

Evolution as a hard combinatorial problem

D. Grigoriev,

CNRS, Université de Lille, dmitry.grigoryev@math.univ-lille1.fr

J.Reinitz,

Dept.of Statistics, Chicago University, reinitz@galton.uchicago.edu

S. Vakulenko,

Institute for Mechanical Engineering Problems, Saint Petersburg, vakulenfr@mail.ru

A. Weber,

Computer Science Department, University of Bonn, weber@cs.uni-bonn.de

We consider complexity emergence problem and formation of complex organs. We exploit recent ideas from theoretical computer science [1, 2] that allows us to formulate the problem in a rigorous mathematical way. The difficulty in organ evolution explanation was understood still by Ch. Darwin. The problem is a simultaneous development of many organism traits. It seems, therefore, that the complex organ formation is not a “feasible” problem. To shed a light on this feasibility problem, we use an analogy between these evolution processes and hard-combinatorial problems, which have been received a great attention of mathematicians and theoretical physicists. The fundamental hard-combinatorial problem is the K-SAT. Let us consider a set of n Boolean variables and a set of m clauses. The clauses are disjunctions involving K variables (or negations of the ones). The problem is to test whether one can satisfy all the clauses by an assignment of boolean variables. The enigma of theoretical computer science, $P \neq NP$, is equivalent to the question: exists there an algorithm solving the K-SAT problem in a polynomial number $Poly(n)$ of steps. A biological interpretation of K-SAT is clear. The number $n \gg 1$ is the gene number. Each gene is involved in formation of many phenotype traits, and the gene can be either turned, or turned off. We have $m \gg 1$ traits that corresponds to a formation of a complex organism with many traits. The parameter K determines genetic

redundancy (a trait can be formed by K different genes). The main difficulty that a logical variable can be involved in many different clauses (gene pleiotropy). Namely, activation of a gene can help to create a useful feature but can become an obstacle for another useful trait formation. We are dealing with a randomly generated K-SAT formula. The parameter $a = m/n$ defines the K-SAT asymptotic behaviour for large n and it can be interpreted as "gene freedom parameter". For values of $a < b(K)$, where $b(K)$ is a critical level, a great ($n \gg 1$) random K-SAT problem has a solution with a probability close to 1 [1]. There is a second threshold value, $d(K) < b(K)$. If $a < d(k)$, all the solutions form a huge cluster. Solutions in this cluster can be found by algorithms of local search running in $Poly(n)$ time such as WalkSat, GSat, DPLL and others, see [3]. Local search algorithms do not work beyond clustering phase transition. This interpretation of the evolution problem as a hard combinatorial problems allows us to formulate rigorously what means feasible evolution. Evolution "in silico" is feasible, if one can find a local search algorithm that resolves the problem in $Poly(n)$ elementary steps (for instance, mutations). Our numerical simulations for special random Boolean formulas (more adapted to biology than K-SAT) show that simple algorithms based on mutations and selection resolve K-SAT for $\log(a) < K$. This result can be generalized for more complicated combinatorial problems associated with genetic networks that were proposed for Drosophila morphogenesis [4]. Our main conclusion is that a genetic redundancy, when K different gens encode the same trait, provides an **exponential effectiveness** of organism morphogenesis and evolution: with n genes one can obtain approximately $n2^K$ traits.

[1] Friedgut E., (1999) J. Amer. Math. Soc., **12**, no. 4, p.1017-1054.

[2] Moore C. and Mertens S., The Nature of Computation, Oxford University Press (2011)

[3] Selman B, Levesque H. and Mitchell D, Proceedings AAAI-92, San Jose, pp. 440-446.

[4] Reinitz J. and Sharp D.H., (1995) Mechanisms of Development, 49, 133-158.