

Поиск генов микроРНК в участках генома, содержащих два очень поздних гена вируса ядерного полиэдроза *Bombyx mori*

Т. В. Ширина, А. А. Висловух, М. Т. Бобровская, Э. А. Козлов

Институт молекулярной биологии и генетики НАН Украины
Ул. Академика Заболотного, 150, Киев, Украина, 03680

e.a.kozlov@imbg.org.ua

Цель. Вирусы ядерного полиэдроза (ВЯП) *B. mori* кодируют два очень поздних гена – *ph* и *p10*. Интерес к поиску генов *miR* в этих участках генома обусловлен тем, что полиэдры, образующиеся на очень поздней стадии инфекции, включают в себя не только вирионы, но и малую РНК длиной 50–60 нуклеотидов. Цель настоящего сообщения состояла в поиске *miRs* в альтернативных транскриптах, синтезируемых не только с ТААГ-промоторного элемента, но и с ТАТА-промоторных элементов, расположенных в участках генома ВЯП *B. mori*, включающих гены *ph* и *p10*. **Методы.** Поиск *miRs* осуществляли с помощью биоинформатических программ предсказания *miR*: *MiPred*, *miRNA SVM*, *Microprocessor SVM* и *RNAfold*. **Результаты.** Предсказано, что участок, содержащий ген *ph*, может кодировать две *miRs* (*bmoNPV-miR-1ph*, *bmoNPV-miR-2ph*) и один потенциальный (С) предшественник *miR* – *bmoNPV-pre-miR-1Cph*, не являющийся субстратом для фермента *Dicer*. Участок, содержащий ген *p10*, может кодировать одну предсказанную *miR* – *bmoNPV-miR-3p10*. **Выводы.** Обсуждается возможность регуляции экспрессии предсказанными *miRs* генов *orf 1629* и *p74*, расположенных в тех же участках комплементарной цепи.

Ключевые слова: вирус ядерного полиэдроза, *Bombyx mori*, микроРНК, биоинформатический подход, предсказание.

Введение. МикроРНК (*miRs*) являются представителями одного из трех наиболее распространенных классов малых некодирующих РНК длиной 20–30 нуклеотидов (*miRNAs*, *siRNAs*, *piRNAs*), инициирующих РНК-интерференцию. *miRs* выполняют роль биорегуляторов экспрессии генов в клетках эукариотов. Биогенез, функционирование, биохимические и биоинформатические подходы к исследованию *miRs*, их участие в регуляции различных клеточных процессов, а также связь с патологиями описаны нами в обзоре [1]. О *siRNA*, *piRNA* и дру-

гих малых некодирующих РНК можно прочитать в публикации [2]. Кроме эукариотов, *miRs* обнаружены у вирусов, в основном, у крупных ДНК-содержащих вирусов [3]. Среди РНК-содержащих вирусов *miRs* выявлены у вируса иммунодефицита человека [4, 5]. Мало что известно о взаимоотношениях вирус–клетка на уровне *miRs*. По этому вопросу опубликовано несколько экспериментальных статей и обзоров [6–8].

Бакуловирусы относят к классу крупных ДНК-содержащих вирусов. Вирусы ядерного полиэдроза (ВЯП) представляют самостоятельную серологическую группу бакуловирусов, вирионы которых

Institute of Molecular Biology and Genetics NAS of Ukraine, 2009

на очень поздних стадиях инфекции встраиваются в тела включения – полиэдры. Белок, образующий полиэдры (полиэдрин), является продуктом одного из двух очень поздних генов. Вторым геном *p10* кодирует белок *p10*. Экспрессия обоих этих генов инициируется поздними промоторными элементами – A/G/T/ TAAG [9].

Интерес к поиску предшественников (precursor) miR (pre-miR) в мРНК, транскрибируемых с двух очень поздних промоторов генома ВЯП *B. mori*, обусловлен тем, что полиэдры ВЯП *B. mori*, образующиеся на очень поздней стадии инфекции, включают в себя не только вирионы, но и малую РНК длиной 50–60 нуклеотидов [10]. На основании размеров мы предположили, что она представляет собой pre-miR (размеры pre-miRs колеблются от 50 до 100 нуклеотидов). Наиболее вероятно, что включенная в полиэдры предполагаемая pre-miR может процессироваться из очень поздних транскриптов и захватываться полиэдрином в процессе формирования полиэдров. К таким очень поздним транскриптам относят мРНК полиэдрина и *p10* [9]. Мы не исключаем также, что полиэдры могут включать pre-miRs, процессируемые из других поздних транскриптов, или хозяйский pre-miR-let7. Всплеск синтеза miR-let7 на стадии превращения гусеницы в куколку отмечен авторами [11]. Это именно та стадия развития насекомого, на которой мы выделяли полиэдры для изучения (коконы, содержащие погибшие гусеницы). Дальнейшие биохимические исследования РНК из полиэдров помогут прояснить вопрос, какая именно малая РНК включается в полиэдры

В настоящем сообщении опубликованы результаты биоинформатического подхода к поиску pre-miRs и miRs не только в транскриптах, синтезируемых с ТААГ-промоторного элемента для двух очень поздних белков, но и в альтернативных транскриптах (alts), синтезируемых с предполагаемых ТАТА-промоторных элементов, расположенных в участках генома ВЯП *B. mori*, включающих гены *ph* и *p10*.

