Chapter 1

Genome structures, operating systems and the image of the machine

ANTOINE DANCHIN* & STANISLAS NORIA
HKU-Pasteur Research Centre, Dexter HC Man Building, 8, Sassoon Road, Pokfulam, Hong Kong
Present address: Genetics of Bacterial Genomes, Institut Pasteur, 23 rue du Docteur Roux, 75724 Paris Cedex 15, France
e-mail: adanchin@pasteur.fr

1. Introduction

Despite scattered reflections about the relationships tying genes and genomes, our view of genomes remains static. Worse, many scientists still speak about "junk" DNA for a large part of the genome, demonstrating that they consider it as a collection of genes rather than an integrated structure. When some speak about the genome "fluidity" or "plasticity" this is just to account for variation in the distribution of genes along the chromosome, not for the time-dependent expression of genes as a function of their distribution in the genome (Brunder and Karch,2000; Dobrint and Hacker, 2001). Our view is commonly to split the genome text into the multiple collections of its individual genes. Some even advocate "draft sequencing" as a sufficient approach for studying genomes (Bhattacharyya, 2002), without understanding that there might exist rules of genome organisation that would only be visible in completed sequences. We ought to think of life, however, more as the set of relationships between objects: genes, gene products and metabolites, than of a simple collection of objects (Danchin, 2003). On august 7th 2003, at least 747 genome sequencing programmes were underway or completed, including 140 complete and 355 ongoing prokaryotic genomes. This allows scientists, despite a significant bias in the species chosen for sequencing (often pathogens, because of our anthropocentric view of the world), to explore the sequences and compare

them with one another. At first sight, especially when the organisms are widely divergent, the structures of their genomes look very different. This is certainly not unexpected in view of the very long time of divergence between organisms. As a case in point it is admitted (but this rests on a very shaky set of interpretations of molecular clocks and rRNA divergence) that the two model Bacteria, *Escherichia coli* and *Bacillus subtilis* would have diverged from a common ancestor some 1-1.5 billion years ago. All genes should have been randomly reshuffled if there had not been some selective pressure keeping some of them together. And, as a matter of fact, at least 100 segments of operons are conserved between these organisms, sometimes as full operons (in the case of molecular machines such as the ribosome or ATP synthase, for example). Does this mean that there is some kind of rule that maintains genes together? Is there a constraint in the organization of the chromosome? At the time when the *B. subtilis* and *E. coli* genome sequences were known we argued that this might be related to the structure of the cell (Danchin and Henaut, 1997). What can we say today, as we have so many genome sequences available?

2. Cells as turing machines

2.1 Data, programme and machine

The discovery of the processes setting up regulation of gene expression, followed by that of the genetic code, spread the representation of life as the result of the expression of a "programme" (Yockey, 1992). This happened at a time when, in what was to become Computer Sciences or using the Gallic neologism "Informatics", the concept of programming (or writing algorithms) was already widely understood. The first computers had already been shown to operate as predicted by Turing, von Neumann and the many theoreticians and scientists who had uncovered the link between the arithmetics of whole numbers and logic. In a famous metaphore - that was to be implemented as a concrete object - Turing proposed that all computations involving integers as well as all operations of logic could be performed by a simple machine reading and modifying a tape carrying a linear sequence of symbols, the Universal Turing Machine. He could show that this required only the physical separation between the string of symbols and the machine. More precisely he showed that the tape was carrying the data that allowed the machine to proceed. In terms of their anthropocentric meaning the data could be split into two types, a programme, that embedded the meaning of the logical sequence recognized by the machine, and the data, that were

needed for the programme to operate (in a way they provided the context). However this distinction, which implies a purpose (in the implicit signification of programme), is not an intrinsic property of the machine: the algorithm that is read can be considered as purely declarative. This has important consequences in terms of the origin of the "genetic programme" since it would not formally require a pre-existing purpose. We shall not discuss this further here.

