J. Appl. Genet. 42(3), 2001, pp. 385-392

Review article

The minimal genome paradox

Grzegorz WĘGRZYN

Department of Molecular Biology, University of Gdańsk, Gdańsk, Poland

Abstract. The concept of a 'minimal genome' has appeared as an attempt to answer the question what the minimum number of genes or minimum amount of DNA to support life is. Since bacteria are cells bearing the smallest genomes, it has been generally accepted that the minimal genome must belong to a bacterial species. Currently the most popular chromosome in studies on a minimal genome belongs to *Mycoplasma genitalium*, a parasite bacterium whose total genetic material is as small as ~580 kb. However, the problem is how we define life, and thus also a minimal genome. *M. genitalium* is a parasite and requires substances provided by its host. Therefore, if a genome of a parasite can be considered as a minimal genome, why not to consider genomes of bacteriophages? Going further, bacterial plasmids could be considered as minimal genomes. The smallest known DNA region playing the function of the origin of replication, which is sufficient for plasmid survival in natural habitats, is as short as 32 base pairs. However, such a small DNA molecule could not form a circular form and be replicated by cellular enzymes. These facts may lead to an ostensibly paradoxical conclusion that the size of a minimal genome is restricted by the physical size of a DNA molecule able to replicate rather, than by the amount of genetic information.

Key words: bacteriophages, *Escherichia coli*, minimal genome, *Mycoplasma genitalium*, plasmids, replicons.

The minimal genome concept

During development of molecular genetics it became clear that not all genes of any organism are necessary for its survival under all environmental conditions.

Received: May 9, 2001. Accepted:
Correspondence: G. WĘGRZYN, Department of Molecular Biology, University of Gdańsk, Kładki 24, 80-822 Gdańsk, Poland, e-mail: wegrzyn@biotech.univ.gda.pl
Presented as a plenary lecture at the XIV Congress of the Polish Society of Genetics, Poznań, Poland, June 11-13, 2001.

Indeed, researchers could inactivate particular genes without any obvious phenotypes of mutants. Functions of many genes are required only under specific conditions, but some genes were found to be essential, i.e. their inactivation was lethal. Therefore, an idea to find a minimum set of genes necessary to support life appeared. This idea is called the minimal genome concept. A minimal genome can be defined as a minimum number of genes or a minimum amount of DNA to support life.

The original concept of a minimal genome was defined as a minimal set of genes that are both necessary and sufficient for life outside any host cell (for a recent review see RILEY, SERRES 2000). Since bacteria are the simplest and the smallest cellular organisms, they were unquestionable candidates for cells bearing a minimal genome. Large-scale sequencing of whole bacterial genomes provided data allowing intensive search for the minimal genome. *Mycoplasma genitalium* became a popular model in studies on the minimal genome concept. This bacterium has a very small genome consisting of about 580 kb (FRASER et al. 1995, HIMMELREICH et al. 1996). For comparison, the genome of *Escherichia coli* consists of about 4,639 kb (BLATTNER et al. 1997). Therefore, a comparison of sizes of genomes of *M. genitalium* and other unrelated bacteria led to a conclusion that a relatively large proportion of *M. genitalium* genes must be essential. If so, identification of these genes could be the first step in finding the minimal genome.

Comparison of *E. coli* and *M. genitalium* genomes

Comparison of *E. coli* and *M. genitalium* genomes (Table 1) can suggests that most of *E. coli* genes are not essential. Indeed, previous studies revealed that many *E. coli* genes could be disrupted and although mutant cells often became sick, they

Table 1. Comparison of genomes of *Escherichia coli* and *Mycoplasma genitalium*

Genome feature	*Escherichia coli*	*Mycoplasma genitalium*
Genome size (bp)	4 639 221	580 070
Total number of genes/ORFs	4406	470
Number of genes/ORFs of unknown functions	1408	170

were able to survive under certain laboratory conditions. Since there are still many *E. coli* and *M. genitalium* genes whose functions are unknown (Table 1), a search for the minimum set of essential genes is complicated. Experiments in which *M. genitalium* was subjected to transposon mutagenesis revealed that the genome

of this bacterium contains over 200 dispensable genes (PETERSON, FRASER 2001). Since there are 470 open reading frames (ORF) in the *M. genitalium* genome, the number of genes in a minimal genome could be roughly estimated to about 250.