Материалы и методы. Нуклеотидная последовательность генома ВЯП *B. mori* взята из ICTVdB Management (2006). 00.006.0.01. Nucleopolyhedrovirus (ICTVdB – The Universal Virus Database, version

4. Columbia University, New York, USA (<http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/00.006.0.01.htm>).

Из существующих программ, предназначенных для предсказания микроРНК, мы выбрали те, в алгоритм которых не заложен критерий консервативности, поскольку вирусные микроРНК в отличие от микроРНК эукариотов неконсервативны. Вторичную структуру alts (гипотетические (hypothetical) первичные (primary) транскрипты – h-pri-miR) исследовали с помощью программы RNAfold (<http://rna.tdi.univie.ac.at/cgi-bin/RNAfold.cgi>) [12]. Поиск в альтернативных транскриптах шпичечно-петлевых структур (sls, stem-loop structure) длиной от 48 до 150 нуклеотидов, являющихся субстратами для ферментов Drosha и Dicer, осуществляли с помощью программы предсказания и процессирования pri-miR (<https://demol.interagon.com/miRNA/>). Предсказанными pre-miR и зрелыми miR считали субстраты с оценкой (score) выше точки пересечения кривых чувствительности (Se) и специфичности (Sp) – $>-0,55$ [13]. Шпильки, процессируемые Drosha, но не процессируемые Dicer, считали потенциальными (candidate, C).

Нуклеотидные последовательности шпичечных структур, обнаруженных в альтернативных транскриптах, исследовали также с помощью программы RNAfold. Процессируемыми шпильками считали sls со значением свободной энергии свертывания $-23,0$ ккал/моль [12] или (в пересчете на кДж) $-96,6$ кДж/моль. Тонг и др. [14] в своих поисках miR в геноме *B. mori* с помощью программы RNAfold приняли в качестве «фильтра» величину свободной энергии «больше 105 кДж/моль». Мы приняли значение -100 кДж/моль. Поиск реальных и псевдошпильек pre-miR осуществляли с помощью программы miPred (<http://www.bioinf.seu.edu.cn/miRNA/index.html>) [15].

Поиск зрелых miR в предсказанных pre-miR проводили, используя программу miRscan (<http://genes.mit.edu/miRscan/>) [16]. В качестве первой и второй последовательностей в miRscan вносили нуклеотидную последовательность исследуемой pre-miR.

Результаты и обсуждение. У бакуловирусов поздняя транскрипция инициируется ТААГ-промоторным элементом и терминируется полиТ-по-

следовательностью [9]. Известно, что у ВЯП *Auto-grapha californica* синтезируются три полиэдриновых транскрипта длиной 1,16; 3,4 и 4,9 тыс. п. н. [17] и два р10 транскрипта длиной 0,75 и 2,5 тыс. п. н. [18]. Аналогичных данных для ВЯП *B. mori* нет. Так как ВЯП *B. mori* является генотипическим вариантом ВЯП *A. californica* можно предположить, что такая же ситуация характерна и для ВЯП *B. mori*. На этом основании мы определили границы участков генома ВЯП *B. mori* для поиска miRs. Для исследований выбраны участки, содержащие только два предполагаемых полиэдриновых транскрипта (1,16 и 3,4 тыс. п. н.) и оба р10 транскрипта. Такой выбор обусловлен тем, что транскрипты 1,16 и 3,4 тыс. п. н. перекрывают ген *orf 1629*, а транскрипт 2,5 тыс. п. н. р10 – ген *p74*, расположенные на комплементарной цепи. Полиэдриновый транскрипт 4,9 тыс. п. н. не исследовали, так как он выходит за пределы выбранной области. Расположение обозначенного полиэдринового участка в геноме ВЯП *B. mori* – 128298–3404, а участка р10 – 108411–110961. Если за точку отсчета принять А в кодоне AUG, то это участки –116–3404 (в дальнейшем ph) и –86–2565 (в дальнейшем р10) соответственно. Транскриптам 1,16 и 3,4 тыс. п. н. отвечают транскрипты –51–1129ph и –51–3404ph; двум транскриптам р10 – транскрипты –71–630р10 и –71–2565р10.

