

Reconstruction of mathematical models of living systems: problems and prospects

Lashin S.A.

Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia

lashin@bionet.nsc.ru

Biological systems are complex hierarchical systems functioning at different levels of biological organization, from the molecular-genetic to the ecosystem and biocenosis. A key feature of biological systems is their ability to evolve, that is, to gradually change their properties over generations of living organisms. The study of the mechanisms underlying evolutionary processes is one of the most important tasks of modern biology, which often involves the apparatus of other sciences – physics, mathematics, and computer science.

Here we present methods for modeling the functioning and evolution of complex biological systems, which take into account the widest possible range of levels of biological organization (molecular-genetic, metabolic, cellular, population and ecological levels) within the framework of a single model. The methods are implemented in the form of software packages “Haploid Evolutionary Constructor 3D” (HEC 3D) and “Diploid Evolutionary Constructor” (DEC), “Population Genetics of Deafness – Simulator” and “Microcosm”. The key feature of the proposed methods is that the constructed models have a multilayer architecture, with each layer being a submodel responsible for describing a certain level of biological organization: genetic, metabolic, population and, finally, ecological levels.

生命系統數學模型的重構：問題與展望

拉欣公司

Lashin S.A.

ICG SB RAS, 新西伯利亞, 俄羅斯

lashin@bionet.nsc.ru

生物系統是複雜的等級系統，在不同層次的生物組織中發揮作用——
從分子遺傳到生態系統和生物群落。

生物系統的關鍵特性是它們的進化能力，即在一系列生物體的世代中逐漸改變它們的特性。
研究進化過程背後的機制是現代生物學最重要的任務之一，為此經常涉及其他科學的設備——
物理學、數學、計算機科學。

該報告介紹了對複雜生物系統的功能和進化進行建模的方法，在一個模型的框架內考慮了生物組織水平的最廣泛範圍（分子遺傳、代謝、細胞、種群和生態水平）。

這些方法以軟件系統“Haploid Evolutionary Constructor 3D” (HEC 3D) 和“Diploid Evolutionary Constructor” (DEC)、

“人口遺傳學耳聾-模擬器”和“Microcosm”的形式實施。

所提出方法的關鍵特徵是構建的模型具有多層架構，每一層都是一個子模型，負責描述特定級別的生物組織：遺傳、代謝、種群，最後是生態級別。

DOI 10.18699/sblai2023-24

Об актуальных задачах моделирования эволюционирующих систем

Пальянов А.Ю.^{1, 2, 3*}, Пальянова Н.В.^{2**}

¹ Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия

² НИИ вирусологии ФИЦ фундаментальной и трансляционной медицины,
Новосибирск, Россия

³ Новосибирский национальный исследовательский государственный университет,
Новосибирск, Россия
* palyanov@iis.nsk.su, ** natalia.palyanova@gmail.com

Возможность компьютерного моделирования жизненного цикла и размножения простейшего биологического организма с детализацией до геномного уровня стала бы научным прорывом, однако это по-прежнему находится далеко за пределами возможностей современных суперкомпьютеров. Процесс естественного отбора наиболее приспособленных особей происходит с учетом огромного количества факторов как внешней, так и внутренней среды. Особенности организма реализуются через наборы особенностей белков, а влияние изменений каждого белка на приспособленность оценить достаточно трудно в связи с необходимостью учитывать все возникающие изменения взаимодействий этого белка со всеми факторами среды и другими белками, число которых весьма значительно.

В компьютерных моделях эволюционирующих объектов, как правило, внесение изменений в геном потомков осуществляется не явным образом (посредством воспроизведения молекулярных механизмов), а лишь имитируется посредством описания алгоритмов внесения изменений в копию генома предков. Однако и сами механизмы внесения мутаций и горизонтального переноса генов являются субъектами эволюции, и среди возможных изменений, не приводящих к гибели или стерильности особи, встречаются и те, что влияют на скорость и точность репликации генома. Благодаря этому возникает внутривидовая конкуренция, в результате которой, например, для Sars-CoV-2 с момента его появления и до настоящего времени длительность инкубационного периода, напрямую связанная со скоростью репликации вируса, постоянно снижается [1].

