A challenge to vaccinology: Living organisms trap information

Antoine Danchin*

CEA/Genoscope, Amabiotics, 2, rue Gaston Crémieux, 91057 Evry Cedex, France

ABSTRACT

Life couples reproduction of the cell machinery with replication of the genetic program. Both processes are linked to the expression of some information. Over time, reproduction can enhance the information of the machine. We show that accumulation of valuable information results from degradative processes required to make room for novel entities. Degradation systems act as Maxwell's demons, using energy not to make room per se, but to prevent degradation of what has some functional features. This myopic process will accumulate information, whatever its source, in a ratchet-like manner. The consequence is that genes acquired by horizontal transfer as well as viruses will tend to perpetuate in niches where they are functional, creating recurrent conditions for emergence of diseases.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Emerging diseases are recurrent [1–3]. We argue that, although human behaviour has also its share in their development [4], this is mainly the consequence of a still poorly explored property of living organisms, their capacity to trap information. Innovation in consubstantial with life. The first step of colonisation of a niche requires access to novel information. Genome analyses show that it is inevitable that pathogens will keep emerging because living organisms are poised to recruit and even create information. Comparative genomics identified some of the genes and processes involved in this ubiquitous faculty. Bacterial genomes are organised in such a way that they are prone to acquire, invent or express genes permitting occupation of any novel niche. Our present view of information trapping is based on the generally intuitive idea that the cell (or the organism) is a material system able to handle information. Briefly, the cell behaves as a computer making computers (for a summary of the history of the conjecture and modern data stemming from the birth of synthetic biology, see [5]). The central tenet of the cell-as-a-computer model is that the genetic program coded in the genome couples information (an authentic category of reality) with the standard categories, matter, energy, space and time [6].

Recent work has identified some of the genes that permit this articulation. In a nutshell, bacterial genomes are split into two categories of genes, which both tend to stick together when they belong to different species [7]. The first category, the *paleome*, comprises

some 500 genes, which can be grouped into a scenario illuminating what could have been the origin of life [8]. These genes, which tend to persist in the genomes of free living bacteria, comprise about 250 "essential" genes [9]. In addition, a set of similar size was not found to be essential when experimental tests for essentiality explored the ability of bacteria to make colonies on rich media [10,11]. The second gene category, named the *cenome*, is not persistent but highly variable. It is made of genes extracted from a gene pool which for the time being, seems unlimited. The cenome genes permit the organism to occupy a particular niche. The genes which permit accumulation of information – essential in the emergence of diseases – belong to the paleome. Remarkably, these genes code for degradative functions that use energy to *prevent* degradation of any functional entity, thereby trapping information from whatever source it may come from. Subsequently, emerging diseases use genes of the cenome to colonise their novel environment.

2. Information is at the heart of life

Although we live in the computer era, we tend to forget the way computers are constructed and run. Briefly, computers are Turing Machines [12]. Their setup involves two separate entities, associated via a read/write process. On the one hand, a machine is moving a device that supports a linear string of symbols written in a finite alphabet; on the other hand the data made of the string of symbols, once read by the machine, triggers its future actions in a way highly reminiscent of the way messenger RNA is read by the ribosomes, as remarked by Woese in 1972 [13]. Many further features of the cell and the organism argue in favour of considering the cell as a (highly parallel) Turing Machine [5].

* Tel.: +331 6087 1158.
  E-mail address: antoine.danchin@normalesup.org.

The focal point of this model is that it assumes the *physical separation* between a *machine* (that the fashionable synthetic biology now names "chassis" [14]) and a *program*, supported by one or several linear strings of symbols. In this context, the crucial prediction of the model is that one should be able to isolate the entity carrying the program, put it back into a recipient host, and observe that the program in its new location begins to be displaying phenomena specific of the information it carries. In addition to the many experiments showing that pieces of program can be handled by cells (namely, viruses and genes moved by horizontal transfer, via conjugation, transformation and viral infection, or genetic engineering), two families of experiments support the model. 1/Animal cloning, first perfomed with the famous ewe Dolly [15] is now commonplace; 2/Lartigue and co-workers claimed success in transplanting the genome of a Mycoplasma species *M. mycoides*, into another Mycoplasma species, *M. capricolum*, and after several rounds of reproduction (reproduction of the machine and replication of the program) the host species was replaced by a colony of the *donor* genome [16]. Finally, the cell-as-a-computer model is also remarkable in that it is so consistent with the way viruses behave (pieces of program meant to self-replicate) that the metaphor went from biology to computer sciences, i.e. in the way opposite to that which made the concept of program come from computer sciences to biology.

