

КРУПНОМАСШТАБНЫЕ ПЕРИОДИЧНОСТИ В ПОСЛЕДОВАТЕЛЬНОСТЯХ НУКЛЕОТИДОВ ЛОКУСОВ ГЕНОВ РАННЕГО РАЗВИТИЯ *Drosophila*

© 2017 г. А.П. Лифанов, Г.И. Кравацкая, Н.Г. Есипова

Институт молекулярной биологии им. В.А. Энгельгардта РАН, 119991, Москва, ул. Вавилова, 32

E-mail: johnnie_me@list.ru

Поступила в редакцию 24.11.16 г.

Для генов раннего развития *Drosophila* с использованием усовершенствованного метода Фурье-анализа и применением вейвлет-преобразования установлено наличие крупномасштабных периодичностей в расположении участков с характерным локальным АТ-ГС-составом последовательности нуклеотидов локусов этих генов. Локализация периодических участков локусов позволила установить, что они располагаются в основном в модулях регуляции транскрипции (энхансерах). Наблюдаемые длины периодов в диапазонах 80–85, 165–180 и 330–400 нуклеотидов хорошо согласуются с длинами, характерными для нуклеосомного уровня организации хроматина. Длины периодов в диапазоне 600–750 нуклеотидов составляют примерно четыре нуклеосомных повтора и близки к минимальной длине энхансера, типичной для генома *Drosophila*.

Ключевые слова: локус, энхансер, модуль регуляции транскрипции, нуклеосома, последовательность нуклеотидов, нуклеотидный состав, периодичность.

Гены раннего развития эмбриона *Drosophila* – хорошо изученная система генов. Для нуклеотидных последовательностей их локусов известно положение не только кодирующего сегмента, но и большинства *цис*-регуляторных модулей (энхансеров) [1]. Для энхансеров этих генов характерно наличие периодичностей, связанных с нуклеосомным уровнем организации хроматина. Эти периодичности наиболее ярко проявляются в размещении консервативных блоков – участков повышенной консервативности, границы которых определены путем сравнения нуклеотидных последовательностей генов насекомых двух видов – *Drosophila melanogaster* и *Drosophila pseudoobscura* [2,3].

В настоящей работе мы показываем, что периодичности, характерные для размещения нуклеосом, могут быть выявлены прямыми методами исследования распределения участков характерного для нуклеосомной ДНК нуклеотидного состава по сегментам выбранных генов.

Для 12 локусов генов раннего развития *D. melanogaster* (*btd*, *ems*, *en*, *eve*, *ftz*, *gsb*, *hairy*, *kni*, *kr*, *sal*, *slp*, *tll*) нами был проведен перевод их нуклеотидных последовательностей в числовой формат путем замены нуклеотидов А и Т числовым значением +1, а С и G – 0 соответственно. Для полученных цифровых последовательностей рассчитаны амплитуды гармоник

преобразования Фурье, определена их статистическая значимость согласно [4] и отброшены статистически незначимые гармоники.

На рис. 1 приведена гистограмма, построенная по длинам периодов статистически значимых гармоник всех исследованных генов; рассмотрены периоды в диапазоне от 10 до 930 нуклеотидов (нт).

Для большинства изученных последовательностей наблюдается выраженное разделение на диапазон гармоник 10–400 нт (субнуклеосомный, нуклеосомный, динуклеосомный масштабы длин) и диапазон гармоник с длинными (600 нт и более) периодами.

Преобразование Фурье последовательностей нуклеотидов локусов исследуемых 12 генов позволяет выявить наличие периодичностей, но не позволяет установить местоположение участков, содержащих периодические структуры. Для локализации участков, содержащих крупномасштабные периодические последовательности нуклеотидов, необходимо в дополнение к преобразованию Фурье применить другие методы; в частности, может быть использована свертка с функциями, подчеркивающими периодичности в сигнале.

