The confines of science: too much information for so little knowledge

Juli Peretó

One of my first activities as vice-rector for Research at the University of Valencia was to accompany my rector, the historian Pedro Ruiz, to a meeting of rectors of Catalan-speaking universities at the University of the Balearic Islands convened by the late rector Nadal Batle. It was July 1994 and the reason for such a pleasant visit, to constitute the Xarxa Vives d'Universitats, which would become a reality the following autumn in Morella. Rector Batle had a bold vision of the future, advanced to that of his other colleagues. He ceaselessly insisted on the need for the connection, the exchange of information between our institutions as a fundamental ingredient, as the indispensable foundation, to build the Catalan language university community. Universities had to work in a network in a networked world and it was necessary to be pioneers in an information society in full construction. The passage of time has not only not denied the vision of Nadal Batle but the expectations have been exceeded, because the network is becoming more compressed and more extensive, and because nobody conceives today the work of scientific research and higher education without the implication of connections and cohesive work, without information and communication technologies. Whether the Xarxa Vives has been successful in this area or not, that is another thing.

Now the speed of communications, the circulation of news, is vertiginous. I still remember the times when we had to go to the library and consult the Current Contents or the Biological Abstracts on Bible paper to keep ourselves imperfectly informed, with a time imposed, in the best of cases, by airmail and the revision of printed texts. Now we receive on the screen of our laptops the virtual indexes of the magazines before their publication and we learn about scientific advances from the press thanks to the competition between the news agencies of the main scientific publishing groups. The abundant and accelerated information is an inseparable part of our lives.

When physics lost its innocence with the warlike applications of what Joan Fuster called the vagaries of the atom, an army of physicists emigrated to the study of life. At that time, molecular-scale biology was in the maturation phase and offered unique opportunities for the brightest minds. The booklet «What is life?» by Erwin Schrödinger, one of the fathers of quantum mechanics, despite the irritation it caused in some chemists, fascinated many young people with the illusion of the power of the physical and reductionist explanation to understand life. In hindsight we realise to what extent those people who discovered the molecular structure of hereditary material or deciphered the genetic code were building the foundations of an intellectual building of still unsuspected proportions.

It does not matter if we move on a geological scale or during the life of the individual, the great enigmas of the origin of life, of embryonic development or of brain activity continue to be excessive unknowns, the last inaccessible borders. Because science has borders but these are mobile. They move away at a speed proportional to the sagacity of our questions, to the sharpness of our searching stilettos, to the penetration of our curiosity into reality. And, undoubtedly, we do not falter in this search because we recognise ourselves as heirs of a modern thought, such as the one represented by Michel de Montaigne when he says that: «Miracles are according to the ignorance we have of nature, not according to the being of the nature».

A few years ago the journalist John Horgan, emulating the Francis Fukuyama of *The end of history*, predicted the end of science in a book where he reviewed several branches of knowledge through interviews with scientists. I get the impression that one of the drawbacks of Horgan's approach was not taking into account that the perception each person has of the progress or limits of knowledge is dependent upon their age and their own contribution to knowledge. If the respondents are eminent and rather mature, they will find it difficult to appreciate the great changes, the radical novelties that may sweep away their vision of the world. Martí Domínguez has skilfully portrayed this situation in his novel *El retorn de Voltaire*. In a critical chapter, an older and sicker Voltaire, having arrived in Paris since his exile, suddenly receives Diderot and this floods him with an uncontrollable torrent of new ideas. Voltaire, bewildered, keeps thinking about them. He has vertigo before a mind more modern than his own: «What was that man talking about? Ah, it was another model of the world that came late, that was already invaluable to him, that no matter how much he wanted to believe in, it would always feel foreign».

However, there are exceptional minds that maintain a youthful vigour despite age. Without mentioning Horgan once, John Maddox, who directed *Nature* magazine for many years, ruthlessly refuted *What remains to be discovered*, a brilliant essay of more than 400 pages. In spite of his advanced age, Maddox's scrupulous and incorruptible critical spirit, who has observed from a privileged vantage point all the great scientific tendencies of the last fifty years, allows him to diagnose the numerous fissures of knowledge, the profound unknowns that remain and that will keep human minds occupied for centuries to come. «Every day I learn and unlearn at the same time», a verse by the poet Jordi de Sant Jordi, reflects the intellectual audacity of Maddox and others like him. Life is complexity that generates perplexity and, therefore, we must live with the permanent revision of our convictions.

