

Conserved Gene Clusters in Bacterial Genomes Provide Further Support for the Primacy of RNA

Janet L. Siefert, Kirt A. Martin, Fadi Abdi, William R. Widger, George E. Fox

Department of Biochemical and Biophysical Sciences, 3201 Cullen Blvd., University of Houston, Houston, TX 77204-5934, USA

Received: 10 March 1996 / Accepted: 20 May 1997

Abstract. Five complete bacterial genome sequences have been released to the scientific community. These include four (eu)Bacteria, *Haemophilus influenzae*, *Mycoplasma genitalium*, *M. pneumoniae*, and *Synechocystis* PCC 6803, as well as one Archaeon, *Methanococcus jannaschii*. Features of organization shared by these genomes are likely to have arisen very early in the history of the bacteria and thus can be expected to provide further insight into the nature of early ancestors. Results of a genome comparison of these five organisms confirm earlier observations that gene order is remarkably unpreserved. There are, nevertheless, at least 16 clusters of two or more genes whose order remains the same among the four (eu)Bacteria and these are presumed to reflect conserved elements of coordinated gene expression that require gene proximity. Eight of these gene orders are essentially conserved in the Archaea as well. Many of these clusters are known to be regulated by RNA-level mechanisms in *Escherichia coli*, which supports the earlier suggestion that this type of regulation of gene expression may have arisen very early. We conclude that although the last common ancestor may have had a DNA genome, it likely was preceded by progenotes with an RNA genome.

Key words: Genomics — Gene regulation — RNA world — Progenote

Introduction

Two major repositories of evolutionary history reside in extant organisms. These are the three-dimensional struc-

tures of macromolecules and their assemblies and the linear sequence of the DNA. With the progress now being made by whole-genome sequencing projects, the latter source of information is being accessed at an unprecedented rate. Already, the complete genome sequences of four Bacteria, *Haemophilus influenzae* (Fleischmann et al. 1995), *Mycoplasma genitalium* (Fraser et al. 1995), *M. pneumoniae* (Himmelreich et al. 1996), and *Synechocystis* PCC 6803 (Kaneko et al. 1996), and one Archaeon, *Methanococcus jannaschii* (Bult et al. 1996), have been released to the scientific community. Within the next year it is possible that at least 20 more bacterial genomes will be completed, annotated, and accessible. Comparison of these genomes will have far-reaching impact upon our understanding of the evolutionary history of bacteria. For example, it will be possible to identify how many distinct gene families exist, and the history of each of these genes families will gradually be deciphered. Moreover, the focus of many studies will gradually change from the evolutionary history of individual genes to the history of genomes. At present, the analysis of genome sequence data is in its infancy, and much of the published analysis focuses on preliminary issues such as gene identification and annotation, etc. This study represents an initial attempt to understand what can be learned about early genomes from local gene order conservation.

Methods and Results

Initially, the complete sequence of the *M. genitalium* genome was compared to that of *H. influenzae* by listing each gene in order of occurrence in the *M. genitalium* genome with the location (i.e., gene number) of the equivalent gene in *H. influenzae*. When two or more