Вторичная структура транскрипта –51–1129ph содержит, как показано нами, две шпилечно-петлевые структуры, из которых одна (sls1ph) процессируется в зрелую miR, а вторая (sls2ph) не проходит фильтры примененных нами программ. Вторичная структура второго транскрипта –51–3404ph содержит 12 шпилек, в числе которых находится sls1ph. Из оставшихся шпилечно-петлевых структур три не проходят фильтры программ, а восемь процессируются только в pre-miRs. Рисунки вторичных структур транскриптов из-за их громоздкости не приведены, а характеристики sls будут рассмотрены ниже в связи с обсуждением альтернативных транскриптов. В полиэдрин, по-видимому, транслируется –51–1129ph [17].

Таким образом, транскрипт –51–3404ph, по нашим предсказаниям, может выступать в качестве h-pri-miR. Аналогичная ситуация наблюдается с

двумя транскриптами участка –86–2565р10. По нашим данным, вторичная структура меньшего транскрипта (0,75 тыс. п. н.) содержит единственную sls1, процессируемую в miR. Более крупный транскрипт (2,5 тыс. п. н.) содержит, кроме sls1, еще шесть шпилек, которые процессируются в pre-miR, но не проходят фильтры остальных программ. Как в случае транскрипта –51–3404ph, транскрипт –71–2565р10 может выступать в качестве h-pri-miR.

Логика подхода к поиску miR в альтернативных транскриптах заключается в следующем. Все существующие программы предсказания потенциальных pre-miR основаны на поиске отдельных шпилечно-петлевых структур, отвечающих определенным требованиям. Реально же шпильки pre-miR процессируются из первичных транскриптов – pri-miR. Сложность поиска h-pri-miR среди alts состоит в том, что промоторы, с которых инициируется транскрипция pri-miR, точно не определены, хотя известно, что pri-miRs транскрибируются в основном РНК-полимеразой II в большинстве случаев с ТАТА-промоторов [19]. Транскрипция pri-miRs может происходить и с других последовательностей [20]. Мы решили начать поиск miRs с предсказания h-pri-miRs среди различных alts. Границы альтернативных транскриптов определены нами от предполагаемых промоторов ТАТА до полиТ (не менее четырех Т) последовательностей.

В участке –116–3404ph насчитывается шесть предполагаемых промоторов и 33 полиТ-терминирующие последовательности, а в участке –86–2565р10 – девять промоторов и 19 полиТ-последовательностей. Исходя из этих данных исследовали 148 alts-ph и 114 alts-p10. 148 alts-ph содержат 19 уникальных sls-ph. Далее уникальные альтернативные транскрипты отбирали по принципу: минимальный по величине alt должен содержать, кроме искомой шпильки, минимальное количество других шпилек. Из 148 alts-ph только 11 уникальных транскриптов содержали все 19 sls-ph, а 16 уникальных из 114 alts-p10 включали 21 sls-p10. Все результаты исследований характеристик sls приведены в табл. 1 и 2.

Как видно из табл. 1, sls3-ph, sls6-ph, sls7-ph, sls10-ph, sls17-ph, sls19-ph процессируются из уни-

Таблица 1

Характеристики шпильчно-петлевых структур (*sls*) в альтернативных транскриптах (*alt*), синтезируемых с участка 128298–3404 генома ВЯП *B. mori*, кодирующего мРНК полиэдрина

Sls	Локализация в alt	Старт	Длина, нуклеотиды	Drosha	Real	-E, кДж/моль	Dicer
1	-7-683	248	75	-0,421	+	114	-0,52
2	-7-683	538	50	-0,504	-	64	-0,504
3	-51-1272	634	68	0,024	+	87	-0,661
4	744-970	775	67	-0,219	-	65	<-1,0
5	744-970	840	75	-0,239	-	73	<-1,0
6	-51-1003	941	60	-0,33	-	61	<-1,0
7	744-1331	1018	60	-0,88	-	36	<-1,0
8	744-2139	1046	56	-0,204	-	43	<-1,0
9	744-1672	1092	73	-0,218	-	70	-0,992
10	744-1407	1284	57	-0,292	-	58	<-1,0
11	744-1672	1305	74	0,392	+	106	-0,425
12	744-2139	1540	69	-0,158	+	140	<-1,0
13	744-2446	1607	65	нет	+	96	Нет
14	744-2446	1842	59	-0,729	+	80	<-1,0
15	744-2446	1905	65	-0,711	Pseudo	136	<-1,0
16	744-2446	2080	56	-0,101	Pseudo	71	-0,967
17	1851-2339	2169	50	-0,22	-	63	<-1,0
18	744-2446	2306	90	Нет	Pseudo	85	Нет
19	1851-3107	2724	56	Нет	-	73	Нет

