

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 577.212.3:593.161.13

ТРИ ТИПА СТРУКТУРНОЙ ОРГАНИЗАЦИИ  
КРИПТОГЕНОВ ТРИПАНОСОМАТИД

© 2012 г. Е. С. Герасимов\*, Н. С. Ефимова, А. А. Колесников\*\*

Биологический факультет Московского государственного университета им. М.В. Ломоносова, Москва, 119992

Поступила в редакцию 05.12.2011 г.

Принята к печати 13.01.2012 г.

Проанализирована структура некоторых криптогенов у неизученных ранее представителей гомоксенозных трипаносоматид разных филогенетических групп. Обнаружены новые случаи редукции длины редактируемых доменов *COIII* и *A6*. Сравнительный анализ позволяет сгруппировать три принципиально различных варианта организации криптогенов, положив в основу деления не только эволюционное изменение длины редактируемого домена, но и степень консервативности его нуклеотидной последовательности. Обсуждаются потенциальные факторы, определяющие варианты организации криптогена. Показана возможность альтернативного редактирования мРНК криптогена *rps12* *Wallaceina* sp. Wsd.

**Ключевые слова:** редактируемый домен, криптоген, трипаносоматиды, ретропозиционная модель, альтернативное редактирование.

THREE PATTERNS OF TRYPANOSOMATID CRYPTOGENE STRUCTURAL ORGANIZATION, by E. S. Gerasimov\*, N. S. Efimova, A. A. Kolesnikov\*\* (Department of Molecular Biology, Moscow State University, Moscow, 119991 Russia; \*e-mail: jalgard@yandex.ru; \*\*aak330@yandex.ru). We sequenced a number of cryptogenes from previously unstudied species of homoxenous trypanosomatids belonging to the different phylogenetic groups and found new examples of editing domain length reduction for *A6* and *COIII*. The comparative analyzes of sequences allows to divide the cryptogenes in three groups (patterns) according to the degree of primary structure conservation and editing domain length variation. We discuss the possible factors which influence the cryptogene's structure and evolutionary behavior. Also we demonstrate alternative editing of *rps12* transcript in *Wallaceina* sp. Wsd.

**Keywords:** editing domain, cryptogene, trypanosomatids, retroposition model, alternative editing.

Кинетопластная ДНК трипаносоматид представлена сетью взаимно катенированных молекул двух типов — максикольцевых и миникольцевых [1, 2]. Максикольцевые ДНК имеют длину около 30 т.п.н., в кинетопласте они представлены примерно 50 копиями. Размер миникольцевых ДНК обычно не превышает 700–2500 п.н. в зависимости от вида. Содержание миникольцевых ДНК в кинетопласте достигает нескольких тысяч копий. У трипаносоматид большинство максикольцевых генов, кодирующих митохондриальные белки, не содержат полноценных открытых рамок считывания (ОРС) и называется криптогенами. Их транскрипты подвергаются уридиловому редактированию, которое осуществляется сложным белковым комплексом — эдитосомой. Локализация и число вставляемых или вырезаемых в процессе

редактирования уридиловых оснований определяется гидовыми РНК (гРНК) [3–5]. Большинство генов гРНК находится на миникольцах, однако известны гРНК, кодируемые максикольцевыми ДНК [6, 7].

При делении трипаносоматид миникольца стохастически распределяются между дочерними клетками [8], в результате чего могут утрачиваться редкие классы миникольцев, а вместе с ними и гены отдельных видов гРНК [9, 10]. Так, у штамма *Leishmania tarentolae* UC, более 40 лет культивируемого в лабораторных условиях, отсутствуют приблизительно 30 классов миникольцев, в результате чего криптогены, редактируемые с участием утраченных гРНК, остаются неотредактированными. Клетки недавно выделенного штамма LEM125, а также длительное время культивируемого штамма *L. mexicana amazonensis* содержат миникольца этих классов [11, 12]. Предполагается, что отсутствие отдельных классов миникольцев возможно потому, что отредактированные продукты рья

Принятые сокращения: РД — редактируемый домен; гРНК — гидовая РНК; ОРС — открытая рамка считывания; АР — альтернативное редактирование.

\* Эл. почта: jalgard@yandex.ru; \*\* aak330@yandex.ru

криптогенов не нужны на промастиготной стадии цикла. Однако штамм *L. donovani*, также утративший часть классов миниколец, способен проходить полный жизненный цикл [13].

Редактирование криптогена происходит полярно — от 3'- к 5'-концу мРНК, потому что сайты, комплементарные якорным участкам последующих гРНК, создаются в результате редактирования предыдущими [14]. Участок пре-мРНК, редактируемый набором гРНК, называется редактируемым доменом (РД). Сравнительный анализ структур криптогенов показал, что длина РД может уменьшаться с 3'-конца [15–17]. Например, редактируемые по всей длине криптогены *ND7*, *A6* и *COIII Trypanosoma brucei* становятся 5'-редактируемыми у *L. tarentolae* и *Crithidia fasciculata*, причем *ND7* содержит два независимых РД, длина которых уменьшается одновременно. Редукция длины РД *ND8* выявлена недавно у эндосимбионт-содержащего вида *Strigomonas oncopelti* [18]. Дальнейшее укорачивание РД *ND8* характерно для представителей родственной эндосимбионт-содержащим видам филогенетической группы, которую мы будем называть группа 1. Редактирование *ND8* в этой группе представляет собой встраивание всего 18 уридиловых оснований [19].