Table 1. Conserved gene clusters^a

Group	Short name, gene names	Type of regulation	Other organisms	Domain conservation
I	1. 16S, 23S, 5S rRNAs	RNA, DNA	<i>B. subtilis</i> , other Gram +, <i>Borrelia burgdorferi</i>	Bacteria/Archaea ^g
	2. S10 ^b "operon" (<i>rps10</i> , <i>rpl3</i> , <i>rpl4</i> , <i>rpl23</i> , <i>rpl2</i> , <i>rps19</i> , <i>rpl22</i> , <i>rps3</i> , <i>rpl16</i> , <i>rpl29</i> , <i>rps17</i>)	RNA	<i>E. coli</i> , <i>B. subtilis</i> , <i>Thermotoga maritima</i>	Bacteria/Archaea ^g
	3. Str "operon" (<i>rps12</i> , <i>rps7</i> , <i>fusA</i>)	RNA	<i>E. coli</i> , <i>B. subtilis</i>	Bacteria/Archaea
	4. Spc ^c "operon" (<i>rpl14</i> , <i>rpl24</i> , <i>rpl5</i> , <i>rps14</i> , <i>rps8</i> , <i>rpl6</i> , <i>rpl18</i> , <i>rps5</i> , <i>rpl30</i> , <i>rpl15</i> , <i>secY</i>)	RNA	<i>E. coli</i> , <i>B. subtilis</i>	Bacteria/Archaea
	5. L13 "operon" (<i>rps9</i> , <i>rpl13</i>)	Not known	<i>E. coli</i> , <i>B. subtilis</i>	Bacteria/Archaea
	6. L11 "operon" (<i>rpl11</i> , <i>rpl1</i>)	RNA	<i>E. coli</i> , <i>B. subtilis</i> , <i>T. maritima</i>	Bacteria
	7. Alpha ^d "operon" (<i>rps13</i> , <i>rps11</i> , <i>rpoA</i> , <i>rpl17</i>)	RNA	<i>E. coli</i> , <i>B. subtilis</i>	Bacteria
	8. L35 ^e "operon" (<i>infC</i> , <i>rpl35</i> , <i>rpl20</i>)	RNA	<i>E. coli</i> , <i>B. subtilis</i>	Bacteria
	9. L34 "operon" (<i>rpl34</i> , <i>mpA</i>)	Not known	<i>E. coli</i> , <i>B. subtilis</i> , <i>B. burgdorferi</i> , other Gram+	Bacteria
	10. L21/L27 (<i>rpl21</i> , <i>rpl27</i>)	Not known	<i>E. coli</i>	Bacteria
	11. L10 "operon" (<i>rpl10</i> , <i>rpl12</i>)	RNA	<i>E. coli</i> , <i>B. subtilis</i> , <i>B. burgdorferi</i> , other Gram+	Bacteria/Archaea ^g
II	12. ATPases ^f (<i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpA</i> , <i>atpG</i> , <i>atpD</i> , <i>atpC</i>)	RNA	<i>B. subtilis</i>	Bacteria/Archaea
III	13. Beta "operon" (RNA polymerase) (<i>rpoC</i> , <i>rpoB</i>)	RNA	<i>B. subtilis</i> , other Gram+	Bacteria/Archaea
	14. Initiation factor (<i>nusA</i> , <i>infB</i>)	RNA	<i>B. subtilis</i> , <i>E. coli</i> , <i>Thermus aquaticus</i>	Bacteria
IV	15. Spermidine/putrescine Transport (<i>potA</i> , <i>potB</i> , <i>potC</i>)	Not known	<i>E. coli</i>	Bacteria
V	16. Chaperones (<i>groEL</i> , <i>groES</i>)	RNA, DNA	<i>B. subtilis</i> , other Gram+	Bacteria

^a Conserved gene clusters identified in this study are numbered 1 through 16 and categorized in five groups according to gene function as studied in *E. coli*. Clusters 1–11 in group I are primarily RNA and protein constituents of the ribosome. Group II contains cluster 12, whose genes are involved in energy metabolism, e.g., the component of the ATP proton motive force interconversion enzyme, ATP synthase. The genes code for the hydrophilic F¹ unit which catalyzes the synthesis of ATP and membrane-bound hydrophobic F₀ unit, which forms the proton channel. Genes in group III are involved in RNA synthesis, modification, transcription, and translation. Cluster 13 codes for the β and β' subunits of the DNA-dependent RNA polymerase. Cluster 14 codes for NusA, which modulates the rate of chain synthesis and IF-2, which binds tRNA to the ribosome complex during initiation of or protein synthesis. In addition it should be noted that cluster 7 includes the gene for the α subunit of DNA-dependent RNA polymerase. Cluster 15, in group IV, is a member of the superfamily of periplasmic binding-protein-dependent (BDP) and ATP-binding cassette (ABC) transporters, e.g., traffic ATPases which transport polyamines into the cell. Group V contains cluster 16, which codes for the molecular chaperones GroEL and GroES. The second column indicates whether the gene expression in the case of *E. coli* is regulated at the RNA or DNA level. The third column indicates other organisms in which the gene order is known to occur and the last column indicates the extent to which phylogenetic conservation of the gene order exists

^b *rps10* is not in this cluster in *Synechocystis* PCC 6803

^c In *Synechocystis* PCC 6803 *rps14* has been relocated elsewhere and a homolog of *rpl30* has not been identified

^d *rps4* is frequently found between *rps11* and *rpoA*

^e *infC* is not located immediately upstream in *Synechocystis* PCC 6803

^f The order in *Synechocystis* PCC 6803 is *atpH*, *atpG*, *atpF*, *atpD*, *atpA*, *atpC* with *atpB* and *atpE* located elsewhere