The metaphore of genetic programme was initially only perceived as a convenient way to describe how cells live and develop. It was however understood early on that some strange organisms (were they living, or not alive?), the viruses, behaved as individual pieces of programmes, using the cell as the machine needed to make them multiply and subsequently propagate (often by destroying the machine): no virus can survive without a host living cell. Interestingly, the phenomenon of lysogeny demonstrated that viruses could be embedded into the very genetic programme of the cell. Later on, when computer programming developed, pieces of programmes were found to behave formally as do biological viruses, and were named "viruses" accordingly. This was a further indication that there was perhaps a deeper meaning of the "programme" metaphore in what is life. Furthermore, when it became possible to manipulate DNA in vitro, the analogy appeared even deeper: working on a string of symbols, on a paper or in a computer (in silico) is enough to allow scientists to construct in vitro, and then in vivo experiments that correspond to the very concrete action of reprogramming cells. Many bacteria, today, produce human proteins. Finally, the identification of widely spread horizontal gene transfer (Medigue *et al.*, 1991), and subsequently nuclear cloning (Wilmut *et al.*, 1997), perfected the analogy to a point where it can be considered as highly revealing, if not (of course) explaining life in totality.

2.2 The first von Neumann's conjecture: operating systems and "housekeeping genes"

At this point it became interesting to explore more in depth the analogy: is it possible to really separate between "data + programme" and "machine" in living organisms? And, if so, can one find more detailed organization of the data + programme in the cell? Can we see how it evolved from the time when life appeared? The first question asks to demonstrate that there is a physical separation between the carrier of the programme and the machine. The DNA molecule is the obvious candidate, as argued frequently. But one must also find particular properties of the programme itself that should be present. It must be possible to represent it as a string of symbols: this is

indeed the case of the DNA when represented as a text written in a four-letter alphabet, and genome sequencing programmes are the ultimate demonstration that this is quite fit to reality. The programme must specify "instructions" that make the machine manipulate it. The basic concepts of what became Molecular Biology are remarkably appropriate for that. "Gene expression" is organized in an algorithmic way, and the coding process introduced at the level of translation allows for recursive processes to be implemented (Hofstadter, 1979). And recursivity, in itself, displays remarkable properties in terms of information, properties that are quite fit to account for the "creative" features of life (Bennett, 1988; Danchin, 2003; Yockey, 1992).

Now, von Neumann looked for a kind of hierarchy in the various processes that needed to be implemented for a Turing Machine to function as a concrete physical machine. And he introduced the concept of what is now named the Operating System (OS), a particular piece of the programme that is indispensible for the machine to operate (Silberschatz *et al.*, 2001). This is defined by a series of functions that can be summarized as follows. A "virtual machine" hides from the "users" all the engineering details of the computer as a physical entity (we shall go to this point again below). Associated to it there must be a "resource manager" providing efficient and effective sharing among "users" of the machine (each one using and creating data and running programmes). It should be noticed at this point that the concept of user is not directly associated to the human interaction with the machine (Man-Machine interaction). In fact users can also be machines (printers, screens, memory storage devices, all kinds of peripherals; the corresponding algorithms are usually named "drivers") and some are programmes. Finally several classes of programmes managed by the OS can be summarized as system programs (compilers, editors, loaders, etc.), application support programs (database management systems, networking systems, etc.) and finally, the programmes that correspond to the goals of the machine, application programs. In different OSs the data can be organized in a variety of ways, and in particular, in "object oriented" OSs resource management can be associated to data serving as output or input of application programmes.

As one can immediately see, there is no reason why there should exist only one type of OS. And, as a matter of fact, many exist and are in competition, since no machine would work without an OS. But an OS is a programme, hence can be separated from a machine, and a given machine can run several OSs. In short, an OS plays the role of a "housekeeping" programme. Do we find similar properties in living organisms? At first sight,

