INSTITUTE OF PHYSICS PUBLISHING PHYSICAL BIOLOGY

Phys. Biol. **3** (2006) 183–189 doi:10.1088/1478-3975/3/3/003

Optimal gene partition into operons correlates with gene functional order

Alon Zaslaver, Avi Mayo, Michal Ronen and Uri Alon

Department of Molecular Cell Biology and Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel

E-mail: urialon@weizmann.ac.il

Received 26 April 2006
Accepted for publication 23 August 2006
Published 18 September 2006
Online at stacks.iop.org/PhysBio/3/183

Abstract

Gene arrangement into operons varies between bacterial species. Genes in a given system can be on one operon in some organisms and on several operons in other organisms. Existing theories explain why genes that work together should be on the same operon, since this allows for advantageous lateral gene transfer and accurate stoichiometry. But what causes the frequent separation into multiple operons of co-regulated genes that act together in a pathway? Here we suggest that separation is due to benefits made possible by differential regulation of each operon. We present a simple mathematical model for the optimal distribution of genes into operons based on a balance of the cost of operons and the benefit of regulation that provides 'just-when-needed' temporal order. The analysis predicts that genes are arranged such that genes on the same operon do not skip functional steps in the pathway. This prediction is supported by genomic data from 137 bacterial genomes. Our work suggests that gene arrangement is not only the result of random historical drift, genome re-arrangement and gene transfer, but has elements that are solutions of an evolutionary optimization problem. Thus gene functional order may be inferred by analyzing the operon structure across different genomes.

M This article features online multimedia enhancements

Introduction

Genes in prokaryotes are arranged into groups called operons [1–3]. The genes in an operon are located adjacently on the DNA and are expressed on the same mRNA. Genes in the same operon usually work together, and their products sometimes physically interact [4–11]. However, partitioning of a given set of genes into operons is not the same in every organism. It has long been known that genes are often partitioned into operons in different ways in different organisms (figure 1) [12–16]. In some species all genes in a certain pathway or system are on the same operon (figure 1(*a*)), while in other species they are found in several operons (figures 1(*b*) and (*c*)), or even each on a separate operon (figure 1(*d*)).

Gene arrangement in bacterial genomes is governed by mutational processes such as gene duplication that can form new operons, genome rearrangements that can split existing operons, and lateral gene transfer that can introduce parts of operons from other organisms. The life-cycle of operons can be studied based on comparing operons in phylogenetically related organisms [13, 17, 18]. Naturally, no absolute rules can precisely describe operon structure in every organism. However, theory might help us to understand the forces that tend to shape operons in general.

To understand operon partitioning, we need to understand the selective forces that separate genes into different operons and the forces that tend to put genes together on the same operon. Several theories have been proposed to explain the latter. The 'selfish operon' theory notes that a gene system in which all of the genes are on a single operon allows for more reliable lateral transfer than a system of several operons distantly placed on the genome [19–21]. The proximity of the genes increases the probability that the entire system will be transferred when a piece of the DNA is incorporated into the genome of another organism. In addition, genes on the same operon are precisely co-regulated and, since they share

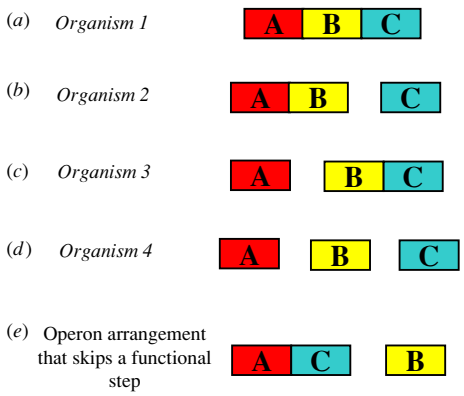

Figure 1. Different operon arrangements of genes in a given system: (*a*) three genes on the same operon; (*b*), (*c*) genes arranged in two operons; (*d*) each gene in a separate operon; (*e*) operon arrangement that skips a step, assuming the gene functional order is A→B→C. This arrangement is predicted in the present study to be much rarer than (*a*)–(*d*).

the same mRNA, may avoid fluctuations in mRNA levels that can lead to imperfect stoichiometry between the protein products [18, 22–25].