^g The match with Archaea has exceptions or unusual features

equivalent genes were adjacent in both genomes it was recorded as a cluster. This process was repeated for the *H. influenzae* and *Synechocystis* pair. Because the three organisms are phylogenetically distantly related (*M. genitalium* and *M. pneumoniae* are considered as essentially a single taxonomic representative in this study), it was considered likely that clusters of genes which occurred in both comparisons would be conserved throughout the (eu)Bacteria. Additional gene order data was sought for each of these clusters and was found in many cases in partially sequenced genomes such as *E. coli* and *Bacillus subtilis*. Occasional anomalies exist in the larger ribosomal protein gene clusters in that in some cases, individual proteins are in either aberrant positions

or not found. In addition, several of the ribosomal protein gene clusters frequently associate into large contiguous clusters in the various organisms. Herein, we consider the largest units of these ribosomal protein gene clusters whose order is conserved among the genomes examined to be separate clusters. Thus, 16 highly conserved gene clusters were identified. In Table 1, the 16 clusters are grouped according to gene functions assigned to them in *E. coli* (Neidhardt 1996). These results extend an earlier study (Mushegian and Koonin 1996a) that reached similar conclusion based on a smaller data set. The two largest clusters each contain 11 ribosomal protein genes; the rest are smaller, ranging from two to eight genes. Although the gene order within each

Table 2. Gene clusters found either in the *Haemophilus/Mycoplasma* comparison or the *Haemophilus/Synechocystis* comparison, but not both^a

Cluster	Organism and gene number		Other organisms	
	<i>M. genitalium/M. pneumoniae</i> (580,000/800,000 bp)	<i>H. influenzae</i> (1,830,000 bp)	<i>Synechocystis</i> PCC 6803 (3,570,000 bp)	
<i>rps6, rps18, rpl9</i>	090,092–093/ G07_215,104b,149	545–544	Not conserved	<i>E. coli</i>
<i>rps16, trmD, rpl19</i>	444–446/K05_119,210,88	201–202,204	Not conserved	<i>E. coli</i>
<i>moaA, moaC, moaE</i>	Genes not found	1676, 1675, 1673	slr0901–0903	<i>E. coli</i>
<i>oppB, dciAC, amiE, amiF</i>	077–080/ G07_376,389a,423,851	1123–1120	No data on <i>oppB</i> , <i>dciAC</i>	
<i>secD, secF</i>	Genes not found	240, 239	slr0774,slr0775	<i>Mycobacterium tuberculosis</i> , <i>Streptococcus coelicolor</i> , <i>E. coli</i>
<i>secE, nusG</i>	<i>secE</i> not found	716, 717	ssl3335, ssl1739	<i>Desulfurolobus ambivalens</i> , <i>B. subtilis</i> , Other Gram+, <i>Rickettsia prowazekii</i> , <i>E. coli</i>
<i>clpX, clpP</i>	Genes not found	715, 714	slI0535, slI0534	
<i>dnaJ, dnaK</i>	Order not conserved	1237, 1238	slI1933, slI1932	<i>B. stearothermophilus</i> , <i>B. subtilis</i> , <i>Mycoplasma capricoleum</i>
<i>pdhD, pdhC, pdhB</i>	271–273/F11_457,402,327	1231–1233	slr1841, all1721	<i>B. subtilis</i> , other Gram+, <i>Acholeplasma laidlawii</i>
<i>dnaA, dnaN</i>	469–001/order not conserved	993–992	Order not conserved	<i>E. coli</i> , <i>B. subtilis</i> , <i>Borrelia burgdorferi</i>
<i>fruA, fruK</i>	062–063/R02_694,300	446–447	<i>fruA</i> not found	<i>Xanthomonas campestris</i>
<i>pheS, pheT</i>	194–195/C09_341,808	1311–1312	Order not conserved	<i>E. coli</i> , other Gram–, <i>B. subtilis</i>
<i>parC, parE</i>	203–204/C09_789,635	1529–1528	Genes not found	<i>Streptococcus pneumoniae</i>
<i>prfA, hemB</i>	258–259/H91_359v,453	1561, 1559	Order not conserved	<i>Salmonella typhimurium</i>

^a Gene numbers are as designated by the investigators of each genome. Other organisms known to have the indicated gene order are indicated. Two clusters, *secE, nusG* and *secD, secF*, are widely conserved across bacterial lineages with the exception of the mycoplasmas and may well be better regarded as members of Table 1, but have been placed here for consistency of definition. The pyruvate dehydrogenase cluster is of special interest. The order *pdhB, pdhC, pdhD* is widely conserved. In the case of *Synechocystis* PCC 6803 *pdhB* and *pdhC* are still neighbors in the genome but are now encoded on opposite strands and hence are not considered as forming a conserved cluster in the sense used here

cluster is highly conserved, no obvious evidence was found for conservation of cluster order between *H. influenzae* and *M. genitalium*. This strongly suggests there are no conserved regulatory mechanisms that rely on spatial location requirements to simultaneously regulate several of the clusters. Finally, we searched for each cluster in the genome of the Archaeon *Methanococcus jannaschii* and found equivalents for eight of the clusters (Table 1).