Согласно ретропозиционной модели, уменьшение длины РД обусловлено замещением полностью редактируемого криптогена частично отредактированной мРНК [8, 14, 15]. Показано, что митохондрии трипаносоматид содержат обратнотранскриптазную активность [20]. Предполагается, что редкие ретропозиции в одну из 50 копий максикольца закрепляются, если за ними следуют описанные выше потери классов миниколец, которые кодируют соответствующие гРНК, и обычно происходят чаще. Наши результаты изучения криптогена *ND8* согласуются с ретропозиционной моделью, указывая как минимум на две ретропозиции в ходе эволюции этого криптогена: один раз у общего предка эндосимбионт-содержащих видов и группы 1, а второй раз — в группе 1 [19].

Показано, что *COIII T. brucei* подвергается альтернативному редактированию (АР), причем нуклеотид в определенной позиции пре-мРНК *COIII* может взаимодействовать с альтернативной гРНК, предотвращая узнавание последующими гРНК [21–23]. В результате образуется частично отредактированный транскрипт, который содержит новую рамку считывания. Альтернативный белок АЕР-I локализуется в области кинетопластного ассоциата. АР 3'-части транскриптов происходит и у криптогенов *G5* и *ND9* [11, 24]. АР выявлено и у представителей других систематических групп [25].

С целью расширения наших знаний о структуре криптогенов трипаносоматид и особенно у представителей группы 1, в которой наблюдается

редукция длины РД *ND8*, в представленной работе изучены криптогены ряда представителей филогенетической группы 1 и так называемой “слабодивергирующей” группы [26] (в эту группу входят *Leishmania*, *Crithidia*, *Leptomonas*, *Wallaceina*). В результате сравнения нуклеотидных последовательностей криптогенов, как установленных в нашей работе, так и представленных в базах данных, мы предлагаем выделить три типа структурной организации криптогенов, а также обсуждаем возможные причины, определяющие структуру их РД.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Культуры** гомоксенных жгутиковых простейших *Lep. collosoma* ATCC 30261, *Lep. seymouri* ATCC 30220, *Lep. rigidus*, *Lep. pyrrocoris*, *Wallaceina* sp. Wsd, *W. podlipaevi*, *Herpetomonas pessoai*, *H. muscarum*, *Angomonas deanei* любезно предоставлены сотрудниками Зоологического института РАН (Санкт-Петербург). Клетки трипаносоматид растили при 26°C на среде Brain Heart Infusion (“Becton, Dickinson and Company”) с добавлением гемина (10 мкг/мл).

**Кинетопластную ДНК выделяли** из 50 мл культуры на 7–10-й день согласно [27]. Суммарную РНК выделяли с использованием набора RNeasy Mini Kit по протоколу фирмы “QIAGEN”. Обратную транскрипцию проводили следующим образом: к 3 мкл препарата РНК добавляли 2.5 мкл праймера GATACCTGCGTTGTCACTCTACTTTTTTTTTTTTTTTT (10 пмоль/мкл) и прогревали смесь при 70°C в течение 10 мин, а затем охлаждали на льду. Смесь (20 мкл) для обратной транскрипции содержал 25 пмоль праймера, 1 мМ каждого dNTP, 1× буфер для обратной транскрипции (“Fermentas”, Литва), 1 ед. ингибитора РНКаз, 2 ед. обратной транскриптазы RevertAid RT (“Fermentas”). Реакцию проводили при 40°C в течение 1 ч, затем смесь прогревали (70°C, 10 мин), разбавляли в 4 раза и использовали далее в ПЦР с праймеров CCAATACCTTCAACTCTATTATCAT и GATACCTGCGTTGTCACTCTACT для получения мРНК криптогена *rps12*. С учетом присутствия бетаина в используемом наборе для ПЦР, температура ренатурации ( $T_{np}$ ) данной пары праймеров составила 52°C.

ПЦР-продукты, соответствующие изучаемым участкам максикольца, получали, используя следующие праймеры:

1) Криптогены *cyb* и *A6 Lep. collosoma*, *Lep. rigidus*, *W. sp. Wsd*, *W. podlipaevi*, *Lep. pyrrocoris*: JCO3F1 – GTTTATATTGACATTTTGTTCGATTG и JA6R1 – AACAAAAACAAAATAAAAAGGCCAT;  $T_{np} = 52^\circ\text{C}$ .

2) Криптоген *COIII Lep. collosoma*, *Lep. rigidus*, *W. sp. Wsd*, *W. podlipaevi*, *Lep. pyrrocoris*: JND7F1 –

TTTGACATGCGTATGAGT и JCO3R1 – AT-CAATAAAATACAAAGCCATAATCACA;  $T_{пр} = 52^{\circ}\text{C}$ .

3) Криптоген *ND7* всех видов лейшманий: 9S5F – TTAAGCTTGTGCCCACCATTCTTTGTAATAAAGACAACGTGCAGT и s960com – TTGGATCCAYTGUTCWACWSTTTTATATTCRCATAAСТTTCTGTACC; для других видов WND8J4 – GGAATCTTTTTTTGTAATATTTT и s960com – TTGGATCCAYTGUTCWACWSTTTTATATTCRCATAAСТTTCTGTACC.  $T_{пр} = 51^{\circ}\text{C}$ .

4) Криптоген *rps12* и участок *G5* для всех видов: ND4F6 – TTGGTRTTAGYATATGAAGYTC и ND5R6 – CCATARAWRCANAYRCTWAAAC, кроме *Lep. rigidus*. В этом случае использовали ND5R1 в качестве второго праймера: TCTAAAATATATAACWAAATTTTCATAYAAA.  $T_{пр} = 48^{\circ}\text{C}$ .

5) Участок *G4* для всех видов: COIF7 – TTTAAGCTTCCWATWAKWGARAGTTCTAATC и ND4R4 – SAYAYATAWAWTSCYAAATAYAAAATATCAAAAWWAC. Для *Lep. pyrrocoris* использовали праймер ND4R5 – TTTAAGCTTTCAAAAYAWATCATAAAC, для *W. sp. Wsd* и *Lep. collosoma* – COIF8 – TTTGGATCCAAATCATTTTATGAGAAAC и ND4R5.  $T_{пр} = 47^{\circ}\text{C}$ .