ДНК энхансеров большинства генов из геномов различных видов упакована в нуклеосомы [5]. Для нуклеосомной ДНК характерна

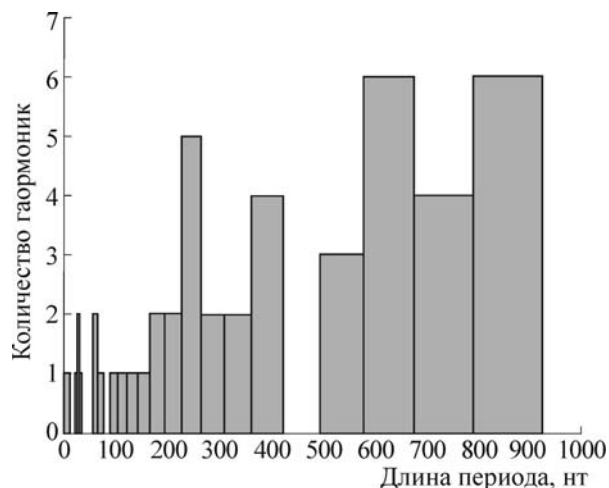


Рис. 1. Гистограмма длин периодов статистически значимых гармоник преобразования Фурье для оцифрованных последовательностей нуклеотидов локусов генов *Drosophila*.

модуляция нуклеотидного состава с периодом 10–11 нт, что близко к длине витка двойной спирали ДНК [6,7]. Для подавления периодичности 10–11 нт нами была произведена низкочастотная фильтрация оцифрованных последовательностей путем суммирования (свертки) в скользящем окне 50 нт, как и в работе [8].

Результат такой свертки для локуса гена *eve* приведен на рис. 2.

В кодирующем сегменте, расположенном в диапазоне 6000–7500 нт, наблюдается пониженное содержание нуклеотидов А и Т. Участки, обогащенные А и Т (положительные пики шириной 200–300 нт, расположенные с шагом в диапазоне 1500–2000 нт), являются местами крепления хроматина к ядерному скелету (scaffold/matrix attachment regions, SAR/MAR) [9]. Остальные участки, как правило, не содержат крупных периодичностей: на графике не наблюдается значительных крупномасштабных отклонений от уровня сигнала, усредненного по рассматриваемому участку.

После низкочастотной обработки выполним свертку сигналов с функцией, являющейся второй производной от кривой нормального распределения (вейвлет «мексиканская шляпа»): $f(z) = c \cdot \exp(-z^2/2) \cdot (1 - z^2)$, $c = 2/(\sqrt{3} \cdot \pi^{1/4})$, $z = x/a$, x – номер позиции в локусе. Данная функция, содержащая два симметричных минимума по бокам от центрального максимума, хорошо подходит для подчеркивания локальных периодичностей в пределах одного–полутора периодов [10]. Для выделения различных периодов будем изменять переменную a , определяющую длину выделяемого периода. Результат такой

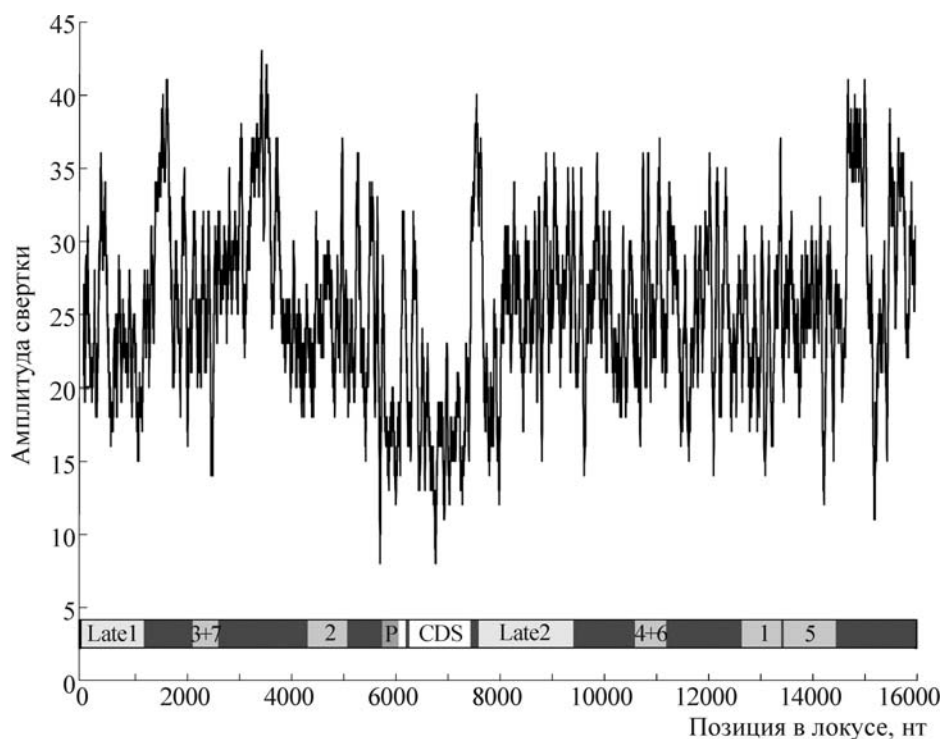


Рис. 2. Результат свертки функции оцифрованной нуклеотидной последовательности для локуса гена *eve* с окном 50 нт. Внизу: расположение функциональных элементов локуса: белые полосы – кодирующий сегмент (CDS), серые полосы – энхансеры.