Discussion

The set of conserved gene clusters must be regarded as minimal at this stage. As gene identification is refined, additional conserved clusters may be located. In addition, a number of gene clusters were found that are in either the *Haemophilus/Mycoplasma* comparison or the *Haemophilus/Synechocystis* comparison (Table 2), but not both. Some of these less conserved clusters may in fact be widely distributed as well. This is especially true of clusters not present in the *M. genitalium/M. pneu-*

moniae comparison as those organisms are intracellular parasites with a reduced set of genes. They may have lost gene clusters late in evolution that have nevertheless been widely conserved. Although the gene clusters identified here likely contain many of the most ancient genes, it is clear that we have not sought in this study to identify all the primitive genes—only those whose gene expression is likely to be coordinated in a way that requires spatial proximity in the genome.

The conserved gene clusters are arranged in five groups (Table 1). These groupings encompass genes that in many respects outline the essentials of a primitive cell. They encompass components of translation, transcription, energy production, and membrane transport. In addition to the likelihood that their own expression is coordinated, many of these proteins are themselves important coordinators of cellular functions. The first group contains 11 clusters which consist almost entirely of components of the translation machinery. In fact, the ribosomal protein gene orders seen in *E. coli* (Keener and

Nomura 1996) are largely preserved throughout the (eu)Bacteria, and it therefore is likely that a portion of the regulatory structure relating to translation was in place at an early stage in the history of life. The second grouping contains the ATP synthase gene cluster which is essential for energy metabolism and the third grouping includes key components of the transcription machinery. Efficient coupling of transcription/translation is ensured in *E. coli* by two genes, *nusA* (modulates the rate of chain synthesis and prolongs pausing at natural pause sites) and *nusG* (enhances the rate of transcription elongation by interaction with RNA polymerase) (Richardson and Greenblatt 1996). *SecE* acts as a membrane-bound transporter and contains a minimal, functional domain homolog conserved in both bacteria and eukaryotes (Murphy and Beckwith 1996). Recognition and export of proteins with aberrant or absent signal by *secY* and *secE* (Kadner 1996) signifies possible utilization of primitive versions of these proteins. No homologs for the *secE* and *nusG* genes have been identified in the mycoplasmas but it is clear that this cluster is otherwise extremely universal and its eventual elevation to the most conserved group is likely. The spermidine/putrescine transport cluster of group IV is a member of a group of membrane-bound transporters that span the periplasm in Gram-negative bacteria. These genes belong to a superfamily of ATP-binding cassette transporters that import polyamines (Boos and Lucht 1996). The molecular chaperones of group V monitor and respond to the state of protein folding for non-native polypeptide substrates and nascent chain polypeptides (Mayhew and Hartl 1996). Described originally as heat-shock proteins, their response provides a necessary advantage for primitive cells evolving in a thermophilic environment.

The regulation of many of the 16 most conserved gene cluster has been studied in detail in at least one case (typically *E. coli*). A remarkable correlation is found in that in essentially every case where regulation is well understood it occurs at the translational and/or transcriptional level with the RNA transcript being involved rather than DNA. Typically this RNA-level regulation can be considered attenuation as broadly defined (Landick et al. 1996). Thus, although the details differ, in the case of the S10, *Str*, *Spc*, L11, and α operons, it is well known that in each case a ribosomal protein product plays a key role as a repressor which binds to the RNA transcript. Recently, the L35 operon has been added to this list with the demonstration that L20 acts as a translation repressor and that a long-range pseudoknot in the RNA is required (Chiaruttini et al. 1996). The *trmD* operon (Table 2) is unique among the ribosomal protein operons in that neither translational nor feedback regulation occurs (Bjork 1996). Rather, a large stem/loop structure in the RNA influences the frequency of translational initiation. The individual products of the ATPase synthase operon are expressed at significantly different