ПЦР-амплификацию проводили, используя наборы GenePak PCR Core Kit (“ИзоГен”, Россия). Условия ПЦР были следующими: первичная денатурация –  $94^{\circ}\text{C}$ , 3 мин; 27 циклов денатурации при  $94^{\circ}\text{C}$  в течение 30 с; ренатурация праймеров при  $T_{пр}$  (указана для каждой пары выше) в течение 45 с; элонгация –  $65^{\circ}\text{C}$ , 1 или 2 мин (в зависимости от длины предполагаемых продуктов реакции). После электрофореза ПЦР-продукты выделяли из агарозного геля, используя набор Gel extraction kit (“QIAGEN”) по протоколу, рекомендованному фирмой. Молекулярное клонирование проводили с использованием набора Instclone PCR Cloning Kit (“Fermentas”) и штамма *Escherichia coli* XL1Blue. Плазмиды выделяли при помощи набора QIAPrep Spin Miniprep Kit (“QIAGEN”) по протоколу производителя. Секвенирование молекулярных клонов и очищенных ПЦР-продуктов проводили на секвенаторе ABI Prism 3100 (“Applied Biosystems”). Полученные нуклеотидные последовательности анализировали, используя программы MEGA5, BioEdit, участки *G4*, *G5* и криптогена *rps12* выравнивали при помощи программы MAFFT.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

### *Криптогены A6 и COIII имеют разную длину редактируемого домена*

Длина РД у криптогена *COIII* *Lep. pyrrocoris* и *W. podlipaevi*, представителей “слабодивергирующей” группы, оказалась такой же, как у криптогена *L. tarentolae* из той же группы. У *Lep. rigidus*, *Lep. collosoma*, *Wallaceina* sp. *Wsd*, входящих в фи-

логенетическую группу 1, отсутствует необходимость в редактировании *COIII*. У них *COIII* является геном, содержащим полную ОРС, включая инициаторный кодон, который находится в таком же положении, как и инициаторные кодоны в мРНК других видов (результаты представлены на рис. 1б).

Аналогичным образом укорачивался РД у криптогена *A6*. У *Lep. pyrrocoris* и *W. podlipaevi* длина РД *A6* примерно такая же, как у других представителей “слабодивергирующей” группы (рис. 1а). У представителей группы 1 РД редуцирован до короткого участка на 5'-конце, сразу за которым располагается предсуществующий на уровне ДНК гипотетический инициаторный кодон (рис. 1а). По структуре РД *A6* напоминает минимально редактируемый ген *ND8*, найденный нами ранее у тех же представителей группы 1 [19]. В обоих случаях в результате редактирования изменяются лишь несколько аминокислотных остатков в слабokonсервативной N-концевой области белка. Криптогены еще до редактирования содержат основную часть ОРС и предсуществующий инициаторный кодон.

### *Криптогены ND7, sub, MURF2 с консервативной структурой короткого редактируемого домена*

Нуклеотидные последовательности РД криптогенов *sub* и *ND7* оказались практически идентичными у всех изученных видов, принадлежащих как к группе 1, так и к “слабодивергирующей” группе. Множественное выравнивание редактируемых доменов криптогенов *sub* и *ND7* показано на рис. 2а,б. Оба криптогена имеют короткие консервативные РД. Приведенные множественные выравнивания построены с учетом того, что при взаимодействии мРНК с гРНК наравне с канонической парой G : C образуется и пара G : U, поэтому транзиции A → G или C → T на уровне РД криптогена оказываются нейтральными. Следовательно, у всех известных видов, включая трипаносом (у которых найдены лишь два–четыре отличия), криптоген *sub* имеет одинаковую структуру и длину РД. Длина РД *ND7* отличается только у представителей трипаносом, у которых этот криптоген редактируется полностью. У *L. tarentolae* (GenBank L07537) и *Crithidia fasciculata* [28], а также у всех изученных нами представителей гомоксенных трипаносоматид РД *ND7* состоит из двух коротких практически идентичных РД (рис. 2б). Идентичным оказывается также участок, непосредственно прилегающий к 3'-РД, который взаимодействует с 5'-концевой частью гРНК (на рис. 2б показан после значка “\*”), обеспечивая узнавание РД.

На рис. 3в приведено множественное выравнивание гена гРНК *gMURF2-II*, отвечающего за 5'-редактирование короткого РД криптогена



<i>a</i>		<i>б</i>	
Ltar	AGCGG AGA <b>GAA</b> AAGAAAAGG CTTT	Ltar	<b>AT</b> AAATATAAAAAAG* <b>ACA</b> CTTGTAT
Ltur	AGCGG AGA <b>GGA</b> AAGAAAAGG CTTT	Ltur	<b>ACA</b> AAATATAAAAAAG* <b>GCA</b> CTTGTAT
Lara	AGCGG AGA <b>GGA</b> AAGAAAAGG CTTT	Lara	<b>G</b> AAAATATAAAAAAT* <b>GCA</b> CTTGTAT
Lcha	AGCGG AGA <b>GGA</b> AAGAAAAGG CTTT	Lcha	<b>A</b> AAAATATAAAAAAG* <b>ACA</b> CTTGTAT
Lmex	AGCGG AGA <b>GAA</b> AAGAAAAGG CTTT	Lmex	<b>A</b> AAAATATAAAAAAG* <b>ACA</b> CTTGTAT
Lpyr	AGCGG AGA <b>AAG</b> AAGAAAAGG CTTT	Lpyr	<b>AT</b> AAATATAAAAAAG* <b>GCA</b> TTTGTAT
Wwsd	AGCGG AGA <b>GAG</b> AAGAAAAGG CTTT	Wwsd	<b>C</b> AAAATATA <b>GAAG</b> * <b>ACA</b> TTTATAT
Lrig	AGCGG AGA <b>GAG</b> AAGAAAAGG CTTT	Lcol	<b>T</b> AAAA <b>C</b> ATAAAAAAG* <b>GCA</b> TTTGTAT
Lcol	AGCGG AGA <b>GAG</b> AAGAAAAGG CTTT		
Wpod	AGCGG AGA <b>AAG</b> AAGAAAAGG CTTT		
Adea	AGCGG AGA <b>GAG</b> AAGAAAAGG CTTT		
Tbru	AGCGG <b>TAG</b> <b>AAA</b> AAGAAA <b>GGGT</b> CTTT		
Tbor	AGCG <b>A</b> <b>CGGGGG</b> GAGAAAAGG <b>T</b> CAGT		