living cells display overall features similar to one another, and the (almost) universal rule of the genetic code, as well as of the DNA replication machineries would argue for universality. However, the process of cell division is remarkably different between the eukaryotes and the prokaryotes, for example. The compartmentalization is also very different in these organisms, with the former having a well-formed nucleus. In the class of prokaryotes, the components of the division machinery are different between Bacteria and Archaea, too. And even Bacteria are not homogeneous (see the interesting - and heated - debate about the origin of prokaryotes (Gupta, 1998; Mayr, 1998; Woese, 1998)). Despite similarities, there are large differences in the housekeeping genes controlling replication, transcription and translation in cells, and these are so large that the question remains: do they all have a common origin, or did they evolve independently? Even in "monoderm" Bacteria, there are at least two classes in terms of replication: one class uses only one housekeeping DNA polymerase, for the management of both DNA strands, while in another class two polymerases are needed (Rocha, 2002). All this points to the idea of a common functional class, that of the analog of an OS (Bozinovski *et al.*, 2001), which would differ in different types of cells, thereby making classes within which horizontal gene transfer is possible, while it would be impossible or quite difficult otherwise. This may have been selected through evolution as a means to screen out a number of viruses, for example. We think that this may account for the many discrepancies observed when scientists compare phylogenetic trees constructed from ribosomal RNA, or from diverse families of proteins (creating virulent controversies, that with the present view would simply look irrelevant).

In summary, a common formal feature is needed for all life to develop, and in particular, to allow for DNA replication and gene expression. This requires the presence of what has precisely been named "housekeeping" functions. If we explore the analogy of the Turing Machine revisited by von Neumann it is easy to place in parallel many of the functions needed by Operating Systems. And, exactly as is the case in the domain of real computers, there can be (and there is) more than one OS. Single cells would need an OS similar to that of personal computer OSs, with some time-sharing properties. In some instances the OS could degenerate to that of a simple batched OS or more often to multiprogrammed batched OSs. For more complex organisms, one would obviously need parallel and/or distributed systems. And in general, because the analog of the OS must manage many nanomachines, it would probably be more of the object-oriented type (i.e. managing resources inside data files). As a consequence, while it is important for each organism to identify its housekeeping genes,

there is no compelling reason that would state that these genes should have exact counterparts in all organisms. The only good reason for universality would be historical: if it is difficult to create this or that function, it is likely that once it has appeared somewhere it will spread everywhere. This implies divergent evolution (but horizontal transfer as well). In contrast, for functions that are more straightforward one could witness diversity and/or convergent evolution.

2.3 The second von Neumann's conjecture: where is the image of the machine?

All this may look fine. However, a Turing machine does not make Turing machines. John von Neumann argued convincingly that, if this were to happen there should be somewhere an image of the machine, and this image should be reproduced in parallel to replication of the programme (von Neumann, 1958). In addition to DNA replication, many structures are duplicated when cells multiply. In some cases, such as that of the centriole, this is particularly remarkable (and still not really understood (Rice and Agard, 2002)). All structures of the cells must at some point duplicate, and one might think that its overall organisation is tightly linked to a duplication process. What would be the link, however? Interestingly there seems to exist a link between the control of centrosome duplication and maintenance of genomic stability that is particularly visible in cancer (Kramer *et al.*, 2002; Wong and Stearns, 2003). Chromosomes of course duplicate, and in eukaryotic cells this means not only DNA structures, but the whole chromatin, with its network of histones and regulatory proteins and RNAs. In bacteria, the daughter cells are not equivalent, as clearly seen in *Caulobacter crescentus* (Judd *et al.*, 2003), and indirectly seen by whole genome analysis (Rocha *et al.*, 2003). It is therefore interesting to explore the conjecture that the very structure for which we know how duplication proceeds, DNA, might also somehow carry the overall map of the cell. In short, there would be a map of the cell in the chromosome. Or stated in another way, the gene order in the chromosome should not be random.

In most Bacteria and some Archaea, the existence of a higher number of genes in the leading strand, relative to the lagging strand has long been known after it was observed that rDNA and ribosomal proteins are systematically coded in the leading strand of the *E. coli* chromosome (Ellwood and Nomura, 1982), then in several genomes (McLean *et al.*, 1998). The preferential positioning of translation-related highly expressed genes (rDNA and ribosomal proteins) in the leading strand has been widely

admitted to result from selection for lower transcription abortion and higher replication rates while avoiding the head-on collisions that happen when genes are located in the lagging strand. Within this view selection would only be effective for highly expressed genes and acting on the positioning of genes in the leading strand, thus creating a gene strand bias proportional to the expression level in replicating bacteria. Because the bias would match the frequency of replication, fast-growing bacteria should display a more important strand bias. It was however demonstrated recently, that, at least in *B. subtilis*, where essential genes have been identified experimentally (Kobayashi *et al*., 2003), the level of expression was not the cause of the bias, but, rather, this bias was due to the essential character of the genes (Rocha and Danchin, 2003). This suggested that replication/transcription collisions are selected against because this creates abortive transcripts, not replication pausing. It also suggests that essentiality is a major determinant of the chromosome structure. That this is a general feature in many bacteria has also been observed (Rocha and Danchin, unpublished).