levels. This regulation is apparently accomplished by differences in translation frequency and to a lesser extent differential mRNA degradation that is dependent on the structure of the RNA (Lagoni et al. 1993). In *E. coli* the L10 operon lies immediately downstream of the L11 operon and the genes for the β and β' subunits of RNA polymerase adjoin it, forming a single transcription unit. The regulation of this operon is complex, involving both transcriptional and translational mechanisms (Keener and Nomura 1996; Dykxhoorn et al. 1996). In *E. coli*, the expression levels of *nusA* and *infB* are modulated by several RNA secondary structures upstream from *nusA* and the NusA protein itself may be involved as well (Plumbridge et al. 1984). The structure of the *groE* operon is known to be highly conserved phylogenetically and typically contains a characteristic inverted repeat in the regulatory region that serves at the DNA level as an operator site and when it is transcribed determines mRNA stability (Segal and Ron 1996). Although, regulation of the rRNA operons in *E. coli* relies in part on DNA/protein interaction (the *fis* protein), it is noteworthy that there is also an important antitermination system involving *nusA* and *nusG* that facilitates transcription of the RNA products.

Clearly, it is premature to suppose that the regulatory schemes seen in *E. coli* will themselves be conserved throughout the (eu) Bacterial Domain. This will certainly not be true for the amalgam of the L11, L10, and the β clusters seen in *E. coli* as these are independent in other organisms. The argument presented here would be strengthened considerably if RNA-level regulation mechanisms were known to exist in other distantly related species such as *B. subtilis*, but information addressing this issues is typically not yet available. Nevertheless, the ubiquitous occurrence of RNA-level regulation in the most conserved clusters is very striking. A majority of the clusters involve ribosomal proteins and this could be interpreted as representing a bias; however, the need for early institution of the ribosomal machinery and the employment of attenuation as a mode of regulation perhaps is more correctly viewed as corroborating evidence. It lends strong support to the earlier suggestion that regulation by transcriptional attenuation might have been among the earliest regulatory mechanisms to have evolved (Landick et al. 1996) and perhaps was the only option when the major portions of the translation apparatus were first emerging. Moreover, in our view, the importance goes beyond this, because we believe the extensive use of RNA-level regulation in the most conserved gene clusters is strong support for the primacy of an RNA genome (Joyce 1989). The early availability of the NusA protein would have provided valuable refinements for attenuation regulation in such a genome.

Earlier comparisons of the gene content of *M. genitalium* and *H. influenzae* have led to the speculation that a minimal ancestral (eu)Bacterial genome would have

had approximately 256 genes (Mushegian and Koonin 1996b). Regulatory proteins with a helix-turn-helix domain are conspicuous by their absence from this minimal genome (Mushegian and Koonin 1996b). The results presented here demonstrate that by the time the level of complexity of such a minimal genome has been reached, significant coordination of gene expression would also have existed. This hypothetical ancestor, although much simpler, would have in its essence not been very different from modern organisms. It is widely believed that an even earlier predecessor would likely have lacked DNA. The current analysis reveals that the earliest efforts to coordinate gene expression were almost certainly mechanisms which operate at the RNA level. The obvious inference is that this reflects the fact that DNA was not available when these mechanisms evolved and hence at some stage an RNA genome existed. This must not be interpreted to mean that the last common ancestor of extant life had an RNA genome. Even among the genes whose coordinated expression is very old, some hint of DNA already exists in that subunits of the DNA-dependent RNA polymerases (at least in their modern form) are present in clusters 7 and 13.

Nevertheless, an RNA-only predecessor of the minimal DNA-based ancestor could clearly be even simpler as all the genes relating to the use of DNA could be dispensed with (Mushegian and Koonin 1996b). Such an RNA-only entity would still have a substantial number of genes, and it is quite possible that RNAs of the required size would be problematic. It is noteworthy that many of the ribosomal proteins that are included among those whose expression has been coordinated since the earliest times may be regarded as having RNA chaperone activity (Herschlag 1995). It also should be appreciated that the size problem would be overcome if a distributed genome comprised of multiple smaller RNAs (in essence a RNA-world equivalent of chromosomes) existed. Such RNAs may or may not have all been in the same cellular compartment. Whereas the last common ancestor based on DNA was likely modern in its essence, earlier RNA-based ancestors would "still be in the process of evolving the relationship between genotype and phenotype" and hence would be better regarded as progenotes (Woese and Fox 1977).

Acknowledgments. This work was supported by NASA grant NAG 5-4004 and a grant from the Institute of Space Systems Operations at the University of Houston.