По сравнению с клеточными формами жизни вирусы представляются значительно более удобными объектами для изучения и компьютерного моделирования эволюции благодаря достаточно простому устройству и значительно меньшему геному при широком спектре взаимодействий с внешней средой и организмом хозяина. До появления технологий быстрого секвенирования геномов эволюцию вирусов можно было рассматривать лишь в рамках моделей «паразит – хозяин», описывающих статистические, но не молекулярные особенности их взаимодействия. С начала пандемии Sars-CoV-2 мировым научным сообществом было получено более 15 млн вариантов геномов данного коронавируса, охватывающих весь временной интервал с момента его появления, предоставляющих беспрецедентно обширные данные о его эволюции с учетом географии и времени появления.

Из этих данных может быть рассчитана динамика распространения и изменения вируса не только в физическом пространстве и времени, но и в многомерном пространстве возможных жизнеспособных вариантов вирусных геномов с метрикой, определяемой минимальным числом единичных изменений (мутация, делеция или вставка), необходимых для преобразования одного генома в другую. При этом вирус изменяется, в том числе в ответ на вакцинацию и формирование иммунитета у переболевших. Также может быть оценен эволюционный потенциал каждого участка генома, рассчитана 3D структура вирусных белков, рассчитано взаимодействие S-белков с разными вариантами рецепторов ACE2 и многое другое. В итоге это позволит построить новую, существенно более реалистичную модель эволюции вируса, способную рассчитывать потенциально возможные, но еще не реализованные в природе варианты новых геномов, чтобы заблаговременно противостоять их появлению.

Исследование выполнено за счет гранта Российского научного фонда (проект № 23-64-00005).

Литература

1. Malone B., Urakova N., Snijder E.J., Campbell E.A. Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Review. Molecular Cell Biology*. 2022, 23(1): 21-39.

On actual problems of evolving systems modeling

Palyanov A.Yu.^{1,2,3*}, Palyanova N.V.^{2**}

¹ A.P. Ershov Institute of Informatics Systems of SB RAS, Novosibirsk, Russia

² Research Institute of Virology, FCR of Fundamental and Translational Medicine,
Novosibirsk, Russia

³ Novosibirsk State University, Novosibirsk, Russia

* palyanov@iis.nsk.su, ** natalia.palyanova@gmail.com

The possibility of computer simulation of the life cycle and reproduction of the simplest biological organism with detailing down to the gene level would be a scientific breakthrough, but this is still far beyond the capabilities of modern supercomputers. The process of natural selection of the fittest individuals takes into account a huge number of both external and internal environment factors. Individual characteristics of an organism are realized as combination of protein features, and the effect of changes in each protein on fitness is difficult to measure due to the need to take into account all the emerging changes in the interactions of this protein with all environmental factors and other proteins, the number of which is tremendous.

Usually in computational models of evolving objects changes in the genome of descendants are not carried out explicitly (by reproducing molecular mechanisms), but only imitated by describing algorithms for making changes to a copy of the ancestral genome(s). However, the mechanisms of introducing mutations and horizontal gene transfer are subjects of evolution as well, and among the possible changes that do not lead to the death or sterility of an individual, there are also those that affect the speed and accuracy of genome replication. Due to this, intraspecific competition arises, and, for example, for Sars-CoV-2 it results in constant decreasing of the rate of virus replication (which is directly related to the duration of the incubation period) from the moment of its appearance to the present time [1].

Compared to whole organisms and unicellular, viruses seem to be much more convenient objects for investigation and computational modeling of evolution due to a fairly simple internal organization and a small genome, along with a wide range of interactions with the environment and the host organism. Before the advent of rapid genome sequencing technologies, the evolution of viruses could only be considered within the framework of “parasite–host” models that describe statistical, but not molecular, features of their interactions. Since the beginning of the Sars-CoV-2 pandemic, more than 15 million variants of the genomes of this coronavirus have been obtained by the global scientific community, providing unprecedentedly extensive data on its evolution taking into account geography and time of occurrence.