Re-definition of a minimal genome

As mentioned above, the original concept of a minimal genome concerned organisms that are able to survive outside any host cell (RILEY, SERRES 2000). However, the question appears whether only such organisms should be considered. If one thinks about a minimal set of genes necessary for life, is it proper to exclude parasites? Paradoxically, *M. genitalium* living in its natural habitat is an obligatory parasite. One might argue that since this bacterium is able to grow in a culture in a cell-free environment under special laboratory conditions, it meets the definition of a free-living organism. However, the generation time of wild-type *M. genitalium* under optimal laboratory conditions is as long as 12 hours (PETERSON, FRASER 2001). In comparison to the generation time of *E. coli* (20-30 min), this is an extremely long period. Moreover, *M. genitalium* needs a very rich medium, including amino acids and other compounds, to grow in a laboratory culture. Hence, is it really a free-living bacterium?

On the basis of the above arguments, it would be possible to propose that if we want to examine genomes of really free-living organisms, we should exclude parasites, and thus also *M. genitalium*. However, this way of thinking is perhaps not proper, as by eliminating parasites researches would have to restrict themselves to consider only cells able to grow in a culture without special requirements. But what do special requirements mean? Even *E. coli* needs a carbon source to replicate in a minimal medium. One might then consider only photosynthetic and chemosynthetic bacteria, but they also need some special substrates.

Assuming that the concept presented in the preceding paragraph is not acceptable, it is necessary to come back to the definition of free-living organisms, or more generally, living organisms. In my opinion, in a search for a really minimal genome, parasites should not be excluded. Thus, the problem whether *M. genitalium* is a really free-living organism appears not important in discussions on a minimal genome. It is clear that this bacterium is able to survive and replicate in an environment containing certain substances. Hence, the definition of a minimal genome can be modified, as no specification whether an organism has to be free-living is necessary in this case. For a definition of a true minimal genome it would be necessary to present a unambiguous definition of life, which still remains a debated subject. However, if it is possible to accept that (i) the most important feature of all living creatures, which is absent in non-living substances, is an ability (at least potential, and concerning at least a part of the body, e.g. some

cells) to give progeny, i.e. to replicate, (ii) the information about functions of an organism is called a genome, and (iii) genomes of all living organisms consist of nucleic acids, then a minimal genome can be defined as a minimal nucleotide sequence that is both necessary and sufficient for life, i.e. for replication.

The smallest genomes

Since the newly proposed definition of a minimal genome, presented in the preceding chapter, is not restricted to non-parasites, genomes significantly smaller than that of *M. genitalium* can be considered. Viruses are classical examples of parasites, and those infecting bacterial cells are called bacteriophages. For replication, they require a special environment, namely a bacterial cell. Nevertheless, having all necessary substances bacteriophages can replicate efficiently, which is analogous to *M. genitalium*. The only considerable difference is that bacteriophages require more complicated substances than mycoplasms. For example bacteriophages need bacterial enzymes, while mycoplasms require only amino acids. However, this is a quantitative difference rather than qualitative, as the requirements still involve more or less complicated substances.

The problem is that while it is possible to prepare artificial conditions allowing growth of mycoplasms, no success in full bacteriophage development outside a host cell has been reported to date. However, genomes of some bacteriophages can exist and replicate as plasmids. For example, bacteriophage P1 has two alternative developmental pathways: one pathway that leads to production of progeny virions, and an alternative pathway in which the P1 genome replicates in the host cell as a plasmid (CHATTORAJ 2000). Similarly, a fragment of bacteriophage λ genome can replicate as a plasmid in *E. coli* cells (TAYLOR, WĘGRZYN 1995). Moreover, it is possible to replicate such plasmids *in vitro* by providing a set of several proteins and nucleotides (DODSON et al. 1986, ZYLICZ et al. 1989, CHATTORAJ 2000).

