Stephan Fischer
INSA-Lyon, CNRS, INRIA, LIRIS, UMR5205, F-69621, France
e-mail: stephan.fischer@insa-lyon.fr
Carole Knibbe
Université Lyon 1, CNRS, INRIA, LIRIS, UMR5205, F-69622, France
e-mail: carole.knibbe@liris.cnrs.fr
Samuel Bernard
Université Lyon 1, CNRS, INRIA, Institut Camille Jordan, UMR 5208, F-69222, France
e-mail: bernard@math.univ-lyon1.fr
Guillaume Beslon
INSA-Lyon, CNRS, INRIA, LIRIS, UMR5205, F-69621, France
e-mail: guillaume.beslon@liris.cnrs.fr

Unravelling laws of genome evolution with both mathematical and individual-based models

In order to investigate laws of evolution of genome organization over large evolutionary time scales, our lab has developed an individual-based model simulating Darwinian selection and most of mutations and rearrangements undergone by a chromosome during asexual reproduction. In particular, the length of the chromosome and the number and lengths of genes are free to vary. It was shown that evolutionary success depends not only on the fitness but also on an appropriate trade-off between genome robustness and variability. This indirect selective pressure regulates the amount of coding DNA, but also, more surprisingly, the amount of non-coding DNA, if large rearrangements are taken into account. The higher the spontaneous rate of duplications and deletions, the more compact the genome in the surviving lineages [1].

This phenomenon is reminiscent of the error-threshold effect described by Eigen in the quasispecies theory [2, 3], where the per-digit mutation rate q sets a maximum number of digits ν that can be reproducibly preserved: $\nu < -\frac{\ln(\sigma_0)}{\ln(q)}$, where σ_0 is a parameter quantifying the fitness superiority of the fittest sequence. If the mutation rate is increased beyond this limit, then the population structure breaks down, and the population disperses over sequence space. However, this effect was mostly studied in the special case where all sequences have an equal length and only point mutations can occur. In these conditions, the maximum chain length ν_{max} applies only to the segments that contribute to fitness [3], and thus cannot directly explain our results regarding the amount of non-coding DNA.

The computational model cannot be considered as an analytic proof of the observed relation. Here, we combine the intuition and power of this model with a mathematical analysis. By relaxing Eigen's hypotheses, we developed simpler dynamical models that exhibit essentially the same behavior as the original computational model as far as genome length and coding/non-coding ratio is concerned. These models yield a better insight on the impact of essential parameters and provide valuable feedback for computational simulations. In return, these computational improvements lead to new relations and limits that can be investigated mathematically, closing the emulation loop.

References

[1] C. Knibbe, A. Coulon, O. Mazet, J.M. Fayard, G. Beslon (2007). A Long-Term Evolutionary Pressure on the Amount of Noncoding DNA. *Molecular Biology and Evolution* **24(10)** 2344–2353.

[2] M. Eigen (1971), Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58(10)** 465–523.

[3] M. Eigen, J. McCaskill, P. Schuster (1988). Molecular Quasi-Species. *Journal of Physical Chemistry* 92:6881-6891.