Evolutionary influence of the protein network topology on gene organisation in artificial organisms

Carole Knibbe[1], Guillaume Beslon[1], Jean-Michel Fayard[2]

Prisma – INSA Lyon – France
BF2I – INSA Lyon – France

Although molecular biology provided us with huge data about individual genes, their roles cannot be fully understood unless we put them into the broader context of their interactions. Multiple links indeed exist between genes: they can share the same ancestral sequence, they can be neighbours on a chromosome, they can be coregulated or regulate each other's expression level, their proteins can interact physically or be involved in the same metabolic pathway. Those various points of view lead to different networks - different but not independent, notably because they are involved in a common evolutionary story. Understanding these structures therefore requires to understand how they build up and interact at the evolutionary time scale.

In particular, are genes randomly distributed on the chromosomes, or does their organisation depend on their functional interactions? It is difficult to answer this question directly for living cells, notably because the notion of gene function is subjective in this context. We therefore chose an artificial life approach, using the fuzzy logic formalism as a generic framework for the functional level.

In our model, called aevol, each organism is able to perform abstract processes with various degrees, depending on the set of interacting functional elements ("proteins") its genome encodes. The genome is a binary string that contains coding sequences (genes) separated by non-coding sequences. Each gene is translated into a protein, able to either realize or inhibit a fuzzy set of processes. Two proteins can have the same process in their fuzzy sets, implying that for the organism, the degree of the process depends on their interaction. Thus, the organism's phenotype is the logic combination of the sets of its proteins. Like in the genetic algorithms used in computer science, selection and variation mechanisms allow the phenotypes to become closer and closer to an arbitrary optimal fuzzy set. When the fittest organisms reproduce, their genomes are replicated with eventual random errors, at the local scale (punctual mutations, small insertions, small deletions) or at the scale of large genomic segments (duplications, translocations, large deletions, inversions).

Since the fitness computation does not include genomic criteria like genome length or gene order, the genome is free to self-organise. Similarly, the functional network is not predefined, and the number of proteins can evolve, as well as the strengths of their interactions. We can however set the maximal interaction potential of the proteins - that is to say the maximal number of processes they can be involved in -, thereby limiting the network's average connectivity. This feature is expected to have evolutionary consequences, since deleting a highly connected gene, for instance, has a much more deleterious effect than deleting an "independent" gene.

To investigate the evolutionary influence of the functional network's topology on the genomic level, we carried out experiments with five different maximal interaction potentials, during 30,000 generations. In most cases, the gene organisation alternates shuffling phases with steady states throughout the evolution, and these steady states seem to depend on the functional network topology. If the maximal interaction potential is high, hence if many genes are highly connected, genes tend to distribute regularly along the chromosome after a shuffling phase. On the contrary, if the interaction potential is low, genes tend to form clusters. For intermediate values of the parameter, gene distribution either remains random, or alternates both steady states.

This emerging relation between the genomic and the functional levels could be the result of selective pressures for both robustness and evolvability, requiring an intermediate effect for the large-scale mutations – not too high, but also not too small. When genes are highly connected, this intermediate effect can be achieved by affecting one gene only, whereas when genes are more independent, affecting several of them at the same time can speed up the discovering of new solutions.

An evolutionary process could therefore make links emerge between different organisation levels in our artificial complex system. To confirm these results, additional experiments will be carried out with extreme values of the parameter, while further analysis of the impact of the mutational events should help us testing our explanatory mechanism.