Among the consequences of the cell-as-a-computer model for the understanding of emerging diseases is the rarely recognised fact that Archaea are not pathogenic. Indeed, within the frame of the model the three domains of life, Archaea, Bacteria and Eukarya, parallel computers with different operating systems (OSs). And, as we all know, many computer programs are impossible to run when going from a platform driven by a particular OS to another one. The same should be expected in genome transplantation experiments (and indeed transplantation of cyanobacteria genomes in Firmicutes did not yield any functional result [17]). However, the Eukarya OS is particular in that these organisms have embedded, first as symbionts and then as organelles, organisms coming from the Bacteria domain. Archaeal symbionts are extremely rare [18]. This supposes some type of interoperability between Eukarya and Bacteria, and might account for the remarkable observation that Bacteria, but not Archaea, can become pathogens of plants and animals.

Can this model give us clues about the way novel behaviour can emerge as the consequence of recruitment or creation of novel information, typically escaping or subverting the immune response of the host? The computer model identifies two parts where this process can be traced back, the chassis/machine, and the genetic program. Describing the former in detail is difficult, as we lack much understanding about the bacterial cell's organisation (despite recent progresses which shows that contrary to expectation the bacterial cytoplasm is certainly not a tiny test tube, but is extremely well organised [19–21]). An essential property of the chassis stems from its capacity of reproduction. Indeed, as shown of non-covalent compositional assemblies made of monomeric mutually catalytic molecules, reproduction may improve over time (this is not the case of replication) [22,23]. Yet, despite its importance, knowledge in the domain is still fairly limited, and will not be further discussed here. We will therefore now focus on the second part of the cell machine, accumulation of information by functions coded in the genetic program.

3. The genetic program and information: Maxwell's demons's genes

Back in 1991, to the great surprise of most investigators, it was found that at least half of the newly discovered genes had no significant counterpart in gene data libraries. Since then, genome programs kept uncovering an evergrowing number of novel genes with unknown functions. However many of these genes were found to correlate with pathogenicity, and they were named "pathogenicity islands" accordingly. In contrast, a limited number of genes, orthologs of which tend to persist in many genomes, form a core set of persistent genes which are often conserved in bacterial genomes [9]. This led some to think that one could define one common universal minimal genome, made of genes deemed essential. However, analysis of gene persistence in a great many genomes showed that the class of persistent genes is much larger than the approximately 250 genes required to permit a cell to make a colony on plates supplemented with rich media [10,11]. The persistent gene class, comprising some 500 genes, was named the *paleome* because it recapitulates the three phases of the origin of life: metabolism of small molecules on mineral surfaces, substitution of surfaces by an RNA-world where transfer RNA played a central role, and invention of template-mediated information transfer [8].

While the paleome contained most essential genes, it also contained a large category of persistent non-essential genes. Some of those coded for metabolic "patches", i.e. metabolic pathways that contribute to resolve chemical conflicts within the cell (many metabolites are incompatible with each other) [24]. The other moiety was made of genes coding for degradative functions, but not those which could easily be expected: many of the corresponding degradation functions degrade RNA or proteins *using energy*, while degradation usually produces energy.

This remarkable observation was placed in perspective when put in parallel with the way information is created in physical systems. Indeed, contrary to intuition, creation of information is reversible, and this implies that creation of information does not require energy [25]. Yet, information needs to be coupled to energy consumption as soon as its accumulation is considered. Accumulation of information requires any physical system to "make room" by erasing the memory used during the process of creation of information, and this requires energy [25–27]. However, making room by random erasure of the intermediary steps required for the creation of information would fail to accumulate valuable information [6,28]. It is therefore necessary that appropriate degradative systems use energy to tell from the bulk what is informative and what is not. These degradation systems would play the role of Maxwell's demons, sorting functional entities from the bulk which is prone to be degraded, by *preventing* degradation of all that has any functional value. We proposed as a basis for experimental validation, the conjecture that this is exactly the function of those genes of the paleome which are often found among persistent non-essential genes. An experimental setup which can select for adaptive mutations has been constructed, with the aim to identify genes responsible for accumulation of adaptive mutations [28] and preliminary results support the conjecture (Sekowska, Martens and Danchin, unpublished observations). The consequence of the existence of these genes is that living organisms are prone to trap information, whatever its source, and use it to build up an adaptive map of their environment.