Other biases exist in the distribution of nucleotides, words or genes in bacterial genomes. We have mentioned distribution of genes in the leading, vs. lagging strand. Another bias between the strands has been known for some time, and used to identify the origins of replication in chromosomes where it was unknown (Lobry, 1996). And this bias is often so large that it could identify a universal principle (A+C enrichment in the lagging strand, vs G+T in the leading strand) that can be seen when genes are shifted from one strand to the complementary one (Rocha and Danchin, 2001). Remarkably, this bias is the same under conditions of DNA shuffling and recombination that are highly variable, including situations where many Insertions Sequences are present or when they are absent (Rocha *et al*., 1999). As a consequence this indicates, at least in Bacteria, that there exist selective constraints that organize the genome irrespective of the way the DNA nucleotide sequence is managed. What could be these constraints?

2.4 Drosophiloculus, homunculus?

Curiously, hints for a possible answer come from the study of eukaryotic organisms, in particular multicellular organisms. In the early eighties extraodinary mutations were discovered in the drosophila fly: mutants arose that had legs in the place of antennae. These mutants carried a mutation at the Antennapedia locus (Bachiller *et al.*, 1994; Balavoine and Telford, 1995; Duboule and Dolle, 1989; Gaunt, 1991; Gehring, 1987; Kaufman *et al*. 1990). Many such homeotic genes were later on discovered, including in plants (for reviews see (Akam, 1998; Brock and van Lohuizen, 2001;

Cobourne, 2000; Jagla *et al.*, 2001; Kammermeier and Reichert, 2001; Lohmann and Weigel, 2002; Reichert, 2002; Stock, 2001)). Interestingly, homeotic genes are clustered on the genome in the same sequence as they are expressed on the anteroposterior axis of the animal, and this is true both for vertebrates and invertebrates (Bachiller *et al.*, 1994; Balavoine and Telford, 1995; Duboule and Dolle, 1989; Gaunt, 1991). In short, they are colinear (as a string of symbols in a computer programme). A knock-out of a homeotic gene often results in the transformation of a segment towards a more anterior type of segment. In general it can be concluded that insects have one such set of homeotic genes, while mammals have four (Bachiller *et al.*, 1994). Whether this is true in the case of fish is a matter of discussion (Holland, 1999). One interesting feature is the comparison between insect and crustacea: the old observation by Geoffroy Saint Hilaire that the body plan of crustacea was reversed under the thorax (the abdomen becomes the back and vice versa) has been substantiated by the observation that modification of homeotic genes between insects and crustacea affected their body plan (Averof, 1997; Averof and Akam, 1995). An interesting summary of all these discussions can be found at (Merks, 1997). There is, as yet, no convincing explanation to account for the selective forces that maintained this order in these control genes.

A general conclusion of an enormous amount of work is that there is, in multicellular organisms, a relationship between the order of at least one family of genes controlling gene expression, and the plan of the organism itself. In short there is a "animalculus" in animals, similar to the "homunculus" that preformists thought to see at the origin of the development of Man (Danchin, 2003). This stucture however, is not a minute animal, but a physical organisation of the programme that makes the animal. We do not have, at this point, any idea of the selective constraints that account for this relationship, or any idea of the reason why it can indeed control the development of the embryo from the egg to the adult animal. We wish however to point out here that this fits very well with the conjecture of von Neumann about computers that would make computers.

2.5 Celluloculus?

Once admitted that the order of the genes making a multicellular organism is not random, it becomes tempting to ask the question in the case of the cell itself. Biochemical studies of cells, usually based on "cell-free" experimental approaches, have the tendency to consider the cell as a tiny test-tube in which biological objects move more or less randomly. Considering the extremely tight level of packaging inside cells suggests that

this is certainly not plausible, and not even an interesting simplification (Ellis, 2001). A cell, even a bacterial cell, must be a highly organized medium (Danchin, 2003; Danchin *et al*., 2000). This must be taken into account in any description of phenomena involving gene expression in the construction or duplication of a cell.