These theories, however, do not explain what causes co-regulated genes to often appear on *separate* operons.

In this paper, we propose that one possible separating force is the benefit of temporal programs of expression that are possible in multi-operon systems [3]. We then consider the implications of an idealized situation, where operon evolution is governed by the selective forces that drive genes into the same operon and those driving to separate them. This idealized model predicts that genes are arranged such that genes on the same operon do not skip functional steps in the pathway. Hence, operon partitioning can be used to infer gene functional order. We find that these predictions are supported by genomic data from a large number of sequenced prokaryotes.

Materials and methods

Operons in sequenced genomes

One hundred and thirty-seven bacterial genomes were obtained from the National Center for Biotechnology Information (NCBI) Entrez Genome website. Best orthologs of all *E. coli* genes were found by BLASTP against all other genomes with E-value cutoff [26] of 10^{-5}. Two genes were considered to be on the same operon if they were less than 40 bp apart and transcribed in the same direction [27]. The present conclusions do not sensitively depend on the 40 bp threshold or E-value cutoff.

Score for putative gene orders

Consider an ordering R of genes $i = 1, \ldots, n$, where R is a permutation of $1, \ldots, n$. For a given operon X of k genes, which contains genes $x_1, \ldots, x_k$, let x'_1, x'_2, x'_k be the same arranged in order of increasing R, and

$$\mathrm{Score}(R, X) = \sum_{j=2}^{k} [R(x'_j) - R(x'_{j-1}) - 1].$$

This score counts the number of skipped steps of the permutation R in the operon. If the operon does not skip steps, or has only a single gene, Score(X) = 0. The score of the putative order R is the sum of the scores of all operons in all genomes. For example, consider five genes A, B, C, D, E partitioned into two operons as follows <u>AB</u>, <u>CDE.</u> There are 5! = 120 different permutation orders. Consider one permutation, R = (A→E→D→B→C). To find the score of operon <u>CDE</u> note that the gene orders in this operon are 5,3,2 according to R, and therefore Score = (3−2−1) + (5−3−1) = 1. The score of operon <u>AB</u> is Score = 4 − 1 − 1 = 2. This permutation thus receives a score of 3 for this genome. To find the total score for R, the score for all operon arrangements observed in the set of genomes is summed.

Each occurrence of an operon arrangement is considered only once, even if the same operon arrangement appears in multiple genomes. This prevents over-counting effects due to phylogenetic relationships between sequenced organisms.

Estimates of parameters in the evolutionary model

The present conclusions do not depend on the precise values of the parameters in the model. However, for completeness, we attempt to estimate the parameters in the theory in the case where fitness is related to the growth rate of the bacteria [28, 29], and where operon cost is related to a putative growth cost of DNA synthesis (in a fraction of the conditions encountered by the cell). To estimate the cost per protein unit synthesized, note that bacteria produce about 10^6 proteins/cell cycle. Thus, a protein unit represents a metabolic load of about $\eta \sim 10^{-6}$. This cost agrees with experimental measurements of growth reduction due to protein expression [3, 29–31]. To estimate the putative cost of synthesizing a promoter region (the cost of an excess operon), note that promoter regions are about 100 bp long, and genomes are on the order of 10^6 bp. Thus a promoter amounts to about $\varepsilon = 10^{-4}$ of the total DNA cost. Since DNA synthesis is thought not to be limiting under rapid growth conditions, we include also the probability that the cell is in an environment where DNA synthesis is limiting for growth, P_D. The overall DNA cost of a promoter is $a = P_D\,\varepsilon$. For example, for a system with $N = 10^3$ protein copies/cell, with delay $\Delta t/\tau = 0.1$ cell cycles, probability that the system is expressed in the environment (demand [3, 28, 32]) $p = 10^{-2}$, and $P_D = 10^{-2}$, one finds $x{\cdot}S = p\eta N\ \Delta t/\tau/a \sim 1$. For this example, the operon cost is approximately equal to the protein cost.

