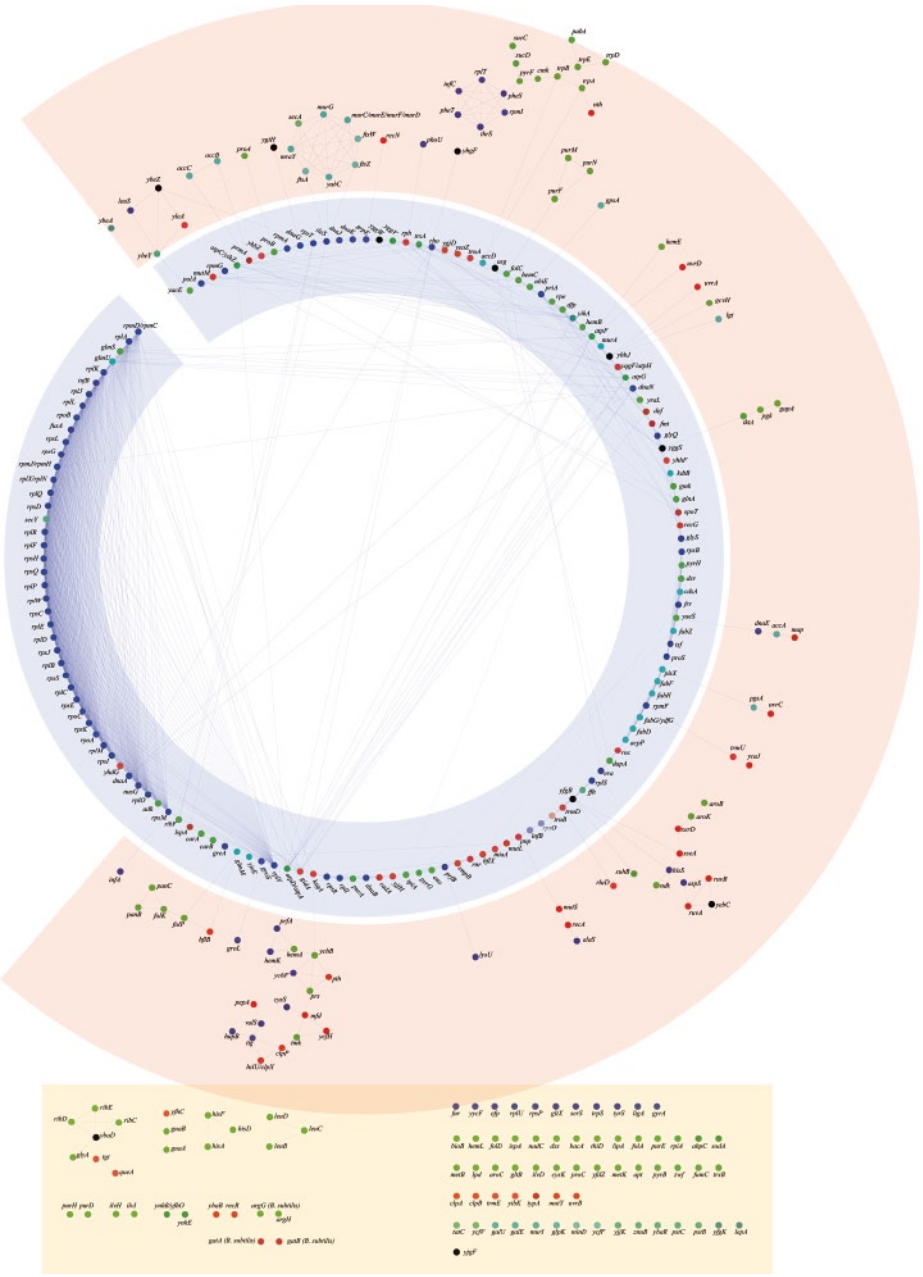
From the Origin of Life to Synthetic Biology

Engineering Cells

Synthetic biology aims at engineering cells to produce systems that behave as living factories. Bacterial genome analysis shows us that cells are already organized as if somewhat engineered, combining an ancestral genome meant to construct, replicate, maintain and repair the cell processes with a set of genes allowing the cell to develop in a particular niche. This organization, which tells us something about the origin of life, should be preserved – and perhaps improved – in our synthetic attempts.

Antoine Danchin, Director, Genomes and Genetics Department, Institut Pasteur

Fifty years ago Samuel Granick proposed that we were still very close to the origin of life. Hence, we needed to explore extant metabolism to understand the first steps of the process leading from minerals to life: *The experimental method whereby it is proposed to find the evolutionary precursors of protoplasm is to examine present-day biochemical reactions in protoplasm and seek to relate them to reactions that may have occurred and may still occur in the minerals around us* [1]. Naturally, at the time, knowledge was so scarce that his prescience stayed as a dead-letter. Suddenly, in the past few years, we contemplate the explosion of genome projects (more than 2,500 are underway or completed) and, with comparative genomics as a Rosetta stone we can identify what is common in genomes, allowing us to identify what makes the core of life. Of course this is not an easy task. Indeed acquisitive evolution has a tendency to preserve functions but not the structures that put them into action: several structures can have the same role in different organisms, preventing us to simply intersect all genomes and look for universal genes. As new genomes keep on being sequenced, the number of conserved genes decreases steadily and will soon get to zero.



Cytoscape [5] connection network of the mutual attraction between genes making the bacterial cell paleome. The inner blue circle is organized around information transfer and RNA, the pink circle connects to tRNA synthetases and features small networks of cell division; the beige rectangle displays genes that control synthesis of the basic coenzymes and building blocks of the cell. This organization is reminiscent of a plausible scenario for the origin of life.