A rapid observation of the position of orthologous genes in various genomes may give the impression that they can be located anywhere, and certainly not always at the same position in different genomes. However, it is clear that genes have a strong tendency to form clusters. This clustering is certainly not random: genes coding for related functions, such as those coding for the subunits of an enzyme for example, are generally located in a same transcription unit. Indeed this was the very basis of the concept of the operon. Another type of clustering, also associated to a correlation between the architecture of the genome and that of the organism was found when lysogenic phages were described. There, genes are clustered into operons that correspond to specialized functions: replication and construction of the head and tail of the virus. We have seen that genes do not move very frequently, or, more exactly that most gene moves are incompatible with the long term survival of the species. What we see as a move represents only what remains, under the constraint of long term survival, preserving usually a privileged orientation (and maybe privileged distances) with respect to the origin of replication. As a consequence, interestingly, one observes in bacteria a strong constraint in the distribution of the genes with respect to the origin and terminus of replication: genome rearrangements tend to be symmetrical around the origin of replication and thus do not change the replicating strand coding for the gene, even though they disrupt synteny (Tillier and Collins, 2000).

Furthermore, although the laterally transferred genes (that contribute much for the apparent genome diversity) appear to be present almost anywhere, they are not randomly located. When comparing strains (for example comparing *E. coli* K12 and pathogenic *E. coli*), one remarks that the foreign genes differ, but that they are placed at the same position in the chromosome, often near tRNA genes (Blum-Oehler *et al.*, 2000; Hou, 1999). This may just be because tRNAs allow recombination hot spots, but this is also compatible with the notion of a strong architectural constraint in the gene distribution, if tRNAs play a role in this organisation. This makes interesting to analyze the distribution of codons in protein coding genes, since this would be directly related to the translation process. Careful studies of the codon usage bias in genomes indeed point to the existence of strong constraints in the gene order along the chromosome. The genes within an

operon have usually a similar codon usage bias. Interestingly, comparison between organisms where genes are clustered into an operon in one and dispersed in the other reveals that the codon usage bias is preserved (Rocha *et al.*, 2000). The usage of a particular codon is coupled to a particular transfer RNA. It is therefore likely that repeating the same codon in a gene tends to maintain the cognate tRNA local concentration at a high level (in the vicinity of the ribosome that translates the messenger RNA). A large bias in the codon usage might therefore imply a local enrichment in some tRNAs, while in the same local environment the corresponding ribosome is depleted in other tRNAs. Another cause could also account for this observation: a gene transcribing a tRNA creates a source for that tRNA that would maintain a certain codon preference in its vicinity. In either case, the fact that genes distant in the chromosome share the same bias strongly suggests that the corresponding mRNAs are sitting next to each other and are translated frrom the same or very closely packed ribosomes. This indicates that there is a relation between the position of some ribosomes in the cell, and the position of some genes in the chromosome.

The most remarkable observation that we could make in favour of this hypothesis was that the codon usage bias in orthologous genes in organisms as different as *E. coli* and *B. subtilis* shows a similar deviation of the bias with respect to its average value (different in both bacteria) for genes with orthologous functions, indicative of the existence of a common selection pressure in these quite different organisms (Danchin *et al.*, 2000). We could not visualize any other cause except for an architectural selection pressure that would cause such a strong correlation. This should prompt further exploration of gene and gene product organization in bacteria.

Of course, one needed to find reasons to account for the selective pressures that would tend to cluster genes together. We have already refered to nanomachines, such as the ribosome, RNA polymerase and ATP synthase: the corresponding genes are indeed clustered into well organized operons. But one would like to see more. Gasses, which diffuse freely, play an important role in life, and not only at its origin. CO_2 and O_2, naturally, but also H_2S, and NO, CO or CH_3SH. Maintaining O_2 and H_2S together, in the presence of metal ions is like having a lit match in a gas station. Hence the idea that there must be compartmentalization of at least some metabolic complexes, in particular those dealing with the sulfur atom, quite prone to oxido-reduction (from -2 to +6). We looked for the distribution of sulfur related genes: they are without doubt clustered together into islands (Rocha *et al.*, 2000). Much more must be performed in this direction, but this is a

hint that some phenotypic characters, such as chemical reactivity, might exert sufficient pressure to get genes clustered together.