Bacteriophage-derived plasmids are only a small group of plasmids. These generally small DNA molecules replicate autonomously is cells, thus one might argue that they can be considered separate genomes. Indeed, plasmids have their own life, in which the most important feature is to replicate and survive inside host cells. To achieve this, many plasmids developed special partition mechanisms ensuring that each living daughter cell of the host harbors at least one plasmid molecule (GERDES et al. 2000). Therefore, it seems clear that plasmids can be treated as separate genomes.

In the light of the minimal genome concept, the most important question is what a minimal amount of DNA for survival is. For many plasmids, 'to survive' means simply 'to replicate'. Thus, one can ask a question what a minimal DNA region ensuring specific replication of a DNA molecule is. In other words, the ques-

tion is what a minimal replicon is. The replication region of bacteriophage λ, which after its excision can replicate in *E. coli* cells as a plasmid, consists of a few thousand base pairs (Figure 1). It may be artificially shortened to several hundred base pairs (HERMAN-ANTOSIEWICZ et al. 1998). Similarly, for replication

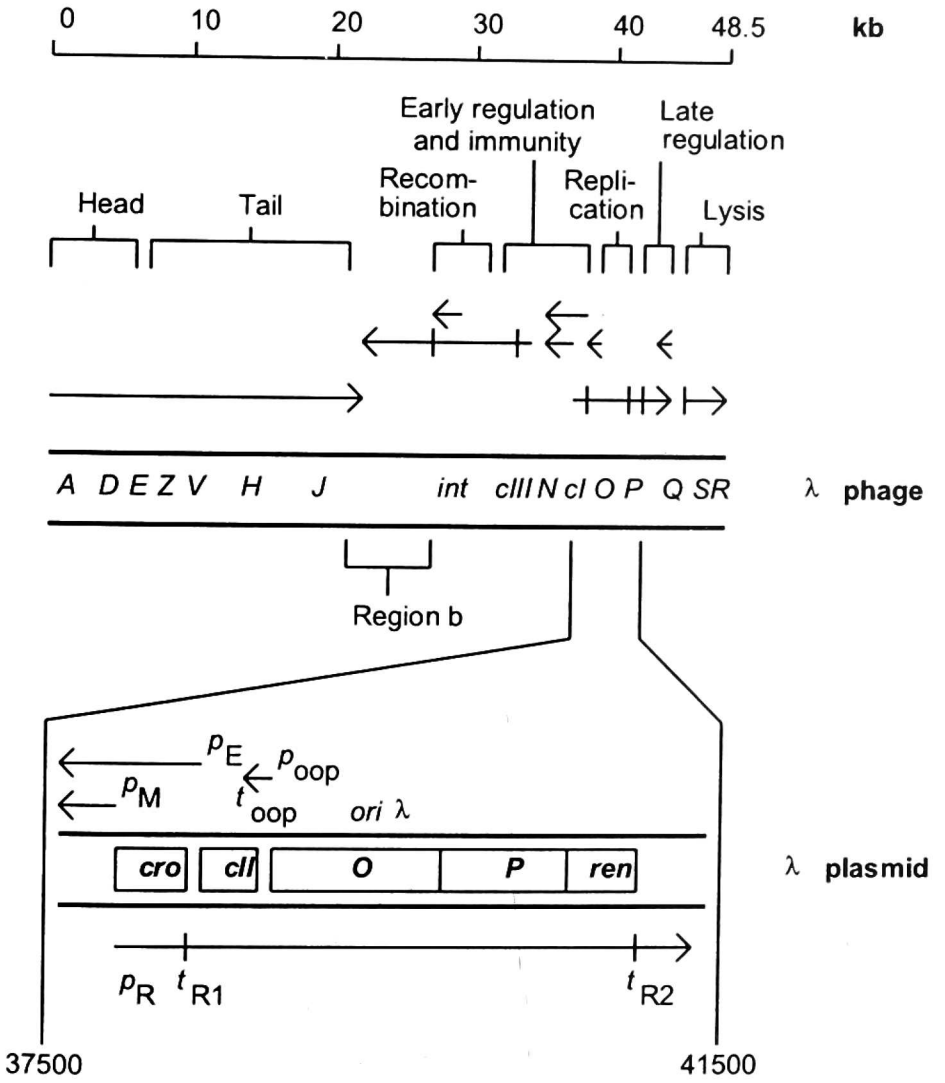

Figure 1. Map of λ phage and λ plasmid DNAs. The scale at the top of the figure is given in kilobases (kb). Regions of the genome which contain genes coding for particular functions are indicated. Region *b* encompasses a nonessential part of the λ genome. Positions of certain important genes are marked. Main transcripts are marked by arrows; arrowheads indicate the direction of transcription. Main terminators are marked by short vertical bars crossing appropriate lines of transcripts. A fragment of the λ genome present in a typical λ plasmid is presented in the lower part of the figure. The *ori*λ sequence (present in the middle of the *O* gene) is a region for initiation of λ DNA replication.