Results and discussion

Selective forces for separation of genes into different operons

Before analyzing the model, we shall discuss the proposed selective force that can drive separation of genes into different operons. This force is the evolutionary benefit of temporal expression patterns that are possible in a multi-operon partitioning. This proposal is based on the experimental observation that many gene systems show a temporal order of expression: operons are expressed one after the other

with timing differences on the order of 0.1 cell-generations [33] when the system is induced or deactivated [3, 33–39]. Importantly, the temporal order correlates with the *functional* order of the gene products. In metabolic pathways, the closer the gene product to the beginning of the pathway, the earlier its promoter is activated [33]. Similar 'just-when-needed' production schedules (also called 'just-in-time' production programs) have also been found in the assembly of large protein machines, such as flagella [36, 37], and in stress-response systems [38].

In these systems, 'just-when-needed' production can be beneficial because it prevents superfluous synthesis of a protein before it is required [33–39]. 'Just-when-needed' production can be achieved by means of differential regulation thresholds for each of the operons. In the simplest case, a gradually changing regulator activity level crosses the different thresholds at different moments, generating a temporal expression pattern [3, 33, 37, 38, 40]. This type of temporal order of transcription cannot be achieved if all of the genes are on the same operon, because such genes all share the same mRNA [17].

Simple model for optimal operon partitioning

We now describe a simple model based on the opposing forces that cluster and separate genes. We begin by quantifying the benefit of 'just-when-needed' temporal expression order that we have just discussed, that is, the benefit of avoiding un-needed protein production. For example, consider the case where gene B is needed only at a delay Δt after gene A. If A and B are on the same operon, they are forced to be expressed together, resulting in un-needed expression of B for a time Δt. The cost of this excess production is $\eta N\ \Delta t/\tau$, where η is the reduction in fitness per unit of protein B, and N is the number of copies of B produced during a cell generation time τ (thus, N/τ is the mean rate of B production). This cost per protein was recently measured directly for the lac system of *E. coli* [29]. The cost occurs only during times when the system is expressed. Thus, the cost occurs with probability per generation, p, corresponding to the probability that the system is induced in the organism's environment [28, 32, 44]. Thus, the cost of putting A and B on the same operon, averaged over many generations, is

$$\text{Cost of A and B on the same operon} = p\eta N\ \Delta t/\tau.$$

The cost of un-needed protein production is avoided when the genes are on different operons by 'just-when-needed' production. Thus the effect we have just examined tends to separate genes into different operons.

The separating force is opposed by a second force that tends to cluster genes into the same operon. This clustering cost can effectively represent two effects: (a) the cost of imbalance (noise) in expression of interacting gene products coded on two different operons, which can lead to deleterious stoichiometric effects [24]; (b) the cost of multiple operons based on the cost of the DNA regulatory region needed for each operon. DNA length may be limiting for cell fitness or growth under some conditions [41]. The cost per operon representing these two effects will be described in this study by the parameter a. If there are m operons, the overall cost is $m{\cdot}a$.

Hence, the overall cost of an arrangement of both genes A and B on a single operon is the sum of the cost of one operon and the cost of un-needed protein production:

$$C_1 = a + p\eta N\ \Delta t/\tau. \tag{1}$$

The cost for arranging the two genes on *separate* operons is only the cost of two operons, because un-needed protein production is avoided by 'just-when-needed' production:

$$C_2 = 2a. \tag{2}$$

The terms with a represent the cost of the operons, and the term with η is the cost of excess protein production, which is avoided in C_2. This fitness function depends on the ecology of the organism through the parameter p which describes the fraction of time in which the system is induced. It also depends on the delays between genes in the system and their production costs. The larger the ratio of protein costs to operon costs, $x = p\eta/a$, the higher the tendency for the genes to be on different operons. (For parameter estimations, see the materials and methods section.)