<i>в</i>	
Ltar	<b>ATTAGTCTCAGT</b> <b>C</b> AAA <b>GT</b> ATAAAAA <b>TAAG</b> AT <b>T</b> AAAT <b>T</b> AAAA <b>TAAT</b> TAAATAT <b>TT</b>
Lsey	<b>ATAAAGCTCAGT</b> <b>C</b> AAAA <b>T</b> ATAAAAA <b>TAGA</b> AT <b>T</b> AAAT <b>T</b> AAAA <b>TAAT</b> TAAATAT <b>GT</b>
Lpyr	<b>GTCGGACTCAGT</b> <b>C</b> AAAA <b>T</b> ATAAAAA <b>TAAAT</b> TAAAT <b>T</b> AAA <b>GTA</b> GT <b>T</b> AAAT <b>AAAG</b>
Wwsd	<b>ATAAAATAAAAT</b> <b>T</b> AAAA <b>C</b> ATAAAAA <b>TAAAT</b> TAAAT <b>T</b> AAA <b>GCAAT</b> <b>C</b> AAA <b>CATAA</b>
Lcol	<b>TTTATTTTAAA</b> <b>C</b> AAAA <b>C</b> ATAAAAA <b>TAAAT</b> TAAAT <b>T</b> AAAA <b>TAGC</b> TAAAT <b>ATA</b>
Lrig	<b>ATTATATTA</b> <b>A</b> <b>C</b> AAAA <b>C</b> ATAAAAA <b>TAAAT</b> TAAAT <b>T</b> AAAA <b>TAGT</b> TAAAT <b>ACA</b>
Hmus	<b>ATGCGATTCAGT</b> <b>T</b> AAAA <b>C</b> ATAAAAA <b>TAAAT</b> <b>C</b> AAAT <b>T</b> AAA <b>GTA</b> GT <b>C</b> AAAT <b>GTAA</b>
Hpes	<b>ACGCGTTCG</b> <b>C</b> AGT <b>C</b> AAAA <b>T</b> ATAAAAA <b>TAAAT</b> TAAAT <b>T</b> AA <b>GATAAT</b> TAAAT <b>ATAA</b>

**Рис. 2.** Выравнивание редактируемых доменов (РД) криптогенов *cyb* (*a*), *ND7* (*б*) и гена *gMURF2-II* (*в*), кодирующего рРНК, соответствующую короткому РД *MURF2*. Wwsd – *Wallaceina* sp. Wsd, Lcol – *Lep. collosoma*, Lrig – *Lep. rigidus*, Lpyr – *Lep. pyrrocoris*, Wpod – *W. podlipaevi*, Adea – *A. deanei*, Tbru – *T. brucei*, Tbor – *T. borreli*. Ltar – *L. tarentolae*, Ltur – *L. turanica*, Lmex – *L. mexicana*, Lara – *L. arabica*, Lcha – *L. chagasi*, Hmus – *H. muscarum*, Hpes – *H. pessoai*. Идентичные позиции обозначены белыми буквами, различающиеся – черными. Поскольку при образовании дуплекса рРНК-мРНК наряду с канонической парой G : C формируется пара G : U, замены A → G и C → T на уровне РД или гена рРНК, показанные полужирным курсивом, рассматриваются как идентичные позиции.

тельностью *gMURF2-II* также свидетельствует в пользу крайней консервативности структуры РД *MURF2*. Таким образом, *MURF2* – третий криптоген с очень коротким консервативным РД.

#### **Полностью редактируемые криптогены *rps12*, *G4*, *G5* с вариательной нуклеотидной последовательностью редактируемого домена**

Криптогены *rps12*, *G5*, *G4* не содержат ОРС и полностью редактируются у всех изученных в нашей работе видов трипаносоматид, а также у видов, нуклеотидные последовательности которых представлены в базах данных (*Trypanosoma*, *Leishmania*). Эти криптогены отличаются крайней вариательностью нуклеотидной последовательности. В таблице приведены значения среднего генетического расстояния между участками максикольца, содержащими данные криптогены. Эти расстояния определяли по алгоритму Tajima-Nei в программе MEGA 5.03 [29] для трех выборок, в которые вошли (1) представители рода *Leishmania*; (2) представители “слабодивергирующей” (по 18S рРНК) группы; и (3) “слабодивергирующей” филогенетической группы 1 рода *Trypanosoma* и рода *Herpetomonas*.

Для сравнения приведены средние значения генетического расстояния между неотредактируемым участком криптогена *cyb* и 9S рРНК аналогичной длины. Из таблицы видно, что среднее генетическое расстояние между вариательными полностью редактируемыми криптогенами сильно возрастает, если переходить от сравнения внутри рода *Leishmania* к сравнению представителей разных родов. Однако для других участков максикольца (*cyb*, 9S) такого увеличения расстояния не наблюдается. В целом, у вариательных криптогенов значения среднего генетического расстояния в 2 раза и более превосходят значения для *cyb* и 9S внутри “слабодивергирующей” группы и *cyb*, 9S, *COIII* – в выборке из представителей всех филогенетических групп. Особенно высоким значением среднего генетического расстояния обладает участок *G4*, функция белкового продукта которого на сегодня не установлена, а гомология с известными белками, представленными в базах данных, отсутствует.