of many plasmids, a DNA region composed of several hundred base pairs is necessary (DEL SOLAR et al. 1998, CHATTORAJ 2000).

Initiation of replication of many plasmids depends on the presence of specific repeated sequences near the physical origin of replication, called iterons, to which a specific initiator protein binds (such an initiator-binding site is called a replicator) (DEL SOLAR et al. 1998, CHATTORAJ 2000). Usually, a replicator

consists of several iterons (Table 2). However, in the region of replication initiation of ColE2 and ColE3 plasmids, there are only two iteron-like sequences (YASUEDA et al. 1989). In fact, these regions are the smallest replicators described so far, consisting of 32 (ColE2) and 33 (ColE3) base pairs (Table 2). For replication initiation these plasmids require a plasmid-encoded protein (Rep), but this protein can be provided *in trans*.

Table 2. Number of iterons in selected plasmids, and the sequence of the replication origin of plasmid ColE2 with iteron-like sequences underlined

Plasmid	Number of iterons or iteron-like regions
R6K	7
P1	5
RK2	5
pSP10	4
λ	4
pSC101	3
ColE2	2
	origin sequence:
	5'-TGAGACCAGATAAGCCTTATCAGATAACAGCG

The physical barrier of a minimal genome

Looking for a minimal genome, I have come to the point when it is possible to define the smallest known DNA region that can assure plasmid replication, and thus its survival. Therefore, could we say that the region of the 32 base pairs of plasmid ColE2, consisting of the two iterons, is a real minimal genome? In the modified definition of a minimal genome (see above) it is important that such a structure should be able to survive. By the way, this is why PCR-amplified DNA fragments cannot be considered as genomes, because they could not survive and replicate in any natural environment although it is possible to replicate them *in vitro*. The minimal sequence necessary for ColE2 replication was deduced on the basis of studies on significantly larger DNA molecules, by deletions of particular DNA regions. Therefore, one can say that the region of 32 base pairs is sufficient for replication of a DNA molecule, but it is not equivalent to a statement that such an extremely short DNA fragment is a minimal genome, because a minimal genome must be able to survive and replicate in a natural environment. In other words, it is obvious that a DNA fragment consisting of 32 base pairs (i.e. having only three helical turns) cannot form a circle that could be then recognized by

a replication protein and replicated by DNA polymerase (Figure 2). Note that ColE2 is a circular plasmid, and opening of a double-stranded structure in circular DNA, usually near the replicator sequence, is necessary for initiation of replication of iteron-containing plasmids (KORNBERG, BAKER 1992, HELINSKI et al. 1996, DEL SOLAR et al. 1998). Clearly, an extra DNA fragment is necessary to ensure mechanically the possibility of enzymatic replication of a small circular DNA

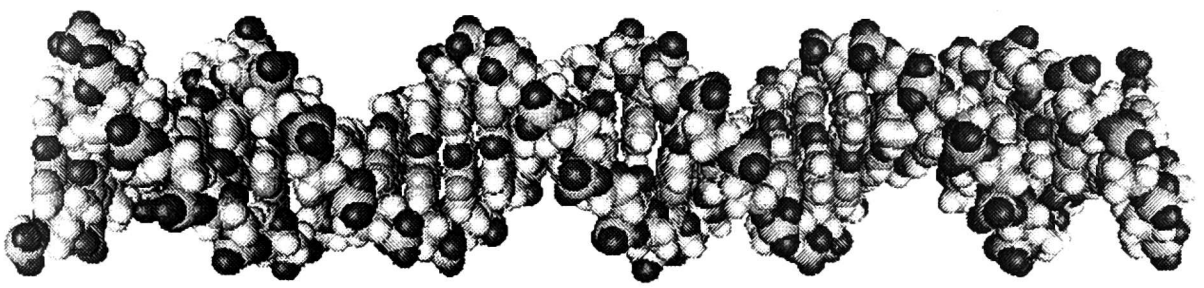

Figure 2. Model of a DNA fragment composed of 32 base pairs

molecule. This additional DNA fragment might be devoid of any genetic function, but it must be the ballast allowing the formation of a proper structure of the whole molecule. Currently it is hard to predict how long such an additional DNA fragment must be.