Generalizing this theory to k genes partitioned into m operons yields a normalized cost:

$$C/a = m + x \sum_{\substack{\text{genes } i \text{ in} \\ \text{same operation}}} S_i \tag{3}$$

where

$$x = p\eta/a$$

is the relative cost of protein production to operon cost, and the excess production of gene i is given by the number of copies produced per cell generation times the delay in production:

$$S_i = N_i\ \Delta t_i/\tau. \tag{4}$$

The excess production S_i is greater the higher the expression level of the gene product N_i and the longer its functional delay Δt_i relative to the gene on the operon whose product acts first.

Note that the present theory does not concern the order of genes *within* an operon (that is, operons ABC and ACB are considered equivalent).

The operon partitioning that minimizes the cost function (equation (3)) can be found analytically. First, consider the case where all proteins have the same delay values, $S_i = S$. This solution can be easily seen graphically by plotting the cost versus $x{\cdot}S$ (figure 2). At each value of $x{\cdot}S$, the operon arrangement with lowest cost is optimal. For the case of two genes, a single operon is optimal when $x{\cdot}S < 1$, and two separate operons are optimal when $x{\cdot}S > 1$ (figure 2(*a*)). For three genes, at low values of $x{\cdot}S$ a single operon is optimal; at intermediate values two operons are optimal, one with a single gene and one with two genes; and a three-gene operon is optimal at high values of $x{\cdot}S$ (figure 2(*b*)). Thus, a three-gene system can have operons of size one, two or three depending on $x{\cdot}S$ (similar solutions are found for any number of genes).

The larger the protein cost of gene i, S_i, relative to the protein costs of the other genes in the system, the higher the probability that it will appear first on an operon or even alone on an operon (figure 2(*c*)). Thus the model predicts that all

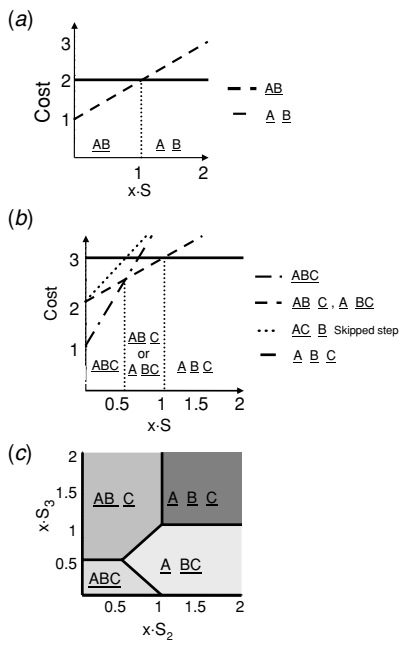

Figure 2. Evolutionary cost-benefit model for operon arrangement as a function of $x{\cdot}S$, the ratio of protein cost to excess operon cost. (*a*) Two-gene system; AB corresponds to a single operon (dashed), and A B to separate operons (full line). (*b*) Three-gene system, with equal protein cost S for all three proteins. The partition which skips a step, AC B, is never optimal (dotted line). (*c*) Three-gene system, with different protein costs S_2 and S_3 for proteins B and C relative to protein A.

things being equal, a highly expressed protein would tend to be on shorter operons more than a low-expressed protein.

Furthermore, genes that have no delay ($\Delta t_i = 0$ and thus $S_i = 0$) should tend to appear together on the same operon. Proteins that are parts of the same protein complex are examples where $\Delta t_i = 0$ is possible.

An important new prediction of the model is that the *optimal operon partitioning does not include an operon that skips functional steps*. Thus, if a pathway's functional order is A→B→C, operons such as ABC, AB, and BC are possible, but not AC (figures 1, 2(*b*)). The latter operon is not optimal because it skips step B in the pathway. This prediction applies to systems with any number of genes (see SOM, where the optimal partitioning is computed for any number of genes N).