From these data, the dynamics of the spread and change of the virus can be calculated not only in physical space and time, but also in the multidimensional space of possible viable variants of viral genomes with a metric determined by the minimum number of single changes (mutation, deletion or insertion) required to transform one genome to another. During this process the viral genome accumulates changes, including those which appear in response to vaccination and the formation of immunity in recovered patients. It is potentially possible to evaluate the evolutionary potential of each region of the genome, to predict the 3D structure of viral proteins, to estimate the interaction of S proteins with different variants of ACE2 receptors, and much more. As a result all these will allow to develop a new, much more realistic model of virus evolution, capable of calculating potential variants of new genomes that have not yet been realized in nature, in order to resist them in case of their emergence.

This work was funded by the Russian Science Foundation grant (project No. 23-64-00005).

References

1. Malone B., Urakova N., Snijder E.J., Campbell E.A. Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Review. Molecular Cell Biology*. 2022, 23(1): 21-39.

關於演化系統建模的熱點問題

Palyanov A.Yu.^{1, 2, 3*}, Palyanova N.V.^{2**}

¹ IIS SB RAS, 新西伯利亞, 俄羅斯

² 俄羅斯新西伯利亞 FRC FTM 病毒學研究所

³ NSU, 新西伯利亞, 俄羅斯

* palyanov@iis.nsk.su, ** natalia.palyanova@gmail.com

計算機模擬最簡單的生物有機體的生命週期和繁殖到基因水平的可能性將是一個科學突破，但這仍然遠遠超出現代超級計算機的能力。

最適者的自然選擇過程考慮了大量因素，包括外部和內部環境。

生物體的特徵是通過一組蛋白質特徵來實現的，由於需要考慮這種蛋白質與所有環境因素相互作用的所有新變化，因此很難評估每種蛋白質的變化對健康的影響其他蛋白質，其中的數量非常重要。

在進化對象的計算機模型中，通常不會明確地（通過複製分子機制）進行後代基因組的變化，而只是通過描述對祖先基因組副本進行更改的算法來模仿。

然而，引入突變和水平基因轉移的機制本身就是進化的主題，在不會導致個體死亡或不育的可能變化中，也有一些會影響基因組複製的速度和準確性。

因此，出現了種內競爭，其結果是，例如，對於 Sars-CoV-

2，從它出現的那一刻到現在，潛伏期的持續時間與病毒的傳播速度直接相關複製，正在不斷減少[1]。

與細胞生命形式相比，病毒似乎是更方便的研究對象和進化的計算機建模，因為它具有相當簡單的設備和更小的基因組，並且與環境和宿主生物體有廣泛的相互作用。

在快速基因組測序技術出現之前，只能在“寄生蟲-

宿主”模型的框架內考慮病毒的進化，該模型描述了它們相互作用的統計而非分子特徵。

自 Sars-CoV-2 大流行開始以來，全球科學界已收到超過 1500

萬個這種冠狀病毒基因組的變體，涵蓋了自其出現以來的整個時間間隔，提供了關於其進化的空前廣泛的數據，考慮到出現的地理和時間。

從這些數據中，不僅可以在物理空間和時間上計算病毒傳播和變化的動態，還可以在病毒基因組可能的可行變體的多維空間中計算，其度量由單個變化的最小數量決定（突變、缺失或插入）將一個基因組轉化為另一個基因組。

在這種情況下，病毒會發生變化，包括對疫苗接種的反應和患病者免疫力的形成。

還可以評估基因組每個區域的進化潛力，計算病毒蛋白的 3D 結構，S 蛋白與不同 ACE2 受體變體的相互作用，等等。

因此，這將有可能建立一個新的、更現實的病毒進化模型，能夠計算尚未在自然界中實現的新基因組的潛在變異，以提前防止它們的出現。

該研究得到了俄羅斯科學基金會的資助（項目編號 23-64-00005）。

文學

1. Malone B., Urakova N., Snijder E.J., Campbell E.A. Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Review. Molecular Cell Biology*. 2022, 23(1): 21-39.