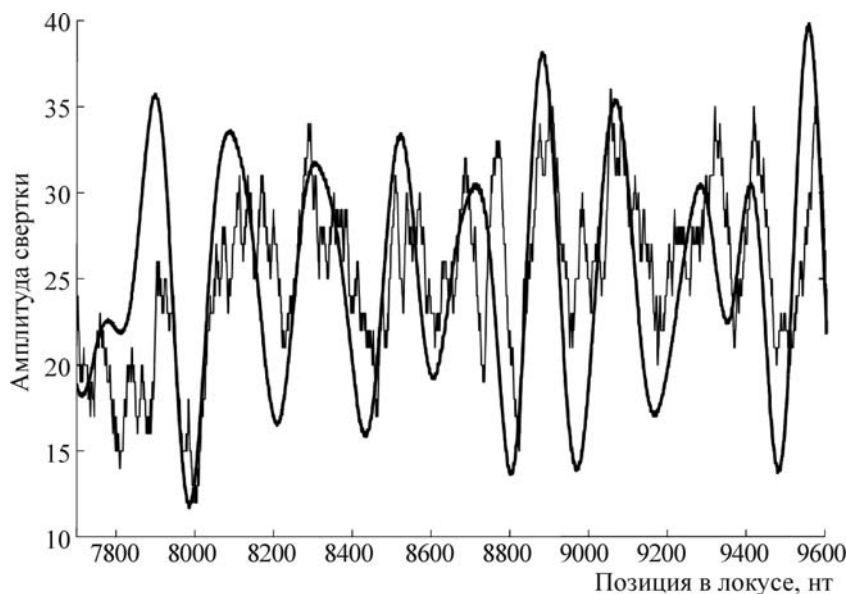


Рис. 3. Результат фильтрации для энхансера, расположенного на отрезке 7700–9600 нт в локусе гена *eve*. Тонкая линия: оцифрованная нуклеотидная последовательность, усредненная в окне 50 нт. Толстая линия: результат свертки усредненной оцифрованной нуклеотидной последовательности локуса с вейвлетом «мексиканская шляпа»; выбранный масштаб выделяет период 175 нт.

фильтрации для одного из энхансеров локуса гена *eve* представлен на рис. 3. Как очевидно, имеет место выраженная периодическая модуляция с периодом 175 нт.

Обобщая результаты фильтрации оцифрованной последовательности нуклеотидов (эквивалентной функции локального АТ-ГС-состава) локусов выбранных генов раннего развития, подчеркивающие периодичность в диапазоне 80–350 нт, можно сделать следующий вывод: периодические участки наблюдаются в основном в энхансерах исследуемых генов; длины периодов приведены в таблице.

Для выбранных генов нуклеосомы в энхансерах располагаются с плотностью, близкой к максимальной; таким образом, участки ДНК, входящие в состав нескольких соседствующих нуклеосом, образуют картину периодического характера, с периодом порядка минимальной длины нуклеосомного повтора, равного согласно работам [2,3] 165 нт. Наблюдаемые нами в энхансерах периоды хорошо согласуются с картиной периодичностей нуклеосомного уровня организации хроматина: длиной витка суперспирали ДНК нуклеосомы (81–84 нт), минимальной длиной нуклеосомного повтора (165 нт) и кратными им величинами [7].

Диапазон более длинных периодов 600–750 нт составляет примерно четыре нуклеосомных повтора и близок к характерной для генов дрозофилы минимальной длине *cis*-регуляторных модулей (энхансеров) [1].

Таким образом, для группы генов раннего развития *Drosophila* удалось продемонстрировать наличие крупномасштабных периодичностей в модулях регуляции транскрипции этих генов, согласующихся с картиной периодичностей, свойственной нуклеосомному уровню организации хроматина.