Analysis of prokaryotic genomes reveals that genes in operons tend not to skip functional steps

We tested the prediction that genes in an operon do not skip functional steps by examining the actual operon partitioning in prokaryotic genomes. We considered all known un-branched metabolic pathways in *E. coli* that consist of three or more gene products (based on the EcoCyc database [42]). We determined the operon partitioning of these genes in 137 fully sequenced genomes. Operon partitioning was based on a stringent criterion: genes were assigned to the same operon if they are spaced less than 40 bp apart and in the same orientation. We find that each of the pathways had several distinct operon partitioning in the different genomes (table 1).

We scored the number of skipped steps in the operon partitioning. For each gene system, we considered all possible putative functional ordering of the genes (permutations of the gene order). Each permutation was scored, with a score equal to the number of skipped steps in all of the operons. Permutations that fit into the operons without skipped steps received a zero score. We found that, in all pathways, the known functional order of the genes is among the lowest scoring permutations, and in many cases it is the lowest scoring permutation (figure 3, table 1). For example, the five genes in the tryptophan biosynthesis pathway show several different operon partitioning across organisms [13]. The actual functional order trpE→trpD→trpC→trpB→trpA receives the lowest score among the 120 possible permutations

Table 1. Metabolic un-branched pathways and their predicted order according to the scoring algorithm. The genes are arranged in the table according to the known *E. coli* pathway order 1, 2, 3, . . . , n. Shown are all pathways in EcoCyc with no branches that consist of at least three enzymes. The deduced order is the order with the lowest score; curly brackets indicate genes that can be interchanged, resulting in permutations with the same lowest score. The number of different operon partitionings found in the genomes is shown.

Product	1	2	3	4	5	6	7	Deduced order	Number of diff. partitionings
Histidine	*hisG*	*hisI*	*hisA*	*hisH/F*	*hisB*	*hisC*	*hisD*	1 2 {3, 4} 5 7 6	10
Lysine	*dapA*	*dapB*	*dapD*	*dapC*	*dapE*	*dapF*		6 {1, 2} 3 5 4	8
Tryptophan	*trpE*	*trpD*	*trpC*	*trpB*	*trpA*			{1, 2} 3 4 5	5
Ubiquinone	*ubiC*	*ubiA*	*ubiX/D*	*ubiB*				1 2 3 4	7
Biotin	*bioF*	*bioA*	*bioD*	*bioB*				1 2 3 4	6
Arginine	*argA*	*argB*	*argC*	*argD*				1 2 3 4	6
Menaquinone	*menD*	*menC*	*menE*	*menB*				1 2 3 4	7
Fatty acids	*fabB/F*	*fabG*	*fabZ*	*fabA*				2 1 3 4	5
Neuroaminate	*nanA*	*yhcH*	*nanK*	*nanE*				1 2 3 4	5
Enterobactin	*entC*	*entB*	*entA*					1 2 3	3
Colanic acid	*cpsB*	*gmd*	*wcaG*					1 2 3	3
Rhamnose	*rhaA*	*rhaB*	*rhaD*					1 2 3	3
Ascorbate	*ulaD*	*ulaE*	*ulaF*					1 2 3	3
Purine	*purK*	*purE*	*purC*					1 2 3	3
Folic acid	*folB*	*folK*	*folP*					1 2 3	3
Heme	*hemL*	*hemB*	*hemC*					1 2 3	3