Примечание. Здесь и в табл. 2 в колонке 1 приведены номера *sls* в порядке удаления их от точки отсчета. В колонках 2 и 3 указана локализация *alt*, содержащего соответствующую *sls*, относительно точки отсчета и положение стартового нуклеотида *sls* соответственно. В колонках 5 и 8 представлены оценки *sls*-субстратов для ферментов Drosha и Dicer соответственно согласно программе предсказания и процессирования *pri-miR* [13]; «нет» обозначает отсутствие в соответствующих *sls* процессируемых центров. В колонке 6 «+» – «реальная» шпилька, «-» – шпилька «не может быть реальной» согласно программе miPred [15]. В колонке 7 приведены величины свободной энергии свертывания *sls* согласно программе RNAfold [12]. Характеристики *sls*, прошедших фильтры соответствующих программ, выделены жирным шрифтом.

кальных *alt-ph*, в то время как остальные *sls-ph* – из двух и более *alt-ph*. *Sls1-p10*, *sls5-p10*, *sls7-p10*, *sls9-p10*, *sls14-p10*, *sls16-p10*, *sls18-p10-sls21-p10* также процессируются из уникальных *alt-p10* (табл. 2). Все *alts*, содержащие единственную процессируемую *sls*, можно рассматривать как кандидаты в *h-pri-miRs*.

Из данных табл. 1, колонка 5, следует, что 13 *sls-ph* (*sls1-sls6*, *sls8-sls12*, *sls16*, *sls17*) являются

субстратами для фермента Drosha, а из 21 *sls-p10* (табл. 2, колонка 5) – 18 (*sls1-sls4*, *sls6-sls13*, *sls15-sls17*, *sls19-sls21*) служат субстратами Drosha. Из 32 отобранных *sls* пять *sls-ph* (*sls1*, *sls3*, *sls11*, *sls12*, *sls16*) и пять *sls-p10* (*sls2*, *sls7*, *sls8*, *sls17*, *sls19*) проходят фильтры «real» и «pseudo» (табл. 1 и 2, колонка 6). Из 10 отобранных реальных и псевдошпилек реальные *sls1-ph*, *sls11-ph*, *sls12-ph* и *sls2-p10* проходят фильтр RNAfold (табл. 1 и 2, ко-

Таблица 2

Характеристики шпилечно-петлевых структур (sls) в альтернативных транскриптах (alt), синтезируемых с участка 108281–110842 генома ВЯП В. тори, кодирующего мРНК p10

Sls	Локализация в alt	Старт	Длина, нуклеотиды	Drosha	Real	-E, кДж/моль	Dicer
1	-86-144	-83	64	-0,316	-	40	<-1,0
2	-30-301	66	65	0,637	+	112	0,009
3	-30-301	214	54	0,054	-	53	<-1,0
4	-30-491	319	76	-0,288	-	52	<-1,0
5	323-1556	326	89	-0,616	-	65	<-1,0
6	-30-491	396	70	0,326	-	70	<-1,0
7	387-611	423	95	-0,35	Pseudo	77	<-1,0
8	323-1405	686	81	-0,426	+	88	<-1,0
9	550-980	742	57	0,093	-	51	-0,529
10	-30-980	808	48	-0,521	-	77	<-1,0
11	323-1405	817	62	-0,142	-	85	0,750
12	-30-980	926	51	-0,477	-	83	<-1,0
13	550-1307	960	47	-0,178	-	83	-0,943
14	-71-1112	989	65	Нет	Pseudo	91	Нет
15	550-1307	1084	50	0,141	-	60	-0,067
16	1203-1307	1203	105	-0,008	-	131	-0,902
17	323-1405	1222	60	-0,008	Pseudo	88	-0,902
18	1203-1556	1472	81	-0,793	Pseudo	102	<-1,0
19	1203-1870	1614	58	-0,108	+	84	0,618
20	411-2565	1952	52	-0,024	-	99	<-1,0
21	2233-2565	2364	83	-0,188	-	48	<-1,0

Примечание. См. табл. 1.

лонка 7). Из четырех отобранных «real» и «pseudo» шпилек три (sls1-ph, sls11-ph, sls2-p10) являются субстратами для Dicer. Поэтому процессируемые из них шпильки мы приняли за предсказанные pre-miR и обозначили их как bmo-pre-miR-1ph, bmo-pre-miR-2ph и bmo-pre-miR-3p10 соответственно, а содержащие их alts посчитали предсказанными h-pri-miRs (их положение см. в табл. 1 и 2). Sls12-ph не проходит фильтр Dicer. Авторы [13] называют шпильки, не являющиеся субстратами для Dicer, потенциальными (candidate, C) miRs. Поэтому sls, проходящие все фильтры, кроме Dicer, мы

обозначили как предсказанные потенциальные предшественники miRs – pre-miR-1Cph, а содержащие их alts (см. табл. 1 и 2) – как соответствующие h-pri-miR.