Bacteria have a variety of shapes. For a long time this was ascribed to some unknown process building up the cell wall, the murein layer(s) in particular. Within this frame of thought the cell's interior was still understood as a simple bag of chemicals, with no specific organisation. Work by many groups however recently challenged this view, when the distribution of Mre and Mre-like proteins was analyzed using GFP fusions (Carballido-López and Errington, 2003; Egelman, 2003). It seems clear that for Bacteria of the Firmicute family, those that have a bacillus form, in contrast to a coccus form possess a kind of cytoskeleton that determines the shape (Carballido-López and Errington, 2003). There are however certainly a variety of means to construct rod-shaped bacteria (Daniel and Errington, 2003), but it is most likely that this corresponds to a highly organized substratum (Errington, 2003; Errington *et al.*, 2003).

3. Conclusions

Many studies are now demonstrating that many gene products are compartmentalized in Bacteria, whether this is related to the organization of the genome is still open to question, but the many new genome sequences recently uncovered, rather than challenging this conjecture, make it all the more intriguing. Furthermore clustering appears not to be restricted to Bacteria but to extend to Eukarya as well (Blumenthal and Gleason, 2003; Lee and Sonnhammer, 2003). Future studies will tell us whether it was worth exploring.

Acknowledgements

We would like to thank all the colleagues who have throughout the years contributed to shape the content of the conjecture presented here. Their contribution is indicated as that of Stanislas Noria, who is the collective Team who creates a seminar in conceptual biology maintained at the Institut Pasteur in Paris, and, for some time, at the Department of Mathematics of the University of Hong Kong.

4. References

Akam, M. (1998). Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. Int. J. Dev. Biol. 42, 445-451.

Averof, M. (1997). Arthropod evolution: same Hox genes, different body plans. Curr. Biol. 7, R634-636.

Averof, M., and Akam, M. (1995). Hox genes and the diversification of insect and crustacean body plans. Nature 376, 420-423.

Bachiller, D., Macias, A., Duboule, D., and Morata, G. (1994). Conservation of a functional hierarchy between mammalian and insect Hox/HOM genes. EMBO J. 13, 1930-1941.

Balavoine, G., and Telford, M. J. (1995). Identification of planarian homeobox sequences indicates the antiquity of most Hox/homeotic gene subclasses. Proc. Natl. Acad. Sci. USA 92, 7227-7231.

Bennett, C. H. (1988). Logical Depth and Physical Complexity. In Herken, R. (ed.), The Universal Turing Machine: a Half-Century Survey. Oxford University Press, Oxford.

Bhattacharyya, A., Stilwagen, S., Reznik, G., Feil, H., Feil, W. S., Anderson, I., Bernal, A., D'Souza, M., Ivanova, N., Kapatral, V., Larsen, N., Los, T., Lykidis, A., Selkov, E., Jr., Walunas, T. L., Purcell, A., Edwards, R. A., Hawkins, T., Haselkorn, R., Overbeek, R., Kyrpides, N. C., and Predki, P. F. (2002). Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights. Genome Res. 12, 1556-1563.

Blum-Oehler, G., Dobrindt, U., Janke, B., Nagy, G., Piechaczek, K., and Hacker, J. (2000). Pathogenicity islands of uropathogenic *E. coli* and evolution of virulence. Adv. Exp. Med. Biol. 485, 25-32.

Blumenthal, T., and Gleason, K. S. (2003). *Caenorhabditis elegans* operons: form and function. Nat. Rev. Genet. 4, 112-120.

Bozinovski, S., Jovancevski, G., and Bozinovska, N. (2001). DNA as a Real Time, Database Operating System. Presented at the International Conference on Information Systems, Analysis and Synthesis, La Habana, Cuba.

Brock, H. W., and van Lohuizen, M. (2001). The Polycomb group--no longer an exclusive club? Curr. Opin. Genet. Dev. 11, 175-181.

