

ПОСТРОЕНИЕ ЛОГИКО-ВЕРОЯТНОСТНОЙ МОДЕЛИ КОДИРУЮЩИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНОМА E-COLI

Г.С.Лбов, В.А.Гусев

Институт математики им. С.Л. Соболева СО РАН, Новосибирск, пр. акад. Коптюга, 4,
630090, Россия, E-MAIL: vgus@math.nsc.ru
Новосибирский государственный университет, г. Новосибирск

Обширная библиотека расшифрованных генетических последовательностей (GenBank) является в настоящее время объектом пристального внимания математиков. Как правило, работы по математическому анализу геномов посвящены применению различных математических методов для выявления регуляторных и кодирующих участков в структуре геномов [1,2]. Этот анализ основан на сопоставлении различных символьных последовательностей исследуемых геномов с паттернами ДНК, функции которых уже известны.

В работах [3,4] авторы проводили анализ структуры генетического кода, то есть кодон-аминокислотного соответствия основываясь на методах теоретико-группового анализа. Это позволило обнаружить неизвестные ранее молекулярным биологам и биохимикам алгебраические и арифметические закономерности в структуре кода.

Логично предположить, что соответствующие алгебраические и арифметические, то есть символьные и числовые закономерности генетического кода должны иметь отображения в нуклеотидной последовательности геномов. Для поиска числовых закономерностей необходимо представить стандартную символьную последовательность из АТ и ГС пар в геноме в цифровом виде. Мы воспользовались данными работы [5], в которой показано, что молекулярные массы АТ и ГС пар в составе двойной спирали равны соответственно 259 и 260 независимо от ориентации, то есть принадлежности нуклеотида к конкретной нити ДНК. Таким образом, чередование АТ и ГС пар в двойной спирали соответствует чередованию только двух чисел. Для анализа закономерностей в этих числовых рядах, их удобно представить в виде последовательностей нулей и единиц. Поиск возможных закономерностей в таком бинарном ряду чисел был проведен в рамках подхода к анализу эмпирической информации, изложенного в монографии [6]. Подход сводится к построению логико-вероятностной модели объекта исследования. Целью данной работы является поиск закономерностей в бинарных последовательностях, сопоставленных кодирующим фрагментам генома в классе логических решающих функций.

THE LOGICAL-AND-PROBABILISTIC SIMULATION OF THE CODING GENOME DNA SEQUENCES

G.S.Lbov, V.A.Gusev

Institute of Mathematics SB RAS, Acad. V. Koptug pr., 4. Novosibirsk, 630090, Russia; e-mail:
vgus@math.nsc.ru

A simplest mathematical model of genome is represented as the sequence of binary numbers. Analysis of the model by the method of logical decision functions demonstrates that the sequence has some regularities unknown before.

Литература

1. Abnizova I., Schilstra M., teBoekhorst R., Nehaniv C.L. A statistical approach to distinguish between different DNA functional parts. //WSEAS Transactions on Computational Methods. 2003. V.2. P. 1188-1196.
2. Abe T., Kanaya S., Kinouchi M., Ichiba Y., Kozuki T. and Ikemura T. Informatics for unveiling hidden genome signatures. //Genome Res. 2003. V. 13. P. 693-702.
3. Jimenez-Montano M.A., de la Mora-Basanez C.R. & Poschel T. The hypercube structure of the genetic code explains conservative and non-conservative aminoacid substitutions in vivo and in vitro. //BioSystems. 1996. V. 39. P. 117-125.
4. Negadi T. Rumer's Transformation, in Biology, as the Negation, in Classic Logic. //Int. Journ. of Quant. Chem. 2003. V.94. P. 65-82.
5. shCherbak V. I. Arithmetic inside the universal genetic code. //BioSystems. 2003. V. 70. P. 187-209.
6. Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопросы статистической устойчивости решений. – Новосибирск: Издательство института математики, 1999. 211 стр.