На рис. 1 представлены вторичные структуры трех h-pri-miRs. Как видно, все три h-pri-miRs содержат по две sls. Однако sls2-ph, хотя и является субстратом для Drosha и Dicer, но она не реальна и не проходит фильтр свободной энергии свертывания (табл. 1). А sls9-ph и sls3-p10 служат субстратами для Drosha, но не проходят фильтры остальных программ. Поэтому представленные на рис. 1

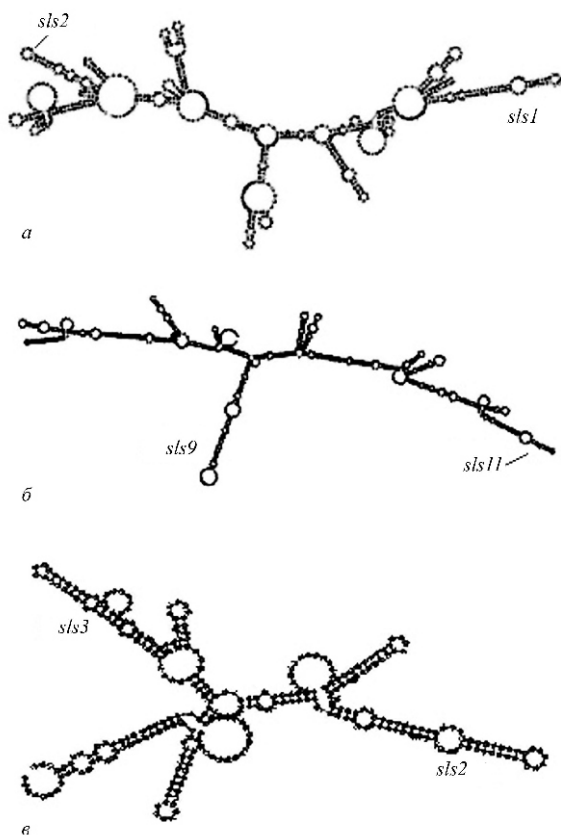


Рис. 1. Вторичные структуры минимальных альтернативных транскриптов, синтезируемых от предполагаемого ТАТА-промотора до полиТ-последовательности, принятых за гипотетические pri-miRs, содержащие минимальное число шпильчнопетлевых структур (*sls*): *a* – alt-ph –7–763, принятый за h-pri-miR-1ph (точкой отсчета здесь и далее считали AUG-кодоны мРНК полиэдрина или p10); до зрелой miR процессируется *sls1*; *б* – alt-ph 744–1672, принятый за h-pri-miR-2ph (до зрелой miR процессируется *sls11*); *в* – alt-p10 –30–301, принятый за h-pri-miR-3p10 (до зрелой miR процессируется только *sls2*)

h-pri-miRs мы обозначили как h-pri-miR-1ph, h-pri-miR-2ph и h-pri-miR-3p10. На рис. 2 представлена вторичная структура h-pri-miR-1Cph. *Sls11* процессируется до зрелой miR-2ph, а *sls12* – до потенциальной pre-miR-1Cph. Рис. 3 иллюстрирует три *sls*, процессируемые до зрелых miRs, и одну *sls*, процессируемую до потенциальной pre-miR-C.

Авторы работы [21] с помощью разработанной ими программы предсказания вирусных miRs (Vir-Mirdb) выявили 11 pre-miR в плюс-цепи генома ВЯП *B. mori*, из которых вырезаются 22 miRs (по одной miR с каждого плеча pre-miR). С этими данными трудно согласиться, поскольку зрелая miR, как правило, вырезается из одного 5'-плеча. Кроме

того, все предсказанные авторами [21] miRs содержат по 26 нуклеотидов, в то время как miRscan предположительно имеет длину miRs, равную 21 нуклеотиду, что более соответствует величине miRs *in vivo*. Нуклеотидную последовательность генома эти авторы брали из того же источника, что и мы, и, тем не менее, они не обнаружили предсказанных нами pre-miR-1ph, pre-miR-2ph и pre-miR-3p10. Однако они нашли шпильку, в которую вписывается предсказанная нами pre-miR-1Cph. Но в отличие от их данных pre-miR-1Cph, как показано выше (табл. 1, *sls12*), не процессируется ферментом Dicer.