Brunder, W., and Karch, H. (2000). Genome plasticity in Enterobacteriaceae. Int. J. Med. Microbiol. 290, 153-165.

Carballido-Lopez, R., and Errington, J. (2003). The bacterial cytoskeleton: in vivo dynamics of the actin-like protein Mbl of *Bacillus subtilis*. Dev. Cell 4, 19-28.

Cobourne, M. T. (2000). Construction for the modern head: current concepts in craniofacial development. J. Orthod. 27, 307-314.

Danchin, A. (2003). The Delphic Boat. Harvard University Press, Cambridge (Mass).

Danchin, A., Guerdoux-Jamet, P., Moszer, I., and Nitschke, P. (2000). Mapping the bacterial cell architecture into the chromosome. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355, 179-190.

Danchin, A., and Henaut, A. (1997). The map of the cell is in the chromosome. Curr. Opin. Genet. Dev. 7, 852-854.

Daniel, R. A., and Errington, J. (2003). Control of cell morphogenesis in bacteria: two distinct ways to make a rod-shaped cell. Cell 113, 767-776.

Dobrindt, U., and Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. Curr. Opin. Microbiol. 4, 550-557.

Duboule, D., and Dolle, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes. EMBO J. 8, 1497-1505.

Egelman, E. H. (2003). Cell walls, cell shape, and bacterial actin homologs. Dev. Cell 5, 4-5.

Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. Trends Biochem. Sci. 26, 597-604.

Ellwood, M., and Nomura, M. (1982). Chromosomal locations of the genes for rRNA in *Escherichia coli* K-12. J. Bacteriol. 149, 458-468.

Errington, J. (2003). Dynamic proteins and a cytoskeleton in bacteria. Nat. Cell Biol. 5, 175-178.

Errington, J., Daniel, R. A., and Scheffers, D. J. (2003). Cytokinesis in bacteria. Microbiol. Mol. Biol. Rev. 67, 52-65.

Gaunt, S. J. (1991). Expression patterns of mouse Hox genes: clues to an understanding of developmental and evolutionary strategies. Bioessays 13, 505-513.

Gehring, W. J. (1987). Homeo boxes in the study of development. Science 236, 1245-1252.

Gupta, R. S. (1998). Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol. Mol. Biol. Rev. 62, 1435-1491.

Hofstadter, D. (1979). Gödel, Escher, Bach: an Eternal Golden Braid. Basic Books, New York.

Holland, P. W. (1999). Gene duplication: past, present and future. Semin. Cell Dev. Biol**.** 10, 541-547.

Hou, Y. M. (1999). Transfer RNAs and pathogenicity islands. Trends Biochem. Sci. 24, 295-298.

Jagla, K., Bellard, M., and Frasch, M. (2001). A cluster of *Drosophila* homeobox genes involved in mesoderm differentiation programs. Bioessays 23, 125-133.

Judd, E. M., Ryan, K. R., Moerner, W. E., Shapiro, L., and McAdams, H. H. (2003). Fluorescence bleaching reveals asymmetric compartment formation prior to cell division in Caulobacter. Proc. Natl. Acad. Sci. USA 100, 8235-8240.

Kammermeier, L., and Reichert, H. (2001). Common developmental genetic mechanisms for patterning invertebrate and vertebrate brains. Brain Res. Bull. 55, 675-682.

Kaufman, T. C., Seeger, M. A., and Olsen, G. (1990). Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. Adv. Genet. 27, 309-362.

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S. C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L. C., Danchin, A., Debarbouille, M., Dervyn, E., Deuerling, E., Devine, K., Devine, S. K., Dreesen, O., Errington, J., Fillinger, S., Foster, S. J., Fujita, Y., Galizzi, A., Gardan, R., Eschevins, C., Fukushima, T., Haga, K., Harwood, C. R., Hecker, M., Hosoya, D., Hullo, M. F., Kakeshita, H., Karamata, D., Kasahara, Y.,