Although a decade is a very short period, events, at least in biology, have categorically denied Horgan – as much or more than the heavy reality that has buried the end of Fukuyama's story. Because while these controversies occurred, biology experienced an unprecedented technological advance: since 1995, complete sequences of genomes of microorganisms, fungi, plants or animals, including humans, have been published. Today we have almost thousands of whole genomes accessible online and thousands and thousands of sequencing projects are underway around the world –a large part of which are executed by North American laboratories. Of course, without counting an undetermined number of finished genomes, or almost finished, kept private by the big biotech or pharmaceutical corporations.

The milestone of the millennium changeover was used to present the first draft of the human genome made by an international consortium led by the United States and

Great Britain. A great media presentation, with the testimony of Bill Clinton and Tony Blair, was accompanied by a bells chiming: we already have the complete encyclopaedia, of more than 3,000 million letters, that defines us as humans. It is obvious that it may sound as if to obtain the sequence of the human genome, or of any other organism, was a scientific achievement of first magnitude when it really is a purely technological milestone. It was the palpable demonstration that we had the tools and methods to spell out a genome. The genome sequence is the starting point of scientific projects, not the goal. Sequencing is routine and, nowadays, increasingly, it is fast and cheap. Nobel laureate Sydney Brenner says that genomes would have to be sequenced by prison tenants: the more abject the crime, the longer the assigned sequence.

Thus, having the sequence of a genome, in itself, has no use. Do we know what does this string of 3,000 million molecules, of repetitions of the four letters of the genetic alphabet mean, apart from representing perhaps the most boring text in history? Do we know how to read and interpret the meaning of this text? Here is one of the most remarkable challenges of contemporary science that must involve the concerted effort, at least, of biologists and computer scientists. In advance, the first tangible result of the new genomics has been to quantify our ignorance: in some cases, more than half of the sequenced genetic material has a totally unknown function for science. Are they remains of evolutionary shipwrecks of the past or structures that we do not yet know how to interpret? G is composed of «islands of meaning in a sea without meaning», using a verse by David Jou. Meaningless because they have lost it or because they have a semantics that is unknown to us.

Genetic information is pure digital information interpreted and executed by an intricate set of molecular machineries that, in turn, are its product. To put it in the Aristotelian way, the efficient causes are contained in the organism itself and that greatly disconcerts physicists. Despite the usefulness of the Cartesian comparison of cells and organisms with machines, this circularity of causes is an essential difference and is the primordial substance of biological complexity.

In the words of Richard Dawkins, genomes are rivers of digital information that were born more than 3,500 million years ago and that have been spreading and branching to reach the current genomes. Time and chance are the great genetic composers. The accumulation of changes and transformations of this digital information, exploring what is possible, crystallises in specific living beings with capacity to adapt to changing environments. Many organisms and their genomes have been the last of their kind and have disappeared. The natural destiny of living beings is extinction.

Now, the combinations of information that survive, perpetuate and reproduce themselves are an infinitesimal fraction of all possible ones. And so it has been, through the eons to this day. Each living being has two biographies: each has an individual one, the result of the deployment of the instructions of the genome, and contains a collective one, the history of their lineage inscribed and configured in their genome, a true archive of evolutionary memory. We, with all the rest of the biosphere, are at the end of an uninterrupted chain of digital information transmission. We are the buds at the end of the many branches of the tree of life. And we are here because life is stubborn but malleable, fragile but robust, clumsy but opportunistic. We are the result of a delicate balance between permanence and change.

There is still a stratum of complexity additional to the genome that must be considered. Think of a multicellular organism like the human being. The genome of our cells forms an essential part of physiology and our relationship with the environment. But the total understanding will come from the integration of this knowledge with the functioning of the organisms that coexist with us: our intestine contains between 10 and 100.000 million microorganisms – this means that there are an equal or greater number of cells in the intestine than the other cells that form tissues and organs.

This microbial community is the one with the highest density known, although it is not very diverse, a few microbial types dominate the whole. But in terms of information, the human microbiome – or set of genomes of the gut microbiota – contains a hundred times more genes than the human genome. For every human gene, a hundred genes of microorganisms work for us. Genomic studies illuminate a fascinating reality:

microbial metabolisms complement the human and interact and adapt to the diet, as a result of millions of years of coevolution. A useful and necessary microscopic universe that will give us many clues about health and disease.

While molecular biology remains essentially an experimental science, or as we say in English a wet science, we witness the emergence of a new biology that has blurred the boundaries between computer science and information technology. This is what Nobel laureate Walter Gilbert called a «paradigm transition» in 1991. Gilbert considered then that the reactive of the future would be the sequence, the series of letters of the genes and the genomes. But the historian and philosopher of biology Evelyn Fox Keller rightly pointed out in 1995 that if we look at it with a little more scrupulosity, once this transition is taken care of, the raw material of research in molecular biology will be networks. A quick glance at traditional scientific journals, such as *Nature* or *Science*, or some very new ones such as *Molecular Systems Biology*, shows us that we are already the future. Networks are everywhere. And, although we still have to wait until much of the literature published sediments and serenes a bit the collective enthusiasm that is felt, reticular thinking could configure, along with phylogenetic thinking, a new mentality that lightens the perplexity that the complexity of life causes us.