Взаимодействие гистонов с ДНК обеспечивает конформационную устойчивость участков ДНК, образующих нуклеосомы, за счет компенсации сильных электростатических взаимодействий, ведущих, согласно теореме Ирншоу,

Длины периодов в энхансерах локусов ранних генов *Drosophila*

Локус	Период в энхансерах
<i>btd</i>	170
<i>ems</i>	165
<i>eng</i>	185
<i>eve</i>	175
<i>ftz</i>	175
<i>gsb</i>	165
<i>hry</i>	170
<i>kni</i>	85 = 170/2
<i>kru</i>	180
<i>sal</i>	165
<i>slp</i>	330 = 165 × 2
<i>tll</i>	350 = 175 × 2

к неустойчивости. Обнаружение крупномасштабных периодичностей означает распространение упорядоченности на более высокие уровни в целых регуляторных модулях – модулях регуляции транскрипции (энхансерах). В системах, содержащих конформационно-стабильные элементы (α -спирали в гистонах и конформационно-устойчивая ДНК), возможно узнавание по механизму теории специфических селективных дальнедействующих взаимодействий [11, 12]. Это отражает функциональную связь процессов компактизации ДНК и процессов репликации генетической информации.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проекты №№ 15-04-99605, 16-54-00219-Бел).

СПИСОК ЛИТЕРАТУРЫ

1. A. P. Lifanov, V. J. Makeev, A. G. Nazina, and D. A. Papatsenko, *Genome Res.* **4**, 579 (2003).
2. А. П. Лифанов, В. Ю. Макеев и Н. Г. Есипова, *Биофизика* **60** (1), 15 (2015).
3. А. П. Лифанов, В. Ю. Макеев и Н. Г. Есипова, *Биофизика* **61** (1), 50 (2016).
4. V. R. Chechetkin and A. Y. Turygin, *J. Theor. Biol.* **175**, 477 (1995).
5. R. C. Hardison and J. Taylor, *Nat. Rev. Genet.* **13** (7), 469 (2012).
6. B. Salih, V. Tripathi and E. N. Trifonov, *J. Biomol. Struct. Dyn.* **33** (1), 1 (2015).
7. L. Marino-Ramirez, M. G. Kann, B. A. Shoemaker, and D. Landsman, *Expert Rev. Proteomics* **2**, 719 (2005).
8. Р. В. Хемминг, *Цифровые фильтры* (Сов. радио, М., 1980).
9. R. U. Pathak, A. Srinivasan, and R. K. Mishra, *BMC Genomics* **15** (1), 1022 (2014).
10. I. Daubechies, *Ten lectures on wavelets* (CBMS-NSF Regional Conf. Series in Applied Math., V. 61; Philadelphia, PA, 1992).
11. V. A. Namiot, A. V. Batyanovskii, I. V. Filatov, et al., *Phys. Lett. A* **375** (32), 2911 (2011).
12. А. В. Батяновский, И. Д. Волотовский, В. А. Намиот, и др., *Биофизика* **60** (3), 437 (2015).

Large-Scale Periodicities in the Nucleotide Sequences of Early Development *Drosophila* Genes Loci

A.P. Lifanov, G.I. Kravatskaya, and N.G. Esipova

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova 32, Moscow, 119991 Russia

For the early development genes of *Drosophila* using the improved Fourier analysis method and the wavelet transform application, large-scale periodicity in the location of sites with the characteristic local AT-GC composition of the nucleotide sequence of the loci of these genes was established. The localization of periodic locus sections implies that they are located mainly in the transcriptional regulation modules (enhancers). The observed lengths of the periods in the ranges of 80–85, 165–180 and 330–400 nucleotides are in good agreement with the lengths typical of the nucleosome level of chromatin organization. Harmonic periods in the range of 600–750 nucleotides are approximately four nucleosome repeats and are close to the minimal length of the enhancer observed in *Drosophila* genome.

Keywords: locus, enhancer, transcription regulatory module, nucleosome, nucleotide sequence, nucleotide content, periodicity

Сдано в набор 14.08.2017	Подписано к печати 16.10.2017	Дата выхода в свет 24.11.2017	Формат 60x88 $\frac{1}{8}$
Цифровая печать	Усл. печ. л. 26,0	Усл. кр.-отт. 3,2 тыс.	Уч.-изд. л. 26,0
	Тираж 113 экз.	Зак. 1622	Цена свободная

Учредители:
Российская академия наук,
Институт биофизики клетки РАН

Издатель: ФГУП «Издательство «Наука», 117997, Москва, Профсоюзная ул., 90

Отпечатано в ФГУП «Издательство «Наука» (Типография «Наука»), 121099, Москва, Шубинский пер., 6