Kawamura, F., Koga, K., Koski, P., Kuwana, R., Imamura, D., Ishimaru, M., Ishikawa, S., Ishio, I., Le Coq, D., Masson, A., Mauel, C., Meima, R., Mellado, R. P., Moir, A., Moriya, S., Nagakawa, E., Nanamiya, H., Nakai, S., Nygaard, P., Ogura, M., Ohanan, T., O'Reilly, M., O'Rourke, M., Pragai, Z., Pooley, H. M., Rapoport, G., Rawlins, J. P., Rivas, L. A., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H. H., Scanlan, E., Schumann, W., Seegers, J. F., Sekiguchi, J., Sekowska, A., Seror, S. J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Tanaka, T., Takeuchi, M., Thomaides, H. B., Vagner, V., van Dijl, J. M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H., Zuber, U., and Ogasawara, N. (2003). Essential *Bacillus subtilis* genes. Proc. Natl. Acad. Sci. USA 100, 4678-4683.

Kramer, A., Neben, K., and Ho, A. D. (2002). Centrosome replication, genomic instability and cancer. Leukemia 16, 767-775.

Lee, J. M., and Sonnhammer, E. L. (2003). Genomic gene clustering analysis of pathways in eukaryotes. Genome Res. 13, 875-882.

Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. Mol. Biol. Evol. 13, 660-665.

Lohmann, J. U., and Weigel, D. (2002). Building beauty: the genetic control of floral patterning. Dev. Cell 2, 135-142.

Mayr, E. (1998). Two empires or three? Proc. Natl. Acad. Sci. USA 95, 9720-9723.

McLean, M. J., Wolfe, K. H., and Devine, K. M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J. Mol. Evol. 47, 691-696.

Medigue, C., Rouxel, T., Vigier, P., Henaut, A., and Danchin, A. (1991). Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol. 222, 851-856.

Merks, R. (1997). The molecular Bauplan.. http://carol.wins.uva.nl/~roel/bauplan/.

Reichert, H. 2002. Conserved genetic mechanisms for embryonic brain patterning. Int. J. Dev. Biol. 46, 81-87.

Rice, L. M., and Agard, D. A. (2002). Centriole duplication: centrin in on answers? Curr. Biol. 12, R618-619.

Rocha, E. (2002). Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? Trends Microbiol. 10, 393-395.

Rocha, E. P., and Danchin, A. (2003). Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat. Genet. 34, 377-378.

Rocha, E. P., and Danchin, A. (2001). Ongoing evolution of strand composition in bacterial genomes. Mol. Biol. Evol. 18, 1789-1799.

Rocha, E. P., Danchin, A., and Viari, A. (1999). Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. Mol. Biol. Evol. 16, 1219-1230.

Rocha, E. P., Fralick, J., Vediyappan, G., Danchin, A., and Norris, V. (2003). A strand-specific model for chromosome segregation in bacteria. Mol. Microbiol. 49, 895-903.

Rocha, E. P., Guerdoux-Jamet, P., Moszer, I., Viari, A., and Danchin, A. (2000). Implication of gene distribution in the bacterial chromosome for the bacterial cell factory. J. Biotechnol. 78, 209-219.

Rocha, E. P., Sekowska, A., and Danchin, A. (2000). Sulphur islands in the *Escherichia coli* genome: markers of the cell's architecture? *FEBS Lett.* 476, 8-11.

Silberschatz, A., Galvin, P. B., and Gagne, G. (2001). Operating System Concepts, 6th Edition ed. John Wiley & Sons, New York.

Stock, D. W. (2001). The genetic basis of modularity in the development and evolution of the vertebrate dentition. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 1633-1653.

Tillier, E. R., and Collins, R. A. (2000). The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. J. Mol. Evol. 50, 249-257.

von Neumann, J. 1958 (reed 1979). The Computer and the Brain. Yale University Press, New Haven.

Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., and Campbell, K. H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature 385, 810-813.

Woese, C. R. (1998). Default taxonomy: Ernst Mayr's view of the microbial world. Proc. Natl. Acad. Sci. USA 95, 11043-11046.

Wong, C., and Stearns, T. (2003). Centrosome number is controlled by a centrosome-intrinsic block to reduplication. Nat. Cell Biol. 5, 539-544.

Yockey, H. P. (1992). Information theory and molecular biology. Cambridge University Press, Cambridge (UK).