Note Added at Proof

As this report was going to press another fully annotated (eu)Bacterial genome sequence, *Helicobacter pylori* (Tomb et al. 1997), has been reported. This diverse bacterium allows further clarification of the extent of conservation of gene clusters. As expected, the vast majority

of the clusters in Table 1 remain conserved. The major exceptions are: *nusA* and *infB* which are both present in *H. pylori*, but no longer clustered; and the genes for the Spermidine/putrescine transport system which were not detected. Although the 16S rRNA and 23S rRNA genes are separated in *H. pylori*, this would appear to be a minor anomaly as the 16S-23S-5S arrangement is well known to be widespread throughout the (eu)Bacteria. Other minor deviations are the absence of *rpl30* in cluster 4 and the fusion of the genes for *rpoC* and *rpoB*. In contrast, the vast majority of the clusters of Table 2 were not found in *H. pylori*. Thus, only the *trmD* operon, the *secD/secE* cluster and the *pheS/pheT* cluster continue to be strong candidates for inclusion in the list of widely conserved clusters. At this stage, there appear to be 19 conserved gene clusters, 14 of which are essentially universal. Components of the translational machinery even more strongly dominate the better defined set of conserved clusters and the new results thus strengthen our view that the expression of translation components was already a highly coordinated process at the time of the last common ancestor of the (eu)Bacteria.

References

- Bjork GR (1996) Stable RNA modification. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 861–886
- Boos W, Lucht JM (1996) Periplasmic binding protein-dependent ABC transporters. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 1175–1209
- Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb J-F, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghegan NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk H-P, Fraser CM, Smith HO, Woese CR, Venter JC (1996) Complete genome sequence of the methanogenic Archaeon, *Methanococcus jannaschii*. Science 273:1058–1073
- Chiaruttini C, Milet M, Springer M (1996) A long-range RNA-RNA interaction forms a pseudoknot required for translational control of the IF3-L35-L20 ribosomal protein operon in *Escherichia coli*. EMBO J 15:4402–4413
- Dykxhoorn DM, St Pierre R, Linn T (1996) Synthesis of the β and β' subunits of *Escherichia coli* RNA polymerase is autogenously regulated *in vivo* by both transcriptional and translational mechanisms. Mol Microbiol 19:483–493
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghegan NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512
- Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman JL, Weidman JF, Small KV, Sandusky M, Fuhrmann J,

- Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb J-F, Dougherty BA, Bott KF, Hu P-C, Lucier TS, Peterson SN, Smith HO, Hutchison III CA, Venter JC (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397–403
- Herschlag DJ (1995) RNA chaperons and the RNA folding problem. *Biol Chem* 270:20871–20874
- Himmelreich R, Hilber H, Plagens H, Pirki E, Li B-C, Herrmann R (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* 24:4420–4449
- Joyce GF (1989) RNA evolution and the origins of life. *Nature* 338: 217–224
- Kadner RJ (1996) Cytoplasmic membrane. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 58–87
- Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. II Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* 3:109–136
- Keener J, Nomura M (1996) Regulation of ribosome synthesis. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 1417–1431
- Lagoni OR, Meyenburg K, Michelsen O (1993) Limited differential mRNA inactivation in the *atp* (*unc*) operon of *Escherichia coli*. *J Bacteriol* 175:5791–5797
- Landick R, Turnbough CL, Yanofsky C (1996) Transcription attenuation. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 1263–1286
- Mayhew M, Hartl F-U (1996) Molecular chaperone proteins. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 922–937
- Murphy CK, Beckwith J (1996) Export of proteins to the cell envelope in *Escherichia coli*. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 967–978
- Mushegian AR, Koonin EV (1996a) Gene order is not conserved in bacterial evolution. *Trends Genet* 12:289–290
- Mushegian AR, Koonin EV (1996b) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* 93:10268–10273
- Neidhardt FC (ed) (1996) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC
- Plumbridge JA, Dondon J, Nakamura Y, Grunberg-Manago M (1984) Effect of NusA protein on expression of the *nusA, infB* operon in *E. coli*. *Nucleic Acids Res* 13:3371–3381
- Richardson JP, Greenblatt J (1996) Control of RNA chain elongation and termination, in *Escherichia coli* and *Salmonella typhimurium*. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium*. ASM Press, Washington, DC, pp 822–848
- Segal R, Ron EZ (1996) Regulation and organization of the *groE* and *dnaK* operons in Eubacteria. *FEMS Microbiol Lett* 138:1–10
- Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539–547
- Woese CR, Fox GE (1977) The concept of cellular evolution. *J Mol Evol* 10:1–6