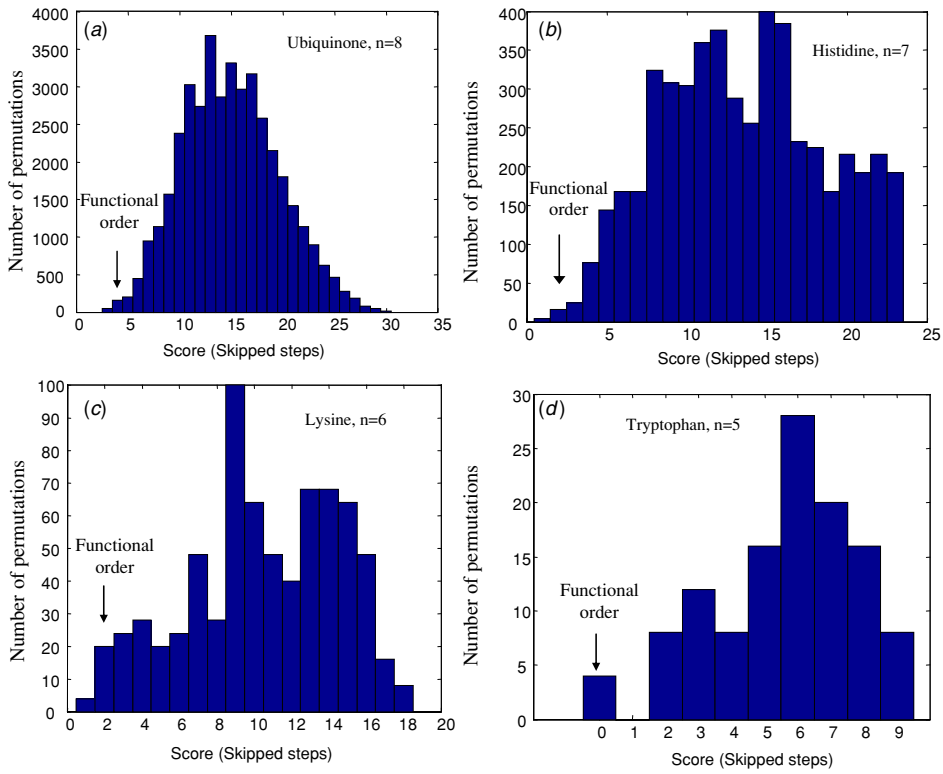

Figure 3. Operons tend not to skip steps with respect to gene functional order. All possible gene orderings of linear metabolic pathways were scored according to the number of skipped steps in the operons of the corresponding genes in 137 different genomes. Each putative gene ordering is one of the $n!$ permutations of the n genes in the pathway. Permutations with low scores correspond to gene orders that fit into the operon partitioning with few skipped steps. Arrows indicate the score of the actual gene order in the *E. coli* metabolic pathway. The biosynthesis pathways are (*a*) ubiquinone (*b*) histidine (*c*) lysine (*d*) tryptophan.

(figure 3(*d*)). It agrees with the prediction that operons tend not to skip functional steps in the pathway.

In most pathways studied, the known order has the least number of skipped steps. In other pathways, such as the seven-gene histidine biosynthesis pathway, only two of the permutations received a lower score than the actual order. Most of these low-scoring permutations are different from the known *E. coli* functional order in only one or two genes (table 1). It is possible that metabolic pathways have variations in different organisms that can explain these few alternative orders. More generally, additional forces and the constraints of evolutionary history may contribute to shape operon partitioning beyond the present considerations.

Conclusion and outlook

The main finding of the present study is that operons contain information about the functional order of genes: operons tend not to skip functional steps in pathways. The gene functional order can be inferred by analyzing the operon structure across different genomes. A simple evolutionary theory was presented that describes operon structure in terms of cost-benefit analysis. The theory indicates which structure is optimal given the demand for the system in the environment, and the cost and delays of the proteins in the system. The theory demonstrates how idealized operon partitioning can be understood based on the benefit of just-when-needed production strategies.

The present 'rules' for operon partitioning have notable exceptions. For example, the highly expressed ribosomal proteins are predicted to be on small operons. This is true for some, but not all ribosomal proteins. Interestingly, ribosomal proteins' operon partitioning does not change substantially between different organisms, suggesting other constraints which might determine the operon partitioning of these genes. A second exception occurs for some proteins which function together in a complex (and thus have $\Delta t = 0$). These proteins are predicted to be on the same operon, and usually are. However, there are cases where such proteins are on different operons. For example, the two subunits of phenylalanyl-tRNA synthetase are on the same operon in about ∼75% of prokaryotes, and on different operons in the remaining ∼25% [43]. Further study is needed to understand these exceptions.