Интересно отметить, что в то время как miR-1ph, miR-2ph и miR-3p10 являются единственными в предсказанных нами h-pri-miRs, они также обнаруживаются во всех исследованных нами алтс, включающих участки их локализации. Можно предположить, что и в других синтезируемых в клетке неизвестных реальных альтернативных транскриптах (не только от промоторов ТАТА и ТААГ до полиТ-последовательности) будут присутствовать и, значит, процессироваться предсказанные зрелые miRs – miR-1ph, miR-2ph и miR-3p10. Предсказанные нами miRs-ph полностью комплементарны мРНК *orf 1629*, а miR-3p10 – мРНК *p74*. Следовательно, если эти miRs существуют, они должны функционировать подобно si-RNA. В этом случае мРНК будут расщепляться в участках, комплементарных предсказанным miRs. Так как miR-1ph и miR-3p10 комплементарны 3'-UTR мРНК *orf 1629* и *p74* соответственно, можно предположить участие этих miRs в регуляции экспрессии генов *orf 1629* и *p74*.

Возможно, такая ситуация наблюдается у ВЯП *A. californica*. Как показано в работе [17], на участке генома ВЯП *A. californica*, включающем ген полиэдрин, с комплементарной цепи синтезируется транскрипт длиной 3,2 тыс. п. н., содержащий две рамки считывания (*orf 1629* и *orf 603*) и перекрывающий ген полиэдрин. Этот транскрипт начинает синтезироваться раньше мРНК полиэдрин и с началом появления полиэдрин исчезает, но наблюдаются его фрагменты.

В 1990 г. авторы объяснили это явление тремя причинами: 1) разрушением промоторных комплексов с 3'-конца полиэдринового гена РНК-поли-

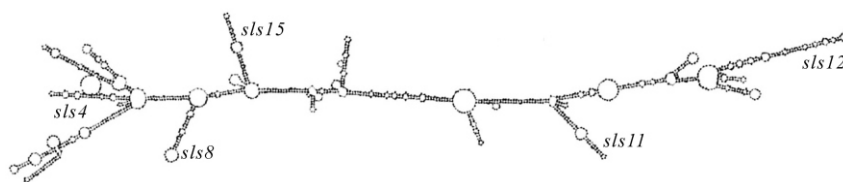


Рис. 2. Вторичная структура alt-ph 744–2139, принятого за h-pri-miR-1Cph (до pre-miR-1Cph процессируется только sls12)

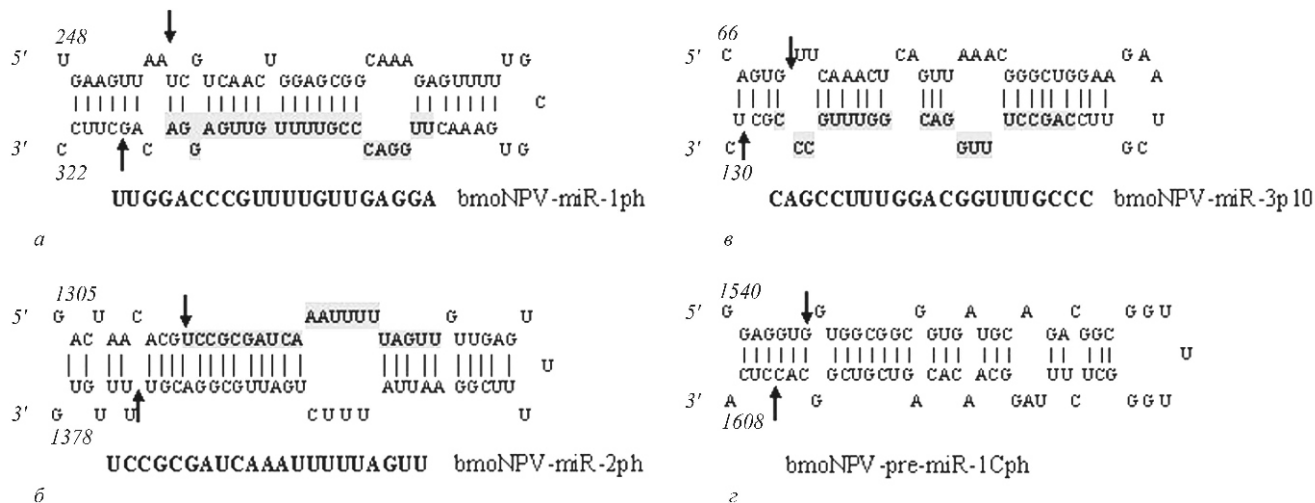


Рис. 3. Вторичные структуры sls1-ph (а), sls11-ph (б), sls2-p10 (в), sls12-ph (г) и предсказанные miRs. Стрелками указаны центры процессирования ферментом Droscha до шпилек pre-miR-1ph, pre-miR-2ph, pre-miR-3p10, pre-miR-1Cph. Присутствующие в pre-miRs зрелые miRs оттенены