#### **Альтернативное редактирование *rps12* *Wallaceina* sp. Wsd**

Клоны мРНК *rps12*, длина которых соответствует отредактированным транскриптам, оказа-

```

rps12 CCAAtACcttCAACtCtAtttAtCattAttA>>ATAATAAtTTGTTTTTTGTCGTAGTTTATATttATA
rps12 I P S T L F I I I N N N L F F C R S .L Y L Y
alt12 CCAAtACcttCAACtCtAtttAtCattAttA>>A AA_AAt_G G CG AG A A ttA A
alt12 I P S T L F I I I K K W R E L S

rps12 TG_GTTTATGtt_GtC_GTTTTTGTtATTATTAGTATATTTTATGTTAGtCCAC_GGTTGCCAAG
rps12 G L C C R F C L F I S I F Y V S P R L P S
alt12 GttG_A_GtttGtCttttG_G A A AG A A A G AGtCCACttGGT_GCCAAG
alt12 W V C L L E R K S P L G A K

rps12 TTCcGGTAAtCGTCGTATTGTTTTA_GCAGt_ATTTTATTATAtAAtTTTGTtTG A GTTATTCGT
rps12 S G N R R I V L A V F Y L Y N F V W V I R
alt12 CCGG_AatCG CG A G AtGCAGtTA ATTTA_AtAA G GTTATTGT ATTcGTT
alt12 P E S R D A V N L I R L F V F V

rps12 GT GTTTTTTGT GT GTTTTTCCGTT G GTTTGTTTTcATTATTATAATGA
rps12 V F F C C V F S V G L F S L F I I E
alt12 GTTTTTGT GTTTTTGTTTTG CCGTTTTGTTGTTG CA ATT A AA GA
alt12 V F V F F V L P F L L F A I K E

rps12 AGGTGGTGGTTTTATTGATTGTCCAGGTTTGAAtTTTACACGtttATATAGTTTTTTGTAG<<
rps12 G G G F I D C P G L K L F T R F I ***
alt12 XGGTGGTGGTTTTATTGATTGTCCAGGTTTGAAtTTTACACGtttATATAAATTTTTTTGTAG<<
alt12 W W F Y W L S R F E I I Y T F Y I I F L ***

rps12 AGGTATTTTAATTAAAA (A) A
alt12 AGGTATTTTAATTAAAA (A) A

```

**Рис. 3.** Альтернативное редактирование криптогена *rps12* у *Wallaceina* sp. Wsd. rps12 – κДНК “нормально” отредактированной мРНК *rps12* и соответствующая ОПС аминокислотная последовательность белка, гомологичного rps12 других видов. alt12 – κДНК “альтернативно” отредактированной мРНК *rps12* и аминокислотная последовательность белка, не имеющего явных гомологов в базах данных. “t” – нуклеотиды, имеющиеся в криптогене; “T” – вставляемые в результате редактирования. Зона, альтернативно отредактированная в клоне alt12, выделена полужирным. Крестом перечеркнут динуклеотид AG, которого нет в клоне alt12. Этот динуклеотид определяет, по-видимому, минорный вариант максикольца, транскрипт которого будет подвергаться альтернативному редактированию.

лись гетерогенными по составу (GenBank JN944131, JN944132). На рис. 3 представлены два варианта отредактированных мРНК: “нормальный” (кодирует белок rps12) и “альтернативный”, его продукт не имеет явных гомологов в базах данных. Альтернативно отредактированные клоны не содержат динуклеотид AG в сайте, с которого начинается AP-участок. Таким образом, поскольку редактирование включает в себя только встраивание или удаление U, можно сделать вывод о гетерогенности кодирующей области максиколецового генома, выявляемой на уровне индивидуальных транскриптов, синтезируемых с максиколец разного класса.

### ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

У эндосимбионт-содержащих видов РД *ND8* редуцируется до 5'-редактируемой формы, а у видов родственной им группы 1 – до формы минимально редактируемого гена [19]. В обоих случаях у видов группы 1 наблюдается совместная редук-

ция криптогенов *A6* и *COIII* до минимально редактируемого криптогена и гена соответственно. У представителей “слабодивергирующей” группы РД этих криптогенов редуцируются до 5'-редактируемой формы. По-видимому, у *A6*, *COIII* и *ND8* ретропозиции происходили, как минимум, дважды. Можно предположить два альтернативных сценария этих событий. По одному из них ретропозиции происходили последовательно при переходе от полностью редактируемого криптогена к минимально редактируемому гену (в группе 1) через промежуточную 5'-редактируемую форму. Согласно другому сценарию, два независимых события привели к переходу от полностью редактируемого криптогена к минимально редактируемому гену в группе 1 и к 5'-редактируемой форме криптогена в “слабодивергирующей” группе (*ND8* у эндосимбионт-содержащих видов).