We also note that the present theory can be extended to include alternative effects that might explain our finding that genes within operons tend not to skip functional steps. One such explanation is that adjacent enzymes in metabolic pathways may channel intermediates to each other whereas non-adjacent ones do not. Therefore it is particularly critical to place adjacent enzymes on the same operon to produce them in stoichiometric amounts and avoid the formation of unproductive complexes (for a more general context on

balance between interacting proteins, see Veitia 2004). Such effects could, in principle, be included in the cost parameter S_i in the theory above.

The present theory helps to explain why gene systems are often partitioned into different operons, and which partitioning is expected to occur based on the organism's environment and metabolic economy. As the number of sequenced genomes grows, this approach could be more thoroughly tested and may help to predict functional gene orders or relative expression levels in novel pathways.

Acknowledgments

We thank M Elowitz, M Surette, R Lenski and all members of our lab for discussions. We thank HFSP, NIH, Minerva and the Clore foundation for support. AM thanks the Pacific Foundation postdoctoral grant for support.

Glossary

Operon. A group of genes transcribed on the same mRNA. Each gene is separately translated.

Promoter. A regulatory region of DNA that controls the transcription rate of a gene. The promoter contains a binding site for RNA polymerase, the enzyme that transcribes the gene to produce mRNA. In addition, each promoter usually contains binding sites for transcription factor proteins; the transcription factors, when bound, affect the probability that RNAp will initiate transcription of an mRNA.

Cost-benefit analysis. A theory that seeks the optimal design such that the difference between the fitness advantage gained by a system (benefit) and fitness reduction due to the cost of its parts is maximal.

Enzyme. A protein that facilitates a biochemical reaction. The enzyme catalyzes the reaction and does not itself become part of the end product.

Genomes. The total genetic information of an organism. Sequenced genomes are genomes of those organisms in which the entire genetic information is available.