меразой, транскрибирующей полиэдрин; 2) образованием двухцепочечной РНК с полиэдриновой мРНК; 3) отрицательной регуляцией промотора *orf 1629* полиэдрином. В то время они не могли предположить участия микроРНК в этом процессе, так как микроРНК были открыты только в начале этого века. Мы предполагаем участие miR, закодированной в гене полиэдрина, в регуляции синтеза транскрипта, содержащего *orf 1629* и *orf 603*. В этом случае будет происходить расщепление транскрипта 3,2 тыс. п. н. В дальнейшем мы планируем исследования по выявлению miRs в участке генома ВЯП *A. californica*, содержащем гены полиэдрина *orf 1629* и *orf 603*.

Какие из четырех обнаруженных нами предсказанных pre-miRs включаются в полиэдры, покажут биохимические исследования РНК из полиэдров.

T. V. Shirina, A. A. Vislovukh, M. T. Bobrovskaja, E. A. Kozlov

The search for microRNA genes in regions of two very late genes of *Bombyx mori* nuclear polyhedrosis virus

Summary

Aim. *B. mori* nuclear polyhedrosis virus (NPV) codes two very late genes – polyhedrin (*ph*) and *p10*. Search for miRs genes in these

regions is of interest because the polyhedra, formed at the very late stage of the virus development, include small RNA of 50–60 nt. The present work was aimed at search for potential precursors of miR transcribed from the late promoter element RTAAG and the TATA promoter elements located in the *ph* and *p10* genes regions. **Methods.** The search was performed using the bioinformatic programs for miR prediction: MiPred, miRNA SVM, Micropocessor SVM, and RNAfold. **Results.** It has been predicted that the region of *ph* gene encodes two predicted miRs (*bmoNPV-miR-1ph*, *bmoNPV-miR-2ph*) and one predicted potential (C) precursor *bmoNPV-pre-miR-1Cph*, which is not a Dicer substrate. The region containing *p10* gene encodes one predicted miR – *bmoNPV-miR-3p10*. **Conclusions.** A possibility of regulation of the genes *orf 1629* and *p74* expression by the predicted miRs, located in the same regions of a complementary chain, is assumed.

Keywords: nuclear polyhedrosis virus, *Bombyx mori*, microRNA, bioinformatic method, prediction.

T. В. Ширина, А. А. Висловух, М. Т. Бобровська, Е. А. Козлов

Пошук генів мікроРНК у ділянках геному, які містять два дуже пізні гени вірусу ядерного поліедрозу *Bombyx mori*

Резюме

Мета. Віруси ядерного поліедрозу (ВЯП) *B. mori* кодують два дуже пізні гени – *ph* і *p10*. Інтерес до пошуку miR у цих ділянках геному обумовлений тим, що поліедри, які утворюються на дуже пізній стадії інфекції, містять у собі не тільки віріони, але й малу РНК довжиною 50–60 нуклеотидів. Мета даного повідомлення полягала в пошуку miRs в альтернативних транскриптах, синтезованих не лише з ТААГ-промоторного елемента, а й з ТАТА-промоторних елементів, розташованих у

ділянках геному ВЯП *B. mori*, що включають гени *ph* і *p10*. **Методи.** Пошук *miR*s здійснювали за допомогою біоінформатичних програм передбачення *miR*: MiPred, *miRNA SVM*, *Microprocessor SVM* і *mFOLD*. **Результати.** Передбачено, що ділянка, в якій локалізується ген *ph*, може кодувати дві *miR*s (втоNPV-*miR-1ph*, втоNPV-*miR-2ph*) та потенційний попередник *miR* – втоNPV-*pre-miR-1Cph*, що не є субстратом для ферменту Dicer. Ділянка, у якій розміщений ген *p10*, може кодувати одну *miR* – втоNPV-*miR-3p10*. **Висновки.** Обговорюється можливість регуляції експресії передбаченими *miR*s генів *orf1629* і *p74*, розташованих на тих же ділянках комплементарного ланцюга.

Ключові слова: вірус ядерного поліедрозу, *Bombyx mori*, мікроРНК, біоінформатичний підхід, передбачення.