*A6*, *COIII*, *ND7* и *ND8* полностью редактируются у представителей рода *Trypanosoma*, базального по отношению к другим членам группы трипано-

Среднее генетическое расстояние между нуклеотидными последовательностями, определенное в программе MEGA 5.03 по алгоритму Tajima-Nei\*

Фрагмент максиколяца	Все группы (1)	“Слабодивергирующая” группа (2)	Внутри рода <i>Leishmania</i> (3)
<i>rps12</i>	0.528	0.252	0.113
<i>G5</i>	0.605	0.476	0.140
<i>G4</i>	0.798 (0.680**)	—	—
<i>cyb</i> (^)	0.190	0.172	0.119
<i>COIII</i>	0.268	—	—
<i>9S</i>	0.179	0.137	0.050

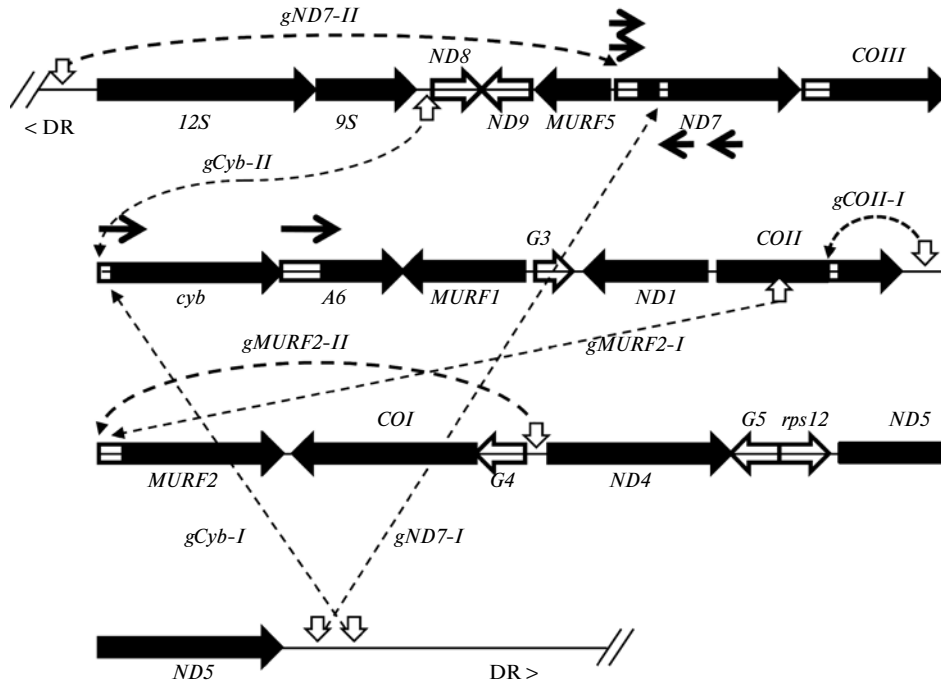
\* Предварительное множественное выравнивание фрагментов выполнено алгоритмом MUSCLE с настройками по умолчанию.

\*\* Взята часть криптогена, не входящая в состав РД.

Примечание. Выборка 1 включает представителей всех групп: *Lep. pyrrocoris*, *W. podlipaevi* (для *9S*, *G4* используется *Lep. seymouri*), *Lep. collosoma*, *Wallaceina* sp. Wsd, *Lep. rigidus*, *L. turanica*, *L. tarentolae* (GenBank acc. no. M10126), *T. brucei* (GenBank acc. no. M94286) и в случае *G4* также включает *H. muscarum* и *H. pessoai*. Выборка 2 представлена видами из группы “слабодивергирующая” – *L. tarentolae* (№ M10126), *L. turanica*, *Lep. pyrrocoris*, *Lep. seymouri* или *W. podlipaevi*, *C. fasciculata* (GenBank acc. no. U44442; X02548). Выборка 3 представлена видами рода *Leishmania* и включает *L. tarentolae*, *L. turanica*, *L. arabica*, *L. chagasi*, *L. mexicana*.

соматид, согласно большинству филогенетических построений. Внутри “слабодивергирующей” группы происходит согласованное укорачивание РД *A6*, *COIII* и *ND7*. Во всех известных нам случаях, за исключением *Trypanosoma*, *ND7* приобретает характерную структуру из двух независимых коротких крайне консервативных РД. Показано, что виды, изученные нами, и виды, нуклеотидные последовательности которых представлены в базах данных, РД *ND7* имеют не только практически одинаковую длину, но и первичную структуру. Структурную организацию, подобную *ND7*, имеют криптогены *cyb* и *MURF2*. Их короткие РД также практически идентичны у видов из разных групп. Подобные результаты получены ранее [30] при изучении *цис*-редактируемого криптогена *COII*, также имеющего короткий консервативный РД. Вероятно, последние три криптогена когда-то могли полностью редактироваться, но у всех современных видов они подверглись редукции РД, как и *ND7*, который все еще полностью редактируется у *Trypanosoma*. Почему эти криптогены не утрачивают РД полностью и остаются консервативными? Мы предполагаем, что основной фактор, определяющий эволюцию криптогенов этой группы, – максиколецевое кодирование гРНК, участвующих в процессе их редактирования. На рис. 4 представлена схема максиколецевого генома, характерного для “слабодивергирующей” группы видов (*L. tarentolae* (GenBank M10126), *C. fasciculata* [28]). На этой схеме участки, кодирующие гРНК, отмечены белыми вертикальными

стрелками, а пунктирными линиями указаны РД-мишени этих гРНК. Видно, что в случае криптогенов с короткими консервативными РД гены гРНК всегда локализованы в максиколецях. Такие гРНК не могут утрачиваться при делении клетки, что не дает преимущества копии максиколеця, в которую происходит ретропозиция. Соотношение числа копий миниколеццевого гена гРНК и редактируемого ею криптогена равно приблизительно 1 : 100, тогда как в случае максиколецевой гРНК это соотношение составляет 1 : 1. Миниколеццевой ген гРНК подвергается многократному дублированию в отличие от криптогена, что позволяет накапливать мутации, не нарушая редактирования. В свою очередь, такие слегка дивергировавшие гРНК могут компенсировать мутации, возникающие в криптогене. Таким образом, кодирование гРНК миниколецевыми генами также повышает вариабельность редактируемых ими участков криптогенов. При этом криптогены с максиколецевыми гРНК остаются очень консервативными. Заметим, что криптогены, у которых происходит редукция длины РД, не имели, насколько нам известно, гРНК, кодируемых максиколецевыми ДНК, за исключением криптогена *ND8*. РД *ND8* не утрачивается полностью, а сокращается до короткого участка на 5'-конце мРНК, который редактируется единственной гРНК, кодируемой максиколецевым геном (по крайней мере, у *Leishmania*) [12]. Следует отметить, что структура короткого РД *ND8* у *Leishmania* оказалась идентичной соответствующему