4. Trapping genes for emerging diseases: the cenome

With the genetic set up just described, we see that cells tend to reproduce while perpetuating a progeny that accumulates information. But what could be the sources of novel information? Part must come from mutations of the pre-existing genetic program, but the likeliness that it will often produce valuable information cannot be high [29,30]. Interestingly, exactly as Maxwell's demons would behave, the active component of the accumulation process

put in action "myopic" functions. Indeed, they can only measure whether a component susceptible to be degraded or not is somehow functional. Many features can be used to this aim, including physico-chemical properties of the amino-acids of the protein build up (such as isomerisation of aspartate and asparagine, a ubiquitous and fast ageing process in proteins), association with specific substrates or changes in the rigidity of active enzyme complexes or simply recent expression [28]. Furthermore, nothing precludes that information coming from outside, such as that resulting from horizontal gene transfer, can be retained in the accumulation process. And as a matter of fact, beside the paleome the genetic program contains another set of genes, often highly specific to a given organism and resulting from horizontal gene transfer [31], the cenome (named after the concept of biocenose, community of organisms sharing a particular niche, proposed in 1877 by Karl Möbius).

As a consequence, colonisation of the organism's niche is performed using the set of genes forming the cenome, made from genes spread in the local environment [8]. The analysis of the cenome of a variety of bacterial species showed us that there are unlimited number of genes in the environment, which can be sampled as a strain-specific setup in a particular strain of a given species. This fact, observed in metagenomic studies, has been recently firmly established in the case of *Escherichia coli*, where individual strains add to a core of 2000 genes, some 2500 genes extracted from a reservoir of much more than the 20,000 *E. coli* cenome genes already identified by genome sequencing [32]. Cells have a genome of a limited size and horizontal gene transfer results in a constant flux of genes, with some entering the genome while others are deleted. In the absence of a selective process it is unlikely that genes would stay for a significant period of time in a particular strain. Hence, when we look for causes of emerging diseases we need not only to identify those genes that have a role in sustaining and propagating the disease, but also for constraints which somehow identify relevant genes and trap them in the genome in a more or less stable way.

Pathogenicity of living organisms often results from discovery of a novel niche, using genes that have been previously used in another context. Indeed, biological systems, which are submitted to the trio variation/selection/amplification, evolve by creating functions which recruit pre-existing structures, as do handymen with the material they collect in a more or less haphazard way. Furthermore nothing forbids that the pathogen be an altered microbial community rather than a single organism [33]. We have summarised above the energy-dependent myopic mechanism permitting selection of traits that cooperate to construct a functional property, such as pathogenicity.

Changes in lifestyle, and in particular in the way patients are treated in hospitals, create new niches for commensal bacteria. This is so general that it seems likely that many bacteria may become emerging pathogens in patient care units. *Acinetobacter baumanii* and *Staphylococcus epidermidis* illustrate this situation [34–36]. Analysis of the single nucleotide polymorphism in the genome of the latter showed that, in addition to genes of the cenome, adaptation involves genes of the paleome, in particular genes involved in stress adaptation, maintenance and repair, i.e. genes belonging to the persistent non-essential class [37]. The conclusion of this observation is that, when analysing emerging diseases, we should not only look for the process making the organism pathogenic or virulent, but we should look for the genes making these processes perennial. If we follow the conjecture presented above, that accumulation of information requires making room *using energy to prevent degradation* of functional entities (processes, pathways, structures), then we should target possible energy sources, which would stabilise the host/pathogen interaction, and identify the relevant degradative processes.

5. Prospective view for the cell-as-a-computer metaphor: emerging viral diseases

At this point we have only considered bacterial pathogens. The case of viruses is somewhat different, and unfortunately more likely than bacteria to be successful in making diseases emerge. The model of the cell-as-a-computer is particularly well suited to viral infection and propagation. Indeed, exactly as with computers, viruses are pieces of genetic program that can readily be expressed in many cell types, and they often need simply to find an entry door to a new cell type to be successful. Because the cell is measuring functional entities, preventing their degradation, many emerging viruses must be variants of existing viruses, preserving essential functions. In particular, shifting from a host to a new one can happen when mutations modify the virus envelope so that it can bind to a new receptor and then is internalized in a host's cell type where it is recognised as functioning. The stepwise capture of information is well illustrated by the influenza virus as it is usually a mild parasite of Anatidae (ducks, geese and the like) which can infect humans, often after having infected pigs as an intermediate host [38]. The structure of the small farm in China, with its pond and its pig under a roof (as in the Chinese character for 'family', 家) is remarkably adapted to this type of transition. It is obvious that large breeding farms, despite considerable control of hygiene, create the conditions appropriate for considerable virus infections. In the same way, we can see pox viruses (infecting buffaloes, sheep and monkeys) as possible sources of emerging pox infections in the future [39].