СПИСОК ЛІТЕРАТУРИ

1. Shirina T. V., Bobrovskaya M. T., Kozlov E. A. MicroRNA: from fundamental research to their application // *Biopolymers and Cell.*–2007.–**23**, N 6.–P. 467–482.
2. Makarova J. A., Kramerov D. A. Noncoding RNAs // *Biochimica.*–2007.–**72**, N 11.–C. 1427–1448.
3. Nelsen J. A. Small RNA and large DNA viruses // *N. Engl. J. Med.*–2007.–**357**, N 26.–P. 2630–2672.
4. Omoto S., Fujii Y. R. Cloning and detection of HIV-1-encoded microRNA // *Meth. Mol. Biol.*–2006.–**342**.–P. 255–265.
5. Klase Z., Kale P., Winograd R., Gupta M. V., Heydarian M., Bezzo R., Mc Caffrey T., Kashanchi F. HIV-1 TAR element is processed by Dicer to yield a viral microRNA involved in chromatin remodeling of the viral LTR // *BMC Mol. Biol.*–2007.–**8**.–P. 63–81.
6. Scaria V., Hariharan M., Pillai B., Maiti S., Brahmachari S. K. Host-virus genome interaction: macro roles for microRNAs // *Cell Microbiol.*–2007.–**9**, N 12.–P. 2784–2794.
7. Yeung M., L., Benkirane M., Jeang K.-T. Small non-coding RNAs, mammalian cells, and viruses: regulatory interaction? // *Retrovirology.*–2007.–**4**.–P. 74–79.
8. Hakim S. T., Alsayari M., Mc Lean D. C., Saleem S., Addanki K. C., Aggarwal M., Mahalingam K., Bagasra O. A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses // *Biochem. and Biophys. Res. Commun.*–2008.–**369**, N 2.–P.357–362.
9. Okano K., Vanarsdall A. L., Mikhailov V. S., Rohrman G. F. Conserved molecular systems of the baculoviridae // *Virology.*–2006.–**344**, N 1.–P. 77–87.
10. Kozlov E. A., Vudmaska M. I., Bobrovskaja M. T., Shirina T. V. Investigation of structure-function interaction of polyhedron of *Bombyx mori* Nuclear Polyhedrosis Virus (NPV) with polyhedra protease and RNA. 3. Polyhedras contain small RNA // *Biopolymers and Cell.*–2007 – **23**, N 4.–P. 301–306.
11. Liu S., Xia Q., Zhao P., Cheng T., Hong K., Xiang Z. Characterization and expression patterns of let-7 microRNA in the silkworm (*Bombyx mori*) // *BMC Develop. Biol.*–2007.–**25**, N 7.–P. 88–104.
12. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction // *Nucl. Acids Res.*–2003.–**31**, N 13.–P. 3406–3415.
13. Helvik S. A., Snove O., Saetrom P. Reliable prediction of Drosha processing sites improves microRNA genes prediction // *Bioinformatics.*–2007.–**23**, N 2.–P. 142–149.
14. Tong C., Jin Y., Zhang Y. Computational prediction of microRNA genes in silkworm genome // *J. Zhejiang Univ. SCIENCE B.*–2006.–**7**, N 10.–P. 806–816.
15. Xue C., Li F., He T., Liu G. P., Li Y., Zhang X. Classification real and pseudo microRNA-precursor using local structure-sequence features and support vector mashine // *Bioinformatics.*–2005.–**6**, N3.–P. 310–316.
16. Lim L. C., Lau N. C., Weinstein E. G., Abdelhakim A., Yekta S., Rhoades M. W., Burge C. B., Bartel D. P. The microRNAs of *Caenorhabditis elegans* // *Genes Develop.*–2003.–**17**, N 8.–P. 991–1008.
17. Ooi B. G., Miller L. K. Transcription of the baculovirus polyhedrin gene reduces the levels of an antisense transcript initiated downstream // *J. Virol.*–1990.–**64**, N 6.–P. 3126–3129.
18. Liu A., Qin J., Rankin C., Hardin S. E., Weaver R. F. Nucleotide sequence of a portion of the *Autographa californica* nuclear polyhedrosis viruses genome containing the *EcoRI* site-rich region (*h2_s*) and open reading frame just 5' *p10* gene // *J. Gen. Virol.*–1986.–**67**, N 11.–P. 2565–2570.
19. Lee J., Kim M., Han J., Yeom K.-H., Lee S., Back S. H., Kim V. N. MicroRNA genes are transcribed by RNA polymerase II // *EMBO J.*–2004.–**23**.–P. 4051–4060.
20. Fujita S., Iba H. Putative promoter regions of microRNAs genes involved in evolutionarily conserved regulatory systems among vertebrates // *Bioinformatics.*–2008.–**24**, N 3.–P. 303–308.
21. Li S.-C., Shiau C.-K., Lin W.-C. Vir-MiR db: prediction of viral microRNA candidate hairpins // *Nucl. Acids Res.*–2008.–**36**, N 1.–P. 184–189.

УДК 577.214:578.841
Надійшла до редакції 09.06.08