**Рис. 4.** Карта-схема кодирующей области максиколеця типичного представителя “слабодивергирующей” группы. Гены и криптогены обозначены горизонтальными объемными стрелками, направление которых показывает положение на ДНК. Черным отмечены нередатируемые области, белым в черном контуре — редактируемые домены. Вертикальными стрелками показаны максиколецевые гены гРНК, пунктиром — редактируемые ими криптогены. Горизонтальными черными стрелками обозначены места транскрипции олиго(U)-РНК [31]. DR — дивергентная область.

участку РД полностью редактируемого *ND8* *T. brucei* [19]. В данном случае *ND8* подобен криптогену *ND7*, что свидетельствует в пользу высказанного нами предположения о зависимости структуры криптогена от места кодирования гРНК. С другой стороны, криптоген *A6*, который как и *ND8*, не утрачивает полностью РД, не консервативен, и его гРНК, насколько нам известно, не кодируется максиколецевым геномом. Другим важным фактором, определяющим структуру РД ряда криптогенов, по нашему мнению, могут быть 3'-олиго(U)-гРНК-подобные транскрипты, обнаруженные недавно [31]. На рис. 4 некоторые из этих транскриптов [31] показаны горизонтальными черными стрелками. Их положение полностью совпадает с границами коротких РД *cyb*, *ND7*, *A6*. Функция этих транскриптов остается неизвестной, однако можно предполагать, что они препятствуют ретропозиции частично отредактированной мРНК в данные области максиколеця.

Криптогены с варибельной структурой РД редактируются преимущественно миниколецевыми гРНК, что и определяет их варибельность, особенно по числу остатков тимидина. Поскольку лишь некоторые полностью редактируемые криптогены имеют одну или две гРНК, кодируемых максиколецевыми генами (*G5*, *ND9*, *G3*), и при этом, насколько нам известно, всегда остают-

ся полностью редактируемыми, можно предположить, что в данном случае реверсия не происходит по другой причине. Многие полностью редактируемые криптогены образуют два типа зрелых матриц, которые кодируют разные белки, за счет АР пре-мРНК разными каскадами гРНК. Ретропозиция одной из альтернативно отредактированных матриц в максиколецю невозможна из-за потери одного из белковых продуктов. АР подвергаются полностью редактируемые криптогены *G5* и *ND9* [11]. Нами впервые показано АР *rps12* *Wallaceina* sp. Wsd. До этого было известно о существовании нескольких форм мРНК *rps12* у *L. tarentolae* [32], предположительно возникающих в результате образования неправильного дуплекса гРНК-мРНК. Наши данные говорят о другом механизме: АР подвергаются РНК-копии криптогена, синтезируемые с минорного класса максиколец. Гетерогенность максиколецевой ДНК в дивергентной области показана ранее [33]. Мы предполагаем, что гетерогенность нуклеотидных последовательностей максиколеця может быть характерна и для альтернативно редактируемых участков кодирующей области, что обнаруживается, в частности, при скрининге транскриптов *rps12*. В данной ситуации смену основных классов максиколец [33] можно рассматривать как механизм регуляции соотношения “нормальной” и



“альтернативной” мРНК. Интересно, что полностью редактируемый *COIII T. brucei* также подвергается АР, причем альтернативный белок АЕР-I синтезируется только на амастиготной стадии. У гомоксенных представителей трипаносоматид *COIII* является не криптогеном, а геном, поэтому белок АЕР-I у них не кодируется этим локусом митохондриального генома. Подобным образом могла происходить эволюция криптогена *ND8* — по пути редукции РД [19], когда на каком-то этапе исчезла необходимость в его альтернативных формах, известных, например, у *L. amazonensis* [11].

Работа получила финансовую поддержку Российского фонда фундаментальных исследований (09-04-01177) и государственного контракта (№ 16.740.11.0009).

