Секция «Математика и механика»

Оптимальные линейные ДНК-коды для обобщенного стебельного сходства

Волкова Юлия Андреевна

Аспирант

Московский государственный университет имени М.В. Ломоносова, Механико-математический факультет, Москва, Россия

E-mail: y.volkova@ymail.com

Рассматривается понятие аддитивного b-стебельного сходства между двумя ДНК-последовательностями длины $n \geq b \geq 2$, которое рассчитывается как максимальное число общих блоков длины b (b-стеблей), содержащих соседние символы в их общей подпоследовательности Хемминга [1].

Одинарная ДНК-цепь представляется как ориентированная последовательность с элементами из ДНК-алфавита $\boldsymbol{A} \triangleq \{A, C, G, T\}$. *Комплементарно сопряженная* (сопряженная по Ватсону-Крику) последовательность для $\boldsymbol{x} \triangleq (x_1 x_2 \ldots x_{n-1} x_n) \in \boldsymbol{A}^n = \{A, C, G, T\}^n$ имеет вид $\widetilde{\bar{\boldsymbol{x}}} \triangleq (\bar{x}_n \bar{x}_{n-1} \ldots \bar{x}_2 \bar{x}_1) \in \boldsymbol{A}^n = \{A, C, G, T\}^n$. Причем, $A = \bar{T}, C = \bar{G}$ и наоборот.

ДНК-код длины n для аддитивного b-стебельного сходства (коротко $(n, d)_b$) - это набор из последовательностей (кодовых слов), инвариантный относительно преобразования Ватсона-Крика, и если пара его слов не образует дуплекс Ватсона-Крика, то b-стебельное расстояние между этими кодовыми словами $\geq d$, где $d = 1, 2, \ldots, n-(b-1)$ есть фиксированный уровень.

$N_b(n, d)$, $d \in [n-(b-1)]$ есть максимальный объем такого кода для расстояния d.

Теорема 1 *Для любого $d \in [n-(b-1)]$, максимальный объем*

$$N_b(n,d) \leq \begin{cases} 4^{n-d+1} - 4^{\frac{n-D+1}{2}}, & \text{если } n\text{-четное}, d\text{-нечетное}, \\ 4^{n-d+1}, & \text{иначе.} \end{cases}$$

$(n, d)_b$-код есть код с максимально достижимым расстоянием(МДР) для аддитивного b-стебельного расстояния, если его объем N достигает этой границы. Для $b = 2$, МДР $(n, d)_2$-коды приведены в [2].

Основным результатом данной работы является

Теорема 2 *Пусть $b = 3$ и $1 \leq d \leq 6$. Тогда для любой длины $n \geq d+2$, существует МДР $(n, d)_3$-код, т.е. максимальный объем*

$$N_3(n,d) = \begin{cases} 4^{n-d+1} - 4^{\frac{n-d+1}{2}}, & \text{если } n\text{-четное}, d\text{-нечетное}, \\ 4^{n-d+1}, & \text{иначе.} \end{cases}$$

Если $d \geq 7$, то $N_3(n, d)$ не достигает максимального значения, начиная с некоторого n. Так, например, в случае $d = 7$ имеем $n = 1303$.