References

[1] Jacob F and Monod J 1961 Genetic regulatory mechanisms in the synthesis of proteins *J. Mol. Biol.* **3** 318–56
[2] Miller J 1981 *The Operon* (New York: Cold Spring Harbor Laboratory Press)
[3] Alon U 2006 *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Boca Raton, FL: CRC Press)
[4] Marcotte E M *et al* Detecting protein function and protein–protein interactions from genome sequences *Science* **285** 751–3
[5] Enright A J *et al* 1999 Protein interaction maps for complete genomes based on gene fusion events *Nature* **402** 86–90
[6] Pellegrini M *et al* 1999 Assigning protein functions by comparative genome analysis: protein phylogenetic profiles *Proc. Natl Acad. Sci. USA* **96** 4285–8
[7] Overbeek R *et al* 1999 The use of gene clusters to infer functional coupling *Proc. Natl Acad. Sci. USA* **96** 2896–901
[8] Huynen M *et al* 2000 Predicting protein function by genomic context: quantitative evaluation and qualitative inferences *Genome Res.* **10** 1204–10
[9] Galperin M Y and Koonin E V 2000 Who's your neighbor? New computational approaches for functional genomics *Nat Biotechnol.* **18** 609–13
[10] Yanai I, Derti A and DeLisi C 2001 Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes *Proc. Natl Acad. Sci. USA* **98** 7940–5
[11] Teichmann S A and Veitia R A 2004 Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective *Genetics* **167** 2121–5
[12] Omelchenko M V *et al* 2003 Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ *Genome Biol.* **4** R55
[13] Xie G *et al* 2003 Ancient origin of the tryptophan operon and the dynamics of evolutionary change *Microbiol. Mol. Biol. Rev.* **67** 303–42 table of contents)
[14] Makarova K S, Mironov A A and Gelfand M S 2001 Conservation of the binding site for the arginine repressor in all bacterial lineages *Genome Biol.* **2**
[15] Tamames J 2001 Evolution of gene order conservation in prokaryotes *Genome Biol.* **2**
[16] Lathe W C III, Snel B and Bork P 2000 Gene context conservation of a higher order than operons *Trends Biochem. Sci.* **25** 474–9
[17] Price M N, Arkin A and Alm E J 2006 The life-cycle of operons *PLoS Genet* **2** e96
[18] Price M N *et al* 2005 Operon formation is driven by co-regulation and not by horizontal gene transfer *Genome Res.* **15** 809–19
[19] Lawrence J G and Roth J R 1996 Selfish operons: horizontal transfer may drive the evolution of gene clusters *Genetics* **143** 1843–60
[20] Lawrence J G 1997 Selfish operons and speciation by gene transfer *Trends Microbiol.* **5** 355–9
[21] Lawrence J 1999 Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes *Curr. Opin. Genet. Dev.* **9** 642–8
[22] McAdams H H and Arkin A 1999 It's a noisy business! Genetic regulation at the nanomolar scale *Trends Genet.* **15** 65–9
[23] Elowitz M B *et al* 2002 Stochastic gene expression in a single cell *Science* **297** 1183–6
[24] Veitia R A 2004 Gene dosage balance in cellular pathways: implications for dominance and gene duplicability *Genetics* **168** 569–74
[25] Kaern M *et al* 2005 Stochasticity in gene expression: from theories to phenotypes *Nat. Rev. Genet.* **6** 451–64
[26] Date S V and Marcotte E M 2003 Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages *Nat. Biotechnol.* **21** 1055–62
[27] Moreno-Hagelsieb G and Collado-Vides J 2002 A powerful non-homology method for the prediction of operons in prokaryotes *Bioinformatics* **18** S329–36
[28] Savageau M A 2001 Design principles for elementary gene circuits: elements, methods, and examples *Chaos* **11** 142–159
[29] Dekel E and Alon U 2005 Optimality and evolutionary tuning of the expression level of a protein *Nature* **436** 588–92
[30] Koch A L 1983 The protein burden of lac operon products *J. Mol. Evol.* **19** 455–62
[31] Lenski R E *et al* 1994 Epistatic effects of promoter and repressor functions of the Tn10 tetracycline-resistance operon of the fitness of Escherichia coli *Mol. Ecol.* **3** 127–35

[32] Shinar G *et al* 2006 Rules for biological regulation based on error minimization *Proc. Natl Acad. Sci. USA* **103** 3999–4004
[33] Zaslaver A *et al* 2004 Just-in-time transcription program in metabolic pathways *Nat. Genet.* **36** 486–91
[34] Spellman P T *et al* 1998 Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization *Mol. Biol. Cell* **9** 3273–97
[35] Endy D *et al* 2000 Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes *Proc. Natl Acad. Sci. USA* **97** 5375–80
[36] Laub M T *et al* 2000 Global analysis of the genetic network controlling a bacterial cell cycle *Science* **290** 2144–8
[37] Kalir S *et al* 2001 Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria *Science* **292** 2080–3
[38] Ronen M *et al* 2002 Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics *Proc. Natl Acad. Sci. USA* **99** 10555–60
[39] McAdams H H and Shapiro L 2003 A bacterial cell-cycle regulatory network operating in time and space *Science* **301** 1874–7
[40] Kalir S and Alon U 2004 Using a quantitative blueprint to reprogram the dynamics of the flagella gene network *Cell* **117** 713–20
[41] Cavalier-Smith T 2005 Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion *Ann. Bot.* **95** 147–75
[42] Karp P D *et al* 2002 The EcoCyc Database *Nucleic Acids Res.* **30** 56–8
[43] Brown J R 2001 Genomic and phylogenetic perspectives on the evolution of prokaryotes *Syst. Biol.* **50** 497–512
[44] Dekel E, Mangan S and Alon U 2005 Environmental selection of the feed-forward loop circuit in gene-regulation networks *Phys. Biol.* **2** 81–8