### СПИСОК ЛИТЕРАТУРЫ

- Shapiro T.A., Englund P.T. 1995. The structure and replication of kinetoplast DNA. *Annu. Rev. Microbiol.* **49**, 117–143.
- Shapiro T.A. 1993. Kinetoplast DNA maxicircles: networks within networks. *Proc. Natl. Acad. Sci. USA.* **90**, 7809–7813.
- Lukes J., Hashimi H., Zikova A. 2005. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr. Genet.* **48**, 277–299.
- Simpson L., Maslov D.A. 1994. RNA editing and the evolution of parasites. *Science.* **264**, 1870–1871.
- Simpson L., Sbicego S., Aphasizhev R. 2003. Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *RNA.* **9**, 265–276.
- Sturm N.R., Simpson L. 1990. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell.* **61**, 879–884.
- Sturm N.R., Simpson L. 1991. *Leishmania tarentolae* minicircles of different sequence classes encode single guide RNAs located in the variable region approximately 150 bp from the conserved region. *Nucl. Acids Res.* **79**, 6277–6281.
- Simpson L., Thiemann O.H., Savill N.J., Alfonzo J.D., Maslov D.A. 2000. Evolution of RNA editing in trypanosome mitochondria. *Proc. Natl. Acad. Sci. USA.* **97**, 6986–6993.
- Maslov D.A., Thiemann O., Simpson L. 1994. Editing and misediting of transcripts of the kinetoplast maxicircle G5 (ND3) cryptogene in an old laboratory strain of *Leishmania tarentolae*. *Mol. Biochem. Parasitol.* **68**, 155–159.
- Thiemann O.H., Maslov D.A., Simpson L. 1994. Disruption of RNA editing in *Leishmania tarentolae* by the loss of minicircle-encoded guide RNA genes. *EMBO J.* **13**, 5689–5700.
- Maslov D.A. 2010. Complete set of mitochondrial pan-edited mRNAs in *Leishmania mexicana amazonensis* LV78. *Mol. Biochem. Parasitol.* **173**, 107–114.
- Gao G.G., Kapushoc S.T., Simpson A.M., Thiemann O.H., Simpson L. 2001. Guide RNAs of the recently isolated LEM125 strain of *Leishmania tarentolae*: an unexpected complexity. *RNA.* **7**, 1335–1347.
- Nebohacova M., Kim C.E., Simpson L., Maslov D.A. 2009. RNA editing and mitochondrial activity in promastigotes and amastigotes of *Leishmania donovani*. *Int. J. Parasitol.* **39**, 635–644.
- Maslov D.A., Simpson L. 1992. The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. *Cell.* **70**, 459–467.
- Landweber L.F. 1992. The evolution of RNA editing in kinetoplastid protozoa. *BioSystems.* **28**, 41–45.
- Simpson L., Maslov D.A. 1999. Evolution of the U-insertion/deletion RNA editing in mitochondria of kinetoplastid protozoa. *Ann. N.Y. Acad. Sci.* **870**, 190–205.
- Колесников А.А., Мерзляк Е.М., Бессолицына Е.А., Федяков А.В., Шониан Г. 2003. Редукция редактируемого домена митохондриального гена *A6* (субъединица 6 АТФазы) у трипаносоматид. *Молекуляр. биология.* **37**, 637–642.
- Aravin A.A., Yurchenko V.Y., Merzlyak E.M., Kolesnikov A.A. 1998. The mitochondrial *ND8* gene from *Crithidia oncopelti* is not pan-edited. *FEBS Lett.* **431**, 457–460.
- Gerasimov E.S., Kostygov A.Y., Shi Yan, Kolesnikov A.A. 2011. From cryptogene to gene? ND8 editing domain reduction in insect trypanosomatids. *Eur. J. Protistol.* doi:10.1016/j.ejop.2011.09.002.
- Gonzalez C.I., Thomas M.C., Martin F., Alami J., Alonso C., Lopez M.C. 1997. Reverse transcriptase-like activity in *Trypanosoma cruzi*. *Acta Tropica.* **63**, 117–126.
- Ochsenreiter T., Anderson S., Wood Z.A., Hajduk S.L. 2008. Alternative RNA editing produces a novel protein involved in mitochondrial DNA maintenance in Trypanosomes. *Mol. Cell Biol.* **28**, 5595–5604.
- Ochsenreiter T., Cipriano M., Hajduk S.L. 2008. Alternative mRNA editing in Trypanosomes is extensive and may contribute to mitochondrial protein diversity. *PLoS One.* **3**, e1566.
- Ochsenreiter T., Hajduk S.L. 2006. Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Rep.* **7**, 1128–1133.
- Read L.K., Wilson K.D., Myler P.J., Stuart K. 1994. Editing of *Trypanosoma brucei* maxicircle CR5 mRNA generates variable carboxy terminal predicted protein sequences. *Nucl. Acids Res.* **22**, 1489–1495.
- Lu B., Wilson R.K., Phreaner C.G., Mulligan M.R., Hanson M.R. 1996. Protein polymorphism generated by differential RNA editing of a plant mitochondrial *rps12* gene. *Mol. Cell Biol.* **16**, 1543–1549.
- Merzlyak E.M., Yurchenko V., Kolesnikov A.A., Alexandrov K., Podlipaev S.A., Maslov D.A. 2001. Diversity and phylogeny of insect trypanosomatids based on small

- subunit rRNA genes: polyphyly of *Leptomonas* and *Blasitocrithidia*. *J. Eukaryot. Microbiol.* **48**, 161–169.
27. Маслов Д.А., Энтелис Н.С., Колесников А.А., Зайцева Г.Н. 1982. ДНК кинетопласта *Crithidia oncopelti*. Рестрикционное картирование максикольцевой ДНК. *Биоорг. химия.* **8**, 676–685.
28. van der Spek H., Arts G.J., Zwaal R.R., van den Burg J., Sloof P., Benne R. 1991. Conserved genes encode guide RNAs in mitochondria of *Crithidia fasciculata*. *EMBO J.* **10**, 1217–1224.
29. Tamura K., Dudley J., Nei M., Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.
30. Golden D.E., Hajduk S.L. 2005. The 3'-untranslated region of cytochrome oxidase II mRNA functions in RNA editing of African trypanosomes exclusively as a cis guide RNA. *RNA.* **11**, 29–37.
31. Madej M.J., Alfonzo J.D., Huttenhofer A. 2007. Small ncRNA transcriptome analysis from kinetoplast mitochondria of *Leishmania tarentolae*. *Nucl. Acids Res.* **1**, 1–11.
32. Sturm N.R., Maslov D.A., Blum B., Simpson L. 1992. Generation of unexpected editing patterns in *Leishmania tarentolae* mitochondrial mRNAs: misediting produced by misguiding. *Cell.* **70**, 469–476.
33. Flegontov P.N., Zhirenkina E.N., Gerasimov E.S., Ponirovsky E.N., Strelkova M.V., Kolesnikov A.A. 2009. Selective amplification of maxicircle classes during the life cycle of *Leishmania major*. *Mol. Biochem. Parasitol.* **165**, 142–152.