches and tools for synthetic biology. *Integr Biol (Camb)*. , 3, 97-108.
- [20] Mc Cutcheon, J., & Moran, N. (2011). Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* , 10, 13-26.
- [21] Mc Cutcheon, J. P., Mc Donald, B. R., & Moran, N. A. (2009). Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet*. 5; e1000565 EOF.
- [22] Miller, S. L. (1953). A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science.* , 117, 528-529.
- [23] Moreno-Hagelsieb, G., & Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatic* , 24, 319-234.
- [24] Murtas, G. (2009). Artificial assembly of a minimal cell. *Mol BioSyst* , 5, 1292-1297.
- [25] Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H. E., Moran, N. A., & Hattori, M. (2006). The 160 -Kilobase Genome of the Bacterial Endosymbiont *Carsonella*. *Science*. 314:267.
- [26] Pósfai, G., Plunkett, G., 3rd , , Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., de Arruda, M., Burland, V., Harcum, S. W., & Blattner, F. R. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* , 312, 1044-1046.
- [27] Quiñones-Valles, C., Espíndola-Serna, L., & Martínez-Antonio, A. (2011). Mechanisms and Controls of DNA Replication in Bacteria. *Fundamental Aspects of DNA Replication*. InTech , 219-244.
- [28] Reyes-Lamothe, R., Sherratt, D. J., & Leake, M. C. (2010). Stoichiometry and Architecture of Active DNA Replication Machinery in *Escherichia coli*. *Science*, 328, 498-501.
- [29] Shimizu, Y., Kuruma, Y., Ying, B. W., Umekage, S., & Ueda, T. (2006). Cell-free translation systems for protein engineering. *FEBS Journal* , 273, 4133-4140.
- [30] Shimizu, Y., & Ueda, T. (2010). PURE technology. *Methods Mol Biol.* , 607, 11-21.
- [31] Siccardi, A. G., Galizzi, A., Mazza, G., Clivio, A., & Albertini, A. M. (1975). Synchronous germination and outgrowth of fractionated *Bacillus subtilis* spores: tool for the analysis of differentiation and division of bacterial cells. *J Bacteriol.* , 121, 13-19.

- [32] Smolke, C. D. (2009). Building outside of the box: iGEM and the BioBricks Foundation. *Nat Biotechnol* , 27, 1099-1102.
- [33] Tabata, S., Oka, A., Sugimoto, K., Takanami, M., Yasuda, S., & Hirota, Y. (1983). The 245 base-pair oriC sequence of the *E. coli* chromosome directs bidirectional replication at an adjacent region. *Nucleic Acids Res.* , 11, 2617-26.
- [34] Tamames, J., Gil, R., Latorre, A., Peretó, J., Silva, F. J., & Moya, A. (2007). The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evolutionary Biology* 7:181.
- [35] Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., & Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
- [36] Ueda, T., Tohda, H., Chikazumi, N., Eckstein, F., & Watanabe, K. (1991). Cell-free translation system using phosphorothioate-containing mRNA. *Nucleic Acids Symp Ser.* , 25, 151-2.
- [37] Williams, K. P., Gillespie, J. J., Sobra, B. W., Nordberg, E. K., Snyder, E. E., Shallom, J. M., & Dickerman, A. W. (2010). Phylogeny of gammaproteobacteria. *J Bacteriol* , 192(9), 2305-1234.