Gene Persistence

This is not the end, however. We can look for genes which tend to stay in many genomes, keeping particular organizational features, such as their orientation with respect to the movement of the replication fork. This makes a set of some 500 „persistent“ genes that we may consider as essential to constitute extant life [2]. We can further explore the way these genes are distributed in genomes and, lo and behold, we observe that despite billions of years of separation from common ancestors, this core genome keeps a well-defined organization, and that it tells us something about the past [3].

Persistent genes behave as forming a network of mutual attraction with explicit relationships with the functions they code for. At the centre of the core genome is a highly connected network organized around genes coding for the process of RNA-mediated information transfer, the ribosome, the transcription machinery and maintenance of its properties. A second, less connected network is organized with domains centered on RNA metabolism, involving tRNA synthetases mostly from class I, the class assumed to be the most ancestral one, as the organizing principle. Furthermore, connected to some of these genes we find genes organizing cell division. Finally, a set of genes that are no longer connected to each other code for synthesis of the basic building blocks of the cell, nucleotides and amino acids, as well as the core catalytic centres of proteins, the coenzymes (with the presence of iron-sulfur enzymes or electron carriers), and genes allowing synthesis of the lipid bilayer that constitutes the cell's membrane. Going in a retrograde way this organization is reminiscent of a scenario of the origin of life, which could then be as follows.

From Surface Metabolism to the RNA World

A first primordial step would be creation of the building blocks and catalytic centres permitting construction of living entities. In this setup, organized around ferrous iron, the most important molecules are the coenzymes (which in fact are required to catalyze the needed chemical reactions), lipids (required for construction of the cell's membrane), some amino acids (by far not the most important pieces of the set up, despite the curious emphasis placed by the mass media on their presence elsewhere in the universe), and of course nucleotides. An apparent driving force is chaining of some building blocks together, with elimination of water molecules: this suggests some kind of surface metabolism, as this type of process would hardly occur freely in solution.

Synthesis of lipids allows formation of primitive cells that continously fuse and split, allowing both local concentration of some molecules and sharing of various metabolic processes. The second step involves tRNA-like molecules, suggesting that construction of some type of an RNA world is underway. Implicitly this assumes both that there was a process for continuous production of nucleotides and that they polymerized via a 3'–5' bond (which requires some kind of interaction with peptide-like molecules). These tRNA ancestors are the place where metabolism is kept ongoing, with creation of primitive tRNA synthetases. Finally, RNA folding uncovers the potentiality offered by complementarity between sequences, and creation of the genetic code. DNA appears only at a late stage, as a molecule that is much more stable in time and less versatile than RNA. It seems likely, from this organization, that the primordial evolution processes involved reticulated populations of cells rather than a single common progenote ancestor. One also observes in this organization that processes involved in maintenance appeared very early on.

The Paleome and the Cenome

Two lessons can be drawn from these observations. First, genomes are organized into two major components. The core genome – named the paleome, to remind us of its deep connection with the onset of life – has a fairly constant organization. It is constituted of a set of genes meant to construct and replicate the cell, associated to a set of similar size meant to maintain, clean up and repair the cell's building blocks and machinery. This genome component aims at propagating life, while fighting weathering and death. But organisms have to live in context, to occupy a particular niche (and this is what tells them from each other). This is controlled by another genome component – which we name the cenome, to remind us that it carries traits specific to a common niche, made of genes that drive sensing, exploration and scavenging. Second, this organization is reminiscent of features that engineers would like to implement in processes using cells as factories. The paleome is the core engine of the factory, it is multipurpose, and aims at making the factory work properly. The cenome codes for the specific purpose of the factory: to live in a particular niche. And, indeed, analysis of the corresponding genes can help us predict, just knowing a genome, what would be the normal environment of the cell which harbours it. Cells which belong to a common niche tend to share their genes, and this explains why it is so difficult to define a bacterial species: while individual organisms have highly related paleomes, their individual cenome can vary enormously. Is this an obstacle for synthetic biology?

Designing Cell Factories

Engineers like nuts and bolts that are as much standardized as possible. Engineering cells asks therefore to identify core components and try to synthesize them in a reproducible way. They also need to be organized into a consistent and constant pattern [4]. The cell factory needs to comprise a cell constructor, a replicator and a maintenance system: this is exactly what we find in the paleome. The cell factory has also a purpose: it can be to degrade xenobiotics, or to make particular fine chemicals or proteins, it can be to produce hydrogen... This is exactly the counterpart of the cenome. However, in the absence of external design the way the paleome and the cenome is organized in extant organisms is plagued with an inevitable noise. In real life, genes constantly come in and out of the cell, disrupting organization that took millenia to be selected for and the organization process has to start over and over again, slowly improving over time via fitness selection. The synthetic cell can get rid of these constraints: we will optimize the distribution of the genes in the constructor/replicator, and place the genes responsible for the designed purpose of the synthetic cell at proper positions, allowing both control of their expression and stability. A new era opens for biology, replacing natural selection by intelligent design…

References

[1] Granick S.: Annals New York Acad. Sci. 69, 292–308 (1957)

[2] Fang G. *et al.*: Mol Biol Evol 22, 2147–2156 (2005)

[3] Danchin A. *et al.*: Proteomics 7, 875–889 (2007)

[4] Pleiss J.: Appl Microbiol Biotechnol 73, 735–739 (2006)

[5] Shannon P. *et al.*: Genome Res 13, 2498–2504 (2003)

▸ **www.eMagazineBIOforum.com**

CONTACT:
Antoine Danchin
Genomes and Genetics Department
Genetics of Bacterial Genomes
Institut Pasteur
28 rue du Docteur Roux
75724 Paris Cedex 15, France
antoine.danchin@normalesup.org