Many viruses responsible of zoonoses are of great concern [40]. SARS, caused by a coronavirus whose origin is probably bats, is a case in point [41]. An epidemic in pigs [42], where there was a tropism shift of a coronavirus between gut and lungs, illustrates what happened in 2002–2004 in humans. The spread of the disease is not easy to understand with simple epidemiological models (why almost no cases in Shanghai and many in Beijing?) and a model suggested a double epidemic where two viruses with different tropisms and/or virulence overlapped [43]. This type of situation should, in any event, be explored further, in particular considering the conjecture presented in this short review.

A further consideration needs to be taken into account, if trapping information is indeed at the heart of what life is. The process of information trapping we have described implies that artifice is considerably *less* dangerous than nature [44]. In the case of genetically modified organisms (GMOs), while the danger of plant GMOs looks fairly limited, the potential risk of using some animal GMOs is evident. In particular, humanisation of organs from animals carrying a wealth of retroviruses is a matter of extreme concern as the information steps needed to adapt from a previous host to humans are in very limited number. In this domain, epidemiological studies of the morbidity and mortality of butchers, slaughterhouse personnel, etc. would be most welcome to permit some evaluation of the potential risks associated with ordinary domestic animals.

Another feature of emergence of diseases is the adaptive power of vectors. In the case of unconventional pathogenic agents, such as prions, not much has been explored about the exact route of contamination. While contaminated food has been incriminated (but with no solid proof of contamination), the existence of affected wild animals and of contaminated pastures suggests that other ways should be explored. The possibility of a new type of vector, that of parasites multiplying intracellularly, has been investigated in a model study [45]. It shows that, depending on the background previous contamination of the host by the vector, one could witness either an epidemic scenario or sporadic cases, exactly as is observed at present. Parasites such as Microsporidia would be consistent with such a scenario.

Conflict of Interest

The authors state that they have no conflict of interest.