The important thing, then, is not the individual gene but its dialogue with the rest of the genome, the genomic conversation in each moment, in each place. In the context of the cell, the genome generates a complex network: thousands of elements that interact in different ways among themselves. The main difficulty we have is not gathering information, which accumulates exponentially in public databases that also have good genome management programs. We have the limitation in the concepts and the theories that we would need to interpret and integrate all this fabulous information in an understandable way. Of the genomic conversation we only catch a little intelligible murmur.

Internet scholars have realised that their methods can also be applied to the scrutiny of biological networks. And so, coinciding with the publication of the human genome, the first works that considered the cellular components as the elements of complex networks appeared. The existence of modules and hierarchies, of protocols, the phenomena of robustness and fragility, the analysis of architecture and the dynamics of metabolic and genetic networks have generated an intense debate and a true industry. The comparison of hackers with pathogens has transcended the mere metaphor and many research groups try to apply these ideas to biomedicine and biotechnology.

We have, then, a remarkable scientific challenge. The way of the solution is not likely to be marked by the physicists' conception of the world. They have tried without success for seventy years, but it is clear that biological phenomena do not follow the pattern of a great theory that explains everything. This is the dream of physicists that works reasonably well in the inanimate world. The great laws of physics, applied successfully by engineers, allow me to teach a class in Burjassot and a few hours later, to give a lecture in the city of Mallorca. And they can do even more spectacular things, such as allowing the exploration of the surface of Mars with a robot commanded by computers that are on Earth.

I have already mentioned it before: living matter has essentially different characteristics that make it refractory to the unifying vision of the world of physicists. One is that the causes are internal. In addition, these causes have been sculpted by time. They have, at the same time, a physical support and a historical origin, with all the contingent component that it implies. I mean, we do not have enough with purely physical principles – a description that has dominated the molecular biology of the twentieth century. In addition, we must look at it through the evolutionary prism.

Although we recognise a molecular, ancient fundamental unit, life is diverse and the solutions found to survive and perpetuate are multiple. The adaptability of living matter is singular. The point is precisely in that unpredictable determinism that characterises and disconcerts us. Therefore, it is more possible that the understanding of life progresses through the approach of engineers more familiar with the notions of control, regulation and sensitivity to disturbances.

A hundred years ago some biologists, inspired by the developments of organic chemistry, had the vision to propose that the understanding of the biological phenomenon would be achieved more by synthesis than by analysis. It is undeniable that the reductionism of biochemistry and the molecular biology of the twentieth century has allowed us to discover the intimacies of cells up to the last atomic limits. We have a very detailed knowledge of the structure and functioning of molecular machineries. Nanobiotechnology tries to exploit these teachings. But there are aspects manifested by the path of synthesis that do not emerge with the analytical approach.

Researchers soon realised that detailed knowledge of the individual elements informed little or nothing of the collective behaviour of the system. The properties of the whole are more than the sum of the properties of the parts. That is characteristic of complex systems and then it was necessary to focus the interest towards the emergence of new properties in living beings. If one hundred years ago the genuinely biochemical approach was called in vitro – that is, studying the parts extracted from the cell in a test tube –, today we demand the return to the living being, in vivo, not only by recomposing the parts in a flask, in a kind of post-genomic genetic engineering – the so-called synthetic biology –, but in the way of study and simulation in silico. Computational models of the genetic and metabolic networks begin to demonstrate their usefulness and power.

The moment we are living in science and, in particular, in biology can be characterised very well with the fortunate expression of John Naisbitt and Patricia Aburdense, the authors of the *Megatrends* series: we get fed up with information while we are famished with knowledge. It is almost like dying of thirst in front of the sea... This is the great paradox, the great dislocation of the contemporary information society. Biology is filled up by a flood of genetic digital information, but biologists lack contrasted theoretical tools, appropriate notions and illuminating concepts. And yet, the impulse of the human brain to understand is so powerful, that nothing will stop this race of humanity towards the confines of knowledge. An infinite but limited career, like the universe itself. A race full of obstacles. But, as Charles Darwin said, the progressive impulse of science is so strong that it comforts us of our mistakes and makes us forget the sacrifices, which are buried by a mountain of new observations and notions that appear every day. From science, there will always be distant horizons for our ignorance.