Concluding remarks

The original concept of a minimal genome concerned only free-living organisms. However, if one wants to consider all known forms of life, a minimal genome may mean a minimal replicon. The smallest known replicator sequence (32 bp region) belongs to plasmid ColE2. However, such a short DNA molecule would not be able to replicate in a natural environment. It needs an additional DNA fragment, even devoid of genes and/or regulatory sequences, to form a structure allowing formation of the replication complex and movement of the replication forks. A minimal genome has been defined as a minimal set of genes that are both necessary and sufficient for life. Paradoxically, it appears that the lowest size of a minimal genome is not limited by the amount of genetic information, but rather by physical properties of DNA molecules.

REFERENCES

BLATTNER F.R., PLUNKETT G., BLOCH C.A., PERNA N.T., BURLAND V. et al. (1997). The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453-1474.

CHATTORAJ D.K. (2000). Control of plasmid DNA replication by iterons: no longer paradoxical. Mol. Microbiol. 37: 467-476.

DEL SOLAR G., GIRALDO R., RUIZ-ECHEVARRIA M.J., ESPINOSA M., DIAZ-OREJAS R. (1998). Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. 62: 434-464.

DODSON M., ECHOLS H., WICKNER S., ALFANO C., MENSA-WILMOT K., GOMES B., LEBOWITZ J., ROBERTS J.D., MCMACKEN R. (1986). Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: localized unwinding of duplex DNA by a six protein reaction. Proc. Natl. Acad. Sci. USA 83: 7638-7642.

FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A., FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M. et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. Science 270: 397-403.

GERDES K., MOLLER-JENSEN J., JENSEN R.B. (2000). Plasmid and chromosome partitioning: surprises from phylogeny. Mol. Microbiol. 37: 455-466.

HELINSKI D.R., TOUKDARIAN A., NOVICK R.P. (1996). Replication control and other stable maintenance mechanisms of plasmids. In: *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (Neidhard F.C., Curtiss III R., Ingraham J.L. Lin E.C.C., Low K.B., Magasanik B., Reznikoff W.S., Riley M., Schaechter M., Umabarger H.E.,eds.). American Society for Microbiology, Washington DC: 2295-2324.

HERMAN-ANTOSIEWICZ A., ŚRUTKOWSKA S., TAYLOR K., WĘGRZYN G. (1998). Replication and maintenance of λ plasmids devoid of the Cro repressor autoregulatory loop in *Escherichia coli*. Plasmid 40: 113-125.

HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.C., HERRMANN R. (1996). Complete sequence analysis of the bacterium *Mycoplasma pneumoniae*. Nucleic Acids Res. 24: 4420-4449.

KORNBERG A., BAKER T. (1992). DNA replication. Freeman, New York.

PETERSON S.N., FRASER C.M. (2001). The complexity of simplicity. Genome Biol. 2: 2002.1-2002.8.

RILEY M., SERRES M.H. (2000). Interim report on genomics of *Escherichia coli*. Annu. Rev. Microbiol. 54: 341-411.

TAYLOR K., WEGRZYN G. (1995). Replication of coliphage lambda DNA. FEMS Microbiol. Rev. 17: 109-119.

YASUEDA H., HORII T., ITOH, T. (1989) Structural and functional organization of ColE2 and ColE3 replicons. Mol. Gen. Genet. 215: 209-216.

ŻYLICZ M., ANG D., LIBEREK K., GEORGOPOULOS C. (1989). Initiation of λ DNA replication with purified host- and bacteriophage-encoded proteins: the role of the DnaK, DnaJ and GrpE heat shock proteins. EMBO J. 8: 1601-1608.