References

[1] Maillard JC, Gonzalez JP. Biodiversity and emerging diseases. Ann N Y Acad Sci 2006;1081:1–16.
[2] Merianos A. Surveillance and response to disease emergence. Curr Top Microbiol Immunol 2007;315:477–509.
[3] Wack A, Rappuoli R. Vaccinology at the beginning of the 21st century. Curr Opin Immunol 2005;17(4):411–8.
[4] Danchin A. Infection of society. As diseases have evolved to exploit the holes in our defences, including weaknesses in society, we have to reconsider our way of life, otherwise they will continue to haunt us. EMBO Rep 2003;4(4):333–5.
[5] Danchin A. Bacteria as computers making computers. FEMS Microbiol Rev 2009;33:3–26.
[6] Danchin A. Information of the chassis and information of the program in synthetic cells. Syst Synth Biol 2009;3(1–4):125–34.
[7] Fang G, Rocha EP, Danchin A. Persistence drives gene clustering in bacterial genomes. BMC Genom 2008;9(1):4.
[8] Danchin A, Fang G, Noria S. The extant core bacterial proteome is an archive of the origin of life. Proteomics 2007;7(6):875–89.
[9] Fang G, Rocha E, Danchin A. How essential are nonessential genes? Mol Biol Evol 2005;22(11):2147–56.
[10] Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2006;2:0008.
[11] Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, et al. Essential Bacillus subtilis genes. Proc Natl Acad Sci USA 2003;100(8):4678–83.
[12] Turing A. A. M. Turing's ACE Report of 1946 and Other Papers. In: Carpenter B, Doran R (eds.). Charles Babbage Institute reprint series for the History of Computing. Cambridge (Mass), MIT Press, 1946 (1986).
[13] Woese CR. The evolution of cellular tape reading processes and macromolecular complexity. Brookhaven Symp Biol 1972;23:326–65.
[14] Keasling JD. Synthetic biology for synthetic chemistry. ACS Chem Biol 2008;3(1):64–76.
[15] Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. Viable offspring derived from fetal and adult mammalian cells. Nature 1997;385(6619):810–3.
[16] Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison 3rd CA, et al. Genome transplantation in bacteria: changing one species to another. Science 2007;317(5838):632–8.
[17] Itaya M, Tsuge K, Koizumi M, Fujita K. Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. Proc Natl Acad Sci USA 2005;102(44):15971–6.
[18] Moya A, Pereto J, Gil R, Latorre A. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 2008;9(3):218–29.
[19] Danchin A, Guerdoux-Jamet P, Moszer I, Nitschké P. Mapping the bacterial cell architecture into the chromosome. Philos Trans R Soc Lond B Biol Sci 2000;355(1394):179–90.
[20] Errington J. Dynamic proteins and a cytoskeleton in bacteria. Nat Cell Biol 2003;5(3):175–8.
[21] Jun S, Mulder B. Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome. Proc Natl Acad Sci USA 2006;103(33):12388–93.
[22] Dyson FJ. Origins of life. Cambridge, UK: Cambridge University Press; 1985.
[23] Segre D, Shenhav B, Kafri R, Lancet D. The molecular roots of compositional inheritance. J Theor Biol 2001;213(3):481–91.
[24] Sekowska A, Danchin A. Frustration: physico-chemical prerequisites for the construction of a synthetic cell. In: Hicks MG, Kettner C, editors. Systems Chemistry. Frankfurt am Main: Beilstein Institut; 2009. p. 1–18.
[25] Landauer R. Irreversibility and heat generation in the computing process. IBM J Res Dev 1961;3:184–91.
[26] Bennett C. Notes on the history of reversible computation. IBM J Res Dev 1988;44:270–7.
[27] Sagawa T, Ueda M. Minimal energy cost for thermodynamic information processing: measurement and information erasure. Phys Rev Lett 2009;102 (E-Print):250602/1–/4.
[28] Danchin A. Natural selection and immortality. Biogerontology 2009;10:503–16.
[29] Orgel L. The maintenance of the accuracy of protein synthesis and its relevance to aging. Proc Natl Acad Sci USA 1963;49:517–21.
[30] Orgel LE. The maintenance of the accuracy of protein synthesis and its relevance to ageing: a correction. Proc Natl Acad Sci USA 1970;67(3):1476.
[31] Médigue C, Rouxel T, Vigier P, Hénaut A, Danchin A. Evidence for horizontal gene transfer in Escherichia coli speciation. J Mol Biol 1991;222(4):851–6.
[32] Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 2009;5(1):e1000344.
[33] Foster JA, Krone SM, Forney LJ. Application of ecological network theory to the human microbiome. Interdiscip Perspect Infect Dis 2008;2008:839501.
[34] Foster TJ. Immune evasion by staphylococci. Nat Rev Microbiol 2005;3(12):948–58.
[35] Massey RC, Horsburgh MJ, Lina G, Hook M, Recker M. The evolution and maintenance of virulence in Staphylococcus aureus: a role for host-to-host transmission? Nat Rev Microbiol 2006;4(12):953–8.
[36] Quinn JP. Clinical problems posed by multiresistant nonfermenting gram-negative pathogens. Clin Infect Dis 1998;27(Suppl. 1):S117–24.
[37] Wei W, Cao Z, Zhu YL, Wang X, Ding G, Xu H, et al. Conserved genes in a path from commensalism to pathogenicity: comparative phylogenetic profiles of Staphylococcus epidermidis RP62A and ATCC12228. BMC Genom 2006; 7:112.
[38] Kaleta EF, Hergarten G, Yilmaz A. Avian influenza A viruses in birds—an ecological, ornithological and virological view. Dtsch Tierarztl Wochenschr 2005;112(12):448–56.
[39] Vorou RM, Papavassiliou VG, Pierroutsakos IN. Cowpox virus infection: an emerging health threat. Curr Opin Infect Dis 2008;21(2):153–6.
[40] De Clercq K, Goris N. Extending the foot-and-mouth disease module to the control of other diseases. Dev Biol (Basel) 2004;119:333–40.
[41] Wong S, Lau S, Woo P, Yuen KY. Bats as a continuing source of emerging infections in humans. Rev Med Virol 2007;17(2):67–91.
[42] Laude H, Rasschaert D, Delmas B, Eleouët J. Le coronavirus respiratoire porcin PRCV: un virus émergent pas comme les autres. Virologie 1998;2: 305–16.
[43] Ng TW, Turinici G, Danchin A. A double epidemic model for the SARS propagation. BMC Infect Dis 2003;3:19.
[44] Danchin A, Nature. artifice and emerging diseases. EMBO Rep 2009;10(5):418–9.
[45] Ng TW, Turinici G, Ching WK, Chung SK, Danchin A. A parasite vector-host epidemic model for TSE propagation. Med Sci Monit 2007;13(3):BR59–66.